

Victor Maia Senna Delgado

Alocação dos alunos nas escolas: uma abordagem de
algoritmos de pareamento para análise do efeito do
cadastro escolar de Belo Horizonte na proficiência dos
estudantes

Belo Horizonte, MG
UFMG/Cedeplar
2013

Victor Maia Senna Delgado

Alocação dos alunos nas escolas: uma abordagem de
algoritmos de pareamento para análise do efeito do
cadastro escolar de Belo Horizonte na proficiência dos
estudantes

Tese apresentada ao curso de doutorado do
Centro de Desenvolvimento e Planejamento
Regional da Faculdade de Ciências
Econômicas da Universidade Federal de Minas
Gerais, como requisito parcial à obtenção do
Título de Doutor em Demografia.

Orientador: Prof. Dr. Eduardo Luís
Gonçalves Rios-Neto

Co-orientador: Prof. Dr. Clodoveu
Augusto Davis Jr.

Belo Horizonte, MG
Centro de Desenvolvimento e Planejamento Regional
Faculdade de Ciências Econômicas - UFMG
2013

Folha de Aprovação

Nota: _____

Prof. Dr. Eduardo Luís Gonçalves Rios-Neto (Orientador)

Nota: _____

Prof. Dr. Clodoveu Augusto Davis Jr. (co-orientador)

Nota: _____

Profª. Dra. Simone Wajnman

Nota: _____

Prof. Dr. José Francisco Soares

Nota: _____

Prof. Dr. Reynaldo Fernandes

Nota: _____

Prof. Dr. Daniel Monte

Dedico essa tese aos meus pais,
Meus irmãos,
Ao meu avô Cid Maia,
E à minha avó Maria Alice Sena Delgado

"Necessitamos sempre de ambicionar alguma coisa que, alcançada, não nos torne sem ambição."
Carlos Drummond de Andrade

QUADRILHA

João amava Teresa que amava Raimundo
que amava Maria que amava Joaquim que amava Lili
que não amava ninguém.

João foi para os Estados Unidos, Teresa para o convento,
Raimundo morreu de desastre, Maria ficou para tia,
Joaquim suicidou-se e Lili casou com J. Pinto Fernandes
que não tinha entrado na história.

Carlos Drummond de Andrade

Andrade, Carlos Drummond de. (1993). *Antologia Poética*. (41ª Ed.). Rio de Janeiro, RJ: Editora Record.

AGRADECIMENTOS

A conclusão desta tese não seria possível sem a ajuda de dezenas de pessoas, muitas delas foram fundamentais para o êxito do trabalho.

Em primeiro lugar agradeço ao meu orientador, prof. Eduardo Luís Gonçalves Rios-Neto, não só pelo que me ensinou ao longo desta tese, mas também pelos ensinamentos de todo período de curso no doutorado, desde a participação na banca de mestrado. Agradeço a confiança em mim depositada e gostaria de destacar sua agudeza de raciocínio para problemas relevantes, inteligência e uma criatividade inigualável. Essas são apenas algumas das características positivas do prof. Rios-Neto, e não posso esquecer de acrescentar o excelente bom humor e a dedicação nas horas extras de fins de semana, que me ajudaram a continuar no caminho.

Agradeço ao prof. Clodoveu Augusto Davis Jr. por ter aceitado essa empreitada, foi muito bom poder contar com sua experiência, solicitude e incentivo, quem dera eu saber um pouquinho mais dessa fascinante área de ciência da computação para fazer um trabalho à altura dos seus demais orientandos, mas espero que eu tenha conseguido atingir os objetivos que foram propostos com esse trabalho. Não posso deixar de agradecer ao Douglas Martins, competente aluno de graduação do professor Davis Jr., o qual eu não tive oportunidade de conhecer e agradecer pessoalmente. Sem o trabalho do Douglas a parte georreferenciada do meu banco seria quase impossível.

Ao professor Chico Soares (como ele gosta de ser chamado), que ao longo dos nossos trabalhos sobre educação na FJP foi quase um segundo orientador. O admiro muito e aqui vai meu muito obrigado ao mestre. Ao professor Irineu Rigotti, por toda ajuda nas etapas do projeto.

Também devo agradecer à Vânia Cristina Liberato da Superintendência de Informações Educacionais e à Maria Inez Barroso Simões da Superintendência de Avaliação Educacional da SEE-MG. Bem como Leila e Vânia Cândido, que prestaram ajuda preciosa para obtenção e trabalho com o banco de dados. Agradeço Denise Nascimento Souza e equipe da diretoria de cadastro da SME-BH pela ajuda com os dados do cadastro.

Ao longo deste trabalho passei por diversas instituições. Agradeço à CAPES pela a bolsa no primeiro ano de doutorado, à equipe da Fundação João Pinheiro-MG nos três anos e meio seguintes. O aprendizado e as amizades de lá foram essenciais para meu desenvolvimento. Em particular, agradeço ao Glauco Umbelino pela ajuda com as coordenadas geográficas de GMS para UTM, muito obrigado, e à amiga Taynãna Simões, pela dica de uso do software Terra View.

Também muito obrigado à Maria Izabel, Adriana, Juliana Riani (duas vezes, por ajudar na banca do projeto e nos trabalhos sobre educação), Juliana Chiari, Paulo César, Giovanni Caixeta, Geovane Máximo, Louis, Wladia, Marcela, Lauren, Roberto Marinho, Maria Luiza, João Batista, Danielle, Luiza, Fernando Prates, Fernando Rodrigues, Olinto, Fátima Beatriz, Carolina Portugal, Murilo Fahel, Bruno Cabral, Bruno Lazzarotti, Agnez, Carla, Silvio Ferreira, Cláudia, Rútila, Bruna e Cláudio Burian, aos colegas de IMRS, de MSF e Observatório da Juventude. Gostaria de nominar a todos os demais colegas, por favor, quem, por ventura, eu tenha esquecido sintam-se representados na figura dos colegas acima elencados. Ao Julio Alfredo Racchumi Romero pela ajuda com o projeto de tese e disponibilidade na banca de defesa do projeto.

Não posso deixar de mencionar o apoio da UFOP e dos vários colegas professores dessa instituição nessa reta final do trabalho. E também ao colegiado da demografia no CEDEPLAR que ajudou incentivando e cobrando os prazos que são necessários, especialmente à professora Simone Wajnman. Aos demais professores do centro, ao Sebastião Guedes, Maristela Dória e a todo pessoal da secretaria de pós-graduação e secretaria geral, o meu muito obrigado.

Agradeço imensamente o apoio da minha família e da minha noiva Lia. Ao longo de todo o doutorado, e especialmente os últimos 6 meses foram muito de trabalho e abnegação do ritmo normal de vida. Sei que para ela e para a família não é fácil estar com essa condicionante de uma tese que tem que ser concluída para agendarmos o casamento. Por tudo isso segue o meu muito obrigado pelo seu amor e por sua paciência. Espero que tenha valido a pena não só para mim como para todos nós.

Por último, mas não menos importante, Aos amigos, colegas de doutorado e amigos de Brasília, que mesmo de longe sempre mandaram muito incentivo e boas vibrações, meu Muito Obrigado!

LISTA DE ABREVIATURAS

ANPEC	-	Associação Nacional de centros de Pós Graduação em Economia.
CEMIG	-	Companhia Energética de Minas Gerais.
DA	-	<i>Deferred Acceptance</i> (algoritmo).
EBTC	-	Empresa Brasileira de Correios e Telégrafos.
ENEM	-	Exame Nacional do Ensino Médio.
GCI	-	<i>Geocoding Certainty Indicator</i> .
ICC	-	<i>Interclass Correlation Coefficient</i> (estatística).
IES	-	Instituições de Ensino Superior.
INEP	-	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
MOE	-	<i>Ministry of Education</i> (singapura).
NRMP	-	<i>National Resident Matching Program</i> .
NSE	-	Nível Socioeconômico (variável).
NYCDOE ou DOE	-	<i>New York City Department of Education</i> .
PBH	-	Prefeitura Municipal de Belo Horizonte.
PISA	-	Programme for International Student Assessment.
PROEB	-	Programa de Avaliação da Rede Pública de Educação Básica.
PRODABEL	-	Empresa de Informática e Informação do Município de Belo Horizonte.
PRODEMGE	-	Companhia de Tecnologia da Informação do Estado de Minas Gerais.
PSLE	-	<i>Primary School Leaving Exam</i> (Singapura).
SAEB	-	Sistema de Avaliação da Educação Básica.
SEE-MG	-	Secretaria Estadual de Educação de Minas Gerais.
SiSU	-	Sistema de Seleção Unificada.
SIMAVE	-	Sistema Mineiro de Avaliação da Educação Pública.
SME-BH	-	Secretaria Municipal de Educação de Belo Horizonte.
TTC	-	<i>Top Trading Cycles</i> (algoritmo).
TRI	-	Teoria de Resposta ao Item.
UEH	-	Unidade Espacial Homogênea.

ÍNDICE DE TABELAS

Tabela 1 – Quais fatores são considerados mais importantes pelos pais	53
Tabela 2 - Resultado de um exemplo de solução estável com homens propondo	95
Tabela 3 - Resultado de um exemplo de solução estável com mulheres propondo	95
Tabela 4 - Exemplo com 4 homens e 4 mulheres, solução estável com mulheres propondo.....	96
Tabela 5 - Exemplo com 4 homens e 4 mulheres, solução estável com homens propondo.....	96
Tabela 6 - Comparando dois algoritmos DA e TTC de acordo com Estável e Pareto	120
Tabela 7 - Resultado Social da troca de dois alunos i1 e i2 entre as escolas s1 e s2.....	141
Tabela 8 - Variáveis do Censo Escolar e do PROEB (nome, tipo e descrição)	150
Tabela 9 - Variáveis no banco de dados de Escolas (nome, tipo e descrição)	152
Tabela 10 - Informações sobre os bancos de dados utilizados	154
Tabela 11 - Informações após a fusão dos bancos de dados por nível de corte do agrep	157
Tabela 12 - Categorias do Nivel Socioeconômico das escolas (NSE)	158
Tabela 13 - Hierarquia de comunicação de endereços comumente utilizada no Brasil	159
Tabela 14 - Valores de corte do GCI, número de alunos e perda resultante	164
Tabela 15 - Presença de bairros no banco de dados.....	170
Tabela 16 - Grupo de bairros com número de alunos moradores vagas e Escolas	176
Tabela 17 - Conjunto de Variáveis do banco por nível de análise	178
Tabela 18 - Média e proporção de alunos nos níveis de proficiência BH e MG	179
Tabela 19 - Resumo estatístico das principais variáveis associadas aos alunos.....	181
Tabela 20 - Matriz de correlação das variáveis do banco, nível 1	182
Tabela 21 - Regressão linear multivaria da proficiência de alunos (nível 1)	184
Tabela 22 - Resumo estatístico das principais variáveis associadas à escola	185
Tabela 23 - Regressão linear multivaria da proficiência nas escolas (nível 2).....	186
Tabela 24 - Proporção de alunos por distância em Km.....	190
Tabela 25 - Resumo das principais estatística da distância.....	192
Tabela 26 - Proporção de alunos por posição da distância da escola	194
Tabela 27 - Análise da variância Intra níveis e Entre níveis	196
Tabela 28 - Saldo “migratório” por regiões, alunos que se movimentam em BH	198
Tabela 29 - Regiões que mais recebem alunos em relação ao total de alunos permanentes.....	199
Tabela 30 - Regiões que mais perdem alunos em relação ao total de alunos permanentes	199
Tabela 31 - Regressão linear simples entre proficiência e distância.....	201
Tabela 32 - Regressão linear simples entre proficiência e log da distância	201
Tabela 33 - Número de migrantes por tipo de escola de acordo com nível de proficiência.....	202
Tabela 34 - Regressão linear multivariada da proficiência com migrante de região.....	203
Tabela 35 - Diferença na proficiência por migração de bairro em um raio de 2 km	205
Tabela 36 - Regressão linear multivariada da proficiência com migrante de região.....	206
Tabela 37 - Modelos Hierárquicos Nulo, modelo 6	211
Tabela 38 - Modelos Hierárquicos de 4 níveis Modelos 7 e 8.....	215
Tabela 39 - Valores dos parâmetros estimados para uso nas simulações	225
Tabela 40 - Proporção de alunos por distância em Km Antes e depois da simulação	235
Tabela 41 - Diferença das proficiências entre a amostra e a simulação do baseline	235
Tabela 42 - Diferença das distâncias entre a amostra e a simulação por distâncias	237
Tabela 43 - Diferença das proficiências entre a amostra e a simulação por distâncias	242
Tabela 44 - Proporção de alunos por distância em Km Antes e depois da simulação μ^1.....	242
Tabela 45 - Diferença das proficiências entre a amostra e a simulação por proficiência.....	245
Tabela 46 - Diferença de proficiência entre amostra e simulação preferências lexicográficas	248

Tabela 47 - Diferença de proficiência entre amostra e simulação preferências com utilidade	252
Tabela 48 - Proporção de alunos por distância em Km Antes e depois da simulação $\mu_1 - \mu_4$	253
Tabela 49 - Resultados das simulações de políticas públicas	257
Tabela 50 - Decomposição de Oaxaca para o diferença de proficiência entre migrantes	260

ÍNDICE DE FIGURAS

Figura 1 - Exemplo simples de puras Trocas do TTC (um ciclo simples)	111
Figura 2 - Exemplo 7 de TTC com 4 alunos e 4 escolas (1º passo)	113
Figura 3 - Exemplo 7 de TTC com 4 alunos e 4 escolas (2º passo)	114
Figura 4 - Exemplo 7 de TTC com 4 alunos e 4 escolas (3º passo)	114
Figura 5 - Exemplo 8 de TTC com 8 alunos e 4 escolas (1º passo)	115
Figura 6 - Exemplo 8 de TTC com 8 alunos e 4 escolas (2º passo)	116
Figura 7 - Exemplo 8 de TTC com 4 alunos e 4 escolas (3º passo)	116
Figura 8 - Exemplo 8 de TTC com 4 alunos e 4 escolas (4º passo)	117
Figura 9 - Ganhos de proficiência (Y) e o conjunto de todas as alocações possíveis ($\mathcal{M}$)	138
Figura 10 - Níveis de proficiência do PROEB em Língua Portuguesa para a 5º ano	148
Figura 11 - Representando a união dos bancos do Censo Escolar e do PROEB	153
Figura 12 - Fluxo das etapas empregadas nas simulações das proficiências médias	231

ÍNDICE DE MAPAS

Mapa 1 - a) Alunos e Escolas de Belo Horizonte representados por pontos e b) Mapa de Densidade aluno por Km ²	165
Mapa 2 - Regiões administrativas de Belo Horizonte	167
Mapa 3 - a) Nível Socioeconômico por Região e b) Mapa da densidade de alunos por km ²	168
Mapa 4 - Relação Aluno/Escola por bairros de BH, com resultado de I-Moran	173
Mapa 5 - Bairros agrupados por características geográficas e número de alunos	174
Mapa 6 - a) Distribuição da Proficiência por bairros e b) por grupos de bairros	180
Mapa 7 - Distribuição do Log da distância por Grupos de bairros	192

ÍNDICE DE GRÁFICOS

Gráfico 1 - Evolução da Demanda do Cadastro Escolar realizado nos Correios de BH	145
Gráfico 2- Distribuição da demanda no Cadastro Escolar de BH no ano de 2012-2013	146
Gráfico 3 - Histograma do Geocoding Certainty Indicador (GCI)	162
Gráfico 4 - Função de Densidade Acumulada do Indicador (GCI)	163
Gráfico 5 - Boxplot da proficiência entre o banco original e o banco final geoprocessado	164
Gráfico 6 - Relação entre NSE escolar do bairro e número de alunos 5º ano per capita	169
Gráfico 7 - Gráfico de Barras do número de bairros por faixas do número de alunos	171
Gráfico 8 - Frequência do número de Bairros por número de escolas	172
Gráfico 9 - Frequência de bairros agrupados por número de alunos	175
Gráfico 10 - Densidade Acumulada da Proficiência BH e MG	179
Gráfico 11 - Densidade das distâncias de casa até a escola para alunos do banco	191
Gráfico 12 - a) Gráfico de dispersão da relação do número de imigrantes e proficiência por regiões e b) por Grupos de Bairros	200

Gráfico 13 - a) Gráfico da relação entre proporção de imigrantes de bairro e proficiência média da escola com $NSE \leq 5,5$ e b) para $NSE > 5,5$	203
Gráfico 14 - Regressão polinomial localmente ponderada da proficiência e migração/região	204
Gráfico 15 - Densidade de kernel para a distribuição das distâncias observada e simulada	234
Gráfico 16 - Densidade de kernel para a distribuição das distâncias observada e da simulação com preferências dadas pelas distâncias	238
Gráfico 17 - Índice de utilização de escolas, simulação por distâncias e com vagas ociosas	241
Gráfico 18 - Relação da proficiência e as alocações simuladas	242
Gráfico 19 - Densidade de kernel para distâncias observadas e da simulação com preferências dadas pelas proficiências	244
Gráfico 20 - Índice de utilização de escolas, simulação por proficiência e com vagas ociosas	245
Gráfico 21 - Relação da proficiência e as alocações simuladas (μ^2) com vagas excedentes	246
Gráfico 22 - Densidade de distâncias observadas e simulação de preferências lexicográficas	248
Gráfico 23 - Índice de utilização de escolas, simulação lexicográfica e com vagas ociosas	249
Gráfico 24 - Relação da proficiência e as alocações simuladas (μ_3) com vagas excedentes	250
Gráfico 25 - Densidade distâncias observadas e da simulação com utilidades	252
Gráfico 26 - Relação entre proficiências médias e alocações simuladas (μ^1 e μ_4)	253
Gráfico 27 - Distribuições acumuladas da proficiência observada e após simulação (μ_1)	259

SUMÁRIO

1. INTRODUÇÃO	17
2. REVISÃO DA LITERATURA.....	23
2.1. Filas nas Escolas (o problema da rede pública de BH pré e pós-1993).....	24
2.2. A literatura de pareamento dos mercados de dois lados (Casamentos à la Roth-Sotomayor)	32
2.2.1. O sistema de alocação de Boston (EUA)	40
2.2.2. O sistema de alocação de Nova York (EUA)	45
2.2.3. O sistema de alocação de Singapura	49
2.2.4. O sistema de alocação de São Paulo (Brasil)	54
2.2.5. O sistema de alocação de Belo Horizonte (Brasil).....	56
2.2.6. Outras cidades?	62
2.3. A literatura de Pareamento na Produção Familiar (casamentos à la Becker).....	65
2.4. A Literatura sobre aquisição de aprendizado na Educação	68
2.5. A Interseção entre as três Literaturas	75
3. OS MÉTODOS DE PAREAMENTO.....	80
3.1. Principais definições e conceitos (Nomenclatura)	81
3.2. O Algoritmo de Gale-Shapley para casamentos (<i>one-to-one</i>).....	89
3.2.1. Manipulação do resultado por parte dos agentes.....	100
3.3. Extensão do Algoritmo de Gale-Shapley para casamentos (<i>many-to-one</i>)	103
3.3.1. Manipulação do resultado por parte dos alunos e escolas	107
3.4. O Algoritmo de <i>Top Trading Cycles</i>	109
3.5. Modelos de casamentos com base em transferência de utilidade (Becker).....	121
3.6. Uma proposição para matrizes de distâncias.....	132
4. A FUNÇÃO DE ALOCAÇÃO E PROFICIÊNCIA	137
5. BANCO DE DADOS DOS ALUNOS E ESCOLAS DE BELO HORIZONTE	143
5.1. Tratamento e Processamento para união dos Bancos de dados.....	152
5.2. Procedimentos para o Georreferenciamento do Banco	159
6. ANÁLISE DESCRITIVA DO BANCO E ALOCAÇÃO OBSERVADA DE ALUNOS E ESCOLAS	166
6.1. Análise exploratória do banco de dados.....	167
6.1.1. Regiões Administrativas	167
6.1.2. Divisão de Bairros	170
6.1.3. Análise descritiva das variáveis	177
6.2. Características de Mobilidade e Alocação Observada	187
6.2.1. Alocação Observada de alunos e escolas	189
6.2.2. Migração de alunos dados e definições	193

6.2.3.	Efeito da migração e do deslocamento na proficiência	200
6.3.	Modelos Hierárquicos e Decomposição de Fatores	207
6.3.1.	Fórmulas de decomposição da proficiência	218
6.4.	Parâmetros estimados para as simulações e conclusões da análise descritiva	225
7.	RESULTADOS DAS SIMULAÇÕES	227
7.1.	Simulação do <i>baseline</i> (<i>Top Trading Cycles</i>)	233
7.2.	Simulação considerando apenas as distâncias.....	236
7.3.	Simulação considerando apenas as proficiências médias.....	243
7.4.	Simulações considerando preferências lexicográficas	247
7.5.	Simulação considerando preferências dadas por utilidades	250
7.6.	Simulações de Políticas Públicas	254
7.7.	Simulações sob a perspectiva dos ganhos para os estudantes.	258
7.8.	Conclusões sobre as simulações.....	260
8.	CONCLUSÕES.....	265
9.	BIBLIOGRAFIA.....	272
10.	ANEXO A.....	282
11.	ANEXO B	292
12.	ANEXO C	308
13.	ANEXO D.....	309
14.	ANEXO E	313
15.	ANEXO F.....	314
16.	ANEXO G	316

RESUMO

O presente trabalho realiza um estudo sobre o sistema de cadastro escolar de Belo Horizonte, que completa, em 2013, 20 anos desde seu início. O sistema de cadastro é um procedimento de sugestão de matrículas para os alunos do ensino básico moradores de Belo Horizonte requisitantes de uma vaga nas escolas públicas municipais e estaduais da capital. Essa tese se propôs a investigar como o sistema de cadastro aloca os alunos nas escolas e quais os efeitos desta alocação sugerida no aprendizado dos alunos, medido pela proficiência do Programa de Avaliação da Rede Pública de Educação Básica (PROEB-MG), prova de língua portuguesa. Foram analisados alunos do 5º ano do ensino fundamental matriculados no ano de 2010.

Para se alcançar esse objetivo, a tese empregou como marco teórico três principais referências da literatura: a literatura de casamentos de Becker (1991) que foi extrapolada para formação de pares entre alunos e escolas, a literatura de algoritmos de pareamento iniciada por Gale & Shapley (1962) e apresentada em detalhes por Roth & Sotomayor (1990), e a literatura sobre, eficácia escolar, *school choice* e efeito do território sobre o aprendizado.

Com uso do PROEB, Censo Escolar e dados das escolas, fornecidos pela Secretaria Municipal de Educação (SME/PBH) e Secretaria Estadual de Educação (SEE/MG), estruturou-se um banco de dados georreferenciado de alunos e escolas e obteve-se a distância euclidiana de cada aluno para cada escola presente no banco. O banco final obteve 16.354 alunos e 296 escolas, com diversas informações características do aluno, da escola e do território. Dividido em bairros, grupos de bairros e regiões administrativas. Por meio do banco foi possível obter informações de mobilidade dos estudantes na cidade.

Por meio dos algoritmos de pareamento, e com o uso dos parâmetros obtidos na análise exploratória, propôs-se novas alocações para os alunos. Verificou-se quais efeitos essas alocações podem acrescentar à proficiência média. Pelas simulações, os efeitos finais dão conta que o sistema de cadastro acrescentou 10 pontos de proficiência média para a prova de português. Outras simulações sugerem que é possível acrescentar, além dos 10 pontos anteriores, cerca de 9 pontos a mais à proficiência, caso se adote outros tipos de alocações que encurtem ainda mais as distâncias dos alunos até as escolas.

ABSTRACT

This thesis conducts a study of the school registration system in Belo Horizonte, Brazil, this system completed, in 2013, 20 years since its inception. The system performs a registration procedure and suggests enrollment places for public elementary schools for students living in Belo Horizonte. This thesis aims to investigate how the registration system allocates students in schools and what effects this suggested allocation has on student learning, measured by their proficiency in PROEB-MG (*Programa de Avaliação da Rede Pública de Educação Básica*), a Portuguese language test for students in the 5th year of the elementary education system, in the term of 2010.

To achieve this goal, three main references in the literature were employed: the literature of marriages of Becker (1991), extended to pairing between students and schools, the matching algorithms literature initiated by Gale & Shapley (1962) and presented in detail by Roth & Sotomayor (1990), and the one of school effectiveness, school choice and effect of territory on learning.

The data from PROEB, School Census and schools, provided by the Municipal Education (SME/PBH) and the State Secretariat for Education (SEE/MG) was merged and structured into a georeferenced database of students and schools with Euclidean distance of each student to actual schools. The final database obtained 16,354 students and 296 schools, with several information characteristics of the student, school and location: city districts, groups of districts and administrative regions. This database allowed obtaining information on student mobility and building indicators for students changing districts and regions to study in different places of the capital.

Matching algorithms and parameters obtained in the exploratory analysis set out new allocations for students. These effects could be added to the proposed allocations average proficiency. Simulations showed that the final proficiency effects could be increased by 10 points on the average. Other simulations suggest that that it is possible to aggregate more 9 points of proficiency if other types of allocations that further shorten the distances of students to schools are used.

1. INTRODUÇÃO

O sistema de cadastro escolar de Belo Horizonte completa 20 anos em 2013 e foi implantado para racionalizar o processo de matrículas em escolas públicas da rede municipal e estadual do município. O sistema se aplica à todas as séries do ensino básico (fundamental e médio), sendo que o primeiro ano do fundamental, o ensino alfabetizante, ocupa um papel importante no processo. O sistema de cadastro coleta dados de moradia fornecidos pelos pais ou responsáveis e sugere, antes do início do ano letivo, em qual escola os alunos devem se matricular.

O sistema foi elaborado ao longo do ano de 1992 e as primeiras turmas cadastradas ocorreram em 1993. Desde aquela data, a mecânica do sistema continua a mesma, houve, porém, aperfeiçoamentos na coleta e registro da informação, no georreferenciamento de escolas e, principalmente, na obtenção do endereço dos alunos. Ocorreram avanços também no tratamento das matrículas para o ensino médio, e nos procedimentos dos casos excepcionais, tais como mudança de endereço, saída e consequente volta para o sistema, alunos com necessidades especiais, ou o problema de inconveniência ou inadequação da escola sugerida.

A proposta deste trabalho é estudar os efeitos da alocação sugerida pelo cadastro escolar no desempenho cognitivo dos estudantes do 5º ano do ensino fundamental.¹ O desempenho cognitivo será medido pelos escores da prova do PROEB-MG (Programa de Avaliação da Rede Pública de Educação Básica), que avalia os estudantes nas disciplinas de língua portuguesa e matemática. Os dados foram coletados para o ano de 2010, e o trabalho se atém aos resultados da prova de português.

Um dos problemas relacionados à captura dos efeitos da alocação na proficiência média dos alunos está no fato de que alocações alternativas não foram testadas. O primeiro ano de avaliação do PROEB foi 2002, quando o sistema de cadastro já havia completado 9 anos desde sua primeira implementação e a primeira prova do SAEB (Sistema de Avaliação da Educação Básica) foi aplicada em 1995, exame amostral e não censitário como o PROEB. Não há, portanto, nenhum

¹ Cabe aqui ressaltar que ao longo da última década um ano a mais de estudo foi acrescido ao ensino fundamental, que possui agora obrigação mínima de nove anos: o ensino alfabetizante para crianças de 6 anos, que pertencia antes ao ensino infantil, foi incluído nos anos de estudo do ensino fundamental. Ao longo deste trabalho estaremos no referindo à antiga 4ª série, ou o atual 5º ano.

levantamento da proficiência anterior ao sistema de cadastro de Belo Horizonte. Por conta dessas particularidades, essa tese se baseia em simulações para captar como poderia ser a proficiência antes da implantação de um cadastro, bem como simular situações alternativas para a alocação de alunos e escolas, que forneçam melhores resultados para o aprendizado.

Para a elaboração das simulações, a tese emprega os algoritmos da literatura de pareamento: a literatura de *two-sided matching*. Particularmente dois algoritmos serão destacados: o algoritmo *deferred-acceptance*, pioneiramente proposto por Gale & Shapley (1962), e o algoritmo de *Top Trading Cycles*, proposto pela primeira vez em artigo seminal publicado por Shapley & Scarf (1974) e para o qual se credita importante contribuição não publicada do professor David Gale. Desde os anos 60 e 70 essa literatura cresceu com bastante vigor, sendo aplicada em diferentes campos, inclusive no estudo de alocação de estudantes e escolas, e possui aplicações em diversas partes do mundo.

No Brasil, essa literatura contou grandes progressos nos trabalhos desenvolvidos pela professora Marilda Antônia de Oliveira Sotomayor, que escreveu, em parceria com Alvin E. Roth, um dos *best-sellers* da área de *two-sided markets* (Roth & Sotomayor, 1990). Em aplicações deste trabalho para o Brasil, tem-se como exemplos Sotomayor (1996a) e Bardella (2005), que estudaram a alocação de estudantes e centros de pós-graduação em economia proporcionados pelo sistema de seleção da ANPEC (Associação Nacional dos Centros de Pós Graduação em Economia). É de particular importância para esta tese o trabalho de dissertação defendido por Fernandes (2007), que analisou o sistema de cadastramento das escolas públicas de São Paulo, sob a perspectiva de teoria de jogos (ver capítulo 2).

Os algoritmos acima mencionados moldaram as simulações realizadas na presente tese, e com isto serviram de base para o estudo das implicações de políticas públicas sobre sistemas alternativos de alocação dos estudantes. Em outras palavras, uma contribuição desta tese é estabelecer a relação entre a alocação e o bem-estar social. E, no caso, mede-se o bem-estar vinculado somente ao quesito educação, melhor dizendo, à aprendizagem, que foi medida pela proficiência média dos estudantes do 5º ano do fundamental.

Duas outras literaturas pertinentes para o tema são a literatura de casamentos proposta por Becker (1973 e 1974) e a literatura sobre os efeitos do espaço urbano e características do território no aprendizado de estudantes, ver Ribeiro & Katzman (2008).

Existe uma interseção entre a bibliografia de casamentos de Becker (1973 e 1974), e a literatura dos casamentos do *two-sided markets*. Essa interseção é explicitamente identificada por Roth & Sotomayor (1990) e Becker (1991), e é mais bem elaborada por Browning, Chiappori & Weiss (2011).

A outra área de estudos mencionada, a da relação entre espaço urbano e desempenho no aprendizado, investiga os possíveis efeitos da região de moradia na proficiência. Investiga também o efeito da localização da escola, procurando compreender em que condições o território contribui para a redução ou perpetuação das desigualdades. Em particular, Soares, Rigotti & Andrade (2008) estudaram os efeitos da segregação espacial urbana de Belo Horizonte no desempenho do SAEB, constituindo-se de um estudo prévio importante para análise exploratória de dados realizados nesta tese.

Como o sistema atual de cadastro de Belo Horizonte estipula uma alocação dos alunos com base nas distâncias, um dos pontos centrais deste trabalho é investigar se tal sistema é o melhor possível, e questionar se ele não penalizaria demasiadamente alunos com potencial para o aprendizado, mas que são moradores de regiões distantes de boas escolas (definindo-se boas escolas como aquelas que possuem maior nível de proficiência).

Nota-se que o sistema de cadastro foi um grande avanço em termos de políticas públicas. Antes do sistema, tinha-se o problema das filas e uma regra que quase sempre resultava em práticas clientelistas para a seleção das vagas nas escolas (Pinto, 1999). O atual sistema foi proposto com o objetivo de mediar os conflitos por vagas advindos da pressão de uma demanda espontânea que ocorria para determinadas escolas (Fonseca & Zuppo, 1997).

O antigo sistema gerava um desperdício de recursos, matrículas duplicadas, e uma dificuldade de sistematização das informações. A implementação do procedimento de cadastro escolar

informatizado e sistematizado foi um avanço reconhecido pela sociedade (PBH, 2003) e trouxe organização e segurança no procedimento de matrículas. Por outro lado, criou um sistema determinístico que não considera as preferências dos pais pelas escolas. Ao designar a escola mais próxima para a matrícula automática, considera que as escolas públicas são todas homogêneas, e que não haveria razão para pais e alunos se deslocarem mais para obterem um tipo de ensino proveniente de outra escola que consideram mais adequada.

A espacialização é uma questão importante, pois as pessoas se movem o tempo inteiro na cidade, e isto não é diferente para o caso dos alunos. Se pais e alunos que desejam uma vaga nas escolas públicas pudessem escolher livremente, e não houvesse limite de vagas, provavelmente eles escolheriam a escola mais conveniente, considerando aspectos variados como a locomoção, qualidade do ensino, facilidade de acesso, relacionamento com professores e funcionários e outros aspectos. Supõe-se então que pais e alunos realizam um *trade-off* entre distância e qualidade do ensino e uma questão que se coloca é saber se vale a pena se locomover mais para obter um ensino melhor.

Essa pergunta relaciona-se ao tema da tese, que é na área de “Demografia da Educação”, que estuda os fluxos de alunos no tempo e espaço, a demanda e a oferta de vagas, as taxas de promoção e repetência, defasagem idade-série, os anos de estudo e os demais indicadores da produção e da qualidade educacional.² O fluxo de alunos no espaço e como isso impacta na proficiência, será estudado neste trabalho.

Formulou-se a hipótese de que uma parte da baixa qualidade e desigualdade do sistema educacional pode ser causada por um problema de alocação entre alunos e escolas. A investigação empírica consiste na busca de uma relação entre a alocação e o desempenho escolar agregado. Para cada alocação, buscou-se mensurar o desempenho total no sentido de identificar qual é a alocação ótima, a que oferece maior desempenho escolar agregado total. Em outras palavras, procura-se responder à pergunta: *poderia a qualidade média do ensino público básico*

² Mais detalhes sobre a área de demografia da educação encontram-se no livro “*Demografia da Educação*” organizado por Rios-Neto & Riani (2004).

em Belo Horizonte ser aumentada por um efeito de alocação, ou seja, por uma alteração na distribuição dos estudantes entre as escolas?

Aprofundando-se sobre esta questão, pode-se presumir que as escolas com melhor qualidade de ensino estejam em zonas mais favorecidas da cidade.³ Sendo assim, o atendimento de qualidade para os que já possuem acesso a mais recursos se perpetuaria. Desse modo, outras perguntas relacionadas ao tema se configuram: 1. *A alocação por distância perpetua as desigualdades de um ambiente de alunos e escolas heterogêneo?* 2. *Quais são as melhores escolas e alunos e como medir essa qualidade?* 3. *Qual é a troca que uma sociedade deve fazer ao alocar um aluno de uma comunidade “desfavorecida” em sua melhor escola?* e 4. *Caso seja feita a troca, para onde deve ir o aluno não-vulnerável desalojado pelo aluno vulnerável?*

A tese elabora um arcabouço teórico-metodológico para responder essas questões e procura chamar a atenção dos demógrafos da educação para o tema do pareamento (casamentos) entre alunos e estudantes. Muitos trabalhos da demografia procuram estimar a oferta necessária para que o número de vagas se equivalha às projeções da demanda futura. Isto é extremamente válido, pois a existência da paridade entre oferta e demanda por vagas é uma condição *necessária* para se prover um ensino de qualidade. Porém, existe mais uma condição também *necessária*: a que trata do melhor uso possível dos recursos, ou seja, uma condição que garanta que os alunos estejam alocados para as vagas em que possam render o seu melhor potencial *vis a vis* os demais estudantes (a alocação ótima). Essa é uma condição de estabilidade dos algoritmos de pareamento que serão estudados nos capítulos a seguir.

Para tratar das propostas discutidas neste capítulo introdutório (capítulo 1) esta tese se divide em outros sete capítulos: 2) capítulo de revisão da literatura, em que se apresentam o problema das filas nas escolas de Belo Horizonte, a literatura de pareamento de Roth & Sotomayor (1990), de Becker (1991), e a literatura de efeitos do território e competição de escolas no aprendizado dos alunos; 3) no terceiro capítulo é apresentada a notação que será utilizada e a formalização dos modelos que serão empregados nas simulações, mecanismos *centralizados* de alocação. Tais

³ Dados de Soares, Rigotti & Andrade (2008) permitem inferir que essa distribuição desigual de escolas no espaço urbano, com melhores escolas situadas em bairros centrais e com maior renda *per capita*, deve ser verdadeira.

mecanismos envolvem pareamentos dados pelos algoritmos *deferred-acceptance* (DA) e *Top Trading Cycles* (TTC), que reúnem todo o conjunto de preferências e emitem um resultado de alocação final; 4) O quarto capítulo aborda a relação entre a alocação e a proficiência; 5) O quinto capítulo apresenta detalhes da construção do banco de dados de alunos e escolas e o seu georreferenciamento; 6) O sexto capítulo faz uma análise exploratória dos dados, uma abordagem da relação entre distâncias e a proficiência e o efeito de alunos migrantes nas escolas, o capítulo elabora também resultados de regressões de modelos hierárquicos que serão utilizadas no capítulo sete; 7) O capítulo sete trata das simulações de resultados que utilizam os algoritmos apresentados no capítulo três. Esse capítulo mostra também os principais resultados de proficiência das simulações e alocações sugeridas; 8) O oitavo e último capítulo tece conclusões e apresenta as principais sugestões para políticas de alocação dos alunos e fornece sugestões para trabalhos posteriores.

2. REVISÃO DA LITERATURA

Esta revisão de literatura segue por cinco linhas principais: a primeira delas ressalta o problema real enfrentado pelas escolas, resgata o histórico de implementação do sistema (em Davis Jr., 1997; Fonseca & Zuppo, 1997; e Pinto, 1999), e apresenta as informações mais atuais que se tem a respeito do sistema de cadastro de Belo Horizonte, ressaltando a importância do problema da alocação entre estudantes e escolas.

A segunda linha enfoca a questão do pareamento tal como pioneiramente tratado por Gale & Shapley (1962) e Shapley & Scarf (1974), a literatura e os avanços teóricos anteriores à década de 90 foram consolidados no livro de Roth & Sotomayor (1990). Roth & Sotomayor apresentam em detalhes uma das aplicações práticas mais famosas do algoritmo nos EUA, o sistema de alocação de estudantes de medicina chamado de *National Residence Matching Program* (NRMP).⁴ Também serão discutidos avanços posteriores da teoria e da aplicação dos algoritmos da área de *two-sided markets*, em particular, os estudos dos sistemas implementados nas cidades de Nova York, Boston, Singapura e São Paulo, que oferecem lições importantes para desenvolvimentos teóricos posteriores.

O terceiro enfoque trata dos pareamentos sob a abordagem econômico-demográfica de Gary Becker (1991), também são abordados desenvolvimentos posteriores dessa literatura. Como a literatura proveniente de Becker se aplica especificamente a situação de casamentos de homens e mulheres, será necessário extrapolar a analogia para “casamentos” entre alunos e escolas, casamentos *many-to-one*, que na literatura de Becker podem ser pensados como casamentos polígamos.

O quarto enfoque se volta para a literatura educacional. Na abordagem da revisão da literatura educacional, procura-se compreender quais os fatores escolares são mais importantes para obtenção de um melhor ensino. Essa parte disserta sobre os elementos já conhecidos e que afetam o aprendizado e procura elementos dessa literatura que subsidiem o papel da alocação dos alunos,

⁴ O NRMP ficou tão identificado com os algoritmos *deferred-acceptance many-to-one* que boa parte da literatura dos anos 90 se refere a esses algoritmos como algoritmos NRMP, mais detalhes sobre a literatura desse algoritmo serão fornecidos mais adiante neste capítulo.

tais como os programas de *school choice* dos EUA e efeitos do território. Serão, portanto, ressaltadas as diversas dimensões que determinam o aprendizado: alunos, escolas e alocação/território.

O quinto e último enfoque se volta para interseção entre essas literaturas. Como abordado na introdução da tese, existe uma intersecção entre a literatura de casamentos dada por Becker (1991) e Becker & Murphy (2000) e a literatura dada por Roth & Sotomayor (1990). Algumas dessas interseções são identificadas explicitamente pelos respectivos autores mencionados e aparecem também em livro recente de Browning, Chiappori & Weiss (2011). Menos destacada é a interseção entre essas duas literaturas e a literatura da educação, no que essa tese procurará apontar caminhos para a união dessas três linhas de pesquisa.

2.1. Filas nas Escolas (o problema da rede pública de BH pré e pós-1993)

Infelizmente, durante muitos anos, o Brasil foi um país em que a grande maioria da população se instruía pouco. Nos anos 70, a média de anos de estudo dos brasileiros com 25 anos ou mais era de 2,4 anos de estudo, e 69% da população nesta faixa etária tinha menos de 4 anos de estudo, o que equivaleria hoje a menos do que os cinco primeiros anos do fundamental.⁵ Naqueles anos, a educação pública brasileira se restringia a um grupo pequeno de alunos com oportunidade de estudo, geralmente, provenientes de famílias de renda média ou superior que prezavam o estudo e almejavam para seus filhos postos de trabalho que exigiam maior escolarização.

No decorrer dos anos 80, ocorreu uma expansão do ensino público para um número maior de alunos, atingindo-se um escopo mais variado de classes, ao mesmo tempo em que houve uma expansão da rede privada de ensino fundamental. O ensino privado foi a solução encontrada para substituir o ensino público que perdia qualidade e não abrangia todos os alunos, fato que ocorria por conta dos poucos investimentos destinados à educação durante as décadas anteriores (para uma revisão da política educacional das décadas anteriores, ver Demo, 2007 e Coelho 2008).

⁵ IBGE, Censo Demográfico de 1970.

No final dos anos 80, com a restauração do regime democrático, a Constituição Federal de 1988 (CF/88) criou um marco para as políticas públicas brasileiras. No caso da educação, se tornou de especial importância o reconhecimento dos princípios básicos compreendidos nos artigos 205 e 214. As normas constitucionais da CF/88 estabeleceram diretrizes para a política educacional dos anos 90 em diante. Depois de estabelecido na carta constitucional o direito à educação universal, um dos maiores problemas da década de 90 foi o de como prover vagas em número suficiente para a inclusão de enorme contingente de crianças e adolescentes nas escolas.

O excesso de demanda pelas escolas públicas, aliado a um conjunto de escolas heterogêneas do lado da oferta, levou a formação de filas para efetuação das matrículas em muitas das capitais brasileiras. Conforme relata Pinto (1999, p. 140): “O crescimento urbano intenso não foi acompanhado pela formação de infra-estrutura de serviços de consumo coletivo. A população contava com número insuficiente de escolas e serviços precários de redes de esgoto, água, luz e postos de saúde”.

Esse número insuficiente de vagas nas escolas originava longas filas de pais tentando matricular seus filhos nas escolas mais disputadas. Em verdade, como destacam Fonseca e Zuppo (1997), a insuficiência de vagas não se manifestava para o sistema como um todo, mas sim para algumas escolas muito demandadas. Segundo os autores (1997, p.16):

“Esta política se mostrou a mais adequada, porque algumas vezes os pais escolhiam escolas com a capacidade já esgotada, enquanto outras ainda dispunham de vagas. Isto ocorria porque algumas escolas eram consideradas melhores do que outras”.

De acordo com Fonseca e Zuppo (1994, 1997) e Pinto (1999), existem três motivos principais que se pode identificar para a formação das filas nas escolas e deficiências no sistema anterior de cadastro. Destaca-se abaixo uma enumeração dos pontos colocados por estes autores, acrescenta-se também um motivo demográfico extraído de Rigotti (2001) e outro estrutural e pedagógico proposto por meio de evidências da literatura, totalizando cinco motivos identificados e aqui reunidos:

1. A crise econômica levou a substituição de escolas particulares por escolas públicas (especialmente em Belo Horizonte no ano de 1993, logo após a crise do biênio 1991-92). (Fonseca e Zuppo, 1997, p. 16). Esse ponto é particularmente importante, no início dos anos 90, com as sucessivas crises econômicas, pais de classe média fizeram um retorno às escolas públicas, acarretando uma pressão na demanda.
2. Muitas décadas de precário investimento na infra-estrutura escolar (Pinto, 1999, p. 140). Esse fator prejudicaria a oferta adequada de boas escolas.
3. Informação, dados insuficientes ou precários (Pinto, 1999, pp. 141-42). A falta de dados não ocasiona as filas, mas torna a busca pela solução do problema mais demorada e difícil. Belo Horizonte possuía informação de cadastro escolar desde 1974, porém, durante muito tempo essas informações eram colhidas de forma descentralizada pelas escolas, impossibilitando o arranjo para solução do problema.
4. Desde a década de 1970, as taxas de fecundidade começaram a dar seus primeiros sinais de queda, acompanhando a taxa bruta de mortalidade, que apresentava queda desde os anos 1940. Nos anos mais recentes, se caminha para uma mais baixa mortalidade e fecundidade. O Brasil passa por uma acelerada transição demográfica e isso impacta no sistema educacional do país e produz impactos na aquisição de anos de estudo (Rigotti, p. 60). Ao longo de toda a década de 90, o elevado contingente de crianças e jovens pressionava a demanda por ensino, e, ao longo dos anos 2000, à medida que o Brasil atinge outra fase de sua transição, com maior proporção da população em idade adulta, essa pressão diminui e as filas surgem em categorias mais elevadas da educação.
5. Uma hipótese formulada por Fernandes (2007) e também mantida neste trabalho é de que as escolas públicas são heterogêneas em sua qualidade. Na implementação do sistema de cadastro de Belo Horizonte, em 1993, fez-se um grande esforço de convencimento da opinião pública para a entrada no novo sistema. Nisso foi necessário desconstruir a visão de que as escolas possuíam qualidades distintas. Em verdade, esse foi um compromisso da administração municipal e estadual: tornar as escolas públicas melhores, sendo que não se

prejudicaria alunos que estudassem nas escolas mais próximas. A política do cadastro precisava, portanto, se basear na premissa de escolas homogêneas. Contudo, evidências apontam que algumas escolas eram mais preferidas do que outras, contribuindo para a formação de longas filas para as escolas mais desejadas.

No ano de 1992, a Secretaria Municipal de Educação (SME-BH), em parceria com a Secretaria Estadual de Educação (SEE-MG), montou uma comissão para analisar o problema das matrículas. A empresa de informática de Belo Horizonte (PRODABEL) participou da criação do sistema de matrícula que, em parte, correspondia à unificação e informatização dos procedimentos de registros das escolas e, de outra parte, de um novo sistema de processamento do local de moradia dos alunos. Tal processo está descrito em maiores detalhes em Pinto (1999).

Na construção do novo sistema, primeiro se verificou se o número de vagas existentes era suficiente para atender a todos que procuravam o sistema público. Como a constatação foi de que havia número suficiente de vagas, concluiu-se que as filas ocorriam principalmente porque determinadas escolas eram mais procuradas do que outras.

O que se tinha era um problema de alocação. A comissão propôs um sistema para racionalizar as matrículas por meio de um cadastro dos alunos. O cadastro de alunos já existia anteriormente, mas não era completamente informatizado. A PRODABEL ficou responsável pelo georreferenciamento de alunos e escolas, processo este que se encontra descrito em Davis Jr. (1997).

Foram construídas “jurisdições” escolares de maneira a adequar o número de vagas ao número de alunos. Tais jurisdições foram construídas levando-se em conta características geográficas tais como barreiras físicas (avenidas, pontes, topografia, etc), a densidade populacional de pessoas em idade escolar nas cercanias das escolas e outros fatores.

O número de vagas em cada jurisdição deve ser suficiente para atender a demanda dos residentes daquela área. Uma mesma jurisdição pode ter mais de uma escola dentro de seus limites, mas

apenas uma escola (que possua o fundamental de 1º ao 5º ano) é a escola responsável daquela jurisdição.⁶

O procedimento para o cadastro requeria que os pais efetuassem o registro em uma das agências dos Correios (Empresa Brasileira de Correios e Telégrafos – EBTC). Essa relação de unidades dos Correios que realizam o cadastro é disponibilizada, todo ano, em uma chamada pública que é divulgada nas escolas, postos de Saúde, unidades dos Correios, e unidades governamentais de atendimento ao público e de grande movimento. Nos primeiros anos, anúncios de rádio e televisão eram também utilizados para veiculação do cadastro.⁷

Para efetuar o cadastro, o responsável deve levar a certidão de nascimento do aluno e o comprovante de residência obtido na conta de energia elétrica da CEMIG – Companhia Energética de Minas Gerais. O número de registro na CEMIG é utilizado para facilitar o georreferenciamento das moradias dos estudantes.⁸ Um avanço mais recente, de 2005, garantiu maior agilidade no preenchimento e processamento dos dados nos Correios. A partir de 2005 o preenchimento passou a ser digital e realizado na agência, com ajuda de um funcionário dos Correios, que repassa as informações no sistema. Antes os pais preenchiam uma ficha cadastral que era enviada para a SME-BH, agora os registros são enviados por sistema eletrônico.

Com o georreferenciamento de alunos e escolas, foi possível fazer um sistema de cadastramento automático, que matriculava os alunos nas escolas mais próximas sem necessidade da criação de filas. Assim, sobrevinha com esse sistema uma alocação de acordo com a localização geográfica, um *matching* entre escolas e alunos que levava em conta a proximidade. É preciso notar que o termo “*matrícula automática*” precisa ser ponderado, a matrícula é apenas sugerida pelo sistema, é necessário que os pais validem a sugestão de matrícula proposta. No final do ano letivo, pais que se cadastraram nos Correios para o sistema de matrículas recebem uma correspondência

⁶ Informações do relatório de Cadastro Escolar 2012/2013 da diretoria de Cadastro da SME-BH. O termo jurisdição é usado pela SME, entretanto é preciso denotar que sem definição jurídica para essas áreas de abrangência das escolas.

⁷ O anexo A, na tabela A.1, apresenta o cronograma das etapas desenvolvidas no cadastro.

⁸ Nos primeiros anos a conta da CEMIG não era exigida, pais poderiam levar qualquer comprovante de endereço com CEP (Código de Endereço Postal). Em 1997, para apurar o sistema de georreferenciamento dos endereços dos alunos, a conta de energia elétrica passou a ser exigida, isso conferiu uma maior padronização da base de endereços, dado que cada ponto de ligação para energia residencial da CEMIG possui um número identificador e esses números estão ligados a uma base georreferenciada, compartilhada entre PBH, CEMIG e várias instituições conveniadas.

indicando em qual escola devem matricular seu(s) filho(s). Os pais não precisam matricular os filhos na escola sugerida, mas ao seguirem a indicação do sistema, a matrícula é garantida.

O sistema de cadastro resultou em altos índices de aprovação pela população, conferindo ao programa o prêmio de políticas sociais de 2001, concedido pela Fundação Ford (informações disponíveis no portal da prefeitura municipal de Belo Horizonte, PBH).⁹ Conforme dados levantados por Fonseca & Zuppo (1997), após a implantação do sistema, houve 2.630 reclamações de pais que tinham preferência por outra escola. É um número extremamente pequeno perto das mais de 30 mil matrículas efetuadas pela primeira rodada do sistema.

O sistema de Cadastro Escolar continua até hoje sem grandes alterações. Desde o início contou com o apoio e colaboração da Secretaria Estadual de Educação e a parceria dos Correios. O sistema realiza a indicação de matrículas para a rede pública estadual e municipal dentro dos limites legais do município de Belo Horizonte. O processo de cadastro e matrículas, em circunstâncias normais, se dá em um prazo não muito maior do que três meses. Os pais são responsáveis por fazer a etapa final de matrícula na escola recomendada pelo sistema, que é informada previamente por carta.¹⁰

Ao longo dos anos, o sistema foi se aperfeiçoando para corrigir erros de informação e incluir necessidades de transferência, readequação de vagas, alunos que não se inscreveram no período adequado, mudança ou inclusão de endereço de famílias que migraram para Belo Horizonte, e outras necessidades relacionadas ao cotidiano escolar. Para as etapas posteriores do aluno já matriculado em uma escola pública, há, dentro do processo, um encaminhamento automático das matrículas para os anos seguintes: um aluno matriculado na série inicial em uma determinada escola tem preferência de matrícula na etapa posterior da mesma escola no ano seguinte. No caso de reprovação, ele também tem sua vaga garantida na mesma escola.

⁹ Link da Prefeitura de Belo Horizonte com informações do sistema de cadastro. Último acesso dia 21/04/2013: <http://portalpbh.pbh.gov.br/pbh/ecp/contents.do?evento=conteudo&idConteudo=22185&chPlc=22185&viewbusca=s>

¹⁰ No Anexo A, após a resolução conjunta que estabelece as normas para o cadastro, há um modelo recente dessa carta de convocação para matrícula. Há também um exemplo de carta enviada para alunos moradores de cidades vizinhas, mas que tentaram realizar o cadastro em Belo Horizonte.

O anexo A apresenta a resolução conjunta da SME-BH e SEE-MG de 15 de Junho de 2012. Esta resolução é geralmente divulgada ao final do primeiro semestre ou início do segundo semestre. A resolução fornece as diretrizes, os prazos e os passos para o procedimento de matrículas. Na seção 2.2.5 esses passos do mecanismo de Belo Horizonte são apresentados em detalhes.

Como se constata nesta parte da revisão, o sistema de cadastro escolar de Belo Horizonte passou da forma descentralizada que vigorava no início da década de 90, com pais fazendo filas para matricular seus filhos nas melhores escolas públicas, gerando grande desconforto e ineficiência do sistema, para uma forma centralizada, ou semi-centralizada em que o processo se dá de uma maneira coordenada e alia a oferta de vagas às demandas existentes.

O cadastro escolar é hoje um sistema misto, ele é centralizado no que diz respeito ao trabalho da comissão de cadastro e das secretarias de educação, mas também conta com uma parte descentralizada, pois a matrícula continua a cargo de pais, alunos e escolas (com grande participação das regionais de ensino). O sistema apenas sugere as vagas para serem alocadas.

Pode se caracterizar o sistema como relativamente estável, pois já se passaram 20 anos desde sua implantação, sendo que o projeto não sofreu discontinuidades, apenas melhoramentos pontuais no cadastro e informatização. Existem situações de descontentamento de parcela dos alunos, dado que não há indicação completa das preferências, mas essa parcela não é grande, não superando 5% dos alunos requerentes.

O autor desta tese ouviu relatos sobre pais (ou responsáveis) que adulteram endereços para conseguirem maiores chances de matricular os filhos em escolas que julgam melhores. Esses procedimentos envolvem o pedido da conta de energia elétrica para familiares, patrões ou pessoas próximas. Essa prática é coibida pela secretaria com a advertência: “Qualquer informação ou documento falso tornará nulo o cadastro do seu filho”. Na situação de identificada a fraude, o responsável é chamado a uma unidade regional de ensino da prefeitura e sua vaga será alocada de acordo com critérios de conveniência da secretaria.

O fato de pais tentarem alterar o endereço denota alguma instabilidade e insatisfação por parte de alguns participantes com o mecanismo atual, porém, pelas evidências do banco de dados (a serem trabalhadas no capítulo 6), a parcela de adulteração de endereços deve ser pequena, indicando mais a favor da relativa estabilidade do atual sistema do que contra.

Na seção a seguir, a revisão se concentra na literatura de *two-sided markets*, que será utilizada nos métodos e nas simulações. Como já apresentado, o artigo de Gale & Shapley de 1962 é considerado a pedra fundamental do desenvolvimento desta literatura, a contar desde a data desta publicação, a literatura de algoritmos determinísticos de pareamentos por preferências possui mais de 50 anos de desenvolvimento, sendo que nos últimos 20 anos observou-se grande aumento das aplicações práticas desses estudos. A revisão da literatura abordará estudos de casos da educação de outras cidades já bastante analisadas pela literatura: Boston, Nova York, Singapura e São Paulo e dará os detalhes do passo a passo do sistema de Belo Horizonte.

Na terceira subseção, a literatura de casamentos de Becker (1991), mais tradicionalmente ligada à demografia, é analisada. Essa literatura, também chamada de literatura de *assortative mating*, contou com desenvolvimentos recentes, sendo ainda hoje empregada para testar hipóteses sobre comportamentos sociais. Apesar disso, até o presente momento, essa literatura não foi apropriadamente extrapolada para o caso de casamentos entre escolas e alunos, fato que pode ser explicado pela longa tradição dessa literatura em analisar a formação de famílias e o contexto familiar.

A quarta subseção trata da literatura de aquisição de ganhos educacionais, como alunos e escolas interagem e quais são as variáveis importantes nesse contexto. Nessa seção realiza uma revisão dos trabalhos de efeito-aluno e efeito-escola, presente em César & Soares (2001), Albernaz, Ferreira & Franco (2002), Machado *et al.* (2006), entre outros. É de particular interesse para o assunto aqui tratado os efeitos de interação entre alunos e escolas, sendo que essa característica será enfatizada sempre que possível. Outro ponto importante da literatura trata dos efeitos do território no aprendizado, um texto de referência utilizado nesse campo foi o artigo de Soares, Rigotti & Andrade (2008), que analisou a segregação urbana de Belo Horizonte e seus aspectos escolares.

A quinta subseção da revisão da literatura explora as interseções entre essas três literaturas. A interseção entre as literaturas de *matching* existe e é identificada explicitamente tanto por Roth & Sotomayor (1990) quanto por Becker (1991) e Becker & Murphy (2000), autores que se referenciam mutuamente em algumas oportunidades. Entretanto, até o conhecimento presente, a interseção entre essas duas literaturas não foi explorada em maior profundidade, sendo que esta tese pretende apontar os caminhos para uma maior interação entre as duas linhas de pensamento, apontando conceitos semelhantes e os pontos em que elas divergem.

Em se tratando da literatura de aquisição de conhecimento na educação e as literaturas anteriormente mencionadas, há uma grande interseção que está implícita nas três teorias. Há, porém, divergências fundamentais na operacionalização dos conceitos.

Contudo, não é objetivo desta tese esgotar todas as possibilidades dessa interseção entre as três áreas (educação, *assortative mating*, e *two-sided markets*), visto que não se trata de uma tese sobre metodologia da ciência ou de técnicas de pesquisa. Apesar disso, assinalando a interseção e a distinção entre as áreas, se torna mais fácil contextualizar o arcabouço teórico do presente trabalho.

2.2. A literatura de pareamento dos mercados de dois lados (Casamentos à lá Roth-Sotomayor)

Para a demografia tradicional o *matching* pode ser também interpretado como um casamento. No que se refere ao arcabouço demográfico geral, esta tese tratará de casamentos de um tipo especial, que ocorrem entre *escolas* e *alunos*. Neste capítulo a revisão da literatura se atém aos aspectos do desenvolvimento da teoria e de seus movimentos históricos até a chegada do reconhecimento mais recente desta área de pesquisa com os trabalhos publicados nas homenagens póstumas ao professor David Gale (Sotomayor 2009) e no reconhecimento público do denominado prêmio Nobel de economia de 2012.¹¹

¹¹ O nome completo da premiação é prêmio Sveriges Riksbank de Ciências Econômicas em Memória de Alfred Nobel, concedido pelo banco central da Suécia e a academia real de ciências daquele mesmo país. Como tal prêmio

O artigo seminal de Gale & Shapley (1962) propõe o tema dos casamentos para tratar da questão teórica de alocação de alunos e escolas. O tema aparece no título do artigo: “*College Admissions and The Stability of Marriage*”. No referido artigo, os autores propõem o modelo que deu origem aos algoritmos *deferred-acceptance* utilizados nesta tese. A ideia proposta pelo artigo tratava do estudo de como alunos e escolas podem ser arranjados. Para tratar melhor o problema, os autores o reduzem para sua versão mais simples, de casamentos entre homens e mulheres, um pareamento um a um (*one-to-one*).

No modelo proposto por Gale & Shapley (1962) há dois conjuntos disjuntos de homens (H) e mulheres (W), os membros de cada um dos conjuntos deseja formar par com um, e somente um, membro do outro conjunto. Assim se estabelecem as preferências, cada indivíduo h do conjunto H de homens possui preferências racionais e completas sobre as todas as mulheres do conjunto W . O algoritmo consiste em cada um dos membros de um dos lados (digamos que os homens) propor ao membro mais preferido do outro lado (as mulheres). No caso de uma mesma mulher receber mais de uma proposta ela deve “segurar” a mais preferida e recusar as que são inferiores à melhor proposta no momento.¹²

O nome *deferred-acceptance* para o algoritmo advém dessa característica: ninguém está definitivamente casado até que o algoritmo termine e o último membro do grupo passivo (mulheres, em continuidade à suposição do parágrafo anterior) receba uma proposta. *Deferred* significa *adiar*, ‘*diferir*’, ou seja, aceitar uma proposta até que surja uma melhor, ou até que o tempo dado pelos passos do algoritmo se esgote. O termo *acceptance* diz respeito a *aceitar*, *admitir*, a aceitação definitiva do compromisso ocorre ao final do algoritmo, quando cada um dos pares se casa com a melhor proposta até o momento.

foi instituído em 1968 e não foi originalmente proposto pelo fundador Alfred Nobel (1833-1896), paira a controvérsia sobre se o prêmio deveria ser denominado como Nobel. Porém, por homenagear o famoso cientista sueco e ser uma denominação mais sucinta, o prêmio é geralmente divulgado como Nobel de Economia pela imprensa e pelo próprio portal oficial da premiação na internet: http://www.nobelprize.org/nobel_prizes/economics/laureates/2012/ último acesso (21/01/2013).

¹² A seção 3.2 do próximo capítulo fornece os passos do algoritmo e exemplos. Além disso, enuncia os principais teoremas e proposições dessa teoria.

Os algoritmos *one-to-one* podem ser facilmente generalizados para situações *many-to-one* que se aplicam ao contexto de escolas, sendo que é preciso somente tratar cada vaga de uma escola como se fosse um indivíduo, respeitando o ordenamento prévio das preferências.

Um dos resultados mais importantes demonstrados por Gale & Shapley (1962) é o de que o algoritmo *deferred-acceptance* no mercado de casamentos (*two-sided*) alcança sempre uma solução e de que esta solução é *estável*, não há como melhorar nenhum dos pares formados os casando novamente sem piorar algum indivíduo já pareado. Outra formulação do problema, proposto por Gale & Shapley no mesmo artigo, envolve a formação de pares entre membros de um único grupo (*one-sided*), essa formulação foi chamada de *roommate problem*, se aplica ao problema de juntar pares de alunos para ocupar um mesmo apartamento, ou salas, por exemplo. No *roommate problem* não há garantias de uma solução estável, pois não há conjuntos disjuntos.

Os anos 60 e 70 foram de grandes avanços para a área de teoria dos jogos, para entender esse desenvolvimento é necessário compreender que os algoritmos de *matching* pertencem a uma área da teoria dos jogos chamada de jogos cooperativos. Shubik (2003) define jogos cooperativos como jogos de coalisção que não possuem um mecanismo explícito para a formação de preços. Segundo o mesmo autor, jogos não-cooperativos são jogos que possuem uma regra explícita para a função de reação dos agentes e, dessa maneira, há garantia da ocorrência de um equilíbrio. No caso, Roth & Sotomayor (1990) e Sotomayor (1998) pontuam que a teoria de *two-sided markets* possui tanto uma parte não-cooperativa, pois agentes de um mesmo grupo (homens ou mulheres) competem entre si para conseguir o melhor par possível e possuem estratégias para tentar conseguí-los, quanto a cooperativa, que se revela no interesse do mercado de formar os pares.

Debreu (1951), Arrow & Debreu (1954), e Debreu & Scarf (1963) analisam garantias para existência do *core* em economia de puras trocas. Em particular, Debreu & Scarf (1963) retomam o trabalho de Edgeworth (1881, *apud*: Debreu & Scarf, 1963) seguindo a sugestão de Shubik (1959, *apud*: Debreu & Scarf, 1963) de que o núcleo poderia ser também interpretado sob a perspectiva de teoria dos jogos (jogos com n -agentes). O artigo de Debreu & Scarf (1963) foi

importante para uma sequência de trabalhos que se seguiram desenvolvendo os conceitos de *core*.¹³

Um dos principais trabalhos, dessa área, da década de 70, é o artigo de Shapley & Shubik (1971) que sedimentou as bases dos *assignment games*, jogos para os quais é possível inserir valores monetários com transferências de bens. Os *assignment games* possuem uma interseção importante com os modelos tradicionais de *two-sided markets* e ajudaram a sedimentar as bases para Shapley & Scarf (1974) propor um mecanismo de alocação para o mercado de imóveis, o algoritmo de *Top Trading Cycles* (TTC) mencionado na introdução e que será detalhadamente apresentado no capítulo 3 e aplicado nas simulações do capítulo 7.

As décadas de 80 e 90 assistiram uma enorme expansão da literatura de *two-sided markets* e dos mecanismos de pareamento. O desenvolvimento foi tanto teórico quanto prático. Pelo lado teórico, em *paper* do começo da década de 80, Dubins & Freedman (1981) acrescentam importante nota ao trabalho de Gale & Shapley (1962) mostrando que, mantidas as preferências dos demais agentes, um determinado participante poderia “lucrar” ao não declarar suas verdadeiras preferências. Esse artigo ganhou notoriedade na comunidade científica dedicada ao assunto, pois esse resultado sinalizava possíveis comportamentos estratégicos e a manipulação dos algoritmos por parte dos agentes.

Na guisa desse mesmo desenvolvimento teórico, Roth (1982) confirma o resultado encontrado por Dubins & Freedman (1981) e estipula um teorema de impossibilidade: “*não há mecanismo de*

¹³ Tesler (1994) defende a importância do uso do conceito de *core* para a formação de economia desde a graduação e apresenta vários exemplos auxiliares a compreensão do conceito. Ressalta-se que, para a demografia da família, tal conceito é também extremamente útil, Becker (1991, p. 111) estuda a condição para a formação de casamentos ótimos, esta condição será apresentada na seção 3.2 do próximo capítulo. A definição de *core* pode ser útil também para temas de estudos migratórios que buscam conceituar a alocação ótima de migrantes. Não há um consenso sobre a tradução das palavras *core* e *nucleolus*, que em teoria dos jogos e de equilíbrio geral são conceitos distintos. Em uma tradução livre, Mas-Colell, Whinston & Green (1995, p.677) definem núcleo como “o conjunto de todos os resultados (*payoffs*) factíveis em que nenhuma coalizão pode, por si mesma, aumentar os resultados de todos os seus membros”. Ou seja, assim como a definição de casamentos, o *core* é uma situação estável em que não é possível para nenhum participante quebrar o *core*. O *núcleo* é uma definição mais estrita do que o *core*. Em determinados casos o *core* existe (é não vazio), porém, o conjunto de alocações que pertencem ao *core* pode ser grande o suficiente de modo que é impossível determinar uma única solução ou um conjunto de soluções manipuláveis (fáceis de serem trabalhadas), nesses casos o conceito de núcleo pode ajudar a reduzir o *core* para uma solução única. Outro conceito interessante para encontrar uma solução única para o *core* é o *valor-Shapley* (*Shapley-value*), definido por Shapley (1951). O conceito de *núcleo* não será mais explorado, pois foge ao escopo dessa tese, o *valor-Shapley* será brevemente mencionado no capítulo 4 a título de apontar futuros caminhos deste trabalho, mas não será aprofundado.

matching estável tal que declarar as verdadeiras preferências é a estratégia dominante para todos os agentes”. Gale & Sotomayor (1985a e 1985b) estendem as implicações do teorema e mostram que “*em qualquer mecanismo de casamentos estável nos quais as preferências são estritas e há mais de um resultado possível (o core não é único), ao menos um agente pode ganhar manipulando (misrepresenting) suas próprias preferências, assumindo que os outros agentes declaram as verdadeiras preferências*”.

Um resultado conhecido desde o artigo de Dubins & Freedman (1981) é que o grupo que propõe primeiro no algoritmo não possui incentivo em subdeclarar suas verdadeiras preferências, ou seja, para o grupo que propõe é interessante revelar as verdadeiras preferências e não há outra estratégia que seja tão boa quanto.¹⁴ Uma série de desenvolvimentos advindos desses textos iniciais da década de 80 surgem posteriormente. Um resultado particularmente importante vem de Demange Gale & Sotomayor (1987), nesse artigo são mostrados os limites para a manipulação dos algoritmos de pareamentos do *two-sided markets*.

Concomitante a esse grande desenvolvimento teórico, vastas descobertas empíricas e práticas também se descortinavam. Roth (1984a) e Gale & Sotomayor (1985b) concomitantemente documentaram que o algoritmo de pareamento de casamentos estáveis já vinha sendo aplicado desde 1951 para alocar estudantes de medicina residentes nos hospitais conveniados ao *National Residence Matching Program* (NRMP).¹⁵

Roth & Sotomayor (1990) consolidaram toda a literatura das décadas anteriores em um livro referência da área. Grande parte da literatura dos anos 60, 70 e 80 se encontra reunida neste volume. As aplicações do algoritmos no mercado de médicos residentes encontra-se

¹⁴ Os teoremas apresentados nesse parágrafo e no anterior serão melhor detalhados no capítulo 3, em 3.3.1.

¹⁵ O nome do programa era National Internal Matching Program (NIMP) e mudou para NRMP tempos depois. Em entrevista disponível na internet, Sotomayor conta que a aplicação e prova de que o NRMP era o mesmo algoritmo do artigo de Gale & Shapley (1962) foi descoberta pelo próprio professor D. Gale. Em 1983 Sotomayor e Gale submeteram o artigo Gale & Sotomayor (1985a) para publicação na *The American Mathematical Monthly* e compartilharam o resultado com Roth que também em 1983 referenciou o achado dos dois autores em seu artigo no *Journal of Political Economy* (Roth, 1984). Como o artigo de Roth tramitou mais rápido para a publicação, seu nome é muitas vezes identificado com a aplicação empírica do método. A confusão ocorre mesmo Roth tendo citado o achado original publicado por Gale & Sotomayor (1985a). A entrevista está disponível em: <http://livrosemanias.economico.sapo.pt/28798.html> (último acesso em 23/04/2013).

detalhadamente descrita no livro, bem como sugestões para aplicações dos métodos para se analisar o mercado de trabalho.

Após a publicação de Roth & Sotomayor (1990) a década de 90 assiste um grande avanço dos estudos e das aplicações dos algoritmos *deferred-acceptance* em diversas áreas, campo de desenvolvimento teórico metodológico que passou a se chamar *desenho de mercados*. No início dos anos 2000, o campo ganhou corpo, Roth (2002) apresenta uma revisão dos desenvolvimentos em vários mercados de alocação centralizada e descentralizada. O NRMP se tornou tão famoso que artigos passaram a referenciar o algoritmo simplesmente por essa sigla advinda do programa de residência dos médicos. Várias interseções surgiram com aplicações no desenho de leilões e formação de mercados para determinados objetivos. Uma das aplicações mais curiosas e de grande impacto social foi a aplicação de *matchings* para transplantes de rins (ver Roth 2012).

De especial interesse para os objetivos do presente trabalho são as aplicações do algoritmo na alocação de estudantes e escolas. E nisso alguns trabalhos publicados no Brasil são valiosos. Em 1996, Sotomayor (1996) analisa o sistema de alocação dos centros de pós-graduação em economia brasileiros, sistema ANPEC. Bardella (2005) em continuidade a este trabalho, analisa mais detidamente o funcionamento deste mecanismo e apresenta sugestões para sua otimização e maior adesão por centros periféricos.

Do trabalho de Bardella (2005) é interessante reportar a revisão da formação de mercados de admissão.¹⁶ O autor apresenta que tais mercados passam tipicamente por estágios bem marcados no que diz respeito a organização e tempo de processamento das propostas. Sobre a organização, os mercados de admissão podem ser descentralizados ou centralizados. Em um mercado descentralizado as decisões são tomadas individualmente pelos agentes, tal como em um mercado de livre concorrência comum, entretanto, nos mercados tipicamente estudados para o

¹⁶ O sistema de alocação de alunos e escolas pode ser pensado como um mercado. Cabe notar que no mercado privado existe uma regra mercadológica para decidir qual será a alocação dos alunos nas escolas. Nessa situação, pais e filhos decidem em qual escola o filho irá estudar por meio de um princípio utilitarista, que pode considerar uma série de fatores de conveniência para a família: localização, método pedagógico, tipo de aprendizado, amizades, custos da matrícula, e uma série de características desejáveis. A escola particular também pode ter por base um critério utilitarista de escolha, além da cobrança da matrícula, a aceitação dos alunos pode estar vinculada ao rendimento acadêmico em provas, histórico escolar, familiar e outras escolhas que sejam convenientes à direção da escola. Dessa forma, da combinação de interesses de pais, alunos e diretores, surge no mercado privado uma alocação entre alunos e escolas.

pareamento determinístico, existem incentivos para a centralização de algumas atividades, principalmente porque não há mecanismos de preços. É assim que Roth & Xing (1994, apud Bardella 2005) identificam uma tendência a centralização nesses mercados.

O sistema de matrículas de Belo Horizonte pré-1993 podia ser visto como descentralizado, pois apesar de existir uma coordenação central, na presença das secretarias de educação, os pais e alunos se dirigiam independentemente para escola que achavam mais conveniente, e as escolas, apesar de proibidas por lei, realizavam determinados acordos para matricular alguns alunos em detrimento de outros. Com isso surgiam todos os problemas já relatados na seção 2.1.

Um mercado centralizado possui regras coordenadas e em geral uma *clearinghouse* para determinar a alocação do mercado. Um sistema plenamente centralizado possui um agente central que coordena a alocação e possui poder coercitivo para fazer cumprir suas medidas (parecido com um sistema de matrículas do ensino básico). No que se pode chamar de sistema semi-centralizado, existe uma coordenação que estipula as regras do sistema, mas a alocação é tomada como uma ação individual entre as partes (a ANPEC pode ser encarada dessa forma).

Ressalta-se que mesmo no sistema centralizado, em muitos casos, é do interesse de todos a cantralização, pois ela serve para regular problemas de congestão e de filas do mercado e é bem vindo para se evitar o descumprimento de regras. O sistema de cadastro de BH não pode ser encarado como extremamente centralizado, dado que há alguma liberdade para a ação dos agentes no momento das matrículas.

Um sistema central ajuda a coordenar a alocação descentralizada em situações em que não há preços. Porém, tal sistema pode se desvirtuar para outro caminho, que é o de centralizar em excesso e diminuir a liberdade dos agentes. Bardella (2005) conclui, em sua revisão, que o estágio mais avançado de um sistema de alocação é aquele em que o sistema volta a ser descentralizado, mas de uma forma bastante ordeira, ou seja, não é preciso de um agente centralizador regulando todo o processo de alocação, é preciso somente que as regras estejam claras e os próprios indivíduos sigam, por interesse próprio, as regras previamente estipuladas no sistema. A

centralização funcionaria de uma forma mais branda, como forma de assegurar compromissos e punir eventuais burlas.

Sobre os estudos dos sistemas de matrículas existentes no Brasil, além do estudo de Fernandes (2007) para as matrículas de São Paulo capital, aqui mencionado mais detidamente na seção 2.6, outro trabalho realizado no Brasil de interessante aplicação é o estudo de Gontijo (2008) que analisou uma aplicação do algoritmo ao sistema de vestibular do estado de São Paulo, que era bastante descentralizado. Observa-se que este estudo foi realizado antes da aplicação do ENEM (Exame Nacional do Ensino Médio) em escala nacional, ou seja, antes da sua adoção como boa parte dos vestibulares em universidades públicas do ensino superior.

O atual SiSU (Sistema de Seleção Unificada) unifica a demanda dos estudantes que prestaram o ENEM e centraliza as ofertas de vagas das IES (Instituições de Ensino Superior). No entanto, o SiSU permite apenas a escolha de duas faculdades pelo estudante, o que é uma grande limitação. Seguindo algumas conclusões do trabalho de Bardella (2005) e Gontijo (2008) pode se sugerir que o sistema SiSU teria muito mais adeptos e geraria muito mais benefícios caso permitisse aos alunos um leque maior de opções. Estudos nesse tema são promissores.

Como são de particular interesse, apresenta-se aqui os estudos de caso de quatro outras cidades importantes para o entendimento mais completo dos sistemas de matrículas de escolas e alunos. Essas cidades são: Boston (EUA), que possui uma lição importante de um mecanismo que era instável. Nova York (EUA), que tem uma lição interessante sobre a competição saudável que o mecanismo pode gerar, a cidade-estado de Singapura, em que se reporta brevemente os resultados do *paper* de Teo, Sethuraman & Tan (2001) e os problemas da escassez de informação na escolhas das preferências e, novamente, o caso de São Paulo, cidade que possui mecanismo de alocação similar ao de Belo Horizonte e para isso se revisa brevemente o trabalho de Fernandes (2007).

2.2.1. O sistema de alocação de Boston (EUA)

O sistema da cidade de Boston se tornou um estudo de caso importante da literatura de *school choice*, ligada aos desenhos dos mecanismos de alocação de alunos e escolas. Abdulkadiroğlu & Sönmez (2003), Abdulkadiroğlu, Pathak, Roth & Sönmez (2005) estudaram em detalhes o mecanismo de alocação da cidade e descreveram algumas das dificuldades que pais e alunos passavam com o antigo mecanismo, que estava em operação desde 1999.

O mecanismo de Boston consistia no seguinte princípio: alocar os alunos pelas suas ordenações de preferências. Isso parece semelhante à proposta de Gale & Shapley (1962), no entanto, diferentemente do algoritmo *deferred-acceptance*, antes mesmo do encerramento de todas as n etapas do algoritmo, uma escola que preenchesse todas as suas vagas era retirada do sistema, não podendo receber nova proposta.

Isso muda fundamentalmente o algoritmo e leva a consequências indesejáveis. Para se compreender melhor o sistema que estava em vigor em Boston, vejamos um passo a passo do algoritmo conforme descrito em Abdulkadiroğlu & Sonmez (2003) e Fernandes (2007):

Passo 1. Cada estudante (pais ou responsáveis) submete ao *Boston School Committee* uma lista de preferências com o *ranking* de, no mínimo, três escolas.¹⁷

Passo 2. Cada escola tinha uma ordem de prioridades (preferências) estabelecida para os alunos, respeitando-se a seguinte hierarquia.

- a. Primeira prioridade: O aluno possui irmãos estudando na escola e estão dentro da *walking zone*.¹⁸
- b. Segunda prioridade: Possui irmãos estudando na escola mas não são da *walk zone*.

¹⁷ Os autores não mencionam se há um número máximo de escolas. Uma lista truncada costuma a ter consequências importantes para alocação de alunos tal como analisado em Bardella (2005) e em diversos outros estudos. No entanto, no caso do algoritmo antigo de Boston outros problemas emergem e a falta dessa informação não compromete a análise do estudo de caso da cidade.

¹⁸ As *walking zones* são correspondentes às jurisdições escolares de Belo Horizonte, mencionados em 2.1. Entretanto, em Boston, essas zonas alteram o seu raio de abrangência para etapas de ensino diferentes. Fernandes (2007, p. 21) descreve que o raio de ação para o ensino infantil é de 1,6 quilômetros, 2,4 km para o correspondente ao ensino fundamental brasileiro, e 3,2 para o ensino médio.

- c. Terceira prioridade: vivem dentro da *walk zone* da escola.
- d. Quarta prioridade: todos os outros estudantes.

Passo 3. Etapa de pareamento seguindo os seguintes passos:

Passo 3.1. As escolas consideram todos alunos que a colocaram em primeiro lugar no *ranking*. Seguindo a regra de prioridades do **passo 2** a escola aloca todos os alunos que a elencaram como primeira escolha. Ao final do processo há duas possibilidades: ou não há mais vagas faltantes ou não há mais alunos que listaram a escola em primeiro lugar.

Passo 3.2. Para os estudantes que não foram alocados no **passo 3.1**. As escolas que ainda possuem vagas (ou seja, não completaram todas as vagas com alunos que a listaram em primeiro lugar no **passo 3.1**) consideram todos os alunos que a listaram como segunda opção e, seguindo o critério de prioridades do **passo 2**, aloca os estudantes. Também ao final deste passo ou as escolas completam todas suas vagas disponíveis ou não há mais alunos que priorizaram a escola em segundo lugar.

Passo 3.k. Esse é um passo genérico, ' k ' é o número da interação atual. Considerando os estudantes que não foram alocados nos passos anteriores. As escolas que ainda possuem vagas consideram todos os alunos que a listaram como a k -ésima opção e prosseguem alocando os estudantes segundo a ordem de prioridades. Ao final do passo, ou a escola fica sem vagas ou sem alunos que a colocaram como k -ésima opção.

Passo 4. O algoritmo encerra no k -ésimo passo em que todos os alunos são alocados.

O algoritmo acima apresentava problemas conhecidos dos moradores da cidade, pois os pais sabiam que era preciso “jogar” com o sistema. Nem sempre era compensador revelar as verdadeiras preferências. Alguns pais preferiam declarar a sua segunda ou terceira escola como a primeira opção. O exemplo abaixo ajuda entender o porque dessa manipulação:

Exemplo.1: (exemplo de Fernandes, 2007, p. 22) Imagine um sistema simples com apenas três escolas: $S = \{s_1, s_2, s_3\}$; e três alunos: $I = \{i_1, i_2, i_3\}$. Suponha que cada escola possua apenas uma vaga. A ordem de prioridades das escolas é a seguinte:

$$\mathcal{P}(s_1) = i_2, i_1, i_3$$

$$\mathcal{P}(s_2) = i_3, i_2, i_1$$

$$\mathcal{P}(s_3) = i_2, i_3, i_1$$

O $\mathcal{P}(s_n)$ significa o ordenamento de preferências de determinada escola s_n . A ordenação da primeira escola, por exemplo, diz que ela prefere primeiro o aluno 2, em seguida o aluno 1 e por último o aluno 3. O capítulo 3 tratará da notação em maiores detalhes, por ora é suficiente tomar o ordenamento como uma informação importante. O ranking de preferências dos alunos é o seguinte:

$$\mathcal{P}(i_1) = s_2, s_1, s_3$$

$$\mathcal{P}(i_2) = s_2, s_3, s_1$$

$$\mathcal{P}(i_3) = s_1, s_2, s_3$$

Aplicando os passos do sistema de Boston, de **3.1** a **3.k**, enunciados acima, teríamos no **primeiro passo** que a escola s_1 considera apenas o aluno i_3 (que a *rankeou* em primeiro lugar) e cede a vaga para esse aluno. A escola s_2 considera os alunos i_1 e i_2 (os dois rankearam a escola 2 em primeiro lugar) e de acordo com suas prioridades, confere a vaga ao aluno i_2 . A escola s_3 não teve nenhum aluno a declarando como primeira escolha. No primeiro passo ela continua com vagas mas sem alunos.

Como é a única com vagas restantes, a escola s_3 segue para o **segundo passo**, o único aluno que a colocou como segunda escolha (i_2) já está alocado e portanto ela não possui nenhum aluno com quem preencher sua vaga.

Finalmente, no **terceiro passo**, a escola s_3 considera os alunos que a elencaram como terceira escolha, que são i_1 e i_3 , como i_3 já está alocado para s_1 , s_3 concede sua vaga para i_1 e o algoritmo pára.

Ao final deste processo nota-se que os alunos i_2 e i_3 conseguiram sua melhor vaga, e i_1 ficou com sua terceira opção. O pareamento final (representado por μ) é:

$$\mu = \begin{array}{ccc} s_3 & s_2 & s_1 \\ i_1 & i_2 & i_3 \end{array}$$

A forma de representação da alocação final acima indica os pares formados: (i_1, s_3) , (i_2, s_2) e (i_3, s_1) . Como dito anteriormente, o algoritmo de Boston permite a manipulação, caso o aluno saiba jogar com as informações fornecidas ao sistema. O aluno i_1 pode falsear suas preferências e dizer que prefere a escola s_1 em primeiro lugar (que é na verdade sua segunda escola em preferência). Com isso, caso os demais alunos continuem declarando as verdadeiras preferências, o resultado final do algoritmo (μ' por conta das preferências alteradas) seria:

$$\mu' = \begin{array}{ccc} s_1 & s_2 & s_3 \\ i_1 & i_2 & i_3 \end{array}$$

O resultado acima pode ser alcançado aplicando-se os passos do algoritmo novamente. Nota-se que ao mascarar suas verdadeiras preferências, o aluno i_1 conseguiu manipular o resultado a seu favor. Nota-se ainda que no caso do exemplo 1 acima, i_3 poderia também maquiagem suas preferências para conseguir o mesmo efeito a seu favor. Caso i_3 falseasse suas preferências para $\mathcal{P}'(i_3) = (s_2, s_1, s_3)$, e o primeiro aluno continuasse falseando com $\mathcal{P}'(i_1) = (s_1, s_2, s_3)$ e i_2 continuasse com as verdadeiras preferências, $\mathcal{P}(i_2) = (s_2, s_3, s_1)$, o algoritmo final seria (μ''):

$$\mu'' = \begin{array}{ccc} s_1 & s_3 & s_2 \\ i_1 & i_2 & i_3 \end{array}$$

Todos os jogadores com sua segunda melhor preferência. Para a alocação acima, i_2 não conseguiria mais alterar o algoritmo a seu favor (isso se i_1 e i_3 continuassem manipulando). Porém, o importante do algoritmo de Boston é que ele carrega uma lição importante: um bom sistema deve evitar a manipulação dos resultados, pois isso pode prejudicar sua estabilidade e gerar insatisfação na população.

Abdulkadiroğlu, Pathak, Roth & Sönmez (2005) realizaram um levantamento de que os pais e alunos realmente jogavam estrategicamente com o algoritmo. Duas estratégias mais características surgiram marcadamente neste estudo: a primeira consistia em declarar uma escola muito popular como primeira escolha e colocar uma boa escola, mas que os pais sabiam ser menos procurada como a segunda escolha, numa estratégia do tipo: “*se conseguirmos, ótimo, caso contrário, ficamos com a segunda opção de segurança*”. Com essa solução, pais passavam bastante tempo procurando uma boa escola que pudesse estar subestimada para declará-la como uma boa opção, contribuía ainda mais para esse processo, a divulgação de uma relação da procura pelas escolas no ano anterior. Os pais tinham que jogar com as expectativas que possuíam sobre a possível ação dos outros pais, adivinhar o comportamento dos outros.

Uma segunda estratégia dos pais, de maior segurança, é desistir da escola mais preferida e popular e declarar como primeira opção uma escola provável de ser obtida. Havia também pais que não jogavam estrategicamente e declaravam as preferências verdadeiras, porém, como vimos no exemplo 1, essa situação poderia levar a uma alocação ruim na situação em que há outros agentes jogando.

Após a publicação dos resultados de Abdulkadiroğlu & Sonmez (2003), Abdulkadiroğlu et. al. (2005) e outros trabalhos apontando achados similares, um grande debate público sobre as possibilidades de alocação de alunos e escolas surgiu nas escolas de Boston. Ainda hoje o sistema é avaliado para aperfeiçoamentos. O atual sistema em vigor se parece mais com um algoritmo que considera as preferências dos alunos segundo regiões de moradia, e as prioridades das escolas de acordo com notas do MCAS (*Massachusetts Comprehensive Assessment System*). O sistema oferece mais previsibilidade para os pais e oferece menor chance para as manipulações intencionais.¹⁹

Tal como apontado por Fernandes (2007, p. 23): “uma das principais razões que explicam o fato do mecanismo de Boston ser manipulável reside justamente na baixa importância dada às

¹⁹ Informações do processo atual podem ser encontradas em: <http://bostonschoolchoice.org/explore-the-proposals/home-based-proposal/> (último acesso 24/04/2013).

prioridades. Um aluno que possui alta prioridade para uma escola ‘s’ pode perdê-la para outro aluno menos prioritário caso a escola não esteja listada como sua escola predileta”.

A correção do algoritmo de Boston foi um uso prático da literatura para a correção de um problema específico e que visava a retificar o erro acima enunciado. As prioridades perdiam a importância, dado que uma mesma escola não poderia ser novamente proposta por um aluno que a priorizasse em uma posição mais desfavorável, mas ainda assim fosse o melhor aluno para aquela escola.

Outras lições que podem ser tiradas do *paper* de Abdulkadiroğlu & Sonmez (2003) são as de que as preferências das escolas podem ser encaradas como prioridades. Para a matemática realizada pelos modelos, os dois conceitos (preferências ou prioridades) são intercambiáveis. No entanto, se considerarmos que todos os alunos das escolas públicas possuem os mesmos direitos, e que, sendo assim, nenhum aluno pode ser preferido a outro, mas sim possuir uma prioridade segundo critérios de alocação e de administração pública, a adoção de prioridades é uma boa ideia.

2.2.2. O sistema de alocação de Nova York (EUA)

Outra cidade importante dos Estados Unidos que é um estudo de caso interessante e que levanta uma lição diferente daquela presente no algoritmo de Boston pré-2005 é a cidade de Nova York. Segundo Abdulkadiroğlu, Pathak & Roth (2005), Nova York é o maior sistema municipal de matrículas dos Estados Unidos e desde 1965 possuía um sistema descentralizado de realização de matrículas, com algumas alterações ao longo das décadas posteriores. Nos anos 90, o sistema da cidade passou a ser centralizado pelo NYCDOE (*New York City Department of Education*), ou simplesmente DOE, e no começo dos anos 2000 uma série de reformas haviam sido sugeridas para o sistema.

A motivação para as mudanças do algoritmo de Nova York são similares aos objetivos propostos desta tese. Como colocam Abdulkadiroğlu, Pathak & Roth (2005): uma das preocupações do DOE era de que alunos de vizinhanças desvantajadas não fossem automaticamente matriculados em escolas de nível socioeconômico mais baixo.

A diferença para Belo Horizonte, no entanto, é que Nova York possuía uma série de programas e critérios para a seleção dos alunos (escolas especializadas, *screened programs* e *unscreened programs*). NY possuía mais de 500 programas e o algoritmo de seleção ocorria com as escolas propondo aos alunos, mas tinha apenas três rodadas.

Os alunos escolhiam os programas ou escolas preferidos e podiam listar no máximo 5 escolhas. Abdulkadiroğlu, Pathak & Roth (2005) mencionam que isso gerava congestionamento no sistema de escolhas. Sendo que com os múltiplos critérios de seleção apenas 50 mil alunos, dos 90 mil existentes, recebiam propostas no primeiro estágio e 17 mil deles recebiam mais de uma proposta. E na conclusão do processo, 30 mil alunos acabavam sendo matriculados em alguma escola que não constava em sua lista inicial de preferências.²⁰

Por ter preferências truncadas e um sistema complexo de propostas e aceites, o mecanismo de Nova York possuía três problemas principais que sugeriam a mudança para o DOE: 1) Os 30 mil alunos que acabavam sendo matriculados em escolas não listadas preocupavam o departamento; 2) alunos agiam estrategicamente, um dos critérios de seleção dos *screened programs* era classificar os alunos pelas notas, alunos que não tinham obtido boas notas eram estimulados a não revelar suas verdadeiras preferências, mas sim procurar uma alternativa mais segura (onde teria maior probabilidade de ser chamado); 3) As escolas também agiam estrategicamente em Nova York. A maneira que as escolas possuíam para desvirtuar o sistema era declarar número de vagas diferente da sua real possibilidade, isso porque escolas muito procuradas sabiam que conseguiriam preencher quase todas suas vagas com os convites, dessa forma, declarar número menor de vagas permitia a essas escolas fechar suas vagas mais cedo sem precisar receber alunos do sistema de zoneamento.

Após aconselhamento do DOE com os autores do artigo acima relatado, chegou-se a conclusão de que o novo algoritmo da cidade poderia se dar, em linhas gerais, da seguinte maneira:

Passo 1. Os alunos submetem uma lista de no máximo 12 escolas e as escolas formulam suas preferências de acordo com um *ranking* de notas (geralmente as escolas

²⁰ Os alunos que não recebiam um convite eram matriculados aleatoriamente por um sistema de zoneamento.

especializadas utilizam esse critério) ou um *ranking* aleatório (geralmente usado por escolas não especializadas).

Passo 2. Cada estudante aplica para sua escola melhor classificada. A escola rejeita os alunos que não estão por ela listados e “segura” aqueles melhor *rankeados* da sua lista, preenche até o limite de sua capacidade e rejeita aqueles que ficaram abaixo do limite de vagas no *ranking*.

Passo 3. Os estudantes rejeitados no **passo 2** aplicam para sua próxima melhor escolha. As escolas que receberam propostas realizam o mesmo procedimento do passo anterior. Neste passo, uma observação se faz necessária: não necessariamente há uma “próxima melhor escola”, os alunos poderiam submeter até uma dúzia de melhores escolas, mas não eram obrigados a listar todas as 12. Caso um aluno listasse apenas uma escola e fosse rejeitado no **passo 2**, ele seria pareado depois administrativamente.²¹

Passo 4. O algoritmo termina quando não há mais rejeições a serem realizadas e cada escola é pareada à turma de alunos que ela estava “segurando” até o fim do algoritmo.

O sistema de Nova York possui mais uma série de detalhes não elucidados nos passos acima. Esses detalhes envolvem algumas escolhas administrativas que o departamento de educação precisou realizar para garantir critérios de distribuição de escores. Existe um programa chamado *Educational Option (EdOpt)* que permite a escola avaliar os alunos individualmente para ocupar metade de suas vagas, com a restrição de que 16% devem ser de percentis superiores da distribuição de um teste padronizado de inglês, 68% devem pertencer à categoria intermediária e outros 16% a categoria dos percentis inferiores. Além disso, um estudante que esteja nos 2% superiores e opte por entrar em uma escola com *EdOpt* deve ser priorizado e matriculado automaticamente.

²¹ Da leitura de Abdulkadiroğlu, Pathak & Roth (2005) abstrai-se que esse não deveria ser um caso frequente, porém, não há maiores detalhes de como seria esse processo administrativo no mencionado artigo. Nesse passo a passo, há uma série de detalhes do algoritmo de Nova York que estão omitidos por simplicidade de exposição. O leitor interessado em mais detalhes pode consultar Abdulkadiroğlu et. al. (2005) ou o site da Secretaria de Educação do Estado de Nova York: <http://schools.nyc.gov/ChoicesEnrollment/ChangingSchools/default.htm> (último acesso 24/04/2013).

Essas características acrescentam detalhes ao algoritmo de Nova York, e algumas dessas restrições diminuem um pouco a eficiência do algoritmo em conseguir uma alocação estável (conceito que se define mais rigorosamente no capítulo 3), porém essa é uma escolha do NYCDOE e dos cidadãos de Nova York.

Como resultados já do primeiro ano do novo algoritmo proposto, em 2006, obteve-se um resultado de mais de 70 mil alunos alocados para sua primeira escolha, 20 mil a mais do que com o mecanismo anterior. Cerca de 7.600 alunos não foram pareados em uma primeira etapa, mas puderam submeter nova listagem de 12 escolas, dentre escolas que ainda possuíam vagas. Ao final, aproximadamente 3.000 alunos não receberam uma escola que haviam escolhido, enquanto no sistema anterior eram 30 mil alunos nessa situação (alunos que recebiam escolas apenas pelo zoneamento).²²

A mudança do algoritmo de Nova York pode ser considerada bem sucedida. Esse caso traz algumas lições importantes para a área de *two-sided markets*. Uma importante lição conhecida desde os trabalhos de Roth & Sotomayor (1990) e o NRMP, é que algoritmos com estudantes propondo levam a melhores resultados. Favorecer instituições pode levar a um resultado estável mas que não é o melhor possível para os estudantes. Como instituições são neutras, e em tese não devem possuir preferências rígidas, ou favorecimentos desequilibrados para determinadas pessoas, os algoritmos em que os alunos propõem para as escolas geram melhor resultado social.²³

Outra lição interessante do algoritmo de NY é que escolas também agem estrategicamente. As escolas agiam assim ao estipularem um número de vagas menor do que o real possível, pois assim evitavam entrar no circuito aleatório do sistema anterior. A curta conclusão de Abdulkadiroğlu, Pathak & Roth (2005, p. 367) pode ser colocada aqui integralmente para ilustrar esse ponto:

²² Abdulkadiroğlu, Pathak & Roth (2005) contam que pouco mais de 5.100 estudantes apelaram da decisão do novo processo, destes, 2.600 foram tratados caso a caso. Curiosamente, 300 apelações vieram de alunos que receberam sua primeira escolha. Os autores julgam que isso pode se dever a falhas de informação e a dificuldades de comunicação e implementação do primeiro ano de mudança.

²³ Pode se usar a analogia de que alunos são demandantes e escolas são ofertantes. Nesses termos, o que está se dizendo aqui é que, na dúvida, é melhor beneficiar o consumidor, conferindo-lhe maior excedente.

“New York City needs more good schools. But for a given stock of school places, more students can be admitted to schools they want if the matching process is free of congestion, so that students’ preferences can be fully taken into account. The new clearinghouse, organized around stable matching mechanism, has helped relieve the congestion of the previous offer/acceptance/wait-list process and provides more straightforward incentives to applicants”.

Em desenvolvimentos posteriores do algoritmo de Nova York, Abdulkadiroğlu, Pathak & Roth (2009) reportam que um resultado importante do novo algoritmo é que ele aumentou a competição entre as escolas públicas da cidade. Um resultado reportado é que escolas muito impopulares foram fechadas e as escolas consideradas boas puderam aumentar o número de vagas ofertadas. Ao contrário do sistema anterior, em que as melhores escolas reportavam um número inferior de vagas, no novo sistema as boas escolas procuravam medir bem sua verdadeira capacidade e tentavam aumentá-la.

Por fim, Abdulkadiroğlu, Pathak & Roth (2009) consideram que uma das principais vantagens do novo algoritmo é ter aumentado o número máximo de opções de 5 para 12. Esse resultado é bastante similar à recomendação de Bardella (2005) para ANPEC e Fernandes (2007) para o sistema de cadastro de São Paulo. Esse é também o principal ponto de Teo, Sethuraman & Tan (2001) que estudaram em detalhes o sistema de Singapura, tema da próxima subseção.

2.2.3. O sistema de alocação de Singapura²⁴

As descrições desta seção se baseiam em Teo, Sethuraman & Tan (2001) e podem estar desatualizadas em relação ao atual sistema. Entretanto, isso não compromete os objetivos da descrição do algoritmo para a revisão de literatura.

O ano letivo em Singapura é similar ao brasileiro: se inicia no começo do ano (no início de Janeiro) e se encerra ao final do ano (perto de dezembro). No sistema de Singapura, há um teste nacional, o *Primary School Leaving Exam* (PSLE), que é realizado quando os alunos atingem o

²⁴ Caso o leitor esteja se perguntando se Singapura é com ‘S’, ou com ‘C’, a nova norma ortográfica recomenda a escrita do nome da cidade-estado asiática com S pelo novo acordo (decreto Nº 6.583, de 29 de Setembro de 2008). “Cingapura” consta nas antigas regras ortográficas adotadas pelo Brasil e ainda hoje se encontra a forma antiga em textos recentemente publicados.

final da sexta série, que compreende as idades de 11 a 12 anos. O resultado desse exame classifica os estudantes e essa classificação é utilizada para admissão em escolas do ensino secundário (grosso modo, essa etapa de ensino correspondente ao antigo segundo grau brasileiro, o que seria, no sistema atual, o correspondente às etapas do 6º ao 9º ano do fundamental).

Antes de obterem o resultado dos exames do PSLE, os pais devem submeter ao ministério da educação (MOE, *Ministry of Education*) uma lista de seis escolas preferidas. Essa lista é solicitada em agosto. O resultado do PSLE sai no final de Novembro. O fato dos pais escolherem as seis escolas antes do resultado do teste indicar quais as reais chances dos alunos em cada escola se constitui um problema, conforme relatado por Teo, Sethuraman & Tan (2001).

Em Singapura há dois tipos de escolas especiais: as independentes (*independent schools*) e as de *Special Assistance Plan* (SAP). As escolas especiais são escolas de elite. Após os resultados dos exames, os 10% dos alunos com melhores notas tem direito a mandar uma segunda listagem dizendo em qual escola especial eles desejam se matricular. Esses alunos são priorizados e as escolas especiais matriculam apenas esse tipo de aluno.

Abaixo são descritos os passos do algoritmo de Singapura tal como funcionava no início dos anos 2000. Foram acrescentadas algumas mudanças em relação ao artigo de Teo, Sethuraman & Tan (2001) para maior facilidade da exposição:

Passo 1. Considere todos alunos elegíveis (top 10% no PSLE) que listaram as escolas independentes como primeira opção na segunda listagem (a que ocorre após resultado do exame em novembro) e matricule os alunos nessas escolas.

Passo 2. Considere todos alunos elegíveis (top 10% no PSLE) que listaram as escolas SAPs como primeira opção na segunda listagem e matricule os alunos nessas escolas.²⁵

²⁵ Não é mencionado no artigo de Teo et. al. (2001), mas ao que parece, as escolas especiais sempre possuem vagas para atender os alunos top 10%. Essa previsão deve ser fácil de ser manejada caso o número de matrículas se apresente estável ou previsível ano a ano.

Passo 3. Considere todos os alunos que listaram como primeira opção (da lista de seis escolas, enviada antes da divulgação do PSLE) uma escola do segundo nível afiliada à escola do nível primário da qual o aluno provém e conceda prioridade de matrícula a esse aluno.²⁶

Passo 4. Considere os alunos restantes um a um, de acordo com o *ranking* descendente de notas do PSLE e conceda a eles uma vaga de acordo com a primeira listagem de preferências submetidas em agosto.

Passo 5. Caso o aluno não seja alocado em algum dos quatro passos anteriores, a alocação final é feita manualmente pela região de residência do aluno, tentando-se alocá-lo para escola mais próxima que possua vagas.

O mecanismo de Singapura possui problemas similares a aqueles reportados pelo sistema de Boston pré-2005, quais sejam: pais possuem incentivos para mascarar suas verdadeiras preferências. Esse problema ocorre muito por conta dos alunos não conhecerem o seu tipo (sua nota no PSLE) antes de submeterem a listagem de escolas. Tal como no algoritmo de Boston, há uma estratégia de segurança: alunos que estudaram em uma escola afiliada ao ensino médio no primário, possuem grandes incentivos a reportar a sua escola atual como primeira opção, mesmo que ela não seja a melhor opção nas verdadeiras preferências.

Essa estratégia é vantajosa, pois caso o aluno se saia bem no PSLE ele poderá ainda submeter uma segunda listagem e garantir sua vagas nas escolas especiais. Caso ele não se saia bem no PSLE, digamos que obtenha um resultado ruim, ainda assim poderá contar com boas chances na escola afiliada à escola em que já estudava.

²⁶ Em paralelo com o sistema brasileiro, esse sistema se parece com as escolas brasileiras que possuem tanto ensino fundamental e médio. Se esse sistema fosse implementado no Brasil, seria como se a escola priorizasse, antes das notas do PSLE, os alunos que já estudavam em suas turmas do ensino fundamental. Teo et. al. (2001) argumentam que isso ocorre por motivos históricos em Singapura, para proporcionar maior integração entre os níveis primário e secundário. Observam também que, na verdade, esse passo confere maior importância à “afiliação” do que o resultado do PSLE. Entretanto, nem todos os estudantes possuem essa chance, pois há escolas sem afiliação, ou em que o número de vagas no secundário é menor do que o número de vagas do primário. E nesse caso, os melhores alunos do ranking PSLE que listaram aquela escola conseguem as vagas, os demais “caem” no **passo 4**.

Porém há problemas: caso o aluno se saia bem no PSLE, mas não bem o suficiente para ingressar nas escolas especiais, listar a escola afiliada como primeira opção pode limitar sua chance de conseguir uma escola que julgue melhor. Nesse caso, não revelando suas verdadeiras preferências, ele não consegue maximizar o resultado.

Uma situação igualmente não desejável é o aluno que estudava em uma escola primária sem afiliação, ou com número de vagas insuficiente. O aluno que se sai mal no PSLE possui grandes chances de ser matriculado em uma escola totalmente fora de suas considerações. Mesmo com o critério das distâncias, o número de vagas faltantes pode ser insuficiente e os alunos podem se situar em escolas distantes e ruins em termos de ensino (o pior dos mundos).

Outro problema levantado no sistema de Singapura é que, tal como no sistema de Nova York, poderia ocorrer a congestão, caso o MOE atendesse aos pedidos dos pais para declararem as preferências após o resultado do PSLE. Isso porque o resultado do exame era liberado em novembro e o período de matrículas para o ano seguinte tinha de estar decidido nas primeiras semanas de dezembro. Seriam apenas duas semanas para os pais mandarem lista, o algoritmo gerar as alocações e os pais efetuarem matrículas. No começo dos anos 2000, o ministério da educação descartava essa possibilidade.

Teo, Sethuraman & Tan (2001) propuseram em seu artigo um algoritmo *deferred-acceptance many to one* nos moldes do modelo de Gale & Shapley (1962). Os autores exploram as possibilidades de burla do algoritmo com alunos propondo, ou escolas propondo. Como visto para Nova York, quando alunos propõem, as escolas teriam incentivos a mascarar suas verdadeiras preferências, o que poderia ser obtido revelando número incorreto de vagas. No caso de escolas propondo, os alunos poderiam desviar, declarando preferências manipuladas.

Uma lição importante do estudo de Singapura é que mesmo no sistema com escolas propondo, e permitindo-se os alunos mudarem a ordem das preferências para conseguir o máximo ganho, observou-se por simulações que a porcentagem de alunos que ganhavam com o *cheating* (trapaceando as preferências) não era muito superior a 10%, e tal número tinha a tendência de diminuir a medida que se inseriam mais escolas no processo. Ou seja, com esse resultado

simulado, o algoritmo pareceria ser, se não totalmente estável, estável a ponto de apenas 10% dos alunos se beneficiarem declarando preferências diferentes das verdadeiras.

Uma tabela de dados interessante sobre Singapura reporta quais são os fatores preponderantes para escolha das escolas pelos pais. Em um questionário em que pais podiam marcar mais de uma opção, apenas 30% dos pais de alunos estudantes da cidade-estado declararam que a distância de casa até a escola é um fator importante na decisão. O fator que ganhou mais importância no critério de decisão dos pais foi a qualidade dos professores, com 88%, seguida de disciplina escolar (com 82%). A TAB. 1 mostra esses resultados:

Tabela 1 – Quais fatores são considerados mais importantes pelos pais

<i>Quesito</i>	<i>Proporção dos respondentes</i>
Qualidade dos Professores	88%
Disciplina escolar	82%
Qualidade do diretor	58%
Reputação da escola	50%
Abrangência de recursos na escola (<i>facilities</i>)	35%
Distância da escola	30%
Amigos na escola	30%

Fonte: Teo, Sethuraman & Tan (2001, p.1263).

Obs: O resultado original foi publicado no *The Straits Times*, 19 de agosto de 1992. O *The Straits Times* é um jornal de Singapura, escrito em inglês e de larga circulação. Infelizmente Teo et. al. (2001) não fornecem maiores detalhes, não é possível saber se os pais puderam marcar duas ou três opções, nem detalhes da amostra da pesquisa. O autor tentou localizar detalhes sobre a matéria, mas ela parece estar acessível apenas em bibliotecas de Singapura.

Teo, Sethuraman & Tan (2001, p.1264) levantam ainda um problema pedagógico, segundo os autores: um ranking único para todas as escolas a partir das notas do PSLE limitava a avaliação mais completa dos alunos, colocando um peso excessivo no resultado acadêmico. E ainda enfraqueceria o poder das escolas de construir uma identidade própria de acordo com interesse e composição de alunos desejada. Nas palavras dos autores:

“Thus, elaborate mechanisms are now being put in place to assess the students in an all-round fashion. In the future, it will no longer be possible to have a single list that ranks all the students according to merit. Also, given that the schools are being encouraged to develop their own identity and strength, it is highly debatable whether the schools will all rank the students in the same way”.

Após esse trabalho, o sistema de Singapura mudou algumas de suas características, a principal delas é que agora a lista de escolas é realmente submetida após os resultados dos exames.²⁷ Em grande parte isso se deve à informatização do sistema de submissão das listas de preferências, que podem também ser enviadas institucionalmente pelas escolas. As escolas especiais continuam recrutando da maneira antiga. Não se encontrou registro de que o algoritmo *deferred-acceptance* está sendo utilizado, todavia, o fato de submeter a lista após o resultados dos exames ameniza o problema do **passo 3** do algoritmo anterior e pode fazer com que o **passo 4** funcione tal como o *deferred-acceptance* para as seis opções listadas.

2.2.4. O sistema de alocação de São Paulo (Brasil)

A descrição do algoritmo de alocação de São Paulo será dada com base na exposição de Fernandes (2007, pp. 53-56). Como enfatizado na seção 2.1, o mecanismo de São Paulo teve início em 1995 (dois anos após a implantação do sistema de Cadastro de Belo Horizonte), e possui algumas similaridades com o mecanismo de BH, pois os períodos e os procedimentos são semelhantes. Uma grande diferença se configura na escala do sistema público de ensino de São Paulo que é superlativo em relação a qualquer cidade do mundo, com mais de 320 mil alunos/ano a serem pareados pelo sistema de matrículas.

Fernandes (2007) identifica quatro fases importantes do mecanismo de São Paulo:

1. Programação da oferta de vagas;
2. Cadastramento dos alunos (ou matrícula antecipada);
3. Compatibilização demanda e oferta; e
4. Efetivação da matrícula.

²⁷ Informações disponíveis no site do MOE de Singapura: <http://www.moe.gov.sg/education/admissions/secondary-one-posting/> (último acesso 10/05/2013).

A primeira fase consiste de um levantamento das vagas disponíveis. Na segunda fase os alunos se inscrevem nas instituições ou em postos de inscrição espalhados pela cidade. Nessa etapa eles fornecem o endereço e outras informações úteis: nome, sexo, data de nascimento, raça, nomes dos pais ou responsáveis. No caso de alunos de ensino médio, os inscritos listam três opções de escolas em ordem de preferências. Alunos do ensino infantil que irão ingressar no fundamental informam apenas os dados cadastrais.

Na terceira fase, os alunos inscritos são pareados de acordo com a distância em um critério de proximidade do endereço de moradia fornecido e a escola mais próxima que possui vagas. A alocação é coordenada em parceria entre ensino da rede municipal e estadual. Esse mecanismo funciona de forma parecida com o de Belo Horizonte, que será descrito em 2.2.5, porém com regiões delimitadas por microrregiões da cidade de São Paulo. No caso do ensino médio, a escola elenca os alunos que a colocaram como primeira opção, juntamente com os que não marcaram opções mas para a qual a escola em questão seja a mais próxima. A escola ordena os alunos do mais velho até o mais novo, preenchendo todas as vagas. Rodadas sucessivas ocorrem para as escolas que possuem ainda vagas e o algoritmo pára quando todos os alunos são alocados.

Na quarta fase os resultados finais da fase 3 são inseridos no sistema informatizado do governo estadual e divulgados publicamente. Pais e alunos realizam as matrículas nas escolas recomendadas pelo sistema e eventuais correções podem surgir nessa fase.

Fernandes (2007) ressalta que o mecanismo de alocação de São Paulo, para o caso do ensino médio, funciona tal como o algoritmo de Gale & Shapley (1962) truncado para a terceira opção. No estudo o autor mostra exemplos de que é possível se conseguir melhorias de Pareto em relação ao atual algoritmo de São Paulo, e realiza simulações comparativas entre o mecanismo atual, a proposta de *deferred-acceptance* (DA), e a proposta de *top trading cycles* (TTC).

A principal variável de resultado observada por Fernandes (2007) é a posição que os alunos conseguem após a utilização do algoritmo. Os resultados foram simulados para as preferências das escolas e dos alunos por meio da geração de números aleatórios. Determinados padrões foram seguidos para se obter maior proximidade das simulações com as situações reais do município. O

estudo de Fernandes mostrou que os resultados simulados de matrículas podem melhorar bastante com a utilização do algoritmo DA ou do TTC. Principalmente para o caso do ensino fundamental, em que o *matching* se dá somente pelas distâncias.

Em relação ao trabalho de Fernandes (2007), a presente tese avança ao incorporar dados de distâncias reais dos alunos. Além disso, a posição das escolas no ordenamento de preferências não é a única variável de resultado analisada. Analisa-se também os resultados em termos de distâncias e da proficiência obtida. A proficiência foi conseguida por meio de simulações desenvolvidas a partir da análise das relações de variáveis envolvidas com a alocação dos alunos.

Uma discussão importante realizada por Fernandes (2007) trata do *trade-off* entre a escolha do algoritmo de DA e o algoritmo de TTC. Abdulkadiroğlu & Sönmez (2003) estipulam que o DA garante a estabilidade, no sentido que nenhum par de alunos e escolas já pareado gostaria de estar com outra escola ou aluno que não o par atual. No entanto, o algoritmo DA para escolas não garante que o resultado será Pareto eficiente (Roth, 1982 e 1985).

Por outro lado, o algoritmo de TTC necessariamente alcança um resultado eficiente de Pareto para os dois lados, mas não necessariamente será estável. No capítulo 3, seção 3.4, serão discutidos mais detalhes da relação entre estes conceitos.

2.2.5. O sistema de alocação de Belo Horizonte (Brasil)

O sistema de cadastro de Belo Horizonte foi implementado em 1992/93. Trata-se de um dos primeiros sistemas de alocação a aparecer no país, o que se tinha anteriormente eram sistemas descentralizados de matrículas por escolas e um processo de filas que repassava o ônus para os cidadãos. O histórico da implementação do sistema em Belo Horizonte foi apresentado em maiores detalhes na seção 2.1.

Os passos do sistema de cadastro de Belo Horizonte estão descritos pela numeração abaixo:

Passo 1. Na metade do ano (Junho ou Julho) é publicada a resolução conjunta da SME-BH e SEE-MG que estabelece os prazos para as matrículas e outras diretrizes.²⁸

Passo 2. Entre os meses de agosto e setembro é aberto o período de cadastramento nos Correios, somente para moradores residentes em Belo Horizonte. Pais interessados em matricular os filhos nas escolas públicas devem se dirigir a uma das agências dos Correios credenciadas portando certidão de nascimento do filho e comprovante de residência, sendo adotada a conta de energia elétrica da CEMIG.

Passo 3. Encerrada a etapa de cadastramento, verifica-se o número de alunos demandantes no sistema. De posse da informação dos alunos demandantes e do endereço georreferenciado dos alunos, a comissão de cadastro, em parceria com técnicos da PRODABEL, apura qual é a jurisdição a que pertence um candidato e faz a recomendação para alguma das escolas de sua jurisdição. Dentro do passo 3, alguns detalhes são importantes:

Passo 3.1. Caso a jurisdição possua apenas uma escola, todos os alunos são direcionados para a única escola da jurisdição. Caso a demanda dos alunos seja maior do que a quantidade de vagas da escola, alunos mais novos são priorizados (ano de nascimento, mês e dia), os alunos que excedem o número de vagas da escola são alocados para a escola mais próxima entre as existentes nas jurisdições vizinhas. O processo continua até que todos os alunos da jurisdição encontrem vagas.

Passo 3.2. Caso a jurisdição possua duas ou mais escolas, um critério de proximidade é determinado para a alocação dos alunos, os alunos são alocados para a escola mais próxima pertencente à jurisdição escolar. Se uma das escolas da jurisdição possui uma demanda maior do que o número de vagas, os alunos mais novos são priorizados e os alunos que excedem o número de vagas são propostos para a segunda

²⁸ A resolução de 2012 se encontra na íntegra no Anexo A.

escola mais próxima da mesma jurisdição. Caso se esgotem as vagas da jurisdição, mas ainda assim exista excedente de alunos não alocados, a escola mais próxima de um distrito vizinho é proposta. Esse sistema continua até que não haja alunos sem vagas.

Passo 4. Com o resultado da alocação do passo 3, a comissão envia para o endereço dos pais (informado pela conta de energia elétrica da CEMIG) uma correspondência identificando a escola para o qual a matrícula é sugerida (o anexo A fornece um modelo da correspondência encaminhada para o aluno).

Passo 5. Do início até o final da terceira semana de dezembro ocorre o período de matrículas para o ano letivo seguinte. Os pais que aceitarem a recomendação do sistema de cadastro fornecida pela comissão de matrículas têm a vaga garantida na escola indicada. Porém, nessa etapa, alguns pais podem se desviar e decidir matricular em outra escola de sua preferência. Isso pode ocorrer, entre outros problemas, por haver pais que julgam existir um erro na informação processada. A qualquer momento, a secretaria da escola pode consultar a comissão de matrícula para verificar o endereço fornecido. No caso de endereço incorreto, a comissão indicará o aluno para escolas de mais fácil acesso onde houver vagas.

Passo 6. Alunos retardatários, que não entraram no processo do sistema de cadastro em tempo hábil, ou não se apresentaram para as matrículas no período estipulado, serão alocados para escolas onde houver vaga.

O sistema encerra quando todos os alunos que procuraram matrículas nas escolas públicas tiverem sido matriculados. Em geral, todos os procedimentos desse sistema duram pouco menos que um semestre (de Agosto a Dezembro). Após a publicação da resolução e a convocação dos pais e responsáveis para as matrículas, o cadastramento nos Correios se inicia, isso ocorre pelo mês de agosto. O processo todo de alocação finaliza até o início do período de matrículas em

dezembro, antes do início do próximo ano letivo em janeiro do ano seguinte. Matrículas extemporâneas podem ser realizadas ao longo de todo o ano, mas são casos de exceção.²⁹

Uma mudança significativa do sistema acima descrito ocorre para matrículas do ensino médio. No ensino médio, os alunos do 9º ano que passaram de ano e ingressarão na primeira série do ensino médio são automaticamente reencaminhados para o ensino médio da escola em que já estudam. Contudo, em alguns casos a escola não possui o ensino médio, nessas circunstâncias, o aluno pode indicar quatro escolas do ensino médio (em ordem de prioridade) para as quais gostaria de se ver matriculado.

Na etapa acima, a capacidade de vagas das escolas listadas é avaliada e caso existam mais alunos demandantes do que vagas, os alunos mais novos são priorizados e os remanescentes são conduzidos para a segunda melhor opção. O procedimento continua até que todos os alunos sejam alocados. Caso não haja vaga em nenhuma das 4 escolas listadas pelo aluno, o sistema o encaminhará para a escola de mais fácil acesso.

Há ainda que se mencionar o caso de escolas de fundamental que possuem o ensino médio, mas que, porém, o número de vagas no ensino médio é menor do que o número de formandos do fundamental daquela escola. Nesses casos, as diretrizes da resolução conjunta estabelecem que se deve sortear os alunos que continuarão na escola e os não sorteados deverão escolher 4 escolas do ensino médio em ordem de prioridade. Estes alunos passam pelo mesmo procedimento descrito acima.

Um caso menos comum, mas que também se enquadra na situação do ensino médio, é o de novos entrantes para o ensino médio no sistema de ensino público da capital. A esses casos também se aplica o processo de escolha de quatro escolas.

O caso do ensino médio possui outra particularidade: desde a lei Nº 9394/96 de diretrizes e bases da Educação e leis posteriores, há uma hierarquia e relação de coordenação entre os diferentes

²⁹ O Anexo A apresenta um cronograma das atividades realizadas pelos diversos órgãos vinculados ao sistema de cadastro de Belo Horizonte.

entes federativos e os níveis de ensino. Os municípios deverão prestar prioritariamente o ensino infantil e fundamental. Os estados colaboram com a educação infantil e fundamental, mas devem priorizar o ensino médio. Por conta dessa diretriz, está se desenvolvendo um sistema de alocação próprio para o ensino médio, a Companhia de Tecnologia da Informação do Estado de Minas Gerais (PRODEMGE), empresa similar à PRODABEL para o Estado de Minas Gerais, participa desse convênio para o desenvolvimento de um projeto para o ensino médio.

Comparando o sistema de cadastro de Belo Horizonte com o mecanismo de São Paulo relatado por Fernandes (2007) notam-se várias semelhanças entre os dois sistemas. A primeira delas é que existe uma estimativa da oferta (uma programação da oferta de vagas) desenvolvida em parceria entre a administração da escola e as secretarias de educação.

A segunda semelhança é que as inscrições em São Paulo são feitas por postos de atendimento espalhados pela cidade, porém não são utilizadas agências dos Correios. O terceiro ponto em comum é que o sistema de São Paulo possui prazos de cadastramento, processamento da informação e matrículas similares aos prazos de Belo Horizonte. A quarta semelhança é que em São Paulo, também para o Ensino Médio, há um processo de escolha das escolas pelos alunos. Entretanto, em São Paulo são escolhidas apenas três escolas e em Belo Horizonte são quatro, como já apresentado.³⁰

As diferenças dizem respeito à escala de São Paulo, uma das maiores redes de escolas públicas urbanas do mundo e que é muito maior do que a de Belo Horizonte. Ainda segundo informações de Fernandes (2007), São Paulo possui 2.497 escolas municipais e estaduais, contra um total de 382 escolas públicas em Belo Horizonte. Em 2007, o total de alunos do ensino básico (fundamental e médio) em São Paulo foi de 2.345.749. Por conta dessa diferença de escala, o sistema de São Paulo é conduzido por regionais.³¹

³⁰ Uma quinta semelhança que pode ser observada é que ambos os sistemas têm abertura para tratamento de casos excepcionais que precisam ser tratados de forma discricionária. E uma sexta é que as duas cidades possuem sistema de transporte de alunos.

³¹ Belo Horizonte também possui participação das regionais, mas o sistema é um pouco mais centralizado para a cidade como um todo e pelo trabalho da SME-BH.

Dados do Censo Escolar de 2007 reportam que Belo Horizonte possuía 593.319 alunos matriculados no ensino básico. Em São Paulo o total de alunos a serem pareados no ensino Fundamental e Médio é de mais de 320 mil alunos, em Belo Horizonte, esse número está por volta de 30 mil alunos (e decrescendo ano a ano, ver capítulo 5). Esses dados fornecem uma dimensão do universo de alunos e uma comparação entre as duas cidades brasileiras.

Para o caso de Nova York, que também possui um grande sistema de matrículas a serem pareadas, Abdulkadirgolu, Pathak & Roth (2005) reportam que o número total de alunos de Nova York a serem pareados é próximo à 90 mil alunos, bem menor do que o de São Paulo e cerca de 3 vezes maior que o de Belo Horizonte.

Além da diferença de escala, outra diferença importante entre o sistema de Belo Horizonte e o de São Paulo é que o sistema de São Paulo foi implementado em 1995 enquanto que o de Belo Horizonte é anterior, de 1993. Mas apesar desta distinção de datas, os dois sistemas apareceram em períodos relativamente próximos e após a instauração do regime democrático, no bojo de uma série de reformas educacionais que ocorreram nos anos 90.

Uma última diferença que se constatou foi que no critério de desempate, o município de São Paulo prioriza o aluno mais velho, enquanto que em Belo Horizonte a prioridade é para o aluno mais jovem.

Afora essas diferenças supracitadas, pode-se concluir que, pelas semelhanças dos dois mecanismos, o sistema de cadastro de Belo Horizonte é também um mecanismo de matrículas que aproxima os alunos da menor distância possível das escolas. O mecanismo de Belo Horizonte funciona como se alunos e escolas tivessem preferências dadas somente pelas distâncias. Não se considerando efeitos de enturmação (formação de turmas), que serão tratados no final deste capítulo, o sistema de Belo Horizonte funciona tal como o algoritmo de Gale & Shapley com escolas propondo primeiro.³²

³² Na realidade, para um sistema de alocação baseado somente nas distâncias, não faz diferença qual lado propõe pois o resultado é único. Uma proposição a ser demonstrada no capítulo 3 (em 3.6) estipula que, sob preferências estritas e dadas pelas distâncias, tem-se que a matriz de preferência de um grupo é determinada pela mesma matriz de preferências que serve ao outro grupo. Para esse caso o resultado do algoritmo de Gale-Shapley é único, o resultado é

Ao longo deste trabalho serão apresentados mais detalhes dos dados de alunos e escolas de Belo Horizonte, bem como detalhes de todo o processo de cadastro e implicações das escolhas feitas pelo município. Como bem ressaltado e enfatizado, uma das principais lições que se pretende conhecer estudando o sistema de Belo Horizonte é desvendar se é possível aperfeiçoá-lo. As evidências desta tese apontam que sim, e um dos propósitos deste trabalho é demonstrar as relações causais entre o *matching* e o bem estar social, medido aqui através da proficiência dos alunos.

2.2.6. Outras cidades?

Outras cidades com sistemas próprios de alocação de alunos aparecem reportadas na literatura. Abdulkadiroğlu e Sönmez (2003), por exemplo, mencionam as cidades norte americanas de Minneapolis e Seattle, que possuem mecanismo similar ao de Boston. De fato, a alocação de alunos e escolas públicas é um problema comum a todas as cidades de grande e médio porte. A questão é saber como as cidades lidam com esse problema e se a solução é a mais conveniente e não causa desconforto para a população.

No Brasil, é possível apanhar relatos de jornais e revistas de cidades que ainda não possuem um sistema com o nível de organização que evite problemas para a população. No começo da etapa de elaboração desta tese, as capitais Vitória-ES e Manaus-AM não contavam com um sistema de matrículas que evitasse as filas. A prefeitura de Vitória descreve no seu portal que o sistema de matrículas é feito por escolas de maneira descentralizada. De acordo com a descrição da prefeitura (Secretaria de Educação de Vitória, acessado da internet)³³:

“As matrículas novas são realizadas por ordem de chegada dos pais ou responsáveis. A prioridade é dada aos moradores de bairros vizinhos às escolas. Os pais das crianças que se enquadram na Chamada Pública Escolar mas que não foram inscritas,

o mesmo independente de qual grupo propõe primeiro. Chamaremos esse tipo de preferências de “preferências dadas pelas distâncias”. É preciso observar que nem todo resultado único advém das distâncias, mas que esse tipo de preferências consiste em considerar que as distâncias do aluno até a escola desempenham um papel cardinal na matriz de preferências, não há dois ou mais alunos equidistantes para qualquer uma das escolas e não há duas ou mais escolas equidistantes para qualquer um dos alunos.

³³ Informações do Portal da prefeitura de Vitória: <http://www.vitoria.es.gov.br/semi.php?pagina=ensinofundamental> (último acesso 24/04/2013).

também podem comparecer aos estabelecimentos de ensino no período da Matrícula Nova. Contudo a vaga não é garantida”.

Os sistemas que impõem a ordem de chegada como critério para a seleção possuem problemas em situações de grande demanda, geram filas e desconforto e podem levar a comportamentos discricionários indesejáveis.

De acordo com dados recentes, obtidos na etapa final de conclusão desse trabalho, a Secretaria Estadual de Educação do Estado do Amazonas trabalha em um sistema centralizado de cadastro, conforme informa o portal da Secretaria.³⁴ Em períodos anteriores, há relatos de grande tempo de espera nas filas (Diário do Amazonas e A Crítica).³⁵ Há até relatos de pagamentos monetários por lugares nas filas.³⁶

Esses inconvenientes surgem em sistemas não monetários descentralizados, a ausência de pagamentos para organizar o processo de alocação de vagas e a descoordenação das decisões leva a resultados de muita demanda. Em certas situações as pessoas podem julgar o sistema injusto (mesmo que alguns consigam a vaga almejada, o fato de alguns não a conseguirem gera insatisfação, pela perda de tempo e outros ônus) o que acaba prejudicando sua credibilidade e estabilidade de todo o sistema.

Trata-se de um problema de grandes capitais e cidades de médio porte do interior que passam por período de rápido crescimento e forte urbanização, essas cidades também enfrentam problemas no momento das matrículas. Isso ressalta a importância do tema de alocação de matrículas e seus

³⁴ Pelas informações presentes no site não é possível saber como é exatamente o sistema, há agora opção de matrículas pela internet, mas o sistema de cadastramento continua também sendo realizado nas escolas. A secretaria alerta que o número de vagas é suficiente para todos: <http://www.matriculas.am.gov.br/index.php/pages/informacoes> (último acesso 24/04/2013). Entretanto, cabe ressaltar que número de vagas suficientes não é o principal requisito para se evitar filas, pois as filas ocorrem pelas boas escolas. O que tem que haver é uma garantia de maior homogeneidade do sistema.

³⁵ O diário do Amazonas de grande circulação divulga a criação de matrículas pela internet no novo sistema e fala do problema anterior das filas: <http://www.d24am.com/noticias/amazonas/em-manaus-matriculas-na-rede-publica-podem-ser-feitas-pela-internet/78015> (último acesso: 24/04/2013). O Jornal “A Crítica”, menciona ocorrência de grandes filas para matrículas no sistema de educação infantil: <http://acritica.uol.com.br/manaus/Amazonia-Amazonas-Manaus-Fila-quilometrica-por-vaga-em-creche-de-Manaus-0-830316980.html> (último acesso: 24/04/2013).

³⁶ Fonte do site G1 da Globo Notícias: <http://g1.globo.com/am/amazonas/noticia/2012/01/mae-garante-vaga-em-escola-publica-pagando-r30-para-furar-fila-no-am.html> (último acesso: 24/04/2013).

impactos no bem estar. O que essa tese pretende desvendar é qual é a relação desses sistemas com o aprendizado dos estudantes, e se isso gera efeitos na aquisição de conhecimento.

2.3. A literatura de Pareamento na Produção Familiar (casamentos à la Becker)

Curiosamente, a literatura que Gary Becker ajudou a desenvolver quando escreveu seus dois artigos sobre casamentos (*A Theory of Marriage Part I* e *Part II*, de 1973 e 1974) trata de mercados para os quais as transferências monetárias são incomuns. São denominados mercados por analogia, em um mercado de casamentos é difícil saber quem é o demandante e quem é o ofertante. O próprio Becker (1991, p.81) no livro *A Treatise on the Family*, escrito posteriormente como um compêndio de diversas das suas teorias desenvolvidas para a família, alerta: “*The phrase ‘marriage market’ is used metaphorically and signifies that the mating of human populations is highly systematic and structured*”.

A preocupação principal de Becker (1973 e 1974) era explicar a formação das famílias por meio do arcabouço econômico dos agentes racionais. Becker intuiu que a racionalidade econômica poderia ajudar a desenvolver modelos preditivos para a formação de pares e, por consequência, influir na constituição de famílias, nas decisões sobre o número de filhos, e uma série de outros temas familiares que passavam despercebidos por grande parte da ciência econômica do seu tempo.³⁷

Em verdade, uma parte dos casamentos de sociedades ancestrais, ou mesmo na era moderna de algumas sociedades, possuíam transferência de renda entre esposos na forma de dote. Becker (1991) menciona evidências de pagamentos monetários em algumas dessas sociedades antigas e Siow (2003) as relaciona mais sistematicamente. Uma hipótese dessa literatura é que o preço dos dotes pagos modificaria de acordo com a fração de homens e mulheres nessa sociedade, que refletiria uma escassez ou abundância relativa (qual lado seria merecedor do dote variava de sociedade para sociedade, mas o mais comum é encontrar dotes sendo pagos aos homens).

Quando não há possibilidades de pagamentos monetários para os casamentos (são legalmente proibidas, ou moralmente inaceitáveis), esse arcabouço teórico menciona que existem outras maneiras de se transferir valores entre esposos dentro de um casamento. Isso envolve, por

³⁷ Becker (1974) realiza uma revisão dos textos clássicos que tratavam de consumo familiar e da formação de família anteriores ao seu trabalho e aponta que no início do século XX a preocupação com tais temas andava em baixa, até identificar um recente desenvolvimento ligado a vertente que procurava entender como o consumo ocorria internamente nas famílias.

exemplo, a fração de bens consumidos no domicílio por cada um dos entes familiares (ver Lundberg & Pollak, 1996; Browning & Chiappori, 1998; ou Chiappori, Fortin & Lacroix, 2002). Essas transferências podem se dar também em valores implícitos, os preço-sombra, tais como o salário-reserva do marido ou da mulher no mercado de trabalho, situações em que compensam a mulher não procurar mercado no trabalho, ou o contrário.

Por conta desse último aspecto levantado, se diz também que, pela teoria de Becker, as utilidades são transferíveis, enquanto que na literatura da seção 2.2, não há, pelo menos em princípio, transferências de utilidades.³⁸ Talvez, uma das maneiras menos óbvias de se transferir utilidades entre casais é procurar emparelhar características comuns. Casais podem buscar se assemelhar em características afins tais como renda, educação, idade, cor, hábitos de consumo, opiniões políticas e demais tipos de preferências sociais.

Casamentos entre pessoas com características afins são chamados por essa literatura de *positive matings*. O termo positivo se refere a características positivamente correlacionadas. Nesse sentido, pessoas de alta renda tenderiam a se casar com pessoas de alta renda, muito educados com muito educados e assim por diante. Quando ocorre a atração dos opostos (alto-baixo) se diz que é um *negative mating*, a correlação é inversa (negativa).

Becker & Murphy (2000) elencam as condições que podem levar à maior ocorrência de *positive matings* e quais as condições que levam aos *negative matings*. Em resumo, essas condições dizem respeito ao que faz o casal derivar mais utilidade, se for mais compensador o casamento entre iguais, isso quer dizer que as transferências que um dos lados do mercado (homens ou mulheres) tem de fazer para o outro lado são menores (mais “baratas”) na situação de características iguais, e, logicamente, maiores no caso inverso. Se for o contrário, se for mais compensador casar entre

³⁸ No capítulo 3, mais detalhes sobre as transferências de utilidades serão apresentados. Por hora pode se pensar no simples exemplo a título de esclarecer a diferença entre as teorias. Suponha que uma mulher w_1 tenha declarado, em um *ranking* de preferências, o homem h_1 como o parceiro mais preferido. Suponha, ademais, que um casamento por preferências (tal como o do modelo *deferred-acceptance*) tenha formado o casal (w_1, h_1) . Porém, suponha que haja mais um homem, h_2 , que prefira a mulher w_1 com bastante estima (em primeiro lugar). Em uma última suposição, suponha-se que h_2 é mais rico que h_1 e que a transferência de recursos monetários seja aceitável nesse casamento. No modelo de transferência de utilidades, se a quantia a mais de riqueza que h_2 possui e transfere para sua pretendente w_1 , compensa a desutilidade que a mulher possui em casar com um homem menos preferido, o par (w_1, h_2) poderá ser formado. Essa conclusão será mais bem formalizada no próximo capítulo. Em termos coloquiais, pode se dizer que o modelo de transferência de utilidades aceita o casamento por conveniência.

desiguais, tem-se a possibilidade dos *negative matings* serem mais frequentes. Em termos gerais, essa literatura é também chamada de *assortative matings*, casamentos ordenados, tanto positivamente quanto negativamente.

Grande parte da evidência encontrada sugere que os casamentos mais comuns são do tipo positivamente correlacionados. É muito mais comum encontrar casamentos com indivíduos de características semelhantes. Uma extensa área da sociologia estuda esse tipo de fenômeno social (Mare, 1990). Mare (2008) mostra evidências de que ao longo do século passado os casamentos foram ficando ainda mais positivos, principalmente nos quesitos de renda e educação.³⁹

Ao contrário da literatura de *two-sided markets*, para o qual a palavra casamentos é mais alusiva, ou seja, casamentos são vistos de uma maneira mais geral para a formação de pares por preferências, ou como uma regra para a formação de pares. Na literatura de *assortative mating*, os casamentos possuem um sentido mais literal e se referem, de fato, às práticas sociais relacionadas à união de casais. Até o presente momento, não se tem notícia da teoria de *assortative matings* sendo aplicada para analisar casamentos mais genéricos entre escolas e alunos, embora o próprio Becker (1975) tenha dado contribuição fundamental para teoria do *capital humano*.

Como defesa do ponto de que a teoria de casamentos de Becker (1991) se aplica à alocação de alunos e escolas, pode-se pensar que o casamento de uma escola entre vários alunos é um casamento poligâmico, tal como descrito por Becker no capítulo 3 (*A Treatise on the Family*), os detalhes do modelo relacionado serão dados no capítulo 3 desta tese. Porém, não é difícil de se extrapolar essa questão para os casamentos abordados pela literatura de *two-sided matchings*, pois grande parte da literatura das decisões familiares já incorporou tanto o modelo de Gale & Shapley (1962) quanto o modelo de Becker (1973 e 1974), ver por exemplo Weiss (1997), Bergstrom (1997), Ermisch (2003) e mais recentemente Browning, Chiappori & Weiss (2011).

³⁹ Para ver uma a trabalhos da literatura de *assortative mating* tratada sob a perspectiva da evolução de casamentos inter-raciais, ver Fryer Jr. (2008) e Christensen (2008). Para estudos da educação, Rios-Neto & Guimarães (2010) empregam os métodos de progressão por série para identificar a seletividade no *matching*, à medida que os alunos progredem na carreira escolar, ou seja, para os últimos anos, o *positive assortative mating* de alunos e escolas se torna ainda mais positivo.

Volta-se a esse último ponto de interseção das literaturas mais à frente na seção 2.5. Por enquanto, na próxima seção, levantaremos trabalhos da demografia da educação e da literatura educacional, que ressaltam evidências importantes a serem consideradas para a tese.

2.4. A Literatura sobre aquisição de aprendizado na Educação

Uma miríade de estudos têm sido realizados para desvendar o impacto de diversos insumos educacionais no aprendizado. Recentemente, em 2010, o instituto Ayrton Senna, em parceria com o Instituto de Estudos do Trabalho e Sociedade (IETS), em uma equipe coordenada pelo pesquisador Ricardo Paes de Barros, divulgou um extenso trabalho de meta-análise dos principais fatores que influenciam a aquisição de conhecimento na educação.⁴⁰

Essa meta-análise organizou uma série de notas técnicas, agrupando diversos estudos da área de aprendizado educacional, e analisou a aquisição de conhecimento sob diversos aspectos agrupados em cinco dimensões: 1) Recursos da escola; 2) Plano e práticas pedagógicas; 3) Gestão da escola; 4) Gestão da rede de ensino; e 5) Condições das famílias. Essa meta-análise pôde ser realizada graças ao grande volume de trabalhos existentes avaliando a educação em diversos aspectos, resgatando textos clássicos e aliando a trabalhos mais recentes.

Antes de detalhar as relações de algumas dessas dimensões mencionadas no parágrafo acima, convém realizar uma revisão dos estudos educacionais anteriores. Brooke & Soares (2008) organizaram um livro que reúne artigos e capítulos importantes de diversos autores da educação, pesquisadores de diversas áreas: pedagogia, sociologia, estatística, economia e psicologia. O livro começa com trechos do relatório Coleman (1966), que é hoje um marco da literatura contemporânea de avaliação da educação. O relatório foi uma demanda do congresso norte-americano para avaliar o quadro educacional das escolas daquele tempo e foi um dos primeiros trabalhos a empregar a estatística para uma ampla base de dados sobre educação.

⁴⁰ A aquisição de conhecimento é medida por meio de testes padronizados tais como o SAEB, Prova Brasil, PISA (*Program for International Student Assessment*) e outros *surveys* independentes. As notas técnicas estão disponíveis no site: <http://www.paramelhoraroaprendizado.org.br/> (último acesso 25/04/2013). Os dados completos da equipe podem ser encontrados em: <http://www.paramelhoraroaprendizado.org.br/Conteudo/Sobre/equipe.aspx>.

O relatório Coleman (1966) colocou uma evidência difícil de ser superada: a de que o desempenho das escolas dependeria mais de características pessoais e familiares dos alunos do que do ambiente escolar.

Uma série de estudos posteriores vieram a confirmar os resultados do relatório Coleman (1966) e o papel das características individuais. Porém, paralelo a esse movimento, surgia a crítica de que as maneiras de se medir o “efeito escola” e outros efeitos institucionais da educação deveriam ser aperfeiçoados. Um argumento desta crítica coloca que se os estudantes negros possuíam desempenho inferior aos alunos brancos, isso ocorria por conta de uma desigualdade anterior proveniente do *background social*, e que diferenças de raça, por exemplo, só refletiriam uma desigualdade inerente do sistema.

Em verdade, a crítica posta dessa forma não pareceria justa, já que muitos autores dessa primeira corrente, e o próprio Coleman, acreditavam que era preciso dirimir as desigualdades para o sistema funcionar de uma maneira geral e para todos. Porém a maneira sistêmica de ver a escola sem ressaltar suas particularidades deixava uma lacuna para um novo programa de pesquisa a ser desenvolvido. Brooke & Soares (2008, p. 218) resumem bem esse ponto: “as escolas pertencentes ao mesmo sistema e com os mesmos recursos poderiam produzir resultados significativamente diferentes”. Medir as particularidades de cada escola se tornou tarefa importante para pesquisadores educacionais da década de 70 e 80.

Em meados dos anos 70 e durante a década de 80 uma série de autores se voltaram para a questão de como medir apropriadamente o efeito escola e os insumos escolares. Importante, sob esse aspecto, foi o surgimento do conceito de *valor adicionado*, ou seja, a escola deve ser avaliada pelo aprendizado que ela agrega ao aluno e não simplesmente pela diferença de patamares entre as escolas. Incorporou-se também, nesta época, o conceito de *modelos hierárquicos*, que são modelos de regressão que levam em conta a estrutura aninhada dos dados. Willms (1992), apud Brooke & Soares (2008), foi o responsável pela introdução do conceito de valor adicionado em

meados dos anos 80. Enquanto que Bryk & Raudenbush (1987) e Lee & Bryk (1989) desenvolveram a teoria dos *modelos hierárquicos* aplicados às escolas.⁴¹

A introdução desses conceitos foi um grande avanço para a literatura educacional, pois foi através deles que se pôde definir melhor o que antes se chamava **efeito escola** e **efeito aluno**. Como colocam Brooke & Soares (2008, p.222), o efeito escola aparece na literatura descrito em duas principais formas:

“A primeira definição do efeito escola capta o efeito global das escolas do grupo de referência com base na divisão da variação total dos desempenhos dos alunos de todas as escolas em dois termos: a variação dentro das escolas e entre as escolas. O efeito escola é definido como a divisão da variação entre escolas pela variação total. Essa medida de efeito assume valores entre 0 e 1, e seu valor absoluto deve ser interpretado como uma medida da heterogeneidade das escolas do grupo de referência. Em outras palavras, se esse efeito é grande, a escolha de uma ou outra escola traz grandes consequências para o aluno, ou seja, tem impacto na vida escolar dos alunos.

O segundo sentido do efeito escola é mais direto e intuitivo. Consiste no aumento no nível de proficiência do aluno que pode ser atribuído às práticas da escola. Por razões técnicas, a média dos efeitos das escolas do conjunto é considerado zero. Os valores positivos do efeito identificam escolas que são melhores do que a média, e os negativos, as escolas piores do que a média. Essa definição é usada com frequência para identificar escolas que merecem estudos de caso”.

Essas duas definições de efeito escola acima apresentadas estão presentes no capítulo 6 desta tese. Em relação à primeira definição do efeito escola, as regressões de *modelos hierárquicos* realizadas com dados de alunos do 5º ano de Belo Horizonte, sugerem um efeito escola máximo próximo de 16% da variância (valores um pouco menores aos de outros estudos no Brasil). Sobre o segundo tipo de efeito escola ele será diretamente empregado nas simulações. A seção 6.3.1 fornece detalhes sobre o uso desse efeito.

No Brasil, os modelos hierárquicos encontraram grande número de aplicações no final dos anos 90 e início dos anos 2000, César & Soares (2001), Albernaz, Ferreira & Franco (2002), Machado

⁴¹ Ressalta-se que o desenvolvimento teórico dos modelos hierárquicos não se restringe apenas aos benefícios acadêmicos. O uso de modelos hierárquicos, ao permitir compreender melhor os efeitos de aluno e escola, auxiliam o entendimento do processo educacional, possibilitando recomendações práticas mais adequadas às políticas de educação.

et al. (2006), Biondi & Felício (2007) e Rodrigues (2009) são algumas delas.⁴² No geral, se encontra que o efeito escola no Brasil é um pouco maior do que os padrões internacionais, ou como colocam Brooke & Soares (2008, p. 464): “O percentual da variância explicado pelo pertencimento escolar, depois de controlada a influência do nível socioeconômico, é um pouco maior do que os valores observados internacionalmente. Ou seja, há hoje uma ampla evidência empírica de que as escolas brasileiras podem ter um papel mais decisivo na melhoria do aprendizado cognitivo dos alunos de ensino básico brasileiro”.

Um desenvolvimento paralelo, relacionado à literatura de economia da educação, procurou olhar com detalhes os insumos e os processos de desenvolvimento educacional. Hanushek (1986 e 1992), e Hanushek & Luque (2002) analisam os insumos relacionados à escola tais como razão professor/aluno, salários dos docentes e tamanho das turmas. As primeiras evidências encontradas por Hanushek (1986) não foram muito animadoras, dado que o investimento por aluno não parecia estar relacionado à melhoria de desempenho, no entanto, isso indicava que escolas e alunos se distinguem essencialmente na questão da qualidade. Estudos posteriores de Hanushek (1992) e Hanushek & Luque (2002) investigaram em maior profundidade esse aspecto.

Alguns autores, como Murillo (2003), contrapõem a visão da escola dada por insumos e produtos, que está relacionada à produtividade escolar, e enfatizam a visão da eficácia escolar, que destaca os processos da escola. Entretanto, essas visões não são tão contrastantes. A abordagem de insumos e produtos também se preocupa por entender os processos já que quanto melhor se conhecem os processos e atividades internas da escola, melhor podem ser especificados os produtos e insumos corretos a serem medidos.

Voltando-se ao estudo de meta-análise abordado no início desta seção, entre os insumos mais importantes analisados (os de maior contribuição para a proficiência) estão a qualidade e experiência do professor, o tempo de exposição do aluno ao professor (interação aluno professor) e o efeito de composição das turmas, literatura de *peer effects* (ver Pinto, 2010).

⁴² Rodrigues, Rios-Neto & Pinto (2010) apresentam resumo do resultado da primeira parte do trabalho de Rodrigues (2009). Nesse artigo os autores não fazem uso dos modelos hierárquicos, mas tratam do problema da queda do desempenho no SAEB no início dos anos 2000. Os autores usam os modelos de decomposição de densidade relativa. Baseando-se nesse trabalho, Delgado, Miranda-Ribeiro & Soares (2013) propuseram um índice para medir a desigualdade de desempenho escolar.

Sobre o impacto qualidade do professor em sala de aula, Rivkin, Hanushek & Kain (2005) estimam impacto positivo, e segundo medições da meta-análise de Barros (2010), transportado para o Brasil, esse efeito corresponderia de 57 a 70% do que um aluno tem de aprender em um ano de escola.

Trabalho recente de Guimarães (2012) levanta maiores detalhes sobre os efeitos da qualificação e seleção dos professores pelos dados do FUNDESCOLA (*Fundo de Fortalecimento da Escola*). A autora encontra efeitos reforçando os impactos positivos da qualidade dos professores e acha também uma relação positiva entre o método de seleção dos professores e o desempenho dos alunos.

Uma parte da literatura dos efeitos educacionais se dedica a mensurar um efeito importante e que pertence ao tema dessa tese: a escolha de escolas (*school choice*) e o efeito da competição entre escolas. A competição pode se dar com escolas públicas competindo entre si, tal como Abdulkadiroğlu, Pathak & Roth (2005) alegam ter ocorrido após a implementação do *matching* em Nova York, ou entre escolas públicas e escolas comunitárias (*charter schools*), ou ainda entre escolas públicas e privadas.

Garoman (1996) estuda um tipo bastante peculiar de escolas do sistema norte americano: as *magnet schools*. Essas escolas foram criadas para serem centros de excelência nos distritos escolares e para impedirem a segregação racial dos Estados Unidos. Foram pensadas já com um currículo diferenciado, na perspectiva de serem um modelo vocacional, acadêmico e de integração. Porém as escolas possuem autonomia para a seleção dos alunos: algumas aplicam provas, enquanto outras selecionam os alunos por sorteio.

Os dados de Garoman (1996) são do National Educational Longitudinal Study (NELS) e o autor analisou os anos de 1988 e 1990 da pesquisa. Os efeitos das escolas *magnet* foram extremamente positivos, comparando com os variados tipos de escolas do banco: escolas católicas privadas, escolas privadas não religiosas e escolas públicas comuns. Entretanto, os fatores limitadores dessa pesquisa são de que o número de escolas *magnet* na pesquisa (48 escolas) é extremamente pequeno perto do universo de escolas dos EUA, sendo que, além disso, tais escolas possuem um

sistema bastante peculiar de funcionamento. Tendo isto em vista, é difícil extrapolar os resultados das *magnet schools*. Ainda assim, esse modelo de escola é importante de ser analisado em detalhes. No capítulo 7 desta tese, uma das simulações para políticas públicas é inspirada nesse tipo de escolas.

Cullen, Jacob & Levit (2006) realizam uma análise do *school choice* para escolas de Chicago nos Estados Unidos. Nesse trabalho os autores compararam estudantes pretendentes ao *school choice* que optaram por determinada escola em que a demanda superou a oferta. Respeitados os critérios da autoridade educacional para a composição das turmas, um número de pretendentes maior do que o número de vagas determina um sorteio aleatório dos alunos. Por este sorteio alguns alunos conseguem a vaga, enquanto que outros são alocados para outras escolas de acordo com critérios de zoneamento.⁴³

Cullen, Jacob & Levit (2006) também realizam um levantamento das distâncias da casa do aluno até a escola, tais como as medições que aqui foram realizadas. Para os alunos que ganham a loteria para escolas comuns (e por isso têm que, em geral, percorrer maiores distâncias) o efeito da distância encontrado pelos autores é negativo. No entanto, para alunos que ganham o sorteio para escolas de alto valor agregado, de média elevada e de alta popularidade (muito demandadas) o efeito da distância passa a ser positivo.

Sobre a concorrência entre as escolas públicas, o resultado final de Cullen et al. (2006) não é conclusivo, o efeito total dos alunos que foram selecionados aleatoriamente para a escola que escolheram no sorteio é ligeiramente negativo em relação à aqueles não sorteados, sendo que aparentemente não encontraram efeito para a escolha de determinadas escolas públicas.

Hoxby (2000) e Hoxby & Muraka (2009) analisam o efeito das escolas comunitárias norte-americanas para os distritos de Washington e Nova York, também tentando controlar por um efeito de seleção. Em Hoxby (2000) esse controle é feito com efeito de borda entre diferentes

⁴³ O leitor pode observar que a descrição do critério de *school choice* de Chicago se assemelha à descrição do critério de seleção de Boston pré-2005. No entanto, pela descrição presente no artigo de Cullen et al. (2006) é possível identificar que um mesmo aluno pode participar de mais de uma loteria para diferentes escolas, mas o artigo não explica se essas escolhas são tomadas uma a uma ou se existiriam mais de uma opção para o aluno ao início do período de matrículas, tal como ocorria no sistema de Boston.

distritos e em Hoxby & Muraka (2009) por uma seleção aleatória dos alunos que foram elegidos para as *charter schools*. Ambos os estudos relatam um efeito positivo para os alunos que frequentam as *charter schools*.

Outra área de estudos da educação investiga os possíveis efeitos da região de moradia do estudante na proficiência. Sobre esse tema, Ribeiro & Katzman (2008) reuniram uma série de trabalhos explorando a relação entre a escola e o seu entorno social e urbano. Investiga-se o efeito de localização da escola, procurando-se entender em que condições o território contribui para a redução ou perpetuação das desigualdades.

Soares, Rigotti & Andrade (2008) estudam os efeitos da segregação espacial urbana de Belo Horizonte no desempenho do SAEB. Na construção dos dados, os autores utilizam dados do Censo Demográfico Brasileiro e uma informação do Nível Sócio Econômico (NSE) das escolas para verificar a relação entre o entorno urbano da escola e os resultados de NSE. Na conclusão deste trabalho, os autores apontam algumas relações espaciais entre a equidade e qualidade das escolas por região de BH e conjecturam várias hipóteses, tais como a do efeito de migração e deslocamento dos alunos, que puderam ser melhor trabalhadas nesta tese.

Por fim, destaca-se que a literatura sobre os efeitos educacionais é importante para a interpretação dos resultados obtidos das simulações do capítulo 7. Além de se destacar a evolução desse campo de estudos, a consulta dessa literatura procurou ressaltar três efeitos importantes para a simulação:

1. o efeito de composição das turmas (enturmação);
2. a interação entre aluno e escola (e interação entre turmas de alunos e a escola, e como medir o efeito escola); e
3. a concorrência entre escolas públicas e a relação da escola e o território.

Uma das principais medições realizadas nessa tese é a distância de casa até a escola para cada um dos alunos. Essa variável permitiu destacar o efeito da alocação como um efeito importante para a literatura educacional.

2.5. A Interseção entre as três Literaturas

As três literaturas anteriormente abordadas possuem cada uma delas uma agenda de pesquisa bem delimitada. A literatura de aprendizado na educação é a mais ampla e congrega conceitos de diversas áreas, sendo que diversos recortes dessa literatura podem ser realizados e com diferentes propósitos. A literatura de *two-sided markets* se encontra em plena expansão, com artigos recentes sendo publicados e com novos achados divulgados, o mesmo se pode dizer da literatura de *assortative matings*, embora, como ressaltado, essa literatura não foi ainda aplicada ao casamento de alunos e escolas.

O intuito desta presente seção é definir a interseção entre essas três literaturas. A tarefa não é fácil, dado os propósitos distintos de algumas delas, mas há alguns indícios dos caminhos que podem ser percorridos.

Em relação às duas literaturas de *matchings* analisadas, a ligada a Gale & Shapley (1962) e a vinculada a Becker (1973 e 1974), Browning, Chiappori & Weiss (2011) realizam a ligação entre as duas vertentes com muita propriedade. Os autores citam que pelo algoritmo de Gale-Shapley, se tanto homens quanto mulheres classificam uns aos outros por meio de uma única característica observada, digamos que a renda ou a educação, o resultado final do algoritmo será o *positive assortative mating*, estipulado por Becker (ver seção 2.3). Em relação ao algoritmo com transferência de utilidades, Browning, Chiappori & Weiss (2011, p. 296) denominam o algoritmo com a alcunha de “Becker-Shapley-Shubik”, em uma clara fusão entre os dois referenciais teóricos, o de Becker (1973) e o de Shapley & Shubik (1971).

Revisitando Roth & Sotomayor (1990, p.247) os autores mencionam a interseção entre as pesquisas, mas também destacam as diferenças entre elas. Apesar das diferenças, Roth & Sotomayor apostam em uma profícua interação entre as duas áreas:

“As mentioned in the guide to the literature following Chapters 2 and 8, highly structured models of the kind explored by Becker (1981), which make strong assumptions about the nature of agents’ preferences, allow one to investigate how agents sort themselves out, that is, who is matched to whom. The models explored in this volume make few assumptions about preferences, but reveal a great deal of

structure to the set of stable outcomes, some of which allow welfare comparisons (e.g., there are optimal stable outcomes for each side of the market). *It seems likely that these two kinds of investigations can be profitably combined.*” (Roth & Sotomayor, 1990, p.247, grifo do autor da tese)⁴⁴

Em outro trecho, Roth & Sotomayor (1990, p.220) apontam uma diferença importante da abordagem entre os dois arcabouços:

“Becker (1981), Who uses the assignment model to study marriage and household economics, makes use of the fact that stable outcomes all correspond to optimal assignments (and that the optimal assignment is typically unique) to study which men are matched to which women, for different assumptions about how the assignment matrix is derived.” (Roth & Sotomayor, 1990, p.220)

Roth & Sotomayor (1990) destacam no ponto acima que a maneira como as preferências dos dois grupos são obtidas se altera em essência entre os modelos. No modelo de Becker (1991) há uma variável, ou um conjunto de duas ou mais variáveis, que ajudam os pares a se decidir sobre as características importantes e que derivam maior utilidade para a união com o parceiro. Em outra explicação Roth & Sotomayor (1990, p.51-52) apresentam que o modelo de Becker se aproxima mais dos modelos de *assignment games* (apresentado no capítulo 8 de Roth & Sotomayor, 1990):

“[Becker (1981)] considers a much more highly structured model of the kind to be considered in Chapter 8, but in which preferences are not taken as primitives, but rather correspond in a given way to attributes of agents (such as wage earning or homemaking ability). Thus although our results focus on things like the existence of stable outcomes for any preferences, his concern is with how the agents sort themselves out (e.g., high wage earners marry good cooks) for particular preferences.” Roth & Sotomayor (1990, p.51-52)

Em seu livro “*A Treatise on the Family*” Becker (1991, p.127 n.23) percebe a complementaridade entre as duas áreas, o autor destaca a relação de uma das condições do *core* dos casamentos de sua teoria com a condição de estabilidade de Gale & Shapley (1962). Essa condição mencionada por Becker é a mesma que está apresentada na equação (7) desta tese:

⁴⁴ O livro de Becker “*A Treatise on the Family*”, possui uma edição anterior de 1981, que é a edição que Roth & Sotomayor (1990) referenciam, mas é essencialmente o mesmo livro. Para essa tese a versão consultada foi a de 1991.

“Gale and Shapley (1962) require optimal assignments to be “stable”; that is, persons not assigned to each other could not be made better off by marrying each other, a requirement that is closely related to condition (4.5).” (Becker, 1991, p. 127n.23).

Dependendo do ponto de vista, pode-se afirmar que a teoria de Becker faz pressuposições mais fortes para as preferências dos casais. No modelo de Becker, os casais concordam entre si sobre quais características devem ser valorizadas para ranquear o outro lado do mercado, isso se dá porque a utilidade não difere entre os indivíduos de um mesmo grupo.⁴⁵

A equação (4.20), mencionada no trecho abaixo, trata a divisão do produto do domicílio entre os membros do casal como fixa, e serve como um comparativo para o modelo de Gale & Shapley (1962), para o qual esse produto do casamento não está sob investigação e pode ser tomado implicitamente como fixo. Becker (1991, p. 127) é incisivo em apontar as diferenças entre as teorias:

“My approach to the marriage market contrasts sharply with other formal models of marital sorting (see Gale and Shapley, 1962; Stoffaës, 1974). These models, like the model given by Eq. (4.20), assume that each person has a *given* ranking of potential mates that determines rather than is determined by the equilibrium sorting. Unlike the rankings implied by (4.20), however, in these models different persons may not rank potential mates in the same way [...]. If rankings were not the same, an “optimal” sorting could only try to minimize the overall conflict between feasible and preferred matches.

These models can be said to assume implicitly, while the model given by (4.20) assumes explicitly, that the division of output in any marriage is not determined by the marriage market and is completely rigid. An individual usually would not prefer the mate assigned him by the optimal sorting, because marital prices are not permitted to eliminate the inconsistencies among the preferred choices of different persons. If the division of marital output were determined by the marriage market, the ranking of potential mates would not be given; it would depend on how the outputs produced with different mates were divided. That is to say, if marital prices were flexible, the problem formulated and solved by these models would be irrelevant to actual marital sortings.” Becker (1991, p. 127)

⁴⁵ Uma possível generalização para o modelo de Becker seria tratar funções de utilidade distintas para pessoas diferentes de um mesmo grupo. Isso poderia adicionar uma complexidade difícil de ser tratada no modelo. Becker & Murphy (2000, p. 30) preferem reduzir o problema por partes, ou seja, dividir o mercado maior em submercados: “Obviously, there are considerable differences in many marital preferences since, for example, Catholics may prefer to marry Catholics, blacks to marry blacks, and tall people to marry tall people. However, strong differences of this type may simply decompose the overall marriage market into various homogeneous submarkets, where participants in each submarket, such as the white or black submarket, marry others in the same submarket”.

O que Becker (1991) ressalta nestes dois parágrafos acima é que o preço de homens e mulheres no mercado de casamentos importa para a sua investigação sobre os casamentos. Becker esperava que em sociedades com oferta de muitos homens o preço cairia, e isso se manifestaria de várias maneiras, inclusive com homens obtendo uma menor repartição do produto no casamento.⁴⁶

No entanto, a nota de rodapé 24 do último parágrafo supracitado de Becker, (Becker, 1991, p.127 n.24) reconhece a relevância da situação proposta pela outra teoria quando não há preços envolvidos:

“It might be relevant, however, to markets that do not use prices to determine assignments. For example, Gale and Shapley (1962) also discuss the assignment of applicants to different universities.” (Becker, 1991, p.127 n.24).⁴⁷

Mercados de alocação que não envolvem mecanismos de preços são justamente o ponto importante para alavancar a teoria da alocação de alunos nas escolas públicas e o seu impacto na proficiência. O tema também serve para introduzir, com maior riqueza de detalhes, a literatura educacional na investigação de como as alocações por algoritmos podem ajudar no bem estar e no desempenho escolar dos alunos.

Como foi visto na seção 2.2, diversos autores estudaram os sistemas de alocação de alunos e escolas empregados em diversas cidades do mundo. Essa literatura trouxe uma série de conhecimentos administrativos e sugestões de procedimentos ótimos sob o ponto de vista das preferências. A literatura educacional aborda o problema do *school choice* e o efeito do território na escola, e enfatiza que esse ponto pode ser importante para o aprendizado.

Ressalta-se novamente que a literatura de casamentos para a formação de domicílios de Becker, muito estudada pela demografia, não parece ter sido ainda expandida para uma analogia de casamento entre alunos e escolas.

⁴⁶ Essa é uma interessante analogia para determinar preços no mercado de escolas particulares.

⁴⁷ O anexo C apresenta duas figuras esquemáticas com um diagrama de fluxo que representa as diferenças entre as duas literaturas.

Propõe-se para esta tese, um trabalho de demografia da educação que una essas três linhas de pesquisa, já que parece existir uma forte área de possível interação entre elas. A tese se propõe estudar os casamentos entre alunos e escolas. Bem como, estudar quais as maneiras que estes “casamentos” possuem para propiciar maiores ganhos de aprendizado aos alunos. Apresenta-se como a *oferta* e a *demanda* interagem em situações em que não há um mecanismo de preços, e a maneira de como essas partes se fundem impactando no produto final (o aprendizado).

Como já exposto, um dos principais motivadores deste trabalho é que o sistema atual pode acarretar numa “armadilha de pobreza” ao designar para um aluno capaz, e morador de um bairro pobre, a escola mais próxima. O procedimento em vigor assume implicitamente que todas as escolas públicas são de igual qualidade, o que não parece ser verificado na prática.

Há que se considerar que propor outras alocações diferentes daquela de menor distância, é um problema logístico bastante complexo. O deslocamento em grandes cidades é um componente importante do orçamento das famílias mais pobres, além de ser muito dispendioso em tempo, e perigoso para crianças pequenas.

Os efeitos negativos de um deslocamento para maiores distâncias podem se sobrepor, em larga medida, aos efeitos positivos de se estudar em uma escola mais qualificada. Principalmente se a diferença de qualidade não for muito grande. Os custos e as implicações oferecidas por outras alocações entre escolas e estudantes deverão também ser abordados na tese.

Outra questão primordial diz respeito a alterações do *status quo* vigente: “o que fazer com os alunos que já se encontram nas escolas ditas melhores?” Designar que esses alunos que têm *background* melhor devam ceder lugar a um colega menos favorecido, porém capaz, pode ser tão injusto quanto o atual sistema. Afinal, quem mais merece o ensino de qualidade? A resposta fácil diria que todos merecem igualmente. Porém, essa é uma resposta que procura melhorar toda a tecnologia de ensino. O tema proposto toma a tecnologia atual como dada e tenta usar o fator de alocação como ferramenta para oferecer melhores possibilidades, e a questão ética por trás de uma alocação diferente é de extrema importância.

3. OS MÉTODOS DE PAREAMENTO

A escolha de uma notação para unificar diferentes literaturas é desafiadora. De um lado, corre-se o risco de sobrecarregar a notação, utilizando uma nova denominação para cada conceito novo utilizado, mas que, porém, não é essencialmente muito diferente dos conceitos utilizados nas outras áreas. Por outro lado, usar uma notação muito sucinta não permite a distinção e que o leitor rapidamente identifique os conceitos das literaturas abordadas. Uma notação é tanto melhor quanto mais rápido e mais economicamente ela permite o leitor abstrair informações precisas sobre o conceito trabalhado, uma tentativa de conciliar economia e precisão.

Entretanto, de uma maneira ou de outra, escolhas para a notação têm de ser feitas, e em certos momentos essas escolhas exigem que se opte por uma alternativa que, naturalmente, possui prós e contras. Um maior detalhamento pode ser oneroso demais em termos de carregar a notação, dificultar a leitura, e não trazer outros ganhos mais imediatos. Em contrapartida, uma notação sucinta demais pode levar a uma notação hermética ou, o que é pior, uma confusão entre os conceitos trabalhados.

A estratégia aqui adotada foi uma tentativa de unificar as nomenclaturas das diversas áreas abordadas no segundo capítulo, procurando respeitar os pontos que elas possuem de diferença. Grande parte da notação aqui utilizada seguiu a padronização de Roth & Sotomayor (1990) e Fernandes (2007), algumas adaptações se fizeram necessárias para melhor transição entre os conceitos de uma área a outra, pois no capítulo de análise exploratória a notação é essencialmente a que comumente se encontra na literatura de estudos da educação.

Destaca-se que em cada uma das literaturas abordadas a nomenclatura pode assumir formas distintas para simbolizar os conceitos. Em Becker (1973, 1974, 1991) e Becker & Murphy (2000) percebe-se algumas mudanças da notação utilizada. O mesmo ocorre nas outras duas literaturas. O objetivo da seção 3.1, a seguir, é o de situar o leitor para os conceitos utilizados no restante da tese, e o restante do capítulo descreve formalmente os fundamentos das teorias utilizadas.

3.1. Principais definições e conceitos (Nomenclatura)

Inicialmente serão apresentados os modelos de casamentos monogâmicos entre homens e mulheres. Cada homem pode se casar com apenas uma mulher, e cada mulher pode casar com apenas um homem (não existe poligamia)⁴⁸, essa suposição inicial facilita a exposição dos mecanismos de *matchings*.

Os modelos de *matching one-to-one* com a analogia de casamentos entre homens e mulheres são dados por dois conjuntos disjuntos e finitos de pessoas: homens (H) e mulheres (W), cada conjunto desses com n indivíduos, ou seja, mesmo número de pessoas em cada grupo.⁴⁹ Pode-se representar o conjunto dos homens como $H = \{h_1, h_2, \dots, h_n\}$ e o das mulheres como $W = \{w_1, w_2, \dots, w_n\}$:

Conjunto dos Homens: $H = \{h_1, h_2, h_3, \dots, h_n\}$

Conjunto das Mulheres: $W = \{w_1, w_2, w_3, \dots, w_n\}$

Neste capítulo as definições estão numeradas e em ordem crescente e são precedidas pela letra D, com letra e número em negrito e o enunciado em itálico. Da mesma forma, os teoremas são precedidos pela letra T e o número em ordem crescente, os algoritmos são precedidos pela letra A e as proposições por P, ambos negritados e com numeração.

Cada pareamento específico é denominado por μ e a denominação formal de *matching* é:

D.1. Um ***matching** μ* é uma correspondência um-a-um de um conjunto $H \cup W$ sobre si mesmo, e de ordem dois (isto é $\mu^2(x) = x$) tal que se $\mu(h) \neq h$ então $\mu(h) \in W$ e se $\mu(w) \neq w$ então $\mu(w) \in H$. Nos referimos $\mu(x)$ como o **parceiro** de x .

⁴⁸ Na seção 3.5, a analogia com a poligamia será discutida em relação aos modelos de Becker (1991, cap. 3).

⁴⁹ O tamanho dos grupos é alterado no modelo *many-to-one*. Percebe-se aqui que, em geral, tanto a literatura de *two-sided markets* quanto a de *assortative matings* utilizam as letras M , do inglês *men* e W , de *women* para designar esses dois grupos. Como em português M pode ser confundido como o designador do grupo de mulheres, e na nomenclatura aqui proposta a variável de alunos migrantes será denominada por M , adotou-se a convenção pouco comum, mas que atende aos propósitos, de chamar o grupo dos homens por H e continuar com W para o grupo das mulheres (observa-se que também se empregou o W para variáveis ligadas ao nível dos bairros, mas essas variáveis foram menos frequentes, quando oportuno, a distinção será feita no capítulo 6).

Exemplificando em detalhes, dado um homem h qualquer, se diz que a sua parceira no casamento μ será representado por $\mu(h) = w$. No caso, isso significa que no pareamento μ , o homem h está casado com a mulher w . Em outro *matching*, digamos que μ' , pode-se ter $\mu'(h) = z$, lê-se: “no pareamento μ' , o homem h está casado com a mulher z ”.

O mesmo vale para as mulheres, se $\mu(h) = w$, tem-se que, para esse mesmo *matching* μ , a mulher w deve estar casada com o homem h , representado por $\mu(w) = h$. Existe ainda a possibilidade de um homem e uma mulher estarem solteiros. Nesse caso, a notação diz que a pessoa está casada consigo mesma. Sendo assim, caso o homem h esteja solteiro, a notação é: $\mu(h) = h$. A definição **D.1** diz que se um homem ou uma mulher não estão solteiros eles necessariamente estão casados com alguém do outro grupo.

Ao elencar todos os pares (ou grupos, para o caso de escolas e alunos que se verá mais à frente) de um *matching* μ , μ será denominado por três principais maneiras (exemplo de três homens e três mulheres):

$$\mu = \{(h_1, w_1), (h_2, w_2), (h_3, w_3)\}$$

ou:

$$\mu = \begin{matrix} w_1 & w_2 & w_3 \\ h_1 & h_2 & h_3 \end{matrix}$$

Ou ainda na forma de uma matriz binária em que 1 denota o par formado e 0 a ausência de pares:

$$\mu = \begin{matrix} & w_1 & w_2 & w_3 \\ \begin{matrix} h_1 \\ h_2 \\ h_3 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

As três formas acima representam o mesmo casamento (são os mesmos pares formados) alterando apenas a forma de apresentação. O primeiro caso mostra o conjunto de todos os pares

homem e *mulher* e será a forma privilegiada para uso no corpo do parágrafo e no texto dos teoremas apresentados.

A segunda forma de apresentação é uma matriz em que, por convenção, mostra os pares com mulheres em cima e os homens em baixo. Essa forma será utilizada essencialmente nos exemplos.

A terceira maneira de representar o *matching* é por meio da matriz em que os homens são representados nas linhas e as mulheres nas colunas, o *matching* é denominado pelo binário, a célula de valor 1 denota um casamento formado, 0 indica que não há casamento. Essa forma foi essencialmente usada para representações de exemplos com mais de quatro indivíduos de cada grupo e no caso de alunos e escolas (os alunos nas linhas e escolas nas colunas).⁵⁰ No caso da matriz, o *matching* é também representado por **M** em negrito.

O conjunto de todos os matchings possíveis é denominado por $\mathcal{M}$, no caso de dois homens e duas mulheres esse conjunto é o conjunto dos conjuntos de pares:

$$\mathcal{M} = \{ \{ (h_1, w_1), (h_2, w_2) \}, \{ (h_1, w_2), (h_2, w_1) \} \}$$

No caso de três homens e três mulheres há seis *matchings* que podem ser formados, no caso de quatro homens e mulheres, são 24 casamentos, e assim por diante para números maiores ($n!$).

Os casamentos (ou matrículas) ocorrem voluntariamente, por isso se diz que os casamentos são racionais e ninguém é obrigado a casar-se com quem não deseja. A definição de casamento racional é a que se segue (conforme fornecida por Roth & Sotomayor, 1990, p. 21):

D.2. O *matching* μ é **individualmente racional** se cada agente é aceitável para o seu parceiro. Isto é, um *matching* é individualmente racional se ele não é bloqueado por qualquer agente (indivíduo).

⁵⁰ A maneira matricial de representação foi também a forma essencialmente utilizada para o desenvolvimento dos algoritmos na linguagem R, ver anexo B.

Um tipo especial de *matching* é o *matching estável*. Os *matchings* estáveis são um subconjunto do conjunto de todos os *matchings* possíveis de um mercado de casamentos ($\mathcal{M}$):

D.3. Um *matching* μ é **estável** se não é bloqueado por nenhum indivíduo ou por nenhum par de agentes.

O par bloqueante se refere a qualquer par de indivíduos, casados ou não, que seja preferível ao par já formado. Um *matching* estável não possui nenhum indivíduo que prefira outro indivíduo do sexo oposto em relação ao seu parceiro atual, e no qual esse indivíduo também o prefira. Na definição formal:

D.4. Um par bloqueante é dado por um par h e w pareados sob um *matching* μ , sendo que existe algum μ' tal que $\mu'(h) \succ_h \mu(h)$ e $\mu'(w) \succ_w \mu(w)$. Os indivíduos h e w bloqueiam o casamento μ pois ambos podem melhorar no casamento μ' e μ é não estável. Um casamento necessita de apenas um par bloqueante para que seja não estável.

O exemplo 2 da seção 3.2 ajuda a compreender a definição de pares bloqueantes. Na definição de par bloqueante empregou-se o conceito de preferências em relação aos pares. Cada homem ou mulher dos conjuntos acima mencionados possui suas preferências, representadas por $\mathcal{P}$. Essas preferências são dadas para cada pessoa e sobre todos os membros do conjunto oposto (são completas). Por exemplo, as mulheres conseguem elencar todos os homens do mais preferido ao menos preferido. As preferências oferecem um ordenamento completo e estrito das opções de parceiros:

$$\mathcal{P}(w) = h_3, h_4, h_5, \dots, h_1$$

Em que $\mathcal{P}(w)$ é o ordenamento de preferências dado por uma mulher qualquer w . No caso, essa mulher prefere o homem h_3 ao h_4 , o h_4 ao h_5 e assim por diante. O último homem em sua lista de preferências é h_1 .

Tais preferências podem ser representadas pelo símbolo de ordem “precede a” ou “mais preferido a” ‘ $\succ$ ’, ou de “sucede a” ou “menos preferido a” ‘ $\prec$ ’. Assim, as preferências da mulher w podem também ser escritas como:

$$\mathcal{P}(w): h_3 \succ h_4 \succ h_5 \succ \dots \succ h_1$$

Em geral, serão abordadas prioritariamente preferências estritas em que se usa o símbolo de fortemente preferido ‘ $\succ$ ’ tal como o que foi usado acima. Em determinadas situações poderão ser mencionadas relações de pares que são “fracamente preferidos” ‘ $\succsim$ ’ ou mesmo o sinal de “indiferença” ‘ $\sim$ ’:

$$\mathcal{P}(w): h_3 \sim h_4 \succsim h_1$$

Que corresponde a dizer que a mulher w é indiferente entre os homens h_3 e h_4 e que os prefere fracamente a h_1 (eles são ao menos tão bons quanto). No entanto, por motivos que ficarão mais claros em 3.2, na maioria das situações, as preferências serão tratadas como preferências estritas.⁵¹

Do mesmo modo, os homens também podem ter suas preferências representadas. Para um homem h :

$$\mathcal{P}(h) = w_9, w_1, w_3, \dots, w_8$$

Ou:

$$\mathcal{P}(h): w_9 \succ w_1 \succ w_3 \succ \dots \succ w_8$$

Como as preferências estão presentes para todos os indivíduos do mercado de casamentos, pode-se representar uma preferência com o subscrito: $h_3 \succ_w h_4$, que significa que a mulher w prefere h_3

⁵¹ Se a pessoa é indiferente entre duas ou mais opções, uma maneira de fazer com que ela decida é estabelecer um sorteio. Já que a pessoa é indiferente ao parceiro, ela também será indiferente a qualquer resultado do sorteio. Para a mulher w , por exemplo, o sorteio pode levar a uma decisão por $h_3 \succ h_4$, ou $h_4 \succ h_3$, e ela estará igualmente satisfeita. Supondo que na ocorrência de indiferença existe uma regra de decisão, fica mais fácil entender o porquê das preferências estritas.

estritamente a h_4 . No caso do homem h , tem-se: $w_3 \succ_h w_8$. A representação acima pode ser usada para *matchings* inteiros: $\mu \succ_w \mu'$, que diz que a mulher w prefere o *matching* μ ao μ' . E pode-se aplicar a grupos inteiros, todo o grupo das mulheres, por exemplo: $\mu \succ_W \mu'$, que quer dizer que $\mu \succ \mu'$ é válida para cada mulher pertencente à W . Uma condição mais fraca é: $\mu \succeq_W \mu'$, que significa que pelo menos uma mulher de W prefere μ estritamente à μ' , e as outras podem ser indiferentes ou não.

As preferências racionais são também transitivas, ou seja, se $w_1 \succ w_2$, e $w_2 \succ w_3$, tem-se que, por consequência da transitividade, $w_1 \succ w_3$.

Em se tratando de *matchings* estáveis propostos pelo algoritmo *deferred acceptance*, a notação μ_H e μ_W significa que o *matching* foi obtido pelo Gale-Shapley com homens e mulheres propondo, respectivamente.

Com os elementos dados, o mercado de casamentos pode ser representado pela tripla: $(H, W, \mathcal{P})$. Um conjunto de *homens*, outro de *mulheres*, e as *preferências racionais* conhecidas de cada um dos agentes para os membros do outro grupo. Essas informações serão trabalhadas para a formalização dos algoritmos e apresentação de exemplos da seção seguinte.

Ao tratar de alunos e escolas a notação muda ligeiramente, os *homens* são substituídos por *alunos*, que são os proponentes nos algoritmos simulados. E as *mulheres* são substituídas pelas *escolas*, que recebem as propostas. O grupo dos alunos é dado por I e o das escolas por S , os conjuntos são finitos (de tamanho n) e disjuntos, ou seja, $I \cap S = \emptyset$. Um exemplo para os conjuntos com n alunos e m escolas ($n \neq m$) segue abaixo:

Conjunto dos alunos: $I = \{i_1, i_2, i_3, \dots, i_n\}$

Conjunto das escolas: $S = \{s_1, s_2, s_3, \dots, s_m\}$

Uma diferença essencial do mercado de alunos e escolas é que os pares não precisam ser tomados *um a um*, uma mesma escola recebe vários alunos até o limite da sua capacidade. Por isso, é necessário especificar um vetor de capacidades, esse vetor foi chamado de Q :

Vetor de capacidades das escolas: $Q = (q_1, q_2, q_3, \dots, q_m)$

O q_1 é um número inteiro que indica quantas vagas a escola s_1 possui, o q_2 indica quantas vagas possui a escola s_2 , e assim por diante, até a informação de capacidades da última escola do conjunto S , q_m . Em um sistema com três escolas em que a primeira possui 5 vagas, a segunda 2 vagas e a terceira 6, tem-se o seguinte vetor Q :

$$Q = (5, 2, 6)$$

De maneira semelhante ao descrito para o mercado de casamentos, cada aluno i possui preferências completas para as escolas presentes no sistema e cada escola possui preferências completas pelos alunos.

Abdulkadiroğlu & Sönmez (2003) ressaltam que o termo “*prioridades*” se adapta melhor para o contexto das escolas, dado que nenhuma escola pública pode discriminar alunos ou preferir uns alunos a outros. As prioridades ressaltam aspectos administrativos, sociais, econômicos e são critérios que a sociedade estipula para priorizar as vagas escassas. Em termos do conceito teórico de prioridade, não há diferença para a determinação do algoritmo, tanto que continua sendo representada por $\mathcal{P}$. Ao se referir às escolas, são utilizados nesta tese tanto os termos preferências como prioridades, não se fazendo distinção entre os conceitos, a não ser quando explicitado textualmente.

Um ponto importante para o entendimento do funcionamento dos algoritmos e dos elementos de teoria dos jogos presentes na teoria de alocação envolve a declaração das preferências pelos agentes. O conceito de preferências até aqui ressaltado se refere às verdadeiras preferências dos indivíduos e escolas. Entretanto, uma ação possível para os jogadores é declarar um ordenamento diferente do verdadeiro. Nesse caso se diz que os agentes repassam um vetor com informações sobre as preferências ($\mathcal{J}$). Em teoria de jogos esse vetor pode ser interpretado tal como uma ação ou estratégia dos agentes.

Um conceito de jogos que ajuda na predição de possíveis equilíbrios estabelece que para compensar declarar sempre as verdadeiras preferências, $\mathcal{R}(i)$ deve ser uma estratégia dominante para o jogador i , ou seja, a ação que o jogador i adota é igual a declarar as verdadeiras preferências, ou seja, escolher a ação em que $\mathcal{R}(i) = \mathcal{R}(i)$.

D.5. Uma *estratégia é dominante* para um agente i ocorre quando $\mathcal{R}^*(i)$ é a melhor resposta para todos os possíveis conjuntos de escolhas estratégicas de $\mathcal{J}_i$ tomada por todos os demais agentes.

O $-i$ representa todos os demais agentes menos o i . Quando todos os jogadores possuem estratégias dominantes, o jogo possui necessariamente um único equilíbrio de Nash.⁵² Na seção 3.3.1, se verá que em um sistema sob o funcionamento do algoritmo *deferred acceptance* com alunos propondo, declarar as verdadeiras preferências é uma estratégia dominante para todos os alunos. Isso não será necessariamente verdadeiro para as escolas, mas o desenho do sistema pode fazer com que as escolas declarem suas verdadeiras preferências seja também a estratégia dominante, proporcionando um funcionamento harmônico dos algoritmos, tal como descrito com verdadeiras preferências e de interesse dos jogadores.

Uma definição que será muito empregada neste capítulo, e também nos posteriores, é a definição de ótimo de Pareto e de uma melhoria de Pareto. Adaptando-se o conceito de Mas Colell et al. (1995, p. 313) para o conceito do mercado de casamentos temos **D.6**:

D.6. Um casamento factível μ é Pareto ótimo (ou Pareto eficiente) se não há nenhum outro casamento factível μ' tal que $\mu' \succeq_h \mu$ para todo $h \in H$ e $\mu' \succ_h \mu$ para pelo menos um h qualquer tal que $h \in H$. Definição que pode ser generalizada também para as mulheres, $w \in W$.

⁵² O equilíbrio de Nash (1951) ocorre quando os agentes adotam a melhor resposta em um jogo não-cooperativo. O fato de alguns participantes poderem declarar preferências não verdadeiras insere elementos de jogos não cooperativos nos modelos de *matching*. Definições formais dos conceitos de teoria dos jogos podem ser encontradas em Mas-Colell, Whinston & Green (1995, cap. 7-9)

Em palavras, uma melhoria de Pareto significa melhorar pelo menos uma pessoa sem piorar nenhuma outra, situações em que um casamento pode melhorar a todos os indivíduos são passíveis de melhorias de Pareto. Na teoria de *two-sided matchings* existem alguns limites para o ponto em que se pode melhorar a todos participantes com um casamento (veja **T.3** mais adiante). O capítulo 4 trata dessas possibilidades e limitações em maiores detalhes.

Um mercado de matrículas, ou sistema de matrículas, pode ser representado pela quintupla $(I, S, \mathcal{P}, Q, \mathcal{J})$, um conjunto finito de alunos, um conjunto finito de escolas, a relação de preferências para cada aluno e escolas, o vetor de vagas das escolas e, por fim, como os participantes declaram suas preferências. Com as informações disponíveis pela tupla é possível a obtenção do *matching*.

Na nomenclatura utilizada, o Y é reservado ao produto do casamento. No caso de interesse, do *matching* entre alunos e escolas, Y representa a soma das proficiências. O resultado da proficiência individual é dado por y ou por y_i . A proficiência por escolas é dada por y_s , e nos dados será geralmente representada pela proficiência média da escola. Nota-se que nos modelos de regressão a serem abordados no capítulo 6, y é a variável independente do modelo, X está ligada ao conjunto de variáveis dos alunos, Z ao conjunto de variáveis das escolas e W com o de regiões e bairros (não confundir com W que define o conjunto das mulheres, o texto faz a distinção quando oportuno).

Finalmente, tem-se que nos modelos de utilidade do casamento cada alocação, denotada de maneira genérica por μ , deriva uma utilidade do casamento, denotada como produto do casamento Y . O produto do casamento de um homem h e uma mulher w é denotado por Y_{hw} , e o mesmo vale para o “casamento” entre alunos e escolas. Y_{is} é o “casamento” entre o aluno i e a escola s . Outros detalhes da notação aparecerão explicitados ao longo do texto.

3.2. O Algoritmo de Gale-Shapley para casamentos (*one-to-one*)

Ao propor inicialmente o modelo que deu origem ao artigo de 1962, os professores e matemáticos David Gale e Lloyd Shapley pensaram em um problema prático de alocação de

alunos e escolas (ou faculdades, uso mais comum da tradução do inglês *colleges*). Os autores observaram que inicialmente seria mais fácil reduzir o problema a uma solução *one to one*, sendo proposto, então, o modelo de casamentos, mais simples que os modelos de alocação de alunos e escolas.

Para essa literatura um casamento (ou pareamento) estável é fornecido quando não há nenhum indivíduo já pareado que prefira estar com algum outro indivíduo também pareado que o prefira.⁵³ Em outras palavras, se temos o *matching* μ dado pelos pares (h_1, w_1) e (h_2, w_2) , se a mulher w_1 prefere o homem h_2 ao seu par atual $(h_2 \succ_{w_1} h_1)$, tem-se que ou h_2 não prefere w_1 ao seu par atual, ou então esse casamento não seria estável. Se h_2 prefere w_1 , μ não seria à prova de divórcio, pois, w_1 e h_2 estariam melhor se estivessem juntos. Em casamentos estáveis não há pares melhores a serem formados (rever definição D.3).

Como visto em D.4, o bloqueio por um ou mais agentes ocorre quando um dos agentes pode melhorar de casamento e algum outro agente o prefere ou é indiferente ao indivíduo atual com quem está casado. Grosso modo, o par bloqueante é um par em que os agentes podem melhorar casando entre si.⁵⁴

Exemplo.2: Um pequeno exemplo pode ajudar a ilustrar esse conceito. Suponha-se três homens, $H = \{h_1, h_2, h_3\}$ e três mulheres, $W = \{w_1, w_2, w_3\}$, e que suas preferências sejam, respectivamente:

Homens:

Mulheres:

$$\mathcal{P}(h_1) = w_1, w_2, w_3$$

$$\mathcal{P}(h_2) = w_2, w_3, w_1$$

$$\mathcal{P}(h_3) = w_2, w_1, w_3$$

$$\mathcal{P}(w_1) = h_2, h_3, h_1$$

$$\mathcal{P}(w_2) = h_3, h_1, h_2$$

$$\mathcal{P}(w_3) = h_1, h_2, h_3$$

⁵³ Inclui o pareamento consigo mesmo, exemplo: $\mu(h) = h$.

⁵⁴ Note-se que casar um *par bloqueante* entre si não é uma condição suficiente para produzir um casamento final e estável. Porém, pares bloqueantes podem casar recursivamente entre si até que não exista nenhum par bloqueante final, o conceito dessa solução foi proposto por Knuth (1976), porém, essa solução necessita da escolha de um casamento correto entre os pares bloqueantes, pois senão se teria um loop infinito tal como mostrado em Knuth (1976, p. 14-15) e reproduzido em Roth & Sotomayor (1990, p.27-28). Roth & Vande Vate (1990) propuseram uma regra de escolha do par bloqueante a ser casado que sempre conduz para a solução estável.

Um dos 6 (3!) pareamentos de casais possíveis do exemplo pode ser dado pelo casamento μ abaixo:

$$\mu = \begin{matrix} w_1 & w_2 & w_3 \\ h_3 & h_1 & h_2 \end{matrix}$$

Porém o *matching* acima não é estável pois o par (h_3, w_2) bloqueia o atual *matching*, pois h_3 prefere w_2 a seu par atual e w_2 prefere h_3 , ao invés do par já formado. Um novo *matching* pode ser proposto:

$$\mu' = \begin{matrix} w_1 & w_2 & w_3 \\ h_2 & h_3 & h_1 \end{matrix}$$

O casamento μ' não possui pares bloqueantes e é a alocação ótima do ponto de vista das mulheres, pois todas elas estão com o par mais desejado. Podemos denotar que $\mu' \succ_W \mu$, ou seja, que alocação μ' é fortemente preferida a μ por todas as mulheres. O mesmo não ocorre para a preferência masculina, dado que sob μ' dois deles (h_1, h_2) estão em pior situação e apenas um (h_3) com melhor alocação do que em μ .

Para encontrar pares estáveis Gale & Shapley (1962) propuseram o algoritmo *deferred-acceptance one-to-one*. Um teorema proposto por Gale Shapley (1962, p.12) mostra que haverá sempre pelo menos uma alocação estável para qualquer mercado de casamentos:

T.1. *Existe um matching estável para todo e qualquer mercado de casamentos.*

A prova de **T.1** pode ser obtida por construção. Os autores construíram um algoritmo com um número finito de passos e que permite sempre alcançar a alocação ótima para um dos lados do mercado (homens ou mulheres). Abaixo está descrito o algoritmo em quatro passos (tal como exposto na forma de parágrafos por Gale & Shapley, p. 12). O algoritmo foi escrito com os homens propondo primeiro:

A.1. Algoritmo de Gale-Shapley:

Passo 1. Cada homem propõe casamento para sua mulher favorita. Cada mulher que tiver recebido mais de uma proposta, compromete-se momentaneamente com o homem mais preferido que tenha lhe proposto, e rejeita todos os outros que estavam abaixo do homem mais bem cotado que lhe ofereceu proposta. Não há casamento ainda, as mulheres que receberam propostas se comprometem, mas podem receber novas propostas de candidatos mais bem *rankeados*.

Passo 2. Um ou mais homens que foram rejeitados no primeiro passo podem propor casamento para sua segunda mulher favorita. Cada mulher que tiver recebido propostas deve se comprometer com o seu par mais preferido. A mulher que já tinha recebido proposta no passo 1 e recebe proposta de novo candidato, deve avaliar se o candidato é mais ou menos preferido do que seu atual compromisso. Caso seja mais preferido, ela deve comprometer-se com o novo par. Caso seja menos preferido, ela deve rejeitá-lo.

Passo 3. O algoritmo segue o mesmo procedimento do passo anterior. Todos aqueles um ou mais homens que foram rejeitados no segundo estágio devem propor para sua próxima escolha, e as mulheres novamente devem rejeitar todos, menos o mais preferido que elas tiveram até então. Caso pelo menos uma mulher tenha recebido mais de uma proposta o algoritmo continua para o passo seguinte.

Passo 4. O algoritmo só finaliza quando todas as mulheres tiverem recebido propostas. Nenhum homem pode propor a uma mesma mulher mais de uma vez. Após a última mulher ter recebido sua proposta, não haverá mais nenhuma rodada.⁵⁵

O algoritmo pode ser escrito para um caso geral e de maneira mais sucinta:

⁵⁵ Se a última mulher receber mais de uma proposta, isto se configurará em uma contradição, pois ela não terá sido realmente a última, pois deverá rejeitar o menos preferido, e o homem rejeitado deverá propor a uma nova mulher, alguma que ainda não tenha recebido propostas.

A.2. Algoritmo de Gale-Shapley (forma geral):

Passo 1. Cada homem propõe casamento para sua mulher favorita. Caso nenhuma mulher tenha recebido mais de uma proposta o algoritmo termina no primeiro passo. Caso pelo menos uma mulher tenha recebido mais de uma proposta, a mulher se compromete momentaneamente com seu par melhor classificado (segundo suas preferências) e rejeita os homens que estão abaixo do homem mais bem cotado que lhe ofereceu proposta.

Passo k. Um passo genérico k , onde $k \geq 2$. Cada homem rejeitado no passo anterior propõe para sua próxima mulher mais preferida. A mulher que receber mais de uma proposta avalia os candidatos e “segura” aquele mais bem cotado pela sua classificação, todos os demais são rejeitados. Caso não ocorra mais de um homem propondo a uma mesma mulher o algoritmo termina e as mulheres se casam com os pares que estavam até então “segurando” como sua melhor opção.

O algoritmo termina quando mais nenhum homem é rejeitado e todas as mulheres tenham recebido ao menos uma proposta. Em um mercado em que todo indivíduo do outro grupo é considerado aceitável, ninguém fica sozinho e todos se casam.

Logicamente, o algoritmo Gale-Shapley pode ser reescrito para mulheres propondo os homens em casamento. O número máximo de passos que o algoritmo pode alcançar é $k = n^2 - 2n + 2$, em que n é o número de jogadores de um determinado conjunto. O número de passos pode ser diferente de acordo com o grupo que propõe. Com sorte, o algoritmo termina na primeira rodada, mas isso se torna cada vez menos provável para conjuntos de tamanhos muito grandes (n grande) e preferências em que as pessoas possuam alguma similaridade de preferências (preferências correlacionadas).⁵⁶

⁵⁶ Para preferências dadas aleatoriamente, com n muito grande, a probabilidade de que ocorra pelo menos um par empatado em pelo menos uma das colunas é virtualmente 1, ou seja, provavelmente ocorre o empate e o algoritmo não termina na primeira rodada. Supondo um algoritmo com homens propondo, essa probabilidade pode ser dada por: $P(\text{dois ou mais homens escolherem a mesma mulher}) = 1 - P(\text{cada homem escolher uma mulher diferente})$. A probabilidade de cada um dos homens escolher uma mulher diferente é dada por: $\frac{n!}{n^n}$, número de *matchings* de casamentos únicos (*one-to-one*) sobre total de combinações possíveis para as proposições que podem ser realizadas na primeira rodada. A probabilidade de empate é $P(E) = 1 - \frac{n!}{n^n}$, $n \in H, W$. Como $\lim_{n \rightarrow \infty} (\frac{n!}{n^n}) = 0$, tem-se que $P(E) \approx 1$, a probabilidade de não ocorrer empate é muito rara para n muito grande. Faz-se a ressalva de que a teoria não se

Compatível com **T.1**, o algoritmo **A.2** sempre acha uma solução estável. O próprio algoritmo ajuda a ver uma prova por construção:

Prova: suponha que o algoritmo proponha um *matching* μ , e suponha-se que o homem h e a mulher w não estão casados um com o outro em μ , mas que h prefira w à parceira que obteve em $\mu(h) = w'$. Sendo assim, a mulher w é aceitável para o homem h , pois ele a prefere e, seguindo a ordem do algoritmo (**A.2**), ele deveria ter proposto a ela antes de w' . Mas como ele não está casado com a mulher w ao final do algoritmo, ou ele foi rejeitado por w , e então ela não o prefere mais do que seu par atual, ou então w e w' são a mesma mulher (opções idênticas), sendo que o homem h não pode dizer que prefere w a w' ($w \succ_h w'$) tal como afirmado no início, pela reflexividade das preferências racionais ($w \sim w$). Pelo algoritmo, o raciocínio se estende para qualquer homem h e mulher w . ■

A prova acima foi dada usando o algoritmo apresentado em **A.2**. Sotomayor (1996b) apresentou a prova analítica da estabilidade do casamento nos denominados mercados de dois lados (*two-sided*), isto é, a autora mostrou que é possível provar a existência do equilíbrio mesmo na ausência do algoritmo.

Exemplo.3: Um exemplo é importante para fixar o funcionamento das regras do *deferred acceptance*. Suponha o conjunto de três homens e três mulheres dados por:

$$H = \{\text{João}, \text{Joaquim}, \text{Raimundo}\} \quad W = \{\text{Lili}, \text{Maria}, \text{Teresa}\}$$

Suponha ademais que as preferências de cada um dos jogadores sejam as seguintes:

Homens:

$$\begin{aligned} \mathcal{P}(\text{João}) &= \text{Teresa}, \text{Maria}, \text{Lili} \\ \mathcal{P}(\text{Joaquim}) &= \text{Lili}, \text{Teresa}, \text{Maria} \\ \mathcal{P}(\text{Raimundo}) &= \text{Maria}, \text{Teresa}, \text{Lili} \end{aligned}$$

Mulheres:

$$\begin{aligned} \mathcal{P}(\text{Lili}) &= \text{João}, \text{Raimundo}, \text{Joaquim} \\ \mathcal{P}(\text{Maria}) &= \text{Joaquim}, \text{Raimundo}, \text{João} \\ \mathcal{P}(\text{Teresa}) &= \text{Raimundo}, \text{João}, \text{Joaquim} \end{aligned}$$

aplica à conjuntos com infinitos elementos. Para matrizes aleatórias, o número de passos do algoritmo não cresce na mesma velocidade que o número máximo de passos possíveis e, com isso, a fração do número de passos observados frente ao total de passos possíveis é decrescente, ver simulações no anexo B.

O algoritmo com os homens propondo às mulheres se soluciona no primeiro passo, pois os homens classificam em primeiro lugar mulheres diferentes. A TAB.2 abaixo traz a solução e o “X” marca os pares formados.

Tabela 2 - Resultado de um exemplo de solução estável com homens propondo

	<i>Lili</i>	<i>Maria</i>	<i>Teresa</i>
<i>João</i>			X
<i>Joaquim</i>	X		
<i>Raimundo</i>		X	

Para as mulheres propondo primeiro, o algoritmo se encerra no primeiro passo, pois os homens mais preferidos pelas mulheres são diferentes entre si e não há empate. Entretanto, nota-se que a solução para as mulheres propondo é diferente daquela dos homens propondo, isso porque não há coincidência das preferências entre os três jogadores de cada conjunto. A TAB.3 mostra a solução:

Tabela 3 - Resultado de um exemplo de solução estável com mulheres propondo

	<i>Lili</i>	<i>Maria</i>	<i>Teresa</i>
<i>João</i>	X		
<i>Joaquim</i>		X	
<i>Raimundo</i>			X

Exemplo. 4: Antes de analisar em mais detalhes as duas soluções propostas, segue outro exemplo para ilustrar o funcionamento do algoritmo com mais passos. Suponha agora quatro indivíduos em cada grupo e que as preferências de cada um dos jogadores são as seguintes:

Homens:

$\mathcal{P}(\text{Carlos}) = \text{Dulce, Lili, Teresa, Maria}$
 $\mathcal{P}(\text{João}) = \text{Lili, Maria, Dulce, Teresa}$
 $\mathcal{P}(\text{Paulo}) = \text{Maria, Teresa, Lili, Dulce}$
 $\mathcal{P}(\text{Raimundo}) = \text{Lili, Teresa, Maria, Dulce}$

Mulheres:

$\mathcal{P}(\text{Dulce}) = \text{João, Carlos, Paulo, Raimundo}$
 $\mathcal{P}(\text{Lili}) = \text{Carlos, João, Paulo, Raimundo}$
 $\mathcal{P}(\text{Maria}) = \text{Paulo, Raimundo, Carlos, João}$
 $\mathcal{P}(\text{Teresa}) = \text{Carlos, João, Raimundo, Paulo}$

No caso do exemplo acima, para as mulheres propondo, temos que no primeiro passo:

Passo 1. *Dulce* propõe à *João*, *Lili* propõe à *Carlos*, *Maria* propõe à *Paulo* e *Teresa* também propõe à *Carlos*. Como *Carlos* obteve duas propostas, de *Lili* e *Teresa*, ele aceita e “segura” a proposta que lhe é mais interessante, e no caso é *Lili*, sua segunda melhor opção, *Teresa* era terceira. Todos os demais homens “securam” a única proposta que tiveram até o momento. *Teresa* foi rejeitada e precisa de uma nova rodada.

Passo 2. *Teresa* propõe para sua segunda opção que é *João*. *João* deve então analisar entre *Teresa* e sua a proposta de *Dulce* que ele estava segurando desde o passo anterior. Como *Dulce* continua sendo sua opção mais preferida, *João* rejeita *Teresa*, que deverá seguir para nova proposta. Todos os demais homens que não receberam proposta nesse passo, continuam segurando sua melhor opção até o momento.

Passo 3. *Teresa* propõe à *Raimundo*, que até então não havia recebido propostas. *Raimundo* não estava “comprometido” com ninguém e ele aceita a proposta de *Teresa* e o algoritmo termina. Todos os homens confirmam a opção que estavam segurando.

Na TAB. 4, representativa da matriz de alocação final, temos o resultado do passo a passo acima. Caso o algoritmo fosse calculado com homens propondo, o resultado seria diferente (TAB. 5) e o algoritmo levaria dois passos:

Tabela 4 - Exemplo com 4 homens e 4 mulheres, solução estável com mulheres propondo

	<i>Dulce</i>	<i>Lili</i>	<i>Maria</i>	<i>Teresa</i>
<i>Carlos</i>		X		
<i>João</i>	X			
<i>Paulo</i>			X	
<i>Raimundo</i>				X

Tabela 5 - Exemplo com 4 homens e 4 mulheres, solução estável com homens propondo

	<i>Dulce</i>	<i>Lili</i>	<i>Maria</i>	<i>Teresa</i>
<i>Carlos</i>	X			
<i>João</i>		X		
<i>Paulo</i>			X	
<i>Raimundo</i>				X

Representando como μ_W o casamento das mulheres e como μ_H , pode-se obter duas outras opções alternativas para representar o *matching* (nomes dados pelas letras iniciais):

$$\mu_W = \begin{matrix} & L & D & M & T \\ C & & J & P & R \end{matrix} \quad \mu_H = \begin{matrix} & D & L & M & T \\ C & & J & P & R \end{matrix}$$

$$\mathbf{M}_W = \begin{matrix} & D & L & M & T \\ C & \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \\ J & \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \\ P & \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \\ R & \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \quad \mathbf{M}_H = \begin{matrix} & D & L & M & T \\ C & \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \\ J & \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \\ P & \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \\ R & \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

As formas de representação de matchings com as matrizes esparsas foi a preferencialmente utilizada para a programação dos algoritmos.⁵⁷ É fácil visualizar que os resultados dos dois *matchings* finais do algoritmo são diferentes no exemplo 3 e 4. Nas preferências novamente representadas abaixo, o resultado para alocação dos homens foi marcada com traço horizontal sobrescrito e para o caso das mulheres o resultado da alocação final está sublinhado.

Preferências do exemplo 3 com os nomes dados somente pela primeira letra (σ para o Joaquim):

Homens:

$$\mathcal{P}(J) = \bar{T}, M, \underline{L}$$

$$\mathcal{P}(\sigma) = \bar{L}, T, \underline{M}$$

$$\mathcal{P}(R) = \bar{M}, \underline{T}, L$$

Mulheres:

$$\mathcal{P}(L) = \underline{J}, R, \bar{\sigma}$$

$$\mathcal{P}(M) = \underline{\sigma}, \bar{R}, J$$

$$\mathcal{P}(T) = \underline{R}, \bar{J}, \sigma$$

Do exemplo 4 (nomes dados pela letra inicial):

Homens:

$$\mathcal{P}(C) = \bar{D}, \underline{L}, T, M$$

$$\mathcal{P}(J) = \bar{L}, M, \underline{D}, T$$

$$\mathcal{P}(P) = \underline{\bar{M}}, T, L, D$$

$$\mathcal{P}(R) = L, \bar{T}, M, D$$

Mulheres:

$$\mathcal{P}(D) = \underline{J}, \bar{C}, P, R$$

$$\mathcal{P}(L) = \underline{C}, \bar{J}, P, R$$

$$\mathcal{P}(M) = \underline{\bar{P}}, R, C, J$$

$$\mathcal{P}(T) = C, J, \underline{\bar{R}}, P$$

⁵⁷ O código do algoritmo *one-to-one* está disponibilizado textualmente no anexo B e o algoritmo *many-to-one* pode ser solicitado ao autor. Para maiores detalhes consulte o anexo B.

Em ambos os exemplos pode se reparar que, para os homens, o traço sobrescrito (indicativo do *matching* favorável aos homens) está sempre à esquerda ou na mesma posição do que o traço sublinhado (indicativo do *matching* favorável às mulheres). Para as mulheres, o traço sobrescrito está sempre à direita ou na mesma posição que o sublinhado. Isso quer dizer que, para os homens, a alocação final com homens propondo é sempre melhor ou tão boa quanto o a das mulheres propondo. E o contrário é também verdadeiro: para as mulheres, o *matching* das mulheres propondo é tão bom ou melhor que o *matching* dos homens propondo primeiro.

O resultado exemplificado acima não é por acaso, há um teorema importante que afirma:

T.2. *Quando todos os agentes possuem preferências estritas, existe sempre um matching estável H-ótimo e um matching estável W-ótimo. Além disso, o matching μ_H produzido pelo algoritmo deferred acceptance com homens propondo é o matching estável H-ótimo. O matching estável W-ótimo é o matching μ_W dado pelo algoritmo quando as mulheres propõem.*

Seguindo a intuição dada por Roth & Sotomayor (1990, p. 32), temos que a prova do teorema **T.2** se baseia no conceito de homens e mulheres atingíveis, isto é, alcançáveis ou factíveis, que podem de fato formar par com determinado indivíduo. Para o algoritmo *deferred-acceptance* com homens propondo primeiro, nenhum homem é rejeitado por uma mulher alcançável. Consequentemente, o casamento estável μ_H , produzido pelo algoritmo DA com os homens propondo, parecia cada homem com a melhor mulher atingível de cada um e é, portanto, o único *matching* estável H-ótimo.

Segue transcrição da prova por indução de Roth & Sotomayor (1990, p. 33) com pequenas alterações textuais para efeito de tradução ou explicações adicionais:

Prova: suponha que para um dado passo do algoritmo nenhum homem tenha sido ainda rejeitado por uma mulher alcançável por ele (pode ser o primeiro passo ou outros de maior índice). Neste passo, suponha que a mulher w rejeita o homem h . Se ela rejeita h como inaceitável, então essa mulher é inatingível para ele, e a análise termina. Se ela o rejeitar em favor de outro homem h' ,

com quem ela fica compromissada, então ela prefere h' à h . E é preciso mostrar que w não era alcançável para o homem h .

Sabe-se que h' prefere w à qualquer mulher, excluídas as que já o rejeitaram, e portanto, por indução (tal como no parágrafo acima), não eram atingíveis pelo homem h' . Considere o *matching* hipotético μ que pareia h a w e todos os demais a um par factível. Sabe-se que h' prefere w às outras mulheres por ele atingíveis e prefere essa mulher ao seu par atual em μ . Portanto, μ não pode ser estável, pois h' e w formam um par bloqueante, não há *matching* estável que una h e w , e, portanto, eles são mutuamente inatingíveis, isso completa a prova. ■

O resultado do T.2 pode ser generalizado. Ocorre que qualquer *matching* estável (não necessariamente *H-ótimo* e *W-ótimo*), terá uma oposição de interesses dos dois grupos. Se um *matching* qualquer μ é preferido pelo grupo dos homens em relação a μ' , então, μ' será preferido à μ pelo grupo das mulheres. Segue o teorema:

T.3. *Quando todos os agentes possuem preferências estritas, as preferências comuns dos dois lados do mercado são opostas no conjunto de matchings estáveis: se μ e μ' são matchings estáveis, então todos os homens preferem μ tão ou mais que μ' se e somente se todas as mulheres preferem μ' tão ou mais que μ . Isto é, $\mu \succ_H \mu'$ se e somente se $\mu' \succ_W \mu$.*⁵⁸

A prova consiste em mostrar que se $\mu \succ_H \mu'$ ocorre, então $\mu' \succ_W \mu$ deve necessariamente ocorrer para as mulheres.

Prova-se pela contradição: suponha que $\mu' \succ_W \mu$ não seja verdade, sendo assim, já que as preferências são estritas, pelo menos alguma mulher w prefere o *matching* μ ao invés do μ' . Se isso ocorre, sabemos que o par para a mulher w em μ é diferente do *matching* μ' , e que a mulher não estava solteira em μ , pois o par aceitável de μ' é preferível a ficar solteira. Então w tem um homem h preferível em μ , $\mu(w) = h$. Como, no caso, μ representa o *matching* proposto pelos homens, sabe-se que h também prefere w à mulher dada em μ' . Desse modo como h e w se

⁵⁸ Roth e Sotomayor (1990, p. 33) atribuem esse teorema a Knuth (1976).

preferem mutuamente, esse casal formaria um par bloqueante ao μ' , contradizendo que μ' é um *matching* estável. Portanto, para ser estável, é preciso que $\mu' \succ_w \mu$ ocorra. ■

Muitos outros exemplos e aplicações interessantes do algoritmo Gale-Shapley estão disponíveis na literatura.⁵⁹ Um survey importante de todo o desenvolvimento desta literatura pode ser encontrado em Roth (2002) e Roth (2008).

Um último teorema apresentado nessa seção estipula a relação entre o equilíbrio ótimo de Pareto e os *matching* estáveis ótimos dados pelos algoritmos *deferred-acceptance*. O teorema se encontra tal como enunciado por Roth & Sotomayor (1990, p.46) com homens propondo:

T.4. *Teorema do Fracamente ótimo de Pareto para os homens. Não há nenhum matching μ de racionalidade individual (estável ou não estável) tal que μ seja preferido ao matching estável ótimo μ_H , simbolicamente: $\nexists \mu$ tal que $\mu \succ_H \mu_H$, para todo homem $h \in H$.*

O teorema **T.4** afirma que não há condição de melhorar todos os homens estritamente em relação a seu *matching* estável ótimo, o *matching* estável *H*-ótimo é o melhor possível ou tão bom quanto qualquer outro que possa ser encontrado. Esse teorema pode ser invertido para o caso das mulheres, e na seção 3.3, será novamente abordado.

3.2.1. Manipulação do resultado por parte dos agentes

Uma extensão importante dos modelos é que nem sempre as preferências são completas ou nem sempre todos os pares são aceitáveis. Como ressaltado em 3.1, em alguns casos, a pessoa preferiria estar solteira a estar casada com as opções disponíveis. Em termos de linguagem de *matching*, uma pessoa solteira está casada consigo mesma. Essas preferências poderiam ser representadas da seguinte maneira:

⁵⁹ Para uma listagem várias aplicações da literatura, a página pessoal do prof. Avin E. Roth tem vários endereços: <http://kuznets.fas.harvard.edu/~aroth/alroth.html> (último acesso 29/04/13). Para exemplos didáticos simples, o site educativo do grupo de estudos em matemática para alunos do fundamental e médio, o [Math Circle](http://mathcircle.berkeley.edu/) da comunidade de Berkeley apoiado pela Universidade e comunidade de Berkeley, mantém um pequeno *software* mostrando os conceitos básicos da solução de Gale Shapley. Os exemplos tratam de pequenos casos para matrizes de 4x4 até 6x6: <http://mathsite.math.berkeley.edu/smp/smp.html> (último acesso 29/04/13).

$$\mathcal{P}(Lili) = Lili$$

No caso acima, *Lili* informa que prefere a si mesma do que estar casada com os outros pares disponíveis, prefere estar solteira. Em outros casos, ela pode considerar apenas dois pares aceitáveis e preferir estar solteira a qualquer outro:

$$\mathcal{P}(Lili) = Carlos, João, Lili$$

Os estudos de Dubins & Freedman (1981), e a literatura que se seguiu posteriormente, enfatizam quais são as possibilidades dos agentes reportarem uma preferência distinta da verdadeira e conseguirem com isso manipular o resultado do algoritmo. Há situações em que as preferências podem ser truncadas, caso do exemplo 5 abaixo.

Exemplo. 5: imagine que *Dulce* possua as verdadeiras preferências dadas por:

$$\mathcal{P}(Dulce) = João, Carlos, Paulo, Raimundo$$

Mas informa as seguntes preferências truncadas:

$$\mathcal{P}(Dulce) = João, Dulce$$

Dulce informa que prefere *João* e mais ninguém, ou que prefere estar casada consigo mesma a estar casada com *Carlos*, *Paulo* ou *Raimundo*. Parte da literatura de *matching* se dedica ao estudo de estratégias de jogadores que revelam preferências diferentes das verdadeiras, tentando influenciar o resultado do algoritmo em favor da solução estável que lhes é mais favorável. Observe-se que o *matching* estável para *Dulce* quando as mulheres propõem é diferente do resultado que ela recebe com os homens propondo. No primeiro caso ela se casaria com *João* e no segundo com *Carlos*.

Em situações com muitos participantes, truncar as preferências é uma estratégia arriscada. Supondo que as preferências completas apresentadas no exemplo 4 são as verdadeiras, *Dulce* poderia tentar censurar suas preferências informadas para tentar se casar com *João*. Jogando

dessa forma, *Dulce* conseguiria “forçar” o algoritmo dos homens a oferecer um resultado favorável para as mulheres. Todos os demais jogadores declaram as preferências verdadeiras. Se *Dulce* declara $\mathcal{P}(Dulce) = João, Dulce$, o algoritmo se encerra em 4 passos:

- Passo 1. *Carlos* propõe à *Dulce*, *João* propõe à *Lili*, *Paulo* propõe à *Maria* e *Raimundo* propõe à *Lili*. Como *Dulce* está maqueando suas preferências, ela prefere estar casada consigo mesma a aceitar a proposta de *Carlos*, enquanto que *Lili*, “segura” a proposta de *João* que é o mais preferido na ordem listada. *Maria* fica momentaneamente comprometida com *Paulo*.
- Passo 2. *Carlos* e *Raimundo* têm que propor para a segunda mais preferida. *Carlos* propõe à *Lili* e *Raimundo* propoe à *Teresa*. *Teresa* ainda não havia recebido nenhuma proposta e então aceita e “segura” a proposta de *Raimundo*. *Lili* recebeu uma proposta mais preferida, ela aceita a proposta de *Carlos* e rejeita a de *João*.
- Passo 3. *João* propõe à *Maria*, sua segunda opção. *Maria* possui agora duas propostas: a nova de *João* e a de *Paulo* (com quem já estava compromissada). No entanto *Maria* prefere *Paulo*, ela mantém o compromisso e rejeita a proposta de *João*.
- Passo 4. *João* propõe à *Dulce* que até então estava casada consigo mesma, como a proposta de *João* foi considerável aceitável por *Dulce*, ela confirma a proposta e o algoritmo se encerra.

O resultado final são os pares: *João* e *Dulce*, *Carlos* e *Lili*, *Paulo* e *Maria* e *Raimundo* e *Teresa*. *Dulce* escolheu declarar apenas *João* como preferido e com isso “forçou” a solução ótima para as mulheres. Com sua escolha *Dulce* praticamente optou por encerrar o algoritmo apenas no momento em que recebesse a proposta feita pelo homem mais desejado.

Entretanto, trata-se de uma estratégia arriscada. Em casos em que há mais mulheres do que homens e algumas mulheres ficam necessariamente solteiras, se o corte das preferências não for suficientemente preciso, no melhor homem alcançável, a estratégia falha. *Dulce* ficaria solteira se errasse nessa estratégia, mesmo que de acordo com suas verdadeiras preferências houvesse um par possível.

Verificar se o mecanismo é passível de manipulação por parte de membros dos dois grupos é uma preocupação de importantes implicações práticas. Em especial, na seção a seguir, se apresenta a extensão do algoritmo para o caso *many-to-one*, caso em que a possível manipulação é importante, considerando aspectos práticos e teóricos.

3.3. Extensão do Algoritmo de Gale-Shapley para casamentos (*many-to-one*)

A mudança de um algoritmo de *one-to-one* para um *many-to-one* se produz alterando um pouco o seu funcionamento e interpretação, imaginando-se o grupo de alunos (I) no lugar do grupo dos homens, e o das escolas (S) no lugar do grupo das mulheres. A diferença é que cada escola possui um limite superior de vagas para aceitar os alunos. A fim de estender o raciocínio da seção anterior para o caso *many-to-one* é possível imaginar cada vaga como uma única escola e os alunos são indiferentes a qualquer vaga de uma mesma escola.⁶⁰

Se pensarmos que uma escola s qualquer aceita uma quantidade ‘ q ’ de alunos, teremos que um casamento não será estável enquanto ainda existirem vagas e candidatos aceitáveis interessados.

Imagine-se que uma aluna chamada *Maria* que prefira as escolas ‘ A ’, ‘ B ’ e ‘ C ’ nesta sequência. Suponha também que cada uma dessas escolas tenha 20 vagas. *Maria* é indiferente a qualquer vaga de ‘ A ’, sendo ela a primeira ou a vigésima: $A_1 \sim_{\text{Maria}} A_{20}$, mas prefere estritamente qualquer vaga de ‘ A ’ a alguma de ‘ B ’: $A_q \succ_{\text{Maria}} B_q$.

O algoritmo para o tratamento do caso *many-to-one* é obtido expandindo-se as preferências de *Maria*, e colocando-se cada vaga como se fosse uma única instituição com uma vaga somente. A preferência de *Maria*, que no modelo *one-to-one* era descrita assim: $\mathcal{P}(\text{Maria}) = A, B, C$. Passa a ser descrita da maneira apresentada abaixo:

$$\mathcal{P}(\text{Maria}) = A_1, A_2, \dots, A_{20}, B_1, B_2, \dots, B_{20}, C_1, C_2, \dots, C_{20}$$

⁶⁰ Essa foi a aproximação utilizada para ir do algoritmo geral *one-to-one* para a versão *many-to-one*, ver anexo B.

As preferências para cada vaga de uma mesma escola são iguais. Ou seja uma escola tem uma preferência única que é copiada para todas as suas vagas. Expandindo-se as preferências dos alunos até o número q de vagas por escola, sempre respeitando o ordenamento conjunto das escolas nas preferências do aluno. A solução do *many-to-one* é essencialmente a mesma do *Gale-Shapley* proposto para casamentos.

A descrição dos passos do algoritmo *many-to-one* com alunos propondo às escolas:

A.3. Algoritmo de Gale-Shapley *many-to-one* (forma geral):

Passo 1. Cada aluno pleiteia a vaga na primeira escola de sua escolha. A escola que tenha recebido um número de proposta superior ao total de vagas disponíveis escolhe os alunos de acordo com sua lista de prioridades e rejeita os demais alunos excedentes.

Passo k . No passo genérico k ($k \geq 2$). Cada aluno rejeitado no passo anterior propõe para sua próxima vaga mais preferida. Cada escola considera os alunos que estava “segurando” em relação às novas propostas que por ventura tenha recebido. Caso algum aluno tenha prioridade maior que os que ela estava retendo, a proposta desse aluno será aceita e o aluno com menor prioridade será rejeitado. Caso não ocorra mais nenhum aluno propondo à uma escola já com todas as vagas preenchidas, o algoritmo termina e as escolas devem, de fato, matricular as propostas recebidas.⁶¹

Exemplo. 6: Pode se ilustrar o funcionamento do algoritmo *many-to-one* com um exemplo famoso de Roth (1985, p. 283) com quatro alunos: i_1, i_2, i_3, i_4 e três escolas: s_1, s_2, s_3 . Como o número de vagas tem de ser igual ou maior ao número de alunos, suponha que a primeira escola tenha duas vagas e a segunda e terceira tenham 1 vaga cada uma, o que fornece o seguinte vetor de vagas: $Q = (q_1 = 2, q_2 = 1, q_3 = 1)$.

As preferências de alunos e escolas são estas abaixo:

⁶¹ Detalhes do código para o algoritmo implementado estão no anexo B.

Alunos:

Escolas:

$$\mathcal{P}(i_1) = s_3, s_1, s_2$$

$$\mathcal{P}(s_1) = i_1, i_2, i_3, i_4$$

$$\mathcal{P}(i_2) = s_2, s_1, s_3$$

$$\mathcal{P}(s_2) = i_1, i_2, i_3, i_4$$

$$\mathcal{P}(i_3) = s_1, s_3, s_2$$

$$\mathcal{P}(s_3) = i_3, i_1, i_2, i_4$$

$$\mathcal{P}(i_4) = s_1, s_2, s_3$$

$$Q = (q_1 = 2, q_2 = 1, q_3 = 1)$$

Operacionalizando a matriz acima, como a escola s_1 possui duas vagas, as preferências dos alunos podem ser escritas da seguinte maneira:

Alunos:

$$\mathcal{P}(i_1): s_3 \succ s_{1,1} \sim s_{1,2} \succ s_2$$

$$\mathcal{P}(i_2): s_2 \succ s_{1,1} \sim s_{1,2} \succ s_3$$

$$\mathcal{P}(i_3): s_{1,1} \sim s_{1,2} \succ s_3 \succ s_2$$

$$\mathcal{P}(i_4): s_{1,1} \sim s_{1,2} \succ s_2 \succ s_3$$

Em que $s_{1,1} \sim s_{1,2}$ indica que os alunos são indiferentes entre qualquer vaga da escola s_1 . Como visto em 3.1, em caso de indiferença, pode-se tomar uma ordem aleatória ou arbitrária para que as preferências continuem estritas e se consiga resolver o algoritmo. Digamos que essa ordem arbitrária seja em favor da vaga 1 para todos os alunos, ou seja, $s_{1,1} \succ s_{1,2}$ para qualquer aluno $i \in I$. As vagas de s_1 possuem as mesmas preferências de $\mathcal{P}(s_1)$. Sendo assim, os passos do algoritmo para o conjunto de preferências ocorre nos passos enunciados à seguir:

Passo 1. O aluno i_1 propõe para a escola s_3 , o aluno i_2 propõe para a escola s_2 e os alunos i_3 e i_4 propõem para a primeira vaga da escola s_1 . Como a primeira vaga da escola s_1 teve duas ofertas, a vaga é garantida para o aluno i_3 que tem maior prioridade e rejeita-se o aluno i_4 .

Passo 2. O aluno i_4 que foi o único rejeitado no passo anterior propõe para a segunda vaga da escola s_1 , e como não havia nenhum aluno a ocupando, a escola aceita o último aluno proponente e o algoritmo termina. As demais escolas matriculam as ofertas recebidas até então.

Agrupando as vagas de uma mesma escola (colapsando-as), o *matching* final do algoritmo com alunos propondo pode ser descrito na matriz $\mathbf{M}$:

$$\mathbf{M} = \begin{matrix} & s_1 & s_2 & s_3 \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{matrix} & \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{matrix}$$

No caso da matriz final do algoritmo *many-to-one* com as escolas representadas nas colunas, a soma das colunas pode ser maior que um, desde que aquela escola possua número de vagas $q > 1$. Como a escola s_1 possui duas vagas, a soma da coluna de s_1 é igual a dois ($q_1 = 2$).

O exemplo 6 acima é interessante, pois Roth (1985) propôs considerar a seguinte alocação alternativa $\mathbf{M}'$:

$$\mathbf{M}' = \begin{matrix} & s_1 & s_2 & s_3 \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{matrix}$$

Na alocação dada por $\mathbf{M}'$ todas as escolas melhoram pois ganham um aluno mais preferido, ou seja, a nova alocação é uma melhoria de Pareto do ponto de vista das escolas (embora não o seja do ponto de vista dos alunos). Esse exemplo é importante pois ele mostra que no caso *many-to-one*, o teorema **T.4** não pode ser generalizado para as escolas, ou seja, no caso *many-to-one* é possível sim melhorar estritamente todas as escolas.

Contudo, o **T.4** continua válido para o caso dos alunos, e esse é um dos principais motivos para o qual é interessante um algoritmo *deferred-acceptance* com alunos propondo primeiro, pois garante-se que o resultado final é, pelo menos, fracamente ótimo de Pareto para os alunos.

Fernandes (2007, p.18) também observa em relação ao exemplo acima que $\mathbf{M}'$ não é estável, pois o aluno i_1 e a escola s_1 formam um par bloqueante, ou uma “inveja justificada” tal como sugerem Abdulkadiroğlu & Sönmez (2003). Isso porque i_1 prefere s_1 e s_1 , por sua vez, confere maior

prioridade à i_1 do que qualquer aluno dado por $\mathbf{M}'$, que são i_2 e i_4 . O aluno i_1 é preferido à qualquer um dos dois.

Os resultados encontrados pela literatura sugerem um *trade-off* entre a otimalidade de Pareto para as escolas e a estabilidade do algoritmo, esse ponto é enfatizado por Abdulkadiroğlu e Sonmez (2003) e será novamente tratado após apresentação do algoritmo de Top Trading Cycles.

3.3.1. Manipulação do resultado por parte dos alunos e escolas

É importante investigar a possibilidade de alunos e escolas falsearem sua declaração de preferências para conseguir uma alocação mais favorável. Tal como *Dulce* do exemplo 5 da seção 3.2.1, que declara como aceitável apenas o *João*, manipulando o algoritmo para chegar no estável ótimo das mulheres, escolas e alunos também podem falsear as preferências. No caso de escolas, as escolas podem truncar sua lista de prioridades de alunos, ou ainda como se descobriu posteriormente pela literatura (Abdulkadiroğlu et al., 2005), manipular a informação prestada sobre a quantidade de vagas.

A estratégia de não declarar as verdadeiras preferências se torna ainda mais difícil de ser implementada em mercados com grande número de competidores e em situações em que não se sabe quantos estão jogando e há informações e preferências incompletas por parte dos jogadores. Repetir a definição **D.5** retirada da literatura de Teoria dos Jogos é oportuna para entender os teoremas advindos desses estudos:

D.5. Uma *estratégia* é *dominante* para um agente i ocorre quando $\mathcal{J}^*(i)$ é a melhor resposta para todos os possíveis conjuntos de escolhas estratégicas de $\mathcal{J}_{-i}$ tomada por todos os demais agentes.

Em que $\mathcal{J}(i)$ é o conjunto de todos os sinais de preferências possíveis de serem declarados por um jogador i qualquer. Uma dessas declarações é a sua verdadeira preferência, mas todas as demais são declarações falsas que podem o beneficiar ou não. Essas são as ações de um jogo em uma

definição formal. $\mathcal{F}_{-i}$ é o conjunto de todas as ações dos outros jogadores, menos o jogador i . Um terceiro e importante teorema, devido à Roth (1984b), estabelece que:

T.5. *Não existe mecanismo de matching estável tal que declarar as verdadeiras preferências seja a estratégia dominante para todos os agentes.*

Ou seja, sempre há ganho possível para um agente quando todos os demais estão falando a verdade. Outro importante teorema da área, proposto por Dubins & Freedman (1981), afirma que o lado que está propondo o casamento (digamos os alunos) não possui incentivos em não declarar suas verdadeiras preferências. Sendo assim, como os alunos possuem a vantagem em propor, eles não teriam incentivos em falsear suas preferências.

T.6. *O mecanismo que leva ao matching estável H-ótimo (em termos das preferências declaradas) faz com que seja uma estratégia dominante para cada aluno dizer suas verdadeiras preferências.*

Partindo do resultado de Dubins & Freedman (1981), Gale & Sotomayor (1985a) mostram que, na validade de **T.6**, compensaria para pelo menos um agente do lado que recebe as propostas declarar um valor não verdadeiro para suas preferências.

T.7. *Se há mais de um matching estável, então há pelo menos uma escola que estará melhor falsificando suas preferências, assumindo que todos os outros agentes dizem a verdade.*

Os teoremas **T.5**, **T.6** e **T.7** são de grande importância por mostrar quais estratégias os agentes podem assumir diante do mecanismo de *matching*. A ideia básica por trás destes teoremas é de que quem propõe o casamento não possui incentivo em falsear suas preferências, e o lado em “desvantagem” (aquele que está sendo proposto) adquire proveito em tentar fazer o algoritmo se aproximar do resultado estável que lhe é mais favorável, tal como *Dulce* fez no exemplo 5.

Mesmo assim, grande parte da literatura mostra que os ganhos dos agentes com os desvios à tática de revelar as verdadeiras preferências são pequenos. Em particular, para simulações do

algoritmo de alocação de estudantes de Singapura, Teo, Sethuraman, & Tan (2001) mostram que as oportunidades de ganhos com a declaração de falsas preferências (preferências em que era permitido alterar a ordem, mas não truncar a lista de escolas) são de apenas 10%. Mostram também que quanto mais escolas houver, menor será o ganho possível com a estratégia de *cheating*. Esse é um resultado importante, pois se aplica aos trabalhos de *matching* em educação.

No que diz respeito às escolas públicas, onde não há processo de barganha que possa estabelecer que um aluno deva estudar em uma escola ‘A’ ou ‘B’, é importante mostrar que as abordagens das preferências se adéqua bem ao caso. Isso porque não há preços envolvidos no mercado.

Como visto nas hipóteses subjacentes ao sistema de cadastro escolar de Belo Horizonte, o pareamento escolar atual leva em conta um sistema de preferências baseado somente na menor distância. Há uma determinação de preferências de pais e alunos com base na distância. O objetivo da tese é relaxar essa hipótese mais forte de preferências e ver quais são os possíveis resultados de outros tipos de alocações plausíveis.

A próxima seção apresenta o algoritmo de *Top Trading Cycles* (TTC) que apresenta *insights* interessantes para a teoria dos algoritmos de pareamento e será utilizado em uma das simulações do capítulo 7.

3.4. O Algoritmo de *Top Trading Cycles*

O algoritmo de *Top Trading Cycles* (TTC) foi originalmente proposto por Shapley & Scarf (1974) no artigo “*On cores and Indivisibility*”. Os autores creditam contribuição não publicada do professor David Gale para a proposição do algoritmo.

O artigo original propunha um exemplo do mercado imobiliário em que o algoritmo TTC permitiria mostrar a existência do núcleo nesse mercado. Em estudos posteriores Abdulkadiroğlu & Sönmez (1999) generalizaram o algoritmo para situações teóricas do mercado imobiliário em que alguns agentes não possuem dotações iniciais. Pápai (2000 e 2002) aplica regras que blindam

o TTC à manipulação das preferências, fazendo uma relação com os algoritmos de pareamento estáveis.

Abdulkadiroğlu & Sönmez (2003) propuseram a aplicação do TTC para escolas. Os autores levantam esse ponto no artigo em que discutem o sistema de Boston, apresentado em 2.2.1. O TTC possui algumas similaridades com o *deferred-acceptance*, mas também algumas diferenças essenciais.

Em comum, há o ponto de que é necessário obter uma relação racional, completa e estrita das preferências dos alunos para as vagas nas escolas. As escolas também possuem preferências racionais, completas, estritas e transitivas sobre os alunos. Outro ponto em comum é que os alunos “propõem” primeiro à escola mais preferida. No entanto, nessa literatura se diz que os alunos “apontam” para a escola mais preferida, por motivos que são esclarecidos logo abaixo.

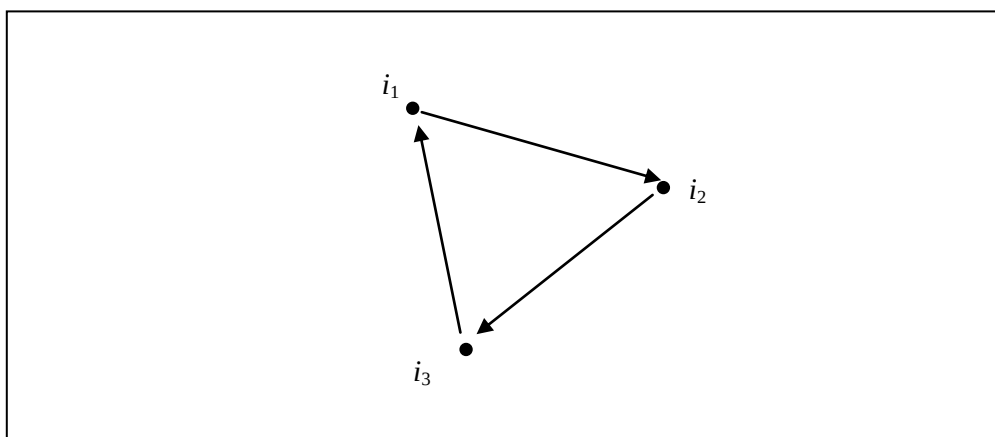
O que é diferente do algoritmo DA, é que no TTC a forma de alocação é distinta. O TTC é mecanismos de trocas puras (de permutas ou escambo), em que um agente para conseguir um bem que não possui precisa requerer que outro agente esteja com vontade de se desfazer do bem de interesse.

A título de ilustração, pode se pensar em um aluno i_1 que já possui uma vaga garantida em uma escola s_1 . Mas esse aluno pode preferir estudar em uma outra escola s_2 que está alocada para o aluno i_2 . Se esses forem os dois únicos alunos e as duas únicas escolas do sistema, para i_1 conseguir uma vaga em s_2 , seria preciso que i_2 estivesse disposto a também trocar de escola.

Suponha que i_2 não estivesse interessado na vaga de s_1 ocupada por i_1 , mas sim de uma outra escola s_3 , que está ocupada pelo aluno i_3 . Dessa maneira i_1 pode conseguir sua vaga em s_2 , basta para isso que i_3 , deseje estar na escola s_1 . Podemos representar graficamente essa situação na FIG. 1:

Na figura abaixo cada aluno é detentor de uma vaga, a seta indica para onde o aluno “aponta”, ou seja, para onde ele deseja ir. Pelo algoritmo de TTC, a troca só é possível quando há um ciclo. Isso quer dizer que existe uma combinação de permutações possíveis que permite a troca.

Figura 1 - Exemplo simples de puras Trocas do TTC (um ciclo simples)



Fonte: Exemplo de elaboração própria.

Em verdade, é preciso se considerar as preferências das escolas. Fernandes (2007, p. 32) ressalta que “intuitivamente, o ponto de partida para o emprego do mecanismo no problema escolar é considerar a prioridade de um aluno em uma escola em relação aos demais como o direito de frequentá-la”.

Os passos do algoritmo de *Top Trading Cycles*, tais como apresentados por Abdulkadiroğlu & Sönmez (2003) são os seguintes:

A.4. O algoritmo *Top Trading Cycles* (TTC):

Passo 1. Cada escola possui um contador do número de vagas. Esse contador registra quantas vagas restam ainda para serem ocupadas. No primeiro passo, cada contador é igual ao total de vagas das escolas. Cada estudante *aponta* para sua escola favorita sob as preferências que foram anunciadas. Cada escola *aponta* para o estudante que possui maior prioridade para aquela escola (melhor classificado no ranque de preferências). Como o número de alunos e escolas é finito, irá ocorrer pelo menos 1 *ciclo*. Além disso, cada

escola pode fazer parte de apenas um ciclo, o que vale também para os estudantes. Cada estudante participante de um ciclo é removido do sistema e o contador das escolas participantes reduz em uma unidade, indicando que uma vaga foi preenchida. Se o contador zerar, a escola também é retirada do sistema.

Passo k. Cada estudante que ainda está no sistema aponta sua escola mais preferida entre as escolas remanescentes (as que ainda possuem vagas, contador maior que zero). E as escolas ainda presentes no sistema apontam para o aluno mais preferido entre os que ainda estão presentes. Ocorrerá pelo menos um ciclo. Cada estudante participante de um ciclo ganha um assento na escola para a qual apontou e é retirado do sistema. O contador de cada escola participante do sistema é reduzido em uma unidade, se o contador zerar a escola também deixa o sistema. Os alunos que não participaram de um ciclo e as escolas que possuem vagas continuam no sistema. O algoritmo termina quando não há mais alunos no sistema.

Apontar para o algoritmo de TTC é similar à *propor* no algoritmo Gale-Shapley. O algoritmo TTC se encerra quando todos os estudantes ganham uma matrícula. Um *ciclo* é uma lista ordenada que intercala alunos e escolas. Um aluno presente no sistema *aponta* para uma escola que *aponta* para um aluno. Sempre que um elemento se repete nessa corrente de invocação, se fecha um ciclo e o algoritmo finaliza um passo, retirando os participantes daquele ciclo do mercado. O algoritmo continua com os alunos e escolas remanescentes.

Um exemplo de ciclo: $i_1 \rightarrow s_1 \rightarrow i_3 \rightarrow s_2 \rightarrow i_2 \rightarrow s_1$

No anexo B, há mais detalhes da implementação do algoritmo. O algoritmo não foi implementado como descrito em A.4, pois para códigos em linguagem R seria difícil estipular contadores de vagas para cada escola funcionando de maneira independente. A solução adotada foi similar a do algoritmo *deferred acceptance*: adotar cada vaga de uma mesma escola como um jogador independente e ao final reagrupar as vagas na matriz. Fazendo isso é preciso respeitar o ordenamento de preferências entre as escolas, uma vaga de uma escola mais preferida é sempre melhor do que o de uma vaga de uma escola menos preferida. Os exemplos a seguir ajudam a compreender o funcionamento do algoritmo na prática.

Exemplo. 7: Suponha-se quatro escolas: s_1, s_2, s_3, s_4 e quatro alunos: i_1, i_2, i_3, i_4 . Cada escola possui apenas uma vaga e as preferências de alunos e escolas estão listadas a seguir:

Alunos:

$$\mathcal{P}(i_1) = s_1, s_2, s_3, s_4$$

$$\mathcal{P}(i_2) = s_1, s_2, s_3, s_4$$

$$\mathcal{P}(i_3) = s_4, s_1, s_3, s_2$$

$$\mathcal{P}(i_4) = s_4, s_3, s_1, s_2$$

Escolas:

$$\mathcal{P}(s_1) = i_3, i_2, i_1, i_4$$

$$\mathcal{P}(s_2) = i_2, i_1, i_3, i_4$$

$$\mathcal{P}(s_3) = i_1, i_2, i_4, i_2$$

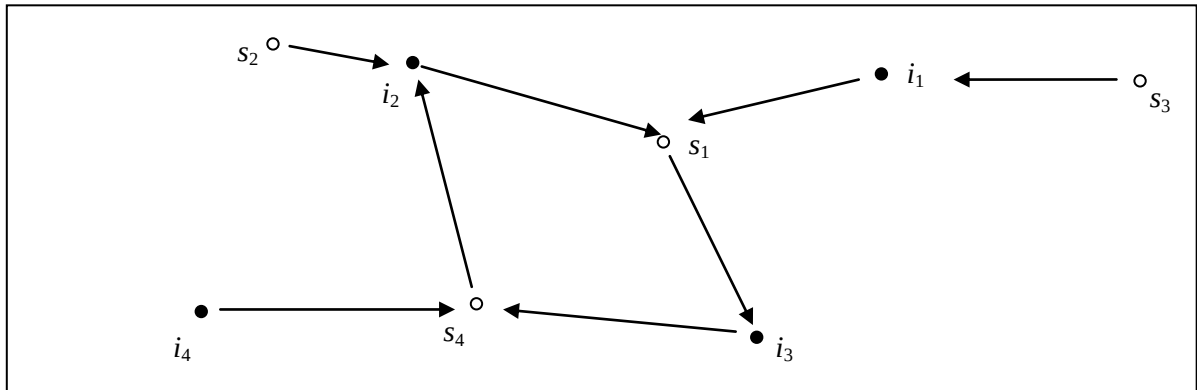
$$\mathcal{P}(s_4) = i_2, i_3, i_4, i_1$$

$$Q = (q_1 = 1, q_2 = 1, q_3 = 1, q_4 = 1)$$

Serão apresentados os três passos do algoritmo de solução para as preferências acima listadas acompanhados de uma figura ilustrativa (FIGs. 2-4).

Passo 1. Os alunos i_1 e i_2 *apontam* para a escola s_1 , a escola s_1 *aponta* para o aluno i_3 , os alunos i_3 e i_4 *apontam* para s_4 e a escola s_4 *aponta* para o aluno i_2 . Fecha-se um ciclo: $(i_2, s_1, i_3, s_4, i_2)$. Os alunos e escolas participantes do ciclo deixam o sistema, e nova rodada será proposta pelos alunos remanescentes.

Figura 2 - Exemplo 7 de TTC com 4 alunos e 4 escolas (1º passo)

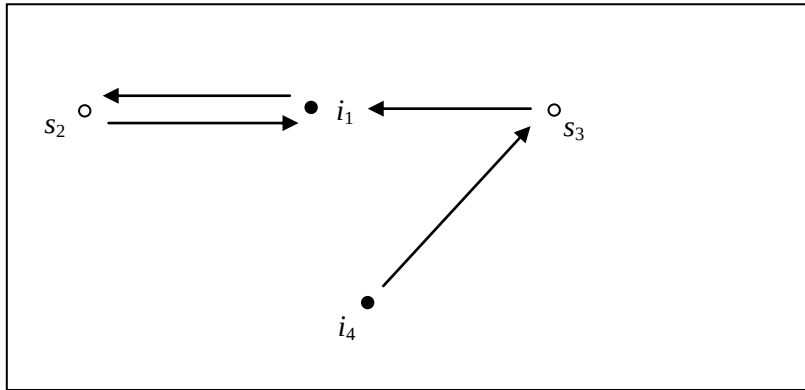


Fonte: Exemplo de elaboração própria.

Passo 2. Os alunos e escolas que não participaram do primeiro ciclo: (i_1, i_4) e (s_2, s_3) *apontam* para as opções mais preferidas entre os remanescentes. O aluno i_1 *aponta* para s_2 , s_2 *aponta* para i_1 e fecha-se mais um ciclo: (i_1, s_2, i_1) . Esse é o tipo de ciclo mais curto

que pode ocorrer pois os agentes se preferem mutuamente. O aluno i_4 *aponta* para s_3 e s_3 continua *apontando* para i_1 .

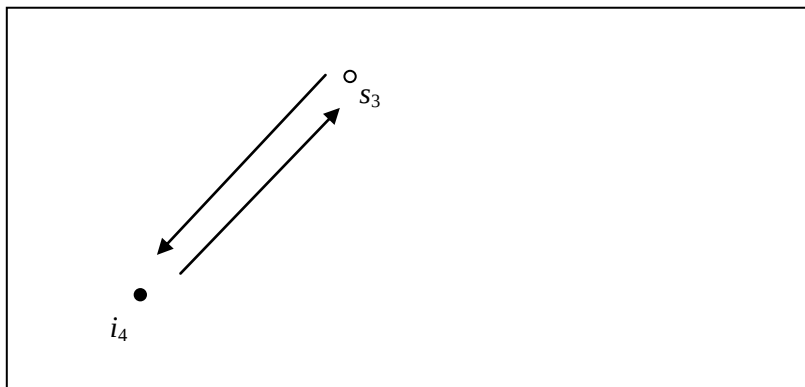
Figura 3 - Exemplo 7 de TTC com 4 alunos e 4 escolas (2º passo)



Fonte: Exemplo de elaboração própria.

Passo 3. O aluno i_4 e a escola s_3 , únicos remanescentes no sistema, se *apontam* mutuamente e o último ciclo se encerra.

Figura 4 - Exemplo 7 de TTC com 4 alunos e 4 escolas (3º passo)



Fonte: Exemplo de elaboração própria.

Exemplo. 8: Outro exemplo pode ser tomado de Abdulkadiroğlu & Sönmez (2003, pp.742-744). Suponha-se que haja 8 alunos $\{i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8\}$ e quatro escolas $\{s_1, s_2, s_3, s_4\}$, sendo que as duas primeiras escolas possuem duas vagas cada uma e as duas últimas possuem 3 vagas cada uma. Esse sistema possui vagas ociosas, 10 vagas para 8 alunos. As preferências são dadas pelas relações abaixo:

Alunos:

$$\mathcal{P}(i_1) = s_2, s_1, s_3, s_4$$

$$\mathcal{P}(i_2) = s_1, s_2, s_3, s_4$$

$$\mathcal{P}(i_3) = s_3, s_2, s_1, s_4$$

$$\mathcal{P}(i_4) = s_3, s_4, s_1, s_2$$

$$\mathcal{P}(i_5) = s_1, s_3, s_4, s_2$$

$$\mathcal{P}(i_6) = s_4, s_1, s_2, s_3$$

$$\mathcal{P}(i_7) = s_1, s_2, s_3, s_4$$

$$\mathcal{P}(i_8) = s_1, s_2, s_4, s_3$$

Escolas:

$$\mathcal{P}(s_1) = i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8$$

$$\mathcal{P}(s_2) = i_3, i_5, i_4, i_8, i_7, i_2, i_1, i_6$$

$$\mathcal{P}(s_3) = i_5, i_3, i_1, i_7, i_2, i_8, i_6, i_4$$

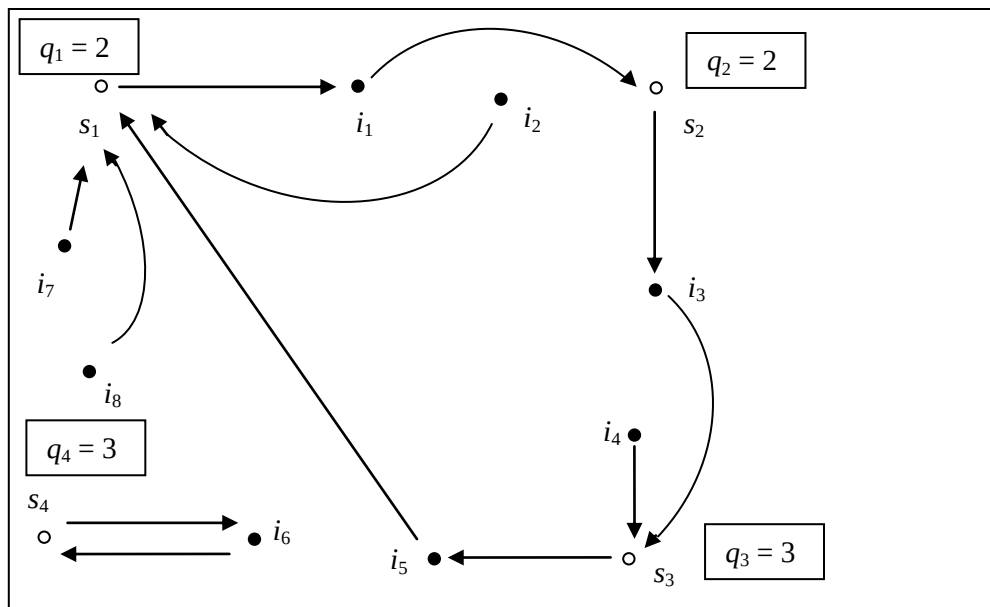
$$\mathcal{P}(s_4) = i_6, i_8, i_7, i_4, i_2, i_3, i_5, i_1$$

$$Q = (q_1 = 2, q_2 = 2, q_3 = 3, q_4 = 3)$$

O exemplo acima se encerra em 4 passos. A FIG. 5 abaixo apresenta o primeiro passo do algoritmo que conta com dois ciclos simultâneos. Os demais passos são representados nas FIGs. 6-8.

Passo 1. No primeiro passo do exemplo, os alunos i_1, i_3, i_5 formam um ciclo com as escolas s_1, s_2, s_3 . Ciclo: $\{i_1, s_2, i_3, s_3, i_5, s_1, i_1\}$. Há outro ciclo de pequeno circuito entre o aluno i_6 e a escola s_4 , $\{i_6, s_4, i_6\}$. Esses alunos são retirados do sistema e todas as quatro escolas tem seu número de vagas diminuído em uma unidade no contador. O próximo passo continua com os alunos que ainda não conseguiram vaga, isto é, os alunos $\{i_2, i_4, i_7$ e $i_8\}$.

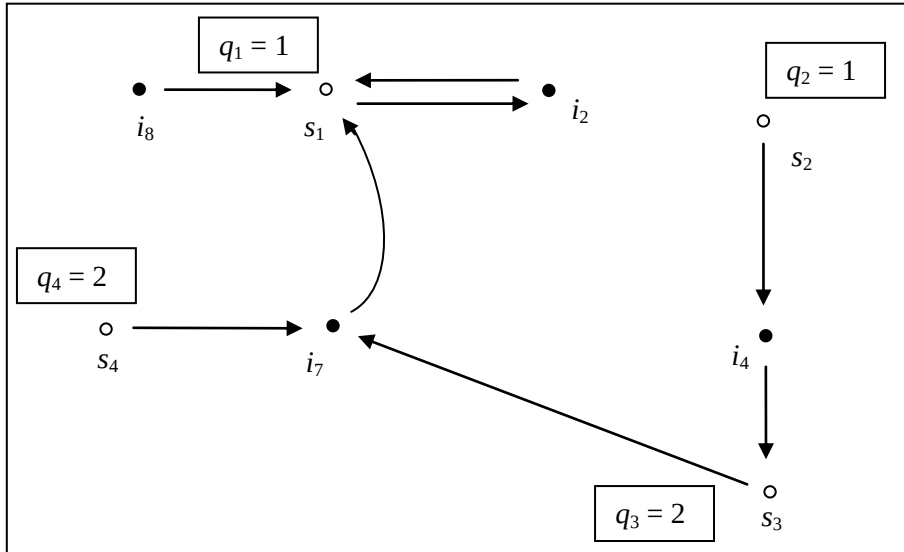
Figura 5 - Exemplo 8 de TTC com 8 alunos e 4 escolas (1º passo)



Fonte: Exemplo adaptado de Abdulkadiroğlu & Sonmez (2003).

Passo 2. No segundo passo apenas um ciclo se fecha, entre o aluno i_2 e a escola s_1 . Nota-se que, como ainda há vagas, todos os alunos continuam apontando para a escola que apontavam já no primeiro passo. As escolas apontam para novos alunos e, ao final, a escola s_1 fica sem mais nenhuma vaga e deixa o sistema.

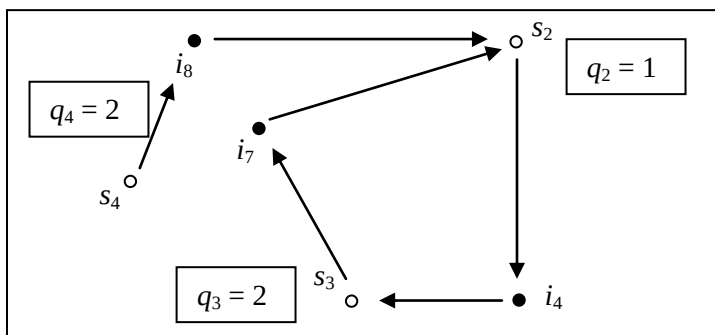
Figura 6 - Exemplo 8 de TTC com 8 alunos e 4 escolas (2º passo)



Fonte: Exemplo adaptado de Abdulkadiroğlu & Sonmez (2003).

Passo 3. Mais um ciclo se fecha, entre os alunos i_4 e i_7 e as escolas s_2 e s_3 . Ao final, a escola s_2 fica sem mais nenhuma vaga e deixa o sistema, os alunos i_4 e i_7 também saem. A escola s_3 tem o seu contador diminuído em uma unidade, restando-lhe mais uma vaga. A escola s_4 permanece no sistema ainda com duas vagas e o aluno i_8 é único ainda sem matrícula.

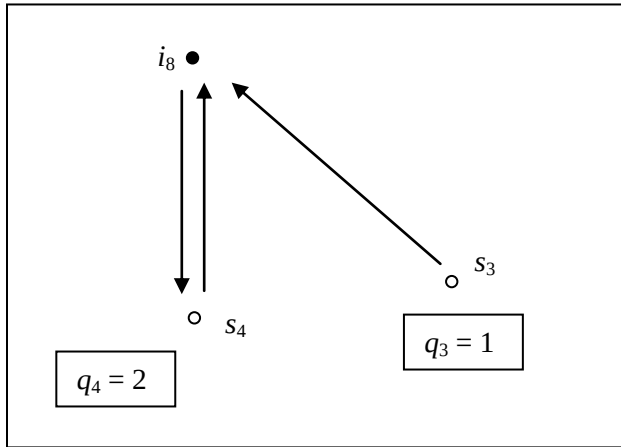
Figura 7 - Exemplo 8 de TTC com 4 alunos e 4 escolas (3º passo)



Fonte: Exemplo adaptado de Abdulkadiroğlu & Sonmez (2003).

Passo 4. Ao final do 4º passo se forma o último ciclo e o aluno i_8 consegue matricular-se na escola s_4 que tinha ainda 2 vagas disponíveis. O algoritmo termina e as escolas s_3 e s_4 ficam com uma vaga sobrando cada uma.

Figura 8 - Exemplo 8 de TTC com 4 alunos e 4 escolas (4º passo)



Fonte: Exemplo adaptado de Abdulkadiroğlu & Sonmez (2003).

O *matching* final gerado pelo mecanismo pode ser apresentado pela matriz **T** abaixo:⁶²

$$\mathbf{T} = \begin{matrix} & s_1 & s_2 & s_3 & s_4 \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \\ i_6 \\ i_7 \\ i_8 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

Do resultado do mecanismo consegue-se visualizar que o resultado final é Pareto ótimo. Cada aluno conseguiu a escola mais preferida para a qual ainda existiam vagas. Como cada aluno está com sua melhor vaga, não é possível propor alguma mudança que melhore um dos participantes sem piorar nenhum outro. Abdulkadiroğlu & Sonmez (2003, p. 737) propõem o teorema abaixo:

⁶² A programação do algoritmo utilizado se deu de uma forma um pouco diferente em relação ao colocado em **A.4**, empregou-se a expansão das vagas, em que cada vaga é como uma escola. Pelo algoritmo empregado, a solução de exemplo ocorre após 10 passos. Para detalhes, ver Anexo B.

T.8. *O mecanismo de Top Trading Cycles é Pareto eficiente.*

Prova-se por indução: considere o algoritmo TTC, qualquer estudante que deixa o processo no 1º passo foi matriculado na melhor escola de sua preferência e, portanto, não se pode propor outra alocação que o melhore. Cada aluno que deixa o sistema no 2º passo foi matriculado para sua melhor escola entre os assentos que ainda permaneciam vagos e, como as preferências são estritas, não se pode fazer os alunos melhorarem no 2º passo sem que isso prejudique os alunos do 1º passo que já estão na melhor alocação que lhes é possível. Procedendo de maneira similar e por indução, nenhum estudante pode ser melhorado sem que isso prejudique algum outro estudante de um passo anterior. Portanto, prova-se que o mecanismo é Pareto eficiente. ■

Fernandes (2007, p. 37) propõem um teorema que pode ser mostrado com um contra exemplo:

T.9. *O mecanismo de Top Trading Cycles não é um mecanismo estável.*

Prova-se com um contra-exemplo: Suponha novamente o exemplo 8, o aluno i_8 e a escola s_2 formam um par bloqueante à $\mathbf{T}$, alocação proposta do TTC. Pode-se verificar que i_8 prefere a escola s_2 em relação à escola s_4 que ele ganhou no TTC. A escola s_2 , por sua vez, prioriza o aluno i_8 em relação ao aluno i_1 ou i_7 que ela obteve na alocação. Por conta desse par bloqueante, nota-se que a alocação do algoritmo de TTC não é estável. ■

O leitor pode conferir que a alocação sugerida pelo *deferred acceptance* seria a da matriz $\mathbf{M}^*$ abaixo:

$$\mathbf{M}^* = \begin{matrix} & s_1 & s_2 & s_3 & s_4 \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \\ i_6 \\ i_7 \\ i_8 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{matrix}$$

Ao tirar a diferença das matrizes $\mathbf{T}$ e $\mathbf{M}^*$, elevar cada um dos termos ao quadrado $(\mathbf{M}^* - \mathbf{T})^2$, e somar o resultado em cada uma das linhas dividindo-o por 2, obtêm-se quantos alunos possuem posições diferentes entre uma alocação e outra:

$$D = \frac{\sum_i^n (\mathbf{M}^* - \mathbf{T})^2}{2}, \quad i \in I \quad (1)$$

Em que i representa a linha para cada aluno do conjunto I e n é número total de alunos no grupo. Para o exemplo acima o resultado de D é igual a 4. Um dos principais pontos dessa diferença é que D ressalta as diferenças entre as alocações com uma medida de quantos alunos trocam de lugar.

Outro aspecto das teorias de *matchings* é que existe um *trade-off* importante entre essas escolhas de estabilidade e eficiência. Com o TTC se consegue uma alocação Pareto ótima tanto para alunos quanto para escolas. Porém, essa escolha não é estável.

No caso do algoritmo TTC em relação ao *deferred acceptance* (DA), no exemplo 8, dos quatro alunos que ganharam alocações diferentes, 3 deles estão melhores sob o TTC, e apenas um deles melhor sob o DA. A estabilidade pode ser um preço alto a pagar em casos assim. No entanto, por outro lado, em algumas situações os alunos que ficam com uma alocação de qualidade inferior na escala de preferências podem ser muito prejudicados. Se esse grupo for numeroso e o prejuízo que eles obtêm no TTC for muito grande, a estabilidade do DA pode ser uma grande vantagem.

Abdulkadiroğlu & Sönmez (2003) denominam os pares bloqueantes como *inveja justificada*. O termo é interessante para o contexto escolar. Sob o conceito de Pareto, não há como propor melhorias para o algoritmo TTC, pois para um aluno ganhar algum outro precisa perder sua posição mais vantajosa (como mostra a prova de **T.8**). Porém, como em várias situações interessantes do estudo de alocações, em alguns casos a eficiência se consegue a um preço de desigualdade muito alto, o que é socialmente não desejado.⁶³

⁶³ Percebe-se pelo argumento do termo que algumas invejas são “*justificadas*” e outras não. O aluno pode ter inveja do aluno que conseguiu uma escola mais desejada por ele, mas se ele não conseguisse uma vaga naquela escola de nenhuma maneira, essa inveja não tem justificativa (injustificada). Noutro caso, o aluno pode preferir mais uma

Se o número de alunos que podem ganhar uma melhor escola for muito grande, a instabilidade de um sistema de alocação, até mesmo o sistema de cadastro de Belo Horizonte pode ficar comprometida. Haveria muitos alunos e pais insatisfeitos o que poderia levar acordos bilaterais entre as partes.⁶⁴ A TAB. 6 apresenta a ideia vinculada a esse *trade-off*, o algoritmo DA leva a um resultado necessariamente estável (sem *pares bloqueantes*), mas só há garantia de que será fracamente Pareto ótimo para o lado que propõe, e não necessariamente leva ao Pareto ótimo das escolas. O algoritmo TTC não garante a estabilidade, como visto pelo contra exemplo (T.9), mas garante um resultado de Pareto ótimo para os dois lados (T.8).

Tabela 6 - Comparando dois algoritmos DA e TTC de acordo com Estável e Pareto

<i>Algoritmo</i>	<i>Estável</i>	<i>Pareto</i>
<i>DA</i>	Sim	Não necessariamente
<i>TTC</i>	Não necessariamente	Sim

Fonte: Elaboração do autor.

Um bom sistema de alocação de alunos precisa conciliar bem esse *trade-off*. O sistema será tanto melhor quanto mais os agentes se sentirem participantes e engajados no processo de escolha. A experiência de BH mostra também que é importante ter uma “válvula de escape” para lidar com as questões particulares de um ou outro caso de insatisfação e necessidade de reencaminhamento.

Um último ponto discutido pela literatura é a questão sobre se as escolas podem ter preferências por determinados alunos. Em se tratando dos direitos constitucionais, todos os alunos requerentes têm o direito a uma vaga na educação básica, não é possível favorecer alunos em detrimento de outros. No entanto, sob os termos da administração pública, é preciso estabelecer prioridades e regras para o registro, tendo em vista recursos de alocação escassos e pressões da demanda.

escola e a escola em contrapartida também o prefere mais do que algum aluno que já lhe esteja alocado, e aí se configura a inveja justificada, pois o aluno teria condições de preencher a vaga.

⁶⁴ No caso de Belo Horizonte, isso ocorre quando pais tentam falsear o endereço de casa. No entanto, como argumentado em 2.1, o sistema é bastante estável, pois essas ocorrências são pontuais. Uma situação instável seria aquela em que esse comportamento se generalizasse e as escolas passassem fazer “vista grossa” para as regras do sistema.

Enfatizar que as escolas priorizam os alunos segundo regras administrativas e pedagógicas é uma maneira importante de se trazer essa questão para um debate amplo.

O capítulo 7 empregou os algoritmos de TTC para simular um processo de alocação por meio de filas que ocorria em Belo Horizonte antes do sistema de cadastro ser implementado. Assumindo-se que os pais são racionais, pais e alunos se dirigiam para a escola mais preferida. No entanto, como não havia critério de entrada, as filas eram a maneira que as escolas obtinham de matricular os alunos. Os primeiros a chegar recebiam as vagas. Na simulação supôs-se que as filas seriam dadas de maneira aleatória, e que os pais escolhem as escolas segundo uma regra lexicográfica de preferências que considera desempenho médio no PROEB e as distâncias.

A seção a seguir apresenta a interpretação e os estudos de uma outra corrente, creditada à Becker (1973 e 1974). Como já evidenciado no capítulo 2, existe um ponto de interseção entre esses dois arcabouços teóricos: um sistema de preferências não dispensa totalmente a existência de utilidades, e um sistema de utilidades tem necessariamente embutido um sistema de preferências.

3.5. Modelos de casamentos com base em transferência de utilidade (Becker)

A metodologia desenvolvida por Becker (1991) e demais teóricos do Capital Social elabora modelos de casamento baseados nas utilidades individuais dos agentes. O casamento confere ao casal determinado produto. Essa teoria ajuda na interpretação de como se dá a formação de pares. Uma aplicação muito útil dessa literatura procura testar empiricamente as trocas intra-familiares e qual a participação do altruísmo nessas transferências.

Os modelos de casamento por utilidades ocorrem sob uma perspectiva de escolha individual e livre dos agentes. Os indivíduos que formam um par consideram o *Capital Social* um do outro e realizam determinada escolha baseados na expectativa de ganhos do casamento.

Os casamentos possuem um produto agregado Y que pode ser medido pela utilidade dos pares formados. Para denotar os *matchings* continuaremos usando a denotação μ da seção anterior. Temos então:⁶⁵

$$Y = U(\mu) \quad (2)$$

Para essa teoria há igualmente dois conjuntos de homens (H) e mulheres (W) que formarão casamentos monogâmicos entre si. Podemos imaginar um casamento μ que forme o par h e w (homem h e mulher w). Relembrando a notação da seção anterior: $\mu(h) = w$ (o parceiro de h é a mulher w na alocação μ), de contraparte temos: $\mu(w) = h$ (o parceiro de w é o homem h).⁶⁶ Dessa forma, é possível reescrever o produto de um casamento como soma de utilidades dos agentes:

$$Y_{hw} = U_h(\mu(h)) + U_w(\mu(w)) \quad (3)$$

O produto do casamento entre o homem h e a mulher w (denotado por Y_{hw}) é a soma da utilidade que o homem h obtém do casamento com w e a soma da utilidade que a mulher w obtém com seu par h . Pode-se simplificar a notação para:

$$Y_{hw} = Y_h + Y_w \quad (4)$$

A utilidade total obtida de todos os casamentos existentes é obtida pelo somatório:

$$Y = \sum_{h \in H, w \in W}^n Y_{hw} \quad (5)$$

Alternativamente, tomando $\mathcal{M}^*$ como todos os conjuntos de todos os casamentos realizados, $\mathcal{M}^* = \{\mu_1, \mu_2, \mu_3, \dots, \mu_n\}$, em que k é o índice de cada par casado:⁶⁷

⁶⁵ Essa seção se baseia principalmente no capítulo 4 de Becker (1991) e no capítulo de mesma numeração de Becker & Murphy (2000).

⁶⁶ Diferentemente da teoria de *two-sided markets*, a alocação de *matching* é única para o caso da teoria de Becker (1991), não sendo necessário distinguir o *matching* de homens e mulheres.

⁶⁷ Repare que são todos os casamentos realizados e não todos os casamentos possíveis.

$$Y = \sum_{k=1}^n U(\mu_k) \quad (6)$$

Para Becker & Murphy (2000), a utilidade de um casamento deve depender não apenas do par formado, ou seja, não apenas com quem se casa ($\mu(h) = w$), mas também da renda proveniente do casamento com o par com quem se casa. Em verdade, essa renda pode ser vista de maneira mais ampla e não apenas como ganhos monetários. Ao abordar aqui o casamento formado por μ , pode-se considerar que por trás do par formado há uma renda que ajuda os indivíduos a derivarem utilidade.

Semelhante ao teorema **T.1**, estipulado por Gale e Shapley (1962), que afirma que existe um casamento estável para todo e qualquer mercado de casamentos, também há em Becker (1991, p. 111) uma condição semelhante que garante casamentos estáveis. A condição postula que tomado um casamento qualquer entre h e x que não seja ótimo, sempre haverá ganho maior na soma desses agentes em outros casamentos que não aquele de (h, x) :

$$Y'_h + Y'_x \geq Y_{hx} \quad (7)$$

Em que Y'_h e Y'_x são os benefícios de h e x advindos de casamentos com outros pares. Suponha que h tenha se casado com a mulher x , e que w tenha se casado com o homem k , tal que $h \neq k$ e $w \neq x$. Em paralelo com a literatura do *matching*, a condição (7) acima equivale a dizer que o casamento de (h, x) foi bloqueado pelo par (h, w) se $U(\mu(h) = w) > U(\mu(h) = x)$ e $U(\mu(w) = h) > U(\mu(w) = k)$. Em equilíbrio, tem-se a condição (7) satisfeita. Assumindo que (h, w) e (k, x) são pares ótimos, reescreve-se as condições como igualdade estritas:

$$Y_h + Y_w = Y_{hw}^* \quad (8)$$

$$Y_k + Y_x = Y_{kx}^* \quad (9)$$

Em que X_h , Y_w e Y_k e Y_x são benefícios advindos dos pares ótimos. Portanto, na versão de Becker (1991), o teorema **T.1** pode ser reescrito da seguinte forma:

T.9. Existe um casamento ótimo para todo e qualquer mercado de casamentos.

A noção de casamento ótimo é mais importante em Becker do que a noção de casamento estável. Se pensarmos que o casamento ótimo é aquele melhor possível, temos que o conceito possui correspondência com o de casamento estável. O próprio Becker (1991, p. 127 n.23) comenta sobre essa semelhança, reproduz-se novamente o trecho:

Gale and Shapley (1962) require optimal assignments to be “stable”; that is, persons not assigned to each other could not be made better off by marrying each other, a requirement that is closely related to condition (4.5) [correspondente à equação (7) deste texto]. (Becker, 1991, p. 127n.23).

Pode-se usar um exemplo de casamentos ótimos para mostrar como funcionam os procedimentos de *matching* segundo essa abordagem. Suponha um conjunto de três homens, $H = \{h, k, m\}$, e três mulheres, $W = \{w, x, z\}$. Os ganhos conjuntos desses casamentos estão representados em uma matriz que mostra todos os pares possíveis, os casamentos ótimos estão nas diagonais e são indicados por um asterisco.⁶⁸ Tem-se a seguinte matriz:

$$\begin{array}{c} \\ \\ \\ \end{array} \begin{array}{ccc} w & x & z \\ \left[\begin{array}{ccc} Y_{hw}^* & Y_{hx} & Y_{hz} \\ Y_{kw} & Y_{kx}^* & Y_{kz} \\ Y_{mw} & Y_{mx} & Y_{mz}^* \end{array} \right] \end{array}$$

Exemplo 9: Suponha que os ganhos máximos individuais que cada agente desse mercado consegue são os seguintes:

Mulheres: $Y'_w = 5$; $Y'_x = 2$; $Y'_z = 1$

Homens: $Y'_h = 3$; $Y'_k = 5$; $Y'_m = 2$

⁶⁸ No modelo original de Becker (1991, cap. 4) os pares ótimos são dados por indivíduos representados com índice de mesma letra, *ii*, *jj*, *ll*, *mm*, e assim por diante. Isso confere a vantagem imediata de sabermos que casamentos de mesmo índice formam um par ótimo, no entanto, essa notação tem a desvantagem de não identificar imediatamente a qual grupo o índice se refere, se dos homens ou das mulheres, ou se um indivíduo está casado com pessoa do outro sexo ou consigo mesmo. Então, para a notação aqui empregada, sempre que o casamento for ótimo, denota-se esta informação por um asterisco junto do Y (Y^*). E empregou-se as aspas simples para denotar o melhor casamento sob a perspectiva individual, ou o produto potencial de um casamento (Y').

Colocando os valores da soma dos casamentos na matriz (homens nas linhas, mulheres nas colunas) tem-se que:

$$\begin{array}{c} w \quad x \quad z \\ h \left[\begin{array}{ccc} 8 & 4 & 2 \end{array} \right] \\ k \left[\begin{array}{ccc} 9 & 7 & 6 \end{array} \right] \\ m \left[\begin{array}{ccc} 5 & 3 & 3 \end{array} \right] \end{array}$$

Na matriz acima, os casamentos ótimos estão na diagonal da matriz (hw , kx , mz). O ganho social agregado dessa alternativa de alocação é $Y = 18$, o maior possível. Nenhuma das outras 5 combinações produziria ganhos maiores. A segunda maior possibilidade seria o *matching* (hw , kz , mx), que possui soma igual a 17.

Existe casamento que individualmente gera produto maior (o par (k,w)), mas não é ótimo, vejamos porque. Os ganhos individuais máximos de cada um dos indivíduos k e w são respectivamente $Y'_k = 5$ e $Y'_w = 5$. Um casamento entre eles, para ser aceito, deveria oferecer benefício mínimo de 10 e não de 9. A restrição da equação (7) está 'folgada' dizendo que o par (kw) é não ótimo (instável):

$$\begin{aligned} Y'_k + Y'_w &\geq Y_{kw} \\ 5 + 5 &\geq 9 \end{aligned}$$

O mesmo o raciocínio se aplica aos demais pares não ótimos. De forma intuitiva, pode-se pensar que ambos os indivíduos só aceitam se casar caso ganhem acima de 5 pontos de utilidade. O casamento de k e w necessariamente os obrigará a dividir um produto de 9 pontos (que é menor do que o produto máximo potencial igual a 10). Caso o produto seja dividido igualmente, os 4,5 que cabem a cada um frustrará as expectativas dos agentes. Se uma das partes fica com 5 e outra com 4, isso manterá apenas um deles atendido, enquanto o outro verá motivos para se desfazer da união.

Ao analisarmos os pares estáveis de (h, w) , (k, x) e (m, z) , vemos que, apesar dos produtos menores do casamento, todas as partes são atendidas e a igualdade de (8) e (9) é satisfeita, e o produto total o maior possível:

$$Y'_h + Y'_w = Y_{hw}$$

$$3 + 5 = 8$$

$$Y'_k + Y'_x = Y_{kx}$$

$$5 + 2 = 7$$

$$Y'_m + Y'_z = Y_{mz}$$

$$2 + 1 = 3$$

O exemplo 9 acima oferece caminhos para se demonstrar **T.9** e destaca a diferença essencial dos modelos de utilidade de Becker dos modelos de *matching* por preferências. A principal diferença está no fato de que os modelos de utilidade consideram a partição do produto no casamento explicitamente, o que está apenas implícito nos modelos de *matching*.

Outro aspecto da teoria dos casamentos de Becker é o papel do *Capital Social* para reduzir custos do *household production* e aumentar os ganhos do casamento. Assim estabelece-se uma relação entre as características dos indivíduos e o produto do casamento. Como colocam Becker & Murphy (2000, p. 29), essas características podem ser elencadas como “renda, educação, idade, saúde, aparência, personalidade, cor, religião, etnia, *background* familiar e outras”. Por simplicidade, as características são resumidas como K para as mulheres e L para os homens.

Também por simplicidade K e L assumem valores no eixo dos valores reais e possuem um *continuum* crescente de características, indo do menor para o maior. É fácil pensar em variáveis contínuas no que diz respeito à renda e idade, no entanto, algumas outras características citadas acima são visivelmente categóricas, tais como educação, cor, religião, e etnia. No entanto, isso não prejudica a análise. Interessa saber se as características estão correlacionadas positiva ou negativamente entre os pares.

As características L e K de homens e mulheres interferem na percepção de como esses grupos derivam suas utilidades. Considerando essas características e as utilidades em todo o conjunto $\mathcal{M}$ de casamentos possíveis (e factíveis), homens e mulheres passam por um mercado que proporciona a escolha em um processo descentralizado. Nesse processo, cada par encontra um casamento ótimo denominado por μ^* e, após o casamento decidido, há a obtenção do produto do casamento (Y).

A função de utilidade individual é dependente dos casamentos e das características dos pares: $U_H(\mu, K)$, $U_W(\mu, L)$. A utilidade conjunta é $U(\mu, K, L)$ e o produto do casamento (Y) depende das

utilidades: $Y(U(\mu, K, L))$.⁶⁹ O importante, no caso, é entender como Y varia de acordo com as características. Becker & Murphy (2000, p. 30) propõem que a primeira derivada de Y em relação à K ou L seja sempre positiva.⁷⁰

$$\frac{\partial Y}{\partial U} \cdot \frac{\partial U}{\partial L} > 0; \frac{\partial Y}{\partial U} \cdot \frac{\partial U}{\partial K} > 0 \quad (10)$$

Tem-se o chamado *positive assortative mating*, quando há o casamento entre iguais e características iguais retornam maior produto. Por exemplo, quanto mais escolarizado é um casal, maior é o benefício do casamento, a segunda derivada é:

$$\frac{\partial^2 Y}{\partial^2 U} \cdot \frac{\partial^2 U}{\partial L \partial K} > 0 \quad (11)$$

Existe também a possibilidade do *negative assortative mating*, quando há casamento entre pessoas com características diferentes. Esse é o casamento em que, por exemplo, uma mulher muito instruída se beneficiaria casando com um homem de pouca instrução. Esse tipo de substituição pode ocorrer por diversos motivos. Pode-se pensar que a mulher teria capacidade de educar bem os filhos, sendo desnecessário que o pai fosse instruído nesse aspecto. E no caso, a mulher pode se dedicar a trabalhos de rendimentos intelectuais e o homem poderia cuidar de trabalhos manuais e afazeres domésticos. Nesse caso obtêm-se:

$$\frac{\partial^2 Y}{\partial^2 U} \cdot \frac{\partial^2 U}{\partial L \partial K} < 0 \quad (12)$$

Um último caso trata de uma situação em que as características não façam nenhuma diferença. Em uma situação em que homens e mulheres são todos iguais no mercado, sem nenhum diferencial, tem-se que:

⁶⁹ Segue-se notação similar à empregada por Becker (1991) e Becker & Murphy (2000), usar Y como uma função e como variável do produto do casamento é um abuso de notação utilizado pelos autores, mas, por outro lado, torna a notação mais econômica.

⁷⁰ Caso se acredite que alguma característica tenha relação negativa, sua polaridade pode ser invertida, e caso se acredite que haja alguma característica que não interfira no produto de um casamento ela pode ser retirada do modelo.

$$\frac{\partial^2 W}{\partial^2 U} \cdot \frac{\partial^2 U}{\partial L \partial K} = 0 \quad (13)$$

Discutindo ainda sobre um resultado importante do modelo de Becker (1991), retorna-se à nota de rodapé 39 para acrescentar um teorema importante, que encontra correspondente na literatura de *matching* (Gale & Sotomayor, 1985b):

T.10. *Uma mulher (ou um homem) acrescida (acrescido) ao mercado faz com que a situação dos homens (ou das mulheres) seja melhor ou tão boa quanto à situação anterior onde havia menos uma mulher (ou homem).*

Em outras palavras, uma mulher acrescida ao mercado não pode piorar a situação dos homens, ou então, no caso contrário, um homem acrescido ao mercado não pode piorar a situação das mulheres. Esse teorema é demonstrado pelas duas literaturas, mas no caso em especial ilustra-se com um exemplo compatível com a literatura de Becker (1991), em que há possibilidade de transferências entre os membros de um casal.

Exemplo. 10: Suponha inicialmente que tenhamos um homem k e uma mulher w que formam um casal do *matching* μ_1 . Digamos que o produto do casamento é $Y_{kw} = 7$, suponha que a repartição do produto nesse casamento seja 5 para o homem e 2 para a mulher.

Suponha agora que haja mais um homem nesse mercado (h) e que a utilidade de permanecer solteiro para os dois homens seja a mesma $Y_{kk} = Y_{hh} = U_k(\mu(k) = k) = U_h(\mu(h) = h) = 1$. A título de exercício, suponha ainda que a união do homem h com a mulher w leve a um casamento com produto de oito, $Y_{hw} = 8$, esse casamento será chamado de μ_2 .

A mulher w tem uma expectativa de ganhos próximos a 6. Caso o casal kw já esteja formado, o homem h , novo entrante, pode oferecer à mulher uma transferência do benefício do casamento (T_h):

Na situação inicial:

$$\mu_1: Y'_k + Y'_w = 5 + 2 = 7 \text{ (situação inicial)}$$

Com a entrada de um novo candidato oferecendo T_h :

$$\mu_2: (Y'_h - T_h) + (Y'_w + T_h) = 6 - T_h + 2 + T_h = 8$$

O novo candidato entra no mercado oferecendo uma transferência T_h para a mulher. Se o valor de T_h for igual à zero, a mulher estaria indiferente entre estar casada com h ou com k , pois ela já ganhava $Y'_w = 2$ na situação anterior. Portanto, tem-se que T_h deve ser maior que zero ($T_h > 0$):

Se a transferência for de uma unidade $T_h = 1$:

$$\mu_2: (Y'_h - T_h) + (Y'_w + T_h) = (6 - 1) + (2 + 1) = 8$$

Transferindo $T_h = 1$ para a mulher o homem h consegue se casar. Mas nesse caso, k pode ainda fazer uma transferência de $T_k = 2$, pois ainda assim terá utilidade maior do que estando solteiro, $(5 - 2) > 1$:

$$\mu_1: (Y'_k - T_k) + (Y'_w + T_k) = (5 - 2) + (2 + 2) = 7$$

Ocorrerá uma competição entre os homens em que, no limite, cada um deles pagará a diferença de utilidade entre estar casado ou solteiro:

$$T_k \leq (Y'_k - Y_{kk}) = (U_k(\mu_1) - U_k(\mu(k) = k))$$

$$T_h \leq (Y'_h - Y_{hh}) = (U_h(\mu_2) - U_h(\mu(m) = m))$$

No exemplo, os limites superiores de T_k e T_h são 4 e 5 ($T_k \leq (5 - 1)$ e $T_h \leq (6 - 1)$). Numa competição entre eles, o homem k pagará em transferências até o seu limite, $T_k = 4$. O homem h não precisará pagar seu valor total de 5, mas sim algo ligeiramente maior do que 4. Ao transferir

esse valor, seu produto obtido com o casamento ficaria pouco menor do que 2, mas ainda assim melhor do que permanecer solteiro, e a mulher receberia um produto pouco maior do que 6, situação bem melhor do que o casamento inicial, onde havia apenas um homem e uma mulher no mercado.⁷¹

A consequência do teorema **T.10** é que com a possibilidade de transferência entre os agentes, permanecerão casados os homens (ou mulheres, no caso contrário) que tiverem maiores possibilidades de transferências.

No exemplo 10, como $T_h \geq T_k$, o homem h conseguiu se casar com a mulher w . Para mostrar que a situação do exemplo (quando há um homem a mais no mercado) é sempre mais preferida ou tão boa quanto a situação contrária, pode-se explorar uma prova pela exaustão dos casos. Suponha que o novo homem entrante h poderia ter produto menor do que k ($T_h \leq T_k$). Ainda assim k , para permanecer casado, terá de transferir para a mulher uma quantia $T_k > 0$, sob pena de perder o casamento para o novo entrante h que pode fazer transferências positivas.

Em último caso, há a situação em que h não ganha nada no casamento com w e, portanto, o valor das transferências é nulo: $T_h = 0$. A mulher não ganha com a entrada no novo homem, mas continuaria na mesma situação inicial, casada com k ($Y'_w = 2$).⁷²

Pode-se, por um lado, considerar que a abordagem de Becker é mais flexível pois permite a acomodação de preferências de acordo com transferência de valores (ou utilidades) e existe um modelo de barganha. Por outro lado, pode-se pensar que este modelo exige condições mais fortes tais como a de que as funções de utilidades sejam iguais para todos os indivíduos de um mesmo grupo, levando-os a classificar os membros de outro grupo sob uma mesma regra. Grosso modo, a forma de classificar os homens do melhor ao pior é um consenso entre as mulheres no modelo

⁷¹ Essa dinâmica lembra bastante o argumento das terras marginais de David Ricardo (1996 [1817] cap. II). Outro ponto que se pode acrescentar é que o produto de estar solteiro pode ser diferente entre os indivíduos, $Y_{hh} \neq Y_{kk}$, as transferências dependem do ganho potencial do casamento e esse valor mínimo de estar solteiro.

⁷² O **T.10** pode ser escrito em termos mais gerais, para isso, consulte-se Becker (1991, p.128). Em verdade, Becker (1991) não enuncia essa consequência do modelo na forma de um teorema como aqui enunciamos, mas prova as propriedades do modelo e suas consequências. Observa-se que apesar das diferenças dos *frameworks* teóricos, esse teorema é similar ao Teorema 2.25 de Roth & Sotomayor (1990, p.44) onde há prova formal para o contexto de mercado de dois lados.

Beckeriano e o mesmo vale para homens classificando as mulheres. No caso de alunos e escolas, seria equivalente a todos os alunos possuírem um consenso absoluto sobre quais as melhores escolas e quais as piores.

Por esse último ponto apresentado, o modelo de preferências do *two-sided markets* faz menores exigências sobre a utilidade dos agentes, mas é também insensível a mudanças nas preferências, pois elas são dadas e determinísticas.⁷³

Para ver este ponto, suponha que uma mulher w declara sua preferência por dois homens (h_1 e h_2) na seguinte ordem: $h_1 \succ_w h_2$. Caso essas três pessoas sejam as únicas no mercado, a solução *Gale Shapley* formaria o par (w, h_1) , e o homem h_2 permaneceria solteiro.

No modelo de Becker (1991), a solução depende das utilidades obtidas por meio da união do casal (do *household production*). O homem com a propensão de permanecer solteiro (h_2) pode aumentar o preço do casamento, compromissando-se então a transferir mais bens para a esposa w . Caso h_1 não consiga cobrir a oferta de h_2 , o casamento que derivará maior utilidade será o par (w, h_2) .

Como as trocas são voluntárias, observa-se que, h_2 nunca oferecerá transferências maiores do que a utilidade conseguida com o casamento. Em outras palavras, o modelo de Becker prevê a ocorrência de casamentos de conveniência e a barganha pode alterar o quadro de preferências original.⁷⁴

Transportando a questão dos casamentos para o problema de admissão dos alunos nas escolas, em que o grupo dos homens representa o grupo dos alunos e o das mulheres o das escolas. Temos que o modelo acima possui relação com o casamento polígamo (ou poliandria, uma mulher

⁷³ Ver a comparação diagramática dos modelos no Anexo C.

⁷⁴ Consistente com as preferências inicialmente declaradas pela mulher w , desconsiderando as transferências, a utilidade de se casar com h_1 deve ser maior do que se casando com h_2 : $U_w(\mu_1) > U_w(\mu_2)$. E assim ela prefere h_1 a h_2 . Entretanto, no modelo de Becker, as transferências devem ser consideradas. Se $U_w(\mu_2 + T_2) > U_w(\mu_1 + T_1)$, o casamento entre w e h_2 será formado. No arcabouço de Roth & Sotomayor (1990), a mulher teria de saber de antemão o valor das transferências que h_2 pode dar. Se ela soubesse, w não deveria declarar $h_1 \succ h_2$, mas sim $h_2 \succ h_1$. A questão de diferentes tipos de preferências não estritas ou como os agentes agem sem informação incompleta foi tratada por Roth (1989), Roth & Sotomayor (1990, cap 4 e 5), e mais recentemente, Liu et al. (2012).

casada com dois ou mais homens). Uma escola admite um número de alunos até o limite de seu benefício marginal. Nesse caso, o modelo de Becker pode ser generalizado para uma situação de uma pessoa de um sexo e muitas do outro, mas as condições postas nesta seção continuam válidas após a generalização.

A generalização acima permite conceber uma oferta mais elástica. Por esses modelos, os alunos julgam as escolas por uma característica comum, e uma escola muito produtiva gera excedente e é preferida por todos. Para aproveitar melhor os recursos dessa escola muito produtiva, pode-se colocar mais alunos nesse tipo de escola do que para uma escola de baixo rendimento.

Um teste empírico para essa predição do modelo de Becker (1991, cap. 3) é verificar se as escolas mais produtivas são as que possuem maior número de alunos (mais turmas). Essa é uma consequência natural do modelo acima proposto, o número de vagas em uma escola seria predito principalmente pela produtividade da escola. Talvez essa condição seja mais fácil de ser testada em escolas particulares, em que a escolha depende mais de critérios utilitaristas do que dentro do sistema público.

Não obstante, a perspectiva de criação de vagas nas escolas públicas serve também como um guia para as políticas públicas. Ao se planejar a criação de vagas em uma escola já existente, o modelo de casamentos poligâmicos sugere que se criem mais vagas nas melhores escolas. Essa perspectiva será retomada para uma das simulações do capítulo 7.

3.6. Uma proposição para matrizes de distâncias.

Um ponto prático do sistema de alocação de Belo Horizonte é a consideração das distâncias dos alunos até a escola. No sistema de cadastro da cidade essa informação é crucial, e para a analogia dos modelos aqui propostos, sugere-se que as distâncias podem moldar preferências tanto das escolas quanto dos pais e alunos.

Exemplo. 11: Suponha a seguinte matriz de distâncias em quilômetros dos alunos até as escolas:⁷⁵

$$\mathbf{D} = \begin{matrix} & \begin{matrix} s_1 & s_2 & s_3 & s_4 \end{matrix} \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{matrix} & \begin{bmatrix} 2,0 & 3,0 & 4,0 & 6,4 \\ 1,0 & 4,5 & 5,4 & 5,8 \\ 4,0 & 2,2 & 2,0 & 3,0 \\ 3,6 & 6,0 & 7,0 & 8,6 \end{bmatrix} \end{matrix}$$

Cada escola possui apenas uma vaga e a matriz de distâncias determina as preferências de alunos e escolas, quanto menor a distância, mais preferida é a escola (ou o aluno, na perspectiva da escola). Dado isso, as preferências de alunos e escolas estão representadas abaixo:

Alunos:

$$\begin{aligned} \mathcal{P}(i_1) &= s_1, s_2, s_3, s_4 \\ \mathcal{P}(i_2) &= s_1, s_2, s_3, s_4 \\ \mathcal{P}(i_3) &= s_3, s_2, s_4, s_1 \\ \mathcal{P}(i_4) &= s_1, s_2, s_3, s_4 \end{aligned}$$

Escolas:

$$\begin{aligned} \mathcal{P}(s_1) &= i_2, i_1, i_4, i_3 \\ \mathcal{P}(s_2) &= i_3, i_1, i_2, i_4 \\ \mathcal{P}(s_3) &= i_3, i_1, i_2, i_4 \\ \mathcal{P}(s_4) &= i_3, i_2, i_1, i_4 \\ Q &= (q_1 = 1, q_2 = 1, q_3 = 1, q_4 = 1) \end{aligned}$$

Aplicando-se o algoritmo *deferred acceptance* com os alunos propondo temos os seguintes 4 passos:

Passo 1. Os alunos i_1 , i_2 e i_4 propõem todos para a escola s_1 . O aluno i_3 propõe para a escola s_3 . A escola s_1 de acordo com sua ordem de prioridades (dada pelas distâncias) “segura” a proposta do aluno i_2 e rejeita as ofertas dos demais.

Passo 2. Os alunos i_1 e i_4 propõem para a segunda opção: a escola s_2 , que é a segunda melhor escola para os dois alunos. Como s_2 recebe duas propostas, de acordo com seu ordenamento de prioridades, ela “segura” o aluno i_1 e rejeita a oferta de i_4 .

⁷⁵ No anexo D apresenta-se um diagrama de pontos de escolas e alunos no plano que gera a matriz de distâncias acima.

Passo 3. O aluno i_4 propõe para a escola s_3 que já havia recebido proposta de i_3 . A escola s_3 analisa a nova oferta e decide continuar com o aluno i_3 e rejeitar i_4 (dado que i_4 é o de menor prioridade, maior distância).

Passo 4. O aluno i_4 propõe para a escola s_4 , que ainda não havia recebido propostas e o algoritmo se encerra. As escolas matriculam os alunos que estavam segurando até este momento.

Ao final do algoritmo tem-se que a matriz de alocação final é:

$$\mathbf{M} = \begin{matrix} & s_1 & s_2 & s_3 & s_4 \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

Analisando-se o caso em que as escolas propõem primeiro no algoritmo *deferred acceptance*, tem-se o algoritmo em 4 passos:

Passo 1. A escola s_1 propõe vaga ao aluno i_2 e as escolas s_2 , s_3 e s_4 propõem todas elas para o aluno i_3 . Como o aluno i_3 recebeu três ofertas ele poderá escolher a que ele mais prefere, no caso é s_3 . As escolas s_2 e s_4 são rejeitadas por i_3 .

Passo 2. A escola s_2 propõe vaga para o aluno i_1 , que ainda não havia recebido proposta e então “segura” a oferta, e a escola s_4 propõe ao aluno i_2 que já possui a oferta de s_1 . Como i_2 recebeu mais de uma oferta, ele analisa e mantém a proposta de s_1 (mais preferida) e rejeita a de s_4 .

Passo 3. A escola s_4 propõe ao aluno i_1 que está “segurando” oferta de s_2 . Entre s_2 e s_4 , o aluno i_1 prefere s_2 , e s_4 é novamente rejeitada, sendo necessário mais um passo.

Passo 4. A escola s_4 propõe ao aluno i_4 , que não havia recebido propostas até então, o aluno i_4 aceita a oferta recebida e o algoritmo termina. Os alunos se matriculam na oferta que estavam segurando até o momento.

Observa-se que a alocação acima é a mesma da matriz **M** que foi obtida com os alunos propondo no algoritmo *deferred acceptance*. O ocorrido parece não ser por acaso tal como a proposição **P.1** abaixo revela:

P.1. *Em escolas com uma vaga, para matrizes de preferências dadas pela matriz de distâncias onde não ocorra dois ou mais alunos equidistantes de uma mesma escola ou não haja duas ou mais escolas a uma mesma distância para um dado aluno, tem-se que o matching estável proposto pelo algoritmo deferred acceptance (escolas propondo ou alunos propondo) será único.*

A proposição acima pode ser provada por contradição:

Prova: Suponha que μ seja o casamento dado pelo algoritmo *deferred acceptance* com os alunos propondo e μ' seja o *matching* dado pelo algoritmo das escolas propondo. Suponha que $\mu \neq \mu'$, se isso ocorre, para casos de escola com apenas uma vaga, há pelo menos dois alunos, digamos que i e i' , cuja escola proposta por μ é diferente de μ' ($\mu(i) \neq \mu'(i)$ e $\mu(i') \neq \mu'(i')$). Considere-se ainda que $\mu(i) = s$ e $\mu'(i) = s'$, e de contraparte, para que o anterior ocorra, que $\mu(i') = s'$ e $\mu'(i') = s$.

Se $\mu \neq \mu'$, sabe-se por **T.2** que $s \succ_i s'$ e que $s' \succ_{i'} s$. Como as preferências foram dadas pelas menores distâncias (quanto mais próximo melhor) isso equivale a dizer que para i as distâncias são $A < B$, e para o i' , são de $D < C$, em que A, B, C, D são valores de $\mathbb{R}$ para as distâncias. A representa a distância do aluno i até a escola s , B é a distância de i até a escola s' , o C representa a distância de i' até a escola s e D é a distância de i' até a escola s' . Por **T.3**, sabe-se que para s : $i' \succ_s i$ e para s' : $i \succ_{s'} i'$, como a matriz de distâncias é a mesma, isso equivale a dizer que para s : $C < A$ e para s' : $B < D$, o que é uma contradição, pois viola a transitividade.

Pela transitividade e pelas três primeiras proposições para as distâncias, tem-se que: $D < C < A < B$. No entanto, a última assertiva de que $B < D$ se constitui uma contradição às afirmações anteriores. Tem-se, portanto, que $\mu \neq \mu'$ gera uma contradição sob as condições impostas em **P.1**. Não é possível, com preferências dadas por uma matriz de distância para alunos e escolas, *matchings* estáveis diferentes dados pelo algoritmo *deferred acceptance*.

Como foi mostrado para dois alunos e escolas quaisquer em que $(i, i') \in I$ e $(s, s') \in S$, para *matchings* não idênticos, pelo menos um par de alunos e escolas do tipo acima ocorre e isso fornece a contradição necessária para a prova, com isso tem-se a prova de **P.1**. ■

Nota-se que a proposição **P.1** fala de escolas com apenas uma vaga, isso faz com que o *matching* da proposição não seja diferente do caso *one-to-one*. Generalizar essa proposição e a prova para o caso *many-to-one* exige que preferências coincidentes sejam estritas para alunos e escolas. Ou seja, a regra de desempate deve ser a mesma para alunos e escolas. Essa extensão pode ser feita em trabalhos futuros.

Pode se imaginar dois ou mais alunos que estão equidistantes da mesma escola, por exemplo, dois irmãos que moram na mesma residência. Sendo assim, o critério de desempate deve priorizar um dos irmãos e ao mesmo tempo esse irmão priorizado deve também ter que priorizar essa escola um pouco mais do que o irmão preterido (o critério de desempate deve funcionar tal como se a distância para um dos irmãos fosse reduzida em pequena fração para gerar o desempate).

Para uma mesma série o empate acima não é comum (teriam de ser irmãos gêmeos ou um dos irmãos estar atrasado nos estudos). Ainda assim, para bancos com muitos alunos, a probabilidade que pelo menos um caso ocorra é grande, sendo necessário um critério para desempatar.

Um ponto importante da proposição é que ela faz com que o *matching* do DA, considerando apenas as distâncias como critério de escolha, seja o mesmo. Esse *matching* não é o que minimiza a distância média dos alunos (ver exemplo no anexo D), no entanto, ele apresenta resultados de distâncias estáveis, de modo que não há *inveja justificada*.

4. A FUNÇÃO DE ALOCAÇÃO E PROFICIÊNCIA

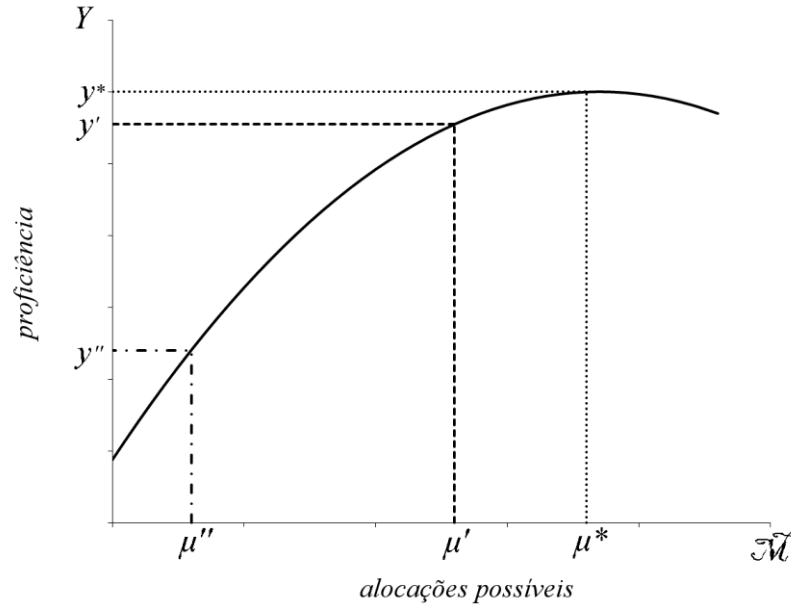
Um tema de investigação da tese é como se relacionam as diferentes alocações propostas com o bem-estar medido pela proficiência dos alunos. A suposição é de que a sociedade não é indiferente à maneira como são alocados os alunos e escolas, e sendo assim, existe uma alocação ótima possível, a que maximiza a proficiência dos alunos no domínio de todas as alocações possíveis $\mathcal{M}$, em que μ representa cada uma dessas alocações ($\mu \in \mathcal{M}$). A relação entre a proficiência y ($y = \frac{Y}{n}$) e as alocações μ , pode ser pensada na forma de uma equação geral representada abaixo, semelhante à equação (2):

$$y = f(\mu) \quad (14)$$

Em que f é uma função contínua com primeira derivada em relação à μ positiva e segunda derivada é negativa. As alocações μ são tomadas de maneira contínua e o resultado da proficiência é dado em relação à alocação observada. A FIG. 9 ilustra o conceito desta função. Na FIG. 9, o μ'' pode ser uma alocação que representa o estágio inicial anterior à implementação do sistema de cadastro, ou seja, uma alocação casuística derivada da baixa organização do processo de matrículas. O μ' pode traduzir uma situação em que algoritmos de pareamento são utilizados e contribuem para elevar a média geral de proficiência dos alunos (y').

O μ^* representa uma possível alocação ótima que pode porventura existir. Uma das tarefas desta tese é descobrir o quanto se pode conseguir com outras alocações, tentando se aproximar do desenho ótimo, em que, controlada pelos demais fatores, a alocação de alunos e escolas oferece o ganho máximo em termos da média de proficiência (y^*).

Figura 9 - Ganhos de proficiência (Y) e o conjunto de todas as alocações possíveis ($\mathcal{M}$)



Fonte: Elaboração do autor.

Tratar dessa relação entre alocações e proficiência da maneira apresentada pela função da equação (14) encerra alguns desafios:

1. Estabelecer uma cardinalidade para todas as alocações é uma questão não trivial. Pode-se indicar uma proposta similar ao modelo de Becker (1991), discutida em 3.4, caso se conheça as características das escolas (tais como o capital, K) e características dos alunos (dadas por L), as melhores alocações são as que combinam positivamente (ou negativamente) as características K e L , na suposição de *positive assortative matings* (ou *negative assortative matings*).⁷⁶
2. A função f deve ser uma função de agregação social de diversas proficiências distintas obtidas para cada aluno em cada escola. Como uma função que agrega vários indivíduos,

⁷⁶ Roth & Sotomayor (1990, pp.36-39) mostram pelo teorema de *Lattice* que é possível ordenar pelos critérios de um dos grupos (homens ou mulheres) os *matchings* estáveis. Isto é, dentro do conjunto de *matchings* estáveis, as alocações terão um ordenamento do menos preferido ao mais preferido. No entanto, em uma extensão do T.3 abordado no capítulo anterior, esse ordenamento é reverso entre os dois lados do mercado, o mais preferido dos homens é o menos preferido das mulheres e vice-versa. Ademais, os autores mostram que na forma ordinal não há como estabelecer esse ordenamento entre *matchings* não estáveis.

são necessárias considerações normativas para a forma funcional desta função. A única exceção seria a situação em que as proficiências médias aumentassem ou diminuíssem igualmente para todos os alunos, não sendo necessário ponderar alunos diferentemente, pode-se adotar um aluno representativo. Porém, é difícil que cada alocação confira os mesmos valores marginais para a proficiência de todos os alunos. O mais provável é que uns alunos ganhem ao serem trocados de escolas e outros não. Para esses casos, a agregação se torna um problema difícil de ser resolvido.

3. Os valores de μ e y podem não ser contínuos e a função pode não ser contínua. Apesar de a variável de proficiência (y), por construção, assumir valores contínuos de 0 até 500,⁷⁷ não há indicativos de como é possível medir μ de maneira contínua, a não ser que se adote uma aproximação tal como levantado no ponto 1. Outra coisa que se pode fazer é ter uma medição da distância entre a alocação ótima e a observada, tal como representada pela equação (1) do capítulo anterior. No entanto, essa ideia precisa ser melhor desenvolvida.

Observando-se em detalhes as questões acima, pode-se dar um exemplo que demonstra o problema de agregação referido e aponta indicativos de como as simulações podem ser encaminhadas.

Exemplo. 12: suponha dois alunos: i_1 e i_2 estudantes das escolas s_1 e s_2 respectivamente, este é o mercado inteiro do exemplo. A nota do aluno i_1 é representada por $y_{1,1}$, que representa a nota do aluno i_1 na escola s_1 (na forma geral: $y_{i,s}$, nota para o aluno i na escola s). Em contrapartida, a nota do aluno i_2 na escola s_2 é dada por $y_{2,2}$.

Nesta tese supõe-se, na maior parte dos casos, um número de vagas constante. Então, para que i_1 mude de s_1 para s_2 é preciso que i_2 mude de s_2 para s_1 . Para tratar de outras alocações possíveis é preciso ter um modelo que diga como ocorreriam as notas dos alunos nas situações alternativas (contrafactuais), que são:

⁷⁷ A variável y é obtida pelo método da Teoria de Resposta ao Item (TRI). Em verdade, valores iguais a zero são para alunos que não fizeram a prova ou erraram todas as questões e geralmente não são tratados, a distribuição começa a apresentar valores acima de zero nos intervalos próximos de 30 a 50 pontos. Outro ponto é que a variável y não é perfeitamente contínua, mas pode ser tratada como tal. No capítulo 5 a seguir serão fornecidos maiores detalhes.

1. $y_{1,2}$ (é a nota do aluno i_1 caso ele estudasse em s_2).
2. $y_{2,1}$ (é a nota do aluno i_2 caso ele estudasse em s_1).

Se a troca de escolas for mutuamente vantajosa, tem-se que $y_{1,2} > y_{1,1}$ e $y_{2,1} > y_{2,2}$. Em palavras, o aluno i_1 consegue melhores resultados na escola s_2 , enquanto que i_2 obtêm notas melhores na escola s_1 , a alocação inicial colocava os alunos nas escolas “erradas”. Neste exemplo a troca entre os alunos é Pareto ótima.

Contudo, outras possibilidades, exploradas na TAB. 7 abaixo podem ocorrer. Se os dois alunos estão estritamente melhores como no exemplo acima, se diz que o a troca é um melhoramento *fortemente preferível* do ponto de vista dos alunos, representantes da sociedade.⁷⁸

Se apenas um dos alunos se favorece com a troca, mas o outro não perde, diz-se que a troca é *fracamente preferível*. Caso não haja resultado favorável, os alunos ficam indiferentes entre a troca. Existem situações reversas: *fracamente não-preferível* e *fortemente não-preferível*. Além disso, podem haver situações *ambíguas*, um dos alunos ganha, mas o outro perde proficiência.

Quando a troca beneficia ambos os alunos, não haverá oposições sociais a essa situação. Ou então, na pior das hipóteses, alcançará resultados *fracamente preferíveis*, em que um aluno melhora com a troca, mas o outro não piora. Todos os casos assinalados em verde na TAB. 7 são melhorias de Pareto, é possível melhorar pelo menos um agente sem piorar nenhum outro. Os casos em que não há nenhuma cor, são indiferentes ou ambíguos e os casos em vermelhos são socialmente indesejáveis.

⁷⁸ Por simplicidade, supõe-se aqui que as escolas são indiferentes, o que vale para a função de bem estar social são as notas dos alunos, mais a frente no capítulo 7, as preferências das escolas serão consideradas na análise.

Tabela 7 - Resultado Social da troca de dois alunos i_1 e i_2 entre as escolas s_1 e s_2 .

		Aluno i_2 mudou-se para a escola s_1		
		$y_{2,1} > y_{2,2}$	$y_{2,1} = y_{2,2}$	$y_{2,1} < y_{2,2}$
Aluno i_1 mudou-se para a escola s_2	$y_{1,2} > y_{1,1}$	Fortemente preferível	Fracamente preferível	Ambíguo
	$y_{1,2} = y_{1,1}$	Fracamente preferível	Indiferente	Fracamente não-preferível
	$y_{1,2} < y_{1,1}$	Ambíguo	Fracamente não-preferível	Fortemente não-preferível

Fonte: Elaboração do próprio autor.

Os casos *ambíguos*, em que um aluno melhora mas o outro piora, possuem por trás um julgamento normativo. A sociedade pode julgar que para o aluno i_2 ganhar mais 20 pontos de proficiência, por exemplo, pode valer à pena o aluno i_1 perder 10 ou menos pontos. Essa é a tarefa que a função de agregação social f mencionada pode ajudar a resolver. Porém, ainda assim, os critérios para tal são subjetivos, é uma ponderação muito difícil entre indivíduos diferentes.

Um critério de repartição do produto de uma coalizão em jogos cooperativos é dado pelo *valor de Shapley*, proposto por Shapley (1951). Nesse critério cada participante receberia em termos de proficiência o valor que o jogador agrega ao resultado geral da coalizão. Seria preciso conhecer o quanto que cada aluno agrega ao estar matriculado em uma nova escola proposta pelas diferentes alocações possíveis. O uso do *valor de Shapley* para o produto educacional tem também dificuldades por conta de não ser o produto educacional transferível facilmente entre os alunos, encarar como podem ser os ganhos em Y e sua distribuição é um desafio.

Uma posição conservadora seria a de só realizar trocas para as células verdes acima indicadas, alcançando-se uma alocação Pareto eficiente. Existem ganhos de proficiência que podem ser arbitrados considerando-se apenas o aspecto de encurtamento das distâncias que os mecanismos

de alocação em geral propiciam, outros advêm de efeito migração e efeito escola que serão detalhados no capítulo 6.

Tais critérios podem ser ainda insuficientes para promover maiores trocas e aproveitar-se dos maiores ganhos possíveis. O capítulo 6, na seção 6.3, voltará a este tema apresentando algumas escolhas dos modelos de inferência que serão utilizados como base para os parâmetros do capítulo 7.

Antes disso, o próximo capítulo apresenta como foi construído o banco de dados para lidar com as informações de alocação dos alunos e escolas. Dados do sistema de cadastro de Belo Horizonte são apresentados, bem como dados do Censo Escolar 2010 e do PROEB de língua portuguesa para o mesmo ano. Detalhes do georreferenciamento e processamento do banco são também fornecidos no capítulo 5.

5. BANCO DE DADOS DOS ALUNOS E ESCOLAS DE BELO HORIZONTE

A etapa de elaboração do banco de dados de alunos e escolas é crucial para se compreender o sistema de cadastramento escolar de Belo Horizonte. Neste capítulo descrevem-se as fontes primárias e as etapas necessárias para a construção de um banco de dados de alunos e escolas conjugados. Por circunscrever-se apenas aos limites do município, o banco de dados possui alunos e escolas situados dentro dos limites legais de Belo Horizonte.⁷⁹

O universo de escolas representadas compõe-se de escolas públicas da rede de ensino municipal e estadual. A rede de escolas públicas federais (apenas quatro em BH) não pertence ao universo de análise aqui examinado, tais escolas possuem sistema de cadastro próprio. Cabe enfatizar que escolas da rede privada não estão no banco de dados, naturalmente, por se tratar de um cadastro do sistema público de ensino, no entanto, a cada ano há uma pequena parcela de alunos novos entrantes provenientes do sistema particular (processo relatado no capítulo 2).

O banco de dados foi construído pelo uso de três principais fontes: Censo Escolar (MEC/INEP); banco do PROEB (Programa de Avaliação da Rede Pública de Educação Básica); e um banco de escolas com o código, nome e endereço, fornecido pela SEE/MG.⁸⁰ Outros bancos menores, mas ainda sim muito importantes, foram utilizados: dados de registro do sistema de cadastro escolar da SME/PBH; um banco de dados com o nível Socioeconômico (NSE) das escolas criado por Soares & Alves (2012); bancos de dados georreferenciados das regiões administrativas e bairros de Belo Horizonte; e um banco de endereços elaborado por Martins, Davis Jr. & Fonseca (2012) utilizado em processo que será mais bem descrito em 5.2.2.

⁷⁹ Muitas zonas limítrofes de Belo Horizonte já se apresentam quase que completamente conurbadas aos municípios vizinhos do entorno: Contagem, Nova Lima, Sabará, Santa Luzia, Ribeirão das Neves e Vespasiano são os mais importantes do entorno. Alunos moradores de tais municípios vizinhos não foram inseridos no banco de dados, alguns destes alunos podem morar bem próximos a escolas de Belo Horizonte, e, em alguns casos pais podem se vir tentados a alterar um pouco o endereço de residência de forma conseguirem matrícula incluída dentro dos limites da capital, o que pode gerar um “efeito borda”.

⁸⁰ O Banco de escolas fornecido pela secretaria de educação estadual tem tanto escolas estaduais quanto escolas municipais.

Sobre o universo discente, foram analisados dados de alunos matriculados em 2010 no 5º ano do ensino fundamental, lembrando que após a lei nº 11.274 de 2006 o ensino fundamental passou a ter nove anos. Em Minas Gerais, segundo dados do censo escolar, a lei já se encontra praticamente universalizada, em 2010 menos de 1% dos alunos continuava no regime seriado de oito anos.⁸¹ Ainda de acordo com dados do censo escolar para o mesmo ano, Belo Horizonte possuía 37.058 estudantes matriculados no 5º ano, matrículas de escolas de todas as redes na modalidade de ensino regular (públicas e privadas), 27.371 eram de escolas públicas municipais ou estaduais.

A série decisiva para entrada dos alunos no sistema de cadastro é o 1º ano do fundamental, série de alfabetização. É nesse ano de ensino que estão localizados maior parte dos alunos novos entrantes, pois se torna obrigatória a matrícula no sistema de ensino básico. A partir dessa matrícula inicial, as matrículas subsequentes são mantidas na mesma escola, a não ser em casos raros de exceção já relatados.

Infelizmente, não há nenhum sistema de avaliação da proficiência realizado nesse primeiro ano de entrada no ensino fundamental, sendo que a etapa mais próxima a ser avaliada é o 5º ano.⁸² Pelo fato de grande parte das matrículas para os anos subsequentes serem automáticas e por ser o 5º ano a primeira etapa para a qual se possui avaliação padronizada da proficiência, os alunos dessa fase foram escolhidos como objeto de análise.

A escolha da prova de Língua Portuguesa ocorreu por conta de o banco contar com mais alunos, tendo-se uma amostra maior de alunos para o processamento dos bancos do Censo Escolar e do PROEB. Além disso, acrescenta-se que os resultados das notas de português e matemática são fortemente correlacionados ($\rho = 0.71$) de modo que não há perda de generalidade em usar apenas uma das provas.

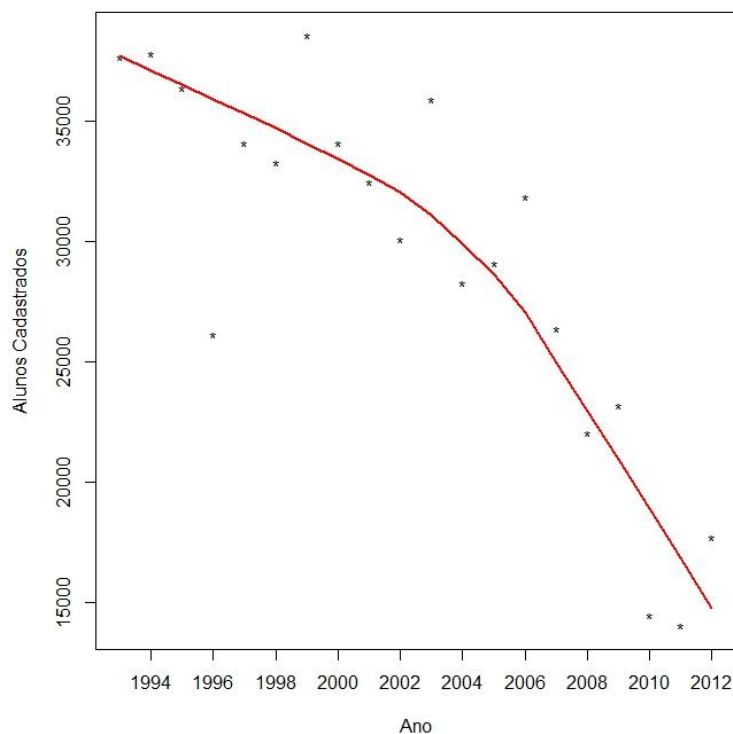
⁸¹ Em números, em 2010 um total de 26.618 alunos do ensino fundamental continuava no sistema seriado de ensino de 8 anos, em um total de 3.037.343 de alunos do fundamental em Minas Gerais.

⁸² Recentemente, em 2012, o MEC criou a Provinha Brasil (<http://provinhabrasil.inep.gov.br/>), um exame em duas etapas para alunos do segundo ano do ensino fundamental, permitindo um acompanhamento longitudinal no segundo ano. No entanto, dado a novidade do banco, não foi possível que ele estivesse disponível nas etapas iniciais de elaboração deste trabalho.

Dados de registro administrativo da gerência de cadastro da PBH estão presentes no GRAF. 1 abaixo. Eles apresentam o número de alunos que requisitaram entrada no sistema para a primeira série desde 1994 até 2013. Os dados se referem à demanda feita pelos pais no ano anterior e não necessariamente correspondem ao total de alunos que efetivamente se matricularam e estão presentes no Censo Escolar.

Essa demanda é realizada nas agências dos Correios, geralmente no mês de agosto, as matrículas são realizadas em dezembro e o início efetivo das aulas se dá no ano seguinte. Dessa forma, os 37.667 alunos requerentes de 1994 que aparecem no gráfico se referem ao biênio 1993-1994, a mesma regra é aplicada para os anos seguintes.

Gráfico 1 - Evolução da Demanda do Cadastro Escolar realizado nos Correios de BH

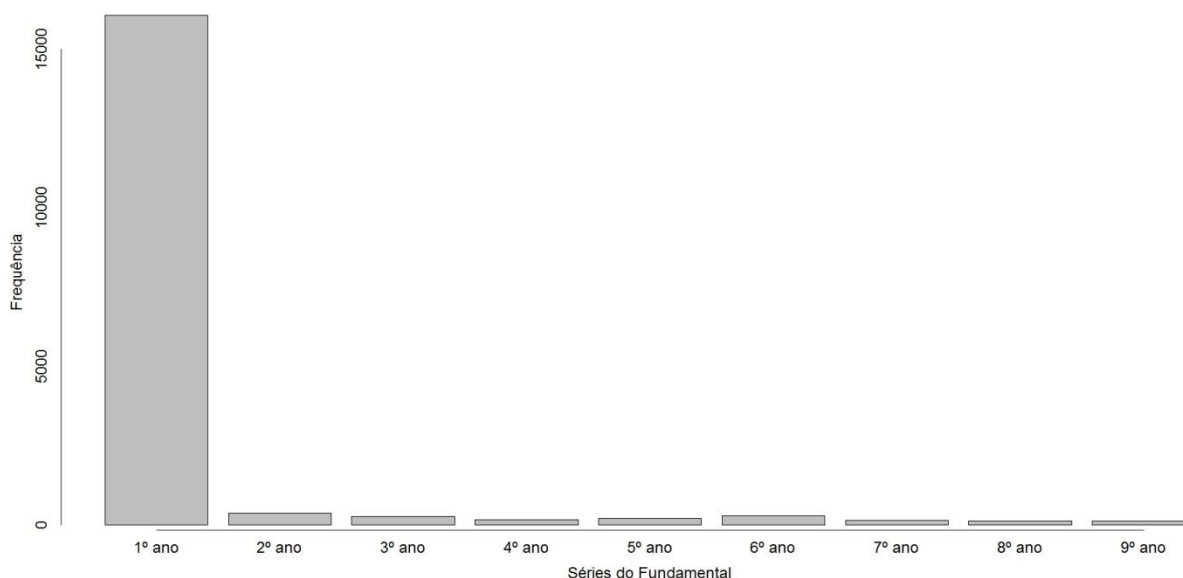


Fonte: Dados de registro da gerência de Cadastro PBH/PRODABEL (2013).

Observa-se que o total da demanda tem apresentado grande tendência de queda (cerca de 50% em 18 anos), muito provavelmente por motivos demográficos. Em 2011, por conta da mudança do ensino fundamental para 9 anos, o registro passou a contar com o 9º ano do ensino fundamental.

O GRAF. 2 abaixo apresenta a distribuição da demanda de acordo com o ano escolar dos alunos para 2012-2013. Nota-se que o primeiro ano conta com o maior número de alunos demandantes. Depois de matriculados em uma escola no primeiro ano ou em qualquer outra etapa de ingresso, as matrículas nos períodos subsequentes são automáticas e mantidas na mesma escola sugerida no processo inicial, a não ser nos casos de exceção relatados no capítulo 2.

Gráfico 2- Distribuição da demanda no Cadastro Escolar de BH no ano de 2012-2013



Fonte: Dados de registro da gerência de Cadastro PBH/PRODABEL (2013).

O Censo Escolar é realizado pelo MEC desde o início dos anos 90 e recentemente, em 2007, passou por uma reformulação (o *Educacenso*) que possibilitou inclusão de mais informações e de maneira mais confiável. Neste trabalho o banco de endereços dos alunos é proveniente do Censo Escolar. Além da informação de endereços, o censo possui algumas características pessoais dos alunos como nome completo, sexo, cor, idade, código da escola e outras informações cadastrais (ver descrição das variáveis na TAB. 8). Cada linha do banco representa um aluno e o censo escolar possui todas as matrículas para o 5º ano no início do ano letivo. No caso, o banco possui 29.256 linhas (entre alunos regulares e especiais).

Para união das bases do Censo Educacional e do PROEB foram necessárias informações confidenciais com os nomes dos alunos e endereços. A informação dos nomes, em geral, não está

disponível no *Censo Escolar* fornecido pelo MEC. Em Junho de 2012, foi assinado um termo de confidencialidade com a Secretaria Estadual de Educação que disponibilizou dados com as informações de nome e endereço dos alunos (as condições do termo assinado estão no Anexo E).

O banco de dados do PROEB traz, além da importante informação de proficiência, informações úteis como a escolaridade da mãe, o padrão de consumo de bens duráveis da família e se a família do aluno é ou não beneficiária do programa federal bolsa família (TAB. 8). O PROEB está inserido no Sistema Mineiro de Avaliação da Educação Pública (SIMAVE), que aplica, a cada ano, provas para alunos do 5º e 9º anos do fundamental e para o 3º ano do ensino médio. As provas abordam conhecimentos das disciplinas de matemática e língua portuguesa.

O SIMAVE foi implantado no ano de 2002 e primeiramente se restringia a escolas estaduais e aplicava provas de matemática e português em anos alternados. Desde então, o sistema ampliou seu uso e aproveitamento, e hoje conta com a participação de todas as escolas públicas estaduais e municipais de Minas Gerais. As provas de ambas as matérias são aplicadas anualmente no final do período letivo.

A proficiência obtida por meio do exame do PROEB é dada em uma escala padronizada, correspondente à escala SAEB (Sistema nacional de Avaliação da Educação Básica). A escala se situa entre 0 (nota mínima) e 500 (nota máxima). Os extremos dessa escala raramente são alcançados pelos alunos. Em verdade, a nota obtida no PROEB não se refere a um sistema de pontuação clássico, mas sim a uma escala padronizada dada pelo conjunto da Teoria de Resposta ao Item (TRI)⁸³.

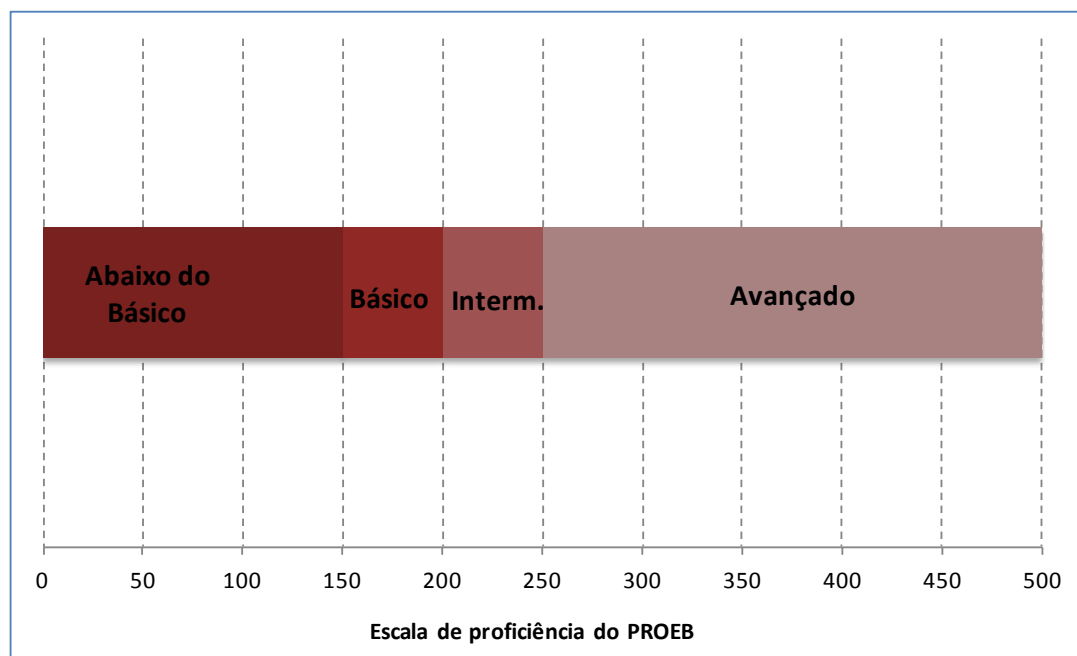
Sendo assim, o nível de proficiência em que o aluno se encontra é mais representativo do que o escore absoluto obtido em uma prova com nota dada pela soma do número de acertos. Além

⁸³ Na Teoria da Resposta ao Item (TRI), também chamada de modelos de características latentes, a medida não é a nota da prova dada pelo número de itens acertados, mas sim um parâmetro ' θ ' que representa uma característica latente do aluno, uma predisposição de acerto de itens que são classificados de acordo com a prova. Em geral, modela-se a TRI de acordo com a probabilidade de acerto $p(\theta) = \frac{1}{1+e^{-(\theta-b)}}$ e um dado parâmetro b_i que revela a dificuldade de um item. No caso do PROEB outros parâmetros são utilizados para maior adequação dos resultados, ver Soares (2005) e Soares & Pereira (2002) para aplicações da TRI nos exames feitos em Minas Gerais e Linden & Hambleton (1997) e Hambleton, Swaminathan, & Rogers (1991) para mais detalhes do método TRI.

disso, o método permite a comparação entre turmas de alunos diferentes, séries e entre diferentes anos e períodos letivos, sendo possível comparar a evolução ao longo dos anos e das etapas. A compatibilidade do PROEB permite comparações com resultados já conhecidos da literatura de proficiência e desempenho educacional. A análise elaborada na tese se delimita a um corte temporal no ano de 2010.

Com o passar do tempo, convencionou-se dividir a escala em 4 níveis: 1) abaixo do básico ($y < 150$); básico ($150 \leq y < 200$); intermediário ($200 \leq y < 250$); e avançado ($y \geq 250$).⁸⁴ A FIG. 10 apresenta a escala e os seus intervalos, destacando os cortes de proficiência mencionados.

Figura 10 - Níveis de proficiência do PROEB em Língua Portuguesa para a 5º ano



Fonte: Elaboração do autor.

Excetuando-se os alunos que não fazem a prova e ficam com informação faltante no escore, notas de proficiência abaixo de 50 pontos são raras. O capítulo 6 mostrará que a menor nota de Língua Portuguesa que ocorre para o 5º ano no banco processado é de 73,73. Também é muito raro um

⁸⁴ O 'y' representa o escore de proficiência $y \in \mathbb{R}$, os valores de corte correspondem à escala do 5º ano na prova de Língua Portuguesa. Para ver a escala de resultados do PROEB para outras séries pode se consultar o portal do SIMAVE na internet: <http://www.simave.caedufjf.net/fique-por-dentro/>. Uma tabela com a escala completa sugerida para o PROEB pode ser também encontrada em Delgado e Miranda-Ribeiro (2012, p. 14).

aluno de 5º ano estar perto dos 500 pontos, nível difícil de ser alcançado até por alunos do 3º ano do ensino médio (para um aluno de 5º ano alcançar esse nível seria necessário ele ter conhecimentos de leitura e interpretação de textos exigidos de um aluno de ensino médio). Em 2010, o máximo alcançado por alunos do 5º ano no banco de dados processado é 349,90.

A TAB. 8, a seguir, apresenta as variáveis disponíveis nos dois bancos de dados do *Censo Escolar* e do PROEB. Na coluna ‘variável’ apresenta-se a ordem e o nome da variável utilizada e na coluna ‘tipo’ apresenta-se a forma de armazenamento da variável no banco (número inteiro, variável categórica, formato de data, alfanumérica, etc.). A ‘descrição’ apresenta uma breve definição da variável. Variáveis que foram utilizadas como variáveis-chave para a fusão dos bancos estão sublinhadas. As variáveis com o prefixo “cod_” representam variáveis de código, ‘cod_entidade’ é a mais importante delas, pois traz o código numérico da escola. Esta variável se encontra em todos os três bancos. Há também dados de endereço dos alunos e das escolas.

Quando o endereço se refere à escola ele possui o sufixo ‘_esc’ para indicar que se trata de um endereço do banco de escolas. Por exemplo, o CEP da escola aparece na variável: ‘cep_esc’. No caso do CEP do aluno, a variável se chama apenas ‘cep’ e está no banco de dados do *Censo Escolar*.

O banco do *Censo Escolar* aqui empregado não traz, todavia, todas as variáveis presentes no levantamento completo. Para o banco aqui utilizado, foram listadas apenas algumas das variáveis mais importantes. Ao final do processo de seleção, o banco do *Censo Escolar* permaneceu com 26 variáveis. Procedimento similar ocorreu com o banco de dados do PROEB, que possui outras informações além das destacadas na TAB. 8. Ao final do processo de seleção das variáveis mais importantes, o banco de dados do PROEB aqui utilizado contou com 32 variáveis ao todo.

O banco de dados de escolas está completo, tal como foi recebido da SEE/MG, e nele constam 9 variáveis (TAB. 9).

Tabela 8 - Variáveis do Censo Escolar e do PROEB (nome, tipo e descrição)

Censo Escolar			Banco do PROEB		
<i>Variável</i>	<i>Tipo</i>	<i>Descrição</i>	<i>Variável</i>	<i>Tipo</i>	<i>Descrição</i>
1.ano	Int.	Ano do censo.	1.cod_proeb	Int.	Código 6 dígitos do aluno no proeb
2.cod_matricula	Int.	Código de 8 dígitos matrícula do aluno	<u>2.cod_entidade</u>	Int.	Código da entidade, 8 dígitos.
3.cod_aluno	Num.	Número de 12 dígitos	3.dependencia	Categ	Estadual ou Municipal
4.dia	Int.	Dia que nasceu	4.cod_regional	Int.	Código 2 dígitos da regional ensino
5.mês	Int.	Mês que nasceu	5.cod_rota	Int.	Código 2 dígitos para o local das provas.
6.ano_nasc	Int.	Ano que nasceu	6.turno	Categ	Integral, Manhã e Tarde
7.idade	Int.	Idade anos completos	<u>7.nome_aluno</u>	Alfab	Nome por extenso do aluno, limite de 54 caracteres.
8.sexo	Categ	Categórica “F” e “M”	8.nome_escola	Alfab	Nome da escola limite de 38 caract.
9.cor	Categ	Cor por extenso.	9.reg_proeb	Categ	Metropolitana A,B, e C
10.cod_munic_nasc	Num.	Código do IBGE 7 dígitos do município onde nasceu	10.proficiência	Num.	Proficiência em Língua Portuguesa na escala PROEB
11.sigla_end	Categ	Sigla do estado onde mora	11.estudo_mae	Int.	6 categorias para os anos de estudo da mãe.
12.cod_municipio_	Int.	Código do IBGE 7 dígitos do município.	12.estudo_pai	Int.	6 categorias para os anos de estudo do pai.
13.id_zona_resid	Categ	Zona urbana ou rural	13.rua_calçada	Int.	Se a rua onde mora é calçada: 0 “Não”; 1 “Sim”.
14.mod_ensino	Categ	Especial ou regular	14.agua	Int.	casas onde mora tem água encanada: 1 “Sim”; 0 “Não”.
15.cod_turma	Int.	Código da turma 7 dígitos	15.banheiro	Int.	Sua casa onde mora tem banheiro: 0 “Nenhum”; 1 “Um”; 2 “Dois”; 3 “Três ou mais”.
16.turma_unificada	Categ	Se é unificada com séries: “Não”, “Multi-fluxo” ou “Fluxo”	16.radio	Int.	Se possui rádio em casa: 0 “Nenhum”; 1 “Um”; 2 “Dois”; 3 “Três ou mais”.
17.cod_tipo_turma	Int.	“0” Normal, “2” Internato	17.geladeira	Int.	Se possui geladeira em casa: 0 “Nenhuma”; 1 “Uma”; 2 “Duas”; 3 “Três ou mais”.
<u>18.cod_entidade</u>	Int.	Código da entidade, 8 dígitos.	18.tvcores	Int.	Se possui TV em cores em casa: 0 “Nenhuma”; 1 “Uma”; 2 “Duas”; 3 “Três ou mais”.
19.id_adm_escola	Categ	Estadual ou Municipal	19.maquina_lavar	Int.	Se possui máquina de lavar roupa em casa: 0 “Nenhuma”; 1 “Uma”; 2 “Duas”; 3 “Três ou mais”.
<u>20.nome_aluno</u>	Alfab	Nome por extenso do aluno, limite de 33 caracteres.	20.automovel	Int.	Se possuem automóvel: “Nenhum”; 1 “Um”; 2 “Dois”; 3 “Três ou mais”.

(Continua)

(Continuação)

Censo Escolar			Banco do PROEB		
<i>Variável</i>	<i>Tipo</i>	<i>Descrição</i>	<i>Variável</i>	<i>Tipo</i>	<i>Descrição</i>
21.dt_nasc	Data	Data de nascimento dd/mm/aaaa	21.dvd	Int.	Se possui aparelho de dvd em casa: 0 “Nenhum”; 1 “Um”; 2 “Dois”; 3 “Três ou mais”.
22.desc_end	Alfab	Endereço, rua, av. etc	22.bolsa_fam	Int.	Família recebe benefício do bolsa família: 0 “Não”; 1 “Sim”.
23.end_comp	Alfan	Complemento do end	23.repetiu	Int.	Se o aluno já repetiu de ano. 0 “Nenhuma”; 1 “Uma vez”; 2 “Duas vezes”; 3 “três ou mais”.
24.bairro	Categ	Bairros	24.computador	Int.	Computador em casa 0 “Não”; 1 “Sim, sem acesso à internet”; 2 “Sim, com acesso à internet”.
25.cep	Int.	Número do CEP, 8 dígitos	25.dever	Int.	Faz dever de casa. 0 “não porque meus professores não passam dever de casa”; 1 “não, meus professores passam dever, mas eu não faço”; 2 “sim, às vezes, faço os deveres de casa”; 3 “sim, sempre faço os deveres”.
26.end_numero	Alfan	Número endereço	26.estuda	Int.	Estudar com frequência: 0 “não, não tenho o costume de estudar”; 1 “estudo apenas quando tem prova”; 2 “faço as tarefas e estudo mais um pouco”; 3 “estudo independente de tarefas ou provas”.
			27.pais_dever	Int.	Pais pedem o dever de casa: 0 “não”; 1 “Às vezes”; 2 “quase sempre”; 3 “sempre”.
			28.pais_ajudam	Int.	Pais ajudam a fazer o dever: 0 “não faço dever de casa”; 1 “não”; 2 “sim, sempre”.
			29.atenção	Int.	Os alunos prestam atenção ao que o professor: 0 “não”; 1 “em algumas aulas”; 2 “na maioria das aulas”; 3 “sempre”.
			30.barulho	Int.	Há barulho em sala de aula: 0 “não”; 1 “em algumas aulas”; 2 “na maioria das aulas”; 3 “sempre”.
			31.falta_prof	Int.	O professor falta às aulas: 0 “nunca”; 1 “em algumas aulas”; 2 “na maioria das aulas”; 3 “sempre”.
			32.educ_física	Int.	Na sua escola existe educação física: 0 “não”; 1 “sim, uma vez por mês”; 2 “sim, uma vez por semana”; 3 “sim, duas vezes por semana ou mais”.

Obs.: Significado as siglas para os tipos usados nessa tabela estão no Anexo F.

Fonte: Censo Educacional, 2010 e PROEB, 2010.

Tabela 9 - Variáveis no banco de dados de Escolas (nome, tipo e descrição)

Banco de Escolas		
<i>Variável</i>	<i>Tipo</i>	<i>Descrição</i>
<u>1.cod_entidade</u>	Int.	Código da entidade, 8 dígitos.
2.nome_escola	Alfab.	Nome da escola limite de 50 caracteres.
3.dependência	Categ.	“Estadual” ou “Municipal”
4.situacao	Categ.	“Ativa” ou “Paralisada”
5.endereco_esc	Alfab	Endereço da escola com limite de 32 caracteres
6.num_esc	Int	Código com número da escola, similar ao cod_escola
7.bairro_esc	Categ.	Bairro da escola
8.cep_esc	Int.	CEP da escola, 8 dígitos
9.telefone	Int.	Telefone da escola, 8 dígitos

Obs.: Significado as siglas para os tipos usados nessa tabela estão no Anexo F.

Fonte: Dados da SEE-MG.

5.1. Tratamento e Processamento para união dos Bancos de dados.

O objetivo da presente seção é descrever os procedimentos de compatibilização dos bancos do censo escolar e do PROEB. Para que se possa avaliar o efeito do local e moradia do aluno na proficiência e em outros fatores do aprendizado é preciso obter uma base que una essas informações que estão separadas nos três bancos. Para realizar a tarefa de união dos bancos é necessário obter variáveis-chave que codifiquem igualmente cada aluno presente nas duas fontes.

Infelizmente as bases do *Censo Escolar* e do PROEB não foram pensadas conjuntamente e há problemas de compatibilização entre os dois bancos. Não existe uma variável-chave exata que possa uni-los, o número do aluno disponível no *Censo Escolar* não é o mesmo número empregado no PROEB. Contudo, felizmente, as duas bases perguntam o nome completo dos alunos e empregando procedimentos de pareamento de edição de caracteres, pode-se fazer a compatibilização entre os dois bancos.

Como ressaltado na introdução do capítulo, a principal variável chave utilizada para a união dos bancos é a informação confidencial do nome completo dos alunos. A informação de nomes é obtida tanto no *Censo Escolar* quanto no PROEB, e de posse dessa informação é possível unir a base que possui os endereços com a que possui a proficiência e outras características dos alunos. A FIG. 11, a seguir, apresenta um esquema representando a união dos dois bancos:

Figura 11 - Representando a união dos bancos do Censo Escolar e do PROEB

	Banco do Censo Escolar		Banco do PROEB		Banco Processado
Variável-chave	<u>nome_aluno</u>		<u>nome_aluno</u>		<u>nome_aluno</u>
Variável-chave	<u>cod_entidade</u>	+	<u>cod_entidade</u>	=	<u>cod_entidade</u>
variáveis	endereço e características pessoais		proficiência, bens de consumo, características familiares		endereço, proficiência, características pessoais e familiares

Fonte: Elaboração do Autor.

A adoção do nome do aluno como variável chave requer alguns procedimentos para a compatibilização dos bancos. Nesse tipo de variável há grande ocorrência de nomes truncados, erros de digitação, abreviações, informações incompletas, variações fonéticas e diversas outras variantes.

Em primeiro lugar temos que o nome dos alunos na base do Censo Escolar está truncado em até 33 caracteres (incluindo espaços). É uma linha grande o suficiente para distinguir grande parte dos nomes dos alunos, mas alguns nomes maiores ficam truncados. A ocorrência de homônimos com nomes maiores do que 33 caracteres é rara (5 duplas em um universo de 29.256 alunos). O banco do PROEB possui limite de 54 caracteres para os nomes e nenhum homônimo com mais de 34 caracteres.

O principal problema que pode ocorrer com nomes homônimos é a troca de homônimos entre escolas, pois o código da escola (*cod_entidade*) é também utilizado como variável-chave para união dos bancos. Caso ocorresse mais de um homônimo em uma mesma escola, não haveria

como distinguir quem é quem na base de dados. Felizmente, não há no banco nenhuma ocorrência de homônimos estudando na mesma escola.

Como ressaltado na introdução do capítulo, o banco do *Censo Escolar* possui 29.256 alunos. O banco do censo possui informações completas e a única variável que possui perda de informação (*missings*) é *raça* com 17% de casos faltantes (4.904 alunos) e *bairro* com perda de menos de 1% dos dados (88 alunos sem bairro), restando 24.264 alunos.

O banco de dados do PROEB para o 5º ano possuía, originalmente, 29.656 alunos. Entretanto, 2.452 alunos não tinham informação de proficiência e outros 9 alunos não possuíam a informação do nome. Restaram 27.195 alunos com informação completa para essas duas variáveis importantes (TAB. 10).

A TAB. 10 abaixo apresenta o número de linhas em cada banco, a ocorrência de homônimos dentro deles e o número de caracteres máximo utilizado para as variáveis de nomes e de endereço.

Tabela 10 - Informações sobre os bancos de dados utilizados

	Número de linhas	Limite de Caracteres do nome do aluno	Homônimos	Limite de caracteres para endereço do aluno	Limite de caracteres para endereço da escola
Censo Escolar 5º ano	29.264	33	144	31	-
Proeb 5º ano	27.195	54	130	-	-
Banco de Escolas	382	-	-	-	32

Obs.: O sinal “-” informa que o dado não se aplica.

Fonte: Banco de dados do Censo Educacional, MEC/INEP 2010; PROEB, 2010; Banco de Escolas da Secretaria Estadual de Educação de Minas Gerais.

Ao realizar a fusão dos dois bancos por meio de um *matching* determinístico dos nomes dos alunos e código de escolas, o banco resultante possui 17.792 alunos. Considerando que o banco do PROEB, o menor, possui 27.195 alunos, isso resultaria em uma perda de 35% da informação.

É possível reduzir o tamanho dessa perda com a fusão dos bancos. O principal motivo dela são os nomes com caracteres ligeiramente diferentes entre os dois bancos. Um exemplo fictício: Imagine-se que uma aluna chamada “Mariana dos Santos Pereira” possui alta probabilidade de ser a “Marina dos Santos Pereira” que estuda na mesma escola e esta listada na outra base. No caso do *matching* determinístico, essas informações de alunos com alta probabilidade de serem a mesma pessoa foram perdidos. Erros de digitação podem resultar em perda de informação relevante nesse tipo de *matching* por caracteres.

Para contornar esse problema empregou-se outra metodologia derivada do trabalho de físico-matemático russo Vladimir Levenshtein (1966) que propôs a distância de Levenshtein para edição de caracteres e para identificar quão diferente é uma sequência de caracteres da outra. O princípio básico é medir quantas operações de edição (inserções, deleções e substituições) são necessárias para transformar uma palavra em outra. Por exemplo, no exemplo acima, a distância de edição do nome “Mariana” para “Marina” é apenas um, pois, para transformar “Mariana” em “Marina” basta apenas apagar o segundo ‘a’ (trocar o ‘a’ por nenhum caractere (ϕ)). Da mesma maneira, a distância de edição entre “Victor” e “Vitor” é apenas 1, dado que é preciso deletar o ‘c’. A distância entre “Nathalia” e “Natally” é 3, pois em “Nathalia” será preciso *deletar* o h, substituir o *i* por *l* e o *a* por *y*.

O máximo a que a distância de Levenshtein pode alcançar é o limite de caracteres da palavra para qual se busca o pareamento e a distância de Levenshtein fornece o menor custo de edição para se chegar de um resultado a outro (ver Levenshtein, 1966).

A distância de Levenshtein é largamente empregada para resolver problemas de *variáveis-chave* com erros de digitação ou com informação imperfeita. É também largamente empregada em situações de transmissão de informação com possível ocorrência de erros, na criptografia para a decodificação de códigos, na computação para comparar diferentes sequências numéricas, ou até mesmo para comparação de diferentes sequências de DNA.

Para reduzir a perda do *matching* pode se determinar um valor máximo para a distância de edição permitida. O meio utilizado para esse propósito foi o comando ‘*agrep*’ presente na linguagem

base do R-CRAN (linguagem usada para trabalhar nos bancos ora apresentados).⁸⁵ Neste comando há duas opções possíveis para escolha do critério de corte: escolher uma fração do comprimento total da sequência, por exemplo, digamos que a fração escolhida seja 10%, no caso do exemplo da “Mariana dos Santos Pereira” que possui 26 caracteres, o limite de 10% equivale a permitir a edição (deletar, inserir ou substituir) de 3 caracteres (por convenção, arredonda-se o valor para cima).

A vantagem de se usar uma fração da extensão total da string é manter fixa a proporção da ocorrência de erros. Outra possibilidade é a de manter fixo o número de operações de edição para qualquer nome, que foi a opção adotada para a tese. Quando o objetivo não é permitir uma margem de erro de digitação aleatório, mas sim um erro sistemático advindo de nomes que podem ser escritos de mais de uma forma, não é necessário supor que a proporção de erros permanece constante, mas sim possibilitar liberdade para edição de fonemas e grafias passíveis de mais de uma representação. Dessa forma, escolheu-se um total de 9 operações de edição dispostas em 5 inserções, 3 remoções (*deletes*) e 3 substituições. Chegou-se a esses valores pela experimentação dos *matchings*.

A TAB. 11 apresenta o resultado para cada uma dessas escolhas de pareamento dos nomes. As linhas apresentam os métodos, a coluna ‘Original’ mostra o número de alunos disponível no banco do PROEB e a coluna ‘Após o Match’ mostra o resultado de quantos alunos tiveram os nomes pareados. A penúltima coluna (perda resultante) mostra em porcentagem qual é a perda da amostra no pareamento e a última coluna mostra o tempo total da operação no R em segundos.⁸⁶

Ocorre um *trade-off* entre aumentar o limite permitido para os erros de edição e a perda de dados resultante, quanto maior o limite de erros permitido, menor será a perda, mas pode se estar incluindo *matchings* não muito precisos no banco de dados. A última linha, que possui perda de 15,6%, foi a opção empregada.

⁸⁵ O R é uma linguagem voltada para aplicação e desenvolvimento de estatística com os mais variados propósitos, o sítio mantenedor se encontra na seguinte URL: <http://www.R-project.org/>. O R foi empregado para maior parte da análise realizado neste trabalho. Para maiores detalhes sobre o comando ‘agrep’ recomenda-se a página de *ajuda*: <http://127.0.0.1:22405/library/base/html/agrep.html> (último acesso 19/02/2013).

⁸⁶ A versão do R é 12.15.2, o Sistema Operacional é Windows 7. Demais especificações do CPU: processador Intel Core, I5 (3,20 GHz) com 12 GB de memória RAM.

Tabela 11 - Informações após a fusão dos bancos de dados por nível de corte do agreg

Banco do 5º ano	Original	Após o Match	Perda Resultante	Tempo computacional
Match determinístico	27.195	17.792	34,6%	0,45s
Limite de 5%		22.891	15,8%	8.604s
Limite de 10%		22.959	15,6%	10.260s
Limite de 15%		19.700	27,6%	12.384s
Limite de 20%		16.310	40,0%	14.328s
Limite de 9 edições*		22.964	15,6%	10.332s

* Limite utilizado.

Fonte: Dados computacionais obtidos da fusão dos bancos do Censo Educacional, MEC/INEP 2010; e do PROEB, 2010.

Outra técnica de pareamento dos dados envolve o reconhecimento por fonemas. Esse método está presente no pacote RecordLinkage do R.⁸⁷ Como, em geral, os pacotes são desenvolvidos para fonemas em inglês ou outras línguas, esse procedimento resultou em perdas muito grandes para grandes amostras, revelando resultados piores do que a junção pelo critério de exatidão. Além disso, Zobel & Dart (1995) mostram que o número de pareamentos obtidos pelo casamento fonético é, no melhor dos casos, apenas igual ao pareamento exato de strings (portanto, o pareamento por edição de caracteres apresenta um limite inferior para a perda de dados nos *matchings*).⁸⁸

Após esta fusão resta ainda mais um passo importante que é unir o banco de escolas. Porém, como esse banco possui a mesma *variável-chave* (*cod_escola*) também presente no *Censo Escolar* e no banco do PROEB, não foi difícil a união dos bancos por meio desta chave, resultando em nenhuma perda e um banco final com endereços dos alunos, das escolas e com proficiência em Língua Portuguesa. Ao fim dessa fusão, por conta da perda de alunos, o banco final permaneceu com **296** escolas (86 escolas a menos que o total de BH).

⁸⁷ As referências sobre este pacote se encontram na página:

<http://cran.r-project.org/web/packages/RecordLinkage/index.html> (último acesso, 19/02/2013).

⁸⁸ O pareamento fonético pode ter maior utilidade em bancos de dados mais incompletos ou imprecisos. Por exemplo, alguns caracteres diferentes possuem muitas vezes um mesmo fonema para determinada língua, tal como muitas palavras com 'ç', 's' e 'ss' do português, ou situações em que o 's' possui o som de 'z'. Nesses casos o *matching* determinístico acrescentará um custo de edição, e dependendo do nome, uma perda desnecessária. Portanto, pode ser interessante ter algoritmos *soundex* (como são chamados os códigos fonéticos de Newcombe et. al. 1959) para o português para determinados usos.

No banco de escolas, além das variáveis descritas na TAB. 9, foi também acrescentado o Nível Socioeconômico (NSE) das escolas de Belo Horizonte calculado por Alves & Soares (2012). O nível socioeconômico calculado pelos autores foi elaborado pelo método de Teoria de Resposta ao Item (TRI) utilizando variáveis presentes nos questionários aplicados aos alunos no SAEB e Prova Brasil, variáveis de consumo de bens duráveis, renda familiar, escolaridade e ocupação dos pais. O NSE utilizado para a escola é, portanto, a média dos NSEs dos alunos que estudam na escola e foi padronizado entre 0 e 10 e dividido nas categorias presentes na TAB. 12.

Tabela 12 - Categorias do Nível Socioeconômico das escolas (NSE)

<i>Valor do NSE</i>	<i>Nome</i>
$NSE < 3,48$	Mais Baixo
$3,48 \leq NSE < 4,10$	Baixo
$4,10 \leq NSE < 4,71$	Médio-Baixo
$4,71 \leq NSE < 5,35$	Médio
$5,35 \leq NSE < 6,18$	Médio-Alto
$6,18 \leq NSE < 7,18$	Alto
$7,18 \leq NSE$	Mais Alto

Fonte: Índice NSE calculado por Alves & Soares (2010).

Na comparação da proficiência entre o banco processado e o banco inicial completo, as estatísticas básicas de medição (média, mediana, quartis, variância, desvio padrão, além de alguns outros indicadores gráficos) não destoaram estatisticamente, ou seja, não há indícios de seletividade na amostra que resultou da fusão.

Para o uso completo dos dados de endereços é importante georreferenciar essa informação. Por meio dos dados georreferenciados é possível descobrir a que distância o aluno está de cada escola. Essa é uma informação valiosa para o banco de dados, pois será uma das principais variáveis objeto de análise e é crucial para a determinação das preferências. No entanto, a transformação de um endereço alfanumérico em um ponto georreferenciado no espaço exige muitos procedimentos que serão elucidados na seção a seguir.

5.2. Procedimentos para o Georreferenciamento do Banco

Os endereços usados cotidianamente podem ser entendidos como um processo heurístico que objetiva dar a exata localização de onde uma pessoa mora, estuda ou trabalha. Permite saber onde está localizado determinado edifício, onde se presta um serviço de atendimento ao público, e onde estão localizados os equipamentos e demais serviços no espaço urbano. O processo de deslocamento das pessoas nas cidades depende dessa comunicação de endereços.

Davis Jr. & Fonseca (2007) desenvolveram trabalhos que auxiliam na codificação de endereços e no georreferenciamento. Para que se possa realizar isso computacionalmente é preciso compreender qual é o padrão de informação de endereços, e esse padrão varia de acordo com os costumes e as normas de cada país. No Brasil, e no banco de dados com o endereço dos alunos, a informação de endereços costuma a ocorrer na seguinte ordem apresentada na TAB. 13:

Tabela 13 - Hierarquia de comunicação de endereços comumente utilizada no Brasil

<i>Tipo</i>	<i>Nome</i>	<i>Número do Edifício</i>	<i>Bairro</i>	<i>Cidade</i>	<i>Estado</i>	<i>País</i>	<i>Código Postal</i>

Fonte: Elaboração do Autor.

Um típico endereço de Belo Horizonte, como o Palácio das Artes, pode aparecer como: “Av. Afonso Pena, 1537, Centro. Belo Horizonte, MG. Brasil CEP: 30130-004”. Existe uma hierarquia de endereços apropriada para a divulgação de uma informação no espaço, em geral, os endereços usados cotidianamente assumem uma característica de posições relativas. A hierarquia segue um padrão de identificação que facilita o processo de referência.

Entretanto, o fornecimento de endereços pelas pessoas podem seguir outros procedimentos, no caso do exemplo mencionado, trata-se de um edifício público conhecido, sendo que em alguns casos, omite-se o número, bairro ou mesmo outras informações. Esse é um padrão pelo qual humanos conseguem identificar e localizar no espaço o “Palácio das Artes”, mas que torna difícil o processamento e codificação dessas informações para uma linguagem computacional.

Outro ponto que se acrescenta é que muitas vezes não é necessário um limite muito preciso de algumas informações para que se localize um endereço, uma dessas informações que as pessoas tipicamente possuem de maneira mais imprecisa são os limites dos bairros, por exemplo. Como o “Palácio das Artes” se situa quase no vértice de um polígono que visualmente limitaria o Centro poderia se ter uma informação imprecisa de que o palácio se situa no bairro Funcionários, no Boa Viagem ou no Santa Efigênia e ainda assim, mesmo com essa imprecisão, pode-se localizar o ponto geográfico do edifício.⁸⁹

O processo de georreferenciamento de um banco de endereços é, de certa maneira, similar ao pareamento de *string* enunciado na seção anterior. Tendo uma base de endereços já georreferenciada, onde cada endereço corresponde a um ponto no espaço (ou a um segmento ou polígono em alguns casos), o que precisa ser realizado é uma compatibilização entre essa base e a base de endereços de alunos e escolas fornecidos. De posse das informações textuais de endereços é possível fazer um pareamento entre os dois *strings*. Ocorre que em caso de bancos de endereços, é preciso de alguns passos a mais, Davis Jr. & Fonseca (2007) apresentam três fases de organização para o georreferenciamento de um banco:

1-Fase: etapa de parsing, essa etapa consiste em investigar qual é o ordenamento de dados de um banco, como ele está hierarquizado e identificar quais informações o banco possui, no caso, de endereços de Belo Horizonte, como destacado, a ordem seguirá tipicamente aquela apresentada na TAB. 13;

2-Fase: etapa de matching. Após o *parsing*, com o banco já padronizado para a informação de endereços, resta fazer o pareamento entre o banco de endereços fornecido e o banco de endereços

⁸⁹ Sobre bairros em Belo Horizonte, ninguém tem ao certo uma medida exata da precisão dos limites e da existência de vários bairros em Belo Horizonte. O anexo G apresenta um mapa oficial da prefeitura de Belo Horizonte com a divisão de bairros disponibilizada no sítio da prefeitura: http://www.pbh.gov.br/smpl/HTA_M007.pdf (último acesso 20/01/2013). Essas múltiplas definições ocorrem por conta das mudanças do espaço urbano, sendo sempre preciso haver uma adaptação entre a maneira como as pessoas entendem uma região e as modificações de ruas, casas, edifícios, que ocorrem naturalmente com a expansão da cidade. Em trabalho recente, Umbelino (2012) estudou como ocorre a expansão da malha urbana. Sobre bairros e endereços um caso interessante é o bairro de Boa Viagem, chamado assim pela proximidade com a Igreja da Boa Viagem, trata-se de um bairro pequeno e frequentemente suprimido na maioria dos endereços informados que o tomam como Funcionários, no Google *maps*, portal de informação geográficas do Google, esse bairro não está delimitado (maps.google.com.br, última pesquisa 14/05/2013).

georreferenciados. Para um maior aproveitamento dessa etapa, o banco de endereços georreferenciado precisa ser o mais completo e preciso possível para garantir a qualidade das informações (adiante serão apresentados mais detalhes desse banco para Belo Horizonte). A etapa de *matching* processa o banco usando a distância de Levenshtein (1966), já abordada na seção anterior;

3-Fase: A terceira fase é a *locating*, essa fase permite estipular a precisão e o tipo de informação georreferenciada que se pode usar. Por exemplo, se o endereço fornecido pode ser identificado como um ponto já georreferenciado. Caso positivo, pode-se vincular o endereço àquele ponto já conhecido. Em alguns casos, ocorrerá localização precisa de uma rua, mas a não identificação do número do edifício (ou casa) no banco georreferenciado, para esses casos adotou-se, quando possível, uma identificação em relação à numeração mais próxima. Em outros, casos, com menor precisão de informação, será possível localizar apenas a quadra ou o bairro (polígono), se não houver nenhuma referência mais precisa sobre o endereço, adota-se o centroide do polígono localizado.⁹⁰

Para o município de Belo Horizonte, Martins, Davis Jr. & Fonseca (2012) desenvolveram um programa que, por meio de uma base de endereços, processa as fases acima estipuladas e cria um banco resultante com endereços georreferenciados. Esse programa também realiza uma medida de precisão (GCI, *Geocoding Certainty Indicator*) proposto por Davis Jr. & Fonseca (2007). O GCI, que varia de 0 a 1, é uma medida da precisão da geocodificação obtida por meio das três etapas acima. Se todas as etapas acima resultaram em sucesso, o índice assume o valor de 1, pois $GCI = 1 \times 1 \times 1 = 1$, em que há um subíndice para cada uma das fases do processamento.

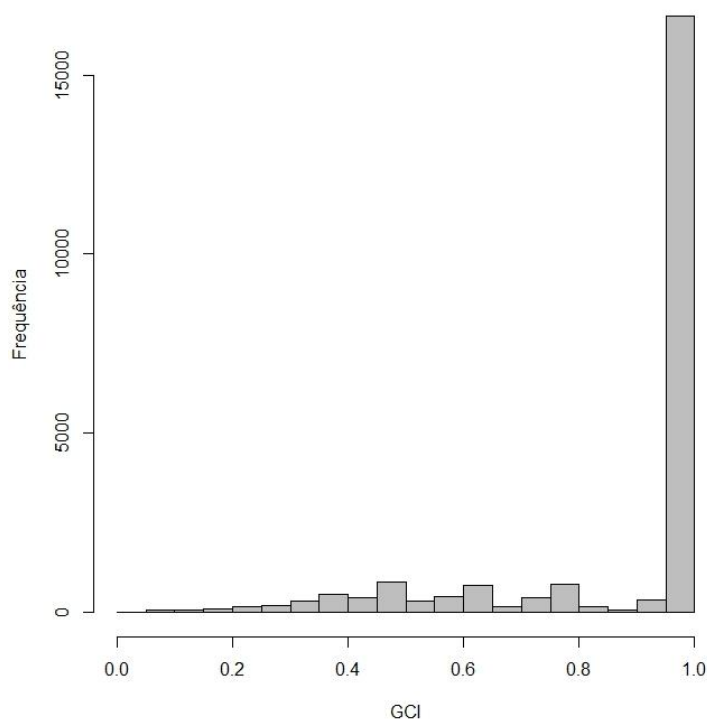
Pode-se obter um nível de corte do GCI para garantir a qualidade da informação no banco. Um GCI alto de 0,95 obteve uma localização com quase 100% de precisão, pode ter ocorrido algum erro na grafia do nome da rua ou na indicação do bairro, mas o processamento conseguiu verificar que aquele endereço, com grande chance, está correto. Uma questão importante, então, trata de delimitar um valor de corte (*threshold*) para o valor do GCI, quanto menor o GCI

⁹⁰ Os passos do georreferenciamento estão aqui enunciados em linhas gerais dado que não é objetivo fornecer contribuições teóricas para esta etapa, para maiores detalhes ver Davis Jr. & Fonseca (2007, p. 116-122).

adotado menor a perda de informação no banco, porém, maior a imprecisão de algumas informações. Do contrário, valores de corte muito altos para o GCI, igual a 1,00, por exemplo, podem propor cortes muito rigorosos, sendo que se consegue uma qualidade boa da informação para GCI menores, mas próximos daquele valor.

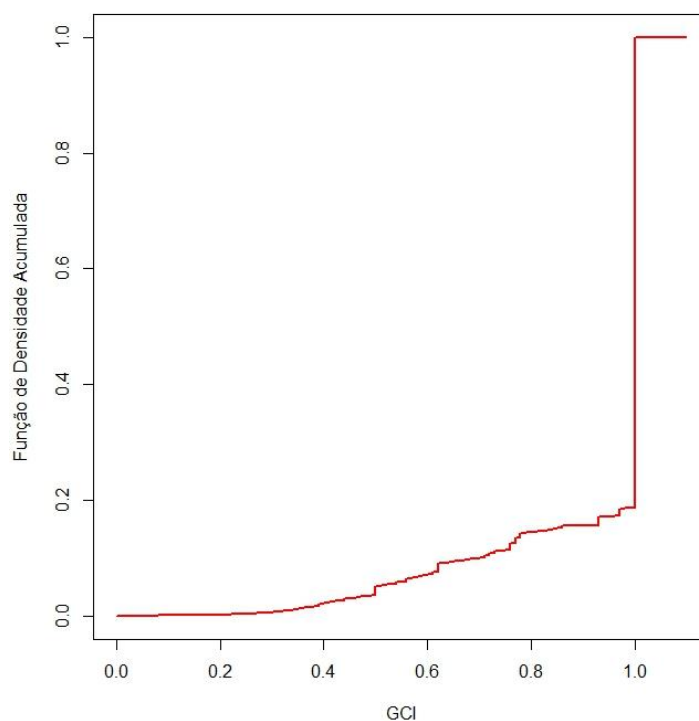
Os GRAFs 3 e 4 abaixo apresentam um histograma do GCI e sua função acumulada. Nota-se que tanto pelo histograma, quanto pela função de densidade acumulada, há uma alta frequência de GCI com valor 1, isso mostra que a maioria (71%) dos alunos está com um endereço para o qual se conseguiu precisão máxima no georreferenciamento.

Gráfico 3 - Histograma do Geocoding Certainty Indicator (GCI)



Fonte: Banco de dados geoprocessado (2010).

Gráfico 4 - Função de Densidade Acumulada do Indicador (GCI)



Fonte: Banco de dados geoprocessado (2010).

O procedimento de georreferenciamento contou com um número de alunos (22.782) um pouco menor do que o obtido após a fusão dos bancos descrita em 5.1, pois 172 alunos não tiveram o endereço identificado ou possuíam variáveis de endereço completamente vazias. Isso ocasionou nova perda no banco, embora muito pequena.

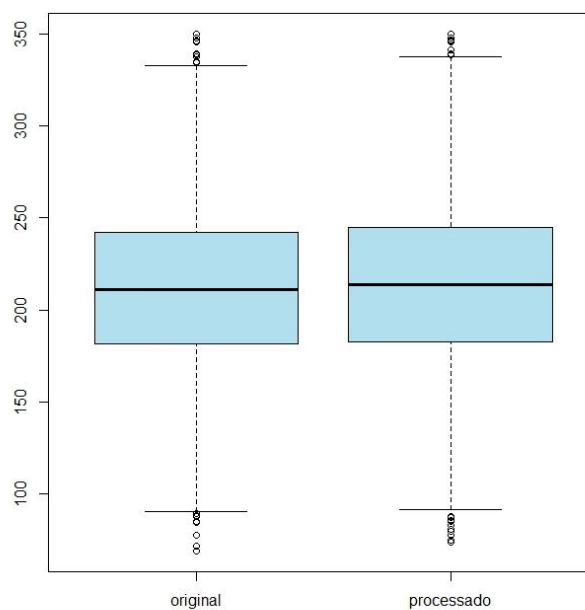
A TAB. 14 abaixo apresenta a relação entre os valores de corte de GCI e a perda de informação do banco. O valor do GCI adotado foi 1,00, por conta de uma necessidade de obter um banco para o qual se possuía informações precisas dos endereços. Optou-se pelo banco de **16.354** alunos com um endereço 100% preciso. Embora essa escolha reduza significativamente o tamanho da amostra em relação ao banco inicial, ela garante ainda grande representatividade dentro do universo de alunos de BH.

Tabela 14 - Valores de corte do GCI, número de alunos e perda resultante

Valor de corte do GCI	Número de alunos após o corte	Perda Total Resultante
0,00	22.781	0%
0,45	20.992	7,8%
0,65	18.642	18,1%
0,85	17.105	22,7%
0,95	16.651	26,9%
1,00	16.354	28,1%

Fonte: Banco de dados geoprocessado (2010).

O GRAF. 5 apresenta o *boxplot* da variável proficiência para o banco original e o banco final processado que será utilizado daqui em diante, a diferença de médias está ainda dentro do intervalo de confiança, sendo um pouco maior para o banco final.

Gráfico 5 - Boxplot da proficiência entre o banco original e o banco final geoprocessado

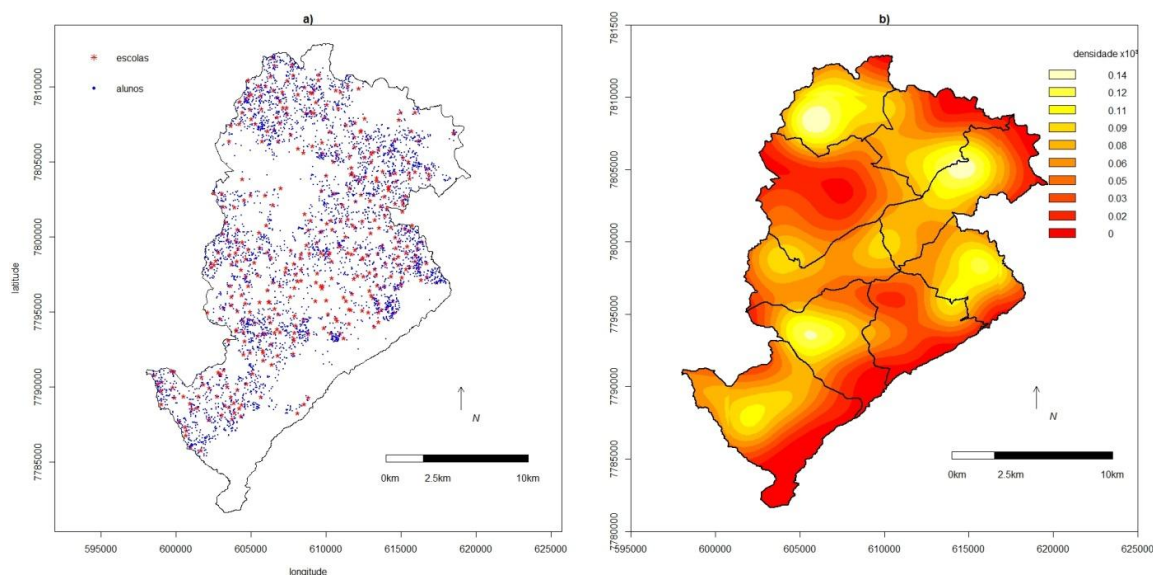
Obs.: Nota-se que não há diferença significativas de médias da proficiência entre os dois bancos.

Fonte: Banco de dados geoprocessado (2010).

Ao final da etapa de geoprocessamento, obtêm-se um sistema completo de informação de onde residem os alunos e onde estão as escolas no espaço urbano. Com isso, calculou-se a distância euclidiana,⁹¹ em Km, da residência até a escola para cada aluno fornecendo uma matriz de distâncias, importante para o estabelecimento de possíveis preferências teóricas. Os dados de distâncias são usados na análise descritiva do próximo capítulo e serão abordados em detalhes.

O mapa 1-a abaixo mostra os alunos e as escolas dentro do limite de Belo Horizonte e o mapa 2-b mostra a densidade de kernel para alunos em um raio de 2 Km, as cores escuras (para o vermelho escuro) indicam uma baixa densidade de alunos e as cores claras (do amarelo ao branco) indicam uma alta densidade (esse mapa possui as divisões administrativas da cidade).

Mapa 1 - a) Alunos e Escolas de Belo Horizonte representados por pontos e b) Mapa de Densidade aluno por Km²



Fonte: Banco de dados geoprocessado (2010).

Obs: Para melhor visualização, o mapa a) apresenta uma amostra de 30% dos alunos.

⁹¹ O programa desenvolvido por Martins, Davis Jr. & Fonseca (2012) estabelece as coordenadas do ponto geográfico em termos de graus de latitude e longitude, que podem ser convertidas na distância em metros. Em geral, para mapas de pequenas áreas, tais como de uma cidade, coordenadas planas são ideais para cálculo de distâncias e demais tarefas de georreferenciamento. No caso, as coordenadas de latitude e longitude foram transformadas em coordenadas UTM (*Universal Transverse Mercator*) um modelo simples de conversão estabelece: $x = R(\theta_{Lon} - 0)$ e $y = R \left(\ln \left(\tan \left(\frac{\pi}{4} + \frac{\theta_{Lat}}{2} \right) \right) \right)$, em que x e y são longitude e latitude em metros do UTM, R é o raio da Terra em metros, θ_{Lon} e θ_{Lat} são os graus de latitude e longitude em radianos. Em verdade, o sistema utilizado para Belo Horizonte é o UTM-23S-WGS84 que possui alguns refinamentos para a conta acima. Por meio dos pontos (x,y) da residência do aluno e da escola, facilmente se consegue a distância euclidiana em metros.

6. ANÁLISE DESCRITIVA DO BANCO E ALOCAÇÃO OBSERVADA DE ALUNOS E ESCOLAS

O banco de dados processado com 16.354 alunos foi dividido em quatro níveis de análise: **alunos, escolas, bairros e regiões**. Além desses níveis de análise, foi construída uma matriz de distâncias dos alunos até as escolas, mensurando a distância em quilômetros da casa de cada aluno para cada escola do banco. Tais informações permitem uma análise “intrabairro”, que enfoca deslocamentos de estudantes dentro de um mesmo bairro, e “entrebairros”, deslocamento de estudantes entre bairros distintos. Do mesmo modo, realizou-se uma análise “intrarregião” e “entrerregiões”, captando a movimentação de alunos dentro de uma mesma região e entre regiões distintas.

Um conceito importante que se insere na análise descritiva é o conceito de **aluno-migrante**, estudante que se desloca “intrabairro” ou “entrebairros” para ir à escola. Esse conceito é diferente do conceito de migrante da demografia populacional. Em primeiro lugar, porque o aluno não está mudando de residência, apenas se deslocando dentro da cidade, em segundo, no sistema de cadastro, o aluno não tem liberdade irrestrita para escolher em qual escola estudar, dado que as matrículas são sugeridas pelo sistema. Apesar disso, a mobilidade existe e o conceito de migração de bairros e regiões ajuda na interpretação do problema de alocação de escolas e estudantes investigado neste trabalho. O conceito de aluno-migrante será formalmente definido mais adiante na seção 6.2.

O presente capítulo está dividido em quatro seções: a seção 6.1, de análise exploratória do banco de dados, e que apresenta a análise descritiva dos dados e expõe as relações entre proficiência e atributos individuais e familiares, apresentando mapas e características espaciais do banco. A seção 6.2, de características de mobilidade e alocação observada, apresenta informações de mobilidade dos alunos, com destaque para a distância euclidiana entre residência do aluno e a escola, desenvolvendo comparações entre grupos de alunos **migrantes** e **não migrantes**. Ainda na seção 6.2, define-se o conceito de “aluno-migrante” e se elabora modelos para o tratamento desta informação. A seção 6.3, de modelos hierárquicos e decomposição de fatores, apresenta um modelo hierárquico completo de 4 níveis: alunos, escolas, bairros e regiões. A seção 6.4 reúne

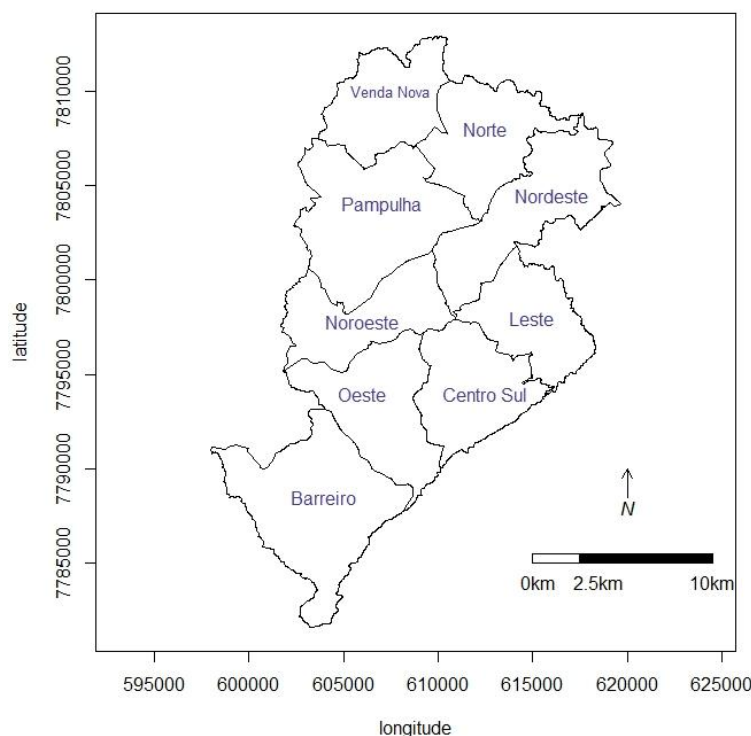
informações das seções anteriores, estabelecendo parâmetros para as preferências de pais e alunos. E propõe os efeitos para as diferentes possibilidades de alocação que serão simuladas no capítulo 7.

6.1. Análise exploratória do banco de dados

6.1.1. Regiões Administrativas

Um ponto importante de se esclarecer desde o início deste capítulo diz respeito à divisão administrativa do município. Belo Horizonte se divide em nove regiões administrativas regulamentadas na lei municipal 10.231/11. Regiões administrativas do Barreiro, Centro-Sul, Leste, Nordeste, Noroeste, Norte, Oeste, Pampulha e Venda Nova (representadas no mapa 2).

Mapa 2 - Regiões administrativas de Belo Horizonte



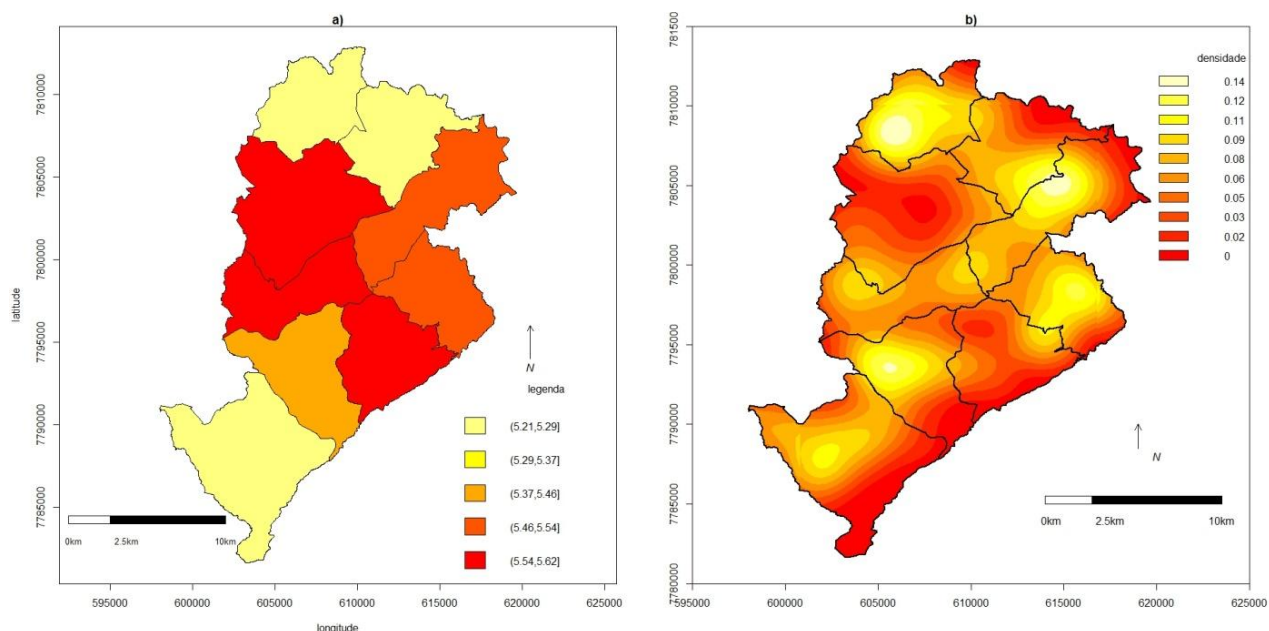
Fonte: Dados administrativos da PBH (2012).

Sabe-se que os alunos não se distribuem igualmente no espaço urbano, a distribuição espacial de estudantes corresponde, em grande medida, às características de desenvolvimento geográfico e econômico da cidade. Há uma relação negativa entre nível de renda familiar e a demanda por

educação pública. Quanto maior a renda familiar, menor a probabilidade de a família demandar educação pública. Dessa forma, a maior ocorrência de alunos matriculados em escolas públicas está localizada em regiões de grande densidade populacional e menor nível de renda. Isso ajuda a explicar o fato de a região Centro-Sul possuir baixo número de alunos, apesar de ser uma região de alta densidade populacional (ver mapa 3b).

O mapa 3-a apresenta a média do nível sócio econômico escolar (NSE) por regiões e o mapa 3-b apresenta o *kernel* para a densidade de alunos em uma área móvel de raio igual a 2 km. Nota-se que as três regionais com maior média de NSE: Centro-Sul, Noroeste e Pampulha, apresentam níveis de densidade de alunos mais baixos. A Pampulha possui a lagoa no centro da região, o aeroporto à Leste, e o campus da UFMG no sudeste, são grandes áreas não habitáveis, mas ainda assim, a densidade de alunos nos bairros do entorno da lagoa é mais baixa do que nas outras regiões. A região Noroeste possui alguns pontos de maior densidade associados a bairros de menor NSE dentro da região.

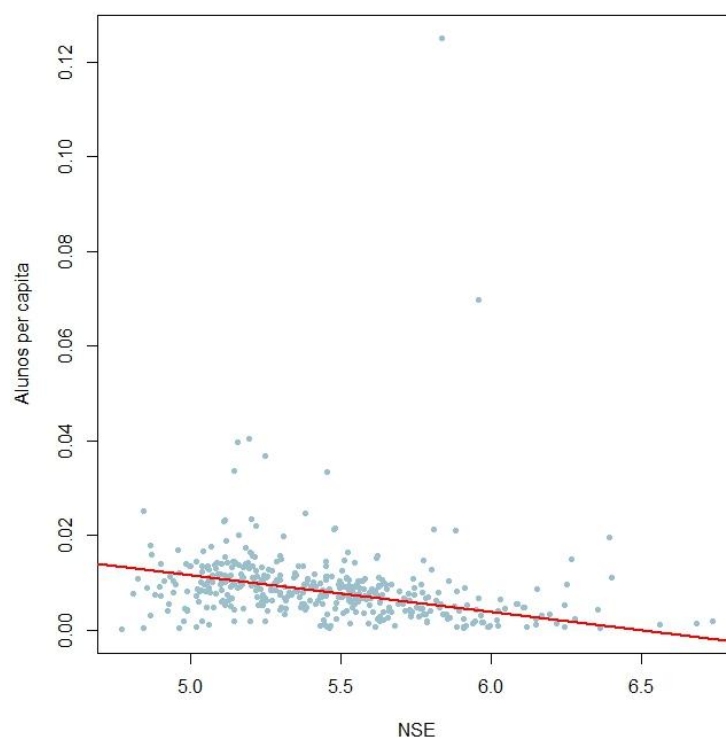
Mapa 3 - a) Nível Socioeconômico por Região e b) Mapa da densidade de alunos por km²



Fonte: Dados de NSE de Alves & Soares (2012) e banco de dados geoprocessado (2010).

O GRAF. 6 apresenta essa mesma correlação entre NSE e número de estudantes, apresentando os dados por bairros. Na obtenção do número de estudantes *per capita*, o número de alunos foi dividido pela população do bairro dada pelo Censo Demográfico de 2010, a divisão da população por bairros foi calculada pela PBH.⁹² Quanto maior o NSE escolar do bairro, menor o número de alunos *per capita*. Isso evidencia as preferências dos pais e uma racionalidade econômica na escolha do ensino público. Pais com renda familiar mais baixa, mesmo que prefiram as escolas privadas, não possuem condições econômicas para matricular seus filhos em tais escolas.⁹³

Gráfico 6 - Relação entre NSE escolar do bairro e número de alunos 5º ano per capita



Fonte: Banco de dados geoprocessado (2010).

⁹² Informações disponíveis em:

http://portalpbh.pbh.gov.br/pbh/ecp/comunidade.do?evento=portlet&pIdPlc=ecpTaxonomiaMenuPortal&app=estatisticaseindicadores&tax=26720&lang=pt_BR&pg=7742&taxp=0& (último acesso em 13/03/2013).

⁹³ Na seção 6.4 abordaremos possíveis determinantes para as preferências dos pais. Uma ideia que se pode apontar desde já, é que escolas privadas podem ser preferências não factíveis para famílias de baixa renda, pois estão fora da restrição orçamentária familiar. Além disso, entre escolas públicas não homogêneas, pais podem ter preferências distintas pelas escolas, tentando ao máximo conseguir matrículas em escolas consideradas de boa qualidade.

6.1.2. Divisão de Bairros

Ao contrário da divisão de regiões administrativas, regulada por lei e que pouco se altera com o passar dos anos, a divisão de bairros de Belo Horizonte é mais imprecisa, uma vez que diversas divisões já foram propostas e os nomes de bairros seguem uma dinâmica desregrada. Nomes e limites de alguns bairros mudam com o tempo e não há uma listagem oficial de bairros ou, na melhor das hipóteses, há mais de uma listagem oficial. Mesmo na PBH não há consenso sobre o número e os limites dos bairros da cidade, sendo que a própria prefeitura realiza diferentes listagens.⁹⁴ Mesmo com tais dificuldades, ainda assim é possível contar com certa estabilidade e confiabilidade na informação disponibilizada, sendo necessários alguns cuidados para a compatibilização desta informação.

O banco de bairros utilizado está relacionado às etapas de geocodificação descritas no capítulo 5 e possui uma listagem de 487 bairros. Esse é um número bastante elevado de bairros, e há casos de pequenas vilas com menos de 1km², geralmente incrustadas em bairros maiores. A TAB. 15 apresenta o número de bairros pela compatibilização de endereços. Em 43 bairros não se conseguiu encontrar nenhum aluno de 5º ano, seja por informação incompleta dos endereços alunos, seja por erro de correspondência na etapa de georreferenciamento.

Tabela 15 - Presença de bairros no banco de dados

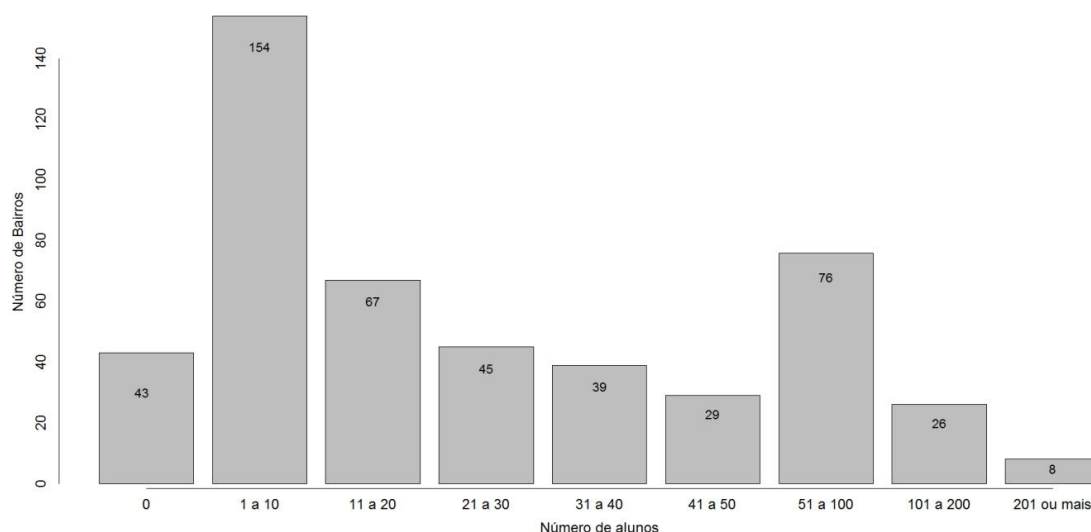
<i>Descrição</i>	<i>Número</i>	<i>Porcentagem</i>
Número de bairros para os quais não se obteve compatibilização.	43	9%
Número de bairros para os quais se obteve a compatibilização.	444	91%
<i>Total</i>	487	100%

Fonte: Banco de dados geoprocessado (2010).

Excetuando-se os 43 bairros para os quais não se encontrou nenhum estudante, a maior parte dos bairros de BH, 154, possui de 1 a 10 estudantes do 5º ano. Apenas oito bairros possuem mais de 200 estudantes. O GRAF. 7 apresenta esses dados.

⁹⁴ No Anexo G, há um mapa oficial da PBH com os bairros e uma listagem dos bairros por regional.

Gráfico 7 - Gráfico de Barras do número de bairros por faixas do número de alunos



Fonte: Banco de dados geoprocessado (2010).

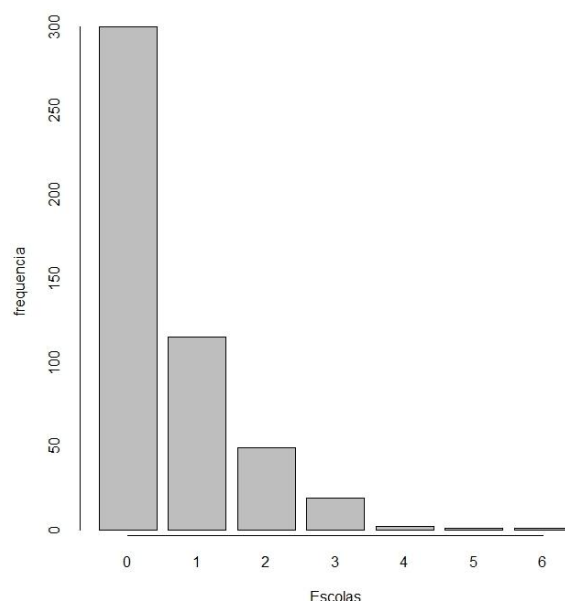
Como Belo Horizonte possui mais bairros do que escolas, os bairros sem escolas são comuns. No banco processado, há 300 pequenos bairros sem nenhuma escola (esse número de bairros sem escolas deve ser um pouco menor, é preciso lembrar que o banco aqui utilizado possui 86 escolas a menos que o total de BH. Perda ocorrida por conta do processo de compatibilização). Há 115 bairros com apenas uma escola, mas, por outro lado, há bairros, como o Serra, que possui 6 escolas (ver GRAF. 8).

Em geral, a maioria dos bairros que não possuem uma escola é vizinho de algum bairro que possui escola, sendo que a distância média percorrida pelos discentes para se chegar à escola é pouco maior do que 1,5 km (ver seção 6.2).

Outra maneira de ver a carência de escolas em bairros da capital mineira envolve observar a relação alunos por escolas. Este indicador é dado por I_k/S_k , em que I_k é o número total de estudantes de 5º ano do bairro k e S_k é o número total de escolas públicas municipais e estaduais do mesmo bairro. Nos casos de indeterminação da divisão (0/0), o indicador foi convencionado

para ser igual a zero, e no caso de bairros que não possuem escolas a convenção adotada foi pelo número total de alunos do bairro ($I_k/1$).⁹⁵

Gráfico 8 - Frequência do número de Bairros por número de escolas



Fonte: Banco de dados geoprocessado (2010).

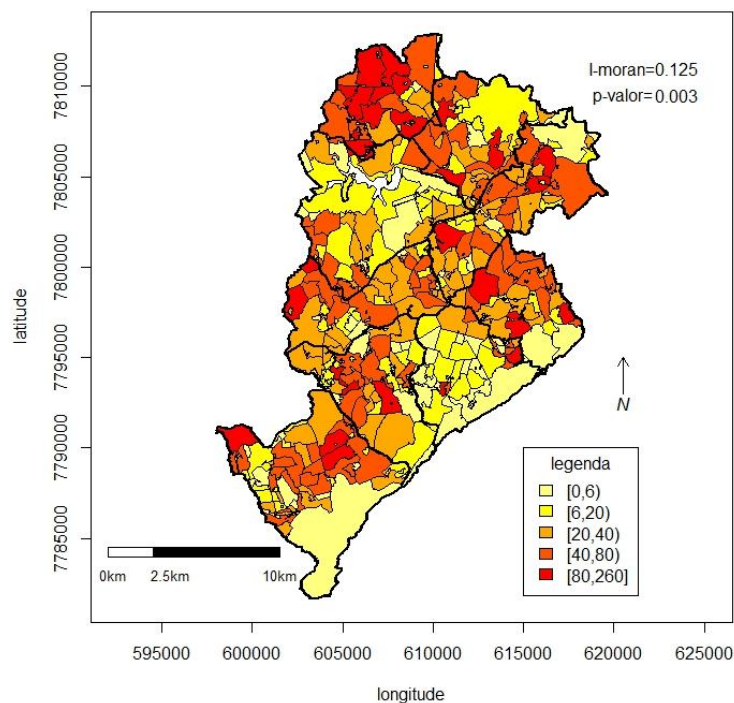
Pelo indicador proposto, as áreas mais claras no mapa 4 (amarelo claro e amarelo ouro) são aquelas melhor atendidas, o número de crianças de 5º ano do bairro não passa de 20 alunos para cada escola (assumindo-se o máximo de 45 alunos por série, isso seria o equivalente a uma única turma de 5º ano).

As áreas de cores escuras (laranja escuro e vermelho) demandam número maior de vagas. Realizou-se também um teste de I-Moran para detecção da correlação espacial, o valor do teste é significativamente diferente de zero (p-valor = 0,003) indicando correlação: bairros que demandam mais turmas apresentam correlação espacial positiva.⁹⁶

⁹⁵ Outra possibilidade de representar essa estatística pode ser pela inversa: S_k/I_k . Como existe um número de alunos maior que zero para quase todos os bairros, esse indicador requisitaria menor uso da convenção de divisão por zero. No entanto, o raciocínio é invertido, levando a uma interpretação inversa. Desse modo, optou-se pela forma direta, que fornece ideia do quantitativo de alunos demandantes por escola em cada bairro.

⁹⁶ Uma observação que se aplica a todos os mapas de bairros da tese: bairros sem informação de alunos, caso como, por exemplo, o Belvedere, que não possuía nenhum aluno matriculado na 5º ano no Censo Escolar 2010, foram imputados pela média dos vizinhos adjacentes nas 8 direções cardiais, N, L, O, S, NE, NO, SO, SE. Esse

Mapa 4 - Relação Aluno/Escola por bairros de BH, com resultado de I-Moran



Fonte: Banco de dados geoprocessado (2010).

As regiões menos atendidas podem fazer com que os alunos se desloquem mais e isso ocasiona impactos na proficiência que são mensurados mais detalhadamente em 6.2 e 6.3 (e simulados no capítulo 7). É natural que algum deslocamento de alunos ocorra, pois escolas precisam de escala para captação de um número ótimo de estudantes. No entanto, grandes distâncias são prejudiciais ao rendimento, existe uma distância ótima e ela depende de uma combinação de variáveis, do aluno, da escola e do entorno.

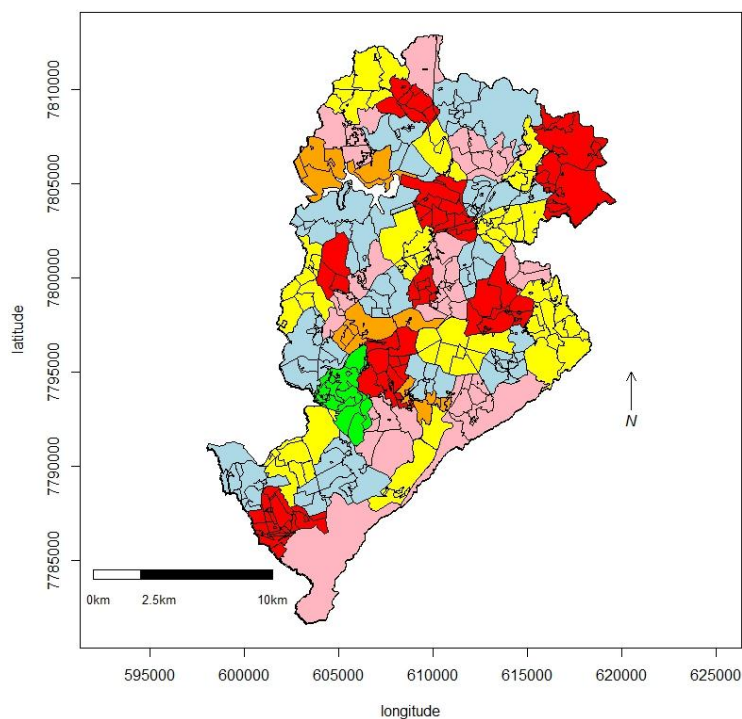
Visto que há bairros muito pequenos e com poucos alunos, a informação de bairros pode ficar muito fragmentada dificultando seu uso e comprometendo sua representatividade. Soares, Rigotti e Andrade (2008) solucionaram esse problema empregando *Unidades Espaciais Homogêneas* (UEH) que “agrupam áreas homogêneas contíguas, com perfil social e urbanístico semelhante”

procedimento de imputação foi realizado tendo em vista apenas melhoria do aspecto estético da apresentação dos mapas e não para as análises estatísticas para o qual informações faltantes não foram imputadas.

(p. 128).⁹⁷ No entanto, a base de UEHs não é compatível com a base de bairros e endereços que foi utilizada na etapa de georreferenciamento descrita no capítulo 5 (infelizmente a compatibilização entre esses dois bancos exigiria um projeto à parte).

Adotou-se a solução de agrupar bairros contíguos com características geográficas semelhantes. O critério para esse agrupamento, além da vizinhança, foi o número de estudantes no bairro, procurou-se agrupar bairros contíguos com número de estudantes menor do que 30. Ao final desse procedimento foram obtidos 41 grupos de bairros (ver o mapa 5 abaixo)⁹⁸. Tais grupos foram utilizados principalmente nos modelos hierárquicos, onde é preciso níveis de agregação maiores, o que seria impossível na amostra anterior, com bairros sem escolas e número não representativo de alunos.

Mapa 5 - Bairros agrupados por características geográficas e número de alunos



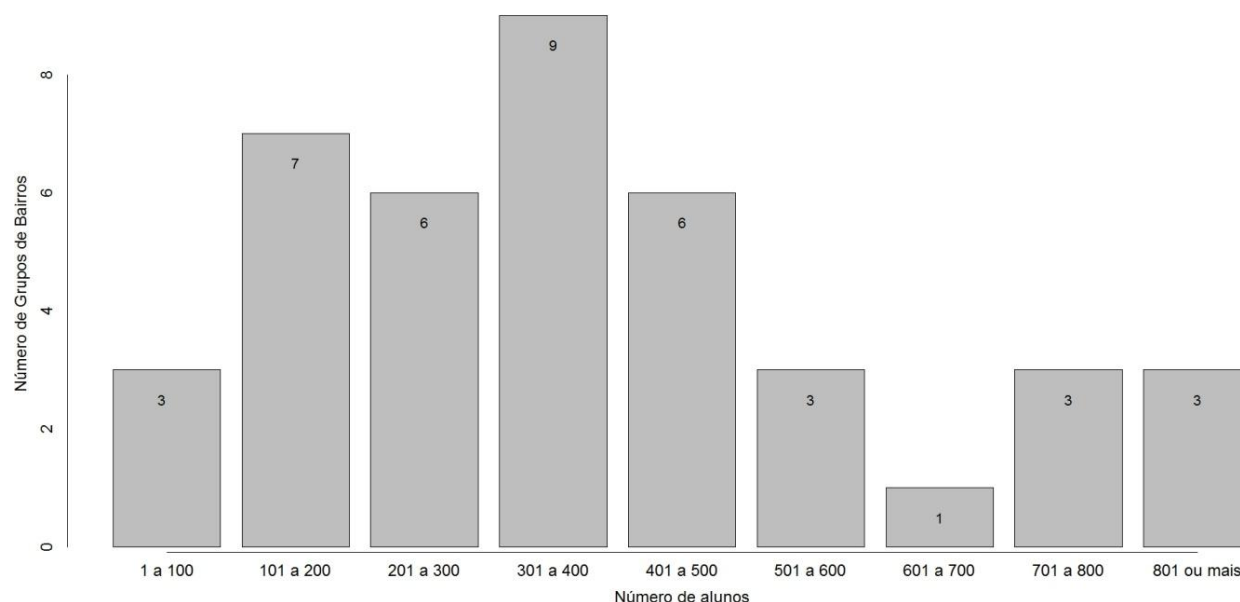
Fonte: Banco de dados geoprocessado (2010).

⁹⁷ O emprego das UEH não foi possível por conta da diferença da unidade de georreferenciamento utilizada. As UEHs foram construídas utilizando-se setores censitários, e foram realizadas para toda a Região metropolitana de Belo Horizonte. Ao final do processo de compatibilização, Soares, Rigotti e Andrade (2008) obtiveram 77 unidades espaciais em Belo Horizonte. No caso do georreferenciamento realizado no capítulo 5, o limite dos bairros é diferente dos limites censitários, exigindo um projeto de compatibilização entre o banco de UEHs e Bairros.

⁹⁸ A lista completa dos bairros pertencentes a cada grupo está no Anexo G.

A distribuição de alunos entre os bairros agrupados está apresentada no GRAF. 9 abaixo. O grupo de bairros com menor número de alunos é o Santo Antônio, com 53 alunos, seguido pelo o grupo Olhos d'Água (64) e Sion (68). Três agrupamentos de bairros possuem mais de 801 alunos: Baleia e Gameleira, ambos com 863, e Mantiqueira que tem 1043 estudantes.

Gráfico 9 - Frequência de bairros agrupados por número de alunos



Fonte: Banco de dados geoprocessado (2010).

A TAB. 16 apresenta os dados de alunos, vagas e escolas por grupos. O grupo de bairros que possui menos escolas é o Santo Antônio da região Centro-Sul, com apenas uma escola, e o grupo que possui mais escolas é o da Baleia, região Leste, com 13 escolas nos bairros que foram agrupados. O número de alunos que mora no grupo de bairros é diferente do número de alunos que estudam nele por conta da emigração e imigração, alguns bairros e regiões não possuem vagas suficientes para todos os alunos moradores, sendo preciso os alunos se deslocarem. Mesmo em bairros que possuem vagas suficientes para todos os seus moradores é comum observar-se tanto movimento de saída de alunos (emigração) quanto o de entrada (imigração), esse tópico será detalhadamente discutido na seção 6.2.

Tabela 16 - Grupo de bairros com número de alunos moradores vagas e Escolas

<i>Grupo de Bairros</i>	<i>Alunos*</i>	<i>Vagas</i>	<i>Escolas</i>
Aeroporto	280	261	6
Aparecida	265	242	6
Baleia	863	626	13
Barragem Santa Lúcia	326	137	3
Barreiro	303	348	7
Boa Vista	347	371	7
Bonsucesso	719	722	10
Braúnas	148	159	3
Buritis	268	227	4
Caiçaras	190	220	5
Capitão Eduardo	531	304	6
Carlos Prates	256	284	6
Castelo	127	251	4
Centro	169	715	10
Céu Azul	742	665	9
Florammar	494	470	10
Floresta	488	652	10
Gameleira	863	838	12
Glória	717	651	9
Independência	641	568	10
Jaqueline	417	386	6
Jardim São José	391	293	6
João Pinheiro	315	256	7
Lagoinha	494	356	9
Lindéia	539	569	10
Mantiqueira	1043	803	11
Maria Goretti	488	459	11
Olhos d'Água	64	140	3
Ouro Minas	366	415	6
Planalto	165	193	5
Prado	349	425	10
Santa Mônica	465	748	9
Santo Antônio	53	10	1
São Francisco	212	50	2
São Gabriel	563	638	11
São Luíz	325	468	10
Serra	620	464	11
Serra Verde	113	98	2
Sion	68	83	3
União	201	344	5
Venda Nova	366	445	8

Alunos*: Alunos que moram no grupo de bairros apresentado na linha.

Fonte: Banco de dados geoprocessado (2010).

6.1.3. Análise descritiva das variáveis

As variáveis do banco de dados processado podem ser interpretadas de acordo com os diversos níveis de análise mencionados na introdução deste capítulo. A variável resposta está associada ao aluno e pode ser agrupada também nos outros níveis: escolas, bairros (grupos de bairros) e regiões.

A TAB. 17 apresenta as variáveis do banco de acordo com o nível de análise. Além das variáveis já descritas no capítulo 5, o banco de dados georreferenciado permite analisar algumas informações espaciais antes não disponíveis nos bancos originais. A mais importante delas é a informação se o aluno estuda no mesmo bairro ou região onde mora, outra informação também importante é a distância da residência do aluno até a sua escola.

As variáveis da TAB. 17 com o asterisco foram obtidas por meio do banco de dados georreferenciado. A variável migrante indica se o aluno é migrante ou não. Há quatro determinações para a variável migrante: se o aluno migra dentro do próprio bairro (quando não vai para a escola mais próxima de casa, mas permanece ainda nas cercanias do bairro), se migra entre bairros, se migra entre grupo de bairros e, finalmente, se migra entre regiões. Essa mesma variável pode ser diferenciada entre alunos que saem (emigrantes) e alunos que entram (imigrantes).

A variável distância é a distância euclidiana em quilômetros da casa até a escola. As variáveis de imigrantes nos níveis de escolas e bairros são informações agregadas, elas fornecem a proporção de alunos imigrantes (no nível 1, a variável é igual a 1 para o caso de ser migrante e 0 para o caso contrário). As variáveis *dummy_escola* e *dummy_bairro* são variáveis indicadoras de bairros e escolas.

Tabela 17 - Conjunto de Variáveis do banco por nível de análise

Variável	Nível 1	Nível 2	Nível 3	Nível 4
resposta	<i>Alunos</i>	<i>Escolas</i>	<i>Bairros</i>	<i>Regiões</i>
<i>proficiência</i>	<i>Idade</i>	<i>Dependência</i>	<i>Cor (% no agregado)</i>	<i>Dummy_região</i>
	<i>Sexo</i>	<i>Turno</i>	<i>Média de anos de</i>	
	<i>Cor</i>	<i>Atenção</i>	<i>Estudo das mães</i>	
	<i>Estudo_mãe</i>	<i>Barulho</i>	<i>Bolsa Família</i>	
	<i>Rua_calçada</i>	<i>Falta_prof</i>	<i>NSE_médio</i>	
	<i>Água</i>	<i>Educação Física</i>	<i>imigrantes (% no agregado)*</i>	
	<i>Rádio</i>	<i>NSE_escola</i>	<i>Dummy_bairro*</i>	
	<i>Geladeira</i>	<i>imigrantes (% no agregado)*</i>		
	<i>Automóvel</i>	<i>Dummy_escola*</i>		
	<i>Computador</i>			
	<i>Bolsa_família</i>			
	<i>Repetiu</i>			
	<i>Dever_de_casa</i>			
	<i>Estuda</i>			
	<i>Pais_dever</i>			
	<i>Pais_ajudam</i>			
	<i>Migrante*</i>			
	<i>Distância*</i>			

Obs: As variáveis com ‘*’ foram obtidas do banco georreferenciado.

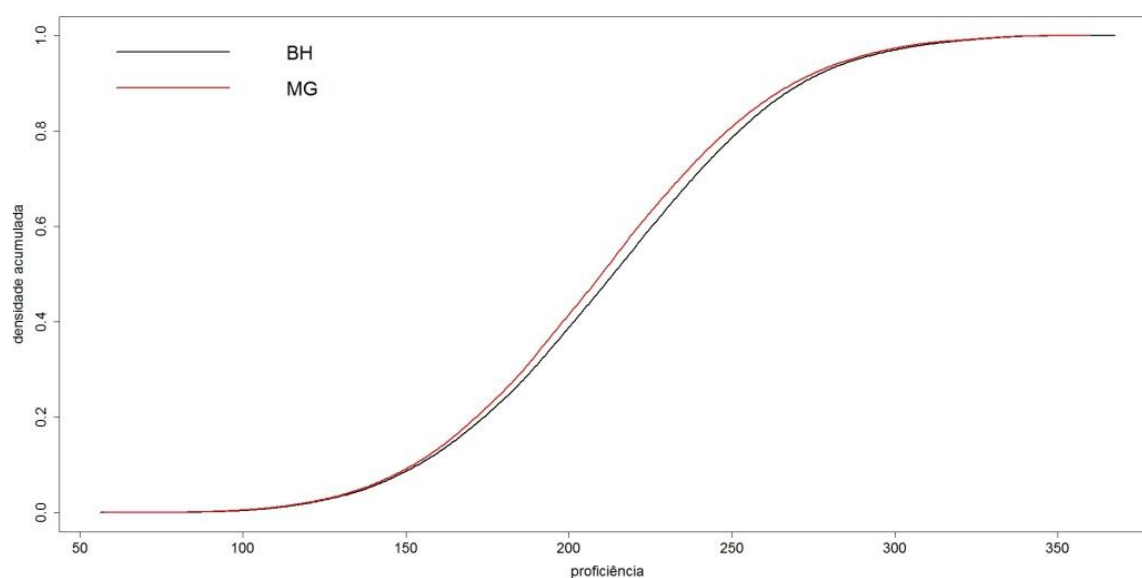
Fonte: Banco de dados geoprocessado (2010).

A proficiência, por ser a principal variável de resultado, convém ser analisada em maiores detalhes. A proficiência média de Belo Horizonte se encontra no nível intermediário da escala de desempenho do PROEB (213,60 pontos), e pouco acima da média de Minas Gerais (210,6), ambas estão no mesmo patamar de desempenho (ver TAB. 18). A distribuição da proficiência de BH tem um pouco mais de alunos nos níveis intermediário e avançado, mas essa diferença é pequena, as distribuições acumuladas das notas estão muito próximas (GRAF. 10).

Tabela 18 - Média e proporção de alunos nos níveis de proficiência BH e MG

	<i>Média</i>	<i>Desvio-padrão</i>	<i>Min</i>	<i>Max</i>	<i>Abaixo do Básico</i>	<i>Básico</i>	<i>Intermediário</i>	<i>Avançado</i>
BH	213,6	45,27	73,77	349,90	8,5%	30,0%	40,0%	21,5%
MG	210,6	45,12	67,47	349,90	9,0%	32,4%	39,5%	19,2%

Fonte: PROEB, 2010.

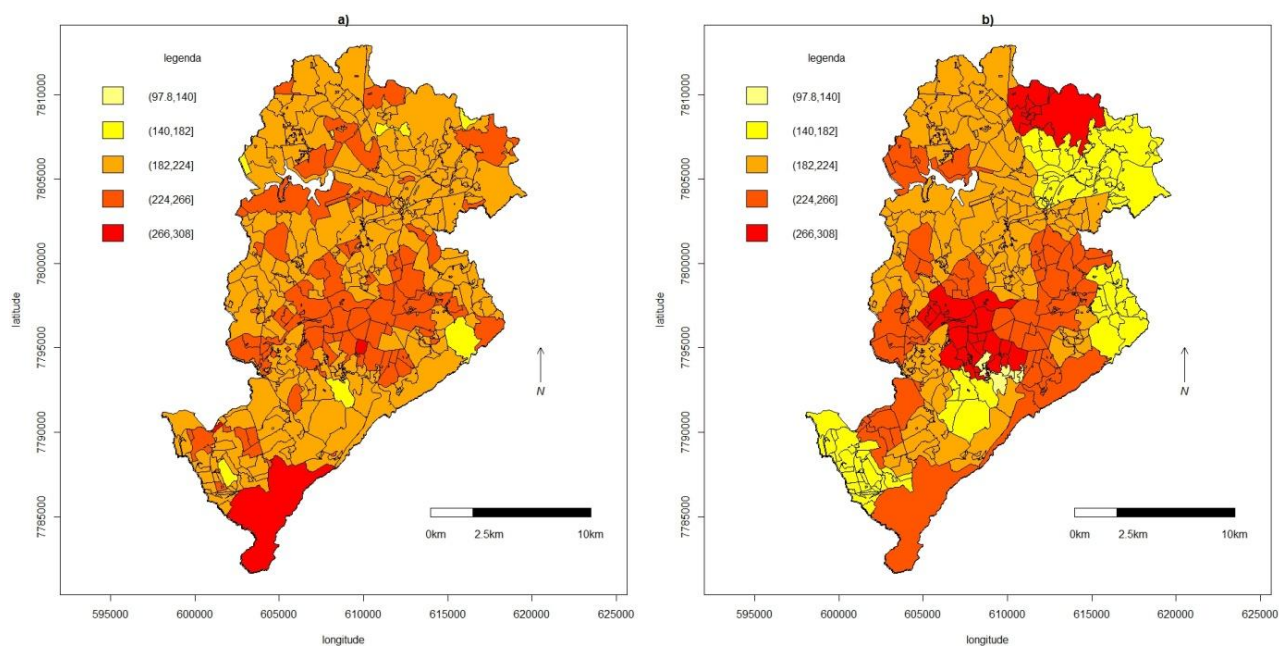
Gráfico 10 - Densidade Acumulada da Proficiência BH e MG

Fonte: PROEB, 2010.

A proficiência segue um padrão espacial similar ao observado para as variáveis socioeconômicas. A região Centro-Sul possui alunos com nível de proficiência consideravelmente acima da média (acima de 224 pontos). No mapa 6-a se visualiza essa distribuição por bairros. Na região Centro-Sul se destaca o bairro Cidade Jardim cuja vizinhança (Lourdes, Santo Agostinho, Gutierrez, Santo Antônio, Monte São José, e Coração de Jesus) também têm média alta para a proficiência.

Na região do Barreiro se destaca o pequeno bairro de João Paulo II à esquerda, e a Serra do Curral no extremo sul à direita, cada um dos bairros com dois e três alunos respectivamente. Para obter maior representatividade dos bairros, a informação de grupos de bairros também foi acrescentada no mapa 6-b.

Mapa 6 - a) Distribuição da Proficiência por bairros e b) por grupos de bairros



Fonte: Banco de dados geoprocessado (2010).

Quanto à divisão do banco por sexo dos alunos, a base possui 8.254 alunas e 8.100 alunos. A média de idade dos alunos do 5º ano é de 10,8 anos e a mediana igual 11 anos. O aluno mais novo declarou ter 6 anos de idade e o aluno mais velho declarou ter 37 anos.⁹⁹ O segundo mais velho declarou ter 18 anos (apenas um caso) e 2,3% do banco (372 alunos) tem idade maior ou igual a 13 anos e cerca de 17% dos alunos já repetiram alguma das séries anteriores ao menos uma vez (o valor da variável é 1 para o caso do aluno que repetiu uma ou mais vezes, e zero para os casos em que não foi reprovado).

A TAB. 19 apresenta o resumo das principais variáveis do nível do aluno. Muitas das variáveis categóricas com mais de uma categoria de resposta foram transformadas em variáveis binárias, esta transformação está indicada por um asterisco. Na variável *sexo*, 1 corresponde à “Feminino”.

Em *cor* 1 foi usado para a cor “branca” ou “amarela”, 0 foi utilizado para as cores “preta”, “parda” ou “indígena”. *Estudo_mãe* está na forma como apresentada no capítulo 5 (PROEB).

⁹⁹ A informação que constava é de que o aluno estava na modalidade de ensino regular e não na Educação de Jovens e Adultos (EJA) como seria de se esperar para esses casos.

Em *dever_de_casa*, que mensura se pais estimulam fazer o dever de casa, a variável possui valor 1 para o caso de alunos que fazem os deveres com regularidade (“sempre” e “quase sempre”) e zero nos outros casos. Em *pais_ajudam*, que pergunta se pais auxiliam no dever, nessa variável as respostas “quase sempre” e “sempre” ganharam valor 1, as demais obtiveram valor zero.

Nas variáveis dos itens do domicílio e de consumo, o 1 indica “sim”, que a família possui um ou mais itens do bem de consumo e 0 para “não”, a família não o possui. No caso do *computador*, o valor é 1 para computador com internet e 0 para o aluno sem computador ou com computador sem internet.

As variáveis do PROEB possuem nível maior de não resposta (*missings*). A variável de maior número de omissões é *automóvel*, as omissões chegam a 21% do banco. A variável do PROEB com menor número de omissões é *repetência*. As variáveis *idade* e *sexo*, provenientes do censo escolar, têm perda de apenas 2% na amostra.

Tabela 19 - Resumo estatístico das principais variáveis associadas aos alunos

	<i>Média</i>	<i>Mediana</i>	<i>Desvio padrão</i>	<i>Min</i>	<i>Max</i>	<i>Missings</i>
<i>Idade</i>	11,77	11	0,756	6	37	0
<i>Sexo*</i>	0,50	1	0,450	0	1	0
<i>Cor*</i>	0,35	0	0,477	0	1	2.518
<i>Estudo_mãe</i>	4,12	4	1,644	0	6	2.085
<i>Dever_de_casa*</i>	0,96	1	0,464	0	1	389
<i>Pais_dever*</i>	0,72	1	0,450	0	1	426
<i>Pais_ajudam*</i>	0,63	1	0,482	0	1	442
<i>Água</i>	0,97	1	0,168	0	1	378
<i>Rua_calçada</i>	0,89	1	0,308	0	1	467
<i>Geladeira*</i>	0,98	1	0,144	0	1	3.282
<i>Automóvel*</i>	0,61	1	0,487	0	1	3.380
<i>Computador*</i>	0,55	1	0,497	0	1	400
<i>Bolsa Família</i>	0,41	0	0,491	0	1	847
<i>Repetiu*</i>	0,17	0	0,373	0	1	332

*Variáveis transformadas para a forma dicotômica.

Fonte: Banco de dados processado (2010), Censo Escolar (2010) e PROEB (2010).

A TAB. 20 fornece a matriz de correlação das variáveis. Especial destaque para a correlação negativa da *idade* com a *proficiência*, a relação positiva da *cor* (“branca” ou “amarela”) com maiores resultados nas notas, a relação positiva de *proficiência* com o nível de *escolaridade da mãe* e as relações negativas para a proficiência na presença de *bolsa família* e *repetiu* de ano, essa última é a correlação mais negativa. Posteriormente, na TAB. 21, que apresenta o modelo da equação 15, são exibidos os valores de teste para as variáveis em um modelo linear multivariado entre proficiência e as demais variáveis apresentadas.

Tabela 20 - Matriz de correlação das variáveis do banco, nível 1

	<i>proficiencia</i>	<i>Idade</i>	<i>sexo</i>	<i>cor</i>	<i>Estudo_mae</i>	<i>computador</i>	<i>Bolsa_fam</i>	<i>repetiu</i>
<i>proficiencia</i>	1,000	-0,165	0,119	0,138	0,166	0,158	-0,236	-0,293
<i>Idade</i>		1,000	-0,067	-0,077	-0,090	-0,092	0,079	0,476
<i>sexo</i>			1,000	0,009	-0,057	-0,029	-0,019	-0,093
<i>Cor</i>				1,000	0,093	0,126	-0,111	-0,076
<i>Estudo_mae</i>					1,000	0,203	-0,167	-0,134
<i>computador</i>						1,000	-0,195	-0,123
<i>Bolsa_fam</i>							1,000	0,116
<i>repetiu</i>								1,000

Fonte: Banco de dados processado (2010).

Algumas variáveis exógenas utilizadas no modelo possuem uma correlação mais expressiva (acima de 20%), como é o caso de *idade* e *repetência* ($\rho = 0,475$) ou de *Estudo_mãe* e *computador* ($\rho = 0,203$), essas são as maiores correlações cruzadas que não envolvem a variável dependente, ambas são significantes pelo teste de Pearson.

A variável *idade* será retirada do modelo, pois *repetência* parece captar melhor a relação com *proficiência*. Em um modelo com a presença de *repetência* e *idade*, a idade deixa de ser significativa. Depois de retirada de *idade*, permanecem algumas correlações entre as variáveis explicativas, no entanto, os valores VIF (*Variance Inflation Factors*) são próximos de 1, não denotando multicolinearidade.

Um primeiro modelo linear multivariado para o banco pode envolver uma análise apenas para o nível do aluno. Esse modelo está descrito na equação 15 abaixo e os valores dos parâmetros estão

na TAB. 21. O R^2 ajustado do modelo é baixo ($R^2_{\text{adj}} = 0,181$), pois se trata de uma regressão de corte temporal, que tipicamente apresenta valores mais baixos para o R^2 . Além disso, grande parte da variância dos erros não está explicada, pois ainda não se refere ainda aos modelos mais completos que serão apresentados na seção 6.3.

$$Y_i = \beta_0 + \beta X_i + \varepsilon_i \quad (15)$$

Modelo em que Y expressa o vetor de proficiência, em que Y_i é o valor da proficiência para de cada aluno i , β_0 é o parâmetro de intercepto, β um vetor de parâmetros em quantidade igual ao número de variáveis da matriz X_i , que possui uma linha para cada aluno. A regressão apresenta ainda erros ε_i normais e não correlacionados, os erros possuem média zero e variância conhecida. Como o modelo possui 14 covariáveis e 16.354 alunos, a equação acima fornece um vetor de parâmetros β de dimensão (1×14) e uma matriz X_i de dimensões (14×16.354) .

Como a variável de estudo da mãe (*estudo_mãe*) é uma variável categórica. Na regressão transformou-se cada uma das categorias de *estudo_mãe* em uma *dummy*, com exceção da primeira, que serviu de referencial (anexo F). Destaca-se na regressão, além das variáveis apresentadas na matriz de correlação, a variável *dever*, que indica se o aluno tem regularidade e costume de fazer o dever de casa. *Pais_dever* representa uma variável categórica binária para o caso dos pais que cobram o dever dos filhos e *pais_ajudam* são pais que auxiliam com regularidade o dever dos filhos, ambas possuem relação negativa com a proficiência.

Tabela 21 - Regressão linear multivaria da proficiência de alunos (nível 1)

<i>Variável</i>	<i>β's</i>	<i>Erro-padrão</i>	<i>t-valor</i>
<i>Intercepto</i>	155,971	4,950	31,510
<i>Sexo</i>	7,644	0,864	8,850
<i>Cor</i>	6,374	0,899	7,085
<i>Estudo_mãe_d1</i>	-4,678	2,071	-2,259
<i>Estudo_mãe_d2</i>	0,028	1,613	0,017
<i>Estudo_mãe_d3</i>	2,342	1,623	1,443
<i>Estudo_mãe_d4</i>	9,957	1,827	5,450
<i>Estudo_mãe_d5</i>	7,693	1,508	5,102
<i>Rua_calçada</i>	11,179	1,461	7,654
<i>Água</i>	27,155	2,941	9,237
<i>Rádio</i>	2,048	0,530	3,865
<i>Geladeira</i>	10,508	2,942	3,571
<i>Automóvel</i>	-4,725	0,913	-5,176
<i>Computador</i>	5,840	0,929	6,286
<i>Bolsa_fam</i>	-15,889	0,909	-17,473
<i>Repetiu</i>	-24,390	1,278	-19,084
<i>Dever</i>	21,480	2,451	8,757
<i>Pais_dever</i>	-5,101	0,960	-5,277
<i>Pais_ajudam</i>	-6,132	0,892	-6,867
			R ² ajustado = 0,182
			F-value = 2,2e ⁻¹⁶

Por fim, apresentam-se aqui nesta seção, as variáveis do nível escola. São variáveis que dizem respeito à administração e ao ambiente disciplinar dentro da escola. Outra variável importante é o nível socioeconômico da escola (NSE). Na TAB. 22 encontra-se o resumo estatístico dessas variáveis. Como no nível anterior, as variáveis sinalizadas com asterisco foram transformadas em binárias: *dependência* ganhou o valor 1 para escolas estaduais e 0 para municipais, *turno* é igual a 1 para turma da “manhã” e 0 para “tarde” ou “integral” (o turno integral possui apenas 27 alunos).

A variável *atenção* indica a percepção do aluno sobre a atenção dos colegas na aula, é igual a 1 para “quase sempre” e “sempre” e 0 para os demais casos. O mesmo ocorre em *barulho*, variável de percepção do aluno sobre o barulho em sala de aula, transformada para 1 nos casos em que o aluno disse ser “sempre” ou “quase sempre”, e 0 nos demais casos. A mesma regra vale para *falta_professor* (que capta a frequência de faltas do professor).

A penúltima variável é *Educ_física*, que capta a frequência de aulas de educação física. Essa variável ganhou valor 1 para frequências de “uma ou mais vezes por semana”, e frequências menores ficaram com zero. A variável NSE varia de 0 a 10 e não foi transformada se encontrado tal como propuseram seus autores: Alves & Soares (2012).

Tabela 22 - Resumo estatístico das principais variáveis associadas à escola

	<i>Média</i>	<i>Mediana</i>	<i>Desvio padrão</i>	<i>Min</i>	<i>Max</i>	<i>Missings</i>
<i>Proficiência</i>	211,2	209,9	18,427	152,0	305,4	0
<i>Dependência*</i>	0,513	1,000	0,501	0	1	0
<i>Turno*</i>	0,598	0,987	0,446	0	1	0
<i>Atenção*</i>	0,668	0,667	0,108	0,33	1	0**
<i>Barulho*</i>	0,473	0,469	0,120	0	0,82	0**
<i>Falta_professor*</i>	0,189	0,178	0,091	0	0,67	0**
<i>Educ_física*</i>	0,895	0,916	0,104	0,06	1	0**
<i>NSE</i>	5,339	5,272	0,431	4,58	6,88	0

*Variáveis transformadas para a forma dicotômica.

** Não há *missing* para o nível da escola, entretanto, por conta dos *missings* no nível dos alunos algumas escolas contam com informação de um número menor de alunos.

Fonte: Banco de dados processado (2010), Censo Escolar (2010) e PROEB (2010).

A equação 16 apresenta o modelo para o nível das escolas:

$$\bar{Y}_s = \gamma_0 + \boldsymbol{\gamma}\mathbf{Z}_s + u_s \quad (16)$$

Em que $\bar{Y}_s$ é o vetor de valores médios da proficiência com s para representar cada escola. Na equação há o parâmetro de intercepto γ_0 e um vetor de parâmetros $\boldsymbol{\gamma}$ de dimensões (1×7) , o modelo possui 7 variáveis. A matriz $\mathbf{Z}_s$ representa as características das escolas, que podem ser características médias das variáveis dos alunos ou variáveis exclusivas para a escola (no modelo acima só foram utilizadas características vinculadas à escola, embora *atenção*, *barulho*, *falta_professor*, *Educ_física*, sejam respostas dos alunos em relação às suas escolas). A matriz $\mathbf{Z}$ possui dimensões (7×296) , 7 variáveis e 296 escolas no modelo. Acrescenta-se, para cada escola s , um erro esperado com média zero e variância conhecida e independente dos erros das demais escolas.

Os resultados da regressão da equação 16 estão na TAB. 23 a seguir. Uma variável de grande magnitude negativa e significância é *Falta_prof*, que é a resposta dos alunos sobre a assiduidade do professor, uma escola com média alta de professor infrequente leva a uma perda de mais de 56 pontos na média de proficiência de cada escola (1,24 vezes maior que o desvio padrão).

Outras variáveis importantes também figuram na análise, a variável *dependência* mostra que há um efeito positivo na média para escolas do sistema estadual de ensino (acréscimo de mais de 5 pontos na média de proficiência). O *turno* em que estudam os alunos não foi significativo, assim como também não foram significativas variáveis indicativas de comportamento: *atenção* e *barulho*. A presença regular de educação física parece ter um efeito bastante positivo na média de proficiência da escola (mais da metade de um desvio padrão da proficiência) e o NSE da escola contribui em cerca de 1/3 do desvio padrão para aumentar a proficiência.

Tabela 23 - Regressão linear multivaria da proficiência nas escolas (nível 2)

<i>Variável</i>	<i>γ's</i>	<i>Erro-padrão</i>	<i>t-valor</i>
Intercepto	118,057	15,021	7,860
Dependência	5,195	1,790	2,902
Turno	-1,650	1,928	-0,856
Atenção	4,446	8,110	0,548
Barulho	-10,315	7,621	-1,354
Falta_prof	-56,028	10,652	-5,260
Educ_física	25,965	8,445	3,075
NSE	15,124	2,165	6,986
R ² ajustado = 0,426			
F-value = 2,2e ⁻¹⁶			

Acrescenta-se que o R² ajustado do modelo de escolas se apresentou maior do que o modelo do nível 1, em parte isso se deve à menor variância do número de observações e por se estar medindo características médias. Ainda assim os dados ainda explicam menos de 50% da variância observada na proficiência média das escolas. Modelos mais completos serão abordados na seção 6.3, que trata dos modelos hierárquicos, nessas situações se explora a inter-relação que os dados possuem entre si, respeitando sua hierarquia.

No que diz respeito aos fatores de interesse neste trabalho, que abordam a formação da preferência por escolas de pais e alunos, e o impacto da alocação das escolas na proficiência e aprendizado, pais precisam estar bem atentos às características tais como dependência da escola para qual o filho foi alocado, média de assiduidade dos professores e a oferta de educação física regular para os alunos. O NSE é uma variável difícil de ser apurada pelos pais, mas pode estar relacionada a características do entorno da escola e informações sobre o corpo discente. Essas características são importantes para moldar preferências dos pais e são auxiliares nos parâmetros finais propostos em 6.4 e nas simulações do capítulo 7.

6.2. Características de Mobilidade e Alocação Observada

Um fator que se acredita importante para a decisão dos pais (e que é central no sistema de cadastro de Belo Horizonte, bem como na maioria das cidades que adota uma alocação centralizada) é a distância da residência do aluno até a escola.

Como relatado na introdução e no capítulo 2, antes da implantação do sistema de cadastro as filas de matrículas eram grandes e geravam um custo social, além de uma desorganização do planejamento educacional do município (escolas estaduais e municipais). No início de cada ano letivo gastava-se um grande esforço humano em termos de tempo e dinheiro público, recursos que poderiam estar voltados para a questão pedagógica.

Tornava-se necessária a racionalização do processo de matrículas. No entanto, uma questão presente neste trabalho é: “na existência de vagas em todas as escolas, por que os pais não escolhiam matricular seus filhos nas escolas mais próximas?”. A resposta a essa pergunta parece estar no *trade-off* entre distância percorrida de casa até a escola (um custo privado) e o benefício (privado) de se estudar em uma escola melhor qualificada na escala de preferências de pais e alunos. Antes da implantação do sistema de matrículas, pais buscavam escolas públicas diferenciadas, mesmo que estivessem distantes da residência, pois os benefícios de tais escolas eram maiores que os custos de deslocamento.

Entretanto, para posterior interpretação dos modelos representativos de escolha e simulações que serão apresentados no capítulo 7, é preciso levar em conta que nem todas as escolas situadas a uma distância equivalente geram o mesmo benefício líquido. Se algumas delas não compensam o custo de deslocamento, teríamos uma preferência desigual no espaço.

O atual sistema de cadastro mimetiza uma preferência dos pais tal como se ela fosse ditada apenas pela distância de casa até a escola, sendo assim, pais não considerariam informações relacionadas ao processo pedagógico na formulação de uma escolha, ou, em outra possibilidade, pais e alunos podem julgar que as distâncias sintetizam as informações necessárias para uma decisão e a distância pode ser a mais importante para a formação das preferências.

Pela descrição já apresentada no capítulo 3, as matrículas anteriores ao sistema de cadastro, deveriam funcionar tal como um sistema de *Top Trading Cycles* (TTC) (Abdulkadiroğlu & Sönmez, 2003), com preferências aleatórias das escolas simulando a ordem das filas, o que equivale a uma indiferença da escola, já que o critério de desempate é aleatório. Sob essa situação, podemos relembrar que Abdulkadiroğlu & Sönmez (2003) mostram que o TCC alcança eficiência de Pareto, mas não é um mecanismo estável, pois apresenta pares bloqueantes. Essa instabilidade devia contribuir para a presença da burla no sistema de filas, caso as escolas (ou os dirigentes responsáveis) não fossem realmente indiferentes, acordos entre escolas e alunos (considerados ilegais) poderiam surgir.

A situação anterior ao cadastro pode ser considerada inferior ao sistema atual por três principais motivos: o primeiro é que a instabilidade do sistema gera um custo que não deve ser desprezível, como já mencionamos no capítulo de revisão da literatura. O segundo diz respeito à hipótese de total indiferença entre quaisquer alunos por parte das escolas, vimos que embora não seja muito adequado para escolas públicas estabelecer preferências, como tais escolas almejam também o bem-estar social, é bem possível se pensar em prioridades, sendo que uma delas que parece plausível no caso é o critério de distância.

O terceiro motivo trata de uma consideração de justiça (tal como denota o argumento da *inveja justificada* elaborado por Abdulkadiroğlu & Sönmez, 2003). Para compreender esse ponto é

preciso ter em mente que pode haver equilíbrios de Pareto que manifestam situações desiguais, tal como o equilíbrio dado pelo exemplo 8 enunciado em 3.4. A intuição para essa ideia pode ser exemplificada na situação de uma mãe que perde a oportunidade de matricular o filho na segunda escola de preferência, por ter esperado na fila por uma oportunidade na escola mais preferida, sendo ainda que tal aluno teria maior prioridade na segunda escola (por estar mais próximo, por exemplo) do que o outro que lhe foi realmente alocado.

Dessa maneira, analisando pelo critério das distâncias e sopesando as soluções de um mercado de TCC descentralizado, o sistema de cadastro de estudantes de Belo Horizonte cumpriu uma bem sucedida etapa. Na subseção 6.2.1 se verá que os alunos, em média, se encontram a uma distância pequena da escola em que estudam, comparando-se com uma situação aleatória que simula a situação anterior ao cadastro, obtêm-se que a distância média cai em mais de 2 Km. No capítulo 7, as consequências de diferentes alocações sobre a proficiência serão analisadas.

6.2.1. Alocação Observada de alunos e escolas

Utilizando-se do banco de dados processado e georreferenciado tem-se a alocação de escolas e alunos empregada nesta tese. Essa alocação será chamada de *real* ou *observada*, sendo que são também possíveis alocações *fictícias* ou *simuladas*, exploradas principalmente no próximo capítulo. A alocação é dada por uma matriz $\mathbf{M}$ em que na linha figuram os alunos e nas colunas as escolas, o 1 representa a presença do aluno ocupando uma vaga de determinada escola, e zeros representam espaços vazios ou inexistentes. O exemplo abaixo de uma alocação entre quatro alunos $I = \{i_1, i_2, i_3, i_4\}$ e três escolas $S = \{s_1, s_2, s_3\}$ esclarece melhor a definição da matriz de alocação:

$$\mathbf{M} = \begin{matrix} & \begin{matrix} s_1 & s_2 & s_3 \end{matrix} \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

Na matriz acima, os alunos i_1 e i_2 estão lotados na escola s_1 (para que isso ocorra, a escola 1 precisa ter, pelo menos, duas vagas), o aluno i_3 está alocado para a escola s_2 e i_4 está na escola s_3 .

A matriz de alocação observada investiga onde estudam os alunos e constrói uma matriz similar com 16.354 linhas e 296 colunas. Vinculadas à matriz, estão as informações dos alunos e das escolas, sendo que se pode construir uma matriz de distâncias observadas. O GRAF. 11 apresenta a densidade das distâncias da residência até a escola em Km.

Observa-se uma distribuição assimétrica. Como pode ser ressaltado na TAB. 24, um total de 82% dos alunos está a menos de 2km da escola. A assimetria da distribuição se deve a alguns poucos casos de informações bem discrepantes, a maior distância encontrada foi de 26,98km, um *outlier*. Para se obter uma ideia, a distância euclidiana do extremo sul da capital até o extremo norte é de 31,9km, esse aluno tem que atravessar quase toda a cidade para chegar até a escola, não é um caso impossível visto que há alunos moradores de outras cidades que estudam em Belo Horizonte, mas é provável que essa discrepância seja fruto de imprecisão da informação da residência do aluno, ou mesmo da escola onde ele estuda. Ou ainda de um terceiro caso, de um aluno que tenha se mudado entre o período de matrícula e o restante do ano letivo.¹⁰⁰

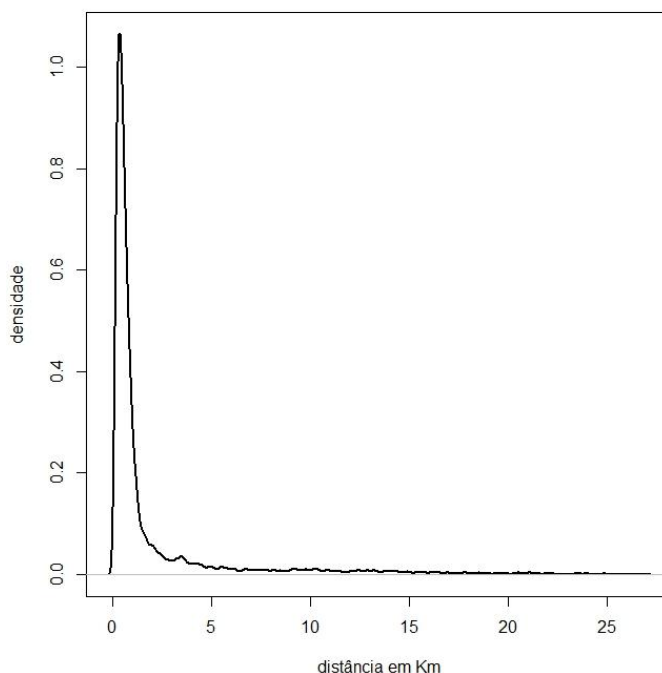
Tabela 24 - Proporção de alunos por distância em Km

Distância (em km)	Proporção acumulada de alunos ($\leq$)
1	71%
2	82%
5	91%
10	95%
15	98%
25	99%
30	100%

Fonte: Banco de dados processado (2010).

¹⁰⁰ O período de coleta de informações para o procedimento de matrículas e resposta do censo escolar não é o mesmo apesar de próximos. No sistema de cadastro os pais informam o endereço, em geral, entre os meses de julho e agosto, o resultado das matrículas é liberado no mês de dezembro, anterior ao início do período do ano letivo. As matrículas são realizadas de janeiro do ano seguinte e o preenchimento do censo escolar pelas escolas ocorre até meados de março.

Gráfico 11 - Densidade das distâncias de casa até a escola para alunos do banco



Fonte: Banco de dados processado (2010).

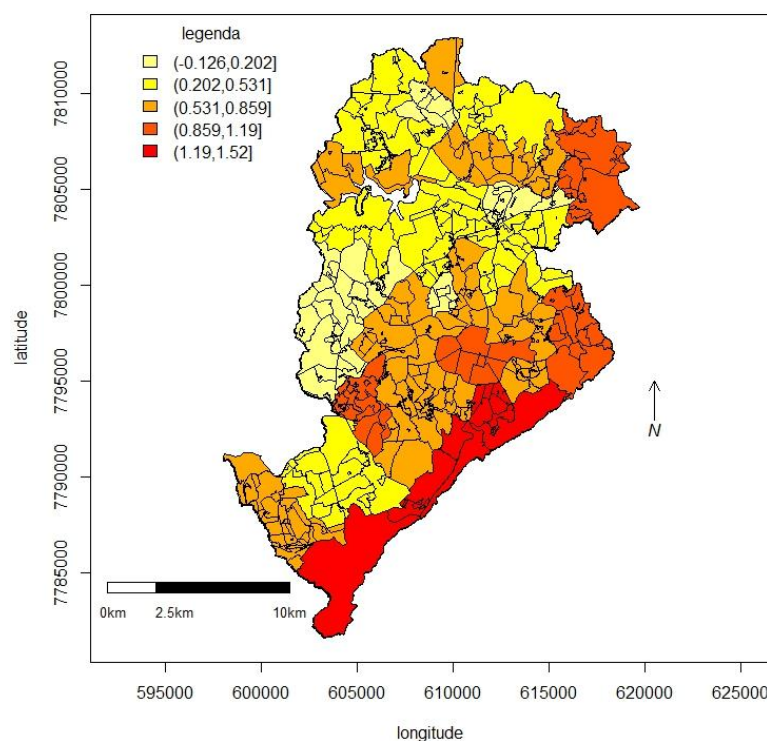
A distância que os alunos percorrem até chegar à escola está relacionada com o bairro em que moram, há alunos que moram regiões mais isoladas, enquanto outros, que moram em regiões centrais, possuem mais opções de escolas.

O mapa 7 apresenta a distância média por grupamento de bairros. Nota-se que bairros perto das regiões limítrofes de BH possuem média de deslocamento mais elevada, entretanto há bairros de regiões com bastante acesso, como a região central, que possuem média de distâncias elevadas. Isso pode ocorrer por endereços incorretos de pais que falsificam o endereço das regiões centrais com objetivo de ter maiores chances de matrícula em determinada escola. Pode ocorrer também pela facilidade de mobilidade no deslocamento urbano em termos de opções de transporte e infraestrutura viária, é mais fácil se deslocar 2km na região central do que na região Norte ou extremo nordeste, ou no leste da Pampulha.

A transformação logarítmica foi utilizada para tratar da distribuição assimétrica das distâncias. Ao longo deste capítulo, se emprega a transformação nas regressões e análises, porém, por

questão de auxílio da interpretação, ao abordar medidas de distância no texto e em tabelas de síntese (excetuando-se as das regressões), a referência será o indicador de distância em quilômetros não logaritmizados.

Mapa 7 - Distribuição do Log da distância por Grupos de bairros



Fonte: Banco de dados processado (2010).

Tabela 25 - Resumo das principais estatística da distância

	<i>Média</i>	<i>Mediana</i>	<i>Desvio padrão</i>	<i>Min</i>	<i>Max</i>
Distância (km)	1,785	0,579	3,446	0,005	26,98

Fonte: Banco de dados processado (2010).

Uma ressalva é que a distância euclidiana não considera os obstáculos físicos do terreno do meio urbano: lagoas, aeroportos, barreiras viárias. O ideal seria obter a distância viária percorrida pelos alunos. Como visto no capítulo 2, o sistema de cadastro de Belo Horizonte considera distâncias de forma compartimentada, fazendo uso das jurisdições escolares, elaborados de modo a evitar

travessia de obstáculos maiores para uma distância à pé. Portanto, com essas considerações, é preciso ter em mente que o ordenamento de distâncias da prefeitura é diferente do aqui adotado.

Apesar disso, com exceção de casos que fogem a regra, a adequação da distância oferece uma distribuição de distâncias que não deve estar distante da realidade da maioria dos alunos do município. Uma comparação da distância real percorrida com a distribuição obtida pode contribuir para desenvolvimentos posteriores mais detalhados do presente trabalho.

Vimos que embora grande parte do banco esteja a menos de 2 Km da escola, o deslocamento é uma variável importante obtida do banco georreferenciado. Em se tratando de crianças de 9 a 11 anos (faixa etária mais comum para o 5º ano), distâncias maiores do que 2,5km ou 3km podem ser grandes distâncias. A próxima subseção se dedicará a conceituação dos alunos que percorrem determinadas distâncias e com isso mudam de bairro, grupo de bairros, ou regiões. O efeito da distância é importante para tentar se determinar os resultados dos diferentes sistemas de alocação possíveis.

6.2.2. Migração de alunos dados e definições

A maioria dos estudantes do banco (79%) estuda na mesma região em que mora. Se considerar os que estudam no mesmo bairro em que moram, esse percentual cai para 37%, principalmente por conta da grande quantidade de pequenos bairros e bairros sem escolas em Belo Horizonte. Mas em grande parte dos casos, os alunos estudam em um bairro vizinho ou próximo. Como observado na TAB. 24 da seção anterior, mais de 80% dos alunos estuda em um raio menor do que 2km de sua casa.

Contudo, outro dado é que a maioria dos estudantes (55%) não estuda na escola mais próxima de sua residência. A TAB. 26 abaixo mostra a evolução da proporção de alunos de acordo com a posição da escola em que estuda no ordenamento das preferências que considera apenas a distância. O número 1 representa a mais próxima e o 296 a mais distante. A proporção de alunos que estuda em escolas para além da 21ª mais próxima é expressiva, constituindo-se 14% dos alunos.

No capítulo 7 será mostrado que pelo algoritmo de Gale-Shapley é possível deixar os alunos ainda mais próximos das escolas, ou seja, em se considerando apenas as distâncias, o algoritmo do sistema atual apresenta pares bloqueantes e, portanto, a alocação não se encontra no *core*. A alocação do sistema de Belo Horizonte pode apresentar instabilidades no sentido econômico do termo dado que, ao menos em tese, um par escola/aluno poderia sair do sistema e formar uma coalizão paralelamente. Isso, de fato, não acontece por ser proibido em termos legais. Porém tentativas de pais tentando alterar o endereço pode ser um indicativo de alguma instabilidade no sistema.¹⁰¹

Tabela 26 - Proporção de alunos por posição da distância da escola

Posição no ordenamento de distâncias	Proporção de alunos
1	45%
2	17%
3	7%
4	4%
5	3%
6 a 10	6%
11 a 20	4%
Acima de 21	14%
Total	100%

Fonte: Banco de dados processado (2010).

Outro ponto importante, relacionado ao fato da alocação atual não estar no *core*, indica que os ganhos agregados dos participantes do sistema podem ser maiores. Caso isso repercuta na proficiência média dos alunos, é possível levar o sistema ao *core* e aumentar o ganho geral agregado (o capítulo 7 mostra mais detalhes sobre esse ponto).

Descobrir para onde vão os alunos que se deslocam é também um fator importante para o entendimento da relação entre o deslocamento e a proficiência. Para isso é preciso definir o que é o aluno-migrante:

¹⁰¹ Em verdade, como se viu no capítulo 2, o sistema é flexível a ponto de permitir que pais possam contestar a sugestão da gerência de cadastro. A troca posterior ao processo não é comum, mas é possível e ocorre em bases regulares ano a após ano.

D.7. (aluno-migrante). *O aluno que se desloca para ir à escola e nesse movimento cruza os limites do bairro, do grupamento de bairros ou da região é chamado de **aluno-migrante**. O aluno-migrante pode ser **migrante de bairro**, quando reside em um bairro e estuda em outro, **migrante de grupo de bairros**, quando reside em um grupo de bairros e estuda em outro grupo. E **migrante de região**, quando mora em uma região, mas estuda em outra. Em termos formais: seja I o conjunto de todos os alunos e E o conjunto de todos os alunos que estudam no bairro em que moram. O conjunto de alunos não migrantes N é dado pela interseção de I e E ($N = I \cap E$). Se N está contido em I ($N \subseteq I$), o conjunto de migrantes (M) será dado pelo complementar de N em relação a I , tem-se a operação de conjuntos: ($M = I - N$).¹⁰²*

O conceito acima é diferente do conceito de migrante populacional. O aluno não precisa fixar residência em outro bairro ou região para frequentar uma escola de outra localidade. É preciso lembrar que no sistema de cadastro o aluno não tem liberdade irrestrita para escolher em qual escola estudar, dado que as matrículas são sugeridas pelo sistema, mas por algum motivo no procedimento (por pedido de alteração dos pais ou outras circunstâncias), alguns alunos foram matriculados em escolas mais distantes e a representatividade deles na amostra não é tão pequena.

As denominações de **intrarregião** e **entrerregiões** serão empregadas para denominar a variância de diversas variáveis dentro de uma mesma região e entre regiões distintas, variação intra e entre grupos:

D.8. (intrarregião ou intrabairros). *A variância é denominada de intrarregião (intrabairros) quando ocorre toda dentro de uma região limitada. Nesse caso, dado um Y como o indicador de proficiência e uma variável $D_r = 1$, indicadora de uma região 'r' (ou bairro k) da cidade, a variância intrarregião (intrabairros) é dada por: $\sum_{r \in R} \text{var}(Y|D_r = 1)$, em que R é o conjunto de todas as regiões (pode igualmente ser feito para os bairros: $k \in K$).*

¹⁰² O conjunto M pode ser também representado por: $M = \{M^1 \cup M^2 \cup \dots \cup M^k\}$, união dos conjuntos de alunos migrantes de todos os bairros $k \in K$. Em que M^k é dado por $M^k = I^k - N^k$, em que I^k é o subconjunto de I de alunos que moram no bairro k , e N^k é o conjunto de alunos que moram e estudam no bairro k , esses são os não migrantes do bairro k , um subconjunto de I^k ($N^k \subseteq I^k$). Tais subdivisões são muito úteis na análise e podem ser também realizadas para regiões (r).

D.9. (intrerregiões ou entrebairros). A variância é denominada de *entrerregiões (entrebairros)* quando ocorre entre regiões (bairros) distintas. Nesse caso, dado um $\bar{Y}_r$ como o indicador médio de proficiência da região 'r' (ou bairro 'k'), a variância *entrerregiões (bairros)* é dada por: $var(\bar{Y}_r)$, é a variância das médias, o $\bar{Y}$ assume valores conhecidos para cada $r \in R$ ($k \in K$).

Abaixo a TAB. 27 apresenta a variância da proficiência intrarregião, enterregiões, intrabairros e entrebairros (grupos de bairros).

Tabela 27 - Análise da variância Intra níveis e Entre níveis

	<i>Intra</i>	<i>Entre</i>	<i>Total</i>
<i>Região</i>	2.043,15 (99,7%)	5,87 (0,3%)	2.049,02 (100%)
<i>Grupos de Bairros</i>	1.983,55 (97%)	65,46 (3%)	2.049,02 (100%)
<i>Bairros</i>	1.559,68 (76%)	489,34 (26%)	2.049,02 (100%)
<i>Escolas</i>	1.709,48 (83%)	339,54 (17%)	2.049,02 (100%)

Fonte: Banco de dados processado (2010).

Os conceitos de *intrarregião* e *entrerregiões* serão abordados mais completamente na seção 6.3. Por ora, outras duas definições uteis são de alunos-emigrantes e alunos-imigrantes, doravante, denominados somente de *emigrantes* e *imigrantes*. Os *emigrantes* são aqueles que saem, deixam a região ou bairro onde moram e vão para outra região ou bairro. Os *imigrantes* (embora em população se reserve o termo mais para mudanças de país) são alunos provenientes de outra região ou bairro que veem estudar no bairro k , todo imigrante é um aluno que não estuda no bairro em que mora.

D.10. (aluno imigrante). Denominando-se por F^k o conjunto de estudantes de um bairro k e por M^j os migrantes de um bairro j , em que $k \neq j$, tem-se que o conjunto de alunos imigrantes do bairro k provenientes de j , representado por $\mathcal{I}^{kj}$, é a interseção dos conjuntos F^k e M^j ($\mathcal{I}^{kj} =$

$F^k \cap M^j$). O conjunto de todos os imigrantes do bairro k ($\mathcal{I}^k$) é dado pela união de todos os imigrantes $\forall j, j \in K$, tal que $j \neq k$, ($\mathcal{I}^k = \{(F^k \cap M^1) \cup (F^k \cap M^2) \cup \dots \cup (F^k \cap M^{k-1}) \cup (F^k \cap M^{k+1}) \cup \dots \cup (F^k \cap M^p)\}$, em que p é o ultimo bairro do conjunto K).¹⁰³

D.11. (aluno emigrante). A definição de aluno emigrante é similar à definição de aluno-migrante de **D.7**, trata-se do aluno que mora em um bairro k (região r), mas que não pertence ao conjunto de alunos que moram e estudam em k (r), o conjunto E^k . Chamando-se o conjunto de todos os alunos emigrantes de $\mathcal{E}$ tem-se que: ($\mathcal{E} = I - N$), em que I é o conjunto de todos os alunos e N é o conjunto dos alunos não migrantes ($N = I \cap E$), dessa forma tem-se $\mathcal{E} = M$, tal como definido em **D.7**.¹⁰⁴

Com o uso de tais definições, a TAB. 28 abaixo apresenta uma matriz de movimentos “migratórios” de alunos para as nove regiões de Belo Horizonte. Na linha apresenta-se a região onde mora o aluno e nas colunas estão as regiões de destino. A soma horizontal na última coluna oferece o total de emigrantes da região. A soma vertical fornece o número de imigrantes que a região recebe, na 11ª linha, *imigrantes*. A última linha apresenta o saldo de *imigrantes – emigrantes* para cada coluna.

A região de maior saldo migratório é a Centro-Sul que recebe 655 e manda para estudar fora 184 alunos. A segunda é a região de Venda Nova, que possui o segundo maior saldo, de 258, e o Barreiro é a terceira região em saldo migratório, com saldo positivo de 8 alunos. As demais regiões possuem saldo migratório negativo, com especial destaque para Pampulha que é a região de menor saldo migratório (negativo em -176).

Caracterizando os diversos tipos de movimentos de alunos em Belo Horizonte. Temos que 21% dos alunos trocam de região, 33% estudam mudam de grupo de bairro e 63% trocam de bairro para ir para escola. Mudar de bairro é uma mudança mais comum, dado que existem muitos bairros sem escolas.

¹⁰³ Aplica-se igualmente para as regiões. O conjunto de alunos imigrantes de um bairro k pode ser também representado por $\mathcal{I}^k = F^k - E^k$, uma operação de conjuntos que retira do total de estudantes do bairro k (F^k), o conjunto dos estudantes que moram e estudam no bairro (E^k).

¹⁰⁴ Para toda a cidade, o número de imigrantes é igual ao de emigrantes, que é por sua vez o mesmo que o conjunto de migrantes, o total de alunos que se deslocam ($\mathcal{I} = \mathcal{E} = M$).

Tabela 28 - Saldo “migratório” por regiões, alunos que se movimentam em BH

	<i>Barreiro</i>	<i>Centro-Sul</i>	<i>Leste</i>	<i>Nordeste</i>	<i>Noroeste</i>	<i>Norte</i>	<i>Oeste</i>	<i>Pampulha</i>	<i>Venda Nova</i>	<i>emigrantes</i>
<i>Barreiro</i>		68	17	22	10	36	51	13	22	239
<i>Centro-Sul</i>	25		76	8	10	11	27	17	10	184
<i>Leste</i>	41	215		80	23	10	37	29	35	470
<i>Nordeste</i>	26	59	113		25	102	71	23	28	447
<i>Noroeste</i>	35	100	24	31		16	71	101	19	397
<i>Norte</i>	25	32	32	72	23		42	96	50	372
<i>Oeste</i>	57	145	21	25	81	35		14	39	417
<i>Pampulha</i>	13	13	24	23	141	25	20		322	581
<i>Venda Nova</i>	25	23	11	17	15	39	25	112		267
<i>imigrantes</i>	247	655	318	278	328	274	344	405	525	3.374
<i>SALDO</i>	8	471	-152	-169	-69	-98	-73	-176	258	

Fonte: Elaboração do autor a partir do banco de dados processado (2010).

Nas tabelas 29 e 30 vemos que a Pampulha recebe muitos alunos, mas ao mesmo tempo envia muitos alunos para estudar fora, principalmente para as regiões vizinhas Noroeste e Venda Nova. Alguns bairros da Pampulha possuem maior carência de escolas, tal como pode ser revisto no mapa 4.

Na contramão desse movimento, a região Centro-Sul é um polo que atrai mais estudantes. O conjunto de alunos que moram e estudam no Centro-Sul (568) é menor do que o conjunto do total de alunos matriculados na região (1.223), de maneira que maior parte dos alunos da região é composta por alunos-imigrantes.

Tabela 29 - Regiões que mais recebem alunos em relação ao total de alunos permanentes

<i>Região</i>	Imigrantes	Alunos da Região	Imigrantes/Alunos* (%)
Centro-Sul	655	568	115%
Pampulha	344	806	43%
Norte	328	1.340	24%
Oeste	274	1.146	24%
Leste	318	1.455	22%
Venda Nova	405	1.954	21%
Noroeste	328	1.951	17%
Nordeste	328	1.713	16%
Barreiro	247	2.047	12%

Fonte: Banco de dados processado (2010).

* Alunos moradores que permanecem na região (não-migrantes).

Tabela 30 - Regiões que mais perdem alunos em relação ao total de alunos permanentes

<i>Região</i>	Emigrantes	Alunos da Região	Emigrantes/Alunos* (%)
Pampulha	518	806	64%
Oeste	417	1.146	36%
Centro-Sul	184	568	32%
Leste	470	1.455	32%
Norte	372	1.340	28%
Nordeste	447	1.713	26%
Noroeste	397	1.951	20%
Venda Nova	267	1.954	14%
Barreiro	239	2.047	12%

Fonte: Banco de dados processado (2010).

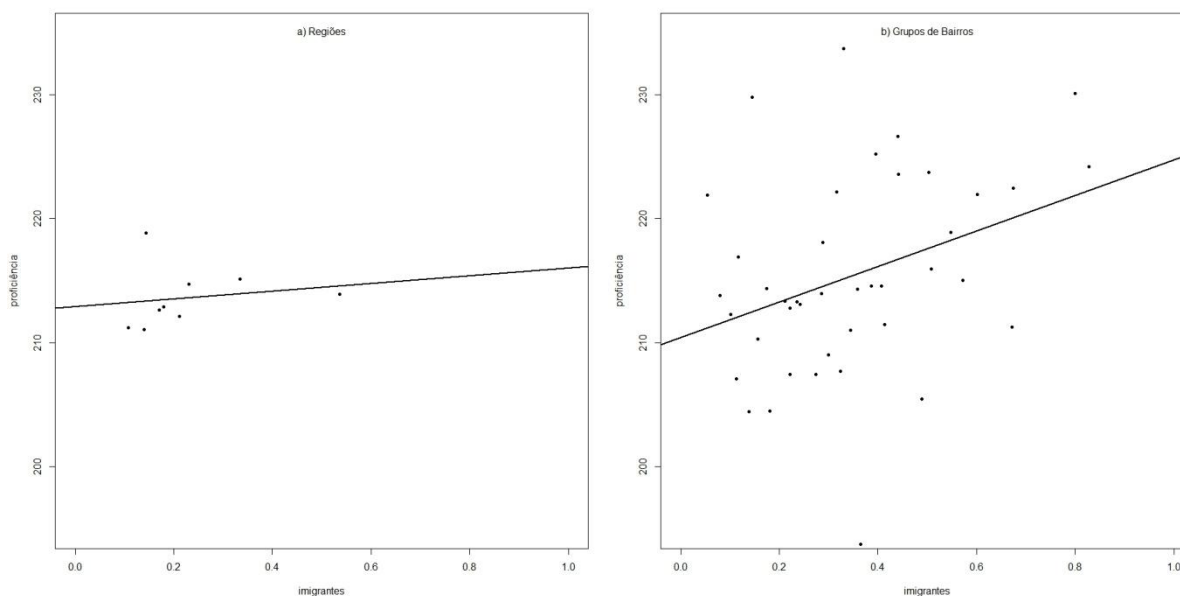
* Alunos moradores que permanecem na região (não-migrantes).

Regiões distantes como Barreiro e Nordeste, que possuem bairros mais isolados, recebem menos estudantes e não enviam muitos emigrantes para as outras regiões. As crianças do 5º ano costumam a permanecer na região em que moram. A mobilidade pode estar relacionada também à disponibilidade de serviço de transporte escolar. A PBH possui um sistema de cadastramento de veículos credenciados por trajeto. Na subseção a seguir serão investigados os efeitos desse movimento de estudantes nas variáveis de desempenho.

6.2.3. Efeito da migração e do deslocamento na proficiência

Resta ainda investigar os efeitos que o deslocamento pode acarretar no aprendizado dos alunos. Comparando-se a informação de migração por bairros e proficiência, tem-se que a imigração possui relação positiva com a proficiência média da região, porém, essa relação é apenas ilustrativa, dado que são apenas 9 regiões em Belo Horizonte (ver GRAF. 12-a). Para a situação de grupos de bairros a relação entre imigração e proficiência passa ser significativa ao nível de 5% (p -valor = 0,0282). Cada 1% a mais de imigração no grupo de bairros, a proficiência aumenta em 0,14 pontos em média (GRAF. 12-b).

Gráfico 12 - a) Gráfico de dispersão da relação do número de imigrantes e proficiência por regiões e b) por Grupos de Bairros



Fonte: Banco de dados processado (2010).

Intuitivamente, pode-se pensar que morar perto da escola é um fator positivo: o aluno perderia menos tempo no deslocamento e ganharia em tempo de estudos. Ter a escola próxima pode contribuir para assiduidade, acesso a atividades extracurriculares, prática de esportes, e outros fatores que auxiliam no desempenho. No entanto, o vínculo entre distância e proficiência no nível do aluno é difícil de ser estabelecido.

Em uma análise de regressão linear simples, temos uma relação negativa e significativa entre distância e proficiência, a cada 1 km acrescido na distância se perde 0,21 pontos de proficiência, o parâmetro possui significância de 5% (TAB. 31). No entanto, além do valor de beta demasiado pequeno e uma significância com alto valor para o critério de corte (5%, com t-valor = -2.085), o resultado não é robusto, com a transformação logarítmica o parâmetro se torna positivo e não significativo (TAB. 32). Os valores do R^2 ajustado são baixos, pois faltam muitas variáveis de controle.

Tabela 31 - Regressão linear simples entre proficiência e distância

Variável	β 's	Erro-padrão	t-valor
Intercepto	213,94	0,40	536,757
Distância	-0,21	0,10	-2,085
R^2 ajustado = 0,0002			
F-value = 0,037			

Tabela 32 - Regressão linear simples entre proficiência e log da distância

Variável	β 's	Erro-padrão	t-valor
Intercepto	213,71	0,37	580,990
Log(distância)	0,45	0,29	1,547
R^2 ajustado = 0,000			
F-value = 0,122			

Ao controlar a regressão por variáveis dos níveis 1 e 2, o valor do parâmetro da distância se torna mais robusto (sofre alterações menores por conta de outras variáveis inseridas nos modelos) e significativo. Adiante, mais modelos para medição do efeito da distância serão apresentados.

Em relação ao nível de proficiência da escola, há uma maior proporção de alunos “migrantes” entre escolas do decil superior de proficiência, com média de proficiência superior a 232 pontos ($\bar{y} \geq 232,3$). Esse padrão se mantém entre os outros segmentos do banco (regiões e grupos de bairros), a TAB. 33 apresenta os detalhes da proporção de migrantes entre as escolas do decil mais alto de proficiência, “*Superior*”, e demais escolas “*Inferior*”, no caso dos alunos migrantes de bairro as escolas do decil superior de BH possuem 73% de alunos migrantes, enquanto que as demais escolas possuem apenas 53% de migrantes neste nível. Resumindo, escolas com nível maior de proficiência possuem proporção ligeiramente maior de alunos migrantes.

Tabela 33 - Número de migrantes por tipo de escola de acordo com nível de proficiência

		<i>Inferior</i>	<i>Superior</i>
<i>Migrante de Região:</i>			
	Não	81,2%	67,3%
	Sim	18,8%	32,7%
		100%	100%
<i>Migrante do Grupo de Bairros:</i>			
	Não	70,0%	51,1%
	Sim	30,0%	48,9%
		100%	100%
<i>Migrante de Bairro*:</i>			
	Não	46,9%	26,7%
	Sim	53,1%	73,3%
		100%	100%

Fonte: Banco de dados processado (2010).

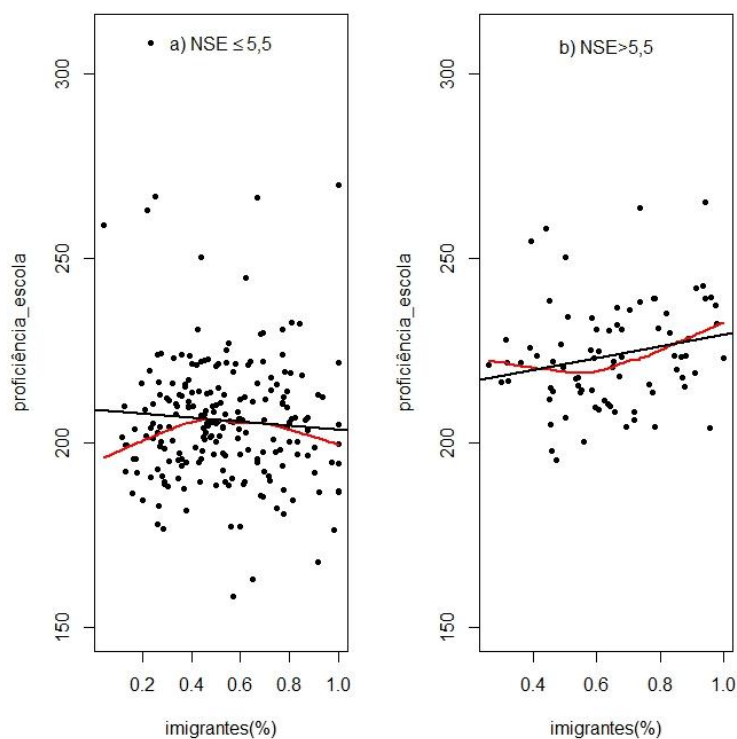
* O indicador de migrante para bairros usa, nesse caso, a informação de alunos que estudam fora de um raio de 2 Km de sua residência.

Obs: Todas as diferenças de da tabela médias são significativas com p-valor < 0,001.

Apesar de haver maior proporção de imigrantes entre escolas de maior nível de proficiência, esse fator não se manifesta homogeneamente nos dados. Entre escolas de NSE mais baixo ($NSE \leq 5,5$), uma maior proporção de imigrantes de bairro contribui negativamente para o resultado médio de proficiência. Em escolas de NSE mais alto ($NSE > 5,5$) o efeito é positivo. O GRAF. 13-a mostra a relação da proporção de imigrantes na escola e a proficiência para escolas de NSE menor ou igual a 5,5, o GRAF. 13-b mostra a mesma relação para escolas com NSE maior que 5,5. O GRAF. 13 apresenta a reta da regressão linear e uma regressão polinomial ponderada, linha em vermelho.

A TAB. 34 após o gráfico mostra a regressão por escolas para a proficiência mediana e a porcentagem de imigrantes sem o controle de NSE (modelo 1), para a qual a proporção de imigrantes é significativa a um t- valor $\geq 1,86$ (p-valor $\leq 0,064$). No modelo 2, com controle pelo NSE, a variável de migração não possui efeito significativo.

Gráfico 13 - a) Gráfico da relação entre proporção de imigrantes de bairro e proficiência média da escola com $NSE \leq 5,5$ e b) para $NSE > 5,5$



Fonte: Banco de dados processado (2010).

Tabela 34 - Regressão linear multivariada da proficiência com migrante de região

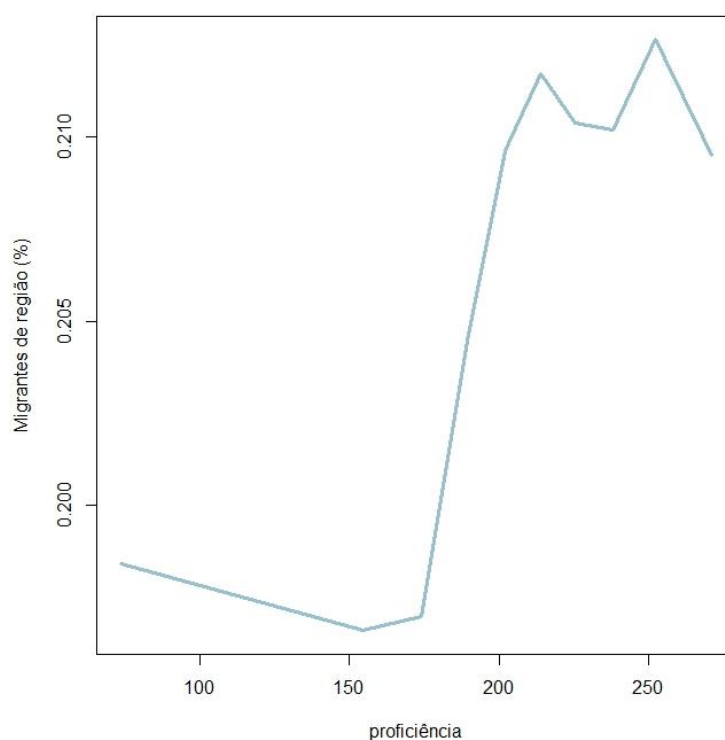
Variável dependente: proficiência de escolas						
Modelo1				Modelo2		
Variável	β 's	Erro-padrão	t-valor	β 's	Erro-padrão	t-valor
Intercepto	211,877	2.909	72,84	86,961	11,479	7,58
Migrante_reg	8,973	4.823	1,86	-4,289	4,277	-1,00
NSE				23,670	2,214	10,69
R ² ajustado = 0,010				R ² ajustado = 0,281		
F-value = 0,064				F-value = 2,2e ⁻¹⁶		

A relação entre a proficiência e a proporção de alunos que migram está presente no GRAF. 14. O gráfico foi construído com uma banda móvel deslocando-se entre os decis de proficiência, para cada decil se obtém a proporção de imigrantes do grupo de alunos. Após essa construção o gráfico foi suavizado tomando-se a média em uma janela de 3 decis. Para o primeiro e último

decil, tomou-se apenas o decil adjacente, de maneira que o 2º decil participou da ponderação da proficiência do primeiro, e de igual maneira, o 9º decil participou da ponderação do 10º.

A diferença na proporção de imigrantes de região entre os alunos de proficiência alta e baixa é um pouco maior do que 1%. O mesmo padrão se repete para migrantes de grupos de bairros e bairros. Entretanto essa diferença é significativa (p-valor menor do que 1%) apenas para bairros, e não é significativa para os demais tipos de migração.¹⁰⁵ A TAB. 35 apresenta essa diferença na proficiência entre os grupos de migrantes, a tabela traz também os intervalos de confiança para a média de proficiência dos grupos.

Gráfico 14 - Regressão polinomial localmente ponderada da proficiência e migração/região



Fonte: Banco de dados processado (2010).

¹⁰⁵ Para alunos que migram entre grupos de bairros o p-valor é 0,088. E para migrantes de regiões é de 0,1621.

Tabela 35 - Diferença na proficiência por migração de bairro em um raio de 2 km

		<i>Média Proficiência</i>	<i>IC_inferior</i>	<i>IC_superior</i>
<i>Migrante de Região:</i>				
	Não	213,31	212,24	214,37
	Sim	214,53	213,47	215,59
<i>Migrante de Grupos de bairros:</i>				
	Não	213,14	212,08	214,20
	Sim	214,43	213,36	215,50
<i>Migrante de Bairro*:</i>				
	Não	212,10	211,02	213,14
	Sim	214,73	213,67	215,80

Fonte: Banco de dados processado (2010).

Obs: intervalo de confiança: $\theta - 3*ep < \theta < \theta + 3*ep$. Em que θ é a média do parâmetro e ep o erro-padrão.

* O indicador de migrante para bairros usa a informação de alunos que estudam fora de um raio de 2 Km de sua residência.

Voltando ao efeito da distância, combinando com o efeito de migração e das demais covariáveis, temos que o valor negativo para o parâmetro da distância se torna mais robusto às mudanças do modelo, bem como o efeito de migração para os diversos níveis de agregação. Partindo de três modelos mais parcimoniosos para captar a distância, contando apenas com as variáveis: *estudo_mãe*, *bolsa_família*, *repetiu*, *falta do professor* e *NSE*, além do controle por regiões, vemos que o efeito do logaritmo da distância é negativo e significativo com o t-valor maior do que 2,00 (p-valor $\leq 0,005$).

A TAB. 36 apresenta 3 modelos que variam apenas com o tipo de variável de migração (Região, Grupo de Bairros e Bairros), cada um dos três modelos explica 20% da variância dos dados, os parâmetros para as variáveis de controle serão explicitadas na seção 6.3.

Por se tratar de um modelo log-linear, a interpretação do coeficiente angular do parâmetro de distância pode ser feita por meio da fórmula $\beta \frac{1}{x_i}$, dessa maneira para um valor de x fixo em 1 km temos que a distância diminui a proficiência em -0,783, no modelo com o maior valor para o parâmetro e em -1,122, modelo com menor valor do parâmetro. Para o coeficiente angular a uma

distância de 0,5 km, esses valores negativos dobram, indicando que a perda de proficiência é maior nas primeiras diferenças de quilômetros percorridos.

Tabela 36 - Regressão linear multivariada da proficiência com migrante de região

<i>Variável</i>	<i>Modelo3</i>			<i>Modelo4</i>			<i>Modelo5</i>		
	β 's	<i>Erro-padrão</i>	<i>t</i> -valor	β 's	<i>Erro-padrão</i>	<i>t</i> -valor	β 's	<i>Erro-padrão</i>	<i>t</i> -valor
<i>Intercepto</i>	152,864	4,433	33,480	153,067	4,428	34,570	152,875	4,434	34,474
<i>Estudo_mãe_d1</i>	-4,206	1,477	-2,849	-4,195	1,477	-2,841	-4,178	1,477	-2,830
<i>Estudo_mãe_d3</i>	1,769	1,050	1,684	1,753	1,051	1,668	1,765	1,050	1,681
<i>Estudo_mãe_d4</i>	7,473	1,271	5,878	7,469	1,271	5,875	7,469	1,271	5,875
<i>Estudo_mãe_d5</i>	4,041	0,922	4,382	4,026	0,922	4,365	4,034	0,922	4,374
<i>Bolsa_fam</i>	-13,422	0,756	-17,750	-13,421	0,756	-17,748	-13,422	0,756	-17,749
<i>Repetiu</i>	-25,753	1,005	-25,624	-25,761	1,005	-25,633	-25,787	1,005	-25,655
<i>Falta_prof</i>	-25,898	0,944	-27,424	-25,898	0,944	-27,422	-25,899	0,944	-27,424
<i>NSE</i>	13,922	0,811	17,155	13,956	0,814	17,135	13,856	0,814	17,024
<i>Log(distancia)</i>	-1,102	0,356	-3,093	-0,769	0,364	-2,112	-1,122	0,385	-2,916
<i>Migrante_reg</i>	1,108	1,079	1,027						
<i>Migrante_grupo</i>				-0,597	0,948	-0,630			
<i>Migrante_bairro</i>							0,856	0,944	0,907
<i>Controles</i>									
<i>Região*</i>									
	R ² ajustado = 0,201			R ² ajustado = 0,201			R ² ajustado = 0,201		
	F-value = 2,2e ⁻¹⁶			F-value = 2,2e ⁻¹⁶			F-value = 2,2e ⁻¹⁶		

* Os valores dos parâmetros para região estão explicitados na seção 6.3.

Concluindo esta seção, temos que os dados mostram que a relação entre a distância de casa até a escola e a proficiência é negativa e significativa. Porém, todo cuidado é pouco com tal afirmação: o efeito da distância aparece marcadamente mesmo nos modelos mais simples, mas ajuda explicar uma parcela ínfima das variâncias dos modelos, ou seja, não é tão robusto quanto os conhecidos efeitos das variáveis de alunos e escolas.

Sobre a variável de migração, torna-se ainda mais difícil captar o efeito. Alguns indicativos dos dados agregados, porém, parecem interessantes. Nota-se que escolas com melhor desempenho na proficiência possuem tendência de ter um pouco mais de alunos imigrantes, a região central de BH, notadamente conhecida por possuir algumas escolas públicas de excelência, possui maior proporção de imigrantes da cidade. No entanto, não há como ser assertivo, pois existe também migração entre escolas de fraco desempenho, e ainda, algumas escolas com pequena porcentagem

de alunos imigrantes que se saem muito bem. Sendo que a migração também carece ainda de explicar maior parte da variância para obter um efeito mais significativo.

Algumas análises da próxima seção indicam que a escola para onde se migra importa. No GRAF. 13-b da presente seção viu-se que a proporção de imigrantes ganha correlação positiva com a proficiência em escolas que possuem NSE maior do que 5,5. A seção a seguir explora modelos de especificações mais detalhados, obtendo parâmetros que subsidiam a simulação do próximo capítulo.

6.3. Modelos Hierárquicos e Decomposição de Fatores

O objetivo desta seção é obter os parâmetros associados à atual alocação de alunos e escolas, esses parâmetros são importantes para a simulação dos modelos de *matching* do capítulo a seguir. Como visto na FIG. 9 do capítulo 4, uma combinação de parâmetros obtida de uma alocação alternativa proporcionam outro patamar de proficiência. Dessa maneira, propõe-se uma relação entre o bem-estar social e a proficiência que, por sua vez, depende da alocação vigente. Como retratado pela equação (2), tem-se que o nível geral de proficiência (Y) depende de uma função ' f ' cuja variável principal é a alocação dos estudantes (μ). Retomando esta equação (14):

$$Y = f(\mu) \quad (14)$$

Logicamente, a equação (14) pode ganhar e muito em termos de sofisticação, como vimos na seção anterior, a distância e a variável migratória possuem algum poder explicativo, mas diversas outras variáveis explicativas já conhecidas da literatura também possuem a propriedade de prever um nível maior de desempenho: escolaridade da mãe, gênero, raça, características do domicílio, nível sócio econômico, etc. Isso para citar apenas as relacionadas diretamente ao aluno, outras variáveis de escola e região também são muito importantes. Dessa forma a equação (14) pode ser desdobrada na equação mais geral (17):

$$Y = f(X, Z, W) \quad (17)$$

As variáveis X são aquelas relacionadas ao aluno, como idade, se repetiu o ano, hábitos de estudo, além de outras de nível individual. As variáveis Z são relacionadas à escola, tais como tipo de administração (estadual ou municipal), assiduidade dos professores, educação física, NSE escolar, e outras. Por fim, as variáveis W que são variáveis de região e podem também depender da alocação dos estudantes. Deste ponto em diante usaremos W para denotar as variáveis de região, não confundir com o W para o conjunto de mulheres no capítulo 3.

Ao mudarmos a alocação dos alunos por simulações, temos algumas características que permanecem fixas com o estudante e outras que se alteram à medida que o aluno troca de escola. Sendo assim, se um aluno i_1 , estudante de uma escola s_1 , passa a estudar em s_2 , ele sofrerá todos os efeitos ligados àquela escola s_2 , e não mais à escola s_1 onde estudava antes. Algumas qualidades relacionadas ao aluno continuarão inalteradas, mas outras variáveis ligadas à escola, bairro e região irão mudar. Estimar os parâmetros relacionados a essas variáveis que impactam na proficiência é o objetivo desta seção.

Tal objetivo, realizado na obtenção dos parâmetros, pode ser interpretado como um exercício de decomposição de fatores, tais como as decomposições clássicas de Oaxaca (1973) e Blinder (1973), ou de outras formas de decomposição desenvolvidas por Juhn, Murphy & Pierce (1993). Da obtenção de uma regressão com os valores dos parâmetros pode se obter um efeito do diferencial de proficiência dos alunos caso eles tivessem alocações coincidentes.

Como já mencionado, as informações do banco de dados foram organizadas em quatro níveis principais: **alunos**, **escolas**, **bairros** e **regiões**. Para uma estrutura de dados como essa, é possível empregar modelos de regressões hierárquicas propostos por Bryk & Raudenbush (1992). Tal método já foi utilizado extensivamente (César & Soares (2001), Albernaz, Ferreira & Franco (2002), Machado *et al.* (2006), dentre outros) na aplicação ao contexto escolar brasileiro. Os modelos hierárquicos têm a vantagem de lidar com a estrutura aninhada dos dados, sendo possível, por meio da correlação intraclasse, captar a variância correspondente a cada nível, auxiliando na decomposição dos fatores.

Os modelos hierárquicos servem para outro importante propósito: captar o efeito escola. Como visto no capítulo 4, pretende-se aqui capturar o efeito de diferentes alocações de alunos e escolas. Essas simulações são sugeridas pelos modelos de pareamento por preferências e utilidades. Para se conhecer o impacto de uma mudança de estudantes nas escolas é preciso ter uma estimativa do efeito-escola, bem como da interação entre alunos e escolas. Os modelos hierárquicos servem para o propósito de se capturar esses efeitos. Ao final da análise, obtém-se um efeito para cada escola, bairro e região da cidade. Os efeitos estimados serão aplicados para capturar o impacto de alocações diferentes na proficiência média em língua portuguesa, resultados das simulações do capítulo 7.

Ao cambiar de escola, o aluno leva consigo suas características pessoais, ou seja, o efeito-aluno permanece o mesmo, ainda que se altere a escola em que o aluno estuda. Essas características relacionadas ao aluno estão presentes nas variáveis do nível 1 de análise, descritas na TAB. 17 na seção 6.1. A escola possui variáveis associadas aos procedimentos educacionais e que podem interferir nos parâmetros do primeiro modelo, este é o nível 2 do modelo hierárquico. O terceiro e quarto níveis seguem lógica semelhante, e denotam bairro e regional respectivamente. Importante ressaltar que o quarto nível (das regiões), só possui nove observações, sendo que será empregado apenas em caráter descritivo, sem se fazer menção à inferência de seus parâmetros.

Dando continuidade à numeração dos modelos de regressão já utilizados, adotaram-se três novos modelos: o modelo 6 (nulo), que apresenta os 4 níveis, mas não possui nenhuma especificação para variáveis exógenas, apenas para os interceptos. Esse modelo é o mais simples e objetiva retratar a decomposição livre de qualquer efeito para as variáveis explicativas.

Modelo 6 (nulo):

$$\text{Nível 1:} \quad Y_{iskr} = \beta_{0jkr} + \varepsilon_{iskr} \quad (18)$$

$$\text{Nível 2:} \quad \beta_{0skr} = \gamma_{0kr} + u_{skr} \quad (19)$$

$$\text{Nível 3:} \quad \gamma_{0kr} = \mu_{0r} + v_{kr} \quad (20)$$

$$\text{Nível 4:} \quad \mu_{0r} = \pi_0 + \xi_r \quad (21)$$

O ‘i’ indica o subscrito para cada aluno, o s para as escolas, k para os bairros¹⁰⁶ e r para as regiões. Y é a variável proficiência, β_0 , γ_0 e μ_0 e π_0 representam os parâmetros de intercepto para cada um dos níveis. Só há um parâmetro para cada nível, dado que em tal modelo não existem variáveis explicativas. A aderência do modelo nulo reside inteiramente nos parâmetros de intercepto, que traduzem médias não condicionadas das informações de proficiências dos alunos, nas escolas, nos bairros e regiões. O restante da variação não explicada está traduzido nos erros dados por: ε_{iskr} , u_{skr} , v_{kr} , e ζ_r . Os erros possuem média 0, variância conhecida e são independentes.

A TAB. 37 apresenta o modelo 6 com apenas os níveis de alunos e escolas e nenhuma especificação. Esse modelo serve de base para os demais, apresentando a variância entre os diversos níveis. Pode-se obter, por esse modelo, o coeficiente de correlação intraclasse (do inglês, ICC, *Intraclass Correlation Coefficient*) dado pela variância observada em cada nível sobre o total da variância dos resíduos:

$$ICC = \frac{\sigma_{us}^2}{(\sigma_{us}^2 + \sigma_{\varepsilon is}^2)} \quad (22)$$

A versão do ICC apresentada acima se refere a um modelo de dois níveis, alunos e escolas. O σ_{us}^2 é a variância dos erros do nível 2, e $\sigma_{\varepsilon is}^2$ é a variância do nível 1, sendo que $(\sigma_{us}^2 + \sigma_{\varepsilon is}^2)$ é a variância total dos resíduos. O ICC pode ser generalizado para modelos com mais níveis. No modelo nulo a variação respondida pelo o nível 1 dos alunos é de 81,48%, o nível 2, das escolas, responde por 15,72% da variância total do modelo, o nível 3, dos grupos de bairros, explica fração pequena da variância 0,28% e o quarto e último nível, das regiões administrativas, que possui 9 unidades, não possui fração explicada da variância relevante no modelo 6, no entanto, continua no modelo em caráter descritivo.

¹⁰⁶ Ao serem mencionados bairros no contexto dos modelos hierárquicos entenda-se como os 41 grupos de bairros criados em 6.1.2. A relação completa está no anexo G.

Tabela 37 - Modelos Hierárquicos Nulo, modelo 6

Modelo 6			
LogLik = -84.615			
Variável	β 's	Erro-padrão	t-valor
Intercepto	211,565	1,106	191,30
	variância	ICC	
Nível 1 (Alunos):	1.474,537,	81,48%	
Nível 2 (Escolas):	275,870	15,72%	
Nível 3 (Bairros):	4,993	0,28%	
Nível 4: (Regiões):	0,000	0,00%	
Total:	1.755,40	100,0	

O modelo 7 (*naïve*) se constitui de uma versão mais simples de um modelo com variáveis explicativas. Os efeitos de escola, bairro e região ocorrem apenas no intercepto da regressão do nível imediatamente acima (ver equações 24,25 e 26). Nesse modelo, ocorre uma diferença de médias entre diferentes escolas, bairros e regiões, ele é dado pelo seguinte conjunto de equações:

Modelo 7 (naïve):

$$\text{Nível 1: } Y_{iskr} = \beta_{0jkr} + \boldsymbol{\beta}_{skr} \mathbf{X}_{iskr} + \varepsilon_{iskr} \quad (23)$$

$$\text{Nível 2: } \beta_{0skr} = \gamma_{0kr} + \boldsymbol{\gamma}_{kr} \mathbf{Z}_{skr} + u_{skr} \quad (24)$$

$$\text{Nível 3: } \gamma_{0kr} = \mu_{0r} + \boldsymbol{\mu}_r \mathbf{W}_{kr} + v_{kr} \quad (25)$$

$$\text{Nível 4: } \mu_{0r} = \pi_0 + \xi_r \quad (26)$$

Novamente os parâmetros dos interceptos estão presentes e o $\boldsymbol{\beta}$ é um vetor de parâmetros para cada variável de aluno disponível em uma matriz $\mathbf{X}$.¹⁰⁷ No segundo nível de análise (eq. 24), a variável regredida é o parâmetro de intercepto obtido para cada escola do nível anterior. Desta feita, aplicam-se neste nível de análise a matriz de variáveis $\mathbf{Z}$ relacionadas à escola. Como no modelo anterior, a equação de cada nível possui seus erros aleatórios de média zero e desvio-padrão constante: ε , u , v , e ξ .

¹⁰⁷ Consultar a tabela 17 da seção 6.1 para ver a lista de variáveis.

Os níveis seguintes seguem a mesma lógica, reparando que o quarto e último nível não possui variável explicativa associada (semelhante ao modelo nulo), possui apenas um intercepto geral e o parâmetro de erro específico da região, ζ_r .

Os modelos hierárquicos podem ser descritos como uma equação única que reúne em uma só regressão todas as variáveis utilizadas. Substituindo (24), (25) e (26) na primeira equação (23) teríamos:

$$Y_{iskr} = \pi_0 + \mu_r W_{iskr} + \gamma_{kr} Z_{iskr} + \beta_{jkr} X_{iskr} + \varepsilon_{iskr} + u_{iskr} + v_{iskr} + \xi_{iskr} \quad (27)$$

Entretanto, a principal vantagem dos modelos hierárquicos é obter progressivamente os erros associados a cada nível, o que seria impossível na forma acima, já que os erros aleatórios e independentes apareceriam somados como um único resíduo. As etapas de agregação não estariam explícitas.

Talvez uma das principais vantagens dos modelos hierárquicos é permitir especificações dos erros diferentes para cada nível de análise. Para esse caso, a matriz de covariância entre os erros deve ser diferente de $\sigma^2 \mathbf{I}$ (em que σ^2 é a variância constante dos erros e $\mathbf{I}$ a matriz identidade). Pressupor a independência requer que os erros de diferentes níveis não estejam correlacionados.

No modelo proposto, uma alteração que se pode fazer é considerar a estrutura espacial de correlação. No entanto, foge aos objetivos mais imediatos deste trabalho, especificar modelos de variância diferentes da identidade, seria preciso uma matriz de vizinhança rica em detalhes, que considerasse como vizinhos não apenas bairros e regiões adjacentes, mas que considerasse o relevo socioeconômico entre bairros próximos, e de diferentes localidades como conjuntos, aglomerados e pequenas vilas.

É necessária uma estimação de valores verossímeis dos parâmetros e uma ideia das variações explicadas em cada nível do modelo. Dessa forma, os modelos aqui expostos não serão exaustivos quanto a predição da proficiência e, portanto, há muito espaço para serem aperfeiçoados.

A principal mudança do modelo 8 em relação ao modelo 7 anterior é a interação entre o aluno e escola. Nesse caso, não só o intercepto do nível 1 é explicado no nível 2, como também o parâmetro associado ao tipo de aluno (T), se migrante ou não (M) e a interação dessas variáveis com a escola são inseridas no 1º nível de análise.

Modelo 8 (interações):

$$\text{Nível 1: } Y_{iskr} = \beta_{0skr} + \beta_{1skr}T_{iskr} + \beta_{2skr}M_{iskr} + \beta_{3skr}Q_{iskr} + \beta_{skr}\mathbf{X}_{iskr} + \varepsilon_{iskr} \quad (28)$$

$$\text{Nível 2: } \beta_{0skr} = \gamma_{0kr} + \gamma_{1kr}\mathbf{Z}_{skr} + u_{0skr} \quad (29)$$

$$\text{Nível 3: } \gamma_{0kr} = \mu_{0r} + \mu_r\mathbf{W}_{kr} + v_{kr} \quad (31)$$

$$\text{Nível 4: } \mu_{0r} = \pi_0 + \xi_r \quad (32)$$

Também em relação ao modelo anterior, o modelo 8 acrescenta um efeito entre a nova variável Q , indicadora da qualidade da escola (Q para *qualidade*).¹⁰⁸ Infelizmente a variável Q , qualificadora do tipo de escola, não é tão geral a ponto de identificar diversas nuances do procedimento pedagógico entre escolas distintas, o Q é um indicador do nível da escola de acordo com a mediana da proficiência, ele é igual para escolas do decil superior de distribuição das notas médias por escola e zero para os demais casos. De fato essa não é o melhor indicador para captar qualidade das escolas já que possui relação óbvia com as variáveis explicadas, e pertence ao nível das escolas e não dos alunos. No entanto, a endogenia e o nível aqui são propositais, pois se procura uma relação entre o aluno e a qualidade da escola, ademais a introdução desta variável não prejudica o desempenho das demais no modelo.

Foram construídas três variáveis: *Melhor_BH*, *Melhor_Região* e *Melhor_Bairro*, na qual as escolas localizadas no decil superior de proficiência de Belo Horizonte, da região administrativa e do grupo de bairros, respectivamente, elas ganham o valor 1 caso estejam no decil superior de proficiência e zero nos demais casos.

A variável T indicativa do tipo do aluno é igual a 1 se o aluno pertence ao decil superior de proficiência, e zero no caso contrário. Sendo que essa variável pode ser repartida entre melhores

¹⁰⁸ Não confundir com o vetor ' Q ' indicativo do quantitativo de vagas do capítulo 3.

alunos do bairro, da região e de toda cidade. Nas regressões reportadas foi usado o tipo para o aluno do decil mais alto de proficiência de toda Belo Horizonte. A variável M é igual a 1 se aluno é migrante, e zero no caso contrário, os migrantes são os que estão reportados nos modelos anteriores.

O modelo 8 se baseia na hipótese de que se mudar para outra escola pode produzir um efeito maior (ou menor) entre o aluno e seu ambiente escolar, alterando a forma como esses dois agentes interagem (foram também estimados parâmetros de interação a variável de migração e distância e as variáveis T e Q). O tipo de escola para o qual se migra importa e os efeitos ganham uma diferente magnitude. Estudar em uma escola do primeiro decil de Belo Horizonte acrescenta, em média, 15,11 pontos na proficiência de português. Em contrapartida, em comparação ao aluno não migrante que estuda em uma escola de qualidade, o migrante leva uma desvantagem, uma perda de mais de 5,39 pontos (ver interação do modelo 8).

A TAB. 38 apresenta os modelos 7 e 8. Os modelos de regressão hierárquicos são estimados por máxima verossimilhança e, em geral, não apresentam estimativa do R^2 . O R^2 não pode ser obtido de maneira direta tal como nos modelos de regressão MQO. A interpretação do R^2 em modelos hierárquicos muda, pois, em alguns casos os valores de R^2 não estão mais contidos no intervalo de 0 a 1 (detalhes do uso e interpretação do R^2 nos modelos multiníveis Luke, 2004, p.33-37).

No caso dos modelos hierárquicos é mais comum utilizar o valor do logaritmo da função de verossimilhança, o *LogLik*. Quanto maior o *LogLik* (menos negativo), melhor a especificação do modelo. No caso dos modelos 7 e 8 os valores do *LogLik* são próximos, o modelo 8 possui valor um pouco maior, indicando que ele adere um pouco mais aos dados, e tanto 7 quanto 8 possuem valores de *LogLik* bem maiores do que o modelo nulo apresentado na TAB. 37.¹⁰⁹

Analisando-se o coeficiente de correlação intraclasse (o ICC), nota-se que a proporção da variância que é explicada pelos dados de alunos é de 96,85% no modelo 7 e 98,31% no modelo 8.

¹⁰⁹ Diferentemente do R^2 , o valor do *LogLik* não pode ser empregado na comparação de modelos distintos, sendo assim seu uso é mais adequado para analisar o ajuste de um modelo a determinado conjunto de variáveis. Em se tratando de modelos hierárquicos é também muito comum a apresentação dos valores AIC, Akaike Information Criterion, e BIC, Bayesian Information Criterion, de Akaike, 1987 e Schwarz, 1978, respectivamente (apud Luke, 2004, p. 34). Estas estatísticas não foram aqui expostas, pois, em geral, elas atuam no mesmo sentido que *Loglik*.

Tabela 38 - Modelos Hierárquicos de 4 níveis Modelos 7 e 8

Variável	Modelo 7			Modelo 8		
	LogLik = -38.010			LogLik = -37.976		
	β 's	Erro-padrão	t-valor	β 's	Erro-padrão	t-valor
Nível 1:						
<i>Intercepto</i>	122,437	7,372	16,61	140,050	7,247	19,33
<i>Sexo</i>	4,144	0,689	6,01	4,158	0,688	6,04
<i>Cor</i>	2,521	0,729	3,46	2,428	0,725	3,35
<i>Estudo_mãe d1</i>	-3,925	1,437	-2,73	-3,765	1,434	-2,63
<i>Estudo_mãe d3</i>	0,493	1,003	0,49	0,491	1,000	0,49
<i>Estudo_mãe d4</i>	4,073	1,209	3,37	4,127	1,205	3,42
<i>Estudo_mãe d5</i>	2,119	0,898	2,36	2,150	0,896	2,40
<i>Rua_calçada</i>	8,039	1,201	6,69	8,194	1,196	6,85
<i>Água</i>	18,472	2,469	7,48	18,400	2,462	7,47
<i>Rádio</i>	1,018	0,420	2,42	0,957	0,419	2,28
<i>Geladeira</i>	2,736	2,403	1,14	2,452	2,397	1,02
<i>Automóvel</i>	-4,833	0,729	-6,63	-4,827	0,728	-6,63
<i>Computador</i>	2,512	0,754	3,33	2,563	0,752	3,41
<i>Bolsa_fam</i>	-8,776	0,738	-11,89	-8,838	0,736	-12,01
<i>Repetiu</i>	-15,679	1,060	-14,79	-15,825	1,056	-14,98
<i>Dever</i>	12,719	2,063	6,16	12,502	2,056	6,08
<i>Pais_dever</i>	-2,720	0,768	-3,54	-2,649	0,766	-3,46
<i>Pais_ajudam</i>	-2,965	0,715	-4,15	-3,144	0,713	-4,41
<i>Migrante</i>	1,622	0,929	1,75	2,634	0,966	2,73
<i>Log(distância)</i>	-1,144	0,392	-2,92	-1,218	0,387	-3,15
<i>Tipo_aluno</i>	68,721	1,070	64,24	70,327	1,646	42,73
<i>Qualidade_escola</i>				15,115	2,214	6,83
<i>Migrante:Tipo</i>				-3,802	2,119	-1,79
<i>Migrante:Qualidade</i>				-5,389	2,372	-2,27
Nível 2:						
<i>Dependência</i>	1,789	1,029	1,74	0,621	0,924	0,67
<i>Turno</i>	-1,613	0,932	-1,73	-1,647	0,860	-1,92
<i>Atenção</i>	0,847	0,728	1,16	0,874	0,726	1,20
<i>Barulho</i>	-5,186	0,701	-7,40	-5,150	0,699	-7,37
<i>Falta_prof</i>	-15,136	0,984	-15,38	-15,204	0,981	-15,49
<i>Ed_física</i>	21,496	1,402	15,34	21,380	1,393	15,35
<i>NSE</i>	7,095	1,215	5,84	3,662	1,193	3,07
Nível 3:						
<i>Imigrantes_bairro</i>	0,859	2,816	0,31	1,309	2,589	0,51
Nível 4:						
<i>Intercepto (ver variância)</i>						
<i>Variância (resíduos)</i>	<i>variância</i>	<i>ICC</i>		<i>variância</i>	<i>ICC</i>	
<i>Nível 1 (Alunos):</i>	858,916	96,86%		872,41	98,31%	
<i>Nível 2 (Escolas):</i>	26,221	2,96%		14,99	1,69%	
<i>Nível 3 (Bairros):</i>	0,556	0,06%		0,00	0,00%	
<i>Nível 4: (Regiões):</i>	1,096	0,12%		0,00	0,00%	
Total:	886,789	100,0%		887,40	100%	

O segundo nível de maior explicação é o das escolas e sua variância fica entre 2,95% e 1,69% nos modelos 7 e 8. Os outros dois níveis possuem variância intraclasse desprezível. Na literatura, também se encontram grande parte da variância intraclasse dada pelo nível dos alunos. Em geral, essa variância é atribuída ao efeito escola é realmente baixa é também encontrada na literatura: César e Soares encontram valores entre e 12,3% e 36,9% entre o modelo mais especificado e o mais simples, o de efeito nulo. Albernaz, Ferreira & Franco (2002) encontram valores entre 6,6% e 28%, sendo este último valor encontrado no modelo de 2 níveis (alunos e escolas) e de menor especificação.

Nos modelos 7 e 8 destacam-se entre as variáveis do aluno a *dummy* de identificação do tipo de aluno (*Tipo_aluno*), presença de água no domicílio (*Água*), o indicador de repetência (*Repetiu*), *dummy* para identificação de qualidade da escola (*Qualidade*), indicada no nível aluno, variável para captar se o aluno faz o dever de casa (*Dever*), Bolsa família (*Bolsa_fam*) e estudo_mãe (*Estudo_mãe*), que acrescenta pontos de proficiência para cada nível a mais de escolaridade da mãe (exceção de *Estudo_mãe_d1*, indicadora de primário completo).

No nível das escolas, as variáveis mais importantes são presença de educação física em periodicidade igual ou acima da semanal (*Ed_física*), as faltas dos professores possuem efeitos negativos na proficiência média da escola (*Falta_prof*), o nível socioeconômico da escola (*NSE*), a frequência com que os estudantes relatam haver barulho na sala de aula (*Barulho*), que influencia negativamente a proficiência. Sobre o nível de bairros não houve variáveis significativas, embora o componente de imigração do parâmetro se apresente positivo.

Para ter uma ideia da dimensão da importância do valor dos parâmetros de cada variável é importante se retomar o valor do desvio-padrão da distribuição da proficiência que é de 45,27 pontos. Segundo o histórico de dados do PROEB de 2006 a 2009, a evolução de um desvio padrão em termos da média de proficiência leva cerca de quatro a cinco anos de estudos. Ou seja, se, por um exemplo, a média de proficiência do 5º ano é de 200,00 pontos, pode-se esperar que essa coorte alcance uma média de 245,27 apenas no 9º ano do fundamental. Isso fornece uma evolução de 9,5 pontos a cada ano de estudos. Aproximando o valor para 10 pontos ao ano, tem-se que a presença de educação física regular na escola avança o rendimento dos alunos em mais

de dois anos. Escolas com grande frequência de falta para os professores atrasam o desempenho em mais de um ano e meio.

Em relação às variáveis de principal interesse na tese: *migração* e *distância*, nota-se que a variável *migrante* têm influencia positiva. Nos dois modelos a variável *migrante* se refere ao aluno que estuda fora de um raio de 2 km de sua casa. Para o migrante de grupos de bairros e região administrativa essa variável não foi significativa. A migração contribui para o desempenho em valores que vão de 1,63 pontos a 2,66. Sobre o logaritmo da distância o resultado é significativo a 1% nos dois modelos ($t\text{-valor} \geq 2,32$) e o valor do parâmetro denota uma relação negativa entre distância e proficiência, tal como nos modelos MQO apresentados em 6.2.3.

Apesar do efeito da migração ser positivo quando analisado isoladamente, defrontando-se a *migração* com os efeitos de interação do tipo de aluno (se pertencente ao decil superior da distribuição de notas) e da qualidade da escola (escolas do decil superior da média de proficiência) nota-se que a interação entre o fato de ser migrante e o tipo de aluno não é significativa. O aluno do decil superior não sofre efeito da migração.

Em resumo, temos que a migração surte um efeito médio positivo, talvez denotando uma busca por escolas que agreguem mais valor da proficiência, dentro das possibilidades permitidas pelo sistema. No entanto, existe a contrapartida da distância que prejudica o desempenho de quem tem que se deslocar mais para ir à escola e o efeito líquido de se estudar em uma boa escola é positivo (por volta de 10 pontos a mais na proficiência, 15 pontos derivados do efeito escola -5 pontos perdidos em comparação aos alunos daquela mesma escola que não precisam migrar). Isso significa um efeito líquido de um ano a mais para o aluno que migra para uma boa escola. Acrescenta-se que os alunos de melhor desempenho não sofrem efeito significativamente diferente com a migração.

A seção 6.3.1. apresenta as fórmulas de decomposição dos fatores de desempenho que são importantes para as simulações do próximo capítulo e na seção 6.4 é apresentado, na TAB. 39, o quadro que resume os parâmetros obtidos neste capítulo.

6.3.1. Fórmulas de decomposição da proficiência

Relembrando as ideias expressas no capítulo 4 e considerando o modelo 7, sem interação entre alunos e escolas, a única coisa que uma alocação diferente, de fato, pode mudar são as distâncias que os alunos percorrem de casa até a escola e a migração.

Se por exemplo tomarmos um aluno i , morador do Norte de Belo Horizonte e matriculado em uma escola s perto de sua residência que não está entre as escolas do melhor decil, e o colocarmos em uma escola t , considerada entre as melhores da região Centro-Sul, tem-se três efeitos advindos dessa mudança: 1) o aluno se torna um migrante de região; 2) a nota do aluno recebe um incremento do efeito escola; e 3) a distância de deslocamento aumenta. Supondo-se que os dois primeiros efeitos sejam positivos, a nota obtida pelo aluno (y_i) será dada pelo efeito líquido da soma dos dois primeiros subtraído do efeito da distância. Representando isso por uma equação:

$$\Delta y_i = \Delta m_i + \Delta e_i - \Delta d_i.$$

Em que Δm_i simboliza o efeito de migração (positivo) do aluno i . O Δe_i (positivo) representa o efeito escola na nota e Δd_i (de efeito negativo) é o efeito da distância maior percorrida. Se $\Delta m_i + \Delta e_i > \Delta d_i$, tem-se um efeito positivo na proficiência do aluno $\Delta y_i > 0$.¹¹⁰

Obtendo-se os parâmetros relacionados às variáveis acima especificadas, ou um conjunto maior de variáveis relacionadas à escola, pode-se obter uma estimativa do efeito de se colocar o aluno i estudando na escola t .

No entanto, o efeito positivo acima pressupõe uma vaga ociosa na escola t , esse pressuposto pode não se verificar. Se o número de vagas estiver fixo, para colocar um aluno i em t , é preciso remanejar um aluno que já estava naquela escola para alguma outra. Suponha que o aluno j antes do remanejamento do exemplo anterior estudasse em t , após o remanejamento este aluno é realocado para s .

¹¹⁰ Os diferenciais são dados pelo resultado da variável do aluno i na nova escola t menos o resultado do mesmo aluno na escola anterior s . Para o caso da variável de proficiência temos: $\Delta y_i = y_{t,i} - y_{s,i}$.

Se o modelo para estimação dos efeitos é linear nos parâmetros e sem interação do aluno com a escola, temos que a mudança de j para a escola s acarreta em efeitos de migração e escola de magnitude inversa daqueles conseguidos pelo aluno i . Em termos formais temos que $\Delta y_j = -\Delta m_j - \Delta e_j - \Delta d_j$, e o efeito da distância continua sendo negativo, de modo que o diferencial de proficiência para o aluno j é negativo $\Delta y_j < 0$.

O efeito líquido da nova alocação é dado por:

$$\Delta y_i - \Delta y_j = \Delta m_i - \Delta m_j + \Delta e_i - \Delta e_j - \Delta d_i - \Delta d_j \quad (33)$$

Como os efeitos de migração e escola são os mesmos para todos os alunos e, no caso da nova alocação, esses alunos estão invertendo de escolas, podem-se anular os efeitos Δm e Δe . Todavia, os efeitos das distâncias operam no mesmo sentido. Por simplicidade, suponha-se que o diferencial da distância do aluno i até a escola t seja o mesmo diferencial do aluno j até a escola s , representado somente por Δd . Sendo assim, o efeito final da proficiência total Y é:

$$\Delta Y = -2\Delta d \quad (34)$$

Acrescentando-se o parâmetro de distância (β_1) tem-se que:

$$\Delta Y = -\beta_1 2\Delta d \quad (35)$$

De uma maneira geral, para casos onde as distâncias são diferentes entre os n estudantes, e i é o índice para os estudantes do conjunto $I = \{i_1, i_2, \dots, i_n\}$, tem-se o somatório:

$$\sum_{i=1}^n \Delta y_i = -\beta_1 \sum_{i=1}^n \Delta d_i \quad (36)$$

O diferencial da proficiência média pode ser obtido dividindo-se ambos os lados da equação (36) pelo número total de estudantes (n):

$$\frac{\sum_{i=1}^n \Delta y_i}{n} = \frac{-\beta_1 \sum_{i=1}^n \Delta d_i}{n} \quad (37)$$

No modelo 7, o diferencial na proficiência média é dado diretamente pelo diferencial na distância média dos alunos até a escola. Segundo esse modelo, a única maneira de se conseguir um efeito positivo na proficiência é reduzindo as distâncias. No exemplo do par de alunos (i, j) e escolas (s, t) as distâncias aumentaram, isso produziria uma perda de proficiência igual ao β_1 multiplicado pelo diferencial médio das distâncias (os valores dos β_1 obtidos para as simulações estão na TAB. 39 na seção 6.4).

Considerando o algoritmo de *Gale-Shapley*, o capítulo 7 apresenta que as distâncias dos alunos até a escola podem ser reduzidas em relação à alocação atual, gerando um efeito positivo na proficiência média. A maior parte dos ganhos com a introdução do sistema de cadastro em Belo Horizonte deve ser proveniente de um ganho de alocação e redução das distâncias dos alunos até as escolas, o capítulo 7 apresenta uma alocação simulada para o caso de preferências aleatórias, caso que pode ser representativo da situação anterior ao cadastro.¹¹¹

O modelo 7 pode ter outros efeitos positivos, desde que existam vagas ociosas nas melhores escolas, de modo que mesmo que os alunos enfrentem maiores distâncias, os efeitos escola compensem a perda devido a elas. Ao final do capítulo de simulações, considerações sobre esses aspectos serão analisadas.

Com ausência de vagas ociosas, o efeito de interação entre estudantes e escolas explicitado no modelo 8 é extremamente importante para a decomposição dos efeitos na proficiência. Nesse caso, desmembrando a equação (33) com parâmetros individuais para os alunos i e j vemos que os efeitos de migração e região não se anulam.

¹¹¹ De fato, sem informações prévias de onde moravam os alunos antes do sistema de cadastro e sem uma medição de proficiência para tal período é impossível se medir ao certo o quanto de impacto na proficiência um sistema de cadastro pode trazer. No caso de capitais brasileiras que ainda não possuem o sistema de cadastro implementado, como Vitória-ES e Manaus-AM (ver 2.6) é possível se fazer uma comparação antes e depois caso seja adotado esse procedimento. Para Belo Horizonte, o contrafactual só pode ser obtido por meio de simulações do que seria a situação prévia.

$$\Delta y_i - \Delta y_j = \beta_{3i}\Delta m_i - \beta_{3j}\Delta m_j + \gamma_{1i}\Delta e_i - \gamma_{1j}\Delta e_j - \beta_1(\Delta d_i + \Delta d_j) \quad (38)$$

Para simplificação da equação acima, suponha-se que a escola t tenha um efeito positivo para o aluno i , mas que a mudança para a escola s não produza impacto para o aluno j . Tem-se β_{3i} e γ_{1i} são positivos e β_{3j} e γ_{1j} são iguais à zero. Dessa maneira a equação (38) simplifica para:

$$\Delta y_i - \Delta y_j = \beta_{3i}\Delta m_i + \gamma_{1i}\Delta e_i - \beta_1(\Delta d_i + \Delta d_j) \quad (39)$$

Se o efeito de $\beta_{3i}\Delta m_i + \gamma_{1i}\Delta e_i > \beta_1(\Delta d_i + \Delta d_j)$ tem-se que o resultado líquido da proficiência será positivo. Se os efeitos positivos da migração e do efeito escola para o aluno i ($\beta_{3i}\Delta m_i + \gamma_{1i}\Delta e_i$) compensam uma possível perda em termos de uma maiores distâncias percorridas para os dois alunos, $\beta_1(\Delta d_i + \Delta d_j)$, tem-se que o efeito no diferencial da proficiência social (no caso de dois alunos i e j) é positivo.

Noutros termos: o que o aluno i ganha em proficiência (Δy_i) compensa sua própria perda com um maior deslocamento ($\beta_1\Delta d_i$) e a perda com deslocamento do outro aluno j , o aluno j perde em notas nesse caso, mas apenas o efeito do deslocamento. Como $\Delta y_i > \Delta y_j$, tem-se que o efeito líquido positivo na média de proficiência. Porém, como se ressaltou no capítulo 2, a questão ética envolvida nessa troca e as dificuldades normativas de se realizar esse *trade-off* entre alunos diferentes (capítulo 4) continuam presentes.

Revisitando o esquema apresentado no quadro da TAB. 7, desconsiderando as distâncias, temos uma situação que está no caso fracamente preferível de Pareto ótimo descrito na tabela, o aluno i tem um ganho de nota na nova escola e o aluno j não perde proficiência se mudando de escola. No entanto, considerando as distâncias, para isso ocorrer, o aluno j teria que ter um ganho na nova escola s que, pelo menos, anulasse sua perda com as maiores distâncias, caso isto não ocorra, tem-se a situação ambígua e que não é ótima de Pareto em relação a alocação inicial.

É claro que ainda possível obter o benefício de um encurtamento das distâncias que percorrem os alunos. Se essa diminuição das distâncias vier acompanhada de um efeito de interação positivo das escolas, o que se obtêm no modelo 8 é um reforço do modelo 7. No outro caso, apenas o

modelo 8 possui flexibilidade de propor alocações que aumentem as distâncias e ao mesmo tempo consigam melhorar o resultados de todos os estudantes.

O caso geral do efeito sob o modelo 8 pode ser descrito pela equação (40) abaixo:

$$\frac{\sum_{i=1}^n \Delta y_i}{n} = \frac{\beta_3 \sum_{i=1}^n \Delta m_i T_i}{n} + \frac{\gamma_1 \sum_{i=1}^n \Delta e_i T_i}{n} - \frac{\beta_1 \sum_{i=1}^n \Delta d_i}{n} \quad (40)$$

Em que i é o índice dos alunos e T_i é uma variável identificadora do tipo do aluno e assume valores 0 e 1, quando zero esse índice zera os efeitos de migração e escola desses alunos, se igual a 1, os efeitos dos parâmetros β_3 e γ_3 continuam presentes.

Outra maneira de se ver o mesmo resultado é por meio das equações de decomposição. Suponha duas regressões: uma para alunos do tipo não migrantes $M_i = 0$, alunos que não mudaram de escolas, e outra para o tipo $M_i = 1$, que mudaram de escolas:

$$Y_{iM_0} = \beta_{M_0} + \beta_{M_0} \mathbf{X}_{iM_0} + \varepsilon_{iM_0} \quad (M_i = 0) \quad (41)$$

$$Y_{iM_1} = \beta_{M_1} + \beta_{M_1} \mathbf{X}_{iM_1} + \varepsilon_{iM_1} \quad (M_i = 1) \quad (42)$$

Representando-se a média das variáveis por $E(\mathbf{X}_{iM_0})$, indicando a esperança das variáveis $\mathbf{X}$ para o grupo de indivíduos de $M_i = 0$, e sendo o mesmo válido para Y . De modo análogo, tem-se $E(\mathbf{X}_{iM_1})$ para o grupo $M_i = 1$. Supondo-se ainda que os erros em cada um dos modelos possuem esperança zero e mesma variância, tem-se que o diferencial entre as duas equações será dado por:

$$\Delta E(Y) = (\beta_{M_1} + \beta_{M_1} E(\mathbf{X}_{iM_1})) - (\beta_{M_0} + \beta_{M_0} E(\mathbf{X}_{iM_0})) \quad (43)$$

Somando e subtraindo $\beta_{M_0} E(\mathbf{X}_{iM_1})$ no lado direito da equação, podemos reescrever (43) como:

$$\Delta E(Y) = (\beta_{M_1} - \beta_{M_0}) + (\beta_{M_1} - \beta_{M_0}) E(\mathbf{X}_{iM_1}) + \beta_{M_0} (E(\mathbf{X}_{iM_1}) - E(\mathbf{X}_{iM_0})) \quad (44)$$

Que é a famosa decomposição de Oaxaca-Blinder.¹¹² O primeiro termo é o diferencial dado pela diferença de tipo entre os grupos. Caso esse diferencial não exista, o efeito será nulo. O segundo termo da equação do lado direito se refere ao diferencial dos efeitos de interação dos parâmetros de cada grupo, se existe efeito de interação ele estará manifesto na diferença entre essas duas variáveis.

O último termo da equação (44) trata de um efeito de composição das variáveis do aluno e da escola, mesmo que não haja interação entre o aluno a escola, o desempenho de um aluno com e uma escola com mais recursos pode ser melhor pelo fato desse diferencial ser mais favorável.

Na decomposição de Juhn, Murphy & Pierce (1993) é realizada uma diferença condicionada na distribuição dos erros das regressões de (41) e (42). Para observar como ocorre essa decomposição é preciso supor duas decomposições contrafactuais para Y_{iM_1} , que podemos denominar $Y_{iM_1}^1$ e $Y_{iM_1}^2$:

$$Y_{iM_1}^1 = \beta_{M_0} + \beta_{M_0} \mathbf{X}_{iM_1} + \vartheta_i \quad (45)$$

$$Y_{iM_1}^2 = \beta_{M_1} + \beta_{M_1} \mathbf{X}_{iM_1} + \vartheta_i \quad (46)$$

Em que ϑ_i representa os erros da equação (41) condicionados nos valores de $\mathbf{X}_{iM_1}$, o que pode ser também descrito como $\vartheta_i = F_{M_0}^{-1}(\theta_{iM_1} | \mathbf{X}_{iM_1})$, em que θ_{iM_1} é o percentil do indivíduo i do grupo M_1 , ou $M_i = 1$. Dado isso, da diferença de Y_{iM_1} e Y_{iM_0} , somando-se e subtraindo $Y_{iM_1}^1$ do lado direito da operação pode se representar a diferença nos valores de Y (ΔY) como:

$$\Delta Y = (\beta_{M_1} - \beta_{M_0}) + (\beta_{M_1} - \beta_{M_0}) \mathbf{X}_{iM_1} + \beta_{M_0} (\mathbf{X}_{iM_1} - \mathbf{X}_{iM_0}) + (\vartheta_i - \varepsilon_{iM_0}) + (\varepsilon_{iM_1} - \vartheta_i) \quad (47)$$

¹¹² Rodrigues (2009) desenvolveu em sua tese varias dessas decomposições para explicar diferenciais do SAEB entre os anos de 1997 e 2005, estudando os motivos da queda no início dos anos 2000. Rodrigues, Rios-Neto & Pinto (2011) empregam o método de densidades relativas para captar os efeitos de polarização na proficiência, respondendo ao questionamento da literatura sobre o efeito do *background* social dos alunos na média do SAEB.

A principal diferença da decomposição de Juhn, Murphy & Pierce (1993) para a decomposição de Oaxaca (1973) e Blinder (1973) é a que a decomposição da equação (47) não é mais realizada para a média dos valores de $\mathbf{X}_{iM0}$ e $\mathbf{X}_{iM1}$ (ou seja, em $E(\mathbf{X}_{iM0})$ e $E(\mathbf{X}_{iM1})$), mas sim para todos os valores das variáveis dos dois grupos, por isso, os termos dos erros não se anulam de imediato e há nesse tipo de decomposição a parcela dos resíduos.

Os dois primeiros termos do lado direito da equação (47) se referem ao diferencial de parâmetros entre as duas equações. O terceiro e quarto termos representam um efeito de composição, como mudam os efeitos das notas caso a composição das variáveis associadas aos alunos e escolas mudasse, nota-se que o efeito composição incorpora a diferença $(\vartheta_i - \varepsilon_{iM0})$, dado que essa composição não é feita nos valores médios de $\mathbf{X}_i$. O último termo $(\varepsilon_{iM1} - \vartheta_i)$ é dos erros não explicados, os resíduos.

A decomposição proposta por Juhn, Murphy & Pierce (1993) faz sentido para médias condicionadas em percentis da distribuição da variável de interesse. Desse modo, pode se apresentar tal decomposição dada em termos da esperança condicionada $E(\mathbf{X}_{iM1}|\theta_{iM1} = 0,9)$ para diferenças entre os grupos no 90º percentil, por exemplo.

Neste trabalho, o efeito de principal interesse é o do diferencial de parâmetros, manifestado em $(\beta_{M1} - \beta_{M0})$ e $(\beta_{M1} - \beta_{M0})$, isso porque o efeito de composição se anula por conta de, ao mesmo tempo, um grupo conseguir a troca de $\mathbf{X}_{iM0}$ para $\mathbf{X}_{iM1}$, e outro fazer a troca inversa, de $\mathbf{X}_{iM1}$ para $\mathbf{X}_{iM0}$. Em outras palavras, o contrafactual não é tão livre quanto em um exercício de decomposição regularmente realizado, ao se propor um aluno i estudando em uma escola s , em que já estuda j , é preciso colocar j na escola t em que i estudava antes da troca, ou fazer outro remanejamento.

Os efeitos de composição serão importantes para as simulações em que boas escolas possuem vagas excedentes. A seção 6.4 a seguir apresenta uma revisão dos parâmetros obtidos nesse capítulo e fornece um guia para as simulações importantes do capítulo seguinte.

6.4. Parâmetros estimados para as simulações e conclusões da análise descritiva

Como visto na seção anterior, os efeitos da troca de alunos e escolas podem proporcionar uma diferença na proficiência média adquirida. Essa diferença advém por meio das variáveis escolares, do efeito da migração, da diferença de distância da residência até a escola e do efeito de interação do aluno e escola. Esses efeitos foram batizados de efeitos *escola* (Δe), *migração* (Δm), *distância* (Δd) e *interação* (ΔInt).

Os modelos 7 (naïve) e 8 (interação) foram utilizados para fornecer os parâmetros para cada uma das simulações possíveis. A TAB. 39 abaixo mostra o valor desses dos parâmetros associados para cada um dos dois tipos de simulação: número de vagas fixo e excedente de vagas. No caso em que as vagas são fixas, tem-se que o efeito escola e migração não atuam, pois os ganhos se anulam. No caso da existência de vagas excedentes, alguns alunos podem ganhar ocupando as vagas ociosas.

Tabela 39 - Valores dos parâmetros estimados para uso nas simulações

		Efeito			
		<i>Escola</i> (Δe)	<i>Migração</i> (Δm)	<i>Distância</i> (Δd)	<i>Interação</i> (ΔInt)
Modelo 7	Vagas fixas	<i>Nulo</i>	<i>Nulo</i>	$-1,14 \cdot \Delta d_{iT_1}$	<i>Ausente</i>
	Vagas ociosas	$\gamma_i \cdot \Delta e_{iT_1}$	$1,62 \cdot \Delta m_{iT_1}$	$-1,14 \cdot \Delta d_{iT_1}$	<i>Ausente</i>
Modelo 8	Vagas fixas	<i>Nulo</i>	<i>Nulo</i>	$-1,22 \cdot \Delta d_{iT_1}$	$-5,39 \cdot \Delta Q_{iT_1}$
	Vagas ociosas	$\gamma_i \cdot \Delta e_{iT_1}$	$2,63 \cdot \Delta m_{iT_1}$	$-1,22 \cdot \Delta d_{iT_1}$	$-5,39 \cdot \Delta Q_{iT_1}$

Não é possível precisar de antemão quais são os valores de γ_i , pois cada escola possui um γ_s específico e esse valor dependerá para qual escola o aluno está se mudando. No anexo G existe a relação dos γ_s para cada escola do conjunto de escolas do banco de dados.

Para o efeito migração tem-se que esse efeito só existe nos modelos com excedente de vagas e os alunos migrantes não se anulam. Para o modelo 7, o valor desse parâmetro é 1,62 para todos os alunos, e no caso do modelo 8 é de 2,63.

Os efeitos das distâncias estão presentes nos dois modelos e assumem os valores apresentados na TAB. 38, apresentada na seção anterior. Os efeitos de interação advêm do fato do aluno aproveitar melhor uma escola de qualidade caso não precise migrar, os efeitos não se anulam, pois eles só valem na condição em que os alunos são migrantes e ao mesmo tempo estão em uma escola de qualidade.¹¹³

Os valores se concentram nos parâmetros de alunos e escolas, pois estes níveis foram os mais explicativos dentre os modelos utilizados, as variáveis de grupos de bairros e regiões explicam somente 0,28% da variância dos dados.

Nas simulações do capítulo 7 são construídos os valores limítrofes dos parâmetros, inserindo um intervalo de confiança. O capítulo abordará em detalhes como os pais dos alunos podem construir suas preferências estabelecendo uma função de preferência que pode levar em conta distâncias, características das escolas e interação de alunos e escolas. Diferentes formulações de preferências dos pais geram resultados diferentes para as alocações.

¹¹³ Da tabela 28 anteriormente apresentada sabe-se que as escolas no corte superior de proficiência possuem tendência de apresentarem proporção maior de alunos imigrantes. Pelos dados dos modelos hierárquicos apresentados na seção 6.3 a migração contribui para a proficiência na média. Porém, em escolas do primeiro decil não são os alunos imigrantes que se saem melhor.

7. RESULTADOS DAS SIMULAÇÕES

Neste capítulo são apresentadas as simulações que geram os resultados das alocações de alunos e escolas de acordo com quatro modelos de preferências dos pais/alunos e um modelo fixo de preferências (prioridades) para as escolas. Dados os modelos de *matching* propostos anteriormente no capítulo 3, e usando os dados elaborados descritos no capítulo 5 com os parâmetros do capítulo 6, têm-se as simulações de impacto na proficiência. Estes resultados podem ser utilizados para maior compreensão do sistema de cadastro escolar de Belo Horizonte e revelam características implícitas por trás dessa política.

Uma das principais simulações realizadas é a do algoritmo *Top Trading Cycles* (TTC). Esta simulação serve para representar a situação inicial, um *baseline* do que seria o estado anterior à implementação do sistema de cadastro escolar. Esse algoritmo simula a formação de filas pelas melhores escolas. Nesta simulação, as preferências dos pais/alunos são lexicográficas e consideram duas informações: os níveis de proficiência (y) e a proximidade da escola (d), tal como representado na expressão (48):

$$\mathcal{P}(i) = \min\{d_{is} | \bar{y}_s \geq 250\} > \min\{d_{is} | 250 > \bar{y}_s \geq 200\} > \min\{d_{is} | 200 > \bar{y}_s \geq 150\} > \min\{d_{is} | \bar{y}_s < 150\} \\ \forall i \in I, \forall s \in S \quad (48)$$

As preferências acima indicam que, dentre as escolas com média de proficiência avançada ($\bar{y}_s \geq 250$), todo e qualquer aluno i prefere a menor distância d_{is} possível entre ele e a escola (s). Esgotadas todas as alternativas das escolas com mais de 250 de proficiência média, os alunos preferem a escola mais próxima do nível intermediário, depois a segunda mais próxima, a terceira e assim sucessivamente. As preferências terminam na escola mais distante do nível abaixo do básico ($\bar{y}_s < 150$).

As preferências lexicográficas serão também utilizadas nas simulações com os algoritmos *deferred-acceptance many-to-one*. A diferença em relação à simulação com o TTC está nas preferências das escolas, que são totalmente aleatórias para esta simulação. As preferências das escolas

simulam uma possível ordem de chegada à fila, um processo similar à aleatoriedade. Pais que chegam primeiro na fila são escolhidos pelas escolas pela ordem de chegada.

O resultado do algoritmo TTC serve para se obter uma situação que seria anterior ao sistema de cadastro, o resultado da proficiência média nesse caso é 204,23 pontos, aproximadamente 9 pontos abaixo da proficiência da amostra. Os procedimentos dessa simulação serão mostrados na seção 7.1.

Todas as outras simulações propõem uma ordem de prioridade fixa para as escolas. Nesta ordem de prioridades proposta, as escolas priorizam os alunos que moram mais próximos. Outros ordenamentos podem ser sugeridos, tais como prioridade maior para os alunos que possuem irmãos que já estudam na escola, prioridades conforme a idade, composição de sexo e cor, ou demais características das crianças. Estes ordenamentos possíveis foram discutidos no capítulo 2. Nas simulações de *deferred acceptance* aqui realizadas, as prioridades das escolas serão únicas:

$$\mathcal{P}(s) = \min\{d_{si}\} \quad \forall i \in I, \forall s \in S \quad (49)$$

A definição da equação (49) acima afirma que para toda e qualquer escola s , as preferências são dadas pela menor distância entre s e o aluno i (d_{si}), em que i é dado pelo conjunto de todos os alunos I e s pelo conjunto de todas as escolas S disponíveis. Para as preferências de pais e alunos temos os seguintes modelos:

$$\mathcal{P}^1(i) = \min\{d_{is}\} \quad (49)$$

$$\mathcal{P}^2(i) = \max\{\bar{y}_s\} \quad (50)$$

$$\mathcal{P}^3(i) = \min\{d_{is} | \bar{y}_s \geq 250\} > \min\{d_{is} | 250 > \bar{y}_s \geq 200\} > \min\{d_{is} | 200 > \bar{y}_s \geq 150\} > \min\{d_{is} | \bar{y}_s < 150\} \quad (51)$$

$$\mathcal{P}^4(i) = \max U(Z_s, d_{is}) \quad \text{tal que: } U(Z_s, d_{is}) = \gamma_{0s} + \gamma \mathbf{Z}_s - \beta_1 d_i \quad (52)$$

$i \in I$ e $s \in S$.

O primeiro modelo de preferências ($\mathcal{P}^1(i)$) considera somente as distâncias, quanto menor a distância da escola s em relação ao aluno i , melhor a escola. Esse modelo é similar ao modelo de

prioridades para as escolas, no entanto, é preciso reparar que o ordenamento de preferências não é simétrico, devido à heterogeneidade da distribuição espacial de alunos e escolas, o aluno i_1 pode ter a escola s como a mais próxima de sua casa, mas ele mesmo não ser o aluno que mora mais próximo daquela escola.¹¹⁴

Este primeiro modelo de preferências é o que está implícito no atual sistema de cadastro da PBH, por ele, tanto pais quanto escolas devem preferir as escolas/alunos mais próximos. Pela proposição **P.1** do capítulo 3, na ausência de empates nas distâncias, esse tipo de preferências leva a uma alocação estável que é única no algoritmo de *deferred-acceptance*.

No segundo modelo, os pais preferem as escolas com média de proficiência mais alta, não considerando as distâncias. As escolas continuam priorizando os alunos que moram mais perto, mas pais e alunos tentam conseguir as escolas que possuem maior média de proficiência, independente das distâncias. Repare que esse tipo de preferência é muito forte, pois equivale a dizer que todos os alunos tem igual ordenamento pelas escolas, as simulações desse modelo com vagas excedentes levam a uma distância média da escola maior do que a observada.

No terceiro modelo, temos preferências lexicográficas de acordo com patamares de proficiência e as distâncias mínimas. Dentro dos níveis de proficiência, os pais e alunos preferem aquelas escolas que estão mais próximas. Nesse modelo os *trade-offs* são estanques, um aluno pode preferir estudar em uma escola muito longe de casa, desde que aquela escola esteja em um patamar superior do que uma escola mais próxima, mas de menor nível de proficiência.

No quarto e último modelo os pais possuem uma função de utilidade que considera parâmetros das escolas (Z_s), parâmetros tais como faltas dos professores, barulho, educação física e os outros descritos nos capítulos anteriores. Esse modelo supõe que os pais possuem acesso a essas informações e com isso decidem suas preferências de acordo com a possibilidade de notas que podem ser obtidas pelos seus filhos. Esse modelo também considera as distâncias, representadas pelo parâmetro negativo β_1 . Mais adiante, na seção 7.4, mais detalhes sobre essa função de utilidade serão fornecidos.

¹¹⁴ Ver um exemplo disso no anexo D.

A simulação proposta segue 7 passos:

Passo 1. Obtém-se o conjunto dos alunos.

Passo 2. Estipulam-se diferentes possibilidades para as preferências dos alunos (a matriz de preferências dos alunos).

Passo 3. Obtém-se um conjunto de escolas e se mantêm fixas as prioridades das escolas, todas iguais à menor distância possível, os alunos mais próximos são priorizados (matriz de prioridades das escolas). Com exceção da simulação de *baseline* em que as escolas não possuem prioridades e o processo de escolha da escola é aleatório.

Passo 4. Dadas as preferências, efetua-se o algoritmo de Gale-Shapley *deferred-acceptance* com alunos propondo (algoritmo TTC para a simulação de *baseline*).

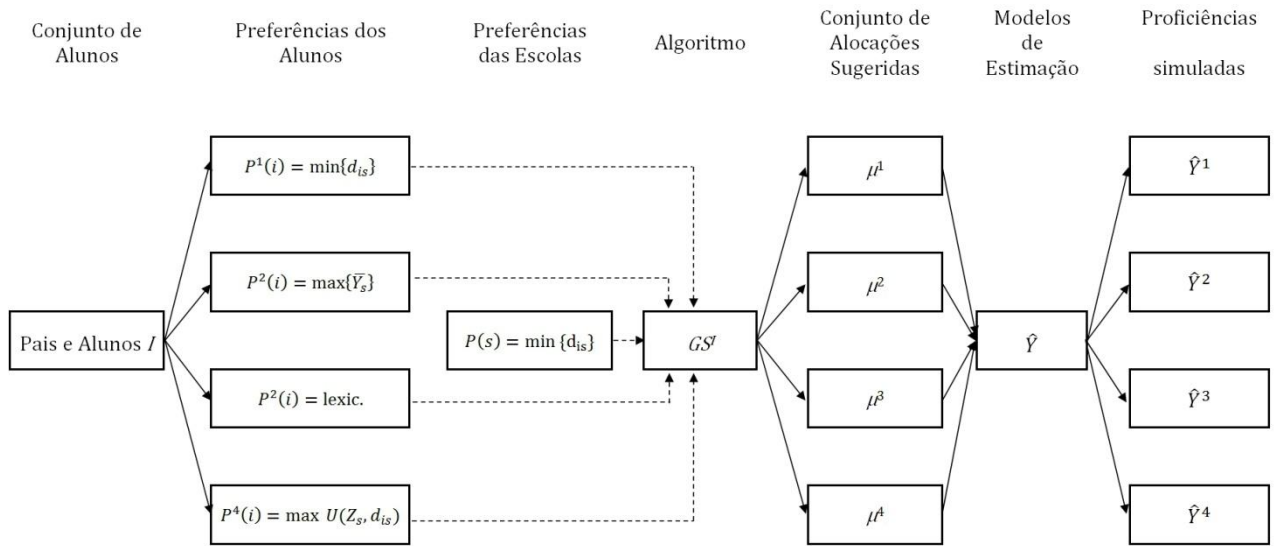
Passo 5. Após a conclusão do algoritmo, alocações novas são indicadas ($\mu^0, \mu^1, \mu^2, \mu^3, \mu^4$), cada uma dessas alocações constitui um conjunto de informações de alunos (i), escolas (s), grupos de bairros (k) e regiões (r) e tem-se uma tupla dada por (i, s, k, r) . O μ^0 representa a alocação obtida do TTC, o μ^1 é para a alocação do DA com preferências dos alunos e escolas dadas pelas distâncias, μ^2 alocação DA dada por preferências dadas inteiramente pelas proficiências das escolas, μ^3 com preferências lexicográficas e μ^4 com preferências dadas por uma função de utilidade.

Passo 6. Com os *matchings* simulados pelo algoritmo, aplicam-se os modelos de estimação com parâmetros para obtenção do $\hat{Y}$, obtém-se um desses para cada alocação do **passo 5**.

Passo 7. O último passo é a obtenção do resultado de proficiências médias a partir do $\hat{Y}$, isso é feito para cada alocação obtida no **passo 5** e para cada tipo de preferência dos pais.

A FIG. 12 abaixo representa esquematicamente o processo de simulação.

Figura 12 - Fluxo das etapas empregadas nas simulações das proficiências médias



Fonte: Elaboração própria.

Concentrando-se apenas na alocação de alunos e escolas (é possível saber o bairro e região a partir das informações de alunos e escolas), cada alocação μ fornece uma matriz $\mathbf{M}$ que reporta os *matchings* realizados, alunos nas linhas e escolas nas colunas. A matriz genérica representada em (53):

$$\mathbf{M} = \begin{matrix} & \begin{matrix} s_1 & s_2 & \cdots & s_m \end{matrix} \\ \begin{matrix} i_1 \\ i_2 \\ i_3 \\ \vdots \\ i_n \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \quad (53)$$

Cada simulação é comparada pela proficiência com a alocação real observada (μ_{Obs}) e com a simulação de TCC (μ^0), que serve de aproximação para uma situação inicial anterior ao sistema de cadastro. Da diferença entre a simulação proposta e os dados observados são tecidas as principais conclusões na seção 7.6.

Além das etapas de simulação descritas acima, cada simulação do algoritmo, sob determinadas preferências, é contrastada com dois diferentes cenários: 1) um cenário de vagas fixas, onde todas

as vagas do ensino público já estão ocupadas e não há como criar vagas excedentes; e 2) um cenário com 300 vagas ociosas em escolas predeterminadas.

Por motivos computacionais, a amostra de alunos nas simulações foi reduzida para 2.173 alunos. Todas as 296 escolas do banco processado estão ainda presentes na amostra, no entanto, a amostra foi reduzida proporcionalmente em todas as escolas. As simulações a seguir consideram essa amostra.¹¹⁵

Por conta dessa amostra menor, os valores das variáveis do banco observado possuem valores diferentes em relação ao banco maior trabalhado no capítulo 6. Em especial, na variável de distância, possui nova média de 4,2 Km, maior do que a média de 1,7 Km da amostra de 16.354 alunos. O valor da média de proficiência é 213,19, que continua bem próximo do valor anterior de 213,06.

Este capítulo apresenta cada simulação separadamente realizando uma pequena discussão para cada uma delas. Ao final, as simulações são agrupadas para uma análise do conjunto. Em 7.6 são propostas mais duas simulações visando projeções para políticas públicas. O resultado para a perspectiva dos alunos é também analisado em maiores detalhes em seção específica em 7.7. Na última seção (7.8), são apontados alguns comentários sobre possíveis consequências e interpretações dos modelos propostos.

¹¹⁵ Com um sistema operacional Windows 7, 64 Bits, processador Intel® Core i5, CPU 650, 3,21 GHz e 12 GB de memória RAM, e versão da linguagem R.2.15.12, cada simulação do algoritmo para a amostra de 2.173 alunos levava em torno de 15 a 30 minutos. Porém o tempo computacional cresce polinomialmente com o tamanho da amostra sendo que a amostra completa de 16.354 alunos levava vários dias para ser completada. Para maiores detalhes dos procedimentos computacionais do algoritmo e uma discussão sobre possíveis maneiras de se aumentar a eficiência do código implementado, ver Anexo B.

7.1. Simulação do *baseline* (*Top Trading Cycles*)

A proposta dessa etapa de simulação é apresentar um modelo que se aproxime melhor de uma situação descentralizada de escolha das matrículas. Com isso, o modelo proposto baseia-se no algoritmo de *Top Trading Cycles* descrito no capítulo 4. Nesse modelo, pais e alunos buscam a escola preferida de acordo com preferências lexicográficas. Assume-se que o processo de formação de preferências é lexicográfico por uma consideração entre proficiência e distância já exposta na introdução do capítulo (Eq. 51). Para esse modelo tem-se a seguinte relação de preferências para os alunos:

$$\mathcal{P}(i) = \min\{d_{is} | \bar{y}_s \geq 250\} > \min\{d_{is} | 250 > \bar{y}_s \geq 200\} > \min\{d_{is} | 200 > \bar{y}_s \geq 150\} > \min\{d_{is} | \bar{y}_s < 150\} \\ \forall i \in I, \forall s \in S \quad (51)$$

Essas preferências retratam que pais preferem, dentre as escolas públicas, aquelas escolas que são mais conceituadas em termos de proficiência e ensino. No entanto, consideram também as distâncias de casa até a escola. As preferências lexicográficas são uma maneira simples de se compor um ordenamento com duas variáveis e livre de parâmetros para as utilidades.

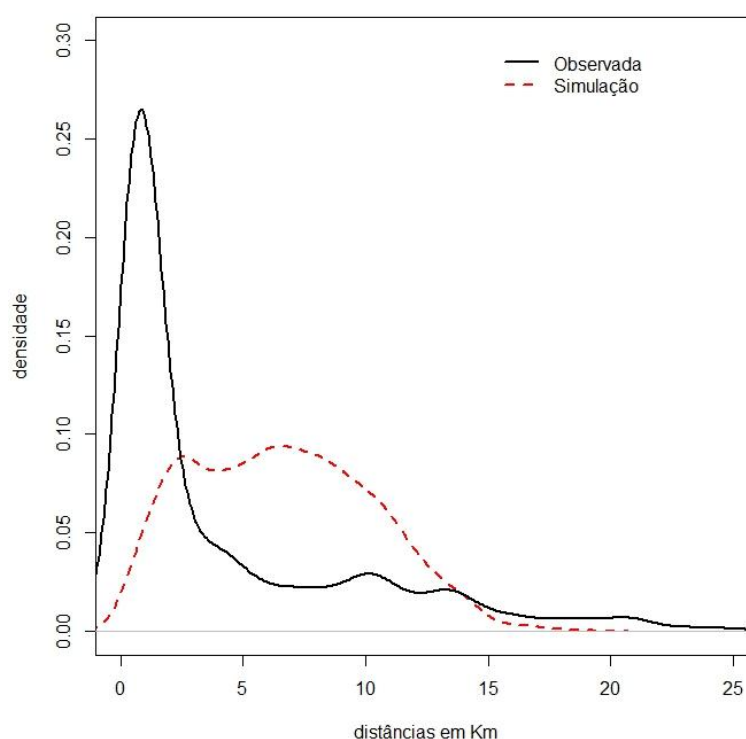
Na simulação de *baseline*, as escolas são indiferentes a qualquer aluno que se apresente para a matrícula. Essa indiferença equivale a sortear os alunos em uma loteria. Em verdade, nunca foi feita uma loteria para a entrada dos alunos, no entanto, o sorteio é aqui proposto como uma ordem aleatória de chegada dos pais nas filas, processo que ocorria antes da implantação do sistema de cadastramento. A posição de uma mãe ou responsável na fila era aleatória e a escola prosseguia preenchendo suas vagas até que as vagas se esgotassem ou até acabar a fila.

O algoritmo TTC obtém uma matriz de alocação para os alunos chamada de μ^0 , por esta alocação obtém-se o valor da proficiência média que obteriam os alunos de 2010 caso as matrículas se dessem por um critério similar ao sistema de filas que vigorava antes do cadastro. A principal diferença de resultado ocorre nas distâncias médias percorridas pelos alunos. Elas aumentam em 2,4 km, e essa diferença impacta na média de proficiência obtida.

Essa diferença ocorre por conta dos pais priorizarem o nível das notas. Porém uma consequência do algoritmo de TTC é que pais que tiveram má sorte no lugar da fila ficam em escolas muito distantes e de nível de proficiência baixo, isso prejudicaria o desempenho médio dos estudantes.

O GRAF. 15 abaixo apresenta as densidades dos dados observados e da simulação. Nota-se que a distribuição de distâncias observada é bastante concentrada nas menores distâncias, uma consequência do sistema de cadastro. Caso os pais adotassem uma preferência lexicográfica, tal como a descrita em (51) e as escolas fossem indiferentes à ordem de chegada dos alunos, teríamos uma distribuição platicúrtica das distâncias (em vermelho no gráfico).

Gráfico 15 - Densidade de kernel para a distribuição das distâncias observada e simulada



Fonte: Cálculos com uso do banco de dados processado.

A TAB. 40 apresenta os resultados para a proporção acumulada das distâncias da distribuição real e da simulada para o *baseline*. Sobre a proficiência, observa-se que o efeito estimado para o sistema de cadastro, dado pela diferença entre a média de proficiência observada e a simulada, chega a 9 pontos em língua portuguesa, ver TAB. 41. Essa diferença pode chegar a quase 12

pontos se se considerar o valor inferior do intervalo de confiança da simulação μ^0 (213,19 – 201,28).

Tabela 40 - Proporção de alunos por distância em Km Antes e depois da simulação

Distância (em km)	Proporção acumulada de alunos ($\leq$) μ_{Obs}	Proporção acumulada de alunos ($\leq$) após simulação μ^0
1	40%	4%
2	58%	11%
5	72%	36%
10	84%	81%
15	95%	99%
25	100%	100%
30	100%	100%
Média da Distância	4,18	6,60

Fonte: Banco de dados processado e amostra do mesmo banco (2010).

Tabela 41 - Diferença das proficiências entre a amostra e a simulação do baseline

	Média Proficiência	Desvio padrão	IC_inferior	IC_superior
Amostra (μ_{Obs})	213,19	45,62	210,25	216,13
Simulação Baseline (μ^0)	204,25	45,79	201,28	207,17

Fonte: Cálculos com o uso do banco de dados processado (2010).

Os resultados da simulação representados na TAB. 41 foram obtidos do modelo 7 com vagas fixas, os parâmetros podem ser novamente consultados na TAB. 39. Para a simulação do *baseline* não se supôs presença de vagas ociosas. O modelo de obtenção da nota é uma relação linear simples que considera a perda de proficiência dado o aumento das distâncias. A formula é dada por:

$$\hat{y} = E(y) - 1,14 \cdot \log(E(d^0) - E(d)) \quad (54)$$

Em que $\hat{y}$ é o valor estimado para a proficiência dada a nova alocação sugerida pelo algoritmo, y é a variável de proficiência original do banco de dados, -1,14 é o valor do parâmetro beta para as

distâncias (obtido do modelo 7 do capítulo anterior), d^0 são as distâncias de casa até a escola obtidas pela nova alocação simulada e d são as distâncias originais da amostra observada.

Pelos resultados obtidos da simulação, temos que a alocação atual (a alocação real observada em 2010 para Belo Horizonte) confere, em média, 9 pontos a mais para cada aluno, pois os aloca mais eficientemente (considerando-se que o efeito da distância é negativo para a proficiência).

Optou-se, nesta simulação, pelo modelo mais simples que considera apenas o efeito negativo das distâncias. Poderiam ser incluídos os efeitos de interação entre alunos e escolas. No entanto, os efeitos dessa interação são mais complexos, sendo que o efeito da distância é o mais importante nesse caso. Como o número de vagas nas escolas está fixo, não há efeito de escolas ou de migração, se um aluno conseguiu uma escola melhor pelo algoritmo de TTC, outro que estava naquela escola teve de ceder a vaga.

É interessante observar que a simulação quase rebaixa a proficiência média de Belo Horizonte do nível intermediário para o básico (valor de corte $y = 200$). O valor de 204,25 obtido para a proficiência será o utilizado para comparar também com outras simulações desenvolvidas nas próximas seções.

7.2. Simulação considerando apenas as distâncias

Supondo-se que os pais e alunos estabelecem suas preferências apenas pela proximidade das escolas, tem-se uma nova conformação de preferências. Esse é o sistema de preferências ideal para o caso em que as escolas são homogêneas e a distância insere um fator de custo em termos de proficiência importante a ser ponderado pelos alunos. Nessa simulação temos que a preferências dos pais é dada de acordo com a equação (49) da introdução deste capítulo:

$$\mathcal{P}^1(i) = \min\{d_{is}\}, \forall i \in I \text{ e } \forall s \in S \quad (49)$$

As escolas possuem prioridades dadas pelas distâncias. A simulação da alocação deste modelo é aquela que minimiza as distâncias possíveis para os alunos em um *matching* estável. A média de

distâncias ao final desta simulação é 3,74 Km, no caso em que o número de vagas é fixo, e 2,61 Km para o caso em que há escolas com excedentes de vagas (TAB.42). O μ_1^* representa a simulação no cenário de vagas ociosas. No GRAF. 16, das densidades, vê-se que na simulação as distâncias ficam muito mais concentradas em valores menores do que 2 km.

Tabela 42 - Diferença das distâncias entre a amostra e a simulação por distâncias

	<i>Média Distância</i>	<i>Desvio padrão</i>	<i>IC_inferior</i>	<i>IC_superior</i>
Simulação Baseline (μ^0)	6,60	3,61	6,37	6,83
Amostra (μ_{Obs})	4,18	5,25	3,84	4,52
Simulação sem vagas ociosas (μ^1)	3,74	6,95	3,29	4,18
Simulação com vagas ociosas (μ^{1*})	2,61	4,69	2,31	2,92

Fonte: Cálculos com o uso do banco de dados processado (2010).

A simulação sem vagas ociosas é obtida rodando o algoritmo de Gale-Shapley com 2.173 alunos e número fixo de vagas nas 296 escolas. O número médio de vagas em cada escola é pouco maior que 7 alunos por escolas.¹¹⁶ Naturalmente algumas escolas possuem mais alunos que a média e outras menos. O resultado final do algoritmo gera uma alocação μ^1 como resultado.

Dada a alocação μ^1 , pode-se obter o valor médio para a estimação de proficiência. Para a estimação utilizam-se os dois modelos e dois cenários apresentados na TAB. 33 do capítulo anterior. O modelo 7 não supõe interação entre se o aluno é migrante (se migrou para uma escola que está a mais de 2 km de casa) e o tipo de escola (se a média de proficiência da escola está no decil superior de proficiência), o modelo 8 supõe que há essa interação:

$$\text{Modelo 7 sem vagas ociosas: } \hat{y} = E(y) + 1,14 \cdot \log(E(d_i - d_i^1)) \quad (55)$$

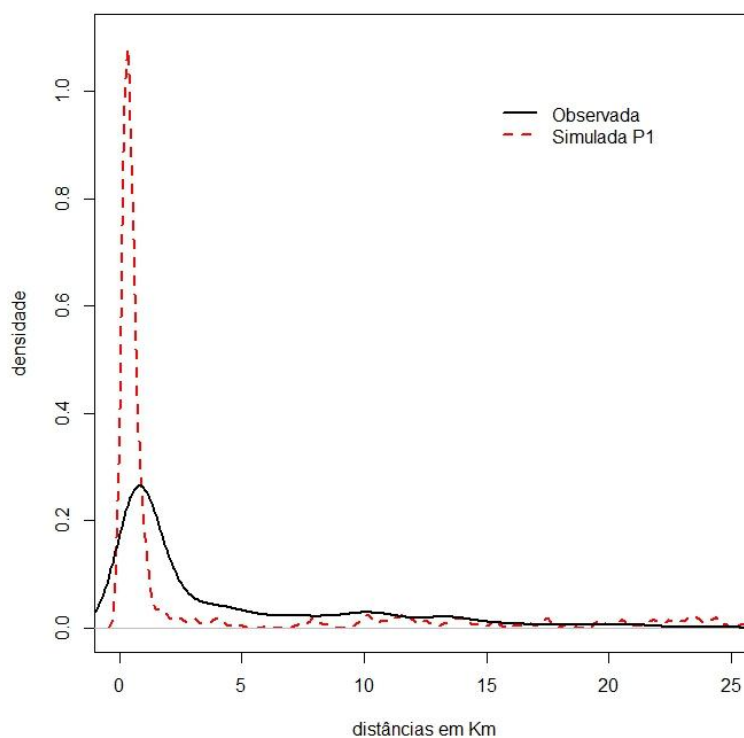
$$\text{Modelo 8 sem vagas ociosas: } \hat{y} = E(y) + 1,22 \cdot \log(E(d_i - d_i^1)) - 5,54 \cdot E(m_i Q_i) \quad (56)$$

¹¹⁶ No banco de dados completo do capítulo 6, o número médio para o tamanho das séries é pouco maior que 55 alunos por escola.

Para os quais m_i é a variável que indica se aluno migra de bairro (que foi a variável de migração considerada para as simulações). Essa variável de migração interage com a qualidade da escola Q_i , que denota se a escola está no decil superior de proficiência média em Belo Horizonte. Vê se que o efeito é negativo, conforme TAB.39, ou seja, o aluno migrante para uma boa escola tem desempenho inferior ao aluno não migrante que estuda em $Q_i = 1$. Lembrando que esse efeito isolado de Q e m não é contabilizado no modelo de vagas fixas, pois enquanto um aluno ganha o efeito o outro perde.

Tem-se que d_i são as distâncias por aluno do banco original, d_i^1 são as distâncias por aluno após a simulação. Como as distâncias após a simulação são menores do que as distâncias originais, pode-se notar que o sinal do parâmetro das distâncias (1,14 e 1,22) é positivo para os dois modelos, pois as simulações possuem efeito de reduzir as distâncias.

Gráfico 16 - Densidade de kernel para a distribuição das distâncias observada e da simulação com preferências dadas pelas distâncias



Fonte: Cálculos com uso do banco de dados processado.

O modelo 8 acrescenta ainda os efeitos de interação entre migração e qualidade das escolas. Para entender melhor o modelo, convém voltar à interpretação do parâmetro de interação entre migração e qualidade da escola. A qualidade esta indicada por escolas que estão no decil superior de proficiência média. Ao migrar para uma boa escola o aluno ganha um efeito escola positivo (rever modelos do capítulo 6), além disso, tem-se que, em geral, alunos migrantes possuem proficiência um pouco mais alta que seus pares não migrantes. Porém, ao migrar para uma escola do decil superior, o aluno médio e migrante, em geral, não se sai tão bem quanto aqueles que são não migrantes que estudam em escolas do primeiro *decil*. Os fatores deste efeito precisam ser mais bem compreendidos.¹¹⁷

Para o modelo com vagas ociosas fez-se a suposição de que escolas do primeiro decil poderiam ter mais vagas (com a ideia de criar mais vagas nas melhores escolas). Criou-se 10 vagas a mais para as 30 melhores escolas do banco.

Modelo 7 com vagas ociosas:

$$\hat{y} = E(y) + E(\Delta\beta_i) + 1,62 \cdot (E(m_i^1 - m_i)) + 1,14 \cdot \log(E(d_i - d_i^1)) \quad (57)$$

Modelo 8 com vagas ociosas:

$$\hat{y} = E(y) + E(\Delta\beta_i) + 2,63(E(m_i^1 - m_i)) + 1,22 \log(E(d_i - d_i^1)) - 5,54 \cdot E(m_i \cdot Q_i) \quad (58)$$

Em relação à (55) e (56) os modelos com vagas ociosas acrescentam dois efeitos que antes se anulavam, quais sejam, o efeito escola representado pela diferença dos Betas que o aluno obtêm na nova escola ($\Delta\beta_i$).¹¹⁸ E o efeito de migração em que m_i^1 é o caso do aluno migrante na simulação, e m_i capta se ele já era migrante ou não na situação anterior.

¹¹⁷ Ocorre também que o fator de redução de distâncias na maioria dos modelos aqui empregados faz com que as distâncias se reduzam e a migração diminua, perdendo se o seu efeito positivo em alguns casos. Ocorre então um *trade-off*, diminuir as distâncias e perder em efeito migração, ou aumentar a migração e perder em termos de efeito distância.

¹¹⁸ Para ver os valores dos Betas consulte o anexo G.

Na simulação com mais vagas nas escolas, um interessante efeito ocorre: a maioria das escolas onde se criaram mais vagas conseguiu preencher todas as vagas excedentes, ou seja, havia ainda demanda pelas melhores e mais bem localizadas escolas. Algumas escolas não preenchem todas as vagas e outras ficam vazias, indicando que elas não estão muito bem posicionadas em termos de distâncias para os alunos da amostra. O GRAF. 17 mostra a utilização das escolas dado por um indicador de aproveitamento que relaciona quantos alunos estão na escola após o algoritmo e quantas vagas a escola possuía antes de serem criadas mais vagas (*índice de utilização = alunos/vagas*). Se o indicador é zero, significa que as escolas ficaram vazias.

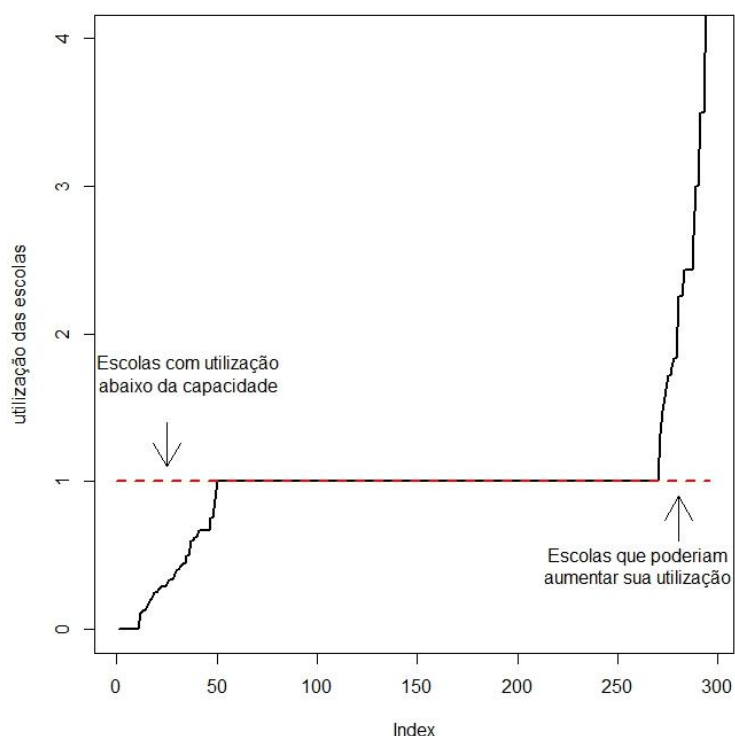
Na simulação, 11 escolas ficam completamente vazias, outras 38 estão abaixo da sua capacidade máxima. Um total de 221 escolas faz uso pleno da sua capacidade e outras 26 são sobredemandadas. Essas escolas mais demandadas possuem procura suficiente para serem utilizadas com mais eficiência caso tivessem, de fato, mais vagas. Observa-se que um total de 49 escolas estão muito distantes dos alunos da amostra, indicando que são pouco preferidas e trabalham com ineficiência. E 4 escolas para as quais se criaram mais vagas não foram sobredemandadas, para essas escolas não faz sentido criar mais vagas do que as já existentes.

A grande maioria da amostra (75%) é formada por escolas que estão com um número adequado de alunos, trabalhando eficientemente. São as escolas que possuem índice de utilização igual a 1.

A consequência dos modelos com mais vagas é que eles permitem verificar onde as escolas poderiam ser criadas para atender melhor as preferências de pais e alunos, além disso, no caso em que existem vagas ociosas, há espaço para ganhos de valor na proficiência, pois a entrada de mais alunos em uma escola não implica remanejamento de alunos que já estavam matriculados nessa escola. Em outras palavras, é possível uma melhoria de Pareto ao alocar mais alunos para escolas que eventualmente são mais demandadas.

Os ganhos de proficiência no modelo com mais vagas podem ser, portanto, maiores. Isso se deve em parte por conta de que os efeitos escola e migração não se anulam, e por outra parte porque é possível reduzir ainda mais as distâncias percorridas pelos alunos.

Gráfico 17 - Índice de utilização de escolas, simulação por distâncias e com vagas ociosas



Fonte: Cálculos com uso do banco de dados processado.

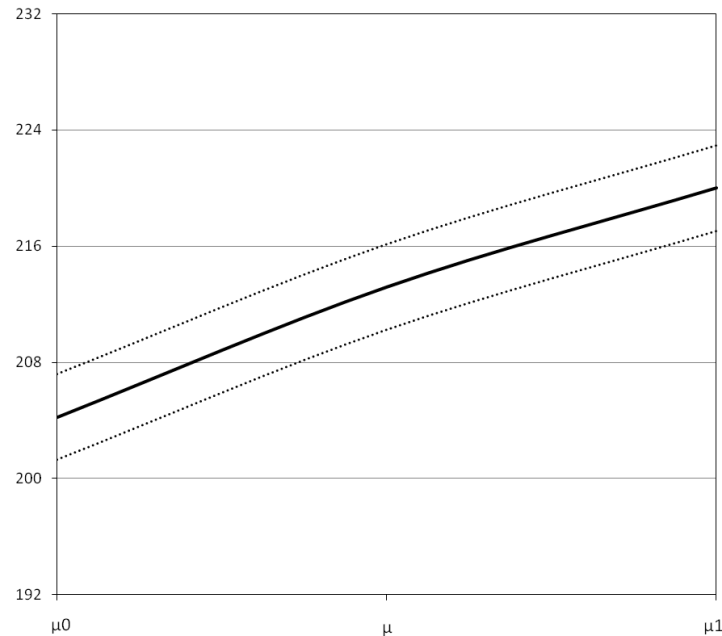
Como conclusão da simulação com $\mathcal{P}^1(i)$ temos que a alocação sugerida μ^1 produz um efeito na proficiência de cerca de 7 pontos para os modelos com vagas fixas (sem vagas ociosas) e de cerca de 10 pontos para o modelo com vagas ociosas. Essa diferença se deve principalmente pela redução das distâncias, e um segundo motivo é o efeito da mudança das escolas (efeito escola).

A TAB. 43 mostra as diferenças na proficiência com a nova simulação para o modelo 8 sem vagas ociosas e compara com a amostra e com a simulação do *baseline*. O GRAF. 18 apresenta os mesmos dados. A TAB. 44 apresenta a distribuição acumulada para as distâncias após a simulação.

Tabela 43 - Diferença das proficiências entre a amostra e a simulação por distâncias

	<i>Média Proficiência</i>	<i>Desvio padrão</i>	<i>IC_inferior</i>	<i>IC_superior</i>
Simulação μ^1	219,99	45,58	217,06	222,93
Amostra (μ_{Obs})	213,19	45,62	210,25	216,12
Simulação Baseline (μ^0)	204,25	45,79	201,28	207,17

Fonte: Cálculos com o uso do banco de dados processado (2010).

Gráfico 18 - Relação da proficiência e as alocações simuladas

Fonte: Cálculos com uso do banco de dados processado.

Tabela 44 - Proporção de alunos por distância em Km Antes e depois da simulação μ^1

Distância (em km)	Proporção acumulada de alunos ($\leq$) μ_{Obs}	Proporção acumulada de alunos ($\leq$) após simulação μ^0	Proporção acumulada de alunos ($\leq$) após simulação μ^1
1	40%	4%	71%
5	72%	36%	81%
10	84%	81%	83%
15	95%	99%	90%
25	100%	100%	98%

Fonte: Banco de dados processado e amostra do mesmo banco (2010).

7.3. Simulação considerando apenas as proficiências médias

Na segunda simulação, a preferência de pais e alunos é estabelecida pelas proficiências médias das escolas. Quanto maior a proficiência média da escola, mais desejada ela se torna. As informações de médias escolares são disponibilizadas na internet. Em tese, se pais considerassem apenas essa informação, teríamos uma preferência do tipo representado por $\mathcal{P}^2(i)$ na equação (50) reescrita abaixo:

$$\mathcal{P}^2(i) = \max\{\bar{Y}_s\}; \forall i \in I, \forall s \in S \quad (50)$$

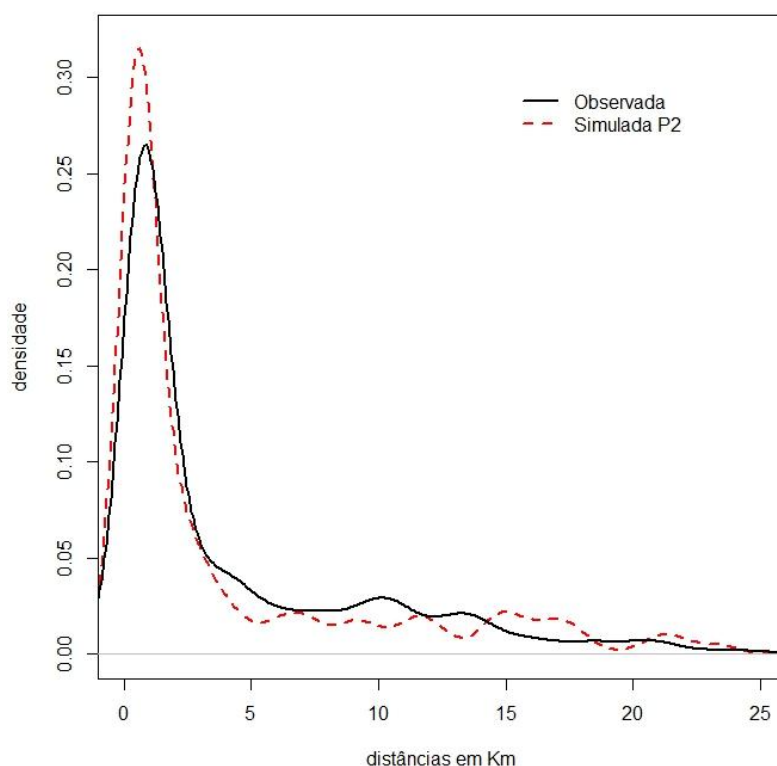
Nesse tipo de preferências, os alunos não se importam com as distâncias a serem percorridas, desde que a média da escola seja mais elevada. Em verdade, se a informação sobre as médias de cada escola são publicamente disponíveis e todos os pais possuem acesso, as preferências de cada aluno se manifestará igual entre todos os participantes do sistema.

Pode se imaginar, por exemplo, pais que ranqueiam as escolas de acordo com o IDEB (*Índice de Desenvolvimento da Educação Básica*), embora o IDEB considere, além da proficiência, a taxa de atendimento, esse indicador daria uma ideia de como pais podem formar preferências sobre escolas públicas.

Considerando que as prioridades das escolas continuam constantes e dadas pelas menores distâncias, o resultado desse algoritmo será próximo de uma alocação que possui as escolas decidindo (propondo primeiro) pelos alunos mais próximos, isso se deve ao fato das preferências de todos os alunos serem iguais.

Como demonstra o GRAF. 19, a distribuição de distâncias obtida da alocação μ^2 é similar à densidade já observada na amostra. As diferenças devem ocorrer principalmente pelas diferenças entre as formas de se obter as distâncias presente nos dados, diferente da dos dados aqui empregados. A média em quilômetros da simulação é de 4,04, pouco menor do que a observada na amostra (4,18).

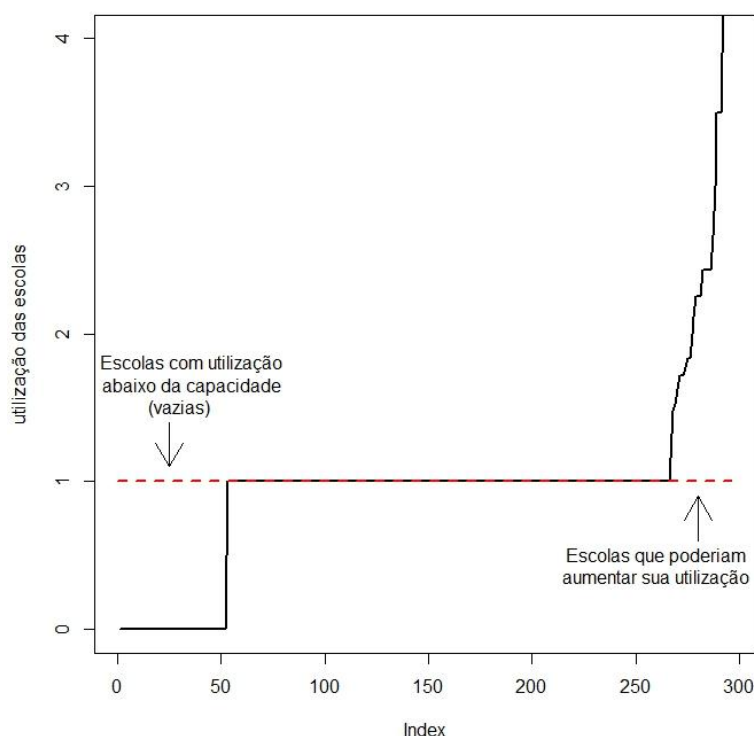
Gráfico 19 - Densidade de kernel para distâncias observadas e da simulação com preferências dadas pelas proficiências



Fonte: Cálculos com uso do banco de dados processado.

Uma consequência do modelo com preferências dadas pelas proficiências médias é que ao se acrescentar mais vagas, a média das distâncias aumenta, e o número de escolas que ficam vazias é maior. Isso ocorre porque agora pais preferem escolas com melhor desempenho, sem considerar as distâncias. Usando novamente o índice de utilização das escolas (alunos/vagas) após se acrescentarem mais vagas, o GRAF. 20 mostra esse resultado da utilização das escolas.

Na simulação com vagas ociosas, 52 escolas ficam vazias, 214 operam em plena capacidade e todas as 30 escolas para as quais se criaram mais 10 vagas aumentam seu coeficiente de utilização. Se as preferências dos pais levam em conta a proficiência média, a simulação sugere que o ideal é criar mais vagas nas escolas com melhor desempenho.

Gráfico 20 - Índice de utilização de escolas, simulação por proficiência e com vagas ociosas

Fonte: Cálculos com uso do banco de dados processado.

Em relação à proficiência na segunda simulação, o ganho é um pouco menor do que do que a simulação por distâncias (ver TAB. 45), isso ocorre principalmente porque as distâncias são um pouco maiores em μ^2 e o efeito de migração não consegue compensar essa perda.

Tabela 45 - Diferença das proficiências entre a amostra e a simulação por proficiência

	<i>Média Proficiência</i>	<i>Desvio padrão</i>	<i>IC_inferior</i>	<i>IC_superior</i>
Simulação μ^2	218,69	45,57	215,76	221,62
Simulação μ^1	219,99	45,58	217,06	222,93
Amostra (μ_{Obs})	213,19	45,62	210,25	216,12
Simulação Baseline (μ^0)	204,25	45,79	201,28	207,17

Fonte: Cálculos com o uso do banco de dados processado (2010).

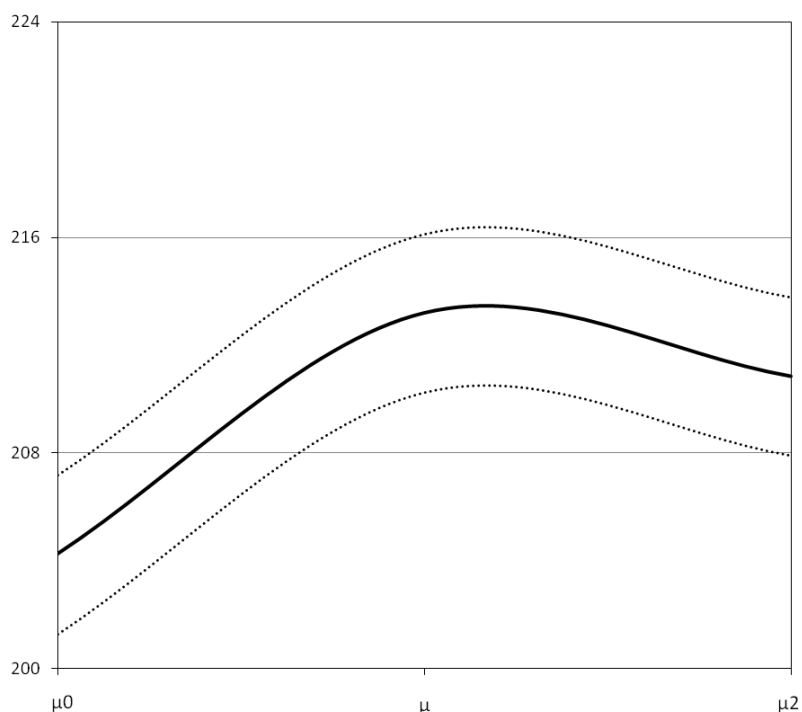
Obs.: A simulação da proficiência foi realizada com o modelo 6 sem vagas ociosas.

Na simulação com excedente de vagas não há ganhos na proficiência média. Ela se torna próxima da proficiência observada na amostra, o GRAF. 21 mostra essa situação. Os alunos que se deslocam para as melhores escolas poderiam ganhar um efeito escola maior, porém estariam nessa situação pagando um custo maior em termos das distâncias. E a grande maioria dos alunos não estaria ganhando o suficiente para compensar o efeito negativo das distâncias.

Esta simulação se emparelha, em termos de distâncias, com a amostra do banco, sendo que seu resultado de proficiência se situa em torno do ganho de proficiência já observado. Dessa maneira, comparando-se a simulação com vagas ociosas com preferências ditadas pelas proficiências médias, tem-se que a alocação observada já é a alocação ótima.

Considerando-se esse resultado, as preferências dadas pelas proficiências médias das escolas não parecem razoáveis do ponto de vista das famílias. Em grande parte, pais sabem que existem escolas que substituem umas as outras, uma escola teria de ter um diferencial de agregar muito valor para que compensasse os alunos perderem tempo se deslocando uma grande distância.

Gráfico 21 - Relação da proficiência e as alocações simuladas (μ^2) com vagas excedentes



Fonte: Cálculos com uso do banco de dados processado.

7.4. Simulações considerando preferências lexicográficas

Preferências lexicográficas são importantes para se determinar uma relação de prioridades, trata-se de um ordenamento pré-estabelecido de acordo com regras específicas. Se os pais tivessem em casa uma planilha do Excel com as informações da proficiência média e a das distâncias de cada escola considerada (escolas nas linhas e as informações de proficiência e distâncias nas colunas), um procedimento de escolha da melhor escola seria: “ordenar em ordem decrescente pela coluna de proficiência e acrescentar um nível de ordenamento para a coluna das distâncias na ordem crescente (esse seria o critério de desempate), a escola na primeira linha é a mais preferida, a da segunda é a segunda mais preferida e assim por diante”.

Dessa forma, para a escolha entre duas escolas com média de proficiência igual a 250, os pais devem indicar como mais preferida aquela que está mais perto. Preferências lexicográficas não diferem muito de uma relação hierárquica de prioridades. Tal como no exemplo acima, digamos que pais e alunos preferem a melhor escola mais próxima, para isso eles precisaram estabelecer qual é o critério para dizer que uma escola é “melhor”.

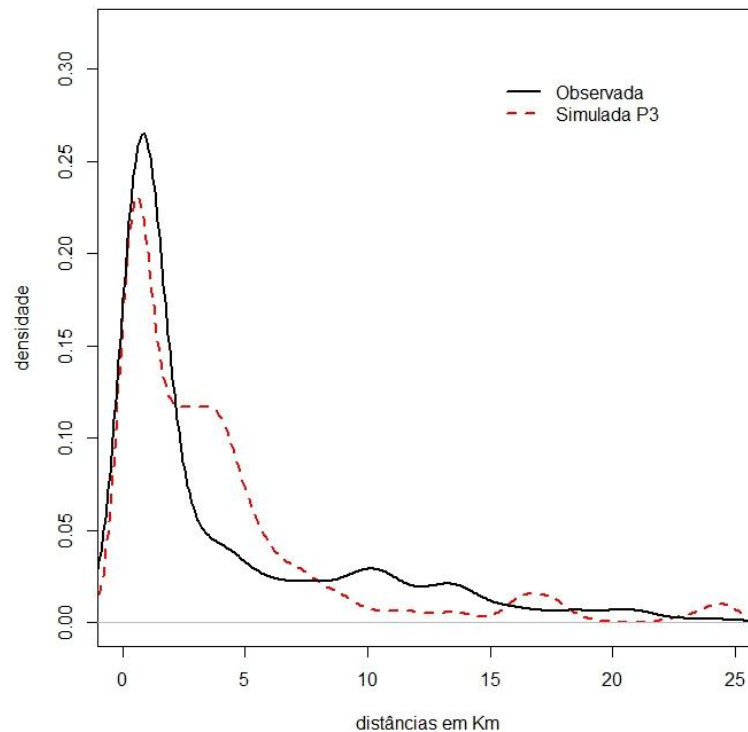
As preferências estabelecidas para esta simulação consideram que pais preferem escolas por nível de proficiência, mesmo que a informação de média não seja totalmente disponível, digamos que os pais conseguem distinguir bem entre uma escola cuja média está no nível avançado, outra que está no intermediário e assim por diante. Dentre todas as escolas de nível avançado, os pais preferem aquela que está mais próxima, esse tipo de preferências pode ser traduzido da forma abaixo, já exposta em (51):

$$\mathcal{P}^3(i) = \min\{d_{is} | \bar{y}_s \geq 250\} > \min\{d_{is} | 250 > \bar{y}_s \geq 200\} > \min\{d_{is} | 200 > \bar{y}_s \geq 150\} > \min\{d_{is} | \bar{y}_s < 150\} \\ \forall i \in I, \forall s \in S \quad (51)$$

As preferências $\mathcal{P}^3(i)$ acima dizem que, para cada aluno i , qualquer escola s com proficiência média ($\bar{y}_s$) maior ou igual a 250 (correspondente ao nível avançado) é preferida às escolas de nível intermediário ($250 > \bar{y}_s \geq 200$). Dentre as escolas de nível avançado, os pais preferem aquelas que possuem a menor distância ($\min\{d_{is} | \bar{y}_s \geq 250\}$).

Por esse modelo temos que os resultados para as distâncias finais são similares àquelas obtidas na simulação considerando apenas as distâncias nas preferências dos pais e alunos (ver GRAF. 22). Entretanto, há uma consideração sobre a proficiência em conjunto com as distâncias o que faz os alunos fiquem nas melhores escolas que estão próximas. Os resultados médios dados pelo modelo 8 sem vagas ociosas estão em μ^3 na TAB. 46 abaixo.

Gráfico 22 - Densidade de distâncias observadas e simulação de preferências lexicográficas



Fonte: Cálculos com uso do banco de dados processado.

Tabela 46 - Diferença de proficiência entre amostra e simulação preferências lexicográficas

	<i>Média Proficiência</i>	<i>Desvio padrão</i>	<i>IC_inferior</i>	<i>IC_superior</i>
Simulação μ^3	219,32	45,50	216,39	222,25
Simulação μ^2	218,69	45,57	215,76	221,62
Simulação μ^1	219,99	45,58	217,06	222,93
Amostra (μ_{Obs})	213,19	45,62	210,25	216,12
Simulação Baseline (μ^0)	204,25	45,79	201,28	207,17

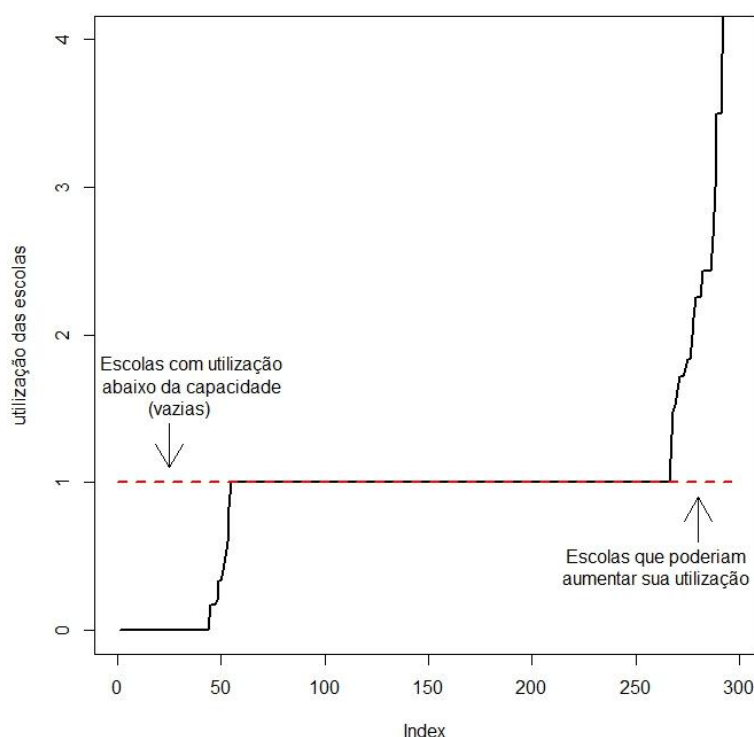
Fonte: Cálculos com o uso do banco de dados processado (2010).

Obs.: A simulação da proficiência foi realizada com o modelo 8 sem vagas ociosas.

Na simulação com vagas ociosas para preferências lexicográficas, observa-se que o número de escolas que ficam completamente vazias é menor do que na simulação anterior, 44 escolas ficam sem alunos, outras 10 não preenchem completamente suas vagas. Um total de 212 escolas continuam com capacidade plena, e todas as 30 escolas que ganharam mais vagas possuem utilização maior do que a prevista pelo número inicial de vagas. O GRAF. 23 mostra o indicador para as simulações lexicográficas.

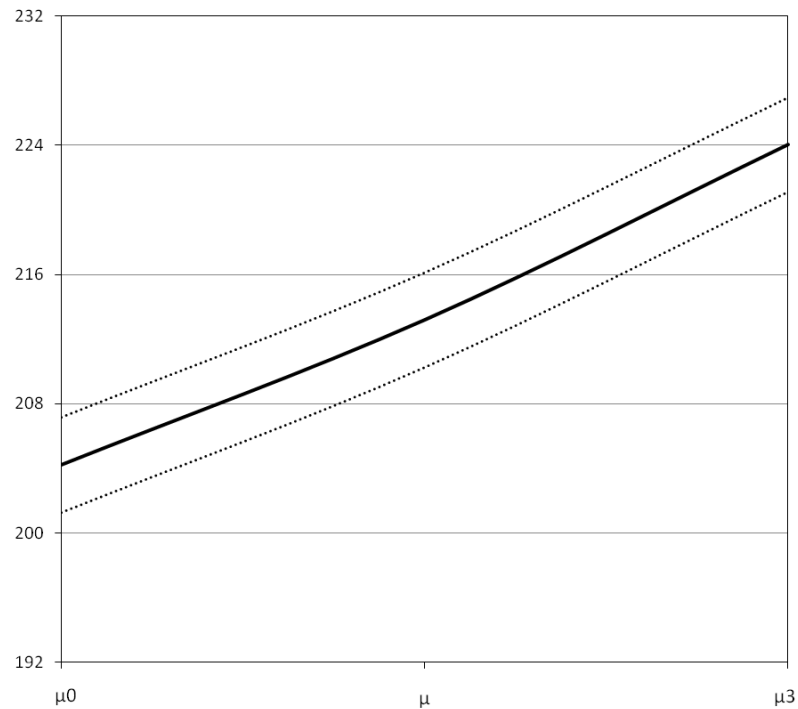
De acordo com o modelo 8 sem vagas ociosas e preferências lexicográficas para as preferências, é possível ganhar, em média, 10,82 pontos de proficiência com a mudança de alocação. Esse modelo é melhor do que o simulado com preferências pela proficiência e tão bom quanto o modelo com preferências pelas distâncias (o GRAF. 24 mostra esses resultados).

Gráfico 23 - Índice de utilização de escolas, simulação lexicográfica e com vagas ociosas



Fonte: Cálculos com uso do banco de dados processado.

Gráfico 24 - Relação da proficiência e as alocações simuladas (μ_3) com vagas excedentes



Fonte: Cálculos com uso do banco de dados processado.

7.5. Simulação considerando preferências dadas por utilidades

Outra maneira de se organizar as preferências pode ocorrer por meio de funções de utilidade que representam como pais/alunos relacionam uma série de características observadas para as escolas. Se além das distâncias e proficiência média, eles considerassem informações tais como rede de ensino, municipal ou estadual, turno (manhã, tarde ou integral), falta dos professores ou se a escola tem educação física.

Porém, para se tratar as informações acima de forma racional e direta é preciso se supor que as preferências dos pais são dadas pela mesma função. Para se considerar todos esses fatores pode se usar uma utilidade que leva em conta essas informações ao obter um ordenamento de utilidades. Na equação (52) foi posta uma especificação geral para essa função de utilidade:

$$\mathcal{P}^4(i) = \max U(Z_s, d_{is}) \quad (52)$$

tal que: $U(Z_s, d_{is}) = \gamma_{0s} + \boldsymbol{\gamma} \mathbf{Z}_s - \beta_1 d_i$

Um modelo simples para a utilidade considerou a equação linear do modelo 7 do capítulo 6 para obtenção dos parâmetros. No fim das contas, os pais estão preocupados com o aprendizado, cuja variável *proxy* é a proficiência. Desse modo, utilizou-se a proficiência como a representação cardinal da utilidade. Estimou-se, portanto, um $\hat{y}$ para cada aluno, dado pela equação (39) abaixo:

$$\hat{y}_{i,s} = 119,75 + 1,8 \cdot \text{dependência}_s - 1,6 \cdot \text{turno}_s + 0,8 \cdot \text{atenção}_s - 5,2 \cdot \text{barulho}_s - 15,1 \cdot \text{falta\_prof}_s + 21,5 \cdot \text{ed\_física}_s + 7,1 \cdot \text{NSE}_s - 1,1 \cdot \log(\text{distância}_{i,s}) \quad (59)$$

Nota-se que a unidade para todas as variáveis acima é a escola, denotada por s . Todas as variáveis são do nível 2 do modelo hierárquico, nível das escolas. A exceção é a variável distância que está por aluno (i) e por escola (s). A ponderação da variável distância é pequena, tal como é pequeno seu efeito nos modelos estimados do capítulo 6, no entanto, esta é a única variável que diferencia as utilidades de dois indivíduos i e j em relação a uma mesma escola, e faz a diferença por conta do grande número de alunos.

Após se obter o $\hat{y}_{i,s}$ para cada um dos alunos, as escolas são ordenadas da maior para a menor utilidade. Para o algoritmo o que importa é ordenação das preferências para as escolas. Pode-se dizer que as preferências são dadas pela utilidade U que depende, em primeira instância, da proficiência estimada:

$$U(Z_s, d_{i,s}) = U(\hat{y}_{i,s}) \quad (60)$$

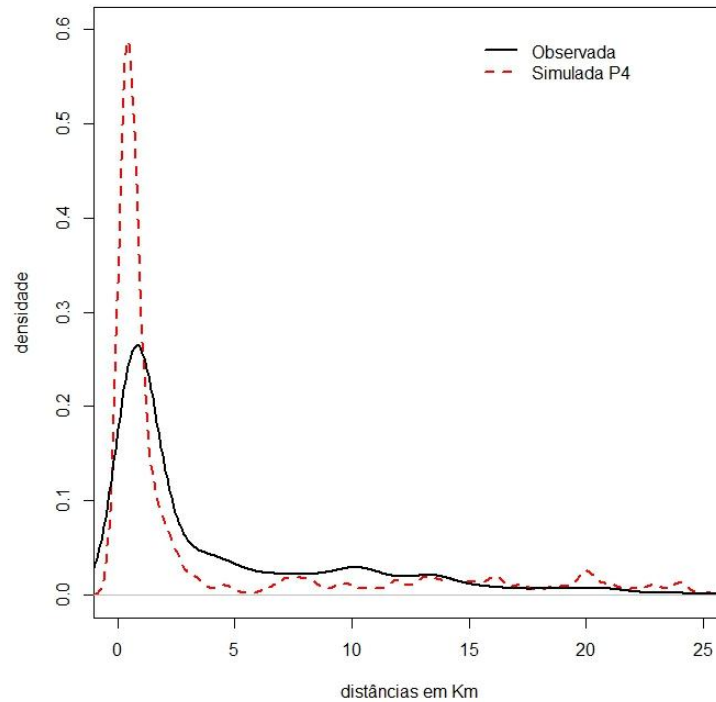
Sendo que as utilidades em (60) podem fornecer um ordenamento de preferências $\mathcal{P}^4$:

$$\mathcal{P}^4(i) = \max U(\hat{y}_{i,s}) \quad (61)$$

A escola mais preferida é a que maximiza utilidade, aqui dependente direta da proficiência. Dadas as preferências de $\mathcal{P}^4(i)$, tem-se a simulação final considerando preferências dadas por utilidades. Esse modelo retorna resultados que estão entre o modelo que considera apenas

distâncias e o modelo que considera preferências lexicográficas, e gera resultados de proficiência semelhantes ao da simulação do modelo lexicográfico (TAB. 47). A densidade de *kernel* para as distâncias aparece no GRAF. 25 abaixo.

Gráfico 25 - Densidade de distâncias observadas e da simulação com utilidades



Fonte: Cálculos com uso do banco de dados processado.

Tabela 47 - Diferença de proficiência entre amostra e simulação preferências com utilidade

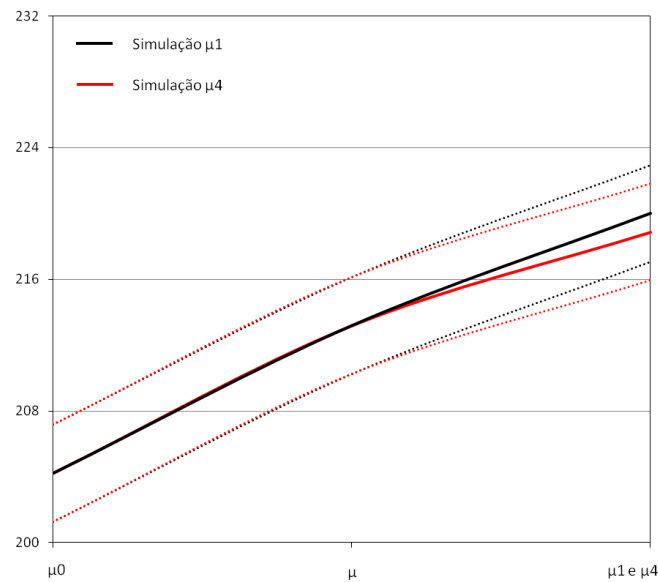
	<i>Média Proficiência</i>	<i>Desvio padrão</i>	<i>IC_inferior</i>	<i>IC_superior</i>
Simulação μ^4	218,87	45,65	215,93	221,81
Simulação μ^3	219,32	45,50	216,39	222,25
Simulação μ^2	218,69	45,57	215,76	221,62
Simulação μ^1	219,99	45,58	217,06	222,93
Amostra (μ_{Obs})	213,19	45,62	210,25	216,12
Simulação Baseline (μ^0)	204,25	45,79	201,28	207,17

Fonte: Cálculos com o uso do banco de dados processado (2010).

Obs.: A simulação da proficiência foi realizada com o modelo 8 sem vagas ociosas.

Na simulação com vagas ociosas, o padrão de utilização é quase o mesmo daquele que foi obtido com as preferências lexicográficas. Quanto aos ganhos de proficiência, eles estão pouco mais de um ponto abaixo da simulação de $\mathcal{P}^1$. O GRAF. 26 abaixo mostra o gráfico da primeira simulação em preto e a nova proficiência da alocação μ^4 de vermelho. A TAB. 48 apresenta as distâncias.

Gráfico 26 - Relação entre proficiências médias e alocações simuladas (μ^1 e μ^4)



Fonte: Cálculos com uso do banco de dados processado.

Tabela 48 - Proporção de alunos por distância em Km Antes e depois da simulação $\mu^1 - \mu^4$

Distância (em km)	Proporção acumulada de alunos ($\leq$)	μ^0	μ^1	μ^2	μ^3	μ^4
	μ_{Obs}					
1	40%	4%	71%	50%	32%	59%
2	58%	11%	77%	61%	43%	71%
5	72%	36%	81%	75%	77%	78%
10	84%	81%	83%	83%	91%	83%
15	95%	99%	90%	91%	94%	89%
25	100%	100%	98%	100%	100%	100%
30	100%	100%	100%	100%	100%	100%
Média das distâncias	4,18	6,60	3,74	4,04	3,98	4,00

Fonte: Banco de dados processado e amostra do mesmo banco (2010).

7.6. Simulações de Políticas Públicas

Nas simulações realizadas até aqui, as preferências das escolas foram postas como se seguissem uma ordem de prioridade imparcial, ou seja, o único critério utilizado para determinar as preferências das escolas foi o das distâncias. Entretanto, cabem também simulações que mimetizam a implementação de critérios mais fortes de seleção por parte das escolas.

Foram executadas duas outras simulações de políticas públicas para educação. Estas simulações fornecem possíveis implicações de processos de seleção diferenciados pelas escolas públicas. Os algoritmos de *matching* permitem sugerir outras alocações possíveis e com implicações mais profundas sobre as políticas educacionais.

Simularam-se então duas situações de políticas educacionais voltadas para as escolas.

1ª Simulação de uma política educacional (escolas polo):

1. Na primeira simulação supõe-se a criação de 9 escolas polo (uma para cada região administrativa da cidade), cada uma com 20 vagas para os alunos de 5º ano. Ao todo serão criadas 180 novas vagas, o que equivale a 8,3% da amostra de 2.173 alunos.
2. O número de vagas das demais escolas continua fixo.
3. As escolas polo preferem os alunos pertencentes à sua regional a qualquer outro aluno de outras regiões. Dentro da região, a escola polo prioriza os alunos segundo critérios de proficiência, da maior para a menor.
4. As preferências das escolas polo por alunos de fora de sua região são dadas aleatoriamente.
5. Todos os alunos de uma mesma regional preferem em primeiro lugar aquela escola polo, independente da distância e não preferem as escolas polo das outras regionais. As demais escolas são preferidas segundo o critério de distâncias.
6. Com as preferências de alunos e escolas dadas pelo procedimento acima, obtêm-se uma nova rodada do algoritmo *deferred-acceptance* com alunos propondo primeiro. O que gera uma nova alocação chamada de μ^5 .

2ª Simulação de uma política educacional (vestibulinho):

1. Nesta segunda simulação não se supõe a criação de novas escolas, mas as três melhores escolas de cada região terão prioridades sobre os alunos dadas pelas proficiências. Pode-se pensar que nesta simulação as escolas que possuem boas notas passam a proceder um sistema de seleção, uma prova para o ingresso dos alunos (um pequeno vestibular, “vestibulinho”).
2. As preferências das três melhores escolas de cada região são dadas pelo resultado de proficiência dos alunos.
3. O número de vagas das escolas, inclusive das três melhores, continua fixo.
4. Todos os alunos de determinada região preferem (antes de qualquer outra escola) as três melhores escolas da regional onde vivem (a ordem arbitrada para o ordenamento dessas três escolas foi a da proficiência média, no entanto, poderia ser uma aleatória dado que se supõe que os alunos sejam indiferentes entre as três primeiras de sua região). As demais escolas, inclusive melhores que não pertencem à mesma região onde vive o aluno, são preferidas segundo o critério de distâncias.
5. Com as preferências de alunos e escolas dadas pelo procedimento descrito, obtêm nova rodada do *deferred-acceptance* com alunos propondo primeiro. Isso gera a alocação μ^6 .

A primeira simulação tem o propósito de compreender como seria a criação de escolas-modelo para cada uma das regionais da cidade, simulação inspirada pelas *magnet schools* estudadas por Garoman (1996). Tais escolas poderiam funcionar próximas de regiões periféricas da cidade e um dos seus principais objetivos seria proporcionar boas condições de aprendizado para alunos com potencial. Em verdade, caso se queira realizar uma política mais meritória para esse tipo de escola, é possível mudar o tipo de preferências, as prioridades dessas escolas-modelo (polo) poderiam ser dadas não somente pela ordenação da proficiência, mas pela inclusão de outros critérios além dos acadêmicos.

A segunda simulação tem a ideia de aproveitar as escolas com bons resultados que já existem em cada regional. Estas escolas conseguem resultados positivos e poderiam, então, selecionar os melhores alunos da regional para aumentar as oportunidades de aprendizado não necessariamente

de quem mora próximo. Esse sistema poderia ser equivalente a aplicação de uma prova de seleção para alunos de quinta série, apelidada de “vestibulinho”. O *vestibulinho* poderia selecionar os alunos de acordo com critérios acadêmicos e também meritórios. Novamente o propósito é permitir que alunos com potencial se desenvolvam plenamente nas capacidades cognitivas.

Logicamente, não se simulou aqui os custos de cada uma dessas alternativas, mesmo assim um indicativo para políticas pode ser obtido: no caso em que os resultados entre as duas alternativas são semelhantes, pode-se optar pela mais barata, de mais rápida implementação e em menor tempo. No caso em que uma delas se mostra muito melhor do que as demais, uma comparação de custo-benefício é necessária.

A TAB. 49 apresenta as informações desses modelos para as políticas públicas. Em particular, observa-se que a simulação de escolas modelo sem vagas a mais de outras escolas (as únicas vagas criadas são das escolas modelo, apresenta os melhores resultados, esses resultados no cenário em que não se mensuram efeitos de migração e escolas é ligeiramente maior do que os obtidos em μ^1 . O modelo do “vestibulinho” não apresentou ganhos maiores e se comportou de maneira semelhante ao modelo μ^2 (de preferências pela proficiência), principalmente quando se considera os efeitos de migração, parece que a alocação de migração proposta por esse modelo não é vantajosa (provavelmente ele tira os alunos de seus melhores lugares possíveis, mesmo em se considerando um critério de prioridades pelas notas). A qualidade (Q_s) se refere ao efeito de qualidade da escola.

Também foram testados isoladamente os efeitos de ganhos para os alunos nos cenários abaixo. Uma simulação que foi realizada é de quanto deveriam ganhar os alunos que migram para as escolas polo para que a proficiência geral se elevasse em um desvio padrão (45,5 pontos). Observou-se que esse ganho deveria ser de 83 pontos por aluno ao ano. Ou seja, esses alunos que migram de escolas precisam mudar de categoria, se tornar aluno do tipo $T_i = 1$. Nos modelos 7 e 8 da TAB. 38. Observou-se que a variável *Tipo_aluno* acrescenta em torno de 70 pontos para alunos que estão no primeiro *decil*. Se cada um dos alunos que conseguiram vagas nas escolas

modelo (180 alunos) conseguirem passar para o primeiro decil de proficiência, o modelo de escolas polo eleva a média de proficiência de Belo Horizonte em um desvio padrão.

Tabela 49 - Resultados das simulações de políticas públicas

<i>Simulação</i>	<i>Cenário</i>	$\hat{y}$	<i>IC_inferior</i>	<i>IC_superior</i>
μ^5	Apenas efeito das distâncias	221,44	218,46	224,41
	Efeito das distâncias e de escolas + efeito migração	213,41	210,42	216,39
	Apenas efeito das distâncias e de interação migração e qualidade	221,34	218,39	224,29
	Efeito das distâncias e de escolas + efeito migração + efeito interação migração e qualidade	219,54	216,59	222,49
μ^6	Apenas efeito das distâncias	219,49	216,54	222,43
	Efeito das distâncias e de escolas + efeito migração	209,64	206,68	212,61
	Apenas efeito das distâncias e de interação migração e qualidade	219,27	216,36	222,19
	Efeito das distâncias e de escolas + efeito migração + efeito interação migração e qualidade	215,74	212,83	218,66

Fonte: Cálculos com o uso do banco de dados processado (2010).

Sob o efeito de interação entre alunos e escolas é possível que existam ganhos maiores não captados pelas simulações aqui realizadas. Pode-se pensar no caso em que a troca de uma das escolas possa alterar de forma fundamental como o aluno se relaciona com o aprendizado, se por exemplo, no efeito interação aluno e escola um aluno passa do tipo $T_i = 0$ para o $T_i = 1$. Ou seja, o aluno mudou de tipo porque mudou de escola. Caso se possa mensurar esse tipo de efeito para as políticas públicas, é possível se obter um efeito maior das simulações de políticas.

Naturalmente esse efeito seria gerado apenas para os alunos que conseguiram vagas e essa política teria de tentar uma maneira de espalhar esses efeitos para os demais alunos que não participam da escola modelo. Uma forma de se conseguir isso é propiciando uma maior

accountability por parte das escolas na provisão e uso de suas vagas bem como para o aprendizado do aluno.

Estímulo a uma concorrência saudável pode ser uma ferramenta para esse processo e as simulações realizadas tentaram dar indicativos para essa direção. Ao serem aplicados algoritmos em que os pais reportam algumas preferências, o sistema permite uma maior liberdade de escolha e pais podem demandar as escolas que julgam mais apropriadas.

7.7. Simulações sob a perspectiva dos ganhos para os estudantes.

Alocações diferentes podem impactar de forma distinta os estudantes. Simulações possuem a vantagem de rapidamente apresentar os resultados contrafactuais comparativos. É possível comparar imediatamente os alunos que mudaram de escola na alocação μ^1 , e comparar com a situação anterior μ_{Obs} (que seria a situação caso eles não tivessem mudado e continuassem na mesma escola).

O GRAF. 27 apresenta essa informação na distribuição acumulada de proficiência antes e depois da simulação. Observa-se que os percentis ganham pontos de forma diferente ao longo da distribuição. Os 20% de alunos de notas mais baixa ganham uma nota igual ou menor com a mudança de escola. Essa diferença se torna maior para os percentis de notas mais altas. Os alunos do percentil de 60% ganham uma diferença em torno de 16 pontos de proficiência. Os alunos de notas 20% mais altas (percentil 80%) ganham em torno de 22 pontos na prova pela mudança de escola pelo novo arranjo.

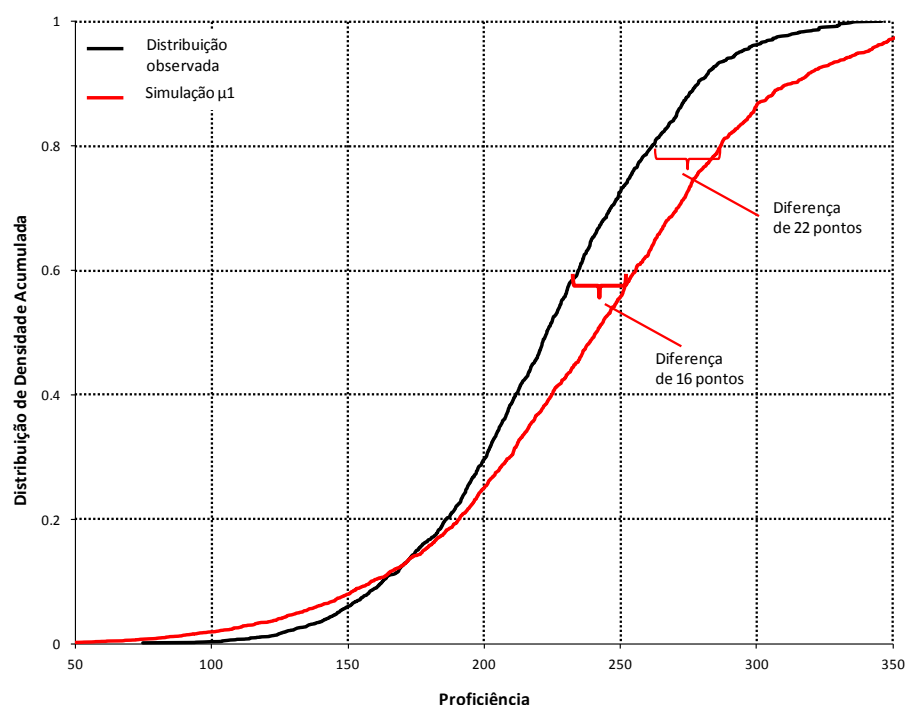
Acima de 80% os ganhos podem se igualar a um desvio padrão de proficiência (em torno de 45 pontos). No entanto, uma explicação para a diferença entre essas duas curvas de distribuição acumulada é também a maior dispersão após a obtenção da proficiência simulada. Nos dados originais o desvio padrão da proficiência é de 45,77 pontos, após a simulação é de 47,42 pontos.

Os efeitos das diferentes alocações e da migração dos estudantes tem de ser melhor conhecidos para se derivar melhor os ganhos sociais com uma nova alocação. Como último exercício a TAB.

50 apresenta uma decomposição de Oaxaca-Blinder (Oaxaca, 1973 e Blinder, 1973) para a diferenciação de alunos migrantes e não migrantes. A equação utilizada para se captar essa diferença é a equação (62), que é similar a Eq. 59, porém sem os valores dos parâmetros:

$$\hat{y}_{i,s} = \gamma_0 + \gamma_1 \cdot \text{dependência}_s - \gamma_2 \cdot \text{turno}_s + \gamma_3 \cdot \text{atenção}_s - \gamma_4 \cdot \text{barulho}_s - \gamma_5 \cdot \text{falta\_prof}_s + \gamma_6 \cdot \text{ed\_física}_s + \gamma_7 \cdot \text{NSE}_s - \gamma_8 \cdot \log(\text{distância}_{i,s}) \quad (62)$$

Gráfico 27 - Distribuições acumuladas da proficiência observada e após simulação ($\mu 1$)



Fonte: Cálculos com uso do banco de dados processado.

A regressão em (62) foi realizada para alunos migrantes e alunos não migrantes. E os procedimentos para a decomposição seguem os passos descritos em 6.3.1 com as equações 41 à 44. Observa-se que a diferença de médias de migrantes e não migrantes é de apenas 2,66 pontos, e por conta disso os valores percentuais são altos para decomposição desse efeito.

Tabela 50 - Decomposição de Oaxaca para o diferença de proficiência entre migrantes

Diferença inicial devido a migração	Diferença dos interceptos	Diferenças nos γ das variáveis explicativas	Diferença dada pelos erros
ΔY	$\Delta \gamma_0$	$\Delta \gamma_{1-8}$	Δu
2.66 (100%)	-12.76 (-480.5%)	24.36 (917.1%)	-8.94 (-336.6%)

Fonte: Cálculos com uso do banco de dados processado.

A diferença dos interceptos é negativa para os alunos migrantes, porém esse efeito é compensado pelas diferenças nas variáveis explicativas. Isso pode ser mais um indicativo de que migrantes tendem a ir para boas escolas e por isso os efeitos mais que compensa o ponto de partida desfavorável. No entanto, novamente se ressalta que a diferença entre migrantes e não migrantes é pequena e testes da robustez dessa diferença precisam também ser implementados em outros bancos de dados de alunos e escolas.

Outras decomposições podem também ser propostas, estudos posteriores podem dar continuidade aos efeitos entre as diversas categorias de aluno, tais como qualidade do aluno, da escola, e outras características diferenciadas. Bem como podem também ser realizadas decomposições entre os percentis das notas, tal como a proposta de Juhn, Murphy & Pierce (1993) podem ser realizadas bem como outros tipos de comparação de grupos antes e depois da mudança da alocação.

7.8. Conclusões sobre as simulações

Excetuando-se a simulação de *baseline*, que possui diferenças maiores para captar a forma de alocação antes do sistema de cadastro da PBH, as quatro simulações com diferentes preferências dos alunos retornam resultados de proficiência finais muito semelhantes, todos eles dentro do intervalo de confiança de estimação (os intervalos se sobrepõem).

A única simulação que destoou do panorama geral foi a simulação com preferências dadas pelas proficiências médias e com excedente vagas. Parece que nesse caso, desprezar o custo de deslocamento não é uma boa proposta, pois os ganhos de migração e de mudança de escola, em média, não compensam os custos de se deslocar maiores distâncias. Como aquele modelo não

considera custos de deslocamento na formação das preferências dos alunos, tem-se um resultado pior, mesmo na situação com excedente de vagas. Sob essa situação, a alocação observada na amostra é a alocação ótima.

De maneira geral, a criação de mais vagas introduz ganhos de proficiência maiores, isso porque as simulações aumentam o número de vagas nas escolas com as maiores médias de proficiência. Isso indica que algumas escolas poderiam, talvez, criar mais turmas enquanto outras poderiam reduzir, pois na existência de mais vagas nas melhores escolas, estas outras se tornam menos demandadas.

De todos os modelos, o mais simples, de preferências pelas distâncias, foi o que ofereceu melhores resultados. Esse modelo sugere que é possível se reduzir ainda mais as distâncias e com isso ganhar cerca de 9 pontos em proficiência da prova de português. O modelo de preferências lexicográficas também oferece bons resultados, principalmente no cenário de vagas excedentes, em que chega a oferecer ganhos de aproximadamente 11 pontos, o que equivale acelerar em um ano o aprendizado dos alunos.

Em relação às simulações, quatro considerações são importantes:

1. O sistema de alocação de Belo Horizonte já consegue um bom resultado em termos de minimizar as distâncias e, até onde se têm notícia, esse é um mecanismo de alocação bastante estável dado que a população recebeu bem a política e os poucos casos dissidentes são tratados de maneira personalizada. Alterar esse equilíbrio já estabelecido ao longo de 20 anos pode ser mais caro do que mantê-lo funcionando com o custo de alguma perda de ineficiência.
2. O ganho de proficiência em cerca de apenas 10 pontos pode parecer pequeno, representa apenas $\frac{1}{4}$ do desvio padrão da proficiência. Porém, observa-se que esse ganho foi obtido para apenas um ano (2010), e uma série (o 5º ano do fundamental) e para apenas uma prova do PROEB (Língua Portuguesa). Realizou uma simulação na série mais próxima do ano de alfabetização para se ter uma ideia de quais são os ganhos para os alunos que estão

começando sua trajetória escolar. Estudos longitudinais podem revelar melhor qual o real efeito da alocação no longo prazo. Se o efeito se acumular à medida que o aluno progride, pode ser que os ganhos de uma alocação diferente sejam maiores do que aqui encontrados.¹¹⁹

3. Mesmo na comparação dos modelos que agregam mais valor à proficiência média, pode ser que a alocação atual já seja a mais eficiente, pois a forma como a prefeitura de Belo Horizonte trabalha com as distâncias é diferente das distâncias euclidianas aqui utilizadas. A PBH estabelece jurisdições escolares que foram elaboradas considerando regiões da cidade e demais obstáculos geográficos (conforme descrito no capítulo 2). Para essa tese não se conseguiu informações georreferenciadas dessas jurisdições, sendo que se optou por utilizar a distância mais simples entre dois pontos, isso pode inserir alguns erros (um aluno não pode atravessar a lagoa da Pampulha em linha reta, por exemplo) sendo necessários aperfeiçoamentos. De maneira geral, parece que o modelo está adequado, pois as alocações simuladas não destoam demais das distâncias que são obtidas da alocação real, mas isso pode dizer que talvez não convenha reduzir muito mais as distâncias em troca de menor segurança para os alunos (fator que não foi considerado neste trabalho).¹²⁰
4. Pode-se dizer que ao ressaltar efeitos que se anulam entre os alunos, as simulações foram conservadoras. Uma limitação das simulações realizadas é que o efeito de interação de alunos e escolas não pôde ser simulado em maior nível de detalhamento. Em teoria, podem existir composições de turmas que devem ser ideais para se alcançar uma maior interação e proficiência dos alunos. Nessas turmas, os ganhos dos alunos que trocam de escolas podem ser maiores do que simplesmente os ganhos estimados com as distâncias, a migração e outros parâmetros que foram utilizados nas simulações. Em parte a dificuldade está na característica não observada desse tipo de informação. A interação de alunos nas escolas precisa ser mais bem compreendida para que se possa empregá-la em um efeito de simulação. Os efeitos dos estudantes uns sobre os outros (*peer effects*) podem inclusive acrescentar novas informações relevantes para as simulações de políticas públicas, mas é

¹¹⁹ A hipótese inversa pode ser a de que o efeito dessa alocação se dissipe com o tempo e prevaleça mais um efeito fixo do aluno devido suas características inatas e familiares, desse modo, a alocação terá seu propósito apenas em alguns momentos específicos, possivelmente no início da trajetória escolar.

¹²⁰ Em alguns casos as barreiras físicas são avenidas movimentadas que oferecem risco na travessia.

difícil dizer de antemão qual composição de turma irá funcionar melhor para ativar esse efeito.

Uma simulação que se realizou e não foi mostrada nos modelos acima, leva em conta a criação de vagas em escolas bem localizadas que possuem o maior número de alunos que moram próximos. Nesse modelo com mais vagas para essas escolas a distância da amostra pode ser reduzida para a média de 1,30 km, uma redução significativa da média em mais de um quilômetro e meio. Curiosamente esse modelo não ofereceu resultados muito melhores do que os já enunciados nos modelos apresentados. Seus resultados ficaram parecidos, mas um pouco abaixo da simulação 1. Uma possibilidade para isso ocorrer é que apesar dos ganhos da redução significativa das distâncias, as escolas para as quais se criou mais vagas podem não ser boas escolas, sendo que está se colocando mais alunos nas escolas piores. Desse modo não compensaria muito, é melhor criar mais vagas nas melhores escolas, o que possui uma interpretação econômica.

Nas simulações de políticas públicas se observou ganhos para a introdução de escolas polo. No entanto, o ganho de proficiência obtido nessa simulação não foram muito maiores do que aqueles obtidos das alocações μ^1 e μ^3 , por exemplo, que simulavam preferências dadas pelas distâncias e pelas preferências lexicográficas. A migração quando não ocorre para uma escola de qualidade não parece levar a ganhos muito maiores. No caso alunos que se deslocam para escolas polo ganham um efeito por essa migração e qualidade, mas os demais alunos não. No entanto, caso esse tipo de escola consiga acrescentar um efeito de interação de 83 pontos a mais por aluno, a média geral de proficiência de BH se elevaria em um desvio. Para que esses ganhos se difundissem para todos os alunos e não ficasse concentrado em tais escolas, outras iniciativas teriam de ser tomadas para promover ganhos não somente dentro das escolas polo.

O modelo do vestibular para alunos de 5º ano, provavelmente é mais barato do que o de criação de escolas polo, mas em compensação, ele não gerou os mesmos efeitos. Teve apenas ganhos pela redução de distâncias com o uso do algoritmo.

Uma das conclusões que se pode esboçar é que o sistema público de ensino tal como funciona hoje no país possui pouca competição. Refere-se aqui a uma competição salutar por resultados.

Em um sistema competitivo, o uso mais eficiente dos recursos é almejado. O modelo com vagas ociosas, ou excedente de vagas, oferece algumas interpretações: os algoritmos *deferred-acceptance*, ao lidarem com as preferências de pais e prioridades de escolas, estão lidando com um sistema em que pais sinalizam preferências por escolas que eles julgam melhor, e desse modo, criam uma demanda, mesmo que artificial, por determinadas escolas. Ao se permitir mais vagas em escolas mais disputadas, a alocação pode ser mais eficiente. Em contrapartida, algumas escolas ficam vazias neste processo, denotando que precisam melhorar para se configurar em patamares mais altos nas preferências dos pais.

Outro ponto que pode ser abstraído à guisa de conclusão, é que o sistema de cadastro escolar da PBH parece considerar um algoritmo em que as preferências de alunos e escolas são as mesmas e dadas pelas distâncias. Nesse tipo de modelo tanto faz quem propõe primeiro, não importa se alunos ou escolas o resultado de alocação é único, como visto em **P.1** no capítulo 3. No entanto, caso se considere implementar sistemas em que pais elencam escolas mais desejadas é preciso propor um sistema estável preferido pelos estudantes e não pelas escolas. Essa é uma perspectiva mais presente na literatura de *matching*, favorecer o algoritmo pelo lado dos agentes hipossuficientes (no caso os alunos).

Um sistema com alunos propondo é a prova de manipulação de estratégias e acrescentaria a vantagem de permitir maior liberdade de escolha. No sistema atual, assume-se que as escolas são homogêneas e as preferências são quase que impostas aos participantes. Os algoritmos *deferred-acceptance* apesar de requererem ainda certa centralização, são uma forma mais democrática de estabelecer uma manifestação de preferências de pais e alunos e criar uma competição salutar por vagas nas escolas públicas.

8. CONCLUSÕES

Essa tese teve como fio condutor a questão sobre qual seria o efeito de diferentes alocações de alunos e escolas sobre a proficiência média dos estudantes. Para se estudar essa relação, tomou-se por base, e como ponto de partida, o sistema de cadastro municipal de matrículas de Belo Horizonte, implantado pela primeira vez em 1993.

Procurou-se desvendar qual seria a melhor alocação possível entre alunos e escolas e com isso tentar saber quão bem tem se desempenhado o sistema de cadastro municipal de matrículas frente às possibilidades de diferentes alocações. Para se estudar este último ponto, introduziu-se no trabalho três vertentes teóricas que visam auxiliar na interpretação das alocações possíveis.

Duas dessas vertentes estão ligadas aos estudos de pareamento ótimo (*matching*), a primeira delas ligada à corrente teórica desenvolvida por Becker (1991), que trata dos *assortative matings*. Realizou-se, nesta tese, a transposição desta teoria para o caso de demografia da educação, abordando os “casamentos” entre alunos e escolas, que podem ser pensados como os casamentos polígamos (uma escola e vários alunos).

A segunda vertente introduzida auxilia na interpretação do sistema de cadastro de matrículas. Trata-se da literatura dos algoritmos de pareamento, pioneiramente desenvolvida por Gale & Shapley de 1962. Uma das principais preocupações teóricas daquele artigo era como alocar eficazmente alunos e escolas de acordo com as preferências de ambos. Tal literatura se desenvolveu enormemente após o seminal artigo, e o algoritmo passou a ser chamado de algoritmo de Gale-Shapley ou de *deferred-acceptance many-to-one*, para o caso de escolas e alunos. Roth & Sotomayor (1990) revisam os principais desenvolvimentos das primeiras três décadas desta área. Essa literatura foi essencialmente utilizada nas simulações de alocações.

Realizou-se também uma revisão das contribuições da literatura educacional que relaciona insumos ao produto escolar (essa é a terceira vertente teórica mencionada), enfatizando o aprendizado como principal produto derivado da relação de alunos e proficiência. Especial

destaque foi dado para a literatura que enfatiza o efeito da escolha das escolas, chamada de literatura do *school choice*, e do papel do território no desempenho escolar.

A teoria desta tese se baseia na área de intersecção entre essas três correntes teóricas. A proposta aqui realizada foi empregar algoritmos de pareamento para propor outras alocações possíveis para alunos e escolas de Belo Horizonte. Para isso construiu-se uma extensa base com dados de moradia dos estudantes, localização das escolas e notas dos alunos na proficiência de língua portuguesa da prova do PROEB. Essa base de dados foi toda georreferenciada de modo que se obteve as distâncias de cada aluno para cada escola do banco. Foram 16.354 alunos amostrados de um universo de aproximadamente 27 mil alunos do 5º ano do fundamental, matriculados no período letivo de 2010. A amostra de escolas contou com 296 escolas municipais e estaduais de um universo de 382 escolas da rede pública dessas duas esferas administrativas.

O sistema de cadastro de Belo Horizonte funciona sob a pressuposição de escolas homogêneas em qualidade e na hipótese que as preferências que consideram apenas as distâncias são as mais adequadas para alocação de estudantes a escolas. Essa pressuposição foi também constatada por Fernandes (2007) ao estudar o sistema de cadastramento e de matrículas para a cidade de São Paulo: os sistemas de alocação de estudantes pelas distâncias, em geral, valem-se dessa suposição. Quebrando-se o pressuposto de escolas homogêneas (que parece pouco plausível), uma questão que aqui se colocou foi a de estudar se não seria uma alocação apenas pelas distâncias uma forma permanente exclusão de alunos que moram em regiões mais pobres da periferia.

Encontrou-se, por meio de simulações, que a situação onde se possui um sistema de cadastro de matrículas é provavelmente muito melhor do que a situação em que pais escolhiam as escolas sob um mecanismo descentralizado. No sistema descentralizado, conseguir uma boa escola era tal como participar de uma loteria. Uma das melhores contribuições para a evolução do aprendizado no que diz respeito à alocação de estudantes é possuir um sistema de alocação determinado. Outra resposta que se obteve para a questão é de que existe a possibilidade de que o sistema atual esteja perpetuando desigualdades para determinados tipos de alunos entre os que estão mais atrasados, no entanto, a solução de propor que esses alunos estudem em outras escolas é delicada,

pois há possibilidades de que o efeito líquido dessa troca seja negativo. A não ser que se consiga reduzir as distâncias e aumentar os efeitos de escola e do aluno.

Infelizmente, Belo Horizonte não dispunha de informações sobre a proficiência antes da implementação do cadastro, uma das primeiras avaliações sistemáticas do ensino básico no Brasil, o SAEB, só foi introduzido em 1995, dois anos depois do sistema de cadastro de Belo Horizonte já estar implantado. Sem contar que o SAEB é uma base amostral e, sendo assim, seria difícil construir um banco que ligasse alunos e escolas aos respectivos endereços. Dessa forma, contar com uma avaliação censitária tal como a do PROEB foi fundamental.

Por meio do banco de dados processado simulou-se uma situação que provavelmente se aproxima da situação anterior ao sistema. Usando o algoritmo de *Top Trading Cycles* (TTC) simulou-se uma situação em que pais/alunos se dirigiriam às escolas mais preferidas e as escolas escolhiam os alunos por um critério de prioridades aleatório (que simularia uma ordem aleatória de chegada nas filas).

Por meio dessas simulações encontrou-se que na situação mais conservadora o sistema não deve ter prejudicado o desempenho dos alunos e no caso mais otimista, este efeito pode ter sido de mais de 10 pontos na proficiência, um grande ganho em se considerando uma mudança apenas de alocação. Os valores médios dessa estimativa sugerem algo em torno de 9 pontos de ganho para o 5º ano. Com dados do PROEB de 2006 a 2009, sabe-se que, em média, cada série ganha algo em torno de 10 pontos de proficiência entre o 5º ano e o 9º ano do fundamental. Com isso, tem-se que o sistema de cadastro de Belo Horizonte pode ter proporcionado um ano de ganho no aprendizado escolar.

Um desenvolvimento importante que pode ser feito deste trabalho é o de estender essas mesmas simulações para mais séries além do 5º ano do fundamental. Se o ganho com de proficiência com a alocação se perpetuar para as demais séries, tem-se que a alocação pode impactar tremendamente a carreira escolar de um estudante. Dessa forma, ter as aplicações aqui realizadas para outros anos de ensino pode trazer um entendimento mais detalhado sobre o tema de alocação de alunos e escolas.

Como alocações diferentes impactam principalmente na distância que o aluno deve percorrer para chegar à escola, foi feita uma análise mais extensiva das relações entre distância de casa até a escola e a proficiência e com relação à migração, definida como o estudante que muda de bairro ou região para estudar. O que se encontrou é uma leve evidência de que a distância prejudica o desempenho, o aluno que está mais distante da escola tende a ter um desempenho pior do que um colega com características semelhantes que estuda na mesma escola, mas mora mais perto.

O efeito de migração aponta o contrário, o aluno que migra tende a apresentar um resultado um pouco melhor do que aqueles que não migraram, mas esse efeito foi pouco robusto e pode estar relacionada a uma heterogeneidade entre os migrantes: há os que migram como uma opção, conseguiram alguma maneira de mudar a alocação obtida ou já seriam privilegiados no sistema, e há os que migram por falta de opção, acabaram-se as vagas em escolas próximas.

Dessa maneira temos que pais podem ter estratégias diferentes para alocar seus filhos nas escolas públicas. Alguns pais podem considerar o local de moradia pensando na influencia que isso pode ter na escolha feita pelo sistema para o seu filho, em outros casos, como reportado no capítulo 2, pais podem falsificar ou solicitar “favores” para adulterar o seu endereço. Fatos como reportar o endereço de parentes ou conhecidos ou o endereço do trabalho e não o da residência. Isso aponta que alguns pais provavelmente consideram preferências diferentes daquelas que levam em conta apenas as distâncias. Isso introduz uma heterogeneidade nos dados que pode estar relacionada aos tipos de alunos migrantes relatados.

Ressalta-se que o sistema de cadastro de Belo Horizonte não é obrigatório, ou seja, ele não obriga nem pais nem escolas a matricularem os alunos nas escolas recomendadas. Porém, os casos que fogem a essa regra não passam de 10% do universo de alunos para quem o sistema recomenda uma escola (conforme relatório do cadastro escolar, SME/PBH 2013). De maneira que, apesar de se encontrar alguns casos pontuais de insatisfação e de tentativa de burla, o sistema já implementado pela PBH é bastante estável.

Com isso, um resultado que se encontrou é que o sistema de cadastro de Belo Horizonte já minimiza muito bem as distâncias. Se o critério das distâncias for o socialmente prioritário, o

algoritmo de Belo Horizonte já faz isso bem e não haveria motivos para mudar. Porém um ponto que esta tese tenta levantar é que outros critérios de escolha sociais são também possíveis. Para ver isso, foram propostas simulações em que outros sistemas de preferências de pais e alunos são propostos, com isso, viu-se que em alguns casos a proficiência pode ter ganhos tão grandes quanto aqueles da primeira implementação do sistema de cadastro.

Um ponto muito importante que se colocou e que impactou mais significativamente nas simulações realizadas foi a presença de vagas ociosas nas escolas do decil superior de proficiência. Se existem tais vagas e o efeito de migração e aprendizado na escola se sobrepõe ao efeito negativo de se percorrer distâncias um pouco maiores, a proficiência pode aumentar em quase 30 pontos no melhor dos casos. Isso pode ocorrer por conta de uma maior competição entre as escolas. Mesmo se tratando de escolas públicas, talvez seja salutar inserir uma competição de vagas entre as escolas.

Um sistema assim proposto teria os alunos escolhendo uma lista de escolas que julgam adequada, o critério de desempate da escola pode continuar sendo a proximidade (maior prioridade para alunos mais próximos) e, nesse sistema, escolas de desempenho muito fraco podem ficar vazias.

Em um sistema dinâmico com o algoritmo acima proposto, pode-se pensar que os gestores educacionais teriam mais uma informação para comparar o desempenho de escolas. Uma escola que recebe poucos alunos teria de se apurar para que seja lembrada na lista de preferência dos pais ou então correria o risco de estar sempre vazia e no limite poderia até fechar.

Analisando-se demograficamente essa questão, a conclusão acima ganha um relevo no contexto de um número decrescente de alunos demandantes. Observou-se que, desde a sua implantação, o número de alunos demandantes no sistema de cadastro passou por uma queda do número usual de alunos, esse número hoje é metade do que se obtinha nos anos 90. Isso se insere no contexto de uma mudança demográfica e econômica pela qual o Brasil passa, e de forma acentuada.

Se por um lado o fato parece positivo, pois se pode ter um excesso de vagas e mais facilidade no manejo, por outro, ao se decidir quais vagas permanecem e quais devem ser extintas é preciso ter

critério, pois é preciso evitar que a distância média se eleve. Isso sem falar que ainda há situações de expansão da demanda que continuam importantes, tais como o surgimento e crescimento de novos bairros, e expansões cíclicas do número de alunos, bem como, a necessidade de introduzir um sistema para o ensino médio que levem em conta as características dessa etapa de ensino. Para todos esses casos um sistema que monitora melhor o preenchimento das vagas nas escolas pode ser oportuno.

Recomenda-se, por outro lado, cuidado em cada nova proposta de alocação que se possa realizar, pois trata-se de um exercício teórico de simulação que visa fornecer maior entendimento do tema e do funcionamento da relação aluno e escola, bem como da relação aluno e território e escola e território. Transplantar-se rapidamente da teoria para sugestões práticas envolve um compromisso ético que precisa ser observado. No caso de se compreender melhor essas alocações, o ideal seria ter observações de linha de base para cidades que ainda não implementaram um sistema de cadastramento.

Muitas capitais e grandes cidades brasileiras funcionam ainda com um sistema de cadastramento descentralizado, tal como era o sistema de Belo Horizonte antes de 1993. Um sistema descentralizado em grandes cidades traz inúmeros problemas já reportados na introdução e no capítulo 2 desta tese. O ideal para toda mudança de política pública que se possa fazer é que o acompanhamento de resultados se faça junto com a mudança de atuação da política. Dessa maneira, pode se pensar em uma cidade que ainda não tenha um sistema de cadastro e que possa implementar um sistema de alocação experimentando seus impactos na proficiência, tal que alunos possam ter maior liberdade de escolha para as escolas.

Tal sistema não pode deixar de levar em conta que os pais/alunos possuem limitações em suas preferências. Ao contrário das simulações aqui realizadas, nenhum pai ou aluno pode listar 296 escolas preferidas, da primeira à última. Isto é, se não tecnicamente impossível, humanamente inviável. O mais próximo que se pode conseguir são pais e alunos que declaram uma lista truncada. Porém, preferências truncadas inserem alguma complexidade maior na solução à prova de estratégias (*gaming*) por parte dos agentes envolvidos, sendo que as aplicações de algoritmos *deferred-acceptance*, ou *Top Trading Cycles*, requerem o máximo de cuidado e estudos para

serem aplicados na prática. Todos os casos de sucesso de implementação de novos sistemas, inclusive o de BH, envolveram uma longa preparação no desenho de mecanismos eficientes.

Por fim, propôs-se com esta tese acrescentar-se o tema de casamentos, bem entendidos como pareamento de alunos e escolas, à área de estudo de demografia da educação. Buscou-se um entendimento teórico e experimental, por meio de simulações, de como alocações diferentes de alunos e escolas podem interferir na obtenção da proficiência média de um sistema de ensino. Foram encontradas evidências de que essa relação deve existir, muito embora, fatores tradicionalmente reportados na literatura educacional tais como efeitos das variáveis familiares e socioeconômicas do aluno, e variáveis escolares continuam sendo importantes.

O presente trabalho pode ter a chave para a compreensão de como os efeitos de alunos e escolas se complementam, e também para entender como estes dois grandes níveis de análise interagem com um terceiro nível que é o espaço. Propor uma primeira análise da interseção entre estes temas foi um dos objetivos da tese, espera-se, com elevada esperança, que este trabalho tenha ajudado a responder as questões a que se propôs, e possa ter acrescentado um ponto importante para a compreensão da função de produção educacional, relações de demografia da educação e a alocação de estudantes, e a influência das alocações na proficiência.

9. BIBLIOGRAFIA

Abdulkadiroğlu, A., Pathak, P. A., & Roth, A. E. (2009). Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *The American Economic Review*, XCIX (5), pp. 1954-1978.

Abdulkadiroğlu, A., Pathak, P. A., & Roth, A. E. (2005). The New York City High School Match. *The American Economic Review*, XCV (2), pp. 364-367.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., & Sönmez, T. (2005). The Boston Public School Match. *The American Economic Review*, XCV (2), pp. 368-371.

Abdulkadiroğlu, A. & Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *The American Economic Review*, XCIII (3), pp. 729-747.

Abdulkadiroğlu, A., & Sönmez, T. (1999). House allocation with existing tenants. *Journal of Economic Theory*, LXXXII (2), 233-260.

Albernaz, A.; Ferreira, F. H. G.; Franco, C. (2002) Qualidade e Equidade na Educação Fundamental Brasileira. *Pesquisa e Planejamento Econômico*, XXXII (2), pp. 453-476.

Alves, M. T., & Soares, J. F. (2007). As pesquisas sobre o efeito das escolas: contribuições metodológicas para a Sociologia da Educação. *Sociedade e Estado*, XXII (2), pp. 435-473.

Arrow, K. J., & Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, XXII (2), pp. 265-290.

Bardella, F. P. (2005). *Alocação de estudantes aos centros de pós-graduação em economia no Brasil: um experimento natural em organização do mercado*. Dissertação de Mestrado em Teoria Econômica do departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (USP).

Barros, R. P. (2010). Qualidade do Professor. Nota técnica do projeto Todos pela Educação para melhorar o aprendizado. Disponível em: <http://www.paramelhoraroaprendizado.org.br/Conteudo/verbete.aspx?canal=20100701145550501160&subtema=20100615161126445512&verbete=20110419145632104555> (ultimo acesso: 15/05/2013).

Becker, G. S. (1991). *A Treatise on the Family* (Enlarged Edition ed.). Cambridge, MA, USA: Harvard University Press.

Becker, G. S. (1975). *Front matter, Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. In *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 2nd ed. (pp. 22-0). NBER.

Becker, G. S. (1974). A Theory of Marriage Part II. *Journal of Political Economy*, LXXXII (2), pp. S11-S22.

Becker, G. S. (1973). A Theory of Marriage Part I. *Journal of Political Economy*, LXXXI (4), pp. 813-846.

Becker, G. S., & Murphy, K. M. (2000). *Social Economics: market behavior in a social environment*. Cambridge, MA, USA: Harvard University Press.

Bergstrom, T. C. (1997). *A survey of theories of the family*. In: Rosenzweig, M. R., & Stark, O. (Orgs.), *Handbook of Population and Family Economics*, Elsevier, 1997, Volume 1, Part A, pp. 21-79.

Biondi, R. L. & Felício, F. (2007) *Atributos escolares e o desempenho dos estudantes: uma análise em painel dos dados do SAEB*. Texto para discussão, 28. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Brasília: 2007.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, VIII (4), pp. 436-455.

BRASIL. Decreto nº 6.583, de 29 de setembro de 2008. Poder Executivo, Brasília, DF.

Disponível em:

http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6583.htm

Brooke, N. & Soares, J. F. (2008). *Pesquisa em Eficácia Escolar: origens e trajetórias*. Belo Horizonte: Editora UFMG, 2008.

Browning, M., Chiappori, P. A., Weiss, Y. (2011). *Family Economics*. Universidade de Tel Aviv University, manuscrito não publicado de livro texto. Disponível em: http://www.cemmap.ac.uk/resources/chiappori/paper_1.pdf (último acesso 15/05/2013).

Browning, M., & Chiappori, P. A. (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, LXVI (6), pp. 1241-1278.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Ed. Sage, Newbury Park: 1992.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, CI (1), pp. 147-158.

César, C. C., Soares, J. F. (2001). Desigualdades acadêmicas induzidas pelo contexto escolar. *Revista Brasileira de Estudos de População*, XVIII (1/2), pp. 97-110.

Chiappori, P. A., Fortin, B., & Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of political Economy*, CX (1), pp. 37-72.

Christensen, F. (2008). *Learning by Dating, Commitment, and Assortative Mating*. Apresentado no North American Winter Meetings of the Econometric Society. Paper não publicado, disponível em Towson University:

<http://pages.towson.edu/FChriste/LearnCommitSort101708.pdf> (último acesso: 15/05/2013).

Coelho, M. I. M. (1998). Vinte anos de avaliação da Educação Básica no Brasil - aprendizagem e desafios. *Revista Ensaio: Avaliação e Políticas Públicas em Educação* XVI (59), pp. 229-258.

Coleman, J. S., Campbell, E., Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., York, R. L. (1966). *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966.

Cullen, J. B. Jacob, B. A., & Levitt, S. (2006). The Effect of School Choice on Choice on Participants: Evidence from Randomized Lotteries. *Econometrica*, Vol. LXXIV (5), pp. 1191–1230.

Davis, Jr. C. A. (1997). GIS: dos conceitos básicos ao estado da arte. *Espaço BH: trabalho, tecnologia e informação na administração municipal*, I (1), pp. 5-8.

Davis Jr, C. A., & Fonseca, F. T. (2007). Assessing the Certainty of Locations Produced by an Address Geocoding System. *Geoinformatica*, XI (1), pp. 103-129.

Debreu, G. (1951). The Coefficient of Resource Utilization. *Econometrica*, XIX (3), pp. 273-292.

Debreu, G. & Scarf, H. (1963). A Limit Theorem on the Core of an Economy. *International Economic Review*, IV (3), pp. 235-246.

Delgado, V. M. S., Miranda-Ribeiro, A., & Soares, J. F. (2013). *Desigualdade Escolar e Desempenho*. In: Fahel, M.; Rambla, X.; Lazarotti, B.; e Bronzo, C. (orgs.) "Desigualdades Educacionais e Pobreza". Editora PUC Minas, Belo Horizonte: 2013, pp. 163-208.

Delgado, V. M. S., & Miranda-Ribeiro, A. (2012). *Medidas comparativas do desempenho e da desigualdade nos exames de proficiência escolar em Minas Gerais por índices unidimensionais*. XV Seminário sobre a Economia Mineira, Diamantina, 2012. Disponível em: <http://web.cedeplar.ufmg.br/cedeplar/seminarios/ecn/ecn-mineira/2012/arquivos/Medidas%20comparativas%20do%20desempenho%20e%20da%20desigualdade%20nos%20exames.pdf> (último acesso: 15/05/2013).

Demange, G., Gale, D., & Sotomayor, M. (1987). A further note on the stable matching problem. *Discrete Applied Mathematics*, XVI (3), pp. 217-222.

Demo, P. (2007). Educação e qualidade. 11ª Edição. Papirus Editora, Campinas: 2007.

Dubins, L. E., & Freedman, D. A. (1981). Machiavelli and Gale-Shapley Algorithm. *The American Mathematical Monthly*, LXXXVIII (7), pp. 485-494.

Ermisch, J. F. (2003). *An Economic Analysis of the Family*. Princeton, NJ, USA: Princeton University Press.

Fernandes, G. A. A. L. (2007) *O sistema de matrícula escolar de São Paulo: uma abordagem à luz da teoria dos jogos*. Dissertação de Mestrado em Teoria Econômica do departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (USP).

Fonseca, F. T., & Zuppo, C. A. (1997). Fim das filas para matrícula em escola pública. *Espaço BH: trabalho, tecnologia e informação na administração municipal*, I (1), pp. 15-20.

Fonseca, F. T., & Zuppo, C. A. (1994). *School Pre-registration and Student Allocation*. In.: Symposium of URISA, pp. 15-20.

Disponível em :

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.8878&rep=rep1&type=pdf> (último acesso 15/05/2013).

Freyer Jr., R. G. (2007). Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century. *The Journal of Economic Perspectives*, XXI, (2), pp. 71-90.

Gale, D., & Shapley, L. S. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, LXIX (1), pp. 9-15.

Gale, D., & Sotomayor, M. (1985a). Ms. Machiavelli and the stable matching problem. *The American Mathematical Monthly*, XCII (4), pp. 261-268.

Gale, D., & Sotomayor, M. (1985b). Some remarks on the stable matching problem. *Discrete Applied Mathematics*, XI (3), pp. 223-232.

Gamoran, A. (1996). Student Achievement in Public Magnet, Public Comprehensive and Private City High Schools. *Educational Evaluation and Policy Analysis*, XVIII (1), pp. 1-18.

Gontijo, M. (2008) *Uma aplicação da Teoria dos Jogos ao mercado do vestibular brasileiro*. Dissertação de Mestrado em Teoria Econômica do departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (USP).

Guimarães, R. R. M. (2012). The Effect of Teacher Qualifications on Student Achievement Gains: Evidence from the FUNDESCOLA Schools in Brazil, 1999-2003. (Tese de Mestrado) Master of Arts Paper International Comparative Education School of Education Stanford University, Stanford: 2012.

Hambleton, R. K., Swaminathan, H., & Rogers, H. (1991). Fundamentals of item response theory. Ed. Sage, Newbury Park: 1991.

- Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, XIX (3), pp.141-164.
- Hanushek, E. A. (1992). The Trade-Off between Child Quantity and Quality. *Journal of Political Economy*. Chicago, C (1), pp. 84-117.
- Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, XXIV (3), pp. 1141-1177.
- Hanushek, E. A. & Luque, J. (2002). *Efficiency and Equity in Schools Around the World*. National Bureau of Economic Research, NBER working paper series. Cambridge Massachusetts 2002.
- Hoxby, C. (2000). Does Competition among Public Schools Benefit Students and Taxpayers? *The American Economic Review*, XC (5), pp. 1209-1238.
- Hoxby, C., & Murarka, S. (2009) Charter School in New York City: Who Enrolls and How They Affect Other Students Achievement. NBER Working Paper, n. 14852, Apr. 2009.
- Juhn, C., Murphy, K. M., & Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of political Economy*, CI (3), pp. 410-442.
- Kesten, O. (2011). *On Two Kinds of Manipulation for School Choice Problems*. Carnegie Mellon University, Tepper School of Business.
- Knuth, D. E. (1976). *Marriages Stables*. Montreal: Les Presses de L'Université de Montreal. Montreal: 1976.
- Lee, V. E., & Bryk, A. S. (1989). A Multilevel Model of the Social Distribution of High School Achievement. *Sociology of Education*, LXII (3), pp. 172-192.
- Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions Insertions and Reversals. *Cybernetics and Control Theory* , X (8), pp. 707-710.
- Linden, W. J. Van der, & Hambleton, R. K. (Orgs). (1997). *Handbook of modern item response theory*. Ed. Springer. New York: 1997.
- Liu, Q., Mailath, G. J., Postlewaite, A., & Samuelson, L. (2012). Stable Matching with Incomplete Information. PIER Working Paper No. 12-032, University of Pennsylvania. Disponível em: http://economics.wustl.edu/files/economics/imce/liu_paper.pdf (último acesso 15/05/2013).
- Luke, D. A. (2004). Multilevel Modeling. Sage University Paper. Series: Quantitative Application in the Social Sciences. Thousand Oaks: 2004.
- Lundberg, S., & Pollak, R. A. (1996). Bargaining and distribution in marriage. *The Journal of Economic Perspectives*, X (4), pp. 139-158.

- Machado, A. F.; Moro, S.; Martins, L.; Rios, J. (2006). *Qualidade do Ensino em Matemática: determinantes do desempenho de alunos em escolas públicas estaduais mineiras*. In: XXXIV Encontro Nacional de Economia, 2006, Salvador. Anais. Belo Horizonte, APEC, 2006 (Disponível em CD-ROM).
- Mare, R. D. (2008). *Educational Assortative Mating in Two Generations*. Texto para discussão do departamento de sociologia, University of California, Los Angeles (2008). Disponível em: http://www.econ.ucla.edu/workshops/papers/History/mare_eam_08.pdf (último acesso em 15/05/2013).
- Mare, R. D. (1991). Five decades of educational assortative mating. *American Sociological Review*, LVI (1), pp. 15-32.
- Martins, D., Davis Jr, C. A., & Fonseca, F. T. (2012). *Geocodificação de endereços urbanos com indicação de qualidade*. Proceedings XIII GEOINFO, (pp. 36-41). Campos do Jordão, SP Brasil.
- Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. Oxford university press, New York: 1995.
- Miranda-Ribeiro, A. (2001) Eficiência no ingresso escolar em Minas Gerais, 1991: uma análise dos diferenciais segundo o status migratório. *Revista Brasileira de Estudos de População*, XVIII (1/2), pp. 75-96.
- Murillo, F. J. T. (2003). Una Panorámica de La Investigación Iberoamericana sobre Eficacia Escolar. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio em Educación*, I (1), pp. 1-14.
- Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics*, LIV (2), pp. 286-295.
- Newcombe, H. B.; Kennedy, J. M.; Axford, S. J.; & James, A. P. (1959). Automatic linkage of vital records. *Science*, CXXX (3381), pp. 954-959.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, XIV (3), pp. 693-709.
- Pápai, S. (2002). Strategyproof and Nonbossy Multiple Assignments. *Journal of Public Economic Theory*, III (3), 257-271.
- Pápai, S. (2000). Strategyproof Assignment by hierarchical Exchange. *Econometrica*, LXVIII (6), pp. 1403-1433.
- Peternelli, L. A. & Mello, M. P. (2011). *Conhecendo o R: uma visão estatística*. Editora UFV, Viçosa: 2011.

Pinto, C. C. X. (2010). Semiparametric Estimation of Peer Effects in Classrooms: Evidence for Brazilian Schools in 2003. Trabalho apresentado nos seminários de UC Irvine School of Social Sciences, 2010. Disponível em:

<http://www.economics.uci.edu/files/economics/docs/econoseminar/w10/pinto.pdf>

Pinto, M. V. (1999). Cadastramento Escolar: democratização do acesso à escola pública. *Informática Pública*, I (2), pp. 139-156.

Pollak, R. A. (1990). Two-Sex Demographic Models. *Journal of Political Economy*, XCVIII (2), pp. 399-420.

Prefeitura Municipal de Belo Horizonte, PBH. (s.d.). *portalpbh.pbh.gov.br*. Acesso em 20 de Novembro de 2012, disponível em:

<http://portalpbh.pbh.gov.br/pbh/ecp/contents.do?evento=conteudo&idConteudo=22185&chPlc=22185&viewbusca=s> (último acesso: 15/05/2015).

Prefeitura Municipal de Belo Horizonte, PBH (2013). Cadastramento Escolar Relatório 2012/2013. Belo Horizonte: 2013. Manuscrito.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>. (último acesso: 15/05/2015).

Ribeiro, L. C., & Kaztman, R. (2008) (orgs.) . *A Cidade contra a Escola Segregação urbana e desigualdades educacionais em grandes cidades da América Latina*. Rio de Janeiro, RJ, Brasil: Letra Capital Editora.

Ricardo, D. (1996 [1817]). Princípios de Economia Política e Tributação. Tradução de Sandroni, P. H. R. Edição original de 1817. Editora Nova Cultural, São Paulo: 1996.

Rigotti, J. I. R. (2001). A transição da escolaridade no Brasil e as desigualdades regionais. *Revista Brasileira de Estudos de População*, XVIII (1/2), pp. 59-73.

Rios-Neto, E. L. G. & Guimarães, R. R. G. M. (2010). *The demography of education in Brazil: inequality of educational opportunities based on Grade Progression Probability (1986-2008)*. Vienna Yearbook of Population Research 2010 (Vol.8), pp. 283-312.

Rios-Neto, E. L. G. & Riani, J. L. (2004) (orgs.). *Introdução à Demografia da Educação*. ABEP - Associação Brasileira de estudos Populacionais. Campinas, 2004.

Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, LXXIII (2), pp. 417-458.

Rodrigues, C. G., Rios-Neto, E. L. G., & Pinto, C. X. (2010). Changes in Test Scores Distribution for Students of the Fourth Grade in Brazil: A Relative Distribution Analysis for the Years 1997 to 2005. Textos para Discussão, 282. C-Micro Working Paper Series, Escola de Economia de São Paulo, FGV-EESP.

Rodrigues, C. G. (2009). *A Relação entre a expansão do acesso ao ensino e o desempenho escolar no Brasil: evidências com base no SAEB para o período de 1997 a 2005*. Tese (Doutorado em Demografia) Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

Roth, A. E. (2012). In 100 Years. Manuscrito não publicado, pp. 1-8. Disponível na internet: <http://www.stanford.edu/~alroth/papers/100%20years.pdf> (último acesso: 08/12/2013).

Roth, A. E. (2008). What have we learned from market design? Hahn Lecture. *Economic Journal* CXVIII (527), pp. 285-310. Disponível em: http://kuznets.fas.harvard.edu/~aroth/papers/2008_Hahn_Lecture_EJ.pdf (último acesso: 15/05/2013).

Roth, A. E. (2002). The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. *Econometrica*, LXX (4), pp. 1341-1378.

Roth, A. E. (1989). Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior*, I (2), pp. 191-209.

Roth, A. E. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of economic Theory*, XXXVI (2), pp. 277-288.

Roth, A. E. (1984a). The evolution of the labor market for medical interns and residents: a case study in game theory. *The Journal of Political Economy*, XCII (6), pp. 991-1016.

Roth, A. E. (1984b). Misrepresentation and stability in the marriage problem. *Journal of Economic Theory*, XXXIV (2), pp. 383-387.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, VII (4), pp. 617-628.

Roth, A. E., & Sotomayor, M. A. (1990). *Two-Sided Matching a Study in Game-Theoretic Modeling and Analysis*. Cambridge, United Kingdom: Cambridge University Press.

Roth, A. E., & Vate, J. H. V. (1990). Random paths to stability in two-sided matching. *Econometrica*, LVIII (6), pp. 1475-1480.

Shapley, L. S. (1951). A value for n-person games. RM-670m Rand Corporation, Santa Monica: 1951. Disponível em: http://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM670.pdf (último acesso: 15/05/2013).

Shapley, L., & Scarf, H. (1974). On cores and indivisibility. *Journal of mathematical economics*, I(1), pp. 23-37.

Shapley, L., & Shubik, M. (1971). Competitive Equilibrium and Game Theory Solutions Part I The core and Value. Cowles Foundation Discussion Paper, no. 316. Cowles Foundation for

Research in Economics Yale University. Disponível em: <http://cowles.econ.yale.edu/P/cd/d03a/d0316.pdf> (último acesso: 15/05/2015).

Shubik, M. (2003). *The Edgeworth, Cournot and Walrasian Cores of an Economy*. Cowles Foundation Discussion Paper, no. 1439. Cowles Foundation for Research in Economics Yale University. Disponível em: <http://cowles.econ.yale.edu/P/cd/d14a/d1439.pdf> (último acesso: 15/05/2013).

Siow, A. (2003) *The Economics of Marriage 30 Years after Becker*. University of Toronto, survey apresentado no encontro de Canadian Economics Association, 2003, Ottawa. Disponível em: <http://homes.chass.utoronto.ca/~siow/papers/becker.pdf> (último acesso: 15/05/2013).

Soares, J. F., & Alves, M. T. (2012). O nível socioeconômico das escolas de educação básica brasileiras. Relatório do Projeto de Pesquisa apoiado pelo Instituto Unibanco realizado pelo Grupo de Avaliação e Medidas Educacionais (GAME) da Universidade Federal de Minas Gerais, Belo Horizonte: 2012. Manuscrito.

Soares, J. F., Rigotti, J. I. R., & Andrade, L. T. (2008). *As desigualdades socioespaciais e o efeito das escolas públicas de Belo Horizonte*. In: Ribeiro, L. C., & Kaztman, R. (organizadores). *A Cidade contra a Escola Segregação urbana e desigualdades educacionais em grandes cidades da América Latina*. Rio de Janeiro, RJ, Brasil: Letra Capital Editora, pp. 119-144.

Soares, T. (2005). Utilização da Teoria da Resposta ao Item na Produção de Indicadores Sócio-Econômicos. *Pesquisa Operacional*, XXV (1), pp. 83-112.

Soares, T.; Pereira, D. (2002). Estudo de critérios de adequação para modelos da teoria da resposta ao item (TRI) aplicado ao caso do ensino fundamental da micro-região de Juiz de Fora em 1999. *Educação em Foco*, VI (2), pp. 91-108.

Sönmez, T. (1995). Implementation in Generalized Matching Problems. *Journal of Mathematical Economics*, XXVI (4), pp. 429-439.

Sotomayor, M. A. O. (1998). The strategy structure of the college admissions stable mechanisms. Mimeo, Universidade de São Paulo. Disponível em: http://www.fea.usp.br/feaecon//media/fck/File/THE_STRUCTURE_STRATEGY_OF_THE_COLLEGE_ADMISSION_STABLE_MECHANISMS.pdf (último acesso: 15/05/2013).

Sotomayor, M. (1996a). Mecanismos de Admissão de Candidatos às Instituições. Modelagem e Análise à Luz da Teoria dos Jogos. *Brazilian Review of Econometrics*, XVI (1), pp. 25-63.

Sotomayor, M. (1996b). A non-constructive elementary proof of the existence of stable marriages. *Games and Economic Behavior*, XIII (1), pp. 135-137.

Telser, L. G. (1994). The Usefulness of Core Theory in Economics. *Journal of Economic Perspectives*, VIII (2), pp. 151-164.

Teo, C.-P., Sethuraman, J., & Tan, W.-P. (2001). Gale-Shapley Stable Marriage Problem Revisited: Strategic Issues and Applications. *Management Science*, XLVII (9), pp. 1252-1267.

Umbelino, G. (2012). Simulações de distribuição espacial domiciliar e projeções demográficas intraurbanas com apoio de geotecnologias. Tese (Doutorado em Demografia) Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

Weiss, Y. (1997). The Formation and Dissolution of Families Why Marry Who Marries and What Happens Upon Divorce. In: M. R. Rosenzweig, & O. Stark, *Handbook of Population and Family Economics* (pp. 81-123). Elsevier Science.

Zobel, J., & Dart. P. (1995). Finding approximate matches in large lexicons. *Software-Practice and Experience*, XXV (3), pp. 331–345.

10. ANEXO A

Resolução Conjunta da SME-BH e SEE-MG Estabelece as normas para a realização do Cadastro Escolar:

Sábado, 28 de Julho de 2012
Ano XVIII - Edição N.: 4121
Poder Executivo

Secretaria Municipal de Educação

RESOLUÇÃO CONJUNTA SEE/MG – SMED/BH N.º 01, DE 15 DE JUNHO DE 2012*

Estabelece normas para a realização do Cadastro Escolar para o ensino fundamental e matrícula na rede pública de ensino de Belo Horizonte e dá outras providências.

A Secretária de Estado de Educação de Minas Gerais e o Secretário Municipal de Educação de Belo Horizonte, no uso de suas atribuições e considerando o disposto no § 3º do artigo 208 e no artigo 211 da Constituição Federal, no § 3º do artigo 198 da Constituição Estadual, na Emenda Constitucional n.º 14/96, no inciso II do § 1º do artigo 5º da Lei n.º 9394/96, Resolução SEE n.º 521/2004, no inciso I do artigo 158, da Lei Orgânica Municipal de Belo Horizonte e Lei Estadual n.º 16056, de 24/04/2006,
RESOLVEM:

Capítulo I Do Cadastramento

Art. 1º O Cadastro Escolar será realizado com o objetivo de proceder à inscrição dos candidatos a vagas no ensino fundamental, em 2013, na Rede Pública de Ensino de Belo Horizonte.

Parágrafo único. O Cadastro Escolar será coordenado por comissão constituída com representantes da Secretaria de Estado de Educação, Superintendências Regionais de Ensino Metropolitana A, B e C e Secretaria Municipal de Educação.

Art. 2º A inscrição para o Cadastro Escolar estará aberta aos candidatos residentes em Belo Horizonte e será realizada no período de 06/08/2012 a 10/08/2012, nas agências dos Correios relacionadas no Anexo desta Resolução, inclusive para os candidatos com deficiência, transtornos globais do desenvolvimento e altas habilidades/superdotação.

Parágrafo único. A inscrição é isenta de pagamento de taxas pelo candidato.

Art. 3º Os alunos da Educação Infantil de instituições municipais terão vaga garantida no ensino fundamental da Rede Municipal de Educação, conforme Instrução de Serviço a ser publicada pela Secretaria Municipal de Educação.

Art. 4º Deverá inscrever-se no Cadastro Escolar:

I – em razão da Decisão Judicial do Tribunal Regional Federal da 5ª Região, publicada no DOU de 02/07/2012, toda criança que completar 6 (seis) anos de idade até 31 de março de 2013, para início do ensino fundamental, conforme as diretrizes das Resoluções CNE/CEB, n.º 01/10 e 06/10;

II - aluno matriculado na Educação Infantil em instituições municipais, cuja família não optar pela escola da rede municipal de educação indicada pela Gerência Regional de Educação;

III - candidato a vaga na rede pública de Belo Horizonte nos demais anos ou ciclos do ensino fundamental, oriundos de outras redes de ensino;

Art. 5º A inscrição no Cadastro Escolar será realizada pelo pai, mãe ou responsável pelo candidato, mediante a apresentação (original e cópia) dos seguintes documentos:

- I - certidão de nascimento do candidato;
- II - conta de luz do mês de maio ou junho de 2012, da residência do candidato em Belo Horizonte;
- III - caderneta, boletim ou documento comprobatório de escolaridade expedido pela escola de origem, nos casos de retorno aos estudos e ingresso na rede pública de Belo Horizonte, quando for oriundo da rede particular de ensino.

Art. 6º O encaminhamento para matrícula dos candidatos inscritos no Cadastro Escolar para o início do ensino fundamental obedecerá, prioritariamente, ao zoneamento escolar.

§ 1º Esgotadas as vagas das escolas de cada zoneamento, o candidato inscrito para os demais anos ou ciclos do ensino fundamental será encaminhado para a escola pública de acesso mais fácil.

§ 2º Não será permitida a realização de exame de seleção.

§ 3º A Comissão do Cadastro Escolar comunicará, por correspondência, ao candidato inscrito, o nome da escola onde deverá efetuar a sua matrícula para 2013.

§ 4º Ocorrendo a necessidade de comprovação de endereço dos candidatos, as informações prestadas pelos pais, por ocasião da inscrição, poderão ser verificadas pelas Gerências Regionais de Educação e Superintendências Regionais de Ensino Metropolitana A, B e C.

I - O candidato com endereço comprovadamente correto terá confirmada sua alocação na escola de sua jurisdição.

II - O candidato que não residir no endereço informado será alocado em escola onde houver vaga.

Art. 7º As Agências dos Correios não receberão inscrições de aluno já matriculado no ensino fundamental da rede pública, ou que já tem garantida a continuidade de estudos na rede pública.

Capítulo II Da Matrícula

Art. 8º A matrícula dos inscritos no Cadastro Escolar será unificada na rede pública de ensino – estadual e municipal – e será realizada no período de 17 a 21 de dezembro de 2012.

§ 1º Terá vaga assegurada o candidato cadastrado que efetuar a matrícula no prazo estabelecido.

§ 2º O candidato que não realizar a matrícula no prazo previsto será reencaminhado, pelas Gerências Regionais de Educação, para escola onde houver vaga remanescente.

§ 3º Os candidatos e os alunos portadores de carteira de identidade deverão apresentá-la no ato da matrícula, ou de sua renovação, cabendo à escola anotar na ficha do aluno o número do respectivo RG, o nome do órgão expedidor do documento e a data de sua expedição.

Art. 9º Para garantia de vaga no ano letivo de 2013, o aluno já matriculado no ensino fundamental e o aluno de educação infantil das escolas municipais, que optarem por permanecer na Rede Municipal de Educação, deverão efetivar ou renovar sua matrícula no período estipulado pela escola, de acordo com Instrução de Serviço a ser publicada pela Secretaria Municipal de Educação.

Art. 10 A escola de ensino fundamental que recebe alunos em prosseguimento de estudos deverá indicar o número de vagas disponíveis, que serão distribuídas proporcionalmente ao número de candidatos de cada escola de origem.

Parágrafo único. Cabe à escola de origem organizar a listagem por escola de destino dos alunos a serem encaminhados, realizando-se sorteio, quando o número de interessados for superior ao número de vagas.

Art. 11 As escolas da rede municipal de Belo Horizonte deverão remeter a relação dos alunos prováveis concluintes do 2º e 3º ciclos do ensino fundamental, no ano de 2012, às escolas da rede estadual desta capital, para que sejam garantidas as vagas necessárias à continuidade de estudos.

Art. 12 A matrícula no ensino médio de concluintes do ensino fundamental da rede pública, para 2013, será efetivada preferencialmente no diurno, de acordo com os seguintes critérios:

I - com prioridade, para os concluintes do ensino fundamental na própria escola, quando esta oferecer o ensino médio;

II - quando o número de vagas for insuficiente para continuidade na própria escola, terá prioridade de opção o aluno mais novo;

III - esgotados os procedimentos previstos nos incisos I e II e constatada demanda excedente para ingresso no ensino médio, será apresentada listagem de escolas da rede pública que ministram o ensino médio para que o aluno escolha, em ordem de preferência, 04 (quatro) escolas com o respectivo turno:

a) o encaminhamento para matrícula na unidade de ensino indicada pelo aluno será feito respeitando-se o limite de vagas e a ordem de preferência e de idade;

b) quando o número de optantes para determinada unidade escolar for superior ao número de vagas existentes, terá prioridade para matrícula o aluno mais novo - considerando dia, mês e ano de nascimento -, obedecida a ordem crescente de idade;

IV - Os alunos a que se refere o inciso III, que não forem contemplados em nenhuma das quatro opções, serão encaminhados, preferencialmente, para escolas de mais fácil acesso, onde houver vagas.

Art. 13 A Comissão de Cadastro e Matrícula deverá indicar a necessidade de criação de novas vagas, quando constatado déficit de oferta.

Art. 14 Em nenhuma hipótese, a matrícula em escola pública poderá ser condicionada:

I - ao pagamento de taxa ou a qualquer forma de contribuição;

II - à realização de exame de seleção.

Art. 15 Esta Resolução entra em vigor na data de sua publicação e revoga a Resolução Conjunta SEE-SMED/BH nº 03, de 13 de maio de 2011.

Secretaria de Estado de Educação, em Belo Horizonte, aos 25 de julho de 2012.

Afonso Celso Renan Barbosa
Secretário Municipal de Educação
Ana Lúcia Almeida Gazzola
Secretária de Estado de Educação

(*) Republicada por sofrer alterações no art. 4º da publicação do “DOM” de 27/6/2012 e da republicação do “DOM” de 19/7/2012.

ANEXO

Relação das Agências de Correios de Belo Horizonte

UNIDADE	LOGRADOURO	BAIRRO
AC AARÃO REIS	RUA RIO DE JANEIRO, 234	CENTRO
AC BAIRRO BELVEDERE	AVE. LUIZ PAULO FRANCO	BELVEDERE
AC BAIRRO CASTELO	AV. MIGUEL PERRELA, 43	CASTELO
AC BAIRRO DA BARROCA	AV. AMAZONAS, 3790	PRADO
AC BAIRRO DE SANTA TEREZA	PRAÇA DUQUE DE CAXIAS, 143, LOJA 03	SANTA TEREZA
AC BAIRRO OURO PRETO	RUA JOÃO ANTONIO CARDOSO	OURO PRETO
AC BAIRRO SANTA EFIGÊNIA	AV. BRASIL, 67	SANTA EFIGÊNIA
AC BAIRRO UNIÃO	RUA ALBERTO CINTRA, 379	UNIÃO
AC BARREIRO	AV. SINFONIA BROCHADO, 550	BARREIRO
AC BARRO PRETO	RUA OURO PRETO, 350/356	BARRO PRETO

AC BH SHOPPING	RODOVIA BR 356 LOJA BH 35	BELVEDERE
AC CONJUNTO IAPI	AV. PRESIDENTE ANTONIO CARLOS 8151	PAMPULHA
AC ENCOMENDAS	RUA GOIÁS, 77	CENTRO
AC FRANCISCO SALES	AV. FRANCISCO SALES, 898 LOJA 18	SANTA EFIGÊNCIA
AC JARAGUA	ANEL RODOVIÁRIO CELSO MELLO AZEVEDO 20901 – BLOCO 02	UNIVERSITÁRIO
AC JUSTINÓPOLIS	RUA PADRE PEDRO PINTO, 6709	VENDA NOVA
AC PARANÁ	AV. PARANÁ, 477	CENTRO
AC PARQUE INDUSTRIAL	AV. CARDEAL EUGÊNIO PACELLI, 1801	CIDADE INDUSTRIAL
AC PEDRO II	AV. PRES. CARLOS LUZ, 126	CAIÇARA
AC PRESIDENTE JUSCELINO KUBITSCHKE	AV. AFONSO PENA, 1270	CENTRO
AC BAIRRO SANTA LÚCIA	RUA KEPLER, 405	SANTA LÚCIA
AC SAVASSI	RUA PERNAMBUCO, 1322	FUNCIONÁRIOS
AC SHOPPING CIDADE	RUA RIO DE JANEIRO, LOJAS 19/20	CENTRO
AC SHOPPING DEL REY	AV. PRES. CARLOS LUZ, LOJA 1147	CAIÇARAS
AC VENDA NOVA	RUA PADRE PEDRO PINTO, 780	VENDA NOVA
AC VIA SHOPPING BARREIRO	AV. AFONSO VAZ DE MELO, LOJA 404 A 407	BARREIRO
ACC CHAPEX	RUA MARIA AMÉLIA MAIA, 620	SÃO BERNARDO
ACC CLIPEL	RUA ÉRICO VERÍSSIMO, 2368, LOJA A	SANTA MÔNICA
ACC EMPÓRIO DO PÃO	AV. ARTUR BERNARDES, 220	SÃO BENTO
ACC JULITA (ANN'S)	AV. SANTA TEREZINHA, 810	ITATIAIA
ACC KITCOLOR	AC. PROF. MARIO WERNECK, 1938 LOJA 04	BURITIS
ACC RICCI	AV. BARÃO HOMEM DE MELO, 2861	ESTORIL
ACC SANTA AMÉLIA	AV. GUARAPARI, 886 LOJA 1	SANTA AMÉLIA
ACC VITÓRIA	RUA PADRE EUSTÁQUIO, 2936 B	PADRE EUSTÁQUIO
ACCA NOVO HORIZONTE	RUA WALDOMIRO LOBO	GUARANI
ACCA VAN GOGH	AV. CORONEL JOSÉ DIAS BICALHO	SÃO JOSÉ PAMPULHA
ACF BAHIA	RUA DA BAHIA, 1161	1161
ACF CAMPOS SALES	RUA JUSCELINO BARBOSA, 280	NOVA SUÍÇA
ACF CAPITAL	AV. AUGUSTO DE LIMA, 595	CENTRO
ACF CARTA & CIA	RUA PLATINA, 1385	CALAFATE
ACF CIDADE NOVA	RUA GONÇALO COELHO, 56	56
ACF GUTIERREZ	RUA MARECHAL HERMES, 280	GUTIERREZ
ACF HABIT	AV. CARANDÁI, 737	FUNCIONÁRIOS
ACF JMA	AV. SILVIANO BRANDÃO, 1878	HORTO
ACF MINAS CENTRO	AV. AUGUSTO DE LIMA, 877	CENTRO

ACF PIO XII	AV. DO CONTORNO, 9159 – LOJA 04	BARRO PRETO
ACF PLANALTO	AV. DR. CRISTIANO GUIMARÃES, 1910	PLANANTO
ACF PONTO 1	RUA DO OURO, LOJA 31 E 35	SERRA
ACF POST FREE	RUA MARTIM DE CARVALHO, 273	SANTO AGOSTINHO
ACF PRUDENTE	AV. PRUDENTE DE MORAIS, 1030	SANTA LÚCIA
ACF RAJA	AV RAJA GABAGLIA, 1097	CIDADE JARDIM
ACF SION	RUA BOA ESPERANÇA, 358	SION
ACF VIA NORTE	AV. DOS BANDEIRANTES	SION
ACF ZERAYD DE MENEZES	RUA PARAÍBA, 1050	FUNCIONÁRIOS
AF BELO HORIZONTE	AV. AFONSO PENA	CENTRO
AGF POMPEIA	AV. ALPHONSUS DE GUIMARÃES, 125	SANTA EFIGÊNIA
AGF ABILIO MACHADO	AV. ABILIO MACHADO	INCONFIDÊNCIA
AGF AUGUSTO DE LIMA	AV. AUGUSTO DE LIMA, 444, LOJAS 01 A 14	CENTRO
AGF CAETÉS	RUA DOS CAETÉS, 939	CENTRO
AGF CARLOS PRATES	RUA PADRE EUSTÁQUIO	CARLOS PRATES
AGF CORAÇÃO EUCARÍSTICO	RUA CORAÇÃO EUCARÍSTICO DE JESUS	CORAÇÃO EUCARÍSTICO
AGF DAVID SARNOFF	AV. GENERAL DAVID SARNOFF, 5230, LOJA 12	CIDADE INDUSTRIAL
AGF ESTAÇÃO CENTRAL	AV. DOS ANDRADAS, LOJAS 103 E 104	CENTRO
AGF EXTRA MINAS SHOPPING	AV. CRISTIANO MACHADO, 4000, LOJA 77	CIDADE NOVA
AGF FLORESTA	AV. DO CONTORNO	FLORESTA
AGF GETÚLIO VARGAS	AV. GETÚLIO VARGAS	FUNCIONÁRIOS
AGF JACUÍ	RUA JACUÍ	CONCÓRDIA
AGF JARDIM MONTANHES	AV. DOM PEDRO II	JARDIM MONTANHES
AGF JOSÉ CANDIDO DA SILVEIRA	AV. JOSÉ CANDIDO DA SILVEIRA	CIDADE NOVA
AGF LIBERDADE	RUA BOAVENTURA	LIBERDADE
AGF PEDRO I	AV. PEDRO I, LOJA 08	ITAPOÃ
AGF SANTA INES	RUA CONTAGEM, LOJAS 01 E 02	SANTA INES
AGF SÃO LUIZ	ALAMEDA DAS ACÁCIAS ESQUINA COM AV. ABRAHÃO CARAM	SÃO LUIZ
AGF TERMINAL RODOVIÁRIO DE BELO HORIZONTE	PRAÇA RIO BRANCO, 325	CENTRO
AGF URSULA PAULINO	RUA URSULA PAULINO, ATE 1209/1210 – LOJA 09	CINQUENTENÁRIO
AGF VILARINHO	AV. VILARINHO, LOJAS 03 A 06	VENDA NOVA

Cronograma do cadastro escolar das atividades realizadas em 2012 para o ano letivo de 2013 conforme relatório de atividades da unidade de cadastro escolar da SME-BH:

Tabela A.1 - Cronograma do Cadastro Escolar 2012/2013 para o ensino fundamental

Nº	ATIVIDADE	INÍCIO	TÉRMINO	RESPONSÁVEL
1	Briefing com agência de publicidade	28/03	28/03	Assessoria de comunicação
2	Distribuição do material de divulgação	07/05	08/05	Assessoria de comunicação
3	Publicação da Resolução Conjunta	17/05	17/05	SEE/SMED
4	Divulgação na mídia	04/06	22/06	SEE/SMED/Ass. Com.
5	Divulgação no jornal do ônibus	01/06	16/06	SEE/SMED/Ass. Com.
6	Impressão dos formulários de inscrição	25/05	28/05	PRODABEL
7	Entrega dos formulários aos Correios	23/07	23/07	Correios
8	Inscrição nos Correios	06/08	10/08	Correios
9	Prazo para prorrogação inscrição	13/08	17/08	Correios
10	Digitização	13/08	24/08	Correios
11	Recepção do arquivo dos Correios	28/08	28/08	Correios
12	Confirmação quadro de vagas	01/08	31/08	SEE/SMED
13	Análise de duplo cadastro	29/08	29/08	Cadastro Escolar
14	Geração e envio do Arquivo à CEMIG	03/09	03/09	Cadastro Escolar
15	Trabalho na CEMIG	03/09	14/09	CEMIG
16	Recebimento e Carga do arquivo da CEMIG	17/09	17/09	Cadastro Escolar
17	Geração e envio do arquivo para alocação	20/09	29/09	Cadastro Escolar
18	Alocação Geográfica	01/10	11/10	PRODABEL/Geo
19	Recebimento e carga do arquivo com as alocações	16/10	16/10	Cadastro Escolar
20	Análise de pontos críticos	17/10	28/11	SMED
21	Emissão de carta de encaminhamento	29/11	30/11	PRODABEL
22	Geração do arquivo de divulgação na internet	10/12	10/12	Cadastro Escolar/Ass. Com.
23	Liberação do programa do sobrecadastro nas regionais	09/12	09/12	Cadastro Escolar
24	Entrega das cartas de encaminhamento	10/12	14/12	Correios
25	Divulgação do resultado na internet	10/12	-	Assessoria
26	Atendimento das retificações e sobrecadastro	10/12	31/01/13	Regionais
27	Matrículas	17/12	21/12	Escolas

Fonte: Relatório do Cadastro Escolar 2012/2013, direção de cadastro Secretaria Municipal de Educação.

Abaixo um modelo de carta enviada a um aluno que teve endereço validado e a matrícula foi recomendada para a escola municipal Ignácio de Andrade Melo, no bairro São José, região noroeste. A carta é nominal para o aluno. Por motivos de sigilo, o nome e número do cadastro do aluno foram ocultados.

Figura A.1 – Carta de Resposta a um aluno que teve o cadastro aceito e matrícula sugerida

CADASTRO ESCOLAR 2012/2013
Secretaria de Estado da Educação de Minas Gerais
Secretaria Municipal de Educação de Belo Horizonte

Em resposta à sua inscrição no Cadastro Escolar **2012/2013** em Belo Horizonte, informamos que você deverá comparecer, no período de **17 a 21 de Dezembro de 2012** na:

E.M. IGNACIO DE ANDRADE MELO
RUA VIOLETA DE MELO, 988 SAO JOSE

Caso você necessite de mais algum esclarecimento, dirija-se à sua Administração Regional, cujo endereço é:

PAMPULHA 3277-7441
AVE ANTONIO CARLOS, 7596 SUBSOLO- SAO LUIZ

Na oportunidade desejamos a você e a toda a sua família um Feliz Natal repleto de novas realizações!

ATENÇÃO

Lembramos que o não comparecimento à escola indicada, na data marcada, poderá implicar na perda da vaga.

MOD-CADESC-(cartaesc_verso.prt)-FSM

Número do seu Cadastro: _____

Abaixo um modelo de carta enviada para um aluno que enviou endereço incorreto. Nesse caso solicita-se aos pais se dirigirem à regional mais próxima para nova readequação de vaga. Por motivos de sigilo, o nome e número do cadastro do aluno foram ocultados.

Figura A.2 – Carta enviada para aluno que teve cadastro rejeitado por endereço incorreto

CADASTRO ESCOLAR 2012/2013
 Secretaria de Estado da Educação de Minas Gerais
 Secretaria Municipal de Educação de Belo Horizonte

Em resposta à sua inscrição no Cadastro Escolar **2012/2013**, feita nos Correios, informamos que você deverá comparecer à Gerência de Educação da Regional abaixo mencionada no período de **17 a 21 de Dezembro de 2012**, de **09:00 às 12:00** horas ou de **14:00 às 17:00** horas levando esta carta.

PAMPULHA
AVE ANTONIO CARLOS, 7596 SUBSOLO- SAO LUIZ 3277-7441

Na oportunidade desejamos a você e a toda sua família um Feliz Natal e um ano Novo repleto de novas realizações!

Atenciosamente
 Comissão de Matrícula

ATENÇÃO

MOD-CADESC-(cadastro+verso).pm)-FSM
 Lembramos que o não comparecimento à Administração Regional acima mencionada, no período indicado, poderá implicar na perda da vaga.

Número do seu Cadastro: _____

Abaixo um modelo de carta enviada para um aluno que mora fora dos municípios de Belo Horizonte, no caso a cidade vizinha de Sabará – MG. Próxima aos limites da região nordeste e leste de Belo Horizonte. Por motivos de sigilo, o nome e número do cadastro do aluno foram ocultados.

Figura A.3 – Carta para aluno que teve cadastro rejeitado por pertencer a outro município

CADASTRO ESCOLAR 2011/2012
 Secretaria de Estado da Educação de Minas Gerais
 Secretaria Municipal de Educação de Belo Horizonte

[Ocultado]

Informamos que sua inscrição ao Cadastro Escolar para 2012, efetuada em agência dos Correios de Belo Horizonte, não poderá ser atendida, pois seu endereço pertence ao município de SABARA

Para que você possa garantir sua vaga para o próximo ano, procure a Secretaria Municipal de Educação do seu município.

Atenciosamente,

Comissão de Matrícula.

Número do seu Cadastro: _____ [Ocultado]

MOD-insc.cadastral-(inscrição cadastral-verso.prm)-FSM

Fonte: diretoria de Cadastro da Secretaria Municipal de Educação de Belo Horizonte Abr./2013.

Abaixo o modelo da ficha do cadastro escolar a ser preenchida pelos pais do momento de cadastro nas agências dos correios listadas na resolução. Após o preenchimento da ficha, um funcionário dos Correios na agência lança as informações cadastrais no computador e as envia para unidade de cadastro da SME-BH.

Figura A.4 – Modelo da ficha cadastral

CADASTRO ESCOLAR 2012/2013 ENSINO FUNDAMENTAL																																																											
SECRETARIA MUNICIPAL DE EDUCAÇÃO Inscrição Nº 21439-5																									CARIMBO DA AGÊNCIA:										SECRETARIA DE ESTADO DE EDUCAÇÃO 203000018																								
NOME DO CANDIDATO																																																											
I. O Cadastro Escolar não é matrícula e destina-se aos residentes no município de Belo Horizonte para: - candidatos nascidos até 31/12/2007, para início do Ensino Fundamental, a serem atendidos nas escolas estaduais e municipais de Belo Horizonte; - candidatos oriundos da rede particular de ensino e - candidatos que desejam retomar aos estudos. II. 1ª etapa: Seu filho receberá uma carta informando a escola em que deverá ser matriculado. A indicação da escola obedecerá, prioritariamente, o zoneamento escolar. Se não receber a carta até 17/12/2012, dirigir-se à Regional mais próxima da residência. 2ª etapa: Período de matrícula nas escolas: de 17/12/2012 a 21/12/2012 <u>Se a matrícula não for realizada neste período, seu filho será considerado desistente e perderá a vaga na escola para onde foi encaminhado.</u> III. A partir de 10/12/2012 a escola para matrícula poderá ser consultada via Internet no endereço www.pbh.gov.br. IV. Lembrem-se: Qualquer informação ou documento falso tornará nulo o cadastro de seu filho.																																																											
Cliente: _____																									TELEFONE _____										RAMAL _____																								
(Assinatura do responsável pelo candidato)																									(VIA DO CANDIDATO)																																		
3< CADASTRO ESCOLAR 2012/2013 (VIA DOS CORREIOS)																																																											
Inscrição Nº 21439-5																																																											
NOME DO CANDIDATO																																																											
DATA DE NASCIMENTO _____																									CONTA DE LUZ: _____																																		
(Não cadastrar nascidos após 31/12/2007)																									(Número da instalação)																																		
CEP residencial: _____																									Número do Imóvel: _____																																		
_____																									TELEFONE _____																																		
(Nome legível do solicitante)																									RAMAL _____																																		
MARQUE COM UM "X" O ANO QUE IRÁ CURSAR EM 2013:																																																											
1 2 3 4 5 6 7 8 9																																																											
Transcrever abaixo os dados do documento escolar:																																																											
NOME DA ESCOLA																																																											
ANO DE EMISSÃO DO DOCUMENTO																									MUNICÍPIO DE LOCALIZAÇÃO DA ESCOLA																																		
_____																									_____																																		
CARIMBO DA AGÊNCIA																									Assinatura do responsável pelo candidato:																																		
203000018																									_____																																		

MOD-Cad Escolar-(cadastroescolarnovo-verso prn) FSM

Fonte: diretoria de Cadastro da Secretaria Municipal de Educação de Belo Horizonte Abr./2013.

11. ANEXO B

Os algoritmos utilizados nessa tese foram elaborados para linguagem R¹²¹, trata-se de uma linguagem livre (*open source*) com inúmeros contribuidores pelo mundo. O R é uma linguagem *object oriented* e é essencialmente voltado para procedimentos estatísticos, sendo eficiente para trabalhar com cálculos estatísticos comumente empregados, e permite trabalhar gráficos com grande maleabilidade. Para uma boa rápida e boa introdução à linguagem, pode se consultar Peternelli & Mello (2011). Em inglês a literatura é bastante ampla, e para quem instalar o *shell* do R no seu sistema operacional o R instantaneamente baixa em sua livreria uma relação de manuais úteis.¹²²

Infelizmente o R não possui uma interface gráfica userfriendly para o usuário não iniciado nesta programação, sendo assim, é preciso ter alguma familiaridade com a linguagem para entender os passos realizados. O algoritmo foi transformado em uma função com 5 parâmetros para o algoritmo *one to one* e 6 parâmetros para o algoritmo *many to one*. A função se chama *'gsa'* e pode ser invocada da seguinte maneira (os inputs da programação estão em fonte *courier.new*):

- `gsa(m, n, preference.row, preference.col, first)`

A implementação se dá na forma matricial e segue alguns princípios:

1. Os homens estão sempre nas linhas e as mulheres estão sempre nas colunas, o número de homens é dado por *m* e o número de mulheres é dado por *n*. São objetos da classe dos números inteiros.
2. O parametro `preference.row` especifica a matriz de preferências para os homens. É um objeto de classe matricial para o R. Em uma matriz 4x4, 4 homens: {A, B, C, D} e 4

¹²¹ Para maiores detalhes: R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

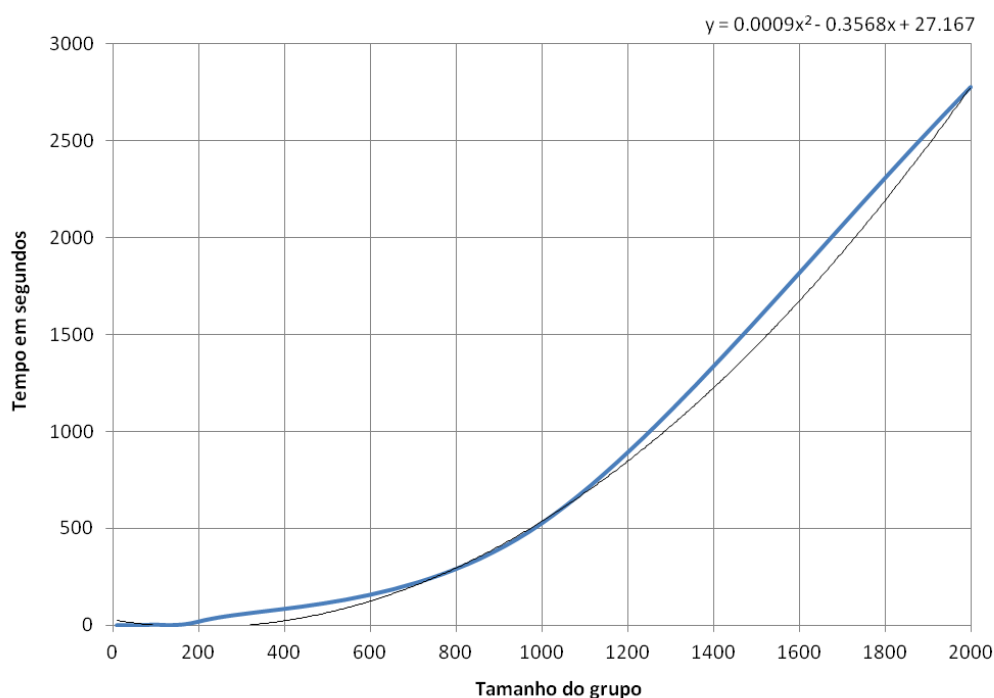
¹²² Também muito úteis são as comunidades na internet para resolução de dúvidas de programação: Este autor usou intensivamente a comunidade: <http://r.789695.n4.nabble.com/>. Em dezembro de 2011 o autor postou sua primeira versão do algoritmo Gale & Shapley (1962) para o R: <http://r.789695.n4.nabble.com/Gale-Shapley-Algorithm-for-R-t4240809.html>. A comunidade Stackoverflow pode ser também de grande auxílio: <http://stackoverflow.com/questions/tagged/r>.

mulheres: {a, b, c, d}. Se por exemplo o homem A, prefere as mulheres na seguinte ordem: {c, a, d, b} a primeira linha da matriz `preference.row` inserida como parâmetro deverá ser: {2, 4, 1, 3}, a posição das mulheres 'a', 'b', 'c', 'd', devem ser mantida e as células recebem o valor do ordenamento para o homem daquela linha, no caso, o homem A. Como 'c' é a mulher mais preferida por ele, a terceira posição da primeira linha de preferência dos homens recebe valor 1, a mulher 'a' é a segunda mais preferida e a primeira posição recebe valor 2, e assim por diante.

3. O parametro `preference.col` especifica a matriz de preferências para as mulheres. Também é um objeto de classe matricial para o R, deve se notar porem que o preenchimento dessa matriz se dá por colunas. Como exemplo, suponha que a mulher 'a' prefira os homens na seguinte ordem: {D, C, B, A}, então, a primeira coluna de preferência das mulheres terá os seguinte números: {4, 3, 2, 1}.
4. O parametro `first` pode assumir apenas dois valores: 1 e 2. Para o caso de ser igual a 1, os homens propõem primeiro, e quando igual a 2 as mulheres propõem primeiro.
5. O princípio básico do algoritmo de pareamento funciona por meio de uma matriz de zeros e uns, tal como apresentado no capítulo 3. Pensando nos homens propondo primeiro, o algoritmo funciona com os homens propondo ao par mais preferido. Caso uma mesma mulher receba mais de uma proposta, ela irá segurar uma, e apenas uma, opção mais preferida e ficará com ela até o final do algoritmo ou até receber outra proposta melhor. Em termos da linguagem, o algoritmo não pára enquanto houver empates, no caso de homens propondo primeiro, enquanto alguma coluna somar mais do que 1 e, no caso das mulheres, enquanto alguma linha for maior do que 1.
6. Como visto no capítulo 3, o número máximo de passos que o algoritmo pode levar é $N^2 - 2N + 2$, em que N é um número inteiro com o número de jogadores em um grupo. Dessa forma, observa-se que para $N \geq 1$, o crescimento do número máximo de passos é quadrático, o de dificulta a implementação do algoritmo para grandes números. Para a amostra completa de 16.354 alunos o algoritmo levava dias sem encontrar solução (detalhes do sistema operacional e especificação da máquina abaixo), a FIG. B.1 mostra algumas simulações para o tempo computacional do algoritmo.

Para os cálculos dos algoritmos foi usada a versão 12.15.2 da linguagem R. O Sistema Operacional é o Windows 7, o CPU possui processador Intel Core, I5, 3,20 GHz, e 12 GB de memória RAM. Para rodar um algoritmo com 2.000 agentes em um grupo, leva-se um pouco mais de meia hora de tempo computacional. Isso depende um pouco dos inputs das matrizes `preference.row` e `preference.col`.

Figura B.1 – Simulações do tempo computacional do algoritmo *deferred-acceptance* no R



Fonte: cálculos próprios.

Desenvolvimentos futuros para esse trabalho podem levar em conta algoritmos computacionalmente eficientes. Raramente o algoritmo *deferred-acceptance* atinge o número máximo de cálculos. Caso a função de progressão acima se mantivesse para valores mais elevados, o cálculo do algoritmo para 16.354 alunos levaria quase três dias, e para a base completa, com aproximadamente 22 mil alunos levaria aproximadamente 5 dias. Entretanto algumas simulações com bases maiores sugerem que para grandes dados o algoritmo fica ainda mais lento (ineficiente). Abaixo segue o código utilizado e alguns exemplos (observação, ao transportar esses códigos para o R, cuidado com possível quebra de linha):

```
#####
# SCRIPT NO R para o algoritmo de Gale-Shapley para casamentos
# Autor: Victor Maia
# última versão: 07/01/2013
#####

gsa <- function(m, n, preference.row, preference.col, first)
{
  # m é Número de Linhas
  # n é Número de Colunas
  # preference.row é a matriz que elenca as preferências dos homens (rows)
  # preference.col é a matriz que elenca as preferências das mulheres (cols)
  # first é 1 for row (men); 2 for column (women)
  #
  ##### UMA FUNÇÃO AUXILIAR MUITO ÚTIL:
  #
  # Função obtida do exemplo de ?match. Serve como negação de %in% NOT IN:
  "%w/o%" <- function(x, y) x[!x %in% y] #-- x without y
  #
  #####

  # first valor é 1 (para homens propondo) e 2 para mulheres:
  s <- NULL
  test_s <- NULL
  loop <- 1 # Contagem das iterações.
  store <- NULL
  r <- NULL

  # Homens propõem primeiro:

  if (first==1)
  {
    step.1 <- preference.row
    step.1[step.1>1] <- 0
    s <- apply(step.1,2,sum)
    test_s <- s>1
    while (any(test_s==TRUE)==TRUE)
    {
      if (any(test_s==TRUE)==TRUE) {
        position1 <- which(s>1)
        position2 <- vector('list')
        position3 <- vector('list')
        position4 <- NULL
        position5 <- 1:n
        aux <- rep(1:m,n)
        for (k in
1:length(position1)){position2[[k]] <- which(step.1[,position1[k]]==1)

position3[[k]] <- which.min(preference.col[position2[[k]],position1[k]])

position3[[k]] <- position2[[k]]%w/o%position2[[k]][position3[[k]]]

step.1[position3[[k]],position1[k]] <- 0}

        position4 <- which(step.1==1)
        position4 <- aux[position4]
        position5 <- position5[-position4]
      }
    }
  }
}
```

```

                                store <- append(position5, store)
                                r <- rle(sort(store))
                                for (j in
position5){step.1[j,][preference.row[j,]==r$lengths[r$values==j]+1] <- 1}
                                s <- apply(step.1,2,sum)
                                test_s <- s>1
                                }else{return(step.1)}

                                loop <- loop + 1
                                cat("interações =",loop,'\n')
                                flush.console()
                                } # terminou o while
                                } # fim IF first==1

# Mulheres propõem primeiro:

if (first==2)
{
  step.2 <- preference.col
  step.2[step.2>1] <- 0
  s <- apply(step.2,1,sum)
  test_s <- s>1
  while (any(test_s==TRUE)==TRUE)
  {
    if (any(test_s==TRUE)==TRUE) {
      position1 <- which(s>1)
      position2 <- vector('list')
      position3 <- vector('list')
      position4 <- NULL
      position5 <- 1:m
      aux <- sort(rep(1:m,n))
      for (k in
1:length(position1)){position2[[k]] <- which(step.2[position1[k],]==1)

      position3[[k]] <- which.min(preference.row[position1[k],position2[[k]])}

      position3[[k]] <- position2[[k]]%w/o%position2[[k]][position3[[k]]]

      step.2[position1[k],position3[[k]]] <- 0}

      position4 <- which(step.2==1)
      position4 <- aux[position4]
      position5 <- position5[-

position4]

      store <- append(position5,

      r <- rle(sort(store))
      for (j in
position5){step.2[,j][preference.col[,j]==r$lengths[r$values==j]+1] <- 1}
      s <- apply(step.2,1,sum)
      test_s <- s>1
      }else{return(step.2)}

      loop <- loop + 1
      cat("interações =",loop,'\n')
      flush.console()
      } # terminou o 2º while
    } # fim IF first==2
  if (first==1) {return(step.1)}
  if (first==2) {return(step.2)}} # FIM da FUNÇÃO!

```



```
#####
#
# EXEMPLOS:
#
#####

##### A,B,C
# Exemplo 1 de Gale&Shapley (1962,p.11) College Admissions And the Stability
# of Marriage
# Valores impostos:

m1 <- c(1,2,3)
m2 <- c(3,1,2)
m3 <- c(2,3,1)

n1 <- c(3,1,2)
n2 <- c(2,3,1)
n3 <- c(1,2,3)

preference.row <- matrix(c(m1, m2, m3), ncol=3, byrow=TRUE)
preference.col <- matrix(c(n1, n2, n3), ncol=3)

gsa(m = 3, n = 3, preference.row = preference.row, preference.col =
preference.col, first=1)
gsa(m = 3, n = 3, preference.row = preference.row, preference.col =
preference.col, first=2)

##### A,B,C,D
# Exemplo 2 de Gale Shapley (1962, p.12):

m1 <- c(1,2,3,4)
m2 <- c(1,4,3,2)
m3 <- c(2,1,3,4)
m4 <- c(4,2,3,1)

n1 <- c(3,4,2,1)
n2 <- c(3,1,4,2)
n3 <- c(2,3,4,1)
n4 <- c(3,2,1,4)

preference.row <- matrix(c(m1, m2, m3, m4), ncol=4, byrow=TRUE)
preference.col <- matrix(c(n1, n2, n3, n4), ncol=4)

gsa(m = 4, n = 4, preference.row = preference.row, preference.col =
preference.col, first=1)
gsa(m = 4, n = 4, preference.row = preference.row, preference.col =
preference.col, first=2)

##### A,B,C,D
# Exemplo 3 de Gale Shapley (1962, p.12):

m1 <- c(1,2,3,4)
m2 <- c(1,2,3,4)
m3 <- c(3,1,2,4)
m4 <- c(2,3,1,4)

n1 <- c(3,4,1,2)
```

```

n2 <- c(2,3,4,1)
n3 <- c(1,2,3,4)
n4 <- c(3,4,2,1)

preference.row <- matrix(c(m1, m2, m3, m4), ncol=4, byrow=TRUE)
preference.col <- matrix(c(n1, n2, n3, n4), ncol=4)

gsa(m = 4, n = 4, preference.row = preference.row, preference.col =
preference.col, first=1)
gsa(m = 4, n = 4, preference.row = preference.row, preference.col =
preference.col, first=2)

# FIM DOS EXEMPLOS!

#####

```

Outros exemplos foram também testados. A função acima foi expandida também para o caso *many-to-one*, que foi chamada de `gsa.many`:

- `gsa.many(m,n, preference.row, preference.col, first, expand)`

Para esse caso, o funcionamento da função continua o mesmo, uma mudança importante é o vetor indicador do número de vagas que cada escola possui, chamado de `expand`. Esse vetor determina o número de vagas. Como a função inteira possui um número de linhas maior do que a função geral `gsa`, ela não será transcrita nesse anexo podendo porém ser solicitada ao autor da tese.¹²³

A solução adotada para o algoritmo *many-to-one* foi tratar cada vaga de escola como uma escola individualizada, porém, determinando-se as preferências dos alunos de maneira que qualquer vaga de uma escola s preferida à s' seja preferida a qualquer vaga de s' . Para as escolas, cada vaga possuía exatamente a mesma preferência das escolas. Com essa regra para as preferências dos alunos e das escolas se obtêm o pareamento tal como no algoritmo acima e depois a função agrega novamente as escolas sendo que cada coluna volta a ser novamente uma única escola.

Por fim, um último detalhe de construção: para lidar com vagas ociosas, a maneira empregada foi criar alunos-vagas, esses alunos são jogadores *dummy*, a regra utilizada para operacionalizá-los foi

¹²³ E-mails de correspondência: ymaia@cedeplar.ufmg.br, ou o vic_senn@yahoo.com.br.

de que os alunos-vaga preferem a escola onde existe a vaga ociosa, porém, em nossos modelos, a escola prefere sempre ficar com vagas ociosas, ou seja, os alunos vaga são os últimos do *ranking* de qualquer escola. Em primeiro lugar os alunos vaga preferem a vaga da escola que eles estão representando, para as demais escolas pode ser aleatória ou alguma outra regra discricionária. No caso das simulações utilizada a regra empregada foi de que depois da sua escola representativa, os alunos vagas preferiam a escola com pior desempenho da média (isso parte do pressuposto que escolas com baixo desempenho, pouco preferidas, possuem mais chance de ficarem com vagas ociosas). Essa escolha não altera demais o resultado do algoritmo.

O algoritmo de *Top Trading Cycles* (TTC) abordado na seção 3.3 tem uma solução gráfica que precisa ser transportada para uma solução computacional. Em verdade, a versão mais simples desse algoritmo é mais fácil de ser implementada, pelo algoritmo cada indivíduo de um grupo aponta para o indivíduo de outro grupo.

Tomando-se o segundo exemplo dado na seção 3.5, temos que o aluno i_1 aponta para a escola s_1 , a escola s_1 por sua vez aponta o aluno i_3 que aponta a escola s_2 , que aponta o aluno i_2 , que termina apontando s_1 novamente, fechando o ciclo. Sempre que o que um elemento se repete nessa corrente de invocação, se fecha um ciclo e o algoritmo para. Os participantes do ciclo deixam o algoritmo que continua com os jogadores remanescentes. Para o ciclo acima temos:

$$i_1 \rightarrow s_1 \rightarrow i_3 \rightarrow s_2 \rightarrow i_2 \rightarrow s_1$$

Ou: $\{i_1, s_1, i_3, s_2, i_2, s_1\}$

Com um input da ordem para cada jogador, é fácil obter computacionalmente uma relação do tipo acima. Como se sabe pelo teorema **T.10** que sempre haverá pelo menos um ciclo, a proposição do algoritmo implementado foi de parar assim que surgir o primeiro jogador repetido. Todos os participantes que estão imediatamente após o jogador que foi repetido (no caso do exemplo, a escola s_1) são retirados do passo seguinte e não podem ser mais escolhidos.

Os participantes à direita do jogador que foi repetido na chamada, inclusive sua repetição são retirados. No caso do exemplo, todos à direita de s_1 : i_3, s_2, i_2, s_1 (este último é a cópia).

Partindo deste princípio, o algoritmo implementado retira um ciclo por vez. Essa solução é ineficiente dado que se sabe que mais de um ciclo pode ocorrer ao mesmo tempo tal como nos exemplos vistos no capítulo 3. Um algoritmo que investiga a ocorrência de mais de um ciclo ao mesmo tempo, ao invés de retirar um por vez, trabalharia mais rápido. No entanto, o TTC, por conta de sua regra de funcionamento, é muito mais rápido do que os algoritmos de Gale-Shapley implementados. A eficiência computacional não chegou a se constituir um grande problema para esse algoritmo.

A função implementada tem argumentos similares a aqueles trabalhados na função GSA para o *deferred-acceptance*:

- `ttc(m, n, preference.row, preference.col)`

São os mesmos argumentos da função GSA, retirando-se o argumento `first`, dado que esse algoritmo foi implementado apenas os algoritmos favoráveis aos alunos (*homens*).

Para tratar com escolas com mais de uma vaga a solução foi similar à aquela praticada para os algoritmos DAs: cada vaga foi tratada como um agente a mais. Como um aluno deve ser indiferente a uma vaga da mesma escola, foi adotada uma regra de acrescentar uma pequena fração crescente (fração $1/(\text{numero de vagas})$) para cada vaga de uma mesma escola. Qualquer vaga de uma escola s que é preferida à qualquer vaga da escola s' , é preferida à qualquer vaga dessa escola.

```
#####
# SCRIPT NO R para o de TTC one-to-one de Abdulkadiroglu & Sonmez (2003)
# Autor: Victor Maia
# última versão: 30/04/2013
#####

ttc <- function(m, n, preference.row, preference.col)
{
  # m é Número de Linhas
  # n é Número de Colunas

  loop <- 1
  R <- NULL
  X <- vector('list')
  M <- matrix(0,nrow=m,ncol=n)

  while(any(preference.row<m+1)==TRUE) {
    i <- 1
    h <- 1
    s <- NULL
    r <- NULL
    p <- NULL
    w <- which(preference.row[h,]==min(preference.row[h,]))
    h <- which(preference.col[,w]==min(preference.col[,w]))
    s <- append(s,c(w,h))
    r <- append(r,w)
    p <- append(p,h)
    w <- which(preference.row[h,]==min(preference.row[h,]))
    h <- which(preference.col[,w]==min(preference.col[,w]))
    s <- append(s,c(w,h))
    r <- append(r,w)
    p <- append(p,h)
    i <- i+1
    while(any((r[i]==r[-i])==TRUE)==FALSE) {
      w <- which(preference.row[h,]==min(preference.row[h,]))
      h <- which(preference.col[,w]==min(preference.col[,w]))
      s <- append(s,c(w,h))
      r <- append(r,w)
      p <- append(p,h)
      i <- i+1} #fim do Pequeno While!

    # Tem que ficar entre os repetidos:
    repetido <- r[duplicated(r)]
    between <- which(r==repetido)
    r <- r[between[1]:between[length(between)]]

    X[[loop]] <- matrix(c(p[between[1]:(between[2]-
1)],r[2:length(r)]),ncol=2,byrow=FALSE)
    preference.row[X[[loop]][,2]] <- m+1
    preference.col[X[[loop]][,1],] <- m+1
    cat("interações =",loop,'\n')
    flush.console()
    loop <- loop+1          } # Fim do While do LOOP

  for(k in 1:length(X)){
    R <- rbind(R,X[[k]])} # fim do for
  R <- R[order(R[,1]),]
```

```

for(k in 1:length(R[,1])){
M[R[k,1],R[k,2]] <- 1} #fim do for
return(M)
} # FIM DA FUNÇÃO!

#####
#
# EXEMPLOS:
#
#####

# Segundo Exemplo de TTC do capítulo 3 (seção 3.5)

# Alunos:

m1 <- c(1,2,3,4)
m2 <- c(1,2,3,4)
m3 <- c(4,1,3,2)
m4 <- c(4,3,1,2)

# Escolas:

n1 <- c(3,2,1,4)
n2 <- c(2,1,3,4)
n3 <- c(1,3,4,2)
n4 <- c(2,3,4,1)

preference.row <- matrix(c(m1, m2, m3, m4), ncol=4, byrow=TRUE)
preference.col <- matrix(c(n1, n2, n3, n4), ncol=4)

ttc(m=4,n=4,preference.row=preference.row,preference.col=preference.col)

## OUTRO EXEMPLO (segundo exemplo de 3.5):

# Alunos:

m1 <- c(1,2,3,4)
m2 <- c(2,1,3,4)
m3 <- c(1,3,2,4)
m4 <- c(4,3,1,2)

# Escolas:

n1 <- c(1,2,3,4)
n2 <- c(2,1,3,4)
n3 <- c(1,3,4,2)
n4 <- c(2,3,1,4)

preference.row <- matrix(c(m1, m2, m3, m4), ncol=4, byrow=TRUE)
preference.col <- matrix(c(n1, n2, n3, n4), ncol=4)

ttc(m=4,n=4, preference.row = preference.row, preference.col = preference.col)

# FIM DOS EXEMPLOS!

```

```
#####
# SCRIPT NO R para o de TTC MANY-TO-ONE
# Autor: Victor Maia
# última versão: 30/04/2013
#####

ttc.many <- function(m, n, preference.row, preference.col, expand)
{
  # m é Número de Linhas
  # n é Número de Colunas

  #####
  #
  ##### TRECHO DO EXPAND
  t <- max(m,n)
  m <- t; n <- t

  preference.row.2 <- vector('list')
  for(i in 1:length(expand)){
    preference.row.2[[i]] <- rep(preference.row[,i], expand[i])
  }
  preference.row.2 <- unlist(preference.row.2)
  preference.row.2 <- matrix(preference.row.2, ncol=sum(expand))

  expand.aux <- NULL
  for(i in 1:length(expand)){
    if(i==1){expand.aux[i] <- i+expand[i]}
    else{expand.aux[i] <- expand.aux[i-1]+expand[i]}
  }
  expand.aux <- expand.aux - 1

  matriz <- matrix(0, nrow=length(preference.row[,1]), ncol=sum(expand))

  for(j in 1:length(matriz[,1])){
    for(i in 1:length(expand)){
      if(i==1){matriz[j,1:expand.aux[i]] <- seq(0,.99,length.out
= expand[i])}
      else{matriz[j, (expand.aux[i-1]+1):expand.aux[i]]<-
seq(0,.99,length.out = expand[i])}
    }
  }

  preference.row.2 <- preference.row.2 + matriz

  # Corrigindo para as novas posições:

  for(j in 1:length(preference.row.2[,1])){preference.row.2[j,] <- rank(preference.row.2[j,])}

  # Replicando preferências das colunas:

  preference.col.2 <- vector('list')
  for(i in 1:ncol(preference.col)){
    preference.col.2[[i]]<-rep(preference.col[,i], expand[i])
  }
  preference.col.2 <- unlist(preference.col.2)
  preference.col.2 <- matrix(preference.col.2, ncol=sum(expand))

  ##### Fim da parte Expand!!!

  loop <- 1
  R <- NULL
  X <- vector('list')
  M <- matrix(0, nrow=m, ncol=n)

  while(any(preference.row.2<m+1)==TRUE) {
    i <- 1
    h <- 1
    s <- NULL
    r <- NULL
    p <- NULL
    w <- which(preference.row.2[h,]==min(preference.row.2[h,]))
    h <- which(preference.col.2[,w]==min(preference.col.2[,w]))
    s <- append(s, c(w,h))
    r <- append(r, w)
  }
}
```

```

p <- append(p,h)
w <- which(preference.row.2[h,]==min(preference.row.2[h,]))
h <- which(preference.col.2[,w]==min(preference.col.2[,w]))
s <- append(s,c(w,h))
r <- append(r,w)
p <- append(p,h)
i <- i+1
while(all(duplicated(r)==FALSE)==TRUE) {
w <- which(preference.row.2[h,]==min(preference.row.2[h,]))
h <- which(preference.col.2[,w]==min(preference.col.2[,w]))
s <- append(s,c(w,h))
r <- append(r,w)
p <- append(p,h)
i <- i+1} #fim do Pequeno While!

# Tem que ficar entre os repetidos:
repetido <- r[duplicated(r)]
between <- which(r==repetido)
r <- r[between[1]:between[length(between)]]

X[[loop]] <- matrix(c(p[between[1]:(between[2]-1)],r[2:length(r)]),ncol=2,byrow=FALSE)
preference.row.2[X[[loop]],2] <- m+1
preference.col.2[X[[loop]],1] <- m+1
cat("interações =",loop,'\n')
flush.console()
loop <- loop+1      } # Fim do While do LOOP

for(k in 1:length(X)){
R <- rbind(R,X[[k]])} # fim do for
R <- R[order(R[,1]),]
for(k in 1:length(R[,1])){
M[R[k,1],R[k,2]] <- 1} #fim do for

##### AGRUPANDO!!!!!!

      expand.aux <- c(1,expand.aux)
      step.1.1 <- vector('list')
      for(k in 1:length(expand)){
M[,expand.aux[k]:expand.aux[k+1]]}
M[, (expand.aux[k]+1):expand.aux[k+1]]}
      if(k==1){step.1.1[[k]]<-
      else{step.1.1[[k]]<-
      }
      step.1.2 <- vector('list')
      for(k in 1:length(step.1.1)){
if(is.array(step.1.1[[k]])==FALSE){step.1.2[[k]] <- step.1.1[[k]]}
      else{step.1.2[[k]]<-rowSums(step.1.1[[k]])}
      }
      step.1.2 <- matrix(unlist(step.1.2),nrow=m, ncol=length(expand))
return(step.1.2)
} # FIM DA FUNÇÃO!

```



```
#####
#
# EXEMPLOS:
#
#####

# Exemplo Abdulkadiroglu e Sonmez(2003) adaptado p/ não ter vagas excedentes:

m1 <- c(2,1,3,4)
m2 <- c(1,2,3,4)
m3 <- c(3,2,1,4)
m4 <- c(3,4,1,2)
m5 <- c(1,4,2,3)
m6 <- c(2,3,4,1)
m7 <- c(1,2,3,4)
m8 <- c(1,2,4,3)

n1 <- c(1,2,3,4,5,6,7,8)
n2 <- c(7,6,1,3,2,8,5,4)
n3 <- c(3,5,2,8,1,7,4,6)
n4 <- c(8,5,6,4,7,1,3,2)

preference.row <- matrix(c(m1,m2,m3,m4,m5,m6,m7,m8), nrow=8, byrow=TRUE)
preference.col <- matrix(c(n1, n2, n3, n4), ncol=4)
exp <- c(2,2,2,2) # Não serão necessárias vagas reservas.

ttc.many(m=8,n=4,preference.row=preference.row, preference.col=preference.col,
expand=exp)

# OBS esse algoritmo de TTC não funciona com vagas excedentes.

# FIM DOS EXEMPLOS!
```

Uma simulação feita com preferências aleatórias e número crescente de participantes em cada grupo foi realizada para verificar quantos passos leva o algoritmo a medida que o tamanho da amostra cresce (n). Para cada valor de n foram realizadas 30 simulações. O valor médio do número de passos foi computado e dividido pelo valor máximo de passos que algoritmo pode levar dado por $N = n^2 - 2n + 2$.

Para preferências dadas aleatoriamente, o número de passos do algoritmo cresce rapidamente mas a uma taxa muito menor do que o valor de N . A TAB. B.1 apresenta os valores médios de n ($\bar{n}$) e os valores máximos de passos que o algoritmo pode alcançar. O θ representa a fração $\bar{n}/N$ e a tabela oferece também o intervalo de confiança. O GRAF. B.1 apresenta ilustra os resultados de TAB B.2.

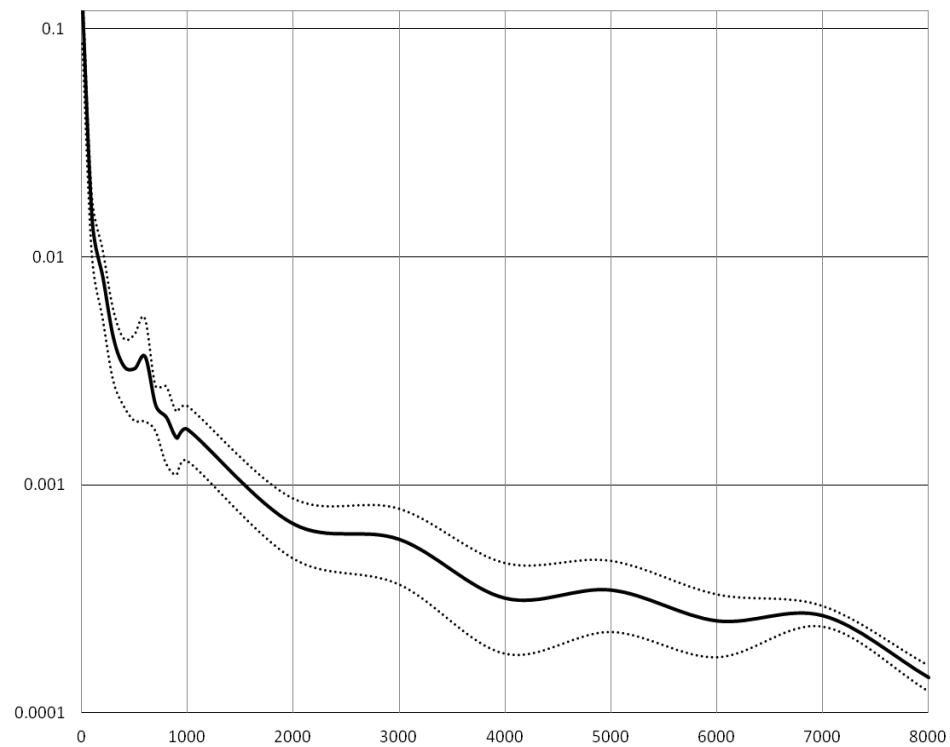
Tabela B.2 – Simulações do número de passos do algoritmo DA, preferências aleatórias

n	N	θ	$IC_{inferior}$	$IC_{superior}$
10	82	0,1203	0,0863	0,1544
100	9.802	0,0143	0,0101	0,0185
200	39.602	0,0081	0,0054	0,0108
300	89.402	0,0044	0,0029	0,0059
400	159.202	0,0033	0,0022	0,0044
500	249.002	0,0032	0,0019	0,0045
600	358.802	0,0036	0,0019	0,0054
700	488.602	0,0022	0,0017	0,0027
800	638.402	0,0020	0,0012	0,0027
900	808.202	0,0016	0,0011	0,0021
1.000	998.002	0,0017	0,0013	0,0022
2.000	3.996.002	0,0007	0,0005	0,0009
3.000	8.994.002	0,0006	0,0004	0,0008
4.000	15.992.002	0,0003	0,0002	0,0005
5.000	24.990.002	0,0003	0,0002	0,0005
6.000	35.988.002	0,0003	0,0002	0,0003
7.000	48.986.002	0,0003	0,0002	0,0003
8.000	63.984.002	0,0001	0,0001	0,0002

Fonte: Dados de Simulações.

OBS: Cada valor de n teve 30 simulações, com exceção dos valores iguais ou maiores à 6000 que possuem apenas 10 simulações.

Gráfico B.1 – Simulações do valor de θ para o algoritmo de Gale-Shapley



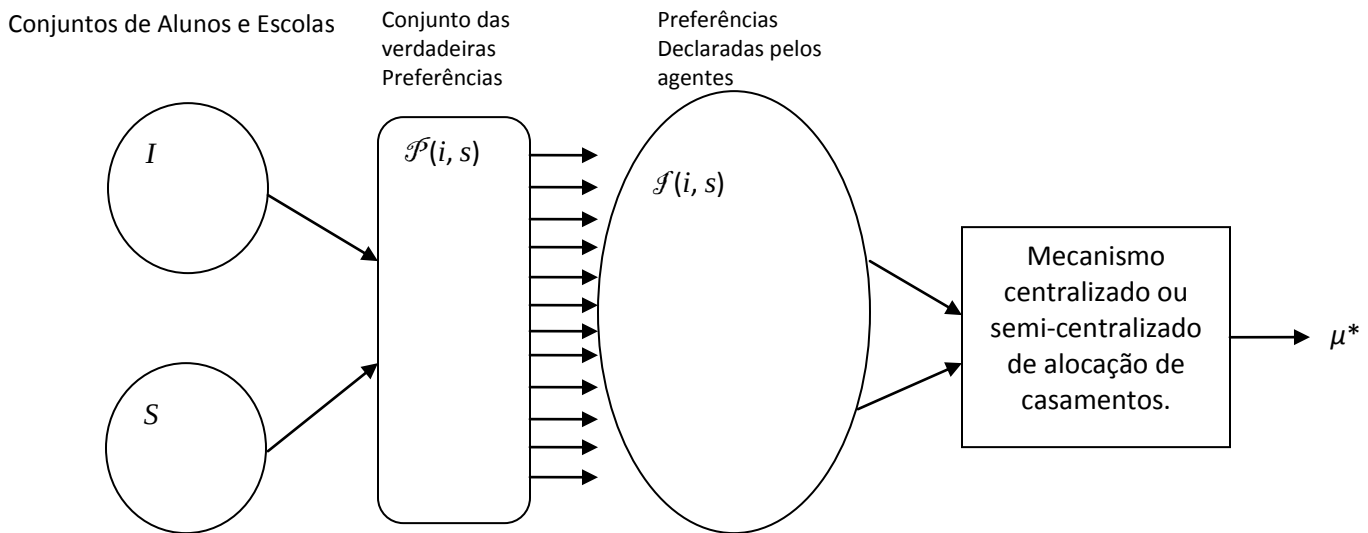
Fonte: Elaboração própria a partir da simulação e dos dados aleatórios para *matching* com matrizes de várias dimensões.

OBS: Cada valor de n teve 30 simulações, com exceção dos valores iguais ou maiores à 6000 que possuem apenas 10 simulações.

12. ANEXO C

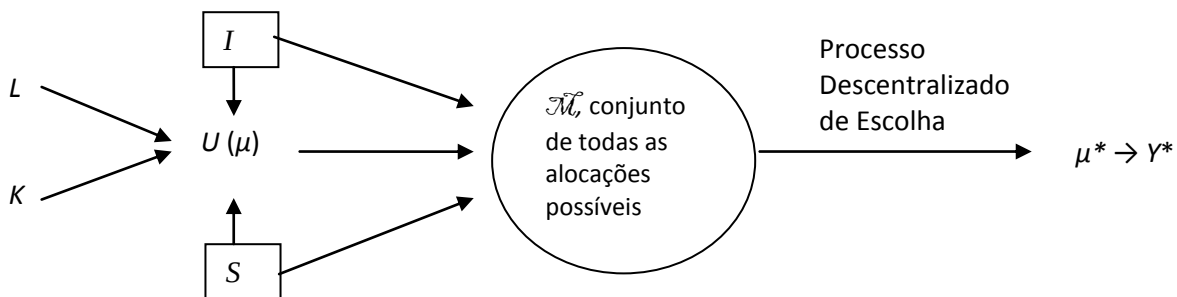
O arcabouço do *two-sided markets* e do *assortative mating* podem ser representados por figuras diagramáticas. No diagrama da teoria das preferências têm-se os dois conjuntos disjuntos de agentes, esses conjuntos possuem preferências dadas, por meio delas decidem como declarar as preferências no sistema. De posse das informações declaradas das preferências, um mecanismo de alocação geralmente centralizado, ou semi-centralizado (isto é, coordenado centralizadamente, mas com alguma liberdade para os agentes tomarem suas ações), fornece o *matching* final para os participantes (μ^*). No modelo da teoria utilitarista os agentes possuem uma função de utilidade para julgar o quanto preferem uma alocação μ qualquer, para isso levam em conta características observadas (L e K) e em um processo descentralizado escolhem o *matching* ótimo (μ^*).

Figura C.1 – Elementos participantes do processo decisório de matching - teoria das preferências



Fonte: elaboração do autor.

Figura C.2 – Elementos participantes do processo decisório de matching - teoria utilitarista

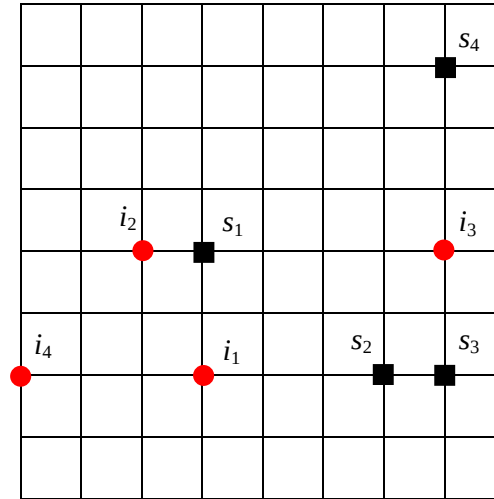


Fonte: elaboração do autor.

13. ANEXO D

O modelo espacial apresentado na matriz de distâncias da seção 3.6 pode ser obtido a partir do seguinte diagrama (FIG. D.1) plano com 4 alunos e 4 escolas:

Figura D.1 – Diagrama espacial para a distribuição de alunos e escolas do exemplo 10.



Fonte: Elaboração do autor.

Cada pequeno quadrado da grade do diagrama acima possui 1 Km de comprimento, e os alunos podem se deslocar para qualquer direção do plano (não precisam se deslocar apenas sobre a grade). Os alunos estão representados pelos círculos vermelhos e as escolas são os pequenos quadrados pretos. Nota-se que o aluno i_1 possui a escola s_1 como a mais próxima, mas ele não é o aluno mais próximo s_1 , e sim s_2 . O aluno i_4 é o mais distante de todas as escolas.

No exemplo 10 as distâncias são únicas, não há nenhum aluno que esteja equidistante de duas escolas diferentes e para cada uma das escolas não há dois ou mais alunos equidistantes. Essa matriz de distâncias determinou o ordenamento de preferências tanto de alunos quanto de escolas do exemplo.

A proposição **P.1** apresenta que o algoritmo de *matching* do *deferred acceptance* na situação descrita pelo exemplo 10 será única. No entanto, explorando todas as 24 possibilidades de *matching* pode se perceber que o resultado do algoritmo não é aquele que retorna as menores

distâncias possíveis. A TAB. D.1 apresenta todas as possibilidades, o conjunto de todas as alocações possíveis ($\mathcal{M}$) e as distâncias médias obtidas em cada uma delas.

Tabela D.1 – Todas as alocações possíveis do exemplo 10 e as distâncias médias em cada alocação

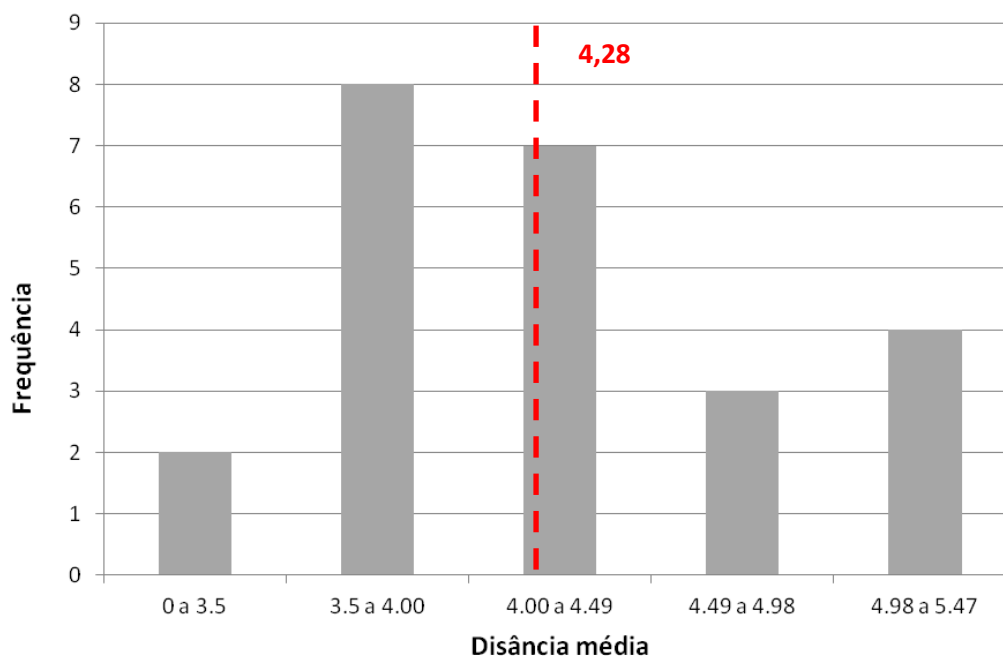
alocação	Escola 1	Escola 2	Escola 3	Escola 4	Distância média
μ_1	i_1s_1	i_2s_2	i_3s_3	i_4s_4	4,27
μ_2	i_1s_1	i_2s_2	i_4s_3	i_3s_4	4,12
μ_3	i_1s_1	i_3s_2	i_2s_3	i_4s_4	4,55
μ_4	i_1s_1	i_3s_2	i_4s_3	i_2s_4	4,25
μ_5	i_1s_1	i_4s_2	i_2s_3	i_3s_4	4,10
μ_6	i_1s_1	i_4s_2	i_3s_3	i_2s_4	3,95
μ_7^*	i_2s_1	i_1s_2	i_3s_3	i_4s_4	3,65*
μ_8	i_2s_1	i_1s_2	i_4s_3	i_3s_4	<u>3,50</u>
μ_9	i_2s_1	i_3s_2	i_1s_3	i_4s_4	3,95
μ_{10}	i_2s_1	i_3s_2	i_4s_3	i_1s_4	4,15
μ_{11}	i_2s_1	i_4s_2	i_3s_3	i_1s_4	3,85
μ_{12}	i_2s_1	i_4s_2	i_1s_3	i_3s_4	<u>3,50</u>
μ_{13}	i_3s_1	i_1s_2	i_2s_3	i_4s_4	5,25
μ_{14}	i_3s_1	i_1s_2	i_4s_3	i_2s_4	4,95
μ_{15}	i_3s_1	i_2s_2	i_1s_3	i_4s_4	5,27
μ_{16}	i_3s_1	i_2s_2	i_4s_3	i_1s_4	5,47
μ_{17}	i_3s_1	i_4s_2	i_1s_3	i_2s_4	4,95
μ_{18}	i_3s_1	i_4s_2	i_2s_3	i_1s_4	5,45
μ_{19}	i_4s_1	i_1s_2	i_2s_3	i_3s_4	3,75
μ_{20}	i_4s_1	i_1s_2	i_3s_3	i_2s_4	3,60
μ_{21}	i_4s_1	i_2s_2	i_1s_3	i_3s_4	3,77
μ_{22}	i_4s_1	i_2s_2	i_3s_3	i_1s_4	4,12
μ_{23}	i_4s_1	i_3s_2	i_1s_3	i_2s_4	3,90
μ_{24}	i_4s_1	i_3s_2	i_2s_3	i_1s_4	4,40

Fonte: Elaboração própria a partir da simulação e dos dados de distâncias do Exemplo. 10.

Obs: Com asterisco está a alocação do algoritmo DA. Sublinhadas as alocações que oferecem as menores distâncias, e em negrito a que oferece a maior distância média.

Duas alocações fornecem a menor distância possível μ_8 e μ_{12} , para a qual a distância média é de 3,5 Km. O *matching* estável dado pelo o algoritmo *deferred acceptance* (ou no TTC) não fornece a menor distância média possível, mas oferece um resultado relativamente próximo do mínimo, 3,65 Km. A alocação que oferece maior distância é μ_{16} , com 5,47 Km. O histograma da FIG. D.2 mostra a distribuição das distâncias e a média das distâncias médias que é 5,47.

Figura D.2 – Histograma das distâncias médias.



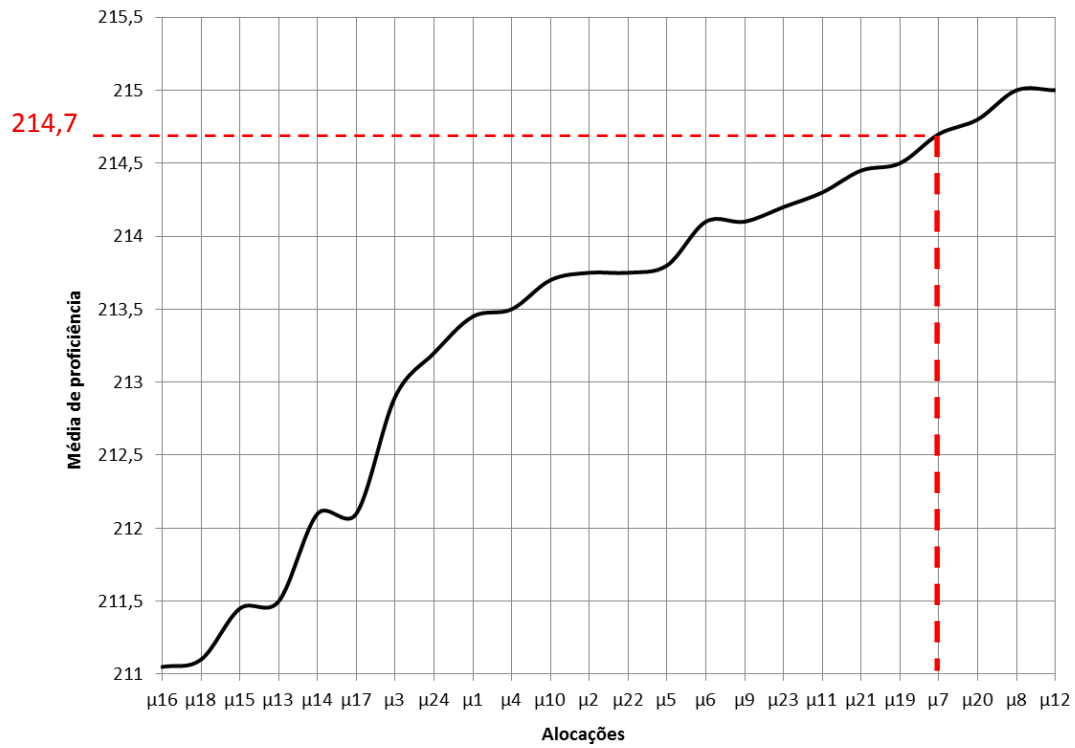
Fonte: Elaboração própria a partir da simulação e dos dados de distâncias do Exemplo. 10.

Pode se fazer uma rápida estimativa do impacto na proficiência para o exemplo. Partindo-se de uma média de 225 pontos caso a distância fosse zero para todos os alunos e supondo-se que cada quilômetro a mais significa uma redução de 2 pontos de proficiência para o aluno médio, tem se que na alocação que retorna as menores distâncias o máximo que pode ser alcançado é de 215 pontos. No caso do matching do algoritmo (μ_7) esse valor é 214,7 e na pior alocação é 211,05.

Entre a pior e a melhor alocação há uma diferença de 4 pontos para o aluno médio. Considerando que pela média de evolução do PROEB (2006-2009) entre o 5º ano e o 9º cada criança ganha em média 12 pontos por ano letivo, essa diferença de 4 pontos significa que apenas pela alocação entre escolas e alunos, os estudantes podem acelerar sua evolução de aprendizado em 1/3 de um ano letivo. Essa conta é feita com maiores detalhes e valores reais no capítulo 7.

A FIG. D.3 apresenta os resultados dessa simulação para o exemplo 10, e destaca também a posição do *matching* obtido com o DA em relação ao máximo.

Figura D.3 – Proficiência simulada para o exemplo 10



Fonte: Elaboração própria a partir da simulação e dos dados de distâncias do Exemplo. 10.

14. ANEXO E

Abaixo segue o termo de confidencialidade assinado pelo autor da tese em Junho de 2012 para uso e sigilo dos dados com nome e endereço dos alunos do banco do Censo Escolar e do PROALFA.

Figura E.1 – Termo de Confidencialidade assinado.

TERMO DE CONFIDENCIALIDADE

Título do projeto: Tese em Demografia: Processos de alocação de matrículas nas escolas públicas de Belo Horizonte, MG

Pesquisador responsável: Victor Maia Senna Delgado

Demais pesquisadores:

Instituição de origem do pesquisador: CEDEPLAR/UFMG

Área de Conhecimento: Demografia da Educação

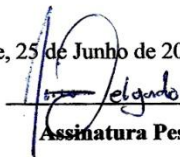
Telefone para contato: (31) 3448-9450; (31) 9222-2531

O(s) pesquisador(es) do projeto acima identificado(s) assume(m) o compromisso de:

- I. Preservar o sigilo e a privacidade dos dados e/ou informações que se terá acesso (Censo Educacional Escolar de 2010 e PROALFA);
- II. Assegurar que as informações e/ou dados serão utilizados, única e exclusivamente, para a execução do projeto em questão;
- III. Manter sigilo quanto a tais dados e informações, utilizando-os exclusivamente para o propósito das atividades relacionadas ao Projeto citado acima e não os distribuindo, revelando ou divulgando a terceiros, em hipótese alguma, sem o prévio e expresso consentimento da Secretaria de Estado de Educação de MG exceto se por exigência de lei ou de quaisquer órgãos governamentais.
- IV. Assegurar que os resultados da pesquisa somente serão divulgados de forma anônima, não sendo usadas iniciais ou quaisquer outras indicações que possam identificar o sujeito da pesquisa.
- V. Citar de maneira correta a fonte dos dados utilizados.
- VI. Compartilhar com a Secretaria de Estado de Educação de MG o resultado final do projeto.

O(s) Pesquisador(es) declara(m) ter conhecimento de que as informações pertinentes às técnicas do projeto de pesquisa somente podem ser acessados por aqueles que assinaram o Termo de Confidencialidade, excetuando-se os casos em que a quebra de confidencialidade é inerente à atividade ou que a informação e/ou documentação já for de domínio público.

Belo Horizonte, 25 de Junho de 2012 .



Assinatura Pesquisador

Nome: Victor Maia Senna Delgado
RG: 1.977.799

15. ANEXO F

Neste apêndice mostramos o significado das siglas sobre o tipo de variáveis utilizadas nas tabelas 2 e 3 e as transformações realizadas na variável *estudo_mãe* do PROEB:

Alfab.: Variável que possui apenas caracteres do “Alfabeto” greco-romano. Alguns softwares denominam esse tipo de variável como “string” que pode ser uma linha numérica ou de caracteres.

Alfan.: Variável “Alfanumérica”, Possui caracteres de texto e de números. Sua ocorrência no banco se dá em casos de descrição de complemento de endereço.

Categ.: Variável “Categórica”, variável que possui quantidade enumerável de categorias. No caso de raça, as categorias são: “AMARELO”, “BRANCO”, “PARDO” e “PRETO”, padrão do IBGE. O caso dos bairros está descrito na seção 4 sobre banco de dados. As demais variáveis categóricas possuem interpretação imediata.

Data: Variável de “Data” no formato dd/mm/aaaa.

Int. : Variável de número “Inteiro”. Variáveis numéricas pertencentes ao conjunto dos números naturais.

Num.: Variável “Numérica”, ao contrário de “Int.” essa denominação serve para números que não são inteiros. Em geral, números que pertencem ao conjunto dos números reais, ou para casos de inteiros de muitos dígitos, tal como ocorre com o código do aluno que possui 12 dígitos e foi guardado na memória como tipo “Num.”.

A variável escolaridade da mãe possuía reportada pelos anos no PROEB possui as 6 categorias abaixo definidas:

1. primário incompleto
2. primário completo.
3. 5ª a 8ª incompleto.
4. 5ª a 8ª completo.
5. 2º grau incompleto.
6. 2º grau completo.

Nos modelos de regressão, cada uma das categorias, com exceção da primeira, foi transformada em uma variável *dummy* Binária (0, 1), e utilizada no modelo. Que compara as categorias de escolaridade em relação a mãe não ter o primário completo.

Em alguns casos, se uma das *dummys* não foi significativa ela foi retirada do modelo.

Para as demais variáveis empregadas ver a tabela 2. E as transformações que estão relatadas no corpo do texto nos capítulos 5 e 6.

16. ANEXO G

Mapa G.1 – Mapa Oficial da Prefeitura de Belo Horizonte com os bairros da capital

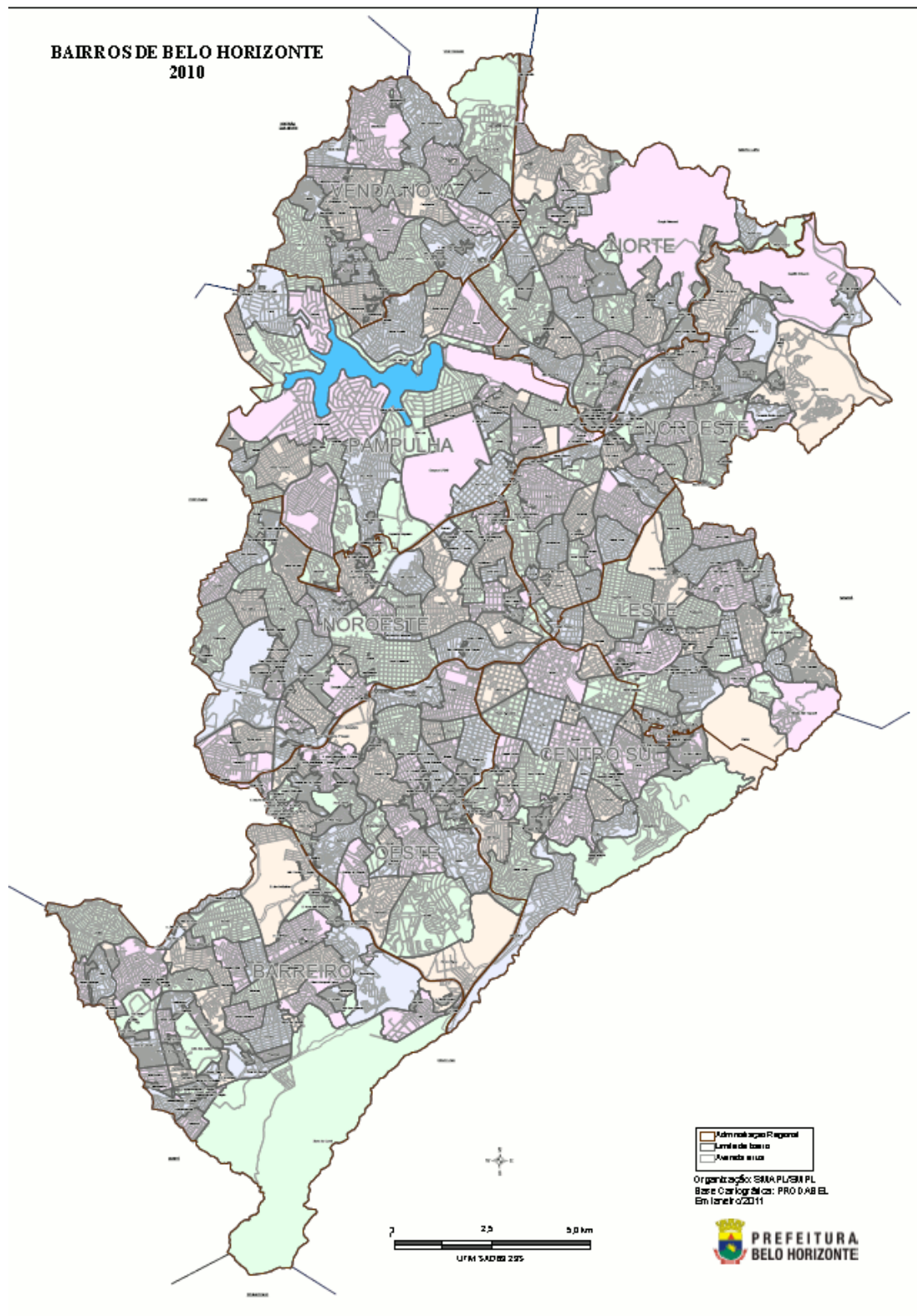


Tabela G.1 - Relação de grupos de bairros e bairros de Belo Horizonte:

Grupo	Bairro_Agrupado	N_bairro	Bairros	Número de Bairros
1	Mantiqueira	1	Lagoa	13
1	Mantiqueira	2	Lagoinha Leblon	13
1	Mantiqueira	3	Vila Santa Mônica	13
1	Mantiqueira	4	Piratininga	13
1	Mantiqueira	5	Piratininga A	13
1	Mantiqueira	6	Flamengo	13
1	Mantiqueira	7	Maria Helena	13
1	Mantiqueira	8	Mantiqueira	13
1	Mantiqueira	9	Nova América	13
1	Mantiqueira	10	Vila Mantiqueira	13
1	Mantiqueira	11	Jardim dos Comerciairos	13
1	Mantiqueira	12	Vila SESC	13
1	Mantiqueira	13	Letícia	13
2	Venda Nova	1	Europa	10
2	Venda Nova	2	Laranjeiras	10
2	Venda Nova	3	Candelária	10
2	Venda Nova	4	Conjunto Minascaixa	10
2	Venda Nova	5	Cenáculo	10
2	Venda Nova	6	Venda Nova	10
2	Venda Nova	7	Minascaixa	10
2	Venda Nova	8	Parque São Pedro	10
2	Venda Nova	9	Vila Satélite	10
2	Venda Nova	10	Satélite	10
3	Serra Verde	1	Serra Verde	5
3	Serra Verde	2	Conjunto Serra Verde	5
3	Serra Verde	3	Vila Serra Verde	5
3	Serra Verde	4	São Damião	5
3	Serra Verde	5	Canaã	5
4	Jaqueline	1	Jaqueline	11
4	Jaqueline	2	Vila Nova	11
4	Jaqueline	3	Juliana	11
4	Jaqueline	4	Xodó-Marize	11
4	Jaqueline	5	Mariquinhas	11
4	Jaqueline	6	Frei Leopoldo	11
4	Jaqueline	7	Etelvina Carneiro	11
4	Jaqueline	8	Madri	11
4	Jaqueline	9	Granja Werneck	11

4	Jaqueline	10	Zilah Spósito	11
4	Jaqueline	11	Solimões	11
5	Planalto	1	Planalto	6
5	Planalto	2	Vila Clóris	6
5	Planalto	3	Bacurau	6
5	Planalto	4	Campo Alegre	6
5	Planalto	5	Biquinhas	6
5	Planalto	6	Heliópolis	6
6	Santa Mônica	1	São João Batista	7
6	Santa Mônica	2	Vila São João Batista	7
6	Santa Mônica	3	Nossa Senhora Aparecida	7
6	Santa Mônica	4	Santa Mônica	7
6	Santa Mônica	5	Santa Branca	7
6	Santa Mônica	6	Itapoã	7
6	Santa Mônica	7	São Tomás	7
7	Céu Azul	1	Céu Azul	11
7	Céu Azul	2	Vila dos Anjos	11
7	Céu Azul	3	Jardim Leblon	11
7	Céu Azul	4	Apolônia	11
7	Céu Azul	5	Vila Jardim Leblon	11
7	Céu Azul	6	Rio Branco	11
7	Céu Azul	7	Várzea da Palma	11
7	Céu Azul	8	Universo	11
7	Céu Azul	9	Copacabana	11
7	Céu Azul	10	Unidas	11
7	Céu Azul	11	Vila Copacabana	11
8	Braúnas	1	Braúnas	9
8	Braúnas	2	Xangri-lá	9
8	Braúnas	3	Nova Pampulha	9
8	Braúnas	4	Bispo de Maura	9
8	Braúnas	5	Trevo	9
8	Braúnas	6	Conjunto São Francisco de Assis	9
8	Braúnas	7	Garças	9
8	Braúnas	8	Jardim Atlântico	9
8	Braúnas	9	Santa Amélia	9
9	Aeroporto	1	Aeroporto	19
9	Aeroporto	2	Vila Aeroporto	19
9	Aeroporto	3	São Bernardo	19
9	Aeroporto	4	Vila Aeroporto Jaraguá	19
9	Aeroporto	5	Vila Rica	19
9	Aeroporto	6	Liberdade	19
9	Aeroporto	7	Vila Santo Antônio	19
9	Aeroporto	8	Indaiá	19

9	Aeroporto	9	Santa Rosa	19
9	Aeroporto	10	Universitário	19
9	Aeroporto	11	Jaraguá	19
9	Aeroporto	12	Dona Clara	19
9	Aeroporto	13	Maria Virgínia	19
9	Aeroporto	14	Vila Universitários	19
9	Aeroporto	15	Palmares	19
9	Aeroporto	16	Suzana	19
9	Aeroporto	17	Vila Suzana II	19
9	Aeroporto	18	São Miguel	19
9	Aeroporto	19	Vila Suzana I	19
10	Floramar	1	Guarani	10
10	Floramar	2	Floramar	10
10	Floramar	3	Jardim Guanabara	10
10	Floramar	4	Conjunto Floramar	10
10	Floramar	5	Jardim Felicidade	10
10	Floramar	6	Tupi A	10
10	Floramar	7	Tupi B	10
10	Floramar	8	Mirante	10
10	Floramar	9	Lajedo	10
10	Floramar	10	Novo Tupi	10
11	Ouro Minas	1	Ouro Minas	7
11	Ouro Minas	2	Beira-Linha	7
11	Ouro Minas	3	Novo Aarão Reis	7
11	Ouro Minas	4	Antônio Ribeiro de Abreu	7
11	Ouro Minas	5	Cocorocó	7
11	Ouro Minas	6	Ribeiro de Abreu	7
11	Ouro Minas	7	Belmonte	7
12	Capitão Eduardo	1	Capitão Eduardo	16
12	Capitão Eduardo	2	Maria Teresa	16
12	Capitão Eduardo	3	Monte Azul	16
12	Capitão Eduardo	4	Conjunto Paulo VI	16
12	Capitão Eduardo	5	Paulo VI	16
12	Capitão Eduardo	6	Acaiaca	16
12	Capitão Eduardo	7	Vista do Sol	16
12	Capitão Eduardo	8	Grotinha	16
12	Capitão Eduardo	9	Jardim Vitória	16
12	Capitão Eduardo	10	Mirtes	16
12	Capitão Eduardo	11	Vila Maria	16
12	Capitão Eduardo	12	Bela Vitória	16
12	Capitão Eduardo	13	Pousada Santo Antônio	16
12	Capitão Eduardo	14	Vitória	16
12	Capitão Eduardo	15	Conjunto Capitão Eduardo	16

12	Capitão Eduardo	16	Beija Flor	16
13	São Gabriel	1	Minaslândia	17
13	São Gabriel	2	Primeiro de Maio	17
13	São Gabriel	3	Vila Primeiro de Maio	17
13	São Gabriel	4	Boa União II	17
13	São Gabriel	5	São Gabriel	17
13	São Gabriel	6	Aarão Reis	17
13	São Gabriel	7	São Gonçalo	17
13	São Gabriel	8	Providência	17
13	São Gabriel	9	Vila Minaslândia	17
13	São Gabriel	10	Conjunto Providência	17
13	São Gabriel	11	Boa União I	17
13	São Gabriel	12	Vila São Gabriel A	17
13	São Gabriel	13	Esplanada A	17
13	São Gabriel	14	Vila São Gabriel B	17
13	São Gabriel	15	Dom Silvério	17
13	São Gabriel	16	Três Marias	17
13	São Gabriel	17	Nazaré	17
14	Maria Goretti	1	Vila São Paulo Modelo	19
14	Maria Goretti	2	Vila Maria Virgínia	19
14	Maria Goretti	3	Vila São Paulo	19
14	Maria Goretti	4	São Paulo	19
14	Maria Goretti	5	Carioca	19
14	Maria Goretti	6	Dom Joaquim	19
14	Maria Goretti	7	Vila de Sá	19
14	Maria Goretti	8	Eymard	19
14	Maria Goretti	9	Pirajá	19
14	Maria Goretti	10	Fernão Dias	19
14	Maria Goretti	11	Penha	19
14	Maria Goretti	12	Vila Aarão Reis	19
14	Maria Goretti	13	Maria Goretti	19
14	Maria Goretti	14	São Marcos	19
14	Maria Goretti	15	Ipê	19
14	Maria Goretti	16	Goiânia	19
14	Maria Goretti	17	Guanabara	19
14	Maria Goretti	18	São Benedito	19
14	Maria Goretti	19	Morro dos Macacos	19
15	São Luíz	1	São Luíz	12
15	São Luíz	2	São José	12
15	São Luíz	3	Ouro Preto	12
15	São Luíz	4	Novo Ouro Preto	12
15	São Luíz	5	Vila Paquetá	12
15	São Luíz	6	Conjunto Lagoa	12

15	São Luíz	7	Santa Terezinha	12
15	São Luíz	8	Bandeirantes	12
15	São Luíz	9	Itatiaia	12
15	São Luíz	10	Urca	12
15	São Luíz	11	Confisco	12
15	São Luíz	12	Paquetá	12
16	Glória	1	Pindorama	9
16	Glória	2	Vila Coqueiral	9
16	Glória	3	Coqueiros	9
16	Glória	4	Novo Glória	9
16	Glória	5	Glória	9
16	Glória	6	São Salvador	9
16	Glória	7	Serrano	9
16	Glória	8	Vila Santo Antônio Barroquinha	9
16	Glória	9	Conjunto Celso Machado	9
17	Castelo	1	Alípio de Melo	3
17	Castelo	2	Manacás	3
17	Castelo	3	Castelo	3
18	Jardim São José	1	Vila Trinta e Um de Março	11
18	Jardim São José	2	Dom Bosco	11
18	Jardim São José	3	Inconfidência	11
18	Jardim São José	4	Jardim São José	11
18	Jardim São José	5	Vila Jardim São José	11
18	Jardim São José	6	Jardim Alvorada	11
18	Jardim São José	7	Vila Antena Montanhês	11
18	Jardim São José	8	Vila Jardim Montanhês	11
18	Jardim São José	9	Vila Alvorada	11
18	Jardim São José	10	Engenho Nogueira	11
18	Jardim São José	11	Vila Engenho Nogueira	11
19	São Francisco	1	Campus UFMG	14
19	São Francisco	2	Sumaré	14
19	São Francisco	3	Aparecida VII	14
19	São Francisco	4	Vila Sumaré	14
19	São Francisco	5	Ermelinda	14
19	São Francisco	6	Vila Nova Cachoeirinha II	14
19	São Francisco	7	Nova Cachoeirinha	14
19	São Francisco	8	Vila Nova Cachoeirinha I	14
19	São Francisco	9	Vila Maloca	14
19	São Francisco	10	São Francisco	14
19	São Francisco	11	Vila São Francisco	14
19	São Francisco	12	Vila Real I	14
19	São Francisco	13	Vila Real II	14
19	São Francisco	14	Vila Santa Rosa	14

20	Caiçaras	1	Caiçaras	5
20	Caiçaras	2	Monsenhor Messias	5
20	Caiçaras	3	Jardim Montanhês	5
20	Caiçaras	4	Caiçara-Adelaide	5
20	Caiçaras	5	Alto Caiçaras	5
21	Aparecida	1	Nova Esperança	5
21	Aparecida	2	Santo André	5
21	Aparecida	3	Prado Lopes	5
21	Aparecida	4	Bom Jesus	5
21	Aparecida	5	Aparecida	5
22	Lagoinha	1	Bonfim	16
22	Lagoinha	2	Lagoinha	16
22	Lagoinha	3	Senhor dos Passos	16
22	Lagoinha	4	São Cristóvão	16
22	Lagoinha	5	Canadá	16
22	Lagoinha	6	Cachoeirinha	16
22	Lagoinha	7	Vila Cachoeirinha III	16
22	Lagoinha	8	Vila Cachoeirinha IV	16
22	Lagoinha	9	Vila Inestan	16
22	Lagoinha	10	Santa Cruz	16
22	Lagoinha	11	Vila da Paz	16
22	Lagoinha	12	Renascença	16
22	Lagoinha	13	Concórdia	16
22	Lagoinha	14	Vila do Pombal	16
22	Lagoinha	15	Tiradentes	16
22	Lagoinha	16	Colégio Batista	16
23	União	1	Graça	7
23	União	2	Nova Floresta	7
23	União	3	Ipiranga	7
23	União	4	Vila Ipiranga	7
23	União	5	Silveira	7
23	União	6	União	7
23	União	7	Cidade Nova	7
24	Floresta	1	Floresta	12
24	Floresta	2	Sagrada Família	12
24	Floresta	3	Horto Florestal	12
24	Floresta	4	Santa Tereza	12
24	Floresta	5	Buraco Quente	12
24	Floresta	6	São Vicente	12
24	Floresta	7	Horto	12
24	Floresta	8	João Alfredo	12
24	Floresta	9	Pompéia	12
24	Floresta	10	Vila São Rafael	12

24	Floresta	11	Vila Nossa Senhora do Rosário	12
24	Floresta	12	Esplanada	12
25	Boa Vista	1	Santa Inês	6
25	Boa Vista	2	Camponesa I	6
25	Boa Vista	3	Camponesa II	6
25	Boa Vista	4	Camponesa III	6
25	Boa Vista	5	Boa Vista	6
25	Boa Vista	6	Nova Vista	6
26	João Pinheiro	1	Califórnia	18
26	João Pinheiro	2	Camargos	18
26	João Pinheiro	3	Virgínia	18
26	João Pinheiro	4	Sport Club	18
26	João Pinheiro	5	Oeste	18
26	João Pinheiro	6	Maravilha I	18
26	João Pinheiro	7	Maravilha II	18
26	João Pinheiro	8	Santa Maria	18
26	João Pinheiro	9	Vila Oeste	18
26	João Pinheiro	10	Alto dos Pinheiros	18
26	João Pinheiro	11	João Pinheiro	18
26	João Pinheiro	12	Delta	18
26	João Pinheiro	13	Conjunto Califórnia I	18
26	João Pinheiro	14	Vila Califórnia	18
26	João Pinheiro	15	Conjunto Califórnia II	18
26	João Pinheiro	16	Novo Dom Bosco	18
26	João Pinheiro	17	Álvaro Camargos	18
26	João Pinheiro	18	Conjunto Filadélfia	18
27	Carlos Prates	1	Coração Eucarístico	10
27	Carlos Prates	2	Dom Cabral	10
27	Carlos Prates	3	Vila PUC	10
27	Carlos Prates	4	Minas Brasil	10
27	Carlos Prates	5	Vila das Oliveiras	10
27	Carlos Prates	6	Padre Eustáquio	10
27	Carlos Prates	7	Lorena	10
27	Carlos Prates	8	Marmiteiros	10
27	Carlos Prates	9	Carlos Prates	10
27	Carlos Prates	10	Peru	10
28	Centro	1	Centro	9
28	Centro	2	Barro Preto	9
28	Centro	3	Santo Agostinho	9
28	Centro	4	Lourdes	9
28	Centro	5	Savassi	9
28	Centro	6	Boa Viagem	9
28	Centro	7	Funcionários	9

28	Centro	8	Santa Efigênia	9
28	Centro	9	Vila União	9
29	Prado	1	Salgado Filho	16
29	Prado	2	Jardim América	16
29	Prado	3	Barão Homem de Melo I	16
29	Prado	4	Barão Homem de Melo II	16
29	Prado	5	Barão Homem de Melo III	16
29	Prado	6	Santa Sofia	16
29	Prado	7	Alpes	16
29	Prado	8	Nova Granada	16
29	Prado	9	Barão Homem de Melo IV	16
29	Prado	10	Grajaú	16
29	Prado	11	Alto Barroca	16
29	Prado	12	Barroca	16
29	Prado	13	Nova Suissa	16
29	Prado	14	Calafate	16
29	Prado	15	Vila Calafate	16
29	Prado	16	Prado	16
30	Santo Antônio	1	Gutierrez	7
30	Santo Antônio	2	Cidade Jardim	7
30	Santo Antônio	3	Monte São José	7
30	Santo Antônio	4	Luxemburgo	7
30	Santo Antônio	5	Vila Bandeirantes	7
30	Santo Antônio	6	Santo Antônio	7
30	Santo Antônio	7	Coração de Jesus	7
31	Sion	1	São Pedro	13
31	Sion	2	Carmo	13
31	Sion	3	Sion	13
31	Sion	4	Mala e Cuia	13
31	Sion	5	Cruzeiro	13
31	Sion	6	Pindura Saia	13
31	Sion	7	Vila FUMEC	13
31	Sion	8	Anchieta	13
31	Sion	9	Comiteco	13
31	Sion	10	Mangabeiras	13
31	Sion	11	Acaba Mundo	13
31	Sion	12	Belvedere	13
31	Sion	13	Serra do Curral	13
32	Serra	1	Serra	17
32	Serra	2	Santa Isabel	17
32	Serra	3	São Lucas	17
32	Serra	4	Nossa Senhora da Aparecida	17
32	Serra	5	Nossa Senhora da Conceição	17

32	Serra	6	Marçola	17
32	Serra	7	Nossa Senhora de Fátima	17
32	Serra	8	Santana do Cafezal	17
32	Serra	9	Cafezal	17
32	Serra	10	Nossa Senhora do Rosário	17
32	Serra	11	Novo São Lucas	17
32	Serra	12	Vila Novo São Lucas	17
32	Serra	13	Fazendinha	17
32	Serra	14	Paraíso	17
32	Serra	15	Cônego Pinheiro A	17
32	Serra	16	Cônego Pinheiro B	17
32	Serra	17	Vila Paraíso	17
33	Baleia	1	Baleia	21
33	Baleia	2	Cidade Jardim Taquaril	21
33	Baleia	3	Pirineus	21
33	Baleia	4	Jonas Veiga	21
33	Baleia	5	Saudade	21
33	Baleia	6	Belém	21
33	Baleia	7	Vera Cruz	21
33	Baleia	8	Vila Alto Vera Cruz	21
33	Baleia	9	São Geraldo	21
33	Baleia	10	Grota	21
33	Baleia	11	Vila São Geraldo	21
33	Baleia	12	Vila Boa Vista	21
33	Baleia	13	Mariano de Abreu	21
33	Baleia	14	Casa Branca	21
33	Baleia	15	Caetano Furquim	21
33	Baleia	16	Alto Vera Cruz	21
33	Baleia	17	Vila da Área	21
33	Baleia	18	Granja de Freitas	21
33	Baleia	19	Taquaril	21
33	Baleia	20	Conjunto Taquaril	21
33	Baleia	21	Castanheiras	21
34	Barreiro	1	Diamante	12
34	Barreiro	2	João Paulo II	12
34	Barreiro	3	Vila Átila de Paiva	12
34	Barreiro	4	Átila de Paiva	12
34	Barreiro	5	Ademar Maldonado	12
34	Barreiro	6	Teixeira Dias	12
34	Barreiro	7	Santa Helena	12
34	Barreiro	8	Barreiro	12
34	Barreiro	9	Santa Margarida	12
34	Barreiro	10	Bairro das Indústrias I	12

34	Barreiro	11	Vila COPASA	12
34	Barreiro	12	Bairro das Indústrias II	12
35	Bonsucesso	1	Cardoso	16
35	Bonsucesso	2	Brasil Industrial	16
35	Bonsucesso	3	Miramar	16
35	Bonsucesso	4	Araguaia	16
35	Bonsucesso	5	Milionários	16
35	Bonsucesso	6	Vila Nova dos Milionários	16
35	Bonsucesso	7	Bairro Novo das Indústrias	16
35	Bonsucesso	8	Alta Tensão	16
35	Bonsucesso	9	Conjunto Bonsucesso	16
35	Bonsucesso	10	Bonsucesso	16
35	Bonsucesso	11	Vila Nova Paraíso	16
35	Bonsucesso	12	Bernadete	16
35	Bonsucesso	13	Esperança	16
35	Bonsucesso	14	Vila CEMIG	16
35	Bonsucesso	15	Flávio Marques Lisboa	16
35	Bonsucesso	16	Flávio de Oliveira	16
36	Buritis	1	Buritis	9
36	Buritis	2	Palmeiras	9
36	Buritis	3	Parque São José	9
36	Buritis	4	Marajó	9
36	Buritis	5	Havaí	9
36	Buritis	6	Vila Havaí	9
36	Buritis	7	Ventosa	9
36	Buritis	8	Estoril	9
36	Buritis	9	Pantanal	9
37	Olhos d'Água	1	Pilar	6
37	Olhos d'Água	2	São João	6
37	Olhos d'Água	3	Olhos d'Água	6
37	Olhos d'Água	4	Vila Pilar	6
37	Olhos d'Água	5	Vila Olhos d'Água	6
37	Olhos d'Água	6	Santa Lúcia	6
38	Independência	1	Vale do Jatobá	26
38	Independência	2	Mangueiras	26
38	Independência	3	Antenas	26
38	Independência	4	Vila Mangueiras	26
38	Independência	5	Independência	26
38	Independência	6	Mineirão	26
38	Independência	7	Vila Independência III	26
38	Independência	8	Vila Independência IV	26
38	Independência	9	Vila Independência I	26
38	Independência	10	Ernesto Nascimento	26

38	Independência	11	Jardim do Vale	26
38	Independência	12	Santa Cecília	26
38	Independência	13	Castanheira	26
38	Independência	14	Vila Formosa	26
38	Independência	15	Petrópolis	26
38	Independência	16	Vila Petrópolis	26
38	Independência	17	Santa Rita	26
38	Independência	18	Distrito Industrial do Jatobá	26
38	Independência	19	Vila Pinho	26
38	Independência	20	Vitória da Conquista	26
38	Independência	21	Águas Claras	26
38	Independência	22	Vila Ecológica	26
38	Independência	23	Novo Santa Cecília	26
38	Independência	24	Solar do Barreiro	26
38	Independência	25	Corumbiara	26
38	Independência	26	Pongelupe	26
39	Lindéia	1	Lindéia	14
39	Lindéia	2	Jatobá	14
39	Lindéia	3	Vila Piratininga	14
39	Lindéia	4	Itaipu	14
39	Lindéia	5	Marilândia	14
39	Lindéia	6	CDI Jatobá	14
39	Lindéia	7	Jatobá IV	14
39	Lindéia	8	Batik	14
39	Lindéia	9	Túnel de Ibirité	14
39	Lindéia	10	Tirol	14
39	Lindéia	11	Marieta II	14
39	Lindéia	12	Marieta I	14
39	Lindéia	13	Olaria	14
39	Lindéia	14	Vila Tirol	14
40	Gameleira	1	Estrela do Oriente	21
40	Gameleira	2	Betânia	21
40	Gameleira	3	Cinquentenário	21
40	Gameleira	4	Vila Vista Alegre	21
40	Gameleira	5	Vista Alegre	21
40	Gameleira	6	Madre Gertrudes	21
40	Gameleira	7	Vila Madre Gertrudes IV e V	21
40	Gameleira	8	Vila Madre Gertrudes I	21
40	Gameleira	9	Vila Madre Gertrudes II	21
40	Gameleira	10	Custodinha	21
40	Gameleira	11	Vila Madre Gertrudes III	21
40	Gameleira	12	Jardinópolis	21
40	Gameleira	13	Cabana do Pai Tomás	21

40	Gameleira	14	Nova Cintra	21
40	Gameleira	15	Nova Gameleira	21
40	Gameleira	16	Vila Nova Gameleira I	21
40	Gameleira	17	Imbaúbas	21
40	Gameleira	18	Vila Nova Gameleira II	21
40	Gameleira	19	Gameleira	21
40	Gameleira	20	Vila Nova Gameleira III	21
40	Gameleira	21	Guaratã	21
41	Barragem Santa Lúcia	1	São Jorge	12
41	Barragem Santa Lúcia	2	Vila Antena	12
41	Barragem Santa Lúcia	3	Leonina	12
41	Barragem Santa Lúcia	4	Chácara Leonina	12
41	Barragem Santa Lúcia	5	São Bento	12
41	Barragem Santa Lúcia	6	Conjunto Santa Maria	12
41	Barragem Santa Lúcia	7	Ápia	12
41	Barragem Santa Lúcia	8	Vila Paris	12
41	Barragem Santa Lúcia	9	Barragem Santa Lúcia	12
41	Barragem Santa Lúcia	10	Vila Barragem Santa Lúcia	12
41	Barragem Santa Lúcia	11	Santa Rita de Cássia	12
41	Barragem Santa Lúcia	12	Estrela	12

Fonte: Construção de grupos de bairros a partir da informação de bairros presente no banco de dados processado.

Relação de escolas e dos betas.

Tabela G.2 – Relação completa das escolas e dos betas utilizados.

cod_entidade	nome_escola	Código Banco processado	Betas
31000027	EE AARÃO REIS	e1	22.210
31000035	EE BUENO BRANDÃO	e2	42.219
31000051	EE GUIA LOPES	e3	-46.279
31000060	EE PERO VAZ DE CAMINHA	e4	13.491
31000078	EE MAJOR DELFINO DE PAULA RICARDO	e5	14.317
31000086	EE PROFESSOR JOSÉ MESQUITA DE CARVALHO	e6	26.233
31000116	EE CABANA DO PAI TOMÁS	e7	19.159
31000124	EE DR SIMÃO TAMM BIAS FORTES	e8	26.366
31000159	EE MANUEL CASASSANTA	e9	28.715
31000191	EE CAIO NELSON DE SENA	e10	77.302
31000205	EE PROFESSORA FRANCISCA MALHEIROS	e11	18.712
31000213	EE HELENA PENA	e12	25.820
31000221	EE PRES ANTÔNIO CARLOS	e13	43.629
31000230	EE MAL DEODORO DA FONSECA	e14	0.000
31000256	EE JOSÉ MIGUEL DO NASCIMENTO	e15	14.103
31000264	EE AFONSO PENA	e16	0.000
31000272	EE PE JOÃO MARIA KOOYMAN	e17	37.407
31000329	EE SANTOS ANJOS	e18	34.488
31000337	EE JOSÉ MENDES JÚNIOR	e19	69.218
31000361	EE DOM BOSCO	e20	14.216
31000370	EE HENRIQUE DINIZ	e21	-5.794
31000396	EE MARIANO DE ABREU	e22	14.715
31000400	EE SÃO BENTO	e23	51.334
31000426	EE ALBERTO DELPINO	e24	-19.366
31000451	EE HUGO WERNECK	e25	74.469
31000469	EE PRINCESA ISABEL	e26	57.564
31000477	EE MARIETA BROCHADO	e27	21.319
31000507	EE DULCE PINTO RODRIGUES	e28	39.826
31000540	EE PROF AFFONSO NEVES	e29	6.642
31000558	EE MÁRIO CASASSANTA	e30	0.367
31000566	EE SÃO SALVADOR	e31	19.621
31000591	EE ÁLVARO LAUREANO PIMENTEL	e32	24.328
31000604	EE DOMINGAS MARIA DE ALMEIDA	e33	33.569
31000621	EE PROF ALCINDO VIEIRA	e34	15.692
31000655	EE DO INST AGRONÔMICO	e35	72.983

31000663	EE ALZIRA ALBUQUERQUE MOSQUEIRA	e36	68.718
31000671	EE CARLOS CAMPOS	e37	9.902
31000701	EE PROF BATISTA SANTIAGO	e38	24.659
31000728	EE SARAH KUBITSCHKE BAIRO GRAÇA	e39	43.502
31000744	EE AMÉLIA JOSEFINA KEESEN	e40	0.000
31000752	EE CARLOS GÓES	e41	24.729
31000761	EE DONA AUGUSTA G NOGUEIRA	e42	15.126
31000809	EE SARAH KUBITSCHKE IPIRANGA	e43	6.516
31000833	EE CARMO GIFFONI	e44	11.084
31000841	EE DR AURINO MORAIS	e45	18.900
31000876	EE NECÉSIO TAVARES	e46	45.637
31000892	EE JOÃO ALPHONSUS	e47	26.980
31000914	EE CARVALHO BRITO	e48	38.848
31000922	EE DR EUZÉBIO DIAS BICALHO	e49	16.289
31000931	EE JORNALISTA JORGE PAES SARDINHA	e50	24.194
31000949	EE PROF CLÓVIS SALGADO	e51	0.000
31000957	EE NOSSA SENHORA APARECIDA	e52	1.081
31000973	EE JOSÉ BONIFÁCIO	e53	31.373
31000990	EE CECÍLIA MEIRELES	e54	19.755
31001015	EE JOSÉ CARLOS DE GUARANÁ MENEZES	e55	-50.865
31001058	EE JOSÉ IZIDORO DE MIRANDA	e56	30.458
31001091	EE JOSÉ MENDES CORREA	e57	33.942
31001121	EE SÉRGIA CALDEIRA ALKIMIN	e58	21.363
31001155	EE CESÁRIO ALVIM	e59	0.000
31001163	EE DUQUE DE CAXIAS	e60	51.276
31001171	EE JÚLIA LOPES DE ALMEIDA	e61	14.401
31001180	EE PROFESSOR LEON RENAULT	e62	67.494
31001201	EE SILVIANO BRANDÃO	e63	26.794
31001210	EE ANITA BRINA BRANDÃO	e64	27.982
31001244	EE EFIGÊNIO SALLES	e65	70.642
31001252	EE LAÍCE AGUIAR	e66	33.353
31001279	EE OLÍMPIA REZENDE PEREIRA	e67	33.321
31001287	EE TIBIRIÇÁ	e68	35.617
31001295	EE LAUDIEME VAZ DE MELO	e69	9.907
31001317	EE CORAÇÃO EUCARÍSTICO	e70	40.511
31001325	EE ELISEU LABORNE E VALE	e71	9.937
31001333	EE LAURA DAS CHAGAS FERREIRA	e72	53.910
31001350	EE OLÍVIA PINTO DE CASTRO LEITE	e73	23.119
31001368	COLÉGIO TIRADENTES PMMG -UNIDADE ARGENTINO MADEIRA	e74	44.237
31001376	EE LÚCIO DOS SANTOS	e75	38.236

31001384	EE BOLIVAR TINOCO MINEIRO	e76	28.744
31001392	EE CORONEL JUCA PINTO	e77	12.352
31001422	EE PROF MAGALHÃES DRUMOND	e78	11.931
31001431	EE ONDINA AMARAL BRANDÃO	e79	62.824
31001449	EE TITO FULGÊNCIO	e80	73.501
31001457	EE MADRE CARMELITA	e81	21.426
31001465	EE PROFA MARIA AUXILIADORA LANNA	e82	17.339
31001473	EE CEL VICENTE TORRES JÚNIOR	e83	0.000
31001481	EE ENGENHEIRO PRADO LOPES	e84	47.076
31001520	EE TOMÁS BRANDÃO	e85	26.123
31001546	EE ARTUR JOVIANO	e86	30.051
31001562	EE ENGENHEIRO SÍLVIO FONSECA	e87	20.443
31001571	EE PE EUSTÁQUIO	e88	37.073
31001597	EE ELPÍDIO ARISTIDES DE FREITAS	e89	29.020
31001601	EE URSULINA DE ANDRADE MELO	e90	-37.666
31001619	EE PADRE JOÃO BOTELHO	e91	33.344
31001627	EE ASSIS CHATEAUBRIAND	e92	19.999
31001678	EE PROF ANTÔNIO J RIBEIRO FILHO	e93	21.612
31001686	EE WALT DISNEY	e94	-4.730
31001708	EE ASSIS DAS CHAGAS	e95	66.534
31001732	EE PANDIÁ CALÓGERAS	e96	45.329
31001741	EE PROFESSORA BENVINDA DE CARVALHO	e97	-23.867
31001759	EE DESEMBARGADOR MÁRIO GONÇALVES MATOS	e98	43.236
31001767	EE DR JOSÉ PATROCÍNIO DA SILVA PONTES	e99	22.522
31001821	EE PROFA MARIA AMÉLIA GUIMARÃES	e100	38.088
31001830	EE DIOGO DE VASCONCELOS	e101	25.958
31001848	EE DR ANTÔNIO AUGUSTO S CANEDO	e102	18.485
31001856	EE GENERAL CARLOS LUIZ GUEDES	e103	28.578
31001864	EE BARÃO DE MACAÚBAS	e104	30.753
31001872	EE DIVINA PROVIDÊNCIA	e105	13.395
31001881	EE FRANCISCO MENEZES FILHO	e106	25.897
31001899	EE PEDRO DUTRA	e107	97.131
31001902	EE PROFA MARIA BELMIRA TRINDADE	e108	5.395
31001911	EE ADALBERTO FERRAZ	e109	26.071
31001937	EE PEDRO FRANCA	e110	14.876
31001953	EE GERALDO TEIXEIRA DA COSTA	e111	15.872
31001988	EE PROFA NAIR DE OLIVEIRA SANTANA	e112	7.623
31001996	INSTITUTO DE EDUCAÇÃO DE MINAS GERAIS	e113	34.196
31002003	EE BENJAMIM GUIMARÃES	e114	54.408
31002011	EE GERALDINA SOARES	e115	-12.595
31002071	EE ISABEL DA SILVA POLCK	e116	0.000

31002097	EE JOSÉ DE ALENCAR	e117	12.414
31002101	EE SAGRADA FAMÍLIA I	e118	0.000
31002127	EE BRITALDO SOARES FERREIRA DINIZ	e119	4.272
31002186	EE SANDOVAL DE AZEVEDO	e120	27.704
31002194	EE PROFA MARIA CECÍLIA DE MELO	e121	15.850
31002208	EE MENINO JESUS DE PRAGA	e122	55.275
31002216	EE ANTENOR PESSOA	e123	-7.569
31002232	EE AFONSO PENA MASCARENHAS	e124	17.414
31002241	EE PASCHOAL COMANDUCCI	e125	104.723
31002259	EE AFRÂNIO DE MELO FRANCO	e126	40.115
31002267	EE PEDRO PAULO PENIDO	e127	16.369
31002283	EE PROF ALBERTO MAZONI ANDRADE	e128	1.361
31002291	EE ANTÔNIO CLEMENTE	e129	20.376
31002305	EE PROF BOLIVAR DE FREITAS	e130	12.018
31002321	EE PROF JOÃO CÂMARA	e131	34.232
31002330	EE SÍRIA MARQUES DA SILVA	e132	15.655
31002348	EE PROFESSORA INÊS GERALDA DE OLIVEIRA	e133	25.060
31002356	EE DEPUTADO ÁLVARO SALLES	e134	20.323
31002372	EE CELMAR BOTELHO DUARTE	e135	10.544
31002381	EE SÃO PEDRO E SÃO PAULO	e136	32.742
31002399	EE ORÔNCIO MURGEL DUTRA	e137	2.107
31002411	EE DEPUTADO MANOEL COSTA	e138	31.097
31002445	EE PROFA MARIA COUTINHO	e139	33.569
31002453	EE GERALDINA ANA GOMES	e140	31.506
31002461	EE PROFESSORA MARIA MUZZI GUASTAFERRO	e141	37.824
31002496	EE JOSÉ HEILBUTH GONÇALVES	e142	29.162
31002518	EE JUSCELINO KUBITSCHK DE OLIVEIRA	e143	25.156
31002526	EE CARLOS DRUMMOND DE ANDRADE	e144	22.158
31002534	EE CORONEL MANOEL SOARES DO COUTO	e145	26.984
31002542	EE GETÚLIO VARGAS	e146	28.857
31002551	EE MARGARIDA DE MELLO PRADO	e147	1.180
31002593	EM DE BELO HORIZONTE	e148	16.613
31002615	EM EMÍDIO BERUTTO	e149	37.245
31002631	EM MONS ARTUR DE OLIVEIRA	e150	39.604
31002658	EM PE FLÁVIO GIAMMETTA	e151	21.813
31002674	EM MARCONI	e152	7.439
31002682	EM FRANCISCA ALVES	e153	16.597
31002691	EM MONS JOÃO R DE OLIVEIRA	e154	5.550
31002704	EM FRANCISCA DE PAULA	e155	13.070
31002739	EM N SRA DO AMPARO	e156	8.606
31002747	EM FRANCISCO BRESSANE DE AZEVEDO	e157	5.341

31002763	EM GEORGE RICARDO SALUM	e158	23.197
31002798	EM AGENOR ALVES DE CARVALHO	e159	34.847
31002801	EM HELENA ANTIPOFF	e160	38.720
31002828	EM HONORINA DE BARROS	e161	25.728
31002836	EM AIRES DA MATA MACHADO	e162	22.968
31002844	EM HONORINA RABELO	e163	14.115
31002852	EM PE GUILHERME PETERS	e164	36.973
31002861	EM HUGO PINHEIRO SOARES	e165	12.922
31002887	EM HUGO WERNECK	e166	18.323
31002895	EM PE HENRIQUE BRANDÃO	e167	29.860
31002917	EM ANA ALVES TEIXEIRA	e168	27.988
31002925	EM ISRAEL PINHEIRO	e169	-0.229
31002933	EM PEDRO ALEIXO	e170	26.255
31002941	EM PROF JOÃO CAMILO O TORRES	e171	35.867
31002950	EM ANTÔNIO MOURÃO GUIMARÃES	e172	29.386
31002968	EM JOÃO DO PATROCÍNIO	e173	21.273
31002976	EM PÉRSIO PEREIRA PINTO	e174	37.049
31002984	EM JOÃO PINHEIRO	e175	21.625
31002992	EM ANTÔNIO SALLES BARBOSA	e176	26.697
31003000	EM JONAS BARCELOS CORREA	e177	18.978
31003018	EM PREF AMINTHAS DE BARROS	e178	14.891
31003026	EM JÚLIA PARAÍSO	e179	29.146
31003034	EM ARTHUR GUIMARÃES	e180	10.527
31003042	EM LEVINDO LOPES	e181	22.143
31003051	EM PREF SOUZA LIMA	e182	14.363
31003069	EM LUIGI TONIOLO	e183	21.412
31003077	EM AUGUSTA MEDEIROS	e184	23.499
31003093	EM PROFA ALCIDA TORRES	e185	12.317
31003107	EM MAGALHÃES DRUMOND	e186	29.683
31003115	EM AURÉLIO PIRES	e187	28.774
31003123	EM MARIA DE MAGALHÃES PINTO	e188	22.624
31003131	EM PROFA ALICE NACIF	e189	15.246
31003140	EM MARIA DAS NEVES	e190	15.421
31003158	EM BENJAMIN JACOB	e191	5.121
31003166	EM MARIA DE REZENDE COSTA	e192	37.198
31003174	EM DINORAH MAGALHÃES FABRI	e193	10.616
31003182	EM PROF MÁRIO WERNECK	e194	30.970
31003191	EM CARLOS GOIS	e195	25.746
31003212	EM PROFA EFIGÊNIA VIDIGAL	e196	16.072
31003221	EM MESTRE PARANHOS	e197	7.961
31003239	EM CARMELITA CARVALHO GARCIA	e198	21.621

31003255	EM CÔNEGO SEQUEIRA	e199	14.733
31003280	EM PROFA MARIA MODESTA CRAVO	e200	40.563
31003298	EM DEP MÍLTON SALLES	e201	7.657
31003301	EM ANTÔNIO ALEIXO	e202	34.052
31003310	EM DOM JAIME DE BARROS CÂMARA	e203	39.570
31003328	EM PROF CLÁUDIO BRANDÃO	e204	36.881
31003336	EM DOM ORIONE	e205	34.618
31003344	EM PROF DOMICIANO VIEIRA	e206	1.752
31003379	EM PROF MELLO CANÇADO	e207	15.393
31003387	EM SANTOS DUMONT	e208	9.949
31003395	EM SÃO RAFAEL	e209	-14.823
31003409	EM SEBASTIÃO GUILHERME OLIVEIRA	e210	15.456
31003417	EM SENADOR LEVINDO COELHO	e211	-3.514
31003425	EM TENENTE MANOEL MAGALHÃES PENIDO	e212	26.122
31003433	EM WLADIMIR DE PAULA GOMES	e213	6.609
31003441	EM IMACO	e214	40.580
31003450	EM ADAUTO LÚCIO CARDOSO	e215	23.265
31003468	EM GRACY VIANNA LAGE	e216	16.470
31003476	EM ANTÔNIA FERREIRA	e217	15.211
31003484	EM HILDA RABELLO MATTA	e218	17.516
31003492	EM ANTÔNIO GOMES HORTA	e219	14.967
31003506	EM JOAQUIM DOS SANTOS	e220	28.631
31003514	EM CÔNEGO RAIMUNDO TRINDADE	e221	35.401
31003522	EM JOSEFINA SOUZA LIMA	e222	6.485
31003531	EM CÔNSUL ANTÔNIO CADAR	e223	28.906
31003549	EM JOSÉ MADUREIRA HORTA	e224	25.910
31003557	EM CORA CORALINA	e225	28.964
31003565	EM JOSÉ MARIA ALKIMIM	e226	15.466
31003581	EM LÍDIA ANGÉLICA	e227	22.164
31003590	EM DORA TOMICH LAENDER	e228	18.240
31003603	EM MARIA SILVEIRA	e229	6.254
31003611	EM ELISA BUZELIN	e230	21.640
31003620	EM MÁRIO MOURÃO FILHO	e231	8.897
31003638	EM FRANCISCO CAMPOS	e232	18.641
31003646	EM MÍLTON CAMPOS	e233	20.628
31003662	EM MINERVINA AUGUSTA	e234	32.881
31003671	EM MOYSÉS KALIL	e235	13.581
31003697	EM PROFA ACIDÁLIA LOTT	e236	8.872
31003701	EM PROFA CONSUELITA CÂNDIDA	e237	24.978
31003719	EM PROFA HELENA ABDALA	e238	13.635
31003727	EM PROFA ONDINA NOBRE	e239	23.401

31003743	EM PROF MOACYR ANDRADE	e240	30.954
31003760	EM DEP RENATO AZEREDO	e241	40.414
31003778	EM SEBASTIANA NOVAIS	e242	35.225
31003786	EM SEC HUMBERTO ALMEIDA	e243	9.212
31003794	EM TANCREDO PHIDEAS GUIMARÃES	e244	27.986
31003808	EM TRISTÃO DA CUNHA	e245	38.434
31003816	EM VICENTE GUIMARÃES	e246	23.229
31205699	EM IGNÁCIO DE ANDRADE MELO	e247	17.190
31212768	EM HENRIQUETA LISBOA	e248	34.061
31212776	EM DULCE MARIA HOMEM	e249	28.362
31212784	EM PROF EDSON PISANI	e250	8.326
31212792	EM JARDIM FELICIDADE	e251	4.336
31217786	EM AURÉLIO BUARQUE DE HOLANDA	e252	20.329
31217794	EM MARLENE PEREIRA RANCANTE	e253	33.580
31217824	EM ELOY HERALDO LIMA	e254	27.744
31217832	EM LUIZ GONZAGA JÚNIOR	e255	27.174
31217841	EM VINÍCIUS DE MORAIS	e256	21.535
31217859	EM DA VILA PINHO	e257	17.541
31217867	EM EDITH PIMENTA DA VEIGA	e258	31.132
31217875	EM FERNANDO DIAS COSTA	e259	22.122
31217883	EM MURILO RUBIÃO	e260	27.061
31217891	EM FRANCISCO MAGALHÃES GOMES	e261	26.418
31218618	EM PE FRANCISCO CARVALHO MOREIRA	e262	17.680
31219185	EM ANNE FRANK	e263	21.280
31219223	EM OSWALDO FRANÇA JÚNIOR	e264	27.690
31219231	EM SANTA TEREZINHA	e265	30.843
31219398	EM PREF OSWALDO PIERUCCETTI	e266	23.624
31219550	EM SOBRAL PINTO	e267	12.908
31219568	EM PROF AMÍLCAR MARTINS	e268	14.956
31223654	EM JOSÉ MARIA DOS MARES GUIA	e269	8.978
31224111	EM GOV OZANAN COELHO	e270	15.896
31231975	EM PEDRO NAVA	e271	23.729
31231983	EM ULISSES GUIMARÃES	e272	7.562
31231991	EM MARIA DA ASSUNÇÃO DE MARCO	e273	15.879
31232009	EM ARMANDO ZILLER	e274	13.986
31232017	EM PRESIDENTE JOÃO PESSOA	e275	58.039
31233366	EM RUI DA COSTA VAL	e276	28.464
31244473	CIAC EM LUCAS MONTEIRO MACHADO	e277	0.000
31246425	EE PROFA ADIR ANDRADE ALBANO	e278	14.915
31246433	EE EMÍLIA CERDEIRA	e279	47.519
31247308	EM HÉLIO PELLEGRINO	e280	14.615

31250651	EM ANÍSIO TEIXEIRA	e281	22.363
31273252	EM UNIÃO COMUNITÁRIA	e282	14.944
31283002	EM PROF EDGAR DA MATTA MACHADO	e283	21.935
31283011	EM FLORESTAN FERNANDES	e284	13.946
31294632	EM PROFESSOR DANIEL ALVARENGA	e285	-14.454
31294641	EM PE MARZANO MATIAS	e286	14.150
31294659	EM PROF HÍLTON ROCHA	e287	36.096
31305481	EM PROF PAULO FREIRE	e288	12.869
31305499	EM JOSÉ DE CALASANS	e289	29.593
31307696	COLÉGIO TIRADENTES PMMG GAMELEIRA	e290	22.479
31307700	COLÉGIO TIRADENTES PMMG MINASCAIXA	e291	22.502
31312614	EM ACADÊMICO VIVALDI MOREIRA	e292	9.763
31312622	EM HERBERT JOSÉ DE SOUZA	e293	17.897
31317357	EE PROFESSOR AGNELO CORREIA VIANA	e294	44.804
31330868	EM VILA FAZENDINHA	e295	17.240
31331589	EM DOM BOSCO	e296	28.069

Fonte: Regressão a partir do banco de dados processado da tese.