

Universidade de Brasília
Instituto de Psicologia
Curso de Pós-Graduação em Psicologia Social,
do Trabalho e das Organizações

Abordagem multinível na avaliação do Programa Bolsa Família

Daniel Ioshiteru Kinpara

Brasília, DF

2013

Universidade de Brasília
Instituto de Psicologia
Curso de Pós-Graduação em Psicologia Social,
do Trabalho e das Organizações

Abordagem multinível na avaliação do Programa Bolsa Família

Daniel Ioshiteru Kinpara

Orientador: Prof. Dr. Jacob Arie Laros

Brasília, DF

2013

Universidade de Brasília
Instituto de Psicologia
Curso de Pós-Graduação em Psicologia Social,
do Trabalho e das Organizações

Abordagem multinível na avaliação do Programa Bolsa Família

Daniel Ioshiteru Kinpara

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações, como requisito parcial à obtenção do grau de Doutor em Psicologia Social, do Trabalho e das Organizações.

Orientador: Prof. Dr. Jacob Arie Laros

Brasília, DF

2013

Abordagem multinível na avaliação do Programa Bolsa Família

Tese de Doutorado avaliada e defendida diante da banca examinadora constituída por:

Prof. Dr. Jacob Arie Laros (Presidente)
Instituto de Psicologia – Universidade de Brasília

Prof^ª. Dr^ª. Gardênia da Silva Abbad (Membro)
Instituto de Psicologia – Universidade de Brasília

Prof^ª. Dr^ª. Girlene Ribeiro de Jesus (Membro)
Faculdade de Educação – Universidade de Brasília

Prof. Dr. Hartmut Günther (Membro)
Instituto de Psicologia – Universidade de Brasília

Prof. Dr. Milton Mattos de Souza (Membro)
Serviço Social da Indústria - SESI

Prof. Dr. Donald Mathew Pianto (Suplente)
Departamento Estatística – Universidade de Brasília

Se partires um dia rumo a Ítaca
Faz votos para que seja longo o caminho,
Repleto de aventuras, repleto de experiências.

Aos Lestrigões, aos Ciclopes
Ou ao colérico Poseidon, nunca temas;
Não encontrarás tais seres no caminho
Se teu pensamento se mantiver elevado,
Se emoção sutil tocar teu corpo e teu espírito.

Nem aos Lestrigões, nem aos Ciclopes,
Nem ao feroz Poseidon vais encontrar,
Se não os lebares na alma,
E se tua alma não os puser diante de ti.

Faz votos de que seja longo o caminho.
Que muitas sejam as manhãs de verão
Em que – com que prazer, com que alegria! –
Entrarás em portos antes nunca vistos.

Detenha-se nos mercados fenícios
Para adquirir finas mercadorias,
Madrepérolas, corais, âmbar e ébanos,
E excitantes perfumes de todos os tipos,
Tantos perfumes excitantes quanto possas.
Vejas as muitas cidades egípcias e aprendas,
Aprendas dos sábios.

Sempre na mente hás de ter Ítaca.
Chegar lá é teu destino.
Mas nunca apresses a viagem.
É melhor que ela dure muitos anos
E que já velho chegues à ilha,
Rico de tudo o que ganhastes pelo caminho,
Sem esperar que Ítaca te dê riquezas.

Ítaca te deu a bela viagem.
Sem ela não terias empreendido o caminho.
Já não tem outra coisa que te dar.

E se a encontras pobre, Ítaca não te enganou.
Sábio como te tornaste, com tantas experiências,
Terás compreendido o que significam as Ítacas.

Konstantínos Kaváfis, “Ítaca”.

Agradecimentos

Gostaria de começar agradecendo ao meu orientador, prof. Jacob Arie Laros, pela sua paciência, conhecimento, presença constante, conselhos e amizade. Alguém que acolheu um profissional de outra área de braços abertos. Alguém que acreditou nas minhas ideias. O senhor foi fundamental para a consecução deste trabalho! *Beste Leraar, dank U wel!*

A todos os professores do Programa de PG do PSTO, pelos seus ensinamentos, pelo apoio incondicional, pelas discussões acadêmicas, pela oportunidade de conhecê-los e por poder dizer orgulhosamente que fui aluno de vocês. Muito obrigado!

Aos membros da banca examinadora, que aceitaram gentilmente o convite e a grande responsabilidade de aferir as minha habilidades e conhecimento. Meu muito e sincero obrigado!

Aos meus *classmates!* A vocês que tornaram o meu curso muito menos árduo, que estiveram presentes nos bons e nos não tão bons momentos, que compartilharam as minhas e as suas ansiedades, cuja confiança se construiu no “fio do bigode”! Um imenso e inestimável obrigado!

Aos funcionários do Instituto de Psicologia, pessoas que trabalham quase “invisíveis”, mas imprescindíveis para que as coisas transcorram da melhor forma possível. Pessoal, meu muito obrigado a vocês e me desculpem a chateação!

À minha família que torceu por mim, que acreditou na minha capacidade, que confiou nas minhas decisões, que deu força para persistir, que me apoiou com suas orações, que esteve presente de maneiras infinitas ao longo destes quatro anos. Obrigado pelo carinho e apoio de todos vocês!

À Universidade de Brasília e ao Instituto de Psicologia, pelo ótimo curso, pelo debate franco, pelo espaço para as diferenças, pelas ideias de um Brasil mais justo e igual. Meu obrigado!

À Embrapa, pela oportunidade de realizar o doutorado.

A todos aqueles que inadvertidamente não mencionei, mas que também contribuíram para que eu alcançasse este título. Obrigado!

Finalmente, mas não menos importante, agradeço a Deus pelas graças recebidas, pela proteção nos momentos de perigo, pelo conforto de minha alma e pela serenidade concedida. Senhor, obrigado por ouvir as minhas preces. Obrigado por me dar forças para persistir. Obrigado por iluminar o meu caminho. Obrigado por estar sempre presente em cada momento de minha vida.

O Eterno é meu pastor e nada me faltará. Far-me-á repousar em pastos verdejantes e me conduzirá por um lugar de plácidas águas. Minha alma será restaurada; guiar-me-á nas veredas da justiça por amor de Seu Nome. Se tiver que seguir pelo sombrio vale da morte, não recearei nenhum mal, porque Tu estarás comigo. Teu apoio depois do Teu castigo ser-me-ão por consolo. Diante de mim preparará uma mesa de delícias na frente dos meus inimigos. Ungiste com óleo de unção a minha cabeça e o meu cálice transborda de fartura. Unicamente a felicidade e a misericórdia me seguirão durante a minha vida. E o meu habitar será por longos dias na mansão do Eterno.

Salmo de Davi

Assim seja.

Sumário

Lista de tabelas	x
Lista de figuras	xii
Lista de abreviações	xv
Resumo	xviii
Abstract	xix
1. Motivação	1
2. Referencial Teórico	2
2.1 Desenvolvimento e educação	2
2.2 Educação fundamental brasileira	6
2.3 O Programa Bolsa Escola.....	10
2.4 O Programa Bolsa Família	13
2.5 Avaliação do Programa Bolsa Família.....	18
2.5.1 Análises descritiva e relacional.....	19
2.5.2 Causalidade	21
2.5.3 Regressão hierárquica	22
2.5.4 Dados.....	23
3. Objetivos	25
3.1 Geral.....	25
3.2 Específicos	25
4. Material e Métodos	25
4.1 Delineamento do estudo	26
4.2 Procedimentos	26
4.3 Pressupostos estatísticos.....	27
4.4 Análise de dados.....	28
4.4.1 Análises descritiva e exploratória	29
4.4.2 Análise multinível	30
4.4.3 Modelos de análise.....	38

5. Resultados	40
5.1 Análises estatísticas descritivas.....	40
5.2 Contexto de pesquisa.....	51
5.3 Resultados dos modelos multinível.....	72
6. Discussão	77
6.1 Os dados	77
6.2 Modelos de regressão multinível.....	83
6.2.1 Modelo com todas as variáveis explicativas	83
6.2.2 Modelo que cobre todo o horizonte de pesquisa.....	88
6.2.3 Modelos antes e depois do evento do PBF.....	90
6.2.4 Modelos de cada componente	95
7. Conclusões	96
7.1 Limitações do estudo.....	97
7.2 Contribuições	99
7.3 Recomendações e sugestões.....	100
8. Referências	102
ANEXO	110

Lista de Tabelas

Tabela 1	Eixos articuladores do “Fome Zero” e seus respectivos programas e ações.	14
Tabela 2	Número de municípios em quatro períodos entre 1997 e 2013.	31
Tabela 3	Descrição das variáveis de nível 1 (medidas repetidas) utilizadas na pesquisa.	31
Tabela 4	Modelos multinível da pesquisa e seus respectivos períodos de análise.	40
Tabela 5	Modelos multinível da pesquisa e seus respectivos componentes.	40
Tabela 6	Estatística de tamanho da população, média e desvio padrão, 2000 a 2010.	41
Tabela 7	Estatísticas descritivas das variáveis de proficiência, escala Saeb, bienais, 2001 a 2009.	43
Tabela 8	Estatísticas descritivas das variáveis de distorção idade-série, percentual, 2000 a 2010.	44
Tabela 9	Estatísticas descritivas das variáveis óbito por diarreia e mortalidade infantil, em número de óbitos e em número de óbitos para cada mil nascidos vivos, respectivamente, 2000 a 2010.	45
Tabela 10	Estatísticas descritivas das variáveis do Programa Bolsa Família, 2004 a 2010, com exceção de 2008.	45
Tabela 11	Estatísticas descritivas das variáveis agropecuárias, em toneladas de grãos e número de cabeças ordenhadas, 2000 a 2010.	47
Tabela 12	Estatísticas descritivas das variáveis econômicas, em mil R\$ por habitante para PIBPC e em porcentagem para os Valores Adicionados Brutos, 2000 a 2010.	48
Tabela 13	Autovalores da análise paralela por permutação da base de dados de 41 variáveis, 2007 e 2009.	49

Tabela 14	Componentes extraídos por Análise dos Componentes Principais, com rotação Varimax.	50
Tabela 15	As variáveis da pesquisa e suas respectivas dimensões.	51
Tabela 16	Modelos multinível M5-Cx, coeficientes padronizados, 2007 e 2009.	72
Tabela 17	Modelos multinível com diferentes períodos de análise longitudinal, coeficientes padronizados.	76
Tabela 18	Coeficientes padronizados β dos modelos multinível, nos diferentes componentes.	76
Tabela 19	Coeficientes não padronizados dos modelos multinível, nos diferentes componentes.	77

Lista de Figuras

Figura 1	Estimativa do percentual de gasto público total em educação em relação ao Produto Interno Bruto (PIB), por nível de ensino, Brasil, 2000-2010.	8
Figura 2	Média do escore na escala Saeb do desempenho escolar das séries iniciais (4S) e finais (8S), para língua portuguesa (P) e matemática (M) em escolas urbanas, Brasil, 1995 a 2011.	9
Figura 3	Valor de investimento do Programa Bolsa Família por região do Brasil, 2003 a 2010.	16
Figura 4	Número de famílias conviventes residentes em domicílios particulares cujo rendimento nominal mensal familiar per capita é de até $\frac{1}{4}$ de salário mínimo (salário mínimo = R\$ 510), por estado, Brasil, 2010.	17
Figura 5	Número de famílias beneficiadas pelo PBF, período de 2004 a 2010.	52
Figura 6	Média dos fundos repassados (mil R\$) para o município, 2004 a 2010.	52
Figura 7	Média do percentual do público elegível que é beneficiário do PBF, 2004 a 2010.	53
Figura 8	Número médio de famílias beneficiadas por município para cada região do Brasil, 2004 a 2010.	54
Figura 9	Número médio de famílias beneficiadas por município para cada unidade federal do Brasil, menos o DF, 2004 a 2010.	54
Figura 10	Média percentual do repasse do PBF em termos de PIB da unidade federal, menos o DF, 2004 a 2010.	55
Figura 11	Média de escores de proficiência em matemática na 4 ^a série/5 ^o ano de escolas urbanas estaduais (a) e municipais (b), nas cinco grandes regiões brasileiras, 2001 a 2009.	56
Figura 12	Média de escores de proficiência em português na 4 ^a série/5 ^o ano de escolas urbanas estaduais (a) e municipais (b), nas 5 grandes regiões brasileiras, 2001 a 2009.	57

Figura 13	Média de escores de proficiência em matemática (a) e português (b) na 8ª série/9º ano de escolas estaduais urbanas, nas 5 grandes regiões brasileiras, 2001 a 2009.	57
Figura 14	Média de escores de proficiência em matemática (a) e português (b) na 4ª série/5º ano de escolas municipais rurais, nas 5 grandes regiões brasileiras, 2001 a 2009.	58
Figura 15	Média de escores de proficiência em Matemática na 4ª série/5º ano de escolas estaduais urbanas (a) e em Português na 8ª série/9º ano de escolas municipais rurais (b), nas 5 grandes regiões brasileiras, 2001 a 2009.	59
Figura 16	Número acumulado de escolas por tipo de dependência, 2000 a 2010.	60
Figura 17	Percentual de escolas por localização, 2000 a 2010.	60
Figura 18	Distorção idade-série (%), de escolas municipais rurais, do 2º ano (a) até o 9º ano (h), 2000 a 2010.	62
Figura 19	Média dos percentuais de DIS em escolas estaduais urbanas, turma de 1ª série/2º ano, nas 5 grandes regiões brasileiras, 2000 a 2010.	63
Figura 20	Média dos percentuais de DIS em escolas estaduais urbanas, turma de 8ª série/9º ano, nas 5 grandes regiões brasileiras, 2000 a 2010.	64
Figura 21	Média de óbitos por diarreia nos municípios, 2000 a 2010.	64
Figura 22	Média da mortalidade infantil nos municípios, 2000 a 2010.	65
Figura 23	Média de óbitos por diarreia por 100 mil habitantes nos municípios, 2000 a 2010.	65
Figura 24	Média de mortalidade infantil nos municípios, 2000 a 2010.	66
Figura 25	Média de produção de feijão nos municípios, 2000 a 2010.	66
Figura 26	Média de produção de milho nos municípios, 2000 a 2010.	67
Figura 27	Média de produção de soja nos municípios, 2000 a 2010.	67
Figura 28	Média de vacas ordenhadas nos municípios, 2000 a 2010.	68

Figura 29	Média de vacas ordenhadas nos municípios, por região, 2000 a 2010.	68
Figura 30	Média de produção de soja nos municípios, por região, 2000 a 2010.	69
Figura 31	Média do percentual dos Valores Adicionados Brutos dos municípios por setor, agrupados por região, 2000 a 2010.	69
Figura 32	Média do percentual dos Valores Adicionados Brutos dos municípios por setor, 2000 a 2010.	70
Figura 33	Média do percentual da inflação anual (INPC) em onze capitais brasileiras em alimentação, educação e saúde, 2000 a 2010.	71
Figura 34	Média do percentual da inflação anual (INPC) em onze capitais brasileiras, 2000 a 2010.	71
Figura 35	Gráficos de resíduos padronizados <i>versus</i> escores normais (a, c, e, g, i) e resíduos padronizados <i>versus</i> valores esperados da parte fixa (b, d, f, h, j) do modelo M5-C1 (a, b), M5-C2 (c, d), M5-C3 (e, f), M5-C4 (g, h) e M5-C5 (i, j), nível 2 (municípios), 2007 e 2009.	74
Figura 36	Gráficos de resíduos padronizados <i>versus</i> escores normais (a, c, e) e resíduos padronizados <i>versus</i> valores esperados da parte fixa (b, d, f) dos modelos M2 (a, b), M3 (c, d) e M4 (e, f), Nível 1 (anos).	75

Lista de Abreviações

ACP	Análise de Componentes Principais
AMN	Análise Multinível
Aneb	Avaliação Nacional do Ensino Básico
Anresc	Avaliação Nacional do Rendimento Escolar
AP	Análise Paralela
BRB	Banco de Brasília
BSP	Benefício para a Superação da Extrema Pobreza na Primeira Infância
BVJ	Benefício Variável vinculado ao Adolescente
CadÚnico	Cadastro Único para Programas Sociais
CAE	Comissão de Assuntos Econômicos do Senado
CCJ	Comissão de Constituição, Justiça e Cidadania
CE	Comissão de Educação
Consad	Consórcio de Segurança Alimentar e Desenvolvimento Local
CRAS	Centro de Referência de Assistência Social
Datasus	Departamento de Informática do SUS
DIS	Distorção Idade-Série
DP	Desvio Padrão
Fundeb	Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação
Fundef	Fundo de Desenvolvimento, Manutenção do Ensino Fundamental e de Valorização do Magistério
GEPE	Grupo de Estudos de Políticas Públicas de Educação
IBGE	Instituto Brasileiro de Geografia e Estatística
ICC	Correlação Intraclasse

Inep	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
INPC	Índice Nacional de Preços ao Consumidor
IPCA	Índice Nacional de Preços ao Consumidor Amplo
MDA	Ministério do Desenvolvimento Agrário
MDS	Ministério do Desenvolvimento Social e Combate à Fome
MEE	Modelagem por Equações Estruturais
MLG	Modelos lineares gerais multivariados
NIS	Número de Inscrição
ODM	Objetivos de Desenvolvimento do Milênio
PAA	Programa de Aquisição de Alimentos da Agricultura Familiar
Pat	Programa de Alimentação do Trabalhador
PBE	Programa Bolsa Escola
PBF	Programa Bolsa Família
Peti	Programa de Erradicação do Trabalho Infantil
PIB	Produto Interno Bruto
Pnad	Pesquisa Nacional por Amostra de Domicílios
Pnae	Programa Nacional de Alimentação Escolar
Pnud	Programa das Nações Unidas para o Desenvolvimento
POF	Pesquisa de Orçamentos Familiares
Pronacampo	Programa Nacional de Educação no Campo
Pronaf	Programa Nacional de Fortalecimento da Agricultura Familiar
Pronatec	Programa Nacional de Acesso ao Ensino Técnico e Emprego
PTR	Programa de Transferência de Renda
RME	Renda Mínima Escolar

Saeb	Sistema Nacional de Avaliação da Educação Básica
Sagi	Secretaria de Avaliação e Gestão da Informação
Secopa	Secretaria Extraordinária da Copa de 2014
Sisvan	Sistema de Vigilância Alimentar e Nutricional
UnB	Universidade de Brasília
VAB	Valor Adicionado Bruto
VD	Variável Dependente

Resumo

O presente trabalho teve por objetivo avaliar o programa de transferência de renda conhecido por Bolsa Família (PBF) por meio de dados secundários oficiais públicos. O período de análise foi de 2000 a 2010. O estudo baseou-se no entendimento de pobreza como medida de desenvolvimento econômico e no conceito de capital humano de Schultz (1961). Nesta perspectiva, a educação é vista como o meio pelo qual se quebra a espiral de pobreza e se proporciona incremento do bem-estar social. No curto prazo, acredita-se que a transferência de renda crie condições básicas para que crianças e adolescentes em idade escolar frequentem a escola. No médio e no longo prazo, espera-se que estejam mais qualificados para o mercado de trabalho, aumentando a renda da família. O método utilizado na presente pesquisa foi a análise de regressão multinível com medidas repetidas. Foram obtidos dados de 5.565 municípios. Estes dados foram agrupados em seis categorias: (1) econômica; (2) agropecuária; (3) demográfica; (4) de saúde; (5) educacional; e (6) do PBF. Os dados econômicos incluíram Produto Interno Bruto e Valor Adicionado Bruto por setor (agropecuário, industrial, serviços e impostos). Os dados agropecuários compreenderam a produção de grãos de feijão, de milho e de soja, bem como o número de cabeças ordenhadas de rebanhos bovinos leiteiros. Os dados demográficos foram a população total municipal e a densidade populacional. Os dados de saúde foram número de óbitos por diarreia e mortalidade infantil. Os dados educacionais foram proficiência escolar e distorção idade-série (DIS). Os dados do PBF foram o número de famílias beneficiadas, o montante em reais repassado para os municípios e o percentual de famílias pobres beneficiadas pelo PBF. Realizou-se a Análise de Componentes Principais (ACP) a fim de determinar como as variáveis explicativas se agrupavam. A análise resultou em cinco componentes: proficiência escolar; DIS de escolas estaduais; DIS de escolas municipais; bem-estar; e produção agropecuária. Foram desenvolvidos cinco modelos: um com todos os componentes, com dados dos anos 2007 e 2009 (M1); um com dados de 2000 a 2010 (M2); um com dados de 2000 a 2003 (M3); um com dados de 2004 a 2010 (M4); e uma análise de cada componente individualmente (M5). O modelo M1 verificou a relação direta entre o PBF e pobreza. Os modelos M3 e M4 foram desenvolvidos para comparar os períodos antes e depois da implementação do PBF. O modelo M2 serviu de referência de comparação para os modelos M3 e M4. O modelo M5 foi desenvolvido para avaliar a relação entre cada um dos cinco componentes e a pobreza, individualmente. O modelo M1 indicou que existe relação significativa entre pobreza e o PBF. Porém, os modelos M2, M3 e M4 não foram conclusivos em apontar esta relação. O modelo M5 verificou que a pobreza está relacionada com: maior desempenho em matemática do que em português, ambos nas séries iniciais; mais DIS; menos óbitos de crianças e adultos; menor produção de riqueza nos setores estudados; e maior produção de feijão. Três conclusões podem ser apontadas: (1) a relação entre PBF e pobreza não é certa; (2) se esta relação de fato existe, o efeito ocorre por meio da educação; e (3) o uso de dados oficiais públicos e a técnica de análise multinível são interessantes como alternativa de avaliação dos impactos de um programa social; mas para as variáveis utilizadas neste estudo, sugere-se um período de pesquisa maior.

Palavras-chave: avaliação de programas sociais, análise multinível, medidas repetidas, programa de transferência de renda.

Abstract

The present study evaluated the Income Transfer Program (ITP) known as *Bolsa Família* (PBF) using secondary data obtained from the Brazilian Government. The analysis period ranged from 2000 to 2010. The study was based on the conception of poverty as a measure of economic development and on Schultz's (1961) concept of human capital. From this perspective, education is considered a means to reduce poverty and to increase welfare. In the short term, income transfer is presumed to establish basic conditions for children and teenagers at school age to attend school. In the long term, it is expected that they will be more qualified for the labor market, thereby increasing family income. The method used in this study was Multilevel Regression Analysis with repeated measures. The study included all 5,565 Brazilian cities. The data were grouped into six categories: (1) economic; (2) agriculture; (3) demographic; (4) health; (5) education; and (6) PBF. Economic data included Gross Domestic Product (GDP) and Gross Added Value by sector (agriculture, industry, services, and taxes). Agriculture data included the production of pinto beans, corn, soybeans, and the number of milking cows. The demographic data were total population and population density. Health data were number of deaths due to diarrhea and infant mortality rate. Education data were school proficiency and age-grade distortion. The PBF data were number of families enrolled in the program, amount of money transferred to the city, and the percentage of poor families benefited by the PBF (coverage). A Principal Components Analysis was performed in order to verify the grouping of the explanatory variables. The analysis resulted in five components: school proficiency; age-grade distortion of State schools; age-grade distortion of school district schools; welfare; and agricultural production. Five models were developed: a model with all components, based on data of 2007 and 2009 (M1); a model based on data of 2000 to 2010 (M2); a model based on data of 2000 to 2003 (M3); a model based on data of 2004 to 2010 (M4); and the analysis of each individual component (M5). The model M1 established a direct relation between PBF and poverty. The M3 and M4 models were designed to compare the periods before and after the implementation of PBF. The model M2 was a reference model to compare regression coefficients of M3 and M4. The model M5 was designed to evaluate the effect of each component on poverty, solely. The model M1 showed that there is a relation between poverty and PBF. However, models M2, M3, and M4 were not conclusive in this respect. The model M5 results indicated that poverty is related to: higher proficiency in Mathematics than in Portuguese in the initial grades of Elementary School; high age-grade distortion; decreased deaths of children and adults; lower production of wealth in the studied economic sectors; and higher production of pinto beans. Three conclusions can be drawn: (1) the relation between PBF and poverty is an uncertain one; (2) if this relation exists, it will take place mediated by education; and (3) the use of governmental data and multilevel analysis are interesting alternatives for impact evaluation of social programs; but for the variables used in this study, a larger period of available data is suggested.

Keywords: social program evaluation, multilevel analysis, repeated measures, income transfer program.

1. Motivação

O ponto de partida deste trabalho foi a agricultura brasileira. O interesse nasceu da percepção do seu papel em prover alimentos baratos e abundantes para a população brasileira, do seu amplo efeito sobre a estabilidade da economia do ponto de vista da inflação e das transformações recentes pelas quais passou. Transformações relacionadas à sua modernização, ao seu novo papel na matriz energética brasileira e às questões de sustentabilidade ambiental.

O homem do campo, o agente por trás desta agricultura destacada, parece gravitar entre dois extremos. De um lado, o arquétipo do “Jeca Tatu” de Monteiro Lobato. Um indivíduo analfabeto, sujo, doente, indolente e pobre da zona rural paulista, esquecido pelo poder público. De outro, o produtor rural “moderno” do Centro-Oeste, grande sojicultor, dono de vastas extensões de terra, afeito à tecnologia, atuando nos mercados internacionais, orientado ao capital, com representantes nas diversas esferas públicas de poder. Afinal, quem é o homem do campo?

Segundo o Ministério do Desenvolvimento Agrário (2009), em 2006, o Brasil contava com mais de 5 milhões de estabelecimentos de produção agropecuária, dos quais 84% eram classificados como agricultura familiar (“pequeno produtor”). A agricultura familiar é caracterizada pelo uso de mão de obra predominantemente da própria família do produtor rural. Estes 84% de agricultores ocupavam apenas 24% da área agrícola total brasileira. A agricultura familiar respondia na época por 70% da produção de feijão, 58% da produção de leite e 59% da produção de suínos no Brasil. Deste total de agricultores familiares, 60% estavam nas regiões Norte e Nordeste, as mais pobres do Brasil (Ministério do Desenvolvimento Agrário, 2009). Enquanto 5% da população urbana é classificada como “extrema pobreza”, no meio rural este percentual é de 25% (Fundo Internacional de Desenvolvimento Agrícola, 2011). Dados de 2008 da Pesquisa Nacional por Amostra de Domicílios – Pnad (Instituto Brasileiro de Geografia e Estatística, 2009) apontavam para outro aspecto diferencial entre o meio urbano e o rural. Cerca de 41,8% da população do meio rural, com idade igual ou acima de 15 anos, eram de analfabetos funcionais, i.e., indivíduos com menos de quatro anos letivos de escolaridade. No meio urbano, o percentual era de 17,2%. Por estes números, percebe-se que existe uma grande parcela da população rural pobre, analfabeta e, paradoxalmente, a principal responsável pelo fornecimento de alimentos importantes da dieta do brasileiro, como o feijão e o leite.

Mas um outro aspecto chama a atenção nesses dados. O Censo Demográfico de 2010 (IBGE, 2010) contabilizava a população urbana em 137 milhões de habitantes. O número de pessoas no meio rural era de 24 milhões. Conseqüentemente, 5% da população urbana representava 6,9 milhões de pessoas em extrema pobreza. No meio rural, 25% representava 6,0 milhões de pessoas. Estes dados apontam para um problema mais grave do que a diferenciação do homem do campo: a pobreza no Brasil. É sobre esta temática que a presente pesquisa passou a se debruçar.

2. Referencial Teórico

O tema pobreza é amplo. O primeiro obstáculo é delimitar uma perspectiva de análise. Neste trabalho, a delimitação começou pela definição dos conceitos de desenvolvimento e pobreza. A partir do entendimento destes conceitos, abordou-se o problema na perspectiva do papel da educação em promover desenvolvimento. Apoiou-se na abordagem da Teoria do Capital Humano (Schultz, 1961). Em seguida, esta perspectiva foi estudada por meio da avaliação de um programa social de transferência de renda denominado de Programa Bolsa Família (PBF). Este estudo foi realizado por meio de dados secundários oficiais e pelo emprego da técnica de análise multinível de medidas repetidas.

2.1 Desenvolvimento e educação

A relação entre desenvolvimento e educação é ainda um tema de acalorados debates. A dificuldade em estabelecer e descrever os mecanismos desta relação parece estar associada a como se entende o desenvolvimento e a educação. A despeito disso, não são poucos os planos de desenvolvimento e políticas de governo que se fiam na assertiva de que aprimorar a educação é o caminho para o desenvolvimento.

Os programas de governo normalmente encaram o desenvolvimento na perspectiva de desenvolvimento econômico. Nesta perspectiva, desenvolvimento está associado a questões como pobreza, desigualdade e exclusão. Na visão de Almeida (2000), a pobreza é uma medida quantitativa da falta de recursos, a desigualdade é a distribuição desigual desses recursos e a exclusão é a marginalização do indivíduo do acesso a esses recursos. Tomando por base esta visão, pode-se definir desenvolvimento econômico como a busca pela diminuição da condição de pobreza do indivíduo, i.e., eliminar a desigualdade por meio da transferência de recursos aos

excluídos. Esta perspectiva de diminuição da pobreza pela simples transferência de recursos já mostrou, no passado, o seu lado negativo na forma de programas de assistência social paternalistas, clientelistas e assistencialistas (Piana, 2009).

Acredita-se que este efeito negativo da transferência de recursos ocorre porque se ignora a complexidade da pobreza. Para Waiselfisz, Abramovay e Andrade (1998), a pobreza é centrada na pessoa e, por isso, trata-se de um conceito multidimensional, onde se destaca a qualidade de vida ao invés dos bens materiais. Os autores definem pobreza como a “negação das escolhas e oportunidades básicas para o desenvolvimento da vida humana, refletida em vida curta, falta de educação, falta de meios materiais, exclusão e falta de liberdade e dignidade”. São três eixos principais: a pobreza material, a pobreza intelectual e a pobreza social. Estes três eixos definem um contexto que impede ou limita o desenvolvimento dos indivíduos. Nesta perspectiva multidimensional do desenvolvimento percebe-se o papel da educação.

Essa mesma perspectiva multidimensional é adotada pelo Programa das Nações Unidas para o Desenvolvimento (Pnud). Um exemplo é a Declaração do Milênio (Nações Unidas, 2000). Os oito Objetivos de Desenvolvimento do Milênio (ODM) referem-se a questões de pobreza, educação, desigualdade social, saúde, sustentabilidade ambiental e de combinação de esforços dispersos pelo mundo.

Esta multidimensionalidade da pobreza cria dificuldades para a pesquisa científica. O maior gargalo está em quantificar elementos como liberdade, dignidade e pobreza intelectual. Uma alternativa à abordagem unidimensional e à abordagem multidimensional foi elaborada a partir da Teoria do Capital Humano, de T. W. Schultz. Esta teoria foi desenvolvida sobre a sua ideia de Capital Educacional (Schultz, 1961). O modelo de Schultz impulsionou as soluções de desenvolvimento econômico baseadas no aprimoramento da educação dos indivíduos. A vantagem deste modelo é a quantificação do desenvolvimento por meio de um conceito ampliado de capital, o qual inclui “capital humano”. Para Schultz (1971), habilidades úteis e conhecimento são formas de capital, frutos de investimento deliberado. Na prática, estes investimentos se traduzem em educação básica e superior, treinamento no trabalho, migração, saúde e informação econômica. Logo, segundo esta visão, a educação não é mais um recurso econômico para o desenvolvimento, mas um investimento em um bem de capital (capital humano).

Entretanto, autores como Spence (1973) e Collins (1979) discordaram de Schultz. Spence (1973) afirma que a educação funciona como um sinalizador da produtividade do indivíduo. Maiores níveis de educação sinalizariam que o indivíduo é provavelmente mais produtivo e, portanto, alguém que devesse receber maiores salários. Contudo, este nível de escolaridade seria apenas um indicador de sua produtividade e não necessariamente a aquisição de habilidades por meio da escola. Para Spence, as habilidades, em grande parte, são inatas do indivíduo. A esta abordagem deu-se o nome de Teoria da Sinalização. Assim como Spence, Collins (1979) não vê como papel da educação dotar o indivíduo de habilidades técnicas. Porém, Collins avança um pouco mais na posição de Spence. Collins afirma que a educação é um meio de adquirir títulos acadêmicos que representam valores de status social. As habilidades técnicas seriam adquiridas na experiência profissional, pois a escola é desconexa do mercado de trabalho e de suas necessidades. Ele aponta para um componente de dominação em que as escolas servem às elites sociais, credenciando as pessoas para desempenharem papéis de dominação ou de dominado. A esta abordagem, deu-se o nome de Teoria Credencialista.

O problema empírico que se apresenta com a Teoria do Capital Humano, a Teoria da Sinalização e a Teoria Credencialista é que todas apontam para um mesmo efeito da educação. Quanto mais educação o indivíduo recebe, maior é a probabilidade dele auferir uma maior renda. Esta dificuldade em distinguir o mecanismo subjacente à relação entre educação e renda já foi apontada por Wu (2004) e Barbosa Filho e Pessôa (2009).

A abordagem de capital humano tem uma visão simples e, por causa disso, mais fácil de ser operacionalizada em termos de intervenção pública. A ideia de capital humano está presente nas políticas públicas brasileiras de desenvolvimento. A partir das discussões ocorridas no âmbito do Grupo de Estudos de Políticas Públicas de Educação (GEPE) sobre a problemática da educação brasileira, Mello (1991) apontou a preocupação das políticas educacionais de aprimorar as competências cognitivas e sociais da população. Segundo o GEPE, estas competências são as matérias-primas das economias modernas e a educação o meio para consegui-las. Além de ampliar a competitividade econômica, a educação promoveria competitividade com equidade social. Aliás, Barbosa e Fernandes (2000, p.136) são enfáticos em afirmar que a

escola é “o instrumento eficaz para a promoção da equidade e mobilidade social”. Esta dimensão social é bem presente no conceito elaborado por Coleman (1988) de “capital social”. A diferença entre capital social e capital humano é o aspecto de bem público do capital social. Como bem público, os atores que geram o capital social recebem apenas frações dos benefícios proporcionados pela sua ação social. Segundo Coleman (1988), este seria um dos motivos do baixo investimento em capital social.

Se por um lado a relação entre educação e desenvolvimento econômico é clara na perspectiva de capital humano, a relação entre educação e pobreza não se mostra tão direta. O trabalho de Yamada e Castro (2011) realizado no Peru mostra que apenas o acesso a níveis educacionais mais elevados (além do primário) permitem quebrar o ciclo de pobreza. Contreras (2008) verificou que o investimento da família na educação das crianças depende principalmente da diferença entre a renda recebida pela criança e a renda recebida pelos adultos (*wage distortion*). No caso de famílias pobres, quanto maior a diferença, menor é o investimento na educação da criança (capital humano). Isto acontece porque a família utiliza o investimento da educação para acumular patrimônio. Utilizando dados empíricos da América Latina para o período de 1970 a 1994, De Janvry e Sedoulet (2000) observaram que o crescimento da renda reduz pobreza, mas não a desigualdade. Porém, nas áreas urbanas, o efeito de redução da pobreza depende do nível de pobreza, do nível de desigualdade e do nível de educação secundária. Estes vários estudos mostram que o efeito da educação não é o único componente da redução da pobreza.

Estudos focando a educação como uma forma de reduzir a pobreza são frequentes. Porém, existe uma outra linha forte de pesquisa que estuda trabalho infantil, pobreza e educação. As pesquisas com trabalho infantil evocam o que se convencionou chamar de “ciclo da pobreza”. Famílias pobres cujos pais não têm educação formal acabam levando os seus filhos a trabalharem em detrimento de sua educação (trabalho infantil). Com isto, quando os filhos chegam à idade adulta, sem educação formal, reproduzem o estado de pobreza da família, fechando o ciclo. Algumas referências sobre o assunto podem ser obtidas no trabalho de Kassouf (2007).

Existem vários exemplos de programas que procuram retirar a criança do mercado de trabalho, como o Programa de Erradicação do Trabalho Infantil (Peti) do

Ministério do Desenvolvimento Social e Combate à Fome (MDS), ou programas que visam alocar o adolescente no mercado de trabalho, como o Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec) do Ministério da Educação e Cultura, o Programa Adolescente Aprendiz do Conselho Nacional do Ministério Público e o QualiCopa da Secretaria Extraordinária da Copa de 2014 (Secopa).

Neste trabalho, será adotada a perspectiva de Schultz de desenvolvimento econômico. Passa-se a seguir à caracterização da educação estudada nesta pesquisa.

2.2 Educação fundamental brasileira

O Programa Bolsa Família (PBF) tem foco nas crianças e adolescentes até os 15 anos, apesar de um novo benefício de 2013 estender esta faixa. Este novo benefício será comentado posteriormente. A permanência do filho do beneficiário na escola é uma condição do programa. Como os levantamentos de proficiência escolar ocorrem na 4^a/5^o e 8^a/9^o séries/anos do ensino fundamental e 3^a série do ensino médio, optou-se nesta pesquisa por limitar o estudo para o ensino fundamental a fim de compatibilizar o PBF e os dados de proficiência.

O Estatuto da Criança e do Adolescente (Lei nº 8.069, 13/07/1990) define criança como o indivíduo com idade até 12 anos incompletos. De 12 a 18 anos, ele é considerado adolescente. A faixa etária entre zero a 17 anos é atendida pela educação básica. Conforme determina a Lei de Diretrizes e Bases da Educação Nacional nº 9.394, de 20 de dezembro de 1996, Art. 21^o, a educação básica é o primeiro nível do ensino escolar. Com a publicação da Emenda Constitucional nº 59, de 11 de novembro de 2009, a educação básica é tornada obrigatória e gratuita dos 4 aos 17 anos. A educação básica ficou dividida em educação infantil, educação fundamental e ensino médio. As idades respectivas para cada etapa são zero a 5 anos, 6 a 14 anos e 15 a 17 anos.

A despeito do Brasil contar com um ensino superior de excelência, o ensino básico se encontra em dificuldades. Ele tem problemas sérios de qualidade de ensino, de acesso a seus serviços, de permanência (evasão) e de progressão dos alunos nos anos (repetência). A repetência e a evasão causam inchaço das primeiras anos. A escola sofre ainda de excesso burocrático, com sobreposição de funções. As características e particularidades regionais não são respeitadas, impondo-se uma uniformização de currículos, de programas, de materiais, de estatutos, de carreiras e

de jornadas de trabalho. A escola ainda mantém uma concepção assistencialista, destinando o seu espaço físico não apenas à educação, mas também se prestando a centro de programas de saúde, de assistência social e nutricional. A expansão da rede física é caótica, levando a unidades congestionadas e outras ociosas. Este investimento em infraestrutura é colocado em risco com a falta de planejamento do custeio das escolas criadas. Há mais de duas décadas, Mello (1991) reportava estes problemas, os quais são encontrados até hoje na educação básica.

Sobre a infraestrutura das escolas, Soares Neto, Jesus, Karino e Andrade (2013) desenvolveram uma escala de “proficiência” em infraestrutura de escolas utilizando a Teoria de Resposta ao Item aplicada a dados do Censo Escolar de 2011. A escala identifica quatro níveis de infraestrutura: elementar (a pior), básica, adequada e avançada (a melhor). Enquanto as regiões Sul, Sudeste e Centro-Oeste apresentavam a maior parte das escolas em um nível básico de infraestrutura, as regiões Norte e Nordeste apresentavam mais de 65% das escolas em nível elementar. Em termos de dependência administrativa (federal, estadual, municipal ou privada), 51% das escolas estaduais encontravam-se em nível básico, enquanto 62% das municipais se encontravam em nível elementar. Também de forma contrastante, em termos de localização da escola (rural ou urbana), enquanto apenas 18% das escolas urbanas estavam em nível elementar, no meio rural, 85% das escolas foram consideradas neste mesmo nível. Assim, as piores condições de infraestrutura escolar encontram-se no Norte e Nordeste, em escolas municipais e em áreas rurais.

Três aspectos melhoraram nos últimos anos: o fluxo escolar da educação fundamental, a taxa de atendimento escolar e o acompanhamento estatístico detalhado da vida escolar na forma do Censo da Educação Básica e do Censo da Educação Superior, realizados desde 1995. Graças a estes acompanhamentos estatísticos, foi possível detectar a melhora do fluxo escolar. Dados do Censo Escolar de 2011 (Inep, 2012a) mostram que existiam 4% de alunos matriculados no ensino fundamental a mais do que o número de indivíduos com a faixa etária correta para ele (de 6 a 14 anos). Em 2000, existiam 20% de alunos matriculados a mais. Em termos de taxa de atendimento escolar na faixa etária de 4 a 17 anos, i.e., percentual da população nessa faixa etária matriculada na escola, os dados do Censo Demográfico de 2000 e 2010

apontavam para 84% e 92% de taxa de atendimento escolar, respectivamente (Todos pela Educação, 2011).

Mas ao que se deve essa profusão de problemas no ensino fundamental? Falta de recursos? Na Figura 1 estão os percentuais de investimento por nível de ensino. Em 2010, verifica-se que 3,5% do PIB foram investidos em ensino fundamental (1º à 9º anos), o equivalente a 140 bilhões de reais. Este montante corresponde a mais de 60% do total gasto na educação brasileira. Desde 2004, os gastos do governo com o ensino fundamental têm sido crescentes.

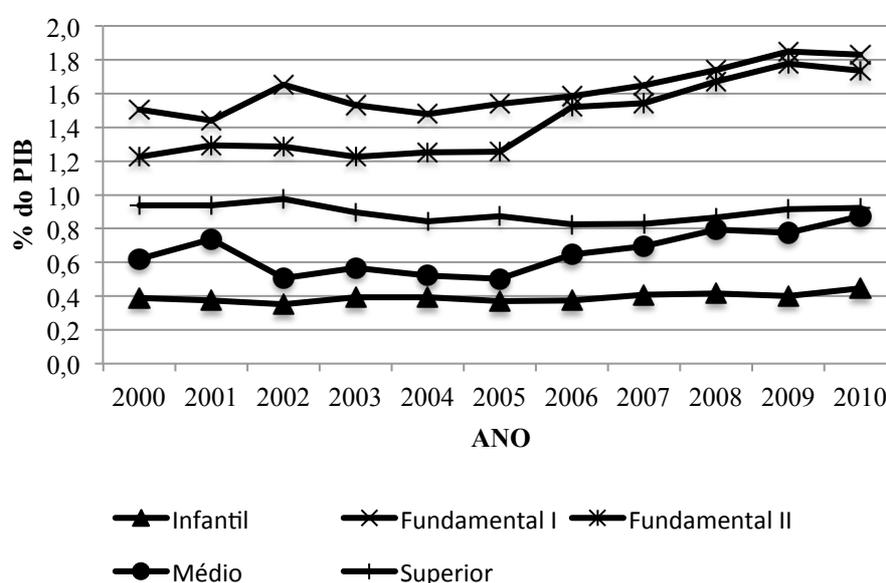


Figura 1. Estimativa do percentual de gasto público total em educação em relação ao Produto Interno Bruto (PIB), por nível de ensino, Brasil, 2000-2010. Adaptado de “Percentual do investimento total em relação ao PIB por nível de ensino”, por Inep, 2011a, recuperado de http://portal.inep.gov.br/web/guest/estatisticas-gastoseducacao-indicadores_financeiros-p.t.i._nivel_ensino.htm

Políticas de ampliação do orçamento em educação não faltaram. Dois exemplos foram o antigo Fundo de Desenvolvimento, Manutenção do Ensino Fundamental e de Valorização do Magistério (Fundef) e o seu substituto, o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb). O primeiro vigorou de 1998 a 2006 e garantiu o repasse de R\$ 35,5 bilhões. O segundo entrou em vigor em 2007 por meio da Emenda Constitucional nº 53, com o repasse total (União, estados e municípios) de R\$ 147,2

bilhões nos três primeiros anos (Ministério da Educação, s.d.). Enquanto o Fundef foi direcionado especificamente ao ensino fundamental, o Fundeb ampliou o seu alcance para a educação infantil e o ensino médio. O que se pode observar na Figura 1 é que de fato o ensino básico recebeu mais. E qual foi o impacto disto na educação?

Verificando os resultados de proficiência escolar da Figura 2, nota-se que o aumento no orçamento da educação coincidiu com a melhoria dos anos iniciais do ensino fundamental, tanto em língua portuguesa como em matemática. O mesmo não se pode dizer dos anos finais. Os maiores orçamentos do ensino fundamental parecem não terem surtido grandes mudanças além da melhoria do fluxo escolar. Aparentemente, o problema do ensino fundamental brasileiro não está na falta de recursos ou, pelo menos, não se limita a ele. Ramos (2009) fez uma breve comparação entre o Fundef e o Fundeb e concluiu que permanecia o problema de como os recursos eram alocados. Oliveira e Teixeira (2008) já criticavam o Fundeb por alocar recursos sem se pautar em indicadores de qualidade, a despeito da discussão em torno deste fundo ter sido com ênfase no fomento de uma educação de qualidade.

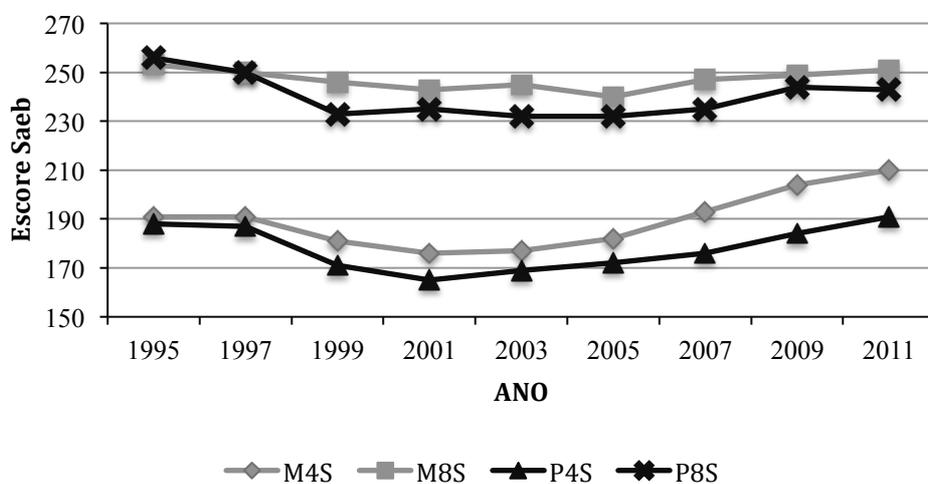


Figura 2. Média do escore na escala Saeb do desempenho escolar das séries iniciais (4S) e finais (8S), para língua portuguesa (P) e matemática (M) em escolas urbanas, Brasil, 1995 a 2011. Adaptado de “Resultados Saeb/Prova Brasil 2011”, por Inep, 2011b, recuperado de http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/resultados/SAEB1995_2005.pdf, e adaptado de “Estatísticas do Ideb: Brasil”, por Inep, 2013, recuperado de http://download.inep.gov.br/educacao_basica/porta_l_ideb/planilhas_para_download/divulgacao-brasil-2011.xls

Essa relação entre gastos em educação e a sua melhoria não se mostra linear. Este aspecto já foi observado por Amaral (2012) ao estudar os gastos de outros países com educação. Não necessariamente os países que tinham maiores orçamentos em educação encabeçavam a lista dos países com melhor pontuação PISA. As discussões de Amaral apontaram para a necessidade de ampliar o orçamento brasileiro para 10% do PIB a fim de elevar o valor gasto por aluno para US\$ 2.398,00, o que é ainda inferior ao valor médio dos 34 países listados por ele, de US\$ 4.456,00. Este percentual de 10% consta no relatório do Plano Nacional de Educação 2011-2020, aprovado no dia 28 de maio de 2013 na Comissão de Assuntos Econômicos (CAE) do Senado. O projeto vai ser ainda analisado pela Comissão de Constituição, Justiça e Cidadania (CCJ) e pela Comissão de Educação (CE) antes de ser votado em plenário no Congresso Nacional.

Alguns autores apontam como a causa do problema da educação o descompasso entre as políticas públicas de educação e a realidade educacional brasileira (Santos, 2002), ou mesmo a falta de políticas públicas (Felicio & Fernandes, 2005), ou a políticas públicas que priorizaram infraestrutura em detrimento do salário dos docentes e do investimento em sua formação (Mendes, 2001), ou uma combinação de elementos como o sistema centralizado de avaliação, a municipalização da educação fundamental, principalmente dos anos iniciais, e a sistemática de financiamento por meio do Fundef (Arelaro, 2005).

O consenso é de que o ensino fundamental não vai bem e as causas são diversas, mas sempre relacionadas com as políticas públicas para a área. Não necessariamente políticas para garantir maiores orçamentos para a educação, mas em como o orçamento é empregado.

Supondo que os recursos fossem alocados corretamente, ainda assim persistiria um problema anterior. A infraestrutura e a qualificação dos professores serão inúteis se as crianças e adolescentes em idade escolar não estiverem presentes nas salas de aula. Assim surgiu o Programa Bolsa Escola.

2.3 O Programa Bolsa Escola

A suposição inicial para um trabalho que se propõe a estudar a relação entre educação e desenvolvimento é de se debruçar sobre programas ou ações de

intervenção que atuem diretamente em educação. Uma escolha natural é o estudo do Programa Bolsa Escola (PBE).

Waiselfisz et al. (1998) fazem um breve histórico do surgimento do PBE. O programa nasceu no Núcleo de Estudos do Brasil Contemporâneo, da Universidade de Brasília (UnB), quando então o seu reitor, Cristovam Buarque, propôs a Renda Mínima Escolar (RME). Em 1994, Cristovam Buarque se tornou governador do Distrito Federal (DF) e implementou o RME. Este programa é transformado em duas ações distintas: o Programa Bolsa Familiar para a Educação (Programa Bolsa Escola) e o Poupança-Escola. Em 1997, os dois programas atenderam a 22.493 famílias e 44.382 crianças e adolescentes entre 7 e 14 anos, com recursos de 1% do orçamento total do DF.

Segundo a Secretaria de Estado de Educação do Distrito Federal (1997), o objetivo do PBE era retirar os menores das ruas, colocá-los na escola e prover uma renda mínima à família. Com isto, esperava-se contornar o problema da exclusão social, da criminalidade juvenil, do trabalho infantil e da miséria. Porém, a participação no programa era condicionada. O programa definia uma faixa etária de elegibilidade dos filhos das famílias beneficiadas, atendia somente a crianças matriculadas em escola pública, definia um teto de renda, condicionava os pais à busca de emprego, no caso de estarem desempregados, e se aplicava somente a residentes do DF. Para o PBE, o aluno não precisava atingir qualquer nível de desempenho escolar. Bastava o comparecimento na escola. Entretanto, o Poupança-Escola tinha como critério principal o desempenho escolar. O indivíduo era elegível ao depósito anual de um salário mínimo em poupança no Banco de Brasília (BRB) caso fosse aprovado na escola. Além disso, a retirada dos recursos só era possível se ele concluísse a 4ª série (primeiro saque), a 8ª série (segundo saque) e o ensino médio (terceiro saque).

Num primeiro momento, o programa não cobriu todo o DF. Para a escolha dos participantes, foram elaborados critérios de prioridade de maneira que as crianças e os adolescentes em pior situação fossem escolhidos. Os indivíduos eram: aqueles em programas de proteção; os que cumpriram medidas socioeducativas; aqueles com dependentes idosos ou portadores de deficiência; os desnutridos; os de famílias com maior número de filhos abaixo dos 14 anos; e os que residissem nas áreas mais carentes do DF. Estes aspectos denotam a escolha de um público com poucas chances

de sucesso em termos de desempenho escolar o que, num primeiro momento, poderia ter representado um risco para a continuidade do programa.

Segundo Glewwe e Kassouf (2012), o PBE se diferenciava dos outros programas por ser um programa não aleatório, i.e., escolhia os grupos em que intervia. Eles o consideraram o maior programa de transferência de renda (PTR) no mundo.

Mas o PBE foi bem sucedido? Glewwe e Kassouf (2012) verificaram que todos os PTR implantados na América Latina melhoraram o desempenho escolar. Do ponto de vista de renda, a avaliação se mostrou mais complicada. Segundo Pochmann (2010), não existem dados de acompanhamento dos beneficiários do programa para todo o seu período de vigência. A solução para o problema proposta por Glewwe e Kassouf (2012) foi estudar o PBE a partir de dados do censo escolar de 1998 a 2005. Porém, fazem uma série de ressalvas: (1) a implantação do programa nos municípios foi feito de maneira gradual ao longo dos anos; (2) algumas escolas foram fechadas em função de avaliações negativas dos seus resultados, desaparecendo da base de dados; (3) dada a desinformação sobre as responsabilidades do programa, em alguns municípios, a escolha dos alunos foi feita em base geográfica e não nos critérios originais de elegibilidade estabelecidos pelo programa; e (4) limitações no orçamento dos municípios determinaram a participação de apenas uma parcela dos alunos elegíveis. Além disso, o dado de quem participou do PBE não é uma porcentagem de alunos da escola, mas apenas uma variável categórica de adesão da escola ao programa. Para contornar parcialmente o problema, os autores utilizaram etnia como um estimador da participação no programa. Apesar desta solução para o problema de acompanhamento dos beneficiários, Glewwe e Kassouf (2012) não foram capazes de verificar o retorno do investimento do programa. A justificativa deles foi que não foi possível identificar qual proporção das melhorias observadas foram resultado só do PBE, uma vez que os indivíduos estão sujeitos ao efeito de diferentes programas sociais.

A partir de 2003, o PBE passou a fazer parte do Programa Bolsa Família, uma dentre várias ações da estratégia “Fome Zero”.

2.4 O Programa Bolsa Família

Diferentemente do Programa Bolsa Escola (PBE), o Programa Bolsa Família (PBF) tem por objetivo combater a fome. Ele é fruto da reunião de vários outros programas de transferência anteriores mantidos em diferentes Ministérios, a citar: Bolsa Escola (Ministério da Educação), Auxílio Gás (Ministério das Minas e Energia), Bolsa Alimentação (Ministério da Saúde) e Cartão Alimentação (Ministério Extraordinário de Segurança Alimentar). Ao incorporar outros programas de transferência de renda, também incorporou as suas condicionalidades. Para uma família receber os recursos do PBF, ela deve atender a três conjuntos de condicionalidades:

a) Saúde:

- no caso da criança: manter atualizado o calendário de vacinação, fazer o monitoramento do peso e do estado de saúde;
- no caso da gestante e da puérpera: acompanhamento pré-natal, acompanhamento pós-parto e participação em atividades educativas sobre aleitamento materno e alimentação saudável.

b) Educação: mínimo de 85% de frequência escolar das crianças e adolescentes entre 6 e 15 anos;

c) Renda: são elegíveis as famílias em situação de pobreza (renda mensal por pessoa de R\$ 70 a R\$ 140) ou de extrema pobreza (menor que R\$ 70).

Estas condicionalidades combinadas fizeram do PBF um programa especial. Elas permitiram precisão na aplicação do dinheiro público (Lindert, Linder, Hobbs & Brière, 2007), aspecto também presente no PBE. Estas três condicionalidades trazem também subjacentes uma lógica de desenvolvimento. Combater a fome por meio da transferência de renda, monitorar a saúde e frequentar obrigatoriamente a escola fazem parte de uma estratégia de intervenção com clara inspiração na teoria do capital humano de Schultz (1961).

O PBF é uma das ações do governo no Eixo 1 de articulação do Fome Zero. Os eixos articuladores são: (1) acesso à alimentação; (2) geração de renda; (3) fortalecimento da agricultura familiar; e (4) articulação, mobilização e controle social (Presidência da República Federativa do Brasil, s.d.). A Tabela 1 apresenta o nome dos vários programas e ações em cada eixo.

Tabela 1

Eixos articuladores do “Fome Zero” e seus respectivos programas e ações.

Eixo 1 Alimentação	Eixo 2 Renda	Eixo 3 Agricultura Familiar	Eixo 4 Organização Social
<ul style="list-style-type: none"> • Bolsa Família 	<ul style="list-style-type: none"> • Programa Nacional de Fortalecimento da Agricultura Familiar (Pronaf) 	<ul style="list-style-type: none"> • Qualificação social e profissional 	<ul style="list-style-type: none"> • Casa das Famílias (Centro de Referência de Assistência Social – CRAS)
<ul style="list-style-type: none"> • Alimentação Escolar (PNAE) 	<ul style="list-style-type: none"> • Garantia-Safra 	<ul style="list-style-type: none"> • Economia solidária e inclusão produtiva 	<ul style="list-style-type: none"> • Mobilização social e educação cidadã
<ul style="list-style-type: none"> • Alimentos a grupos populacionais específicos 	<ul style="list-style-type: none"> • Seguro da Agricultura Familiar 	<ul style="list-style-type: none"> • Consórcio de Segurança Alimentar e Desenvolvimento Local (Consad) 	<ul style="list-style-type: none"> • Capacitação de agentes públicos e sociais
<ul style="list-style-type: none"> • Cisternas 	<ul style="list-style-type: none"> • Programa de Aquisição de Alimentos da Agricultura Familiar (PAA) 	<ul style="list-style-type: none"> • Organização produtiva de comunidades 	<ul style="list-style-type: none"> • Mutirões e doações
<ul style="list-style-type: none"> • Restaurantes populares 		<ul style="list-style-type: none"> • Desenvolvimento de cooperativas de catadores 	<ul style="list-style-type: none"> • Parcerias com empresas e entidades
<ul style="list-style-type: none"> • Bancos de alimentos 		<ul style="list-style-type: none"> • Microcrédito produtivo orientado 	<ul style="list-style-type: none"> • Controle das ações do Fome Zero pela sociedade (Controle Social)
<ul style="list-style-type: none"> • Agricultura urbana / hortas comunitárias 			
<ul style="list-style-type: none"> • Sistema de vigilância alimentar e nutricional (Sisvan) 			
<ul style="list-style-type: none"> • Distribuição de Vitamina A (Vitamina A+) 			
<ul style="list-style-type: none"> • Distribuição de ferro (Saúde de Ferro) 			
<ul style="list-style-type: none"> • Alimentação e nutrição de povos indígenas 			
<ul style="list-style-type: none"> • Educação alimentar, nutricional e para consumo 			
<ul style="list-style-type: none"> • Alimentação Saudável / Promoção de Hábitos Saudáveis 			
<ul style="list-style-type: none"> • Alimentação do Trabalhador (PAT) 			

Nota. Adaptado de “Fome Zero: programas e ações”, por Presidência da República Federativa do Brasil, s.d, recuperado de <http://www.fomezero.gov.br/programas-e-acoas>

Destaca-se a concentração de programas e ações no Eixo 1, que se preocupa com o acesso aos alimentos, e a predominância de atividades que envolvem a agricultura familiar. Esta ênfase na agricultura familiar é compreensível pelo fato de $\frac{1}{4}$ da população rural se encontrar em situação de “extrema pobreza”, como já mencionado. Esta ênfase no meio rural ficou mais evidente com o lançamento, em 20 de março de 2012, do Programa Nacional de Educação no Campo (Pronacampo). Um programa para lidar com 76 mil escolas rurais existentes, 6,2 milhões de matrículas e 342 mil professores (Ministério da Educação, 2012).

Mais recentemente, dois novos benefícios foram incorporados ao PBF. O Benefício Variável vinculado ao Adolescente (BVJ) estende o benefício do PBF ao jovens com 16 e 17 anos. No caso de crianças de zero a 6 anos, existe o Benefício para a Superação da Extrema Pobreza na Primeira Infância (BSP).

Em oito anos de PBF, o programa gastou 76 bilhões de reais, esteve presente em 5.560 municípios brasileiros e beneficiou 11,1 milhões de famílias, o que correspondeu a 45 milhões de pessoas (Ministério do Desenvolvimento Social e Combate à Fome, 2011). Por ano, o PBF aplicou o equivalente a 0,2% do PIB, enquanto o orçamento público em educação foi da ordem de 5,8% do PIB, segundo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep, 2011a). A comparação é puramente financeira e ignora a complexidade das estruturas de custos do PBF e da educação. De qualquer forma, a comparação permite constatar que o gasto com o PBF é relativamente pequeno, considerando o seu impacto sobre $\frac{1}{4}$ da população total do Brasil.

O número de ações e programas dentro da estratégia do Fome Zero evidencia o problema levantado anteriormente por Glewwe e Kassouf (2012). Como isolar o efeito de apenas um programa? A este problema, junta-se agora o tamanho do orçamento. Com um gasto no PBF dez vezes menor que o gasto total em educação, a tarefa de verificar o impacto do programa sobre a fome parece ser ainda mais difícil.

Para tentar determinar o tamanho dessa dificuldade, serão apresentados alguns dados sobre o orçamento do PBF. Na Figura 3 está a distribuição do investimento por região. Chama a atenção o resultado da região Sudeste. A despeito de ser a região de maior desenvolvimento econômico, ela recebeu o segundo maior montante de fundos. Mas isto pode ser explicado pela distribuição da pobreza na população. Como visto anteriormente, a divisão da pobreza em termos populacionais é quase igual entre meio

urbano e meio rural. O Sudeste é a região com maior concentração de população urbana, o que justificaria a sua segunda posição na Figura 3.

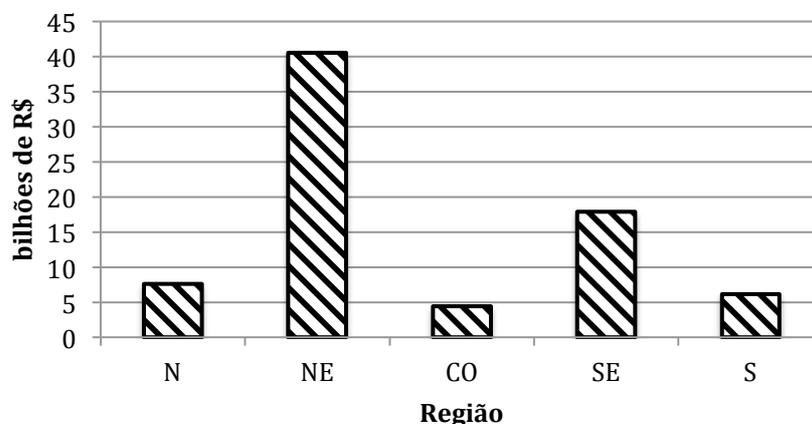


Figura 3. Valor de investimento do Programa Bolsa Família por região do Brasil, 2003 a 2010. Adaptado de “Bolsa Família completa 8 anos com investimentos de R\$ 76 bilhões”, por Ministério do Desenvolvimento Social e Combate à Fome, 2011. Recuperado de <http://www.mds.gov.br/saladeimprensa/noticias/2011/outubro/bolsa-familia-completa-8-anos-com-investimentos-de-r-76-bilhoes>

Para ratificar essa informação, a Figura 4 apresenta o número médio de famílias cujo rendimento mensal nominal é de até $\frac{1}{4}$ do salário mínimo (à época, R\$ 510), por unidade federal, para o ano de 2010. Neste mapa, nota-se que São Paulo e Minas Gerais estão na mesma faixa da Bahia, Ceará, Maranhão e Pará. Estes estados apresentam entre 47 a 86 mil famílias nessa faixa de renda. Rio de Janeiro está na mesma faixa de Piauí, Paraíba e Pernambuco, entre 21 a 45 mil famílias. De fato, os estados do Sudeste, com exceção de Espírito Santo, apresentam quantidades semelhantes de famílias em situação de pobreza e extrema pobreza quando comparado com alguns estados do Norte e Nordeste.

Estes dados confirmam que o PBF de fato está atingindo as regiões mais pobres do país. Portanto, pode-se supor que, apesar de relativamente pouco o valor gasto com o PBF, ele esteja sendo utilizado de forma eficiente no combate à fome. Mas como discernir os impactos do PBF dos demais programas e ações?

A resposta talvez esteja nas próprias condicionalidades do PBF. Elas são baseadas em um modelo teórico de ação. Este modelo estabelece que a melhoria da educação dos membros da família, a exigência do comparecimento aos postos de

saúde e o provimento de renda implicam na diminuição da fome e na sua não perpetuação (bem-estar). Assim, verificar o impacto do PBF não se definiria apenas por seus impactos diretos sobre a fome, mas também pela mensuração da educação, da saúde e da renda das famílias e a sua relação com a pobreza.

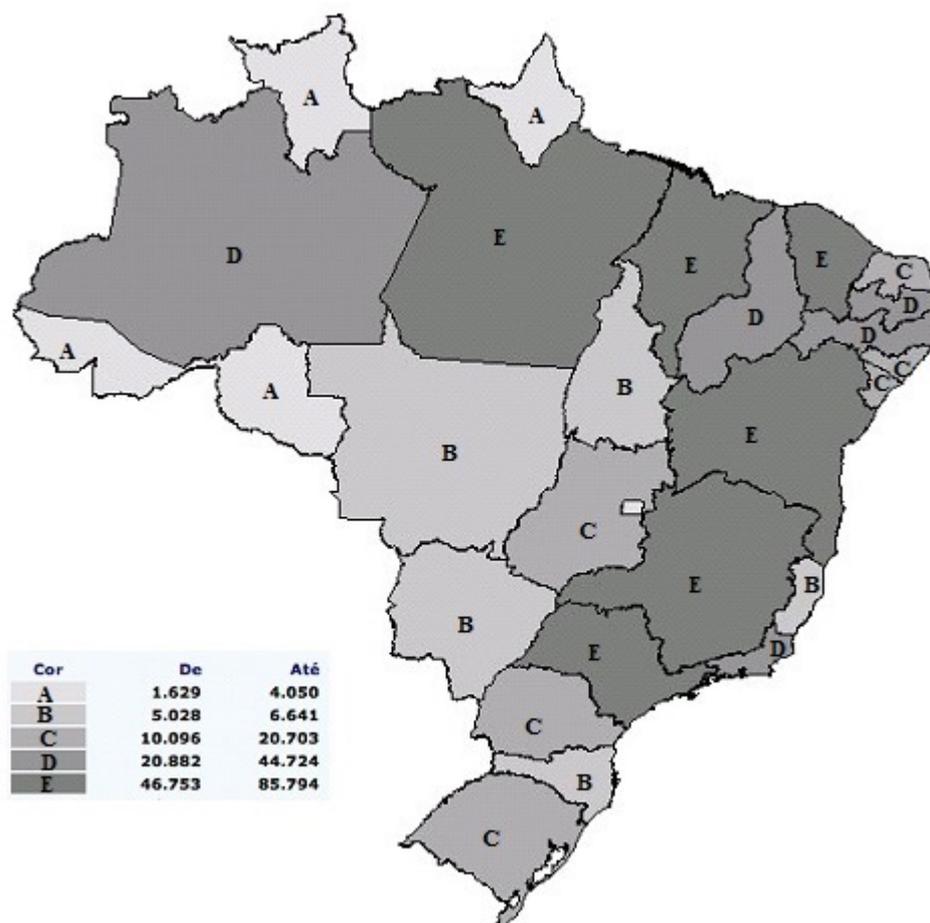


Figura 4. Número de famílias conviventes residentes em domicílios particulares cujo rendimento nominal mensal familiar per capita é de até $\frac{1}{4}$ de salário mínimo (salário mínimo = R\$ 510), por estado, Brasil, 2010. Elaborado por meio do Sistema IBGE de Recuperação Automática (Sidra) a partir de dados do Censo Demográfico de 2010. Recuperado de <http://www.sidra.ibge.gov.br/bda/tabela/listabl3.asp?c=3520&n=0&u=0&z=t&o=25&i=P>

Este modelo subjacente ao PBF permite estabelecer uma estratégia de análise deste programa. Este trabalho pretende analisar o desenvolvimento econômico brasileiro a partir deste *framework* de ação do PBF.

2.5 Avaliação do Programa Bolsa Família

O Programa Bolsa Família (PBF) já foi objeto de vários estudos. Um dos mais recentes que se propõe a estudar o período de 2003 a 2010 é o de Castro e Modesto (2010a) e Castro e Modesto (2010b). Nos dois volumes da obra, os autores partem de aspectos conceituais do programa até o seu futuro, sem deixar de passar pela sua avaliação. Para esta pesquisa, são de maior interesse os capítulos 1 e 2 do segundo volume. O primeiro escrito por Soares, Souza, Osório e Silveira (2010) e o segundo por Silveira Neto (2010). Ambos partem de dados da Pnad. O primeiro trabalho estuda o impacto do PBF sobre a pobreza e desigualdade e utiliza a mesma estratégia de decomposição fatorial do Índice de Gini utilizada por Barros, Carvalho e Franco, (2006) e Helfand, Rocha e Vinhais (2009). A conclusão a que chegaram é que o PBF pode ser considerado grande em termos de número famílias cobertas pelo programa, modesto em termos de recursos mobilizado e grande em termos de efeito sobre a desigualdade na distribuição de renda. Já o trabalho de Silveira Neto (2010) procura estudar o impacto do PBF sobre a frequência escolar. Ele utiliza a técnica de pareamento por *propensity score* para definir os grupos de participantes e não participantes do programa. A conclusão a que o autor chega é de que os impactos do PBF sobre a frequência foram maiores no Nordeste do que no Sudeste. Ele atribui isto ao maior atraso escolar no Nordeste.

A abordagem desta pesquisa é quantitativa. Porém, não ignorando a existência de abordagens complementares, cita-se alguns trabalhos em uma perspectiva qualitativa de avaliação do PBF, seguindo uma vertente sociológica. São exemplos os trabalhos de Silva (2008), Weissheimer (2010) e Silva (2011). O primeiro trabalho é mais específico, estudando os estados de Maranhão e Piauí. Os dois outros são de abrangência nacional.

Independente da abordagem adotada, existe uma questão que precede a escolha da abordagem. É primordial definir o conceito de avaliação. Neste trabalho, o termo “avaliação” seguirá a definição de Fitzpatrick, Sanders e Worthen (2004). Para eles, avaliar significa identificar, clarificar e aplicar critérios defensáveis para determinar o valor do objeto avaliado frente a estes critérios definidos. Os autores comentam que a principal dificuldade no processo de “examinar e julgar” (avaliar) é a definição e a comunicação dos critérios usados pelos avaliadores. Eles recomendam três etapas de questionamento e julgamento: 1) determinar os padrões de julgamento da qualidade e

decidir se tais padrões serão absolutos ou relativos; 2) coletar informação relevante; e 3) aplicar os padrões a fim de determinar a qualidade, a utilidade, a efetividade ou a significância. Por esta definição, a constituição do critério utilizado para valorar é o maior problema da avaliação e não a “valoração” em si.

O primeiro passo da avaliação é examinar (analisar). Esta pesquisa se propõe a analisar a pobreza na perspectiva do PBF. Ao fazer isso, não reduz a relação entre pobreza e educação a uma relação de proporcionalidade entre melhoria e gastos públicos. Isto diferencia esta pesquisa de outros trabalhos onde se procura quantificar proporções entre educação e renda, como faz Barros e Mendonça (1995), Barbosa Filho e Pessoa (2008) ou Curi e Menezes Filho (2009).

O problema que agora se impõe é como operacionalizar esta análise. A seguir serão apresentadas algumas estratégias e técnicas de análise de programas sociais.

2.5.1 Análises descritiva e relacional

É mais comum observar em literatura uma postura descritiva das análises. Esta postura pode ser tanto a partir da visão de quem avalia, como do ponto de vista do avaliado. Um exemplo de avaliação pela visão do avaliador é o trabalho de Araújo e Luzio (2005). Eles realizaram uma avaliação quantitativa da educação básica brasileira a partir da comparação de dados do Saeb de 2003 com anos anteriores. Não estudaram as causas dos baixos resultados observados de desempenho escolar. Limitaram-se a constatar a “má qualidade” do ensino nas séries iniciais.

Um exemplo de avaliação a partir dos sujeitos avaliados é o trabalho de Waiselfisz et al. (1998). Eles procederam à avaliação do Programa Bolsa Escola (PBE) no DF por duas abordagens complementares as quais chamaram de extensiva e compreensiva. Na extensiva, por meio de método indutivo, procuraram inferir o universo a partir de subamostras da população. Na compreensiva, por meio de grupos focais, procuraram recompor a morfologia sociocultural dos alunos. Novamente, os resultados obtidos se limitaram a constatar a estigmatização de alguns alunos bolsistas por professores e por alunos não bolsistas e a desinformação de algumas famílias sobre os critérios de entrada, de permanência no programa e de seu funcionamento.

O que se percebe nos dois trabalhos é que a análise descritiva estabelece as bases empíricas para a discussão, a problematização e a elaboração de hipóteses.

Porém, elas não permitem compreender os mecanismos que levam aos resultados observados.

Um complemento à perspectiva descritiva é a perspectiva relacional. Ainda na avaliação da educação, a partir de dados da Pnad de 1996, Machado e Gonzaga (2007) procederam a uma análise inferencial por meio de modelos econométricos baseados em três grupos de variáveis instrumentais. O objetivo foi mostrar o efeito da renda familiar e do nível educacional dos pais sobre a distorção idade-série (DIS). Chegaram a resultados descritivos tais como constatar que a DIS foi mais acentuada nos meninos, não brancos e amarelos, moradores de regiões de menor oferta de escolas, de famílias grandes, com pais com menor nível de instrução e menor renda per capita. O modelo econométrico permitiu ainda inferir que um programa de transferência de renda parece servir como um redutor significativo da DIS.

As análises descritiva e relacional já foram empregadas no PBF. Um trabalho de avaliação do PBF na linha descritiva do ponto de vista do avaliador pode ser visto em Lindert et al. (2007). Este trabalho documentou a experiência brasileira do PBF para o Banco Mundial. Os autores ressaltaram a complexidade do programa ao integrar quatro outros programas de transferência de renda (PTR), o seu tamanho e o seu rápido crescimento, a precisão do PBF em atingir as famílias-alvo por meio do Cadastro Único para Programas Sociais (CadÚnico), a redução da desigualdade e da extrema pobreza, a unificação horizontal e vertical da política social brasileira e o seu uso como um laboratório natural de inovações.

Na perspectiva descritiva do avaliado, existe um trabalho recente de análise realizado pelo próprio Ministério de Desenvolvimento Social e Combate à Fome (2010). A análise consistiu de três avaliações distintas: (1) do impacto epidemiológico e social em um município da Bahia; (2) da implementação do PBF em 269 municípios; e (3) do impacto sobre o bem-estar de 15.426 famílias em 23 estados e no Distrito Federal (Segunda Rodada). As avaliações foram feitas por meio de entrevistas das famílias de participantes e não participantes, coletas de sangue das crianças acompanhadas, medidas antropométricas das crianças, coleta de dados junto aos gestores municipais do programa e dados do CadÚnico. Entre os resultados, destacam-se: um “efeito” maior do PBF nas faixas etárias mais baixas nos anos iniciais de implementação do programa; “efeito” do PBF sobre o papel da mulher

como gestora da família e melhoria de sua autoestima; aumento no consumo de proteína animal; disponibilidade de acesso e utilização de serviços de saúde, mas que não se refletiram em aumento da busca desses serviços pelos indivíduos portadores de doenças ou queixosos; pior gestão local do PBF observada na região Nordeste, principalmente nos municípios com menos de 50 mil habitantes; e quase 100% de acompanhamento da condicionalidade de vacinação nos municípios estudados.

Estes dois esforços descritivos do PBF trouxeram indagações sobre as particularidades regionais. Estas diferenças regionais foram estudadas a partir da perspectiva relacional por Helfand et al. (2009). Eles estudaram a diferença entre o público rural e o público urbano. Eles decompueram o Índice de Gini, um indicador de concentração de renda, e detectaram que 43% do declínio da pobreza rural decorreu não da diminuição da desigualdade, mas do crescimento da renda. Renda oriunda de rendimentos não relacionados ao trabalho, mas da previdência, das pensões e da categoria “outros rendimentos” da base de dados da Pnad, a qual inclui o Bolsa Família.

Higgins (2012) também se interessou pelas diferenças regionais brasileiras e o efeito delas sobre o impacto do PBF. Ele propôs uma avaliação do PBF levando em conta as diferenças de preços regionais. Como o PBF é um programa que tem a renda da família como uma das condicionalidades e o custo de vida varia em cada município estudado, o efeito da transferência de renda é variável, uma vez que o benefício é o mesmo para todo o Brasil. O autor utilizou uma regressão *probit* para a análise. Os resultados do seu estudo indicam que as famílias pobres em áreas rurais se beneficiam mais do PBF do que as famílias pobres em áreas urbanas.

No presente trabalho, a avaliação do PBF será realizada em uma perspectiva relacional.

2.5.2 Causalidade

Ainda que brevemente, é importante salientar que estudos relacionais permitem estabelecer a dependência entre duas ou mais variáveis, mas não a direção da relação, ou seja, a causalidade.

As técnicas gerais de regressão multivariada não determinam causalidade, mas a dependência entre variáveis. Algumas técnicas multivariadas específicas como a de Modelagem por Equações Estruturais (MEE) permitem testar causalidades. Trata-se

de uma causalidade determinada por um modelo teórico a priori e não uma causalidade estabelecida a partir da análise dos dados empíricos.

Se não é possível atribuir toda a causalidade de uma variável por outra por meio da regressão, é possível determinar parte dela. Em séries temporais econômicas é comum a aplicação da técnica conhecida como Causalidade de Granger (Granger, 1969). Esta técnica é um teste de hipótese da anterioridade do evento de uma variável em relação ao evento de outra por meio de um modelo de auto-regressão.

A questão de causa e efeito é complexa e tem sido debatida na filosofia por séculos. Para fins desta pesquisa, a relação causal, ou a relação entre causa e efeito, ou causalidade, será entendida como definida por Shaddish, Cook e Campbell (2002). Eles citam as características da causalidade propostas por John Stuart Mill: (1) a causa tem que preceder o efeito; (2) a causa tem que estar relacionada com o efeito; e (3) não se pode encontrar outra explicação alternativa plausível para o efeito que não seja a causa.

2.5.3 Regressão hierárquica

A regressão hierárquica é também conhecida como Análise Multinível (AMN) ou modelo linear hierárquico. A grande vantagem desta técnica é poder considerar variáveis em outros níveis de análise que podem se relacionar com as variáveis do nível em estudo. Por isso da denominação de “hierárquica”. Esta característica se adapta bem aos problemas sociais, onde existem diferentes níveis de organização. Um país contém cidades, cidades contêm organizações, organizações contêm grupos sociais e assim por diante.

A aplicação da AMN para o estudo em educação é bastante difundida na literatura. Barbosa e Fernandes (2000) avaliaram um modelo multinível a partir de dados do Saeb do ano de 1997. Utilizando a Modelagem por Equações Estruturais multinível (MEE), Nasser e Hagtvvet (2006) estudaram como as características do aluno, do professor e do curso afetavam a avaliação do aluno. Andrade e Laros (2007) estudaram fatores associados ao desempenho escolar com dados do Saeb de 2001 por meio de AMN. Bagaka's (2011) estudou características dos professores e práticas que poderiam aumentar a auto-eficácia de estudantes secundaristas no Quênia. Liu (2012) utilizou modelos lineares hierárquicos para estudar os fatores que impactam a avaliação dos estudantes em cursos à distância.

A AMN é mais comumente utilizada em análises transversais, i.e., desenvolve as relações entre as variáveis e os níveis num determinado ponto no tempo. Porém, ela pode ser utilizada também em análises longitudinais, com séries de observações. A análise longitudinal é também conhecida como regressão de medidas repetidas. A modificação entre o modelo transversal e o modelo longitudinal é pequena. No modelo transversal, o indivíduo que antes era definido como nível 1 (o mais baixo) passa a ser considerado nível 2. As observações ao longo do tempo de um dado indivíduo são definidas como nível 1. Em outras palavras, os dados longitudinais podem ser vistos como dados multinível com medidas repetidas aninhadas em cada indivíduo (Hox, 2010).

A regressão hierárquica longitudinal foi usada por Alves e Soares (2007) para determinar o efeito das escolas e dos alunos (gênero, atraso escolar e nível socioeconômico) sobre o desempenho escolar dos alunos. Empregaram técnicas de Teoria de Resposta ao Item para quantificação do desempenho escolar e associaram uma pesquisa qualitativa, basicamente 22 entrevistas com profissionais da escola e 38 entrevistas com famílias dos alunos.

Na presente pesquisa, a regressão hierárquica será empregada em um estudo longitudinal por meio de medidas repetidas.

2.5.4 Dados

Fitzpatrick et al. (2004) mencionam a coleta de informações como a segunda etapa do processo de avaliação. Obter informações envolve o processo de coletar dados e combiná-los de maneira que tenham alguma utilidade para o avaliador. Para a avaliação de programas sociais, um caminho desejável é poder coletar os seus próprios dados. A razão é a necessidade de controle experimental. É pelo controle que se constrói um “bom” contrafactual. Segundo Khandker, Koolwal e Samad (2010), este é o maior desafio de uma avaliação de impactos de programas sociais. A fim de contornar o problema, foram desenvolvidas várias técnicas de pareamento (*matching*) para que se pudesse constituir um contrafactual representativo (grupo controle) e comparável. São exemplos de variantes da avaliação de impacto a avaliação randomizada, pareamento por escore de propensão (*propensity score*), métodos de diferenças-duplas, uso de variáveis instrumentais e abordagens de regressão de dados descontínuos e “pipeline” (Khandker et al., 2010). Entretanto, permanece a questão:

como submeter o mesmo sujeito ao tratamento e ao controle? No caso de fenômenos sociais, tais métodos são sempre aproximações da realidade.

Mesmo que fosse possível submeter o mesmo indivíduo às duas situações, o controle experimental está associado a custos elevados, não só financeiros, mas de recursos técnicos, de energia, de material e de tempo. A fim de contornar o problema de custos, pode-se optar por usar fontes secundárias de dados. Isto não diminui o custo da coleta em si, mas o dilui. A principal desvantagem de usar dados secundários é exatamente o menor controle sobre eles. Isto traz duas consequências comuns: (1) adaptação dos métodos de análise às limitações impostas pelos dados; e (2) eventual correção de erros inerentes aos dados, uma vez que os dados não podem ser substituídos por nova coleta.

Como exemplo da adaptação de métodos, alguns trabalhos que estudaram o Programa Bolsa Família (PBF) lançaram mão de técnicas de desagregação de dados da Pnad (Barros, Carvalho & Franco, 2006) e de agregação (Soares, Osório, Soares, Medeiros & Zepeda, 2007) para inferir dados sobre os beneficiários.

Quanto à correção, um exemplo comum é a discrepância de dados observada durante o cruzamento de dados. Soares, Soares, Medeiros e Osório (2006) apresentaram uma comparação entre os dados de número de beneficiários do PBF informados pela Pnad e os informados pelos registros administrativos do Ministério do Desenvolvimento Social e Combate à Pobreza (MDS). Os dados da Pnad apresentam 58% menos beneficiários do que os registros administrativos do MDS.

A despeito destes problemas, as bases secundárias proporcionam uma fonte rápida e igualmente rica de dados. Quando se fala de bases secundárias de dados oficiais, existe ainda o aspecto de “legalidade”, dando um caráter representativo importante aos resultados decorrentes.

Usar uma base secundária é também uma maneira de aproveitar investimentos já realizados na sua coleta. No caso brasileiro, a quantidade de dados oficiais disponíveis é enorme. Com a diretriz de transparência do governo federal, os vários órgãos buscam disponibilizar as suas próprias bases para o grande público. Um caso importante e útil para esta pesquisa é o sistema da Secretaria de Avaliação e Gestão da Informação (Sagi), ligado ao Ministério do Desenvolvimento Social e Combate à

Pobreza (MDS). Uma das ferramentas disponibilizada por eles é o TabSocial. Os dados do PBF por município estão disponíveis nele (<http://www.mds.gov.br/sagi>).

3. Objetivos

3.1 Geral

O presente trabalho tem por objetivo avaliar o impacto do Programa Bolsa Família (PBF) sobre a pobreza.

3.2 Específicos

- Desenvolver um modelo de componentes com as variáveis do estudo;
- Desenvolver um modelo de regressão hierárquica com todos os componentes do modelo de componentes obtido;
- Desenvolver um modelo de regressão hierárquica antes da intervenção do PBF;
- Desenvolver um modelo de regressão hierárquica depois da intervenção do PBF;
- Desenvolver um modelo de regressão hierárquica para cada componente individualmente.

4. Material e Métodos

Esta pesquisa procurou avaliar o impacto do Programa Bolsa Família (PBF) por meio de análise multinível de medidas repetidas de 68 variáveis distribuídas em seis categorias. Três categorias visam cobrir os três grandes pilares do programa: (1) educação; (2) saúde e (3) renda. Outras três categorias cobrem aspectos relacionados a: (4) produção agropecuária, pois são nas populações rurais onde se encontram os maiores percentuais de pobreza no Brasil; (5) características demográficas, as quais permitirão ponderar os impactos do PBF e o efeito das variáveis explicativas aqui utilizadas; e (6) indicadores do PBF. Esta pesquisa baseia-se exclusivamente em dados secundários coletados em bases de dados públicas e oficiais, cobrindo o período de 2000 a 2010.

Os dados foram submetidos a uma verificação de pressupostos univariados e remoção de dados atípicos. Em seguida, uma análise de componentes principais foi realizada com o intuito de verificar se o agrupamento de variáveis seguiu o agrupamento em categorias inicialmente proposto.

A partir dos componentes identificados, foram desenvolvidos quatro modelos de análise com diferentes cortes temporais e diferentes combinações de variáveis explicativas. Os dados foram padronizados e então convertidos do formato *wide* (um caso por linha) para o formato *long* de dados (uma variável explicativa por coluna). Eles foram submetidos ao pacote estatístico de análise multinível MLwiN da Universidade de Bristol.

Os pressupostos multivariados foram então analisados. Os modelos foram comparados em termos de ajuste (*fit*) e dos coeficientes de regressão.

4.1 Delineamento do estudo

O uso somente de dados secundários não permite um desenho experimental ou quase-experimental, uma vez que não se tem controle dos dados coletados e de como são coletados. Por serem dados oficiais em nível nacional, eles dizem respeito à população e não a uma amostra dela. Logo, não existe necessidade de randomização para composição de amostras representativas. Entretanto, não se trata de análise descritiva, pois tenta inferir relações de causalidade entre o PBF e os pilares deste programa. Na classificação apresentada por Khandker et al. (2010), o presente estudo é categorizado como um Levantamento Quantitativo de Impacto *Ex-Post* de um programa, o mesmo tipo de método de avaliação utilizado na experiência mexicana PROGRESA.

4.2 Procedimentos

Uma grande vantagem deste estudo é contar com uma base de dados de âmbito nacional, de toda a população brasileira. Porém, esta vantagem vem carregada de um ônus: a baixa qualidade de uma parte considerável dos dados. Existem perdas significativas de dados por problemas de coleta, por mudança dos critérios e métodos de coleta e por erros na compilação, armazenamento e distribuição destes dados.

Todos os dados utilizados provieram de fontes secundárias e oficiais. Os dados econômicos, agropecuários e demográficos foram obtidos junto ao Instituto Brasileiro de Geografia e Estatística (IBGE). Os dados de saúde foram obtidos junto ao Banco de Dados do Sistema Único de Saúde, mantido pelo Departamento de Informática do SUS (Datasus), do Ministério da Saúde. As variáveis educacionais vieram do Sistema Nacional de Avaliação da Educação Básica (Saeb), do Prova Brasil e do Censo

Escolar, todos mantidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Os dados do Programa Bolsa Família (PBF) foram obtidos junto à Secretaria de Avaliação e Gestão da Informação (Sagi) do Ministério do Desenvolvimento Social e Combate à Fome (MDS).

Os dados estão prontamente disponíveis para a pesquisa, mas o seu tratamento antes da realização das análises é moroso. A princípio, existem dados para todos os 5.565 municípios brasileiros nas 68 variáveis aqui listadas para o período de onze anos. Estas dimensões combinadas podem gerar mais de 4,2 milhões de casos possíveis. Apesar dos modelos aqui apresentados não incluírem ao mesmo tempo todas as variáveis para todos os anos, o número de casos efetivamente disponíveis variou de 3.220 a 60.660.

Para tentar minimizar as perdas de casos, foi necessário proceder à modificação dos dados e à sua reorganização, não só com o intuito de melhorar as suas qualidades estatísticas (linearidade, normalidade e homocedasticidade), mas para permitir as inferências necessárias para responder às perguntas de pesquisa.

4.3 Pressupostos estatísticos

A análise multinível (AMN) é considerada menos exigente em termos de pressupostos que os modelos lineares gerais multivariados (MLG). Os critérios de linearidade, normalidade, homocedasticidade e independência das observações são menos críticos na análise, principalmente estes dois últimos. Como a AMN supõe diferentes níveis de organização, onde os níveis superiores agrupam os níveis inferiores, espera-se que exista alguma relação intraclasse e entreclasse. A consequência disso é que há algum grau de dependência entre as variáveis (principalmente as intraclasses) e diferentes variâncias quando lidamos com comparações entreclasses (heteroscedasticidade).

Tabachnick e Fidell (2007) afirmam que normalidade, linearidade e homocedasticidade melhoram a análise estatística, mesmo que a técnica de análise seja robusta à falha de algum desses pressupostos. Porém, na análise multivariada, em alguns casos pode ser impraticável garantir estes pressupostos. É o que ocorre, por exemplo, com a normalidade multivariada. Não existe um teste específico para isso, sendo necessário testar para normalidade cada combinação linear de variáveis.

Dependendo do número de variáveis envolvidas, a quantidade de combinações pode ser enorme.

Também afirma-se que a AMN é robusta para lidar com dados *missing* (Hox, 2010, p. 106). Contudo, os modelos multinível são igualmente exigentes e sensíveis a dados *missing* como os MLG. A diferença é que a AMN permite atribuir interceptos e coeficientes de inclinação para cada sujeito da análise. No caso de medidas repetidas, isto implica em dizer que o importante é o número de medidas e não a regularidade dos intervalos entre as medidas. Se num determinado período são esperados dez eventos, a AMN necessitará de dez eventos, podendo, contudo, existir um ano sem medidas e outro ano com duas medidas. O ano sem medida poderia ser interpretado como um dado *missing* e, assim, afirmar-se que a AMN é “robusta” a dados *missing*.

Em vista destas peculiaridades da AMN, a verificação dos pressupostos foi feita em duas etapas. A primeira etapa, antes da análise dos modelos, verificou a presença de *outliers* e a normalidade univariada, eliminando-se os casos atípicos e corrigindo eventuais assimetrias. A segunda etapa se mistura com a análise de dados em si. Durante a AMN, verificou-se os resíduos dos modelos multinível, conforme sugerido por Hox (2010), a fim de identificar problemas nos pressupostos multivariados tais como dados atípicos, não normalidade, não linearidade e heteroscedasticidade.

4.4 Análise de dados

A análise foi realizada em três partes distintas: (1) análises descritiva e exploratória dos dados; (2) análise multinível das seis categorias de variáveis; e (3) análise multinível de um subconjunto de variáveis antes e depois da implementação do programa.

A primeira parte teve dois propósitos: verificar os pressupostos univariados de análise; e estabelecer um contexto para o estudo ao realizar cortes temporais e espaciais. A segunda parte procurou pela relação entre as seis categorias estudadas e pobreza. Na terceira parte, o intuito foi fazer um corte temporal, definindo um período antes e outro depois da implementação do PBF. Desta forma, foi possível comparar os modelos de cada período entre si a fim de se observar algum tipo de mudança. Em havendo mudança, procurou-se estabelecer sua relação com o PBF.

4.4.1 Análise descritiva e exploratória

Foram calculadas as estatísticas de dispersão e assimetria. Estabeleceu-se como critério para a determinação de valor atípico se o escore z do dado fosse superior a $z \geq 2,58$. A linearidade das variáveis foi verificada por inspeção do gráfico de resíduos padronizados e valores estimados. A normalidade foi verificada pelos valores de assimetria calculados pelo SPSS, versão 18. Valores iguais ou próximos a 0,00, com variação de $\pm 1,00$, foram considerados de distribuição normal. Nos casos em que ocorreram problemas de não normalidade, a variável foi corrigida por transformação dos dados utilizando a transformação proposta por Box-Cox (Osborne, 2010). A transformação de Box-Cox trata da determinação do coeficiente lambda (λ) na Equação 1 ou Equação 2.

$$y_i^\lambda = \frac{(y_i^\lambda - 1)}{\lambda} \quad \text{para } \lambda \neq 0 \quad (1)$$

$$y_i^\lambda = \log_e(y_i) \quad \text{para } \lambda = 0 \quad (2)$$

Onde:

Y_i : variável a ser transformada;

λ : coeficiente da transformação de Box-Cox.

Alguns casos de transformação exemplificando o efeito do coeficiente:

- Se $\lambda = 1,0$: nenhuma transformação é necessária;
- Se $\lambda = 0,5$: transformação raiz quadrada;
- Se $\lambda = 0,0$: transformação logarítmica;
- Se $\lambda = -1,0$: transformação inversa.

Por meio de uma sintaxe do SPSS, gerou-se valores de λ dentro de um intervalo pré-definido e se calculou a respectiva assimetria (*skewness*). O valor de assimetria que mais se aproximou de zero indicou o valor de λ apropriado para a transformação. A vantagem desta estratégia foi determinar com mais precisão o valor de potência à qual a variável a ser transformada seria submetida.

A transformação requereu alguns cuidados. No caso de λ positivo ou igual a zero (logaritmo), os dados foram corrigidos por meio de adição ou subtração de um valor de maneira que o valor mínimo da variável fosse igual a 1. Osborne (2010) afirma que dados próximos ao valor 1,00 permitem melhores efeitos da transformação. No caso de logaritmo, valores entre 0,00 e 1,00 geram resultados negativos, o que não era desejado para esta pesquisa. Em caso de valor negativo de λ (inversa), a ordem dos dados foi ajustada por meio de subtração do dado após transformação, de maneira a tornar o valor mínimo igual a 1,00. Alguns valores mostrados como “0,00” pelo SPSS tiveram que ser corrigidos para zero absoluto, pois tratavam-se de arredondamentos de resultados negativos muito pequenos, o que afeta os cálculos da AMN.

Mesmo com indicação de problemas em alguns dos pressupostos verificados, é preciso considerar que se trata de dados da população, de maneira que se pode ser mais leniente em relação a estes pressupostos. Avaliações multivariadas são baseadas em estimativas, cujos modelos não se sabe a priori. Nesta pesquisa, o critério principal para ajustes ou cortes de casos foi a impossibilidade de obter convergências (matriz definida semipositiva) durante o cálculo dos coeficientes dos modelos.

Por meio de Análise dos Componentes Principais (ACP), com rotação Varimax, obteve-se a matriz de correlação e covariância e a tabela “estrutura” (*pattern*) para análise do agrupamento das variáveis. Estas tabelas permitiram o estudo do nível de multicolinearidade. O objetivo desta análise foi antever eventuais necessidades de combinação ou eliminação de variáveis em função de correlações elevadas e para verificar se as variáveis se agrupavam dentro das seis categorias definidas na pesquisa.

4.4.2 Análise multinível

Este trabalho emprega a análise multinível de medidas repetidas. Nesta configuração, o município se encontra no nível 2 (i) e as suas respectivas medidas em anos diferentes no nível 1 (t). A *Equação 3* representa a equação geral deste modelo.

$$Y_{ti} = \pi_{0i} + \pi_{2i} \cdot X_{ti} + e_{ti} \quad (3)$$

Onde:

Y_{ti} : variável dependente do município i no evento t ;

π : coeficiente de regressão de nível 1;

X_{ti} : variável explicativa do município i no evento t ;

e_{ti} : resíduo do município i no evento t .

Como já mencionado, os municípios brasileiros estão no nível 2. O número de municípios variou entre 5.507 e 5.565 ao longo do período de estudo (Tabela 2).

Tabela 2
Número de municípios em quatro períodos entre 1997 e 2013.

Período	1997 a 2000	2001 a 2004	2005 a 2008	2009 a 2013
Municípios	5.507	5.560	5.564	5.565

O período de análise deste trabalho foi de 2000 a 2010, num total de onze eventos. Ao longo deste horizonte de análise, os modelos foram analisados com intercepto e coeficientes de inclinação fixos. Foram utilizados dados das seis categorias de variáveis: (1) dados educacionais; (2) dados de saúde; (3) dados econômicos; (4) dados agropecuários; (5) dados demográficos e (6) dados do programa PBF. A codificação, categorias e grandezas das variáveis estão descritas na Tabela 3.

Tabela 3
Descrição das variáveis de nível 1 (medidas repetidas) utilizadas na pesquisa.

Categoria	Nome	Código ¹	Métrica
Educação	Proficiência escolar, onde: <ul style="list-style-type: none"> • p: P = Língua Portuguesa, M = Matemática; • q: E = Estadual, M = Municipal; • r: U = Urbana, R = Rural; • ss: 4S = 4ª Série/5º Ano, 8S = 8ª Série/9º Ano. 	Ppqrssyy	Escala Saeb
	Distorção idade-série, onde: <ul style="list-style-type: none"> • q: E = Estadual, M = Municipal; • r: U = Urbana, R = Rural; • ss: 1S a 9S = 1º Ano ao 9º Ano. 	Dqrssyy	Percentual de alunos na idade INCORRETA ao ano cursado.

(continua)

Categoria	Nome	Código¹	Métrica
Saúde	Óbitos por diarreia	DIAYy	Número de mortes por diarreia de causas diversas
	Mortalidade Infantil	IMFyy	Número de óbitos infantis por mil nascimentos.
Econômica	Produto Interno Bruto Municipal Real per capita	PIBPCyy	mil R\$ por habitante
	Valor Adicionado Bruto Real Municipal, onde <i>x</i> representa os setores: <ul style="list-style-type: none"> • A = Agropecuário; • I = Industrial; • S = Serviços; • E = Educação, Saúde, Seguro e Administração; • M = Impostos, do Setor Agropecuário 	VxRPyy	Percentual obtido pela divisão entre VARP pelo PIB real do município.
	Valor Adicionado Bruto Real Municipal Total, fruto da somatória de VSRP e VERP	VSRTPy	Percentual obtido pela divisão entre VSRTP pelo PIB real do município.
	Indicador de inflação INPC, onde <i>x</i> representa as seguintes categorias de despesas: <ul style="list-style-type: none"> • G = Geral; • A = Alimentos; • H = habitação; • R = Gastos Residenciais; • V = Vestuário; • T = Transporte; • S = Saúde; • P = Despesas Pessoais; • E = Educação; • C - Comunicação. 	IPCxyy	Percentual de inflação acumulada anual.
Agropecuária	Produção municipal de feijão	FEIyy	Toneladas de grãos.
	Produção municipal de milho	MILyy	Toneladas de grãos.
	Produção municipal de soja	SOJyy	Toneladas de grãos.
	Cabeças ordenhadas de vaca leiteira	LEIyy	Número de cabeças.
Demográfica	População total municipal	POPyy	Número de habitantes
	Densidade populacional municipal	DENSyy	Número de habitantes por km ² .
Programa Bolsa Família	Famílias Beneficiadas pelo PBF,	BFFyy	Número de famílias
	Valor deflacionado (real) repassado a cada município pelo PBF,	BFRyy	Valor em mil R\$
	Famílias atendidas pelo PBF do total de famílias elegíveis ao programa.	BFCyy	Percentual

Nota. (1) yy = dois dígitos variando de 00 a 10, representando os anos de 2000 a 2010, respectivamente.

Abaixo estão listados alguns exemplos de decodificação do nome das variáveis:

- PPMU4S00 – variável de proficiência escolar (P) de língua portuguesa (P) de escola municipal (M) de meio urbano (U) da turma de 4ª série (4S) do ano 2000 (00);
- DER5S10 – variável de distorção idade-série (D) de escola estadual (E) do meio rural (R) do 5ª ano (5S) do ano de 2010;
- DIA02 – variável de óbitos por diarreia (DIA) do ano 2002;
- IMF09 – variável de indicador de mortalidade infantil do ano de 2009;
- PIBPC03 – variável de produto interno bruto per capita do ano de 2003;
- VARP08 – variável de Valor Adicionado Bruto real em percentagem do setor agropecuário do ano 2008;
- VSRTP04 – variável de Valor Adicionado Bruto real em percentagem do setor de serviços total do ano 2004;
- IPCG00 – variável de índice de inflação do INPC geral do ano de 2000;
- FEI01 – variável de produção de grãos de feijão do ano de 2001;
- BFF04 – variável do Programa Bolsa Família (BF) de número de famílias de beneficiários (F) do ano de 2004.

Os dados de proficiência foram compilados de modo que cada variável expressasse uma categoria específica. Foram computados apenas os dados de proficiência do ensino fundamental. As combinações foram: proficiência em língua portuguesa e matemática; escolas urbanas e rurais; e escolas municipais e estaduais. As escolas particulares foram retiradas da base pois não possuem levantamento para todo o período de estudo. As escolas federais foram excluídas porque são em número muito reduzido para o ensino fundamental. A partir de 2004, a duração do ensino fundamental mudou de oito para nove anos por meio do Projeto de Lei nº 3.675/04, transformado em Lei Ordinária 11.274/2006. Houve apenas a inclusão da classe de alfabetização, de maneira que o 5º e o 9º anos correspondem à 4ª e 8ª séries do sistema de oito anos. Os estados e municípios tiveram até 2010 para se adaptarem ao novo sistema. Com isso, o registro de desempenho escolar e de matrícula ficaram divididos em dois sistemas de ensino fundamental durante esse período de 2004 a 2010. O problema de análise que se configurou é que cada escola adotou o seu próprio esquema de transição. Por exemplo, as escolas do Mato Grosso do Sul converteram-se

quase que instantaneamente para o novo sistema. A combinação dos dados de proficiência de 4ª série/5º ano e 8ª série/9º ano não foi feita. Uma média ou ponderação não teriam qualquer significado analítico. Optou-se por manter estes registros separados. Um problema sério dos dados de proficiência é a sua característica amostral até o ano de 2005 (Saeb). Em 2005, cerca de 4 mil instituições de ensino foram levantadas de um total de mais de 240 mil em todo o Brasil. Os dados de proficiência de 2000 são poucos e, nos anos seguintes, nem sempre a mesma escola / município eram levantados. Além disso, os critérios de escolha das escolas que seriam submetidas à avaliação mudaram ao longo dos anos. São exemplos o número de alunos mínimo por turma para ser elegível ao levantamento e a escolha por conveniência das escolas rurais a serem incluídas no levantamento. Outros problemas encontrados foram: dicionário de variáveis com codificação incorreta da dependência das instituições (federal, estadual, municipal ou particular), diferença nos dados de dependência e de localização da instituição (rural ou urbana) entre o censo escolar e o Saeb, erro na sintaxe do SPSS fornecida pelo Inep para extração de dados e erro no descritor da linha de dado do arquivo de extensão “TXT” (indicação do dado e suas respectivas colunas).

O dado de distorção idade-série (DIS) foi modificado para as necessidades do presente trabalho. Ele não corresponde ao indicador de mesmo nome calculado pelo Inep. O indicador do Inep considera o aluno em distorção se a sua idade supera em mais de dois anos a idade correta para a respectiva série/ano. Para o 1º ano, a idade ideal é seis anos (sistema de nove anos). Para o Inep, o aluno é considerado em distorção apenas com a idade de nove anos ou mais. No presente estudo, a DIS foi calculada conforme a *Equação 4*.

$$DIS_a = \frac{(MT_a - A_a)}{MT_a} \times 100 \quad (4)$$

onde:

a = série/ano do ensino fundamental

DIS_a = Percentual de alunos com a idade incorreta para a série/ano a

MT_a = matrículas totais para a série/ano a

A_a = alunos matriculados na série/ano a em idade correta

A finalidade de diminuir a tolerância deste indicador é torná-lo mais sensível a mudanças positivas e mais precisa quanto ao ano que começa a melhoria. Os dados de matrícula foram obtidos a partir do censo escolar.

No caso do DIS, os dados dos sistemas de oito e nove anos foram justapostos. A grande maioria das escolas pararam o registro das turmas no sistema antigo assim que passaram a registrar no novo sistema. O cálculo da DIS do município foi feito pela média simples de DIS, considerando escolas de mesma dependência, localização e série/ano.

Mortalidade infantil foi calculada a partir dos dados de óbitos de crianças com menos de um ano dividido pelo número de nascidos vivos. A razão é então multiplicada por mil para ajustar o indicador para o padrão utilizado no Índice de Mortalidade Infantil (IMF).

A inclusão de dados de óbitos por diarreia teve o intuito de incluir a faixa adulta da população e detectar áreas de pobreza. Óbitos por diarreia estão normalmente relacionados com falta de infraestrutura básica de água e esgoto, típico de áreas mais pobres. Também está relacionado com subnutrição, com falta de recursos para acesso a alimentos de melhor qualidade e com falta de assistência médica. Estas relações foram estudadas por Heller (1998). A relação entre pobreza e mortalidade infantil foi estudada por Tejada, Jacinto e Santos (2008). A relação entre mortalidade infantil e o PBF foi estudada por Camelo, Priscilla e Saiani (2009). O estudo do impacto do PBF em variáveis de mortalidade infantil e diarreia também foi feito anteriormente por Rasella, Aquino, Santos, Paes-Sousa e Barreto (2013).

Os dados econômicos de Produto Interno Bruto (PIB) e valor agregado foram deflacionados utilizando o Índice Nacional de Preços ao Consumidor Amplo (IPCA) compilado pelo Instituto de Pesquisa Econômica Aplicada (2013) e produzido pelo IBGE. Todos os dados de PIB foram trazidos a valores do mês de janeiro de 2013. Desta forma, os valores econômicos tornaram-se comparáveis entre si. Utilizou-se a *Equação 5*.

$$\text{Valor deflacionado}_{2013} = \frac{\text{valor nominal}_i \times \text{IPCA}_{2013}}{\text{IPCA}_i} \quad (5)$$

onde:

i = ano do valor nominal coletado

Os dados de pobreza não são coletados regularmente pelo governo. As medidas utilizadas são faixas de renda determinadas a partir de valores internacionais de linha de pobreza e de extrema pobreza, como os apresentados por Revallion (2010). Optou-se pelo uso do PIB como uma forma indireta de medir a pobreza. Ela foi medida de forma comparativa com os demais municípios em termos de produção de riqueza. A falta de um dado de pobreza medido diretamente é um aspecto complicador para aferir o impacto de programas de transferência de renda. Sabe-se quanto o programa transfere de renda, mas não se sabe o quanto o beneficiário já possui (ou não).

As variáveis de produção anual de riqueza são Valores Adicionados Brutos do setor agropecuário (VAR), do setor industrial (VIR), do setor de serviços (VSR), da arrecadação de impostos (VMR) e das atividades ligadas à educação, saúde, seguros e administração (VER). Estes valores foram desmembrados do PIB. Apesar de ser uma medida macroeconômica comumente calculada para o nível de país, o IBGE faz o cálculo em nível de município. A intenção em utilizar Valores Adicionados Brutos (VAB) é para indiretamente medir qual o setor mais importante para a economia do município e o relacionar com o PBF. É também uma forma de estimar qual a maior fonte de emprego. Outro motivo para a escolha de VAB foi a falta de dados de renda das famílias com frequência de coleta anual. As medidas de renda são feitas nos censos demográficos, pela Pesquisa Nacional por Amostra de Domicílio (Pnad) e na Pesquisa de Orçamentos Familiares (POF). O primeiro tem frequência decenal, o segundo tem frequência anual, mas apenas a partir de 2003, e o terceiro tem frequência irregular. Os dados de emprego também são escassos e normalmente levantados por entidades de classe nas grandes capitais do país impondo uma limitação geográfica ao dado.

A *Equação 6* relaciona os Valores Adicionados Brutos com o PIB.

$$PIB = VAR + VIR + VSR + VMR + VER \quad (6)$$

As produções agropecuárias foram escolhidas pela sua representatividade de dois sistemas produtivos distintos, um mais intensivo em capital e outro menos intensivo. As produções de feijão e leite são típicas de pequenos produtores rurais, a sua maioria de estrutura familiar. Espera-se que municípios com algum registro de produção de feijão ou leite venham a identificar áreas mais pobres. O feijão é

intensivo em mão de obra, principalmente durante a colheita. Máquinas para colher o feijão foram desenvolvidas, mas são caras e somente se justificam para colheita de grandes áreas. Além disso, o feijão é produto básico da dieta do brasileiro. As famílias mantêm pequenas roças de feijão para autoconsumo e, eventualmente, para venda. No caso do leite, a vaca funciona não só como um bem de capital produzindo leite e carne, mas é uma espécie de “poupança” da família. Em época de baixos preços da carne ou do leite, os animais podem ser deixados por conta própria nos pastos. Apesar de magros e com baixa produtividade, dada a sua rusticidade, continuarão vivos e se reproduzindo sem muita assistência. Na sinalização de preços melhores, o rebanho pode ser rapidamente mobilizado para incrementar a produção pela introdução de dietas mais calóricas e proteicas.

As culturas de milho e soja são culturas consideradas intensivas em insumo, i.e., demandam sementes melhoradas, fertilizantes, melhor preparo do solo, mais emprego de máquinas e implementos agrícolas, controle e combate de pragas e doenças e áreas maiores de cultivo para que gerem retorno do investimento. Estas culturas caracterizam produtores de maior capital, logo, municípios menos pobres.

A partir dos dados de área do município e população total, calculou-se a densidade populacional em termos de habitantes por km^2 . Também utilizou-se o dado populacional para ponderar o PIB. O objetivo foi detectar o nível de concentração da renda do município. Diferentemente de dados de renda do censo, o PIB não pondera a produção de riqueza. Dois municípios de área e população diferentes podem apresentar o mesmo valor de PIB, de maneira que uma comparação direta de PIB entre municípios pode não refletir a realidade de pobreza de cada um. Em vista disso, na presente pesquisa optou-se por ponderar a produção de riqueza pelo tamanho da população do município, na forma de PIB per capita (PIBPC). Esta é também uma forma indireta de se medir a renda per capita.

Os dados do Programa Bolsa Família (PBF) foram apenas recentemente reunidos e disponibilizados (2012) no sítio do Ministério do Desenvolvimento Social e Combate à Fome por meio da Secretaria de Avaliação e Gestão da Informação – Sagi (MDS, 2013).

Os modelos foram comparados pelo seu valor de *deviance* e pelo seu grau de parcimônia. O *deviance* é uma medida do desajuste do modelo em relação aos dados.

Quanto menor o *deviance*, menor o desajuste. A determinação do grau de desajuste do modelo de regressão foi feita pela comparação do modelo denominado “vazio” (apenas com os resíduos) com o modelo “regredido” (com os coeficientes de regressão). Quando o *deviance* do modelo regredido é menor do que o *deviance* do modelo vazio, pode-se afirmar que o modelo regredido é melhor que o modelo vazio. Um valor de zero de *deviance* indica ajuste perfeito do modelo. A significância do *deviance* do modelo regredido é determinada pela diferença dos *deviances* do modelo vazio e do modelo regredido. Os valores obtidos são comparados com o *p-valor* de uma distribuição qui-quadrado unicaudal (todos os valores desta pesquisa são positivos) com graus de liberdade igual à diferença no número de variáveis de cada modelo.

A significância dos coeficientes das variáveis explicativas e dos valores de resíduos obtidos foi verificada por meio do teste *t* de Wald (*t*-Wald), conforme Wald (1943). O seu cálculo está na *Equação 7*. Se o valor de *Z* obtido é *p-valor* > 1,96, diz-se que o coeficiente é significativo a 95%.

$$Z = \frac{\pi}{EP} \quad (7)$$

Onde:

Z = valor do teste *t*-Wald;

π = valor do coeficiente de regressão ou resíduo;

EP = erro padrão do coeficiente π .

4.4.3 Modelos de análise

Foram desenvolvidos nove modelos de regressão multinível divididos em cinco grupos de análise. Todos os modelos têm como variável dependente o Produto Interno Bruto per capita, deflacionado (PIBPC). Os modelos variam em termos de variáveis explicativas incluídas no modelo e o período de análise. As quatro principais análises foram realizadas por meio de quatro modelos denominados de M1, M2, M3 e M4. Uma quinta análise foi feita por meio de cinco modelos (M5-Cx). Esta quinta análise teve o único propósito de avaliar individualmente cada um dos componentes discriminados na análise de componentes principais. Isto permitiu que as variáveis de cada componente expressassem toda a sua variância sem interferência de outros

componentes. Os modelos deste quinto grupo de análise foram identificados por M5-Cx, onde Cx representa cada um dos componentes identificados na análise de componentes principais (ACP). Por exemplo, o modelo de regressão multinível do Componente 1 (C1) será identificado por M5-C1.

O modelo M1 é o único modelo que reúne todas as seis categorias de variáveis. Para que isso fosse possível, cortes temporais foram necessários. Os dados do Programa Bolsa Família (PBF) estão disponíveis apenas para o período de 2004 a 2010. Os dados de proficiência escolar são medidos apenas nos anos ímpares. A interseção destas duas variáveis fez permanecer no modelo os anos 2005, 2007 e 2009. Analisando os dados de proficiência para 2005, o número de casos disponíveis era muito baixo, dado o caráter amostral do levantamento. Em razão disso, o modelo M1 incluiu apenas dados de 2007 e 2009.

Dada a limitação do número de anos de análise dos modelos M1 e M5, o Modelo M2 foi desenvolvido de maneira que todos os onze anos do levantamento estivessem presentes no modelo. Este modelo servirá de referência para a análise dos dois modelos posteriores (M3 e M4).

O Modelo M3 representa o modelo antes da intervenção do PBF. Ele incluiu dados dos anos 2000 a 2003. Evidentemente, as variáveis explicativas ligadas ao programa não estão presentes neste modelo. Também ficaram de fora os dados de proficiência. Ainda assim, as variáveis de distorção idade-série (DIS) serviram como fonte de variância para o fenômeno educação. A prioridade foi incluir o maior número de anos possíveis.

O Modelo M4 representa o modelo com a intervenção do PBF. Ele incluiu dados dos anos 2004 a 2010. As variáveis referentes ao PBF e à proficiência escolar ficaram de fora. Este modelo foi mantido o mais próximo possível do Modelo M3 em termos de variáveis explicativas. O objetivo foi comparar M3 e M4 a fim de se observar diferenças que pudessem ser atribuídas ao PBF.

Os modelos com os respectivos períodos de análise estão na Tabela 4. Os modelos com os respectivos componentes estão na Tabela 5.

Tabela 4*Modelos multinível da pesquisa e seus respectivos períodos de análise.*

Ano	Modelos								
	M1	M2	M3	M4	M5-C1	M5-C2	M5-C3	M5-C4	M5-C5
2000		X	X			X	X		X
2001		X	X		X	X	X		X
2002		X	X			X	X		X
2003		X	X		X	X	X		X
2004		X		X		X	X	X	X
2005		X		X	X	X	X	X	X
2006		X		X		X	X	X	X
2007	X	X		X	X	X	X	X	X
2008		X		X		X	X	X	X
2009	X	X		X	X	X	X	X	X
2010		X		X		X	X	X	X

Nota. X = ano incluído na análise.**Tabela 5***Modelos multinível da pesquisa e seus respectivos componentes.*

Componente	Modelos								
	M1	M2	M3	M4	M5-C1	M5-C2	M5-C3	M5-C4	M5-C5
C1 – proficiência	X				X				
C2 – DIS estadual	X	X	X	X		X			
C3 – DIS municipal	X	X	X	X			X		
C4 – bem-estar	X	- PBF	- PBF	- PBF				X	
C5 - agropecuária	X	X	X	X					X

Nota. X = componente incluído na análise; -PBF = as variáveis do Programa Bolsa Família foram retiradas do componente.

5. Resultados

Os resultados foram organizados em três sessões. A primeira sessão trata da análise estatística descritiva dos dados empregados na pesquisa, incluindo a Análise dos Componentes Principais (ACP). A segunda, apresenta o contexto da pesquisa por meio da organização temporal e espacial dos dados utilizados. A terceira sessão traz os resultados das análises multinível propriamente dita, assim como a verificação de pressupostos multivariados de normalidade, homocedasticidade e linearidade.

5.1 Análises estatísticas descritivas

A análise exploratória dos dados começou com a inspeção de valores atípicos univariados por meio de cálculo do escore z de cada variável. Os dados com resultados $z \geq 2,58$ foram substituídos por *missing*. Um problema decorrente da retirada de dados atípicos foi a perda dos dados de Brasília. Por se tratar do único

município do DF, os seus dados se sobressaem dentre os demais municípios brasileiros.

Após a retirada dos dados atípicos, calculou-se os dados de média e desvio padrão das variáveis do estudo. Os dados estão na Tabela 6. Os códigos das variáveis estão descritos na Tabela 3.

Após inspeção das assimetrias, nos casos em que se observou problemas procedeu-se à correção por meio da transformação de Box-Cox. Os resultados destas transformações estão nas análises descritivas a seguir, agrupadas por tipo de variável.

O primeiro grupo analisado foi o de proficiência escolar (Tabela 7). Os dados compreendem os anos ímpares entre 2001 a 2009, das escolas estaduais e municipais, em áreas urbanas e rurais, das disciplinas de matemática e língua portuguesa, da 4ª série/5º ano e 8ª série/9º ano.

Tabela 6

Estatística de tamanho da população, média e desvio-padrão, 2000 a 2010.

Variável	N	Média	Desvio padrão
<i>Agropecuárias</i>			
FEIJAO	60.456	368,51	794,52
LEITE	59.653	2.966,56	3.236,95
MILHO	60.069	5.414,47	10.920,97
SOJA	60.373	4.831,19	15.386,38
<i>Bolsa Família</i>			
BFC	38.069	93,58	25,43
BFF	38.452	1.284,81	1.930,45
BFR	38.560	1.466,02	2.319,51
<i>Saúde</i>			
DIARREIA	60.353	0,55	1,34
IMF	59.708	6,79	9,12
<i>Econômicas</i>			
PIBPC	60.072	10,74	7,87
VARP	60.660	23,48	15,55
VERP	60.935	26,88	14,26
VIRP	59.080	13,42	9,38
VMRP	59.735	5,35	3,22
VSRP	60.189	27,93	10,05
VS RTP	60.702	55,74	13,93
<i>Distorção idade-série</i>			
DER1S	11.192	60,91	26,37
DER2S	14.974	60,57	25,39
DER3S	15.101	65,39	25,31
DER4S	15.190	68,30	25,18
DER5S	14.039	66,38	22,80
DER6S	13.787	68,55	21,34
DER7S	13.426	67,62	19,27

(continua)

Variável	N	Média	Desvio padrão
DER8S	13.199	66,39	19,38
DER9S	8.547	62,35	19,86
DEU1S	26.844	59,74	23,18
DEU2S	35.571	57,34	21,44
DEU3S	36.593	62,09	21,62
DEU4S	37.329	64,36	22,16
DEU5S	41.935	64,74	19,59
DEU6S	52.195	68,46	19,48
DEU7S	52.522	67,16	17,47
DEU8S	52.691	66,49	17,21
DEU9S	33.502	62,51	16,16
DMR1S	37.079	60,91	23,53
DMR2S	48.945	64,27	21,01
DMR3S	48.998	69,46	20,18
DMR4S	48.912	72,09	20,44
DMR5S	39.600	71,45	19,88
DMR6S	26.430	78,50	15,83
DMR7S	25.442	77,32	15,25
DMR8S	24.453	77,14	16,23
DMR9S	16.753	72,83	17,07
DMU1S	41.873	56,31	24,91
DMU2S	54.916	57,92	22,49
DMU3S	54.439	63,32	21,02
DMU4S	54.263	65,87	21,19
DMU5S	46.766	66,62	19,88
DMU6S	34.670	76,49	15,63
DMU7S	33.721	75,03	14,82
DMU8S	32.789	74,36	15,89
DMU9S	22.327	69,97	16,15
<i>Proficiência</i>			
PMER4S	698	180,03	25,89
PMER8S	706	236,52	19,25
PMEU4S	6.996	190,18	23,60
PMEU8S	10.313	239,70	18,29
PMMR4S	2.884	176,84	24,42
PMMR8S	1.645	226,21	18,20
PMMU4S	11.185	189,34	23,77
PMMU8S	7.265	233,92	19,30
PPER4S	700	162,25	22,14
PPER8S	707	230,51	17,95
PPEU4S	7.023	172,54	19,84
PPEU8S	10.319	231,06	16,77
PPMR4S	2.893	159,40	20,75
PPMR8S	1.648	220,89	18,58
PPMU4S	11.242	171,02	19,51
PPMU8S	7.297	226,11	19,01

O número de casos *missing* é elevado por causa do caráter amostral dos dados, principalmente nos anos iniciais deste estudo. O número de casos válidos pelo critério *listwise* (sem dados *missing* em qualquer das 16 variáveis nos onze anos) é de apenas 95 de um total de 27.825 casos, representando apenas 0,3% do total. Os índices de normalidade (assimetria e curtose) de todas as variáveis de proficiência indicaram

distribuição normal, não sendo necessário empregar qualquer transformação. As escolas rurais são as que apresentaram menor número de casos, variando de 700 a 2.893 casos. As escolas urbanas variaram de 6.996 a 11.242 casos. As escolas municipais são as que têm maior número de casos na 4ª série/5º ano e as estaduais as que têm maior número de casos na 8ª série/9º ano. As escolas rurais têm menores escores de proficiência que as escolas urbanas. Os escores das 8ª séries/9º ano são sempre maiores que os da 4ª série/5º ano, o que indica que existiu melhora educacional com a progressão nas séries pelo aluno.

Tabela 7

Estatísticas descritivas das variáveis de proficiência, escala Saeb, bienais, 2001 a 2009.

Variável	N	Min	Max	Média		DP	Assimetria	Curtose
				Valor	EP			
PMER4S	698	101,79	246,76	180,03	0,98	25,89	0,06	-0,26
PMER8S	706	186,36	288,49	236,52	0,72	19,25	0,13	-0,45
PMEU4S	6996	117,47	264,21	190,18	0,28	23,60	0,19	-0,16
PMEU8S	10.313	175,88	292,68	239,70	0,18	18,29	0,13	-0,36
PMMR4S	2884	105,74	244,07	176,84	0,45	24,42	0,24	-0,13
PMMR8S	1645	177,05	278,77	226,21	0,45	18,20	0,54	-0,04
PMMU4S	11.185	114,19	264,38	189,34	0,22	23,77	0,31	-0,15
PMMU8S	7265	177,40	290,66	233,92	0,23	19,30	0,36	-0,30
PPER4S	700	88,58	215,36	162,25	0,84	22,14	-0,24	-0,13
PPER8S	707	180,82	279,05	230,51	0,68	17,95	-0,17	-0,21
PPEU4S	7.023	106,68	231,82	172,54	0,24	19,84	0,02	-0,18
PPEU8S	10.319	175,35	281,04	231,06	0,17	16,77	-0,05	-0,26
PPMR4S	2.893	82,98	213,22	159,40	0,39	20,75	-0,03	0,18
PPMR8S	1648	150,35	271,45	220,89	0,46	18,58	0,17	-0,22
PPMU4S	11.242	104,61	229,67	171,02	0,18	19,51	0,15	-0,35
PPMU8S	7.297	167,21	284,14	226,11	0,22	19,01	0,12	-0,28

Nota. EP = erro padrão; DP = desvio padrão.

O segundo grupo de variáveis analisado foi de distorção idade-série (DIS). Os seus resultados estão na Tabela 8.

Os dados de DIS também não apresentaram problemas de normalidade. O número de casos válidos é bem maior que o de proficiência. A explicação está na forma de cálculo da DIS, feita a partir dos dados de matrícula do Censo Escolar, a qual não é amostral. Destaca-se que em todos os anos, indiferente à dependência administrativa (estadual ou municipal) e à localização (urbana ou rural), ocorreram casos em que todas as escolas do município tinham 100% dos alunos em DIS. Notou-se também que nas séries/anos finais não há municípios que não tivesse algum grau de DIS. Já nas séries/anos iniciais, temos um resultado inverso. Existem municípios em

que não ocorre DIS nas séries/anos iniciais de suas escolas. As escolas municipais rurais foram as que apresentaram maior média de DIS. Porém, nas escolas estaduais, a diferença de médias entre escolas rurais e urbanas não é tão evidente. As variâncias são maiores nas séries/anos iniciais, decrescendo conforme o aluno avança nos anos. Isto pode ser fruto do menor número de instituições oferecendo os anos mais avançadas, ou um processo seletivo no qual os alunos que permanecem nas escolas são exatamente aqueles que não sofreram descontinuidade na vida acadêmica.

Tabela 8

Estatísticas descritivas das variáveis de distorção idade-série, percentual, 2000 a 2010.

Variável	N	Min	Max	Média		DP	Assimetria	Curtose
				Valor	EP			
DER1S	11.192	0,00	100	60,91	0,25	26,37	-0,40	-0,63
DER2S	14.974	0,00	100	60,57	0,21	25,39	-0,43	-0,53
DER3S	15.101	0,00	100	65,39	0,21	25,31	-0,60	-0,40
DER4S	15.190	0,00	100	68,30	0,20	25,18	-0,57	-0,51
DER5S	14.039	0,00	100	66,38	0,19	22,80	-0,48	-0,38
DER6S	13.787	8,33	100	68,55	0,18	21,34	-0,29	-0,77
DER7S	13.426	10,71	100	67,62	0,17	19,27	-0,32	-0,63
DER8S	13.199	10,00	100	66,39	0,17	19,38	-0,18	-0,69
DER9S	8.547	8,82	100	62,35	0,21	19,86	-0,07	-0,67
DEU1S	26.844	0,00	100	59,74	0,14	23,18	-0,24	-0,63
DEU2S	35.571	0,00	100	57,34	0,11	21,44	-0,38	-0,41
DEU3S	36.593	0,00	100	62,09	0,11	21,62	-0,44	-0,50
DEU4S	37.329	2,17	100	64,36	0,11	22,16	-0,26	-0,61
DEU5S	41.935	6,25	100	64,74	0,10	19,59	-0,32	-0,56
DEU6S	52.195	14,83	100	68,46	0,09	19,48	-0,22	-0,68
DEU7S	52.522	17,07	100	67,16	0,08	17,47	-0,33	-0,65
DEU8S	52.691	17,65	100	66,49	0,07	17,21	-0,23	-0,66
DEU9S	33.502	18,99	100	62,51	0,09	16,16	-0,10	-0,63
DMR1S	37.079	0,00	100	60,91	0,12	23,53	-0,63	-0,09
DMR2S	48.945	0,00	100	64,27	0,09	21,01	-0,66	-0,06
DMR3S	48.998	5,00	100	69,46	0,09	20,18	-0,76	-0,11
DMR4S	48.912	8,33	100	72,09	0,09	20,44	-0,71	-0,24
DMR5S	39.600	11,62	100	71,45	0,10	19,88	-0,65	-0,31
DMR6S	26.430	22,40	100	78,50	0,10	15,83	-0,77	0,00
DMR7S	25.442	25,00	100	77,32	0,10	15,25	-0,65	-0,14
DMR8S	24.453	25,00	100	77,14	0,10	16,23	-0,57	0,02
DMR9S	16.753	21,05	100	72,83	0,13	17,07	-0,42	0,02
DMU1S	41.873	0,00	100	56,31	0,12	24,91	-0,44	0,01
DMU2S	54.916	0,00	100	57,92	0,10	22,49	-0,54	0,01
DMU3S	54.439	3,70	100	63,32	0,09	21,02	-0,52	0,01
DMU4S	54.263	6,15	100	65,87	0,09	21,19	-0,38	0,01
DMU5S	46.766	10,37	100	66,62	0,09	19,88	-0,39	0,01
DMU6S	34.670	30,09	100	76,49	0,08	15,63	-0,61	0,01
DMU7S	33.721	29,51	100	75,03	0,08	14,82	-0,54	0,01
DMU8S	32.789	27,27	100	74,36	0,09	15,89	-0,44	0,01
DMU9S	22.327	24,29	100	69,97	0,11	16,15	-0,32	0,02

Nota. EP = erro padrão; DP = desvio padrão.

O terceiro grupo analisado é das variáveis de saúde: óbitos por diarreia (DIARREIA) e mortalidade infantil (IMF). Os dados do grupo estão na Tabela 9.

Tabela 9

Estatísticas descritivas das variáveis óbito por diarreia e mortalidade infantil, em número de óbitos e em número de óbitos para cada mil nascidos vivos, respectivamente, 2000 a 2010.

Variável	N	Min	Max	Média		DP	Assimetria	Curtose
				Valor	EP			
DIARREIA	60.353	0	11	0,55	0,01	1,34	3,67	16,28
IMF	59.708	0,00	56,77	6,79	0,04	9,12	1,58	2,44

Nota. EP = erro padrão; DP = desvio padrão.

Os dados coletados apresentaram muitos municípios com valores zero, o que diminuiu acentuadamente a média. Aparentemente, os indivíduos saem de seus municípios de origem e se dirigem a municípios maiores para tratamento. No evento de óbito, é nesses locais que é registrado o evento. Este fenômeno foi mais perceptível em óbito por diarreia, o que se refletiu na média próxima de 0,00 e alta curtose e assimetria. Esta variável apresentou 45.289 casos com dados zero, ou 74% do total de casos. Em menor grau, mortalidade infantil apresentou 28.228 casos zero, ou 46% do total de casos. Andrade e Szwarcwald (2007) já reportavam o problema de subenumeração de nascidos vivos pelos dados do Ministério da Saúde, o que pode explicar parte dos casos zero. Tanto em óbitos por diarreia, como em mortalidade infantil, dada a grande quantidade de valores zero nenhum ajuste por transformação foi efetivo. Decidiu-se por manter as variáveis com os seus valores originais.

O quarto grupo de análise compreendeu as variáveis do Programa Bolsa Família (PBF). Os resultados descritivos estão na Tabela 10.

Tabela 10

Estatísticas descritivas das variáveis do Programa Bolsa Família, 2004 a 2010, com exceção de 2008.

Variável	N	Min	Max	Média		DP	Assimetria	Curtose
				Valor	EP			
BFF (famílias)	33.021	1	20.139	1.496,12	11,04	2.005,85	3,29	14,91
BFR (mil R\$)	33.022	0,27	25.107,37	1.711,87	13,32	2.421,07	3,38	16,14
BFC (%)	32.670	3,87	197,93	93,34	0,15	26,58	-0,41	0,64

Nota. EP = erro padrão; DP = desvio padrão.

A variável BFF representa o número de famílias beneficiadas pelo Programa Bolsa Família (PBF). A variável BFR representa o valor em mil reais repassado ao município pelo PBF. A variável BFC é um valor percentual que indica o número de famílias elegíveis ao PBF que são beneficiadas por este programa. Um problema encontrado nas três variáveis foi a inconsistência dos dados em 2008. Por exemplo, a variável BFF apresentou 5.431 municípios com nenhuma família beneficiada pelo PBF, ou 98% do total de municípios. Os dados de 2008 das três variáveis foram todos descartados. Os dados de BFR foram deflacionados utilizando o mesmo método empregado no deflacionamento do PIB. Utilizou-se o IPCA para trazer a valores de janeiro de 2013. Dada a semelhança dos desvios padrão e erros padrão de BFF e BFR, calculou-se a correlação bivariada de Pearson entre estas variáveis. Obteve-se 0,975, IC 99% [0,974, 0,976]. Os valores de assimetria e curtose para BFF estavam elevados. Procedeu-se à transformação de Box-Cox. O valor de $\lambda = 0$ sugeriu transformação logarítmica. Como o menor valor de BFF é igual a 1,00, não foi necessário ajustar BFF para aplicar o logaritmo. Após ajuste, a assimetria ficou em -0,33 e a curtose em 0,38.

Inspecionando os dados de cobertura do PBF (BFC), notou-se a existência de valores acima de 100%. Aparentemente, trata-se de um problema de cadastro de pessoas elegíveis ao programa. A partir de 2010, o MDS passou a disponibilizar dados com base no CadÚnico. Os valores de BFC do CadÚnico não ultrapassam a 100%. Comparando os dados de 2010 da base antiga com os dados da base do CadÚnico, a correlação de Pearson foi de 0,791, IC 99% [0,786, 0,796]. Assim, apesar dos valores da base antiga apresentarem valores maiores que os do CadÚnico, ambas medem o mesmo fenômeno. BFC apresentou distribuição normal, não exigindo transformação dos dados.

O próximo grupo analisado é de variáveis da agropecuária. Os dados estão na Tabela 11.

Feijão, milho e soja foram medidos em toneladas de grão produzidos no município. Leite foi medido por número de cabeças de vacas ordenhadas. Os valores de média baixos em relação ao mínimo e máximo e a variância elevada indicam o mesmo problema observado com os dados de saúde com valor zero. Muitos municípios não apresentam produção agropecuária para os produtos aqui listados. O

número de municípios sem produção de feijão, milho e leite são inferiores a 14%. Por sua vez, soja apresenta um valor alto de 69%. Esta alta porcentagem está relacionada com a exigência da cultura de soja por grandes áreas mecanizáveis e de condições edafoclimáticas (solo e clima) específicas. Aplicando a transformação de Box-Cox, obteve-se $\lambda = 0$, o que implica na aplicação de transformação logarítmica. Como os valores mínimos das quatro variáveis é zero, foi preciso adicionar a constante 1,00 a fim de gerar resultados positivos. As assimetrias finais ficaram -0,03, -0,16, 1,09 e 0,14 para feijão, milho, soja e leite, respectivamente. Mesmo após a transformação, soja ainda apresenta alguma assimetria positiva.

Tabela 11

Estatísticas descritivas das variáveis agropecuárias, em toneladas de grãos e número de cabeças ordenhadas, 2000 a 2010.

Variável	N	Min	Max	Média		DP	Assimetria	Curtose
				Valor	EP			
FEIJAO	60.456	0	8.651	369	3,23	795	4,43	24,68
MILHO	60.069	0	104.002	5.414	44,56	10.921	3,71	17,37
SOJA	60.373	0	161.200	4.831	62,62	15.386	4,74	26,64
LEITE	59.653	0	20.280	2.967	13,25	3.237	2,00	4,54

Nota. EP = erro padrão; DP = desvio padrão.

O último grupo analisado consiste nos dados de Valor Adicionado Bruto (VAB). Os dados estão na Tabela 12.

Os valores originais são em reais e foram deflacionados para janeiro de 2013 utilizando o IPCA. As variáveis estão expressas como percentual do PIB do município. Apenas três variáveis apresentaram problemas de normalidade: PIB per capita (PIBPC), a variável de VAB industrial (VIRP) e a de VAB oriundo da arrecadação de impostos (VMRP). Aplicando a transformação de Box-Cox, obteve-se o valor de $\lambda = -0,3$ para VIRP e $\lambda = 0$ para PIBPC e VMRP. Calculou-se a inversa da potência 0,3 dos valores de VIRP. Foi necessário ajustar os valores calculados para reverter a ordem. Para VMRP foi necessário adicionar 0,98 ao valor original antes de calcular o logarítmico. Para PIBPC, os valores de ajustes foram variáveis de ano para ano. Desta maneira, obteve-se apenas valores positivos após a transformação. Os novos valores de assimetria foram -0,04, 0,09 e 0,24 para PIBPC, VIRP e VMRP, respectivamente.

Tabela 12

Estatísticas descritivas das variáveis econômicas, em mil R\$ por habitante para PIBPC e em porcentagem para os Valores Adicionados Brutos, 2000 a 2010.

Variável	N	Min	Max	Média		DP	Assimetria		Curtose	
				Valor	EP		Valor	EP	Valor	EP
PIBPC	60.072	0,00	1,85	0,90	0,00	0337	1,74	0,01	4,23	0,02
VARP	60.660	0,00	71,53	23,48	0,06	15,55	0,48	0,01	-0,58	0,02
VIRP	59.080	0,75	49,82	13,42	0,04	9,38	1,64	0,01	2,12	0,02
VSRP	60.189	1,45	58,01	27,93	0,04	10,05	0,55	0,01	-0,19	0,02
VSRTP	60.702	14,96	91,02	55,74	0,06	13,93	-0,20	0,01	-0,71	0,02
VERP	60.935	0,98	67,54	26,88	0,06	14,26	0,50	0,01	-0,74	0,02
VMRP	59.735	0,02	17,62	5,35	0,01	3,22	1,07	0,01	0,71	0,02

(*) EP = erro padrão; DP = desvio padrão.

Corrigido os problemas de dados atípicos e assimetria, verificou-se o agrupamento das variáveis do estudo. Para esta análise, utilizou-se a Análise de Componentes Principais (ACP). Porém, algumas variáveis possuem cortes temporais diferentes da maioria. Os dados de proficiência escolar são coletados a cada dois anos, em anos ímpares. Os dados do PBF começaram a ser coletados a partir de 2004. Os dados do 9º ano, tanto de proficiência como de DIS, só começaram a ser coletados a partir de 2004. Dada estas diferenças de disponibilidade de dados dentro do horizonte da pesquisa, foi necessário selecionar um conjunto específico de anos de maneira que todas as variáveis estivessem presentes a fim de permitir a ACP. Os anos que permitem reunir todas as variáveis são 2005, 2007 e 2009. Verificando o número de casos válidos pelo critério *listwise*, 2005 resultou em nenhum caso válido. Assim, apenas os anos de 2007 e 2009 foram selecionados para a ACP. Conforme mencionado anteriormente, percebeu-se que o número de casos das variáveis educacionais (proficiência e DIS) para o meio rural é bem menor comparativamente aos dados do meio urbano. Por exemplo, a proficiência em português da 4ª série/5º ano de uma escola estadual do meio rural (PPER4S) possui 700 casos na base. O mesmo dado, mas de uma escola do meio urbano (PPEU4S) é de 7.023 casos, mais de dez vezes maior. Por causa desta diferença muito grande de casos, os dados relativos ao meio rural foram também removidos da ACP.

Aplicou-se a Análise Paralela (AP) para determinar o número de dimensões do problema, pois é o método mais preciso de avaliação da dimensionalidade de uma matriz (Ledesma & Valero-Mora, 2007; Timmerman & Lorenzo-Seva, 2011). Por meio de AP aplicada à ACP, utilizando o método de permutação para geração dos autovalores esperados, discriminou-se sete fatores (Tabela 13).

Tabela 13*Autovalores da análise paralela por permutação da base de dados de 41 variáveis, 2007 e 2009.*

Raiz	Autovalor Empírico	Autovalor AP	95º percentil
1	13,544	1,366	1,403
2	4,483	1,327	1,355
3	3,229	1,299	1,323
4	2,094	1,273	1,295
5	2,037	1,250	1,271
6	1,569	1,229	1,248
7	1,355	1,209	1,227
8	1,178	1,191	1,209
9	0,917	1,173	1,190
10	0,870	1,156	1,173

Nota. Autovalor AP = autovalor obtido por meio de análise paralela.

Por meio da ACP, com extração de sete fatores e emprego de rotação Varimax, obteve-se os componentes da Tabela 14. Os valores de coeficientes menores que 0,32 foram suprimidos.

As considerações deste parágrafo dizem respeito às denominações de componente da Tabela 14. O Componente 6 (C6) agrupou variáveis que não exprimem qualquer significado teórico. Soja apresentou coeficiente tanto em C6 como em C7. Apesar do valor maior do coeficiente em C6, a variável faz mais sentido agrupada em C7, com outros produtos agropecuários. VIRP e PIBPC também se agruparam em C1, onde predomina as variáveis ligadas à proficiência escolar. Como estas duas variáveis não apresentam significado em C1, foram mantidas em C6. Entretanto, optou-se por eliminar C6, movendo SOJA para C7, eliminando VERP e PIBPC da ACP. Apesar da remoção de PIBPC da ACP, ela será utilizada como variável dependente nas análises multinível. C5 detectou diferenças entre a distorção idade-série (DIS) observada nas primeiras séries em relação à DIS das demais séries. Como já existe dois componentes de DIS e apenas quatro variáveis se agruparam em C5, optou-se por descartar C5.

Tabela 14*Componentes extraídos por Análise dos Componentes Principais, com rotação Varimax.*

Variáveis	C1	C2	C3	C4	C5	C6	C7
PMMU4S	0,875						
PPMU4S	0,869						
PMEU4S	0,826	-0,349					
PPEU4S	0,820	-0,360					
PPMU8S	0,811						
PPEU8S	0,801	-0,327					
PMMU8S	0,783						
PMEU8S	0,769	-0,383					
DEU6S		0,821					
DEU5S		0,771	0,325				
DEU7S		0,757					
DEU8S		0,756					
DEU4S		0,741					
DEU9S		0,727					
DEU3S		0,534			0,518		
DMU6S		0,375	0,755				
DMU8S			0,753				
DMU7S		0,341	0,726				
DMU5S		0,429	0,717				
DMU9S	-0,332		0,702				
DMU4S		0,381	0,675				
DMU3S			0,553		0,506		
BFF				0,802			
BFR				0,758			
DIARREIA				0,726			
IMF				0,560			
VSRP	0,438			0,552			
VMRP	0,328			0,506		-0,495	
VIRP				0,462			-0,394
DMU1S					0,765		
DMU2S			0,361		0,747		
DEU1S					0,668		
DEU2S		0,357			0,668		
PIBPC	0,448					-0,711	
VERP	-0,502					0,585	
SOJA						-0,514	0,476
BFC						0,333	
MILHO							0,728
LEITE							0,623
FEIJAO						0,331	0,611
VARP				-0,566			0,579

Assim, definiram-se cinco dimensões finais: (C1) proficiência escolar de escolas do meio urbano; (C2) distorção idade-série (DIS) do 3º ano ao 9º ano de escolas estaduais; (C3) DIS da 3º ano ao 9º anos de escolas municipais; (C4) bem-estar (economia, saúde e variáveis do Programa Bolsa Família); e (C5) agropecuária. O conjunto final de fatores e suas respectivas variáveis estão na Tabela 15. Estas denominações de C1 a C5 é que serão utilizadas daqui para frente.

Tabela 15

As variáveis da pesquisa e suas respectivas dimensões.

Proficiência (C1)	DIS estadual (C2)	DIS municipal (C3)	Bem-estar (C4)	Agropecuária (C5)
PMEU4S	DEU3S	DMU3S	BFF	FEIJAO
PMEU8S	DEU4S	DMU4S	BFR	LEITE
PMMU4S	DEU5S	DMU5S	DIARREIA	MILHO
PMMU8S	DEU6S	DMU6S	IMF	SOJA
PPEU4S	DEU7S	DMU7S	VSRP	VARP
PPEU8S	DEU8S	DMU8S	VMRP	
PPMU4S	DEU9S	DMU9S	VIRP	
PPMU8S				

5.2 Contexto da pesquisa

Os resultados anteriormente apresentados incluem todos os 5.565 municípios sem distinção do ano ou da região. A seguir estão os resultados de recortes temporais e espaciais das variáveis estudadas. Para a montagem dos gráficos abaixo, utilizou-se a base de dados original, sem retirada dos dados atípicos e sem aplicação das transformações. Desta forma preservou-se as grandezas e escalas originais dos dados.

O número de famílias atendidas pelo PBF cresceu no período de 2004 a 2010, com diminuição no ritmo apenas em 2007 (Figura 5). Já o montante repassado pelo PBF foi sempre crescente (Figura 6). Como os dados de repasse foram deflacionados, o ganho foi real e acima da inflação apurada pelo IPCA no período. O descompasso entre o número de famílias atendidas pelo PBF e o repasse feito aos municípios pode ser consequência do tamanho da população elegível ao programa.

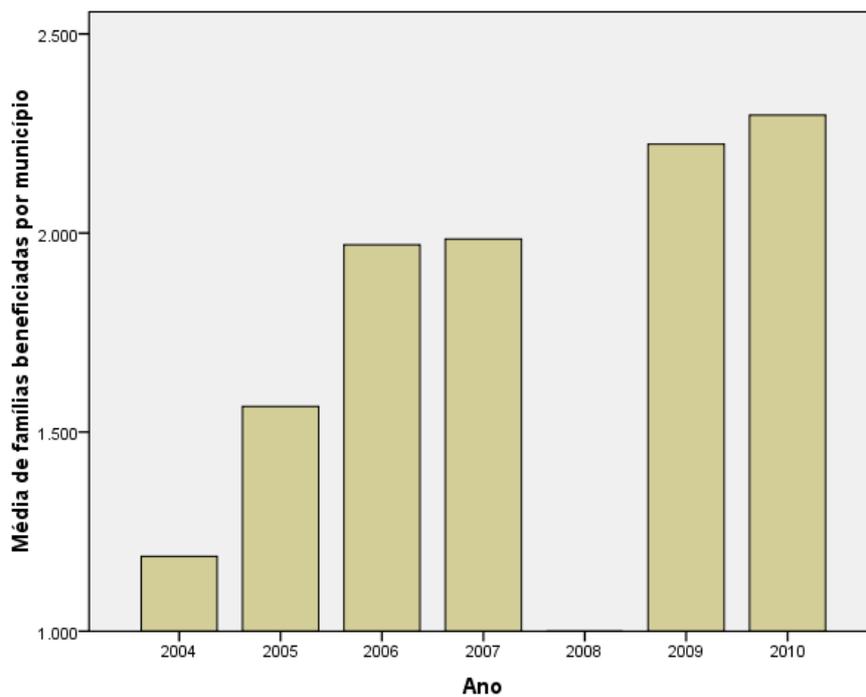


Figura 5. Número de famílias beneficiadas pelo PBF, 2004 a 2010.

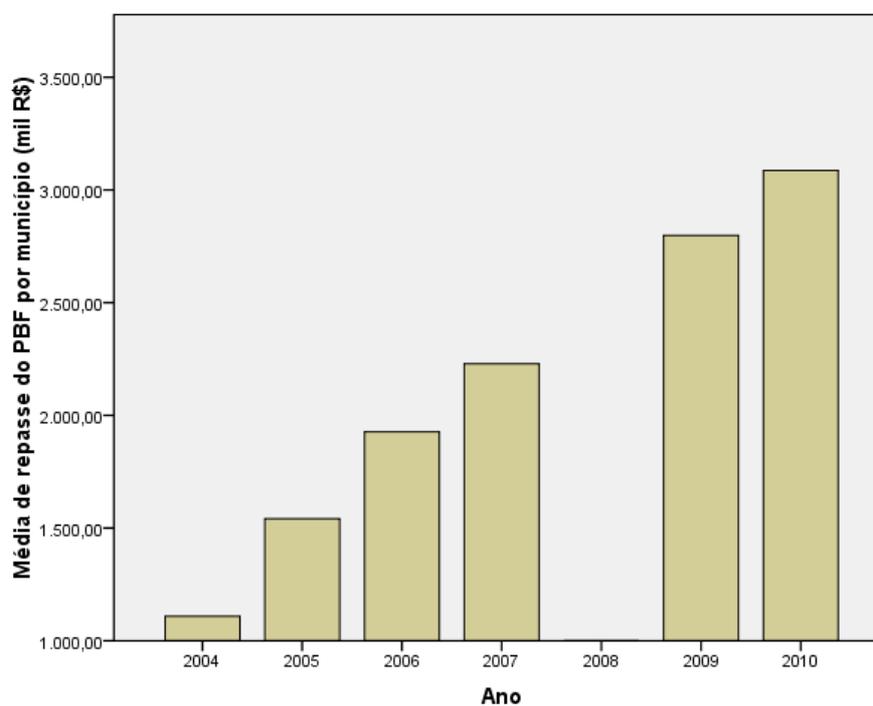


Figura 6. Média dos fundos repassados (mil R\$) para o município, período de 2004 a 2010.

Observando a Figura 7, nota-se queda na cobertura do PBF entre o período de 2006 a 2008. A cobertura (BFC) representa o número de famílias atendidas dividido

pelo número de famílias elegíveis ao programa. Como o número de famílias foi sempre crescente, supõe-se portanto que a população elegível aumentou. Nesse período, não houve alteração dos critérios de elegibilidade. Como já mencionado anteriormente, os dados de BFC apresentam alguma inconsistência em função da presença de percentuais acima de 100%, o que reforça a suspeita de viés no cadastramento da população.

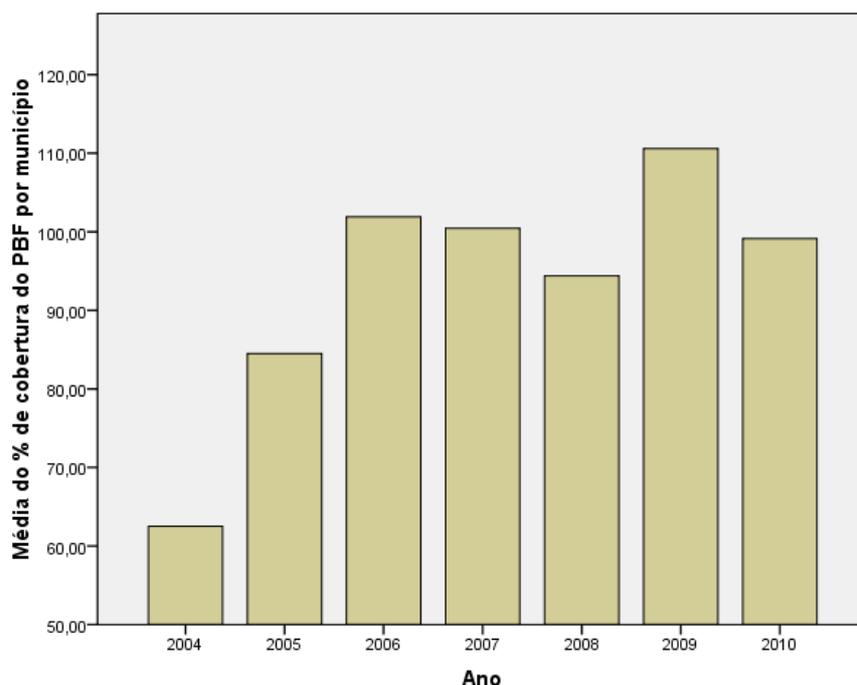


Figura 7. Média do percentual do público elegível que é beneficiário do PBF, 2004 a 2010.

O PBF tem uma ação diferenciada por região do país. O maior número de famílias beneficiadas está nas regiões Norte e Nordeste, o que é condizente com os dados de pobreza disponíveis pelo Censo Demográfico (Figura 8).

Esta mesma informação em nível de estado está disponível na Figura 9. O Distrito Federal foi propositalmente retirado do gráfico, pois os seus dados são muito atípicos em relação aos estados. No DF, o nível de famílias beneficiadas é mais de dez vezes maior que o número médio dos municípios dos outros estados. Esta discrepância ocorre pelo fato do DF, administrativamente, ter apenas um município. As administrações regionais como Ceilândia, Samambaia etc. não são classificadas como município pelo IBGE.

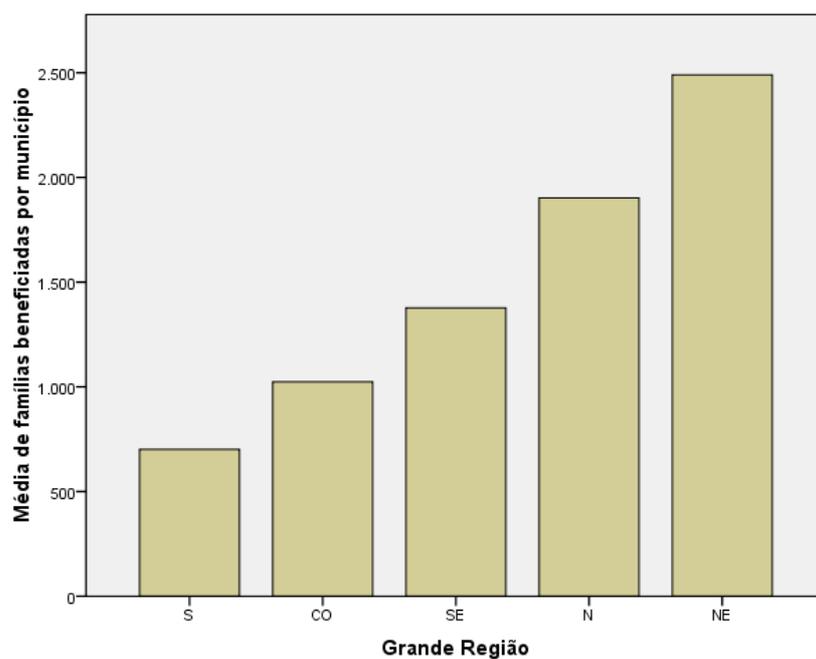


Figura 8. Número médio de famílias beneficiadas por município para cada região do Brasil, 2004 a 2010.

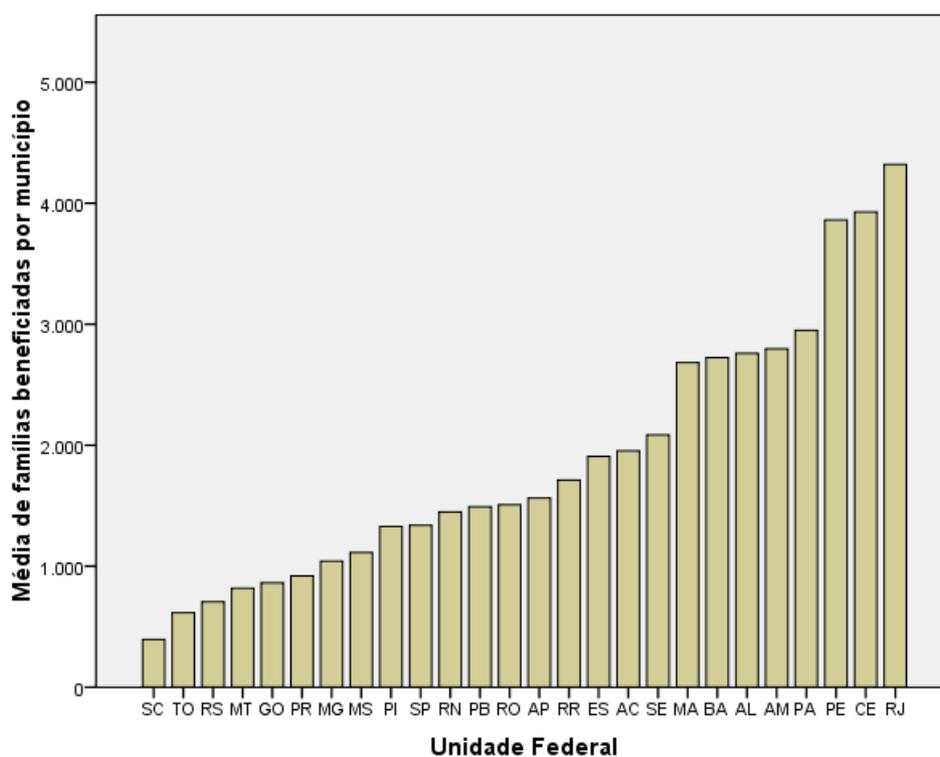


Figura 9. Número médio de famílias beneficiadas por município para cada unidade federal do Brasil, menos o DF, 2004 a 2010.

Nos dados por unidade federal, nota-se que o Rio de Janeiro encabeça a média de beneficiados por município. Comparado com estados economicamente pujantes como São Paulo, Minas Gerais e Paraná, o Rio de Janeiro apresenta números de beneficiados de duas a quatro vezes maior. Será que o PBF estaria equivocado de alguma forma? A Figura 10 permite dirimir esta dúvida.

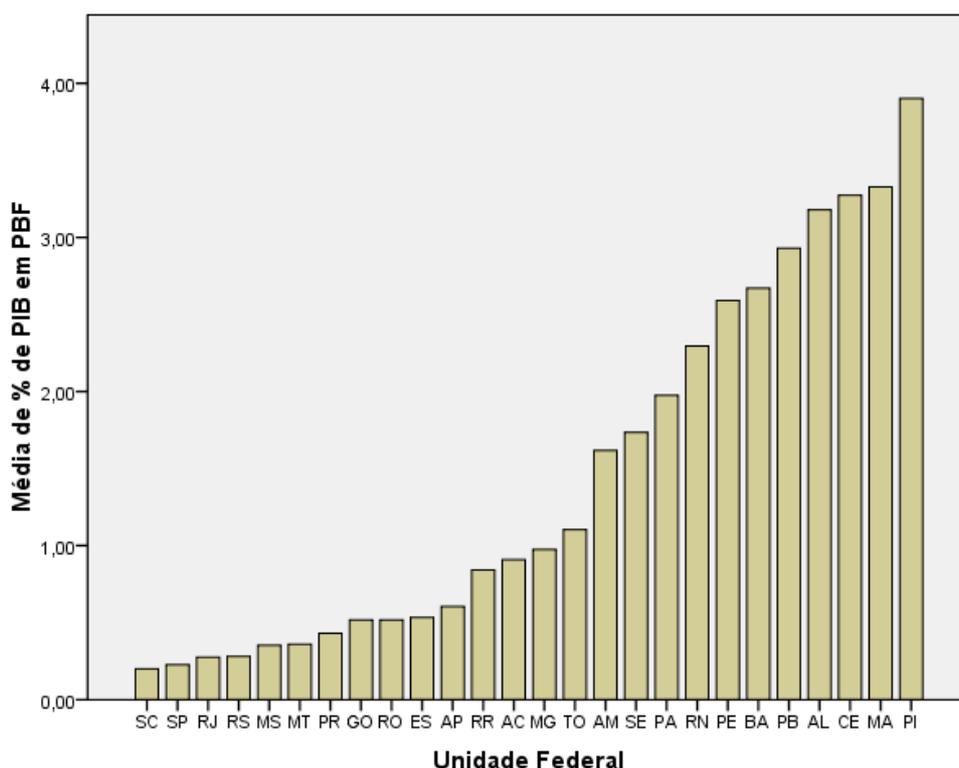


Figura 10. Média percentual do repasse do PBF em termos de PIB da unidade federal, menos o DF, 2004 a 2010.

Na Figura 10 estão os dados ajustados de tal forma que reflitam o fluxo monetário do estado, isto é, a saída de riqueza produzida pelo estado e a entrada de recursos por meio do PBF. Trata-se do percentual obtido da divisão entre o montante repassado pelo PBF ao estado e o valor do PIB correspondente. Neste novo gráfico, nota-se que Piauí passa a encabeçar a lista e o Rio de Janeiro se aproxima de São Paulo. Isto mostra que estados que produzem menos riqueza (menor PIB) estão recebendo maiores montantes do PBF, relativamente. O PIB do estado do Piauí é o menor do Brasil.

Outro grupo de dados estudados na pesquisa está relacionado com os indicadores de educação. Foram escolhidos dois: a proficiência escolar (desempenho) e a distorção idade-série (DIS).

Os dados de proficiência são medidos na 4ª série e na 8ª série para as disciplinas de língua portuguesa e matemática. Depois da mudança da educação fundamental de oito para nove anos, a medida passou a ser dos discentes do 5º e do 9º anos. Estes dados estão divididos por dependência da escola (municipal ou estadual) e sua localização (rural ou urbana). Os dados são bienais, ocorrendo nos anos ímpares. Nos gráficos a seguir, optou-se por não apresentar todas as combinações possíveis de dependência, localização e série/ano. Existe uma certa repetição de resultados entre eles. Escolheu-se algumas combinações que mostram o efeito de dependência, localização, série/ano e disciplina sobre a proficiência escolar.

As ordenadas das Figuras 11 a 14 conservam a mesma grandeza. Nas Figuras 11 e 12, os gráficos representam as proficiências de discentes da 4ª série/5º ano em matemática (Figura 11) e português (Figura 12) em escolas localizadas no meio urbano. As letras (a) e (b) indicam a dependência da escola, estadual e municipal, respectivamente.

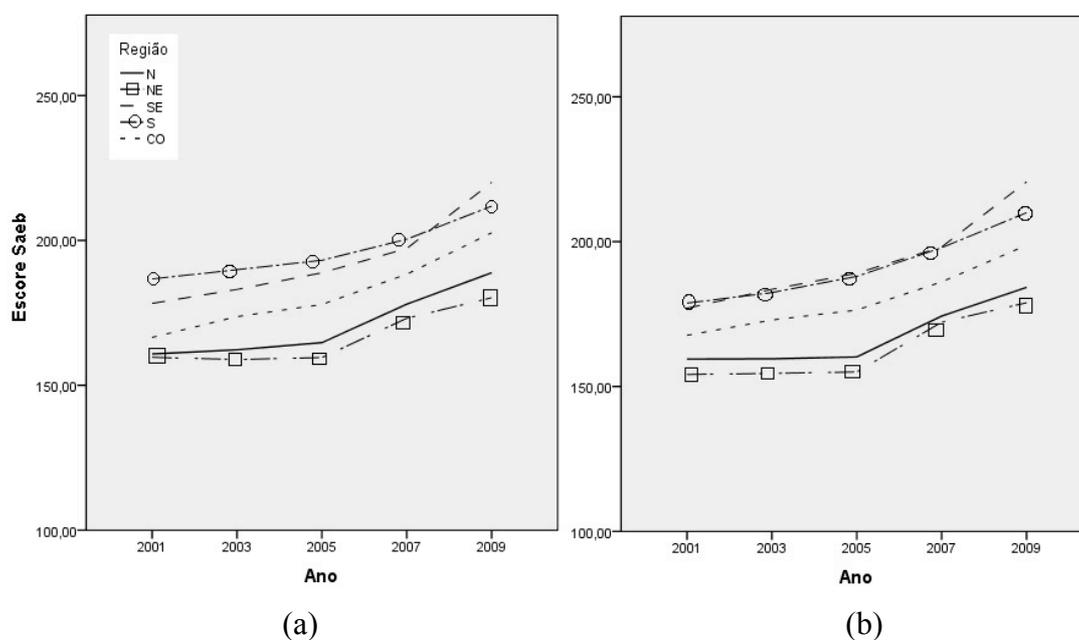


Figura 11. Média de escores de proficiência em **matemática** na 4ª série/5º ano de escolas urbanas **estaduais** (a) e **municipais** (b), nas cinco grandes regiões brasileiras, 2001 a 2009.

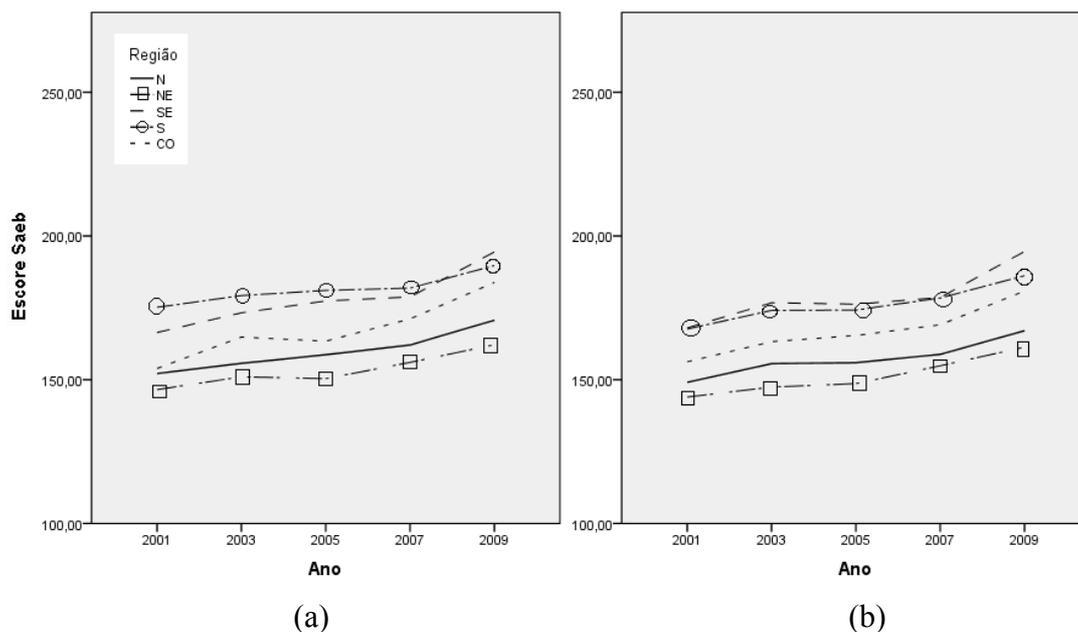


Figura 12. Média de escores de proficiência em **português** na 4ª série/5º ano de escolas urbanas **estaduais** (a) e **municipais** (b), nas 5 grandes regiões brasileiras, 2001 a 2009.

O que se pode notar em todos os quatro gráficos é que predomina a tendência crescente das proficiências, com alguns poucos eventos de estagnação ou leve queda. As escolas estaduais apresentam escores maiores quando comparadas às municipais.

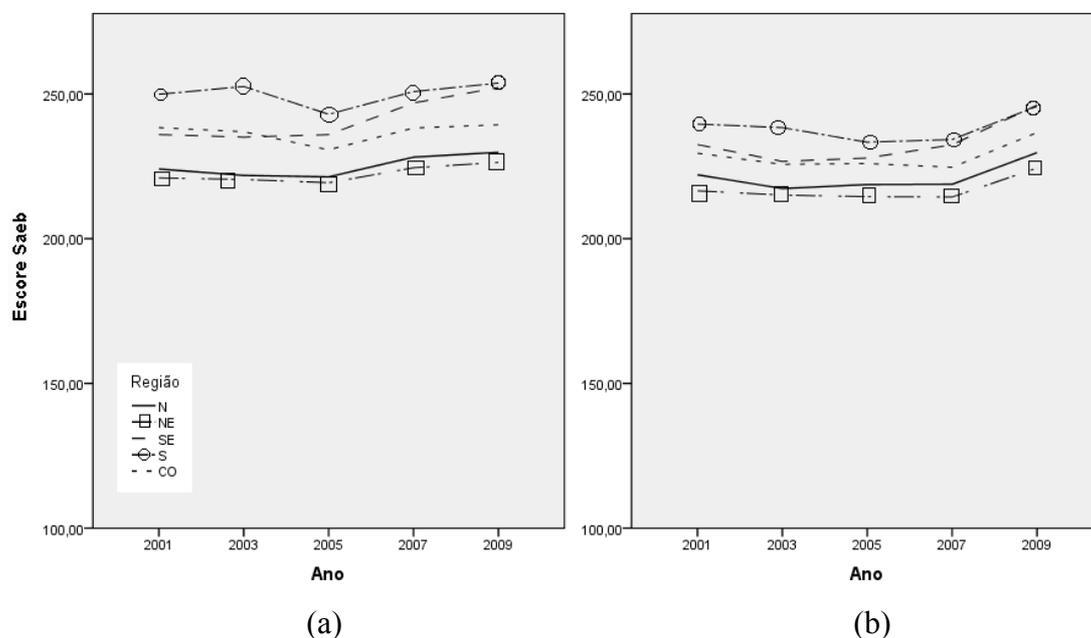


Figura 13. Média de escores de proficiência em **matemática** (a) e **português** (b) na 8ª série/9º ano de escolas **estaduais** urbanas, nas 5 grandes regiões brasileiras, 2001 a 2009.

A Figura 13 traz os dados de proficiência em matemática (a) e português (b) de discentes da 8ª série/9º ano de escolas estaduais urbanas. As notas são bem maiores

que as notas da 4ª série/5º ano, indistintamente da disciplina, indicando consistente melhora do desempenho do discente com a progressão nas séries/anos. Porém, as curvas crescem numa taxa menor que as curvas dos discentes da 4ª série/5º ano. As quedas de proficiência são mais frequentes na 8ª série/9º ano.

Na Figura 14 estão os dados de escolas rurais da 4ª série/5º ano. Existe uma descontinuidade em algumas linhas do gráfico em 2007, um sinal de problema nos dados. Os dados de proficiência dos anos de 2001 a 2005 foram obtidos por amostragem por conveniência. Isto dificulta uma análise mais precisa do escore das instituições rurais. Ignorando a representatividade destes dados, fica evidente o escore menor, comparado ao das instituições em meio urbano. Também se percebe que as diferentes regiões possuem curvas de proficiência mais próximas entre si do que as do meio urbano. Isto indica maior semelhança entre as regiões em termos de desempenho escolar. Isto também pode apontar para uma realidade indistinta entre as populações localizadas no meio rural do Brasil. Se por um lado a linha da fatura pode ser bastante variável de região para região, a linha de pobreza tem um limite bem definido e comum a todos que são pobres.

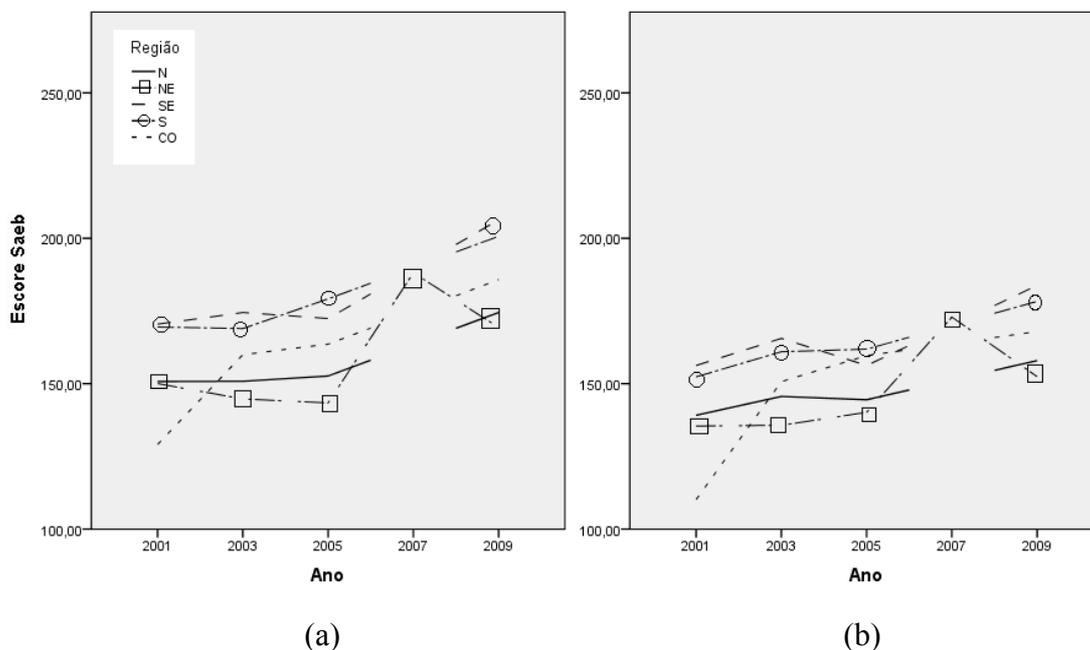


Figura 14. Média de escores de proficiência em **matemática** (a) e **português** (b) na 4ª série/5º ano de escolas **municipais rurais**, nas 5 grandes regiões brasileiras, 2001 a 2009.

A Figura 15 traz a informação das médias totais de desempenho por grande região ao longo do período de 2001 a 2009. Propositadamente, escolheu-se duas realidades bem distintas em termos de dependência, localização, série/ano e disciplina. Como se pode observar, repete-se a ordem das grandes regiões nas duas situações. Isto deixa claro que as regiões Norte e Nordeste, notadamente as mais pobres, são as que apresentam os menores escores médios de proficiência. De forma oposta, as regiões economicamente mais desenvolvidas apresentam as melhores proficiências. Isto aponta para uma relação entre desenvolvimento econômico e desempenho escolar.

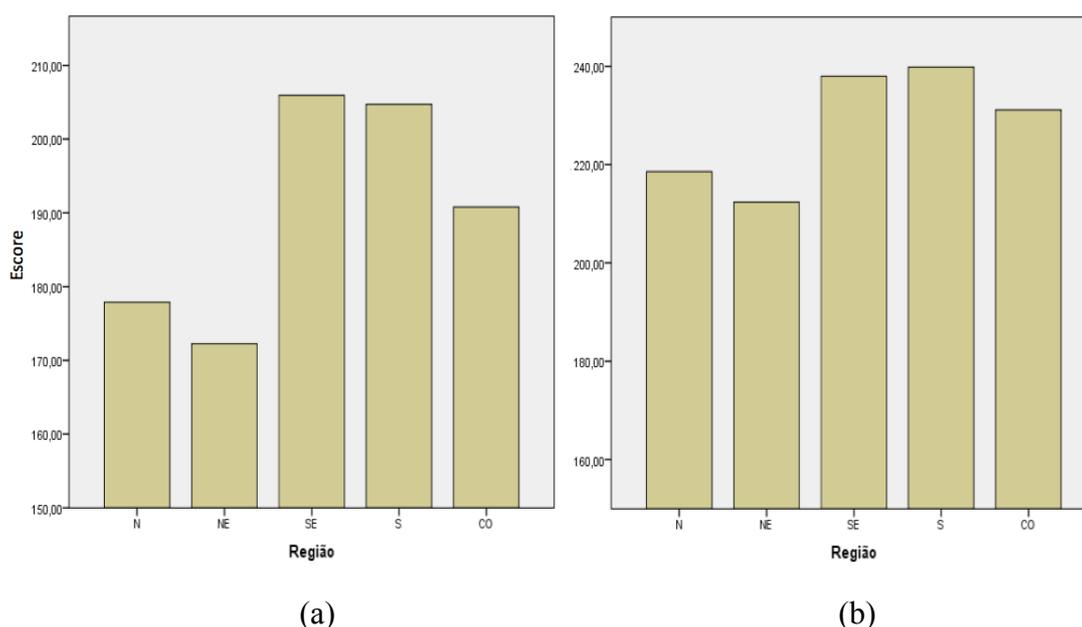


Figura 15. Média de escores de proficiência em **Matemática** na 4ª série/5º ano de escolas **estaduais urbanas** (a) e em **Português** na 8ª série/9º ano de escolas **municipais rurais** (b), nas 5 grandes regiões brasileiras, 2001 a 2009.

Uma outra medida de educação é a distorção idade-série (DIS). Mas antes de apresentar os gráficos referentes à DIS, é preciso compreender a dinâmica das relações municipal / estadual e urbana / rural.

Na Figura 16 observa-se que o número total de instituições de ensino vem caindo desde 2000. Uma queda abrupta ocorreu entre 2006 e 2007, coincidindo com alterações significativas nos indicadores de proficiência. Também chama a atenção o número de instituições municipais, de três a cinco vezes maior que o número de instituições estaduais. A queda no número total de instituições estava intimamente relacionada com a diminuição do número de instituições municipais. Em 2000, eram

150 mil escolas municipais. Em 2010, esse número caiu para 101 mil escolas. Porém, este número de instituições não refletiu o número de matrículas. Em 2005, haviam 28 milhões de alunos matriculados em escolas urbanas e 6 milhões em escolas rurais (Inep, 2007).

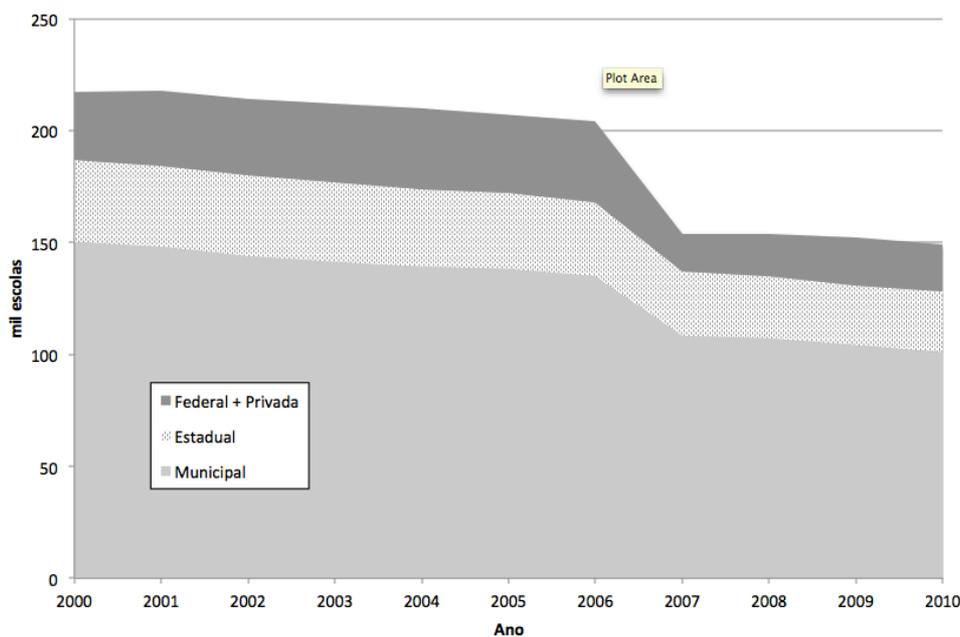


Figura 16. Número acumulado de escolas por tipo de dependência, 2000 a 2010.

Os dados da Figura 17 mostram que praticamente metade das escolas são rurais e outra metade urbana. Dado o número de matrículas mencionado anteriormente, depreende-se que as escolas rurais são menores e mais numerosas.

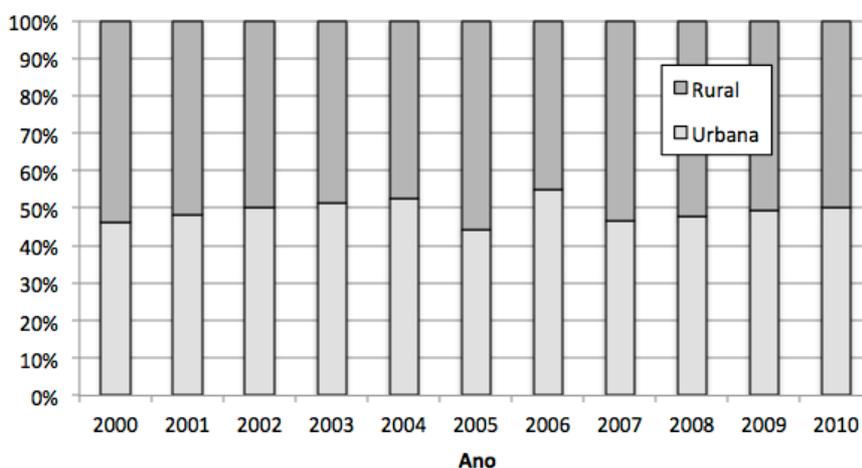


Figura 17. Percentual de escolas por localização, 2000 a 2010.

Esta diferença entre rural e urbano também fica evidente nos dados de DIS. Os dados da presente pesquisa cobrem do 1º ano ao 9º ano, de escolas municipais e estaduais, de áreas rurais e urbanas. Ao todo são 36 combinações de variáveis. Em vista dos dados serem baseados no censo escolar, as informações referentes ao meio rural são mais completas do que os dados rurais de proficiência.

De forma geral, mantém-se a tendência anterior de escolas municipais terem piores indicadores que escolas estaduais. O mesmo acontece em relação ao aspecto rural / urbano. A fim de explorar estes dados, montou-se os gráficos da Figura 18.

Os dados da Figura 18 representam a DIS para escolas municipais rurais, do 2º ano (a) até o 9º ano (h). Não foram utilizados os dados do 1º ano por conterem apenas dados de 2004 a 2010, o que prejudicaria a comparação. A escolha das escolas municipais está na sua maior capilaridade em áreas mais distantes dos grandes centros. Estes dados mostram um aspecto novo. Observa-se uma diminuição na distorção idade-série entre 2002 e 2004, em todos os gráficos. Este período coincide com o lançamento oficial do PBF. Esta tendência também se observa em escolas estaduais e em escolas urbanas, sempre até o 5º ano. Também se observa que existe uma similaridade entre os gráficos de (a) a (d), representando os anos iniciais, e uma similaridade entre os gráficos de (e) a (h), representando os anos finais do ensino fundamental. No primeiro grupo, os anos iniciais mostram um aumento da DIS com o passar do tempo (não se trata de avanço no ano escolar, mas do mesmo ano escolar comparado em diferentes anos). Já no segundo grupo, os anos finais tem a tendência oposta, de diminuir a DIS conforme o tempo passa. Nota-se que existe um ponto de inflexão no segundo grupo, em 2008. Parece um efeito da diminuição no número de estabelecimentos de ensino observado em 2007, conforme a Figura 16. Mas deve-se ressaltar que em 2007, por coincidência, os dados dos alunos fornecidos pelo Censo Escolar passam a ser disponibilizados de forma diferente, organizados por estado. O Manual do Usuário não menciona qualquer modificação metodológica. Não se sabe o motivo do pico da distorção idade-série (DIS) para o ano de 2001. Este pico se observa principalmente nos anos iniciais, tanto de escolas urbanas e rurais, municipais e estaduais. Averiguou-se por erros de cálculo ou de recuperação do dados da base, mas não se encontrou qualquer problema.

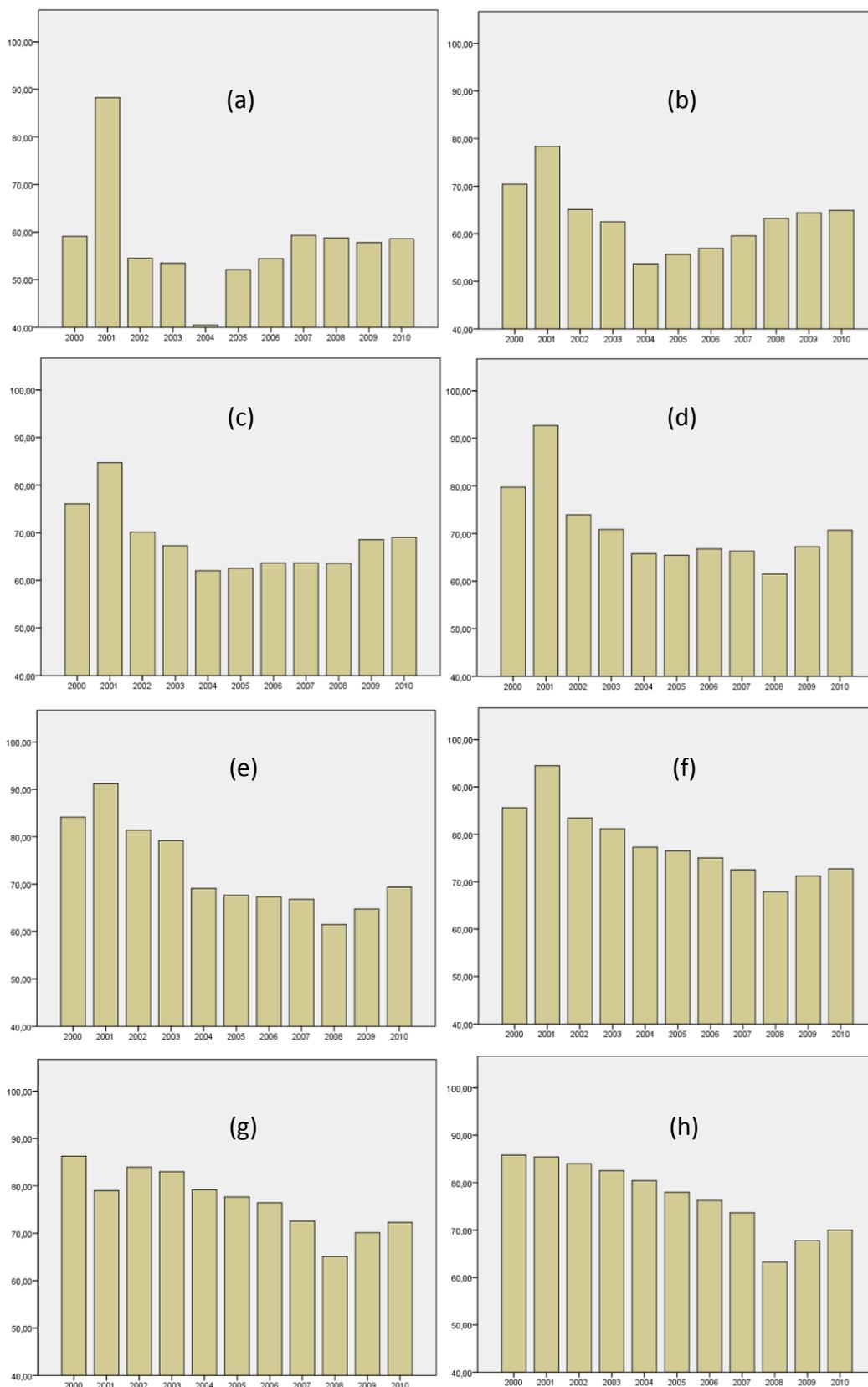


Figura 18. Distorção idade-série (%), nas escolas municipais rurais, do 2º ano (a) até o 9º ano (h), 2000 a 2010.

Do ponto de vista geográfico, há diferenças entre as regiões brasileiras. Porém esta diferença varia dependendo da série/ano analisada. Na Figura 19 estão os dados da 1ª série/2º ano de escolas estaduais urbanas. Uma observação que surpreende é que a DIS teve a região Sul com uma das piores desde 2007. Também surpreende que a região Norte tenha valores similares à região Sudeste desde 2006. E o Centro-Oeste apresentou melhora notável em anos recentes, apresentando as menores DIS do país.

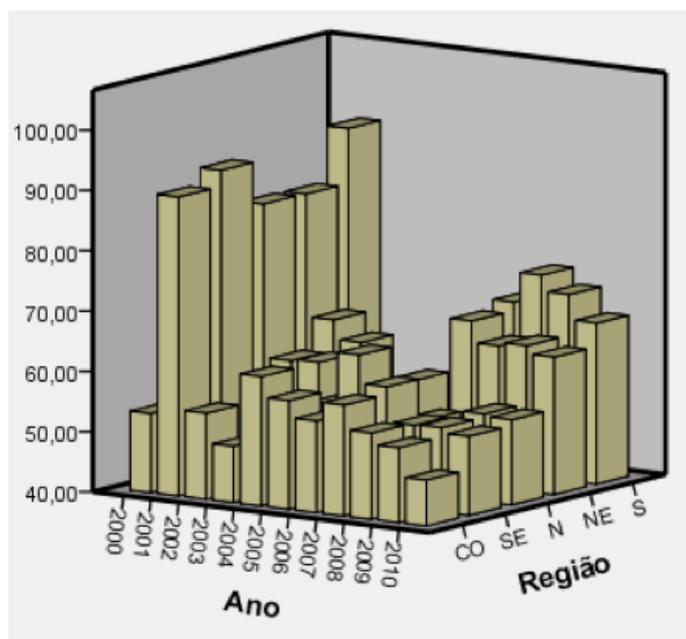


Figura 19. Média dos percentuais de DIS em escolas estaduais urbanas, turma de 1ª série/2º ano, nas 5 grandes regiões brasileiras, 2000 a 2010.

Na Figura 20 estão os dados da 8ª série/9º ano desse mesmo tipo de instituição. Nota-se que a DIS diminuiu para todas as regiões. A ordem das regiões também acompanha a ordem dos dados de proficiência, o que pode indicar uma relação entre estas duas variáveis. As regiões Norte e Nordeste apresentam os piores indicadores, o Centro-Oeste fica intermediário e as regiões Sul e Sudeste ficam com os melhores resultados. Nas regiões Sul, Sudeste e Centro-Oeste, a DIS vem diminuindo ao longo dos anos. Norte e Nordeste tiveram esta tendência interrompida em 2008.

Do ponto de vista da saúde, apesar da tendência de queda nos óbitos por diarreia, entre 2000 e 2006, os valores se mantiveram relativamente os mesmos (Figura 21). Já os valores de mortalidade infantil foram sempre decrescentes (Figura 22).

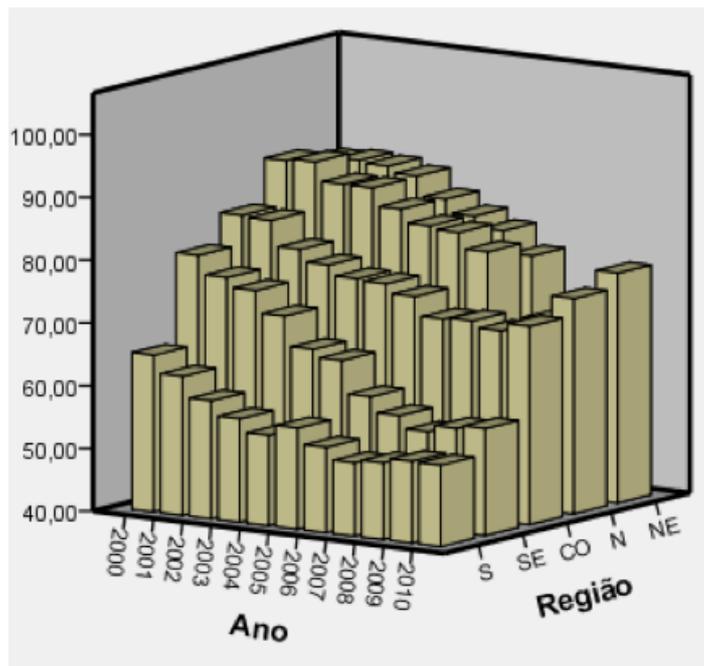


Figura 20. Média dos percentuais de DIS em escolas estaduais urbanas, turma de 8ª série/9º ano, nas 5 grandes regiões brasileiras, 2000 a 2010.

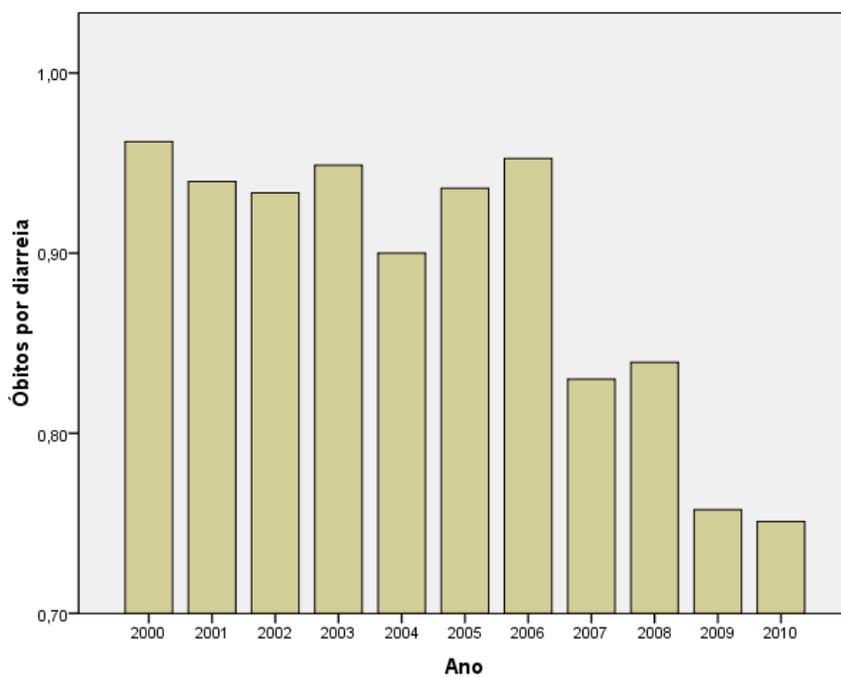


Figura 21. Média de óbitos por diarreia nos municípios, 2000 a 2010.

Do ponto de vista geográfico, as regiões Sul e Sudeste apresentaram os melhores indicadores de saúde, com o Centro-Oeste numa faixa intermediária e o

Norte e Nordeste com os piores resultados. Porém, desmembrando os indicadores por unidade federal, nota-se alguns resultados inesperados. Ponderando os dados por tamanho da população do município, obteve-se o gráfico da Figura 23. Neste gráfico a posição do DF chama a atenção próximo a estados do Norte e Nordeste e a posição de Goiás com o menor número de óbitos.

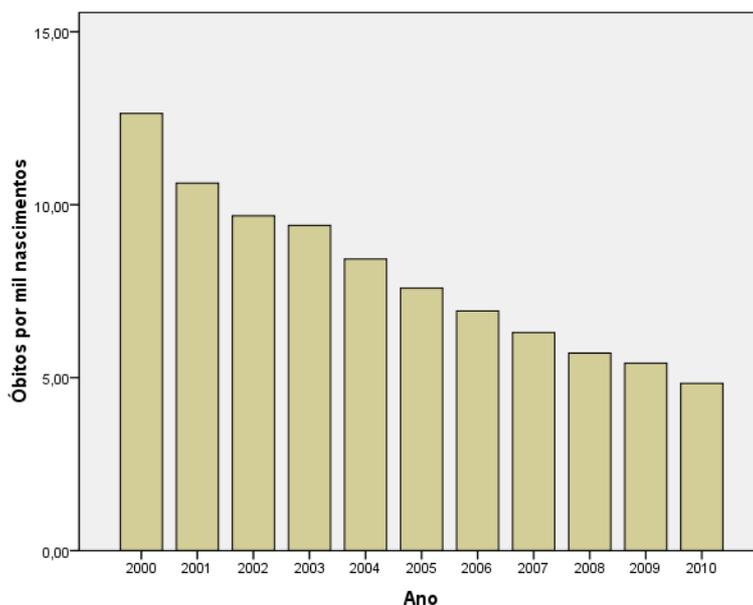


Figura 22. Média da mortalidade infantil nos municípios, 2000 a 2010.

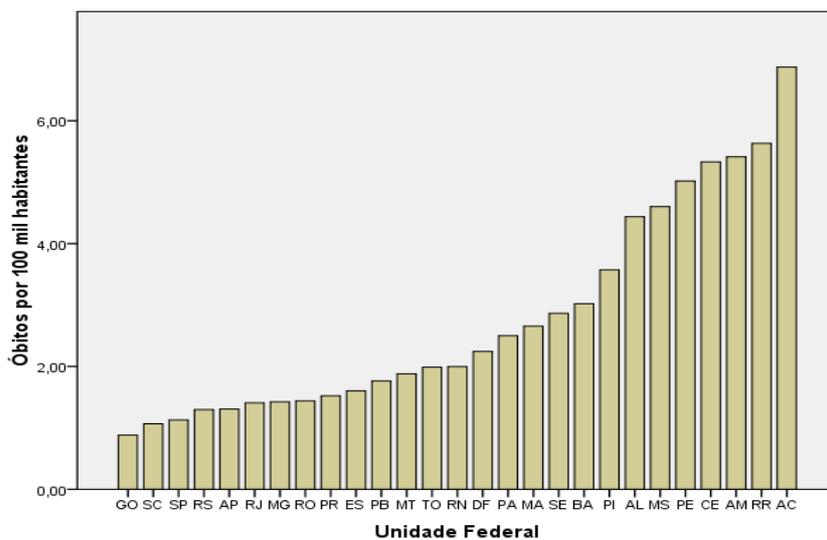


Figura 23. Média de óbitos por diarreia por 100 mil habitantes nos municípios, 2000 a 2010.

Na Figura 24, novamente o DF aparece em destaque, com a maior média de mortalidade infantil para o período. Goiás repete o bom desempenho no indicador de mortalidade infantil, seguido de estados do Nordeste como Rio Grande do Norte, Sergipe e Paraíba.

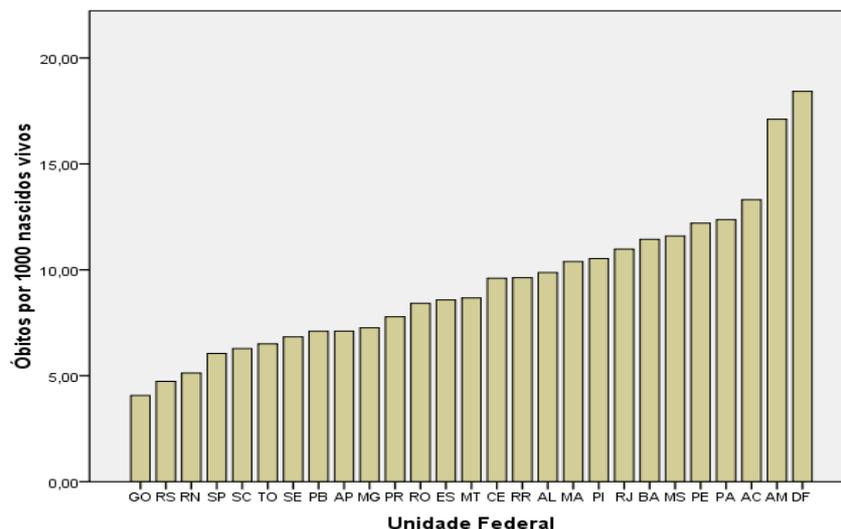


Figura 24. Média de mortalidade infantil nos municípios, 2000 a 2010.

Entretanto, há que se mencionar que o DF sofre de um problema antigo com as cidades goianas e mineiras da Região do Entorno, em termos de saúde. É sabido a vinda de ambulâncias trazendo pacientes para os hospitais do DF devido à falta de médicos, recursos ou especialidades nas cidades de origem. Isto talvez seja um dos responsáveis pelos indicadores ruins apresentados.

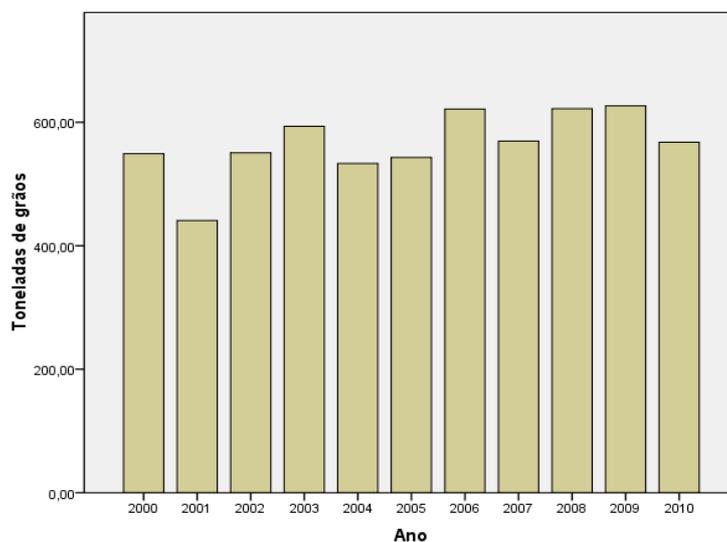


Figura 25. Média de produção de feijão nos municípios, 2000 a 2010.

De forma geral, os produtos agropecuários aumentaram a produção ao longo dos anos, com exceção do feijão. Ele se manteve relativamente estável no período, com uma queda mais acentuada em 2001 (Figura 25).

Apesar de leite, milho e soja estarem crescentes, o milho apresentou maiores variações para o período, alternando anos de aumento e diminuição da produção, com exceção de 2005 (Figura 26).

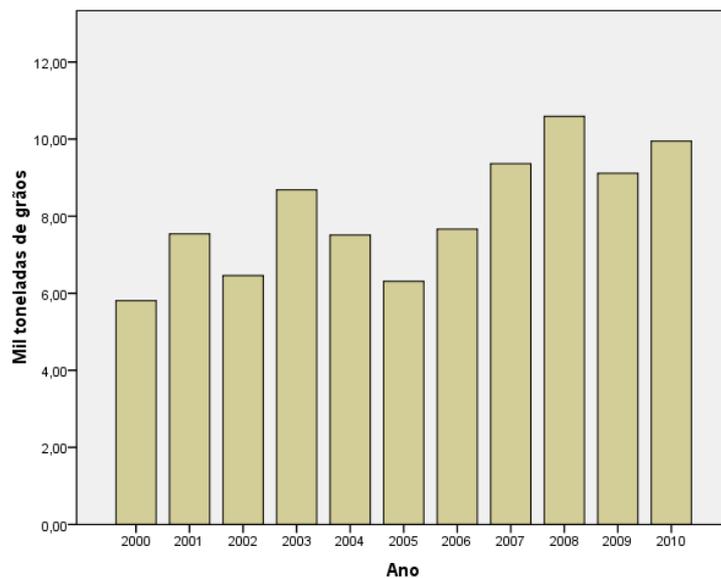


Figura 26. Média de produção de milho nos municípios, 2000 a 2010.

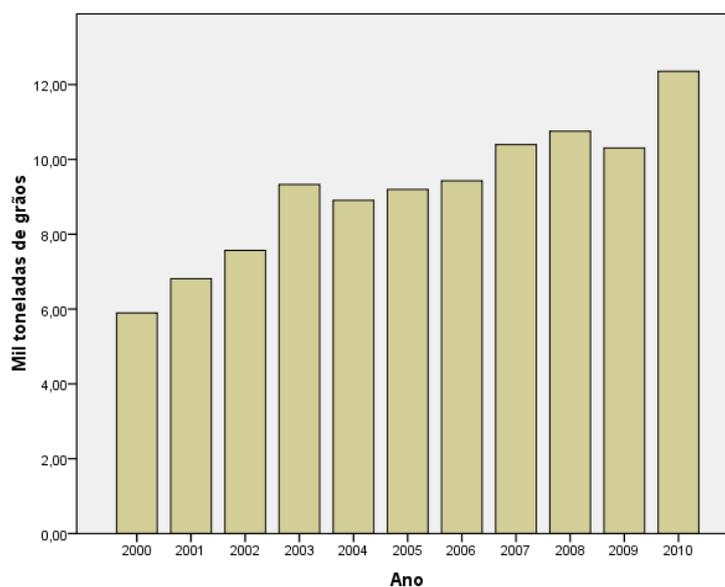


Figura 27. Média de produção de soja nos municípios, 2000 a 2010.

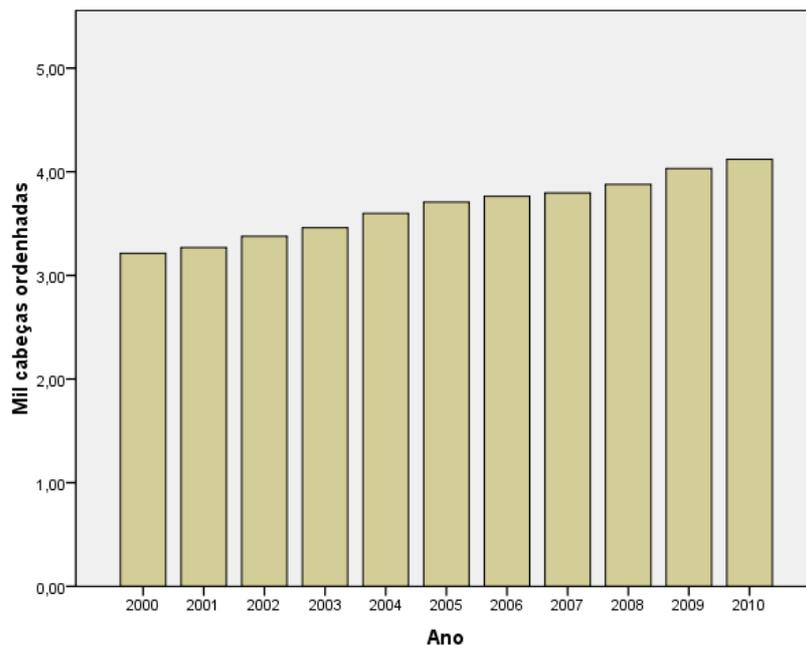


Figura 28. Média de vacas ordenhadas nos municípios, 2000 a 2010.

Soja (Figura 27) e leite (Figura 28) apresentaram produções crescentes e mais estáveis. A produção agropecuária é dependente de clima e solo para o seu desenvolvimento. Quanto mais tecnificada a produção, mais específica a produção é para uma dada região. Independente da quantidade produzida, as culturas existem em praticamente todas as regiões brasileiras. A produção de leite é a melhor distribuída pelo país (Figura 29), enquanto a produção de soja é a mais concentrada, principalmente no Centro-Oeste (Figura 30).

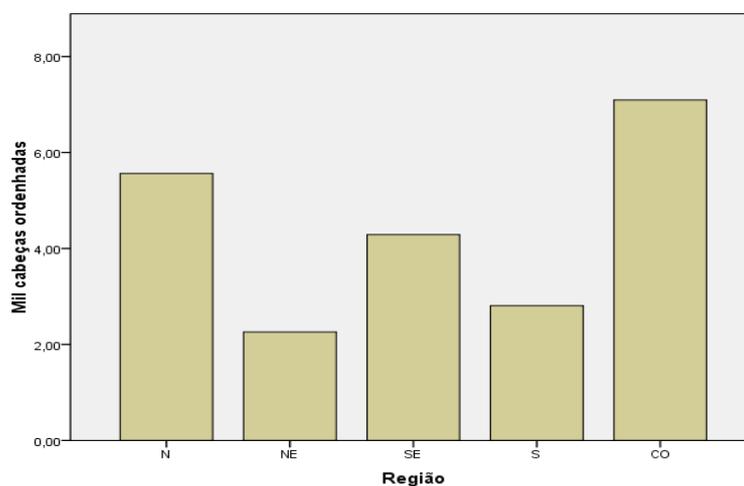


Figura 29. Média de vacas ordenhadas nos municípios, por região, 2000 a 2010.

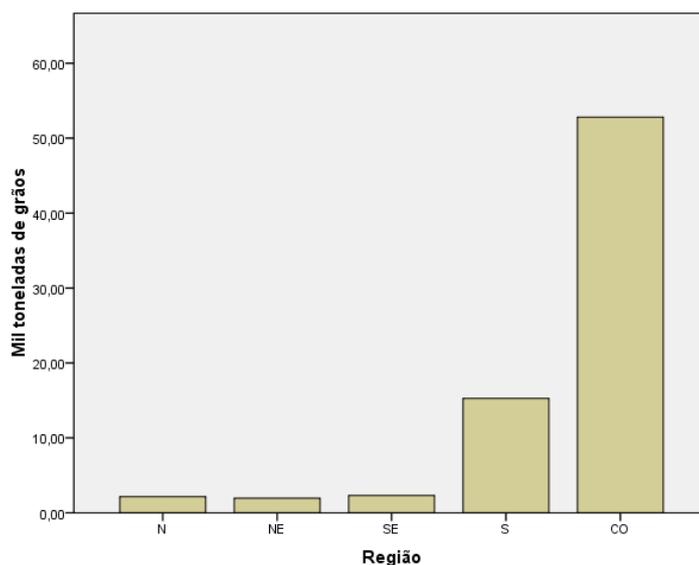


Figura 30. Média de produção de soja nos municípios, por região, 2000 a 2010.

O último grupo é o das variáveis econômicas. Combinando as variáveis referentes ao Valor Adicionado Bruto, obteve-se a Figura 31. Regiões Sul, Norte e Centro-Oeste possuem os maiores percentuais de PIB oriundo do setor agropecuário (VARP). Sudeste destaca-se no setor industrial (VIRP) e de serviços (VSRP). Junto com o Centro-Oeste, são as regiões que mais riqueza produzem com a arrecadação de impostos. Em termos de riqueza oriunda de serviços ligados a seguros, administração pública, educação e saúde, as regiões Nordeste e Norte se destacam.

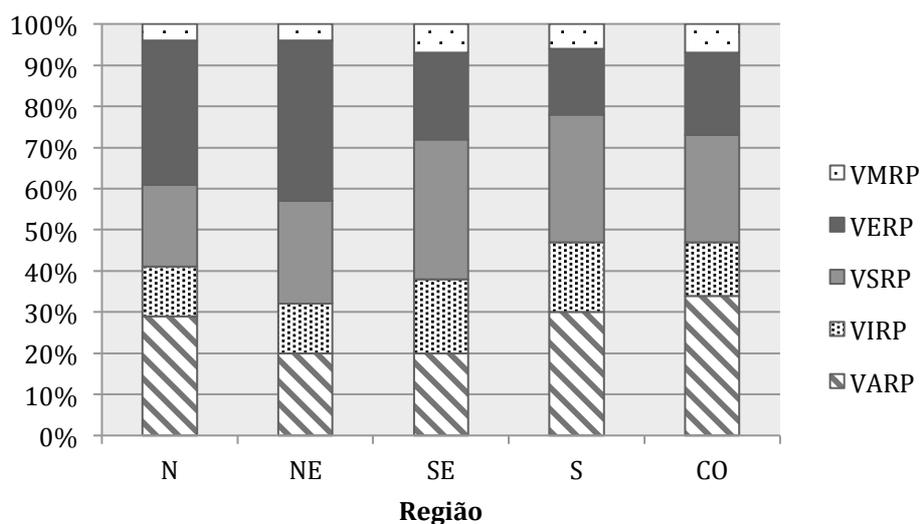


Figura 31. Média do percentual dos Valores Adicionados Brutos dos municípios por setor, agrupados por região, 2000 a 2010.

Na Figura 32 é possível observar a variação dos Valores Adicionados no período estudado. A riqueza oriunda do setor industrial e da arrecadação de impostos foram os que menos variaram ao longo do tempo. O único que sofreu queda nos percentuais do PIB foi o setor agropecuário. Isso se deve à queda dos preços agrícolas. Apesar dos recordes seguidos de produção, a riqueza gerada é menor, pois o preço pago por unidade do produto é menor. Graças à tecnificação da produção agrícolas (maior emprego de técnicas modernas de produção), os custos de produção caem. Os que sofreram algum incremento foram os setores de serviços (comércio, alojamento, alimentação, transporte, comunicação e imobiliário) e os que envolvem outros serviços (seguro, saúde, administração pública e educação).

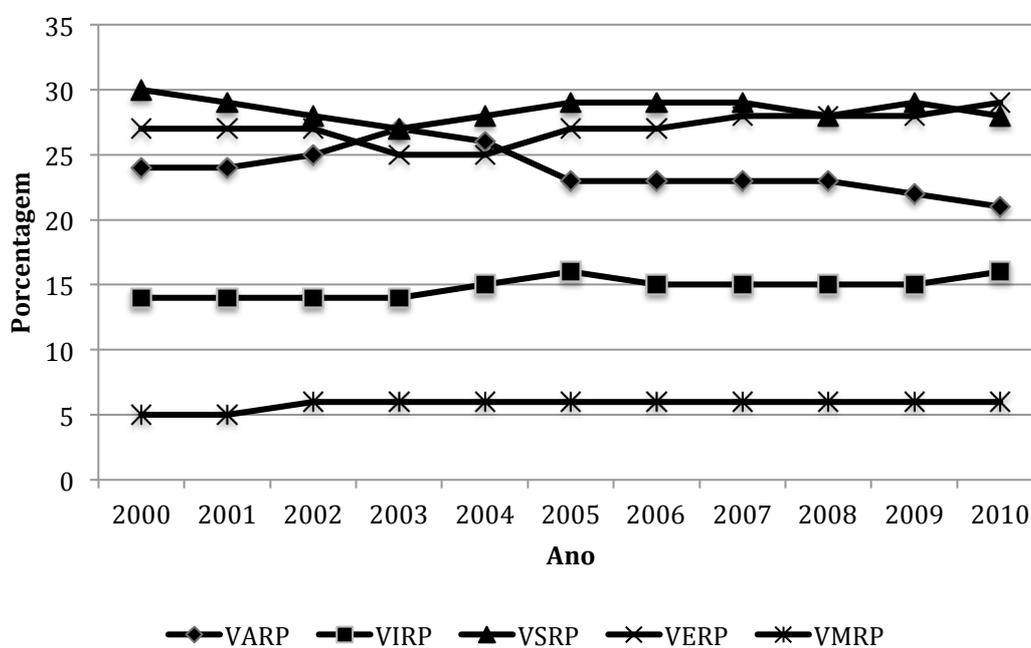


Figura 32. Média do percentual dos Valores Adicionados Brutos dos municípios por setor, 2000 a 2010.

Um aspecto marcante da década de 2000 é a manutenção da inflação a patamares relativamente baixos. Uma medida da inflação é o Índice Nacional de Preços ao Consumidor (INPC), calculado pelo IBGE a partir de levantamentos de preços realizados em onze capitais brasileiras. Na Figura 33 apresenta-se três categorias de despesas das nove levantadas pelo INPC e suas respectivas inflações. Nota-se que o INPC para a educação e saúde é parecido. Alimentação é o que possui maiores picos de alta das nove categorias. Normalmente, a alta dos alimentos está relacionada com quebra de safra e a baixa está relacionada com recordes de produção

agropecuária. Chama a atenção a coincidência dos picos de alta dos alimentos com os dados de educação. Ocorreram nos anos 2002 e 2007/2008, anos próximos aos anos em que se observou inflexões nos dados de DIS, proficiência e dados do PBF.

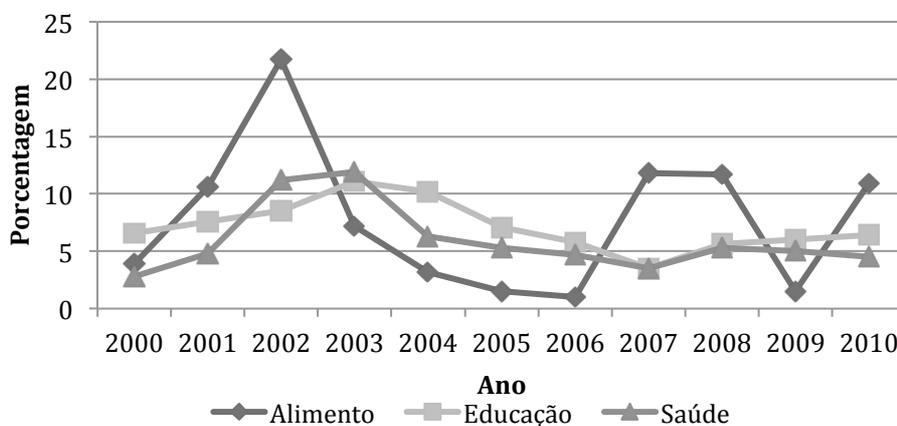


Figura 33. Média do percentual da inflação anual (INPC) em onze capitais brasileiras em alimentação, educação e saúde, 2000 a 2010.

A inflação não é uniforme nas onze capitais. Observando-se a Figura 34, nota-se diferenças que chegam a mais de dois pontos percentuais. A princípio não se pode inferir uma relação geográfica com as variações da inflação. As capitais mais próximas ou situadas na mesma região brasileira não se agrupam.

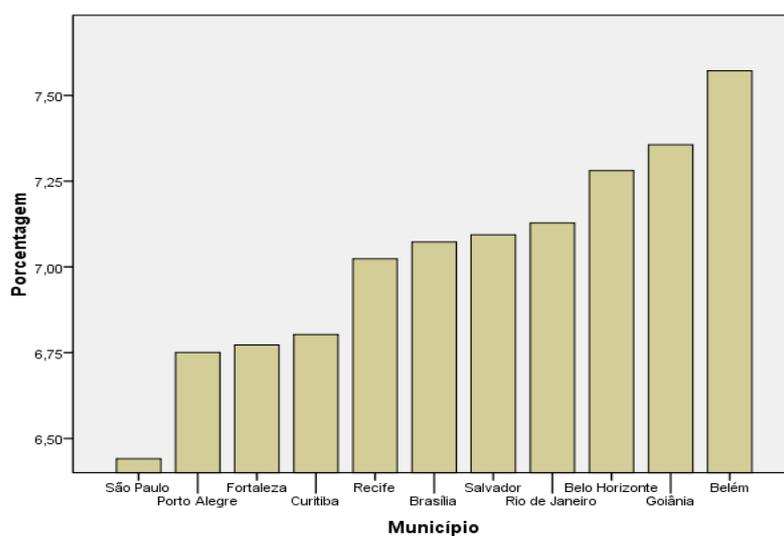


Figura 34. Média do percentual da inflação anual (INPC) em onze capitais brasileiras, 2000 a 2010.

5.3 Resultados dos modelos multinível

Antes de introduzir os resultados dos modelos, sugere-se que o leitor estude o capítulo 2 de Hox (2010) para se familiarizar com os aspectos básicos de um modelo de regressão multinível de dois níveis. Também sugere-se a leitura do capítulo 5, subitem 5.1, sobre análise de dados longitudinais com eventos fixos e variáveis. Isto ajudará no entendimento dos resultados a seguir.

A primeira análise incluiu cinco modelos denominados de Modelo M5-Cx. Estes modelos relacionaram o PIB per capita (variável dependente) a cada um dos componentes apontado na Tabela 15, a saber: proficiência escolar do meio urbano (C1); distorção idade-série das escolas estaduais (C2); distorção idade-série das escolas municipais (C3); bem-estar (C4); e agropecuária (C5). Na Tabela 16 estão os resultados dos cinco modelos.

Tabela 16

Modelos multinível M5-Cx, coeficientes padronizados, 2007 e 2009.

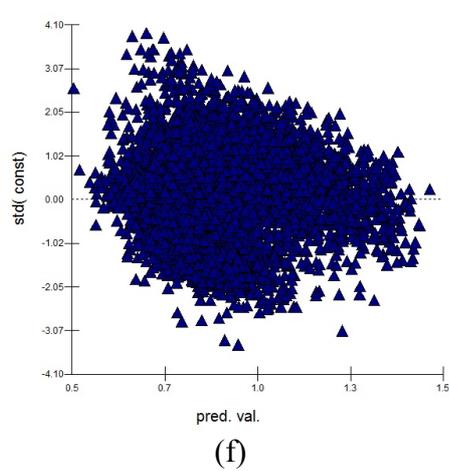
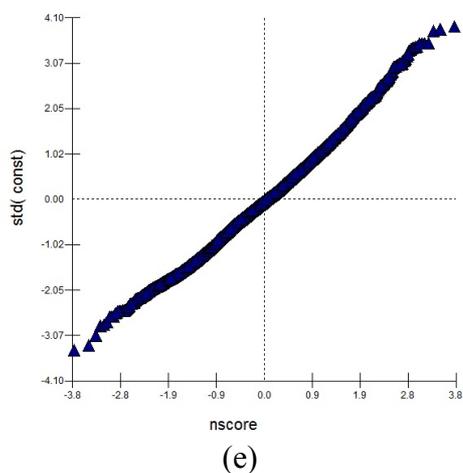
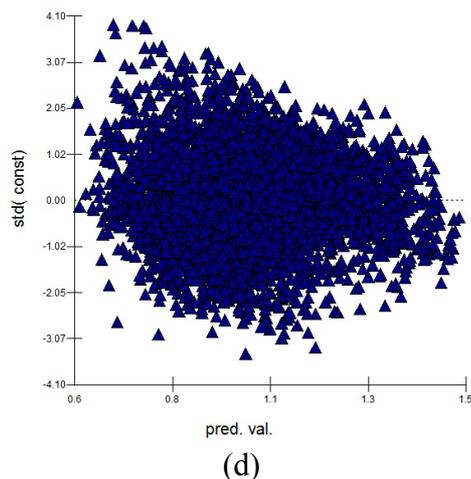
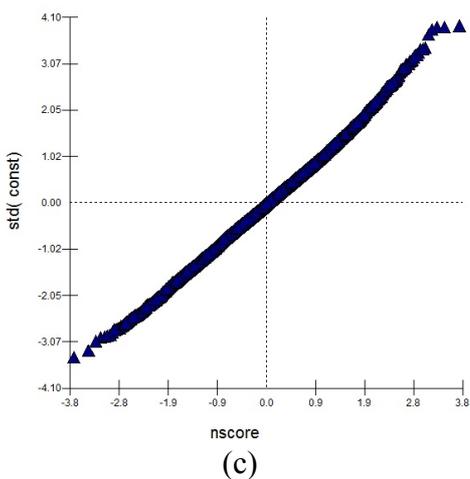
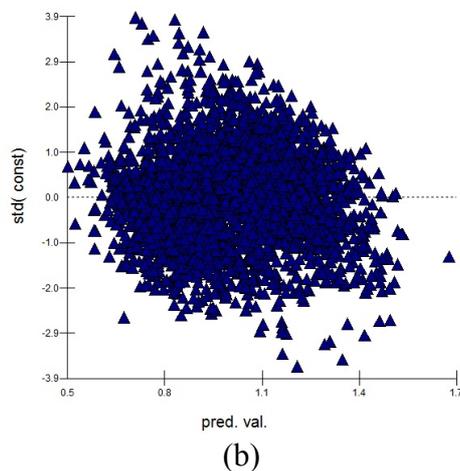
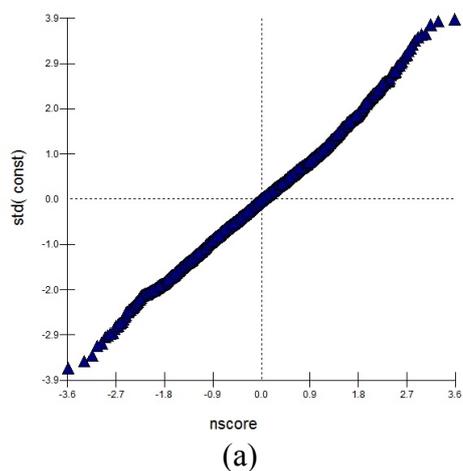
Modelo	Vazio	M5-C1	M5-C2	M5-C3	M5-C4	M5-C5
Resíduo N2	0,8207*	0,4638*	0,5592*	0,6186*	0,3391*	0,5176*
Resíduo N1	0,0484*	0,0921*	0,0545*	0,0946*	0,0552*	0,0564*
ICC	0,9443					
Deviance	29.505	6.712	12.310	15.309	19.067	23.532
R ² nível 2	-	0,4349	0,3186	0,2463	0,5868	0,3693
Casos válidos	10.957	2.988	5.251	6.134	10.007	10.326
Casos totais	11.130	11.130	11.130	11.130	11.130	11.130

Nota. (*) teste *t*-Wald significativo a $p > 1,96$.

O que mais chama a atenção em todos os modelos M5-Cx são os resíduos do nível 1 (N1) inferiores a 0,1. Praticamente não existe variância entre os anos 2007 e 2009. Ainda assim, foi possível calcular os resíduos para o nível 2 (N2), o que permitiu avaliar a linearidade, normalidade e homocedasticidade multivariadas. Os gráficos de resíduos estão na Figura 35.

Os componentes analisados apresentam relação linear e distribuição normal multivariada de maneira geral. Observa-se leve tendência de heteroscedasticidade, principalmente nos gráficos (d), (f) e (j) da Figura 35. Em torno de oito valores atípicos multivariados também ficam evidentes no gráfico (h). Existe grande variação no número de casos válidos, principalmente nos modelos M5-C1 (proficiência) e M5-C2 (distorção idade-série de escolas estaduais). O modelo M5-C5 apresentou o maior *deviance*, logo, o pior ajuste dos cinco modelos. A correlação intraclasse (ICC) foi de 94% por causa da baixa variância entre os anos (nível 1). Assim, toda a variabilidade

ficou carregada na diferença entre os municípios e não na diferença entre os anos. Este valor de ICC justifica a análise em dois níveis, mas alerta que a pesquisa longitudinal aqui proposta não traz maiores vantagens que uma pesquisa transversal, com apenas um ano. Na Tabela 17 estão os resultados dos modelos M1, M2, M3 e M4.



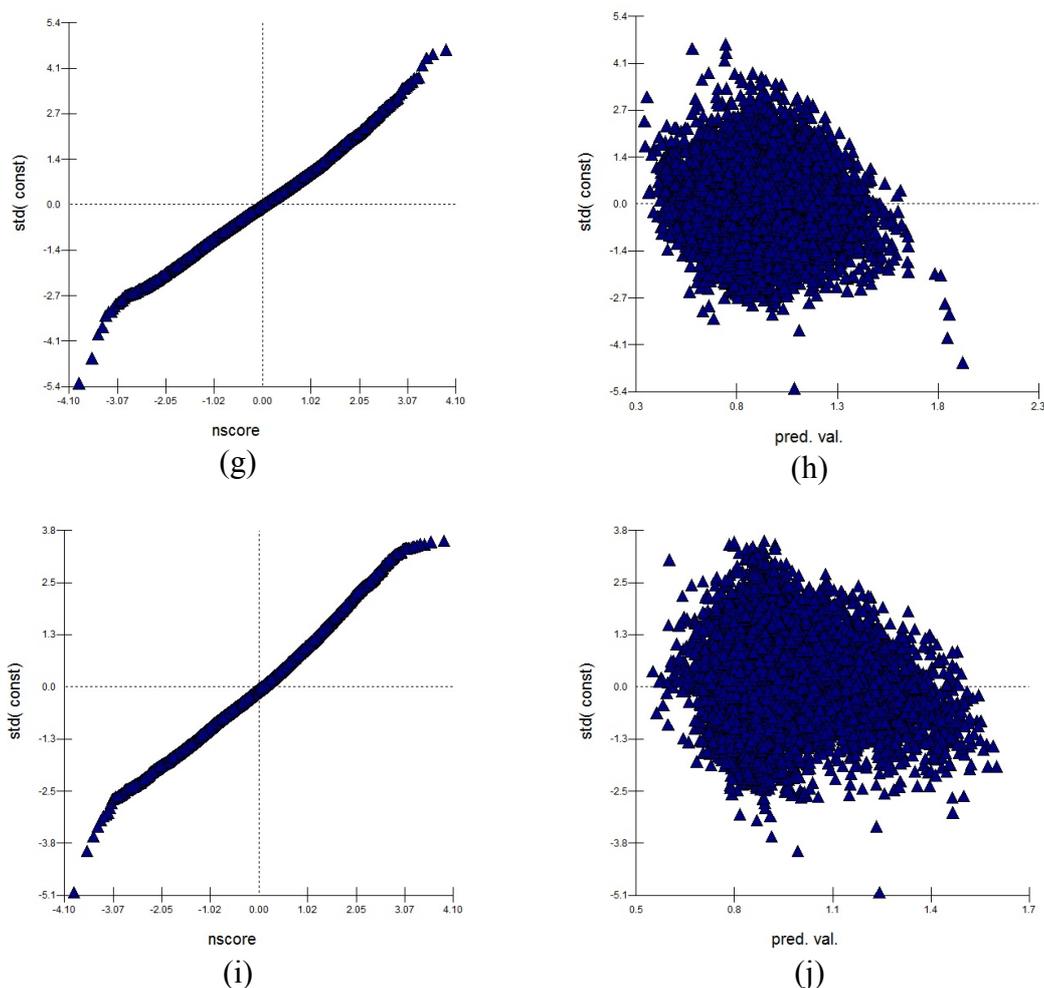


Figura 35. Gráficos de resíduos padronizados *versus* escores normais (a, c, e, g, i) e resíduos padronizados *versus* valores esperados da parte fixa (b, d, f, h, j) do modelo M5-C1 (a, b), M5-C2 (c, d), M5-C3 (e, f), M5-C4 (g, h) e M5-C5 (i, j), nível 2 (municípios), 2007 e 2009.

Para os modelos M2, M3 e M4, os gráficos de resíduos estão na Figura 36. Os resíduos de M1 foram estudados por meio dos seus componentes na Figura 35. Os pressupostos de linearidade e normalidade multivariada estão atendidos de forma geral. O gráfico (d) apresenta alguma heteroscedasticidade. Gráficos (b) e (f) apresentam também dois valores atípicos em torno da abscissa. De qualquer forma, as análises destes modelos não ficam comprometidas, podendo se assumir linearidade na relação entre as variáveis dependentes e explicativas e distribuição normal dos modelos estudados. A variância entre os anos é relativamente homogênea, com um algum problema quando o período de análise é mais curto, que é o caso do gráfico (d) do modelo M3 (2000 a 2003). Os coeficientes β das variáveis estão na Tabela 18.

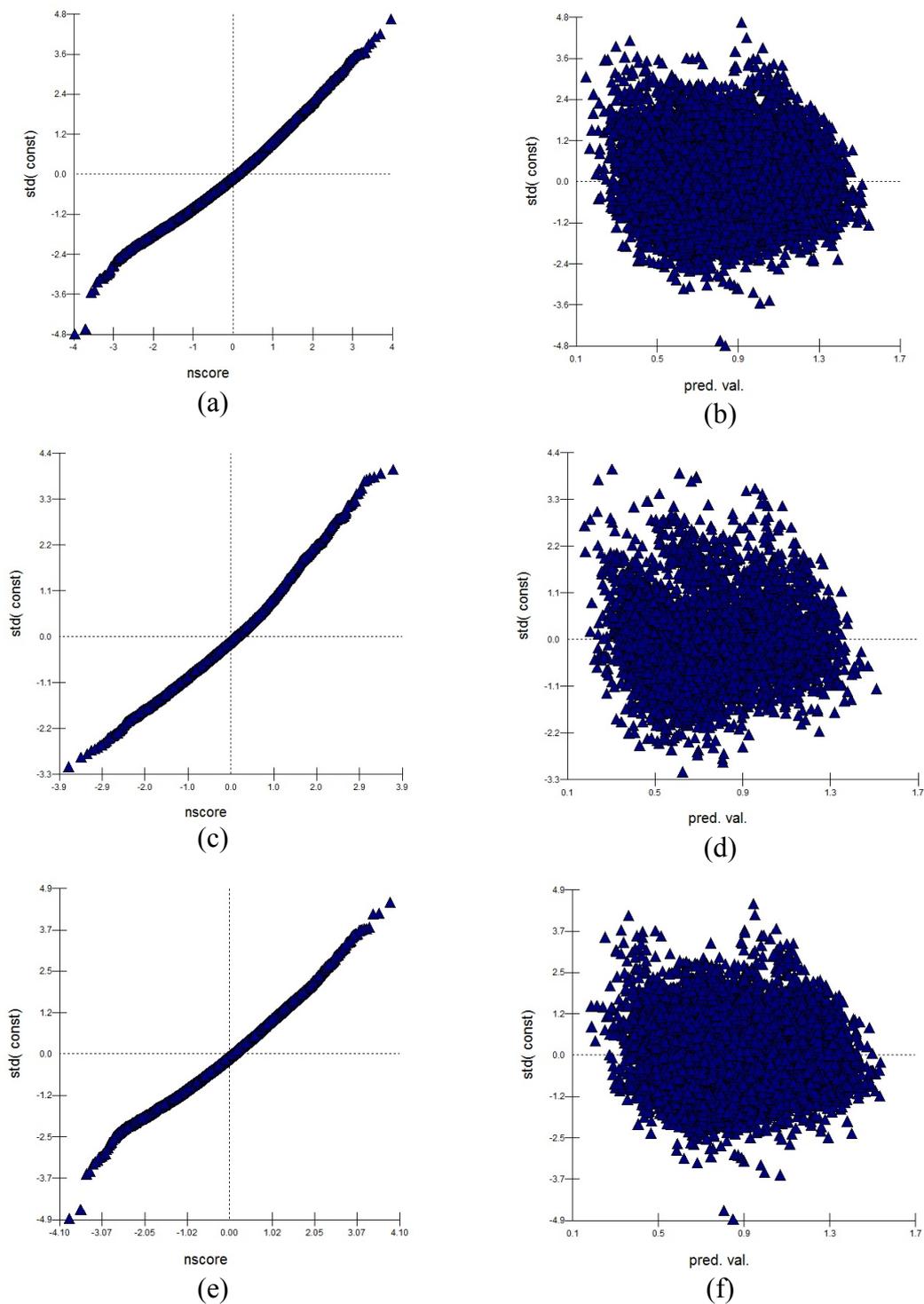


Figura 36. Gráficos de resíduos padronizados *versus* escores normais (a, c, e) e resíduos padronizados *versus* valores esperados da parte fixa (b, d, f) dos modelos M2 (a, b), M3 (c, d) e M4 (e, f), Nível 1 (anos).

Tabela 17*Modelos multinível com diferentes períodos de análise longitudinal, coeficientes padronizados.*

Modelo	2007 e 2009		2000 a 2010		2000 a 2003		2004 a 2010	
	Vazio	M1	Vazio	M2	Vazio	M3	Vazio	M4
Resíduo N2	0,8233*	0,1549*	0,9523*	0,2597*	0,9648*	0,2506*	0,9397*	0,2646*
Resíduo N1	0,0434*	0,0241	0,0465*	0,0271*	0,0111*	0,0212*	0,0431*	0,0228*
ICC	0,9499	-	0,9534	-	0,9886	-	0,9561	-
Deviance	29.413	2.189	170.071	21.972	61.222	6.912	107.648	14.822
R ² nível 2	-	0,8119	-	0,7273	-	0,7403	-	0,7184
Casos válidos	10.957	1.964	60.072	13.853	21.816	4.514	38.256	9.339
Casos totais	11.130	11.130	61.215	61.215	22.260	22.260	38.955	38.955

Nota. (*) teste *t*-Wald significativo a $p > 1,96$.**Tabela 18***Coefficientes padronizados β dos modelos multinível, nos diferentes componentes.*

Comp.	Coefficiente	M1	M2	M3	M4	Componentes
	Intercepto	0,1813	-0,1475	-0,0857	-0,1523	0,3381 (C1) 0,1906 (C2) 1,8351 (C3) 0,2205(C4) 0,1496 (C5)
C1	PPMU4S	0,1596*				0,5253*
	PPEU4S	0,0413				0,2936*
	PMMU4S	-0,0571				-0,3081*
	PMEU4S	-0,0158				-0,1395*
	PPMU8S	-0,1171*				-0,1292*
	PPEU8S	-0,0883*				-0,1555*
	PMMU8S	0,1097*				0,2846*
	PMEU8S	0,1125*				0,3062*
C2	DEU3S	-0,0088	0,0256*	-0,0135	0,0283*	-0,0327
	DEU4S	-0,0940*	-0,0046	0,0195	-0,0176	-0,1416*
	DEU5S	-0,0030	-0,0414*	-0,0618*	-0,0179	-0,1237*
	DEU6S	-0,0491	-0,0495*	0,0173	-0,0707*	-0,2004*
	DEU7S	0,0127	-0,0133	-0,0292	-0,0166	0,0208
	DEU8S	0,0216	-0,0660*	-0,0780*	-0,0427*	-0,0476*
	DEU9S	0,0685*				-0,0723*
C3	DMU3S	-0,0194	0,0343*	-0,0169	0,0505*	-0,0595*
	DMU4S	-0,0498	0,0422*	0,0181	0,0364*	-0,1287*
	DMU5S	0,0279	-0,0404*	-0,1903*	0,0080	-0,1157*
	DMU6S	-0,0355	-0,0604*	0,0156	-0,0806*	-0,2209*
	DMU7S	-0,0137	-0,0290*	0,0135	-0,0583*	-0,1712*
	DMU8S	0,0320	-0,1069*	-0,1053*	-0,0846*	0,0273
	DMU9S	0,0188				-0,0443*
C4	BFF	-0,1567*				-0,4719*
	BFR	-0,0447*				-0,0552*
	DIARREIA	0,0240*	-0,0231*	-0,0169*	-0,0255*	0,0854*
	IMF	0,0281*	-0,0157*	0,0071	-0,0315*	0,0814*
	VSRP	0,1939*	0,1329*	0,1630*	0,1247*	0,0824*
	VMRP	0,3657*	0,4043*	0,3803*	0,4090*	0,4890*
	VIRP	0,3390*	0,2933*	0,2426*	0,3178*	-0,0546*
C5	FEIJAO	-0,0261	-0,1137*	-0,1191*	-0,1071*	-0,3244*
	LEITE	0,0259*	0,0309*	0,0255*	0,0327*	0,0646*
	MILHO	-0,0510*	0,0531*	0,0699*	0,0405*	0,3013*
	SOJA	0,1110*	0,1296*	0,0978*	0,1395*	0,2973*
	VARP	0,4835*	0,4308*	0,4228*	0,4391*	-0,1098*

Nota. (*) teste *t*-Wald significativo a $p > 1,96$.

Para se verificar a grandeza das relações, os coeficientes padronizados foram convertidos de volta ao seu valor original utilizando os dados de média e desvio padrão da Tabela 6. Os valores dos coeficientes significativos estão na Tabela 19.

Tabela 19

Coeficientes não padronizados dos modelos multinível, nos diferentes componentes.

Componente	Coeficiente	M1	M2	M3	M4
VD	Intercepto	0,1813	-0,1475	-0,0857	-0,1523
	PIBPC	18,61	18,61	18,61	18,61
C1	PPMU4S	174,14			
	PPEU4S				
	PMMU4S				
	PMEU4S				
	PPMU8S	-228,34			
	PPEU8S	-232,54			
	PMMU8S	228,20			
	PMEU8S	241,75			
C2	DEU3S		62,65		62,70
	DEU4S	-66,44			
	DEU5S		-65,55	-65,95	
	DEU6S		-69,42		-69,84
	DEU7S				
	DEU8S		-67,62	-67,83	-67,22
	DEU9S	63,62			
C3	DMU3S		64,04		64,38
	DMU4S		66,77		66,64
	DMU5S		-67,42	-70,40	
	DMU6S		-77,43		-77,75
	DMU7S		-75,46		-75,89
	DMU8S		-76,05	-76,03	-75,70
	DMU9S		-69,97		
C4	BFF	-1587,31			
	BFR	-1569,70			
	DIARREIA	0,58	-0,58	-0,57	-0,58
	IMF	7,04	-6,93		-7,07
	VSRP	29,88	29,27	29,57	29,19
	VMRP	6,53	6,65	6,58	6,67
	VIRP	16,60	16,17	15,70	16,40
C5	FEIJAO		-458,84	-463,13	-453,60
	LEITE	3050,40	3066,58	3049,10	3072,41
	MILHO	-5971,44	5994,37	6177,84	5856,77
	SOJA	6539,08	6825,27	6335,98	6977,60
	VARP	31,00	30,18	30,05	30,31

6. Discussão

6.1 Os dados

A primeira questão desta pesquisa diz respeito aos dados utilizados e à sua quantidade e qualidade: existem dados suficientes? E se existem, são representativos do fenômeno?

A Variável Dependente (VD) desta pesquisa é a pobreza estudada em nível de município, a qual é medida indiretamente pela produção de riqueza. A Tabela 6 mostra que a variável Produto Interno Bruto per capita (PIBPC) tem 60.072 casos válidos (98,1 %) de um total possível de 61.215 casos (5.565 municípios para um período de 11 anos – de 2000 a 2010). Logo, pode-se afirmar que a quantidade de dados de PIB é suficiente. Do ponto de vista da qualidade, existia um problema de assimetria e curtose univariados que foram corrigidos por meio de transformação dos escores originais. A Tabela 6 traz os escores originais de média e desvio padrão (DP). A Figura 35 e a Figura 36 mostram que não existem problemas acentuados de linearidade, normalidade multivariada ou heteroscedasticidade multivariada.

Das variáveis explicativas, o primeiro grupo a se discutir é o grupo das variáveis do Programa Bolsa Família (PBF). Compõem este grupo o número de famílias beneficiadas pelo PBF (BFF), o valor de repasse do PBF ao município (BFR) e o percentual de famílias pobres do município beneficiadas pelo PBF (BFC). Comparado ao PIBPC, este grupo de variáveis tem a mesma abrangência geográfica, mas difere em termos de período de coleta. O PBF começou oficialmente em 2003. A partir de 2004 os dados do PBF ficaram disponíveis. Apesar de anteriormente a esta data o programa já existir na forma de outras quatro ações federais (Bolsa Escola, Auxílio Gás, Bolsa Alimentação e Cartão Alimentação), os dados disponíveis destas ações não compartilhavam o mesmo método e período de coleta, o que não permitiu unificar os dados. Um problema observado nos dados do PBF foi a ausência de dados do ano de 2008 para o número de famílias beneficiárias - BFF (Figura 5) e para o montante repassado ao município pelo programa – BFR (Figura 6). Apenas os dados de cobertura do programa - BFC (percentual de famílias beneficiárias do total de famílias elegíveis ao programa) estavam disponíveis. Entretanto, BFC apresentava um outro problema. Até o ano de 2007, alguns municípios apresentavam percentual acima de 100%, o que indica que o PBF estaria beneficiando mais famílias pobres do que existiriam no município. Tanto pode ser um número subestimado da população de pobres (maior número de famílias elegíveis), como pode ser um número superestimado de beneficiários do PBF. Aparentemente, o problema era de superestimação de beneficiários. Esta suspeita foi reforçada pela nota divulgada pela Caixa Econômica Federal (CEF) em 27 de maio de 2013. Esta nota procurava justificar uma corrida aos caixas eletrônicos em vários estados brasileiros por conta de

boatos sobre o término do PBF. A nota afirmou que ocorreu uma liberação adiantada de valores para o saque dos beneficiários como resultado da correção de duplicidade do Número de Inscrição (NIS). Segundo a CEF, existiam cerca de 700 mil cadastros em duplicidade no sistema de Cadastro de Informações Sociais (“Caixa admite equívoco”, 2013). Porém, em 2007 foi realizado um recadastramento dos beneficiários do PBF e a sua migração para o atual sistema de Cadastro Único (CadÚnico). A partir de 2009, os dados já não ultrapassavam a 100%. Dada a nota divulgada pela CEF em 2013, é certo que o cadastro de beneficiários estava superestimado para todo o período de 2004 a 2010. Os dados dos repasses do PBF (BFR) foram deflacionados da mesma forma que os dados de PIB e de Valor Adicionado Bruto. Em 2010, dois dados de BFC foram divulgados, um baseado no sistema CadÚnico e outro no sistema de cadastro antigo. A correlação bivariada entre estes dois dados foi de 0,791, IC 99% [0,786, 0,796]. Assim, apesar dos valores da base antiga apresentarem valores maiores que os do CadÚnico, ambas parecem medir o mesmo fenômeno. A variável que mede a cobertura do PBF (BFC) apresentou uma distribuição normal, não exigindo transformação dos dados. De qualquer forma, BFC não foi utilizado em função dos resultados da Análise dos Componentes Principais (ACP).

O grupo dos dados agropecuários foi um caso diferente. Apesar da disponibilidade de dados para todos os municípios e para todo o período do estudo, muitos dados apresentaram o valor zero de produção agropecuária. Este problema foi mais acentuado na soja por conta da característica técnica desta cultura. Ela exige grandes áreas mecanizáveis e com condições especiais de solo e clima. Estes valores zero determinaram a assimetria positiva persistente observada nas variáveis deste grupo. Mesmo com a transformação, a correção da assimetria não foi suficiente, conforme mostra a Tabela 11. Outras culturas poderiam ter sido escolhidas, mas o problema teria permanecido, pois trata-se de uma característica comum aos produtos agropecuários em geral. A retirada dos casos com valores zero mostraram que os dados remanescentes se distribuíram normalmente. Mas esta solução tiraria a representatividade dos dados. A retirada de municípios também não resolveria. Culturas como a soja, que ocupam grandes áreas, mas em poucos municípios, faria com que o número remanescente de municípios para análise seria muito pequeno

frente ao número total de municípios. Esta solução introduziria algum grau de violação dos pressupostos.

O grupo das variáveis de proficiência trouxe um outro tipo de problema. Os critérios de coleta se modificaram muito ao longo dos anos, conforme informa o Manual do Usuário do Saeb e do Prova Brasil. Em 2001, a amostra probabilística era composta de 288.331 alunos da 4ª e 8ª séries do fundamental e 3ª série do ensino médio, em escolas públicas e particulares (Inep, 2006a). Foram escolhidas escolas urbanas, excetuando escolas federais e turmas multisseriadas. Para a 4ª série, foram mantidos os alunos das escolas rurais do Nordeste, Minas Gerais e Mato Grosso do Sul. Em 2003, a amostra de alunos era composta de matriculados em escolas urbanas constantes no censo escolar de 2002, com exceção das turmas multisseriadas e de aceleração (Inep, 2006b). Foi incluído o universo de alunos da 4ª série do fundamental de escolas rurais com 10 ou mais alunos por série, com exceção das rurais federais. Em 2005, a amostragem foi praticamente igual a 2003, com exceção das escolas de educação especial, das escolas localizadas em áreas indígenas e quilombolas e escolas com menos de 10 alunos por série (Inep, 2008). Em 2007, houve uma grande mudança com a implementação do Prova Brasil. Os levantamentos passaram a ser de escolas públicas urbanas com mais de 20 alunos por série (Inep, 2009). Em 2009, ficou mais explícita a diferença entre a Avaliação Nacional do Ensino Básico (Aneb) e a Avaliação Nacional do Rendimento Escolar (Anresc). A primeira manteve foco na educação básica, utilizando amostras de escolas com 10 alunos ou mais por série, como proposto pelo Saeb. A segunda, chamada de Prova Brasil, manteve foco na demanda dos gestores públicos, educadores e pesquisadores. A Anresc adotou o caráter censitário com escolas públicas de 20 ou mais alunos por série, da zona rural e urbana (Inep, 2011b). Em 2011, o levantamento compreendeu 4,3 milhões de alunos em 56,6 mil escolas com pelo menos 20 alunos por série (Inep, 2012b). Esta grande gama de alterações nos critérios e focos das coletas teve impacto nas análises. O problema mais grave foi o pouco número de casos válidos, i.e., municípios que apresentassem dados de proficiência para todos os anos de levantamento dentro do horizonte da pesquisa. O número de casos variou de 698 (proficiência de matemática na 4ª série/5º ano em escolas estaduais rurais – PMER4S) a 11.242 casos (proficiência de português na 4ª série/5º ano em escolas municipais urbanas – PPMU4S), conforme mostra a Tabela 7. O corte temporal também foi outro

fator complicador, pois os levantamentos de proficiência são bienais. Isto exigiu recortes nos modelos de regressão para que os outros dados pudessem ser calculados e mantivessem a mesma base de comparação. Do ponto de vista de qualidade estatística, os dados se mostraram bons. Não houve necessidade de qualquer ajuste de assimetria ou de curtose (Tabela 7). Porém, existiam inconsistências nos dados de dependência (municipal, estadual, federal ou privada) e localização (rural ou urbana) em alguns anos do levantamento, quando comparado com os dados do censo escolar.

O grupo de dados de distorção idade-série (DIS) foi calculado a partir dos dados do censo escolar. O número de casos foi sensivelmente maior. As qualidades de assimetria e curtose também evidenciaram a boa representatividade dos dados, não tendo sido necessária qualquer transformação dos dados. Contudo, este conjunto de dados foi afetado com a mudança do sistema do ensino fundamental de oito para nove anos. Notou-se que as estratégias de transição entre o antigo e o novo sistema não foram uniformes. Cada instituição escolheu a sua solução. Assim, estados como o Mato Grosso do Sul, em 2011, já não contabilizavam qualquer escola no sistema de oito anos, mas foi a única em todo o Brasil. Esta diferença na celeridade de transição entre sistemas causou um sério problema de comparabilidade dos dados. O Inep manteve registros separados para alunos matriculados no sistema de oito anos e no sistema de nove anos. Porém, verificando os dados individualmente, os percentuais de DIS não eram numericamente próximos quando comparando séries/anos equivalentes entre um sistema e outro (por exemplo, 3ª série do sistema antigo e 4º ano do sistema novo). Por conta disso, não fazia sentido calcular a média entre séries/anos equivalentes. Também notou-se que, em algumas escolas, o sistema antigo era adotado para as turmas até a 4ª série e o sistema novo era adotado de maneira complementar para as turmas de 5º ao 9º anos. A solução para este problema foi a justaposição dos dados dos dois sistemas. Do ponto de vista de DIS, os dados foram representativos uma vez que as idades corretas foram calculadas com base nos respectivos sistemas. Contudo, quando comparado com os dados de proficiência, existiu diferenças. Os dados de proficiência combinam as duas informações de série escolar (4ª/5º e 8ª/9º), enquanto os registros do censo escolar são separados.

Os dados de saúde tentaram captar informações de aspectos básicos de saneamento. Apenas depois da coleta dos dados ficou evidente que estes dados se concentraram nos maiores centros urbanos. Normalmente, as regiões afastadas têm

menos recursos de saúde. As famílias vão aos centros urbanos maiores em busca de tratamento. Este deslocamento dos indivíduos faz o cômputo do óbito em favor do local de falecimento e não do local cujas condições sanitárias pudessem estar relacionadas ao óbito. Outros registros de saúde como cobertura vacinal, número de leitos hospitalares, número de centros de saúde, disponibilidade de rede de esgoto ou de sistema de tratamento de água foram experimentados. Porém, esbarraram em problemas de disponibilidade de dados para todo o período de estudo, bem como careciam de uma metodologia de coleta melhor definida. Por exemplo, as coberturas vacinais variavam em função das campanhas promovidas pelos governos. Hospitais não estavam presentes em todos os municípios e os centros de saúde tinham registro junto às secretarias municipais de saúde, algumas com bons registros e outras com absolutamente nenhuma informação. Os dados de esgoto e água tratada estão listados nos censos demográficos, cuja periodicidade é decenal.

Finalmente, os dados de Valor Adicionado Bruto (VAB) foram os menos problemáticos. Como são a base para o cálculo do Produto Interno Bruto (PIB), o método utilizado e a frequência de coleta seguem o mesmo rigor observado no PIB. Os dados foram úteis e precisos em identificar os setores produtivos mais dinâmicos de cada município. Porém, os resultados da Análise de Componentes Principais (ACP) da relação entre estes setores e as demais variáveis explicativas nem sempre fizeram muito sentido. Talvez pelo fato do conceito de setor produtivo ser por si só multidimensional, concatenando sempre educação, saúde, renda e agronegócio em diferentes proporções.

A estratégia de correção das assimetrias e curtoses por meio da transformação de Box-Cox foi efetiva. A grande facilidade desta técnica é a possibilidade de utilizar uma sintaxe do SPSS para procurar a solução de potência mais acertada, economizando tempo e evitando aproximações pelo método de tentativa e erro com as quatro transformações básicas da literatura (exponencial, logarítmica, raiz e inversa). A sintaxe se encontra no Anexo 1. A transformação não foi apenas efetiva para os pressupostos univariados, mas também permitiu o atendimento dos pressupostos multivariados (Figura 35 e Figura 36).

A despeito das limitações de tempo e cobertura geográfica dos dados, do ponto de vista estatístico, os dados coletados foram suficientes para desenvolver os modelos

desta pesquisa. O caráter amostral de proficiência, o comportamento dos indivíduos doentes e a concentração da produção agrícola em algumas regiões do país foram algumas das causas dos problemas observados de representatividade dos dados.

6.2 Modelos de regressão multinível

Para fins da discussão a seguir, será considerado “rico” aquele município com elevado PIB per capita anual (PIBPC). De forma análoga, será considerado “pobre” aquele município de baixo PIB per capita. Evidentemente, são adjetivos relativos, pois “elevado” e “baixo” são resultados de comparações entre dois ou mais municípios quaisquer.

6.2.1 Modelo com todas as variáveis explicativas

Este modelo é denominado modelo M1. A principal característica do modelo M1 é conter as variáveis do Programa Bolsa Família (PBF). É também o único modelo que inclui todas as variáveis explicativas indicadas pela Análise de Componentes Principais (ACP). O *trade-off* para permitir a inclusão de todas as variáveis foi a realização de cortes temporais. No modelo M1, apenas os dados dos anos de 2007 e 2009 foram utilizados. O principal impacto desta decisão foi reduzir acentuadamente a variância observada para o nível 1 (N1), o qual representa os dados longitudinais dos municípios. Quando M1 é rodado utilizando dados não padronizados, as variâncias em N1 ficam iguais a zero. Mesmo utilizando dados em escores z (Tabela 17), o resíduo N1 é bem menor que o resíduo N2 e ainda não significativamente diferente de zero, $\sigma_e^2 = 0,0241$, $p < 1,96$. Isto indica que a abordagem por análise multinível (AMN) não foi vantajosa para este modelo e que uma simples regressão para qualquer um dos dois anos seria suficiente para apontar os coeficientes de regressão. A correlação intraclasse (ICC) também indica o nível 2 (N2) como a maior fonte de variância. Segundo os cálculos de ICC para M1, existe uma correlação de 95% entre os dois anos de um mesmo município. Verificando o R^2 de N2, obteve-se 0,812, IC 90% [0,799, 0,824] da variância sendo explicada por N2. Logo, a variância entre municípios é maior do que a variância entre os anos.

Levando em conta que são 5.565 municípios e apenas dois anos de dados, estes resultados não são de todo surpreendentes. Os municípios têm uma progressão muito mais lenta de desenvolvimento do que teria, por exemplo, um indivíduo. Ficou claro

que dois anos de interstício entre os dados, é um período pequeno para se observar mudanças quando se trata de municípios.

Ainda assim, a *deviance* mostrou redução acentuada. Por se tratar de dados de população, o teste qui-quadrado não foi apropriado para verificar a significância da diferença entre *deviances* de dois modelos. Segundo Hox (2010), o grande número de casos implica na expressão de um valor de estatística qui-quadrado muito baixo, tornando qualquer pequena diferença de resultados sempre significativa. Assume-se aqui que uma redução de dez vezes na *deviance* não deixa margens à dúvida de que o ajuste de M1 é melhor do que o do modelo vazio.

Outro aspecto a se salientar é o número total de casos válidos. Para 5.565 municípios e dois anos de dados, esperava-se um total de 11.130 casos. M1 apresentou pouco mais de 17% desse total. Foram 1.964 casos válidos para análise. Ainda assim, um número expressivo de casos.

O modelo final de M1 (Tabela 18) apresentou 18 coeficientes significativos. Todos os coeficientes de distorção idade-série (DIS) da rede municipal foram não significativos. Apenas dois coeficientes de DIS de escolas estaduais urbanas, do 4º e do 9º anos, foram significativos. Por outro lado, todas os coeficientes da Componente 4 (C4), referente ao grupo de variáveis de bem-estar, foram significativos. As da Componente 5 (C5), grupo das variáveis agropecuárias, somente FEIJAO ficou fora do modelo. De forma geral, depreende-se que municípios mais ricos estão relacionados com setores de agropecuária, de serviços e da indústria mais pujantes, com maior arrecadação de impostos, maior número de casos de óbito por diarreia e maior mortalidade infantil. Do ponto de vista da educação, os alunos de tais municípios se saem melhor em matemática do que em língua portuguesa, com menores DIS nas séries iniciais e maior DIS nas séries finais. Destaca-se ainda que os municípios mais ricos são também onde se encontram o menor número de famílias beneficiárias do PBF (BFF) e com os menores repasses do programa (BFR).

Desapontador foram os coeficientes não significativos de DIS das escolas municipais urbanas. Como comentado anteriormente, as escolas municipais predominam longe dos grandes centros urbanos. Talvez pelo fato do modelo conter apenas dados de escolas urbanas, isto tenha prejudicado o efeito destas variáveis. Contudo, verificando a Tabela 6, não se nota problemas evidentes em termos de casos válidos quando se compara a DIS de escolas estaduais urbanas e a DIS de escolas

municipais urbanas. Mesmo que o número de escolas municipais fosse menor nas áreas urbanas, o número de casos não seria um motivo para não se verificar efeitos significativos.

Dos resultados obtidos, duas informações são interessantes. A primeira diz respeito às variáveis de saúde. As variáveis escolhidas não se mostraram bons indicadores das condições de saúde da população. Municípios mais ricos deveriam apresentar melhores condições de saúde, o que implicaria em menor incidência de óbitos por diarreia e mortalidade infantil. Assim, ou as pessoas de fato estão se deslocando de regiões menos favorecidas em infraestrutura de saúde para as mais favorecidas, ou as cidades mais ricas (maior PIBPC) têm condições ruins de saneamento básico e de atendimento à gestante e aos recém-nascidos. A primeira hipótese parece a mais razoável. A segunda informação interessante é que o número de famílias beneficiadas pelo PBF (BFF) e o repasse para o município de fundos pelo PBF (BFR) se comportaram como esperado, confirmando que cidades mais ricas não se associam com maiores intervenções do PBF. Isto é um indício de que o PBF possui precisão em atingir as regiões menos favorecidas, conforme afirma Lindert et al. (2007), Castro e Modesto (2010a) e Soares et al. (2010).

Os resultados obtidos em escore z (Tabela 18) foram convertidos para os escores originais (Tabela 19) por meio das médias e dos desvios padrão (DP) da Tabela 6. A regressão não estabelece a causalidade da relação. Assim, o resultado da equação pode ser lido como pobreza (PIBPC) afetando o Programa Bolsa Família (PBF), ou como o programa afetando a pobreza. Contudo, pensando em termos de anterioridade, a pobreza antecede o PBF. A pobreza da família (baixo PIB per capita) é que torna a família elegível. Porém, não se pode esquecer que a pobreza está sendo medida indiretamente pelo PIB, que é uma medida de produção de riqueza (renda). Assim, é também uma relação causal o PBF aumentar a renda, uma vez que é um programa de transferência de renda. A princípio, poderia se assumir uma relação de bi-causalidade diferida no tempo, i.e., a pobreza desencadeia o PBF mas, posteriormente, PBF afeta a renda.

A partir dos coeficientes convertidos a escores originais, percebe-se que o aumento de R\$ 1,00 no PIBPC do município implica na saída de 85 famílias do programa (BFF). Este valor foi obtido pela divisão do coeficiente de BFF (-1.587,31) pelo coeficiente de PIBPC (18,61). É preciso lembrar que o coeficiente da variável

dependente também foi padronizado. Por estes valores, fica evidente que aumentar a produção de riqueza do município (PIBPC) é uma maneira eficaz de combater a pobreza uma vez que o número de famílias que deixam de depender do programa é muito maior que o valor de incremento na riqueza per capita. Então o PBF não é importante? Pelo raciocínio inverso, cada R\$ 1,00 de queda em PIBPC implica em um aumento de apenas R\$ 84 no repasse do programa. O valor foi obtido pela divisão do coeficiente de BFR (-1.569,70) pelo coeficiente de PIBPC (18,61). Logo, é uma solução relativamente barata de combate à pobreza. Esta conclusão é a mesma à qual chegou Soares et al. (2010). Uma outra consequência desta afirmação é que, apesar da relação entre famílias beneficiadas (BFF) e pobreza (PIBPC) ter sido considerada bi-causal, como o efeito de aumento de repasse é pequeno, pode-se admitir que a relação causal entre BFF e PIBPC é mais forte no sentido de PIBPC para BFF. Esta constatação pode também indicar que o programa está num estágio ainda inicial, a despeito de em 2010 ter sete anos de implementado.

É preciso ter em mente as limitações destas afirmações. Estes valores foram obtidos pela observação do efeito em apenas dois anos. Nestes dois anos, a diferença entre eles é pequena (vide a alta correlação intraclasses – ICC), de maneira que grande parte da variação provém das diferenças entre municípios. Outro ponto importante é que estes valores são válidos para um contexto onde se leva em conta aspectos educacionais, de saúde, de produção agropecuária e de percentual de participação de cada setor na produção de riqueza. Os coeficientes de regressão das variáveis do PBF foram retirados de uma equação completa.

Um outro efeito que poderia ser atribuído ao programa são as fortes oscilações da inflação dos alimentos (Figura 33). Dada a situação de pobreza, estas famílias não poupam. Assim, ao receberem o repasse do PBF, o incremento no consumo de bens essenciais é instantâneo, principalmente o incremento no consumo de alimentos. Uma elevação tão rápida num curto espaço de tempo causa a chamada inflação de demanda. Se o aumento no consumo de um determinado bem não é acompanhado pelo pronto abastecimento deste bem no mercado, a tendência é de aumento no preço do bem a fim de equalizar a demanda aos estoques disponíveis. Isto talvez explique a curva crescente e o pico de inflação dos alimentos entre 2000 e 2002 (Figura 33). Apesar do programa PBF ter sido lançado em 2003, ele já existia na forma de quatro programas de transferência de renda independentes. A queda posterior da inflação dos alimentos

pode ser entendida como um ajuste do mercado ao aumento da produção de alimentos. Isto também pode ser um aspecto positivo que indicaria a abertura de mais postos de trabalho a fim de aumentar a produção e atender à demanda elevada.

A escolha dessas duas variáveis explicativas (BFF e BFR) do PBF foi útil. Percebe-se que o efeito da pobreza sobre as famílias é muito forte. É preciso criar mecanismos que ampliem a produção de riquezas nos municípios pobres. Em outras palavras, apoiar ações que ampliem o número de postos de trabalho, que aumente o consumo de bens e serviços, que torne a economia local mais dinâmica. Pequenas quedas nesta taxa de crescimento do município ampliam em muito o número de famílias que passam a depender do PBF num curto espaço de tempo. Esta afirmação encontra respaldo nos coeficientes positivos das variáveis de Valor Adicionado Bruto (VAB). Estes coeficientes mostraram que aumentos em qualquer setor produtivo estão relacionados com melhores PIBPC. Um aumento de PIBPC implica em menos beneficiários do PBF e menos repasse das bolsas. Outro aspecto importante do PBF é saber que ele tem um custo relativamente baixo caso seja necessário lançar mão dele. Baseado nos resultados do presente estudo, a sugestão para os gestores públicos é primeiro estimular a produção de riqueza nos municípios. Além do efeito de tirar um grande número de beneficiários da linha de pobreza, a alocação destas famílias no mercado de trabalho ajuda a ampliar a oferta de serviços e produtos, o que diminui o impacto de uma inflação de demanda. Somente depois disso, em sendo necessário, realizar a intervenção pelo PBF. Isto maximizaria o investimento público no combate à pobreza. Na perspectiva do modelo, o programa se mostrou um mecanismo efetivo de redução da pobreza no curto prazo.

Cabe ainda mencionar algumas relações das variáveis educacionais com a produção de riqueza do município. Municípios mais “ricos” (maior PIBPC) apresentaram menor distorção idade-série (DIS) no 4º ano e maior DIS no 9º ano. Estes mesmos municípios apresentaram os maiores escores de proficiência em matemática e menor em língua portuguesa, quando comparados com os municípios mais pobres. Estas variáveis serão melhor exploradas nos modelos subsequentes.

6.2.2 Modelo que cobre todo o horizonte da pesquisa

O segundo modelo (modelo M2) foi pensado com o objetivo de pesquisar o que acontece com as variáveis ao longo de todo o horizonte temporal da pesquisa. Para isso, foi necessário abrir mão das variáveis de proficiência em Matemática e Português e das variáveis relacionadas ao Programa Bolsa Família (PBF).

Pelos resultados na Tabela 17, a primeira diferença que se nota é o resíduo de nível 1 (N1) significativo. Os dois anos estudados no modelo M1 não foram suficientes para capturar diferenças ao longo do tempo, mas onze anos foram. A participação do nível N1 em explicar a variância total se confirma com a menor variância explicada R^2 de nível 2 do modelo M2 em comparação com o modelo M1 (73% *versus* 81%). Nota-se também o aumento no número de casos válidos, que passam de pouco menos de 2 mil em M1 para 13,8 mil em M2.

O maior número de casos válidos se refletiu em maior número de variáveis explicativas de distorção idade-série (DIS) significativas. Enquanto em M1 nenhuma variável de DIS de escolas municipais foi significativo (Componente 3 – C3), em M2, todas as variáveis de C3 foram significativas. Com exceção de duas variáveis da Componente 2 (C2 – DIS de escolas estaduais urbanas), todas as demais variáveis entraram no modelo final de M2.

Verificando os sinais dos coeficientes de regressão, confirma-se a expectativa de que os municípios mais ricos têm menores percentuais de alunos em DIS. A exceção fica para as medidas do 3º e 4º anos. Tanto em escolas estaduais, como em escolas municipais, a DIS no 3º ano se relacionou positivamente com a riqueza do município. Observando os gráficos da Figura 18, existem duas observações a se fazer que podem ajudar a explicar esta diferença no comportamento das variáveis de séries iniciais e variáveis de séries finais do ensino fundamental. A primeira observação é que a DIS de 2º ao 5º anos apresentam um curva mais plana entre 2002 a 2010, com leve piora no 3º e 4º anos (maior DIS) e leve melhoria no 2º ano. Já os anos de 6º ao 9º, apresentaram queda na DIS. A segunda observação é que a DIS média é menor nos anos iniciais. O gráfico (a) da Figura 18 mostra que, com exceção de 2002, todos os demais anos apresentaram DIS menor que 60%. Porém, nos gráficos (f) a (h) da Figura 18, respectivamente 7º ao 9º anos, os valores percentuais de DIS são acima de 70%, com exceção de 2008 no 6º ano. O que se depreende é que mais alunos estavam entrando na idade correta na escola. Porém, com a progressão no curso, há elevação

de DIS, apesar de decrescente ao longo dos anos. A elevação da DIS entre anos iniciais e anos finais pode ser explicada pela repetência do aluno e sua permanência por mais tempo em um determinado ano. Com o crescente insucesso, o aluno desiste da escola (evasão), ocorrendo uma seleção não proposital de alunos melhores, o que reflete na queda do DIS ao longo dos anos. Outra explicação plausível para a queda de rendimento escolar e atraso escolar é que os municípios mais pobres se diferenciam dos demais municípios por terem menos escolas que oferecem os anos finais (5º ao 9º anos) e por mais crianças se tornarem mão de obra infantil para ajudar no orçamento familiar.

A ausência das variáveis de proficiência e do PBF em M2 teve um benefício. Além do maior número de variáveis de DIS no modelo, os sinais dos coeficientes das variáveis de bem-estar (Componente 4 – C4), especificamente as variáveis ligadas à saúde, mostraram-se mais coerentes. Ambos os coeficientes foram negativos e significativos. Quanto mais rico o município, menores são os casos de diarreia e menores são os indicadores de mortalidade infantil. Deve-se ser cauteloso na análise destas variáveis, pois em M1 elas foram significativas. Como discutido anteriormente, as pessoas tendem a ir para os grandes centros urbanos na falta de recursos de saúde em suas cidades de origem. Com isso, o registro de óbito tende a ser maior nas cidades com mais recursos. Logo, também é justificável coeficientes positivos para estas variáveis. Este comportamento dos coeficientes leva a duas discussões pertinentes. De um lado, modelos com mais variáveis explicativas tendem a ser “super-ajustados” (*overfitted*), o que poderia explicar o comportamento fora do esperado do modelo M1 (modelo com todas as variáveis explicativas). Por outro lado, as variáveis de saúde já apresentavam problemas de assimetria causada pelo grande número de municípios com o valor zero de registro de óbitos. Desta forma, esta oscilação no comportamento entre um modelo e outro pode estar refletindo a má qualidade destas variáveis. Será possível avaliar melhor isto com os próximos modelos, M3 e M4.

Não foi surpresa que todas as variáveis que medem o tamanho dos setores produtivos em termos da riqueza total do município tinham um coeficiente de regressão positivo. Isto confirma o que se observou em M1. Independente de qual seja o setor produtivo mais pujante no município, percentuais altos destas variáveis estão

relacionados com municípios mais ricos. Isto também é válido para a variável VARP, ligada à produção de riqueza no setor agropecuário.

Quanto às produções agropecuárias em si, apenas FEIJAO mostrou-se negativo. Isto era esperado já que se trata de uma cultura típica de pequena produção (Ministério do Desenvolvimento Agrário, 2009). Confirma-se que municípios ricos não apresentam expressiva produção de feijão e, provavelmente, possuem menos produtores rurais do tipo familiar. Era esperado o mesmo comportamento para a variável LEITE. Porém, a proporção de produtores familiares que produzem leite é menor que os que produzem feijão, 58% e 70% do total de produção nacional, respectivamente. Isto pode ter afetado a relação deste coeficiente com a variável dependente (VD). Tanto em M1, como em M2, este coeficiente se apresentou positivo. Logo, é mais provável que LEITE seja uma escolha ruim para expressar a pequena produção e a pobreza associada a ela. MILHO e SOJA comportaram-se como esperado, ambos positivos. Milho e soja são culturas associadas à mecanização da produção e mais intensivas em insumos, o que demanda maior capital para investimento. Municípios mais ricos se relacionam com a produção de soja e milho. A ressalva fica para milho, que oscilou de negativo para positivo. Diferente da soja, milho é uma cultura relativamente bem distribuída geograficamente, de maneira que o registro de municípios com nenhuma produção é raro. Portanto, a fonte de oscilação do coeficiente desta variável não parece estar relacionada à qualidade do dado. Além disso, os coeficientes padronizados são praticamente iguais, apenas de sinais trocados. Assim, não parece se tratar de algum tipo de comportamento associado ao tamanho da amostra. Se assim fosse, esperaria-se uma mudança relativamente proporcional. Será observado o comportamento desta variável nos modelos M3 e M4.

6.2.3 Modelos antes e depois do evento do PBF

A fim de estudar o impacto do PBF, duas estratégias poderiam ter sido adotadas: (1) incluir uma variável dicotômica (*dummy*) na forma de T_{ij} na equação de regressão multinível, representando a ocorrência ou não do evento; ou (2) realizar um *split* da base de dados, com dados antes da ocorrência do PBF e outra base com os dados depois da ocorrência do PBF. A opção desta pesquisa pela segunda estratégia reside na possibilidade de comparar separadamente os coeficientes de ambas as equações. Caso se optasse pela primeira estratégia, ter-se-ia um coeficiente para T_{ij} e

uma significância associada, o que apenas informaria se existiu um efeito ou não do evento. Tendo os coeficientes separadamente para cada equação de regressão permitirá não só comparar o efeito do PBF, como comparar com os modelos finais de M1 (todas as variáveis) e de M2 (todos os anos), verificando o comportamento de variáveis para períodos de análises diferentes. Observando M1, já foi possível notar que a variância de nível 1 (N1) não foi significativa. Com onze anos, já se percebe algum efeito de N1. Como seria com quatro anos (modelo M3) e com sete anos (modelo M4)? Também vimos que alguns coeficientes de regressão das variáveis de M1 e M2 mostraram comportamento oscilatório. Será que estas variações se mantêm nestes outros modelos?

Pela Tabela 17, observa-se que a variância inicial do nível micro (N1) do modelo M4 (0,0431) é quase quatro vezes maior do que a variância inicial do nível micro do modelo M3 (0,0111), mas é menor que de modelo M2 (0,0465), indicando que existe relação entre variância do nível micro e o tamanho do período de análise.

Assim, período igual ou superior a quatro anos é recomendável para análises longitudinais com séries históricas como as séries utilizadas aqui. Os outros indicadores do modelo também apoiam esta afirmação. A correlação intraclasse (ICC) mostra também que a correlação entre os anos caiu com o incremento dos anos analisados, mantendo M3 como o modelo com maior ICC, M4 em posição intermediária e M2 no extremo oposto. Entretanto, R^2 de nível de municípios (N2) tem essa ordem alterada. M4 é o modelo que apresenta o nível de município (N2) como o que menos explica a variância total, inclusive menos que M2, que é o modelo com o maior período de análise e do qual se esperava que fosse o que apresentasse menos variância explicada de N2. A interpretação deste resultado é de que o período de 2004 a 2010 explica mais variância do que o período de 2000 a 2010. Algumas razões podem ser levantadas: (a) sete anos pode ser um número de anos ótimo de análise; (b) a qualidade dos dados antes de 2004 pode ser baixa, diminuindo o poder explicativo do modelo pelos anos (N1); (c) os dados podem ser de boa qualidade, mas os municípios apresentam pouca variância entre os anos para o período antes de 2004; (d) de fato existe diferença entre a variância observada antes de 2004 e depois de 2004, o que poderia indicar algum efeito associado ao PBF. O número de casos válidos não teria influência sobre isso, já que mantém a mesma ordem anterior de $M2 > M4 > M3$. Ou seja, modelo M4 permanece um modelo intermediário.

Para verificar as possibilidades da causa desse R^2 inesperado, deve-se estudar os dados descritivos e os resultados da regressão combinadamente. Seria sete anos um período ótimo de análise? Se fosse, o ajuste do modelo M4 deveria ser melhor que o ajuste do modelo M2. Como o indicador de ajuste aqui utilizado (*deviance*) é uma medida relativa ao modelo vazio, não é possível comparar diretamente os ajustes de cada modelo entre si. Contudo, como os modelos são idênticos (mesmas variáveis explicativas), os graus de liberdade dos dois modelos são iguais. Assim, verificar percentualmente quanto cada modelo “melhorou” em relação ao respectivo modelo vazio pode dar uma medida indireta de comparação entre os ajustes dos dois modelos. No modelo M2 (todas os anos), o ajuste apresentou uma melhora de 87% em relação ao modelo vazio. Já o modelo M4 (após intervenção do PBF) apresentou uma melhora de 86%. Apesar dos valores muito próximos, ainda assim M2 tem melhor ajuste que M4. Logo, sete anos não é um número de anos ótimo para análise.

Seria então os dados antes de 2004 de baixa qualidade? As variáveis explicativas de bem-estar (Componente 4) e agropecuárias (Componente 5) são semelhantes entre M3 e M4, com exceção apenas da variável de mortalidade infantil (IMF). Resta então as variáveis de distorção idade-série (DIS) de escolas estaduais (Componente 2) e as variáveis de DIS de escolas municipais (Componente 3). Os coeficientes de DIS apresentam algumas diferenças de sinais, especificamente nos anos iniciais. Estes dados são baseados no censo escolar e, portanto, relativamente completos, cobrindo a grande maioria das instituições de ensino. Verificando a Tabela 6, o menor número de casos para escolas estaduais urbanas foi 26.844 e o maior 52.691. Para escolas municipais urbanas, o menor número de casos foi 22.327 e o maior 54.916. Estes intervalos são relativamente parecidos. Em termos de normalidade univariada, também estão dentro do aceitável. Os coeficientes de DIS dos anos iniciais do modelo M3 (antes da intervenção do PBF) apresentaram erro padrão elevado, o que resultou em coeficientes não significativos. Isto pode ser um indicador de dados ruins, porém também indica que a variação entre as médias das amostras é grande. Logo, se a variação é maior, o resíduo de N1 deveria ser maior, o que não aconteceu.

Será então que existe pouca variância dos dados de DIS? As médias de DIS para cada ano mostram um padrão até 2003 e outro padrão a partir de 2004, principalmente nas séries iniciais. Se observa maiores variações entre os anos de 2000 a 2003, o que

implicaria numa variância de N1 maior, o que se confirmou com um erro padrão elevado. Aparentemente, a grande variação observada em DIS para o período 2000 a 2003 não influenciou ou foi compensada no modelo. Um possível candidato para a compensação é a variável de mortalidade infantil (IMF), a única variável da Componente 4 (C4) que não foi significativa em M3. Mas é apenas uma variável contra quatro variáveis de DIS que se diferenciam nos outros modelos. Assim, existe sim variância maior nos dados de DIS do período de 2000 a 2003 quando comparado ao período de 2004 a 2010.

Resta supor que existe alguma diferença entre M3 e M4 que não está relacionado com o tamanho do período de análise e nem com qualidade ou falta de variância das variáveis de M3, modelo com menor resíduo de N1.

Por meio dos escores padronizados da Tabela 17, é possível comparar uma a uma as variáveis explicativas de M2 (todos os anos), M3 (antes da intervenção) e M4 (depois da intervenção). As variáveis da Componente 5 (C5), relacionadas à produção agropecuária, apresentaram coeficientes muito parecidos entre M2 e M4. M3 teve coeficientes numericamente distintos, porém as variáveis mantiveram a mesma ordem de grandeza dos outros dois modelos, bem como os mesmos sinais dos coeficientes. Os coeficientes das variáveis da Componente 4 (C4 – bem-estar) também são semelhantes para M2 e M4, com exceção de IMF, a mesma variável que foi não significativa em M3. Isto aponta para problemas de qualidade dos dados de IMF, que já foi mencionado anteriormente. As componentes C2 (DIS estadual) e C3 (DIS municipal) são as que apresentam mais variáveis com valores de coeficientes diferentes. Mas antes de avaliar os seus valores, deve-se observar o conjunto das variáveis. O primeiro aspecto que fica evidente é que existe uma relação entre número de anos analisados e número de variáveis significativas de distorção idade-série (DIS). O modelo M2 tem dez variáveis explicativas, M3 tem quatro e M4 tem oito. Em média uma variável significativa por ano analisado no modelo. O segundo aspecto é o sinal dos coeficientes. Comparando apenas as variáveis significativas dos três modelos, todos os sinais são iguais entre os modelos. Isto é um bom indicador de consistência da variável e mostra a importância de avaliar a significância do coeficiente como critério para colocar ou retirar a variável do modelo final. Quanto aos valores, os coeficientes significativos de M2 e M4 são relativamente semelhantes entre si. Já em M3, os coeficientes são bem diferentes, com apenas uma exceção, a

variável de DIS da turma de 8º ano de escola municipal urbana (DMU8S). O terceiro aspecto é que os valores dos coeficientes significativos de M2 são todos maiores, em módulo, que os respectivos valores dos coeficientes dos outros dois modelos.

O que se pode depreender destas comparações é que parece existir algum efeito do PBF e ele foi principalmente por meio das variáveis de educação e não tanto pela produção de riqueza ou melhoria de saúde. M2 parece fazer uma “média” de dois períodos distintos medidos por M3 e M4. O modelo M4 se parece mais com o modelo M2, talvez pelo maior número de anos que estes dois modelos têm. Como já mencionado, o número de anos definitivamente tem efeito sobre os indicadores de ajuste e de explicação da variância do modelo. Porém, com a determinação dos períodos de análise do Modelo M3 e do Modelo M4 exatamente no momento do evento do PBF, percebe-se que se identificou dois modelos distintos, cuja maior diferença reside nos indicadores de educação. Esta diferença de indicadores não coincidiu com a queda abrupta do número de instituições de ensino que se observou entre 2006 e 2007 (Figura 16), ou com a queda no número de óbitos por diarreia em 2007 (Figura 21), ou com a queda de produção de milho em 2005 (Figura 26). Porém, em 2003 ocorreu um pico no percentual de riqueza oriunda do setor agropecuário, que também coincidiu com a queda de participação do setor de serviços (VSRP) e de “outros serviços” (VERP). Outro dado interessante é que em 2002 ocorreu um pico substancial da inflação do grupo “alimentos” medido pelo INPC, o maior em todo o período do estudo. Os grupos “saúde” e “educação” tiveram um pico em 2003. Entretanto, apesar da coincidência dos diversos picos em torno de 2003, nos demais anos os valores permaneceram relativamente estáveis, ou com oscilações menores. Assim, mesmo que os picos tivessem efeitos posteriores na população, ele foi pontual e não perdurou de maneira a diferenciar 2000/2003 de 2004/2010. Uma grande exceção está no Valor Adicionado Bruto Real da Produção Agropecuária (VARP). Ele vem perdendo terreno desde 2003, enquanto os demais setores se mantêm estáveis (Figura 32). No modelo de regressão, o seu coeficiente é estável entre os três modelos, sempre positivo, acompanhando as demais variáveis de Valor Adicionado Bruto. De qualquer forma, existe aqui espaço para um estudo mais aprofundado de uma possível relação entre alimentos e educação.

6.2.4 Modelos de cada componente

Do ponto de vista prático, os modelos com apenas um componente têm pouco poder explicativo já que não computam interação com variáveis de outros componentes. Isto pode ser observado pela última coluna da Tabela 18. Nota-se que os coeficientes são bem diferentes, assim como os sinais, sem que seja possível estabelecer alguma explicação plausível para todos os modelos.

Ao todo foram desenvolvidos cinco modelos, um para cada componente identificado pela Análise de Componentes Principais (ACP). Estes modelos tiveram por finalidade verificar o efeito individual de cada componente sobre a Variável Dependente (VD). Os resultados de ajuste destes modelos estão na Tabela 16. Todos os resíduos foram significativos, inclusive de N1 (ano). Isto é uma informação importante, pois os períodos de análise são diferentes para alguns componentes, conforme a limitação das variáveis explicativas que os compõem. O menor período foi de cinco anos, com os dados bienais de C1 (proficiência). Em termos de *deviance*, C1 (proficiência) foi o que apresentou o melhor ajuste dos cinco componentes. O pior ajuste foi obtido por C5 (variáveis agropecuárias). O pior ajuste de C5 pode ser explicado pela pouca produção de riqueza do setor agropecuário em comparação aos outros setores, o que proporciona pequeno impacto sobre o PIB per capita (VD). A qualidade dos dados também pode explicar este pior ajuste. Os dados de C5 possuem problemas com valores zero em grande frequência, causando assimetria positiva persistente e de difícil correção por meio de transformação. É o caso das variáveis de saúde de C4. Não por acaso, a *deviance* de C4 (bem-estar) indica o segundo pior ajuste. Em termos de variância explicada, C3 é o que apresenta menor variância explicada por N2 (município). Ou seja, a variância de DIS entre os anos explica uma boa parte da variância total (dado que se trata apenas de variáveis de C3). Já C4 é a que tem maior variância explicada pelo N2. A variação na produção de riqueza (VD) é menos relacionada com a variação de “bem-estar” ao longo dos anos e mais relacionada com as diferenças entre os municípios. Novamente, isto pode ser fruto da qualidade ruim dos dados de saúde.

De forma geral, os componentes ligados à educação (C1, C2 e C3) são os que apresentaram melhor ajuste dos modelos. “Bem-estar” (C4) e “agropecuária” (C5) foram os de pior ajuste. Este problema de ajuste parece estar relacionado com os problemas de assimetria das variáveis explicativas. Também é digno de nota o fato do

número de casos válidos ter alguma relação com o *deviance*. Os modelos com menos casos válidos apresentaram os menores *deviance*, enquanto os com maior número de casos, apresentaram maior *deviance*. De certa forma, faz sentido, uma vez que o maior número de casos aumenta a variância dentro do modelo, o que implica em maior dificuldade para encontrar uma solução convergente. Um outro aspecto interessante pode ser verificado pelo R^2 de N2. Os componentes C1, C4 e C5 têm nos municípios a maior fonte de explicação da variância, enquanto que C2 e C3 (DIS) têm os anos como a maior fonte de explicação da variância total. Isto é razoável, na medida que DIS é uma variável intimamente ligada ao fator tempo.

7. Conclusões

Por esta pesquisa não se tratar de um desenho experimental ou quase-experimental, não se tinha um grupo controle ou um grupo contrafactual com o qual se comparar e, assim, determinar o impacto do PBF. Este seria o caminho “fácil” de fazer ciência, controlando as variáveis e o seu contexto.

O caminho escolhido no presente estudo foi o de observar o comportamento dos dados ao longo de um período e, a partir de eventuais mudanças, procurar associar com possíveis causas. A dificuldade residiu em determinar que dado coletar de maneira que permitisse inferir o impacto da intervenção. Uma vez escolhidos os dados, a segunda maior dificuldade foi lidar com as suas limitações, sejam temporais, geográficas ou de representatividade. Para resolver este problema, foram utilizadas duas estratégias de análise: (1) determinar a relação direta entre as variáveis do PBF e a VD (modelo M1); e (2) verificar as diferenças entre dois modelos, um antes (M3) e outro depois (M4) da implementação do PBF.

A primeira estratégia permitiu concluir que pequenas quedas na taxa de crescimento do município ampliam em muito, e num curto espaço de tempo, o número de famílias que passam a depender do PBF. Logo, como estratégia de intervenção do governo, é mais interessante procurar ampliar a produção de riqueza a fim de evitar a dependência ao PBF. Porém, caso seja necessária a intervenção do PBF, ele se mostra um mecanismo barato de redução da pobreza.

Quanto à segunda estratégia de análise, ele foi inconclusivo em determinar o impacto do PBF na diminuição da pobreza. O que se pode afirmar é que existe sim uma diferença entre o estado das variáveis antes e depois de 2003. Dadas as variáveis

utilizadas no modelo, estas diferenças se concentram nos percentuais de distorção idade-série (DIS). As variáveis ligadas à saúde, agropecuária e setor produtivo se mostraram relativamente estáveis nos dois períodos. Logo, o que causou a diferença está ligado às variáveis educacionais. Não só o PBF tem algum efeito sobre estas variáveis por meio da exigência de frequência escolar dos filhos dos beneficiários. Também afetam estas variáveis os investimentos diretos em educação, não só em infraestrutura, mas em ações sociais ligadas às escolas, como os programas de merenda escolar, vacinação, alfabetização de adultos e outros. Portanto, as causas parecem ser de natureza difusa. O fato do PBF ter como bases de ação ampliar a renda, aprimorar a saúde e melhorar a escolaridade e os modelos não identificarem diferenças em saúde e produção de riqueza (renda), é bem improvável que o PBF seja de fato a principal causa das diferenças observadas em DIS.

Além da natureza difusa do que afeta a educação no Brasil, também deve-se lembrar que mudanças de hábitos, seja de consumo, de alimentação, ou de saúde, não ocorrem instantaneamente. É importante frisar que as diferenças observadas podem na realidade não indicar mudanças, mas apenas se tratar de picos transitórios. Da mesma forma, a estabilidade pode ser vista como uma mudança, só que de maneira muito lenta. Tudo depende de o que se considera como “período longo”. Uma diferença observada em dois anos num intervalo de cem anos de fato pode ser considerado um “pico”. Mas em um período de quatro anos, não. Isto é um ponto a favor para se atribuir causalidade ao PBF. As suas bases de ação lidam com variáveis que demandam tempo para mudanças. Assim, talvez não fosse apenas a técnica ou o modelo errados, mas um intervalo de tempo curto estudado. Suporta esta hipótese o fato da causalidade de BFF sobre PIBPC ser pequena, o que indicaria um estágio inicial do programa.

7.1 Limitações do estudo

O uso de bases de dados públicas e oficiais foi a maior vantagem e desvantagem desta pesquisa. Estes dados estão prontamente disponíveis. Usá-los ajuda a amortizar o investimento público feito em sua coleta, armazenamento e distribuição. Estas bases permitem fazer inferências que tem força e respaldo oficiais. Finalmente, tem uma abrangência nacional, permitindo comparar diferentes realidades regionais. Entretanto, usar tais dados em pesquisa científica requer muitos cuidados.

O primeiro cuidado diz respeito ao seu caráter político. Os levantamentos nem sempre fazem parte de uma política nacional, mas sim de governo. Alternâncias de poder definem mudanças nas diretrizes dos levantamentos de dados que afetam os métodos de coleta, a sua frequência, a escolha dos indicadores e o nível de agregação das informações. Um exemplo são os dados do PBF, que apenas recentemente foram tabulados e disponibilizados ao grande público por causa do esforço do atual governo federal em ampliar as ações deste programa. Outro exemplo são os dados do Inep dos levantamentos do Saeb e do Prova Brasil. A quantidade de dados é imensa e, aparentemente, parte disso é consequência da interferência dos vários governos que se sucederam e seus respectivos programas para a educação. Porém, não há como imaginar que dados oficiais fossem despidos de tais influências. Nem no Brasil e nem em governos de outros países.

O segundo cuidado faz referência à qualidade dos dados. À guisa de exemplos, listam-se alguns casos. Foram observadas inconsistências entre os dados do censo escolar e dos sistemas de avaliação do Saeb. Os dados da zona rural foram coletados por conveniência entre 2000 a 2006, dificultando qualquer inferência sobre o estado da educação no meio rural. Os dados do PBF não apresentavam dados confiáveis para 2008 (Figura 5 e Figura 6). Os dados de cobertura do PBF apresentavam percentuais acima de 100% (Figura 7), indicando que os dados estavam subestimando a população de pessoas elegíveis ao programa ou que estavam superestimando o número de beneficiários. Dada a nota de esclarecimento da CEF (“Caixa admite equívoco”, 2013), admite-se que os dados estão superestimando o número de beneficiários. Os dados de saúde computam óbitos nos locais de tratamento dos pacientes e de nascimento dos recém-nascidos e não no local de origem destes que procuram atendimento médico. Assim, os dados ficaram inflados nos grandes centros urbanos, não refletindo a qualidade da saúde da população em seus respectivos locais de origem, o que não foi útil a este trabalho. Alguns problemas de qualidade podem ser contornados, outros não. O importante é compreender que os dados apresentam algum grau de limitação para inferências e que se deve respeitá-los.

O terceiro cuidado está relacionado com a expectativa de que os bancos de dados têm um número elevado de casos. Por causa dos problemas de métodos empregados na coleta e da qualidade dos dados, apesar da disponibilidade de grandes bases oficiais, o número de casos efetivamente válidos é baixo. O número de dados

missing é grande. Se as perdas de casos válidos fossem aleatórias, um certo grau de inferência estatística poderia ser associado aos dados. Porém, normalmente os dados mais abundantes provêm de locais com maiores facilidades de coleta, o que se limita geograficamente a capitais e grandes centros urbanos. Para pesquisas com foco nas porções mais pobres da população, os dados são escassos.

O quarto cuidado com os dados é que nem sempre o significado dos dados é explícito nas bases de dados. Enquanto os dados de IBGE e Inep se encontram relativamente bem documentados (a despeito de erros encontrados nos dicionários de variáveis do Inep e na falta de descrição de algumas variáveis como as de Valor Adicionado Bruto do IBGE), os dados de saúde e do PBF não apresentam qualquer descrição dos métodos de coleta, o que dificulta determinar o que os dados representam e quão bem representam.

Quanto à técnica de regressão multinível, ela é uma opção interessante para a análise de séries históricas. Porém, modelos com dois anos de dados não justificam o emprego desta análise. Mesmo com um horizonte de onze anos, a variância pode não ser suficiente para se detectar relações entre a variável dependente e as explicativas. Dados em nível nacional, como os utilizados no presente estudo, que se expressam na forma de pequenas mudanças em largos períodos, precisam do maior número de anos possível para análise. Onze anos de horizonte permitiram variâncias significativas em nível 1 (N1), mas ainda com correlações intraclasse (ICC) elevadas. Se por um lado ICC elevadas justificam o uso de análise multinível em modelos convencionais (sujeito no nível 1 e grupo de sujeitos no nível 2), no caso de medidas repetidas é interessante que o nível 1 tenha mais participação na variância, pois é onde se encontram as medidas. A técnica também é muito sensível a dados *missing*. Os dados oficiais são fartos de casos *missing* devido às características amostrais de algumas bases. Diferentemente do que se encontra em literatura sobre a robustez da técnica de análise multinível a casos *missing*, ela é um tipo de regressão multivariada e, como tal, sofre com a perda de dados.

7.2 Contribuições

Duas contribuições importantes são apresentadas neste trabalho. A primeira, o uso de dados oficiais para as análises. A segunda, o emprego da técnica de análise multinível de medidas repetidas sobre séries históricas. Normalmente, os dados

oficiais são utilizados pontualmente ou como informação de contexto para pesquisas científicas. Poucas vezes são de fato utilizados como a fonte principal de dados. Em parte, esta tendência ocorre pela impossibilidade de criar um ambiente de controle experimental a fim de estabelecer os casos necessários para comparação dos resultados. Além disso, os dados oficiais são sempre *ex-post* o fenômeno. A técnica de análise multinível de medidas repetidas é comumente utilizada para avaliação de testes psicométricos aplicados em diferentes momentos dos indivíduos. Esta pesquisa adaptou esta técnica para a análise de séries históricas, substituindo a abordagem tradicional da Econometria de usar modelos lineares gerais de regressão (MLG). A desvantagem dos MLG convencionais é a adição de erro Tipo I por causa do uso de contrastes para medir as diferenças entre grupos.

Assim, para avaliações de impactos de programas sociais de tipo *ex-post* e onde o grupo controle (contrafato) e o grupo tratamento não são definidos, a análise multinível de medidas repetidas mostrou-se mais uma opção interessante e viável de análise.

7.3 Recomendações e sugestões

Para usar dados oficiais é preciso que o pesquisador tenha em mente que os dados têm algum nível de influência política. Não no sentido de dados manipulados, mas na definição das prioridades de coleta, dos níveis de análise, da frequência e mesmo da escolha de quais aspectos de um dado fenômeno serão registrados. Os dados públicos oficiais têm uma finalidade de fonte de dados para a gestão e não especialmente para a pesquisa acadêmica. Assim, nem sempre os dados oficiais possuem as características necessárias a uma análise científica mais rigorosa. Cabe mencionar a útil diferenciação feita pelo Inep entre a Avaliação Nacional do Ensino Básico (Aneb), com finalidade gerencial, e a Avaliação Nacional do Rendimento Escolar (Anresc), com finalidade de estudos científicos e de pesquisa em geral. A bem da “neutralidade científica” é preciso manter o foco nos pressupostos científicos de replicabilidade, em inferências à luz de teorias e especial atenção no cruzamento com outras fontes de dados.

A despeito destas recomendações, é preciso sim estimular o uso destes dados na pesquisa científica. O principal motivo é se aproximar mais da realidade, crítica constante impingida à academia por seus ensaios e experimentos em ambientes

rigorosamente controlados. Não deve ser o dado a trabalhar em favor da técnica, mas o contrário. A técnica deve se sujeitar ao dado. E isso só é possível com a criatividade do pesquisador em combinar estas técnicas a fim de desenhar a realidade do dado tal como é. É estimular o pesquisador a encarar o desafio da pesquisa como uma pequena obra de arte, de genialidade. Uma consequência desta “ditadura” do dado é a esterilização da riqueza das pesquisas. Os dados da realidade são “sujos” por natureza, mas talvez o que se denomine sujeira, como nos programas de reciclagem, seja apenas um ponto de vista. Não será na complexidade do dado, nas suas infrações à norma, nos seus melindres, que reside de fato o principal tesouro da ciência: o diferente, o novo?

8. Referências

- Almeida, E. (2000). *Programas de garantia de renda mínima: inserção social ou utopia?* São Paulo: EDUC-FAPESP.
- Alves, M. T. G., & Soares, J. F. (2007) As pesquisas sobre o efeito das escolas: contribuições metodológicas para a sociologia da educação. *Sociedade e Estado*, 22(2), 435-473.
- Amaral, N. C. (2012). *O Brasil e o novo PNE: expansão-qualidade-financiamento*. Artigo apresentado no VII Seminário da ANPAE Centro-Oeste, Políticas e Gestão da Educação no Centro-Oeste do Brasil: Pesquisas e Práticas, Goiânia, GO. Recuperado de <http://www.anpae.org.br/seminario/ANPAE2012/Textos/NelsonAmaral.pdf>
- Andrade, C. L. T. de, & Szwarcwald, C. L. (2007). Desigualdades sócio-espaciais da adequação das informações de nascimentos e óbitos do Ministério da Saúde, Brasil, 2000-2002. *Cad. Saúde Pública*, 23(5), 1207-1216.
- Andrade, J. M. de, & Laros, J. A. (2007). Fatores associados ao desempenho escolar: estudo multinível com dados do Saeb/2001. *Psicologia: Teoria e Pesquisa*, 23(1), 33-42.
- Araújo, C. H., & Luzio, N. (2005). *Avaliação da educação básica: em busca da qualidade e equidade no Brasil*. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
- Arelaro, L. R. G. (2005). O ensino fundamental no Brasil: avanços, perplexidades e tendências. *Educ. Soc.*, 26(92), 1039-1066. Recuperado de <http://www.scielo.br/pdf/%0D/es/v26n92/v26n92a15.pdf>
- Bagaka's, J. G. (2011). The role of teacher characteristics and practices on upper secondary school student's mathematics self-efficacy in Nyanza province of Kenya: A multilevel analysis. *International Journal of Science and Mathematics Education*, 9, 817-842.
- Barbosa Filho, F. H., & Pessôa, S. (2008). Retorno da educação no Brasil. *Pesquisa e Planejamento Econômico*, 38(1): 97-125.
- Barbosa Filho, F. H., & Pessôa, S. (2009). Educação, crescimento e distribuição de renda: a experiência brasileira em perspectiva comparada. Em: F. Veloso, S. Pessôa, R. Henriques, & F. Giambiagi (Org.), *Educação básica no Brasil: construindo o país do futuro* (pp. 51-72). Rio de Janeiro: Elsevier.
- Barbosa, M. E. F., & Fernandes, C. (2000). Modelo multinível: uma aplicação a dados de avaliação educacional. *Estudos em Avaliação Educacional*, 22, 135-154.
- Barros, R. P. de, & Mendonça, R. S. P. de. (1995). *Os determinantes da desigualdade no Brasil* (Texto para Discussão nº. 377). Rio de Janeiro: Ipea. Recuperado de http://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_0377.pdf

- Barros, R. P. de, Carvalho, M. de, & Franco, S. (2006). O papel das transferências públicas na queda recente da desigualdade de renda brasileira. In R. P. de Barros, M. M. Foguel, & G. Ulyssea (Orgs.), *Desigualdade de Renda no Brasil: uma análise da queda recente* (Vol. 2). Brasília, DF: Instituto de Pesquisa Econômica Aplicada.
- Caixa admite equívoco em informação sobre Bolsa Família e pede desculpas. (27 de maio de 2013). *Folha de São Paulo*. Recuperado de <http://www1.folha.uol.com.br/poder/2013/05/1285790-caixa-admite-equivoco-em-informacao-sobre-bolsa-familia-e-pede-desculpas.shtml>
- Camelo, R. de S., Priscilla, A. T., & Saiani, C. C. S. (2009). Alimentação, nutrição e saúde em programas de transferência de renda: evidências para o Programa Bolsa Família. *Economia*, 10(4), 685-713. Recuperado de http://www.anpec.org.br/revista/vol10/vol10n4p685_713.pdf
- Castro, J. A., & Modesto, L. (Orgs.). (2010a). *Bolsa Família 2003 – 2010: avanços e desafios* (Vol. 1). Brasília, DF: Instituto de Pesquisa Econômica Aplicada. Recuperado de http://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/livro_bolsafamilia_vol1.pdf
- Castro, J. A., & Modesto, L. (Orgs.). (2010b). *Bolsa Família 2003 – 2010: avanços e desafios* (Vol. 2). Brasília, DF: Instituto de Pesquisa Econômica Aplicada. Recuperado de http://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/livro_bolsafamilia_vol2.pdf
- Coleman, J. S. (1988). Social capital in the creation of human capital. *The American Journal of Sociology (Supplement)*, 94, S95-S120.
- Collins, R. (1979). *The credential society: An historical sociology of education and stratification*. Nova Iorque: Academic.
- Contreras, S. (2008). Child labor participation, human capital accumulation, and economic development. *Journal of Macroeconomics*, 30(1), 499-512.
- Curi, A. Z., & Menezes Filho, N. A. (2009). A relação entre educação pré-primária, salários, escolaridade e proficiência escolar no Brasil. *Estudos Econômicos, São Paulo*, 39(4), 811-850. doi: 10.1590/S0101-41612009000400005
- De Janvry, A., & Sadoulet, E. (2000). Growth, poverty, and inequality in Latin America: A causal analysis, 1970-94. *Review of Income and Wealth*, 46(3), 267-287.
- Felício, F., & Fernandes, R. (2005). O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental no estado de São Paulo. *Anais do Encontro Nacional de Economia* 33. Recuperado de <http://EconPapers.repec.org/RePEc:anp:en2005:157>.

- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3ªed.). Boston: Pearson Education.
- Fundo Internacional de Desenvolvimento Agrícola. (2011). *Habilitando os pobres rurais a superar a pobreza no Brasil*. Recuperado de http://www.ifad.org/operations/projects/regions/PL/factsheet/brazil_p.pdf
- Glewwe, P., & Kassouf, A. L. (2012). The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of Development Economics*, 97, 505-517. doi:10.1016/j.jdeveco.2011.05.008
- Granger, C. W. J. (1969). Investigating causal relations by econometric models in cross-spectral methods. *Econometrica*, 37(3), 424-438.
- Helfand, S. M., Rocha, R., & Vinhais, H. E. F. (2009). Pobreza e desigualdade de renda no Brasil rural: uma análise da queda recente. *Pesquisa e Planejamento econômico*, 39(1), 59-80.
- Heller, L. (1998). Relação entre saúde e saneamento na perspectiva do desenvolvimento. *Ciência & Saúde Coletiva*, 3(2), 73-84.
- Higgins, S. (2012). The impact of Bolsa Família on poverty: Does Brazil's conditional cash transfer program have a rural bias? *Journal of Politics Society*, 23(1), 88-125.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2ª ed.). New York, NY: Routledge.
- Instituto Brasileiro de Geografia e Estatística. (2009). *Pesquisa Nacional por Amostra de Domicílio: síntese de indicadores sociais – Educação*. Recuperado de ftp://ftp.ibge.gov.br/Indicadores_Sociais/Sintese_de_Indicadores_Sociais_2009/Tabelas/educacao.zip
- Instituto Brasileiro de Geografia e Estatística. (2010). *Censo Demográfico 2010: Trabalho e rendimento – Resultados da amostra*. Recuperado de ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Trabalho_e_Rendimento/censo_trabalho_e_rendimento.pdf
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2006a). *Microdados do Saeb 2001: manual do usuário*. Brasília, DF: Ministério da Educação.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2006b). *Microdados do Saeb 2003: manual do usuário*. Brasília, DF: Ministério da Educação.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2007). *Panorama da educação no campo*. Brasília: Inep.

- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2008). *Microdados do Saeb 2005: manual do usuário*. Brasília, DF: Ministério da Educação.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2009). *Microdados da Prova Brasil 2007: manual do usuário*. Brasília, DF: Ministério da Educação.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep. (2011a). *Percentual do Investimento Total em Relação ao PIB por Nível de Ensino*. Recuperado de http://download.inep.gov.br/download/estatisticas/gastos_educacao/docs/indicadores_financeiros/pib_total_nivel.xls
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2011b). *Microdados da Prova Brasil 2009: manual do usuário*. Brasília, DF: Ministério da Educação.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2012a). *Censo escolar 2011*. Recuperado de ftp://ftp.inep.gov.br/microdados/micro_censo_escolar_2011.zip
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2012b). *Microdados Prova Brasil 2011: manual do usuário*. Brasília, DF: Ministério da Educação.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2013). *Estatísticas do Ideb: Brasil*. Brasília, DF: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Recuperado de <http://portal.inep.gov.br/web/portal-ideb/planilhas-para-download>
- Instituto de Pesquisa Econômica Aplicada - IPEA (2013). *Ipeadata: IPCA – Geral*. Recuperado de <http://www.ipeadata.gov.br>
- Kassouf, A. L. (2007). O que conhecemos sobre o trabalho infantil? *Nova Economia*, 17(2), 323-350.
- Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook of impact evaluation: Quantitative methods and practices*. Washington, DC: The World Bank. doi: 10.1596/978-0-8213-8028-4
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: an easy-to-use computer program carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2). Recuperado de <http://pareonline.net/getvn.asp?v=12&n=2>
- Lindert, K., Linder, A., Hobs, J., & Brière, B. de la. (2007). *The nuts and bolts of Brazil's Bolsa Família Program: Implementing conditional cash transfers in a decentralized context* (Social Protection Discussion Paper nº 0709). Washington, DC: The World Bank.

- Liu, O. L. (2012). Student evaluation of instruction: In the new paradigm of distance education. *Research in High Education*, 53, 471-486. doi: 10.1007/s11162-011-9236-1
- Machado, D. C., & Gonzaga, G. (2007). *O impacto dos fatores familiares sobre a defasagem idade-série de crianças no Brasil* (Texto para Discussão nº 546). Rio de Janeiro, RJ: Departamento de Economia da PUC-Rio.
- Mello, G. N. de. (1991). Políticas públicas de educação. *Estudos Avançados*, 5(13), 7-47.
- Mendes, M. (2001). Descentralização do ensino fundamental: avaliação de resultados do Fundef. *Planejamento e Políticas Públicas*, 24, 27-51.
- Ministério da Educação. (2012). *Pronacampo*. Recuperado de http://portal.mec.gov.br/index.php?option=com_content&view=article&id=17608
- Ministério do Desenvolvimento Agrário – MDA. (2009). *Agricultura familiar no Brasil e o censo agropecuário 2006*. Recuperado de http://www.mda.gov.br/portal/publicações/download_orig_file?pageflip_id=3697318
- Ministério do Desenvolvimento Social e Combate à Fome – MDS (2013). *Secretaria de Avaliação e Gestão da Informação – SAGI: Gestão da Informação*. Recuperado de <http://aplicacoes.mds.gov.br/sagi/FerramentasSAGI/index.php?group=1>
- Ministério do Desenvolvimento Social e Combate à Fome – MDS (2011). *Bolsa Família completa 8 anos com investimentos de R\$ 76 milhões*. Recuperado em <http://www.mds.gov.br/saladeimprensa/noticias/2011/outubro/bolsa-familia-completa-8-anos-com-investimentos-de-r-76-bilhoes>
- Ministério do Desenvolvimento Social e Combate à Fome – MDS. Secretaria de Avaliação e Gestão da Informação (2010). *Síntese das pesquisas de avaliação de programas sociais do MDS: versão atualizada e revisada 2006-2010* (Cadernos de Estudos: Desenvolvimento Social em Debate nº 13). Brasília, DF: Ministério do Desenvolvimento Social e Combate à Fome, Secretaria de Avaliação e Gestão da Informação.
- Ministério da Educação, Assessoria de Comunicação Social (s.d.). Quadro comparativo – Fundeb-Fundef. Recuperado de <http://www.fnde.gov.br/arquivos/category/167-fundeb?download=3018:quadro-comparativo-fundeb-fundef.pdf>
- Nações Unidas (2000). *Declaração do Milênio* (Resolução A/RES/55/2). Recuperado de <http://www.unric.org/html/portuguese/uninfo/DecdoMil.pdf>
- Nasser, F., & Hagtvet, K. A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, 47(5), 559-590. doi: 10.1007/s11162-005-9007-y

- Oliveira, R. de F., & Teixeira, B. de B. (2008). As políticas de financiamento da educação básica na última década: do Fundef ao Fundeb. *Anais do Simpósio Internacional: o Estado e as Políticas Educacionais no Tempo Presente*, 4. Recuperado de <http://www.simpósioestadopoliticas.ufu.br/imagens/anais/pdf/DC10.pdf>
- Osborne, J. Q. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, 15(12), 1-9.
- Piana, M. C. (2009). *A construção do perfil do assistente social no cenário educacional*. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica,. Recuperado de <http://books.scielo.org/id/vwc8g/pdf/piana-9788579830389-02.pdf>
- Pochmann, M. (2010). Apresentação Ipea. In J. A. de Castro & L. Modesto (Eds.), *Bolsa Família 2003-1020: avanços e desafios* (Vol. 1, pp. 7-10). Brasília: Ipea.
- Presidência da República Federativa do Brasil. (s.d.). *Fome Zero: estrutura*. Recuperado de <http://www.fomezero.gov.br/estrutura>
- Ramos, G. P. (2009). Fundef versus Fundeb: uma análise introdutória sobre as continuidades e as descontinuidades da política de valorização docente. *Revista Ibero-Americana de Estudos em Educação*, 4(1). Recuperado de <http://seer.fclar.unesp.br/iberoamericana/article/download/2696/2406>
- Rasella, D., Aquino, R., Santos, C. A. T., Paes-Sousa, R., & Barreto, M. L. (2013). Effect of a conditional cash transfer programme on childhood mortality: a nationwide analysis of Brazilian municipalities. *The Lancet*, doi:10.1016/S0140-6736(13)60715-1
- Revallion, M. (2010). *Poverty lines across the World* (Policy Research Working Paper n° 5284). Washington, DC: Development Research Group, World Bank.
- Santos, L. L. de C. P. (2002). Políticas públicas para o ensino fundamental: parâmetros curriculares nacionais e Sistema Nacional de Avaliação (Saeb). *Rev. Educ. & Soc.*, 23(80), 346-367.
- Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1), 1-17.
- Schultz, T. W. (1971). *Investment in human capital*. New York: The Free Press.
- Secretaria de Estado de Educação do Distrito Federal (1997). *Programa Bolsa-Escola e Poupança-Escola: governo do Distrito Federal*. Brasília: Governo do Distrito Federal.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs: For generalized causal inference*. Belmont, CA: Wadsworth.
- Silva, M. O. da S. e (Coord.). (2008). *O Bolsa Família: no enfrentamento à pobreza no Maranhão e Piauí*. São Paulo, SP: Cortez.

- Silva, M. O. da S. e (Coord.). (2011). *Avaliando o Bolsa Família*. São Paulo, SP: Cortez.
- Silveira Neto, R. da M. (2010). Impacto do Programa Bolsa Família sobre a frequência à escola: estimativas a partir de informações da Pesquisa Nacional por Amostra de Domicílio (Pnad). Em J. A. Castro & L. Modesto (Orgs.), *Bolsa Família 2003 – 2010: avanços e desafios* (Vol. 2, pp. 53-71). Brasília, DF: Instituto de Pesquisa Econômica Aplicada.
- Soares, F. V., Soares, S., Medeiros, M., & Osório, R. G. (2006, Dezembro). *Programas de transferências de renda no Brasil: impactos sobre a desigualdade*. Artigo apresentado no XXXIV Encontro Nacional de Economia, Salvador, BA.
- Soares, S., Souza, P. H. G. F. de, Osório, R. G., & Silveira, F. G. (2010). Os impactos do benefício do Programa Bolsa Família sobre a desigualdade e pobreza. Em J. A. Castro & L. Modesto (Orgs.), *Bolsa Família 2003 – 2010: avanços e desafios* (Vol. 2, pp. 27-52). Brasília, DF: Instituto de Pesquisa Econômica Aplicada.
- Soares Neto, J. J., Jesus, G. R. de, Karino, C. A., & Andrade, D. F. de (2013). Uma escala para medir a infraestrutura escolar. *Estudos em Avaliação Educacional*, 24(54), 78-99.
- Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3), 355-74.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, Massachusetts: Allyn and Bacon.
- Tejada, C. A. O., Jacinto, P. de A., & Santos, A. M. A. dos (2008). Pobreza e saúde: evidências de causalidade em painel de dados para o Brasil. Em M. A. Serra (Coord.), *XI Encontro Regional de Economia – Anpec Sul*. PPGDE – UFPR, Curitiba.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220.
- Todos pela Educação (2011). *De olho nas metas 2011: quarto relatório de monitoramento das 5 metas do Todos pela Educação*. São Paulo: Moderna.
- Waiselfisz, J. J., Abramovay, M., & Andrade, C. (1998). *Bolsa-Escola: melhoria educacional e redução da Pobreza*. Brasília, DF: Unesco.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 462-482.
- Weissheimer, M. A. (2010). *Bolsa Família: avanços, limites e possibilidades do programa que está transformando a vida de milhões de famílias no Brasil* (2ª ed.). São Paulo, SP: Fundação Perseu Abramo.

Wu, D. Y. T. (2004). *Teoria da sinalização e educação* (Dissertação de mestrado). Recuperado de <http://www.ppge.ufrgs.br/giacomo/arquivos/eco02036/diana-yi-ting-wu-sinalizacao-educacao.pdf>

Yamada, G., & Castro, J. F. (2011). Educational attainment, growth and poverty reduction within the MDG framework: Simulations and costing for the Peruvian case. *Journal of Economic Policy Reform*, 12(1), 57-73.

Anexo

Sintaxe para cálculo de lambda da transformação de Box-Cox, adaptado de Osborne (2010).

```
COMPUTE var1 = sua_variavel.
VECTOR lam(31) /x1(31).
LOOP idx=1 to 31.
  COMPUTE lam(idx)= -2.1 + idx * .1.
  DO IF lam(idx) = 0.
    COMPUTE x1(idx) = ln(var1).
  ELSE.
    COMPUTE x1(idx) = (var1**lam(idx) - 1) / lam(idx).
  END IF.
END LOOP.

FREQUENCIES VARIABLES = var1 x11 TO x131
  /format = notable
  /statistics = minimum maximum SKEWNESS
  /histogram normal
  /order=analysis.

EXECUTE .
```