UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

**IMPACTOS DO SANEAMENTO SOBRE SAÚDE E EDUCAÇÃO: UMA ANÁLISE
ESPACIAL**

Juliana Souza Scriptore

Orientador: Prof. Dr. Carlos Roberto Azzoni

SÃO PAULO

2016

Prof. Dr. Marco Antonio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Hélio Nogueira da Cruz
Chefe do Departamento de Economia

Prof. Dr. Márcio Issao Nakane
Coordenador do Programa de Pós-Graduação em Economia

JULIANA SOUZA SCRIPTORE

**IMPACTOS DO SANEAMENTO SOBRE SAÚDE E EDUCAÇÃO: UMA ANÁLISE
ESPACIAL**

Tese apresentada ao Programa de Pós-Graduação do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Carlos Roberto Azzoni

VERSÃO CORRIGIDA

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

SÃO PAULO

2016

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Scriptore, Juliana Souza.

Impactos do saneamento sobre saúde e educação: uma análise espacial / Juliana Souza Scriptore. -- São Paulo, 2016.
206 p.

Tese (Doutorado) – Universidade de São Paulo, 2016.

Orientador: Carlos Roberto Azzoni.

1. Saneamento básico. 2. Educação. 3. Saúde. 4. Doenças relacionadas ao saneamento ambiental inadequado – DRSAI. 5. Econometria espacial. I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 628

AGRADECIMENTOS

Agradeço, em primeiro lugar, ao meu orientador Carlos Roberto Azzoni por compartilhar sua sabedoria de forma tão generosa e por todo aprendizado que tive com os seus ensinamentos, exemplos e convivência. Além de todo apoio acadêmico, agradeço por ter me estimulado a não desistir de prestar o doutorado, pelo apoio nos momentos difíceis de viagens a Varginha e pela sua orientação exigente, mas que, ao mesmo tempo, sempre me inspirou tranquilidade. Agradeço à professora Paula Pereda pelas valiosas contribuições e por toda ajuda despendida com esse trabalho, sempre de forma acolhedora e receptiva. Agradeço também todo conhecimento que me proporcionou no período da monitoria e nas aulas de Econometria Espacial; mas, principalmente, agradeço sua amizade e apoio nos momentos mais críticos do doutorado, desde o processo seletivo até os momentos finais. Agradeço também a receptividade e importante contribuição acadêmica do professor Naércio Menezes e também por me incentivar a desenvolver o artigo da disciplina que contribuiu para um dos capítulos dessa Tese. Agradeço ao professor Alexandre Nunes de Almeida pelo auxílio nos momentos de dúvidas e incertezas ao longo do doutorado e, principalmente, pelo encorajamento presente em cada conversa. Agradeço ao professor André Chagas pelas suas colaborações dadas no exame de qualificação e seminários; além dos conselhos sobre caminhos futuros. Ao professor Rudinei Toneto Júnior por ter me iniciado nos estudos sobre Saneamento Básico, por ter sido sempre tão atencioso e preocupado e por ter me indicado para a orientação com o professor Azzoni. Aos professores do NEREUS, dedico agradecimento especial ao professor Joaquim Guilhoto e ao professor Eduardo Haddad.

Agradeço o apoio institucional do IPE-USP e toda ajuda recebida pela seção de Pós-Graduação: Pinho, Alessandra e Adriana Miranda. Igualmente agradeço ao ICSA-UNIFAL pela concessão do período de afastamento para o Doutorado e colaboração do colega Leandro Riveli com todos os trâmites, até então desconhecidos, para viabilizar a licença de vários professores.

Não tenho palavras para agradecer o apoio nos mais diversos momentos, da minha grande amiga Denise Imori, tanto em relação às incansáveis conversas e desabafos quanto em relação à ajuda acadêmica. Agradeço o apoio nas inúmeras consultas (muitas vezes, inoportunas) solicitadas ao Diogo, Lucas e Moisés mas, acima de tudo, agradeço a valiosa amizade de cada um de vocês. Aos amigos de Varginha, agradeço, em especial, ao Deive e ao Leandro Ferreira que também acompanharam de perto e torceram por cada etapa concluída do Doutorado. Dario, Andréa Lepine Ana Barufi, Joelson, Diana e Leopoldo também foram pessoas importantes nessa jornada.

Agradecimento especial devo aos grandes responsáveis por cada passo do meu trajeto. A minha mãe Célia e ao meu pai Genaro pelo suporte, carinho, amor e também pelo investimento em minha educação que, por ter sido tão dificultoso, me fez valorizá-los ainda mais e me fez mais forte para lutar em busca de novas conquistas. Ao meu namorado João Paulo Soares Moreira agradeço o amor, o companheirismo, a paciência tantas vezes testada nos meus momentos de desespero, por sempre acreditar em mim, me fazer uma pessoa melhor e por estar ao meu lado mesmo quando eu não merecia. Agradeço à minha irmã Gabriela pela preocupação constante, pelos conselhos, pela torcida e por todo seu amor e dedicação. Aos tios Carmen, Sandrinha, Carlinhos e cunhado Alan agradeço a torcida, as lembranças e todo carinho que vocês me deram e me fornecem diariamente.

RESUMO

Este trabalho investiga o efeito do saneamento sobre educação, ou seja, as consequências educacionais das crianças que apresentam estado de saúde debilitado por terem contraído Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI). Nesse sentido, condições de infraestrutura dos domicílios, expressas pela variável de acesso a saneamento básico, por impactar a saúde dos indivíduos, é mais um dos determinantes dos indicadores educacionais. A partir de dados municipais e distritais, o efeito do saneamento sobre educação foi obtido em duas etapas. Na primeira, utilizou-se dados por município do Censo Demográfico (IBGE), Censo Escolar (INEP), Sistema Nacional de Informações Sobre Saneamento (SNIS), entre outros, para anos 2000 e 2010, com o objetivo de avaliar o efeito do saneamento sobre educação. Na segunda etapa, verificou-se a hipótese que dá suporte ao efeito da etapa anterior: o efeito do saneamento sobre a saúde. Essa investigação foi realizada por meio de dados por distrito construídos a partir do Censo Demográfico do Universo por setor censitário (IBGE), Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE, 2011), Diretório Nacional de Endereços (DNE, 2015) e Departamento de Informática do Sistema Único de Saúde (DATASUS, 2010). Ambas etapas fizeram uso de econometria espacial e não-espacial para estimação dos dois efeitos.

Os resultados da primeira etapa indicam que o efeito do aumento de uma unidade percentual no acesso a saneamento está associado a um aumento de 0,11 pontos percentuais na taxa de frequência escolar, a uma queda de 0,31 pontos percentuais na taxa de distorção idade-série e a também uma queda de 0,12 pontos percentuais na taxa de abandono escolar. Os resultados da segunda etapa indicam que, em geral, o acesso aos serviços de saneamento inadequados causa impacto positivo na taxa de incidência de internações por DRSAI (por mil habitantes), a variável de saúde considerada nesse estudo. Para a população de todas as idades, o impacto total médio do aumento de 1% na parcela de domicílios servidos por água da chuva armazenada em cisterna ou caixa de cimento é 0,217 casos por mil habitantes. Quando se considera a porcentagem dos domicílios que tem banheiro conectado a uma fossa rudimentar, esse impacto é 0,035 casos por 1000 habitantes. Por outro lado, no que diz respeito à variável adequada de saneamento, o impacto total médio de 1% na porcentagem de domicílios particulares permanentes conectados à rede de distribuição de água e que, além disso, na quadra onde estão localizados é inexistente condições de esgoto a céu aberto é -0,166 casos por mil habitantes. A variável de maior impacto total médio refere-se à porcentagem de domicílios que não tem banheiro e, além disso, possuem presença de esgoto a céu aberto em suas condições de entorno. O aumento em uma unidade percentual nessa variável gera aumento de 3,281 na taxa de incidência de internações por DRSAI. Outro resultado desse estudo é que a população mais jovem é a mais atingida pelo acesso a condições inadequadas dos serviços de saneamento básico tanto em relação à saúde quanto à educação. Por fim, a recomendação de política pública apontou que, se fossem implementadas políticas de mitigação das condições inapropriadas de saneamento, iniciando por diminuir em um por cento a porcentagem de domicílios que não têm banheiro e estão expostos a condições de esgoto a céu aberto o setor público da saúde economizaria recursos num valor correspondente a 4,4% do gasto federal total investido por ano em saneamento. Dessa forma, essa política pode ser uma estratégia para acelerar as metas de universalização do acesso aos serviços de saneamento básico no Brasil.

ABSTRACT

This dissertation studies the effect of sanitation on education, in other words, the educational consequences for children whose health has been affected by Diseases Related to Inadequate Environmental Sanitation (DRIES). In this sense, infrastructure conditions of dwellings, expressed by the variable of access to basic sanitation, for impacting the health of individuals, is one of the determinants of educational indicators. The effect of sanitation on education was obtained in two steps, from municipality and district level data. At the first step, we applied data at municipality level, from the Population Census (IBGE), the School Census (INEP), the National Information System on Sanitation (SNIS), among others, for the years 2000 and 2010, with the goal of evaluating the effect of sanitation on education. At the second step, we analyzed the hypothesis that supports the effect of the previous step: the effect of sanitation on health. This analysis was conducted applying data at district level, which was constructed from the results from the universe of the Population Census (IBGE) by census tract, the National Register of Addresses for Statistical Purposes (CNEFE, 2011), the National Address Directory (DNE, 2015) and the Department of Informatics of the Unified Health System (DATASUS, 2010). Both steps made use of spatial and non-spatial econometrics techniques for estimating the two effects. The results of the first step indicate that the effect of one percentage increase in access to sanitation is associated with an increase of 0.11 percentage points in school attendance rate, a decrease of 0.31 percentage points in age-grade distortion rate, and also a decrease of 0.12 percentage points in the school dropout rate. The results of the second step indicate that, in general, access to inadequate sanitation services has a positive impact on the incidence of hospitalizations because of DRIES (per thousand inhabitants), the health variable considered in this study. For the population of all ages, the average total impact of a 1% increase in the share of dwellings served by rain water (stored in cisterns or cement tanks) is 0.217 cases per thousand inhabitants. Considering the percentage of dwellings with toilets connected to rudimentary cesspools, this impact is 0.035 cases per thousand inhabitants. On the other hand, with regard to the appropriate sanitation variable, the average total impact of 1% increase in the share of permanent dwellings connected to the water distribution network and whose block does not present open sewage conditions is - 0.166 cases per thousand inhabitants. The variable with highest average total impact refers to the percentage of dwellings that have no toilet and, moreover, have the presence of open sewage in their surrounding conditions. The increase in a percentage unit in this variable generates an increase of 3.281 in the rate of incidence of hospitalizations because of DRIES. Another result of this study is that the younger population is the most affected by the poor conditions of access to basic sanitation services, both in terms of health on education. Finally, the public policy recommendations pointed out that, if the policies for mitigating inadequate sanitation conditions were implemented, starting with a one percentage decrease in the share of dwellings that do not have toilets and are exposed to open sewage, the health sector would save an amount corresponding to 4.4% of total federal spending invested annually in sanitation. Therefore, this policy can be a strategy to accelerate the targets for universal access to basic sanitation services in Brazil.

SUMÁRIO

1	INTRODUÇÃO.....	11
2	INDICADORES DE SANEAMENTO BÁSICO, EDUCAÇÃO E SAÚDE.	17
2.1	Saneamento Básico.....	17
2.2	Educação.....	20
2.3	Saúde	22
2.3.1	As Doenças de Saneamento Ambiental Inadequado (DRSAI) e os canais de transmissão.....	23
3	DADOS.....	31
3.1	Dados municipais	31
3.2	Estatísticas descritivas dos dados municipais.....	33
3.3	Dados distritais	37
3.4	Estatísticas Descritivas dos dados distritais.....	47
4	EFEITOS DO SANEAMENTO SOBRE EDUCAÇÃO.....	55
4.1	Saneamento básico e educação.....	55
4.2	Metodologia.....	60
4.2.1	Análise Exploratória dos Dados Espaciais (ESDA).....	66
4.2.2	Metodologia espacial.....	68
4.3	Resultados não-espaciais	70
4.4	Resultados espaciais	78
4.5	Comparação com o método de Máxima Verossimilhança.	86
5	EFEITO DO SANEAMENTO SOBRE A SAÚDE.....	89
5.1	O modelo teórico de Grossman (1972).....	90
5.2	Os impactos do saneamento básico sobre a saúde dos indivíduos	96
5.2.1	Saneamento Básico e Mortalidade	97
5.2.2	Saneamento Básico e Morbidade	99
5.2.3	Mais uma razão para investir em saneamento básico	103
5.3	Análise exploratória dos dados espaciais (ESDA)	104

5.4	Metodologia.....	110
5.5	Modelos e Hipóteses.....	117
5.6	Resultados.....	122
5.6.1	Modelo 1 – população de todas as idades.	123
5.6.2	Modelo 1 – população de zero a 14 anos.	127
5.6.3	Modelo 2 – população de todas as idades.	129
5.6.4	Modelo 2 – população de zero a 14 anos.	132
5.7	A Censura dos Dados	134
5.7.1	Metodologia Bayesiana	134
5.7.2	Resultados	137
5.8	Recomendação de política pública	145
6	CONCLUSÃO	149
	REFERÊNCIAS	155
	APÊNDICE.....	165

LISTA DE SIGLAS

ABCON – Associação Brasileira das Concessionárias Privadas de Serviços Públicos de Água e Esgoto
AIH – Autorização de Internação Hospitalar
BIRD – Banco Internacional para Reconstrução e Desenvolvimento
CID-10 – Classificação Internacional de Doenças Capítulo 10
CNEFE – Cadastro Nacional de Endereços para Fins Estatísticos
CNES – Cadastro Nacional de Estabelecimentos de Saúde
DNE – Diretório Nacional de Endereços
DRSAI – Doenças Relacionadas ao Saneamento Ambiental Inadequado
DATASUS – Departamento de Informática do Sistema Único de Saúde
FUNASA – Fundação Nacional da Saúde
FINBRA – Finanças do Brasil
IBGE – Instituto Brasileiro de Geografia e Estatística
INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IPEA – Instituto de Pesquisa Econômica Aplicada
MC – Ministério das Cidades
MPO – Ministério do Planejamento, Orçamento e Gestão
MS – Ministério da Saúde
OMS – Organização Mundial da Saúde
PAC – Programa de Aceleração do Crescimento
PBF – Programa Bolsa Família
PISA – Programa Internacional de Avaliação dos Estudantes
PNAD – Pesquisa Nacional de Amostras de Domicílios
PNUD – Programa das Nações Unidas para o Desenvolvimento
PNRS – Política Nacional de Resíduos Sólidos
RDH – Relatório de Desenvolvimento Humano
SEPURB – Secretaria de Política Urbana
SIASUS – Sistema de Informações Ambulatoriais
SIM – Sistema de Informações de Mortalidade
SINASC – Sistema de Informações de Nascidos Vivos
SNIS – Sistema Nacional de Informações sobre Saneamento
STN – Secretaria do Tesouro Nacional
TDI – Taxa Distorção Idade-Série
UNICEF – United Nations Children's Fund
WHO – World Health Organization

LISTA DE QUADROS

Quadro 2.1 – Descrição e tipos das DRSAIs	25
Quadro 3.1 – Variáveis dependentes, variável instrumental e de saneamento	32
Quadro 3.2 – Variáveis independentes.....	33
Quadro 3.3 – Sinais esperados - relação entre Saneamento e Internações por DRSAI	51
Quadro 4.1 – Sinal e significância e sinais das variáveis explicativas.....	77
Quadro 4.2 – Sinal e significância dos determinantes espaciais - frequência escolar	80
Quadro 4.3 – Sinal e significância determinantes espaciais da distorção idade-série.....	83
Quadro 4.4 – Sinal e significância dos determinantes espaciais do abandono escolar	86

LISTA DE TABELAS

Tabela 2.1 – Acesso à água e esgoto da população urbana e rural por regiões mundiais (%)	17
Tabela 2.2 – Acesso à água e esgoto canalizados, esgoto tratado e perdas na distribuição (%), 2013 e 2014.....	18
Tabela 2.3 – Indicadores educacionais para as regiões brasileiras, 2000 e 2010.....	22
Tabela 2.4 – Número de casos de internações por DRSAIs e parcela do total do país.....	25
Tabela 2.5 – Taxa de incidência DRSAIs - grupos de transmissão e faixas de idade.....	27
Tabela 2.6 – Taxa de incidência de DRSAI pelas grandes regiões (2010)	28
Tabela 2.7 – Taxa de incidência de DRSAI e valores gastos por estado (2010).....	29
Tabela 3.1 – Estatísticas Descritivas – dependentes, saneamento e instrumental.....	34
Tabela 3.2 – Estatísticas Descritivas das variáveis independentes (painel não espacial)	35
Tabela 3.3 – Estatísticas Descritivas – variáveis dependentes e outras	36
Tabela 3.4 – Estatísticas Descritivas – variáveis independentes (painel espacial)	36
Tabela 3.5 – Número - municípios, distritos, subdistritos com e sem dados omitidos.	39
Tabela 3.6 – Distribuição dos endereços do CNEFE por espécie nas grandes regiões.....	42
Tabela 3.7 – Número de distritos e subdistritos brasileiros	42
Tabela 3.8 – Regiões postais brasileiras.....	43
Tabela 3.9 – Número de CEPs encontrados pelo CNEFE, Correios e agregações	45
Tabela 3.10 – Estatísticas descritivas, população do número de distritos por estado	46
Tabela 3.11 – Estatísticas descritivas para as variáveis dependentes	47
Tabela 3.12 – Acesso à abastecimento de água, lixo e drenagem de águas pluviais	49
Tabela 3.13 – Acesso aos tipos de coleta de esgoto.....	50
Tabela 3.14 – Estatísticas Descritivas das variáveis de interesse e de controle.....	52
Tabela 4.1 – Empresas e municípios atendidos por empresas públicas e privadas.....	63
Tabela 4.2 – Estatísticas I de Moran para os resíduos e defasagem espacial.....	67
Tabela 4.3 – Resultados do primeiro estágio	70
Tabela 4.4 – Impacto do saneamento em frequência escolar 6-14 anos	71
Tabela 4.5 – Impacto do saneamento na taxa de abandono escolar.....	72
Tabela 4.6 – Impacto do saneamento na taxa de Distorção Idade-Série.....	73
Tabela 4.7 – Impacto do saneamento na TDI (1º e 2º ciclos)	74
Tabela 4.8 – Impacto do saneamento na taxa de Abandono Escolar (1º e 2º ciclos).....	75
Tabela 4.9 – Análise de robustez para o instrumento	76
Tabela 4.10 – Determinantes espaciais para a frequência escolar (2000-2010)	79

Tabela 4.11 – Determinantes espaciais da taxa de distorção idade-série (2000-2010).....	82
Tabela 4.12 – Determinantes espaciais da taxa de abandono escolar (2000-2010)	85
Tabela 4.13 – Comparação entre os modelos estimados por GMM e MV	87
Tabela 5.1 – Coeficiente I de Moran para taxa de incidência de DRSAI	106
Tabela 5.2 – Coeficiente I de Moran para Incidência de DRSAI, Modelo 1	123
Tabela 5.3 – Testes LM e TRMV – pop de todas as idades.....	125
Tabela 5.4 – Efeitos marginais - SDM (W contígua) e SDM (W distância).....	126
Tabela 5.5 – Testes de Multiplicador de Lagrange, Modelo 1	128
Tabela 5.6 – Efeitos marginais, SDM e SDEM	128
Tabela 5.7 – Novo I de Moran	129
Tabela 5.8 – Coeficiente I de Moran para Incidência de DRSAI, Modelo 2	130
Tabela 5.9 – Testes de Multiplicador de Lagrange, Modelo 2.....	130
Tabela 5.10 – Efeitos marginais - SDM (W distância) e SDEM (W contígua)	131
Tabela 5.11 – Testes de Multiplicador de Lagrange, Modelo 2.....	132
Tabela 5.12 – Efeitos marginais - SDEM (W distância) e SDM (W contígua)	133
Tabela 5.13 – Novo I de Moran	134
Tabela 5.14 – Comparação entre modelos: MQO, SDM e SDMT (todas as idades).....	138
Tabela 5.15 – Comparação entre modelos: MQO, SDM e SDMT (de 0 a 14 anos).....	139
Tabela 5.16 – Comparação entre efeitos marginais (SDM e SDMT)	140
Tabela 5.17 – Comparação entre efeitos marginais (SDM e SDMT)	142
Tabela 5.18 – Comparação entre modelos: MQO, SDM e SDMT (de 0 a 14 anos).....	143
Tabela 5.19 – Comparação entre efeitos marginais (SDM e SDMT)	144
Tabela 5.20 – Efeitos marginais do modelo SDMT com matriz W contígua	145
Tabela 5.21 – População, incidência DRSAI, número de casos e custos (2010).....	146
Tabela 5.22 – Análise do novo número de casos e valores gastos com as internações	147

LISTA DE FIGURAS

Figura 2.1 – Mapas de acesso aos serviços de saneamento básico (2000 e 2010).....	19
Figura 2.2 – Diagrama-F	24
Figura 2.3 – Evolução DRSAl e Gastos Federais com saneamento (1995-2010)	27
Figura 3.1 – Distritos paulistas: acesso à rede geral de esgoto e sem esgoto a céu aberto nas condições de entorno (%).....	40
Figura 3.2 – Valores nulo da taxa de incidência de internações por DRSAl.....	48
Figura 4.1 – Tempo de exposição das empresas privadas nos municípios	64
Figura 5.1 – Função de produção de dias saudáveis	94
Figura 5.2 – A eficiência marginal do investimento em saúde (EMH).....	95
Fonte: Grosman (1972a).....	95
Figura 5.3 – Incidência por internações de DRSAl nos distritos brasileiros	104
Figura 5.4 – Distribuição espacial das aglomerações das taxas de incidência de internações por DRSAl para a população de todas as idades (2010).....	108
Figura 5.5 – Valores da taxa de incidência DRSAl por <i>clusters</i>	109

1 INTRODUÇÃO

A importância do saneamento básico e da sua associação com a saúde remonta a tempos mais antigos. É reconhecida e consolidada a literatura que trata dos efeitos positivos do acesso aos serviços de abastecimento de água, tratamento e coleta de esgoto tanto em relação ao impacto sobre mortalidade infantil (Cutler e Miler, 2005; Watson, 2006; Geruso e Spears, 2015) quanto no que diz respeito aos indicadores de morbidez (Bleakley, 2007; Duflo *et al*, 2015; Fewtrell *et al*. 2005).

As precárias condições de saneamento básico propiciam a transmissão de bactérias, vírus e parasitas, que estão presentes nas fezes, urina ou vômito do doente ou portador, causadores de diversas doenças infectocontagiosas. A diarreia é a mais conhecida dentre elas, pelo seus efeitos devastadores sobre a taxa de mortalidade infantil em diversos países do mundo. No entanto, existem diversas outras que também merecem preocupação quanto aos danos causados à população, em especial às crianças: esquistossomose, hepatite A, febre amarela, leishmaniose, malária, febre tifoide, entre outras.

A grande importância em diminuir as taxas de incidência de tais doenças levou a Fundação Nacional da Saúde (Funasa, 2004) a classificá-las em cinco grupos. O primeiro é o de transmissão feco-oral, o segundo é composto por doenças transmitidas por inseto-vetor, o terceiro contém as doenças transmitidas via contato com a água, o quarto diz respeito às doenças relacionadas a higiene e, o último, denomina-se o grupo dos geo-helminthos e teníases. O conjunto de doenças de todos esses grupos foi denominado de Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI) baseado na proposta de classificação das doenças de Cairncross e Feachem (1990; 1993) e Mara e Feachem (1999).

O objetivo dessa tese, a partir do reconhecimento milenar desta associação entre saneamento e saúde, é investigar as consequências educacionais sobre crianças que apresentam estado de saúde debilitado por terem contraído as DRSAI. A hipótese estabelecida é que os sintomas de tais doenças lhes proporcionam um estado de saúde insuficiente para alcançar bons resultados escolares. Segundo Padro *et al*. (2001), as infecções parasitárias estão entre as mais disseminadas desordens que afetam crianças em idade escolar que vivem em áreas pobres de grandes centros urbanos. Os efeitos patológicos diretos destes parasitas são diarreia crônica, má absorção dos alimentos e anemia ferropriva, que é decorrente da privação e deficiência de

ferro dentro do organismo (KUNZ et al., 2008). As consequências dessas desordens tendem a comprometer o comportamento e a função cognitiva das crianças, principalmente quanto à capacidade de atenção e concentração, que afetam o rendimento escolar e dificultam o aprendizado (ARAÚJO et al, 2009 e KUNZ et al., 2008).

Em outras palavras, quando a incidência das DRSAs não leva a população infantil a óbito, quais são os impactos gerados sobre a acumulação de capital humano dos sobreviventes em termos de perda de rendimento escolar?

O objetivo dessa pergunta é ressaltar que, além de destruir capital humano, os efeitos de adversas condições de saneamento podem impactar o perfil de morbidade de uma região no longo prazo, com graves consequências para a aquisição de anos de escolaridade e, dessa forma, comprometimento de ganhos salariais e de produtividade no futuro. Tais efeitos deletérios da falta dessa infraestrutura essencial propagam as sementes do subdesenvolvimento social e educacional, pois não haverá políticas de educação que sejam capazes de alterar tais resultados escolares se não acompanhadas de políticas de infraestrutura.

Dessa forma, as perguntas dessa tese são formuladas em duas etapas. Na primeira etapa, o principal efeito de interesse é investigado, ou seja, o efeito indireto do saneamento sobre educação via saúde. Na segunda etapa, o segundo efeito de interesse que dá suporte ao primeiro se refere ao efeito direto da falta de saneamento sobre a saúde.

A primeira delas investiga questões acerca do efeito de saneamento sobre educação. As condições de infraestrutura dos domicílios, expressas pela variável de acesso a saneamento básico, podem ser mais um dos determinantes dos indicadores educacionais? Em outras palavras, o maior acesso aos serviços de saneamento básico nos domicílios impacta os indicadores educacionais de frequência, distorção idade-série e abandono escolar da população infantil?

Na segunda etapa de investigação, a hipótese sustentada nesse estudo é que o efeito destacado acima se dá pelo estado de saúde debilitado dos indivíduos. Segundo Relatório de Desenvolvimento Humano (2006), infecções parasitárias transmitidas pela água ou pelas más condições de saneamento provocam 443 milhões de faltas escolares por ano no mundo. Miguel e Kremer (2004) também encontraram um efeito associado ao motivo doença para

explicar a evasão escolar. Nesse sentido, a pergunta que se pretende responder é se, de fato, as precárias condições de saneamento básico se traduzem em maior incidência de DRSAL. Além disso, esse impacto pode ser ampliado pelos efeitos de transbordamento espacial relacionados à variável de saúde dos vizinhos? Se condições inadequadas de saneamento afetam a saúde dos indivíduos, essas mesmas condições de saneamento dos vizinhos também pode ser um fator que alastre o efeito sobre a saúde? Em outras palavras, quais são os efeitos espaciais presentes nesse fenômeno em estudo?

A justificativa da importância de tais perguntas pode ser vista sob três dimensões. O maior acesso aos serviços essenciais de saneamento, além de gerar impactos sobre a saúde, meio ambiente, turismo, entre outros, ocasiona uma nova externalidade de longo prazo por impactar indicadores educacionais de performance. Na área de educação, as condições precárias de saneamento onde a escola está localizada e onde os alunos residem importam para o rendimento escolar dos mesmos, além do *background* familiar e do efeito escola. Na área da saúde, as consequências dos precários ou ausentes serviços de saneamento básico são maiores do que a literatura dessa área documenta, pois vão além dos impactos em termos de maiores taxas de mortalidade infantil, ou seja, afetam o perfil de morbidez das crianças sobreviventes gerando deletérios efeitos socioeconômicos de longo prazo para o país.

Apesar do efeito direto do saneamento sobre a saúde ocorrer antes do efeito indireto do saneamento sobre educação via saúde, esta tese investigará tais efeitos na ordem contrária. Em primeiro lugar, o efeito de maior interesse, que é o indireto e, depois, o efeito direto, que dá suporte ao primeiro. A justificativa dessa ordem encontra-se na forma de utilização dos dados das amostras: do mais agregado para o menos agregado. A estratégia de resposta de tais perguntas foi organizada de acordo com a apresentação dos capítulos descritos a seguir.

No capítulo 2 são apresentados alguns indicadores de saneamento básico, educação e saúde para traçar o cenário da realidade socioeconômica predominante no país em tais áreas. A análise dá enfoque às diferenças regionais e ressalta alguns dados de gastos do setor público nas mesmas.

O terceiro capítulo apresenta e descreve os dois tipos de dados, municipais e distritais, que serão utilizados nas análises dos capítulos subsequentes. Ou seja, as estatísticas descritivas de todos eles, as fontes nas quais tais dados foram obtidos e, para o caso dos dados por distritos,

o procedimento pelo qual tais dados foram criados, uma vez que não estão disponíveis em domínio público dados nesse nível de desagregação. Dessa forma, os dados dos municípios brasileiros obtidos para os anos 2000 e 2010 são utilizados na investigação do efeito de saneamento sobre educação. Os dados criados para cada distrito do território nacional são utilizados para verificação da hipótese associada ao efeito na saúde que a falta de saneamento é capaz de gerar.

No quarto capítulo são analisados empiricamente os impactos do saneamento sobre a educação, utilizando dados municipais dos anos de 2000 e 2010. Além da análise econométrica convencional para dados em painel com presença de endogeneidade, adotou-se também metodologia para o controle das características espaciais presentes nesse tipo de dados. A partir do referencial teórico dos determinantes educacionais, buscou-se identificar se as características domiciliares onde os indivíduos residem são importantes para explicar variáveis de frequência escolar, distorção idade-série e abandono escolar.

No entanto, para validar esse efeito é necessário que a associação entre saneamento e saúde seja também comprovada de forma empírica, objetivo ao qual se dedica o quinto capítulo. Este se inicia com o desenvolvimento do arcabouço teórico do modelo de Grosman (1972), que foi utilizado para definir as variáveis utilizadas nas estimações subsequentes, bem como identificar os canais de transmissão entre as variáveis. Em seguida, realizou-se revisão bibliográfica acerca das relações entre saneamento, mortalidade e morbidade. Nas seções posteriores, foram realizadas a Análise Exploratória de Dados Espaciais, estabelecidas as hipóteses acerca dos possíveis efeitos espaciais existentes, que podem ser mais realisticamente justificadas dada a utilização dos dados distritais, a escolha entre os diversos modelos, a descrição sobre as metodologias utilizadas e a discussão dos resultados. O modelo escolhido, de acordo com a primeira metodologia, foi o Modelo Durbin Espacial (SDM) com matriz contígua de ponderação espacial.

Uma característica importante da amostra utilizada nesse capítulo deve-se à quantidade de zeros na variável dependente. Tais zeros referem-se aos distritos saudáveis que não tiveram ocorrências de internações por DRSAI no ano de 2010. O controle da censura dos dados foi realizado pela estimação do Modelo SDM Tobit Espacial de acordo com a abordagem Bayesiana.

Nos dois últimos capítulos, questões específicas foram levantadas acerca da sensibilidade de tais efeitos para a população mais jovem. Quanto mais novas as crianças que, em geral, são os organismos mais suscetíveis às DRSAs, maior o impacto na saúde e maior o efeito sobre piores indicadores educacionais?

Por fim, na última seção do capítulo 5 foi realizada uma recomendação de política pública. Segundo esta, investir recursos em saneamento gera economia de recursos no setor da saúde que poderiam ser investidos em saneamento para acelerar o cumprimento das metas de universalização.

O capítulo 6 conclui o trabalho. Alguns mapas, quadros e tabelas de cada capítulo, que não foram explicitados no decorrer do texto, são mostrados no Apêndice no final da Tese. Além destes, estão presentes no mesmo outros tópicos que detalham ou elucidam certas questões mais específicas.

2 INDICADORES DE SANEAMENTO BÁSICO, EDUCAÇÃO E SAÚDE.

O objetivo deste capítulo é apresentar alguns indicadores de saneamento básico, educação e saúde, que serão utilizados nesse estudo, para traçar o cenário da realidade socioeconômica predominante no país. A análise dá enfoque às diferenças regionais e ressalta alguns dados de gastos do setor público em tais áreas.

2.1 Saneamento Básico

Apesar do progresso mundial no acesso aos serviços de abastecimento de água e esgotamento sanitário, há ainda 768 milhões de pessoas sem possibilidade de consumo de água segura e 2,5 bilhões que não possuem serviços adequados de coleta e tratamento de esgoto (WHO – UNICEF, 2014). Na América Latina, 117 milhões de pessoas não possuem acesso aos serviços de esgotamento sanitário, 22 milhões de pessoas registram a existência de esgoto a céu aberto e apenas 20% das águas servidas são tratadas (BIRD, 2014).

As informações da Tabela 2.1, referentes ao sistema de esgoto, representam a porcentagem da população urbana e rural que tem acesso às instalações sanitárias por algum tipo de sistema de coleta de esgoto, seja por meio de rede geral, fossa séptica ou outro tipo de fossa. O indicador relativo à água refere-se ao acesso a água encanada dentro do domicílio, poço artesiano, torneiras públicas ou nascentes, incluindo drenagem de águas pluviais.

Tabela 2.1 – Acesso à água e esgoto da população urbana e rural por regiões mundiais (%)

Indicadores de Acesso	Esgoto				Água			
	pop urb		pop rur		pop urb		pop rur	
Regiões	1990	2015	1990	2015	1990	2015	1990	2015
Leste Asiático e Pacífico	68,0	85,0	40,0	64,0	91,4	95,9	75,1	86,8
Europa e Ásia Central	95,0	97,0	74,0	89,0	98,6	99,3	92,6	97,1
América Latina e Caribe	78,0	86,0	34,0	62,0	93,8	96,5	71,7	88,5
Brasil	79,0	88,0	31,0	52,0	95,8	100,0	67,7	87,0
Oriente Médio e Norte da África	86,0	92,0	53,0	86,0	94,4	94,9	79,1	89,9
América do Norte	96,0	97,0	91,0	93,0	99,9	99,7	96,5	98,6
Sul da Ásia	51,0	65,0	9,0	35,0	93,2	93,1	70,6	87,7
África Subsariana	39,0	40,0	91,0	93,0	80,8	89,1	45,6	64,0

Fonte: WHO/UNICEF e BIRD (2014).

O Brasil apresenta média superior em relação aos países da América Latina e Caribe para os indicadores de acesso da população urbana aos serviços de água e esgoto. Em relação a essa população, a situação mais crítica encontra-se no acesso ao esgotamento sanitário. No *ranking*

de todos os países que possuem informações para essa variável, que totalizam 192, o Brasil fica na 101ª posição, atrás de países como Angola, Chile, Argentina e Cuba.

A insuficiência do setor de saneamento é histórica na realidade brasileira, com os indicadores relativos aos serviços de água e esgoto ainda aquém do necessário para se alcançar as metas de universalização. Na Tabela 2.2 os indicadores de acesso aos serviços de saneamento básico para o país como um todo, e segmentado para as cinco grandes regiões, apontam que os valores nacionais encobrem diferenças regionais significativas. Nessa tabela, dados recentes da Pesquisa Nacional de Amostras de Domicílios (PNAD) de 2013 mostram a porcentagem dos domicílios atendidos por rede geral de abastecimento de água e a porcentagem daqueles atendidos por rede coletora de esgoto. Os dados do Sistema Nacional de Informações sobre Saneamento (SNIS) de 2014 indicam a porcentagem de esgoto tratado referente à água consumida e a porcentagem de perdas na distribuição de água¹.

Tabela 2.2 – Acesso à água e esgoto canalizados, esgoto tratado e perdas na distribuição (%), 2013 e 2014.

Regiões	Rede Geral de Água	Rede Coletora de Esgoto*	Tratamento de Esgoto**	Água - Perdas na distribuição**
Norte	57,35	19,75	14,40	47,90
Nordeste	80,07	41,40	31,40	46,90
Sudeste	91,5	88,16	45,70	32,60
Sul	87,82	59,95	36,90	33,40
Centro-Oeste	84,57	47,60	46,40	34,20
Brasil	84,62	62,95	40,80	36,70

Fonte: Elaboração própria a partir de dados da: *PNAD (2013) e **SNIS (2014).

Observa-se que somente as regiões Sul e Sudeste apresentam médias superiores à nacional em relação aos domicílios que possuem distribuição de água por canalização interna. Em relação à provisão da rede coletora de esgoto, a situação é ainda mais preocupante: pouco mais da metade da população possui acesso a esse serviço essencial. Além disso, para esse indicador, a região Norte apresenta valor muito baixo e nem metade dos domicílios da região Nordeste e Centro-Oeste é atendida por rede coletora de esgoto. No tocante à porcentagem de tratamento de esgoto, o melhor desempenho é das regiões Centro-Oeste e Sudeste. Deve-se ressaltar que a região Sul apresenta elevado índice de esgotamento sanitário por fossa séptica como será apresentado mais adiante, o que justifica o baixo nível de tratamento de esgoto pela forma

¹Segundo SNIS (2014), a partir da parcela referente ao volume total de água (VT), que é igual ao volume de água tratada importado mais o volume de água produzido, menos, o volume de serviço (valor da soma dos volumes anuais de água usados para atividades operacionais e especiais), o índice de perdas na distribuição é calculado em porcentagem e corresponde ao $[(VT - \text{Volume de água consumido})/VT]$. O índice de esgoto tratado é dado por $[\text{Volume de esgotos tratado}/(\text{Volume de água consumido} - \text{Volume de água tratada exportado})]$.

convencional. No entanto, no que se refere ao índice de perdas na distribuição de água, as regiões Sudeste, Sul e Centro-Oeste voltam a ganhar destaque em detrimento das outras duas.

Em outras palavras, em torno de 35 milhões de pessoas não possuem acesso adequado a abastecimento de água. Cerca de 100 milhões de brasileiros são desprovidos de rede coletora de esgotos. Somente as capitais brasileiras lançaram 1,2 bilhão de m³ de esgotos não tratados na natureza em 2013 e a cada 100 litros de água coletados e tratados, em média, apenas 67 litros são consumidos (TRATA BRASIL, 2015).

Os mapas abaixo, expressos na Figura 2.1, mostram a evolução espacial dos dados dos Censos Demográficos do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2000 e 2010 no tocante ao percentual de pessoas que residem em domicílios que possuem banheiro e água canalizada por rede geral de distribuição.

Figura 2.1 – Mapas de acesso aos serviços de saneamento básico (2000 e 2010).

Os mapas apontam a concentração espacial entre as regiões e estão em consonância com as conclusões obtidas da Tabela anterior. Pode-se notar a evolução dos indicadores de saneamento nesse período pelo espreadimento da mancha azul. No entanto, a desigualdade estrutural entre as regiões se manteve, uma vez que os menores indicadores de acesso ainda estão nas regiões Nordeste e Norte, com avanço considerável da região Centro-Oeste. Segundo Toneto e Saiani (2006), o déficit de acesso no setor de saneamento é concentrado nas regiões mais pobres, na zona rural, nos municípios de menor porte, naqueles que apresentam menor taxa de urbanização e menor renda per capita e nos domicílios de baixa

renda. Ou seja, os resultados apresentados para as regiões vão ao encontro das conhecidas características econômicas e sociais das mesmas. Ademais, o acesso aos serviços de saneamento básico tornou-se uma questão de saúde pública no Brasil. A preocupação com o setor deve ser priorizada, pois o mesmo pode gerar uma série de externalidades negativas para a sociedade, como, por exemplo, a degradação do meio ambiente, a proliferação de inúmeras doenças infecciosas e parasitárias. Além disso, pode permitir que regiões deficientes dessa infraestrutura fiquem impedidas de praticar atividades relativas à indústria, comércio e turismo.

O Relatório de Desenvolvimento Humano (RDH) de 2006 e o estudo do Trata Brasil (2008) adicionaram um novo elemento a essa lista: os indicadores educacionais. Segundo o último, a universalização do acesso à coleta de esgoto e à água tratada reduziria em 6,8% o atraso escolar dos alunos que vivem em regiões sem saneamento. Em tais regiões não é incomum que crianças se afastem das escolas ou tenham o aprendizado comprometido devido a viroses, verminoses e infecções causadas pela falta de água tratada e por um sistema de esgotamento sanitário inadequado.

No presente estudo, a relação entre saneamento e educação será investigada à luz do estado de saúde da população de todas as idades e das crianças de zero a catorze anos. Portanto, deve-se ter em consideração alguns indicadores acerca da educação e saúde.

2.2 Educação

O Brasil tem avançado na cobertura do sistema educacional nos últimos anos e vem experimentando consistente aumento dos níveis de escolaridade ao longo do tempo. No entanto, quando é comparado a países com nível de desenvolvimento inferior ou semelhante, verifica-se que tais avanços têm-se dado de forma lenta, desigual e concentrada. Os dados do último PISA (Programa Internacional de Avaliação dos Estudantes) mostraram, por exemplo, que o país está em 38º lugar entre 44 países avaliados quanto à capacidade de alunos de 15 anos resolverem problemas de matemática aplicados à vida real². Os indicadores educacionais utilizados nesse estudo dizem respeito à taxa de frequência escolar, à taxa de distorção idade-

² No restante das disciplinas, de 65 países analisados, o Brasil ocupou o 58º em matemática, o 55º em leitura e o 59º lugar em ciências (PISA, 2012).

série e ao abandono escolar da Educação Primária ou Ensino Fundamental, que deve ser cursado pelas crianças e adolescentes brasileiros de 6 a 14 anos.

A escolha pela taxa de frequência escolar se deu em razão do fato de que o primeiro impacto da saúde debilitada das crianças, por terem contraído alguma doença relacionada ao saneamento básico inadequado, é o afastamento da mesma da escola. Os dados relativos às taxas de rendimento que fornecem informações sobre a eficiência do sistema escolar, são de três tipos: aprovação, reprovação e abandono. Este último indicador foi selecionado pois, de alguma forma, irá refletir se o afastamento da criança da escola implicou em abandono definitivo, consequência grave e que dificulta a retomada dos estudos posteriormente.

Por fim, a taxa de distorção idade-série (TDI) foi escolhida pois reflete o comprometimento do aprendizado do aluno quando este não está na série adequada à sua idade. É importante destacar que a queda da taxa de distorção idade-série não implica necessariamente em aproveitamento escolar, principalmente quando se tem em vigência no país a progressão continuada ou o regime de ciclos³. Essa forma de organização pressupõe que a classificação em qualquer série ou etapa, exceto na primeira série do ensino fundamental, pode ser feita independentemente de escolarização anterior.

A Tabela 2.3 apresenta os dados que mostram a evolução entre os anos 2000 e 2010 das cinco grandes regiões para as três variáveis de educação ressaltadas anteriormente. A taxa de frequência escolar refere-se à porcentagem da população de 6 a 14 anos frequentando qualquer nível ou série. A taxa de distorção idade-série (TDI) diz respeito à porcentagem de discentes cuja idade está acima da adequada para a série que frequenta nas duas fases do Ensino Fundamental. Por fim, a taxa de abandono escolar é definida como a porcentagem de alunos que deixaram de frequentar a escola, e tiveram a matrícula cancelada, nas duas fases do Ensino Fundamental.

³Impactos negativos sobre rendimento escolar (abandono) e proficiência foram confirmados por Menezes-Filho *et al* (2008), mas não por Neves & Pazello (2012) e Carvalho & Firpo (2014) quando avaliaram o desempenho escolar

Tabela 2.3 – Indicadores educacionais para as regiões brasileiras, 2000 e 2010.

Regiões	Frequência Escolar		Variação (%)	Distorção Idade-Série		Variação (%)	Abandono Escolar		Variação (%)
Anos	2000	2010		2000	2010		2000	2010	
Norte	82,94	92,61	11,66	59,51	34,65	-41,78	18,36	4,88	-73,44
Nordeste	90,22	96,07	6,49	63,63	35,93	-43,53	15,61	5,70	-63,48
Sudeste	93,98	97,48	3,72	30,56	18,16	-40,59	8,97	1,81	-79,87
Sul	92,84	96,35	3,78	24,12	17,96	-25,54	5,57	1,39	-75,00
Centro-Oeste	92,96	96,97	4,31	44,94	23,36	-48,03	16,22	2,72	-83,20
Brasil	92,85	97,22	4,71	38,39	23,17	-39,65	11,08	2,84	-74,37

Fonte: Elaboração própria a partir de Censo Demográfico (IBGE) e Censo Escolar (INEP)

Os dados mostram as mesmas diferenças regionais explicitadas nos indicadores de acesso aos serviços de saneamento básico. De forma geral, as regiões Sul, Sudeste e Centro-Oeste apresentam valores superiores para os indicadores educacionais em relação às regiões Norte e Nordeste. No que se refere à variação percentual entre os anos de 2000 e 2010 para a taxa de distorção idade-série e abandono escolar, a região Centro-Oeste apresentou maior queda.

É importante destacar que tais variáveis representam apenas o fluxo escolar das crianças e adolescentes do Ensino Fundamental, ou seja, não serão analisadas questões relativas à qualidade do ensino. Dessa forma, pretende-se investigar o impacto inicial da falta de saneamento sobre o distanciamento da criança da escola, o atraso que pode ser gerado por um estado de saúde comprometido e a possibilidade de tais crianças abandonarem os estudos, o que pode gerar graves consequências no longo prazo em termos de inserção no mercado de trabalho. No entanto, ainda que sejam analisadas variáveis de fluxo escolar, é razoável supor que a retenção do conteúdo pelo aluno esteja comprometida, devido a condições inadequadas de saneamento em seu domicílio. Esse é um fator não considerado na literatura do setor.

2.3 Saúde

A análise do impacto do acesso aos serviços de saneamento básico sobre a educação, por se dar via estado de saúde do indivíduo que contraiu alguma doença relacionada à falta ou inadequada provisão desse bem essencial, requer a identificação dos agentes patogênicos que causam tais doenças, bem como os efeitos que geram nos organismos que os contraem. Nesta seção trata-se desse assunto por meio da análise dos indicadores de saúde.

2.3.1 As Doenças de Saneamento Ambiental Inadequado (DRSAI) e os canais de transmissão

O reconhecido Diagrama F, de Wagner e Lanoix (1958), mostra os canais de transmissão dos agentes patogênicos que entram no meio ambiente através das fezes e a associação destes com os sistemas de saneamento básico. Os agentes patogênicos presentes nas fezes do indivíduo doente podem ser transmitidos pela água (flúidos), por hábitos higiênicos inadequados (tais como não lavar as mãos), por insetos vetores (moscas) ou no momento da irrigação agrícola com água contaminada⁴. Tais agentes podem chegar a um novo hospedeiro de forma direta (eles mesmos serão ingeridos) ou pela comida.

De acordo com esse modelo, as barreiras primárias são sintetizadas nos sistemas de coleta adequada de esgoto, pois irão evitar a propagação das fezes contaminadas, isolando a matéria fecal da oferta de água, bem como de moscas e outros insetos transmissores. Outra forma de evitar a entrada de tais agentes na localidade se dá por meio de práticas higiênicas, tais como lavar as mãos após o uso do banheiro ou fossa. As barreiras secundárias são aquelas que impedem os agentes patogênicos que já entraram no meio ambiente de avançar em direção ao novo hospedeiro. Destaca-se entre elas o sistema de tratamento de água, pois consegue removê-los, a proteção das fontes de água e práticas higiênicas referentes ao preparo e armazenamento dos alimentos. A oferta de água via abastecimento por rede geral causa um efeito direto em melhores práticas de higiene pessoal, doméstica e comunitária. Além disso, a melhor qualidade da água ingerida e utilizada para consumo interrompe a transmissão de novos agentes patogênicos que geram as doenças de saneamento (CVJETANOVIC, 1986; HELLER, 1997).

O Diagrama F, representado abaixo, identifica quais as intervenções e tecnologias são plausíveis de serem aplicadas para evitar doenças de saneamento que estão associadas aos indicadores de morbidade de uma região.

⁴ A denominação Diagrama-F vem da primeira letra de cada termo em inglês: os agentes patogênicos adentram o ambiente pelas fezes (*feces*) e são transmitidos por meio dos flúidos (*fluids*), dedos (*fingers*), moscas (*flies*) ou irrigação (*fields crops*). Essa transmissão pode ocorrer de forma direta ou pela comida (*food*).

Figura 2.2 – Diagrama-F
 Fonte: Wagner e Lanoix (1958)
 Observação: Figura A3 de Duflo *et al* (2015)

O período de 1981-1990, denominado pela Organização das Nações Unidas como a Década Internacional do Abastecimento de Água e do Esgotamento Sanitário, marcou o início de uma compreensão mais aprofundada da relação entre condições sanitárias e saúde (HUTTLY, 1990). Cvjetanovic (1986) considerou os modelos que relacionavam as ações de saneamento com as enfermidades diarreicas limitados, diante do caráter amplo da definição de saúde formulada pela Organização Mundial da Saúde (OMS).

De fato, o Diagrama F enfatizava os canais de transmissão das doenças do grupo feco-oral, com destaque para a diarreia. Cairncross (1989) enunciou a importância de se chegar a um consenso científico sobre as relações entre o abastecimento de água e o esgotamento sanitário, por um lado, e doenças infecciosas, por outro. Em 2001, a Fundação Nacional da Saúde (Funasa) elaborou uma seleção das Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI) baseada na proposta de classificação das doenças de Cairncross e Feachem (1990; 1993) e Mara e Feachem (1999). Tais doenças são características de localidades que possuem abastecimento de água deficiente, esgotamento sanitário

inadequado, rios e lençóis freáticos contaminados por resíduos sólidos e precárias condições habitacionais. O quadro 2.2 apresenta a classificação elaborada pela Funasa das DRSAIs.

Tipos de Doenças (CID-10)	Doenças Capítulo I	Descrição e Subcategorias
Doenças de transmissão feco-oral	Diarréias	Cólera; Outras infecções por Salmonella; Outras infecções intestinais bacterianas; Amebíase; Outras doenças intestinais por protozoários; Infecções intestinais virais, outras e as não especificadas; Diarréia e gastroenterite de origem infecciosa presumível
	Febres Entéricas	Febre tifóide e paratifóide
	Hepatite A	Hepatite aguda A
Doenças transmitidas por inseto vetor	Dengue	Dengue [dengue clássico]
	Febres Amarela	Febre hemorrágica devida ao vírus do dengue Febre amarela
	Leishmaniose	Leishmaniose
	Filariose Linfática	Filariose
	Malária	Malária por Plasmodium falciparum; Malária por Plasmodium vivax; Malária por Plasmodium malariae; Outras formas de malária confirmadas por exames parasitológicos; Malária não especificada
Doenças transmitidas via contato com a água	Doença de Chagas	Doença de Chagas
	Esquistossomose	Esquistossomose [bilharziose] [Schistosomíase]
	Leptospirose	Leptospirose
Doenças relacionadas com a higiene	Doenças dos Olhos	Tracoma; Conjuntivite
	Doenças da Pele	Dermatofitose; Outras micoses superficiais
Geo-helmintos e Teníases	Helmintíases	Infestação por Taenia; Cisticercose; Outras infestações por cestóides; Ancilostomíase; Ascaridíase; Estrongiloidíase; Tricuríase; Oxiuríase; Outras helmintíases intestinais, não classificadas em outra parte; Parasitose intestinal não especificada
	Teníase	Equinococose

Quadro 2.1 – Descrição e tipos das DRSAIs

Fonte: Elaboração própria a partir de Funasa (2004).

A sistematização de tais doenças ampliou o escopo dos impactos do saneamento ao demonstrar os diversos alcances, em termos epidemiológicos, sobre a saúde do indivíduo.

Na Tabela 2.4 são apresentados os dez tipos de DRSAI com os maiores números de casos de internações que correspondem a 95% do total de casos para o Brasil (599.499 casos), no ano de 2010.

Tabela 2.4 – Número de casos de internações por DRSAIs e parcela do total do país

DRSAI	Num casos	%
Diarréia e gastroenterite de origem infecciosa presumível	232.770	38,83
Outras infecções intestinais bacterianas	122.017	20,35
Infecções intestinais virais, outras e as não especificadas	104.172	17,38
Dengue [dengue clássico]	88.554	14,77
Outras infecções por Salmonella	7.895	1,32
Febre hemorrágica devida ao vírus do dengue	6.323	1,05
Cólera	3.265	0,54
Malária por Plasmodium vivax	2.972	0,50
Leishmaniose	2.431	0,41
Leptospirose	2.191	0,37
Total Brasil (todas)	599.499	95,52

As diarreias, infecções virais, bacterianas e a dengue são os grupos de doenças que apresentam os maiores números de internações e que, conjuntamente, correspondem a 91% do total das doenças apresentadas no Quadro 2.1.

Dado que esses dois grupos são que mais se destacam, na Figura 2.3 é possível observar a evolução da taxa de incidência de internações por DRSAl para ambos ao longo do tempo, por 100.000 habitantes. Essa taxa de incidência é calculada pela razão entre o número de casos de internações por DRSAl que ocorreram em uma determinada localidade no ano de 2010 e a população residente nesse período. O resultado dessa razão é multiplicado por 100.000. Além disso, a título de comparação, foram adicionados no mesmo gráfico os dados do Instituto de Pesquisa Econômica Aplicada (IPEA)⁵, referentes aos valores de gastos federais com saneamento, em 10 milhões de reais de 2010, valores deflacionados pelo deflator implícito do PIB do IBGE. O gráfico indica que, em alguns períodos, tais como 1995-98, a queda da curva da taxa de incidência foi acompanhada pelo aumento da curva de gastos federais com saneamento. No entanto, a partir do valor da taxa de incidência de aproximadamente 300 casos por 100.000 habitantes, os aumentos dos gastos federais não foram capazes de diminuir de forma significativa a incidência de internações por DRSAl. Isso ocorre porque muitos dos aportes dos recursos liberados pelo PAC (Programa de Aceleração do Crescimento)⁶ não foram efetivamente realizados, em decorrência da participação tímida da iniciativa privada no setor e devido às dificuldades de captação de tais recursos pelas empresas públicas.

⁵ Tais dados foram obtidos da Diretoria de Estudos e Políticas Sociais (DISOC), estimativas anuais a partir dos dados do Sistema Integrado de Administração Financeira e Dados Orçamentário (SIAFI/SIDOR) e das Contas Nacionais do IBGE.

⁶ No período de 2003-06, a média de desembolso federal em saneamento foi de R\$1,57 bilhão ao ano. Esse aporte anual passou para média de R\$7,1 bilhões para o período 2007-2015 (Ministério das Cidades, 2016).

Figura 2.3 – Evolução DRSAI e Gastos Federais com saneamento (1995-2010)

Fonte: Elaboração própria a partir de IBGE, MS e IPEA (1995-2010)

A taxa de incidência de internações por DRSAIs no Brasil em 2010 foi de 316 casos por 100.000 habitantes. A Tabela 2.5 apresenta a distribuição dessa taxa pelos diferentes grupos especificados que constam no quadro 2.1, para populações de diferentes faixas etárias. Devido à sua importância, a diarreia e a dengue foram destacadas. Na classificação original, estas fazem parte do grupo de transmissão feco-oral e do grupo de doenças transmitidas por inseto vetor, respectivamente. Os dados da Tabela mostram que o grupo das doenças de transmissão feco-oral é o que mais se destaca, acometendo de forma mais intensiva a população de 0 a 6 anos de idade. As internações por diarreia também são as que mais atingem os organismos mais suscetíveis, que são as crianças. O segundo grupo de destaque são as doenças transmitidas por inseto vetor. A dengue atinge de forma mais expressiva a população na faixa de idade de 6 a 14 anos, ou seja, crianças em idade escolar.

Tabela 2.5 – Taxa de incidência DRSAIs - grupos de transmissão e faixas de idade

Incidência DRSAI Brasil	Todas idades	Inc 0a14	Inc 0a6	Inc 6a14
De transmissão feco-oral	139,36	316,71	545,85	185,46
Transmitidas por inseto vetor	5,28	7,16	11,15	4,87
Transmitidas através do contato com a água	1,48	0,70	0,34	0,90
Relacionadas com a higiene	0,52	1,28	2,32	0,69
Geo-helmintos e teníases	0,44	0,65	0,96	0,47
Diarreia	119,73	280,69	494,90	158,00
Dengue	49,66	55,45	37,76	65,58
Todos os grupos	316,47	662,63	1.093,28	415,95

Nessa tese, os diferentes tipos de doenças serão tratados com um grupo único. A justificativa dessa escolha está baseada no fato de que todas essas doenças estão associadas ao abastecimento de água deficiente, ao esgotamento sanitário inadequado, à contaminação por resíduos sólidos ou às condições precárias de moradia. Caso houvesse segmentação entre os grupos, outras especificidades epidemiológicas deveriam ser consideradas, o que está fora do escopo dessa tese. Dos três grupos de faixas de idades especificados na Tabela anterior, os que serão considerados para fins de estimação são a população de todas as idades e a população de 0 a 14 anos. As taxas de incidência das internações por DRSAl entre as macrorregiões brasileiras, para a população dessas duas faixas etárias, por 100 mil habitantes, são apresentadas na Tabela 2.6:

Tabela 2.6 – Taxa de incidência de DRSAl pelas grandes regiões (2010)

Regiões	Inc DRSAl todas	Inc DRSAl 0a14
Norte	688,84	1208,03
Nordeste	521,77	984,94
Sudeste	120,38	283,16
Sul	207,19	451,94
Centro-Oeste	434,25	797,28
Total	316,47	662,63

Mais uma vez, como era esperado, e também de acordo com os resultados para os indicadores de saneamento básico, as regiões Sul e Sudeste apresentam taxas de incidência por internações por DRSAl abaixo da taxa nacional. No entanto, as regiões Norte, Nordeste e Centro-Oeste apresentam valores acima da mesma. A mesma conclusão vale para as duas faixas etárias. Na Tabela 2.7 são apresentadas as taxas de incidência por estado. A coluna referente aos valores diz respeito ao valor total gasto com as internações por DRSAl no ano de 2010, em R\$ milhões. As colunas 5 e 6 referem-se ao *ranking* da maior taxa de incidência e do maior gasto total com as internações⁷.

⁷ Após classificar em ordem decrescente as colunas 5 e 6 e ordená-las de 1 a 27, foram destacados os dez maiores valores em cada uma delas. O estado de São Paulo, por exemplo, apesar de ser o quarto colocado em termos de gastos realizados com as internações por DRSAl é o estado que apresenta a menor taxa de incidência de internações por DRSAl.

Tabela 2.7 – Taxa de incidência de DRSAl e valores gastos por estado (2010)

Sigla	uf	Inc	DRSAI	Valores	Rank. Inc	Rank. Val.	Pop (%)
RO		864,45	4,227		3	17	0,82
AC		598,38	1,345		5	23	0,39
AM		208,07	2,699		19	21	1,83
RR		597,30	0,965		6	24	0,24
PA		966,06	24,714		2	2	3,98
AP		193,19	0,442		20	26	0,35
TO		498,39	2,305		11	22	0,73
MA		849,96	18,636		4	3	3,45
PI		967,66	9,932		1	9	1,63
CE		359,76	10,465		15	7	4,41
RN		438,29	4,593		14	16	1,66
PB		541,30	6,974		8	11	1,98
PE		289,60	9,471		16	10	4,62
AL		490,48	5,571		12	14	1,64
SE		121,40	0,900		24	25	1,09
BA		591,33	28,085		7	1	7,36
MG		177,88	13,851		21	5	10,28
ES		230,19	2,891		18	20	1,84
RJ		108,77	6,539		26	13	8,39
SP		88,21	14,084		27	4	21,63
PR		284,46	10,454		17	8	5,47
SC		145,71	3,590		23	19	3,26
RS		167,42	6,778		22	12	5,59
MS		451,18	4,002		13	18	1,28
MT		504,89	5,426		10	15	1,59
GO		529,84	10,534		9	6	3,15
DF		111,97	0,126		25	27	1,35

Os estados que possuem maiores gastos com as internações proporcionais às maiores taxas de incidência são Pará, Piauí, Bahia e Goiás.

3 DADOS

Nos capítulos subsequentes, a investigação dos impactos do saneamento sobre a educação irá se valer de dois tipos de dados: municipais e distritais. Assim sendo, neste capítulo, apresentam-se os dados municipais utilizados na investigação do efeito de saneamento sobre educação e os dados distritais utilizados para verificação da hipótese associada ao motivo saúde, pois a análise do primeiro efeito se vale da hipótese de que os indicadores educacionais da população de 6 a 14 anos são afetados porque o estado de saúde desse público-alvo fica comprometido ao contrair doenças de saneamento.

3.1 Dados municipais

Os dados municipais utilizados nesse estudo são provenientes de diversas fontes de dados e compreendem os anos de 2000 e 2010. No modelo a ser estimado, que será apresentado no próximo capítulo, as variáveis dependentes são provenientes do Censo Demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE)⁸ e do Censo Escolar do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

A variável explicativa de interesse, referente a saneamento, provém também do Censo Demográfico (IBGE). A variável instrumental, utilizada para identificar o impacto da falta de saneamento nos indicadores educacionais, foi construída a partir de informações obtidas pela ABCON (Associação Brasileira das Concessionárias Privadas de Serviços Públicos de Água e Esgoto)⁹. O SNIS (Sistema Nacional de Informações sobre Saneamento) foi consultado para confirmação das informações da ABCON, bem como para identificar outros dados dos municípios atendidos por cada empresa. O Quadro 3.1 apresenta tais variáveis, a descrição de cada uma delas, a unidade e a fonte dos dados.

⁸ Dados obtidos via Atlas do Desenvolvimento Humano do Programa das Nações Unidas para o Desenvolvimento (PNUD).

⁹ A ABCON congrega empresas privadas prestadoras de serviços públicos de água e saneamento básico bem como diversas informações acerca do tipo de contrato de concessão, data em que o mesmo passou a vigorar, município em que a empresa privada oferece os serviços de saneamento básico, população atendida, investimentos previstos e realizados entre outros. A partir de tais informações, foi possível determinar a partir de que ano a empresa privada de saneamento começou a atuar nos municípios em questão de forma a construir a variável instrumental.

Variáveis	Descrição	Unidade	Fonte
Frequência Escolar (t_freq_6a14)	População de 6 a 14 anos frequentando a escola em qualquer nível ou série	%	Censo Dem.
Abandono escolar (abandono_fund)	Alunos que deixaram de frequentar a escola, e tiveram a matrícula cancelada, nas duas fases do Ensino Fundamental (1ª a 8ª série ou 1º ao 9º ano)	%	Censo Escolar
Abandono escolar fase 1 (abandono_fund_1)	Alunos que deixaram de frequentar a escola, e tiveram a matrícula cancelada, na primeira fase do Ensino Fundamental (1ª a 4ª série ou 1º ao 5º ano)	%	Censo Escolar
Abandono escolar fase 2 (abandono_fund_2)	Alunos que deixaram de frequentar a escola, e tiveram a matrícula cancelada, na segunda fase do Ensino Fundamental (5ª a 8ª série ou 6º ao 9º ano)	%	Censo Escolar
Distorção Idade-Série (tdi_fund)	Discentes cuja idade está acima da adequada para a série que frequenta nas duas fases do Ensino Fundamental (1ª a 8ª série ou 1º ao 9º ano)	%	Censo Escolar
Distorção Idade-Série fase 1 (tdi_fund_1)	Discentes cuja idade está acima da adequada para a série que frequenta na primeira fase do Ensino Fundamental (1ª a 4ª série ou 1º ao 5º ano)	%	Censo Escolar
Distorção Idade-Série fase 2 (tdi_fund_2)	Discentes cuja idade está acima da adequada para a série que frequenta na segunda fase do Ensino Fundamental (5ª a 8ª série ou 6º ao 9º ano)	%	Censo Escolar
Variável de Saneamento			
Banheiro e água (t_banagua)	População que vive em domicílios com banheiro e água encanada	%	Censo Dem.
Variável Instrumental			
Privatização saneamento (priv5anos)	Indicadora se a empresa de saneamento atuante no município é privada há pelo menos cinco anos	0 ou 1	Abcon/SNIS

Quadro 3.1 – Variáveis dependentes, variável instrumental e de saneamento

As variáveis explicativas restantes utilizadas nas estimações foram obtidas das mesmas fontes de dados do Quadro 3.1, além de algumas novas tais como o IPEA (Instituto de Pesquisa Econômica Aplicada), Finanças do Brasil (FINBRA), da Secretaria do Tesouro (STN), e Departamento de Informática do Sistema Único de Saúde (DataSUS). Tais variáveis foram organizadas em três vetores, denominados características socioeconômicas (1), características da escola e professores (2) e saúde e educação (3). De forma semelhante ao Quadro 3.1, encontram-se, no Quadro 3.2, a apresentação de tais variáveis, a descrição de cada uma delas, a unidade e também a fonte dos dados de onde são provenientes.

Variáveis	Descrição	Unidade	Fonte
Características Socioeconômicas			
Escolaridade pop. adulta (t_analf25m)	População de 25 anos ou mais analfabeta	%	Censo Dem.
Etnia (pop_negra_parda)	População negra ou parda	%	Censo Dem.
Renda per capita (renda)	Renda por habitante média	R\$/hab	Censo Dem.
Índice de gini (gini)	Grau de desigualdade existente na distribuição de indivíduos segundo a renda domiciliar per capita	0-1	Pnud
Distribuição etária população (pop15a29 e pop55mais)	População de 15 a 29 anos e população de 55 anos ou mais	%	Censo Dem.
População total (poptot)	População total (habitantes)	Hab	Censo Dem.
Bolsa Família (bolsafam)	Número de famílias beneficiárias do Programa Bolsa Família em relação ao número total de famílias	%	IPEA
Taxa de urbanização (urb)	População urbana em relação à população total	%	Censo Dem.
Características da escola e professores			
Alunos_turma (alunos_turma_fund)	Número Médio de Alunos por Turma no ensino fundamental	Al/turma	Censo Escolar
Docente_Superior (doc_sup_fund)	Percentual de Docentes com Curso Superior no ensino fundamental	%	Censo Escolar
Média_horas_aula (horas_aula_fund)	Número Médio de Horas-Aula Diária no ensino fundamental	Hs/aula	Censo Escolar
Infra Tecnológica (iit)	Índice de infraestrutura escolar tecnológica	0-1	Censo Escolar
Infra Básica (iiba)	Índice de infraestrutura escolar básica	0-1	Censo Escolar
Escolas públicas (p_pub_esc)	Porcentagem de escolas públicas	%	Censo Escolar
Diretoria (p_diretoria)	Porcentagem de escolas com existência de diretoria	%	Censo Escolar
Saúde e educação			
Desp_saude (desp_saude)	Despesa com saúde por habitante	R\$/hab	Finbra
Desp_ed_cult (desp_educ_cult_hab)	Despesa com educação e cultura por habitante	R\$/hab	Finbra
Imunização (doses_calculos_cv)	Doses de vacinas aplicadas	Doses	DataSUS

Quadro 3.2 – Variáveis independentes

O primeiro vetor de observáveis engloba variáveis demográficas, de renda, urbanização e escolaridade da população adulta. O segundo refere-se às características dos professores e da escola. O último apresenta informações de despesa municipal com saúde, educação e cultura, além da variável relativa às doses de vacinas aplicadas na população de crianças (imunização). Essas variáveis, mais a variável de saneamento (denominada variável de interesse), compõe o vetor de variáveis explicativas das estimações com dados municipais.

3.2 Estatísticas descritivas dos dados municipais

A partir da compatibilização dessas diversas fontes de dados por município e para os anos 2000 e 2010 foram calculadas as estatísticas descritivas que serão apresentadas nesta seção. A tabela 3.3 exibe informações sobre as principais variáveis de educação utilizadas nesse estudo, para os anos 2000 e 2010. O primeiro e o segundo ciclo correspondem, respectivamente, da primeira à quarta série ou do primeiro ao quinto ano, e da quinta à oitava série, ou do quinto ao nono ano. Além disso, foram apresentadas as mesmas estatísticas descritivas para as variáveis de acesso aos serviços de saneamento básico e da variável que será utilizada como instrumento. Todas as variáveis apresentaram evolução positiva ao longo década, com destaque para a redução da taxa de abandono escolar, particularmente no primeiro ciclo. Contudo, o avanço na universalização do acesso não pode ter eliminado as

deficiências do sistema educacional, que se refletem na baixa qualidade do ensino e na deficitária aprendizagem dos alunos.

Tabela 3.1 – Estatísticas Descritivas – dependentes, saneamento e instrumental

Variáveis	Obs		Média		Desvio Padrão		Mínimo		Máximo	
Períodos	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
Variáveis Dependentes										
t_freq6a14	4369	4369	92,85	97,22	6,12	2,54	43,68	51,77	100	100
abandono_fund	4314	4369	11,08	2,84	7,14	2,93	0	0	53,40	34,10
abandono_fund_1	4313	4369	8,35	1,31	7,93	2,16	0	0	59,90	25,40
abandono_fund_2	4312	4369	14,10	4,67	8,12	4,51	0	0	100	45,10
tdi_fund	4314	4275	38,39	23,17	18,23	11,63	6,00	1,30	87,00	66,10
tdi_fund_1	4314	4267	29,92	16,93	19,70	11,23	0,90	0,40	86,70	64,40
tdi_fund_2	4313	4275	48,86	30,35	20,07	13,70	8,70	1,40	96,60	79,40
Variável de Saneamento										
t_banagua	4369	4369	71,62	86,73	27,18	18,01	1,26	6,68	100	100
Variável instrumental										
priv5anos	4369	4369	0,002	0,04	0,05	0,20	0	0	1	1

A Tabela 3.2 apresenta as estatísticas descritivas para as variáveis independentes dos vetores de características socioeconômicas, da escola e professores e da saúde e educação. Os dados da tabela abaixo mostram que significativas mudanças demográficas ocorreram no período: queda da população cuja faixa etária demanda atendimento da educação básica e superior (de 0 a 14 anos e de 15 a 29 anos) e, por outro lado, envelhecimento populacional, caracterizado pelo aumento da quantidade de pessoas com 55 anos ou mais. Além disso, o país tornou-se mais urbano, o valor da renda real aumentou e a mesma distribui-se de forma menos desigual (menor coeficiente de Gini). As iniciativas de transferência direta de renda às camadas da população abaixo da linha de pobreza, o Programa Bolsa Família (PBF), podem ter contribuído para esse cenário.

Além disso, houve aumento real de recursos financeiros destinados à educação e à saúde, evidência que pode ser constatada pelo aumento real do valor das despesas nessas áreas. Além disso, a redução da população na faixa escolar significou mais recursos para o atendimento de crianças, jovens e adultos. Registrou-se aumento da porcentagem de docentes com nível superior, do índice de infraestrutura escolar tecnológica e básica, da porcentagem de escolas que têm diretoria e ligeiro aumento do número médio de horas-aula diárias no ensino fundamental. Por outro lado, a razão alunos por turma diminuiu, juntamente com a porcentagem de escolas públicas. Como consequência dessa evolução, além de outros fatores

específicos relacionados às políticas públicas específicas, os indicadores educacionais apresentaram evolução positiva.

Tabela 3.2 – Estatísticas Descritivas das variáveis independentes (painel não espacial)

Variáveis	Obs		Média		Desvio Padrão		Mínimo		Máximo	
Períodos	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
Características socioeconômicas										
t_analf25m	4369	4369	22,13	16,67	13,15	10,72	0,95	1,10	69,38	54,57
pop_negra_parda	4312	4369	41,20	46,93	25,10	24,23	0	0,25	94,81	93,43
rendapc	4369	4369	391,69	556,40	184,25	234,37	89,78	141,88	1759,76	2043,74
gini	4369	4369	0,54	0,48	0,07	0,07	0,30	0,28	0,87	0,80
pop55mais	4315	4369	13,09	17,01	3,27	4,30	2,70	4,30	29,60	38,30
pop15a29	4315	4369	26,85	25,54	2,27	2,61	16,20	14,90	34,90	43,80
pop0a14	4315	4369	30,49	24,36	5,15	4,76	17,30	7,30	52,80	45,60
bolsafam	4369	4369	0	19,51	0	13,77	0	0,08	0	100
poptot	4315	4369	32695	36077	204357	220636	795	805	10400000	11300000
t_urb	4315	4369	61,46	66,45	23,49	22,09	0	4,18	100	100
Características da escola e professores										
alunos_turma_fund	4314	4276	31,73	21,77	9,06	3,84	12,50	6,50	276,30	37,70
doc_sup_fund	3978	4188	42,49	75,97	27,98	21,94	0	0	100	100
horas_aula_fund	4314	4275	4,31	4,43	0,32	0,50	3,60	3,70	6,80	9,20
iiet	4315	4369	0,09	0,21	0,10	0,13	0	0,0001	0,67	0,73
iiba	4315	4369	0,47	0,50	0,19	0,17	0,03	0,02	1	0,98
p_pub_esc	4315	4369	91,54	89,13	11,11	12,50	32,86	25,70	100	100
p_diretoria	4315	4369	32,79	53,38	26,56	26,68	0	0	100	100
Saúde e educação										
desp_educ_cult_hab	4203	4323	187,20	234,44	329,12	95,48	0	0	17754,50	1383,97
desp_saude_hab	4155	4196	97,3463	475,7884	136,4799	256,49	0	0	6096,55	4549,625
doses_calculos_cv	4316	4369	12719,54	13512,81	75567,96	77356,83	275	259	3967753	3986617

Os dados municipais descritos acima foram utilizados com análise econométrica convencional: estimação em dois estágios de um modelo de efeitos fixos para dados em painel com variável instrumental¹⁰. A abordagem espacial utilizou os mesmos dados, porém perdeu algumas informações, pelo fato de que o painel espacial utilizado para estimação do mesmo modelo deveria ser balanceado, sem *missings* e sem “ilhas”¹¹. A Tabela 3.3 mostra a evolução das variáveis dependentes utilizadas nas estimações econométricas espaciais. Os resultados apontam para a mesma tendência de evolução dos dados das seções anteriores. Os valores para a variável de frequência escolar de 6 a 14 anos, de acesso aos serviços de saneamento básico adequado e o instrumento relativo às empresas que privatizaram os serviços de saneamento apresentaram aumento entre os anos de 2000 e 2010. Os valores relativos à taxa de distorção idade-série e ao abandono escolar mostraram queda no mesmo período.

¹⁰ Os detalhes metodológicos desse tipo de estimação serão descritos no próximo capítulo.

¹¹ As “ilhas” são definidas como localidades que não tem nenhum vizinho.

Tabela 3.3 – Estatísticas Descritivas – variáveis dependentes e outras

Variáveis	Obs	Média		Desvio Padrão		Mínimo		Máximo	
		2000	2010	2000	2010	2000	2010	2000	2010
t_freq_6a14	3612	92,18	96,66	5,96	2,94	43,68	51,77	100	100
abandono_fund	3612	10,65	2,81	7,03	2,94	0	0	53,40	34,10
tdi_fund	3612	37,87	23,01	18,34	11,60	6,00	1,30	82,70	66,10
t_banagua	3612	72,07	86,83	27,15	17,75	1,56	6,68	100	100
priv5anos	3612	0,002	0,04	0,05	0,21	0	0	1	1

Os valores das estatísticas descritivas das variáveis explicativas utilizadas no painel espacial, demonstrados na Tabela 3.4, também estão de acordo com os resultados anteriores.

Tabela 3.4 – Estatísticas Descritivas – variáveis independentes (painel espacial)

Variáveis	Obs	Média		Desvio Padrão		Mínimo		Máximo	
		2000	2010	2000	2010	2000	2010	2000	2010
t_analf25m	3612	22,19	16,74	13,58	11,10	0,95	1,10	69,38	54,57
pop_negra_parda	3612	3.469	4.869	20.854	28.377	0	2	811.090	1.113.998
renda per capita	3612	396,81	559,55	190,83	240,90	98,43	141,88	1759,76	2043,74
gini	3612	0,54	0,48	0,069	0,066	0,30	0,28	0,87	0,80
pop55mais	3612	13,20	17,05	3,31	4,24	2,70	4,8	29,60	34,00
pop15a29	3612	26,82	25,60	2,30	2,58	16,20	14,9	34,90	43,80
bolsafam	3612	0	19,15	0	13,31	0	0,08	0	100
poptot	3612	35.952	40.002	222.300	241.247	795	805	10.400.000	11.300.000
t_urb	3612	61,75	67,14	23,83	22,21	0	4,18	100	100
alunos_turma_fund	3612	31,73	21,89	8,94	3,82	12,50	7,80	276,30	37,70
doc_sup_fund	3612	44,88	76,38	27,61	21,68	0	0	100	100
horas_aula_fund	3612	4,32	4,44	0,33	0,50	3,60	3,70	6,80	9,20
iiet	3612	0,10	0,21	0,10	0,13	0	0,0001	0,67	0,73
iiba	3612	0,48	0,50	0,20	0,18	0,025	0,022	1,00	0,98
p_pub_esc	3612	91,40	88,87	11,46	12,92	32,86	25,70	100	100
p_diretoria	3612	33,28	52,95	26,83	26,60	0	0	100	100
desp_educ_cult_hab	3612	189,11	234,72	353,05	93,35	1,64	0	17.755	1.384
desp_saude_hab	3612	88,51	181,43	51,38	79,49	10,26	42,96	613,34	849,06
doses_calculos_cv	3612	13.911	14.906	82.153	84.522	275	259	3.967.753	3.986.617

De acordo com os dados apresentados acima, novamente tem-se um retrato de um país com mais pessoas idosas, menor população adulta em termos percentuais, mais urbano, com maior renda per capita distribuída de forma menos concentrada ao longo do período e com mais beneficiários do programa bolsa família. No que se refere ao vetor de características relacionadas à educação, houve aumento do número de docentes com ensino superior, da quantidade média de horas por aula, maior porcentagem de escolas que têm diretoria e maiores índices de infraestrutura escolar básica e tecnológica. Por outro lado, o número de escolas públicas diminuiu assim como a razão do número de alunos por turma. Em relação às variáveis de despesa todas registraram maiores recursos destinados à saúde, educação e

cultura. Além disso, a cobertura das taxas de imunização também aumentou ao longo desse período.

3.3 Dados distritais

Essa seção descreve as fontes dos dados utilizados para análise espacial dos determinantes das DRSAs. Os dados são provenientes do Censo Demográfico do Universo por setor censitário (IBGE, 2010), Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE, 2011), Diretório Nacional de Endereços (DNE, 2015) e Departamento de Informática do Sistema Único de Saúde (DATASUS, 2010).

Os dados por setor censitário compreendem características dos domicílios particulares e das pessoas que foram investigadas para a totalidade da população e são denominados, por convenção, resultados do universo (IBGE, 2011). O setor censitário é a menor unidade territorial, formada por área contínua, integralmente contida em área urbana ou rural, com dimensão adequada à realização da coleta de dados por um pesquisador que vai a campo por ocasião do censo. O setor constitui um conjunto de quadras, no caso de área urbana, ou uma área do município, no caso de uma área não urbanizada. Ao contrário dos dados da amostra, os dados do universo não são desagregáveis até o nível do indivíduo. Para a preservação da privacidade do indivíduo e do domicílio que fornece a informação, cada variável é agregada como uma média do setor. Por exemplo, não é possível conhecer a renda de um determinado domicílio inserido em um setor censitário específico, apenas a média da renda de todos os domicílios que compõem o setor em questão.

Nesse estudo utilizaram-se dados por setor censitário, agregados por distritos ou subdistritos, com o objetivo de se alcançar um nível de desagregação menor que o municipal. O menor nível de desagregação é relevante para o problema em análise desta tese, pois o deslocamento dos agentes patogênicos pode ocorrer mais facilmente entre distritos, mas menos provavelmente entre municípios. Os distritos são unidades administrativas dos municípios e criados por legislação municipal, devendo observar a continuidade territorial e os requisitos previstos em lei complementar estadual (IBGE, 2011). Todo município tem ao menos um distrito, que é chamado de distrito-sede. Os distritos, por sua vez, podem ser subdivididos em unidades administrativas denominadas subdistritos, regiões administrativas, zonas ou outra denominação específica. Os subdistritos, normalmente estabelecidos nas grandes cidades são

criados através de leis ordinárias das Câmaras Municipais e sancionadas pelo prefeito (IBGE, 2011).

A base de dados do Censo Demográfico do Universo possui 10.899 distritos, a menor unidade geográfica depois do setor censitário. Com o objetivo de atender à legislação estatística brasileira, que assegura a proteção dos dados dos informantes das pesquisas, o IBGE optou pela restrição dos dados dos setores censitários que registrassem menos de cinco domicílios particulares permanentes. A Tabela 3.5 indica o quantitativo de municípios, distritos, subdistritos e setores de cada macrorregião e de cada Unidade da Federação, ressaltando também o número de setores e subdistritos que tiveram seus dados omitidos por terem poucos domicílios particulares permanentes.

A análise dos dados da tabela abaixo mostra que, após ajustes para dados omitidos, que correspondem a aproximadamente 2% do total de informações, o número total de subdistritos e distritos é 10.860. Quando o estado ou o município não possui nenhum subdistrito, o número de subdistritos é igual ao número de distritos. Exemplo desse caso ocorre no estado de São Paulo, onde não há subdistritos. A mesma situação ocorre para estados ou cidades que não possuem distritos. O estado que exemplifica essa situação é Roraima, pois não apresenta nenhum município subdividido em distrito e, conseqüentemente, nenhum subdistrito.

Tabela 3.5 – Número - municípios, distritos, subdistritos com e sem dados omitidos.

UF	Municípios	Distritos	Subdistritos	Setores censitários	Setores censitários dados omitidos	Subdistritos dados omitidos	Subdistritos informações válidas
RO	52	99	104	2346	65	0	104
AC	22	23	23	874	20	0	23
AM	62	85	94	5641	205	0	94
RR	15	15	15	824	28	0	15
PA	143	250	250	8769	78	0	250
AP	16	33	33	810	12	0	33
TO	139	153	153	2101	29	0	153
Norte	449	658	672	21365	437	0	672
MA	217	243	243	8802	107	0	243
PI	224	224	229	5251	104	0	229
CE	184	839	845	13276	372	1	844
RN	167	183	186	4289	38	0	186
PB	223	288	288	5548	33	0	288
PE	185	391	404	12379	102	0	404
AL	102	115	122	3724	34	0	122
SE	75	83	83	3297	27	0	83
BA	417	844	869	23782	405	1	868
Nordeste	1794	3210	3269	80348	1222	2	3267
MG	853	1633	1767	32564	537	7	1760
ES	78	275	299	6380	101	0	299
RJ	92	293	379	27769	384	2	377
SP	645	1036	1036	66096	1379	1	1035
Sudeste	1668	3237	3481	132809	2401	10	3471
PR	399	757	820	17465	423	3	817
SC	293	461	461	11882	529	1	460
RS	496	1230	1369	22332	822	22	1347
Sul	1188	2448	2650	51679	1774	26	2624
MS	78	162	174	4207	55	1	173
MT	141	249	255	5929	234	0	255
GO	246	317	379	9434	136	0	379
DF	1	1	19	4349	43	0	19
Centro Oeste	466	729	827	23919	468	1	826
Total	5565	10282	10899	310120	6302	39	10860

É importante fazer um esclarecimento acerca da nomenclatura que será utilizada no decorrer dessa pesquisa. Apesar das diferenças existentes e anteriormente ressaltadas entre distritos e subdistritos, o fato de que o número destes (617) é menor do que o número daqueles (10.282) e pelo fato de que ambos se referem à desagregação municipal, será utilizada a convenção de denominar a menor unidade geográfica utilizada nessa tese apenas de distrito.

A justificativa dessa desagregação, além do menor raio de deslocamento dos agentes patogênicos citado anteriormente, encontra-se na percepção de que as condições de saneamento básico de um município tendem a ser muito heterogêneas entre essas diferentes unidades geográficas, em especial nas grandes cidades. Por exemplo, São Paulo, o maior

município brasileiro em população, possui 96 distritos, que apresentam distintas taxas de acesso a saneamento básico. A média dos domicílios paulistanos cuja coleta de esgoto é feita via rede geral ou fossa séptica, e que não apresentam em suas condições de entorno esgoto a céu aberto, é de 88%. No entanto, a média elevada encobre uma concentração espacial desse indicador no município, como apresentado no mapa da Figura 3.1.

Figura 3.1 – Distritos paulistas: acesso à rede geral de esgoto e sem esgoto a céu aberto nas condições de entorno (%)

Assim, as variáveis explicativas do banco de dados são provenientes do Censo Demográfico do Universo por setor censitário, agregadas por distritos para melhor caracterizar a distribuição e os determinantes espaciais da variável dependente, relativa à taxa de incidência de DRSAIs. No entanto, a maior parte dos dados disponíveis para pesquisa empírica aplicada em economia, em especial as análises em economia regional, apresentam desagregação máxima em nível municipal. Logo, a análise proposta nessa tese para todo Brasil em nível distrital exige a construção de uma nova variável dependente, que não está disponível nas bases de dados comumente utilizadas pelos pesquisadores. Esta variável é proveniente do Departamento de Informática do Sistema Único de Saúde (DATASUS), cujos dados permitem a mensuração do estado da saúde da população por meio de diversos sistemas, tais como o Sistema de Informações Ambulatoriais (SIASUS), Sistema de Informações de Mortalidade (SIM), Sistema de Informações de Nascidos Vivos (SINASC), Cadastro Nacional de Estabelecimentos de Saúde (CNES), entre outros.

Neste estudo, serão utilizados os registros de Autorização de Internação Hospitalar (AIH) extraídos do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIHSUS). Esses dados são informados mensalmente por todos os estabelecimentos públicos, conveniados e contratados, que realizam internações financiadas pelo SUS. Um dos objetivos deste processamento é possibilitar o repasse do governo federal às Secretarias de Saúde e ao gestor local para que possam fazer os pagamentos aos estabelecimentos de saúde. O acesso aos microdados de cada AIH Reduzida (RD-AIH)¹² faz parte de um contexto de descentralização do SIHSUS pelo Ministério da Saúde (MS), que flexibilizou a gerência e autonomia do gestor local (estadual e municipal), além de facilitar o conhecimento e o direcionamento adequado das ações de prevenção e promoção da saúde.

Com o auxílio do programa TabWin, que facilita o trabalho de tratamento dos dados e tabulação dos mesmos¹³, foi possível construir uma base de dados com todas as internações ocorridas no ano de 2010. Cada uma delas acompanha informações referentes ao paciente (idade, sexo, etnia), o valor gasto com os procedimentos, o diagnóstico principal e secundário pelo qual a internação foi realizada, o Código de Endereçamento Postal (CEP) do local da residência de cada paciente, entre outros.

Entre os diversos tipos de internações disponíveis, foram selecionadas as internações cujo diagnóstico principal fosse típico das DRSAIs, de acordo com classificação da FUNASA (2004) pelo CID-10, capítulo 10 da Classificação Internacional de Doenças (OMS, 1997). A compatibilização realizada consta dos Quadros 1A e 2A do Apêndice deste trabalho. Para a construção da nova variável dependente em nível distrital foi necessário estabelecer uma lista de CEPs do local de residência de cada paciente internado por DRSAI para cada distrito, com o objetivo de realizar um “merge” e identificar em quais distritos os casos de DRSAI ocorreram. As faixas de CEP de cada distrito foram construídas utilizando-se duas bases de dados: o CNEFE e o DNE. O Cadastro Nacional de Endereços para fins Estatísticos (CNEFE) é uma base de dados disponibilizada também pelo IBGE, que contém aproximadamente 80 milhões de endereços urbanos e rurais distribuídos pelos 310 mil setores censitários. Os endereços disponíveis podem ser de domicílio particular, coletivo, estabelecimento agropecuário, estabelecimento de ensino, estabelecimento de saúde, estabelecimento de outras

¹² Obtidos em <http://www2.datasus.gov.br/DATASUS/index.php?area=0901&item=1&acao=25>

¹³ Algumas tabulações específicas de tais dados podem ser geradas pelo aplicativo TABNET, disponível na Internet em <http://www2.datasus.gov.br/DATASUS/index.php?area=02>

finalidades e edificação em construção. Tais endereços são identificados pelo logradouro, número, complemento, localidade e CEP. Na Tabela 3.6 abaixo apresenta-se o número de endereços por cada tipo de estabelecimento em cada região geográfica brasileira.

Tabela 3.6 – Distribuição dos endereços do CNEFE por espécie nas grandes regiões

Espécie endereço	Norte	Nordeste	Sudeste	Sul	Centro-oeste	Total
Domicílio Particular	4.608.946	17.861.904	29.537.906	10.409.455	5.039.307	67.457.518
Domicílio Coletivo	11.672	24.482	45.540	16.287	12.659	110.640
Estabelecimento Agropecuário	278.548	831.320	626.116	683.037	292.483	2.711.504
Estabelecimento de Ensino	26.203	99.412	67.851	27.713	12.707	233.886
Estabelecimento de Saúde	9.642	37.800	75.911	28.066	11.673	163.092
Estabelecimento de outras finalidades	599.374	2.448.199	3.490.352	1.210.935	621.681	8.370.541
Edificação em construção	196.168	813.925	1.047.210	269.335	176.768	2.503.406
Total	5.730.553	22.117.042	34.890.886	12.644.828	6.167.278	81.550.587

Um dos objetivos dessa base de dados é apoiar a realização de pesquisas domiciliares pelo IBGE, uma vez que a mesma facilitará aos recenseadores a localização dos domicílios a serem entrevistados. Os microdados do CNEFE estão disponíveis em 10.903 arquivos, que correspondem, cada um deles, a um subdistrito ou distrito. A Tabela 3.7 expõe a distribuição dos distritos e subdistritos entre as cinco regiões brasileiras.

Tabela 3.7 – Número de distritos e subdistritos brasileiros

Regiões	Norte	Nordeste	Sudeste	Sul	Centro-oeste	Total
Distrito	663	3222	3392	2567	764	10.608
Subdistrito	9	47	91	85	63	295
Total	672	3269	3483	2652	827	10.903

Os números de distritos e subdistritos dessa base de dados são similares àqueles apresentados pelo Censo Demográfico do Universo (2010), possuindo inclusive os mesmos códigos. Uma vez que a base do CNEFE será utilizada para a construção da faixa de CEPs de cada distrito, as informações de distritos que serão utilizadas para esse objetivo serão aquelas que constam no Censo Demográfico do Universo (2010). Logo, serão desconsideradas da análise dois do Nordeste, 12 do Sudeste, 28 do Sul e apenas um no Centro-Oeste, totalizando 43 unidades geográficas que não foram consultadas. De posse de tais informações, é possível identificar os CEPs existentes dentro de cada um dos 10.860 distritos, que são as unidades de análise para este estudo.

No entanto, os CEPs do CNEFE não cobrem todo o território brasileiro. Há uma grande quantidade de CEPs iguais para distritos diferentes e outros com códigos errados e inexistentes. Isso ocorre, pois, os mesmos foram obtidos a partir de declaração dos próprios

moradores. A verificação da validade de tais informações por parte do IBGE foi realizada apenas para os municípios pequenos de CEP único. De forma a complementar e identificar os CEPs das regiões faltantes, utilizou-se a base de dados dos Correios, o Diretório Nacional de Endereços (DNE). O DNE é um banco de dados que contém mais de 900 mil CEPs de todo o Brasil, constituído de elementos de endereçamento (descrição de logradouros, bairros, municípios, vilas, povoados) e Códigos de Endereçamento Postal (CEP). O Brasil foi dividido pelos Correios em dez regiões postais, segundo o desenvolvimento socioeconômico e fatores de crescimento demográfico de cada unidade da federação. A Tabela 3.8 apresenta essa divisão e o número de municípios por região.

Tabela 3.8 – Regiões postais brasileiras

Regiões	Estados	Total de municípios
Região 0	Grande SP	39
Região 1	SP Interior	606
Região 2	RJ; ES	170
Região 3	MG	853
Região 4	BA; SE	492
Região 5	AL; PE; PB; RN	677
Região 6	CE; PI; MA; PA; AP; AM; RR; AC	883
Região 7	DF; GO; TO; MT; RO; MS	657
Região 8	PR; SC	692
Região 9	RS	496
Total	27 UF	5565

Fonte: DNE (Correios, 2011)

Os Correios definem localidade de três formas distintas: município, distrito ou povoado. Cada uma delas é codificada por logradouro ou por CEP único. Em geral, o primeiro caso ocorre quando se trata de uma cidade grande, que necessita de códigos específicos para ruas de grande extensão e com elevada densidade populacional. Quando o município possui um único número de CEP, o código é válido para todo o território daquela localidade. O número de cidades nessa situação totaliza 5.194, na maioria com menos de 50.000 habitantes. A média da população desse grupo de cidades é de 15 mil habitantes. Por outro lado, há apenas 371 municípios que possuem codificação por logradouro, a grande maioria deles (366) com mais de 50.000 habitantes. A média da população dos municípios desse grupo é de 290 mil habitantes. Para os distritos também ocorre essa organização: há 4.196 distritos com CEP único e 203 distritos codificados por logradouro. Em relação aos povoados, existem 768 que possuem codificação única e 24 que apresentam codificação por logradouro.

O primeiro passo para a busca dos CEPs dos 10.860 distritos foi a separação de todos aqueles que possuíam distrito único, ou seja, aqueles em que o município todo era o próprio distrito sede. Essa separação foi realizada pois a essas unidades (3.465) foram atribuídas diretamente o CEP único do município. Os CEPs dos distritos restantes (7.395) deveriam ser encontrados pelo CNEFE ou pela base dos Correios. No entanto, a compatibilização da base do IBGE com a do DNE não é direta. Os distritos podem divergir de nome e de extensão territorial, em especial para aqueles de pequeno porte. Além disso, a definição de determinadas localidades pelo IBGE tem a função de cumprir o objetivo da pesquisa domiciliar em razão do Censo Demográfico. Ainda que se pudesse utilizar dos mapas disponibilizados por esse órgão de pesquisa, houve distritos não encontrados em nenhuma das duas bases de dados. O número de tais casos pelas macrorregiões brasileiras é evidenciado na Tabela 3.9 da página seguinte. Na segunda coluna é descrito o número de distritos não encontrados em cada estado. Estes, ao serem agregados, foram transformados em novos distritos, que tiveram uma denominação genérica, cuja quantidade é expressa na terceira coluna. A quarta e quinta colunas referem-se à quantidade de CEPs encontrada pela base do CNEFE e dos Correios. A sexta coluna refere-se aos casos de distritos que apresentam CEP único. A penúltima coluna, que é o somatório da segunda, quarta, quinta e sexta colunas, apresenta o número de distritos antes da agregação, cuja soma totaliza os 10.860 distritos com informações válidas, como evidenciado na Tabela 3.7. A última coluna, que corresponde ao somatório da terceira, quarta, quinta e sexta colunas, diz respeito à quantidade final de unidades distritais que serão utilizadas nessa tese (10.225), após os ajustes decorrentes das agregações mencionadas.

Em síntese, 9,8% dos distritos não foram encontrados e, como solução, foram agregados em novos distritos genéricos. Ou seja, 1.065 distritos transformaram-se em 635 novos distritos, que se constituem em diferentes agregações daqueles (dois a dois, três a três e assim por diante). Em relação àqueles que foram encontrados, 10,1% dos distritos foram identificados pela base do CNEFE, 31,9% tiveram seus CEPs identificados por ambas as bases de dados, uma vez que apresentam CEPs únicos para todo seu território, e 48,1% tiveram seus CEPs encontrados pela base dos Correios.

Tabela 3.9 – Número de CEPs encontrados pelo CNEFE, Correios e agregações

Cod_uf	Distritos agregados	Novos distritos	CNEFE	Correios	Distritos únicos	Distritos antes agregação	Distritos pós agregação
RO	31	20	9	29	35	104	84
AC	0	0	0	2	21	23	23
AM	0	0	11	35	48	94	94
RR	0	0	0	0	15	15	15
PA	27	17	15	118	90	250	233
AP	4	2	1	19	9	33	31
TO	9	5	2	14	128	153	148
Norte	71	44	38	217	346	672	628
MA	0	0	1	48	194	243	243
PI	0	0	6	0	223	229	229
CE	104	67	28	691	21	844	777
RN	2	1	4	28	152	186	185
PB	40	21	5	71	172	288	267
PE	36	22	40	250	78	404	382
AL	2	1	9	22	89	122	121
SE	0	0	0	13	70	83	83
BA	118	66	43	512	195	868	802
Nordeste	302	178	136	1635	1194	3267	3089
MG	273	162	149	872	466	1760	1598
ES	34	19	59	195	11	299	280
RJ	38	26	136	179	24	377	351
SP	34	16	208	339	454	1035	1019
Sudeste	379	223	552	1585	955	3471	3248
PR	44	24	92	470	211	817	793
SC	37	20	28	197	198	460	440
RS	177	117	142	790	238	1347	1230
Sul	258	161	262	1457	647	2624	2463
MS	9	5	11	118	35	173	168
MT	34	18	12	122	87	255	237
GO	12	6	73	93	201	379	373
DF	0	0	19	0	0	19	19
Centro_Oeste	55	29	115	333	323	826	797
Total	1065	635	1103	5227	3465	10860	10225

A Tabela 3.10 apresenta algumas estatísticas descritivas de dois grupos de distritos. O primeiro deles é formado por distritos que coincidem com o município, ou seja, o município tem um único distrito, que é o distrito-sede. O segundo grupo diz respeito aos distritos que pertencem a um município que é subdividido por mais de um distrito. Para cada um desses grupos em cada estado foram calculados o número de distritos, a média da população de tais distritos, a população mínima e a máxima de cada grupo de distritos. Esses dois grupos compõem as 10.225 unidades de análise dessa tese.

Tabela 3.10 – Estatísticas descritivas, população do número de distritos por estado

Cod_uf	Distrito Único				Mais de um distrito			
	Num	Média	Mínimo	Máximo	Num	Média	Mínimo	Máximo
RO	35	16.043	2.315	90.353	49	20.362	450	112.408
AC	21	34.112	4.691	336.038	2	8.605	2.009	15.200
AM	48	24.113	7.326	85.141	46	50.469	629	487.919
RR	15	30.032	6.750	284.313	143	28.000	618	342.742
PA	90	39.683	3.431	471.980	22	25.357	438	376.127
AP	9	12.341	3.793	39.942	20	19.601	538	222.006
TO	128	7.745	1.037	150.484	49	20.393	1.146	150.626
MA	194	28.730	4.020	1.014.837	6	135.436	46.637	178.631
PI	223	10.332	1.253	145.705	756	10.719	207	326.397
CE	21	14.301	4.629	46.033	33	28.473	389	303.543
RN	152	14.650	1.618	259.815	95	15.490	280	359.992
PB	172	13.333	1.256	723.515	304	22.241	56	382.650
PE	78	26.000	2.630	377.779	32	41.106	1.311	238.680
AL	89	20.249	2.866	214.006	13	9.969	903	40.823
SE	70	27.692	2.275	571.149	607	16.276	73	468.360
BA	195	21.169	2.612	204.667	1.132	13.058	42	591.334
MG	466	10.270	815	152.435	269	12.537	37	223.085
ES	11	11.862	4.516	23.843	327	44.269	91	572.599
RJ	24	62.254	5.269	487.562	565	47.206	104	970.065
SP	454	32.029	805	666.740	582	13.566	52	314.915
PR	211	11.972	1.409	212.967	242	15.621	245	494.004
SC	198	12.350	1.465	183.373	992	6.611	17	229.712
RS	238	17.201	1.216	1.409.351	133	14.802	207	184.838
MS	35	13.581	2.928	46.424	-	-	-	-
MT	87	11.247	1.096	113.099	150	13.655	91	182.621
GO	201	13.051	1.020	159.378	172	19.645	48	329.262
DF	-	-	-	-	19	135.238	15.924	402.721
Total	3465	18.452	805	1.409.351	6760	18.678	17	970.065

Cabe ressaltar que as médias da população de cada grupo de distritos de cada estado são semelhantes e a média da população entre os dois grupos no total também¹⁴. Os municípios maiores foram desagregados em unidades geográficas menores e, por outro lado, os municípios menores, ou com apenas um único distrito, foram preservados. Alguns valores discrepantes devem-se às cidades que são distrito-único, apesar de possuírem mais de 500 mil habitantes: Porto Alegre (1.409.351), São Luís (1.014.837), João Pessoa (723.515), Osasco (666.740), Sorocaba (586.625) e Aracaju (571.149).

Com esse procedimento, após obtenção dos CEPs para os 10.225 distritos, obteve-se, a partir do “merge” mencionado anteriormente, o número de casos ocorridos para duas populações de

¹⁴ No Brasil, a média da população de todos os distritos (3.465) que coincidem com o município (18.452 habitantes) é bastante semelhante à média da população dos distritos (6.760) que pertencem à municípios que são subdivididos em mais de um distrito (18.678 habitantes).

interesse, a de todas as idades e a de 0 a 14 anos, em nível distrital. A partir da população residente em cada distrito, obtida do Censo Demográfico do Universo (2010), foram calculadas as taxas de incidência de tais acometimentos para essas duas populações, por 1.000 habitantes.

$$(3.1) \quad INC_DRSAI_{x,i} = \left(\frac{\text{Número de casos internação } DRSAI_{x,i}}{\text{População Residente}_{x,i}} \right) * 1.000$$

Em que,

$i = 1, \dots, 10.225$ e $x = \text{todas idades}; 0a14_anos$.

Na próxima seção, serão exploradas as estatísticas descritivas das variáveis explicadas e explicativas, em nível distrital, que serão utilizadas nas estimações econométricas espaciais. Além disso, serão ressaltados os sinais esperados para cada uma delas, justificados pelo modelo teórico.

3.4 Estatísticas Descritivas dos dados distritais

Outra convenção utilizada para o cálculo da taxa de incidência por DRSAI é considera-la por mil habitantes. Essa escolha ocorre porque se utilizará a unidade geográfica distrital, que apresenta menores médias populacionais. Além disso, mais uma particularidade dos dados, que se refere à variável dependente, diz respeito à quantidade de zeros existente na amostra. As considerações e o tratamento parcial dessa característica serão realizados em seção à parte. No entanto, essas informações são apresentadas nas estatísticas descritivas das duas variáveis dependentes para as 10.225 unidades distritais.

Tabela 3.11 – Estatísticas descritivas para as variáveis dependentes

Var. Dependentes	Observações	Zeros na amostra	Media	Desv. Padrão	Mín	Máx
Inc_todas_idades	10225	3703	3,36	9,25	0	458,65
Inc_0a14	10225	4174	5,86	19,36	0	1000

Os dados da Tabela 3.11 mostram que a porcentagem de zeros na amostra é elevada. Para a população de todas as idades, a porcentagem de zeros corresponde a 36,2% e, para a população de 0 a 14 anos, essa porcentagem é de 40,8%. Os dois mapas abaixo mostram a distribuição dos distritos que registraram zero internações de DRSAI para o ano de 2010. O

mapa à esquerda refere-se à população de todas as idades e o mapa à direita refere-se à população de 0 a 14 anos. Tais mapas mostram que a distribuição dos zeros não está localizada em uma região específica do país. Ou seja, não se trata de valores faltantes, mas sim de características intrínsecas aos distritos “saudáveis”.

Figura 3.2 – Valores nulo da taxa de incidência de interações por DRSAl

As variáveis explicativas de interesse desse estudo captam todas as dimensões do setor de saneamento básico. A lei federal 11.445, de janeiro de 2007, estabelece saneamento básico como o setor que contempla os serviços de abastecimento de água, esgotamento sanitário, limpeza urbana e manejo de resíduos sólidos, drenagem e manejo de águas pluviais. Haverá uma variável representativa para cada uma dessas dimensões (água, esgoto, lixo e drenagem de águas pluviais), de forma a considerar o maior alcance do setor em termos de impactos sobre a saúde pública.

As próximas duas Tabelas, 3.12 e 3.13, apontam os valores das variáveis de saneamento básico segmentados pelos grupos acima destacados para as cinco regiões brasileiras. As colunas hachuradas tratam das variáveis escolhidas para serem utilizadas nas regressões. As demais variáveis são utilizadas a título de comparação.

Tabela 3.12 – Acesso à abastecimento de água, lixo e drenagem de águas pluviais

Regiões	Abastecimento de Água					Lixo		Drenagem de águas pluviais
	Ag rede	Ag nao ceuab	D poço	D cist	D outros	D lixo	Lixo mor inad	D não bueiro
Norte	54,50	34,03	31,59	0,07	13,84	74,29	0,30	54,02
Nordeste	76,63	47,12	7,91	2,07	13,39	75,00	0,12	57,41
Sudeste	90,31	81,51	6,74	0,07	2,87	94,99	0,03	45,30
Sul	85,56	74,15	10,74	0,02	3,68	91,67	0,02	29,33
Centro-Oeste	81,81	76,66	14,92	0,15	3,12	89,71	0,03	57,68
Brasil	82,89	67,76	10,01	0,59	6,51	87,44	0,07	47,52

O primeiro grupo é referente aos tipos de abastecimento de água dos domicílios. A variável da segunda coluna (**ag_rede**) indica a porcentagem de domicílios particulares permanentes conectados à rede geral de distribuição de água e a variável da terceira coluna (**ag_nao_ceuab**) se refere à porcentagem de domicílios particulares permanentes conectados à rede geral de distribuição de água e que, além disso, na quadra onde estão localizados é inexistente qualquer tipo de esgotamento sanitário a céu aberto. Ou seja, além de tais domicílios possuírem o acesso adequado à água, via rede geral, ainda apresentam as melhores condições de entorno, pois na região em que estão localizados não existe esgoto a céu aberto. Todas as regiões apresentaram valores menores para a variável “ag_nao_ceuab” em relação à “ag_rede”. Esse resultado era esperado, pois para a primeira variável, além dos domicílios terem que possuir acesso à água via rede geral também não podem apresentar esgoto a céu aberto em seu entorno. Esta variável representa as melhores e mais desejáveis condições de acesso à água, porém a exigência de ambos atributos implica em menor porcentagem de domicílios em tal situação. Para essas duas variáveis, a região Sudeste apresenta os melhores indicadores, seguida das regiões Sul, Centro-Oeste, Nordeste e Norte.

As outras variáveis tratam dos tipos alternativos de abastecimento de água considerados inadequados: porcentagem dos domicílios servidos por água de poço ou nascente localizado no terreno da propriedade (**d_poço**), porcentagem dos domicílios servidos por água de chuva armazenada em cisterna ou caixa de cimento (**d_cist**) e a porcentagem dos domicílios abastecidos de outra forma diferente das descritas anteriormente (**d_outros**). No tocante à porcentagem dos domicílios com coleta de lixo por empresa especializada (**d_lixo**), notam-se taxas elevadas para todas as regiões. Por conseguinte, a porcentagem dos domicílios que apresentam moradia inadequada e que em sua face ou quadra existe lixo acumulado (**lixo_mor_inad**) é pequena. Porém, essa variável é representativa de uma precária condição de limpeza urbana e manejo de resíduos sólidos, uma vez que diz respeito ao acúmulo de lixo

nas moradias, que também se constitui em foco de propagação de doenças. A variável escolhida para representar a inapropriada condição de drenagem de águas pluviais é a porcentagem dos domicílios que não possuem bueiro ou boca de lobo, ou seja, a abertura que dá acesso a caixas subterrâneas, por onde escoam a água proveniente de chuvas (**d_nao_bueiro**). A não existência dessa infraestrutura provoca alagamentos, assim como a transmissão de diversas doenças, tais como a esquistossomose.

É importante destacar que, dessa Tabela 3.13, as variáveis escolhidas para a análise espacial econométrica foram aquelas hachuradas: “ag_nao_ceuab”, “d_cist”, “lixo_mor_inad” e “d_nao_bueiro”. A Tabela subsequente expõe as variáveis dos diversos tipos de esgotamento sanitário para os domicílios que têm banheiro ou sanitário¹⁵.

Tabela 3.13 – Acesso aos tipos de coleta de esgoto

Regiões	Esgotamento sanitário										
	Esg_rede	D_fsep	D_red_fsep	Esg_nao_ceuab	D_frud	D_vala	D_riomar	D_outros	D_sbanh	Sbceuab	
Norte	13,99	18,85	32,83	21,60	48,25	5,84	2,64	5,82	4,62	0,49	
Nordeste	34,00	11,24	45,24	33,26	39,94	3,27	1,47	2,27	8,06	0,44	
Sudeste	81,10	5,44	86,54	78,68	7,61	1,89	3,09	0,48	0,38	0,01	
Sul	45,88	25,64	71,52	61,85	24,59	2,05	0,88	0,50	0,46	0,05	
Centro-Oeste	38,43	13,15	51,58	48,64	46,70	0,37	0,25	0,48	0,63	0,01	
Brasil	55,50	11,59	67,09	58,02	24,43	2,44	2,08	1,32	2,71	0,16	

A variável **esg_rede** refere-se à porcentagem dos domicílios que têm acesso à rede geral de coleta de esgoto¹⁶ e a variável **d_fsep** indica a porcentagem dos domicílios que têm banheiro ou sanitário ligado a uma fossa séptica, que consiste em um sistema que leva a matéria esgotada para uma fossa próxima para passar por um processo de tratamento ou decantação¹⁷. No tocante à primeira variável, o baixo valor desse indicador para o Brasil (55,5%) revela uma situação preocupante no que diz respeito à coleta de esgoto. A região Sudeste é a única que apresenta valor maior que o nacional. Quanto ao acesso à coleta de esgoto por fossa séptica, cabe destacar que a região Sul exibe o maior valor para esse tipo de esgotamento sanitário. Ao considerar os dois tipos de coleta acima mencionados, obtém-se a variável

¹⁵ Segundo IBGE (2010), considerou-se como banheiro o cômodo que dispunha de chuveiro (ou banheira) e vaso sanitário (ou privada) e sanitário o local limitado por paredes de qualquer material, coberto ou não por um teto, que dispunha de vaso sanitário ou buraco para dejeções.

¹⁶ O banheiro ou sanitário do domicílio está ligado à rede geral de esgoto quando a canalização das águas servidas e dos dejetos está ligada a um sistema de coleta que os conduzia a um desaguadouro geral da área, região ou município, mesmo que o sistema não dispusesse de estação de tratamento da matéria esgotada (IBGE, 2010).

¹⁷ Também neste caso, a parte líquida é conduzida em seguida para um desaguadouro geral da área, região ou município sem garantias de que seja tratada.

d_red_fsep, que denota a porcentagem dos domicílios com acesso a coleta de esgoto por rede geral ou fossa séptica.

A análise da coleta de esgoto nos domicílios também se vale de uma variável que traduz as melhores condições de acesso a esse serviço de saneamento básico: a porcentagem dos domicílios que possuem esgotamento sanitário por rede geral ou fossa séptica e que não apresentam em suas condições de entorno (face ou quadra) esgoto a céu aberto (**esg_nao_ceuab**). Como era de se esperar, há uma redução deste valor expresso na quinta coluna (**esg_nao_ceuab**) com respeito à quarta coluna (**d_red_fsep**). Há ainda as variáveis típicas das formas de coleta de esgoto consideradas inadequadas ao meio ambiente bem como à saúde pública: quando o banheiro ou sanitário está ligado a uma fossa negra, poço ou buraco (**d_frud**), quando o mesmo está ligado diretamente a uma vala a céu aberto (**d_vala**), quando as águas servidas e dejetos são jogados diretamente no rio, lago ou mar (**d_riomar**) ou nenhuma das alternativas anteriores (**d_outros**).

Existe ainda a variável que diz respeito à porcentagem de domicílios que não têm banheiro ou sanitário (**d_sbanh**) e uma variável que reflete as piores condições de saneamento, pois trata-se da porcentagem de domicílios que não têm banheiro e, além disso, possuem presença de esgoto a céu aberto em suas condições de entorno (**sbceuab**). Apesar dos baixos valores, as regiões Norte e Nordeste são as que possuem maior porcentagem de domicílios que vivem na ausência total de esgotamento sanitário (**sbceuab**) e, em contrapartida, aquelas que apresentam as menores taxas de esgotamento sanitário considerado “ideal”, por não serem circundados por condições de entorno à céu aberto (**esg_nao_ceuab**). Ao relacionar tais variáveis de interesse às internações por DRSAI, os sinais esperados de cada uma dessas variáveis para os dois modelos estão expressos no Quadro 3.3.

Modelos	Saneamento	Internações por DRSAI
Modelo 1	D_cist	+
	Sbceuab	+
	D_frud	+
Modelo 2	Ag_nao_ceuab	-
	Esg_nao_ceuab	-
	Lixo_mor_inad	+
	D_não_bueiro	+

Quadro 3.3 – Sinais esperados - relação entre Saneamento e Internações por DRSAI

No capítulo 5, os Modelos 1 e 2, estimados com cada grupo de variáveis descrito acima, serão apresentados formalmente. Além disso, os sinais de tais variáveis serão justificados de acordo com os estudos empíricos e o modelo teórico. No entanto, de forma geral, espera-se maior taxa de incidência de internações por DRSAI para as variáveis que indicam condições inadequadas de acesso aos serviços de saneamento básico (“d_cist”, “sbceuab”, “d_frud”, “lixo_mor_inad” e “d_nao_bueiro”). Por outro lado, as duas variáveis indicativas das melhores condições de entorno e de acesso à água e à esgoto (“ag_nao_ceuab” e “esg_nao_ceuab”) devem proporcionar menores taxas de incidência de internações por DRSAI. Logo, justifica-se o sinal positivo e negativo no primeiro e segundo caso respectivamente. As estatísticas descritivas de todas as variáveis explicativas da amostra utilizadas nesse estudo encontram-se na Tabela 3.14:

Das variáveis de controle, apresentadas nesta tabela abaixo, excluindo as variáveis de interesse explicitadas acima, três delas foram selecionadas de acordo com o arcabouço teórico apresentado anteriormente. A primeira delas é uma *proxy* de cuidados médicos (est_saude), que é dada pelo número de estabelecimentos de saúde em cada distrito da amostra. A outra é a renda per capita (rendapc), análoga à influência do salário para a produção de saúde do indivíduo¹⁸. A variável de educação (mu_resp_alfab) corresponde à taxa de alfabetização das mulheres que são chefes de família.

Tabela 3.14 – Estatísticas Descritivas das variáveis de interesse e de controle

Variáveis	Obs	Média	Desv. Pad	Mín	Max
dcist	10225	1,41	5,03	0	85,97
sbceuab	10225	0,19	0,69	0	14,89
dfrud	10225	44,77	30,66	0	100
esg_nao_ceuab	10225	30,25	31,44	0	100
ag_nao_ceuab	10225	45,16	30,31	0	100
lixo_mor_inad	10225	0,06	0,41	0	13,71
d_nbueiro	10225	15,87	20,96	0	100
rendapc	10225	430,19	305,38	28,76	5373,99
gini	10225	0,60	0,06	0,30	0,92
mu_resp_alfab	10225	76,24	14,62	0	100
popurb	10225	56,89	29,98	0	100
dom_kmq	10225	153,46	681,49	0,02	16023,88
est_saude	10225	16,11	49,44	0	1259
dummy_cap	10225	0,04	0,19	0	1
dummy_reg_met	10225	0,05	0,22	0	1

¹⁸ No capítulo 5 será desenvolvido o modelo de Grosman (1972) que dá suporte teórico aos canais de transmissão do efeito de saneamento sobre saúde.

De acordo com Kassouf (1994) e Fewtrell *et al.* (2005), a educação da mãe é o principal determinante para a produção de saúde das crianças. Segundo Grosman (1972), as variáveis relacionadas à educação, renda e cuidados médicos implicam em maior estoque de saúde. De acordo com a variável de saúde considerada nesse estudo, o maior estoque de saúde implicaria em menor taxa de incidência de internações por DRSAI. Logo, os sinais esperados são negativos, como implicação direta do modelo teórico. Deaton (2003) propõe em seu modelo teórico que, à medida em que a renda é redistribuída, há melhores resultados de saúde. Logo, espera-se que o indicativo de maior concentração de renda, dado pelo coeficiente de Gini, resulte em maior taxa de incidência de internações por DRSAI. As duas variáveis *dummies* que indicam se o distrito pertence à região metropolitana e à capital foram adicionadas para captar o efeito da infraestrutura de saneamento básico e de saúde mais organizada e homogênea nessas localidades. As variáveis referentes à taxa de urbanização (popurb) e ao número de domicílios por km² (dom_kmq) também dizem respeito à organização prévia da oferta de tais serviços, bem como possíveis ganhos de escala que os ofertantes de tal serviço podem obter.

4 EFEITOS DO SANEAMENTO SOBRE EDUCAÇÃO

No capítulo anterior foram descritos os dois tipos de dados utilizados neste estudo. Neste capítulo serão analisados os impactos do saneamento sobre a educação, utilizando dados municipais dos anos de 2000 e 2010. Além da análise econométrica convencional para dados em painel com presença de endogeneidade, adotou-se também metodologia para o controle das características espaciais presentes nesse tipo de dados. A partir do referencial teórico dos determinantes educacionais buscou-se identificar se as características domiciliares onde os indivíduos residem são importantes para explicar variáveis de frequência escolar, distorção idade-série e abandono escolar. Ademais, é descrita a metodologia utilizada nos dois tipos de análises, bem como apresentam-se os resultados obtidos com sua utilização.

4.1 Saneamento básico e educação

Antes de definir os canais de transmissão pelos quais o acesso aos serviços de saneamento básico pode afetar indicadores educacionais, é preciso descrever quais são os principais determinantes dos indicadores de educação considerados pela literatura. Diversos estudos tratam da relação positiva entre anos de escolaridade e salário dos indivíduos ou sucesso no mercado de trabalho (Card, 2001; Kassouf, 2001). Além disso, a educação gera diferenças no nível de produtividade das pessoas, que perduram durante todo seu ciclo de vida (Menezes-Filho, 2001). Nesse sentido, a distribuição da educação também explica boa parte da desigualdade da renda do país: Barros e Mendonça (1996) argumentam que esta diminuiria em até 50% se a desigualdade educacional fosse eliminada. Os investimentos em educação, que geram notas positivas em testes de proficiência e aumentam a probabilidade de os alunos prosseguirem com os estudos, ocasionam crescimento econômico para um país, dada a importância do capital humano (BISHOP, 1989; TOPEL, 1999; HANUSHEK e KIMKO, 2000).

A importância da educação, tanto para o crescimento econômico de longo prazo quanto para obtenção de resultados individuais em termos de rendimentos salariais no futuro, propulsionou a investigação das várias dimensões que afetam os indicadores de rendimento escolar. Uma delas é referente ao *efeito-escola*, que se materializa nas características atribuídas ao ambiente físico em que ocorre o aprendizado e àquelas atribuídas aos profissionais que conduzem esse processo, os professores. Como ressaltado por Duflo (2001)

e Menezes-Filho (2007), as variáveis relativas à infraestrutura da escola são importantes para explicar o aumento da taxa de escolaridade, seja por meio de construção de novas escolas ou pela melhora na infraestrutura física das mesmas. Nesse sentido, melhores instalações no interior das escolas, um quadro de pessoal formado por professores qualificados e instrumentos de gestão eficientes, tais como a presença de diretoria na escola e funcionamento adequado dos trâmites administrativos, são fatores relevantes dentro dessa dimensão. Dessa forma, estudos nacionais e internacionais investigaram formas de melhorar os resultados escolares por meio desse tipo de investimento¹⁹.

A outra dimensão ressaltada pela literatura refere-se ao que se denomina de *background familiar*. Os estudos de Barros *et al* (2001), Vasconcellos (2003) e Menezes-Filho (2001) indicaram a importância das características familiares (principalmente a escolaridade dos pais) e dos fatores socioeconômicos dos alunos nos níveis de escolaridade dos mesmos. Tais autores apontam a família como principal determinante dos resultados educacionais dos filhos. Em relação a essas duas dimensões, Menezes-Filho (2007) destaca que as características relacionadas à escola têm efeitos importantes na explicação do desempenho escolar, porém reduzidos quando comparados às características familiares e socioeconômicas dos alunos.

No entanto, seja pela ação e influência dos pais ou pelo ambiente físico em que a escola funciona, o primeiro passo para elevar o nível médio de escolaridade de um país é elevar a frequência escolar e manter a criança na escola, garantindo-lhe o avanço de seus níveis educacionais (PONTILI e KASSOUF, 2007). Uma das explicações para a baixa frequência escolar, com consequente impacto em outros indicadores de rendimento escolar tais como abandono e distorção idade-série, é a saúde deficiente. De acordo com Banerjee e Duflo (2011), dentre as prováveis explicações para evasão escolar infantil em vários países do mundo, além da falta de exigência dos pais em obrigar as crianças a continuarem os estudos, está a falta de estímulo das mesmas em frequentarem as escolas. Essa falta de estímulo pode estar associada a um estado de saúde comprometido.

¹⁹ Para resultados da literatura internacional, ver Coleman (1966); Hanushek (1986); Card e Krueger (1992), Heckman, Layne-Farrar e Todd (1996); Rivkin, Hanushek e Kain (2005). Para o caso brasileiro, ver Ferrão et al (2001), Albernaz, Ferreira e Franco (2002); Macedo (2004).

Miguel e Kremer (2004) encontraram um efeito associado ao motivo doença para explicar a evasão escolar. O estudo de tais autores indicou efeito positivo da medicação contra *helminths* (“vermes”) intestinais, alojados em crianças quenianas, sobre a frequência escolar. Os autores ressaltam a importância da adequada provisão de bens públicos de saneamento, uma vez que a transmissão de tais agentes patogênicos ocorre pela matéria fecal não tratada ou não disposta corretamente pelos mesmos. Ou seja, oferecer tais serviços é uma forma de impactar positivamente o desempenho das crianças na escola.

O estudo de Neri et al. (2008), ao fazer análises de correlações brutas entre variáveis de infraestrutura e desempenho escolar em 2001 e 2003, encontrou que o acesso à água apresenta correlação positiva com a proficiência escolar e negativa com o índice de reprovação. Neste último caso, segundo os autores, melhor infraestrutura, tanto na casa como na escola, deve melhorar o rendimento dos estudantes, reduzindo a repetência.

A literatura da área médica documenta as possíveis consequências de um estado de saúde debilitado decorrente das parasitoses intestinais e outras doenças que estão relacionadas à falta ou inadequada provisão dos serviços de saneamento básico (DRSAI). Segundo Padro *et al.* (2001), as infecções parasitárias estão entre as mais disseminadas desordens que afetam crianças em idade escolar que vivem em áreas pobres de grandes centros urbanos. Os efeitos patológicos diretos destes parasitas são diarreia crônica, má absorção dos alimentos e anemia ferropriva, que é decorrente da privação e deficiência de ferro dentro do organismo (KUNZ et al., 2008). Segundo ARAÚJO *et al.* (2009), a presença de anemia associada a enteroparasitoses é resultante da subnutrição (*Ascaris lumbricoides*), da ação hematofágica (*Ancylostoma sp.*) e da ulceração das mucosas intestinais (*Entamoeba histolytica*). O indivíduo sofre constantes perdas sanguíneas, além do agravamento do quadro patológico, a depender da carga parasitária, da idade, do estado nutricional e fisiológico do organismo, bem como da associação com outras espécies parasitárias patogênicas.

As consequências tendem a comprometer o comportamento e a função cognitiva das crianças, principalmente quanto à capacidade de atenção e concentração, que afetam o rendimento escolar e dificultam o aprendizado (ARAÚJO et al, 2009 e KUNZ et al., 2008). Outra consequência destacada por Brito *et al* (2003) é que a anemia tem efeitos negativos sobre o crescimento ponderal e estatural na infância e no período escolar (fases em que a necessidade nutricional é maior). Logo, o desenvolvimento físico e intelectual da população pediátrica é prejudicado (GURGEL et al., 2005; LUDWIG et al., 1999) e, nos adultos, o principal impacto

da anemia é a diminuição da capacidade produtiva, que se reflete na menor capacidade de trabalho.

Quando os fatores acima citados estão associados às condições precárias de saneamento básico, em que a ocorrência de verminoses se torna comum, crianças e adolescentes ficam expostos a uma situação de desequilíbrio crônico (DOMENE, 2004). Outro agravante de tal desequilíbrio crônico deve-se à perda dos maiores retornos do investimento em educação no início da infância. Segundo Heckman (2005), todo tipo de ação voltada para a primeira infância do indivíduo constitui o melhor investimento social existente e quanto mais baixa for a idade, mais alto é o retorno do investimento.

Carneiro, Cunha e Heckman (2003) encontraram que nos Estados Unidos o retorno de investimento em educação é muito maior para idades mais precoces. Para o Brasil, Curi e Menezes-Filho (2009) apontam que a educação pré-primária, voltada à população de zero a seis anos de idade, tem relação positiva e significativa com a conclusão dos ciclos escolares, com a escolaridade média e, conseqüentemente, com a renda do trabalho principal, indicando efeitos socioeconômicos positivos no futuro. Schady (2006) reforça tais conclusões, ao destacar que os prejuízos causados pelo desenvolvimento inadequado nas idades iniciais das crianças afetam não somente o desempenho escolar, mas também emprego, salário, criminalidade e medidas de integração social de adultos.

Os resultados encontrados pelos autores acima são importantes para esse estudo, pois a relação entre as condições sanitárias de uma localidade e os níveis educacionais da mesma será investigada à luz do estado de saúde dos indivíduos, assunto a ser tratado no próximo capítulo.

A última dimensão dos determinantes educacionais que será tratada nesse capítulo diz respeito aos possíveis efeitos espaciais presentes nas variáveis de educação. Ou seja, os municípios que apresentam altas taxas de rendimento escolar são capazes de gerar algum tipo de *spillover* espacial para os municípios vizinhos? Sabe-se que a literatura dessa área considera inúmeras variáveis não observadas presentes no termo de erro, tais como habilidade dos alunos e dos professores e características de gestão educacional específicas das escolas e municípios. Caso exista autocorrelação das variáveis não observadas no espaço, esse efeito pode ser controlado ao considerar a estrutura espacial dos dados, de forma a fornecer estimativas mais consistentes dos parâmetros. Por outro lado, ignorar efeitos espaciais

significativos pode proporcionar estimativas inconsistentes e viesadas dos parâmetros. Portanto, deve-se avaliar, assim como no estudo de Baylis et al. (2011), se a introdução de tais efeitos altera significativamente os resultados obtidos por meio da análise econométrica convencional.

Os efeitos da interação no espaço entre alunos, escolas e professores tem sido tema de estudo de alguns pesquisadores, principalmente estudos acerca da relação entre educação (e/ou sua distribuição) e renda (e/ou desigualdade de renda) em um país ou região. Rodriguez-Pose e Tselios (2008) apontam para uma relação positiva e robusta entre a desigualdade educacional e a desigualdade de renda entre as regiões da União Europeia. Umar *et al* (2014) encontraram efeitos positivos da distribuição equitativa da educação sobre o nível de renda regional nigeriano.

Na dimensão dos determinantes dos indicadores educacionais referentes ao *efeito-escola*, Milimet e Rangaprasad (2007) mostraram que os administradores de escolas públicas em Illinois, nos Estados Unidos, levam em consideração nas suas decisões acerca dos insumos escolares destinados a melhorar a qualidade educacional a escolha pelos mesmos insumos feita pelos administradores das escolas públicas dos distritos vizinhos. Ou seja, as decisões das escolas públicas relativas à razão aluno-professor, média salarial dos professores, gastos por aluno e tamanho da turma influenciam estrategicamente a escolha por tais insumos das escolas públicas vizinhas que se encontram no mesmo município. Plenzler (2004) também destaca a importância da escala espacial na análise da relação entre desempenho do aluno e recursos educacionais. Naidoo *et al* (2014), ao utilizar o método de regressão ponderada geograficamente, encontraram forte relação entre desempenho escolar e características socioeconômicas da localidade onde a escola está localizada na África do Sul. Dentro de tais características, ressaltam aquelas relativas às características dos pais e dos domicílios em que as famílias residem. Nesse contexto, Borjas (1995) encontrou que o nível educacional e de rendimentos alcançados pelos jovens dos Estados Unidos eram impactados pela renda dos pais e também pela renda média dos vizinhos de mesma etnia da geração de seus pais.

Os exemplos de interações espaciais demonstradas em estudos acima apontam a relevância de se levar em conta os efeitos de um espaço geográfico mais abrangente do que o municipal para a determinação de políticas educacionais.

4.2 Metodologia

O modelo geral para dados em painel é representado da seguinte forma:

$$(4.1) \quad EDUC_{it} = \beta_{0it} + \beta_{1i} SAN_{1it} + \dots + \beta_{kit} X_{kit} + a_i + \varepsilon_{it}$$

O subscrito i indica município ($i = 1, \dots, 5.565$), o subscrito t denota o ano ($t = 2.000$ e 2.010); $EDUC_{it}$ são os indicadores de educação (taxa de frequência escolar, de distorção idade-série e de abandono escolar), X_{kit} é o valor da k -ésima variável explicativa para o município i no instante t ($k = 1, \dots, 19$); os β_{ki} são os parâmetros a serem estimados (β_0 é o intercepto da equação); o termo a_i capta todos os fatores não observados dos municípios, constantes no tempo, que afetam $EDUC_{it}$; ε_{it} é o termo de erro para o i -ésimo município em t .

Para que esse modelo seja estimado consistentemente é necessário estabelecer hipóteses acerca do termo a_i . Neste estudo, considera-se que a estimação dos indicadores educacionais é condicional aos efeitos específicos dos municípios, ou seja, os a_i são tratados como parâmetros a serem estimados. Além disso, permite-se que esse termo seja correlacionado com as variáveis explicativas: $E(a_i X_{it}) \neq 0$. É razoável supor que existam características não-observáveis dos municípios que também influenciem os indicadores educacionais e que podem estar correlacionadas com as variáveis explicativas, ou seja, características comunitárias (entre elas, o acesso aos serviços de saneamento), características socioeconômicas das famílias e os insumos escolares. A estimação do modelo de efeitos fixos para dados em painel necessita de três hipóteses. A primeira delas é a necessidade de exogeneidade forte ou estrita

$$(4.2) \quad E(\ddot{X}_{it} \ddot{\varepsilon}_{it}) = 0, \text{ em que } \ddot{X}_{it} = X_{it} - \bar{X}_t \text{ e } \ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_t.$$

Dado que a implementação desse método subtrai a média ao longo do tempo para cada município i , não é possível que as variáveis explicativas sejam correlacionadas com o termo de erro contemporaneamente e nem em qualquer outro período do tempo. A outra hipótese necessária é a inexistência de perfeita multicolinearidade, ou seja,

$$(4.3) \quad \text{Posto } [E(X_{it} X_{it})] = K$$

A terceira hipótese refere-se à presença de erros homocedásticos e não autocorrelacionados, conforme enunciado abaixo

$$(4.4) \quad E(\varepsilon_{it}\varepsilon_{it}'/X_{it}'a_i) = \sigma_u^2 I.$$

No entanto, caso se assuma que os efeitos individuais são aleatórios, de forma que se entenda que o termo específico não observado (a_i) seja não correlacionado com X_{it} , a variância desse termo poderá ser estimada de acordo com o modelo de efeitos aleatórios. Esse modelo também necessita da hipótese de exogeneidade forte, $E(\varepsilon_{it}/X_{it}, a_i) = 0$, e da hipótese de não correlação de a_i para cada valor de X_{it} , $E(a_i/X_{it}) = 0$. Para a escolha entre o modelo de efeitos fixos ou aleatórios, foi utilizado o teste de especificação de Hausman (1978), que é utilizado para comparar os coeficientes estimados no modelo com efeitos fixos com aqueles estimados no modelo com efeitos aleatórios. Sob a hipótese nula deste teste, se os efeitos específicos não estão correlacionados com as variáveis explicativas, o estimador de efeitos aleatórios é consistente e eficiente, enquanto que o estimador de efeitos fixos produz estimativas consistentes, mas não eficientes. Por outro lado, caso tais efeitos sejam correlacionados com as variáveis explicativas (ou seja, se a hipótese nula é rejeitada), o estimador de efeitos fixos será consistente, enquanto que o de efeitos aleatórios será inconsistente (o mais grave problema dentre as características desejáveis para um estimador).

O teste de Hausman apontou o modelo de efeitos fixos como o mais adequado para todas as especificações. O modelo de efeitos fixos controla as características não observáveis de cada município e constantes no tempo e, além disso, pode minorar os inúmeros problemas que cabem no termo de endogeneidade nas estimações com dados prospectivos (Franco e Menezes-Filho, 2009).

Mesmo considerando as vantagens de se aplicar a metodologia de efeitos fixos para dados em painel, a variável de saneamento é endógena nesse modelo: $Cov(SAN_{it}, \varepsilon_{it}) \neq 0$.

O teste de Durbin-Wu-Hausman de endogeneidade dos regressores para o modelo de efeitos fixos rejeitou a hipótese nula de que a variável dependente é exógena. Além desse teste detectar a endogeneidade, a literatura do setor também traz argumentos que reforçam essa hipótese. A presença de saneamento afeta os indicadores de educação, pois melhores condições sanitárias nos domicílios ou escolas fornecem condições de infraestrutura física

mais adequadas que, por sua vez, se refletem em superiores indicadores de desempenho escolar²⁰. Por outro lado, maiores níveis de escolaridade levam a melhores condições de saneamento, seja por meio do conhecimento sobre as condições de higiene ou sobre as formas de se evitar e prevenir a incidência de doenças relacionadas a saneamento²¹.

De acordo com Wooldridge (2002), há três possíveis fontes para a existência de endogeneidade nos dados: simultaneidade, erro de medida e a omissão de variáveis correlacionadas com pelo menos uma das variáveis explicativas. Neste estudo ocorre tanto o viés de simultaneidade, em que a variável explicativa afeta a dependente e vice-versa, quanto a existência de variáveis omitidas na explicação das variáveis de desempenho educacionais. A solução para resolver o problema de endogeneidade é a utilização do método de Mínimos Quadrados em Dois Estágios (MQ2E ou 2SLS) com variáveis instrumentais (Greene, 2003; Davidson & Mackinnon, 2004).

Para entender a construção do instrumento utilizado nesse estudo é necessário investigar a estrutura de provisão dos serviços de saneamento no Brasil. A privatização dos serviços de saneamento foi possível com a Lei de Concessões de 1995. Antes disso, os serviços eram oferecidos somente por empresas públicas, municipais ou estaduais. O governo federal implementou em 1971 o primeiro plano voltado para eliminação do déficit de abastecimento de água e serviços de esgoto, o Plano Nacional de Saneamento, Planasa (MPO e Sepurb, 1995). Segundo Galvão Jr. *et. al.* (2009), o Planasa constituiu marco da participação dos estados na provisão e operação da infraestrutura de saneamento no Brasil, uma vez que foram criadas as Companhias Estaduais de Saneamento Básico (CESBs), que deveriam obter as concessões dos serviços diretamente do poder concedente, as autoridades municipais.

A estrutura de provisão do setor de saneamento é caracterizada pelo predomínio das empresas públicas regionais, refletindo os diversos incentivos concedidos no Planasa para esse tipo de organização. Mas existem outros tipos de provedores que apresentam diferentes formas de atuação geográfica e diversas naturezas jurídicas na forma de gestão. Segundo Seroa da Motta (2006), o setor de saneamento básico apresenta uma complexa estrutura de provisão, em que predominam diversos tipos de prestadores com objetivos diferentes. O SNIS (2015) permite

²⁰ Os artigos que ressaltam o *efeito-escola* mostram que a infraestrutura física das escolas, ou seja, a adequada provisão dos serviços de saneamento básico, é um fator que pode afetar o rendimento escolar.

²¹ Kassouf (1995).

identificar basicamente três grupos de prestadores, agrupados a partir de sua abrangência geográfica, que são: prestadores regionais, microrregionais e locais, cuja natureza jurídica pode ser tanto pública quanto privada²².

As empresas que atendem a mais de um município, distribuídas por estado, e que geralmente atendem as regiões metropolitanas, são chamadas pelo SNIS de prestadores regionais. Existem empresas regionais públicas e privadas. As do primeiro tipo são as que atendem a maior parte da população brasileira. A Companhia de Saneamento Básico do Estado de São Paulo (Sabesp) e a Companhia de Saneamento de Minas Gerais (Copasa) são dois exemplos de empresas desse grupo. Existe uma única empresa regional privada no país, a Companhia de Saneamento do Tocantins (Saneatins), cuja natureza jurídica é de sociedade de economia mista, porém com administração privada. As empresas locais fornecem serviços de saneamento básico exclusivamente a um único município e também podem ser públicas ou privadas. A Tabela 4.1 apresenta o número de empresas nessas quatro categorias.

Tabela 4.1 – Empresas e municípios atendidos por empresas públicas e privadas					
		2000		2010	
		Número de empresas	Número de municípios	Número de empresas	Número de municípios
PÚBLICAS	REGIONAL	27	3386	27	3522
	LOCAL	2047	2047	1813	1813
	TOTAL PÚBLICAS	2074	5433	1840	5335
PRIVADAS	REGIONAL	1	102	1	138
	LOCAL	28	28	90	90
	TOTAL PRIVADAS	29	130/9*	91	228/186*
PÚBLICAS + PRIVADAS TOTAL		2103	5565	1931	5565

Fonte: Elaboração própria a partir de dados da Abcon e SNIS (2013)

*municípios que possuem empresas privadas de saneamento há cinco anos

De acordo com essa Tabela, no ano 2000, do total de 130 municípios atendidos por empresas privadas de saneamento básico, em apenas nove deles essas empresas atuavam há pelo menos cinco anos. Em 2010, esse número aumentou para 186, de um total de 228 municípios que tinham seus serviços de saneamento providos por empresa privada há qualquer tempo.

O mapa da Figura 4.1 apresenta a distribuição espacial de tais municípios para o ano de 2010. O tempo de exposição refere-se ao número de anos que a empresa privada, que fornece os

²² Nesse trabalho convencionou-se que as empresas do grupo “privadas” são aquelas classificadas pelo SNIS como *empresa privada* e aquelas do grupo “públicas” são as empresas que exibem as seguintes naturezas jurídicas: *sociedade de economia mista com administração pública*, *empresa pública*, *autarquia* e *administração pública direta*.

serviços de saneamento, atua no município. Nota-se que as empresas que ofertam tais serviços nos municípios do Tocantins operam nestes há mais de uma década.

Figura 4.1 – Tempo de exposição das empresas privadas nos municípios

Portanto, o instrumento utilizado nesse estudo será o número de empresas que foram privatizadas há mais de 5 anos para os dois períodos de tempo. Para que o instrumento seja válido, é importante que duas hipóteses sejam satisfeitas. A primeira delas refere-se à relevância do instrumento: $E(PRIV_{it}, SAN_{it}) \neq 0$. Ou seja, o mesmo deve ser correlacionado com a variável endógena (saneamento). Os resultados a serem apresentados adiante indicam que os municípios com serviços de água e esgoto fornecidos por uma empresa privada há mais de cinco anos exibem melhores indicadores de acesso a saneamento básico. No entanto, o problema de instrumento fraco pode surgir, mesmo quando a correlação entre a variável endógena e o instrumento é significativa aos níveis convencionais de 5% e 1%. O teste de Stock-Yogo²³ (2005), que verifica se o instrumento é fraco ou o estimador é fracamente identificado, rejeitou a hipótese nula de que o instrumento é fraco.

²³ O teste de Stock-Yogo (Stock e Yogo, 2005) é calculado tendo por base a estatística F de Cragg e Donald (1993). Sob a hipótese nula o estimador é fracamente identificado no sentido de que o viés verificado é inaceitavelmente grande. Os valores críticos deste teste são obtidos a partir de uma distribuição não padrão.

A outra hipótese necessária para assegurar a validade do instrumento diz respeito à exogeneidade do mesmo: $E(PRIV_{it}, \varepsilon_{it}) = 0$. Ou seja, a decisão de privatizar, que envolve um contrato entre o município e a empresa privada, não deve ser correlacionada com fatores não observados dos indicadores de educação. Apesar dessa hipótese não ser passível de teste, é possível garantir sua validade, pois a literatura mostra que essa decisão é política. Saiani e Azevedo (2014) apontam que as privatizações podem ser entendidas como estratégias políticas. O primeiro fato desse resultado é que os riscos eleitorais percebidos pelos prefeitos, diante da perspectiva de insucesso no próximo pleito, motivariam a decisão pela privatização, que seria adotada para “amarrar as mãos” da próxima administração. Ou seja, os prefeitos privatizariam para diminuir a discricionariedade²⁴ de um eventual oponente incumbente (mesmo que isso implique em reduzir o seu próprio escopo de atuação). Em relação ao segundo fato, considerando a existência de provisão dos serviços de saneamento básico pelos governos estaduais, o outro fato está relacionado a reduzir (ou não aumentar) o escopo de atuação dos governadores estaduais nos municípios nos quais os prefeitos não pertencem a partidos da coligação de tais governadores (SAIANI e FURQUIM, 2014). Em outras palavras, esta ação se constitui em estratégia política para diminuir a importância do governo do estado na economia do município e até mesmo do próprio estado.

A partir de tais argumentos e dos testes realizados, que são apresentados na seção de resultados, é possível garantir que o instrumento apresentado é válido como forma de solucionar a endogeneidade por meio do método de Mínimos Quadrados em dois estágios para dados em painel.

A partir de tal instrumento e das variáveis exógenas, que são instrumentos delas mesmas, uma matriz Z é utilizada para realizar o procedimento em dois estágios. No primeiro estágio, estima-se a equação na forma reduzida, ou seja, a variável explicativa endógena é regredida em função da matriz de instrumentos Z , de tal forma que:

$$(4.5) \quad San_{it} = Z_{it}\gamma + \vartheta_{it}$$

²⁴ Ações discricionárias em diversas áreas de serviços públicos, entre elas, a de saneamento, podem ser favoráveis para garantir votos a políticos que desejam se reeleger. Por exemplo, evitar a interrupção de serviços por falta de pagamento e assegurar empregos no setor são medidas populares que podem render votos (principalmente quando ações como essa recaem sobre a população de baixa renda).

$$(4.6) \quad San_{it} = \widehat{San_{it}} + \widehat{\vartheta_{it}} \rightarrow \widehat{\vartheta_{it}} = San_{it} - \widehat{San_{it}}$$

O parâmetro γ precisa ser significativo a fim de que seja válida a relevância do instrumento. A variável endógena tem duas partes: uma que não é correlacionada com o erro ε_{it} e a segunda que é correlacionada com o mesmo. No segundo estágio, estima-se o modelo para a variável dependente em função dos valores previstos do primeiro estágio, ou seja, utiliza-se $\widehat{San_{it}}$ como instrumento para San_{it} .

$$(4.7) \quad EDUC_{it} = \widehat{San_{it}} + \widehat{\vartheta_{it}} + a_i + \varepsilon_{it}$$

A intuição do método é que o instrumento afete a variável de resposta somente por meio da correlação que apresenta com a variável endógena. Em outras palavras, o instrumento impacta a variável endógena via parte que não é correlacionada com erro. A estimação não realizou o segundo estágio manualmente, e sim de uma única vez, de forma a evitar que os erros-padrão e as estatísticas de teste obtidos estivessem invalidados²⁵.

4.2.1 Análise Exploratória dos Dados Espaciais (ESDA)

Anteriormente à análise empírica dos dados espaciais, é necessário verificar se existem evidências de dependência espacial no modelo. No contexto desse capítulo, é preciso investigar se os resíduos do modelo convencional estimado em dois estágios com efeitos fixos e variável instrumental mostram-se autocorrelacionados espacialmente. Ou seja, deve-se aplicar o teste I de Moran para os resíduos das unidades de corte transversal para cada período (ALMEIDA, 2012). A matriz de ponderação espacial (w_{ij}^*) contígua e normalizada na linha e o coeficiente de autocorrelação espacial I de Moran são definidos a seguir:

$$(4.8) \quad w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}}$$

$$(4.9) \quad w_{ij}^* = \begin{cases} 0, & \text{se não há fronteira entre } i \text{ e } j \\ 1, & \text{se há fronteira entre } i \text{ e } j. \end{cases}$$

²⁵ As estimações foram realizadas pelo *software Stata* por meio do comando “xtivreg”.

$$(4.10) \quad I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} Z_i Z_j}{\sum_{i=1}^n Z_i^2}$$

Na equação 7, n é o número de regiões, z denota os valores das variáveis de interesse padronizadas e Wz representa os valores médios das variáveis de interesse padronizadas dos vizinhos. As variáveis de interesse são os resíduos da regressão do modelo não espacial e as variáveis dependentes de educação. Dado que a matriz de pesos espaciais é normalizada na linha, tem-se que $n = S_0$.

Duas formas de inferência foram adotadas: a primeira delas (“Res_1” e “Lag”) foi baseada em um processo que assume 99 permutações aleatórias de Monte Carlo; a segunda (“Res_2”) assume que a variável padronizada de interesse $Z(I)$ segue uma distribuição normal com média zero e variância unitária.

$$(4.8) \quad Z(I) = \frac{[I - E(I)]}{DP(I)}$$

Na Tabela 4.2 encontram-se, para cada ano, as estatísticas I de Moran calculadas com os resíduos, segundo os dois tipos de inferência, e com a variável dependente segundo inferência via permutações aleatórias.

Tabela 4.2 – Estatísticas I de Moran para os resíduos e defasagem espacial				
I de Moran		Frequência Escolar	Distorção Idade-Série	Abandono Escolar
2000 (Res_1)	Estatística	0,0037	0,0432	0,054
	P-Valor	0,37	0,001	0,001
2000 (Res_2)	Estatística	0,00374932	0,043228031	0,05398291
	P-Valor	0,717	0,0001	0,000001
2010 (Res_1)	Estatística	0,0037	0,0432	0,054
	P-Valor	0,37	0,001	0,001
2010 (Res_2)	Estatística	0,00374937	0,043228032	0,05398290
	P-Valor	0,717	0,0001	0,000001
2000 (Lag)	Estatística	0,0514	0,365	0,0336
	P-Valor	0,001	0,001	0,001
2010 (Lag)	Estatística	0,0166	0,0722	0,2196
	P-Valor	0,07	0,001	0,001

Os dados da Tabela 4.2 mostram que a hipótese nula de aleatoriedade espacial com os resíduos do modelo não espacial foi rejeitada com segurança para as variáveis de distorção idade-série e abandono escolar pelos dois tipos de testes. No entanto, as três variáveis apresentaram evidências de presença de autocorrelação espacial na forma de defasagem

espacial da variável dependente. Tais conclusões valem para os dois períodos de tempo em análise. Portanto, pelos testes acima apresentados, os efeitos espaciais precisam ser controlados nas estimações econométricas para garantir que os resultados não estejam sujeitos a viés ou inconsistência.

4.2.2 Metodologia espacial

Apesar de ser recente a aplicação de modelos espaciais em painel, diversos estudos empíricos, em diversas áreas, adotam essa metodologia. Há dois principais métodos desenvolvidos para a estimá-los: o método da Máxima Verossimilhança (MV)²⁶ e o método das Variáveis Instrumentais ou dos Momentos Generalizados (GMM)²⁷. O método de estimação para os dados em painel com erros espacialmente correlacionados e defasagem espacial desse estudo foi o GMM, de Piras e Millo (2012)²⁸. Segundo Almeida (2012), Kelejian and Prucha (1998), Brueckner (2003) e Milimet e Rangaprasad (2007), a vantagem deste método, em comparação ao de Máxima Verossimilhança (MV), deve-se à estimação consistente de modelos espaciais que apresentam variáveis endógenas entre as variáveis explicativas, além da defasagem espacial, e aqueles que também apresentam erros espacialmente correlacionados.

De acordo com o apresentado na seção anterior, esse é exatamente o caso deste estudo. Há endogeneidade na variável explicativa de saneamento e, além disso, efeitos espaciais detectados pela Análise Exploratória dos Dados Espaciais (ESDA). Não considerar tanto os efeitos fixos quanto a dependência espacial leva a uma ampliação do viés das estimativas em relação ao modelo convencional de dados em painel.

Outro fator que favorece a utilização do GMM é que, em modelos estimados por máxima verossimilhança, os parâmetros espaciais são estimados no intervalo $\left(\frac{1}{\alpha_{min}}, \frac{1}{\alpha_{max}}\right)$. O método IV/GMM não restringe esses parâmetros a este intervalo no processo de estimação (GOLGHER, 2015). No entanto, o problema do GMM é que este método não consegue determinar a significância estatística para o parâmetro ρ , que é considerado um parâmetro de distúrbio, porque é apenas usado para filtrar as variáveis (ALMEIDA, 2012). Piras e Millo (2012) se basearam em um modelo do tipo Cliff e Ord, analisado em Kapoor *et al.* (2007), e

²⁶ Elhorst (2014), Kalenkoski e Lacombe (2012) e, para o caso brasileiro, Chagas *et al.* (2014).

²⁷ Baylis *et al.* (2011).

²⁸ A estimação foi realizada pelo software R e o pacote utilizado foi o *splm*.

em uma extensão do estimador generalizado dos momentos (GMM) proposto por Kelejian e Prucha (1999) para estimar o modelo de erro espacialmente correlacionado. O modelo de defasagem espacial segue os estudos de Mutl e Pfaffermayr (2011) e Piras (2011). A especificação de tais modelos assume a seguinte forma:

$$(4.9) \quad y_N(t) = \lambda W y + X_N(t)\beta + u_N(t)$$

$$(4.10) \quad u_N(t) = \rho W_N u_N(t) + \varepsilon(t)$$

$$(4.11) \quad \varepsilon_N = (e_T \otimes I_N)\mu_N + v_N$$

Nas equações acima, o termo $y_N(t)$ é um vetor de dimensão $(N * t \times 1)$, $X_N(t)$ é uma matriz $(N * t \times k)$ de variáveis explicativas, I_N é uma matriz identidade $(N \times N)$, W_N é a matriz de pesos espaciais cuja dimensão também é $(N \times N)$, o vetor $u_N(t)$ de erros $(N * t \times 1)$ e $\otimes$ produto Kronecker entre duas matrizes.

Quando $\lambda = 0$, os parâmetros ρ , σ_μ^2 e σ_v^2 são estimados por GMM e os coeficientes restantes do modelo por um estimador de Mínimos Quadrados Generalizados (MQG ou GLS) factível. No entanto, ao se considerar incluídos no modelo tanto o termo da variável dependente defasada espacialmente (WY_{educ}) quanto os efeitos fixos, o procedimento passa a ser o de Mínimos Quadrados em Dois Estágios Espacial (Baltagi and Liu 2011).

De acordo com esse procedimento, além da variável dependente defasada ser instrumentalizada pelo vetor de parâmetros sugerido pelos autores, existe a possibilidade de instrumentalizar outro vetor de variáveis explicativas endógenas. Neste estudo, esse vetor é composto de uma única variável explicativa endógena, *saneamento*, que será instrumentalizada por *priv5anos*. Como apresentado em seção 3.2.1 pelos argumentos teóricos, testes implementados e exercício de robustez, esse instrumento possui as propriedades desejáveis de relevância e exogeneidade e por isso pode ser considerado também na análise espacial. Logo, reescrevendo a equação 4.9, o modelo de efeitos fixos pode ser expresso por:

$$(4.12) \quad y_N(t) = \lambda W y + X_N(t)\beta + D_N\gamma + u_N(t) = Z_N\delta + u_N(t)$$

Com tais modelos, tanto o problema das variáveis endógenas (a dependente defasada que é endógena por natureza e outras variáveis explicativas que também podem ser) quanto o problema de autocorrelação espacial dos erros são tratados. Segundo Anselin (1998), apesar dos problemas causados pela heterogeneidade espacial poderem ser abordados por técnicas econométricas não espaciais, o conhecimento teórico sobre a estrutura espacial dos dados pode tornar os procedimentos de estimação mais eficientes.

4.3 Resultados não-espaciais

As estimações do modelo não espacial descrito acima foram realizadas para sete diferentes variáveis dependentes: taxa de frequência escolar, taxa de abandono escolar no ensino fundamental e taxa de distorção idade-série também no ensino fundamental. Essas duas últimas foram segmentadas para o primeiro e segundo ciclo. Os resultados do primeiro estágio são apresentados na Tabela 4.3.

Tabela 4.3 – Resultados do primeiro estágio	
Variável Dependente	Saneamento
priv5anos - Variável Instrumental	11,18*** (0,839)
Constante	66,06*** (5,370)
Observações	7.959
R-quadrado (within)	0,7748
R-quadrado (between)	0,5572
R-quadrado (overall)	0,5999
Número de municípios	4.347
Características socioeconômicas	SIM
Características da escola e professores	SIM
Saúde e educação	SIM
Efeito fixo de município	SIM
Erro-padrão robusto entre parênteses	
*** p<0,01; ** p<0,05; * p<0,1	

O primeiro estágio mostra a significância da variável instrumental em relação à variável endógena de interesse. Esse é um dos indícios de que o instrumento é relevante: ter serviços de saneamento oferecidos por uma empresa privada há mais de cinco anos é correlacionado positivamente com a variável de acesso a saneamento, como era esperado. No entanto, como o problema de instrumento fraco pode surgir mesmo quando a correlação entre a variável

endógena e o instrumento é significativo aos níveis convencionais de 5% e 1%, foi realizado o teste de Stock-Yogo (2005). O resultado do teste indicou a relevância do instrumento, demonstrando que não se trata de um instrumento fraco.

Nas tabelas seguintes, a primeira estimação refere-se a um painel de efeitos fixos sem nenhum controle. Na segunda, terceira e quarta estimatórias, os controles foram sendo adicionados. A quinta estimação foi realizada em dois estágios, utilizando a variável instrumental descrita anteriormente. As estatísticas dos testes de Hausman, para escolha de se estimar os dados em painel por efeito fixo ou aleatório, validade do instrumento e endogeneidade, são descritas nas últimas linhas das tabelas. A Tabela 4.4 mostra as estimatórias descritas acima para a variável de frequência escolar.

Tabela 4.4 – Impacto do saneamento em frequência escolar 6-14 anos					
Variável Dependente	Frequência escolar da população de 6 a 14 anos				
	Efeito Fixo				IV Efeito Fixo
	(1)	(2)	(3)	(4)	(5)
Saneamento	0,224*** (0,0041)	0,040*** (0,0078)	0,022*** (0,0085)	0,024*** (0,0092)	0,118*** (0,0312)
Constante	76,61*** (0,329)	88,83*** (1,926)	93,45*** (2,656)	93,96*** (2,746)	87,94*** (3,005)
Observações	8.738	8.681	8.163	7.959	7.959
R-quadrado	0,485	0,628	0,651	0,642	-
Número de municípios	4.369	4.369	4.365	4.347	4.347
Características socioeconômicas	NÃO	SIM	SIM	SIM	SIM
Características da escola e professores	NÃO	NÃO	SIM	SIM	SIM
Saúde e educação	NÃO	NÃO	NÃO	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM	SIM
priv5anos - Variável Instrumental	-	-	-	-	SIM
Hausman - Chi2	-	-	-	832,2***	-
Durbin Wu Hausman - Chi2 (18)	-	-	-	-	36,79
Durbin Wu Hausman p-value	-	-	-	-	(0,0056)
Yogo and Stock (2005) - F statistic	-	-	-	-	154,96***
Erro-padrão robusto entre parênteses					
*** p<0,01; ** p<0,05; * p<0,1					

Os resultados da tabela acima mostram que o aumento de um por cento na proporção de pessoas que vivem em domicílios com acesso a água e esgoto gera aumento de 0,11 pontos percentuais na variável de frequência escolar da população de 6 a 14 anos. Esse resultado é compatível com o encontrado por Neri et al. (2008), de que o acesso aos serviços públicos (tanto água quanto luz elétrica) gera maiores índices de matrícula. A estimação de dados em painel por efeitos fixos, em dois estágios, utilizando variável instrumental, gerou coeficiente

4,9 vezes maior em relação ao coeficiente do modelo de efeitos fixos. O coeficiente deste modelo que era 0,024 (quarta coluna) foi para 0,118 no modelo completo expresso na quinta coluna, indicativo de que este tipo de controle pela endogeneidade é relevante para a análise em questão.

A Tabela 4.5 apresenta os mesmos tipos de estimação para a variável taxa de abandono escolar no ensino fundamental. O resultado obtido anteriormente é corroborado: um crescimento marginal na variável de saneamento provoca uma queda na taxa de abandono escolar em 0,49 pontos percentuais. Quando se controla a endogeneidade pelo uso da variável instrumental, o coeficiente torna-se oito vezes maior, em módulo, em relação àquele da estimação sem o controle. Ou seja, o coeficiente da variável de saneamento, que contém os três vetores de variáveis explicativas e o efeito fixo de município, porém, não faz o controle pela endogeneidade, é -0,0576 enquanto que o coeficiente desta mesma variável para o modelo completo da última coluna é -0,499.

Tabela 4.5 – Impacto do saneamento na taxa de abandono escolar					
Variável Dependente	Taxa de abandono escolar - ensino fundamental				
	Efeito Fixo				IV Efeito Fixo
	(1)	(2)	(3)	(4)	(5)
Saneamento	-0,371*** (0,00577)	-0,106*** (0,0101)	-0,0715*** (0,0106)	-0,0576*** (0,0109)	-0,499*** (0,0528)
Constante	36,35*** (0,458)	-15,60*** (2,615)	-12,09*** (3,406)	-14,45*** (3,432)	13,99*** (5,080)
Observações	8.683	8.680	8.163	7.959	7.959
R-quadrado	0,542	0,706	0,725	0,724	-
Número de municípios	4.369	4.369	4.365	4.347	4.347
Características socioeconômicas	NÃO	SIM	SIM	SIM	SIM
Características da escola e professores	NÃO	NÃO	SIM	SIM	SIM
Saúde e educação	NÃO	NÃO	NÃO	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM	SIM
priv5anos - Variável Instrumental	-	-	-	-	SIM
Hausman - Chi2	-	-	-	408,99***	-
Durbin Wu Hausman - Chi2 (18)	-	-	-	-	895,12
Durbin Wu Hausman p-value	-	-	-	-	(0,000)
Yogo and Stock (2005) - F statistic	-	-	-	-	154,9***
Erro-padrão robusto entre parênteses					
*** p<0,01; ** p<0,05; * p<0,1					

A Tabela 4.6 também confirma os resultados anteriores, agora com respeito à taxa de distorção idade-série (TDI). O resultado indica que um aumento de um por cento na população com adequado acesso a saneamento gera uma queda de 0,96 pontos percentuais na

taxa de distorção idade-série no ensino fundamental. Nesse caso, o controle da endogeneidade (modelo completo) apresentou coeficiente quatro vezes maior em relação ao modelo da quarta coluna, sem o controle pelo instrumento.

Tabela 4.6 – Impacto do saneamento na taxa de Distorção Idade-Série

Variável Dependente	Taxa de Distorção Idade-Série Ensino Fundamental				
	Efeito Fixo				IV Efeito Fixo
	(1)	(2)	(3)	(4)	(5)
Saneamento	-0,741*** (0,00917)	-0,310*** (0,0156)	-0,259*** (0,0168)	-0,234*** (0,0180)	-0,965*** (0,0847)
Constante	89,42*** (0,725)	11,24*** (4,346)	19,36*** (5,529)	18,28*** (5,652)	65,39*** (8,147)
Observações	8.589	8.586	8.163	7.959	7.959
R-quadrado	0,666	0,784	0,808	0,806	-
Número de municípios	4.369	4.369	4.365	4.347	4.347
Características socioeconômicas	NÃO	SIM	SIM	SIM	SIM
Características da escola e professores	NÃO	NÃO	SIM	SIM	SIM
Saúde e educação	NÃO	NÃO	NÃO	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM	SIM
priv5anos - Variável Instrumental	-	-	-	-	SIM
Hausman - Chi2	-	-	-	541,45***	-
Durbin Wu Hausman - Chi2 (18)	-	-	-	-	1455,68
Durbin Wu Hausman p-value	-	-	-	-	(0,000)
Yogo and Stock (2005) - F statistic	-	-	-	-	154,9***
Erro-padrão robusto entre parênteses					
*** p<0,01; ** p<0,05; * p<0,1					

A segmentação das duas últimas variáveis em primeiro e segundo ciclos é importante porque estes abrangem população de idades diferentes e com organismos e comportamentos diversos.

O primeiro ciclo é referente à população de seis a dez anos que frequenta a primeira fase do Ensino Fundamental (de primeira à quarta série ou do primeiro ao quinto ano). O segundo diz respeito à população de dez a catorze anos que frequenta a segunda fase do Ensino Fundamental (de quinta à oitava série ou do sexto ao nono ano).

Os resultados da Tabela 4.7 indicam que não houve importantes diferenças entre os ciclos em relação à magnitude dos coeficientes para a taxa de distorção idade-série. Enquanto a queda do primeiro ciclo foi de 1,0 ponto percentual, a do segundo ciclo foi de 1,04 pontos percentuais, considerando o modelo completo de efeitos fixos com variável instrumental.

Tabela 4.7 – Impacto do saneamento na TDI (1º e 2º ciclos)

Variável Dependente	Distorção Idade-Série		Distorção Idade-Série	
	Efeito Fixo	IV Efeito Fixo	Efeito Fixo	IV Efeito Fixo
	(1)	(2)	(3)	(4)
Saneamento	-0,271*** (0,0196)	-1,004*** (0,0873)	-0,259*** (0,0207)	-1,049*** (0,100)
Constante	20,58*** (5,872)	67,65*** (8,422)	26,40*** (7,031)	77,27*** (9,644)
Observações	7.951	7.951	7.959	7.959
R-quadrado	0,77	-	0,783	-
Número de municípios	4.347	4.347	4.347	4.347
Características socioeconômicas	SIM	SIM	SIM	SIM
Características da escola e professores	SIM	SIM	SIM	SIM
Saúde e educação	SIM	SIM	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM
Efeito fixo de ano	SIM	SIM	SIM	SIM
Hausman - Chi2	582,61***	-	502,87***	-
priv5anos - Variável Instrumental	-	SIM	-	SIM
Erro-padrão robusto entre parênteses				
*** p<0.01, ** p<0.05, * p<0.1				

Para a variável abandono escolar, os resultados da Tabela 4.8 mostram que o primeiro ciclo registrou redução ligeiramente maior: neste, um aumento percentual de saneamento gera uma queda de 0,52 pontos percentuais; no segundo ciclo, essa queda é de 0,47 pontos percentuais.

Em que pese a pequena diferença quantitativa, essa evidência vai ao encontro da hipótese estabelecida de que a população de seis a dez anos pode estar mais sujeita às doenças infecto-parasitárias transmitidas pela água do que os alunos mais velhos (crianças de dez a catorze anos). Quando estão menos expostas, uma vez inseridas em melhores condições de saneamento básico, ficam menos doentes, frequentando mais as aulas, repetindo menos e, portanto, tendo menos fatores associados ao estado de saúde para abandonar os estudos.

Tabela 4.8 – Impacto do saneamento na taxa de Abandono Escolar (1º e 2º ciclos)

Variável Dependente	Abandono escolar		Abandono escolar	
	Efeito Fixo	IV Efeito Fixo	Efeito Fixo	IV Efeito Fixo
	(1)	(2)	(3)	(4)
Saneamento	-0,112*** (0,0118)	-0,527*** (0,0528)	-0,00484 (0,0139)	-0,479*** (0,0664)
Constante	-2,694 (3,580)	24,00*** (5,083)	-25,17*** (4,358)	5,433 (6,390)
Observações	7.958	7.958	7.958	7.958
R-quadrado	0,678	-	0,633	-
Número de municípios	4.347	4.347	4.347	4.347
Características socioeconômicas	SIM	SIM	SIM	SIM
Características da escola e professores	SIM	SIM	SIM	SIM
Saúde e educação	SIM	SIM	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM
Efeito fixo de ano	SIM	SIM	SIM	SIM
Hausman - Chi2	373,2***	-	455,2***	-
priv5anos - Variável Instrumental	-	SIM	-	SIM
Robust standard errors in parentheses				
*** p<0,01; ** p<0,05; * p<0,1				

O teste de robustez apresentado abaixo procura verificar se há correlação entre o instrumento e a porcentagem do ensino cuja provisão é municipal, buscando determinar se a privatização dos serviços de saneamento básico poderia estar associada à descentralização também do ensino no nível municipal. Ou seja, caso a correlação seja positiva e alta, governos que privatizam os serviços de saneamento poderiam ter assumido maior protagonismo também na área de ensino, encampando atividades que antes eram estaduais ou municipais. Isso poderia ocorrer por serem mais dinâmicos e eficientes, ou mais alinhados ideologicamente com a ideia do aumento do poder decisório no nível municipal. Nessa situação, o instrumento poderia ter algum impacto também nos fatores não observados das variáveis educacionais.

O teste foi feito tanto com o instrumento utilizado nas regressões (município tem os serviços de saneamento básico oferecidos por uma empresa privada há mais de cinco anos) quanto com uma variável mais geral, que se refere aos municípios que têm os serviços de saneamento básico oferecidos por uma empresa privada, independentemente do tempo. Os resultados encontram-se na Tabela 4.9 abaixo:

Tabela 4.9 – Análise de robustez para o instrumento

Períodos	2000			2010		
Variáveis	priv5anos	priv	p_escol_mun	priv5anos	priv	p_escol_mun
priv5anos	1,0000	-	-	1,0000	-	-
priv	0,2594	1,0000	-	0,8987	1,0000	-
p_escol_mun	-0,0546	0,0001	1,0000	-0,0588	-0,0597	1,0000

Fonte: Elaboração própria a partir de dados da pesquisa.

A conclusão do teste é que, à medida que cresce o número de municípios que têm seus serviços oferecidos por uma empresa privada há mais de cinco anos, decresce a porcentagem do ensino que é responsabilidade do governo municipal. Mesmo que essa correlação seja de baixa magnitude nos dois anos (5,4 e 5,9%), não existe evidência de que a privatização dos serviços de saneamento básico esteja associada à descentralização do ensino no nível municipal.

A Tabela 4.10 apresenta os sinais e significâncias das variáveis explicativas utilizadas nas regressões, separadas nos três principais vetores descritos anteriormente: o primeiro deles refere-se às características socioeconômicas, o segundo às características das escolas e dos professores e o terceiro refere-se às variáveis de imunização e despesas com saúde e educação. Células em branco indicam que o coeficiente estimado não se mostrou estatisticamente diferente de zero; um sinal positivo indica que foi significativo e positivo; um sinal negativo indica que foi significativo e negativo. O número de sinais indica o nível de significância (um sinal, 10%; dois sinais, 5%; três sinais, 1%).

A comparação entre o modelo de efeitos fixos sem variável instrumental e o modelo de efeitos fixos estimado em dois estágios com esta variável permite concluir que controlar a endogeneidade altera os sinais das estimativas dos coeficientes em alguns casos (variáveis bolsa família e número de docentes com ensino superior). Apesar da robustez do sinal, significância e magnitude da variável de interesse para os sete tipos de especificações realizadas, algumas variáveis explicativas apresentaram resultados contra-intuitivos, problema que será contornado ao se controlar os efeitos espaciais na análise.

Variáveis Explicativas	Variáveis Dependentes					
	Frequência escolar		Distorção idade-série		Abandono escolar	
	EF	IV EF	EF	IV EF	EF	IV EF
t_analfabetismo_25m	---	---	+++	+++	+++	+++
pop_negra_parda	+++	+++	-	---	---	---
renda_per_capita	---	---	-	-	---	---
(1) gini	+	+	+	+++	+++	+++
pop_55mais			+++	+++	+++	+++
pop_15a29	--	---	+++	+++	+++	+++
bolsa_familia	+++	+++	---	+++	---	+++
pop_total			-	--	---	---
urbanização	+++	+++		+++	++	+++
alunos_turma_fundamental	---	---	+++	+++	++	+
docente_ensino_superior_fundamental	+++		---	+++	---	+
horas_aula_fundamental			---	---	-	-
(2) indice de infraestrutura escolar tecnológica				---	---	---
indice de infraestrutura escolar básica				+++		+++
porcentagem de escolas públicas						
porcentagem de escolas com diretoria	+		+++	+++	++	+++
despesa_educacao_cultura_por_habitante	+++	+++			--	
(3) despesa_saude_por_habitante			--		--	
imunização	+					
Observações	7,959	7,959	7,959	7,959	7,959	7,959
Número de municípios	4,347	4,347	4,347	4,347	4,347	4,347
Erro-padrão robusto entre parênteses						
() não significativa, (+ + +) ou (- - -) p<0.01, (+ +) ou (- -) p<0.05, (+) ou (-) p<0.1						

Quadro 4.1 – Sinal e significância e sinais das variáveis explicativas

Para o primeiro grupo de variáveis, a maior a taxa de analfabetismo da população adulta (de 25 anos ou mais) está associada a piores indicadores educacionais para as três variáveis: menor frequência escolar, maior abandono e maior taxa de distorção idade-série. Os municípios mais desiguais (maior índice de Gini) apresentaram piores indicadores para taxa de abandono e TDI. Para a variável de frequência escolar, esses mesmos municípios (com alto grau de concentração de renda) apresentaram maiores valores para a mesma. Por outro lado, esse indicador de frequência escolar tem associação positiva com o número de beneficiários do programa Bolsa Família no município, como era esperado²⁹. Em relação ao vetor de características das escolas e dos professores, a maior razão alunos por turma impacta negativamente todos os indicadores educacionais. Por outro lado, a maior quantidade de horas-aula diminui a taxa de distorção idade-série e o abandono, mas não afeta a frequência; assim como o índice de infraestrutura tecnológica. As maiores despesas com educação e cultura aumentam a frequência escolar e diminuem a taxa de abandono escolar no ensino fundamental.

²⁹ A condicionalidade do Programa Bolsa Família (PBF) estabelece que crianças e adolescentes de 6 a 15 anos devem possuir 85%, no mínimo, de frequência escolar para receber o Benefício Variável Jovem (BVJ). Além disso, a família assume compromissos nas áreas de saúde, tais como: acompanhamento pré-natal e acompanhamento nutricional (Campello e Neri, 2013).

4.4 Resultados espaciais

Na Tabela 4.10 são apresentados os resultados de quatro estimações para dados em painel. A primeira delas não leva em consideração o espaço. Ou seja, trata-se do resultado obtido em seção anterior, que se utilizou de um modelo de efeitos fixos para dados em painel, com variável instrumental, em dois estágios. As outras três partem do método de estimação GMM/IV para dados espaciais relativo ao período 2000-2010. Apesar desses resultados não poderem ser analisados como o impacto marginal da variável de saneamento sobre os indicadores educacionais³⁰, nota-se a diferença do controle da natureza espacial dos dados em painel.

Na quase totalidade dos casos (com exceção da variável de frequência escolar), os modelos espaciais resultaram em menor magnitude para as estimativas dos coeficientes. No contexto do impacto de mudanças climáticas sobre a agricultura, Baylis et al. (2011) também encontrou redução da magnitude das estatísticas dos coeficientes ao comparar o modelo de dados em painel com efeitos fixos não-espacial com o modelo também de efeitos fixos, mas que considera as características espaciais dos dados.

Apesar da variável da frequência escolar não apresentar evidências de autocorrelação espacial nos resíduos, esse indicador varia de acordo com a realidade socioeconômica dos estados e regiões (PONTILI & KASSOUF)³¹. São tais diferenças que justificam a necessidade de se analisar essa variável, pois ainda que o país apresente médias elevadas para ela³², há diversas regiões que apresentam baixas taxas para a mesma, carecendo de políticas educacionais direcionadas. A ESDA apontou que o efeito espacial, neste caso, ocorre sobre a forma de variável dependente defasada. Assim, a estimação do modelo de erro foi realizada apenas a título de comparação com os outros modelos.

³⁰ Não foi possível calcular os efeitos marginais, pois o comando não está implementado quando há variável endógena entre as explicativas.

³¹ A heterogeneidade espacial não observada admite situação de instabilidade estrutural, de forma que, a resposta de certos indicadores varia consideravelmente de acordo com a região de estudo.

³² Segundo Menezes-Filho (2007), desde meados da década de 90 o Brasil conseguiu aumentar significativamente a frequência escolar em todos os níveis, de forma que, segundo o autor, o desafio agora está em melhorar a qualidade da educação que é oferecida para alunos na rede pública.

Tabela 4.10 – Determinantes espaciais para a frequência escolar (2000-2010)

Variável Dependente	Frequencia escolar da população de 6 a 14 anos			
	EF Não Espacial	EF Lag	EF Erro	EF Lag e Erro
	(1)	(2)	(3)	(4)
Saneamento	0,118*** (0,0312)	0,0351*** (0,0130)	0,2422*** (0,0339)	0,0262** (0,0123)
Observações	7.959	7.224	7.224	7.224
Número de municípios	4.347	3.612	3.612	3.612
Características socioeconômicas	SIM	SIM	SIM	SIM
Características da escola e professores	SIM	SIM	SIM	SIM
Saúde e educação	SIM	SIM	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM
Efeito fixo de ano	NÃO	NÃO	NÃO	NÃO
priv5anos - Variável Instrumental	SIM	SIM	SIM	SIM
Lambda	-	0,1428*** (0,0321)	-	0,1686*** (0,0308)
Rho	-	-	0,1099	-0,0729
Sigma^2	-	-	-	14,74
Erro-padrão entre parênteses				
*** p<0,01; ** p<0,05; * p<0,1				

Os três últimos modelos da Tabela 4.10 apresentam resultados condizentes àqueles encontrados com o uso de econometria convencional: o acesso adequado aos serviços de saneamento básico municipal impacta positivamente o indicador de frequência escolar da população de 6 a 14 anos. Logo, há evidências de que dentro do grupo de fatores relacionados às características familiares e comunitárias na determinação dos níveis de escolaridade alcançado pelos indivíduos, o acesso aos serviços de saneamento básico no domicílio mostrou-se importante para alcançar esse objetivo.

No entanto, com exceção do modelo de erro espacial, a magnitude dos efeitos se alteraram. No Quadro 4.2, são apresentados o sinal e a significância das variáveis explicativas quando se faz o controle pelo efeito espacial. Os resultados corrigem alguns resultados contra intuitivos que foram anteriormente obtidos no modelo em que se utilizou econometria convencional. Pontili e Kassouf (2007) estimaram um modelo *probit* para duas variáveis qualitativas de frequência e atraso escolar dos estados de São Paulo e Pernambuco para o ano 2000, indicou os fatores que aumentam ou diminuem a probabilidade dos alunos de escolas públicas se matricularem nas mesmas e se manterem na série adequada a sua idade. De forma geral, os resultados encontrados por tais autoras são corroborados pelos resultados apresentados nesse estudo para essas duas variáveis.

Variáveis Explicativas	IV EF NE	EF Lag	EF Erro	EF Lag e Erro
t_analfabetismo_25m	---	---	---	---
pop_negra_parda	+++	+++	+++	+++
renda_per_capita	---	+++	+	+++
gini	+	---	--	---
pop_55mais		+++	+++	+++
pop_15a29	---	+++	+++	+++
bolsafam	+++	+++	+++	+++
poptot			-	
urbanização	+++	+++	+++	+++
alunos_turma_fundamental	---	---	---	---
docentes_com_ens_superior		---	---	---
h_a_total_fundamental		+++		+++
índice infraestrutura tecnológica		---	---	---
índice infraestrutura básica		+++	+++	+++
porcent. escolas com diretoria		+		+
porcent. escolas públicas			-	
desp_educ_cult_hab	+++	+++	++	++
desp_saude_hab		+++	+++	++
imunização			+	

Quadro 4.2 – Sinal e significância dos determinantes espaciais - frequência escolar

Em relação ao grupo das variáveis socioeconômicas, algumas delas merecem destaque na análise, por terem seus sinais alterados ou por terem adquirido significância estatística.

A primeira delas é a renda per capita. Ao se controlar o efeito espacial, o aumento da renda per capita aumenta a frequência escolar em todos os modelos espaciais. Este resultado corrobora os resultados encontrados por Pontili e Kassouf (2007) e Vasconcellos (2003), que encontraram efeito positivo e significativo para a variável renda em relação à probabilidade de as famílias matricularem seus filhos na escola. A outra refere-se ao índice de Gini, cujo sinal passa a ser o esperado: quanto maior a concentração de renda, menor será a taxa de frequência escolar. E, por fim, as outras duas variáveis dizem respeito à estrutura etária da população. A maior porcentagem da população adulta de 15 a 29 anos e da população de 55 anos ou mais estão associadas, individualmente, a maiores taxas de frequência escolar da população de 6 a 14 anos. Nesse caso, pode ser que os fatores que explicam a transição demográfica, tais como a maior expectativa de vida, o planejamento familiar, as quedas das taxas de fertilidade e o próprio aumento dos anos de escolaridade da população, impactem também o indicador de frequência escolar.

As variáveis relativas ao bolsa família, à taxa de urbanização e à taxa de analfabetismo da população de 25 anos ou mais apresentaram resultados esperados para todos os modelos estimados. A primeira deve-se às condicionalidades do programa Bolsa-Família; a segunda, às evidências de que em áreas urbanas a frequência escolar é maior do que em áreas rurais e a

terceira, ao fato de que quanto mais analfabeta a população nessa faixa etária, principalmente em relação à escolaridade do chefe de família, menor é a frequência escolar.

O fato da maior porcentagem da população negra e parda estar associada a maiores indicadores de frequência escolar pode ser reflexo das políticas de inclusão ocorridas nos últimos anos. Tal resultado pode evidenciar um fenômeno recente, pois Pontili e Kassouf (2007) encontraram que os negros têm menor probabilidade de frequentar a escola que os brancos, ocorrendo o mesmo com os pardos da área rural de Pernambuco e urbana de São Paulo. Vasconcellos (2003) também encontrou resultados similares, indicando que as diferenças inter-raciais no Brasil se estendem à área de educação.

Quanto maior o número médio de horas-aula, o índice de infraestrutura escolar básica e a porcentagem de escolas com diretoria, maior é a taxa de frequência escolar da população na faixa etária de 6 a 14 anos. Além do melhor desempenho escolar obtido pelos alunos que passam entre quatro e cinco horas ou mais de cinco horas na sala de aula do que aqueles que ficam menos de quatro horas (MENEZES-FILHO, 2007), esse estudo mostra que o maior tempo de aula, por aumentar o tempo que o aluno permanece na escola, também tem um efeito na taxa de frequência escolar.

Em consonância com Pontili e Kassouf (2007), os resultados indicam que a maior proporção de escolas com biblioteca, que faz parte do índice de infraestrutura escolar básica formulado nesse estudo, afeta de maneira positiva a frequência de alunos na escola. E, por fim, a variável *proxy* de gestão escolar, a porcentagem de escolas com diretoria, também apresentou efeito positivo em frequência escolar.

A variável de despesa com educação e cultura se mostrou relevante para incrementar positivamente o nível de frequência escolar. No mesmo sentido, a variável de gastos com saúde também se mostrou importante. Ou seja, o aumento da despesa com saúde gera condições mais propícias para que os alunos frequentem as escolas. Nesse sentido, essa evidência pode contribuir para a explicação da hipótese do próximo capítulo, de que o motivo doença, provocado pelas inadequadas condições de saneamento, afetam os indicadores educacionais.

No grupo de variáveis de controle educacionais, o maior número de docentes com ensino superior e o maior índice de infraestrutura tecnológica diminui a frequência escolar de forma robusta e significativa. Em relação à última variável, de acordo com Menezes-Filho (2007), há muita discussão acerca da inclusão digital, pois os impactos da presença de computadores para alunos, diretores e professores são dúbios sobre o desempenho dos alunos.

O outro resultado contra intuitivo é aquele relativo ao impacto negativo da variável de porcentagem de docentes com curso superior sobre a frequência escolar. Uma das explicações pode ser que a decisão dos alunos frequentarem e permanecerem na escola esteja mais relacionada às características familiares do que a questões relacionadas à didática e motivação do professor em mantê-lo na mesma. Ou ainda, como ressalta Hanushek e Rivkin (2006), os anos de escolaridade, experiência e capacitação muitas vezes não traduzem a qualidade do professor, no sentido de que podem existir características que são não observáveis nos dados. A Tabela 4.11 apresenta os resultados das estimações para os mesmos modelos espaciais em relação à taxa de distorção idade-série da população de 6 a 14 anos do ensino fundamental.

Tabela 4.11 – Determinantes espaciais da taxa de distorção idade-série (2000-2010)				
Variável Dependente	Taxa de Distorção Idade-Série Ensino Fundamental			
	EF Não Espacial	EF Lag	EF Erro	EF Lag e Erro
	(1)	(2)	(3)	(4)
Saneamento	-0,965*** (0,0847)	-0,2065*** (0,0312)	-0,3114*** (0,0589)	-0,2000** (0,0296)
Observações	7.959	7.224	7.224	7.224
Número de municípios	4.347	3.612	3.612	3.612
Características socioeconômicas	SIM	SIM	SIM	SIM
Características da escola e professores	SIM	SIM	SIM	SIM
Saúde e educação	SIM	SIM	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM
Efeito fixo de ano	NÃO	NÃO	NÃO	NÃO
prv5anos - Variável Instrumental	SIM	SIM	SIM	SIM
Lambda	-	0,0806*** (0,0166)	-	0,0842*** (0,0162)
Rho	-	-	0,1269	0,03
Sigma^2	-	-	-	56,5181
Erro-padrão entre parênteses				
*** p<0,01; ** p<0,05; * p<0,1				

No que diz respeito à taxa de distorção idade-série, os resultados também se mantiveram e indicaram que o maior acesso às condições adequadas de saneamento básico reduz a distorção idade-série. Além do acesso a esse serviço essencial manter as crianças na escola, ainda fornece melhores condições cognitivas, por estarem livres das doenças relacionadas ao

saneamento inadequado (DRSAI), de forma que sejam aprovadas e não sejam retidas em séries inadequadas à sua idade.

Neste caso, todos os modelos espaciais apresentaram coeficientes de menor magnitude em relação ao modelo não espacial da primeira coluna. No Quadro 4.3 estão dispostos os sinais e significâncias dos efeitos das variáveis de controle.

Variáveis Explicativas	IV	EF	NE	EF Lag	EF Erro	EF Lag e Erro
t_analfabetismo_25m	+++			+++	+++	+++
pop_negra_parda	---			+++	+++	+++
renda_per_capita	-			---	-	-
gini	+++			+++	+++	+++
pop_55mais	+++			+++	+++	+++
pop_15a29	+++			+++	+++	+++
bolsafam	+++			---	---	---
poptot	--					
urbanização	+++			--	-	-
alunos_turma_fundamental	+++			+++	+++	+++
docentes_com_ens_superior	+++			---	---	---
h_a_total_fundamental	---			---	---	---
índice infraestrutura tecnológica	---			---	---	---
índice infraestrutura básica	+++			---	--	--
porcent. escolas com diretoria				+++	+++	+++
porcent. escolas públicas	+++			++	-	-
desp_educ_cult_hab						
desp_saude_hab				+	-	+
imunização						

Quadro 4.3 – Sinal e significância determinantes espaciais da distorção idade-série

Para a taxa de distorção idade-série (TDI), o coeficiente da renda per capita, o índice de Gini e a taxa de analfabetismo não apresentaram reversão de sinais. No entanto, em relação à variável étnica referente à porcentagem da população negra e parda, quanto maior a porcentagem dessa população maior a taxa de distorção idade-série, revelando que as diferenças étnicas-raciais são relevantes para explicar o atraso escolar. Além disso, o maior número de famílias beneficiadas pelo programa Bolsa Família indicou menor taxa de distorção idade-série, ou seja, além das crianças frequentarem mais as escolas, o atraso escolar também diminui. Ainda que essa evidência não implique em qualidade do ensino, tais resultados revelam o alcance do programa em termos educacionais.

A taxa de urbanização também teve seu sinal alterado: de fato, regiões mais urbanizadas tendem a ter melhores indicadores educacionais que, nesse caso, refletem-se em menores taxas de distorção idade-série. No tocante à estrutura etária da população, conforme a porcentagem da população adulta e da população de 55 anos ou mais aumenta, a taxa de

distorção idade-série aumenta também, pois a probabilidade de atraso escolar é maior conforme aumenta a idade do estudante (PONTILI e KASSOUF, 2007).

A maior razão de alunos por turma leva a menor taxa de frequência escolar, para todos os modelos, indicando que turmas muito numerosas podem captar a deficiência do aprendizado individual, que levaria a reprovações e aumentos no atraso escolar. De forma contrária ao resultado encontrado para a taxa de frequência escolar, quando se considera a taxa de distorção idade-série, a porcentagem de docentes com ensino superior é importante para diminuir o atraso escolar entre as séries no ensino fundamental. Além disso, o índice de infraestrutura escolar tecnológica, o índice de infraestrutura escolar básica e o número médio de horas-aula também se mostraram fatores importantes para diminuir essa variável.

Neste caso, o resultado que não era esperado diz respeito à variável que representa a gestão da escola (porcentagem de escolas com diretoria): esta contribui de forma a aumentar o atraso dos alunos de ensino fundamental. O efeito da porcentagem de escolas públicas e das despesas com saúde por habitante, por terem apresentado sinais inconclusivos entre os três modelos, não foram consideradas na análise. Em linha com o formato das estimações espaciais das duas variáveis dependentes anteriores, a Tabela 4.12 mostra os resultados das estimações para a variável referente à taxa de abandono escolar da mesma população do ensino fundamental.

Tabela 4.12 – Determinantes espaciais da taxa de abandono escolar (2000-2010)

Variável Dependente	Taxa de abandono escolar - ensino fundamental			
	EF Não Espacial	EF Lag	EF Erro	EF Lag e Erro
	(1)	(2)	(3)	(4)
Saneamento	-0,499*** (0,0528)	-0,0563*** (0,0148)	-0,1223*** (0,0304)	-0,0497** (0,0142)
Observações	7.959	7.224	7.224	7.224
Número de municípios	4.347	3.612	3.612	3.612
Características socioeconômicas	SIM	SIM	SIM	SIM
Características da escola e professores	SIM	SIM	SIM	SIM
Saúde e educação	SIM	SIM	SIM	SIM
Efeito fixo de município	SIM	SIM	SIM	SIM
Efeito fixo de ano	NÃO	NÃO	NÃO	NÃO
priv5anos - Variável Instrumental	SIM	SIM	SIM	SIM
Lambda	-	-0,0154 (0,0247)	-	-0,0108 (0,0243)
Rho	-	-	0,1269	0,0489
Sigma^2	-	-	-	14,9124
Erro-padrão entre parênteses				
*** p<0,01; ** p<0,05; * p<0,1				

Novamente, a variável relativa ao maior acesso aos serviços de saneamento básico adequados impacta de forma negativa a taxa de abandono escolar. Ou seja, trata-se de resultado positivo em termos da melhora dos indicadores educacionais, haja vista os resultados também favoráveis obtidos anteriormente. Nota-se que, para essa variável, a magnitude dos coeficientes diminuiu de forma mais intensa. Ou seja, no modelo de Efeito Fixo não espacial o valor era de -0,49 e, com a introdução dos efeitos espaciais, esse valor passou para -0,056 e -0,049.

O abandono escolar está associado com altas taxas de reprovação e, consequentemente, com o atraso escolar e com menores ganhos no mercado de trabalho (EIDE e SHOWALTER, 2001). Por outro lado, Jacob e Lefgren (2009) sugerem que a reprovação no ensino fundamental não afeta a probabilidade do aluno concluir o ensino médio, porém a reprovação associada ao baixo rendimento no último ano do ensino fundamental aumenta substancialmente a probabilidade de que esses alunos abandonem os estudos no ensino médio. O fato dos municípios que apresentam adequadas condições de saneamento básico também apresentarem menores taxas de abandono escolar pode significar que, uma vez garantidas melhores condições de saúde às crianças, essas terão melhor rendimento escolar, o que pode implicar em menores taxas de abandono escolar.

O Quadro 4.4 apresenta os resultados referentes aos sinais e níveis de significância das variáveis explicativas após o controle pelos efeitos espaciais.

Variáveis Explicativas	IV	EF	NE	EF Lag	EF Erro	EF Lag e Erro
t_analfabetismo_25m	+++		+++	+++	+++	+++
pop_negra_parda	---		+++	++	+++	+++
renda_per_capita	---		---	---	---	---
gini	+++		+++	+++	+++	+++
pop_55mais	+++					
pop_15a29	+++		+++	+	+++	+++
bolsafam	+++		---	---	---	---
poptot	---					
urbanização	+++		+++	+++	+++	+++
alunos_turma_fundamental	+		+++	+++	+++	+++
docentes_com_ens_superior	+		---	---	---	---
h_a_total_fundamental	-		---	---	---	---
índice infraestrutura tecnológica	---		---	---	---	---
índice infraestrutura básica	+++		--			-
porcent. escolas com diretoria			+++	+++	+++	++
porcent. escolas públicas	+++		+	+		
desp_educ_cult_hab						
desp_saude_hab						
imunização						

Quadro 4.4 – Sinal e significância dos determinantes espaciais do abandono escolar

Os resultados dos sinais e significância das variáveis explicativas para a variável dependente taxa de abandono escolar foram semelhantes à análise para a variável anterior (taxa de distorção idade-série), mas com duas exceções. Em relação à primeira delas, a maior taxa de urbanização gerou maior abandono. A segunda faz parte do vetor de características educacionais. Segundo resultado expresso no Quadro 4.4 acima, a maior porcentagem de escolas públicas mostrou-se favorável para o registro de maiores taxas de abandono escolar. Além disso, cabe observar que nenhum tipo de despesa com saúde, educação, cultura e nem a cobertura de imunização afetam estatisticamente as taxas de abandono escolar do ensino fundamental.

4.5 Comparação com o método de Máxima Verossimilhança.

Millo e Piras (2012), a partir de um exemplo de modelo sem endogeneidade, compararam os coeficientes estimados por Máxima Verossimilhança (MV) e pelo método generalizado dos momentos (GMM). Segundo os autores, nesse caso, os parâmetros ficaram bem próximos, ressaltando pouca diferença entre eles. Espera-se que, pelo controle da endogeneidade, as estimativas dos coeficientes sejam diferentes nesse estudo. Na Tabela 4.13 são comparados o

modelo SAR de efeitos fixos estimado por GMM, com a identificação da variável endógena entre as explicativas além da variável dependente defasada, com o modelo SAR de efeitos fixos estimado por MV. As outras duas comparações, terceira e quarta colunas, referem-se aos modelos SAC de efeitos fixos estimados por GMM e MV para a taxa de distorção idade-série (TDI) por GMM. A mesma comparação foi feita para as duas últimas colunas considerando a taxa de abandono escolar.

Tabela 4.13 – Comparação entre os modelos estimados por GMM e MV

Variável Dependente	Frequência escolar (GMM)	Frequência escolar (MV)	TDI (GMM)	TDI (MV)	Abandono Escolar (GMM)	Abandono Escolar (MV)
	(1)	(2)	(3)	(4)	(5)	(6)
Saneamento	0,0351*** (0,0130)	0,0241*** (0,0046)	-0,20*** (0,0296)	-0,23*** (0,0097)	-0,049*** (0,0142)	-0,058*** (0,0062)
Observações	7.959	7.959	7.224	7.959		7.224
Número de municípios	4.347	4.347	3.612	4.347		3.612
Características socioeconômicas	SIM	SIM	SIM	SIM		SIM
Características escola e professores	SIM	SIM	SIM	SIM		SIM
Saúde e educação	SIM	SIM	SIM	SIM		SIM
Efeito fixo de município	SIM	SIM	SIM	SIM		SIM
Efeito fixo de ano	NÃO	NÃO	NÃO	NÃO		NÃO
priv5anos - Variável Instrumental	SIM	NÃO	SIM	NÃO		SIM
Lambda	0,1428*** (0,0321)	0,0180 (0,0125)	0,0842*** (0,0162)	-0,007 (0,0118)	-0,0108 (0,0243)	0,019 (0,0147)
Rho	-	-	0,03	0,011 (0,021)	0,0489	0,0053 (0,023)

Erro-padrão entre parênteses

*** p<0,01; ** p<0,05; * p<0,1

As estimativas dos coeficientes espaciais se mostraram diferentes, como era esperado, pois na estimação por MV tais coeficientes não foram significativos. Com exceção da variável de frequência escolar, a magnitude dos coeficientes estimados por MV foram ligeiramente maiores.

5 EFEITO DO SANEAMENTO SOBRE A SAÚDE

O capítulo anterior apresentou evidências robustas de que o saneamento importa para a educação. A análise não espacial apontou que o melhor modelo para análise de dados em painel é o de efeitos fixos, detectou a presença de endogeneidade nos dados e verificou a validade do instrumento utilizado na estimação em dois estágios. O teste dos resíduos para os modelos das três variáveis de educação detectou a presença de efeitos espaciais, que foram controlados pelo método GMM para dados em painel. Não controlar esse efeito poderia ter dado origem à viés e inconsistência nas estimativas do modelo convencional.

Para validar esse efeito, encontrado no capítulo 4, é necessário que a associação entre saneamento e saúde seja também comprovada de forma empírica, objetivo ao qual se dedica esse capítulo. Desta forma, em primeiro lugar, será realizado o desenvolvimento do arcabouço teórico do modelo de Grosman (1972), utilizado para definir as variáveis utilizadas nas estimações subsequentes bem como identificar os canais de transmissão entre elas e a variável de interesse. Em seguida, realizou-se revisão bibliográfica acerca das relações entre saneamento, mortalidade e morbidade. Nas seções posteriores, foram realizadas a Análise Exploratória de Dados Espaciais, estabelecidas as hipóteses acerca dos possíveis efeitos espaciais, a escolha entre os diversos modelos, a descrição sobre as metodologias utilizadas e a discussão dos resultados. O modelo escolhido, de acordo com a primeira metodologia, trata-se de um Modelo Durbin Espacial (SDM) com matriz contígua de ponderação espacial.

Uma característica importante da amostra utilizada nesse capítulo deve-se a quantidade de zeros na variável dependente. Tais zeros referem-se aos distritos saudáveis que não tiveram ocorrências de internações por DRSAI no ano de 2010. O controle da censura dos dados foi realizado pela estimação do Modelo SDM Tobit Espacial de acordo com a abordagem Bayesiana.

Por meio desta estratégia foi possível comprovar a associação entre saneamento e saúde. Por fim, na última seção foi realizada uma recomendação de política pública que poupa recursos no setor da saúde e que poderiam ser investidos no setor de saneamento para acelerar o cumprimento das metas de universalização.

5.1 O modelo teórico de Grossman (1972).

A grande questão que permeou os estudos da economia da saúde desde a segunda metade do século passado refere-se a como os cuidados com a saúde diferem entre os indivíduos e quais as falhas de mercado que surgem em razão dessa diversidade. As características dos cuidados de saúde apresentam quatro aspectos fundamentais ressaltados por diversos autores. O primeiro deles, de acordo com Grossman (1972), é que a procura por cuidados médicos é derivada da demanda por saúde, que é fundamental para os indivíduos participarem do mercado de trabalho e auferirem rendimentos no futuro. Em outras palavras, os bens e serviços apropriados para o cuidado da saúde (medicamentos, exames clínicos e procedimentos médicos) são considerados fatores de produção necessários para o indivíduo produzir saúde e, portanto, atender a essa demanda.

O segundo aspecto dos cuidados ou serviços de saúde diz respeito às externalidades geradas pelos mesmos. Tais externalidades ocorrem porque serviços de saúde não são somente consumidos, mas trazem também características de bens de investimento (MUSHKIN, 1962). O retorno desse investimento pode gerar benefícios para o próprio indivíduo e para os outros que os circundam. Exemplos disso são a vacinação e a oferta de serviços de saneamento básico. Neste último caso, a provisão de tais serviços, na maior parte dos países do mundo, é fornecida exclusivamente pelo Estado ou com grande participação deste, pelo caráter social do investimento.

A terceira e quarta características, apesar de não serem abordadas neste estudo, dizem respeito à assimetria de informações entre os prestadores de serviços médicos e os pacientes e à incerteza em relação à necessidade e eficácia dos cuidados de saúde. A saúde é considerada por Grossman (1972) como um bem de consumo, pois entra diretamente na função utilidade e proporciona, por si só, bem-estar aos indivíduos. Além disso, pode ser vista também como um bem de investimento, porque as pessoas demandam saúde com o intuito de ganhar renda no futuro ao participarem do mercado de trabalho. Por último, a saúde é também bem de capital, pelo fato de a mesma durar mais que um período de tempo e constituir-se em estoque acumulado, que se deprecia com o aumento da idade. Ou seja, o nível de saúde acumulado caracteriza-se como uma variável de estoque enquanto que investimentos em saúde, em determinados períodos de tempo, podem ser entendidos como variáveis de fluxo. Outra

contribuição de Grosman (1972) foi a separação artificial entre consumo e produção, ao estabelecer que o indivíduo é também produtor de saúde e não somente consumidor.

A partir de uma função utilidade intertemporal de um consumidor típico, expressa na equação 5.1., o autor demonstra como os indivíduos escolhem a quantidade ótima de saúde:

$$(5.1) \quad U = U(\phi_0 H_0, \dots, \phi_n H_n; Z_0, \dots, Z_n)$$

Nessa função, H_0 é o estoque de saúde inicial herdado pelo indivíduo, H_i é o estoque de saúde em qualquer período de tempo, ϕ_i é o fluxo de serviços de saúde por unidade de estoque de saúde, em qualquer período de tempo, $\phi_i H_i$ é o consumo total de serviços de saúde, Z_i é consumo total de outras mercadorias não relacionadas com a saúde e $i: 1, \dots, n = i$ são os períodos de tempo.

Portanto, a utilidade depende da quantidade de saúde ótima acumulada até determinado período de tempo e da quantidade de outras mercadorias que não a saúde. O investimento bruto em saúde, ou a função de produção de saúde (I_i), é obtido de forma similar a um processo de acumulação de capital, conforme equação 5.2 abaixo:

$$(5.2) \quad H_{i+1} - H_i = I_i - \delta_i H_i$$

Nesta equação, δ_i é a depreciação do estoque de saúde em qualquer período do tempo. A função de produção de saúde pode assumir o seguinte formato:

$$(5.3) \quad I_i = I_i(M_i, TH_i; E_i, A_i, ST_i, SOC_i)$$

As variáveis dessa função de produção são resultado do processo de escolha individual. A primeira delas diz respeito ao consumo de bens e serviços médicos apropriados, denominados cuidados médicos (M_i) e a segunda é referente ao tempo e ao esforço que cada pessoa empreende para melhorar a saúde (TH_i). A variável idade (A_i) demonstra que a idade do indivíduo pode afetar a capacidade de produção de saúde. A saúde de pessoas idosas, por se deteriorar mais rapidamente em relação à saúde das mais jovens, requer um volume maior de cuidados médicos. A variável educação (E_i) evidencia que pessoas educadas produzem

investimento em saúde de forma mais eficiente. Ou seja, compreendem a importância de comportamentos não saudáveis, conhecem atitudes de prevenção em relação a determinadas doenças, possuem melhor capacidade de se comunicar com profissionais de saúde e, por se cuidarem mais, conseguem detectar precocemente diagnósticos que tenham maiores chances de cura.

Berndt *et al.* (2000) ressaltam que a função de produção de saúde apresenta uma série de argumentos que podem ser interdependentes entre si. Um dos elementos que os autores relatam afetar a saúde refere-se ao ambiente. As mudanças ambientais (poluição do ar, água, clima, entre outros) podem melhorar ou retardar a saúde e, portanto, devem ser contempladas nesta função.

Em consonância com o estudo de Sousa e Leite Filho (2008), que introduziram variáveis relativas aos serviços de saneamento e saúde pública para explicar o *status* de saúde de uma população, neste trabalho serão introduzidas variáveis de saneamento na função de produção de saúde. O acesso aos diversos tipos de serviços de saneamento (água, esgoto, lixo e drenagem de águas pluviais) será considerado variável exógena importante na função de produção de saúde que, nesse estudo, é medida pela taxa de incidência de internações por Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI). Dessa forma, a variável que será adicionada na função de produção é denominada “tipo do serviço de saneamento básico (ST_i)” no período de tempo i . A outra variável adicionada diz respeito ao vetor de características socioeconômicas (SOC_i) dos distritos. Entre essas variáveis, Grossman (1972) resalta a renda ou salário do agente como um fator que afeta a produção de saúde.

Dado que o tempo gasto com saúde é o tempo gasto adquirindo serviços médicos apropriados, $TH_i = t_i M_i$, e que a função de produção domiciliar é homogênea de grau 1,³³ a equação (5.3) pode ser reescrita como:

$$(5.4) \quad I_i = M_i g(t_i; E_i, A_i, ST_i, SOC_i)$$

$$(5.5) \quad \frac{\partial I_i}{\partial TH_i} = \frac{\partial g}{\partial t_i} = g' > 0; \quad \frac{\partial I_i}{\partial M_i} = g - t_i g' > 0$$

³³ Uma função $g(M_i t_i; E_i, A_i, ST_i, SOC_i)$ é homogênea de grau k se $g(M_i t_i; E_i, A_i, ST_i, SOC_i) = M_i^k g(t_i; E_i, A_i, ST_i, SOC_i)$. Logo, o pressuposto da homogeneidade de grau 1 garante que $g(M_i t_i; E_i, A_i, ST_i, SOC_i) = M_i g(t_i; E_i, A_i, ST_i, SOC_i)$.

Uma das implicações do modelo é que os produtos marginais do tempo gasto com saúde e dos cuidados médicos são ambos positivos, pois o aumento de ambos leva a incrementos adicionais positivos do investimento em saúde.

O outro componente da função utilidade que fornece bem-estar para o indivíduo é a produção de um bem de consumo puro (Z_i). Para produzi-lo, o indivíduo também utiliza tempo (T_i) e outros bens, que adquire no mercado (X_i), de acordo com a equação abaixo:

$$(5.6) \quad Z_i = Z_i(X_i, T_i; E_i)$$

Os indivíduos enfrentam um *trade-off* em relação ao tempo T_i , que pode ser utilizado em horas de trabalho para obter renda (TW_i) e/ou lazer (TL_i). Logo, o indivíduo aloca seu tempo total (Ω) em horas para o trabalho (TW_i), lazer ou produção de bens domésticos (TL_i), produção de saúde (TH_i) ou perda de uma parcela de tempo em decorrência de doenças ou falta de saúde (TD_i), conforme equação abaixo:

$$(5.7) \quad \Omega = TW_i + TL_i + TH_i + TD_i$$

A restrição orçamentária de bens iguala o valor presente dos gastos com bens ao valor presente dos rendimentos auferidos ao longo do ciclo de vida mais os ativos iniciais conforme equação 5.8:

$$(5.8) \quad \sum \frac{P_i M_i + F_i X_i}{(1+r)^i} = \sum \frac{W_i TW_i}{(1+r)^i} + A_0$$

Na equação acima, P_i e F_i são os preços de M_i e X_i respectivamente, W_i é a taxa salarial e r é a taxa de juros. Dado que $\Omega = 365$ dias, h_i é definido como o número de dias saudáveis em um dado ano:

$$(5.9) \quad h_i = \Omega - TD_i$$

A relação entre o estoque de saúde (H_i) e o número de dias saudáveis (h_i) é denominada função de produção de dias saudáveis:

Figura 5.1 – Função de produção de dias saudáveis

Fonte: Grosman (1972a)

Ao substituir TW_i da equação (5.7) na equação (5.8) obtêm-se a restrição de riqueza completa, que ocorre quando gasta todo seu tempo em trabalho, descrita abaixo:

$$(5.10) \quad \sum \frac{P_i M_i + F_i X_i + W_i (TH_i + TL_i + TD_i)}{(1+r)^i} = \sum \frac{W_i \Omega}{(1+r)^i} + A_0$$

As quantidades de equilíbrio de H_i e Z_i são encontradas pela maximização da utilidade familiar, equação (5.1), sujeito às restrições dadas pelas equações (5.2), (5.3), (5.6) e (5.10). As condições de primeira ordem deste problema de maximização desembocam na seguinte relação:

$$(5.11) \quad G_i \left[W_i + \left(\frac{U_{h_i}}{\lambda} \right) (1+r)^i \right] = \pi_{i-1} (r - \tilde{\pi}_{i-1} + \delta_i)$$

Na equação acima, $\pi_{i-1} = \frac{\partial H_n}{\partial I_{i-1}}$ é o custo marginal do investimento bruto em saúde no período $i-1$, $\tilde{\pi}_{i-1}$ é a variação percentual de uma mudança no custo marginal entre o período $i-1$ e o período i , U_{h_i} é a utilidade marginal de dias saudáveis, λ é a utilidade marginal da riqueza, $G_i = \frac{\partial h_i}{\partial H_i} = -\frac{\partial T}{\partial H_i}$ é o produto marginal do estoque de saúde na produção de dias saudáveis. Ao dividir ambos os lados dessa equação pelo custo marginal do investimento bruto no período $i-1$ obtêm-se:

$$(5.12) \quad \gamma_i + a_i = r - \tilde{\pi}_{i-1} + \delta_i$$

Em que, $\gamma_i = \frac{W_i G_i}{\pi_{i-1}}$ é a taxa marginal em valores monetários do retorno do investimento unitário em saúde e $a_i = \frac{\left[\left(\frac{U_{h_i}}{\lambda} \right) (1+r)^i (G_i) \right]}{\pi_{i-1}}$ é a taxa psicológica do retorno de uma unidade de investimento em saúde³⁴.

Quando a aquisição de saúde é considerada como um bem puro de investimento tem-se que $U_{h_i} = 0$ e, conseqüentemente, $a_i = 0$. Nesse caso, os custos de capital de saúde são os custos de se abdicar de tempo e dinheiro para se obter saúde. Desde que o estoque de saúde inicial herdado pelo indivíduo e as taxas de depreciação são dadas, a quantidade ótima de investimento bruto em saúde (π_{i-1}^*) determina a quantidade ótima de capital de saúde (H^*).

A relação desse custo com o estoque de saúde do indivíduo descreve o padrão das taxas de retorno do investimento em saúde, ou seja, a eficiência marginal do investimento em saúde (EMH), que exibe retornos marginais decrescentes para a função de produção de dias saudáveis. A curva S, que mostra a relação entre o estoque de saúde e o custo de capital, é infinitamente elástica. O ponto ótimo (A) de demanda de saúde é obtido pela interseção da curva EMH com a curva S.

Figura 5.2 – A eficiência marginal do investimento em saúde (EMH)

Fonte: Grosman (1972a)

À medida em que a educação, o acesso aos serviços adequados de saneamento básico e as variáveis socioeconômicas (tal como a renda e escolaridade) aumentam, a curva EMH se

³⁴ No artigo original, Grossman (1972) denomina essa taxa de “*psychic rate of return on an investment in health*”.

desloca para a direita afetando o capital de saúde do indivíduo. Esse deslocamento pode se dar pelo fato das pessoas utilizarem os insumos de produção de forma mais eficiente (educação), devido à presença de externalidades ambientais que reforçam os retornos obtidos dos dias saudáveis (saneamento) ou em decorrência do investimento em novos insumos de saúde (variáveis socioeconômicas). No caso da renda, duas situações podem ocorrer: o maior valor da hora do indivíduo no mercado de trabalho o leva a desejar não perder tempo com a falta de saúde e, portanto, investe na mesma. De forma alternativa, por trocar um pouco do seu rendimento por lazer terá mais tempo para investir em saúde que, conseqüentemente, reforça o primeiro efeito e ainda lhe causa bem-estar.

Por outro lado, com o avanço da idade há um deslocamento para a esquerda da curva EMH, pois os retornos dos investimentos perdurarão por um período mais curto de tempo, de forma que a saúde ótima cairá enquanto a pessoa envelhecer até que $H^* < H_{MIN}$. Apesar desse modelo ter sido formulado em nível individual e a partir de uma função de utilidade familiar, pode ser adaptado para analisar o estado de saúde de uma população, se forem agregadas tais funções individuais. Logo, a escolha das variáveis utilizadas nesse estudo terá como respaldo o arcabouço teórico acima descrito.

5.2 Os impactos do saneamento básico sobre a saúde dos indivíduos

A importância do saneamento básico e da sua associação com a saúde remonta aos tempos mais antigos. Da civilização greco-romana aos dias atuais, a vida em comunidade pode ocasionar condições propícias à disseminação de doenças. Diversos tipos de bactérias, vírus e parasitas estão presentes nas fezes, urina ou vômitos do indivíduo doente ou portador. A depender da forma como a matéria fecal e outros rejeitos humanos chegam ao meio ambiente, tais agentes patogênicos podem ser transmitidos para seres humanos saudáveis com muita facilidade. Logo, o abastecimento de água encanada e segura, o adequado sistema de coleta e tratamento das águas servidas, o manejo dos resíduos sólidos e a drenagem de águas pluviais são fatores essenciais que, aliados às práticas higiênicas, impedem a propagação de doenças.

5.2.1 Saneamento Básico e Mortalidade

A literatura da saúde reconhece evidências da relação entre a situação dos serviços de saneamento básico e indicadores epidemiológicos de uma localidade. Em particular, as crianças são as mais atingidas e, pela fragilidade dos seus organismos, em grande parte dos casos podem vir a óbito. A associação entre saneamento e mortalidade foi constatada e é bem consolidada em diversos países. A diferença entre eles está no período de tempo em que as altas taxas para este indicador de vulnerabilidade social foram superadas: enquanto que para alguns países desenvolvidos esse é um problema do século passado (como para os Estados Unidos, por exemplo), para outros (tais como Índia, Argentina e Brasil) é um desafio da realidade atual.

Para o primeiro grupo de países, Fair et al (1966) mostraram a similaridade e tendência histórica entre duas curvas, no período de 1885 a 1940 nos Estados Unidos: a curva que se referia ao decréscimo da mortalidade por febre tifoide e a que tratava da diminuição da parcela populacional sem acesso ao abastecimento de água. Ainda no contexto norte-americano, Cutler e Miller (2005) mostraram que a introdução da tecnologia de água limpa, iniciada por volta de 1900, reduziu a taxa de mortalidade infantil por volta de 46% e foi responsável por 74% do total da queda que ocorreu no país ao longo do século. Watson (2006), ao estudar os efeitos de programas sanitários na década de 1960, mostrou que o aumento de 10 pontos percentuais nos serviços de saneamento de residências de reservas indígenas americanas levou a uma redução da taxa de mortalidade infantil de 2,5%. Além disso, a autora encontrou que intervenções de saneamento explicaram quase 40% da convergência das taxas de mortalidade infantil em tais reservas desde 1970.

Os dados são mais recentes quando se consideram os países em desenvolvimento: dos nove milhões de crianças que morrem no mundo todos os anos antes do seu quinto aniversário, a maioria é formada por crianças pobres do Sul da Ásia e da África Subsaariana e por volta de uma em cada cinco morre de diarreia (BANERJEE; DUFLO, 2011). Na Argentina, Galiani *et al* (2005) apresentaram evidências empíricas de que a privatização do setor de saneamento básico, ao aumentar o acesso da população aos serviços e melhorar a qualidade dos mesmos, reduziu a taxa de mortalidade infantil por todas as causas de doenças em torno de 5% a 7%. Nas áreas mais pobres do país esse valor atingiu 26%. Na Índia, mais da metade da população (600 milhões de pessoas) considera "impuro" o uso de banheiros e fossas e defende que

defecar a céu aberto faz parte de um estilo de vida. Geruso e Spears (2015) encontraram evidências do impacto causal da prática de defecação a céu aberto nesse país sobre a mortalidade infantil. Os autores utilizaram a variação que surge da composição religiosa dos vizinhos como instrumento para identificar esse efeito. Segundo eles, reduzir em média essa prática em 10 pontos percentuais reduziria a taxa de mortalidade infantil média por volta de 4%. No contexto indiano, no qual nascem 26 milhões de crianças por ano, esse resultado equivale a 78.000 mortes por ano. O estudo de tais autores destaca que a magnitude desse impacto revela a necessidade de investigações acerca das externalidades da falta de saneamento nos países em desenvolvimento.

Para a realidade brasileira, Alves e Belluzzo (2004) constataram, ao estimar um modelo de dados em painel estático e outro dinâmico (para os anos de 1970 a 2000), que a mortalidade infantil no Brasil vem diminuindo, mas que essa taxa continua alta quando comparada a outros países e apresenta desigualdades entre os estados e regiões brasileiras, que ficam ainda mais visíveis quando comparadas em termos municipais. Os autores concluíram que o aumento de uma unidade adicional de escolaridade reduziu a mortalidade infantil por volta de 7%. No entanto, a proporção de residências com saneamento básico no município teve impacto muito pequeno na redução da taxa de mortalidade infantil, comparado ao efeito da variável de educação. Barufi *et al* (2012) mostraram que a infraestrutura de saúde perdeu importância na explicação do padrão de mortalidade infantil ao longo do período de 1980 a 2000. Os autores, ao utilizarem o modelo de determinantes proximais, verificaram a influência dos vizinhos na análise e destacaram que as variáveis socioeconômicas, em especial o maior acesso aos serviços de saneamento básico, foram variáveis relevantes e significativas para explicar a redução da mortalidade infantil nesse período.

Mendonça e Motta (2007), por meio do uso da metodologia de dados em painel com efeito fixo para o período 1981-2001, a partir de dados agregados por estado, demonstraram que a redução da mortalidade infantil associada às doenças de veiculação hídrica foi alcançada com a melhoria na cobertura dos serviços de saneamento (queda de 1,29 a 1,33) e também devido ao acesso aos serviços de educação (queda de 0,26) e saúde (queda de 0,0001). Além disso, mostraram que o baixo acesso a rede de esgoto e a água tratada é uma das principais causas da mortalidade na infância.

Os resultados de Sousa e Leite Filho (2008), que analisaram o impacto do saneamento sobre a mortalidade infantil na região Nordeste, estão em consonância com as conclusões anteriores: por meio de metodologia de dados em painel com informações agregadas para municípios, concluíram que a mortalidade de menores de um ano de idade nesta região foi reduzida em 31,8% no período de 1991 a 2000 para os estados que apresentaram maiores ganhos no acesso à água tratada.

O resultado desse estudo é corroborado pelo de Gamper-Rabindran et al. (2010). Para estes últimos, intervenções de abastecimento de água encanada em áreas que se localizam nos quantis superiores da distribuição condicional da taxa de mortalidade infantil, quando acompanhadas de outros insumos básicos de saúde pública, podem alcançar maiores reduções na taxa de mortalidade infantil.

Rocha e Soares (2012), sob outro ponto de vista, também demonstraram o efeito do saneamento sobre a taxa de mortalidade infantil. Os autores encontraram que o aumento da cobertura de saneamento básico reduz a resposta da mortalidade infantil às flutuações de chuvas na região do semiárido nordestino brasileiro. A maior parte dos estudos, portanto, apontam que o acesso aos serviços de saneamento básico tem efeito sobre a queda da taxa de mortalidade infantil no Brasil e em outros países do mundo. A próxima seção tratará do impacto do mesmo sobre os indicadores de morbidade de uma região.

5.2.2 Saneamento Básico e Morbidade

O objetivo desta seção é enfatizar a relação entre serviços de saneamento básico e as doenças descritas anteriormente (DRSAI), mas que não levam o indivíduo a óbito. Em outras palavras, destacar o papel do saneamento sobre o perfil de morbidade de uma região, de forma a enfatizar os impactos gerados sobre a acumulação de capital humano dos sobreviventes: as pessoas que conseguiram vencer a incidência dos agentes patogênicos, porém com perda de produtividade, rendimento escolar e outros impactos sociais.

Desde a década de 1970, é reconhecido que crianças que sofrem de desnutrição em seus primeiros anos terão seu desenvolvimento físico e mental reduzido, diminuindo o retorno ao investimento em capital humano, como educação e treinamento profissional (SELOWSKY e

TAYLOR, 1973). De acordo com Fellenberg (1980), moscas e mosquitos, que se desenvolvem nos restos de alimento depositados no lixo e em ambientes que acumulam água, podem invadir áreas habitadas próximas e transmitir doenças como Dengue, Salmonelas, estreptococos, estafilococos, bacilo do tétano, entre outros.

Os núcleos urbanos desprovidos de rede coletora ou com cobertura insuficiente acabam por eliminar resíduos líquidos em fossas ou valas, que se direcionam para o meio ambiente. A infiltração no solo pode atingir o lençol freático e contaminá-lo. Se o esgoto estiver correndo na superfície, a massa de resíduo orgânico pode acumular-se na margem do fluxo. Existe risco de surgirem focos de transmissão de leptospirose e, entre outras infecções entéricas, a hepatite tipo A (NATAL, *et. al.*, 2005). A convivência da população em ambientes com alto risco de contaminação, tais como os descritos pelos autores acima, deu origem a diversos estudos, que consolidaram a relação da falta de saneamento com a incidência de tais doenças, bem como com as consequências que elas poderiam causar aos indivíduos.

Os resultados empíricos de Snow (1990), em 1854, mostraram a associação existente entre a fonte de água consumida pela população de Londres e a incidência de cólera. Em relação à malária, dentre as principais causas dessa doença estão a má qualidade da água e as poças de água parada que afetam principalmente as crianças, que se tornam adultos menos produtivos (OMS, 2004). Gallup e Sachs (2001) mostraram que países como Costa do Marfim ou Zâmbia, em que 50% da população está exposta a essa doença, tem rendimentos per capita que são um terço daqueles países em que as pessoas estão livres da contaminação pela mesma. Bleakley (2010) encontrou evidências, em vários países da América Latina, de que uma criança que cresce livre da malária ganha 50% mais por ano durante toda sua vida adulta, em comparação com uma criança que tenha contraído a doença. Banerjee e Duflo (2011) ressaltam a gravidade da incidência de tais doenças, uma vez que têm o potencial de se tornarem fonte de uma variedade de armadilhas: trabalhadores que vivem em ambiente insalubre podem perder dias de trabalho, crianças que frequentemente estão doentes podem ser incapazes de ter bons resultados escolares e mães doentes podem gerar filhos com sérias complicações de saúde.

Bleakley (2007) estimou o impacto dos esforços da Fundação Rockefeller, na virada do século XX, para promover um programa em larga escala de erradicação do parasita da ancilostomose (“amarelão”) no Sul dos Estados Unidos. O agente patogênico, que se

desenvolve em condições sanitárias inadequadas, foi espalhado pelo contato com as fezes humanas e causou anemia, letargia e raquitismo nas crianças. O autor encontrou que crianças infectadas com o parasita tinham uma probabilidade 20 por cento menor de matricular-se na escola. Além disso, pelo fato do programa ter ocorrido no período 1910-1915, os efeitos de longo prazo mostraram que a população infectada durante a infância teve uma redução de 40% em seus salários após se tornarem adultos. Em outras palavras, a incidência da doença causou resultados insatisfatórios em relação à aquisição de escolaridade e menores ganhos salariais no futuro.

Miguel e Kremer (2004) encontraram um efeito positivo da medicação contra *helminths* (“vermes”) intestinais sobre a frequência escolar de crianças quenianas. Os autores ressaltam a importância da adequada provisão de bens públicos de saneamento, uma vez que a transmissão de tais agentes patogênicos ocorreu pela matéria fecal não tratada ou não disposta corretamente pelos mesmos. Ou seja, oferecer tais serviços é uma forma de impactar positivamente o desempenho das crianças na escola. Duflo *et al* (2015) estimaram o impacto de um programa integrado de oferta dos serviços de água encanada e fossas higiênicas, com custo razoável, para cada residência de um conjunto de vilas na área rural da Índia. A complementariedade de tais serviços, juntamente com as externalidades que geram, mostrou-se efetiva para a redução de doenças transmitidas pela água. Os resultados encontrados pelos autores indicaram redução em torno de 30 a 50% dos episódios de diarreia aguda e, ainda que em menor magnitude e precisão, também houve diminuição da incidência de malária e febre.

Fewtrell *et. al.* (2005) realizaram uma coletânea de 47 estudos científicos que tiveram como variável explicada a incidência de doenças diarreicas e cujas intervenções foram relativas às ações de saneamento básico. Concluíram que abastecimento de água, coleta de esgoto e práticas higiênicas e também a combinação destas ações foram efetivas na redução de tais doenças. Em relação às intervenções associadas à qualidade da água, estas se mostraram ainda mais eficientes. Porém, estudo de Esrey (1991) destaca superioridade da intervenção de esgotamento sanitário comparativamente à oferta de água na queda de incidência de diversas doenças relacionadas à saneamento.

Para o caso brasileiro, um dos determinantes que explica a evolução dos indicadores de saúde e nutrição infantil são as melhorias no sistema de água e saneamento (VICTORA *et al.*, 2011). Kassouf (1994) destacou o papel da infraestrutura residencial sobre a saúde das

crianças em idade escolar. A água encanada foi importante para a melhoria da saúde das crianças na região Nordeste e no setor urbano. Na região Sul e no setor rural, o sistema de esgoto afetou positivamente a saúde das crianças. Heller (1997) fez uma revisão de diversos tipos de estudos epidemiológicos e mostrou que um número importante deles aponta para a associação entre condições de saneamento e saúde. Andreazzi *et al.* (2007), revisaram a literatura publicada entre 1995 e 2004 acerca da relação entre saneamento e saúde. Encontraram que em 15 estudos foi observada uma associação positiva entre a variável de acesso a saneamento e a melhora da variável de saúde investigada.

Barreto *et al.* (2010) investigaram por meio de um modelo hierárquico os efeitos epidemiológicos da intervenção de um programa de saneamento em Salvador no ano de 1997. Apontaram queda de 21% na prevalência de diarreia nas crianças com menos de 3 anos de idade entre o período pré e pós introdução do programa. O estudo ainda ressaltou que essa queda ocorreu via aumento da cobertura na vizinhança, já que o programa conseguiu diminuir a poluição fecal dos vizinhos como um todo.

Além disso, o aumento na taxa de consumo de água de uma comunidade depois da disponibilização de água encanada, cuja fonte principal anteriormente era o poço raso ou caminhões-pipa, é normalmente acompanhado por aumento dos efluentes, com extravasamento das fossas negras e, em geral, lançamento nas sarjetas, dentre outros problemas decorrentes. Um benefício à saúde pública, que é o abastecimento de água potável, poderá ter um saldo negativo, na ausência da implantação conjunta e concomitante do esgotamento sanitário (Phillippi Jr. e Mallheiros, 2005). Os dois estudos a seguir tiveram como região de estudo o estado da Bahia e, além disso, trataram de duas variáveis de saneamento pouco estudadas, manejo de resíduos sólidos e de águas pluviais, em termos de impactos que ambas exercem sobre os indicadores de saúde.

Moraes (2003) *et al.*, a partir de três regiões pré-estabelecidas em Salvador, mediu os impactos do saneamento ambiental sobre doenças diarreicas. A incidência de diarreia nas crianças em vizinhança que contavam com apenas os serviços de drenagem de águas pluviais foi de dois terços da presente nas regiões que não tinham esse serviço. A incidência daquelas que tinham drenagem e esgotamento sanitário foi menos que um terço também em comparação a regiões que não tinham nenhum desses dois serviços. Rego *et al.* (2004) mostrou que crianças

expostas ao lixo no ambiente em que vivem tiveram probabilidade 3,98 vezes maior de ter diarreia, quando comparadas a crianças que não são expostas.

De forma geral, os estudos apontam o efeito positivo de ações de saneamento sobre a queda da incidência de DRSAl, com destaque para a diarreia. Além disso, tais estudos ressaltam as dificuldades associadas ao motivo doença que, conforme prevê o modelo de Grossman (1972), podem comprometer o rendimento das crianças no futuro.

5.2.3 Mais uma razão para investir em saneamento básico

No final da década de 1980, Briscoe (1987) destacava que, por mais que avanços nas áreas de abastecimento de água e esgotamento sanitário nos países em desenvolvimento resultassem na redução das taxas de mortalidade, a exemplo do ocorrido em outros países, havia um falacioso argumento de que o custo de cada disfunção infantil prevenida por meio do acesso adequado aos serviços de abastecimento de água e esgoto, seria superior ao custo correspondente ao de outras medidas de atenção primária (por exemplo, terapia de reidratação oral, vacinas e tratamento contra malária). No entanto, esse argumento baseava-se somente nos custos brutos das políticas de saneamento e não nos custos líquidos. Ou seja, naqueles custos não se levava em consideração os valores tradicionalmente pagos pela população por tais serviços. Além disso, a análise também não reconhecia que os efeitos benéficos de longo prazo dessa infraestrutura são substancialmente superiores aos de intervenções médicas, uma vez que é capaz de gerar efeitos multiplicadores (BRISCOE, 1987).

Segundo Watson (2006), os investimentos em saúde pública, incluído o acesso a bens públicos de saneamento, têm efeitos benéficos sobre resultados econômicos, presumivelmente mediante os seus impactos sobre a saúde. Rocha e Soares (2012) estimaram que, para a região do semiárido nordestino brasileiro, considerando somente os impactos sobre mortalidade infantil, a universalização combinada dos serviços de saneamento básico é eficaz do ponto de vista de custo-benefício e, portanto, deveria ser implementada. Para Mendonça e Motta (2007), é mais viável economicamente investir em saneamento do que executar gastos defensivos em saúde para tratar doenças relacionadas à falta de tais serviços.

Portanto, além do investimento no setor de saneamento básico poupar vidas de muitas crianças e evitar os gastos relativos à cura de inúmeras doenças infectocontagiosas, é fator essencial para a acumulação de capital humano no longo prazo.

5.3 Análise exploratória dos dados espaciais (ESDA)

Segundo Anselin (1998), a análise exploratória dos dados espaciais (ESDA) descreve distribuições espaciais, identifica localidades atípicas, agrupamentos de observações semelhantes (*clusters*) e aponta formas de heterogeneidade espacial. Logo, antes de se estimar um modelo espacial deve-se fazer tal análise.

A ESDA identificará se há algum padrão espacial sistemático ou se os dados espaciais estão distribuídos aleatoriamente. Em outras palavras, verificará se a taxa de incidência de internações por DRSAl de um distrito depende dos valores das taxas de incidência das localidades vizinhas. As áreas distritais que apresentam elevados valores para esse indicador de saúde pública podem gerar externalidades negativas aos distritos próximos, por estarem associadas ao efeito contágio característico da natureza de tais doenças. Ou seja, por serem doenças infectocontagiosas, é esperado que distritos que cercam localidades que são focos dos agentes patogênicos sofram com os efeitos de transbordamento. A análise será aplicada aos dados que compõem os mapas da figura 5.3 (população de todas as idades, à direita, e população de 0 a 14 anos, à esquerda, ano de 2010).

Figura 5.3 – Incidência por internações de DRSAl nos distritos brasileiros

O cálculo das estatísticas da ESDA necessita de uma matriz quadrada, W , denominada matriz de vizinhança ou matriz de ponderação espacial, de dimensão (10.225×10.225) . Essa matriz registra critérios geográficos de vizinhança entre as observações, tal que $w_{ij} > 0$ se os distritos i e j são vizinhos e $w_{ij} = 0$ caso não forem de acordo com o critério estabelecido.

Neste estudo, dois critérios de vizinhança foram utilizados na construção da matriz W . O primeiro deles, que deu origem à matriz de contiguidade, utilizou as fronteiras dos distritos para a construção dos pesos espaciais, em que:

$$(5.13) \quad w_{ij}^* = \begin{cases} 0, & \text{se não há fronteira entre } i \text{ e } j \\ 1, & \text{se há fronteira entre } i \text{ e } j. \end{cases}$$

O segundo critério foi baseado na distância geográfica dos centroides dos 100 vizinhos mais próximos³⁵, a partir do inverso da distância Euclidiana.

$$(5.14) \quad w_{ij}^* = \begin{cases} \frac{1}{(d_{ij})^\alpha}, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases}$$

$$(5.15) \quad d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

A vantagem do uso de uma matriz inversa da distância com $\alpha > 1$ é que faz com que a influência dos vizinhos diminua rapidamente com a distância.

A primeira estatística de teste global elaborada para a análise em questão é o I de Moran, coeficiente de autocorrelação espacial cuja hipótese nula é de aleatoriedade espacial, ou seja, variáveis espaciais aleatórias não apresentam dependência em relação a seus vizinhos. Moran (1948) propôs esta estatística de teste utilizando uma medida de autocovariância na forma de produto cruzado da seguinte forma:

$$(5.16) \quad I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

³⁵ Em primeiro lugar, verificou-se a distância mínima (370 km) para a qual os distritos tivessem ao menos um vizinho. A análise do histograma dessa matriz auxiliar mostrou que o número máximo de vizinhos dentro dessa distância era de, aproximadamente, na média, 100 vizinhos.

$$(5.17) \quad I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i x_i \sum_j w_{ij} x_j}{\sum_i x_i^2}$$

O símbolo n da equação 5.16 indica o número de unidades espaciais, x é a variável de interesse, $\bar{x}$ sua média e w_{ij} são elementos da matriz de pesos espaciais. A equação 5.17 é a equação 5.16 escrita em termos matriciais, e denota uma medida de autocorrelação de x com sua defasagem espacial. O índice de Moran varia entre -1 e 1, sendo que o primeiro corresponde à máxima correlação espacial negativa, enquanto o segundo diz respeito à máxima correlação espacial positiva. O valor zero para essa estatística indica que os dados não apresentam nenhum padrão espacial sistemático. Na tabela 5.1 abaixo encontram-se os valores do I de Moran para as duas variáveis dependentes desse estudo, associadas às duas matrizes de ponderação espacial (a de contiguidade e a de distância inversa). Os resultados mostram que a estatística I de Moran é significativa e positiva para as duas variáveis e matrizes de pesos. Logo, existe um padrão de concentração espacial através dos distritos, embora a força desta concentração seja fraca.

Tabela 5.1 – Coeficiente I de Moran para taxa de incidência de DRSAl

Variáveis Dependentes	I Moran (W_cont)		I Moran (W_distInv)	
	Estatística	p-valor	Estatística	p-valor
Inc_todas_idades	0,1190	0,01	0,0924	0,01
Inc_0 a 14 anos	0,0641	0,01	0,0473	0,01

Outra forma possível de visualizar a autocorrelação espacial é por meio do diagrama de dispersão de Moran, que plota o valor da variável de interesse no eixo horizontal e o valor da sua defasagem espacial no eixo vertical. O coeficiente I de Moran pode ser interpretado também como o coeficiente angular da reta de regressão dessas duas variáveis. Além disso, permite identificar quatro tipos de associação linear espacial: alto-alto (AA), baixo-baixo (BB), alto-baixo (AB) e baixo-alto (BA). As variáveis plotadas no diagrama estão padronizadas: o valor nulo representa a média e as unidades do gráfico denotam desvios-padrão. Os diagramas de dispersão de Moran para todas as especificações encontram-se no anexo.

Pode-se concluir de tais gráficos que a maioria dos distritos encontra-se no quadrante alto-alto (AA). Logo, os distritos exibem valores altos para a taxa de incidência por DRSAl, ou seja, acima da média, rodeados por regiões que apresentam valores também altos. Em ambos os diagramas, a inclinação da reta de regressão é positiva, conforme evidenciado pelas

estatísticas de teste do I de Moran apresentados na Tabela anterior. Os dados dos diagramas, para a população de 0 a 14 anos, mostram maior concentração dos distritos no quadrante alto-alto (AA). Além disso, observa-se que a reta de regressão, neste caso, embora positiva apresenta-se menos inclinada do que no caso anterior.

O valor positivo e estatisticamente significativo da estatística I de Moran é global, pois leva em conta todas as regiões. No entanto, é possível encontrar alguns agrupamentos espaciais locais que apresentam autocorrelação espacial negativa. Segundo Anselin (1995), um *Local Indicator of Spatial Association* (LISA) tem o objetivo de detectar padrões locais de autocorrelação espacial nas variáveis. Estas estatísticas identificam, para cada observação, os *clusters* espaciais significativos estatisticamente, em função da média dos vizinhos. Além disso, o somatório dos indicadores locais é proporcional ao indicador global de associação espacial. O coeficiente I_i de Moran local é uma das estatísticas LISA proposta por Anselin (1995), que tem a particularidade de fazer uma decomposição do indicador global em quatro tipos de aglomerações no território (Alto-Alto, Baixo-Baixo, Alto-Baixo e Baixo-Alto), que serão destacadas adiante. A forma de estimação desse coeficiente se dá através da correlação do valor padronizado (desvio da média global) da variável z do distrito i em relação ao valor padronizado da variável z do vizinho j conforme a equação 5.18:

$$(5.18) \quad I_i = z_i \sum_{j=1}^J w_{ij} z_j$$

Neste estudo, calculou-se o indicador LISA para a taxa de incidência de DRSAI por mil habitantes para o ano de 2010.

$$(5.19) \quad I_i = (Inc_{DRSAI_i} - \overline{Inc_{DRSAI}}) \sum_{j=1}^N w_{ij} (Inc_{DRSAI_j} - \overline{Inc_{DRSAI}})$$

Os distritos que possuem altas taxas de incidência de DRSAI e que são cercados por distritos cuja média das taxas de incidência de DRSAI também é alta fazem parte da aglomeração ou *cluster* denominado Alto-Alto. O *cluster* Baixo-Baixo indica que os distritos pertencentes as estas aglomerações têm baixas taxas de incidência de DRSAI e possuem vizinhos cuja média também é baixa. Logo, nesses dois casos, há autocorrelação espacial positiva. Ao contrário, a autocorrelação espacial negativa ocorre quando um determinado distrito possui alta (baixa) taxa de incidência e é cercado por vizinhos que, conjuntamente, apresentam um valor baixo

(alto) para a média das taxas de incidência de DRSAl. Ou seja, trata-se do *cluster* Alto-Baixo (Baixo-Alto).

Para a análise desta tese, é interessante a identificação de *clusters* Alto-Alto (AA). No entanto, os *clusters* Alto-Baixo (AB) e Baixo-Alto (BA) também podem apresentar riscos para as regiões de baixa taxa de incidência de DRSAl, uma vez que os agentes patogênicos de tais doenças, presentes de forma mais intensiva nas regiões de alta taxa de incidência, podem atingir tais regiões menos infectadas devido ao efeito contágio.

Neste estudo, o LISA foi calculado com um nível de significância de 5% e os vizinhos foram definidos conforme uma matriz de pesos de contiguidade de primeira ordem. O mapa da figura 5.4 mostra a análise espacial de *cluster* para o país como um todo, de forma a identificar os principais distritos e regiões que podem evidenciar situações preocupantes em termos de saúde pública.

Figura 5.4 – Distribuição espacial das aglomerações das taxas de incidência de internações por DRSAl para a população de todas as idades (2010)

A análise do mapa da distribuição espacial da estatística LISA para a taxa de incidência de internações por DRSAI por distrito permite identificar a existência de dois aglomerados bem definidos. Nas regiões Sul e Sudeste destaca-se a presença majoritária de *clusters* do tipo Baixo-Baixo, enquanto nas regiões Norte, Nordeste e Centro-Oeste é visível a prevalência de *clusters* Alto-Alto. Esse resultado é bastante similar às tendências apresentadas por essas regiões para as variáveis anteriormente apresentadas, que dizem respeito aos indicadores de acesso à saneamento básico.

Para cada região brasileira foi realizado um “zoom” para o mapa de clusters, de forma a facilitar a identificação dos grupos de distritos mais problemáticos. Tais mapas são encontrados na seção 7A do Apêndice. O gráfico a seguir indica os valores das taxas de incidência por DRSAI para cada estado bem como o número de distritos presentes em cada cluster de interesse (Alto-Alto, Baixo-Alto e Alto-Baixo).

Figura 5.5 – Valores da taxa de incidência DRSAI por clusters

Na região norte é possível constatar que todos os estados apresentam altos valores para a taxa de incidência de internações por DRSAI. O estado do Pará apresenta o maior número de distritos que fazem parte de um *cluster* Alto-Alto (66) em relação aos outros estados da região norte. Em termos de taxa de incidência de internações por DRSAI em 2010, esse estado é o segundo colocado no ranking brasileiro, com 966 casos por 100 mil habitantes.

O estado de Rondônia possui 27 distritos nesse tipo de aglomeração (32% do seu número total de distritos) e é o terceiro estado com a maior taxa de incidência por DRSAI em termos nacionais (864,45). Fonseca e Vasconcelos (2011) encontraram resultados semelhantes em relação ao estado de Rondônia quando analisaram os coeficientes *I* de Moran local para a taxa média de internação por DRSAI para o período de 2000 a 2010. As autoras encontraram que Rondônia é o estado brasileiro que possui a maior média de taxa de internação por DRSAI e quase a totalidade dos seus municípios têm altos valores para essa variável, ou seja, os municípios desse estado são cercados por outros municípios na mesma situação.

A região Nordeste possui os três estados que mais possuem distritos inseridos no cluster Alto-Alto (AA): Piauí (106), Maranhão (103) e Bahia (101). O estado do Piauí apresenta 46,29% dos seus distritos cercados por distritos vizinhos com altas taxas de internação por DRSAI. Para o estado do Maranhão essa porcentagem é de 42,39%. Os valores das taxas de incidência de internações por DRSAI também são elevadas para os dois estados, de tal modo que o estado do Piauí se encontra em primeiro lugar (967,66) no *ranking* nacional, o Maranhão em quarto (849,96) e a Bahia na oitava posição (591,33).

De acordo com Funasa (2004), para o ano de 2000, o risco de ocorrer uma internação por DRSAI na região Norte é mais de cinco vezes maior em Rondônia do que em Roraima, enquanto que no Nordeste os estados com maior e menor taxa eram Piauí (1035,20) e Sergipe (339,90). Isto é, após uma década, os valores continuam praticamente os mesmos assim como a posição dos estados na classificação regional.

As regiões Sudeste e Sul estão em situação privilegiada comparativamente às outras regiões: tanto apresentam valores baixos para as variáveis de taxas de incidência por DRSAI (barras azuis no gráfico) quanto indicam valores também baixos do número de distritos que têm altas taxas dessa variável, cercados por vizinhos que também apresentam altas taxas. A próxima seção tratará sobre a metodologia que será utilizada para a estimação dos fatores associados à incidência de DRSAI.

5.4 Metodologia

A relação entre a incidência de DRSAI e as variáveis de saneamento de uma localidade é estimada a partir de $n = 10.225$ distritos ($i = 1, \dots, 10225$), de acordo com o seguinte modelo

$$(5.19) \text{ IncDrsai}_i = \text{San}_i\gamma + X_i\beta + \varepsilon_i$$

Em notação matricial, *IncDrsai* é um vetor (10.225×1) de valores de taxas de incidência de internações por DRSAL. *San* é uma matriz ($10.225 \times (k + 1)$) de variáveis de acesso aos serviços de saneamento básico³⁶. *X* é uma matriz (10.225×8) de outras características exógenas dos distritos que afetam o valor da taxa de incidência de internações, ε é o vetor do termo de erros aleatórios e γ e β são os vetores de parâmetros a serem estimados.

A estimação de tais parâmetros por Mínimos Quadrados Ordinários (MQO) com dados em corte transversal deve garantir, entre outras hipóteses, que os erros estocásticos sejam independentes.

$$(5.20) E(\varepsilon_i\varepsilon_j/X) = 0 \text{ para } i \neq j$$

No entanto, de acordo com Almeida (2012), quando se trabalha com unidades geográficas, muito provavelmente essa hipótese é violada, pois a interação socioeconômica e a interdependência no espaço entre os agentes implica em correlação linear do erro de uma região *i* com o erro de outra região *j*. Além disso, a amostra esgota a população e, portanto, não é aleatória. As consequências de se ignorar os efeitos espaciais na modelagem econométrica são a ineficiência do estimador de MQO ou, um problema mais grave, o viés e a inconsistência do mesmo. Logo, deve-se considerá-los na análise, pois estes problemas estão presentes na maior parte dos dados de saúde e de saneamento.

O primeiro efeito espacial ocorre quando alterações em uma observação da variável *IncDrsai* (*IncDrsai_i*), causadas por choques endógenos e/ou exógenos ao modelo, extravasam para os vizinhos, causando alterações nas observações dos mesmos (*IncDrsai_j*)³⁷. A natureza infectoparasitária das DRSAL, assim como as características dos insetos vetores que transmitem os agentes patogênicos que as causam, justifica a inclusão desse efeito na modelagem. O mesmo pode ser denominado dependência espacial ou efeito transbordamento.

³⁶ Na primeira especificação adotada, que leva em consideração as variáveis de saneamento inadequadas, o valor de *k* é 3. Na segunda especificação, que inclui lixo e drenagem de águas pluviais, *k*=4. Nos dois casos, a constante não está incluída.

³⁷ Em outras palavras, o dado espacial apresenta a característica de autocorrelação espacial.

Além disso, pode-se fazer uma analogia ao que Haining (1990) denomina de processo de difusão, que se caracteriza pela adoção de um atributo de interesse por parte dos elementos de uma população fixa: o exemplo típico é o efeito dos transbordamentos tecnológicos decorrentes de uma inovação. Contudo, para este caso, o processo de difusão é compulsório, o atributo não é de interesse da população e é adquirido involuntariamente em decorrência de externalidades negativas causadas pelas características de transmissão de tais doenças.

O modelo de Defasagem Espacial (SAR) adaptado às variáveis deste trabalho é dado, em termos matriciais, pela equação a seguir (Anselin, 1988):

$$(5.21) \text{ IncDrsai}_i = \rho W \text{ IncDrsai}_i + \gamma S \text{ an}_i + X_i \beta + \varepsilon_i; \quad |\rho| < 1$$

$$(5.22) \varepsilon/X \approx i.i.d. N(0, \sigma_{un}^2 I_n)$$

Em que,

ρ é o parâmetro auto-regressivo espacial, que mede a influência média dos vizinhos sobre observações no vetor *IncDrsai*;

$W \text{ IncDrsai}_i$ é o termo de defasagem espacial, que representa, para cada distrito, a taxa de incidência de internações por DRSAI dos distritos vizinhos.

No modelo SAR, há um problema de endogeneidade intrínseco: a variável defasada de *IncDrsai* é correlacionada com o termo de erro, independente da estrutura de correlação de erros e da matriz de vizinhança utilizada. O modelo da equação 5.21 pode ser reescrito de acordo com a equação 5.23 e a correlação citada acima, pela equação 5.24:

$$(5.23) \text{ IncDrsai}_i = (I - \rho W)^{-1}(\gamma S \text{ an}_i + X_i \beta) + (I - \rho W)^{-1} \varepsilon$$

$$(5.24) E[(W \text{ IncDrsai}_i) \varepsilon^T] = \sigma_{\varepsilon}^2 W (I - \rho W)^{-1} I \neq 0$$

Na presença de endogeneidade, o método de MQO não é apropriado porque gera estimativa viesada e inconsistente do parâmetro autoregressivo. Uma das soluções propostas na literatura é usar máxima verossimilhança.

O segundo efeito espacial é a heterogeneidade espacial, que significa que podem existir respostas distintas decorrentes de um estímulo do fenômeno em estudo, dependendo da região onde ocorrem. O fato de que as condições de infraestrutura de saúde e de saneamento são heterogêneas no território justifica a presença desse efeito espacial. A consequência dessa característica do dado espacial para a estimação é o surgimento de erros heterocedásticos. Neste caso, ainda de acordo com Anselin e Bera (1998), a especificação mais comum desse efeito assume a seguinte forma:

$$(5.25) \text{ IncDrsai}_i = \text{San}_i\gamma + X_i\beta + \varepsilon; \quad \varepsilon = \lambda W\varepsilon + \xi$$

Em que,

λ é o coeficiente auto-regressivo espacial para a defasagem do termo de erro;

ξ é o termo de erro não-correlacionado e homocedástico.

Por fim, o último componente espacial a ser apresentado trata da defasagem do vetor de variáveis exógenas (San e β). Em outras palavras, esse efeito ocorre quando as variáveis contidas na matriz X transbordam espacialmente. De acordo com uma das hipóteses deste estudo, as condições de infraestrutura de saneamento dos vizinhos podem ser determinantes para o valor da taxa de incidência de internações por DRSAI do distrito em questão. Esse efeito é bastante intuitivo quando se verifica a solução dada à disposição final dos dejetos pelos municípios que não possuem infraestrutura de saneamento adequada. Em geral, essa solução consiste em afastar os rejeitos para as localidades vizinhas, ampliando o efeito negativo do saneamento sobre a saúde. Se a localidade que recebe tal externalidade possui saneamento insuficiente, sua situação agrava-se. Caso possua alguma qualidade no acesso a tais serviços, os efeitos positivos que seriam gerados podem ser cancelados ou amenizados. O terceiro componente espacial pode ser modelado da seguinte forma:

$$(5.26) \text{ IncDrsai}_i = W\text{San}_i\theta_1 + WX_i\theta_2 + \gamma\text{San}_i + X\beta + \varepsilon$$

Em que,

θ_1 e θ_2 são os parâmetros dos vetores das variáveis explicativas defasadas espacialmente.

Os três tipos de componentes espaciais mencionados acima dão origem, respectivamente, ao Modelo de Defasagem Espacial (SAR), ao Modelo de Erro Espacial (SEM) e ao Modelo Regressivo Cruzado Espacial (SLX).

O modelo SAR é considerado um modelo de dependência espacial de alcance global, pois o multiplicador espacial ρ impacta todas as regiões de forma sistêmica, ou seja, faz com que o impacto gerado pela taxa incidência de internação por DRSAI em uma região seja transmitido para as regiões vizinhas que, por sua vez, o transmitem para as regiões vizinhas das regiões vizinhas e assim por diante.

O modelo SEM também é considerado um modelo de dependência espacial de alcance global, porém relacionado com os resíduos. Os erros associados com qualquer observação são uma média dos erros nas regiões vizinhas mais um componente de erro aleatório. Segundo este modelo, a influência sobre a variável dependente não é resultado apenas do choque específico a uma região, mas também de transbordamentos de choques de regiões vizinhas, vizinhas das regiões vizinhas e assim por diante.

Existem fatores não observáveis específicos de uma região que podem ter impacto na taxa de incidência de internações das regiões vizinhas e que podem não estar corretamente incluídos em modelos não espaciais. Em relação ao modelo SLX, os impactos do transbordamento são localizados e não afetam todo o sistema. A combinação do modelo SAR ($\rho \neq 0$) com o modelo SLX ($\theta \neq 0$) resulta no modelo de Durbin Espacial (SDM), que inclui dois tipos de defasagem espacial: da variável dependente e das variáveis independentes:

$$(5.27) \text{ IncDrsai}_i = \rho W \text{ IncDrsai}_i + W \text{ San}_i \theta_1 + W X_i \theta_2 + \gamma \text{ San}_i + X_i \beta + \varepsilon$$

O valor da taxa de incidência de internação por DRSAI de um distrito depende da taxa de incidência de internação das regiões vizinhas, das variáveis de saneamento da própria região, das variáveis de saneamento das regiões vizinhas, de características socioeconômicas do distrito (variáveis de controle) e de características socioeconômicas dos distritos vizinhos.

Outra combinação importante neste estudo é a do modelo SEM ($\lambda \neq 0$) com o modelo SLX ($\theta \neq 0$), que resulta no modelo Durbin Espacial do Erro (SDEM), que inclui a defasagem do termo de erro e também a defasagem das variáveis independentes.

Os outros modelos estimados nesse estudo para comprovar a robustez dos resultados foram o modelo de Defasagem Espacial com Erro Autoregressivo Espacial (SAC) e o modelo de Dependência Espacial Geral (GSM). O primeiro deles assume $\lambda \neq 0$ e $\rho \neq 0$ e o segundo, o mais completo entre todos, apresenta os três efeitos espaciais descritos anteriormente: $\lambda \neq 0$, $\rho \neq 0$ e $\theta \neq 0$.

A decisão de escolha entre os modelos para especificação final envolve a análise dos testes de Multiplicador de Lagrange (LM), em suas versões tradicionais e robustas.

Em primeiro lugar, realiza-se a estimação do modelo de regressão não-espacial por meio de Mínimos Quadrados Ordinários (MQO). No segundo passo é feito o teste da hipótese de ausência de dependência espacial devido à omissão de termos espaciais (defasagem espacial ou erros auto-regressivos espaciais) por meio dos testes LM-ERRO e LM-LAG (inclusive em suas versões robustas).

O teste LM-erro assume que $\rho = 0$ e testa a hipótese nula $H_0: \lambda = 0$. Se esta hipótese nula for rejeitada, a escolha recai sobre o modelo SEM. Analogamente, o teste LM-lag assume que $\lambda = 0$ e a hipótese nula a ser testada é $H_0: \rho = 0$. Neste caso, se a hipótese nula for rejeitada, existem evidências de que se trata de um modelo de defasagem espacial (FLORAX et al., 2003). As versões robustas de ambos os testes estabelecem, no primeiro caso, a mesma hipótese nula, porém não assume $\rho = 0$. Ou seja, permite que esse parâmetro seja de qualquer valor e, portanto, admite a presença da defasagem espacial, ainda que o valor desta seja desconhecido. No segundo caso, não assume $\lambda = 0$. Dessa forma, o teste dirá se o termo de lag espacial deve ser incluído num modelo em que está presente uma correlação espacial nos erros de valor desconhecido. O teste SARMA diagnostica os dois tipos de dependência espacial, pois sua hipótese nula é que λ e ρ são zero conjuntamente. Ou seja, este último teste identifica os dois tipos de efeitos espaciais (defasagem - ρ - e heterogeneidade espacial - λ). Os resultados de tais decisões de escolha serão apresentados na próxima seção, acrescidos da estatística I de Moran utilizada como medida de dependência espacial, sob a hipótese nula de ausência de associação espacial (ocorrência das observações em qualquer lugar do espaço com igual probabilidade).

Outra estatística que auxilia na escolha entre modelos é o Critério de Informação Akaike (AIC). Este teste utiliza o valor da função de verossimilhança (LIK) para os modelos estimados por Máxima Verossimilhança (MV), utilizando a seguinte fórmula de cálculo:

$$(5.28) \text{ AIC} = -2\text{LIK} + 2k$$

O teste de Razão de Verossimilhança (RV) é baseado na função de verossimilhança e pode ser utilizado para verificar a autocorrelação espacial, tanto na forma de defasagem quanto de erro. A estatística do teste é dada por:

$$(5.29) \text{ RV} = 2[\text{LIK}_x - \text{LIK}_y]$$

Os subscritos x e y são os modelos espaciais e o teste segue RV uma distribuição qui-quadrado com q restrições.

A interpretação dos coeficientes em um modelo espacial é diferente da forma usual, ou seja, não se interpreta mais como o impacto marginal da variável independente sobre a variável dependente. Quando os efeitos da interação entre os distritos são levados em consideração, o efeito marginal passa a contar com a matriz da equação 5.30, de acordo com o modelo SDM:

$$(5.30) \frac{\partial \text{IncDrSai}}{\partial \text{San}} = (I - \rho W)^{-1} [W\theta_1 + \gamma I]$$

O cálculo das derivadas parciais desse modelo são feitos por uma matriz não diagonal, que é produto de duas outras, também não diagonais: $(I - \rho W)^{-1}$ e $[W\theta_1 + \gamma I]$. Quando há *spillovers* no modelo ($\rho \neq 0$), os termos fora da diagonal principal são diferentes de zero. Logo, quando a taxa de internações por DRSAI de um distrito varia, todas as observações do sistema são influenciadas. O diferencial do modelo SDM é que além dos efeitos associados à correlação espacial endógena ($\text{IncDrSai}_i \rightarrow \text{IncDrSai}_j$), há também os efeitos da correlação espacial exógena ($\text{San}_i \rightarrow \text{IncDrSai}_j$). Esses dois efeitos implicam em *spillovers* espaciais para os demais distritos, tanto via variável dependente diretamente como via variável independente.

LeSage e Pace (2009) definem o efeito direto (ED) como a média dos elementos da diagonal principal da matriz de derivadas parciais (efeitos da variação de uma variável explicativa na variável dependente da própria observação). Logo, é o impacto médio na taxa de incidência de internações por DRSAl ($IncDrsai_i$) quando ocorre variação de 1% na variável de saneamento (San_i). O efeito indireto (EI) é a média das somas dos elementos não diagonais em cada uma das linhas (ou colunas) da matriz (efeitos da variação de uma variável explicativa na variável dependente de outras observações). Dessa forma, tem-se o impacto médio na taxa de incidência de internações por DRSAl ($IncDrsai_i$) quando ocorre variação de 1% na variável de saneamento (San_j). O efeito total (ET) consiste na soma dos efeitos diretos e indiretos.

No modelo SDEM, por não possuir o termo representativo do efeito transbordamento ($\rho = 0$), os parâmetros são estimados de forma direta. Ou seja, γ é o impacto direto das variáveis de saneamento sobre a taxa de incidência de internações por DRSAl de um mesmo distrito ($San_i \rightarrow IncDrsai_i$) e θ_1 é o impacto do saneamento dos vizinhos sobre essa variável de saúde do distrito em questão ($San_j \rightarrow IncDrsai_i$).

Quanto à significância dos efeitos marginais, pode acontecer de alguns parâmetros dos modelos espaciais serem não significativos ao passo que os efeitos marginais sejam significativos (LESAGE e PACE, 2009; ELHORST, 2010). As variações também podem ocorrer quando se trata da significância dos efeitos diretos e indiretos conjuntamente frente à significância do efeito total. Nem sempre se os dois primeiros forem significativos, o terceiro será necessariamente. Ou seja, essa significância dependerá dos erros padrão estimados para cada caso.

No caso do modelo SDEM, pelo fato deste não apresentar efeito *feedback*, os efeitos direto e indireto são obtidos diretamente das estimativas dos coeficientes do modelo. Logo, não há cálculo para o erro-padrão do efeito total e, conseqüentemente, não se pode inferir a significância do mesmo aos níveis usuais. Por esse motivo, na exposição dos efeitos marginais para esse modelo, o efeito total não aparece como para os restantes. Na próxima seção são estabelecidas as hipóteses que serão testadas neste estudo.

5.5 Modelos e Hipóteses

Nessa tese, serão estimados dois tipos de especificação, como inicialmente enunciado em seção 3.4 e expresso no quadro 3.3, os quais serão denominados Modelo 1 e Modelo 2. Para cada um deles, haverá duas variáveis dependentes representativas de duas faixas etárias da população: a de todas as idades e a de 0 a 14 anos. O vetor X , referente às outras variáveis de controle, é o mesmo para os dois modelos. No primeiro deles, o vetor relativo às variáveis que expressam o tipo de acesso aos serviços de saneamento básico contém três formas inadequadas de acesso à água e esgoto (“ d_cist ” e “ d_frud ”) sendo que uma delas refere-se às piores condições em termos de serviços de esgotamento sanitário (“ $sbceuab$ ”). O Modelo 1 é expresso conforme a equação 5.31:

$$(5.31) \text{ IncDrsai}_i = D_cist_i \gamma_1 + Sbceuab_i \gamma_2 + D_frud_i \gamma_3 + X_i \beta + \varepsilon_i$$

O segundo modelo contém duas variáveis representativas das melhores condições de acesso aos serviços de água e esgoto (“ ag_nao_ceuab ” e “ esg_nao_ceuab ”) e duas outras representativas das condições inadequadas de acesso à coleta de lixo e moradia inadequada e drenagem de águas pluviais (“ $lixo_mor_inad$ ” e “ d_nao_bueiro ”). A equação (5.32) expressa a especificação do Modelo 2:

$$(5.32) \text{ IncDrsai}_i = Ag_nao_ceuab_i \alpha_1 + Esg_nao_ceuab_i \alpha_2 + Lixo_mor_inad_i \alpha_3 + D_nao_bueiro_i \alpha_4 + X_i \beta + \varepsilon_i$$

No primeiro modelo, a primeira e a terceira variáveis expressam formas alternativas de inadequada provisão dos serviços de abastecimento de água e coleta de esgoto, respectivamente. A segunda variável, por sua vez, reflete ausência total do serviço de esgotamento sanitário: além dos domicílios não terem banheiro, estão inseridos em condições adversas de esgoto a céu aberto. Tanto os estudos de Bleakey (2007), Miguel e Kremer (2004) e Esrey (1991) para outros países, quanto os estudos de Kassouf (1994), Moraes *et al.* (2003) e Barreto *et al.* (2007) para o Brasil, defendem a relação positiva entre variáveis de saúde (na maior parte medida pela incidência de doenças diarreicas) e o acesso aos serviços de esgotamento sanitário.

Em relação aos serviços de abastecimento de água (via quantidade ou qualidade), resultados nessa direção também foram encontrados para outros países por Snow (1990) e Fewtrell *et al.* (2005) e, para o caso brasileiro, por Kassouf (1994) e Andreazzi *et al.* (2007). Além disso, Duflo *et al.* (2015), Fewtrell *et al.* (2005), Phillilippi e Malheiros (2005) e Heller (1997), este último para o Brasil, apontam que as duas formas principais de intervenção em saneamento (água e esgoto) devem ser combinadas em ação conjunta e efetiva para a redução de incidência de DRSAs.

O modelo teórico desenvolvido na primeira seção deste capítulo também corrobora essa associação. Uma das variáveis que pode determinar a produção de saúde dos indivíduos é aquela relacionada ao ambiente em que o mesmo reside, ou seja, as condições de infraestrutura de saneamento básico do domicílio. Em relação a essa variável a implicação do modelo é que, por ser uma variável exógena, pode melhorar ou retardar a saúde (BERNDT, 2000). Nessa situação, por se tratar de variáveis “adversas” para a saúde humana, o deslocamento da curva de eficiência marginal do investimento em saúde (EMH), conforme Figura 5.2 deste estudo, desloca-se para a esquerda.

Logo, de acordo com os estudos mencionados e com o modelo teórico, é esperado que piores condições de acesso a tais serviços correspondam a uma maior propagação de DRSAs e, consequentemente, maiores taxas de internação, o que justifica o esperado sinal positivo para as variáveis, como apresentado no Quadro 3.3.

No segundo modelo, há dois efeitos a serem analisados: o efeito das variáveis que refletem as melhores condições de saneamento (“ag_nao_ceuab” e “esg_nao_ceuab”) e o efeito daquelas que refletem ausência de políticas de manejo de resíduos sólidos e drenagem de águas pluviais. Em relação às duas primeiras variáveis, a análise é complementar à realizada para o primeiro modelo, porém com sinal contrário. Espera-se que os domicílios que tenham acesso a água e esgoto e não estejam circundados por esgoto a céu aberto apresentem menor taxa de internações por DRSAs.

As duas últimas variáveis representam dois problemas pouco estudados na literatura que trata da associação destas variáveis com os indicadores de saúde: o acesso ao lixo e à drenagem de águas pluviais. Segundo Rego *et al.* (2004) e Moraes *et al.* (2003), acredita-se que a falta ou inadequado fornecimento desses dois serviços gere maiores taxas de internações por DRSAs,

dado que também são serviços inseridos na concepção abrangente de saneamento básico pela nova lei.

Em relação aos efeitos espaciais, as hipóteses estabelecidas nessa tese, que podem ser justificadas pelo modelo teórico anteriormente apresentado e pela literatura, definem a existência de três deles:

$$(5.33) H_0: \rho \neq 0 \text{ e } \lambda = 0$$

$$(5.34) H_0: \theta \neq 0$$

$$(5.35) H_0: \lambda \neq 0 \text{ e } \rho = 0$$

O primeiro deles, expresso na equação 5.33, refere-se ao efeito transbordamento, associado à taxa de incidência de internações por DRSAI de um distrito. A própria definição de tais doenças, infectocontagiosas, pressupõe a possibilidade de existir interações endógenas entre diferentes distritos. Uma dessas interações pode ocorrer por meio do deslocamento do inseto de um distrito a outro. Os insetos são considerados como vetores mecânicos que carregam uma diversidade de microrganismos transmissores e podem invadir áreas habitadas próximas e transmitir doenças. A outra forma de interação entre os distritos é via água contaminada dos rios e mares que atravessam diversas localidades. Em geral, a transmissão dos agentes patogênicos que causam as DRSAI ocorre pela ingestão de água contaminada por fezes, urina ou vômitos de doente ou portador.

Outro feito espacial que será considerado é o representado na equação 5.34. A contribuição para a incidência das internações por DRSAI de um determinado distrito pode estar associada à infraestrutura de saneamento básico do distrito vizinho. Natal *et al.* (2005) defendem que, do ponto de vista ecológico, a cidade não está isolada e tanto pode importar materiais e alimentos como pode devolver seus resíduos para as áreas vizinhas.

O terceiro efeito espacial possível, que também será testado neste estudo, é a presença da heterogeneidade espacial, última equação 5.35. De acordo com Golgher (2015), regiões próximas tendem a ter características similares, o que pode ser causado por externalidades ou podem representar um equilíbrio de longo prazo devido a fatores históricos em comum

ocorridos no desenvolvimento das localidades. Ou seja, a heterogeneidade espacial é relacionada com a estabilidade observada no espaço com relação a variáveis que apresentam estrutura espacial desigual, fato observado em diversos dos indicadores socioeconômicos.

A situação acima descrita pode se aplicar ao setor de saneamento, pois as empresas (públicas ou privadas) que oferecem os serviços à população atuam em diversos municípios, com diferentes níveis de eficiência, o que pode fazer com que desigualdades estruturais se propaguem e afetem tanto os indicadores de acesso e, conseqüentemente, as taxas de internações por DRSAL.

No Brasil, há vários arranjos institucionais na estrutura de oferta dos serviços de saneamento básico: empresas que abastecem um único município, outras que atendem um pequeno grupo deles (consórcios) e outras que atendem uma grande quantidade de municípios tais como a SABESP, a COPASA, entre outras. A formação destes grupos ocorreu por razões históricas e razões econômicas ligadas ao interesse da empresa em obter economias de escala ou se valer de subsídios cruzados. Logo, os indicadores de acesso a saneamento podem ser mais ou menos homogêneos tanto dentro de tais grupos quanto entre os municípios, estados e regiões.

Além da composição desta estrutura de oferta dos serviços, a qualidade e o alcance do atendimento realizado à população por estas empresas também está determinado pelas características das mesmas. Tais características podem ser observáveis, relativas ao seu desempenho, eficiência e nível de investimento realizado, e não observáveis, quando variáveis políticas associadas à natureza jurídica da empresa (pública ou privada) podem justificar um melhor ou pior desempenho na provisão dos serviços aos municípios.

Em outras palavras, o comportamento do acesso aos serviços de abastecimento de água, coleta e tratamento de esgoto, lixo e drenagem de águas pluviais não é uniforme no espaço por diversas razões. O mesmo ocorre com relação à área da saúde. Portanto, no contexto da problemática desta tese, os efeitos espaciais a serem testados são relativos à defasagem espacial, à defasagem do vetor de variáveis explicativas e à defasagem do termo de erro.

5.6 Resultados

A especificação de um modelo econométrico-espacial, em geral, depende de considerações teóricas acerca do fenômeno estudado. De acordo com Anselin (1988), essa abordagem parte de um modelo teórico bem definido e este é utilizado para derivar o modelo espacial das análises empíricas. Além disso, se a situação for intuitivamente óbvia quanto à dependência espacial ou heterogeneidade espacial, os modelos espaciais tendem a ser os mais apropriados (GOLGHER, 2015).

Segundo Almeida (2012), um dos procedimentos de especificação com maior potencial para se chegar em um modelo espacial mais representativo do mecanismo gerador de dados leva em consideração as seguintes etapas: estimar o modelo de regressão linear por Mínimos Quadrados Ordinários (sem componente espacial) e testar os resíduos por algum teste de autocorrelação espacial difuso (*I* de Moran). Caso existam evidências de efeitos espaciais, ou seja, se a hipótese nula for rejeitada, deve-se estimar os modelos espaciais.

O melhor modelo dentre todos os estimados precisa atender sequencialmente a duas condições: o modelo mais adequado não pode apresentar evidências de autocorrelação espacial em seus resíduos e, uma vez atendida essa condição, o modelo escolhido será aquele que apresentar o menor Critério de Informação (AIC). Ainda que este procedimento seja válido, a escolha do melhor modelo também deverá levar em conta o sinal, a magnitude e a significância estatística dos coeficientes que estiverem alinhados com a expectativa teórica e com as hipóteses estabelecidas. Além disso, é importante que os resultados sejam invariantes à utilização de ao menos duas diferentes matrizes de pesos espaciais.

As estimações realizadas a seguir foram feitas para duas populações: a de todas as idades e a população de 0 a 14 anos. Para cada caso, foram estimados modelos de regressão linear não espaciais por Mínimos Quadrados Ordinários (MQO) e os modelos SLX, que apresentam o vetor de variáveis explicativas defasado para as duas diferentes matrizes. Com os resíduos das estimações foram calculados os testes *I* de Moran. De acordo com a Tabela 5.2, conclui-se que há efeitos espaciais que precisam ser controlados, pois as estatísticas de todos os testes foram rejeitadas ao nível de significância de 1%. Caso tais efeitos não sejam controlados, pode ocorrer viés e ineficiência nas estimativas dos coeficientes, tanto para os modelos de alcance global quanto local.

Tabela 5.2 – Coeficiente I de Moran para Incidência de DRSAl, Modelo 1

Testes	MQO - Wcont		MQO - Wdist_inv		SLX - Wcont		SLX - Wdist_inv	
	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor
I de Moran	INC DRSAl (todas as idades)				INC DRSAl (todas as idades)			
	0,094	0,01	0,066	0,01	0,088	0,01	0,058	0,01
I de Moran	INC DRSAl (de 0 a 14 anos)				INC DRSAl (de 0 a 14 anos)			
	0,056	0,01	0,037	0,01	0,052	0,01	0,030	0,01

Nas próximas tabelas, para encontrar os resultados dos impactos espaciais do saneamento inadequado (modelo 1), foram estimados oito modelos espaciais de regressão para a taxa de incidência de internações por DRSAl da população de todas as idades para cada uma das duas diferentes matrizes de pesos espaciais. O mesmo procedimento foi realizado ao considerar a população de 0 a 14 anos. Em seção seguinte, a mesma análise foi realizada para a segunda especificação (modelo 2), que introduziu as variáveis de drenagem de águas pluviais, lixo e variáveis “limpas” de saneamento (acesso à água e esgoto sem a presença de esgoto a céu aberto em suas condições de entorno).

Os modelos estimados nos dois casos foram o de Mínimos Quadrados Ordinários (MQO), Modelo Regressivo Cruzado Espacial (SLX), Modelo de Erro Autoregressivo Espacial (SEM), Modelo de Durbin Espacial do Erro (SDEM), Modelo de Defasagem Espacial (SAR), Modelo de Durbin Espacial (SDM), Modelo de Defasagem Espacial com Erro Autoregressivo espacial de Kejian-Prucha (SAC) e o Modelo de Dependência Espacial Geral ou Generalizado (GNS) ou Modelo de Manski (1993). O método de estimação de todos eles foi por Máxima Verossimilhança (MV). De acordo com esse método, mesmo que a hipótese da normalidade do erro não seja respeitada, a estimação por MV conduz às propriedades desejáveis (consistência, normalidade assintótica e eficiência) quando se tem uma grande amostra e desde que o modelo represente o verdadeiro processo gerador de dados (LEE, 2004).

5.6.1 Modelo 1 – população de todas as idades.

Na Tabela 8.A1 do Apêndice são apresentados os modelos espaciais MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAl da população de todas as idades, utilizando a matriz de contiguidade de pesos espaciais. A regressão estimada por MQO aponta que há evidências de que as variáveis de saneamento inadequado e precário (*dcist*, *sbceuab* e *dfrud*) tenham contribuição relevante e positiva para a taxa de incidência de internações por DRSAl. O primeiro modelo econométrico espacial estimado (SLX) revela que ao menos duas

variáveis de saneamento dos vizinhos (*sbceuab* e *dfrud*) têm impacto positivo sobre a taxa de internações por DRSAI. Logo, há efeitos de transbordamento espaciais do vetor de variáveis exógenas do saneamento sobre a variável de saúde que precisam ser consideradas. Os outros dois modelos estimados (SEM e SDEM) apresentaram resultados similares quanto aos sinais e significância dos coeficientes. Além disso, segundo estes últimos dois modelos, o λ foi estatisticamente diferente de zero no nível de significância de 1%, indicando a presença de heterogeneidade espacial.

A Tabela 8.2 do Apêndice apresenta os resultados dos modelos SAR e SDM, que consideraram a inclusão da dependência espacial. O modelo de Kelejian-Prucha (SAC), que considera tanto a presença de *spillovers* entre as observações quanto a correlação espacial nos erros por causa de variáveis omitidas e não observadas, também foi estimado. Por fim, o último modelo estimado é aquele que contém todos os termos dos demais modelos (GNS).

De acordo com o critério de melhor ajuste do modelo (menor valor para o AIC), o modelo escolhido é o GNS. Porém, segundo Manski (1993), os efeitos endógenos da defasagem espacial, exógenos do vetor X e de correlação dos erros não podem ser separados de forma efetiva nesse modelo. Isso ocorre devido ao efeito reflexão, que se dá simultaneamente entre tais componentes espaciais e faz com que eles não possam ser interpretados de forma relevante, pois não podem ser distinguidos com efetividade. Dessa forma, é necessário impor alguma restrição para que os parâmetros sejam corretamente especificados.

Nesse sentido, de acordo com o segundo menor AIC, o modelo escolhido será o SDM. Para reafirmar essa escolha, adicionalmente foi testada a estratégia específica-geral de acordo com Elhorst (2010). Na primeira etapa dessa estratégia, estima-se o modelo mais específico dentre todos, que é o de MQO. O segundo passo é comparar esse modelo com o SAR e SEM por meio dos testes do Multiplicador de Lagrange (LM) clássicos (LM-erro e LM-lag) e robustos (RLM-erro e RLM-lag). Tais testes são realizados com os resíduos do modelo MQO. A Tabela 5.3 abaixo mostra os resultados de tais testes para a população de todas as idades e para as duas matrizes da primeira especificação (modelo 1 - efeitos de saneamento precário).

Tabela 5.3 – Testes LM e TRMV – pop de todas as idades

Testes	INC DRSAI (todas as idades)			
	MQO - Wcont		MQO - Wdist_inv	
	Estatística	p-valor	Estatística	p-valor
LM-erro	243,741	0,00	1417,555	0,00
LM-lag	243,979	0,00	1314,918	0,00
RLM-erro	1,698	0,19	133,531	0,00
RLM-lag	1,937	0,16	30,894	0,00
SARMA	245,678	0,00	1448,449	0,00
TRMV (SEM X SDM)	108,980	0,00	46,397	0,00
TRMV (SAR X SDM)	114,50	0,00	68,800	0,00
TRMV (SLX X SDM)	211,92	0,00	382,930	0,00
TRMV (SAR X SAC)	50,720	0,00	32,470	0,00
TRMV (SEM X SAC)	45,210	0,00	10,063	0,00

Na segunda e terceira colunas dessa tabela, vê-se que as hipóteses nulas dos testes RLM-lag e RLM-erro foram rejeitadas e, portanto, escolhe-se momentaneamente os dois modelos SEM e SAR. A próxima etapa dessa estratégia é estimar esses dois modelos por Máxima Verossimilhança juntamente com o modelo SDM (seção 8 do Apêndice). Por fim, os modelos SEM e SAR são comparados com o modelo espacial de Durbin por meio do teste de razão de máxima verossimilhança (TRMV).

Como ambas as hipóteses nulas foram rejeitadas, conclui-se que o modelo espacial de Durbin explica melhor a variabilidade dos dados do que os modelos SAR e SEM. No entanto, ainda falta testar os modelos SLX, SDEM e SAC. Prossegue-se com as comparações, por meio da TRMV, do modelo SDM com o modelo SLX. O resultado foi a rejeição da hipótese nula e, portanto, o procedimento é encerrado com o modelo SDM escolhido.

Quando se comparam os modelos SAR e SEM com o modelo SAC, ambas as hipóteses nulas da TRMV foram rejeitadas, de forma a apontar evidências de que o modelo SAC fosse estimado. No entanto, entre o modelo SDM e SAC, como ressaltado anteriormente, tanto pela justificativa teórica quanto pelo critério de informação, a escolha final do modelo recai no SDM.

Ao levar em consideração a matriz de distância inversa na estimação dos oito modelos acima mencionados, os testes são mais inconclusivos, pois todos foram rejeitados: os testes LM são rejeitados nas suas duas versões e os testes TRMV também. Ao utilizar o mesmo

procedimento anterior, de acordo com a mesma estratégia de Elhorst (2010)³⁸, chega-se em duas opções: modelo SAC e SDM novamente. Pelo motivo do menor Critério de Informação e pelas expectativas teóricas, o modelo SDM é o escolhido. Neste caso, o modelo que possui o menor Critério de Informação, que é o GNS, apresenta lambda não significativo, ou seja, mais um argumento a favor da escolha do SDM. Portanto, a correlação espacial nos termos de erro é preterida em favor da defasagem espacial. Os resultados das estimações para a população de 0 a 14 anos também se encontram nas tabelas da seção 8 Apêndice.

As estimações de todos esses modelos com uma matriz de vizinhança diferente destacam que a maioria dos sinais e significância dos coeficientes é invariante ao tipo de matriz de pesos espaciais adotada. No entanto, a magnitude dos mesmos se altera ligeiramente.

Conforme apresentado na seção de metodologia desse estudo, os coeficientes de modelos espaciais, em particular aqueles que apresentam o termo de defasagem espacial, não podem ser comumente interpretados como se interpreta a derivada parcial de um modelo clássico de MQO. Logo, para tais modelos, calculam-se os efeitos marginais, que levam em consideração o efeito *spillover*. Tais efeitos são exibidos na Tabela 5.4. Foram estimados os efeitos marginais desses dois modelos escolhidos com as duas matrizes de ponderação consideradas nesse estudo (SDM com matriz W contígua e SDM com matriz W inverso da distância).

Tabela 5.4 – Efeitos marginais - SDM (W contígua) e SDM (W distância)

Inc_DRSAI_todas_idades	SDM - W contígua			SDM - W distância		
	ED	EI	ET	ED	EI	ET
dcist	0,0586***	0,0282	0,08684	0,0597***	0,107	0,166
sbceuab	0,2653**	1,7087***	1,97399***	0,3156**	4,162***	4,477***
dfrud	0,0137***	0,011	0,02467*	0,0126***	0,013	0,026*
rendapc	0,0002	-0,0011	-0,00089	-0,0001	0,001	0,001
gini	4,589**	8,3185	12,90743	4,269**	12,472	16,741
mu_resp_alfab	0,007	-0,0258	-0,01879	0,001	0,015	0,016
tpopurb	0,0597***	-0,0318*	0,02788	0,0595***	-0,04*	0,019
dom_kmq	-0,0002	0,0003	0,00003	-0,0003	0,001	0,0000
est_saude	-0,0017	-0,0039	-0,00567	-0,0034	-0,04	-0,044
dummy_capital	-1,6937**	0,4412	-1,25248	-1,8889**	-0,345	-2,233
dummy_rm	-1,0109	-1,5515	-2,56243*	-0,9961	-2,695	-3,691*

Nota-se que, para as duas diferentes matrizes de vizinhança consideradas, os coeficientes das variáveis de saneamento inadequado se mostraram positivos e significativos. O impacto direto

³⁸ Outra vantagem de se estimar o SDM é que, ainda que o processo gerador de dados (DGP) verdadeiro não corresponda a esse modelo, os coeficientes serão não viesados.

médio sobre a taxa de incidência de internações por DRSAI decorrente de 1% do aumento da inadequada oferta de água à população (*dcist*) da própria região foi de 0,058 casos por 1000 habitantes quando se considerou a matriz contígua e 0,059 casos por 1000 habitantes ao utilizar a matriz de distância. Outra evidência importante é a ausência do efeito transbordamento para essa variável, pois os efeitos indiretos não foram significativos em ambos os casos.

Para a variável relativa ao inadequado acesso aos serviços de esgoto (*dfrud*), porcentagem de domicílios que tem coleta de esgoto via fossa rudimentar, o efeito direto valor de 0,012 casos por 1000 habitantes para a primeira matriz e 0,013 para a segunda. Neste caso, apesar do efeito indireto não ter sido significativo para nenhum dos dois casos, o efeito total, em termos de número de casos por 1000 habitantes, foi de 0,024 a 0,026.

A ausência de acesso à coleta de esgoto é representada pela variável *sbceuab*, que indica a porcentagem dos domicílios sem banheiro e que estão em regiões em que existe esgoto a céu aberto. O efeito direto dessa variável foi 0,26 e 0,31 casos por 1000 habitantes respectivamente para a primeira e segunda matriz, porém o destaque deve-se ao efeito indireto, em particular, para a matriz de distância (que capta a maior amplitude do efeito quando estende o critério de vizinhança em comparação ao critério de contiguidade). Essa variável apresentou efeito transbordamento altamente significativo e de alta magnitude frente aos demais. Ao utilizar a matriz de contiguidade, o efeito indireto foi de 1,7, porém foi de 4,1 com a matriz de distância. O efeito total no primeiro caso foi de 1,9 e de 4,4 no segundo.

5.6.2 Modelo 1 – população de zero a 14 anos.

As próximas estimações serão realizadas para a taxa de incidência de internações por DRSAI da população de 0 a 14 anos. Novamente, os primeiros oito modelos estimados consideram a matriz de contiguidade e, posteriormente, o mesmo exercício será feito para a matriz inversa da distância. Todos os resultados se encontram na seção 9 do Apêndice. Os testes LM com os resíduos do modelo MQO para essa população foram menos inconclusivos.

Tabela 5.5 – Testes de Multiplicador de Lagrange, Modelo 1

Testes	INC DRSAI (de 0 a 14 anos)				INC DRSAI (de 0 a 14 anos)			
	MQO - Wcont		MQO - Wdist_inv		SLX - Wcont		SLX - Wdist_inv	
	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor
LM-erro	85,693	0,00	447,706	0,00	75,955	0,00	302,951	0,00
LM-lag	79,708	0,00	415,267	0,00	78,829	0,00	306,964	0,00
RLM-erro	8,882	0,00	34,704	0,00	8,037	0,00	0,179	0,67
RLM-lag	2,897	0,09	2,265	0,13	10,912	0,00	4,191	0,04
SARMA	88,590	0,00	449,971	0,00	86,867	0,00	307,143	0,00

Os testes do Multiplicador de Lagrange, ao se adotar a matriz contígua, apontam escolha do SDM, ainda que o RM-lag tenha sido rejeitado a 9%. Quando os testes são repetidos para considerar os resíduos do modelo SLX, há evidência em favor da defasagem espacial. Logo, o modelo SDM foi estimado com a matriz de contiguidade. Quando se leva em consideração a matriz de distância inversa, os testes LM com os resíduos de MQO apontam que o componente espacial assume forma de correlação dos erros. No entanto, quando os mesmos testes são feitos com os resíduos do SLX, há evidências em favor da defasagem espacial. Devido a essa divergência, serão estimados também tanto o SDEM quanto o SDM com a matriz inversa da distância.

Tabela 5.6 – Efeitos marginais, SDM e SDEM

INC DRSAI (0 a 14 anos)	SDM - W contígua			SDM - W distância			SDEM - W distância	
	ED	EI	ET	ED	EI	ET	ED	EI
dcist	0,1049*	0,0137	0,1185*	0,0918**	0,1197	0,2115	0,0767	0,1384
sbceuab	0,1827	2,0347***	2,2174***	0,2126	4,4638***	4,6764***	-0,0206	3,8996***
dfrud	0,0381***	-0,0055	0,0325***	0,0295***	-0,0009	0,0286	0,031***	-0,0045
rendapc	-0,0001	-0,0019	-0,002	-0,0001	-0,0011	-0,0013	-0,0001	-0,0007
gini	6,1355	13,0663**	19,2018***	6,3469	15,9736	22,3206	5,3657	16,265
mu_resp_alfab	0,0559**	-0,0826**	-0,0267	0,0387	-0,0145	0,0241	0,0453*	-0,0507
tpopurb	0,1075***	-0,0516***	0,0559***	0,1078***	-0,0676**	0,0402	0,1108***	-0,0704**
dom_kmq	-0,0015***	0,0024***	0,0009	-0,0008*	0,0034**	0,0026*	-0,0012**	0,0039***
est_saude	0,0009	-0,0153*	-0,0144	-0,0032	-0,0911**	-0,0943**	0,0016	-0,0798**
dummy_capital	-7,4961***	6,6577**	-0,8384	-2,8753*	-0,1867	-3,062	-3,0833*	-0,4661
dummy_rm	-3,4182	-1,4363	-4,8545***	-1,5911	-6,3724**	-7,9635***	-1,4612	-5,5428*

Espera-se que a população de 0 a 14 anos seja mais sensível à influência do inadequado acesso aos serviços de saneamento. É o que ocorre, de fato, ao analisar os efeitos marginais. Para o efeito direto, as magnitudes dos coeficientes são maiores para *dcist* relativamente (0,10 e 0,09; quando se considera o modelo SDM), *dfrud* (0,038, 0,029 e 0,031; quando se considera o modelo SDEM) e, para o efeito total, os coeficientes de *sbceuab* são 2,21; 4,67 e 3,89; para os três modelos.

O aumento do Gini da própria região também afeta consideravelmente a incidência das taxas de internações por DRSAI da população dessa faixa etária. Em relação à concentração de domicílios por km quadrado, se os vizinhos exibem alta concentração, há um efeito positivo e significativo (embora de pequena magnitude) na própria região. Porém, esse efeito é negativo quando se considera o efeito direto. Esse fato pode ocorrer porque a concentração de domicílios também pode significar a obtenção de economias de escala por parte das empresas fornecedoras dos serviços de saneamento e, portanto, maior alcance no acesso, caso a mesma opere de forma eficiente.

No modelo SDM com a utilização da matriz de contiguidade, o efeito marginal da variável de educação das mulheres chefes de família apresenta resultado condizente com o modelo teórico. A Tabela 5.7 apresenta os testes que mostram que a autocorrelação dos dados foi eliminada. Com o novo cálculo do I de Moran, após os efeitos espaciais serem controlados, as estatísticas dos testes não são rejeitadas, como era esperado.

Tabela 5.7 – Novo I de Moran

Modelos e matrizes W - INC DRSAI (todas as idades)		
	Estatística	p-valor
SDM W contígua	-0,0055	0,873
SDM W distância	-0,0052	0,866
Modelos e matrizes W - INC DRSAI (de 0 a 14 anos)		
	Estatística	p-valor
SDM W contígua	-0,0018	0,585
SDM W distância	-0,0012	0,740
SDEM W distância	-0,0027	0,688

Nas próximas seções serão estimados os mesmos modelos espaciais para o modelo 2. Neste modelo são incluídas as variáveis “ideais” de saneamento assim como dois outros serviços cujos impactos são pouco estudados na literatura do setor: a drenagem de águas pluviais e o acesso à coleta de lixo.

5.6.3 Modelo 2 – população de todas as idades.

As estimações realizadas do Modelo 2 foram feitas para as mesmas duas populações: a de todas as idades e a população de 0 a 14 anos, iniciando pela primeira. Para cada caso, foram

estimados modelos de regressão linear não espaciais por MQO e os modelos SLX. Com os resíduos das estimações foram calculados os testes I de Moran.

Tabela 5.8 – Coeficiente I de Moran para Incidência de DRSAI, Modelo 2

Testes	MQO - Wcont		MQO - Wdist_inv		SLX - Wcont		SLX - Wdist_inv	
	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor
I de Moran	INC DRSAI (todas as idades)				INC DRSAI (todas as idades)			
	0,093	0,01	0,065	0,01	0,088	0,01	0,060	0,01
I de Moran	INC DRSAI (de 0 a 14 anos)				INC DRSAI (de 0 a 14 anos)			
	0,056	0,01	0,037	0,01	0,049	0,01	0,032	0,01

De acordo com os resultados mostrados na Tabela 5.8, conclui-se que há efeitos espaciais que precisam ser controlados, pois as estatísticas de todos os testes foram rejeitadas ao nível de significância de 1%.

A partir dos procedimentos de escolha de modelos relatados em seções anteriores, foi realizada a análise dos testes de Multiplicador de Lagrange. Quando se utiliza da matriz de distância, o modelo escolhido é o SDM, dado que o teste RLM-erro não foi significativo, ao passo que o RLM-lag foi significativo ao nível de significância de 7,9%. Tal conclusão foi obtida com a utilização dos resíduos de MQO. Por outro lado, quando se considera a matriz de contiguidade, ambos os testes (RLM-lag e RLM-erro) são significativos. Nesse caso, quando os testes robustos são inconclusivos, indica-se utilizar o modelo com a maior estatística de teste (OSLAND, 2010). Logo, para essa matriz contígua, será estimado o modelo SDEM.

Tabela 5.9 – Testes de Multiplicador de Lagrange, Modelo 2

Testes	INC DRSAI (todas as idades)				INC DRSAI (todas as idades)			
	MQO Wdist		MQO Wcont		SLX Wdist		SLX Wcont	
	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor
LM-erro	239,27	0,000	196,328	0,0000	1104,50	0,000	215,550	0,0000
LM-lag	241,51	0,000	187,225	0,0000	1092,08	0,000	224,385	0,0000
RLM-erro	0,83	0,362	12,563	0,0004	14,99	0,000	16,951	0,0000
RLM-lag	3,08	0,079	3,459	0,0629	2,57	0,109	25,786	0,0000
SARMA	242,35	0,000	199,787	0,0000	1107,08	0,000	241,337	0,0000

Na Tabela 5.10 são apresentados os efeitos marginais para esses dois modelos: o modelo SDM com a matriz de pesos espaciais inversa da distância e o SDEM com a matriz de contiguidade.

Tabela 5.10 – Efeitos marginais - SDM (W distância) e SDEM (W contígua)

Inc DRSAI (todas idades)	SDM - W dist			SDEM - W cont	
	ED	EI	ET	ED	EI
<i>esg_ao_ceuab</i>	0,0122**	-0,0624**	-0,0502*	0,0031	-0,0361***
<i>ag_ao_ceuab</i>	-0,0223***	0,046	0,0237	-0,024***	0,0077
<i>lixo_mor_inad</i>	0,1062	7,4499*	7,556*	0,1618	0,6574
<i>d_nbueiro</i>	0,0108	0,0477	0,0585*	0,0053	0,0435***
<i>rendapc</i>	-0,0001	0,0033	0,0032	0,0003	0,0001
<i>gini</i>	2,7831	6,4128	9,1959	4,2464**	3,8306
<i>mu_resp_alfab</i>	-0,0014	-0,0447	-0,0462	0,0057	-0,0445***
<i>tpopurb</i>	0,0624***	-0,0818*	-0,0194	0,0689***	-0,0427***
<i>dom_kmq</i>	-0,0003	0,0014	0,001	-0,0002	0,0005
<i>est_saude</i>	-0,0012	-0,0354	-0,0366	-0,0019	-0,0013
<i>dummy_capital</i>	-1,9754**	-1,4183	-3,3937	-1,8137	0,2758
<i>dummy_rm</i>	-0,9165	-2,9096	-3,8261*	-1,1808	-1,5283

As estimações do modelo 2 para a população de todas as idades indica que o efeito total do acesso ao esgoto “ideal” (*esg_ao_ceuab*) é negativo para a taxa de incidência de internações por DRSAI (para o SDM). O efeito indireto, negativo e significativo (tanto para o SDM quanto para SDEM), evidencia que o fato dos vizinhos terem condições de acesso ao esgotamento sanitário adequadas implica em menor taxa de DRSAI para o distrito em questão. Em outras palavras, novamente a externalidade pode ser detectada (embora de menor magnitude quando comparada ao modelo anterior).

Em consonância com o modelo 1, o impacto marginal indireto da variável de acesso adequado a água (*ag_ao_ceuab*) também não foi significativo. Apenas o efeito direto foi negativo e significativo, de forma a indicar que intervenções relacionadas ao abastecimento de água diminuem a taxa de incidência por DRSAI, porém com limitado alcance espacial. Em outras palavras, essa ação de saneamento tem efeito somente sobre a própria região.

O próximo resultado que merece destaque é o efeito indireto dos domicílios que apresentam moradia inadequada e lixo acumulado nos logradouros (*lixo_mor_inad*). Em decorrência do efeito de transbordamento espacial pela presença do termo de defasagem ($\rho \neq 0$), esse efeito desfavorável do lixo para a saúde é amplificado pela interdependência entre os distritos. A magnitude de tal externalidade, quando desconsiderada, torna mais uma vez subestimados os impactos das ações de saneamento básico. Ainda que exista a Política Nacional de Resíduos Sólidos (PNRS), voltada para as questões do manejo desse tipo de resíduos, tanto a definição abrangente de saneamento básico estabelecida pela lei federal número 11.445/2007 como a própria definição das DRSAI pela Funasa (2004), que inclui as condições inapropriadas de

moradia e condições de entorno, justificam que essa variável seja contemplada como variável determinante dessa variável de saúde.

A última variável de interesse, que apresentou sinal positivo e significativo (porém de menor magnitude), é a drenagem de águas pluviais, representada pela porcentagem de domicílios em que em sua quadra não havia bueiro ou boca-de-lobo (*d_nbueiro*). Quando não há esse escoamento das águas da chuva nas condições de entorno dos domicílios, podem ocorrer alagamentos e outras inundações que transmitem inúmeros agentes patogênicos das DRSAI.

Os sinais e magnitudes de todas as variáveis de interesse apresentados foram os esperados, apontados no Quadro 5.1 e justificados na seção correspondente. No que diz respeito às variáveis de controle, o coeficiente de Gini, a taxa de urbanização, a variável de educação das mulheres chefes de família e a variável *dummy* de capital apresentaram os sinais e significâncias no mesmo sentido dos que foram relatados em seção anterior.

5.6.4 Modelo 2 – população de zero a 14 anos.

A escolha entre os modelos espaciais estimados para a população de zero a 14 anos será realizada novamente por meio dos testes de Multiplicador de Lagrange. Quando se utiliza a matriz de distância, o modelo escolhido é o SDEM. Dado que os dois testes foram significativos, RLM-erro e RLM-lag, indica-se utilizar o modelo com a maior estatística de teste que, nesse caso, deve-se ao modelo de erro. Tal conclusão é confirmada quando o teste do RMLag não é significativo para os resíduos do modelo SLX. Por outro lado, quando se considera a matriz de contiguidade, novamente ambos os testes (RLM-lag e RLM-erro) fazem com que a escolha recaia sobre o modelo com a maior estatística de teste, que é o SAR. Logo, para essa matriz contígua, será estimado o modelo SDM (também confirmado quando se estima o SLX).

Tabela 5.11 – Testes de Multiplicador de Lagrange, Modelo 2

Testes	INC DRSAI (de 0 a 14 anos)				INC DRSAI (de 0 a 14 anos)			
	MQO Wcont		MQO Wdist		SLX Wcont		SLX Wdist	
	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor	Estatística	p-valor
LMerr	215,55	0,000	449,985	0,000	76,560	0,000	339,023	0,000
LMlag	224,385	0,000	422,706	0,000	79,463	0,000	335,824	0,000
RLMerr	16,951	0,000	30,242	0,000	7,784	0,005	3,702	0,054
RLMlag	25,786	0,000	2,964	0,085	10,686	0,001	0,504	0,478
SARMA	241,337	0,000	452,949	0,000	87,247	0,000	339,527	0,000

Na Tabela 5.12 foram estimados os efeitos marginais para esses dois modelos: o modelo SDM, com a matriz de pesos espaciais contígua, e o SDEM, com a matriz inversa da distância.

Tabela 5.12 – Efeitos marginais - SDEM (W distância) e SDM (W contígua)					
INC DRSAI (de 0 a 14 anos)	SDEM - W distância		SDM - W contígua		
	ED	EI	ED	EI	ET
<i>esg_ao_ceuab</i>	0,0125	-0,0639*	-0,002	-0,0443**	-0,0463***
<i>ag_ao_ceuab</i>	-0,0728***	0,1275**	-0,0663***	0,0429	-0,0234
<i>lixo_mor_inad</i>	-0,1229	7,0452	-0,0547	0,9661	0,9114
<i>d_nbueiro</i>	-0,0119	0,0694	-0,0211	0,0683***	0,0472**
<i>rendapc</i>	-0,0001	0,0005	0,0002	-0,0006	-0,0004
<i>gini</i>	4,114	10,302	4,1053	10,7831	14,8883**
<i>mu_resp_alfab</i>	0,0495*	-0,1134*	0,0576**	-0,1076***	-0,05*
<i>tpopurb</i>	0,1605***	-0,1846***	0,1644***	-0,0991***	0,0653**
<i>dom_kmq</i>	-0,0012***	0,0045***	-0,0015***	0,0026***	0,0012*
<i>est_saude</i>	0,0011	-0,0715**	0,0001	-0,0131	-0,013
<i>dummy_capital</i>	-3,7201**	-0,1261	-8,2409***	6,9385**	-1,3024
<i>dummy_rm</i>	-1,8186	-4,9044	-3,7947*	-1,3476	-5,1423***

O modelo 2 estimado para a população de zero a 14 anos apresentou efeitos negativos quando se consideram as condições “ideais” de acesso a esgotamento sanitário. O efeito direto das condições ideais de acesso a água também foi negativo e significativo, porém o efeito indireto no modelo SDEM apresentou um resultado à primeira vista contra-intuitivo, ou seja, positivo e significativo. Quando se faz o controle pela censura dos dados, na próxima seção, esse resultado contra intuitivo desaparece.

O efeito de não se ter bueiro ou boca de lobo nos domicílios aumenta, na média, a taxa de incidência de internações por DRSAI, indicando que a população de crianças é também afetada pelas doenças que podem ser transmitidas por alagamentos ou inundações. Os sinais e magnitudes de quase todas as variáveis de interesse (com exceção de uma) são os esperados. No que diz respeito às variáveis de controle, o coeficiente de Gini, a taxa de urbanização, a concentração de domicílios por quilômetro quadrado e as variáveis *dummies* de capital e região metropolitana apresentaram os sinais e significâncias no mesmo sentido dos que foram relatados em seção anterior. No que diz respeito a variável de educação das mulheres chefes de família, no modelo SDM o efeito total foi negativo e significativo, de acordo com a predição do modelo teórico e dos estudos apresentados na revisão da literatura. A última tabela desta seção, Tabela 5.13, apresenta os testes de que a autocorrelação dos dados foi eliminada por meio do controle de tais efeitos espaciais nas regressões.

Tabela 5.13 – Novo I de Moran

Modelos e matrizes W - INC DRSAl (todas as idades)		
	Estatística	p-valor
SDEM W contígua	0,0008	0,361
SDM W distância	0,0001	0,432
Modelos e matrizes W - INC DRSAl (de 0 a 14 anos)		
	Estatística	p-valor
SDM W contígua	0,00001	0,409
SDEM W distância	-0,0008	0,644

Com o novo cálculo do I de Moran, após os efeitos espaciais serem controlados, as estatísticas dos testes são rejeitadas, como era esperado

5.7 A Censura dos Dados

Em muitas análises empíricas, as informações da variável dependente não são observadas por causa de um determinado tipo de censura. Em relação ao fenômeno de estudo dessa tese, a censura dos dados deve-se à não ocorrência de casos de internações por DRSAl nos distritos. Nesse contexto, são distritos saudáveis que não registraram nenhuma internação por DRSAl para qualquer uma das duas populações analisadas. Conforme descrito anteriormente, a base de dados utilizada nesse estudo apresenta 3.703 observações (distritos), de um total de 10.225 observações, que apresentam valor igual a zero para a taxa de internações por DRSAl no ano de 2010 em relação à população de todas as idades. Quando se trata da população de 0 a 14 anos esse valor é de 4.174. Para tais situações, LeSage e Pace (2009) sugerem o uso de um modelo Tobit Espacial.

5.7.1 Metodologia Bayesiana

Antes de prosseguir com a apresentação do modelo Tobit Espacial convém realizar uma breve análise de um modelo Tobit convencional. De acordo com Tobin (1958), a variável explicada neste caso é contínua ao longo dos valores positivos, mas parte da amostra é composta por observações que têm valores nulos para essa variável. Se a variável latente (y^*) supera um nível crítico (c) os valores da variável latente são observados, caso contrário, observam-se valores nulos.

$$(5.36) \quad y = \begin{cases} y^* & \text{se } y^* > c \\ 0 & \text{se } y^* \leq c \end{cases}$$

Neste estudo, a censura ocorre à esquerda, nos valores iguais a zero ($c=0$). Caso um modelo Tobit seja estimado por Mínimos Quadrados Ordinários (MQO), as estimativas dos coeficientes serão viesadas e inconsistentes em decorrência do aparecimento da razão inversa de Mills, o segundo termo do lado direito da equação abaixo.

$$(5.37) \quad E[y/X] = X\beta + \sigma \left[\frac{\phi(X\beta/\sigma)}{\Phi(X\beta/\sigma)} \right]$$

A estimação consistente do parâmetro β é realizada por meio do método de máxima verossimilhança (MV). Almeida (2012) ressalta que existem duas complicações para um modelo Tobit com dados espaciais: a primeira delas refere-se à dificuldade computacional de se utilizar o estimador de máxima verossimilhança num contexto de dependência espacial. A segunda complicação refere-se ao fato do modelo Tobit ser muito sensível a hipóteses sobre distribuição dos termos de erro. Se houver heterocedasticidade ou não normalidade dos erros, a estimação torna-se inconsistente.

Segundo Lynch (2006), se existe incerteza acerca de um vetor de parâmetros de, por exemplo, um modelo Durbin Espacial (SDM), $\emptyset = (\rho, \beta, \theta, \sigma^2)$, é possível representá-lo por uma distribuição de probabilidade, denominada **distribuição de probabilidade a priori**, $f(\emptyset)$. O principal objetivo da abordagem Bayesiana, de acordo com o mesmo autor, é atualizar essa incerteza inicial, representada por $f(\emptyset)$, por meio dos dados empíricos (D) a fim de produzir uma **distribuição de probabilidade a posteriori**, com menor incerteza, representada por $f(\emptyset/D)$. A partir do teorema de Bayes³⁹, a seguinte relação pode ser obtida:

$$(5.38) \quad f(\emptyset/D) = \frac{f(D/\emptyset)f(\emptyset)}{f(D)}$$

Em que,

$f(D/\emptyset)$ = probabilidade de ocorrência dos dados condicionada nos parâmetros.

³⁹ O teorema de Bayes é um corolário do teorema da probabilidade total e mostra que a probabilidade de um evento (A) dado a ocorrência de um outro evento (B), $P(A/B)$, depende não somente da probabilidade do relacionamento entre esses eventos, $P(B/A)$, mas também da probabilidade marginal de ocorrência de cada um deles separadamente: $P(A)$ e $P(B)$. Logo, $P(A/B) = \frac{P(B/A)*P(A)}{P(B)}$.

$f(D)$ = probabilidade marginal (ou probabilidade simples) dos dados empíricos.

A equação 5.38 pode ser reescrita após assumir-se $\alpha = \frac{1}{f(D)}$, pois trata-se de uma constante de normalização para a distribuição a posteriori:

$$(5.39) \quad f(\emptyset/D) = \alpha f(D/\emptyset) f(\emptyset)$$

Se for possível, e teoricamente plausível, deve-se escolher o formato da distribuição a priori de forma que a distribuição a posteriori resultante seja de um formato conhecido. No entanto, quando a distribuição a posteriori não pode ser obtida analiticamente ou se a utilização desta é complicada, em alguns casos utiliza-se o algoritmo de Gibbs⁴⁰, o mais básico dentre os métodos bayesianos com Monte Carlo via cadeias de Markov (MCMC), para gerar as estimativas de forma indireta (GOLGHER, 2015).

A partir de uma distribuição conjunta, o algoritmo encontra as distribuições condicionais em formato fechado. A implementação desse algoritmo parte de um valor inicial de uma das variáveis e, por meio de um valor conhecido para ρ obtém-se um conjunto de valores de x e outro de y . Dado que ambos os valores de cada grupo dependem do valor inicial, deve-se descartar os primeiros valores de cada grupo.

A estimativas bayesianas do modelo heterocedástico Tobit Durbin Espacial (SDMT) foram obtidas a partir das funções disponibilizadas no site do LeSage e Pace (2009)⁴¹. É importante destacar que a distribuição $f(\emptyset)$ para o parâmetro de lag espacial ρ é uniforme, definida no intervalo $\rho \in (-1; 1)$ e, portanto, todos os valores nesse intervalo têm a mesma probabilidade inicial, $f(\rho) = 1/2$. O procedimento foi realizado 2500 vezes (*ndraw*) mas, destas, as 500 primeiras (*nomit*) obtiveram valores que foram descartados. Logo, o procedimento de obter um número aleatório de uma distribuição com média e variância conhecidas foi realizado 2000 vezes. Os seguintes conjuntos de valores foram obtidos para 1500 betas.

$$(5.40) \quad X = \{x_{501}, \dots, x_{2000}\}$$

$$(5.41) \quad Y = \{y_{501}, \dots, y_{2000}\}$$

⁴⁰ Para maiores detalhes ver Lacombe (2008).

⁴¹ <http://www.spatial-econometrics.com/>

O p-valor obtido para essa estimação é um pouco diferente do convencional, pois é determinado, segundo LeSage e Pace (2009), como:

$$(5.42) \quad p - level = 1 - \left(\frac{\text{número de draws em que o beta é } \neq 0}{ndraw - nomi} \right)$$

Dessa forma, quanto mais próximo de 0,25 o p-valor se encontra mais retiradas produziram betas diferentes de zero e, portanto, menor o nível de significância do parâmetro. Caso as estimativas dos parâmetros não difiram muito em termos de sinal, magnitude e significância dos coeficientes em relação aos modelos estimados sem o controle da razão inversa de Mills, pode-se concluir que este controle não tem influência direta sobre os resultados.

Em cada etapa da amostragem por MCMC obtém-se a matriz com as derivadas parciais da variável dependente em relação a uma variável independente específica e, portanto, é possível estimar os efeitos marginais do modelo espacial em cada uma dessas etapas. Os intervalos de confiança desses efeitos permitem que se verifiquem a significância dos mesmos. Os resultados da estimação dos modelos⁴² apresentam os limites superiores e inferiores de dois intervalos de confiança, 95% e 99%, cujos níveis de significância correspondem a 5% e 1% respectivamente. Caso tais intervalos de confiança contenham o zero, as variáveis não serão significantes. Logo, intervalos de confiança que contenham somente valores positivos ou apenas valores negativos apresentam efeitos estatisticamente significativos ao nível estabelecido.

5.7.2 Resultados

As estimações espaciais para a taxa de incidência de internações por DRSAI das seções anteriores foram importantes para nortear a escolha do melhor modelo e da verificação se os resultados iriam variar ao se utilizar diferentes matrizes de ponderação. Nestas estimações, o modelo SDM com a matriz de contiguidade se mostrou o mais adequado, na maior parte dos casos. Logo, estes modelos foram escolhidos para servir de comparação ao SDM Tobit.

A estimação do SDM Tobit foi realizada para os modelos 1 e 2 para as duas populações de interesse nesse estudo. Nas próximas tabelas, foram realizadas comparações entre o modelo

⁴² Tais estimações foram realizadas pelo *software Matlab*

MQO (sem qualquer dependência espacial), o modelo SDM (dependência espacial na forma de defasagem da variável dependente e das exógenas) e o modelo SDMT (SDM com controle dos dados censurados estimado por um Tobit Bayesiano). As tabelas completas com todas as informações das estimações encontram-se na seção 8 do Apêndice.

A Tabela 5.14 apresenta e compara tais modelos para a primeira especificação (modelo 1), somente com variáveis inadequadas de saneamento, para a população de todas as idades.

Tabela 5.14 – Comparação entre modelos: MQO, SDM e SDMT (todas as idades)

Variáveis	MQO	SDM	SDMT
dcist	0,0823***	0,0577**	0,1429***
sbceuab	0,7554***	0,2074	-0,0015
dfrud	0,023***	0,0133***	-0,0024
rendapc	-0,0012**	0,0002	-0,0003
gini	12,3859***	4,3069**	15,9892***
mu_resp_alfab	-0,0307***	0,0079	0,2269***
tpopurb	0,0536***	0,0607***	0,1821***
dom_kmq	-0,0003	-0,0003	-0,0001
est_saude	0,0003	-0,0016	-0,0027
dummy_capital	-1,6921**	-1,7087	-3,658**
dummy_rm	-2,7477***	-0,9583	-3,7953***
Wdcist		0,0096	0,0216
Wsbceuab		1,3223***	2,481***
Wdfrud		0,0058	0,0283***
Wrendapc		-0,0009	-0,0006
Wgini		5,695**	6,1999*
Wmu_resp_alfab		-0,0225	-0,2444***
Wtpopurb		-0,0391***	-0,0908***
Wdom_kmq		0,0003	0,0002*
West_saude		-0,0028	0,0042
Wdummy_capital		0,7381	1,7968
Wdummy_rm		-1,0273	0,3859
Constante		-4,2876***	-19,2627***

A análise da comparação entre modelos SDM e SDMT mostra que o sinal e a significância da maior parte dos coeficientes se mantiveram, com exceção da variável *dfrud*, que não foi significativa no coeficiente representativo do efeito direto, mas o foi em relação à variável defasada. O contrário se evidencia no modelo SDM para essa variável: o coeficiente da variável não defasada foi significativo, mas o da defasada não. A magnitude dos coeficientes se alteraram em todos os casos. Ou seja, as primeiras estimações, sem o controle dos zeros, estão subestimando o efeito do saneamento sobre a saúde. Segundo LeSage e Pace (2009), as estimativas de um modelo com dependência espacial na forma de defasagem tendem a ter os coeficientes subestimados pela existência de um viés para baixo quando não se faz o controle

da censura dos dados. A Tabela 5.15 apresenta os resultados do mesmo exercício para a população de 0 a 14 anos, igualmente para o modelo 1.

Tabela 5.15 – Comparação entre modelos: MQO, SDM e SDMT (de 0 a 14 anos)

Variáveis	MQO	SDM	SDMT
dcist	0,1372***	0,1055*	0,243***
sbceuab	0,879***	0,1315	-0,444
dfrud	0,0403***	0,0405***	0,004
rendapc	-0,0019**	-0,0001	-0,001
gini	18,5096***	6,0992	43,309***
mu_resp_alfab	-0,0264	0,0647**	0,334***
tpopurb	0,0996***	0,1082***	0,43***
dom_kmq	-0,0005	-0,0015***	-0,001**
est_saude	0,0025	0,0012	0,005
dummy_capital	-2,5055*	-7,6224***	-13,102***
dummy_rm	-4,8263***	-3,366	-9,753***
Wdcist		-0,0051	0,128
Wsbceuab		1,7543***	4,827***
Wdfrud		-0,0127	0,04**
Wrendapc		-0,0017	-0,002
Wgini		10,622*	7,813
Wmu_resp_alfab		-0,0847**	-0,37***
Wtpopurb		-0,0614***	-0,203***
Wdom_kmq		0,0023***	0,001***
West_saude		-0,0134*	0,001
Wdummy_capital		6,8834**	11,027***
Wdummy_rm		-0,8319	2,051
Constante		-6,7549*	-47,407***

A mesma conclusão obtida anteriormente pode ser aplicada à análise dos dados da tabela acima. Os sinais e significância da maior parte das estimativas são mantidos, mas a magnitude das mesmas é maior para o modelo SDMT. Em geral, como visto em seções anteriores, a população de 0 a 14 anos é mais afetada pelos efeitos deletérios da inadequada provisão dos serviços de saneamento básico.

O efeito de transbordamento da taxa de incidência de internações por DRSAI foi positivo e significativo com maior magnitude para o modelo SDMT. Esse resultado era esperado pela própria definição de doenças infectocontagiosas e, portanto, a hipótese estabelecida na seção 5.6 foi confirmada. A transmissão dos agentes patogênicos causadores das DRSAI que pode ocorrer tanto pelo deslocamento dos insetos de um distrito para o outro quanto pela água contaminada dos rios e mares que atravessam diversas localidades, gera o efeito de *spillover* espacial de tal forma que a incidência de DRSAI em um determinado distrito seja influenciada pela média da taxa de incidência de DRSAI dos distritos vizinhos. Além disso, a

hipótese estabelecida nessa mesma seção, e expressa na equação 5.34, de que a defasagem do vetor de variáveis explicativas pode importar para explicar a taxa de incidência de internações por DRSAI foi confirmada. As condições de infraestrutura de saneamento básico da vizinhança de um determinado distrito são capazes de afetar essa variável, conforme o modelo Durbin prevê. De acordo com Natal *et al.* (2005), há diversos exemplos de distritos ou municípios que lançam águas servidas sem tratamento, dejetos e outros resíduos coletados em suas regiões para fora de seus limites territoriais e que, conseqüentemente, atingem seus vizinhos mais próximos.

O mesmo argumento da seção anterior de que os coeficientes de um modelo espacial não são informativos em termos de análise do impacto marginal, também vale para esse caso. Os efeitos marginais também são comparados com relação aos obtidos nas estimações das seções anteriores deste capítulo para os dois modelos e duas populações. A Tabela 5.16 apresenta os resultados dos efeitos marginais da população de todas as idades do modelo 1.

Tabela 5.16 – Comparação entre efeitos marginais (SDM e SDMT)

Inc DRSAI (todas idades)	ED		EI		ET	
	SDMT	SDM	SDMT	SDM	SDMT	SDM
dcist	0,145***	0,0586***	0,072	0,0282	0,217***	0,08684
sbceuab	0,113	0,2653**	3,168***	1,7087***	3,281***	1,97399***
dfrud	-0,001	0,0137***	0,036***	0,011	0,035***	0,02467*
rendapc	-0,0003	0,0002	-0,0009	-0,0011	-0,001	-0,00089
gini	16,41***	4,589**	12,864***	8,3185	29,274***	12,90743
mu_resp_alfab	0,218***	0,007	-0,242***	-0,0258	-0,024	-0,01879
tpopurb	0,18***	0,0597***	-0,059***	-0,0318*	0,121***	0,02788
dom_kmq	-0,0001	-0,0002	0,0002	0,0003	0,0001	0,00003
est_saude	-0,0026	-0,0017	0,005	-0,0039	0,002	-0,00567
dummy_capital	-3,632***	-1,6937**	1,152	0,4412	-2,48	-1,25248
dummy_rm	-3,807***	-1,0109	-0,693	-1,5515	-4,499***	-2,56243*

Os efeitos marginais para a população de todas as idades do modelo SDMT também apresentaram magnitudes superiores em relação ao modelo SDM. Porém, na coluna dos efeitos diretos, comparando esses dois modelos, duas das variáveis de interesse perderam significância: *sbceuab* e *dfrud*. Apesar disso, a variável *dfrud* foi significativa para os efeitos indiretos e, para esses mesmos efeitos, a variável *sbceuab* manteve-se significativa, embora com maior magnitude. A variável *dcist* tornou-se significativa nos efeitos totais, dado que no SDM não era.

Os três efeitos (direto, indireto e total) para a variável *sbceuab* vão ao encontro ao resultado de Barreto *et al* (2007). Os autores apontam que ter banheiro no interior do domicílio (80% dos domicílios da cidade de Salvador possuem banheiro) não explica a redução da diarreia. O aumento da cobertura de esgotamento sanitário de toda a vizinhança foi capaz de reduzir a poluição fecal do domínio público e, portanto, explicar a queda nesse indicador de morbidade. Ou seja, segundo os autores, a prevenção da transmissão da diarreia se deu principalmente na vizinhança e não no ambiente doméstico dos domicílios cujos fatores de risco apresentaram pequena importância. As contribuições do estudo de tais autores e dos resultados apresentados acima encontram-se na amplitude espacial das externalidades geradas aos vizinhos pela ausência de esgotamento sanitário. Caso esse efeito espacial não seja considerado, o impacto positivo do saneamento inadequado sobre o indicador de morbidade será subestimado.

Além destes, outros autores tais como Bleakley (2007), Miguel e Kremer (2004), Esrey (1991), Kassouf (1994), Moraes *et al.* (2003), defendem a relação positiva entre variáveis de saúde (na maior parte medida pela incidência de doenças diarreicas) e o acesso aos serviços de esgotamento sanitário⁴³.

O modelo teórico desenvolvido na primeira seção deste capítulo também corrobora essa associação. Uma das variáveis que pode determinar a produção de saúde dos indivíduos é aquela relacionada ao ambiente em que o mesmo reside, ou seja, as condições de infraestrutura de saneamento básico do domicílio (" ST_i ", conforme equação 5.3). Em relação a essa variável, a implicação do modelo é que, por ser uma variável exógena, pode melhorar ou retardar a saúde (BERNDT, 2000). No contexto dos resultados apresentados, pelo fato das variáveis de condições de saneamento desse estudo se tratarem de variáveis "adversas" para a saúde humana, o deslocamento da curva de eficiência marginal do investimento em saúde (EMH), conforme Figura 5.2, desloca-se para a esquerda gerando queda da quantidade ótima de capital de saúde (H_i^*).

Em relação ao efeito total, algumas variáveis de controle tornaram-se importantes para explicar a taxa de incidência de internações por DRSAI: *gini*, *tpopurb* e *dummy_rm*. As duas primeiras apresentaram coeficientes positivos e significativos e, a última, coeficiente negativo

⁴³ Alguns desses estudos indicaram o impacto de saneamento em educação, mas por meio do estado de saúde do indivíduo que deteriora o perfil de morbidade dos indivíduos tais como Bleakley (2007), Miguel e Kremer (2004) e Kassouf (1994).

e significativo. Quanto maior o grau de urbanização do distrito em questão e dos distritos vizinhos, maior a incidência de internações por DRSAI desse distrito. A maior concentração da renda, medida pelo coeficiente Gini, leva a maior incidência de DRSAI em valores significativos e de grande magnitude. Por fim, o efeito da *dummy* de região metropolitana se mostrou significativo e negativo, como era esperado, dado que em tais regiões a provisão dos serviços de saneamento e saúde tende a ser melhor e mais homogênea. Na Tabela 5.17, a mesma comparação foi repetida para a população de 0 a 14 anos de idade para o mesmo modelo 1.

Tabela 5.17 – Comparação entre efeitos marginais (SDM e SDMT)

Inc DRSAI (de 0 a 14 anos)	ED		EI		ET	
	SDMT	SDM	SDMT	SDM	SDMT	SDM
dcist	0,25***	0,1049*	0,215	0,0137	0,465***	0,1185*
sbceuab	-0,267	0,1827	5,733***	2,0347***	5,466***	2,2174***
dfrud	0,006	0,0381***	0,049**	-0,0055	0,055***	0,0325***
rendapc	-0,002	-0,0001	-0,003	-0,0019	-0,004	-0,002
gini	43,908***	6,1355	20,266	13,0663**	64,174***	19,2018***
mu_resp_alfab	0,324***	0,0559**	-0,368***	-0,0826**	-0,044	-0,0267
tpopurb	0,426***	0,1075***	-0,142***	-0,0516***	0,284***	0,0559***
dom_kmq	-0,001	-0,0015***	0,001***	0,0024***	0,001***	0,0009
est_saude	0,005	0,0009	0,002	-0,0153*	0,007	-0,0144
dummy_capital	-12,8	-7,4961***	10,169**	6,6577**	-2,631	-0,8384
dummy_rm	-9,815***	-3,4182	0,138	-1,4363	-9,677***	-4,8545***

A comparação dos efeitos marginais dos dois modelos (SDM e SDMT) seguiu a mesma tendência da análise dos coeficientes realizada na tabela anterior com a população de todas as idades no tocante à magnitude e significância dos coeficientes. A única diferença deve-se ao efeito total da variável *dcist* que se manteve significativa no SDMT, pois já era no modelo SDM. Logo, a hipótese de que a população de zero a catorze anos é mais atingida por DRSAI por estar mais sujeita às doenças infecto-parasitárias transmitidas pela água, esgoto não tratado ou pelos insetos vetores foi verificada pela magnitude dos coeficientes apresentados.

Esse resultado está em consonância com o estudo de Geruso e Spears (2015) que, apesar de ser aplicado para outro país (Índia) e a variável de análise se tratar da mortalidade infantil, investigam as externalidades do esgoto a céu aberto e mostram que é mais provável que crianças muçulmanas sobrevivam até seu primeiro aniversário em comparação às crianças hindus, pois a vizinhança das crianças muçulmanas é formada em grande parte por vizinhos

muçulmanos, que são menos propensos a defecar a céu aberto⁴⁴. As próximas estimações foram realizadas para o modelo 2 e, inicialmente, para a população de 0 a 14 anos. A Tabela 5.18 compara os sinais e significância dos três modelos, MQO, SDM e SDMT, como realizado anteriormente.

Tabela 5.18 – Comparação entre modelos: MQO, SDM e SDMT (de 0 a 14 anos)

Variáveis	MQO	SDM	SDMT
esg_nao_ceuab	-0,0356***	-0,0056	0,0044
ag_nao_ceuab	-0,0486***	-0,0588***	-0,0583***
lixo_mor_inad	0,0517	-0,1026	-0,5544
d_nbueiro	0,0154	-0,0182	0,0107
rendapc	-0,0009	0,0002	-0,0009
gini	16,2374***	1,6238	12,8452***
mu_resp_alfab	-0,038**	0,0919***	-0,0089
tpopurb	0,1385***	0,1557***	0,1417***
dom_kmq	-0,0004	-0,0015***	0,0003
est_saude	0,0027	0,0001	0,0065*
dummy_capital	-3,1044**	-7,683***	-0,8451
dummy_rm	-5,4576***	-3,5105	-2,3734**
Wesg_nao_ceuab	-	-0,0347*	-0,0495**
Wag_nao_ceuab	-	0,0304	-0,0172
Wlixo_mor_inad	-	0,9136	0,1172
Wd_nbueiro	-	0,0523**	-0,0002
Wrendapc	-	-0,0007	-0,0028
Wgini	-	10,443	57,3316***
Wmu_resp_alfab	-	-0,1361***	-0,2743***
Wtpopurb	-	-0,0888***	0,3249***
Wdom_kmq	-	0,0026***	-0,0014**
West_saude	-	-0,0129*	-0,0055
Wdummy_capital	-	6,5008	-4,5744*
Wdummy_rm	-	-1,0463	-10,8401***
Constante	-	-0,8823	-42,1422***
Rho	-	0,14212***	0,6237***

Nesse caso, não há diferenças significativas, com exceção do coeficiente de “*Wd_nbueiro*” que perdeu significância no modelo SDMT. Os efeitos marginais para essa população de 0 a 14 anos do modelo 2 são apresentados na tabela 5.19.

⁴⁴ Os autores solucionam um "puzzle" dos dados indianos: é mais provável que crianças muçulmanas sobrevivam até seu primeiro aniversário do que crianças hindus, mesmo que as primeiras tenham menor riqueza, consumo, frequência escolar e pior acesso aos serviços públicos em comparação à maioria hindu. Isso ocorre por dois motivos. Um deles deve-se ao fato da população hindu considerar "impuro" o uso de banheiros e fossas, defendendo que defecar a céu aberto faz parte de um estilo de vida: mais da metade da população indiana, 600 milhões de pessoas, tem essa prática. O segundo motivo deve-se ao efeito vizinhança já relatado.

Tabela 5.19 – Comparação entre efeitos marginais (SDM e SDMT)

INC DRS AI (de 0 a 14 anos)	ED		EI		ET	
	SDMT	SDM	SDMT	SDM	SDMT	SDM
esg_nao_ceuab	-0,0032	-0,002	-0,1174	-0,0443**	-0,1206	-0,0463***
ag_nao_ceuab	-0,0674***	-0,0663***	-0,1339	0,0429	-0,2013**	-0,0234
lixo_mor_inad	-0,5859	-0,0547	-0,5616	0,9661	-1,1476	0,9114
d_nbueiro	0,0117	-0,0211	0,0162	0,0683***	0,0279	0,0472**
rendapc	-0,0015	0,0002	-0,0085	-0,0006	-0,01	-0,0004
gini	23,4857***	4,1053	163,1399***	10,7831	186,6257***	14,8883**
mu_resp_alfab	-0,0533**	0,0576**	-0,6952***	-0,1076***	-0,7485***	-0,05*
tpopurb	0,2089***	0,1644***	1,0317***	-0,0991***	1,2406***	0,0653**
dom_kmq	0,0002	-0,0015***	-0,003	0,0026***	-0,0029	0,0012*
est_saude	0,0064	0,0001	-0,0028	-0,0131	0,0036	-0,013
dummy_capital	-1,6916	-8,2409***	-12,7709	6,9385**	-14,4625	-1,3024
dummy_rm	-4,3958***	-3,7947*	-30,8404***	-1,3476	-35,2362***	-5,1423***

Os efeitos das variáveis relativas à lixo e drenagem de águas pluviais passaram a ser não significantes, na estimação do SDMT, pois os efeitos indiretos e totais eram significantes no modelo SDM. Os resultados do modelo SDMT apontam que localidades que têm acesso à água por rede geral e não apresentam em suas condições de entorno esgoto a céu aberto têm menores taxas de incidência de internações, considerando o efeito direto e total.

Esse resultado pode estar relacionado ao fenômeno apontado por Phillippi Jr. e Mallheiros (2005). Segundo os autores, o abastecimento de água potável, quando não acompanhado da implantação conjunta e concomitante do esgotamento sanitário, pode vir a ter um saldo negativo para os vizinhos, em face do lançamento de efluentes sem tratamento ou com tratamento inadequado que poder ser devolvido ao meio ambiente em condições nocivas que ainda apresentem agentes patogênicos. A variável “ag_nao_ceuab” pode ser entendida como uma variável que integre essas duas características destacadas pelos autores. Duflo et al. (2015) também defendem a provisão integrada de tais serviços como a forma mais eficiente de afetar os indicadores de saúde.

Os resultados da estimação para a população de todas as idades para o segundo modelo, na versão SDMT, encontram-se na Tabela 5.20. No entanto, não foi possível fazer comparação com o modelo SDM da seção anterior, pois este foi estimado com a matriz inversa da distância.

Tabela 5.20 – Efeitos marginais do modelo SDMT com matriz W contígua

INC DRSAl - pop todas idades	ED	EI	ET
esg_nao_ceuab	-0,0004	-0,055	-0,056
ag_nao_ceuab	-0,02683***	-0,139***	-0,166***
lixo_mor_inad	-0,07773	-1,204	-1,281
d_nbueiro	-0,00478	-0,076	-0,081
rendapc	0,00007	0,001	0,001
gini	3,15531	60,557***	63,712***
mu_resp_alfab	-0,04657***	-0,464***	-0,51***
tpopurb	0,05809***	0,565***	0,623***
dom_kmq	-0,00004	-0,001	-0,001
est_saude	0,00004	-0,002	-0,002
dummy_capital	-0,65814	-8,036	-8,694
dummy_rm	-1,3788***	-13,591***	-14,97***

Tanto os resultados da Tabela 5.19 quanto os da Tabela 5.20 apontam que localidades que têm acesso à água por rede geral e não apresentam em suas condições de entorno esgoto a céu aberto têm menores taxas de incidência de internações. Novamente, a magnitude do coeficiente para o efeito total é maior para a população de 0 a 14 anos.

Neste caso também, as variáveis relativas ao acesso inadequado à drenagem de águas pluviais e ao lixo não foram significativas ao se controlar a censura dos dados. De fato, a maior parte dos estudos que relatam queda da taxa de incidência de diarreia, por exemplo, no Brasil e em outros países do mundo, apontam ser o esgotamento sanitário ou o processo de distribuição e tratamento de água como as ações de saneamento mais eficientes na promoção da saúde.

Por outro lado, a variável relativa à escolaridade das mulheres chefes de família (“mu_resp_alfab”) se mostrou relevante para diminuir a taxa de incidência de internações por DRSAl, como prevê a literatura. O coeficiente de Gini apresentou significância estatística na grande parte dos resultados apresentados nesse estudo. Ou seja, quanto maior a concentração de renda no distrito maior é a incidência de DRSAl no mesmo, indicando que variáveis relacionadas à distribuição de renda nos distritos pode ser uma política pública importante para a diminuição da incidência de DRSAl.

5.8 Recomendação de política pública

Essa seção tem o objetivo de fazer uma recomendação de política pública: investir em saneamento básico significa economia de recursos para a saúde. Nesse sentido, a partir dos coeficientes estimados, será mensurado o quanto o setor de saúde economizaria caso não precisasse gastar com as internações por DRSAl que seriam evitadas pelo maior acesso da

população aos serviços de saneamento básico. A economia de recursos da saúde poderia ser destinada a novos investimentos no setor de saneamento, de forma a acelerar o processo de universalização e diminuir o tempo de exposição das pessoas à falta de tais serviços essenciais de infraestrutura.

De acordo com o Plano Nacional de Saneamento Básico do Ministério das Cidades, a universalização dos serviços de água e esgotos custará ao país R\$ 303 bilhões em 20 anos. Ou seja, R\$ 15 bilhões por ano⁴⁵. A partir das informações do valor gasto por cada internação na base de dados do DATASUS, foi possível calcular o custo médio de cada internação no ano de 2010. Além disso, de posse do valor médio para a taxa de incidência de internações por DRSAI⁴⁶, do número total de casos ocorridos, da população residente no país em cada faixa etária e dos valores estimados dos efeitos marginais⁴⁷ realizou-se a estimativa de quanto o setor de saúde gasta em decorrência do aumento de um ponto percentual de condições adversas de saneamento básico e, analogamente, quanto economiza por proporcionar as condições adequadas. A Tabela 5.21 apresenta os dados que foram utilizados nas estimativas dos custos das internações.

Tabela 5.21 – População, incidência DRSAI, número de casos e custos (2010)

Variáveis	Todas as idades	População de 0 a 14 anos
População Residente (habitantes)	190.308.447	45.838.773
Inc DRSAI Média (internações/1000 habitantes)	3,15	6,58
Número de casos (internações)	599.499	302.024
Custo Total (R\$)	209.599.214	121.254.421
Custo por internação (R\$/internação)	349,62	401,47

A partir dos coeficientes dos efeitos marginais obtidos do modelo SDMT foi possível estimar o número incremental de casos decorrentes do aumento de um ponto percentual nas condições inadequadas de saneamento (“dcist”, “sbceuab” e “dfrud”). Analogamente, foi possível encontrar a diminuição do número de casos pelo aumento de um ponto percentual nas condições adequadas de saneamento básico (“ag_nao_ceuab”). As etapas desse cálculo estão descritas na seção 10 do Apêndice. A Tabela 5.22 apresenta os coeficientes utilizados no

⁴⁵ Devido à magnitude dos recursos, esse aporte não deve vir somente do governo federal, mas também da iniciativa privada e outras formas de parcerias.

⁴⁶ Esse valor foi obtido de acordo com a equação 3.1.

⁴⁷ Devido ao efeito positivo do parâmetro autoregressivo para a variável dependente defasada, o efeito total leva em consideração o impacto da variável de saneamento (independente) sobre a variável de internações por DRSAI (dependente) incluindo o efeito *spillover* das internações dos distritos vizinhos e o efeito do nível de acesso aos serviços de saneamento dos mesmos.

cálculo da nova taxa de incidência, a diferença no número de casos em termos do aumento de um ponto percentual nas condições de saneamento (em relação à taxa média) e o valor gasto com as internações do número incremental de casos (calculados a partir do custo por internação expresso na Tabela 5.21).

Tabela 5.22 – Análise do novo número de casos e valores gastos com as internações

Variáveis	População de todas as idades			
	Coefficientes	Nova taxa inc	Numero casos (var)	Valores Totais (R\$)
dcist	0,217***	3,367	41.270	14.428.820,32
sbceuab	3,281***	6,431	624.375	218.296.327,32
dfrud	0,035***	3,185	6.633	2.319.196,86
ag_nao_ceuab	-0,166***	2,984 -	31.619 -	11.054.759,95
Variáveis	População de 0 a 14 anos			
	Coefficientes	Nova taxa inc	Numero casos (var)	Valores Totais (R\$)
dcist	0,465***	7,045	20.910	8.394.858,78
sbceuab	5,466***	12,046	250.150	100.428.364,43
dfrud	0,055***	6,635	2.116	849.620,37
ag_nao_ceuab	-0,2**	6,38	9.573 -	3.843.149,87

Tanto para a população de todas as idades quanto para a população de 0 a 14 anos, o aumento de um por cento na porcentagem dos domicílios que não têm banheiro e estão expostos a condições de esgoto à céu aberto (“sbceuab”) implica em aumento de 104% e 83,06%, respectivamente, na taxa de incidência média de internações por DRSAl.

Analogamente, se fossem implementadas políticas de mitigação das condições inapropriadas de saneamento, iniciando por diminuir em um por cento a porcentagem de domicílios que não têm banheiro e estão expostos a condições de esgoto a céu aberto o setor público da saúde teria uma economia de R\$ 218.296.327, por não ter que incorrer em gastos com as internações por doenças relacionadas a saneamento para a população de todas as idades.

Quando se consideram as estimativas do Ministério das Cidades para universalização do saneamento, de R\$15 bilhões por ano em média, esse valor corresponde somente a 1,45% do total anual. No entanto, quando se consideram os gastos federais com saneamento que vêm sendo realizados nos últimos anos, essa economia é bastante relevante. Em 2010, segundo dados do IPEA, apresentadas no capítulo 2 na Figura 2.4, o gasto federal com saneamento foi R\$ 4.994.000.000. Ao considerar esse valor, a contribuição da economia do setor público da saúde corresponde a 4,4% dos gastos federais com saneamento.

É importante considerar que esse exercício leva em conta somente o serviço de maior impacto para a taxa de incidência de internações por DRSAl representado pela variável “sbceuab”. A

política integrada dos serviços de saneamento (tais como abastecimento de água por rede geral, coleta e tratamento de esgoto, drenagem de águas pluviais e coleta de lixo) traria maior impacto e, portanto, maior economia de recursos para o setor da saúde que poderiam ser destinados para acelerar o cumprimento das metas de universalização do acesso a esses serviços.

6 CONCLUSÃO

No Brasil, as elevadas taxas de incidência de internações por Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI), os insatisfatórios indicadores educacionais e de acesso aos serviços de saneamento básico apontam para uma realidade socioeconômica desigual e concentrada através das regiões. Apesar da evolução positiva de tais indicadores ao longo do tempo, os desafios a serem superados em tais áreas ainda são de grande magnitude. A identificação do efeito do acesso aos serviços de saneamento básico sobre os indicadores educacionais, por meio do impacto que tal acesso gera no perfil de morbidade da população em idade escolar, é uma contribuição dessa tese para superar tais desafios.

A partir de dados municipais e distritais, o efeito do saneamento sobre educação foi obtido em duas etapas. Na primeira delas, os dados municipais foram utilizados para identificar o efeito do maior percentual de pessoas que residem em domicílios que possuem banheiro e água canalizada por rede geral de distribuição sobre indicadores educacionais da população de 6 a 14 anos. A identificação desse efeito se valeu de análise não espacial e espacial para dados em painel (2000 e 2010) em dois estágios com variável instrumental.

A segunda etapa, que utilizou dados por distrito, confirmou a hipótese que dá suporte aos resultados da primeira etapa: o efeito do saneamento sobre educação ocorre via efeito do saneamento sobre a saúde. Ou seja, o estado de saúde debilitado das crianças infectadas por agentes patogênicos causadores de doenças relacionadas à falta de saneamento é o motivo que as impossibilita de obter bons resultados escolares. Nesse sentido, a estimação do efeito do saneamento sobre a saúde foi realizada para o ano de 2010 e se valeu de análise espacial com e sem o controle da censura dos dados. A utilização de dados distritais foi necessária para justificar o efeito espacial de transbordamento de tais agentes patogênicos entre as regiões.

Os resultados finais da identificação do efeito do saneamento sobre educação, após todos os procedimentos de estimação serem realizados, indicam que o efeito do aumento de 1% no acesso a saneamento está associado a um aumento de 0,11 pontos percentuais na taxa de frequência escolar, a uma queda de 0,31 pontos percentuais na taxa de distorção idade-série e a também a uma queda de 0,12 pontos percentuais na taxa de abandono escolar. O alcance de tais resultados se valeu de vários passos.

Em primeiro lugar, os dados municipais utilizados foram obtidos de diversas fontes para os anos 2000 e 2010. A análise não espacial de dados em painel com modelo de efeitos fixos, ao utilizar variável instrumental em dois estágios, indicou que saneamento importa para educação. Um aumento de 1% no acesso aos serviços de saneamento básico está associado a um aumento de 0,11 pontos percentuais na taxa de frequência escolar da população de 6 a 14 anos e a diminuições de 0,49 pontos percentuais na taxa de abandono escolar e de 0,96 pontos percentuais na taxa de distorção idade-série no ensino fundamental para essa mesma população.

Os testes dos resíduos do modelo não espacial, que geraram os resultados acima, foram implementados para identificar possíveis efeitos espaciais. Os resultados de tais testes indicaram erros autocorrelacionados espacialmente para as variáveis de distorção idade-série e abandono escolar e evidências da presença de autocorrelação espacial na forma de defasagem espacial da variável dependente para as três variáveis de educação (incluindo frequência escolar). Dado que ignorar os efeitos espaciais pode gerar viés e ineficiência das estimativas dos coeficientes, o método mais adequado para o controle dos efeitos espaciais, neste caso, é o método das Variáveis Instrumentais ou dos Momentos Generalizados (GMM). A vantagem deste método é estimar consistentemente os modelos espaciais que apresentam variáveis endógenas entre as variáveis explicativas, além da defasagem espacial, e aqueles que também apresentam erros espacialmente correlacionados.

Na quase totalidade dos casos, os Modelos de Erro Autoregressivo Espacial (SEM), de Defasagem Espacial (SAR) e o de Defasagem Espacial com Erro Autoregressivo Espacial (SAC) estimados por GMM apresentaram menor magnitude para as estimativas dos coeficientes em relação aos modelos não-espaciais, igualmente à evidência obtida por Baylis et al. (2011). Porém, por ser área recente de estudo e poucos trabalhos terem usado esse método, não foi possível comparar os efeitos marginais dos modelos não espaciais com os modelos SAR e SAC, por uma limitação computacional do método⁴⁸. Dessa forma, para esses dois últimos modelos não foi factível precisar o impacto do saneamento sobre educação somente com os resultados dos coeficientes, pois o parâmetro positivo e significativo (ρ), que mede o *spillover* das variáveis de educação, é capaz de potencializar o impacto de uma

⁴⁸Para o GMM, o cálculo do efeito marginal quando existe endogeneidade entre as explicativas não está implementado no *software* R ou em outro programa estatístico.

variável independente sobre a variável dependente no efeito marginal. Com as informações obtidas, foi possível concluir somente sobre a direção do sinal.

Frente a essa situação, a escolha pelo modelo espacial recaiu sobre o modelo de erro (SEM) para as variáveis de abandono escolar e distorção idade-série. Pelo fato desse modelo não possuir o termo que determina o *feedback* na variável dependente, os parâmetros estimados sob tal modelo possuem interpretação direta. Além disso, teoricamente é razoável supor, pela literatura apresentada, que os indicadores educacionais apresentam respostas distintas de acordo com as diversas realidades socioeconômicas dos estados e regiões. Apesar dos problemas causados pela heterogeneidade espacial poderem ser abordados por técnicas econométricas não espaciais, considerar a estrutura espacial dos dados torna os procedimentos de estimação mais eficientes. Além disso, vários sinais e significância dos coeficientes das variáveis explicativas deixaram de ser inconclusivos, como eram na estimação do modelo não espacial.

Para a variável de frequência escolar, o modelo escolhido é o não-espacial, pelo fato de não existir evidências de efeitos espaciais sobre a forma de autocorrelação dos resíduos. Logo, o efeito do aumento de 1% no acesso a saneamento está associado ao aumento de 0,11 pontos percentuais na taxa de frequência escolar e a quedas de 0,31 e 0,12 pontos percentuais na taxa de distorção idade-série e abandono escolar, respectivamente.

No tocante à segunda etapa, que mensura a associação do efeito do saneamento sobre a saúde, as variáveis de saneamento utilizadas representam o acesso inadequado aos serviços de água e esgoto e a variável de saúde é representada pela taxa de incidência de internações por DRSAI por mil habitantes. O modelo de Grosman (1972) foi utilizado como arcabouço teórico para a escolha das variáveis dos modelos que foram estimados para populações de duas faixas etárias diferentes: a de todas as idades e a de 6 a 14 anos.

Os resultados indicam, para a população de todas as idades, que o impacto total médio do aumento de 1% na parcela de domicílios servidos por água da chuva armazenada em cisterna ou caixa de cimento é de 0,217 casos por mil habitantes. Quando se considera a porcentagem dos domicílios que têm banheiro conectado a uma fossa rudimentar, esse impacto é de 0,035 casos por 1000 habitantes. Por outro lado, o impacto total médio de 1% na porcentagem de domicílios particulares permanentes conectados à rede de distribuição de água e que, além

disso, na quadra onde estão localizados é inexistente condições de esgoto a céu aberto é de uma queda de 0,166 casos por mil habitantes. Por fim, a variável de maior impacto total médio refere-se à porcentagem de domicílios que não têm banheiro e, além disso, possuem presença de esgoto a céu aberto em suas condições de entorno. O aumento em uma unidade percentual nessa variável gera aumento de 3,281 na variável de saúde, a taxa de incidência de internações por DRSAI.

Neste estudo, as informações da variável dependente não são observadas por causa de um tipo de censura relacionado a não ocorrência de casos de internações por DRSAI em alguns distritos. Ou seja, são distritos saudáveis que não registraram nenhuma internação por DRSAI. A obtenção dos resultados relatados acima envolveu, inicialmente, a estimação de vários modelos espaciais, sem que fossem controlados os zeros da amostra, para nortear a escolha do melhor modelo e para verificar se os resultados iriam mudar ao se utilizar diferentes matrizes de ponderação. Nestas estimações, o modelo Spatial Durbin (SDM) com a matriz de contiguidade se mostrou o mais adequado, na maior parte dos casos. Para controlar os zeros mencionados acima, foi estimado um modelo Spatial Durbin Tobit Espacial (SDMT) por meio da abordagem Bayesiana.

O efeito de transbordamento espacial da taxa de incidência de internações por DRSAI foi positivo e significativo, com maior magnitude para o modelo SDMT (0,2) em comparação ao modelo SDM (0,14). Esse resultado era esperado pela própria definição de doenças infectocontagiosas das DRSAI. A transmissão dos agentes patogênicos causadores de tais doenças pode ocorrer tanto pelo deslocamento dos insetos de um distrito para o outro quanto pela água contaminada dos rios e mares que atravessam diversas localidades, gerando o efeito de *spillover* espacial, de tal forma que a incidência de DRSAI em um determinado distrito seja influenciada pela média dessa taxa nos distritos vizinhos. Além disso, as condições de infraestrutura de saneamento básico da vizinhança de um determinado distrito também são capazes de afetar essa variável de saúde, devido à importância de algumas variáveis explicativas defasadas espacialmente, conforme o modelo Durbin prevê. É o caso de distritos que lançam águas servidas sem tratamento, dejetos e outros resíduos coletados em suas regiões para fora de seus limites territoriais.

Em relação à sensibilidade dos efeitos para as populações das duas diferentes faixas etárias, pela magnitude das estimativas dos parâmetros dos efeitos marginais, a população de zero a

14 anos é a mais atingida pelo acesso a condições inadequadas dos serviços de saneamento básico, o que confirma uma das hipóteses estabelecidas nesse estudo, de que essa população é a mais frágil e a mais suscetível a contrair as DRSAl.

Nesse sentido, outra evidência que pode ser destacada refere-se ao modelo não espacial da primeira etapa, quando se desagrega o ensino fundamental em dois ciclos, sendo o primeiro formado pela população de 6 a 10 anos e o segundo constituído pela população de 10 a 14 anos. Comparativamente, o primeiro ciclo apresenta maior redução na taxa de abandono escolar (0,52 pontos percentuais, contra 0,47 pontos percentuais do segundo ciclo). Para a taxa de distorção idade-série, os resultados foram similares. Em que pese a pequena diferença quantitativa, essa evidência vai ao encontro do resultado anterior. A população mais jovem, de seis a dez anos, está mais sujeita às DRSAl do que os alunos mais velhos.

Por fim, a recomendação de política pública apontou que, se fossem implementadas políticas de mitigação das condições inapropriadas de saneamento, iniciando por diminuir em um por cento a porcentagem de domicílios que não têm banheiro e estão expostos a condições de esgoto a céu aberto, o setor público da saúde teria uma economia de R\$ 218.296.327, por não ter que incorrer em gastos com as internações por doenças relacionadas a saneamento para a população de todas as idades. Ao se considerar os gastos federais com saneamento que vêm sendo realizados nos últimos anos, essa economia é bastante relevante e corresponde a 4,4% do gasto total para 2010.

Nesse contexto, políticas públicas voltadas para a mitigação das condições precárias de saneamento básico devem levar em consideração os impactos negativos que tais condições exercem sobre a saúde dos indivíduos, em particular das crianças, e os impactos educacionais decorrentes das dificuldades que as mesmas têm de se manter na escola, avançar nos estudos e não desistir dos mesmos, por terem a saúde comprometida em razão de condições impróprias de abastecimento de água, coleta e tratamento de esgoto.

REFERÊNCIAS

- ALBERNAZ, A.; FERREIRA, F.H.G.; FRANCO, C. *Qualidade e equidade no ensino fundamental brasileiro*. **Pesquisa e Planejamento Econômico**, v.32, n.3, p.453-476, 2002.
- ALMEIDA, E. *Econometria Espacial Aplicada*, Campinas, SP: Editora Alínea, 2012.
- ALVES, D.; BELLUZZO, W. *Infant mortality and child health in Brazil*. **Econ Hum Biol**, 2(3): 391-410, 2004.
- ANSELIN, L. *Spatial econometrics: methods and models*. Boston: **Kluwer Academic**, 1988.
- ANSELIN, L. *Local indicators of spatial association – LISA*. **Geographical Analysis**, v.27, n.2, p. 93-115, 1995.
- ANSELIN, L.; BERA, A. *Spatial dependence in linear regression models with an introduction to spatial econometrics*. In: ULLAH A.; GILES D. E. (eds). **Handbook of applied economic statistics**, Nova York: Marcel Dekker, p. 237-289, 1998.
- ANDREAZZI, M. A. R.; BARCELLOS, C.; HACON, S. *Velhos indicadores para novos problemas: a relação entre saneamento e saúde*. **Revista Panam Salud Publica**, n. 22, 2007.
- ANSELIN, L. *Spatial econometrics: methods and models*. Boston: Kluwer Academic, 1988.
- ANSELIN, L.; BERA, A. *Spatial dependence in linear regression models with an introduction to spatial econometrics*. In: ULLAH A.; GILES D. E. (eds). **Handbook of applied economic statistics**, Nova York: Marcel Dekker, p. 237-289, 1998.
- ARAÚJO, B. S.; SANTOS, J. F.; NEIVA, T. S.; MAGALHÃES-FILHO, R. R. e RIOS, D. S. *Associação das parasitoses intestinais com anemia e eosinofilia em escolares do povoado de Matilha dos Pretos, Feira de Santana, Bahia, Brasil*. **Sitientibus Série Ciências Biológicas**, 9 (1): 3-7, 2009.
- BALTAGI, B. H.; Liu L. *Instrumental Variable Estimation of a Spatial Autoregressive Panel Model with Random Effects*, **Economics Letters**, 111, 135-137, 2011.
- BANERJEE, A. V.; DUFLO, E. A. *Economia dos Pobres*, 1 ed, 2011.
- BARRETO, M. L. et al. *Impact of a Citywide Sanitation Program in Northeast Brazil on Intestinal Parasites Infection in Young Children*, **Environ Health Perspect**, Nov; 118(11): 1637–1642, 2010.
- BARROS, R. P. de; MENDONÇA, R. Os determinantes da desigualdade no Brasil. A economia brasileira em perspectiva: 1996. Rio de Janeiro, **IPEA**, v. 2, p. 421- 474, 1996.
- BARROS, R.P. de; MENDONÇA, R.; SANTOS, D.D.; QUINTAES, G. Determinantes do desempenho educacional no Brasil. Rio de Janeiro: **IPEA**, 33p., Texto para discussão, n. 834, 2001.

BARUFI, A. M.; HADDAD, E.; PAEZ, A. *Infant mortality in Brazil, 1980-2000: a spatial panel data analysis*, **BMC Public Health**, 17, 12:181, 2012.

BAYLIS, K, PAULSON, N. E PIRAS, G. *Spatial Approaches to Panel Data in Agricultural Economics: A Climate Change Application*, **Journal of agricultural and applied economics**, v.3, n.43, p.325-338, 2011.

BERNDT, E.R.; Cutler, D.; Frank, R.G.; Griliches, Z.; Newhouse, J.; Triplett, J.E. *Medical care prices and output*. In: **Handbook of Health Economics**, vol. 1. Amsterdam: Elsevier, Cap. 03, p. 119-180, 2000.

BISHOP, J. *Incentives for learning: Why American high school students compare so poorly to their counterparts overseas*. Ithaca, NY: Cornell University, **Center for Advanced Human Resource Studies**, 1989.

BLEAKLEY, H. *Disease and Development: Evidence from Hookworm Eradication in the American South*, **The Quarterly Journal of Economics**, 122 (1): 73-117, 2007.

BLEAKLEY, H. *Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure*, **American Economic Journal: Applied Economics** v.2 (2), 1-45, 2010.

BORJAS, G. *Ethnicity, Neighborhoods, and Human-Capital Externalities*. **American Economic Review**, v.3, n.85, p.365-390, 1995.

BRUECKNER, J. K. *Strategic Interaction Among Governments: An Overview of Empirical Studies*. **International Regional Science Review**, Champaign, IL, 2003.

BRISCOE, J. *Abastecimiento de agua y servicios de saneamiento; su funcion en la revolucion de la supervivência infantil*. **Boletín de la Oficina Sanitária Panamericana** 103 (4): 325-339, 1987.

BRITO, L. L.; Mauricio, L. B.; SILVA, R. C. R.; ASSIS, A. M. O.; REIS, M. G.; PARRAGA, I. e BLANTON, R. E. *Fatores de risco para anemia, por deficiência de ferro em crianças e adolescentes parasitados por helmintos intestinais*. **Revista Panam Salud Publica**, Pan AM J Public Health, 14(6), 2003.

CAIRNCROSS, S. *Water supply and sanitation: an agenda for research*. **Journal of Tropical Medicine and Hygiene**, v.92, p.301-314, 1989.

CAIRNCROSS, S.; FEACHEM, R. G. *Environmental health engineering in the tropics: an introductory text*. John Wiley & Sons, Chichester, 1990.

CAIRNCROSS, S.; R. FEACHEM. *Environmental Health Engineering in the Tropics*. 2. ed. Chichester, U.K.: John Wiley & Sons, 1993.

CARD, D.; KRUEGER, A. B. *Does school quality matter? returns to education and the characteristics of public schools in the United States*, **The Journal of Political Economy**, v. 100, n. 1, 1-40, 1992.

CARNEIRO, P., CUNHA, F. e HECKMAN, J. *Interpreting The Evidence of Family Influence on Child Development*, **Economics of early childhood conference**, Minneapolis Fed, 2003

CARVALHO, S.; FIRPO, S. *O regime de Ciclos de aprendizagem e a Heterogeneidade de seus efeitos sobre a proficiência dos alunos*. **Revista de Economia Aplicada**, v. 18, n. 2, pp. 199-214, 2014.

CHAGAS, ANDRÉ L. S.; ALMEIDA, ALEXANDRE N.; AZZONI, CARLOS R. *Sugar Cane Burning and Human Health: a Spatial Difference-in-Difference Analysis*, Department of Economics – FEA/USP. **Working Paper Series** N° 2014-20. 2014.

COLEMAN, J. S. et al. *Equality of educational opportunity*. Washington, 1966.

CRAGG, J. G., e DONALD, S. G., *Testing Identifiability and Specification in Instrumental Variable Models*, **Econometric Theory**, 9, 222–40., 1993.

CURI, A.; MENEZES-FILHO, N. *A relação entre educação pré-primária, salários, escolaridade e proficiência escolar no Brasil*, **Estudos Econômicos**, São Paulo, vol.39 no.4, São Paulo, 2009.

CUTLER, D., & MILLER, G. *The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States*. **Demography**, 42 (1), 1-22, 2005.

CVJETANOVIC, B. *Health effects and impact of water supply and sanitation*. **World Health Statistics Quarterly**, 39: 105-117, 1986.

CARD, D. *Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems*. **Econometrica**, Vol. 69, No. 5, p. 1127-1160, 2001.

DAVIDSON, R. and MACKINNON, J. G., *Econometric Theory and Methods*, Oxford University Press, 2004.

DEATON, A. *Health inequality and economic development*. **Journal of Economic Literature**, 41: 113-158, 2003.

DOMENE, S. M. A. *O Papel do ferro sobre a nutrição e a saúde*. Serviço de Informação da Carne, **Comitê Técnico do SIC**, PUC, Campinas, 2004.

DUFLO, E. *Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment*. **American Economic Review**, Vol. 91 (4), pp 795-813, 2001.

DUFLO, E.; MICHAEL GREENSTONE, RAYMOND GUITERAS, AND THOMAS CLASEN. *Toilets Can Work: Short and Medium Run Health Impacts of Addressing Complementarities and Externalities in Water and Sanitation*, **NBER Working Paper**, No. 21521, 2015.

EIDE, E. R.; SHOWALTER, M. H. *The effect of grade retention on educational and labor market outcomes*. **Economics of Education Review**, v. 20, pp. 563–576, 2001.

ELHORST, J. *Applied spatial econometrics: Raising the bar*. **Spatial Economic Analysis**, v.5, n.1, p.9-28, 2010.

ELHORST, J. P. *Spatial econometrics: from cross-sectional data to spatial panels*. Heidelberg, New York, Dordrecht, London: Springer-Verlag. 2014.

ESREY, S.; POTASH, J., ROBERTS, L.; SHIFF, C. *Effects of Improved Water Supply and Sanitation on Ascariasis, Diarrhoea, Dracunculiasis, Hookworm Infection, Schistosomiasis, and Trachoma*. **Bulletin of the World Health Organization**, 69 (5), 1991

FAIR, G. M., GEYER, J.C., OKUN, D.A. *Water and wastewater engineering*. Nova York: Wiley, 1966.

FELLENBERG, G. *Introdução aos problemas da poluição ambiental*. Trad. Juergen Heinrich Maar, São Paulo, **Pedagógica e Universitária**, 1980.

FERRÃO, M. E. et. al. *Sistema Nacional de Avaliação da Educação Básica: objetivos, características e contribuições na investigação da escola eficaz*. **Revista Brasileira de Estudos de População**, v. 18, n.1/2, jan/dez, 2001.

FEWTRELL, L.; COLFORD, JR. *Water, Sanitation and Hygiene: Interventions and Diarrhea*, **HNP Discussion Paper**, 2004.

FEWTRELL, L., KAUFMANN, R. *Water Sanitation and Hygiene Interventions to Reduce Diarrhoea in Developed Countries: A Systematic Review and Meta-Analysis*. **Lancet Infectious Diseases**, 5, 2005.

FLORAX.; FOLMER, H.; REY,S. *Specification searches in spatial econometrics: the relevance of Hendry's methodology*. **Regional Science Urban Economics**, v.22, n.5, p. 557-579, 2003.

FONSECA, R. F.; VASCONCELOS, C. H. *Análise espacial das Doenças Relacionadas ao Saneamento Ambiental Inadequado no Brasil*, **Caderno de Saúde Coletiva**, Rio de Janeiro, 19 (4): 448-53, 2011.

FRANCO, A.M de P; MENEZES FILHO, N.A. Os determinantes do aprendizado com dados de um painel de escolas do SAEB. **Encontro ANPEC**, 2009.

GALIANI, S.; GERTLER, P. & SCHARGRODSKY, E. *Water for Life: the Impact of the Privatization of Water Services on Child Mortality*, **Journal of Political Economy**, v. 113, n.1, 2005.

GALLUP, J.; SACHS, J. *The Economic Burden of Malaria*, **American Journal of Tropical Medicine and Hygiene**, 64, 1, 85-96, 2001.

GALVÃO JR., A. C. et al. *Marcos regulatórios estaduais em saneamento básico no Brasil*. **Revista de Administração Pública**, Rio de Janeiro, v. 43, n. 1, p. 207-227, 2009.

GAMPER-RABINDRAN, S. et al. *The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach*. **Journal of Development Economics**, 92(2): 188-200, 2010.

GERUSO, M.; SPEARS, D. *Neighborhood Sanitation and Infant Mortality*, **National Bureau of Economic Research Working Paper Series**, n. 21184, 2015.

GOLGHER, A. B. *Introdução à Econometria Espacial*, Jundiaí, Paco Editorial: 2015.

GREENE, W. H. *Econometric Analysis*, 7 Edition, Prentice Hall, 2003.

GROSSMAN, M. *On the concept of health capital and the demand for health*. **Journal of Political Economy**, 80: 2, 223-255, 1972.

GURGEL, R. Q. et al. *Creche: ambiente expositor ou protetor nas infestações por parasitas intestinais em Aracaju, SE*. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba, v. 38, n. 3, p. 267-269, maio/jun. 2005.

HAINING, R. *The nature of geo referenced data*. In: FISCHER, M. M.; GETIS, A. (ed). **Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications**. Berlin: Springer, 2010.

HANUSHEK, E. *The Economics of Schooling: Production and Efficiency in Public Schools*, **Journal of Economic Literature**, v. 24, No. 3, p 1141-1177, 1986.

HANUSHEK, E. e KIMKO, D. *Schooling, Labor-Force Quality, and the Growth of Nations*, **The American Economic Review**, 90, 1184-1208, 2000.

HANUSHEK, E. A.; RIVKIN, S. G. *Teacher Quality* in: **Handbook of the Economics of Education**, v.2, Amsterdam: North Holland, pp. 1052-1078, 2006.

KELEJIAN, H. H.; PRUCHA, I. R. *Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model*. **International Economic Review**, Vol. 40, No. 2, pp. 509-533, 1999.

HAUSMAN, J. A. *Specification Tests in Econometrics*, **Econometrica**, Vol. 46, No. 6, p. 1251-1271, 1978.

HECKMAN, J. *The Lessons from technology of skill formation*, **NBER Working Paper**, 11142, 2005.

HECKMAN, J., LAYNE-FARRAR, A.; TODD, P. *The schooling quality-earnings relationship: using economic theory to interpret functional forms consistent with the evidence*. **National Bureau of Economic Research Working paper**, n. 5288, 1996.

HELLER L. *Saneamento e Saúde*. Brasília: OPAS; 1997.

HUTTLY, S.R.A. *The impact of inadequate sanitary conditions on health in developing countries*. **World Health Statistics Quarterly**, v.43, p.118-126, 1990.

JACOB, B. A.; LEFGREN, L. *The Effect of Grade Retention on High School Completion*. **CLOSUP Working Paper Series**, n. 12, 2009.

KALENKOSKI, C.; LACOMBE, D. *Minimum Wages and teen employment: a spatial panel approach*. **Regional Science**, 2012.

KAPOOR M, KELEJIAN H.H, PRUCHA I.R. *Panel Data Model with Spatially Correlated Error Components*. **Journal of Econometrics**, 140(1), 97-130, 2007.

KASSOUF, A. *A demanda de saúde infantil no Brasil por região e setor*. **Pesquisa e Planejamento Econômico**, n. 24, 235-260, 1994.

KASSOUF, A. L. *Saneamento e educação: bens substitutos ou complementares*, **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v.25, n.2, p. 359-372, ago 1995.

KASSOUF, A. L. *Trabalho infantil: escolaridade x emprego*, **Economia**, v. 2, n. 2, p. 549-586, jul./dez. 2001.

KELEJIAN, H. H.; PRUCHA, I. R. *A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances*. **J. Real State Finance Econ.**, Dordrecht, v.17, n.1, p.99-121, 1998.

KELEJIAN, H. H.; PRUCHA, I. R. *A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model*, **International Economic Review**, Vol. 40, No. 2, pp. 509-533, May, 1999.

KUNZ, J. M. O.; VIEIRA, A. S.; VARVAKIS, T.; GOMES, G. A.; ROSSETO, A. L.; BERNARDINI, O. J.; ALMEIDA, M. S. S.; ISHIDA, M. M. I. *Parasitas intestinais em crianças de escola municipal de Florianópolis, SC*, Educação ambiental e em saúde, **Biotemas**, v. 21, n. 4, p. 157-162, 2008.

LEE. *Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models*. **Econometrica**, v.72, n.6, p. 1899-1925, 2004.

LESAGE, J.; PACE, R. K. *Introduction to spatial econometrics*. Boca Raton: Taylor & Francis Group, 2009.

LUDWIG, K. M.; FREI, F.; ALVARES FILHO, F.; RIBEIRO-PAES, J. T. *Correlação entre condições de saneamento básico e parasitoses intestinais na população de Assis, Estado de São Paulo*, **Rev. Soc. Brasileira de Medicina Tropical**, 32(5):547-55, 1999.

LYNCH, S. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*, New York: Springer, 2006.

MACEDO, G. A. *Fatores associados ao rendimento escolar de alunos da 5ª série (2000): uma abordagem do valor adicionado e da heterogeneidade*. 124 f. Dissertação (Mestrado em Demografia), Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2004

MACEDO, H. S. *Prevalência de parasitoses e comensais intestinais em crianças de escolas da rede pública municipal de Paracatu (MG)*. **Rev. bras. anal. clin; RBAC** 37: 209-213, 2005.

MANSKI, C. *Identification of endogenous social effects: the reflection problem*. **The Review of Economic Studies**, v.3, n.60, p. 531-542, 1993.

MARA, D. D.; FEACHEM, R. G. A. *Water and excreta related diseases: unitary environmental classification*. **Journal of Environmental Engineering**, n.125, 1999.

MENDONÇA, M. J. C.; MOTTA, R. S. *Saúde e saneamento no Brasil*. **Planejamento e Políticas Públicas**, v.30, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília, 2007.

MENEZES-FILHO, N.A. *Os determinantes do desempenho escolar no Brasil*. Instituto Futuro Brasil, IBMEC São Paulo e Faculdade de Economia e Administração da Universidade de São Paulo. Sumário Executivo. 2007.

MENEZES-FILHO, N. A. *Educação e desigualdade*. In: MENEZES-FILHO, N. A.; LISBOA, M. (Eds.). **Microeconomia e sociedade**. Rio de Janeiro: Contracapa, 2001.

MENEZES-FILHO, N.; VASCONCELLOS, L.; WERLANG, S. *Avaliando o impacto da progressão continuada no Brasil*. São Paulo: **Fundação Itaú Social**, 2008.

MIGUEL, E.; KREMER, M. *Worms: identifying impacts on education and health in the presence of treatment externalities*, **Econometrica**, 72 (1), 159-217, jan 2004.

MILLO E PIRAS, *splm: Spatial Panel Data Models in R*, **Journal of Statistical Software**, v. 47, Issue 1, 2012.

MILLIMET, D. L.; RANGAPRASAD, V. *Strategic Competition Amongst Public Schools*, **Regional Science and Urban Economics**, 37, 199-219, 2007.

MORAES et al, L. R. S., CANCIO, J. A., CAIRNCROSS, S.; HUTTLY, S., *Impact of drainage and sewerage on diarrhoea in poor urban areas in Salvador, Brazil*, **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 97, 153- 158, 2003.

MUSHKIN, S. J.; *Investment in Human Beings*, **Journal of Political Economy**, v. 70, n. 5, Part 2, pp. 129-157, 1962.

MUTL, J.; PFAFFERMAYR, M. *The Hausman test in a Cliff and Ord panel model*, **Econometrics Journal**, volume 14, pp. 48-76, 2011.

NAIDOO A.G.V, VAN EEDEN, MUNCH Z 2014. *Spatial variation in school performance, a local analysis of socio-economic factors in Cape Town*. **South African Journal of Geomatics**, 3(1): 78-94, 2014.

NATAL, D.; MENEZES, R. M. T. de; MUCCI, J. L. N. *Fundamentos de Ecologia Humana*, cap3, p57-86 in: **Saneamento, Saúde e Ambiente: fundamentos para um desenvolvimento sustentável**, Arlindo Philippi Jr.(editor), Barueri, SP: Manole, 2005.

NERI *et. al.*, TRATA BRASIL: *Saneamento, Educação, Trabalho e Turismo*, 2008.

NEVES, D. P. *Parasitologia Humana*. 12^a ed. Rio de Janeiro: Atheneu, 2011.

NEVES, R.; PAZELLO, E. T. *O efeito de políticas de não repetência sobre o desempenho dos estudantes do Ensino Fundamental*. In: Anais 40 **Encontro Nacional de Economia**, 2012, Porto de Galinhas - RE.

OSLAND, L. *An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling*. **Journal of Real Estate Research**, vol. 32 (3), 2010, p. 289-320.

PEDROSO, R. S.; SIQUEIRA, R.V. *Pesquisa de cistos de protozoários, larvas e ovos de helmintos em chupetas*. **Jornal da Pediatria**, 73 (1): 21-25, 1997.

PHILIPPI JR. A. *Saneamento, Saúde e Ambiente: fundamentos para um Desenvolvimento Sustentável*, Barueri, SP: Manole, 2005.

PONTILI, R. M.; KASSOUF, A. L. *Fatores que afetam a frequência e o atraso escolar, nos meios urbano e rural, de São Paulo e Pernambuco*, **Revista de Economia e Sociologia Rural**, v. 45, n. 1, Brasília, jan/mar 2007.

PRADO, M. S.; BARRETO, L. M.; STRINA, A.; FARIA, J. A. S.; NOBRE, A. A. e JESUS, S. R. *Prevalência e intensidade da infecção por parasitas intestinais em crianças na idade escolar na Cidade de Salvador (Bahia, Brasil)*. **Revista da sociedade Brasileira de Medicina Tropical**, 34 (1): 99-101, Jan/Fev 2001.

REGO, R.F.A., MORAES L.R.S., DOURADO, C. *Diarrhoea and garbage disposal in Salvador, Brazil*, **Trans R Soc Trop Med Hyg**, 99(1):48-54, jan 2005.

RIVKIN, S. G.; HANUSHEK, E. A.; KAIN, J. F. *Teachers, schools, and academic achievement*. **Econometrica**, v. 73, n. 2, p. 417-458, 2005.

ROCHA & SOARES (2012), *Water Scarcity and Birth Outcomes in the Brazilian Semiarid*, **IZA Discussion Paper**, No. 6773, Transactions of the Royal Society of Tropical Medicine and Hygiene (2005) 99, 48—54

RODRIGUEZ-POSE, A.; TSELIOS, V. *Education and Income Inequality in the Regions of the European Union*, **Spatial Economics Research Centre (SERC)**, SERC Discussion paper 11, November 2008.

UMAR, H. M. Regional Inequality of Educational Attainment in Nigeria, **International, British Journal of Economics, Management & Trade**, 4(3): 420-430, 2014.

SAIANI, C. C. S.; FURQUIN, de A. P. *Privatização Como Estratégia Política: Evidências Para O Saneamento Básico Brasileiro*, Anais do XL **Encontro Nacional de Economia**, 2006.

SAIANI, C. C. S.; AZEVEDO, P. F. *Privatização como estratégia política: evidências para o saneamento básico brasileiro*. In: ENCONTRO NACIONAL DE ECONOMIA, 40, 2012. Anais... Porto de Galinhas, Pernambuco: ANPEC, 2012.

SCHADY, N. *Early childhood development in Latin America and the Caribbean*, **World Bank Policy Research**, Working Paper 3869, March 2006.

SEROA DA MOTTA, R. As opções de marco regulatório de Saneamento no Brasil. **Plenarium**, Brasília, n.3, p. 100-116, set. 2006.

SELOWSKY, M., TAYLOR, L. *The Economics of Malnourished Children: An Example of Disinvestment in Human Capital*, **Economic Development and Cultural Change**, v. 22, n. 1, pp. 17-30, Oct., 1973.

SNOW, J. *Sobre a maneira de transmissão do cólera*, 2. ed. São Paulo: Hucitec/ABRASCO, 249p, 1990.

SOUSA, T. R. V.; LEITE FILHO, P. A. M. *Análise por Dados em painel do status de saúde no Nordeste brasileiro*, **Revista Saúde Pública**, v.45(2), 796-804, 2008.

STOCK, J. H.; YOGO, M. *Testing for weak instruments in linear IV Regression. Identification and Inference for Econometric Models*, Cambridge, **UK. Cambridge University Press**, 2005.

PHILIPPI JR; MALHEIROS, T. F. *Saneamento e Saúde Pública: integrando homem e ambiente*, p.3-31 in: PHILIPPI JR. A (ed). *Saneamento, Saúde e Ambiente: fundamentos para um Desenvolvimento Sustentável*, Barueri, SP: Manole, 2005.

PIRAS, G. *Estimation of Random Effects Spatial Panel Data Models: Some Additional Evidence*. **Unpublished Manuscript**, 2011.

PLENZLER, N. L. *Student performance and educational resources: a spatial econometric examination*, Theses and Dissertations, **The University of Toledo Digital Repository**, 2004.

TOBIN, J. *Estimation of Relationships for Limited Dependent Variables*, **Econometrica**, 26 (1): 24-36, 1958.

TONETO, R. T.; SAIANI, C. S. *Restrições à Expansão dos Investimentos no Saneamento Básico Brasileiro*, **Revista Econômica do Nordeste**, Fortaleza, v. 37, nº 4, out-dez. 2006.

TOPEL, R. *Labor Markets and Economic Growth* in: **Handbook of Labor Economics**, Amsterdam: Elsevier Science B.V., 2943-29, 1999.

VASCONCELLOS, C. dos S. *Avaliação da aprendizagem: práticas de mudanças por uma práxis transformadora*, 5. ed., São Paulo: **Libertad**, 2003.

VICTORA, C. G., BARRETO M. L., DO CARMO LEAL, et al. *Health conditions and health-policy innovations in Brazil: the way forward*. **Lancet**, 2011.

WAGNER E. G.; LANOIX, JN. *Excreta disposal for rural areas and small communities*. **Monogr Ser World Health Organ**, 39:1-182, 1958.

WATSON, T. *Public Health Investments and the Infant Mortality Gap: Evidence from Federal Sanitation Interventions on U.S. Indian Reservations*, **Journal of Public Economics**, 90: 1537-1560, 2006.

WOOLDRIDGE, J. M. *Introductory Econometrics*. Pioneira Thompson Learning, 20^a ed., 2002.

APÊNDICE

1A. Principais mecanismos de transmissão das DRSAl

As DRSAl são doenças que podem estar associadas ao abastecimento de água deficiente, ao esgotamento sanitário inadequado, a contaminação por resíduos sólidos ou às condições precárias de moradia.

As principais doenças desse tipo que mais acometeram a população brasileira no ano de 2010 foram a diarreia e gastroenterite de origem infecciosa presumível, outras infecções intestinais bacterianas, a Dengue, as infecções intestinais virais, febre hemorrágica devido ao vírus do dengue e infecções por *Salmonella*.

Apesar da investigação da etiologia e descrição detalhada das características epidemiológicas de tais doenças estarem fora do escopo dessa tese, é importante ressaltar ao menos algumas das formas de transmissão das mesmas.

Em geral, a transmissão das DRSAls ocorre pela ingestão de água contaminada por fezes, urina ou vômitos de doente ou portador. As águas servidas, não tratadas ou insuficientemente tratadas, carregam inúmeros parasitas, vírus e bactérias, muitos deles resistentes à luz solar e diversas condições de umidade, que entram facilmente em contato com os seres humanos e animais. Além disso, os alimentos que são irrigados por água contaminada também podem atingir a população, caso esta não tenha higiene ou o conhecimento necessário para desinfetar o mesmo. A falta de higiene também pode fazer com que alimentos e utensílios sejam contaminados pelo manuseio ou por moscas e, principalmente, pode também levar ao surgimento de locais propícios para a criação de insetos vetores. O caso mais típico no Brasil e que, recentemente, tem se constituído epidemia em nível nacional é o alastramento da fêmea do mosquito *Aedes aegypti* que transmite a dengue entre outras doenças.

2A. Compatibilização DRSAl com Capítulo 1 – CID 10

Nos dois quadros abaixo é apresentado a compatibilização das DRSAl elaborada pela Funasa com a classificação de doenças de acordo com o capítulo 1 do CID-10. Os valores entre parênteses indicam o código que aparece no sistema do DATASUS. No quadro 2.1 segue a compatibilização dos dois grupos que mais se destacam entre os casos de internações: o grupo de doenças de transmissão feco-oral e o de doenças transmitidas por inseto vetor.

Grupo DRSAI	Doenças CID-10	Categoria	Descrição e Subcategorias
Doenças de transmissão feco-oral	Diarréias	A00	Cólera A000 A001 A009
		A02	Outras infecções por Salmonella A020 A021 A022 A028 A029
		A04	Outras infecções intestinais bacterianas A040 A041 A042 A043 A044 A045 A046 A047 A048 A049
		A06	Amebíase A060 A061 A062 A063 A064 A065 A066 A067 A068 A069
		A07	Outras doenças intestinais por protozoários A070 A071 A072 A073 A078 A079
		A08	Infecções intestinais virais, outras e as não especificadas A080 A081 A082 A083 A084 A085
		A09	Diarréia e gastroenterite de origem infecciosa presumível A009 (=A09)
	Febres Entéricas	A01	Febres tifóide e paratífóide A010 A011 A012 A013 A014
	Hepatite A	B15	Hepatite aguda A B150 B159
	Dengue	A90	Dengue [dengue clássico] A090 (=A90)
Doenças transmitidas por inseto vetor	Febres Amarela	A91	Febre hemorrágica devida ao vírus do dengue A091
		A95	Febre amarela A950 A951 A959
	Leishmaniose	B55	Leishmaniose B550 B551 B552 B559
	Filariose Linfática	B74	Filariose B740 B741 B742 B743 B744 B748 B749
	Malária	B50	Malária por Plasmodium falciparum B500 B508 B509
		B51	Malária por Plasmodium vivax B510 B518 B519
		B52	Malária por Plasmodium malariae B520 B528 B529
		B53	Outras formas de malária confirmadas por exames parasitológicos B530 B531 B538
		B54	Malária não especificada B054 (=B54)
	Doença de Chagas	B57	Doença de Chagas B570 B571 B572 B573 B574 B575

Quadro 1A: Compatibilização dos grupos das DRSAI (I)

A seguir a compatibilização utilizada é apresentada para o restante dos grupos: o grupo de doenças transmitidas através do contato com a água, doenças relacionadas com a higiene e o grupo dos geo-helminthos e teníases.

Grupo DRSI	Doenças CID-10	Categoria	Descrição e Subcategorias
Doenças transmitidas através do contato com a água	Esquistossomose	B65	Esquistossomose [bilharziose] [Schistosomíase] B650 B651 B652 B653 B658 B659
	Leptospirose	A27	Leptospirose A270 A278 A279
Doenças relacionadas com a higiene	Doenças dos Olhos	A71	Tracoma A710 A711 A719
		H10	Conjuntivite H100 H101 H102 H103 H104 H105 H108 H109
	Doenças da Pele	B35	Dermatofitose B350 B351 B352 B353 B354 B355 B356 B358 B359
		B36	Outras micoses superficiais B360 B361 B362 B363 B368 B369
Geo-helminths e Teníases	Helminthiases	B68	Infestação por Taenia B680 B681 B689
		B69	Cisticercose B690 B691 B698 B699
		B71	Outras infestações por cestóides B710 B711 B718 B719
		B76	Ancilostomíase B760 B761 B768 B769
		B77	Ascaridíase B770 B778 B779
		B78	Estrongiloidíase B780 B781 B787 B789
		B79	Tricuríase B079
		B80	Oxiuríase B080
		B81	Outras helmintíases intestinais, não classificadas em outra parte B810 B811 B812 B813 B814 B818
		B82	Parasitose intestinal não especificada B820 B829
		B83	Outras helmintíases B830 B831 B832 B833 B834 B838 B839
	Teníase	B67	Equinococose B670 B671 B672 B673 B674 B675 B676 B677 B678 B679

Quadro 2A: Compatibilização dos grupos das DRSI (II)

3A. Faixas de CEPs dos distritos brasileiros

Um resumo das faixas de CEPs de cada distrito segundo as quais a variável de saúde foi construída, de acordo com o descrito no capítulo 3, encontra-se no site do Nereus: <http://www.usp.br/nereus/>

4A. Descrição das fórmulas das variáveis utilizadas nas regressões

- **Inc_todas_idades** = razão do número de casos de pessoas de todas as idades internadas por DRSI pela população de todas as idades; multiplicado por 1000.
- **Inc_0a14** = razão do número de casos de pessoas de 0 a 14 anos de idade internadas por DRSI pela população de 0 a 14 anos; multiplicado por 1000.

- **Dcist** = razão do número total de domicílios particulares permanentes com abastecimento de água da chuva armazenada em cisterna ou caixa de cimento pelo número total de domicílios particulares permanentes; multiplicado por 100.
- **Sbceuab** = razão do número de domicílios particulares permanentes que não tinham banheiro ou sanitário e que suas condições de entorno apresentavam esgoto a céu aberto (e1v154) pelo número total de domicílios particulares permanentes; multiplicado por 100.
- **Dfrud** = razão do número de domicílios particulares permanentes com banheiro de uso exclusivo dos moradores ou sanitário ligado a uma fossa rudimentar, ou seja, fossa negra, poço ou buraco pelo número total de domicílios particulares permanentes; multiplicado por 100.
- **Esg_nao_ceuab** = razão do somatório do número de domicílios particulares permanentes que tinham banheiro ou sanitário e esgotamento sanitário da rede geral de esgoto ou rede pluvial e onde não existia em sua face/quadra esgoto a céu aberto (e1v151) e do número de domicílios particulares permanentes que tinham banheiro ou sanitário e esgotamento sanitário de fossa séptica e onde não existia em sua face/quadra esgoto a céu aberto (e1v153) pelo número total de domicílios particulares permanentes; multiplicado por 100.
- **Ag_nao_ceuab** = razão do número de domicílios particulares permanentes com rede geral de distribuição de água e onde não existia em sua face/quadra esgoto a céu aberto pelo número total de domicílios particulares permanentes; multiplicado por 100.
- **Lixo_mor_inad** = razão do número de domicílios particulares permanentes com moradia inadequada e onde em sua face/quadra existia lixo acumulado nos logradouros pelo número total de domicílios particulares permanentes; multiplicado por 100.
- **D_nao_bueiro** = razão do somatório do número dos domicílios particulares permanentes cedidos e onde em sua face/quadra não existia bueiro/boca-de-lobo (e1v037), do número dos domicílios particulares permanentes alugados e onde em sua face/quadra não existia bueiro/boca-de-lobo (e1v035) e do número dos domicílios particulares permanentes próprios e onde em sua face/quadra não existia bueiro/boca-de-lobo (e1v033) pelo número total de domicílios particulares permanentes; multiplicado por 100.

- **Rendapc** = razão do total do rendimento nominal mensal dos domicílios particulares permanentes pela população residente.
- **Gini** = índice obtido da renda domiciliar per capita que representa a completa igualdade de renda quando assume valor igual a zero e a completa desigualdade quando assume valor igual a 1.
- **Mu_resp_alfab** = razão do número de pessoas do sexo feminino, responsáveis pelo domicílio e alfabetizadas pelo número de mulheres responsáveis pelo domicílio particular; multiplicado por 100.
- **Popurb** = razão do número de pessoas residentes em setores censitários urbanos pelo total da população residente; multiplicado por 100.
- **Dom_kmq** = razão do número total de domicílios particulares permanentes pela área em quilômetros quadrados.
- **Est_saude** = número de estabelecimentos de saúde.
- **Dummy_cap** = variável binária que indica se o distrito é capital
- **Dummy_reg_met** = variável binária que indica se o distrito pertence a uma região metropolitana oficial, estabelecida pelo IBGE.

5A. Diagramas de Dispersão de Moran (variáveis educação)

Os gráficos abaixo representam os diagramas de dispersão de Moran das estatísticas I de Moran Global. Tais estatísticas são obtidas a partir de uma regressão da defasagem espacial da variável dependente padronizada contra essa própria variável padronizada e uma constante: $Wy_{educ} = a + by_{educ} + e$. O mesmo cálculo foi realizado para os resíduos das regressões não espaciais, ou seja, modelos de dados em painel com efeitos fixos estimados em dois estágios com variável instrumental. Logo, a equação nesse caso será $WRes_{nao-esp} = a + bRes_{nao-e} + e$. A significância do parâmetro b é encontrada via técnica de *bootstrap*: a partir da atribuição de um valor aleatório a cada região, compara-se o I de Moran da amostra artificial com o observado. Caso sejam diferentes, o pseudo p-valor será rejeitado apresentando evidências de não aleatoriedade espacial. Nos diagramas de dispersão de Moran abaixo os dois primeiros referem-se ao “lag” da variável dependente e os dois últimos referem-se aos resíduos.

Gráfico A1: Dispersão de Moran - Variável Dependente: taxa de distorção idade-série (à esquerda) frequência escolar (no centro) e abandono escolar (à direita) para o ano de 2000.

Gráfico A2: Dispersão de Moran - Variável Dependente: taxa de distorção idade-série (à esquerda) frequência escolar (no centro) e abandono escolar (à direita) para o ano de 2010.

Gráfico A3: Dispersão de Moran - Resíduos: taxa de distorção idade-série (à esquerda), frequência escolar (no centro) e abandono escolar (à direita) para o ano de 2000.

Gráfico A.7: Diagrama de dispersão de Moran da Taxa de Incidência por DRSAI para população de 0 a 14 anos com matriz inversa da distância.

7A. Mapa de clusters segmentado pelas regiões brasileiras (“zoom regional”)

Figura A.1: Clusters região Norte

Figura A.2: Clusters região Nordeste

Figura A.3: Clusters região Sudeste

Figura A.4: Clusters região Sul

Figura A.5: Clusters região centro-oeste

8A. Estimação dos modelos espaciais

Tabela 8A.1: Modelos MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAI da população de todas as idades (matriz contígua)

	Y = Incidência DRSAI (todas as idades) - Wcont			
	MQO	SLX	SEM	SDEM
dcist	0,0823***	0,0588**	0,0819***	0,0602**
sbceuab	0,7554***	0,2522*	0,5294***	0,2947**
dfrud	0,023***	0,0145**	0,0206***	0,0141****
rendapc	-0,0012**	0,0002	-0,0012***	0,0001
gini	12,3859***	4,3251**	11,5606***	5,1074**
mu_resp_alfab	-0,0307***	0,009	-0,0274***	0,0064
tpopurb	0,0536***	0,0594	0,0586***	0,059***
dom_kmq	-0,0003	-0,0002	-0,0004**	-0,0002
est_saude	0,0003	-0,0018	0,0009	-0,0017
dummy_capital	-1,6921**	-1,5652	-1,9327***	-1,6948
dummy_rm	-2,7477***	-0,9152	-2,6551***	-1,0739
Wdcist		0,0244		0,0236
Wsbceuab		1,758***		1,4966***
Wdfrud		0,0108*		0,0107*
Wrendapc		-0,001		-0,0011
Wgini		8,4593***		7,4435**
Wmu_resp_alfab		-0,0309*		-0,0283*
Wtpopurb		-0,0298***		-0,028***
Wdom_kmq		0,0002		0,0002
West_saude		-0,0041		-0,0031
Wdummy_capital		0,3107		0,4476
Wdummy_rm		-1,7033		-1,4814
Constante	-5,3463	-5,3132**	-5,3217***	-5,2404***
Observações	10225	10225	10225	10225
R2	0,04792	0,0603	-	-
R2 Ajustado	0,0469	0,05831	-	-
Estatística F	46,74	29,77	-	-
Lambda	-	-	0,2405***	0,2220***
Teste de Wald	-	-	269,7***	225,05***
Teste LR	-	-	237,07***	203,09**
Criterio de Inf. Akaike	-	-	73788	73710

Tabela 8A.2: Modelos SAR, SDM , SAC e GNS para a taxa de incidência de internações por DRSAI da população de todas as idades (matriz contígua)

	Y = Incidência DRSAI (todas as idades) - Wcont			
	SAR	SDM	SAC	GNS
dcist	0,0731***	0,0577**	0,082***	0,0522*
sbceuab	0,6135***	0,2074	0,3206**	0,0622
dfrud	0,0189***	0,0133***	0,0188***	0,0123**
rendapc	-0,001**	0,0002	-0,0012**	0,0002
gini	10,3359***	4,3069**	10,4776***	2,8569
mu_resp_alfab	-0,0215***	0,0079	-0,0261**	0,0084
tpopurb	0,0519***	0,0607***	0,0608***	0,0636***
dom_kmq	-0,0003**	-0,0003	-0,0005**	-0,0003
est_saude	0,0004	-0,0016	0,0011	-0,0013
dummy_capital	-1,5806***	-1,7087	-2,4718***	-1,675
dummy_rm	-2,2672***	-0,9583	-2,8239***	-0,8421
Wdcist		0,0096		-0,0099
Wsbceuab		1,3223***		0,9346***
Wdfrud		0,0058		-0,0014
Wrendapc		-0,0009		-0,0006
Wgini		5,695**		3,3535
Wmu_resp_alfab		-0,0225		-0,0152
Wtpopurb		-0,0391***		-0,0533***
Wdom_kmq		0,0003		0,0003
West_saude		-0,0028		-0,0019
Wdummy_capital		0,7381		1,0978
Wdummy_rm		-1,0273		-0,3099
Constante	-5,4833***	-4,2876***	-3,3308**	-2,7327**
Rho	0,2317***	0,2251***	-0,42864***	0,5394***
Lambda	-	-	0,5602***	-0,4294***
Teste de Wald	255,44***	211,92***	-	-
Teste LR	231,55***	233,09***	282,28***	262,32***
Critério de Inf. Akaike	73794	73701	73745	73653

Tabela 8A.3: Modelos MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAI da população de todas as idades (matriz de distância)

	Y = Incidência DRSAI (todas as idades) - Wdist_inv			
	MQO	SLX	SEM	SDEM
dcist	0,0823***	0,0534**	0,0654***	0,0458*
sbceuab	0,7554***	0,0221	0,1655	0,072
dfrud	0,023***	0,0116**	0,0139***	0,0119***
rendapc	-0,0012***	-0,0001	-0,0004	-0,0001
gini	12,3859***	3,156	5,7467***	3,455*
mu_resp_alfab	-0,0307***	-0,0006	-0,0136	0,0031
tpopurb	0,0536***	0,0621***	0,0616***	0,0619***
dom_kmq	-0,0003*	-0,0003	-0,0003	-0,0003
est_saude	0,0003	-0,001	-0,0004	-0,0012
dummy_capital	-1,6921***	-1,777**	-2,1283***	-1,5502*
dummy_rm	-2,7477***	-0,1846	-1,6383***	-0,7438
Wdcist		0,0789		0,1251
Wsbceuab		3,984***		2,9283***
Wdfrud		0,0147**		0,0128
Wrendapc		-0,0005		0,0006
Wgini		11,77***		10,7602
Wmu_resp_alfab		0,0248		-0,0143
Wtpopurb		-0,0351***		-0,0457**
Wdom_kmq		0,0005		0,001
West_saude		-0,026***		-0,0268
Wdummy_capital		0,7962		-1,5846
Wdummy_rm		-3,567***		-1,8517
Constante	-5,3464***	-10,2***	-2,8066*	-6,7158
Observações	10225	10225	-	-
R2	0,04792	0,06831	-	-
R2 Ajustado	0,0469	0,0663	-	-
Estatística F	46,74***	34,0***	-	-
Lambda	-	-	0,73091***	0,65446***
Teste de Wald	-	-	927,02***	528,07***
Teste LR	-	-	557,86***	389,76***
Critério de Inf. Akaike	-	-	73468	73436

Tabela 8A.4: Modelos SAR, SDM , SAC e GNS para a taxa de incidência de internações por DRSAl da população de todas as idades (matriz de distância)

	Y = Incidência DRSAl (todas as idades) - Wdist_inv			
	SAR	SDM	SAC	GNS
dcist	0,0581***	0,0553**	0,0636***	0,0546**
sbceuab	0,3688***	0,1318	0,1209	0,0419
dfrud	0,0142***	0,0126***	0,0127***	0,0116***
rendapc	-0,0005	0,0002	-0,0003	-0,0001
gini	5,9052***	3,4506*	4,8283**	3,4798*
mu_resp_alfab	-0,0078	0,0069	-0,0086	-0,0007
tpopurb	0,0495***	0,0592***	0,0618***	0,0619***
dom_kmq	-0,0003*	-0,0003	-0,0003	-0,0003
est_saude	-0,0001	-0,0014	-0,0006	-0,0007
dummy_capital	-1,6458***	-1,7994	-2,1598***	-1,9081**
dummy_rm	-1,5223***	-1,0805	-1,415**	-0,8428
Wdcist		-0,0102		-0,0065
Wsbceuab		0,6916***		1,3219***
Wdfrud		-0,0002		-0,0041
Wrendapc		-0,0004		0,0005
Wgini		1,21		1,3382
Wmu_resp_alfab		-0,0129		0,0053
Wtpopurb		-0,0389***		-0,0564***
Wdom_kmq		0,0003		0,0005
West_saude		-0,0022		-0,0129
Wdummy_capital		0,7037		1,2181
Wdummy_rm		-0,0298		-0,2494
Constante	-5,0747***	-2,7808*	-1,0562	-3,0864
Rho	0,68281***	0,66717***	-0,39705***	0,7119*
Lambda	-	-	0,84821***	-0,16071
Teste de Wald	795,38***	618,28***	-	
Teste LR	535,45***	470,13***	567,92***	393,49***
Critério de Inf. Akaike	73490	73443	73460	73435

Tabela 8A.5: Modelos MQO, SLX , SEM e SDEM para a taxa de incidência de internações por DRSAl da população de 0 a 14 anos (matriz contígua)

	Y = Incidência DRSAl (de 0 a 14 anos) - Wcont			
	MQO	SLX	SEM	SDEM
dcist	0,1372***	0,10465*	0,141***	0,1068*
sbceuab	0,879***	0,16671	0,696**	0,2114
dfrud	0,0403***	0,03866***	0,04***	0,0385***
rendapc	-0,0019**	-0,00004	-0,002*	-0,0001
gini	18,5096***	5,90067	17,889***	6,7056
mu_resp_alfab	-0,0264	0,05783**	-0,019	0,0549**
tpopurb	0,0996***	0,10712***	0,105***	0,1068***
dom_kmq	-0,0005	-0,00146***	-0,001*	-0,0014***
est_saude	0,0025	0,00071	0,004	0,0009
dummy_capital	-2,5055*	-7,18673***	-2,886**	-7,6891***
dummy_rm	-4,8263***	-3,21287	-4,678***	-3,5614
Wdcist		0,0103		0,0098
Wsbceuab		2,08113***		1,8716***
Wdfrud		-0,00452		-0,0055
Wrendapc		-0,0018		-0,0019
Wgini		13,07835**		12,2746*
Wmu_resp_alfab		-0,08782**		-0,0837**
Wtpopurb		-0,0487***		-0,0478***
Wdom_kmq		0,00231***		0,0022***
West_saude		-0,01541**		-0,0137*
Wdummy_capital		6,12333*		6,8601**
Wdummy_rm		-1,7123***		-1,2372
Constante	9,9562***	-7,40049**	-10,492***	-7,4677*
Observações	10225	10225	-	-
R2	0,02799	0,0361	-	-
R2 Ajustado	0,02695	0,03402	-	-
Estatística F	26,74***	17,37***	-	-
Lambda	-	-	0,1647***	0,1536***
Teste de Wald	-	-	116,75***	100,51***
Teste LR	-	-	95,174***	83,462***
Critério de Inf. Akaike	-	-	89258	89206

Tabela 8A.6: Modelos SAR, SDM , SAC e GNS para a taxa de incidência de internações por DRSAI da população de 0 a 14 anos (matriz contígua)

	Y = Incidência DRSAI (de 0 a 14 anos) - Wcont			
	SAR	SDM	SAC	GNS
dcist	0,1322***	0,1055*	0,1444***	0,1046*
sbceuab	0,7904***	0,1315	0,6828**	0,1335
dfrud	0,0374***	0,0405***	0,0419***	0,0382***
rendapc	-0,0017*	-0,0001	-0,002*	-0,00001
gini	17,456***	6,0992	18,541***	5,8223
mu_resp_alfab	-0,0138	0,0647**	-0,0152	0,0579**
tpopurb	0,0977***	0,1082***	0,106***	0,1088***
dom_kmq	-0,0005	-0,0015***	-0,0007	-0,0015***
est_saude	5,2328	0,0012	0,0041	0,0012
dummy_capital	-2,4501*	-7,6224***	-2,9773**	-7,6599***
dummy_rm	-4,3338***	-3,366	-4,8306***	-3,3844
Wdcist		-0,0051		-0,0044
Wsbceuab		1,7543***		1,7406***
Wdfrud		-0,0127		-0,0107
Wrendapc		-0,0017		-0,0017
Wgini		10,622*		10,412
Wmu_resp_alfab		-0,0847**		-0,0805**
Wtpopurb		-0,0614***		-0,0615***
Wdom_kmq		0,0023***		0,0023***
West_saude		-0,0134*		-0,0135*
Wdummy_capital		6,8834**		6,9509**
Wdummy_rm		-0,8319		-0,7198
Constante	-11,009***	-6,7549*	-10,962***	-6,4139*
Rho	0,1441***	0,1447***	-0,0469	0,1545
Lambda	-	-	0,19579**	0,0011402
Teste de Wald	89,311***	88,703***	-	-
Teste LR	72,957***	72,347***	80,349***	86,225***
Critério de Inf. Akaike	90310	90253	90304	89205

Tabela 8A.7: Modelos MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAI da população de todas as idades (matriz de distância)

	Y = Incidência DRSAI (de 0 a 14 anos) - Wdist_inv			
	MQO	SLX	SEM	SDEM
dcist	0,1372***	0,0861*	0,1209***	0,0767
sbceuab	0,879***	-0,0652	0,3128	-0,0206
dfrud	0,0403***	0,0290***	0,0347***	0,031***
rendapc	-0,0019**	0,0000	-0,0011	-0,0001
gini	18,5096***	5,1190	12,2907***	5,3657
mu_resp_alfab	-0,0264	0,0375	-0,0077	0,0453*
tpopurb	0,0996***	0,1122***	0,1101***	0,1108***
dom_kmq	-0,0005	-0,0009*	-0,0009**	-0,0012**
est_saude	0,0025	0,0014	0,0036	0,0016
dummy_capital	-2,5055*	-2,4230	-3,2823**	-3,0833*
dummy_rm	-4,8263***	-0,1961	-3,5875***	-1,4612
Wdcist		0,1081		0,1384
Wsbceuab		4,5240***		3,8996***
Wdfrud		0,0029		-0,0045
Wrendapc		-0,0023		-0,0007
Wgini		17,2800*		16,265
Wmu_resp_alfab		0,0027		-0,0507
Wtpopurb		-0,0630***		-0,0704**
Wdom_kmq		0,0030***		0,0039***
West_saude		-0,0808***		-0,0798**
Wdummy_capital		0,8159		-0,4661
Wdummy_rm		-7,6240***		-5,5428*
Constante	9,9562***		-8,256***	-9,7031
Observações	10225	10225	-	-
R2	0,02799	0,0393	-	-
R2 Ajustado	0,02695	0,03723	-	-
Estatística F	26,74***	18,97***	-	-
Lambda	-	-	0,5349***	0,4655***
Teste de Wald	-	-	237,46***	149,98***
Teste LR	-	-	214,95***	151,39***
Critério de Inf. Akaike	-	-	89138	89104

Tabela 8A.8: Modelos SAR, SDM, SAC e GNS para a taxa de incidência de internações por DRSAl da população de 0 a 14 anos (matriz de distância)

	Y = Incidência DRSAl (de 0 a 14 anos) - Wdist_inv			
	SAR	SDM	SAC	GNS
dcist	0,1113***	0,10025*	0,1081**	0,0659
sbceuab	0,5716**	0,07365	0,0715	-0,0096
dfrud	0,0311***	0,03713***	0,0278***	0,0311***
rendapc	-0,0012	-0,00004	-0,0004	-0,0005
gini	11,7738***	4,79376	7,2349*	5,4112
mu_resp_alfab	-0,0048	0,05524**	0,0235	0,0484*
tpopurb	0,0935***	0,10638***	0,1107***	0,1081***
dom_kmq	-0,0007*	-0,00161***	-0,0008*	-0,0013***
est_saude	0,0034	0,00207	0,0023	0,0013
dummy_capital	-2,6399**	-7,78022***	-3,0413*	-3,2921*
dummy_rm	-3,2952***	-3,62583	-2,1702	-2,283
Wdcist		-0,02759		0,515*
Wsbceuab		1,21357**		3,6424
Wdfrud		-0,01671		-0,0329
Wrendapc		-0,00115		0,0008
Wgini		5,22349		18,0514
Wmu_resp_alfab		-0,07131**		-0,36***
Wtpopurb		-0,06322***		0,0167
Wdom_kmq		0,00215***		0,0079***
West_saude		-0,01012		-0,0851*
Wdummy_capital		6,81338**		-6,4951
Wdummy_rm		0,71823		-1,3418
Constante		-4,62007		
Rho	0,49805***	0,48548***	-1,2196***	-1,2513
Lambda	-	-	0,8923***	0,88172**
Teste de Wald	212,69***	178,07***	-	-
Teste LR	202,5***	183,39***	310,49***	232,98***
Critério de Inf. Akaike	89151	89106	89045	89025

Tabela 8A.9: Modelos MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAl população de todas as idades (matriz contígua)

	Y = Incidência DRSAl (todas as idades) - Wcont			
	MQO	SLX	SEM	SDEM
esg_nao_ceuab	-0,0219***	0,0037	-0,0139***	0,0031
ag_nao_ceuab	-0,0186***	-0,0229***	-0,0207***	-0,024***
lixo_mor_inad	0,2599	0,1724	0,1698	0,1618
d_nbueiro	0,0268***	0,0054	0,0196***	0,0053
rendapc	-0,0005	0,0003	-0,0007	0,0003
gini	10,762***	3,5665	10,57***	4,2464**
mu_resp_alfab	-0,0367***	0,0088	-0,0323***	0,0057
tpopurb	0,0576***	0,0678***	0,0646***	0,0689***
dom_kmq	-0,0002	-0,0002	-0,0004*	-0,0002
est_saude	0,0009	-0,0021	0,0013	-0,0019
dummy_capital	-1,9186***	-1,7493	-2,1236***	-1,8137
dummy_rm	-2,9914***	-1,0515	-2,9208***	-1,1808
Wesg_nao_ceuab		-0,0393***		-0,0361***
Wag_nao_ceuab		0,0079		0,0077
Wlixo_mor_inad		1,0637**		0,6574
Wd_nbueiro		0,0448***		0,0435***
Wrendapc		0,0003		0,0001
Wgini		4,5374		3,8306
Wmu_resp_alfab		-0,0493***		-0,0445***
Wtpopurb		-0,0437***		-0,0427***
Wdom_kmq		0,0005		0,0005
West_saude		-0,0019		-0,0013
Wdummy_capital		0,1746		0,2758
Wdummy_rm		-1,7177		-1,5283
Constante	-2,7751**	-0,1136	-3,2137***	-0,2452
Observações	10225	10225	-	-
R2	0,04657	0,05993	-	-
R2 Ajustado	0,04545	0,05771	-	-
Estatística F	41,56***	27,09***	-	-
Lambda	-	-	0,24078***	0,22404***
Teste de Wald	-	-	270,58***	207,08***
Teste LR	-	-	235,03***	229,87***
Critério de Inf. Akaike	-	-	73807	73715

Tabela 8A.10: Modelos SAR, SDM, SAC e GNS para a taxa de incidência de internações por DRSAI da população de todas as idades (matriz contígua)

	Y = Incidência DRSAI (todas as idades) - Wcont			
	SAR	SDM	SAC	GNS
esg_nao_ceuab	-0,016***	0,006	-0,0052	0,0103
ag_nao_ceuab	-0,0187***	-0,0244***	-0,0216***	-0,0253***
lixo_mor_inad	0,1952	0,1321	0,1045	0,0254
d_nbueiro	0,0197***	0,0033	0,0139**	0,0023
rendapc	-0,0004	0,0003	-0,0009*	0,0003
gini	9,0839***	3,7689*	9,8214***	2,8611
mu_resp_alfab	-0,0275***	0,0084	-0,0289***	0,0108
tpopurb	0,0582***	0,0707***	0,0666***	0,0728***
dom_kmq	-0,0003	-0,0002	-0,0005**	-0,0003
est_saude	0,0007	-0,002	0,0014	-0,0019
dummy_capital	-1,7934***	-1,8006	-2,6163***	-1,7356
dummy_rm	-2,505***	-1,0513	-3,0564***	-0,9124
Wesg_nao_ceuab		-0,0323***		-0,0257***
Wag_nao_ceuab		0,0127		0,0202*
Wlixo_mor_inad		0,7282		0,7237
Wd_nbueiro		0,0353***		0,0208**
Wrendapc		0,0001		0,0003
Wgini		2,4849		0,9443
Wmu_resp_alfab		-0,0382**		-0,0285*
Wtpopurb		-0,0539***		-0,0673***
Wdom_kmq		0,0005		0,0005
West_saude		-0,001		-0,0003
Wdummy_capital		0,6056		1,0417
Wdummy_rm		-1,0378		-0,2812
Constante	-3,2912***	-0,1185	-1,5396	-0,0226
Rho	0,23188***	0,22635***	-0,43783***	0,54101***
Lambda	-	-	0,56777***	-0,42975***
Teste de Wald	255,73***	236,09***	-	-
Teste LR	230,43***	213,96***	281,91***	263,34***
Criterio de Inf. Akaike	73812	73708	73762	73660

Tabela 7A.11: Modelos MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAI da população de todas as idades (matriz de distância)

	Y = Incidência DRSAI (todas as idades) - Wdist_inv			
	MQO	SLX	SEM	SDEM
esg_nao_ceuab	-0,0356***	0,0136**	0,0065	0,0122**
ag_nao_ceuab	-0,0486***	-0,0224***	-0,0204***	-0,0232***
lixo_mor_inad	0,0517	0,1046	0,0746	0,0988
d_nbueiro	0,0154	0,0113	0,0138**	0,0092
rendapc	-0,0009	-0,0001	-0,0002	-0,0001
gini	16,2374***	2,4774	4,9626**	2,7138
mu_resp_alfab	-0,038**	-0,0026	-0,0151	0,002
tpopurb	0,1385***	0,0618***	0,0612***	0,0643***
dom_kmq	-0,0004	-0,0003	-0,0003	-0,0003
est_saude	0,0027	-0,0011	-0,0005	-0,0014
dummy_capital	-3,1044**	-1,7668**	-2,2618***	-1,6692*
dummy_rm	-5,4576***	-0,1275	-1,7618***	-0,8188
Wesg_nao_ceuab		-0,0619***		-0,0516**
Wag_nao_ceuab		0,0083		0,0407
Wlixo_mor_inad		4,3269***		4,590*
Wd_nbueiro		0,0449***		0,0518*
Wrendapc		0,002		0,0022
Wgini		5,5025		5,3698
Wmu_resp_alfab		-0,0275		-0,055
Wtpopurb		-0,0378**		-0,09**
Wdom_kmq		0,0009**		0,0015*
West_saude		-0,0184*		-0,0211
Wdummy_capital		-0,5203		-1,9097
Wdummy_rm		-3,8954***		-1,544
Constante	-5,4944**	-1,1899	-1,3301	1,2436
Observações	10225	10225	-	-
R2	0,04657	0,06437	-	-
R2 Ajustado	0,04545	0,06217	-	-
Estatística F	41,56***	29,24***	-	-
Lambda	-	-	0,74092***	0,66662***
Teste de Wald	-	-	1003,9***	430,14***
Teste LR	-	-	565,49***	575,08***
Criterio de Inf. Akaike	-	-	73477	73443

Tabela 8A.12: Modelos SAR, SDM, SAC e GNS para a taxa de incidência de internações por DRSAI da população de todas as idades (matriz de distância)

	Y = Incidência DRSAI (todas as idades) - Wdist_inv			
	SAR	SDM	SAC	GNS
esg_nao_ceuab	-0,0076	0,0126**	0,0097	0,0122**
ag_nao_ceuab	-0,0155**	-0,0226***	-0,0209***	-0,0231***
lixo_mor_inad	0,12749	0,0591	0,0642	0,1102
d_nbueiro	0,01218**	0,0105	0,0128*	0,0093
rendapc	-0,00009	-0,0001	-0,0002	-0,0001
gini	5,1173***	2,7425	4,0184**	2,7108
mu_resp_alfab	-0,01295	-0,0012	-0,0098	0,0026
tpopurb	0,05424***	0,0629***	0,0609***	0,0639***
dom_kmq	-0,0003*	-0,0004*	-0,0003	-0,0003
est_saude	-0,00002	-0,001	-0,0008	-0,0014
dummy_capital	-1,801***	-1,9664**	-2,2986***	-1,5585*
dummy_rm	-1,7134***	-0,8981	-1,5165**	-0,8375
Wesg_nao_ceuab		-0,0297***		-0,053*
Wag_nao_ceuab		0,0306*		0,0464
Wlixo_mor_inad		2,5193*		5,0321*
Wd_nbueiro		0,0094		0,068**
Wrendapc		0,0012		0,0025
Wgini		0,3954		5,7838
Wmu_resp_alfab		-0,0146		-0,0811*
Wtpopurb		-0,0695***		-0,1037**
Wdom_kmq		0,0007*		0,0018*
West_saude		-0,0115		-0,0222
Wdummy_capital		0,8084		-3,211
Wdummy_rm		-0,4075		-1,2064
Constante	-3,4187***	-0,1212		
Rho	0,6861***	0,6587***	-0,4228***	-0,2504**
Lambda	-	-	0,8578***	0,7654**
Teste de Wald	802,38***	556,6***	-	-
Teste LR	535,72***	427,87***	578,24***	432,82***
Critério de Inf. Akaike	73506	73445	73466	73442

Tabela 7A.13: Modelos MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAI da população de 0 a 14 anos (matriz contígua)

	Y = Incidência DRSAI (de 0 a 14 anos) - Wcont			
	MQO	SLX	SEM	SDEM
esg_nao_ceuab	-0,0356***	-0,0031	-0,0334***	-0,0081
ag_nao_ceuab	-0,0486***	-0,066***	-0,0503***	-0,058***
lixo_mor_inad	0,0517	-0,0409	-0,017	-0,0775
d_nbueiro	0,0154	-0,0208	0,0096	-0,0166
rendapc	-0,0009	0,0001	-0,0013	0,0002
gini	16,2374***	4,9289	15,2052***	2,304
mu_resp_alfab	-0,038**	0,064**	-0,017	0,0872***
tpopurb	0,1385***	0,1636***	0,1449***	0,1537***
dom_kmq	-0,0004	-0,0014***	-0,0006	-0,0014***
est_saude	0,0027	-0,0002	0,0044	-0,0002
dummy_capital	-3,1044**	-7,6224***	-3,2738**	-7,6203***
dummy_rm	-5,4576***	-3,4563	-5,4435***	-3,6577
Wesg_nao_ceuab		-0,0472**		-0,0381*
Wag_nao_ceuab		0,044*		0,0231
Wlixo_mor_inad		1,1593		0,8086
Wd_nbueiro		0,0706***		0,0555**
Wrendapc		-0,0003		-0,0007
Wgini		7,5990		11,6001
Wmu_resp_alfab		-0,1199***		-0,1388***
Wtpopurb		-0,0968***		-0,0719**
Wdom_kmq		0,0026***		0,0026***
West_saude		-0,0125*		-0,0133*
Wdummy_capital		6,1212		6,2039
Wdummy_rm		-1,7318		-1,619
Constante	-5,4944***	-0,1710	-6,3937**	-1,0367
Observações	10225	10225	-	-
R2	0,02742	0,03669	-	-
R2 Ajustado	0,02627	0,03442	-	-
Estatística F	23,99***	16,19***	-	-
Lambda	-	-	0,15128***	0,14025***
Teste de Wald	-	-	97,321***	82,809***
Teste LR	-	-	79,043***	68,868***
Criterio de Inf. Akaike	-	-	89978	89914

Tabela 8A.14: Modelos SAR, SDM, SAC e GNS para a taxa de incidência de internações por DRSAl da população de 0 a 14 anos (matriz contígua)

	Y = Incidência DRSAl (de 0 a 14 anos) - Wcont			
	SAR	SDM	SAC	GNS
esg_nao_ceuab	-0,0339***	-0,0056	-0,0326***	-0,0053
ag_nao_ceuab	-0,0469***	-0,0588***	-0,0516***	-0,0589***
lixo_mor_inad	-0,001	-0,1026	-0,0255	-0,1066
d_nbueiro	0,0108	-0,0182	0,0088	-0,0184
rendapc	-0,0011	0,0002	-0,0014	0,0002
gini	13,9876***	1,6238	15,5611***	1,5408
mu_resp_alfab	-0,0183	0,0919***	-0,0157	0,0924***
tpopurb	0,1362***	0,1557***	0,1479***	0,1559***
dom_kmq	-0,0005	-0,0015***	-0,0007*	-0,0015***
est_saude	0,0034	0,0001	0,0048	0,0001
dummy_capital	-2,8909**	-7,683***	-3,4619**	-7,6827***
dummy_rm	-5,004***	-3,5105	-5,6163***	-3,4923
Wesg_nao_ceuab		-0,0347*		-0,0343*
Wag_nao_ceuab		0,0304		0,0313
Wlixo_mor_inad		0,9136		0,9246
Wd_nbueiro		0,0523**		0,0519**
Wrendapc		-0,0007		-0,0007
Wgini		10,443		10,3185
Wmu_resp_alfab		-0,1361***		-0,1358***
Wtpopurb		-0,0888***		-0,0906***
Wdom_kmq		0,0026***		0,0026***
West_saude		-0,0129*		-0,0128*
Wdummy_capital		6,5008		6,5285
Wdummy_rm		-1,0463		-0,9824
Constante	-6,2343**	-0,8823		-0,8643
Rho	0,14339	0,14212***	-0,0742***	0,158
Lambda	-	-	0,2188****	0,0173
Teste de Wald	88,405	85,399***	-	-
Teste LR	73,781	71,270***	79,805***	71,302***
Critério de Inf. Akaike	89984	89911	89980	89913

Tabela 8A.15: Modelos MQO, SLX, SEM e SDEM para a taxa de incidência de internações por DRSAI da população de 0 a 14 anos (matriz de distância)

	Y = Incidência DRSAI (de 0 a 14 anos) - Wdist_inv			
	MQO	SLX	SEM	SDEM
esg_nao_ceuab	-0,0356***	0,01644	-0,0058	0,0125
ag_nao_ceuab	-0,0486***	-0,07188***	-0,0608***	-0,0728***
lixo_mor_inad	0,0517	-0,12233	-0,1113	-0,1229
d_nbueiro	0,0154	-0,00745	0,0019	-0,0119
rendapc	-0,0009	-0,00002	-0,0008	-0,0001
gini	16,2374***	3,83142	10,9418***	4,114
mu_resp_alfab	-0,038**	0,04053	-0,0096	0,0495*
tpopurb	0,1385***	0,15641***	0,1504***	0,1605***
dom_kmq	-0,0004	-0,00091**	-0,0009**	-0,0012***
est_saude	0,0027	0,00098	0,0035	0,0011
dummy_capital	-3,1044**	-2,92363*	-3,9489**	-3,7201**
dummy_rm	-5,4576***	-0,44734	-4,1757***	-1,8186
Wesg_nao_ceuab		-0,07952***		-0,0639*
Wag_nao_ceuab		0,08336**		0,1275**
Wlixo_mor_inad		5,75485*		7,0452
Wd_nbueiro		0,05536**		0,0694
Wrendapc		-0,0002		0,0005
Wgini		11,02036		10,302
Wmu_resp_alfab		-0,06999		-0,1134*
Wtpopurb		-0,11765***		-0,1846***
Wdom_kmq		0,00345***		0,0045***
West_saude		-0,07187***		-0,0715**
Wdummy_capital		0,31883		-0,1261
Wdummy_rm		-7,50241***		-4,9044
Constante	-5,4944**	-2,55939		0,6105
Observações	10225	10225	-	-
R2	0,0274	0,03878	-	-
R2 Ajustado	0,0262	0,03652	-	-
Estatística F	23,99***	17,15***	-	-
Lambda	-	-	0,54752***	0,4803***
Teste de Wald	-	-	257,89***	165,58***
Teste LR	-	-	220,85***	165,4***
Criterio de Inf. Akaike	-	-	89140	89100

Tabela 7A.16: Modelos SAR, SDM, SAC e GNS para a taxa de incidência de internações por DRSAI da população de 0 a 14 anos (matriz de distância)

	Y = Incidência DRSAI (de 0 a 14 anos) - Wdist_inv			
	SAR	SDM	SAC	GNS
esg_nao_ceuab	-0,0201**	0,0143	0,0134	0,013
ag_nao_ceuab	-0,0478***	-0,072***	-0,0657***	-0,0691***
lixo_mor_inad	-0,068	-0,1657	-0,1598	-0,0615
d_nbueiro	0,0046	-0,0096	-0,0046	-0,0097
rendapc	-0,0005	-0,0001	-0,0004	-0,0005
gini	10,2032***	4,1771	5,9162	4,127
mu_resp_alfab	-0,0152	0,043*	0,0251	0,0525**
tpopurb	0,132***	0,1584***	0,1498***	0,1532***
dom_kmq	-0,0007*	-0,0011**	-0,0008*	-0,0013***
est_saude	0,0033	0,0016	0,0018	0,001
dummy_capital	-3,1808**	-3,4557**	-3,6584**	-4,049**
dummy_rm	-3,8795***	-1,5857	-2,5558*	-2,7278
Wesg_nao_ceuab		-0,0463*		0,0458
Wag_nao_ceuab		0,1031***		0,1905*
Wlixo_mor_inad		4,6984		9,0729
Wd_nbueiro		0,0365		0,2187**
Wrendapc		0,0001		0,002
Wgini		3,4424		10,6937
Wmu_resp_alfab		-0,0671		-0,436***
Wtpopurb		-0,1672***		-0,3065***
Wdom_kmq		0,0028***		0,0085***
West_saude		-0,0477**		-0,0723
Wdummy_capital		1,5019		-3,8383
Wdummy_rm		-2,6201		1,1159
Constante	-6,4549***	-0,2022		25,9276
Rho	0,50265***	0,47009***	-1,228***	-1,2557***
Lambda	-	-	0,89643**	0,88697***
Teste de Wald	217,06***	157,52***	-	-
Teste LR	205,65***	162,62***	323***	251,23
Critério de Inf. Akaike	89156	89102	89040	89016

9A – Resultados Tobit Durbin Espacial Bayesiano

Tabela 9A.1: Tobit Durbin Espacial Bayesiano – Inc. DRSAl (todas as idades)

Variáveis	Coefficiente	Desvio-Padrão	P-valor
dcist	0,1429***	0,039	0,250
sbceuab	-0,0015	0,210	0,626
dfrud	-0,0024	0,007	0,532
rendapc	-0,0003	0,001	0,517
gini	15,9892***	3,456	0,250
mu_resp_alfab	0,2269***	0,028	0,250
tpopurb	0,1821***	0,007	0,250
dom_kmq	-0,0001	0,000	0,385
est_saude	-0,0027	0,003	0,378
dummy_capital	-3,658**	1,711	0,263
dummy_rm	-3,7953***	1,550	0,254
Wdcist	0,0216	0,055	0,516
Wsbceuab	2,481***	0,381	0,250
Wdfrud	0,0283***	0,009	0,252
Wrendapc	-0,0006	0,001	0,465
Wgini	6,1999*	4,583	0,313
Wmu_resp_alfab	-0,2444***	0,032	0,250
Wtpopurb	-0,0908***	0,010	0,250
Wdom_kmq	0,0002*	0,000	0,334
West_saude	0,0042	0,005	0,381
Wdummy_capital	1,7968	2,041	0,397
Wdummy_rm	0,3859	1,768	0,556
constant	-19,2627***	2,413	0,250
rho	0,2446***	0,016	0,250
Número de censurados	3703	-	-
N_draws	2500	-	-
N_omit	500	-	-
N_obs	10225	-	-

Tabela 9A.2: Tobit Durbin Espacial Bayesiano – Inc DRSAI (de 0 a 14 anos)

Variáveis	Coeficiente	Desvio-Padrão	P-valor
dcist	0,243***	0,0883	0,2525
sbceuab	-0,444	0,4593	0,3750
dfrud	0,004	0,0161	0,5585
rendapc	-0,001	0,0019	0,4200
gini	43,309***	7,8996	0,2500
mu_resp_alfab	0,334***	0,0491	0,2500
tpopurb	0,43***	0,0158	0,2500
dom_kmq	-0,001**	0,0003	0,2630
est_saude	0,005	0,0060	0,4045
dummy_capital	-13,102***	3,8370	0,2505
dummy_rm	-9,753***	3,5243	0,2515
Wdcist	0,128	0,1187	0,3540
Wsbceuab	4,827***	0,8396	0,2500
Wdfrud	0,04**	0,0199	0,2680
Wrendapc	-0,002	0,0026	0,4245
Wgini	7,813	10,4811	0,4210
Wmu_resp_alfab	-0,37***	0,0595	0,2500
Wtpopurb	-0,203***	0,0225	0,2500
Wdom_kmq	0,001***	0,0004	0,2510
West_saude	0,001	0,0107	0,6125
Wdummy_capital	11,027***	4,5022	0,2555
Wdummy_rm	2,051	4,0073	0,4820
constant	-47,407***	5,6919	0,2500
rho	0,2***	0,0175	0,2500
Número de censurados	3703	-	-
N_draws	2500	-	-
N_omit	500	-	-
N_obs	10225	-	-

Tabela 9A.3: Tobit Durbin Espacial Bayesiano – Inc DRSAI (de 0 a 14 anos)

Variáveis	Coefficiente	Desvio-Padrão	P-valor
esg_nao_ceuab	0,0044	0,0115	0,5200
ag_nao_ceuab	-0,0583***	0,0154	0,2500
lixo_mor_inad	-0,5544	0,5570	0,3715
d_nbueiro	0,0107	0,0144	0,4100
rendapc	-0,0009	0,0010	0,3925
gini	12,8452***	4,3448	0,2500
mu_resp_alfab	-0,0089	0,0228	0,5095
tpopurb	0,1417***	0,0155	0,2500
dom_kmq	0,0003	0,0004	0,3815
est_saude	0,0065*	0,0045	0,3025
dummy_capital	-0,8451	1,4070	0,4595
dummy_rm	-2,3734**	1,0388	0,2620
Wesg_nao_ceuab	-0,0495**	0,0256	0,2690
Wag_nao_ceuab	-0,0172	0,0330	0,4710
Wlixo_mor_inad	0,1172	1,2300	0,5955
Wd_nbueiro	-0,0002	0,0315	0,6295
Wrendapc	-0,0028	0,0024	0,3505
Wgini	57,3316***	9,2779	0,2500
Wmu_resp_alfab	-0,2743***	0,0448	0,2500
Wtpopurb	0,3249***	0,0311	0,2500
Wdom_kmq	-0,0014**	0,0009	0,2910
West_saude	-0,0055	0,0103	0,4785
Wdummy_capital	-4,5744*	3,0740	0,3025
Wdummy_rm	-10,8401***	2,2412	0,2500
Constante	-42,1422***	6,6647	0,2500
rho	0,6237***	0,010	0,250
Número de censurados	3703	-	-
N_draws	2500	-	-
N_omit	500	-	-
N_obs	10225	-	-

10.A - Etapas do Cálculo dos custos com internação

O primeiro passo desse cálculo foi estimar o novo valor da taxa de incidência média de internações por DRSAL. Por exemplo, o aumento de um ponto percentual na porcentagem dos domicílios que não tem banheiro e estão expostos às condições de esgoto à céu aberto dos distritos gerou nova taxa de incidência de 6,431 (3,281+3,15). A partir da nova taxa de incidência e de posse da informação da população residente foi possível encontrar o número de casos correspondente a essa nova taxa (de acordo com a fórmula da taxa de incidência formula na equação 3.1). A diferença do número de casos da taxa antiga com a nova taxa deu o valor do número de casos decorrente desse aumento percentual. Dado que o custo médio por internação foi calculado pela razão entre o valor total gasto pelo SUS com todas as internações e o número de casos registrado para o ano de 2010, foi possível calcular o valor gasto com a variação do número de casos.