



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE ECONOMIA
BACHARELADO EM CIÊNCIAS ECONÔMICAS

Luiza Lins Bellon

UMA ANÁLISE ESPACIAL DA RELAÇÃO ENTRE SANEAMENTO BÁSICO E
EDUCAÇÃO NOS MUNICÍPIOS BRASILEIROS

Rio de Janeiro

2022

Luiza Lins Bellon

UMA ANÁLISE ESPACIAL DA RELAÇÃO ENTRE SANEAMENTO BÁSICO E
EDUCAÇÃO NOS MUNICÍPIOS BRASILEIROS

Trabalho de conclusão de curso apresentado ao Instituto de Economia da Universidade Federal do Rio de Janeiro como exigência para obtenção do título de Bacharel em Ciências Econômicas.

Orientadora: Professora Dra. Ana Carolina da Cruz Lima

Rio de Janeiro

2022

Ficha catalografica

LUIZA LINS BELLON

UMA ANÁLISE ESPACIAL DA RELAÇÃO ENTRE SANEAMENTO BÁSICO E
EDUCAÇÃO NOS MUNICÍPIOS BRASILEIROS

Trabalho de conclusão de curso apresentado ao Instituto de Economia da Universidade Federal do Rio de Janeiro, como requisito para a obtenção do título de Bacharela em Ciências Econômicas.

Rio de Janeiro, 5/9/2022.

ANA CAROLINA DA CRUZ LIMA - Presidente
Professora Dra. do Instituto de Economia da UFRJ

KEULER HISSA TEIXEIRA
Professor Dr. da Faculdade de Economia, Administração e Contabilidade da UFAL

VALÉRIA LÚCIA PERO
Professora Dra. do Instituto de Economia da UFRJ

AGRADECIMENTOS

Apesar da jornada um tanto solitária da escrita da monografia, não teria chegado aqui sem o apoio de uma série de pessoas. Primeiramente, agradeço aos meus pais, Ernani e Yeda, por sempre terem priorizado a minha formação, independentemente das circunstâncias; e ao meu irmão, Eduardo, a quem devo boa parte de quem sou hoje.

No âmbito acadêmico, agradeço a todo o corpo docente e funcionários do Instituto de Economia da UFRJ, em particular ao Prof. Eduardo Pontual Ribeiro, pelos ensinamentos em mais de 2 anos de bolsa PIBIC, e à Profa. Ana Carolina da Cruz Lima, pela paciência com prazos apertados e dedicação na orientação desta monografia. Agradeço também aos meus chefes na Empresa de Pesquisa Energética e na Inter.B, e às amigas que fiz nesta empresa.

Esta jornada com certeza não teria sido a mesma sem o CDE: Betina, Bia, Bianca, Clara, Fernanda, Maria, Mylena, Nathalia, Paula, Sofia e Victoria. Poder contar com uma rede de suporte de 11 mulheres, em um curso como o de Economia, foi essencial. Agradeço também ao meu amigo Leonardo, pelas infintas conversas sobre economia, ANPEC e a vida. Por fim, um obrigada especial às amigas de longa e longuíssima data, Julia Saraiva e Julia Mejias, que em breve se tornarão uma jurista e psicóloga brilhantes, e ao meu namorado, Lucas, por ser sinônimo de companheirismo e carinho antes mesmo de eu sonhar em seguir pelos caminhos da Economia.

RESUMO

Este trabalho utiliza dados de 2017 do SNIS, PNSB, IBGE e RAIS para investigar a associação entre a cobertura de esgoto municipal e indicadores educacionais no contexto brasileiro – a taxa de defasagem serial e a nota padronizada do Saeb – sob uma perspectiva espacial. A análise exploratória de dados espaciais (AEDE) identificou a presença da dependência espacial positiva para as três variáveis de interesse da monografia e para os resíduos dos modelos de regressão por MQO. Os critérios de informação Akaike e BIC, a log-verossimilhança e teste de razão de verossimilhança indicaram a seleção do Modelo do Erro Espacial de Durbin (SDEM) para o Saeb e para a distorção idade-série. No modelo do Saeb, não foi identificada significância estatística da variável local ou de lag da cobertura de esgoto. Já a dummy que denota a presença de infraestrutura de esgotamento (seja ela alta ou baixa) foi positiva e significativa, permitindo concluir que a existência de saneamento local está associada a um aumento da nota municipal do Saeb. Paralelamente, nas cidades onde a infraestrutura existe, a ampliação da cobertura de esgoto está relacionada a uma redução de cerca de 0.017% da distorção idade-série, e a diminuição alcança 0.043% quando se eleva a cobertura de municípios vizinhos, caracterizando a existência de transbordamentos espaciais locais. O Índice de Moran indicou que o modelo SDEM foi eficaz em eliminar a dependência espacial dos resíduos em ambas as regressões.

Palavras-chave: Saneamento Básico; Educação; Análise Espacial.

ABSTRACT

This paper uses 2017 data from SNIS, PNSB, IBGE and RAIS to investigate the association between municipal sewage coverage and educational indicators in the Brazilian context – the age-grade distortion and the standardized SAEB score – from a spatial perspective. The exploratory spatial data analysis (ESDA) identified the presence of positive spatial dependence for the three variables of interest in this paper as well as for the residuals of the OLS regression models. The Akaike and BIC information criterion, log-likelihood and likelihood ratio test pointed to the selection of the Spatial Durbin Error Model (SDEM) for the SAEB and age-grade distortion regressions. In the SAEB model, no statistical significance was found for the local or lagged sewage coverage rate. However, it was found for the dummy that identifies the presence of sewage infrastructure (either high or low), allowing the conclusion that the existence of local sanitation is associated with an increase in the SAEB's municipal score. In cities where the infrastructure exists, the expansion of sewage coverage is related to a reduction of 0.017% of age-grade distortion. This decrease reaches 0.043% when the sewage coverage of neighboring cities is expanded, which characterizes local spatial spillovers. The Moran's I showed that the SDEM model was effective in eliminating the spatial dependence of residuals in both regressions.

Keywords: Sanitation; Education; Spatial Analysis.

LISTA DE FIGURAS

Figura 1 – População atendida com abastecimento de água (Brasil, 2017)	28
Figura 2 – População atendida com coleta de esgoto (Brasil, 2017).....	28
Figura 3 – Nota padronizada do Saeb, por município (Brasil, 2017) ²⁹ Erro! Indicador não definido.	
Figura 4 – Distorção idade-série, por município (Brasil, 2017).....	30
Figura 5 – Convenções de contiguidade mais comuns.....	32
Figura 6 – Modelos de regressão espacial	36
Figura 7 – Diagramas de dispersão de Moran: Saeb e distorção idade-série	41
Figura 8 – Índice de Moran local, notas do Saeb (Brasil, 2017)	42
Figura 9 – Índice de Moran local, distorção idade-série (Brasil, 2017)	43
Figura 10 – Índice de Moran local bivariado: Saeb (X) e coleta de esgoto (Y) (Brasil, 2017)	44
Figura 11 – Índice de Moran local bivariado: distorção idade-série (X) e coleta de esgoto (Y) (Brasil, 2017).....	45
Figura 12 – Resultados do modelo SDEM para as notas do Saeb.....	47
Figura 13 – Resultados do modelo SDEM para a distorção idade-série	50
Figura 14 – SDM: notas do Saeb.....	60
Figura 15 – SLX: notas do Saeb.....	61
Figura 16 – SEM: notas do Saeb	62
Figura 17 – SAR: notas dos Saeb	63
Figura 18 – SARAR: notas do Saeb	64
Figura 19 – MQO: notas do Saeb	65
Figura 20 – SDM: distorção idade-série.....	66
Figura 21 – SLX: distorção idade-série	67
Figura 22 – SEM: distorção idade-série	68
Figura 23 – SAR: distorção idade-série.....	69
Figura 24 – SARAR: distorção idade-série	70
Figura 25 – MQO: distorção idade-série	74

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas	24
Tabela 2 – População atendida com abastecimento de água e coleta de esgoto	27
Tabela 3 – Índices de Moran dos resíduos para seleção da matriz de vizinhança.....	39
Tabela 4 – Índices de Moran globais: indicadores educacionais.....	40
Tabela 5 – Resultados dos testes multiplicadores de Lagrange – Saeb.....	46
Tabela 6 – Estatísticas de ajuste dos modelos – Saeb	46
Tabela 7 – Resultados dos testes multiplicadores de Lagrange – distorção idade-série	49
Indicador não definido.	
Tabela 8 – Estatísticas de ajuste dos modelos – distorção idade-série.....	49

SUMÁRIO

INTRODUÇÃO	11
1 SANEAMENTO BÁSICO E EDUCAÇÃO	13
2 BASE DE DADOS E METODOLOGIA	21
2.1. Base de dados e modelos econométricos	21
2.2. Contextualização e caracterização espacial	25
2.3. Metodologia	31
3. RESULTADOS	39
3.1. Escolha da matriz de pesos espaciais	39
3.2. Análise exploratória de dados espaciais	40
3.3 modelos econométricos.....	45
CONSIDERAÇÕES FINAIS.....	53
REFERÊNCIAS	55
APÊNDICE A – RESULTADOS DOS MODELOS ECONOMÉTRICOS PARA O SAEB.....	60
APÊNDICE B – RESULTADOS DOS MODELOS ECONOMÉTRICOS PARA DISTORÇÃO IDADE-SÉRIE	66

INTRODUÇÃO

No Brasil, o direito ao saneamento básico é assegurado pela Constituição Federal¹ e pela Lei nº 11.445/2007², que estabelece as diretrizes nacionais do setor. Contudo, segundo dados do Sistema Nacional de Informações de Saneamento (SNIS), cerca de 40 milhões de brasileiros não têm ligação à rede de abastecimento de água, e 94,1 milhões encontram-se marginalizados da infraestrutura de esgotamento sanitário (SNIS, 2020). Tal cenário compõe o legado histórico de décadas de uma regulação permissiva sobre as companhias de saneamento, ausência de metas operacionais vinculantes e iniciativas desarticuladas no âmbito das políticas públicas no setor (TUROLLA, 2002)

A insuficiência da cobertura leva a problemas de primeira ordem sobre o bem-estar das famílias e a produtividade dos trabalhadores. O contato com água contaminada por bactérias, parasitas e insetos vetores, assim como a exposição ao esgoto a céu aberto, estão associados a doenças cujos efeitos imediatos incluem febres, hemorragias e anemia (HUTTON et al., 2012), impondo a interrupção de atividades de estudo e trabalho e um aumento da morbidade hospitalar. Em casos mais graves, podem gerar sequelas ou levar ao óbito.

As crianças configuram um dos recortes da população mais afetados pela ausência do saneamento, dado que são mais suscetíveis à contaminação por doenças infecciosas (NGURE et al., 2014) e podem ter seu desenvolvimento prejudicado por períodos contínuos de baixa absorção de nutrientes, com efeitos de longo prazo sobre habilidades sociais e cognitivas. Ademais, a contaminação frequente por doenças gastrointestinais pode reduzir o aproveitamento escolar ao impor o absenteísmo.

De fato, há evidências de que a ampliação do saneamento é um preditor para uma elevação da frequência escolar (DREIBELBIS et al., 2013) e redução da taxa de distorção idade-série (AGOL et al., 2018), inclusive por meio de uma relação causal. Ademais, a instalação da infraestrutura tem efeitos de longo prazo sobre o total de anos de estudo em locais previamente desabastecidos por água encanada (ZHANG; XU, 2016), e estudantes mulheres costumam ser mais afetadas pela ausência do serviço, visto que a inadequação de instalações sanitárias eleva o absenteísmo escolar feminino durante o período menstrual (RYLEY; JEWITT, 2014).

¹ BRASIL. [Constituição (1988)]. Constituição da República Federativa do Brasil de 1988. Brasília, DF. Disponível em http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

² BRASIL. Lei nº 11.445, de 05 de janeiro de 2007. Disponível em http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/lei/111445.htm

No caso brasileiro, há estudos que relacionam o saneamento ao aumento da incidência de doenças gastrointestinais (SANTIAGO, 2018; VENTURA; LOPES, 2017; UHR et al., 2016) e à mortalidade infantil (OLIVEIRA, 2016; BARUFI et al., 2012; SILVA, ESPEREDIÃO, 2017), mas há escassa literatura sobre a sua associação com indicadores educacionais. Scriptore (2016) identifica que a ampliação do saneamento leva a uma maior frequência escolar e redução da distorção idade-série, e Ortiz-Correa et al. (2016) demonstram que a conexão residencial duradoura com água e esgoto contribui para elevar os anos médios de estudo das crianças do domicílio,

Tendo em vista os potenciais ganhos de bem-estar associados ao saneamento, assim como a heterogeneidade na provisão do serviço nos municípios brasileiros, a monografia busca estudar a relação entre saneamento básico e educação nos municípios brasileiros. Particularmente, é investigada a interação entre a cobertura de esgotamento municipal e dois indicadores educacionais: a taxa de distorção idade-série e a nota padronizada do Sistema de Avaliação da Educação Básica (Saeb).

A existência da autocorrelação espacial entre os indicadores educacionais e sanitários – isto é, a correlação entre o valor de uma observação em um local e o valor de observações de espaços vizinhos – requer o uso de ferramentas econométricas que adereçam esse fenômeno de maneira apropriada (ALMEIDA, 2012), uma vez que a independência entre variáveis compõe uma hipótese comum para a consistência e ausência de viés em procedimentos de estimação.

Portanto, esta monografia analisa os dados sob o prisma da econometria espacial. Será realizada uma análise exploratória de dados espaciais (AEDE) para identificar o padrão da autocorrelação espacial nas variáveis do Saeb, distorção idade-série e cobertura de esgoto. Em seguida, os modelos econométricos incorporam componentes de defasagem espacial na variável dependente, variáveis explicativas e termo de erro, e é escolhido aquele com o melhor ajuste dos dados. Esta abordagem também é utilizada por Scriptore (2016), que encontra transbordamentos espaciais significativos da provisão de água e esgoto sobre indicadores educacionais de municípios vizinhos.

A monografia está estruturada da seguinte forma. O Capítulo 1 apresenta o marco teórico utilizado para analisar os mecanismos de interação entre saneamento básico e educação, bem como as principais evidências empíricas sobre o tema. O Capítulo 2 descreve as características da base de dados e a metodologia utilizada, além de prover uma contextualização sobre a problemática do saneamento básico no Brasil e caracterização espacial dos dados. O Capítulo 3 fornece os resultados da análise exploratória de dados espaciais e dos modelos econométricos estimados. Em seguida são expostas as considerações finais do estudo.

1 SANEAMENTO BÁSICO E EDUCAÇÃO

Os efeitos do saneamento básico em indicadores socioeconômicos são amplamente documentados, sobretudo no âmbito da saúde pública. O consumo de água não tratada e a exposição ao esgoto a céu aberto constituem fatores de risco para uma série de enfermidades, convencionalmente chamadas Doenças Relacionadas ao Saneamento Ambiental Inadequado (DRSAI), tais como cólera, febre tifoide, leptospirose, teníase, dengue etc. (HUTTON et al., 2012). Os efeitos imediatos dessas doenças incluem diarreia, febre, dores abdominais, hemorragias e anemia, podendo levar a óbito em casos mais graves.

De fato, em estudo recente sobre o descontingenciamento de crédito para projetos de saneamento entre 2007 e 2017, Sant’Anna e Rocha (2021) apontam reduções significativas nas internações hospitalares de crianças de até 05 (cinco) anos e idosos associadas à expansão da cobertura dos serviços, resultados similares aos de Teixeira et al. (2012) e Uhr et al. (2016). A associação entre falta de saneamento e mortalidade – tanto adulta quanto infantil – encontrada é evidenciada, na literatura internacional, por Hutton et al. (2007), Banerjee e Duflo (2011) e Benova et al. (2014), e, no caso do Brasil, por Barufi et al. (2012) e Silva e Esperedião (2017). Quando não causam óbito, as DRSAI podem gerar sequelas, como atrasos cognitivos, má formação fetal e susceptibilidade a outras doenças (KOTHARI et al., 2019), além de compelir o afastamento das atividades de trabalho e estudo, prejudicando a produtividade.

Estima-se que as DRSAI sejam responsáveis por cerca de 270 mil internações anuais na rede pública brasileira e gastos que superam R\$ 100 milhões (TRATA BRASIL, 2021). Neste sentido, Pessanha et al. (2011) demonstram que o investimento na construção de uma estação de tratamento de esgoto em Campo dos Goytacazes (RJ) seria recuperado em 09 (nove) anos e 10 (dez) meses em custos evitados com tratamento de enfermos. Outros impactos econômicos da provisão do serviço incluem a valorização de imóveis e a viabilização de atividades econômicas que dependem de condições ambientais adequadas, como o turismo e a pesca (TRATA BRASIL, 2018).

Uma das implicações mais relevantes do saneamento básico para as políticas públicas reside em sua relação com o desempenho escolar. Dado que as crianças são mais suscetíveis ao contágio por doenças infecciosas (NGURE et al., 2014; UHR et al., 2016), elas são particularmente afetadas pela ausência do serviço: a desidratação e anemia, características de doenças diarreicas, estão associadas à baixa absorção de nutrientes, comprometendo o desenvolvimento físico e cognitivo. Cabe destacar que a nutrição adequada durante a gestação

e os dois primeiros anos de vida estabelece as fundações para habilidades sociais e intelectuais futuras, rendimento escolar e produtividade (NGURE et al., 2014).

Uma fundamentação teórica sobre a relação entre o saneamento básico e indicadores educacionais pode ser extraída de Glewwe e Miguel (2007), que propõem um modelo para analisar o impacto da saúde e da nutrição sobre a educação. O modelo assume dois períodos: o primeiro engloba a faixa etária de 0 (zero) a 6 (seis) anos, quando as crianças costumam ser matriculadas na escola, e o segundo vai de 6 (seis) até 11 (onze) anos de idade, fase em que as crianças passam pela educação primária. Os autores formulam uma função de produção para as habilidades acadêmicas, denotadas por T_2 e mensuradas pelas notas escolares no período 2:

$$T_2 = T_{2,P}(H_1, H_2, EI_1, EI_2, \alpha, SC, YS) \quad (1)$$

Na qual o subscrito P indica a função de produção; H_t é a saúde da criança no período t; EI_t configura o “esforço educacional dos pais” no período t – em termos de materiais escolares, livros, jogos lúdicos e tempo gasto em atividades pedagógicas; α é a habilidade inata; SC representa a infraestrutura escolar; e YS os anos de estudo no período 2. Todas as variáveis têm impacto positivo direto em T_2 , e assume-se que as características escolares são mantidas constantes para fins de simplificação.

Dados os esforços educacionais dos pais, a infraestrutura escolar, as habilidades inatas e os anos de estudo, a capacidade de aprendizado da criança é afetada por seu estado de saúde. Dessa forma, a equação (1) mensura relações ‘estruturais’, pois representa o impacto direto das variáveis nas habilidades acadêmicas. Tais relações podem ser aprofundadas ao se considerar que os pais têm preferências em relação aos gastos pecuniários e de tempo, por meio de uma função de utilidade:

$$U = U(C_1, C_2, HI_1, HI_2, \alpha, T_2) \quad (2)$$

Na qual C_t é o consumo parental de um bem de consumo agregado no período t. Já a saúde da criança nos períodos 1 e 2 assume duas funções de produção:

$$H_1 = H_{1,P}(C_1^C, M_1, HE_1, \eta) \quad (3)$$

$$H_2 = H_{2,P}(C_2^C, M_2, HE_2, \eta) \quad (4)$$

Nas equações (3) e (4), C_t^C é o consumo do bem de consumo agregado no período t ; M_t são os gastos com a saúde da criança no período t ; HE_t é o ambiente sanitário em que a criança vive no período t – como, por exemplo, qualidade da água ingerida, existência de saneamento básico e incidência de doenças infecciosas; e η é a saúde inata da criança. Todas as variáveis em (3) e (4) têm impacto positivo direto na saúde, e assume-se que tanto HE_t como η estão fora do controle dos pais.

O modelo pressupõe que o montante de lazer consumido é fixo, de modo que o uso do tempo – com, por exemplo, cuidados com a saúde e educação da criança – é limitado por uma restrição orçamentária intertemporal. Sendo W_0 o nível inicial de riqueza do domicílio, que pode emprestar e tomar emprestado entre os dois períodos à taxa de juros r :

$$W_0 = p_{C,1}(C_1 + C_1^C) + p_{C,2} \frac{(C_2 + C_2^C)}{1+r} + p_{M,1}M_1 + p_{EI}EI_1 + \frac{p_{M,2}M_2 + p_{EI}EI_2 + p_S YS}{1+r} \quad (5)$$

Na qual $p_{C,t}$ representa o preço do bem de consumo agregado no período t ; $p_{M,t}$ é o preço de medicamentos em t , que também inclui custos de transporte; EI_t são os insumos educacionais adquiridos em t ; p_{EI} o preço de tais insumos (por hipótese, constante); e PS é o preço do ano de estudo em t_2 .

Assim, a utilidade é maximizada tendo em vista a restrição orçamentária, a função de produção das habilidades acadêmicas e as duas funções de produção para a saúde da criança. A otimização da função de utilidade (2) sujeita às restrições (1), (3), (4) e (5) fornece as respectivas funções de demanda relativas às nove variáveis endógenas³ que podem ser compradas no mercado:

$$C_t = H_{t,D}(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{EI}, p_S; HE_1, HE_2, SC, PS, \alpha, \eta, \sigma, \tau), t = 1, 2 \quad (6) \text{ e } (7)$$

$$C_t^C = C_{t,D}^C(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{EI}, p_S; HE_1, HE_2, SC, PS, \alpha, \eta, \sigma, \tau), t = 1, 2 \quad (8) \text{ e } (9)$$

$$M_t = M_{t,D}(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{EI}, p_S; HE_1, HE_2, SC, PS, \alpha, \eta, \sigma, \tau), t = 1, 2 \quad (10) \text{ e } (11)$$

$$EI_t = M_{t,D}(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{EI}, p_S; HE_1, HE_2, SC, PS, \alpha, \eta, \sigma, \tau), t = 1, 2 \quad (12) \text{ e } (13)$$

$$YS = YS_D(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{EI}, p_S; HE_1, HE_2, SC, PS, \alpha, \eta, \sigma, \tau) \quad (14)$$

³ Nesse caso, as variáveis endógenas referem-se àquelas que podem ser influenciadas pelo comportamento dos pais, não tendo relação com o significado econométrico da palavra “endógeno”.

Nas equações acima, D indica as funções de demanda; PS é a escolaridade dos pais (uma vez que essa pode impactar a efetividade dos gastos com educação dos filhos); σ representa as preferências dos pais sobre a educação da criança; e τ são as preferências dos pais sobre a saúde da criança. Todas as variáveis do lado direito das equações são exógenas, isto é, não podem ser controladas pelos pais.

A inserção das equações (12), (13) e (14) em (1), assim como a inserção de (8) e (10) em (3); (3), (9) e (11) em (4) e (3) e (4) em (1) fornece a seguinte função de demanda pelas habilidades acadêmicas da criança:

$$T_2 = T_{2,D}(W_0; r, p_{C,1}, p_{C,2}, p_{M,1}, p_{M,2}, p_{EI}, p_S; HE_1, HE_2, SC, PS, \alpha, \eta, \sigma, \tau) \quad (15)$$

A equação (15) demonstra como mudanças em variáveis exógenas afetam o nível de saúde dos jovens e, conseqüentemente, seu desempenho acadêmico. A vantagem de estimar a função de demanda (15) em detrimento da função de produção (1) é que, enquanto (1) mensura apenas o impacto direto (estrutural) das variáveis, (15) permite analisar todas as mudanças comportamentais decorrentes de alterações no ambiente sanitário. Por este motivo, Glewwe e Miguel (2007) recomendam que a análise do impacto de políticas sanitárias em indicadores educacionais siga este referencial. Os autores tomam como exemplo o potencial efeito da melhoria da infraestrutura de saneamento básico, por meio da diferenciação de (15) em relação à variável do ambiente sanitário no período 2 (HE_2):

$$\begin{aligned} \frac{\partial T_{2,D}}{\partial HE_2} &= \frac{\partial T_P}{\partial HE_2} \left[\frac{\partial H_{2,P}}{\partial HE_2} + \frac{\partial H_{2,P}}{\partial M_2} \frac{\partial M_{2,D}}{\partial HE_2} + \frac{\partial H_{2,P}}{\partial C_2^C} \frac{\partial C_{2,D}^C}{\partial HE_2} + \frac{\partial H_{2,P}}{\partial H_1} \left(\frac{\partial H_{1,P}}{\partial M_1} \frac{\partial M_{1,D}}{\partial HE_2} + \frac{\partial H_{1,P}}{\partial C_1^C} \frac{\partial C_{1,D}^C}{\partial HE_2} \right) \right] + \\ &\frac{\partial T_{CD}}{\partial H_1} \left[\frac{\partial H_{1,P}}{\partial M_1} \frac{\partial M_{1,D}}{\partial HE_2} + \frac{\partial H_{1,P}}{\partial C_1^C} \frac{\partial C_{1,D}^C}{\partial HE_2} \right] + \frac{\partial T_P}{\partial EI_1} \frac{\partial EI_{1,D}}{\partial HE_2} + \frac{\partial T_P}{\partial EI_2} \frac{\partial EI_{2,D}}{\partial HE_2} + \frac{\partial T_P}{\partial YS} \frac{\partial YS_D}{\partial HE_2} = \\ &\frac{\partial T_P}{\partial H_2} \left[\frac{\partial H_{2,P}}{\partial HE_2} + \frac{\partial H_{2,P}}{\partial M_2} \frac{\partial M_{2,D}}{\partial HE_2} + \frac{\partial H_{2,P}}{\partial C_2^C} \frac{\partial C_{2,D}^C}{\partial HE_2} \right] + \frac{\partial T_P}{\partial EI_2} \frac{\partial EI_{2,D}}{\partial HE_2} + \frac{\partial T_P}{\partial YS} \frac{\partial YS_D}{\partial HE_2} \end{aligned} \quad (16)$$

A primeira conclusão que pode ser obtida a partir deste exercício é que a provisão de saneamento básico terá efeitos diferentes sobre crianças em idade pré-escolar e crianças em uma fase posterior da educação. Embora o choque em HE afete tanto H_1 quanto H_2 , os domicílios que já se encontram no período 2 – isto é, nos quais moram crianças acima de 6 (seis) anos – não poderão alterar M_1 , C_1^C , C_1 e EI_1 frente às mudanças nas condições sanitárias, e, portanto, experimentarão um impacto apenas de curto prazo. Em contrapartida, domicílios

que ainda estão em t_1 no momento da provisão do serviço poderão alterar suas decisões de consumo no curto e no longo prazo.

No curto prazo, o impacto positivo estrutural de HE_2 em T_2 , via $\left(\frac{\partial T_P}{\partial H_2}\right)\left(\frac{\partial H_{2,P}}{\partial HE_2}\right)$, aumenta as possibilidades de consumo no período 2, de forma a gerar um efeito renda positivo. Contudo, a depender do efeito marginal de M_2 e C_2^C em H_2 , o impacto resultante em T_2 é ambíguo. Por exemplo, a intervenção pode mitigar a incidência de doenças infecciosas, e, conseqüentemente, reduzir a demanda por medicamentos ($\frac{\partial M_{2,D}}{\partial HE_2} < 0$). Em contrapartida, pode-se argumentar que a redução da exposição a fatores de riscos tem o potencial de aumentar os benefícios de tais remédios de forma a elevar sua demanda ($\frac{\partial M_{2,D}}{\partial HE_2} > 0$).

Mesmo no caso de um efeito marginal negativo de M_2 sobre a saúde H_2 , melhorias no ambiente sanitário podem aumentar a absorção de nutrientes e tornar a alimentação infantil mais efetiva, estimulando uma maior demanda por C_2^C de maneira que $\frac{\partial C_{2,D}^C}{\partial HE_2} > 0$. Como resultado, o sinal do termo $\frac{\partial H_{2,P}}{\partial M_2} \frac{\partial M_{2,D}}{\partial HE_2} + \frac{\partial H_{2,P}}{\partial C_2^C} \frac{\partial C_{2,D}^C}{\partial HE_2}$ é incerto.

Similarmente, o choque de HE_2 na demanda por EI_2 e YS_D pode gerar sinais positivos ou negativos. O efeito renda cria incentivos a desviar recursos de produção de T_p para maior consumo de C_2 , enquanto o efeito preço – ou seja, a redução do preço sombra de T_2 – estimula um aumento de T_2 , de modo que EI_2 e YS_D (duas variáveis que afetam T_2) podem aumentar ou diminuir em resposta à mudança de HE_2 . Conclui-se que não é claro se o impacto de curto prazo de uma melhoria de HE_2 na demanda por T_p será superior ou inferior ao impacto estrutural positivo medido por $\frac{\partial T_P}{\partial H_2} \frac{\partial H_{2,P}}{\partial HE_2}$.

No longo prazo, um choque em HE pode ser descrito como:

$$\begin{aligned} \frac{\partial T_{2,D}}{\partial HE_2} = & \frac{\partial T_P}{\partial H_2} \left[\frac{\partial H_{2,P}}{\partial HE} + \frac{\partial H_{2,P}}{\partial M_2} \frac{\partial M_{2,D}}{\partial HE} + \frac{\partial H_{2,P}}{\partial C_2^C} \frac{\partial C_{2,D}^C}{\partial HE} + \frac{\partial H_{2,P}}{\partial H_1} \left(\frac{\partial H_{1,P}}{\partial HE} + \frac{\partial H_{1,P}}{\partial M_1} \frac{\partial M_{1,D}}{\partial HE} + \frac{\partial H_{1,P}}{\partial C_1^C} \frac{\partial C_{1,D}^C}{\partial HE} \right) \right] \\ & + \frac{\partial T_{CD}}{\partial H_1} \left[\frac{\partial H_{1,P}}{\partial M_1} \frac{\partial M_{1,D}}{\partial HE} + \frac{\partial H_{1,P}}{\partial C_1^C} \frac{\partial C_{1,D}^C}{\partial HE} \right] + \frac{\partial T_P}{\partial EI_1} \frac{\partial EI_{1,D}}{\partial HE} + \frac{\partial T_P}{\partial EI_2} \frac{\partial EI_{2,D}}{\partial HE} + \frac{\partial T_P}{\partial YS} \frac{\partial YS_D}{\partial HE} \end{aligned} \quad (17)$$

Novamente, embora o efeito estrutural $\frac{\partial T_P}{\partial H_2} \frac{\partial H_{2,P}}{\partial HE} + \frac{\partial H_{2,P}}{\partial H_1} \frac{\partial H_{1,P}}{\partial HE}$ seja positivo, não é possível determinar o sinal exato de $\frac{\partial H_{1,P}}{\partial M_1} \frac{\partial M_{1,D}}{\partial HE} + \frac{\partial H_{1,P}}{\partial C_1^C} \frac{\partial C_{1,D}^C}{\partial HE}$. Logo, é preciso analisar os efeitos renda e preço. Como o domicílio poderá ajustar as preferências de consumo ainda no período

1, o efeito renda tem o potencial de resultar em um aumento de C_1 de forma a desviar recursos que ampliariam T_2 . Além disso, repete-se o problema de indefinição do sinal relativo ao efeito marginal de EI_1 .

Assim, tanto no curto como no longo prazo não é claro se uma mudança no ambiente sanitário levará a um impacto na demanda por T_2 positivo ou negativo, ou se esse superará o efeito estrutural da saúde sobre a educação, que é positivo. Ademais, há a possibilidade de os efeitos estruturais serem mais relevantes em um período que no outro, com implicações para o desenho de políticas públicas: intervenções no setor de saneamento podem ser mais efetivas no curto ou no longo prazo.

Os autores destacam que, apesar das dificuldades envolvidas na mensuração desses efeitos, a evidência empírica tem apontado para uma relação positiva significativa – inclusive causal – entre a saúde infantil e indicadores educacionais. Neste sentido, cabe apresentar uma revisão da literatura recente sobre o impacto do ambiente sanitário – especificamente, das condições de saneamento básico – na educação.

Um dos indicadores educacionais mais utilizados para essa investigação é frequência escolar, uma vez que a falta de saneamento pode impor o afastamento constante das atividades de estudo. Dreibelbis et al. (2013) apontam que a ausência do serviço em residências e escolas é um preditor significativo para o absenteísmo escolar no Quênia, e que esse resultado pode ser decorrente tanto da incidência de DRSAI, quanto do fato de que estudantes tendem a evitar completamente o ambiente escolar se a infraestrutura de saneamento não for adequada, de modo que a qualidade do esgotamento em escolas – em termos de limpeza e integridade – é mais relevante para a redução do absenteísmo que sua quantidade. Similarmente, a instalação de latrinas em escolas indianas foi responsável pelo aumento na taxa de matrículas e redução da evasão estudantil (ADUKIA, 2016).

Zhang e Xu (2016) examinam os efeitos de longo prazo da construção da infraestrutura de abastecimento de água nos domicílios rurais da China nos anos 1980 através de um modelo de variáveis instrumentais com efeitos fixos de ano e condado. Os resultados estatísticos indicam que o programa foi responsável por elevar o tempo médio de estudo em 1.1 anos, e que o aumento foi superior para crianças mais novas (0 a 2 anos) em comparação àquelas que só obtiveram acesso ao serviço em fases posteriores da educação.

As consequências da falta de saneamento sobre a educação também têm um viés por sexo. Isso ocorre devido à inadequação geral da infraestrutura escolar para higiene pessoal durante o período menstrual (JEWITT; RYLEY, 2014), que pode levar a maiores taxas de absenteísmo entre estudantes mulheres. Além disso, nos locais sem conexão com a rede de

abastecimento, o fardo da coleta e armazenamento de água recai sobre as mulheres (UNICEF; ONU, 2012), o que poderia configurar outro fator para o absentéismo feminino (GARN et al., 2013). Freeman et al. (2012) e Agol et al. (2018) apontam evidências nesse sentido para o tratamento de água em escolas no Zâmbia e para a infraestrutura de esgotamento escolar no Quênia, respectivamente. Carecem estudos similares para o Brasil e América Latina.

As evidências apresentadas estão em consonância com o conhecimento acumulado sobre determinantes do desempenho escolar, tipicamente associados a condições escolares, *background* familiar e estado de saúde do estudante (SCRIPTORE, 2016). De maneira geral, a literatura internacional tem focado nas externalidades da infraestrutura de saneamento escolar em detrimento da domiciliar, enquanto na literatura nacional o inverso ocorre.

Ortiz-Correa et al. (2016) usam dados de 2010 do Censo Demográfico, fornecidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE), e da Agência Nacional de Águas (ANA) para investigar como o acesso aos serviços de água e esgoto afeta o nível de escolaridade, por meio de um modelo de variáveis instrumentais com efeitos fixos de município. Os autores concluem que o total de anos de estudo cresce de acordo o tempo de conexão do indivíduo com a rede de água e esgoto, indicando que os serviços contribuem para efeitos de bem-estar cumulativos com o decorrer do tempo. O investimento em saneamento é, portanto, essencial para quebrar o ciclo vicioso de pobreza: a precariedade das redes de água e esgoto se traduzem em piores indicadores educacionais e, conseqüentemente, em menor acumulação de capital humano.

Scriptore (2016) utiliza dados do Censo Demográfico e do Censo Escolar de 2010 e identifica que a elevação de 1 p.p. no acesso a serviços de saneamento está associada a um aumento de 0,11 p.p. na frequência escolar e a reduções de 0,12 e 0,31 p.p. na taxa de evasão e distorção idade-série no ensino fundamental, respectivamente. A autora insere componentes da econometria espacial na análise dos efeitos do saneamento na educação, como a realização de uma análise exploratória de dados espaciais (AEDE) e a estimação por meio de modelos espaciais⁴.

Adicionalmente, são identificados efeitos de transbordamento espacial para a incidência de DRSAI nos municípios, o que pode ser explicado pelo deslocamento de insetos vetores de um distrito para outro, municípios que despejam esgoto em cidades vizinhas, rios e

⁴ Outros autores utilizaram AEDE para investigar a distribuição espacial do saneamento básico no território e sua relação com indicadores sanitários – como Cruz et al. (2021) e Costa et al. (2020) –, mas carecem estudos que explorem a associação espacial entre saneamento e indicadores educacionais. Similarmente, Barufi et al. (2012) utilizam modelos espaciais com efeitos fixos de municípios para investigar os efeitos do saneamento básico para o capital humano, mas a análise se limita ao impacto sobre a mortalidade infantil.

correntes marítimas que atravessam localidades e pela própria definição de doenças infectocontagiosas. Em trabalho posterior, Scriptore e Azzoni (2020) apontam transbordamentos espaciais significativos na taxa de matrícula escolar como consequência da ampliação do saneamento básico, o que também pode ser explicado pela redução da incidência de DRSAI entre crianças e adolescentes. Dessa forma, a coordenação do investimento entre regiões próximas é fundamental para incrementar a eficácia de políticas públicas em saneamento.

Nessa perspectiva, a monografia contribui para o tema de duas maneiras. Primeiramente, pela utilização de uma base de dados composta por uma combinação de informações do Sistema Nacional de Informações sobre Saneamento (SNIS) e da Pesquisa Nacional de Saneamento Básico (PNSB) para o ano de 2017, o que permitirá identificar tendências mais recentes e compatíveis com as mudanças socioeconômicas enfrentadas pelo país em comparação aos estudos existentes, que utilizam dados do Censo Demográfico 2010. Ademais, além da taxa de distorção idade-série – usualmente empregada na literatura internacional sobre as externalidades do saneamento –, é analisada a relação da ausência do serviço com uma medida quantitativa da avaliação do ensino: a nota padronizada do Saeb.

2 BASE DE DADOS E METODOLOGIA

O objetivo geral da monografia é investigar a relação entre saneamento básico e educação nos municípios brasileiros sob uma perspectiva espacial. A análise divide-se em dois objetivos específicos: (i) examinar como os indicadores de cobertura de esgoto e de qualidade do ensino – particularmente, a nota padronizada do Saeb e a taxa de distorção idade-série – variam no território, e se tal distribuição pode ser resultado da dependência espacial entre as observações; e (ii) investigar a associação entre a ampliação da cobertura do saneamento (variável dependente) e os índices educacionais (variáveis de interesse).

A seguir são apresentadas as principais características da base de dados e da metodologia utilizada, e além de uma breve contextualização sobre o arcabouço regulatório e panorama atual da provisão de saneamento no Brasil.

2.1. Base de dados e modelos econométricos

Todos os dados utilizados são georreferenciados e possuem desagregação municipal. Os dados sobre a cobertura de esgoto, utilizados para a construção da variável dependente do modelo espacial, foram obtidos por intermédio do Sistema Nacional de Informações sobre Saneamento (SNIS). As informações sobre indicadores educacionais, utilizadas para a construção das variáveis de interesse, provêm do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Informações da Relação Anual de Informações Sociais (RAIS) e do Instituto Brasileiro de Geografia e Estatística (IBGE) são utilizadas para construir variáveis de controle no âmbito municipal.

Embora constitua a principal fonte oficial de dados atualizados sobre saneamento, o SNIS possui uma alta quantidade de dados faltantes, visto que a pesquisa acontece em caráter voluntário para as prestadoras de serviço e carece de instrumentos de auditoria. Além disso, a forma de exibição dos dados faltantes não permite diferenciar municípios nos quais não há oferta de serviço de localidades nas quais as informações existem, mas não foram divulgadas.

Para contornar esse problema, as informações do SNIS foram combinadas com a Pesquisa Nacional de Saneamento Básico (PNSB), cuja última edição ocorreu em 2017. Entre outros objetivos, a PNSB identifica quais municípios dispõem de abastecimento de água e coleta de esgoto, de modo que para as cidades onde esses serviços são ausentes a cobertura foi computada como 0% na base de dados. Já nas cidades em que a PNSB reporta haver serviços,

mas que o SNIS não dispõe de valores, as observações permaneceram como dados faltantes (NA). Dessa forma, foi possível reduzir o número de NAs da variável de 3.128 para 843. A fim de manter a consistência, optou-se pelo ano de 2017 para coleta de todas as informações do trabalho.

A importância desta etapa reside no fato que a análise econométrica espacial requer a identificação da dependência espacial entre observações de diferentes pontos geográficos e dos efeitos da interação entre municípios vizinhos. Assim, a falta de dados pode reduzir a precisão das estimativas e impossibilitar a estimação correta de transbordamentos espaciais, sendo mais grave para observações que compõem *clusters* no território (ARBIA et al, 2016).

A manipulação dos dados faltantes resultou em grande quantidade de zeros para a variável da cobertura de esgoto: cerca de 42% das observações, gerando receios de que a elevada distorção na distribuição não permitisse que o efeito linear do modelo fosse devidamente acomodado pela variável independente da coleta de esgoto. A fim de adereçar esse problema, foi introduzida uma *dummy* igual a 1, caso a cobertura de saneamento municipal seja nula, e igual a 0, caso ela exista. Dessa forma, a coeficiente da variável contínua da cobertura passa a representar o efeito da ampliação do saneamento em cidades em que a infraestrutura é presente.

Na esfera dos dados educacionais, foram coletadas informações agregadas sobre os anos iniciais⁵ do Ensino Fundamental na rede pública relativas à preparação docente, infraestrutura escolar e qualidade do ensino – como as notas do Sistema de Avaliação da Educação Básica (Saeb). O Saeb consiste em um exame padronizado anual que avalia o desempenho dos alunos do 5º e 9º anos do Ensino Fundamental e 3ª série do Ensino Médio em português e matemática.

Alguns indicadores do Inep foram modificados para facilitar a análise. Cabe comentar sobre três deles: adequação, esforço e regularidade docentes. A adequação docente representa a porcentagem de professores com formação superior na área que lecionam, enquanto o esforço docente diz respeito à parcela de docentes nos níveis 5 e 6 do indicador homônimo do Inep. O Quadro 1 fornece mais detalhes sobre as categorias do Esforço Docente. Para os anos iniciais do Ensino Fundamental, consideram-se os níveis 5 e 6 como indicativos de sobrecarga na profissão (INEP, 2014).

⁵ A escolha pelos anos iniciais do Ensino Fundamental (5 a 11 anos), em detrimento dos anos finais ou do Ensino Médio, se deu em virtude da maior susceptibilidade das crianças a doenças relacionadas à falta de saneamento, como citado na Seção 2.

Quadro 1: Detalhamento do indicador de Esforço Docente

Níveis	Descrição
Nível 1	Docentes que têm até 25 alunos e atuam em um único turno, escola e etapa.
Nível 2	Docentes que têm entre 25 e 150 alunos e atuam em um único turno, escola e etapa.
Nível 3	Docentes que têm entre 25 e 300 alunos e atuam em um ou dois turnos em uma única escola e etapa.
Nível 4	Docentes que têm entre 50 e 400 alunos e atuam em dois turnos, em uma ou duas escolas e em duas etapas.
Nível 5	Docentes que têm mais de 300 alunos e atuam nos três turnos, em duas ou três escolas e em duas etapas ou três etapas.
Nível 6	Docentes que têm mais de 400 alunos e atuam nos três turnos, em duas ou três escolas e em duas etapas ou três etapas.

Fonte: Inep.

Paralelamente, o Indicador de Regularidade Docente (IRD) mensura a capacidade de criação de vínculos com alunos e o desenvolvimento de projetos escolares de longo prazo: quanto menor o tempo de permanência médio dos professores em uma escola, menor é esta capacidade. O indicador valoriza docentes que lecionam no mesmo local por anos consecutivos e estavam presentes em anos mais recentes, por meio de uma escala de pontuação de 0 a 5. Neste trabalho, a variável foi registrada como a porcentagem de docentes com IRD maior ou igual a 3, o que equivale aos níveis alto e médio-alto de regularidade.

A RAIS serviu como fonte de dados para indicadores municipais relacionados à parcela de pessoas ocupadas nos municípios e outras características socioeconômicas, calculadas como proporção do total de vínculos ativos em 2017. É o caso da porcentagem de analfabetos, população feminina, emprego no setor agropecuário e população com renda de até um salário-mínimo. A variável relativa à parcela de vínculos de pessoas negras e pardas foi calculada diretamente dos microdados da RAIS, visto que as informações de raça não constam no sistema Dardo.

Por fim, o PIB *per capita* (a valores de 2017) e nível hierárquico urbano dos municípios têm como fonte, respectivamente, as Contas Nacionais e o estudo sobre as Regiões de Influência das Cidades, ambos fornecidos pelo IBGE. O Quadro 2 fornece uma breve definição sobre as variáveis utilizadas na monografia.

Quadro 2: Descrição das variáveis utilizadas

Variável	Descrição	Fonte
perc_esg	Percentual da população do município com ligação à rede geral de coleta de esgoto.	SNIS e PNSB
saeb	Nota padronizada do Saeb.	Inep
dist	Distorção idade-série: percentual de alunos com idade superior à recomendada para a série frequentada.	
ad	Adequação docente: percentual de professores com ensino superior na área que lecionam.	
atu	Média de alunos por turma.	
ed	Esforço docente: percentual de professores que se enquadram na categoria de sobrecarga de trabalho, segundo o Inep.	
horas	Média de horas de aula diárias.	
rd	Regularidade docente: percentual de professores com indicador alto e médio-alto.	
rd	Regularidade docente: percentual de professores com indicador alto e médio-alto.	
pib_pc	PIB per capital municipal (R\$), a valores de 2017.	IBGE
hier	Nível de hierarquia urbana do município.	
perc_vinc	Percentual de vínculos trabalhistas ativos em 31/12/2017 na população total do município.	RAIS
F	Percentual de mulheres no total de vínculos ativos.	
agro	Percentual de trabalhadores do setor agropecuário no total de vínculos ativos.	
analf	Percentual de pessoas analfabetas no total de vínculos ativos.	
renda	Percentual de pessoas com renda até 1 salário-mínimo no total de vínculos ativos.	
pard	Percentual de negros e pardos no total de vínculos ativos.	
dummy	<i>Dummy</i> igual a 1 para municípios sem cobertura de esgoto, e igual a 0 para locais onde o serviço existe.	-

Fonte: Elaboração própria a partir de SNIS, PNSB, IBGE, Inep e RAIS

Serão estimados dois modelos. Um deles terá a nota do Saeb municipal como variável dependente e o outro terá a taxa de distorção idade-série, de acordo com as equações (18) e (19) abaixo. Ambos possuem as mesmas variáveis explicativas: a cobertura de esgotamento, cuja relação com a educação configura a principal motivação do trabalho, e as variáveis de controle relativas à infraestrutura escolar e demografia dos municípios, agrupadas no vetor X_i .

$$saeb_i = \alpha + perc_esg_i + X_i\beta + \varepsilon_i \quad (18)$$

$$dist_i = \alpha + perc_esg_i + X_i\beta + \varepsilon_i \quad (19)$$

A Tabela 1 provê as estatísticas descritivas dos dados.

Tabela 1: Estatísticas descritivas

Variável	N	Média	Desvio padrão	Min	Pectl. 25	Pectl. 75	Max
perc_esg	4727	28.86	36.27	0	0	60.44	100
saeb	5462	5.07	0.61	2.62	4.65	5.49	7.24
dist	5570	12.24	9.2	0	4.7	18.5	57.6
ad	5570	63.52	22.25	0	48.7	80.88	100
atu	5570	19.77	3.82	4.3	17.5	22.4	33.9
horas	5570	4.46	0.74	3.6	4	4.5	9.9

rd	5570	49.47	25.16	0	30.7	66.7	100
ed	5570	4.9	5.33	0	0	7.2	50
pib_pc	5570	21990.76	20946.32	3285.04	9492.89	27051.62	344847.17
hier	5570						
... Capital Regional	189	3%					
... Centro de Zona	561	10%					
... Centro Local	4479	80%					
... Centro Subregional	164	3%					
... Metr�pole	177	3%					
perc_vinc	5570	13.99	9.58	0.16	7	18.21	128.61
F	5570	46.05	9.52	6.31	40.28	52.06	96.24
agro	5570	10.34	12.61	0	1.11	14.76	87.68
analf	5570	0.6	1.36	0	0.13	0.67	66.76
renda	5570	9.03	8.08	0	3.81	11.53	85.31
pard	5570	21.91	17.37	0	7.36	33.33	94.41
dummy	5570						
... 0	3206	58%					
... 1	2364	42%					

Fonte: Elabora  o pr pria a partir de SNIS, PNSB, IBGE, Inep e RAIS.

A seguir, ser  apresentada uma breve contextualiza  o sobre a oferta de servi os de saneamento e abastecimento nos munic pios brasileiros, bem como ser  realizada uma caracteriza  o espacial dos dados. A contextualiza  o contribuir  para uma melhor compreens o da problem tica em an lise e para a correta interpreta  o dos resultados que ser o obtidos por interm dio do exerc cio emp rico proposto.

2.2. Contextualiza  o e caracteriza  o espacial

At  o in cio dos anos 1970, os munic pios eram os principais respons veis pela provis o do saneamento b sico (CUNHA, 2011). Em 1971, a institui o do Plano Nacional de Saneamento B sico (Planasa) fortaleceu as companhias estaduais de saneamento (CESBs) ao garantir acesso a recursos federais – com intermedia  o do Banco Nacional de Habita  o (BNH) – para as empresas que prestassem os servi os regionalmente. O arranjo induziu um amplo movimento de terceiriza  o do saneamento para as CESBs, que atingiu 75% das cidades brasileiras. Al m disso, a lei possibilitava que essas empresas assinassem contratos de programa, isto  , sem ocorr ncia de processo licitat rio ou necessidade de atendimento de metas de cobertura ou efici ncia.

Embora tenha sido eficaz em ampliar a cobertura de esgotamento e, sobretudo, de abastecimento de água, o modelo centralizador do Planasa, caracterizado pela forte dependência de recursos da União, começou a apresentar limitações em virtude da deterioração financeira das CESBs (TUROLLA, 2002) e da extinção do BNH, em 1986 (SOUSA; COSTA, 2016). Paralelamente, o foco cedido à construção da infraestrutura, em detrimento da operação do saneamento, resultava na degradação do sistema e alto índice de perdas de água.

O Planasa foi extinto oficialmente em 1992 e as iniciativas governamentais nesta década voltadas ao setor de saneamento foram pontuais e desarticuladas (TUROLLA, 2002). Cabe citar o Programa de Saneamento para Populações em Áreas de Baixa Renda (Prosanear) e o Programa de Saneamento para Núcleos Urbanos (Pronurb), que objetivavam levar o serviço a favelas e assentamentos precários. A Lei de Concessão de Serviços Públicos⁶, aprovada em 1995, permitiu a privatização dos serviços de água e esgoto nos municípios não operados pelas CESBs, ainda que a transformação tenha sido moderada: em 2017, as empresas privadas estavam presentes em apenas 6% dos municípios (ABCON SINDCON⁷, 2017).

Em 2007, a Lei nº 11.445 estabeleceu novas diretrizes nacionais para o setor, com destaque para a necessidade de inclusão de metas vinculantes de cobertura e eficiência nos contratos e para a maior separação entre os papéis de execução e fiscalização do saneamento, visto que anteriormente a regulação podia ser feita pelas administrações locais ou CESBs (SANTOS, 2014). Também foi determinada a extinção dos contratos de programa considerados irregulares – isto é, aqueles com prazos vencidos ou não formalizados.

Contudo, a Lei previa o prosseguimento dos contratos de programa regulares e vigentes, sem ordenar sobre a possibilidade (ou não) de prorrogação (LOUREIRO, 2020). Ademais, em 2017 – 10 anos após a aprovação da Lei – apenas 30% das cidades brasileiras haviam cumprido a determinação legal de elaboração de um plano municipal de saneamento básico. A dificuldade de coordenação entre os diferentes entes é agravada pela difusão da competência legal do saneamento entre os 5.570 municípios, que, desde a Constituição de 1988, detêm a titularidade sobre o planejamento e fornecimento de saneamento básico (CUNHA, 2011), podendo executá-lo através de autarquias municipais ou delegá-los às CESBs e empresas privadas.

O arcabouço regulatório do setor continuou marcado pela imprevisibilidade e ausência de um esforço nacional em prol da universalização. Dados do SNIS evidenciam que, em 2017, o abastecimento de água por empresas de saneamento básico atingia 80,72% da população e a

⁶ BRASIL. Lei n.º 8.987, de 13 de fevereiro de 1995. Disponível em http://www.planalto.gov.br/ccivil_03/leis/18987cons.htm

⁷ Associação e Sindicato Nacional das Concessionárias Privadas de Serviços Públicos de Água e Esgoto.

coleta de esgoto alcançava pouco mais da metade do país (Tabela 2). Apesar do cenário mais otimista para o abastecimento de água, estima-se quase 40 milhões de brasileiros marginalizados do serviço.

Tabela 2: População atendida com abastecimento de água e coleta de esgoto (Brasil, 2017)

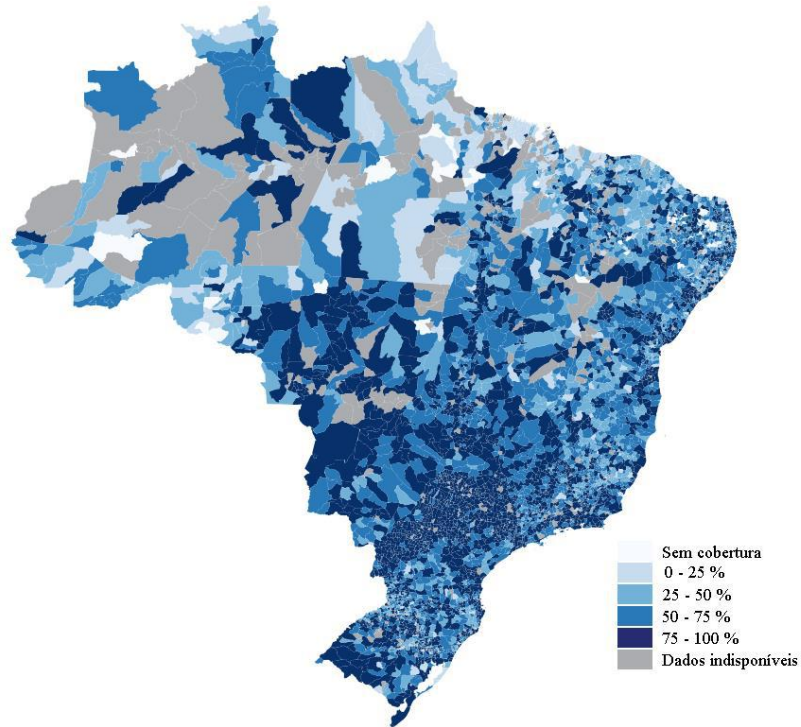
Região	População atendida com abastecimento de água (%)	População atendida com coleta de esgoto (%)
Centro-Oeste	88,07	52,70
Nordeste	69,81	25,25
Norte	48,62	7,87
Sudeste	90,56	77,61
Sul	88,42	42,99
Total Geral	80,72	50,30

Fonte: Estimativas próprias, com base em dados do SNIS e PNSB.

Os dados da Tabela 2 realçam a desigualdade regional na provisão de saneamento básico. Ao passo que as regiões Sudeste e Centro-Oeste apresentaram os melhores níveis de atendimento dos serviços, na região Norte somente 48,62% da população possuía ligação à rede de abastecimento de água, e 7,87% ao esgotamento. O Nordeste também se situava abaixo da média nacional em ambos os serviços, e o Sul detinha a segunda maior abrangência do abastecimento de água, mas estava abaixo do *benchmark* nacional de cobertura de esgoto.

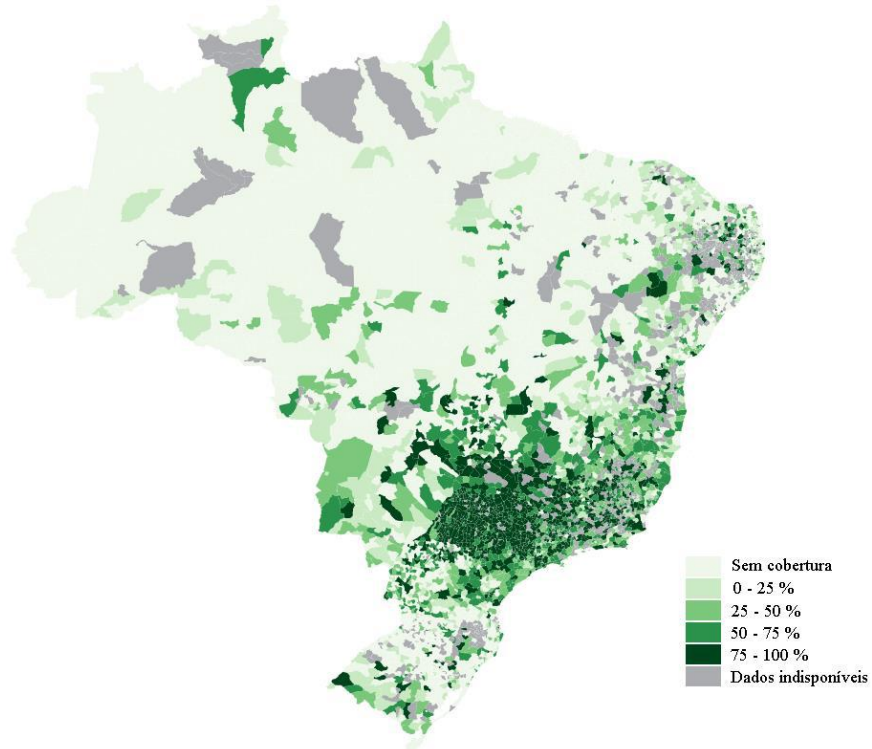
Os mapas das Figuras 1 e 2 permitem uma melhor visualização da distribuição do saneamento do território. No caso da coleta de esgoto, há alta concentração dos serviços no Centro-Sul do país, sobretudo no estado de São Paulo, enquanto porção considerável do Norte e Nordeste sequer apresenta infraestrutura de esgotamento.

Figura 1: População atendida com abastecimento de água (Brasil, 2017)



Fonte: Elaboração própria, com base em dados do SNIS e PNSB.

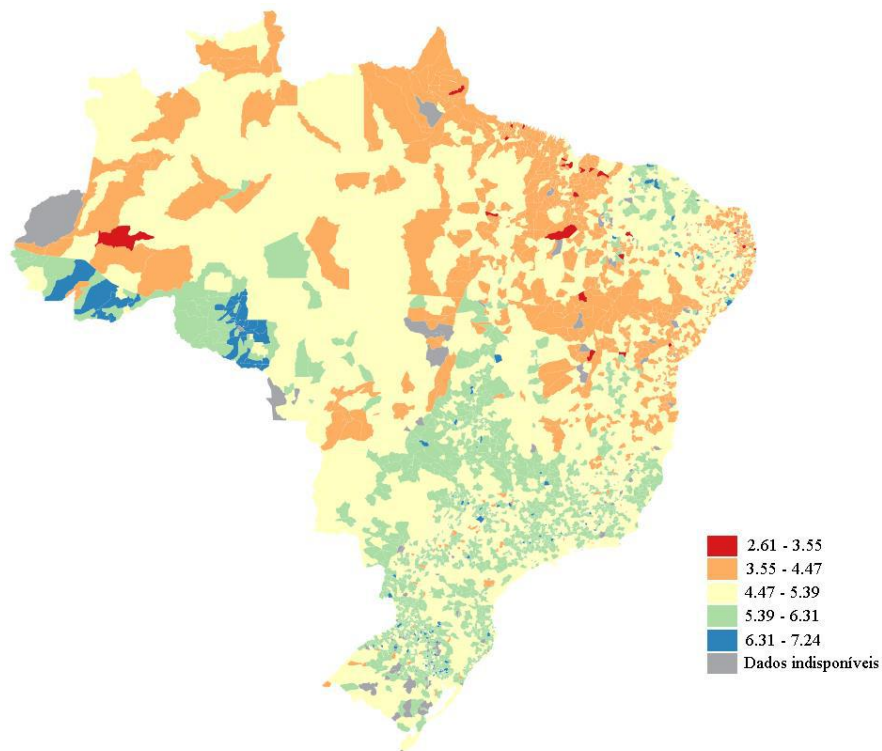
Figura 2: População atendida com coleta de esgoto (Brasil, 2017)



Fonte: Elaboração própria, com base em dados do SNIS e PNSB.

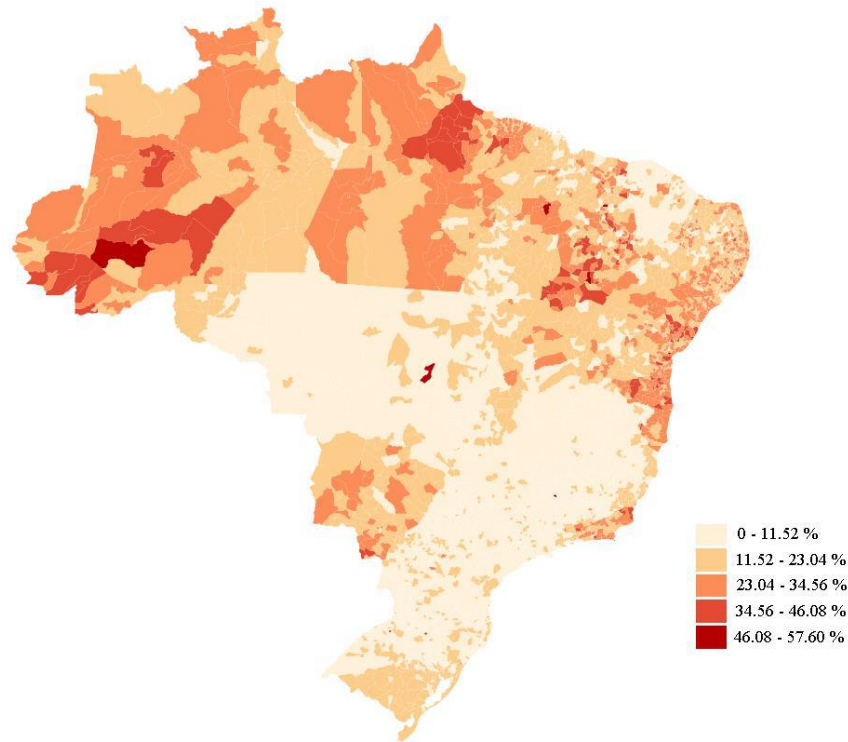
O Capítulo 1 desenvolveu uma revisão teórica e empírica a respeito dos mecanismos de interação entre saneamento básico e aproveitamento escolar. Nesse contexto, ao visualizar a desagregação dos principais indicadores educacionais no território (Figuras 3 e 4), identificam-se padrões similares aos do saneamento: em média, as maiores notas do Saeb, assim como os menores índices de defasagem idade-série, ocorrem no Sul, Sudeste e parte da região Centro-Oeste – ainda que com elevado grau de heterogeneidade entre os estados e municípios.

Figura 3: Nota padronizada do Saeb, por município (Brasil, 2017)



Fonte: Elaboração própria, com base em dados do Inep.

Figura 4: Distorção idade-série, por município (Brasil, 2017)



Fonte: Elaboração própria, com base em dados do Inep.

A distribuição desigual dos indicadores apresentados configura a principal motivação para o uso de ferramentas da econometria espacial na análise dos dados e estimação dos modelos de regressão deste trabalho. Atualmente, é amplamente aceito que a simples visualização de mapas não é um método suficientemente rigoroso para identificar padrões no espaço e a existência de *clusters* (LESAGE, 2014). Ademais, a autocorrelação espacial, quando não aferida propriamente, pode enviesar estimativas e reduzir o poder de explicação de modelos econométricos.

Dessa forma, cabe indagar: seria o espaço determinante para explicar as desigualdades regionais entre indicadores educacionais e de saneamento básico? Mais além: estaria o desempenho escolar em uma cidade condicionado não apenas ao contexto local, mas também à situação da educação e saneamento em municípios vizinhos?

Não se objetiva, aqui, reduzir o problema educacional brasileiro à falta de saneamento básico. O subaproveitamento escolar e a precariedade da infraestrutura de serviços públicos estão ligados ao processo histórico do desenvolvimento econômico do país e agem simultaneamente como causa e consequência de uma série de problemas sociais, como a pobreza e a desigualdade. Contudo, tendo em vista a escassez da literatura empírica sobre o

tema no Brasil, é pertinente investigar a relação particular entre ambos os fatores à luz das ferramentas da econometria espacial.

A próxima seção apresenta a metodologia empregada para a análise exploratória de dados espaciais (AEDE) e para a estimação de modelos econométricos que buscam responder tais questionamentos.

2.3. Metodologia

Um aspecto importante de estudos que envolvem unidades espaciais configura as potenciais relações entre as observações, decorrentes da interação de fatores econômicos, sociais e ecológicos que transpassam fronteiras políticas. Nesse contexto, a autocorrelação espacial é definida como a correlação de uma variável em diferentes espaços geográficos, isto é, a dependência entre o valor de uma observação e os valores de observações vizinhas (LESAGE; PACE, 2009).

Do ponto de vista estatístico, uma série de instrumentos – como modelos de regressão linear – são baseados em hipóteses de independência entre as variáveis, de modo que a análise da autocorrelação espacial permite identificar a violação desse pressuposto e quantificar a estrutura do fenômeno espacial estudado. O primeiro passo para mensurar a dependência espacial consiste na imposição de um arranjo para a ocorrência das interações espaciais, por meio da especificação de uma matriz de pesos espaciais (W), ou matriz de vizinhança (ALMEIDA, 2012).

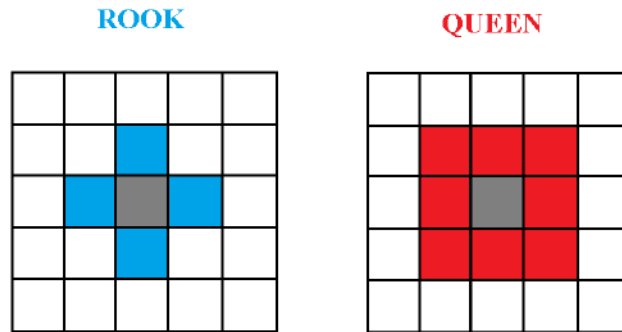
Assumindo a existência de n objetos espaciais (cidades, estados, países ou outra forma de subdivisão geográfica), W é uma matriz quadrada de dimensão $n \times n$ que define quais observações serão classificadas como vizinhas (SARRIAS, 2020). Cada elemento (i, j) de W expressa a intensidade do relacionamento do par i, j , como retratado abaixo:

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \quad (20)$$

Por convenção, os elementos da diagonal principal de W são iguais a zero, visto que, de maneira geral, regiões não são registradas como vizinhas de si mesmas (ALMEIDA, 2012). Há diversas formas de especificar os pesos da matriz de vizinhança. Uma das mais comuns configura a contiguidade direta das observações: apenas locais imediatamente próximos são

classificados como vizinhos, de forma que w_{ij} é igual a 1, caso i e j sejam contíguos, e igual a zero, caso não sejam. Em alusão às peças do jogo de xadrez, a convenção de contiguidade é do tipo *rook* (torre) se considerar apenas fronteiras com extensão diferente de zero ou é do tipo *queen* (rainha) se também incluir os vértices contíguos. A Figura 5 retrata ambas as convenções.

Figura 5: Convenções de contiguidade mais comuns: *rook* (esquerda) e *queen* (direita)



Fonte: Elaboração própria, com base em Almeida (2012).

O critério de distância geográfica também é comumente utilizado para a elaboração de matrizes de vizinhança. Sua intuição é que regiões mais próximas sofrerão interações mais intensas que locais distantes geograficamente. No caso da matriz de k vizinhos mais próximos:

$$w_{ij}(k) = \begin{cases} 1, & \text{se } d_{ij} \leq d_i(k) \\ 0, & \text{se } d_{ij} > d_i(k) \end{cases} \quad (21)$$

Em que d_{ij} é a distância entre as regiões i e j , e $d_i(k)$ é a distância mínima para que a observação i tenha exatamente k vizinhos. Há uma série de outras abordagens para a especificação de W , como a imposição de distâncias fixas, distância inversa e matrizes de pesos espaciais endógenos, relacionadas ao fenômeno estudado⁸.

A vantagem das matrizes de contiguidade ou baseadas na distância geográfica é a manutenção de uma estrutura de pesos espaciais exógena ao fenômeno estudado, hipótese assumida em grande parte das aplicações econométricas (ANSELIN, 2013). O uso de matrizes simples gera uma estrutura de pesos esparsa, com grande quantidade de zeros, permitindo distinguir entre efeitos de transbordamento locais e globais (LESAGE, 2014), que serão vistos adiante. Ademais, estudos demonstraram que matrizes com alto grau de conectividade reduzem

⁸ Uma discussão mais aprofundada sobre os tipos de matrizes de vizinhança foge do escopo do presente trabalho, mas pode ser encontrada em Almeida (2012).

a probabilidade de especificação correta do modelo e induzem um viés de subestimação dos parâmetros (SMITH, 2008).

O próximo passo reside na mensuração da autocorrelação espacial. A ferramenta mais comumente usada para este fim é o Índice de Moran, que assume a forma global, local ou bivariada. O Índice de Moran global mede o nível de autocorrelação geral nos dados e é dado por:

$$I = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^n (x_i - \bar{x})^2 / n} = \frac{n \sum_i \sum_j w_{ij} z_i z_j}{S_0 \sum_{i=1}^n z_i^2} \quad (22)$$

Em que n é o número de regiões, z representa os valores da variável de interesse x , padronizada, e S_0 é igual à operação $\sum_i \sum_j w_{ij} z_i z_j$ (ALMEIDA, 2012). Reescrevendo I na forma matricial e padronizando W pela linha:

$$I = \frac{z' W z}{z' z} \quad (23)$$

No teste de hipótese, o Índice de Moran é padronizado, variando entre -1 e 1, e testa a hipótese nula de aleatoriedade espacial: a estatística estimada é comparada com a distribuição obtida em sucessivas permutações dos dados. Valores de I que excedem o valor esperado de $-[1/(n-1)]$ são positivos e – desde que significativos estatisticamente – indicam a autocorrelação espacial positiva. Isto é, altos (baixos) valores da variável de interesse tendem a estar circundados por altos (baixos) valores dessa variável em regiões vizinhas. Já nas situações em que I é menor que o valor esperado, a estimativa é negativa, e aponta para maior grau de dispersão da variável. Quanto mais próximo de 1 (em módulo), mais forte é a autocorrelação.

O Índice de Moran local, por sua vez, permite analisar as relações de dependência espacial para cada observação. É útil na medida em que porções do território podem apresentar padrões de autocorrelação que diferem da estatística global ou que sequer apresentam autocorrelação espacial. Dessa forma, o I de Moran local atende aos critérios de um indicador LISA (*Local Indicator of Spatial Association*): (i) capacidade de identificação de *clusters* significativos estatisticamente, e (ii) o somatório dos indicadores locais é proporcional ao indicador global correspondente (ANSELIN, 1996). Sua fórmula é dada por:

$$I_i = z_i \sum_j w_{ij} z_j \quad (24)$$

Na qual o somatório dos pesos abrange apenas observações vizinhas de i . O indicador pode ser visualizado na forma de um mapa, que agrupa as observações significativas estatisticamente em quatro grupos: Alto-Alto, Baixo-Baixo, Alto-Baixo e Baixo-Alto.

O diagrama de dispersão de Moran configura outra ferramenta para visualização dos padrões de autocorrelação. Nele, o eixo X mede o valor das observações de cada região e o eixo Y representa o valor médio da variável para seus vizinhos, permitindo a identificação da estatística global (a inclinação da reta gerada pelas observações) e o padrão de associação espacial local (os quadrantes do diagrama de dispersão) (ANSELIN, 1996).

O Índice de Moran pode ser utilizado, ainda, em sua forma bivariada, que quantifica o grau de associação linear (positiva ou negativa) entre o valor de uma variável em uma dada região e a média de outra variável em regiões vizinhas. O indicador bivariado global na forma matricial e normalizado na linha é dado como na fórmula (25) abaixo. Este índice também assume uma forma local, representado em (26):

$$I^{z_1 z_2} = \frac{z_1' W z_2}{z_1' z_1} \quad (25)$$

$$I_i^{z_1 z_2} = z_{1i} W z_{2i} \quad (26)$$

Nas quais z_1 e z_2 são duas variáveis padronizadas. Na monografia, esse indicador é utilizado para avaliar a associação entre educação e saneamento básico.

Apesar de o instrumental teórico da econometria espacial enfatizar a importância da seleção da matriz de vizinhança, há pouca informação sobre os critérios ideais para essa escolha (HARRIS et al., 2011), além de divergência entre autores (KOPCZEWSKA, 2021). Burridge (2015) ressalta que para alguns pesquisadores, esta escolha deve feita *a priori*, com base no conhecimento teórico do objeto de estudo. Outros defendem a estimação do modelo econométrico com diferentes tipos de estruturas de pesos espaciais e seleção da matriz que gera o modelo com melhor ajuste de acordo com critério de informação Akaike (STAKHOVYCH, 2009; DUNCAN, 2017). Há também métodos que misturam estratégias de otimização dinâmica (KOPCZEWSKA, 2021).

Cabe ressaltar que mesmo a importância desta etapa gera divergências. Autores como Kostov (2010) e Corrado e Fingleton (2012) consideram que as estimativas dos modelos de regressão espacial são sensíveis à má especificação de W . Por outro lado, Stakhovych (2009), Lesage e Pace (2014) e Rahal (2017) defendem que, *para modelos de regressão bem especificados*, a escolha da matriz de pesos espaciais não altera significativamente os

coeficientes encontrados – embora possa afetar em maior grau a estatística de autocorrelação espacial.

A monografia adota a abordagem de Almeida (2012), que sugere selecionar, dentre uma série de matrizes de contiguidade e k vizinhos mais próximos, aquela que gera o maior valor do Índice de Moran estatisticamente significativo para os resíduos dos modelos de regressão por mínimos quadrados ordinários.

Desta forma, o Índice de Moran pode ser utilizado tanto para a especificação adequada da matriz de vizinhança quanto para a análise exploratória espacial das variáveis de interesse. Ademais, quando o I de Moran dos resíduos de um modelo MQO denota significância estatística, há indicação para estimação por meio de um modelo de regressão espacial (ALMEIDA, 2012), visto que os dados violam pressupostos básicos do MQO de aleatoriedade da amostra e de normalidade do termo do erro, para testes de hipótese.

Os modelos de regressão espacial se diferenciam de acordo com o componente em que há defasagem espacial: essa pode estar presente na variável dependente, por meio da multiplicação de um fator autorregressivo espacial ρ e da matriz de ponderação espacial W ($\rho W y$); nas variáveis explicativas ($XW\theta$); ou no termo do erro ($\lambda W u$); ou ainda em uma combinação dessas possibilidades. ρ , θ e λ variam entre -1 e 1, sendo que sinais positivos indicam autocorrelação positiva e sinais negativos apontam dispersão dos dados. O modelo mais geral de todos – o modelo de Manski, ou modelo de dependência espacial generalizada (GNS) – possui os três tipos de defasagem e é retratado em (27):

$$y = \rho W y + X\beta + WX\theta + u, \quad u = \lambda W u + \varepsilon \quad (27)$$

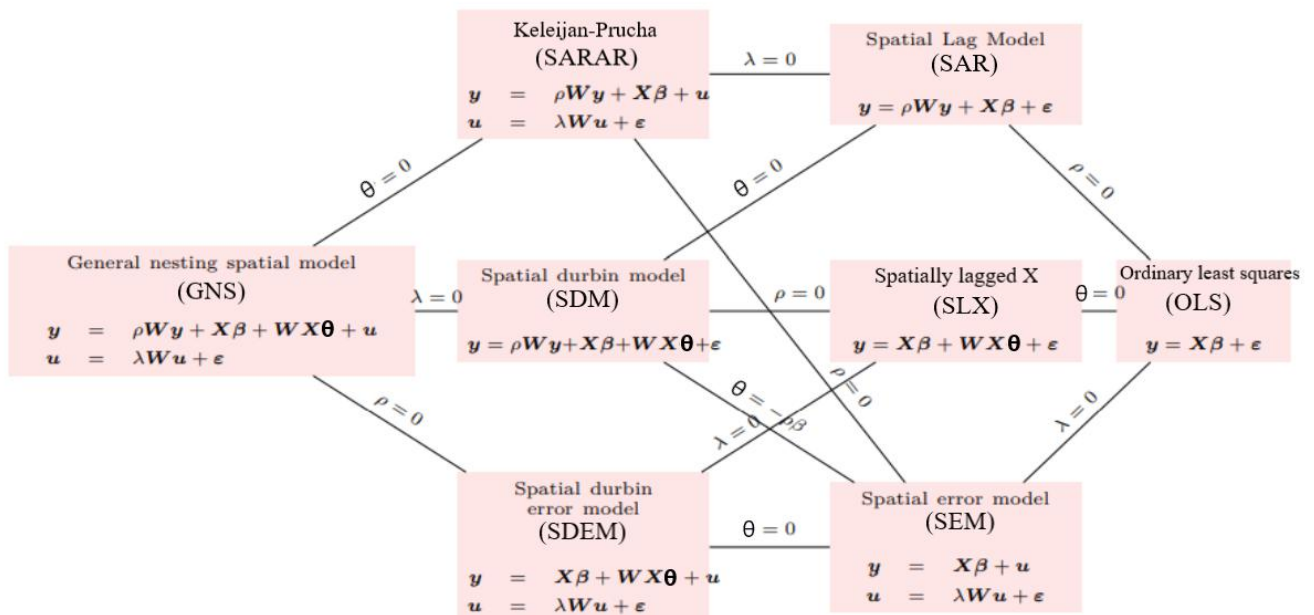
Em (5), o termo $X\beta$ representa o mesmo que em um modelo de regressão linear usual: como uma alteração na variável explicativa em uma dada região se associa com mudanças na variável dependente desta região, isolando-se o efeito das demais variáveis. Já $WX\theta$ parte do princípio de que as observações de X são espacialmente correlacionadas, de forma que a variável dependente de um local é afetada por distúrbios das variáveis explicativas de locais vizinhos. O termo $\lambda W u$ é o componente autorregressivo do erro u e permite identificar como fatores não observados de regiões próximas se relacionam com a variável de interesse y . Finalmente, $\rho W y$ caracteriza a associação da variável dependente em uma região com a média dos valores da variável dependente em regiões vizinhas.

Tanto $WX\theta$ quanto $\lambda W u$ modelam efeitos *spillover* (de transbordamento) locais, isto é, y é afetado apenas por distúrbios em X ou u provenientes de observações vizinhas nos critérios

determinados pela matriz de vizinhança. Já o termo $\rho W y$ caracteriza um efeito *feedback* global ou endógeno, visto que mudanças em y dariam início a uma reação em cadeia para (potencialmente) toda a amostra (LESAGE, 2014).

O GNS é raramente usado para estudos empíricos porque não é possível separar de maneira apropriada os efeitos resultantes de cada tipo de defasagem (MANSKI, 1993). A solução é a imposição de restrições nos parâmetros ρ, θ e λ , dando origem às principais especificações de modelos de regressão espacial *cross-section*, como retratado na Figura 6.

Figura 6: Modelos de regressão espacial



Fonte: Sarrias (2020), modificado.

A escolha do modelo de regressão, assim como a seleção da matriz de pesos espaciais, ainda é um tema em discussão na literatura. Uma opção é fazer uso dos testes multiplicadores de Lagrange (LM), desenvolvidos por Anselin (1988; 1996), com o objetivo de identificar a origem autocorrelação espacial do modelo. O teste LM-erro tradicional assume $\rho = 0$ e testa a hipótese nula $H_0: \lambda = 0$, de modo que se esta for rejeitada, há indicativos de defasagem no termo do erro, e o modelo SEM é o mais adequado. Por sua vez, o teste LM-lag assume $\lambda = 0$ e testa $H_0: \rho = 0$; se rejeitado, deve-se escolher o modelo SAR, pois haveria defasagem na variável dependente.

Se ambos rejeitarem H_0 , é preciso aplicar as versões robustas desses testes, que não assumem inicialmente $\rho = 0$ e $\lambda = 0$. Portanto, o LM-erro robusto testa a presença de defasagem

no erro quando há defasagem na variável dependente e o LM-lag faz o inverso. Caso os dois testes rejeitem novamente a hipótese nula, Anselin (2017) recomenda selecionar aquele que produziu o maior valor de teste significativo estatisticamente ou considerar outros tipos de especificação, como a adoção múltiplas matrizes de vizinhança. Há ainda o teste SARMA, que tem como hipótese nula $\rho = \lambda = 0$; ou seja, identifica a presença dos dois tipos de defasagem.

Elhorst (2010) também faz uso dos testes multiplicadores de Lagrange, mas adota outra estratégia: se houver indicação da presença de *lag* na variável dependente, no termo de erro ou em ambos, deve-se estimar o modelo SDM, em seguida, realizar testes de razão de verossimilhança a fim de verificar se o modelo deve ser restrito para um SEM ou SAR. Os testes de razão de verossimilhança em questão terão como hipóteses nulas $H_0: \theta = 0$ e $H_0: \theta = -\rho\beta$. Se a primeira não for rejeitada, o SAR deve ser escolhido (desde que o LM-lag também dê essa indicação); se a segunda não for rejeitada, o SEM configura o modelo mais adequado; e se os testes recusarem ambas as hipóteses, estima-se o SDM.

Lesage (2014) defende que os principais modelos a serem considerados são o SDEM (no caso de embasamento teórico para efeitos *spillover* locais) e o SDM (quando a teoria dá suporte a efeitos de transbordamento globais). Ademais, transbordamentos globais seriam mais raros empiricamente, além de difíceis de estimar de maneira adequada⁹, embora sejam os mais frequentemente modelados pelos pesquisadores de estudos regionais. Similarmente, Cook et al. (2015) apresentam argumentos em favor da utilização do SDEM e do SARAR.

a monografia opta por uma abordagem mais abrangente, porém consistente com as possibilidades de especificação definidas no arcabouço teórico de econometria espacial. Serão estimados os principais modelos de regressão espacial *cross-section* e a escolha se dará de acordo com aquele que obtiver o melhor ajuste de acordo com a estatística de *log-verossimilhança* e os critérios de informação Akaike (AIC) e bayesiano (BIC). Em seguida, a necessidade de restrição do modelo será testada pela razão de verossimilhança. Por fim, aplica-se o Índice de Moran sobre os resíduos para verificar se a autocorrelação espacial foi devidamente capturada pelo modelo.

Os critérios de informação Akaike e BIC podem ser definidos como as fórmulas (28) e (29), respectivamente:

⁹ Não obstante, para o autor, nos casos em que a teoria por trás dos modelos não é tão clara, a escolha entre efeitos globais e locais pode ser feita por meio de uma comparação *bayesiana* das probabilidades em favor de cada especificação. Silva et al. (2016) configura um exemplo de utilização desse método para modelar o crescimento populacional das cidades brasileiras entre 1970 e 2010.

$$AIC = 2k - 2 \ln(L') \quad (28)$$

$$BIC = k \ln(n) - 2 \ln(L') \quad (29)$$

Em que k é a quantidade de parâmetros estimados, n é o tamanho da amostra e L' é o valor máximo da função de *log-verossimilhança*. Os critérios de informação permitem escolher, entre um conjunto de modelos regressão, o que possui o melhor ajuste *relativo*. Ambas as estatísticas premiam a qualidade do ajuste (via função de verossimilhança) e penalizam a inclusão de parâmetros estimados, de modo que o modelo com melhor ajuste terá o menor valor para o AIC e BIC e o maior para L' .

O teste de razão de verossimilhança é utilizado para modelos aninhados, ou seja, quando a imposição de restrições aos parâmetros de um dos modelos origina o outro. A estatística assume que a *log-verossimilhança* do modelo não restrito é maior que a do restrito e testa a significância da diferença entre ambas. Se for significativa, o modelo mais simples deve ser descartado; caso contrário, o modelo ilimitado não agrega valor, então não deve ser utilizado (KOPCZEWSKA, 2021). O teste é definido da seguinte maneira:

$$RV = -2 \left(\ln \frac{L'_{\text{modelo restrito}}}{L'_{\text{modelo irrestrito}}} \right) \quad (30)$$

Estabelecidos os procedimentos metodológicos, a próxima seção expõe e interpreta os resultados obtidos. A AEDE investigará a presença de autocorrelação no esgotamento, notas do Saeb e taxa de distorção idade-série, além da existência de *clusters* abaixo ou acima da média nacional para essas variáveis de interesse. Em seguida, os modelos econométricos buscam quantificar a associação entre a ampliação da cobertura de saneamento e os indicadores educacionais, a fim de testar tanto a hipótese de que essa relação existe no âmbito local, como também de que gera transbordamentos espaciais significativos.

3. RESULTADOS

3.1. Escolha da matriz de pesos espaciais

A escolha pela matriz de pesos espaciais seguiu os critérios sugeridos por Almeida (2012), a saber:

1. Estimação do modelo clássico de regressão linear;
2. Teste do índice de Moran global para os resíduos por meio de várias especificações de matrizes espaciais;
3. Seleção da matriz que gera o maior valor do Índice de Moran significativo estatisticamente.

Dessa forma, para os resíduos dos dois modelos a serem estimados (SAEB e distorção idade-série), o Índice de Moran foi aplicado com base em 22 matrizes de vizinhança, a fim de selecionar aquela que melhor explica o padrão de defasagem espacial nos dados: contiguidade do tipo *queen*, *rook* e uma série de matrizes de *k* vizinhos mais próximos, nas quais *k* vai de 1 a 20. A Tabela 3 detalha os resultados.

Tabela 3: Índices de Moran dos resíduos para seleção da matriz de vizinhança

Tipo de contiguidade	Modelo 1 - Saeb	Modelo 2 – distorção idade-série	p-valor
queen	0,386	0,481	<0.001
rook	0,387	0,480	<0.001
k1	0,289	0,369	<0.001
k2	0,292	0,366	<0.001
k3	0,287	0,356	<0.001
k4	0,283	0,350	<0.001
k5	0,278	0,353	<0.001
k6	0,275	0,352	<0.001
k7	0,275	0,353	<0.001
k8	0,274	0,352	<0.001
k9	0,275	0,352	<0.001
k10	0,276	0,352	<0.001
k15	0,279	0,353	<0.001
k20	0,276	0,353	<0.001

Fonte: Elaboração própria, a partir de dados do Inep e IBGE.

Para todas as matrizes, os índices de Moran globais são positivos e significativos, indicando a presença de dependência espacial positiva. Rejeita-se a hipótese nula de distribuição aleatória dos resíduos, a uma significância de 0,1%. Tendo em vista que as matrizes

de contiguidades do tipo *rook* e *queen* apresentaram o maior valor do Índice de Moran para os modelos do Saeb e da distorção idade-série, respectivamente, essas serão utilizadas para a estimação dos resultados econométricos do trabalho. Os resíduos do modelo da distorção idade-série, com I de Moran igual a 0,481 (em uma escala que vai de -1 a 1), aparentam possuir uma autocorrelação espacial mais forte (dados mais concentrados no território) que a nota padronizada do Saeb, cuja estatística é de 0,387.

3.2. Análise exploratória de dados espaciais

Além de sua aplicação para escolha da matriz de pesos espaciais, o índice de Moran configura ferramenta útil para a análise da distribuição de variáveis específicas no território, bem como a identificação de *clusters* de valores abaixo ou acima da média amostral. Esta subseção explora a autocorrelação espacial nas três principais variáveis de interesse do trabalho: as notas do Saeb, defasagem idade-série e a cobertura de esgotamento.

A Tabela 4 apresenta os índices de Moran globais para os indicadores educacionais. Os valores são positivos e significativos estatisticamente, isto é, há indícios de que municípios com indicadores de qualidade do ensino acima da média são vizinhos de municípios com indicadores também favoráveis. Analogamente, o contrário – proximidade geográfica entre cidades com indicadores abaixo da média – também ocorre.

Tabela 4: Índices de Moran globais: indicadores educacionais

I de Moran - Saeb	I de Moran - Distorção idade-série	p-valor
0,581	0,727	<0.001

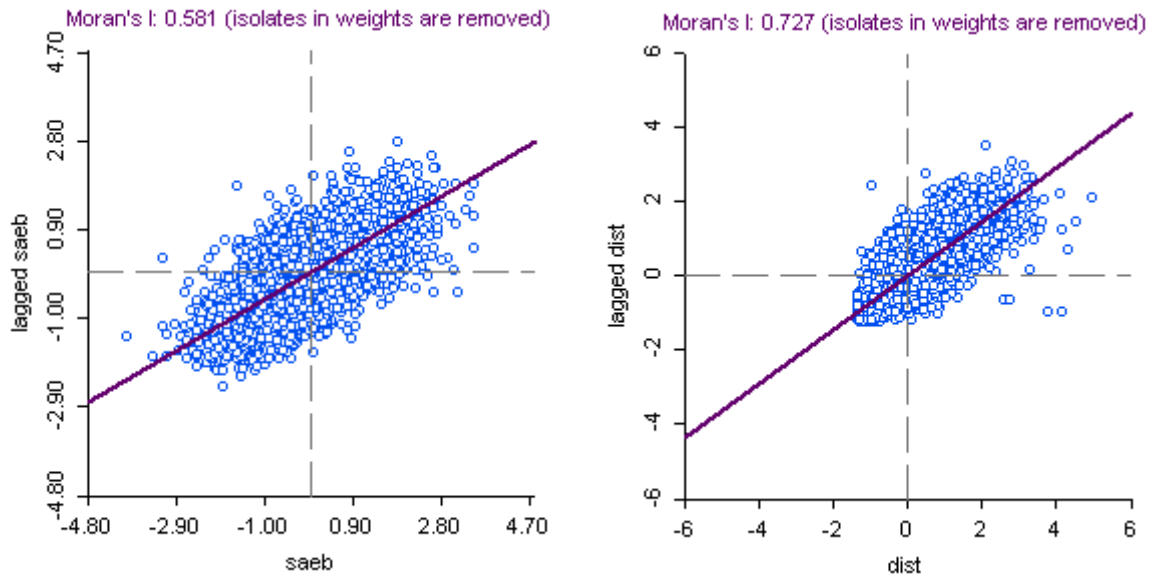
Fonte: Elaboração própria.

O índice de Moran global fornece uma perspectiva geral da dependência espacial nos dados. Contudo, em todo o território, é possível encontrar conjuntos de observações que vão contra a tendência de autocorrelação positiva, ou nas quais a dependência espacial sequer está presente. A Figura 7 exhibe os diagramas de dispersão de Moran para as variáveis educacionais.

O quadrante superior direito (Alto-Alto) exhibe observações acima da média que estão cercadas por outras acima da média, enquanto o quadrante inferior esquerdo (Baixo-Baixo) mostra locais que formam *clusters* de indicadores abaixo da média. Ambos os casos refletem uma dependência espacial positiva. Já os quadrantes superior esquerdo e inferior direito

indicam tendências de autocorrelação espacial negativa: cidades com valores abaixo do *benchmark* nacional e que se avizinham de outros acima do padrão (Baixo-Alto), ou vice-versa (Alto-Baixo).

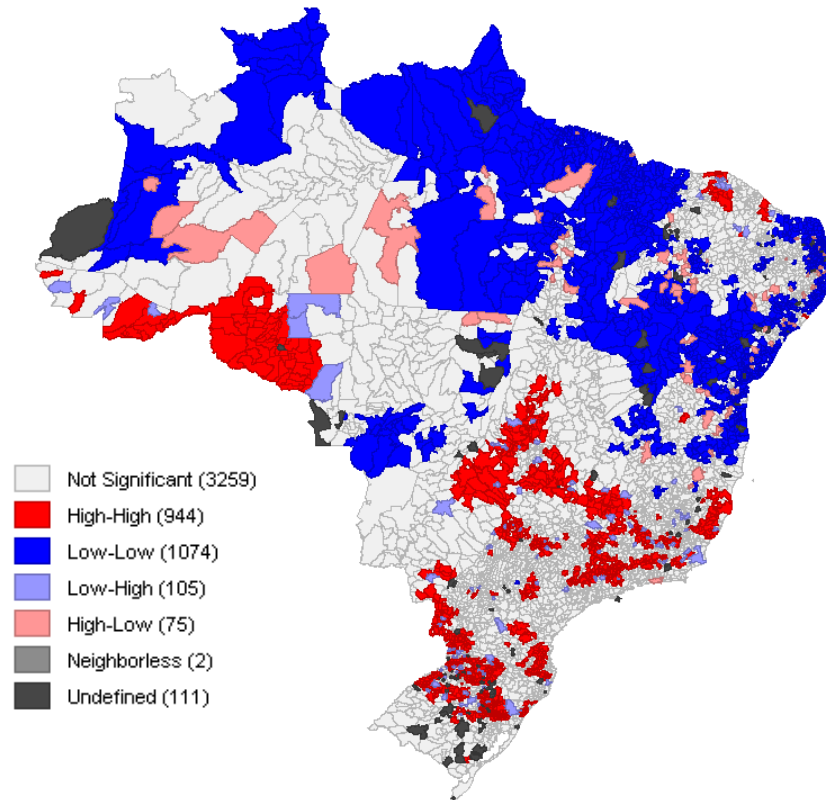
Figura 7: Diagramas de dispersão de Moran – Saeb (esquerda) e Distorção idade-série (direita)



Fonte: Elaboração própria, a partir de dados do Inep e IBGE.

O índice de Moran local foi utilizado para facilitar a visualização dos agrupamentos no território e sua significância estatística. No caso das notas do Saeb, a autocorrelação espacial abrange 2.198 municípios, a uma significância de 5% (Figura 8). *Clusters* Alto-Alto podem ser observados principalmente em cidades do Centro-Sul, Ceará, Rondônia e Acre, embora entre os locais com notas acima da média também se identifiquem cidades destoantes do padrão (*clusters* Baixo-Alto). Já os *clusters* Baixo-Baixo estão presentes no Mato Grosso e em maior proporção no Norte e Nordeste, onde também são mais comuns os agrupamentos Alto-Baixo – uma das exceções é o município do Rio de Janeiro (RJ).

Figura 8: Índice de Moran local, notas do Saeb (Brasil, 2017)

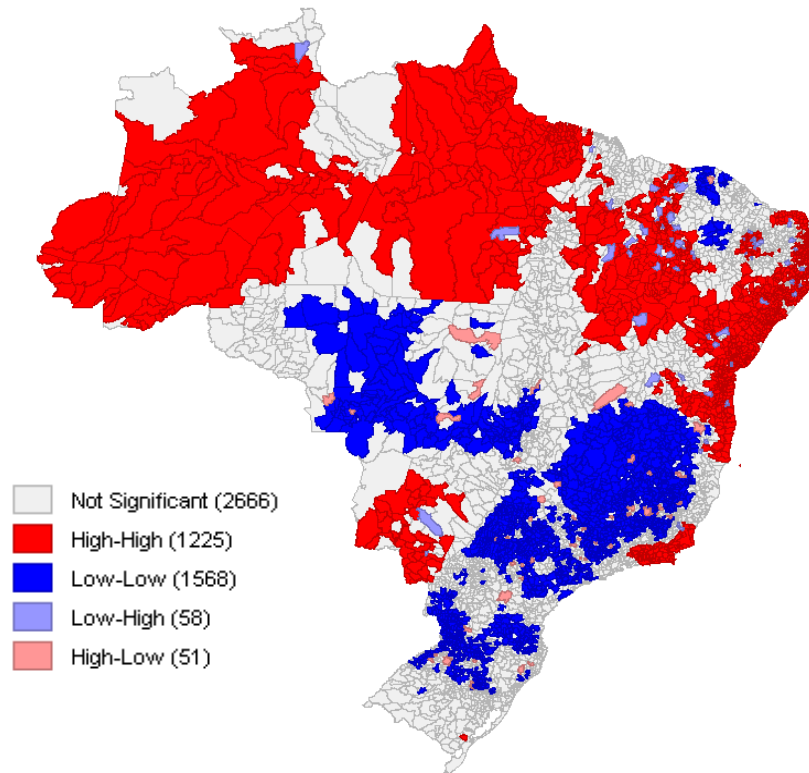


Fonte: Elaboração própria, a partir de dados do Inep e IBGE.

Para a distorção idade-série (Figura 9), *clusters* Baixo-Baixo são observados principalmente em Mato Grosso, Goiás, São Paulo, Minas Gerais, Paraná, Santa Catarina e Ceará¹⁰, enquanto agrupamentos com valores acima da média estão presentes em parte considerável do Nordeste e Norte, além de no Mato Grosso do Sul e em quase a totalidade do Rio de Janeiro. Agrupamentos do tipo Baixo-Alto ou Alto-Baixo são menos comuns, contabilizando 109 municípios. No total, 2.904 municípios apresentaram algum tipo de dependência espacial para a desfasagem idade-série.

¹⁰ Os resultados positivos para o Ceará, em contraposição ao restante do Nordeste, podem ser reflexo das políticas educacionais implementadas no estado a partir dos anos 2000. Uma análise aprofundada sobre este tema está presente em Verdiano (2019).

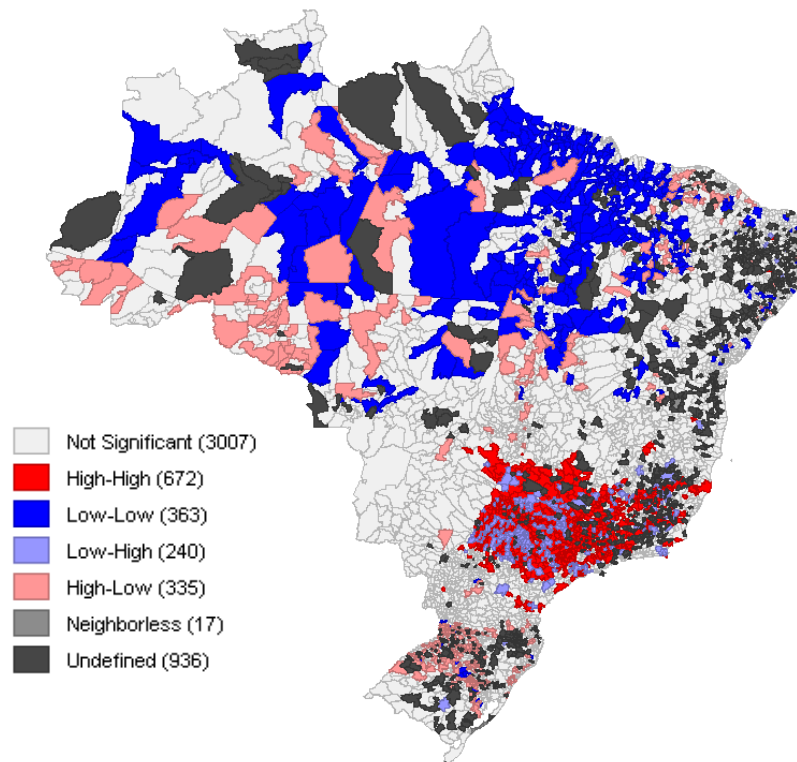
Figura 9: Índice de Moran local, distorção idade-série (Brasil, 2017)



Fonte: Elaboração própria, a partir de dados do Inep e IBGE.

Por fim, cabe analisar como ambas as variáveis se relacionam com a cobertura de esgoto no território. Em consonância com o esperado, o Índice de Moran bivariado global encontrou uma autocorrelação espacial positiva (0.188) entre o Saeb e a coleta de esgoto, sendo que a estatística local foi significativa em 1.610 municípios (Figura 10). Há agrupamentos Baixo-Baixo concentrados na região Norte e agrupamentos Alto-Alto principalmente no Sudeste, onde também estão a maior parte dos municípios do padrão Baixo-Alto. Já *clusters* Alto-Baixo são mais comuns no Norte, Mato Grosso e no Rio Grande do Sul.

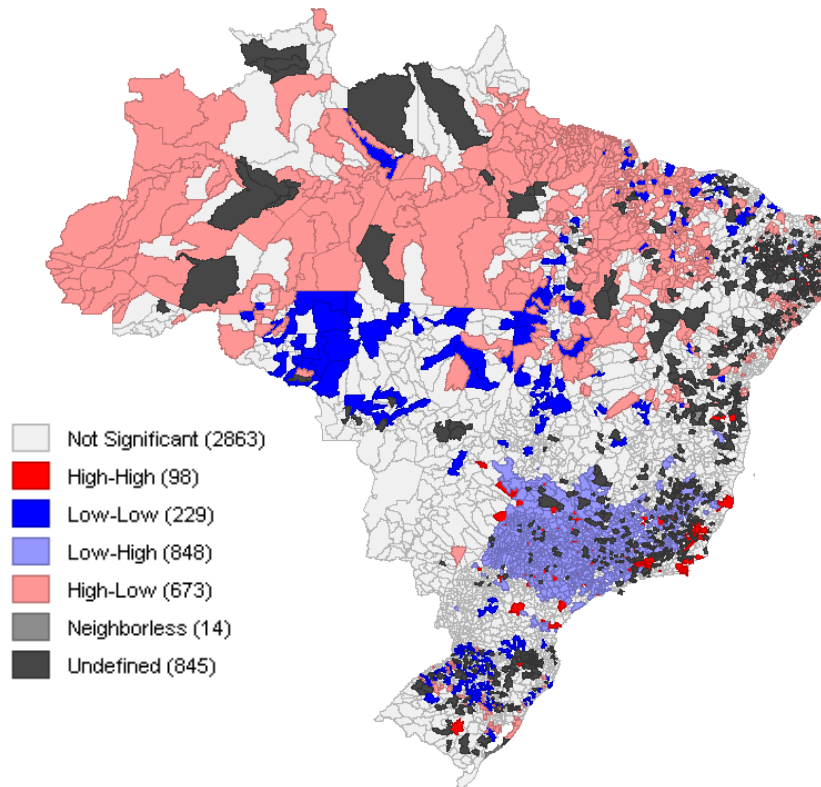
Figura 10: Índice de Moran local bivariado: Saeb (X) e coleta de esgoto (Y) (Brasil, 2017)



Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS e PNSB.

Paralelamente, o Índice de Moran bivariado indica uma autocorrelação espacial negativa (-0.273) entre a defasagem idade-série e a coleta de esgoto, e a estatística local identificou significância em 1.848 cidades (Figura 11). Em outras palavras, a taxa de distorção idade-série em um município guarda uma associação negativa com a cobertura de esgoto em regiões vizinhas. A Figura 8 mostra agrupamentos Alto-Baixo em quase a totalidade da região Norte e parte da região Nordeste, e *clusters* Baixo-Alto no Sudeste. A grande quantidade de dados faltantes da cobertura de esgoto para o Nordeste pode ter prejudicado a identificação da dependência espacial bivariada tanto para o Saeb quanto para a defasagem idade-série nesta região.

Figura 11: Índice de Moran local bivariado: distorção idade-série (X) e coleta de esgoto (Y) (Brasil, 2017)



Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS e PNSB.

A próxima seção se destina a identificar e estimar um modelo de regressão espacial que capture com eficácia os tipos de interações espaciais entre os indicadores municipais de saneamento e qualidade do ensino, além de outras variáveis de controle.

3.3 Modelos econométricos

- Cobertura sanitária e notas do Saeb

Os testes LM-erro e LM-lag tradicionais foram executados para o modelo do Saeb e ambos rejeitaram a hipótese nula. Ou seja, o teste LM-erro indicou a presença da defasagem na variável dependente, enquanto o teste LM-lag favoreceu a existência de defasagem no termo do erro. A seguir, executaram-se as versões robustas desses testes e, novamente, ambos rejeitaram as hipóteses nulas de ausência de *lag* na variável dependente e no erro, respectivamente. O teste SARMA também apontou para a presença de ambos os tipos de defasagem no modelo. A Tabela 5 apresenta os resultados.

Tabela 5: Resultados dos testes multiplicadores de Lagrange - Saeb

Teste	Estatística	p-valor
LM-erro	1591,5	< 0.001
LM-lag	1178	< 0.001
LM-erro robusto	497,57	< 0.001
LM-lag robusto	84,073	< 0.001
SARMA	1675,5	< 0.001

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

Foram estimados os principais modelos lineares de regressão espacial *cross-section* (SDM, SDEM, SLX, SEM, SAR e SARAR), além do modelo por mínimos quadrados ordinários (MQO). A Tabela 6 detalha os valores encontrados.

Tabela 6: Estatísticas de ajuste dos modelos - Saeb

Estatística	SDM	SLX	SEM	SDEM	SAR	SARAR	MQO
AIC	5833,89	6838,449	5880,628	5664,974	6275,422	5860,203	7301,072
BIC	6085,154	7083,23	6015,902	5916,196	6410,696	6001,918	7429,904
<i>log</i> - verossimilhança	-2877,966	-3381,224	-2919,314	-2793,487	-3116,711	-2908,101	-3630,536

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

O modelo SDEM aparenta possuir o melhor ajuste entre os estimados na medida em que apresentou os menores valores para os critérios de informação e o maior valor para a estatística de *log*-verossimilhança, e, portanto, foi o selecionado para avaliar a associação entre as notas do Saeb e a cobertura de saneamento básico. A Figura 12 apresenta as estimativas obtidas pelo SDEM. Os resultados completos de todos os modelos estão presentes no Apêndice A.

Figura 12: Resultados do modelo SDEM para as notas do Saeb

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.72079903725	0.13402634802	35.2229	< 0.00000000000000022
perc_esg	-0.00047852776	0.00036882138	-1.2975	0.194476
ad	0.00215707577	0.00041273696	5.2263	0.0000001729614716
rd	0.00060323172	0.00028526860	2.1146	0.034463
ed	0.00252893464	0.00135817756	1.8620	0.062602
atu	-0.00406182639	0.00222095156	-1.8289	0.067419
horas	0.01263836998	0.00980938035	1.2884	0.197608
pib_pc	0.00000015068	0.00000043428	0.3470	0.728625
hierCapital Regio-1	0.06603195839	0.04106228816	1.6081	0.107815
hierCentro de Zo-	0.05510063672	0.02432566427	2.2651	0.023505
hierCentro subregio-1	0.04931567228	0.04151175914	1.1880	0.234836
hierMetrópole	0.00824642464	0.06257364598	0.1318	0.895152
pard	-0.00050622124	0.00055243812	-0.9163	0.359489
perc_vinc	0.01015785220	0.00109300539	9.2935	< 0.00000000000000022
F	-0.00007151182	0.00100401261	-0.0712	0.943218
agro	-0.00228060909	0.00071196109	-3.2033	0.001359
analf	-0.01250903593	0.00532962680	-2.3471	0.018921
renda	0.00069984642	0.00101776974	0.6876	0.491687
dummy0	0.07399326127	0.02402751748	3.0795	0.002073
lag.perc_esg	-0.00018679176	0.00067799844	-0.2755	0.782929
lag.ad	0.00387697492	0.00075167944	5.1578	0.0000002499345118
lag.rd	0.00061561086	0.00061503114	1.0009	0.316855
lag.ed	0.00446533789	0.00309487942	1.4428	0.149073
lag.atu	-0.02171105369	0.00426716721	-5.0879	0.0000003619904065
lag.horas	0.00204003455	0.01999135313	0.1020	0.918720
lag.pib_pc	0.00000017001	0.00000104592	0.1625	0.870875
lag.hierCapital Regio-1	-0.24868154015	0.07681391432	-3.2375	0.001206
lag.hierCentro de Zo-	0.05576256074	0.05031655990	1.1082	0.267760
lag.hierCentro subregio-1	-0.02841508187	0.08124439381	-0.3497	0.726528
lag.hierMetrópole	-0.22710765950	0.09592100569	-2.3677	0.017901
lag.pard	-0.00417733443	0.00096945291	-4.3090	0.0000164023567271
lag.perc_vinc	0.01851649782	0.00256706341	7.2131	0.00000000000005469
lag.F	0.00011966764	0.00187417021	0.0639	0.949089
lag.agro	-0.00100384319	0.00146239037	-0.6864	0.492436
lag.analf	-0.02799300188	0.01332207998	-2.1012	0.035619
lag.renda	0.00241319804	0.00231210271	1.0437	0.296613
lag.dummy0	0.07115315473	0.04984058804	1.4276	0.153403

Lambda: 0.56754, LR test value: 1177.4, p-value: < 0.000000000000000222
Asymptotic standard error: 0.015095
z-value: 37.597, p-value: < 0.000000000000000222
wald statistic: 1413.5, p-value: < 0.000000000000000222

Log likelihood: -2792.532 for error model
ML residual variance (sigma squared): 0.18052, (sigma: 0.42488)
Number of observations: 4636
Number of parameters estimated: 39
AIC: 5663.1, (AIC for lm: 6838.4)

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.

A análise dos *p*-valores das estimativas permite constatar que, *para as cidades com infraestrutura de saneamento*, não foi encontrada significância estatística entre a nota padronizada do Saeb e a ampliação da cobertura de esgoto, nem entre o Saeb local e o esgotamento de cidades vizinhas. Por outro lado, a existência de saneamento básico – em comparação ao cenário de ausência do serviço – está associada a um efeito positivo de 0.074

na nota do Saeb (erro padrão de 0.024), isolando-se os efeitos das demais variáveis. A estimativa do *lag* desta variável não foi significativa. Portanto, descarta-se a hipótese de transbordamentos espaciais das externalidades da cobertura de esgoto sobre este indicador educacional.

No âmbito das variáveis controle, a adequação docente, a regularidade docente e o percentual de vínculos ativos têm o sinal esperado positivo e significativo sobre a nota municipal do Saeb. A interpretação se dá da seguinte maneira: a ampliação de 1% em alguma dessas variáveis, isolando-se os efeitos das demais, está associada a um aumento de (coeficiente) da nota padronizada municipal do Saeb. O modelo também encontra uma associação positiva caso o município configure um Centro de Zona, em comparação a cidades classificadas como Centro Local. Já o aumento de vínculos no setor agropecuário e vínculos analfabetos guarda uma relação negativa com a nota da educação municipal, respeitando os sinais esperados¹¹ para esses indicadores.

Transbordamentos espaciais *positivos* sobre o Saeb foram identificados para a adequação docente e proporção de vínculos ativos na população. Já efeitos *spillover* espaciais negativos ocorrem para a quantidade de alunos por turma, proporção de pessoas negras e pardas nos vínculos ativos e percentual de analfabetos. Isso significa que a ampliação desses fatores em cidades vizinhas está associada a uma redução local na nota do Saeb. Similarmente, transbordamentos espaciais negativos ocorrem quando um município se situa próximo a capitais regionais e metrópoles, em comparação à vizinhança com centros locais.

O coeficiente Lambda é estatisticamente significativo, corroborando a presença de dependência espacial no termo do erro. Isto é, o componente da variabilidade das notas do Saeb não explicada pelo modelo é positivamente correlacionado no espaço, o que pode ser resultado da omissão de variáveis explicativas relevantes.

Ao aplicar o índice de Moran sobre os resíduos do SDEM, obtém-se um *p*-valor não significativo (~ 0.99), o que indica que o modelo foi eficaz em capturar a dependência espacial existente nos dados. Com exceção do SARAR e do SDM, todos os outros modelos falharam neste teste. Ademais, o *p*-valor encontrado para o SDM foi consideravelmente mais próximo da significância estatística (0.07).

Por fim, os testes de razão de verossimilhança rejeitaram a hipótese nula de que a restrição do SDEM para modelos mais simples, tais como o SLX, SEM ou MQO, seria mais adequada.

¹¹ Para um estudo mais aprofundado da relação entre emprego agrícola e nível de desenvolvimento humano, ver Köle et al. (2021).

- Cobertura sanitária e taxa de distorção idade-série

Os mesmos procedimentos seguidos no modelo com a variável dependente do Saeb foram aplicados para o modelo com a distorção idade-série como variável de interesse. Tanto os testes multiplicadores de Lagrange tradicionais como os robustos, além do teste SARMA, rejeitaram as hipóteses nulas, o que indica a presença de defasagem no termo de erro e na variável dependente. Dessa forma, modelos que não possuem nenhum desses tipos de defasagem, como o SLX e o MQO, podem não ser os mais adequados. A Tabela 7 expõe os resultados.

Tabela 7: Resultados dos testes multiplicadores de LaGrange – Distorção idade-série

Teste	Estatística	p-valor
LM-erro	2516	< 0.001
LM-lag	3080,9	< 0.001
LM-erro robusto	10,977	< 0.001
LM-lag robusto	575,95	< 0.001
SARMA	3091,9	< 0.001

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

Ao comparar os critérios de informação e a estatística de *log*-verossimilhança obtidos nos diferentes modelos espaciais e no MQO, novamente o modelo SDEM aparenta possuir o melhor ajuste, sendo selecionado para a análise das estimativas (Tabela 8). Os resultados do SDEM estão presentes na Figura 13, enquanto as estimativas obtidas nos outros modelos estão no Apêndice B.

Tabela 8: Estatísticas de ajuste dos modelos – Distorção idade-série

Estatística	SDM	SLX	SEM	SDEM	SAR	SARAR	MQO
AIC	34112,41	34885,98	33049,83	32870	32951,69	33998,93	35758,23
BIC	34364,39	35131,50	33185,51	32869,84	33087,37	34141,07	35887,45
<i>log</i> - verossimilhança	-17017,02	-17404,99	-16503,91	-16395,92	-16454,81	-16977,46	-17859,12

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS.

Figura 13: Resultados do modelo SDEM para a distorção idade-série

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	33.3030270227	2.5003291779	13.3195	< 0.00000000000000022
perc_esg	-0.0175155405	0.0067377615	-2.5996	0.0093330
ad	-0.1023472995	0.0094476859	-10.8331	< 0.00000000000000022
rd	0.0105914104	0.0051283766	2.0653	0.0388988
ed	0.1516669073	0.0273253791	5.5504	0.0000000285009
atu	0.1356673293	0.0315829675	4.2956	0.0000174233064
horas	-0.5601889913	0.1880871115	-2.9783	0.0028981
pib_pc	0.0000043045	0.0000078484	0.5485	0.5833771
hierCapital Regio-1	1.4582615459	0.7285741762	2.0015	0.0453355
hierCentro de Zo-	0.1794882969	0.4353748896	0.4123	0.6801478
hierCentro subregio-1	1.3521992075	0.7390203807	1.8297	0.0672921
hierMetrópole	1.7300517164	1.0663479852	1.6224	0.1047160
pard	0.0009431161	0.0099545130	0.0947	0.9245193
perc_vinc	-0.0705346755	0.0198397661	-3.5552	0.0003777
F	0.0128822296	0.0176827586	0.7285	0.4662958
agro	-0.0007819663	0.0127459244	-0.0614	0.9510802
analf	0.0823109291	0.0968114616	0.8502	0.3952034
renda	-0.0219892541	0.0182899562	-1.2023	0.2292634
dummy0	-0.5674464616	0.4382483125	-1.2948	0.1953872
lag.perc_esg	-0.0433096749	0.0137906201	-3.1405	0.0016865
lag.ad	-0.1248563259	0.0195336629	-6.3919	0.000000001639
lag.rd	0.0326281476	0.0121570774	2.6839	0.0072773
lag.ed	0.2588368327	0.0688168561	3.7612	0.0001691
lag.atu	0.3052048640	0.0687391737	4.4400	0.0000089941046
lag.horas	-0.2204937229	0.4074278159	-0.5412	0.5883802
lag.pib_pc	0.0000187364	0.0000202410	0.9257	0.3546211
lag.hierCapital Regio-1	4.0777012112	1.4763738269	2.7620	0.0057454
lag.hierCentro de Zo-	0.6462185825	0.9601582545	0.6730	0.5009260
lag.hierCentro subregio-1	4.3065097292	1.5127735802	2.8468	0.0044166
lag.hierMetrópole	1.2096041823	1.8326478819	0.6600	0.5092340
lag.pard	0.0224188633	0.0203205261	1.1033	0.2699134
lag.perc_vinc	-0.2576613499	0.0506779548	-5.0843	0.0000003690064
lag.F	0.0186535702	0.0362789112	0.5142	0.6071323
lag.agro	-0.1155830181	0.0284357590	-4.0647	0.0000480929151
lag.analf	0.5259274573	0.2553303502	2.0598	0.0394184
lag.renda	-0.0557937304	0.0451290283	-1.2363	0.2163412
lag.dummy0	-1.0763912316	1.0005122102	-1.0758	0.2819987

Lambda: 0.70461, LR test value: 2018.1, p-value: < 0.000000000000000222
Asymptotic standard error: 0.012143
z-value: 58.024, p-value: < 0.000000000000000222
wald statistic: 3366.8, p-value: < 0.000000000000000222

Log likelihood: -16395.92 for error model
ML residual variance (sigma squared): 53.026, (sigma: 7.2819)
Number of observations: 4727
Number of parameters estimated: 39
AIC: 32870, (AIC for lm: 34886)

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.

Os resultados apontam significância estatística tanto para a variável contínua da cobertura de esgoto local, quanto para a defasagem espacial desse componente. Para as cidades em que a infraestrutura existe, a ampliação de 1% na cobertura de esgoto municipal estaria associada a uma redução de 0.017% na taxa de distorção idade-série (erro padrão igual a 0.006), enquanto o incremento percentual da cobertura em cidades vizinhas associa-se a uma

diminuição de 0.043% do indicador (erro padrão de 0.013). Dessa forma, os efeitos relacionados aos transbordamentos espaciais do saneamento seriam mais fortes que a ampliação da cobertura local, isolando-se a influência das demais variáveis. Em contrapartida, o modelo não encontrou significância estatística para a *dummy* que representa a presença (ou ausência) do serviço no município ou em cidades próximas.

Ademais, a ampliação da adequação docente, média de horas-aula diárias e percentual de vínculos empregatícios locais e de municípios vizinhos se relaciona com uma redução da defasagem idade-série, sendo que para a variável das horas-aula o coeficiente aponta uma diminuição expressiva, de 0.56%. Já a ampliação da média de alunos por turma e do indicador de esforço (sobrecarga) docente – tanto na esfera municipal quanto de cidades próximas – está associada a um aumento relativo de alunos com idade superior à recomendada para a série que frequentam.

A maior defasagem serial também é observada para cidades classificadas como capitais regionais, em comparação aos centros locais. Já para municípios vizinhos de capitais regionais ou de centros sub-regionais, associa-se uma elevação da ordem de 4% deste indicador educacional, comparativamente à proximidade com centros locais. A ampliação dos vínculos de pessoas analfabetas também está relacionada com um aumento da distorção idade-série, enquanto esta decresce com o aumento de vínculos no setor agropecuário. Em um resultado contraintuitivo, o modelo aponta um coeficiente positivo e significativo para a proporção de docentes com IRD alto e médio-alto.

A defasagem no termo de erro é corroborada pelo valor do Lambda, significativo estatisticamente. O teste do índice de Moran sobre os resíduos não rejeitou a hipótese nula de distribuição aleatória no espaço (p-valor = ~0.99), de forma que o modelo aparenta ter capturado com eficácia a autocorrelação espacial presente nos dados. Os únicos modelos que também passaram neste teste foram o SEM e o SAR. O teste de razão de verossimilhança rejeitou a necessidade de restringir o SDEM para um SEM, SLX ou MQO.

- Discussão

Os resultados encontrados foram, de certa forma, heterogêneos. Nos dois modelos estimados, há associação entre a ampliação do serviço de coleta de esgoto e melhorias no rendimento escolar, embora no caso do Saeb esta interação ocorra via instalação de sistemas de saneamento, enquanto para a distorção idade-série a externalidade existe para a ampliação da infraestrutura existente.

A literatura internacional sobre a relação do saneamento com as notas de estudantes é escassa, o que dificulta traçar paralelos entre o modelo do Saeb e evidências empíricas. Contudo, tais resultados são consistentes com o modelo teórico de Glewwe e Miguel (2007), que argumentam que choques positivos no ambiente sanitário têm efeitos incertos (positivos ou negativos) sobre as habilidades acadêmicas das crianças, a depender de fatores como a função de utilidade dos pais, impacto marginal sobre os gastos com saúde, efeito renda e efeito preço. Já os resultados sobre a distorção serial têm correspondência com aqueles de trabalhos mencionados no capítulo 2, como Scriptori (2016) e Agol et al. (2018), que reportam evidências para o Brasil e Zâmbia, respectivamente.

A seleção pelo modelo SDEM – que modela efeitos de transbordamento espacial nas variáveis explicativas (saneamento) e termo do erro (fatores não observáveis) – pode encontrar respaldo no fato de que resíduos sólidos são levados por rios e correntes marítimas, afetando a população de cidades com sistema próprio de saneamento. Insetos vetores, como mosquitos e parasitas, também se deslocam além dos limites municipais, trazendo consigo doenças infecciosas. Esses são argumentos apresentados por Scriptori (2016).

Ademais, os modelos estimados podem ter omitido variáveis relevantes na determinação de indicadores educacionais – como, por exemplo, a parcela da população municipal que recebe Bolsa Família –, mas cujos dados inexistem ou não estavam disponíveis para o ano analisado. Uma evidência nesse sentido é o baixo valor para a estatística do R^2 ajustado nos modelos estimados por MQO (0,27 para o Saeb e 0,49 para a distorção). A seleção do SDEM também é respaldada por Lesage (2014), que argumenta que esta é a especificação mais comum para estudos empíricos na ciência regional

No âmbito as variáveis de controle da distorção serial, Scriptori (2016) também encontra significância estatística e sinal positivo para o analfabetismo e média de alunos por turma, e negativo para a adequação docente. No entanto, obteve resultados significativos para a população negra e parda e renda *per capita*, o que não ocorreu nesta monografia¹².

¹² Cabe ressaltar que a autora utiliza dados do Censo Demográfico 2010, enquanto nesta monografia os dados socioeconômicos provêm da RAIS, que inclui apenas informações de vínculos trabalhistas.

CONSIDERAÇÕES FINAIS

A associação entre saneamento básico e saúde pública, como a incidência de doenças infecciosas e morbidade hospitalar, é amplamente evidenciada na literatura. Uma face menos intuitiva, mas com fortes implicações para as políticas públicas, é sua relação com a qualidade da educação, tendo em vista as externalidades da ampliação do saneamento sobre o desenvolvimento cognitivo de crianças, aumento da frequência escolar e redução da desigualdade de gênero. Este trabalho utiliza dados do SNIS, PNSB, IBGE e RAIS para investigar a associação entre a cobertura de esgoto municipal e indicadores educacionais no contexto brasileiro, a saber: a taxa de defasagem serial e a nota padronizada do Saeb. A análise foi realizada com dados de 2017, de modo a compatibilizar as informações de diferentes bases municipais.

A análise exploratória de dados espaciais, que visa identificar e quantificar a autocorrelação espacial no território, indicou a presença da dependência espacial positiva para as três variáveis de interesse da monografia e para os resíduos dos modelos de regressão por MQO. A distorção idade-série apresentou o maior valor do Índice de Moran global, da ordem de 0.727, seguido do Saeb (0.581). Em ambos os casos, pode-se observar *clusters* de indicadores educacionais favoráveis (altos para o Saeb, e baixos para a defasagem serial) no Centro-Sul, e resultados desfavoráveis nas regiões Norte, Nordeste e em grande porção do Mato-Grosso.

Todavia, mesmo nestes agrupamentos, identificam-se regiões que vão contra a tendência. O estado do Ceará, por exemplo, possui *clusters* de notas mais altas que a média nacional do Saeb e menores taxas de distorção idade-série; no estado do Rio de Janeiro, por outro lado, o alto patamar de distorção serial chama atenção. O Índice de Moran bivariado global denotou que municípios com notas municipais do Saeb acima da média nacional ou taxas de distorção abaixo dela tendem a ter como vizinhos municípios em que a cobertura de esgoto municipal se situa acima da média. Mais uma vez, o indicador bivariado local identifica agrupamentos favoráveis no âmbito das três variáveis de interesse em estados do Centro-Sul, e desfavoráveis no Norte, Nordeste e parte do Centro-Oeste, embora a falta de dados da cobertura de esgoto e Saeb possa ter prejudicado a estimação dos resultados locais.

Tendo em vista a presença de autocorrelação espacial nos resíduos, a estimação dos modelos econométricos ocorreu pela inserção de componentes de defasagem espacial (*lag*), de modo que foram avaliados os principais tipos de modelos de regressão espacial *cross-section*. A escolha final foi orientada pelos critérios de informação Akaike e BIC, valor da *log-*

verossimilhança e teste de razão de verossimilhança, metodologia análoga à adotada por Scriptore (2016). Dessa forma, a especificação selecionada – tanto para o Saeb quanto para a distorção idade-série – foi o Modelo do Erro Espacial de Durbin (SDEM), que possui defasagem nas variáveis explicativas e no termo do erro, dando origem a efeitos de transbordamento espacial locais.

Se por um lado não foi encontrada associação significativa entre ampliação do saneamento existente e melhorias no Saeb, o modelo encontrou que a existência da infraestrutura, mesmo que insuficiente, se relaciona com um melhor rendimento das notas da Educação Básica. Já para a distorção idade-série, constata-se que, nas cidades onde a infraestrutura existe, a ampliação da cobertura de esgoto está relacionada a uma redução de cerca de 0.017% desse indicador, e a diminuição é ainda mais expressiva quando se eleva a cobertura de municípios vizinhos: da ordem de 0.043%, isolando-se o efeito das demais variáveis.

Em ambos os casos, o Índice de Moran denotou a eliminação da autocorrelação espacial dos resíduos, indicando que os modelos foram especificados adequadamente. Contudo, as estimativas encontradas podem ter sido enviesadas pela endogeneidade, isto é, a presença de autocorrelação das variáveis explicativas com o termo de erro. Esta, por sua vez, pode ser resultado da omissão de variáveis explicativas relevantes, visto que uma série de fatores socioeconômicos contribuem para a variabilidade dos indicadores educacionais. Ademais, os critérios de informação AIC, BIC e a *log*-verossimilhança dão suporte à especificação dentro de um conjunto de modelos estimados, mas não necessariamente refletem o modelo mais adequado dentro da miríade de modelos econométricos existentes.

Por fim, por resultarem de modelos de regressão linear, as estimativas encontradas demonstram apenas a associação entre saneamento e educação, e não uma relação causal entre os fatores. Futuros trabalhos sobre o tema devem considerar a estimação com o uso de variáveis instrumentais, a fim de eliminar o viés da endogeneidade e aprofundar o conhecimento sobre as formas de interação entre o saneamento básico e o aproveitamento escolar. Não obstante, os resultados encontrados para a distorção idade-série indicam que iniciativas públicas eficazes para a ampliação da qualidade da educação devem ser coordenadas a nível regional, e não conduzidas de maneira isolada pelos municípios.

REFERÊNCIAS

ABCON SINDCON. Panorama da participação privada no saneamento – Brasil 2017.

ADUKIA, Anjali. Sanitation and education. **American Economic Journal: Applied Economics**, v. 9, n. 2, p. 23-59, 2017.

AGOL, Dorice; HARVEY, Peter; MAÍLLO, Javier. Sanitation and water supply in schools and girls' educational progression in Zambia. **Journal of Water, Sanitation and Hygiene for Development**, v. 8, n. 1, p. 53-61, 2018.

ALMEIDA, E. (2012). **Econometria Espacial Aplicada**. Campinas, SP: Editora Alínea.

ANSELIN, Luc. Interactive techniques and exploratory spatial data analysis. 1996.

ANSELIN, Luc. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. **Geographical analysis**, v. 20, n. 1, p. 1-17, 1988.

ANSELIN, Luc. Specification tests (2). 2017. Disponível em https://spatial.uchicago.edu/sites/spatial.uchicago.edu/files/10_specification_tests_2_slides.pdf

ANSELIN, Luc; FLORAX, Raymond; REY, Sergio J. (Ed.). **Advances in spatial econometrics: methodology, tools and applications**. Springer Science & Business Media, 2013.

ARBIA, Giuseppe; ESPA, Giuseppe; GIULIANI, Diego. Dirty spatial econometrics. **The Annals of Regional Science**, v. 56, n. 1, p. 177-189, 2016.

BANERJEE, A. V., & DUFLO, E. (2011). **Poor economics: A radical rethinking of the way to fight global poverty**. Nova York: PublicAffairs.

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

BENOVA, Lenka; CUMMING, Oliver; CAMPBELL, Oona MR. Systematic review and meta-analysis: association between water and sanitation environment and maternal mortality. *Tropical medicine & international health*, v. 19, n. 4, p. 368-387, 2014.

BRASIL. INEP. Nota técnica nº 039 de 17 de dezembro de 2014. Indicador de Esforço Docente. Disponível em https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_esforco/nota_tecnica_indicador_docente_esforco.pdf

COOK, Scott J.; HAYS, Jude C.; FRANZESE, Robert J. Model specification and spatial interdependence. In: **APSA Conference Paper**. 2015.

CORRADO, Luisa; FINGLETON, Bernard. Where is the economics in spatial econometrics?. **Journal of Regional Science**, v. 52, n. 2, p. 210-239, 2012.

- COSTA, Djalma Ribeiro et al. Associação entre taxa de mortalidade infantil e indicadores de habitação e saneamento: um guia para tomada de decisão em medicina preventiva e social. **Revista de Administração em Saúde**, v. 20, n. 79, 2020.
- CUNHA, Alexandre dos Santos. Saneamento básico no Brasil: desenho institucional e desafios federativos. In: **Saneamento básico no Brasil: desenho institucional e desafios federativos**. 2011. p. 26-26.
- DA CRUZ, Felipe Ponciano et al. A Regionalização dos Serviços de Saneamento Básico e os Desafios da Universalização no Brasil: Uma Análise Exploratória de Dados Espaciais para os anos de 2010 e 2018 *Área 10: Economia Regional e Urbana*, 2021.
- DE MENDONÇA, Mário Jorge Cardoso; DA MOTTA, Ronaldo Seroa. Saúde e saneamento no Brasil. **Planejamento e Políticas Públicas**, n. 30, 2009.
- DOS SANTOS LIMA, Jarleson et al. Morbidade e mortalidade por doenças relacionadas ao saneamento ambiental inadequado em Belém do Pará. **Ciência E Sustentabilidade**, v. 4, n. 2, p. 115-132, 2018.
- DREIBELBIS, Robert et al. Water, sanitation, and primary school attendance: A multi-level assessment of determinants of household-reported absence in Kenya. **International Journal of Educational Development**, v. 33, n. 5, p. 457-465, 2013.
- DUNCAN, Earl W.; WHITE, Nicole M.; MENGERSEN, Kerrie. Spatial smoothing in Bayesian models: a comparison of weights matrix specifications and their impact on inference. **International journal of health geographics**, v. 16, n. 1, p. 1-16, 2017.
- ELHORST, J. Paul. Applied spatial econometrics: raising the bar. **Spatial economic analysis**, v. 5, n. 1, p. 9-28, 2010.
- FIRMINO COSTA DA SILVA, Diego; ELHORST, J. Paul; SILVEIRA NETO, Raul da Mota. Urban and rural population growth in a spatial panel of municipalities. **Regional Studies**, v. 51, n. 6, p. 894-908, 2017.
- FREEMAN, Matthew C. et al. Assessing the impact of a school-based water treatment, hygiene and sanitation programme on pupil absence in Nyanza Province, Kenya: a cluster-randomized trial. **Tropical medicine & international health**, v. 17, n. 3, p. 380-391, 2012.
- GARN, J. V., GREENE, L. E., DREIBELBIS, R., SABOORI, S., RHEINGANS, R. D., & FREEMAN, M. C. (2013). A cluster-randomized trial assessing the impact of school water, sanitation and hygiene improvements on pupil enrolment and gender parity in enrolment. **Journal of Water, Sanitation and Hygiene for Development**, 3(4), 592–601.
- GLEWWE, Paul; MIGUEL, Edward A. The impact of child health and nutrition on education in less developed countries. **Handbook of development economics**, v. 4, p. 3561-3606, 2007.
- HARRIS, Richard; MOFFAT, John; KRAVTSOVA, Victoria. In search of ‘W’. **Spatial Economic Analysis**, v. 6, n. 3, p. 249-270, 2011.

HUTTON, Guy et al. Global costs and benefits of drinking-water supply and sanitation interventions to reach the MDG target and universal coverage. World Health Organization, 2012.

JEWITT, Sarah; RYLEY, Harriet. It's a girl thing: Menstruation, school attendance, spatial mobility and wider gender inequalities in Kenya. **Geoforum**, v. 56, p. 137-147, 2014.

KOPCZEWSKA, Katarzyna. **Applied spatial statistics and econometrics: data analysis in R**. Routledge, 2020.

KOSTOV, Philip. Model boosting for spatial weighting matrix selection in spatial lag models. **Environment and Planning B: Planning and Design**, v. 37, n. 3, p. 533-549, 2010.

KOTHARI, Monica T. et al. Exploring associations between water, sanitation, and anemia through 47 nationally representative demographic and health surveys. *Annals of the New York Academy of Sciences*, v. 1450, n. 1, p. 249, 2019.

LESAGE, James P. What regional scientists need to know about spatial econometrics. **Available at SSRN 2420725**, 2014.

LESAGE, James P.; PACE, R. Kelley. The biggest myth in spatial econometrics. **Econometrics**, v. 2, n. 4, p. 217-249, 2014.

LESAGE, James; PACE, Robert Kelley. **Introduction to spatial econometrics**. Chapman and Hall/CRC, 2009.

LOUREIRO, Gustavo Kaercher. Observações sobre a prorrogação de contratos de programa sob o Marco do Saneamento. 2020.

MANSKI, Charles F. Identification of endogenous social effects: The reflection problem. **The review of economic studies**, v. 60, n. 3, p. 531-542, 1993.

NGURE, Francis M. et al. Water, sanitation, and hygiene (WASH), environmental enteropathy, nutrition, and early child development: making the links. *Annals of the New York Academy of Sciences*, v. 1308, n. 1, p. 118-128, 2014.

ORTIZ-CORREA, Javier Santiago; RESENDE FILHO, Moises; DINAR, Ariel. Impact of access to water and sanitation services on educational attainment. **Water Resources and Economics**, v. 14, p. 31-43, 2016.

PESSANHA, F. B., Alves, K.S.C., Shimoda, E., Erthal Jr., M. Benefícios econômicos do saneamento básico para o município de Campos dos Goytacazes, RJ. XXXI Encontro Nacional De Engenharia De Produção. Inovação Tecnológica e Propriedade Intelectual: Desafios da Engenharia de Produção na Consolidação do Brasil no Cenário Econômico Mundial. 2011.

RAHAL, Charles. A Grid Based Approach to Analysing Spatial Weighting Matrix Specification. 2017.

SANT'ANNA, André Albuquerque; ROCHA, Romero. Saneamento e morbidade: evidências a partir de variações exógenas no tempo de execução dos investimentos. 2021.

SANTOS, Angela Moulin S. Penalva; SANTOS, Juliana. Saneamento Básico no Estado do Rio de Janeiro: longo percurso rumo à regulação. **Cadernos do Desenvolvimento Fluminense**, n. 5, p. 1-18, 2014.

SARRIAS, Mauricio. Notes on Spatial Econometrics. 2020.

SCRIPTORE, Juliana Souza. Impactos do saneamento sobre saúde e educação: uma análise espacial. São Paulo: USP, 2016.

SCRIPTORE, Juliana Souza; AZZONI, Carlos. A spatial estimation of the impacts of sanitation on education. **Regional Science Policy & Practice**, v. 12, n. 1, p. 159-166, 2020.

SILVA, Valéria Andrade; ESPERIDIÃO, Fernanda. Saneamento básico e seus impactos na mortalidade infantil e no desenvolvimento econômico da região Nordeste. **Scientia Plena**, v. 13, n. 10, 2017.

SMITH, Tony E. Estimation bias in spatial models with strongly connected weight matrices. **Geographical Analysis**, v. 41, n. 3, p. 307-332, 2009.

SOUSA, Ana Cristina A. de; COSTA, Nilson do Rosário. Política de saneamento básico no Brasil: discussão de uma trajetória. **História, Ciências, Saúde-Manguinhos**, v. 23, p. 615-634, 2016.

STAKHOVYCH, Stanislav; BIJMOLT, Tammo HA. Specification of spatial models: A simulation study on weights matrices. **Papers in Regional Science**, v. 88, n. 2, p. 389-408, 2009.

TEIXEIRA, J. C.; GOMES, M. H. R.; SOUZA, J. A. Associação entre cobertura por serviços de saneamento e indicadores epidemiológicos nos países da América Latina: estudo com dados secundários. **Revista Panamericana de Salud Pública**, v. 32, n. 6, p. 419-425, 2012.

TRATA BRASIL. Benefícios econômicos e sociais da expansão do saneamento brasileiro. São Paulo, 2018.

TRATA BRASIL. Saneamento e Doenças de Veiculação Hídrica – ano base 2019. São Paulo, 2021.

TUROLLA, Frederico Araújo. Política de saneamento básico: avanços recentes e opções futuras de políticas públicas. **Política de saneamento básico: avanços recentes e opções futuras de políticas públicas**, p. 26-26, 2002.

UHR, Júlia Gallego Ziero; SCHMECHEL, Mariana; UHR, Daniel de Abreu Pereira. Relação entre saneamento básico no Brasil e saúde da população sob a ótica das internações hospitalares por doenças de veiculação hídrica. **RACEF – Revista de Administração, Contabilidade e Economia da Fundace**, v. 7, n. 2, p. 01-16, 2016.

VENTURA, Grace Kelly Amaral; LOPES, Frederico Wagner Azevedo. Infraestrutura de saneamento básico e incidência de doenças associadas: uma análise comparativa entre Belo Horizonte e Ribeirão das Neves-Minas Gerais/Sanitation infrastructure and associated diseases: a comparative analysis between Belo Horizonte (...). **Caderno de Geografia**, v. 27, n. 51, p. 788-805, 2017.

VERDIANO, Aurilano de Araújo. Um modelo explicativo do avanço na qualidade da educação do ensino fundamental no estado do Ceará. 2019.

WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation. (2012). Progress on drinking water and sanitation: 2012 update. World Health Organization.

ZHANG, Jing; XU, Lixin Colin. The long-run effects of treated water on education: The rural drinking water program in China. **Journal of Development Economics**, v. 122, p. 1-15, 2016.

APÊNDICE A – RESULTADOS DOS MODELOS ECONÔMICOS PARA O SAEB

Figura 14: Resultados do modelo SDM para as notas do Saeb

Coefficients: (asymptotic standard errors)					
	Estimate	Std. Error	z value		Pr(> z)
(Intercept)	3.64645672194	0.11815885112	30.8606	< 0.00000000000000022	
perc_esg	-0.00051851736	0.00037895469	-1.3683		0.1712234
ad	0.00197751941	0.00043776675	4.5173	0.000006263607700507	
rd	0.00062154812	0.00028290198	2.1970		0.0280173
ed	0.00164080371	0.00132348424	1.2398		0.2150639
atu	0.00465826084	0.00226839430	2.0535		0.0400193
horas	0.03167487812	0.00951927157	3.3274		0.0008765
pib_pc	0.00000025733	0.00000042193	0.6099		0.5419401
hierCapital Regio-1	0.07166981799	0.04221259274	1.6978		0.0895398
hierCentro de Zo-	0.03330115019	0.02265031839	1.4702		0.1414998
hierCentro subregio-1	0.00568100025	0.03847964878	0.1476		0.8826297
hierMetrópole	0.01144278952	0.06975004593	0.1641		0.8696885
pard	0.00073334624	0.00059125680	1.2403		0.2148579
perc_vinc	0.00864541601	0.00108463510	7.9708	0.000000000000001554	
F	0.00196944579	0.00098970096	1.9899		0.0465975
agro	-0.00158819957	0.00072915580	-2.1781		0.0293960
analf	-0.01123923788	0.00513633194	-2.1882		0.0286562
renda	0.00140081878	0.00099129856	1.4131		0.1576219
dummy0	0.07038936763	0.02334163329	3.0156		0.0025646
lag.perc_esg	0.00043837064	0.00053784490	0.8151		0.4150435
lag.ad	-0.00024537307	0.00060086426	-0.4084		0.6830043
lag.rd	-0.00064972549	0.00046873595	-1.3861		0.1657095
lag.ed	-0.00211677293	0.00236450072	-0.8952		0.3706639
lag.atu	-0.03002418207	0.00326802618	-9.1873	< 0.00000000000000022	
lag.horas	-0.08648651584	0.01601538577	-5.4002	0.000000066561322454	
lag.pib_pc	-0.00000111679	0.00000080076	-1.3947		0.1631185
lag.hierCapital Regio-1	-0.24809757640	0.06427201463	-3.8601		0.0001133
lag.hierCentro de Zo-	0.03299956585	0.04040615803	0.8167		0.4141019
lag.hierCentro subregio-1	-0.10001565409	0.06593717203	-1.5168		0.1293089
lag.hierMetrópole	-0.18389461231	0.08669571801	-2.1211		0.0339092
lag.pard	-0.00300074049	0.00076614954	-3.9167	0.000089787516006634	
lag.perc_vinc	0.00897925612	0.00199378282	4.5036	0.000006680310132623	
lag.F	-0.01340450992	0.00155760916	-8.6058	< 0.00000000000000022	
lag.agro	-0.00458731047	0.00114102565	-4.0203	0.000058114331646486	
lag.analf	-0.04683633492	0.01054158883	-4.4430	0.000008871098459151	
lag.renda	0.00220948996	0.00180770309	1.2223		0.2216079
lag.dummy0	0.02319997571	0.03755378239	0.6178		0.5367204

Rho: 0.48316, LR test value: 1006.5, p-value: < 0.000000000000000222

Asymptotic standard error: 0.014539

z-value: 33.233, p-value: < 0.000000000000000222

wald statistic: 1104.4, p-value: < 0.000000000000000222

Log likelihood: -2877.966 for mixed model

ML residual variance (sigma squared): 0.19188, (sigma: 0.43804)

Number of observations: 4636

Number of parameters estimated: 39

AIC: 5833.9, (AIC for lm: 6838.4)

LM test for residual autocorrelation

test value: 10.015, p-value: 0.0015531

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.

Figura 15: Resultados do modelo SLX para as notas do Saeb

```

Coefficients:
              Estimate      Std. Error t value      Pr(>|t|)
(Intercept)    4.64708875335    0.12586323845    36.922 < 0.00000000000000002
perc_esg      -0.00076459483    0.00043583855    -1.754    0.079444
ad             0.00192593345    0.00050333005     3.826    0.000132
rd             0.00059647690    0.00032532695     1.833    0.066797
ed            0.00234159132    0.00152214486     1.538    0.124032
atu          -0.00203445757    0.00258114256    -0.788    0.430620
horas         0.01432382510    0.01088349770     1.316    0.188204
pib_pc        0.00000001645    0.00000048510     0.034    0.972954
hierCapital.Regio.1 0.06035981724    0.04854828233     1.243    0.213823
hierCentro.de.Zo. 0.05677889164    0.02603147939     2.181    0.029222
hierCentro.subregio.1 0.03519972470    0.04423536011     0.796    0.426226
hierMetrópole -0.03645249056    0.08021848170    -0.454    0.649552
pard         -0.00011632320    0.00067879268    -0.171    0.863942
perc_vinc     0.01026996646    0.00124647727     8.239    0.000000000000000224
F             0.00019992168    0.00112980035     0.177    0.859553
agro         -0.00257079812    0.00083692582    -3.072    0.002141
analf        -0.01580003142    0.00590452272    -2.676    0.007479
renda         0.00120386333    0.00113995813     1.056    0.290997
dummy0       0.08070886602    0.02684441851     3.007    0.002657
lag.perc_esg -0.00118531833    0.00061719553    -1.920    0.054858
lag.ad        0.00407765301    0.00067656658     6.027    0.000000001801058515
lag.rd        0.00072086890    0.00053759150     1.341    0.180012
lag.ed        0.00330054750    0.00271381589     1.216    0.223971
lag.atu      -0.02258866679    0.00375523988    -6.015    0.000000001935471018
lag.horas    -0.01413745367    0.01826038981    -0.774    0.438844
lag.pib_pc   -0.00000020012    0.00000092032    -0.217    0.827872
lag.hierCapital.Regio.1 -0.39797933008    0.07384864090    -5.389    0.000000074340946634
lag.hierCentro.de.Zo. 0.08092921685    0.04645962023     1.742    0.081588
lag.hierCentro.subregio.1 -0.13280896285    0.07583479387    -1.751    0.079962
lag.hierMetrópole -0.29542681093    0.09965741060    -2.964    0.003048
lag.pard     -0.00317336886    0.00088090332    -3.602    0.000319
lag.perc_vinc 0.02739860985    0.00222823221    12.296 < 0.00000000000000002
lag.F        -0.00057002999    0.00174073616    -0.327    0.743331
lag.agro     -0.00200850985    0.00130916505    -1.534    0.125051
lag.analf    -0.06928626931    0.01212071589    -5.716    0.000000011571211198
lag.renda    0.00509966847    0.00207808949     2.454    0.014164
lag.dummy0   0.15146079183    0.04303188689     3.520    0.000436
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5038 on 4599 degrees of freedom
Multiple R-squared:  0.348,    Adjusted R-squared:  0.3429
F-statistic: 68.18 on 36 and 4599 DF,  p-value: < 0.00000000000000022

```

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com p -valores significativos a 5% foram realçadas em amarelo.

Figura 16: Resultados do modelo SEM para as notas do Saeb

```

Coefficients: (asymptotic standard errors)
              Estimate      Std. Error  z value      Pr(>|z|)
(Intercept)    4.86113060280  0.08793168077  55.2830 < 0.000000000000000022
perc_esg       0.00010440174  0.00036040279   0.2897      0.772061
ad             0.00274152891  0.00041510187   6.6045  0.0000000000039893422
rd             0.00044048004  0.00027600064   1.5959      0.110503
ed             0.00272485537  0.00128728419   2.1167      0.034281
atu           -0.00448488040  0.00219920303  -2.0393      0.041418
horas          0.01446130329  0.00930050944   1.5549      0.119971
pib_pc         0.00000028787  0.00000041068   0.7010      0.483324
hierCapital Regio-1  0.09953176965  0.04118672580   2.4166      0.015666
hierCentro de Zo-  0.04552746899  0.02053207546   2.2174      0.026597
hierCentro subregio-1  0.04999431983  0.03455306288   1.4469      0.147929
hierMetrópole   -0.00742836046  0.05818161453  -0.1277      0.898406
pard          -0.00149891603  0.00056024729  -2.6755      0.007463
perc_vinc      0.00866080114  0.00105285209   8.2260  0.0000000000000000222
F             -0.00140192511  0.00095571187  -1.4669      0.142406
agro          -0.00157583466  0.00070305086  -2.2414      0.024999
analf         -0.01037019140  0.00498356146  -2.0809      0.037445
renda         -0.00118644032  0.00095342052  -1.2444      0.213351
dummy0        0.05722707669  0.02271532365   2.5193      0.011758

Lambda: 0.64884, LR test value: 1422.4, p-value: < 0.0000000000000000222
Asymptotic standard error: 0.013369
z-value: 48.532, p-value: < 0.0000000000000000222
wald statistic: 2355.4, p-value: < 0.0000000000000000222

Log likelihood: -2919.314 for error model
ML residual variance (sigma squared): 0.18512, (sigma: 0.43025)
Number of observations: 4636
Number of parameters estimated: 21
AIC: 5880.6, (AIC for lm: 7301.1)

```

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.

Figura 17: Resultados do modelo SAR para as notas do Saeb

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.70047948950	0.11006848217	24.5345	< 0.00000000000000022
perc_esg	-0.00041777568	0.00032987243	-1.2665	0.2053425
ad	0.00346137386	0.00036054090	9.6005	< 0.00000000000000022
rd	0.00057524512	0.00027836444	2.0665	0.0387796
ed	0.00351080434	0.00133381324	2.6322	0.0084845
atu	-0.00908371399	0.00209095834	-4.3443	0.00001397315
horas	0.01633468348	0.00968088519	1.6873	0.0915432
pi_b_pc	0.00000037504	0.00000043363	0.8649	0.3871110
hierCapital Regio-1	-0.00341174425	0.03874162192	-0.0881	0.9298258
hierCentro de Zo-	0.05997081127	0.02328654828	2.5753	0.0100141
hierCentro subregio-1	0.03604483330	0.04028553529	0.8947	0.3709294
hierMetrópole	-0.10666363189	0.03953069573	-2.6982	0.0069705
pard	-0.00251614859	0.00046603190	-5.3991	0.00000006698
perc_vinc	0.01339195499	0.00106923336	12.5248	< 0.00000000000000022
F	0.00007889109	0.00099860323	0.0790	0.9370315
agro	-0.00203539457	0.00066953586	-3.0400	0.0023657
analf	-0.01935584820	0.00530804824	-3.6465	0.0002658
renda	-0.00064005004	0.00098061968	-0.6527	0.5139500
dummy0	0.08506657038	0.02336846369	3.6402	0.0002724

Rho: 0.41069, LR test value: 1027.6, p-value: < 0.000000000000000222

Asymptotic standard error: 0.012967

z-value: 31.671, p-value: < 0.000000000000000222

wald statistic: 1003, p-value: < 0.000000000000000222

Log likelihood: -3116.711 for lag model

ML residual variance (sigma squared): 0.21621, (sigma: 0.46499)

Number of observations: 4636

Number of parameters estimated: 21

AIC: 6275.4, (AIC for lm: 7301.1)

LM test for residual autocorrelation

test value: 124.83, p-value: < 0.000000000000000222

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com p -valores significativos a 5% foram realçadas em amarelo.

Figura 18: Resultados do modelo SARAR para as notas do Saeb

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.38938356551	0.12616724703	42.7162	< 0.00000000000000022
perc_esg	0.00006852194	0.00035682255	0.1920	0.84772
ad	0.00240378610	0.00041311666	5.8187	0.0000000059320653
rd	0.00043444930	0.00027004889	1.6088	0.10766
ed	0.00239337296	0.00125828556	1.9021	0.05716
atu	-0.00346515591	0.00216604702	-1.5998	0.10965
horas	0.01287727356	0.00908509640	1.4174	0.15636
pi_b_pc	0.00000024363	0.00000040090	0.6077	0.54337
hierCapital Regio-1	0.10333730115	0.04050170259	2.5514	0.01073
hierCentro de Zo-	0.04205235328	0.01990985819	2.1121	0.03467
hierCentro subregio-1	0.04939078538	0.03339916095	1.4788	0.13919
hierMetrópole	0.01574349142	0.05993783138	0.2627	0.79281
pard	-0.00098030450	0.00055871545	-1.7546	0.07933
perc_vinc	0.00763300003	0.00103352266	7.3854	0.0000000000001519
F	-0.00137298681	0.00093487124	-1.4686	0.14193
agro	-0.00152783324	0.00069393347	-2.2017	0.02769
analf	-0.00879950251	0.00486543592	-1.8086	0.07052
renda	-0.00089952649	0.00093410508	-0.9630	0.33556
dummy0	0.05169712654	0.02216656700	2.3322	0.01969

Rho: -0.10622

Asymptotic standard error: 0.018204

z-value: -5.8349, p-value: 0.000000005383

Lambda: 0.7194

Asymptotic standard error: 0.014376

z-value: 50.043, p-value: < 0.000000000000000222

LR test value: 1444.9, p-value: < 0.000000000000000222

Log likelihood: -2908.101 for sac model

ML residual variance (sigma squared): 0.17801, (sigma: 0.42191)

Number of observations: 4636

Number of parameters estimated: 22

AIC: 5860.2, (AIC for lm: 7301.1)

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.

Figura 19: Resultados do modelo MQO para as notas do Saeb

```

Coefficients:
              Estimate      Std. Error t value      Pr(>|t|)
(Intercept)    4.8523329829    0.0985815218   49.222 < 0.000000000000000002
perc_esg      -0.0000401356    0.0003758067   -0.107    0.914953
ad             0.0055759298    0.0004065217   13.716 < 0.000000000000000002
rd             0.0005671195    0.0003176350    1.785    0.074255
ed            0.0058412177    0.0015204045    3.842    0.000124
atu           -0.0185381238    0.0023737213   -7.810    0.0000000000000000704
horas         0.0108379834    0.0110461476    0.981    0.326568
pib_pc        0.0000004541    0.0000004948    0.918    0.358820
hierCapital Regio-1 -0.0214827621    0.0441957034   -0.486    0.626932
hierCentro de Zo-  0.0799577734    0.0265650320    3.010    0.002628
hierCentro subregio-1 0.0322310889    0.0459683089    0.701    0.483239
hierMetrópole   -0.1211202278    0.0451055417   -2.685    0.007273
pard          -0.0044017672    0.0005275059   -8.344 < 0.000000000000000002
perc_vinc      0.0190355500    0.0012138211   15.682 < 0.000000000000000002
F             -0.0010891963    0.0011381838   -0.957    0.338638
agro          -0.0014166739    0.0007637299   -1.855    0.063668
analf         -0.0315071777    0.0060534943   -5.205    0.00000020263918867
renda         -0.0018359994    0.0011175126   -1.643    0.100465
dummy0        0.1024719105    0.0266646019    3.843    0.000123
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5306 on 4617 degrees of freedom
Multiple R-squared:  0.2739,    Adjusted R-squared:  0.2711
F-statistic: 96.77 on 18 and 4617 DF,  p-value: < 0.000000000000000022

```

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com p -valores significativos a 5% foram realçadas em amarelo.

APÊNDICE B – RESULTADOS DOS MODELOS ECONÔMETRICOS PARA DISTORÇÃO IDADE-SÉRIE

Figura 20: Resultados do modelo SDM para a distorção idade-série

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	22.76565513902	2.76519443069	8.2329	0.000000000000000222
perc_esg	-0.01344175360	0.00678011508	-1.9825	0.047420
ad	-0.09576427458	0.00995339068	-9.6213	< 0.00000000000000022
rd	0.00783077786	0.00527148540	1.4855	0.137412
ed	0.17400224720	0.02791614458	6.2330	0.000000000457490046
atu	0.18221544267	0.03241385779	5.6215	0.000000018927468037
horas	-0.70245112987	0.19842957012	-3.5401	0.000400
piB_pc	0.00000028345	0.00000803096	0.0353	0.971845
hierCapital Regio-1	2.78237696902	0.71284438819	3.9032	0.000094927680399426
hierCentro de Zo-	-0.21939550740	0.43207567362	-0.5078	0.611614
hierCentro subregio-1	0.73545699561	0.74235150011	0.9907	0.321826
hierMetrópole	3.19357644072	0.7333236460	4.3549	0.000013313818119309
pard	-0.00086013852	0.01010428006	-0.0851	0.932161
perc_vinc	-0.12458599397	0.02011219105	-6.1946	0.000000000584513993
F	-0.03446325343	0.01864968293	-1.8479	0.064613
agro	-0.01190307716	0.01288621334	-0.9237	0.355639
analf	0.21721315983	0.09960180282	2.1808	0.029197
renda	-0.02527822666	0.01863912369	-1.3562	0.175038
dummy0	0.64819136035	0.44111080808	1.4695	0.141710
lag. perc_esg	-0.05140244096	0.01041141108	-4.9371	0.000000792825386675
lag.ad	-0.03881707617	0.01358665951	-2.8570	0.004277
lag.rd	0.04964974621	0.00807935868	6.1453	0.000000000798334732
lag.ed	0.26194061442	0.04459649929	5.8736	0.000000004265114706
lag.atu	0.05561140444	0.04759819980	1.1684	0.242665
lag.horas	0.60915018454	0.29896340229	2.0375	0.041596
lag.piB_pc	0.00001108662	0.00001341742	0.8263	0.408642
lag.hierCapital Regio-1	0.95409070826	1.24072392699	0.7690	0.441906
lag.hierCentro de Zo-	1.11472098460	0.73992680288	1.5065	0.131931
lag.hierCentro subregio-1	1.13163911289	1.27164142939	0.8899	0.373517
lag.hierMetrópole	-3.58932125529	1.21799034111	-2.9469	0.003210
lag.pard	0.00181529241	0.01386986028	0.1309	0.895870
lag. perc_vinc	-0.16614021501	0.03312354171	-5.0158	0.000000528202585226
lag.F	-0.04492597091	0.03141430983	-1.4301	0.152685
lag.agro	-0.14074672824	0.02017203086	-6.9773	0.000000000003008704
lag.analf	0.45193561122	0.19975925992	2.2624	0.023673
lag.renda	-0.04410600507	0.03049636527	-1.4463	0.148101
lag.dummy0	-0.25527001599	0.69633283136	-0.3666	0.713923

Rho: 0.42978, LR test value: 871.5, p-value: < 0.000000000000000222

Asymptotic standard error: 0.013359

z-value: 32.171, p-value: < 0.000000000000000222

wald statistic: 1035, p-value: < 0.000000000000000222

Log likelihood: -17017.2 for mixed model

ML residual variance (sigma squared): 73.76, (sigma: 8.5884)

Number of observations: 4727

Number of parameters estimated: 39

AIC: 34112, (AIC for lm: 34982)

LM test for residual autocorrelation

test value: 193.62, p-value: < 0.000000000000000222

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com p -valores significativos a 5% foram realçadas em amarelo.

Figura 21: Resultados do modelo SLX para a distorção idade-série

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	38.209569921	2.432933466	15.705	< 0.00000000000000002
perc_esg	-0.020895844	0.008342463	-2.505	0.012287
ad	-0.096002747	0.012671009	-7.577	0.00000000000000425
rd	0.019984901	0.006133350	3.258	0.001128
ed	0.194713597	0.032401469	6.009	0.0000000020029599
atu	0.105946317	0.039288809	2.697	0.007030
horas	-0.737468767	0.229318942	-3.216	0.001309
pib_pc	0.000010431	0.000009273	1.125	0.260727
hierCapital.Regio.1	1.502069810	0.925809493	1.622	0.104777
hierCentro.de.Zo.	0.336445436	0.493953513	0.681	0.495824
hierCentro.subregio.1	2.078601560	0.840004636	2.475	0.013377
hierMetrópole	1.842611947	1.518821460	1.213	0.225120
pard	-0.010364123	0.013025082	-0.796	0.426244
perc_vinc	-0.090914185	0.023742992	-3.829	0.000130
F	-0.020511580	0.021353082	-0.961	0.336807
agro	-0.005997277	0.015914848	-0.377	0.706313
analf	0.098849248	0.113320282	0.872	0.383090
renda	-0.042730884	0.021706148	-1.969	0.049057
dummy0	0.125677254	0.511487606	0.246	0.805918
lag. perc_esg	-0.079450104	0.012056720	-6.590	0.0000000000489262
lag. ad	-0.177649178	0.017335318	-10.248	< 0.00000000000000002
lag. rd	0.118100666	0.010299923	11.466	< 0.00000000000000002
lag. ed	0.759197784	0.058464665	12.986	< 0.00000000000000002
lag. atu	0.399250328	0.058407735	6.836	0.0000000000092140
lag. horas	-0.648320675	0.354866107	-1.827	0.067772
lag. pib_pc	0.000061124	0.000017842	3.426	0.000618
lag. hierCapital.Regio.1	9.629338121	1.403875634	6.859	0.0000000000078293
lag. hierCentro.de.Zo.	1.147190964	0.906442856	1.266	0.205720
lag. hierCentro.subregio.1	7.380963126	1.458048886	5.062	0.0000004302660225
lag. hierMetrópole	0.889213056	1.881561990	0.473	0.636526
lag. pard	-0.013679968	0.017270614	-0.792	0.428345
lag. perc_vinc	-0.508822602	0.042944806	-11.848	< 0.00000000000000002
lag. F	-0.047987338	0.033874618	-1.417	0.156662
lag. agro	-0.190789726	0.025111617	-7.598	0.00000000000000362
lag. analf	1.109095212	0.241005164	4.602	0.0000042958755151
lag. renda	-0.149231310	0.040914394	-3.647	0.000268
lag. dummy0	1.776072387	0.842253966	2.109	0.035022

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.651 on 4690 degrees of freedom
 Multiple R-squared: 0.4947, Adjusted R-squared: 0.4908
 F-statistic: 127.5 on 36 and 4690 DF, p-value: < 0.000000000000000022

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.

Figura 22: Resultados do modelo SEM para a distorção idade-série

```

Coefficients: (asymptotic standard errors)
              Estimate   Std. Error   z value   Pr(>|z|)
(Intercept)   32.2637839435  1.5598494404  20.6839 < 0.000000000000000022
perc_esg     -0.0121241344  0.0061611739  -1.9678  0.0490878
ad           -0.1073634444  0.0091632349 -11.7168 < 0.000000000000000022
rd            0.0028768061  0.0045669295  0.6299  0.5287461
ed            0.0978547376  0.0240724689  4.0650  0.00004803
atu           0.0968951290  0.0295621158  3.2777  0.0010466
horas        -0.5969281375  0.1735588618  -3.4393  0.0005831
pib_pc       -0.0000017110  0.0000068556  -0.2496  0.8029134
hierCapital Regio-1  0.5142662553  0.6943997385  0.7406  0.4589414
hierCentro de Zo-  0.0329607286  0.3350824440  0.0984  0.9216417
hierCentro subregio-1 0.1677717283  0.5631267027  0.2979  0.7657574
hierMetrópole  1.4436461128  1.0748101349  1.3432  0.1792189
pard         0.0086871638  0.0096751882  0.8979  0.3692492
perc_vinc    -0.0270334863  0.0176578387  -1.5310  0.1257787
F            0.0237375494  0.0158704237  1.4957  0.1347293
agro         0.0111628218  0.0118646961  0.9408  0.3467851
analf        -0.0084388063  0.0832794229  -0.1013  0.9192875
renda        0.0025699541  0.0158865023  0.1618  0.8714872
dummy0       -0.4186444221  0.3761600965  -1.1129  0.2657333

Lambda: 0.791, LR test value: 2710.4, p-value: < 0.0000000000000000222
Asymptotic standard error: 0.0097327
z-value: 81.272, p-value: < 0.0000000000000000222
wald statistic: 6605.1, p-value: < 0.0000000000000000222

Log likelihood: -16503.91 for error model
ML residual variance (sigma squared): 52.812, (sigma: 7.2672)
Number of observations: 4727
Number of parameters estimated: 21
AIC: 33050, (AIC for lm: 35758)

```

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.

Figura 23: Resultados do modelo SAR para a distorção idade-série

```

Coefficients: (asymptotic standard errors)
              Estimate   Std. Error   z value   Pr(>|z|)
(Intercept)  11.1659322975  1.3984360574  7.9846  0.000000000000001332
perc_esg     -0.0174700924  0.0053645479  -3.2566  0.0011276
ad           -0.0886431104  0.0071049362 -12.4763 < 0.000000000000000022
rd           0.0126664151  0.0043518380  2.9106  0.0036075
ed           0.1645098687  0.0232612523  7.0723  0.000000000001524114
atu          0.1329394530  0.0255577025  5.2015  0.000000197642199495
horas        -0.4588067531  0.1584929071  -2.8948  0.0037939
pib_pc       0.0000042791  0.0000068455  0.6251  0.5319075
hierCapital Regio-1  1.4512441775  0.6068026182  2.3916  0.0167740
hierCentro de Zo-  -0.2499590500  0.3653202090  -0.6842  0.4938368
hierCentro subregio-1  0.4533821484  0.6317464331  0.7177  0.4729640
hierMetrópole  1.0169760094  0.6243543359  1.6288  0.1033460
pard         -0.0026420888  0.0074746691  -0.3535  0.7237344
perc_vinc    -0.0654587320  0.0169898243  -3.8528  0.0001168
F            0.0104171678  0.0156749521  0.6646  0.5063229
agro         -0.0141496805  0.0105575462  -1.3402  0.1801663
analf        0.1107036630  0.0843060471  1.3131  0.1891437
renda        -0.0193147879  0.0155303043  -1.2437  0.2136160
dummy0      -0.1772743755  0.3672684024  -0.4827  0.6293205

Rho: 0.70763, LR test value: 2808.5, p-value: < 0.0000000000000000222
Asymptotic standard error: 0.011037
      z-value: 64.115, p-value: < 0.0000000000000000222
wald statistic: 4110.7, p-value: < 0.0000000000000000222

Log likelihood: -16454.85 for lag model
ML residual variance (sigma squared): 54.155, (sigma: 7.359)
Number of observations: 4727
Number of parameters estimated: 21
AIC: 32952, (AIC for lm: 35758)
LM test for residual autocorrelation
test value: 150.49, p-value: < 0.0000000000000000222

```

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com p -valores significativos a 5% foram realçadas em amarelo.

Figura 24: Resultados do modelo SARAR para a distorção idade-série

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	10.9854146280	1.1769517420	9.3338	< 0.00000000000000022
perc_esg	-0.0213588282	0.0045292162	-4.7158	0.0000024077584699
ad	-0.0692152262	0.0058656836	-11.8000	< 0.00000000000000022
rd	0.0156485850	0.0036329421	4.3074	0.0000165174392377
ed	0.1824557866	0.0203273526	8.9759	< 0.00000000000000022
atu	0.1114282048	0.0209669565	5.3145	0.0000001069697781
horas	-0.3859692758	0.1312222451	-2.9413	0.0032680
pib_pc	0.0000050544	0.0000061760	0.8184	0.4131301
hierCapital Regio-1	2.3154017080	0.5505267838	4.2058	0.0000260168252901
hierCentro de Zo-	0.0642360565	0.3336447480	0.1925	0.8473284
hierCentro subregio-1	0.6671841495	0.5666205197	1.1775	0.2390041
hierMetrópole	1.8378764596	0.5530600378	3.3231	0.0008902
pard	-0.0067243714	0.0057760350	-1.1642	0.2443492
perc_vinc	-0.1102096089	0.0152126876	-7.2446	0.0000000000004339
F	-0.0268612328	0.0139026485	-1.9321	0.0533478
agro	-0.0354429837	0.0091063535	-3.8921	0.0000993737309811
analf	0.2155208117	0.0765898200	2.8140	0.0048935
renda	-0.0291920547	0.0137223882	-2.1273	0.0333926
dummy0	0.2301816250	0.3114211214	0.7391	0.4598262

Rho: 0.76434

Asymptotic standard error: 0.0094281

z-value: 81.071, p-value: < 0.000000000000000222

Lambda: -0.70192

Asymptotic standard error: 0.027675

z-value: -25.363, p-value: < 0.000000000000000222

LR test value: 1763.3, p-value: < 0.000000000000000222

Log likelihood: -16977.46 for sac model

ML residual variance (sigma squared): 52.402, (sigma: 7.239)

Number of observations: 4727

Number of parameters estimated: 22

AIC: 33999, (AIC for lm: 35758)

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com p -valores significativos a 5% foram realçadas em amarelo.

Figura 25: Resultados do modelo MQO para a distorção idade-série

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	37.399408604	1.933287923	19.345	< 0.0000000000000002
perc_esg	-0.068821245	0.007685217	-8.955	< 0.0000000000000002
ad	-0.232737016	0.009722160	-23.939	< 0.0000000000000002
rd	0.041619263	0.006269012	6.639	0.00000000000352
ed	0.394719967	0.033433921	11.806	< 0.0000000000000002
atu	0.318783194	0.036667081	8.694	< 0.0000000000000002
horas	-0.859156781	0.228135640	-3.766	0.000168
pib_pc	0.000008533	0.000009863	0.865	0.387006
hierCapital Regio-1	4.931363240	0.873456785	5.646	0.0000000174065
hierCentro de Zo-	0.469226985	0.526206833	0.892	0.372591
hierCentro subregio-1	2.661514325	0.909825206	2.925	0.003458
hierMetrópole	2.324531920	0.899031123	2.586	0.009751
pard	0.033446965	0.010726407	3.118	0.001831
perc_vinc	-0.222906633	0.024418271	-9.129	< 0.0000000000000002
F	-0.021116770	0.022584009	-0.935	0.349820
agro	-0.136185322	0.015149534	-8.989	< 0.0000000000000002
analf	0.552445830	0.121378652	4.551	0.0000054626003
renda	0.009859194	0.022351205	0.441	0.659158
dummy0	-0.241279495	0.529008921	-0.456	0.648341

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.6 on 4708 degrees of freedom
 Multiple R-squared: 0.3877, Adjusted R-squared: 0.3853
 F-statistic: 165.6 on 18 and 4708 DF, p-value: < 0.00000000000000022

Fonte: Elaboração própria, a partir de dados do Inep, IBGE, SNIS, PNSB e RAIS. Nota: estimativas com *p*-valores significativos a 5% foram realçadas em amarelo.