

**A influência da regularidade docente a partir de modelos  
econométricos espaciais para Minas Gerais**

**The influence of teaching regularity from spatial  
econometric models for Minas Gerais**

**La influencia de la regularidad docente a partir de modelos  
econométricos espaciales para Minas Gerais**

**Vanessa Ragone Azevedo<sup>1</sup>**

**Resumo:** A importância da busca pelos determinantes educacionais é amplamente debatida na teoria econômica, e os resultados dos estudos nas últimas décadas do século passado indicavam apenas a relevância das características socioeconômicas do aluno e a escolaridade da mãe como fatores relevantes. Porém, a utilização de outros modelos e aplicações vêm demonstrando que características do professor, da escola e da turma são relevantes na modelagem do desempenho discente. A partir de dados do Saeb e do Censo da Educação Básica foi construída uma base de dados com o objetivo de analisar as características espaciais desta questão, assim como a influência da regularidade docente no desempenho dos estudantes do ensino fundamental (anos finais). Os resultados obtidos a partir da aplicação de métodos de georreferenciamento indicaram a forte influência das características socioeconômicas e da escolaridade materna, mas também reforçaram a importância da distorção idade-série e do parâmetro do multiplicador espacial, indicando que as notas dos alunos são influenciadas pela vizinhança. Outro resultado importante foi em relação ao impacto dos gastos governamentais em atividades de educação que indicou ser significativo em todos os modelos estimados.

**Palavras-chave:** Educação, Determinantes do Desempenho, Econometria Espacial.

---

<sup>1</sup>) Doutora em Economia pela Universidade Federal de Juiz de Fora e Professora do Departamento de Economia da Universidade Federal do Maranhão. ORCID: <https://orcid.org/0000-0001-6998-7323>; Email: [vanessa.ragone@gmail.com](mailto:vanessa.ragone@gmail.com).

**Abstract:** The importance of the search for educational determinants is widely debated in economic theory, the results of studies between the 60's and 90's indicated only the relevance of the student's socioeconomic characteristics and the mother's education as relevant factors, but the use of other models and Applications have shown that teacher, school and class characteristics are relevant in modeling student performance. From data from the Saeb and the Censo da Educação Básica, a database was built to analyze the spatial characteristics of this issue, as well as the influence of teacher regularity on the performance of elementary school students (final years). The results obtained from the application of georeferencing methods indicated the strong influence of socioeconomic characteristics and maternal education, but also indicated the importance of age-grade distortion and spatial multiplier parameter, indicating that students' grades are influenced by neighborhood. Another important result was in relation to the impact of government spending on education activities which indicated to be significant in all estimated models.

**Keywords:** Education, Educational Determinants, Spatial Econometric.

**Resumen:** La importancia de la búsqueda de determinantes educativos es ampliamente debatida en la teoría económica, los resultados de estudios en las últimas décadas del siglo pasado solo indicaron la relevancia de las características socioeconómicas del estudiante y la educación de la madre como factores relevantes. Sin embargo, el uso de otros modelos y aplicaciones ha venido demostrando que las características del docente, la escuela y la clase son relevantes para modelar el desempeño de los estudiantes. A partir de datos de la Saeb y del Censo de Educación Básica, se construyó una base de datos con el objetivo de analizar las características espaciales de este tema, así como la influencia de la regularidad docente en el desempeño de los alumnos de la enseñanza básica (últimos años). Los resultados indicaron la fuerte influencia de las características socioeconómicas y la educación materna, pero también reforzaron la importancia de la distorsión edad-grado y el parámetro multiplicador espacial, indicando que las calificaciones de los estudiantes están influenciadas por el barrio. Otro resultado importante estuvo relacionado con el impacto del gasto público en actividades educativas, que indicó ser significativo en todos los modelos estimados.

**Palabras clave:** Educación, Determinantes Educativos, Econometría Espacial.

## Introdução

O desenvolvimento econômico está ligado a melhora do capital humano local. Por isso, um sistema educacional qualificado está entre os principais objetivos de um governo em busca de crescimento e desenvolvimento econômico. A busca por melhora no sistema educacional está ligada aos benefícios que a educação proporciona no campo individual e no campo social. Entre os benefícios individuais, destaca-se a capacidade de geração de renda futura, aumento da expectativa de vida (Murnane et al, 1995; Menezes- Filho, 2001; Murphy e Peltzman, 2004).

A partir do entendimento da importância da educação e de investimentos nessa área tanto na perspectiva individual como do desenvolvimento econômico, a literatura há anos esforça-se para identificar os determinantes da educação, utilizando como proxy o desempenho escolar para mensurar a qualidade da educação. Coleman et al. (1966) encontra resultados que indicam que cerca de 80% do desempenho escolar pode ser explicado pelo ambiente familiar. Trabalhos posteriores, inclusive aplicados a dados brasileiros, encontram resultados semelhantes ao estudo de Coleman et al. (1966), reforçando a importância do background familiar no desempenho acadêmico (Albernaz, Ferreira e Franco, 2002; Soares, 2005; Menezes-Filho, 2007; Machado et al., 2008).

Com o surgimento de novos modelos e formas de análise, diversos estudos indicaram que quando controlada as características do aluno e do ambiente familiar, as características da escola e do contexto escolar possuem influência significativa no desempenho dos alunos. Diversos estudos aplicando modelos hierárquicos que permitem o controle e análise de diferentes níveis (escola, turma, professor, aluno) obtiveram resultados demonstrando que a influência do contexto escolar, do desempenho de alunos próximos e nível socioeconômico da turma possuem influência nos resultados educacionais a nível individual (Felício e Fernandes, 2005; Alves e Soares, 2007; Riani e Rios-Neto, 2008).

Vernier, Bagolon e Jacinto (2015) utilizam uma análise mais agregada, introduzindo além das características socioeconômicas dos alunos, as características dos professores, diretores e da escola, encontram impacto para todas as dimensões, sendo significativos a questões ligadas à região, de modo que a localização importa quando pretende-se analisar desempenho escolar.

Utilizando como base a teoria desenvolvida por Fujita, Krugman e Venables (1999), é possível entender como a localização possui um papel importante e o efeito proximidade torna-se mais intenso à medida que as trocas de informações e conhecimento são mais fluídas e eficazes.

Considerando que o professor em sala de aula possui papel significativo no desempenho de seus alunos, analisar a relação entre o índice de regularidade docente (isto é, índice que possibilita analisar lugares que possuem menor fluxo de professores nos últimos 5 anos) e sua relação com a proficiência discente é importante para entender como a troca de professores, ou seja, alterações no corpo docente podem surtir efeito no desempenho dos alunos.

O presente artigo possui o objetivo geral de fazer uma análise do impacto de capital humano a nível municipal para o estado de Minas Gerais, analisando a proficiência escolar, variável proxy para qualidade de ensino e o aspecto espacial desta variável, assim como a

regularidade docente afeta o rendimento escolar. O presente artigo propõe uma investigação do transbordamento do capital humano baseado na abordagem utilizada por Ertur e Koch (2007), em que o estoque de conhecimento acumulado em um país depende do estoque de conhecimento dos países vizinhos. A partir de um modelo de interdependência tecnológica, os resultados obtidos indicam que o estoque de conhecimento de um país gera transbordamentos que ultrapassam as fronteiras e afetam os países vizinhos, sendo tal efeito decrescente conforme a distância geográfica aumenta.

A partir desse olhar para o espaço geográfico e como ele influencia o desenvolvimento e questões de acordo com seus limites, pretende-se responder a seguinte questão: Como as externalidades educacionais medidas pelos gastos em educação, regularidade docente e esforço docente impactam no desempenho escolar? A hipótese adotada inicialmente é que existe impacto das externalidades educacionais sobre o desempenho discente.

Sendo assim, o objetivo geral deste estudo é investigar o impacto das externalidades educacionais a partir de modelos espaciais. Para atender tal demanda, em termos de objetivos específicos, serão utilizados modelos de econometria espacial que auxiliam nos estudos de externalidades e permitindo verificar se características institucionais, econômicas, culturais e sociais de determinada região possuem efeito transbordamento para as regiões vizinhas. Diversos estudos que utilizam econometria espacial analisam como desempenho e crescimento econômico são influenciados pela vizinhança (Easterly e Levine, 1995; Moreno e Trehan, 1997; Silveira Neto, 2001; Hewings, Magalhães e Azzoni; 2005), outros estudos analisam como os gastos e investimentos na área educacional afetam os vizinhos (Case e Rosen, 1993).

Os resultados encontrados avançam na literatura em questão ao utilizar o modelo de autocorrelação espacial dos dados para verificar se há influência na regularidade docente. Apesar da variável regularidade docente apresentar-se como não significativa nos modelos analisados, os demais resultados demonstram compatibilidade com os achados já consolidados na literatura sobre o tema. Torna-se relevante, portanto, a investigação de variáveis que captem o efeito da relação professor-aluno para elaboração de políticas públicas efetivas na melhoria do desempenho.

O estudo está estruturado em 6 seções. Além dessa introdução, a próxima seção descreve os dados utilizados. Nas seções 3 e 4, são apresentados a análise exploratória de dados espaciais e a estratégia empírica para estimação dos modelos. A sexta seção apresenta os resultados obtidos a partir das estimações dos modelos, e, por fim, são apresentadas as considerações finais do estudo.

## Base de Dados e Estratégia Empírica

Os dados utilizados no presente estudo são de duas fontes distintas. o primeiro conjunto de dados é proveniente dos indicadores educacionais disponíveis no site do INEP para o ano de 2017, no caso específico deste estudo serão analisadas apenas as escolas públicas.

De acordo com o INEP (2020), os indicadores sociais proporcionam atribuir valores estatísticos à qualidade de ensino, atendo-se não apenas a proficiência discente, mas também ao contexto socioeconômico em que as escolas estão inseridas.

Além disso, dados sobre os gastos per capita com atividades de educação e renda per capita do mercado formal foram dados obtidos a partir do site do Índice Mineiro de Responsabilidade Social (IMRS) da Fundação João Pinheiro que possui diversos indicadores para todas as cidades de Minas Gerais.

Na tabela 1 estão descritos os indicadores utilizados no presente artigo, assim como uma breve descrição e a fonte dos dados em questão, todos os dados estão à nível municipal.

Tabela 1 – Descrição das variáveis de interesse.

Variáveis	Ano	Descrição	Fonte
Nota Padronizada	2017	Nota padronizada do Saeb dos anos finais do ensino fundamental.	Saeb
Escolaridade da mãe	2017	Proporção de alunos que possuíam mãe com ensino médio ou superior completo.	Saeb
Cor	2017	Proporção de alunos brancos.	Saeb
Sexo	2017	Proporção de alunos do sexo masculino.	Saeb
Gastos em Educação	2018	Gasto per capita com atividades de educação.	IMRS
Renda per capita	2018	Renda média per capita do mercado formal.	IMRS
Regularidade docente (> 3)	2018	Média de anos que os docentes atuaram na escola nos últimos 5 anos (Varia de 0 a 5).	Censo da Educação Básica
Distorção Idade-Série	2018	Taxa de Distorção Idade-Série.	Censo da Educação Básica
Média Hora/Aula Diária	2018	Média de Horas Aula diária	Censo da Educação Básica
Esforço docente (<=3)	2018	Quanto menor o indicador de esforço, menor a carga horária do docente (Varia de 1 a 6).	Censo da Educação Básica
Média de Aluno/Turma	2018	Média de alunos por turma	Censo da Educação Básica

Fonte: Elaborada pela autora

Todos os dados foram coletados para os 853 municípios de Minas Gerais. O Sistema de Avaliação da Educação Básica – SAEB – é um sistema de avaliação das escolas brasileiras realizado pelo INEP. O exame analisa o desempenho dos alunos nas disciplinas de matemática e português. Além das provas, são aplicados questionários para os diretores, professores e alunos, permitindo uma avaliação mais ampla das escolas.

O desempenho escolar municipal, obtido através da nota padronizada nas disciplinas de

português e matemática, será a variável dependente deste estudo. Os dados sobre cor, sexo e escolaridade da mãe foram introduzidas no modelo para não haver problemas de variável importante omitida, uma vez que características socioeconômicas e a escolaridade da mãe são variáveis com importância significativa no desempenho escolar.

### **Análise Exploratória dos Dados Espaciais**

A aplicação de um modelo de regressão espacial deve ser precedida de uma investigação da existência de dependência espacial. De acordo com Gibbons e Overman (2012) é importante se atentar a necessidade da aplicação de regressões espaciais dada a questão de pesquisa e não somente para o tratamento de problemas de endogeneidade. Assim, o presente artigo foca em uma abordagem experimentalista que possui maior enfoque nas questões de identificação e causalidade, assim como é indicado pelos autores.

Vernier, Bagolin e Fochezatto (2017) destacam que a localização assume um papel importante, e o efeito proximidade torna-se mais intenso à medida que as trocas de informações e conhecimentos são mais fluídas e eficazes. Os modelos espaciais permitem estudar as externalidades, possibilitando, assim, identificar as variáveis institucionais, econômicas, culturais e sociais de determinada região têm efeitos de transbordamento para as regiões vizinhas.

A partir da ideia deste efeito transbordamento é necessário realizar a análise exploratória de dados espaciais (AEDE) que permite avaliar se há algum padrão de associação espacial entre as regiões. Como a variável de interesse nota padrão do ensino fundamental nos anos finais possui regiões sem informação (missing values), para realizar a AEDE foram removidos da análise os seguintes municípios: Fervedouro, São Geraldo do Baixo, Santana do Manhuaçu, Pedra do Anta e Lagoa Grande.

### **Determinação da Matriz de Pesos**

O primeiro passo para a implementação da AEDE é a construção de uma matriz de pesos espaciais, denominada matriz “W”. Apesar de existirem diversas formas possíveis de construção de matrizes (matriz socioeconômica, matriz transacional, etc.), Rêgo e Pena (2012) citam que majoritariamente os trabalhos utilizam quatro tipos de matrizes: torre (rook), rainha (queen), distância/distância inversa e “k vizinhos mais próximos”.

Como o objetivo do presente artigo é analisar o transbordamento do efeito educacional a partir das interações entre municípios de Minas Gerais, torna-se necessário especificar como os municípios estão conectados. A matriz de pesos espaciais deve ser escolhida de acordo com a estrutura da amostra com o objetivo de garantir que não existam “ilhas” – regiões sem vizinhança, que possuam pesos não negativos e finitos e pesos exógenos.

O uso da matriz distância, apesar de permitir a identificação da proximidade dos municípios, não possibilita verificar a existência de fronteira entre eles. E, uma vez que os tamanhos dos municípios brasileiros não são homogêneos, o uso de matriz de ponderação com base na distância ou contiguidade pode gerar uma estrutura não equilibrada. Uma solução comum para este problema consiste em considerar as matrizes de pesos com base nos vizinhos mais próximos, pois, dessa forma, forçaria cada unidade ter o mesmo número de vizinhos (Anselin, 2002; Dominicus et. al. 2013).

Apesar da escolha da matriz ser baseada em critérios teóricos é importante ressaltar um importante resultado encontrado por LeSage e Pace (2014) que concluiu que a especificação da matriz de pesos espaciais não interfere nos fatores dos modelos de regressão espacial.

Com o intuito de especificar de que ordem será a matriz de vizinhança, será utilizado o critério de Almeida (2012). De acordo com este critério, após testar o I de Moran para um conjunto de matrizes, seleciona-se a matriz que tenha gerado o valor mais alto, e que seja estatisticamente significativo. O coeficiente do I de Moran conforme o número de vizinhos é apresentado na tabela a seguir.

Tabela 2 – I de Moran para matrizes de k vizinhos mais próximos

Variáveis	Coeficiente do I de Moran		
	5 vizinhos	10 vizinhos	15 vizinhos
Nota Padronizada (anos finais ensino fundamental)	0,387	0,364	0,362
Escolaridade da mãe	0,313	0,280	0,255
Cor	0,770	0,747	0,738
Sexo	0,008	0,004	0,005
Gastos em Educação	0,112	0,094	0,093
Renda per capita (mercado formal)	0,238	0,202	0,203
Regularidade docente (média/alta)	0,131	0,118	0,108
Distorção Idade-Série	0,338	0,290	0,270
Média Hora/Aula Diária	0,111	0,096	0,086
Esforço docente (1+2+3)	0,080	0,060	0,070
Média de Aluno/Turma	0,180	0,148	0,133

Fonte: Elaborado pela autora.

A partir dos dados apresentados pela tabela 2 é possível definir a matriz de pesos espaciais com os três vizinhos mais próximos (k=5) como a matriz a ser utilizada para a AEDE e futuras estimações.

## Método de Análise de Autocorrelação Espacial dos Dados

A dependência espacial surge quando o valor da variável em um determinado local depende do valor dessa variável em regiões vizinhas, isto é, se os dados se distribuem ou não de forma aleatória no espaço. O *I de Moran* é um método que permite verificar e fazer inferência à dependência espacial.

Nesse estudo, inicialmente serão empregadas as estatísticas que permitem a identificação de padrões espaciais: *I de Moran* global e local. O *I de Moran* global oferece um sumário da distribuição espacial dos dados, isto é, um único valor (médio) para todas as regiões. O *I de Moran* Local (LISA), ao calcular um valor para cada unidade de observação, permite identificar diferentes padrões de distribuição espacial (*clusters* ou *outliers*). De acordo com Almeida (2012), o *I de Moran* global pode ser definido matricialmente como:

$$I = \left( \frac{n}{S_0} \right) \left( \frac{z'Wz}{z'z} \right) \quad (1)$$

Em que  $n$  é o número de regiões,  $z$  indica os valores da variável de interesse padronizada,  $Wz$  representa os valores médios da variável de interesse padronizada nos vizinhos, definidos segundo uma matriz de ponderação espacial  $W$ .  $S_0$  é igual à operação  $\sum \sum w_{ij}$ , significa que todos os elementos da matriz de pesos espaciais  $W$  devem ser somados.

A hipótese nula sendo testada é a de aleatoriedade espacial. Quando o valor calculado excede o valor esperado (dentro do limite da significância estatística) indica que existe autocorrelação espacial positiva (similaridade) e valores abaixo do esperado indicam autocorrelação espacial negativa (dissimilaridade).

A verificação dos padrões locais e a determinação das regiões que mais contribuem para a autocorrelação espacial podem ser feitas por meio do emprego do *LISA*. Esse indicador foi inicialmente sugerido por Anselin (1995), e pode ser calculado da seguinte forma:

$$I_l = \frac{y_j \sum_{j=1}^n w_{ij} y_i}{\sum_{i=1}^n y_i^2} \quad (2)$$

Em que  $n$  indica o número de regiões,  $w_{ij}$  são os elementos da matriz de pesos espaciais,  $y_i$  e  $y_j$  são os valores da variável utilizada, enquanto  $i$  e  $j$  referem-se às diferentes localidades.

Um instrumento adicional para análise da autocorrelação espacial é baseada no diagrama de dispersão de Moran, que mostra a defasagem espacial da variável de interesse no eixo vertical e o valor da variável de interesse no eixo horizontal. Esse tipo de diagrama fornece evidências do tipo de autocorrelação linear (negativa ou positiva). Quando o coeficiente angular

é negativo, existem evidências de autocorrelação espacial negativa (presença de *outliers*) e coeficiente angular positivo, autocorrelação espacial positiva (presença de *clusters*).

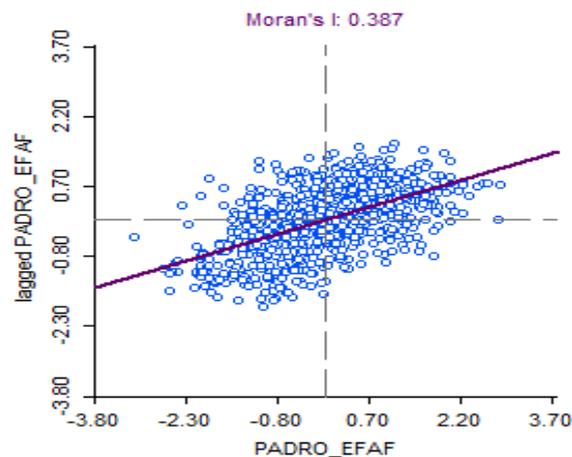
Além disso, este diagrama fornece informações sobre os tipos de associação linear espacial, de acordo com a distribuição dos dados nos quatro quadrantes, podendo haver associações do tipo Alto-Alto (AA), Baixo-Baixo (BB), Alto-Baixo (AB) e Baixo-Alto (BA).

A partir desses métodos é possível fazer a análise exploratória dos dados espaciais, entendendo como os dados se distribuem no espaço e a presença ou não de padrões e estudar o melhor modelo a ser aplicado juntamente com o arcabouço teórico.

## Resultados da Análise Exploratória de Dados Espaciais

Os resultados da tabela 2 que indicam o I de Moran para o teste de autocorrelação espacial univariada indicam que todas as variáveis (exceto a variável sexo) apresentaram resultados positivos e significativos, a partir disso, rejeitamos a hipótese nula de aleatoriedade espacial, ou seja, existe autocorrelação espacial nos dados para as variáveis investigadas a nível municipal em Minas Gerais. O diagrama de dispersão de Moran, ilustrado no gráfico 1, reforça os resultados indicados pelo I de Moran global para a variável dependente.

Gráfico 1 – Diagrama de dispersão de Moran para a Nota padronizada do Ensino Fundamental – Anos Finais.

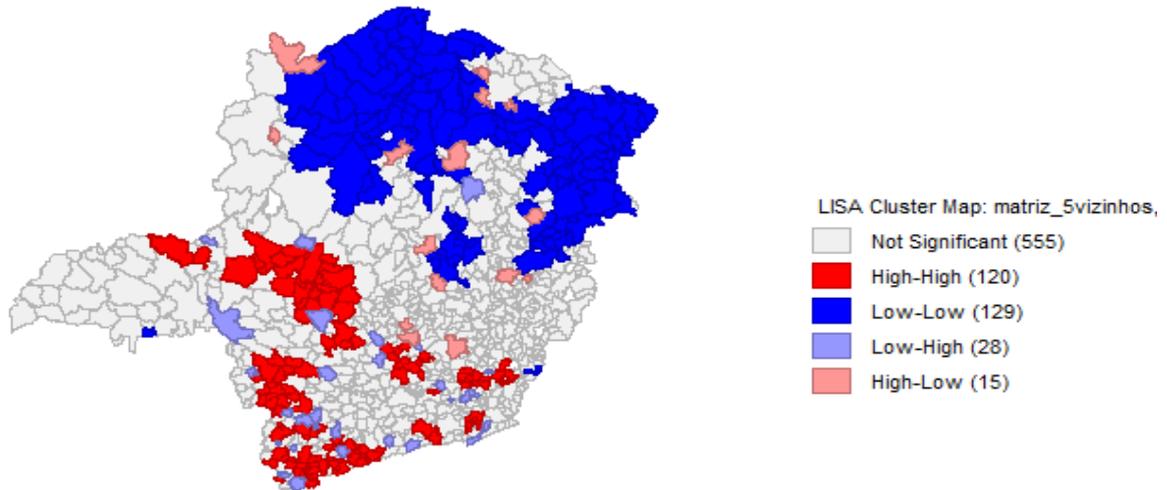


Fonte: Elaboração Própria.

Considerando que o I de Moran foi positivo para todas as variáveis, é possível concluir que existe um padrão de concentração espacial, indicando que municípios que possuem indicadores com altos valores têm vizinhos com indicadores altos e municípios com baixos indicadores possuem vizinhos com baixos indicadores.

Com o auxílio do I de Moran local, é possível analisar melhor os padrões espaciais locais. Abaixo na figura 1 encontra-se o mapa de clusters LISA com base no I de Moran local univariado a 5% de significância com 999999 permutações para a nota padronizada do Ensino Fundamental – Anos Finais do Saeb.

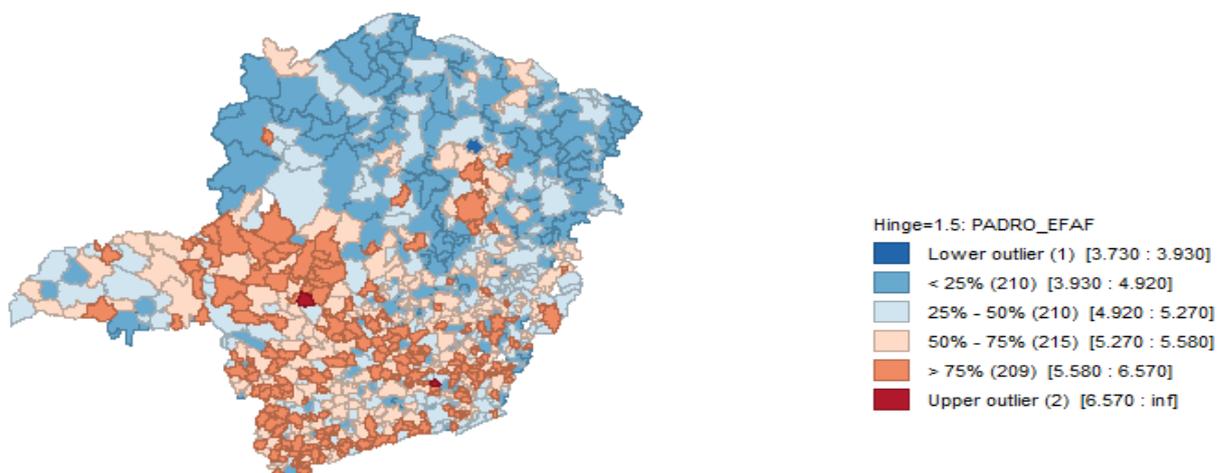
Figura 1 – LISA para Nota Padronizada do Ensino Fundamental – Anos Finais



Fonte: Elaboração Própria

Apesar do I de Moran ter sido significativo e positivo, existem algumas regiões que fogem ao padrão notas altas com vizinhos com notas altas e notas baixas com vizinhos com notas baixas. As regiões determinadas pelas cores azul claro e vermelho claro indicam, respectivamente, regiões que possuem notas baixas e que os vizinhos possuem notas altas e regiões com notas altas com vizinhos com notas baixas. Outro padrão perceptível é que existe uma divisão norte e sul no mapa, com maior concentração de notas altas na metade inferior do mapa e concentração de notas baixas na parte superior do mapa.

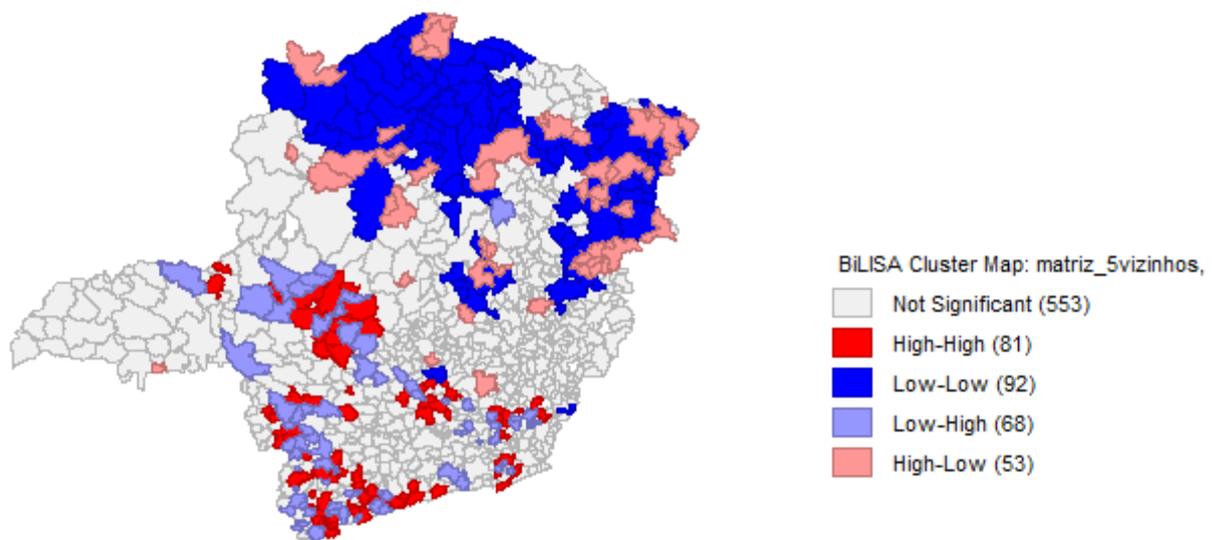
Figura 2 – Box Map utilizando hinge 1,5 para Nota Padronizada do Ensino Fundamental - Anos Finais



Fonte: Elaboração Própria

Analisando o mapa de outliers a partir da figura 2 a divisão norte e sul fica mais discrepante. Utilizando um hinge de 1,5, ou seja, multiplicando o intervalo interquartilico uma vez e meia, é identificado apenas um município caracterizado como outlier inferior (na região norte de MG) e 2 municípios caracterizados como outliers superiores (na região sul de MG). Este resultado foi obtido após o tratamento das regiões que não possuíam valores para a variável de interesse nota padronizada do ensino fundamental anos finais. Com o objetivo de verificar a existência de autocorrelação espacial entre a variável de regularidade docente e de desempenho escolar será analisado o mapa de clusters bivariado a partir do I de Moran bivariado das variáveis regularidade docente e nota padronizada.

Figura 3 – BLISA: Mapa de Clusters Bivariado para Regularidade Docente e Nota Padronizada do Ensino Fundamental - Anos Finais



Fonte: Elaboração Própria

O mapa de clusters bivariado possibilita analisar a associação entre a regularidade docente e a nota padronizada do Saeb dos municípios vizinhos.

A predominância de clusters foi do tipo baixo-baixo, indicando que cidades que possuem baixo índice de regularidade docente estão rodeadas de vizinhos com notas padronizadas mais baixas; seguido pelos clusters alto-alto que indicam que municípios que possuem altos índices de regularidade estão cercados de vizinhos com altas notas e os clusters do tipo baixo-alto que indicam municípios com baixo índice de regularidade rodeados de vizinhos com altas notas e alto-baixo indicam municípios com alta regularidade e nota baixa nos vizinhos.

## Estratégia Empírica

A partir da Análise Exploratória de Dados Espaciais e mensurada a heterogeneidade espacial, deve-se incluir a dependência espacial no modelo a ser estudado. Geralmente, o ponto de partida dos modelos de análise econométrica é o modelo clássico de regressão linear, estimado por mínimos quadrados ordinários (MQO). No entanto, Ertur *et al.* (2006) sugere que a dependência ou heterogeneidade espacial pode levar a estimações de MQO não confiáveis, devido à possibilidade de heterocedasticidade gerada por alterações nos coeficientes ou na variância do erro entre as observações.

Assim, para considerar e explorar a natureza espacial do problema em questão devemos considerar o uso de modelos econométrico-espaciais. Partindo da equação 3:

$$\begin{aligned}y_i &= \alpha + \beta X_i + \rho W_{ij} y_j + \delta_i \\ \delta_i &= \lambda W_{ij} \delta_j + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2)\end{aligned}\tag{3}$$

em que  $y_i$  ( $y_j$ ) é a variável explicada na região  $i$  ( $j$ ),  $\alpha$  é o intercepto,  $X$  é a matriz  $n \times k$  de variáveis explicativas,  $\beta$  é o vetor  $k \times 1$  de coeficientes,  $\rho$  é o parâmetro relacionado com a defasagem espacial da variável explicada,  $\lambda$  é o parâmetro de variância do ruído,  $W$  é a matriz de pesos espaciais  $n \times n$ , com  $W_{ij} > 0$ , quando a região  $j$  é vizinha da região  $i$ .

Quando os valores dos parâmetros  $\rho$  e  $\lambda$  são alterados, obtém-se diferentes modelos. No caso em que não exista dependência espacial, nem na variável dependente ou nos distúrbios ( $\rho=0$  e  $\lambda=0$ ), o modelo seria o tradicional de MQO, podendo ser representado da seguinte forma:

$$y_i = \alpha + \beta X_i + \delta_i\tag{4}$$

Em casos que  $\rho \neq 0$  e  $\lambda = 0$ , o modelo a ser estimado é o *Spatial Autoregressive* (SAR). Assim, a dependência espacial é incluída no modelo por meio dos valores espacialmente defasados da variável dependente, como descrito na equação a seguir:

$$y_i = \alpha + \beta X_i + \rho W_{ij} y_j + \delta_i\tag{5}$$

Outro modelo possível é o *Spatial Error Model* (SEM), por sua vez, reflete a dependência espacial nos resíduos ( $\rho=0$  e  $\lambda \neq 0$ ). Essa especificação indica que um choque aleatório introduzido em determinada região afeta as demais através da estrutura espacial. O modelo pode ser especificado da seguinte maneira:

$$\begin{aligned}y_i &= \alpha + \beta X_i + \delta_i \\ \delta_i &= \lambda W_{ij} \delta_j + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2)\end{aligned}\tag{6}$$

Segundo Anselin (1988), a estimação mais apropriada para esses modelos é a de máxima verossimilhança ou a de variáveis instrumentais, dado que as estimações via MQO geram resultados viesados e inconsistentes devido à simultaneidade na natureza de autocorrelação causada pela defasagem espacial.

Além dos modelos usualmente utilizados na econometria espacial, também existem os modelos SLX ou Modelo Regressivo Cruzado Espacial que possui apenas a matriz de pesos nas explicativas, não possuindo multiplicador espacial. Outro modelo é o SDM ou Modelo Durbin Espacial que pode ser descrito de acordo com a equação 7:

$$y_i = \alpha + \beta X + WX\tau + \rho W_{ij}y_j + \delta_i \quad (7)$$

O Modelo SDM é muito utilizado em estudos que analisam o impacto educacional devido a sua característica de capturar o alcance local através de  $WX$  e o alcance global através do  $Wy$ . De acordo com a literatura este seria o modelo mais indicado para a análise do transbordamento educacional.

## Modelo Aplicado

Com o objetivo de identificar o impacto da regularidade docente na proficiência discente e o efeito transbordamento da educação será estimado um modelo que considera as características espaciais das variáveis e os principais determinantes da proficiência para garantir que não haverá viés de variável omitida, construindo estimadores confiáveis. O modelo base, teórico, será construído pela seguinte equação, que também será analisada por MQO:

$$\begin{aligned} edu(Nota) = & \beta_0 + \beta_1 RegDocente + \beta_2 Cor + \beta_3 Sexo + \beta_4 EscMãe + \beta_5 DistIdSer + \beta_6 Renda \\ & + \beta_7 HAula + \beta_8 AlunoTurma + \beta_9 EsforçoDoc + \beta_{10} GEduc + \rho_i \end{aligned}$$

em que  $edu(Nota)$  é a nota padronizada do Saeb no ensino fundamental (anos finais) de cada município,  $RegDocente$  é o índice de regularidade docente por município,  $EscMãe$  é a escolaridade média das mães dos alunos por município,  $DistIdSer$  é a distorção idade-série por município para os anos finais do ensino fundamental,  $HAula$  é a média de horas aula diária por município para os anos finais do ensino fundamental,  $AlunoTurma$  é a média de alunos por turma em cada município para os anos finais do ensino fundamental,  $EsforçoDoc$  é a média de escolas que possuem docentes com esforço no nível 1, 2 e 3 e, por fim,  $GEduc$  é a média dos gastos em atividades de educação por município. Os demais modelos com defasagens

espaciais serão especificados ao decorrer do desenvolvimento dos resultados para melhor compreensão e interpretação dos modelos.

## Resultados

O primeiro modelo a ser testado foi o Modelo de Mínimos Quadrados Ordinários, de acordo com a teoria, quando se utiliza um modelo de MQO para analisar modelos que deveriam utilizar econometria espacial são geradas estimativas não confiáveis, devido à possibilidade de heterocedasticidade gerada por alterações nos coeficientes ou na variância do erro entre as observações. A tabela 3 possui os resultados obtidos a partir da aplicação do modelo de MQO.

Tabela 3 – Estimativas do modelo utilizando MQO.

Variáveis	Coefficiente
Regularidade Docente	0,0005
Cor	1,8194***
Sexo	-0,4989**
Escolaridade da Mãe	0,5091***
Distorção Idade Série	0,0040**
Renda	0,0001*
Média Hora Aula	-0,0275
Média Aluno por Turma	-0,0029
Esforço Docente	0,0002
Gasto em Educação	0,0001*
Constante	4,8094***

Fonte: Elaboração própria. Notas: \*, \*\*, \*\*\* denotam valor-p menor do que 10%, 5% e 1%.

O modelo MQO demonstra que apenas as variáveis que denotam características socioeconômicas (cor, sexo, escolaridade da mãe) e as variáveis distorção idade série, renda e gasto em educação foram significativas no modelo. Além disso, o modelo de MQO indica que o sexo masculino está negativamente associado as notas padronizadas do Saeb, já as demais variáveis significativas no modelo estão associadas positivamente as notas padronizadas do ensino fundamental. A seguir serão analisados os modelos clássicos de análise espacial, modelos SAR, SEM e SAC que respectivamente consideram: a defasagem espacial da variável dependente, a defasagem espacial nos erros e a defasagem na dependente e no erro, simultaneamente.

Tabela 4 – Resultados dos Modelos SAR, SEM e SAC.

Variáveis	SAR	SEM	SAC
Regularidade Docente	0,0004	0,0003	0,0002
Cor	1,6334***	1,5737***	1,3576***
Sexo	-0,5088**	-0,4644*	-0,3890*
Escolaridade da Mãe	0,7584***	0,7201***	0,7464***
Distorção Idade Série	0,0013	0,0023	0,0041**
Renda	0,0000	0,00003	0,0000
Média Hora Aula	0,0432	0,0483	0,06980
Média Aluno por Turma	-0,0039	-0,00424	-0,0054
Esforço Docente	0,0009	0,0008	0,0008
Gasto em Educação	0,0001**	0,0001**	0,0001**
Constante	4,1276***	4,2180***	3,9578***
$\rho$	0,0860***		0,1035***
$\lambda$		1,0478***	3,0046***
Log likelihood (LIK)	-410,3891	-407,8721	-395,9727
Akaike info criterion (AIC)	421,93	474,29	419,38

Fonte: Elaboração própria. Notas: \*, \*\*, \*\*\* denotam valor-p menor do que 10%, 5% e 1%.

Os resultados da tabela 4 indicam que os coeficientes dos modelos econométrico espaciais clássicos SAR, SEM e SAC obtiveram coeficientes muito próximos aos do modelo de Mínimos Quadrados Ordinários, sendo a regularidade docente não significativa em todos eles. O principal destaque é para  $\rho$  e  $\lambda$  que são estatisticamente significantes nos modelos espaciais, indicando que a nota dos alunos está correlacionada com a nota dos vizinhos e que parâmetro de erro autorregressivo nos modelos SEM e SAC também é estatisticamente significativo. Cabe a análise dos impactos diretos, indiretos e totais do modelo SAR e SAC, no caso do modelo SEM por não haver multiplicador espacial, não há Wy, não existe impacto direto apenas indireto sendo igual ao total.

Tabela 5 – Impactos direto, indireto e total dos Modelos SAR e SAC

Variáveis	Modelo SAR			Modelo SAC		
	Direto	Indireto	Total	Direto	Indireto	Total
Regularidade Docente	0,0004	0,0000	0,0004	0,0002	0,0000	0,0002
Cor	1,6338***	0,1407***	1,7745***	1,3576***	0,1434***	1,5011***
Sexo	-0,5088**	-0,0439**	-0,5527**	-0,3890*	-0,0411	-0,4301*
Escolaridade da Mãe	0,7585***	0,0653***	0,8638***	0,7464***	0,0789***	0,8253***
Distorção Idade Série	0,0013	0,0001	0,0015	0,0041**	0,0004*	0,0045**
Renda	0,000	0,0000	0,0001	0,0000	0,0000	0,0000
Média Hora Aula	0,0432	0,003725	0,4697	0,0698	0,0074	0,7718
Média Aluno/Turma	-0,0039	-0,0003	-0,00428	-0,0054	-0,0006	-0,0059
Esforço Docente	0,0009	0,0001	0,0009	0,0008	0,0001	0,0008
Gasto em Educação	0,0001**	0,00001**	0,0002**	0,0001**	0,0001*	0,0001**

Fonte: Elaboração própria. Notas: \*, \*\*, \*\*\* denotam valor-p menor do que 10%, 5% e 1%.

O efeito transbordamento é dado pelo impacto indireto que indica em ambos os modelos que as variáveis que controlam as características socioeconômicas dos alunos possuem impacto

direto, indireto e total significativos (cor, sexo, escolaridade da mãe), nota-se que a predominância do sexo masculino está associada a notas mais baixas, o que já é esperado de acordo com a literatura.

Além disso, apenas no modelo SAC a variável distorção idade-série apresentou os impactos significativos, sendo apenas o direto e total significativos à 5%. A única variável introduzida no modelo clássico de determinante de desempenho que apresentou impactos significativos em ambos os modelos descritos na tabela 5 foi o gasto em educação, apresentando impacto direto, indireto e total significativos.

A seguir serão analisados os modelos mais indicados para tratar este tipo questão na literatura: modelos SDM e SLX. Ambos os modelos introduzem defasagem espacial nas variáveis explicativas sendo que o modelo SLX não possui  $W_y$ , logo não possui multiplicador espacial, sendo um modelo mais simples e o modelo SDM que possui multiplicador espacial e defasagem das explicativas. No caso do modelo de determinantes do desempenho em questão, não foram defasadas espacialmente as variáveis esforço docente e sexo por não apresentarem autocorrelação espacial.

Tabela 6 – Resultados dos Modelos SDM e SLX

Variáveis	SDM	SLX
Regularidade Docente	0,0004	0,0004
Cor	1,3748***	1,3994***
Sexo	-0,4239*	-0,4454*
Escolaridade da Mãe	0,6984***	0,6973***
Distorção Idade Série	0,0040*	0,0036*
Renda	0,0000	0,0000
Média Hora Aula	0,0672	0,0501
Média Aluno por Turma	-0,0054	-0,0055
Esforço Docente	0,0005	0,0004
Gasto em Educação	0,0001**	0,0001**
Constante	4,0802***	4,1087***
WRegularidade	0,0086	0,0107
WCor	-1,3630	-0,1505
WEscolaridade Mãe	-7,7446*	-7,8138**
WDistorção Idade-Série	-0,1093***	-0,1085***
WRenda	0,0018	0,0021
WMédia Hora Aula	0,6308	1,4917**
WMédia Aluno por Turma	-0,1274*	-0,1371*
WGasto em Educação	0,0010	0,0013
$\rho$	0,8045***	
Log likelihood (LIK)	-397,0511	-400,8238
Akaike info criterion (AIC)	476,23	451,04

Fonte: Elaboração própria. Notas: \*, \*\*, \*\*\* denotam valor-p menor do que 10%, 5% e 1%.

Analisando os resultados da tabela 6, modelos SDM e SLX, os coeficientes encontrados se aproximam muito aos dos demais modelos, porém como estes modelos consideram a

defasagem espacial das explicativas é possível notar que a variável WDistorção idade-série é significativa em ambos os modelos à 1%, a variável WMédia Aluno por Turma também é significativa em ambos os modelos, além disso, no modelo SDM o  $\rho$  permanece significativo à 1%. Abaixo a tabela com os impactos dos modelos SDM e SLX (sendo o impacto do modelo SLX composto por impacto direto  $\beta$ , impacto indireto  $\tau$  e o total igual a  $\beta + \tau$ ).

Tabela 7 – Impactos direto, indireto e total dos Modelos SDM e SLX

Variáveis	Modelo SDM			Modelo SLX		
	Direto	Indireto	Total	Direto	Indireto	Total
Regularidade Docente	0,0004	0,0411	0,0415	0,0003	0,0099	0,0100
Cor	1,3732***	-1,1768	0,1964	1,3994***	-0,1380	1,2614*
Sexo	-0,4261*	-1,5619	-1,9880	-0,4454*	0,0000	-0,4454*
Escolaridade da Mãe	0,6526***	-32,8923	-32,2396	0,6973***	-7,1676*	-6,4703*
Distorção Idade Série	0,0033	-0,4857	-0,4823	0,0036*	-0,0995***	-0,0959***
Renda	0,0000	0,0083	0,0083	0,0000	0,0019	0,0019
Média Hora Aula	0,0716	3,1365	3,2080	0,0501	1,3684**	1,4185**
Média Aluno/Turma	-0,0062*	-0,6032	-0,6094	-0,0055	-0,1258*	-0,1313*
Esforço Docente	0,0005	0,0018	0,0023	0,0004	0,0000	0,0004
Gasto em Educação	0,0001**	0,0049	0,0050	0,0001**	0,0012	0,0014

Fonte: Elaboração própria. Notas: \*, \*\*, \*\*\* denotam valor-p menor do que 10%, 5% e 1%.

Os modelos SDM e SLX indicaram impacto direto significativos das variáveis que compõem as características socioeconômicas e apenas impacto direto do gasto em educação, no modelo SLX as variáveis: média hora aula e média aluno turma tiveram impacto indireto e total significativo, assim como distorção idade série.

Utilizando o critério AIC para escolha do modelo, podemos indicar que os modelos que apresentaram melhor ajuste foram respectivamente: SAC, SLX e SDM.

## Conclusões

Os determinantes educacionais são foco de estudos há muitos anos pela teoria econômica. O desempenho escolar é utilizado como proxy de qualidade de ensino. É de amplo conhecimento na literatura da economia da educação que as características socioeconômicas dos alunos possuem grande influência no desempenho, principalmente a escolaridade da mãe, cor e sexo do aluno.

Porém, quando se trata da análise econométrica espacial ainda existem poucos estudos que abordam os determinantes educacionais. A grande questão em observar os determinantes do desempenho através de modelos tradicionais de econometria espacial é o problema da falácia

ecológica em que se leva uma informação do todo para o nível individual, perdendo as nuances envolvidas na análise microeconômica.

Com o desenvolvimento de modelos que utilizam georreferenciamento este problema pode ser solucionado, sendo possível analisar a nível individual e considerando o espaço e a distribuição espacial dos dados na análise de dados individuais, porém as bases de dados brasileiras ainda não permitem este tipo de aplicação, apesar dos esforços constantes e a realidade cada vez mais próxima com o desenvolvimento de bases georreferenciadas em grandes centros de estudos.

Voltando a questão do artigo, os resultados indicaram que a regularidade docente não foi significativa em nenhum dos modelos analisados. Apesar disso, todos os modelos apresentaram resultados compatíveis com os indicados pela literatura, com grande influência da escolaridade materna no desempenho escolar, sexo e cor, além disso, algumas outras variáveis externas (que não estão nos modelos tradicionais de determinantes da educação) apresentaram efeito significativo quando a análise foi a partir do modelo espacial.

Primeiramente, a nota padronizada apresentou autocorrelação espacial, sendo o  $\rho$ , parâmetro que está ligado ao multiplicador espacial significativo em todos os modelos que foi estimado. Além disso, os gastos com atividades de educação foi uma variável significativa em todos os modelos, indicando que os gastos governamentais influenciam o desempenho escolar dos alunos das escolas públicas de Minas Gerais. A distorção idade-série também foi significativa nos modelos que apresentaram melhores ajustes a partir do critério AIC, com destaque para a grande significância da variável WDistorção idade-série.

A partir dos resultados é possível afirmar que características do corpo docente como esforço e regularidade em modelos espaciais não influenciaram o desempenho discente, a utilização de modelos hierárquicos pode ser uma melhor sugestão para a análise de diferentes níveis possibilitando melhor análise de como professor, escola e turma impactam nas notas dos alunos.

## Referências

ALBERNAZ, Ângela; FERREIRA, Francisco HG; FRANCO, Creso. **Qualidade e equidade na educação fundamental brasileira**. Texto para discussão, 2002.

ALMEIDA, Eduardo. **Econometria Espacial Aplicada**. Campinas: Alínea, 2012.

ALVES, M. T. G.; SOARES, J. F. Efeito-escola e estratificação escolar: o impacto da composição de turmas por nível de habilidade dos alunos. **Educação em Revista**, v. 45, p. 25-28, jun. 2007.

- BARRO, R. J. e Xavier SALA-I-MARTIN. Technological Diffusion, Convergence, and Growth. **Journal of Economic Growth**, vol. 2, pp.1-26, 1997.
- BARROS, R. P; MENDONÇA, R. Investimentos em Educação e Desenvolvimento Econômico. Rio de Janeiro, **IPEA**, 1997. (Texto para Discussão No. 525)
- BIONDI, Roberta L.; FELÍCIO, Fabiana. Atributos escolares e o desempenho dos estudantes: uma análise de painel dos dados do SAEB. Brasília: **INEP**, 2007.
- BIONDI, Roberta L.; VASCONCELLOS, Lígia; MENEZES-FILHO, Naércio A. Avaliando o Impacto da Olimpíada Brasileira de Matemática das Escolas Públicas no desempenho de matemática nas avaliações educacionais. In: 31º Encontro da Sociedade Brasileira de Econometria, 2009, Foz do Iguaçu. **Encontro Brasileiro de Econometria - SBE**, 2009.
- BIONDI, Roberta L.; VASCONCELLOS, Lígia; MENEZES-FILHO, Naércio A. Avaliando o impacto do programa Escrevendo o Futuro no desempenho das escolas públicas na Prova Brasil e na renda futura dos estudantes. Anais do **XXXVII Encontro Nacional de Economia da ANPEC**, Foz do Iguaçu, Paraná, 2009.
- BISHOP, J. Is the test score decline responsible for the productivity growth decline? **American Economic Review**. Vol. 79 (1), p. 178-97, 1989.
- CÉSAR, C. C.; SOARES, J. Desigualdades acadêmicas induzidas pelo contexto escolar. **Revista Brasileira de Estudos de População**, v. 18, n. 1/2, p. 97-110, 2001.
- COLEMAN, J. S. et al. Equality of educational Opportunity. U.S. **Government printing office**. Washington, DC, 1966.
- ERTUR, C; KOCH, W. Growth, technological interdependence and spatial externalities: theory and evidence. **Journal of Applied Econometrics**, v.22, issue 6, 2007.
- FELÍCIO, F.; FERNANDES, R. O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental no Estado de São Paulo. Anais do **XXXIII Encontro Nacional de Economia**, Natal/RN, Brasil. ANPEC, 2005.
- FUJITA, M.; KRUGMAN, P. & VENABLES, A. The spatial economy. Cambridge, **MIT**, 1999.
- HANUSHEK, E. A.; KIMKO, D. D. Schooling, labor-force quality and the growth of nations? **The American Economic Review**, 90(5):1184–1208.2000
- HOLLAND, P. W. Statistics and casual inference. **Journal of the American Statistical Association** , v. 81, n. 306, p. 945-960, 1986.
- HEWINGS, G., MAGALHAES, A., AZZONI, C. R. Spatial dependence and regional inequality in Brazil. **Investigaciones Regionales**, Madrid, v. 6, p. 5-20, 2005.
- MACHADO, Ana Flávia et al. Qualidade do ensino em matemática: determinantes do desempenho dos alunos em escolas públicas estaduais mineiras. **Revista Economia**, v. 9, n. 01, p. 23-45, jan/abr., 2008
- MENEZES-FILHO, Naércio. A Evolução da Educação no Brasil e seu Impacto no Mercado de Trabalho. São Paulo, **FEA/USP**, 2001.
- MENEZES-FILHO, Naércio. Os determinantes do desempenho escolar no Brasil. Instituto Futuro Brasil, **IBMEC São Paulo e Faculdade de Economia e Administração da Universidade de São Paulo**. Sumário Executivo, 2007.
- MORENO, R., TREHAN, B. Location and the Growth of Nations, **Journal of Economic Growth**, 2, pp.399-418, 1997.
- MORETTI, E. Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-sectional Data. **Journal of Econometrics**, n.121, p.175-

212, 2004.

MURNANE, R. J., WILLETT, J. B. AND LEVY, F. The growing importance of cognitive skill in wage determination. **Review of Economics and Statistics**, vol. 77 (2) p. 251-66, Maio, 1995.

MURPHY, K. M. AND PELTZMAN, S. School Performance and the Youth Labor Market. **Journal of Labor Economics**, vol.22 (2), pp.299-325, The University of Chicago. 2004.

PALERMO, G. A.; DO NASCIMENTO SILVA, D. B.; NOVELLINO, M. S. F. Fatores associados ao desempenho escolar: uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro. **Revista Brasileira de Estudos de População**, v. 31, n. 2, p. 367-394, 2014.

RAMOS, R., SURIÑACH, J., ARTÍS, M. Human capital spillovers, productivity and regional convergence in Spain. **Papers in Regional Science**, Volume 89 n2 June 2010.

RIANI, J. L. R.; RIOS-NETO, E. L. G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros? **Revista Brasileira de Estudos de População**, v. 25, n. 2, p. 251-269, jul./dez. 2008.

RIVKIN, S. G. Black/white differences in schooling and employment. **Journal of Human Resources**, vol.30 (4), p. 826-52, 1995.

RIVKIN, Steven G.; HANUSHEK, Eric A.; e KAIN, John F. Teachers, Schools, and Academic. **Achievement. Econometrica**. vol. 73, No.2, p417-58. 2005

ROSENTHAL S., STRANGE W. The attenuation of human capital externalities. **Journal of Urban Economics** 64: 373–389, 2008.

SILVEIRA NETO, R. M. Crescimento e spillovers: a localização importa? Evidências para os estados brasileiros. **Revista Econômica do Nordeste**, Fortaleza, v. 32, p. 524-545, 2001.

SOARES, Tufi M. Modelo de três níveis hierárquicos para a proficiência dos alunos de 4ª série avaliados no teste de língua portuguesa do SIMAVE/PROEB 2002. **Revista Brasileira de Educação**, n.29, p.73-87. 2005

SOUZA, M. R. P. Análise da variável escolaridade como fator determinante do crescimento econômico. **FAE**, Curitiba, Vol. 2, No. 3, p.47-56. 1999

VERNIER, Laura. D. S.; BAGOLIN, Izete. P.; JACINTO, Paulo. A. Fatores que Influenciam o Desempenho Escolar no Estado do Rio Grande do Sul: Uma Análise com Regressões Quantílicas. **Análise Econômica**, v. 33, n.64, 2015.

VERNIER, Laura Desirée Silva; BAGOLIN, Izete Pengo; FOCHEZATTO, Adelar. Distribuição e disseminação espacial da educação nos municípios brasileiros. **45º Encontro Nacional de Economia**, p. 18, 2017