

Evasão Escolar no Ensino Médio Seriado Brasileiro Através de Modelos de Regressão Espacial: Metodologia Potencial para Aplicação no Varejo

RESUMO

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, INEP, coleta e disponibiliza dados da educação em parceria com secretarias estaduais e municipais de educação do Brasil. Utilizando estas informações, a presente pesquisa busca identificar as variáveis que impactam na evasão escolar no ensino médio seriado brasileiro e a influência de municípios vizinhos. Os dados dos cadastros de matrículas, turmas, escolas, profissionais escolares e indicadores do Censo Escolar, nível socioeconômico dos alunos e remuneração dos professores, foram filtrados, normalizados, agregados por município e georreferenciados. A taxa da evasão escolar é resultado da taxa de abandono somada à taxa de não-matrícula, que foi calculada através da comparação das matrículas dos anos de 2015 e 2016.

As variáveis consideradas, 104 no total, foram testadas em um modelo de regressão linear múltipla otimizado com *stepwise regression*, mantendo 62 variáveis. A significância do coeficiente *Moran's I* aplicado sobre os resíduos do modelo comprovou que a utilização de modelos de regressão espacial poderia ser promissora. O teste *Lagrange Multiplier* indicou o uso de modelos de defasagem espacial, em detrimento de modelos de erro espacial. Com isso, foram adotados modelos de alcance global, *Spatial Autoregressive* (SAR), alcance local, *Spatial Lagged X* (SLX), e alcance misto, *Spatial Durbin* (SDM). Os modelos de defasagem espacial apresentaram melhor ajuste que o modelo de regressão linear múltipla e, entre eles, o *Spatial Durbin* mostrou o melhor desempenho, com um ajuste 12% melhor, demonstrando que a evasão escolar no ensino médio seriado de um município sofre influência de municípios vizinhos.

Estudos quantitativos na área de Varejo, em especial, considerando áreas de influência de pontos de oferta de serviços (lojas próprias ou franquias) podem se beneficiar de modelos como estes, visando melhorar a predição do desempenho através da influência espacial de variáveis da vizinhança (desempenho de lojas vizinhas e variáveis sócio-econômico-demográficas do público-alvo).

Palavras-chave: Evasão Escolar. Geoanálise. Regressão Espacial.

1. CONTEXTO

1.1 Evasão Escolar no Ensino Médio Brasileiro

Em 2016, tínhamos cerca de 10 milhões de jovens entre 15 e 17 anos no Brasil. Todos deveriam estar cursando o ensino médio, porém, todo ano, cerca de 1,5 milhão de jovens não se matriculam no início do ano letivo e cerca de 700 mil abandonam a escola durante o ano letivo. São 2,2 milhões de jovens fora da escola. A taxa de evasão escolar está praticamente estagnada, desde a virada do milênio, em cerca de 25%.

Um estudo do economista Ricardo Paes de Barros fez uma comparação entre moradores de zonas urbanas da região Sudeste e detectou que uma pessoa formada no ensino médio ganha, aos 35 anos, 122% a mais do que uma pessoa que não se formou, e tem 54% mais chances de um emprego formal com carteira de trabalho assinada. Além disso, o estudo mostra que filhos de pessoas que se formaram no ensino médio têm quase o dobro de chance de se formarem também, quando comparados a filhos de pais com ensino fundamental ou médio incompleto, gerando um ciclo virtuoso (BARROS, 2017).

No Brasil, o prejuízo causado por jovens fora da escola foi estimado em R\$ 200 bilhões anuais. O prejuízo para a sociedade está relacionado a questões de segurança pública, saúde e

desenvolvimento econômico. Falando da segurança pública, muitos jovens que abandonam a escola se envolvem com o tráfico de drogas e organizações criminosas e estima-se que o país gasta R\$ 18 mil a mais no combate à violência e criminalidade para cada jovem que não conclui o ensino médio. Quanto à saúde, a educação reduz a dependência de serviços públicos e aumenta a expectativa de vida da pessoa, uma vez que quem estuda tende a cuidar mais do seu bem-estar, e estima-se que o custo social na área da saúde é de R\$ 28 mil por jovem que não conclui o ensino médio. No âmbito da economia, os jovens que não concluem o ensino médio reduzem a produção de riquezas do país, com custo de oportunidade estimado na ordem de R\$ 49 mil por pessoa. O prejuízo somado e aplicado aos 2,2 milhões de jovens fora da escola é digno de nota (BARROS, 2017).

Sendo assim, a redução da evasão escolar não é só uma questão de elevar os níveis de habilidade e ajudar as pessoas a levar uma vida mais produtiva e saudável, é também um elemento de recuperação econômica.

1.2 Principais Motivos da Evasão Escolar

Buscar soluções para tornar a escola um ambiente atrativo e enriquecedor é um grande desafio. Muitos motivos impedem ou desmotivam os jovens para frequentar a escola:

- Pobreza: falta de condições básicas para frequentar a escola;
- Dificuldade de acesso à escola: restrições do transporte público;
- Necessidades especiais: limitações dos alunos, seja por doença grave ou deficiência; falta de profissionais qualificados, infraestrutura e recursos nas escolas para o atendimento a portadores de necessidades especiais;
- Reprovações cumulativas: endereçamento inadequado ao baixo desempenho acadêmico;
- Gravidez na adolescência: limitação no tempo disponível para os estudos e possíveis constrangimentos sociais (em 2016, este foi o motivo de 18% da evasão escolar de meninas, sendo que, entre as adolescentes que engravidam, apenas 2% continuam os estudos);
- Atividades ilícitas: rivalidade da escola com o consumo ou envolvimento no tráfico de drogas ou outras atividades criminosas;
- Mercado de trabalho: entrada precoce e/ou em intensidade inadequada no mercado de trabalho, seja para garantir o sustento da família, ou visando ganhos financeiros em curto prazo, sem se atentar aos prejuízos que esta decisão pode trazer em longo prazo;
- Intimidação, discriminação e violência: ocorrências que podem trazer sérias consequências e traumas, comprometendo o aprendizado e tornando a escola uma experiência insuportável;
- Responsabilidades familiares: cuidado de irmãos menores, membros idosos da família ou pessoas da família com problemas de saúde, gerando muitas dificuldades para frequentar a escola e manter o ritmo das tarefas escolares em casa;
- Percepção de relevância da escola: crença na desconexão entre a maioria do conteúdo visto em sala de aula e o “mundo real”; prestação de serviços de baixa qualidade, com muitas greves e/ou falta de professores, infraestrutura e equipamentos deficitários e desatualizados, contribuindo para o pensamento de que o que a escola ensina não é essencial e útil para a vida (GESTA, 2017);
- Problemas psicológicos: transtornos de aprendizagem como déficit de atenção; transtornos emocionais e comportamentais, incluindo a chamada crise da adolescência; depressão causada pela evolução dos transtornos (BONACHERA, 2015).

Os motivos da evasão escolar podem ser categorizados em dois grupos, chamados de *push* (empurrão) e *pull* (atração). Um aluno é empurrado para fora da escola quando normas e políticas de atendimento e disciplina e até mesmo consequências de mau comportamento levam ao abandono. Por outro lado, os alunos podem ser atraídos para fora da escola quando fatores

como preocupações financeiras, atratividade do mercado de trabalho, responsabilidades familiares e gravidez na adolescência, afastam os alunos da escola (JORDAN, 1994).

Os autores Watt e Roessingh acrescentaram uma terceira categoria, *falling out* (queda), que ocorre quando um aluno não apresenta progresso acadêmico significativo no trabalho escolar e se torna apático ou mesmo desiludido com a escola. Não é necessariamente uma decisão ativa, mas sim um efeito colateral de apoio familiar e/ou educacional insuficiente. O fator *falling out* destaca um processo de abandono pelo qual o aluno demonstra gradualmente um comportamento de desengajamento acadêmico (WATT, ROESSINGH, 1994).

2. FOCO DA PESQUISA

A presente pesquisa busca identificar quais são as características do aluno, da escola e da formação e atuação dos professores que impactam na evasão escolar no ensino médio seriado brasileiro. Além disso, a verificação da influência da vizinhança, determinada pela localização geográfica dos municípios, é parte fundamental da pesquisa.

O termo evasão escolar é utilizado em diversos contextos, muitas vezes com significados diferentes. Na presente pesquisa, evasão escolar é a situação do aluno que abandonou a escola, o que pode ser caracterizado de duas formas distintas:

- o aluno realizou a matrícula, mas não frequentou ou deixou de frequentar as aulas durante o ano letivo;
- o aluno concluiu um ano letivo, mas não se matriculou no ano subsequente.

A lei nº 9.394 que norteia a educação básica esclarece que o ensino médio é a etapa final da educação básica, com duração mínima de três anos, abrangendo adolescentes de 15 a 17 anos. Existem cursos profissionalizantes que podem ser articulados com o ensino médio e incluem um ano a mais de estudo. Como o quarto ano do ensino médio é opcional, este estudo considera apenas as três primeiras séries do ensino médio. Além disso, o ensino médio pode ser seriado ou não seriado. O ensino médio seriado, objeto do presente estudo, é organizado em séries anuais, enquanto o ensino não seriado envolve grupos que podem ser organizados com base na idade, na competência e em outros critérios de interesse para o processo de aprendizagem. No Censo Escolar de 2016, o ensino médio não seriado representa apenas 0,2% das matrículas no ensino médio (BRASIL. Lei nº 9394, 1996).

O escopo da presente pesquisa inclui todas as escolas do ensino médio em território nacional. As matrículas de alunos, cujo país de origem não é o Brasil, também são consideradas na pesquisa, uma vez que 99,87% das matrículas no ensino médio são de alunos, cujo país de origem é o Brasil.

2.1 Características da Amostra

Os dados da pesquisa foram obtidos a partir do Censo Escolar e de indicadores adicionais disponibilizados pelo INEP.

2.1.1 Cadastro de alunos (matrículas)

O cadastro de alunos ou matrículas contém informações sobre o aluno, como idade, sexo, nacionalidade, local de residência, tipo de necessidade especial, entre outras, e informações sobre a matrícula, como a etapa de ensino (série), metodologia de ensino (regular ou profissionalizante), tipo de mediação didático-pedagógica (ensino presencial, semipresencial ou a distância), entre outras. Além disso, o cadastro explicita o tipo de necessidade especial do aluno e o recurso necessário para a escola, caso o aluno seja portador de necessidade especial. A necessidade de transporte público para o deslocamento até a escola e o tipo de transporte utilizado pelo aluno também fazem parte deste cadastro.

Para o presente estudo, foram filtradas as matrículas do ensino médio seriado da 1ª a 3ª séries, regular e profissionalizante. Além disso, foram filtradas apenas as turmas que não são específicas de atividades complementares e atendimento educacional especializado:

Região	Quantidade de Matrículas	
	Total	Selecionadas
Centro-Oeste	3.924.436	591.303
Nordeste	22.609.439	2.211.168
Norte	5.356.369	787.559
Sul	7.066.943	1.031.773
Sudeste	20.741.766	3.398.550
Total	59.698.953	8.020.353

A obtenção do cadastro de matrículas do Censo Escolar do ano de 2015 também foi necessária para o presente estudo, uma vez que uma parcela da evasão escolar é composta por alunos que concluíram o ano letivo de 2015, mas não se matricularam em 2016. Para o cadastro de matrículas de 2015, o filtro das etapas de ensino inclui as matrículas do último ano do ensino fundamental, além das matrículas do ensino médio seriado:

Região	Quantidade de Matrículas	
	Total	Selecionadas
Centro-Oeste	4.064.193	824.280
Nordeste	16.669.798	2.973.225
Norte	5.866.810	1.053.627
Sul	21.026.814	4.527.803
Sudeste	7.223.607	1.415.649
Total	54.851.222	10.794.584

Com as matrículas e as taxas de aprovação de 2015, é possível identificar a quantidade de alunos que deveriam ter se matriculado no ensino médio em 2016, mas não o fizeram.

2.1.2 Cadastro de turmas

O cadastro de turmas traz informações sobre o horário, a carga horária das aulas e os dias de atividade por semana, além das disciplinas, atividades complementares e atendimentos especiais ministrados.

O cadastro de turmas de 2016 foi relacionado ao cadastro de alunos já filtrado, para seleção das turmas do ensino médio seriado. O total de turmas registradas no Censo Escolar de 2016 é 2.430.211, sendo que o presente estudo selecionou 262.072 turmas.

2.1.3 Cadastro de escolas

O cadastro de escolas contém informações sobre a localização da escola, a dependência administrativa (federal, estadual, municipal, privada), o local de funcionamento, sua infraestrutura (abastecimento de água, energia, esgoto e coleta de lixo), instalações, dependências e equipamentos. Além disso, o cadastro traz informações sobre as etapas de ensino que a escola atende.

O cadastro de escolas de 2016 foi relacionado ao cadastro de alunos já filtrado, para seleção das escolas do ensino médio seriado. O total de escolas registradas no Censo Escolar de 2016 é 279.358, sendo que o presente estudo selecionou 28.129 escolas.

2.1.4 Cadastro de profissionais escolares

O cadastro de profissionais escolares contém informações sobre o profissional, tais como, idade, sexo, nacionalidade, local de residência, entre outras, e sobre a sua formação, tais como,

grau de escolaridade, curso superior e ano de conclusão, especializações, realização de mestrado e/ou doutorado. Além disso, o cadastro traz as disciplinas que o docente leciona. O cadastro de profissionais escolares foi relacionado ao cadastro de alunos com os filtros já aplicados, para seleção dos profissionais escolares das turmas do ensino médio seriado. Além disso, foram filtrados os profissionais escolares que atuam nas funções de Docente e de Docente Titular Coordenador de Tutoria de Módulo ou Disciplina do Ensino a Distância:

Região	Quantidade de Docentes	
	Total	Selecionados
Centro-Oeste	893.615	254.125
Nordeste	2.930.354	757.517
Norte	1.066.913	283.487
Sudeste	4.734.613	1.297.668
Sul	1.846.143	457.304
Total Geral	11.471.638	3.050.101

2.1.5 Indicadores do censo escolar

Taxas de Rendimento Escolar: contêm as taxas de aprovação, reprovação e abandono por escola e por série.

AFD – Adequação da Formação do Docente: é categorizada em 5 grupos, de acordo com a formação do docente e a disciplina lecionada (INEP, 2018):

Grupo	Curso Superior	Licenciatura (ou Complementação Pedagógica)	Formação na Mesma Área que Leciona	Para cada escola, o INEP disponibiliza o percentual de docentes em cada grupo, somando 100% no ensino médio.
1	Sim	Sim	Sim	
2	Sim	Não	Sim	
3	Sim	Sim	Não	
4	Sim	Não	Não	
5	Não	Não	Não	

ICG – Indicador de Complexidade de Gestão da Escola: é identificado numa escala de 1 a 6, para cada escola, de acordo com as suas características predominantes (INEP, 2018):

Nível	Matrículas	Turnos	Séries
1	Menos de 50	1	1
2	50 a 300	2	2
3	50 a 500	2	2 a 3
4	150 a 1000	2 a 3	2 a 3
5	150 a 1000	3	2 a 3
6	Mais de 500	3	Mais de 3

IED – Indicador do Esforço Docente: classifica o docente numa escala de 1 a 6, de acordo com o esforço empreendido no exercício da profissão, sendo que níveis elevados indicam maior esforço (INEP, 2018):

Nível	Alunos	Turnos	Escolas	Séries	Para cada escola, o INEP disponibiliza o percentual de docentes em cada nível de esforço, somando 100% no ensino médio.
1	Até 25	1	1	1	
2	25 a 150	1	1	1	
3	25 a 300	1 a 2	1	1	
4	50 a 400	2	1 a 2	2	
5	Mais de 300	3	2 a 3	2 a 3	
6	Mais de 400	3	2 a 3	2 a 3	

HAD – Média de Horas-Aula Diária: é disponibilizada por escola.

TDI – Taxa de Distorção Idade-Série: é disponibilizada por escola.

INSE – Indicador do Nível Socioeconômico dos Alunos: foi apurado através de uma pesquisa realizada pelo INEP em 2013. A pesquisa foi realizada com um questionário contextual com questões referentes à posse de bens no domicílio, contratação de serviços de mensalista ou diarista, renda familiar mensal, em salários mínimos, e escolaridade dos pais. O indicador é disponibilizado por escola em valor absoluto, sendo a média aritmética simples do nível socioeconômico dos seus alunos.

Remuneração Média dos Docentes: foi apurada através de uma pesquisa realizada pelo INEP em 2014 por município e por nível de escolaridade dos docentes, sendo “com curso superior” e “sem curso superior”. A remuneração média é obtida através do RAIS - Relação Anual de Informações Sociais e padronizada para uma carga horária de 40 horas semanais de trabalho.

3. ABORDAGEM

3.1 Unidade de Análise

O presente estudo apresenta uma análise da evasão escolar no ensino médio seriado brasileiro por município e, para tanto, consolida os dados das escolas, turmas, alunos e docentes por município.

Existem escolas que só atendem o ensino médio, sendo que, de um total de 28.129 escolas do ensino médio brasileiro, 22% não atendem o ensino fundamental. A maior taxa de não-matrícula ocorre na primeira série do ensino médio, ou seja, alunos que concluíram o ensino fundamental não se matriculam no ensino médio no ano subsequente. A não-matrícula na primeira série do ensino médio representa 83% do total da taxa de não-matrícula no ensino médio. Sendo assim, não é possível determinar a escola que o aluno evadiu, uma vez que o aluno concluiu o ensino fundamental e, em 22% dos casos, o aluno obrigatoriamente precisa escolher outra escola para se matricular no ensino médio. Este é o principal motivo pelo qual a análise da presente pesquisa se dá no âmbito do município e não da escola.

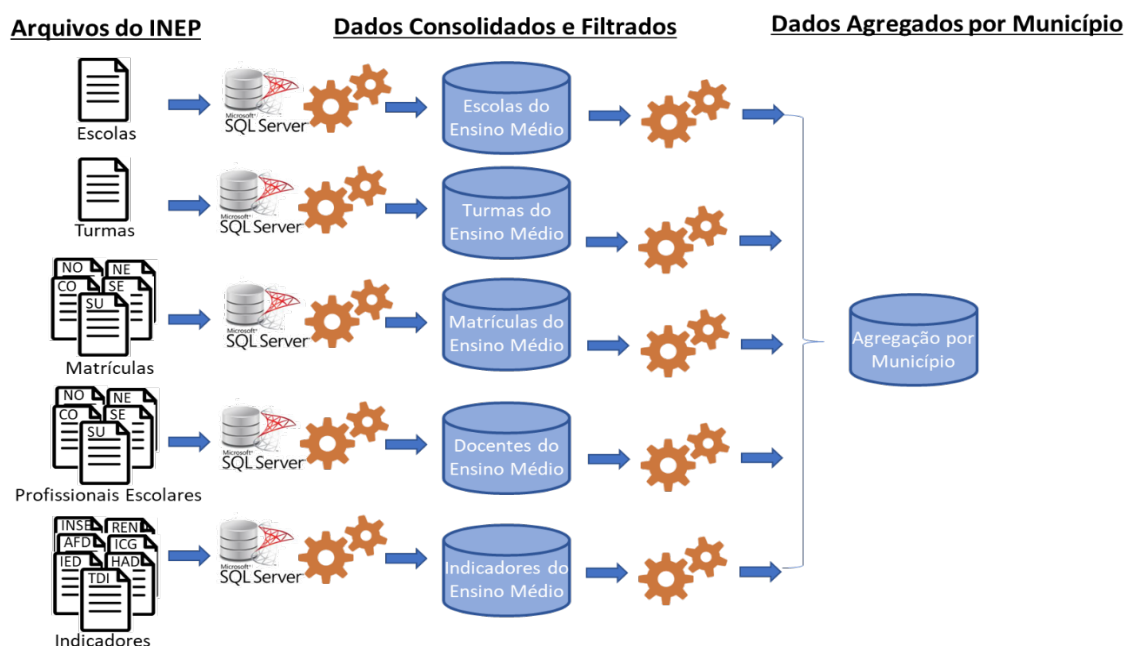
Ainda assim, considerando que 49% dos municípios possuem uma única escola para atender o ensino médio, o estudo acaba direcionado para a escola nestes municípios. Em linhas gerais, este estudo contribui com os órgãos governamentais federais, estaduais e municipais, voltados para a educação, além da administração das escolas e profissionais da educação.

3.2 Método

Os dados obtidos do censo escolar do INEP foram preparados para a análise: consolidados, filtrados, calculados, agregados por município e georreferenciados. Após a preparação dos dados, as análises incluíram duas abordagens: modelos de regressão linear para facilitar a identificação de variáveis que impactam a evasão escolar e modelos de regressão espacial para avaliar o impacto da vizinhança.

3.2.1. Preparação dos dados

Passo 1: Consolidação, filtro e agregação dos dados por município



Para o armazenamento e a preparação dos dados, foi utilizado o Microsoft SQL Server. A linguagem utilizada para as instruções ao banco de dados foi a linguagem SQL - Structured Query Language. Os arquivos disponibilizados pelo INEP foram importados no SQL Server e, em seguida, filtrados e agregados por município.

A agregação por município foi realizada em duas etapas. Na primeira etapa, as variáveis foram transformadas em quantidades. Na segunda etapa, as quantidades foram transformadas em médias ou percentuais, visando à atribuição de uma escala única para todos os municípios, independente da quantidade de escolas, turmas, alunos e/ou docentes. A agregação dos indicadores por município foi realizada através da média ponderada do indicador, de acordo com seu nível de granularidade e aplicação.

Algumas variáveis dos cadastros de escolas, turmas, matrículas e profissionais escolares foram desconsideradas do presente estudo porque: não são aplicáveis com a agregação por município, são muito específicas e não são aplicáveis a mais de 90% dos registros, são irrelevantes no contexto do ensino médio e/ou podem suscitar polêmicas sobre segregação.

Passo 2: Cálculo da taxa de evasão escolar por município

A evasão escolar é composta da taxa de não-matrícula e da taxa de abandono.

Taxa de não-matrícula

A taxa de não-matrícula contempla as seguintes situações:

- 1ª série: considera os alunos que foram aprovados na última série do ensino fundamental mais os alunos que foram reprovados na 1ª série do ensino médio no ano de 2015, e não realizaram a matrícula na 1ª série do ensino médio em 2016;
- 2ª série: considera os alunos que foram aprovados na 1ª série do ensino médio mais os alunos que foram reprovados na 2ª série do ensino médio no ano de 2015, e não realizaram a matrícula na 2ª série do ensino médio em 2016;
- 3ª série: considera os alunos que foram aprovados na 2ª série do ensino médio mais os alunos que foram reprovados na 3ª série do ensino médio no ano de 2015, e não realizaram a matrícula na 3ª série do ensino médio em 2016.

As taxas de aprovação e reprovação do ano de 2015 foram aplicadas na quantidade total de matrículas do ano de 2015 por escola e série, gerando a quantidade prevista para matrículas em 2016. As taxas de rendimento não existiam para 0,7% dos registros e, nestes casos, foi considerada uma taxa de aprovação de 100%. A quantidade prevista foi totalizada por município.

Para a obtenção da quantidade real de matrículas em 2016, foi verificada a existência da matrícula em 2016 para o universo de alunos de 2015 (através do código da pessoa física). A diferença entre a quantidade prevista e a quantidade real por município é a quantidade de não-matricula. Logo, a quantidade de não-matricula dividida pela quantidade prevista é a taxa de não-matricula.

Taxa de abandono

A taxa de abandono refere-se aos alunos que realizaram a matrícula em 2016, mas não concluíram o ano letivo. Esta taxa é parte integrante das taxas de rendimento escolar do censo escolar de 2016 disponibilizadas pelo INEP por escola e série. Portanto, a agregação por município foi realizada através de média ponderada. As taxas de rendimento não existiam para 0,7% dos registros e, nestes casos, foi considerada uma taxa de abandono de 0%.

Taxa de evasão escolar

A taxa de evasão escolar por município é a soma da taxa de não-matricula com a taxa de abandono.

É importante ressaltar que a taxa de evasão no presente estudo não leva em consideração a taxa de mortalidade, ou seja, alunos que faleceram no período de férias escolares entre os anos de 2015 e 2016, constam na taxa de não-matricula. Da mesma forma, alunos que faleceram ao longo do ano letivo de 2016, constam na taxa de abandono. Como referência, a taxa de mortalidade para jovens de 15 a 19 anos, em 2015, foi de 21.356 óbitos, o que representa 0,3% das matrículas no ensino médio em 2016. Porém, para o cálculo correto, seria necessária a taxa de mortalidade para jovens de 15 a 17 anos, o que reduziria a representatividade, mas esta informação não está disponível.

Passo 3: Georreferenciamento dos municípios e geração do arquivo shapefile

Para o modelo de regressão espacial, foi necessário adicionar as informações de latitude e longitude dos 5.570 municípios brasileiros. Os dados agregados por município foram exportados do SQL Server em um arquivo texto (.txt). A ferramenta QGIS foi utilizada para a transformação do arquivo texto em um arquivo shapefile (.shp), através das seguintes etapas:

- Na primeira etapa, o arquivo texto com as variáveis agregadas por município foi importado na ferramenta QGIS e um arquivo shapefile de pontos foi obtido a partir da latitude e longitude;
- Na segunda etapa, o arquivo shapefile de polígonos dos municípios brasileiros, obtido através do site www.codegeo.com.br, foi importado na ferramenta QGIS (CODEGEO, 2018);
- Por fim, o shapefile de pontos foi unido ao shapefile de polígonos, sendo que o resultado final é um shapefile de polígonos de municípios brasileiros contendo todas as variáveis deste estudo. Foram excluídos 9 municípios que não possuíam escola para atender o ensino médio seriado em 2016.

3.2.2. Análise dos dados

Para a identificação das variáveis que impactam a evasão escolar, foram criados modelos de regressão linear múltipla no software R.

A regressão linear múltipla, objeto deste estudo, determina a relação entre variáveis independentes e uma variável dependente que é representada através da seguinte equação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

Sendo que: y = variável dependente que, no presente estudo, é a taxa de evasão; β_0 = é uma constante que graficamente representa a interceptação da reta com o eixo vertical; β = são as constantes que representam o declive (coeficiente angular) da reta; x = são as variáveis independentes ou explicativas; p = número total de variáveis independentes; ε = variável que inclui todos os fatores residuais e os possíveis erros de medição, sendo que seu comportamento é aleatório e representa a variabilidade em y que não pode ser explicada pelo efeito linear das p variáveis independentes (SWEENEY; WILLIAMS; ANDERSON, 2015).

Para medir a qualidade do ajuste da equação, o presente estudo usa o coeficiente de determinação múltiplo, R^2 , que pode ser interpretado como “a proporção de variabilidade da variável dependente que pode ser explicada pela equação de regressão múltipla estimada” (SWEENEY; WILLIAMS; ANDERSON, 2015). O coeficiente é calculado da seguinte forma:

$$R^2 = \frac{SQReg}{SQTot}$$

Sendo que: $SQTot = SQReg + SQRes$; $SQReg$ = soma dos quadrados da regressão; $SQRes$ = soma dos quadrados dos resíduos.

Os “quadrados” referem-se à distância entre os pontos e a reta estabelecida pela equação. A regressão linear utiliza um método, conhecido como método dos mínimos quadrados, para criar a equação que melhor aproxima a relação linear entre as variáveis dependente e independentes, minimizando a soma dos quadrados dos resíduos.

O procedimento adotado no presente estudo para seleção das variáveis é conhecido como *stepwise regression*. Num modelo de regressão linear múltipla, quando uma variável é adicionada ou removida, os coeficientes das variáveis remanescentes podem sofrer alterações. As mudanças são causadas pela correlação entre as variáveis independentes. Se não existe correlação entre as variáveis independentes, o coeficiente de cada variável permanece o mesmo quando outra variável é incluída ou excluída. Porém, poucas variáveis são puramente independentes. É importante que o modelo inclua todas as variáveis relevantes para que os impactos na variável dependente e os impactos entre as variáveis independentes sejam refletidos nas estimativas. Por outro lado, a inclusão de variáveis irrelevantes pode causar problemas em predições, uma vez que a flutuação do coeficiente de variáveis irrelevantes tende a ser maior do que a flutuação do coeficiente das variáveis relevantes. Quando há um grande número de variáveis candidatas ao modelo, o procedimento *stepwise regression* pode ser utilizado para a seleção do melhor conjunto de variáveis (WANG; JAIN, 2003).

A correlação entre duas variáveis y e x sem a influência de outras variáveis é chamada de correlação simples e é a divisão entre a covariância e o produto do desvio padrão das variáveis:

$$corr(y, x) = \frac{cov(y, x)}{\sigma_y * \sigma_x}$$

Quando o coeficiente de correlação entre y e x é calculado com a eliminação do efeito de outras variáveis, é chamado de coeficiente de correlação parcial. Por exemplo, se temos três variáveis y , x_1 e x_2 e precisamos calcular o coeficiente de correlação parcial entre y e x_1 , é necessário remover o impacto de x_2 em y e x_1 :

$$corr(y, x_1.x_2) = \frac{corr(y, x_1) - corr(y, x_2) * corr(x_1, x_2)}{\sqrt{(1 - corr(y, x_2)^2) * (1 - corr(x_1, x_2)^2)}}$$

O cálculo dos coeficientes de correlação parcial pode ficar muito complexo quando mais e mais variáveis são envolvidas. O procedimento *stepwise regression* calcula o coeficiente de correlação simples e correlação parcial entre a variável dependente e cada uma das variáveis independentes consideradas. No software R, este procedimento pode ser realizado através de uma única linha de comando. No presente estudo, o resultado do *stepwise regression* é o modelo de regressão linear múltipla com o melhor subconjunto de variáveis que explicam a evasão escolar.

Além do modelo de regressão linear múltipla com *stepwise regression*, o presente estudo considerou a análise do impacto das características da vizinhança na evasão escolar do município, através de modelos de regressão espacial. Primeiramente, o coeficiente Moran's *I* foi utilizado para checar se um modelo de regressão espacial seria estatisticamente válido. O coeficiente criado por Patrick Alfred Pierce Moran, que mede a autocorrelação espacial, foi aplicado aos resíduos do modelo de regressão linear. Sua fórmula é a seguinte:

$$I = \frac{N}{W} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Sendo que: N = quantidade de unidades espaciais indexadas por i e j (neste estudo, são os municípios); x = variável de interesse (neste estudo, é a taxa de evasão escolar); $\bar{x}$ = média de x; w_{ij} = matriz de pesos espaciais com zeros na diagonal (por exemplo, $w_{ii} = 0$); W = somatória de todos os w_{ij} (LI; CALDER; CRESSIE, 2007).

Moran's *I* é uma estatística inferencial e, para sua interpretação, é necessário checar a sua significância estatística.

Uma vez confirmada a hipótese de que a regressão espacial é aplicável ao presente estudo, o teste Lagrange Multiplier (LM) foi utilizado para determinar que tipo de modelo seria mais apropriado. As estatísticas do teste Lagrange Multiplier permitem uma distinção entre modelos de erro espacial e modelos de defasagem espacial. Foram realizados os seguintes testes de dependência espacial em modelos lineares:

- teste LM simples para dependência de erro: LMErr;
- teste LM simples para variável dependente com defasagem espacial: LMLag;
- teste robusto para dependência de erro com a possível presença de variável dependente com defasagem espacial: RLMErr;
- teste robusto para variável dependente com defasagem espacial com a possível presença de dependência de erro: RLMLag.

O teste LM demonstrou que modelos de defasagem espacial são mais indicados para o presente estudo. Sendo assim, foram utilizados três diferentes tipos de modelos de defasagem espacial para análise do impacto das características da vizinhança na evasão escolar do município: o Spatial Autoregressive ou SAR, o Spatial Lagged X ou SLX, e o Spatial Durbin ou SDM.

O modelo Spatial Autoregressive, SAR, considera o efeito da variável dependente observada em unidades vizinhas, ou seja, o quanto a taxa de evasão escolar de municípios vizinhos influenciam na taxa de evasão escolar do município. O modelo é representado pela seguinte equação:

$$y = \rho Wy + X\beta + \varepsilon$$

Sendo que: y = vetor com as observações sobre a variável dependente que, no presente estudo, é a taxa de evasão; ρ = coeficiente autoregressivo espacial, que varia entre -1 e 1; Wy = vetor de defasagens espaciais para a variável dependente; X = matriz de observações das variáveis independentes (mais a constante); β = vetor com os coeficientes de regressão β ; ε = variável que inclui todos os fatores residuais e os possíveis erros de medição. (ALMEIDA, 2012)

O modelo SAR é considerado um modelo de alcance global. Essa classe de modelos é caracterizada por considerar a dependência espacial através da interferência do multiplicador espacial, que faz com que o impacto sobre a variável dependente seja refletido para todas as regiões da área em estudo. O alcance global faz com que o impacto gerado para a taxa da evasão escolar num município seja transmitido para o restante dos municípios por meio do multiplicador espacial. (ALMEIDA, 2012)

Enquanto o modelo Spatial Autoregressive, SAR, considera o efeito da variável dependente observada em unidades vizinhas, o modelo de regressão espacial Spatial Lagged X, SLX, considera o efeito das variáveis independentes observadas em unidades vizinhas. Trata-se de um modelo de regressão linear ampliado que inclui o impacto de algumas ou todas as variáveis

independentes observadas em unidades vizinhas. O modelo é representado pela seguinte equação:

$$y = X\beta + WX\theta + \varepsilon$$

Sendo que: y = vetor com as observações sobre a variável dependente que, no presente estudo, é a taxa de evasão; X = matriz de observações das variáveis independentes (mais a constante); β = vetor com os coeficientes de regressão β ; $WX\theta$ = média de X na vizinhança; ε = variável que inclui todos os fatores residuais e os possíveis erros de medição (BURKEY, 2018).

O modelo SLX é considerado um modelo de alcance local. Essa classe de modelos é caracterizada por considerar que o alcance da dependência espacial é localizado, no sentido de que o impacto é observado apenas para algumas regiões da área em estudo, sobretudo os vizinhos diretos. Neste estudo, o alcance local faz com que o impacto das variáveis independentes seja observado entre municípios vizinhos (ALMEIDA, 2012).

O modelo Spatial Durbin, SDM, é a junção dos modelos SAR e SLX, considerando o efeito da variável dependente e das variáveis independentes observadas em unidades vizinhas. O modelo é representado pela seguinte equação:

$$y = \rho Wy + X\beta + WX\theta + \varepsilon$$

O modelo SDM é considerado um modelo de alcance global e local. O alcance global é dado pelo multiplicador espacial da variável dependente defasada espacialmente e o alcance local é dado pelas defasagens espaciais das variáveis independentes (ALMEIDA, 2012).

A matriz W indica a contiguidade entre duas unidades espaciais que, neste caso, são municípios. Quando dois municípios compartilham uma borda comum de comprimento diferente de zero, podemos dizer que são contíguos. Para este estudo, foi utilizado o critério conhecido como *queen*. Trata-se de uma referência a rainha no jogo de xadrez, que pode se movimentar em qualquer direção, inclusive nas diagonais. O critério da rainha é o mais abrangente e define vizinhos como unidades espaciais compartilhando uma aresta comum ou um vértice comum.

No contexto deste estudo, portanto, foi utilizado um modelo espacial de alcance global, o SAR, um modelo espacial de alcance local, o SLX, e um modelo misto de alcance global e local, o SDM.

Para a escolha do modelo apropriado, do ponto de vista estatístico, foi utilizado o Akaike Information Criterion, AIC. Na busca do modelo que melhor explica o comportamento da variável resposta, o modelo com menor AIC é considerado o modelo de melhor ajuste.

A estimativa AIC para um determinado modelo é dada por:

$$AIC = 2k - 2\ln$$

Sendo que: k = número de parâmetros estimados no modelo; $\ln$ = valor máximo da função de verossimilhança para o modelo. (AKAIKE, 1974)

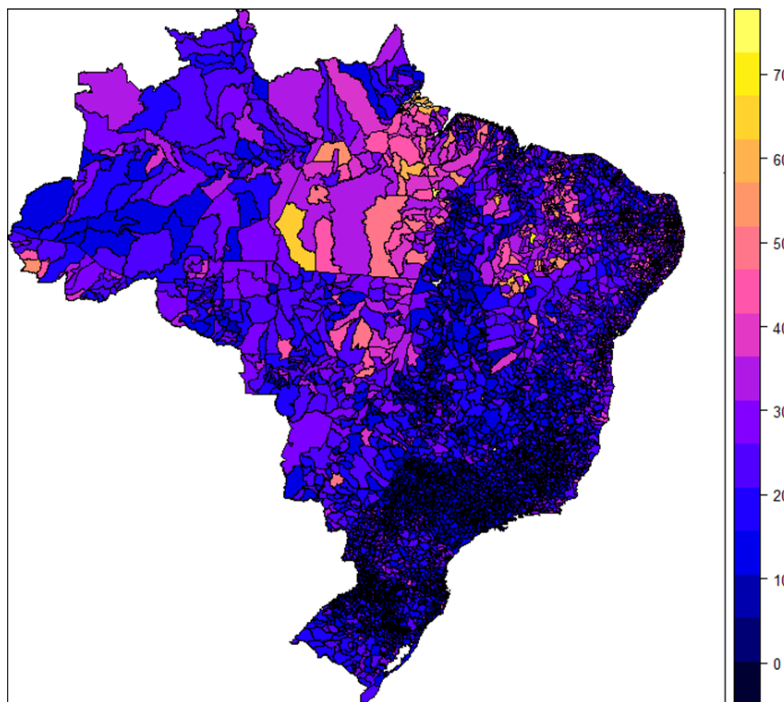
Com o uso do AIC, considera-se que nenhum modelo avaliado é “real”, ou seja, descreve com perfeição a relação entre a variável dependente e as variáveis independentes. Então, o modelo escolhido é o que minimiza a divergência, lidando com o conflito entre a qualidade do ajuste do modelo e a simplicidade do modelo. Por um lado, o critério AIC recompensa a qualidade do ajuste do modelo (*goodness of fit*), que é avaliada pela função de verossimilhança. Por outro lado, inclui uma penalidade em função do número de parâmetros estimados. Desta forma, o critério desincentiva a inclusão de parâmetros no modelo que, apesar de melhorar a qualidade do ajuste, podem causar *overfitting*, ou seja, podem fazer com que o modelo seja ineficaz para novos conjuntos de dados.

4. ANÁLISE DOS DADOS

A análise dos dados é iniciada com um modelo de regressão linear múltipla, seguida de *stepwise regression* para sua otimização, identificando as variáveis da escola, turma, aluno e/ou

matrícula e corpo docente que são relevantes na evasão escolar. Na sequência, o coeficiente Moran's I é calculado sobre os resíduos do modelo, confirmando a validade de modelos de regressão espacial para a evasão escolar. O teste Lagrange Multiplier mostra que modelos de defasagem espacial são mais indicados que modelos de erro espacial. Então, a análise segue com os modelos de defasagem espacial de alcance global Spatial Autoregressive, SAR, alcance local Spatial Lagged X, SLX, e alcance misto (global e local) Spatial Durbin, SDM. O Akaike Information Criterion, AIC, é utilizado confirmando que o modelo Spatial Durbin, SDM, representa a melhor alternativa entre os modelos analisados neste estudo.

Mapa da evasão escolar no ensino médio seriado brasileiro:



4.1. Regressão linear múltipla (OLS)

O modelo de regressão linear múltipla, Ordinary Least Squares, OLS, para a taxa de evasão escolar no ensino médio seriado brasileiro foi, inicialmente, criado com todas as variáveis. Em seguida, para eliminar as variáveis irrelevantes, o procedimento *stepwise regression* foi utilizado, com os seguintes resultados:

Residual standard error: 6,982 on 5498 degrees of freedom Multiple R-squared: 0,5086, Adjusted R-squared: 0,503 F-statistic: 91,77 on 62 and 5498 DF, p-value: < 0,00000000000000022

A condição dos resíduos foi satisfatória para que o modelo fosse considerado válido estatisticamente e não houveram casos extremos que pudessem influenciar a linha de regressão. Com a eliminação de variáveis irrelevantes, o R^2 ficou em 50,3%. No total, 42 variáveis foram eliminadas, enquanto 62 variáveis foram mantidas, sendo que existem variáveis com correlação positiva e negativa. As variáveis mantidas seguem abaixo, sendo que as variáveis com maior relevância estão destacadas em negrito:

- **Escola:** tipo de dependência estadual, localização na zona rural, existência de água filtrada, abastecimento de água e esgoto da rede pública, existência de serviço de coleta periódica de lixo, existência de quadra de esportes e biblioteca, adaptação a portadores de necessidades especiais, existência de refeitório e oferta de alimentação, existência de equipamentos de retroprojeção, som e multimídia, acesso à internet com banda larga,

quantidade média de alunos por funcionário, índice de complexidade de gestão entre os níveis 2 e 5 (ICG), **taxas de aprovação e reprovação do ano anterior**;

- **Turma**: quantidade de turmas, quantidade média de alunos por turma, quantidade média de dias de atividade, **média de horas-aula diária (HAD)**;
- **Aluno**: quantidade de alunos, porte de necessidades especiais, matrícula em curso profissionalizante, residência na zona rural e em município diferente da escola, matrícula em curso presencial ou EAD, **taxa de distorção idade-série (TDI)**, **indicador socioeconômico (INSE)**;
- **Corpo docente**: professores com menos de 25 anos de idade, com idade entre 25 e 40 anos e com mais de 55 anos de idade, residência na zona rural e em município diferente da escola, professores com curso superior completo, especialização e mestrado, professores especializados no ensino médio e em direitos da criança e do adolescente, professores graduados em instituições públicas, professores que lecionam em apenas uma escola, em 3 ou mais escolas, em apenas uma série, em apenas uma disciplina, em duas disciplinas, em 3 ou mais disciplinas, índice de esforço do docente entre os níveis 2 e 5 (IED), adequação da formação do docente entre os grupos 2 e 5 (AFD).

4.2. Coeficiente Moran's I

O coeficiente Moran's I foi calculado sobre os resíduos do modelo de regressão linear múltipla. O resultado mostra que o coeficiente é estatisticamente significativo, indicando que um modelo de regressão espacial poderia ser aplicado ao presente estudo:

Moran I statistic standard deviate = 26,828, p-value < 0.0000000000000022		
alternative hypothesis: two.sided		
sample estimates:		
Moran I statistic	Expectation	Variance
0,21505676285	-0,00017995321	0,00006436505

4.3. Teste Lagrange Multiplier

O teste Lagrange Multiplier foi utilizado para determinar que tipo de modelo seria mais apropriado: modelos de erro espacial ou modelos de defasagem espacial:

LMerr = 718.54, df = 1, p-value < 0.00000000000000022
RLMerr = 2.5534, df = 1, p-value = 0.1101
LMlag = 944.62, df = 1, p-value < 0.00000000000000022
RLMlag = 228.63, df = 1, p-value < 0.00000000000000022

Comparando o resultado dos testes robustos RLMerr e RLMLag, o teste Lagrange Multiplier demonstra que os modelos de defasagem espacial são mais apropriados para este estudo.

4.4. Regressão espacial

Os modelos de regressão espacial, Spatial Autoregression, SAR, Spatial Lagged X, SLX, e Spatial Durbin, SDM, foram criados para a taxa de evasão escolar com as variáveis significativas identificadas previamente no item "4.1 Regressão linear múltipla (OLS)". Todos apresentaram condição dos resíduos satisfatória para que o modelo fosse considerado válido estatisticamente

4.4.1. Spatial Autoregression (SAR)

O modelo Spatial Autoregression, SAR, obteve melhor desempenho que o modelo de regressão linear múltipla:

Rho: 0,45636, LR test value: 813,64, p-value: < 0,000000000000000222
 Approximate (numerical Hessian) standard error: 0,014857
 t-value: 30,717, p-value: < 0,000000000000000222
 Wald statistic: 954,36, p-value: < 0,000000000000000222
 Log likelihood: -18258,95 for lag model
 ML residual variance (sigma squared): 39,99 (sigma: 6,3238)
Nagelkerke pseudo-R-squared: 0,57545
 Number of observations: 5561
 Number of parameters estimated: 65
 AIC: 36648 (AIC for lm: 37460)

4.4.2. Spatial Lagged X (SLX)

O modelo Spatial Lagged X, SLX, obteve melhor desempenho que o modelo de regressão linear múltipla e que o modelo SAR:

Residual standard error: 6.405 on 5436 degrees of freedom
 Multiple R-squared: 0.591, **Adjusted R-squared: 0,5817**
 F-statistic: 63.35 on 124 and 5436 DF, p-value: < 0.00000000000000022

4.4.3. Spatial Durbin (SDM)

O modelo Spatial Durbin, SDM, obteve o melhor desempenho entre todos os modelos analisados:

Rho: 0.37418, LR test value: 400.41, p-value: < 0.000000000000000222
 Approximate (numerical Hessian) standard error: 0.017264
 t-value: 21.673, p-value: < 0.000000000000000222
 Wald statistic: 480.54, p-value: < 0.000000000000000222
 Log likelihood: -17954.88 for mixed model
 ML residual variance (sigma squared): 36.358, (sigma: 6.0298)
Nagelkerke pseudo-R-squared: 0.61943
 Number of observations: 5561
 Number of parameters estimated: 127
 AIC: 36164, (AIC for lm: 36562)

4.5. Comparação entre os Modelos Explicativos

Uma comparação entre os modelos de regressão linear múltipla e de regressão espacial empregados neste estudo mostra que o modelo Spatial Durbin, SDM, apresenta o melhor resultado.

	OLS	SAR	SLX	SDM
R²	50,3%	57,54%	58,17%	61,94%
AIC	37459,53	36647,89	36562,17	36163,76

Com isso, concluímos que há impacto dos municípios vizinhos na evasão escolar no ensino médio seriado e que este impacto ocorre de duas formas. A taxa de evasão escolar em um município sofre impacto da taxa de evasão escolar em municípios vizinhos. Além disso, as variáveis independentes de municípios vizinhos ajudam a explicar a evasão escolar do município.

Em 2015, um estudo divulgado pelo Instituto Brasileiro de Geografia e Estatística, IBGE, mostrou que, “em todo o Brasil, 7,4 milhões de pessoas se deslocam para cidades vizinhas de

onde moram para trabalhar ou estudar”. Segundo o estudo, habitantes de municípios menores e mais pobres buscam alternativas de emprego e estudo em municípios vizinhos. Sendo assim, parece plausível e justificada a influência de municípios vizinhos na evasão escolar. (IBGE, 2015)

5. CONSIDERAÇÕES FINAIS E APROPRIAÇÃO PELO VAREJO

A principal contribuição do presente estudo reside na constatação do impacto da vizinhança na evasão escolar no ensino médio seriado brasileiro. Os objetivos propostos pelo presente estudo foram alcançados com a identificação de que a evasão escolar no ensino médio seriado em um município é impactada não somente por variáveis que representam as condições da escola, turma, aluno e corpo docente no município, mas também pela evasão escolar e pelas variáveis que impactam a evasão escolar nos municípios vizinhos.

A importância do tema para a sociedade brasileira, entretanto, demanda pesquisas futuras que, como sugestão, podem ser aprofundadas nos seguintes aspectos:

- realização de testes de novas variáveis não disponíveis no Censo Escolar do INEP;
- redimensionamento do mapa: redução da vizinhança de município para distrito ou bairro;
- especialização do estudo: realização de estudos isolados por tipo de dependência (pública ou privada) e/ou para cada série do ensino médio, sendo que o estudo da segunda e terceira séries poderia ser realizado por escola.

O melhor ajuste dos modelos do presente estudo, cerca de 62%, sugere que a evasão escolar não pode ser muito bem explicada só pelas características da matrícula, turma, escola e corpo docente. A falta de interesse e o desengajamento do aluno pode ser um processo gradativo e lento que culmina na evasão escolar e seus sintomas são importantes na detecção e tratamento da questão. O acompanhamento do comportamento e desempenho do aluno ao longo do ano letivo poderia trazer indícios relevantes a serem considerados em estudos futuros da evasão escolar, uma vez que a taxa de distorção idade-série (TDI) e a taxa de reprovação do ano anterior estão entre as variáveis mais significativas para a taxa de evasão escolar.

Estudos quantitativos na área de Varejo, em especial, considerando áreas de influência de pontos de oferta de serviços (lojas próprias ou franquias) podem se beneficiar de modelos como estes, visando melhorar a predição do desempenho através da influência espacial de variáveis da vizinhança (desempenho de lojas vizinhas e variáveis sócio-econômico-demográficas do público-alvo). As características sociodemográficas, como escolaridade, renda e número de pessoas que compõem uma família, influenciam as atitudes dos consumidores em relação a produtos e marcas, e podem ser utilizadas para segmentar mercados (WEDEL; KAMAKURA, 2000; KAMAKURA; MAZZON, 2013). Estudos quantitativos que levem em conta ferramentas geográficas muito têm contribuído para estratégias de marketing e de localização de varejistas (BAVIERA-PUIG *et al.*, 2013), porém ainda são incipientes na apropriação de variáveis de localização e predicados espaciais em modelos preditivos, tornando o estudo aqui apresentado de alto potencial para o Varejo, de maneira geral.

REFERÊNCIAS

- ALMEIDA, Eduardo. *Econometria Espacial Aplicada*. São Paulo: Editora Alínea, 2012.
- AKAIKE, Hirotugu. A New Look at the Statistical Model Identification. IEEEExplore Digital Library, 1974. Disponível em <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1100705&isnumber=24140>. Acesso em 01/mai/2018.
- BAVIERA-PUIG, A.; ROIG-TIERNO, N.; BUILTRAGO-VERA, J.; MAS-VERDU, F. Comparing trade areas of technology centres using ‘Geographical Information Systems’. *The Service Industries Journal*, 33(7-8), 789-801, 2013.
- BARROS, Ricardo Paes, 2016. Políticas Públicas para Redução do Abandono e Evasão Escolar de Jovens. Patrocínio de Fundação BRAVA, Instituto UNIBANCO, Instituto de Ensino e Pesquisa

INSAPER, Instituto AYRTON SENNA. Disponível em <https://gesta.org.br/wp-content/uploads/2017/09/Brava_COMPLETA_V11.pdf>. Acesso em 30/mar/2018.

BONACHERA, Ángela R., 31 de julho de 2015. Los Principales Problemas Psicológicos en Adolescentes. Disponível em <http://www.hacerfamilia.com/psicologia/noticia-principales-problemas-psicologicos-adolescentes-20150731084431.html>. Acesso em 30/mar/2018.

BRASIL. Lei nº 9394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional, Brasília, DF, dez 1996. Disponível em http://www.planalto.gov.br/ccivil_03/leis/19394.htm>. Acesso em 05/fev/2018.

BRASIL. Ministério da Saúde, DATASUS Tecnologia da Informação a Serviço do SUS. Mortalidade Brasil. Disponível em <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10br.def>>. Acesso em: 30/mar/2018.

BURKEY, Mark L, Janeiro de 2018. The Mother of All R Spatial Econometrics Handouts. Disponível em <<http://spatial.burkeyacademy.com/>>. Acesso em 01/mai/2018.

_____. , Janeiro de 2018. Spatial Regression in R: The Four Simplest Models. Disponível em <<https://www.youtube.com/watch?v=b3HtV2Mhmvk&feature=youtu.be>>. Acesso em 01/mai/2018.

CODEGEO, 2018. Shapefiles do Brasil para Download. Autores: JUNIOR, Benicio; ESTRELA, Fred; CRUZ, Gledson; LIMA, Marcelo. Disponível em <<http://www.codegeo.com.br/2013/04/shapefiles-do-brasil-para-download.html>>. Acesso em 30/mar/2018.

GESTA GALERIA DE ESTUDOS E AVALIAÇÃO DE INICIATIVAS PÚBLICAS, set 2017. Disponível em <<http://gesta.org.br/tema/engajamento-escolar>>. Acesso em 30/mar/2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 25 de março de 2015. IBGE Mostra que 7,4 mi de Brasileiros Estudam ou Trabalham Fora da Cidade em que Residem. Disponível em <<http://www.brasil.gov.br/cidadania-e-justica/2015/03/IBGE-mostra-que-7-4-mi-de-brasileiros-estudam-ou-trabalham-fora-da-cidade-em-que-residem>>. Acesso em 01/mai/2018.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Censo Escolar 2016. Brasília, DF. Disponível em <http://inep.gov.br/microdados>. Acesso em: 06/jan/2018.

_____. Censo Escolar 2015. Brasília, DF. Disponível em <<http://inep.gov.br/microdados>>. Acesso em 06/jan/2018.

_____. , 2018. Indicadores Educacionais. Brasília, DF. Disponível em <<http://inep.gov.br/indicadores-educacionais>>. Acesso em 20/jan/2018.

JORDAN, Will J. Exploring the Complexity of Early Dropout Causal Structures. Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students, The John Hopkins University. Google Scholar, 1994.

KAMAKURA, W. A.; MAZZON, J. A. Socioeconomic status and consumption in an emerging economy. International Journal of Research in Marketing, 30, 4-18. 2013.

LI, Hongfei; CALDER, Catherine A.; CRESSIE, Noel., Setembro de 2007. Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. Disponível em <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.2007.00708.x>. Acesso em 01/mai/2018.

SWEENEY, Dennis J.; WILLIAMS, Thomas A.; ANDERSON, David R. Estatística Aplicada à Administração e Economia. Tradução da 6ª edição americana, 3ª edição brasileira. São Paulo: Cengage Learning, 2015.

WANG, George C. S.; JAIN, Chaman L. Regression Analysis: Modeling & Forecasting. 1st Edition. New York: Graceway Publishing Company, 2003.

WATT, David; ROESSINGH, Hetty. Some You Win, Most You Lose: Tracking ESL Dropout in High School (1988-1993). English Quarterly, 26, 5-7. Google Scholar, 1994.

WEDEL, M.; KAMAKURA, W. A. Market segmentation: Conceptual and methodological foundations. Boston: Kluwer Academic. 2000.