



Desempenho dos alunos de escolas públicas no Exame Nacional do Ensino Médio (2012-2018): análise e representação

Performance of public-school students in the National High School Exam (2012-2018): analysis and representation

Eduardo Felix Justiniano*  , Alfredo Pereira de Queiroz  

Departamento de Geografia, Faculdade de Filosofia, Letras e Ciências Humanas,
Universidade de São Paulo, São Paulo, SP, Brasil

E-mail: aqueiroz@usp.br

*E-mail para correspondência: e.justiniano@usp.br

Recebido (Received): 15/05/2021

Aceito (Accepted): 16/08/2021

Resumo: Dados de 5,6 milhões de alunos de escolas públicas que pontuaram em todas as provas do Exame Nacional do Ensino Médio (ENEM), de 2012 a 2018, foram analisados. O desempenho dos alunos foi agrupado por escola e por zona, correlacionado com variáveis socioeconômicas e escolares, e representado em diferentes unidades administrativas (município, micro e mesorregiões, estados e regiões). O problema das unidades de áreas modificáveis e as falácias também foram avaliados na comparação entre os diferentes níveis de agregação. Constatou-se que a pontuação média dos concluintes melhorou no período, mas o número total de alunos com elevada proficiência não aumentou da mesma forma. Dentre as variáveis analisadas, a renda familiar *per capita* apresentou maior correlação com o desempenho neste teste educacional. Embora a quantidade de dados analisados seja expressiva e seus resultados tenham permitido identificar as áreas de melhor desempenho, considerou-se que o baixo percentual de participação no exame pode comprometer o potencial de utilização deste teste educacional para avaliar a qualidade educacional das escolas públicas brasileiras. De modo geral, as menores proficiências do Brasil ocorreram nos municípios e microrregiões do Norte e do Nordeste do Brasil. Nestes agrupamentos territoriais, a renda *per capita* média é frequentemente menor. Apesar dos valores de correlação terem aumentado nos zoneamentos de maior agregação, a percepção desta associação entre as variáveis diminuiu nas representações cartográficas.

Palavras chave: correlação; ENEM; zoneamento; educação.

Abstract: Data was analyzed from 5.6 million public school students who scored in all tests of the National High School Exam (ENEM), from 2012 to 2018. Student performance was grouped by school and by zone and correlated with socioeconomic and school variables. Student performance was also represented in different administrative units (municipality, micro and mesoregions, states and regions). The modifiable areal unit problem and fallacies were also evaluated in the comparison in different levels of aggregation. The analysis determined that the average score of graduates improved in the time period, but the total number of students with high proficiency did not increase in the same way. Of the variables analyzed, family income *per capita* showed a greater correlation with performance in the educational tests. An analysis was also conducted for the problem of modifiable areal unit problem that results from the change in average results and correlations in the aggregation of Federation Units in Regions. The variation of means and correlations in the different administrative zonings of the Brazilian territory was analyzed as well as a determination of some of the possible fallacies that could occur in the interpretation of the results. Although the amount of data analyzed was substantial and results of the analysis have allowed identification of the areas of best performance, it was determined that the low percentage of participation in the exam compromises the potential of using this test to assess the educational quality of Brazilian public schools. In most cases, the lowest proficiencies in Brazil occurred in the municipalities and micro-regions of the North and Northeast of Brazil, where the average *per capita* income is often lower. Although correlation values increased with a greater aggregation level, the perception of this association between variables on maps decreased.

Keywords: correlation; ENEM; zoning; education.

1. Introdução

O Exame Nacional do Ensino Médio (ENEM), organizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), foi instituído em 1998. Seus objetivos são: possibilitar a autoavaliação do cidadão, criar uma referência nacional dos egressos do ensino médio, fornecer subsídios para o acesso à educação superior e se constituir como acesso aos cursos profissionalizantes pós-ensino médio (BRASIL, 1998). O ENEM passou a ter, também, o objetivo de avaliar o desempenho acadêmico dos ingressantes nos cursos de graduação do país (BRASIL, 2009).

A aceitação do ENEM tanto pelos estudantes quanto pelas instituições de ensino aumentou de forma progressiva. Em 1998, dos 1.330.150 alunos concluintes do ensino médio somente 115.575 (8,7%) do total participaram do exame. Esse número evoluiu até 2014, com mais de 8,7 milhões de inscritos, porém 2,5 milhões não compareceram às provas. O custo do exame não pode ser considerado como justificativa para a baixa participação, pois a inscrição é gratuita para alunos concluintes das escolas públicas e para pessoas de baixa renda (BRASIL, 2013).

O ENEM é um dos maiores testes educacionais do mundo em relação ao número de candidatos (TRAVITZKI, 2017). Os microdados do ENEM contêm dados do desempenho escolar e as características dos alunos, de suas famílias e residências. Informações relativas às características das escolas e quantidade de alunos podem ser obtidas nos censos escolares.

Os microdados do ENEM e os censos escolares são duas valiosas fontes de dados. Quando integrados, permitem que os dados individuais dos alunos sejam agrupados em diferentes recortes administrativos. No entanto, as medidas estatísticas de descrição e de associação entre variáveis se alteram numérica e visualmente conforme o tipo de agregação (escolas, zonas, municípios, etc.). Além da agregação, os resultados também variam com o uso de zoneamentos alternativos (OPENSHAW, 1984).

Esta variação começou a ser discutida na década de 1930 e, posteriormente, passou a ser conhecida como o problema das unidades de área modificáveis ou “*modifiable areal unit problem*” (MAUP). Esse problema é perceptível quando se comparam as médias, as variâncias e as correlações em zoneamentos distintos de uma mesma área.

Este estudo utiliza os microdados do ENEM e dos censos escolares elaborados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Seu objetivo é analisar os resultados do desempenho no ENEM dos estudantes de escolas públicas concluintes do ensino médio no Brasil, entre 16 e 18 anos, no período de 2012 a 2018. Visa também avaliar suas correlações com os dados socioeconômicos, em diferentes níveis de agregação, e abordar os possíveis problemas de interpretação decorrentes do processo de representação dos resultados (MAUP e falácias).

2. Embasamento teórico

O estudo pioneiro sobre a força da correlação entre variáveis é atribuído a Galton (1889). Em sua pesquisa, o autor analisou a relação entre o tamanho do cúbito e a estatura de homens adultos. No entanto, o coeficiente de correlação foi formalizado matematicamente por Pearson (1896), e mede o grau da correlação (positiva ou negativa) entre duas variáveis em uma escala que varia entre -1 e 1. A correlação foi considerada como uma função que descreve a relação entre variáveis.

Na área da educação, alguns dos pioneiros do uso da correlação foram Burt (1911), Duff e Thomson (1923), Jones e Carr-Saunders (1927), Collins e Douglass (1937) e Fleming (1943). Em período mais recente, White (1982) identificou diversas variáveis utilizadas em pesquisas que mensuraram o *status* socioeconômico dos alunos, como renda familiar, tamanho da família, instrução dos pais, ocupação, etnia, atividades culturais e intelectuais, salário de professores e proporção aluno-professor. Para Contreras et al. (2015), a localização dos alunos teria influência na variação de resultados das correlações entre desigualdade social, *bullying* e desempenho escolar, em diferentes países.

A quantificação do grau de associação entre duas variáveis é feita pelo coeficiente de correlação linear de Pearson que descreve, por um único número, a associação ou dependência entre duas variáveis (BUSSAB; MORETTIN, 2010). A identificação do grau de dependência entre variáveis pode auxiliar na formulação de hipóteses sobre a dependência espacial entre fenômenos geográficos (FERREIRA, 2014). O coeficiente de correlação é dado pela fórmula:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]}}$$

Onde \bar{X} e \bar{Y} são médias aritméticas das variáveis X e Y , respectivamente, e o coeficiente de correlação (r) varia de -1 a 1. Acima de zero, quanto maior for o valor de r , maior será a correlação positiva entre as variáveis; abaixo de zero, quanto menor for o valor de r , maior será a correlação negativa. Os valores próximos de 0 indicam ausência de correlação entre as variáveis.

As correlações espaciais são frequentemente associadas às áreas ou zonas. Os cientistas sociais, por inúmeras razões, trabalham com dados agregados e os representam na forma de mapas coropléticos. A principal desvantagem de utilizar dados agregados, contudo, está na dificuldade de fazer inferências válidas para vários níveis (escala de agrupamento), baseadas no resultado de um único nível (CLARK; AVERY, 1976).

Na Geografia, em particular nos zoneamentos, a variação de resultados de correlação e de dispersão em diferentes níveis de agregação de área passou a ser observada a partir da década de 1930. Gehlke e Biehl (1934) estudaram as correlações entre criminalidade de jovens do sexo masculino e os valores mensais de locação e, também, entre o valor dos produtos agrícolas e o número de fazendeiros. Demonstraram que: 1) a variação dos coeficientes de correlação estaria condicionada ao tamanho das unidades de área utilizadas; 2) quanto maior a área e o nível de agrupamento, maior seria a correlação entre as variáveis. Os autores creditaram essa perspectiva a Henry Sheldon que, em 1931, demonstrou o aumento do valor do coeficiente de correlação na medida em que as dimensões das áreas analisadas aumentavam e o número de áreas analisadas diminuía.

O coeficiente de correlação não seria apenas uma medida do grau de relacionamento entre variáveis, mas uma medida do grau de correspondência em suas respectivas distribuições geográficas e, para uma área heterogênea, a correlação poderia ser alta numa parte e ausente em outra (NEPRASH, 1934).

Yule e Kendall (1950) estudaram a correlação entre a produtividade de batata e trigo. Os autores estabeleceram as correlações em unidades de área modificáveis da produção agrícola de 1936, em 48 condados da Inglaterra e o agrupamento destes em 24, 12, 6 e 3 zonas. Os valores de correlação obtidos foram 0,2189, 0,2963, 0,5757, 0,7649 e 0,9902, respectivamente. Os autores consideraram que estas áreas geográficas foram escolhidas, portanto são passíveis de modificação, assim como são as unidades de tempo consideradas em uma observação.

Robinson (1950) estabeleceu uma diferenciação entre a correlação ecológica e a correlação individual. Na correlação individual, o objeto estatístico seria indivisível e as variáveis seriam propriedades descritivas do indivíduo. Seu valor era o da correlação de Pearson entre duas variáveis de todos os indivíduos, desconsiderando-se o posicionamento geográfico. Na correlação ecológica, o objeto estatístico seria um grupo de pessoas e as variáveis poderiam ser expressas em porcentagens, medianas e, principalmente, médias. O valor da correlação ecológica dependeria do número de zonas, e certamente seria diferente do valor da correlação individual correspondente. Em decorrência, as correlações ecológicas não poderiam ser usadas como substitutas das correlações individuais.

A falácia ecológica consiste na atribuição de valores das correlações ecológicas aos indivíduos (ALKER, 1969). Consiste em um equívoco de atribuição da correlação do todo às suas partes. As falácias também poderiam ocorrer na atribuição de valores de correlação (ou covariância) de um indivíduo para uma zona (falácia universal), da zona para o indivíduo (falácia seletiva) e de uma zona para o zoneamento ou grupo de zonas e vice-versa (falácia entre escalas). Além disso, a falácia ocorreria também por fatores contextuais, quando a estrutura social altera a força ou a forma das relações causais ou estatísticas.

Para Openshaw (1984), esta questão da variação de resultados em unidades de áreas modificáveis seria composta por dois problemas distintos, mas muito relacionados: escala e agregação. No primeiro problema, a variação nos resultados seria consequência da redução progressiva do número de unidades de análise e do aumento das suas respectivas áreas e agrupamento de dados. O segundo decorreria da modificação dos limites das áreas e manutenção do número de unidades de análise.

Os zoneamentos poderiam atuar como detectores dos padrões espaciais e, esses padrões dependeriam do desenho do zoneamento (OPENSHAW, 1984). As zonas seriam constituídas por objetos ou áreas geográficas individualizadas, que seriam a unidade básica para a observação e para a medição de fenômenos espaciais. No agrupamento (redução do número de zonas), ocorreria a diminuição da variância dos valores por área e, como consequência, a heterogeneidade espacial seria distorcida ou perdida (JELINSKI; WU, 1996).

Para Manley (2013), no entanto, é improvável encontrar uma solução global para o MAUP. O problema decorrente da agregação de dados deve ser sempre ponderado, mas não poderia ser eliminado sob o risco de negar a inerente espacialidade dos dados. Nesse contexto, o autor recomenda considerar uma escala analítica explícita ou, quando possível, adotar uma abordagem multiescalar que possibilite uma rigorosa reflexão sobre as distintas manifestações escalares do fenômeno.

3. Procedimentos metodológicos

Foram selecionados os inscritos que declararam estar matriculados no último ano letivo do ensino médio, das escolas públicas, entre 16 e 18 anos (idade escolar ideal para o último ano do ensino médio). Os microdados do ENEM e dos censos escolares de 2012 a 2018 estão disponíveis no sítio eletrônico do INEP (<http://inep.gov.br/microdados>).

Os programas SPSS e Tableau foram utilizados para gerenciar os dados e, os aplicativos MsOffice Access e Excel, para recortar e gerenciar as tabelas, descrever, agrupar e obter medidas estatísticas. O *software* ArcMap foi utilizado para importar e gerenciar os arquivos vetoriais das malhas territoriais, extraídos do IBGE. Esses arquivos foram concatenados com as tabelas para a produção dos mapas utilizados na avaliação de zoneamentos (**Figura 1**).

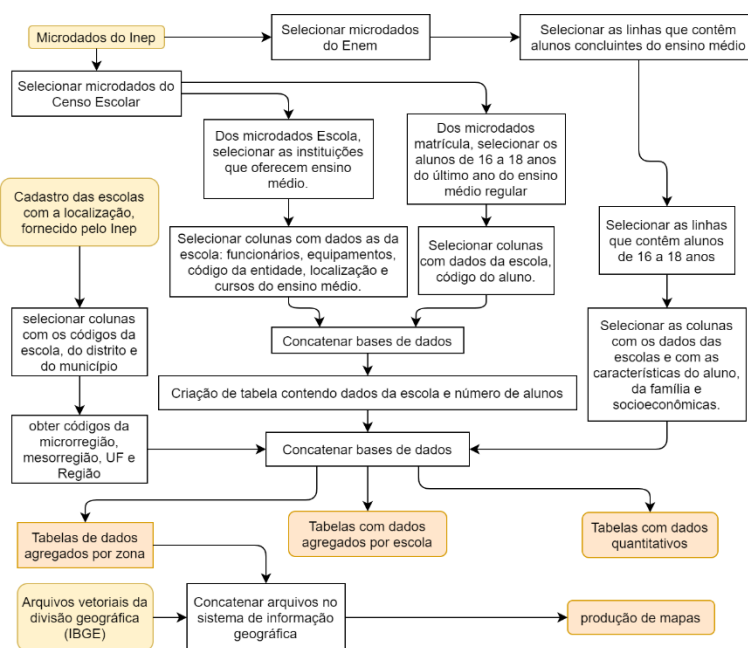


Figura 1 – Organograma de processamento de dados.

As tabelas produzidas possibilitaram estabelecer correlações de forma a identificar as variáveis mais associadas com o desempenho no ENEM. Os resultados estatísticos e a espacialização destes nos diversos níveis de agregação foram comparados. Buscou-se compreender as diferenças nos resultados médios e nas correlações em diferentes zoneamentos. Dados agregados e desagregados foram confrontados, possibilitando identificar eventuais falácias que podem ocorrer na interpretação dos dados.

4. Resultados e discussão

4.1 Análise exploratória e descritiva de dados

De 2012 a 2017, a maioria dos concluintes do ensino médio de até 18 anos da rede pública se inscreveu e pontuou em todas as provas do ENEM. Em 2018, a quantidade de não inscritos e dos que faltaram ou foram eliminados na prova foi superior ao número dos que pontuaram em todas as provas e na redação (**Tabela 1**). Nesse período, 5.561 municípios (99,8% do total de municípios brasileiros) possuíam escolas que ofereceram o último ano do ensino médio regular, para os concluintes que prestaram o ENEM.

Tabela 1: Alunos concluintes do ensino médio no Brasil de até 18 anos e a participação no ENEM (x 1.000).

Ano	Concluintes	Candidatos que pontuaram em todas as provas	Participação (%)
2012	1.350,8	711,7	52,7
2013	1.370,1	783,9	57,2
2014	1.411,5	803,8	56,9
2015	1.398,8	862,3	61,6
2016	1.482,7	918,3	61,9
2017	1.443,9	822,1	56,9
2018	1.434,4	702,9	49,0
Total	9.889,4	5605,0	56,7

Fonte: INEP (2013, 2015a, 2015b, 2016a, 2016b, 2016c, 2016d, 2016e, 2017a, 2017b, 2018a, 2018b, 2019a, 2019b).

A avaliação das competências e das habilidades dos alunos foi mensurada por provas objetivas de cada área do conhecimento: linguagem e códigos e suas tecnologias (LC), matemática e suas tecnologias (MT), ciências humanas e suas tecnologias (CH) e ciências da natureza e suas tecnologias (CN) e pela redação (RED), cujas pontuações variam de 0 a 1.000. A pontuação média dos concluintes pode ser observada na **Tabela 2**. O aluno foi considerado proficiente em uma área do conhecimento quando obteve 450 pontos ou mais; na redação, foi proficiente o aluno que obteve pelo menos 500 pontos. A expressão proficiência plena foi utilizada para caracterizar os alunos proficientes em todas as provas.

Tabela 2: Pontuação média dos alunos concluintes do ensino médio da rede pública.

Ano	CN	CH	LC	MT	RED	Média
2012	509,6	458,7	484,2	495,6	497,1	489,0
2013	503,2	461,3	482,0	501,5	511,7	491,9
2014	541,2	477,0	508,4	466,7	496,5	498,0
2015	548,1	466,0	497,8	459,6	533,9	501,1
2016	527,9	468,8	516,6	476,9	539,0	505,8
2017	509,8	496,7	502,5	507,5	551,5	513,6
2018	557,7	480,9	518,4	522,2	509,9	517,8
Média	528,2	472,8	501,4	490,0	519,9	502,5
Desvio padrão	19,8	12,2	13,4	21,2	20,0	9,9
Coeficiente de variação (x100)	3,7	2,6	2,7	4,3	3,8	2,0

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

Para estabelecer as correlações do desempenho no ENEM com as características socioeconômicas, foram selecionadas variáveis que pudessem ser comparadas ao longo do tempo, como renda familiar, quantidade de moradores na residência, anos de estudo dos pais e número de professores na escola. O grau de instrução dos pais foi dividido em nove categorias nos microdados das edições do ENEM de 2012 a 2014 e em oito categorias nas edições de 2015 a 2018. Para possibilitar a correlação com o desempenho no ENEM, um valor médio de anos de estudo para cada classe foi atribuído (**Tabela 3**).

Tabela 3: Grau de instrução dos pais segundo os microdados do ENEM.

2012 a 2014	2015 a 2018	Anos de estudo
Não estudou.	Nunca estudou.	0
Da 1ª à 4ª série do ensino fundamental (antigo primário).	Não completou a 4ª série/5º ano do ensino fundamental.	2,5
Da 5ª à 8ª série do ensino fundamental (antigo ginásio).	Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do ensino fundamental.	7
Ensino médio (antigo 2º grau) incompleto.	Completou o ensino fundamental, mas não completou o ensino médio.	10,5
Ensino médio (antigo 2º grau).	-	12
Ensino superior incompleto.	Completou o ensino médio, mas não completou a faculdade.	14
Ensino Superior.	-	16
-	Completou a faculdade, mas não completou a pós-graduação.	17,5
Pós-graduação.	Completou a pós-graduação.	19
Não sabe.	Não sabe.	Nulo

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

Nos microdados do ENEM, a renda familiar dos inscritos, de 2012 a 2018, está dividida em 17 faixas. Para definir a renda familiar *per capita*, o valor médio de cada faixa de renda foi dividido pelo número de moradores na residência. Os concluintes com renda familiar acima de 20 salários-mínimos (SM) foram excluídos da amostra pela impossibilidade de estabelecer o valor médio para essa faixa de renda (0,11% do total). Esses dados foram reagrupados em quatro faixas de renda familiar *per capita* (**Tabela 4**).

Tabela 4: Alunos concluintes do ensino médio da rede pública que pontuaram em todas as provas do ENEM, de 2012 a 2018, por faixas de renda *per capita* (x 1.000 alunos).

Renda <i>per capita</i> (SM)	Alunos	Frequência relativa (%)
até 0,2	1.648,2	29,4
Acima de 0,2 até 0,4	1.291,1	23,1
Acima de 0,4 até 0,8	1.708,8	30,5
Acima de 0,8	950,4	17,0

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

A participação dos alunos no ENEM não se mostrou homogênea no país entre 2012 e 2018. Em 291 municípios, mais de 80% dos concluintes pontuaram no exame, contudo, menos de 20% dos concluintes pontuaram em 20 municípios. Mesmo quando se compara as capitais, a variação na participação no ENEM é grande: de 31,1% em Maceió (AL) a 66,3% em Belém (PA).

Observou-se que, apesar do aumento constante da pontuação média dos concluintes do ensino médio no período (**Tabela 2**), a evolução do número total de proficiências nas áreas do conhecimento não ocorreu da mesma forma. Houve uma constante diminuição do percentual de alunos da rede pública não proficientes e dos que obtiveram uma única proficiência. Porém, isso não resultou no aumento contínuo dos plenamente proficientes, os quais corresponderam a menos de 40% do total de alunos. Variaram positiva e negativamente os percentuais de alunos parcialmente proficientes que obtiveram nota mínima em três e quatro provas e dos plenamente suficientes (**Figura 2**).

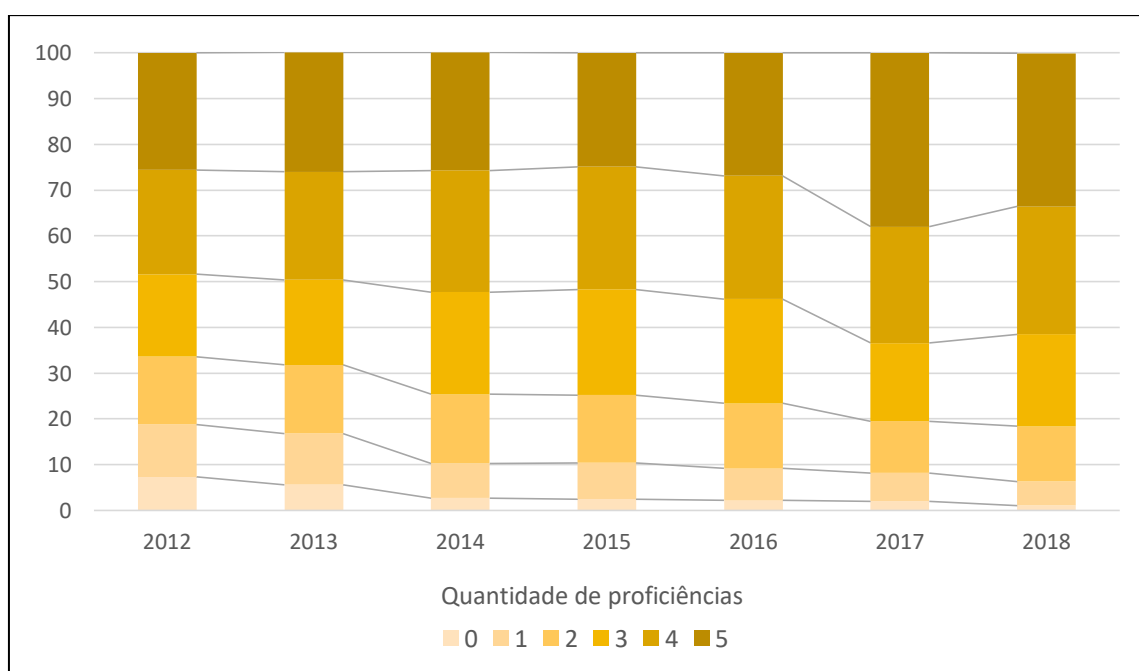


Figura 2: Variação percentual das proficiências obtidas pelos concluintes do ensino médio da rede pública. Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

A análise da relação entre proficiência e renda permitiu constatar que a proficiência média aumentou em todas as faixas de renda familiar *per capita* de 2012 a 2017. Em 2018, no entanto, houve uma redução da proficiência média em todas as faixas. A proficiência média anual mais elevada para os alunos de maior renda foi considerado um indicativo da correlação positiva entre essas duas variáveis (**Figura 3**).

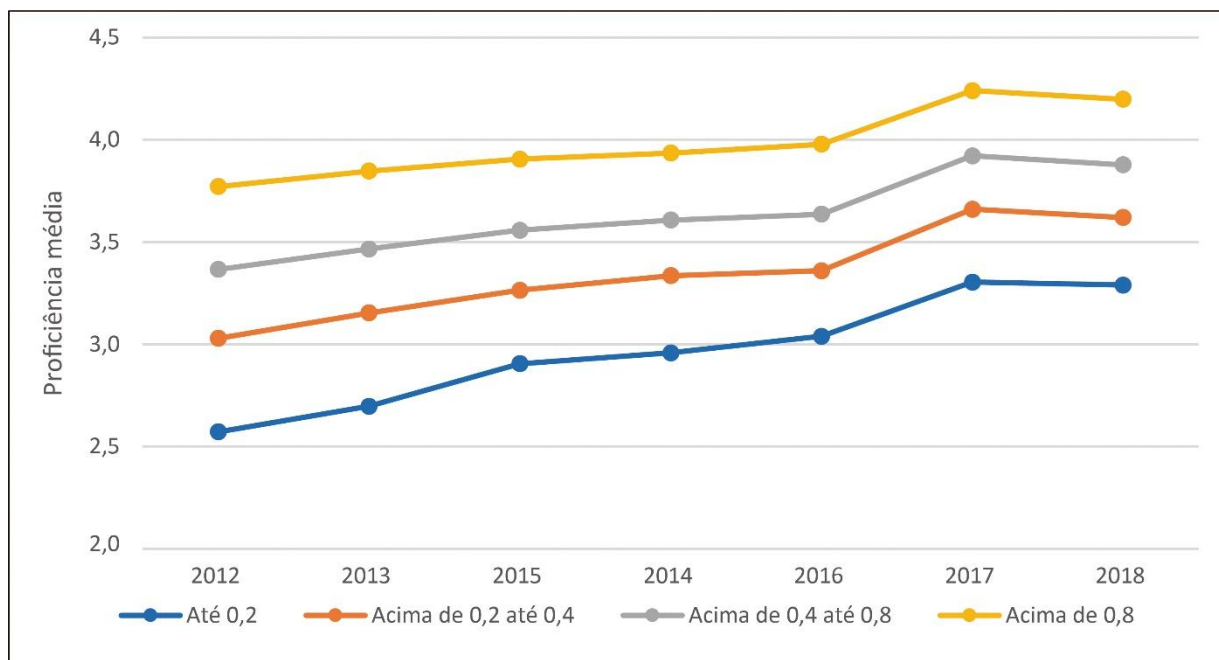


Figura 3: Evolução das proficiências médias dos alunos por faixa de renda familiar *per capita* (2012-2018).
Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

A participação dos estudantes no teste educacional durante o período analisado foi considerada baixa (**Tabela 1**), apesar da gratuidade da inscrição no ENEM para os alunos concluintes de escolas públicas (Ministério da Educação, 2010). Este aspecto reduz o potencial de uso dos dados agregados de desempenho por unidade de área para mensurar a qualidade de ensino. Além disso, corrobora a afirmação de Travitzki *et al.* (2014), sobre a possibilidade da classificação das escolas no ENEM representar uma amostra distorcida do ensino médio no Brasil, provável decorrência do caráter voluntário do exame.

4.2 Associação entre variáveis

Os coeficientes de correlação individual dos alunos podem ser observados na **Tabela 5** (exceto renda superior a 20 SM). Observou-se uma forte correlação individual positiva entre a pontuação e a proficiência. As variáveis também foram agrupadas e correlacionadas por escolas (**Tabela 6**).

Tabela 5: Correlações individuais com a variável pontuação (2012-2018)

Variável	Correlação
Proficiências	0,836201
Renda <i>per capita</i>	0,284918
Renda familiar	0,280623
Anos de estudo - Pai	0,249858
Anos de estudo - Mãe	0,228373
Moradores	-0,103530

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

Ser ou não proficiente em uma ou mais áreas do conhecimento foi um parâmetro estabelecido pelo MEC, com base no desempenho no ENEM. Em razão desse estabelecimento e da alta correlação com a pontuação, optamos por utilizar a proficiência média como parâmetro de desempenho.

A renda familiar *per capita* foi selecionada para estabelecer a correlação com o desempenho no ENEM por ser a variável socioeconômica que apresentou os maiores valores de correlação individual e agrupada com a pontuação no ENEM.

Desta forma, elegemos a proficiência e a renda *per capita* para avaliar o comportamento da correlação e dos resultados médios em diferentes níveis de agregação: regiões, unidades federativas, mesorregiões, microrregiões e municípios. Os distritos, que são unidades menos agregadas, não foram analisados pela inexistência de escolas em 35% deles.

Tabela 6: Variáveis agrupadas por escolas correlacionadas com a pontuação no ENEM (2012-2018).

Variável	Correlação
Proficiência média	0,950339
Renda <i>per capita</i> média	0,602034
Renda familiar média	0,584255
Média de moradores por família	-0,410568
Média dos anos de estudo dos pais	0,493646
Média dos anos de estudo das mães	0,435501
Participação no exame	0,394337
Número de alunos concluintes da escola	0,152946
Porcentagem de alunos homens na escola	0,079376
Porcentagem de alunas mulheres na escola	-0,079376
Número de computadores para alunos	0,443638
Número de funcionários da escola	0,272662
Número de equipamentos multimídia	0,372857

Fonte: INEP (2013, 2015a, 2015b, 2016a, 2016b, 2016c, 2016d, 2016e, 2017a, 2017b, 2018a, 2018b, 2019a, 2019b).

Os valores das correlações ecológicas aumentaram na medida em que se agrupavam áreas (redução do total de zonas). Por outro lado, os valores das correlações individuais por zona, entre estas variáveis, indicaram a existência de áreas onde não há correlação entre renda e desempenho no ENEM. No zoneamento municípios, há zonas onde a correlação entre estas variáveis é negativa (**Tabela 7**). Dos municípios às regiões, o aumento do valor da correlação ecológica observada é compatível com o obtido por Gehlke e Biehl (1934), enquanto a variação da correlação individual nas diferentes zonas corroborou a pesquisa de Neprash (1934).

Tabela 7: Correlações ecológicas e individual entre proficiência e renda *per capita* (2012-2018).

Zoneamento	Zonas	Zonas com escolas	Correlação ecológica	Correlação individual por zona
Regiões	5	5	0,933589	0,244213 a 0,277545
UF	27	27	0,814791	0,200971 a 0,355581
Mesorregiões	137	137	0,776775	0,091754 a 0,391367
Microrregiões	558	558	0,761100	0,046786 a 0,454123
Municípios	5.570	5.561	0,662224	-0,304343 a 0,533128

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

4.3 Variação da correlação

As amplitudes da proficiência e da renda familiar *per capita* diminuíram na medida em que se reduziu o número de zonas, de municípios para regiões, como ilustra a **Figura 4**.

A diferença de correlação entre variáveis foi perceptível na comparação entre zoneamentos. Esse problema diz respeito à variação das medidas estatísticas quando ocorre o agrupamento espacial dos dados (redução das zonas). O entendimento desse problema pode auxiliar na interpretação e evitar conclusões equivocadas, relativas às falácias descritas por Robinson (1950) e Alker (1969).

A variação da correlação e da média nos zoneamentos avaliados estaria relacionada com o problema de escala, descrito por Openshaw (1984), pois, dos municípios para as regiões, os conjuntos de unidades de área são progressivamente agrupados em menos unidades, porém com maiores extensões. No agrupamento de áreas, do município para a região (**Figura 4**), ocorreu a diminuição da diferenciação zonal da proficiência média, que corrobora as considerações de Jelinski e Wu (1996).

A diferença na visualização das médias nos mapas e dos gráficos de dispersão é característica das comparações multiescalares. Para Nelson e Brewer (2017), as operações e interações não são restritas a uma única escala, mas acontecem e respondem de maneira diferente em escalas diversas.

As microrregiões são compostas por municípios que compartilham certas especificidades relacionadas à estrutura de produção, que podem ser resultado do ambiente natural e das relações sociais e econômicas (IBGE, 1990). Dessa forma, os municípios próximos que foram agregados em uma microrregião estariam mais relacionados que os municípios mais distantes e que foram agregados em outra zona. As mesorregiões, por sua vez, são definidas pelo processo social, pelo quadro natural e pela rede de comunicação e de lugares (IBGE, 1990). As microrregiões de uma mesorregião seriam mais relacionadas do que as microrregiões de

outra mesorregião. Neste caso, a característica de agregação das bases de estruturas territoriais, proposta por IBGE (1994), poderia ser considerada compatível com os princípios da primeira lei da geografia, proposta por Tobler (1970).

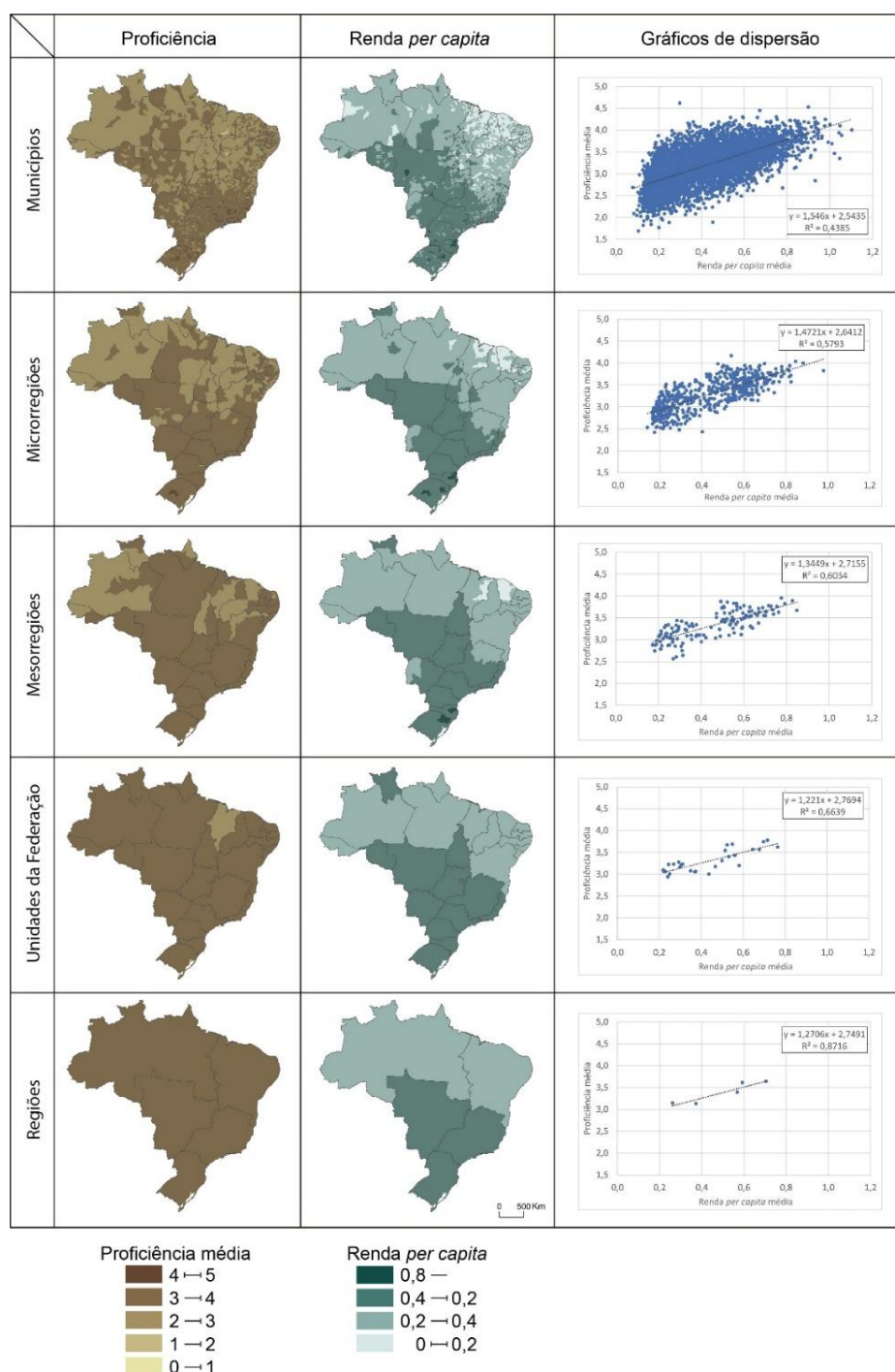


Figura 4: Proficiência no ENEM e renda *per capita* em diferentes zoneamentos (2012-2018). Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

Em cada zoneamento abordado neste estudo, o coeficiente de correlação linear de Pearson entre as variáveis proficiência e renda familiar *per capita* foi calculado. Na correlação individual, o valor do coeficiente independe da localização de cada indivíduo. Ao agrupar os concluintes em zonas, o posicionamento geográfico assumiu maior importância.

No agrupamento das Unidades da Federação (UF) em regiões, os estados das regiões Norte e Nordeste apresentaram renda *per capita* e proficiências médias inferiores às dos alunos das outras regiões (**Tabela 8**). Na mudança de escala de agrupamento, os valores da correlação ecológica ampliaram-se, evidenciando que a localização das zonas interferiu no aumento do valor da correlação ecológica.

Tabela 8: Correlações ecológicas entre a médias da renda *per capita* e proficiência por UF e por região.

Estado (UF)	Média (UF)		Região	Média (Região)	
	Renda <i>per capita</i>	Proficiência		Renda <i>per capita</i>	Proficiência
Alagoas	0,25	3,01	Nordeste	0,26	3,15
Bahia	0,31	3,22			
Ceará	0,22	3,09			
Maranhão	0,24	2,94			
Paraíba	0,24	3,23			
Pernambuco	0,29	3,28			
Piauí	0,23	3,06			
Rio Grande do Norte	0,31	3,23			
Sergipe	0,27	3,23			
Acre	0,37	3,06	Norte	0,37	3,13
Amapá	0,35	3,08			
Amazonas	0,37	3,05			
Pará	0,30	3,16			
Rondônia	0,53	3,40			
Roraima	0,47	3,17			
Tocantins	0,44	3,00			
Distrito Federal	0,72	3,78			
Goiás	0,50	3,31			
Mato Grosso	0,58	3,20	Centro Oeste	0,57	3,39
Mato Grosso do Sul	0,56	3,43			
Espírito Santo	0,51	3,54			
Minas Gerais	0,52	3,67			
Rio de Janeiro	0,55	3,69			
São Paulo	0,65	3,57			
Paraná	0,68	3,56			
Rio Grande do Sul	0,70	3,74			
Santa Catarina	0,76	3,62			
Correlação ecológica por UF	0,814791		Correlação ecológica por região	0,933589	

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

O agrupamento das unidades administrativas diminuiu a amplitude das médias da renda familiar *per capita* e da proficiência (**Tabela 8**) e resultou no aumento dos valores das correlações. Isto corrobora as observações de Clark e Avery (1976), segundo os quais a variação da escala de agrupamento, aparentemente, tende a diminuir a heterogeneidade entre os pares de variáveis e reduzir a autocorrelação espacial, quando regiões maiores são formadas.

Neste contexto, a pesquisa observou as recomendações de Manley (2013): adotar escalas analíticas explícitas e avaliar o fenômeno em múltiplas escalas. Também foi coerente com as divisões territoriais brasileiras, que agregou áreas a partir de características comuns (IBGE, 1984). A “organização do espaço refere-se às diferentes estruturas espaciais resultantes da dinâmica da sociedade sobre um suporte territorial” (IBGE, 1990, p. 9).

4.4 Falácias

De maneira geral, os concluintes das escolas públicas de até 18 anos das regiões Norte e Nordeste apresentaram menor renda *per capita* e menor desempenho no ENEM do que os estudantes de outras regiões. Porém, atribuir ao indivíduo o resultado médio de uma área (região) pode caracterizar a falácia seletiva. A falácia ecológica, descrita por Robinson (1950), refere-se à incorreta suposição de que a relação entre as variáveis observadas na forma agrupada ou ecológica, em um zoneamento, é a mesma que a dos dados individuais (JARGOWSKY, 2005).

Assim, quando se escolhe uma zona específica (ex.: município), a atribuição do desempenho médio deste a um indivíduo, pode também resultar numa falácia seletiva. Em Ipanguaçu (RN), por exemplo, a proficiência média dos alunos concluintes do ensino médio no ENEM foi 4,3. Possíveis erros de interpretação podem ocorrer porque o desempenho elevado derivou, principalmente, das proficiências

obtidas pelos alunos de uma única escola. Houve também alunos com pleno desempenho, em escolas de proficiência média baixa, e alunos pouco proficientes na escola de maior proficiência média (**Tabela 9**).

Tabela 9: Proficiências médias por escola do município de Ipanguaçu - RN (2012-2018).

Escola	alunos	Proficiência		
		Média	Mínima	Máxima
EE João Francisco da Costa	14	2,29	0	5
EE Manoel de Melo Montenegro	215	2,83	0	5
EE Maria da Glória de A. Luna	21	1,90	0	3
IF de Educação, Ciência e Tecnologia do RN – Campus Ipanguaçu	478	4,68	2	5

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

Com resultados médios extremos, os municípios de Cocal dos Alves (PI) e de Jatobá (MA) indicam a polarização da proficiência no ENEM, com elevado e baixo desempenhos dos alunos. Assim, a partir da **Tabela 10**, não seria possível fazer qualquer afirmação individual de desempenho: de que não houve alunos com proficiência plena no município de pior rendimento e de que o município de melhor rendimento médio no Brasil não teve alunos de baixo desempenho (falácia ecológica).

Tabela 10: Municípios com maior e menor proficiência média (2012-2018).

Município	Escolas	Alunos	Proficiência		
			Média	Mínima	Máxima
Cocal dos Alves (PI)	1	105	4,6	2	5
Jatobá (MA)	3	154	1,7	0	5

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

Os melhores exemplos da possibilidade da ocorrência de falácias de escala, considerando as variáveis deste estudo, podem ser encontrados no estado de Roraima. Seus 15.255 alunos concluintes do ensino médio pontuaram em todas as provas e na redação e obtiveram média de 3,17 de proficiência. A microrregião de Boa Vista, composta por quatro municípios, contém 85,3% dos concluintes do ensino médio do estado que pontuaram no ENEM. A proficiência média acima de três revelou o melhor desempenho médio dos concluintes do município de Boa Vista, que corresponde a 80,5% dos concluintes desta microrregião. Destarte, a proficiência média acima de três da zona não poderia ser atribuída aos outros três municípios que integram a referida microrregião (**Tabela 11**).

Tabela 11: Proficiência média da microrregião de Boa Vista - RR (2012-2018).

Município	Alunos	Proficiência média
Alto Alegre	401	2,57
Amajari	133	2,87
Pacaraima	193	2,97
Boa Vista	12.281	3,30

Fonte: INEP (2013, 2015a, 2015b, 2017a, 2017b, 2018a, 2019a).

Da mesma forma, a mesorregião Norte de Roraima apresenta proficiência média de 3,23 e é composta por duas microrregiões: Boa Vista, cujos 13.008 alunos obtiveram a proficiência média de 3,26, e o Nordeste de Roraima, cuja proficiência média de 564 alunos foi de 2,59.

5. Conclusões

Uma avaliação do desempenho de estudantes brasileiros de escolas públicas concluintes do ensino médio, de 16 a 18 anos, no ENEM, no período de 2012 a 2018 foi realizada. A análise e representação dos dados individuais dos candidatos e dos agrupamentos em unidades administrativas permitiu observar a variação das médias e das correlações em unidades de áreas modificáveis, de municípios a regiões, e identificar possíveis falácias na interpretação de dados agrupados.

Constatou-se que a pontuação média dos alunos concluintes melhorou no período analisado, mas o total de alunos com elevada proficiência não aumentou na mesma proporção. Das variáveis analisadas, a renda

familiar *per capita* apresentou maior correlação com o desempenho no ENEM. Verificou-se também que a maioria dos concluintes do ensino médio de até 18 anos não obteve proficiência plena, que seria pré-requisito para a obtenção do certificado do ensino médio. A certificação da proficiência plena ou parcial, foi determinada pelo INEP (2014) para os alunos acima de 18 anos que não concluíram o ensino médio, com base nos resultados obtidos nas provas objetivas e na redação no ENEM.

Também foi constatado que apenas 56,7% dos alunos concluintes do ensino médio de até 18 anos participaram do ENEM, de 2012 a 2018. Por essa razão, acredita-se que a baixa participação poderia reduzir o potencial de uso dos resultados do teste educacional como parâmetro de referência da educação no Brasil.

Uma associação entre renda e desempenho no ENEM foi observada. De modo geral, as menores proficiências do Brasil ocorreram nos municípios e microrregiões do Norte e do Nordeste do Brasil. Nestes agrupamentos territoriais, a renda *per capita* média é frequentemente menor (até dois salários mínimos). Apesar dos valores de correlação terem aumentado nos zoneamentos de maior agregação (ex.: estados), a percepção desta associação entre as variáveis diminuiu nas representações cartográficas. A análise em múltiplas escalas permitiu observar essas diferenças e refletir sobre o MAUP e as falácias inerentes à representação espacial em distintas unidades de agregação.

Agradecimentos

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela bolsa de mestrado que possibilitou o desenvolvimento da pesquisa, e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pela bolsa de produtividade (processo: 307725/2020-0).

Referências bibliográficas

ALKER, H. R. A Typology of Ecological Fallacies. In: DOGAN, M.; ROKKAN, S. (Eds.). **Quantitative Ecological Analysis in the Social Sciences**. Cambridge: MIT Press, 1969. p. 69-86.

BRASIL. Ministério da Educação. Portaria n° 438 de 28 de maio de 1998. Diário Oficial da União, Brasília, 1º maio 1998. Seção 1, p. 5.

BRASIL. Portaria n° 462 de 27 de maio de 2009. Diário Oficial da União, Brasília, 28 maio 2009. Seção 1, p. 54.

BRASIL. Lei n° 12.799 de 10 de abril de 2013. Dispõe sobre a isenção de pagamento de taxas para inscrição em processos seletivos de ingresso nos cursos das instituições federais de educação superior. Diário Oficial da União, Brasília, 10 abr. 2013. Seção 1, p. 1.

BURT, C. Experimental Tests of Higher Mental Processes and their Relation to General Intelligence. **Journal of Experimental Pedagogy**, v. 1, p. 93-112, 1911.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 6.ed. São Paulo: Saraiva, 2010.

CLARK, W. A. V; AVERY, K. L. The Effects of Data Aggregation in Statistical Analysis. **Geographical Analysis**, v. 8, n. 4, p. 428-438, 1976.

COLLINS, J. H.; DOUGLASS, H. R. The Socio-Economic Status of the Home as a Factor in Success in the Junior High School. **The Elementary School Journal**, v. 38, n. 2, p. 107-113, out. 1937. DOI: 10.1086/462139.

CONTRERAS, D. et al. Income Inequality or Performance Gap? A Multilevel Study of School Violence in 52 Countries. **Journal of Adolescent Health**, v. 57, n. 5, p. 545-552, 2015. DOI: 10.1016/j.jadohealth.2015.08.002.

DUFF, J. F.; THOMSON, G. H. The Social and Geographical Distribution of Intelligence in Northumberland. **British Journal of Psychology**, General Section, v. 14, n. 2, p. 192-198, out. 1923. DOI: 10.1111/j.2044-8295.1923.tb00126.x.

FERREIRA, M. C. **Iniciação à análise geoespacial: teoria, técnicas e exemplos para geoprocessamento**. São Paulo: Editora Unesp, 2014.

FLEMING, C. M. Socio-Economic Level and Test Performance. **British Journal of Educational Psychology**, v. 13, n. 2, p. 74-82, 1943. DOI: 10.1111/j.2044-8279.1943.tb02724.x.

GALTON, F. Co-Relations and Their Measurement, Chiefly from Anthropometric Data. **Proceedings of the Royal Society of London**, v. 45, p. 135-145, 1889. DOI: 10.1098/rspl.1888.0082.

GEHLKE, C. E.; BIEHL, K. Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. **Journal of the American Statistical Association**, v. 29, n. 185, p. 169, 1934. DOI: 10.1080/01621459.1934.10506247.

IBGE. **Divisão do Brasil em mesorregiões e microrregiões geográficas**. Departamento de Geografia. Rio de Janeiro, 1990. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv2269_1.pdf>. Acesso em: 12 dez. 2019.

IBGE. **Geoprocessamento no IBGE**. Rio de Janeiro, 1994. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv24556.pdf>>. Acesso em: 2 set. 2018.

INEP. Portaria nº 179, de 28 de abril de 2014. **Diário Oficial da União**, Brasília, 29 abr. 2014. Seção 1, p. 40

INEP. Microdados do ENEM 2012. Brasília, 2013. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 1º mar. 2017.

INEP. Microdados do ENEM 2013. Brasília, 2015a. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 7 jun. 2017.

INEP. Microdados do ENEM 2014. Brasília, 2015b. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 7 jun. 2017.

INEP. Microdados do Censo da Educação Básica 2012. Brasília, 2016a. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 21 maio. 2018.

INEP. Microdados do Censo da Educação Básica 2013. Brasília, 2016b. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 21 maio. 2018.

INEP. Microdados do Censo da Educação Básica 2014. Brasília, 2016c. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 19 maio. 2018.

INEP. Microdados do Censo da Educação Básica 2015. Brasília, 2016d. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 21 maio. 2018.

INEP. Microdados do Censo da Educação Básica 2016. Brasília, 2016e. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 21 maio. 2018.

INEP. Microdados do ENEM 2015. Brasília, 2017a. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 29 maio. 2017.

INEP. Microdados do ENEM 2016. Brasília, 2017b. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 10 maio. 2018.

INEP. Microdados do ENEM 2017. Brasília, 2018a. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 21 maio. 2018.

INEP. Microdados do Censo da Educação Básica 2017. Brasília, 2018b. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 26 maio. 2018.

INEP. Microdados do ENEM 2018. Brasília, 2019a. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 15 jul. 2019.

- INEP. Microdados do Censo da Educação Básica 2018. Brasília, 2019b. Disponível em: <<http://inep.gov.br/microdados>>. Acesso em: 15 jul. 2019.
- MINISTÉRIO DA EDUCAÇÃO. Portaria nº 807 de 18 jun 2010. Disponível em: <https://download.inep.gov.br/educacao_basica/enem/legislacao/2010/portaria807_180610.pdf>. Acesso em: 8 maio 2021.
- JARGOWSKY, P. A. Ecological Fallacy. **Encyclopedia of Social Measurement**, v. 1, p. 715-722, 2005.
- JELINSKI, D. E.; WU, J. The Modifiable Areal Unit Problem and Implications for Landscape Ecology. **Landscape Ecology**, v. 11, n. 3, p. 129-140, jun. 1996. DOI: 10.1007/BF02447512.
- JONES, D. C.; CARR-SAUNDERS, A. M. The Relation between Intelligence and Social Status among Orphan Children. **British Journal of Psychology, General Section**, [S. l.], v. 17, n. 4, p. 343-364, 1927. DOI: 10.1111/j.2044-8295.1927.tb00437.x.
- MANLEY, D. Scale, Aggregation, and the Modifiable Areal Unit Problem *in* Fischer, M.; NIJKAMP, P. (eds). *Handbook of Regional Science*, p. 1157-1171. Berlin, 18 jul. 2013. DOI: 10.1007/978-3-642-23430-9_69.
- NELSON, J.K.; BREWER, C.A. Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem. **Cartography and Geographic Information Science**, v. 44, p. 35-50, 12 out. 2015. DOI: 10.1080/15230406.2015.1093431.
- NEPRASH, J. A. Some Problems in the Correlation of Spatially Distributed Variables. **Journal of the American Statistical Association**, v. 29, n. 185, p. 167-168, 1934. DOI: 10.1080/01621459.1934.10506246.
- OPENSHAW, S. The Modifiable Areal Unit Problem. **Concepts and Techniques in Modern Geography**, v. 38, p. 169-174, 1984.
- PEARSON, K. VII. Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. **Philosophical Transactions of the Royal Society of London**. Series A, Containing Papers of a Mathematical or Physical Character, v. 187, p. 253-318, 31 dez. 1896.
- ROBINSON, W. S. Ecological Correlations and the Behavior of Individuals. **American Sociological Review**, v. 15, n. 3, p. 351-357, jun. 1950. DOI: 10.2307/2087176.
- TOBLER, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. **Economic Geography**, v. 46, n. 332, p. 234-240, 14 jun. 1970. DOI: 10.2307/143141.
- TRAVITZKI, R. Avaliação da qualidade do ENEM 2009 e 2011 com técnicas psicométricas. **Estudos em Avaliação Educacional**, v. 28, n. 67, p. 256-288, 2017. DOI: 10.18222/eae.v28i67.3910.
- TRAVITZKI, R.; CALERO, J.; BOTO, C. ¿Qué información proporciona el examen nacional de enseñanza media (Enem) a la sociedad Brasileña? **Cepal Review**, n. 113, p. 163-181, 2014. DOI: 10.18356/a8e61670-es.
- WHITE, K. R. The Relation between Socioeconomic Status and Academic Achievement. **Psychological Bulletin**, v. 91, n. 3, p. 461-481, 1982. DOI: 10.1037/0033-2909.91.3.461.
- YULE, G. U.; KENDALL, M. G. **An Introduction to the Theory of Statistics**. 5.ed. London: Charles Griffin, 1950. DOI: 10.1017/S0020268100011033.



Este artigo é distribuído nos termos e condições do *Creative Commons Attributions/Atribuição-NãoComercial-CompartilhaIgual* (CC BY-NC-SA).