

Artigos Originais

AVALIAÇÃO DE IMPACTO DO PROGRAMA ENSINO MÉDIO INOVADOR SOBRE A FREQUÊNCIA ESCOLAR DOS ALUNOS EM SANTA CATARINA

Original Articles

IMPACT EVALUATION OF THE INNOVATIVE HIGH SCHOOL PROGRAM - PROEMI ABOUT STUDENTS' FREQUENCY IN SCHOOLS AT SANTA CATARINA STATE

Jeniffer Gonçalves*

jeniffer.g@posgrad.ufsc.br

<http://lattes.cnpq.br/3078402175699546>

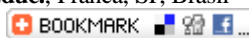
Francis Carlo Petterini Lourenço**

f.petterini@ufsc.br

<http://lattes.cnpq.br/3396205737621572>



CAMINE: Cam. Educ. = CAMINE: Ways Educ., Franca, SP, Brasil - eISSN 2175-4217 -
está licenciada sob [Licença Creative Commons](#)



RESUMO: Em 2009 o Ministério da Educação lançou o Programa Ensino Médio Inovador (ProEMI), uma política pública com proposta de apoiar e fortalecer o desenvolvimento de currículos inovadores nas escolas de ensino médio estaduais. Um dos objetivos do ProEMI é a redução da taxa de abandono escolar, buscando tornar o ensino médio um lugar conectado com as necessidades dos jovens, contribuindo assim para a permanência do aluno na escola. O abandono escolar não é uma atitude repentina, o estudante costuma faltar as aulas com muita frequência antes do abandono completo, de modo que a frequência escolar funciona como um bom preditivo para o abandono. Considerando isso, este trabalho se propôs a investigar o efeito da participação no ProEMI na frequência escolar dos alunos de Santa Catarina no ano de 2011. Foram utilizados três modelos econométricos comuns na análise de dados de contagem, são eles: Poisson, binomial negativa e Poisson inflado de zeros. Foram consideradas apenas as faltas nas disciplinas de Língua Portuguesa e Matemática por serem, geralmente, as disciplinas de maior destaque ao longo do ciclo escolar, e entram nos exames de proficiência padrão. Os resultados estimados mostram que o ProEMI tem efeitos positivos sobre a frequência escolar nos três modelos, reduzindo o número de faltas nas

* Doutoranda do PPGeco/UFSC

** Doutorado em Economia, professor na Universidade Federal de Santa Catarina (UFSC), vice-coordenador do Programa de Pós-Graduação em Economia da UFSC e bolsista de produtividade em pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico.

duas disciplinas analisadas, indicando que o programa contribui para manter os alunos na escola.

Palavras-chave: Avaliação de impacto. Modelos de dados de contagem. Educação. ProEMI.

ABSTRACT: In 2009 the Ministry of Education launched the *Innovative High School Program* (ProEMI), a public policy with the purpose of supporting and strengthen the development of innovative curricula of state high schools. One of the objectives of ProEMI is to reduce the dropout rate, seeking to make high school a place connected with the needs of young people, thus contributing to the student's permanence in school. Dropout is not a sudden attitude, the student often skips classes before dropping out, therefore school attendance works as a good predictor for dropouts. This study aimed to investigate the effect of participation at ProEMI in the school attendance of students from Santa Catarina in 2011. Three common econometric models were used in the analysis of counting data (Poisson, Negative Binomial and Zero-Inflated Poisson). Only absences in the Portuguese and Mathematics subjects were considered because they are generally the most prominent subjects throughout the school cycle, and they enter at most of the standard proficiency exams. The estimated results show that ProEMI has positive effects on school attendance in all three models, reducing the number of absences in both subjects, indicating that the program contributes to keeping students in school.

Keywords: Impact evaluation. Count data models. Education. ProEMI.

INTRODUÇÃO

A Constituição Federal de 1988 instituiu a progressiva extensão da obrigatoriedade e gratuidade ao ensino médio no Brasil. Em uma Emenda Constitucional de 1996 foi determinada a progressiva universalização dessa modalidade de ensino, passando a ser considerada como etapa final do ensino básico. Em 2009, a educação básica obrigatória e gratuita dos 4 aos 17 anos de idade passou a ser assegurada, inclusive para todos os que não tiveram acesso a ela em idade própria (BRASIL, 1988; 1996).

Reconhecer o ensino médio como parte integrante da educação básica, é incluí-lo no processo de escolarização que tem como finalidade principal o desenvolvimento do indivíduo, que lhe assegura a formação comum indispensável para o exercício da cidadania, oferecendo a base necessária para progredir no trabalho e estudos posteriores. Não há dúvidas sobre a importância da educação geral como meio para preparar os indivíduos para o trabalho e formar pessoas capacitadas para a inserção na sociedade.

Entretanto, apesar do esforço despendido na expansão de todas as modalidades de ensino, a etapa do ensino médio ainda se encontra distante da universalização proposta na Constituição Federam de 1988.

Em 2010, o Instituto Unibanco publicou um compilado de estudos sobre os problemas de evasão escolar, intitulado *A Crise de Audiência no Ensino Médio* (INSTITUTO UNIBANCO, 2010), mostrando que em 2007 cerca de 20% dos jovens de 15 a 17 anos encontrava-se fora da escola e dentre aqueles que terminaram o ensino fundamental, mas não ingressaram no médio, mais da metade não estava trabalhando, o que, a princípio, contrapõe a hipótese geralmente sugerida de que os jovens abandonam a escola para trabalhar, assim, se estes jovens não estão trabalhando, poderiam estar frequentando o ensino médio.

Bridgeland, DiIulio Jr e Morison (2006) buscando entender as raízes do abandono escolar, entrevistaram pessoas que abandonaram a escola durante o ensino médio nos Estados Unidos. Dos entrevistados, 47% falaram que a principal razão para terem abandonado a escola foi a de que as aulas não eram interessantes; 69% falaram que não sentiram-se motivados a se esforçar (2/3 afirmaram que teriam se esforçado se os professores tivessem demandado mais deles, e 70% asseguraram que teriam terminado em ensino médio se tivessem tentado).

Alguns entrevistados apresentaram motivos pessoais: 32% falaram que tiveram que parar para trabalhar, 26% falaram que tiveram filho(os) e 22% tiveram que cuidar de algum membro da família. O estudo mostrou também que o abando escolar não costuma ser um ato repentino, pelo contrário, antes do abandono completo há um processo gradual de afastamento, de modo que o padrão de frequência escolar fornece um claro preditivo para o abandono. Dos entrevistados, 65% afirmaram que faltavam as aulas frequentemente no ano antes do abandono.

Analisando as taxas de abandono escolar em Portugal, Caetano (2005) encontra que as causas que levam um aluno a abandonar a escola dependem principalmente de questões geográficas e econômicas. Alunos de região urbana tendem a abandonar por motivos de trabalho, já aqueles de zonas rurais por motivos relacionados a dificuldade econômica. Mas a questão da falta de motivação pelos estudos também é apontada pelos alunos.

Em um estudo sobre as taxas de evasão escolar nos países da América Latina, Kattan e Székely (2015) constataram que a evasão no ensino médio aumentou consideravelmente, apesar de a taxa de matrícula ter crescido. Entre os motivos encontrados para o aumento na taxa de evasão estão o despreparo das escolas em conseguir atender as necessidades individuais da crescente porção de jovens ingressando no ensino médio. Para os autores, a eficácia dos sistemas educacionais em equipar os jovens com capital humano valorizado pelo mercado de trabalho para que sejam engajados no setor produtivo é o segredo para manter os jovens na escola.

Buscando tornar o ensino médio um lugar conectado com as necessidades dos jovens, contribuindo assim para a permanência do aluno na escola, o Ministério da Educação propôs em 2009 o Programa Ensino Médio Inovador (ProEMI), como parte do Plano de Desenvolvimento da Educação (PDE), para desenvolver ações voltadas para a melhoria do ensino médio. O Programa tem como objetivo apoiar e fortalecer o desenvolvimento de propostas curriculares inovadoras nas escolas de ensino médio estaduais. Buscando promover a formação integral dos estudantes melhorando a qualidade do ensino médio.

A adesão ao ProEMI se dá de forma voluntária pelas Secretarias de Educação Estaduais e Distritais, que escolhem as unidades escolares que receberão apoio técnico e financeiro, por meio do Programa Dinheiro Direto na Escola (PDDE), para elaboração e implementação das Propostas de Redesenho Curricular (PRC). Além do conteúdo-base para a etapa de ensino médio, as propostas devem atender também as necessidades reais da escola, contemplando as especificidades do contexto local/regional.

Conforme o documento orientador (MINISTÉRIO DA EDUCAÇÃO, 2009), as ações propostas podem apresentar diferentes formatos, como: disciplinas optativas, oficinas, clubes de interesse, seminários integrados, grupos de pesquisa, trabalhos de campo, e outras ações interdisciplinares. Para conseguir atender às propostas, as escolas devem ampliar gradualmente a carga horária de 2.400 para 3.000 distribuídas ao longo dos três anos, considerando 200 dias letivos por ano, transformando em ensino de período integral.

Em 2010, 18 escolas estaduais participavam do ProEMI em Santa Catarina, em 2016 já eram 155 unidades escolares, com mais de 15.000 alunos sendo atendidos. Dado esse crescimento na adesão das escolas ao programa, pressupõe-se que este vem

atendendo às expectativas que a ele foram atribuídas. Desta forma, o presente trabalho propõe uma análise do efeito do Programa Ensino Médio Inovador na frequência escolar dos alunos do ensino médio. Como já apresentado, a frequência escolar funciona como um preditivo para o abandono escolar, assim espera-se que se a participação no ProEMI tem um impacto positivo, reduzindo o número de faltas dos alunos em comparação com as escolas que não participam, então deve reduzir a taxa de abandono escolar também.

Este trabalho divide-se em mais 4 seções. A Seção 0 apresenta a descrição dos dados que serão utilizados na análise. A Seção 0 apresenta as metodologias para modelos com dados de contagem e de avaliação de impacto. A Seção 0 traz os principais resultados encontrados com a análise. Na Seção 0 são feitas algumas considerações finais. O trabalho também conta com um Apêndice com os resultados completos das estimações.

DESCRIÇÃO DOS DADOS

Para conseguir estimar o efeito do ProEMI na frequência dos alunos serão utilizados dados referentes as características das escolas e dos alunos do estado de Santa Catarina no programa, disponibilizados pela Secretaria do Estado de Educação (SED) e pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O Quadro 1 apresenta as variáveis utilizadas, relacionadas por escola e aluno.

A variável “ProEMI” corresponde a uma *dummy* para as escolas que participaram do Programa Ensino Médio Inovador no estado de Santa Catarina em 2011. O número de funcionários será utilizado como uma *proxy* para o tamanho da escola. Também será considerada a estrutura física das escolas, representada na presença de biblioteca, laboratórios de ciência e de informática, quadra esportiva e na proporção média de alunos por turma em cada escola. Além disso, tem-se também a taxa média de abandono e a distorção média idade-série por escola em 2011. Já em relação aos estudantes, a base de dados apresenta o número de faltas que cada aluno teve em cada disciplina no ano de 2011.

Serão consideradas neste trabalho as disciplinas de Língua Portuguesa e Matemática por serem, geralmente, as disciplinas de maior destaque ao longo do ciclo escolar, e entram nos exames de proficiência padrão. Também será considerada a etnia

do aluno (branca, preta, parda, indígena ou outra), se o aluno mora na zona urbana ou rural e a idade de cada estudante em 2011.

Quadro 1 - Covariadas referentes as escolas e alunos

Escola	Aluno
ProEMI	Nº de faltas
Nº de funcionários	Cor
Biblioteca	Zona
Laboratório de Ciências	Idade
Laboratório de Informática	
Quadra Esportiva	
Aluno por turma	
Taxa de abandono	
Distorção idade-série	

Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço

Comparando o número de faltas entre os alunos que estudam em escolas que participam do ProEMI e aqueles que estudam em outras escolas, nas disciplinas de Língua Portuguesa e Matemática, percebe-se pela *Tabela 1* que o percentual de faltas vai ficando maior para as escolas sem ProEMI, em relação as escolas com ProEMI, conforme o número de faltas aumenta. Por exemplo, na disciplina de língua portuguesa, o percentual de alunos que sem faltas ao longo do ano é maior nas escolas fora do programa, e tem percentual menor de alunos com até 9 faltas, por outro lado, quando considerado os alunos com 10 faltas ou mais, as escolas sem ProEMI superam as escolas com ProEMI. Já com relação a disciplina de matemática não é possível identificar um padrão, pois o percentual de faltas oscila entre os dois grupos de escolas, sendo que o percentual de alunos com 11 faltas ou mais é superior nas escolas fora do programa.

Tabela 1 - Percentual de faltas por disciplina no ano de 2011

Português			Matemática		
Faltas	Sem ProEMI	Com ProEMI	Faltas	Sem ProEMI	Com ProEMI
0	12,77	10,36	0	12,12	11,02
1	7,45	8,04	1	7,35	8,44
2	8,81	9,45	2	9,16	11,76
3	7,95	9,63	3	7,77	8,38
4	7,36	8,10	4	7,47	7,37
5	6,35	6,83	5	6,40	6,69
6	5,92	7,04	6	5,73	6,69
7	5,00	5,57	7	5,00	4,79
8	4,34	4,49	8	4,48	3,99
9	4,06	4,61	9	4,01	3,87
10	3,60	3,34	10	3,34	3,59
≥11	26,39	22,54	≥11	27,17	23,41

Fonte: Secretaria da Educação. Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço

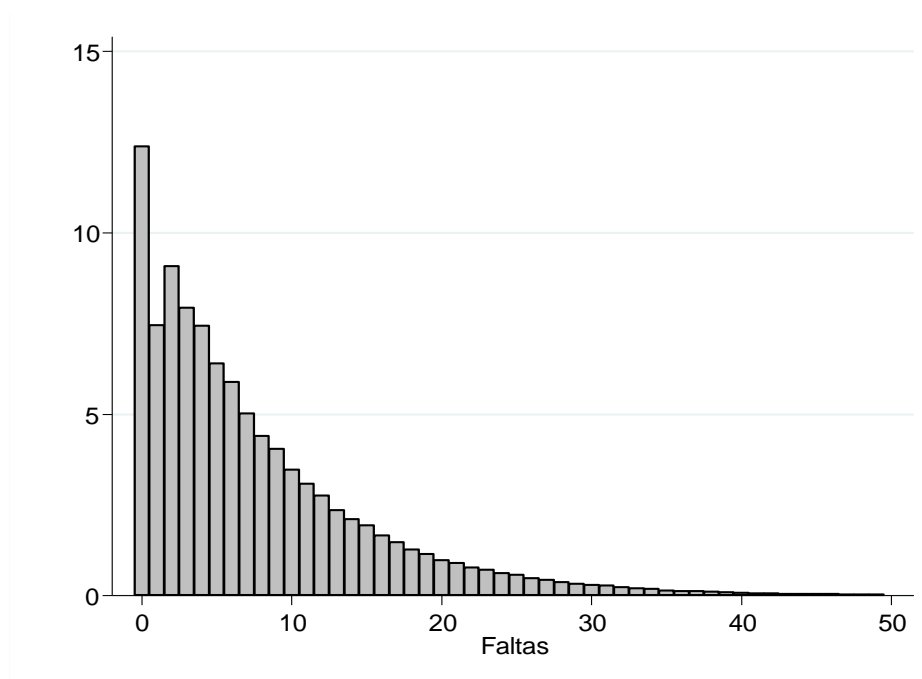
Assim, não se pode afirmar de forma direta que os alunos que frequentam escolas participantes do ProEMI têm assiduidade maior às aulas. Será necessária uma análise mais profunda, que consiga isolar o efeito do ProEMI na frequência dos alunos. A próxima seção apresenta a metodologia que será utilizada para conseguir estimar o efeito do programa.

METODOLOGIA

A variável dependente nesta análise é o número de faltas dos alunos ao longo do ano, um valor inteiro e não negativo (i.e. 0,1,2,3... faltas). Esse tipo de dado é conhecido como dado de contagem e existem metodologias apropriadas para abordar este tipo variável, as principais são os modelos Poisson ou binomial negativa. A regressão binomial negativa é uma generalização da regressão de Poisson, relaxando a suposição restritiva feita pelo modelo de Poisson de que a variância é igual a média.

É muito comum em modelos com dados de contagem a grande frequência do número zero. Quando a variável dependente do modelo é inflada de zeros, os modelos Poisson e binomial negativa iram subestimar a previsão. A *Figura 1* traz um histograma da frequência de faltas nas disciplinas de Língua Portuguesa e Matemática.

Figura 1 - Frequência do número de faltas



Fonte: Secretaria da Educação. Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço.

A regressão Poisson Inflado de Zeros (*Zero-Inflated Poisson – ZIP*) é utilizada em modelos com dados de contagem que exhibe um número excessivo de zeros. A distribuição dos dados combina uma distribuição Poisson e uma distribuição logística. Já a regressão binomial negativa é baseada em uma mistura das distribuições Poisson e gamma.

Além disso, para conseguir estimar o efeito do ProEMI será necessário utilizar uma técnica de pareamento estatístico, selecionando as escolas que não aderiram ao programa, mas que apresentam características semelhantes àquelas que participaram. O pareamento busca reduzir o viés, causado por variáveis “confundidoras”, que pode ser encontrado em uma estimativa do efeito da adesão ao programa obtida a partir de uma simples comparação entre as escolas que aderiram ao programa e as que não aderiram.

Portanto antes de apresentar os detalhes sobre os modelos de contagem, será abordada a técnica a ser utilizada para parear as escolas. Assim, a seção 0 apresenta a técnica de pareamento e a seção 0 apresenta os modelos de contagem (Poisson, binomial negativa e ZIP), conforme apresentado por Cameron e Trivedi (2005).

AVALIAÇÃO DE IMPACTO

Conforme Cameron e Trivedi (2005), a forma mais natural de mensurar o efeito de um tratamento é construindo uma medida que compara os resultados médios dos grupos tratados e não tratados. Quando a escolha do grupo a ser tratado é feita de forma aleatória, a simples diferença de médias dos resultados entre os dois grupos pode ser considerada para verificar o efeito do tratamento. Entretanto, as escolas que aderiram ao ProEMI não foram escolhidas aleatoriamente, partiu delas o interesse em participar, o que pode gerar o chamado “viés de seleção”. A comparação entre os dois grupos necessita, então, de algumas hipóteses.

O resultado para as escolas que aderiram ao programa será denotado como y_1 e para as que não aderiram como y_0 . Uma hipótese importante a ser assumida é a de independência condicional, que afirma que dado o conjunto de variáveis explicativas x , os resultados são independentes do tratamento, isto é

$$y_0, y_1 \perp T \mid x \quad (1)$$

onde T é uma *dummy* que identifica o tratamento. Se válida, a hipótese (1) pode ser utilizada para a identificação de alguns parâmetros de impacto, pois afirma que uma vez controlados os efeitos dos regressores em x , o tratamento e o resultado são independentes.

Uma hipótese mais fraca do que (1) é:

$$y_0 \perp T \mid x, \quad (1)$$

que implica independência condicional de y_0 . Essa hipótese é utilizada para identificar a média populacional do efeito do tratamento no grupo tratado (*average treatment effect on the treated* – ATET). A validade da hipótese (1) implica que não há viés por variável omitida, uma vez que X é incluída na regressão.

Uma segunda hipótese é necessária para identificar algumas medidas de impacto populacional, a hipótese de suporte comum:

$$0 < Pr[T = 1 | x] < 1. \quad (2)$$

Essa hipótese garante que para cada valor X existem casos tratados e não tratados. Isto é, para cada escola que aderiu ao ProEMI existe uma outra escola que não participou do programa com um X similar. Essa hipótese é necessária, pois para identificar o efeito do tratamento deve-se ter, para cada participante do programa, um não participante.

Uma terceira hipótese é a independência da média condicional:

$$E[y_0 | T = 1, x] = E[y_0 | T = 0, x] = E[y_0 | x], \quad (3)$$

que implica que y_0 não determina a participação no tratamento.

Pareamento por Escore de Propensão

Quando a participação no tratamento não é feita de forma aleatória, o conceito de escore de propensão é necessário (ROSENBAUM; RUBIN, 1983). O escore de propensão é uma medida de probabilidade condicional da participação no tratamento dado X e é denotada por

$$p(x) = Pr[T = 1 | X = x]. \quad (4)$$

A condição de balanceamento é uma hipótese com um papel importante na avaliação de tratamento. Ela pressupõe

$$T \perp x | p(x), \quad (5)$$

isto é, para as escolas com o mesmo escore de propensão, a participação no programa é aleatória e deve parecer idêntica em termos do vetor x .

Assim, essa hipótese postula que, ao comparar duas escolas, uma no grupo de controle (não tratada) e outra no grupo de tratamento, com características observáveis semelhantes, o único fator que diferencia os resultados entre as escolas é a participação ou não no programa. Segundo Rosenbaum e Rubin (1983), o ajuste pelo escore de propensão é suficiente para remover o viés causado pelas variáveis observáveis.

O escore de propensão pode ser estimado a partir de (T_i, x_i) , em geral via *logit* ou *probit*. A escolha das variáveis que farão o pareamento deve ser feita com muita cautela, o uso de um conjunto muito grande de variáveis deve ser evitado, pois embora a inclusão de variáveis não significativas na especificação do escore de propensão não viole as estimativas ou as torne inconsistentes, pode aumentar sua variância.

Um dos métodos mais comuns de pareamento é o do "vizinho mais próximo" (*nearest neighbor – NN matching*). Esse método é eficaz em selecionar indivíduos do grupo não tratado que estão pareados com o grupo de tratado e descarta os que não estão. Primeiro deve-se escolher quantas escolas do grupo de não tratadas deverão ser pareadas com cada escola tratada, e então elas serão selecionadas de acordo com a proximidade com os valores $p(x_i)$ do grupo de tratamento. Isso é feito da seguinte forma:

- a) Calcular $|p(x_i) - p(x_j)|$ para todas as observações i do grupo tratado e para todas as comparações j no grupo não tratado;
- b) Ordenar as observações j em termos de $|p(x_i) - p(x_j)|$ do menor para o maior;
- c) Selecionar o conjunto de observações x com os menores valores de $|p(x_i) - p(x_j)|$. Estes são os chamados vizinhos próximos.

O método de pareamento a ser utilizado será uma variação do método do vizinho mais próximo. Serão escolhidas duas escolas que não participaram do ProEMI para cada escola que participou, e a escolha será feita da seguinte forma:

- i. Calcular $\rho(x)$ a partir de um *probit* para todas as observações;
- ii. Ordenar as observações em termos de $\rho(x)$ do menor para o maior.
- iii. Selecionar, entre as escolas que não participaram do ProEMI, o conjunto de observações x com os valores de $\rho(x_j)$ imediatamente anteriores e imediatamente posteriores ao $\rho(x_i)$ das escolas que participaram.

Após feito o pareamento, tem-se um grupo tratado e um grupo não tratado, mas com características semelhantes que pode ser considerado com o contrafactual do grupo tratado.

MODELOS DE DADOS DE CONTAGEM

O modelo Poisson é o ponto inicial para a análise de dados de contagem, entretanto nem sempre ele é o mais adequado. Em alguns casos a alta proporção de zeros na amostra pode coexistir com valores muito altos de contagem, tornando difícil a modelagem. Primeiro será apresentado o modelo Poisson, depois as suas variações, para verificar quais são as principais diferenças.

Poisson

A regressão de Poisson é um modelo básico sob o qual vários modelos de contagem são baseados. É derivado da função de probabilidade de Poisson, que pode ser expressa por

$$f(y_i) = \frac{e^{-\mu_i} (\mu_i)^{y_i}}{y_i!}, \quad y = 0, 1, 2, \dots, \quad (6)$$

onde y_i corresponde a resposta da contagem, μ_i o parâmetro predito de contagem. Com $E[y] = \mu$ e $Var[y] = \mu$. A hipótese padrão de parametrização com média exponencial é

$$\mu_i = \exp(x_i' \beta), \quad i = 1, \dots, N. \quad (7)$$

A função log-verossimilhança é dada por

$$\ln L(\beta) = \sum_{i=1}^N [y_i x_i' \beta - \exp(x_i' \beta) - \ln y_i!]. \quad (8)$$

A condição de primeira ordem para a máxima verossimilhança é

$$\sum_{i=1}^N [y_i - \exp(x_i' \beta)] x_i = 0. \quad (9)$$

A característica chave do modelo de Poisson é a igualdade entre a média e a variância. Quando a variância de um modelo Poisson excede a sua média o modelo é determinado com grande dispersão. A regressão binomial negativa tem sido utilizada para acomodar essa dispersão.

Binomial Negativa

A regressão binomial negativa é uma generalização do modelo Poisson que afrouxa a suposição restritiva de que a variância é igual à média. O modelo de regressão binomial negativa tradicional é baseado em uma mistura das distribuições Poisson e gamma.

Suponha que a distribuição de uma variável de contagem aleatória Y é Poisson, condicional ao parâmetro λ , de forma que $f(y|\lambda) = \exp(-\lambda) \lambda^y / y!$. Suponha agora que o parâmetro λ é aleatório. Em particular, considere $\lambda = \mu V$, onde μ é uma função determinística de X , por exemplo $\exp(X' \beta)$, e $V > 0$ é iid com densidade $g(v|\alpha)$. Se $E[V] = 1$ então $E[\lambda V \mu] = \mu$, então a interpretação do parâmetro de inclinação permanece como no modelo de Poisson.

A densidade marginal de Y condicional aos parâmetros μ e α é obtida através da integração em V , chegando a

$$h(y | \mu, \alpha) = \int f(y | \mu, v)g(v | \alpha)dv, \quad (10)$$

onde $g(v | \alpha)$ é chamada de distribuição mista (*mixing distribution*) e α denota um parâmetro desconhecido dessa distribuição. Dependendo das escolhas específicas de $f(\cdot)$ e $g(\cdot)$, a integral terá uma solução explícita ou de forma fechada.

Os primeiros dois momentos da distribuição binomial negativa são:

$$\begin{aligned} E[y | \mu, \alpha] &= \mu, \\ \text{Var}[y | \mu, \alpha] &= \mu(1 + \alpha\mu) \end{aligned} \quad (11)$$

Como pode ser visto, a variância é maior do que a média, desde que $\alpha > 0$ e $\mu > 0$.

Poisson Inflado de Zeros

Os zeros nos modelos de dados de contagem podem ter mais de uma interpretação. Por exemplo, o número de faltas ser zero não implica, necessariamente, que tal escola é melhor do que aquelas com mais faltas. O modelo de Poisson inflado de zeros (ou ZIP) é utilizado para dados de contagem com alta dispersão e excesso de zeros. Esse excesso de zeros pode ter sido gerado a partir de processo de valores de contagem separado e o excesso de zeros pode ser modelado de forma independente. Assim, o modelo ZIP tem duas partes, um modelo de contagem de Poisson e um modelo logit para prever o excesso de zeros.

Suponha que para cada observação existam dois casos possíveis. Se o caso 1 acontecer, a contagem é zero, entretanto, se o caso 2 ocorrer, as contagens (incluindo zeros) são geradas de acordo como o modelo Poisson. Suponha que o caso 1 ocorra com probabilidade π e o caso 2 ocorra com probabilidade $(1 - \pi)$. Então, a distribuição de probabilidade da variável aleatória Y_i pode ser escrita como

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1 - \pi) \exp(-\mu_i) & \text{se } j = 0 \\ (1 - \pi_i)(\mu_i^{y_i} \exp(-\mu_i)) / (y_i!) & \text{se } j > 0 \end{cases}, \quad (12)$$

onde o Y_i corresponde ao número de faltas de cada aluno ($Y \geq 0$), μ é o parâmetro da média que pode ser escrito como

$$\mu_i = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}), \quad (13)$$

e π_i é a função logística que define a proporção de zeros, dada por

$$\pi_i = \frac{\lambda_i}{(1 + \lambda_i)}, \quad (14)$$

onde

$$\lambda_i = \exp(\gamma_1 z_{1i} + \gamma_2 z_{2i} + \dots + \gamma_m z_{mi}). \quad (15)$$

Os vetores de variáveis X 's e Z 's podem ou não incluir termos em comum.

PRINCIPAIS RESULTADOS

Em 2011, 18 escolas de Santa Catarina aderiram ao ProEMI, após tirar da base as observações com dados incompletos, restaram 8 escolas. O Pareamento foi feito com base nessas 8 escolas e seguindo os passos descritos na Seção 3.1.1 de forma que a base pareada é composta por dados de 24 escolas.

Para estimar o escore de propensão ($P(X)$), foi utilizado um modelo probit com a variável dependente sendo a *dummy* de tratamento (participação no ProEMI) e variáveis explicativas foram as características das escolas - número de funcionários, biblioteca, laboratório de ciências, laboratório de informática, quadra de esporte, distorção idade-série, taxa de abandono e a proporção aluno/turma. As escolas do grupo de tratamento e do grupo de controle são apresentadas na *Tabela 2*, juntamente com o escore de propensão estimado. As *Figura 2* e *Figura 3* (no Apêndice) apresentam, respectivamente, as densidades por escore de propensão antes e depois do pareamento das escolas entre o grupo de tratamento e o grupo de controle.

Tabela 2 - Grupos de tratamento e de controle pareados

Tratamento	E.P. (%)	Controle	E.P. (%)
EEB Gov. Heriberto Hulse	1,49	EEB Prof. Antonia Gasino de Freitas	1,46
		EEB Dr. Paulo Medeiros	1,49
EEB São Miguel	1,99	EEB São Francisco	1,98
		EEB Rodolfo Zipperer	1,99
EEB Osvaldo Aranha	4,62	EEB Prof. Osni Paulino da Silva	4,54
		EEB Paulo Blasi	4,63
EEB Prof. Jandira D. Ávila	6,90	EEB Gertrud Aichinger	6,77
		EEB Prof. Olegário Bernardes	6,97
EEB Almirante Barroso	7,21	EEB Santa Teresinha	7,16
		EEB Gomes Carneiro	7,34
EEB Luiz Delfino	9,47	EEB Prof. Jurema Savi Milanez	9,42
		EEB Alfredo Zimmermann	9,52
EEB N. Sra. da Salete	10,96	EEB Jose do Patrocinio	10,95
		EEB Aderbal Ramos Da Silva	11,79
EEB Prof. Nelson Horostecki	14,72	EEB São Ludgero	14,65
		EEB Olavo Bilac	15,23

Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço a partir dos resultados do modelo.

A partir disso, foram feitas regressões de Poisson, binomial negativa e Poisson inflada de zeros tanto para os dados pareados quanto para os não pareados, para as duas disciplinas separadamente. Todos os modelos utilizaram as mesmas variáveis explicativas – as características da escola: ProEMI, Número de funcionários, biblioteca, laboratório de ciências, laboratório de informática, quadra de esporte e aluno por turma; e características dos alunos: zona, idade e cor. A taxa de abandono escolar foi utilizada no modelo ZIP para diferenciar a origem da inflação de zeros, de forma que nas escolas que tiveram menor taxa de abandono, espera-se que os zeros correspondam, de fato, a 100% de presença dos alunos.

A *Tabela 3* apresenta os resultados para as estimativas do efeito do ProEMI no número de faltas em cada disciplina, os resultados completos estão no Apêndice. Para que a participação no ProEMI tenha o efeito de reduzir o número de faltas, o sinal do

parâmetro deve ser negativo. Pode-se ver pelos resultados da *Tabela 3* que a participação no ProEMI reduz o número de faltas nas duas disciplinas.

Analisando os resultados dos modelos Poisson, percebe-se que para a disciplina de Matemática o efeito nos dados pareados é um pouco maior do que nos dados não pareados, mas em ambos o sinal é negativo. Já para Língua Portuguesa, que também apresentou sinais negativos, o efeito no modelo não pareado mostrou-se maior do que naquele pareado. A interpretação destes parâmetros é de semi-elasticidade, mantendo todas as outras variáveis constantes. Então considerando o modelo pareado, as escolas que participaram do ProEMI têm em média uma redução de 11,9% nas faltas em matemática em relação as que não participaram. Em português essa redução é de 5,27%.

Tabela 3 - Efeito do ProEMI na frequência escolar por disciplina

Modelo	Pareamento	Português	Matemática
Poisson	Não	-0.0823*** (0.00685)	-0.0944*** (0.00687)
	Sim	-0.0527*** (0.0105)	-0.119*** (0.0104)
Binomial Negativa	Não	-0.0783*** (0.0190)	-0.0939*** (0.0192)
	Sim	-0.0802*** (0.0266)	-0.0923*** (0.0282)
ZIP	Não	-0.0958*** (0.0167)	-0.0941*** (0.0180)
	Sim	-0.0732*** (0.0277)	-0.125*** (0.0291)

Erro padrão em parênteses

*** p<0.01, ** p<0.05, * p<0.1

Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço a partir dos resultados do modelo.

Nos modelos de regressão binomial negativa, os efeitos estimados nos modelos pareados são menores do que os estimados nos modelos não pareados para a disciplina de matemática, mas maiores para língua portuguesa, sendo que a participação no ProEMI reduz o número de faltas em aproximadamente 8% em português e 9,2% em matemática, considerando os dados pareados.

A *Tabela 4*, no Apêndice, apresenta, além das estimativas das variáveis explicativas comuns a todos os modelos, as estimativas do parâmetro de dispersão alpha. Se o parâmetro de dispersão é igual a zero, então o modelo se reduz a um Poisson simples, mas se alpha for estatisticamente maior do que zero, então os dados apresentam alta dispersão, por consequência são melhores estimados por um modelo binomial negativa do que por um modelo Poisson. Para todas as regressões o teste de dispersão dos dados deu positivo indicando que a variável dependente tem alta dispersão e não é suficientemente descrita por uma distribuição Poisson simples.

Com relação aos parâmetros estimados com modelos ZIP, tem-se que a participação no programa reduz o número de faltas em 7,32% nas aulas de língua portuguesa e em 12,5% nas aulas de matemática, em comparação com as escolas que não participam e considerando a base com pareamento. A *Tabela 5* mostra os resultados das regressões ZIP com dados não pareados e pareados, respectivamente. Os últimos dois parâmetros em cada modelo ZIP (taxa de abandono e constante) correspondem as estimativas por um modelo logit que prediz se os estudantes tiveram zero faltas ou não. Os coeficientes negativos de inflação mostram que quanto maior a taxa de abandono escola, menor será a probabilidade de o aluno ter zero faltas.

CONSIDERAÇÕES FINAIS

Após a grande expansão da oferta do ensino público, o governo brasileiro começou a concentrar as políticas públicas educacionais em aspectos relacionados à atração e permanência de estudantes nas escolas, e à qualidade dos serviços oferecidos. O Programa Ensino Médio Inovador foi criado com o propósito de transformar as escolas em ambientes mais conectados com a realidade moderna da sociedade a partir de propostas inovadoras vindas das próprias unidades escolares. A expectativa era de que com um currículo focado às necessidades locais/regionais os estudantes se sentiriam mais motivados a frequentar às aulas e por consequência haveria um queda nas taxas de abandono escolar.

Este trabalho se propôs a avaliar o impacto do ProEMI na frequência escolar dos alunos do ensino médio nas escolas de Santa Catarina. Foi utilizada uma técnica de pareamento por escore de propensão para selecionar as escolas que não aderiram ao

programa, mas que tinham características semelhantes às aquelas que participaram. Foram selecionadas duas escolas fora do programa para cada escola dentro do programa. Dessa forma, o grupo de tratamento foi composto por 8 escolas e o grupo de controle por 16, totalizando 24 escolas. Foram utilizados três modelos de regressão que abordam dados de contagem: Poisson, binomial negativa e Poisson inflado de zeros (ZIP). As regressões foram rodadas tanto para a amostra pareada, quanto para a não pareada, e individualmente para cada disciplina (*Tabela 6*).

Os resultados estimados mostram que o ProEMI tem efeitos positivos nos três modelos, reduzindo o número de faltas nas duas disciplinas analisadas. Isto é, os resultados estimados pelos modelos mostram que o Programa Ensino Médio Inovador teve o efeito de redução no número de faltas nas disciplinas de Língua Portuguesa e Matemática nas escolas estaduais de Santa Catarina que aderiram ao programa no ano de 2011. Portanto, considerando que a frequência escolar é um preditivo para o abandono escolar, os resultados encontrados fornecem indícios de que o ProEMI contribui para a permanências dos alunos nas escolas analisadas.

Resultados mais robustos podem ser obtidos com uma análise longitudinal, pois o efeito de programas como o Ensino Médio Inovador é muito subjetivo para ser avaliado em apenas um ano. Quando se trata de educação, as mudanças ocorrem gradualmente, de modo que a percepção dos alunos com relação a atratividade da escola mude lentamente, portanto, para conseguir estimar o efeito do ProEMI na frequência escolar seria mais interessante analisar dados dentro de um período maior.

REFERÊNCIAS

BRASIL. Constituição da República Federativa do Brasil de 1988. **Diário Oficial da União**, Brasília, DF, 5 out. 1988. Atual. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Acesso em: 12 nov. 2018.

BRASIL. Emenda Constitucional n. 14, de 12 de setembro de 1996. Modifica os arts. 34, 208, 211 e 212 da Constituição Federal e dá nova redação ao art. 60 do Ato das Disposições constitucionais Transitórias. **Diário Oficial da União**, Brasília, DF, 13 set. 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Emendas/Emc/emc14.htm#art2. Acesso em: 12 nov. 2018.

BRASIL. Emenda Constitucional n. 59, de 11 de novembro de 2009. Acrescenta § 3º ao art. 76 do Ato das Disposições Constitucionais Transitórias para reduzir, anualmente, a partir do exercício de 2009, o percentual da Desvinculação das Receitas da União incidente sobre os recursos destinados à manutenção e desenvolvimento do ensino de que trata o art. 212 da Constituição Federal, dá nova redação aos incisos I e VII do art. 208, de forma a prever a obrigatoriedade do ensino de quatro a dezessete anos e ampliar a abrangência dos programas suplementares para todas as etapas da educação básica, e dá nova redação ao § 4º do art. 211 e ao § 3º do art. 212 e ao **caput** do art. 214, com a inserção neste dispositivo de inciso VI. **Diário Oficial da União**, Brasília, DF, 12 nov. 2009. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Emendas/Emc/emc59.htm#art1. Acesso em: 12 nov. 2018.

BRIDGELAND, John M.; DIJULIO JR, John J.; MORISON, Karen Burke. **The silent epidemic: perspectives of high school dropouts**. Washington, DC: Civic Enterprises, 2006.

CAETANO, Lucília. Abandono escolar: repercussões sócio-económicas na região Centro. Algumas reflexões. **Finisterra**, Lisboa, v. 79, p. 163-176, 2005.

CAMERON, Adrian Colin; TRIVEDI, Pravin K. **Microeconometrics: methods and applications**. Cambridge: Cambridge University Press, 2005.

INSTITUTO UNIBANCO. **A crise de audiência no ensino médio**. São Paulo: Grandini & GCP, 2010.

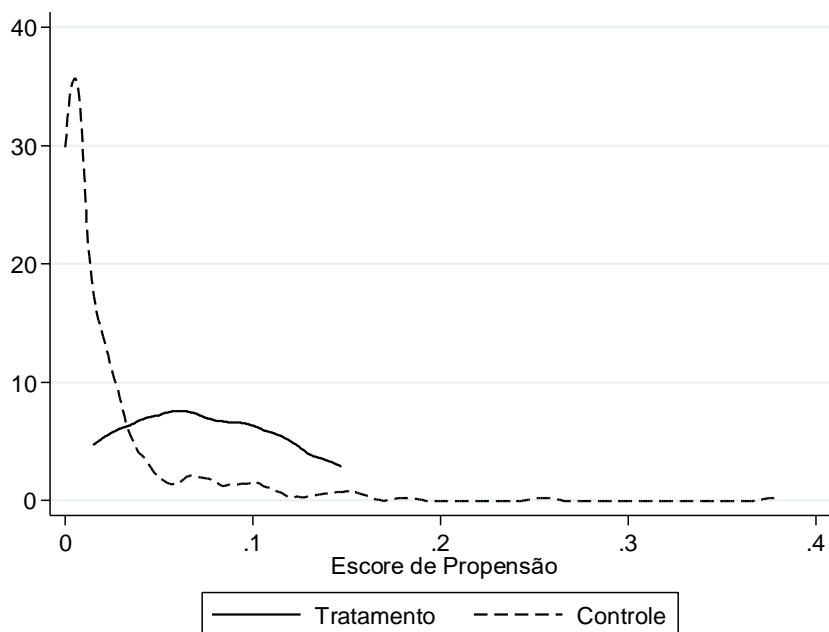
KATTAN, Raja B; SZÉKELY, Miguel. **Analyzing the dynamics of school dropout in upper secondary education in Latin America: a cohort approach**. Washington, DC: The World Bank, 2015.

MINISTÉRIO DA EDUCAÇÃO. **Programa Ensino Médio Inovador: documento orientador**. Brasília, DF, 2009. Disponível em: http://portal.mec.gov.br/dmdocuments/documento_orientador.pdf. Acesso em: 12 nov. 2018.

ROSENBAUM, Paul R.; RUBIN, Donald B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, Oxford, v. 70, n. 1, p. 41-55, 1983.

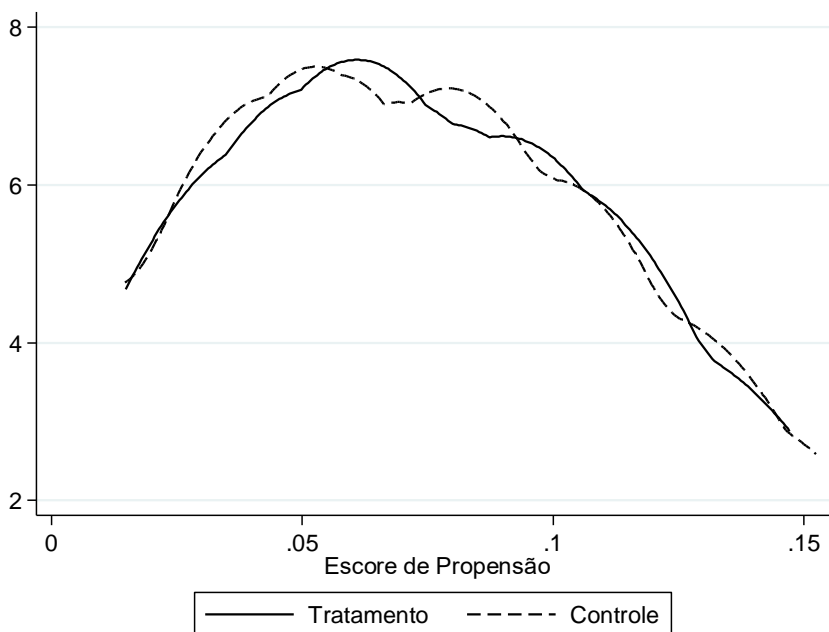
APÊNDICE A – RESULTADOS

Figura 2 – Escore de propensão dos grupos tratado e controle antes do pareamento



Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço a partir dos resultados do modelo.

Figura 3 – Pareamento por escore de propensão dos grupos tratado e controle



Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço a partir dos resultados do modelo.

Tabela 4 - Poisson

	Sem Pareamento		Com Pareamento	
	Português	Matemática	Português	Matemática
ProEMI	-0.0823*** (0.00685)	-0.0944*** (0.00687)	-0.0527*** (0.0105)	-0.119*** (0.0104)
Nº funcionários	0.000990*** (7.66e-05)	-0.000227*** (7.56e-05)	-0.00999*** (0.000388)	-0.0164*** (0.000375)
Biblioteca	-0.0107*** (0.00362)	-0.00186 (0.00358)	-0.201*** (0.0172)	-0.104*** (0.0168)
Lab. Ciências	-0.151*** (0.00356)	-0.169*** (0.00351)	-0.227*** (0.0160)	-0.403*** (0.0154)
Lab. Informática	0.00429 (0.00441)	0.114*** (0.00448)	0.286*** (0.0152)	0.166*** (0.0148)
Quadra esporte	-0.0543*** (0.00383)	0.0100*** (0.00367)	0.257*** (0.0139)	0.454*** (0.0137)
Zona aluno	0.334*** (0.00428)	0.401*** (0.00425)	0.205*** (0.0135)	0.267*** (0.0137)
Idade	0.0369*** (0.000351)	0.0371*** (0.000339)	0.0253*** (0.000898)	0.0233*** (0.000899)
Aluno/Turma	0.0173*** (0.000418)	0.0191*** (0.000411)	0.0611*** (0.00138)	0.0443*** (0.00127)
i.Cor				
Preta	0.198*** (0.0164)	0.234*** (0.0158)	0.0677 (0.0533)	0.0371 (0.0517)
Parda	0.0732*** (0.00779)	0.125*** (0.00745)	0.172*** (0.0289)	0.111*** (0.0284)
Indígena	-0.0413 (0.0510)	-0.139*** (0.0530)	-1.160*** (0.302)	-0.646*** (0.218)
Outra	0.362*** (0.0284)	0.462*** (0.0273)	-0.753*** (0.200)	-0.228 (0.153)
Constante	0.698*** (0.0130)	0.578*** (0.0128)	0.391*** (0.0484)	1.393*** (0.0457)
Observações	64,194	64,633	7,596	7,573

Erro padrão em parênteses

*** p < 0.01, ** p < 0.05, * p < 0.1

Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço a partir dos resultados do modelo.

Tabela 5 – Binomial Negativa

	Sem Pareamento		Com Pareamento	
	Português	Matemática	Português	Matemática
ProEMI	-0.0783*** (0.0190)	-0.0939*** (0.0192)	-0.0802*** (0.0266)	-0.0923*** (0.0282)
Nº funcionários	0.000963*** (0.000224)	8.66e-05 (0.000222)	-0.00929*** (0.000998)	-0.0128*** (0.000990)
Biblioteca	-0.00976 (0.0105)	-0.0153 (0.0106)	-0.203*** (0.0481)	-0.137*** (0.0505)
Lab. Ciências	-0.141*** (0.00985)	-0.159*** (0.00985)	-0.212*** (0.0401)	-0.284*** (0.0406)
Lab. Informática	0.0315** (0.0128)	0.145*** (0.0129)	0.382*** (0.0436)	0.211*** (0.0453)
Quadra esporte	-0.0481*** (0.0108)	0.0207* (0.0106)	0.250*** (0.0352)	0.378*** (0.0356)
Zona aluno	0.339*** (0.0108)	0.401*** (0.0109)	0.244*** (0.0323)	0.287*** (0.0339)
Idade	0.137*** (0.00320)	0.134*** (0.00319)	0.132*** (0.00920)	0.119*** (0.00936)
Aluno/Turma	0.0197*** (0.00117)	0.0197*** (0.00114)	0.0603*** (0.00339)	0.0430*** (0.00343)
i.Cor				
Preta	0.154*** (0.0520)	0.172*** (0.0522)	-0.0222 (0.149)	-0.0234 (0.153)
Parda	0.101*** (0.0225)	0.145*** (0.0225)	0.184** (0.0800)	0.158* (0.0813)
Indígena	-0.0972 (0.139)	-0.180 (0.140)	-1.212** (0.518)	-0.651 (0.477)
Outra	0.493*** (0.0885)	0.567*** (0.0886)	-0.0556 (0.467)	0.0870 (0.465)
Constante	-1.048*** (0.0626)	-1.086*** (0.0619)	-1.498*** (0.192)	-0.468** (0.192)
ln(alpha)	-0.1025 (0.0067)	-0.0911 (0.0066)	-0.1905 (0.0197)	-0.1256 (0.0194)
alpha	0.9026 (0.0060)	0.9129 (0.0060)	0.8265 (0.0163)	0.8820 (0.0171)
Observações	64,194	64,633	7,596	7,573

Erro padrão em parênteses

*** p < 0.01, ** p < 0.05, * p < 0.1

Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço a partir dos resultados do modelo.

Tabela 6 – Poisson Inflado de Zeros (ZIP)

	Sem Pareamento		Com Pareamento	
	Português	Matemática	Português	Matemática
ProEMI	-0.0958*** (0.0167)	-0.0941*** (0.0180)	-0.0732*** (0.0277)	-0.125*** (0.0291)
Nº funcionários	0.00044** (0.000202)	-0.00091*** (0.000215)	-0.00948*** (0.00101)	-0.0141*** (0.00107)
Biblioteca	-0.0155 (0.00950)	0.0168* (0.00983)	-0.167*** (0.0399)	-0.0247 (0.0392)
Lab. Ciências	-0.139*** (0.00923)	-0.134*** (0.00965)	-0.171*** (0.0407)	-0.325*** (0.0424)
Lab. Informática	-0.0231* (0.0122)	0.0202 (0.0137)	0.194*** (0.0351)	0.0497 (0.0347)
Quadra esporte	-0.0107 (0.0100)	0.0143 (0.0101)	0.224*** (0.0373)	0.378*** (0.0423)
Zona aluno	0.269*** (0.0112)	0.335*** (0.0113)	0.125*** (0.0373)	0.221*** (0.0348)
Idade	0.0346*** (0.00398)	0.0368*** (0.00424)	0.0208*** (0.00491)	0.0236*** (0.00620)
Aluno/Turma	0.0114*** (0.00118)	0.0132*** (0.00130)	0.0452*** (0.00374)	0.0379*** (0.00359)
i.Cor				
Preta	0.181*** (0.0449)	0.208*** (0.0500)	0.0825 (0.103)	0.0120 (0.148)
Parda	0.0389* (0.0224)	0.0820*** (0.0234)	0.160** (0.0773)	0.101 (0.0863)
Indígena	-0.109 (0.116)	-0.217** (0.101)	-1.248*** (0.274)	-0.642*** (0.0594)
Outra	0.228* (0.131)	0.323** (0.137)	-0.781 (0.555)	-0.429 (0.471)
Constante	1.126*** (0.0745)	1.017*** (0.0788)	1.090*** (0.150)	1.589*** (0.163)
Inflação				
Taxa de abandono	-0.0248*** (0.00232)	-0.0433*** (0.00227)	-0.0497*** (0.00718)	-0.0455*** (0.00689)
Constante	-1.741*** (0.0213)	-1.660*** (0.0203)	-1.571*** (0.0719)	-1.594*** (0.0702)
Observações	64,194	64,633	7,596	7,573

Erro padrão robusto em parênteses

*** p < 0.01, ** p < 0.05, * p < 0.1

Fonte: Elaborado por Jeniffer Gonçalves e Francis Carlo Petterini Lourenço a partir dos resultados do modelo.