



Universidade de Brasília - UnB
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

**Fatores associados ao abandono e reprovação escolar dos alunos
do 5º e 9º ano no Distrito Federal em 2015**

Helen Cristina de Acypreste Rocha

Prof. Dr. Luís Gustavo do Amaral Vinha

Brasília
2018

Helen Cristina de Acypreste Rocha

**Fatores associados ao abandono e reprovação escolar dos alunos do 5º e 9º
ano no Distrito Federal em 2015**

Orientador:

Prof. Dr. Luís Gustavo do Amaral Vinha

Projeto apresentado para obtenção do título
de Bacharel em Estatística.

Brasília

2018

Sumário

1 Introdução	5
2 Referencial Teórico	7
2.1 Modelos Lineares Generalizados	7
2.2 Regressão Logística	7
2.2.1 Regressão Logística Simples	8
2.2.2 Regressão Logística Múltipla	9
2.2.3 Estimação	10
2.3 Regressão Multinível	10
2.3.1 Regressão Logística Multinível	11
3 Resultados	12
3.1 Análise Descritiva	12
3.1.1 Quantidade de alunos por Coordenação Regional de Ensino (CRE). 12	
3.1.2 Percentual de abandono e reprovação por CRE	14
3.1.3 Relação do percentual de abandono e reprovação e nível socio-econômico da escola	17
3.1.4 Relação do percentual de abandono e reprovação e tamanho da escola	19
3.1.5 Relação do percentual de abandono e reprovação e escore de violência	21
3.2 Análise bivariada	23
3.2.1 Variáveis dos alunos do 5º ano e abandono e reprovação	23
3.2.2 Variáveis dos alunos do 9º ano e abandono e reprovação	24
3.3 Análise dos modelos	26
3.3.1 Reprovação no 5º ano	27
3.3.2 Abandono no 5º ano	28
3.3.3 Reprovação no 9º ano	29
3.3.4 Abandono no 9º ano	31
4 Conclusão	33

Referências. 35

1 Introdução

A educação tem papel fundamental para o desenvolvimento humano e o entrelaçamento das relações sociais. Segundo a Constituição Federal Brasileira de 1988, artigo 6º, deve ser garantido, para melhoria da condição social de todos os indivíduos, além de outros direitos, o acesso à educação. Além disso, a constituição reafirma o direito à educação básica para todos, ao estabelecer, no seu artigo 208 que "o dever do Estado com a educação será efetivado mediante a garantia de ensino fundamental, obrigatório e gratuito, inclusive para os que a ele não tiveram acesso na idade própria". Acrescenta ainda, no mesmo parágrafo, que "o não oferecimento do ensino obrigatório pelo Poder Público, ou sua oferta irregular, importa responsabilidade da autoridade competente".

Ainda que seja determinada a garantia dos direitos educacionais à sociedade, existe um distanciamento entre a especificação das leis e sua prática efetiva.

Sabe-se que a educação brasileira apresenta dificuldades e obstáculos a serem superados. Um deles é a violência no ambiente escolar. Frequentemente são reportadas notícias sobre infrações cometidas por crianças e adolescentes dentro das escolas, entre elas agressões verbais, físicas, *bullying*, furtos e roubos, além de portarem armas ou estarem sob efeitos de drogas. "Vítima da violência, a criança, além de reproduzi-la, pode reagir através de uma mudança brusca de comportamento. Falta de atenção, baixa auto-estima, variação de humor e agressividade são alguns sinais aos quais pais e educadores devem estar sempre atentos. No entanto, apesar da violência ocorrer dentro das escolas, não é gerada pela escola em si, mas por fatores externos, como famílias desestruturadas, narcotráficos, conflitos sociais, etc." (SOUZA, 2008).

Um outro obstáculo a ser considerado para a discussão na educação é o fracasso escolar, que inclui o número de alunos que abandonaram e reprovaram e segundo Bica (2013), é um dos temas mais discutidos na área da educação. Afinal, trata-se de uma problemática que se estende desde o início da história educacional brasileira até hoje. Ele pode ser percebido sobre dois aspectos: questões internas à escola ou questões externas à ela.

Segundo Ortigão e Aguiar (2013), apesar da taxa de reprovação/repetência escolar ter caído de 30% na década de 1990 para 18,7%, essa taxa ainda coloca o Brasil na condição de país que mais reprova na educação básica entre os 41 países da América Latina e Caribe, segundo relatório da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) sobre a educação mundial publicado em 2010. Um estudo de Silva e Araújo (2017) aponta que "o Brasil tem a terceira

maior taxa de abandono escolar entre os 100 países com maior IDH e no PNUD e a menor média de anos de estudo entre os países da América do Sul".

A reprovação e o abandono escolar têm impactos sociais negativos e portanto devem ser tratados em coletivo com incentivo e aplicação de práticas que visam à sua diminuição. É intuitivo crer que características dos alunos, aspectos familiares e situações escolares estão diretamente ligadas ao fracasso escolar.

O estudo proposto nesse documento objetiva mensurar alguns fatores que estão ligados diretamente ao abandono e reprovação escolar, considerando os alunos de 5º e 9º ano do ensino fundamental, do Distrito Federal, que participaram da Sistema Nacional de Avaliação da Educação Básica (SAEB) em 2015. Para isso, foram utilizadas técnicas descritivas e inferenciais considerando as variáveis relacionadas aos alunos como sexo, idade, cor/raça, nível socioeconômico, entre outras. Além disso, considerará o escore de violência interna obtido a partir da resposta dos diretores e professores no questionário da Prova Brasil referentes a violência dentro da escola.

2 Referencial Teórico

2.1 Modelos Lineares Generalizados

Modelos estatísticos são utilizados para analisar a associação entre uma variável de interesse Y , chamada de variável resposta, e outras variáveis X_1, X_2, \dots, X_k , chamadas de variáveis explicativas. A partir da estimação dos parâmetros e ajuste do modelo é possível avaliar a força e importância dos efeitos. Com isso, podem-se fazer inferências para identificar quais variáveis explicativas afetam a variável resposta. Por fim, os valores preditos do modelo fornecem estimativas otimizadas da média de Y considerando possíveis valores das variáveis explicativas X_1, X_2, \dots, X_k .

Os Modelos Lineares Generalizados (MLG) são uma extensão do modelo linear clássico, onde o valor esperado da variável resposta é uma função linear das covariáveis. Aqui, a distribuição considerada não precisa ser normal, pode ser qualquer distribuição da família exponencial. Por outro lado, a função que relaciona o valor esperado e as covariáveis pode ser qualquer função diferenciável.

Os MLGs possuem 3 componentes:

1) Componente Aleatório

Determina a distribuição de probabilidades da variável resposta Y . O conjunto de observações (y_1, \dots, y_n) são tratados como independentes.

2) Componente Sistemático

Especificam as variáveis explicativas. A combinação linear das k variáveis X_1, X_2, \dots, X_k são denominadas preditores lineares e é dado por $\alpha + \beta_1 x_1 + \dots + \beta_k x_k$.

3) Função de ligação

Especifica uma função $g(\cdot)$ que relaciona o valor esperado $\mu = E(Y)$ ao preditor linear $g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$.

O modelo considerado no trabalho é o modelo logístico $g(\mu) = \log[\mu/(1 - \mu)]$, onde Y tem distribuição binomial.

2.2 Regressão Logística

Os modelos de regressão logística são utilizados quando se deseja avaliar a relação entre uma variável resposta categórica e variáveis explicativas quantitativas e/ou categóricas. Neste caso, será considerada variável resposta binária, ou seja,

assume valor 0, indicando fracasso ou 1, indicando sucesso. Vale ressaltar que a probabilidade de sucesso/fracasso depende de outras variáveis.

Para melhor compreensão desse modelo é importante considerar o conceito de chance (do inglês *odds*) que indica a chance de sucesso da variável em estudo. Para a probabilidade $P(Y = 1) = \pi$, a chance de sucesso é

$$odds = \frac{\pi}{1 - \pi}. \quad (1)$$

A razão de chances (do inglês *odds ratio*) se uma variável binária Y em função de uma variável X também binária é dada por

$$\theta = \frac{odds_1}{odds_2} = \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_2}{1-\pi_2}}, \quad (2)$$

onde $odds_1$ é chance de sucesso para a variável Y na categoria 1 de X e $odds_2$ é a chance de sucesso para a variável Y na categoria 2 de X. Se $\theta > 1$ a chance de sucesso na categoria 1 de X é maior que a chance de sucesso na categoria 2 de X. Se $\theta < 1$ a chance de sucesso na categoria 1 de X é menor que a chance de sucesso na categoria 2 de X (Rocha, 2014).

2.2.1 Regressão Logística Simples

Baseado em Agresti (1996) e Hox (2010), vamos considerar aqui uma variável resposta Y com duas categorias que depende de apenas uma variável explicativa X. O modelo de regressão logística tem forma linear para o logito da probabilidade de Y em função de X, $\pi(x)$, como

$$\text{logito}(\pi(x)) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x. \quad (3)$$

Isolando $\pi(x)$ e fazendo as manipulações necessárias, é obtida a probabilidade de sucesso

$$\exp(\beta_0 + \beta_1 x) = \frac{\pi(x)}{1 - \pi(x)}$$

$$\frac{1}{\pi(x)} = \frac{1 + \exp(\beta_0 + \beta_1 x)}{\exp(\beta_0 + \beta_1 x)} \quad (4)$$

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

Na equação (4), β_1 determina o crescimento ou decréscimo da curva que tem o formato de "S" alongado.

Há duas opções no modelo de regressão logística, sendo $Y = 1$ ou $Y = 0$. A razão dessas duas opções é conhecida como razão de chances (*odds ratio*) designada por Ψ , que é a base para a interpretação dos coeficientes do modelo de regressão logística. A razão de chances é a probabilidade de Y ser um membro de uma classe relativa à outra classe. Por exemplo, para uma variável binária independente (X : 0 ou 1), Ψ pode ser expresso como

$$\Psi = \frac{\frac{\pi(x+1)}{1-\pi(x+1)}}{\frac{\pi(x)}{1-\pi(x)}} = \frac{e^{\beta_0} (e^{\beta_1})^{(x+1)}}{e^{\beta_0} (e^{\beta_1})^x} = \exp(\beta_1 x + \beta_1 - \beta_1 x) = e^{\beta_1}. \quad (5)$$

O *odds ratio* é a medida de quanto mais provável ou improvável é para o resultado de estar presente entre aqueles com $X = 1$ do que entre aqueles com $X = 0$. Assim, depois de estimar os parâmetros, o efeito da variável independente na variável de resultado pode ser medido através do *odds ratio*.

2.2.2 Regressão Logística Múltipla

A regressão logística múltipla se estende para mais de uma variável independente, podendo ser quantitativa e/ou qualitativa. O processo para obtenção do modelo é o mesmo usado na regressão simples. Suponha um vetor com p variáveis independentes $X = (x_1, x_2, \dots, x_p)$. O logito do modelo é dado por

$$\text{logito}(\pi(X)) = g(X) = \beta_0 + \sum_{j=1}^p \beta_j x_j \quad (6)$$

e o modelo logístico é

$$\pi(X) = \frac{\exp(g(X))}{1 + \exp(g(X))}. \quad (7)$$

2.2.3 Estimação

Os parâmetros do modelo logístico são estimados por meio do método de máxima verossimilhança. A ideia é maximizar o logaritmo da função de verossimilhança, uma vez que $0 < P[Y_i = y_i|x_i] < 1$. A variável resposta tem distribuição Bernoulli e tem a seguinte forma

$$P(Y = y) = \pi(x)^y(1 - \pi(x))^{1-y}, \quad (8)$$

onde $y = 0, 1$.

Dado Y_i a resposta da i -ésima observação e o vetor de covariáveis $x_i = (x_{1i}, x_{2i}, \dots, x_{ki})$. A função de máxima verossimilhança é dada por

$$L(\beta|Y) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}, \quad (9)$$

em que

$$\pi(x_i) = \frac{e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ji}}}{1 + e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ji}}}, \quad (10)$$

onde β_1, \dots, β_k são os parâmetros a serem estimados.

Para obter os estimadores, deriva-se a função em relação a cada parâmetro e iguala a zero. Essas equações são não-lineares nos parâmetros e para resolvê-las é preciso utilizar métodos numéricos iterativos, como Newton-Raphson.

2.3 Regressão Multinível

Nos estudos onde a população de interesse apresenta estrutura hierárquica, isto é, observações de nível mais baixo aninhadas em unidades de nível mais alto, utilizamos análise multinível. Aqui, temos interesse em avaliar os fatores que estão

associados a alguma medida do nível mais baixo (no caso do estudo, reprovação e abandono dos alunos).

Portanto, a regressão multinível permite verificar a relação entre variáveis de diferentes níveis. Isso acontece porque o comportamento dos indivíduos são interferidos pelo grupo em que estão inseridos. Nesse estudo, por exemplo, temos os alunos como nível baixo e as escolas como o nível alto.

2.3.1 Regressão Logística Multinível

Quando a variável resposta é categórica e os dados apresentam estrutura hierárquica, utilizamos o modelo logístico multinível. Esse modelo é semelhante ao modelo logístico, incluindo os efeitos aleatórios e variáveis explicativas dos demais níveis.

Na regressão logística, temos que $\pi(x)$ é modelado usando a função de ligação *logito*. Considere y_{ij} a variável resposta binária associada ao indivíduo i do nível 1 considerada dentro do grupo de nível 2 j . O modelo de regressão logística de dois níveis pode ser escrito como

$$\ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 X_{ij} + u_j, \quad (11)$$

onde u_j é o efeito aleatório do nível 2. Podemos obter o modelo de regressão logística retirando u_j .

Aqui, u_j é uma erro aleatório e segue distribuição $N(0, \sigma_u^2)$. O modelo (11) pode ser escrito dividindo-se em dois modelos, um para o nível 1 e outro para o nível 2:

$$\ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_{0j} + \beta_1 X_{ij}, \quad (12)$$

onde $\beta_{0j} = \beta_0 + u_j$

Uma questão que surge é cálculo do coeficiente de correlação intra-classe. Como não se tem a variância individual, a variância que se deve usar é dada por $\pi^2/3 \approx 3,29$ (Evans, Hastings e Peacock, 2000). Dessa forma, temos que $\rho = \frac{\sigma_G^2}{\sigma_G^2 + 3,29}$.

3 Resultados

Para a realização do estudo, utilizaram-se dados¹ dos alunos do 5º e 9º ano das escolas públicas do Distrito Federal referentes ao ano de 2015. Além disso, para medir tamanho da escola, considerou-se o Censo Escolar de 2015. As variáveis consideradas são descritas na Tabela 1.

Tabela 1: Descrição das variáveis.

Variável	Descrição e Codificação
Reprovação	0 = Não; 1 = Sim.
Abandono	0 = Não; 1 = Sim.
Sexo	1 = Feminino; 0 = Masculino.
Raça/cor ²	1 = Não negros; 0 = Negros.
Escolaridade da mãe	1 = Ensino Médio ou Superior; 0 = Outros.
Faz dever de português	0 = Não; 1 = Sim.
Faz dever de matemática	0 = Não; 1 = Sim.
Nível socioeconômico do aluno	Obtida por meio de TRI ³ utilizando as respostas dos alunos referentes a posse de bens da família.
Nível socioeconômico da escola	Variável calculada pelo INEP com os níveis: Muito baixo, Baixo, Médio Baixo, Médio, Médio Alto, Alto, Muito Alto.
Tamanho da escola	Quantidade de matrículas na escola (Dados do Censo Escolar).
Escore de violência	Obtida por meio de TRI utilizando a resposta dos professores e diretores sobre violência escolar do SAEB 2015.

3.1 Análise Descritiva

Inicialmente foi feita uma análise descritiva das variáveis do nível da escola com reprovação e abandono dos alunos do 5º e 9º ano.

3.1.1 Quantidade de alunos por Coordenação Regional de Ensino (CRE).

As escolas do Distrito Federal são divididas e administradas por 14 Coordenadorias Regionais de Ensino. Cada CRE engloba determinadas Regiões Administrativas (RA).

A Tabela 2 mostra a quantidade de alunos separados por sexo que responde-

¹Disponível em: <http://portal.inep.gov.br/microdados>.

²Alunos que responderam "não sei" ou não responderam foram retirados do banco. A categoria "Não negros" engloba brancos, amarelos e indígenas.

³Teoria de Resposta ao Item.

ram o questionário do SAEB de 2015 e as Regiões Administrativas de cada CRE.

Tabela 2: Tabela quantidade de alunos por Coordenação Regional de Ensino.

CRE	RA	Feminino	Masculino	Total
Ceilândia	Ceilândia	4570 (51,4%)	4319 (48,6%)	8889
Planaltina	Planaltina	2099 (50,7%)	2044 (49,3%)	4143
Taguatinga	Taguatinga, Águas Claras, Vicente Pires	2052 (51,3%)	1946 (48,7%)	3998
Plano Piloto	Brasília, Cruzeiro, São Sebastião, Lago Sul, Lago Norte, Sudoeste/Octogonal, Varjão, Jardim Botânico	1943 (51,6%)	1821 (48,4%)	3764
Samambaia	Samambaia	1947 (51,8%)	1811 (48,2%)	3758
Gama	Gama, Recanto das Emas (1)	1720 (51,3%)	1634 (48,7%)	3354
Recanto das Emas	Recanto das Emas, Samambaia (1)	1574 (51,8%)	1465 (48,2%)	3039
Sobradinho	Sobradinho, Planaltina, Lago Norte, Sobradinho II, Fercal	1391 (51,8%)	1294 (48,2%)	2685
Santa Maria	Santa Maria	1267 (49,4%)	1300 (50,6%)	2567
Núcleo Bandeirante	Núcleo Bandeirante, Riacho Fundo, Candangolândia, Riacho Fundo II e Park Way	1163 (51,1%)	1112 (48,9%)	2275
São Sebastião	São Sebastião	1082 (47,8%)	1182 (52,2%)	2264
Brazlândia	Brazlândia	932 (48,5%)	988 (51,5%)	1920
Guará	Guará, SCIA e SIA	964 (50,7%)	936 (49,3%)	1900
Paranoá	Paranoá, Sobradinho (1) e Itapoã	940 (49,5%)	960 (50,5%)	1900

Nota-se que Ceilândia tem maior quantidade de alunos, enquanto Paranoá, Guará e Brazlândia têm menor quantidade. Além disso, a maioria das CREs têm percentual de meninas maior que de meninos.

3.1.2 Percentual de abandono e reprovação por CRE

A Figura 1 mostra que as escolas de Brazlândia, Planaltina e Guará têm maior percentual de reprovação dos alunos do 9º ano. Já Samambaia, Santa Maria e Paranoá têm menor percentual.

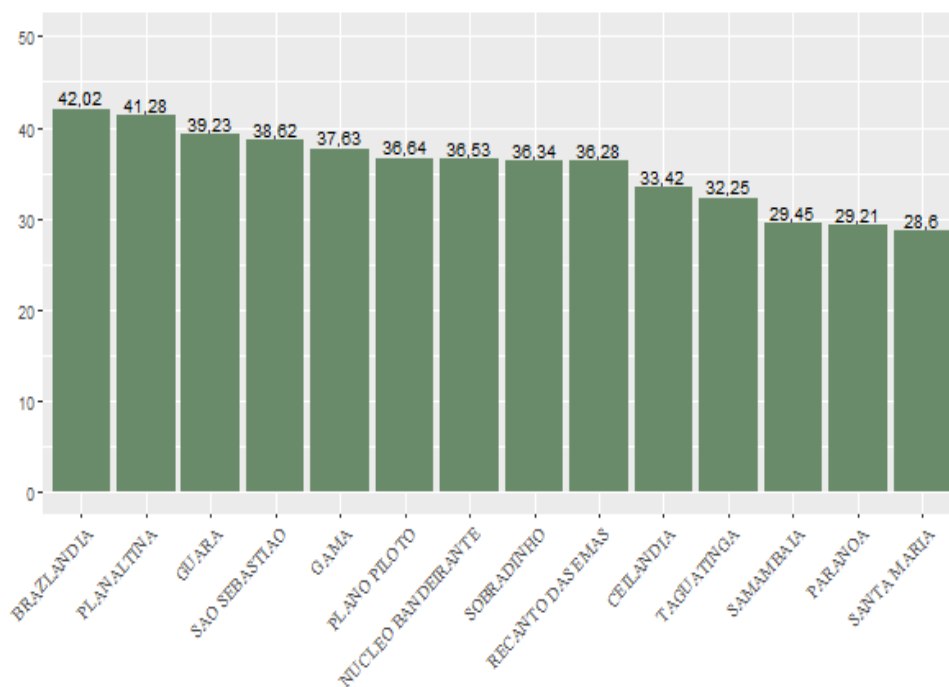


Figura 1: Percentual de reprovação do 9º ano por CRE.

Quanto ao abandono dos alunos do 9º ano, as escolas do Guará, Samambaia e Gama apresentam maior percentual. Taguatinga, Sobradinho e Paranoá apresentam menor percentual.

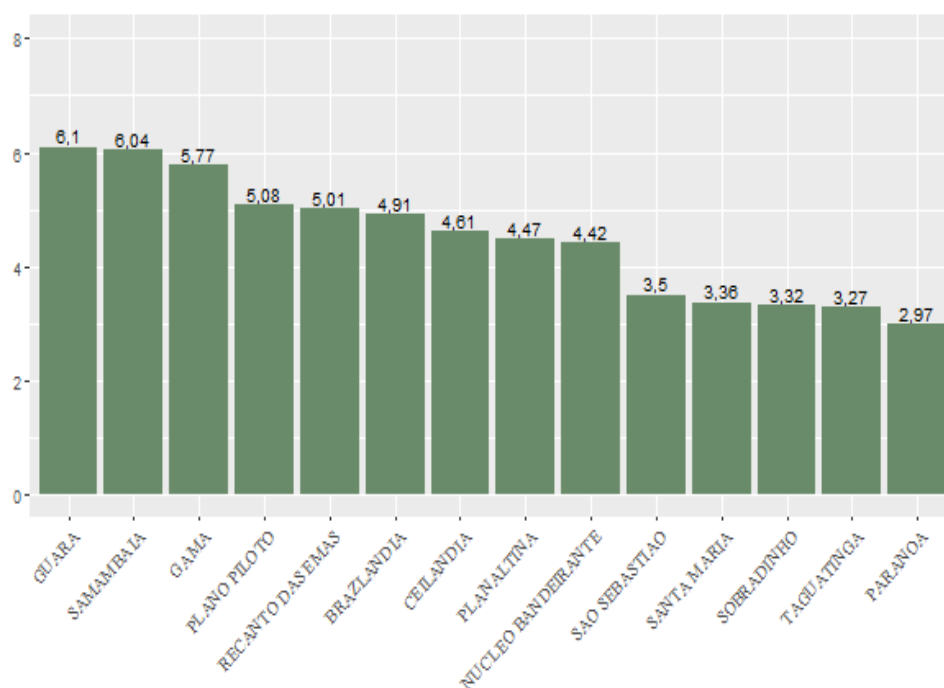


Figura 2: Percentual de abandono do 9º ano por CRE.

Considerando o 5º ano do ensino fundamental, tem-se que o percentual de reprovação é maior em São Sebastião, Brazlândia e Paranoá e menor em Taguatinga, Plano Piloto e Guará.

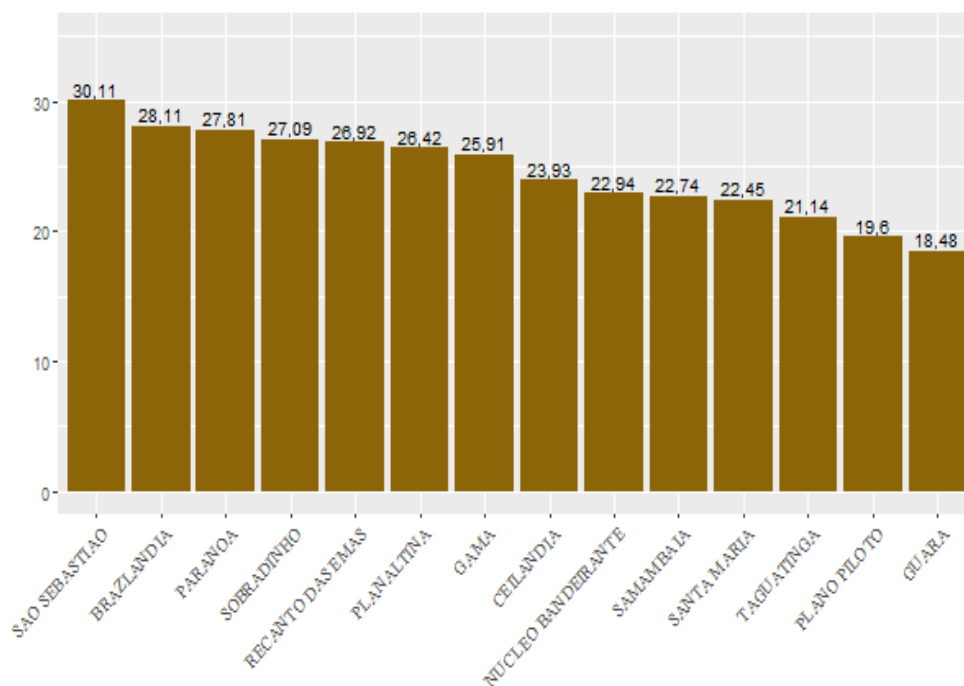


Figura 3: Percentual de reprovação do 5º ano por CRE.

O percentual de abandono é maior em São Sebastião, Paranoá e Sobradinho e menor em Taguatinga, Plano Piloto e Santa Maria.

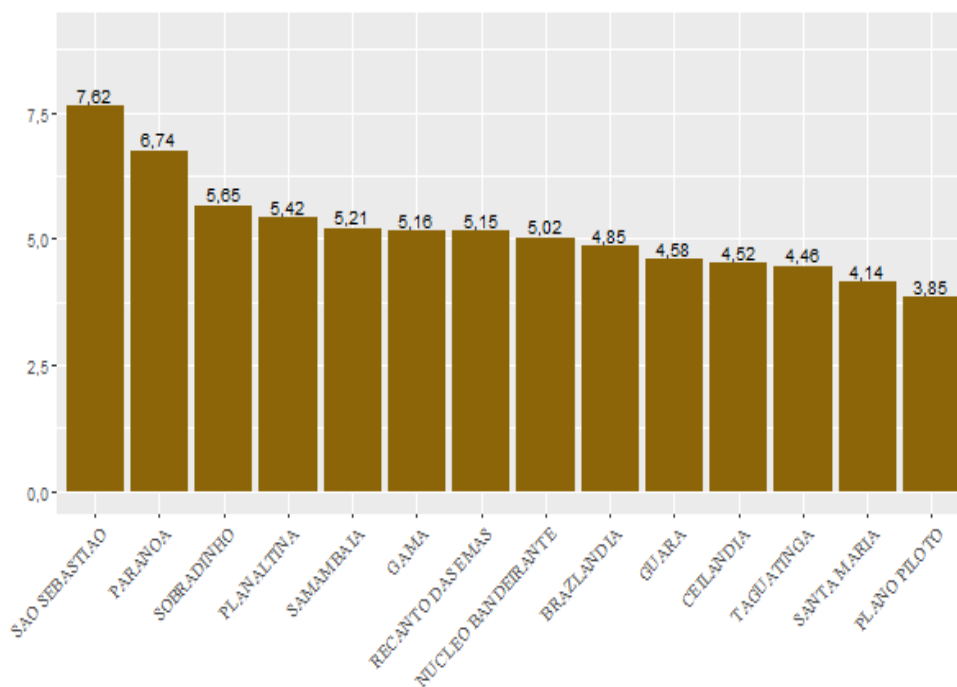


Figura 4: Percentual de abandono do 5º ano por CRE.

3.1.3 Relação do percentual de abandono e reprovação e nível socioeconômico da escola

Considerando o percentual de reprovação do 9º ano, pela Figura 5, percebe-se que o percentual de reprovação é menor quando o nível socioeconômico da escola aumenta. Só há 1 escola com nível socioeconômico muito alto, por isso não houve variação.

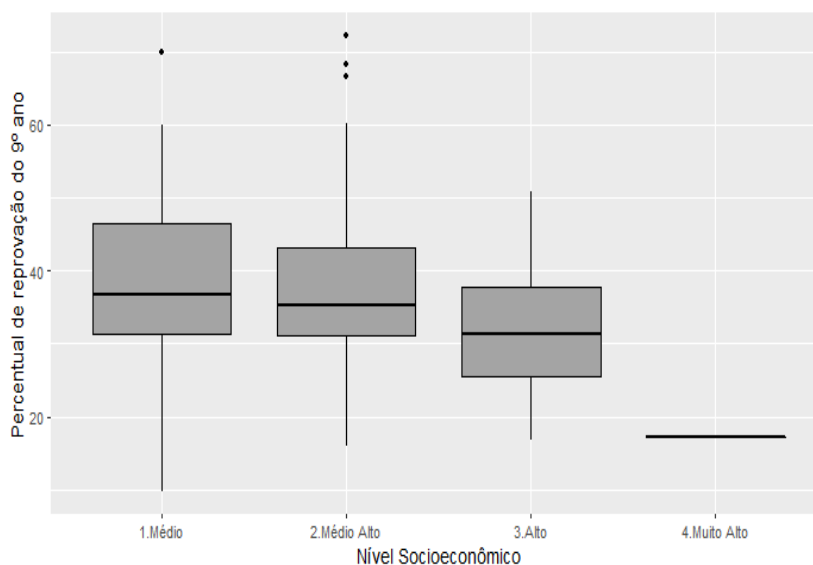


Figura 5: Percentual de reprovação do 9º ano por nível socioeconômico.

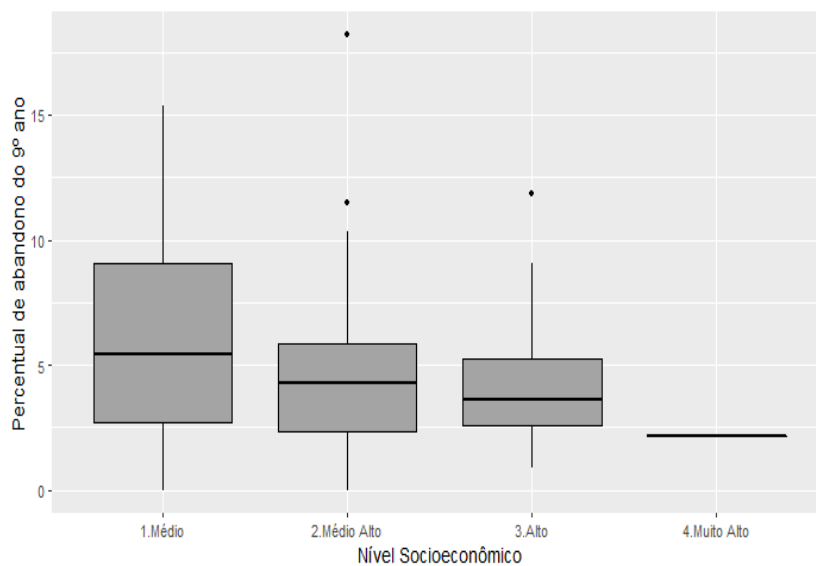


Figura 6: Percentual de abandono do 9º ano por nível socioeconômico.

Pela Figura 6, pode-se verificar que o percentual de abandono é menor quando o nível socioeconômico da escola aumenta. Quanto ao nível muito alto, apenas uma escola se encaixa no perfil.

Considerando os alunos do 5º ano, o percentual de reprovação e abandono são menores quanto maior for o nível socioeconômico da escola. Ainda temos apenas uma escola que se encaixa no perfil de nível socioeconômico muito alto (Figuras 7 e 8).

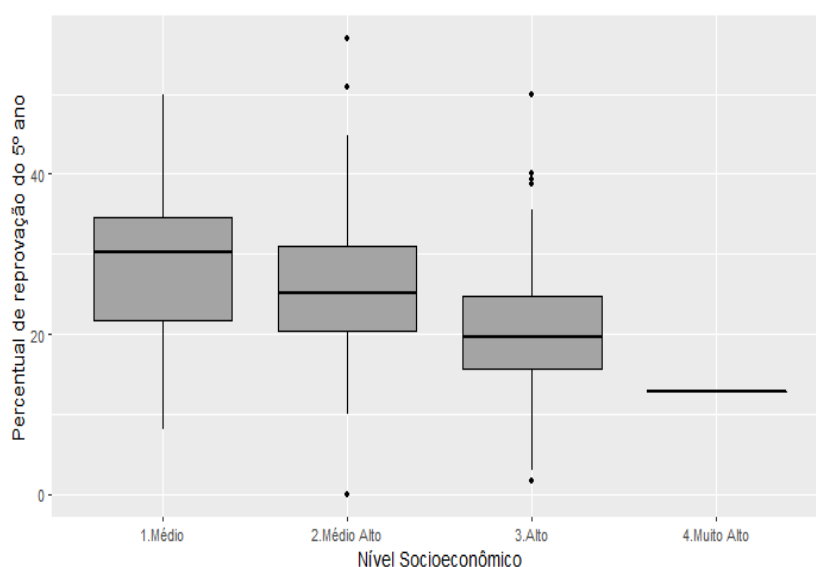


Figura 7: Percentual de reprovação do 5º ano por nível socioeconômico.

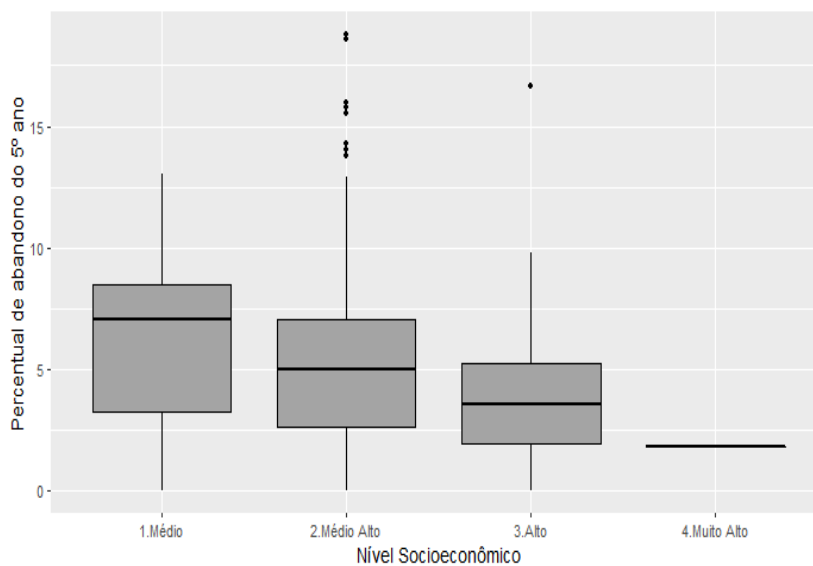


Figura 8: Percentual de abandono do 5º ano por nível socioeconômico.

3.1.4 Relação do percentual de abandono e reprovação e tamanho da escola

A fim de analisar a relação entre o número de alunos que abandonaram ou reprovaram o 9º ano e o tamanho da escola, foi calculado o coeficiente de correlação de Pearson. Os valores obtidos são 0,39 e 0,48 que indicam uma relação moderada. Pode-se dizer que a medida que o tamanho da escola aumenta, o número de alunos que reprovam ou abandonam a escola aumenta (Figuras 9 e 10).

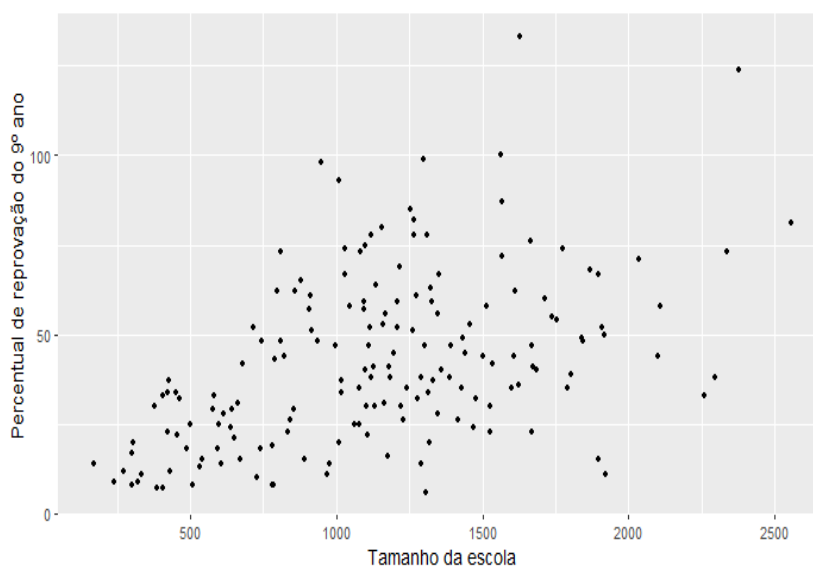


Figura 9: Relação entre percentual de reprovação do 9º ano e tamanho da escola.

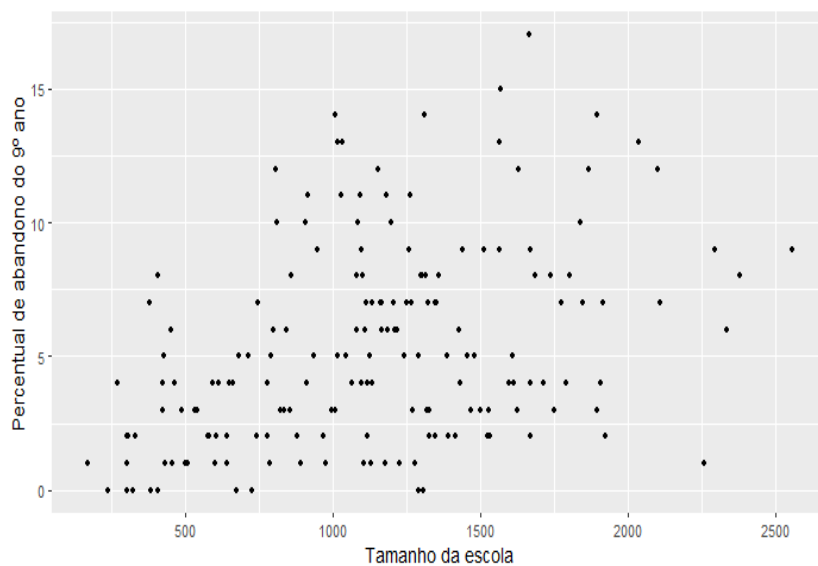


Figura 10: Relação entre percentual de abandono do 9º ano e tamanho da escola.

Os coeficientes de correlação de Pearson entre número de alunos que reprovaram e/ou abandonaram o 5º ano por tamanho da escola são 0.64 e 0.51, respectivamente, sendo mais fortes que para o 9º ano. Assim, as Figuras 11 e 12 mostram que a medida que o tamanho da escola aumenta, o número de alunos que abandonam e/ou reprovam aumenta.

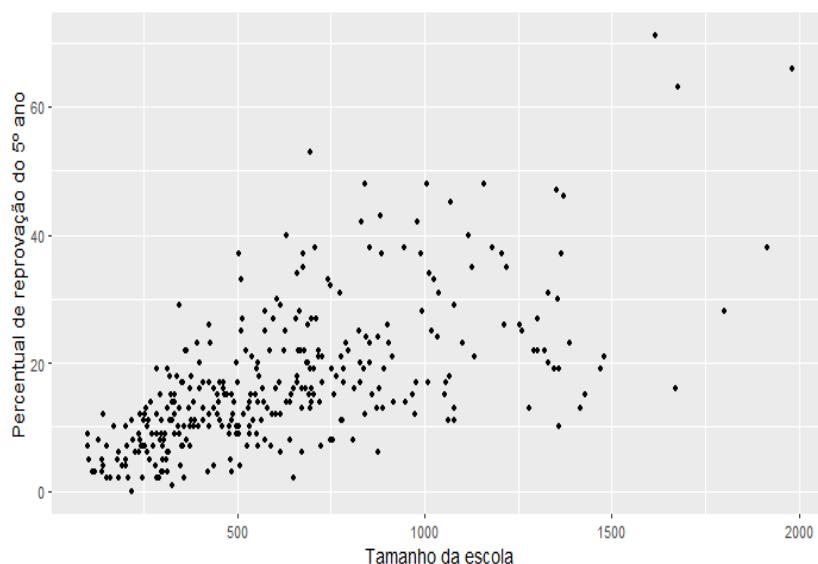


Figura 11: Relação entre percentual de reprovação do 5º ano e tamanho da escola.

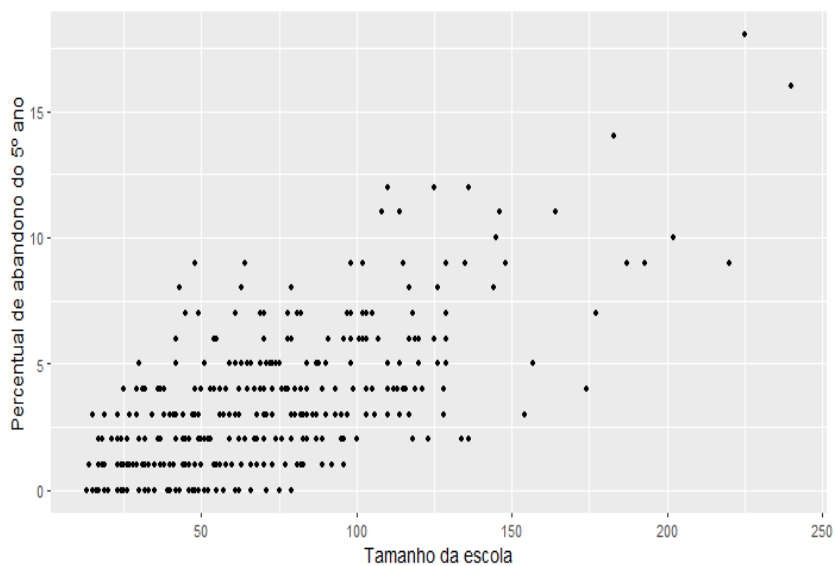


Figura 12: Relação entre percentual de abandono do 5º ano e tamanho da escola.

3.1.5 Relação do percentual de abandono e reprovação e escore de violência

Temos que o coeficiente de correlação de Pearson para o percentual de reprovação do 9º ano de cada escola e escore de violência é 0,13, indicando uma correlação fraca positiva. Para o percentual de abandono do 9º ano e escore de violência (Figura 14), a correlação é 2 vezes maior comparada à reprovação (0,27).

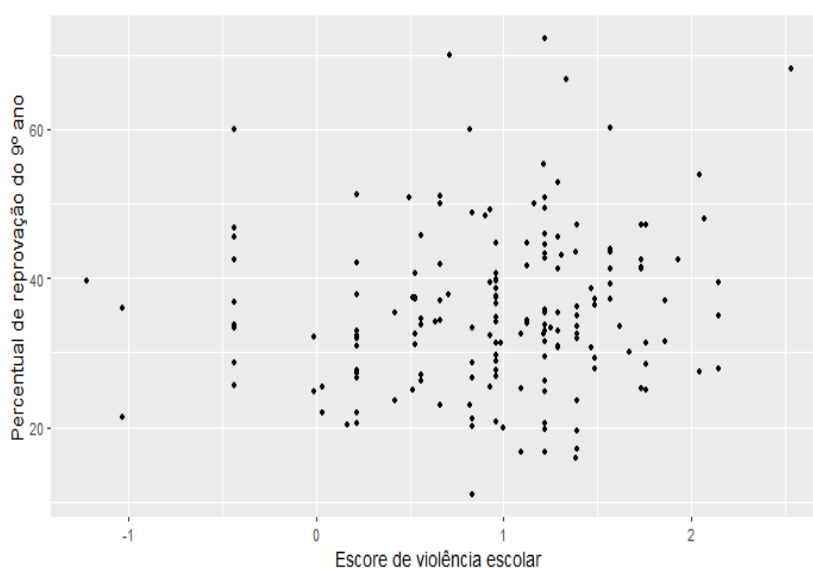


Figura 13: Percentual de reprovação do 9º ano por escore de violência.

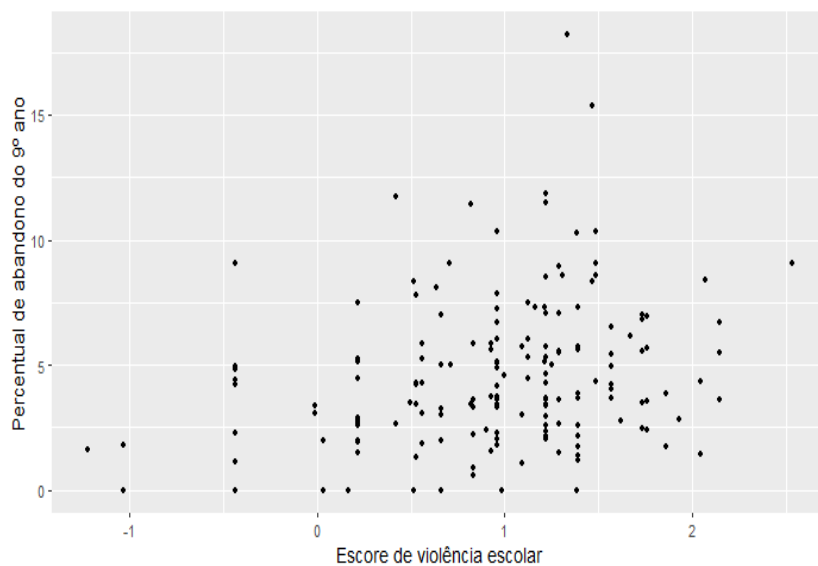


Figura 14: Percentual de abandono do 9º ano por escore de violência.

Temos que o coeficiente de correlação de Pearson do percentual de reprovação dos alunos do 5º e escore de violência foi de 0,15, indicando uma fraca correlação. A Figura 15 mostra a associação entre as variáveis. Para o total de abandono e escore de violência, obteve-se correlação fraca 0,09. Percebe-se que houve uma inversão quando comparado com os alunos do 9º ano. Pode-se pensar que a violência implica em maior abandono dos alunos do 9º ano e maior reprovação para alunos do 5º ano.

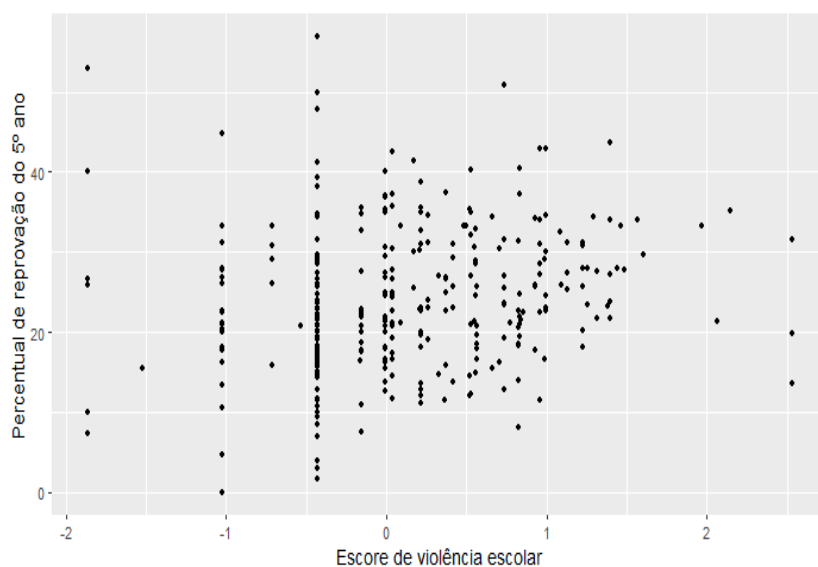


Figura 15: Percentual de reprovação do 5º ano por escore de violência.

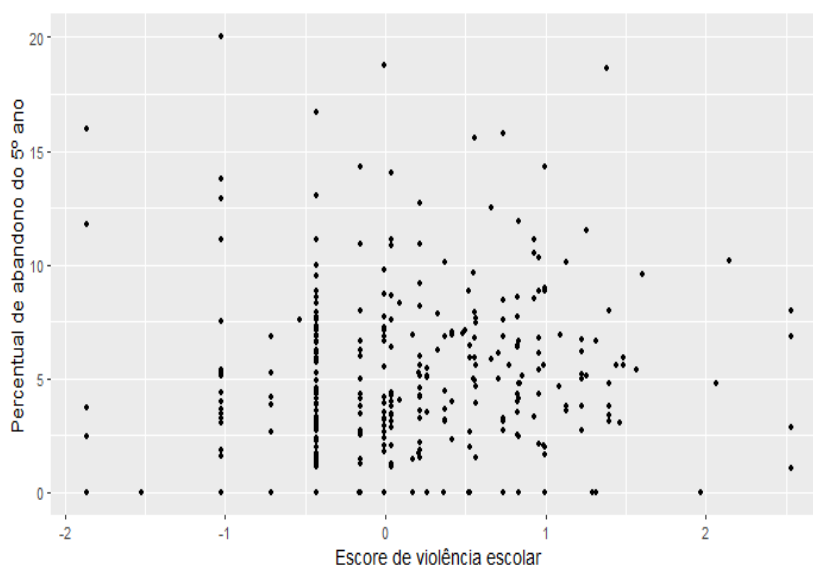


Figura 16: Percentual de abandono do 5º ano por score de violência.

3.2 Análise bivariada

3.2.1 Variáveis dos alunos do 5º ano e abandono e reprovação

A partir da Tabela 3 é possível avaliar a associação entre o abandono e características dos alunos do 5º ano do ensino fundamental. Considerando $\alpha = 0,05$, todas as variáveis têm associação com abandono dos alunos.

Tabela 3: Relação entre o abandono e características dos alunos.

Variável		Abandono			<i>p</i> -valor
		Não	Sim	Total	
Sexo	Masculino	11150 (93,9%)	721 (6,1%)	11871	< 0,001
	Feminino	11013 (96,4%)	415 (3,6%)	11428	
Raça/cor ⁴	Não negros	7423 (95,3%)	366 (4,7%)	7789	0,0292
	Negros	15119 (94,9%)	808 (5,1%)	15927	
Escolaridade da mãe	Ensino Médio ou Superior	8228 (96,5%)	302 (3,5%)	8530	< 0,001
	Outros	13873 (94,3%)	846 (5,7%)	14719	
Dever de português	Sim	21523 (95,5%)	1021 (4,5%)	22544	< 0,001
	Não	600 (83,8%)	116 (16,2%)	716	
Dever de matemática	Sim	21771 (95,4%)	1050 (4,6%)	22821	< 0,001
	Não	455 (82,7%)	95 (17,3%)	550	

⁴Alunos que responderam "não sei" ou não responderam foram retirados do banco. A categoria "Não negros" engloba brancos, amarelos e indígenas.

Pela tabela, nota-se que 96,4% das meninas e 93,9% dos meninos disseram não ter abandonado a escola. Percebe-se que quanto maior a escolaridade da mãe, menos alunos abandonam a escola. Há uma diferença de mais de 10% entre alunos que responderam fazer ou não dever de português ou matemática e não ter abandonado a escola. Esse resultado pode ser indicativo de que se existe engajamento e interesse dos alunos na realização das tarefas, a chance de abandonar a escola é menor.

Pela Tabela 4, temos que 71,3% dos meninos e 81,5% das meninas disseram não ter reprovado. Existe uma diferença de 10,2% entre a não reprovação entre meninos e meninas. Ainda, 83% dos alunos que a mãe tem pelo menos ensino médio disseram não para reprovação. Percebe-se que existe cerca de 13% de diferença entre alunos que disseram fazer ou não dever de matemática ou português, indicando mais uma vez que o esforço dos alunos com as atividades escolares principais é um forte indicativo de não reprovação.

Tabela 4: Relação entre a reprovação e características dos alunos.

Variável		Reprovação			<i>p</i> -valor
		Não	Sim	Total	
Sexo	Masculino	8400 (71,3%)	3382 (28,7%)	11782	< 0,001
	Feminino	9274 (81,5%)	2111 (18,5%)	11385	
Raça/cor	Não negros	6099 (78,5%)	1668 (21,5%)	7767	< 0,001
	Negros	11807 (74,6%)	4000 (25,4%)	15807	
Escolaridade da mãe	Ensino Médio ou Superior	7045 (83%)	1440 (17%)	8485	< 0,001
	Outros	10535 (72%)	4088 (28%)	14623	
Dever de português	Sim	17147 (76,5%)	5275 (23,5%)	22422	< 0,001
	Não	431 (62,2%)	262 (37,8%)	693	
Dever de matemática	Sim	17323 (76,3%)	5379 (23,7%)	22702	< 0,001
	Não	336 (63,5%)	193 (36,5%)	529	

3.2.2 Variáveis dos alunos do 9º ano e abandono e reprovação

Agora temos a análise de associação entre as variáveis dos alunos do 9º com abandono e reprovação. Considera-se $\alpha = 0.05$.

A partir da Tabela 5 é possível avaliar a associação entre o abandono e características dos alunos do 9º ano do ensino fundamental. Percebe-se que 95% dos meninos e 96,3% das meninas responderam que não abandonaram a escola. Apenas 4,3% dos alunos que se consideram negros disseram ter abandonado a escola, e 4,3% dos não negros responderam sim. Além disso, 96% dos alunos que disseram fazer o dever de matemática e/ou português, responderam não terem abandonado a escola.

Tabela 5: Relação entre a abandono e características dos alunos.

Variável		Abandono			<i>p-valor</i>
		Não	Sim	Total	
Sexo	Masculino	9806 (95%)	510 (5%)	10316	< 0,001
	Feminino	11189 (96,3%)	430 (3,7%)	11619	
Raça/cor	Não negros	6903 (95,7%)	313 (4,3%)	7216	0,0397
	Negros	14016 (95,7%)	636 (4,3%)	14652	
Escolaridade da mãe	Ensino Médio ou Superior	10848 (96,7%)	370 (3,3%)	11218	< 0,001
	Outros	10258 (94,6%)	589 (5,4%)	10847	
Dever de português	Sim	19983 (95,9%)	853 (4,1%)	20836	< 0,001
	Não	1214 (91,8%)	109 (8,2%)	1323	
Dever de matemática	Sim	19560 (96%)	813 (4%)	20373	< 0,001
	Não	1619 (91,5%)	150 (8,5%)	1769	

Na Tabela 6 é possível avaliar a associação entre reprovação e características dos alunos do 9º ano do ensino fundamental. Percebe-se que há uma diferença de mais de 10% entre os meninos e as meninas que disseram nunca ter reprovado. Essa diferença também existe entre alunos que a mãe tem pelo menos ensino médio ou têm no máximo ensino fundamental completo. Portanto, uma melhor qualificação na educação das mães interfere no rendimento dos alunos na escola. Além disso, 66% dos alunos que disseram fazer o dever de português e 66,6% que fazem o dever de matemática responderam não ter reprovado.

Tabela 6: Relação entre a reprovação e características dos alunos.

Variável		Reprovação			<i>p</i> -valor
		Não	Sim	Total	
Sexo	Masculino	6124 (59,5%)	4172 (40,5%)	10296	< 0,001
	Feminino	8287 (71,4%)	3312 (28,6%)	11599	
Raça/Cor	Não negros	4963 (68,9%)	2238 (31,1%)	7201	< 0,001
	Negros	9670 (64,1%)	5420 (35,9%)	15090	
Escolaridade da mãe	Ensino Médio ou Superior	7978 (71,2%)	3221 (28,8%)	11199	< 0,001
	Outros	6066 (60,3%)	3995 (39,7%)	10061	
Dever de português	Sim	13729 (66%)	7056 (34%)	20785	< 0,001
	Não	737 (61,3%)	465 (38,7%)	1202	
Dever de matemática	Sim	13549 (66,6%)	6782 (33,4%)	20331	< 0,001
	Não	962 (54,5%)	804 (45,5%)	1766	

3.3 Análise dos modelos

Para fazer o ajuste dos modelos foram utilizadas as variáveis explicativas, ao nível do aluno, idade, sexo, etnia, escolaridade da mãe, se faz dever de matemática e/ou português. Para o nível da escola temos o escore de violência interno. O objetivo é determinar se esses fatores influenciam a reprovação e o abandono dos alunos do 5º e 9º ano. Nesse caso, temos variáveis binárias como resposta, onde 1 indica reprovação/abandono e 0 indica não reprovação/abandono.

O critério utilizado para a escolha do modelo foi baseado no valor do coeficiente intraclasse. O modelo logístico multinível será aplicado para variável resposta que obteve o maior valor do coeficiente. Como a análise foi feita somente com escolas do Distrito Federal, acredita-se que a variância total é pouco explicada pelas escolas.

3.3.1 Reprovação no 5º ano

Inicialmente foi obtido o modelo nulo para variável resposta reprovação considerando a escola como nível 2. A Tabela 7 apresenta o resultado.

Tabela 7: Modelo nulo - Reprovação 5º ano

	Valor	Desvio padrão	p-valor
Intercepto	-1,202	0,029	<0,001
σ_u^2	0,1908	-	-
AIC	25703,0	-	

A variância entre grupos σ_u^2 foi igual a 0,1908. Assim, pode-se calcular o coeficiente de correlação intra-classe

$$\rho = \frac{\sigma_G^2}{\sigma_G^2 + 3,29} = \frac{0,1908}{0,1908 + 3,29} = 0,055. \quad (13)$$

Admitiu-se que o valor foi baixo, ou seja, a variância total é pouco explicada pela escola. Portanto, será utilizado modelo logístico sem considerar a estrutura multi-nível. Para composição do modelo, foram consideradas as variáveis de sexo, raça/cor, escolaridade da mãe, faz dever de matemática, faz dever de português e nível socioeconômico do aluno (Tabela 8).

Tabela 8: Modelo - Reprovação 5º ano

	Valor	p-valor	Razão de chances	IC [95%]
Intercepto	-0,8157	<0,001	0,442	[0,413; 0,474]
Sexo(Feminino)	-0,5993	<0,001	0,549	[0,515; 0,586]
Raça/cor(Negros)	0,1945	<0,001	1,214	[1,133; 1,302]
Escolaridade mãe (EM)	-0,5843	<0,001	0,557	[0,519; 0,599]
Dever português (não)	0,4125	<0,001	1,511	[1,246; 1,832]
Dever matemática (não)	0,3091	0,007	1,362	[1,088; 1,705]
Nível socioeconômico	-0,1939	<0,001	0,824	[0,807; 0,869]
AIC	23203			

Pode-se inferir que não fazer dever de português ou matemática tem forte relação com reprovação. Além disso, como colocado na análise descritiva, meninas têm menor probabilidade de reprovar. Analisando a razão de chance (*odds ratio*), temos que a chance de um aluno negro reprovar é 21,4% maior do que alunos não negros. Além disso, meninas têm 45,1% menos chance de reprovar em relação aos meninos. Quanto a escolaridade da mãe, estudantes que a mãe tem pelo menos ensino médio têm 44,3% menos chance de reprovar. Alunos que não fazem dever de português ou matemática têm 51,1% e 36,2% de chance de reprovar, respectivamente. A chance de um aluno reprovar é 17,6% menor quando há um aumento de 1 unidade no nível socioeconômico.

3.3.2 Abandono no 5º ano

O mesmo procedimento feito para modelo de reprovação do 5º foi aplicado para estudar a relação do abandono com as variáveis. A Tabela 9 apresenta o resultado para o modelo nulo.

Tabela 9: Modelo nulo - Abandono 5º ano

	Valor	Desvio padrão	p-valor
Intercepto	-3,0122	0,0396	<0,001
σ_u^2	0,1344	-	-
AIC	9326,5	-	

A variância entre grupos σ_u^2 foi igual a 0,1344. Agora pode-se calcular o coeficiente de correlação intra-classe

$$\rho = \frac{\sigma_G^2}{\sigma_G^2 + 3,29} = \frac{0,1344}{0,1344 + 3,29} = 0.039. \quad (14)$$

Como o valor obtido também é pequeno, será utilizado modelo logístico múltiplo. A Tabela 10 mostra o resultado para o modelo.

Tabela 10: Modelo - Abandono 5º ano

	Valor	p-valor	Razão de chance	IC [95%]
Intercepto	-2,6788	<0,001	0,069	[0,063; 0,075]
Sexo(Feminino)	-0,5247	<0,001	0,592	[0,520; 0,673]
Escolaridade mãe (EM)	-0,4319	<0,001	0,649	[0,563; 0,748]
Dever português (não)	0,9705	<0,001	2,639	[2,005; 3,474]
Dever matemática (não)	0,7665	<0,001	2,152	[1,575; 2,941]
Nível socioeconômico	-0,1123	0,0054	0,894	[0,826; 0,967]
AIC	8277,3	-		

Pode-se inferir que não fazer dever de português ou matemática implica em abandono do aluno. Além disso, meninas têm menor probabilidade de reprovar.

Pela razão de chance (*odds ratio*), percebe-se que meninas têm 40,8% menos chance de abandonar os estudos em relação aos meninos. Quanto a escolaridade da mãe, estudantes que a mãe tem pelo menos ensino médio têm 35,1% menos chance de abandono. Alunos que não fazem dever de português ou matemática têm 163,9% e 115,2% de chance de abandonar a escola, respectivamente. A chance de um aluno abandonar é 10,6% menor quando há um aumento de 1 unidade no nível socioeconômico.

3.3.3 Reprovação no 9º ano

A Tabela 11 apresenta o resultado para o modelo nulo para variável resposta reprovação dos alunos do 9º ano.

Tabela 11: Modelo nulo - Reprovação 9º ano

	Valor	Desvio padrão	p-valor
Intercepto	-0,709	0,044	<0,001
σ_u^2	0,3107	-	-
AIC	23586,8	-	-

A variância entre grupos σ_u^2 é 0,3107. Assim, temos que o coeficiente de correlação intra-classe é

$$\rho = \frac{\sigma_G^2}{\sigma_G^2 + 3,29} = \frac{0,3107}{0,3107 + 3,29} = 0.086. \quad (15)$$

Temos que 8,6% da variância total é explicada pela escola. Esse foi o modelo que apresentou a maior correlação intraclass (ICC). Portanto, para esse caso, será estimado o modelo logístico multinível.

Para fazer a seleção das variáveis foi utilizado o índice AIC e o p-valor ao nível de significância de 5%. A Tabela 12 mostra o resultado.

Tabela 12: Modelo - Reprovação 9º ano

	Valor	p-valor	Razão de chances	IC [95%]
Intercepto	-0,5384	<0,001	0,584	[0,511; 0,667]
Sexo(Feminino)	-0,5810	<0,001	0,559	[0,524; 0,597]
Raça/cor(Negros)	0,1195	0,002	1,127	[1,051; 1,209]
Escolaridade mãe (EM)	-0,3843	<0,001	0,681	[0,637; 0,728]
Nível socioeconômico	-0,1726	<0,001	0,841	[0,807; 0,877]
Dever matemática (não)	0,4665	<0,001	1,594	[1,419; 1,791]
Escore violência	0,2826	<0,001	1,327	[1,203; 1,462]

Temos que a chance de um aluno negro reprovar é 12,7% maior do que alunos não negros. Além disso, meninas têm 44,1% menos chance de reprovar em relação aos meninos. Quanto a escolaridade da mãe, estudantes que a mãe tem pelo menos ensino médio têm 31,9% menos chance de reprovar. Alunos que não fazem dever de matemática têm 59,4% de chance de reprovar. A chance de um aluno reprovar é 15,9% menor quando há um aumento de 1 unidade no nível socioeconômico. Ainda, a chance de um aluno reprovar é 32,7% maior com aumento de 1 unidade no escore

de violência.

Fazer ou não dever de português não foi significativo para dizer sobre reprovação. Uma hipótese a se considerar é que realizar as tarefas escolares pode não ser exatamente uma variável explicativa, mas um resultado no desempenho do aluno, em especial na reprovação.

3.3.4 Abandono no 9º ano

A Tabela 13 apresenta o resultado para o modelo nulo de abandono do 9º ano.

Tabela 13: Modelo nulo - Abandono 9º ano

	Valor	Desvio padrão	p-valor
Intercepto	-3,162	0,049	<0,001
σ_u^2	0,138	-	-
AIC	6698,8	-	

A variância entre grupos σ_u^2 é 0,138. Assim, temos que o coeficiente de correlação intra-classe é

$$\rho = \frac{\sigma_G^2}{\sigma_G^2 + 3,29} = \frac{0,1385}{0,1385 + 3,29} = 0,040. \quad (16)$$

O valor é baixo, ou seja, a variância total não é explicada pela escola. Portanto, será utilizado modelo logístico múltiplo.

As variáveis significativas no modelo são sexo, escolaridade da mãe, faz dever de matemática, faz dever de português e nível socioeconômico do aluno. A Tabela 14 mostra o resultado.

Tabela 14: Modelo nulo - Abandono 9º ano

	Valor	p-valor	Razão de chances	IC [95%]
Intercepto	-2,803	<0,001	0,061	[0,054; 0,068]
Sexo (Feminino)	-0,337	<0,001	0,714	[0,616; 0,830]
Escolaridade mãe (EM)	-0,468	<0,001	0,626	[0,537; 0,730]
Dever português (não)	0,460	<0,001	1,585	[1,225; 2,050]
Dever matemática (não)	0,556	<0,001	1,744	[1,389; 2,189]
Nível socioeconômico	-0,223	<0,001	0,799	[0,730; 0,876]

Pode-se inferir que não fazer dever de português ou matemática implica em reprovação. Além disso, como suposto na análise descritiva, meninas têm menor probabilidade de reprovar.

Temos que a chance de meninas do 9º ano abandonarem a escola é 29,6% menor do que os meninos. Quanto a escolaridade da mãe, estudantes que a mãe tem pelo menos ensino médio têm 37,4% menos chance de abandonar. Alunos que não fazem dever de português e/ou matemática têm 58,5% e 74,4% de chance de abandonar a escola, respectivamente. A chance de um aluno abandonar os estudos é 20,13% menor quando há um aumento de 1 unidade no nível socioeconômico.

4 Conclusão

Buscar compreender fatores que estão relacionados ao fracasso escolar é um dos maiores desafios da educação brasileira, uma vez que tem impacto sociais negativo. Esse é um problema que se estende desde o início da história educacional brasileira até hoje. Como foi mencionado no estudo de Silva e Araújo (2017), o Brasil tem a terceira maior taxa de abandono escolar entre os 100 países com maior IDH e no PNUD e a menor média de anos de estudos entre os países da América Latina.

Este estudo explorou fatores que podem estar associados ao abandono e/ou reprovação dos alunos do 5º e 9º anos das escolas públicas do Distrito Federal que participaram da Prova Brasil do Sistema Nacional de Avaliação da Educação Básica de 2015.

O escore de violência foi obtido por meio da Teoria de Resposta ao Item (TRI) com as respostas dos professores e diretores das escolas. Uma vez que as respostas são subjetivas, não podemos garantir que os valores dizem realmente sobre violência na escola. Ainda assim, vimos que violência implica em maior abandono dos alunos do 9º ano e maior reprovação para alunos do 5º ano.

Os resultados apontam que existe uma diferença de mais de 10% na reprovação dos alunos em que a mãe tem pelo menos ensino médio e aqueles que a mãe tem no máximo ensino fundamental completo. A base familiar é importante para um melhor desempenho dos alunos na vida acadêmica. Vimos que negros são maioria nas escolas públicas e o percentual dos negros de reprovação e abandono também é maior para as duas séries analisadas, ficando evidente que ainda existe discriminação.

As escolas públicas de Brazlândia e Planaltina têm maior percentual de reprovação dos alunos do 9º ano. Quanto ao abandono de alunos dessa mesma série, as escolas do Guará e Samambaia apresentam maior percentual. Agora, considerando alunos do 5º ano, as escolas públicas de São Sebastião apresentam maior percentual tanto de abandono como de reprovação, enquanto as do Guará e Plano Piloto apresentam menor percentual.

Vimos ainda que quanto maior o nível socioeconômico da escola, menor é o percentual médio de reprovação e abandono dos alunos do 5º e 9º ano. Outro fator importante a ser considerado é o tamanho da escola. Quanto maior o número de alunos em uma escola, maior o percentual de alunos que abandonam e/ou reprovam. Isso porque diminui a atenção individualizada dos docentes com os alunos. Fazer dever de português e/ou matemática interferem no abandono e reprovação. Vale questionar se essa é uma variável explicativa para reprovação e abandono ou se é resultante no desempenho dos alunos.

De todos os modelos apresentados, apenas reprovação do 9º ano foi considerado multinível, isto é, a escola interfere no processo. Foi decidido dessa forma pois foi o modelo que obteve maior coeficiente intraclasse. Aqui, estudantes que a mãe tem pelo menos ensino médio têm 31,9% menos chances de reprovar. Alunos que se consideram negros têm 12,7% mais chance de reprovar do que alunos que se autodeclaram não negros. Ainda, a chance de um aluno reprovar é 32,7% maior com aumento de 1 unidade no escore de violência.

A educação brasileira, de uma maneira geral, enfrenta diversas dificuldades que precisam ser estudadas. Nesse trabalho, consideramos apenas alguns fatores que podem interferir no fluxo escolar dos alunos com objetivo de fornecer informações para possíveis questionamentos e análises futuras.

Referências

- [1] A. AGRESTI. *An Introduction to Categorical Data Analysis*, volume 2.ed. Wiley-Interscience, 2007.
- [2] P. H. F. Araújo. Regressão logística multinível: um estudo sobre o consumo de álcool e inalantes entre crianças e adolescentes, 2014.
- [3] C. M. D. A. BICA. Evasão escolar: Os comprometimentos da má qualidade da escola. *APEOC*, 2013.
- [4] M. R. de Souza. Violência nas escolas: Causas e consequências. *Caderno Discente do Instituto Superior de Educação*, 2008.
- [5] N. A. M.-F. Fernanda Leite Lopez de Leon. Reprovação, avanço e evasão escolar no brasil. *Revista Pesquisa e Planejamento Econômico*, 32(3), 2002.
- [6] J. J. Hox. *Multilevel Analysis: Techniques and Applications. Quantitative Methodology Series*, volume 2.ed. Routledge, 2010.
- [7] D. W. Iijima and T. M. R. Schroeder. Pesquisa sobre violência escolar no brasil. *Travessias*, 6(3), 2015.
- [8] G. L. S. M. Antonio Amaral Turkman. Modelos lineares generalizados - da teoria à prática. *PRAXIS XXI- FEDER*, 2000.
- [9] G. S. A. Maria Isabel Ramalho Ortiglão. Repetência escolar nos anos iniciais do ensino fundamental: evidências a partir dos dados da prova brasil 2009I. *Revista Brasileira de Estudos Pedagógicos.*, 94(237), 2013.
- [10] L. C. Marriel, S. G. de Asis, J. Q. Avanci, and R. V. Oliveira. Violência escolar e auto-estima de adolescentes. *Cadernos de pesquisa*, 36(127):35–50, 2006.
- [11] R. M. d. L. A. Raimundo Barbosa Silva Filho. Evasão e abandono escolar na educação básica no brasil: fatores, causas e possíveis consequências. *Revista eletrônica Educação Por Escrito*, 8(1), 2017.
- [12] A. L. M. M. Rocha. Regressão logística multinível: Uma aplicação de modelos lineares generalizados mistos, 2017.
- [13] L. G. A. Vinha. Validação das escalas de percepção de violência e conflitos nas escolas. manuscrito não publicado. 2018.