

Desigualdade de educação no Brasil. evidências para o ano de 2023

André Santos

Universidade Federal da Paraíba

`andre.brandao@academico.ufpb.br`

January 31, 2025

Abstract

Este artigo investiga as desigualdades de gênero e raça na educação, analisando a probabilidade de conclusão dos níveis fundamental, médio e superior no Brasil. Utilizando dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) de 2023, foram estimados modelos de regressão logística para avaliar o impacto das variáveis `HOMEM` (sexo masculino) e `BRANCO` (raça branca) sobre a conclusão dos três níveis educacionais. Os resultados revelaram padrões de desigualdade que favorecem majoritariamente pessoas brancas e do sexo feminino no nível superior. Os resultados destacam a necessidade de políticas públicas específicas para reduzir as desigualdades de gênero e raça na educação, como programas de retenção escolar para homens jovens e ações afirmativas para populações não brancas no ensino superior. O estudo também aponta limitações, como a ausência de variáveis socioeconômicas e a não consideração de interações entre gênero e raça, sugerindo direções para pesquisas futuras. Conclui-se que a promoção da equidade educacional requer intervenções direcionadas e contínuas, garantindo oportunidades iguais para todos os indivíduos, independentemente de gênero ou raça.

1 Introdução

A desigualdade na educação é uma das principais barreiras para o desenvolvimento socioeconômico do Brasil, um país caracterizado por sua vasta diversidade cultural e regional. Embora o acesso à educação tenha aumentado significativamente nas últimas décadas, permanecem disparidades marcantes em termos de qualidade, infraestrutura e resultados educacionais entre diferentes grupos socioeconômicos e regiões do país. Essas desigualdades são reflexo de fatores históricos, econômicos e políticos que moldaram o sistema educacional brasileiro. No contexto de 2023, novas evidências destacam a persistência e, em alguns casos, o agravamento dessas desigualdades, especialmente em um cenário de recuperação econômica e social após os impactos da pandemia de COVID-19. A pandemia aprofundou lacunas já existentes, como o acesso desigual à tecnologia e a recursos de ensino remoto, afetando especialmente estudantes em situações de vulnerabilidade. Este artigo busca explorar as evidências mais recentes sobre as desigualdades educacionais no Brasil em 2023, analisando aspectos como desigualdade de gênero e raça.

2 Revisão da Literatura

A educação é um dos principais indicadores de desenvolvimento e bem-estar social, sendo central nos Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas. Dentro dessa perspectiva, diversos estudos analisam diferentes aspectos da gestão educacional e da desigualdade na educação, fornecendo evidências empíricas sobre o impacto de políticas públicas e dos fatores estruturais na aprendizagem dos alunos.

Um primeiro eixo de análise se refere à gestão escolar e seus impactos na qualidade da educação. Estudo realizado sobre um programa piloto de gestão escolar na rede estadual de São Paulo, implantado em 2008, demonstrou efeitos positivos sobre o desempenho dos alunos em matemática na 8^a série, mas sem impacto significativo nas notas de língua portuguesa. O programa envolveu treinamento de gestores, estabelecimento de metas e monitoramento de indicadores de aprendizagem. A análise, conduzida por meio da metodologia de regressão com descontinuidade, sugere que mudanças na gestão podem melhorar aspectos específicos do desempenho estudantil.

Outro aspecto relevante é o impacto da remuneração dos professores sobre a aprendizagem dos alunos. Na rede estadual paulista, professores recebem aumentos salariais regulares baseados no tempo de serviço, além de bonificações vinculadas ao desempenho dos estudantes em testes padronizados. No entanto, estudos que utilizaram estratégias de diferenças duplas e triplas não encontraram evidências conclusivas de que o aumento dos salários fixos ou a presença de bonificações influenciem significativamente a proficiência dos alunos.

A desigualdade educacional também é um tema central na literatura. Uma análise da desigualdade de oportunidades na educação identificou que fatores relacionados às redes de ensino e às escolas têm um papel substancial na formação do aprendizado dos alunos. Estima-se que 50

No contexto internacional, a modelagem da distribuição dos anos de escolaridade permite explorar tendências da desigualdade educacional desde 1970 e prever o progresso em direção às metas dos ODS para 2030. Projeções indicam que a educação primária universal estará próxima de ser atingida em 2030, mas desafios persistem nas taxas de conclusão do ensino secundário e superior. Além disso, observa-se uma redução na desigualdade educacional global desde o pico atingido em 2017, e a expectativa é de uma tendência decrescente até 2030.

No que tange à desigualdade de gênero, estudos apontam que a diferença na escolarização entre homens e mulheres foi praticamente eliminada até 2018 em escala global. Contudo, persistem disparidades significativas em regiões como África Subsaariana, Norte da África e Oriente Médio. As projeções indicam que, até 2030, as mulheres terão alcançado maior nível educacional do que os homens em 18 países.

Dessa forma, a literatura revisada destaca que a gestão escolar, a remuneração docente e as políticas educacionais exercem impactos variados sobre a qualidade do ensino e a desigualdade educacional. Compreender essas relações é fundamental para o desenvolvimento de políticas públicas mais eficazes que promovam uma distribuição mais equitativa do capital humano e a construção de sociedades mais justas.

3 Metodologia e base de dados

3.1 Modelo Econométrico

Para analisar os determinantes da conclusão dos níveis de educação fundamental, médio e superior, utilizamos um modelo de regressão logística (*logit*). A escolha do modelo *logit* é apropriada, uma vez que a variável dependente é binária, assumindo o valor 1 se o indivíduo completou o nível educacional em questão e 0 caso contrário. Formalmente, o modelo é especificado da seguinte forma:

Seja Y_i uma variável binária que indica se o indivíduo i completou um determinado nível educacional. A probabilidade de $Y_i = 1$ condicional às variáveis explicativas é modelada como:

$$P(Y_i = 1|X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 HOMEM_i + \beta_2 BRANCO_i)}}$$

onde:

- $HOMEM_i$ é uma variável binária que assume o valor 1 se o indivíduo i é do sexo masculino e 0 caso contrário;
- $BRANCO_i$ é uma variável binária que assume o valor 1 se o indivíduo i se declara branco e 0 caso contrário;
- β_0 é o intercepto do modelo;
- β_1 e β_2 são os coeficientes associados às variáveis $HOMEM_i$ e $BRANCO_i$, respectivamente.

A função de verossimilhança para o modelo *logit* é dada por:

$$\mathcal{L}(\beta) = \prod_{i=1}^N [P(Y_i = 1|X_i)^{Y_i} \cdot (1 - P(Y_i = 1|X_i))^{1-Y_i}]$$

onde N é o número de observações na amostra. Os parâmetros β são estimados por máxima verossimilhança (*Maximum Likelihood Estimation* - MLE), maximizando a função de log-verossimilhança:

$$\ln \mathcal{L}(\beta) = \sum_{i=1}^N [Y_i \ln P(Y_i = 1|X_i) + (1 - Y_i) \ln(1 - P(Y_i = 1|X_i))]$$

3.2 Especificações do Modelo

O modelo *logit* é estimado de forma equivalente para três variáveis dependentes distintas, representando a conclusão dos seguintes níveis educacionais:

1. **Fundamental Completo:** $Y_i = 1$ se o indivíduo completou o ensino fundamental e $Y_i = 0$ caso contrário.
2. **Ensino Médio Completo:** $Y_i = 1$ se o indivíduo completou o ensino médio e $Y_i = 0$ caso contrário.
3. **Ensino Superior Completo:** $Y_i = 1$ se o indivíduo completou o ensino superior e $Y_i = 0$ caso contrário.

Para cada uma dessas variáveis dependentes, as variáveis explicativas $HOMEM_i$ e $BRANCO_i$ são incluídas no modelo, conforme descrito anteriormente.

3.3 Bases de Dados

A base de dados utilizada neste estudo é a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), referente ao ano de 2023. A PNAD Contínua é uma pesquisa domiciliar realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que fornece informações detalhadas sobre características demográficas, socioeconômicas e educacionais da população brasileira.

Para este estudo, foram selecionados os indivíduos com idade superior a 25 anos, uma vez que essa faixa etária é considerada adequada para analisar a conclusão dos níveis educacionais de interesse. A amostra final consiste em N observações, após a exclusão de casos com informações faltantes (*missing values*) nas variáveis de interesse.

3.4 Tratamento dos Dados

As variáveis utilizadas no modelo foram construídas da seguinte forma:

- **Fundamental Completo:** Indivíduos que declararam ter concluído o 9º ano do ensino fundamental, sem nível médio ou superior.
- **Ensino Médio Completo:** Indivíduos que declararam ter concluído o 3º ano do ensino médio, sem nível superior.
- **Ensino Superior Completo:** Indivíduos que declararam ter concluído um curso de graduação.
- **HOMEM:** Indivíduos do sexo masculino (categoria de referência: sexo feminino).
- **BRANCO:** Indivíduos que se declararam brancos (categoria de referência: não brancos).

3.5 Testes de Diagnóstico do Modelo

Para garantir a robustez e a adequação do modelo de regressão logística, foram realizados testes de diagnóstico para avaliar a presença de multicolinearidade e heterocedasticidade.

3.5.1 Teste de Multicolinearidade

A multicolinearidade ocorre quando há alta correlação entre as variáveis explicativas, o que pode inflacionar os erros padrão dos coeficientes e comprometer a precisão das estimativas. Para detectar a presença de multicolinearidade, utilizou-se o *Fator de Inflação da Variância* (VIF). O VIF é calculado para cada variável explicativa, e valores superiores a 10 indicam problemas de multicolinearidade [?]. No presente estudo, todas as variáveis apresentaram valores de VIF inferiores a 10, indicando que a multicolinearidade não é uma preocupação no modelo estimado, conforme demonstrado na tabela abaixo:

Table 1: Resultados do Teste de Fator de Inflação da Variância (VIF) para os Modelos de Ensino Fundamental, Médio e Superior

Nível de Ensino	HOMEM	BRANCO
Ensino Fundamental	1.000187	1.000187
Ensino Médio	1.000206	1.000206
Ensino Superior	1.000054	1.000054

Nota: Valores de VIF próximos a 1 indicam ausência de multicolinearidade.

3.5.2 Teste de Heterocedasticidade

A heterocedasticidade ocorre quando a variância dos resíduos do modelo não é constante ao longo das observações, o que pode levar a inferências estatísticas incorretas. Para verificar a presença de heterocedasticidade, foi utilizado o teste de Breusch-Pagan [?]. A hipótese nula do teste é que os resíduos são homocedásticos. No presente estudo, o teste de Breusch-Pagan rejeitou a hipótese nula ($p > 0,05$), indicando que há evidências de heterocedasticidade nos resíduos do modelo. Como medida adicional de robustez, foram utilizados erros padrão robustos nas estimativas dos coeficientes.

Table 2: Resultados do Teste de Breusch-Pagan para Heterocedasticidade nos Modelos de Ensino Fundamental, Médio e Superior

Nível de Ensino	Estatística BP	Graus de Liberdade (df)	Valor-p
Ensino Fundamental	787.99	2	$< 2.2 \times 10^{-16}$
Ensino Médio	135.87	2	$< 2.2 \times 10^{-16}$
Ensino Superior	37918	2	$< 2.2 \times 10^{-16}$

Nota: Um valor-p < 0.05 indica evidências de heterocedasticidade.

4 Resultados

Nesta seção, apresentamos os resultados dos modelos de regressão logística estimados para os três níveis de ensino: fundamental, médio e superior. Os coeficientes estimados, erros padrão, valores-z e valores-p são analisados para avaliar o impacto das variáveis explicativas **HOMEM** e **BRANCO** sobre a probabilidade de conclusão de cada nível educacional. Adicionalmente, discutimos a significância estatística dos coeficientes e suas implicações práticas.

4.1 Modelo para o Ensino Fundamental

O modelo de regressão logística para o ensino fundamental apresentou os seguintes resultados:

$$\text{Logito}(Y) = -1.9688 + 0.0880 \cdot \text{HOMEM} - 0.1256 \cdot \text{BRANCO} \quad (1)$$

- **Intercepto:** O coeficiente para o intercepto é -1.9688 ($z = -447.659$, $p < 0.001$), indicando que, quando todas as variáveis explicativas são zero, a probabilidade de conclusão do ensino fundamental é baixa.
- **HOMEM:** O coeficiente para a variável **HOMEM** é 0.0880 ($z = 16.165$, $p < 0.001$), indicando que ser do sexo masculino aumenta a probabilidade de conclusão do ensino fundamental.
- **BRANCO:** O coeficiente para a variável **BRANCO** é -0.1256 ($z = -22.463$, $p < 0.001$), sugerindo que indivíduos brancos têm menor probabilidade de conclusão do ensino fundamental.

4.2 Modelo para o Ensino Médio

O modelo de regressão logística para o ensino médio apresentou os seguintes resultados:

$$\text{Logito}(Y) = -0.8002 - 0.0451 \cdot \text{HOMEM} - 0.0012 \cdot \text{BRANCO} \quad (2)$$

- **Intercepto:** -0.8002 ($z = -256.8108$, $p < 0.001$), indicando um logito negativo da probabilidade de conclusão do ensino médio.
- **HOMEM:** -0.0451 ($z = -11.6520$, $p < 0.001$), sugerindo menor probabilidade de conclusão do ensino médio para homens.
- **BRANCO:** -0.0012 ($z = -0.3132$, $p = 0.7541$), sem significância estatística para a variável **BRANCO**.

4.3 Modelo para o Ensino Superior

O modelo de regressão logística para o ensino superior apresentou os seguintes resultados:

$$\text{Logito}(Y) = -1.8466 - 0.3733 \cdot \text{HOMEM} + 0.8487 \cdot \text{BRANCO} \quad (3)$$

- **Intercepto:** -1.8466 ($z = -447.304$, $p < 0.001$), sugerindo um logito negativo na ausência dos efeitos das variáveis explicativas.

- **HOMEM:** -0.3733 ($z = -75.919$, $p < 0.001$), indicando menor probabilidade de conclusão do ensino superior para homens.
- **BRANCO:** 0.8487 ($z = 174.759$, $p < 0.001$), indicando maior probabilidade de conclusão do ensino superior para indivíduos brancos.

4.4 Tabela Resumo dos Coeficientes

	Estimate	Std. Error	z value	Pr(> z)
Ensino Fundamental				
(Intercept)	-1.9688	0.0044	-447.659	< 2.2e-16
HOMEM	0.0880	0.0054	16.165	< 2.2e-16
BRANCO	-0.1256	0.0056	-22.463	< 2.2e-16
Ensino Médio				
(Intercept)	-0.8002	0.0031	-256.811	< 2e-16
HOMEM	-0.0451	0.0039	-11.652	< 2e-16
BRANCO	-0.0012	0.0039	-0.3132	0.7541
Ensino Superior				
(Intercept)	-1.8466	0.0041	-447.304	< 2.2e-16
HOMEM	-0.3733	0.0049	-75.919	< 2.2e-16
BRANCO	0.8487	0.0049	174.759	< 2.2e-16

Table 3: Resultados do teste z para os coeficientes dos modelos de regressão logística.

Os modelos de regressão logística revelam padrões importantes sobre desigualdades de gênero e raça na educação. As políticas públicas devem considerar essas diferenças para promover maior equidade no acesso e conclusão dos níveis educacionais. Estudos futuros podem explorar outros fatores explicativos e interações entre variáveis para aprofundar a compreensão desses fenômenos.

5 Discussão

5.1 Discussão dos Resultados

Os resultados dos modelos de regressão logística para os três níveis de ensino (fundamental, médio e superior) revelam padrões interessantes e significativos sobre as desigualdades de gênero e raça na educação. A seguir, discutimos as implicações dos coeficientes estimados, sua significância estatística e as possíveis razões para os padrões observados.

5.1.1 Desigualdades de Gênero (Variável HOMEM)

A variável **HOMEM** apresentou efeitos distintos nos três níveis de ensino:

- **Ensino Fundamental:** O coeficiente positivo (0.0880, $p < 0.001$) indica que os homens têm maior probabilidade de concluir o ensino fundamental em comparação com as mulheres. Esse resultado pode refletir diferenças culturais ou socioeconômicas que favorecem os homens no acesso e permanência na educação básica. Em

algumas regiões, meninos podem ter maior incentivo para permanecer na escola, enquanto meninas podem enfrentar barreiras como trabalho doméstico ou casamento precoce.

- **Ensino Médio:** O coeficiente negativo (-0.0451 , $p < 0.001$) sugere que os homens têm menor probabilidade de concluir o ensino médio em comparação com as mulheres. Esse padrão pode estar relacionado a fatores como evasão escolar, engajamento em atividades laborais precoces ou maior exposição a riscos sociais entre os homens jovens.
- **Ensino Superior:** O coeficiente negativo (-0.3733 , $p < 0.001$) indica que os homens têm menor probabilidade de concluir o ensino superior em comparação com as mulheres. Esse resultado está alinhado com tendências globais que mostram maior participação feminina no ensino superior, possivelmente devido a políticas de inclusão, maior desempenho acadêmico das mulheres ou mudanças culturais que incentivam a educação feminina.

5.1.2 Desigualdades Raciais (Variável BRANCO)

A variável **BRANCO** também apresentou efeitos variados nos três níveis de ensino:

- **Ensino Fundamental:** O coeficiente negativo (-0.1256 , $p < 0.001$) sugere que indivíduos brancos têm menor probabilidade de concluir o ensino fundamental em comparação com não brancos. Esse resultado pode refletir desigualdades socioeconômicas, onde populações não brancas, em alguns contextos, podem ter maior acesso à educação básica devido a políticas de inclusão ou menor exposição a barreiras educacionais.
- **Ensino Médio:** O coeficiente para **BRANCO** não foi estatisticamente significativo (-0.0012 , $p = 0.7541$), indicando que não há diferenças significativas na probabilidade de conclusão do ensino médio entre brancos e não brancos. Esse resultado sugere que, nesse nível de ensino, as desigualdades raciais podem ser menos pronunciadas ou que políticas de inclusão têm sido eficazes em reduzir essas disparidades.
- **Ensino Superior:** O coeficiente positivo (0.8487 , $p < 0.001$) indica que indivíduos brancos têm maior probabilidade de concluir o ensino superior em comparação com não brancos. Esse padrão reflete desigualdades estruturais profundas, onde populações brancas têm maior acesso a recursos educacionais, melhores escolas e maior suporte familiar para continuar os estudos em nível superior.

5.1.3 Significância Estatística e Magnitude dos Efeitos

Todos os coeficientes, exceto para **BRANCO** no ensino médio, foram estatisticamente significativos ($p < 0.001$), indicando que as variáveis **HOMEM** e **BRANCO** têm efeitos robustos e consistentes sobre a probabilidade de conclusão dos diferentes níveis de ensino. A magnitude dos coeficientes também revela que:

- No ensino fundamental, o efeito de **HOMEM** é positivo, mas relativamente pequeno (0.0880), enquanto o efeito de **BRANCO** é negativo e de magnitude moderada (-0.1256).

- No ensino médio, o efeito de **HOMEM** é negativo e pequeno (-0.0451), enquanto o efeito de **BRANCO** é praticamente nulo (-0.0012).
- No ensino superior, o efeito de **HOMEM** é negativo e de magnitude considerável (-0.3733), enquanto o efeito de **BRANCO** é positivo e muito forte (0.8487).

5.1.4 Limitações e Sugestões para Estudos Futuros

Embora os modelos tenham fornecido insights valiosos, algumas limitações devem ser consideradas:

- O estudo não inclui outras variáveis explicativas relevantes, como renda familiar, localização geográfica ou qualidade da escola.
- A análise não considera interações entre variáveis, como o efeito combinado de gênero e raça.

6 Considerações finais

Os modelos de regressão logística revelaram padrões significativos de desigualdade de gênero e raça na educação brasileira. Homens têm maior probabilidade de concluir o ensino fundamental, mas menor probabilidade nos níveis médio e superior. Indivíduos brancos apresentam menor probabilidade de conclusão no ensino fundamental, mas maior no ensino superior. Esses resultados destacam a necessidade de políticas públicas direcionadas, como programas de retenção escolar para homens jovens e ações afirmativas para populações não brancas no ensino superior. A promoção da equidade educacional requer intervenções contínuas e específicas, garantindo oportunidades iguais para todos, independentemente de gênero ou raça.