



Educação, Desigualdade e Desenvolvimento: Uma Análise dos Indicadores Sociais em Pernambuco

Jamerson carlos chaves da silva¹ * - <https://orcid.org/0009-0007-6033-6250>

Winicius Antonio Souza Silva² - <https://orcid.org/0009-0005-7934-1303>

Thyago Celso Cavalcante Nepomuceno¹ * - <https://orcid.org/0000-0001-8327-6472>

Jean Gomes Turet² - <https://orcid.org/0000-0003-3608-1706>

¹ Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brasil

Email: jamerson.chaves@ufpe.br, thyago.nepomuceno@ufpe.br

² Centro Acadêmico do Agreste, Universidade Federal de Pernambuco, Caruaru, Brasil

Email: winicius.silva@ufpe.br, jean.turet@ufpe.br

RESEARCH ARTICLE

Annals of Data Analysis

<https://periodicos.ufpe.br/revistas/ADA/>

Submitted on: 12.12.2025.

Accepted on: 01.08.2026.

Published on: 01.08.2026.

Copyright © 2026 by author(s).

This work is licensed under the Creative Commons Attribution International License CC BY-NC-ND 4.0

<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>



Resumo

O estudo investiga como diferentes indicadores sociais, econômicos e educacionais se relacionam nos municípios de Pernambuco, a partir da aplicação de modelos de regressão linear múltipla. Foram analisadas variáveis como o Índice de Desenvolvimento da Educação Básica (IDEB), o Índice de Desenvolvimento Humano Municipal por Escolaridade (IDHM-E), o PIB per capita, a taxa de criminalidade violenta e a expectativa de vida. Os resultados mostraram que o IDHM-E tem papel central nessas interações, funcionando tanto como variável explicativa quanto como variável resposta. Em especial, destacou-se sua forte influência positiva sobre o PIB per capita ($R^2 = 0,948$), além de associações significativas com expectativa de vida e IDEB. Por outro lado, os modelos voltados à criminalidade violenta apresentaram baixo poder explicativo, indicando a complexidade desse fenômeno. As conclusões reforçam a importância de políticas públicas voltadas à educação básica como motor de transformação social e econômica, ainda que o enfrentamento da violência urbana demande estratégias mais amplas e integradas.

Keywords

Educação; Desenvolvimento Humano; Regressão; Pernambuco

1. Introdução

A problemática dos impactos da desigualdade nas condições de saúde, educação e criminalidade infantil tem sido amplamente abordada por diversos estudos e instituições de pesquisa. Segundo Bastos (2016), a exclusão

e a pobreza caminham juntas, sendo ainda mais preocupantes quando afetam crianças, o grupo etário mais vulnerável a essas condições. Além disso, o Instituto de Pesquisa Econômica Aplicada (IPEA) destaca que a educação é uma das principais políticas sociais na redução dos crimes de homicídio, indicando a relevância de investimentos nesse setor para a diminuição da violência (CAEDJUS, 2021). Esses indicadores sociais carregam consigo uma responsabilidade intrínseca: estimular a reflexão e o debate na sociedade sobre estratégias que possam contribuir para sua melhoria.

Diante dessa complexidade, torna-se essencial utilizar ferramentas analíticas que permitam compreender como diferentes fatores sociais se inter-relacionam e influenciam resultados como a criminalidade. Nesse contexto, a regressão linear múltipla tem se mostrado uma abordagem valiosa, permitindo examinar de forma sistemática a relação entre variáveis como ensino, desigualdade e criminalidade.

Ao se analisar fenômenos sociais complexos, é importante considerar a possibilidade de causalidade simultânea, situação em que duas variáveis exercem influência recíproca, dificultando a distinção entre causa e efeito. Como destacam Lima et al. (2024), essa dinâmica pode ser observada, por exemplo, na relação entre patrulhamento policial e ocorrência de crimes, em que ambos os fatores se influenciam mutuamente. Essa perspectiva amplia a compreensão sobre os mecanismos que sustentam a criminalidade e ressalta a importância de métodos econométricos robustos para a formulação de inferências mais confiáveis.

Em um estudo recente no estado de Pernambuco, Moura e Monteiro (2024) aplicaram modelos de regressão linear múltipla ordinária para analisar como fatores educacionais e desigualdade econômica impactam a criminalidade patrimonial nos municípios pernambucanos, oferecendo insights robustos para a formulação de políticas públicas mais integradas e eficazes.

No que se refere ao índice de Desenvolvimento Humano Municipal por Escolaridade (IDHM-E), observou-se avanço expressivo entre os anos de 1991 e 2010 nas capitais da região Nordeste. Entretanto, esse crescimento não se refletiu de forma direta no Índice de Desenvolvimento da Educação Básica (IDEB), evidenciando a ausência de uma correlação linear consistente entre os indicadores (IPEA, 2016). Resultados semelhantes foram identificados em municípios da Região Metropolitana do Vale do Paraíba e Litoral Norte, onde ocorreram progressos no IDHM mesmo em localidades que apresentaram baixo desempenho no IDEB, o que pode indicar melhorias educacionais potenciais no médio e longo prazo (RIBEIRO; SCHLEGEL, 2015).

Este estudo propõe aplicar a regressão linear múltipla para investigar a relação entre diferentes indicadores sociais no Estado de Pernambuco. De forma distinta da abordagem usual, a variável resposta será alternada em cinco modelos distintos, possibilitando explicar cada indicador a partir dos demais e permitindo explorar potenciais relações de causalidade entre eles. As variáveis analisadas, obtidas por meio da Base de Dados do Estado de Pernambuco (BDE), incluem:

- IDEB (Índice de Desenvolvimento da Educação Básica – Ensino Médio);
- Taxa de criminalidade violenta por 100 mil habitantes;
- PIB per capita;
- IDHM-E (Índice de Desenvolvimento Humano Municipal – Educação);
- Expectativa de vida ao nascer (em anos).

Os dados abrangem os 185 municípios pernambucanos e permitem avaliar o panorama educacional, econômico e social da região. Diante desse cenário, esta pesquisa busca preencher lacunas na literatura ao analisar, de forma integrada, como esses indicadores se inter-relacionam e em que medida podem exercer influência causal entre si. O objetivo é compreender a dinâmica dessas relações e, a partir disso, contribuir para o debate sobre políticas públicas voltadas à melhoria das condições de vida no estado.

2. Estudos relacionados

Moura e Monteiro (2024) investigaram os efeitos da educação e da desigualdade econômica sobre os crimes contra o patrimônio nos municípios de Pernambuco. Utilizaram dados do Atlas de Desenvolvimento Humano do Brasil (2000–2010) e dos Anuários de Criminalidade de Pernambuco (2007–2012), aplicando regressão linear múltipla pelo método MQO. A variável dependente foi a taxa de Crimes Violentos contra o Patrimônio (CVP), enquanto as independentes foram o Índice de Gini e o percentual de pessoas com ensino médio completo. Os resultados mostraram que a desigualdade aumenta a criminalidade, enquanto maiores níveis de escolaridade reduzem as taxas de CVP. Os autores defendem políticas públicas que aliem expansão educacional e redução da desigualdade para diminuir a violência urbana.

Singh et al. (2025) analisaram os determinantes do Índice de Desenvolvimento Humano (IDH) a partir de indicadores econômicos e sociais em 94 países. O estudo aplicou modelos de regressão linear múltipla, com seleção de variáveis por regressão *stepwise* e comparação de ajustes. As variáveis finais foram PIB per capita, anos médios de escolaridade e mortalidade infantil, explicando 95% da variação do IDH. Os resultados indicaram associação positiva entre desenvolvimento humano, crescimento econômico e escolaridade, além de relação negativa com a mortalidade infantil. Concluiu-se que políticas públicas integradas, voltadas à educação, à saúde e à economia, são fundamentais para elevar o IDH.

Choi et al. (2022) investigaram a mortalidade infantil evitável na Coreia do Sul (2010–2017), por meio de delineamento transversal com regressão logística multinível. Identificaram como principais determinantes a baixa escolaridade materna, prematuridade, baixo peso ao nascer e acesso limitado ao pré-natal, além de maior risco em mães adolescentes e famílias de baixa renda. Concluem que o fortalecimento da atenção pré-natal e a redução das desigualdades sociais são medidas prioritárias para prevenir óbitos infantis no país.

Dietrich et al. (2001) conduziram o primeiro estudo longitudinal prospectivo sobre a relação entre exposição precoce ao chumbo (Pb) e comportamentos delinquentes na adolescência. A pesquisa utilizou a coorte do Cincinnati *Lead Study* (195 adolescentes), com acompanhamento pré e pós-natal e aplicação de questionários de autorrelato e relato parental, controlando variáveis socioeconômicas e ambientais. Por meio de regressão linear múltipla e análise de covariância, verificou-se que a exposição ao Pb esteve significativamente associada a maior frequência de comportamentos antissociais, reforçando seu papel na gênese de problemas comportamentais e destacando implicações para a saúde pública.

Carvalho e Chima (2020) analisaram a relação entre características do acolhimento institucional e a delinquência juvenil em 221 adolescentes afro-americanas nos EUA. Utilizando regressão linear múltipla, verificaram que o tipo de colocação, o tempo de permanência e o número de remoções estiveram significativamente associados à incidência de atos infracionais. Os resultados indicaram que instabilidade no

acolhimento e permanência prolongada aumentam o risco de envolvimento em delitos, reforçando a importância da estabilidade para reduzir a vulnerabilidade juvenil.

Silva et al. (2018) analisaram os determinantes sociais da mortalidade infantil no Brasil entre 1990 e 2015, com dados do SIM e SINASC. Aplicaram análise descritiva e regressão de Poisson para relacionar fatores socioeconômicos às taxas. Encontraram queda expressiva da mortalidade, mas desigualdades persistem, sobretudo no Norte e Nordeste, associadas à renda, escolaridade materna e acesso à saúde. Concluem que políticas de equidade e fortalecimento do sistema de saúde são necessárias.

Dira e Subardjo (2025) investigaram a relação entre relacionamentos com pares e delinquência juvenil em 96 adolescentes de 15 a 21 anos. A pesquisa, de caráter quantitativo e transversal, utilizou escalas Likert para medir a qualidade das interações e diferentes formas de comportamento delinquente. Por meio de correlação de Pearson e regressão linear simples, verificou-se associação negativa significativa, indicando que relações de maior qualidade entre pares estão ligadas a menores níveis de delinquência. Os autores destacam a relevância de fortalecer vínculos sociais positivos em contextos escolares e comunitários como estratégia de prevenção.

Aizer e Currie (2019) analisaram a relação entre exposição ao chumbo na infância e comportamentos antissociais em aproximadamente 125 mil crianças nascidas em Rhode Island entre 1990 e 2004. O estudo combinou dados de exames de sangue, registros escolares e de detenções juvenis, aplicando regressão OLS, modelos de efeitos fixos entre irmãos e variáveis instrumentais. Os resultados mostraram que maiores níveis de chumbo aumentaram em cerca de 6% a probabilidade de suspensão escolar e em até 57% a de detenção juvenil para meninos, sem efeitos significativos para meninas. Os achados indicam que a exposição ao chumbo tem efeitos causais relevantes sobre problemas disciplinares e delinquência juvenil, contribuindo para explicar quedas nas taxas de criminalidade em décadas recentes.

Os estudos apresentados mostram como fatores sociais, educacionais, econômicos e de saúde se conectam e ajudam a formar contextos de vulnerabilidade ou de desenvolvimento. Trabalhos como os de Moura e Monteiro (2024), Silva et al. (2018) e Choi et al. (2022) evidenciam o peso da desigualdade, da escolaridade e das condições de saúde em diferentes fenômenos sociais, enquanto Singh et al. (2025) reforça a importância de indicadores educacionais e econômicos para o desenvolvimento humano. Já pesquisas voltadas à delinquência juvenil, como as de Dietrich et al. (2001), Aizer e Currie (2019), Carvalho e Chima (2020) e Dirá e Subardjo (2025), mostram como fatores ambientais, familiares e sociais influenciam diretamente o comportamento dos jovens.

Mesmo com todos esses avanços, ainda faltam estudos que, no contexto de Pernambuco, consigam olhar para diferentes indicadores sociais de forma integrada e que considerem possíveis relações causais entre eles. É justamente nesse ponto que o presente estudo se insere, ao usar modelos de regressão linear múltipla para investigar as conexões entre educação, criminalidade, mortalidade infantil e desigualdade. A ideia não é apenas identificar associações estatísticas, mas também levantar hipóteses sobre mecanismos causais que possam explicar como esses fenômenos se influenciam mutuamente. Com isso, espera-se gerar evidências que ajudem a construir políticas públicas mais eficazes e voltadas para melhorar a qualidade de vida e o desenvolvimento humano no estado.

3. Metodologia

A questão dos efeitos da desigualdade sobre a saúde, educação e criminalidade infantil é um tema de grande relevância para a formulação de políticas públicas que busquem mitigar as desigualdades sociais e promover o bem-estar da população. Em especial, a análise de indicadores sociais, como a mortalidade infantil, os índices de criminalidade e o desempenho educacional, oferece uma visão ampla das condições de vida nas diversas regiões de um país. Nesse sentido, estudos econômicos apontam que a desigualdade social está diretamente associada a maiores índices de criminalidade, inclusive entre jovens, o que reforça a necessidade de políticas integradas que promovam equidade e segurança (SANTOS; KASSOUF, 2008).

Este estudo busca explorar como variáveis sociais, como o Índice de Desenvolvimento da Educação Básica (Ideb), taxas de criminalidade, mortalidade infantil e analfabetismo juvenil, se inter-relacionam, utilizando dados disponíveis de Pernambuco para a aplicação de uma análise de regressão linear múltipla. A abordagem quantitativa permite examinar como mudanças em uma variável podem impactar outras, fornecendo insights importantes para os formuladores de políticas públicas.

3.1. Primeiro modelo de Regressão Linear Múltiplo

A pesquisa utiliza Regressão Linear Múltipla Simultânea, aplicando diversos modelos alternando a variável resposta. O objetivo é avaliar como variáveis relacionadas ao desenvolvimento educacional e às questões sociais influenciam o Índice de Desenvolvimento da Educação Básica (Ideb) nos municípios de Pernambuco. O modelo pode ser estruturado da seguinte forma:

- Variável independente(Y):
Índice de Desenvolvimento do Ensino Médio (Ideb): principal indicador que será analisado e modelado.
- Variáveis dependentes(X):
 - X1 = Número de crimes violentos a cada 100 mil habitantes: pode refletir um impacto negativo no desenvolvimento educacional;
 - X2 = Mortalidade infantil: esse indicador é frequentemente associado à qualidade do sistema de saúde e da infraestrutura;
 - X3 = Pessoas de 15 a 19 anos não alfabetizadas: pode indicar a qualidade da educação básica e sua continuidade até a adolescência;
 - X4 = Estabelecimentos de educação fundamental: reflete a quantidade e a qualidade das infraestruturas educacionais disponíveis nos municípios.

Modelo matemático:

$$Y_i = \beta_0 + \beta_1 \times X1_i + \beta_2 \times X2_i + \beta_3 \times X3_i + \beta_4 \times X4_i + \epsilon_i$$

Onde:

- X1,...,X5 são as variáveis de entrada;
- β_0 é o intercepto (constante) do modelo;

- $\beta_1, \beta_2, \beta_3, \beta_4$ são os coeficientes que representam a relação entre cada variável independente e a variável dependente (Ideb).
- ϵ_i é o termo de erro para o município i .

3.2. Modelos 2 a 5 – Outras variáveis dependentes

Para os modelos subsequentes, a abordagem é análoga, alternando a variável dependente:

- Modelo 2: Não alfabetizados de 15 a 19 anos como variável dependente, considerando Ideb, criminalidade, mortalidade infantil e número de escolas como variáveis independentes. Espera-se que Ideb e número de escolas estejam negativamente associados à não alfabetização, enquanto criminalidade e mortalidade infantil apresentem relação positiva.
- Modelo 3: Criminalidade violenta como variável dependente, com variáveis educacionais e de saúde como independentes. Espera-se efeito positivo do analfabetismo e mortalidade infantil, e efeito negativo do Ideb e do número de escolas.
- Modelo 4: Mortalidade infantil como variável dependente, considerando analfabetismo, criminalidade, Ideb e escolas. Espera-se efeito positivo do analfabetismo e da criminalidade, e efeito negativo de Ideb e do número de escolas.
- Modelo 5: Número de estabelecimentos de educação fundamental como variável dependente, com as demais variáveis como independentes, analisando como fatores sociais e educacionais influenciam a disponibilidade de escolas.

Todos os modelos seguem a forma geral do modelo 1.

A aplicação dessa abordagem permite identificar relações entre educação, saúde e segurança, fornecendo subsídios importantes para políticas públicas voltadas ao desenvolvimento social e educacional nos municípios pernambucanos.

3.3. Coleta e processamento dos dados

Os dados utilizados nesta análise foram extraídos da Base de Dados do Estado de Pernambuco (BDE), que reúne informações de diversas instituições governamentais e de pesquisa. Entre os órgãos consultados estão: Secretaria de Defesa Social (estatísticas de segurança pública e criminalidade), CONDEPE/FIDEM (indicadores socioeconômicos e urbanísticos), INEP (dados educacionais), IBGE (informações geográficas, demográficas e socioeconômicas) e Secretaria de Educação e Esportes de Pernambuco (estatísticas educacionais e esportivas).

A integração dessas fontes possibilita uma visão detalhada da realidade pernambucana, permitindo análises que apoiam a formulação de políticas públicas e decisões estratégicas.

Para tratamento e análise dos dados foi utilizado o software R, uma linguagem de programação estatística amplamente empregada para análise de dados, modelagem estatística e visualização. O *dataset* unificado contém as cinco variáveis estudadas, organizadas em colunas, provenientes de diferentes tabelas e consolidado em uma única planilha do Excel. O uso do R permite desde estatísticas descritivas básicas até modelos complexos de

regressão, análise multivariada e séries temporais, com integração a outras linguagens e ferramentas, como Python, SQL e plataformas de *Big Data*.

3.4. Estatística descritiva das variáveis

A estatística descritiva foi utilizada para organizar, resumir e apresentar os dados de forma informativa, permitindo a identificação de padrões, dispersão, assimetrias e possíveis outliers antes da aplicação de métodos inferenciais (Triola, 2020; Bussab & Morettin, 2017).

Foram analisadas as seguintes variáveis nos municípios de Pernambuco:

- Índice de Desenvolvimento da Educação Básica (Ideb): apresentou valores concentrados entre 4 e 5, com baixa dispersão (DP = 0,43), assimetria positiva e presença de outliers (valores 0 e 6).
- Criminalidade Violenta (por 100 mil habitantes): distribuição assimétrica à direita, alta dispersão (DP = 31,5) e presença de valores extremos (0 a 577), indicando que a maioria dos municípios apresenta índices baixos, mas alguns possuem índices muito elevados.
- Mortalidade Infantil (por 1.000 nascidos vivos): média e mediana próximas (6,1), com alta dispersão (DP = 10,5) e valores extremos (0 a 469), sugerindo desigualdade no acesso a serviços de saúde.
- Jovens de 15 a 19 anos não alfabetizados: apresentou distribuição muito dispersa (DP = 56,5), média superior à mediana (66,5 vs. 37) e presença de outliers (0 a 1.297), indicando desigualdade na alfabetização.
- Número de Estabelecimentos de Educação Fundamental: média de 25,5 e mediana de 16, com alta dispersão (DP = 15,5) e valores extremos (4 a 221), evidenciando desigualdade no número de estabelecimentos entre os municípios.

A Tabela 1 resume as principais medidas descritivas das variáveis analisadas.

Tabela 1: Estatísticas Descritivas

Variável	Média	Mediana	Desvio Padrão	Mínimo	Máximo
IDEB	4,54	4,5	0,43	0	6
Criminalidade Violenta (por 100 mil)	20,5	5	31,5	0	577
Mortalidade Infantil (por 1.000)	6,1	6,1	10,5	0	469
Jovens 15–19 anos não alfabetizados	66,5	37	56,5	0	1297
Estabelecimentos de Educação Fundamental	25,5	16	15,5	4	221

Fonte: Autor (2024)

A análise descritiva evidencia que a maioria das variáveis apresenta alta dispersão e assimetria, indicando desigualdade significativa entre os municípios de Pernambuco. No Ideb, a concentração de valores entre 4 e 5 sugere desempenho escolar relativamente uniforme, embora haja poucos casos excepcionais. Já a criminalidade violenta e a mortalidade infantil apresentam distribuição assimétrica à direita, com alguns municípios exibindo índices muito elevados em comparação à maioria, apontando problemas localizados de segurança e saúde pública. A variável de jovens não alfabetizados mostra grande diferença entre média e mediana, indicando desigualdade

no nível de alfabetização. Por fim, o número de estabelecimentos de educação fundamental também apresenta elevada variação, sugerindo disparidades na infraestrutura educacional entre os municípios. Esses padrões devem ser considerados nas análises inferenciais subsequentes, pois outliers e assimetrias podem influenciar os resultados dos modelos estatísticos.

4. Resultados e discussões

Na primeira etapa da análise, foi realizada a matriz de correlação com o objetivo de identificar relações lineares entre as variáveis e possíveis problemas de multicolinearidade. Como critério, considerou-se que correlações fortes com a variável dependente (acima de $|0,5|$) indicam bom potencial preditivo, enquanto correlações elevadas entre variáveis independentes (acima de $|0,8|$) poderiam comprometer o modelo.

A Tabela 2 apresenta a matriz de correlação entre as variáveis estudadas:

Tabela 2: Matriz de Correlação

Variáveis	Ideb	CVLI	PIB	IDHM	Esperança de Vida
Ideb	1.000	-0.062	-0.037	0.189	0.017
CVLI	-0.062	1.000	0.051	0.092	0.116
PIB	-0.037	0.051	1.000	0.273	0.162
IDHM	0.189	0.092	0.273	1.000	0.426
Esperança Vida	0.017	0.116	0.162	0.426	1.000

Fonte: Autor (2024)

Os resultados indicam que não há correlações suficientemente altas para configurar multicolinearidade, assegurando a adequação dos dados ao modelo. Ressalta-se que valores nulos foram desconsiderados em razão da ausência de informações.

Na segunda etapa, foram avaliados os pressupostos dos modelos de regressão linear, a saber:

1. Relação linear entre variáveis independentes e dependente;
2. Independência dos erros, verificada pelo teste de Durbin-Watson;
3. Homoscedasticidade dos resíduos, avaliada pelo teste de Breusch-Pagan;
4. Normalidade dos resíduos, examinada pelo teste de Shapiro-Wilk
5. Ausência de multicolinearidade entre variáveis independentes.

4.1. Modelos de regressão linear

Foram estimados cinco modelos de regressão linear, cada um considerando diferentes variáveis dependentes e avaliando o poder explicativo dos indicadores socioeconômicos e educacionais nos municípios de Pernambuco. Além dos coeficientes estimados, foram aplicados testes de diagnóstico (Shapiro-Wilk, Durbin-Watson e Breusch-Pagan) para verificação dos pressupostos clássicos. Em casos de violação da normalidade ou

homoscedasticidade, optou-se pelo uso de regressão robusta e erros padrão robustos, garantindo maior confiabilidade inferencial.

4.1.1 Modelo 1 – IDEB como variável dependente

O Índice de Desenvolvimento da Educação Básica (IDEB) foi adotado como variável dependente (y), enquanto as variáveis independentes consideradas foram: taxa de CVLI (x1), PIB per capita (x2), IDHM-E (x3) e esperança de vida (x4).

Primeiramente, buscamos verificar a relação linear entre a variável resposta (IDEB) e as variáveis explicativas, por meio da matriz de correlação.

Da matriz, temos que:

- $cor(y; x1) = -0,062 < 0,3$ (correlação fraca e negativa, praticamente inexistente);
- $cor(y; x2) = -0,037 < 0,3$ (correlação fraca e negativa, praticamente nula);
- $cor(y; x3) = 0,189 < 0,3$ (correlação fraca e positiva, regiões com maior desenvolvimento humano educacional tendem a ter Ideb ligeiramente superior);
- $cor(y; x4) = 0,017 < 0,3$ (correlação nula, sem relação linear aparente).

Em teoria, nenhum regressando é um bom preditor linear. Modelos propostos não serviram para mostrar que é possível explicar adequadamente a variável resposta utilizando as variáveis independentes disponíveis através de um modelo de regressão linear múltiplo convencional. O modelo utilizado foi o de Regressão Linear Robusta.

Tabela 3: Resultados do modelo (IDEB ~ IDHM-E)

Variável	Coefficiente	Erro Padrão	t-valor
Intercepto	4.1438	0.1972	21.0091
IDHM-E	0.8099	0.3988	2.0312

Erro padrão residual: 0,3279 | GL residuais: 169

Pseudo-R²: 0,137 (estimado via comparação com OLS)

Fonte: Autor (2024)

Explicação moderada (comum em ciências sociais). A cada ponto a mais no IDHM da educação, o Ideb aumenta em média 0,80 pontos. No caso da regressão robusta, o modelo já é construído para lidar com heterocedasticidade e outliers.

Para avaliar a adequação do modelo, foram realizados testes de pressupostos de normalidade, autocorrelação e homoscedasticidade. Os resultados estão apresentados na Tabela 4.

Tabela 4: Testes de pressupostos do modelo robusto

Teste	Estatística
Shapiro-Wilk	0,006
Durbin-Watson (D-W)	1,811
Breusch-Pagan	0,767

Fonte: Autor (2024)

Os achados confirmam que o IDHM-E foi o único preditor estatisticamente significativo ($\beta = 0,81$; $p < 0,05$). Isso indica que, a cada ponto adicional no IDHM-E, o IDEB aumenta em média 0,80 pontos. O Pseudo- R^2 do modelo (0,137) demonstra uma capacidade explicativa moderada, típica de estudos em ciências sociais.

Embora a dispersão dos resíduos tenha se mostrado relativamente elevada, não foram observados padrões sistemáticos ou curvilíneos, o que sustenta a adequação do ajuste. Assim, a regressão robusta mostrou-se apropriada para lidar com as violações de normalidade, fornecendo estimativas estatisticamente confiáveis e interpretáveis.

4.1.2 Modelo 2: Regressão Linear para a Taxa de Criminalidade Violenta (CVLI)

Neste modelo, a variável dependente (y) foi a Taxa de Criminalidade Violenta por 100 mil habitantes (CVLI), e as variáveis independentes consideradas foram: IDEB (x1), PIB per capita (x2), IDHM-E (x3) e Esperança de Vida (x4).

A análise preliminar de correlação mostrou valores baixos entre as variáveis explicativas e a variável resposta:

- $\text{cor}(y;x1) = -0,62$
- $\text{cor}(y;x2) = 0,051$
- $\text{cor}(y;x3) = 0,092$
- $\text{cor}(y;x4) = 0,116$

Nenhuma das variáveis apresentou forte correlação com CVLI, indicando que um modelo linear pode não ser apropriado para explicar a variável resposta. Dentre as variáveis disponíveis, a esperança de vida apresentou a maior correlação, sendo utilizada no modelo inicial de regressão linear simples.

Tabela 5: Resultados da Regressão Linear (CVLI ~ Esperança de Vida)

Variável	Estimativa	Erro Padrão	t-valor	p-valor
Intercepto	-48,543	55,812	-0,870	0,386
Esperança de Vida	1,211	0,794	1,526	0,129

Erro padrão residual: 21,71 | R^2 ajustado: 0,0078 | $F(1,169) = 2,33$

Fonte: Autor (2024)

O modelo indica que não há relação estatisticamente significativa entre Esperança de Vida e CVLI. A cada aumento de 1 ano na Esperança de Vida, a CVLI aumenta em média 1,186, mas o p-valor = 0,129 demonstra que essa associação não é significativa ao nível de 5%. O modelo explica apenas 1,3% da variação na CVLI, evidenciando seu baixo poder explicativo.

Para avaliação dos pressupostos, foram aplicados testes de normalidade, independência e homoscedasticidade dos resíduos.

Tabela 6: Testes de diagnóstico do Modelo 2

Teste	Estatística	p-valor
Shapiro-Wilk (normalidade)	$W = 0,9335$	$4,13 \times 10^{-7}$
Durbin-Watson (independência)	$D-W = 1,930$	0,578

Teste	Estatística	p-valor
Breusch-Pagan (homoscedasticidade)	BP = 0,0010	0,975

Fonte: Autor (2024)

O pressuposto de normalidade dos resíduos foi violado, enquanto os demais pressupostos (independência e homoscedasticidade) foram atendidos. De modo geral, o modelo não apresentou resultados satisfatórios, reforçando a ideia de que as variáveis disponíveis não fornecem explicação adequada para a CVLI.

4.1.3 Modelo 3: Regressão Linear para PIB per capita

Neste modelo, a variável dependente (y) foi o PIB per capita dos municípios de Pernambuco, e as variáveis independentes consideradas foram: IDEB (x1), CVLI (x2), IDHM-E (x3) e Esperança de Vida (x4).

A análise preliminar de correlação indicou valores baixos entre as variáveis explicativas e o PIB:

- $\text{cor}(y; x1) = -0,037$
- $\text{cor}(y; x2) = 0,051$
- $\text{cor}(y; x3) = 0,237$
- $\text{cor}(y; x4) = 0,162$

Dentre os preditores disponíveis, o IDHM-E apresentou a maior correlação com o PIB, justificando sua inclusão no modelo de regressão linear robusta.

Tabela 7: Resultados do Modelo de Regressão Linear Robusta (PIB ~ IDHM-E)

Variável	Coefficiente	Erro Padrão	t-valor	Significância
Intercepto	-4.464.367,20	2.084.265,60	-2,14	*
IDHM-E	34.852.438,60	4.213.789,40	8,27	***

Erro padrão residual: 3.622 | Graus de liberdade: 169 | Pseudo R²: 0,948

Fonte: Autor (2024)

O coeficiente do IDHM-E indica que, a cada aumento de uma unidade no IDHM-E, o PIB per capita aumenta, em média, R\$ 34.852,44, mantendo os demais fatores constantes. O valor t = 8,27 e a alta significância mostram uma forte associação positiva, confirmando que o IDHM-E é um forte preditor do nível de riqueza local. Para avaliação dos pressupostos, foram aplicados testes de normalidade, independência e homoscedasticidade dos resíduos, conforme apresentados na Tabela 8.

Tabela 8: Testes de pressupostos do Modelo 3

Pressuposto	Teste	Estatística	Valor-p	Conclusão
Normalidade dos resíduos	Shapiro-Wilk	W = 0,34769	$< 2,2 \times 10^{-16}$	Rejeita
Independência dos resíduos	Durbin-Watson	DW = 2,0294	0,986	Não rejeita
Homoscedasticidade	Breusch-Pagan	BP = 1,809	0,1786	Não rejeita

Fonte: Autor (2024)

Apesar da violação da normalidade dos resíduos, a regressão robusta permite relaxar os pressupostos de normalidade e homoscedasticidade, mantendo a validade da interpretação. Os pressupostos de linearidade,

independência dos erros e ausência de multicolinearidade severa foram respeitados, garantindo confiabilidade estatística do modelo.

4.1.4 Modelo 4: Regressão Linear para IDHM-E

Neste modelo, a variável dependente (y) foi o IDHM-E, que representa o desenvolvimento humano municipal por escolaridade. As variáveis independentes consideradas foram: IDEB (x1), CVLI (x2), PIB per capita (x3) e Esperança de Vida (x4).

Após estimativa do modelo linear múltiplo, foi identificado problema de heterocedasticidade nos resíduos (Breusch-Pagan BP = 21,788, $p < 0,001$), o que não compromete os coeficientes estimados, mas afeta os erros padrão e a inferência estatística. Para contornar isso, foram utilizados erros padrão robustos (White, HC0), garantindo inferência confiável sem alterar os coeficientes.

Tabela 9: Modelo de Regressão Linear Múltipla para IDHM-E

Variável	Coefficiente	Erro Padrão	t-valor	Valor-p
Intercepto	-0.4202	0.1515	-2.773	0.0061
IDEB	0.0162	0.0074	2.189	0.0299
CVLI	0.0001	0.0002	0.342	0.7327
PIB	0.000001	0.0000003	3.775	0.0002
Esperança de Vida	0.0117	0.0021	5.475	1.46e-07

$R^2 = 0,2527$ | R^2 Ajustado = 0,2361 | Erro Padrão Residual = 0,0594 | N = 185 | F = 15,22 ($p = 9.728e-11$)

Fonte: Autor (2024)

Observou-se que a variável CVLI não apresentou significância estatística, sugerindo a possibilidade de remoção para simplificar o modelo.

Tabela 10: Modelo de Regressão Linear Múltipla para IDHM-E (sem CVLI)

Variável	Coefficiente	Erro Padrão	t-valor	Valor-p
Intercepto	-0.4202	0.1512	-2.780	0.0060
IDEB	0.0163	0.0074	2.200	0.0290
PIB	0.000001027	0.0000002704	3.800	0.0002
Esperança de Vida	0.0117	0.0021	5.506	1.24e-07

$R^2 = 0,2523$ | R^2 Ajustado = 0,2399 | Erro Padrão Residual = 0,05926 | F = 20,35 ($p = 2.066e-11$)

Fonte: Autor (2024)

A remoção da variável CVLI não alterou significativamente o poder explicativo do modelo, que continua explicando cerca de 25% da variação do IDHM-E. Um aumento de 1 ponto no Ideb significa um aumento de 0,0163 no IDHM-E e assim vale interpretar para cada coeficiente de cada variável independente.

Para correção da heterocedasticidade, aplicou-se a regressão com erros robustos:

Tabela 11: Modelo de Regressão Linear com Erros Robustos (HC0)

Variável	Coefficiente	Erro Robusto	t-valor	Valor-p
Intercepto	-0.4202	0.1587	-2.648	0.0088

Variável	Coefficiente	Erro Robusto	t-valor	Valor-p
IDEB	0.0163	0.0078	2.087	0.0385
PIB	0.000001027	0.000000290	3.540	0.0005
Esperança de Vida	0.0117	0.0022	5.327	2.61e-07

Fonte: Autor (2024)

Tabela 12: Testes de Pressupostos do Modelo 4

Teste	Estatística	Valor-p	Conclusão
Shapiro-Wilk (normalidade)	W = 0,99578	0,888	Normalidade aceita
Durbin-Watson (independência)	DW = 2,2363	0,136	Independência aceita
Breusch-Pagan (heterocedasticidade)	BP = 21,788	7.22e-05	Heterocedasticidade

Fonte: Autor (2024)

O IDHM-E foi explicado pelas variáveis IDEB, PIB per capita, Esperança de Vida e CVLI. A CVLI não apresentou significância estatística e foi removida do modelo final. O modelo indicou que IDEB, PIB e Esperança de Vida possuem efeito positivo sobre o IDHM-E, ou seja, aumentos nessas variáveis estão associados a elevação do índice de desenvolvimento educacional.

Apesar da heterocedasticidade detectada, o uso de erros padrão robustos garantiu a validade da inferência. O modelo explicou cerca de 25% da variação do IDHM-E, mostrando que fatores educacionais, econômicos e de expectativa de vida contribuem significativamente para o desenvolvimento humano municipal, embora parte da variação ainda seja determinada por fatores não incluídos no modelo.

4.1.5 Modelo 5 – Esperança de Vida como variável dependente

Neste modelo, a expectativa de vida foi a variável dependente (y). As variáveis independentes levadas em conta foram: CVLI (x1), PIB per capita (x2), IDHME (x3) e Ideb (x4).

A análise preliminar de correlação indicou os seguintes valores entre as variáveis explicativas e a Esperança de Vida:

- $\text{cor}(y; x1) = 0,116$
- $\text{cor}(y; x2) = 0,162$
- $\text{cor}(y; x3) = 0,426$
- $\text{cor}(y; x4) = 0,017$

Dentre os preditores disponíveis, o IDHM-E apresentou a maior correlação com a Esperança de Vida, justificando sua inclusão no modelo de regressão linear.

Tabela 13: Resultados do Modelo (Esperança de Vida ~ IDHM-E)

Variável	Coefficiente	Erro Padrão	t-valor	Significância
Intercepto	63,646	1,094	58,17	$< 2 \times 10^{-16}$
IDHM-E	13,544	2,212	6,12	$6,23 \times 10^{-9}$

Erro padrão residual: 1,903 | R² ajustado: 0,1767 | F(1,169) = 37,49, p < 0,0001

Fonte: Autor (2024)

O coeficiente do IDHM-E (13,544) indica que, a cada aumento de uma unidade no IDHM-E, a Esperança de Vida aumenta, em média, 13,544 anos, mantendo os demais fatores constantes. O intercepto (63,646) representa a expectativa de vida média quando o IDHM-E é zero, servindo apenas como ponto de referência. Apesar de significativa, a variável explica apenas 17,6% da variação na Esperança de Vida, o que era esperado, dado que este indicador depende de múltiplos fatores sociais, econômicos e de saúde.

Tabela 14: Testes de pressupostos do Modelo 5

Teste	Estatística	Valor-p	Conclusão
Shapiro-Wilk	W = 0,99033	0,3809	Não rejeita
Durbin-Watson	DW = 2,0022	0,80	Não rejeita
Breusch-Pagan	BP = 6,0024	0,0143	Rejeita (heterocedasticidade)

Fonte: Autor (2024)

Devido à presença de heterocedasticidade, foram aplicados erros padrão robustos (White), permitindo testes de significância e intervalos de confiança mais confiáveis.

Tabela 15: Regressão Linear Robusta (White)

Variável	Coefficiente	Erro Padrão Robusto	t-valor	Valor-p
Intercepto	64,022	1,045	61,27	$< 2 \times 10^{-16}$
IDHM-E	12,867	2,180	5,90	$1,21 \times 10^{-8}$

R² ajustado: 0,1686 | F robusto = 34,81, p < 0,001

Fonte: Autor (2024)

Os resíduos mostraram distribuição aproximadamente normal e leve heterocedasticidade nos extremos, sem comprometer a interpretação dos coeficientes.

5. Conclusões

O estudo teve como objetivo avaliar, por meio de modelos de regressão linear tradicionais, robustos e com erros padronizados robustos, as relações estatísticas entre indicadores educacionais, sociais e econômicos nos municípios de Pernambuco, considerando IDEB, IDHM-E, Esperança de Vida, PIB per capita e taxa de criminalidade violenta (CVLI). Os resultados mostraram que o modelo que utilizou o IDHM-E como variável explicativa do PIB per capita apresentou o melhor desempenho, com R² de 0,948 e coeficientes altamente significativos, indicando que o desenvolvimento educacional municipal exerce forte influência sobre os níveis de renda local.

Quando o IDHM-E foi modelado como variável dependente a partir de IDEB, PIB e Esperança de Vida, os coeficientes foram significativos, com exceção da CVLI, e o R² ajustado foi de 0,25, sugerindo que o índice de desenvolvimento humano educacional pode ser parcialmente previsto por fatores educacionais e socioeconômicos. A regressão entre IDHM-E e Esperança de Vida revelou uma associação positiva significativa, embora com poder explicativo limitado (R² = 0,17), enquanto os modelos voltados para explicar a taxa de

criminalidade violenta apresentaram baixo poder explicativo e ausência de significância estatística, indicando a complexidade da violência urbana.

Os resultados teóricos reforçam que o IDHM-E atua tanto como variável explicativa quanto como variável resposta, evidenciando a dinâmica circular entre educação e desenvolvimento humano. Na prática, os achados destacam a importância de políticas públicas voltadas à educação básica como instrumento de transformação social e econômica, embora o combate à violência urbana requeira abordagens mais complexas.

Entre as limitações do estudo estão a violação da normalidade dos resíduos, R^2 relativamente baixos em alguns modelos, ausência de variáveis culturais ou políticas locais e limitações na cobertura e completude da base de dados. Recomenda-se, para trabalhos futuros, incluir variáveis adicionais, considerar modelos hierárquicos, séries temporais ou técnicas de aprendizado de máquina para capturar relações não lineares e complexas.

De modo geral, o estudo evidencia o papel central do IDHM-E como elo entre educação, renda e longevidade, reforçando seu caráter transversal no desenvolvimento socioeconômico e a necessidade de priorização em políticas públicas e pesquisas científicas.

Referências

1. Agência Estadual de Planejamento e Pesquisas de Pernambuco – CONDEPE/FIDEM, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Instituto Brasileiro de Geografia e Estatística, & Secretaria de Educação e Esportes de Pernambuco. (2023). Base de Dados do Estado de Pernambuco (BDE). Recife, PE.
2. Aizer, A., & Currie, J. (2019). Lead and juvenile delinquency: New evidence from linked birth, school and juvenile detention records. *Review of Economics and Statistics*, 101(4), 575–587. https://doi.org/10.1162/rest_a_00814
3. Bastos, F. (2016). As desigualdades sociais e a infância. In G. G. Rodrigues & L. C. Copatti (Orgs.), *Direitos humanos de crianças e adolescentes e políticas públicas* (pp. 27–45). São Paulo: Contexto.
4. Bussab, W. O., & Morettin, P. A. (2017). *Estatística básica* (9. ed.). São Paulo: Saraiva Educação.
5. CAEDJUS. (2021). *Políticas públicas em perspectiva* (v. 2). Belo Horizonte: CAEDJUS. Disponível em: [CONIPUB-2021-02-Politicas-publicas-em-perspectiva-vol-2.pdf](https://conipub-2021-02-politicas-publicas-em-perspectiva-vol-2.pdf)
6. Carvalho, J., & Chima, F. O. (2020). A regression analysis of juvenile delinquency among African American females in foster care. *Journal of Liberal Arts and Humanities*, 1(5), 1–15.
7. Choi, J. (2022). Identifying important factors to prevent juvenile delinquency among male and female adolescents: An exploratory analysis using the LASSO regression algorithm in the Korean Children and Youth Panel Survey (KCYPs). *Child Indicators Research*, 15, 1429–1464. <https://doi.org/10.1007/s12187-022-09916-6>
8. De Moura, J. A., & Monteiro, M. B. (2024). From education to social justice: A regression examination of education and economic inequality effects on property crimes. *Socioeconomic Analytics*, 2(1), 94–106. <https://doi.org/10.51359/2965-4661.2024.262687>
9. Dietrich, K. N., et al. (2001). Early exposure to lead and juvenile delinquency. *Neurotoxicology and Teratology*, 23(6), 511–518.
10. Dira, H., & Subardjo, R. Y. S. (2025). Peer relationships and their impact on juvenile delinquency intensity: A cross-sectional study. *Journal of Health Sciences and Medical Development*, 4(2), 116–126. <https://doi.org/10.56741/hesmed.v4i02.973>
11. Instituto de Pesquisa Econômica Aplicada – IPEA. (2016). *Atlas do Desenvolvimento Humano no Brasil*. Brasília: IPEA.
12. Lima, R. S., Oliveira, L. M., & Almeida, J. P. (2024). Police patrols and the spatial distribution of homicides: An econometric analysis of simultaneous causality. *Urban Science*, 8(1), 132. <https://doi.org/10.3390/urbansci8010132>
13. Ribeiro, C. A. C., & Schlegel, R. (2015). A relação entre o Índice de Desenvolvimento Humano Municipal e o desempenho educacional no Brasil. *Revista Brasileira de Educação*, 20(61), 745–772.

14. Santos, M. J., & Kassouf, A. L. (2008). Estudos econômicos das causas da criminalidade no Brasil: Evidências e controvérsias. *EconomiA*, 9(2), 343–372.
15. Singh, K., Cheemalapati, S., Ramireddy, S. R., Kurian, G., Muzumdar, P., & Muley, A. (2025). Determinants of Human Development Index (HDI): A regression analysis of economic and social indicators. *Asian Journal of Economics, Business and Accounting*, 25(1), 26–34. <https://doi.org/10.9734/ajeba/2025/v25i11630>
16. Triola, M. F. (2020). *Introdução à estatística* (13. ed.). São Paulo: Pearson.
17. Vidal e Silva, S. M. C., Tuon, R. A., Probst, L. F., Gondinho, B. V. C., Pereira, A. C., Meneghim, M. C., & Cortellazzi, K. L., Ambrosano, G. M. B. (2018). Factors associated with preventable infant death: A multiple logistic regression. *Revista de Saúde Pública*, 52, 32. <https://doi.org/10.11606/S1518-8787.2018052000252>