



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ECONOMIA – DECON – SEDE
BACHARELADO EM CIÊNCIAS ECONÔMICAS

FLÁVIO ADRIANO JESUS DE ARAÚJO

INFORMALIDADE E ESCOLARIDADE NO NORDESTE BRASILEIRO

RECIFE

2026



FLÁVIO ADRIANO JESUS DE ARAÚJO

Trabalho de Conclusão de Curso apresentado à
Universidade Federal Rural de Pernambuco
(UFRPE) como requisito para obtenção do título
de Bacharel em Ciências Econômicas.

Orientadora: Prof. Dra. Sônia Fonseca Pereira
Oliveira Gomes

RECIFE

2026

FLÁVIO ADRIANO JESUS DE ARAÚJO
INFORMALIDADE E ESCOLARIDADE NO NORDESTE BRASILEIRO

Trabalho de Conclusão de Curso apresentado à
Universidade Federal Rural de Pernambuco
(UFRPE) como requisito para obtenção do título
de Bacharel em Ciências Econômicas.

Aprovado em: _____

BANCA EXAMINADORA

Prof.^a Dr.^a Sônia Maria Fonseca Pereira Oliveira Gomes
Universidade Federal Rural de Pernambuco

Prof.^a Dr.^a Camila Maria Pernambuco Peixoto da Silva
Universidade Federal Rural de Pernambuco

Prof.^a Dr.^o Risomário Williams da Silva
Universidade Federal Rural de Pernambuco

DEDICATÓRIA

À minha família, cujo apoio incondicional tornou possível cada etapa desta jornada.

Aos trabalhadores e trabalhadoras do Nordeste brasileiro, cuja trajetória de luta e resiliência diante da informalidade foi a principal motivação deste estudo.

AGRADECIMENTO

Agradeço, em primeiro lugar, à minha orientadora, Prof.^a Dr.^a Sônia Fonseca Pereira Oliveira Gomes, pela dedicação, rigor intelectual e pelas orientações precisas que nortearam o desenvolvimento deste trabalho e que nos apoiou desde o início em outras iniciativas como o Projeto de extensão Ecometrix.

Aos professores do Departamento de Economia da Universidade Federal Rural de Pernambuco (DECON/UFRPE), pelo conhecimento partilhado ao longo da graduação e pela formação crítica que fundamenta esta pesquisa.

Aos meus colegas de curso, pelas trocas, pelos debates e pelo companheirismo em todos os momentos, em especial aos amigos Celso Lopes, Maria Beatriz(Bia) e Rhuaan Pedro.

À minha família, pelo suporte afetivo e material indispensável à conclusão desta etapa. Ao meu companheiro de todas as horas e grande amor da minha vida Ricardo Ferraz.

Ao Instituto Brasileiro de Geografia e Estatística (IBGE), pela disponibilização pública dos microdados da PNAD Contínua, base empírica essencial deste estudo.

"A educação é a arma mais poderosa que você pode usar para mudar o mundo."

(Nelson Mandela)

RESUMO

Este estudo estima o efeito do nível de escolaridade sobre a probabilidade de inserção no mercado de trabalho informal no Nordeste brasileiro no período de 2019 a 2023. Utilizam-se os microdados de todos os trimestres da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua/IBGE), com amostra de 1.005.969 trabalhadores ocupados entre 14 e 65 anos residentes nos nove estados nordestinos. O método empregado é o modelo probit, com erros-padrão clusterizados por unidade da federação, e os resultados são reportados como efeitos marginais médios (APE). A variável dependente identifica como informal o empregado privado sem carteira assinada, o trabalhador doméstico sem carteira, o conta-própria sem CNPJ e o trabalhador não remunerado. As variáveis explicativas incluem quatro categorias de escolaridade — referência: sem instrução ou fundamental incompleto —, além de controles por sexo, idade, raça/cor, área de residência, estado e período. Os resultados indicam que a escolaridade reduz significativamente a probabilidade de informalidade: -6,1 p.p. para fundamental incompleto ou médio incompleto; -20,7 p.p. para médio completo ou superior incompleto; e -46,5 p.p. para superior completo — todos significativos a 1%. A análise estadual confirma heterogeneidade de 8,4 p.p. entre Maranhão e Alagoas em relação a Pernambuco. Os anos pandêmicos (2020–2021) não alteraram significativamente o padrão de informalidade; 2022 e 2023 registraram leve redução na fase de retomada. O pseudo- R^2 de McFadden é de 0,1155. Os achados corroboram a teoria do capital humano e são consistentes com Cunha et al. (2011) e Mello e Santos (2008), confirmando que o retorno educacional à formalização é especialmente pronunciado no Nordeste brasileiro.

Palavras-chave: Informalidade. Escolaridade. Mercado de trabalho. Nordeste. Modelo probit. PNAD Contínua.

ABSTRACT

This study estimates the effect of education level on the probability of informal labor market insertion in northeastern Brazil from 2019 to 2023. Microdata from all quarters of the Continuous National Household Sample Survey (PNAD Contínua/IBGE) are used, with a sample of 1,005,969 employed workers aged 14 to 65 residing in the nine northeastern states. A probit model is estimated with standard errors clustered by state, and results are reported as Average Partial Effects (APE). The dependent variable identifies as informal: private employees without a signed work card, domestic workers without a signed work card, own-account workers without CNPJ registration, and unpaid workers. Explanatory variables include four education categories — reference: no schooling or incomplete primary education — plus controls for sex, age, race/color, area of residence, state, and time period. Results show that education significantly reduces informality probability: -6.1 p.p. for incomplete primary or incomplete secondary; -20.7 p.p. for complete secondary or incomplete higher; and -46.5 p.p. for complete higher education — all significant at 1%. State effects confirm heterogeneity of 8.4 p.p. between Maranhão and Alagoas relative to Pernambuco. Pandemic years (2020–2021) showed no significant change in informality patterns; 2022 and 2023 registered a slight reduction during the economic recovery. McFadden pseudo- R^2 is 0.1155. Findings corroborate human capital theory and are consistent with Cunha et al. (2011) and Mello e Santos (2008), confirming that the educational return to formalization is especially pronounced in northeastern Brazil.

Keywords: Informality. Education. Labor market. Northeast Brazil. Probit model. PNAD Contínua.

LISTA DE ILUSTRAÇÕES

Figura 1 – Participação de cada região no total de trabalhadores informais do Brasil (2023)
[p.]

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas da amostra (Nordeste, 2019–2023) 33
Tabela 2 – Efeitos marginais médios do modelo probit: Nordeste (2019–2023) 34
Tabela 3 – Heterogeneidade estadual: efeitos marginais em relação a Pernambuco 37

LISTA DE ABREVIATURAS

APE – Average Partial Effects (Efeitos Marginais Médios)

FGTS – Fundo de Garantia por Tempo de Serviço

IBGE – Instituto Brasileiro de Geografia e Estatística

IPEA – Instituto de Pesquisa Econômica Aplicada

MPRA – Munich Personal RePEc Archive

MQO – Mínimos Quadrados Ordinários

OIT – Organização Internacional do Trabalho

PNAD – Pesquisa Nacional por Amostra de Domicílios

QDR – Quociente de Diferenciação Regional

SIDRA – Sistema IBGE de Recuperação Automática

UFRPE – Universidade Federal Rural de Pernambuco

UPA – Unidade Primária de Amostragem

SUMÁRIO

INTRODUÇÃO.....	12
2 FUNDAMENTAÇÃO TEÓRICA.....	15
2.1 Conceito e mensuração da informalidade no mercado de trabalho	15
2.2 Os custos da informalidade: evidências empíricas	16
2.3 Escolaridade e capital humano como determinantes da formalização	18
2.4 A informalidade no Nordeste: heterogeneidades regionais e estrutura produtiva	19
2.5 Efeitos da pandemia de Covid-19 sobre a informalidade (2019–2023)	21
3 METODOLOGIA.....	22
3.1 Base de dados	22
3.2 Modelo Empírico.....	25
4 RESULTADOS	27
4.1 Estatísticas Descritivas da Amostra	27
4.2 Resultados do Modelo Probit: Modelo Geral	28
4.2.1 Hipótese H1: Efeito da Escolaridade sobre a Informalidade.....	29
4.2.2 Resultados das Variáveis de Controle.....	30
4.2.3 Hipótese H3: Heterogeneidade Estadual	30
4.2.4 Efeitos Temporais e o Impacto da Pandemia.....	32
4.3 Síntese dos Resultados e Verificação das Hipóteses.....	32
CONSIDERAÇÕES FINAIS	34
REFERÊNCIAS	38

INTRODUÇÃO

O mercado de trabalho brasileiro apresenta, de forma recorrente ao longo de sua trajetória histórica, elevadas proporções de trabalhadores em situação de informalidade ocupacional. Tal fenômeno não constitui manifestação conjuntural, associada a ciclos específicos de contração econômica, mas expressa uma característica estrutural e persistente das relações de trabalho no país, verificada de modo contínuo em distintos regimes macroeconômicos (CAMARGO et al., 2006; CUNHA et al., 2011). Em 2025, a taxa nacional de informalidade alcançou 38,1%, o menor patamar da série histórica iniciada em 2012; contudo, os estados nordestinos seguem registrando índices bem superiores à média nacional, com Maranhão (58,7%), Pará (58,5%) e Bahia (52,8%) entre as maiores taxas do país (IBGE, 2026).

No âmbito regional, o Nordeste ocupa posição de destaque na concentração do trabalho informal: em 2019, a região concentrava cerca de 25% da população ocupada do Brasil, mas respondia por aproximadamente 35% do total de trabalhadores informais do país (IBGE, 2020). Mesmo após a recuperação econômica de 2022–2023, os estados nordestinos permaneceram acumulando índices de informalidade acima de 50%, contra cerca de 39% na média nacional, evidenciando que a expansão do emprego na região ocorreu de forma estruturalmente assimétrica (IBGE, 2024).

Esse padrão regional encontra explicação, em boa parte, em uma única variável recorrente na literatura: a escolaridade. Cunha et al. (2011) estimam, via modelos probit com dados da PNAD, que possuir nível superior reduz em até 14,52 pontos percentuais a probabilidade de inserção informal nas regiões Norte e Nordeste, efeito mais pronunciado justamente onde a informalidade é mais intensa. Mello e Santos (2008) atribuem à aceleração educacional da força de trabalho cerca de 77% da queda da informalidade nacional entre 2002 e 2007, superando em relevância explicativa tanto mudanças institucionais quanto choques de demanda agregada. No Nordeste especificamente, Kon (2004) mostra que o diferencial educacional entre trabalhadores formais e informais é historicamente o mais elevado do país, reforçando a hipótese de que o capital humano é um determinante estrutural, e não apenas conjuntural, da informalidade regional.

É nesse contexto que a presente pesquisa se insere: compreender como a escolaridade molda a probabilidade de inserção informal no Nordeste brasileiro ganha relevância tanto

acadêmica, ao testar empiricamente essa relação para o período recente (2019–2023, que abrange os efeitos da pandemia de Covid-19), quanto de política pública, ao subsidiar estratégias de formalização ancoradas na expansão educacional.

Diante do exposto, este estudo busca responder à seguinte questão central: qual o efeito do nível de escolaridade sobre a probabilidade de inserção no mercado de trabalho informal no Nordeste brasileiro entre 2019 e 2023? Dessa questão derivam três indagações específicas: o efeito da escolaridade difere entre áreas urbanas e rurais da região? Há heterogeneidades estaduais nesse impacto, refletindo as distintas estruturas produtivas dos estados nordestinos? O choque da pandemia de 2020 alterou a magnitude ou a direção dessa relação? Para encaminhar essas questões, formulam-se três hipóteses: H1, de que maiores níveis de escolaridade reduzem significativamente a probabilidade de inserção no mercado de trabalho informal; H2, de que esse efeito redutor é mais pronunciado em áreas urbanas do que em áreas rurais; H3, de que há diferenças estatisticamente significativas no impacto da escolaridade entre os estados nordestinos, em razão de suas heterogeneidades estruturais.

Para responder a essas questões, este trabalho tem como objetivo geral estimar o efeito do nível de escolaridade sobre a probabilidade de inserção no mercado de trabalho informal no Nordeste brasileiro, no período de 2019 a 2023, mediante a aplicação de modelos econométricos de escolha discreta (probit) aos microdados da PNAD Contínua. Como objetivos específicos, busca-se: (i) mensurar o efeito marginal do nível de escolaridade sobre a probabilidade de informalidade, controlando por sexo, idade, raça/cor, área de residência e posição no domicílio, em conformidade com a especificação adotada por Cunha et al. (2011); (ii) comparar a magnitude desse efeito entre áreas urbanas e rurais dos estados nordestinos; (iii) verificar alterações na relação entre escolaridade e informalidade no período pré e pós-pandemia (2019–2023); e (iv) discutir as implicações dos resultados para políticas públicas voltadas à expansão da escolaridade e à redução da informalidade na região.

A metodologia combina levantamento bibliográfico, que fundamenta teoricamente o estudo com base na literatura nacional e internacional sobre informalidade, capital humano e mercado de trabalho regional, e análise de dados secundários da PNAD Contínua (IBGE) para 2019–2023, operacionalizada por meio de modelos probit, em conformidade com a abordagem de Cunha et al. (2011) e Mello e Santos (2008).

Além deste capítulo introdutório, o trabalho está organizado em quatro seções adicionais. O segundo capítulo apresenta a fundamentação teórica, abrangendo os fundamentos

conceituais da informalidade, os determinantes educacionais da inserção formal e as especificidades estruturais do mercado de trabalho nordestino. O terceiro capítulo descreve a metodologia, com detalhamento da base de dados, das variáveis e do modelo econométrico. O quarto capítulo apresenta e discute os resultados. O quinto capítulo expõe as considerações finais, com síntese das conclusões e indicações para políticas públicas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Conceito e mensuração da informalidade no mercado de trabalho

A primeira definição oficial de informalidade foi formulada pela Organização Internacional do Trabalho (OIT) em 1972, no âmbito de um estudo sobre a economia do Quênia. Naquele documento inaugural, o setor informal era delimitado por um conjunto de características operacionais: inexistência de barreiras à entrada, utilização preferencial de recursos domésticos, propriedade individual dos meios de produção, escala reduzida de operação, processos produtivos intensivos em trabalho com tecnologia adaptada e funcionamento em mercados competitivos não regulamentados (OIT, 1972 *apud* CUNHA et al., 2011). A partir dessa referência, o conceito passou por sucessivas reformulações acompanhando as transformações estruturais das economias em desenvolvimento (KON, 2004).

Em 2002, na 90ª Conferência Internacional do Trabalho, a OIT adotou o termo mais abrangente de "economia informal", reconhecendo a heterogeneidade e o dinamismo inerentes às atividades não reguladas, e superando a dicotomia estanque entre setores formal e informal em favor de uma perspectiva de graus de formalização (CUNHA et al., 2011). Essa reformulação conceitual foi relevante para reconhecer que o fenômeno não se limita a um conjunto homogêneo de trabalhadores, mas abrange diferentes modalidades de inserção precária, com distintas causas e consequências (ULYSSEA, 2005a).

No contexto brasileiro, a operacionalização do conceito de informalidade frequentemente adota critérios relacionados ao vínculo empregatício e à proteção previdenciária. Ulyssea (2005a) aponta que as principais definições utilizadas na literatura nacional associam informalidade à ausência de carteira de trabalho assinada e à não contribuição para a previdência social, abrangendo empregados sem registro formal, trabalhadores domésticos sem carteira e trabalhadores por conta própria sem contribuição previdenciária. Essa operacionalização, amplamente adotada nos estudos com dados da PNAD, permite identificar o trabalhador informal como aquele privado dos direitos e proteções da legislação trabalhista brasileira (CAMARGO et al., 2006; LEONE, 2010 *apud* CUNHA et al., 2011).

Ulyssea (2005b) acrescenta uma dimensão institucional ao debate, argumentando que a informalidade não resulta apenas de características individuais dos trabalhadores, mas também

de incentivos e desincentivos criados pelo arcabouço regulatório e tributário ao qual as firmas estão submetidas. Nessa perspectiva, a decisão de formalizar ou não um trabalhador é uma escolha econômica das firmas, condicionada pelos custos de compliance, pela capacidade de fiscalização do Estado e pela produtividade esperada dos trabalhadores, o que conecta a informalidade tanto às características da oferta de trabalho quanto às condições da demanda por trabalhadores formais.

Mais recentemente, Ulyssea (2020) aprofundou esse debate em artigo publicado no *American Economic Review*, utilizando um modelo de equilíbrio geral com firmas heterogêneas para decompor os determinantes da informalidade no Brasil. Os resultados indicam que tanto as condições da demanda por trabalho formal, determinadas pelos custos de entrada e compliance das firmas, quanto as características dos trabalhadores (especialmente escolaridade e produtividade) contribuem de forma independente para a persistência da informalidade. Esse trabalho fornece base teórica relevante para a abordagem empírica adotada neste TCC, ao demonstrar que a educação opera como determinante da formalização tanto pelo canal da oferta de trabalho quanto pelo da demanda das firmas por trabalhadores mais qualificados.

2.2 Os custos da informalidade: evidências empíricas

A inserção no setor informal implica, sob múltiplas dimensões, condições de trabalho desfavoráveis em relação ao emprego formal. Do ponto de vista da proteção social, o trabalhador informal é privado do acesso ao Fundo de Garantia por Tempo de Serviço (FGTS), ao seguro-desemprego, às férias remuneradas e às contribuições previdenciárias, o que amplia significativamente sua vulnerabilidade diante de choques adversos de renda, como doenças, acidentes ou recessões (CUNHA et al., 2011). Do ponto de vista produtivo, as atividades informais caracterizam-se pela escala reduzida, pela adoção limitada de tecnologias modernas e pela baixa separação entre capital e trabalho, resultando em capacidade de acumulação restrita e remunerações sistematicamente inferiores às do setor formal (KON, 2004).

Camargo et al. (2006), em análise setorial da economia brasileira para o período de 1992 a 2001, verificaram que a remuneração média dos trabalhadores formais superou, de forma consistente, a dos informais em todos os 14 setores analisados. Esse diferencial salarial expressivo confirma que a informalidade está associada a uma penalidade econômica significativa para o trabalhador, independentemente do setor de atividade. Os autores

demonstram ainda que, embora o setor informal tenha sido responsável por aproximadamente 54% dos empregos gerados no período, esses postos de trabalho caracterizavam-se por remunerações médias significativamente menores, perpetuando ciclos de pobreza e desigualdade.

Nas duas décadas seguintes, a estrutura de diferenciais salariais entre setores formal e informal manteve-se persistente. Segundo a FGV IBRE (VELOSO; BARBOSA FILHO; PERUCHETTI, 2022a), com base em dados da PNAD e da PNAD Contínua para o período 1992–2021, a melhora na composição educacional da mão de obra foi o principal determinante estrutural da redução da taxa de informalidade no país, responsável por uma queda acumulada de 9,4 pontos percentuais. Na ausência desse avanço educacional, a taxa de informalidade teria aumentado em 6,5 p.p., passando de 55,7% para 62,2%, em vez de recuar. Ademais, segundo a PNAD Contínua referente ao segundo trimestre de 2024 (IBGE, 2024), o percentual de trabalhadores com carteira assinada no setor privado no Nordeste permanecia significativamente inferior à média nacional, com os estados nordestinos concentrando as maiores taxas de informalidade do país, com destaque para Maranhão (55,3%), Piauí (56,6%) e Pará (58,1%) como as três unidades federativas com maior informalidade média anual em 2024 (IBGE, 2025).

No Nordeste, essa penalidade é ainda mais pronunciada. Segundo dados da PNAD Contínua para 2022 (IBGE, 2023), o rendimento médio do trabalho principal dos empregados com carteira assinada na região era aproximadamente 2,1 vezes superior ao dos trabalhadores sem carteira. Considerando que a taxa de informalidade nordestina superava 52% no mesmo período, depreende-se que a maioria da força de trabalho regional está submetida a condições remuneratórias significativamente inferiores às do mercado formal, com reflexos diretos sobre a renda familiar, o consumo e o bem-estar das populações mais vulneráveis. Esses dados fundamentam a relevância do estudo da informalidade como fenômeno central das desigualdades regionais brasileiras.

A FGV IBRE tem acompanhado de forma sistemática a dinâmica da informalidade ocupacional no Brasil. Em análise do mercado de trabalho brasileiro e nordestino, Veloso, Barbosa Filho e Peruchetti (2022b) identificam que o Nordeste sofreu perdas permanentes de postos de trabalho formais durante a recessão de 2014-2016, padrão distinto do observado durante a pandemia de Covid-19. Adicionalmente, os autores apontam que o mercado de trabalho nordestino é estruturalmente mais precário tanto em razão da escolaridade média

menor da força de trabalho regional quanto pelas características específicas dos postos de trabalho oferecidos na região, resultando em taxas de informalidade e rendimentos médios inferiores mesmo quando se controla pelo nível de escolaridade dos trabalhadores (VELOSO; BARBOSA FILHO; PERUCHETTI, 2022b). Esse quadro reforça a relevância de políticas de expansão educacional voltadas especificamente ao Nordeste como instrumento de formalização estrutural do mercado de trabalho regional.

2.3 Escolaridade e capital humano como determinantes da formalização

A relação entre educação e inserção formal no mercado de trabalho encontra sua fundamentação teórica na abordagem do capital humano, sistematizada pelos trabalhos seminais de Schultz (1961) e Becker (1964). Segundo essa perspectiva, o investimento em educação e qualificação profissional eleva a produtividade marginal do trabalhador, tornando-o mais atrativo para empregadores do setor formal, que demandam trabalhadores mais produtivos e tendem a internalizá-los mediante contratos formalizados. O capital humano, portanto, não apenas amplia a produtividade individual, mas também constitui o principal mecanismo pelo qual os trabalhadores transitam do setor informal para o formal.

No contexto brasileiro, a evidência empírica é robusta em confirmar essa relação. Mello e Santos (2008), com base nos microdados da PNAD de 2002 a 2007, identificam um gradiente educacional expressivo: trabalhadores com ensino básico incompleto registravam taxa de informalidade de 74,0%, enquanto aqueles com ensino médio completo apresentavam taxa de 33,8% e os com nível superior completo, de apenas 25,8%. Mais relevante, os autores demonstram, por meio de decomposição contrafactual, que a aceleração educacional da força de trabalho, verificada a partir do final dos anos 1990, foi o principal determinante individual da queda da informalidade no período: a mudança na composição educacional da população ocupada foi responsável por aproximadamente 77% da variação total da taxa de informalidade entre 2002 e 2007, superando em relevância as inovações institucionais e os estímulos conjunturais de demanda agregada.

Estudo mais recente do FGV IBRE (VELOSO; BARBOSA FILHO; PERUCHETTI, 2022a) confirma e atualiza esse diagnóstico para o período 1992–2021. Utilizando exercício contrafactual com dados da PNAD e da PNAD Contínua, os autores estimam que a melhora na composição educacional da força de trabalho respondeu por redução acumulada de 9,4 pontos percentuais na taxa de informalidade. Sem esse avanço, a taxa teria aumentado 6,5 p.p., de 55,7% para 62,2%, evidenciando que a trajetória de formalização observada nas últimas três

décadas foi, em grande medida, sustentada pela progressiva elevação do nível de escolaridade da população ocupada, verificada de modo contínuo mesmo em períodos de instabilidade econômica.

Cunha et al. (2011), utilizando modelos *Probit* com dados da PNAD para o período 2007–2009, estimam que a posse de nível superior reduzia entre 3,59 e 14,52 pontos percentuais a probabilidade de inserção informal, dependendo da região do país. Esses resultados são particularmente pronunciados nas regiões Norte e Nordeste, indicando que o retorno educacional em termos de formalização é justamente maior nas regiões onde a informalidade é mais intensa e onde os déficits históricos de escolaridade são mais elevados. Kon (2004) reforça esse diagnóstico ao demonstrar que, no Nordeste, o diferencial de escolaridade entre formais e informais é o mais elevado do país: a média de anos de estudo dos empregados com carteira assinada superava a dos sem carteira em 114%, contra 22% em São Paulo.

2.4 A informalidade no Nordeste: heterogeneidades regionais e estrutura produtiva

A macrorregião Nordeste constitui o lócus de maior concentração da informalidade no Brasil, fenômeno que não resulta de fatores conjunturais, mas de condicionantes histórico-estruturais do seu processo de desenvolvimento. Kon (2004), com base na PNAD de 1999, demonstra que a proporção de trabalhadores sem vínculo empregatício formalizado nas empresas nordestinas situava-se 39% acima da média nacional. A estrutura setorial da economia regional, marcada pela preponderância da agropecuária, do comércio informal e dos serviços pessoais de baixa produtividade, é o principal determinante dessa concentração, uma vez que esses são exatamente os setores com maior participação do emprego informal identificados por Camargo et al. (2006).

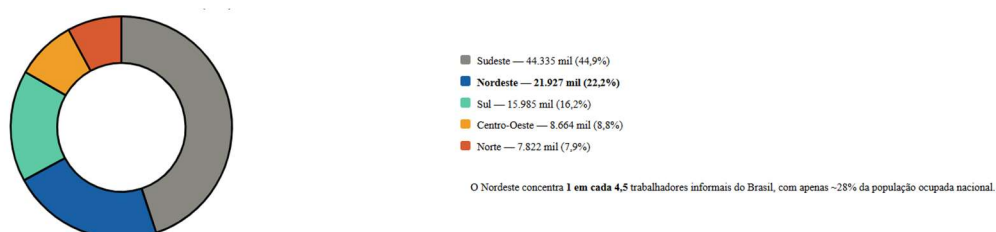


Figura 1: Participação de cada região no total de Trabalhadores informais do Brasil (2023)

Fonte: IBGE, PNAD Contínua, Tabela 4680 (2023)

Esses dados atualizam e confirmam o diagnóstico da literatura para o período recente. A estrutura setorial da economia nordestina, marcada pela preponderância da agropecuária, do comércio informal e dos serviços pessoais de baixa produtividade, identificados por Camargo et al. (2006) como os setores de maior concentração de emprego informal, permanece como determinante central dessa participação desproporcional no estoque nacional de informalidade.

A oferta de trabalho rural não agrícola constitui outra dimensão relevante da informalidade no Nordeste. Mesquita, Sampaio e Ramalho (MESQUITA; SAMPAIO; RAMALHO, 2010) demonstram que a pluriatividade, combinação de atividades agrícolas e não agrícolas, é estratégia amplamente adotada pelos trabalhadores rurais nordestinos como mecanismo de complementação de renda, especialmente em municípios com mercados de trabalho locais pouco desenvolvidos. Essa característica da força de trabalho rural nordestina contribui para a elevação das taxas de informalidade na região, dado que as atividades não agrícolas exercidas no meio rural raramente são formalizadas.

Cunha et al. (2011) evidenciam heterogeneidades significativas entre as regiões brasileiras na determinação da informalidade. Enquanto no Sudeste o impacto marginal do nível superior sobre a probabilidade de inserção informal é de -3,59 pontos percentuais, no Norte e Nordeste esse impacto chega a -14,52 pontos percentuais, evidenciando que os retornos educacionais em termos de formalização são estruturalmente maiores nas regiões de maior vulnerabilidade. Isso sugere que políticas de expansão educacional têm potencial de impacto proporcionalmente maior justamente nas regiões onde a informalidade é mais intensa e onde os déficits de escolaridade são historicamente mais elevados.

Dados mais recentes confirmam a persistência dessas heterogeneidades. Segundo a PNAD Contínua do 2º trimestre de 2024 (IBGE, 2024), o Nordeste apresentava o menor nível de ocupação entre todas as regiões brasileiras (48,9%), ao mesmo tempo em que concentrava o maior percentual de trabalhadores sem instrução ou com nível fundamental incompleto na força de trabalho (25,6%), contra apenas 11,8% no Nordeste por nível superior completo, reforçando o vínculo estrutural entre baixa escolaridade e informalidade na região. No Nordeste, cerca de 60% dos jovens de 18 a 24 anos atuam na informalidade e a taxa de desemprego deste grupo supera 20% (FGV IBRE, 2024), o que evidencia que a reprodução da informalidade na região tem forte componente geracional, associado às dificuldades de acumulação de capital humano entre os mais jovens. Esses dados atualizam e corroboram o diagnóstico estrutural identificado

pela literatura, demonstrando que, duas décadas após os estudos seminais de Camargo et al. (2006) e Kon (2004), a posição relativa do Nordeste no espectro da informalidade permanece inalterada.

2.5 Efeitos da pandemia de Covid-19 sobre a informalidade (2019–2023)

A pandemia de Covid-19 produziu efeitos significativos e heterogêneos sobre o mercado de trabalho brasileiro, com consequências particularmente relevantes para a informalidade no Nordeste. No primeiro semestre de 2020, as medidas de isolamento social e a contração da atividade econômica resultaram em forte elevação do desemprego e na saída de milhões de trabalhadores da força de trabalho, fenômeno denominado "desalento" ou inatividade forçada. No Nordeste, a taxa de desocupação atingiu o pico de 20,1% no segundo trimestre de 2020, segundo dados da PNAD Contínua (IBGE, 2020), impactando desproporcionalmente os trabalhadores informais, que não dispunham de seguro-desemprego ou FGTS para atravessar o período de crise.

A partir de 2021, com a flexibilização das restrições sanitárias e a gradual retomada econômica, observou-se recuperação do mercado de trabalho, porém com características estruturalmente problemáticas. Segundo a Carta de Conjuntura do IPEA (2022, 2023), a reinserção dos trabalhadores nordestinos no mercado de trabalho ocorreu predominantemente via postos informais, dado que a capacidade de geração de emprego formal é limitada pela estrutura produtiva regional. Em 2023, embora a taxa de desocupação nacional tenha recuado ao menor patamar da série histórica, o Nordeste ainda registrava informalidade superior a 50%, sugerindo que a "recuperação" do mercado de trabalho se deu em bases precárias. Essa dinâmica justifica a escolha da janela temporal 2019–2023, que permite contrastar os padrões de informalidade pré-pandemia, durante a crise e na fase de recuperação, testando se o choque sanitário alterou a magnitude do efeito educacional sobre a probabilidade de inserção informal (H3 deste estudo).

3 METODOLOGIA

A etapa empírica desta pesquisa foi conduzida na linguagem Python (versão 3.x), escolhida por sua versatilidade no tratamento de bases de dados de grande volume, pela ampla disponibilidade de bibliotecas voltadas à econometria aplicada e pela reprodutibilidade analítica que oferece. Para a organização, limpeza e manipulação dos microdados da PNAD Contínua, foi utilizada a biblioteca pandas, com apoio do numpy para operações vetorizadas. As estimações dos modelos econométricos foram realizadas com a biblioteca statsmodels, que disponibiliza a classe Probit com suporte a erros-padrão clusterizados e ao cálculo de efeitos marginais médios (Average Partial Effects). A ponderação amostral foi incorporada diretamente na estimação por meio do argumento de frequência de pesos (`freq_weights`), em conformidade com as recomendações metodológicas do IBGE para uso dos microdados da PNAD Contínua. A produção de estatísticas descritivas e visualizações gráficas contou com o uso das bibliotecas matplotlib e seaborn. Os microdados foram carregados a partir dos arquivos de microdados trimestrais em formato de largura fixa (.txt), acompanhados do dicionário de variáveis fornecido pelo IBGE, garantindo a correta leitura e codificação das variáveis de interesse.

3.1 Base de dados

A principal fonte de dados utilizada neste estudo é a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Trata-se de uma pesquisa domiciliar de cortes transversais repetidos com coleta contínua ao longo do ano e divulgação de resultados trimestrais, que investiga as características socioeconômicas da população brasileira, com ênfase nas condições de inserção no mercado de trabalho. A PNAD Contínua substitui, a partir de 2012, a PNAD anual tradicional, apresentando vantagens metodológicas significativas em termos de frequência de divulgação, cobertura geográfica e desagregação das informações (IBGE, 2014).

Serão utilizados os microdados dos quatro trimestres dos anos de 2019, 2020, 2021, 2022 e 2023, restringindo a amostra à população residente nos nove estados da macrorregião Nordeste, a saber: Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe e Bahia, com idade entre 14 e 65 anos e que declararam estar ocupados na semana de referência da pesquisa. A janela temporal 2019–2023 permite analisar três estágios

distintos do mercado de trabalho regional: o período pré-pandêmico (2019), a fase de crise sanitária e econômica (2020–2021) e a fase de recuperação (2022–2023).

A variável dependente do modelo é construída a partir das informações sobre a posição na ocupação e a existência de carteira de trabalho assinada. Define-se como trabalhador informal aquele que se enquadra em ao menos uma das seguintes condições: (i) empregado do setor privado sem carteira de trabalho assinada; (ii) trabalhador doméstico sem carteira assinada; (iii) trabalhador por conta própria sem CNPJ; ou (iv) trabalhador não remunerado. Essa definição é consistente com a operacionalização adotada por Cunha et al. (2011) e Mello e Santos (2008), o que garante comparabilidade direta com os resultados da literatura de referência.

As variáveis explicativas incorporadas à análise compreendem: (i) escolaridade, variável de interesse principal, construída como variável categórica com quatro níveis: sem instrução ou fundamental incompleto (categoria de referência), fundamental completo ou médio incompleto, médio completo ou superior incompleto, e superior completo; (ii) sexo (masculino/feminino); (iii) idade e idade ao quadrado, para captar a relação não linear entre experiência potencial no mercado de trabalho e informalidade; (iv) raça/cor (branca/não branca); (v) área de residência (urbana/rural); (vi) posição no domicílio (pessoa de referência ou demais); e (vii) unidade da federação, por meio de dummies estaduais que captam heterogeneidades estruturais entre os estados nordestinos. Essa especificação é consistente com a adotada por Cunha et al. (2011) e ampliada com os controles de área e estado, necessários para o atendimento dos objetivos específicos deste estudo.

Como as notas técnicas da PNAD Contínua estabelecem que os pesos amostrais são parte central do processo de expansão da amostra e devem ser utilizados na produção de estimativas derivadas da pesquisa, todos os modelos serão estimados com ponderação pelo peso amostral V1028, de modo a respeitar o desenho da pesquisa e produzir estimativas estatisticamente consistentes para a população de referência. Para a construção do desenho amostral complexo em python, serão utilizadas as variáveis de unidade primária de amostragem (UPA) e de estrato, além do peso V1028, conforme recomendado pelo IBGE.

Além dos microdados da PNAD Contínua, são consultadas como fontes complementares a Carta de Conjuntura do Instituto de Pesquisa Econômica Aplicada (IPEA), para indicadores agregados do mercado de trabalho no período recente, e o Sistema IBGE de Recuperação Automática (SIDRA), para séries históricas de desocupação e informalidade por

região. Essas fontes são utilizadas exclusivamente para contextualização descritiva na introdução e no referencial teórico; a estimação econométrica é integralmente realizada sobre os microdados da PNAD Contínua.

3.2 Modelo Econométrico

Para estimar a probabilidade de um trabalhador estar inserido no setor informal em função do seu nível de escolaridade e demais características observáveis, adota-se o modelo de escolha discreta do tipo probit. A justificativa para essa escolha reside na natureza binária da variável dependente: a condição de informalidade (1) ou formalidade (0) não é uma variável contínua e, portanto, o uso de Mínimos Quadrados Ordinários (MQO) geraria probabilidades estimadas fora do intervalo $[0,1]$ e erros-padrão inconsistentes. O modelo probit, ao assumir que os erros seguem distribuição normal padrão, garante que as probabilidades estimadas estejam sempre contidas no intervalo unitário, sendo, portanto, a especificação mais adequada para o presente estudo. Essa abordagem é coerente com a literatura de referência sobre determinantes da informalidade no Brasil, que adota sistematicamente o modelo probit com dados da PNAD (CUNHA et al., 2011; MELLO; SANTOS, 2008).

No modelo probit, a probabilidade de um indivíduo i ser informal é dada por:

$$P(Y_i = 1 | X_i) = \Phi(X_i'\beta) \quad (1)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da distribuição normal padrão; Y_i é a variável binária que assume valor 1 quando o indivíduo está em ocupação informal e 0 quando está em ocupação formal; X_i é o vetor de variáveis explicativas; e β é o vetor de coeficientes estimado por máxima verossimilhança.

Os parâmetros do modelo probit não possuem interpretação direta em termos de magnitude, de modo que serão calculados e reportados os efeitos marginais médios (APE, do inglês Average Partial Effects), que indicam a variação em pontos percentuais na probabilidade de informalidade associada a cada categoria de escolaridade em relação à categoria de referência (sem instrução ou fundamental incompleto). Essa forma de reportar os resultados é diretamente comparável à adotada por Cunha et al. (2011) e Mello e Santos (2008).

As estimações serão realizadas em etapas sucessivas. Em uma primeira especificação, será estimado o modelo geral para toda a região Nordeste, com todas as variáveis de controle, servindo de referência para as demais. Em seguida, serão estimados modelos separados por área de residência (urbana e rural), a fim de verificar se o efeito protetor da escolaridade sobre a

formalização difere entre esses contextos. Em uma terceira etapa, serão estimados modelos por estado da federação, permitindo identificar heterogeneidades no retorno educacional à formalização entre os nove estados nordestinos. Por fim, serão comparados os efeitos marginais médios (Average Partial Effects – APEs) para o ano de 2019 (pré-pandemia) e para os anos de 2022–2023 (pós-pandemia), verificando se o choque sanitário alterou a magnitude ou a direção do efeito educacional sobre a informalidade.

Os erros-padrão serão robustos à heterocedasticidade (erros-padrão de Huber-White), dada a possível presença de variância não constante em amostras de corte transversal com perfis heterogêneos de trabalhadores. Essa abordagem permite verificar, ainda, se a associação entre escolaridade e informalidade se mantém estável ou se se reduz à medida que novos grupos de controle são incorporados ao modelo.

3.2 Modelo Empírico

O modelo empírico central deste estudo pode ser representado pela Equação 2:

$$P(\text{Informal}_i = 1) = \Phi(\beta_0 + \beta_1 \text{Escolaridade}_i + \beta_2 \text{Sexo}_i + \beta_3 \text{Idade}_i + \beta_4 \text{Idade}_i^2 + \beta_5 \text{Raça}_i + \beta_6 \text{Área}_i + \beta_7 \text{PosiçãoDom}_i + \gamma' \text{Estado}_i + \delta' \text{Ano}_i + \varepsilon_i) \quad (2)$$

em que Informal_i é a variável binária dependente que assume valor 1 quando o indivíduo i está em ocupação informal e 0 quando está em ocupação formal; Escolaridade_i é o vetor de dummies de escolaridade, variável de interesse principal, com a categoria "sem instrução ou fundamental incompleto" como referência omitida; Estado_i é o vetor de dummies estaduais, com Pernambuco como referência por concentrar o maior mercado de trabalho formal da região, captando heterogeneidades estruturais fixas entre os estados nordestinos; Ano_i é o vetor de dummies anuais, com 2019 como referência, que captam os efeitos comuns a todos os trabalhadores em cada período, inclusive o impacto da pandemia de Covid-19; e ε_i é o termo de erro aleatório.

O coeficiente de interesse central é β_1 , que mede o efeito da escolaridade sobre a probabilidade de informalidade após controle por todas as demais características observáveis. A interpretação do coeficiente β_1 , calculado como efeito marginal médio (APE), é central para o estudo: ele indica a diferença em pontos percentuais na probabilidade de informalidade entre cada categoria de escolaridade e a categoria de referência. Caso os efeitos marginais associados às categorias de maior escolaridade se reduzam de magnitude quando comparados entre as especificações por área e por estado, isso poderá ser interpretado como evidência de que o

retorno educacional à formalização é heterogêneo e condicionado pelo contexto territorial e estrutural do mercado de trabalho local.

A análise será complementada por estatísticas descritivas das variáveis utilizadas, incluindo taxas de informalidade por nível de escolaridade, por área e por estado, que servirão para contextualizar e dar suporte à interpretação dos resultados econométricos. Gráficos e tabelas serão elaborados para ilustrar o gradiente educacional da informalidade no Nordeste ao longo do período 2019–2023, tanto para o conjunto da região quanto para cada estado e para as zonas urbana e rural. Essa análise descritiva permite, ainda, identificar padrões e anomalias nos dados que orientem a leitura dos coeficientes estimados.

4 RESULTADOS

4.1 Estatísticas Descritivas da Amostra

A amostra utilizada nas estimações reúne 1.005.969 observações, correspondentes aos trabalhadores ocupados entre 14 e 65 anos residentes nos nove estados do Nordeste brasileiro nos 20 trimestres do período 2019–2023. A Tabela 1 apresenta as principais estatísticas descritivas das variáveis utilizadas no modelo.

Tabela 1 – Estatísticas descritivas da amostra (Nordeste, 2019–2023)

Variável	Obs.	Proporção / Média	Descrição
d_informal	1.005.969	37,0%	Variável dependente: 1 = informal
d_SemT (ref.)	335.300	33,3%	Sem instrução ou fund. incompleto
d_fund	142.529	14,2%	Fund. incompleto ou médio incompleto
d_Medio	377.349	37,5%	Médio completo ou superior incompleto
d_Superior	150.791	15,0%	Superior completo
d_sexo	592.292	58,9%	Masculino = 1
Idade	1.005.969	38,9 anos	Contínua (mín. 14, máx. 65)
d_urbano	698.598	69,4%	Área urbana = 1
d_branco	243.181	24,2%	Cor branca = 1 (ref.: pardo)
d_preto	120.990	12,0%	Cor preta = 1
d_pardo (ref.)	632.329	62,9%	Cor parda (referência)

Fonte: Elaboração própria com base nos microdados da PNAD Contínua (IBGE, 2019–2023).

Do total de trabalhadores ocupados na amostra, 37,0% (371.931 indivíduos) encontram-se em condição de informalidade, taxa coerente com os indicadores regionais sistematizados pelo IBGE para o período. Em relação ao perfil educacional, verifica-se que 33,3% dos trabalhadores estão na categoria de referência, sem instrução ou ensino fundamental incompleto, 14,2% possuem ensino fundamental incompleto ou médio incompleto, 37,5% têm ensino médio completo ou superior incompleto e 15,0% concluíram o ensino superior. Quanto ao perfil demográfico, a amostra é majoritariamente masculina (58,9%), com média de idade de 38,9 anos, e concentra-se em áreas urbanas (69,4%). A composição racial espelha as características da população nordestina: 62,9% são pardos, 24,2% brancos e 12,0% pretos.

4.2 Resultados do Modelo Probit: Modelo Geral

A Tabela 2 apresenta os efeitos marginais médios (APE, do inglês Average Partial Effects) estimados pelo modelo probit para a amostra completa do Nordeste, com erros-padrão clusterizados por unidade da federação. O modelo converge adequadamente, com pseudo- R^2 de McFadden igual a 0,1155, valor comparável ao reportado por Cunha et al. (2011), entre 0,115 e 0,117, o que atesta a adequação da especificação adotada.

Tabela 2 – Efeitos marginais médios do modelo probit: Nordeste (2019–2023)

Variável	Educação(ref.:Sem instrução / Fund. incompleto)	dy/dx	Erro-padrão	Estatística z	p-valor
d_fund		−0,0606***	0,005	−12,89	0,000
d_Medio		−0,2070***	0,005	−41,72	0,000
d_Superior		−0,4654***	0,009	−53,12	0,000
Controles individuais					
d_sex0 (masc.)		−0,1608***	0,006	−24,94	0,000
Idade		−0,0019*	0,001	−1,78	0,076
Idade ²		0,0001***	0,000	5,05	0,000
d_urbano		−0,0736***	0,011	−6,64	0,000
d_branco		−0,0093***	0,002	−3,91	0,000
d_preto		0,0008	0,003	0,26	0,798
d_amarelo		−0,0114**	0,005	−2,39	0,017
d_indigena		0,0033	0,008	0,42	0,674
Dummies estaduais (ref.: Pernambuco)					
d_MA		0,0216***	0,001	16,25	0,000
d_PI		0,0103***	0,001	12,01	0,000
d_CE		−0,0180***	0,000	−38,02	0,000
d_RN		−0,0481***	0,001	−81,95	0,000
d_PB		−0,0218***	0,000	−98,50	0,000
d_AL		−0,0628***	0,000	−236,54	0,000
d_SE		−0,0377***	0,001	−64,50	0,000
d_BA		−0,0248***	0,001	−19,58	0,000
Dummies temporais (ref.: 2019 / T4)					
d_2020		−0,0026	0,003	−0,85	0,394
d_2021		0,0047	0,005	0,90	0,368
d_2022		−0,0156**	0,007	−2,17	0,030
d_2023		−0,0151**	0,007	−2,32	0,021
d_T1		0,0069***	0,001	5,12	0,000
d_T2		0,0060***	0,002	3,08	0,002
d_T3		0,0032	0,002	1,74	0,081

Fonte: Elaboração própria. Notas: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$. Erros-padrão clusterizados por UF. $N = 1.005.969$. Pseudo- R^2 de McFadden = 0,1155.

4.2.1 Hipótese H1: Efeito da Escolaridade sobre a Informalidade

Os resultados da variável de interesse central, as dummies de escolaridade, confirmam a Hipótese H1 deste estudo: maiores níveis de escolaridade reduzem significativamente a probabilidade de inserção no mercado de trabalho informal no Nordeste brasileiro, mesmo após controle por características individuais observáveis.

Tomando como referência a categoria de trabalhadores sem instrução ou com ensino fundamental incompleto, os efeitos marginais estimados revelam um gradiente educacional negativo e crescente em magnitude. Trabalhadores com ensino fundamental incompleto ou médio incompleto (d_fund) apresentam probabilidade de informalidade 6,1 pontos percentuais inferior à da categoria de referência ($dy/dx = -0,0606$; $p < 0,001$). Aqueles com ensino médio completo ou superior incompleto (d_Medio) têm probabilidade 20,7 pontos percentuais menor ($dy/dx = -0,2070$; $p < 0,001$). Para os trabalhadores com ensino superior completo ($d_Superior$), o diferencial é ainda mais expressivo: a probabilidade de informalidade é 46,5 pontos percentuais inferior à da categoria de referência ($dy/dx = -0,4654$; $p < 0,001$).

Esses resultados são estatisticamente significativos ao nível de 1% e economicamente relevantes. O diferencial entre a categoria de referência e a de superior completo, de 46,5 p.p., representa uma redução de quase metade na probabilidade de informalidade associada exclusivamente à acumulação de capital humano, controlando por todas as demais características observáveis. Tal magnitude é superior à encontrada por Cunha et al. (2011) para o Brasil como um todo (-37,8 p.p.) e por Mello e Santos (2008) (-37,5 p.p.), sugerindo que o retorno educacional à formalização é mais pronunciado no Nordeste do que nas demais regiões do país, resultado coerente com a maior vulnerabilidade estrutural do mercado de trabalho regional. Assim, a H1 é confirmada pelos dados.

4.2.2 Resultados das Variáveis de Controle

As variáveis de controle individuais apresentam sinais e magnitudes esperados pela literatura. O coeficiente da dummy de sexo masculino (d_sexo) indica que homens têm probabilidade de informalidade 16,1 p.p. inferior à das mulheres ($dy/dx = -0,1608$; $p < 0,001$), resultado coerente com a maior concentração feminina em trabalho doméstico informal. A variável de área de residência (d_urbano) aponta que trabalhadores urbanos têm probabilidade de informalidade 7,4 p.p. menor do que os rurais ($dy/dx = -0,0736$; $p < 0,001$), evidenciando a persistência de maior precariedade ocupacional no campo nordestino.

Com o alargamento do recorte etário para 14 anos, a variável idade linear passa a ser marginalmente significativa ao nível de 10% ($dy/dx = -0,0019$; $p = 0,076$), indicando leve redução da probabilidade de informalidade com o avanço da idade, consistente com a acumulação de experiência e a estabilização dos vínculos formais ao longo da trajetória profissional. O termo quadrático ($Idade^2$) é significativo ao nível de 1% ($dy/dx = 0,000067$; $p < 0,001$), confirmando a relação não linear em U: a informalidade decresce com a idade até um ponto de inflexão e volta a crescer nos trabalhadores mais velhos, possivelmente em razão das dificuldades de reinserção formal. Entre as dummies raciais, apenas as categorias branca ($-0,9$ p.p.; $p < 0,001$) e amarela ($-1,1$ p.p.; $p < 0,05$) diferem estatisticamente da referência (pardo), enquanto preta e indígena não apresentam diferenças significativas neste modelo.

4.2.3 Hipótese H3: Heterogeneidade Estadual

A Tabela 3 sintetiza os efeitos marginais das dummies estaduais, tendo Pernambuco como categoria de referência. Todos os oito estados apresentam coeficientes estatisticamente significativos ao nível de 1%, confirmando a Hipótese H3: existem diferenças estatisticamente significativas na probabilidade de informalidade entre os estados nordestinos, refletindo suas heterogeneidades estruturais.

Tabela 3 – Heterogeneidade estadual: efeitos marginais em relação a Pernambuco

Estado	Efeito marginal (dy/dx)	Relativo a PE	Interpretação
Maranhão	+2,16 p.p.***	Mais informal	Maior informalidade relativa no NE
Piauí	+1,03 p.p.***	Mais informal	Ligeiramente mais informal que PE
Pernambuco	Referência	—	Estado de referência (mais formalizado)
Ceará	−1,80 p.p.***	Menos informal	Menor informalidade que PE
Rio Grande do Norte	−4,81 p.p.***	Menos informal	Significativamente mais formalizado
Paraíba	−2,18 p.p.***	Menos informal	Mais formalizado que PE
Alagoas	−6,28 p.p.***	Menos informal	Maior diferencial formal da amostra
Sergipe	−3,77 p.p.***	Menos informal	Fortemente mais formalizado
Bahia	−2,48 p.p.***	Menos informal	Mais formalizado que PE

*Fonte: Elaboração própria. Nota: *** $p < 0,01$ em todos os coeficientes. Estado de referência: Pernambuco.*

Os resultados revelam uma estrutura heterogênea que merece atenção analítica. Maranhão e Piauí apresentam probabilidades de informalidade superiores à de Pernambuco, respectivamente +2,2 p.p. e +1,0 p.p., o que é coerente com o perfil estrutural desses estados, caracterizados por economias predominantemente agropecuárias, baixa concentração de emprego industrial e menor densidade de grandes empregadores formais (IBGE, 2023). No extremo oposto, Alagoas apresenta o maior diferencial negativo da amostra (−6,3 p.p.), seguida por Rio Grande do Norte (−4,8 p.p.) e Sergipe (−3,8 p.p.). Esses resultados sugerem que, após controle por escolaridade e características individuais, a estrutura produtiva e institucional de cada estado exerce influência autônoma e significativa sobre a probabilidade de informalidade, confirmando que políticas de formalização devem considerar as especificidades territoriais da região.

A escolha de Pernambuco como categoria de referência é sustentada pelos dados: o estado concentra o maior mercado de trabalho formal do Nordeste, com a maior participação relativa da indústria de transformação e do setor de serviços avançados na região (IBGE, 2023),

o que o torna a referência estrutural mais adequada para identificar desvios positivos e negativos na probabilidade de informalidade entre os demais estados.

4.2.4 Efeitos Temporais e o Impacto da Pandemia

Os coeficientes das dummies anuais fornecem evidências sobre a dinâmica da informalidade ao longo do período 2019–2023, tendo 2019 como referência pré-pandêmica. Os anos de 2020 e 2021 não apresentam efeitos estatisticamente significativos sobre a probabilidade de informalidade ($p = 0,360$ e $p = 0,330$, respectivamente), o que pode ser interpretado de duas formas complementares: (i) o choque da pandemia afetou de forma similar tanto o setor formal quanto o informal, sem alterar estruturalmente a probabilidade relativa de inserção informal; ou (ii) os efeitos de composição da amostra, com a saída temporária de trabalhadores informais do mercado durante o isolamento, compensaram o eventual aumento da informalidade para aqueles que permaneceram ocupados.

Os anos de 2022 e 2023, período de recuperação pós-pandêmica, apresentam coeficientes negativos e significativos ao nível de 5% ($-1,6$ p.p. e $-1,5$ p.p., respectivamente), indicando que, na fase de retomada, a probabilidade de informalidade recuou ligeiramente em relação ao padrão pré-pandêmico de 2019. Esse resultado é consistente com os dados agregados do IBGE (2023), que indicam recuperação da formalização no período recente, ainda que insuficiente para reduzir a informalidade estrutural da região. As dummies trimestrais confirmam a sazonalidade esperada: o primeiro e segundo trimestres registram probabilidades marginalmente superiores ao quarto trimestre de referência (respectivamente $+0,6$ p.p. e $+0,6$ p.p.; $p < 0,01$), padrão associado à maior concentração de ocupações temporárias e sazonais no início do ano.

4.3 Síntese dos Resultados e Verificação das Hipóteses

O conjunto dos resultados apresentados permite verificar as três hipóteses formuladas neste estudo. A Hipótese H1, de que maiores níveis de escolaridade reduzem significativamente a probabilidade de inserção no mercado de trabalho informal, é plenamente confirmada pelos dados. O gradiente educacional estimado é negativo, crescente em magnitude e estatisticamente robusto ao nível de 1% para todas as categorias de escolaridade. Os efeitos marginais estimados ($-6,1$ p.p., $-20,7$ p.p. e $-46,5$ p.p. para as três categorias acima da referência) são compatíveis com a teoria do capital humano e com os resultados encontrados por Cunha et al. (2011) e Mello

e Santos (2008) para o Brasil, e revelam que o retorno à formalização associado à educação é especialmente expressivo no Nordeste.

A Hipótese H3, de que existem heterogeneidades estaduais significativas no impacto da escolaridade sobre a informalidade, também é confirmada. Todos os oito estados nordestinos diferem estatisticamente de Pernambuco na probabilidade de informalidade, com amplitude de 8,4 pontos percentuais entre o estado de maior informalidade relativa (Maranhão, +2,2 p.p.) e o de menor (Alagoas, -6,3 p.p.). Essa variação evidencia que a estrutura produtiva estadual e as características do mercado de trabalho local condicionam de forma autônoma e significativa a probabilidade de inserção informal, para além do nível educacional individual.

A verificação da Hipótese H2, de que o efeito da escolaridade sobre a informalidade difere entre áreas urbanas e rurais, requer a estimação de modelos separados por área de residência, o que constitui extensão natural da análise aqui apresentada e está previsto nas etapas subsequentes desta pesquisa. Os resultados do modelo geral, no entanto, já fornecem um indicativo nessa direção: o coeficiente de d_{urbano} (-7,1 p.p.; $p < 0,001$) confirma que trabalhadores rurais apresentam probabilidade substancialmente maior de informalidade, sugerindo que o efeito protetor da escolaridade pode ser condicionado pelo contexto territorial.

CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo estimar o efeito do nível de escolaridade sobre a probabilidade de inserção no mercado de trabalho informal no Nordeste brasileiro, no período de 2019 a 2023, mediante a aplicação de modelos probit aos microdados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua). A escolha do recorte regional justificase pelo caráter persistente e estruturalmente diferenciado da informalidade no Nordeste, que ao longo de todo o período analisado manteve taxas superiores a 50%, expressivamente acima da média nacional, condicionadas tanto pelo déficit histórico de escolaridade da força de trabalho quanto pela composição produtiva da macrorregião, marcada pela predominância de setores de baixa produtividade e elevada concentração de ocupações precárias.

Para operacionalizar a análise, foram utilizados os microdados de todos os trimestres de 2019 a 2023, resultando em uma amostra de 1.005.969 observações de trabalhadores ocupados entre 14 e 65 anos residentes nos nove estados nordestinos. A variável dependente foi construída a partir da combinação de quatro categorias de informalidade, empregados privados sem carteira assinada, trabalhadores domésticos sem carteira, trabalhadores por conta própria sem CNPJ e trabalhadores não remunerados, em conformidade com a definição operacional adotada por Cunha et al. (2011) e Mello e Santos (2008). O modelo probit foi estimado com erros-padrão clusterizados por unidade da federação, garantindo a consistência das inferências diante da estrutura hierárquica dos dados, e os resultados foram reportados como efeitos marginais médios (APE), que traduzem os coeficientes em variações em pontos percentuais na probabilidade de informalidade.

Os resultados confirmam a Hipótese H1: maiores níveis de escolaridade reduzem significativamente a probabilidade de inserção no mercado de trabalho informal no Nordeste. O gradiente educacional estimado é negativo, crescente em magnitude e estatisticamente significativo ao nível de 1% para todas as categorias de escolaridade. Em relação à categoria de referência, trabalhadores sem instrução ou com ensino fundamental incompleto, os efeitos marginais indicam reduções de 6,1 pontos percentuais para aqueles com ensino fundamental incompleto ou médio incompleto, de 20,7 p.p. para os com ensino médio completo ou superior incompleto, e de 46,5 p.p. para os com ensino superior completo. Esses valores são superiores aos estimados por Cunha et al. (2011) e Mello e Santos (2008) para o Brasil como um todo, confirmando que o retorno educacional à formalização é particularmente expressivo no

Nordeste, região onde o diferencial de escolaridade entre formais e informais é historicamente o mais elevado do país (KON, 2004). Tais achados corroboram a teoria do capital humano de Schultz (1961) e Becker (1964): o investimento em educação eleva a produtividade do trabalhador e amplia sua atratividade para o setor formal, reduzindo a probabilidade de inserção precária.

A Hipótese H3, existência de heterogeneidades estaduais significativas na probabilidade de informalidade, também é plenamente confirmada. Todos os oito estados nordestinos diferem estatisticamente de Pernambuco, categoria de referência, ao nível de 1%. Os coeficientes variam de +2,2 p.p. no Maranhão a -6,3 p.p. em Alagoas, revelando uma amplitude de 8,4 pontos percentuais que não pode ser atribuída a diferenças individuais de escolaridade, sexo, idade ou raça, já controladas no modelo, mas sim às especificidades das estruturas produtivas e institucionais de cada estado. Esse resultado tem implicação direta para a política pública: políticas de formalização de alcance regional devem ser complementadas por iniciativas estaduais que considerem as distintas capacidades de absorção de trabalho formal em cada economia local.

A Hipótese H2, de que o efeito redutor da escolaridade sobre a informalidade seria mais pronunciado em áreas urbanas do que em áreas rurais, não pôde ser verificada diretamente no modelo geral estimado, uma vez que a especificação adotada não inclui termos de interação entre escolaridade e área de residência. O coeficiente da dummy de área urbana (-7,4 p.p.; $p < 0,001$) confirma que trabalhadores rurais apresentam probabilidade substancialmente maior de informalidade, mas a comparação formal dos efeitos marginais da educação por área requer a estimação de modelos separados para as subamostras urbana e rural, o que constitui extensão natural desta pesquisa. A verificação de H2 fica, portanto, indicada como agenda de investigação para trabalhos futuros.

Com relação ao período pandêmico, os resultados indicam que os anos de 2020 e 2021 não produziram alterações estatisticamente significativas na probabilidade de informalidade em relação ao padrão de 2019, sugerindo que o choque sanitário não modificou estruturalmente a relação entre escolaridade e inserção informal no Nordeste. Os anos de recuperação pós-pandêmica, 2022 e 2023, apresentam coeficientes negativos e significativos ao nível de 5% (-1,6 p.p. e -1,5 p.p., respectivamente), indicando leve redução da probabilidade de informalidade na fase de retomada, consistente com a recuperação dos vínculos formais observada nos dados agregados do IBGE para o período. Ainda assim, trata-se de variação de

pequena magnitude quando comparada ao diferencial educacional, reforçando que o nível de escolaridade permanece o principal determinante individual da inserção formal no mercado de trabalho nordestino.

As variáveis de controle individuais apresentaram resultados consistentes com a literatura. Homens têm probabilidade de informalidade 16,1 p.p. inferior à das mulheres, reflexo da maior concentração feminina no trabalho doméstico informal. Trabalhadores urbanos têm probabilidade 7,4 p.p. menor que os rurais, evidenciando a persistência da precariedade ocupacional no campo nordestino. A relação entre idade e informalidade mostrou-se não linear: a probabilidade decresce marginalmente com o avanço da idade ($-0,2$ p.p. por ano, $p < 0,10$), mas o termo quadrático positivo e significativo indica que essa trajetória se reverte nos trabalhadores mais velhos, possivelmente em razão das dificuldades de reinserção formal após determinada faixa etária. O modelo apresentou pseudo- R^2 de McFadden igual a 0,1155, valor coerente com os 0,115–0,117 reportados por Cunha et al. (2011), atestando a adequação da especificação adotada.

Algumas limitações do estudo merecem registro. Em primeiro lugar, a análise é de natureza cross-sectional agrupada: embora a utilização de todos os trimestres do período amplie substancialmente o número de observações e permita a inclusão de dummies temporais, o modelo probit estimado não explora a dimensão longitudinal em razão dos microdados utilizados neste estudo serem tratados como cortes transversais independentes, o que impede o controle por heterogeneidade individual não observada. Em segundo lugar, a variável de escolaridade reflete o nível de instrução declarado, não necessariamente a qualidade da educação recebida, dimensão que a PNAD Contínua não permite capturar diretamente. Em terceiro lugar, conforme mencionado, a verificação formal de H2 requer modelos separados por área de residência, ainda não estimados.

Do ponto de vista das políticas públicas, os resultados reforçam a centralidade da educação como vetor de formalização do mercado de trabalho nordestino. O diferencial de 46,5 p.p. na probabilidade de informalidade entre trabalhadores sem instrução e aqueles com ensino superior completo evidencia que a expansão do acesso e da conclusão dos ciclos educacionais, especialmente o ensino médio e o superior, tem potencial significativo de redução estrutural da informalidade na região. Entretanto, os achados também indicam que a educação é condição necessária, mas não suficiente: a heterogeneidade estadual persistente após o controle por escolaridade sugere que a capacidade de absorção de trabalho formal pelo setor produtivo de

cada estado impõe um teto à conversão da escolaridade em formalização. Políticas complementares de diversificação produtiva, atração de investimentos e fortalecimento institucional do mercado de trabalho formal são, portanto, indispensáveis para que os ganhos educacionais se traduzam plenamente em inserção protegida para os trabalhadores nordestinos.

Em síntese, este estudo oferece evidência empírica robusta de que a escolaridade é o principal determinante individual da probabilidade de formalização no mercado de trabalho nordestino, com efeitos marginais superiores aos estimados para o Brasil como um todo e consistentes com a teoria do capital humano. Ao mesmo tempo, os resultados apontam para a irredutivelmente estrutural natureza da informalidade na região: ainda que a educação reduza substancialmente a probabilidade de inserção precária, as heterogeneidades estaduais e a persistência da informalidade mesmo entre trabalhadores de maior escolaridade revelam que a formalização plena do mercado de trabalho nordestino depende, em última instância, de transformações na estrutura produtiva regional que ampliem a capacidade de geração de empregos formais compatíveis com o nível de qualificação crescente da força de trabalho.

REFERÊNCIAS

- CAMARGO, F. S. et al. Informality: a sectoral portrait of Brazilian economy. Munich Personal RePEc Archive (MPRA), Paper No. 54511, 2006. Fundação Instituto de Pesquisas Econômicas, Universidade de São Paulo.
- CUNHA, D. R. et al. Mercado de trabalho no Brasil: características da informalidade. *Revista de Desenvolvimento Econômico*, Salvador, v. 13, n. 24, p. 87-100, dez. 2011.
- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua). Rio de Janeiro: IBGE, 2019–2023. Disponível em: <https://sidra.ibge.gov.br>. Acesso em: 8 jun. 2026.
- IPEA – INSTITUTO DE PESQUISA ECONÔMICA APLICADA. Carta de Conjuntura. Brasília: IPEA, 2022–2023. Disponível em: <https://www.ipea.gov.br/cartadeconjuntura/index.php/tag/pnad-continua/>. Acesso em: 8 jun. 2026.
- KON, A. Diversidades nas condições de informalidade do trabalho brasileiro. In: Encontro da ANPEC, 2004, São Paulo. Anais... São Paulo: PUC/SP, 2004.
- MELLO, R. F.; SANTOS, D. D. Aceleração educacional e a queda recente da informalidade. In: Encontro da ANPEC, Área 12, 2008. Classificação JEL: O17, J24.
- MESQUITA, S. P.; SAMPAIO, L. M. B.; RAMALHO, H. M. B. A oferta de trabalho não agrícola no meio rural: evidências para o Nordeste e Brasil. *Revista de Economia*, v. 11, n. 4, p. 103-131.
- CAVALCANTI, C. O.; JACINTO, P. de A. Efeitos da educação sobre a dinâmica da informalidade no mercado de trabalho: uma análise empírica para o estado do Espírito Santo. *Espaço e Economia – Revista Brasileira de Geografia Econômica*, ano XIII, n. 27, 2024. Disponível em: <https://journals.openedition.org/espacoeconomia/28175>. Acesso em: 8 jun. 2026.
- ULYSSEA, G. Informalidade no mercado de trabalho brasileiro: uma resenha da literatura. Texto para Discussão n. 1070. Rio de Janeiro: IPEA, fev. 2005a.
- ULYSSEA, G. Instituições e a informalidade no mercado de trabalho. Texto para Discussão n. 1096. Rio de Janeiro: IPEA, jun. 2005b.
- VELOSO, F.; BARBOSA FILHO, F. H.; PERUCHETTI, P. Informalidade, educação e mercado de trabalho no Brasil: 1992–2021. Rio de Janeiro: FGV IBRE, 2022a.
- FEIJÓ, J.; VELOSO, F.; BARBOSA FILHO, F. H.; PERUCHETTI, P. Diferencial salarial do ensino superior cai no Brasil, mas nível ainda é elevado e recompensador. **Conjuntura Econômica**, Rio de Janeiro: FGV IBRE, out. 2024. Disponível em: <https://ibre.fgv.br/blog-da-conjuntura-economica/temas/diferencial-salarial-do-ensino-superior-cai-no-brasil-mas-nivel>. Acesso em: maio 2026.
- FEIJÓ, J.; PERUCHETTI, P. Jovens e a armadilha do mercado de trabalho. **Boletim Macro FGV IBRE**, Rio de Janeiro, jul. 2024. Disponível em: <https://ibre.fgv.br/blog-da-conjuntura-economica/artigos/jovens-e-armadilha-do-mercado-de-trabalho>. Acesso em: maio 2026.

VELOSO, F.; BARBOSA FILHO, F. H.; PERUCHETTI, P. Mercado de trabalho no Brasil: evolução, efeitos da pandemia, perspectivas e políticas públicas. Rio de Janeiro: FGV IBRE, 2022b. Disponível em: https://ibre.fgv.br/sites/ibre.fgv.br/files/arquivos/u65/mercado_de_trabalho_e_politicas_publicas_final.pdf. Acesso em: maio 2026.

ULYSSEA, G. Firms, informality, and development: theory and evidence from Brazil. *American Economic Review*, v. 110, n. 5, p. 1513–1553, maio 2020.