
DETERMINANTES DA INFORMALIDADE NO MERCADO DE TRABALHO BRASILEIRO: UMA ANÁLISE BASEADA NA PNAD ANUAL DE 2019 E 2022

*Determinants of informality in the Brazilian labor market: an analysis based on the 2019 and 2022
Annual PNAD*

Susan Schommer

Economista. Doutora em Economia Matemática pelo Instituto de Matemática Pura e Aplicada. Professora
Adjunta do Instituto de Economia da Universidade Federal do Rio de Janeiro. Campus Praia Vermelha,
Av. Pasteur, 250, 22290-902, Rio de Janeiro, RJ, Brasil. susan.schommer@ie.ufrj.br

Catarina Luz Olivares

Economista. Universidade Federal do Rio de Janeiro. Campus Praia Vermelha, Av. Pasteur,
250, 22290-902, Rio de Janeiro, RJ, Brasil. olivares.catarina@gmail.com

Resumo: Este estudo tem por objetivo analisar os determinantes da informalidade no mercado de trabalho brasileiro nos anos de 2019 e 2022, utilizando dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua Anual. O modelo foi estimado através da regressão logística generalizada com o intuito de mensurar como fatores individuais e regionais influenciam a probabilidade de inserção de um indivíduo no mercado informal. Os resultados revelam padrões persistentes, indicando que características como sexo, raça ou cor, idade, condições do domicílio, escolaridade e região continuam influenciando a probabilidade de trabalhar informalmente ao longo do tempo. A análise identificou que a baixa escolaridade e a localização nas Regiões Norte e Nordeste são os fatores mais significativos para a determinação da ocupação informal do trabalhador, ressaltando a importância da implementação de políticas públicas voltadas para a promoção da educação e à redução das disparidades regionais.

Palavras-chave: Mercado de trabalho, informalidade, PNAD, regressão logística.

Abstract: This study aims to analyze the determinants of informality in the Brazilian labor market in the years 2019 and 2022, using data from the Continuous Annual National Household Sample Survey (PNAD). The model was estimated through generalized logistic regression to measure how individual and regional factors influence the probability of an individual entering the informal labor market. The results reveal persistent patterns, indicating that characteristics such as gender, race or color, age, household conditions, education level, and region continue to influence the probability of informal employment over time. The analysis identified that low educational attainment and residence in the North and Northeast regions are the most significant factors determining informal employment, underscoring the importance of public policies aimed at promoting education and reducing regional disparities.

Keywords: Labor Market, informality, PNAD, logistic regression.

1 INTRODUÇÃO

A informalidade no mercado de trabalho é um fenômeno complexo e multifacetado que tem despertado crescente interesse, tornando-se alvo de debates acadêmicos e políticos nas últimas décadas. O presente estudo busca responder à questão: quais são os principais fatores individuais e regionais que determinam a inserção dos trabalhadores brasileiros no mercado de trabalho informal, e se esses fatores exerceram influências diferenciadas nos anos de 2019 e 2022, especialmente após o choque da pandemia de covid-19.

Os estudos sobre a informalidade no mercado de trabalho destacaram-se em períodos de elevação das taxas de trabalho informal, especialmente no que se refere à disparidade salarial entre os setores formal e informal¹. Com a pandemia de covid-19, o interesse por análises das características individuais e regionais do mercado de trabalho informal foi intensificado, visando compreender os impactos diferenciados da pandemia na condição laboral da população. Em uma perspectiva regional, conforme discutido em Costa *et al.* (2023), Borchers *et al.* (2022) e Simões *et al.* (2023), evidencia-se que a residência nas Regiões Norte e Nordeste aumenta a probabilidade de inserção no mercado de trabalho informal, tanto em períodos anteriores quanto durante a pandemia. Assim, questiona-se se essa desigualdade regional se tornou mais acentuada após o impacto inicial da pandemia.

Ao considerar a característica racial, observa-se que os achados variam conforme o período analisado. Duarte (2020; 2021), ao investigar determinantes da informalidade em Recife e São Paulo, com base no Censo de 2010, constatou maior probabilidade de inserção no setor informal para indivíduos negros ou pardos em comparação aos brancos. Em contraste, Tavares *et al.* (2023) observaram que, em 2020, ser negro não afeta significativamente a probabilidade de informalidade em relação aos brancos. Esse efeito pode ser explicado pelo isolamento social de 2020, que gerou uma intensa transição dos ocupados para a inatividade (Costa *et al.*, 2022).

O alto índice de informalidade, que foi de 39,2% da população ocupada no Brasil em 2023, de acordo com Instituto Brasileiro de Geografia e Estatística (IBGE), reforça a importância de investigar as variáveis que contribuem para esse fenômeno. A escolha dos anos de 2019 e 2022 para análise deve-se à proximidade temporal e à possibilidade de investigar alterações significativas, especialmente em decorrência da pandemia de covid-19, além de a coleta de dados ter sido realizada totalmente de forma presencial². Logo, é possível observar o comportamento dos coeficientes ao longo de um período marcado por grandes mudanças no mercado de trabalho, decorrentes das consequências da pandemia.

A hipótese deste estudo é que com a pandemia de covid-19 ocorreram alterações significativas nos determinantes da informalidade no mercado de trabalho brasileiro. Assim sendo, o estudo tem por objetivo analisar os determinantes da informalidade no mercado de trabalho brasileiro nos anos de 2019 e 2022. Para alcançar tal objetivo, propõe-se: 1) identificar os fatores individuais e regionais que influenciam a probabilidade de inserção de um indivíduo no mercado informal; 2) examinar possíveis mudanças significativas nos determinantes da informalidade entre 2019 e 2022, considerando os impactos da pandemia de covid-19; e 3) estimar os coeficientes dos determinantes da informalidade por meio da aplicação de modelos Logit, usando os dados da Pesquisa Nacional por Amostra de Domicílios Contínua Anual (PNAD).

1 Ver Ulysseia (2006) para uma revisão da literatura.

2 Segundo o IBGE (2021), entre março de 2020 e julho de 2021, a coleta foi realizada exclusivamente por telefone, o que reduziu consideravelmente a taxa de aproveitamento da amostra. Apesar disso, verificou-se que as estatísticas não apresentaram viés nos principais indicadores. Para uma discussão mais detalhada sobre o impacto da redução no número de entrevistas na PNAD Contínua e suas implicações para as estatísticas do mercado de trabalho, consulte Corseuil e Russo (2022).

2 A INFORMALIDADE NO MERCADO DE TRABALHO

A definição de informalidade na literatura carece de consenso. Desde a década de 1970, a Organização Mundial do Trabalho (OIT) utiliza o termo setor informal para descrever atividades com características como facilidade de entrada, uso de recursos locais, gestão familiar e operações em pequena escala. Esse termo tornou-se comum nas análises do mercado de trabalho, especialmente em países em desenvolvimento (Cacciamali, 2016).

Vargas (2016) propõe, em seu ensaio, fornecer insumos teóricos para uma compreensão mais ampla da informalidade no trabalho, examinando três conceitos: setor informal, economia informal e trabalho decente. Apesar dos esforços para adaptar esses conceitos à realidade brasileira, persistem imprecisões que dificultam análises precisas do tema. A tentativa de conceitualização não encerra o debate, mas evidencia a complexidade de estabelecer um critério único e preciso para definir a informalidade no mercado de trabalho brasileiro.

Carvalho e Nogueira (2022) analisam o conceito de precariedade e trabalho precário na literatura, ressaltando a importância de distingui-los da informalidade, especialmente em países como o Brasil. O trabalho precário é marcado pela insegurança e falta de proteção, diferindo do trabalho informal, que ocorre fora das regulamentações legais e sem garantias como previdência social ou contratos formais.

Roubaud *et al.* (2020) analisam quatro definições de setor informal, sendo as duas primeiras do IBGE: a primeira inclui trabalhadores sem carteira assinada, autônomos e empregadores que não contribuem para a previdência; e a segunda, aqueles sem Cadastro Nacional de Pessoa Jurídica (CNPJ). A terceira inclui todos os trabalhadores que não contribuem para a previdência e familiares não remunerados, enquanto a quarta abrange empresas com até cinco pessoas. Na PNAD Contínua, as três primeiras definições produzem resultados semelhantes em nível e estrutura da informalidade, enquanto a quarta apresenta menor precisão.

A complexidade em definir o setor informal é evidente, dada a diversidade de abordagens conceituais e metodológicas, assim como a falta de consenso na literatura. A legislação brasileira simplifica a questão ao requerer carteira de trabalho assinada para vínculos empregatícios, associando a informalidade à ausência desse documento. Neste estudo, adota-se a definição oficial do IBGE como *proxy* para trabalho informal, incluindo empregados privados e domésticos sem carteira assinada, empregadores e autônomos (conta própria) sem CNPJ, e trabalhadores familiares auxiliares.

A informalidade no mercado de trabalho é um desafio persistente que impacta milhões de trabalhadores brasileiros. Dados da PNAD Contínua de 2022 indicam uma população ocupada total de 99,6 milhões de pessoas, sendo que cerca de 39 milhões desses trabalhadores se encontram no setor informal. O país tem experimentado mudanças nas últimas décadas em relação à formalização do emprego, mas a informalidade continua a ser uma característica intrínseca do mercado de trabalho brasileiro.

Ulyssea (2006) analisa o aumento da informalidade no Brasil desde a década de 1990, apontando fatores estruturais, institucionais e individuais. No aspecto estrutural, o autor destaca o crescimento do setor de serviços, mais informal que a indústria, como um fator relevante. Entre os aspectos institucionais, as leis trabalhistas rígidas podem restringir contratações formais, enquanto a baixa qualificação dos trabalhadores é um fator individual que dificulta a inserção no mercado formal. Ulyssea questiona se a informalidade é uma escolha ou uma imposição, concluindo que ambos os casos existem: alguns trabalhadores preferem a informalidade pela flexibilidade e autonomia, enquanto outros se veem obrigados a aderir devido à escassez de empregos formais.

Com o intuito de analisar os dados mais recentes, a Tabela 1 apresenta as informações sobre os trabalhadores informais e suas posições ocupacionais nos anos de 2019 e 2022.

Tabela 1 – Trabalhadores informais por posição de ocupação em 2019 e 2022 no Brasil

Posição na ocupação	Ocupados 2019		Ocupados 2022	
	(Mil pessoas)	(%)	(Mil pessoas)	(%)
Empregado no setor privado sem carteira assinada	12.316	31,7	13.570	34,2
Trabalhador doméstico sem carteira assinada	4.382	11,3	4.446	11,2
Empregador sem CNPJ	839	2,2	936	2,4
Conta própria sem CNPJ	19.229	49,5	19.025	48
Trabalhador familiar auxiliar	2.048	5,3	1.691	4,3
Total	38.814	100,0	39.668	100,0

Fonte: Elaboração própria a partir dos dados da PNAD Contínua.

Destaca-se que os trabalhadores por conta própria sem CNPJ constituem a maioria significativa entre os informais, representando 49,5% em 2019 e 48% em 2022 do total de trabalhadores informais no Brasil. Em comparação com 2019, observa-se um acréscimo de 854 mil trabalhadores informais em 2022, sendo esse crescimento atribuído, principalmente, ao aumento do emprego no setor privado sem carteira assinada.

Nesse contexto, a realização de uma análise sobre a informalidade, baseada em um amplo conjunto de dados representativos que capturem a diversidade da população, revela-se uma abordagem crucial para uma compreensão mais aprofundada dos fatores que influenciam a inserção no trabalho informal. Ademais, torna-se essencial examinar as transformações no mercado de trabalho após a pandemia, visando identificar as mudanças estruturais e suas implicações para a informalidade.

Tavares *et al.* (2023) analisaram as probabilidades de mulheres brancas, mulheres negras e homens negros estarem desempregados ou no setor informal durante a pandemia de covid-19, utilizando dados da PNAD 2020 e um modelo Logit. O estudo considerou variáveis como gênero, raça, escolaridade, região, localização do domicílio, idade, filhos menores de cinco anos e responsabilidade pelo domicílio. Os resultados mostraram que mulheres brancas tiveram menor probabilidade de estar na informalidade em comparação aos homens brancos, enquanto mulheres negras e homens negros apresentaram maior propensão ao trabalho informal. Regionalmente, a Região Norte teve a maior taxa de informalidade, seguida pelo Nordeste, Centro-Oeste, Sudeste e Sul.

Borchers *et al.* (2022) analisaram os determinantes do desemprego e da informalidade juvenil no Brasil, focando na faixa etária de 20 a 34 anos, com dados da PNAD de 2012, 2016, 2019 e 2021. Utilizando um modelo Logit, consideraram variáveis como gênero, raça, responsabilidade pelo domicílio, escolaridade e região. Os resultados mostram que níveis educacionais acima do primário estão associados a um menor risco de emprego informal, e que jovens homens, especialmente nas Regiões Norte e Nordeste, têm maior probabilidade de trabalhar sem carteira assinada. Além disso, características como gênero e educação impactam significativamente as taxas de desemprego, com os homens enfrentando maiores riscos. Em geral, jovens não brancos das regiões Nordeste e Norte são os mais propensos ao desemprego e à informalidade.

Costa *et al.* (2023) investigaram os determinantes do acesso ao trabalho informal entre mulheres brasileiras em 2020, utilizando dados da PNAD covid-19. Com um Modelo de Probabilidade Linear, o estudo analisou variáveis como escolaridade, raça, idade, chefia domiciliar e renda. Os resultados indicam que mulheres menos escolarizadas, mais velhas e chefes de domicílio têm maior probabilidade de estar em empregos informais, enquanto aquelas casadas, com menos filhos e residentes em áreas urbanas apresentam menor propensão à informalidade. A educação, experiência e a presença de cônjuge e filhos reduzem significativamente essa probabilidade, e residir nas Regiões Sul, Sudeste e Nordeste diminui a chance de estar no setor informal em comparação às mulheres da Região Norte.

Simões *et al.* (2023) investigaram a vulnerabilidade dos responsáveis pelo domicílio no mercado de trabalho, com foco nas desigualdades de sexo e raça, utilizando dados da PNAD de 2016 a 2023. Através de um modelo Logit, analisaram variáveis como idade, escolaridade, região, situação do domicílio e presença de crianças. Os resultados mostram que a crise econômica pós-2015, agravada pela pandemia, afetou mais intensamente as mulheres responsáveis pelo domicílio, especialmente as negras e pardas. Fatores como escolaridade, idade, residência em áreas urbanas e presença de cônjuge reduziram a probabilidade de informalidade, enquanto residir nas Regiões Norte e Nordeste e ter crianças aumentaram a vulnerabilidade à informalidade.

Uma das principais políticas públicas implementadas para mitigar os impactos da pandemia no mercado de trabalho foi o auxílio emergencial, focado em trabalhadores informais e famílias de baixa renda. Estudos de Russo *et al.* (2022) e Carvalho (2022) destacam que essa medida foi crucial para manter o rendimento das camadas mais baixas, especialmente nas Regiões Norte e Nordeste. Souza (2022) aponta que o auxílio contribuiu para reduzir a disparidade de renda entre negros e brancos, com a diferença na renda média caindo de quase 90% para 70% após sua implementação, embora ainda permaneça elevada.

Tanto Duarte (2020) quanto Duarte (2021), ao analisarem as condições de informalidade na região metropolitana do Recife e no município de São Paulo com dados do Censo de 2010, encontraram que trabalhadores negros e pardos têm maior probabilidade de estar no setor informal ou desempregados em comparação aos brancos. A escolaridade se destaca como uma variável crucial, já que sua elevação reduz a probabilidade de informalidade. Ademais, ambos os estudos apontam que homens têm menor chance de informalidade do que mulheres. Condições domiciliares, como a presença de melhores infraestruturas e o papel de chefe de domicílio, estão associadas a menores chances de informalidade, enquanto um maior número de pessoas no domicílio aumenta essa probabilidade.

3 BASE DE DADOS

3.1 Fonte de dados e variáveis

Os dados utilizados neste estudo foram extraídos da PNAD Contínua, uma pesquisa realizada pelo IBGE que fornece um extenso conjunto de informações sobre a população brasileira. A PNAD Contínua adota um plano amostral estratégico baseado no Sistema Integrado de Pesquisas Domiciliares (SIPD), que define os parâmetros para a seleção da amostra e orienta todo o processo de coleta de dados, garantindo a representatividade geográfica e populacional (IBGE, 2023, p. 7-8).

O SIPD estabelece as bases metodológicas da PNAD Contínua, definindo sua abrangência geográfica, população-alvo e plano amostral. A pesquisa abrange todo o território nacional, exceto áreas especiais, focando nos residentes de domicílios particulares permanentes. A amostra é conglomerada em dois estágios de seleção, estratificada em unidades primárias de amostragem (UPAs) com critérios específicos para garantir representatividade (IBGE, 2023, p. 10).

A amostra mestra, formada por UPAs selecionadas trimestralmente, é renovada a cada cinco trimestres, com a seleção ocorrendo de forma proporcional ao número de domicílios, mantendo a estratificação do SIPD. No segundo estágio, 14 domicílios são escolhidos aleatoriamente em cada UPA. A estratificação considera características como região geográfica e tamanho dos setores censitários, melhorando a eficiência do plano amostral e possibilitando análises mais robustas (IBGE, 2023, p. 10).

A PNAD Contínua adota um esquema de rotação 1-2(5), no qual entrevista um domicílio por um mês e o exclui nos dois meses seguintes, repetindo essa sequência cinco vezes para capturar mudanças nos indicadores trimestrais e manter uma amostra sobreposta. Os pesos, calibrados pelos totais populacionais por sexo e grupos de idade, são fundamentais para ajustar as estimativas

e garantir representatividade, corrigindo desvios amostrais e aprimorando a precisão dos dados (IBGE, 2023, p. 11). Dessa forma, a PNAD Contínua, ao integrar elementos como SIPD, amostra mestra, pesos amostrais e estratificação, estabelece um plano amostral robusto que assegura uma visão abrangente do mercado de trabalho brasileiro.

O modelo proposto tem como objetivo específico identificar os fatores individuais e regionais que afetam a probabilidade de um indivíduo estar inserido no mercado informal, utilizando a regressão logística com dados da PNAD Contínua. A variável dependente, chamada informal, é dicotômica, assumindo os valores binários 0 ou 1, em que 0 indica ausência de atividades informais e 1 a participação em empregos informais, como: empregado no setor privado sem carteira assinada, empregado doméstico sem carteira, empregador sem registro no CNPJ, trabalhador autônomo sem registro e trabalhador familiar auxiliar, categorias adotadas pelo IBGE para calcular a *proxy* da informalidade. Essa escolha de variável permite modelar a probabilidade de um indivíduo pertencer a uma das duas categorias com base nas variáveis explicativas selecionadas.

As variáveis explicativas selecionadas para o modelo baseiam-se no estudo de Borchers *et al.* (2022), no qual incluem: sexo, cor ou raça, região geográfica (Norte, Nordeste, Sul, Sudeste e Centro-Oeste), escolaridade, condição no domicílio e composição familiar. Adicionalmente, foi incorporada a variável que representa experiência no trabalho, composta por idade e idade², conforme utilizada em diversos estudos, como Fernandes (1996), Curi e Menezes-Filho (2004), Ramalho e Silveira (2012) e Cirino e de Lima (2012). Por fim, incluiu-se a variável região metropolitana, considerada nos trabalhos de Pereira *et al.* (2018) e Simões *et al.* (2023). O quadro abaixo apresenta todas as variáveis utilizadas no modelo, juntamente com suas descrições.

Quadro 1 – Variáveis utilizadas no modelo e suas descrições

Variável	Descrição
Informal (Variável Dependente)	Variável categórica que terá valor 1, caso considerado trabalhador informal e 0, caso contrário
Sexo	Variável categórica: Feminino; Masculino
Idade	Variável contínua: Idade do trabalhador
Idade ²	Variável contínua: Idade ao quadrado como <i>proxy</i> de experiência
Cor ou raça	Variável categórica: Branca; Não-branca*
Região	Variável categórica: Norte; Nordeste; Sul; Sudeste; Centro-Oeste
Tipo de Área	Variável categórica: Região Metropolitana ou Região Integrada de Desenvolvimento Econômico; Resto da UF**
Escolaridade	Variável categórica: Sem instrução e menos de 1 ano de estudo; Fundamental incompleto ou equivalente; Fundamental completo e médio incompleto ou equivalente; Médio completo e superior incompleto ou equivalente; Superior completo
Condição no Domicílio	Variável categórica: Responsável pelo domicílio; Outros
Componentes no Domicílio	Variável contínua: Quantidade de pessoas que vivem no mesmo domicílio.

Fonte: Elaboração própria.

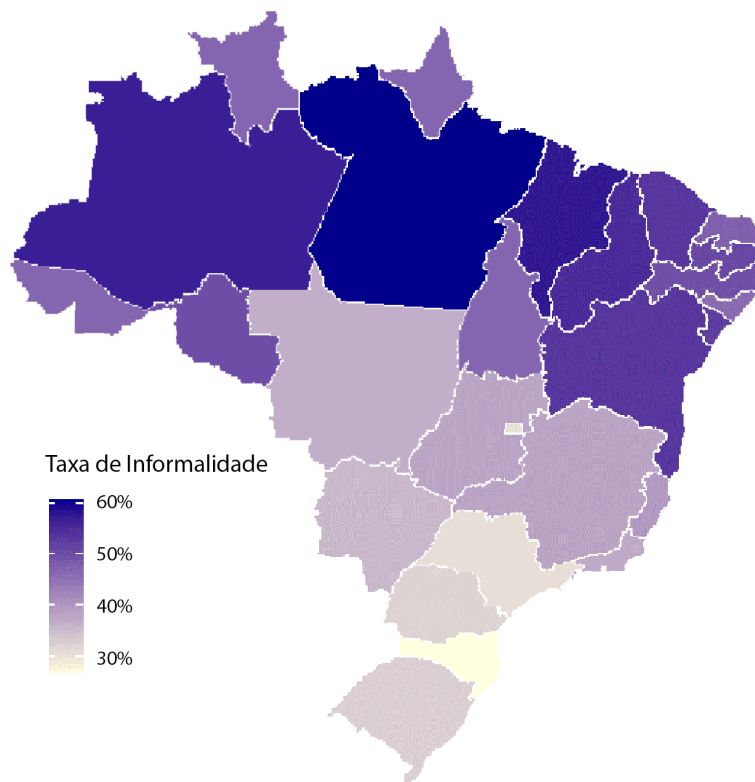
Nota: Todos os dados utilizados foram obtidos na PNAD referente aos anos de 2019 e 2022. *Não-branca inclui preta, amarela, parda, indígena e ignorada. O IBGE pesquisa a cor ou raça da população brasileira com base na autodeclaração. **UF significa unidade federativa.

As análises foram restritas aos indivíduos ocupados, conforme definido pelo IBGE, e àqueles que responderam às perguntas relacionadas às variáveis mencionadas.

3.2 Estatísticas descritivas

Para compreender os fatores que contribuem para a informalidade, é essencial realizar análises estatísticas descritivas dos dados, permitindo identificar as principais características dos trabalhadores informais e suas diferenças em relação aos formais. A Figura 1 apresenta a taxa de informalidade por estado brasileiro, proporcionando uma representação visual da distribuição regional desse fenômeno.

Figura 1 – Mapa do Brasil com as taxas de informalidade por estado



Fonte: Elaboração própria a partir dos dados da PNAD Continua 2022.

Os dados apresentados na Figura 1 indicam que a informalidade é um fenômeno heterogêneo, com características distintas em diferentes regiões do Brasil. A informalidade é mais prevalente nas Regiões Norte e Nordeste do Brasil, onde as taxas de informalidade são de 55,7% e 52,4%, respectivamente. Essas regiões também abrigam os três estados com as maiores taxas: Pará (60,6%), Maranhão (57,7%) e Amazonas (56,3%). A região Sul, por outro lado, apresenta a menor taxa do país, com 30,7%. Os estados com as menores taxas foram em Santa Catarina (26,6%), Distrito Federal (30,3%) e São Paulo (30,8%), evidenciando disparidades regionais.

As seguintes estatísticas descritivas da Tabela 2 são de dados para 2022, que totalizam um tamanho amostral (n) de 163.942 observações e um tamanho populacional (N) de 99,6 milhões de trabalhadores, com um número de observações menor do que as usadas no modelo de 2019, que contam com 189.811 observações em seu tamanho amostral e 94 milhões de trabalhadores ocupados compondo seu tamanho populacional, dos quais aproximadamente 38,7 milhões são informais.

A Tabela 2 apresenta todas as variáveis e suas respectivas categorias em dois cenários: um para a população ocupada em atividades informais e outro para a totalidade dos trabalhadores, incluindo formais e informais. As análises revelam que a idade média dos trabalhadores informais é de 39,6 anos, enquanto a de todos os trabalhadores é de 39,1 anos, indicando um envelhecimento geral da força de trabalho. Apesar da diferença de 0,5 anos, essa variação é mínima.

Tabela 2 – Médias e Proporções das variáveis utilizadas no modelo de 2022 no Brasil.

Variável		Informais	População total
Idade (média)		39,6	39,1
Idade ² (média)		1702,8	1778,7
Componentes no domicílio (média)		3,4	3,3
Sexo	Homem	0,59	0,57
	Mulher	0,41	0,43
Cor ou raça	Não branca	0,63	0,55
	Branca	0,37	0,45
Região	Norte	0,11	0,08
	Nordeste	0,3	0,23
	Sul	0,12	0,16
	Sudeste	0,39	0,44
	Centro-Oeste	0,08	0,09
Tipo de área	RM+RIDE*	0,38	0,43
	Resto da UF	0,62	0,57
Escolaridade	Sem instrução e menos de 1 ano de estudo	0,04	0,02
	Fundamental incompleto ou equivalente	0,31	0,2
	Fundamental completo ou médio incompleto ou equivalente	0,19	0,14
	Médio completo ou superior incompleto ou equivalente	0,36	0,42
	Superior completo	0,1	0,22
Condição no domicílio	Responsável	0,45	0,45
	Outros	0,55	0,55

Fonte: Elaboração própria a partir dos dados da PNAD Continua 2022.

Nota: *RM significa região metropolitana e RIDE região integrada de desenvolvimento.

Além disso, há pequenas diferenças entre os trabalhadores informais e a população total em relação ao número médio de membros no domicílio e às proporções de homens e mulheres. No entanto, a diferença na variável cor ou raça é significativa: entre os trabalhadores informais, 63% são não brancos, em comparação a 55% entre todos os trabalhadores, enquanto 37% dos informais são brancos, contra 45% na população total.

Essas disparidades indicam uma relação entre cor ou raça e a informalidade no mercado de trabalho brasileiro. Os não brancos totalizam cerca de 29 milhões de ocupados formais e 25 milhões de informais, enquanto os brancos apresentam números semelhantes entre formais (29 milhões) e uma proporção consideravelmente menor entre informais (14 milhões). Essas diferenças ressaltam a relevância da variável racial no modelo, considerando as profundas implicações históricas na

formação econômica e social do Brasil³, tornando a variável racial uma característica relevante a ser incorporada ao modelo.

Simões *et al.* (2023) evidenciam que a raça se apresenta como um fator importante na configuração da informalidade no mercado de trabalho, sendo mais significativa que o sexo em sua influência sobre a força de trabalho potencial. Essa constatação revela uma situação de vulnerabilidade acentuada para mulheres negras e pardas, que são as principais responsáveis por domicílios, e enfrentam maior exposição a condições precárias de emprego, bem como a um risco elevado de pobreza e vulnerabilidade monetária.

Com relação às regiões do país, também é possível observar consideráveis diferenças nas proporções de informalidade. Conforme mencionado anteriormente, a disparidade torna-se mais evidente ao constatar que as regiões Norte e Nordeste são as únicas que registram uma maior proporção de trabalhadores informais em relação às outras regiões e ao total de trabalhadores. Essa mudança é notável, passando de 8% para 11% no caso da Região Norte e de 23% para 30% no caso da Região Nordeste. Em contrapartida, as Regiões do Sudeste, Sul e Centro-Oeste apresentam menor proporção de informais com relação ao seu total. Esse fenômeno pode indicar diferenças estruturais nas economias regionais e nas características do mercado de trabalho.

As áreas Região Metropolitana (RM) e Região Integrada de Desenvolvimento do Distrito Federal e Entorno (RIDE), que concentram urbanizações significativas, apresentam uma maior integração econômica e social. Ao analisar os dados, nota-se uma diferença na proporção de trabalhadores informais entre essas áreas e o restante da unidade federativa, que possui uma população maior de trabalhadores. Essa disparidade é ainda mais acentuada no que se refere aos informais e pode ser atribuída a fatores como maior disponibilidade de empregos formais, infraestrutura desenvolvida e acesso a serviços e oportunidades de capacitação. Considerar as particularidades regionais é, portanto, essencial para compreender as nuances específicas de cada área.

Os dados de escolaridade mostram uma tendência clara: à medida que o nível educacional aumenta, a incidência de informalidade diminui. Comparando os níveis de ensino, 31% dos trabalhadores informais possuem até o ensino fundamental incompleto, enquanto essa proporção é de 20% entre todos os trabalhadores. Para aqueles com ensino médio completo e superior incompleto, a proporção de informais é de 36%, contra 42% do total. Além disso, a discrepância na proporção de trabalhadores com ensino superior completo é significativa: apenas 10% entre os informais, em comparação a 22% entre todos os trabalhadores, sugerindo que a educação superior é um fator relevante para a formalização do emprego.

Ao analisar os dados sobre a condição do domicílio dos trabalhadores, categorizados entre responsáveis e outros, nota-se que a proporção entre essas categorias é consistente. O trabalhador responsável é aquele que sustenta a residência, enquanto a categoria outros inclui situações não abrangidas pela primeira. Essa equivalência indica que, independentemente da responsabilidade pelo domicílio, a incidência de informalidade no trabalho não apresenta variações significativas.

Em suma, as estatísticas descritivas apresentadas nesta seção fornecem um panorama geral da informalidade no mercado de trabalho brasileiro. A partir da análise das variáveis mencionadas, é possível delinear algumas hipóteses pertinentes sobre os trabalhadores informais, incluindo a existência de desigualdades regionais, a vulnerabilidade associada a questões raciais e de gênero, bem como a influência de níveis mais baixos de escolaridade, sendo essencial conduzir investigações mais aprofundadas com rigor estatístico para gerar evidências mais consistentes.

3 Para um estudo profundo sobre como a discriminação racial impacta as desigualdades sociais e econômicas no Brasil, ver Hasenbalg (1979).

4 METODOLOGIA

Na metodologia adotada, optou-se pelo emprego do modelo econométrico de regressão logística, uma escolha comum para a modelagem de variáveis dicotômicas, ou seja, aquelas que apresentam apenas dois possíveis resultados. O modelo tem como objetivo prever a probabilidade de um resultado binário (geralmente denotado como 0 ou 1) com base em uma ou mais variáveis preditoras. Hosmer *et al.* (2013), representam como $Pr(Y = 1 \vee x) = \pi(x)$ a probabilidade condicional de o resultado ser igual a 1, dado o vetor de k variáveis independentes $x' = (x_1, x_1, \dots, x_k)$. O modelo de regressão logística pode ser escrito a partir da forma:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad , \quad (1)$$

em que a função logit $g(x)$ é definida como:

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad , \quad (2)$$

em que, os $\beta_0, \beta_1, \dots, \beta_k$ são os coeficientes das variáveis independentes ou preditoras.

O termo da função $\ln()$ é chamado de chances ou *odds* (probabilidade de ocorrência do evento dividida pela probabilidade de não ocorrência do mesmo evento) e com o logaritmo é chamado de *log* das chances.

Para interpretar os coeficientes da regressão logística é preciso analisar a diferença nas probabilidades logarítmicas do resultado associado a uma diferença de 1 unidade no x_j , no qual mostraremos que é o exponencial e^{β_j} chamado de razão de chances (*odds ratio*). Para ver isso basta aplicar a função exponencial em ambos os lados da equação (2) e obter $odds = \left(\frac{\pi(x)}{1 - \pi(x)}\right) = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}$. Para um preditor x_j , a proporção das probabilidade em $x_j + 1$ para as probabilidades de x_j (uma diferença de uma unidade) pode ser expressa como a seguinte razão $\frac{odds_{x_j+1}}{odds} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_j (x_j+1) + \dots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k}}$. Aplicando a regra $\frac{e^a}{e^b} = e^{a-b}$ temos que vários termos são removidos e assim obtemos a razão de chances ou *odds ratio* (OR):

$$OR_j = \frac{odds_{x_j+1}}{odds} = e^{\beta_j(x_j+1) - \beta_j x_j} = e^{\beta_j} \quad . \quad (3)$$

No modelo em estudo o vetor das variáveis independentes é definido por $x' = (constante, idade, idade^2, compdom, sexo, raça, região, área, escolaridade, conddom)$ em que *compdom* significa componentes do domicílio e *conddom* significa a condição do domicílio, conforme descrito no Quadro 1. As variáveis qualitativas possuem m categorias, sendo então necessário estabelecer a sua categoria base e considerar no modelo apenas as $m-1$ categorias dessa variável. A categoria base foi escolhida pela sua maior frequência na amostra e são as seguintes: homem, não brancos, Região Sudeste, resto da UF, escolaridade médio completo ou superior incompleto e outros responsáveis pelo domicílio.

A variável dependente desse estudo é uma *proxy* para o caso de o trabalhador fazer ou não parte do mercado de trabalho informal, assumindo o valor 1 quando o indivíduo está no setor informal

e 0 quando não está. O modelo permite a avaliação das probabilidades de sucesso e fracasso e a aplicação do Método de Máxima Pseudo-Verossimilhança (MPV) para estimar os coeficientes das variáveis independentes.

A representação da variável Y_i como 1 para a presença no setor informal e 0 para a ausência simplifica a análise, tornando-a coerente com a modelagem de regressão logística. A escolha por um modelo de regressão logística binária ponderada⁴, destaca-se pela capacidade de lidar adequadamente com o desenho complexo da amostra. Ao utilizar dados de dois anos distintos, 2019 e 2022, mas mantendo constantes as variáveis utilizadas, possibilitamos a observação de possíveis mudanças nos coeficientes ao longo do período.

A aplicação da regressão logística no presente estudo, conforme descrita por Hosmer *et al.* (2013), proporciona uma ferramenta adequada para modelar a relação entre as variáveis preditoras e a variável dependente binária, permitindo a estimativa das chances de inserção no mercado informal. Contudo, a validade e precisão das estimativas dependem de uma adequada consideração do plano amostral, dada a complexidade dos dados utilizados. Nesse sentido, Silva *et al.* (2002) ressaltam a importância de incorporar corretamente a estrutura do plano amostral nos procedimentos de análise estatística, especialmente em pesquisas como a PNAD, que utilizam amostras complexas. Essa integração é essencial para garantir que os resultados obtidos pela regressão logística reflitam de maneira precisa as características da população e não sejam enviesados por possíveis distorções amostrais. Assim, no tópico seguinte, detalha-se o plano amostral seguido neste estudo.

4.1 Declaração do plano amostral

Diferentemente de amostras simples, os dados da PNAD não podem ser tratados como observações independentes e identicamente distribuídas. Ignorar esses aspectos na análise estatística pode levar a resultados incorretos para estimativas pontuais, desvios padrão e níveis de significância. A atenção crescente a esse tema resultou em recursos que facilitam a aplicação de técnicas de análise capazes de incorporar a complexidade dos planos amostrais (Silva *et al.*, 2002).

Seguindo as etapas descritas por Lumley (2010, p. 20)⁵, iniciou-se o processo com a descrição da estrutura do plano amostral⁶, a qual é complexa, e busca considerar a diversidade da população brasileira. Após a importação dos microdados, foram utilizadas algumas variáveis para a declaração do plano amostral, dentre elas, a variável da unidade primária de amostragem (UPA), que permite a identificação única de cada unidade amostral, composta por informações como o código da Unidade Federativa (UF), número sequencial e dígito verificador.

A estratificação, baseada na variável Estrato, considera o código da UF para garantir que a amostra reflita a distribuição geográfica do país. Além disso, são utilizadas variáveis de projeção da população, refletindo as características demográficas da população em diferentes níveis geográficos, bem como por sexo e idade. Por último, os pesos do domicílio e das pessoas são necessários para garantir que a amostra seja representativa, visando corrigir vieses amostrais ao considerar a probabilidade de seleção e ajustes para não resposta.

4.2 Método de estimação considerando o plano amostral

A estimativa do modelo de regressão logística generalizada, considerando as características específicas do desenho amostral adotado, foi feita através da função de Máxima Pseudo-Verossi-

4 O modelo foi estimado utilizando a linguagem de programação R (R Core Team, 2024) para a estimativa da regressão logística foi usada a função *svyglm* do *software* R (Lumley; Gao; Schneider, 2024).

5 Criador do pacote *survey* no *software* R. Esse pacote é utilizado para o tratamento dos dados e estimativa do modelo.

6 Implementado através da função *svydesign* do *software* R (Lumley; Gao; Schneider, 2024).

milhança (MPV)⁷. Essa função assemelha-se à da Máxima Verossimilhança (MV), com a principal distinção de considerar os pesos e plano amostral na estimativa dos parâmetros. Tal abordagem possibilita que, ao aplicar o método de MPV ao invés do usual de MV, lide eficazmente com dados provenientes de amostragens complexas, integrando elementos como estratificação, conglomerados e ponderação.

O cerne da MPV reside na maximização de uma função de verossimilhança modificada, conhecida como pseudo-verossimilhança, construída de forma a incorporar as características do desenho amostral. Dessa forma, a aplicação do método de MPV proporciona estimativas que consideram a estrutura amostral na avaliação da variância dos parâmetros. Em situações em que os dados são provenientes de pesquisas amostrais complexas, como a PNAD, a obtenção de estimadores de MPV e da respectiva matriz de covariância para os parâmetros em modelos de regressão logística oferece estimativas robustas (Pessoa; Silva, 1998).

Seguindo a discussão apresentada por Pessoa e Silva (1998), que adotam técnicas comumente usadas para o tratamento de amostras complexas, seja $U = \{1, \dots, N\}$ o conjunto de N elementos de uma população finita U e $s = \{i_1, \dots, i_n\} \subset U$ uma amostra com n elementos. Tem-se que cada observação i está associada a um valor y_i e a um vetor x composto por variáveis explanatórias ou independentes. Chamaremos Y_i e X_i como a variável e o conjunto de fatores que geram y_i e x_i respectivamente, para $i \in U$. Então, o modelo de regressão logística pode ser expresso da seguinte forma:

$$f(y_i \vee x_i, \beta) = [p(x_i' \beta)]^{y_i} [1 - p(x_i' \beta)]^{1-y_i} \quad , \quad (4)$$

em que $p(x_i' \beta) = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} = \pi(x)$, conforme descrito na equação (1).

A função escore denotada por $u_i(\beta)$ é definida por:

$$u_i(\beta) = \frac{\partial \log(y_i \vee x_i, \beta)}{\partial \beta} = [y_i - p(x_i' \beta)] x_i \quad (5)$$

e, portanto, a equação de verossimilhança é dada por:

$$\sum_{i \in U} u_i(\beta) = \sum_{i \in U} [y_i - p(x_i' \beta)] x_i = 0 \quad . \quad (6)$$

As estimativas do vetor de coeficientes β são calculadas pelo método de MPV, resolvendo a equação:

$$\sum_{i \in s} w_i u_i(\beta) = \sum_{i \in s} w_i [y_i - p(x_i' \beta)] x_i = 0 \quad , \quad (7)$$

em que w_i é o peso da i -ésima observação amostral.

⁷ Devido a Binder (1983).

Neste contexto, a abordagem do MPV busca encontrar os valores dos coeficientes β que maximizam a pseudo-verossimilhança, considerando os pesos w_i para lidar com a complexidade do desenho amostral. Em essência, ela representa a condição na qual as diferenças entre as respostas observadas e as probabilidades preditas, ajustadas pelos pesos e variáveis explicativas, anulam-se, indicando um ajuste apropriado do modelo aos dados da PNAD.

Portanto, o método adotado para realizar as estimativas e calcular as variâncias do modelo reflete uma abordagem que considera a representatividade das unidades amostrais por meio de elementos fundamentais, como estratos, conglomerados e pesos de amostragem. A utilização da maximização da pseudo-verossimilhança, em conjunto com a ponderação pelos pesos amostrais, assegura que as estimativas dos coeficientes se ajustem de maneira precisa aos dados, levando em conta a complexidade do desenho amostral. Adicionalmente, o cálculo dos desvios padrões dos coeficientes, por meio do método de linearização, contribui para uma avaliação precisa da variabilidade das estimativas.

Conforme Pessoa e Silva (1998), o método MPV apresenta desvantagens, especialmente pela necessidade de informações detalhadas sobre a amostra, como estratos, conglomerados e pesos amostrais. A ausência ou incorreta incorporação dessas informações pode resultar em inferências equivocadas. Ademais, as propriedades dos estimadores MPV são pouco conhecidas para pequenas amostras, o que compromete a confiabilidade das análises em subdomínios reduzidos. Também há limitações na aplicação de métodos tradicionais de diagnóstico e testes de inferência, como gráficos de resíduos e Razões de Verossimilhança, dificultando a avaliação e a interpretação dos resultados em certas situações.

Em relação a essa última limitação, Pessoa e Silva (1998) indicam o uso da estatística Wald como alternativa mais adequada para testes de inferência. O teste de Wald é uma abordagem estatística comumente utilizada para avaliar a significância estatística dos parâmetros estimados em modelos de regressão. Esse teste é particularmente empregado em modelos de regressão logística, inclusive para pesquisas amostrais complexas, como é o caso da PNAD.

A realização do teste visa verificar se cada coeficiente é estatisticamente igual a zero, testando a seguinte hipótese nula para cada parâmetro: $H_0: \beta_j = 0$ onde, $j = 0, \dots, k$. Em outras palavras, o teste de Wald permite avaliar se as variáveis consideradas no modelo têm impacto significativo na variável dependente.

5 RESULTADOS E DISCUSSÃO

A Tabela 3 apresenta os resultados das estimativas do modelo econométrico para os anos de 2019 e 2022. Para facilitar a interpretação, foi calculada a Razão de Chances ou *Odds Ratio* (OR), descrita na equação (3).

Tabela 3 – Resultados do modelo logístico

Variável	2019		2022	
	Coeficiente OR	Erro padrão [#]	Coeficiente OR	Erro padrão [#]
Mulher	1,2***	0,01	1,18***	0,02
Brancos	0,96*	0,02	0,94***	0,02
Responsável pelo domicílio	1,08***	0,02	1,1***	0,02
Idade	0,9***	0,003	0,9***	0,003
Idade ²	1,001***	0,00004	1,001***	0,00004
Sem instrução e menos de 1 ano de estudo	5,31***	0,05	4,02***	0,05

Variável	2019		2022	
	Coefficiente OR	Erro padrão [#]	Coefficiente OR	Erro padrão [#]
Fundamental incompleto ou equivalente	3,6***	0,02	3,36***	0,02
Fundamental completo ou médio incompleto ou equivalente	2,15***	0,02	2,15***	0,02
Superior completo	0,47***	0,03	0,47***	0,03
Norte	2,29***	0,03	2,18***	0,03
Nordeste	1,92***	0,02	1,87***	0,03
Sul	0,79***	0,03	0,76***	0,03
Centro-Oeste	1,08**	0,03	1,01	0,03
RM+RIDE	0,94***	0,02	0,93***	0,02
Componentes no domicílio	1,03***	0,01	1,0003	0,01
Constante	2,7***	0,07	3,35***	0,07
Número de observações	189.811		163.942	

Fonte: Elaboração própria a partir dos dados da PNAD Continua.

Nota: # Erro Padrão robusto a heterocedasticidade. O ***, ** e * denota significância a 1%, 5% e 10%, respectivamente.

A interpretação das OR sugere como a chance de pertencer ao setor informal é afetada por um aumento unitário nas variáveis independentes, facilitando a interpretação dos resultados em termos de chances proporcionais. Valores de OR superiores a 1 indicam maior probabilidade de informalidade, enquanto valores abaixo de 1 indicam diminuição na razão de chances.

A direção dos coeficientes se mantém constante em ambos os modelos, evidenciando a consistência nas relações entre as variáveis independentes e a probabilidade de informalidade. Características associadas positivamente à informalidade, como ser do sexo feminino ou de certas regiões, mantêm essa tendência nos dois períodos analisados. Ao interpretar os p-valores, a maioria dos coeficientes apresenta significância estatística, embora algumas variáveis mostrem menor significância em um ano específico.

A variável sexo apresenta uma associação positiva e significativa com a informalidade em ambos os modelos, indicando que as mulheres têm aproximadamente 20% e 18% mais chances de ocupar posições informais em relação aos homens, em 2019 e 2022, respectivamente. Esse resultado difere dos encontrados por Tavares *et al.* (2023) para os meses de maio, agosto e novembro de 2020, período em que as mulheres apresentaram menor probabilidade de informalidade em comparação aos homens, atribuída pelos autores à redução das mulheres na força de trabalho em meio ao *lockdown*. Em contraste, para períodos anteriores à pandemia, os achados são consistentes com Fernandes (1996) e Duarte (2020, 2021). Assim, o comportamento observado durante o *lockdown* pareceu atípico, enquanto em 2022, mesmo com uma leve queda em relação a 2019, a desigualdade de gênero ainda persiste.

Os resultados deste estudo indicam que a probabilidade de trabalhadores autodeclarados brancos estarem na informalidade é menor, com uma redução de 4% em 2019 e 6% em 2022, em comparação a trabalhadores de outras raças ou cores. Tais achados estão em consonância com Ramalho e Silveira (2012), que observaram uma chance 16% menor de ingresso no trabalho assalariado informal para brancos em 2010, e com Simões *et al.* (2023), que, em dados de 2016 a 2023, evidenciaram uma intensificação das desigualdades de gênero e raça, onde mulheres e homens pretos e pardos mostraram-se mais vulneráveis do que os brancos, especialmente após 2022. Em contraste, Tavares *et al.* (2023) identificaram que, em 2020, ser homem ou mulher negra não afetou significativamente a probabilidade de estar na informalidade em relação a homens brancos. Similarmen- te, Costa *et al.* (2023) concluíram que, em 2020, ser mulher branca ou não branca não teve impacto

sobre a probabilidade de inserção no setor informal. Esses achados enfatizam a complexidade das relações entre raça, gênero e informalidade, que parecem ter tido um comportamento diferenciado com o choque inicial da crise sanitária.

Costa *et al.* (2022) analisaram as desigualdades relacionadas a sexo, raça ou cor e idade em relação à perda e ao ganho de ocupação no Brasil em 2020. Os autores concluíram que houve uma transição significativa dos ocupados para a inatividade, especialmente entre indivíduos negros. Saboia *et al.* (2020) destacaram que, durante a crise de 2020, houve uma redução acentuada na população ocupada, com maior impacto sobre trabalhadores informais e de menor rendimento. Powell *et al.* (2022) analisaram o impacto da pandemia no mercado de trabalho do Reino Unido, enfatizando que ela afetou desproporcionalmente os grupos étnicos minoritários.

A variável escolaridade tem um papel significativo nas probabilidades de um indivíduo integrar o mercado informal. A categoria sem instrução apresenta um aumento expressivo na razão de chances de participação no setor informal em comparação com aqueles que completaram o ensino médio ou possuem ensino superior incompleto, destacando-se como a que mais contribui para a informalidade. Entre os grupos com diferentes níveis de escolaridade, observa-se uma variação nas chances de informalidade, indicando uma associação proporcional. Para trabalhadores com ensino fundamental completo ou médio incompleto, as chances de estar no setor informal são 215% maiores do que para aqueles com ensino médio completo ou superior incompleto; por outro lado, concluir o ensino superior reduz em 53% a chance de estar no setor informal. Esses achados, de que a probabilidade de estar no setor informal diminui à medida que aumenta a escolaridade, estão em consonância com diversos autores (por exemplo: Ramalho; Silveira, 2012; Pereira; Galvão; Maxir, 2018; Duarte, 2020; Duarte, 2021; Borchers *et al.*, 2022; Simões *et al.*, 2023), tanto para o período antes e durante a pandemia.

O fato de a idade do indivíduo estar negativamente relacionada a estarem em ocupações informais, significa que, de acordo com o modelo quanto maior a idade menor a chance do trabalhador ser informal. No entanto, o coeficiente positivo da idade ao quadrado faz com que a relação entre idade e informalidade tenha uma curvatura em forma de U, sugerindo que, embora a probabilidade de informalidade diminua com a idade, ela pode aumentar novamente para faixas etárias mais avançadas. Essa tendência em forma de U é consistente com achados observados na literatura, como é visto por Fernandes (1996).

O coeficiente positivo associado à variável componentes do domicílio indica que, conforme aumenta o número de pessoas na residência, a razão de chances de o trabalhador estar envolvido em ocupações informais também aumenta. No entanto, em 2022, essa variável teve um valor igual a 1, o que significa que as chances de estar no setor informal ou formal são iguais. Além disso, ser responsável pelo domicílio está positivamente associado à probabilidade de não ter um emprego formal com carteira assinada em comparação com outras condições de responsabilidade domiciliar. Essa relação destaca a importância de considerar não apenas características individuais, mas também ambientes familiares e responsabilidades ao compreender as dinâmicas do mercado de trabalho informal, evidenciando que fatores como responsabilidades familiares e a contribuição para a renda do domicílio podem influenciar as escolhas ocupacionais.

A análise das características regionais revela que indivíduos em regiões metropolitanas ou áreas integradas de desenvolvimento têm menor probabilidade de estarem no setor informal, com taxas de 6% em 2019 e 7% em 2022, em comparação aos que vivem no restante da unidade federativa. Esse resultado é consistente com estudos anteriores, como os de Ulyssea (2006), que mostram menores chances de trabalhadores urbanos estarem na informalidade. A tendência também se aplica às mulheres, com Costa *et al.* (2023) destacando que aquelas em áreas urbanas têm menos chances de participar do mercado informal, e Simões *et al.* (2023) apontando que residir em regiões metropolitanas reduz as chances de informalidade para mulheres pretas e pardas.

Cabe dizer que as possibilidades de residentes em regiões metropolitanas apresentam maior mais chances de estarem no emprego formal deve-se a fatores como maior urbanização, infraestrutura desenvolvida e presença de setores econômicos dinâmicos, como serviços e indústrias de grande porte. A economia formal nessas áreas é fortalecida por uma fiscalização mais rigorosa e regulamentações que incentivam a formalização. Em contraste, as regiões Norte e Nordeste, com maiores taxas de informalidade (ver Figura 1), enfrentam desafios estruturais como baixo dinamismo econômico e menor industrialização, o que favorece a prevalência do trabalho informal.

Observando a variável região, nota-se que Norte e Nordeste apresentam coeficientes bastante superior a um, indicando uma maior probabilidade de os trabalhadores dessas áreas estarem no setor informal, em comparação com aqueles que vivem no Sudeste (categoria base). Em termos de variação percentual, as chances de estar no setor informal para a Região Norte são 118% maiores e para a Região Nordeste são 87% maiores do que as chances para a Região Sudeste em 2022. A Região Centro-Oeste não apresentou diferença significativa em 2022, enquanto a Região Sul foi a única que demonstrou uma razão de chances menor de informalidade em relação ao Sudeste. Essas constatações estão em consonância com as observações feitas na Figura 1, que revela baixos índices de informalidade na Região Sul e altos índices de informalidade nas Regiões Norte e Nordeste.

Os resultados obtidos estão em consonância com Borchers *et al.* (2022), que destacam que jovens residentes nas Regiões Nordeste e Norte têm maior chance de estarem empregados em trabalhos informais. Ademais, Costa *et al.* (2023) apontam que ser mulher e residir nas Regiões Sul, Sudeste ou Nordeste reduz a probabilidade de estar no setor informal em comparação com mulheres residentes no Norte do país. Simões *et al.* (2023) também ressaltam que residir nas Regiões Norte e Nordeste aumenta as chances de informalidade, especialmente entre mulheres pretas e pardas. Esses autores também analisaram os dados relacionando escolaridade e regiões, concluindo que, mesmo com o aumento da escolaridade contribuindo para a redução da inserção no mercado informal, a maior probabilidade de estarem na informalidade é mais expressiva para as mulheres que residem na Região Nordeste.

A Tabela 4 apresenta os p-valores obtidos por meio do teste de Wald, por meio do qual se permite avaliar se as variáveis consideradas no modelo têm impacto significativo na variável dependente.

Tabela 4 – Teste de Wald

Variáveis	Teste de Wald: p-valor	
	2019	2022
Sexo	2E-16	2E-16
Cor ou raça	0,034736	0,000482
Condição do domicílio	4E-07	2E-08
Idade	2E-16	2E-16
Idade2	2E-16	2E-16
Escolaridade	2E-16	2E-16
Região	2E-16	2E-16
Tipo de área	0,000577	0,000204
Componentes no domicílio	3,26E-09	0,960567

Fonte: Elaboração própria a partir dos dados da PNAD Continua.

Os dados da Tabela 4 mostram que a maioria dos p-valores é inferior a 5%, indicando que muitas variáveis são estatisticamente significativas. Sexo, cor ou raça, condição do domicílio, idade, escolaridade, região e tipo de área têm associações significativas com a informalidade em ambos os anos, rejeitando a hipótese nula. No entanto, há variações na importância de certas variáveis ao longo do tempo, como o número de componentes no domicílio, que foi significativo em 2019,

mas não em 2022. Embora a variável cor ou raça apresente diferenças no p-valor entre os anos, ela continua significativa.

A análise geral das variáveis confirma a consistência e robustez do modelo ao captar fatores determinantes para a inserção de trabalhadores no mercado informal, evidenciando o impacto significativo de características individuais e regionais. A análise dos períodos de 2019 e 2022 indica que o sexo feminino, a menor escolaridade e a residência em regiões como o Norte e o Nordeste constituem fatores que contribuem para a informalidade, destacando a vulnerabilidade de determinados grupos. A relação em forma de U entre idade e informalidade sugere que trabalhadores mais jovens e mais velhos tendem a estar mais inseridos no mercado informal, enquanto a responsabilidade domiciliar e o número de componentes no domicílio elevam essa probabilidade. Apesar de algumas variáveis, como componentes no domicílio, apresentarem flutuações na significância estatística entre os anos, os principais fatores determinantes para a informalidade não foram alterados de forma distinta pela pandemia de covid-19.

6 CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo analisar os determinantes da informalidade no mercado de trabalho brasileiro. Especificamente, buscou-se examinar as influências das características individuais e regionais sobre o ingresso no setor informal, bem como identificar possíveis mudanças nesses determinantes entre 2019 e 2022. Para tal propósito, estimou-se um modelo Logit, utilizando dados da PNAD Contínua de 2019 e 2022. A análise parte do pressuposto de que as características individuais e regionais dos trabalhadores são importantes para compreender as dinâmicas do mercado de trabalho informal.

Os principais resultados indicam a relevância das características individuais e regionais como determinantes da informalidade. Entre as características individuais, observou-se que o risco de inserção no setor informal é maior para mulheres em comparação aos homens. O nível educacional destacou-se como uma variável significativa, contribuindo para a redução da informalidade, em consonância com a literatura existente. As variáveis idade e idade² apresentaram um formato em U, sugerindo maior risco de informalidade entre os mais jovens e os mais velhos. Quanto às características relacionadas ao domicílio, a condição de responsável pelo domicílio aumentou a probabilidade de inserção na informalidade. De modo geral, essas variáveis mantiveram padrões semelhantes entre 2019 e 2022, com exceção da educação nos níveis sem instrução e fundamental incompleto, que apresentaram uma queda no risco de informalidade, possivelmente associada ao aumento da escolaridade na população brasileira.

No âmbito regional, os resultados indicam que indivíduos residentes em regiões metropolitanas ou áreas integradas de desenvolvimento têm menor probabilidade de inserção no setor informal em comparação àqueles que vivem em outras áreas da mesma unidade federativa, sem diferenças relevantes entre 2019 e 2022. Considerando as cinco grandes regiões brasileiras, verificou-se que os residentes das Regiões Norte e Nordeste apresentam maior probabilidade de inserção na informalidade em relação à Região Sudeste, um resultado amplamente corroborado pela literatura. Apesar de certa melhora de 2019 para 2022 — com uma redução de 129% para 118% na Região Norte e de 92% para 87% no Nordeste —, o nível de informalidade permanece elevado, refletindo persistentes desigualdades regionais no mercado de trabalho. A Região Sul mostrou menor chance de inserção no setor informal, enquanto a Centro-Oeste não apresentou diferença significativa em relação à Sudeste em 2022.

Quanto aos testes de significância, a maioria das variáveis do modelo demonstrou ser estatisticamente significativa, indicando a relevância das características individuais e regionais na explicação da informalidade. Outro ponto relevante é a coerência observada nas direções dos coeficientes, indicando que certas dinâmicas do mercado de trabalho informal se mantêm relativamente

estáveis. Assim, a hipótese de que a pandemia de covid-19 teria provocado alterações significativas nos determinantes da informalidade no mercado de trabalho brasileiro não foi confirmada. Observa-se, no entanto, uma leve redução nas desigualdades educacionais e regionais, embora insuficiente para alterar de forma significativa a estrutura do mercado de trabalho no Brasil antes do choque da pandemia.

Diante dessas descobertas, embora as limitações inerentes aos dados da PNAD Contínua restrinjam a generalização dos resultados, as variáveis selecionadas desempenham um papel relevante ao oferecer um panorama das relações entre características individuais, familiares e regionais e a presença no setor informal. Além disso, é notável que as variáveis de escolaridade e região se destacam como fatores preponderantes para a informalidade, sugerindo a necessidade de políticas públicas direcionadas para a melhoria da educação e a diminuição das disparidades econômicas regionais, em especial para o Norte e Nordeste.

REFERÊNCIAS

- BINDER, D. A. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, v. 51, n. 3, p. 279-292, 1983.
- BORCHERS, J.; ROSALEM, L.; LEITE, T.; ARAÚJO, E. Análise dos determinantes do desemprego e da informalidade juvenil no brasil (2012-2021). In: ENCONTRO DE ECONOMIA DA REGIÃO SUL, 25. **Anais...** Porto Alegre: ANPEC, 2022. Disponível em: https://www.anpec.org.br/sul/2022/submissao/files_I/i7-225b94780b62341ff3208a1f9b0e005d.pdf. Acesso em: 9 out. 2024.
- CACCIAMALI, M. Globalização e processo de informalidade. **Economia e sociedade**, v. 9, n. 1, p. 153–174, 2016.
- CARVALHO, S. S. de. Os efeitos da pandemia sobre os rendimentos do trabalho e o impacto do auxílio emergencial: os resultados dos microdados da PNAD COVID-19 de novembro de 2020. In: SILVA, S. P.; CORSEUIL, C. H.; COSTA, J. S. (org.). **Impactos da pandemia de COVID-19 no mercado de trabalho e na distribuição de renda no Brasil**. Brasília: Instituto de Pesquisa Econômica Aplicada (IPEA), 2022. p. 505-528.
- CARVALHO, S. S. de; NOGUEIRA, M. O. O trabalho precário e a pandemia: os grupos de risco na economia do trabalho. In: SILVA, S. P.; CORSEUIL, C. H.; COSTA, J. S. (org.). **Impactos da pandemia de COVID-19 no mercado de trabalho e na distribuição de renda no Brasil**. Brasília: Instituto de Pesquisa Econômica Aplicada (IPEA), 2022. p. 101-124.
- CIRINO, J. F.; DE LIMA, J. E. Diferenças de rendimento entre as regiões metropolitanas de Belo Horizonte e Salvador: uma discussão a partir da decomposição de Oaxaca-Blinder. **Revista econômica do Nordeste**, v. 43, n. 2, p. 371–390, 2012.
- CORSEUIL, C. H.; RUSSO, F. M. A redução no número de entrevistas na PNAD contínua durante a pandemia e sua influência para a evolução do emprego formal. In: SILVA, S. P.; CORSEUIL, C. H.; COSTA, J. S. (org.). **Impactos da pandemia de COVID-19 no mercado de trabalho e na distribuição de renda no Brasil**. Brasília: Instituto de Pesquisa Econômica Aplicada (IPEA), 2022. p. 583-598.
- COSTA, M. C. R.; ALMEIDA, A. C.; BRAGA, C. A. S. Determinantes do trabalho informal das mulheres durante o período de pandemia da COVID-19. **Revista Brasileira de Estudos Regionais e Urbanos**, v. 17, n. 4, p. 606–629, 2023.

COSTA, J. S.; BARBOSA, A. L. N. H.; HECKSHER, M. Desigualdades no mercado de trabalho e pandemia da COVID-19. In: SILVA, S. P.; CORSEUIL, C. H.; COSTA, J. S. (org.). **Impactos da pandemia de COVID-19 no mercado de trabalho e na distribuição de renda no Brasil**. Brasília: Instituto de Pesquisa Econômica Aplicada (IPEA), 2022. p. 307-334.

CURI, A. Z.; MENEZES-FILHO, N. A. Os determinantes das transições ocupacionais no mercado de trabalho brasileiro. In: XXXII ENCONTRO NACIONAL DE ECONOMIA. **Anais...** João Pessoa: Anpec, 2004. Disponível em: <https://www.anpec.org.br/encontro2004/artigos/A04A141.pdf> Acesso em: 22 jan. 2024.

DUARTE, L. B. Determinantes da informalidade e desemprego no mercado de trabalho da região metropolitana do Recife no período de 2010. **Revista de Estudos Sociais**, v. 22, n. 45, p. 37–62, 2020.

_____. Utilização do modelo Logit para analisar os determinantes do desemprego e informalidade no município de São Paulo. **Revista Estudo & Debate**, v. 28, n. 1, 2021.

FERNANDES, R. Mercado de trabalho não-regulamentado: participação relativa e diferenciais de salários. **Pesquisa e Planejamento Econômico**, v. 26, n. 3, p. 417-442, 1996.

HASENBALG, C. **Discriminação e desigualdades raciais no Brasil**. Rio de Janeiro: Graal, 1979.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied Logistic Regression**. EUA: John Wiley & Sons, 2013.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD). Notas técnicas - Versão 1.15**. Rio de Janeiro: IBGE, 2023.

_____. **Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD): Dicionário de variáveis**. 2022.

_____. **Pesquisa Nacional por Amostra de Domicílios Contínua – PNAD Contínua: Nota técnica 03/2021**. Nota Técnica 03/2021, 2021.

LUMLEY, T. **Complex Surveys: a guide to analysis using R**. Hoboken, New Jersey: John Wiley & Sons, 2010.

LUMLEY, T.; GAO, P.; SCHNEIDER, B. survey: analysis of complex survey samples. **R package version 4.4-2**, 2024. Disponível em: <https://cran.r-project.org/web/packages/survey/index.html> Acesso em: 28 out. 2024.

PEREIRA, R. M.; GALVÃO, M. C.; MAXIR, H. DOS S. Determinantes do emprego secundário e informalidade: evidências adicionais para o mercado de trabalho brasileiro. **Revista brasileira de estudos de população**, v. 35, n. 3, p. 1–23, 2018.

PESSOA, D. G. C.; SILVA, P. L. **Análise de Dados Amostrais Complexos**. São Paulo: ABE, 1998.

POWELL, A.; FRANCIS-DEVINE, B.; CLARK, H. Coronavirus: Impact on the labour market. **Commons Library Research Briefing**, 2022.

R Core Team. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2024. Disponível em <https://www.R-project.org/> Acesso em: 30 out. 2024.

RAMALHO, H. M. DE B.; SILVEIRA NETO, R. DA M. A Inserção do migrante rural no mercado de trabalho urbano no Brasil: uma análise empírica da importância dos setores informal e formal. **Estudos econômicos**, v. 42, n. 4, p. 731–771, 2012.

ROUBAUD, F.; RAZAFINDRAKOTO, M.; SABOIA, J.; SIMÕES, A.; NETO, J. H.; BOTELHO, L. **Conceitos, definições e mensuração do trabalho informal no Brasil**. Rio de Janeiro: IE-UFRJ, 2020. (Texto para Discussão n. 031)

RUSSO, F. M.; SILVA, S. P.; CORSEUIL, C. H. Programas federais de manutenção de empregos e garantia de renda no contexto da pandemia em 2020: panorama geral de implementação e cobertura. In: SILVA, S. P.; CORSEUIL, C. H.; COSTA, J. S. (org.). **Impactos da pandemia de COVID-19 no mercado de trabalho e na distribuição de renda no Brasil**. Brasília: Instituto de Pesquisa Econômica Aplicada (IPEA), 2022. p. 485-504.

SABOIA, J.; ROUBAUD, F.; RAZAFINDRAKOTO, M. **A Pandemia e o Mercado de Trabalho no Brasil. Valor Econômico**. São Paulo, 6 de Julho, 2020. Disponível em: <https://www.ie.ufrj.br/images/IE/IEnaMidia/07/Valor%20Econ%C3%B4mico%2006-07%20Jo%C3%A3o%20Saboia.pdf> Acesso em: 16 out. 2024.

SILVA, P. L. DO N.; PESSOA, D. G. C.; LILA, M. F. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. **Ciência & saúde coletiva**, v. 7, n. 4, p. 659–670, 2002.

SIMÕES, A. C. R.; BORGES, C. G. M.; BALBINO, T. de F.; ANTIGO, M. F. Vulnerabilidade do responsável pelo domicílio no mercado de trabalho: uma análise por sexo e por cor para o Brasil entre 2016 e 2023. **Revista de Economia do Centro-Oeste**, v.9, n.1, p. 29-54, 2023.

SOUZA, P. H. G. F. A pandemia de COVID-19 e a desigualdade racial de renda. In: SILVA, S. P.; CORSEUIL, C. H.; COSTA, J. S. (org.). **Impactos da pandemia de COVID-19 no mercado de trabalho e na distribuição de renda no Brasil**. Brasília: Instituto de Pesquisa Econômica Aplicada (IPEA), 2022. p. 473-481.

TAVARES, M. R. G.; COSTA, L. V.; ALMEIDA, A. C. Uma análise sobre as desigualdades raciais e de gênero no mercado de trabalho durante a pandemia de covid-19. **Estudos Econômicos (São Paulo)**, v. 53, n. 4, p. 691–725, 2023.

ULYSSEA, G. Informalidade no mercado de trabalho brasileiro: uma resenha da literatura. **Revista de Economia Política**, v. 26, n. 4, p. 596–618, 2006.

VARGAS, J. O mundo, o Brasil e a informalidade do trabalho: uma abordagem conceitual. **Revista Capital Científico – Eletrônica (RCCe)**, v. 14, n. 3, p. 1-13, 2016.