



Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas
Programa de Pós-Graduação em Economia do Setor Público

**Impacto da presença de crianças em creches no
domicílio sobre a condição do jovem nem-nem:
Uma análise para o Brasil no ano de 2022**

Anderson Henrique Fabião Cavalcanti Lima

Anderson Henrique Fabião Cavalcanti Lima

**Impacto da presença de crianças em creches no domicílio
sobre a condição do jovem nem-nem:
Uma análise para o Brasil no ano de 2022**

Dissertação apresentada ao Programa de Pós-Graduação em Economia do Setor Público (PPESP) do Centro de Ciências Sociais Aplicadas da Universidade Federal da Paraíba (UFPB), como requisito para obtenção do grau de Mestre em Economia do Setor Público.

Orientador: Prof. Dr. Magno Vamberto Batista da Silva

Coorientadora: Prof^ª. Dra. Liédje Bettizaide Oliveira de Siqueira

Catálogo na publicação
Seção de Catalogação e Classificação

L732i Lima, Anderson Henrique Fabiao Cavalcanti.

Impacto da presença de crianças em creches no domicílio sobre a condição do jovem nem-nem : uma análise para o Brasil no ano de 2022 / Anderson Henrique Fabiao Cavalcanti Lima. - João Pessoa, 2025.
61 f. : il.

Orientação: Magno Vamberto Batista da Silva.

Coorientação: Liédje Bettizaide Oliveira de Siqueira.

Dissertação (Mestrado) - UFPB/CCSA.

1. Políticas públicas - Educação. 2. Educação infantil. 3. Geração nem-nem. 4. PNADc/IBGE. 5. Escore de propensão. I. Silva, Magno Vamberto Batista da. II. Siqueira, Liédje Bettizaide Oliveira de. III. Título.

UFPB/BC

CDU 37.014.5(043)



Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas
Programa de Pós-Graduação em Economia do Setor Público

Ata da reunião da Banca Examinadora designada para avaliar o trabalho de dissertação do mestrando **Anderson Henrique Fabião Cavalcanti Lima** submetido para obtenção do grau de mestre em Economia área de concentração em **Economia do Setor Público**.

Aos 16 dias do mês de dezembro do ano dois mil e vinte e quatro, às 15 horas e 0 minutos, na Sala 08, do bloco da pós antiga do CCSA, reuniu-se em cerimônia pública os membros da Banca Examinadora, constituída pelos professores doutores Magno Vamberto Batista da Silva (Orientador), Liédje Bettizaide Oliveira de Siqueira (Co-orientadora), Aléssio Tony Cavalcanti de Almeida (Examinador Interno) e Adriano Firmino Valdevino de Araújo (Examinador Externo), a fim de examinarem o candidato ao grau de mestre em Economia do Setor Público, área de concentração em **Economia do Setor Público**, Anderson Henrique Fabião Cavalcanti Lima (aluno). Iniciando a sessão, o professor Magno Vamberto Batista da Silva (Orientador), na qualidade de presidente da Banca Examinadora, comunicou aos presentes qual o fim da reunião e os procedimentos de encaminhamento da mesma. A seguir, concedeu a palavra ao candidato, para que fizesse oralmente a exposição do trabalho, apresentado sob o título: **“IMPACTO DA PRESENÇA DE CRIANÇAS EM CRECHES NO DOMICÍLIO SOBRE A CONDIÇÃO DO JOVEM NEM-NEM: UMA ANÁLISE PARA O BRASIL NO ANO DE 2022”**. Concluída a exposição, o senhor presidente solicitou que fosse feita a arguição por cada um dos examinadores. Após o que foi concedida a palavra ao candidato para que respondesse e esclarecesse as questões levantadas. Terminadas as arguições, a Banca Examinadora passou a proceder à avaliação e ao julgamento do candidato. Em seguida, o senhor presidente comunicou aos presentes que a Banca Examinadora, por unanimidade, aprovou a dissertação apresentada e defendida com o conceito **APROVADO**, concedendo assim, o grau de **Mestre em Economia do Setor Público** ao mestrando **Anderson Henrique Fabião Cavalcanti Lima**. E, para constar, eu, Magno Vamberto Batista da Silva, lavrei a presente ata, que assino junto com os membros da Banca Examinadora. João Pessoa, 16 de dezembro de 2024.

Prof. Dr. Magno Vamberto Batista da Silva
Orientador

Prof. Dra. Liédje Bettizaide Oliveira de Siqueira
Co-orientadora

Prof. Dr. Aléssio Tony Cavalcanti de Almeida
Examinador Interno

Prof. Dr. Adriano Firmino Valdevino de Araújo
Examinador Externo

Emitido em 16/12/2024

ATA Nº 0/2024 - CCSA - DE (11.01.13.02)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 16/12/2024 21:59)
MAGNO VAMBERTO BATISTA DA SILVA
PROFESSOR DO MAGISTERIO SUPERIOR
1285539

(Assinado digitalmente em 16/12/2024 22:05)
LIEDJE BETTIZAIDE OLIVEIRA DE SIQUEIRA
PROFESSOR DO MAGISTERIO SUPERIOR
1466441

(Assinado digitalmente em 17/12/2024 09:45)
ALESSIO TONY CAVALCANTI DE ALMEIDA
PROFESSOR DO MAGISTERIO SUPERIOR
1813840

(Assinado digitalmente em 17/12/2024 09:37)
ADRIANO FIRMINO VALDEVINO DE ARAUJO
PROFESSOR DO MAGISTERIO SUPERIOR
1412731

Para verificar a autenticidade deste documento entre em <https://sipac.ufpb.br/documentos/> informando seu número: 0, ano: 2024, documento (espécie): ATA, data de emissão: 16/12/2024 e o código de verificação: 3fb6dd44ed

Dedico este trabalho à minha esposa,
Juliana, cujo amor sempre me fortalece e
sustenta. Ao nosso querido filho, Benjamin,
pois sua presença ilumina nosso lar com a
mais pura alegria.

Agradecimentos

Agradeço às pessoas que tornaram possível a realização deste trabalho acadêmico.

Em primeiro lugar, expresso minha sincera admiração e gratidão à minha esposa, e ao nosso filho de nasceu ao longo deste curso. Suas constantes demonstrações de amor, paciência e apoio foram a base que sustentou minha jornada acadêmica. Obrigado por compreenderem as demandas deste desafio e por estarem sempre ao meu lado, oferecendo conforto, motivação e alegria, mesmo nos momentos mais exigentes.

Aos meus orientadores, expresso meu profundo agradecimento pela cuidadosa orientação, pelos valiosos conselhos e pela inspiração ao longo deste processo. Sua dedicação e sabedoria foram fundamentais para o desenvolvimento deste trabalho.

À minha família, em especial aos meus pais e aos meus irmãos, por seu apoio incondicional, incentivo e compreensão ao longo de toda a minha trajetória acadêmica.

Aos amigos e colegas que estiveram presentes, oferecendo palavras de encorajamento, apoio emocional, compartilhando alegrias e desafios, meu sincero agradecimento.

Agradeço também à instituição, pelas oportunidades de aprendizado que tornaram possível a realização deste estudo.

Cada um de vocês teve um papel fundamental nesta jornada e sou imensamente grato por todas as formas de apoio e inspiração que recebi ao longo do caminho.

Resumo

Este estudo investiga o impacto da participação de crianças entre 0 a 5 anos em creche ou pré-escola sobre a condição dos jovens entre 15 e 29 anos que não estudam nem trabalham (nem-nem), residentes no mesmo domicílio. Utilizando dados do segundo trimestre de 2022 da Pesquisa Nacional por Amostra de Domicílios Contínua - PNADc/IBGE, foi aplicado o método de escore de propensão *Propensity Score Matching*, com distância de Mahalanobis e balanceamento por entropia, sob um modelo de escolha discreta binária com a hipótese *logit*. Esses procedimentos garantiram o equilíbrio entre os grupos tratado (domicílios com crianças matriculadas) e controle (domicílios sem crianças matriculadas), minimizando o viés de seleção. Os resultados indicam que a presença no domicílio de crianças na educação infantil reduz, em média, 38,05% a probabilidade de jovens no mesmo lar estarem na condição de nem-nem. Em 2022, 22,3% dos jovens brasileiros estavam fora do mercado de trabalho e do sistema educacional, totalizando 10,9 milhões de pessoas, sendo 63,4% mulheres, muitas vezes responsáveis por cuidados familiares. A análise destaca que o acesso à educação infantil não apenas favorece a inclusão social das crianças, mas também contribui para a permanência de jovens no mercado de trabalho ou na educação, ao liberar tempo de cuidadores, principalmente mães, para investir em suas carreiras e formação. Os achados reforçam que a escolarização infantil gera benefícios familiares de longo prazo, ao preparar crianças para o sucesso educacional e profissional e influenciar positivamente o engajamento econômico dos jovens. A importância de políticas públicas que ampliem o acesso à educação infantil é sublinhada como estratégia para combater a condição nem-nem, promovendo inclusão social e desenvolvimento econômico no Brasil.

Palavras-chave: Nem-nem, Educação Infantil, Escore de Propensão, PNADc, Políticas Públicas.

Abstract

This study investigates the impact of children aged 0 to 5 attending daycare or preschool on the condition of young individuals aged 15 to 29 who are neither in education nor employment (NEET) and reside in the same household. Using data from the second quarter of 2022 from the Continuous National Household Sample Survey (PNADc/IBGE), the Propensity Score Matching method was applied, incorporating Mahalanobis distance and entropy balancing, under a binary discrete choice model with the logit hypothesis. These procedures ensured balance between the treated group (households with enrolled children) and the control group (households without enrolled children), minimizing selection bias. The results indicate that the presence of children enrolled in early childhood education reduces the probability of NEET youth in the same household by an average of 38.05%. In 2022, 22.3% of Brazilian youth were NEET, totaling 10.9 million individuals, 63.4% of whom were women, often responsible for caregiving tasks that limit their educational and professional opportunities. The analysis highlights that access to early childhood education not only promotes children's social inclusion but also contributes to keeping young individuals in education or the labor market by freeing caregivers, especially mothers, to invest in their careers and education. The findings underscore that early childhood education generates long-term family benefits by preparing children for academic and professional success and positively influencing young people's economic engagement. Expanding access to early childhood education is emphasized as a critical public policy strategy to reduce NEET rates, fostering social inclusion and economic development in Brazil.

Keywords: NEET, Early Childhood Education, Propensity Score Matching, PNADc, Public Policy.

Lista de tabelas

Tabela 1 – Composição dos domicílios na base de dados	24
Tabela 2 – Número de casos	29
Tabela 3 – Motivo da condição de ser nem-nem	31
Tabela 4 – Comparação entre os grupos	33
Tabela 5 – Modelo de Regressão Logística Múltipla	34
Tabela 6 – Estatísticas Descritivas do Escore de Propensão	36
Tabela 7 – Métricas de Avaliação da Matriz de Confusão	37
Tabela 8 – Amostra, Pareamento e Balanceamento entre Grupos	39
Tabela 9 – Amostra e Balanceamento entre os Grupos, Distância de Mahalanobis .	43
Tabela 10 – Síntese do balanceamento entre os grupos, Entropia	45
Tabela 11 – Resultado das Regressões	47
Tabela 12 – Resultados das Regressões em percentual	48
Tabela 13 – Sumário antes do balanceamento 1:1 com reposição.	54
Tabela 14 – Sumário após o balanceamento 1:1 com reposição.	55
Tabela 15 – Sumário antes do balanceamento 1:3 com reposição.	56
Tabela 16 – Sumário após o balanceamento 1:3 com reposição.	57
Tabela 17 – Sumário antes do balanceamento 1:1 com reposição e tolerância de 5%. .	58
Tabela 18 – Sumário após o balanceamento 1:1 com reposição e tolerância de 5%. .	59
Tabela 19 – Sumário antes do balanceamento 1:1 com reposição - Distância de Mahalanobis.	60
Tabela 20 – Sumário após o balanceamento 1:1 com reposição - Distância de Maha- lanobis.	61
Tabela 21 – Sumário das Variáveis com Diferenças Ajustadas - Entropia.	62
Tabela 22 – Resultados do Teste de Sensibilidade de Rosenbaum	63

Lista de ilustrações

Figura 1 – Esquema rotacional das entrevistas da PNADc nos domicílios	23
Figura 2 – Fluxo de Decisão: Efeito da presença em creche na probabilidade de ter nem-nem no domicílio	26
Figura 3 – Percentual de domicílios com jovens nem-nem por região	30
Figura 4 – Percentual de nem-nem por sexo em domicílios de acordo com a presença ou não da criança na escola	32
Figura 5 – Gráficos de densidade dos escores de propensão	36
Figura 6 – Gráfico antes e depois do balanceamento - 1:1 com reposição	40
Figura 7 – Gráfico antes e depois do balanceamento - 1:3 com reposição	41
Figura 8 – Gráfico antes e depois do balanceamento - 1:1 com reposição e 5% de tolerância	41
Figura 9 – Gráfico antes e depois do balanceamento - Mahalanobis	43
Figura 10 – Gráfico antes e depois do balanceamento - Entropia	45

Sumário

1	INTRODUÇÃO	13
1.1	Objetivo	16
1.2	Objetivos Específicos	16
2	REVISÃO DA LITERATURA	17
2.1	Literatura aos jovens nem-nem	17
2.2	Literatura à primeira infância	20
3	METODOLOGIA	22
3.1	Modelo empírico	25
3.2	Estratégia Empírica	27
4	RESULTADOS	29
4.1	Análise preliminar	29
4.2	Estimações	32
4.2.1	Regressão Múltipla	33
4.2.2	Propensity Score Matching (PSM)	35
4.2.2.1	Escore Propensão	35
4.2.2.2	Pareamento	37
4.2.2.3	Distância de Mahalanobis	41
4.2.2.4	Entropia	44
4.3	Sensibilidade	46
4.4	Efeito Médio do Tratamento sobre os Tratados	46
5	CONSIDERAÇÕES FINAIS	49
	REFERÊNCIAS	51
	APÊNDICE A – PAREAMENTO POR VIZINHO MAIS PRÓXIMO	54
	APÊNDICE B – DISTÂNCIA DE MAHALANOBIS	60
	APÊNDICE C – ENTROPIA	62
	APÊNDICE D – SENSIBILIDADE DE ROSENBAUM	63

1 Introdução

Há algumas décadas, o tema primeira infância começou a ganhar destaque na literatura em diversas áreas do conhecimento. No Brasil, essa mudança na percepção sobre a educação infantil ocorreu principalmente após a promulgação da Constituição de 1988, que estabeleceu o aprendizado até os 5 anos de idade como um direito social (Art. 7º, inciso XXV) e dever do Estado (Art. 208, inciso IV) (BRASIL, 1988). Anteriormente, a escolaridade infantil era considerada uma etapa anterior ao ensino formal. Posteriormente, em 2006, o texto constitucional foi atualizado para abranger as creches e pré-escolas como parte da educação infantil, além de exigir qualificações mínimas para os docentes do ensino primário (BRASIL, 2006).

Somente em 2009, com a publicação da Emenda Constitucional número 59, foi estabelecida a obrigatoriedade do ensino básico a partir dos 4 anos de idade, garantindo sua oferta gratuita (BRASIL, 2009). A Lei de Diretrizes e Bases da Educação Nacional de 2013 organizou a formação infantil até os 5 anos completos, dividindo-a entre creche (0 a 3 anos) e pré-escola (4 a 5 anos), mantendo a obrigatoriedade (BRASIL, 2013a).

De acordo com o Plano Nacional de Educação (PNE), estabeleceu-se como "meta 1" a proposta de universalizar, até 2016, a educação infantil na pré-escola para as crianças de 4 a 5 anos de idade e ampliar a oferta de vagas em creches de forma a atender, no mínimo, 50% das crianças de até 3 anos ao final da vigência deste Plano, cuja validade foi prorrogada para até 31 de dezembro de 2025¹ (BRASIL, 2014).

A frequência à creche e ou à pré-escola compreende um importante componente de política de desenvolvimento nacional. De modo geral, há consenso na literatura de que o ensino primário tem efeitos positivos e significativos no que diz respeito ao desenvolvimento cognitivo e aprendizado das crianças. Por outro lado, alguns pais enfrentam dificuldades ao tentar colocar seus filhos nessas instituições, seja por falta de vagas, incorrendo em ter que deixar de trabalhar para cuidar dos menores, seja devido a alguma dificuldade de acesso e viabilidade de locomoção ou até por opção própria.

Ter a necessidade de cuidar dos filhos ou parentes, traz dificuldades de investimento no capital humano das pessoas cuidadores e afeta indiretamente sua trajetória. Em especial, pode ser observado no fenômeno dos jovens "nem-nem", que remete aos indivíduos fora do mercado de trabalho e que também não frequentam instituições educacionais, pois, a dificuldade de acesso ao sistema educacional infantil, muitas vezes leva os pais, especialmente as mães, a serem forçados a permanecer em casa para cuidar dos filhos, o que pode limitar suas oportunidades de emprego e continuidade na sua própria formação

¹ O Projeto de Lei nº 5.665/2023, prorrogou até 31 de dezembro de 2025, a vigência do Plano Nacional de Educação, aprovado por meio da Lei nº 13.005, de 25 de junho de 2014.

educacional.

Esse cenário é particularmente relevante quando se considera que uma parte significativa dos jovens nem-nem mencionam a responsabilidade de cuidar dos parentes ou filhos como um dos motivos principais para sua situação atual. Assim, a falta de infraestrutura adequada, vagas disponíveis e acesso a creches e pré-escolas podem contribuir para a perpetuação do ciclo de exclusão educacional e econômica, alimentando a categoria dos jovens que não estão nem estudando nem trabalhando (IBGE, 2023).

A terminologia nem-nem foi adaptada do inglês *NEET*, que significa *not in education, employment, or training*. É um indicador relativamente novo e ganhou, durante as últimas décadas, uma importância e ênfase cada vez maior no que diz respeito ao mercado de trabalho desses indivíduos. Tem sua origem no Reino Unido no final dos anos 80, devido principalmente às mudanças no regime de benefícios, o que deixou a maioria das pessoas com idade entre 16 e 18 anos sem acesso ao seguro desemprego (BALAN, 2016).

Este conceito abrange grupos diversos, seja por evasão escolar, dificuldade na inserção no mercado de trabalho e outras vulnerabilidades sociais, cancelando as desvantagens competitivas associadas aos requisitos necessários para que essa condição seja apenas transitória. O lapso temporal estendido entre a escola e o trabalho tende a produzir fatores negativos não só para o jovem em questão, mas também para a sociedade em que vive, haja visto que produz altos índices de informalidade, remuneração mais baixa, capital humano de menor habilidade, além de repercussão futura na aposentadoria.

Como parte de uma agenda até 2030 sugerida pelas Nações Unidas, o Brasil adotou a medida de jovens nem-nem como indicador para os Objetivos de Desenvolvimento Sustentável (ODS) na meta 8.6 ², com o intuito de reduzir a proporção desses jovens sem trabalho, educação ou formação. Sendo alcançado por meio da promoção de emprego decente e crescimento econômico, uma vez que essa medida tem impacto no futuro ocupacional desse grupo populacional, pois eles não estão se qualificando nem adquirindo experiência profissional. Este indicador compreende aqueles que não frequentam a escola, sem incluir a participação em treinamento ou qualificação profissional e simultaneamente, estão desocupados ³ ou fora da força de trabalho ⁴.

De acordo com a síntese de indicadores divulgada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), para o Brasil observou-se, em números absolutos, que em 2022 o total de jovens que não estudavam e nem estavam ocupados atingiu o patamar de 10,9 milhões, o que corresponde a 22,3% dos jovens entre 15 e 29 anos. Ainda é possível notar uma forte relação desse grupo com o sexo feminino, na qual, para o mesmo ano, do total de jovens nem-nem citados acima, as mulheres representavam 6,9 milhões (63,4%),

² Disponível em: <https://odsbrasil.gov.br/objetivo8/indicador861>

³ De acordo com o IBGE, a definição de desocupados refere-se às pessoas que estão sem trabalho, mas procuraram emprego ativamente no período de referência da pesquisa.

⁴ São aqueles que não tomaram atitude para conseguir emprego, porém estão aptos ao trabalho ou não querem trabalhar.

enquanto os homens somavam 3,9 milhões (36,6%). Em paralelo, 2,3 milhões de crianças até 3 anos de idade não frequentavam creches, representando 20% dessa população (IBGE, 2023).

Diante do exposto, os jovens poderiam estar disponíveis para o trabalho e/ou qualificação, caso seus filhos ou parentes estivessem sob os cuidados de creche ou pré-escola. Assim, qual relação poderia ser estimada entre a existência de um jovem nem-nem com a presença de criança na escola no mesmo domicílio?

Buscando responder a essa indagação, o objetivo principal deste estudo é investigar o efeito da presença de crianças até 5 anos no ambiente escolar (creche ou pré-escola) sobre os jovens entre 15 e 29 anos, de estar na condição de não trabalhar nem estudar, residindo no mesmo domicílio.

Utilizando dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADc) de 2022⁵ do IBGE, será construído um painel para estimar esse efeito comparando dois grupos: o grupo tratado (domicílios com crianças em creche ou pré-escola) e o grupo de controle (domicílios sem crianças em creche ou pré-escola).

Aplicaremos o método de *escore de propensão* sob um modelo de escolha discreta binária, com a hipótese *logit*, para construir um *contrafactual* adequado que permita uma comparação legítima entre os grupos. Esse procedimento visa isolar o efeito da educação infantil sobre a condição nem-nem dos jovens, proporcionando uma análise mais precisa e detalhada do impacto dessa política pública na inclusão social e produtividade futura dos jovens.

Esta dissertação está estruturada em cinco capítulos. O primeiro capítulo apresenta os elementos introdutórios, a contextualização e os objetivos que orientam a pesquisa. O segundo capítulo compreende a revisão da literatura, organizada em seções que abordam, inicialmente, a temática dos jovens em condição nem-nem, seguida pela análise da relevância da educação na primeira infância.

O terceiro capítulo descreve a metodologia adotada, com destaque para o modelo empírico, as fontes de dados, os procedimentos de análise e as abordagens estatísticas utilizadas. No quarto capítulo, são apresentados os resultados, abrangendo uma análise preliminar, as estimativas derivadas da regressão múltipla, o método de Propensity Score Matching (PSM) e testes de robustez, bem como subseções que exploram os métodos de balanceamento aplicados, as estimativas do efeito médio do tratamento e a análise de sensibilidade.

Por fim, o quinto capítulo reúne as considerações finais, sintetizando as principais conclusões do estudo, as implicações práticas e sugestões para futuras pesquisas.

⁵ Foi utilizada a base de dados de 2022, devido esta ser a última atualização dos microdados contendo a primeira visita e o agrupamento no segundo trimestre como entrada de dados até a elaboração deste documento.

1.1 Objetivo

O objetivo geral deste estudo é investigar o efeito da presença de crianças em creche ou pré-escola sobre a condição de jovens entre 15 e 29 residentes no mesmo domicílio, utilizando dados da PNADc de 2022 e o método de escore de propensão com a hipótese *logit* para estimar a relação entre a educação infantil e a redução da probabilidade de jovens estarem nessa situação.

1.2 Objetivos Específicos

- Estimar a probabilidade de tratamento, ou seja, calcular as chances de cada domicílio ter ao menos uma criança em creche ou pré-escola (ser tratado), considerando uma série de covariáveis de pré-tratamento definidas na literatura, utilizando o método de escore de propensão.
- Realizar o pareamento, aplicando o algoritmo de *matching* (pareamento), especificamente o método *nearest neighbor propensity score matching* (vizinho mais próximo), para formar pares entre os domicílios tratados e os de controle com base no escore de propensão.
- Avaliar o balanceamento das covariáveis entre os grupos pareados, assegurando que as diferenças entre os domicílios tratados e de controle sejam minimizadas e que a comparação seja válida.
- Aplicar outras modelagens para verificar a robustez dos resultados utilizando a distância de Mahalanobis e o método de balanceamento por entropia.
- Estimar o efeito do tratamento, calculando o impacto da presença de crianças em creche ou pré-escola sobre a probabilidade de jovens, do mesmo domicílio, estarem na condição nem-nem, utilizando as estimativas obtidas após o pareamento e analisando o efeito médio do tratamento.

2 Revisão da Literatura

2.1 Literatura aos jovens nem-nem

Embora, de maneira simples e com base na percepção cotidiana, seja possível identificar a população que não trabalha nem estuda, este indicador revela uma população muito heterogênea. Segundo o estudo de Mascherini et al. (2012), estes jovens podem se enquadrar em alguns grupos principais, os quais podem ser identificados como:

- Os desempregados: seja de longa ou curta duração;
- Os indisponíveis: são os jovens com responsabilidades familiares, doentes ou deficientes;
- Os descomprometidos: aqueles que não procuram emprego ou educação e não estão constrangidos de fazê-lo por diversas razões e acabam seguindo estilos de vida perigosos e anti-sociais;
- Os voluntários: enquadram os jovens que estão envolvidos em outras atividades não formais, como aprendizagem autodirigida ou informalidade;

Estas categorias incluem uma combinação de jovens vulneráveis e não vulneráveis, desde pessoas desfavorecidas até aquelas que optam, por conta própria, se afastar do mercado de trabalho e da educação. Mesmo períodos curtos na condição de *NEET* podem ter consequências duradouras, como efeitos negativos no futuro desempenho profissional e nos rendimentos, além de repercussões na saúde física e mental. Essas consequências podem incluir relacionamentos conflituosos, abuso de substâncias e envolvimento em atividades ilícitas, impactando tanto o indivíduo quanto sua família. Ainda segundo Mascherini et al. (2012), os reflexos econômicos são particularmente relevantes para países da União Europeia, onde a desconexão prolongada com o mercado de trabalho pode comprometer a experiência profissional, aumentando a probabilidade de salários reduzidos no futuro a chamada “penalidade salarial” destacada no estudo.

Em seu trabalho, Coles, Bradshaw e Hutton (2002) sugerem que existem dois fatores de risco principais relacionados à vulnerabilidade do status de não estudar e não trabalhar. A primeira indica uma desvantagem competitiva e a segunda o descontentamento. Enquanto a desvantagem educacional esteja associada a fatores sociais como a condição familiar, a escola e as características pessoais do jovem, o descontentamento preocupa-se com as atitudes que os jovens têm em relação à educação e ao trabalho. Em seu estudo, eles afirmam que essa característica não é de permanência, expondo a existência

de transitoriedade nessa faixa etária. Os autores os classificam como desocupados de curto prazo, possuindo maior probabilidade de encontrarem emprego ou estudo ao longo de doze meses, além de que estes indivíduos estão mais disponíveis para participar de programas de treinamento profissional e têm mais informações sobre a relação de oferta de emprego. Em contrapartida, o segundo grupo, que não busca trabalho nem estudo, forma um contingente difícil de ser alcançado por políticas públicas.

Sabe-se que os jovens enfrentam uma transição não linear entre a escola e o trabalho. Guimarães, Marteleto e Alves (2016) revelam, para o caso brasileiro, que mudanças entre situações de emprego, desemprego e inatividade são movimentos habituais para esses jovens, enfatizando a necessidade de pesquisas longitudinais para compreender melhor os movimentos desses jovens. No entanto, demonstram preocupação com a tendência de ampliação ao desalento¹, especialmente no grupo de jovens mulheres negras e de baixa escolaridade, para as quais ocorrem a exclusão simultânea da escola e do mercado, fazendo um contraponto com a ligação à necessidade de desempenhar atividades de cuidados aos membros da família.

Essa complexidade no percurso de transição educacional e laboral ressalta a relevância de políticas de apoio que auxiliem a juventude nesse processo. Em um contexto de exclusão socioeconômica, como observado, políticas de transferência de renda, a exemplo do Programa Bolsa Família, mostram-se importantes para atenuar a condição nem-nem. Nessa direção, Vasconcelos et al. (2017) utilizam o método *Propensity Score Matching* (PSM) para investigar a relação entre o Programa Bolsa Família e a probabilidade de jovens entre 18 e 29 anos se enquadrarem na geração nem-nem, concluindo que ser beneficiário deste programa implica em um efeito redutor nas chances de que esses jovens não estudem e não trabalhem, resultado que é evidenciado em maior proporção para os indivíduos classificados como extrema pobreza. Também observaram um efeito médio maior para os homens no que diz respeito à redução da probabilidade de pertencerem à geração nem-nem.

É importante ressaltar o caso para o sexo feminino, já que representam maior parcela do grupo nem-nem. Em seu trabalho, Silva (2019) corrigindo o viés de auto seleção da amostra se valendo do método *Propensity Score Matching*, percebeu que mulheres de características similares, a qual em seu domicílio possui criança e o único diferencial entre elas é se seu filho frequenta creche, há um efeito sobre a jornada de trabalho positivo e consequentemente sobre a renda, chegando ao aumento de 8%, demonstrando o efeito da instituição educacional sobre o posicionamento laboral feminino.

A maternidade é frequentemente apontada como uma das principais causas para que mulheres jovens se tornem nem-nem, não apenas no Brasil. Dicks et al. (2022) investigaram as chances de jovens mães se tornarem e permanecerem nem-nem, relacionando com a

¹ O IBGE define desalento como aquelas pessoas que abandonaram a busca por uma vaga de emprego ou sequer iniciaram a procura, seja por se considerarem muito jovens, muito idosos, com pouca experiência ou por acreditarem que não encontrarão oportunidades em sua região.

disponibilidade dos membros da sua rede de apoio social e familiar, especialmente parceiros e avós, para ajudar nos cuidados com suas crianças, na população dos Países Baixos. Os resultados mostram que as mães jovens casadas têm menos probabilidades de se tornarem nem-nem do que as mães solteiras. Ter avós e viver nas proximidades está associado a uma menor probabilidade de se tornar *NEET*, ainda que as jovens as quais moram com os pais economicamente inativos possuem maior probabilidade de se tornarem nem-nem e menos probabilidades de sair.

Apesar de que para alguns jovens essa condição seja um período de tempo relativamente curto, fruto da mudança entre a escola e o emprego, para outros sua saída desse status pode ser mais demorado. Neste sentido, a literatura tende a discutir os custos econômicos associados ao nem-nem no Brasil, elencando algumas consequências repercutidas oriunda do ingresso tardio no mercado de trabalho, no qual varia da redução da produtividade, até possível ingresso na criminalidade. Shirasu e Arraes (2020) utilizam o método *Propensity Score Matching* para obter resultados que revelaram os custos econômicos associados à ociosidade dos nem-nem, concluindo que esses jovens representaram um custo de 0,6% do PIB em 2015, ainda demonstram que mais de R\$ 36 bilhões poderiam ser investidos em diversas áreas sociais que implicariam a melhoria de bem-estar de uma parcela substancial da sociedade.

Não obstante, durante a pandemia do vírus Sars-COV-2 houve consequências na diminuição do consumo, investimento e a perda de postos de trabalho. Fatalmente, houve um aumento do número de jovens desalentados e inativos, já que foram impactados no mercado de trabalho e na continuidade da formação escolar ou profissional. Recentemente, Silva e Vaz (2022) contextualizaram o posicionamento dos jovens nem-nem na pandemia, para os quais o Auxílio Emergencial assistiu esse grupo durante o auge dessa crise sanitária com cerca de 80% da categoria dos desencorajados e dos que estavam nessa situação por afazeres domésticos. Também enfatizam a preocupação das perspectivas para o posterior, pois um dos efeitos é a tendência de ampliação do grupo dos jovens nem-nem situados fora da força de trabalho, assim, levando em consideração as altas vulnerabilidades desses jovens, eles sugerem implementações de políticas públicas capazes de enfrentar os dois principais desafios, que são a desocupação de longo prazo e a elevada proporção de nem-nem que já encontra-se desengajada do mercado de trabalho e da educação.

No entanto, deve ser enfatizado que muitas vezes não é fácil diferenciar, entre esses fatores, as causas que levam a esse status e quais estão correlacionados ao jovem ser nem-nem, uma vez que a literatura descreve as características, mas dizem pouco sobre como as pessoas chegaram à essa situação de vulnerabilidade. Sabendo que a compreensão desse fenômeno ajuda no estabelecimento das políticas públicas, cabe ressalvas quanto a delimitação desse grupo.

2.2 Literatura à primeira infância

Pesquisas sobre a educação infantil no Brasil apontam que a frequência à creche pode ter impacto positivo nas etapas mais avançadas da escolaridade. Curi e Filho (2009) identificaram uma correlação significativa entre a frequência à creche e a conclusão do ensino médio e universitário, destacando que crianças as quais frequentaram creches apresentaram uma probabilidade maior de completar essas etapas educacionais em comparação com aquelas que não tiveram essa experiência. No entanto, o impacto sobre a conclusão do ensino fundamental não foi estatisticamente relevante. Em contrapartida, a frequência à pré-escola mostrou efeitos positivos em todas as fases do ciclo escolar, assim como em rendimentos futuros.

Correa, Comim e Tai (2014) analisaram o impacto da creche no desenvolvimento cognitivo e socioemocional de crianças entre 4 e 5 anos, destacando-se, como os efeitos podem variar dependendo das características familiares, condição socioeconômico e o grau de sensibilidade e envolvimento parental. Os autores demonstraram que as crianças de famílias nas quais as estruturas socioeconômicas são mais baixas, são as maiores beneficiadas em termos de habilidades cognitivas ao frequentarem a creche, enquanto efeitos negativos são observados quando a participação e sensibilidade parental são reduzidas.

Pinto e Guimarães (2017) investigaram o impacto da frequência das crianças às instituições de educação infantil no desempenho em testes de matemática no quarto ano do ensino fundamental. Os autores encontraram evidências de que a frequência à creche eleva as pontuações nos testes de proficiência, embora com magnitude modesta, e que esses efeitos são heterogêneos, variando de acordo com o nível de escolaridade materna. Além disso, destacam que a frequência à creche pode ter efeitos cumulativos ao longo da educação básica, resultando em ganhos de produtividade e maiores salários na vida adulta. O estudo também aponta que a creche beneficia as famílias vulneráveis, ao permitir maior participação no mercado de trabalho e reduzir a pobreza, gerando vantagens tanto para as crianças quanto para os pais ou cuidadores.

Dados da síntese de indicadores do IBGE de 2023, revelam que mais de 2,5 milhões de mulheres não buscaram trabalho porque precisavam cuidar de filhos ou parentes, neste sentido, é notório a concentração desse público em estudos como Campos e Silva (2020) que procuraram analisar como a frequência de crianças à escolarização afeta a alocação de tempo das mães, seja por emprego ou trabalho não remunerado. Valendo-se de dados da PNAD de 2011 a 2015 e aplicando métodos de controle para fatores observados e não observados, os resultados indicaram que o acesso à creche contribui para um aumento na participação das mães no mercado de trabalho formal em aproximadamente 2 pontos percentuais, sem, no entanto, reduzir de forma significativa o tempo dedicado às tarefas domésticas. Isso sugere que, apesar da inserção maior no mercado de trabalho, as mães continuam a desempenhar grande parte das responsabilidades no âmbito doméstico.

Este conflito entre o trabalho e deixar os filhos aos cuidados de terceiros, que afetam principalmente as mulheres, reflete também em diferentes contextos internacionais. Na Alemanha, Bick (2016) sugere que o aumento da oferta de cuidados infantis subsidiados para crianças de zero a dois anos há uma tendência de elevação na participação materna no mercado de trabalho, da mesma forma Bauernschuster e Schlotter (2015) exploraram a expansão do cuidado infantil em escolas e observaram que a matrícula em creches elevou a participação das mães jovens no mercado de trabalho.

Em um estudo na Argentina, Berlinski e Galiani (2007) avaliaram o impacto de um programa de frequência pré-escolar e constataram que a ampliação da infraestrutura educacional, que aumentou a oferta de vagas, resultou em um aumento do emprego materno. Brilli, Boca e Pronzato (2016) examinou os efeitos da disponibilidade de creches públicas na Itália, com dados de 2009 a 2010, sobre o emprego das mães. O estudo encontrou que um aumento na cobertura de creches públicas elevou a probabilidade das mães trabalharem em 1,3 pontos percentuais.

Givord e Marbot (2015) avaliou o impacto, no curto prazo, de um aumento nos subsídios para utilização de creches privadas ou serviços de cuidados pagos na taxa de participação das mães de crianças em idade pré-escolar na França. A análise revelou um aumento significativo tanto na utilização de serviços pagos de creche, devido à nova política, bem como a participação no mercado de trabalho das mães, sendo de maior magnitude em famílias numerosas.

3 Metodologia

Esta pesquisa estabelece um recorte etário para dois grupos distintos. O primeiro, segue a Lei de Diretrizes e Bases da Educação Nacional de 2013, organizando a formação infantil até os 5 anos completos. O segundo é em consonância à definição da Organização Internacional do Trabalho (OIT) e a indicação pela Organização das Nações Unidas (ONU), bem como embasado no Estatuto da Juventude ¹, estabeleceu-se o parâmetro delimitador de faixa etária para os jovens com idade entre 15 e 29 anos que não estudam e não trabalham. Este último, formam uma categoria que contém uma variedade entre subgrupos, alguns dos quais são vulneráveis e outros não, com diferentes experiências, características sociais e necessidades individuais ou familiares.

A implementação empírica deste trabalho se dá a partir do uso dos microdados da PNADc de 2022. A escolha desse período se justifica pela realização da Pesquisa Suplementar de Educação durante esse trimestre, que amplia a investigação sobre aspectos educacionais, abrangendo desde a educação infantil (creches e pré-escolas) até a educação profissional e tecnológica, além de características relacionadas à frequência escolar e cursos (IBGE, 2022). Essa suplementação de dados é essencial para capturar informações detalhadas e relevantes ao tema deste estudo.

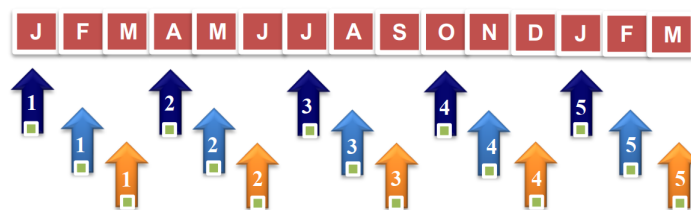
Vale salientar que a incorporação amostral com seus pesos não foi realizada nesta análise, uma vez que o objetivo principal é a execução de modelos de pareamento pela estimação do escore de propensão, focando na comparação entre domicílios. O uso dos pesos amostrais, que ajustam a representatividade da amostra para a população, seria incompatível com essa estratégia, pois o pareamento requer uma amostra mais balanceada e uma correspondência precisa entre os diferentes tipos de domicílios.

A PNAD Contínua oferece diversas informações socioeconômicas do país, sendo realizada trimestralmente através de uma amostra probabilística e complexa de domicílios. Esta pesquisa tem caráter longitudinal e como definido pelo Manual Básico da Entrevista² permite, através de um método contínuo, o monitoramento de variáveis ao longo do tempo, uma vez que o mesmo domicílio é entrevistado cinco vezes, com um esquema rotacional a cada dois meses, ou seja, a residência é entrevistada no primeiro mês, fica ausente da amostra nos dois meses seguintes, para ser novamente visitada no quarto mês. Assim segue, até completar o ciclo das cinco entrevistas ao longo do ano, conforme esquema na Figura 1 (IBGE, 2022).

¹ O Estatuto da Juventude foi instituído pela Lei n. 12.852 de 05/08/2013 Brasil (2013b), consolidando os direitos dos jovens.

² O Manual Básico da Entrevista da PNADc 2022 do IBGE fornece orientações sobre os procedimentos de coleta de dados, definição de variáveis e codificação das respostas, essenciais para garantir a qualidade e a consistência das informações coletadas. Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=55642>>.

Figura 1 – Esquema rotacional das entrevistas da PNADc nos domicílios



Fonte: Extraído do Manual Básico da Entrevista, PNADc 2022.

O critério de escolha da base trimestral deu-se de acordo com o Livro de Notas técnicas Versão 1.12³, o qual informa que os dados de características gerais dos moradores são obtidos em todas as cinco visitas da pesquisa, independentemente do trimestre, sendo a escolha determinada pelo melhor aproveitamento da amostra e sugere a primeira entrevista como aquela de melhor utilização, além de, dessa forma, fechar o ciclo de entrevistas no mesmo domicílio.

Além disso, optou-se pelo segundo trimestre para capturar as informações do módulo de Educação⁴ para crianças menores de 5 anos, realizadas simultaneamente com a coleta de dados sobre as condições de habitação, durante a primeira visita⁵. Essa abordagem permite integrar dados sobre educação infantil e características do domicílio, assegurando a sincronização temporal das informações com o ciclo anual da investigação, conforme o padrão estabelecido pela PNADc.

Para realizar este estudo, foi necessária a identificação dos domicílios na entrevista, a fim de agrupá-los com base no corte etário dos indivíduos ali residentes e evitar a dupla contagem da mesma habitação. Esse processo foi viabilizado pela geração de um número índice para a subamostra de dados longitudinais integrada ao banco de dados, em conformidade com o documento Chaves da PNAD Contínua⁶, que detalha a composição das chaves de identificação dos domicílios e indivíduos na base de dados. Esse documento explica que a chave de domicílio é composta pelas variáveis descritas no Quadro 1, sendo essenciais para realizar análises longitudinais e identificar registros específicos dentro da pesquisa.

³ O livro discute aspectos centrais da metodologia utilizada no levantamento e inclui conceitos e definições essenciais para entender seus indicadores.

⁴ O Módulo Educação da PNAD Contínua consiste em uma entrevista suplementar concentrada no segundo trimestre, abrangendo todos os moradores do domicílio, independentemente da idade. Diferencia-se das demais edições da pesquisa anual por incluir informações adicionais, como dados específicos sobre a Educação Infantil, voltados para crianças de 0 a 5 anos.

⁵ O módulo de características do domicílio, que aborda aspectos como abastecimento de água, esgotamento sanitário e coleta de lixo, é aplicado em todas as visitas trimestrais, sendo especialmente relevante na primeira visita para consolidar as informações socioeconômicas e habitacionais básicas do domicílio.

⁶ Disponível em: PNAD Contínua – Chaves

Quadro 1 – Descrição das variáveis para identificação

Variável	Descrição
UPA	Unidade Primária de Amostragem*
V1008	Número de seleção do domicílio
V1014	Painel do grupo da amostra

Fonte: Elaboração própria.

Nota 1: A UPA é composta pela UF, Número Sequencial do domicílio e um dígito verificador.

Assim, algumas informações importantes são apresentadas na Tabela 1, que ilustra a composição dos domicílios entrevistados. A amostra é composta por 6.930 domicílios, incluindo jovens entre 15 e 29 anos, além de crianças com até 5 anos. A distribuição dos dados revela padrões importantes sobre a relação entre a condição de nem-nem e a presença de crianças na escola. Observa-se que a maior parte (31%, ou 2.147 habitações) possuem crianças matriculadas em creche ou pré-escola e não incluem jovens nem-nem. Em contrapartida, 1.822 domicílios (26,3%) têm jovens nem-nem, e, nesses lares, as crianças não frequentam a educação infantil. Esses dados sugerem que a inclusão de crianças na educação infantil é mais comum em domicílios sem jovens que não trabalham nem estudam, indicando uma possível associação entre o acesso à escola e melhores condições sociais para os moradores.

Tabela 1 – Composição dos domicílios na base de dados

Domicílios possui		Número de domicílios	Percentual (%)
Nem-Nem	Criança na Escola		
Não	Não	1.364	19,7
Não	Sim	2.147	31,0
Sim	Não	1.822	26,3
Sim	Sim	1.597	23,0
Total		6.930	100

Fonte: Elaboração própria.

Por fim, verificou-se a influência da presença de criança em creche ou pré-escola, localizada no mesmo domicílio, sobre o status de nem-nem do jovem, utilizando a estimação do escore de propensão sob o modelo de escolha discreta binária, conforme a hipótese *logit*.

Contudo, a decisão de matricular a criança em uma instituição de ensino não é aleatória, pois está influenciada por diversos fatores que levam os responsáveis a essa escolha. Entre esses fatores, destaca-se a busca por cuidado individualizado, que pode levar alguns responsáveis a optar por não trabalhar para dedicar-se exclusivamente ao cuidado da criança. Além disso, a localização da residência é crucial, pois a falta de

infraestrutura nas áreas em que as famílias vivem pode resultar em escassez de vagas em creches gratuitas, dificultando ainda mais o acesso à educação infantil.

Essa situação incide diretamente no viés de auto-seleção, uma vez que famílias em diferentes contextos socioeconômicos podem tomar decisões divergentes sobre a matrícula em creches com base em suas circunstâncias pessoais. Assim, aqueles que conseguem matricular seus filhos em creches podem ter características que os diferenciam dos que não o fazem, o que pode afetar os resultados em estudos sobre a relação entre a educação infantil e o status de nem-nem dos jovens.

Para mitigar este viés, é fundamental estabelecer um contrafactual robusto. Isso envolve emparelhar os domicílios dos grupos de controle e tratamento com características observáveis semelhantes, como renda *per capita* domiciliar, nível de escolaridade dos responsáveis e condições da moradia. Esse emparelhamento garante que as comparações entre os grupos sejam mais justas e válidas, permitindo uma avaliação mais precisa do efeito da educação infantil.

Assim, ao garantir que os grupos sejam comparáveis, é possível reduzir a influência de fatores não observáveis que poderiam distorcer os resultados, oferecendo uma base sólida para as conclusões da pesquisa.

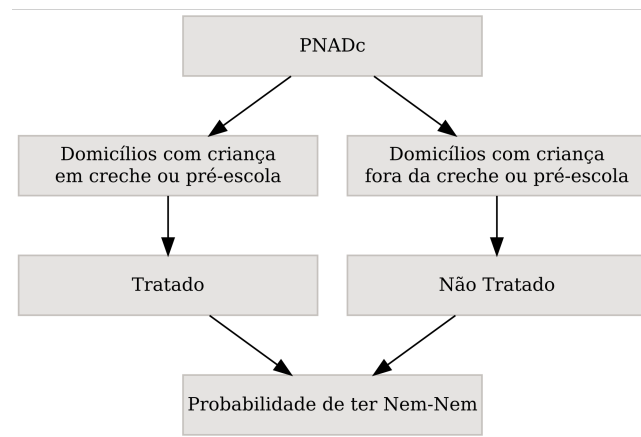
3.1 Modelo empírico

O trabalho pretende investigar se a presença de criança na escola afeta as chances de ter jovens na condição de nem-nem residente no mesmo domicílio. Será utilizado o processo de pareamento dos dados pela probabilidade predita *Propensity Score* do domicílio ter criança matriculada em creche ou pré-escola (tratamento), dado os atributos observados, identificando um grupo de referência e encontrando seus pares segundo as características elencadas.

Além do pareamento, é utilizado modelos multivariados, regressões logísticas com efeitos fixos, para mensurar a influência das covariadas em relação ao domicílio com nem-nem. A ideia do PSM é contrastar os domicílios que foram submetidas a uma ação, neste caso a presença de filho(s) ou parente(s) em creche ou pré-escola (tratado), com domicílios não submetidos (controle), porém com características semelhantes em termos de probabilidade de terem passado por essa ação.

A Figura 2 simplifica o fluxo interpretativo desse trabalho, o qual descreve como a presença de uma criança em creche ou pré-escola, influencia a probabilidade de existir um jovem na condição de nem-nem no domicílio. A análise compara os dois grupos ("Tratado" e "Não Tratado") para avaliar o impacto dessa variável específica (presença em escola) na probabilidade de existir um jovem que não estuda nem trabalha no lar.

Figura 2 – Fluxo de Decisão: Efeito da presença em creche na probabilidade de ter nem-nem no domicílio



Fonte: Elaboração Própria.

Ainda, é necessário respeitar duas hipóteses a fim da utilização do processo de pareamento *matching*. Primeiramente é a ignorabilidade, isto é, a seleção dos atributos observáveis do domicílio independem do status do tratamento. A segunda alternativa é de suporte comum na região do vetor de variáveis explicativas, que comporta as características dos tratados e não tratados, não sendo possível a identificação de qual grupo o domicílio pertence.

Logo, para mensurar o efeito da creche no domicílio sobre a probabilidade do jovem nem-nem a ele pertencente, esta dissertação emprega a abordagem centrada na escolha de características percebidas do domicílio e individuais do responsável pela habitação, de forma a tornar-se um contrafactual apropriado e a média do impacto da creche nos tratados possa ser calculado, já que, como mencionado, o método de pareamento torna os grupos de controle e de tratamento semelhantes.

Os efeitos da interação são incluídos ao banco de dados e para melhorar o equilíbrio das covariáveis, é estimado usando a rotina *pscore* e balanceamento do R *bal.tab*⁷ via RStudio, deixando o conjunto de variáveis correspondentes comuns para o grupo tratado e controle.

Para realizar o pareamento, foi aplicado o algoritmo de pareamento *MatchIt*⁸, pelo método *nearest neighbour matching (NNM)*, indicando o emparelhamento de vizinho mais próximo, com reposição. Neste caso, um participante do grupo de controle é selecionado como correspondente a um tratado com pontuação de propensão mais próxima. Também é

⁷ A função *bal.tab* faz parte do pacote *cobalt* no R. Ela realiza a análise de balanceamento de forma automatizada, utilizando o escore de propensão estimado pela função *pscore* (RSTUDIO, 2020).

⁸ O pacote *MatchIt* é utilizado no R para realizar o pareamento de unidades em estudos observacionais, ajustando as covariáveis e permitindo comparações justas entre grupos tratado e controle.

aplicada a modelagem utilizando a distância de Mahalanobis e o método de balanceamento por entropia.

Por fim, é estimado o efeito do tratamento, calculando o impacto da presença de crianças na escola sobre a probabilidade dos domicílios possuírem jovens fora do mercado de trabalho e não estudando, utilizando as estimativas obtidas após o pareamento e analisando o efeito médio do tratamento.

3.2 Estratégia Empírica

A partir dos dados da PNADc para o segundo trimestre de 2022 do IBGE, propõe-se solucionar o problema de viés de seleção e obter estimativas não enviesadas do efeito médio do tratamento. Para isso, comparam-se domicílios do grupo de controle e do grupo de tratamento, que possuem atributos observáveis similares, exceto pelo critério de que, no grupo de tratamento, o domicílio tem crianças matriculadas em creche ou pré-escola. Dessa forma, é possível inferir sobre a correlação (ou a falta dela) entre a presença de jovens na condição de nem-nem e a situação educacional das crianças no domicílio.

Nesta metodologia, a seleção do grupo de controle considera diversas características observáveis, conforme identificado em pesquisas anteriores, para assegurar comparabilidade com o grupo de tratamento. Assim, a análise incorpora variáveis de natureza socioeconômica e demográfica, além de aspectos específicos dos domicílios e dos chefes de família, que podem influenciar tanto a decisão de matricular uma criança em creche quanto a condição dos jovens em relação ao status de nem-nem. Essas características foram cuidadosamente escolhidas, com base nos estudos de Shirasu e Arraes (2020) e Mascherini et al. (2012), para capturar determinantes relevantes na composição do grupo de controle.

A estimação do escore de propensão realizada através do modelo de escolha discreta binária sob a hipótese *logit*, tem equação geral é escrita como:

$$P(Y = 1|X) = \frac{1}{1 + e^{-X\beta}} \quad (3.1)$$

Onde, $P(Y = 1|X)$ representa a probabilidade de um domicílio possuir jovem na condição de nem-nem, dado o conjunto de características X . A variável Y é binária, indicando a presença (1) ou ausência (0) de jovens nem-nem no domicílio. Já o vetor $X\beta$ combina as características observadas e os coeficientes estimados, representando o peso relativo de cada variável. O termo $e^{-X\beta}$ deriva da função logística, que transforma os valores do modelo em probabilidades dentro do intervalo $[0, 1]$.

A cesta de variáveis X é baseada nos estudos de Shirasu e Arraes (2020) e Mascherini et al. (2012), incluindo a localização (urbana ou rural), presença de coleta de esgoto ou lixo, internet no domicílio, renda domiciliar *per capita*, número de pessoas no domicílio, gênero e anos de estudo do chefe da família, conforme detalhado no Quadro 2.

O componente de maior interesse é a variável de intervenção, uma *dummy* que indica a presença de crianças de até 5 anos matriculadas em creche ou pré-escola no domicílio (1 para sim; 0 caso contrário). Desta forma, é possível averiguar se há e qual é o efeito da escolarização infantil sobre a condição do domicílio ter jovem e este não estar no mercado de trabalho nem estudando.

Quadro 2 – Descrição das variáveis para o modelo

Categoria	Descrição das variáveis	Tipo	Resultados
Dependente	Domicílio com nem-nem	<i>dummy</i>	(1) Possui nem-nem (0) Caso contrário
Intervenção	Domicílio com criança em escola	<i>dummy</i>	(1) Sim (0) Não
Variáveis explicativas	Situação do domicílio	<i>dummy</i>	(1) Urbano (0) Rural
	Unidade da Federação	inteiro	Estados e Região
	Pessoas no domicílio	inteiro	Número de pessoas no domicílio
	Esgoto do domicílio	<i>dummy</i>	(1) Rede geral (0) Caso contrário
	Lixo do domicílio	<i>dummy</i>	(1) Coleta por serviço de limpeza (0) Caso contrário
	Domicílio com internet?	<i>dummy</i>	(1) Sim (0) Não
	Renda <i>per capita</i> domiciliar	<i>float</i>	Valores em Reais
	Sexo do chefe de família	<i>dummy</i>	(1) Mulher (0) Homem
	Anos de estudo do chefe de família	inteiro	0 a 16 anos

Fonte: Dicionário da PNADc. Elaboração própria.

4 Resultados

Este capítulo apresenta as principais constatações da pesquisa, começando com a identificação dos domicílios na PNAD Contínua de 2022. Em seguida, realiza-se uma análise preliminar para explorar características gerais da amostra. Posteriormente, são conduzidas as estimações, que incluem modelos de regressão múltipla e análise por escore de propensão, utilizando métricas como a distância de Mahalanobis e a Entropia para melhorar a qualidade das comparações. Por fim, são discutidas as análises de sensibilidade e os valores do efeito médio do tratamento, buscando estabelecer alguma relação entre a presença de jovens na condição de nem-nem e a existência de crianças em escola no mesmo domicílio, conforme descrito na subamostra construída.

Para tanto, são produzidas uma transformação e uma filtragem na PNADc. A Tabela 2 apresenta a quantidade de casos encontrados na base de dados após a aplicação dos filtros, incluindo apenas os domicílios que contêm crianças e/ou jovens, com remoção dos valores ausentes ("NA"). Assim, as análises foram elaboradas com os dados efetivamente presentes.

Tabela 2 – Número de casos

Variável	Total
Domicílios com criança(s) e jovens	6.930
Domicílios com criança(s) em escola	3.744
Domicílios com jovens nem-nem	3.419

Fonte: Dados da PNADc 2º trimestre de 2022. Elaboração própria.

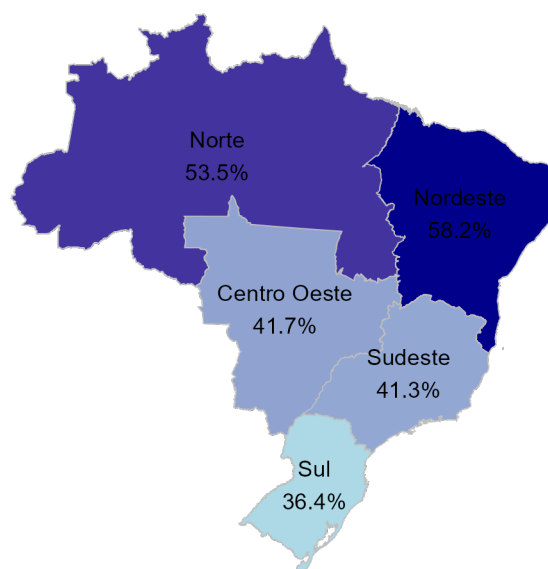
4.1 Análise preliminar

De acordo com a síntese de indicadores divulgada pelo IBGE, observou-se, em números absolutos, que em 2022 o total de jovens que não estudavam e não estavam ocupados atingiu o patamar de 10,9 milhões, o que corresponde a 22,3% dos jovens entre 15 e 29 anos. Ainda é possível notar uma forte relação desse grupo com o sexo feminino, no qual, para o mesmo ano, do total de jovens nem-nem citados acima, as mulheres representavam 6,9 milhões (63,4%), enquanto os homens representavam 3,9 milhões (36%) (IBGE, 2023).

Com a base de dados filtrada para as estimações e a segregação dos domicílios com jovens na condição de nem-nem por região, observa-se uma predominância dessas situações no Norte e no Nordeste do país. Conforme apresentado na Figura 3, jovens que

não trabalham nem estudam representaram 53,5% e 58,2% dos domicílios nessas regiões, respectivamente.

Figura 3 – Percentual de domicílios com jovens nem-nem por região



Fonte: Elaboração Própria.

Na Tabela 3, se encontra os diversos motivos apontados pelos jovens em estarem nessa condição. Alguns não tomam providências para conseguir emprego, enquanto outros gostariam de trabalhar, mas por alguma razão não o fazem. Há também aqueles que alegam não haver oportunidades de trabalho na região, sendo este o motivo de maior recorrência, com 44,9%. Um segundo fator relevante, que os levam à condição de não trabalhar nem estudar, está relacionado à necessidade de cuidar de filhos em seu domicílio, apresentando um percentual de 21,5% .

Tabela 3 – Motivo da condição de ser nem-nem

Legenda	Amostra*	(%)
Conseguiu proposta de trabalho à iniciar	70	1,1
Estava aguardando resposta para conseguir trabalho	237	3,6
Estava estudando por conta própria	291	4,4
Não conseguia trabalho adequado	677	10,2
Não conseguia trabalho, considerado muito jovem ou idoso	49	0,7
Não havia trabalho na localidade	2.994	44,9
Não tinha experiência profissional	328	4,9
Outros motivos	198	3,0
Por problema de saúde ou gravidez	384	5,8
Tinha que cuidar do(s) filho(s) ou parente(s)	1433	21,5

Fonte: Dados da PNADc 2º trimestre de 2022. Elaboração própria.

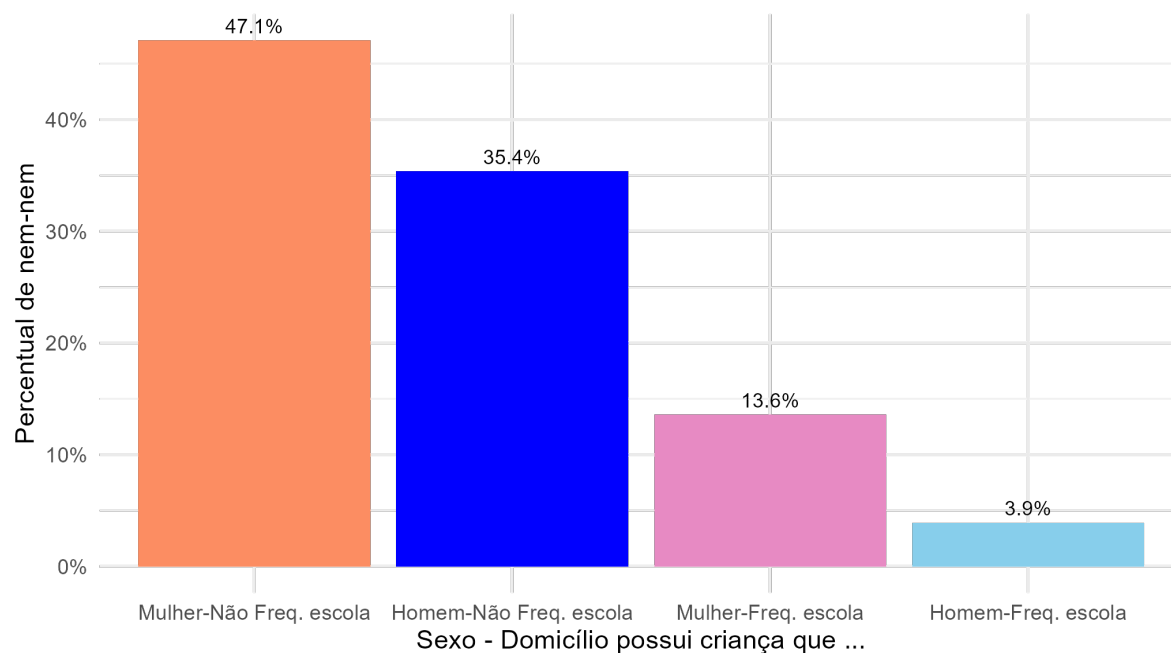
Nota 1: Refere-se a quantidade de respostas individuais.

Nota 2: Os dados são de acordo com a filtragem apresentada na Metodologia.

Apesar de não ser a maior parcela dos motivos da população nem-nem, este último será o objeto primordial para o estudo. Dados da Síntese de Indicadores Sociais do IBGE, utilizando informações da PNAD Educação, destacam a magnitude do problema: cerca de 1,6 milhão de crianças entre 0 e 5 anos no Brasil não estão matriculadas em instituições de ensino devido à falta de vagas. Esse cenário reforça a relevância de investigar o impacto da ausência de acesso à educação infantil (IBGE, 2023).

Abordando a presença de crianças em instituição educacional e relacionando com o jovem nem-nem residindo no mesmo domicílio, é apresentada a figura 4, na qual é perceptível o alto número desses jovens em residências onde as crianças estão fora da creche. Neste caso, 47,1% das mulheres e 35,4% dos homens estão na condição de nem-nem. Por outro lado, quando as crianças no domicílio, sejam filhos ou parentes, têm a chance de ingressar em estabelecimento de ensino na primeira infância, o percentual desses jovens cai para 13,6% entre as mulheres e 3,9% entre os homens. Sugerindo uma forte correlação entre a presença à escola e o *status* de ser nem-nem.

Figura 4 – Percentual de nem-nem por sexo em domicílios de acordo com a presença ou não da criança na escola



Fonte: Elaboração Própria.

Contudo, isso sugere que a oferta de vagas em creches pode ser um fator importante, já que pertencer a esse grupo pode não ser apenas resultado de escolhas pessoais, mas também da falta de opções para deixar os filhos ou parentes, dificultando que esses jovens trabalhem ou estudem.

4.2 Estimações

Primeiramente, realizamos uma comparação entre o grupo de controle e o grupo de tratamento em relação às variáveis do modelo, abrangendo o parâmetro de corte etário das crianças entre 0 e 5 anos e jovens entre 15 e 29 anos no domicílio. Dessa forma, é possível obter uma visão geral das variáveis com diferenças significativas e aquelas sem tais diferenças, bem como a apresentação de suas proporções ou médias em cada grupo.

A Tabela 4 apresenta uma comparação detalhada entre o grupo de controle e o grupo de tratamento em relação a diversas variáveis socioeconômicas e demográficas. Os valores médios de cada variável para ambos os grupos são listados, juntamente com o p-valor para a comparação estatística, indicando se as diferenças entre os grupos são estatisticamente significativas. A coluna "Iguais" aponta se as características dos grupos são consideradas equivalentes, adotando um nível de significância de 5%¹ para essa avaliação.

¹ Se o p-valor for maior que 0,05, as características dos grupos são consideradas equivalentes ("Sim"); caso contrário, são consideradas diferentes ("Não").

Tabela 4 – Comparação entre os grupos

Variáveis	Grupo de controle	Grupo de tratamento	P-Valor	Iguais
Urbano	0,649	0,725	0,000	Não
Esgoto do domicílio	0,325	0,400	0,000	Não
Lixo do domicílio	0,753	0,834	0,000	Não
Internet do domicílio	0,875	0,895	0,010	Não
Norte	0,237	0,151	0,000	Não
Nordeste	0,356	0,391	0,003	Não
Sudeste	0,176	0,218	0,000	Não
Sul	0,122	0,150	0,001	Não
Centro Oeste	0,109	0,090	0,010	Não
Renda <i>per capita</i> domiciliar	698,695	807,077	0,000	Não
Anos de estudo do chefe	9,023	9,447	0,000	Não
Gênero do chefe (mulher)	0,524	0,575	0,000	Não
Número de pessoas no domicílio	4,413	4,571	0,000	Não

Fonte: Elaboração própria.

Assim, o PSM é uma estratégia válida para este estudo, uma vez que os grupos iniciais apresentam diferenças. O objetivo do pareamento é justamente equalizar esses grupos em relação às variáveis observáveis, permitindo uma comparação mais precisa.

4.2.1 Regressão Múltipla

Inicialmente, utilizamos um modelo ingênuo, contendo uma regressão múltipla, que é uma técnica estatística usada para entender a relação entre uma variável resposta e várias variáveis independentes simultaneamente. Neste caso, é realizada uma regressão logística múltipla, pois estamos lidando com uma variável dependente binária do domicílio possuir ou não nem-nem e utilizando a função *glm()* com a família binomial e o *link* de *logit*.

A Tabela 5 expressa se as diversas variáveis demográficas e socioeconômicas exercem uma influência significativa na probabilidade do domicílio ter nem-nem. A participação no tratamento, neste caso, o domicílio possuir criança matriculada em creche ou pré-escola, tem um efeito negativo indicando que, para cada aumento unitário na variável *tratado*, as chances deste domicílio possuir indivíduo que não trabalha nem estuda diminui.

Tabela 5 – Modelo de Regressão Logística Múltipla

	Variável Dependente Domicílios com Nem Nem
Tratado	−0,510*** (0,052)
Urbano	−0,181** (0,083)
Esgoto do domicílio	−0,015 (0,063)
Lixo do domicílio	−0,190** (0,092)
Internet do domicílio	−0,180** (0,087)
Norte	0,047 (0,103)
Nordeste	0,194** (0,095)
Sudeste	−0,036 (0,102)
Sul	−0,098 (0,109)
Centro Oeste	
Renda <i>per capita</i> domiciliar	−0,001*** (0,0001)
Anos de estudo do chefe	−0,027*** (0,007)
Gênero do chefe (mulher)	−0,019 (0,053)
Número de pessoas no domicílio	−0,029* (0,017)
Constante	1,689*** (0,165)
Observações	6.930

Notas:

*p<0,1; **p<0,05; ***p<0,01

Fonte: Elaboração própria.

Nota 1: O Centro-Oeste é a categoria de referência para a variável "Região" neste modelo de regressão.

O coeficiente representa o logaritmo da razão de chances, contudo, é possível a conversão para termos percentuais, utilizando a equação *Odds Ratio*²

$$Oddsratio = e^{\beta_i} \quad (4.1)$$

Onde, β_i é o coeficiente estimado para a variável preditora X_i no modelo. e^{β_i} representa o *Odds Ratio*, ou seja, a razão entre as chances do evento ocorrer ($P(Y = 1)$) versus as chances do evento não ocorrer ($P(Y = 0)$) para cada aumento unitário em X_i , sabendo que e é a base do logaritmo natural, aproximadamente 2,71828.

No primeiro caso, para a variável *Tratado*, no qual o domicílio possui criança de até 5 anos em escola, tem-se $Odds\ ratio\ tratado = e^{-0,51} = 0,6005$. Esse valor implica que, ao ter uma criança matriculada em creche ou pré-escola, as chances do domicílio possuir um jovem na condição de nem-nem são 39,95% menores. O cálculo é obtido por meio da fórmula $((1 - 0,6005) = 0,3995)$, ou seja, a redução das chances é de aproximadamente 40%

² Uma forma alternativa de expressar a probabilidade em relação a eventos complementares, conforme discutido em Jr, Lemeshow e Sturdivant (2013)

4.2.2 Propensity Score Matching (PSM)

É uma abordagem estatística introduzida na literatura por Rosenbaum e Rubin (1983) e comum em estudos observacionais para estimar o impacto de uma intervenção, tratamento ou exposição sobre uma variável de resultado.

Consiste em estabelecer grupos de tratamento e controle comparáveis com um conjunto de covariáveis observadas. Para atingir esse objetivo, a cada domicílio é atribuída uma pontuação de propensão, que representa a probabilidade de receber o tratamento, com base nas características que foram observadas.

Neste trabalho, o grupo de tratamento é composto por domicílios com crianças matriculadas em creche ou pré-escola, enquanto o grupo de controle inclui domicílios em que as crianças não estão matriculadas em tais instituições. Ambos os grupos foram comparados com base em características observadas, como renda familiar, composição do domicílio, localização geográfica e perfil do chefe de família, para assegurar condições similares e reduzir o viés de seleção.

Em seguida, os indivíduos com pontuações de propensão comparáveis são emparelhados para formar grupos de tratamento e controle comparáveis. O modelo pode fornecer estimativas não enviesadas do efeito médio do tratamento e resolver o problema do viés de seleção.

4.2.2.1 Escore Propensão

O escore de propensão é uma probabilidade condicional que estima as chances de um participante receber o tratamento com base nas covariáveis observadas. Esse escore reflete a probabilidade de participação no tratamento, dado o perfil das características do domicílio. Portanto, o escore de propensão atua como um mecanismo de balanceamento das covariáveis entre os grupos de controle e tratamento. Quando esses grupos apresentam escores de propensão semelhantes, eles se tornam mais comparáveis, facilitando uma avaliação mais precisa do efeito do tratamento.

Inicialmente, ajustamos um modelo de regressão logística para modelar a propensão de receber o tratamento com base em um conjunto de covariáveis mencionadas anteriormente. Em seguida, calculamos as pontuações de propensão para cada indivíduo e as adicionando ao conjunto de dados como uma nova variável denominada: *pscore*.

A Tabela 6 informa que os escores de propensão variam entre 0,252 e 0,893, com a maioria dos valores concentrados no 3º quartil (0,604), indicando que em maior parte estão acima da mediana, sugerindo que há uma tendência na amostra para escores mais altos de propensão. A mediana está localizada no valor de 0,554, indicando que metade dos valores do *pscore* estão abaixo desse número e a outra metade está acima. O valor médio é de 0,540. A análise desses valores ajuda a entender a distribuição da propensão ao tratamento entre os indivíduos no estudo, sendo esses escores essenciais para o pareamento

no PSM.

Tabela 6 – Estatísticas Descritivas do Escore de Propensão

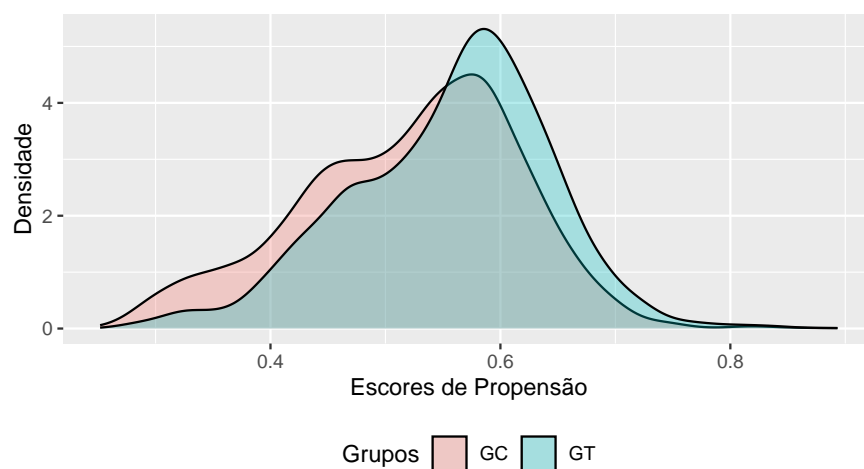
Estatística	Valor
Mínimo	0,252
1º Quartil (Q1)	0,477
Mediana	0,554
Média	0,540
3º Quartil (Q3)	0,604
Máximo	0,893

Fonte: Elaboração Própria.

Também se faz necessário avaliar a hipótese de suporte comum, ou *overlap*. É fundamental para a análise de pareamento por escore de propensão, pois garante que cada domicílio do grupo tratado tenha um correspondente com características similares no grupo de controle. Isso significa que as distribuições dos escores de propensão entre os grupos precisam ter uma boa sobreposição, formando uma área comum onde ambos compartilham probabilidades similares de receber o tratamento.

Esse suporte comum é visualmente verificável no gráfico de densidade da Figura 5 percebendo uma boa sobreposição, garantindo que a comparação entre os grupos seja válida, onde a curva verde representa o grupo de tratamento e a curva rosa representa o grupo de controle. A sobreposição dessas curvas indica que existem indivíduos em ambos os grupos com escores semelhantes, reforçando a capacidade de criar pares válidos e, assim, um contrafactual robusto.

Figura 5 – Gráficos de densidade dos escores de propensão



Fonte: Elaboração Própria.

Também é proposto o processo de produção da matriz de confusão, como apresentado no livro de Steiner e Cook (2013), pelo qual ajuda a avaliar o desempenho do modelo de previsão dos escores de propensão. Nela é possível extrair métricas relevantes como:

- Acurácia: indica a proporção de previsões corretas, para os dois grupos;
- Sensibilidade: infere sobre a capacidade do modelo de identificar corretamente os indivíduos tratados;
- Especificidade: consiste na capacidade do modelo de identificar corretamente os indivíduos não tratados;
- Acurácia Balanceada: média entre a sensibilidade e a especificidade.

No contexto desse estudo, a matriz de confusão avalia o quão bem os indivíduos do grupo de controle correspondem aos indivíduos do grupo de tratamento com base em suas pontuações de propensão.

A tabela 7 apresenta os resultados obtidos. Com uma acurácia de 57,59%, o modelo acerta mais da metade das previsões. A sensibilidade de 52,17% revela a capacidade do modelo de identificar corretamente os casos positivos (neste contexto, os domicílios com crianças em creche). A especificidade de 62,21% indica que o modelo consegue identificar corretamente os casos negativos (domicílios não tratados). Por fim, a acurácia balanceada de 57,19% complementa a análise, proporcionando uma avaliação mais justa do desempenho do modelo, especialmente em situações em que as classes são desiguais.

Em resumo, de acordo com Fawcett (2006) valores entre 50% e 70% na matriz de confusão indicam que o modelo apresenta um desempenho moderado e aceitável, o que se alinha com os resultados obtidos na análise.

Tabela 7 – Métricas de Avaliação da Matriz de Confusão

Métricas	Valores
Acurácia	57,59
Sensibilidade	52,17
Especificidade	62,21
Acurácia Balanceada	57,19

Fonte: Elaboração própria.

4.2.2.2 Pareamento

Nos estudos observacionais, o pareamento é uma técnica essencial para tornar os grupos comparados mais semelhantes em relação às variáveis de interesse, permitindo uma avaliação precisa do efeito do tratamento. Neste trabalho, foi utilizado o método

de emparelhamento de vizinho mais próximo (*nearest neighbour matching*, NNM) com reposição. Essa metodologia emparelha cada unidade tratada com uma unidade do grupo de controle que apresenta a pontuação de propensão mais próxima, podendo uma unidade de controle ser pareada com várias tratadas.

Diante da diferença na quantidade amostral entre o grupo de controle e o de tratamento, observada principalmente na Tabela 8, e da necessidade de maximizar o poder estatístico da análise, optou-se pelo pareamento com reposição em todos os cenários. A utilização de um pareamento sem reposição acarretaria que muitos domicílios do grupo de tratamento ficassem sem um par adequado, comprometendo a representatividade da amostra e a precisão das estimativas.

Neste sentido, a flexibilidade da reposição possibilitou a experimentação com diferentes razões de pareamento, o que resultou em três abordagens relacionadas ao método do vizinho mais próximo:

- **1:1 com reposição:** Cada unidade do grupo tratado é emparelhada com uma unidade do grupo de controle ($ratio = 1$). Com a opção de reposição ($replace = TRUE$), uma unidade de controle pode ser reutilizada em outros pareamentos.
- **1:3 com reposição:** Cada unidade do grupo tratado é emparelhada com até três unidades do grupo de controle ($ratio = 3$), útil em casos de desequilíbrio no tamanho dos grupos. Mantém-se a reposição para permitir o uso repetido das unidades de controle.
- **1:1 com reposição e tolerância de 5%:** Utiliza-se a configuração 1:1 com reposição, mas adiciona-se uma tolerância de 5% nas pontuações de propensão ($caliper = 0,05$), para que as unidades pareadas apresentem maior similaridade nas características observadas. Essa restrição evita pareamentos de unidades muito distantes em termos de propensão.

A Tabela 8 apresenta uma visão geral do processo de amostragem, pareamento e balanceamento entre os grupos de controle e tratamento em diferentes configurações de pareamento, sendo elas: (1:1) com reposição, (1:3) com reposição, e (1:1) com reposição e uma tolerância de 5%. Para cada tipo de pareamento, os itens descritos incluem:

- Amostra Total: Representa o número inicial de observações em ambos os grupos antes do pareamento.
- Após Pareamento com Estimador (ESS^3): Exibe o tamanho efetivo da amostra nos grupos de controle e tratamento após o pareamento.

³ ESS significa *Effective Sample Size* (Tamanho Efetivo da Amostra). Esse termo é utilizado para indicar o tamanho da amostra ajustado após a aplicação dos pesos no processo de balanceamento.

- Após Pareamento Sem Pesos (*Unweighted*): Refere-se ao tamanho da amostra após o pareamento, desconsiderando o uso de pesos, fornecendo uma estimativa sem ajustes.
- Não Pareadas: Indica a quantidade de observações que não encontraram correspondência, permanecendo não pareadas no grupo de controle ou tratamento.
- Balanceamento das Diferenças Médias ($<0,1$): Mostra o número de variáveis balanceadas e não balanceadas, considerando uma diferença de média inferior a 0,1 como indicativo de balanceamento. Em todas as configurações, as 13 variáveis avaliadas estão balanceadas.

Tabela 8 – Amostra, Pareamento e Balanceamento entre Grupos

	Controle	Tratamento
1:1 com reposição		
Amostra Total	3.186	3.744
Após Pareamento com estimador (ESS)	1.269,24	3.744
Após Pareamento sem pesos (<i>Unweighted</i>)	1.896	3.744
Não Pareadas	1.290	0
Balanceamento das Diferenças Médias (<0,1)	13 balanceadas, 0 não balanceadas	
1:3 com reposição		
Amostra Total	3.186	3.744
Após Pareamento com estimador (ESS)	1.979,07	3.744
Após Pareamento sem pesos (<i>Unweighted</i>)	2.899	3.744
Não Pareadas	287	0
Balanceamento das Diferenças Médias (<0,1)	13 balanceadas, 0 não balanceadas	
1:1 com reposição e tolerância de 5%		
Amostra Total	3.186	3.744
Após Pareamento com estimador (ESS)	1.275,88	3.727
Após Pareamento sem pesos (<i>Unweighted</i>)	1.894	3.727
Não Pareadas	1.292	17
Balanceamento das Diferenças Médias (<0,1)	13 balanceadas, 0 não balanceadas	

Fonte: Elaboração própria.

Nota 1: Método do estimador é *nearest* (vizinho mais próximo).

Esses resultados destacam a eficácia dos diferentes métodos de pareamento em reduzir as diferenças entre os grupos, proporcionando uma base sólida para inferências causais na análise do impacto.

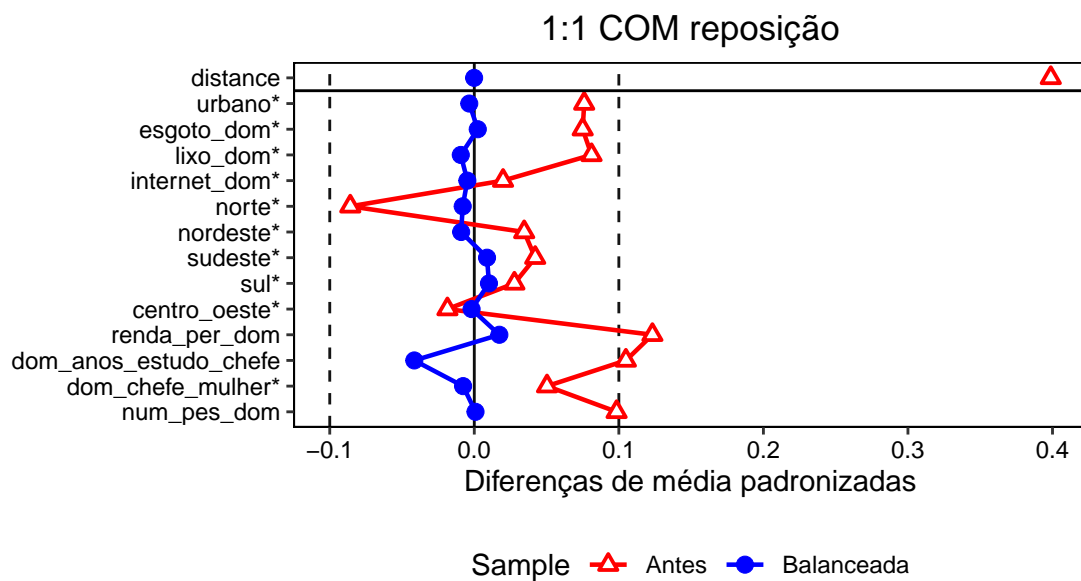
Ainda, é estabelecido o processo de balanceamento entre os grupos, nele é examinado as estatísticas das covariáveis em cada grupo após o pareamento. Se as médias, medianas ou outras medidas das covariáveis forem semelhantes entre os grupos tratado e controle, então, pode-se sugerir que o pareamento foi eficaz em equilibrar as características observáveis entre os grupos.

O resultado deste balanceamento corrobora com a indicação de que o emparelhamento foi bem-sucedido, pois todas as variáveis analisadas estão bem balanceadas, com diferenças médias inferiores a 0,1, ou seja, os grupos de controle e tratado são comparáveis em termos das variáveis consideradas após o emparelhamento.

Nas figuras 6, 7 e 8, são representações do balanceamento entre os grupos tratado e controle após o pareamento. O Apêndice A fornece um resumo detalhado das estatísticas das variáveis consideradas para o balanceamento, demonstrando as diferenças médias, razões de variância e outros indicadores importantes antes e após o pareamento.

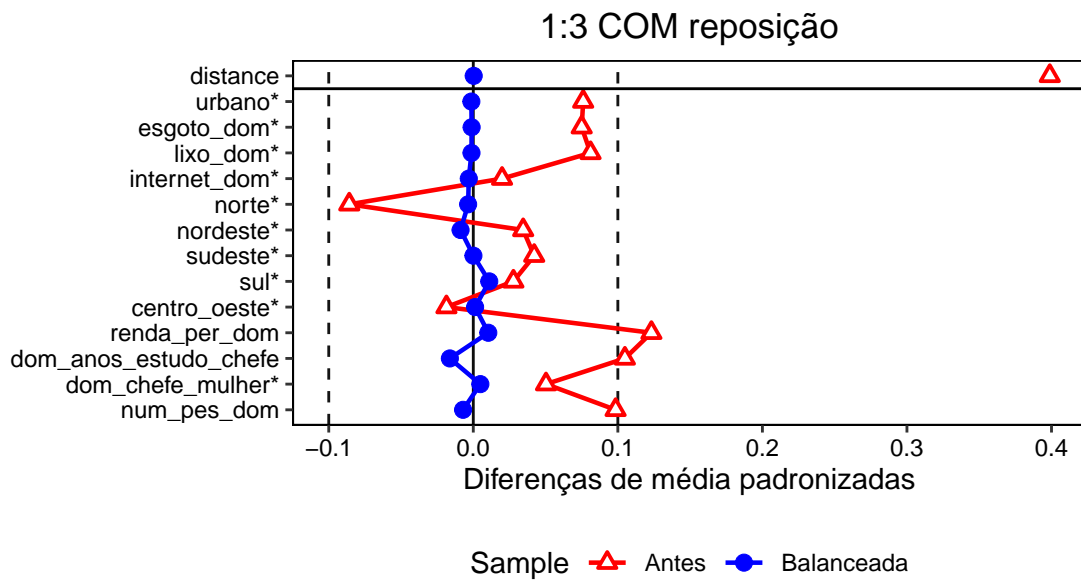
As diferenças de média padronizadas entre os grupos tratado e controle para cada covariável são representadas no eixo horizontal. A linha com pontos triangulares corresponde às covariáveis antes do balanceamento, enquanto a linha com pontos circulares refere-se ao cenário após o balanceamento. Observa-se que, após o procedimento, os valores das covariáveis se aproximam do eixo de referência zero, indicando um bom balanceamento.

Figura 6 – Gráfico antes e depois do balanceamento - 1:1 com reposição



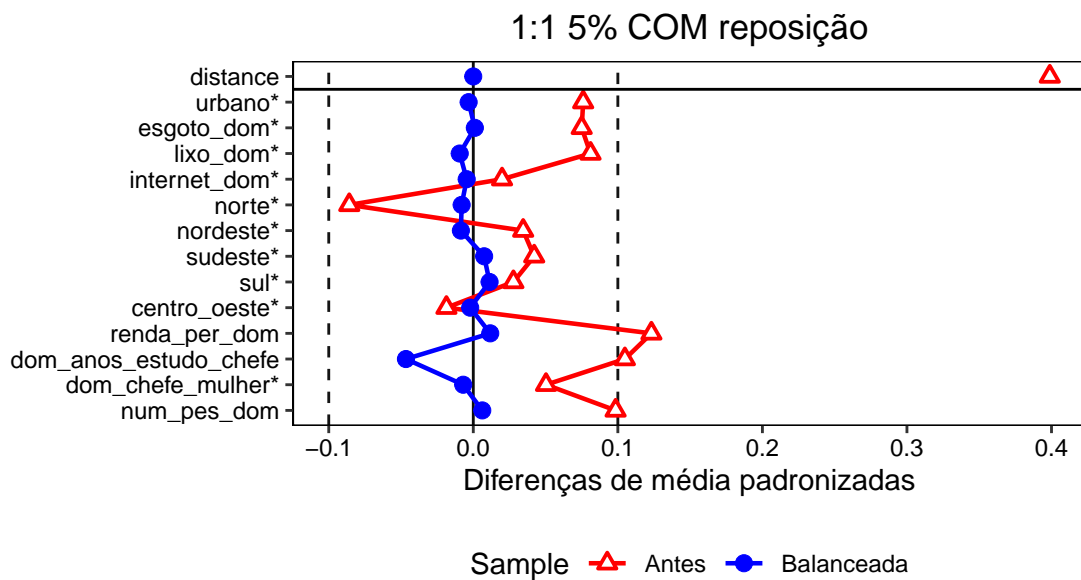
Fonte: Elaboração Própria.

Figura 7 – Gráfico antes e depois do balanceamento - 1:3 com reposição



Fonte: Elaboração Própria.

Figura 8 – Gráfico antes e depois do balanceamento - 1:1 com reposição e 5% de tolerância



Fonte: Elaboração Própria.

4.2.2.3 Distância de Mahalanobis

Apresentada no trabalho de Mahalanobis (1936), é uma medida que avalia a distância entre um ponto e uma distribuição em um conjunto de dados, independentemente da escala das variáveis. Diferentemente da distância Euclidiana, ela é invariante à escala e

considera as correlações entre as variáveis do conjunto de dados, utilizando cálculos vetoriais. É usado na estatística para avaliar como uma amostra conhecida e uma desconhecida se assemelham, na utilização do PSM, como neste caso, é importante para encontrar correspondências entre indivíduos tratados e não tratados com características próximas, sendo ajustada pela variância e correlação das variáveis.

Valendo-se dessa aplicação, foram ajustados os parâmetros com o método de vizinho mais próximo, a razão de emparelhamento $ratio = 1$, significando que uma unidade tratada será emparelhada com uma unidade não tratada, e, por fim, a reposição dos não tratados.

A Tabela 9 apresenta um resumo do balanceamento entre os grupos de controle e tratamento utilizando a Distância de Mahalanobis, ela é organizada em duas seções principais:

- Diferença de Médias: Demonstra o número de variáveis que alcançaram o balanceamento adequado, com uma diferença de médias menor que 0,1. Todas as 13 variáveis analisadas estão balanceadas.
- Tamanho da Amostra: A segunda seção descreve o tamanho da amostra para os grupos de controle e tratamento em diferentes estágios do processo de pareamento:
 - Amostra Total: O número inicial de observações nos grupos de controle e tratamento antes do pareamento.
 - Pareadas com Estimador (*ESS*): O tamanho efetivo da amostra após o pareamento, considerando o estimador ajustado.
 - Pareadas Sem Pesos (*Unweighted*): O tamanho da amostra pareada, desconsiderando o uso de pesos.
 - Não Pareadas: O número de observações que não foram pareadas, o que representa dados do grupo de controle que ficaram sem correspondência no grupo de tratamento.

Tabela 9 – Amostra e Balanceamento entre os Grupos, Distância de Mahalanobis

Diferença de Médias		Contagem
Variáveis Balanceadas ($< 0,1$)		
Variáveis Não Balanceadas ($> 0,1$)		0

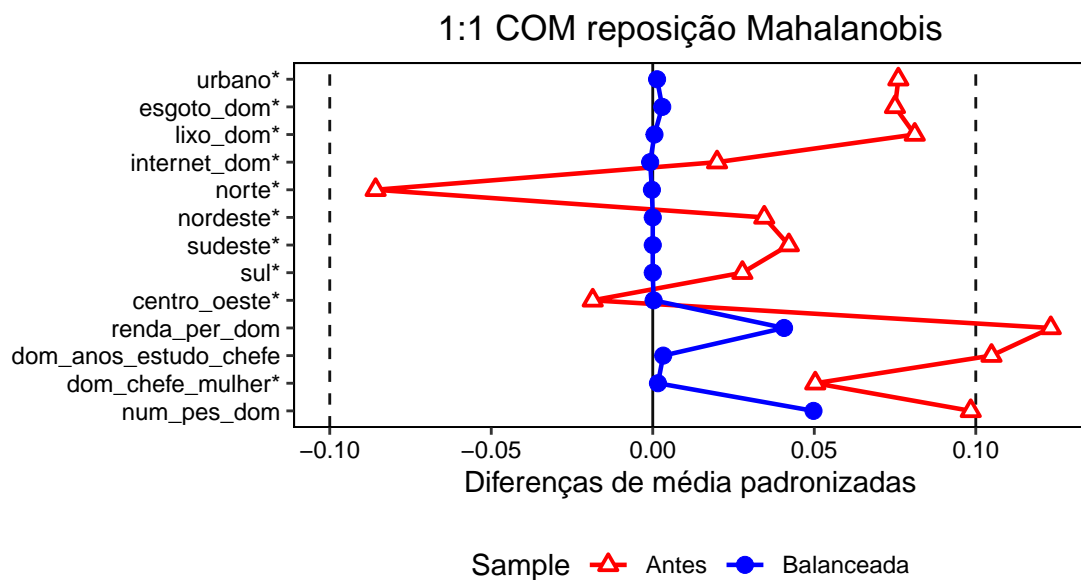
Tamanho da Amostra		
	Controle	Tratamento
Amostra Total	3.186	3.744
Pareadas com Estimador (ESS)	1.259,89	3.744
Pareadas Sem Pesos (<i>Unweighted</i>)	1.860	3.744
Não Pareadas	1.326	0

Fonte: Elaboração própria.

Nota 1: Método do estimador é *nearest* (vizinho mais próximo) e distância de Mahalanobis para medir a proximidade entre as unidades tratadas e de controle.

Também é providenciado a apresentação gráfica do resultado de balanceamento entre os grupos na Figura 9, demonstrando que o pareamento foi eficaz em equilibrar as características observáveis entre os grupos.

Figura 9 – Gráfico antes e depois do balanceamento - Mahalanobis



Fonte: Elaboração própria.

No Apêndice B, é fornecido um resumo detalhado das distâncias de Mahalanobis calculadas para as variáveis consideradas, permitindo avaliar a eficácia do pareamento.

4.2.2.4 Entropia

O modelo de Entropia, também conhecido como *Weighting Adjustment (WeightIt)*, é um método estatístico utilizado para ajustar os pesos de uma amostra para que seja representativa da população de interesse. O objetivo é reduzir a distorção entre a amostra e a população, minimizando a “desordem” entre as duas (GREIFER, 2024). Para este estudo, iremos manter a modelagem binomial para avaliar a relação entre a variável explicada e a variável preditora.

A Tabela 10 fornece uma síntese do balanceamento entre os grupos de controle e tratamento após o ajuste, utilizando como critério do modelo de Entropia. Ela está dividida em três seções:

- **Diferença de Médias:** Mostra quantas variáveis apresentaram diferenças de médias dentro de um limite aceitável ($< 0,1$), indicando que essas variáveis estão balanceadas entre os grupos. Todas as 13 variáveis analisadas estão balanceadas. A variável com a maior diferença de média é "Centro-Oeste", que ainda se encontra dentro do limite estabelecido para considerar o balanceamento.
- **Razão de Variância:** Verifica o balanceamento usando a razão de variâncias, com um limite de $< 3,33$. Das variáveis analisadas, três atendem a esse critério de balanceamento.
- **Tamanho Efetivo da Amostra (ESS):** Mostra os tamanhos das amostras de controle e tratamento antes e após o ajuste. Houve uma redução no tamanho efetivo da amostra de controle, mas manteve o grupo de tratamento inalterado, sugerindo um maior alinhamento entre os grupos em termos de variáveis analisadas.

Essa tabela é importante para confirmar se os grupos foram balanceados de forma adequada após o processo de pareamento, garantindo a comparabilidade entre eles.

Tabela 10 – Síntese do balanceamento entre os grupos, Entropia

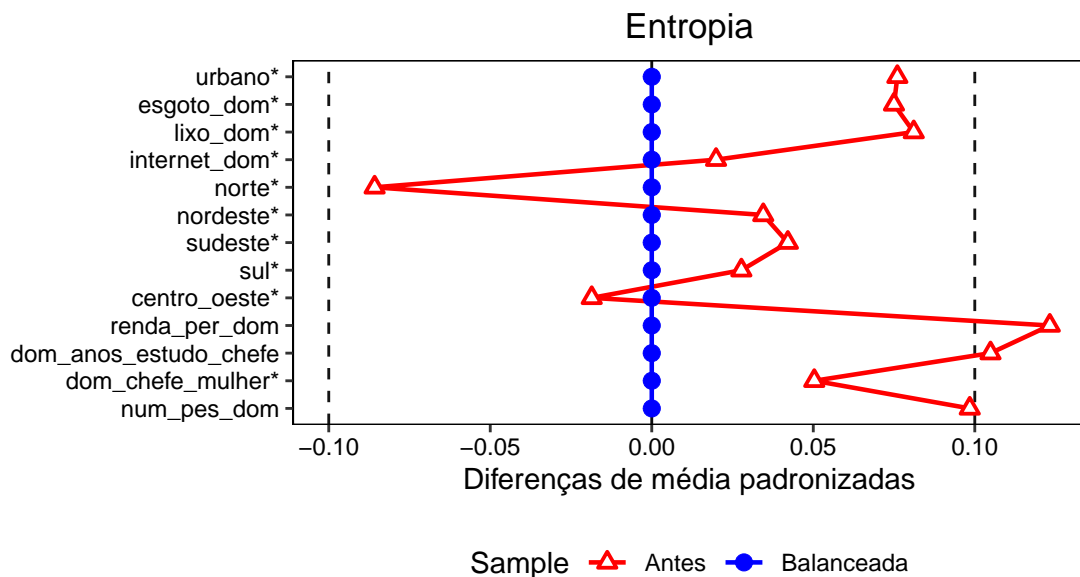
Diferença de Médias		
	Contagem	
Variáveis Balanceadas (< 0,1)	13	
Variáveis Não Balanceadas (> 0,1)	0	
Razão de Variância		
	Contagem	
Variáveis Balanceadas (< 3,333)	3	
Variáveis Não Balanceadas (> 3,333)	0	
Tamanho Efetivo da Amostra		
	Controle	Tratamento
Não Ajustado	3.186	3.744
Ajustado	2.786,78	3.744

Fonte: Elaboração própria.

Nota 1: Método utilizado é *ebal*, refere-se ao pareamento por balanceamento de entropia (*Entropy Balancing*)

O gráfico de balanceamento revela a eficácia do processo de pareamento entre os grupos, como apresentado na Figura 10. Em outras palavras, os grupos de controle e tratado são comparáveis em termos das variáveis consideradas após o emparelhamento.

Figura 10 – Gráfico antes e depois do balanceamento - Entropia



Fonte: Elaboração própria.

A tabela 21 no Apêndice C apresenta um sumário detalhado das medidas calculadas para as variáveis analisadas, incluindo as diferenças não ajustadas, as razões de variância

e as diferenças ajustadas, além dos limites médios e das razões de variância ajustadas, mostrando as comparações antes e após o pareamento por entropia.

4.3 Sensibilidade

A análise de sensibilidade foi conduzida de acordo com os procedimentos descritos por Rosenbaum e Rubin (1983), utilizando o método *binarysens*, para avaliar se fatores não observados poderiam impactar significativamente a estimativa do efeito do tratamento. Os resultados indicaram que, na ausência de viés ($\text{Gamma} = 1.0$), não há efeito significativo, com ambos os limites superior e inferior iguais a zero.

À medida que o parâmetro Gamma aumenta, indicando maior probabilidade de viés devido a fatores não observados, observa-se que os resultados tornam-se sensíveis a partir de $\text{Gamma} = 1.3$, momento em que o limite superior se aproxima de 1 (ou seja, marginalmente significativo). A partir de $\text{Gamma} = 1.4$, o limite superior atinge o valor máximo de 1.0, sugerindo que uma pequena magnitude de viés não observado pode alterar os resultados.

Adicionalmente, foi aplicado o teste de ranqueamento *Wilcoxon Signed Rank*, conforme recomendado por Woolson (2005), para investigar possíveis desequilíbrios entre os grupos tratado e controle em relação a fatores não observados. O valor obtido foi 1, indicando que não há diferenças significativas entre os grupos em termos desses elementos, o que reforça a confiabilidade e robustez dos resultados apresentados.

Portanto, os resultados da análise de sensibilidade mostram que os efeitos estimados são sensíveis a um viés não observado de magnitude $\text{Gamma} = 1.3$, conforme detalhado na Tabela 22, no Apêndice D.

4.4 Efeito Médio do Tratamento sobre os Tratados

Originalmente denominado *Average Treatment Effect on the Treated (ATT)*, é uma medida fundamental em estudos que utilizam o PSM para avaliar o impacto médio de uma intervenção, especificamente no grupo de indivíduos que efetivamente receberam o tratamento. Esse conceito é crucial para entender como a intervenção influencia aqueles que a utilizaram, comparando-os com um grupo de controle que, embora não tenha recebido o tratamento, é similar em características relevantes.

Neste estudo, a variável de interesse *Tratado* refere-se ao domicílio com a presença de criança em creche ou pré-escola residente no mesmo lar de jovens com idade entre 15 e 29 anos. Os coeficientes estimados para esta variável mostram como a probabilidade de um domicílio possuir criança em instituições educacionais muda em relação a situação dos jovens.

A Tabela 11 mostra os coeficientes estimados para a variável dependente *Domicílios com nem-nem* em diferentes parâmetros do PSM. Em todos os modelos, o coeficiente para a variável *Tratado* (Domicílios com criança em escola) é negativo e estatisticamente significativo, indicando que o tratamento está associado a uma redução na probabilidade de um domicílio ter jovens nessa condição.

Os coeficientes variam entre -0,442 a -0,504, todos com significância a 1%, sugerindo um efeito consistente do tratamento sobre a redução da condição do domicílio possuir nem-nem. Os valores entre parênteses representam os erros padrão das estimativas, indicando a precisão com que os coeficientes foram estimados.

Tabela 11 – Resultado das Regressões

	Variável Dependente: Domicílios com Nem Nem				
	(1)	(2)	(3)	(4)	(5)
Tratado	-0,504*** (0,057)	-0,452*** (0,050)	-0,500*** (0,057)	-0,442*** (0,057)	-0,499*** (0,049)
Observações	5.640	6.643	5.621	5.604	6.930
<p><i>Notas:</i> *p<0,1; **p<0,05; ***p<0,01 (1): PSM-ATT - NNM 1:1 COM Reposição. (2): PSM-ATT - NNM 1:3 COM Reposição. (3): PSM-ATT - NNM 1:1 COM Reposição 5% Tolerância. (4): PSM-ATT - Mahalanobis 1:1 COM Reposição. (5): Entropia.</p>					

Fonte: Elaboração própria.

Outra forma de entender o impacto do tratamento é calculando a razão de chances *Odds Ratio* associada aos coeficientes da regressão logística. Essa abordagem nos permite quantificar a mudança nas chances do evento, considerando a intervenção, oferecendo uma visão detalhada do efeito da variável *Tratado* (presença de criança matriculada em creche ou pré-escola no domicílio). Para calcular a razão de chances dos coeficientes, utiliza-se a função exponencial do coeficiente, conforme descrito na equação 4.1.

Para transformar o *Odds Ratio* em termos percentuais e interpretar a redução ou aumento nas chances, utilizamos a fórmula:

$$Percentual = (1 - OddsRatio) * 100 \quad (4.2)$$

Ela expressa a mudança nas chances do evento (domicílio ter um jovem na condição de nem-nem) após a introdução da variável explicativa. A interpretação disso é que, ao introduzirmos a variável *Tratado*, que indica se o domicílio possui criança matriculada em

creche ou pré-escola, as chances de um domicílio ter um jovem nem-nem diminuem em uma certa porcentagem.

A Tabela 12 apresenta a redução percentual na probabilidade de um domicílio ter jovens nem-nem em decorrência da presença de crianças em creches ou pré-escola no mesmo lar. Por exemplo, no modelo (1), a razão de chances de 0,604 indica que, para cada domicílio com criança em creche ou pré-escola, a probabilidade de ter um jovem sem trabalhar e nem estudar é 39,59% menor em comparação com um domicílio sem essa característica, mantendo as demais condições constantes. Isso sugere que a matrícula de crianças em creches pode ser um fator de proteção contra a condição de nem-nem, possivelmente ao liberar o tempo dos responsáveis para buscar emprego e proporcionar um ambiente estimulante para as crianças.

Tabela 12 – Resultados das Regressões em percentual

Modelos	Coefficiente	Odds Ratio	Percentual (%)
(1)	-0,504	0,604	-39,59
(2)	-0,452	0,637	-36,33
(3)	-0,500	0,607	-39,33
(4)	-0,442	0,643	-35,69
(5)	-0,499	0,607	-39,28

(1): PSM-ATT - NNM 1:1 COM Reposição.
(2): PSM-ATT - NNM 1:3 COM Reposição.
(3): PSM-ATT - NNM 1:1 COM Reposição 5% Tolerância.
(4): PSM-ATT - Mahalanobis 1:1 COM Reposição.
(5): Entropia.

Fonte: Elaboração própria.

Com um percentual médio de redução na probabilidade de um domicílio ter jovens nem-nem, devido à presença de crianças em creches, de 38,05%, os resultados evidenciam que a matrícula de crianças em creches pode ser uma estratégia eficaz para facilitar a inserção dos jovens no mercado de trabalho ou para mantê-los no sistema educacional. Além disso, esse impacto da educação infantil traz oportunidades aos pais ou responsáveis, oferecendo-lhes tempo para procurar emprego ou estudar, o que pode contribuir para a melhoria das condições socioeconômicas da família.

5 Considerações finais

Este estudo teve como objetivo principal avaliar o impacto da matrícula de crianças em creches ou pré-escolas na probabilidade dos domicílios possuírem jovens que não trabalham nem estudam. Utilizando o método de *Propensity Score Matching* e realizando análises de robustez, os resultados indicaram uma redução na probabilidade de o domicílio possuir jovens nem-nem quando há criança matriculada em creche ou pré-escola na mesma residência.

Foi observado, através dos dados do IBGE, que 22,3% dos jovens entre 15 e 29 anos estavam fora da escola e do mercado de trabalho em 2022, com um número expressivo de mulheres afetadas por essa condição. Das 10,9 milhões de pessoas nessa situação, 6,9 milhões são mulheres, representando 63,4% do total de jovens nem-nem.

As estatísticas regionais reforçam que as regiões Norte e Nordeste concentram a maior parte dos domicílios com jovens nem-nem, com índices de 53,5% e 58,2%, respectivamente. Além disso, a análise de motivos para a condição nem-nem revela que a falta de acesso ao trabalho na localidade e a necessidade de cuidar de filhos são fatores predominantes.

Os resultados deste estudo são consistentes com a literatura, como achados de Campos e Silva (2020), que indicam um aumento na participação das mães no mercado de trabalho formal após o acesso à creche. Nossos resultados sugerem um efeito mais direto na redução da condição de nem-nem dos jovens. Essa diferença pode ser explicada pelo fato de que o presente estudo se concentrou especificamente na relação entre os domicílios com matrícula em creches e os domicílios com jovens nem-nem.

Esta pesquisa corroborou essa realidade, ao mostrar que a presença de crianças em creches reduz, em média, 38,05% a probabilidade de um domicílio ter jovens nem-nem.

Assim, possíveis mecanismos podem explicar tal relação. Por exemplo, com a matrícula em creches, há uma possível liberação do tempo dos responsáveis para buscar emprego ou qualificação profissional, aumentando a renda familiar e proporcionando um modelo de trabalho para os filhos. Além de que, no longo prazo, a escolarização na primeira infância proporciona um ambiente estimulante que favorece o desenvolvimento de habilidades cognitivas e socioemocionais, preparando as crianças para o sucesso escolar e profissional.

Destaca-se que para realizar o pareamento, foi utilizado o algoritmo de *matching* pelo método de vizinho mais próximo, que emparelha o grupo de controle com o grupo tratado com base na pontuação de propensão mais próxima, com reposição. Foram aplicadas modelagens utilizando a distância de Mahalanobis e o método de balanceamento por entropia. Além disso, a análise de sensibilidade, realizada por meio do teste de ranquea-

mento de Wilcoxon e do método proposto por Rosenbaum, não evidenciou desequilíbrios significativos entre os grupos tratados e controle em relação a fatores não observados, reforçando a robustez dos resultados obtidos.

Embora os resultados pareçam promissores, é importante ressaltar algumas limitações deste estudo. A generalização dos resultados pode ser limitada devido à natureza da amostra e à possibilidade de viés de seleção, mesmo após o controle por fatores observados e não observados. Além disso, o estudo não permite estabelecer uma relação de causalidade direta entre a matrícula em creches e a redução da condição de nem-nem, sendo necessário aprofundar a investigação sobre os mecanismos subjacentes a essa relação.

A investigação realizada contribui para a literatura ao explorar o impacto da matrícula em creches na redução da condição de nem-nem em nível domiciliar, uma questão ainda pouco explorada na pesquisa brasileira. Utilizando o método de PSM, os resultados obtidos oferecem evidências sólidas sobre a relação entre o acesso à educação infantil e a diminuição da probabilidade dos domicílios possuírem jovens nessa condição. Essas evidências ressaltam a importância de políticas públicas que ampliem o acesso à educação infantil, promovendo o desenvolvimento humano.

Referências

- BALAN, M. Economic and social consequences triggered by the neet youth. *Knowledge Horizons - Economics*, v. 8, p. 80–87, 6 2016.
- BAUERNSCHUSTER, S.; SCHLOTTER, M. Public child care and mothers' labor supply—evidence from two quasi-experiments. *Journal of Public Economics*, v. 123, n. C, p. 1–16, 2015. Disponível em: <<https://EconPapers.repec.org/RePEc:eee:pubeco:v:123:y:2015:i:c:p:1-16>>.
- BERLINSKI, S.; GALIANI, S. The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment. *Labour Economics*, v. 14, n. 3, p. 665–680, June 2007. Disponível em: <<https://ideas.repec.org/a/eee/labeco/v14y2007i3p665-680.html>>.
- BICK, A. The quantitative role of child care for female labor force participation and fertility. *Journal of the European Economic Association*, Oxford University Press, v. 14, n. 3, p. 639–668, 2016. ISSN 15424766, 15424774. Disponível em: <<http://www.jstor.org/stable/43965320>>.
- BRASIL. *Constituição da República Federativa do Brasil*. Brasília, DF: Senado Federal, 1988. Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/518231/CF88_Livro_EC91_2016.pdf>.
- BRASIL. *Emenda Constitucional nº 53, de 19 de dezembro de 2006*. Brasília, DF: [s.n.], 2006. Dá nova redação aos arts. 7º, 23, 30, 206, 208, 211 e 212 da Constituição Federal e ao art. 60 do Ato das Disposições Constitucionais Transitórias.
- BRASIL. *Emenda Constitucional nº 59, de 11 de novembro de 2009*. Brasília, DF: [s.n.], 2009. Dá nova redação aos incisos I e VII do art. 208, ampliando a obrigatoriedade e a abrangência da educação básica.
- BRASIL. *Lei nº 12.796, de 4 de abril de 2013*. Brasília, DF: [s.n.], 2013. Altera a Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional.
- BRASIL. Lei nº 12.852 de 05 de agosto de 2013. *Diário Oficial da República Federativa do Brasil*, 2013. Disponível em: <<https://legis.senado.leg.br/norma/591300/publicacao/15637530>>.
- BRASIL. *Lei nº 13.005, de 25 de junho de 2014*. Brasília, DF: [s.n.], 2014. Aprova o Plano Nacional de Educação – PNE e dá outras providências.
- BRILLI, Y.; BOCA, D. D.; PRONZATO, C. D. Does child care availability play a role in maternal employment and children's development? evidence from Italy. *Review of Economics of the Household*, Springer, v. 14, p. 27–51, 2016.
- CAMPOS, A. L. M. de; SILVA, M. M. da C. Access to early childhood education and time allocation of mothers. 2020.

- COLES, B.; BRADSHAW, J.; HUTTON, S. Literature review of the costs of being 'not in education, employment or training' at age 16-18. Stationery Office, v. 347, 6 2002. ISBN: 1-84185-743-2.
- CORREA, E.; COMIM, F.; TAI, S. H. T. Impactos da creche na primeira infância: efeitos dependendo das características da família e do grau de exposição ao centro de cuidado. *Anais da Associação Nacional de Pós-Graduação em Economia ANPEC*, 2014.
- CURI, A. Z.; FILHO, N. A. M. A relação entre educação primária, salários, escolaridade e proficiência escolar no brasil. *Estudos Econômicos*, SciELO Brasil, v. 39, p. 811–850, 2009.
- DICKS, A.; LEVELS, M.; VELDEN, R. der; MILLS, M. C. How young mothers rely on kin networks and formal childcare to avoid becoming neet in the netherlands. *Frontiers in Sociology*, Frontiers Media SA, v. 6, 2022. DOI: <https://doi.org/10.3389/fsoc.2021.78753210.3389/fsoc.2021.787532>.
- FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006.
- GIVORD, P.; MARBOT, C. Does the cost of child care affect female labor market participation? an evaluation of a french reform of childcare subsidies. *Labour Economics*, v. 36, p. 99–111, 2015. ISSN 0927-5371. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0927537115000779>>.
- GREIFER, N. *Using WeightIt to Estimate Balancing Weights*. 2024. Source: vignettes/WeightIt.Rmd. Disponível em: <<https://ngreifer.github.io/WeightIt/articles/WeightIt.html>>.
- GUIMARÃES, N. A.; MARTELETO, L.; ALVES. Trajetórias e transições: os múltiplos e difíceis caminhos dos jovens brasileiros no mercado de trabalho. *Brasília*, 2016. DOI: 10.25091.
- IBGE. *Pesquisa nacional por amostra de domicílios contínua: Manual Básico da Entrevista*. Rio de Janeiro: IBGE, 2022. 262 p. <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=55642>>. Disponível somente em meio digital.
- IBGE. *Síntese de Indicadores Sociais: Uma análise das condições de vida da população brasileira*. [s.n.], 2023. 152 p. ISBN 978-85-240-4598-1. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/protecao-social/9221-sintese-de-indicadores-sociais.html>>.
- JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. [S.l.]: John Wiley & Sons, 2013.
- MAHALANOBIS, P. C. Mahalanobis distance. In: *Proceedings National Institute of Science of India*. [S.l.: s.n.], 1936. v. 49, n. 2, p. 234–256.
- MASCHERINI, M.; SALVATORE, L.; MEIERKORD, A.; JUNGBLUT, J.-M. Neets: Young people not in employment, education or training: Characteristics, costs and policy responses in europe. Publications Office of the European Union Luxembourg, 2012. ISBN: 978-92-897-1094-7 DOI: <https://doi.org/10.2806/4157810.2806/41578>.
- PINTO, D. S. Cristine Campos de X.; GUIMARÃES, C. The impact of daycare attendance on math test scores for a cohort of fourth graders in brazil. *The Journal of Development Studies*, Routledge, v. 53, n. 9, p. 1335–1357, 2017.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, Oxford University Press, v. 70, n. 1, p. 41–55, 1983.

RSTUDIO. *RStudio: Integrated Development Environment for R*. 2020. <<https://www.rstudio.com>>.

SHIRASU, M. R.; ARRAES, R. de A. Avaliação dos custos econômicos associados aos jovens nem-nem no brasil. *Brazilian Journal of Political Economy*, SciELO Brasil, v. 40, p. 161–182, 2020. DOI: <https://doi.org/10.1590/0101-31572020-2902>.

SILVA, E. R. A. da; VAZ, F. M. Os jovens que não trabalham e não estudam no contexto da pandemia da covid-19 no brasil. In: _____. [S.l.]: Instituto de Pesquisa Econômica Aplicada (Ipea), 2022. p. 335–354. ISBN: 978-65-5635-042-4 DOI: <http://dx.doi.org/10.38116/978-65-5635-042-4/capitulo15>10.38116/978-65-5635-042-4.

SILVA, M. B. *Mulher e trabalho no Brasil: efeitos da creche na jornada de trabalho e na renda laboral*. 2019.

STEINER, P. M.; COOK, D. *Matching and Propensity Scores*. [S.l.]: Oxford University Press, 2013. 237-258 p.

VASCONCELOS, A. M.; RIBEIRO, F. G.; GRIEBELER, M. de C.; CARRARO, A. Programa bolsa família e geração "nem-nem": Evidências para o brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 71, p. 233–257, 2017.

WOOLSON, R. F. Wilcoxon signed-rank test. In: _____. [S.l.]: Wiley, 2005.

APÊNDICE A – Pareamento por vizinho mais próximo

Tabela 13 – Sumário antes do balanceamento 1:1 com reposição.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Distância	0,556	0,521	0,399	0,839	0,105	0,156	NA
Urbano	0,725	0,649	0,170	NA	0,076	0,076	NA
Esgoto no domicílio	0,400	0,325	0,153	NA	0,075	0,075	NA
Lixo no domicílio	0,834	0,753	0,218	NA	0,081	0,081	NA
Internet no domicílio	0,895	0,875	0,065	NA	0,020	0,020	NA
Norte	0,151	0,237	-0,240	NA	0,086	0,086	NA
Nordeste	0,391	0,356	0,071	NA	0,035	0,035	NA
Sudeste	0,218	0,176	0,102	NA	0,042	0,042	NA
Sul	0,150	0,122	0,078	NA	0,028	0,028	NA
Centro Oeste	0,090	0,109	-0,065	NA	0,019	0,019	NA
Renda <i>per capita</i> domiciliar	807,077	698,695	0,123	1,242	0,040	0,083	NA
Anos de estudo do chefe de família	9,447	9,023	0,105	0,967	0,025	0,041	NA
Gênero do chefe (mulher)	0,575	0,524	0,102	NA	0,050	0,050	NA
Número de pessoas no domicílio	4,571	4,413	0,098	1,116	0,011	0,069	NA

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

Tabela 14 – Sumário após o balanceamento 1:1 com reposição.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Distância	0,556	0,556	-0,000	0,999	0,000	0,003	0,001
Urbano	0,725	0,729	-0,008	NA	0,003	0,003	0,728
Esgoto no domicílio	0,400	0,397	0,005	NA	0,002	0,002	0,787
Lixo no domicílio	0,834	0,844	-0,025	NA	0,009	0,009	0,504
Internet no domicílio	0,895	0,900	-0,016	NA	0,005	0,005	0,572
Norte	0,151	0,159	-0,022	NA	0,008	0,008	0,446
Nordeste	0,391	0,400	-0,019	NA	0,009	0,009	0,908
Sudeste	0,218	0,209	0,021	NA	0,009	0,009	0,740
Sul	0,150	0,140	0,028	NA	0,010	0,010	0,668
Centro Oeste	0,090	0,092	-0,007	NA	0,002	0,002	0,522
Renda <i>per capita</i> domiciliar	807,077	791,819	0,017	0,958	0,012	0,044	0,714
Anos de estudo do chefe de família	9,447	9,615	-0,042	1,045	0,010	0,024	1,028
Gênero do chefe (mulher)	0,575	0,583	-0,016	NA	0,008	0,008	0,864
Número de pessoas no domicílio	4,571	4,569	0,001	0,915	0,006	0,045	0,914

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

Tabela 15 – Sumário antes do balanceamento 1:3 com reposição.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Distância	0,556	0,521	0,399	0,839	0,105	0,156	NA
Urbano	0,725	0,649	0,170	NA	0,076	0,076	NA
Esgoto no domicílio	0,400	0,325	0,153	NA	0,075	0,075	NA
Lixo no domicílio	0,834	0,753	0,218	NA	0,081	0,081	NA
Internet no domicílio	0,895	0,875	0,065	NA	0,020	0,020	NA
Norte	0,151	0,237	-0,240	NA	0,086	0,086	NA
Nordeste	0,391	0,356	0,071	NA	0,035	0,035	NA
Sudeste	0,218	0,176	0,102	NA	0,042	0,042	NA
Sul	0,150	0,122	0,078	NA	0,028	0,028	NA
Centro Oeste	0,090	0,109	-0,065	NA	0,019	0,019	NA
Renda <i>per capita</i> domiciliar	807,077	698,695	0,123	1,242	0,040	0,083	NA
Anos de estudo do chefe de família	9,447	9,023	0,105	0,967	0,025	0,041	NA
Gênero do chefe (mulher)	0,575	0,524	0,102	NA	0,050	0,050	NA
Número de pessoas no domicílio	4,571	4,413	0,098	1,116	0,011	0,069	NA

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

Tabela 16 – Sumário após o balanceamento 1:3 com reposição.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Distância	0,556	0,556	0,000	1,001	0,000	0,003	0,003
Urbano	0,725	0,727	-0,003	NA	0,001	0,001	0,721
Esgoto no domicílio	0,400	0,401	-0,002	NA	0,001	0,001	0,806
Lixo no domicílio	0,834	0,836	-0,003	NA	0,001	0,001	0,503
Internet no domicílio	0,895	0,898	-0,010	NA	0,003	0,003	0,582
Norte	0,151	0,155	-0,010	NA	0,004	0,004	0,446
Nordeste	0,391	0,400	-0,018	NA	0,009	0,009	0,921
Sudeste	0,218	0,218	0,000	NA	0,000	0,000	0,757
Sul	0,150	0,139	0,031	NA	0,011	0,011	0,662
Centro Oeste	0,090	0,089	0,004	NA	0,001	0,001	0,514
Renda <i>per capita</i> domiciliar	807,077	797,946	0,010	0,871	0,015	0,041	0,716
Anos de estudo do chefe de família	9,447	9,513	-0,016	1,022	0,005	0,012	1,040
Gênero do chefe (mulher)	0,575	0,570	0,010	NA	0,005	0,005	0,908
Número de pessoas no domicílio	4,571	4,582	-0,007	0,960	0,007	0,036	0,934

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

Tabela 17 – Sumário antes do balanceamento 1:1 com reposição e tolerância de 5%.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Distância	0,556	0,521	0,399	0,839	0,105	0,156	NA
Urbano	0,725	0,649	0,170	NA	0,076	0,076	NA
Esgoto no domicílio	0,400	0,325	0,153	NA	0,075	0,075	NA
Lixo no domicílio	0,834	0,753	0,218	NA	0,081	0,081	NA
Internet no domicílio	0,895	0,875	0,065	NA	0,020	0,020	NA
Norte	0,151	0,237	-0,240	NA	0,086	0,086	NA
Nordeste	0,391	0,356	0,071	NA	0,035	0,035	NA
Sudeste	0,218	0,176	0,102	NA	0,042	0,042	NA
Sul	0,150	0,122	0,078	NA	0,028	0,028	NA
Centro Oeste	0,090	0,109	-0,065	NA	0,019	0,019	NA
Renda <i>per capita</i> domiciliar	807,077	698,695	0,123	1,242	0,040	0,083	NA
Anos de estudo do chefe de família	9,447	9,023	0,105	0,967	0,025	0,041	NA
Gênero do chefe (mulher)	0,575	0,524	0,102	NA	0,050	0,050	NA
Número de pessoas no domicílio	4,571	4,413	0,098	1,116	0,011	0,069	NA

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

Tabela 18 – Sumário após o balanceamento 1:1 com reposição e tolerância de 5%.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Distância	0,555	0,555	-0,000	0,999	0,000	0,003	0,001
Urbano	0,725	0,728	-0,007	NA	0,003	0,003	0,731
Esgoto no domicílio	0,398	0,397	0,002	NA	0,001	0,001	0,785
Lixo no domicílio	0,834	0,843	-0,025	NA	0,009	0,009	0,506
Internet no domicílio	0,895	0,900	-0,015	NA	0,005	0,005	0,574
Norte	0,152	0,160	-0,022	NA	0,008	0,008	0,448
Nordeste	0,391	0,400	-0,018	NA	0,009	0,009	0,908
Sudeste	0,218	0,210	0,018	NA	0,008	0,008	0,741
Sul	0,150	0,138	0,032	NA	0,011	0,011	0,668
Centro Oeste	0,090	0,092	-0,008	NA	0,002	0,002	0,523
Renda <i>per capita</i> domiciliar	792,407	782,093	0,012	0,927	0,012	0,043	0,702
Anos de estudo do chefe de família	9,441	9,630	-0,047	1,052	0,011	0,025	1,023
Gênero do chefe (mulher)	0,574	0,581	-0,014	NA	0,007	0,007	0,866
Número de pessoas no domicílio	4,557	4,547	0,006	0,947	0,006	0,045	0,904

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

APÊNDICE B – Distância de Mahalanobis

Tabela 19 – Sumário antes do balanceamento 1:1 com reposição - Distância de Mahalanobis.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Urbano	0,725	0,649	0,170	NA	0,076	0,076	NA
Esgoto no domicílio	0,400	0,325	0,153	NA	0,075	0,075	NA
Lixo no domicílio	0,834	0,753	0,218	NA	0,081	0,081	NA
Internet no domicílio	0,895	0,875	0,065	NA	0,020	0,020	NA
Norte	0,151	0,237	-0,240	NA	0,086	0,086	NA
Nordeste	0,391	0,356	0,071	NA	0,035	0,035	NA
Sudeste	0,218	0,176	0,102	NA	0,042	0,042	NA
Sul	0,150	0,122	0,078	NA	0,028	0,028	NA
Centro Oeste	0,090	0,109	-0,065	NA	0,019	0,019	NA
Renda <i>per capita</i> domiciliar	807,077	698,695	0,123	1,242	0,040	0,083	NA
Anos de estudo do chefe de família	9,447	9,023	0,105	0,967	0,025	0,041	NA
Gênero do chefe (mulher)	0,575	0,524	0,102	NA	0,050	0,050	NA
Número de pessoas no domicílio	4,571	4,413	0,098	1,116	0,011	0,069	NA

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

Tabela 20 – Sumário após o balanceamento 1:1 com reposição - Distância de Mahalanobis.

Variável	Média Tratados	Média Controle	Dif. de Média Padronizada	Razão de Variância	eCDF Média	eCDF Máx	Dist. Par Std.
Urbano	0,725	0,724	0,003	NA	0,001	0,001	0,004
Esgoto no domicílio	0,400	0,397	0,006	NA	0,003	0,003	0,011
Lixo no domicílio	0,834	0,834	0,001	NA	0,001	0,001	0,001
Internet no domicílio	0,895	0,896	-0,003	NA	0,001	0,001	0,003
Norte	0,151	0,151	-0,001	NA	0,000	0,000	0,001
Nordeste	0,391	0,391	0,000	NA	0,000	0,000	0,001
Sudeste	0,218	0,218	0,000	NA	0,000	0,000	0,001
Sul	0,150	0,150	0,000	NA	0,000	0,000	0,000
Centro Oeste	0,090	0,090	0,001	NA	0,000	0,000	0,001
Renda <i>per capita</i> domiciliar	807,077	771,322	0,041	1,105	0,018	0,040	0,192
Anos de estudo do chefe de família	9,447	9,434	0,003	1,079	0,008	0,022	0,176
Gênero do chefe (mulher)	0,575	0,573	0,003	NA	0,002	0,002	0,016
Número de pessoas no domicílio	4,571	4,491	0,050	1,246	0,006	0,019	0,138

Fonte: Elaboração própria.

Notas: *eCDF Média*: Função de Distribuição Empírica Acumulada Média; *eCDF Máx*: Valor Máximo da Função de Distribuição Empírica Acumulada; *Dist. Par Std.*: Distância Padrão entre Pares.

APÊNDICE C – Entropia

Tabela 21 – Sumário das Variáveis com Diferenças Ajustadas - Entropia.

Variável	Dif. Un	Razão de Variância Un	Dif. Ajustada	Limite	Razão de Variância Ajustada
Urbano	0,0760	.	0	Balanceada, <0,1	.
Esgoto no domicílio	0,0750	.	0	Balanceada, <0,1	.
Lixo no domicílio	0,0811	.	0	Balanceada, <0,1	.
Internet no domicílio	0,0199	.	-0	Balanceada, <0,1	.
Norte	-0,0858	.	0	Balanceada, <0,1	.
Nordeste	0,0345	.	0	Balanceada, <0,1	.
Sudeste	0,0421	.	0	Balanceada, <0,1	.
Sul	0,0277	.	0	Balanceada, <0,1	.
Centro Oeste	-0,0186	.	-0	Balanceada, <0,1	.
Renda <i>per capita</i> domiciliar	0,1232	1,2418	-0	Balanceada, <0,1	0,6401
Anos de estudo do chefe de família	0,1049	0,9667	0	Balanceada, <0,1	1,0055
Gênero do chefe (mulher)	0,0503	.	0	Balanceada, <0,1	.
Número de pessoas no domicílio	0,0984	1,1164	0	Balanceada, <0,1	0,9817

Fonte: Elaboração própria.

Notas: *Dif. Un*: Diferença não ajustada; *Razão de Variância Un*: Razão de variância não ajustada; *Dif. Ajustada*: Diferença ajustada; *M. Threshold*: Limite médio; *Razão de Variância Ajustada*: Razão de variância ajustada; *V. Threshold*: Limite de Variância.

APÊNDICE D – Sensibilidade de Rosenbaum

Tabela 22 – Resultados do Teste de Sensibilidade de Rosenbaum

Gamma	Lower Bound	Upper Bound
1.0	0	0.00000
1.1	0	0.00000
1.2	0	0.00012
1.3	0	0.98840
1.4	0	1.00000
1.5	0	1.00000
1.6	0	1.00000
1.7	0	1.00000
1.8	0	1.00000
1.9	0	1.00000
2.0	0	1.00000
2.1	0	1.00000
2.2	0	1.00000
2.3	0	1.00000
2.4	0	1.00000
2.5	0	1.00000
2.6	0	1.00000
2.7	0	1.00000
2.8	0	1.00000
2.9	0	1.00000
3.0	0	1.00000

Fonte: Elaboração própria.

Nota 1: *Gamma* é a razão das chances de atribuição diferencial ao tratamento devido a fatores não observados.

Nota 2: *Lower Bound* refere-se ao limite inferior da estimativa de sensibilidade.

Nota 3: *Upper Bound* refere-se ao limite superior da estimativa de sensibilidade.