



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE APERFEIÇOAMENTO DE ECONOMISTAS DO NORDESTE - CAEN
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

EDUARDO CÂNDIDO DE OLIVEIRA

ENSAIOS EM ECONOMETRIA, ESTATÍSTICA E APRENDIZADO DE MÁQUINA
APLICADOS À VIOLÊNCIA DOMÉSTICA

FORTALEZA

2024

EDUARDO CÂNDIDO DE OLIVEIRA

ENSAIOS EM ECONOMETRIA, ESTATÍSTICA E APRENDIZADO DE MÁQUINA
APLICADOS À VIOLÊNCIA DOMÉSTICA

Tese apresentada ao Programa de Pós-Graduação em Economia do Centro de Aperfeiçoamento de Economistas do Nordeste - CAEN da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Economia. Área de Concentração: Econometria.

Orientador: Prof. Dr. José Raimundo Carvalho.

Coorientador: Prof. Dr. Diego de Maria André.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- O46e Oliveira, Eduardo Cândido de.
 Ensaio em econometria, estatística e aprendizado de máquina aplicados à violência doméstica /
 Eduardo Cândido de Oliveira. – 2024.
 158 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e
 Contabilidade, Programa de Pós-Graduação em Economia, Fortaleza, 2024.
 Orientação: Prof. Dr. José Raimundo Carvalho.
 Coorientação: Prof. Dr. Diego de Maria André.
1. Violência doméstica. 2. Dados longitudinais. 3. Atrito. 4. Reposição. 5. Inconsistência lógica. I.
 Título.

CDD 330

EDUARDO CÂNDIDO DE OLIVEIRA

ENSAIOS EM ECONOMETRIA, ESTATÍSTICA E APRENDIZADO DE MÁQUINA
APLICADOS À VIOLÊNCIA DOMÉSTICA

Tese apresentada ao Programa de Pós-Graduação em Economia do Centro de Aperfeiçoamento de Economistas do Nordeste - CAEN da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Economia. Área de Concentração: Econometria.

BANCA EXAMINADORA

Prof. Dr. José Raimundo Carvalho (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Diego de Maria André (Coorientador)
Universidade Federal do Rio Grande do Norte
(UFRN)

Prof. Dr. Daniel Tomaz de Sousa
Universidade Federal do Ceará (UFC/FEAAC)

Profa. Dra. Mércia Santos da Cruz
Universidade Federal da Paraíba (UFPB)

Dr. Victor Hugo de Oliveira
Instituto de Pesquisa e Estratégia Econômica do Ceará
(IPECE)

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me sustentar nos momentos mais difíceis, por me escutar nas diversas vezes que estava com dificuldades e não me deixar desistir, além de ter me dado força e, principalmente, saúde para superar as dificuldades.

Aos meus pais, Régio e Rozivania pela base familiar que me proporcionaram, servindo-me como exemplo de vida. Sou grato pelo carinho de vocês e pela dedicação com a qual me criaram, obrigado por estarem sempre presentes na minha vida escolar e acadêmica e por nunca medirem esforços para me proporcionar melhores condições de estudo.

Agradeço a minha esposa Késsia que sempre esteve ao meu lado, escutando minhas angústias, meus desabafos, muitas vezes sendo paciente e compreensiva em momentos que eu não estava presente. Obrigado por me incentivar a sentar na cadeira para estudar ou escrever algo, mesmo nos dias que eu chegava do trabalho sem querer ver a tela do notebook (eu queria mesmo era ficar deitado em uma rede sem fazer absolutamente nada).

Quero agradecer, ao meu orientador Prof. Dr. José Raimundo que, mesmo ele não sabendo foi parte essencial para eu voltar a ingressar na vida acadêmica. Obrigado principalmente pela sua paciência e atenção, dedicou do seu tempo para me orientar neste trabalho, o qual sem seu apoio não obteria êxito, até mesmo aos sábados e feriados. Assim como, o Prof. Dr. Diego André que nesses dois últimos anos teve papel fundamental na finalização deste trabalho, além de todos os demais membros da banca examinadora pelo aceite em participar e pelas contribuições realizadas.

Agradeço a todos os colegas da Vigilância em Saúde do município de São Gonçalo do Amarante-CE e do rodízio de caronas, começar a trabalhar em meio uma vida acadêmica foi um episódio desafiador e eu não sabia o quanto era complicado, apesar disso vocês deixaram o ambiente mais leve, facilitando a adaptação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Aos professores e pesquisadores da Rede EvA – Rede de Evidências e Análises Interdisciplinares sobre Violência Doméstica e Familiar que dedicam seu tempo em um assunto tão importante e delicado.

“Sonhos determinam o que você quer, ação
determina o que você conquista.” (Aldo Novak)

RESUMO

Esta tese apresenta três ensaios sobre a temática da violência doméstica, utilizando os dados da Pesquisa sobre Condições Socioeconômicas e Violência Doméstica e Familiar contra a Mulher (PCSVDF-Mulher). O primeiro capítulo busca analisar a ocorrência da perda de unidades amostrais, isto é, o atrito amostral, na PCSVDF-Mulher. Durante as duas primeiras ondas da pesquisa, que ocorreram entre 2016 e 2017, foi encontrado um percentual de 55,13% de atrito. Para entender melhor o fenômeno, foram utilizadas técnicas de machine learning e um modelo de escolha discreta, capaz de identificar variáveis e a sua importância na ocorrência do atrito. Os resultados mostraram que características da moradia da mulher (tipo de domicílio - próprio ou não - e se há a pretensão de mudar-se) têm influência na participação dela em outra onda, além do fator ter filhos. Características do setor censitário em que a participante reside e da relação entre entrevistada e entrevistadora também são explorados. Sabendo da existência do atrito amostral, é possível que seja necessário substituir algumas mulheres durante o estudo longitudinal. O segundo capítulo visa, então, verificar se o protocolo de substituição adotado pela PCSVDF-Mulher foi cumprido pela empresa contratada e se a amostra de mulheres que são classificadas como reposição tem o mesmo perfil das que deixaram de participar da pesquisa (atritadas), tendo em vista que é preciso manter a representatividade estatística da amostra. Para isso, são utilizadas técnicas espaciais e estatísticas com o objetivo de verificar se a empresa cumpriu o protocolo no sentido de repor a amostra dentro do setor censitário ou em setores mais próximos e para comparar as amostras de reposição e atritadas. Os achados mostram que a empresa apresentou um comportamento otimizador em relação a reposição espacial. Em setores censitários que registraram déficit de domicílios de uma onda para outra estão, especialmente, mais próximos de setores censitários que obtiveram superávit de domicílios, com evidências de que as reposições tiveram dependência espacial. Além disso, a maior parte das variáveis comparadas entre as amostras, atritadas e reposição, tiveram a distribuição estatisticamente iguais, concluindo, assim, que a reposição foi cumprida de forma satisfatória. Por fim, o terceiro capítulo analisa outra problemática em pesquisas longitudinais, que é a inconsistência lógica na resposta, que pode ser explicada por diversas áreas, sejam elas psicológicas, cognitivas, por tipos de importâncias de eventos, entre outros. Em especial, na temática da violência doméstica, relatos inconsistentes de violência por parceiro íntimo (VPI) levam potencialmente a estimativas imprecisas da prevalência de tais comportamentos. Para a inconsistência sobre a violência geral, na PCSVDF-Mulher, entre as ondas 1 e 2, foi encontrada uma taxa de 18,08%, e, quando

desagregado por tipo de violência, a inconsistência emocional, física e sexual, atingia percentuais iguais a 17,17%, 12,57% e 6,54%, respectivamente, refletindo um perfil diferente. Os resultados encontrados indicam que variáveis de conhecimento, seja da escolaridade da mulher ou acerca do conhecimento sobre a Lei Maria da Penha, têm relação negativa com o relato inconsistente. Em contrapartida, caso a mulher tenha filhos ou foi vítima recente de violência, tem relação positiva com a inconsistência.

Palavras-chave: Violência doméstica. Dados longitudinais. Atrito. Reposição. Inconsistência lógica

ABSTRACT

This thesis presents three essays on the theme of domestic violence, using data from the Survey on Socioeconomic Conditions and Domestic and Family Violence against Women (PCSVDF-Mulher). The first chapter seeks to analyze the occurrence of loss of sampling units, that is, sampling attrition, in PCSVDF-Mulher. During the first two waves of the research, which took place between 2016 and 2017, a percentage of attrition of 55.13% was found. To better understand the phenomenon, machine learning techniques and a discrete choice model were used, capable of identifying variables and their importance in the occurrence of friction. The results showed that characteristics of the woman's housing (type of home - owned or not - and whether she intends to move) have an influence on her participation in another wave, in addition to the factor of having children. Characteristics of the census sector in which the participant resides and the relationship between interviewee and interviewer are also explored. Knowing the existence of sample attrition, it is possible that it will be necessary to replace some women during the study longitudinal. The second chapter then aims to verify whether the replacement protocol adopted by PCSVDF-Mulher was complied with by the contracted company and whether the sample of women who are classified as replacement has the same profile as those who stopped participating in the research (attrited), considering that it is necessary to maintain the statistical representativeness of the sample. For this, spatial and statistical techniques are used with the aim of verifying whether the company complied with the protocol in order to replace the sample within the sector census or in closer sectors and to compare replacement and attrited samples. The findings show that the company presented an optimizing behavior in relation to spatial replacement. In census tracts that recorded a deficit of households from one wave to the next, they are spatially closer to census tracts that had a surplus of households, with evidence that the replacements were spatially dependent. Furthermore, most of the variables compared between the samples, rubbed and replacement, had statistically equal distribution, thus concluding that the replacement was carried out satisfactorily. Finally, the third chapter analyzes another problem in longitudinal research, which is the logical inconsistency in the response, which can be explained by several areas, whether psychological, cognitive, types of importance of events, among others. In particular, on the topic of domestic violence, inconsistent reports of intimate partner violence (IPV) potentially lead to inaccurate estimates of the prevalence of such behaviors. For inconsistency regarding general violence, in PCSVDF-Mulher, between waves 1 and 2, a rate of 18.08% was found, and, when disaggregated by type of violence, emotional,

physical and sexual inconsistency reached percentages equal to 17.17%, 12.57% and 6.54%, respectively, reflecting a different profile. The results found indicate that knowledge variables, whether regarding the woman's education or knowledge about the Maria da Penha Law, have negative relationship with inconsistent reporting. On the other hand, if the woman has children or has been a recent victim of violence, it has a positive relationship with inconsistency.

Keywords: Domestic violence. Longitudinal data. Attrition. Replacement. Logical inconsistency

LISTA DE FIGURAS

Figura 1 – Comparação da taxa de atrito de acordo com os dados da base de Andre e Carvalho (2023) e da base atual	31
Figura 2 – Fluxograma dos motivos de atrito na PCSVDF-Mulher	36
Figura 3 – Fluxograma das ondas da PCSVDF	39
Figura 4 – Esquemático da Onda 1 e 2 e o tipo de variável que pode ser obtida em cada seção	42
Figura 5 – Diagrama de seleção das variáveis e estimação do modelo	47
Figura 6 – Diagrama com as variáveis selecionadas pelo método LASSO	50
Figura 7 – Variáveis, de acordo com a importância, método LASSO	51
Figura 8 – Esquemático dos tipos de substituição em uma pesquisa.	62
Figura 9 – Fluxograma do processo de sucesso, atrito ou substituição PCSVDF-Mulher	70
Figura 10 – Valor de N(g) de acordo com a desagregação geográfica - Bairro e Setores censitários.	74
Figura 11 – Mapa setores censitários Fortaleza-CE de acordo com a classificação do setor	99
Figura 12 – Mapa setores censitários Recife-PE de acordo com a classificação do setor .	99
Figura 13 – Mapa setores censitários Salvador-BA de acordo com a classificação do setor	100
Figura 14 – Mapa setores censitários Aracajú-SE de acordo com a classificação do setor	100
Figura 15 – Mapa setores censitários João Pessoa-PB de acordo com a classificação do setor	101
Figura 16 – Mapa setores censitários Maceió-AL de acordo com a classificação do setor	101
Figura 17 – Mapa setores censitários Natal-RN de acordo com a classificação do setor .	102
Figura 18 – Mapa setores censitários São Luís-MA de acordo com a classificação do setor	102
Figura 19 – Mapa setores censitários Teresina-PI de acordo com a classificação do setor	103

LISTA DE TABELAS

Tabela 1 – Coletado, atritado e pareado, Onda 1, de acordo com a base de dados utilizada	30
Tabela 2 – Variáveis do modelo logit e sua descrição	44
Tabela 3 – Estatística descritiva das variáveis quantitativas do modelo	45
Tabela 4 – Estatísticas de precisão do modelo para prever atrito. LASSO, SVM, Random Forest e Regressão Logística.	55
Tabela 5 – Estimação logit do modelo	57
Tabela 6 – Características de uma pesquisa adequada para substituições de campo . . .	71
Tabela 7 – Taxa de reposição por município, em 2016 e 2017	73
Tabela 8 – Resumo dos testes estatísticos utilizados	76
Tabela 9 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD ²⁰¹⁶ , PCSVDF-Mulher ²⁰¹⁶)	77
Tabela 10 – Frequência absoluta e relativa, PNAD ²⁰¹⁶ E PCSVDF-Mulher ²⁰¹⁶ de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.	78
Tabela 11 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD ²⁰¹⁶ , PCSVDF-Mulher ^{ATT})	79
Tabela 12 – Frequência absoluta e relativa, PNAD ²⁰¹⁶ e PCSVDF-Mulher ^{atr} de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.	80
Tabela 13 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD ²⁰¹⁷ , PCSVDF-Mulher ²⁰¹⁷)	81
Tabela 14 – Frequência absoluta e relativa, PNAD ²⁰¹⁷ e PCSVDF-Mulher ²⁰¹⁷ de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.	82
Tabela 15 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD ²⁰¹⁷ , PCSVDF-Mulher ^{REP})	83
Tabela 16 – Frequência absoluta e relativa, PNAD ²⁰¹⁷ e PCSVDF-Mulher ^{REP} de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.	84
Tabela 17 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição.	86

Tabela 18 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição. Fortaleza-CE e Recife-PE	88
Tabela 19 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição. Salvador-BA e Maceió-AL	89
Tabela 20 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição. João Pessoa-PB e Natal-RN	90
Tabela 21 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição, Aracaju-SE e Teresina-PI	91
Tabela 22 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição, São Luís-MA	92
Tabela 23 – Tabela resumo das classificações	94
Tabela 24 – Quantidade de domicílios, segundo o município, em que ocorreu reposição dentro do domicílio	95
Tabela 25 – Quantidade de domicílios, segundo o município, em que ocorreu reposição dentro do domicílio	95
Tabela 26 – Teste join count múltiplo para os municípios.	98
Tabela 27 – Resumo dos mecanismos de inconsistência e teorias	117
Tabela 28 – Perguntas do questionário da PCSVDF-Mulher sobre episódios de violência	118
Tabela 29 – Combinações de respostas totais de acordo com os cortes no tempo, somente com as combinações que ocorreram nas respostas	120
Tabela 30 – Combinações de respostas de acordo com os cortes no tempo	121
Tabela 31 – Combinações de respostas de acordo com os cortes no tempo e o tipo de violência	123
Tabela 32 – Tipo de violência e porcentagem de respostas inconsistentes	124
Tabela 33 – Variáveis para o modelo e sua descrição	125
Tabela 34 – Conhecimento da LMP de acordo com a pergunta e o ano	126
Tabela 35 – Análise descritiva das variáveis da mulher e do parceiro	127
Tabela 36 – Análise de frequência sobre a Lei Maria da Penha (LMP)	128
Tabela 37 – Modelo logit para inconsistência na resposta sobre violência geral	134
Tabela 38 – Modelo logit para inconsistência na resposta sobre violência emocional	134
Tabela 39 – Modelo logit para inconsistência na resposta sobre violência física	135
Tabela 40 – Modelo logit para inconsistência na resposta sobre violência sexual	135

Tabela 41 – Combinações de respostas de acordo com os cortes no tempo	136
Tabela 42 – Tipo de violência e porcentagem de respostas inconsistentes	136
Tabela 43 – Modelo multilogit para inconsistência na resposta sobre violência emocional	139
Tabela 44 – Modelo multilogit para inconsistência na resposta sobre violência física . .	140
Tabela 45 – Modelo multilogit para inconsistência na resposta sobre violência sexual . .	141
Tabela 46 – Combinações de respostas totais de acordo com os cortes no tempo	152
Tabela 47 – Perguntas sobre a Lei Maria da Penha	154
Tabela 48 – Percentual de inconsistência de acordo com as variáveis com características da mulher e do parceiro (Inconsistência geral)	155
Tabela 49 – Percentual de inconsistência de acordo com as variáveis com características da mulher e do parceiro (Inconsistência tipo 1)	156
Tabela 50 – Percentual de inconsistência de acordo com as variáveis com características da mulher e do parceiro (Inconsistência tipo 2)	157

SUMÁRIO

1	INTRODUÇÃO GERAL	16
2	CAPÍTULO 1 - UTILIZANDO MACHINE LEARNING PARA ANALISAR ATRITO EM UMA PESQUISA LONGITUDINAL COM DESENHO COMPLEXO: A PCSVDF-MULHER	19
2.1	Introdução	19
2.2	Revisão de literatura	22
2.2.1	<i>O processo de resposta de uma pesquisa</i>	<i>22</i>
2.2.2	<i>Atrito: Definição, Conceitos e Mecanismos</i>	<i>23</i>
2.2.3	<i>Estimação de Modelos de Atrito</i>	<i>24</i>
2.3	Soluções para o Atrito: Preventivas e Corretivas	27
2.3.1	<i>Medidas preventivas</i>	<i>27</i>
2.3.2	<i>Medidas corretivas</i>	<i>28</i>
2.4	A PCSVDF-Mulher	29
2.4.1	<i>Origem do estudo</i>	<i>29</i>
2.4.2	<i>Metodologia da PCSVDF-Mulher (Desenho amostral)</i>	<i>30</i>
2.4.3	<i>Instrumentos de pesquisa (Questionário)</i>	<i>32</i>
2.5	O atrito na PCSVDF-Mulher	34
2.5.1	<i>Análise das causas e o protocolo do atrito na PCSVDF-Mulher</i>	<i>34</i>
2.5.2	<i>Cálculo de alguns indicadores de atrito na PCSVDF-Mulher</i>	<i>37</i>
2.6	Métodos	40
2.6.1	<i>Mecanismo/Modelo de atrito escolhido</i>	<i>40</i>
2.6.2	<i>Atrito na Perspectiva de Aprendizado de Máquina: Estimação e Previsão</i>	<i>48</i>
2.6.2.1	<i>Método LASSO</i>	<i>48</i>
2.6.2.2	<i>Random Forests</i>	<i>52</i>
2.6.2.3	<i>Support Vector Machines</i>	<i>53</i>
2.7	Resultados	54
2.8	Considerações finais	57
3	CAPÍTULO 02 - AVALIAÇÃO DO PROTOCOLO DE SUBSTITUIÇÃO DE UMA PESQUISA LONGITUDINAL COMPLEXA: O CASO DA PCSVDF-MULHER	59

3.1	Introdução	59
3.2	Revisão de literatura	60
3.3	Exemplos de Protocolos de substituição	63
3.3.1	<i>Protocolo de substituição Belgian Health Interview Survey (BHIS) - Bélgica</i>	63
3.3.2	<i>Protocolo de substituição Progressi delle Aziende Sanitarie per la Salute in Italia (PASSI) - Itália</i>	65
3.3.3	<i>Protocolo de substituição Estudo Longitudinal da Saúde dos Idosos Brasileiros (ELSI-Brasil) - Brasil</i>	66
3.3.4	<i>Protocolo de substituição AmericasBarometer 2016/17 - América do Norte, Central, do Sul e parte do Caribe</i>	67
3.3.5	<i>Protocolo de substituição Russia Longitudinal Monitoring Survey (RLMS)- Rússia</i>	68
3.3.6	<i>Protocolo de substituição Healthy Chicago Survey HCS - Estados Unidos</i>	68
3.3.7	<i>Protocolo de substituição da PCSVDF-Mulher</i>	69
3.4	Análise preliminar do processo de reposição amostral	72
3.4.1	<i>Reposição do tamanho original: Desagregação por cidades</i>	72
3.4.2	<i>Reposição do tamanho original: Desagregação por bairros e setores censitários</i>	74
3.5	Metodologia I: Representatividade não espacial	75
3.5.1	<i>PCSVDF-Mulher 2016 versus PNAD-2016</i>	76
3.5.2	<i>PCSVDF-Mulher^{ATT} versus PNAD-2016</i>	79
3.5.3	<i>PCSVDF-Mulher 2017 versus PNAD-2017</i>	81
3.5.4	<i>PCSVDF-Mulher^{REP} versus PNAD-2017</i>	83
3.5.5	<i>PCSVDF-Mulher^{ATT} versus PCSVDF-Mulher^{REP}</i>	85
3.6	Metodologia II: Representatividade estatística espacial	92
3.6.1	<i>Definindo uma linguagem operacional</i>	92
3.6.2	<i>Teste Join Count Múltiplo</i>	96
3.7	Considerações Finais	104
4	CAPÍTULO 03 - MODELANDO INCONSISTÊNCIAS LONGITUDINAIS NAS RESPOSTAS SOBRE VITIMIZAÇÃO POR VIOLÊNCIA DOMÉSTICA.	106
4.1	Introdução	106

4.2	Referencial Teórico	109
4.2.1	<i>Inconsistência de resposta: Definição</i>	109
4.2.2	<i>Psicologia</i>	110
4.2.3	<i>Estatística (Desenho de surveys)</i>	113
4.2.4	<i>Economia/Econometria</i>	113
4.2.5	<i>Violência doméstica e saúde pública</i>	115
4.3	Análise Exploratória dos dados	117
4.3.1	<i>A PCSVDF-Mulher</i>	117
4.3.2	<i>Inconsistência: Operacionalização</i>	119
4.3.3	<i>Análise das inconsistências</i>	122
4.4	Modelagem econométrica	124
4.4.1	<i>Variáveis do modelo</i>	125
4.4.2	<i>Análise descritiva</i>	126
4.4.3	<i>Inconsistência e variáveis</i>	129
4.4.4	<i>Modelo logit padrão</i>	131
4.4.5	<i>Modelo multi-logit</i>	136
4.5	Considerações finais	142
	REFERÊNCIAS	144
	APÊNDICE A –	151
	APÊNDICE B – APÊNDICE B - FIGURAS E TABELAS ADICIONAIS	
	DO CAPÍTULO 03	152

1 INTRODUÇÃO GERAL

De acordo com dados da Organização Mundial da Saúde (OMS), em 2021, a violência contra as mulheres começa bastante cedo, uma em cada quatro mulheres jovens - idade entre 15 e 24 anos - que estiveram em um relacionamento já teriam sofrido violência cometida pelo parceiro até os 20 anos.

Ao longo da vida, uma em cada três mulheres, é submetida à violência física ou sexual por parte de seu parceiro ou violência sexual por parte de um não parceiro. A violência praticada pelo parceiro íntimo (VPI) é a forma mais prevalente de violência contra a mulher em todo o mundo, , contudo, ainda cerca de 6% relatam algum tipo de abuso sexual por alguma pessoa que não seja o parceiro íntimo.

De acordo com Sardhina *et al.* (2022) essa prevalência é maior entre as faixas etárias de 15 a 19 anos e 20 a 24 anos e os autores encontram resultados e que países de alta renda tendem a possuir menores prevalências de violência contra a mulher, tendo em vista que a violência foi maior na Oceania (49%) e na África Central Subsaariana (44%).

No Brasil, a violência doméstica também está presente, de três a cada dez brasileiras já foram vítimas de violência doméstica provocada por homens, de acordo com DataSenado (2023). A pesquisa apontou que o tipo de violência mais recorrente contra a mulher é a violência psicológica, seguida pela moral, física, patrimonial e pela sexual.

Metade das mulheres agredidas sofreram violência praticada pelo marido ou companheiro, e 15%, pelo ex-marido, ex-namorado ou ex-companheiro, ainda de acordo com os dados encontrados em DataSenado (2023). As mulheres vítimas, em sua maioria conseguem terminar relacionamentos abusivos, entretanto existe um número alto de mulheres vítimas que disseram que houve uma quebra de medidas protetivas de urgência, ocorrendo assim, uma falha no sistema de proteção a mulheres que já foram vítima. Além disso, por mais que o aumento pela procura de ajuda tenham aumentado, a falta de serviços especializados, como por exemplo, delegacias da mulher, em muitas cidades, dificulta o acesso ao serviço.

Outra questão a ser destacada com relação a violência doméstica, diz respeito à exposição de crianças, direta ou indiretamente, à violência doméstica que pode ocorrer muito cedo na vida (talvez ainda no útero). Pesquisas apontam que crianças expostas à violência doméstica estão mais propensas a se envolverem em relações violentas durante a vida adulta Roberts *et al.* (2010). Assim, a violência doméstica não ameaça somente a saúde e o bem-estar da mulher, pode também gerar graves consequências para as gerações futuras, sendo um dos

potenciais influenciadores no desenvolvimento infantil.

É sábio que a desigualdade entre gêneros é evidenciada nos salários, empregos, acesso à educação, oportunidades de tratamento de saúde, utilização, entre outras. Porém, a face mais marcante da desigualdade entre gêneros se reflete na experiência cotidiana da violência doméstica e familiar contra a mulher.

Apesar da importância de se estudar a problemática relativa à violência doméstica, existem alguns impeditivos para que o nível de análise sobre o tema no Brasil alcance uma certa excelência. Essa falta de capacidade analítica se dá por falta de bases de dados que consigam mensurar o nível de violência doméstica associada a fatores econômicos e sociais a nível agregado e, principalmente, a nível domiciliar e familiar.

Para preencher essa lacuna viu-se a necessidade de se criar uma base de dados longitudinal, que permita compreender a dinâmica da violência doméstica e familiar contra a mulher no Brasil e sua relação com os processos de desenvolvimento econômico e social passados e vigentes. Foi onde surgiu a Pesquisa de Condições Socioeconômicas e Violência Doméstica e Familiar contra a Mulher - PCSVDF-Mulher.

A importância de se ter uma investigação longitudinal é o fato que a mesma coleta dados qualitativos e quantitativos e é responsável pelo uso de medidas contínuas ou repetidas para rastrear indivíduos por um período prolongado de tempo, geralmente anos ou décadas.

No entanto, as pesquisas longitudinais ou de painel enfrentam um desafio fundamental relacionado ao desgaste do painel (atrito amostral), isto é, entrevistados que participam de uma entrevista, mas não das etapas subsequentes. O capítulo 01 do presente estudo destaca justamente esse desgaste do painel que pode introduzir viés nas estimativas da pesquisa, especialmente se as pessoas que pararam de participar do estudo longitudinal forem sistematicamente diferentes daquelas que permanecem no painel.

É necessário modelar o desgaste amostral, haja vista que esse fenômeno pode causar diminuição da variância ou seletividade. Com isso, o capítulo 01 identifica fatores que estão ligados ao desgaste amostral na PCSVDF-Mulher, utilizando análises estatísticas exploratórias, modelagem estatística e de aprendizagem de máquina.

Os problemas causados pelo atrito amostral tendem a crescer à medida que aumenta a não resposta da unidade amostral. Portanto, ao planejar e conduzir o desenho do projeto de pesquisa, é necessário tomar algumas medidas para alcançar uma taxa de resposta elevada e que consiga lidar com os efeitos da não resposta.

Duas abordagens são essenciais para se alcançar uma taxa de resposta satisfatória. A primeira se dá na melhora do processo de coleta de dados para reduzir a proporção de unidades não contactadas ou que se recusam a participar; a segunda é abordado por métodos diferentes que visam minimizar os efeitos da não resposta nos resultados.

Dentre os métodos que são comumente adotados para lidar com esses efeitos está a substituição de campo. A substituição de campo é utilizada durante o trabalho de campo para substituir unidades que não respondem por outras, na maioria das vezes, que já são pré-definidas de acordo com um protocolo de substituição no processo de construção da pesquisa. Geralmente, essa substituição ocorre dentro do mesmo estrato.

É necessário estudar o processo de substituição da PCSVDF-Mulher, haja vista que é de suma importância que o processo tenha sido bem feito e de acordo com o protocolo definido na pesquisa, uma vez que o não cumprimento do protocolo poderá causar viés nas estimativas. Com isso, o objetivo do capítulo 02 é verificar se o processo de substituição da PCSVDF-Mulher entre as ondas 1 e 2 foi cumprido e se foi eficaz.

Outra problemática que ocorre em dados de painel se dá em compreender como a resposta dada muda ao longo da vida e se ela permanece consistente, logicamente, no decorrer das ondas. Relatos inconsistentes podem ocorrer ao longo das ondas, os entrevistados podem relatar um fato somente anos após a experiência inicial, quando tiveram tempo de processá-la. Outros entrevistados, por sua vez, que já relataram algum fato anterior, podem querer deixá-lo “para trás” não relatar no momento posterior.

Com relação ao estudo longitudinal da violência a notificação inconsistente está associada a problemas de saúde e bem-estar. A divulgação da violência, portanto, pode ser um processo dinâmico que muda ao longo do tempo em resposta a uma série de fatores individuais, relacionais, situacionais e sociais.

O capítulo 03 tem como objetivo estudar respostas dadas pelas entrevistadas que são classificadas como inconsistentes no que rege a violência doméstica nos municípios do Nordeste, a partir de dados oriundos da Pesquisa de Condições Socioeconômicas e Violência Doméstica e Familiar contra a Mulher (PCSVDF-Mulher), durante os anos de 2016 e 2017.

2 CAPÍTULO 1 - UTILIZANDO MACHINE LEARNING PARA ANALISAR ATRITO EM UMA PESQUISA LONGITUDINAL COM DESENHO COMPLEXO: A PCSVDF-MULHER

2.1 Introdução

Dados em painel ou dados longitudinais referem-se a observações de séries temporais de vários indivíduos. Dentre as suas vantagens, pode ser destacada uma inferência mais precisa dos parâmetros do modelo, haja vista que dados em painel dotam a modelagem estatística de mais graus de liberdade e maior capacidade de refletir a complexidade dos fenômenos sob análise. Esse tipo de evidência empírica envolve pelo menos duas dimensões: uma dimensão transversal (cross-section) e uma dimensão temporal. Hsiao (2006)

Apesar de suas vantagens, como a detecção de mudanças, grande volume de dados e a sua precisão, as pesquisas longitudinais sofrem com o desgaste, ou mais precisamente com o atrito da amostra, que pode ser definido como a perda de uma unidade amostral ao longo do tempo – isto é, participantes deixam de comparecer a uma ou mais ondas de entrevistas. Lynn (2018) apresenta duas razões principais para justificar que o desgaste em painel é problemático. Primeiro, uma alta taxa de atrito pode fazer com que o tamanho da amostra das ondas subsequentes da pesquisa diminuam rapidamente, dificultando ou impedindo que se produzam estimativas precisas para a população (viés de seleção, imprecisão). Em segundo lugar, semelhante ao viés de não-resposta, o desgaste do painel pode introduzir viés nas estimativas da pesquisa, especialmente se as pessoas que pararam de participar (atritaram) do estudo longitudinal forem sistematicamente diferentes daquelas que permaneceram no painel.

No planejamento de pesquisas longitudinais, é fundamental manter as taxas de atrito baixas para preservar o poder estatístico e minimizar o risco de viés de atrito. Esse viés pode surgir se houver diferenças significativas entre os participantes que abandonam o estudo e aqueles que permanecem, especialmente se não forem adotadas medidas adicionais para controlar essa variação, Sánchez e Escobal (2020).

A Pesquisa de Condições Socioeconômicas e Violência Doméstica e Familiar contra a Mulher (PCSVDF-Mulher) é uma pesquisa de caráter longitudinal iniciada em 2016 nas nove capitais dos estados do Nordeste que tem como objetivo desenvolver uma base de dados longitudinais sobre condições socioeconômicas, violência doméstica e familiar, e desenvolvimento cognitivo infantil. De acordo com Antonino (2020) nas duas primeiras ondas (2016 e 2017),

mais de 10 mil mulheres foram entrevistadas em cada ano. A partir do ano de 2019 a pesquisa aumentou sua abrangência nacional; continuando com apenas três capitais da região nordeste, Fortaleza-CE, Recife-PE e Salvador-BA e adicionando as cidades de Belém-PA, Goiânia-GO, Porto Alegre-RS e São Paulo-SP; coletando portanto, dados nas cinco regiões brasileiras.

A taxa de desgaste amostral, entre os anos de 2016 e 2017, encontrada na PCSVDF-Mulher foi de 55,13%. Segundo Andre e Carvalho (2023) esse valor poderia ser interpretado elevado, entretanto os autores destacam que a pesquisa trata de questões sensíveis como violência perpetrada pelo parceiro, aborto, consumo de álcool e drogas, e não fornece incentivos monetários para as entrevistadas, concluindo, assim, que a taxa de atrito encontrada foi satisfatória.

Apesar das considerações feitas por esses autores, entende-se ser necessário modelar o atrito amostral no contexto pouco explorado de violência doméstica, haja vista que esse fenômeno pode distorcer estudos que utilizam a capacidade longitudinal da PCSVDF-Mulher. Por conseguinte, o objetivo do presente capítulo é identificar e analisar fatores que estão ligados ao desgaste amostral na PCSVDF-Mulher, utilizando análises estatísticas exploratórias e modelagem estatística e econométrica.

Para estimação do desgaste amostral foram utilizadas características do setor censitário em que a entrevistada reside, características da entrevistada e da entrevistadora, entre outras. Por exemplo, dentro das características da entrevistada, aspectos de gênero como a influência de eventos relacionados à experiência de violência doméstica sofrida nos últimos 12 meses. Modelos de escolha discreta foram estimados para detectar as principais variáveis que impactam na não participação das entrevistadas na onda subsequente.

Note, no entanto, que os modelos foram estimados sem se identificar qual o conjunto de variáveis independentes deveria ser utilizado. Isso é relevante no contexto da PCSVDF-Mulher devido ao número muito grande de variáveis coletadas que potencialmente poderiam ser inseridas no modelo. Contudo, existem na literatura métodos que podem estimar modelos estatísticos e selecionar as variáveis que mais contribuem para o poder explicativo do modelo. Um desses métodos é o Least Absolute Shrinkage and Selection Operator (LASSO). Técnicas de machine learning como o LASSO, ainda são pouco exploradas em um contexto de modelagem econométrica do atrito. Portanto, além de ser utilizada uma abordagem tradicional para modelagem de atrito comumente usada na prática (regressão logística); foi utilizado o método LASSO para selecionar variáveis em ambos os modelos.

De acordo com a classificação com relação a importância das variáveis para se

explicar o atrito pelo método LASSO, foram utilizadas as variáveis de residência da mulher, taxa de pessoas brancas e taxa de domicílios com abastecimento de água no setor censitário, a entrevistada não pretender se mudar e ter domicílio próprio, assim como a sua idade e ela possuir filhos, e a experiência da entrevistadora; essas foram importantes em todos os modelos. De posse desse vetor de variáveis independentes advindo da metodologia do LASSO, foram estimado tanto o modelo logit quanto outros modelos de machine learning e comparadas as métricas de desempenho, ou seja, acurácia, sensibilidade, especificidade e acurácia balanceada.

Os resultados do modelo logit mostram o quanto variáveis de domicílio da mulher, como a sua pretensão em mudar, ou se o domicílio é próprio afetam o atrito. Além disso, variáveis da relação entre entrevistada e entrevistadora e sobre experiências anteriores de violência também tem influência sobre o atrito em ondas subsequentes.

Logo, acredita-se que o presente trabalho possa contribuir para o entendimento do atrito amostral em pesquisas longitudinais de violência doméstica uma vez que existe uma carência de dados e uma carência de modelagem de atrito nesse contexto.

O capítulo é dividido em sete seções. A segunda seção aborda a literatura teórica sobre o processo de resposta em pesquisas, com aspectos econométricos e psicológicos, unit nonresponse em estudos longitudinais, bem como suas consequências, seus modelos e métodos, e estimação (abordagem econométrica e de aprendizagem de máquina), e por fim, sugestões de solução para o atrito amostral seja por meio de medidas preventivas ou corretivas. Na terceira seção, é descrita a PCSVDF-Mulher, evidência empírica utilizada na modelagem do presente estudo, apresentando a origem do estudo, metodologia da pesquisa, assim como a definição de atrito na pesquisa, o protocolo de atrito e o cálculo de algumas taxas de atrito entre as ondas. A seção quatro apresenta a descrição das variáveis que serão utilizadas no modelo com uma breve análise exploratória e o resultado da estimação. A abordagem de machine learning é apresentada no capítulo cinco, com a descrição dos métodos que serão utilizados. A seção seis aborda os resultados comparativos entre as abordagens de machine learning e o método de regressão logística, com algumas estatísticas de performance a partir de uma previsão de um banco de dados de treino e de teste. Por fim, a última seção traz as considerações finais do trabalho.

2.2 Revisão de literatura

2.2.1 O processo de resposta de uma pesquisa

O primeiro passo antes do entendimento do conceito do desgaste amostral (ou atrito amostral) em uma pesquisa é entender o processo de resposta de uma pesquisa domiciliar, ou seja, quais fatores têm influência em um entrevistado(a) para aceitar ou não responder o instrumento. Segundo Tourangeau *et al.* (2000, p. 03), dois processos de respostas de pesquisa são os pioneiros em metodologias de pesquisa: Teorias Psicométricas e a Teoria do processo de Cannell.

Os primeiros psicometristas a estudarem modelos do processo de resposta à pesquisa foram Likert (1932), Guttman (1950) e Guilford (1960). Produzindo as primeiras métricas de mensuração de atitudes, técnicas que quantificaram a força da convicção de uma pessoa em uma opinião. Os modelos se baseavam em uma teoria psicológica do questionamento com a ideia de que as pessoas representam estímulos como pontos ou regiões em uma dimensão interna. Assim, seus modelos psicológicos distinguiram vários processos compostos, incluindo a identificação da dimensão do questionamento

O modelo proposto por Cannell *et al.* (1981) identifica dois caminhos distintos para a formulação de respostas. O primeiro caminho é fundamentado no processamento detalhado da pergunta, enquanto o segundo se baseia em características superficiais da situação da entrevista, como a aparência do entrevistador. De acordo com esse modelo, a resposta é o resultado de cinco conjuntos de processos.

1. Compreensão da questão;
2. Processamento cognitivo;
3. Avaliação da precisão da resposta;
4. Avaliação da resposta com base em objetivos diferentes da precisão; e
5. Resposta precisa.

Alguns aspectos são essenciais para a participação em uma pesquisa, Groves e Couper (2000) apresentam alguns deles e fazem uma divisão em características que podem ser controladas pelo pesquisador (design da pesquisa, o tema da pesquisa, escolha dos entrevistadores e a interação entre entrevistador/entrevistado) e outras que fogem do seu controle (características do ambiente social e do domicílio que estão sob estudo).

2.2.2 Atrito: Definição, Conceitos e Mecanismos

Wittwer e Hubrich (2015) definem *unit nonresponse* como sendo considerado a “não resposta da unidade”, que resulta de uma perda da unidade amostral completa (domicílio, pessoa ou qualquer objeto de estudo selecionado). Cochran (1977) elenca as razões para a ocorrência da *unit nonresponse*: Ausência no domicílio, pessoa incapaz de responder e a não cobertura, embora a não cobertura possa ser classificada mais apropriadamente como um erro de design de pesquisa.

Segundo Silva (2011), o atrito em estudos longitudinais poderá ocorrer em algumas situações:

1. O respondente decide abandonar o estudo ou vem a falecer;
2. Por algum motivo o respondente é desligado da amostra;
3. O respondente mudou de endereço residencial e não é localizado ou sua nova área geográfica de residência agora não abrange mais a área de estudo necessária para o painel;
e
4. O assunto/tema do estudo poderá inviabilizar a continuidade do entrevistado na pesquisa.

Além disso, Ribas e Soares (2008) afirmam que o atrito pode ser: a) aleatório, que não gera qualquer tipo de viés em estimações de parâmetros; b) função de variáveis observadas, cujo viés pode ser corrigido mediante ponderação adequada; e c) função de variáveis não-observadas.

O trabalho de Ahern e Brocque (2005) analisa se a questão do atrito tem recebido a devida atenção em relatórios de pesquisas longitudinais. Os autores afirmam que poucos artigos investigam e ajustam os efeitos do atrito em pesquisas longitudinais. Alguns estudos ignoram, outros citam o fenômeno em um pequeno trecho e outros fazem algum tipo de ajustes simples. Segundo Ahern e Brocque (2005), menos de, 25% dos trabalhos que descrevem o tratamento dos dados na presença do atrito.

Groves (2002, p. 259-272) investiga a não resposta em ondas subsequentes, e para isso cria uma estrutura de modelo que divide o processo de pesquisa em três etapas: localização da unidade amostral, o contato (entrevistadora/entrevistada) devido à localização e cooperação da entrevistada devido ao contato.

A localização da unidade leva em consideração a propensão da unidade amostral em se locomover. Se a unidade for uma pessoa, a atenção será concentrada nesse indivíduo. Entretanto, em algumas pesquisas longitudinais, a unidade é um domicílio ou família e, com isso, as mudanças de toda a família representam um problema de localização. Para fatores

relacionados ao contato, a dificuldade se dá a um fator significativo de não resposta nas primeiras ondas, porém é relativamente pequena nas ondas posteriores, visto que ocorreu sucesso na localização da unidade amostral.

A disposição da unidade amostral em cooperar nas ondas subsequentes será definida por uma variedade de fatores no momento do contato (pesquisas transversais), e em pesquisas longitudinais pela sua experiência da onda anterior. O efeito da experiência da onda anterior no entrevistado em cooperar na onda posterior pode depender de vários fatores: a relevância da experiência, o tempo entre as ondas, se há contato entre as ondas da organização da pesquisa e se o mesmo ou um entrevistador diferente é usado Zabel (1998). Com relação a experiências em pesquisas anteriores, alguns estudos mostram que a avaliação do entrevistador sobre o prazer e a compreensão do respondente sobre uma entrevista de onda anterior está fortemente relacionada à cooperação, Kalton *et al.* (1990).

Levin *et al.* (2000) e Ahern e Brocque (2005) mencionam variáveis previamente identificadas na literatura como associadas ao atrito, incluindo idade avançada dos respondentes, solidão, problemas de saúde, baixa escolaridade, baixo nível socioeconômico e menor capacidade cognitiva.

2.2.3 Estimação de Modelos de Atrito

É de suma importância verificar na literatura o que já é pontuado sobre a problemática do atrito amostral. Alguns trabalhos, na abordagem econométrica e de machine learning (aprendizado de máquina), já dão ênfase a essa temática.

Kelfve *et al.* (2017) investigaram como a mortalidade e o atrito em pesquisas afetam a composição social das amostras longitudinais ao longo do tempo. A análise foi conduzida com uma amostra de 1132 indivíduos suecos, nascidos entre 1924 e 1934, acompanhados de 1968 a 2011. A pesquisa abrangeu seis ondas (1968, 1974, 1981, 1991, 2000/2002 e 2004) e utilizou dados dos estudos *Swedish Level-of-Living Survey* (LNU) e *The Swedish Panel Study of Living Conditions of the Oldest Old* (SWEOLD).

Os resultados indicaram que a mortalidade modificou a composição da amostra, aumentando progressivamente a proporção de mulheres, trabalhadores manuais, indivíduos mais jovens, com maior escolaridade e casados no início do estudo. Quando os desistentes também foram excluídos das análises, a seleção da amostra para as variáveis demográficas não apresentou mudanças substanciais. Após a exclusão tanto dos falecidos quanto dos desistentes, a

principal alteração foi observada em duas variáveis: o estado civil dos homens e o ano médio de nascimento das mulheres. Especificamente, ao longo do tempo, houve uma redução na participação de homens solteiros em comparação com os casados, e uma diminuição na presença de mulheres mais velhas em relação às mais jovens.

Régnier-Loilier (2012) analisa o problema do atrito na Pesquisa de Gerações e Gênero (PGS), realizada em cerca de 20 países. A PGS é composta por três ondas aplicada a cada três anos. O objetivo do estudo é permitir comparações internacionais. Através de análises descritivas e modelagens logísticas, foram identificados diversos fatores associados ao atrito: características sociodemográficas dos entrevistados, condições do domicílio, saúde e sociabilidade dos participantes, além de aspectos relacionados à experiência dos entrevistados durante a primeira onda (como a duração do questionário, recusa em responder a determinadas perguntas, entre outros).

Eisner *et al.* (2019) fazem um estudo do atrito longitudinal a partir de dados reais de uma pesquisa que tem o objetivo de medir o comportamento problemático classificado por crianças, pais e professores, com o foco no desenvolvimento do crime e da agressão. A pesquisa tem início a partir do ano de 2004, com o público alvo sendo crianças de 7 anos de idade, que ingressaram no primeiro ano do ensino fundamental em escolas na cidade de Zurique, Suíça. Ao longo dos 10 anos de estudo foram fornecidos dados pelos pais, professores e pelos próprios alunos.

Algumas variáveis para explicar o atrito foram avaliadas: 1) Tipo da classe em que a criança está inserido, por exemplo, turmas especiais destinam-se a atender crianças com algum tipo de necessidade; 2) Idioma do responsável; 3) Classe social do bairro e da família; 4) Nível de escolaridade dos pais; 5) Composição familiar: Se os dois responsáveis estão morando na mesma casa; 6) Comportamento infantil: Utilizando o Questionário de Comportamento Social que engloba aspectos de agressão, ansiedade, depressão e transtorno de déficit de atenção e hiperatividade (TDAH).

Os resultados do trabalho de Eisner *et al.* (2019) mostraram que a não resposta e o desgaste foram maiores entre certos pais de origem imigrante, não nativos. O comportamento infantil também foi considerado um preditor significativo de atrito nas análises. Agressividade, problemas de conduta não-agressiva, TDAH e internalização, avaliados pelos professores, foram todos preditores do atrito.

Com relação aos efeitos não observáveis, Fitzgerald, Gottschalk e Moffitt (1997)

observa-se que esses efeitos estão ligados às características do entrevistador e ao próprio processo de entrevista. Zabel (1998) identifica evidências de que tais efeitos impactam o desgaste, destacando, por exemplo, o tempo de aplicação do questionário. Quando o processo é menos cansativo, o participante tende a desenvolver um maior comprometimento com o estudo. Contudo, muitos estudos não levam em conta esses efeitos a menos que sejam fornecidas informações adicionais. Portanto, avaliar o desgaste através de efeitos observáveis torna-se uma abordagem mais prática. (Outes-Leon e Dercon (2008) e Alderman et al. (2001)).

Alguns pesquisadores discutem que o atrito não é um problema grave que impeça a modelagem do fenômeno de interesse. Em seu estudo, Fitzgerald, Gottschalk e Moffitt (1997) não encontraram fortes evidências de que o atrito tenha impactado a representatividade da amostra, e também observaram que o atrito estava concentrado entre indivíduos com baixo nível socioeconômico. De forma semelhante, Falaris (2003) e Cheng e Trivedi (2015) também encontraram poucas evidências de viés de atrito nos coeficientes estimados. No entanto, Baigrie e Eyal (2013) e Maluccio (2000) detectaram evidências de viés nos coeficientes devido ao atrito.

Liu (2020) explora alguns métodos que estimam os determinantes do atrito em uma pesquisa longitudinal. O autor explora a probabilidade de atrito na onda 2 para cada respondente da onda 1 com base em quatro métodos de modelagem diferentes: regressão logística (abordagem tradicional para modelagem de atrito), e outros três métodos de aprendizagem de máquina: random forest, SVM e LASSO.

Ainda do trabalho de Liu (2020), o autor mostra que, além da previsão de atrito, é possível com os métodos random forest e LASSO fornecer uma medida de importância de variável que pode ajudar a priorizar ou entender melhor quais utilizar para prever com precisão o atrito do painel. Nos quatro métodos usados neste estudo, o LASSO mostra uma pequena vantagem na previsão da participação na pesquisa da onda 2, conforme indicado pela acurácia. A vantagem não é substancial, no entanto, e os outros dois métodos (random forest e SVM) mostram desempenho preditivo semelhante, a julgar por várias estatísticas. Ao comparar esses três métodos com o modelo de regressão logística, os três métodos não apresentam uma vantagem significativa sobre o modelo de regressão logística.

O trabalho de Nicholson *et al.* (2017) tem o objetivo de testar de forma empírica se o atrito pode ser previsto com mais precisão por meio de algoritmos de machine learning que são capazes de incorporar efeitos não lineares ou de interação de preditores heterogêneos. Comparando a precisão preditiva de um algoritmo de machine learning baseado em árvore, GBM,

e um modelo de regressão logística.

Para validar os modelos de previsão, são empregadas abordagens de validação temporal com dados de treinamento e teste que são disjuntos. Por fim, são comparados os resultados da previsão de atrito em dois estudos longitudinais de grande escala que diferem em seus objetivos de estudo, amostra e período de tempo. O primeiro está ligado ao desenvolvimento da saúde e bem-estar de uma população dos Estados Unidos com um onda a cada 9 anos, a outra é uma pesquisa anual alemã sobre parceria e dinâmica familiar.

Os resultados de desempenho de um modelo básico de regressão logística com um algoritmo de aprendizado de máquina foram comparados e mostram quase nenhuma diferença nas precisões para ambas as abordagens de modelagem

2.3 Soluções para o Atrito: Preventivas e Corretivas

No campo do design de pesquisas longitudinais existem algumas soluções que podem ocorrer no desenvolvimento do plano amostral para conseguir manter a unidade amostral por um período maior do estudo. Essas medidas podem ocorrer antes ou durante o contato com o entrevistado (medidas preventivas) ou podem ocorrer após a percepção da ocorrência do atrito amostral (medidas corretivas).

2.3.1 Medidas preventivas

Hill (2004) cita algumas medidas preventivas que podem reduzir o atrito. A escolha, treinamento, supervisão e motivação dos entrevistadores tem uma relação direta com o sucesso da pesquisa. Alguns estudos descobriram que uma seleção cuidadosa do entrevistador é essencial para isso.

Thomas *et al.* (2001) e Chamratrithirong e Sethaput (1997) encontraram na The Indonesian Family Life Study que o sucesso nas pesquisas estava associado à qualidade do entrevistador (medida pelas pontuações em matemática), o salário do emprego anterior e o contato com o supervisor. Os entrevistadores devem ter características pessoais como entusiasmo, flexibilidade, engenhosidade e perseverança. O sucesso na entrevista é facilitado se os entrevistadores conhecerem o local de trabalho (comunidade/bairro/setor censitário), pois isso os ajudará a localizar domicílios e rastrear contatos. É também desejável, embora muitas vezes difícil, utilizar os mesmos entrevistadores em cada onda de inquérito, uma vez que estão

familiarizados com a área de estudo e é fomentada uma boa relação entre os entrevistadores e os entrevistados.

As sugestões de medidas preventivas para reduzir o atrito são: 1.) coletar o máximo possível de informações sobre o entrevistado; 2.) considerar todas as diferentes opções de rastreamento e usar aquelas que são apropriadas localmente; 3) regras e limites de rastreamento; 4.) intervalo menor entre as ondas; 5.) fazer com que os entrevistados se sintam parte do estudo, e 6.) escolha, treinamento, supervisão e motivação dos entrevistadores.

Lima-Costa *et al.* (2018) destacam algumas medidas preventivas para se reduzir o atrito na pesquisa. 1.) durante a primeira entrevista, foram registrados números de telefone e endereços da rede do participante (familiares, amigos e vizinhos); 2.) entre as entrevistas serão realizadas ligações telefônicas para acompanhamento de mudanças de endereço e estado vital; 3.) será realizada vinculação ao Sistema Nacional de Mortalidade. Para rastrear os participantes que não foram localizados através dos procedimentos acima mencionados, além de uma visita ao setor censitário (domicílio do participante e domicílios vizinhos, se necessário) está planejada para as ondas subsequentes.

Ribisl *et al.* (1996) fornecem uma visão de como o atrito pode ameaçar a validade do painel e discutem oito métodos para minimizá-lo por meio do uso de estratégias de retenção e acompanhamento.

2.3.2 Medidas corretivas

Normalmente, item nonresponse foram tratados por métodos imputacionais, enquanto unit nonresponse geralmente é tratado por ponderação (ou seja, certas os registros dos respondentes recebem maior importância como compensação para não respondentes Lepkowski (1989). Também forneceu uma discussão convincente sobre a vantagens e desvantagens relativas do uso de imputação e técnicas de ponderação para lidar com a falta de resposta da unidade e descreve os procedimentos para integrar o duas abordagens.

Segundo a maioria dos relatos, quando os vieses de atrito são detectados, as correções estatísticas não podem anular todas as fontes de erro e viés. Métodos estatísticos, ou ajustes de ponderação para não respondentes, normalmente não podem eliminar todos os vieses introduzidos por atrito e não resposta.

Os entrevistados que, por ventura, são atritados podem ser excluídos da amostra, substituídos, ou rastreados. A estratégia apropriada depende do objetivo do estudo. As reduções

no poder estatístico podem ser superadas recrutando mais participantes ou substituindo aqueles que estão perdidos por entrevistados com características sociodemográficas semelhantes, mas em muitos casos o rastreamento dos entrevistados é a estratégia mais apropriada para minimizar o viés. Alguns estudos em que não foi utilizado o rastreamento tiveram de 9 a 21% de atrito ao ano, em comparação com apenas 1 a 8% para aqueles com rastreamento.

2.4 A PCSVDF-Mulher

2.4.1 Origem do estudo

A Pesquisa de Condições Socioeconômicas e Violência Doméstica e Familiar contra a Mulher (PCSVDF-Mulher) nasceu da necessidade de se ter uma base de dados longitudinais sobre condições socioeconômicas, violência doméstica e familiar, e desenvolvimento cognitivo infantil nas regiões do Brasil, por meio de coleta de dados primários. O estudo restringiu-se a entrevistar mulheres com idade igual ou maior a 15 anos e menor que 50 anos ($15 \leq \text{idade} < 50$). Até o presente momento, foram realizadas quatro ondas de entrevistas, nas quais as duas primeiras ondas foram realizadas nas nove capitais do nordeste, e a terceira e quarta ondas foram realizadas nas cidades de Fortaleza-CE, Recife-PE, Salvador-BA, Belém-PA, Goiânia-GO, Porto Alegre-RS e São Paulo-SP. As unidades amostrais são os domicílios e a unidade de observação é a mulher residente no domicílio dentro da faixa etária previamente estabelecida.

De acordo com Andre e Carvalho (2023) na onda 1 da PCSVDF-Mulher, a amostra original era composta por 10094 observações. Entretanto, para o processo de pós-estratificação, infelizmente, foram perdidos algumas observações tendo em vista que algumas variáveis como coordenadas do GPS do domicílio, idade, raça e escolaridade tinham missing. Com isso, a amostra final foi de 9618 observações.

Silva (2018) utiliza os dados da PCSVDF-Mulher para modelar o atrito em três aspectos diferentes. A primeira é modelar o atrito com o objetivo de abordar os determinantes e efeitos da violência doméstica. A segunda é classificar o atrito de diversas maneiras e não somente de forma binária. O terceiro aspecto é a abordagem do impacto das relações de gênero entre os membros do domicílio podem impactar a decisão de participar das demais ondas.

O presente trabalho tem um diferencial do trabalho de Silva (2018) no aspecto de que não serão abordadas somente as características da entrevistada e do seu parceiro. Também serão consideradas na modelagem do atrito amostral características do setor censitário em que o

domicílio está inserido, características do domicílio e da entrevistadora que fez a entrevista, as variáveis serão descritas de acordo com a Tabela 2.

Importante destacar que um dos diferenciais da presente estimação do atrito é a inclusão de dados das entrevistadoras que realizaram a pesquisa (ver Tabela 2). Com isso, as informações do banco de dados atual será menor que os dados oficiais (descritos em Andre e Carvalho (2023)), tendo em vista que ao se fazer o link entre a base de dados oficial e a base de dados das entrevistadoras algumas observações não foram correspondidas. Entretanto, não houve perda significativa com relação ao atrito e as variáveis, como apresentado na Tabela 1 e na Figura 1

Tabela 1 – Coletado, atritado e pareado, Onda 1, de acordo com a base de dados utilizada

Cidade	Andre e Carvalho (2023)			Base atual		
	W2016	ATT	PAIR	W2016	ATT	PAIR
Maceio/AL	943	505	438	932	499	441
Salvador/BA	1104	559	545	948	417	541
Fortaleza/CE	1172	504	668	704	296	419
São Luís/MA	1082	616	466	1075	611	490
João Pessoa/PB	1072	672	400	934	546	421
Recife/PB	1245	633	612	1200	622	611
Teresina/PI	962	531	431	935	512	431
Natal/RN	1052	688	364	1048	684	389
Aracaju/SE	986	594	392	1031	620	429
Total	9618	5302	4316	8807	4807	4000

Fonte: Elaborado pelo autor com base nos dados Andre e Carvalho (2023) e da base atual.

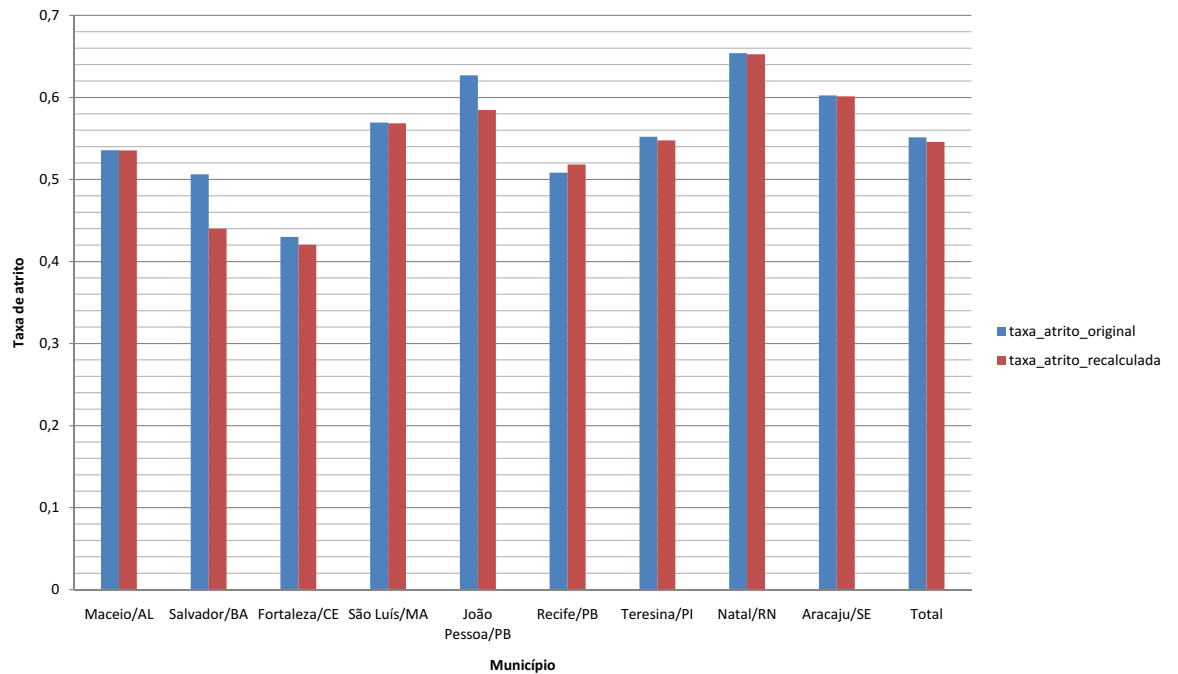
W2016: Entrevistas válidas 2016/ ATT: Domicílios atritados/ PAIR: Domicílios pareados

2.4.2 Metodologia da PCSVDF-Mulher (Desenho amostral)

Alguns conceitos com relação ao processo amostral precisam ser definidos em um primeiro momento. Segundo Lohr (2010):

- Unidade de observação: o objeto do qual a coleta de dados será feita. Esta é a unidade básica de observação, às vezes chamada de elemento. Ao estudar populações de humanos, as unidades de observação são normalmente um indivíduo.
- Unidade amostral: é a unidade que pode ser selecionada para uma amostra. Pode-se pretender estudar determinados indivíduos, mas talvez não tenhamos uma lista de todos os elementos na população-alvo. Em vez disso, os domicílios servem como unidades amostrais, e as unidades de observação são os indivíduos vivendo no domicílio.

Figura 1 – Comparação da taxa de atrito de acordo com os dados da base de Andre e Carvalho (2023) e da base atual



Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

- Arcabouço amostral: uma lista, mapa, ou outra especificação de unidades amostrais na população da qual a amostra será selecionada.
- Amostra: um subconjunto da população.
- População amostrada: a coleção de todas as possíveis unidades de observação que têm possibilidade de ser selecionadas para a amostra; é a população da qual a amostra será coletada.

O presente estudo propôs-se a analisar a população dos 9 estados brasileiros situados na região nordeste. Conforme dados da sinopse do Censo IBGE 2010, esta população é de 53.050.120 brasileiros.

Para delimitação do arcabouço amostral foram utilizados do Censo 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística – IBGE, bem como os mapas cartográficos dos setores censitários. As seguintes condições foram observadas para determinar o conjunto de setores censitários elegíveis ao estudo:

- Considerados apenas os setores pertencentes às nove capitais das Unidades Federativas da Região Nordeste do Brasil;

- Considerados apenas os setores cuja situação é tipificada como “Setor censitário de área urbanizada de cidade ou vila”;
- Considerados apenas os setores com dados agregados disponíveis (existem setores cujos dados são anônimos, isto é, cujos quantitativos não são publicados nas tabelas fornecidas pelo IBGE, sendo representados apenas como X).

Seria desejável ter um método de amostragem que impedisse que amostras que produzissem estimativas ruins e que aumentassem a precisão dos estimadores. A amostragem estratificada pode atingir esses objetivos. De acordo com Antonino (2020) a amostra foi estratificada em três estágios:

1. Primeiro estágio: seleção de uma amostra de setores censitários em cada uma das nove Capitais dos Estados pré-determinados para o Estudo;
2. Segundo estágio: seleção de uma amostra de domicílios em cada um dos setores selecionados na etapa [1.];
3. Terceiro estágio: seleção de uma pessoa do sexo feminino ($15 \leq \text{idade} < 50$) em cada domicílio para responder ao questionário da pesquisa;

2.4.3 Instrumentos de pesquisa (Questionário)

Inicialmente, o questionário do PCSVDF-Mulher deveria fazer perguntas sobre Histórico Socioeconômico, Poder de Negociação e Traços Culturais, Experiências de Vitimização, Expectativa de Vitimização, Avaliação Contingente e Risco, Descrição da comunidade, Informações gerais. Com isso, o instrumento foi finalizado contendo a seguinte estrutura:

1. Formulário de seleção de residência: coleta o endereço da residência e identifica de forma única cada respondente;
2. Formulário de administração - números aleatórios: captura automaticamente as coordenadas geográficas do endereço do domicílio, cria números aleatórios para auxiliar nas seguintes randomizações, bem como obtém mais informações sobre a composição do domicílio.
3. Formulário de Seleção De Mulheres (FSM)-caracterização dos membros do domicílio: permite que os entrevistadores construam uma lista de todos os membros da família e reúnam informações socioeconômicas básicas e informações necessárias para selecionar aleatoriamente uma mulher para continuar o processo de entrevista. Início de um conjunto específico de questões relacionadas a mulher. Além de coletar informações detalhadas so-

bre Bolsa Família, seguro-desemprego, pensão alimentícia e outros para todos os membros do domicílio.

4. Questionário da Mulher (QM) - saúde geral e reprodutiva: sucesso (parcial) no processo de entrevista, ou seja, atingir esse ponto durante o processo de pesquisa é condição necessária para fazer com que a empresa de pesquisa pague um entrevistador por aquele questionário específico. Coleta informações sobre o medo do crime, saúde geral e reprodutiva, informações detalhadas sobre o uso de anticoncepcionais, bem como histórico de gravidez, que contém informações sobre o peso do bebê, mortalidade (se aplicável) e tipo de parto. Há também uma subseção especial relacionada a muitas questões relacionadas à epidemia do vírus Zika no nordeste do Brasil.
5. Normas, conscientização / conhecimento sobre a violência contra as mulheres e a lei “MARIA DA PENHA”: discute a percepção das mulheres sobre o que constitui VIOLÊNCIA DOMÉSTICA, a ocorrência dessa violência em seu bairro, suas atitudes sobre as relações de gênero e a Lei Maria da Penha.
6. Entrevistada e seu parceiro: coleta informações sobre qualquer relacionamento com seu atual/mais recente marido/parceiro. Além disso, contém mais informações sobre as características do parceiro/ex-parceiro, bem como informações sobre a “Escala de tom de cor da pele” (ver, Massey e Martin (2003) e Hersch (2008)).
7. Poder de barganha: trata da descrição de todos os aspectos relativos ao desenvolvimento dos processos de negociação no domicílio. Perguntas sobre geração de renda, propriedade de ativos, alocação de tempo dos membros da família (informações sobre desigualdade de gênero na alocação de tempo entre a produção doméstica, trabalho e atividades de lazer), preferências e estrutura de tomada de decisão são coletadas, oferecendo uma perspectiva única para apoiar mais estudos precisos sobre as inter-relações entre a estrutura de negociação doméstica e a violência doméstica.
8. Experiências de violência doméstica (parceiro atual, ex-parceiro mais recente ou qualquer outro ex-parceiro: episódios detalhados de vitimização de violência doméstica de gênero. Os episódios de vitimização de violência são desagregados em físicos, psicológicos e sexuais. Não apenas dados sobre a prevalência de violência são coletados, mas também informações sobre estratégias de enfrentamento, avaliação de instituições que lidam com a violência feminina e histórico de relacionamentos das mulheres.
9. Experiências de violência não-relacionadas ao parceiro: recolhe dados sobre vivências

de diferentes formas de violência de familiares, de outras pessoas que conheçam e/ou de estranhos.

10. Avaliação do matching, expectativas subjetivas e contrafactuais: reúne informações inco-muns para calcular o valor do casamento, usando um conjunto de perguntas de avaliação subjetivas associadas a um conjunto experimental de perguntas “factuais”. Incluímos uma subseção inteiramente dedicada ao conhecimento sobre os aspectos e contrafactuais do casamento. Informações para estimar o custo da violência doméstica de gênero (Expectati-vas Subjetivas e Avaliação Contingente) também são coletadas junto com perguntas sobre bem-estar ou felicidade geral.
11. Seção complementar: é uma seção curta com avaliação do entrevistado sobre a dura-ção do processo de entrevista e avaliação da cor da pele do entrevistado realizada pelo entrevistador.
12. Resultados: seção de encerramento que coleta informações para auxiliar no processo de avaliação da qualidade dos dados.

2.5 O atrito na PCSVDF-Mulher

Assim como em toda pesquisa longitudinal, a PCSVDF-Mulher está sujeita a um fenômeno de perdas de unidades amostrais, que pode ser definido como atrito amostral ou como o fato da não resposta ao questionário.

Para fins do estudo, o atrito na PCSVDF-Mulher pode ser definido como o fato de uma entrevista ser considerada válida em uma onda e na onda subsequente ser considerada não válida. Uma resposta é considerada válida quando a entrevistada aceita participar da seção de saúde geral e reprodutiva.

As duas seções que seguem mostram a descrição das causas e o protocolo de atrito na PCSVDF-Mulher e o cálculo dos indicadores do atrito.

2.5.1 Análise das causas e o protocolo do atrito na PCSVDF-Mulher

De acordo com DATAINFO (2017) podem ser identificados quatro perfis de entre-vista:

- Sucesso (mulher da Onda 1): quando é entrevistada a mesma mulher que participou da Onda 1 e ela responde pelo menos até a seção de saúde geral e reprodutiva;

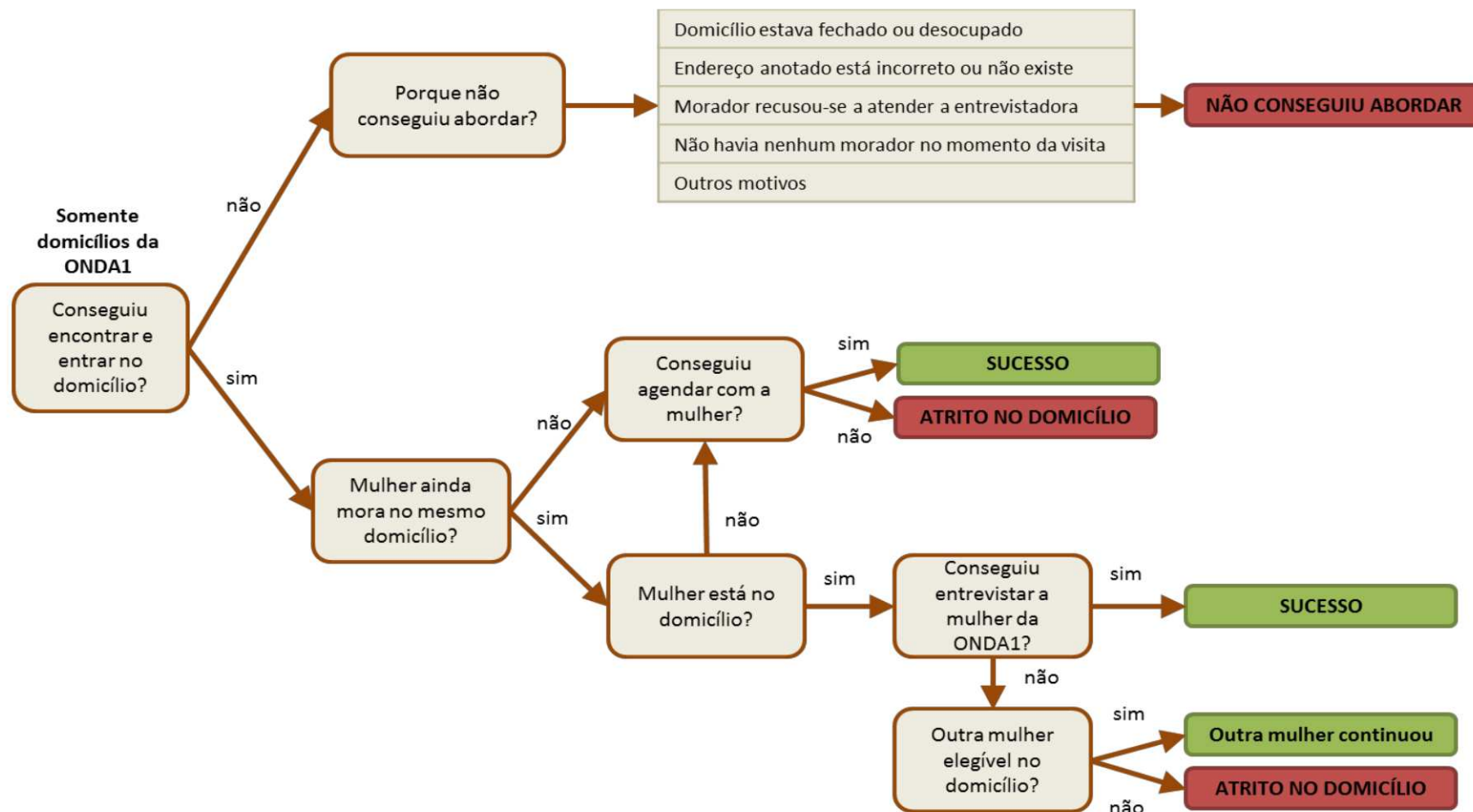
- Sucesso parcial (outra mulher continuou): quando outra mulher que mora no mesmo domicílio visitado na Onda 1 responde à entrevista;
- Não conseguiu abordar: quando ocorrem situações que impedem que a entrevistadora encontre e entre num domicílio visitado na Onda 1;
- Atrito no domicílio: quando ocorrem situações em que a entrevistadora entra num domicílio visitado na Onda 1, mas a mulher da Onda 1 não é entrevistada ou não chega a seção de saúde geral e reprodutiva.

As principais causas de atrito que ocorreram na PCSVDF-Mulher:

1. Causas de problemas em questionários com perfil - Não conseguiu abordar
 - O domicílio estava fechado ou desocupado
 - Não havia nenhum morador do domicílio no momento da visita
 - Endereço não encontrado
 - Falta de padronização na numeração das ruas
 - Necessidade de melhor identificação dos domicílios
 - Baixa fidelidade das entrevistadas ao número de telefone celular
 - Um morador recusou-se a atender a entrevistadora
2. Causas de problemas em questionários com perfil “Atrito no domicílio”
 - Havia alguém no domicílio, mas não conseguiu encontrar a mulher no domicílio na primeira visita.
 - Mulher saiu para trabalhar, estudar ou realizar alguma atividade.
 - Mulher mudou de endereço
 - Não consegue encontrar a mulher no domicílio nas visitas posteriores
 - Recusa da mulher
3. Outras causas de problemas identificadas:
 - Rotatividade das entrevistadoras/desmotivação em continuar o trabalho de campo
 - Baixa produtividade de questionários por dia
 - Remuneração
 - Impossibilidade de treinar novas entrevistadoras
 - Violência / Insegurança

O fluxograma apresentado na Figura 2 detalha de maneira os diversos passos que identificam o perfil de entrevistas da PCSVDF-Mulher. Essa representação foi elaborada com base nas informações fornecidas DATAINFO (2017).

Figura 2 – Fluxograma dos motivos de atrito na PCSVDF-Mulher



Adaptado de DATAINFO (2017)

Além das causas, é necessário descrever o protocolo específico de atrito, que é dado da seguinte forma:

- a) Um endereço visitado será declarado “endereço anotado incorreto/não existe” se atestado pela visita da entrevistadora de campo responsável pelo endereço e confirmado por sua coordenadora de campo;
- b) Um endereço visitado será declarado como “não havia ninguém no momento” ou “domicílio fechado ou desocupado” se atestado por duas visitas da entrevistadora de campo (em dias da semana e turnos diferentes) responsável pelo endereço e confirmado por uma terceira visita da sua coordenadora de campo;
- c) Um endereço visitado será declarado “morador recusou” se atestado por duas visitas da entrevistadora de campo (em dias da semana e turnos diferentes) responsável pelo endereço e confirmado por uma terceira visita da sua coordenadora de campo;
- Outros motivos podem ser definidos como:
- d) Um endereço visitado será declarado “impossibilidade de participar” se atestado por duas visitas da entrevistadora de campo (em dias da semana e turnos diferentes) responsável pelo endereço e confirmado por uma terceira visita da sua coordenadora de campo;
- e) Um endereço visitado será declarado “impossibilidade de se comunicar” se atestado pela entrevistadora de campo responsável pelo endereço e confirmado por sua coordenadora de campo;
- f) Um endereço visitado será declarado “não completou a entrevista” se atestado pela entrevistadora de campo responsável pelo endereço e confirmado por sua coordenadora de campo.

2.5.2 Cálculo de alguns indicadores de atrito na PCSVDF-Mulher

De acordo com Andre e Carvalho (2023) na PCSVDF-Mulher nas ondas 1 e 2 foram consideradas entrevistas válidas um total de 9618 e 9845, respectivamente. Apesar de modelar apenas no atrito entre 2016 e 2017, será realizado uma análise geral da perda amostral da PCSVDF-Mulher através da Figura 3.

A Figura 3 representa o quantitativo de domicílios entrevistados e válidos nas quatro ondas da PCSVDF-Mulher. É possível calcular as fórmulas elencadas nos itens abaixo que representam as taxas de atrito entre as ondas da PCSVDF.

$$\begin{aligned} \text{A.) } 1 - \frac{\#PAR_{16,17}}{\#N_{16}} &= 1 - \frac{4316}{9618} = 0,5512 \\ \text{B.) } 1 - \frac{\#PAR_{17,19}}{\#REP_{CE,PE,BA(17)}} &= 1 - \frac{853}{1841} = 0,5367 \\ \text{C.) } 1 - \frac{\#PAR_{16,17,19}}{\#N_{CE,PE,BA(16)}} &= 1 - \frac{962}{3606} = 0,7322 \end{aligned}$$

Considerando as quantidades:

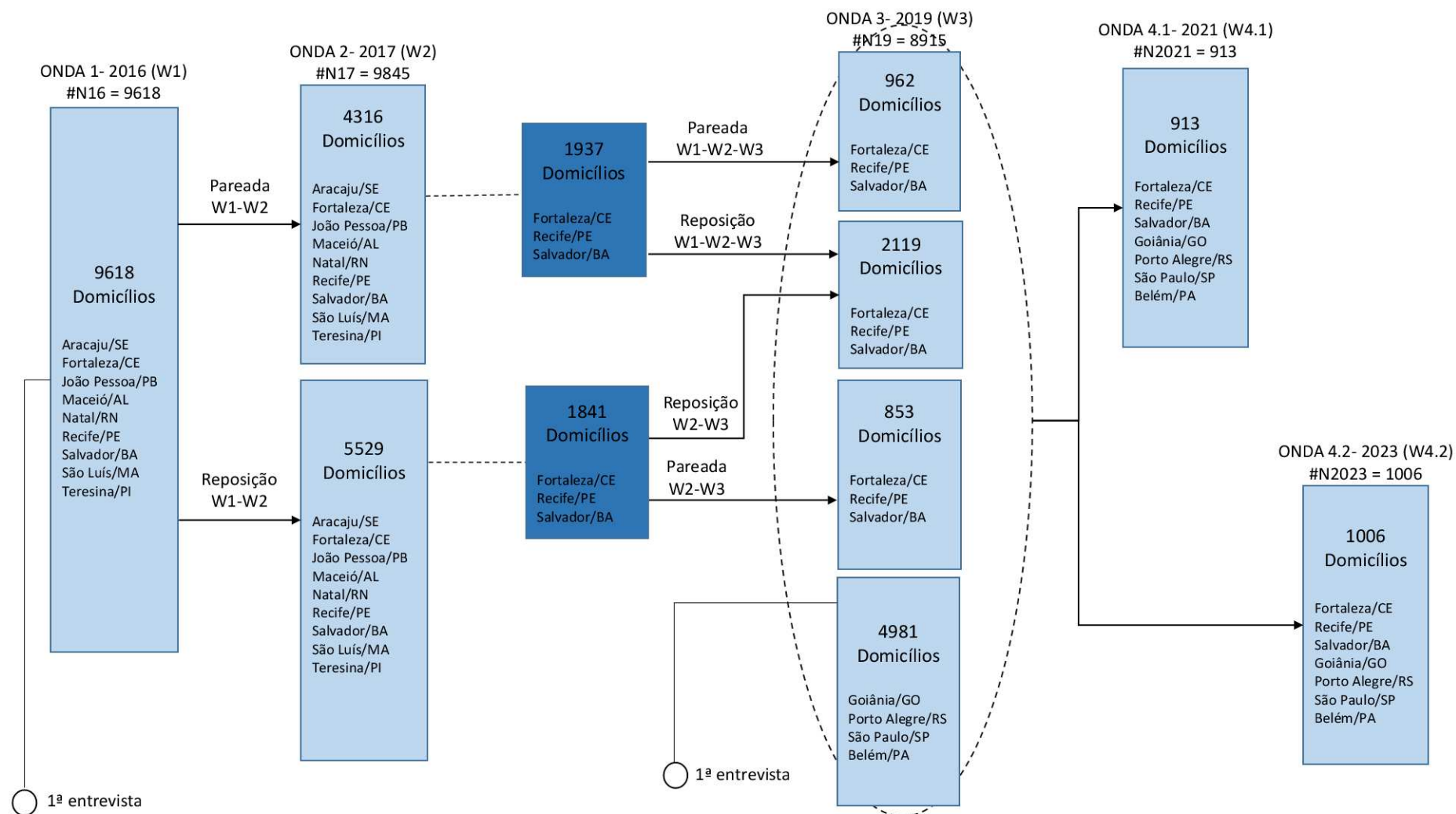
- $\#PAR_{16,17}$: Quantidade de pareados entre as ondas 1 e 2;
- $\#PAR_{17,19}$: Quantidade de pareados entre as ondas 2 e 3;
- $\#PAR_{16,17,19}$: Quantidade de pareados entre as ondas 1, 2 e 3;
- $\#N_{16}$: Quantidade de entrevistas na onda 1;
- $\#REP_{CE,PE,BA(17)}$: Quantidade de reposições na onda 2 nos municípios em específico (1ª visita);
- $\#N_{CE,PE,BA(16)}$: Quantidade de entrevistas na onda 1 nos municípios em específico.

O item A refere-se à taxa de atrito observada entre a primeira e a segunda ondas de coleta de dados. Como todos os municípios estão representados em ambas as ondas, pode-se calcular essa taxa de forma abrangente, utilizando o quantitativo total de participantes sem restrições. Essa abordagem permite uma análise mais precisa e representativa, já que a presença contínua dos mesmos municípios facilita a comparação direta entre os dados coletados em diferentes momentos.

A análise da taxa de atrito entre as ondas 2 e 3 exige uma consideração específica, uma vez que apenas os municípios de Fortaleza, Recife e Salvador participaram dessas fases. Dessa forma, para calcular essa taxa, é considerado somente o número de entrevistas realizadas nesses municípios, levando em consideração a quantidade de domicílios pareados da onda 2 e da onda 3 (que servirá como numerador), e a quantidade de entrevistas que foram reposição na onda 2, nesses município, como denominador. Essa metodologia está detalhada no item B.

O item C apresenta a taxa de atrito observada desde a primeira até a terceira onda da pesquisa. Para isso, utilizamos dados pareados das três ondas, focando nos quantitativos coletados nos três municípios durante a onda 1. Essa abordagem permite uma análise abrangente das mudanças ao longo do tempo, destacando a evolução da taxa de atrito e proporcionando insights sobre a continuidade e a desistência dos participantes nas diferentes fases da pesquisa. A metodologia adotada visa garantir a consistência e a representatividade dos dados, fundamentais para a interpretação dos resultados.

Figura 3 – Fluxograma das ondas da PCSVDF



Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

2.6 Métodos

2.6.1 *Mecanismo/Modelo de atrito escolhido*

Não existe uma literatura consolidada para selecionar um conjunto de variáveis que afetam a resposta em pesquisas de painel. Watson e Wooden (2009), no entanto, elencam alguns fatores que influenciam na resposta.

- A. Localização do entrevistado: Propensão a mudar de endereço, rastreamento dos entrevistados;
- B. Busca do contato com o entrevistado: Acessibilidade do endereço, disponibilidade do entrevistado, número de tentativas do contato e o momento do dia da tentativa;
- C. Aceite do entrevistado: utilização de incentivos, experiência e continuidade do entrevistador, identificação do entrevistado com o estudo, tema da pesquisa e experiências em entrevista anteriores;
- D. Características do entrevistador: Idade e experiência do entrevistador;
- E. Experiência na entrevista: Cooperação do entrevistado, o quanto gostou da entrevista e a facilidade para se responder as perguntas;
- F. Características do entrevistado: Sexo, idade, raça/etnia, estado civil, educação, propriedade de casa própria, renda, trabalho;
- G. Características da localização do domicílio.

Da PCSVDF-Mulher foram, inicialmente, selecionados os seguintes blocos para se obter variáveis independentes (ver tabela 2). Note que para obtermos essas variáveis independentes foi preciso entender como o fluxo de aplicação do questionário se desenvolveu entre 2016 e 2017. Esse fluxo aparece na Figura 4 onde distinguimos claramente os 5 blocos de variáveis para o estudo do atrito amostral.

Bloco 1. CS - características do setor censitário em que o domicílio está inserido;

Bloco 2. X_{vizin} - características da vizinhança do domicílio;

Bloco 3. X_{mulher} - características da mulher que iniciou a pesquisa;

Bloco 4. Entrev - características da entrevistadora que realizou a pesquisa;

Bloco 5. X_{genero} - características da mulher que iniciou a pesquisa;

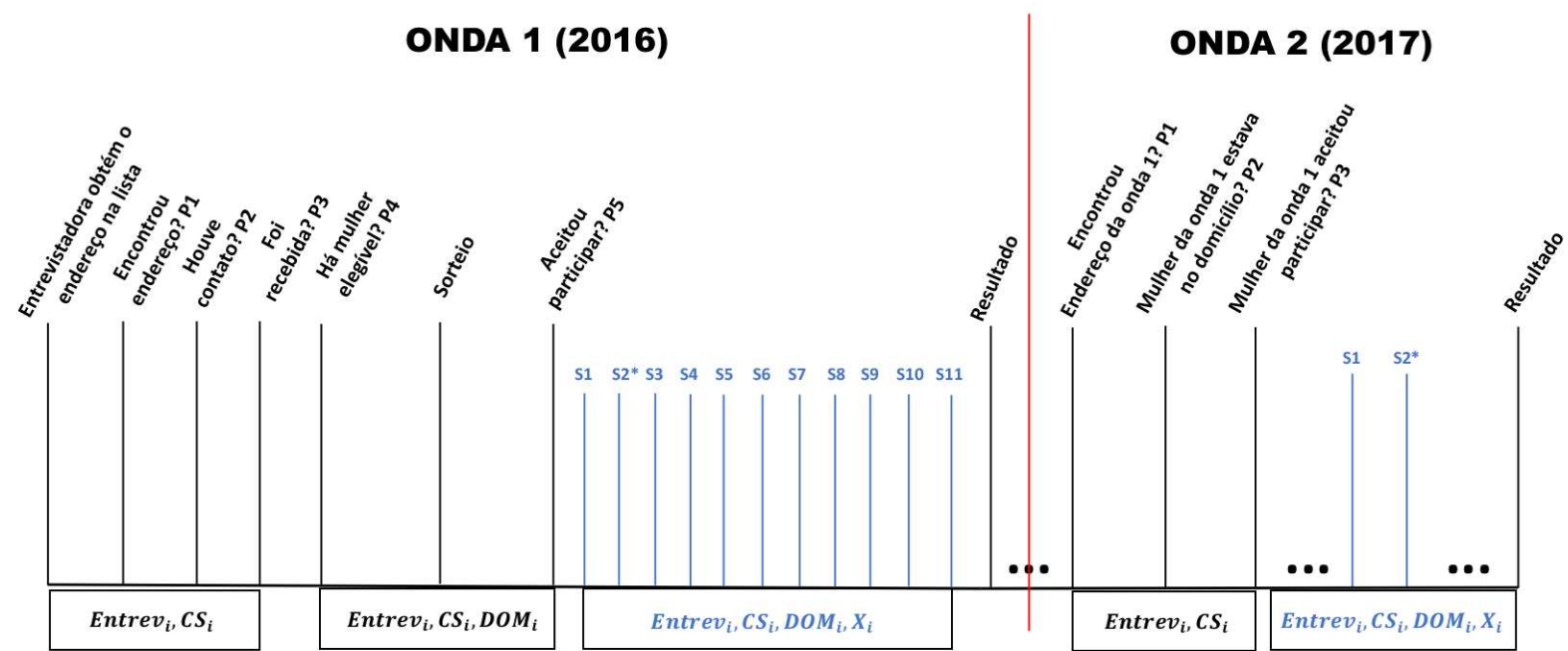
Os blocos que abordam as características da vizinhança do domicílio, as particularidades das mulheres e as variáveis de gênero são extraídos do questionário da PCSVDF-Mulher. Compreender esses aspectos é crucial para a continuidade das pesquisas, especialmente no que diz respeito à participação das mulheres no estudo. Tendo em vista que a cada nova coleta de

dados existe a possibilidade da ocorrência do atrito amostral. Dessa forma, as próximas pesquisas podem se tornar mais robustas e eficazes na formulação de estratégias focadas nas variáveis de características da mulher que influenciam mais no atrito amostral.

Os dados referentes aos setores censitários são provenientes do Censo 2010, especificamente da publicação intitulada “Características da População e dos Domicílios: Resultados do Universo”. Este censo, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), oferece um panorama abrangente sobre as condições sociais e demográficas da população brasileira na época. As informações foram agregadas por setores censitários, permitindo uma análise detalhada das características da população e dos domicílios em diferentes regiões do país.

As variáveis relacionadas às entrevistadoras são coletadas a partir de um questionário específico, que é aplicado no momento da contratação. Esse questionário visa capturar informações relevantes sobre as entrevistadoras, com aproximadamente 60 questões, incluindo dados demográficos, formação acadêmica, experiência profissional e habilidades interpessoais. Essas características são essenciais, pois podem influenciar a qualidade da coleta de dados e a interação com os participantes da pesquisa.

Figura 4 – Esquemático da Onda 1 e 2 e o tipo de variável que pode ser obtida em cada seção



Fonte: Elaborado pelo autor.

A Figura 2 detalha as seções dos questionários das ondas 1 e 2 percorridas pela entrevistada ao longo da sua participação nas duas ondas que são importantes para entender a análise a ser executada no atrito.

Mais especificamente,

- S1- Perfil socioeconômico e mercado de trabalho;
- S2* – Questionário da mulher, saúde geral e reprodutiva (Questionário válido);
- S3 – Conhecimento sobre a violência contra mulher, Lei Maria da Penha;
- S4 – Entrevistada e seu parceiro;
- S5 – Barganha e empoderamento;
- S6 – Experiências de violência doméstica relacionada ao parceiro;
- S7 – Experiências de violência doméstica não-relacionadas ao parceiro;
- S8 – Acesso e utilização da rede de atendimento a mulher;
- S9 – Avaliação do relacionamento;
- S10 – Avaliação do relacionamento;
- S11 – Expectativas subjetivas.

Vale salientar que a escolha de cada bloco de variáveis para a modelagem e as variáveis da entrevistada e do domicílio serão derivados de cada etapa da Figura 2. Na seção S1 são obtidos variáveis de características da mulher, tanto com relação ao seus aspectos socioeconômicos e do mercado de trabalho, e de sua vizinhança; a seção S2 é de fundamental importância, tendo em vista que é a partir dela que um questionário é considerado válido; e por fim, as características com relação a episódios de violência doméstica são obtidos na seção S6.

Tabela 2 – Variáveis do modelo logit e sua descrição

Variável	Descrição
Setor Censitário	
qtd_morad	média do número de moradores em domicílios particulares permanentes
taxa_branco	taxa de residentes que se denominam de cor branca
abast_agua	percentual de domicílios particulares permanentes com abastecimento de água da rede geral
resp_mulh	percentual de pessoas responsáveis pelo domicílio, do sexo feminino
rend_nominal (sal. mínimos)	renda nominal média dos residentes do setor censitário em salários mínimos
Vizinhança	
ilumin_ruas	variável dummy. 0-não percebe má iluminação onde mora, 1-percebe sempre ou frequentemente má iluminação das ruas
patrulha_policial	variável dummy. 0-não percebe patrulha onde mora, 1-percebe sempre ou frequentemente patrulha policial
roubos_assaltos	variável dummy. 0-não percebe roubos ou assaltos onde mora, 1-percebe sempre ou frequentemente roubos ou assaltos onde mora
Entrevistada	
idade_ent	idade em anos da entrevistada
branca	variável dummy 0-não branca, 1-branca
religião_catolica	variável categórica. 0-não católica, 1-católica
escolaridade	variável categórica. 1-sem escolaridade, 2-ensino fundamental, 3-ensino médio/técnico, 4-ensino superior
trabalho_remunerado	variável dummy. 0-não trabalha, 1-trabalha
residencia_atual	variável dummy. 1-não pretende mudar de endereço, 0-talvez/prende mudar de endereço.
domic_propr	variável dummy. 0-casa alugada/cedida, 1-casa própria
cohabita	variável dummy. 0-não mora com o parceiro, 1-mora com o parceiro
filhos	variável dummy. 0-não possui filhos, 1-possui filhos
Entrevistadora	
idade	idade em anos da entrevistadora
branca	variável dummy 0-não branca, 1-branca
experiencia	variável categórica. 0-não, 1-sim
Match entrevistada/entrevistadora	
match_raca	variável dummy. 0-se a raça entre a entrevistada e entrevistadora forem diferentes e 1-se a raça entre a entrevistada e entrevistadora forem iguais
match_idade	variável dummy 1- idade (entrevistada - entrevistadora) ≤ 5 , 0-idade (entrevistada - entrevistadora) > 5
Questões de gênero	
violence	variável dummy. 0-a entrevistada não sofreu nenhum tipo de violência ao longo da vida. 1- a entrevistada sofreu algum tipo de violência ao longo da vida
vio_emo	variável dummy. 0-a entrevistada não sofreu violência emocional ao longo da vida. 1- a entrevistada sofreu violência emocional ao longo da vida
violence12	variável dummy. 0-a entrevistada não sofreu nenhum tipo de violência nos últimos 12 meses. 1- a entrevistada sofreu algum tipo de violência nos últimos 12 meses
vio_emo12	variável dummy. 0-a entrevistada não sofreu violência emocional nos últimos 12 meses. 1- a entrevistada sofreu violência emocional nos últimos 12 meses

Fonte: Elaborado pelo autor.

A Tabela 3 fornece uma análise exploratória, com as medidas de média, desvio padrão, mínimo e máximo, das variáveis pré-selecionadas para os modelos, como já citado por Watson e Wooden (2009) não existe uma literatura consolidada para selecionar um conjunto de variáveis que explicam o atrito amostral em pesquisas longitudinais, no entanto, os seguintes fatores influenciam na resposta: Localização do entrevistado, aceite do entrevistado, características do entrevistador, experiência na entrevista, características do entrevistado, características da localização do domicílio. .

Tabela 3 – Estatística descritiva das variáveis quantitativas do modelo

	Média	Desvio padrão	Mínimo	Máximo
Setor				
qtd_morad	3,4620	0,2957	2,0900	5,9600
taxa_branco	0,3322	0,1310	0,0000	0,9274
abast_agua	0,9324	0,1786	0,0000	1,0000
resp_mulh	0,4602	0,0974	0,0952	0,8507
taxa_alfabe	0,8602	0,0680	0,4508	1,0000
rend_nominal (salario minimo)	1,3010	0,9818	0,1612	11,7300
Entrevistada				
idade_entrevis	32,2000	9,9747	15	49
branca	0,2417	0,4281	0	1
catolica	0,5072	0,4999	0	1
escolaridade	2,9480	0,6857	1	4
cohabit	0,4915	0,4999	0	1
trabalho_remunerado	0,3887	0,4874	0	1
domic_proprio	0,6356	0,4812	0	1
filhos	0,5808	0,4934	0	1
ilumina_ruas	0,3048	0,4603	0	1
patrulha_policial	0,2659	0,4418	0	1
roubos_assaltos	0,5639	0,4959	0	1
Entrevistadora				
escolaridade	1,9200	0,7046	1	3
experiencia	0,4779	0,4995	0	1
Match entrevistada/entrevistadora				
match_idade	0,4541	0,4979	0	1
match_raca	0,4004	0,4900	0	1
Gênero				
violencia	0,3640	0,4811	0	1
vio_emo	0,3050	0,4605	0	1
violencia_12	0,1600	0,3668	0	1
vio_emo12	0,1370	0,3442	0	1

Fonte: Elaborado pelo autor.

Como o questionário da PCSVDF-Mulher é extenso, existem dezenas de variáveis independentes passíveis de serem utilizadas para se prever o atrito. Não necessariamente todas essas variáveis podem ser empregadas para o modelo de forma adequada. Apesar da tabela 2

apresentar um conjunto de variáveis pré-selecionadas para estimar o atrito, tem-se a necessidade de encontrar técnicas que estimam modelos de escolha discreta e ao mesmo tempo escolher de maneira ótima as variáveis.

De acordo com Bag *et al.* (2022) os métodos baseados em penalidades são as técnicas mais discutidas na última década para selecionar um conjunto de variáveis independentes para um modelo. Tibshirani (1996) desenvolve o Least Absolute Shrinkage and Selection (LASSO), que foi avaliado positivamente por conta das melhores propriedades de seleção de variáveis. Neste método, são impostas restrições aos valores dos coeficientes de regressão através de uma função de penalidade (λ). O método LASSO reduz os valores próximos de zero dos coeficientes de regressão para precisamente zero, sugerindo assim que essas variáveis independentes não têm qualquer efeito no modelo.

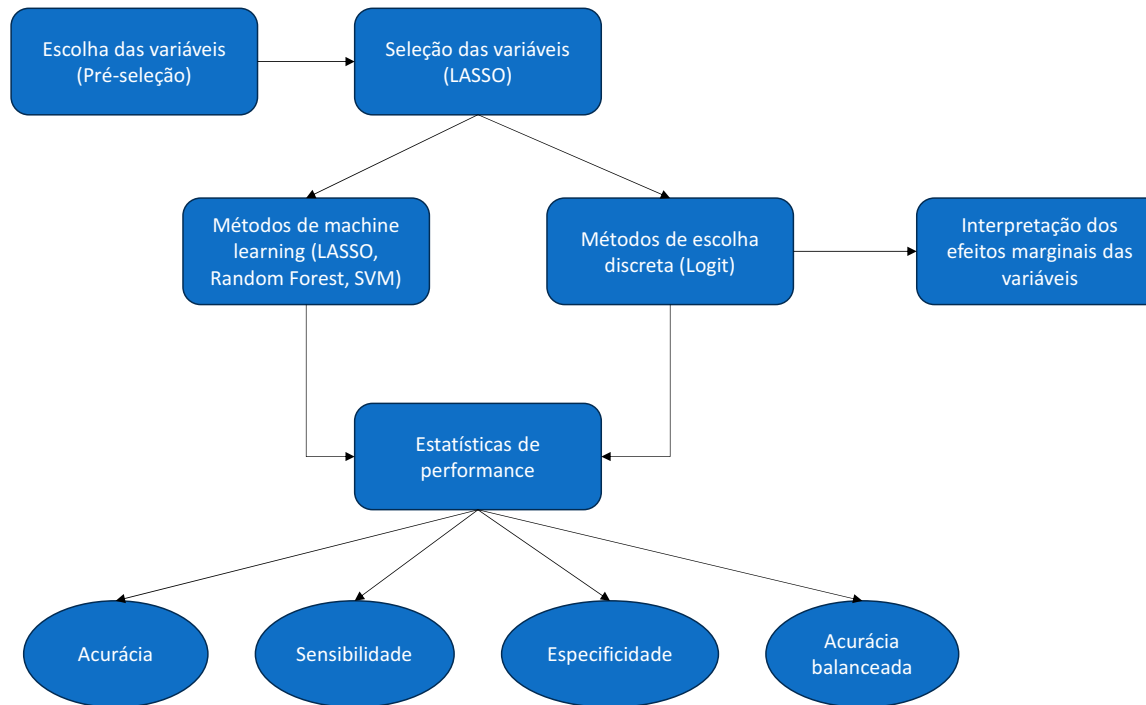
Com isso, a abordagem de machine learning do presente estudo se dará da seguinte forma:

1. Seleção das variáveis de acordo com o método LASSO no modelo especificado;
2. Utilização das variáveis selecionadas no método LASSO nas demais técnicas de machine learning;
3. Cálculo das estatísticas de performance dos métodos nos quatro modelos.

Tendo em vista que a abordagem de machine learning possui limitações na interpretabilidade dos seus resultados, após a sua análise serão selecionadas as variáveis independentes de cada modelo e estimado o modelo logit com essas variáveis pré-selecionadas, com isso será possível interpretar os resultados dos coeficientes estimados.

Serão utilizados três métodos de machine learning para verificar a previsão do atrito na pesquisa. Os métodos utilizados serão Random Forests, Suport Vector Machine (SVM) e LASSO. Além dos métodos de machine learning, será feito o mesmo estudo comparativo com a regressão logística. Os três métodos são descritos na seção a seguir.

Figura 5 – Diagrama de seleção das variáveis e estimação do modelo



Fonte: Elaborado pelo autor.

2.6.2 Atrito na Perspectiva de Aprendizado de Máquina: Estimção e Previsão

2.6.2.1 Método LASSO

A vantagem do presente trabalho é da utilização de dados relativamente ricos coletados na primeira onda para prever a segunda onda. No entanto, nem todas as variáveis são igualmente importantes para prever a participação na onda 2.

O método LASSO tem como objetivo encontrar um equilíbrio entre simplicidade e precisão do modelo. Ele consegue isso adicionando um termo de penalidade ao modelo de regressão linear tradicional, em que alguns coeficientes são forçados a ser exatamente zero. Esse recurso torna o LASSO particularmente útil para seleção de recursos, pois pode identificar e descartar automaticamente variáveis irrelevantes ou redundantes.

Esse modelo também é conhecido como modelo de regressão penalizado porque LASSO restringe ou regulariza as estimativas de coeficiente para zero. Isso é obtido por meio de um parâmetro de ajuste. Um grande definirá alguns coeficientes exatamente iguais a zero, de modo que essas variáveis sejam automaticamente eliminadas do modelo, deixando apenas aquelas que são mais preditivas das variáveis de resultado. Além disso, restringe a magnitude das correlações entre as variáveis; Chudik *et al.* (2023)

A regressão de mínimos quadrados tenta encontrar estimativas de coeficientes que minimizem a soma dos quadrados dos resíduos (RSS):

$$RSS = \sum_{i=1}^n [y_i - \hat{y}_i]^2$$

A regressão LASSO procura minimizar a seguinte função:

$$RSS + \lambda \sum |\beta_j|$$

Em que j varia de 1 a n variáveis explicativas e $\lambda \geq 0$. O termo $\lambda \sum |\beta_j|$ é definido como regularização L1.

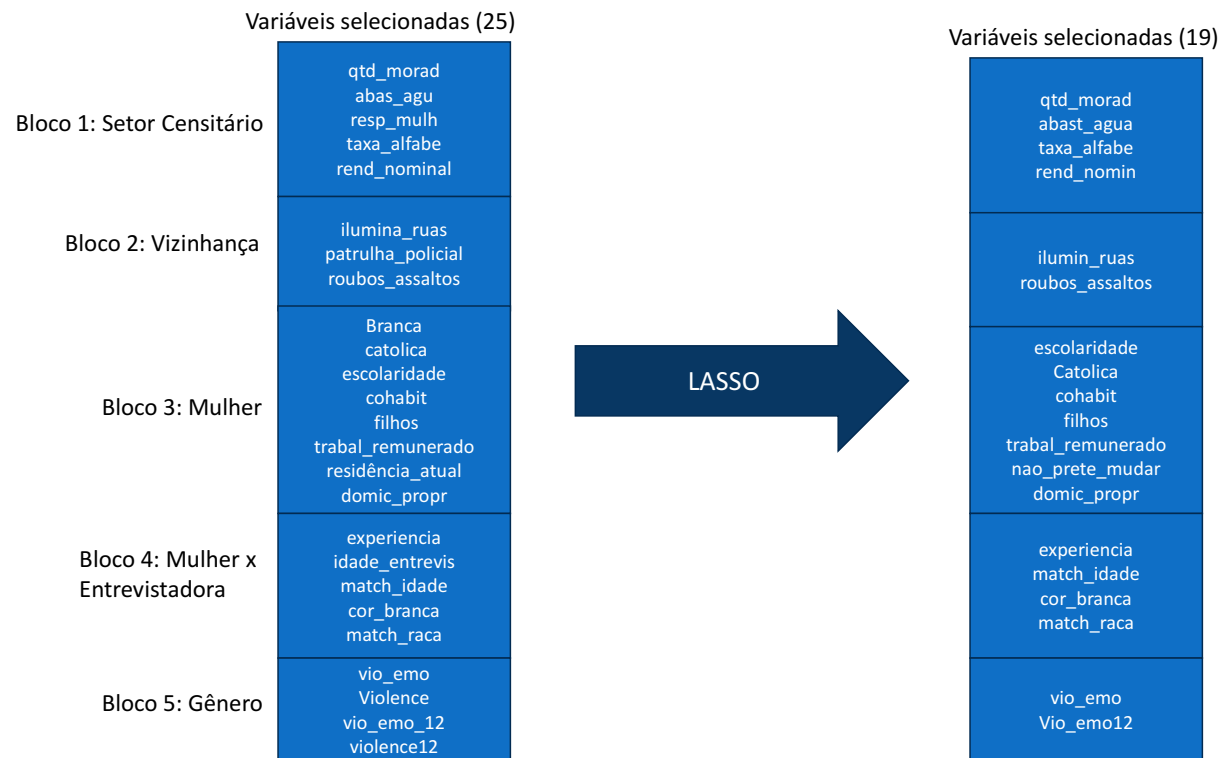
Além de diminuir a variância do modelo, essa regularização tem uma importante aplicação em machine learning. Quando há variáveis altamente correlacionadas, que se comportam da mesma maneira, a regularização LASSO seleciona apenas uma delas e zera os coeficientes das outras, de forma a minimizar a penalização L1. Com isso, são gerados vários coeficientes com peso zero, ou seja, que são ignorados pelo modelo. A figura 7 indica as variáveis que apresentam valor diferente de zero e que foram selecionadas para a previsão do atrito. Vale ressaltar que o

valor de λ escolhido para o método LASSO é definido de forma que ocorra um maior valor da acurácia do método, foi encontrado o valor de $\lambda = 0,0058$.

O método LASSO selecionou, do conjunto de variáveis independentes pré-selecionadas, de um total de 25 variáveis 19 tiveram seus coeficientes diferentes de zero nos modelos. A Figura 6 apresenta o conjunto de variáveis que foram pré-selecionadas para se prever o atrito amostral e as variáveis que foram selecionadas pelo método LASSO. O modelo utilizado a partir de então levará em consideração as 19 variáveis selecionadas pelo LASSO.

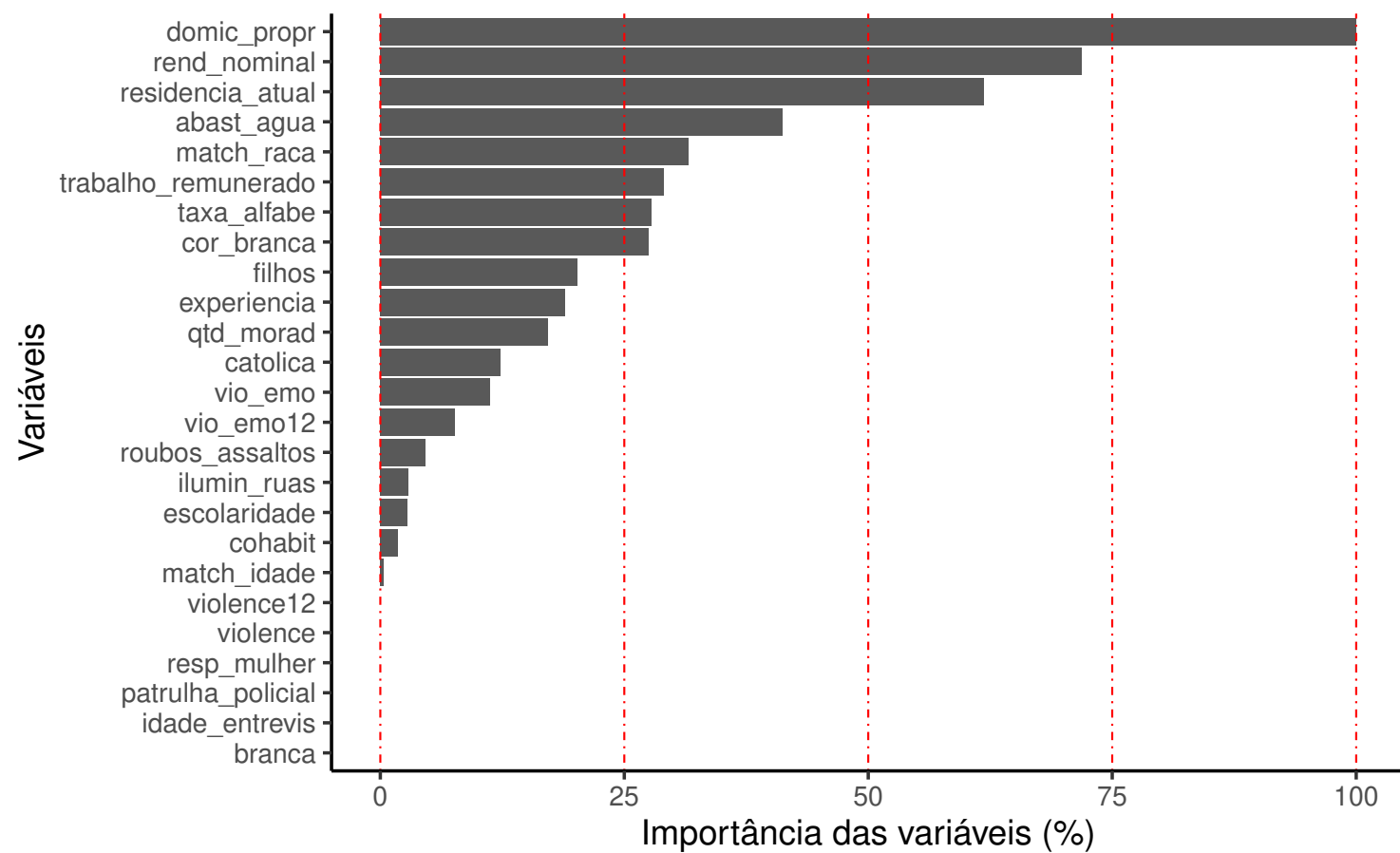
A Figura 7 mostra o gráfico de importância de variáveis do modelo em questão que é essencial para entender como diferentes características influenciam as previsões de um modelo de machine learning. Esse tipo de abordagem tem como objetivo mensurar globalmente a importância de cada variável para o modelo, em que mostra a ordem das variáveis do maior para o menor. As variáveis `domic_propr`, `rend_nominal` e `residencia_atual` foram as principais variáveis que contribuíram para a previsão do atrito no modelo. A partir da variável `match_idade` as demais tem pouca ou nenhuma importância, tendo em vista que a sua importância foi igual a 0.

Figura 6 – Diagrama com as variáveis selecionadas pelo método LASSO



Fonte: Elaborado pelo autor.

Figura 7 – Variáveis, de acordo com a importância, método LASSO



Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

2.6.2.2 *Random Forests*

Para entender o método de random forest é necessário compreender o significado de árvores de decisões, que são estruturas similares a um fluxograma, que tem nós com uma determinada condição verificada, e caso seja atendida, o fluxo segue por determinado ramo, caso contrário, irá para outro ramo, sempre para o próximo nó, até finalizar a árvore.

Uma das técnicas de machine learning é o aprendizado conjunto (ensemble learning), que são combinações de um número finito de modelos para obter uma melhor performance preditiva. O método random forest é um exemplo disso. Em tradução para o português, florestas aleatórias, tem em sua base a criação de um conjunto, de forma aleatória, de árvores de decisões, obtidas a partir de amostras da base de dados.

Cada árvore teria a sua própria base de dados e, com isso, teria previsões diferentes, entretanto cada árvore ficaria com poucos dados de treino. Dessa forma, a técnica de bootstrapping seria a solução, um método de reamostragem, em que as amostras selecionadas podem ser repetidas na seleção. Segundo Louppe (2015) é introduzida uma aleatoriedade nas árvores de decisão e com isso reduz os erros de previsão ao descorrelaciona-la.

Para a construção de uma árvore de decisão é preciso definir o nó raiz, que será a primeira condição verificada, dando origem aos dois primeiros ramos. Com isso, será utilizado o algoritmo de entropia ou o índice Gini, para a escolha da melhor variável para compor o nó raiz, variando de acordo com o método utilizado.

- Passo 1: No Random Forest, a definição desta variável não acontece com base em todas as variáveis disponíveis. O algoritmo irá escolher de maneira aleatória (random) duas ou mais variáveis, e então realizar os cálculos com base nas amostras selecionadas, para definir qual dessas variáveis será utilizada no primeiro nó.

- Passo 2: Para escolha da variável do próximo nó, novamente serão escolhidas duas (ou mais) variáveis, excluindo as já selecionadas anteriormente, e o processo de escolha se repetirá. Desta forma, a árvore será construída até o último nó. A quantidade de variáveis a serem escolhidas pode ser definida na criação do modelo.

Na construção da próxima árvore, os dois processos anteriores se repetirão, levando a criação de uma nova árvore. Provavelmente essa árvore será diferente da primeira, pois tanto na seleção das amostras, quanto na seleção das variáveis, o processo acontece de maneira aleatória. Esse processo ocorre de forma que um número finito, definidos pelo pesquisador, de árvores sejam criadas.

Cada árvore criada irá apresentar o seu resultado e, quando apresentar problemas de regressão, será realizada a média dos valores previstos, e esta média informada como resultado final, e em problemas de classificação o resultado que mais vezes foi apresentado será o escolhido.

2.6.2.3 *Support Vector Machines*

Outro método de classificação é o algoritmo Support Vector Machine (SVM). De acordo com Bertsimas e Margaritis (2023) SVM são modelos de machine learning que usam um hiperplano adequado para fazer previsões, seja para classificação ou regressão. Nesse algoritmo, cada item da base de dados é plotado como um ponto no espaço n-dimensional (em que n é o número de variáveis do modelo), com o valor de cada variável sendo uma determinada coordenada. Para o exemplo de classificação do presente estudo ser binário- presença ou não de atrito- o objetivo é encontrar um hiperplano que melhor diferencie as duas classificações. Isto é, o SVM escolhe a reta - ou hiperplano em maiores dimensões - entre os dois grupos que se distancia mais de cada um.

De acordo com Liu (2020) esse tipo de modelo se destaca na classificação de dados espalhados de maneira não regular, já que a separação não precisa ser linear e nem a mesma para todos os dados. Além disso, o SVM é eficiente em espaços com muitas dimensões (muitas variáveis) e converge para o melhor hiperplano possível. Contudo, o seu resultado é dificilmente interpretável (mas possível) e, conforme o tamanho da base de dados for aumentando, o tempo necessário para fazer os cálculos cresce rapidamente e a interpretabilidade cai.

Segundo Lorena *et al.* (2007) SVMs podem ser classificados em lineares e não lineares. Dentro da classe dos lineares existem os SVMs com margens rígidas, que recebem essa nomenclatura pois suas restrições são impostas de maneira a assegurar que não haja dados de treinamento entre as margens de separação das classes. E existem também os SVMs com margem suaves que permitem que algumas restrições impostas sejam violadas para tratar um conjunto de dados de treinamento mais geral (presença de ruído ou outliers). A aplicação desse procedimento suaviza as margens do classificador linear, permitindo que alguns dados permaneçam entre os hiperplanos.

2.7 Resultados

A seção de resultados é dividida em duas partes, a primeira engloba do cálculo das métricas de desempenho do modelo de acordo com os quatro métodos. A segunda etapa trata da estimação de um modelo logit com as variáveis selecionadas, de acordo com a Figura 7.

De forma usual, para avaliar o desempenho dos modelos de machine learning, é necessário a divisão, de forma aleatória, do conjunto de dados em dois grupos, o conjunto de treinamento (para desenvolver os modelos) e o conjunto de testes (para avaliar o desempenho do modelo). Para o presente estudo, a base de dados foi dividida da forma que 70% dos respondentes da onda 1 foram classificados como conjunto de treinamento e os demais 30% para o conjunto de teste.

A partir do conjunto de dados de teste são calculadas algumas métricas para avaliar a performance dos modelos. As métricas abaixo, de acordo com a definição de Mariano (2021), serão utilizadas para realizar a avaliação.

- A acurácia tem o objetivo de avaliar o percentual de acertos, ou seja, é obtida pela razão entre a quantidade de acertos e o total de observações;
- A especificidade avalia a capacidade do método de detectar resultados negativos, ou seja, entrevistadas que participaram da pesquisa nas duas ondas (não atritatos);
- A sensibilidade avalia a capacidade do método de detectar com sucesso resultados classificados como positivos, ou seja, entrevistadas não participaram da pesquisa na onda 2 (atritados);
- Por fim, a acurácia balanceada é dada pela média da especificidade e a sensibilidade.

O resultado dessa estatística está apresentada na Tabela 4 e destacado os melhores métodos para as diferentes estatísticas. De acordo com James *et al.* (2013) as estatísticas são calculadas a partir da matriz de confusão de cada modelo, que apresentam os resultados reais nas colunas e os resultados previstos nas linhas para os quatro métodos usados na análise. Vale ressaltar novamente que os dados são baseados apenas no conjunto de teste.

Os resultados apresentados mostram que o método RANDOM FOREST previu 61,99% das observações, conforme indicado pela acurácia. Além disso, destacou-se na previsão correta de "positivos reais"(atritados), com uma sensibilidade de 36,97%, superando outros métodos.

A especificidade, presença de “negativos reais” (não atritatos), foi a única estatística de performance que o RANDOM FOREST não foi superior, sendo para essa estatística o método

Tabela 4 – Estatísticas de precisão do modelo para prever atrito. LASSO, SVM, Random Forest e Regressão Logística.

	LASSO	LOGIT	RANDOM	SVM
Acurácia	0,6168	0,6118	0,6199	0,6142
Sensibilidade	0,3208	0,3103	0,3697	0,2739
Especificidade	0,8104	0,8312	0,8020	0,8619
Acurácia Balanceada	0,5656	0,57075	0,5858	0,5679

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

SVM superior aos demais, especificidade igual a 86,19%. E, por fim, a acurácia balanceada foi superior no método RANDOM FOREST, aproximadamente igual a 60%.

O método RANDOM FOREST demonstrou ser o mais eficaz na previsão do atrito, apresentando resultados superiores em comparação com outras abordagens, Tabela 4. A precisão das previsões geradas por esse método destaca sua utilidade em contextos onde a identificação de fatores que afetam o atrito amostral é crucial para a realização de estudos longitudinais mais robustos e confiáveis.

Embora os resultados do método RANDOM FOREST tenham mostrado um desempenho superior, será estimado também o modelo logit devido à sua maior interpretabilidade. O modelo logit permite uma compreensão mais clara das relações entre as variáveis independentes e a variável dependente, facilitando a identificação de quais fatores têm um impacto significativo nas previsões. Essa transparência é crucial para a tomada de decisões informadas e a comunicação dos resultados, especialmente em contextos onde a interpretação dos dados é essencial.

Os resultado da estimação logit do modelo são exibidos pela Tabela 5. Dados que indicam as características do setor censitário do domicílio estão inseridos no primeiro bloco. A variável *abast_agua* (indicador de saneamento básico) apresentou valor significativo com efeito marginal igual a 0,1357 e *rend_nominal*, que representa a renda média do setor censitário apresentou efeito marginal igual a 0,0469. Esse resultado indica que setores censitários mais ricos apresentaram uma taxa de atrito maior na pesquisa, sugerindo que indivíduos em áreas de maior renda podem estar mais propensos a se desengajar ao longo do tempo. Essa tendência pode ser atribuída a fatores como mobilidade social e mudanças frequentes de residência, que são mais comuns em populações com maior poder aquisitivo.

Em contrapartida, entrevistadas residentes de setores com maiores taxas de alfabetização tem uma probabilidade menor de atritar, valor marginal igual a -0,3637. A taxa de alfabetização do setor censitário teve um efeito negativo na taxa de atrito da pesquisa, indicando

que áreas com maior nível de alfabetização tendem a apresentar menor perda de participantes ao longo do tempo. Esse fenômeno pode ser explicado pela maior capacidade dos indivíduos alfabetizados em compreender a importância da pesquisa e em manter um compromisso com a participação. Assim, a relação entre alfabetização e atrito sugere que um maior nível educacional pode contribuir para a retenção de participantes em estudos longitudinais.

Em relação as percepções sobre a vizinhança feita pela entrevistada, questões sobre a má iluminação das ruas, patrulha policial e roubos/assaltos não foram relevantes para explicar a ocorrência do atrito na entrevista, haja vista que todas as estimativas terem sido estatisticamente não significantes.

Para o bloco de características da entrevistada, as variáveis escolaridade, religião católica e cohabitar com o parceiro não apresentaram valor significativo para o modelo. As variáveis possuir filhos, domicílio próprio e não pretende mudar de endereço apresentaram efeitos marginais negativos. A questão da moradia em domicílio próprio indica que ela tem uma probabilidade maior de estar presente na mesma residência na entrevista seguinte, diminuindo o atrito, além da variável não pretender mudar de endereço nos próximos 12 meses que também se mostrou negativamente correlacionada com o atrito.

Outro fator importante correlacionado com o atrito, desta vez positivamente, é o fato da mulher trabalhar e ter uma maior probabilidade de não estar em casa na próxima visita. Essa relação pode ser atribuída a fatores como a maior carga de trabalho e a mobilidade profissional, que podem dificultar a continuidade da participação. Além disso, a busca por melhores oportunidades de emprego pode levar os participantes a se afastarem da pesquisa.

Por fim, no bloco 4, a experiência da entrevistadora está associada ao aumento do atrito, sugerindo que entrevistadoras mais experientes podem, paradoxalmente, provocar uma maior desengajamento dos participantes, vale salientar que essa experiência não necessariamente se dá em pesquisas sobre violência da mulher. Isso pode ocorrer devido a uma abordagem mais incisiva ou a expectativas elevadas durante as entrevistas, levando os indivíduos a se afastarem.

Além disso, mulheres da cor branca e a relação da raça entre a entrevistada e a entrevistadora possuem efeito marginal igual a -0,0415 e -0,0480, respectivamente. A semelhança racial entre entrevistadora e entrevistadoa tende a reduzir o atrito, pois essa identificação pode criar um ambiente mais confortável e de confiança durante as entrevistas. Quando ambos compartilham a mesma raça, pode haver uma maior empatia e compreensão das experiências, o que facilita a comunicação e o engajamento. Essa conexão pode incentivar os participantes a

permanecerem na pesquisa, resultando em menores taxas de atrito ao longo do tempo.

Tabela 5 – Estimação logit do modelo

	Ef.marginal	error	t.value	p-value
Bloco 1: Setor Censitário				
qtd_morad	0,0329	0,0280	1,1751	0,2400
abast_agua	0,1357	0,0501	2,7104	0,0067**
taxa_alfabe	-0,3297	0,1446	-2,2802	0,0226**
rend_nominal	0,0469	0,0108	4,3642	<0,001***
Bloco 2: Vizinhança				
ilumin_ruas	-0,0028	0,0170	-0,1638	0,8699
roubos_assaltos	-0,0134	0,0159	-0,8432	0,3992
Bloco 3: Entrevistada				
escolaridade	0,0021	0,0046	0,4577	0,6472
catolica	-0,0122	0,0158	-0,7728	0,4397
filhos	-0,0476	0,0183	-2,5956	0,0095**
trabalho_remunerado	-0,0400	0,0162	-2,4714	0,0135**
residencia_atual	-0,1126	0,0183	-6,1427	<0,001***
domic_propr	-0,1261	0,0170	-7,4044	<0,001***
cohabit	0,0145	0,0167	0,8689	0,3850
Bloco 4: Entrevistada x Entrevistadora				
experiencia	0,0327	0,0164	1,9954	0,0461**
match_idade	0,0156	0,0164	0,9520	0,3411
cor_branca	-0,0415	0,0192	-2,1595	0,0309**
match_raca	-0,0480	0,0190	-2,5265	0,0116**
Bloco 5: Gênero				
vio_emo	-0,0009	0,0222	-0,0395	0,9685
vio_emo12	0,0235	0,0293	0,8030	0,4220

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001; **p-valor<0,05; *p-valor<0,1

2.8 Considerações finais

O presente estudo investigou o atrito na PCSVDF-Mulher, pesquisa que tem como objetivo principal estudar a violência doméstica e seus efeitos. São utilizadas duas abordagens para estudar o atrito na PCSVDF-Mulher. A primeira modelagem, utilizando modelos de regressão logística tradicional e uma segunda abordagem utilizando métodos de machine learning (aprendizagem de máquina).

Para atingir esse objetivo foram analisadas características socioeconômicas dos setores censitários, características do domicílio, características das mulheres (com informações de questões de gênero), da entrevistadora e variáveis que consideram o match entre entrevistada

e entrevistadora (idade e raça).

A modelagem logística ocorreu através da estimação de um modelo, considerando a variável resposta sendo binária em que ocorreu atrito ou não entre as duas primeiras ondas da PCSVDF-Mulher, a seleção de variáveis ocorre por um método de machine learning que seleciona as variáveis que apresentam uma maior importância na previsão da variável resposta.

Os métodos de machine learning utilizados são random forest, SVM e LASSO. A base de dados foi dividida de forma que 70% dos dados fiquem na base de treino e 30% na base de teste, com isso serão calculadas algumas estatísticas a partir da previsão utilizando a base de teste, além de um estudo que determina quais variáveis foram mais importantes para o desgaste amostral.

O conjunto de variáveis escolhido, de acordo com o método LASSO, apresentou em sua seleção duas variáveis de questões de gênero (que é de fundamental importância, tendo em vista a temática da pesquisa), episódios de violência geral e violência emocional.

Com relação a variáveis do setor censitário, a taxa de abastecimento de água e a taxa de pessoas brancas foram positivamente correlacionadas com o atrito. A taxa de domicílios com abastecimento de água é um indicador de saneamento básico e de desenvolvimento social e econômico, com isso o resultado mostra que mulheres residentes em setores mais desenvolvidos tem maior probabilidade de atrito.

O principal determinante das características da mulher foi se ela tem filhos, se ela tem domicílio próprio, intenção de mudar, se ela tem trabalho remunerado. o primeiro indicando que ela tem maior propensão a não está na mesma residência na próxima onda e o segundo ter mais chance de não está em casa no momento da segunda entrevista.

Entrevistadoras com experiência em pesquisas anteriores tiveram relação positiva com o atrito, além disso o match entre a raça da entrevistada e entrevistadora também apresentou valor negativo, ou seja, se ambas possuírem a mesma raça a probabilidade de atrito entre as ondas tende a diminuir.

O objetivo da pesquisa está em entender como os fatores selecionados afetaram o atrito na PCSVDF-Mulher. Com esses resultados, é possível reduzir o viés de atrito por meio de uma melhor compreensão desse fenômeno. Consequentemente, elaborar melhores estratégias para minimizar o desgaste da entrevistada em futuras ondas.

3 CAPÍTULO 02 - AVALIAÇÃO DO PROTOCOLO DE SUBSTITUIÇÃO DE UMA PESQUISA LONGITUDINAL COMPLEXA: O CASO DA PCSVDF-MULHER

3.1 Introdução

Em pesquisas longitudinais, as medidas de interesse são estimadas a partir de observações coletadas ao longo do tempo. A precisão dessas estimativas pode ser afetada, para além de erros aleatórios e sistemáticos, pelo fenômeno do atrito amostral.

Como já discutido no capítulo anterior, o atrito amostral pode afetar as estimativas, e esse problema tende a crescer à medida que aumenta a taxa de não resposta da unidade amostral. Outro fator advindo do atrito amostral é a ampliação da variância das estimativas, tendo em vista que o tamanho da amostra se reduz em relação ao tamanho original.

Portanto, para evitar esses problemas, ao planejar e conduzir o desenho amostral de um projeto de pesquisa envolvendo coleta primária de dados é necessário tomar algumas medidas para alcançar uma taxa de resposta elevada e lidar com os efeitos da não resposta. A primeira medida foca na melhora do processo de coleta de dados para reduzir a proporção de unidades não contactadas ou que se recusam a participar. A segunda medida utiliza diferentes métodos que visam minimizar os efeitos da não resposta nos resultados. Ou seja, enquanto a primeira medida atua na fase do desenho amostral, antes da coleta de dados, a segunda medida é corretiva, e é aplicada após a coleta.

Dentre os métodos que são comumente adotados para lidar com esses efeitos está a substituição de campo, que terá um papel importante no presente estudo. A substituição de campo é utilizada para substituir unidades que não respondem por outras. Na maioria das vezes, estas são pré-definidas de acordo com um protocolo de substituição no processo de construção da pesquisa.

A substituição de campo é vantajosa porque mantém o tamanho e estrutura da amostra, garantindo o número originalmente planejado de observações para cada parte da amostra. Isso é especialmente adequado para pesquisas que envolvem uma amostra extensivamente estratificada. Além disso, quando são encontradas unidades inelegíveis, estas são substituídas, enriquecendo o conjunto de dados com informações que, caso contrário, estariam perdidas.

A PCSVDF-Mulher possui um protocolo de substituição que foi criada pela equipe de pesquisadores e foi acordada com a empresa contratada responsável pela execução da coleta de dados. Pelo exposta anteriormente faz-se necessário estudar o processo de substituição

empregado na PCSVDF-Mulher, haja vista que é de suma importância que o processo tenha sido bem feito e de acordo com o protocolo definido na pesquisa, uma vez que o não cumprimento do protocolo poderá causar viés nas estimativas. Com isso, o objetivo do presente estudo é verificar se o processo de substituição da PCSVDF-Mulher entre as ondas 1 e 2 foi cumprido e se foi eficaz.

O presente trabalho analisa o protocolo de substituição de acordo com três critérios. Critério 1: Manutenção do tamanho amostra; Critério 2: Perfil espacial de substituição; Critério 3: Preservação populacional da representatividade amostral da PCSVDF-Mulher. O critério 1 se dá no sentido de verificar se o quantitativo de entrevistas válidas foi semelhante entre as ondas. Para isso, foi analisado o número de entrevistas válidas nas duas ondas a nível de pesquisa, no nível de cidade, no nível de bairro e no nível de setor censitário. O critério 2 ocorre na forma de verificar, no menor nível de desagregação geográfica (setor censitário), se as substituições ocorreram no mesmo setor censitário ou em setores próximos - abordagem espacial - e o critério 3 vai de encontro a testar se a amostra de substituição possui o mesmo perfil da amostra que atritou na pesquisa, para isso serão realizados testes estatísticos de acordo com o tipo da variável.

O presente capítulo é dividido em oito seções. A segunda seção aborda a literatura teórica sobre o processo de substituição em pesquisas, evidenciando sua importância. A terceira seção analisa alguns exemplos de protocolos de substituição utilizados em diferentes pesquisas ao redor do mundo, bem como o protocolo de substituição da PCSVDF-Mulher. A partir da seção quatro se dá o início do processo de adequabilidade ao protocolo de substituição, realizando uma análise para verificar os três critérios definidos anteriormente. A seção cinco engloba o critério 1, com o objetivo de determinar se, em termos quantitativos, a pesquisa manteve sua representatividade. Na seção seis, o critério 2 é explorado abordando aspectos espaciais, com o objetivo de verificar se a empresa contratada seguiu o protocolo e a seção sete, critério 3, versa sobre um comparativo entre as variáveis do recorte de pessoas que saíram da PCSVDF-Mulher e as que as substituíram, e com isso um processo de verificação de eficiência do protocolo. E, finalmente, a última seção traz as considerações finais.

3.2 Revisão de literatura

Nishimura (2015) afirma que a não resposta é um problema significativo para pesquisas de opinião. O fenômeno abrange a não resposta tanto no nível da unidade quanto no nível do item. A não resposta da unidade refere-se à ausência completa de uma entrevista por parte de

um domicílio amostrado, enquanto a não resposta do item refere-se à ausência de respostas a perguntas específicas na entrevista depois que um integrante do domicílio amostrado concordar em participar da pesquisa. A não resposta da unidade é geralmente considerada uma ameaça muito maior para as pesquisas de opinião do que a não resposta do item, introduzindo distorções na inferência estatística; como o viés de seleção, a perda de representatividade, a subestimação ou superestimação dos parâmetros; e o aumento da variabilidade.

Segundo Baldissera *et al.* (2014), os métodos de sobreamostragem, ponderação e substituição de campo são os mais amplamente adotados para lidar com o fenômeno da não resposta. Tais ferramentas usam abordagens distintas, mas que podem ser combinadas na mesma pesquisa.

Os pesquisadores enfrentam desafios ao tentar avaliar o impacto da não resposta nas análises de casos completos, utilizando apenas os dados coletados, já que isso limita a compreensão da magnitude da degradação das análises. Para superar essa limitação, é essencial recorrer a fontes externas de informação. Nesse contexto, as amostras de reposição ou substituição surgem como uma alternativa valiosa, proporcionando dados adicionais que podem enriquecer a análise e minimizar os efeitos da não resposta Deng *et al.* (2013).

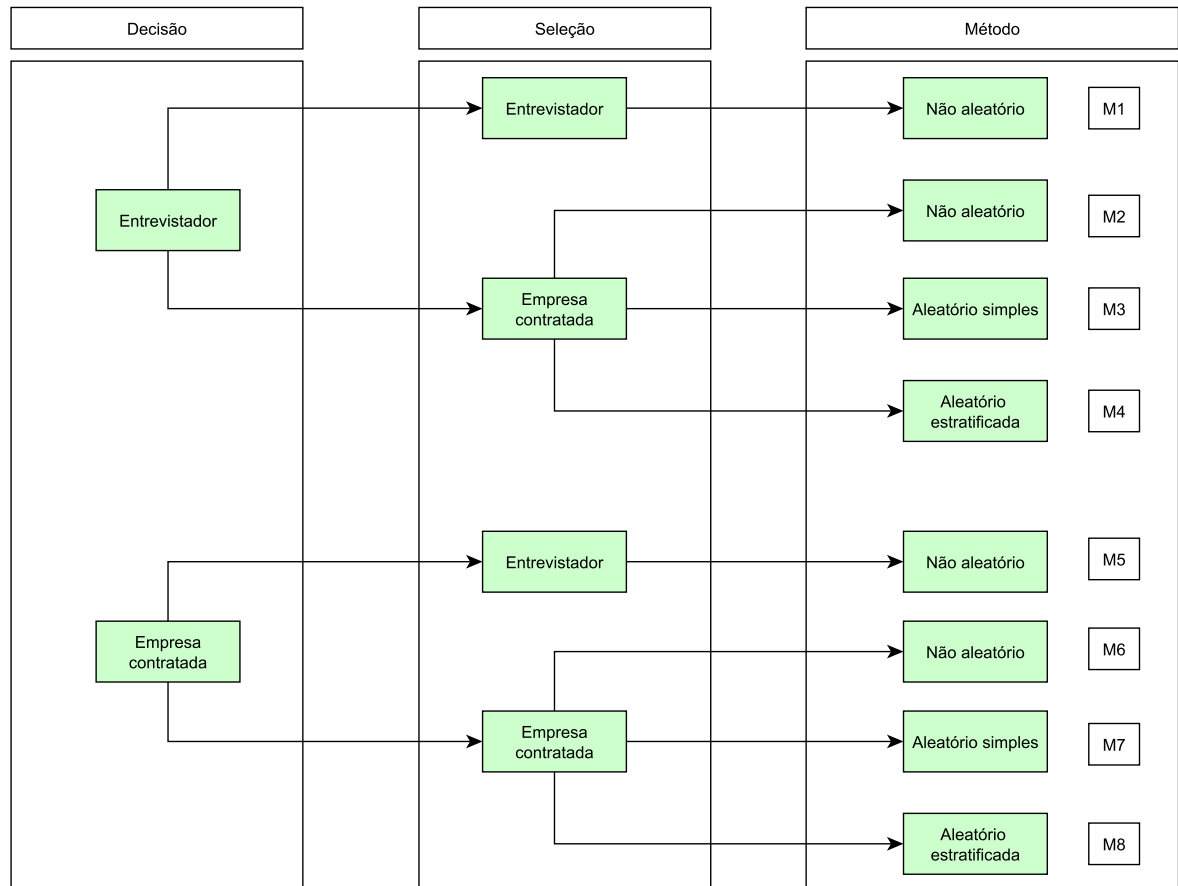
Segundo Vehovar (1999), a substituição de campo acontece quando uma unidade que não fornece resposta é trocada por outra. Esse procedimento é uma ferramenta específica para tratar a falta de resposta, funcionando como uma forma de imputação de dados, na qual uma unidade substituta é inserida no lugar da que não respondeu. De acordo com Baldissera *et al.* (2014), a substituição de campo é benéfica pois preserva o tamanho e a estrutura da amostra, assegurando o número previamente estabelecido de observações para cada segmento, conforme o planejamento inicial.

Lynn (2021) cita alguns dos motivos para adicionar uma amostra de substituição: Melhorar a representatividade da amostra e restaurar o tamanho da amostra; aumentar o tamanho geral da amostra; aumentar o tamanho da amostra de subgrupos específicos; e aumentar o tamanho da amostra e abordar a não cobertura no desenho inicial da amostra.

Segundo Lynn (2004), os métodos de substituição podem ser classificados em três dimensões: Se a decisão de substituição de uma unidade é tomada pelo entrevistador ou pela empresa responsável pela pesquisa (empresa responsável é qualquer processo que o entrevistador desconhece); se a seleção é feita pelo entrevistador ou pela empresa; como é feita a seleção: aleatório simples, aleatório estratificado, não aleatório. A Figura 8 a seguir mostra o esquemático

das três dimensões da classificação de acordo com o autor.

Figura 8 – Esquemático dos tipos de substituição em uma pesquisa.



Fonte: Adaptado de Lynn (2004)

Ainda de acordo com Deng *et al.* (2013), um aspecto crucial ao recrutar a amostra de substituição para o estudo é em relação ao cronograma do trabalho de campo e ao método de substituição. As decisões sobre o desenho da amostra podem impactar diretamente a execução do trabalho de campo. Por exemplo, se a nova amostra for integrada a listas já existentes dos entrevistadores, pode ser necessário aumentar o tempo concedido a eles. Dessa forma, o tempo de trabalho dos entrevistadores pode ser ampliado para que a coleta de dados da nova amostra se encaixe no cronograma já estabelecido.

Contudo, a substituição pode adicionar um viés extra se os procedimentos apropriados não forem seguidos no processo, fazendo com que as unidades que são mais facilmente rastreáveis e unidades cooperativas sejam preferencialmente entrevistadas, aumentando assim a não resposta e reduzindo a representatividade tanto na amostra original como entre os substitutos.

Apesar dos trabalhos citados anteriormente, Nishimura (2015) afirma que, frequen-

temente, utilizada na prática, a substituição é muitas vezes negligenciada na literatura sobre amostragem em pesquisas. O autor cita que poucos estudos apresentam um quadro formal para descrever e avaliar os métodos de substituição, e pouca investigação foi feita para melhorar as estimativas obtidas através da utilização da substituição como procedimento de ajustamento de não resposta.

Por mais escassa que seja a literatura sobre a interligação entre a não resposta ao item e a não resposta a unidade, alguns trabalhos mostram que parece existir uma relação entre a propensão de alguns entrevistados em não participar de uma pesquisa e a quantidade de não respostas ao item [Mason e Traugott (2010) e Stinchcombe e Jones (1981)]. Entrevistados que inicialmente recusaram, mas que posteriormente foram convertidos em sucesso, tiveram taxas de dados faltantes mais altas do que os entrevistados que cooperaram desde o início.

O presente trabalho busca preencher uma lacuna significativa na pesquisa nacional relacionada à reposição amostral, abordando uma temática de grande relevância e sensibilidade: os aspectos de gênero.

3.3 Exemplos de Protocolos de substituição

Para entender o processo de substituição em uma pesquisa foram elencados e apresentados alguns protocolos de substituições de estudos longitudinais, sendo definidos das mais diversas formas. Nesta seção serão descritos de forma sucinta esses protocolos.

3.3.1 Protocolo de substituição *Belgian Health Interview Survey (BHIS)* - Bélgica

Demarest *et al.* (2017) avaliam se o processo de reposição tem impacto nas taxas de resposta e nas estimativas de indicadores de saúde a partir da pesquisa *Belgian Health Interview Survey* (BHIS). O objetivo do trabalho é verificar se a reposição, conforme aplicada no BHIS, contribui para atingir a meta de tamanho de amostra líquida pré-definida, tanto em tamanho quanto em composição e avaliar o impacto da reposição no viés de não resposta para várias estimativas de saúde.

É preciso minuciar o protocolo de substituição da BHIS, pois é um protocolo novo, bem documentado e bem avaliado em todos os aspectos, ver [Demarest *et al.* (2022)]. Além disso, parte do presente capítulo se aproxima do que os autores fizeram, com o objetivo de mostrar o protocolo de reposição da PCSVDF-Mulher, analisar sua eficácia e efetividade em

campo.

Para entender o processo de substituição é preciso, em um primeiro momento, explicitar o processo de amostragem da pesquisa. Na primeira etapa foi realizada uma amostragem sistemática baseada em domicílios ordenados por setor censitário, tamanho do domicílio e faixa etária da pessoa de referência. Para cada domicílio selecionado, três domicílios consecutivos na lista ordenada também foram selecionados, criando um cluster de quatro domicílios no mesmo setor censitário, com o mesmo tamanho de domicílio e com a mesma faixa etária do responsável pelo domicílio.

Uma vez selecionados, é aplicado aos clusters de domicílios um processo de aleatoriedade vertical e horizontal. O processo vertical altera a ordem original deles (entre os clusters), enquanto o embaralhamento horizontal muda a sequência em que os domicílios dentro de um cluster específico são ativados, seja como domicílio inicial ou como primeiro, segundo ou terceiro substitutos.

No início de cada trimestre, todos os domicílios na primeira posição dos clusters são ativados. Nesse momento, são disparados convites para as reposições dos domicílios, convidando-os para participar da pesquisa. No mesmo momento, os entrevistadores recebem os endereços dos domicílios para que possam iniciar o contato com os domicílios. O protocolo da pesquisa estabelece que cada tentativa de contato tem de ser registrada (dia/hora, forma de contato e o resultado do contato) e para cada agregado familiar devem ser efetuadas pelo menos cinco tentativas de contato.

Se após cinco tentativas a entrevistadora não obtiver êxito, o entrevistador indica que o domicílio era “não contactável” e interrompe as tentativas de contato. No caso de um domicílio estar “não contactável”, recusar-se a participar ou a informação do endereço estar incorreta, é ativado o primeiro domicílio substituto e feito o mesmo processo de tentativas de contato do domicílio inicial.

Caso as tentativas de contato com o primeiro domicílio substituto não resultarem em participação, o segundo domicílio substituto será ativado. Esse processo será repetido até que um domicílio participasse da pesquisa ou até que os dois clusters sejam esgotados.

De acordo com Demarest *et al.* (2022), embora deva ser admitido que a aplicação da substituição de campo não elimina o viés - um viés residual permanece - o viés sem o uso da substituição de campo seria (substancialmente) maior. Este estudo é, ao nosso conhecimento, o primeiro que avalia a evolução da utilização da substituição de campos ao longo dos sucessivos

anos de levantamento em que esta técnica foi aplicada. Demarest *et al.* (2022) mostraram que a substituição de campo é indispensável para se obter o tamanho amostral solicitado e garante que a amostra líquida reflita a composição da amostra bruta de acordo com os critérios de seleção.

Além disso, há evidências de que o método funciona bem, pois as subamostras inicial e substituta não diferem significativamente. Além disso, nunca se pode ter certeza e, portanto, não se pode afirmar que o viés de não resposta foi totalmente removido, pela simples razão de que não se tem os dados que faltam. Evidências úteis são fornecidas comparando os respondentes da primeira onda com os substitutos.

3.3.2 *Protocolo de substituição Progressi delle Aziende Sanitarie per la Salute in Italia (PASSI) - Itália*

A *Progressi delle Aziende Sanitarie per la Salute in Italia* (PASSI) consiste em uma pesquisa com população-alvo de residentes entre 18 e 69 anos que possuam um número de telefone disponível, seja fixo ou celular. A unidade amostral é o indivíduo que é extraído a partir de uma amostra aleatória retirada da *Local Health Units* (LHUs), que mantem listas de matrículas dos pacientes atualizadas.

O entrevistador utiliza entrevistas telefônicas assistidas por computador ou questionários impressos. Os registros são classificados com uma identificação única. Caso a unidade amostral não possa ser entrevistada, ela será substituída por uma unidade de reserva, pré-selecionada aleatoriamente no mesmo estrato. A não resposta pode ser oriunda de três motivos, se a unidade amostral for: 1) contactada, elegível, mas recusou participar; 2) contactada, inelegível devido a incapacidade de comunicação (não falar italiano), estar hospitalizado ou institucionalizado, ou ter idade fora da faixa etária; 3) não contactada.

A substituição é realizada por entrevistadores sob a supervisão de Coordenadores da LHU. Vários substitutos para cada unidade inicialmente amostrada são sorteados em uma ordem predefinida (nível de substituição). Se uma unidade reserva não pode ser entrevistada, o substituto com melhor classificação é selecionado. O processo continua até que uma unidade seja entrevistada.

Baldissera *et al.* (2014) avalia o efeito da substituição na pesquisa, em que são comparados os resultados das estimativas ponderadas obtidas na amostra original e na subamostra dos substitutos. Os indicadores escolhidos são: características sociodemográficas, fatores de riscos comportamentais e alguns indicadores do estado de saúde da população.

Um primeiro resultado foi que, como o processo de substituição seguiu de perto a não resposta, as taxas de não resposta tiveram os mesmos valores que as taxas de substituição. O teste de Wald não mostrou diferenças significativas para qualquer variável. Discrepâncias entre as estimativas pontuais das duas subamostras eram pequenas em termos absolutos, especialmente para risco de fatores e indicadores de saúde. Se o processo de substituição produzisse importantes efeitos de distorção nas estimativas, seria esperado diferenças maiores nas regiões com maiores taxas de substituição.

3.3.3 *Protocolo de substituição Estudo Longitudinal da Saúde dos Idosos Brasileiros (ELSI-Brasil) - Brasil*

Lima-Costa *et al.* (2022) descrevem o Estudo Longitudinal da Saúde dos Idosos Brasileiros, primeiro estudo de envelhecimento de base populacional em grande escala no Brasil. O estudo tem o intuito de fornecer determinantes sociais e biológicos do envelhecimento e suas consequências para o indivíduo e a sociedade, sendo uma ferramenta importante para a promoção de políticas públicas da população mais idosa.

O estudo visa entrevistar residentes com 50 anos ou mais de idade de 70 municípios das cinco regiões do Brasil, sendo 10.000 o número de entrevistas planejadas. Além disso, dada a complexidade da amostra, foram utilizados pesos amostrais para levar em consideração as diferentes probabilidades de seleção e a não resposta.

Foram realizadas duas ondas de entrevistas: a primeira, entre 2015 e 2016, com 9412 participantes e a segunda entre 2019-2021, com 9949 participantes. Dos 9412 participantes da onda 1, 6172 foram entrevistados novamente, 2270 não participaram da segunda onda e 970 vieram a óbito, com isso é observado uma taxa de atrito de 34,42%. Isso leva a necessidade da reposição da amostra.

Lima-Costa *et al.* (2018) destacam algumas medidas preventivas para se reduzir o atrito na pesquisa. 1) durante a primeira entrevista, foram registrados os números de telefone e endereços das pessoas próximas ao entrevistado, incluindo familiares, amigos e vizinhos.; 2) entre as entrevistas foram realizadas ligações telefônicas para acompanhamento de mudanças de endereço e estado de saúde do entrevistado; 3) foi realizada vinculação ao Sistema Nacional de Mortalidade com o objetivo de rastrear os participantes que não foram localizados através dos procedimentos mencionados anteriormente, além de uma visita ao setor censitário (domicílio do participante e domicílios vizinhos, se necessário). Caso o participante não seja entrevistado na

onda posterior, a substituição do mesmo se dará por um morador, com idade acima de 50 anos, de um domicílio previamente selecionado no setor censitário correspondente.

No estudo foram encontrados que entrevistados com idade superior a 80 anos tem uma menor propensão de serem entrevistados novamente, tendo em vista que eles tinham uma maior probabilidade de falecer. Moradores do Nordeste e Centro-Oeste eram menos propensos a atritarem. Assim, como já previsto, foi realizada uma reposição da amostra para garantir a sua representatividade.

3.3.4 Protocolo de substituição AmericasBarometer 2016/17 - América do Norte, Central, do Sul e parte do Caribe

AmericasBarometer é um estudo periódico que tem o objetivo de estudar valores e comportamentos democráticos que abrangem todos os países independentes da América do Norte, Central e do Sul, além de um número significativo de países do Caribe.

A pesquisa extrai amostras nacionalmente representativas de adultos em idade eleitoral por meio de entrevistas presenciais. As amostras em cada país seguem um desenho probabilístico de múltiplos estágios e são estratificadas pelas principais regiões do país, tamanho do município e por áreas urbanas e rurais dentro dos municípios.

De certa forma, os entrevistadores precisam substituir algumas unidades amostrais preselecionadas, devido a problemas com segurança na área das unidades ou por conta da dificuldade em encontrar os entrevistados.

Kobilanski e Pizzolitto (2019) descrevem o protocolo de substituição do estudo, que tem em um primeiro momento uma solicitação escrita para a empresa responsável pela pesquisa justificando detalhadamente a substituição. Após a análise da solicitação, é feito um trabalho para identificar uma unidade amostral substituta, assegurando que a unidade se encontre numa área com uma dimensão populacional semelhante, um nível de urbanização semelhante e características socioeconômicas semelhantes às da seleção original. Além disso, as amostras de reposição devem estar dentro da mesma unidade primária de amostragem.

As taxas de substituição foram consideravelmente mais altas na Venezuela, Guatemala e México. As principais razões para a substituição de pontos de amostragem são preocupações de segurança, embora questões como infraestrutura inadequada e escassez de habitação também desempenhem um papel importante. Em conclusão, o protocolo inclui um estudo sobre o impacto das substituições na pesquisa, focando na taxa de vitimização do crime.

Os pedidos de substituição não introduziram viés que comprometa as comparações entre os dados criminais.

3.3.5 Protocolo de substituição *Russia Longitudinal Monitoring Survey (RLMS)*- Rússia

O *Russia Longitudinal Monitoring Survey (RLMS)* é composto por uma série de questionários representativos a nível nacional com o objetivo de monitorar os efeitos das reformas russas na saúde e no bem-estar econômico dos domicílios e dos indivíduos na Federação Russa.

A pesquisa engloba os anos de 1994-2022, contendo dados de todos os anos, tanto individuais (adultos e crianças) quanto para os dados domiciliares. Cada observação é um indivíduo ou um agregado familiar com todas as variáveis disponíveis para esse ano, portanto, tem-se dados longitudinais.

Heeringa (1997) afirma que a pesquisa enfrenta problemas de não resposta e atrito com as unidades amostrais preestabelecidas, entre as ondas 5 e 7. Mais uma vez, os efeitos do atrito e a mudança dos ocupantes das unidades residenciais resultam em uma redução entre as ondas na proporção de amostra principalmente na região de Moscou/São Petersburgo. Além disso, foi encontrada também uma redução na amostra na proporção de homens com idade entre 0-19 anos entre as ondas 5 e 7.

A reposição da amostra tem como objetivo evitar problemas mais sérios de desgaste amostral entre os domicílios que continuam a residir na amostra original de unidades habitacionais. O protocolo de substituição da pesquisa se dá a partir de uma nova amostra de residências que são pré selecionadas a partir das listas de enumeração originais da onda anterior, dando prioridade a domicílios dentro das unidades habitacionais ou próximos.

3.3.6 Protocolo de substituição *Healthy Chicago Survey HCS* - Estados Unidos

Lewis e McMichael (2023) apresentam uma avaliação de uma abordagem de substituição na pesquisa *Healthy Chicago Survey (HCS)* no ano de 2021, uma pesquisa autoadministrada por contato por correio, com início em 2014 como uma pesquisa telefônica anual, com adultos em Chicago. Os dados da pesquisa foram utilizados para apoiar a implementação do *Healthy Chicago* e para desenvolver diversas intervenções de saúde pública e moldar políticas para abordar as desigualdades em saúde.

As unidades de domicílio que não foram contactadas foram substituídas pelas unidades mais próximas com o mesmo tamanho. O substituto mais próximo foi identificado

deterministicamente usando a função GEODIST no SAS com base nas coordenadas de latitude e longitude dos endereços. Um substituto era sempre identificado dentro do mesmo bloco. A distância máxima foi de pouco menos de 0,3 milhas, o que equivale a cerca de três quarteirões.

Comparando as distribuições agregadas da amostra em que ocorreu atrito e dos seus substitutos, Lewis e McMichael (2023) observam certas diferenças no que diz respeito à idade, situação profissional, estado civil e posse de habitação, mas não há diferenças substanciais nos principais resultados de saúde.

Apresentados os protocolos de substituição ao redor do mundo faz-se necessário apresentar o protocolo de substituição utilizado na PCSVDF-Mulher

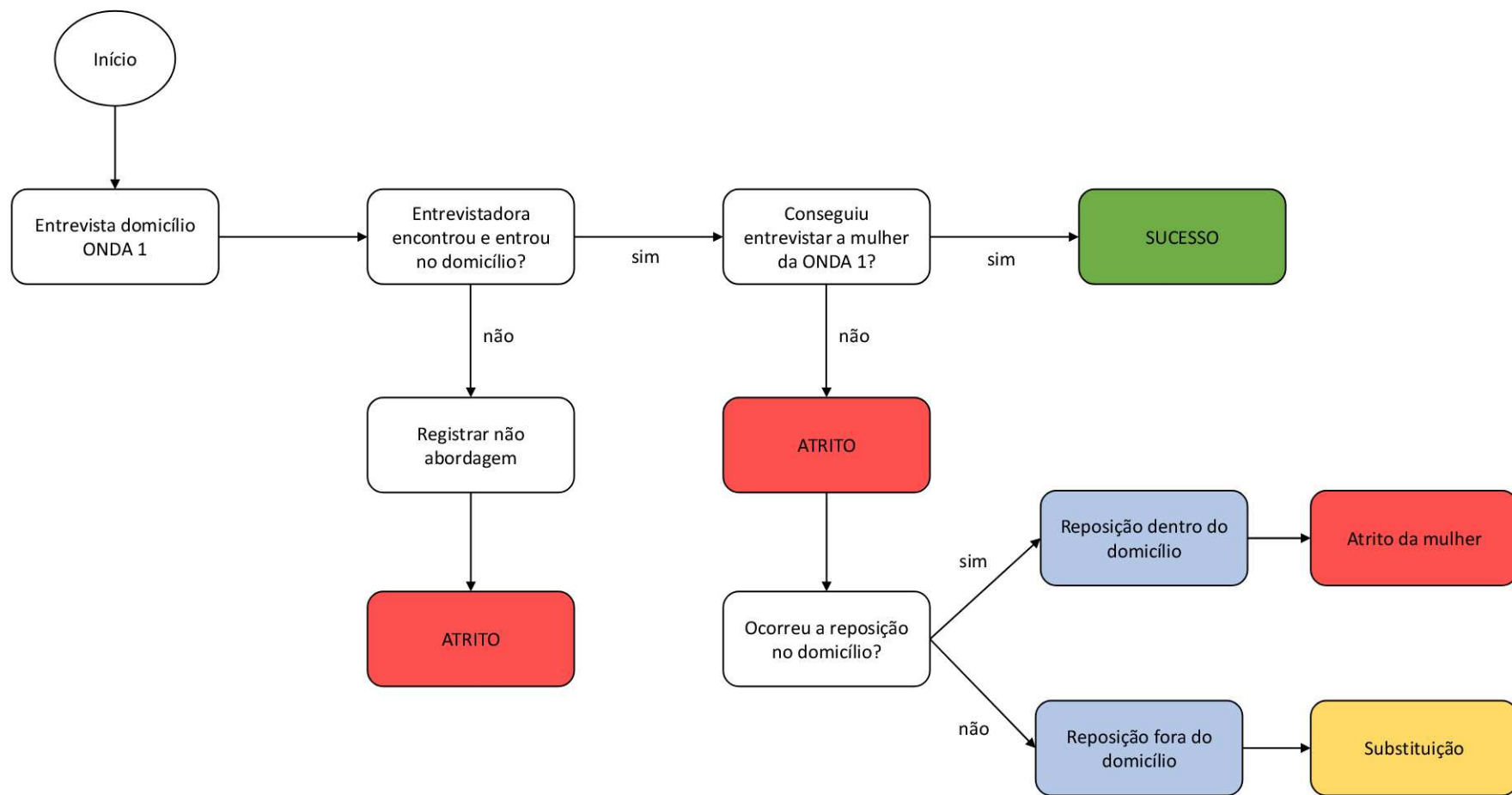
3.3.7 Protocolo de substituição da PCSVDF-Mulher

Antes de apresentar o protocolo de substituição da PCSVDF-Mulher é fundamental entender a dinâmica das situações possíveis no campo e as complexidades decorrentes dela. De maneira geral, segundo DATAINFO (2017), uma entrevista na onda 2 pode ser classificada como (a) um sucesso, (b) um atrito da mulher ou (c) um questionário de substituição, conforme as definições a seguir:

- a. Sucesso: Entende-se como sucesso os casos nos quais a mesma mulher da Onda 1 é entrevistada agora na Onda 2 e responde pelo menos até a seção de saúde geral e reprodutiva;
- b. Atrito da mulher: Entende-se como atrito da mulher os casos nos quais não foi possível entrevistar a mesma mulher da Onda 1 ou a entrevista inicia, mas é encerrada antes da mulher responder a seção de saúde geral e reprodutiva ou outra mulher do domicílio responde o questionário pelo menos até a seção de saúde reprodutiva;
- c. Substituição: Entende-se como substituição os novos domicílios que são acrescentados à base da pesquisa em substituição aos casos de atrito ocorridos durante a coleta de dados.

O protocolo de substituição da PCSVDF-Mulher não é feito em tempo real. As entrevistadoras irão informar para a suas coordenadoras de campo e estas relatam para a gestora de projetos da empresa contratada os casos de atrito ou não resposta. Com isso, a decisão de substituição é da empresa contratada, a seleção é feita pela entrevistadora, sendo dado prioridade para reposições ou dentro do domicílio ou no mesmo setor censitário. Portanto, o processo M5 da Figura 8 se assemelha ao tipo de substituição da PCSVDF-Mulher. O fluxograma, apresentado na Figura 9, indica uma visão geral de como uma entrevista pode tornar-se uma das categorias acima definidas.

Figura 9 – Fluxograma do processo de sucesso, atrito ou substituição PCSVDF-Mulher



Fonte: Adaptado de DATAINFO (2017)

Segundo DATAINFO (2017), a substituição para qualquer caso de atrito - reposição - se dará através da escolha, de forma aleatória, de um outro domicílio localizado no mesmo setor censitário daquele que houve o atrito.

Haverá a possibilidade de, após a “exaustão” das tentativas de se entrevistar a mulher entrevistada em 2016, a entrevistadora estará autorizada a repetir o protocolo para a próxima mulher elegível no domicílio. Não tendo sucesso com essa segunda mulher, segue o protocolo original, ou seja, não haverá substituição por outra mulher no domicílio, e o domicílio deverá ser substituído.

Não há consenso claro na literatura sobre o entendimento da substituição de campo. De acordo com Baldissera e Ferrante (2019) não há um referencial teórico geral que sustente ou contrarie o uso da substituição de campo. Na verdade, a decisão de optar pela substituição depende da sua utilidade prevista, levando em conta os objetivos e contextos específicos de cada pesquisa.

Não obstante, Baldissera e Ferrante (2019) oferecem uma referência útil para usar como uma espécie de lista de verificação para avaliar a adequação da aplicação de uma estratégia de substituição de campo num inquérito domiciliar. A Tabela 6, apresenta o check-list e vinculá-lo às características da PCSVDF-Mulher.

Tabela 6 – Características de uma pesquisa adequada para substituições de campo

Característica	PCSVDF-Mulher
Design	
Design extensivamente estratificado, abrangendo diversas áreas geográficas	Sim
Pesquisas ou vigilâncias estruturadas e recorrentes (que possam justificar investimentos adequados na organização, gestão e controle)	Sim
Seleção aleatória estratificada dos substitutos, da mesma lista utilizada para a amostra original	Sim
Entrevista por telefone	Não
Organização	
Protocolo do estudo definindo procedimentos detalhados para busca, contato e abordagem dos indivíduos selecionados, bem como os critérios para substituição de não respondentes	Sim
Seleção de operadores experientes e atividades estruturadas para treinamento/reciclagem	Sim
Coleta e análise de dados	
Supervisão e controle contínuos das atividades de pesquisa e adesão ao protocolo	Sim
Monitorização da qualidade dos dados e dos indicadores do processo, em particular as taxas de não resposta e de substituição	Sim
Avaliação do viés de substituição nos dados coletados	Sim

Fonte: Elaborado pelo autor adaptado de Baldissera e Ferrante (2019)

Portanto, a PCSVDF-Mulher possui quase toda a totalidade de características, que Baldissera e Ferrante (2019) citam para avaliar a adequação da implementação de uma estratégia

de reposição amostral em um inquérito domiciliar.

Como a coleta de dados ocorre em um país em desenvolvimento, apesar de existirem algumas escolhas metodológicas ou decisões mais rigorosas e melhores (exemplo o protocolo *Belgian Health Interview Survey* (BHIS) apresentado anteriormente), vale salientar que o custo que viabiliza o projeto é um critério muito importante, tendo em vista a limitação financeira de tais países.

Apresentado o protocolo de substituição da PCSVDF-Mulher são necessárias métricas para avaliá-lo, especialmente em relação a sua eficácia no que se refere a recomposição do tamanho da amostra inicial e mantendo a representatividade estatística da população alvo sob análise, tanto do ponto de vista espacial quanto não espacial. Para isso, será desenvolvida uma metodologia que analisa quantitativamente os aspectos do protocolo de substituição da PCSVDF-Mulher. Antes, faremos uma análise descritiva preliminar do processo de reposição amostral.

3.4 Análise preliminar do processo de reposição amostral

Inicialmente, foi analisado se o tamanho original da amostra foi preservado, ao longo de diferentes níveis de agregação espacial (pesquisa, cidades, bairros e setores censitários). Segundo, foi desenvolvido e aplicado uma metodologia para inspecionar qual critério de reposição espacial foi realmente aplicado pela empresa. Por fim, foi verificado se o processo de substituição utilizado pela PCSVDF-Mulher manteve a representatividade populacional, comparando-a com a PNAD.

Para o primeiro bloco, como o principal objetivo é verificar se a reposição foi igual ao atrito, será criado um indicador (taxa de reposição amostral) como a razão do número de questionários válidos em 2017 e em 2016, ou seja, $N(g) = \#N17/\#N16$.

3.4.1 Reposição do tamanho original: Desagregação por cidades

Para nível de pesquisa e de cidades, o valor de $N(g)$ é apresentado na Tabela 7. Como seria extensivo mostrar todos os bairros e setores censitários em forma de tabela, para as duas desagregações geográficas a representação é apresentada na Figura 10. De acordo com Andre e Carvalho (2023), na PCSVDF-Mulher nas ondas 1 e 2 foram consideradas entrevistas válidas um total de 9618 e 9845, respectivamente e distribuídas da forma que segue na Tabela 7.

Tabela 7 – Taxa de reposição por município, em 2016 e 2017

Nível de agregação	#N16	#N17	N(g)
PCSVDF-Mulher	9618	9845	1,0236
Aracaju/SE	986	1001	1,0152
Fortaleza/CE	1172	1233	1,0520
João Pessoa/PB	1072	995	0,9281
Maceio/AL	943	1073	1,1378
Natal/RN	1052	1091	1,0370
Recife/PE	1245	1212	0,9734
Salvador/BA	1104	1187	1,0751
São Luís/MA	1082	1097	1,0138
Teresina/PI	962	956	0,9937

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

Notas: #N16: Domicílios em 2016, #N17: Domicílios em 2017, $N(g) = \#N17 / \#N16$

A Tabela 7 fornece o quantitativo de domicílios válidos no ano de 2016 e 2017 de acordo com os recortes, o primeiro da pesquisa completa e o segundo com os recortes dos municípios. Espera-se que o quantitativo de domicílios em 2017 seja igual ou maior que o número de domicílios em 2016, tendo em vista que houve atrito amostral e tem-se a necessidade de se fazer as substituições dessas unidades. Com isso, os valores de $N(g)$ devem ser iguais ou maiores a 1,00.

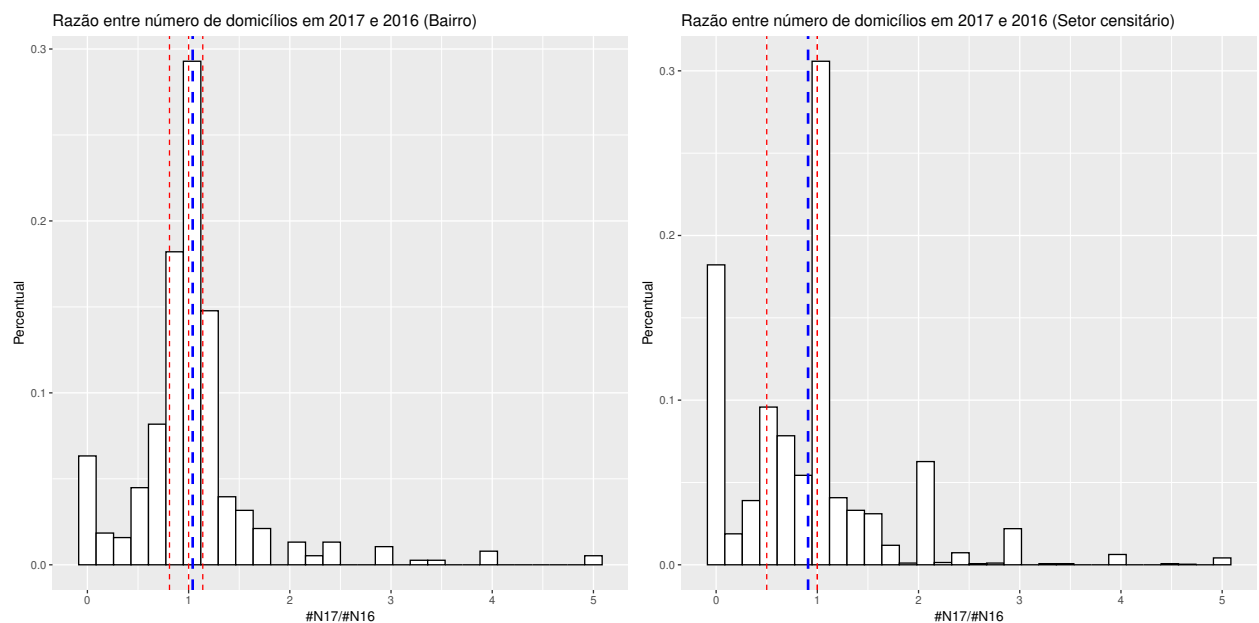
Como se pode ver, em 2017 foram entrevistados 2,36% mulheres a mais que em 2016. Com isso, a nível de pesquisa, o valor de $N(g) = 1,0236$ corrobora com o que foi preconizado no protocolo de substituição, que é a substituição do domicílio atritado. Entretanto, isso não ocorre em sua totalidade quando se faz a primeira desagregação territorial, a nível de município, pois os municípios de João Pessoa/PB, Recife/PE e Teresina/PI (apesar de $N(g)$ ser bem próximo de 1,00) tiveram um quantitativo de domicílios entrevistados em 2017 inferior ao entrevistado em 2016. O município de João Pessoa-PB teve cerca de 8% de entrevistadas a menos do que em 2016, Recife-PE cerca de 3% e Teresina-PI com somente 6 mulheres a menos que 2016, o que representa uma diminuição de aproximadamente 0,62%. Por outro lado, as cidades de Maceió/AL e Salvador/BA, tiveram mais entrevistas em 2017 com relação a 2016, sendo superior em 13,78% e 7,51%, respectivamente. Essas cidades se destacaram por apresentar os maiores valores de $N(g)$.

A próxima análise apresenta as desagregações da PCSVDF-Mulher em nível de bairro e setor censitário, com o objetivo de investigar o comportamento do valor de $N(g)$ em ambas as divisões territoriais.

3.4.2 Reposição do tamanho original: Desagregação por bairros e setores censitários

A Figura 10 mostra a distribuição do valor de $N(g)$ de acordo com os quantitativos de domicílios entrevistados nos bairros e nos setores censitários em 2016 e em 2017. O gráfico (a) indica que a PCSVDF-Mulher conseguiu manter as substituições dentro do nível de bairro, tendo em vista que o histograma de $N(g)$ se concentra mais próximo ao valor 1,00, sendo a média de $N(g)$ igual a 1,0659 (linha azul). Os resultados dos quartis e do intervalo interquartílico (IQR) indicados pelas linhas vermelhas, mostram que o primeiro quartil ($Q1 = 0,75$) indica que 25% dos valores são menores ou iguais a 0,75, enquanto o terceiro quartil ($Q3 = 1,1$) mostra que 75% dos valores são menores ou iguais a 1,1. O intervalo interquartílico (IQR), calculado como a diferença entre $Q3$ e $Q1$ ($1,1 - 0,75 = 0,35$), representa a amplitude dos 50% centrais dos dados, ou seja, a região onde está a maior concentração de valores, minimizando o impacto de extremos.

Figura 10 – Valor de $N(g)$ de acordo com a desagregação geográfica - Bairro e Setores censitários.



Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

O gráfico (b) fornece a distribuição de $N(g)$ de acordo com o setor censitário. Diferentemente do comportamento de $N(g)$ a nível de bairro, nos setores censitários a distribuição se dá de uma forma mais dispersa, além de uma frequência considerável de valores de $N(g) = 0$, que representa os setores que foram visitados em 2016 mas não tiveram entrevistas válidas em 2017. A média de $N(g)$ - linha azul - foi igual a 0,8649, e as linhas indicadas em vermelho indicam os intervalos dos quartis. Para o nível de setor censitário o 1º quartil igual a 0,5000 e o 2º e 3º

quartis iguais a 1,0000, com isso mais de 50% dos setores conseguiram repor as entrevistadas.

A próxima seção realizará uma análise comparativa entre a amostra da Pesquisa Nacional de Domicílios (PNAD) dos anos de 2016 e 2017 e a amostra completa da PCSVDF-Mulher, incluindo os dados atritados de 2016 e a reposição de 2017. O objetivo é avaliar como a amostra coletada se comporta em relação à amostra nacional da PNAD, proporcionando uma visão mais aprofundada das semelhanças e diferenças entre os conjuntos de dados.

3.5 Metodologia I: Representatividade não espacial

A presente seção tem como objetivo verificar como está distribuída as variáveis de raça, idade e escolaridade da PCSVDF-Mulher fazendo um comparativo com a amostra da Pesquisa Nacional por Amostra de Domicílios (PNAD). A PNAD é uma pesquisa realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Seu objetivo principal é coletar informações sobre as condições de vida da população. A PNAD fornece uma visão abrangente das características da população brasileira, permitindo análises sobre emprego, rendimento, educação e outras questões sociais. Com isso, o intuito é utilizar a base da PNAD como um direcionamento para verificar a adequação dos dados da base da PCSVDF-Mulher.

Para atingir tal objetivo, será utilizado o teste qui-quadrado de proporção para investigar se duas bases de dados apresentam distribuições semelhantes. O teste permite avaliar se as proporções observadas em diferentes categorias são iguais entre as duas amostras, ajudando a identificar possíveis diferenças significativas. Através dessa análise, são explorados padrões e tendências, fornecendo insights valiosos sobre as características de cada base de dados. Ao aplicar esse teste, é possível verificar se as proporções observadas em diferentes categorias são iguais entre as bases de dados analisadas.

O processo envolve a formulação de hipóteses: a hipótese nula (H_0) afirma que não há diferença significativa nas proporções entre as duas bases, enquanto a hipótese alternativa (H_1) sugere que existe uma diferença. Ao calcular o valor do qui-quadrado são comparadas a frequência observada nas categorias com a frequência esperada sob a hipótese nula.

As variáveis raça, idade e escolaridade serão utilizadas no comparativo entre as bases PNAD e PCSVDF-Mulher. Tendo em vista que os dados da PNAD estão disponibilizados em forma categórica, o teste a ser utilizado para comparar as distribuições será o teste de homogeneidade (teste qui quadrado) das variáveis, sendo o teste feito entre as bases:

– PNAD-2016 versus PCSVDF-Mulher 2016

- PNAD-2016 versus PCSVDF-Mulher^{ATT}
- PNAD-2017 versus PCSVDF-Mulher 2017
- PNAD-2017 versus PCSVDF-Mulher^{REP}

Após a análise comparativa das amostras da PNAD e da PCSVDF-Mulher, será realizada a avaliação da eficácia da reposição. A eficácia da substituição da PCSVDF-Mulher será determinada pela verificação da representatividade estatística da sua amostra. Para isso, algumas variáveis, tanto socioeconômicas quanto relacionadas a episódios de violência, serão comparadas entre as mulheres que saíram da pesquisa e aquelas que foram substituídas.

Serão realizados testes estatísticos para analisar as variáveis do estudo. Para as variáveis numéricas, como número de filhos e idade, serão aplicados o teste t de Student para comparar médias e o teste de Levene para verificar a homogeneidade das variâncias. No caso das variáveis categóricas — raça, escolaridade e religião — será utilizado o teste qui-quadrado de homogeneidade, com o objetivo de determinar se essas variáveis se distribuem de maneira semelhante entre as populações de atrito e reposição. Por fim, para as variáveis dicotômicas, o teste de proporção para verificar se os percentuais das variáveis de interesse são iguais entre os grupos analisados.

Tabela 8 – Resumo dos testes estatísticos utilizados

Bases de dados	Testes
PNAD-2016 versus PCSVDF-Mulher 2016	Teste qui-quadrado
PNAD-2016 versus PCSVDF-Mulher ^{ATT}	Teste qui-quadrado
PNAD-2017 versus PCSVDF-Mulher 2017	Teste qui-quadrado
PNAD-2017 versus PCSVDF-Mulher ^{REP}	Teste qui-quadrado
PCSVDF-Mulher ^{ATT} versus PCSVDF-Mulher ^{REP}	Teste qui-quadrado, Teste t-student, Teste de LAVENE, Teste de proporções

Fonte: Elaborado pelo autor.

3.5.1 PCSVDF-Mulher 2016 versus PNAD-2016

As análises das Tabelas 9 e 10 consideram a comparação das distribuições das três variáveis—raça, idade e escolaridade—entre as duas bases de dados: PNAD e PCSVDF-Mulher. Para verificar se as distribuições percentuais dessas variáveis são idênticas em ambas as bases, utilizou-se o teste qui-quadrado de homogeneidade.

No que diz respeito à raça, a análise revela que, embora as mulheres pardas tenham uma maior frequência na PCSVDF-Mulher e na PNAD-2016, conforme mostrado na Tabela 9, há uma proporção relativamente maior de mulheres pretas na PCSVDF-Mulher. Esse fenômeno

pode ter influenciado o teste, que rejeitou a hipótese de que as distribuições de raça são iguais. Essa tendência é observada em todos os municípios pesquisados, conforme indicado na Tabela 10, exceto em Salvador-BA, onde o percentual de mulheres pretas superou o de mulheres pardas.

A análise etária revela que, na faixa de 15 a 17 anos, as idades são semelhantes entre a PCSVDF-Mulher e a PNAD-2016. No entanto, ao observar os grupos etários de 18 a 39 anos, a PCSVDF-Mulher apresenta um percentual maior, enquanto na faixa de 40 a 59 anos, a PNAD-2016 tem uma proporção superior. Isso sugere que a população da PCSVDF-Mulher é mais jovem. Esse padrão se mantém em todos os municípios analisados, exceto em Teresina-PI, onde os perfis etários nas duas pesquisas mostram resultados quase idênticos, sendo a única situação em que a hipótese de homogeneidade nas distribuições não foi rejeitada, conforme demonstrado nas Tabelas 9 e 10.

Por fim, ao comparar a escolaridade da base completa da PCSVDF-Mulher 2016 com a da PNAD-2016, observa-se uma diferença significativa nas faixas de escolaridade de Ensino Fundamental (EF) e Ensino Médio (EM). Na PCSVDF-Mulher 2016, a maioria das participantes possui Ensino Médio (56,10%), enquanto na PNAD-2016 esse percentual é de 39,58%.

O perfil de escolaridade em cada município, revela que a frequência de mulheres com Ensino Médio é superior na maioria dos cenários, exceto em Maceió-AL, onde os dados da PNAD-2016 indicam que a maior proporção é de mulheres com Ensino Fundamental.

Tabela 9 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD²⁰¹⁶, PCSVDF-Mulher²⁰¹⁶)

	PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶	
Branca	1940(0,2924)	2385(0,2427)	<0,001 ¹
Parda	3800(0,5726)	5207(0,5299)	
Preta	896(0,135)	2234(0,2274)	
15 a 17 anos	277(0,0615)	578(0,0588)	<0,001 ¹
18 a 24 anos	723(0,1605)	2131(0,2169)	
25 a 29 anos	508(0,1127)	1528(0,1555)	
30 a 39 anos	1167(0,259)	2878(0,2929)	
40 a 49 anos	1831(0,4063)	2711(0,2759)	
SE	254(0,045)	101(0,0103)	<0,001 ¹
EF	1747(0,3098)	2234(0,2274)	
EM	2232(0,3958)	5512(0,561)	
ES	1406(0,2493)	1979(0,2014)	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Tabela 10 – Frequência absoluta e relativa, PNAD²⁰¹⁶ E PCSVDF-Mulher²⁰¹⁶ de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.

Aracaju-SE				Fortaleza-CE				João Pessoa-PB			
	PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶			PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶			PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶	
15 a 17 anos	13(0,0551)	50(0,0514)	<0,001 ¹	15 a 17 anos	60(0,0647)	113(0,0957)	<0,001 ¹	15 a 17 anos	17(0,0616)	56(0,0518)	<0,001 ¹
18 a 24 anos	34(0,1441)	225(0,2312)		18 a 24 anos	150(0,1616)	259(0,2193)		18 a 24 anos	41(0,1486)	247(0,2285)	
25 a 29 anos	23(0,0975)	144(0,148)		25 a 29 anos	115(0,1239)	185(0,1566)		25 a 29 anos	33(0,1196)	167(0,1545)	
30 a 39 anos	68(0,2881)	290(0,298)		30 a 39 anos	242(0,2608)	326(0,276)		30 a 39 anos	68(0,2464)	322(0,2979)	
40 a 59 anos	98(0,4153)	264(0,2713)		40 a 59 anos	361(0,389)	298(0,2523)		40 a 59 anos	117(0,4239)	289(0,2673)	
Branca	88(0,2659)	206(0,2117)	<0,001 ¹	Branca	484(0,348)	268(0,2269)	<0,001 ¹	Branca	172(0,3963)	353(0,3265)	<0,001 ¹
Parda	214(0,6465)	548(0,5632)		Parda	832(0,5981)	777(0,6579)		Parda	239(0,5507)	567(0,5245)	
Preta	29(0,0876)	219(0,2251)		Preta	75(0,0539)	136(0,1152)		Preta	23(0,053)	161(0,1489)	
SE	14(0,0483)	11(0,0113)	<0,001 ¹	SE	60(0,0516)	13(0,011)	<0,001 ¹	SE	16(0,0448)	8(0,0074)	<0,001 ¹
EF	92(0,3172)	267(0,2744)		EF	408(0,3508)	327(0,2769)		EF	114(0,3193)	268(0,2479)	
EM	104(0,3586)	498(0,5118)		EM	454(0,3904)	684(0,5792)		EM	125(0,3501)	519(0,4801)	
ES	80(0,2759)	197(0,2025)		ES	241(0,2072)	157(0,1329)		ES	102(0,2857)	286(0,2646)	
Maceió-AL				Natal-RN				Recife-PE			
	PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶			PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶			PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶	
15 a 17 anos	23(0,0637)	44(0,0548)	<0,001 ¹	15 a 17 anos	21(0,0675)	68(0,0647)	<0,001 ¹	15 a 17 anos	33(0,056)	61(0,0483)	<0,001 ¹
18 a 24 anos	64(0,1773)	21(0,0262)		18 a 24 anos	47(0,1511)	244(0,2322)		18 a 24 anos	97(0,1647)	265(0,2097)	
25 a 29 anos	37(0,1025)	191(0,2379)		25 a 29 anos	32(0,1029)	136(0,1294)		25 a 29 anos	60(0,1019)	207(0,1638)	
30 a 39 anos	89(0,2465)	284(0,3537)		30 a 39 anos	74(0,2379)	304(0,2892)		30 a 39 anos	147(0,2496)	378(0,2991)	
40 a 59 anos	148(0,41)	263(0,3275)		40 a 59 anos	137(0,4405)	299(0,2845)		40 a 59 anos	252(0,4278)	353(0,2793)	
Branca	142(0,2606)	284(0,2857)	<0,001 ¹	Branca	200(0,4237)	381(0,3625)	<0,001 ¹	Branca	371(0,425)	397(0,3141)	<0,001 ¹
Parda	374(0,6862)	511(0,5141)		Parda	248(0,5254)	539(0,5128)		Parda	435(0,4983)	619(0,4897)	
Preta	29(0,0532)	199(0,2002)		Preta	24(0,0508)	131(0,1246)		Preta	67(0,0767)	248(0,1962)	
SE	31(0,069)	28(0,0282)	<0,001 ¹	SE	14(0,0347)	5(0,0048)	<0,001 ¹	SE	43(0,0578)	13(0,0103)	<0,001 ¹
EF	177(0,3942)	255(0,2565)		EF	132(0,3267)	225(0,2141)		EF	209(0,2809)	315(0,2492)	
EM	159(0,3541)	496(0,499)		EM	149(0,3688)	601(0,5718)		EM	267(0,3589)	720(0,5696)	
ES	82(0,1826)	215(0,2163)		ES	109(0,2698)	220(0,2093)		ES	225(0,3024)	216(0,1709)	
Salvador-BA				São Luís-MA				Teresina-PI			
	PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶			PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶			PNAD ²⁰¹⁶	PCSVDF-Mulher ²⁰¹⁶	
15 a 17 anos	56(0,0517)	78(0,0659)	<0,001 ¹	15 a 17 anos	35(0,0864)	71(0,0639)	<0,001 ¹	15 a 17 anos	16(0,0539)	37(0,0375)	0,4891
18 a 24 anos	166(0,1531)	253(0,2137)		18 a 24 anos	75(0,1852)	238(0,2142)		18 a 24 anos	46(0,1549)	188(0,1905)	
25 a 29 anos	113(0,1042)	167(0,141)		25 a 29 anos	46(0,1136)	178(0,1602)		25 a 29 anos	44(0,1481)	153(0,155)	
30 a 39 anos	290(0,2675)	365(0,3083)		30 a 39 anos	103(0,2543)	338(0,3042)		30 a 39 anos	83(0,2795)	271(0,2746)	
40 a 59 anos	459(0,4234)	321(0,2711)		40 a 59 anos	146(0,3605)	286(0,2574)		40 a 59 anos	108(0,3636)	338(0,3425)	
Branca	235(0,1503)	73(0,0617)	<0,001 ¹	Branca	141(0,2456)	213(0,1917)	<0,001 ¹	Branca	104(0,2353)	210(0,2128)	<0,001 ¹
Parda	792(0,5064)	461(0,3894)		Parda	368(0,6411)	602(0,5419)		Parda	295(0,6674)	583(0,5907)	
Preta	537(0,3434)	650(0,549)		Preta	65(0,1132)	296(0,2664)		Preta	43(0,0973)	194(0,1966)	
SE	36(0,026)	6(0,0051)	<0,001 ¹	SE	18(0,0383)	1(0,0009)	<0,001 ¹	SE	17(0,0468)	16(0,0162)	<0,001 ¹
EF	365(0,2641)	252(0,2128)		EF	125(0,266)	127(0,1143)		EF	121(0,3333)	198(0,2006)	
EM	591(0,4276)	737(0,6225)		EM	234(0,4979)	710(0,6391)		EM	146(0,4022)	547(0,5542)	
ES	390(0,2822)	189(0,1596)		ES	93(0,1979)	273(0,2457)		ES	79(0,2176)	226(0,229)	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

3.5.2 PCSVDF-Mulher^{ATT} versus PNAD-2016

Em relação à raça, embora a maior proporção de mulheres na PCSVDF-Mulher^{atrato} e na PNAD-2016 seja de mulheres pardas, como mostrado na Tabela 11, há uma proporção relativamente maior de mulheres pretas na PCSVDF-Mulher^{atrato}, seguindo o mesmo resultado encontrado na comparação entre a base completa da PCSVDF-Mulher e da PNAD-2016. .

A Tabela 12 mostra que a situação não se repete quando é realizado o recorte dos municípios. Essa situação é particularmente evidente em Salvador-BA, onde, ao contrário do que ocorre na maioria dos municípios analisados, a proporção de mulheres pretas é superior à de mulheres pardas na pesquisa. Esse fenômeno pode refletir características socioeconômicas e culturais específicas da região, que influenciam a composição racial da população feminina.

A idade é semelhante na faixa etária de 15 a 17 anos, entre PCSVDF-Mulher^{atrato} e a PNAD-2016, entretanto nos recortes etários que se estendem da idade de 18 a 39 anos percebe-se um percentual superior na PCSVDF-Mulher, e na faixa etária de 40 a 59 anos um percentual maior da PNAD-2016.

Quando comparado a amostra da PNAD com a população de mulheres atritadas, a escolaridade de Ensino Médio é bem semelhante nas duas populações. A maior diferença ocorre entre as mulheres com Ensino Superior, sendo 24,93% e 36,84%, na PNAD e na PCSVDF-Mulher^{ATT}, Tabela 11. A Tabela 12 mostra o perfil de escolaridade em cada município, percebe-se que em todos ocorre diferença na distribuição entre os dados da PCSVDF-Mulher^{ATT} e PNAD-2016.

Tabela 11 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD²⁰¹⁶, PCSVDF-Mulher^{ATT})

	PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}	
Branca	1940(0,2924)	1366 (0,2576)	<0,001 ¹
Parda	3800(0,5726)	2781 (0,5245)	
Preta	896(0,135)	1126 (0,2124)	
15 a 17 anos	277(0,0615)	297(0,0562)	<0,001 ¹
18 a 24 anos	723(0,1605)	1258(0,2381)	
25 a 29 anos	508(0,1127)	884(0,1672)	
30 a 39 anos	1167(0,259)	1502(0,2843)	
40 a 49 anos	1831(0,4063)	1343(0,2542)	
SE	254(0,045)	54 (0,0102)	<0,001 ¹
EF	1747(0,3098)	1128 (0,2127)	
EM	2232(0,3958)	2167 (0,4087)	
ES	1406(0,2493)	1953 (0,3684)	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Tabela 12 – Frequência absoluta e relativa, PNAD²⁰¹⁶ e PCSVDF-Mulher^{atr} de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.

Aracaju-SE				Fortaleza-CE				João Pessoa-PB			
	PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}			PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}			PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}	
15 a 17 anos	13(0,0551)	29(0,0508)	<0,001 ¹	15 a 17 anos	60(0,0647)	47(0,0948)	<0,001 ¹	15 a 17 anos	17(0,0616)	29(0,0454)	<0,001 ¹
18 a 24 anos	34(0,1441)	144(0,2522)		18 a 24 anos	150(0,1616)	125(0,252)		18 a 24 anos	41(0,1486)	148(0,2316)	
25 a 29 anos	23(0,0975)	93(0,1629)		25 a 29 anos	115(0,1239)	84(0,1694)		25 a 29 anos	33(0,1196)	124(0,1941)	
30 a 39 anos	68(0,2881)	169(0,296)		30 a 39 anos	242(0,2608)	123(0,248)		30 a 39 anos	68(0,2464)	191(0,2989)	
40 a 59 anos	98(0,4153)	136(0,2382)		40 a 59 anos	361(0,389)	117(0,2359)		40 a 59 anos	117(0,4239)	147(0,23)	
Branca	88(0,2659)	133 (0,2239)	<0,001 ¹	Branca	484(0,348)	124 (0,2460)	<0,001 ¹	Branca	172(0,3963)	225 (0,3348)	<0,001 ¹
Parda	214(0,6465)	333 (0,5606)		Parda	832(0,5981)	332 (0,6587)		Parda	239(0,5507)	349 (0,5193)	
Preta	29(0,0876)	122 (0,2054)		Preta	75(0,0539)	46 (0,0913)		Preta	23(0,053)	94 (0,1399)	
SE	14(0,0483)	07 (0,0118)	<0,001 ¹	SE	60(0,0516)	06 (0,0119)	<0,001 ¹	SE	16(0,0448)	05 (0,0074)	<0,001 ¹
EF	92(0,3172)	158 (0,2660)		EF	408(0,3508)	136 (0,2698)		EF	114(0,3193)	143 (0,2128)	
EM	104(0,3586)	304 (0,5118)		EM	454(0,3904)	292 (0,5794)		EM	125(0,3501)	307 (0,4568)	
ES	80(0,2759)	125 (0,2104)		ES	241(0,2072)	70 (0,1389)		ES	102(0,2857)	217 (0,3229)	
Maceió-AL				Natal-RN				Recife-PE			
	PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}			PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}			PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}	
15 a 17 anos	23(0,0637)	23(0,0406)	<0,001 ¹	15 a 17 anos	21(0,0675)	39(0,0586)	<0,001 ¹	15 a 17 anos	33(0,056)	43(0,0705)	<0,001 ¹
18 a 24 anos	64(0,1773)	164(0,2892)		18 a 24 anos	47(0,1511)	162(0,2436)		18 a 24 anos	97(0,1647)	146(0,2393)	
25 a 29 anos	37(0,1025)	101(0,1781)		25 a 29 anos	32(0,1029)	92(0,1383)		25 a 29 anos	60(0,1019)	92(0,1508)	
30 a 39 anos	89(0,2465)	152(0,2681)		30 a 39 anos	74(0,2379)	177(0,2662)		30 a 39 anos	147(0,2496)	177(0,2902)	
40 a 59 anos	148(0,41)	127(0,224)		40 a 59 anos	137(0,4405)	195(0,2932)		40 a 59 anos	252(0,4278)	152(0,2492)	
Branca	142(0,2606)	151 (0,2990)	<0,001 ¹	Branca	200(0,4237)	246 (0,3576)	<0,001 ¹	Branca	371(0,425)	227 (0,3586)	<0,001 ¹
Parda	374(0,6862)	246 (0,4871)		Parda	248(0,5254)	349 (0,5073)		Parda	435(0,4983)	288 (0,4550)	
Preta	29(0,0532)	105 (0,2079)		Preta	24(0,0508)	90 (0,1308)		Preta	67(0,0767)	114 (0,1801)	
SE	31(0,069)	13 (0,0257)	<0,001 ¹	SE	14(0,0347)	04 (0,0058)	<0,001 ¹	SE	43(0,0578)	08 (0,0126)	<0,001 ¹
EF	177(0,3942)	131 (0,2594)		EF	132(0,3267)	141 (0,2049)		EF	209(0,2809)	146 (0,2306)	
EM	159(0,3541)	252 (0,4990)		EM	149(0,3688)	400 (0,5814)		EM	267(0,3589)	362 (0,5719)	
ES	82(0,1826)	109 (0,2158)		ES	109(0,2698)	143 (0,2078)		ES	225(0,3024)	117 (0,1848)	
Salvador-BA				São Luís-MA				Teresina-PI			
	PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}			PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}			PNAD ²⁰¹⁶	PCSVDF-Mulher ^{atr}	
15 a 17 anos	56(0,0517)	33(0,0535)	<0,001 ¹	15 a 17 anos	35(0,0864)	34(0,0556)	<0,001 ¹	15 a 17 anos	16(0,0539)	20(0,0366)	0,2555
18 a 24 anos	166(0,1531)	150(0,2431)		18 a 24 anos	75(0,1852)	147(0,2406)		18 a 24 anos	46(0,1549)	110(0,2015)	
25 a 29 anos	113(0,1042)	96(0,1556)		25 a 29 anos	46(0,1136)	109(0,1784)		25 a 29 anos	44(0,1481)	93(0,1703)	
30 a 39 anos	290(0,2675)	183(0,2966)		30 a 39 anos	103(0,2543)	181(0,2962)		30 a 39 anos	83(0,2795)	149(0,2729)	
40 a 59 anos	459(0,4234)	155(0,2512)		40 a 59 anos	146(0,3605)	140(0,2291)		40 a 59 anos	108(0,3636)	174(0,3187)	
Branca	235(0,1503)	34 (0,0608)	<0,001 ¹	Branca	141(0,2456)	118 (0,1916)	<0,001 ¹	Branca	104(0,2353)	108 (0,2034)	<0,001 ¹
Parda	792(0,5064)	232 (0,4150)		Parda	368(0,6411)	327 (0,5308)		Parda	295(0,6674)	325 (0,6121)	
Preta	537(0,3434)	290 (0,5188)		Preta	65(0,1132)	168 (0,2727))		Preta	43(0,0973)	97 (0,1827)	
SE	36(0,026)	03 (0,0054)	<0,001 ¹	SE	18(0,0383)	00 (0,0000)	<0,001 ¹	SE	17(0,0468)	08 (0,0151)	<0,001 ¹
EF	365(0,2641)	113 (0,2021)		EF	125(0,266)	57 (0,0925)		EF	121(0,3333)	103 (0,1940))	
EM	591(0,4276)	342 (0,6118)		EM	234(0,4979)	391 (0,6347)		EM	146(0,4022)	292 (0,5499)	
ES	390(0,2822)	101 (0,1807)		ES	93(0,1979)	168 (0,2727)		ES	79(0,2176)	128 (0,2411)	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

3.5.3 PCSVDF-Mulher 2017 versus PNAD-2017

Assim como na análise das amostras de 2016, o cenário de 2017 também apresenta semelhanças. Observa-se uma diferença na distribuição em todos os contextos, conforme a Tabela 13. Na variável raça, a maior discrepância continua sendo entre as mulheres pretas. Em 2017, o percentual de mulheres pretas na PCSVDF-Mulher foi de 22,74%, o que representa uma diferença de cerca de 10% em relação ao percentual de mulheres pretas na PNAD-2017, que é de 13,50%.

No que diz respeito à faixa etária, a PCSVDF-Mulher 2017 focou predominantemente na entrevista de mulheres com idades entre 18 e 39 anos. Em contraste, a PNAD-2017 registrou que o percentual de mulheres na faixa etária de 40 a 49 anos é de 40%. Por outro lado, na amostra da PCSVDF-Mulher, esse percentual foi significativamente menor, atingindo apenas 25,42%.

Por fim, na escolaridade, a diferença se deu por conta do quantitativo de entrevista entre mulheres com ensino fundamental e com ensino superior. A escolaridade da população amostrada na PNAD-2017 apresenta um percentual próximo entre mulheres com Ensino Médio 39,58% e de mulheres com Ensino Fundamental 30,98%. Entretanto a população da PCSVDF-Mulher 2017 possui mais da metade das mulheres com Ensino Médio 56,10%.

Tabela 13 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD²⁰¹⁷, PCSVDF-Mulher²⁰¹⁷)

	PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷	
Branca	1940(0,2924)	2385(0,2427)	<0,001 ¹
Parda	3800(0,5726)	5207(0,5299)	
Preta	896(0,1350)	2234(0,2274)	
15 a 17 anos	277(0,0615)	578(0,0588)	<0,001 ¹
18 a 24 anos	723(0,1605)	2131(0,2169)	
25 a 29 anos	508(0,1127)	1528(0,1555)	
30 a 39 anos	1167(0,259)	2878(0,2929)	
40 a 49 anos	1831(0,4063)	2711(0,2759)	
SE	254(0,045)	101(0,0103)	<0,001 ¹
EF	1747(0,3098)	2234(0,2274)	
EM	2232(0,3958)	5512(0,561)	
ES	1406(0,2493)	1979(0,2014)	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Tabela 14 – Frequência absoluta e relativa, PNAD²⁰¹⁷ e PCSVDF-Mulher²⁰¹⁷ de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.

Aracaju-SE				Fortaleza-CE				João Pessoa-PB			
	PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷			PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷			PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷	
15 a 17 anos	16(0,0687)	34(0,0333)	0,004 ¹	15 a 17 anos	58(0,0624)	71(0,0564)	<0,001 ¹	15 a 17 anos	20(0,0707)	59(0,0538)	0,0374 ¹
18 a 24 anos	35(0,1502)	229(0,2241)		18 a 24 anos	159(0,1712)	284(0,2258)		18 a 24 anos	44(0,1555)	233(0,2124)	
25 a 29 anos	25(0,1073)	113(0,1106)		25 a 29 anos	127(0,1367)	187(0,1486)		25 a 29 anos	33(0,1166)	136(0,124)	
30 a 39 anos	62(0,2661)	310(0,3033)		30 a 39 anos	230(0,2476)	356(0,283)		30 a 39 anos	70(0,2473)	311(0,2835)	
40 a 59 anos	95(0,4077)	336(0,3288)		40 a 59 anos	355(0,3821)	360(0,2862)		40 a 59 anos	116(0,4099)	358(0,3263)	
Branca	82(0,2363)	227(0,2221)	<0,001 ¹	Branca	463(0,3466)	238(0,1892)	<0,001 ¹	Branca	174(0,3937)	326(0,2972)	<0,001 ¹
Parda	242(0,6974)	559(0,547)		Parda	815(0,61)	829(0,659)		Parda	233(0,5271)	592(0,5397)	
Preta	23(0,0663)	236(0,2309)		Preta	58(0,0434)	191(0,1518)		Preta	35(0,0792)	179(0,1632)	
SE	13(0,0451)	10(0,0098)	<0,001 ¹	SE	53(0,0454)	8(0,0064)	<0,001 ¹	SE	15(0,0412)	14(0,0128)	<0,001 ¹
EF	90(0,3125)	246(0,2407)		EF	395(0,3385)	366(0,2909)		EF	111(0,3049)	336(0,3063)	
EM	105(0,3646)	536(0,5245)		EM	461(0,395)	699(0,5556)		EM	120(0,3297)	471(0,4294)	
ES	80(0,2778)	230(0,225)		ES	258(0,2211)	185(0,1471)		ES	118(0,3242)	276(0,2516)	
Maceió-AL				Natal-RN				Recife-PE			
	PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷			PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷			PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷	
15 a 17 anos	22(0,0595)	33(0,029)	<0,001 ¹	15 a 17 anos	20(0,0627)	61(0,0545)	<0,001 ¹	15 a 17 anos	31(0,0534)	67(0,0523)	<0,001 ¹
18 a 24 anos	64(0,173)	250(0,2195)		18 a 24 anos	51(0,1599)	279(0,2491)		18 a 24 anos	85(0,1463)	237(0,185)	
25 a 29 anos	40(0,1081)	161(0,1414)		25 a 29 anos	38(0,1191)	175(0,1563)		25 a 29 anos	58(0,0998)	207(0,1616)	
30 a 39 anos	91(0,2459)	327(0,2871)		30 a 39 anos	68(0,2132)	308(0,275)		30 a 39 anos	146(0,2513)	367(0,2865)	
40 a 59 anos	153(0,4135)	368(0,3231)		40 a 59 anos	142(0,4451)	297(0,2652)		40 a 59 anos	261(0,4492)	403(0,3146)	
Branca	131(0,2404)	325(0,2853)	<0,001 ¹	Branca	179(0,3683)	470(0,4196)	<0,001 ¹	Branca	331(0,3844)	326(0,2545)	<0,001 ¹
Parda	389(0,7138)	579(0,5083)		Parda	283(0,5823)	533(0,4759)		Parda	428(0,4971)	721(0,5628)	
Preta	25(0,0459)	235(0,2063)		Preta	24(0,0494)	117(0,1045)		Preta	102(0,1185)	234(0,1827)	
SE	32(0,0691)	26(0,0228)	<0,001 ¹	SE	15(0,0369)	8(0,0071)	<0,001 ¹	SE	41(0,055)	8(0,0062)	<0,001 ¹
EF	186(0,4017)	289(0,2537)		EF	142(0,3498)	218(0,1946)		EF	231(0,3097)	347(0,2709)	
EM	166(0,3585)	632(0,5549)		EM	148(0,3645)	664(0,5929)		EM	270(0,3619)	744(0,5808)	
ES	79(0,1706)	192(0,1686)		ES	101(0,2488)	230(0,2054)		ES	204(0,2735)	182(0,1421)	
Salvador-BA				São Luís-MA				Teresina-PI			
	PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷			PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷			PNAD ²⁰¹⁷	PCSVDF-Mulher ²⁰¹⁷	
15 a 17 anos	73(0,0672)	47(0,0378)	<0,001 ¹	15 a 17 anos	30(0,075)	60(0,0516)	0,002 ¹	15 a 17 anos	19(0,0627)	31(0,0315)	0,007 ¹
18 a 24 anos	164(0,1509)	248(0,1992)		18 a 24 anos	71(0,1775)	211(0,1814)		18 a 24 anos	48(0,1584)	192(0,1953)	
25 a 29 anos	121(0,1113)	179(0,1438)		25 a 29 anos	50(0,125)	170(0,1462)		25 a 29 anos	33(0,1089)	142(0,1445)	
30 a 39 anos	285(0,2622)	399(0,3205)		30 a 39 anos	95(0,2375)	382(0,3285)		30 a 39 anos	79(0,2607)	293(0,2981)	
40 a 59 anos	444(0,4085)	372(0,2988)		40 a 59 anos	154(0,385)	340(0,2923)		40 a 59 anos	124(0,4092)	325(0,3306)	
Branca	295(0,1872)	77(0,0618)	<0,001 ¹	Branca	124(0,2112)	233(0,2003)	<0,001 ¹	Branca	92(0,204)	177(0,1801)	<0,001 ¹
Parda	712(0,4518)	551(0,4426)		Parda	392(0,6678)	648(0,5572)		Parda	311(0,6896)	615(0,6256)	
Preta	569(0,361)	617(0,4956)		Preta	71(0,121)	282(0,2425)		Preta	48(0,1064)	191(0,1943)	
SE	50(0,0365)	8(0,0064)	<0,001 ¹	SE	25(0,0533)	4(0,0034)	<0,001 ¹	SE	16(0,0437)	8(0,0081)	<0,001 ¹
EF	326(0,2378)	268(0,2153)		EF	114(0,2431)	134(0,1152)		EF	131(0,3579)	172(0,175)	
EM	614(0,4478)	756(0,6072)		EM	236(0,5032)	739(0,6354)		EM	142(0,388)	590(0,6002)	
ES	381(0,2779)	213(0,1711)		ES	94(0,2004)	286(0,2459)		ES	77(0,2104)	213(0,2167)	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

3.5.4 PCSVDF-Mulher^{REP} versus PNAD-2017

A Tabela 15 representa o comparativo entre a distribuição das variáveis raça, escolaridade e idade entre a PCSVDF-Mulher^{REP} e PNAD-2017. O percentual de mulheres pretas na PCSVDF-Mulher^{REP} (22,34%) é quase 10% superior ao percentual encontrado na PNAD-2017 (13,50%).

Em relação à idade, a PCSVDF-Mulher^{REP} focou principalmente na entrevista de mulheres com idades entre 30 e 49 anos, que correspondem a aproximadamente 55% do total. Em contraste, na PNAD-2017, o percentual de mulheres nessa faixa etária é superior a 65%.

Por fim, a análise da escolaridade entre as amostras revela que a diferença observada se deve ao número de entrevistas realizadas com mulheres que possuem apenas o ensino fundamental em comparação àquelas que têm ensino superior. Essa disparidade sugere que a amostra da PCSVDF-Mulher^{REP} tende a ser mais educada.

Tabela 15 – Frequência absoluta e relativa de acordo com a raça, escolaridade e idade em dois cenários (PNAD²⁰¹⁷, PCSVDF-Mulher^{REP})

	PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}	
Branca	1940(0,2924)	1380 (0,2512)	<0,001 ¹
Parda	3800(0,5726)	2852 (0,5192)	
Preta	896(0,1350)	1227 (0,2234)	
15 a 17 anos	277(0,0615)	297(0,0562)	<0,001 ¹
18 a 24 anos	723(0,1605)	1258(0,2381)	
25 a 29 anos	508(0,1127)	884(0,1672)	
30 a 39 anos	1167(0,2590)	1502(0,2843)	
40 a 49 anos	1831(0,4063)	1343(0,2542)	
SE	254(0,045)	57 (0,0103)	<0,001 ¹
EF	1747(0,3098)	1178 (0,2131)	
EM	2232(0,3958)	2347 (0,4245)	
ES	1406(0,2493)	1947 (0,3521)	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

As diferenças nas variáveis entre os conjuntos de dados da PNAD-2016 e PNAD-2017, em comparação com a PCSVDF-Mulher, tanto na amostra completa quanto nas subamostras de atrito e reposição, destacam a necessidade de aplicar a ponderação dos dados. Isso se deve às diferenças significativas observadas entre as variáveis. A ponderação permitirá uma melhor representatividade dos resultados para a população em geral. Mais informações sobre a ponderação da PCSVDF-Mulher estão disponíveis em Andre e Carvalho (2023).

Tabela 16 – Frequência absoluta e relativa, PNAD²⁰¹⁷ e PCSVDF-Mulher^{REP} de acordo com a idade, raça e escolaridade, segundo o município em que ocorreu a pesquisa.

Aracaju-SE			Fortaleza-CE			João Pessoa-PB		
	PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}		PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}		PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}
15 a 17 anos	16(0,0687)	21(0,034)	15 a 17 anos	58(0,0624)	30(0,0535)	15 a 17 anos	20(0,0707)	42(0,0646)
18 a 24 anos	35(0,1502)	143(0,2318)	18 a 24 anos	159(0,1712)	143(0,2549)	18 a 24 anos	44(0,1555)	140(0,2154)
25 a 29 anos	25(0,1073)	67(0,1086)	25 a 29 anos	127(0,1367)	91(0,1622)	25 a 29 anos	33(0,1166)	88(0,1354)
30 a 39 anos	62(0,2661)	191(0,3096)	30 a 39 anos	230(0,2476)	135(0,2406)	30 a 39 anos	70(0,2473)	180(0,2769)
40 a 59 anos	95(0,4077)	195(0,316)	40 a 59 anos	355(0,3821)	162(0,2888)	40 a 59 anos	116(0,4099)	200(0,3077)
Branca	82(0,2363)	145(0,2366)	Branca	463(0,3466)	108(0,1943)	Branca	174(0,3937)	174(0,3062)
Parda	242(0,6974)	335(0,5413)	Parda	815(0,61)	360(0,6471)	Parda	233(0,5271)	312(0,5262)
Preta	23(0,0663)	126(0,222)	Preta	58(0,0434)	86(0,1586)	Preta	35(0,0792)	103(0,1677)
SE	13(0,0451)	05 (0,0082)	SE	53(0,0454)	5(0,0088)	SE	15(0,0412)	12 (0,0202)
EF	90(0,3125)	134 (0,2200)	EF	395(0,3385)	172(0,3044)	EF	111(0,3049)	180 (0,3025)
EM	105(0,3646)	325 (0,5337)	EM	461(0,395)	306(0,5416)	EM	120(0,3297)	256 (0,4303)
ES	80(0,2778)	145 (0,2381)	ES	258(0,2211)	82(0,1451)	ES	118(0,3242)	147 (0,2471)
Maceió-AL			Natal-RN			Recife-PE		
	PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}		PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}		PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}
15 a 17 anos	22(0,0595)	24(0,0359)	15 a 17 anos	20(0,0627)	39(0,0528)	15 a 17 anos	31(0,0534)	58(0,0934)
18 a 24 anos	64(0,173)	163(0,244)	18 a 24 anos	51(0,1599)	205(0,2778)	18 a 24 anos	85(0,1463)	137(0,2206)
25 a 29 anos	40(0,1081)	96(0,1437)	25 a 29 anos	38(0,1191)	130(0,1762)	25 a 29 anos	58(0,0998)	104(0,1675)
30 a 39 anos	91(0,2459)	176(0,2635)	30 a 39 anos	68(0,2132)	192(0,2602)	30 a 39 anos	146(0,2513)	150(0,2415)
40 a 59 anos	153(0,4135)	209(0,3129)	40 a 59 anos	142(0,4451)	172(0,2331)	40 a 59 anos	261(0,4492)	172(0,277)
Branca	131(0,2404)	197(0,3039)	Branca	179(0,3683)	292(0,4214)	Branca	331(0,3844)	208(0,4098)
Parda	389(0,7138)	299(0,494)	Parda	283(0,5823)	338(0,4661)	Parda	428(0,4971)	281(0,4564)
Preta	25(0,0459)	129(0,2021)	Preta	24(0,0494)	88(0,1125)	Preta	102(0,1185)	96(0,1562)
SE	32(0,0691)	12 (0,0189)	SE	15(0,0369)	05 (0,0069)	SE	41(0,055)	8 (0,0133)
EF	186(0,4017)	160 (0,2520)	EF	142(0,3498)	122 (0,1678)	EF	231(0,3097)	126 (0,2100)
EM	166(0,3585)	357 (0,5622)	EM	148(0,3645)	454 (0,6245)	EM	270(0,3619)	353 (0,5883)
ES	79(0,1706)	106 (0,1669)	ES	101(0,2488)	146 (0,2008)	ES	204(0,2735)	113 (0,1883)
Salvador-BA			São Luís-MA			Teresina-PI		
	PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}		PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}		PNAD ²⁰¹⁷	PCSVDF-Mulher ^{REP}
15 a 17 anos	73(0,0672)	19(0,0279)	15 a 17 anos	30(0,075)	31(0,0473)	15 a 17 anos	19(0,0627)	20(0,0372)
18 a 24 anos	164(0,1509)	146(0,2141)	18 a 24 anos	71(0,1775)	127(0,1939)	18 a 24 anos	48(0,1584)	120(0,223)
25 a 29 anos	121(0,1113)	110(0,1613)	25 a 29 anos	50(0,125)	99(0,1511)	25 a 29 anos	33(0,1089)	82(0,1524)
30 a 39 anos	285(0,2622)	222(0,3255)	30 a 39 anos	95(0,2375)	212(0,3237)	30 a 39 anos	79(0,2607)	174(0,3234)
40 a 59 anos	444(0,4085)	185(0,2713)	40 a 59 anos	154(0,385)	186(0,284)	40 a 59 anos	124(0,4092)	142(0,2639)
Branca	295(0,1872)	36(0,0674)	Branca	124(0,2112)	128(0,2031)	Branca	92(0,204)	92(0,1729)
Parda	712(0,4518)	261(0,4062)	Parda	392(0,6678)	329(0,5527)	Parda	311(0,6896)	337(0,632)
Preta	569(0,361)	337(0,5264)	Preta	71(0,121)	168(0,2443)	Preta	48(0,1064)	94(0,1952)
SE	50(0,0365)	04 (0,0062)	SE	25(0,0533)	2(0,0031)	SE	16(0,0437)	01 (0,0016)
EF	326(0,2378)	139 (0,2165)	EF	114(0,2431)	79(0,1206)	EF	131(0,3579)	74 (0,1173)
EM	614(0,4478)	370 (0,5763)	EM	236(0,5032)	414(0,6321)	EM	142(0,388)	403 (0,6387)
ES	381(0,2779)	129 (0,2009)	ES	94(0,2004)	160(0,2443)	ES	77(0,2104)	153 (0,2425)

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

3.5.5 *PCSVDF-Mulher^{ATT} versus PCSVDF-Mulher^{REP}*

Neste momento, é fundamental investigar se a base de dados de reposição apresenta um perfil semelhante ao da base de dados das mulheres atritadas. A presente seção será dedicada a uma análise detalhada, abordando cada aspecto relevante dessa comparação.

Do ponto de vista teórico, para verificar a eficiência da substituição era necessário examinar todas as variáveis da PCSVDF-Mulher, porém isso não é factível de se fazer, de modo que serão escolhidos um conjunto de variáveis usualmente utilizadas na pesquisa de modelagem dos determinantes da violência por parceiro íntimo.

A Tabela 17 sugere que somente as variáveis idade, não pretende mudar de domicílio e domicílio próprio mostraram diferenças significativas em relação a amostra de atritadas e de reposição, as demais variáveis tiveram comportamentos semelhantes, o que é de certa forma importante para se mostrar que apesar de a camada de sofisticação de design de substituição da pesquisa não ser tão grande, por exemplo, em comparação aos protocolos apresentados na seção (3.3 Exemplos de Protocolos de substituição), ela consegue ser eficaz.

Além disso, a Tabela 17 fornece um fato interessante, com relação aos episódios de violência. A violência emocional (valor significativo) apresenta uma ligeira proporção maior na reposição em comparação as mulheres atritadas. Em contrapartida, apesar do valor das proporções não serem significativamente diferentes, as proporções na violência física e sexual é menor no recorte de reposições em comparação ao recorte de atrito, ou seja, para esses tipos de violência a nossa nova amostra é ligeiramente menos vitimizada. Embora ainda haja controle sobre a reposição, percebe-se que a amostra está se tornando cada vez menos vitimizada.

Considerando a amostra da pesquisa inteira a reposição foi eficaz, no sentido que grande parte das variáveis tiveram os seus testes, seja de igualdade de média, seja de homogeneidade e seja de proporção não rejeitados e, com isso, dando indícios que as variáveis da população atritada e de reposição possuem valores semelhantes.

Tabela 17 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição.

	ATT (N = 5302)	REP (N = 5529)	p-valor	ATT (N = 5302)	REP (N = 5529)	p-valor
Teste médias				Teste variância		
Idade	31,4717	32,1030	<0,001***	98,7185	98,9604	0,5774
Número de filhos	1,8371	1,8016	0,2511	1,5659	1,4484	0,8601
Teste homogeneidade						
Raça						
Branca	1366 (0,2576)	1380 (0,2512)				
Preta	1126 (0,2124)	1227 (0,2234)	0,5088	-	-	-
Parda	2781 (0,5245)	2852 (0,5192)		-	-	-
Outras	29 (0,0055)	34 (0,0062)		-	-	-
Escolaridade						
Sem escolaridade	54 (0,0102)	57 (0,0103)		-	-	-
Ensino fundamental	1128 (0,2127)	1178 (0,2131)	0,3085	-	-	-
Ensino médio	2167 (0,4087)	2347 (0,4245)		-	-	-
Ensino superior	1953 (0,3684)	1947 (0,3521)		-	-	-
Religião						
Sem religião	816 (0,1542)	905 (0,1637)		-	-	-
Católica	2640 (0,4990)	2733 (0,4943)	0,5034	-	-	-
Evangélica	1500 (0,2835)	1563 (0,2827)		-	-	-
Outras	335 (0,0633)	328 (0,0593)		-	-	-
Teste proporção						
Cohabita	0,4683	0,4663	0,8622	-	-	-
Trabalho remunerado	0,4102	0,4166	0,5146	-	-	-
Não pretende mudar	0,6742	0,6925	0,0469**	-	-	-
Domicílio próprio	0,5681	0,5921	0,0122**	-	-	-
Episódios de Violência						
Violência	0,3559	0,3599	0,7846	-	-	-
Violência emocional	0,2977	0,3200	0,0783*	-	-	-
Violência física	0,1905	0,1890	0,9128	-	-	-
Violência sexual	0,0813	0,0755	0,4605	-	-	-

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

ATT: PCSVDF-Mulher (atritadas) e REP: PCSVDF-Mulher (reposição)

Agora o mesmo estudo será replicado em cada cidade, com o objetivo de verificar se as variáveis apresentam diferenças significativas quando se reduz o escopo espacial amostral. Por uma questão de segregação, por causa da uma redução das amostras, muito provavelmente em alguns municípios o número de rejeições dos testes pode aumentar. E com isso tem-se o seguinte questionamento, a eficácia da coleta de dados será feita com base no dado inteiro ou cidade por cidade.

Os testes das variáveis de interesse estão apresentados na Tabela 18 a 22 com a divisão dos estratos geográficos. Observa-se, com relação as variáveis numéricas, a idade apresentou diferença nas médias nos municípios de Fortaleza-CE, João Pessoa-PB, Natal-RN, Aracajú-SE e São Luís-MA, com exceção do município de Natal-RN em que a média de idade foi inferior no recorte de reposições, os demais municípios em que as médias apresentaram valores

diferentes e significativos, tiveram a média de idade superior na reposição em comparação a atritadas. A média de idade variou com valor mínimo de 30,8165 anos e valor máximo de 33,2203 anos.

O *número de filhos* apresentou diferença significativa somente no município de Maceió-AL (a um nível de significância de 10%) e no município de Aracajú-SE. Em ambos a média de filhos tendeu a diminuir na população nova em comparação a população atritada na PCSVDF-Mulher.

As variáveis categóricas apresentam um comportamento bem semelhante dentro dos municípios, como já observado no aspecto geral da pesquisa, em que raça, escolaridade e religião não foram rejeitadas no teste de homogeneidade entre as distribuições. Fortaleza-CE foi o único município que apresentou diferença entre a variável raça e observando os percentuais da variável a raça branca, sendo a principal diferença entre os percentuais nas raças branca e preta.

A principal diferença na distribuição da escolaridade ocorreu no município de João Pessoa-PB. Enquanto na população dos atritados a escolaridade de ensino superior apareceu como a segunda escolaridade com o maior percentual, entretanto na população de reposição o segundo percentual foi o da escolaridade ensino fundamental.

Dentre as variáveis categóricas, a *religião* apresentou um maior número de municípios com diferença na distribuição da variável entre os atritados e reposições. Dos nove municípios nordestinos, quatro apresentaram diferença na distribuição da variável.

O teste de proporção para as variáveis dummy indicam que *não pretende mudar de endereço* só apresentou diferença significativa na proporção no município de Fortaleza-CE. No geral, as pessoas que são reposição tendem a não pretender se mudar em comparação as que foram atritadas.

A variável de *cohabitação* apresentou diferença significativa entre as proporções nos municípios de João Pessoa-PB, Natal-RN e Teresina-PI, sendo os dois últimos municípios com redução na proporção da variável e o município de João Pessoa-PB com aumento na proporção da variável. Isso indica que nos municípios que houve a redução existe uma menor propensão a se viver com o parceiro nas novas mulheres que estão entrando no painel.

Para as variáveis de violência, apesar de poucos recortes serem significativos enquanto a diferença de proporções, é importante detalhar minuciosamente esse processo. Tendo em vista que a amostra nova (substituição) pode está deixando a amostra menos ou mais vitimizada.

Tabela 18 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atriuto e na base de reposição. Fortaleza-CE e Recife-PE

Fortaleza-CE	ATT (N = 504)	REP (N = 565)	p-valor	ATT (N = 504)	REP (N = 565)	p-valor
Teste médias				Teste variância		
Idade	30,3948	31,7992	0,0257**	104,0008	106,2784	0,4993
Número de filhos	1,9315	2,0512	0,2103	1,4626	1,5173	0,8922
Teste homogeneidade						
Raça						
Branca	124 (0,2460)	108 (0,1942)	0,0079**	-	-	-
Preta	46 (0,0913)	86 (0,1547)		-	-	-
Parda	332 (0,6587)	360 (0,6475)		-	-	-
Outras	02 (0,0040)	02 (0,0036)		-	-	-
Escolaridade						
Sem escolaridade	06 (0,0119)	05 (0,0088)	0,3085	-	-	-
Ensino fundamental	136 (0,2698)	172 (0,3044)		-	-	-
Ensino médio	292 (0,5794)	306 (0,5416)		-	-	-
Ensino superior	70 (0,1389)	82 (0,1451)		-	-	-
Religião						
Sem religião	62 (0,1233)	70 (0,1239)	0,9359	-	-	-
Católica	276 (0,5487)	318 (0,5628)		-	-	-
Evangélica	144 (0,2863)	152 (0,2690)		-	-	-
Outras	21 (0,0417)	25 (0,0442)		-	-	-
Teste proporção						
Cohabita	0,5167	0,5128	0,9536	-	-	-
Trabalho remunerado	0,4102	0,4265	0,6354	-	-	-
Não pretende mudar	0,6598	0,7307	0,0173**	-	-	-
Domicílio próprio	0,5264	0,5433	0,6248	-	-	-
Episódios de Violência						
Violência	0,3710	0,3688	0,9845	-	-	-
Violência emocional	0,3024	0,3300	0,4877	-	-	-
Violência física	0,2256	0,1990	0,4497	-	-	-
Violência sexual	0,0740	0,1024	0,2442	-	-	-
Recife-PE	ATT (N = 633)	REP (N = 600)	p-valor	ATT (N = 633)	REP (N = 600)	p-valor
Teste médias				Teste variância		
Idade	30,8988	31,6450	0,1842	99,2207	94,9605	0,7208
Número de filhos	1,8141	1,9068	0,3258	1,2193	1,5925	0,2154
Teste homogeneidade						
Raça						
Branca	227 (0,3586)	208 (0,3502)	0,3697	-	-	-
Preta	114 (0,1801)	96 (0,1616)		-	-	-
Parda	288 (0,4550)	281 (0,4731)		-	-	-
Outras	04 (0,0063)	09 (0,0152)		-	-	-
Escolaridade						
Sem escolaridade	08 (0,0126)	08 (0,0133)	0,8564	-	-	-
Ensino fundamental	146 (0,2306)	126 (0,2100)		-	-	-
Ensino médio	362 (0,5719)	353 (0,5883)		-	-	-
Ensino superior	117 (0,1848)	113 (0,1883)		-	-	-
Religião						
Sem religião	129 (0,2041)	135 (0,2250)	0,5133	-	-	-
Católica	245 (0,3877)	213 (0,3550)		-	-	-
Evangélica	211 (0,3339)	213 (0,3550)		-	-	-
Outras	47 (0,0744)	39 (0,0650)		-	-	-
Teste proporção						
Cohabita	0,4331	0,4711	0,2396	-	-	-
Trabalho remunerado	0,4076	0,4124	0,9102	-	-	-
Não pretende mudar	0,6633	0,7076	0,1154	-	-	-
Domicílio próprio	0,6172	0,6516	0,2330	-	-	-
Episódios de Violência						
Violência	0,3563	0,4100	0,3029	-	-	-
Violência emocional	0,3147	0,3729	0,2394	-	-	-
Violência física	0,1727	0,2352	0,1423	-	-	-
Violência sexual	0,0561	0,0905	0,2390	-	-	-

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

ATT: PCSVDF-Mulher (atritadas) e REP: PCSVDF-Mulher (reposição)

Tabela 19 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição. Salvador-BA e Maceió-AL

Salvador-BA	ATT (N = 559)	REP (N = 642)	p-valor	ATT (N = 559)	REP (N = 642)	p-valor
Teste médias				Teste variância		
Idade	31,7227	32,3551	0,2541	94,8064	89,1747	0,4446
Número de filhos	1,7771	1,8265	0,5899	1,5759	1,5657	0,4033
Teste homogeneidade						
Raça						
Branca	34 (0,0608)	36 (0,0565)				
Preta	290 (0,5188)	337 (0,5290)	0,9776	-	-	-
Parda	232 (0,4150)	261 (0,4097)		-	-	-
Outras	03 (0,0054)	03 (0,0047)		-	-	-
Escolaridade						
Sem escolaridade	03 (0,0054)	04 (0,0062)		-	-	-
Ensino fundamental	113 (0,2021)	139 (0,2165)	0,6578	-	-	-
Ensino médio	342 (0,6118)	370 (0,5763)		-	-	-
Ensino superior	101 (0,1807)	129 (0,2009)		-	-	-
Religião						
Sem religião	174 (0,3113)	211 (0,3287)		-	-	-
Católica	160 (0,2862)	188 (0,2928)	0,2025	-	-	-
Evangélica	177 (0,3166)	172 (0,2679)		-	-	-
Outras	48 (0,0859)	71 (0,1106)		-	-	-
Teste proporção						
Cohabita	0,4341	0,4586	0,4368	-	-	-
Trabalho remunerado	0,4424	0,4423	0,9604	-	-	-
Não pretende mudar	0,6821	0,6523	0,3151	-	-	-
Domicílio próprio	0,6540	0,6791	0,3916	-	-	-
Episódios de Violência						
Violência	0,3172	0,4154	0,0058**	-	-	-
Violência emocional	0,2388	0,3673	< 0,001***	-	-	-
Violência física	0,1755	0,2093	0,2684	-	-	-
Violência sexual	0,0981	0,0958	0,9839	-	-	-
Maceió-AL	ATT (N = 505)	REP (N = 635)	p-valor	ATT (N = 505)	REP (N = 635)	p-valor
Teste médias				Teste variância		
Idade	31,4415	32,5779	0,1842	96,2907	104,6954	0,1323
Número de filhos	1,6972	1,5411	0,0722*	1,4472	1,0840	0,1916
Teste homogeneidade						
Raça						
Branca	151 (0,2990)	197 (0,3127)				
Preta	105 (0,2079)	129 (0,2048)	0,9329	-	-	-
Parda	246 (0,4871)	299 (0,4746)		-	-	-
Outras	03 (0,0059)	05 (0,0079)		-	-	-
Escolaridade						
Sem escolaridade	13 (0,0257)	12 (0,0189)		-	-	-
Ensino fundamental	131 (0,2594)	160 (0,2520)	0,0964*	-	-	-
Ensino médio	252 (0,4990)	357 (0,5622)		-	-	-
Ensino superior	109 (0,2158)	106 (0,1669)		-	-	-
Religião						
Sem religião	52 (0,1032)	100 (0,1575)		-	-	-
Católica	260 (0,5159)	318 (0,5008)	0,0085**	-	-	-
Evangélica	154 (0,3056)	190 (0,2992)		-	-	-
Outras	38 (0,0754)	27 (0,0425)		-	-	-
Teste proporção						
Cohabita	0,4771	0,4864	0,8074	-	-	-
Trabalho remunerado	0,4649	0,4532	0,7405	-	-	-
Não pretende mudar	0,6369	0,6515	0,6716	-	-	-
Domicílio próprio	0,4650	0,5417	0,0122**	-	-	-
Episódios de Violência						
Violência	0,3839	0,3468	0,4173	-	-	-
Violência emocional	0,3347	0,3128	0,6457	-	-	-
Violência física	0,2085	0,1888	0,6394	-	-	-
Violência sexual	0,1130	0,0593	0,0350**	-	-	-

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

ATT: PCSVDF-Mulher (atritadas) e REP: PCSVDF-Mulher (reposição)

Tabela 20 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição. João Pessoa-PB e Natal-RN

João Pessoa-PB	ATT (N = 672)	REP (N = 595)	p-valor	ATT (N = 672)	REP (N = 595)	p-valor
Teste médias				Teste variância		
Idade	31,2782	32,6655	0,0123**	91,5871	102,5091	0,0519
Número de filhos	1,7543	1,8847	0,1152	1,5759	1,5657	0,7112
Teste homogeneidade						
Raça						
Branca	225 (0,3348)	174 (0,0565)				
Preta	94 (0,1399)	103 (0,5290)	0,2526	-	-	-
Parda	349 (0,5193)	312 (0,4097)		-	-	-
Outras	04 (0,0060)	04 (0,0047)		-	-	-
Escolaridade						
Sem escolaridade	05 (0,0074)	12 (0,0202)		-	-	-
Ensino fundamental	143 (0,2128)	180 (0,3025)	<0,001***	-	-	-
Ensino médio	307 (0,4568)	256 (0,4303)		-	-	-
Ensino superior	217 (0,3229)	147 (0,2471)		-	-	-
Religião						
Sem religião	104 (0,1548)	79 (0,1328)		-	-	-
Católica	340 (0,5060)	275 (0,4622)	0,0786*	-	-	-
Evangélica	190 (0,2827)	208 (0,3496)		-	-	-
Outras	38 (0,0565)	33 (0,0555)		-	-	-
Teste proporção						
Cohabita	0,4810	0,5384	0,0517*	-	-	-
Trabalho remunerado	0,3991	0,4067	0,8272	-	-	-
Não pretende mudar	0,6515	0,6961	0,1108	-	-	-
Domicílio próprio	0,4790	0,5126	0,2566	-	-	-
Episódios de Violência						
Violência	0,3895	0,3750	0,7304	-	-	-
Violência emocional	0,3454	0,3324	0,7572	-	-	-
Violência física	0,1975	0,1977	0,9967	-	-	-
Violência sexual	0,0965	0,0560	0,0512*	-	-	-
Natal-RN	ATT (N = 688)	REP (N = 727)	p-valor	ATT (N = 688)	REP (N = 727)	p-valor
Teste médias				Teste variância		
Idade	32,1293	30,8624	0,0190**	108,0167	98,2179	0,0605*
Número de filhos	1,9952	1,8655	0,1575	1,4472	1,0840	0,1916
Teste homogeneidade						
Raça						
Branca	246 (0,3576)	292 (0,4056)				
Preta	90 (0,1308)	88 (0,1222)	0,3072	-	-	-
Parda	349 (0,5073)	338 (0,4694)		-	-	-
Outras	03 (0,0044)	02 (0,0028)		-	-	-
Escolaridade						
Sem escolaridade	04 (0,0058)	05 (0,0069)		-	-	-
Ensino fundamental	141 (0,2049)	122 (0,1678)	0,2773	-	-	-
Ensino médio	400 (0,5814)	454 (0,6245)		-	-	-
Ensino superior	143 (0,2078)	146 (0,2008)		-	-	-
Religião						
Sem religião	112 (0,1633)	100 (0,1774)		-	-	-
Católica	350 (0,5102)	318 (0,4856)	0,8068	-	-	-
Evangélica	193 (0,2813)	190 (0,2916)		-	-	-
Outras	31 (0,0452)	27 (0,0454)		-	-	-
Teste proporção						
Cohabita	0,5251	0,4864	0,0114**	-	-	-
Trabalho remunerado	0,3543	0,3319	0,4078	-	-	-
Não pretende mudar	0,6937	0,6924	0,9578	-	-	-
Domicílio próprio	0,5379	0,4759	0,0228**	-	-	-
Episódios de Violência						
Violência	0,4343	0,3620	0,1172	-	-	-
Violência emocional	0,3928	0,3322	0,1819	-	-	-
Violência física	0,2201	0,1827	0,3504	-	-	-
Violência sexual	0,0751	0,0701	0,9714	-	-	-

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

ATT: PCSVDF-Mulher (atritadas) e REP: PCSVDF-Mulher (reposição)

Tabela 21 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição, Aracaju-SE e Teresina-PI

Aracaju-SE	ATT (N = 594)	REP (N = 609)	p-valor	ATT (N = 594)	REP (N = 609)	p-valor
Teste médias				Teste variância		
Idade	31,3585	32,7668	0,0136**	95,9369	99,3106	0,5551
Número de filhos	1,9662	1,7699	0,0245**	1,5087	1,1185	0,4089
Teste homogeneidade						
Raça						
Branca	133 (0,2239)	145 (0,2381)	0,7051	-	-	-
Preta	122 (0,2054)	126 (0,2069)		-	-	-
Parda	333 (0,5606)	335 (0,5501)		-	-	-
Outras	06 (0,0101)	03 (0,0049)		-	-	-
Escolaridade						
Sem escolaridade	07 (0,0118)	05 (0,0082)	0,2306	-	-	-
Ensino fundamental	158 (0,2660)	134 (0,2200)		-	-	-
Ensino médio	304 (0,5118)	325 (0,5337)		-	-	-
Ensino superior	125 (0,2104)	145 (0,2381)		-	-	-
Religião						
Sem religião	104 (0,1760)	67 (0,1100)	0,0039**	-	-	-
Católica	300 (0,5074)	356 (0,5846)		-	-	-
Evangélica	135 (0,2284)	142 (0,2332)		-	-	-
Outras	52 (0,0880)	44 (0,0722)		-	-	-
Teste proporção						
Cohabita	0,4683	0,4278	0,1921	-	-	-
Trabalho remunerado	0,3855	0,4417	0,0564*	-	-	-
Não pretende mudar	0,7005	0,6582	0,1364	-	-	-
Domicílio próprio	0,5287	0,5566	0,3609	-	-	-
Episódios de Violência						
Violência	0,3166	0,3450	0,5213	-	-	-
Violência emocional	0,2523	0,2866	0,3817	-	-	-
Violência física	0,1612	0,1619	0,9967	-	-	-
Violência sexual	0,0911	0,0885	0,9844	-	-	-
Teresina-PI	ATT (N = 531)	REP (N = 525)	p-valor	ATT (N = 531)	REP (N = 525)	p-valor
Teste médias				Teste variância		
Idade	33,2203	32,2304	0,1031	101,5834	92,6891	0,0720*
Número de filhos	2,0066	1,8441	0,1308	2,2340	1,2850	0,3595
Teste homogeneidade						
Raça						
Branca	108 (0,2034)	92 (0,1756)	0,3072	-	-	-
Preta	97 (0,1827)	94 (0,1794)		-	-	-
Parda	325 (0,6121)	337 (0,6431)		-	-	-
Outras	01 (0,0019)	01 (0,0019)		-	-	-
Escolaridade						
Sem escolaridade	08 (0,0151)	05 (0,0095)	0,0435**	-	-	-
Ensino fundamental	103 (0,1940)	71 (0,1352)		-	-	-
Ensino médio	292 (0,5499)	323 (0,6152)		-	-	-
Ensino superior	128 (0,2411)	126 (0,2400)		-	-	-
Religião						
Sem religião	20 (0,0377)	43 (0,0819)	0,0023**	-	-	-
Católica	367 (0,6911)	371 (0,7067)		-	-	-
Evangélica	119 (0,2241)	84 (0,1600)		-	-	-
Outras	25 (0,0471)	27 (0,0514)		-	-	-
Teste proporção						
Cohabita	0,5172	0,4407	0,0255**	-	-	-
Trabalho remunerado	0,3787	0,3709	0,8421	-	-	-
Não pretende mudar	0,7159	0,7554	0,1742	-	-	-
Domicílio próprio	0,6792	0,6933	0,6696	-	-	-
Episódios de Violência						
Violência	0,4215	0,3369	0,0728*	-	-	-
Violência emocional	0,3380	0,3024	0,4578	-	-	-
Violência física	0,2303	0,1697	0,1255	-	-	-
Violência sexual	0,0947	0,0602	0,2055	-	-	-

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher,
ATT: PCSVDF-Mulher (atritadas) e REP: PCSVDF-Mulher (reposição)

Tabela 22 – Testes estatísticos para verificar a semelhança entre as variáveis na base de dados de atrito e na base de reposição, São Luís-MA

São Luís-MA	ATT (N = 616)	REP (N = 631)	p-valor	ATT (N = 616)	REP (N = 631)	p-valor
Teste médias				Teste variância		
Idade	30,8165	32,2297	0,0115**	93,3597	100,9360	0,1636
Número de filhos	1,5731	1,5212	0,5878	1,6884	1,3566	0,4802
Teste homogeneidade						
Raça						
Branca	118 (0,1916)	128 (0,2032)	0,8601	-	-	-
Preta	168 (0,2727)	168 (0,2667)		-	-	-
Parda	327 (0,5308)	329 (0,5222)		-	-	-
Outras	03 (0,0049)	05 (0,0079)		-	-	-
Escolaridade						
Sem escolaridade	00 (0,0000)	01 (0,0016)	0,2715	-	-	-
Ensino fundamental	57 (0,0925)	74 (0,1173)		-	-	-
Ensino médio	391 (0,6347)	403 (0,6387)		-	-	-
Ensino superior	168 (0,2727)	153 (0,2425)		-	-	-
Religião						
Sem religião	59 (0,0962)	71 (0,1125)	0,5994	-	-	-
Católica	342 (0,5579)	341 (0,5404)		-	-	-
Evangélica	177 (0,2887)	190 (0,3011)		-	-	-
Outras	35 (0,0571)	29 (0,0460)		-	-	-
Teste proporção						
Cohabita	0,3926	0,4166	0,4254	-	-	-
Trabalho remunerado	0,4641	0,4715	0,8400	-	-	-
Não pretende mudar	0,6588	0,7037	0,1057	-	-	-
Domicílio próprio	0,6311	0,7004	0,0111**	-	-	-
Episódios de Violência						
Violência	0,2576	0,2679	0,8430	-	-	-
Violência emocional	0,2117	0,2352	0,5394	-	-	-
Violência física	0,1447	0,1526	0,8718	-	-	-
Violência sexual	0,0256	0,0487	0,1829	-	-	-

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

ATT: PCSVDF-Mulher (atritadas) e REP: PCSVDF-Mulher (reposição)

3.6 Metodologia II: Representatividade estatística espacial

Os resultados encontrados na análise do valor de $N(g)=N_{17}/N_{16}$ para a desagregação de bairro e setor censitário dão indícios que a reposição preservou a área até o nível de bairro, e que quando desagrega para o próximo nível - setor censitário - ocorre uma maior variação de $N(g)$. Por esse motivo, é necessário um estudo mais detalhado do processo de reposição à nível d setor censitário. Para isso, é necessário definir uma linguagem que irá classificar o setor censitário a partir do quantitativo de domicílios atritados, pareados e de reposição entre os anos de 2016 e 2017.

3.6.1 Definindo uma linguagem operacional

Essa subseção irá definir uma linguagem para o estudo de atrito e reposição da PCSVDF-Mulher. Será definido como $D(g)$ o número de domicílios entrevistados dentro de uma geografia “g”, seja a geografia o município, bairro/subdistrito ou o setor censitário. A partir das

coletas de 2016 e 2017, os domicílios podem ser classificados de três formas: pareados, atritados e de reposição.

- Domicílios pareados são domicílios válidos em 2016 e válidos ¹ em 2017 ;
- Domicílios atritados são domicílios válidos em 2016 e não válidos em 2017;
- Domicílios de reposição são domicílios, participantes apenas em 2017, que irão repor os domicílios onde houve atrito de 2016 para 2017.

$$\Delta = D_{ent}^{17}(g) - D_{ent}^{16}(g) \leq 0 \quad (3.1)$$

$$D_{ent}^{16}(g) = D_{atr}^{16}(g) + D_{par}^{16}(g) \geq 0 \quad (3.2)$$

$$D_{ent}^{17}(g) - \Delta = D_{atr}^{16}(g) + D_{par}^{16}(g) \quad (3.3)$$

$$D_{ent}^{17}(g) = (D_{atr}^{16}(g) + \Delta) + D_{par}^{16}(g) \quad (3.4)$$

$$D_{ent}^{17}(g) = D_{rep}^{17}(g) + D_{par}^{16}(g) \quad (3.5)$$

Onde:

- $D_{ent}^{16}(g)$ Número de domicílios entrevistados em 2016 na geografia;
- $D_{ent}^{17}(g)$ Número de domicílios entrevistados em 2017 na geografia;
- $D_{atr}^{16}(g)$ Número de domicílios atritados 2016-2017 na geografia;
- $D_{par}^{16}(g)$ Número de domicílios pareados 2016-2017 na geografia;
- $D_{rep}^{17}(g)$ Número de domicílios reposição 2017 na geografia.

A equação 3.1 indica a definição de Δ como a subtração da quantidade de domicílios visitados na geografia em 2017 e em 2016. Esse valor pode ser maior, igual ou menor a zero. A equação 3.2 descreve a definição da quantidade de domicílios visitados na geografia em 2016. Combinando 3.1 e 3.2 pode-se escrever a equação 3.4, que é a quantidade de domicílios na geografia no ano de 2017. Entretanto, a soma $(D_{atr}^{16}(g) + \Delta)$, que é quantidade de atritos em 2016 mais um número Δ , é considerada o número de reposições em 2017 naquela geografia $D_{rep}^{17}(g)$, equação 3.5.

Após a definição das quantidades, são considerados dois valores fundamentais $D_{ent}^{16}(g)$ e Δ que vão definir dois cenários com cinco situações distintas em relação a qualquer geografia:

Cenário 1 - $D_{ent}^{16}(g) > 0$

¹ É considerado válido o domicílio em que a entrevistada residente aceita participar da seção de saúde reprodutiva

Essa premissa indica que, naquela determinada geografia, houve pelo menos um domicílio entrevistado no ano de 2016. Por definição desse cenário, tem-se $D_{ent}^{16}(g) > 0 \Rightarrow D_{ent}^{17}(g) > 0$ e com isso precisa-se analisar a seguinte equação $D_{rep}^{17}(g) = D_{atr}^{16}(g) + \Delta$

- S1)** Se $\Delta > 0$ então $D_{ent}^{17}(g) > D_{ent}^{16}(g)$ e $D_{rep}^{17}(g) > D_{atr}^{16}(g)$ com isso na geografia a quantidade de domicílios atriados foi reposta e ainda houve excedente. Essa geografia pode doar domicílios para outras geografias, portanto ela é uma DOADORA 1 - (DD1);
- S2)** Se $\Delta = 0$ então $D_{ent}^{17}(g) = D_{ent}^{16}(g)$ e $D_{rep}^{17}(g) = D_{atr}^{16}(g)$ com isso na geografia a quantidade de domicílios atriados foi reposta de forma exata, portanto ela é NEUTRA - (NEU);
- S3)** Se $\Delta < 0$ então $D_{ent}^{17}(g) < D_{ent}^{16}(g)$ e $D_{rep}^{17}(g) < D_{atr}^{16}(g)$ com isso na geografia a quantidade de domicílios atriados não foi reposta de forma exata, portanto ela é a RECEPTORA - (REC).

Cenário 2 - $D_{ent}^{16}(g) = 0$

O cenário 2 inicia com a premissa de que não houve domicílios entrevistados na determinada geografia. Então, pode-se ter os dois cenários a seguir:

- S4)** $D_{ent}^{17}(g) = 0$. A primeira expressão indica que a geografia não participou da PCSVDF-Mulher em qualquer uma das duas ondas - (N.VIS);
- S5)** $D_{ent}^{17}(g) > 0$. Isso implica que $\Delta = D_{ent}^{17}(g) = D_{rep}^{17}(g) > 0$. Isso significa que a geografia participou apenas da onda de 2017 da PCSVDF-Mulher e, portanto, a geografia tem domicílios apenas de reposição. Nesse sentido, essa geografia pode ser classificada como uma DOADORA 2 - (DD2).

Tabela 23 – Tabela resumo das classificações

Classificação			Descrição
DD1	$D_{ent}^{16}(g) > 0$	$D_{ent}^{17}(g) > D_{ent}^{16}(g)$	DOADORA
NEU		$D_{ent}^{17}(g) = D_{ent}^{16}(g)$	NEUTRO
REC		$D_{ent}^{17}(g) < D_{ent}^{16}(g)$	RECEPTORA
N.VIS	$D_{ent}^{16}(g) = 0$	$D_{ent}^{17}(g) = 0$	NÃO VISITADA (ONDA 1 E 2)
DD2		$D_{ent}^{17}(g) > 0$	DOADORA

Fonte: Elaborado pelo autor.

Reposição dentro do domicílio

Tabela 24 – Quantidade de domicílios, segundo o município, em que ocorreu reposição dentro do domicílio

Cidade	W2016	ATT_fora	ATT_dent	ATT	PAREA	REP_fora	W2017
Aracaju/SE	986	506	88	594	392	521	1001
Fortaleza/CE	1172	473	31	504	668	534	1233
João Pessoa/PB	1072	599	73	672	400	522	995
Maceio/AL	943	469	36	505	438	599	1073
Natal/RN	1052	582	106	688	364	621	1091
Recife/PE	1245	301	332	633	612	268	1212
Salvador/BA	1104	506	53	559	545	589	1187
São Luís/MA	1082	504	112	616	466	519	1097
Teresina/PI	962	428	103	531	431	422	956
Total	9618	4368	934	5302	4316	4595	9845

Fonte: Extraído de Andre e Carvalho (2023)

Notas: W2016 - entrevistas em 2016, ATT_fora - atrito fora do domicílio, ATT_dent - atrito dentro do domicílio, ATT - domicílios atritados
PAREA - domicílios pareados, REP_fora - reposição fora do domicílio, W2017 - entrevistas em 2017

A Tabela 24 oriunda de Andre e Carvalho (2023) mostra que dos 5529 domicílios definidos como reposição na onda 2 (REP_fora + ATT_dentro), 934 são domicílios classificados como “Domicílio da listagem” na pergunta tipo de entrevista (onda 2). Com isso, deduz-se que essas reposições ocorrem dentro do domicílio, pois os domicílios são de reposição e já estavam na listagem definida, pois foram visitados em 2016. Nesse caso, avalia-se definir os setores censitários com esse fenômeno de uma outra forma. A Tabela 25 indica o quantitativo de reposições dentro do domicílio de acordo com a cidade em que foi realizada a pesquisa, assim como o percentual referente ao total de reposições dentro do domicílio em cada município com relação ao total.

Tabela 25 – Quantidade de domicílios, segundo o município, em que ocorreu reposição dentro do domicílio

Município	Reposição dentro do domicílio	%
Maceió-AL	36	3,8544
Salvador-BA	53	5,6745
Fortaleza-CE	31	3,3191
São Luís-MA	112	11,9914
João Pessoa-PB	73	7,8158
Recife-PE	332	35,5460
Teresina-PI	103	11,0278
Natal-RN	106	11,3490
Aracajú-SE	88	9,4218
Total	934	

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

Um total de 934 substituições foram realizadas dentro do próprio domicílio da entrevistada que atritou. Deste total, os municípios de Recife-PE, São Luís-MA, Teresina-PI e Natal-RN, concentram aproximadamente 70% destes domicílios, o que indica que, nessas residências, o protocolo de substituição que preconiza a 1ª tentativa ser dentro do domicílio foi de certa forma efetivo. Em contrapartida, a cidade de Fortaleza-CE e Maceió-AL tiveram uma pouca representatividade de reposições dentro do domicílio, sendo a soma de ambas inferior a 8% do total, como mostrado na Tabela 25

3.6.2 *Teste Join Count Múltiplo*

A geografia principal da PCSVDF-Mulher são os setores censitários e para efeitos de análise dos quantitativos de setores, a apresentação dos resultados em forma de tabela seria extensa, cansativa e ineficaz, pois os setores são as menores unidades territoriais e o quantitativo de linhas nas tabelas iria ultrapassar 100 unidades. Por isso, foi elaborado uma série de figuras e mapas para que essa análise possa ser feita de forma concisa.

A classificação que tem a maior frequência é a de setores censitários não visitados. Em seguida, aparece a classificação REC, que são setores que tiveram mais domicílios válidos no ano de 2016 do que em 2017 e com isso são chamados de potenciais receptores. As classificações DD1 e DD2, DOADORES, aparecem logo em seguida, e percebe-se que somadas as quantidade DD1 e DD2 o valor é aproximado da quantidade REC.

De acordo com o protocolo da PCSVDF-Mulher, as reposições deveriam ser realizadas primeiramente dentro do domicílio, e caso não conseguisse, a entrevistada deveria realizá-lo dentro do setor censitário. Esse protocolo gera um processo com menos custos para a empresa, tendo em vista que a locomoção para repôr esse atrito amostral se dá de forma mais próxima. Entretanto, algumas vezes as reposições são realizadas não dentro do setor censitário e sim em setores próximos.

Com isso, será realizada uma análise visual se os setores classificados como REC fazem fronteiras com setores DD1 e DD2, além de um teste join count, que é uma forma de contabilizar os setores, onde as relações de vizinhança são definidas. Dadas as contagens globais em cada categoria, as contagens e variações esperadas são calculadas e calculada uma estatística para definir se existe uma dependência espacial entre as categorias ou não.

O teste join count tem em sua essência o objetivo de testar se duas variáveis qualitativas binárias são correlacionadas espacialmente, em que seus dados podem ser mapeados em duas

cores. Upton e Fingleton (1985) adaptaram esse teste para contagens de junções multicoloridas.

Quando a variável de interesse é categórica, o grau de associação local é medido analisando as estatísticas de contagem de vizinhos (junções). Essas categorias podem ser binárias ou não. No caso do presente trabalho, o teste será realizado de acordo com as cinco classificações definidas na linguagem (REC, DD1, DD2, NEU e N.VIS).

As Figuras 11 a 19 representam os mapas das cidades de acordo com os setores censitários, e são coloridos com cores que representam a classificação do setor censitário oriunda da linguagem, definido na seção *Definindo uma linguagem operacional*. A função join count contabiliza a quantidade de vizinhos que possuem a mesma cor e cores diferentes.

Dadas as contagens gerais em cada cor, as contagens e variações esperadas são calculadas e um valor z é relatado. Como existem cinco níveis (REC, DD1, DD2, NEU e N.VIS), a saída é uma matriz de resultados estimados e as contagens de vizinhos são de acordo com as combinações da classificação, seja da mesma cor ou de cores diferentes. O valor z é obtido da maneira usual, dividindo a diferença entre as contagens de junção observadas e esperadas pela raiz quadrada da variância.

A hipótese nula do teste explora se os valores da variável nominal categórica estão distribuídos de forma aleatória espacialmente e, com isso, rejeita-se a hipótese nula quando o número de vizinhos/junções é maior que o seu valor esperado. Esse teste fornece evidência sobre o comportamento efetivo dos responsáveis pela coleta em campo. Em outras palavras, supondo-se que a empresa tenha seguido as determinações do Termo de Referência, a reposição da mulher atirada deve ser feita prioritariamente dentro do domicílio, caso não consiga será feito em domicílios dentro do setor censitário, e por fim em setores próximos.

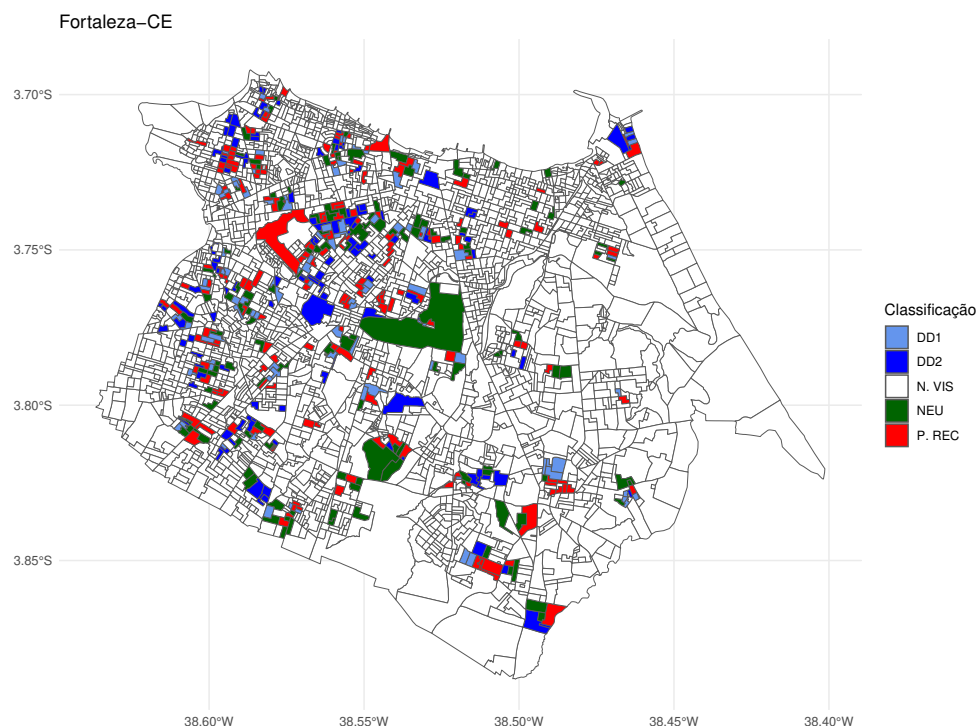
A Tabela 26 fornece os resultados do teste *join count* para os diferentes municípios que participaram da PCSVDF-Mulher. Para fins de observação, só foram disponibilizadas na tabelas as contagens das junções REC:DD1, REC:DD2 e REC:NEU, tendo em vista que o interesse do trabalho é avaliar se existe dependência espacial entre setores classificados de acordo com essas categorias.

Tabela 26 – Teste join count múltiplo para os municípios.

	Joincount	Esperado	Variância	z-value
Fortaleza-CE				
REC:DD1	105,00	32,62	32,02	12,79***
REC:NEU	145,00	47,32	46,30	14,36***
REC:DD2	93,00	42,30	41,43	7,88***
João Pessoa-PB				
REC:DD1	111,00	69,24	64,90	5,18**
REC:NEU	115,00	71,14	66,62	5,37**
REC:DD2	103,00	80,63	75,19	2,58**
Teresina-PI				
REC:DD1	112,00	43,63	43,61	10,35***
REC:NEU	99,00	47,09	47,03	7,57***
REC:DD2	53,00	34,62	34,68	3,12**
São Luís-MA				
REC:DD1	131,00	64,17	62,94	8,42***
REC:NEU	88,00	46,35	45,78	6,16**
REC:DD2	64,00	60,61	59,53	0,44
Maceió-AL				
REC:DD1	108,00	52,27	52,99	7,66***
REC:NEU	126,00	71,49	72,21	6,42**
REC:DD2	90,00	94,55	95,08	-0,47
Natal-RN				
REC:DD1	120,00	61,10	58,18	7,72***
REC:NEU	101,00	54,63	52,19	6,42**
REC:DD2	101,00	62,54	59,51	4,99**
Recife-PE				
REC:DD1	95,00	34,98	33,89	10,30***
REC:NEU	131,00	67,94	65,10	7,81***
REC:DD2	79,00	46,64	45,01	4,82**
Salvador-BA				
REC:DD1	96,00	29,13	30,74	12,06***
REC:NEU	177,00	53,84	57,25	16,28***
REC:DD2	105,00	71,17	76,08	3,88**
Aracaju-SE				
REC:DD1	131,00	64,17	62,94	8,42***
REC:NEU	88,00	46,35	45,78	6,16**
REC:DD2	64,00	60,61	59,53	0,44

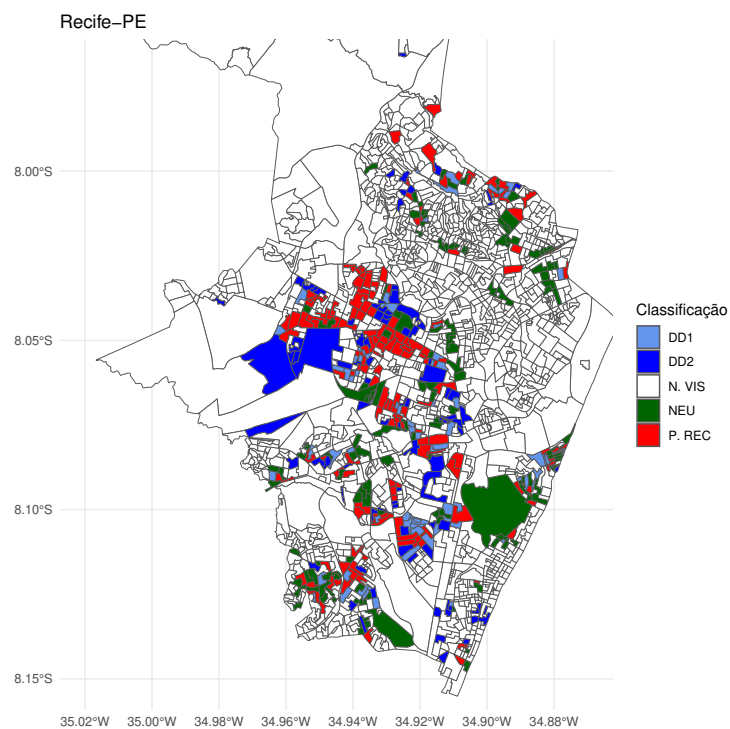
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

Figura 11 – Mapa setores censitários Fortaleza-CE de acordo com a classificação do setor



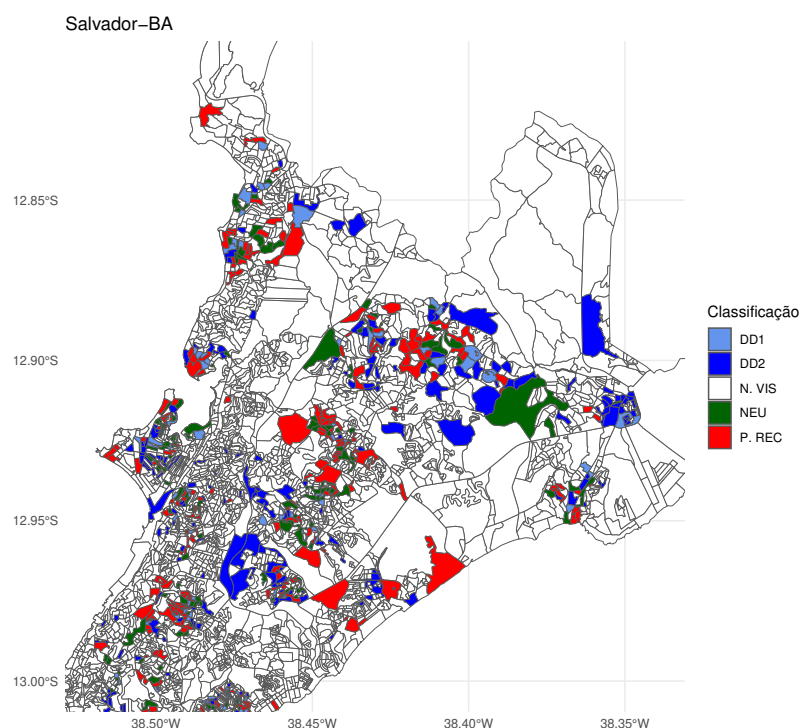
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 12 – Mapa setores censitários Recife-PE de acordo com a classificação do setor



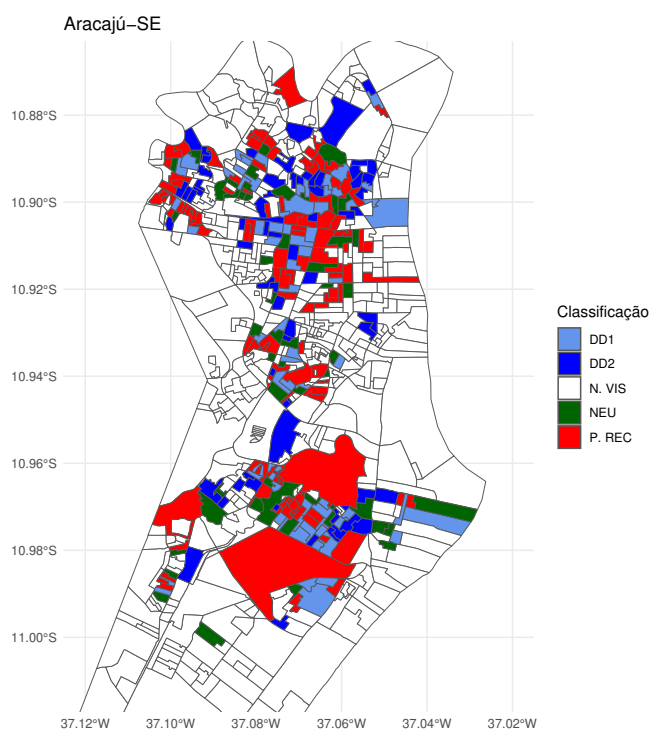
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 13 – Mapa setores censitários Salvador-BA de acordo com a classificação do setor



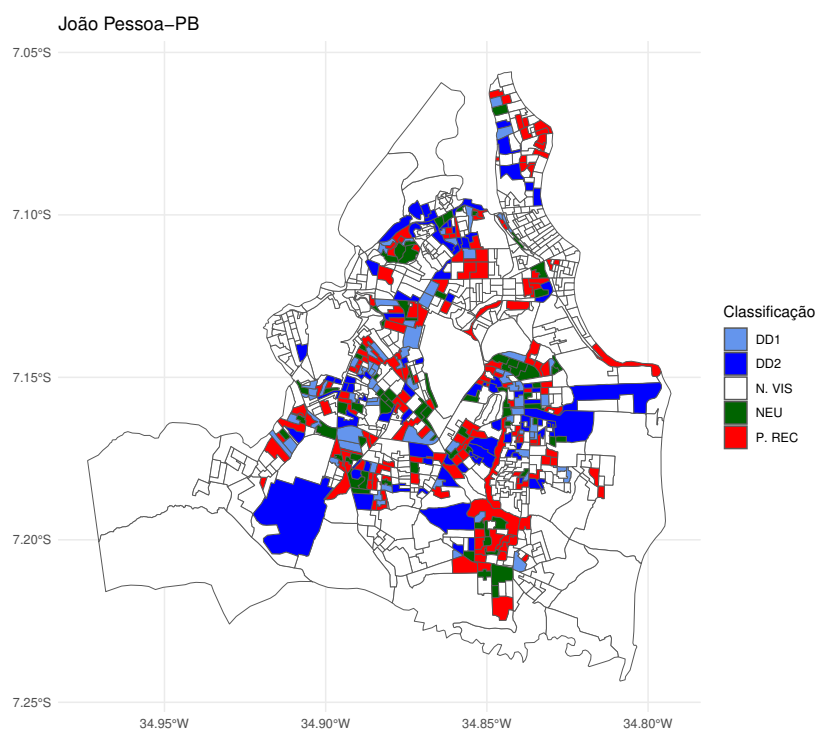
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 14 – Mapa setores censitários Aracaju-SE de acordo com a classificação do setor



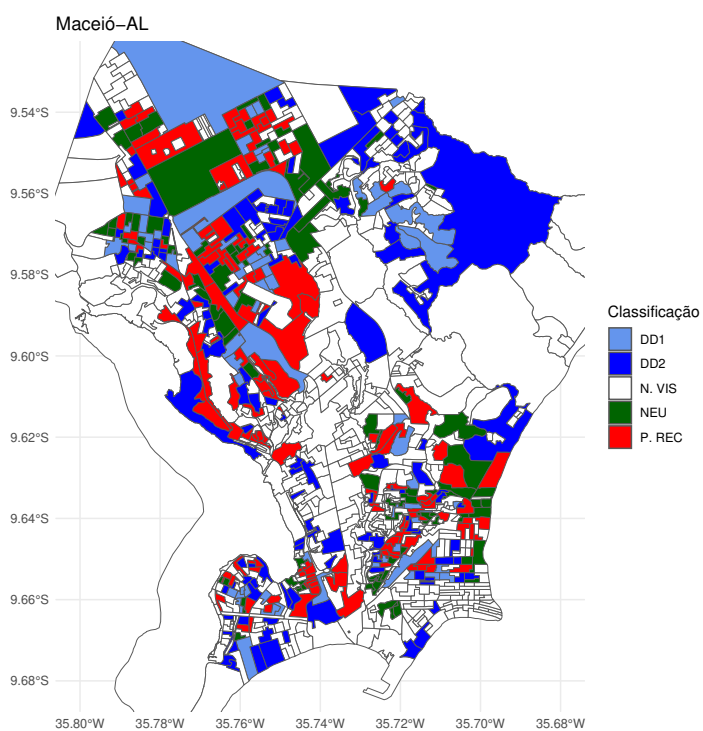
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 15 – Mapa setores censitários João Pessoa-PB de acordo com a classificação do setor



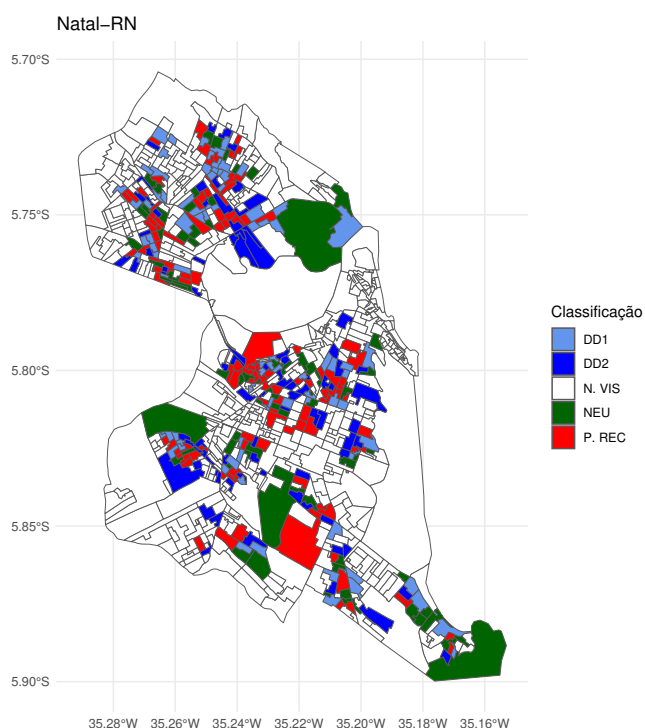
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 16 – Mapa setores censitários Maceió-AL de acordo com a classificação do setor



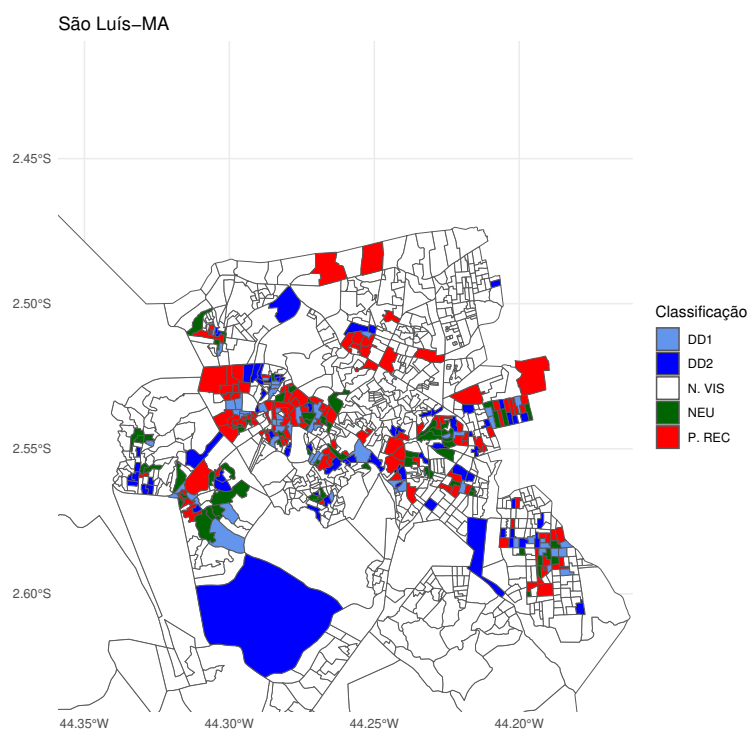
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 17 – Mapa setores censitários Natal-RN de acordo com a classificação do setor



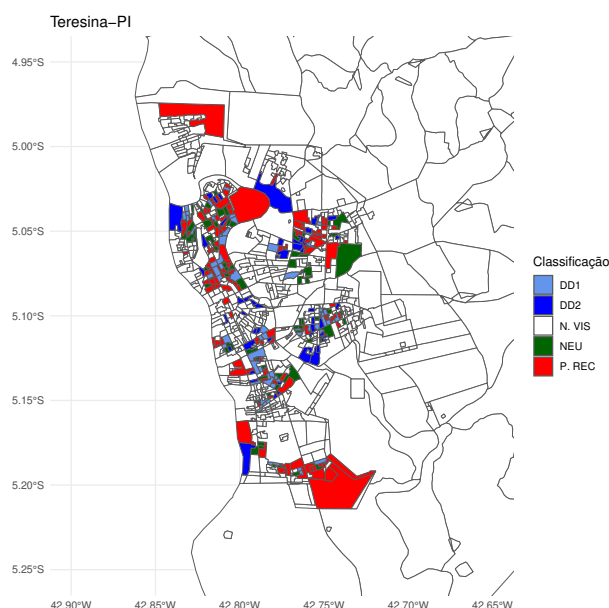
Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 18 – Mapa setores censitários São Luís-MA de acordo com a classificação do setor



Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Figura 19 – Mapa setores censitários Teresina-PI de acordo com a classificação do setor



Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher

A análise espacial e o teste de join count indicam uma dependência entre as classificações REC (RECEPTORA) e DD1/DD2 (DOADORAS). A tabela 26 mostra a quantidade de vizinhos (real e esperado) de acordo com a classificação. Por exemplo, na cidade de Fortaleza-CE era esperado 32,62 vizinhanças com a classificação REC:DD1 e foram encontrados 105 vizinhanças nessa classificação.

Vale destacar que a maior parcela dos setores censitários dados como DOADORES x RECEPTORES é na relação REC:DD1. Esse fator evidencia que a entrevistadora irá procurar repôr o atrito em setores que já haviam sido entrevistados no ano de 2016 na PCSVDF-Mulher.

Foi descoberto um padrão regular de substituição. Vale salientar que a hipótese nula testada é que a escolha é aleatória e o teste sugere uma rejeição a favor de um comportamento de reposições mais próximas e isso mostra um comportamento otimizador por conta da empresa contratada no sentido de fazer com que suas entrevistadoras minimizem a distância percorrida. Em apenas três cenários o teste não foi estatisticamente significativo.

A conclusão sugere que a empresa contratada conseguiu repôr os atritos amostrais, seguindo o protocolo que é a reposição dentro do domicílio e, caso não consiga, dentro do setor. No entanto, quando não foi possível conseguir uma dessas duas alternativas, a entrevistadora teve que sair do setor censitário e, para minimizar os seus custos, privilegiou nas suas movimentações e suas rotas setores próximos. Isso dá uma economia maior no momento da pesquisa.

3.7 Considerações Finais

Este capítulo teve como objetivo verificar se o protocolo de substituição da PCSVDF-Mulher foi seguido pela empresa contratada e se o mesmo foi eficaz para repôr a amostra original. Para isso, foram utilizadas três abordagens com o recorte dos dados das mulheres que atritaram nos anos de 2016 e 2017 e as que foram amostra de reposição em 2017.

A primeira análise se deu de forma exploratória, contabilizando as quantidades de amostras coletadas em 2016 e em 2017 e fazendo um comparativo de acordo com quatro cenários - na pesquisa completa, no município da entrevista, no bairro e no setor censitário, dessa forma é possível verificar uma das premissas do protocolo de substituição que seria substituir a mulher atritada dentro do domicílio, caso não obtivesse êxito, essa substituição se daria dentro do setor censitário. A partir do resultado encontrado verificou-se que a substituição foi efetiva, quantitativamente, até o nível de bairro.

Apesar da substituição ocorrer na forma numérica - sendo o número de substituições igual ou maiores que o número de atritados - a segunda análise examinou se essa substituição ocorreu sem causar viés nas variáveis finais de interesse.

Uma análise preliminar entre as amostras da PNAD-2017 e da PCSVDF-Mulher (completa) e PCSVDF-Mulher (reposição) mostraram que as variáveis analisadas possuem diferença entre suas distribuições, de acordo com o teste qui quadrado de homogeneidade. As variáveis idade, escolaridade e raça se mostraram diferentes em todos os cenários (pesquisa e nos municípios), mostrando a importância dos pesos amostrais para calcular medidas estatísticas de forma representativa e precisa.

Por fim, a análise entre as amostras de atrito e reposição mostrou que a maior parte das variáveis possuem igualdade entre suas médias e/ou variâncias, distribuições e proporções. Foram utilizados testes para os diferentes tipos de variáveis, sejam elas numéricas, categóricas ou dummies. Com isso, o protocolo de substituição conseguiu manter a representatividade estatística da amostra original.

A partir da primeira análise, viu-se a necessidade de se criar uma linguagem para definir os setores censitários a partir da quantidade de domicílios visitados em 2016 e 2017. Eles foram divididos em DOADORES, NEUTROS, RECEPTORES e NÃO VISITADOS, e analisados espacialmente com o intuito de verificar se existe uma dependência espacial entre as classificações.

A análise teve o objetivo de verificar, de forma espacial, se existiu dependência entre

as classificações, ou seja, era esperado que setores censitários que tiveram déficit de domicílio de um ano para outro (RECEPTORA) estivessem mais próximos de setores que eram classificados com DOADORES. Utilizou-se o teste join count múltiplo que visa testar se as classificações estão dispostas de forma aleatória espacialmente versus existir alguma dependência entre setores vizinhos. O resultado do teste mostra evidências de que os setores não estão dispostos de forma aleatória, além disso, o quantitativo contabilizado de setores classificados como POTENCIAIS RECEPTORES e DOADORES foi superior ao quantitativo esperado, concluindo assim que a empresa contratada após exaustada as tentativas de repor as mulheres atritadas dentro do próprio domicílio ou dentro do setor censitário ela tende a substituir em setores vizinhos ou próximos.

É importante observar que os resultados encontrados no presente estudo são importantes para preencher a lacuna de estudos que visam o estudo sobre protocolos de substituição em pesquisas longitudinais. Além disso, tendo em vista que o protocolo utilizado foi de certa forma mais simples e menos oneroso que outros de trabalhos em países desenvolvidos e mesmo assim manteve o seu objetivo final que era ser eficaz perante a substituição.

Como sugestão para pesquisas futuras, o mesmo estudo poderá ser executado para as ondas futuras da PCSVDF-Mulher, levando em consideração que já foram realizadas entrevistas em 2019, 2021 e 2023, além do fato de na onda 3 identificar qual reposição está ligada a tal atrito. Além de verificar se a pandemia da COVID-19 poderia ter influência na mudança de perfil da coleta de dados dos anos subsequentes a ela.

4 CAPÍTULO 03 - MODELANDO INCONSISTÊNCIAS LONGITUDINAIS NAS RESPOSTAS SOBRE VITIMIZAÇÃO POR VIOLÊNCIA DOMÉSTICA.

4.1 Introdução

Violência por parceiro íntimo (VPI), definida como um conjunto de comportamentos de violência física, sexual, psicológica ou perseguição perpetrados pelo parceiro íntimo (atual ou ex-cônjuge, namorado(a), parceiro sexual ou outra pessoa com quem a vítima mantenha ou tenha mantido um relacionamento pessoal próximo) é um importante problema de saúde pública no mundo, Mendonça e Ludermir (2017).

No Brasil, de acordo com dados do Sistema de Agravos e Notificações (SINAN), a taxa de prevalência de VPI no ano de 2022 foi igual a 28,9%. Dados provenientes dos Estados Unidos sugerem que existe um valor semelhante, com até 27% das mulheres jovens com idades compreendidas entre os 18 e os 23 anos, a denunciarem violência por parte de um parceiro Halpern *et al.* (2009). De acordo com a Organização Mundial da Saúde (OMS), no mundo, em 2021, as estimativas sugerem que pelo menos um quarto das mulheres sofreram experiências de violência físicas e/ou sexual por parte do parceiro durante a sua vida.

A mensuração da violência perpetrada pelo parceiro íntimo (VPI) é fundamental tanto para compreender a extensão do problema quanto para agir de forma eficaz na sua erradicação. Esse processo contribui para a promoção de relacionamentos saudáveis e uma sociedade mais justa. Por meio dessa abordagem, é possível identificar a magnitude da violência, planejar políticas públicas, desenvolver estratégias de intervenção, proteger grupos vulneráveis e monitorar a eficácia de programas e ações preventivas.

Mensurar a violência perpetrada pelo parceiro íntimo (VPI) envolve diversos desafios, como o estigma social e a vergonha que contribuem para a subnotificação, o medo de retaliações por parte do agressor, barreiras culturais que legitimam ou normalizam a violência e a falta de confiança na segurança e confidencialidade dos dados coletados. A variabilidade na definição de violência também dificulta a identificação de suas diferentes formas, como física, psicológica, sexual ou econômica. Além disso, há obstáculos metodológicos, como o acesso a grupos vulneráveis, a influência do agressor durante as entrevistas e o impacto emocional sobre as vítimas ao relatar experiências traumáticas. Por fim, restrições de recursos, incluindo treinamento inadequado e falta de financiamento, comprometem ainda mais a precisão e a abrangência da coleta de informações.

Embora alguns estudos tenham investigado quais fatores estão associados aos relatos de VPI, Sylaska e Edwards (2013), há pouco conhecimento sobre a consistência com que, especificamente, mulheres relatam VPI ao longo do tempo. Serão consideradas consistentes os padrões de respostas intertemporais que mantêm uma sequência lógica sem contradições. Respostas inconsistentes, por sua vez, apresentam rupturas nessa lógica. Na seção 4.2.1 - Inconsistência de Resposta: Definição, esses conceitos serão detalhados com maior precisão.

A maior parte dos estudos sobre relatos inconsistentes são decorrentes de estudos com adultos que recordam abusos sexuais na infância [Aalsma *et al.* (2002), Langeland *et al.* (2014), Mckinney *et al.* (2008)]. Estes sugerem que a denúncia inconsistente de abuso é comum e que são necessárias múltiplas avaliações do abuso ao longo do tempo para identificá-lo, Mckinney *et al.* (2008).

Queiroz e Biasoli-Alves (1987) explora métodos qualitativos em pesquisas baseadas em entrevistas, com foco nos desafios relacionados à interpretação das respostas e ao impacto que eventos marcantes podem exercer ao longo do tempo no processo de coleta e análise de dados. Os entrevistados relatam eventos salientes somente anos após a experiência inicial, quando tiveram tempo de processá-lo. Outros, por sua vez, que já denunciaram experiências anteriormente, podem querer deixá-las “para trás” e podem decidir não denunciá-los posteriormente.

O estigma social também pode suscitar vergonha e dissuadir as pessoas a relatarem novamente as suas experiências de VPI. O medo da estigmatização é frequentemente relatado em estudos com mulheres que sofreram VPI. Tendo em vista que a vergonha sentida diante dessas violências sofridas fragiliza a autoimagem dessas mulheres, que se sentem humilhadas por assumir que sofrem violências, Dourado e Noronha (2015).

A divulgação da violência, portanto, é um processo dinâmico que muda ao longo do tempo em resposta a uma série de fatores individuais, relacionais, situacionais e sociais. A investigação longitudinal é essencial para compreender como a notificação de VPI muda durante a vida e se a notificação inconsistente está associada a questões econômicas, psicológicas, sociais, cognitivas ou outras.

Só existe um estudo abordando inconsistência de VPI. O estudo de Loxton *et al.* (2019) que acompanha mulheres australianas utilizando dados de três ondas de pesquisa. Dado o exposto, entende-se ser importante a contribuição desta pesquisa para a literatura sobre o tema, referente a estudos que buscam compreender os motivos que levam à inconsistência do autorrelato de eventos salientes de saúde pública ao longo da vida.

O presente capítulo tem como objetivo analisar o padrão de respostas inconsistentes a respeito de vitimização de violência doméstica nos municípios do Nordeste, a partir de dados oriundos da Pesquisa de Condições Socioeconômicas e Violência Doméstica e Familiar contra a Mulher (PCSVDF-Mulher), durante os anos de 2016 e 2017.

A base de dados da PCSVDF-Mulher permite obter dados socioeconômicos da mulher, dados sobre saúde geral e reprodutiva, experiências com violência doméstica, além de conhecimento sobre normas e sobre a Lei Maria da Penha. Além da violência geral, é possível classificar a violência doméstica em 3 tipos: emocional, física e sexual. Com isso, a inconsistência no relato de experiências, além de ser estudado no âmbito geral, também será analisado de acordo com cada tipo de violência.

Utilizando os dados da PCSVDF-Mulher, foi encontrada uma taxa de inconsistência igual a 18,08% no autorrelato das entrevistadas entre os anos de 2016 e 2017. Na estimativa do modelo logit, algumas variáveis aparecem com maior destaque na influência tanto de forma positiva quanto de forma negativa da inconsistência (mulher ter filhos, o seu local de residência e a mulher ter sofrido algum tipo de violência no último ano). Vale destacar que o percentual de respostas inconsistentes é diferente quando se compara os diferentes tipos de violências, emocional, física e sexual, sendo igual a 17,17%, 12,57% e 6,54%, respectivamente.

Além disso, foi possível diferenciar as sequências de respostas inconsistentes em diferentes padrões, com base no perfil do sequenciamento das respostas dadas. A análise revelou que um dos padrões de inconsistência ocorre com uma frequência aproximadamente duas vezes maior que o outro, comportamento observado de forma consistente em todos os tipos de violência analisados — geral, emocional, física e sexual.

O capítulo é dividido em cinco seções. A segunda seção aborda a literatura teórica sobre o assunto, dividida em diferentes domínios do estudo, Psicologia, Estatística (Desenho da pesquisa), Economia/Econometria, Violência doméstica e Saúde pública. Na terceira seção, é feita uma análise descritiva do processo de inconsistência na PCSVDF-Mulher, apresentando a definição e operacionalização do problema, assim como uma análise de frequência das inconsistências encontradas. A seção quatro apresenta a modelagem econométrica com a descrição dos modelos que serão utilizados na modelagem, assim como as variáveis que estarão presentes nos modelos com uma breve estatística descritiva das mesmas, e os resultados econométricos encontrados. Por fim, a última seção traz as considerações finais do trabalho.

4.2 Referencial Teórico

A inconsistência temporal de um relato de vida é influenciada por efeitos de respostas que podem ocorrer devido a problemas no entendimento da pergunta, na lembrança de informações relevantes, na produção de uma resposta apropriada ou em outros processos mentais. Existem na literatura áreas que elencam esse efeitos, que são capazes de influenciar em relatos inconsistentes dos(as) entrevistados(as). Destacam-se a área da Psicologia, Estatística (desenho de survey), Economia/Econometria e, por fim, direcionado para a área de VPI Violência doméstica e saúde pública.

Antes de iniciar a discussão sobre as perspectivas teóricas, é preciso definir minuciosamente o conceito de inconsistência adotado em nossa análise.

4.2.1 *Inconsistência de resposta: Definição*

O entendimento do princípio da contradição precisa ser determinado antes da definição de uma resposta inconsistente. Esse princípio estabelece que duas afirmações contraditórias não podem ser verdadeiras ou falsas, uma delas é verdadeira enquanto a outra é falsa. Por exemplo, as duas afirmações seguintes não podem ser verdadeiras ao mesmo tempo: “X é Y” e “X não é Y”. Ou é ou não é.

O princípio da não contradição foi formulado primeiramente por Aristóteles, quando afirmou que uma proposição não pode ser verdadeira e falsa ao mesmo tempo. Logo, na lógica clássica, nenhuma afirmação pode ser falsa e verdadeira concomitantemente. A lei do terceiro excluído é um complemento da lei da não contradição. Uma sequência de respostas no questionário da PCSVDF-Mulher é inconsistente se, e somente se, essa sequência dá origem a uma situação onde o "Princípio da Não Contradição" não é obedecido.

Em um ano específico, as respostas sobre violência doméstica, pode assumir três valores: 1- se ela já sofreu algum tipo de violência, 0 - se ela não sofreu nenhum tipo de violência e NA(NOT APPLICABLE). A resposta NA (não aplicável) pode ocorrer por 3 motivos:

1. nunca ter tido parceiro na vida;
2. entrevistada teve algum parceiro na vida, porém

- 2.1 chegou na seção EXPERIÊNCIAS DE VIOLÊNCIA DOMÉSTICA (PARCEIRO ATUAL, EX-PARCEIRO (MAIS RECENTE) OU QUALQUER OUTRO EX-PARCEIRO) e decidiu não responder a seção;

2.2 entrevistada desiste de responder as perguntas no meio da seção EXPERIÊNCIAS DE VIOLÊNCIA DOMÉSTICA (PARCEIRO ATUAL, EX-PARCEIRO (MAIS RECENTE) OU QUALQUER OUTRO EX-PARCEIRO).

A Tabela 46, que encontra-se no APÊNDICE B, indica o limite da evidência empírica, ou seja, são todas as combinações de respostas possíveis que a mulher pode ter. Utilizando os dois anos e os dois períodos de tempo em que são questionados se a mulher já sofreu algum episódio de violência, se esse evento ocorreu ao longo da vida (LV) e deu-se nos últimos 12 meses (<12 meses).

4.2.2 *Psicologia*

Tourangeau *et al.* (2000) examina os processos psicológicos envolvidos na resposta a diferentes tipos de perguntas. Alguns mecanismos são explorados que podem influenciar na consistência da resposta do participante:

- Desejabilidade social;
- Compreensão de perguntas (complexidade cognitiva);
 - Ausência de conhecimento;
- Telescoping.

Desejabilidade social

A desejabilidade social reflete uma propensão por parte das pessoas a dar respostas consideradas como socialmente mais aceitáveis e a negar associação pessoal com opiniões ou comportamentos considerados socialmente negativos.

Uma resposta inconsistente se dá muitas vezes em perguntas sensíveis, que questionam se o entrevistado já violou algum tipo de norma. Isso pressupõe que os entrevistados acreditam que existem normas que definem atitudes e comportamentos desejáveis e estão preocupados com elas, com isso distorcem suas respostas com o intuito de evitar se apresentar de acordo com um aspecto que não seja favorável perante a sociedade.

A inconsistência nos relatos em eventos salientes decorrentes da desejabilidade social podem estar associadas a diferentes processos e ter diferentes origens, como características pessoais (por exemplo, estados de humor, autopercepções), segundo Richman *et al.* (1999).

Compreensão de perguntas (complexidade cognitiva)

A compreensão de uma pergunta é dividida em duas partes, uma obrigatória e outra opcional. Consiste em representações do que o entrevistado acabou de ler ou ouvir, mas

são diferentes em relação ao conteúdo. A parte obrigatória é uma especificação da estrutura gramatical e lógica da sentença. A opcional consiste em inferências que o entrevistado extrai da sentença, em conjunto com outros conhecimentos que ele já possui.

Em muitas perguntas, os respondentes seguem a sequência lógica de compreender primeiro o item, recuperar os fatos relevantes, depois fazer o julgamento necessário e, finalmente, relatar. Entretanto, de acordo com Anderson (1983), pode não ocorrer necessariamente nesta ordem, a recuperação pode começar antes que a compreensão seja concluída; o próprio ato de entender as palavras-chave na interpretação de uma pergunta pode desencadear na recuperação de memórias.

Além disso, vale destacar que os itens anteriores podem mudar a maneira como os entrevistados interpretam as perguntas posteriores, que considerações eles recuperam ao formular suas respostas, quais padrões ou normas eles aplicam no julgamento da questão e como relatam suas respostas. Ademais, diferentes mecanismos que produzem efeitos de contexto influenciam as respostas de maneiras diferentes. Os itens anteriores, às vezes, influenciam as respostas posteriores na direção da consistência, mas, por vezes, têm o efeito oposto, produzindo aparentes inconsistências. Para complicar ainda mais, o contexto de um item pode alterar a direção geral das respostas ou alterar as correlações entre respostas para itens diferentes.

Ausência de conhecimento

A ausência de conhecimento descreve uma pessoa, em seu pleno estado de consciência ou de conhecimento, que deliberadamente ignoram ou desconsideram informações ou fatos importantes, ou indivíduos que desconhecem informações ou fatos importantes.

Por exemplo, sendo “S” um sujeito, “S sabe que Salvador é a capital da Bahia”, entretanto ao perguntar para “S” qual a capital de Samoa ele não sabe, simplesmente por ausência de conhecimento. Ele entende a pergunta, sabe o conceito de capital, porém não sabe a capital de Samoa.

Telescoping

Na psicologia cognitiva, o efeito telescópico (ou viés telescópico) descreve um distúrbio na percepção temporal dos eventos. Nesse fenômeno, as pessoas tendem a perceber acontecimentos recentes como se fossem mais distantes no tempo e eventos passados como mais próximos do que realmente são. A expansão temporal, em que eventos distantes parecem mais recentes, é denominada telescopia direta. Jackson *et al.* (2016) verificam o efeito *telescoping*, relatando que os participantes afirmam incorretamente eventos passados mais distantes como

ocorrendo mais recentemente.

Como as pessoas têm dificuldade em lembrar datas exatas [Friedman (2013), Thompson *et al.* (1996), Wagenaar (1986)], elas podem relatar eventos que ocorreram antes do período de referência especificado. As pessoas podem fazer inferências incorretas sobre o tempo com base na acessibilidade da memória ou com base no destaque dos eventos (saliência). Eventos mais salientes tendem a ser lembrados e ditos ser mais recentes quando comparado a eventos menos salientes que tendem a ser lembrados como mais antigos.

Norris (1992) relatou sobre a inconsistência de eventos traumáticos com relação a idade do entrevistado e o período que o evento ocorreu. O autor sugeriu que os benefícios da recordação devido à idade como uma possível explicação. Os efeitos da inconsistência foram mais pronunciados nas mulheres mais jovens do que nas mulheres adultas de meia idade, e ocorreram mais inconsistências para o intervalo de 6 anos do que para o intervalo de 3 anos.

Além desses três mecanismos Pachana *et al.* (2011) elenca mais dois, que são: Saliência e memória.

Saliência

Os trabalhos de Casey *et al.* (1967) e Glickman Michael Hubbard e Valciukas (1990) citam que os eventos que são caracterizados como mais importantes ou significativos na vida da pessoa tendem a ser reportados de maneira mais consistente, principalmente ao longo do tempo.

Os resultados apresentam uma queda na ocorrência de eventos durante os primeiros doze meses. Alguns eventos altamente salientes (morte do cônjuge, casamento, nascimento do filho) não demonstram essencialmente nenhuma diminuição no seu relato, enquanto outros (morte de amigos, doença na família) apresentam diminuição imediata no relato.

Além disso, o autor mostra que em uma pesquisa feita com estudantes universitários para prever eventos que provavelmente serão narrados por um adulto, com eventos emocionais mais salientes, percebe-se uma maior frequência na lembrança de eventos positivos, mas não negativos.

Memória

Raphael *et al.* (1991) descobriram que, à medida que o tempo decorrido do evento aumentava, a chance de eventos serem relatos pelos entrevistados diminuía.

Existem outros problemas com a precisão dos acontecimentos de vida auto-relatados ao longo do tempo. Por exemplo, Raphael *et al.* (1991) perceberam que apenas aproximadamente 25% dos eventos mencionados pelos informantes durante entrevistas mensais ao longo de 10

meses foram relatados retrospectivamente na conclusão do seu estudo. O preditor mais forte do relato de eventos foi há quanto tempo o evento ocorreu: à medida que o tempo desde o evento aumentava, a chance de ele ter sido relatado diminuía.

4.2.3 Estatística (*Desenho de surveys*)

De acordo com Tourangeau *et al.* (2000), os desenhos de surveys são de extrema importância para o êxito em uma resposta consistente das pesquisas, mesmo que sejam pela redação das perguntas, em que os entrevistados podem interpretar mal, até as que foram bem formuladas e, quando isso acontece, a pergunta respondida pode não ser a que o pesquisador pretendia fazer. Além disso, podem ser consideradas até situações mais amplas como em uma relação, no momento do contato da pesquisa, entre entrevistado e entrevistador.

As perguntas longas ou complexas podem exceder a capacidade dos respondentes de processá-las, resultando em interpretações erradas. O significado (ou semântica) da pergunta pode confundir os respondentes se eles não entenderem termos vagos, desconhecidos ou ambíguos ou se forem enganados por pressupostos inaplicáveis.

O autor ainda destaca que a colaboração entre o entrevistador e o entrevistado pode convergir para uma nova interpretação de uma palavra ou frase durante o processo de resposta a perguntas, mas nem tudo acontece. Às vezes, esses significados estarão incorretos.

Clark e Schober (1992) e Schwarz (1996) sugerem que aspectos das perguntas da pesquisa podem produzir resultados semelhantes e, portanto, afetar as respostas. A sobreposição de significado entre os itens da pesquisa, o sequenciamento dos itens, o intervalo numérico das alternativas de resposta e o intervalo e a rotulagem das escalas de classificação podem moldar o caminho dos entrevistados nas perguntas.

4.2.4 Economia/Econometria

Dentro da economia/econometria um dos mecanismos encontrados é a mentira (racional) na resposta de uma pesquisa.

Mentira (racional)

Em muitas situações, é possível que uma pessoa possa mentir sobre suas informações privadas. Embora o ato da mentira seja comum, evidências convincentes mostram que as pessoas, às vezes, evitam contar mentiras que aumentariam suas recompensas. Gneezy *et al.* (2018) Um comportamento honesto é uma resposta otimizada a incentivos materiais; as pessoas são

honestas porque a desonestidade pode levar à algum tipo de punição.

Os autores formalizam um aspecto importante dos custos intrínsecos da mentira – como o tamanho da mentira afeta a decisão de mentir. São discutidos três tipos possíveis de custos mentirosos: um custo relacionado com a distância entre o resultado verdadeiro e o que é relatado; um custo relacionado aos ganhos monetários gerados pela mentira; e um custo associado à probabilidade de uma declaração ser considerada desonesta.

Em uma pesquisa sobre o consumo de drogas ilícitas, a precisão das informações coletadas depende de respostas confiáveis e precisas dos entrevistados. Pode acontecer, no entanto, que estes tenham incentivo para informar erradamente o seu consumo de drogas, uma vez que as drogas, em particular as de natureza ilícita, estão associadas a riscos legais, bem como ao estigma ou à desaprovação social, Greene *et al.* (2017). Tais respostas inconsistentes podem levar à classificação errada da informação nos dados das pesquisas, o que pode mascarar a incidência de tais comportamentos e levar a estimativas tendenciosas e inconsistentes nas análises estatísticas.

Glickman Michael Hubbard e Valciukas (1990) comparam a frequência de eventos de vida relatados nos dois intervalos de 6 meses para avaliar a inconsistência em relatos de eventos. A hipótese nula de que a frequência relatada de eventos no intervalo mais recente é igual à frequência relatada de eventos no intervalo anterior de 6 meses será testada. Menos eventos são relatados no intervalo mais distante da aplicação do questionário, sugerindo uma redução na recordação dos eventos ao longo do tempo.

Foi realizada uma análise de regressão, em que idade, número de eventos relatados no intervalo mais próximo à administração do questionário e o número de eventos associados à mudança de vida acentuada relatada no intervalo mais recente foram incluídos como preditores. A variável resposta foi a razão obtida dividindo o número de eventos relatados no intervalo mais distante pelo número de eventos no intervalo mais recente.

Ainda do trabalho de Glickman Michael Hubbard e Valciukas (1990), o primeiro modelo, somente com a idade, revelou uma significância entre esta última e a proporção de eventos relatados nos dois intervalos de tempo. No segundo modelo, a idade e o número de eventos relatados no intervalo mais próximo à administração do questionário foram significantes à proporção de eventos relatados. Os resultados do terceiro modelo testado mostraram efeito marginal para a idade e nenhum efeito para o número de eventos associados à mudança de vida acentuada relatada.

Meyer *et al.* (2020) mostram que tanto os relatórios falsos negativos quanto os falsos positivos estão associados a uma variedade de características do domicílio e das entrevistas, incluindo rendimento e raça. Do ponto de vista metodológico, isto mostra que os erros das pesquisas não são aleatórios. Assim, os erros também levam a vieses complicados, direcionando a respostas inconsistentes.

Os autores ainda afirmam que, se ignorados, os relatos inconsistentes levam potencialmente a estimativas imprecisas da prevalência de tais comportamentos e, em última análise, pode levar alguém a questionar a validade de quaisquer conclusões tiradas com base nestas, o que por sua vez levanta preocupações quanto ao ponto em que esses dados dados são realmente úteis para os decisores políticos.

4.2.5 Violência doméstica e saúde pública

O Estudo Longitudinal Australiano sobre Saúde da Mulher (ALSWH) investiga fatores relacionados à saúde e bem-estar das australianas, referentes a mulheres de dois grupos, nascidas em 1973-1978 e 1946-1951. As primeiras ondas foram conduzidas em 1996 e a onda 2 foi realizada em 1998 para o grupo de 1946 a 1951 e em 2000 para o grupo de 1973 a 1978. Todas as ondas subsequentes foram conduzidas com 3 anos de intervalo.

Loxton *et al.* (2019) utilizam dados da Pesquisa 5 (2009), Pesquisa 6 (2012) e Pesquisa 7 (2015). A quantidade e o tipo de perguntas feitas às mulheres também influenciaram nas estatísticas de prevalência. Foi encontrado que, ao se fazer uma única pergunta sobre violência, gera menos relatos do que perguntar às mesmas mulheres sobre violência com um instrumento mais abrangente. Experiências caracterizadas como “violentas” ou “abusivas” podem levar a menos respostas do que itens que perguntam sobre comportamentos específicos.

O objetivo do estudo está em determinar por que há inconsistência na resposta a itens de violência. A consistência do autorrelato de violência doméstica foi avaliada por respostas à Escala Comunitária de Abuso Composto. A escala foi usada pela primeira vez na Pesquisa 5 (2009) e depois na Pesquisa 6 (2012) e na Pesquisa 7 (2015). As respostas foram classificadas como "Inconsistentes" quando os participantes responderam "Sim" a um item no CCAS e depois responderam "Não" ao mesmo item em uma pesquisa subsequente.

Mulheres que eram consideradas inconsistentes eram menos propensas a viver com crianças, em comparação a mulheres que eram consistentes. Além disso, mulheres inconsistentes eram um pouco menos propensas a ter dificuldade em gerenciar sua renda disponível do que

mulheres consistentes. De todas as subescalas, o abuso emocional gerou a maior porcentagem de respostas inconsistentes, seguidas de abuso físico, assédio e abuso sexual.

Os resultados quantitativos vão de acordo com o estudo qualitativo, em as mulheres questionaram se o abuso emocional seria visto como violento e o tipo, a frequência e a gravidade do abuso estavam abertos à interpretação e reinterpretação ao longo do tempo, levando a atrasos e relatos inconsistentes de abuso. Isso está de acordo com pesquisas anteriores, que mostraram que o abuso emocional, como controlar o comportamento, geralmente não é identificado como abuso. [Webster *et al.* (2018) e Catallozzi *et al.* (2011)]

Vale destacar que, quando observados os itens em separado, as experiências de violência doméstica que eram mais obviamente violentas ou abusivas por natureza tinham menos probabilidade de serem relatadas inconsistentemente.

O estudo de Funch e Marshall (1984) se baseia em uma entrevista de eventos estressantes da vida em uma amostra de 3692 indivíduos, maiores de 18 anos, que tem como objetivo principal responder aos seguintes questionamentos: (1). Qual é a taxa de queda de recordação de eventos de vida em geral - como ela varia ao longo do tempo?; (2) eventos diferentes têm padrões diferentes de queda?; e (3) as características demográficas dos entrevistados (idade, sexo, educação, renda, estado civil) afetam as taxas de desistência?

Muitos trabalhos estudam os efeitos dos eventos da vida no humor, bem como o impacto dos estados de humor atuais e passados no relato de eventos da vida. Vale ressaltar que o humor não é estável ao longo do tempo Pachana *et al.* (2011)

Nas pesquisas, as mulheres receberam uma lista de itens de eventos da vida e foram solicitadas a indicar quais havia experimentado dentro de um período de tempo especificado. Para calcular o humor da mulher, foi utilizada a subescala MH do SF-36, que varia de 0 a 100, classificada em cada onda como baixo MH (<53) ou alto MH (≥ 53).

Na onda 1, as participantes foram questionadas: “Nos últimos 12 meses, você experimentou algum dos seguintes eventos?” Nas próximas ondas, também foram perguntadas se haviam experimentado os eventos há mais de 12 meses. As respostas à primeira pergunta foram usadas para definir a incidência de 12 meses para o item a cada onda. As respostas a ambos os itens foram usadas para calcular a incidência acumulada ao longo da vida das ondas 2, 3 e 4.

Itens que são difíceis com relação à sua interpretação (Assédio angustiante no trabalho) têm uma maior probabilidade de serem relatados de maneira inconsistente do que itens

menos ambíguos e mais objetivos (Pais se divorciaram/se casaram novamente). O mesmo ocorre para eventos salientes, que tendem a ser relatados com mais consistência do que eventos menos importantes.

A Tabela 27 apresenta um resumo detalhado de cada mecanismo que pode explicar um relato inconsistente, incluindo as explicações teóricas propostas e os principais autores responsáveis por essas contribuições.

Tabela 27 – Resumo dos mecanismos de inconsistência e teorias

Mecanismo	Autores	Teoria
Desejabilidade social	Richman <i>et al.</i> (1999) e Tourangeau <i>et al.</i> (2000)	Respostas mais aceitáveis ao julgamento da sociedade e negar opiniões ou comportamentos negativos;
Cognição	Anderson (1983) e Tourangeau <i>et al.</i> (2000)	Dificuldade de entender a pergunta, seja por aspectos gramaticais, comunicação ou por influência de perguntas anteriores;
Ausência de conhecimento	Tourangeau <i>et al.</i> (2000)	Entende a pergunta, mas pela sua falta de conhecimento sobre o assunto não é capaz de opinar ou responder sobre o questionamento;
Telescoping	Jackson <i>et al.</i> (2016); Friedman (2013); Tourangeau <i>et al.</i> (2000);	Deslocamento temporal da ocorrência dos eventos, fatos que ocorreram mais recentes como mais remotos ou eventos distantes como mais recentes;
Saliência	Casey <i>et al.</i> (1967) e Glickman Michael Hubbard e Valciukas (1990)	Eventos mais importantes (salientes) tendem a ser lembrados com uma maior frequência em comparação a eventos menos salientes;
Memória	Raphael <i>et al.</i> (1991)	Esquecimento da existência de determinado evento, principalmente a medida que o seu tempo decorrido aumenta;
Mentira (“Rational lie”)	Gneezy <i>et al.</i> (2018) e Greene <i>et al.</i> (2017)	Mentira sobre suas informações privadas, principalmente em situações mais salientes, por exemplo, consumo de drogas.
Desenho de survey	Tourangeau <i>et al.</i> (2000)	Redação das perguntas, desenho amostral e relação entre entrevistado e entrevistador;
Gênero	Loxton <i>et al.</i> (2019); Webster <i>et al.</i> (2018) e Catallozzi <i>et al.</i> (2011)	Eventos inconsistentes podem estar ligados a questões de gênero, por exemplo: Patriarcalismo, Dominação, Violência doméstica, entre outros.

Fonte: Elaborado pelo autor.

4.3 Análise Exploratória dos dados

4.3.1 A PCSVDF-Mulher

A Pesquisa de Condições Socioeconômicas e Violência Doméstica e Familiar contra a Mulher (PCSVDF-Mulher) tem em questão um conjunto de dados longitudinal, que permite o

estudo de várias temáticas com relação a mulher, inclusive violência doméstica.

Como já explicitado nos capítulos anteriores, na PCSVDF-Mulher, a população amostrada é formada por mulheres com idade maior ou igual a 15 anos e menor que 50 anos residentes das nove capitais das Unidades Federativas da Região Nordeste do Brasil. As unidades amostrais são os domicílios e a unidade de observação é a mulher residente no domicílio dentro da faixa etária previamente estabelecida.

O objetivo do presente estudo é estudar respostas inconsistentes sobre VPI, a PCSVDF-Mulher, por ser uma pesquisa longitudinal, permite a realização de combinações de respostas sobre eventos de VPI que possam ser inconsistentes. Para identificar se as entrevistadas sofreram algum tipo de violência, foram questionadas se elas já sofreram algum tipo de evento pelo seu atual ou ex-parceiro mais recente, a tabela 28 apresenta todos os eventos, sejam de violência emocional, física e sexual que serviram como base para identificar se a entrevistada já foi vítima, seja ao longo da vida, seja nos últimos 12 meses.

Tabela 28 – Perguntas do questionário da PCSVDF-Mulher sobre episódios de violência

Emocional	1. Insultou você ou te fez sentir mal consigo mesma? 2. Menosprezou ou te humilhou na frente da sua família? 3. Te menosprezou ou te humilhou na frente de outras pessoas? 4. Fez coisas para te assustar ou te intimidar de propósito (ex: pela forma como ele te olhou, por gritar ou quebrar coisas)? 5. Ameaçou te ferir ou ferir alguém importante pra você?
Física	1. Te deu um tapa ou jogou algo em você que poderia machucá-la? 2. Empurrou-a ou puxou seu cabelo? 3. Te deu um soco ou fez algo que poderia machucá-la? 4. Chutou-a, arrastou-a ou a espancou? 5. Estrangulou-a? 6. Queimou-a de propósito? 7. Ameaçou usar uma arma de fogo, faca, pedaço de madeira, ferro, machado ou outra arma contra você? 8. Chegou a realmente usar uma arma de fogo, faca, pedaço de madeira, ferro, machado ou outra arma contra você?
Sexual	1. Forçou a ter relação sexual com você quando você não queria? 2. Teve relação sexual com você, SEM USAR VIOLÊNCIA FÍSICA, porque você teve medo da reação dele se você dissesse NÃO? 3. Forçou a fazer algo durante uma relação sexual que você achou degradante ou humilhante?

Fonte: Elaborado pelo autor.

4.3.2 Inconsistência: Operacionalização

Vale salientar que o ponto principal da pesquisa é o de avaliar a inconsistência na resposta de mulheres que sofreram algum tipo de violência cometida pelo seu parceiro atual ou ex-parceiro. Com isso, serão consideradas somente as respostas de mulheres que já possuíram algum parceiro ao longo da vida e que tenham aceitado participar da seção de violência. Portanto, mulheres com NA's motivados por 1.) e 2.1), anteriormente citados, serão desconsideradas da nossa amostra.

Por questões metodológicas, serão consideradas apenas combinações de respostas que possam ser classificadas como CONCLUSIVAS. Para isso, é necessário avaliar as respostas que contêm "NA" (não aplicável) em cada ano, a fim de classificar a resposta como CONCLUSIVA ou INCONCLUSIVA. A análise será realizada considerando a sequência das respostas, da esquerda para a direita, começando com a resposta dada pela mulher sobre episódios de violência ao longo da vida em 2016 e terminando com a resposta sobre episódios de violência nos últimos 12 meses em 2017. A sequência das respostas será a seguinte: (LV₂₀₁₆/_{<12}₂₀₁₆/LV₂₀₁₇/_{<12}₂₀₁₇)

1. Dois NA's dentro do mesmo ano levará a uma resposta INCONCLUSIVA, pois não poderemos afirmar se a mulher sofreu algum tipo de violência no determinado ano e com isso não poderemos considerar uma resposta como CONSISTENTE ou INCONSISTENTE.
Ex: NA/NA/NA/NA, NA/NA/1/0, 1/1/NA/NA;
2. NA aparecendo na resposta de violência ao longo da vida e a mulher afirmar que não sofreu violência nos últimos 12 meses também produz uma resposta INCONCLUSIVA,
Ex: NA/0/1/0. NA/0/0/0, 1/1/NA/0

Em uma combinação CONCLUSIVA é possível classificá-la como CONSISTENTE ou INCONSISTENTE. Uma resposta inconsistente se define como uma combinação que não tem sentido lógico quando combinamos os dois anos, 2016 e 2017, e os cortes dentro de cada ano, ao longo da vida e nos últimos 12 meses. Um exemplo de inconsistência é a mulher reportar que sofreu violência ao longo da vida em 2016 e, ao ser perguntada em 2017 se sofreu violência, dizer que nunca sofreu.

Todas as combinações de respostas são apresentadas na Tabela 46 (APÊNDICE), com uma coluna classificando cada combinação e informado se é CONCLUSIVA. Além dessa coluna, será adicionada também uma outra para esclarecer se a combinação é possível de ocorrer por conta de restrições no momento do preenchimento do questionário. Com isso, por mais que a combinação seja CONCLUSIVA, ela pode não ocorrer por conta de restrições do questionário.

Por exemplo, a combinação 0/1/1/NA é CONCLUSIVA (CONSISTENTE), porém ela não é possível de ocorrer tendo em vista que se a mulher afirmou que sofreu algum tipo de violência nos últimos 12 automaticamente ela sofreu violência ao longo da vida.

A tabela 29 mostra uma análise de frequência com todas as combinações que ocorreram a partir das respostas das mulheres sobre ter sofrido algum tipo de violência nas ondas 1 e 2. É possível perceber que algumas das combinações classificadas como possíveis na tabela 46 não apareceram na tabela 29.

Tabela 29 – Combinações de respostas totais de acordo com os cortes no tempo, somente com as combinações que ocorreram nas respostas

ID	2016		2017		Frequência	%	Conclusiva
	LV	<12 meses	LV	<12 meses			
81	0	0	0	0	825	43,86%	1
54	1	0	0	0	121	6,43%	1
51	1	0	1	0	114	6,06%	1
41	1	1	1	1	114	6,06%	1
77	0	0	1	1	92	4,89%	1
78	0	0	1	0	88	4,68%	1
9	NA	NA	0	0	83	4,41%	0
73	0	0	NA	NA	73	3,88%	0
45	1	1	0	0	67	3,56%	1
1	NA	NA	NA	NA	58	3,08%	0
50	1	0	1	1	52	2,76%	1
42	1	1	1	0	41	2,18%	1
36	1	NA	0	0	23	1,22%	0
6	NA	NA	1	0	22	1,17%	0
5	NA	NA	1	1	17	0,90%	0
76	0	0	1	NA	14	0,74%	1
31	1	NA	1	NA	14	0,74%	1
46	1	0	NA	NA	10	0,53%	0
32	1	NA	1	1	10	0,53%	1
37	1	1	NA	NA	9	0,48%	0
40	1	1	1	NA	8	0,43%	1
33	1	NA	1	0	8	0,43%	1
49	1	0	1	NA	7	0,37%	1
28	1	NA	NA	NA	7	0,37%	0
4	NA	NA	1	NA	4	0,21%	0

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

Apesar de serem retiradas observações de mulheres que nunca possuíram parceiro ou que tiveram algum parceiro, mas não aceitaram participar do questionário sobre violência, ainda sim algumas combinações com NA's aparecem, como pode ser observado na tabela 29. Nestes casos precisa-se avaliar minuciosamente o que as questões em que ainda ocorrem o NA

querem nos dizer. Por exemplo, a combinação NA/NA/NA/NA indica que a mulher chegou na seção de violência e desistiu de responder durante a seção, no entanto, ainda continuamos sem condições de afirmar que uma combinação dessas apresenta problema lógico e classificá-la em consistente ou inconsistente.

A tabela 29 possui um total de 16 combinações que são consideradas CONCLUSIVAS, entretanto, dentre essas combinações, sete possuem o NA na resposta de algum episódio de violência no recorte do tempo. No primeiro momento, essas sete combinações serão desconsideradas e trabalharemos somente com combinações em que ocorrem resposta afirmativa ou negativa sobre a ocorrência de violência. A tabela 30 indica todas as nove combinações e suas respectivas combinações.

Tabela 30 – Combinações de respostas de acordo com os cortes no tempo

ID	Resposta	2016		2017		Resposta analisada
		LV	< 12 MESES	LV	< 12 MESES	
81	Consistente	0	0	0	0	*
78	Inconsistente	0	0	1	0	*
77	Consistente	0	0	1	1	*
54	Inconsistente	1	0	0	0	*
51	Consistente	1	0	1	0	*
50	Consistente	1	0	1	1	*
45	Inconsistente	1	1	0	0	*
42	Consistente	1	1	1	0	*
41	Consistente	1	1	1	1	*

Fonte: Elaborado pelo autor.

As três combinações de inconsistências, identificadas pelas ID's 78, 54 e 45, podem ser agrupadas em duas categorias. A ID 78 representa uma direção contrária ao evento ocorrido no ano anterior. Já as ID's 54 e 45 podem ser combinadas e classificadas como uma supressão da informação sobre violência. Para fins de análise, essas inconsistências serão tratadas como inconsistência tipo 2 (para a ID 78) e inconsistência tipo 1 (para as ID's 54 e 45).

Como já definido e apresentado na revisão literária a ocorrência de inconsistência se dará por pelos seguintes mecanismos:

1. Desejabilidade social
2. Não entende a pergunta (cognição)
3. Esquece (telescoping)
4. Saliência
5. Memória

6. Mentira
7. Desenho de survey
8. Gênero

Considera-se que todas as respostas das mulheres, em 2016, são verdadeiras, ou seja, exclui-se a possibilidade de ocorrência de inconsistência nas respostas devido a eventuais mentiras por parte das entrevistadas. A próxima seção analisará os percentuais encontrados das combinações de respostas e, em consequência, os percentuais de combinações inconsistentes.

4.3.3 *Análise das inconsistências*

A primeira análise se dará de acordo com todas as combinações conclusivas de respostas dentro de cada tipo de violência, com o percentual em que a combinação aparece e sua determinada classificação.

Das nove possíveis combinações de respostas, três indicam inconsistência e seis indicam consistência na resposta das entrevistadas. Somente as respostas com Nº 54, 78 e 45 são consideradas inconsistentes, com isso, no decorrer do texto essas inconsistências serão classificadas de Inconsistência 1, 2 e 3 respectivamente.

É possível definir as combinações inconsistentes da seguinte forma:

- Inconsistência 1: quando a mulher afirma que em 2016 SOFREU algum tipo de violência ao longo da vida, porém (NÃO SOFREU violência no ano de 2016); e em 2017 afirma que NUNCA sofreu violência ao longo da vida;
- Inconsistência 2: quando a mulher afirma em 2016 que NUNCA sofreu violência doméstica nem ao longo da vida nem nos últimos 12 meses; entretanto, quando é feita a mesma pergunta em 2017, ela responde que SOFREU violência doméstica ao longo da vida.
- Inconsistência 3: quando a mulher afirma que, em 2016, SOFREU algum tipo de violência tanto ao longo da vida quanto nos últimos 12 meses e em 2017 afirma que NUNCA sofreu nenhum tipo de violência.

A tabela 31 mostra as nove sequências de respostas que ocorrem na PCSVDF-Mulher e suas respectivas frequências percentuais de acordo com o total de respostas, em ordem decrescente. Além disso, são destacadas as respostas que são classificadas como inconsistentes, de acordo com o seu tipo, como definido acima.

É possível observar um determinado padrão na combinação de respostas em todos os tipos de violência, principalmente na frequência das duas maiores combinações. As ID's 81

(mulher não ter sofrido nenhuma violência nas duas ondas) e 54 (INCONS 1) aparecem com maior frequência nos quatro recortes.

Tabela 31 – Combinações de respostas de acordo com os cortes no tempo e o tipo de violência

VIO GER		2016		2017		
ID	Resposta	LV	<12 meses	LV	<12 meses	Freq (%)
81	INCONS 1	0	0	0	0	54,49
54		1	0	0	0	7,99
51		1	0	1	0	7,53
15		1	1	1	1	7,53
77	INCONS 2	0	0	1	1	6,08
78		0	0	1	0	5,81
45	INCONS 3	1	1	0	0	4,43
50		1	0	1	1	3,43
42		1	1	1	0	2,71
VIO EMO		2016		2017		
ID	Resposta	LV	<12 meses	LV	<12 meses	Freq (%)
81	INCONS 1	0	0	0	0	59,01
54		1	0	0	0	7,06
77		0	0	1	1	6,44
51		1	0	1	0	6,38
78	INCONS 2	0	0	1	0	6,32
41		1	1	1	1	6,01
45	INCONS 3	1	1	0	0	4,15
50		1	0	1	1	2,66
42		1	1	1	0	1,98
VIO FIS		2016		2017		
ID	Resposta	LV	<12 meses	LV	<12 meses	Freq (%)
81	INCONS 1	0	0	0	0	75,59
54		1	0	0	0	5,91
51		1	0	1	0	5,09
78		0	0	1	0	4,32
45	INCONS 3	1	1	0	0	2,54
41		1	1	1	1	2,10
77		0	0	1	1	1,97
42		1	1	1	0	1,46
50		1	0	1	1	1,02
VIO SEX		2016		2017		
ID	Resposta	LV	<12 meses	LV	<12 meses	Freq (%)
81	INCONS 1	0	0	0	0	88,99
54		1	0	0	0	3,43
78		0	0	1	0	2,08
51		1	0	1	0	1,96
77	INCONS 3	0	0	1	1	1,59
45		1	1	0	0	1,28
41		1	1	1	1	0,92
50		1	0	1	1	0,31
42		1	1	1	0	0,18

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

A tabela 32 mostra o quantitativo de respostas consistentes e inconsistentes de acordo com as violências geral, emocional, física e sexual. A frequência máxima ocorre na violência geral, em que 18,08% das entrevistadas apresentaram inconsistência na sua resposta. Esse valor pode ser considerado alto, tendo em vista que a PCSVDF-Mulher investiu recursos e treinamento com as entrevistadoras, além de pré-testes para minimizar esse tipo de ocorrência, já que respostas inconsistentes levam potencialmente a estimativas imprecisas da prevalência.

De acordo com a Tabela 32, a violência emocional é o tipo de violência que apresenta um maior percentual de inconsistências, visto que a violência emocional não está bem enraizada no contexto da população, quando comparada a violência física e sexual que são melhores definidas. Esse resultado se repetiu em pesquisas anteriores, pois a violência emocional muitas vezes não é identificada como tal, [Webster *et al.* (2018), Catallozzi *et al.* (2011) e Loxton *et al.* (2019)].

Tabela 32 – Tipo de violência e porcentagem de respostas inconsistentes

Tipo violência	Respostas	Inconsistências	Percentual em relação ao total (%)
Violência geral	1530	247	18,08
Violência emocional	1450	249	17,17
Violência física	1416	178	12,57
Violência sexual	1446	96	6,54

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

4.4 Modelagem econométrica

De acordo com Andre e Carvalho (2023), na PCSVDF-Mulher há um total de 4316 mulheres pareadas entre os anos de 2016 e 2017. Entretanto, como explicado anteriormente, serão desconsideradas mulheres que não tiveram parceiro ou que não aceitaram participar do questionário sobre violência. Com isso, o total de mulheres que tiveram parceiros e aceitaram participar da seção sobre violência foi de 1687.

Antes de apresentar a análise descritiva com o percentual de inconsistências em cada variável explicativa que posteriormente serão utilizadas para se estimar um modelo logit e um modelo multilogit, é preciso defini-las e descrever o tipo de variável, como consta na tabela 33.

4.4.1 Variáveis do modelo

Tabela 33 – Variáveis para o modelo e sua descrição

Variável	Descrição
residente(CE/PE/BA)	variável dummy que indica a capital nordestina que ocorreu a pesquisa, recebendo valor 1 se a mulher residir nos municípios de Fortaleza-CE, Recife-PE e Salvador-BA e 0 caso contrário
jovem_mul	variável dummy que recebe valor 1 se a mulher é jovem [15,24] anos e 0 caso ela seja adulta [25,52] anos
escolari_mul	variável categórica que indica a escolaridade da mulher. 1-ensino fundamental, 2-ensino médio/técnico e 3-ensino superior/pós graduação
branca_mul	variável dummy que indica a raça da mulher. 1-branca e 0-não branca
filhos	variável dummy que indica se a mulher tem pelo menos 1 filho. 1-tem filho e 0-não tem filho
status_trabalho	variável categórica que indica se a mulher mudou de status com relação ao trabalho de 2016 para 2017. 1-Diminuiu (Sim2016/Não2017), 2- Permaneceu (Não2016/Não2017) (Sim2016/Sim2017), 3-Aumentou (Não2016/Sim2017),
branco_parc	variável dummy que indica a raça do parceiro atual. 1-branca e 0-não branca
escolari_parc	variável categórica que indica a escolaridade do parceiro. 1-ensino fundamental, 2-ensino médio/técnico e 3-ensino superior/pós graduação
trabalho_parc	variável dummy que indica se o parceiro trabalha. 1-sim e 0-não
parceiro_w1	variável dummy que indica se a mulher continua com o mesmo parceiro da onda 1. 1-sim e 0-não
barganha	variável categórica que indica a relação de trabalho entre a mulher e o parceiro, 1-parceiro trabalha e mulher não trabalha, 2-ambos trabalham ou ambos não trabalham, 3-mulher trabalha e parceiro não trabalha
LMP_falar	variável dummy que indica se a mulher já ouviu falar da Lei Maria da Penha alguma vez na vida. 1-sim e 0-não
conhec_LMP	variável dummy que indica se o conhecimento sobre a Lei Maria da Penha diminuiu ou permaneceu e aumentou. 1-Aumentou e 0-Diminuiu/Permaneceu
violence	variável dummy se a mulher sofreu algum tipo de violência nos últimos 12 meses. 1-sim e 0-não

Conhecimento Lei Maria da Penha

Um indicador que fornece o conhecimento da mulher sobre a Lei Maria da Penha (LMP) foi obtido a partir de seis perguntas sobre o entendimento da entrevistada sobre a lei, tendo em vista que essas perguntas foram realizadas em 2016 e 2017 foi feito um score que iria de 0 a 6, dependendo da quantidade de respostas corretas que elas responderam nos dois anos.

Para as perguntas q404_1, q404_4, q404_5, q404_6, q404_7 e q404_8, tabela 47 APÊNDICE B, as respostas que foram dadas como “concordo completamente” até “discordo completamente” foram definidas como perguntas respondidas corretamente, pois como se trata de uma opinião sobre a Lei é possível considerar que as respondentes possuem conhecimento e podem concordar ou não com ela, caso as respondentes afirmem “não sabe/não respondeu” será considerado como um não conhecimento para aquela afirmação.

Com isso, cada pergunta respondida corretamente irá receber um score 1 e cada pergunta respondida errada receberá score 0, por fim serão somados os scores nos anos de 2016 e 2017. Em última instância, é comparado o score em 2016 e em 2017 e as mulheres entrevistadas são classificadas de três formas:

- Conhecimento diminuiu, se $\text{score}_{2017} < \text{score}_{2016}$;
- Conhecimento permaneceu, se $\text{score}_{2017} = \text{score}_{2016}$;
- Conhecimento aumentou, se $\text{score}_{2017} > \text{score}_{2016}$;

A tabela 34 fornece o quantitativo de mulheres que possuem um determinado conhecimento de acordo com cada pergunta sobre a LMP,

Tabela 34 – Conhecimento da LMP de acordo com a pergunta e o ano

Pergunta sobre LMP	Possui conhecimento	Não possui conhecimento	Possui conhecimento	Não possui conhecimento
	2016		2017	
q404_1	1672 (0,9929)	12 (0,0071)	1679 (0,9953)	08 (0,0047)
q404_4	1652 (0,9810)	32 (0,0190)	1658 (0,9828)	29 (0,0172)
q404_5	1602 (0,9519)	81 (0,0481)	1622 (0,9615)	65 (0,0385)
q404_6	1504 (0,8931)	180 (0,1069)	1530 (0,9069)	157 (0,0931)
q404_7	1647 (0,9780)	37 (0,0220)	1668 (0,9887)	19 (0,0113)
q404_8	1643 (0,9757)	41 (0,0243)	1658 (0,9828)	29 (0,0172)

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

4.4.2 Análise descritiva

A presente seção apresenta uma análise exploratória de cada variável que será utilizada no modelo representado na tabela 35, além da mesma análise sobre variáveis que foram criadas a partir da Lei Maria da Penha (LMP). Finalizando com uma análise associativa da inconsistência nos quatro recortes de violência (geral, emocional, física e sexual) associado as variáveis da mulher, sobre a LMP e do parceiro.

A tabela 35 mostra a análise de cada variável, das 1687 mulheres que participaram

Tabela 35 – Análise descritiva das variáveis da mulher e do parceiro

Variável	Média	Mínimo	Máximo	Desvio
Características mulher				
Município residência				
Residente(CE/PE/BA)	0,4428	0	1	0,4968
Idade				
jovem_mul	0,1731	0	1	0,3784
Escolaridade				
escolari_mul	1,9410	1	3	0,6630
Raça				
branca_mul	0,2015	0	1	0,4012
Possui filhos				
Sim	0,7789	0	1	0,4151
Trabalho remunerado				
status_trabalho	2,0470	1	3	0,4518
Características parceiro				
Raça parceiro				
branco_parc	0,3183	0	1	0,4659
Escolaridade do parceiro				
escolari_parc	1,7890	1	3	0,5964
Trabalho do parceiro				
trabalho_parc	0,8074	0	1	0,3944
Mulher/Parceiro				
Parceiro_w1				
Sim	0,6242	0	1	0,4844
Barganha mulher/parceiro				
barganha	1,6230	1	3	0,6133

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

da PCSVDF-Mulher em 2016 e 2017.¹ O município de residência (CE/PE/BA) mostrou que 44,28% das mulheres residem nesse município. Tal divisão ocorreu pois os três municípios em questão possuem uma maior população e, com isso, tem-se o objetivo de verificar se mulheres residentes desse tipo de município, mais populoso, influenciam na inconsistência.

A média de idade das mulheres foi de aproximadamente 35 anos e desvio igual a 9,37, sendo a idade mínima de 15 anos e idade máxima de 52 anos. As mulheres com escolaridade de ensino médio/técnico representam mais da metade, 55,72%, do total de mulheres, sendo seguido por mulheres com ensino fundamental (25,07%) e aproximadamente 20% das mulheres entrevistadas possuíam ensino superior ou pós graduação, com isso a média para a variável escolaridade foi igual a 1,9410, além disso mulheres não brancas representam aproximadamente 80% do total das entrevistadas.

Adicionalmente, aproximadamente 80% delas possuem pelo menos um filho, sendo

¹ Mulheres que tiveram parceiro e aceitaram participar da seção de violência em 2016 e 2017

a média do número de filhos igual a 1,55 e desvio de 1,37, mínimo de 0 filhos e máximo de 11 filhos. Por fim, das características da mulher, foi realizada uma forma de verificar se houve mudança dos status quo da renda entre os anos de 2016 e 2017. As mulheres foram classificadas de acordo com a situação, se estavam em algum trabalho remunerado durante os anos, sendo possível três classificações: diminuiu status, permaneceu e aumentou. De acordo com os resultados, observou-se que grande parte das mulheres não tiveram mudanças no status quo da renda entre anos, tendo em vista que 49,02% delas não trabalhavam em 2016 e não trabalhavam em 2017, adicionalmente 30,35% das mulheres trabalhavam em 2016 e 2017, a média para a variável status_trabalho foi igual a 2,0470.

A raça do parceiro se mostrou semelhante ao comportamento da raça das mulheres, com cerca de 31,83% dos homens sendo brancos. A escolaridade do parceiro também se mostrou muito semelhante ao comportamento da escolaridade das mulheres entrevistadas. A maior parte são de homens com ensino médio/técnico 60%, seguido de parceiros com ensino fundamental 30,55% e somente 9,45% com ensino superior/pós graduação. Isso representou para a variável uma média de 1,7890 e, além disso, aproximadamente 80,74% dos parceiros tem ou tinham trabalho remunerado.

A análise de variáveis da mulher/parceiro mostra que mais da metade das mulheres, 62,42%, estão com o mesmo parceiro da primeira entrevista, as demais, ou estão solteiras no momento ou estão com outro parceiro. E com relação a barganha da mulher com o parceiro, a variável apresentou média igual a 1,6230 aproximando da resposta que ambos possuem o mesmo status, ou não trabalham ou trabalham.

Tabela 36 – Análise de frequência sobre a Lei Maria da Penha (LMP)

Variável	Freq.absol.(percent.)		
Ouviu falar LMP	Nunca2017	Pouco2017	Muito 2017
Nunca2016	00 (0,0000)	03 (0,0018)	11 (0,0066)
Pouco2016	01 (0,0006)	77 (0,0459)	154 (0,0917)
Muito2016	05 (0,0030)	168 (0,1001)	1260 (0,7504)
Variação conhecimento LMP			
Diminuiu (2016/2017)	141 (0,0838)		
Manteve (2016/2017)	1365 (0,8111)		
Aumentou (2016/2017)	177 (0,1052)		

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

A análise com relação a frequência em que a entrevistada ouviu falar sobre a Lei Maria da Penha (LMP) e a variação do seu conhecimento sobre algumas afirmações referentes

a lei são apresentadas na tabela 36. É possível observar que não houve mulheres que nunca ouviram falar da LMP em ambos os anos, e que a grande maioria, 75,04%, ouviram falar muito sobre a lei em ambos os anos. Além disso, 10% das mulheres relataram que ocorreu um aumento na frequência sobre ouvir falar da LMP, assim como também 10% delas relataram que ocorreu uma redução sobre terem ouvido falar sobre a LMP de 2016 para 2017.

Com relação a variação do conhecimento das mulheres sobre a LMP, explicado anteriormente que tem o objetivo de verificar se o que ela afirmou conhecer sobre a lei em 2016 sofreu alguma variação em 2017, a variável não tem o intuito de saber se a mulher conhece muito ou pouco a LMP. Percebe-se que um pouco mais de 80% das entrevistadas manteve o seu conhecimento sobre a lei, e as que tiveram uma variação positiva e negativa do conhecimento ficaram com um percentual bem próximo, sendo 10,52% e 8,38%, respectivamente.

4.4.3 Inconsistência e variáveis

A tabela 48 presente no APÊNDICE B apresenta os dados de inconsistência nas respostas sobre violência geral, emocional, física e sexual de acordo com as variáveis socioeconômicas da mulher e do seu parceiro, além da relação de aspectos sobre a LMP e a inconsistência nas respostas.

Como já explicitado anteriormente, existe um certo padrão em que as inconsistências nas respostas sobre violência emocional é superior a violência física que é superior a violência sexual, e esse padrão é observado também, na maior parte, dentro das variáveis.

Com relação ao local de residência da mulher, existe uma menor prevalência da inconsistência em residentes nas capitais de CE/PE/BA. Essa diferença é significativa dentro das violências geral, emocional e física, entretanto, com relação a violência sexual, o percentual de residentes de CE/PE/BA é estatisticamente semelhantes com relação as residentes dos demais municípios.

Percebe-se que mulheres jovens apresentam um maior percentual de inconsistência para violência geral, emocional e violência física quando comparada a mulheres adultas. Entretanto, esse cenário é diferente para respostas inconsistentes da violência sexual, apesar dos percentuais serem bem semelhantes para jovens e adultas, o percentual de inconsistência de mulheres adultas é ligeiramente superior em relação as mulheres adultas.

O grau de instrução das entrevistadas apresentam um perfil parecido de respostas inconsistentes para violência geral e emocional. Quando visualizamos as colunas que representam

as inconsistências dentro de cada tipo de violência percebe-se um decréscimo do percentual de respostas inconsistentes a medida que aumenta o grau de instrução das mulheres. Vale ressaltar que isso é esperado, tendo em vista que pessoas que possuem um grau de instrução maior conseqüentemente tendem a reportar as vivências de vida com mais consistência.

Com relação a etnia, por mais que a diferença percentual de respostas inconsistentes entre mulheres brancas e não brancas seja mínima, percebe-se uma porcentagem maior de mulheres não brancas com respostas inconsistentes para violência emocional.

A situação de trabalho da mulher durante os anos de 2016 e 2017 mostra que mulheres que não mudaram de status, ou seja, estavam sem emprego em 2016 e continuam sem emprego em 2017 ou estavam empregadas nos dois anos, tendem a ser menos inconsistentes em relação as mulheres que mudaram de status de trabalho durante os anos. Além disso, o fato das mulheres possuírem pelo menos um filho as tornam mais inconsistentes nas suas respostas, isso pode ser explicado pelo fato delas quererem resguardar os filhos de algum episódio de violência.

As características do parceiro são medidas a partir da raça e de sua escolaridade. A raça do parceiro não se mostrou relevante para a mulher reportar alguma resposta inconsistente, esse comportamento também foi encontrado com a raça da mulher. E a escolaridade do parceiro seguiu o mesmo padrão da escolaridade da mulher, isto é, mulheres com parceiros melhores instruídos tendem a reportar os episódios de violência de uma forma mais consistente. Esse fato ocorre com exceção da violência sexual, em que não existe diferença entre as escolaridades dos parceiros.

A relação entre mulher e parceiro mostra que mulheres que continuam com o mesmo parceiro da onda 1 tem um percentual maior de inconsistência na violência emocional e sexual. Com relação a barganha entre a mulher e o parceiro, os percentuais são estatisticamente semelhantes.

A última análise se dá com as variáveis que envolvem a LMP. Mulheres que nunca ouviram falar ou ouviram falar pouco da LMP tendem a reportar respostas inconsistentes, sendo a diferença percentual maior na violência geral e na violência emocional.

A variação do conhecimento mostrou que mulheres que tiveram variação sobre o conhecimento de algumas afirmações sobre a LMP se mostraram mais inconsistentes em relação as mulheres que não sofreram variação. Isso ocorre nas categorias violência geral, emocional e física. Para a violência sexual o cenário é diferente. Mulheres que mantiveram o conhecimento sobre a LMP foram mais inconsistentes, apesar de, nesse caso, a diferença entre os percentuais

ser menor do que nas outras violências.

4.4.4 Modelo logit padrão

Nessa seção são apresentados os efeitos marginais dos modelos logit, para os diferentes tipos de violência. Em posse dos valores dos efeitos marginais das variáveis explicativas é possível analisar o efeito dessas variáveis sobre a probabilidade da ocorrência de inconsistência na resposta sobre eventos de violência. Para isso, serão agrupados todos os tipos de inconsistência, sendo a variável resposta e as demais variáveis descritas na tabela 33 como as variáveis de controle. Os resultados das Tabelas 37, 38, 39 e 40 indicam os resultados estimados das variáveis para a estimação da inconsistência na resposta

VIOLÊNCIA GERAL

A tabela 37 apresenta os resultados para a violência geral. Levando em consideração o município de residência das entrevistas observa-se que residentes de Fortaleza-CE, Recife-PE e Salvador-BA tiveram valor significativo comparado a residentes dos demais municípios, conclui-se, então, que residir nesses municípios aumenta a probabilidade de inconsistência na resposta.

No bloco de características da mulher, não se mostraram relevantes as variáveis de idade e raça, entretanto a escolaridade apresentou valor significativo e negativo, mostrando que quanto maior a escolaridade da mulher menor será a sua chance de reportar uma inconsistência e escolaridade, em contrapartida ter filhos aumenta a probabilidade de inconsistência.

Com relação ao mercado de trabalho, o único status de trabalho remunerado foi das mulheres que ficaram desempregadas de 2016 para 2017. As mulheres com esse perfil tendem a reportar mais inconsistências em relação a variável de referência, que são mulheres que não mudaram de status com relação ao trabalho.

As características do parceiro, em que todos os efeitos marginais foram negativos, não se mostraram significativos para explicar a inconsistência nesse tipo de violência, além delas as variáveis que dizem a respeito a LMP também não se mostraram significantes.

A mulher continuar com o mesmo parceiro desde a onda 1 se mostrou negativamente relacionado com a inconsistência, mostrando que mulheres que estão com o mesmo parceiro tendem a ser menos inconsistentes. Por fim, a mulher ter sofrido alguma violência nos últimos 12 meses contribui positivamente para a mulher reportar uma resposta inconsistente.

VIOLÊNCIA EMOCIONAL

Com relação a inconsistência sobre eventos de violência emocional, igualmente da violência geral, as mulheres residentes dos municípios menos populosos foram significativas, como mostra a tabela 38. As mulheres residentes dos municípios com menor população apresentaram probabilidade menor de inconsistência quando comparadas a residentes de municípios mais populosos.

Para as variáveis que contemplam as características da mulher, o comportamento foi semelhante ao da violência geral, em que somente as variáveis filhos tiveram relação para a entrevistada apresentar uma resposta inconsistente. Entretanto, diferentemente do primeiro modelo, a variável que relaciona com o mercado de trabalho não apresentou relevância para a resposta inconsistente da mulher.

As características do parceiro, em que a escolaridade apresentou valor significativo e negativo, indica que quanto maior for o grau de instrução do parceiro as mulheres tendem a ser menos inconsistentes para a violência emocional, além disso, a mulher continuar com o mesmo parceiro desde a onda 1 se mostrou negativamente relacionado com a inconsistência.

A variável que indica a ocorrência da mulher ter ouvido falar sobre a LMP mostrou que as entrevistadas que escutaram falar da LMP com uma maior frequência tendem a ser menos inconsistentes na resposta sobre episódios de violência emocional. Sobre as variáveis que explicam a inconsistência da violência emocional mulheres que relataram sofrer algum tipo de episódio de violência emocional nos últimos 12 meses estas contribuem positivamente para a mulher reportar uma resposta inconsistente.

VIOLÊNCIA FÍSICA

As mulheres residentes da cidade de Maceió-AL, São Luís-MA e Teresina-PI foram menos inconsistentes nos relatos de episódios de violência física. Esse fato pode ser levado em consideração, tendo em vista que a variável que indica residência das entrevistadas constatou que residentes de Fortaleza-CE, Recife-PE e Salvador-BA aponta valor positivo e significativo para a inconsistência no relato da violência física.

Ainda sobre a tabela 39 das demais variáveis de características da mulher, se a entrevistada possuir filhos então há relação para apresentar uma resposta inconsistente. Entretanto, diferentemente do modelo de violência emocional, a variável que relaciona com o mercado de trabalho não apresentou relevância para a resposta inconsistente da mulher. Igualmente, a violência emocional a mulher continuar com o mesmo parceiro desde a onda 1 se mostrou negativamente relacionado com a inconsistência.

Mulheres que relataram sofrer algum tipo de episódio de violência nos últimos 12 meses tiveram uma probabilidade maior de reportar respostas inconsistentes. Esse fato ocorre em todas as outras violências e nos leva a deduzir que mulheres vítimas recentes tendem a se resguardar e omitir alguma resposta sobre os episódios de violência.

VIOLÊNCIA SEXUAL

Para finalizar as análises das inconsistências por tipo de violência, tem-se a violência sexual, na tabela 40. Nas características da mulher, a dummy ter filhos apresentou valor significativo, além disso, mulheres que começaram a trabalhar em 2017 se mostraram mais inconsistentes para esse tipo de violência. Como a violência sexual é mais saliente e bem definida, pode-se deduzir que esse tipo de inconsistência no novo trabalho pode ser explicado como uma omissão por conta do novo emprego, tendo em vista que em 2016 as mulheres não eram empregadas.

Ainda sobre a tabela 40, a variável raça do parceiro, diferentemente das violências emocional, física e geral, se mostrou significativa para a inconsistência na resposta sobre violência sexual e como a variável foi definida como uma dummy em que o valor de referência é de parceiros brancos, tem-se um efeito marginal negativo e com isso mulheres com parceiros brancos tendem a reportar respostas menos inconsistentes.

Ademais as variáveis de relação da mulher com o mesmo parceiro da onda 1 e ter sofrido violência recente, se mostraram negativamente e positivamente, respectivamente, relacionados com a inconsistência. A variação de conhecimento sobre a LMP mostrou que mulheres que aumentaram o seu conhecimento adquirido sobre a lei são menos propensas a serem inconsistentes.

Tabela 37 – Modelo logit para inconsistência na resposta sobre violência geral

	Ef. Marginal	error	Estatística t	p-valor
Entrevistada				
muni(CE/PE/BA)	0,0448	0,0210	2,1301	0,0334**
idade(jovem)	-0,0158	0,0298	-0,5297	0,5964
escolaridade	-0,0302	0,0178	-1,6998	0,0894*
raça	0,0004	0,0264	0,0153	0,9878
filhos	0,0555	0,0254	2,1828	0,0292**
status_trab(diminuiu)	0,0865	0,0466	1,8580	0,0634*
status_trab(aumentou)	0,0284	0,0361	0,7868	0,4315
Parceiro				
raça	-0,0207	0,0218	-0,9497	0,3425
escolaridade	-0,0072	0,0194	-0,3725	0,7096
trabalho_parceiro	-0,0162	0,0375	-0,4309	0,6666
Mulher/Parceiro				
mulher_parceiro_w1	-0,0405	0,0226	-1,7928	0,0732*
barganha(diminuiu)	-0,0278	0,0270	-1,0295	0,3035
barganha(aumentou)	0,0012	0,0495	0,0241	0,9808
LMP				
imp_falar	-0,0391	0,0312	-1,2558	0,2094
var_con_lmp_v3	0,0245	0,0340	0,7219	0,4705
Violência				
violence16_12	0,1282	0,0342	3,7504	<0,001***

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001 **p-valor<0,05 *p-valor<0,1

Tabela 38 – Modelo logit para inconsistência na resposta sobre violência emocional

	Ef. Marginal	error	Estatística t	p-valor
Entrevistada				
muni(CE/PE/BA)	0,0391	0,0200	1,9579	0,0504*
idade(jovem))	-0,0287	0,0272	-1,0552	0,2915
escolaridade	-0,0083	0,0170	-0,4920	0,6228
raça	0,0088	0,0256	0,3452	0,7300
filhos	0,0661	0,0235	2,8170	0,0049**
status_trab(diminuiu)	0,0423	0,0418	1,0122	0,3116
status_trab(aumentou)	0,0380	0,0351	1,0839	0,2786
Parceiro				
raça	-0,0104	0,0210	-0,4975	0,6189
escolaridade	-0,0369	0,0185	-1,9960	0,0461**
trabalho_parceiro	-0,0083	0,0353	-0,2360	0,8135
Mulher/Parceiro				
mulher_parceiro_w1	-0,0617	0,0220	-2,8108	0,0050**
barganha(diminuiu)	-0,0243	0,0256	-0,9492	0,3427
barganha(aumentou)	0,0014	0,0471	0,0305	0,9757
LMP				
imp_falar	-0,0677	0,0314	-2,1591	0,0310**
var_con_lmp_v3	0,0101	0,0316	0,3205	0,7486
Violência				
violence16_12	0,1639	0,0347	4,7285	<0,001***

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001 **p-valor<0,05 *p-valor<0,1

Tabela 39 – Modelo logit para inconsistência na resposta sobre violência física

	Ef. Marginal	error	Estatística t	p-valor
Entrevistada				
muni(CE/PE/BA)	0,0560	0,0170	3,2945	0,0010**
idade_mul(jovem)	-0,0308	0,0216	-1,4225	0,1551
escolaridade	-0,0159	0,0139	-1,1484	0,2510
raça	-0,0043	0,0208	-0,2045	0,8380
filhos	0,0741	0,0181	4,1027	<0,001***
status_trab(diminuiu)	0,0048	0,0302	0,1575	0,8749
status_trab(aumentou)	0,0284	0,0313	0,9092	0,3634
Parceiro				
raça	-0,0190	0,0170	-1,1202	0,2628
escolaridade	-0,0152	0,0151	-1,0049	0,3151
trabalho_parceiro	-0,0252	0,0315	-0,7985	0,4247
Mulher/Parceiro				
mulher_parceiro_w1	-0,0361	0,0182	-1,9853	0,0473**
barganha(diminuiu)	0,0030	0,0215	0,1401	0,8886
barganha(aumentou)	-0,0476	0,0271	-1,7584	0,0789*
LMP				
imp_falar	0,0001	0,0226	0,0056	0,9956
var_con_lmp_v3	0,0357	0,0289	1,2348	0,2171
Violência				
violence	0,1327	0,0315	4,2116	<0,001***

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001 **p-valor<0,05 *p-valor<0,1

Tabela 40 – Modelo logit para inconsistência na resposta sobre violência sexual

	Ef. Marginal	error	Estatística t	p-valor
Entrevistada				
muni(CE/PE/BA)	0,0129	0,0115	1,1287	0,2592
idade_mul(jovem)	-0,0057	0,0149	-0,3828	0,7020
escolaridade	-0,0028	0,0095	-0,2930	0,7696
raça	-0,0151	0,0132	-1,1476	0,2513
filhos	0,0229	0,0130	1,7633	0,0781*
status_trab(diminuiu)	-0,0168	0,0168	-1,0009	0,3171
status_trab(aumentou)	0,0537	0,0295	1,8176	0,0693*
Parceiro				
raça	-0,0228	0,0111	-2,0531	0,0403**
escolaridade	0,0016	0,0105	0,1489	0,8817
trabalho_parceiro	-0,0113	0,0237	-0,4748	0,6350
Mulher/Parceiro				
mulher_parceiro_w1	-0,0336	0,0133	-2,5204	0,0118**
barganha(diminuiu)	0,0244	0,0160	1,5260	0,1273
barganha(aumentou)	-0,0158	0,0220	-0,7171	0,4734
LMP				
imp_falar	-0,0054	0,0163	-0,3316	0,7403
var_con_lmp_v3	-0,0291	0,0137	-2,1203	0,0342**
Violência				
violence	0,0974	0,0257	3,7900	0,0002***

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001 **p-valor<0,05 *p-valor<0,1

4.4.5 Modelo multi-logit

Também foram examinados se as respostas consistentes e inconsistentes (tipo 1 e 2) de VPI diferiam entre si, de acordo com as variáveis também estimadas no modelo logit. De acordo com a tabela 42, a inconsistência do tipo 1 aparece com maior frequência em todos os tipos de violência, sendo aproximadamente o dobro da inconsistência 2, com exceção a violência emocional.

Foi utilizada a regressão logística multinomial para determinar a associação entre características da mulher e do parceiro com relatos inconsistentes sobre VPI. Esses modelos estimam simultaneamente as variáveis explicativas de cada Grupo VPI (Inconsistente), relativo ao grupo VPI Consistente.

Serão agrupados duas respostas inconsistentes que possuem o mesmo perfil de resposta, ID 54 e 45, como inconsistência 1 e a inconsistência com ID 78 será considerada como inconsistência 2. A tabela 41 indica as combinações de cada grupo e a tabela 42 mostra a frequência que cada grupo inconsistente ocorre. Percebe-se que a frequência da inconsistência do tipo 1 é aproximadamente o dobro da ocorrência da inconsistência do tipo 2, isso ocorre em todas as violências.

Tabela 41 – Combinações de respostas de acordo com os cortes no tempo

ID	Resposta	2016		2017		Tipo inconsistencia
		LV	< 12 MESES	LV	< 12 MESES	
81	Consistente	0	0	0	0	*
78	Inconsistente	0	0	1	0	Tipo 2
77	Consistente	0	0	1	1	*
54	Inconsistente	1	0	0	0	Tipo 1
51	Consistente	1	0	1	0	*
50	Consistente	1	0	1	1	*
45	Inconsistente	1	1	0	0	Tipo 1
42	Consistente	1	1	1	0	*
41	Consistente	1	1	1	1	*

Fonte: Elaborado pelo autor.

Tabela 42 – Tipo de violência e porcentagem de respostas inconsistentes

Tipo violência	Respostas	Consistentes	Inconsistentes (tipo 1)	Inconsistentes (tipo 2)
Violência geral	1530	1219 (0,8192)	164 (0,1201)	83 (0,0608)
Violência emocional	1450	1201 (0,8283)	156 (0,1076)	93 (0,0641)
Violência física	1416	1238 (0,8743)	113 (0,0798)	65 (0,0459)
Violência sexual	1467	1371 (0,9346)	65 (0,0443)	31 (0,0211)

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

As tabelas 43 a 45 mostram os resultados estimados do modelo multilogit de acordo com os dois tipos de inconsistência, já explicitado anteriormente. As variáveis utilizadas são as mesmas do modelo logit da seção anterior, somente com a diferença que nesta seção não será contemplada no modelo multilogit a violência geral.

VIOLÊNCIA EMOCIONAL

A tabela 43 mostra o resultado da estimação do modelo multilogit para a inconsistência na violência emocional. Para a inconsistência do tipo 1 somente a ocorrência de violência no último mês se mostrou significativa para a ocorrência de um relato inconsistente.

Ainda sobre a tabela 43, para o tipo 2, o resultado se aproximou mais do que aquele encontrado no modelo logit. As variáveis explicativas (município de residência, filhos, ter o mesmo parceiro da onda 1, ouviu falar da LMP e ter sofrido violência) mostraram relação com esse tipo de inconsistência.

Para as variáveis município de residência, filhos, ter o mesmo parceiro da onda 1, ouviu falar da LMP, têm magnitude semelhante ao que já foi encontrado na inconsistência quando foram considerados todas os tipos.

A diferença que pode ser considerada mais significativa se dá na relação da resposta inconsistente e na variável violence. Para o tipo 1, essa variável apresentou valor positivo, pela definição do tipo de inconsistência (1/0/0/0 e 1/1/0/0), entretanto, no tipo 2, a variável apresentou efeito marginal negativo, sendo a sua definição (0/0/1/0)

VIOLÊNCIA FÍSICA

Com relação a violência física, a tabela 44 fornece a estimação do modelo multilogit para a inconsistência. Mulheres que relataram ter sofrido violência ao longo da vida em 2016 e não ter sofrido violência ao longo da vida em 2017 garantem, assim, uma resposta inconsistente do tipo 1, e receberam influência das variáveis residência, filhos, variação do conhecimento da LMP e ter sofrido violência recentemente.

No modelo de violência emocional, somente a variável de violência mostrou efeito significativo de influência em uma resposta inconsistente do tipo 1 com relação a uma resposta consistente, no entanto, na violência física percebe-se que o município, fator ter filhos e se o seu conhecimento da LMP aumentou tem efeito positivo na inconsistência.

Para as mulheres que foram inconsistentes do tipo 2, além das variáveis município e filhos que já tiveram influência no tipo 1, vale ser destacado o status do trabalho, sendo que mulheres que começaram a trabalhar em 2017 foram mais propensas a reportar uma resposta

inconsistente. Mulheres com parceiros não brancos tem uma probabilidade menor de inconsistência, além disso, ela ter o mesmo parceiro da onda 1 e se o seu poder de barganha aumentar também se mostraram negativamente correlacionados.

Apesar de nos outros modelos estimados, tanto no logit quanto no multilogit para violência emocional, ter a influência da violência na inconsistência, para o tipo 2 da violência física essa variável não se mostrou relevante.

VIOLÊNCIA SEXUAL

Para finalizar as análises das inconsistências por tipo de violência, tem-se a violência sexual, representada na tabela 45. O município de residência das entrevistadas teve influência somente na inconsistência do tipo 2, perfazendo, portanto, mulheres residentes de CE/PE/BA foram mais inconsistentes.

O resultado estimado do efeito marginal das variáveis, da inconsistência 1, filhos, status do trabalho (ambos com efeito positivo) e raça do parceiro (efeito negativo) se mostraram semelhantes ao encontrado na estimação da violência emocional para a inconsistência do tipo 2, indicada na tabela 44

Ainda sobre a tabela 45, é possível verificar a influência da conhecimento sobre a LMP, em que mulheres que aumentaram seu conhecimento foram menos inconsistentes, para o grupo específico, e se caso elas sofreram episódios de violência recente teve efeito positivo.

De acordo com os resultados, na violência sexual, a inconsistência que omitiu a violência ocorrida em 2016 teve influência do município de residência da mulher, se ela continua com o mesmo parceiro e seu poder de barganha.

Tabela 43 – Modelo multilogit para inconsistência na resposta sobre violência emocional

Inconsistência 1				
	Efeito margin	Desvio	Estat. Z	p-valor
(CE/PE/BA)	-0,0059	0,0154	-0,3800	0,7020
idade_jovem	-0,0244	0,0204	-1,2000	0,2300
escolaridade	-0,0087	0,0133	-0,6600	0,5120
raça	0,0195	0,0207	0,9400	0,3460
filhos	0,0199	0,0199	1,0000	0,3180
trabalho_status	-0,0106	0,0188	-0,5700	0,5720
raça_parcei	0,0084	0,0168	0,5000	0,6170
escolaridade_parcei	-0,0130	0,0144	-0,9000	0,3660
trabalho_parcei	0,0130	0,0237	0,5500	0,5840
parceiro_w1	-0,0027	0,0164	-0,1600	0,8700
barganha	0,0200	0,0184	1,0900	0,2760
lmp_falar	-0,0285	0,0245	-1,1600	0,2450
var_conhec_lmp	0,0100	0,0258	0,3900	0,6980
violence	0,2258	0,0342	6,6100	<0,001***
Inconsistência 2				
	Efeito margin	Desvio	Estat. Z	p-valor
(CE/PE/BA)	0,0395	0,0109	3,6200	<0,001***
idade_jovem	-0,0008	0,0141	-0,0500	0,9570
escolaridade	-0,0006	0,0083	-0,0700	0,9460
raça	-0,0096	0,0119	-0,8100	0,4200
filhos	0,0341	0,0100	3,4200	<0,001***
trabalho_status	0,0127	0,0118	1,0800	0,2790
raça_parcei	-0,0152	0,0098	-1,5500	0,1210
escolaridade_parcei	-0,0174	0,0090	-1,9300	0,053*
trabalho_parcei	-0,0108	0,0179	-0,6100	0,5450
parceiro_w1	-0,0514	0,0131	-3,9200	<0,001***
barganha	0,0003	0,0118	0,0300	0,9790
lmp_falar	-0,0332	0,0174	-1,9100	0,056*
var_conhec_lmp	-0,0024	0,0138	-0,1800	0,8590
violence	-0,0439	0,0088	-4,9900	<0,001***

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001

**p-valor<0,05

*p-valor<0,1

Tabela 44 – Modelo multilogit para inconsistência na resposta sobre violência física

Inconsistência 1				
	Efeito margin	Desvio	Estat. Z	p-valor
(CE/PE/BA)	0,0214	0,0130	1,6400	0,1000*
idade_jovem	-0,0097	0,0180	-0,5400	0,5890
escolaridade	-0,0091	0,0108	-0,8400	0,3990
raça	0,0040	0,0164	0,2500	0,8060
filhos	0,0423	0,0140	3,0200	0,003**
trabalho_status	-0,0104	0,0156	-0,6700	0,5040
raça_parcei	0,0047	0,0137	0,3400	0,7340
escolaridade_parcei	-0,0025	0,0117	-0,2100	0,8320
trabalho_parcei	0,0135	0,0189	0,7100	0,4760
parceiro_w1	-0,0135	0,0138	-0,9800	0,3270
barganha	0,0096	0,0151	0,6300	0,5260
lmp_falar	0,0067	0,0171	0,3900	0,6950
var_conhec_lmp	0,0468	0,0260	1,8000	0,0720*
violence	0,1527	0,0303	5,0400	<0,001***
Inconsistência 2				
	Efeito margin	Desvio	Estat. Z	p-valor
(CE/PE/BA)	0,0322	0,0099	3,2600	<0,001***
idade_jovem	-0,0158	0,0102	-1,5500	0,1210
escolaridade	-0,0069	0,0073	-0,9400	0,3480
raça	-0,0082	0,0107	-0,7600	0,4450
filhos	0,0259	0,0097	2,6700	0,0080**
trabalho_status	0,0190	0,0109	1,7400	0,0801*
raça_parcei	-0,0205	0,0086	-2,3700	0,0180**
escolaridade_parcei	-0,0101	0,0080	-1,2600	0,2060
trabalho_parcei	-0,0292	0,0220	-1,3300	0,1840
parceiro_w1	-0,0212	0,0103	-2,0700	0,0309**
barganha	-0,0194	0,0108	-1,8000	0,0720*
lmp_falar	-0,0073	0,0126	-0,5700	0,5650
var_conhec_lmp	-0,0092	0,0112	-0,8200	0,4140
violence	-0,0134	0,0100	-1,3400	0,1820

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001

**p-valor<0,05

*p-valor<0,1

Tabela 45 – Modelo multilogit para inconsistência na resposta sobre violência sexual

Inconsistência 1				
	Ef.marginal	Erro	Estatística z	p-valor
(CE/PE/BA)	-0,0090	0,0080	-1,1200	0,2610
idade_jovem	-0,0023	0,0107	-0,2100	0,8300
escolaridade	0,0019	0,0069	0,2800	0,7800
raça	-0,0140	0,0087	-1,6100	0,1070
filhos	0,0152	0,0091	1,6700	0,0950*
trabalho_status	0,0216	0,0102	2,1200	0,0340**
raça_parcei	-0,0150	0,0080	-1,8700	0,0610*
escolaridade_parcei	-0,0016	0,0076	-0,2100	0,8370
trabalho_parcei	-0,0081	0,0158	-0,5200	0,6060
parceiro_w1	-0,0146	0,0094	-1,5600	0,1190
barganha	-0,0074	0,0100	-0,7400	0,4600
lmp_falar	-0,0109	0,0129	-0,8500	0,3980
var_conhec_lmp	-0,0178	0,0099	-1,8000	0,0720*
violence	0,1014	0,0245	4,1400	<0,001***
Inconsistência 2				
	Ef.marginal	Erro	Estatística z	p-valor
(CE/PE/BA)	0,0231	0,0072	3,1900	<0,001***
idade_jovem	-0,0020	0,0067	-0,3000	0,7610
escolaridade	-0,0044	0,0044	-1,0100	0,3110
raça	0,0017	0,0072	0,2300	0,8170
filhos	0,0048	0,0060	0,8000	0,4230
trabalho_status	0,0087	0,0065	1,3200	0,1850
raça_parcei	-0,0046	0,0050	-0,9200	0,3570
escolaridade_parcei	0,0026	0,0048	0,5600	0,5780
trabalho_parcei	0,0005	0,0099	0,0500	0,9630
parceiro_w1	-0,0142	0,0070	-2,0500	0,0400**
barganha	-0,0113	0,0065	-1,7400	0,0820*
lmp_falar	0,0054	0,0061	0,9000	0,3700
var_conhec_lmp	-0,0062	0,0064	-0,9700	0,3320
violence	-0,0049	0,0061	-0,8000	0,4220

Fonte: Elaborado pelo autor com base nos dados da PCSVDF-Mulher.

***p-valor<0,001

**p-valor<0,05

*p-valor<0,1

4.5 Considerações finais

Este capítulo teve como objetivo verificar o processo de inconsistência na resposta de episódios de violência contra a mulher a partir da PCSVDF-Mulher. A inconsistência lógica da resposta se dá quando existe uma situação onde o “Princípio da Não Contradição” não é obedecido, no caso estudado foram explicitados que sequências do tipo (1/0/0/0), (1/1/0/0) são respostas inconsistentes, por mulheres que afirmaram que foram vítimas de alguma violência em um ano e no ano seguinte forneceram uma resposta contrária àquela do período anterior.

Na literatura, existem pelo menos oito mecanismos que exploram a inconsistência na resposta, e estão divididos em diversas áreas, sejam elas questões psicológicas, questões de desenho de pesquisa, economia/econométricas e na área mais específica de VPI (violência doméstica e saúde pública).

Para isso, foram utilizadas abordagens exploratórias e inferências com a utilização da base de dados da PCSVDF-Mulher, a partir do recorte dos dados das mulheres nos anos de 2016 e 2017, levando em consideração somente as mulheres que já tiveram parceiro e que aceitaram participar da seção EXPERIÊNCIAS DE VIOLÊNCIA DOMÉSTICA (PARCEIRO ATUAL, EX-PARCEIRO (MAIS RECENTE) OU QUALQUER OUTRO EX-PARCEIRO).

A primeira análise se deu de forma exploratória, contabilizando as sequências de respostas que foram encontradas nas duas ondas. Logo após, foi realizada uma análise das inconsistências de acordo com o tipo de violência que a mulher relatou. Ao final do levantamento, foi encontrado um percentual maior de inconsistências em respostas sobre violência emocional, tendo em vista que não é um tipo de violência bem definido e que causa confusão nas pessoas quanto a sua classificação.

É plausível que eventos mais sérios ou comportamentos extremos sejam mais facilmente lembrados e considerados dignos de relato do que daqueles que tiveram menos impacto na vida de um indivíduo. Mulheres que sofreram VPI emocional eram muito mais propensas a fornecer relatos inconsistentes do que aquelas que experimentaram VPI física ou sexual. Este é um lembrete dos impactos profundos e duradouros que a VPI física ou sexual pode ter sobre saúde e bem-estar das mulheres.

A segunda análise teve o objetivo de estudar, de forma exploratória, a frequência de respostas inconsistentes em cada variável explicativa. Para isso, foram retiradas da base de dados da PCSVDF-Mulher aspectos sobre a característica da mulher, do parceiro, características da relação dos dois, conhecimento sobre a Lei Maria da Penha, referentes a mulher e a episódios

recentes de violência. Além da análise conjunta da inconsistência, essa relação também foi feita por tipo de inconsistência, dividindo-a em dois tipos de acordo com o seu conceito.

Por fim, a análise inferencial se deu de forma estimada a partir de um modelo logit com a variável resposta sendo a mulher ter relatado uma sequência de respostas inconsistentes ou não e as variáveis de controle característica da mulher, do parceiro, características da relação dos dois, conhecimento sobre a Lei Maria da Penha, referente a mulher e questões de gênero (episódios recentes de violência). Além dessa análise, como a resposta poderia ser separada em dois tipos de inconsistência, foi estimado um modelo multilogit, com três grupos: o de mulheres consistentes, o de mulheres inconsistentes do tipo 1 e do tipo 2, sendo as mulheres consistentes como referencial.

Alguns fatores são importantes para destacar como a mulher ter filhos, o seu local de residência (município), ela viver com o parceiro nas duas ondas e a escolaridade do parceiro. Enquanto ter filho e morar em municípios mais populosos do nordeste aumenta a probabilidade de inconsistência na resposta, ela viver com o mesmo parceiro nas duas ondas e quanto maior for a escolaridade dele tem influência negativa com a probabilidade de inconsistência.

É importante observar que os resultados encontrados no presente estudo são importantes para preencher a lacuna de produção científica acerca dos relatos inconsistentes em pesquisas longitudinais, e até, principalmente, por respostas inconsistentes levarem potencialmente a estimativas imprecisas da prevalência de tais comportamentos e, em última análise, podem levar alguém a questionar a validade de quaisquer conclusões tiradas com base nestas.

Nossos resultados também são importantes para a literatura na área de “Qualidade de survey”, mas especificamente na questão de “Atrito” e “Missing data”. Ao apontar a prevalência do fenômeno da inconsistência e mecanismos plausíveis de explicação, coloca-se a necessidade de não tratar tais dados como aberrações empíricas e simplesmente descartáveis. De fato, as estimativas sugerem importantes estruturas subjacentes que merecem ser consideradas na análise econométrica.

Como sugestão para pesquisas futuras, o mesmo estudo poderá ser executado para as ondas futuras da PCSVDF-Mulher, levando em consideração que já foram realizadas coletadas de dados em 2019, 2021 e 2023

REFERÊNCIAS

- AALSMA, M.; ZIMET, G.; FORTENBERRY, J.; BLYTHE, M.; ORR, D. Reports of childhood sexual abuse by adolescents and young adults: Stability over time. **Journal of sex research**, v. 39, p. 259–63, 12 2002.
- AHERN, K.; BROCCO, R. L. Methodological issues in the effects of attrition: Simple solutions for social scientists. **Field Methods**, v. 17, n. 1, p. 53–69, 2005. Disponível em: <https://doi.org/10.1177/1525822X04271006>.
- ANDERSON, J. R. **The Architecture of Cognition**. Cambridge: Harvard University Press, 1983.
- ANDRADE, J. d. O.; CASTRO, S. d. S.; HEITOR, S. F. D.; ANDRADE, W. P. d.; ATIHE, C. C. Indicators of violence against women according to the reports of health services in the state of minas gerais-brazil. **Texto Contexto - Enfermagem**, Universidade Federal de Santa Catarina, Programa de Pós Graduação em Enfermagem, v. 25, n. 3, p. e2880015, 2016. ISSN 0104-0707. Disponível em: <https://doi.org/10.1590/0104-07072016002880015>.
- ANDRE, D. de M.; CARVALHO, J. R. **Sample Design and Cross-sectional Weights of the Brazilian PCSVDF-Mulher Study (Waves 2016 and 2017): Integrating a Refreshment Sample with an Ongoing Longitudinal Sample to Calculate IPV Prevalence**. [S. l.], 2023.
- ANDRÉ, D. de M.; CARVALHO, J. R. **Termo de referência PCSVDF₃^{Mulher}**. [S. l.], 2019.
- ANTONINO, M. **Relatório metodológico sobre desenho da amostra e ponderação amostral**. [S. l.], 2020.
- BAG, S.; GUPTA, K.; DEB, S. A review and recommendations on variable selection methods in regression models for binary data. In: _____. [S. l.: s. n.], 2022.
- BALDISSERA, S.; FERRANTE, G. Field substitution in surveys. In: _____. [S. l.: s. n.], 2019. p. 1–7. ISBN 9781118445112.
- BALDISSERA, S.; FERRANTE, G.; QUARCHIONI, E.; MINARDI, V.; POSSENTI, V.; CARROZZI, G.; MASOCCO, M.; SALMASO, S. Field substitution of nonresponders can maintain sample size and structure without altering survey estimates—the experience of the italian behavioral risk factors surveillance system (passi). **Annals of Epidemiology**, v. 24, n. 4, p. 241–245, 2014. ISSN 1047-2797. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1047279713004699>.
- BERTSIMAS, D.; MARGARITIS, G. **Global Optimization: A Machine Learning Approach**. 2023.
- CANNELL, C. F.; MILLER, P. V.; OKSENBURG, L. Research on interviewing techniques. **Sociological Methodology**, v. 12, p. 389, 1981. Disponível em: <https://api.semanticscholar.org/CorpusID:120680256>.
- CASEY, R. L.; MASUDA, M.; HOLMES, T. H. Quantitative study of recall of life events. **Journal of Psychosomatic Research**, v. 11, n. 2, p. 239–247, 1967. ISSN 0022-3999. Disponível em: <https://www.sciencedirect.com/science/article/pii/002239996790013X>.

CATALLOZZI, M.; SIMON, P. J.; DAVIDSON, L. L.; BREITBART, V.; RICKERT, V. I. Understanding Control in Adolescent and Young Adult Relationships. **Archives of Pediatrics Adolescent Medicine**, v. 165, n. 4, p. 313–319, 04 2011. ISSN 1072-4710. Disponível em: <https://doi.org/10.1001/archpediatrics.2011.32>.

CHAMRATRITHIRONG, A.; SETHAPUT, C. Fieldwork experiences related to the longitudinal study of the demographic responses to a changing environment in nang rong, 1994. In: . [S. n.], 1997. Disponível em: <https://api.semanticscholar.org/CorpusID:127987446>.

CHUDIK, A.; PESARAN, M. H.; SHARIFVAGHEFI, M. **Variable Selection in High Dimensional Linear Regressions with Parameter Instability**. 2023.

COCHRAN, W. G. Sampling techniques. John Wiley Sons, New York, p. 448, 1977.

DATAINFO. **Relatório de campo Segunda onda da PCSVDF-Mulher**. [S. l.], 2017.

DATASENADO. **Pesquisa Nacional de Violência contra a Mulher**. [S. l.], 2023.

DEMAREST, S.; MOLENBERGHS, G.; BERETE, F.; CHARAFEDDINE, R.; OYEN, H. V.; HAL, G. V. Time trends in the use of field-substitution in the belgian health interview survey. **Archives of public health = Archives belges de sante publique**, v. 80, n. 1, p. 229, November 2022. ISSN 0778-7367. Disponível em: <https://europepmc.org/articles/PMC9644564>.

DEMAREST, S.; MOLENBERGHS, G.; HEYDEN, J. Van der; GISLE, L.; OYEN, H.; WALEFFE, S.; HAL, G. V. Sample substitution can be an acceptable data-collection strategy: the case of the belgian health interview survey. **International journal of public health**, v. 62, 05 2017.

DENG, Y.; HILLYGUS, D. S.; REITER, J. P.; SI, Y.; ZHENG, S. Handling attrition in longitudinal studies: The case for refreshment samples. **Statistical Science**, Institute of Mathematical Statistics, v. 28, n. 2, p. 238–256, 2013. ISSN 08834237, 21688745. Disponível em: <http://www.jstor.org/stable/43288490>.

DOURADO, S. M.; NORONHA, C. V. Marcas visíveis e invisíveis: danos ao rosto feminino em episódios de violência conjugal. **Ciência Saúde Coletiva**, Associação Brasileira de Pós-Graduação em Saúde Coletiva (ABRASCO), v. 20, n. 9, p. 2911–2920, 2015. Disponível em: <https://www.scielo.br/j/csc/a/wjXrtsVWg3rWDQKZjtLMWGH/?format=pdf&lang=pt>.

EISNER, N. L.; MURRAY, A. L.; EISNER, M.; RIBEAUD, D. A practical guide to the analysis of non-response and attrition in longitudinal research using a real data example. **International Journal of Behavioral Development**, v. 43, n. 1, p. 24–34, 2019. Disponível em: <https://doi.org/10.1177/0165025418797004>.

ELLSBERG, M.; ARANGO, D.; MORTON, M.; GENNARI, F.; KIPLESUND, S.; CONTRERAS, M.; WATTS, C. Prevention of violence against women and girls: What does the evidence say? **The Lancet**, 11 2014.

FRIEDMAN, W. The development of memory for the times of past events. p. 394–407, 09 2013.

FUNCH, D. P.; MARSHALL, J. R. Measuring life stress: Factors affecting fall-off in the reporting of life events. **Journal of Health and Social Behavior**, [American Sociological Association, Sage Publications, Inc.], v. 25, n. 4, p. 453–464, 1984. ISSN 00221465. Disponível em: <http://www.jstor.org/stable/2136382>.

- GLICKMAN MICHAEL HUBBARD, T. L. L.; VALCIUKAS, J. A. Fall-off in reporting life events: Effects of life change, desirability, and anticipation. **Behavioral Medicine**, Taylor & Francis, v. 16, n. 1, p. 31–38, 1990. PMID: 2322655. Disponível em: <https://doi.org/10.1080/08964289.1990.9934589>.
- GNEEZY, U.; KAJACKAITE, A.; SOBEL, J. Lying aversion and the size of the lie. **American Economic Review**, v. 108, p. 419–453, 02 2018.
- GREENE, W.; HARRIS, M.; SRIVASTAVA, P.; ZHAO, X. Misreporting and econometric modelling of zeros in survey data on social bads: An application to cannabis consumption. **Health Economics**, v. 27, 08 2017.
- GROVES, R.; COUPER, M. Non-response in household interview surveys. **Population (French Edition)**, v. 55, 03 2000.
- GROVES, R.; PRESSER, S.; DIPKO, S. The role of topic interest in survey participation decisions. **Public Opinion Quarterly**, v. 68, p. 2–31, 04 2004.
- GROVES, R. M. Book. **Survey nonresponse / edited by Robert M. Groves ... [et al.]**. Wiley New York ; Chichester, 2002. xv, 500 p. : p. ISBN 0471396273. Disponível em: <http://www.loc.gov/catdir/toc/wiley021/2001045522.html>.
- GROVES, R. M.; PEYTCHEVA, E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. **The Public Opinion Quarterly**, [Oxford University Press, American Association for Public Opinion Research], v. 72, n. 2, p. 167–189, 2008. ISSN 0033362X, 15375331. Disponível em: <http://www.jstor.org/stable/25167621>.
- GROVES, R. M.; SINGER, E.; CORNING, A. Leverage-saliency theory of survey participation: Description and an illustration. **The Public Opinion Quarterly**, [Oxford University Press, American Association for Public Opinion Research], v. 64, n. 3, p. 299–308, 2000. ISSN 0033362X, 15375331. Disponível em: <http://www.jstor.org/stable/3078721>.
- HALPERN, C.; MADKOUR, A.; MARTIN, S.; KUPPER, L. Patterns of intimate partner violence victimization from adolescence to young adulthood in a nationally representative sample. **The Journal of adolescent health : official publication of the Society for Adolescent Medicine**, v. 45, p. 508–16, 11 2009.
- HEERINGA, S. G. Russia longitudinal monitoring survey (rlms) sample attrition, replenishment, and weighting in rounds v-vii. p. 1–8, 03 1997.
- HILL, Z. Reducing attrition in panel studies in developing countries. **International Journal of Epidemiology**, v. 33, n. 3, p. 493–498, 05 2004. ISSN 0300-5771. Disponível em: <https://doi.org/10.1093/ije/dyh060>.
- HSIAO, C. Panel data analysis - advantages and challenges. **SSRN Electronic Journal**, 01 2006.
- JACKSON, C.; DOBSON, A.; TOOTH, L.; MISHRA, G. Lifestyle and socioeconomic determinants of multimorbidity patterns among mid-aged women: A longitudinal study. **PLOS ONE**, v. 11, p. e0156804, 06 2016.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications in R**. 1st. ed. New York: Springer, 2013.

- KALTON, G.; MONTANARI, G.; MALIGALIG, D. Characteristics of second wave nonrespondents in a panel survey. **Proceedings of the Survey Research Methods Section, American Statistical Association**, 01 1990.
- KELFVE, S.; FORS, S.; LENNARTSSON, C. Getting better all the time? selective attrition and compositional changes in longitudinal and life-course studies. **Longitudinal and Life Course Studies**, v. 8, 01 2017.
- KOBILANSKI, F. S.; PIZZOLITTO, G. Sample substitutions in the americasbarometer 2016/17. **Vanderbilt University**, p. 1–22, 2019. ISSN 1047-2797.
- LANGELAND, W.; SMIT, J.; MERCKELBACH, H.; VRIES, G.; HOOGENDOORN, A.; DRAIJER, N. Inconsistent retrospective self-reports of childhood sexual abuse and their correlates in the general population. **Social Psychiatry and Psychiatric Epidemiology**, v. 50, 11 2014.
- LEPKOWSKI, J. M. "the treatment of wave nonresponse in panel surveys" in panel surveys. In: . [S. l.]: John Wiley and Sons, 1989.
- LEVIN, B.; KATZEN, H.; KLEIN, B.; LLABRE, M. Cognitive decline affects subject attrition in longitudinal research. **Journal of clinical and experimental neuropsychology**, v. 22, p. 580–6, 11 2000.
- LEWIS, T.; MCMICHAEL, J. Evaluating substitution as a strategy for handling u.s. postal service drop points in self-administered address-based sampling frame surveys. **Sociological Methodology**, v. 53, p. 008117502211475, 01 2023.
- LIMA-COSTA, M. F.; ANDRADE, F. B. de; SOUZA, P. R. B. d.; NERI, A. L.; DUARTE, Y. A. d. O.; CASTRO-COSTA, E.; OLIVEIRA, C. de. The Brazilian Longitudinal Study of Aging (ELSI-Brazil): Objectives and Design. **American Journal of Epidemiology**, v. 187, n. 7, p. 1345–1353, 01 2018. ISSN 0002-9262. Disponível em: <https://doi.org/10.1093/aje/kwx387>.
- LIMA-COSTA, M. F.; MAMBRINI, J. V. de M.; ANDRADE, F. B. de; SOUZA PAULO ROBERTO BORGES, J. de; VASCONCELLOS, M. T. L. de; NERI, A. L.; CASTRO-COSTA, E.; MACINKO, J.; OLIVEIRA, C. de. Cohort Profile: The Brazilian Longitudinal Study of Ageing (ELSI-Brazil). **International Journal of Epidemiology**, v. 52, n. 1, p. e57–e65, 06 2022. ISSN 0300-5771. Disponível em: <https://doi.org/10.1093/ije/dyab132>.
- LIU, M. Using machine learning models to predict attrition in a survey panel. In: _____. [S. l.: s. n.], 2020. p. 415–433. ISBN 9781118976357.
- LORENA, A.; COGNiÇÃO, C.; CARVALHO, A.; COMPUTAÇÃO, D.; COMPUTAÇÃO, I.; USP. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada; Vol. 14, No 2 (2007); 43-67**, v. 14, 12 2007.
- LOUPPE, G. **Understanding Random Forests: From Theory to Practice**. 2015.
- LOXTON, D.; POWERS, J.; TOWNSEND, N.; HARRIS, M. L.; FORDER, P. Longitudinal inconsistency in responses to survey items that ask women about intimate partner violence. **BMC Med Res Methodol**, v. 27, 02 2019.
- LOZIER, J. D.; ANIELLO, P.; HICKERSON, M. J. Predicting the distribution of sasquatch in western north america: anything goes with ecological niche modelling. **Journal of Biogeography**, v. 36, p. 1623–1627, 2009.

- LYNN, P. The use of substitution in surveys. **Surv Stat**, v. 49, p. 14–16, 06 2004.
- LYNN, P. Tackling panel attrition. In: _____. **The Palgrave Handbook of Survey Research**. Cham: Springer International Publishing, 2018. p. 143–153. ISBN 978-3-319-54395-6. Disponível em: https://doi.org/10.1007/978-3-319-54395-6_19.
- LYNN, P. Advances in longitudinal survey methodology. 04 2021.
- LYNN, P.; WATSON, N. Refreshment sampling for longitudinal surveys. 04 2021.
- MARIANO, D. Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e f-score. In: _____. [S. l.: s. n.], 2021. ISBN 9786599275326.
- MASON, R.; TRAUGOTT, M. The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective. **International Journal of Public Opinion Research**, v. 22, n. 4, p. 535–551, 10 2010. ISSN 0954-2892. Disponível em: <https://doi.org/10.1093/ijpor/edq037>.
- MCKINNEY, C.; CAETANO, R.; RAMISETTY-MIKLER, S.; NELSON, S. Childhood family violence and perpetration and victimization of intimate partner violence: Findings from a national population-based study of couples. **Annals of epidemiology**, v. 19, p. 25–32, 11 2008.
- MENDONÇA, M. F. S. de; LUDERMIR, A. B. Violência por parceiro íntimo e incidência de transtorno mental comum. **Revista de Saúde Pública**, v. 51, p. 32, 2017.
- MEYER, B.; MITTAG, N.; GOERGE, R. Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. **Journal of Human Resources**, v. 57, p. JHRv58n01_MeyerMittagGeorge, 082020.
- NICHOLSON, J. S.; DEBOECK, P. R.; HOWARD, W. Attrition in developmental psychology: A review of modern missing data reporting and practices. **International Journal of Behavioral Development**, v. 41, n. 1, p. 143–153, 2017. Disponível em: <https://doi.org/10.1177/0165025415618275>.
- NISHIMURA, R. Substitution of nonresponding units in probability sampling. 01 2015.
- NORRIS, F. Epidemiology of trauma: frequency and impact of different potentially traumatic events on different demographic groups. **Journal of consulting and clinical psychology**, v. 60, n. 3, p. 409–418, June 1992. ISSN 0022-006X. Disponível em: <https://doi.org/10.1037//0022-006x.60.3.409>.
- PACHANA, N.; BRILLEMAN, S.; DOBSON, A. Reporting of life events over time: Methodological issues in a longitudinal sample of women. **Psychological assessment**, v. 23, p. 277–81, 03 2011.
- PENDSE, G. V. **BSVM: A Banded Suport Vector Machine**. 2011.
- QUEIROZ, M. I. M. de; BIASOLI-ALVES, Z. M. M. Análise qualitativa de dados de entrevista: uma proposta. **Cadernos de Pesquisa**, v. 39, p. 27–38, 1987. Disponível em: <https://www.scielo.br/j/cp/a/KkVcMtT8TTTL7xprgXPnszL/abstract/?lang=pt>.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>.

RAPHAEL, K.; CLOITRE, M.; DOHRENWEND, B. Problems of recall and misclassification with checklist methods of measuring stressful life events. **Health psychology : official journal of the Division of Health Psychology, American Psychological Association**, v. 10, p. 62–74, 01 1991.

RIBAS, R. P.; SOARES, S. S. D. O atrito nas pesquisas longitudinais: O caso da pesquisa mensal de emprego (pme/ibge). 08 2008.

RIBISL, K. M.; WALTON, M. A.; MOWBRAY, C. T.; LUKE, D. A.; DAVIDSON, W. S.; BOOTSMILLER, B. J. Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. **Evaluation and Program Planning**, v. 19, n. 1, p. 1–25, 1996. ISSN 0149-7189. Disponível em: <https://www.sciencedirect.com/science/article/pii/0149718995000372>.

RICHMAN, W.; KIESLER, S.; WEISBAND, S.; DRASGOW, F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. **Journal of Applied Psychology**, v. 84, p. 754–775, 10 1999.

ROBERTS, A.; AUSTIN, S. B.; CORLISS, H.; VANDERMORRIS, A.; KOENEN, K. Pervasive trauma exposure among us sexual orientation minority adults and risk of posttraumatic stress disorder. **American journal of public health**, v. 100, p. 2433–41, 12 2010.

RÉGNIER-LOILLIER, N. G. A. Sample attrition and distortion over the waves of the french generations and gender survey. **Institut national d'études démographiques**, 2012.

SARDHINA, L.; MAHEU-GIROUX, M.; STÖCKL, H.; MEYER, S.; GARCIA-MORENO, C. Global, regional, and national prevalence estimates of physical or sexual, or both, intimate partner violence against women in 2018. **The Lancet**, v. 399, 02 2022.

SILVA, A. B. R. da. **Essays on the econometric of gender, domestic and intimate partner violence**. Tese (PhD thesis) – Universidade Federal do Ceará, Fortaleza-CE, 12 2018.

SILVA, V. G. da. **O ATRITO E SEUS EFEITOS NOS ESTUDOS LONGITUDINAIS**. Tese (Doutorado), 2011.

SKOGAN, W. Citizen reporting of crime. some national panel data. **Criminology**, v. 13, p. 535 – 549, 03 2006.

STINCHCOMBE, A. L.; JONES, C. Nonresponse bias for attitude questions. **The Public Opinion Quarterly**, Oxford University Press, American Association for Public Opinion Research, v. 45, n. 3, p. 359–375, 1981.

SYLASKA, K.; EDWARDS, K. Disclosure of intimate partner violence to informal social support network members: A review of the literature. **Trauma, violence abuse**, v. 15, 07 2013.

SÁNCHEZ, A.; ESCOBAL, J. Survey attrition after 15 years of tracking children in four developing countries: The Young Lives study. **Review of Development Economics**, v. 24, n. 4, p. 1196–1216, November 2020. Disponível em: <https://ideas.repec.org/a/bla/rdevec/v24y2020i4p1196-1216.html>.

T, H.; M, N.; MN, W.; GBM, H.; B, G. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. **PeerJ Preprints** 6, v. 03, 2018.

TARLING, R.; MORRIS, K. Reporting crime to the police. **British Journal of Criminology - BRIT J CRIMINOL**, v. 50, p. 474–490, 04 2010.

THOMAS, D.; FRANKENBERG, E.; SMITH, J. P. Lost but not forgotten: Attrition and follow-up in the indonesia family life survey. **The Journal of Human Resources**, [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], v. 36, n. 3, p. 556–592, 2001. ISSN 0022166X. Disponível em: <http://www.jstor.org/stable/3069630>.

THOMPSON, C. P.; SKOWRONSKI, J. J.; LARSEN, S. F.; BETZ, A. Autobiographical memory: Remembering what and remembering when. Lawrence Erlbaum Associates, 1996.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society. Series B (Methodological)**, [Royal Statistical Society, Wiley], v. 58, n. 1, p. 267–288, 1996. ISSN 00359246. Disponível em: <http://www.jstor.org/stable/2346178>.

TOURANGEAU, R.; RIPS, L. J.; RASINSKI, K. **The Psychology of Survey Response**. [S. l.]: Cambridge University Press, 2000.

UPTON, G. J. G.; FINGLETON, B.patial data analysis by example: Point pattern and qualitative data. **New York, Wiley**, 1985.

VEHOVAR, V. Field substitution and unit nonresponse. 1999. Disponível em: <https://api.semanticscholar.org/CorpusID:201633880>.

WAGENAAR, W. A. My memory: A study of autobiographical memory over six years. **Cognitive Psychology**, v. 18, n. 2, p. 225–252, 1986. ISSN 0010-0285. Disponível em: <https://www.sciencedirect.com/science/article/pii/0010028586900137>.

WATSON, N.; WOODEN, M. Identifying factors affecting longitudinal survey response. In: _____. [S. l.: s. n.], 2009. v. 1, p. 157 – 181. ISBN 9780470743874.

WEBSTER, K.; WARD, A.; DIEMER, K.; FLOOD, M.; POWELL, A.; FORSTER, K.; HONEY, N. Attitudinal support for violence against women: What a population-level survey of the australian community can and cannot tell us. **Australian Journal of Social Issues**, v. 54, 11 2018.

WITTEW, R.; HUBRICH, S. Nonresponse in household surveys: A survey of nonrespondents from the repeated cross-sectional study “mobility in cities – srv” in germany. **Transportation Research Procedia**, v. 11, p. 66–84, 12 2015.

ZABEL, J. E. An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. **The Journal of Human Resources**, [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], v. 33, n. 2, p. 479–506, 1998. ISSN 0022166X. Disponível em: <http://www.jstor.org/stable/146438>.

APÊNDICE A –

**APÊNDICE B – APÊNDICE B - FIGURAS E TABELAS ADICIONAIS DO
CAPÍTULO 03**

Tabela 46 – Combinações de respostas totais de acordo com os cortes no tempo

ID	2016		2017		Conclusiva	Possível
	LV	<12 meses	LV	<12 meses		
1	NA	NA	NA	NA	0	1
2	NA	NA	NA	1	0	0
3	NA	NA	NA	0	0	0
4	NA	NA	1	NA	0	1
5	NA	NA	1	1	0	1
6	NA	NA	1	0	0	1
7	NA	NA	0	NA	0	1
8	NA	NA	0	1	0	0
9	NA	NA	0	0	0	1
10	NA	1	NA	NA	0	0
11	NA	1	NA	1	1	0
12	NA	1	NA	0	0	0
13	NA	1	1	NA	1	0
14	NA	1	1	1	1	0
15	NA	1	1	0	1	0
16	NA	1	0	NA	1	0
17	NA	1	0	1	1	0
18	NA	1	0	0	1	0
19	NA	0	NA	NA	0	0
20	NA	0	NA	1	0	0
21	NA	0	NA	0	0	0
22	NA	0	1	NA	0	0
23	NA	0	1	1	0	0
24	NA	0	1	0	0	0
25	NA	0	0	NA	0	0
26	NA	0	0	1	0	0
27	NA	0	0	0	0	0
28	1	NA	NA	NA	0	1
29	1	NA	NA	1	1	0
30	1	NA	NA	0	0	0
31	1	NA	1	NA	1	1
32	1	NA	1	1	1	1
33	1	NA	1	0	1	1
34	1	NA	0	NA	1	1
35	1	NA	0	1	1	0
36	1	NA	0	0	1	1
37	1	1	NA	NA	0	1
38	1	1	NA	1	1	0
39	1	1	NA	0	0	0
40	1	1	1	NA	1	1

	2016		2017			
ID	LV	<12 meses	LV	<12 meses	Conclusiva	Possível
41	1	1	1	1	1	1
42	1	1	1	0	1	1
43	1	1	0	NA	1	1
44	1	1	0	1	1	0
45	1	1	0	0	1	1
46	1	0	NA	NA	0	1
47	1	0	NA	1	1	0
48	1	0	NA	0	0	0
49	1	0	1	NA	1	1
50	1	0	1	1	1	1
51	1	0	1	0	1	1
52	1	0	0	NA	1	1
53	1	0	0	1	1	0
54	1	0	0	0	1	1
55	0	NA	NA	NA	0	1
56	0	NA	NA	1	1	0
57	0	NA	NA	0	0	0
58	0	NA	1	NA	1	1
59	0	NA	1	1	1	1
60	0	NA	1	0	1	1
61	0	NA	0	NA	1	1
62	0	NA	0	1	1	0
63	0	NA	0	0	1	1
64	0	1	NA	NA	0	0
65	0	1	NA	1	1	0
66	0	1	NA	0	0	0
67	0	1	1	NA	1	0
68	0	1	1	1	1	0
69	0	1	1	0	1	0
70	0	1	0	NA	1	0
71	0	1	0	1	1	0
72	0	1	0	0	1	0
73	0	0	NA	NA	0	1
74	0	0	NA	1	1	0
75	0	0	NA	0	0	0
76	0	0	1	NA	1	1
77	0	0	1	1	1	1
78	0	0	1	0	1	1
79	0	0	0	NA	1	1
80	0	0	0	1	1	0
81	0	0	0	0	1	1

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Tabela 47 – Perguntas sobre a Lei Maria da Penha

q404_1 - É uma lei para punir o agressor por cometer violência doméstica	Concordo completamente Concordo um pouco Nem concordo nem discordo Discordo um pouco Discordo completamente Não sabe/não respondeu
q404_4 - Nos termos dessa lei, apenas a vítima pode denunciar o agressor	Concorda completamente Concordo um pouco Nem concordo nem discordo Discordo um pouco Discordo completamente Não sabe/não respondeu
q404_5 - Nos termos dessa lei, depois de denunciar uma agressão, uma mulher, em alguns casos, não pode retirar sua queixa	Concorda completamente Concordo um pouco Nem concordo nem discordo Discordo um pouco Discordo completamente Não sabe/não respondeu
q404_6 - A Lei Maria da Penha aplica-se igualmente a lésbicas, travestis, transsexuais e heterossexuais	Concorda completamente Concordo um pouco Nem concordo nem discordo Discordo um pouco Discordo completamente Não sabe/não respondeu
q404_7 - A Lei Maria da Penha aplica-se a casos de violência cometida por ex-maridos/ex-parceiros/examantes	Concorda completamente Concordo um pouco Nem concordo nem discordo Discordo um pouco Discordo completamente Não sabe/não respondeu
q404_8 - Mulheres vítimas de violência doméstica que não tenham condições de contratar um advogado podem ir a uma Delegacia Especializada no Atendimento à Mulher	Concorda completamente Concordo um pouco Nem concordo nem discordo Discordo um pouco Discordo completamente Não sabe/não respondeu

Tabela 48 – Percentual de inconsistência de acordo com as variáveis com características da mulher e do parceiro (Inconsistência geral)

	VIO GER		VIO EMO		VIO FIS		VIO SEX	
	INC	p-valor	INC	p-valor	INC	p-valor	INC	p-valor
Mulher								
Município								
(CE/PE/BA)	0,1561	0,0141	0,1514	0,0333	0,0976	<0,001***	0,0552	0,1137
demais	0,2087		0,1950		0,1572		0,077	
Idade								
jovem	0,1858	0,3360	0,1784	0,173	0,133	0,0909*	0,0636	0,6404
adulta	0,1568		0,1406		0,0916		0,0739	
Escolaridade								
ens fundamental	0,2116	0,0132**	0,1928	0,0868*	0,1473	0,0091**	0,0728	0,685
ens médio tecnico	0,1888		0,176		0,1367		0,0665	
ens superior/pós	0,1211		0,1282		0,0714		0,0556	
Raça								
branca	0,1831	0,734	0,1707	0,903	0,1292	0,4907	0,0711	0,1119
não branca	0,1720		0,1759		0,1119		0,0435	
Filhos								
sim	0,1229	0,0040***	0,1080	<0,001***	0,0541	<0,001***	0,0401	0,0499
não	0,1972		0,1901		0,1461		0,0726	
Trabalho								
diminuiu status	0,2692	0,0199**	0,2162	0,1063	0,1604	0,4455	0,0522	0,1828
permaneceu	0,1674		0,161		0,1207		0,0619	
aumentou	0,2099		0,2105		0,1366		0,0957	
Parceiro								
Raça								
branca	0,1868	0,4464	0,1772	0,4685	0,1323	0,3179	0,0728	0,1192
não branca	0,1682		0,1603		0,1118		0,0496	
Escolaridade								
ens fundamental	0,2053	0,2038	0,2164	0,0078**	0,1578	0,0217**	0,0714	0,7997
ens médio tecnico	0,1725		0,1537		0,1161		0,0632	
ens superior/pós	0,145		0,1314		0,0763		0,0584	
Trabalho								
sim	0,2143	0,1358	0,1964	0,2576	0,1367	0,6063	0,0662	0,9798
não	0,1727		0,1658		0,123		0,0653	
Mulher_parceiro								
sim	0,1996	0,1861	0,1993	0,0378**	0,134	0,5209	0,0853	0,0228**
não	0,1696		0,1553		0,1208		0,0535	
Barganha								
menor	0,1661	0,4277	0,1598	0,4468	0,1218	0,8655	0,0669	0,9571
mesma	0,1909		0,1783		0,1305		0,0636	
maior	0,2041		0,2039		0,1176		0,0693	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Tabela 49 – Percentual de inconsistência de acordo com as variáveis com características da mulher e do parceiro (Inconsistência tipo 1)

	VIO GER		VIO EMO		VIO FIS		VIO SEX	
	INC	p-valor	INC	p-valor	INC	p-valor	INC	p-valor
Mulher								
Município								
(CE/PE/BA)	0,1160	0,6862	0,1100	0,8205	0,0695	0,1577	0,0462	0,8025
demais	0,1246		0,1049		0,0913		0,0422	
Idade								
jovem	0,1248	0,2872	0,1139	0,1175	0,0841	0,2447	0,0438	0,9700
adulta	0,0975		0,0781		0,0598		0,0467	
Escolaridade								
ens fundamental	0,1246	0,0639*	0,1129	0,3963	0,0822	0,0954*	0,0431	0,7397
ens médio técnico	0,1330		0,1124		0,0907		0,0480	
ens superior/pós	0,0781		0,0842		0,0489		0,0370	
Raça								
branca	0,1168	0,5351	0,1026	0,2614	0,0788	0,8687	0,0488	0,1348
não branca	0,1326		0,1276		0,0839		0,0268	
Filhos								
sim	0,0831	0,0326**	0,0679	0,0119**	0,0382	0,0030**	0,0247	0,0732*
não	0,1305		0,1190		0,0917		0,0499	
Trabalho								
diminuiu status	0,2404	<0,001***	0,1982	<0,001***	0,1415	0,0229**	0,0435	0,0394**
permaneceu	0,1008		0,0923		0,0710		0,0387	
aumentou	0,1657		0,1474		0,0984		0,0798	
Parceiro								
Raça								
branca	0,1177	0,7655	0,1049	0,6968	0,0792	0,9816	0,0508	0,0983*
não branca	0,1250		0,1132		0,0811		0,0302	
Escolaridade								
ens fundamental	0,1159	0,5431	0,1207	0,3595	0,0905	0,4609	0,0491	0,7653
ens médio técnico	0,1247		0,1028		0,0750		0,0421	
ens superior/pós	0,0916		0,0803		0,0611		0,0365	
Trabalho								
sim	0,1316	0,5898	0,1179	0,6100	0,0791	0,9878	0,0523	0,5684
não	0,1173		0,1051		0,0800		0,0424	
Mulher_parceiro								
sim	0,1096	0,4041	0,1052	0,8869	0,0830	0,8071	0,0563	0,1108
não	0,1263		0,1090		0,0779		0,0371	
Barganha								
menor	0,1036	0,2462	0,1020	0,5812	0,0744	0,6552	0,0395	0,6097
mesma	0,1333		0,1087		0,0865		0,0466	
maior	0,1327		0,1359		0,0686		0,0594	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher

Tabela 50 – Percentual de inconsistência de acordo com as variáveis com características da mulher e do parceiro (Inconsistência tipo 2)

	VIO GER		VIO EMO		VIO FIS		VIO SEX	
	INC	p-valor	INC	p-valor	INC	p-valor	INC	p-valor
Mulher								
Município								
(CE/PE/BA)	0,0401	0,001**	0,0414	<0,001***	0,0281	<0,001***	0,0090	0,0011**
demais	0,0841		0,0901		0,0659		0,0349	
Idade								
jovem	0,0611	0,9765	0,0645	0,9899	0,0489	0,3150	0,0198	0,6096
adulta	0,0593		0,0625		0,0319		0,0272	
Escolaridade								
ens fundamental	0,0870	0,0543*	0,0799	0,1863	0,0652	0,0443**	0,0296	0,4383
ens médio técnico	0,0559		0,0637		0,0460		0,0185	
ens superior/pós	0,0430		0,0440		0,0226		0,0185	
Raça								
branca	0,0662	0,1256	0,0681	0,2719	0,0504	0,1432	0,0223	0,7123
não branca	0,0394		0,0483		0,0280		0,0167	
Filhos								
sim	0,0399	0,1137	0,0401	0,061*	0,0159	0,0064**	0,0154	0,5557
não	0,0667		0,0710		0,0544		0,0227	
Trabalho								
diminuiu status	0,0288	0,1852	0,0180	0,1139	0,0189	0,3038	0,0087	0,5110
permaneceu	0,0666		0,0688		0,0497		0,0232	
aumentou	0,0442		0,0632		0,0383		0,0160	
Parceiro								
Raça								
branca	0,0691	0,0795*	0,0723	0,0848*	0,0531	0,0804*	0,0219	0,9052
não branca	0,0432		0,0470		0,0307		0,0194	
Escolaridade								
ens fundamental	0,0894	0,0156**	0,0957	0,0065**	0,0673	0,0226**	0,0223	0,9883
ens médio técnico	0,0479		0,0508		0,0411		0,0211	
ens superior/pós	0,0534		0,0511		0,0153		0,0219	
Trabalho								
sim	0,0827	0,1269	0,0786	0,3362	0,0576	0,3813	0,0139	0,4740
não	0,0555		0,0607		0,0431		0,0229	
Mulher_parceiro								
sim	0,0900	<0,001***	0,0941	<0,001***	0,0509	0,5689	0,0290	0,1483
não	0,0433		0,0463		0,0429		0,0164	
Barganha								
menor	0,0625	0,8416	0,0578	0,6710	0,0475	0,9441	0,0274	0,2942
mesma	0,0576		0,0696		0,0440		0,0169	
maior	0,0714		0,0680		0,0490		0,0099	

Elaborado pelo autor com base nos dados da PCSVDF-Mulher