



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Análise da Exclusão Escolar das Famílias Pobres:
Estudo de Caso do Público Beneficiário do Programa
Bolsa Família**

Ereny Nunes Sena

Dissertação apresentada como requisito parcial para qualificação do
Mestrado Profissional em Computação Aplicada

Orientador
Prof. Dr. Marcelo Ladeira

Brasília
2024

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

Na NUNES SENA, ERENY
Análise da Exclusão Escolar de Famílias Pobres: Um Estudo de Caso do Público Beneficiário do Programa Bolsa Família / ERENY NUNES SENA; orientador Marcelo Ladeira; co-orientador Luís Henrique da Silva de Paiva. -- Brasília, 2024.
142 p.

Dissertação(Mestrado Profissional em Computação Aplicada)
-- Universidade de Brasília, 2024.

1. Qui-Quadrado de Independência. 2. Programa Bolsa Família. 3. Exclusão Escolar. 4. Integração de Bases de Dados. 5. Discretização de Dados. I. Ladeira, Marcelo, orient. II. da Silva de Paiva, Luís Henrique, co-orient. III. Título.

Resumo

Esta pesquisa tem como objetivo estudar os fatores que contribuem para a exclusão escolar do público de crianças e adolescentes beneficiárias do Programa Bolsa Família (PBF), com base nos dados fornecidos pelo Ministério do Desenvolvimento Social, Ministério da Educação e do Instituto Brasileiro de Geografia e Estatística. Os dados utilizados são do período de 2014 a 2019. O estudo parte da análise quantitativa dos dados das famílias beneficiárias do PBF que apresentam registros de faltas na escola, cuja reincidência leva ao descumprimento da condicionalidade de educação e, conseqüentemente, à suspensão do benefício. Como resultado, pretende-se compreender os fatores que ajudem a explicar a exclusão escolar deste público, possibilitando caracterizar o perfil de famílias em risco de vulnerabilidade social, dado que a ocorrência e reiteração de faltas é um alerta à trajetória escolar adequada ou pode indicar risco social acentuado.

O perfil visa caracterizar o ambiente domiciliar, econômico e social em que vivem famílias que tiveram suspensão de benefício por descumprimento da condicionalidade de educação. As questões de pesquisa abordadas são: a) "o arranjo familiar monoparental está associado à suspensão de benefício PBF decorrente de exclusão escolar?", b) "a suspensão de benefício PBF decorrente de exclusão escolar tem associação com localidades que possuem maiores PIB Per Capita municipal?", c) "infraestrutura domiciliar mais precária ou membros familiares em estado de maior vulnerabilidade estão associados à suspensão do benefício do PBF decorrente da exclusão escolar?", e d) "altos índices de violência estão associados à suspensão de benefício PBF decorrente de exclusão escolar?".

As discrepâncias nas métricas de absentismo entre os estados brasileiros sugerem diferentes dinâmicas regionais que influenciam as suspensões, indicando a existência de duas macrorregiões que denominamos SERRP e NÃO-SERRP. SERRP compreende os estados de São Paulo, Espírito Santo, Rio de Janeiro, Rio Grande do Sul e Paraná. NÃO-SERRP compreende os demais estados brasileiros.

Palavras-chave: Exclusão, Programa Bolsa Família, Condicionalidade Educação, Trajetória Escolar

Abstract

This research aims to study the contributing factors to school exclusion among children and adolescents who are beneficiaries of the Bolsa Família Program (PBF), based on data provided by the Ministry of Social Development, Ministry of Education, and the Brazilian Institute of Geography and Statistics. The data analyzed covers the period from 2014 to 2019. The study employs quantitative analysis of data from PBF beneficiary families with records of school absenteeism, whose recurrence leads to non-compliance with the educational conditionality, resulting in the suspension of the benefit. The objective is to understand the factors that help explain school exclusion among this population, enabling for the characterization of families at risk of social vulnerability, as recurring absenteeism signals either a warning for proper school attendance or indicates heightened social risk.

The study aims to characterize the household, economic, and social environment of families whose benefits were suspended due to non-compliance with the educational conditionality. The research questions addressed include: a) "Is single-parent family arrangement associated with the suspension of PBF benefits due to school exclusion?" b) "Is the suspension of PBF benefits due to school exclusion associated with municipalities with higher PIB per capita?" c) "Are precarious household infrastructure or family members in greater states of vulnerability associated with the suspension of PBF benefits due to school exclusion?" d) "Are high rates of violence associated with the suspension of PBF benefits due to school exclusion?"

Discrepancies in absenteeism metrics across Brazilian states suggest distinct regional dynamics influencing benefit suspensions, pointing to the existence of two macro-regions referred to as SERRP and NON-SERRP. SERRP comprises the states of São Paulo, Espírito Santo, Rio de Janeiro, Rio Grande do Sul, and Paraná, while NON-SERRP encompasses the remaining Brazilian states.

Keywords: Exclusion, Family Purse Program, Education Conditionality, School Trajectory

Sumário

1	Introdução	1
1.1	Definição do Problema	1
1.2	Hipóteses de Pesquisa	4
1.3	Justificativa	5
1.4	Contribuição Esperada	6
1.5	Objetivos	7
1.6	Organização do Trabalho	8
2	Fundamentação Teórica e Revisão do Estado da Arte	9
2.1	Conceitos de Programas de Transferência Condicionada de Renda	9
2.1.1	A Trajetória Escolar na Educação Pública Brasileira	10
2.2	Exclusão Escolar do Programa Bolsa Família	11
2.3	Testes de Independência	13
2.4	Teste de Significância	15
2.4.1	Testes de Independência e Conceitos do Poder de uma Análise	17
2.4.2	Testes de Qui-quadrado em Tabelas de Contingência e o Tamanho de Efeito Ômega	21
2.5	Mineração de Dados	27
2.5.1	Conceitos de Mineração de Dados	27
2.5.2	CRISP-DM	27
2.6	Técnicas de Discretização de Dados	28
3	Metodologia	32
3.1	Obtenção e Preparação do Dados	34
3.2	Integração e Discretização de Dados	37
3.3	Modelo de Análise das Condicionais da Educação	47
3.3.1	Delimitação das Macrorregiões	47
3.3.2	Definição de Variáveis	48
3.3.3	Método Estatístico	48

4	Resultados Obtidos	50
4.1	Estudo do Absenteísmo Escolar	50
4.1.1	Análise das Faltas	50
4.1.2	Macroregiões SERRP e NÃO-SERRP	54
4.1.3	Análise da Monoparentalidade	56
4.1.4	Análise da Violência	57
4.1.5	Análise do PIB Per Capita	65
4.2	Associações e o Tamanho do Efeito	67
4.2.1	Associação entre Monoparentalidade e a Exclusão Escolar	69
4.2.2	Associação entre Violência e a Exclusão Escolar	71
4.2.3	Associação entre PIB Per Capita, Violência e a Exclusão Escolar	76
4.2.4	Associação entre o Perfil Familiar e a Exclusão Escolar	83
4.2.5	Associação entre Monoparentalidade, PIB Per Capita, Violência e a Exclusão Escolar	92
4.2.6	Monoparentalidade, Perfil Familiar e a Exclusão Escolar	98
5	Conclusões e Trabalhos Futuros	108
5.1	Conclusões	108
5.2	Trabalhos Futuros	116
	Referências	118
	Anexo	120
I	Indicadores de Faltas	121

Lista de Figuras

1.1	Efeitos Gradativos por Descumprimento das Condicionalidades do PBF. . .	3
2.1	Modelo de Referência CRISP-DM	29
3.1	Representação do Projeto Conceitual de Integração de Bases de Dados . . .	35
3.2	<i>Workflow</i> de Integração e Pareamento.	38
3.3	Seleção de Famílias Suspensas.	42
3.4	Obtenção do Dados do IBGE de Localização Geográfica.	43
3.5	Obtenção dos Dados de Família e Pessoas da Família e Marcação de Suspensão.	43
3.6	Marcação da Composição Familiar Quanto a Monoparentalidade.	44
3.7	Obtenção do Dados de Violência.	45
3.8	Obtenção do PIB Per Capita Municipal Médio	45
3.9	Discretização de Dados.	46
3.10	Quadro Resumido de Variáveis.	49
4.1	Brasil: Famílias Beneficiárias <i>versus</i> Famílias Suspensas.	52
4.2	Proporção das Famílias Suspensas Monoparentais em cada Estado	57
4.3	(%) de Famílias Suspensas e Não-suspensas por de Famílias por Faixas de Suicídio. Proporção Nacional	60
4.4	(%) de Famílias Suspensas e Não-suspensas por Faixas de Homicídio de Homens. Proporção Nacional.	61
4.5	(%) de Famílias Suspensas e Não-suspensas por Faixas de Mortes Violentas. Proporção Nacional.	61
4.6	(%) de Famílias Suspensas e Não-suspensas por Faixas de Homicídio de Mulheres. Proporção Nacional.	62
4.7	Distribuição de Famílias Suspensas e Não-suspensas por Faixas de PIB per Capital. Proporções: Brasil e por Suspensão.	66
4.8	Monoparentalidade e Macrorregião	69
4.9	Suicídio e Macrorregião.	72

4.10	Morte Violenta por Causa Indeterminada e Macrorregião.	73
4.11	Crosstab de Homicídio de Mulheres e Macrorregião.	74
4.12	Homicídio de Homens e Macrorregião	75
4.13	Suicídio e Média Categorizada de PIB Per Capita de NÃO-SERRP	76
4.14	Suicídio e Média Categorizada de PIB Per Capita de SERRP	77
4.15	Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita de NÃO-SERRP	78
4.16	Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita de SERRP	79
4.17	Homicídio de Mulheres e Média Categorizada de PIB Per Capita de NÃO- SERRP	80
4.18	Homicídio de Mulheres e Média Categorizada de PIB Per Capita de SERRP	81
4.19	Homicídio de Homens e Média Categorizada de PIB Per Capita de NÃO- SERRP	82
4.20	Homicídio de Homens e Média Categorizada de PIB Per Capita de SERRP.	83
4.21	Escoamento Sanitário e Macrorregião	84
4.22	Raça-Cor e Macrorregião	85
4.23	Infraestrutura Domiciliar: Iluminação do Domicílio e Macrorregião.	86
4.24	Material Predominante nas Paredes do Domicílio e Macrorregião	87
4.25	Abastecimento de Água do Domicílio e Macrorregião.	89
4.26	Quantidade de Cômodos do Domicílio e Macrorregião.	90
4.27	Quantidade de Dormitórios do Domicílio e Macrorregião	91
4.28	Raça-Cor, Monoparentalidade e Macrorregião	98
4.29	Escoamento Sanitário, Monoparentalidade e Macrorregião	100
4.30	Iluminação do Domicílio, Monoparentalidade e Macrorregião	102
4.31	Abastecimento de Água, Monoparentalidade e Macrorregião	103
4.32	Material Predominante nas Paredes, Monoparentalidade e Macrorregião . .	104
4.33	Quantidade de Cômodos do Domicílio, Monoparentalida e Macrorregião .	105
4.34	Quantidade de Dormitórios do Domicílio, Monoparentalidade e Macrorregião	106
5.1	Relação de Efeitos de Características Pessoais de Membros da Família, Infraestrutura Familiar e Violência	113
5.2	Relação de Efeitos de Características Pessoais de Membros da Família, Infraestrutura Familiar e Violência Controlado para Monoparentalidade . .	114
5.3	Relação de Efeitos de PIB Per Capita e Violência.	115
5.4	Relação de Efeitos de Efeitos de PIB Per Capita Violência Controlado para Monoparentalidade	115

I.1	Crosstab de Suicídio e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.	125
I.2	Crosstab de Suicídio e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.	126
I.3	Crosstab de Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.	127
I.4	Crosstab de Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.	128
I.5	Crosstab de Homicídio de Mulheres e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.	129
I.6	Crosstab de Homicídio de Mulheres e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.	130
I.7	Crosstab de Homicídio de Homens e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.	131
I.8	Crosstab de Homicídio de Homens e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.	131

Lista de Tabelas

2.1	Proporções P1 (Hipótese 1) na Distribuição Conjunta de Local de Residência (Urbano/Rural) e Suspenso	23
2.2	Proporções P0 (Hipótese H0) da Distribuição Conjunta de Local de Residência e Suspenso	24
2.3	Tabela de Equivalência de Medidas Simétricas: ω , C , ϕ e ϕ'	25
3.1	Calendário de Registro de Frequências no Sistema Presença	36
3.2	Variáveis para Estudo da Frequência dos Beneficiários	39
4.1	Frequência do Indicador de Faltas por Motivo	51
4.2	Faltas de Suspensos: Proporção por Região e Zonas Geográficas	52
4.3	Distribuição de Motivos de Faltas por Região (%)	53
4.4	Análise do Absenteísmo Escolar no Brasil	55
4.5	Detalhamento das Variáveis da Dimensão Social Municipal	58
4.6	Estatística Descritiva do Modelo de Discretização por Entropia MDL	59
4.7	Detalhamento da Variável da Dimensão Econômica Municipal.	65
4.8	Tabela de Equivalência de Medidas Simétricas: ω em Termos de C , ϕ e ϕ'	68
I.1	Distribuição Nacional do Público Suspenso e do Público Beneficiário	121
I.2	Estatística Descritiva da Discretização por Entropia MDL das Taxas de Suicídio	122
I.3	Estatística Descritiva da Discretização por Entropia MDL das taxas de Morte Indeterminada	123
I.4	Estatística Descritiva da Discretização por Entropia MDL das Taxas de Homicídio de Mulheres	123
I.5	Estatística Descritiva da Discretização por Entropia MDL das Taxas de Homicídio de Homens	124
I.6	Resumo da Discretização do Valor Médio de PIB Per Capital Municipal no Período de 2014 a 2019	124

Lista de Abreviaturas e Siglas

CRISP-DM *Cross Industry Standard Process for Data Mining.*

Capítulo 1

Introdução

Neste Capítulo são apresentados a definição do problema, a justificativa para estudar a exclusão escolar de crianças e adolescentes beneficiárias do Programa Bolsa Família, a contribuição esperada com esta pesquisa, as hipóteses a serem investigadas e os objetivos geral e específicos.

1.1 Definição do Problema

O programa Bolsa Família (PBF), criado em 2003, é o principal programa de transferência de renda com condicionalidades do Governo Federal, cujo principal objetivo é o combate à pobreza e à desigualdade da distribuição da renda no Brasil. A seleção das famílias potencialmente elegíveis para o programa é feita a partir do Cadastro Único para Programas Sociais (CadÚnico), cujo preenchimento está sob a responsabilidade dos municípios. A gestão do programa é descentralizada entre União, estados, municípios e o Distrito Federal.

O PBF é uma transferência condicionada de renda na qual há condições que as famílias precisam atender para se tornarem ou se manterem beneficiárias. Os critérios de elegibilidade do PBF dependem da condição de renda da família. Entre as condicionalidades para sua permanência no programa, estão o cumprimento de certas tarefas relacionadas à frequência escolar e aos cuidados com a saúde de seus membros. Segundo a Lei 10.836/2004, ao ingressar no programa, a família beneficiária se compromete a cumprir as condicionalidades do Bolsa Família nas áreas de saúde e educação, que são: a) manter as crianças e adolescentes em idade escolar na escola, com frequência mínima, durante o período letivo, de 85% para os com idade entre 6 e 15 anos e de 75% para os com idade de 16 e 17 anos; b) cumprir com os seguintes cuidados básicos em saúde: atendimento do calendário de vacinação para as crianças entre 0 e 7 anos, acompanhamento do seu

crescimento e desenvolvimento a cada 6 meses, e cumprimento da agenda pré-natal para as gestantes.

No ano de 2017, o PBF teve mais de 13,8 milhões de famílias beneficiadas em quase todos os municípios brasileiros. Neste ano, o programa repassou aproximadamente R\$ 2,5 bilhões ao mês aos municípios¹, 16,7 milhões de crianças e adolescentes tiveram informações de frequência escolar registradas, e mais de 3 milhões de crianças e adolescentes faltaram às aulas em, pelo menos, um dos cinco períodos de acompanhamento anual da frequência escolar. Segundo a iniciativa *Trajetórias Escolares: Desigualdades e Diversidades*² desenvolvido pelo MEC e executado pela Universidade Federal de Santa Catarina - UFSC, a maioria das crianças e dos adolescentes, na faixa etária de 6 a 17 anos que está fora da escola ou em risco de exclusão escolar³ integra os 25% da população mais pobre do País. Esses estudantes correspondem àqueles que têm apresentado maior dificuldade de progressão e conclusão dos estudos na idade adequada. Neste sentido, segundo os dados do Censo Escolar de 2017, a taxa de reprovação no 3º ano do Ensino Fundamental é de 9,7% no Brasil. No 1º ano do Ensino Médio chega a 15,8%. Essas populações compreendem as mesmas consideradas nas políticas de inclusão social, as que são beneficiárias do Programa Bolsa Família, sobretudo por representarem as camadas mais pobres da população do País.

A concessão da transferência de renda às famílias não está vinculada à comprovação de conformidade às condicionalidades, a família incluída no Bolsa Família é que inicia a verificação das condicionalidades para a continuidade do recebimento do benefício. São necessários reiterados não cumprimentos das condicionalidades, ao longo de um período de um ano e meio, para que ocorra o cancelamento da transferência de renda do programa. O cancelamento do benefício poderá ocorrer após quatro descumprimentos consecutivos em um prazo total de 18 meses. É necessário estar fora da escola ao longo de todo um ano para ocorrer o cancelamento da transferência de renda. Passados os 18 meses, os não-cumprimentos anteriores deixam de ser considerados.

Quando uma família descumpra os compromissos do Bolsa Família, são aplicados efeitos gradativos que variam conforme o histórico de descumprimento da família⁴. As consequências dos descumprimentos são:

- Advertência: a família é comunicada de que algum integrante deixou de cumprir condicionalidades, mas não deixa de receber o benefício;

¹Disponível em Painel de Dados Abertos da Matriz de Informações Sociais

²Disponível em Projeto Trajetórias Escolares: Desigualdades e Diversidades (2017-2018)

³Disponível em Sinopse Estatística da Educação Básica 2017, INEP.

⁴Disponível no site da Secretaria de Desenvolvimento Social do DF - Condicionalidades do Programa Bolsa Família

- Bloqueio: o benefício fica bloqueado por um mês, mas pode ser sacado no mês seguinte junto com a nova parcela;
- Suspensão: o benefício fica suspenso por dois meses, e a família não poderá receber os valores referentes a esse período;
- Cancelamento: a família deixa de participar do PBF.

Para a progressão de um efeito para o seguinte, considera-se o intervalo de seis meses. O prazo de seis meses, no entanto, não vale para a progressão da suspensão para o cancelamento, que obedece a regras específicas, conforme pode ser visto na Figura 1.1.



Figura 1.1: Efeitos Gradativos por Descumprimento das Condicionalidades do PBF.

As famílias em descumprimento são notificadas pela gestão federal do PBF, com indicação de qual integrante da família descumpriu algum dos compromissos (se foi relacionado à área de saúde ou de educação) e o efeito aplicado. Portanto, para que uma família tenha seu benefício cancelado, são necessárias as seguintes ocorrências:

1. a família estar em fase de suspensão;
2. a família estar em Acompanhamento Familiar (AF);

3. se após 12 meses, contados do dia em que tiver começado a vigorar a suspensão e o AF, a família apresentar novo descumprimento com efeito de suspensão.

O acompanhamento da condicionalidade da educação é realizado por meio do monitoramento da frequência escolar, ao longo de cinco períodos do ano, no âmbito do projeto Presença⁵ do Ministério da Educação e envolve uma rede de mais de 56 mil profissionais da educação, em todos os estados, DF e municípios do País. Anualmente, são recolhidos bimestralmente as frequências de mais de 16 milhões de alunos. O Sistema de Condicionalidades (Sicon) do Ministério do Desenvolvimento Social é uma ferramenta de apoio à gestão intersetorial que integra as informações do acompanhamento de condicionalidades nas áreas de Saúde e Educação, consolidando as informações de frequência escolar, do calendário de vacinação e das consultas pré-natais oriundas dos sistemas dos Ministérios da Educação e da Saúde, possibilitando monitorar as famílias beneficiárias para uma gestão mais eficiente e eficaz do PBF.

1.2 Hipóteses de Pesquisa

O presente projeto de pesquisa, além de procurar identificar os fatores que estão associados à exclusão escolar de crianças e adolescentes do Programa Bolsa Família, introduz uma inovação: a utilização de dados massivos, contendo todo o público beneficiário do PBF entre os anos de 2014 a 2019, integrado à base de dados do Cadastro Único, à base de frequência escolar dos beneficiários PBF, à base do IBGE sobre tipologias do território e Produto Interno Bruto - PIB - per capita municipal e à base do IPEA sobre violência. Estes dados, integrados, totalizam mais de 82 milhões de registros de família com 23 atributos que permitem compor o perfil das famílias, sob diferentes perspectivas, com faltas registradas pela gestão federal do programa.

Esta pesquisa busca evidências para as seguintes hipóteses (H):

1. Família monoparental enfrenta mais dificuldades para acompanhar a educação dos filhos? Neste sentido, é verificado se o arranjo familiar monoparental está associado à suspensão de benefício PBF, decorrente de exclusão escolar.
 - H1: A exclusão escolar não está associada à existência de monoparentalidade na família.

⁵O sistema Presença é um Sistema de Acompanhamento da Frequência Escolar do Bolsa Família, gerenciado pelo Ministério da Educação, em parceria com as secretarias municipais e estaduais de Educação e do Distrito Federal, que é responsável pela coleta e registro da frequência escolar dos beneficiários.

2. Famílias que residem em localidade economicamente mais desenvolvidas, com grande disponibilidade de ofertas de empregos, podem ser mais afetadas pela exclusão escolar em função de priorizarem assumir postos de trabalho? Nesta pesquisa é testada a hipótese da suspensão de benefício PBF decorrente de exclusão escolar ter associação com localidades que possuem maiores PIB Per Capita municipal.
 - H2: A exclusão escolar não está associada às famílias residirem em localidades com maiores PIB Per Capita municipal.
3. Famílias residentes em domicílios mais precários ou cujos membros estão em situações de maior vulnerabilidade (situação de rua ou deficiência), podem estar em condições acentuadas de risco social que afetam a permanência das crianças na escola? No estudo é avaliado se infraestrutura domiciliar mais precária ou membros familiares em estado de maior vulnerabilidade estão associados à suspensão do benefício PBF, decorrente da exclusão escolar.
 - H3: A exclusão escolar não está associada às famílias possuírem infraestrutura domiciliar mais precária ou terem membros em situação mais vulnerável.
4. Famílias residentes em áreas com maior índice de violência têm suas rotinas familiares negativamente afetadas por crimes que impactam a continuidade das crianças na escola? Aqui é analisado se altos índices de violência estão associados à suspensão de benefício PBF, decorrente de exclusão escolar.
 - H4: A exclusão escolar não está associada às famílias residirem em localidades com maior índice de violência.

1.3 Justificativa

A efetividade dos programas de transferência condicionada de renda, como instrumento para o aumento do capital humano dos beneficiários, tem sido estudado em diversos países de baixa e média renda. De acordo com Currello et al.[1], as regras estabelecidas nas condicionalidades baseiam-se na perspectiva de que elas sejam utilizadas para induzir comportamentos que contribuem para a promoção social das famílias. O aumento da escolarização e cumprimento de agendas de saúde promovem o aumento do capital humano das populações mais pobres, possibilitando novas perspectivas de inserção socioeconômica e a superação do ciclo de pobreza entre as gerações.

Esta pesquisa propõe identificar o perfil das famílias suspensas, visando auxiliar no entendimento dos fatores associados à exclusão escolar dos alunos beneficiários do PBF, compreendendo as razões que levam os alunos a faltarem à escola e sofrerem suspensão

de benefícios. Esta compreensão é também um importante mecanismo de diagnóstico dos riscos para a trajetória escolar adequada do público suspenso do PBF, dado que a transferência de renda e suas condicionalidades miram num objetivo de longo prazo, visando a ruptura do ciclo intergeracional da pobreza das populações mais pobres e excluídas. O entendimento dos principais fatores relacionados à suspensão por faltas à escola permite localizar e identificar famílias que necessitem de apoio prioritário do poder público. Millán et al.[2] argumentam que programas de transferência condicionada de renda, adotados ao redor do mundo há mais de 20 anos, objetivam a redução da pobreza em curto prazo, por meio de transferências de renda, e redução da pobreza em longo prazo, por meio do investimento em capital humano, tendo as séries completas de escolaridade dos beneficiários como elemento de eficácia de adoção destes programas. Esta pesquisa procura identificar o perfil das famílias com exclusão escolar, numa abordagem ampliada para além das informações contidas no Cadastro Único, investigando os fatores associados às famílias suspensas do PBF, em função da interrupção escolar de seus filhos, embora eles tenham recebido benefícios financeiros para motivar a presença na escola.

1.4 Contribuição Esperada

A análise da exclusão escolar do público beneficiário do Programa Bolsa Família é importante na busca de subsídios para compreender o fenômeno, com vistas a minimizá-lo. Num âmbito mais sistêmico, permite compreender mais detalhadamente a exclusão escolar de crianças e adolescentes cujas famílias sofreram a suspensão do benefício em função de faltar às aulas e subsidiar a implementação de estratégias mais focalizadas, no sentido de evitar ou minimizar a exclusão de alunos nas escolas. Ao se conhecer melhor as variáveis com grande influência na exclusão escolar, pode-se adequar os serviços assistenciais às necessidades e peculiaridades das famílias. O poder público poderá atuar preventivamente, já nos primeiros sinais de alerta presentes na família. Além disso, a partir do conhecimento prévio do perfil das famílias envolvidas no problema, há possibilidade de aperfeiçoar o processo de acompanhamento familiar feito pelo poder público, visando otimizar o enfrentamento de dificuldades sociais e o vínculo de confiança entre os assistentes e os beneficiários.

No que se refere à área de Computação, essa pesquisa inicia com o enriquecimento da base de dados do PBF com colunas das bases de dados do IBGE (tipologias do território e PIB per capita municipal) e da base de dados do IPEA (índices municipais de violência), totalizando mais de 82 milhões de registros de família, com 23 atributos. A seguir, é feita uma análise de associações de variáveis em toda a população beneficiária entre os anos de 2014 a 2019, utilizando o método do Qui-Quadrado para verificar a existência

de associação e o algoritmo V de Crámer para medir o tamanho da associação, num conjunto de dados populacionais, visando identificar o perfil de alunos suspensos e o de não-suspensos. Será analisado o evento suspensão de benefício PBF, considerando aspectos familiares, sociais e econômicos do aluno e quais variáveis estão associadas à suspensão do benefício.

1.5 Objetivos

O objetivo geral desta pesquisa consiste em identificar os fatores que estão associados aos beneficiários do PBF suspensos em função da exclusão escolar, utilizando-se das bases de dados fornecidas pelo Ministério do Desenvolvimento Social, Ministério da Educação, IBGE e IPEA, focalizando na população com alguma suspensão válida de benefício entre os anos de 2014 e 2019⁶.

Para a consecução do objetivo desta pesquisa, serão considerados os seguintes objetivos específicos:

- integrar em um único ambiente as bases de dados governamentais Cadastro Único incremental com marcação PBF, SICON, Presença, Divisão Territorial Brasileira, PIB Per Capita e Mapa da Violência e preparar os dados para aplicação de testes de independência;
- desenvolver e disponibilizar um modelo estatístico aplicado a grandes volumes de dados, lançando mão de estratégias que garantam alta confiabilidade dos resultados ao utilizar dados populacionais massivos;
- verificar as características sociais, econômicas e familiares que possam explicar a suspensão escolar de crianças e adolescentes beneficiárias do PBF por meio da abordagem de associação e traçar o perfil das famílias que descumprem as condicionalidades da educação por estarem em situação acentuada de vulnerabilidade e risco social.
- verificar se existe variação nos graus de associação entre as subpopulações considerando critérios de distribuição regional, local de residência e gênero;

⁶Os dados posteriores a 2019 não foram utilizados nesta pesquisa em função da interrupção do registro de frequência escolar, conforme Portaria Ministério do Desenvolvimento Social nº 335, de 20 de março de 2020, que suspendeu a aplicação dos efeitos decorrentes do descumprimento das condicionalidades, a contar do início de abril de 2020, dentre outras medidas em decorrência da emergência em saúde pública de importância nacional do COVID-19

1.6 Organização do Trabalho

O restante desta dissertação está estruturada em quatro capítulos. O Capítulo 2 apresenta a fundamentação teórica e a revisão do estado da arte, O Capítulo 3 apresenta a metodologia e as etapas para o desenvolvimento do estudo. No Capítulo 4 são discutidos os resultados obtidos e no Capítulo 5 são apresentadas as conclusões e trabalhos futuros.

Capítulo 2

Fundamentação Teórica e Revisão do Estado da Arte

Este capítulo apresenta alguns conceitos importantes para a compreensão deste trabalho, como programas de transferência condicionada de renda, conceitos e as tarefas comuns de análise de dados, a aplicação do teste de independência com 5% de significância e da medida do tamanho de efeito por meio do algoritmo V de Cramér.

2.1 Conceitos de Programas de Transferência Condicionada de Renda

As lições globais e as evidências sobre programas de rede de segurança social estimulam mais de 130 países emergentes a introduzir e expandir este tipo de programa, inspirados pela forte base de evidências internacionais sobre sua eficácia, incluindo documentos sobre impactos em dimensões econômicas e sociais. Dentre estes, estão os programas de transferência condicionada de renda (*Conditional Cash Transfer - CCT*), que fornecem dinheiro para as famílias pobres dependendo de determinado comportamento, geralmente investimentos em capital humano, como enviar crianças à escola ou levá-las aos centros de saúde.

Banerjee et al.[3] afirmam que as redes de segurança social estão entre as intervenções mais bem avaliadas de desenvolvimento humano, com evidências que confirmam impactos na redução da pobreza, redução da desigualdade e promoção da prosperidade compartilhada, por meio do estímulo às economias locais; investimento do capital humano em educação, saúde e nutrição; e na promoção de melhores oportunidades de emprego no futuro e no aumento da resiliência. Eles afirmam que há evidência sustentando que ao relaxar as múltiplas restrições enfrentadas pelos pobres, as transferências previsíveis,

mesmo em pequenas quantidades, têm efeitos que vão muito além de apenas abordar a pobreza e as desigualdades de renda, permitindo aumento do investimento em infraestrutura familiar e atividades produtivas. Para Pires [4], em relação a educação, a exigência da frequência escolar mínima propõe que o incremento da frequência escolar de crianças e jovens de famílias pobres promova melhores condições de ingresso no mercado de trabalho após finalizada sua trajetória escolar, aumentando a renda e, conseqüentemente, fazendo com que saiam da condição de pobreza em relação à geração de seus pais.

Millian et al.[2] revisaram a literatura sobre os impactos de longo prazo de programas de transferência condicional de renda (ou CCTs) no capital humano e seus resultados observados em dois estágios do ciclo de vida: do útero à primeira infância e da primeira infância/adolescência à idade adulta. Eles concluíram que as evidências de curto prazo deixam claro que os CCTs podem levar a ganhos em nutrição e saúde para crianças pequenas e ganhos na escolaridade para crianças mais velhas ou adolescentes ao possibilitar níveis de escolaridade mais elevados. Afirmam ainda que há evidências de impactos positivos sobre a aprendizagem em todos os países avaliados.

Na avaliação de CCTs realizada por Das et al.[5], sugere-se que a maior parte da renda transferida por CCTs alcançam dois principais objetivos. O primeiro objetivo é restaurar a eficiência social gerando mudanças mais amplas no nível da comunidade - as chamadas externalidades - uma vez que, associada às condicionalidades, induzem mudanças nas preferências das famílias e modificam suas atitudes e comportamentos. O segundo objetivo é melhorar a equidade entre diferentes classes sociais utilizando as transferências condicionadas como mecanismo para induzir a auto-seleção de público alvo, direcionando renda adicional às famílias conforme as condições de seleção impostas. Debrauw et al.[6] avaliaram dados domiciliares para analisar o impacto do PBF sobre os resultados de escolaridade dos beneficiários em idade escolar de 6 a 17 anos e concluíram que a participação escolar é significativamente maior entre as meninas, principalmente as mais velhas, residentes das áreas urbanas, diferenciando os resultados por sexo, idade e localização.

2.1.1 A Trajetória Escolar na Educação Pública Brasileira

Em 2018, o Ministério da Educação lançou a *Iniciativa Trajetórias Escolares: Desigualdades e Diversidades*, que surgia da constatação de que a maioria das crianças e dos adolescentes, na faixa etária de 6 a 17 anos, que está fora da escola ou em risco de exclusão escolar (repetência, abandono e evasão escolar)⁷ integra os 25% da população mais pobre do País e corresponde àqueles que têm apresentado maior dificuldade de progressão e conclusão dos estudos na idade adequada. Ao comparar as informações de matrículas

⁷Segundo dados do Censo Escolar de 2017, a taxa de reprovação no 3º ano do Ensino Fundamental é de 9,7% no Brasil. No 1º ano do Ensino Médio chega a 15,8%. Fonte: INEP 2017

divulgadas pelo INEP para 2017, constatou-se que metade deles corresponde as mesmas populações consideradas nas políticas de inclusão do MEC e, mais especificamente, àquelas que são beneficiárias do PBF, sobretudo por representarem as camadas mais pobres da população do País.

O MEC é o órgão responsável pelo acompanhamento da frequência escolar de beneficiários do PBF por meio do sistema Presença. Esse acompanhamento é registrado bimestralmente, em cinco períodos do ano. Nos casos em que crianças e adolescentes não cumprem a frequência escolar mínima requerida pelo PBF, a escola identifica e registra os motivos de baixa frequência.

Conforme disposto nas Diretrizes da Iniciativa⁸, os dados produzidos a partir do acompanhamento escolar dos beneficiários do PBF representam relevantes diagnósticos socioeducacionais e territoriais, que podem subsidiar o poder público na proposição de soluções para a superação de fenômenos da educação como a repetência, o abandono e a evasão escolar. Nesse contexto, a proposta da *Iniciativa Trajetórias Escolares: Desigualdades e Diversidades* faz conexão com o PBF, pois tem o propósito de mitigar estes fenômenos enfrentados por crianças e adolescentes em idade escolar de 6 a 17 anos, em situação de pobreza, beneficiários do PBF, objetivando fomentar estratégias que promovam a superação dos principais desafios que dificultam o acesso, a permanência e do sucesso escolar na idade adequada.

2.2 Exclusão Escolar do Programa Bolsa Família

O Instituto de Pesquisa Econômica Aplicada (IPEA) tem dedicado diversos estudos aos temas relacionados a entender os efeitos das condicionalidades da educação sobre o público beneficiário do PBF. Dentre estes, o estudo de Paiva et al.[7] dedicou-se a avaliar se a cobertura e o monitoramento da condicionalidade relacionada à educação do Bolsa Família estão associados a qualquer mudança nos resultados educacionais. Como resultado, concluiu-se que não foi encontrado um resultado de associação positiva entre a cobertura do Bolsa Família e as taxas de abandono e progressão. Entretanto, o monitoramento da frequência escolar foi negativamente associado às taxas de evasão e positivamente associado à progressão escolar. Eles afirmam que esses resultados parecem fazer sentido no contexto brasileiro, onde o acesso à educação (nos primeiros nove anos letivos) é quase universal, as taxas de evasão têm nível muito baixo, a progressão é relativamente alta e um conjunto de políticas para apoiar a frequência escolar está em vigor há muito tempo.

O impacto do PBF sobre a educação de seus beneficiários também foi pesquisado por Amaral et al.[8], que usaram dados de pesquisas realizadas em 2005 e 2009 pelo *Centro*

⁸Disponível no sítio trajetoriasescolares.mec.gov.br

de Desenvolvimento e Planejamento Regional para avaliar o impacto do programa na permanência dos jovens na escola, utilizando uma amostra de 15.426 famílias. Reynolds[9] avaliou o impacto do PBF sobre a escolaridade, lançando mão de dados obtidos pela *Pesquisa Nacional por Amostra de Domicílios* (PNAD), envolvendo aproximadamente 150.000 famílias. Diferente desses estudos, nossa pesquisa utiliza toda a base de dados do PBF e não apenas uma amostra dos dados.

Cireno et al.[10] compararam o desempenho e o percurso escolar de beneficiários e não beneficiários do PBF do 5º e 9º ano de 2011, separadamente, usando dados do Sistema Presença do MEC e da Prova Brasil, cruzados com dados do Cadastro Único. A partir desse cruzamento, foi analisada a associação por meio do teste t de significância das diferenças entre ser ou não beneficiário do PBF, o desempenho na Prova Brasil, a reprovação, o abandono escolar e a distorção idade-série. Os resultados indicaram que os beneficiários do programa apresentam menor taxa de abandono e distorção idade-série no 5º e no 9º anos, sugerindo que o Programa exerce a importante função de manter na escola alunos advindos de famílias pobres. Este público têm, em média, um desempenho pior na Prova Brasil que os não-beneficiários e evidências sugeriram que a condicionalidade de educação tem efeitos positivos sobre a trajetória escolar destas crianças, o que significa que a alta exposição ao PBF apresenta uma diminuição na desigualdade educacional destes alunos, quando comparados ao restante dos alunos de escola pública, ajudando na diminuição do ciclo intergeracional da pobreza e diminuindo a desvantagens que essa condição tem gerado no percurso escolar dessas crianças.

Oliveira e Soares[11] estudaram o impacto do PBF na repetência escolar de seus beneficiários e concluíram que dentre os principais motivos relacionados à repetência dos alunos pobres, destacam-se altas taxas de retenção de alunos do sexo masculino e portadores de necessidades especiais (PNEs) oriundos de famílias com menor escolaridade, de escolas com piores infraestrutura e que dependem administrativamente dos municípios ou estados. Afirmam que o perfil de residências mais precárias está associado aos piores retornos para suas crianças, mas que o PBF ajuda tais famílias na garantia de condições mais dignas para seus filhos. Foram encontradas evidências de que o PBF contribui para a redução da repetência escolar de quem o recebe.

Para Amaral e Monteiro [8], ao avaliar o impacto das condicionalidades da educação do PBF sobre a evasão escolar de crianças beneficiadas, utilizando dados de 2005 e 2009 sobre a Avaliação de Impacto do Programa Bolsa Família (AIBF) conduzido pelo Ministério do Desenvolvimento Social, estimaram a chance de as crianças abandonarem a escola em 2005 e 2009 para três limites de renda familiar diferentes, levando em conta características relacionadas à família, mãe e filho. As análises concluíram que crianças que viviam em domicílios beneficiados pelo Bolsa Família tiveram uma chance significativamente menor

de abandono escolar em 2005. Entretanto, os dados de 2009 não foram estatisticamente significativos, embora os resultados apontassem para uma diminuição nas taxas de evasão devido ao impacto do Bolsa Família.

2.3 Testes de Independência

Uma das formas de testar a diferença entre grupos e sua significância é utilizar teste de hipóteses. Existem duas principais categorias de hipóteses: hipótese nula e hipótese alternativa[12]. A hipótese nula sugere a ausência de efeito ou associação, enquanto a hipótese alternativa argumenta que existe uma diferença entre pelo menos duas populações em estudo. Os testes estatísticos de hipóteses calculam a probabilidade do evento pesquisado ocorrer assumindo-se que a hipótese nula seja verdadeira. Isso é conhecido como valor de **p** (ou **p-value**). Se o valor de **p** calculado for menor que o nível de significância, pode-se rejeitar a hipótese nula, favorecendo a hipótese alternativa, indicando a presença de diferença ou associação entre os grupos estudados. Este estudo adotou o teste não paramétrico, ou seja, que não depende de parâmetros populacionais e as hipóteses foram definidas para avaliar a existência ou não de associação entre duas variáveis, sendo verificadas por meio da aplicação do teste de independência com nível de significância de 5%.

O modelo aplicado explora a associação bivariada entre características das famílias e a suspensão do benefício. Do ponto de vista social e econômico, analisa aspectos como a violência e o PIB per capita municipal. Do ponto de vista familiar, analisa as características da infraestrutura domiciliar e características de vulnerabilidade dos membros das famílias, como presença de monoparentalidade, portador de deficiência, morador de rua, entre outras. A associação é verificada por meio de teste de independência a um nível de significância de 5%.

O tipo de teste não-paramétrico escolhido foi o teste de qui-quadrado a um nível de significância de 5% a fim de examinar a associação entre características da família e a suspensão do benefício. Trata-se de uma decisão metodológica que se fundamenta nos tipos de variáveis envolvidas, no tamanho dos dados disponíveis e no objetivo do estudo. Esta abordagem estatística é utilizada para esclarecer e validar hipóteses de associação em variáveis descritivas, ou seja, a existência ou não de relações subjacentes aos dados categóricos em estudo. Dentre os motivos relacionados ao desenho metodológico utilizado, pode-se citar:

- **Identificação de Relações Potenciais:** atributos sociais ou comportamentais assim como a suspensão do programa Bolsa Família são variáveis complexas que podem estar relacionadas de várias maneiras. Através da análise bivariada, pode-se

investigar a relação direta entre variáveis, o que nos permite identificar tendências e padrões nos dados.

- **Hipótese de Independência:** A hipótese nula em um teste Qui-quadrado assume que não há associação entre as variáveis em estudo (no caso desta pesquisa, entre vários atributos e suspensão do Bolsa Família). A hipótese alternativa, por outro lado, sugere que existe uma associação. Ao aplicar o teste Qui-quadrado são avaliadas se a probabilidade de as diferenças observadas nas proporções de duas ou mais populações são estatisticamente significativas ou se poderiam ter ocorrido ao acaso.
- **Validação Estatística:** O teste Qui-quadrado fornece uma base sólida para não refutar ou para refutar hipóteses e conclusões. A decisão por rejeitar ou não rejeitar a hipótese nula H_0 é sempre do tipo sim ou não, mutuamente exclusiva, complementar e exaustiva, mas sempre associada a riscos na tomada de decisão[13]. A utilização de um nível de significância de 5% implica na disposição de aceitar um risco de 5% de cometer o erro de rejeitar a hipótese nula quando ela é verdadeira. Isso torna a análise construída robusta e confiável.
- **Interpretação Significativa:** Resultados que demonstrem uma associação estatisticamente significativa, como por exemplo, associação entre altos níveis de violência municipal e a suspensão do Bolsa Família - considerando tratar-se de uma pesquisa baseada em dados de magnitude populacionais ao invés de dados amostrais - suscita suposições robustas para o apoio às famílias alvo, suporte às políticas públicas e orientação de futuras intervenções.

Um algoritmo de discretização baseado em MDL (*Minimum Description Length*) por entropia categorizou as variáveis numéricas, quais sejam, as taxas médias de violência e o valor médio PIB Per Capita do período entre 2014 a 2019. O procedimento foi configurado para gerar até 10 faixas para as variáveis de violência e até 3 faixas para a variável de PIB Per Capita, variando do menor ao maior valor encontrado. Ao produzir categorias para os valores médios de violência e de PIB Per Capita, o algoritmo discriminou a população residente em áreas mais ou menos violentas e mais ou menos desenvolvidas, permitindo a utilização do teste de qui-quadrado de independência usando **crosstabs** ou tabelas de contingência. A discretização dos valores dos dados da violência e do PIB per capita foi motivada por:

- **Simplificação:** uma das características desafiadoras do conjunto de dados desta pesquisa é a presença de milhares de valores contínuos que medem diferentes aspectos da violência e do PIB per capita dos municípios ao longo de seis anos (2014 a

2019). Lidar com essa quantidade maciça de valores de maneira direta seria complexo e poderia dificultar a interpretação e a análise. A discretização, ou seja, a categorização desses valores emerge como uma solução prática para simplificar a análise.

- **Redução de Dimensionalidade:** a discretização não apenas simplifica o conjunto de dados, mas também ajuda na redução da dimensionalidade, tornando-o mais gerenciável. Ao agrupar os valores contínuos em faixas categóricas, pode-se criar uma estrutura mais clara e acessível para avaliar a associação com a suspensão do Bolsa Família.
- **Interpretação Significativa:** a categorização de variáveis como violência e PIB per capita permite avaliar a relação entre atributos em diferentes níveis, como por exemplo baixa, média e alta e a suspensão do benefício PBF de forma mais fácil de interpretação. Essa abordagem facilita a comunicação de resultados complexos.
- **Facilidade na Comparação:** a discretização permite ainda a comparação direta entre subgrupos de populações com diferentes níveis de violência ou PIB per capita. Isso ajuda a entender se as variações na violência ou no desenvolvimento municipal estão relacionadas às diferenças na suspensão do benefício PBF.

Portanto, a escolha pelo modelo metodológico aplicado neste estudo não apenas permite a análise das associações, mas também facilita a interpretação e a compreensão da relação entre os complexos aspectos populacionais e a suspensão PBF. Essa abordagem é guiada para o contexto da pesquisa, envolvendo um conjunto de dados volumoso e heterogêneo, ao mesmo tempo em que mantém a solidez metodológica.

2.4 Teste de Significância

A análise estatística em estudos que utilizam dados populacionais difere dos procedimentos usuais aplicados em dados amostrais, principalmente no que diz respeito à relevância do teste de hipótese. Em pesquisas amostrais, conforme descrevem [14] e [15], a incerteza estatística decorre do fato de a amostra ser apenas uma representação da população, o que torna necessário formular hipóteses (nula e alternativa) para, então, inferir parâmetros populacionais. Entretanto, quando se dispõe de toda a população, não há incerteza quanto à representatividade, tornando-se menos pertinente o procedimento tradicional de hipótese nula versus hipótese alternativa [16, 17].

Nesse contexto, ganha destaque o teste de significância, cujo objetivo é avaliar a força estatística dos resultados observados sem a pretensão de inferir sobre parâmetros desco-

nhecidos — afinal, os dados englobam a totalidade da população [18]. Diferentemente dos testes de hipótese, em que se busca rejeitar ou não rejeitar H_0 com base em uma amostra, a aplicação de um teste de significância em dados populacionais enfoca a magnitude ou a relevância estatística de uma possível associação, diferença ou comportamento no conjunto de dados disponível.

Nesta pesquisa, por não se tratar de uma amostra, mas da população como um todo, o procedimento clássico de formular uma hipótese nula (assumindo a inexistência de efeito ou diferença) e uma hipótese alternativa (sugerindo a existência de efeito) perde parte de seu sentido inferencial. Em estudos populacionais, a pergunta “Isso é verdadeiro para a população?” já está respondida pela própria natureza dos dados, pois não se está inferindo para além do que já se conhece — toda a população está contemplada no estudo [16]. Portanto, o que se faz necessário é quantificar e avaliar a relevância estatística dos resultados encontrados. Por exemplo, se se observa uma diferença entre dois subgrupos da população, o teste de significância vem para evidenciar se essa diferença é estatisticamente relevante ou apenas um padrão numérico que, embora real, pode não ser considerado substancial para uma determinada aplicação ou hipótese de trabalho.

Autores como [14] e [17] ressaltam que os testes de significância podem ser aplicados tanto em contextos amostrais quanto em contextos populacionais. A diferença crucial está na finalidade:

- **Quando há amostras:** Busca-se *inferir* se determinado resultado pode ser extrapolado para a população maior, lidando com a incerteza amostral.
- **Quando há a totalidade da população:** Não há incerteza de representatividade, de modo que o foco recai sobre *constatar ou avaliar a intensidade* dos padrões ou associações verificadas, sem a necessidade de inferir algo além do conjunto de dados analisado [18].

Ou seja, se já se conhece a totalidade de indivíduos ou unidades que compõem o fenômeno em estudo, o teste de significância serve para demonstrar se certos achados, como diferenças estatisticamente relevantes entre subgrupos, são robustos o suficiente para serem considerados “significativos” dentro desse universo de observação [16]. Ao utilizar testes de significância em vez do tradicional teste de hipótese, o pesquisador foca na robustez do resultado, ao invés de formular um cenário de incerteza amostral. Além disso, evita equívocos teóricos, pois não se está testando algo “desconhecido” sobre a população, e sim quantificando a relevância dos padrões já observados nos dados [17]. Também mantém o rigor metodológico e estatístico, uma vez que a análise de significância continua válida para confirmar se a diferença ou relação encontrada é passível de ser

considerada estatisticamente forte — ainda que não seja inferência em seu sentido mais estrito [14].

Cohen[12] argumenta que antes que os dados sejam coletados, faz-se necessário o estabelecimento dos padrões de probabilidades apropriados para os resultados da investigação de modo que forneça uma base para a rejeição ou não da hipótese nula, e portanto, para a corroborar ou não com a existência de evidências do fenômeno sob teste. Deve-se, portanto, selecionar um nível de probabilidade *alfa* (α) suficientemente pequeno (como por exemplo 0,01 ou 0,05), de modo que seja possível fazer a seguinte afirmação sobre os dados amostrais: “Se a hipótese nula for verdadeira, a probabilidade do resultado obtido não é maior do que α ”, ou seja, é um resultado estatisticamente significativo. Se for possível fazer essa afirmação, já que α é pequeno, diz-se ter rejeitado a hipótese nula “com um critério de significância α ” ou “no nível de significância α ”. Se, por outro lado, o estudo concluir que a probabilidade é maior que α , não é possível fazer a afirmação acima e nem rejeitar a hipótese nula, ou seja, deve-se considerá-la “sustentável” ou “aceita-a”, tudo em um nível de significância definido antecipadamente. No caso do teste de hipótese não-paramétrico, [19] Kevin L. Delucci apresenta ainda os principais erros e acertos ao se utilizar os testes de qui-quadrado para a análise de tabelas de contingência (análise bivariada), apontando abordagens complementares e alternativas para sua modelagem em dados qualitativos ordenados ou não.

2.4.1 Testes de Independência e Conceitos do Poder de uma Análise

O procedimento *crosstabs* é a ferramenta básica para examinar hipóteses, ou seja, a existência ou não de relação entre duas variáveis categóricas, possivelmente controlando para variáveis adicionais de controle (também chamadas de camadas). O procedimento de *crosstabs* oferece testes de independência e medidas de associação de acordo com os dados, se nominais ou ordinais. Neste estudo, o teste de independência foi utilizado como teste estatístico para avaliar a associação entre variáveis de características da família, de violência e de PIB Per Capita e a suspensão de benefícios em diferentes subpopulações: público suspenso e não-suspenso, monoparental e não-monoparental, residente em SERRP e NÃO-SERRP. O teste de independência é um método comumente usado para testar a associação entre duas variáveis categóricas e determinar se elas são independentes uma da outra. Ele é aplicável tanto em amostras quanto em populações completas (censos), como é o caso desta pesquisa.

A ideia central de um teste de hipótese sobre um parâmetro ou comparação de parâmetros de populações é a de supor verdadeira a hipótese em questão e verificar se a

amostra observada é verossímil ou mais provável, ou seja, se está refletindo o que há de mais próximo do real estado das condições testadas dentro do ambiente em que se conduz a pesquisa [13]. O tipo não-paramétrico ou de livre distribuição, utilizado nesta pesquisa, tem por objetivo auxiliar o pesquisador na resolução de problemas nos quais o pesquisador deseja testar uma hipótese baseada em alguma distribuição teórica de probabilidade, homogeneidade de amostras ou elaborada por raciocínio inferencial, bem como na verificação ou ajustamento de frequências ou proporções populacionais. No caso de teste de independência, procura-se medir o grau de associação entre as variáveis e colocar à prova as seguintes hipóteses:

1. H_0 = As variáveis são independentes ou não existe associação entre as variáveis.
2. H_1 = As variáveis não são independentes, ou seja, elas apresentam algum grau de associação entre si

O cálculo efetuado baseia-se na Equação 2.1. Esse teste foi proposto pelo estatístico Karl Pearson [20].

$$\chi_t^2 = \chi_c^2 = \sum_{i=1}^k \left(\frac{(f_{oi} - f_{ei})^2}{f_{ei}} \right) \quad (2.1)$$

onde χ_t^2 é a estatística do teste, χ_c^2 é o qui-quadrado calculado, f_{oi} é a frequência observada e f_{ei} é a frequência esperada ou teórica e K é o número de categorias ou atributos em que foi dividida a amostra. As frequências observadas são obtidas na amostra e as frequências esperadas, teóricas ou calculadas são resultantes da hipótese H_0 formulada. As frequências esperadas encontram-se sempre sujeitas a um particular H_0 e uma hipótese comumente admitida é que as duas classificações são independentes uma da outra.

Para aplicação desse teste, os dados são dispostos em uma *crosstabs* de i linhas e k colunas, ou seja, é uma tabela que mostra como duas características ou atributos dependem uma da outra. A estatística do teste Qui-quadrado possui distribuição teórica de probabilidade de qui-quadrado com o número de graus de liberdade ν . O grau de liberdade ν para $i > 1$ e $k > 1$, onde “ i ” é o número de linhas e “ k ” é o número de colunas da tabela de contingência. A definição $\nu = (i - 1)(k - 1)$, nas tabelas $i \times k$, ocorre quando as frequências esperadas podem ser calculadas sem que se tenha que estimar os parâmetros populacionais por meio das estatísticas amostrais do estimador sob estudo. No caso deste estudo, dado que o conjunto de dados tem magnitude populacional, a análise com o Qui-quadrado fornece uma forte evidência sobre a realidade, em vez de inferências sobre os resultados de uma amostra.

Outro componente do nível de significância diz respeito a definição exata da natureza de existência do fenômeno, como por exemplo, a direcionalidade/não-direcionalidade (unicaudal/bicaudal) da afirmação da alternativa da hipótese nula. Quando a pesquisa

trabalha no contexto de comparação de parâmetros (por exemplo média, proporção ou coeficiente de correlação) para duas populações A e B, pode-se definir a existência de um fenômeno de duas maneiras diferentes:

1. Considera-se que o fenômeno existe se os parâmetros de A e B diferem. Nenhuma direção na diferença, como A maior que B, é especificada, de modo que desvios em qualquer direção da hipótese nula constitui evidência contra ela. Como qualquer cauda das diferenças da distribuição pode contribuir com alfa, isso geralmente é chamado de teste bicaudal ou bilateral.
2. O fenômeno só é considerado existente apenas se os parâmetros A e B diferem em uma direção especificada antecipadamente, por exemplo, A maior que B. Neste caso, desvios da hipótese nula apenas na direção especificada constituem uma evidência contra ela. Como apenas uma cauda das diferenças da distribuição pode contribuir com alfa, isso geralmente é chamado de teste unicaudal ou unilateral.

Quando a pesquisa adota o componente bicaudal para avaliação das diferenças nas distribuições populacionais, como no caso desta pesquisa, o critério de significância em um teste bicaudal no nível de significância de 0,05, simbolizado por alfa (α), diz duas coisas:

1. o fenômeno cuja existência está em questão é entendido como ser manifestado por qualquer diferença entre os valores dos parâmetros das duas populações;
2. que o padrão de prova é um resultado amostral que ocorreria menos que 5% do tempo se a hipótese nula fosse verdadeira.

A combinação da probabilidade e da lateralidade numa única entidade como o nível de significância é conveniente porque essa combinação define antecipadamente a “região crítica”, isto é, o intervalo de valores do resultado que leva a rejeição da hipótese nula e, conseqüentemente, o intervalo de valores que leva à sua não-rejeição. O critério de significância estabelecido, digamos 0,05, especifica o risco de rejeitar erroneamente a hipótese nula (ou seja, risco não superior a 0,05).

Conceitos de Poder do Teste

O poder de um teste é a probabilidade de que ele rejeitará a hipótese nula quando a mesma for falsa, ou seja, a probabilidade de que o teste resulte na conclusão de que o fenômeno existe. Esta medida é dada por $(1 - \beta)$, onde β é a probabilidade de se cometer um erro tipo II. O poder de um teste estatístico depende de três parâmetros: o critério de significância, confiabilidade (ou precisão) dos resultados da amostra e do “tamanho do efeito”, ou seja, o grau em que o fenômeno existe [12]. O nível de significância *alfa*,

também é conhecido como erro Tipo I ou a probabilidade de se rejeitar erroneamente a hipótese nula quando ela é verdadeira. O complemento do poder $(1 - \beta)$, simbolizado com Beta (β), é denominado erro Tipo II e representa a taxa de “erro” de não rejeitar uma hipótese nula quando a mesma é falsa.

Para [12], a confiabilidade de um valor amostral é a proximidade com o qual pode-se esperar que ele se aproxime de um valor populacional. Ela sempre depende do tamanho da amostra (e pode ou não ser diretamente dependente da unidade de medida, do valor populacional e do formato da distribuição populacional). Por exemplo, considerando a média aritmética de uma variável $X(\bar{X})$, sua confiabilidade pode ser estimada pelo erro padrão da média conforme a fórmula 2.2:

$$SE_{\bar{X}} = \sqrt{\frac{s^2}{n}} \quad (2.2)$$

onde s^2 é a estimativa usual (da amostra aleatória) da variância populacional de X e n , é o número de unidades independentes na amostra (isto é, o tamanho da amostra). Nem todos os testes estatísticos envolvem a definição explícita do erro padrão de um valor amostral. Entretanto, qualquer que seja a confiabilidade amostral de que se dependa, ela sempre depende do tamanho da amostra.

A dependência entre confiabilidade e n é evidente pela fórmula 2.2. Quanto maior o tamanho da amostra (dado que todas as outras condições são iguais), menor o erro e maior a confiabilidade ou precisão dos resultados. A relação com o poder fica clara: quanto maior a precisão dos resultados da amostra (dado que todas as outras condições são iguais) mais claramente o fenômeno testado pode se manifestar, podendo assim, ser formulada a relação direta entre tamanho da amostra e poder. Por conseguinte, aumentos no tamanho da amostra aumentam o poder estatístico. Dado que esta pesquisa se baseia em dados de tamanho populacional, a confiabilidade e o poder são assegurados.

Para o tamanho do efeito (ES do inglês *Effect Size*), em circunstâncias em que duas populações estão sendo comparadas, a hipótese nula geralmente assume a forma “a diferença no valor dos parâmetros relevantes é zero”. Um teste para verificar se um conjunto de $k \geq 2$ proporções populacionais é igual pode ser realizado por meio da estatística qui-quadrado. A hipótese nula aqui é que a variância das proporções populacionais é zero (um valor exato), uma condição que só pode ocorrer quando todas forem iguais. No caso desta pesquisa, uma das hipóteses nulas poderia ser descrita como: “a diferença nas proporções de suspensos e não suspensos que possuem escoamento sanitário do tipo fossa rudimentar é zero”, ou de forma equivalente, que o tamanho do efeito do escoamento sanitário do tipo fossa rudimentar sobre a suspensão é zero. Assim, pode-se pensar na hipótese nula como a circunstância em que as diferenças nas variáveis independentes, as k populações,

não tem efeito (tem um tamanho de efeito zero) nas proporções da variável dependente. Assim, a ausência de um fenômeno em estudo é expressa por uma hipótese nula que especifica um valor exato para um parâmetro populacional, adequado à forma como o parâmetro populacional se manifesta. Sem pretender qualquer implicação necessária de causalidade, é conveniente usar a expressão “tamanho de efeito” para significar “o grau em que um fenômeno está presente na população” ou “o grau em que a hipótese nula é falsa”. A hipótese nula (não existência do fenômeno) significa que o tamanho do efeito é zero. Assim, o ES pode ser tratado como um parâmetro que assume o valor zero quando a hipótese nula é verdadeira e algum valor específico não nulo quando a hipótese nula é falsa. Equivalentemente, o tamanho do efeito (ES) para a existência do fenômeno em teste é algum valor diferente de zero na população. Quanto maior for esse valor, maior será o grau do fenômeno em estudo. Desta forma, o ES serve como um índice do grau de afastamento da hipótese nula.

2.4.2 Testes de Qui-quadrado em Tabelas de Contingência e o Tamanho de Efeito Ômega

Dentre as aplicações mais comuns do teste de Qui-quadrado está o teste de “Contingência”, “Independência” ou “Associação”. Assim, observadas frequências, estas são classificadas simultaneamente por meio de duas variáveis ou princípios de classificação diferentes, ou seja, em uma tabela bidirecional. As frequências conjuntas são testadas contra uma hipótese nula que não especifica nenhuma associação entre as duas bases de classificação [21][22][23]. O teste do Qui-quadrado em frequências é particularmente apropriado com variáveis como escalas nominais ou categorias não ordenadas. Quando usado para comparações de frequência, o teste Qui-quadrado é um teste não-paramétrico, pois compara distribuições inteiras em vez de parâmetros (médias, variâncias) de distribuições. Assim, além da necessidade de evitar frequências hipotéticas muito pequenas [22], o teste é relativamente livre de suposições restritivas.

Segundo Cohen [12], exige-se para um índice ES um número “puro” que aumenta com o grau de discrepância entre a distribuição especificada pela hipótese alternativa e aquela que representa a hipótese nula. Alcança-se “pureza” trabalhando com frequências relativas, ou seja, proporções. Tomando como exemplo esta pesquisa, existem células para a categoria no caso 0 (por ex: para valores de não-suspensão) e para categoria conjunta no caso 1 (por ex, os valores como “Suspensão”). Para cada célula, existem duas proporções populacionais, uma dada pela hipótese nula e outra pela alternativa. O índice ES (ω), também conhecido como Ômega de Cohen, mede a discrepância entre essas proporções

pareadas nas células da seguinte maneira:

$$\omega = \sqrt{\sum_{i=1}^m \left(\frac{(p_{1i} - p_{0i})^2}{p_{0i}} \right)} \quad (2.3)$$

Onde:

- p_{0i} é a proporção na célula i postulada pela hipótese nula, e
- p_{1i} é a proporção na célula i postulada pela hipótese alternativa que reflete o efeito para aquela célula e “ m ” é o número de células.

Para cada célula, a diferença entre os dois P’s é elevada ao quadrado e dividida pela p_{0i} associada à hipótese nula; os valores resultantes são então somados sobre as células e a raiz quadrada obtida. Analisando a Equação 2.3 é possível observar a identidade na sua estrutura com a da fórmula de cálculo padrão para χ^2 com frequências. Em ω , as proporções são usadas no lugar das frequências (para generalização) e os valores da população substituem os valores amostrais.

Se as *proporções da amostra* forem usadas na Equação 2.3 no lugar das p_{1i} ’s, e o “ ω ” resultante for elevado ao quadrado e multiplicado por N (o tamanho total da amostra), o resultado é o valor χ^2 da amostra. ω varia de zero, quando os P’s pareados em todas as células são iguais e, portanto não há efeito e a hipótese nula é verdadeira, até um limite superior que depende da natureza do problema.

A estrutura dos testes χ^2 sobre distribuições (e, portanto, ω) é naturalmente não direcional. Ele simplesmente mede a magnitude total do desvio, independentemente da direção. Por causa do quadrado das diferenças, qualquer desvio — positivo ou negativo — contribui igualmente para o valor do χ^2 . Dessa forma, não é possível determinar a direção do desvio, apenas sua magnitude. Se há apenas 1 grau de liberdade, existem apenas duas direções possíveis para as discrepâncias entre H_0 e H_1 : as diferenças podem ocorrer em uma direção positiva ou negativa. Isso é mais simples porque há um único parâmetro a ser comparado (por exemplo, proporções em uma tabela 2×2). Para tabelas maiores ($r \times k$, onde $r > 2$ e $k > 2$), existem múltiplas direções possíveis para que discrepâncias ocorram. Isso significa que os desvios entre H_0 e H_1 podem ocorrer em várias células simultaneamente, levando a uma combinação complexa de diferenças. Os resultados de todos esses desvios da hipótese nula são agregados em um único valor de χ^2 e comparados a um valor crítico da distribuição de χ^2 , analisando o resultado em uma cauda superior única da distribuição χ^2 e, portanto, são não direcionais (ou bilaterais).

A aplicação dos testes de independência também podem ser vistos como testes de igualdade de duas ou mais distribuições sobre um conjunto de duas ou mais categorias. Consideremos a Tabela 2.1 em que existem duas variáveis ou esquemas de classificação,

cada um composto por categorias mutuamente exclusivas e exaustivas. Adotemos uma das variáveis como “suspensão” = R , composta por $r \geq 2$ categorias, e a outra como “local de residência” = K , composta por $k \geq 2$ categorias. Se todos os membros de uma população forem caracterizados simultaneamente no que diz respeito à sua atribuição de categoria em R ou K , os resultados podem ser expressos numa tabela bidirecional de dimensão $r \times k$, com células r e k . Em cada célula podemos escrever a proporção de observações na população que ela contém. A partir desta tabela, pode-se determinar se R está associado (ou depende de, ou não é independente de) a K na população, ou, equivalentemente, se as r subpopulações na variável R têm distribuições diferentes nas k categorias de K .

Tabela 2.1: Proporções P1 (Hipótese 1) na Distribuição Conjunta de Local de Residência (Urbano/Rural) e Suspensão

Suspensão?	Urbano	Rural	Total
Não	70,0%	27,1%	97,1%
Sim	2,5%	0,4%	2,9%
Total	72,5%	27,5%	100,0%

Na Tabela 2.1 uma subpopulação de famílias foi caracterizada conjuntamente em relação a “suspensão” = R ($r = 2$) e “local de residência” = K ($k = 2$). Observe que a distribuição marginal para suspensão é 0,971 e 0,029, e para local de residência é 0,725 e 0,275. Observe que, embora a proporção marginal entre não-suspensos e suspensão seja de 0,971 e 0,029, a proporção entre os rurais é de 0,271:0,04. Da mesma forma, pode-se notar que, há mais urbanos do que rurais entre os não-suspensos em uma proporção de 2,58:1 e a desigualdade se mantém entre os suspensos, mas com proporção de 6,25:1. Esta desigualdade de proporções dentro de uma coluna (ou linha) da tabela com as proporções marginais da coluna (ou linha) constitui evidência de que R e K não são independentes um do outro, ou que estão associados.

Embora o acima exposto tenha sido descrito em termos de associação entre R e K , também poderia ser entendido como uma investigação sobre se os diferentes grupos R (de suspensão) têm a mesma distribuição proporcional nas várias categorias de K (de local de residência). Na análise de tabelas de contingência, a hipótese nula convencionalmente testada é a de não associação. Assim, para a questão da associação entre suspensão e local de residência, a hipótese nula é representada pelos valores de P_0 da Tabela 2.2. Pequenos desvios destes valores representariam uma associação (ou dependência) fraca, enquanto grandes desvios representariam uma associação forte. O grau de desvio ou índice ES é dado por ω , conforme definido na Equação 2.3, aplicado em tabelas de contingência $r \times k$. Cada uma das células $r \cdot k = m$ tem um valor nulo-hipotético P_0 dado pelo produto das

proporções marginais, como na Tabela 2.2 e um valor P1 alternativo-hipotético refletindo a associação postulada.

Tabela 2.2: Proporções P0 (Hipótese H0) da Distribuição Conjunta de Local de Residência e Suspenso

Família possui beneficiário suspenso?	Urbano	Rural	Suspensão
Não	70,4%	26,7%	97,1%
Sim	2,1%	0,8%	2,9%
Local	72,5%	27,5%	100,0%

Para o problema considerado, aplicando a Equação 2.3 usa-se os valores destas tabelas, $\omega = 0,0533$. Esse valor indexa a quantidade de afastamento de *nenhuma associação* (hipótese nula) ou o grau de associação entre suspenso e local de residência nesta população. De forma equivalente, pode ser entendido como uma indexação da diferença entre suspensos e não-suspensos na sua distribuição em relação ao local de residência. No caso em consideração, o número de graus de liberdade associados ao χ^2 para uma tabela de contingência $r \times k$ é dado por $u = (r - 1)(k - 1)$. Nas tabelas de contingência, o valor máximo de ω depende de r , k e das condições marginais. Se r e k forem atribuídos de forma que r não seja maior que k (é o menor dos dois valores) e nenhuma restrição for colocada nas marginais, o máximo ω é $\sqrt{r - 1}$. Assim, no exemplo, não há valores P1 que resultem em ω maior que $\sqrt{2 - 1} = 1$. Se para ambos as marginais as classes tiverem proporções iguais, ou seja, $\frac{1}{r}$ para um conjunto e $\frac{1}{k}$ para o outro, o máximo $\omega = \frac{\sqrt{(r(r-1))}}{k}$.

Omega e Outra Medidas de Associação.

Embora omega (ω) seja um índice **ES** útil na análise de poder de tabelas de contingência, como medida de associação, seu valor máximo é $\sqrt{r - 1}$; portanto, ω varia com o tamanho da menor das duas dimensões da tabela. Existem vários índices de associação para tabelas de contingência $r \times k$ que estão relacionados com ω . Na Tabela 2.3, os valores equivalentes para esses outros índices são fornecidos para os valores de ω . As equações e a tabela possibilitam a indexação do ES em termos dessas outras medidas.

Tabela 2.3: Tabela de Equivalência de Medidas Simétricas: ω , C, ϕ e ϕ' .

Tamanho do Efeito	ϕ'										
	w	c	$r = 2^*$	3	4	5	6	7	8	9	10
pequeno	0,10	0,100	0,100	0,071	0,058	0,050	0,045	0,041	0,038	0,035	0,033
medio	0,30	0,287	0,300	0,212	0,173	0,150	0,134	0,122	0,113	0,106	0,100
grande	0,50	0,447	0,500	0,354	0,289	0,250	0,224	0,204	0,189	0,177	0,167

Fonte: Baseada em Cohen [12]

*Esta coluna fornece os equivalentes em termos de ϕ para uma tabela 2×2 .

O coeficiente de contingência de Pearson (C) é a medida de associação mais amplamente utilizada em tabelas de contingência [22]. A relação entre C, χ^2 e ω é dada por :

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{\omega^2}{\omega^2 + 1}} \quad (2.4)$$

onde, a primeira expressão fornece o valor C da amostra, a segunda o da população. Para os dados populacionais da Tabela 2.1, por exemplo, onde $\omega^2 = 0,002845$, o valor C é igual a $\frac{\sqrt{0,002845}}{(0,002845+1)} = 0,053259$. Para expressar ω em termos de C:

$$\omega = \sqrt{\frac{C^2}{1 - C^2}} \quad (2.5)$$

Onde $C=0$ quando $\omega = 0$, indicando ausência de associação. O valor máximo de C não é 1, mas aumenta em direção a 1, à medida que ω máximo aumenta. Tem-se que o valor máximo ω é igual a $\sqrt{r - 1}$. Portanto, substituindo na Equação 2.4, o máximo de C é $C = \sqrt{\frac{(r-1)}{r}}$. Por exemplo, uma tabela $2 \times k$ ($k \geq 2$) tem um C máximo de $\sqrt{\frac{(2-1)}{2}} = \sqrt{\frac{1}{2}} = 0,71$, enquanto uma tabela $5 \times k$ ($k \geq 5$) tem um C máximo de $\sqrt{\frac{(5-1)}{5}} = \sqrt{\frac{4}{5}} = 0,89$. Essa dependência variável do limite superior de r é geralmente considerada uma deficiência na medida, tornando-se particularmente estranho quando se deseja comparar valores de C provenientes de tabelas de tamanhos diferentes. À medida que ω aumenta, C aumenta, mas com incrementos progressivamente menores.

Phi (ϕ) é o coeficiente de correlação utilizado em tabelas 2×2 . Em tabelas 2×2 , pode-se conceber cada uma das dimensões dicotômicas R e K como 0 para uma categoria e 1 para a outra (ou quaisquer outros valores distintos) e calcular um coeficiente de correlação produto-momento entre as duas dimensões. Nessas circunstâncias, o coeficiente de correlação é denominado ϕ [12] [24]. Sua relação com ω é de identidade:

$$\phi = \sqrt{\frac{\chi^2}{N}} = \omega \quad (2.6)$$

V de Cramer (ϕ') é uma generalização de ϕ útil para tabelas de contingência de qualquer dimensionalidade [22] [21], sendo:

$$\phi' = \sqrt{\frac{\chi^2}{N(r-1)}} = \frac{\omega}{\sqrt{r-1}}, \quad (2.7)$$

Onde r é não maior que k . A fórmula 2.7 fornece o valor para a amostra e a fórmula 2.8 fornece o valor para a população. ω em termos de ϕ' e r é dado por:

$$\omega = \phi' \sqrt{r-1} \quad (2.8)$$

Naturalmente, ϕ' não pode ser interpretado como uma correlação produto-momento, uma vez que nem R nem K são em geral, métricos ou mesmo ordenados. Mas ele tem um intervalo entre zero e um limite superior uniforme de 1 (um). A última opção é verdadeira porque, conforme explicado anteriormente, o limite superior de ω numa tabela de contingência é $\sqrt{r-1}$. Que ϕ' é uma generalização de ϕ pode ser vista quando nota-se que para uma tabela 2 x 2, $r = 2$; a fórmula 2.7 então fornece $\phi' = \omega / \sqrt{(2-1)} = \omega (= \phi)$. É por isso que os equivalentes ϕ da Tabela 2.3 são dados em ϕ' para $r = 2$. O ϕ' é mais geral, uma vez que se aplica não apenas a tabelas 2 x 2, mas a tabelas 2 x k . Por exemplo, para a associação entre “suspenso” e “local de residência” da Tabela 2.1, $\phi' = \frac{0,0533}{\sqrt{2-1}} = 0,0533$

Valores ω ”Pequeno”, ”Médio” e ”Grande”.

Como ω não é um índice amplamente conhecido, torna-se importante ter alguma orientação sobre a sua magnitude para efeitos de análise de poder ou de estimativa do tamanho da amostra necessária ou ambos. Para Cohen[12], o melhor guia neste assunto é o desenvolvimento de algum senso de magnitude *ad hoc* para um problema específico ou um campo específico. Como ω é uma função de proporções, espera-se ser capaz de expressar o tamanho do efeito que deseja detectar escrevendo um conjunto de proporções alternativas hipotéticas para o caso e, com as proporções nulas hipotéticas, calcular ω . Alguma experimentação nesse sentido deverá fornecer uma “intuição” de ω . Os valores de ω para um ES “pequeno”, “médio” e “grande” são oferecidos para servir como convenções para esses adjetivos qualitativos. Seu uso requer cautela especial, uma vez que, além de sua possível inaptidão em qualquer dado contexto substantivo, pois o que é subjetivamente o “mesmo” grau de associação, pode produzir valores variados de ω , conforme muda o valor de r , k ou ν (graus de liberdade). Por exemplo, na Tabela 2.3, a constante ϕ' diminui à medida que r aumenta. Por isso, utiliza-se as definições convencionais como regra geral de referência para ES, mas não literalmente.

Tamanho de Efeito Pequeno. A Tabela 2.3 fornece os equivalentes de $\omega = 0,10$ para C e ϕ' . O que é definido como um pequeno grau de associação implica um C de 0,100 e, para uma tabela 2 x 2, um ϕ também de 0,100. Para tabelas maiores, o V de Cramer ϕ' diminui, de modo que quando a menor dimensão é 6 (de r categorias), $\phi' = 0,045$. Os valores P_1 que relacionam “suspensão” ao “local de residência” na Tabela 2.1 produziram um $\omega = 0,0533$, na faixa operacional de efeito pequeno.

Tamanho de Efeito Médio. A equivalência na Tabela 2.3 a $\omega = 0,30$ é $C = 0,287$. Para ϕ' em tabelas maiores do que 2 x 2, a constante $\omega = 0,30$ implica valores decrescentes, por exemplo, $\phi' = 0,134$ para $r = 6$.

Tamanho de Efeito Grande. A equivalência na Tabela 2.3 a $\omega = 0,50$ é $C = 0,447$. Para tabelas maiores do que 2 x 2, os valores de ϕ' diminuem a medida que r aumenta, por exemplo, para $r = 6$, $\phi' = 0,224$.

2.5 Mineração de Dados

A mineração de dados tem grande relevância no processo de descoberta de conhecimento em bases de dados. Nesta seção são explorados o conceito e as tarefas de Mineração de Dados e o modelo de referência CRISP-DM para o processo de mineração de dados.

2.5.1 Conceitos de Mineração de Dados

Quando nos deparamos com uma grande base de dados podem existir ali muitas informações interessantes e relevantes que ainda não são conhecidas, ocultas em meio a uma massa de dados. De acordo com Agarwal e Shivam [25], a mineração de dados é um campo que une a Ciência da Computação e a Estatística usada para descobrir padrões no banco de informações com objetivo principal de extrair as informações úteis dos dados e moldá-las em uma estrutura compreensível para uso.

2.5.2 CRISP-DM

O *Cross Industry Standard Process for Data Mining* (CRISP-DM)[26] constitui um modelo de referência para guiar o processo de mineração de dados, independentemente da tecnologia utilizada. O modelo descreve o processo de mineração de dados em fases e define algumas iterações nas quais uma fase pode ser executada mais de uma vez.

O CRISP-DM possui seis fases (Figura 2.1), as quais não são rígidas quanto à sequência de execução, pois pode-se avançar ou retroceder nelas sempre que necessário:

- Entendimento do negócio: é necessário compreender os objetivos de acordo com as regras de negócio, definir o problema a ser tratado com mineração de dados e criar

um plano preliminar. Neste estudo, esta etapa é contemplada nos subtópicos deste capítulo 2, onde estão descritos os aspectos específicos da operacionalização do PBF e sua conexão intersetorial com a educação e saúde.

- Entendimento dos dados: é necessário conhecer aspectos dos dados como a qualidade dos mesmos, os subconjuntos de dados e as variáveis que interessam e formular hipóteses. Esta etapa esta descrita no capítulo 3 sobre a metodologia utilizada, no qual são descritos detalhadamente as bases de dados utilizadas.
- Preparação dos dados: após o entendimento dos dados, procede-se com ações como limpeza dos dados, transformação de variáveis a fim de preparar os dados para aplicação das técnicas de mineração de dados. Esta etapa esta descrita no capítulo 3 sobre a metodologia utilizada, no qual são descritos detalhadamente o tratamento aplicado às bases de dados para se obter uma base de dados única e integrada.
- Modelagem: nessa etapa as técnicas de mineração são aplicadas para a obtenção de bons resultados e descoberta de novo conhecimento. Esta etapa esta descrita no capítulo 3 sobre a metodologia utilizada, no qual são descritos detalhadamente como os modelos de associação foram desenhados.
- Avaliação: Os resultados obtidos na fase de modelagem são analisados. Os passos anteriores também são revisados para se verificar se cumprem os objetivos do negócio. Assim, pode-se tomar a melhor decisão quanto ao uso dos resultados. Esta etapa esta descrita no capítulo 4 contendo os resultados obtidos, no qual se descreve detalhadamente os resultados das análises.
- Implantação: Compreende as ações a serem tomadas diante dos resultados obtidos com a mineração de dados já avaliados. Envolve desde a produção de relatórios até a repetição do processo de mineração de dados se assim for necessário. Os conceitos precisam ser apresentados de forma inteligível para o cliente. Esta etapa esta descrita no capítulo 5 sobre a conclusões e resultados futuros, no qual são descritas as interpretações dos resultados em termos do PBF.

2.6 Técnicas de Discretização de Dados

O uso da heurística *Minimum Description Length Principle* (MDPL) para discretização por minimização da entropia (Fayyad et. al[27]) foi utilizada com o objetivo de possibilitar a aplicação do teste Qui-quadrado baseado em associação de variáveis qualitativas. Foram realizadas comparações entre os resultados de discretizações utilizando métodos

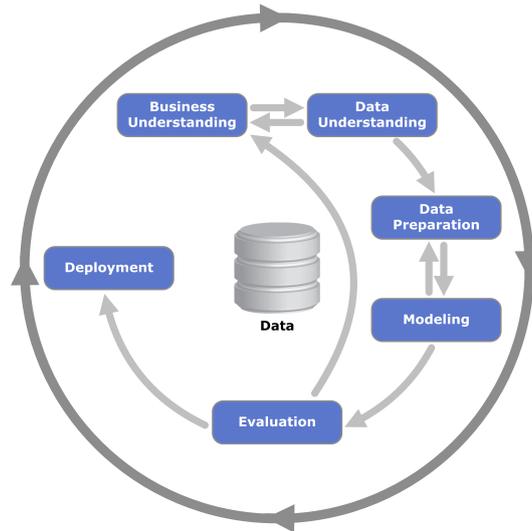


Figura 2.1: Modelo de Referência CRISP-DM
 Fonte: CRISP-DM 1.0 [26]

supervisionados e não-supervisionados [28] [29] e uma comparação específica sobre esses efeitos em algoritmos de associação foi descrita conforme [30]. Um estudo específico sobre algoritmos não-supervisionados baseados em *Equal Width* e *Equal Frequency* foi feito com base nas referências encontradas em [31] e em [32].

Esta pesquisa utiliza a discretização recomendada por McHugh[33] para categorizar dados quantitativos a fim de se obter categorias mutuamente exclusivas e possibilitar o uso de métodos de associação. Garcia et. al [29] afirmam que a discretização é uma técnica de pré-processamento essencial em tarefas de descoberta de conhecimento e mineração de dados, cujo objetivo é transformar um conjunto de atributos numéricos em nominais, associando valores categóricos a intervalos e, assim, transformando dados quantitativos em dados qualitativos para permitir que algoritmos de mineração de dados simbólicos, como indução e previsão, possam ser aplicados sobre dados numéricos e a representação da informação seja simplificada, tornando-a mais concisa e específica.

Diferentes técnicas para discretização de variáveis nominais dão suporte às pesquisas não-paramétricas [29]. O discretizador MDLP é um método supervisionado, ou seja, utiliza informações sobre a distribuição do atributo de classe (variável alvo) para determinar automaticamente o melhor número de intervalos para cada atributo. O processo de discretização aqui descrito foi totalmente automatizado utilizando a ferramenta analítica IBM SPSS 29.0.0 por meio da função *Binning Optimal* que implementa o algoritmo de Entropia MDL¹⁴. Este requer a indicação da variável alvo e nesta pesquisa, propôs-se uma alteração ao uso do MDLP[34], adicionando uma etapa de pré-processamento. Esta alteração estabelece uma restrição baseada em *Equal Frequency* que limita o número máximo

¹⁴IBM SPSS.

de k intervalos de discretizações possíveis, sendo necessária em função do grande número de categorias geradas nas primeiras tentativas. A entropia é medida em *bits* conforme a Equação 2.9 de [27] como: dado um conjunto de dados contendo exemplos rotulados com as classes positiva e negativa, a entropia de S relativa a essa classificação booleana é:

$$\text{Entropia}(S) = - \sum_{i=1}^{N_{cl}} p_i \log_2 p_i \quad (2.9)$$

onde N_{cl} é o número de classes e p_i é a proporção da classe i .

É possível observar que o valor de entropia tem valor mínimo de 0 bits quando a proporção de uma das classes for 1. Nesse caso, todos os exemplos pertencem a essa classe.

A tarefa de discretização MDLP ocorre de forma recursiva, inicialmente todos os valores do atributo X a ser discretizado pertencem a um mesmo intervalo. Todos os possíveis pontos de corte, dados pelo valor médio entre cada par de valores adjacentes do atributo, são testados. Esse teste calcula, para cada ponto de corte pc_j , a quantidade de informação necessária esperada para a classificação dos exemplos, considerando somente o atributo X , discretizado pelo ponto de corte. Essa quantidade de informação é calculada por meio da Fórmula 2.10.

$$\text{Info}_f(pc_j) = \frac{\text{size}(S_{j1})}{\text{size}(X)} \text{Entropia}(S_{j1}) + \frac{\text{size}(S_{j2})}{\text{size}(X)} \text{Entropia}(S_{j2}) \quad (2.10)$$

onde S_{j1} e S_{j2} são dois conjuntos de exemplos, tal que S_{j1} é composto pelos exemplos cujos valores de X são menores ou iguais a pc_j e S_{j2} é composto pelos exemplos cujos valores de X são maiores que pc_j . A função *size* retorna o número de exemplos pertencentes a uma coleção ou atributo.

O melhor ponto de corte pc_j a ser selecionado é aquele que minimiza o valor de *Info*. O menor valor que pode ser obtido é zero e ocorre quando o ponto de corte separa os exemplos de X em dois conjuntos e todos os exemplos de cada um desses conjuntos pertencem a uma única classe.

Após o primeiro ponto de corte ser selecionado, o processo é aplicado recursivamente para cada um dos conjuntos de exemplos S_{j1} e S_{j2} , de maneira a encontrar novos pontos de corte caso necessário. O critério de parada padrão baseia-se no MDLP e compara se o ganho de informação obtido com a criação de um novo ponto de corte é maior do que o ganho de informação com o ponto de corte anterior.

A literatura mostra ainda a aplicação de métodos de discretização não supervisionada como *EqualFrequency* (EFB) e o *EqualWidth* (EWB) além dos métodos supervisionados como o método baseado em entropia. O EFB utiliza **pontos de corte determinados por percentis iguais** para gerar categorias com base no agrupamento com igual número

de casos em cada categoria e o EWB utiliza **pontos de corte determinados por intervalos de largura igual** para gerar categorias de agrupamento de largura igual (por exemplo, renda de R\$ 0,00–89,00 reais) com base na primeira localização de ponto de corte (0,00) e na largura do intervalo (89,00) [28].

Capítulo 3

Metodologia

A pesquisa possui desenho de corte transversal, pois seu propósito é descrever variáveis e analisar suas associações em um período especificado do tempo. A unidade de análise é a família (e seus beneficiários), verificando as interações de suas características e suas relações com variáveis socioeconômicas como violência e PIB per capita no nível municipal. A escolha metodológica de variáveis para representar o conceito de *taxas de suspensão* deu-se sobre a proporção de famílias beneficiárias que sofreram suspensão de benefício por descumprimento das regras de educação. A análise de fatores associados é realizada sobre o histórico das informações das famílias, considerando haver atualização dos dados das famílias a cada dois anos. O tema principal deste estudo está centrado na análise dos fatores associados às diferentes taxas de suspensão entre as famílias, no período de 2014 a 2019. Os dados dos anos posteriores a 2019 não foram utilizados nesta pesquisa em função da publicação de medidas governamentais que afetaram as regras normativas de gestão e operação do PBF para atender as mudanças decorrentes da emergência em saúde pública nacional do COVID-19 como, por exemplo, a interrupção do registro de frequência escolar. As regras de gestão do PBF foram modificadas pela Portaria do Ministério do Desenvolvimento Social nº 335, de 20 de março de 2020, que também suspendeu a aplicação dos efeitos decorrentes do descumprimento das condicionalidades pelo prazo de 120 dias.

A metodologia utilizada nesta pesquisa corresponde a uma adaptação do processo recomendando por Chapman[26], dada a necessidade de inserir um detalhamento das fases relativas às etapas de obtenção e preparação dos dados, uma vez que a utilização de grandes conjuntos de dados governamentais traz complexidade adicional a estas fases. Para tanto, foram adicionadas fases sobre os parâmetros utilizados no desenvolvimento dos algoritmos, haja vista o volume massivo de dados envolvidos.

As seções desse capítulo apresentam as fases da metodologia utilizada no presente projeto de pesquisa, contemplando todas as atividades do projeto, incluindo a obtenção

e integração dos dados, desenvolvimento de algoritmos e a realização das análises de qui-quadrado de independência.

Para que este projeto fosse executado foi necessário gerar uma base de dados integrando dados das seguintes bases de dados do Governo Federal:

- Sistema de Condicionalidades do Programa Bolsa Família. Registra o acompanhamento das condicionalidades de educação e saúde do Programa Bolsa Família. A gestão das condicionalidades do PBF é um trabalho de parcerias entre os três níveis de governo e entre vários setores. Esta base é decorrente do Sistema de Condicionalidades (SICON), que é a ferramenta desenvolvida para apoiar a gestão intersetorial do PBF, integrando diversas informações da gestão de condicionalidades vindas de vários outros sistemas. Dessa forma, é possível monitorar o cumprimento das condicionalidades pelas famílias, bem como sua gestão no âmbito municipal e estadual.
- Sistema de Frequência Escolar do Ministério da Educação. Contêm informações de acompanhamento e monitoramento da frequência escolar de estudantes beneficiários do PBF. Ao registrar os casos e os motivos de crianças e adolescentes que não cumprem os percentuais mínimos de frequência escolar, os gestores públicos podem atuar intersetorialmente, buscando soluções para a melhoria da qualidade do acesso, da permanência e do sucesso escolar dessa população em contexto de vulnerabilidade social.
- Sistema do Cadastro Único do Ministério do Desenvolvimento Social com marcação PBF. Identifica e caracteriza as famílias de baixa renda, permitindo que o Governo conheça melhor a realidade socioeconômica dessa população. Nele são registradas informações como: características do domicílio, identificação de cada pessoa, escolaridade, situação de trabalho e renda, entre outras. Nesta base há marcações que indicam quais pessoas são ou não beneficiárias do Programa Bolsa Família.
- Dados do IBGE sobre Divisão Territorial Brasileira. Disponibiliza dados associados às tipologias Classificação e Caracterização dos Espaços Rurais e Urbanos do Brasil e os Arranjos Populacionais. A Classificação e Caracterização dos Espaços Rurais e Urbanos do Brasil visa apresentar uma proposta de classificação, por município, dos espaços rurais e urbanos, identificados a partir de critérios comuns para todo o Território Nacional. A partir desta base de dados foi possível enriquecer as informações territoriais da zona residencial das famílias, agregando-as à base do Cadastro Único por meio do código IBGE do município.

- Dados do IBGE sobre PIB per capita municipal. Apresenta o PIB per capita a preços correntes (em Reais) entre os anos de 2010 a 2021 e é disponibilizada no endereço eletrônico do IBGE.¹¹
- Dados do IPEA disponível no Atlas da Violência¹², um portal que reúne, organiza e disponibiliza informações sobre violência no Brasil, bem como reúne publicações do IPEA sobre violência e segurança pública, criado em 2016.

Para que se possa alcançar as metas estipuladas neste projeto de pesquisa, propôs-se como abordagem as etapas abaixo:

1. obter os dados de frequência escolar, cadastro único, localização geográfica, PIB per capita municipal, violência municipal e condicionalidades do PBF;
2. realizar limpeza, integração e enriquecimento das bases de dados;
3. aplicar técnica de discretização para transformar dados quantitativos em dados qualitativos para permitir aplicar análise de associação e estimar o tamanho do efeito (ES);
4. utilizar técnicas estatísticas para entender o comportamento dos dados;
5. aplicar algoritmos de Qui-quadrado com Tamanho de Efeito;
6. realizar a análise dos modelos e validar sua precisão e seu desempenho.

Explicaremos cada uma das tarefas propostas nas próximas seções. A metodologia para atividades de análise de dados é baseada no modelo de referência *Cross Industry Standard Process for Data Mining* (CRISP-DM) [35].

3.1 Obtenção e Preparação do Dados

Nesta seção é descrito o processo de obtenção e preparação de dados até a montagem dos *datasets* adequados à realização da análise de associação bivariada sobre o PBF. As tarefas realizadas nesta fase implicaram a construção de uma série de *scripts* para formatação dos dados fornecidos em formato CSV pelos Ministério da Educação (MEC), Instituto Brasileiro de Geografia (IBGE), Instituto de Pesquisa Aplicada (IPEA) e Ministério do Desenvolvimento Social (MDS) e construir um modelo relacional em um banco de dados único. Com o apoio dos dicionários de dados disponibilizados conjuntamente com os

¹¹PIB per capita municipal acessada em 09/2023

¹²Atlas da Violência acessado em 11/2023

arquivos de dados, pode-se então estabelecer a semântica e a sintaxe das tabelas, de modo a realizar o processo de cruzamento de dados nas fases posteriores. Por tratar-se de uma pesquisa populacional e transversal, foram utilizados dados cujos modelos de gestão de dados permitem ter repositórios de dados individualizados e identificados por mês e ano de referência. Após o download dos dados, os mesmos foram carregados em tabelas do banco Postgres para facilitar a execução das fases posteriores.

A Figura 3.1 apresenta o projeto conceitual com a representação gráfica das principais bases de dados necessárias para o estudo, integradas uma modelagem única utilizada nesta pesquisa. Ela representa uma modelagem conceitual de alto nível no qual são apresentadas as entidades, seus relacionamentos e cardinalidade, representados por um Diagrama de Entidade Relacionamento (DER) simplificado. É importante frisar que as entidades criadas representam diferentes políticas públicas ao longo do tempo, em diferentes bancos de dados que foram internalizadas no ambiente centralizado mediante importação de arquivos texto. Por isto, a unicidade do registro de família não é o objetivo final deste modelo.

DER do Projeto Conceitual de Integração das Bases de Dados

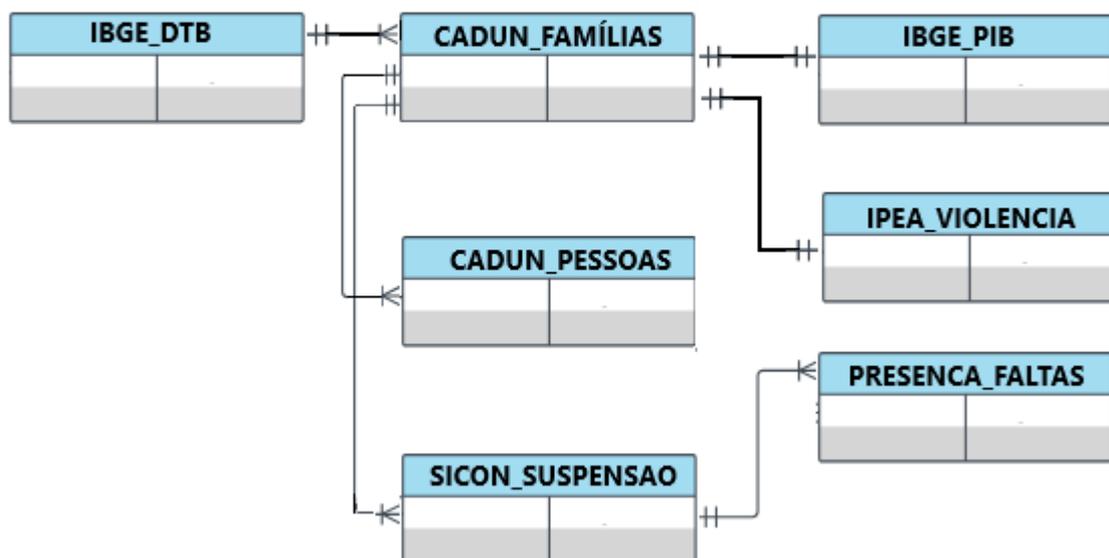


Figura 3.1: Representação do Projeto Conceitual de Integração de Bases de Dados

- **TABELA FAMÍLIA CADASTRO ÚNICO.** Representa a base incremental dos dados históricos da família, provenientes do MDS. Estes dados foram estruturados em duas tabelas. A tabela `cadun_familias` contém 53 informações cadastrais e

incrementais com histórico das famílias no período entre 2014 a 2019, dispostas no Caderno Verde do questionário de atualização cadastral do Cadastro Único. A segunda tabela contém informações do questionário suplementar (conhecido como Caderno Branco) sobre famílias com algum membro em situação de rua.

- **TABELA PESSOA CADASTRO ÚNICO.** Representa a base incremental dos dados históricos das pessoas da família. Estes dados reúnem 48 informações cadastrais e incrementais com histórico sobre cada membro da família, no período entre 2014 e 2019, dispostas no Caderno Verde do questionário de atualização cadastral do Cadastro Único. A identificação do grau de parentesco com o Responsável Familiar (RF) permitiu classificar as famílias segundo critérios de monoparentalidade.
- **TABELAS ACOMPANHAMENTO FAMILIAR SICON.** Reúnem 36 informações sobre o acompanhamento da condicionalidade de educação dos beneficiários do PBF, entre elas a frequência escolar dos beneficiários disponibilizada pelo MEC. É possível identificar quais famílias estão sendo acompanhadas pela Assistência Social Municipal (pré-requisito para eventual cancelamento do PBF por descumprimento da frequência escolar mínima). É importante ressaltar que, no SICON, as informações relativas à repercussão se referem apenas aos descumprimentos que geraram aplicação de algum efeito, visto que alguns descumprimentos não geram repercussão, como por exemplo, se o aluno não foi às aulas porque está doente. Sua estrutura permite identificar os descumprimentos dos beneficiários PBF nos 5 períodos do ano, além de manter o histórico de repercussão das famílias. Neste estudo, por convenção, o termo “faltas” refere-se ao registro de descumprimento bimestral da frequência mínima escolar coletado no processo de acompanhamento de condicionalidades, ou seja, o registro de que o aluno não obteve a frequência mínima do bimestre, conforme calendário operacional do SICON da Tabela 3.1. As tabelas da base de dados do SICON são integradas pelo código familiar, constituindo uma única tabela contendo os dados de acompanhamento da frequência dos beneficiários entre 2014 a 2019 e com alguma suspensão válida;

Tabela 3.1: Calendário de Registro de Frequências no Sistema Presença

Acompanhamento da Família	Registro no Presença	Repercussão no SICON
Fevereiro e março	Abril	Maiο
Abril e maio	Junho	Julho
Junho e julho	Agosto	Setembro
Agosto e setembro	Outubro	Março

- **IBGE.** A tabela IBGE de divisão territorial brasileira (DTB) é composta por 12 informações geográficas sobre os municípios. O campo REGIÃO foi derivado para enriquecer os dados originais. A tabela referente ao PIB per capital Municipal reúne 42 informações municipais para compôr o PIB per capita municipal do ano dividido pela população do mesmo ano.
- **IPEA.** As tabelas do IPEA contêm 4 atributos para caracterizar as taxas de violência por 100 mil habitantes em nível municipal e compreendem o período entre 2014 a 2019. Desta fonte, derivou-se 4 variáveis com a média municipal das taxas de violência para o período pesquisado, sendo uma média para cada tipo de violência. Foram analisadas 4 tipos de violência: suicídio, morte violenta por causa indeterminada, homicídio de mulheres e homicídio de homens. Todas as taxas são padronizadas por 100 mil habitantes.

Alguns dados foram transformados para obter índices derivados. Quando necessário, adaptações em variáveis foram realizadas visando atender aos pressupostos do modelo estatístico adotado.

3.2 Integração e Discretização de Dados

Os desafios de integração de grandes volumes de dados são muitos, dentre eles pode-se destacar: a acomodação do escopo dos dados, a inconsistência dos dados, a otimização de consultas e a adequação de recursos [36]. No caso deste estudo, que aborda bases de dados heterogêneas e com volume massivo, diversas fases fizeram-se necessárias para o cumprimento desta fase. Este tópico trata de descrever as fases relativas ao processo de integração desenvolvido, que considera tanto da integração semântica dos dados envolvidos (junção de escopo), quanto do desenvolvimento de algoritmos específicos para derivar informações adicionais às bases de dados originais.

Em função do grande volume de dados, os procedimentos foram fragmentados em fases com estágios intermediários que consolidaram informações visando possibilitar o acompanhamento dos resultados e atender a limitação dos recursos disponíveis a cada nova junção. Todas as fases de junção foram implementadas no SGBD DBVisualizer em Postgres SQL¹², R Studio e IBM-SPSS 29.0. A Figura 3.2 é uma representação gráfica do *workflow* de integração e discretização, cujos estágios envolvem a seleção, integração, padronização, agregação e parametrização. As medidas de associação tratadas neste estudo referem-se à análise da resposta dicotômica da variável "Suspenso- (Sim/Não) com

¹²DBVIS é uma ferramenta gratuita de banco de dados para coletar e manipular dados variadas as suas fontes.

agrupamento por macrorregião de localização residencial aqui denominadas SERRP e NÃO-SERRP, cuja agregação será tratada adiante.

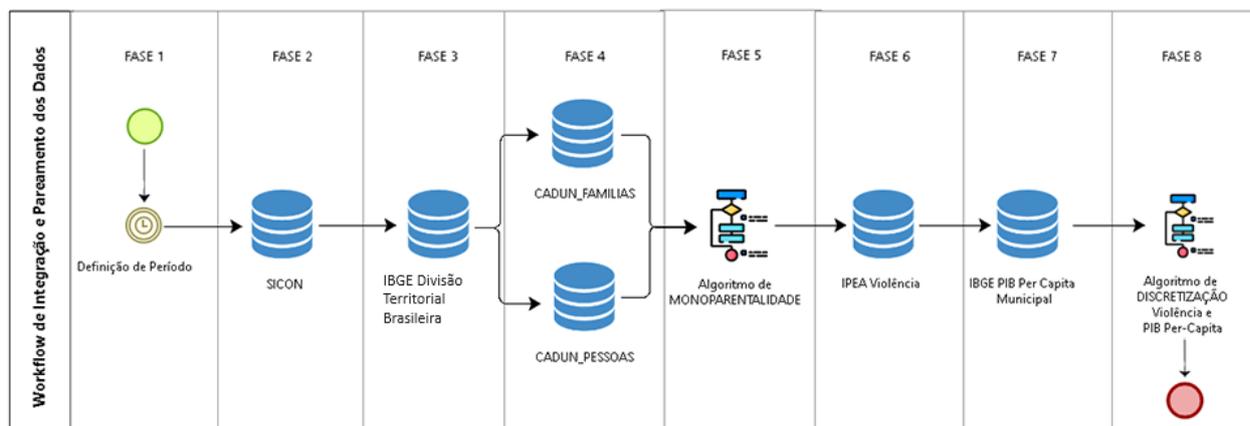


Figura 3.2: *Workflow* de Integração e Pareamento.

Fase 1 - Definição do Período

A primeira fase, representada pelo relógio na Figura 3.2 define o período de pesquisa do estudo, compreendido entre os anos de 2014 a 2019. Embora não represente manipulação dos dados, define variáveis que serão utilizadas nas fases subsequentes. Essa fase define o período da frequência dos beneficiários a serem analisadas e restringiu às famílias que, durante o período, **tiveram alguma suspensão válida em função do descumprimento da frequência escolar mínima** exigida para a permanência no Programa.

Fase 2 - Seleção de Famílias Suspensas

A segunda fase envolve a seleção das famílias segundo os critérios da condicionalidade de educação. Neste caso, as famílias são selecionadas inicialmente no SICON segundo critérios de suspensão: todas as famílias com alguma suspensão válida entre 2016 e 2019 e, em seguida, obtêm-se o histórico de frequências escolares entre 2014 e 2019, uma vez que famílias que estão em fase de suspensão a partir de 2016, estão nesta fase em decorrência das frequências de até 2 anos anteriores. Optou-se também por selecionar apenas famílias que efetivamente possuem acompanhamento familiar pela educação, dado que as famílias que por algum motivo não tenham sido acompanhadas não sofrem efeitos no benefício em decorrência do descumprimento de condicionalidades, e portanto, não se pode afirmar que estas se encontram em situação de descumprimento.

No intuito de comparar o cumprimento dos compromissos de educação assumidos pelas famílias entre municípios e estados, foram criadas as **Taxa de Risco de Suspensão (R)**, **Taxas de Suspensão (S)**, **Dispersão de Absenteísmo dos Suspensos**

(DAS), Indicador de Frequência Máxima Esperada (F) e a Dispersão da Taxa de Suspensão (DTS), cujas fórmulas serão explicadas a seguir. A análise padronizada mostra-se essencial para a compreensão dos dados ao uniformizar as medidas comparativas, independentemente da quantidade de beneficiários de cada estado. A Tabela 3.2 resume os indicadores relacionados ao estudo da frequência escolar:

Tabela 3.2: Variáveis para Estudo da Frequência dos Beneficiários

Variável	Descrição
Indicador de Faltas dos Alunos Suspensos	Quantidade de vezes que a baixa frequência foi registrada pelo recenseador no público suspenso
Total de Beneficiários	Quantidade de beneficiários (suspensos e não-suspensos)
Total de Beneficiários Suspensos	Quantidade de beneficiários suspensos
Indicador de Frequência Máxima Esperada	Cálculo que representa a quantidade de presenças que um aluno pode alcançar (máximo de cinco presenças por aluno por ano). A frequência máxima de 5(cinco) reportada no sistema Presença, equivale a 5(cinco) bimestres nos quais o aluno atendeu a frequência mínima requerida pelo PBF (mínima de 75%).
Taxa de Risco de Suspensão	A proporção entre o número de faltas registradas e o indicador de frequência máxima esperada de um aluno. Mede o risco do aluno não-suspenso, mas que poderá ser suspenso se continuar a faltar.
Taxa de Suspensão	Proporção de alunos suspensos em relação ao total de alunos por estado. Mede a taxa de alunos suspensos.
Dispersão de Absenteísmo dos Suspensos (DAS)	Representa o quão maior é a Taxa de Risco de Suspensão de um estado quando comparado à menor taxa estadual encontrada.
Dispersão da Taxa de Suspensão (DTS)	Representa quão maior é a Taxa de Suspensão de um estado quando comparado à menor taxa estadual encontrada.

A coluna “Indicador de Faltas” dos alunos suspensos refere-se à quantidade de registros de baixa frequência (aqui chamadas de “faltas”, computadas para efeito de repercussão do benefício) de cada aluno suspenso. O registro ocorre conforme Calendário Operacional do Programa, numa frequência bimestral. Logo, a cada bimestre, a gestão avalia se o aluno obteve frequência mínima esperada para o período, onde espera-se uma frequência superior a 85% (6 a 15 anos) ou 75% (16 e 17 anos) nas aulas do bimestre. Caso a frequência mínima não tenha sido atingida pelo aluno, computa-se um registro (ou uma "falta"). Ou seja, um indicador de falta equivale a um bimestre no qual o aluno não

frequentou às aulas conforme o esperado pelo PBF. A Fórmula 3.1 foi utilizada para o cálculo do “Indicador de Frequência Máxima Esperada”:

$$F = TotaldeBeneficiarios * 5(bimestres) * 6(anos) \quad (3.1)$$

Onde, F corresponde ao “Indicador de Frequência Máxima Esperada” dos alunos entre 2014 e 2019. Ela é o resultado da multiplicação entre “Total de Beneficiários”, o número de bimestres e o número de anos. Equivale dizer que o esperado é que cada aluno (suspenso ou não) tenha frequência máxima esperada igual a cinco (5) por ano, ou seja, tenha presença regular em sala de aula durante os cinco períodos de coleta de frequência de cada ano.

A Fórmula 3.2 foi utilizada para o cálculo da “Taxa de Risco de Suspensão”:

$$R = (IndicadordeFaltasdosAlunosSuspensos/F) \quad (3.2)$$

Onde R é a “Taxa de Risco de Suspensão” é obtida a partir da divisão entre o “Indicador de Faltas” dos alunos suspensos e “Indicador de Frequências Máxima Esperada” de todos os beneficiários PBF. Esta taxa permite entender a proporção de frequências perdidas pelos beneficiários suspensos em relação ao total de frequências esperadas para todos os alunos, permitindo analisar a prevalência de faltas entre os alunos suspensos em cada estado. É uma métrica para o absenteísmo escolar instalado entre os beneficiários e representa quão frequentemente os alunos suspensos faltam à escola. A comparação das “Taxa de Risco de Suspensão” entre estados permite identificar aqueles com alto grau de risco de suspensão, cujas faltas dos alunos suspensos são maiores em comparação com o total de presenças esperadas. Isso pode sugerir que, nestes estados, os alunos enfrentam dificuldades mais significativas que os impedem de comparecer à escola regularmente. Já as baixas “Taxas de Risco de Suspensão” em um estado podem indicar que os alunos conseguem comparecer à escola mais vezes, sendo mesmo suspensos.

A “Taxa de Suspensão” mede a proporção de beneficiários PBF de um estado que tiveram pelo menos uma suspensão válida por falta em relação à quantidade total de beneficiários do estado. A variável “Taxa de Suspensão” foi obtida a partir da divisão entre a quantidade de beneficiários suspensos e a quantidade de beneficiários total do estado. A Fórmula 3.3 foi utilizada para o cálculo da Taxa de Suspensão:

$$S = (TotaldeSuspensos/TotaldeBeneficiários) \quad (3.3)$$

Onde S corresponde à “Taxa de Suspensão” no período, “Total de Suspensos” corresponde à quantidade de beneficiários com alguma suspensão válida no período e e “Total de

Beneficiários” equivale à quantidade de beneficiários PBF por estado. A taxa de suspensão nos dá uma ideia da proporção de todos os beneficiários em um determinado estado que tiveram seu benefício suspenso devido à inobservância da condição de frequência escolar. Por exemplo, se a Taxa de Suspensão em um estado for de 0,20 (ou 20%), isso significa que 20% dos beneficiários nesse estado receberam pela menos uma suspensão do PBF por causa de faltas escolares. Essa métrica permite identificar estados onde as suspensões são particularmente prevalentes, o que pode indicar problemas mais amplos com o estado de risco social familiar.

Índices de dispersão foram criados para refletir as diferenças estaduais. Optando-se por medir o grau de variação das taxas de suspensão e de risco de suspensão com base nos menores valores encontrados dentre todos os estados. O estado do Amapá é quem apresentou os menores valores. Foram criadas as variáveis “Dispersão de Absenteísmo” dos Suspensos (DAS) que mede o grau de variação da “Taxa de Risco de Suspensão” comparando-se com o Amapá e a “Dispersão da Taxa de Suspensão (DTS)” que mede o grau de variação da “Taxa de Suspensão” quando comparado com o Amapá. Portanto, o Amapá possui DTS e DAS igual a um. As Fórmulas 3.4 e 3.5 foram utilizadas para o cálculo da “Dispersão de Absenteísmo dos Suspensos (DAS)” e a “Dispersão da Taxa de Suspensão (DTS)”, respectivamente:

$$DAS = (R/MENOR(R)) * 100 \quad (3.4)$$

e

$$DTS = (S/MENOR(S)) * 100 \quad (3.5)$$

Para realizar e registrar o acompanhamento, o município utiliza o Sistema Presença¹³, ferramenta desenvolvida pelo Ministério da Educação. A frequência escolar das crianças e adolescentes é apurada mensalmente e o registro no Sistema Presença é feito bimestralmente. O percentual da frequência escolar dos beneficiários do PBF é calculado com base nos dias letivos do calendário escolar de cada sistema ou estabelecimento de ensino. No Sistema Presença deve-se indicar o motivo nos casos de frequência abaixo da estabelecida pelo Programa. Essas frequências são enviadas ao Ministério do Desenvolvimento Social que identifica as famílias com descumprimento de condicionalidade – ou seja, aquelas que têm um ou mais integrantes que deixaram de cumprir os compromissos assumidos na educação – e aplica os efeitos previstos na legislação num processo chamado de repercussão.

A Figura 3.3 representa a fase de repercussão das faltas, a qual inicia com a junção das tabelas CID_EDUCACAO e CID_FAMÍLIA_FASE_SUSPENSAO. A exclusão de famí-

¹³Sistema de Acompanhamento da Frequência Escolar do Bolsa Família.

lias sem acompanhamento familiar pela assistência social (CD_SITUACAO_ACOMPANHAMENTO) é feita na junção inicial, enquanto a frequência do período é feita em filtro colocado na junção (NU_ANO_MES_REF_FREQUENCIA). O resultado final é uma tabela temporária `FREQ_FAMILIAS_SUSPENSAO` contendo famílias suspensas, suas frequências (ou faltas) escolares, além de informações sobre os motivos destas faltas.

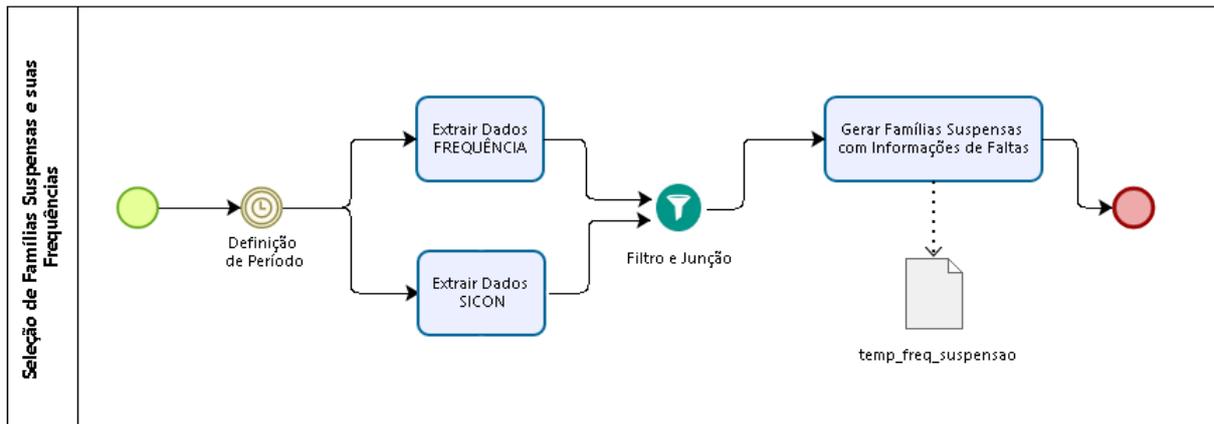


Figura 3.3: Seleção de Famílias Suspensas.

Fase 3 - Obtenção dos Dados de Localização Geográfica e Marcação SERRP

Na terceira fase (Figura 3.4) a base de dados do CADÚNICO é enriquecida com os dados de localização geográfica do IBGE, sendo adicionadas as informações sobre nome do município, estado e região por meio do cruzamento de dados, via chave `COD_MUNICIPIO_IBGE` (padrão IBGE) presente em ambas as bases. A maior parte da população suspensa reside nos estados de São Paulo, Espírito Santo, Rio de Janeiro, Rio Grande do Sul e Paraná, por isso uma *flag* foi criada para diferenciar a população residente nestes estados daquela residente nos demais estados do país. A *flag* assume então o valor SERRP como uma referência às iniciais dos estados com maior população suspensa e o valor NÃO-SERRP como referência aos demais estados do país.

Fase 4 - Obtenção de Famílias e Pessoas da Família e Marcação de Suspensão

A quarta fase foi responsável por desenvolver um algoritmo para classificar as famílias de acordo com a existência de suspensão. A Figura 3.5 detalha esta fase. Inicialmente, todas as famílias e pessoas das famílias que passaram pelo Programa Bolsa Família entre os anos de 2016 a 2019 são selecionadas. Os atributos selecionados para compor o perfil das famílias e pessoas da família basearam-se na perspectiva de se analisar a infraestrutura domiciliar das famílias e o perfil de características pessoais dos membros segundo critérios de localização geográfica, escolaridade, etnia e gênero. Inicialmente, são selecionadas

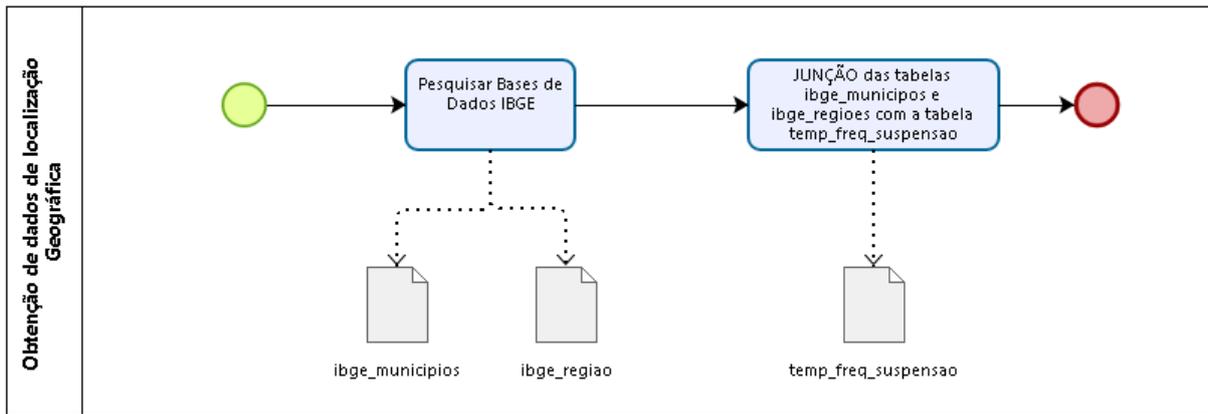


Figura 3.4: Obtenção do Dados do IBGE de Localização Geográfica.

12 variáveis dos membros da família e 22 variáveis das características da infraestrutura de moradia familiar. Em seguida, cria-se uma nova variável SUSPENSO nas tabelas de família e pessoas da família, preenchendo-a por meio de uma junção entre a tabela intermediária do SICON. A junção realizada é uma junção externa à esquerda, que assegura que todas as famílias e pessoas selecionadas no Cadastro Único serão marcadas pela junção, mesmo que não exista registros complementares de suspensão e suas frequências. Além disso, a marcação PBF existente na base de dados do Cadastro Único assegura que apenas a composição cadastral histórica das famílias identificadas como beneficiárias do PBF entram nesta junção. As tabelas resultantes foram CADUN_PESSOAS_PBF e CADUN_FAMILIAS_PBF com marcação de suspensão.

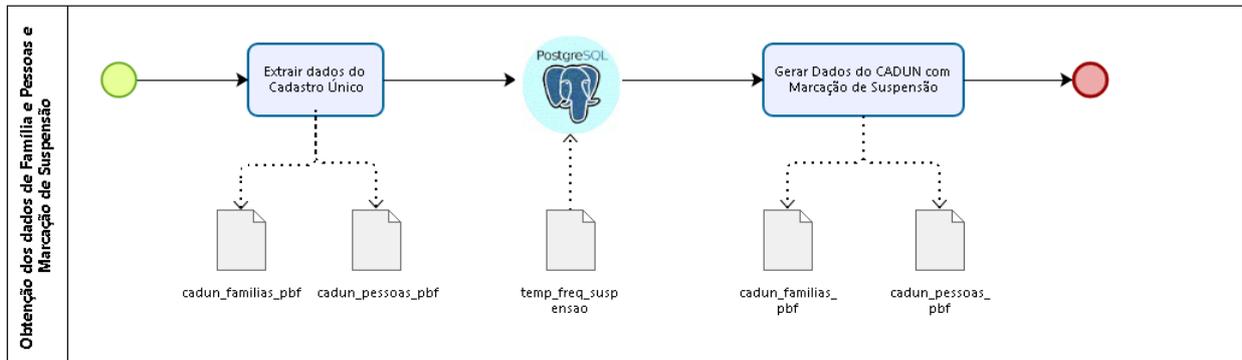


Figura 3.5: Obtenção dos Dados de Família e Pessoas da Família e Marcação de Suspensão.

Fase 5 - Marcação Monoparental

Na quinta fase, após coletados os dados de família e pessoas (membros) das famílias, é desenvolvido um algoritmo para categorizar os membros das famílias quanto a composição familiar de monoparentalidade. Para marcar cada membro da família, utiliza-se

a função OVER com PARTITION BY da classe SQL WINDOW FUNCTION. Essas funções criam uma nova coluna com base nos cálculos executados em um subconjunto ou “window” dos dados e aliadas à cláusula PARTITION BY funcionam em dados particionados (dados agrupados), onde primeiro classifica os dados e depois aplica a função de agregação nesses grupos, colocando o resultado na nova coluna para cada linha desse grupo. Os membros de cada família são agrupados por meio do campo COD_FAMILIAR_FAM e seu parentesco com o responsável familiar foi identificado por meio do campo COD_PARENTESCO_RF_PESSOA. Para famílias beneficiárias PBF cujo responsável familiar não possui cônjuge, marca-se o campo MONOPARENTAL com SIM e as demais com NÃO. A Figura 3.6 mostra a nova classificação armazenada na tabela CADUN_PESSOAS_PBF.

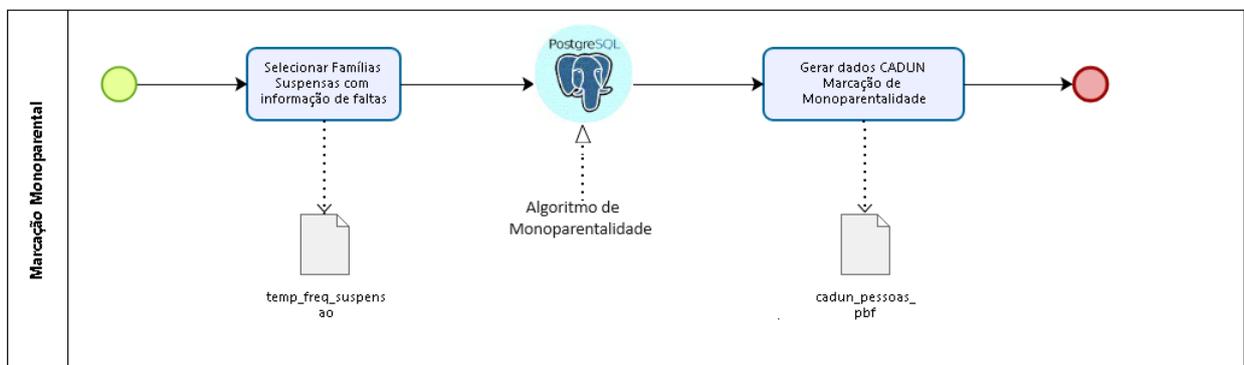


Figura 3.6: Marcação da Composição Familiar Quanto a Monoparentalidade.

Fase 6 - Obtenção dos Dados de Violência

A Figura 3.7 detalha a sexta fase realizada sobre as tabelas de seleção de famílias e membros da família do Cadastro Único. Essas tabelas são enriquecidas com dados do **Mapa de Violência** do IPEA. Inicialmente são utilizadas as tabelas contendo o histórico de taxas municipais por cem mil habitantes de suicídio, morte violenta, homicídios de mulheres e homicídios de homens. A seguir são calculadas as médias históricas para cada município. Os anos analisados correspondem ao mesmo período de faltas, ou seja, são taxas anuais de violência entre os anos de 2014 e 2019. A tabela IPEA_MEDIA consolida o valor médio das taxas anuais de violência para cada município e então realiza-se a junção destas com as tabelas CADUN_PESSOAS_PBF e CADUN_FAMILIAS_PBF.

Fase 7 - Obtenção dos Dados de PIB Per Capita Municipal

A Figura 3.8 detalha a sétima fase para o enriquecimento dos dados do Cadastro Único. A tabela de membros da família e a da família são enriquecidas com dados do IBGE sobre

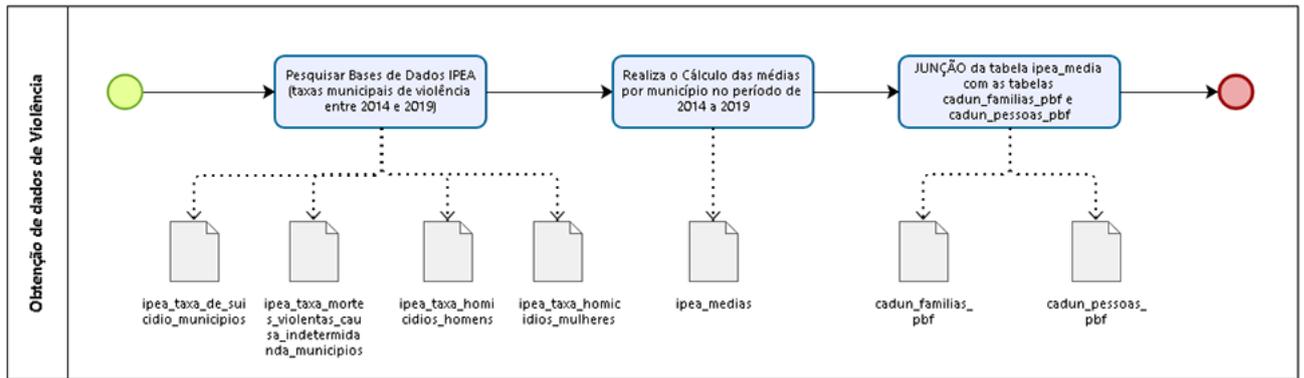


Figura 3.7: Obtenção do Dados de Violência.

o valor médio anual de PIB per capita municipal. O valor utilizado corresponde à média do valor entre os anos de 2014 a 2019. A tabela IBGE_PIB_PERCAPITA_MUNIC_MEDIAS contém o valor do PIB per capita municipal médio deste período e é realizada a junção destes com as tabelas CADUN_PESSOAS_PBF e CADUN_FAMILIAS_PBF.

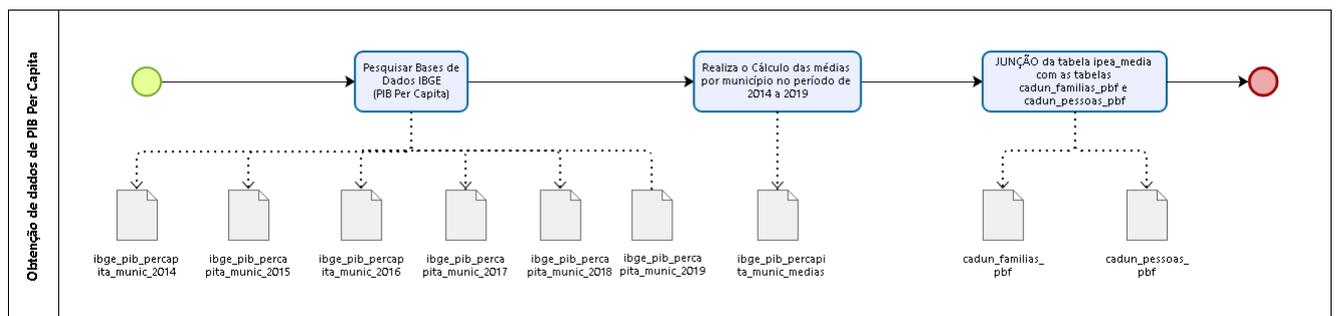


Figura 3.8: Obtenção do PIB Per Capita Municipal Médio

Fase 8 - Limpeza e Discretização dos Dados

O software R Studio foi utilizado para limpeza de variáveis não relevantes ao estudo ou valores de variáveis considerados inválidos, como aqueles majoritariamente sem preenchimento. Na oitava fase também é realizada a discretização recomendada por [33] para categorizar dados quantitativos a fim de se obter categorias mutuamente exclusivas e possibilitar o uso de métodos de associação.

A solução adotada, apresentada na Figura 3.9, mostra a aplicação dos métodos de discretização não supervisionada *EqualFrequency* (EFB) e *EqualWidth* (EWB) e do método supervisionado MDLP baseado em entropia. EWB utiliza pontos de cortes pré-determinados para categorização e pode ser utilizado para representar regras pré-estabelecidas ou para minimizar a perda de informação decorrente da discretização. No atributo referente a quantidade de dormitório no domicílio, QTD_COMODOS_DORMITO-

RIO_FAM utilizou-se a abordagem *EWB*, pois a tentativa de discretizá-lo com base em *EqualFrequency* apresentou grande perda de informação. A discretização utilizou valor inicial um (dormitório) e a largura de intervalo também de um, ou seja, foram formadas categorias com incrementos de um dormitório em cada categoria, até que todos os dados fossem discretizados.

A abordagem baseada em *EqualFrequency* divide a população em pontos de corte com a mesma quantidade de registros em cada corte. Neste estudo, utiliza-se quatro pontos de corte que geram quatro categorias de percentil (quartis), cada um contendo aproximadamente 25% dos casos para cada classe do atributo discretizado. Essa abordagem tem vantagem de otimizar a dispersão de valores, criando uma distribuição uniforme em todos os intervalos criados. Esta foi utilizada nos atributos QTD_COMODOS_DOMIC_FAM (quantidade de cômodos), QTD_PESSOAS_DOMIC_FAM (quantidade de pessoas), QTD_FAMILIAS_DOMIC_FAM (quantidade de famílias).

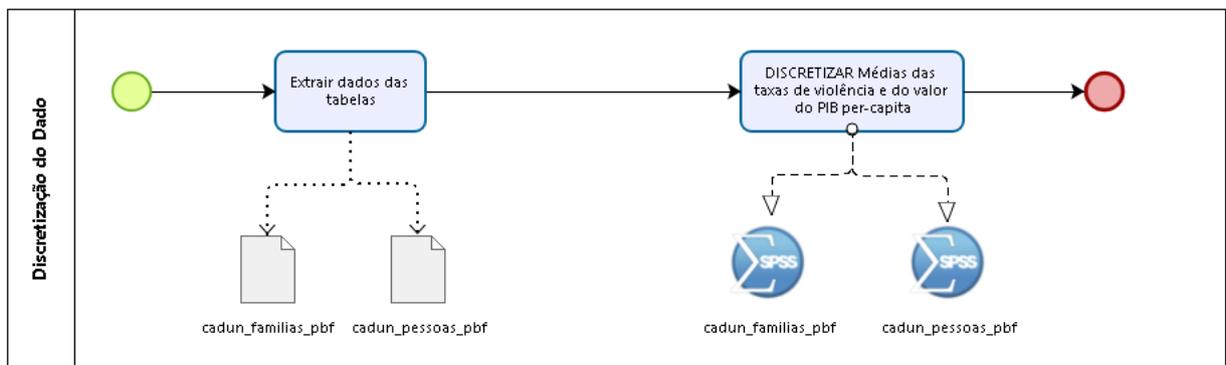


Figura 3.9: Discretização de Dados.

O atributo de violência dos municípios brasileiros possuem milhares de valores, assim como o PIB per capita, exigindo um método mais sofisticado para categorização destas variáveis. A cada município foi atribuído a taxa média da violência entre os anos de 2014 a 2019, uma para cada um dos quatro tipos de violência analisados: TX_MEDIA_SUICIDIO, TX_MEDIA_MORTE_INDETERMINADA, TX_MEDIA_HOMICIDIO_MULHERES e TX_MEDIA_HOMICIDIO_HOMENS. Em seguida, o algoritmo MDLP foi aplicado em cada agrupamento de população nas macrorregiões SERRP e NÃO-SERRP separadamente, a fim de minimizar a perda de informações inerentes ao processo de discretização, especialmente aquelas relacionadas ao uso de técnicas não-supervisionadas [34]. É importante destacar que essa abordagem se baseia no grau de relação da variável a ser discretizada e a variável alvo, não sendo possível caso em que não seja encontrada nenhuma correlação entre estas. Fayyad argumenta que [27] que o algoritmo MDLP é baseado em pureza (entropia) e que faz uso da heurística *Princípio do Comprimento de Descrição Mínima* (MDLP) para minimização de entropia (pureza). Neste estudo, o método utiliza

a variável SUSPENSO como alvo, escolhendo as divisões das taxas de violência que maximizam o ganho de informação sobre SUSPENSO. A tentativa de discretizar a violência e o PIB per capita, utilizando somente o MDLP como critério de parada, produz centenas de categorias em função do grande número de valores contido nestas variáveis. Portanto, para o atributo da violência também é adicionada uma restrição de pré-processamento via *EqualFrequency* com “k” igual a 10, uma abordagem semelhante àquela encontrada em Agree et. al[34]. Para o atributo PIB Per Capita, optou-se pela definição de “k” igual a 3 pois esta configuração manteve os valores de entropia iguais ao “k” igual a 10, ao passo que simplifica a representação da informação.

3.3 Modelo de Análise das Condicionais da Educação

Nas análises por meio de Testes de Hipóteses utiliza-se modelos bivariados analisados por meio de tabelas de contingência *Crosstable* e testes qui-quadrado de independência. Os modelos são aplicados na população de beneficiários suspensos e não-suspensos entre 2016 e 2019, com frequência escolar acompanhada entre 2014 e 2019 (período de 6 anos). São avaliados aproximadamente 155 milhões de registros históricos cadastrais dos membros da família, que equivalem a aproximadamente 82 milhões de registros históricos de famílias. O estudo contempla ainda a análise de 9.172.198 milhões de registros de faltas (frequências escolares para efeito de repercussão do benefício). O software utilizado para obtenção dos resultados foi o IBM-SPSS 29. Foram aplicados o teste estatístico qui-quadrado com nível de significância de 5%.

3.3.1 Delimitação das Macrorregiões

A análise das faltas dos beneficiários suspensos por descumprimento da condicionalidade da educação constatou padrões populacionais que estabelecem dois agrupamentos distintos de beneficiários, dada a similaridade das características dos alunos suspensos em cada grupo. Cada agrupamento é composto por um conjunto de estados brasileiros. Desta forma, o comportamento do absenteísmo escolar foi analisado sob a ótica de duas subpopulações. A primeira subpopulação é formada pelos cinco estados com as maiores Taxas de Suspensão, **S**ão Paulo, **E**spírito Santo, **R**io de Janeiro, **R**io Grande do Sul e **P**araná, identificada pelo acrônimo **SERRP**. A segunda subpopulação é formada pelos demais estados brasileiros, identificada pelo acrônimo **NÃO-SERRP**. As métricas utilizadas para avaliar o período 2014-2019 estão representadas na Tabela 3.2, já apresentada, e foram analisadas por macrorregião, estado e motivos de faltas.

3.3.2 Definição de Variáveis

Essa pesquisa admite que a Taxa de Suspensão PBF por descumprimento das condicionalidades da educação é afetada por fatores socioambientais e econômicos que podem influenciar o aumento do risco da vulnerabilidade social das famílias pobres. Para isso, foram selecionadas as variáveis disponíveis que trazem informações sobre as hipóteses estabelecidas, conforme as dimensões social, econômica e familiar. Exclui-se da modelagem as variáveis não representativas de informação como id, códigos de integridade referencial e variáveis com alto percentual de nulos. A dimensão econômica é representada pela variável de PIB per capita municipal e pretende verificar se o alto desenvolvimento econômico municipal contribui para o aumento dos níveis de suspensão decorrente de exclusão escolar. A dimensão social, representada pela variável de violência, classificada nos tipos Morte Violenta por Causa Indeterminada, Homicídio de Mulheres, Homicídio de Homens, além da visão sobre Suicídio, busca avaliar se condições de violência municipal afetam a dinâmica familiar ao ponto de ter grande associação com a exclusão escolar. A dimensão familiar é representada pelas variáveis que compreendem as características da infraestrutura domiciliar da família e do perfil dos membros da família e analisa se a exclusão escolar dos alunos suspensos do PBF é influenciada por fatores que representam maior precariedade do domicílio ou por fatores que representam condições de maior vulnerabilidade nos membros da família.

3.3.3 Método Estatístico

Para avaliar as hipóteses desta pesquisa, faz-se necessário um método que represente o contexto associado à variação da Taxa de Suspensão. O método escolhido foi o Teste de Independência. A ideia central do teste sobre um parâmetro ou comparação de parâmetros de populações é a de supor verdadeira a hipótese em questão e verificar se a amostra observada é verossímil ou mais provável, ou seja, se está refletindo o que há de mais próximo do real estado das condições testadas dentro do ambiente em que se conduz a pesquisa[37]. No caso de teste de independência, procura-se medir o grau de associação entre as variáveis e colocar à prova as seguintes hipóteses:

1. H_0 : as variáveis são independentes (não existe associação entre as variáveis).
2. H_1 : as variáveis não são independentes (elas apresentam grau de associação entre si).

A associação entre as variáveis foi medida pelo qui-quadrado (χ^2) de independência com nível de significância de 5%. Para aplicação desse teste, os dados são dispostos em

Figura 3.10: Quadro Resumido de Variáveis.

DIMENSÃO	FATOR	VARIÁVEL	CLASSES
FAMILIAR	PERFIL DOS MEMBROS DA FAMÍLIA	Morador de Rua	(Sim; Não)
		Raça-Cor	(Branca; Preta; Amarela; Parda; Indígena;)
		Sexo	(Masculino; Feminino)
		Pessoa Frequenta Escola?	(Sim, rede pública; Sim, rede particular; Não, já frequentou ; Nunca frequentou)
		Deficiente	(Sim; Não)
		Sabe ler e escrever?	(Sim; Não)
		Local de Nascimento	(Deste município - Onde está sendo feito o cadastramento ; Em outro município)
	Família Monoparental	(Sim; Não)	
	INFRAESTRUTURA DOMICILIAR	Local de Residência	(rural ou urbano)
		Espécie Domicílio	(Coletivo, Particular permanente, Particular improvisado)
		Água Canalizada	(Sim; Não)
		Escoamento Sanitário	(Rede coletora de esgoto ou pluvial; Fossa séptica; Fossa rudimentar; Vala a céu aberto; Direto para um rio, lago ou mar; Outra forma)
		Iluminação	(Elétrica com medidor próprio; Elétrica com medidor comunitário; Elétrica sem medidor; Óleo ou querosene ou gás; Vela; Outra forma)
		Material Paredes Domicílio	(Alvenaria/tijolo com revestimento; Alvenaria/tijolo sem revestimento; Madeira aparelhada; Taipa revestida; Taipa não revestida; Madeira aproveitada; Palha; Outro Material)
		Abastecimento de Água	(Rede geral de distribuição; Poço ou nascente; Cisterna; Outra forma)
		Quantidade de Pessoas no Domicílio	(5 faixas categorizadas por Equal Frequency)
		Quantidade de Famílias no Domicílio	(6 faixas categorizadas por Equal Frequency)
		Quantidade de Cômodos no Domicílio	(4 faixas categorizadas por Equal Width)
		Quantidade de Dormitórios no Domicílio	(4 faixas categorizadas por Equal Frequency)
SOCIAL		VIOLÊNCIA MUNICIPAL (Discretização por Entropia MDL)	Taxa de Suicídio
	Taxa de Morte Violenta por Causa Indeterminada		Média Municipal do período - 10 faixas SERRP e 10 faixas NÃO-SERRP. Faixas crescentes.
	Taxa de Homicídio de Mulheres		Média Municipal do período - 9 faixas NÃO-SERRP e 10 faixas SERRP. Faixas crescentes.
	Taxa de Homicídio de Homens		Média Municipal do período - 8 faixas NÃO-SERRP e 8 faixas SERRP. Faixas crescentes.
ECONÔMICA	PIB PERCAPITA MUNICIPAL (Discretização por Entropia MDL)	Faixa 1, 2 e 3	Média Municipal do período - 3 faixas SERRP e 3 faixas NÃO-SERRP. Faixas crescentes.

uma tabela de contingência constituída de h linhas e k colunas, ou seja, é uma tabela que mostra como duas características ou atributos dependem uma da outra.

A estatística do teste qui-quadrado possui distribuição teórica de probabilidade de qui-quadrado com o número de graus de liberdade V. Em uma tabela de contingência h x k, com h > 1 e k > 1, $V = (h-1) (k-1)$ se as frequências esperadas podem ser calculadas sem que se tenha que estimar os parâmetros populacionais por meio das estatísticas amostrais do estimador em questão ou sob estudo [13]. Para medir o grau de associação entre as variáveis é necessário calcular o V de Cramer (ϕ'), sendo feito após verificada a significância da associação com o teste qui-quadrado. Isso ocorre porque a significância do teste de qui-quadrado depende das diferenças entre as proporções e também do tamanho da amostra. O grau de associação, no entanto, independe do tamanho da amostra, pois ele é função das proporções observadas.

Capítulo 4

Resultados Obtidos

Neste capítulo são apresentados os resultados empíricos dos estudos realizados sobre o conjunto de dados conforme tratamento detalhado na Metodologia. O capítulo se inicia com a estatística descritiva das variáveis da pesquisa e em seguida apresenta uma análise abreviada das associações, buscando identificar a existência e o grau de força das associações encontradas. Isso permite esclarecer os fatores que influenciam a exclusão escolar dos beneficiários suspensos do Programa Bolsa Família.

Por último, são construídos modelos de associações considerando a interação entre as diversas variáveis estudadas, que permitirá, no capítulo final, avaliar as hipóteses de pesquisa. A análise será dividida em duas Seções: "Estudo do Absenteísmo Escolar", focada na análise exploratória e "Associações e o Tamanho do Efeito", que trata dos modelos que avalia a existência e o grau das associações.

4.1 Estudo do Absenteísmo Escolar

Nesta seção são apresentados os indicadores criados para avaliar a relação entre as faltas e o comportamento escolar dos beneficiários suspensos e não-suspensos. As análises basearam-se no estudo do Indicador de Faltas, Taxa de Risco de Suspensão, Taxa de Suspensão, Dispersão do Absenteísmo dos Suspensos (DAS) e Dispersão da Taxa de Suspensão (DTS), considerando as subpopulações por critérios de distribuição regional e local de residência. Em seguida é feito um detalhamento sobre os principais motivos de faltas registrados na frequência escolar dos beneficiários.

4.1.1 Análise das Faltas

A Tabela 4.1 apresenta a distribuição dos tipos de motivos das ausências no período de 2014 a 2019 e indica que a maior parte das ausências (47,2%) dos alunos suspensos não

possuem motivos qualificados, impossibilitando a análise da motivação das ausências a partir dos registros de frequência escolar. As motivações mais frequentes para as faltas escolares foram caracterizadas como desinteresse ou abandono escolar dos próprios aluno ou pela negligência dos familiares. O tratamento de doenças também aparece entre os cinco motivos mais frequentes, junto com a falta de matrícula do aluno. Os 14 motivos menos frequentes foram agregados no código 100 para facilitar as análises por tipo de motivo.

Tabela 4.1: Frequência do Indicador de Faltas por Motivo

Cod. Motivo	Descrição Motivo	Quantidade	%
58	Motivo da baixa frequência não foi informado	4.329.735	47,20%
64	Desinteresse/Desmotivação pelos estudos	1.044.660	11,39%
53	Negligência dos pais ou responsáveis	1.012.269	11,04%
65	Abandono Escolar/ Desistência	812.004	8,85%
72	Beneficiário sem vínculo/matricúla escolar	756.505	8,25%
1	Tratamento de doença e de atenção à saúde do aluno	748.026	8,16%
68	Questões sociais, educacionais e/ou familiares	181.945	1,98%
4	Fatos que impedem o deslocamento/acesso do aluno à escola	132.820	1,45%
100	Outros Motivos	154.234	1,68%
Total		9.172.198	100%

A análise das faltas registradas no SICON indicou 9.172.198 milhões de faltas (registradas para fins de repercussão de benefício) de beneficiários que sofreram alguma suspensão no período entre 2016 a 2019. As subpopulações foram agrupadas segundo critérios de classificação geográfica definidos pelo IBGE, que divide o território brasileiro em cinco classes de zonas em cada região¹⁵, também chamadas de zonas geográficas:

- Predominantemente Rural Remota ou Rural Remota
- Predominantemente Rural Próxima a uma Cidade ou Rural Adjacente
- Intermediária Remota
- Intermediária Próxima a uma Cidade ou Intermediária Adjacente
- Predominantemente Urbana ou Urbana

Esta classificação leva em conta critérios como densidade demográfica, tamanho da população e localização em relação aos principais centros urbanos (ou acessibilidade, definida pelo tempo de deslocamento entre os centros urbanos e as áreas não urbanas). A

¹⁵Classificação e Caracterização dos Espaços Rurais e Urbanos do Brasil. Uma Primeira Aproximação. Publicado pelo IBGE-2017.

Tabela 4.2 mostra como as faltas ocorrem segundo as características fisiogeográficas definidas pelo IBGE, mostrando importantes diferenças regionais como nas regiões Norte e Nordeste que possuem as menores populações urbanas e as maiores populações de beneficiários suspensos em áreas rurais.

Tabela 4.2: Faltas de Suspensos: Proporção por Região e Zonas Geográficas

	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Intermediário adjacente	6,47%	9,82%	4,58%	3,60%	7,95%
Intermediário remoto	3,96%	0,06%	3,57%	0,02%	0,02%
Rural adjacente	11,64%	28,98%	15,41%	6,56%	15,79%
Rural remoto	5,13%	0,87%	10,67%	0,13%	0,00%
Urbano	72,80%	60,27%	65,77%	89,69%	76,24%

A Figura 4.1 apresenta, lado a lado, as proporções de famílias suspensas versus proporções de famílias beneficiárias, em cada região do país. Quase o triplo dos alunos beneficiários do PBF residem nos estados das regiões Norte, Nordeste e Centro-Oeste quando comparado aos estados das regiões Sul e Sudeste. Entretanto, estes estados respondem por apenas cerca de 1/4 (28,2%) dos alunos suspensos do país quando comparados aos dos estados das regiões Sul e Sudeste (71,8%).

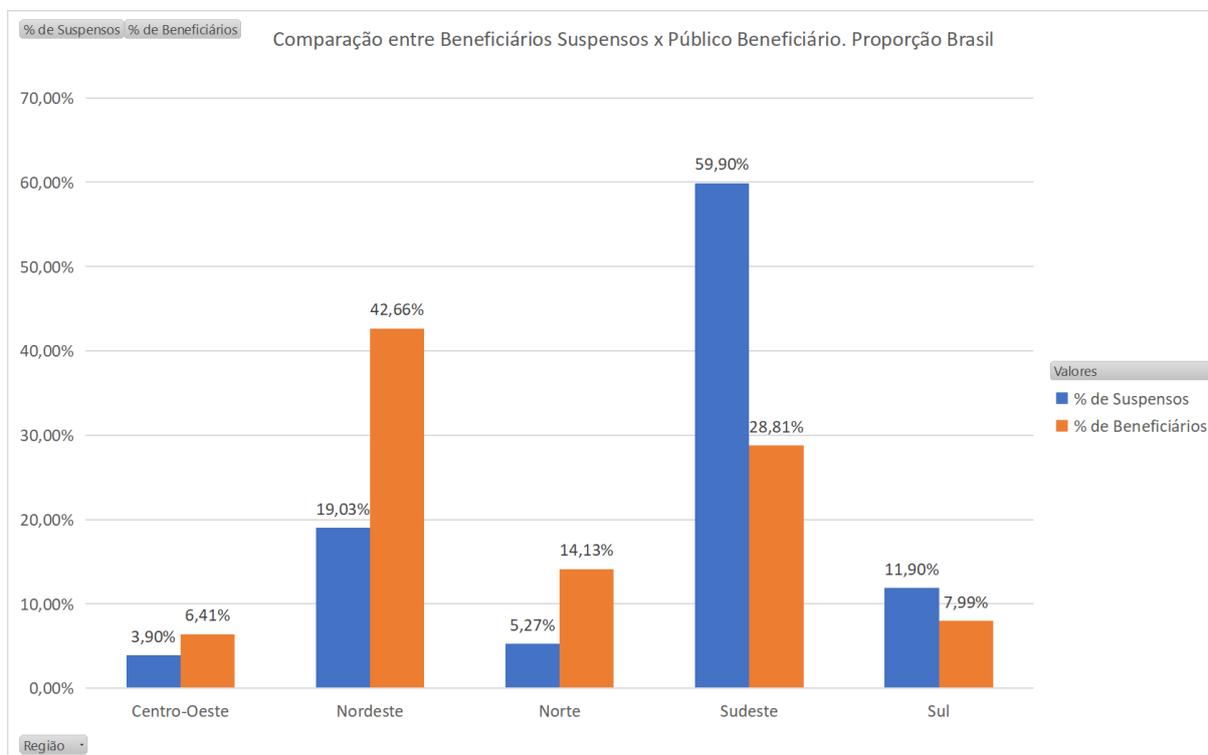


Figura 4.1: Brasil: Famílias Beneficiárias *versus* Famílias Suspensas.

Se a Taxa de Suspensão fosse uma simples consequência do volume de benefícios, não deveria ser observada uma aparente relação inversa entre os estados com a maior população beneficiária e menor população suspensa. Os dados sugerem que os percentuais de suspensão por descumprimento das condicionalidades de educação são menores nas regiões com maior número de beneficiários, as mesmas economicamente mais pobres, onde políticas como o PBF são mais sobressalentes. A Tabela I.1 do Anexo I oferece o detalhamento das distribuições dos beneficiários suspensos em cada estado, revelando que as regiões Sul e Sudeste lideram em termos de suspensões, sendo as maiores taxas nos estados de Santa Catarina, Espírito Santo, Rio de Janeiro, Rio Grande do Sul e Paraná. As diferenças entre as regiões conduziram a definição de duas macrorregiões distintas: SERRP (compreendendo os cinco estados com as maiores taxas de suspensão) e NÃO-SERRP (englobando os demais estados). Esta divisão viabiliza uma avaliação mais aprofundada e padronizada dos fatores associados às características específicas de cada macrorregião, respeitando suas singularidades socioeconômicas, geográficas e culturais de cada agrupamento.

As faltas foram analisadas segundo a frequência de motivos identificados durante a coleta municipal. É importante ressaltar a qualidade da identificação dos motivos depende de cada recenseador municipal, podendo ser uma importante ferramenta de identificação de vulnerabilidade, se apropriadamente marcados. A Tabela 4.3 detalha os motivos identificados nas faltas dos alunos suspensos no período 2014-2019. Os percentuais referentes ao tipo de motivo “Motivo da baixa frequência não foi informado” foi retirado para simplificar a análise.

Tabela 4.3: Distribuição de Motivos de Faltas por Região (%)

Motivos de Faltas	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Tratamento de doença e de atenção à saúde do aluno	15,39%	13,32%	11,46%	17,80%	10,97%
Fatos que impedem o deslocamento/ acesso do aluno à escola	3,23%	1,96%	3,01%	4,78%	2,61%
Negligência dos pais ou responsáveis	23,29%	16,68%	25,25%	21,23%	22,79%
Desinteresse/Desmotivação pelos estudos	19,42%	19,70%	13,43%	22,24%	23,65%
Abandono Escolar/Desistência	13,31%	21,47%	21,54%	12,99%	17,12%
Questões sociais, educacionais e/ou familiares	2,68%	4,04%	3,29%	2,15%	3,18%
Beneficiário sem vínculo/matricula escolar	19,78%	19,91%	19,73%	15,79%	15,38%
Outro Motivos	2,91%	2,92%	2,30%	3,01%	4,30%

A Tabela 4.3 revela disparidades significativas nos motivos que levam à ausência escolar em diferentes partes do Brasil. Dos sete motivos identificáveis, observa-se, por exemplo, que o Sudeste possui os maiores índices motivados por tratamento de doença e de atenção à saúde do aluno, com 17,80%. O Sudeste apresenta também o maior percentual de faltas devido a dificuldades de deslocamento ou acesso à escola, com 4,78%. O Sul é a região

com o maior índice de faltas por desinteresse ou desmotivação, atingindo 23,65%. O Nordeste lidera as faltas por problemas sociais, educacionais e familiares com 4,04%. O Centro-Oeste tem o maior percentual de ausências devido à falta de vínculo ou matrícula escolar, com 19,78%. O Norte lidera um motivo com 25,25%, indicando que a negligência parental pode ser um fator significativo na ausência dos alunos.

4.1.2 Macroregiões SERRP e NÃO-SERRP

Para avaliar o absentismo escolar no país, optou-se por comparar a variação das medidas entre os estados a partir de indicadores padronizados, quais sejam: “Taxa de Suspensão”, “Indicador de Frequência Máxima Esperada”, “Taxa de Risco de Suspensão”, “Dispersão de Absenteísmo dos Suspensos (DAS)” e a “Dispersão da Taxa de Suspensão (DTS)”. Estes indicadores foram comparados entre si, considerando o menor valor encontrado para a “Taxa de Suspensão” e para a “Taxa de Risco de Suspensão” que foram os do estado do Amapá, os quais foram padronizados com o valor um. Essa abordagem permite uma análise comparativa que abstrai os valores quantitativos que são muito diferentes entre os estados. Um estado com alto grau de risco de suspensão entre alunos, em comparação àquele com a menor taxa, pode fornecer um guia de quais estados podem estar em estado de alerta em relação a um absentismo ainda não instalado, mas em risco de concretizar-se. Estas comparações avaliam o quão frequentemente as faltas dos alunos derivaram suspensões nos estados, delimitando o caminho para as demais avaliações desenvolvidas no estudo. A Tabela 4.4 apresenta a distribuição das métricas de absentismo nos estados brasileiros. Pode-se observar que os estados de São Paulo, Espírito Santo, Rio de Janeiro, Rio Grande do Sul e Paraná possuem os maiores índices de DAS e DTS. O Rio Grande do Sul apresenta índice DAS de 1,74 ou seja, possui um índice 74% maior de absentismo do que o do Amapá de beneficiários suspensos. São Paulo tem índice DTS de 26,48 ou seja, uma taxa 2548% maior para os beneficiários suspensos do PBF do que a do Amapá.

É importante notar que a diferença no número de beneficiários do programa entre os municípios é levada em conta quando se calcula a “Dispersão de Absenteísmo dos Suspensos” e a “Taxa de Suspensão”. Portanto, essas razões podem ser comparadas de maneira justa entre municípios ou estados, independentemente do tamanho da população beneficiária em cada um. Ao utilizar um estado com a menor taxa de suspensão e o menor grau de risco de suspensão por aluno suspenso como referência, pode-se comparar os outros estados a essa base, fornecendo uma visão mais clara das disparidades nacionais existentes.

Analisando os dados da Tabela 4.4, vemos que São Paulo apresenta a maior Taxa de Suspensão Estadual (13,07%) e a maior Taxa de Risco de Suspensão (3,40%). Isso significa que São Paulo não só tem a maior proporção de beneficiários suspensos em relação ao total

de beneficiários do estado, mas também a maior proporção de frequências perdidas em relação ao total de presenças esperado. Em contraste, o Amapá tem as menores taxas de suspensão e de risco de suspensão, com valores de referência para as métricas de dispersão (DAS e DTS iguais a 1). Esses dados mostram que o estado do Amapá tem a menor taxa de problemas de frequência e suspensão dentre os beneficiários analisados.

Tabela 4.4: Análise do Absenteísmo Escolar no Brasil

Estado	Indicador de Frequência Máxima Esperada	Taxa de Suspensão (%)	Taxa de Risco de Suspensão (%)	Absenteísmo dos Suspensos (DAS)	Dispersão da Taxa de Suspensão (DTS)	Macrorregião
São Paulo	110.291.520	13.07	3,40	1.56	26.48	SERRP
Paraná	29.649.480	7.06	1,73	1.47	14.30	SERRP
Rio de Janeiro	53.774.490	7.01	1,76	1.51	14.21	SERRP
Rio Grande do Sul	28.494.390	6.91	2,00	1.74	14.00	SERRP
Espírito Santo	13.804.740	6.58	1,81	1.65	13.33	SERRP
Minas Gerais	73.705.920	4.13	1,00	1.45	8.37	NÃO-SERRP
Mato Grosso do Sul	10.395.210	3.78	0,85	1.35	7.66	NÃO-SERRP
Amazonas	28.925.220	3.07	0,58	1.14	6.23	NÃO-SERRP
Santa Catarina	11.641.170	2.88	0,67	1.40	5.83	NÃO-SERRP
Sergipe	14.310.030	2.69	0,69	1.55	5.44	NÃO-SERRP
Mato Grosso	13.553.190	2.68	0,61	1.36	5.43	NÃO-SERRP
Distrito Federal	6.951.690	2.42	0,44	1.08	4.91	NÃO-SERRP
Pernambuco	59.019.900	2.40	0,54	1.36	4.87	NÃO-SERRP
Ceará	56.553.270	2.35	0,58	1.49	4.76	NÃO-SERRP
Bahia	92.388.660	2.31	0,53	1.38	4.69	NÃO-SERRP
Goiás	25.090.140	2.06	0,46	1.35	4.17	NÃO-SERRP
Rondônia	7.761.870	2.01	0,40	1.18	4.07	NÃO-SERRP
Rio Grande do Norte	19.140.030	2.00	0,48	1.46	4.05	NÃO-SERRP
Paraíba	25.284.030	1.93	0,44	1.39	3.91	NÃO-SERRP
Alagoas	23.870.910	1.43	0,33	1.37	2.89	NÃO-SERRP
Tocantins	9.540.360	1.43	0,29	1.23	2.89	NÃO-SERRP
Roraima	3.888.360	1.24	0,23	1.09	2.52	NÃO-SERRP
Pará	61.156.890	1.02	0,20	1.20	2.07	NÃO-SERRP
Acre	6.827.550	0.97	0,18	1.13	1.96	NÃO-SERRP
Maranhão	58.631.490	0.72	0,15	1.25	1.45	NÃO-SERRP
Piauí	23.272.050	0.57	0,11	1.20	1.16	NÃO-SERRP
Amapá	5.286.990	0.49	0,08	1.00	1.00	NÃO-SERRP

Através destes indicadores é possível observar que há uma variação significativa entre os estados em termos de suspensão e risco de suspensão dos alunos beneficiários, o que pode refletir diferenças na aplicação do programa Bolsa Família ou nas condições socioeconômicas ou no sistema educacional dos estados. Em consequência disso, os estados foram divididos em SERRP e NÃO-SERRP como guia para as análises subsequentes. As regiões SERRP (São Paulo, Espírito Santo, Rio de Janeiro, Rio Grande do Sul e Paraná) apresentam taxas de suspensão mais elevadas quando comparadas com a maioria dos es-

tados da região NÃO-SERRP, sugerindo a existência de fatores associados diversos no que se refere às suspensões nestes locais.

4.1.3 Análise da Monoparentalidade

Estudos do IPEA apontam que a vulnerabilidade de uma família pode ser medida como o volume adicional de recursos necessários para satisfação de suas condições básicas de sobrevivência e que um dos componentes que representa a vulnerabilidade familiar é a insuficiência financeira decorrente de arranjos monoparentais com filhos pequenos (de 5 anos ou menos), que precisam optar entre trabalhar ou cuidar dos filhos, ficando sem fonte de renda para o atendimento das necessidades mais básicas de sua família[38]. Os registros foram classificados por meio de um algoritmo que, após agrupar todos os membros de uma família, verifica se há cônjuge entre eles. A Figura 4.2 mostra a distribuição de famílias monoparentais suspensas do PBF em cada estado. Verifica-se a prevalência da monoparentalidade entre as famílias suspensas, sendo aproximadamente 62% das famílias suspensas do país. Os cinco maiores índices estaduais são: Rio de Janeiro, Distrito Federal, Pernambuco, São Paulo e Rio Grande do Sul. Três deles compõem a macrorregião SERRP. Percebe-se também que, dos 26 estados mais Distrito Federal, 21 deles possuem índices de monoparentalidade superiores a 50% em suas populações suspensas.

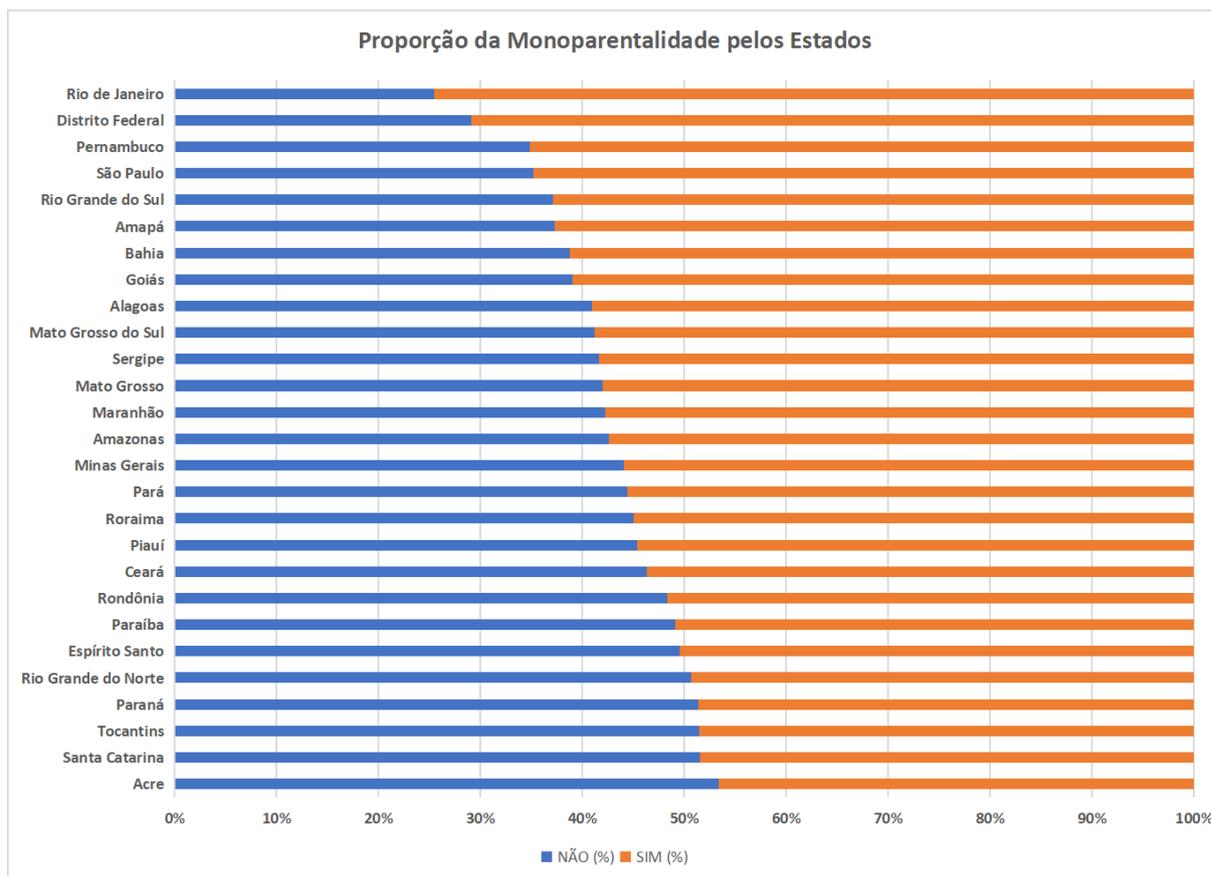


Figura 4.2: Proporção das Famílias Suspensas Monoparentais em cada Estado

4.1.4 Análise da Violência

Os indicadores de violência considerados são: taxas de suicídio, morte violenta por causa indeterminada, homicídio de mulheres e homicídio de homens por 100 mil habitantes. Os dados foram coletados do Mapa da Violência do IPEA, um compêndio de dados do Ministério da Saúde utilizando-se os códigos CID 10 para classificar os tipos de mortes violentas (X60 a X99 e Y00 a Y099) no período de 2014 e 2019. As taxas de violência, além de estarem associadas às características sociais locais, também estão relacionadas ao ambiente familiar. A violência constitui um dos fatores destacados na literatura como fontes de choques entre os membros das famílias [39], especialmente entre os mais pobres, o que pode estar associado a uma maior taxa de suspensão entre os beneficiários do PBF. Estas taxas foram agregadas aos registros das famílias dos beneficiários.

Tabela 4.5: Detalhamento das Variáveis da Dimensão Social Municipal

Dimensão	Variável	Indicador	Descrição	Forma de mensuração	Transformação
Social Municipal	Taxa de Suicídio	Taxa Média de Suicídio	Taxa média municipal de suicídio entre os anos de 2014 a 2019	Taxa de Suicídio por 100 mil habitantes	Discretização da média do Período
	Taxa de Morte Violenta	Taxa Média de Morte Violenta	Taxa média municipal de morte violenta entre os anos de 2014 a 2020	Taxa de Morte Violenta por 100 mil habitantes	Discretização da média do Período
	Taxa de Homicídio de Mulheres	Taxa Média de Homicídio Mulheres	Taxa média de homicídio de mulheres municipal entre os anos de 2014 a 2021	Taxa de Homicídio de Mulheres por 100 mil habitantes	Discretização da média do Período
	Taxa de Homicídio de Homens	Taxa Média de Homicídio de Homens	Taxa média municipal de homicídio de homens entre os anos de 2014 a 2022	Taxa de Homicídio de Homens por 100 mil habitantes	Discretização da média do Período

Fonte: Elaboração própria conforme detalhado na Metodologia

Para possibilitar o uso das taxas de violência no modelo de associação proposto, as variáveis de violência municipal são transformadas em variáveis categóricas utilizando-se técnicas de discretização, conforme explicado anteriormente. Cada partição produz faixas de violência diferentes entre si, de acordo com a homogeneidade de cada partição para a variável alvo SUSPENSO.

As etapas básicas do algoritmo do SPSS de *Binning Optimal* podem ser caracterizadas da seguinte forma:

1. **Preprocessamento (opcional).** A variável de entrada de classe (taxa de violência) é dividida em n partições (onde n pode ser especificado) e cada partição contém o mesmo número de casos ou o mais próximo do mesmo número de casos possível.
2. **Identificação de potenciais pontos de corte.** Cada valor distinto da entrada de classe que não pertence à mesma categoria da variável guia (suspensio), uma vez que o próximo valor distinto maior da variável de entrada de classe é um potencial pontos de corte.
3. **Selecionando recortes.** O potencial ponto de corte que produz o maior ganho de informação é avaliado pelo critério de aceitação do MDLP. Essa avaliação é repetida recursivamente até que nenhum potencial pontos de corte seja aceito. Os recortes aceitos definem os terminais das partições.

A Tabela 4.6 fornece informações sobre as variáveis de entrada de partição. As quatro primeiras colunas dizem respeito aos valores pré-processados. A coluna 5 fornece o valor da entropia do modelo de discretização.

1. A coluna N é o número de casos utilizados na análise. Uma vez que foi utilizado a exclusão *pairwise* de valores ausentes, este valor pode não ser constante. Como esse *dataset* não tem valores ausentes, o valor é simplesmente o número de casos.
2. As colunas **Mínimo** e **Máximo** mostram os menores valores (pré-binning) mínimos e máximos no *dataset* para cada variável de entrada de partição (taxa de violência). Além de dar uma noção da faixa observada de valores para cada variável, estes podem ser úteis para captura de valores fora do intervalo esperado.
3. A coluna **Número de Valores Distintos** informa quais variáveis foram pré-processadas usando o algoritmo *Equal Frequency*. Por padrão, as variáveis com mais de 1000 valores distintos são pré-binadas em 1000 bins distintos. Esses bins pré-processados são então vinculados contra a variável guia usando MDLP. O recurso de pré-processamento é parametrizável (número máximo de n bins especificado).
4. A coluna **Número de Categorias** é o número final de partições geradas pelo procedimento e é muito menor do que o número de valores distintos.
5. A coluna **Entropia do Modelo** dá uma ideia de quão útil cada variável de violência poder ser para prever a variável alvo. Quanto menor a entropia de uma classe, melhor ela pode prever a variável alvo associada.

Tabela 4.6: Estatística Descritiva do Modelo de Discretização por Entropia MDL

	N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo	Macrorregião
Taxa média de suicídio municipal de 2014 a 2019	63 250 143	0.00	44.10	3.080	10	0.115	NÃO-SERRP
	19 210 881	0.00	59.08	1.512	10	0.373	SERRP
Taxa média de Morte Indeterminada municipal de 2014 a 2019	63 250 143	0.00	91.37	2.150	10	0.115	NÃO-SERRP
	19 210 881	0.00	33.47	998	10	0.374	SERRP
Taxa média de homicídio de mulheres municipal de 2014 a 2019	63 250 143	0.00	69.55	1.672	9	0.115	NÃO-SERRP
	63 250 143	0.00	212.14	866	8	0.372	SERRP
Taxa média de homicídio de Homens municipal de 2014 a 2019	19 210 881	0.00	33.13	3.568	10	0.116	NÃO-SERRP
	19 210 881	0.00	151.45	1.523	8	0.372	SERRP

Observando as entropias dos modelos para cada macrorregião, vemos que na macrorregião NÃO-SERRP a entropia do modelo é de 0,115 para todos os tipos de violência. Esta entropia relativamente baixa sugere que, após a discretização, as categorias de violência estão relativamente bem homogêneas em relação à variável SUSPENSO. Isso significa que

existe uma distribuição mais previsível de violência nas categorias criadas pelo modelo de discretização nesta macrorregião.

Para SERRP, a entropia do modelo é aproximadamente 0,373 para todas as violências, o que é significativamente maior do que na macrorregião NÃO-SERRP. Uma entropia maior implica uma maior variação nas taxas de violência municipal em relação à variável SUSPENSO. Isso pode indicar que a relação entre as taxas de violência é mais complexa na macrorregião SERRP ou que os dados são mais heterogêneos, dificultando a categorização.

A macrorregião NÃO-SERRP tem maiores valores no campo **Valores Distintos** do que SERRP para duas categorias de violência, sugerindo que há mais variação nas taxas de suicídio e morte indeterminada antes da discretização. A macrorregião SERRP tem maiores valores **Máximo** de taxa de suicídio, homicídio de mulheres e de homens em comparação com a NÃO-SERRP, o que indica incidências mais altas destas categorias nessa região. As Figuras 4.3 a 4.6 apresentam a distribuição de famílias por faixas discretizadas de violência.

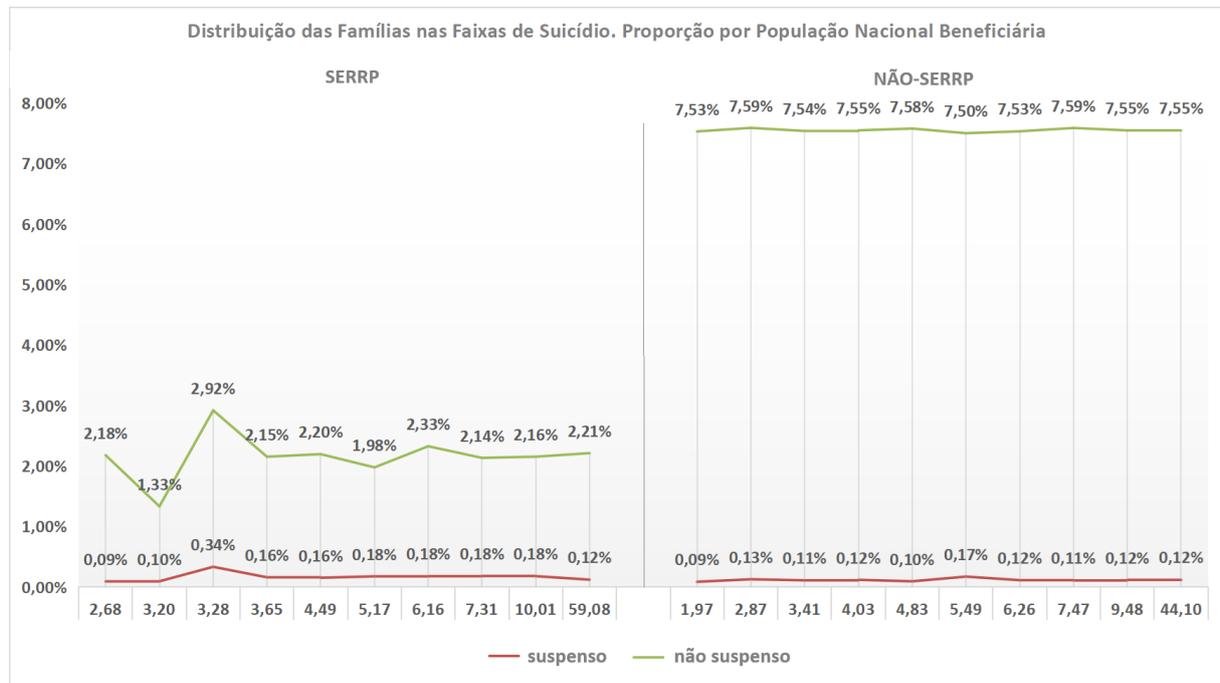


Figura 4.3: (%) de Famílias Suspensas e Não-suspensas por de Famílias por Faixas de Suicídio. Proporção Nacional

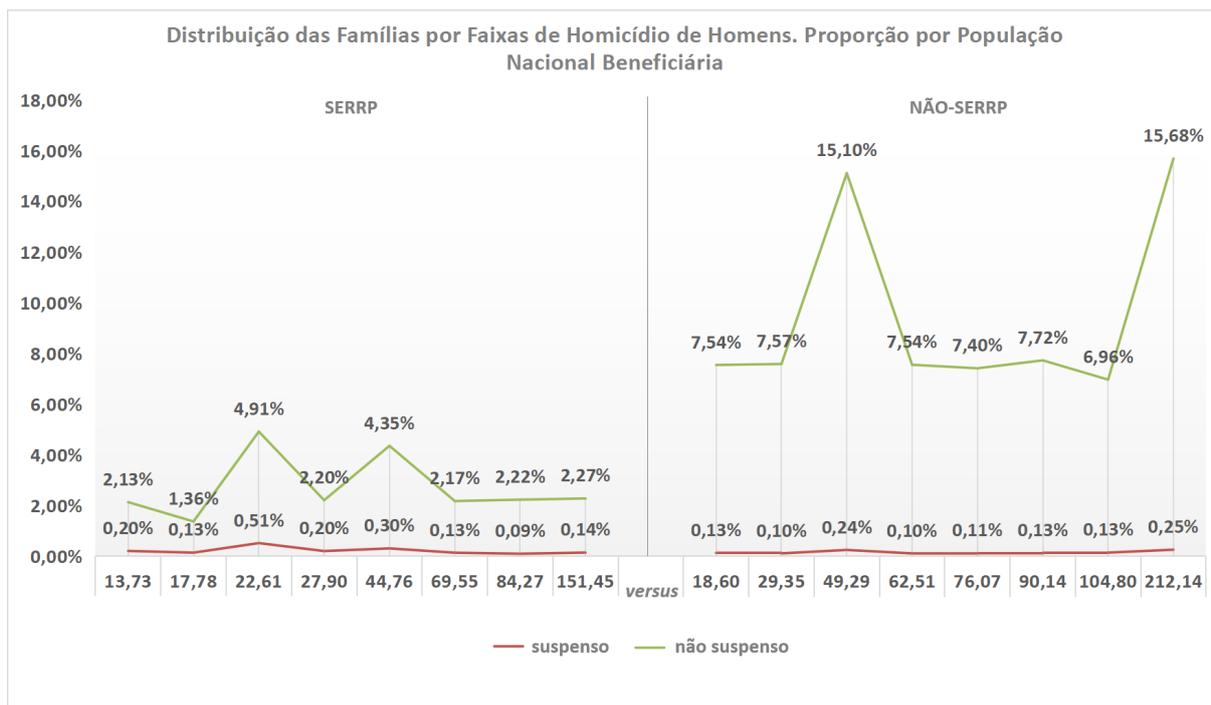


Figura 4.4: (%) de Famílias Suspensas e Não-suspensas por Faixas de Homicídio de Homens. Proporção Nacional.

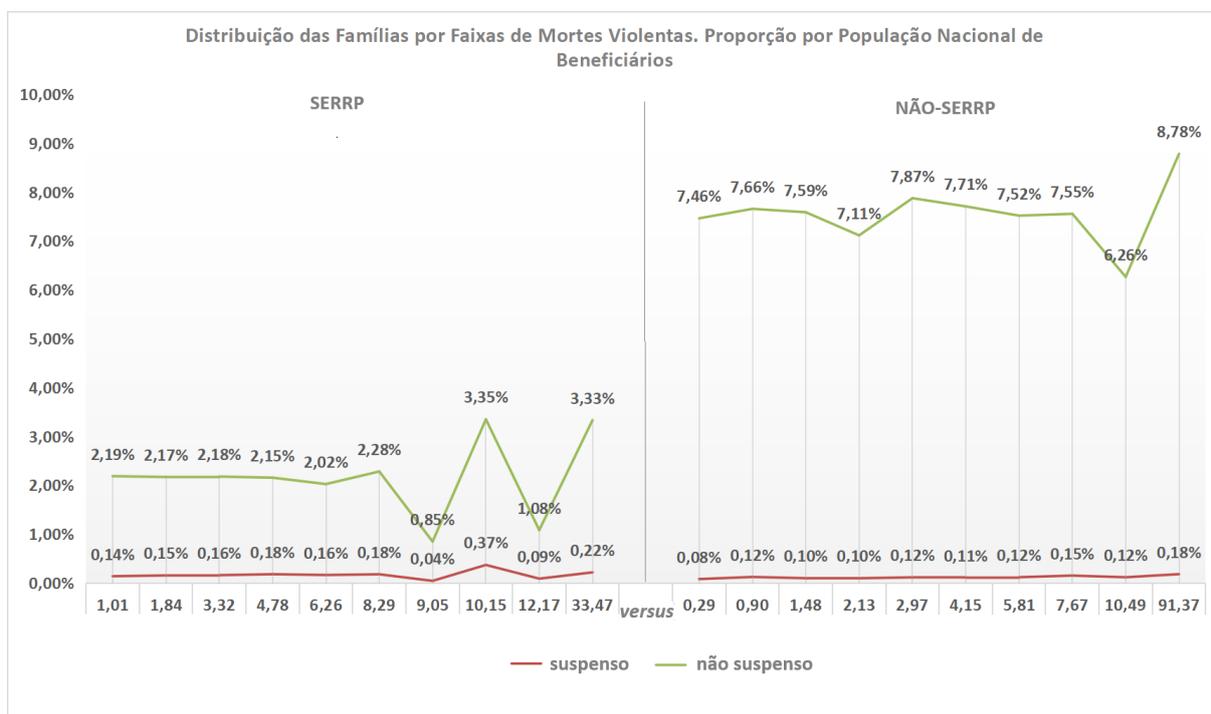


Figura 4.5: (%) de Famílias Suspensas e Não-suspensas por Faixas de Mortes Violentas. Proporção Nacional.

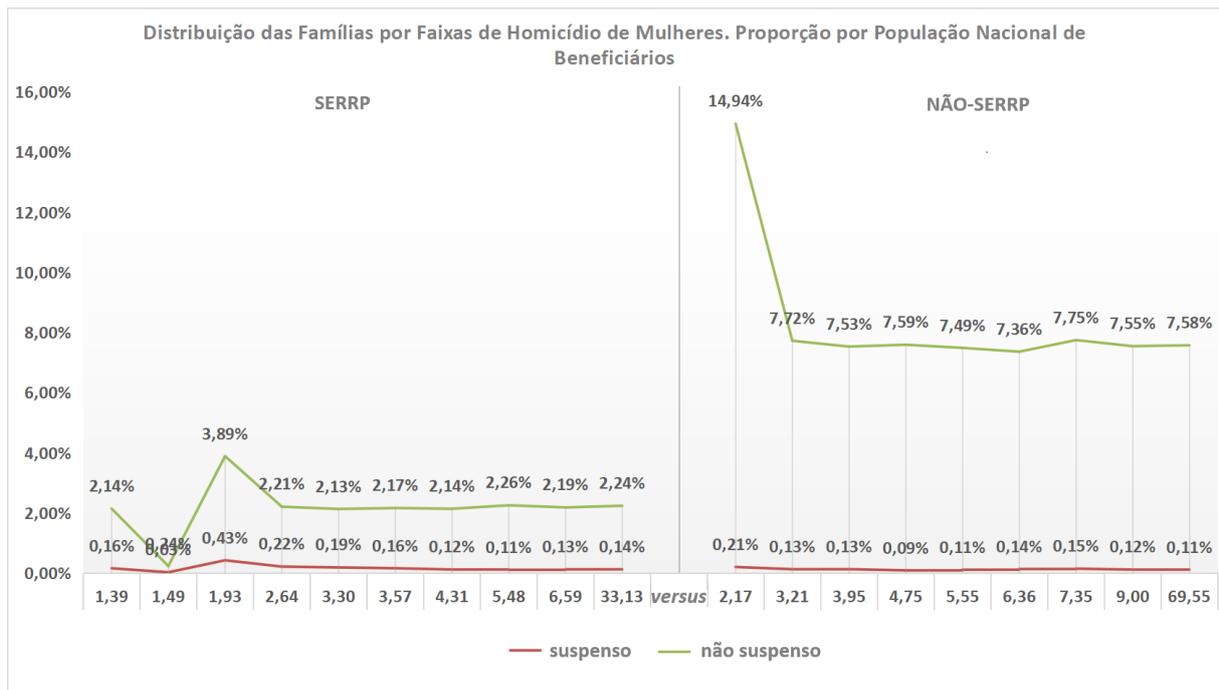


Figura 4.6: (%) de Famílias Suspensas e Não-suspensas por Faixas de Homicídio de Mulheres. Proporção Nacional.

Suicídio

A Tabela I.2 do Anexo I mostra o detalhamento da discretização por Entropia MLD para as médias das taxas municipais de suicídio. Nela vemos que a região NÃO-SERRP tem uma entropia do modelo de 0,115, que é relativamente baixa, sugerindo que a variável de suspensão pode ser bem explicada pelas faixas de suicídio nesta região. Em NÃO-SERRP, a faixa com a maior porcentagem de famílias suspensas (0,17%) é a faixa 6, que vai de 4,83 a 5,49. A região SERRP apresenta uma entropia de modelo mais alta de 0,373, indicando que a associação entre as faixas de suicídio e a suspensão é menos previsível. As porcentagem de famílias suspensas são mais variáveis nesta região, com a faixa 3 (de 3,20 a 3,28) mostrando a maior porcentagem de casos suspensos (0,34%). Em comparação com NÃO-SERRP, a SERRP tem uma proporção significativamente maior de famílias suspensas em todas as faixas.

A Figura 4.3 apresenta as Faixas de Suicídio e Suspensões em cada macrorregião. Nele vemos que na região SERRP as faixas de suicídio mais baixas têm uma menor porcentagem de casos suspensos (por exemplo, 0,09% para a faixa de 2,68) em comparação com as faixas mais altas (por exemplo, 0,34% para a faixa de 3,28). Isso pode sugerir uma associação entre taxas de suicídio mais altas e suspensões. Já na região NÃO-SERRP, as porcentagens de casos suspensos são mais uniformes e baixas em todas as faixas, mas também apresenta diferenças entre a distribuições de famílias suspensas com 0,9% na

faixa 1,97 e 0,17% na faixa 5,49.

Morte Violenta por Causa Indeterminada

A Tabela I.3 do Anexo I mostra o detalhamento da discretização por Entropia MDL para as médias das taxas municipais de morte violenta por causa indeterminada. Pode-se observar que a coluna "Mínimo" e "Máximo" definem os limites das taxas de morte indeterminada, com a máxima em NÃO-SERRP alcançando 91,37 e na SERRP 33,47. As faixas com a maior porcentagem de famílias suspensas em NÃO-SERRP estão concentradas nas faixas mais violentas entre 8 e 10, indicando que as famílias mais vulneráveis residem nos locais mais violentos. Em SERRP, a maior porcentagem de famílias suspensas também está nas faixas 8 e 10.

A Figura 4.5 mostra a proporção de famílias suspensas e não-suspensas do programa bolsa família em cada faixa de taxa de morte indeterminada para as macrorregiões SERRP e NÃO-SERRP. A maior discrepância entre as distribuições de famílias suspensas (SIM) e não-suspensas (NÃO) aparece nas extremidades das taxas de morte indeterminada, indicando que municípios com taxas de morte indeterminada muito altas têm uma proporção maior de suspensões. Para a macrorregião NÃO-SERRP, a maior proporção de famílias suspensas ocorre nas faixas mais altas de taxas de morte indeterminada (>7), indicando uma possível associação entre altas taxas de morte indeterminada e suspensões do programa bolsa família. Para a macrorregião SERRP, há um pico na proporção de suspensões na faixa de 9,05 a 10,15, seguido de uma queda e outro pico na faixa máxima (12,17 a 33,47). A volatilidade nas proporções, especialmente na macrorregião SERRP, pode ser explicada pela maior entropia, que sugere uma distribuição mais ampla das taxas de morte indeterminada.

Homicídio de Mulheres

Análise detalhada da entropia MDL disposta da Tabela I.4 do Anexo I mostra que em NÃO-SERRP o valor da entropia de modelo é 0,115 com 9 categorias e, portanto, a distribuição da variável "suspenso" em NÃO-SERRP é menos incerta (menor entropia) comparada com a SERRP. O percentual de suspensão varia de 1,2% a 1,9% nas diferentes faixas, sendo a maior distribuição de famílias suspensas na faixa 7 (6,36 a 7,35), seguida pela faixa 6 (5,55 a 6,36). Em SERRP, a entropia de modelo é de 0,372 e há 10 categorias e isso pode refletir uma maior variação na forma como a violência contra as mulheres afeta a suspensão dos benefícios nesta macrorregião. A maior distribuição de famílias suspensas ocorre na faixa 3 (1,49 a 1,93), o que representa um pico em faixas de homicídio de mulheres mais baixas que em NÃO-SERRP.

A Figura 4.6 apresenta a distribuição das famílias nas faixas de homicídio de mulheres em SERRP, para as famílias suspensas (SIM), mostra que as porcentagens de famílias suspensas variam de 0,03% a 0,43%, com um pico na faixa de homicídio de 1,93 (0,43%). Isso pode indicar uma associação entre essa faixa de taxa de homicídio e a suspensão de benefícios. Para as famílias não-suspensas (NÃO) a distribuição de famílias não-suspensas é relativamente estável, variando de 2,14% a 3,89%, exceto por um valor mais baixo de 0,24% na faixa de 1,49. A faixa de homicídio de 1,93 se destaca também entre as famílias não-suspensas, com 3,89%, o que é o maior valor entre as faixas. Em NÃO-SERRP, para as famílias suspensas (SIM) as porcentagens de famílias suspensas varia de 0,09% a 0,21%. O valor mais alto de distribuição de famílias suspensas, 0,21%, ocorre na faixa de 2,17, que é consideravelmente menor do que a porcentagem correspondente de famílias não-suspensas. Para as famílias não-suspensas (NÃO) existe uma proporção significativamente maior de famílias não-suspensas na primeira faixa de 2,17, com 14,94%, o que destoa significativamente dos outros valores que variam de 7,36% a 7,75%. Estes valores são mais uniformes e consistentes nas demais faixas em comparação com a SERRP.

Comparando as duas macrorregiões, em SERRP há uma aparente associação entre famílias suspensas e a faixa de taxa de homicídio de 1,93. Esta faixa se destaca tanto para as famílias suspensas quanto para as não-suspensas. Em NÃO-SERRP é alta a proporção de famílias não-suspensas na faixa mais baixa de taxa de homicídio (2,17).

Homicídio de Homens

A Tabela I.5 do Anexo I mostra em detalhes o resultado da discretização por entropia MDL das taxas médias de homicídio de homens por município de 2014 a 2019 com alvo na variável SUSPENSO. O número de categorias é limitado a 8 para ambas as macrorregiões, embora a restrição para o número máximo categorias seja 10, indicando que menos categorias foram necessárias para capturar a variação nos dados. O valor máximo da taxa de homicídio de homens é 212,14 em NAO-SERRP e 151,45 em SERRP, mostrando que NAO-SERRP é mais violenta por 100 mil habitantes.

A Figura 4.4 mostra a distribuição das famílias por faixas de homicídio de homens nas macrorregiões SERRP e NÃO-SERRP, discriminando entre famílias com benefícios suspensos (SIM) e não-suspensos (NÃO). A análise de SERRP mostra que há uma variação na porcentagem de famílias suspensas em relação às faixas de homicídio de homens, com o maior valor sendo 0,51% na faixa de 22,61. O percentual de famílias não-suspensas tende a aumentar com a taxa de homicídio de homens, com um pico na faixa de 22,61 (4,91%) e um segundo pico menos acentuado na faixa de 44,76 (4,35%).

Na análise de NÃO-SERRP (Demais Estados) nota-se que as taxas de suspensão de famílias são estáveis em torno de 0,25% em todas as faixas, sendo a mais alta (0,25%)

na faixa de 212,14. A porcentagem de famílias não-suspensas é muito mais alta nas extremidades da distribuição, particularmente na faixa de 49,29 (15,10%) e na faixa de 212,14 (15,68%).

Comparando as macrorregiões SERRP e NÃO-SERRP, observa-se que SERRP apresenta maior variação nas taxas de não-suspensão em relação às faixas de homicídio de homens, enquanto a NÃO-SERRP mantém taxas de suspensão relativamente constantes. Este gráfico confirma que as famílias de NÃO-SERRP têm uma maior possibilidade de não ter o benefício suspenso em comparação com a SERRP nas faixas mais altas de homicídio.

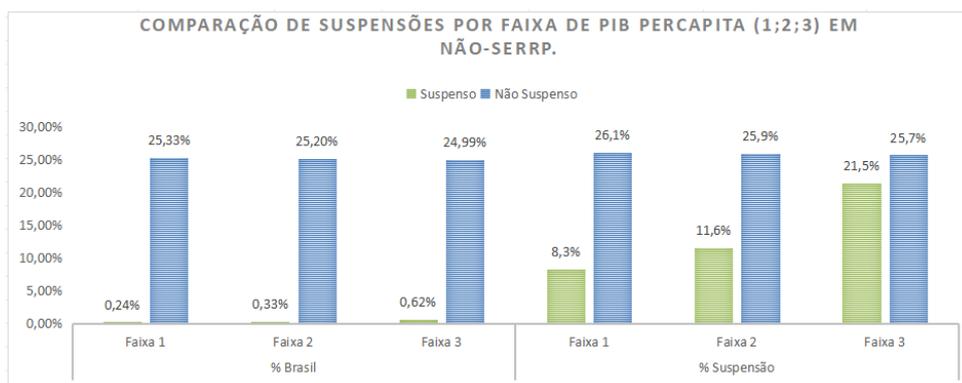
4.1.5 Análise do PIB Per Capita

As variáveis relacionadas ao fator econômico buscam avaliar a premissa de que localidades com melhor desenvolvimento econômico apresentem maior quantidade de acesso a fontes formais de renda, tais como aquelas provenientes de carteira assinada e, conseqüentemente, menor público PBF além de menor proporção de suspensões. O conjunto de dados contempla os valores anuais de PIB per capita municipal entre os anos de 2014 a 2019, mesmo período considerado para a análise de suspensões PBF. Após isso, calcula-se a média municipal do período, obtendo-se o valor médio anual de PIB per capita de cada município dos anos entre 2014 a 2019. A seguir, um modelo de discretização baseado em algoritmo de entropia MDL foi aplicado aos valores médios de PIB per capita dos municípios, gerando 3 faixas de PIB para cada macrorregião SERRP e NÃO-SERRP. A Tabela I.6 do Anexo I apresenta as categorias resultantes do algoritmo de discretização.

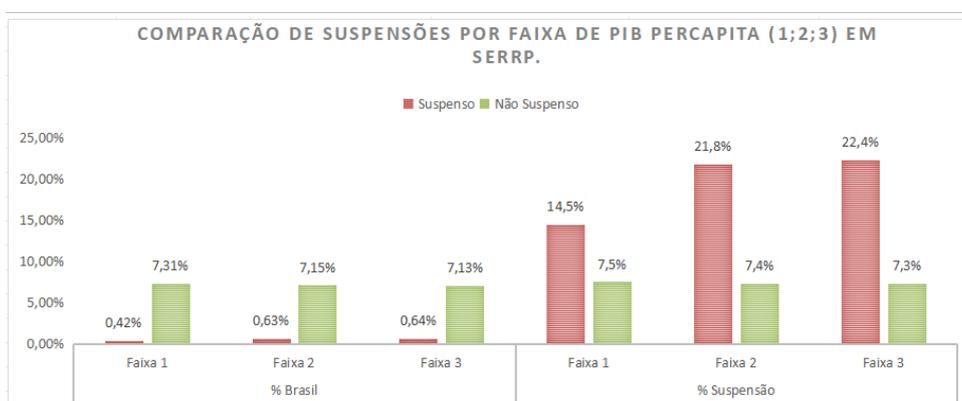
Tabela 4.7: Detalhamento da Variável da Dimensão Econômica Municipal.

Dimensão	Variável	Indicador	Descrição	Forma de mensuração	Transformação
Econômica Municipal	Valor anual de PIB per capita Municipal	Média do PIB Per capita do Período	Média dos valores municipais de PIB per capita entre os anos de 2014 a 2019	Valor do PIB municipal per capita, calculado como sendo o PIB municipal do ano dividido pela população do mesmo ano	Discretização da média do Valor de PIB per capita no período de 2014 a 2019

Fonte: Elaboração própria conforme detalhado na Metodologia.



(a) Distribuição de Famílias por Faixas de PIB per Capita em NÃO-SERRP



(b) Distribuição de Famílias por Faixas de PIB per Capita em SERRP.

Figura 4.7: Distribuição de Famílias Suspensas e Não-suspensas por Faixas de PIB per Capital. Proporções: Brasil e por Suspensão.

Analisando a Tabela I.6 do Anexo I e os Gráficos 4.7 é possível notar que na macrorregião NÃO-SERRP (proporção Brasil) apresenta na Faixa 1 (PIB baixo de 3,85 a 9,19 milhares de reais) a proporção de famílias suspensas com 0,24% das famílias. Na Faixa 2 (PIB médio de 9,19 a 16,95 milhares de reais) a proporção de suspensão aumenta ligeiramente para 0,33% e na Faixa 3 (PIB alto de 16,95 a 304,77 milhares de reais) a proporção de famílias suspensas continua aumentando (0,62%). Considerando proporções por suspensão, o público não-suspense possui uma distribuição homogênea entre as faixas, enquanto o público suspenso apresenta o mesmo padrão com uma distribuição que aumenta progressivamente, conforme aumenta os valores das faixas de PIB per capita, sendo 8,3% no Faixa 1 (PIB baixo de 3,85 a 9,19 milhares de reais), 11,6% na Faixa 2 (PIB médio de 9,19 a 16,95 milhares de reais) e 21,5% no Faixa 3 (PIB alto de 16,95 a 304,77 milhares de reais).

Para a macrorregião SERRP, em relação as proporções Brasil, o padrão é similar à NÃO-SERRP, mas com percentuais muitos maiores para o público suspenso, sendo

0,42% para Faixa 1 (PIB baixo de 7,63 a 26,09 milhares de reais), de 0,63% para o Faixa 2 (PIB médio de 26,09 a 49,68 milhares de reais) e de 0,64% para o Faixa 3 (PIB alto de 49,68 a 473,08 milhares de reais). Já quando comparadas as populações com as proporções por SUSPENSÃO o padrão é inverso a NÃO-SERRP. Pode-ser observar que na Faixa 1 (PIB baixo de 7,63 a 26,09 milhares de reais) a proporção de famílias suspensas em comparação com as não-suspensas é significativamente maior na macrorregião SERRP em comparação com a NÃO-SERRP, com 14,5% das famílias suspensas em SERRP frente a 7,5% não-suspensas em NÃO-SERRP. O mesmo vale para a Faixa 2 (PIB médio de 26,09 a 49,68 milhares de reais) onde as famílias suspensas representam 21,8% frente a 7,4% não-suspensas e para o Faixa 3 (PIB alto de 49,68 a 473,08 milhares de reais), cujas famílias suspensas representam a maior proporção de todas, com 22,4% em comparação com 7,3% de não-suspensas.

O valor de entropia nas tabelas é uma medida da incerteza ou impureza da distribuição das famílias suspensas e não-suspensas dentro de cada faixa de PIB. Observa-se que a entropia é menor na NÃO-SERRP (0,114) do que na SERRP (0,374), indicando uma maior homogeneidade em NÃO-SERRP em relação à suspensão de benefícios e as faixas de PIB.

Com base nas proporções fornecidas, é possível observar que, de fato, tanto em NÃO-SERRP quanto em SERRP **se destaca a tendência de maior suspensão do benefício em famílias situadas em faixas de PIB mais altas**, independentemente do tipo de proporção avaliada.

4.2 Associações e o Tamanho do Efeito

Um dos problemas com o teste de independência é que o tamanho da estatística qui-quadrado pode não fornecer um guia confiável para a força da relação estatística entre as duas variáveis. Quando duas tabelas de classificação cruzada diferentes têm o mesmo tamanho de amostra, as duas variáveis na tabela com o maior valor qui-quadrado estão mais fortemente relacionadas do que as duas variáveis na tabela com o menor valor qui-quadrado. Mas quando os tamanhos de amostra para duas tabelas diferem, o tamanho da estatística qui-quadrado é um indicador enganoso da extensão da relação entre duas variáveis [40]. As medidas de associação coeficiente de contingência e V de Cramer visam ajustar a estatística qui-quadrado pelo tamanho da amostra.

O V de Cramer (ϕ') é uma medida de associação baseada em qui-quadrado que usa as informações sobre as dimensões da tabela, corrigindo o problema de que medidas de associação para tabelas de dimensões diferentes podem ser difíceis de serem diretamente comparadas. O V de Cramer é igual a 0 quando não há relação entre as duas variáveis,

e geralmente tem valor máximo de 1, independentemente da dimensão da tabela ou do tamanho da amostra.

Neste estudo, a representatividade populacional da amostra exige o uso de métricas baseadas em cálculos que utilizam proporções relativas e corrigem as distorções provocadas pelo valor do qui-quadrado, dado que este utiliza cálculos baseados em valores absolutos. Por isso, as medidas de associação adotadas são valores tabelados propostos por Cohen[12], que representam uma equivalência entre Ω (ω) em termos V de Cramer (ϕ'). Propõe-se em seguida uma alteração destas faixas, calculando-se tamanhos intermediários para uma melhor compreensão da extensão da relação entre as duas variáveis. O valor das medidas de associação avalia o poder das associações nas tabelas de contingência, ou seja, o tamanho do efeito de uma variável sobre a outra (pequeno, médio ou grande). Por convenção, o efeito de tamanho “irrelevante” é adotado para especificar tamanhos menores que a menor faixa de “pequeno”. O resumo pode ser visto na Tabela 4.8:

Tabela 4.8: Tabela de Equivalência de Medidas Simétricas: ω em Termos de C, ϕ e ϕ' .

Tamanho do Efeito	ϕ'										
	w	c	$r = 2^*$	3	4	5	6	7	8	9	10
Irrelevante	-	-	-	-	-	-	-	-	-	-	-
pequeno-	0,10	0,100	0,100	0,071	0,058	0,050	0,045	0,041	0,038	0,035	0,033
pequeno	0,150	0,148	0,150	0,106	0,087	0,075	0,067	0,062	0,057	0,053	0,050
pequeno+	0,20	0,196	0,200	0,141	0,115	0,100	0,089	0,082	0,076	0,071	0,067
medio-	0,30	0,287	0,300	0,212	0,173	0,150	0,134	0,122	0,113	0,106	0,100
medio	0,350	0,329	0,350	0,248	0,202	0,175	0,157	0,143	0,132	0,124	0,117
medio+	0,40	0,371	0,400	0,283	0,231	0,200	0,179	0,163	0,151	0,141	0,133
grande-	0,50	0,447	0,500	0,354	0,289	0,250	0,224	0,204	0,189	0,177	0,167
grande	0,70	0,573	0,700	0,495	0,404	0,350	0,313	0,286	0,245	0,247	0,233
grande+	0,90	0,669	0,900	0,636	0,52	0,450	0,402	0,367	0,327	0,318	0,300

Fonte: Baseada em Cohen, 1992.

*Esta coluna fornece os equivalentes em termos de ϕ para a tabela (2×2).

À medida que aumenta a extensão da relação entre as variáveis, cada uma dessas medidas também aumenta, embora em quantidades diferentes. Onde essas medidas diferem é em seu valor máximo. Se houver uma relação perfeita entre as duas variáveis de uma tabela de classificação cruzada, seria preferível que a medida de associação tivesse o valor 1. O V de Cramer é a medida preferida entre essas medidas baseadas ϕ . Ele geralmente tem um valor máximo de 1 quando existe uma relação muito forte entre duas variáveis. Quando é calculado, o V de Cramer leva em consideração as dimensões da tabela, o que implica que o V para tabelas de diferentes dimensões pode ser comparado de forma significativa. Ao comparar várias tabelas da mesma dimensão e tamanho de amostra semelhante, faz pouca diferença qual dessas medidas é usada.

4.2.1 Associação entre Monoparentalidade e a Exclusão Escolar

No contexto da monoparentalidade, a hipótese nula testa as chances de exclusão escolar por descumprimento das condicionalidades da educação não estarem associadas à existência de arranjos familiares monoparentais.

Crosstab						
				MACRORREGIÃO		
				NAO-SERRP	SERRP	Total
Não-Suspensão	Monoparental?	Não	% em Monoparental?	81,6%	18,4%	100,0%
			% em MACRORREGIÃO	48,1%	37,9%	45,8%
			% do Total	37,4%	8,4%	45,8%
		Sim	% em Monoparental?	74,5%	25,5%	100,0%
			% em MACRORREGIÃO	51,9%	62,1%	54,2%
			% do Total	40,3%	13,8%	54,2%
	Total	% em Monoparental?	77,8%	22,2%	100,0%	
		% em MACRORREGIÃO	100,0%	100,0%	100,0%	
		% do Total	77,8%	22,2%	100,0%	
Suspensão	Monoparental?	Não	% em Monoparental?	45,8%	54,2%	100,0%
			% em MACRORREGIÃO	42,9%	35,7%	38,6%
			% do Total	17,7%	20,9%	38,6%
		Sim	% em Monoparental?	38,5%	61,5%	100,0%
			% em MACRORREGIÃO	57,1%	64,3%	61,4%
			% do Total	23,6%	37,7%	61,4%
	Total	% em Monoparental?	41,3%	58,7%	100,0%	
		% em MACRORREGIÃO	100,0%	100,0%	100,0%	
		% do Total	41,3%	58,7%	100,0%	
Total	Monoparental?	Não	% em Monoparental?	80,8%	19,2%	100,0%
			% em MACRORREGIÃO	48,0%	37,7%	45,6%
			% do Total	36,9%	8,8%	45,6%
		Sim	% em Monoparental?	73,3%	26,7%	100,0%
			% em MACRORREGIÃO	52,0%	62,3%	54,4%
			% do Total	39,9%	14,5%	54,4%
	Total	% em Monoparental?	76,7%	23,3%	100,0%	
		% em MACRORREGIÃO	100,0%	100,0%	100,0%	
		% do Total	76,7%	23,3%	100,0%	

Figura 4.8: Monoparentalidade e Macrorregião

A *Crosstab* 4.8 mostra a análise bivariada, que considera a interação entre o local de residência em SERRP ou NÃO-SERRP e a existência ou não de monoparentalidade na família, controlando o efeito da variável “suspensão”. O controle é feito utilizando a estratificação, uma estratégia que visa excluir o efeito confundidor numa associação quando há classes muito desequilibradas no conjunto de dados, como das classes “SIM” (suspensão) e “NÃO” (não-suspensão) da suspensão, que possuem proporções aproximadas de 3,00% e 97,00%, respectivamente. Analisando a Figura 4.8, tem-se que:

1. Famílias Não-Suspensas

(a) Famílias Não-Monoparentais

- i. NAO-SERRP: 81,6% das famílias não-monoparentais pertencem à região NÃO-SERRP. Contribuem com 48,1% das famílias da região NÃO-SERRP. Representam 37,4% do total de famílias na base de dados.

- ii. SERRP: 18,4% das famílias não-monoparentais pertencem à região SERRP. Contribuem com 37,9% das famílias da região SERRP. Representam 8,4% do total de famílias.

(b) Famílias Monoparentais

- i. NAO-SERRP: 74,5% das famílias monoparentais pertencem à região NÃO-SERRP. Contribuem com 51,9% das famílias da região NÃO-SERRP. Representam 40,3% do total de famílias.
- ii. SERRP: 25,5% das famílias monoparentais pertencem à região SERRP. Contribuem com 62,1% das famílias da região SERRP. Representam 13,8% do total de famílias.

2. Famílias Suspensas

(a) Famílias Não-Monoparentais

- i. NAO-SERRP: 45,8% das famílias não-monoparentais estão na região NÃO-SERRP. Contribuem com 42,9% das famílias da região NÃO-SERRP. Representam 17,7% do total de famílias.
- ii. SERRP: 54,2% das famílias não-monoparentais estão na região SERRP. Contribuem com 35,7% das famílias da região SERRP. Representam 20,9% do total.

(b) Famílias Monoparentais

- i. NAO-SERRP: 38,5% das famílias monoparentais estão na região NÃO-SERRP. Contribuem com 57,1% das famílias da região NÃO-SERRP. Representam 23,6% do total.
- ii. SERRP: 61,5% das famílias monoparentais estão na região SERRP. Contribuem com 64,3% das famílias da região SERRP. Representam 37,7% do total.

3. Total Geral

(a) Famílias Não-Monoparentais

- i. NAO-SERRP: 80,8% das famílias não-monoparentais estão na região NÃO-SERRP. Contribuem com 48,0% das famílias da região NÃO-SERRP. Representam 36,9% do total.
- ii. SERRP: 19,2% das famílias não-monoparentais estão na região SERRP. Contribuem com 37,7% das famílias da região SERRP. Representam 8,8% do total.

(b) Famílias Monoparentais

- i. NAO-SERRP: 73,3% das famílias monoparentais estão na região NÃO-SERRP. Contribuem com 52,0% das famílias da região NÃO-SERRP. Representam 39,9% do total.
- ii. SERRP: 26,7% das famílias monoparentais estão na região SERRP. Contribuem com 62,3% das famílias da região SERRP. Representam 14,5% do total.

A análise das proporções permite concluir que, acerca das diferenças regionais, as famílias monoparentais são mais representativas na região SERRP do que na NÃO-SERRP, especialmente entre as famílias suspensas (61,5% na SERRP contra 38,5% na NÃO-SERRP). No que se refere ao impacto da suspensão, os resultados sugerem que as famílias monoparentais são mais afetadas pela suspensão, representando uma proporção maior entre famílias suspensas (61,4%) em comparação às famílias não suspensas (54,2%). O binômio monoparentalidade *versus* suspensão revela que a maior vulnerabilidade de famílias monoparentais em SERRP sugere a necessidade de intervenções regionais para mitigar os impactos sociais e econômicos da suspensão do benefício. Esses resultados podem ajudar a orientar políticas públicas, com foco em famílias monoparentais e no contexto regional das suspensões do Bolsa Família.

Olhando para monoparentalidade, em relação às famílias **não-monoparentais suspensas**, a região NÃO-SERRP (74,5%) ultrapassa a SERRP (25,5%) em termos de proporção. Isso mostra que na região NÃO-SERRP, suspensões estão mais associadas às famílias não-monoparentais. No caso das famílias **monoparentais com beneficiário suspenso**, vemos uma inversão: a região SERRP mostra uma proporção maior (61,5%) que a NÃO-SERRP (38,5%). Esse achado reforça a ideia de que as famílias monoparentais na região SERRP estão mais sujeitas a terem membros suspensos ou que elas enfrentam desafios mais significativos que as levam a ter maiores índices de suspensões no programa.

4.2.2 Associação entre Violência e a Exclusão Escolar

Esta seção apresenta os resultados do tamanho do efeito (força) das relações entre violência e a suspensão de benefício onde, combinados com a análise do qui-quadrado, são cruciais para compreender a relevância prática dos resultados estatísticos destas relações, destacando que as associações variam significativamente entre as subpopulações. O valor do qui-quadrado de cada categoria foi calculado individualmente, para identificar qual categoria, dentre as existentes, mais contribui para o valor total do χ^2 .

Para fazer a análise da associação bivariada das taxas de suicídio e da variável macroregião, faz-se necessário controlar o efeito da variável SUSPENSO, dado que devemos

considerar o desequilíbrio nas proporções de famílias suspensas e não-suspensas. Analisando a Tabela 4.9, algumas observações específicas podem ser mencionadas. Para NÃO-SERRP, a faixa 6 (que tem a maior porcentagem de famílias suspensas, 14,6%) apresenta uma discrepância em comparação com a porcentagem de famílias não-suspensas (9,9%). Para SERRP, a faixa 3 se destaca com a maior porcentagem de famílias suspensas (19,8%) em comparação com a porcentagem de famílias não-suspensas (13,52%). Essas diferenças já indicam que há uma associação significativa entre a macrorregião e as taxas de suicídio categorizadas. O valor de qui-quadrado alto sugere que essa associação é muito forte.

Suicídio

Figura 4.9: Suicídio e Macrorregião.

VARIÁVEIS	FAIXA DE SUICÍDIO	Crosstab						MEDIDAS		
		NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	6.210.637	10,0%	43,79	1.798.105	10,1%	153,04	0,071	406.357,7500	0,000
	2	6.259.152	10,1%	50.320,90	1.100.599	6,2%	175.885,19			
	3	6.217.029	10,0%	35.846,70	2.409.329	13,5%	125.293,93			
	4	6.225.433	10,0%	1,63	1.777.012	10,0%	5,68			
	5	6.249.924	10,0%	57,18	1.812.458	10,2%	199,86			
	6	6.186.441	9,9%	1.897,62	1.631.813	9,2%	6.632,71			
	7	6.212.299	10,0%	1.947,26	1.920.055	10,8%	6.806,20			
	8	6.257.843	10,0%	80,43	1.761.572	9,9%	281,13			
	9	6.224.714	10,0%	1,04	1.777.619	10,0%	3,65			
	10	6.223.681	10,0%	200,17	1.826.123	10,3%	699,64			
SUSPENSO	1	73.041	7,4%	1.779,81	78.218	5,6%	1.253,07	0,184	80.655,2240	0,000
	2	107.059	10,9%	11.596,64	79.667	5,7%	8.164,60			
	3	92.857	9,4%	23.496,33	276.900	19,8%	16.542,56			
	4	99.575	10,1%	96,11	134.060	9,6%	67,66			
	5	79.310	8,1%	469,03	127.989	9,2%	330,22			
	6	143.514	14,6%	4.593,49	147.008	10,5%	3.234,04			
	7	94.980	9,7%	345,84	149.202	10,7%	243,49			
	8	90.999	9,3%	870,06	151.866	10,9%	612,56			
	9	98.805	10,1%	137,40	149.424	10,7%	96,74			
	10	102.850	10,5%	3.946,82	101.862	7,3%	2.778,75			

Utilizando o guia de interpretação para o tamanho do efeito (Tabela de Equivalência 4.8) vemos que os valores de V de Cramer indicam que na subpopulação não suspensa (V de Cramer = 0,071) há uma associação pequena entre as taxas de suicídio e macrorregião. Embora a associação seja estatisticamente significativa (como mostrado pelo p-valor do qui-quadrado), a intensidade dessa associação é fraca. Na subpopulação de suspensos (V de Cramer = 0,184) há uma associação forte entre as taxas de suicídio e a suspensão de benefícios. A medida de V de Cramer de 0,184 sugere que a variável de taxa de suicídio tem uma influência substancial nas chances de uma família ter benefícios suspensos, especialmente nos municípios onde as taxas de suicídio variam entre 2,87 e 3,28 (faixas 3 de suicídio, com a taxa variando no intervalo [2.87, 3.41) para NÃO-SERRP e [3.20, 3.28) para SERRP).

Morte Indeterminada por Causa Violenta

Figura 4.10: Morte Violenta por Causa Indeterminada e Macrorregião.

VARIÁVEIS	FAIXA DE MORTE INDETERMINADA	Crosstab						MEDIDAS		
		NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	6153275	9,9%	171,84	1802388	10,1%	600,62	0,129	1.338.462,2050	0,000
	2	6312554	10,1%	31,57	1787891	10,0%	110,35			
	3	6255358	10,0%	2,95	1795187	10,1%	10,30			
	4	5865248	9,4%	992,13	1776800	10,0%	3.467,76			
	5	6492842	10,4%	3.399,20	1668716	9,4%	11.881,14			
	6	6354284	10,2%	382,84	1881646	10,6%	1.338,13			
	7	6197980	10,0%	130.040,97	699223	3,9%	454.528,43			
	8	6229121	10,0%	83.050,39	2762140	15,5%	290.283,61			
	9	5164540	8,3%	44.027,72	891965	5,0%	153.888,81			
	10	7241951	11,6%	35.649,34	2748729	15,4%	124.604,12			
SUSPENSO	1	66523	6,8%	893,33	114257	8,2%	628,95	0,215	110.470,7180	0,000
	2	101661	10,3%	520,32	127410	9,1%	366,33			
	3	80683	8,2%	596,31	132124	9,5%	419,83			
	4	80058	8,1%	2.141,43	148099	10,6%	1.507,67			
	5	95164	9,7%	0,11	134916	9,7%	0,08			
	6	90146	9,2%	613,21	146792	10,5%	431,73			
	7	97786	9,9%	33.002,87	35650	2,6%	23.235,63			
	8	124868	12,7%	15.573,55	304588	21,8%	10.964,54			
	9	96875	9,9%	10.152,20	72991	5,2%	7.147,64			
	10	149226	15,2%	1.335,05	179369	12,8%	939,94			

Observando a Figura 4.10, verifica-se que na subpopulação não-suspensa a maior contribuição para o qui-quadrado em NAO-SERRP é vista na faixa 7, seguida pela faixa 8. As faixas 9 e 10 também apresentam contribuições importantes para o qui-quadrado em ambas as macrorregiões, implicando diferenças importantes nas proporções esperadas. Já para a subpopulação **suspensa**: a faixa 7 em NAO-SERRP e em SERRP têm as maiores contribuições para o qui-quadrado. Isso sugere que, para as famílias suspensas, as taxas mais altas de morte indeterminada estão associadas com a suspensão do benefício, particularmente nas faixas 7, 8 e 9, nas duas macrorregiões.

No que se refere ao tamanho do efeito, o valor de V de Cramer 0,129 para não-suspensos indica uma força de associação moderada com a suspensão, enquanto um V de Cramer de 0,215 para suspensos sugere que a força da associação é grande com a suspensão. Esses valores mostram que existe uma associação estatisticamente significativa entre as faixas de morte indeterminada e a suspensão do benefício, sendo mais pronunciada na macrorregião SERRP em função dos seus altos valores de qui-quadrado. A significância estatística é confirmada pelos p-valor de 0,000 em ambos os grupos (não-suspensa e suspensa), o que indica que as diferenças observadas não são prováveis de ocorrer por acaso.

Homicídio de Mulheres

Figura 4.11: Crosstab de Homicídio de Mulheres e Macrorregião.

VARIÁVEIS		Crosstab						MEDIDAS			
		HOMICÍDIO DE MULHERES	NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	12318434	19,8%	170.189,37	1768373	9,9%	594.857,98	0,343	9.396.952,0340	0,000	
	2	6370069	10,2%	313.777,57	194827	1,1%	1.096.737,67				
	3	6211210	10,0%	168.732,61	3206626	18,0%	589.766,22				
	4	6261572	10,1%	92,84	1822510	10,2%	324,51				
	5	6175750	9,9%	4,57	1760055	9,9%	15,99				
	6	6070745	9,7%	237,69	1785852	10,0%	830,77				
	7	6388999	10,3%	367,44	1765820	9,9%	1.284,30				
	8	6222812	10,0%	622,55	1860802	10,4%	2.175,98				
	9	6247562	10,0%	21,97	1802514	10,1%	76,79				
	10	*	*	*	1847306	10,4%	5.020.473,50				
SUSPENSO	1	169723	17,3%	15.608,66	133957	9,6%	10.989,26	0,355	300.555,2140	0,000	
	2	108026	11,0%	51.823,17	24504	1,8%	36.486,03				
	3	109360	11,1%	34.966,30	353182	25,3%	24.617,98				
	4	76614	7,8%	8.122,36	179821	12,9%	5.718,54				
	5	94381	9,6%	868,06	157038	11,2%	611,16				
	6	114043	11,6%	1.179,92	135297	9,7%	830,72				
	7	123996	12,6%	10.537,36	100458	7,2%	7.418,81				
	8	97528	9,9%	4.768,15	91783	6,6%	3.357,01				
	9	89319	9,1%	998,98	105179	7,5%	703,33				
	10	*	*	*	114977	8,2%	33.445,25				

A Figura 4.11 oferece uma visão detalhada do relacionamento entre as taxas de homicídio de mulheres e o status de suspensão nas macrorregiões SERRP e NÃO-SERRP, com base em um teste de qui-quadrado de independência. O valor p-value aponta para a significância estatística do relacionamento testado. A análise dos resultados do teste de independência com qui-quadrado mostra que para a subpopulação de não-suspensos em NÃO-SERRP há um alto valor de qui-quadrado em certas faixas de homicídio, como na faixa 1 e 2, indicando as faixas que mais contribuíram para a forte associação entre as taxas de homicídio de mulheres e a não-suspensão. A variação nos valores de χ^2 em cada categoria sugere diferentes níveis de associação entre as taxas de homicídio de mulheres e o status de não-suspensa. Em SERRP, assim como em NÃO-SERRP, observa-se uma forte associação na faixa 2, mas a faixa 3 é particularmente sobressalente, com 18% e um valor de qui-quadrado muito alto. O V de Cramer de 0,343 indica uma associação de magnitude forte no contexto geral de não-suspensos.

Já para a subpopulação de suspensos em NÃO-SERRP o valor de qui-quadrado varia, com a faixa 2 tendo o maior valor de χ^2 , seguido pela faixa 3, sugerindo pontos de interesse específicos onde a taxa de suspensão é mais fortemente associada às taxas de homicídio de mulheres. Em SERRP, a faixa 3 se destaca novamente com uma alta porcentagem de suspensão (25,3%) e um segundo valor de qui-quadrado mais alto, destacando uma possível maior contribuição para a associação importante entre as taxas de homicídio e suspensões de benefícios nesta região. O V de Cramer de 0,355 indica uma associação

de magnitude alta para os suspensos, ligeiramente mais forte do que na população não-suspensa.

Ambos os segmentos (suspensos e não-suspensos) indicam que há uma diferença estatisticamente significativa na distribuição das taxas de homicídio de mulheres entre as macrorregiões para os beneficiários do programa, tanto suspensos quanto não-suspensos, o que é confirmado pelo valor do tamanho de efeito com V de Cramer que é grande.

Homicídio de Homens

Figura 4.12: Homicídio de Homens e Macrorregião

Crosstab								MEDIDAS		
VARIÁVEIS	HOMICÍDIO DE HOMENS	NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	6213553	10,0%	58,45	1753231	9,8%	204,29	0,169	2.280.904,421	0,00
	2	6245769	10,0%	46.446,90	1123442	6,3%	162.344,51			
	3	12453026	20,0%	11.020,72	4046411	22,7%	38.520,41			
	4	6220491	10,0%	120,18	1814929	10,2%	420,05			
	5	6105664	9,8%	270.695,04	3583458	20,1%	946.152,54			
	6	6364137	10,2%	104,59	1787671	10,0%	365,57			
	7	5737192	9,2%	3.772,58	1833067	10,3%	13.186,20			
	8	12927321	20,8%	175.182,40	1872476	10,5%	612.309,98			
SUSPENSO	1	103844	10,6%	452,58	164645	11,8%	318,64	0,261	162.122,997	0,00
	2	86181	8,8%	352,68	109484	7,8%	248,31			
	3	197249	20,1%	13.052,61	419782	30,1%	9.189,68			
	4	84651	8,6%	2.814,55	161155	11,5%	1.981,58			
	5	89906	9,1%	17.618,32	247680	17,7%	12.404,16			
	6	105137	10,7%	3.444,79	107251	7,7%	2.425,30			
	7	110817	11,3%	15.472,51	74168	5,3%	10.893,40			
	8	205205	20,9%	41.931,84	112031	8,0%	29.522,05			

A Figura 4.12 apresenta a distribuição das taxas de homicídio de homens em relação à suspensão do benefício em famílias das macrorregiões SERRP e NÃO-SERRP, com base no teste qui-quadrado de independência. Os resultados indicam que para a subpopulação de não-suspensos NÃO-SERRP apresenta porcentagens estáveis de não-suspensos (cerca de 10%), com um pico significativo de 20,8% na faixa 8. A contribuição para o qui-quadrado varia consideravelmente, sendo maior na faixa 5. Em SERRP nota-se também um padrão estável, mas tem um aumento significativo na porcentagem de não-suspensos nas faixas 3 e 5 (22,7% e 20,1%, respectivamente). As contribuições para o qui-quadrado refletem essas porcentagens, sendo especialmente alta na faixa 5. A maior contribuição para o qui-quadrado entre o homicídio de homens e a não-suspensão está na faixa 5 de homicídio de homens em SERRP e em NAO-SERRP.

Para a subpopulação de suspensos, NÃO-SERRP apresenta faixa 8 com a maior porcentagem de suspensos (20,9%) e a maior contribuição para o qui-quadrado. As porcentagens de suspensos nas outras faixas são mais baixas, sugerindo uma associação específica nessa faixa. Em SERRP, a maior porcentagem de suspensos está na faixa 3 (30,1%) que também possui uma grande contribuição para o qui-quadrado, mas não a maior e mais

relevante que é a faixa 9. Em SERRP nota-se uma variação nas porcentagens de suspensos indicando uma possível associação mais forte em faixas específicas de homicídio de homens. A maior contribuição para o qui-quadrado entre o homicídio de homens e a suspensão está na faixa mais alta (faixa 8) de homicídio de homens em SERRP e em NAO-SERRP.

Avaliando as medidas de associação, o V de Cramer possui valores de 0,169 para não-suspensos e 0,261 para suspensos, sugerindo importante associação entre as faixas de homicídio de homens e o status de suspensão, sendo moderada no grupo não-suspensos e forte no grupo de suspensos, ou seja, é estatisticamente significativa em ambas as macrorregiões, especialmente para o público suspenso.

4.2.3 Associação entre PIB Per Capita, Violência e a Exclusão Escolar

Para entender como as variáveis sociais e econômicas podem afetar as chances de suspensão de benefício PBF decorrentes de exclusão escolar, esta seção apresenta os resultados do tamanho do efeito das relações entre a média categorizada da violência e a média categorizada de PIB per capita, separadas por macrorregião SERRP e NÃO-SERRP. São combinadas as análises do qui-quadrado e do V de Cramer em cada região, fornecendo uma visão relevante da estatística destas relações entre as subpopulações de suspensos e não-suspensos.

Figura 4.13: Suicídio e Média Categorizada de PIB Per Capita de NÃO-SERRP

Crosstab NÃO-SERRP											MEDIDAS		
VARIÁVEIS	FAIXA DE SUICÍDIO	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	3332408	16,0%	752.959,06	2021432	9,7%	1.304,28	856797	4,2%	699.850,55	0,163	3.318.833,5040	0,000
	2	2160909	10,4%	1.948,85	2044314	9,8%	982,82	2053929	10,0%	167,09			
	3	1975643	9,5%	5.520,03	1972290 _a	9,5%	5.139,13	2269096 _b	11,0%	21.525,73			
	4	1780691	8,5%	44.599,90	1579096	7,6%	119.939,63	2865646	13,9%	313.798,16			
	5	1824545	8,7%	34.646,90	2642025	12,7%	147.876,04	1783354	8,6%	39.560,38			
	6	1334356	6,4%	262.969,44	1786927	8,6%	37.535,15	3065158	14,9%	504.608,31			
	7	1631333	7,8%	97.274,81	2398775	11,5%	50.860,98	2182191	10,6%	7.616,23			
	8	2231592	10,7%	8.699,17	1973021	9,5%	6.460,11	2053230	10,0%	171,93			
	9	2255868	10,8%	13.927,15	2212994	10,6%	8.751,58	1755852	8,5%	45.216,67			
	10	2333801	11,2%	29.664,39	2157157	10,4%	3.031,64	1732723	8,4%	52.227,40			
SUSPENSO	1	31156	15,8%	18.700,15	24355	8,8%	742,70	17530	3,4%	10.997,79	0,192	72.643,9650	0,000
	2	19763	10,0%	129,72	25739	9,3%	601,19	61557	12,0%	628,36			
	3	20367	10,4%	170,37	27785	10,1%	121,54	44705	8,8%	262,22			
	4	14289	7,3%	1.597,71	22295	8,1%	1.121,93	62991	12,3%	2.439,77			
	5	19458	9,9%	808,33	29847	10,8%	2.623,66	30005	5,9%	3.051,98			
	6	12815	6,5%	8.814,22	27473	10,0%	4.026,60	103226	20,2%	10.992,43			
	7	14354	7,3%	1.141,41	29977	10,9%	428,26	50649	9,9%	33,35			
	8	19229	9,8%	56,39	23979	8,7%	89,19	47791	9,4%	5,17			
	9	22056	11,2%	262,36	30893 _a	11,2%	374,67	45856 _b	9,0%	588,60			
	10	23281	11,8%	352,32	32970 _a	12,0%	601,94	46599 _b	9,1%	879,64			

Figura 4.14: Suicídio e Média Categorizada de PIB Per Capita de SERRP

Crosstab SERRP											MEDIDAS		
VARIÁVEIS	FAIXA DE SUICÍDIO	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	1546788	25,6%	1.445.615,14	190787	3,2%	275.261,98	60530	1,0%	478.762,18	0,567	11.438.861,7070	
	2	324070	5,4%	6.318,39	577743	9,8%	124.537,52	198786	3,4%	74.543,82			
	3	6870	0,1%	801.956,76	18250	0,3%	762.159,30	2384209	40,5%	3.173.257,33			
	4	142870	2,4%	349.768,80	181118	3,1%	282.229,06	1453024	24,7%	1.279.287,57			
	5	485412	8,0%	26.771,69	775256	13,1%	50.863,10	551790	9,4%	3.628,87			
	6	691470	11,5%	34.998,34	608966	10,3%	8.634,79	331377	5,6%	79.820,08			
	7	537112	8,9%	19.606,71	931703	15,8%	137.325,21	451240	7,7%	52.640,19			
	8	419825	7,0%	52.253,69	1110705	18,8%	475.995,76	231042	3,9%	211.288,65			
	9	904674	15,0%	152.450,87	772830	13,1%	57.411,62	100115	1,7%	403.735,15			
	10	971779	16,1%	202.223,27	734879	12,5%	27.873,20	119465	2,0%	387.642,67			
SUSPENSO	1	62920	18,3%	98.590,66	12089	2,3%	9.954,25	3209	0,6%	23.733,77	0,574	918.934,8310	
	2	18851	5,5%	33,05	41746	8,0%	4.933,17	19070	3,6%	4.197,49			
	3	198	0,1%	67.926,80	1534	0,3%	100.015,20	275168	51,7%	272.763,74			
	4	8893	2,6%	17.682,81	21590	4,2%	16.058,20	103577	19,5%	53.935,20			
	5	26615	7,7%	780,59	64301	12,4%	5.829,54	37073	7,0%	2.806,41			
	6	41934	12,2%	883,59	78105	15,0%	9.998,51	26969	5,1%	15.065,25			
	7	38915	11,3%	119,90	81381	15,7%	12.032,17	28906	5,4%	13.739,97			
	8	26747	7,8%	3.069,33	101675	19,6%	36.067,37	23444	4,4%	20.480,43			
	9	69923	20,3%	29.634,18	71789	13,8%	4.703,98	7712	1,4%	42.561,04			
	10	49501	14,4%	23.625,13	45445	8,7%	1.496,62	6916	1,3%	26.216,47			

A partir das Figuras 4.13 e 4.14 é possível observar que no grupo de beneficiários não-suspensos da macrorregião NÃO-SERRP, para a faixa de suicídio 1 no grupo não-suspensos, observa-se a maior contribuição para o valor do qui-quadrado nas faixas de PIB 1 (baixo) e 3 (alto), sugerindo as maiores contribuições para uma forte associação entre as faixas de renda mais baixa e mais alta com as menores taxas de suicídio nessa população. As faixas de suicídio 5 também têm contribuições significativas principalmente na segunda faixa de PIB (médio), indicando uma tendência de maiores taxas de suicídio em áreas de PIB mais intermediário.

No grupo de não-suspensos da macrorregião SERRP, de maneira similar, para a faixa de PIB 1 (baixo), a maior contribuição do qui-quadrado para a força da associação está na faixa de suicídio 1, sugerindo uma forte associação entre a faixa de renda mais baixa e as menores taxas de suicídio. As faixas de suicídio 3 também têm contribuições significativas para a força da associação, principalmente nas segunda e terceira faixa de PIB, o que pode indicar uma tendência de que maiores taxas de suicídio ocorrem em áreas de PIB intermediário a alto.

Na subpopulação de suspensos da macrorregião NAO-SERRP a primeira (baixo) e a terceira faixa de PIB (alto) tem o qui-quadrado que mais contribui para a força da associação na faixa de suicídio 1, indicando que as populações de renda mais baixa e mais alta convivem com menores taxas de suicídio. Já a segunda faixa de PIB (médio) tem a maior contribuição de qui-quadrado na faixa de suicídio 6, o que significa que esta faixa de suicídio está associada às rendas intermediárias de PIB. Já para a subpopulação de suspensos da macrorregião SERRP, a primeira faixa de PIB (baixo) tem a maior contribuição para a força da associação no qui-quadrado da faixa de suicídio 1, ou seja, novamente as populações de renda mais baixa são mais afetadas por menores taxas de suicídio. A faixa de suicídio 3 tem um valor de qui-quadrado significativo na segunda

(média) e na terceira (alta) faixa de PIB, o que indica que esta faixa de suicídio cresce associada às localidades mais ricas.

Em relação à força da associação encontrada na população de NAO-SERRP, o V de Cramer varia de 0,163 para os não-suspensos e 0,192 para os suspensos, o que sugere uma associação moderada entre as faixas de suicídio e as faixas de PIB. Para a população de SERRP, o V de Cramer é de 0,567 para os não-suspensos e 0,574 para os suspensos, indicando uma associação forte especialmente para os suspensos, o que sugere que a relação entre as faixas de suicídio e as faixas de PIB é mais pronunciada entre as famílias que residem em SERRP e tiveram o benefício suspenso.

Figura 4.15: Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita de NÃO-SERRP

VARIÁVEIS	FAIXA DE MORTE VIOLENTA	Crosstab NÃO-SERRP									MEDIDAS		
		PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	3330353	16,0%	780.962,52	1808445	8,7%	29.420,09	1014477	4,9%	513.641,89	0,184	4.203.191,5610	0,000
	2	2250834	10,8%	8.740,69	1660127	8,0%	94.951,86	2401593	11,6%	46.383,99			
	3	2032722	9,7%	1.893,19	2226495	10,7%	9.136,21	1996141	9,7%	2.725,92			
	4	2140343	10,3%	15.643,88	1810695	8,7%	11.100,42	1914210	9,3%	400,75			
	5	2265193	10,9%	3.716,96	2477455	11,9%	44.277,94	1750194	8,5%	74.318,71			
	6	2427628	11,6%	41.931,59	1831478	8,8%	39.620,32	2095178	10,2%	37,31			
	7	2204460	10,6%	7.886,86	2651544	12,8%	163.885,35	1341976	6,5%	245.841,20			
	8	1565242	7,5%	130.406,75	2002545	9,6%	2.855,56	2661334	12,9%	173.805,09			
	9	1417292	6,8%	56.608,65	2196090	10,6%	129.153,92	1551158	7,5%	14.770,58			
	10	1227079	5,9%	592.685,37	2123157	10,2%	35.893,09	3891715	18,9%	930.494,90			
SUSPENSO	1	27879	14,2%	15.926,42	21955	8,0%	592,82	16689	3,3%	9.252,86	0,221	95.703,2860	0,000
	2	16681	8,5%	661,43	17825	6,5%	3.981,97	67155	13,1%	3.879,15			
	3	18106	9,2%	236,76	17501	6,4%	1.149,42	45076	8,8%	235,27			
	4	22414	11,4%	2.546,80	19414	7,1%	403,64	38230	7,5%	274,58			
	5	28602	14,5%	4.790,47	37114	13,5%	4.105,58	29448	5,8%	8.098,01			
	6	20189	10,3%	254,79	22437	8,1%	312,93	47520	9,3%	9,48			
	7	22441	11,4%	419,89	38887	14,1%	4.828,30	36458	7,1%	4.060,85			
	8	13098	6,7%	5.662,82	32146	11,7%	228,47	79624	15,6%	3.340,40			
	9	15123	7,7%	939,69	32284	11,7%	978,10	49468	9,7%	15,48			
	10	12235	6,2%	10.412,40	35750	13,0%	874,26	101241	19,8%	7.230,25			

De acordo com a Figura 4.15, a análise do subgrupo de não-suspensos (NÃO-SERRP) mostra que a faixa 1 de Morte Violenta tem o valor de qui-quadrado mais elevado, contribuindo em especial para uma forte associação entre a menor taxa de mortes violentas e a primeira (baixo) e terceira (alto) faixas de PIB. Isso sugere que, na macrorregião NÃO-SERRP, as taxas mais baixas de morte violenta estão mais fortemente associadas com as faixas mais baixas e mais altas de PIB per capita municipal. A faixa 10 de Morte Violenta também apresenta um valor de qui-quadrado substancialmente alto para a terceira faixa de PIB, o que indica uma associação mais importante entre a maior taxa de mortes violentas e a faixa de PIB mais intermediária.

A análise do subgrupo de suspensos em NÃO-SERRP mostra que a faixa 1 de Morte Violenta repete o padrão dos não-suspensos, tendo as contribuições mais altas de qui-quadrado nas faixas iniciais e finais de PIB, com destaque para a faixa 1 de PIB, o que sugere que as taxas de morte violenta são mais baixas entre os suspensos que vivem em locais com PIB per capita municipal mais alto ou baixo. Os locais com PIB intermediário

tem a maior contribuição de qui-quadrado para a força da associação na faixa 7 de morte violenta.

O V de Cramer para os não-suspensos é de 0,184, o que indica uma associação forte, mas menor que para os suspensos. Para suspensos, o V de Cramer é de 0,221, indicando uma associação ainda mais forte entre as mortes violentas e as faixas de PIB para este grupo. Com base nesta análise, constata-se que há uma associação forte entre a faixa de PIB per capita municipal e as taxas de morte violenta, com essa relação sendo um pouco mais pronunciada entre as famílias que tiveram o benefício suspenso.

Figura 4.16: Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita de SERRP

Crosstab SERRP											MEDIDAS		
VARIÁVEIS	FAIXA DE MORTE VIOLENTA	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	865816	14,4%	107.110,35	765693	13,0%	47.567,57	170879	2,9%	302.376,14	0,465	7.719.725,6690	
	2	453075	7,5%	38.265,54	1080334	18,3%	402.002,60	254482	4,3%	191.027,03			
	3	593450	9,8%	335,60	732405	12,4%	31.850,56	469332	8,0%	25.673,68			
	4	430090	7,1%	48.850,07	1087325	18,4%	422.387,62	259385	4,4%	182.539,09			
	5	738026	12,2%	53.046,61	747019	12,7%	68.180,68	183671	3,1%	244.822,83			
	6	776051	12,9%	30.353,28	738747	12,5%	21.336,94	366848	6,2%	104.165,78			
	7	644257	10,7%	701.676,30	43729	0,7%	152.458,07	11237	0,2%	208.923,80			
	8	166567	2,8%	631.614,16	189753	3,2%	574.972,25	2405820	40,9%	2.447.236,63			
	9	370161	6,1%	15.403,87	134186	2,3%	88.077,18	387618	6,6%	29.453,81			
	10	993377	16,5%	4.243,61	383046	6,5%	305.711,01	1372306	23,3%	238.062,99			
SUSPENSO	1	46124	13,4%	11.406,39	55916	10,8%	4.216,27	12217	2,3%	22.533,59	0,506	714.588,3080	
	2	34232	9,9%	248,48	84397	16,2%	28.831,27	8781	1,7%	32.577,85			
	3	39591	11,5%	1.499,09	62466	12,0%	3.591,86	30067	5,7%	8.169,57			
	4	28453	8,3%	1.790,56	98505	19,0%	34.145,13	21141	4,0%	22.073,12			
	5	45159	13,1%	4.232,42	75561	14,5%	12.793,57	14196	2,7%	26.939,85			
	6	49439	14,4%	4.824,96	69733	13,4%	4.172,21	27620	5,2%	14.335,35			
	7	30765	8,9%	54.867,01	3908	0,8%	6.603,71	977	0,2%	11.701,30			
	8	10433	3,0%	55.736,28	16301	3,1%	83.107,60	277854	52,2%	225.510,02			
	9	20532	6,0%	353,23	11714	2,3%	8.789,71	40745	7,7%	6.011,23			
	10	39769	11,5%	455,21	41154	7,9%	9.821,24	98446	18,5%	13.250,23			

Na análise das taxas de morte violenta e PIB per capita municipal na macrorregião SERRP, a Figura 4.16 mostra que para os não-suspensos a faixa de Morte Violenta 7 possui o maior valor de qui-quadrado na faixa 1 do PIB, indicando que neste ponto está a maior contribuição para a forte associação entre taxas de morte violenta (neste caso, alta) e os níveis de PIB per capita (neste caso, baixo). Isso sugere que, na macrorregião SERRP, as taxas mais altas de morte violenta são observadas em municípios com menores PIBs per capita. A faixa de Morte Violenta 8 mostra um valor extremamente alto de qui-quadrado na faixa 3 do PIB (alto) e o terceiro maior valor na faixa 2 de PIB (médio), destacando uma relação forte entre altas taxas de mortes violentas e as faixas mais altas de PIB per capita.

Para os suspensos, o padrão se repete exatamente como para os não-suspensos, onde a faixa de Morte Violenta 7 apresenta um qui-quadrado alto na faixa 1 do PIB (baixo) e a faixa de Morte Violenta 8 traz um valor muito elevado de qui-quadrado na faixa 3 do PIB (alto) e um terceiro valor significativo de qui-quadrado na faixa 2 de PIB. Portanto, são as faixas 7 e 8 de morte violenta que contribuem mais para a força da associação.

O valor V de Cramer é de 0,465 para não-suspensos e 0,506 para suspensos. Ambos indicam uma associação muito forte, o que sugere que a relação entre a faixa de morte violenta e as faixas de PIB é mais expressiva na macrorregião SERRP, especialmente em zonas mais ricas. O valor ainda mais alto para os suspensos indica uma associação mais forte nesse subgrupo. Esta análise mostra que a relação entre o nível econômico medido pelo PIB per capita municipal e as altas taxas de morte violenta na macrorregião SERRP é mais acentuada do que em NÃO-SERRP, com a força da associação sendo maior para aqueles com o benefício suspenso.

Figura 4.17: Homicídio de Mulheres e Média Categorizada de PIB Per Capita de NÃO-SERRP

Crosstab NÃO-SERRP											MEDIDAS		
VARIÁVEIS	FAIXA DE HOMICÍDIO DE MULHERES	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	7393071	35,4%	2.584.735,57	3477332	16,7%	98.111,15	1448031	7,0%	1.696.892,23	0,268	8.972.860,2860	
	2	2784720	13,3%	198.324,15	2055693	9,9%	2.368,32	1529656	7,4%	159.271,90			
	3	2268559	10,9%	16.919,65	2023385	9,7%	1.217,30	1919266	9,3%	9.178,98			
	4	1992087	9,5%	5.326,39	2409167	11,6%	48.595,91	1860318	9,0%	21.886,49			
	5	1596681	7,7%	107.838,86	2240379	10,8%	15.469,42	2338690	11,3%	42.201,95			
	6	1499780	7,2%	140.246,54	1904895	9,2%	7.324,24	2666070	12,9%	214.027,41			
	7	1186475	5,7%	425.199,67	1703723	8,2%	86.387,60	3498801	17,0%	904.466,96			
	8	998444	4,8%	566.088,08	1930479	9,3%	10.404,41	3293889	16,0%	738.282,93			
	9	1141329	5,5%	432.786,90	3042978	14,6%	439.292,95	2063255	10,0%	14,33			
SUSPENSO	1	69680	35,4%	37.526,45	57987	21,1%	2.297,93	42056	8,2%	24.151,82	0,256	129.070,9670	
	2	26292	13,4%	1.007,74	29362	10,7%	26,39	52372	10,3%	253,75			
	3	15247	7,7%	2.016,43	31228	11,3%	11,71	62885	12,3%	642,93			
	4	18699	9,5%	737,44	29308	10,6%	2.871,92	28607	5,6%	3.157,55			
	5	19865	10,1%	50,06	24496	8,9%	142,08	50020	9,8%	19,00			
	6	12731	6,5%	4.466,20	24554	8,9%	1.708,32	76758	15,0%	5.157,35			
	7	11522	5,9%	7.125,28	20308	7,4%	5.987,86	92166	18,0%	11.922,18			
	8	10898	5,5%	3.810,04	21228	7,7%	1.356,60	65402	12,8%	4.269,82			
	9	11834	6,0%	2.043,99	36842	13,4%	5.590,35	40643	8,0%	719,78			

Na Figura 4.17, analisando homicídios de mulheres em relação à média categorizada do PIB per capita municipal na macrorregião NÃO-SERRP, observa-se que para a subpopulação de não-suspensos a faixa 1 de Homicídio de Mulheres possui o maior valor de qui-quadrado nas faixas 1 (baixo) e 3 (alto) de PIB, contribuindo em maior intensidade com a forte associação entre baixas taxas de homicídio de mulheres e os níveis de menor e maior PIB per capita. Especificamente, há uma associação significativa nas faixas de menor PIB, indicando que essa taxa de homicídio pode ser elevada tanto em áreas mais pobres quanto mais ricas. Neste subgrupo, as famílias com PIB intermediário possuem maior valor de qui-quadrado na maior faixa de homicídio de mulheres entre todas, evidenciando o ponto onde há maior contribuição para o grau de associação forte.

Para a subpopulação de suspensos de NÃO-SERRP a análise para beneficiários com suspensão ainda apresenta um valor de qui-quadrado na mesma composição dos não-suspensos. Há uma contribuição de qui-quadrado significativa na Faixa 1 de Homicídio de Mulheres relacionada às faixas mais baixas e mais altas de PIB. A faixa de PIB intermediária tem maior qui-quadrado na faixa 7 de Homicídio de Mulheres, indicando que famílias de rendas médias se relacionam de forma mais significativa com taxas mais alta de violência de mulheres.

O V de Cramer para os não-suspensos é de 0,268 e para os suspensos é de 0,256, ambos indicando uma associação forte entre a faixa de homicídio de mulheres e o PIB per capita municipal. Para este tipo de violência, a associação é ainda mais relevante nas famílias sem suspensão.

Figura 4.18: Homicídio de Mulheres e Média Categorizada de PIB Per Capita de SERRP

Crosstab SERRP											MEDIDAS		
VARIÁVEIS	FAIXA DE HOMICÍDIO DE MULHERES	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	903817	15,0%	155.556,73	642480	10,9%	5.466,93	222076	3,8%	224.154,45	0,505	9.083.619,8290	
	2	27203	0,5%	22.769,23	162012	2,7%	147.161,57	5612	0,1%	53.588,42			
	3	174316	2,9%	764.910,13	293702	5,0%	556.187,50	2738608	46,6%	2.665.742,39			
	4	485454	8,0%	28.038,68	1190002	20,2%	569.055,52	147054	2,5%	343.538,96			
	5	438053	7,3%	41.783,26	951278	16,1%	232.424,66	370724	6,3%	76.155,48			
	6	260767	4,3%	195.512,47	201056	3,4%	257.884,37	1324029	22,5%	914.813,39			
	7	986154	16,4%	252.307,93	649764	11,0%	7.160,73	129902	2,2%	352.132,13			
	8	1156890	19,2%	440.788,39	406647	6,9%	71.437,21	297265	5,1%	163.657,37			
	9	663231	11,0%	4.606,69	657126	11,1%	6.014,04	482157	8,2%	21.437,33			
	10	934985	15,5%	153.281,14	748170	12,7%	30.279,77	164151	2,8%	325.772,96			
SUSPENSO	1	56214	16,3%	16.230,37	54348	10,5%	404,37	23395	4,4%	14.978,66	0,539	810.874,6570	
	2	3722	1,1%	893,38	20210	3,9%	13.484,64	572	0,1%	8.228,70			
	3	9633	2,8%	68.942,87	30286	5,8%	77.857,79	313263	58,9%	237.212,64			
	4	30597	8,9%	4.274,78	138402	26,6%	76.328,13	10822	2,0%	48.588,94			
	5	40060	11,6%	44,46	85842	16,5%	12.838,70	31136	5,9%	13.770,15			
	6	17723	5,1%	7.346,22	22086	4,3%	15.871,39	95488	17,9%	37.432,50			
	7	37324	10,8%	6.341,11	55132	10,6%	8.419,01	8002	1,5%	23.949,89			
	8	49424	14,3%	31.662,00	26553	5,1%	1.694,39	15806	3,0%	10.506,46			
	9	41652	12,1%	9.498,18	40569	7,8%	51,66	22958	4,3%	7.314,60			
	10	58148	16,9%	31.257,83	46227	8,9%	275,45	10602	2,0%	25.175,37			

Na Figura 4.18 a análise revela que para os não-suspensos (SERRP) a categoria com o maior valor de qui-quadrado na Faixa 1 e 3 do PIB é a faixa 3 de homicídio de mulheres. Esta faixa indica uma forte relação entre um maior homicídio de mulheres e os níveis econômicos mais baixos e mais altos das localidades. A faixa 2 de PIB apresenta a maior contribuição para o valor de qui-quadrado na faixa 4 de homicídio de mulheres. Isso sugere que há uma associação significativa entre a incidência de homicídios de mulheres e regiões com maiores rendas per capita. Para os suspensos (SERRP), semelhantes aos não-suspensos, a faixa 3 de homicídios de mulheres mostra um valor alto de qui-quadrado em todas as faixas de PIB.

O V de Cramer para os suspensos é de 0,539, que é um pouco maior do que para os não-suspensos (0,505), indicando uma associação muito forte entre as faixas de homicídio de mulheres e as faixas de PIB per capita. Os valores de V de Cramer acima de 0,5 em ambos os grupos (suspensos e não-suspensos) sugerem uma associação consistentemente forte, o que implica que as diferenças no PIB per capita tem uma relação relevante com a incidência de homicídios de mulheres na macrorregião SERRP.

Figura 4.19: Homicídio de Homens e Média Categorizada de PIB Per Capita de NÃO-SERRP

Crosstab NÃO-SERRP											MEDIDAS		
VARIÁVEIS	FAIXA DE HOMICÍDIO DE HOMENS	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	3488271	16,7%	950.388,44	1536955	7,4%	139.247,68	1188327	5,8%	367.134,12	0,313	12.238.162,7890	
	2	3665599	17,6%	1.182.627,82	1468164	7,1%	182.570,40	1112006	5,4%	442.012,26			
	3	5745047	27,5%	593.031,33	3466768	16,7%	114.750,24	3241211	15,7%	188.765,07			
	4	2696584	12,9%	180.047,02	2572491	12,4%	118.351,64	951416	4,6%	596.374,58			
	5	2247624	10,8%	19.960,40	2627882	12,6%	170.478,18	1230158	6,0%	309.916,29			
	6	1105314	5,3%	494.523,93	2504634	12,0%	67.946,08	2754189	13,4%	198.578,33			
	7	1074938	5,2%	373.392,86	2017066	9,7%	5.399,01	2645188	12,8%	292.541,76			
	8	837769	4,0%	2.817.511,46	4594071	22,1%	17.940,08	7495481	36,4%	2.414.673,78			
SUSPENSO	1	30467	15,5%	4.508,02	31516	11,4%	203,31	41861	8,2%	2.718,00	0,258	130.786,0140	
	2	36015	18,3%	20.409,35	19397	7,0%	930,95	30769	6,0%	4.390,47			
	3	51924	26,4%	3.919,46	47315	17,2%	1.138,27	98010	19,2%	198,41			
	4	26984	13,7%	5.947,81	33173	12,0%	3.777,96	24494	4,8%	8.645,54			
	5	22170	11,3%	967,73	37561	13,6%	6.086,98	30175	5,9%	5.864,13			
	6	10699	5,4%	5.086,66	30022	10,9%	11,25	64416	12,6%	1.747,16			
	7	10004	5,1%	6.686,22	26661	9,7%	617,07	74152	14,5%	4.758,29			
	8	8505	4,3%	25.827,47	49668	18,0%	1.060,00	147032	28,8%	15.285,50			

No cruzamento entre as faixas de homicídio de homens e a média categorizada do PIB per capita municipal em NÃO-SERRP, observamos que para as famílias não-suspensas, nas faixas 1 e 3 de PIB se localizam as categorias com as maiores contribuições para o valor do qui-quadrado relacionadas à faixa 8 de homicídio, sugerindo que são os locais com as maiores taxas de homicídio de homens que mais contribuem para a associação forte com os grupos residentes nos níveis de PIB baixo e alto. A faixa intermediária de PIB dos não-suspensos de NÃO-SERRP apresenta maior qui-quadrado na faixa 2 de homicídio masculino, ou seja, a forte associação do PIB intermediário tem maior contribuição de faixas menores de violência .

Para as famílias suspensas de SERRP, todos os níveis de PIB (baixo, médio e alto) estão fortemente associados à violência, sendo encontradas as maiores contribuições de qui-quadrado em categorias de altas taxas de violência, como as faixas 5 e 8 de homicídio de homens. Essa relação sugere que as famílias suspensas de SERRP sofrem mais influência de locais mais violentos.

O V de Cramer para os suspensos é 0,258, indicando uma força de associação ligeiramente menor que os não-suspensos, mais ainda forte. O V de Cramer de 0,313 dos não-suspensos indica uma associação de força alta entre a faixa de homicídio de homens e as faixas de PIB per capita municipal.

Figura 4.20: Homicídio de Homens e Média Categorizada de PIB Per Capita de SERRP.

Crosstab SERRP											MEDIDAS		
VARIÁVEIS	FAIXA DE HOMICÍDIO DE HOMENS	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	1	701071	11,6%	19.486,15	731097	12,4%	38.853,35	321063	5,5%	114.793,41	0,341	4.152.398,3750	
	2	397052	6,6%	735,86	550945	9,3%	85.827,29	175445	3,0%	103.006,21			
	3	675314	11,2%	352.138,06	900970	15,3%	144.185,87	2470127	42,0%	962.911,92			
	4	614933	10,2%	0,44	841121	14,3%	95.640,43	358875	6,1%	96.391,83			
	5	984465	16,3%	43.098,41	997093	16,9%	30.455,57	1601900	27,2%	148.256,81			
	6	541304	9,0%	6.743,38	871696	14,8%	131.819,78	374671	6,4%	78.710,12			
	7	1317342	21,8%	782.384,31	123687	2,1%	385.135,13	392038	6,7%	75.075,44			
	8	799389	13,3%	43.205,66	885628	15,0%	113.413,21	187459	3,2%	300.129,71			
SUSPENSO	1	52987	15,4%	3.762,08	78908	15,2%	5.071,06	32750	6,2%	14.335,90	0,362	364.954,8770	
	2	25781	7,5%	56,28	64447	12,4%	13.781,47	19256	3,6%	12.096,24			
	3	46600	13,5%	31.342,57	88722	17,1%	29.177,50	284460	53,5%	96.889,97			
	4	39781	11,5%	0,01	94274	18,1%	19.606,62	27100	5,1%	19.169,79			
	5	57999	16,8%	158,62	74274	14,3%	3.479,96	115407	21,7%	4.683,32			
	6	30922	9,0%	751,31	52906	10,2%	4.225,78	23423	4,4%	7.447,82			
	7	46688	13,6%	44.036,03	8793	1,7%	12.819,69	18687	3,5%	3.244,50			
	8	43739	12,7%	9.373,16	57331	11,0%	5.861,68	10961	2,1%	23.583,54			

Com base nos dados cruzados fornecidos para SERRP por meio da Figura 4.20, verifica-se que para o público não-suspensão, as maiores contribuições para o valor do qui-quadrado estão na faixa 7 de homicídio de homens para os grupos de PIB baixo e médio e na faixa 8 para o grupo de PIB alto, indicando uma forte relação entre PIB e altas taxas de homicídio masculino.

Para o público suspensão de SERRP, para o grupo de PIB baixo, o maior valor de qui-quadrado está na faixa 7 de homicídio de homens, semelhantes à categoria não suspensão. Já para os grupos de PIB médio e alto a maior contribuição do qui-quadrado está na faixa 3 de violência de homens, indicando que em SERRP, as faixas mais ricas de PIB são menos afetadas por homicídio de homens.

O valor do V de Cramer é 0,362 de SERRP, indica associação forte entre as taxas de homicídios e as categorias do PIB para o grupo suspensão, ligeiramente superior à NÃO-SERRP com V de Cramer de 0,341.

Há uma diferença marcante entre as taxas de homicídio de homens e os níveis econômicos na região SERRP para suspensos e não-suspensos. Para o público suspensão, isto é realçado pelos valores mais significativos do qui-quadrado nas extremidades das categorias do PIB, particularmente nos escalões mais elevados do PIB ao tempo que são categorias mais baixas de violência. Os valores do V de Cramer sugerem uma relação forte, indicando que a variação nas suspensões os beneficiários do Programa Bolsa Família pode ser explicadas, até certo ponto, pelas diferenças nas taxas de violência e PIB per capita da região de residência.

4.2.4 Associação entre o Perfil Familiar e a Exclusão Escolar

Nesta seção são apresentadas as relações entre as características pessoais dos membros familiares e da infraestrutura domiciliar com a suspensão de benefício. Para melhor entendimento do perfil familiar, as variáveis foram divididas em dois grupos: um grupo

com as características de infraestrutura do domicílio das famílias dos beneficiários e o outro com as características pessoais dos membros da família do beneficiário. Combinada com a análise do qui-quadrado, estão as medidas simétricas (ou tamanho do efeito), representadas pelo V de Cramer que são cruciais para constatar a relevância prática dos resultados estatísticos destas relações.

São apresentados apenas as associações com tamanho de efeito (ES) grande ou moderado, mensurado pelo V de Cramer. As associações com tamanho de efeito pequeno não serão abordadas.

Os Grandes Efeitos do Perfil da Família sobre a Suspensão

Figura 4.21: Escoamento Sanitário e Macrorregião

VARIÁVEIS	ESCOAMENTO SANITÁRIO DO DOMICÍLIO	Crosstab						MEDIDAS		
		NÃO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	Rede coletora de esgoto ou pluvial	16411222	31,7%	1.263.220,56	12220153	73,1%	3.909.649,73	0,369	9.340.821,32	0,000
	Fossa séptica	9862058	19,1%	179.835,71	1539383	9,2%	556.588,97			
	Fossa rudimentar	22577878	43,6%	795.457,36	2189412	13,1%	2.461.929,26			
	Vala a céu aberto	1807940	3,5%	14.579,01	378772	2,3%	45.121,83			
	Direto para um rio, lago ou mar	477791	0,9%	18.039,73	289725	1,7%	55.832,71			
	Outra forma	591621	1,1%	9.906,37	96205	0,6%	30.660,08			
SUSPENSO	Rede coletora de esgoto ou pluvial	380641	43,9%	60.203,05	1065405	78,8%	38.616,83	0,385	329.045,99	0,000
	Fossa séptica	136334	15,7%	27.161,44	87744	6,5%	17.422,52			
	Fossa rudimentar	309693	35,7%	110.845,64	129750	9,6%	71.101,17			
	Vala a céu aberto	21636	2,5%	217,79	28447	2,1%	139,70			
	Direto para um rio, lago ou mar	11011	1,3%	1.843,32	31292	2,3%	1.182,38			
	Outra forma	8290	1,0%	190,18	9945	0,7%	121,99			

A Figura 4.21 apresenta uma análise bivariada que explora a relação entre o tipo de escoamento sanitário de água no domicílio e a macrorregião, comparando famílias com suspensão de benefícios e sem suspensão, usando o teste de qui-quadrado de independência.

Da análise bivariada apresentada na Figura 4.21, conclui-se que a maioria das famílias não-suspensas na macrorregião SERRP possui rede coletora de esgoto ou pluvial (73,1%), enquanto na NÃO-SERRP esse número é significativamente menor (31,7%). Em NÃO-SERRP há uma prevalência maior de fossa rudimentar (43,6%) comparada com a SERRP (13,1%). O V de Cramer de 0,369 indica uma associação de magnitude grande, sendo puxado para cima principalmente por domicílios tratados por rede coletora de esgoto ou pluvial.

Para a subpopulação de suspensos, a diferença é ainda mais acentuada, com 78,8% na SERRP tendo rede coletora de esgoto ou pluvial, em comparação com 43,9% na NÃO-SERRP. A proporção de famílias suspensas com fossa rudimentar é alta na NÃO-SERRP

(35,7%) e mais baixa na SERRP (9,6%). Entretanto, a diferença substancial na infraestrutura sanitária entre as macrorregiões é refletida no alto valor de qui-quadrado presente no tipo fossa rudimentar, o tipo que mais contribui para uma associação forte entre a macrorregião e o tipo de escoamento sanitário. O V de Cramer de 0,385 para suspensos indica uma associação de magnitude grande entre o tipo de escoamento sanitário e a macrorregião, semelhante ao observado para os não-suspensos. Entretanto, essa magnitude é puxada para cima por domicílios com escoamento sanitário do tipo fossa rudimentar.

As diferenças marcantes no tipo de escoamento sanitário entre as macrorregiões refletem desigualdades na infraestrutura básica que impactam a qualidade de vida e saúde das famílias beneficiárias. A associação entre a infraestrutura de escoamento sanitário e a macrorregião é estatisticamente significativa, com forte associação tanto para famílias suspensas quanto para as não-suspensas, como evidenciado pelos altos valores do V de Cramer. O fato de uma grande proporção de famílias suspensas em SERRP ter acesso à rede coletora de esgoto ou pluvial indica que, apesar da melhor infraestrutura, é o escoamento do tipo fossa rudimentar que mais contribui para a suspensão dos benefícios nesta região.

Figura 4.22: Raça-Cor e Macrorregião

VARIÁVEIS		Crosstab						MEDIDAS		
		NÃO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
RAÇA-COR		N	%	X2	N	%	X2			
NÃO-SUSPENSO	Branca	20143166	17,6%	2.394.071,07	16841575	48,6%	7.904.433,69	0,314	14.680.779,45	0,000
	Preta	6970162	6,1%	34.222,75	2770116	8,0%	112.992,25			
	Amarela	637088	0,6%	2.344,01	144115	0,4%	7.739,16			
	Parda	85286675	74,5%	937.831,91	14774144	42,6%	3.096.411,89			
	Indígena	1370782	1,2%	44.339,22	121562	0,4%	146.393,49			
SUSPENSO	Branca	476296	19,3%	160.430,97	1411651	46,4%	129.799,10	0,294	474.790,05	0,000
	Preta	197758	8,0%	244,02	260236	8,5%	197,43			
	Amarela	12166	0,5%	96,92	12715	0,4%	78,42			
	Parda	1754944	71,3%	96.224,36	1351553	44,4%	77.851,77			
	Indígena	21533	0,9%	5.454,23	7725	0,3%	4.412,83			

A Tabela 4.22 apresenta uma análise bivariada entre a variável raça-cor dos membros familiares dos beneficiários do programa e a macrorregião, estratificada por beneficiários suspensos e não-suspensos, com a aplicação do teste de independência.

Ao analisar o público de beneficiários não-suspensos, nota-se que há uma grande discrepância entre as macrorregiões em relação à cor “Branca”, com 48,6% em SERRP quando comparado com 17,6% na NÃO-SERRP, também refletido pelo alto valor de qui-quadrado que mais contribui para a força da associação entre raça-cor e macrorregião. A categoria “parda” constitui a maioria em ambas as macrorregiões, com 74,5% em NÃO-SERRP e 42,6% em SERRP. Apesar de ser a categoria predominante, a proporção é significativamente menor na SERRP. O V de Cramer de 0,314 indica uma associação de magnitude forte entre raça-cor e macrorregião na subpopulação de não-suspensos.

Analisando o subpopulação de beneficiários suspensos é possível observar que a categoria “branca” representa 19,3% em NÃO-SERRP e 46,4% em SERRP, mostrando uma proporção semelhante à observada entre os não-suspensos. A categoria “parda” é a mais comum entre os suspensos, sendo 71,3% em NÃO-SERRP e 44,4% em SERRP, refletindo uma tendência similar à observada entre os não-suspensos. O V de Cramer de 0,294 para suspensos é ligeiramente menor do que para não-suspensos, ambos indicando uma associação de magnitude forte influenciada principalmente pela categoria de brancos.

A análise sugere que a cor “branca” é maioria em SERRP em comparação com NÃO-SERRP, tanto para beneficiários suspensos quanto para não-suspensos. A cor “parda” é a mais comum em ambas as macrorregiões, mas a diferença na proporção entre as macrorregiões é mais pronunciada entre os não-suspensos. A grandeza do impacto da raça-cor é constatada tanto sobre a suspensão quanto não-suspensão, sendo estatisticamente significativa, com altos valores de qui-quadrado e V de Cramer, refletindo desigualdades raciais estruturais. O V de Cramer grande sugere que a relação entre raça-cor branca e a macrorregião é um dos fatores mais importantes para determinar a suspensão dos benefícios.

Os Efeitos Moderados do Perfil da Família sobre a Suspensão

Figura 4.23: Infraestrutura Domiciliar: Iluminação do Domicílio e Macrorregião.

VARIÁVEIS	ILUMINAÇÃO DO DOMICÍLIO	Crosstab						MEDIDAS		
		NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	Elétrica com medidor próprio	49847013	84,2%	49.010,01	12353065	72,6%	170.431,29	0,206	3.235.793,45	0,000
	Elétrica com medidor comunitário	2191964	3,7%	497.946,32	2333740	13,7%	1.731.597,93			
	Elétrica sem medidor	3774275	6,4%	86.986,30	1881224	11,1%	302.493,03			
	Óleo, querosene ou	1075794	1,8%	62.274,73	13784	0,1%	216.559,08			
	Vela	618450	1,0%	24.038,84	35567	0,2%	83.594,55			
	Outra forma	1664061	2,8%	2.425,78	398267	2,3%	8.435,59			
SUSPENSO	Elétrica com medidor próprio	771165	81,3%	7.025,89	944902	68,8%	4.851,88	0,203	95.326,10	0,000
	Elétrica com medidor comunitário	50520	5,3%	29.894,59	211681	15,4%	20.644,36			
	Elétrica sem medidor	75911	8,0%	7.298,76	177171	12,9%	5.040,32			
	Óleo, querosene ou	11466	1,2%	8.362,48	774	0,1%	5.774,90			
	Vela	7558	0,8%	3.281,39	2125	0,2%	2.266,04			
	Outra forma	31469	3,3%	523,78	36251	2,6%	361,71			

A Figura 4.23 exibe os resultados de uma análise bivariada sobre a iluminação do domicílio e a macrorregião, categorizando as famílias em suspensas e não-suspensas. O teste de independência foi utilizado para avaliar a existência de associação entre as variáveis e o V de Cramer para medir o tamanho do efeito dessa associação.

A análise sobre os beneficiários não-suspensos mostra que em ambas as macrorregiões, a maior contribuição de qui-quadrado para a força da associação está na categoria de

medidor comunitário. Uma proporção maior de SERRP (13,7%) possui eletricidade com medidor comunitário em relação à NÃO-SERRP (3,7%). A força da associação entre iluminação do domicílio e macrorregião é moderada, tendo um V de Cramer de 0,206, valor significativo do ponto de vista da estatística. A diferença entre as macrorregiões é acentuada, com SERRP mostrando uma maior dependência de medidores comunitários entre os não-suspensos. Isso reflete uma disparidade no fornecimento de eletricidade nas diferentes macrorregiões.

A análise sobre os beneficiários suspensos é semelhante aos não-suspensos com a maior contribuição de qui-quadrado para a força da associação está na categoria de medidor comunitário. Em SERRP, 5,3% possui medidor comunitário. Já em NÃO-SERRP, a proporção de suspensos com medidor comunitário é de 15,4% . O V de Cramer é de 0,203 para os suspensos, confirmando uma associação de grau moderado. Entre os suspensos, a diferença entre as macrorregiões é semelhante à observada entre os não-suspensos.

A eletricidade com medidor próprio é o tipo de iluminação mais comum em ambas as macrorregiões, tanto para suspensos quanto para não-suspensos. Há uma discrepância significativa entre as macrorregiões em termos de acesso à eletricidade com medidor próprio, com a SERRP tendo uma proporção menor nessa categoria, especialmente entre os suspensos. O acesso à eletricidade com medidor comunitário é muito mais comum na NÃO-SERRP, indicando diferentes modelos de infraestrutura de eletricidade nas macrorregiões. O V de Cramer sugere que há associação moderada entre o tipo de iluminação e a macrorregião afetando as suspensões, sendo a maior influência encontrada nas diferenças entre as proporções do medidor do tipo comunitário.

Figura 4.24: Material Predominante nas Paredes do Domicílio e Macrorregião

VARIÁVEIS	MATERIAL PREDOMINANTE NAS PAREDES DO DOMICÍLIO	Crosstab						MEDIDAS		
		NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	Alvenaria/tijolo com revestimento	38453700	65,0%	12.078,04	11943211	70,2%	42.001,12	0,135	1.388.800,85	0,000
	Alvenaria/tijolo sem revestimento	10161216	17,2%	3.268,27	3158759	18,6%	11.365,36			
	Madeira aparelhada	4172317	7,1%	6.032,61	1408004	8,3%	20.978,28			
	Taipa revestida	1449834	2,5%	87.521,20	11164	0,1%	304.353,17			
	Taipa não revestida	1854201	3,1%	113.059,37	11989	0,1%	393.161,62			
	Madeira aproveitada	1138965	1,9%	2.417,29	396660	2,3%	8.406,08			
	Palha	152221	0,3%	9.402,38	741	0,0%	32.696,60			
	Outro Material	1789092	3,0%	76.395,54	85114	0,5%	265.663,92			
SUSPENSO	Alvenaria/tijolo com revestimento	635837	67,1%	525,58	966141	70,4%	362,95	0,166	63.636,45	0,000
	Alvenaria/tijolo sem revestimento	173167	18,3%	192,61	265135	19,3%	133,01			
	Madeira aparelhada	63949	6,7%	68,83	97824	7,1%	47,53			
	Taipa revestida	15424	1,6%	11.575,88	856	0,1%	7.993,97			
	Taipa não revestida	19629	2,1%	15.177,19	844	0,1%	10.480,93			
	Madeira aproveitada	17851	1,9%	706,21	35449	2,6%	487,69			
	Palha	1524	0,2%	1.234,01	36	0,0%	852,17			
	Outro Material	20707	2,2%	8.161,68	6619	0,5%	5.636,22			

A tabela 4.24 apresenta a análise bivariada entre a material predominante nas paredes do domicílio e a macrorregião. Para o grupo de beneficiários não-suspensos, o material do tipo “alvenaria/tijolo com revestimento” é o material mais comum tanto na NÃO-SERRP (65%) quanto na SERRP (70,2%). A diferença entre as macrorregiões indica que uma proporção maior de domicílios na SERRP possui um acabamento melhor nas paredes. O material “taipa não revestida” mostra as maiores diferenças percentuais entre as macrorregiões e a maior contribuição do qui-quadrado para a força da associação. Na NÃO-SERRP, 3,1% das casas usa esse material em comparação com apenas 0,1% na SERRP. Ao comparar NÃO-SERRP com SERRP, especificamente no material “Taipa não revestida”, percebe-se que a contribuição para o qui-quadrado é bastante alta, especialmente considerando a discrepância entre as macrorregiões. Um qui-quadrado de 113.059,37 de NÃO-SERRP e de 393.161,62 de SERRP, considerando uma porcentagem de casos muito menor na SERRP, indica uma diferença significativa nas condições de habitação entre as macrorregiões. A utilização de taipa não revestida, sendo maior na NÃO-SERRP, reflete condições habitacionais mais precárias quando comparada com a SERRP. O V de Cramer de 0,135 do grupo não-suspense indica uma associação de magnitude moderada.

Analisando o grupo de beneficiários suspensos, vê-se que o material do tipo “Alvenaria/tijolo com revestimento” continua sendo o material mais comum, com 67,1% em NÃO-SERRP e 70,4% em SERRP. Comparando o resultado da “taipa não revestida” em NÃO-SERRP vs. SERRP, é possível observar que embora a porcentagem para suspensos seja menor, a contribuição para o qui-quadrado total deste grupo é alta, com 15.177,19 no NÃO-SERRP e 10.480,93 no SERRP. Isso sugere que a qualidade do material de construção das paredes é uma variável importante na diferenciação entre as macrorregiões também entre o público suspenso. O V de Cramer é maior para os suspensos (0,166), sugerindo que a diferença entre as macrorregiões para os suspensos é mais acentuada em comparação com os não-suspensos, ao tempo que indica uma associação moderada entre a variável de material das paredes e a macrorregião para esse subconjunto da população.

A análise comparativa mostra que para ambas as populações, suspensas e não-suspensas, as casas de alvenaria/tijolo com revestimento são predominantes. No entanto, há uma proporção significativamente maior de casas sem revestimento e de materiais menos duráveis (taipa e madeira) na NÃO-SERRP em comparação com a SERRP. As altas contribuições na categoria “taipa não revestida” para o valor do qui-quadrado dos grupos destacam a disparidade significativa nas condições de moradia entre as macrorregiões. A análise indica que a condição material das habitações, que é um reflexo da situação socioeconômica, difere substancialmente entre NÃO-SERRP e SERRP, tanto para o público suspenso quanto para o não-suspensos. A maior contribuição para o qui-quadrado da taipa não revestida, em ambas as regiões, significa que este tipo de material apresenta

as maiores diferenças entre os casos observados e os esperados e são os mais importantes para a significância da associação.

Figura 4.25: Abastecimento de Água do Domicílio e Macrorregião.

VARIÁVEIS	ABASTECIMENTO DE ÁGUA DO DOMICÍLIO	Crosstab						MEDIDAS		
		NÃO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	Rede geral de distribuição	38239569	64,6%	175.004,01	14441721	84,9%	608.572,97	0,191	2.789.182,20	0,000
	Poço ou nascente	13275324	22,4%	146.587,52	2113266	12,4%	509.755,19			
	Cisterna	3167915	5,4%	177.863,49	52232	0,3%	618.516,76			
	Outra forma	4488749	7,6%	123.480,67	408423	2,4%	429.401,58			
SUSPENSO	Rede geral de distribuição	694565	73,3%	9.304,93	1214308	88,4%	6.425,72	0,213	105.363,55	0,000
	Poço ou nascente	152616	16,1%	16.760,25	116014	8,5%	11.574,14			
	Cisterna	33118	3,5%	23.379,15	2688	0,2%	16.144,97			
	Outra forma	67789	7,2%	12.879,91	39894	2,9%	8.894,50			

Analisando a Figura 4.25 entre a característica de infraestrutura domiciliar "abastecimento de água do domicílio" e "macrorregião", estratificada por beneficiários suspensos e não-suspensos, observamos os seguintes pontos: para o público de não-suspensos: comparando (NÃO-SERRP vs. SERRP) vemos que “rede geral de distribuição” tem a maior diferença nos valores do qui-quadrado entre as macrorregiões, com 175.004,01 para NÃO-SERRP e 608.572,97 para SERRP. A porcentagem de domicílios conectados à rede geral de distribuição é significativamente maior na SERRP (84,9%) do que na NÃO-SERRP (64,6%). Para Cisterna, na comparação NÃO-SERRP vs. SERRP, observamos uma contribuição significativa para o qui-quadrado dos não-suspensos nesta categoria, com 177.863,49 para NÃO-SERRP e uma muito alta de 618.516,76 para SERRP, apesar da baixa porcentagem de casos em SERRP. Isso sugere que o uso de cisternas como principal meio de abastecimento de água é mais comum na NÃO-SERRP e, embora menos prevalente, a presença de cisternas na SERRP é um fator significativo de diferenciação. O V de Cramer de 0,191 indica uma associação moderada entre a macrorregião e o tipo de abastecimento de água, sugerindo que esta característica influencia a relação.

Para o público de suspensos, o abastecimento de água por “rede geral de distribuição” na comparação (NÃO-SERRP com SERRP), vê-se que entre os beneficiários suspensos, a contribuição do qui-quadrado é menor do que entre os não-suspensos, mas ainda indica uma diferença significativa entre as macrorregiões, com 9.304,93 para NÃO-SERRP e 6.425,72 para SERRP. A porcentagem de domicílios suspensos com acesso à rede geral é maior na SERRP (88,4%) do que na NÃO-SERRP (73,3%). O abastecimento por cisterna apresenta a maior contribuição para o qui-quadrado dos suspensos, com 23.879,91 para NÃO-SERRP e 16.144,97 para SERRP, apesar da baixa porcentagem de casos em SERRP. Similar ao grupo de não-suspensos, esse resultado indica que o uso de cisternas é mais comum em NÃO-SERRP e, embora menos prevalente, a presença de cisternas na SERRP

é o principal fator das diferenças entre casos observados e esperados. O V de Cramer de 0,213 indica uma associação moderada entre a macrorregião e o abastecimento de água para a população suspensa.

A maior contribuição para o qui-quadrado dos suspensos na categoria de "Cisterna" evidencia uma disparidade marcante no acesso a infraestrutura de abastecimento de água entre as macrorregiões. O V de Cramer moderado em ambos os grupos (suspensos e não-suspensos) sugere que, há uma importante associação entre a macrorregião e o tipo de abastecimento de água, sendo um dos fatores determinantes.

Figura 4.26: Quantidade de Cômodos do Domicílio e Macrorregião.

VARIÁVEIS	QUANTIDADE DE CÔMODOS DO DOMICÍLIO	Crosstab						MEDIDAS		
		NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	<= 3	13522078	22,9%	263.456,54	6494210	38,2%	916.164,53	0,172	2.250.609,08	0,000
	4 - 4	15394904	26,0%	4.645,02	4774345	28,1%	16.152,95			
	5 - 5	20326898	34,4%	57.600,73	4488670	26,4%	200.305,31			
	6+	9927720	16,8%	176.948,69	1258436	7,4%	615.335,31			
SUSPENSO	<= 3	219666	23,2%	24.771,76	531529	38,7%	17.106,67	0,202	95.050,90	0,000
	4 - 4	242919	25,6%	484,10	378912	27,6%	334,31			
	5 - 5	349413	36,9%	9.998,81	372999	27,2%	6.904,89			
	6+	136090	14,4%	20.969,46	89464	6,5%	14.480,91			

A Figura 4.26 apresenta a análise entre a característica de infraestrutura domiciliar “quantidade de cômodos do domicílio” e “macrorregião”, com foco nas categorias com as maiores contribuições para o qui-quadrado. Para o grupo de beneficiários não-suspensos, a categoria ≤ 3 cômodos apresenta a maior contribuição para o valor de qui-quadrado em ambas as macrorregiões, com 263.456,54 para NÃO-SERRP e 916.164,53 para SERRP. Este resultado indica que há uma maior proporção de domicílios com ≤ 3 cômodos em SERRP (38,2%) em comparação com NÃO-SERRP (22,9%), o que sugere uma diferença significativa na distribuição do tamanho do domicílio entre as duas macrorregiões para os não-suspensos. A categoria “6+” também apresenta diferenças importantes, sendo a segunda maior contribuição para o qui-quadrado dos não-suspensos. O V de Cramer de 0,172 indica uma associação pequena entre a quantidade de cômodos e a macrorregião, sugerindo que outros fatores podem estar influenciando a associação.

Já para o grupo de beneficiários suspensos, a categoria ≤ 3 também se destaca com a maior contribuição para o qui-quadrado, mostrando que 38,7% dos domicílios suspensos em SERRP têm ≤ 3 cômodos, em comparação com 23,2% em NÃO-SERRP. Esta categoria apresenta um importante valor de qui-quadrado de 24.771,76 para NÃO-SERRP e 17.106,67 para SERRP. A categoria “6+” para suspensos apresenta um importante valor de qui-quadrado, sugerindo uma discrepância onde menos domicílios suspensos em SERRP têm um número maior de cômodos. O V de Cramer para suspensos é de 0,202, o que indica uma associação moderada, ligeiramente mais forte entre a quantidade de

cômodos e a macrorregião para os beneficiários suspensos do que para os não-suspensos, o que pode refletir vulnerabilidades no acesso a habitação maior que afetam a estabilidade educacional dessa população.

As maiores contribuições para o qui-quadrado nas categorias ≤ 3 e “6+” cômodos destacam diferenças significativas na infraestrutura habitacional entre as macrorregiões para suspensos e não-suspensos. O maior V de Cramer entre os suspensos indica uma associação mais substancial, sugerindo que as diferenças habitacionais entre as macrorregiões são mais pronunciadas nesse grupo. Isso pode ser indicativo de desigualdades na distribuição de moradias adequadas e acessíveis. Em ambos os grupos, a tendência é de que os domicílios na SERRP são mais propensos a ter menos cômodos do que aqueles na NÃO-SERRP. Valores significativos nas contribuições do qui-quadrado da categoria ≤ 3 indicam áreas onde as características do domicílio são mais proeminentes para representar o perfil das famílias suspensas do programa Bolsa Família.

Figura 4.27: Quantidade de Dormitórios do Domicílio e Macrorregião

VARIÁVEIS	QUANTIDADE DE DORMITÓRIOS DO DOMICÍLIO	Crosstab						MEDIDAS		
		NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
		N	%	X2	N	%	X2			
NÃO-SUSPENSO	<= 1	19647169	33,3%	420.752,64	9631463	56,8%	1.463.405,58	0,205	3.185.690,89	0,000
	2 - 2	30496234	51,7%	146.312,09	6141049	36,2%	508.883,14			
	3 - 3	7769672	13,2%	114.470,79	1091132	6,4%	398.136,98			
	4+	1079120	1,8%	29.863,27	97561	0,6%	103.866,42			
SUSPENSO	<= 1	301840	31,9%	36.536,45	742814	54,2%	25.228,62	0,229	121.228,93	0,000
	2 - 2	507327	53,6%	16.203,99	531694	38,8%	11.188,94			
	3 - 3	121008	12,8%	14.640,69	88573	6,5%	10.109,48			
	4+	16379	1,7%	4.330,52	7732	0,6%	2.990,25			

Analisando a Figura 4.27, entre "quantidade de dormitórios do domicílio" e "macrorregião", considerando a estratificação por beneficiários suspensos e não-suspensos, pode-se destacar que para os beneficiários não-suspensos a categoria ≤ 1 dormitório é a que mais contribui para o valor do qui-quadrado, especialmente na macrorregião SERRP, com um valor de 1.463.405,58. Isso indica que uma proporção significativamente maior de domicílios com um único dormitório está presente em SERRP (56,8%) em comparação com NÃO-SERRP (33,3%). O V de Cramer de 0,205 sugere uma associação moderada entre a quantidade de dormitórios e a macrorregião, indicando que a variável macrorregião tem um efeito razoável sobre o número de dormitórios em um domicílio, pelo menos para a população não-suspensa.

Para a população de suspensos a mesma tendência é observada, onde a categoria ≤ 1 também apresenta a maior contribuição para o qui-quadrado, mostrando que 54,2% dos beneficiários suspensos em SERRP têm domicílios com um único dormitório, contra 31,9% em NÃO-SERRP. O V de Cramer para suspensos é moderado com valor de 0,229, mais alto do que para os não-suspensos, o que também significa uma associação um pouco mais

forte entre a quantidade de dormitórios e a macrorregião para este grupo. Isso reflete uma concentração maior de vulnerabilidade e um impacto mais significativo da suspensão de benefícios nesta dimensão da qualidade de vida.

As categorias com um único dormitório em ambas as macrorregiões, para suspensos e não-suspensos, são as que mais contribuem para a estatística de qui-quadrado, refletindo desigualdades significativas na infraestrutura habitacional entre as macrorregiões. Os valores de V de Cramer indicam que a associação é moderada e significativa para justificar atenção, especialmente no que diz respeito a políticas públicas sociais.

4.2.5 Associação entre Monoparentalidade, PIB Per Capita, Violência e a Exclusão Escolar

Nesta seção analisaremos, mais especificamente, como a relação entre violência e PIB per capita afeta o público suspenso monoparental. Nisto, busca-se descobrir como os fatores sociais e econômicos se entrelaçam na rotina das famílias monoparentais pobres dificultando o cumprimento da presença escolar e conseqüentemente, gerando suspensão do benefício PBF.

Do Suicídio e PIB Per Capita

A Figura I.1 do Anexo I, apresenta a relação entre as variáveis de suicídio e a média categorizada do PIB per capita municipal para a região NÃO-SERRP, considerando o agrupamento de monoparentalidade (SIM/NÃO) e o agrupamento de suspensão (suspenso/não-suspenso). A partir da análise, vê-se que para as famílias não-monoparentais não-suspensas, a categoria com o maior valor de qui-quadrado está na faixa 1 de suicídio para o subgrupo de faixa 3 do PIB (alto), com um qui-quadrado de 224.666,70 indicando a maior contribuição para o qui-quadrado que contribui para a forte associação entre os suicídios e PIB per capita. O V de Cramer geral para este grupo é de 0,136, indicando uma associação moderada. Para as famílias monoparentais não-suspensas, a faixa 1 de suicídio também se destaca com os maiores qui-quadrados nos subgrupos 1 (baixo) e 3 (alto) do PIB, tendo valor de 604.123,93 no subgrupo 1 de PIB e 448.862,73 no subgrupo 3 de PIB, mostrando maiores contribuições para o qui-quadrado e para a associação forte nos locais mais pobres e mais ricos. O V de Cramer é um pouco maior, com um valor de 0,183, sugerindo uma associação forte.

Analisando as famílias não-monoparentais suspensas de NÃO-SERRP, vê-se que aqui, a faixa 1 de suicídio tem um qui-quadrado mais alto no subgrupo de PIB per capita 1 (baixo), com um valor de 5.739,62, a maior contribuição do qui-quadrado esta subpopulação. O V de Cramer é de 0,174, sugerindo uma associação forte, maior do que para as

famílias não-monoparentais e não-suspensas. Para as famílias monoparentais suspensas, novamente, a faixa 1 de suicídio tem o maior qui-quadrado no subgrupo 1 do PIB, com um valor de 12.852,78 indicando o ponto de maior contribuição para o qui-quadrado que influencia na força da associação. O V de Cramer aumenta para 0,201 entre os suspensos, implicando que é na população monoparental suspensa onde está a associação com a maior força entre todas.

Pode-se resumir, portanto, que na região NÃO-SERRP, a associação mais forte entre a faixa de suicídio e o PIB per capita municipal sobre maior influência dos altos valores de qui-quadrado da faixa de suicídio 1 para todas as combinações de monoparentalidade e suspensão, com uma associação mais fraca em famílias não monoparentais não-suspensas e mais forte em famílias monoparentais suspensas.

A Figura I.2 detalha para SERRP a relação entre suicídio e as três faixas categorizadas de PIB per capita. Entre as famílias não-suspensas e não monoparentais de SERRP, a faixa 3 de suicídio tem o maior qui-quadrado associado ao subgrupo 3 do PIB, indicando que as localidades mais ricas estão mais afetadas por menores índices de suicídio. O V de Cramer é super alto em 0,498, sugerindo uma associação muito forte dentro deste grupo. Entre as famílias não-suspensas monoparentais de SERRP, a faixa 3 de suicídio também tem o maior qui-quadrado do grupo e é bem maior do que no grupo não-suspensos, sugerindo um padrão semelhante de associação. O V de Cramer é muito superior em 0,603, sendo o maior de todos os grupos.

Em relação às famílias não-monoparentais suspensas de SERRP, a faixa 3 de suicídio apresenta os maiores valores de qui-quadrado para todos os PIBs (faixa 1, 2 e 3), enfatizando uma associação forte nesta categoria. O V de Cramer de 0,525, mostra uma força de associação muito forte. A relação das famílias monoparentais suspensas de SERRP evidencia que a faixa 3 de suicídio novamente apresenta o maior qui-quadrado, especialmente para a faixa 3 do PIB, mantendo a maior contribuição para o valor do qui-quadrado e para a força da associação, mas agora somente com localidades mais ricas. O V de Cramer aumenta para 0,597, indicando a associação mais forte entre todos os subgrupos de suspensos.

As associações mais fortes entre as taxas de suicídio e o PIB municipal per capita são observadas em famílias com situação monoparental, tanto suspensas como não-suspensas. Os valores V de Cramer refletem que a força da associação é mais pronunciada nestes grupos em comparação com os seus homólogos não-monoparentais.

Morte Violenta Por Causa Indeterminada e PIB Per Capita

A Figura I.3 do Anexo I apresenta a análise do cruzamento entre a variável morte indeterminada categorizada e a variável categorizada de média de PIB per capita municipal

na região NÃO-SERRP. Ela mostra que nas famílias não-monoparentais não-suspensas a faixa 10 de morte indeterminada associada à terceira faixa de PIB (alto) tem o valor de qui-quadrado mais alto do grupo contribuindo para a força da associação, ou seja, os valores mais altos de violência e PIB per capita municipal. O V de Cramer com valor de 0,148 indica uma associação moderada entre as variáveis. Nas famílias não-suspensas monoparentais, a faixa 1 de morte indeterminada se destaca com a contribuição alta para o qui-quadrado relacionada à faixa 1 de PIB (baixo), sugerindo uma associação significativa para famílias monoparentais nas menores faixas de PIB e morte violenta por causa indeterminada. O V de Cramer de 0,205 indica uma associação forte.

A análise de NÃO-SERRP de famílias suspensas, para o grupo não monoparental, apresenta a faixa 1 de morte indeterminada como a de maior contribuição para o qui-quadrado do grupo, associada ao subgrupo 1 de PIB (baixo), ou seja, as famílias de localidades mais pobres estão mais associadas à menor faixa de morte indeterminada. O V de Cramer, com um valor de 0,188, possui força da associação forte, mas menor que nos grupos monoparentais. Para o grupo monoparental suspenso, novamente a faixa 1 de morte indeterminada exibe a mais alta contribuição para o qui-quadrado no subgrupo 1(baixo) de PIB, com o maior V de Cramer entre todos, no valor de 0,236, o que indica a associação mais forte de todos os subgrupos apresentados.

De forma geral, na região NÃO-SERRP, o V de Cramer varia entre 0,188 e 0,236 confirmando que há uma associação forte entre as variáveis de monoparentalidade, faixa de morte indeterminada e PIB per capita municipal. O maior valor de força aparece nas famílias monoparentais, independentemente da suspensão.

Ao analisar a Tabela I.4, focando nas categorias com os maiores valores de qui-quadrado para cada camada de monoparentalidade (Sim/Não) e suspensão (Suspenso/Não-Suspenso) na macrorregião SERRP, vemos que para os beneficiários não-suspensos de famílias não-monoparentais, a faixa 8 de morte indeterminada tem os maiores valores de qui-quadrado em todas as três faixas de PIB, indicando uma associação significativa em todas as categorias de PIB. O V de Cramer de 0,402 sugere uma associação muito forte. Já para as famílias monoparentais, a faixa 8 de morte indeterminada possui a mais alta contribuição para o qui-quadrado (1.459.965) associado ao subgrupo 3 de PIB (alto), o que indica uma associação muito significativa na terceira faixa de PIB per capita municipal. Equivale dizer que as famílias monoparentais não-suspensas são mais afetadas por localidades muito violentas quando economicamente ricas. O V de Cramer de 0,499 apresenta uma medida ainda mais forte para a significância da relação.

Para os beneficiários de SERRP suspensos, quando em famílias não-monoparental, a associação com a faixa 8 de morte indeterminada é similar a para os beneficiários não-suspensos, com alto valor de qui-quadrado em todas as faixas de PIB. O V de Cramer com

valor 0,453 indica uma associação forte. Para as famílias monoparentais, a faixa 8 de morte indeterminada novamente apresenta um alto qui-quadrado, especialmente na segunda e terceira faixas de PIB, o que indica que para as famílias monoparentais suspensas de SERRP, a relação entre localidades ricas e mais violentas é alta. O V de Cramer dos monoparentais suspensos é de 0,532, sendo o mais alto entre todos os grupos, indicando a associação mais forte observada.

Estes resultados refletem como o status socioeconômico e a monoparentalidade, combinados às faixas específicas de morte indeterminada, podem estar relacionados com o status de suspensão.

Homicídio de Mulheres e PIB Per Capita

Para a macrorregião NÃO-SERRP, analisando a Figura I.5 é possível observar que para o agrupamento de famílias não-suspensas não-monoparental, a faixa 1 de homicídio de mulheres tem o maior χ^2 com 823.581,61 na faixa 1 do PIB. O V de Cramer de 0,241 indica uma grande força de associação entre as variáveis. No agrupamento de monoparental sem suspensão, a faixa 1 de homicídio de mulheres tem χ^2 de 1.724.651,61 na faixa 1 do PIB, o que é muito alto e significa uma forte associação. O V de Cramer de 0,282 indica que a força de associação é grande.

Já para os grupos de famílias suspensas NAO-SERRP, quando não monoparentais, tem-se na faixa 1 de homicídio de mulheres os maiores valores de χ^2 relacionados as faixas 1 e 3 do PIB. O V de Cramer é 0,240, comparável ao grupo não monoparental e não-suspensão, sugerindo uma associação semelhante em força. Quando as famílias são suspensas e monoparentais, novamente a faixa 1 de homicídio de mulheres apresenta os maiores χ^2 com 23.835,30 na faixa 1 do PIB e 11.400,75 na faixa 3 do PIB. O V de Cramer de 0,256, um pouco maior do que nos não-suspensos, indica uma ligeira diferença na grandeza da força de associação.

Observamos que as maiores associações em NÃO-SERRP tendem a ocorrer na primeira faixa de PIB, o que pode refletir desigualdades socioeconômicas relacionadas ao homicídio de mulheres nas regiões mais pobres, especialmente nas famílias monoparentais. As associações mais fortes encontradas em famílias monoparentais podem refletir tensões e desafios adicionais enfrentados por essas famílias, especialmente em relação à violência contra as mulheres.

Para a macrorregião SERRP e famílias não-suspensão não-monoparental, a categoria com o maior χ^2 é a faixa 3 de homicídio de mulheres com um χ^2 de 229.340,19 no subgrupo 1 do PIB, seguida por um valor extremamente alto de 1.055.445,86 no subgrupo 3 do PIB que indica uma relação muito significativa. O V de Cramer de 0,444 sugere uma associação forte entre as variáveis. Para as famílias monoparental sem suspensão,

o padrão se repete, onde a faixa 3 de homicídio de mulheres apresenta o maior χ^2 com 1.551.443,50 no subgrupo 3 do PIB, o que é um indicativo de maior relação nesta categoria. O V de Cramer é de 0,538, o que indica uma associação de força ainda maior, sendo mais substancial do que no grupo monoparental não-suspenso.

Quando observados os suspenso não-monoparentais de SERRP, o padrão se mostra ainda mais consistente, tendo na faixa 3 de homicídio de mulheres os maiores valores de χ^2 associados aos subgrupos 1, 2 e 3 do PIB, sendo o subgrupo 3 o de maior contribuição nesta população. O V de Cramer de 0,501 constata uma associação também forte. Para o grupo de famílias monoparentais suspensas, o maior χ^2 é novamente encontrado na faixa 3 de homicídio de mulheres com 136.245,98 relacionado ao subgrupo 3 do PIB. Nas outras faixas de PIB os maiores qui-quadrados também se relacionam ao nível 3 de homicídio de mulheres. Com um V de Cramer de 0,559, a força da associação é a mais forte entre todas.

Em todos os grupos, as maiores contribuições para o qui-quadrado estão principalmente nas categorias de homicídio de mulheres e PIB faixa 3, especialmente no grupo de suspensos, sendo maiores nos grupos suspensos monoparentais. O V de Cramer mais alto no grupo monoparental indica que a condição monoparental pode ter uma interação mais significativa com a incidência de homicídios de mulheres, em comparação com os não-monoparentais.

Homicídio de Homens e PIB Per Capita

Análise dos dados apresentados na Figura I.7, para a macrorregião NÃO-SERRP relacionada ao homicídio de homens e as faixas de PIB per capita municipal, nota-se que na subpopulação de não-suspenso:

- Família Não-monoparental: O maior valor de χ^2 está na faixa 8 de homicídio de homens com um χ^2 de 1.144.807,90 relacionada ao subgrupo 3 do PIB, indicando uma alta contribuição para o qui-quadrado nesta categoria. A próxima maior associação é no mesmo nível de homicídio de homens com o subgrupo 1 do PIB. O V de Cramer é de 0,290, que sugere uma associação grande entre as variáveis.
- Família Monoparental: O maior χ^2 está na faixa 8 de homicídio de homens com 1.524.939,36 no subgrupo 1 do PIB, com a maior significância nesta categoria. O V de Cramer aumenta para 0,315, o que indica uma associação mais forte do que para os não-monoparentais.

Para a subpopulação de suspensos, temos:

- Família Não-monoparental: O maior χ^2 é para a faixa 8 de homicídio de homens com um χ^2 de 9.631,94 no subgrupo 1 do PIB. O V de Cramer de 0,243 mostra uma associação forte.
- Família Monoparental: A maior associação está na faixa 8 de homicídio de homens com um χ^2 de 14.058,32 no subgrupo 1 do PIB. O V de Cramer é de 0,254, indica uma associação forte.

As maiores associações em cada grupo tendem a estar no subgrupo 1 do PIB, sugerindo que a renda pode ter uma relação importante com a incidência de homicídios de homens em ambas as categorias de monoparentalidade e suspensão. Nota-se que as maiores forças das associações estão nos grupos monoparentais, o que pode refletir condições socioeconômicas e dinâmicas familiares específicas sendo afetadas por maiores taxas de homicídio de homens.

Analisando o detalhamento da Figura I.8 para a região SERRP, são observadas as seguintes evidências para a subpopulação sem suspensões:

- Família Não-monoparental: O maior valor de χ^2 é observado na faixa 3 de homicídio de homens dentro do subgrupo 3 do PIB, com um valor de χ^2 de 307.579,26. O valor do V de Cramer é 0,260, indica uma grande associação.
- Família Monoparental: A categoria com o maior valor de χ^2 aqui é a faixa 7 de homicídio de homens relacionado ao subgrupo 1 do PIB, com χ^2 de 804.594. O V de Cramer para essa categoria é de 0,393, sugerindo uma associação mais forte, maior do que para os não-monoparentais.

Ao analisar o grupo de famílias com suspensões, observa-se que:

- Família Não-monoparental: O maior valor de χ^2 é encontrado na faixa 3 de homicídio de homens para todas as faixas do PIB, sendo o maior valor de χ^2 de 36.395,49 no subgrupo 3 do PIB. O V de Cramer é 0,307, que indica uma associação forte.
- Família Monoparental: A faixa 3 de homicídio de homens se destaca com as maiores contribuições para o qui-quadrado daqueles com suspensão, especialmente para o subgrupo 3 do PIB o qual tem o maior valor de χ^2 de 57.457,81. O V de Cramer aumenta para 0,395, indicando uma associação ainda mais forte.

Em resumo, as faixas menores de homicídio de homens se destacam na macrorregião SERRP e as faixas maiores de homicídio de homens se destacam na macrorregião NÃO-SERRP. Em todos os casos, as famílias monoparentais apresentam as associações mais significativas com as maiores contribuições de qui-quadrado. O V de Cramer sugere que enquanto há associações significativas, elas são grandes em força, com destaque para as famílias monoparentais.

4.2.6 Monoparentalidade, Perfil Familiar e a Exclusão Escolar

A seção anterior apresentou os resultados da análise da associação entre **monoparentalidade, PIB per capita, violência e a exclusão escolar**. Nesta seção serão apresentadas as associações entre **monoparentalidade, perfil familiar, infraestrutura domiciliar e a exclusão escolar**.

As variáveis foram modeladas em grupos, com um grupo contendo as características de infraestrutura do domicílio das famílias e outro contendo as características dos membros das famílias, ambos segmentados em monoparentais e não monoparentais. Combinada com a análise de independência por meio do cálculo do qui-quadrado, utiliza-se a medida de tamanho do efeito por subgrupo monoparental, representadas pelo V de Cramer. Juntas, constata-se a relevância prática dos resultados estatísticos destas relações.

Grandes Efeitos do Perfil dos Membros Familiares e da Infraestrutura Domiciliar sobre a Suspensão em Famílias Monoparentais

Crosstab									MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	RAÇA-COR	NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2			
NÃO-SUSPENSOS	NÃO	Branca	11481296	17,9%	1.284.474,81	8707787	52,2%	4.931.371,07	0,327	8.645.055,07	0,000
		Preta	3566590	5,6%	5.258,29	1104952	6,6%	20.187,68			
		Amarela	339023	0,5%	1.121,40	64425	0,4%	4.305,30			
		Parda	47717091	74,5%	472.296,46	6735338	40,4%	1.813.246,24			
		Indígena	973995	1,5%	23.308,30	77901	0,5%	89.485,51			
	SIM	Branca	8661870	17,2%	1.115.789,98	8133788	45,3%	3.126.620,61	0,301	6.191.713,20	0,000
		Preta	3403572	6,8%	29.513,97	1665164	9,3%	82.702,82			
		Amarela	298065	0,6%	1.388,73	99352,77009	0,4%	3.891,43			
		Parda	37569584	74,6%	465.730,12	8038806	44,8%	1.305.049,72			
		Indígena	396787	0,8%	16.050,31	43661	0,2%	44.975,51			
SUSPENSOS	NÃO	Branca	239426	20,1%	80.717,79	633928	50,3%	76.353,08	0,323	255.418,66	0,000
		Preta	87525	7,3%	15,52	90147	7,1%	14,68			
		Amarela	5473	0,5%	39,87	4865	0,4%	37,72			
		Parda	845381	70,9%	47.619,77	527040	41,8%	45.044,80			
		Indígena	14974	1,3%	2.865,18	4984	0,4%	2.710,25			
	SIM	Branca	236870	18,7%	81.250,46	777723	43,6%	57.872,29	0,273	227.135,85	0,000
		Preta	110233	8,7%	348,56	170089	9,5%	248,27			
		Amarela	6693	0,5%	68,43	7850	0,4%	48,74			
		Parda	909563	71,6%	49.113,42	824513	46,2%	34.982,03			
		Indígena	6559	0,5%	1.871,00	2741	0,2%	1.332,66			

Figura 4.28: Raça-Cor, Monoparentalidade e Macrorregião

A Figura 4.28 apresenta a interação entre a raça-cor dos membros familiares dos beneficiários e a macrorregião em que residem, estratificada por beneficiários suspensos e não-suspensos, bem como por famílias monoparentais e não-monoparentais.

Para os beneficiários não-suspensos de famílias não-monoparentais, a maior contribuição para o qui-quadrado vem da categoria racial “branca” na macrorregião SERRP. Em NÃO-SERRP, apesar dos “pardos” terem a maior proporção, a contribuição para o qui-quadrado é menor do que a categoria “branca”, indicando que, embora haja mais indivíduos pardos em NÃO-SERRP, a associação entre raça-cor e macrorregião é mais

pronunciada entre os brancos em SERRP. Isso é destacado pelo V de Cramer de 0,327, sugerindo uma associação forte entre raça-cor e macrorregião.

Para os beneficiários não-suspensos e monoparentais, a categoria “branca” também tem a maior contribuição para o qui-quadrado em SERRP, mas com um V de Cramer ligeiramente menor de 0,301, implicando que a associação é um pouco menos intensa do que nas famílias não-monoparentais. De forma semelhante aos não-monoparentais, a cor “parda” é predominante em NÃO-SERRP com 74,6% e menos comum em SERRP com 44,8%. Novamente, a contribuição para o qui-quadrado de pardos é alta, mas não excede a da categoria “branca” em nenhuma das macrorregiões.

Para os beneficiários suspensos de famílias não-monoparentais, a categoria racial “branca” tem uma maior contribuição para o qui-quadrado em NÃO-SERRP. Isso indica que a suspensão do benefício está fortemente relacionada com a raça-cor, especialmente para as famílias brancas em NÃO-SERRP. Aqui a categoria “parda” também é a maioria em NÃO-SERRP com 70,9%, mas cai para 41,8% em SERRP. O V de Cramer é de 0,323, o que indica uma associação forte, sendo a categoria “branca” a maior diferença entre as duas macrorregiões, o que pode refletir disparidades no impacto da suspensão do benefício entre diferentes raças.

Para os beneficiários suspensos de famílias monoparentais o padrão se mantém, com a raça “branca” mostrando a maior contribuição para o qui-quadrado. Para estas famílias monoparentais suspensas, vemos um padrão similar na categoria “parda”, sendo a maior categoria em NÃO-SERRP com 71,6% e menor em SERRP com 46,2%. Aqui, o V de Cramer é de 0,273 sugere uma associação ligeiramente menos intensa, mas ainda forte, em comparação com os não monoparentais suspensos.

Existe uma associação estatisticamente significativa entre a raça-cor e a macrorregião para todos os estratos da população, com a raça “branca” apresentando as maiores contribuições para o qui-quadrado. A cor “parda” é a categoria mais numerosa em NÃO-SERRP em comparação com SERRP, independentemente do status de suspensão ou monoparentalidade. A associação é mais forte para as famílias não-monoparentais do que para as monoparentais. Para todos os grupos, SERRP tem uma proporção maior de indivíduos brancos em comparação com NÃO-SERRP. Em todos os casos de “parda”, o qui-quadrado indica que a associação é substancial, mas não tão forte quanto para a categoria “branca”.

VARIÁVEIS	MONOPARENTAL?	ESCOAMENTO SANITÁRIO DE ÁGUA NO DOMICÍLIO	Crosstab						MEDIDAS		
			NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2			
NÃO-SUSPENSOS	NÃO	Rede coletora de esgoto ou pluvial	6292856	25,7%	482.042,15	4182511	64,9%	1.830.857,32	0,343	3.636.531,99	0,000
		Fossa séptica	4646457	19,0%	34.116,79	741467	11,5%	129.579,91			
		Fossa rudimentar	11900578	48,6%	217.300,56	1234401	19,2%	825.335,13			
		Vala a céu aberto	1056830	4,3%	12.605,90	140184	2,2%	47.878,82			
		Direto para um rio, lago ou mar	237971	1,0%	5.257,27	110782	1,7%	19.967,79			
		Outra forma	331190	1,4%	6.583,89	32220	0,5%	25.006,45			
	SIM	Rede coletora de esgoto ou pluvial	10118366	37,1%	714.187,68	8037642	78,2%	1.895.489,88	0,380	5.423.158,24	0,000
		Fossa séptica	5215601	19,1%	164.558,53	797916	7,8%	436.746,57			
		Fossa rudimentar	10677300	39,2%	587.735,93	955011	9,3%	1.559.880,59			
		Vala a céu aberto	751110	2,8%	1.447,89	238588	2,3%	3.842,76			
		Direto para um rio, lago ou mar	239820	0,9%	13.610,28	178943	1,7%	36.122,36			
		Outra forma	260431	1,0%	2.609,65	63985	0,6%	6.926,13			
SUSPENSOS	NÃO	Rede coletora de esgoto ou pluvial	135185	37,2%	26.033,14	350707	72,8%	19.639,51	0,372	116.654,81	0,000
		Fossa séptica	59383	16,3%	6.523,70	39904	8,3%	4.921,51			
		Fossa rudimentar	147244	40,5%	32.324,08	67983	14,1%	24.385,41			
		Vala a céu aberto	12043	3,3%	846,43	9458	2,0%	638,55			
		Direto para um rio, lago ou mar	5227	1,4%	357,43	10550	2,2%	269,65			
		Outra forma	4242	1,2%	407,78	3002	0,6%	307,63			
	SIM	Rede coletora de esgoto ou pluvial	245456	48,7%	32.284,76	714698	82,1%	18.692,29	0,392	210.826,14	0,000
		Fossa séptica	76951	15,3%	21.263,34	47840	5,5%	12.311,09			
		Fossa rudimentar	162449	32,2%	78.298,90	61767	7,1%	45.333,64			
		Vala a céu aberto	9593	1,9%	75,15	18989	2,2%	43,51			
		Direto para um rio, lago ou mar	5784	1,1%	1.598,08	20742	2,4%	925,26			
		Outra forma	4048	0,8%	0,08	6943	0,8%	0,05			

Figura 4.29: Escoamento Sanitário, Monoparentalidade e Macrorregião

A Figura 4.29 apresenta a relação entre escoamento sanitário do domicílio nas macrorregiões SERRP e NÃO-SERRP, agrupando a população por famílias suspensas e não-suspensas e monoparentais e não monoparentais. Dentre os principais pontos, baseados nos maiores valores de qui-quadrado e do V de Cramer, tem-se que para os beneficiários não-suspensos de famílias não-monoparentais existe uma diferença significativa nas distribuições da categoria “rede coletora de esgoto ou pluvial”, onde é bem mais comum em SERRP (64,9%) em comparação com NÃO-SERRP (25,7%). Isso é refletido no alto valor de qui-quadrado e um V de Cramer de 0,343, sugerindo uma associação forte.

Para a subpopulação de não-suspensos monoparentais, o padrão é amplificado nas famílias monoparentais, com a “rede coletora de esgoto ou pluvial” ainda mais predominante em SERRP (78,2%) em relação a NÃO-SERRP (37,1%). O valor de qui-quadrado desta categoria é ainda maior para este grupo e o V de Cramer aumenta para 0,380, indicando uma associação ainda mais robusta.

Para os beneficiários suspensos de famílias não-monoparentais, “fossa rudimentar” é a categoria com a maior contribuição para o valor do qui-quadrado, o que indica uma variação significativa entre as macrorregiões NÃO-SERRP e SERRP. Em NÃO-SERRP, observa-se que uma parcela substancial de famílias (40,5%) depende de fossas rudimentares, enquanto em SERRP, essa porcentagem é significativamente menor (14,1%). O V de Cramer é de 0,372, mantém uma associação forte.

Para a subpopulação suspensa de famílias monoparentais, a categoria “fossa rudimentar” também mostra uma grande disparidade entre as regiões, com 32,2% em NÃO-

SERRP em comparação com 7,1% em SERRP. O V de Cramer é o mais alto de todos, com 0,392, mostrando que a questão do saneamento inadequado é particularmente prevalente entre as famílias suspensas monoparentais e que a maior contribuição para o χ^2 está na categoria “fossa rudimentar” de NÃO-SERRP.

O V de Cramer nos diz que, quanto maior o valor, mais forte é a associação entre as variáveis. Nos casos apresentados, os valores de 0,343 a 0,392 mostram que há uma relação clara entre a macrorregião e o tipo de escoamento sanitário dos domicílios em todos os estratos de monoparentalidade. Para alunos não-suspensos, as famílias em SERRP têm muito mais probabilidade de ter acesso a uma rede coletora de esgoto ou pluvial, enquanto em NÃO-SERRP, métodos menos centralizados, como fossas sépticas e rudimentares, são mais comuns. As fossas rudimentares mostram-se particularmente associadas à suspensão de benefícios, em ambas as regiões e em famílias monoparentais ou não. Além disso, a categoria de fossa séptica também contribui significativamente para o alto valor do qui-quadrado em todos os grupos, principalmente entre os monoparentais suspensos, destacando as diferenças no acesso ao saneamento básico entre os beneficiários. A disparidade entre as macrorregiões é particularmente pronunciada nas famílias monoparentais, sugerindo que elas podem enfrentar desafios adicionais relacionados à infraestrutura e serviços básicos. Um p-valor de 0,000 em todas as estratificações indica que as diferenças observadas são estatisticamente significativas e não ocorrem por acaso.

Os Efeitos Moderados da Infraestrutura Domiciliar sobre a Suspensão em Famílias Monoparentais

Crosstab														
VARIÁVEIS	MONOPARENTAL?	ILUMINAÇÃO DO DOMICÍLIO	NAO-SERRP			SERRP			MEDIDAS					
			N	%	X2	N	%	X2	V DE CRAMER	X2 TOTAL	P-VALOR			
NÃO-SUSPENSOS	NÃO	Elétrica com medidor próprio	24333788	84,4%	7.637,31	5026846	76,4%	33.475,74	0,192	1.311.296,09	0,000			
		Elétrica com medidor comunitário	980337	3,4%	191.444,84	886111	13,5%	839.137,52						
		Elétrica sem medidor	1663280	5,8%	6.209,76	508156	7,7%	27.218,53						
		Óleo, querosene ou gás	686765	2,4%	26.600,65	6218	0,1%	116.595,46						
		Vela	360462	1,3%	9.572,19	15735	0,2%	41.956,65						
	Outra forma	800814	2,8%	2.126,52	133309	2,0%	9.320,91							
	SIM	Elétrica com medidor próprio	25513225	84,1%	47.669,68	7326219	70,2%	138.571,87				0,214	1.866.109,06	0,000
		Elétrica com medidor comunitário	1211627	4,0%	297.306,78	1447629	13,9%	864.246,52						
		Elétrica sem medidor	2110995	7,0%	89.361,00	1373068	13,2%	259.765,13						
		Óleo, querosene ou gás	389029	1,3%	29.908,95	7566	0,1%	86.942,88						
Vela		257988	0,9%	12.720,26	19832	0,2%	36.976,76							
Outra forma	863247	2,8%	675,53	264958	2,5%	1.963,71								
SUSPENSOS	NÃO	Elétrica com medidor próprio	335541	82,3%	1.459,20	356212	72,7%	1.213,86	0,202	36.488,69	0,000			
		Elétrica com medidor comunitário	20026	4,9%	11.846,85	73458	15,0%	9.855,01						
		Elétrica sem medidor	28013	6,9%	1.027,99	46678	9,5%	855,15						
		Óleo, querosene ou gás	7469	1,8%	4.201,24	430	0,1%	3.494,87						
		Vela	4067	1,0%	1.226,64	1139	0,2%	1.020,41						
	Outra forma	12711	3,1%	156,92	12338	2,5%	130,54							
	SIM	Elétrica com medidor próprio	435624	80,6%	5.608,63	588690	66,7%	3.432,99				0,198	55.846,86	0,000
		Elétrica com medidor comunitário	30494	5,6%	17.587,67	138223	15,7%	10.765,26						
		Elétrica sem medidor	47898	8,9%	5.808,45	130493	14,8%	3.555,30						
		Óleo, querosene ou gás	3997	0,7%	3.347,08	344	0,0%	2.048,72						
Vela		3491	0,6%	1.887,31	986	0,1%	1.155,21							
Outra forma	18758	3,5%	403,35	23913	2,7%	246,88								

Figura 4.30: Iluminação do Domicílio, Monoparentalidade e Macrorregião

Na Figura 4.30, vê-se que a categoria “Elétrica com medidor comunitário” tem a maior contribuição para o valor do qui-quadrado, especialmente quando comparada com “Elétrica com medidor próprio”. Entre os não-suspensos, para famílias não-monoparentais, há uma porcentagem maior de uso de medidor elétrico comunitário (3,4% em NÃO-SERRP vs. 13,5% em SERRP) e os maiores qui-quadrado das categorias (191.444,84 em NÃO-SERRP vs. 839.137,52 em SERRP). A força da associação é confirmada com um V de Cramer moderado de 0,192.

Para famílias não-suspensas e monoparentais, a diferença nas proporções de medidor comunitário é ainda mais marcante (4,0% em NÃO-SERRP vs. 13,9% em SERRP), com contribuições para o qui-quadrado ainda maiores (297.306,78 em NÃO-SERRP e 864.246,52 em SERRP). Uma moderada associação é mostrada com um V de Cramer de 0,214.

Entre os suspensos, as famílias não-monoparentais mostram um qui-quadrado alto na categoria de medidor comunitário (11.846,85 em NÃO-SERRP e 9.855,01 em SERRP) e a diferença percentual entre as regiões ainda é marcante (4,9% em NÃO-SERRP vs. 15,0% em SERRP). O V de Cramer é de 0,202, indica uma associação moderada.

As famílias monoparentais suspensas apresentam um padrão similar às não-monoparentais, com proporções de medidor comunitário de 5,6% em NÃO-SERRP e 15,7% em SERRP. As contribuições para qui-quadrado nesta categoria de iluminação são maiores do que aquelas da mesma categoria nas famílias não-monoparentais suspensas (17.587,67 em NÃO-SERRP e 10.765,26 em SERRP) e o V de Cramer de 0,198 reflete uma associação moderada.

O V de Cramer varia de 0 a 1, onde valores mais próximos de 0 indicam uma associação mais fraca e valores mais próximos de 1 indicam uma associação mais forte. Neste caso, de acordo com a Tabela de Equivalência de Medida Simétricas 4.8 adotada nesta pesquisa, para variáveis com 6 categorias, todos os valores de V de Cramer a partir de 1,34 são moderados, mostrando que a força dessas associações é moderada. Em todas as categorias, famílias de SERRP têm maior associação com eletricidade com medidor comunitário do que famílias de NÃO-SERRP. A dependência é mais acentuada entre as famílias suspensas, indicando que os desafios relacionados à suspensão são críticos para essas famílias, especialmente nos estados de SERRP.

VARIÁVEIS	MONOPARENTAL?	ABASTECIMENTO DE ÁGUA DO DOMICÍLIO	Crosstab						MEDIDAS		
			NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2			
NÃO-SUSPENSOS	NÃO	Rede geral de distribuição	16.661.619	57,8%	77.031	5.240.730	79,7%	337.640,40	0,191	1.289.536,48	0,000
		Poço ou nascente	7.795.322	27,0%	32.405	1.180.774	18,0%	142.036,81			
		Cisterna	2.041.021	7,1%	77.579	22.268	0,3%	340.041,94			
		Outra forma	2.327.479	8,1%	52.534	132.600	2,0%	230.268,26			
	SIM	Rede geral de distribuição	21.577.950	71,1%	76.422,79	9.200.991	88,1%	222.154,87	0,178	1.292.375,91	0,000
		Poço ou nascente	5.480.002	18,1%	105.305,93	932.492	8,9%	306.115,83			
		Cisterna	1.126.894	3,7%	82.289,55	29.964	0,3%	239.209,08			
		Outra forma	2.161.270	7,1%	66.773,29	275.823	2,6%	194.104,57			
SUSPENSOS	NÃO	Rede geral de distribuição	270.957	66,4%	5.421,82	416.306	84,9%	4.510,23	0,243	52.876,94	0,000
		Poço ou nascente	86.335	21,2%	5.855,59	60.301	12,3%	4.871,07			
		Cisterna	20.067	4,9%	11.537,33	992	0,2%	9.597,52			
		Outra forma	30.468	7,5%	6.050,32	12.656	2,6%	5.033,06			
	SIM	Rede geral de distribuição	423.608	78,4%	3.487,92	798.002	90,4%	2.134,92	0,180	46.267,96	0,000
		Poço ou nascente	66.281	12,3%	8.602,37	55.713	6,3%	5.265,43			
		Cisterna	13.051	2,4%	9.917,15	1.696	0,2%	6.070,19			
		Outra forma	37.321	6,9%	6.693,16	27.238	3,1%	4.096,82			

Figura 4.31: Abastecimento de Água, Monoparentalidade e Macrorregião

Na Figura 4.31 pode-se observar as maiores contribuições do qui-quadrado e a estatística V de Cramer para a variável "Abastecimento de Água do Domicílio" e sua relação com a macrorregião (SERRP e NÃO-SERRP) por situação de monoparentalidade e por status de suspensão. Sua análise mostra que para os domicílios não-suspensos, nos agregados familiares não monoparentais, a categoria "Cisterna" apresenta o maior valor de qui-quadrado, particularmente na região SERRP, indicando diferenças significativas nas frequências esperadas desta macrorregião. Para as famílias monoparentais, é observada uma alta contribuição para qui-quadrado na categoria "Poço ou nascente", mostrando que a baixa condição de acesso a água é igualmente pronunciada para as famílias monoparentais. Percentuais maiores de famílias dependentes de "poços ou nascentes" em

NÃO-SERRP (27,0% e 18,1%) em comparação com SERRP (18,0% e 8,9%), evidenciam o acesso precário ao abastecimento de água nessa localidade.

Em relação aos domicílios suspensos, entre os agregados familiares não monoparentais, o maior valor qui-quadrado está na categoria "Cisterna" em NÃO-SERRP (4,9%) em comparação com SERRP(0,2%), o que mostra uma associação semelhante aos agregados familiares não monoparentais não-suspensos. Nos agregados familiares monoparentais, vemos novamente um padrão semelhante onde a "Cisterna" detém o maior valor de qui-quadrado, especialmente em NÃO-SERRP (2,4%) quando comparado à SERRP (0,2%), implicando uma discrepância significativa nas fontes de abastecimento de água dessas localidades.

Os valores de V de Cramer, neste caso, superiores a 0,173, são suficiente para concluir uma associação moderada entre tipos de abastecimento de água e macrorregiões, independentemente dos arranjos familiares existentes. O fato de o V de Cramer ser maior nos domicílios suspensos indica uma relação mais forte entre abastecimentos de água precarizados e a macrorregião para essas famílias. Isto mostra também vulnerabilidades adicionais nos serviços de acesso à água das famílias suspensas e como isso se dá de forma polarizada regionalmente (mais comum em NÃO-SERRP), potencialmente devido a problemas infra-estruturais desse local. A tendência geral mostra que cisternas e poços ou nascentes são mais prevalentes na região NÃO-SERRP em todas as categorias.

VARIÁVEIS	MONOPARENTAL?	MATERIAL PREDOMINANTE NAS PAREDES DO DOMICÍLIO	Crosstab						MEDIDAS		
			NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2			
NÃO-SUSPENSOS	NÃO	Alvenaria/tijolo com revestimento	18161230	63,0%	3.806,61	4468652	68,0%	16.685,05	0,137	665.737,63	0,000
		Alvenaria/tijolo sem revestimento	4777839	16,6%	572,98	1154651	17,6%	2.511,50			
		Madeira aparelhada	2365217	8,2%	9.680,88	731493	11,1%	42.433,06			
		Taipa revestida	802606	2,8%	31.863,05	5303	0,1%	139.661,56			
		Taipa não revestida	997305	3,5%	40.375,07	4634	0,1%	176.971,29			
		Madeira aproveitada	573097	2,0%	2.187,23	175595	2,7%	9.587,01			
		Palha	95782	0,3%	3.936,72	299	0,0%	17.255,37			
		Outro Material	1052359	3,7%	31.247,36	35744	0,5%	136.962,90			
	SIM	Alvenaria/tijolo com revestimento	20292470	66,9%	6.534,14	7474559	71,6%	18.994,21	0,131	700.190,91	0,000
		Alvenaria/tijolo sem revestimento	5383377	17,7%	2.332,76	2004108	19,2%	6.781,14			
		Madeira aparelhada	1807100	6,0%	901,50	676511	6,5%	2.620,59			
		Taipa revestida	647228	2,1%	53.543,18	5861	0,1%	155.645,67			
		Taipa não revestida	856896	2,8%	71.121,68	7355	0,1%	206.744,94			
		Madeira aproveitada	565868	1,9%	659,10	221065	2,1%	1.915,95			
Palha	56439	0,2%	4.708,93	442	0,0%	13.688,46					
Outro Material	736733	2,4%	39.416,91	49370	0,5%	114.581,75					
SUSPENSOS	NÃO	Alvenaria/tijolo com revestimento	263611	64,6%	251,96	335117	68,4%	209,59	0,181	29.344,38	0,000
		Alvenaria/tijolo sem revestimento	72324	17,7%	59,05	91558	18,7%	49,12			
		Madeira aparelhada	33938	8,3%	182,60	46484	9,5%	151,90			
		Taipa revestida	8143	2,0%	4.754,18	356	0,1%	3.954,85			
		Taipa não revestida	10229	2,5%	6.118,19	355	0,1%	5.089,52			
		Madeira aproveitada	8111	2,0%	366,34	13971	2,8%	304,74			
		Palha	922	0,2%	577,67	16	0,0%	480,55			
		Outro Material	10549	2,6%	3.708,86	2398	0,5%	3.085,28			
	SIM	Alvenaria/tijolo com revestimento	372226	68,9%	198,49	631024	71,5%	121,49	0,151	32.471,02	0,000
		Alvenaria/tijolo sem revestimento	100843	18,7%	107,76	173577	19,7%	65,96			
		Madeira aparelhada	30011	5,6%	24,90	51340	5,8%	15,24			
		Taipa revestida	7281	1,3%	6.336,40	500	0,1%	3.878,45			
		Taipa não revestida	9400	1,7%	8.487,72	489	0,1%	5.195,25			
		Madeira aproveitada	9740	1,8%	376,70	21478	2,4%	230,58			
Palha	602	0,1%	566,70	20	0,0%	346,87					
Outro Material	10158	1,9%	4.043,51	4221	0,5%	2.475,00					

Figura 4.32: Material Predominante nas Paredes, Monoparentalidade e Macrorregião

A Figura 4.32 apresenta uma análise detalhada da relação entre o material predominante nas paredes do domicílio e a macrorregião, considerando também se os beneficiários são suspensos ou não e se as famílias são monoparentais ou não. Nos dados apresentados, para as famílias não-suspensas, tanto não monoparentais como monoparentais, vê-se as maiores contribuições para o qui-quadrado na categoria "Taipa não revestida". Isso indica que, independentemente da suspensão do benefício, as famílias dessa categoria são as que maior contribuem para a significância da associação.

De maneira similar, nas famílias não-suspensas, tanto com status de não monoparentais como as monoparentais, a categoria "Taipa não revestida" possui a mais alta contribuição para o qui-quadrado, sugerindo que uma proporção significativa dessas famílias vive em condições de moradia menos estruturadas. Com o maior V de Cramer de todos os grupos (1,81), o grupo suspenso pode refletir uma situação de maior vulnerabilidade socioeconômica.

Esses achados indicam que as famílias beneficiárias, em todos os tipos de arranjos familiares, vivem em uma situação de moradia não ideal que responde pela força da associação entre suspensão e não suspensão de benefício, possivelmente enfrentando desafios relacionados à qualidade do abrigo. O valor do V de Cramer em todas as subcategorias é moderado, o que mostra a força da associação entre o material das paredes do domicílio e as macrorregiões. A associação é mais forte em famílias suspensas não monoparentais, conforme indicado pelo V de Cramer de 0,181 e é menos expressiva nas famílias não-suspensas monoparentais, com um V de Cramer de 0,131.

VARIÁVEIS	MONOPARENTAL?	Crosstab							MEDIDAS		
		QUANTIDADE DE CÔMODOS DO DOMICÍLIO	NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2			
NÃO-SUSPENSOS	NÃO	<= 3	5445415	18,9%	52.774,98	1933932	29,4%	231.322,22	0,131	605.240,93	0,000
		4 - 4	7273001	25,2%	3.096,43	1845514	28,1%	13.572,21			
		5 - 5	10098337	35,0%	3.043,56	2090434	31,8%	13.340,46			
		6+	6008713	20,8%	53.516,89	706502	10,7%	234.574,18			
	SIM	<= 3	8076663	26,6%	186.938,60	4560278	43,7%	543.415,30	0,188	1.436.363,67	0,000
		4 - 4	8121903	26,8%	1.224,19	2928831	28,1%	3.558,61			
		5 - 5	10228561	33,7%	73.977,14	2398236	23,0%	215.045,53			
		6+	3919007	12,9%	105.506,24	551934	5,3%	306.698,05			
SUSPENSOS	NÃO	<= 3	81014	19,9%	5.378,03	149655	30,5%	4.473,81	0,166	24.733,15	0,000
		4 - 4	99773	24,5%	430,74	134857	27,5%	358,32			
		5 - 5	152451	37,4%	873,16	158800	32,4%	726,35			
		6+	74589	18,3%	6.819,67	46943	9,6%	5.673,06			
	SIM	<= 3	138652	25,7%	17.604,23	381874	43,3%	10.775,38	0,215	65.539,77	0,000
		4 - 4	143146	26,5%	101,84	244055	27,7%	62,33			
		5 - 5	196962	36,5%	10.688,88	214199	24,3%	6.542,56			
		6+	61501	11,4%	12.260,20	42521	4,8%	7.504,35			

Figura 4.33: Quantidade de Cômodos do Domicílio, Monoparentalidade e Macrorregião

Analisando a Figura 4.33, vê-se que a relação mostra força de associação significativa apenas para os arranjos monoparentais, com V de Cramer superior a 0,173, sendo 0,188 para os monoparentais não-suspensos e 0,215 para os monoparentais suspensos. Para as famílias não monoparentais, independentemente do status de suspensão, a força da associação é pequena.

Para as famílias não-suspensas monoparentais, a categoria «= 3" cômodos tem alta contribuição para o qui-quadrado, com um V de Cramer de (0,188), sugerindo uma associação moderada entre as variáveis. Isso pode refletir uma dinâmica socioeconômica diferente para as famílias monoparentais em relação ao número de cômodos em seus domicílios.

Para as famílias suspensas monoparentais, a categoria «= 3" cômodos é a que mais contribui para o valor do qui-quadrado, com o maior V de Cramer de 0,215, o que é uma associação moderada mais forte. Isso implica que, entre as famílias suspensas monoparentais, a diferença na proporção de famílias com ≤ 3 cômodos entre as macrorregiões é mais pronunciada é a mais significativamente entre as macrorregiões.

Há uma associação estatisticamente significativa entre a quantidade de cômodos do domicílio e a macrorregião entre todos os subgrupos monoparentais, com a força dessa associação mais forte entre os suspensos, conforme indicado pelos valores de V de Cramer. A prevalência da categoria «= 3" cômodos como contribuinte significativo para o qui-quadrado em vários subgrupos sugere que essa característica de moradia está relacionada a questões socioeconômicas que afetam as condições de vida dessas famílias e influencia a permanência na escola, variando de uma região para outra.

VARIÁVEIS	MONOPARENTAL?	QUANTIDADE DE DORMITÓRIOS DO DOMICÍLIO	Crosstab						MEDIDAS		
			NAO-SERRP			SERRP			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2			
NÃO-SUSPENSOS	NÃO	≤ 1	7796553	27,1%	109.182,64	2979908	45,5%	478.803,49	0,160	905.167,67	0,000
		2 - 2	15612172	54,3%	20.122,33	2884019	44,0%	88.243,36			
		3 - 3	4677488	16,3%	28.464,69	635661	9,7%	124.827,47			
		4+	658136	2,3%	10.310,15	55051	0,8%	45.213,54			
	SIM	≤ 1	11850616	39,2%	266.518,04	6651555	63,9%	774.664,43	0,220	1.966.617,23	0,000
		2 - 2	14884062	49,2%	142.458,89	3257030	31,3%	414.072,66			
		3 - 3	3092184	10,2%	77.622,31	455471	4,4%	225.617,90			
		4+	420984	1,4%	16.808,17	42510	0,4%	48.854,83			
SUSPENSOS	NÃO	≤ 1	102662	25,2%	11.123,86	211095	43,1%	9.252,64	0,200	35.749,40	0,000
		2 - 2	227391	55,8%	2.209,21	226391	46,2%	1.837,58			
		3 - 3	67269	16,5%	4.312,31	47814	9,8%	3.586,91			
		4+	9871	2,4%	1.870,80	4242	0,9%	1.556,10			
	SIM	≤ 1	199178	36,9%	22.102,97	531719	60,3%	13.527,60	0,233	77.279,09	0,000
		2 - 2	279936	51,9%	15.005,87	305303	34,6%	9.183,98			
		3 - 3	53739	10,0%	8.892,46	40759	4,6%	5.442,42			
		4+	6508	1,2%	1.937,81	3490	0,4%	1.185,99			

Figura 4.34: Quantidade de Dormitórios do Domicílio, Monoparentalidade e Macrorregião

A análise da Figura 4.34 mostra que a relação da variável "quantidade de dormitórios" tem associação moderada em todos os grupos, com exceção das famílias não-monoparentais não-suspensas. Para este subgrupo em específico, o V de Cramer de 0,160 é pequeno e apesar de possuir um qui-quadrado alto, não mostra aplicação prática para a significância como as demais subpopulações.

No caso das famílias não-suspensas monoparentais, a categoria «= 1 dormitórios" se destaca, com o segundo maior V de Cramer de 0,220 entre os graus moderados dos agrupamentos, indicando uma associação um pouco mais forte. Isso pode sugerir que a monoparentalidade tem uma relação mais significativa com a quantidade de dormitórios do

domicílio, possivelmente devido a limitações de espaço ou recursos financeiros mais restritos nessas famílias.

Quanto às famílias suspensas não monoparentais, novamente a categoria «= 1 dormitórios» é a mais significativa em termos de contribuição para o qui-quadrado, com um V de Cramer de 0,200. Este valor aponta para uma associação substancial, reforçando a ideia de que a quantidade de dormitórios é um fator distintivo entre as macrorregiões para essas famílias.

Por fim, para as famílias suspensas monoparentais, a maior contribuição vem da categoria «= 1 dormitórios» com o maior V de Cramer de 0,233 entre todos os grupos de força moderada. Esta é a associação mais forte entre todos os grupos analisados, o que pode refletir como as condições de moradia das famílias suspensas monoparentais são particularmente afetadas pelas características econômicas regionais.

A recorrência da categoria «= 1 dormitórios» como a principal contribuinte para o qui-quadrado em quase todos os subgrupos destaca a importância do tamanho da moradia como um indicador da situação educacional das famílias em diferentes macrorregiões. As diferenças regionais nesta categoria podem ser influenciadas por uma série de fatores e, dado a maior influência dos grupos monoparentais, o fator de composição e recursos das famílias parece ser determinante.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

A exclusão escolar do público beneficiário do Programa Bolsa Família se mostrou um problema complexo, que sofre a influência de diversas variáveis. De forma abrangente, foram encontradas evidências que fatores como o perfil familiar descrito em termos de infraestrutura domiciliar e características dos membros da família, assim como os fatores socioeconômicos influenciam as altas taxas de exclusão escolar, especialmente em arranjos familiares monoparentais de localidades mais desenvolvidas.

Sobre os motivos de ausência, vê-se que uma parcela significativa das faltas carece de motivos detalhados, o que dificulta a análise das motivações do absenteísmo. A análise destaca que a maioria das ausências dos alunos são inexplicáveis devido a motivos categorizados inadequadamente, tornando difícil analisar diretamente as motivações a partir dos registros de frequência escolar. As motivações mais frequentes para as faltas escolares foram registradas como desinteresse ou abandono escolar dos próprios aluno ou pela negligência dos familiares. O tratamento de doenças também aparece entre os 5 (cinco) motivos mais frequentes junto com a falta de matrícula do aluno.

A análise do público segundo classificação geográfica mostra que os beneficiários estão segmentados em cinco zonas geográficas definidas pelo IBGE. Estas zonas ajudam a compreender as diferentes condições socioeconômicas das diferentes regiões. Estas zonas são rurais (remotas ou adjacentes), intermediárias (remotas ou adjacentes) e urbanas. A classificação é baseada em fatores como densidade populacional e proximidade de centros urbanos, fundamentais para avaliar o impacto das políticas públicas. O Sudeste possui os maiores índices motivados por tratamento de doença e de atenção à saúde do aluno e também o maior percentual de faltas devido a dificuldades de deslocamento ou acesso à escola. O Sul é a região com o maior índice de faltas por desinteresse ou desmotivação. O Nordeste lidera as faltas por problemas sociais, educacionais e familiares. O Centro-Oeste

tem o maior percentual de ausências devido à falta de vínculo ou matrícula escolar. O Norte lidera faltas por negligência parental.

As discrepâncias nas suspensões apresentam um paradoxo, onde regiões com maior número de beneficiários (muitas vezes regiões mais pobres como o Norte e o Nordeste) têm menos suspensões em comparação com regiões mais ricas como o Sul e o Sudeste. Uma conclusão interessante é a proporcionalidade inversa entre o número de beneficiários e as suspensões entre regiões. Regiões com maior pobreza e mais beneficiários, como Norte e Nordeste, têm menos suspensões em comparação com regiões mais ricas, como Sul e Sudeste. Isto sugere que as suspensões não são apenas uma função do número de beneficiários, mas são influenciadas por outros fatores. Comparando os estados entre si, pode-se observar que os estados de São Paulo, Espírito Santo, Rio de Janeiro, Rio Grande do Sul e Paraná possuem as maiores taxas de suspensão, tendo medida de dispersão até 26 vezes maior que o estado do Amapá que apresenta os menores valores para “Taxa de Suspensão”, “Indicador de Frequência Máxima Esperada”, “Taxa de Risco de Suspensão”, “Dispersão de Absenteísmo dos Suspensos (DAS)” e a “Dispersão da Taxa de Suspensão (DTS)”. Em termos simples, ao utilizar um estado com a menor taxa de suspensão e o menor grau de risco de suspensão por aluno suspenso como referência, pode-se comparar os estados entre si de forma isonômica, fornecendo uma visão mais clara das disparidades nacionais existentes.

Duas macrorregiões distintas, SERRP e NÃO-SERRP, são estabelecidas para uma análise mais detalhada e multidimensional. As discrepâncias nas métricas de absenteísmo entre os estados sugerem diferentes dinâmicas regionais que influenciam as suspensões. As macrorregiões SERRP (que compreende os cinco estados com maiores taxas de suspensão e mais ricos) e NÃO-SERRP (os demais estados) são analisadas para compreender as diferenças na conformidade educacional.

O estudo das famílias monoparentais dentro do grupo beneficiário mostra uma maior vulnerabilidade considerando a insuficiência econômica já instalada, influenciando as suspensões. Os resultados mostram riscos acentuados nas famílias monoparentais entre os beneficiários, constatando que estas famílias enfrentam maiores dificuldades socioeconômicas. A prevalência da monoparentalidade está ligada ao aumento das suspensões, levando a uma maior vulnerabilidade e a um menor cumprimento das condições do programa. O resultado aponta a prevalência da monoparentalidade entre as famílias suspensas nos estados, afetando aproximadamente 62% das famílias suspensas do país. Os cinco maiores índices estaduais de monoparentalidade são, respectivamente, Rio de Janeiro, Distrito Federal, Pernambuco, São Paulo e Rio Grande do Sul, três deles na macrorregião SERRP. Percebe-se também que, dos 26 estados mais Distrito Federal, 21 deles possuem índices de monoparentalidade superiores a 50% em suas populações suspensas.

Os indicadores de violência são analisadas por meio de taxas de vários tipos de violência (incluindo suicídio e homicídio). Taxas de violência mais elevadas associam-se com taxas de suspensão mais elevadas, sugerindo interações complexas entre ambientes familiares e cumprimento do programa. Observando as entropias dos modelos de discretização das taxas de violência para cada macrorregião, temos que na macrorregião NÃO-SERRP a entropia do modelo é de 0,115 para todos os tipos de violência enquanto em SERRP, a entropia do modelo é aproximadamente 0,373 para todas as violências. Isso significa que existe uma distribuição mais previsível de violência em NAO-SERRP (regiões mais pobres) e que em SERRP (regiões mais ricas) há uma maior variação nas taxas de violência municipal em relação à variável “suspenso”. A macrorregião SERRP tem maiores valores máximo de taxa de suicídio e homicídio de mulheres em comparação com a NÃO-SERRP, o que indica incidências mais altas desse tipo de violência nos estados dessa macrorregião, sugerindo que fatores socioambientais externos das regiões mais ricas influenciam o cumprimento das condições de bem-estar, especificamente por serem mais violentas com mulheres e estimularem maior suicídio, podendo estar mais propensas a suspensões de programa.

Os índices de fatores econômicos mostram que as condições econômicas, representadas pelo PIB municipal per capita, indicam que um maior desenvolvimento econômico está relacionado com taxas de suspensão mais elevadas. A análise das condições econômicas com base no PIB per capita municipal mostra que as regiões com maior desenvolvimento econômico tendem a ter taxas de suspensão mais elevadas, especialmente na região SERRP. Isto pode indicar que, à medida que as condições econômicas melhoram, a dependência dos benefícios sociais diminui, conduzindo a taxas mais elevadas de descumprimento ou de abandono do programa. Com base nas análises fornecidas, é possível observar que SERRP se destaca com uma tendência de maior suspensão do benefício em famílias situadas em faixas de PIB mais altas, independentemente do tipo de associação avaliada.

As análises das associações entre as variáveis utilizando o teste de independência foram complementadas com a medida de força associação utilizando o V de Cramer permitiu avaliar diversas associações estatísticas e suas relevâncias no que se refere ao contexto de exclusão educacional e aos efeitos de múltiplas variáveis socioeconômicas e familiares na suspensão dos benefícios. Essa medida ajusta o valor do χ^2 ao tamanho da amostra e às dimensões da tabela, fornecendo uma indicação mais precisa da força da associação. Especificamente, o V de Cramer é preferido pela sua consistência em tabelas de diferentes dimensões, fornecendo uma medida padronizada da força da associação. A estratificação é utilizada para controlar o desequilíbrio da amostra, permitindo uma compreensão mais clara de como o status de monoparental impacta os resultados educacionais.

Primeiramente, a análise dos agregados familiares monoparentais revela associações

significativas com a exclusão escolar. A análise bivariada quando controlada pela monoparentalidade mostra variações importantes nas taxas de suspensão entre regiões (SERRP e NÃO-SERRP), onde vê-se que de maneira geral, a força das associações é maior quando controlada pela monoparentalidade onde apresentam maior prevalência de suspensões PBF, enquanto estão sob influência de precariedades mais marcantes como abastecimento de água por meio de poço ou nascente e escoamento sanitário via fossa rudimentar. Desta forma, a hipótese H1 (a exclusão escolar não está associada à existência de monoparentalidade na família) foi rejeitada.

Em seguida, diversas formas de violência, incluindo taxas de suicídio, são analisadas quanto à sua associação com a suspensão de benefícios. A análise mostra que taxas mais elevadas de violência na região SERRP estão fortemente relacionadas com taxas mais elevadas de suspensão. Valores elevados de V de Cramer sugerem associações fortes, particularmente na região SERRP, onde as taxas de suspensão são mais altas. Aqui constata-se uma exceção nas forças de associação, onde o suicídio possui força pequena entre os beneficiários não-suspensos. A violência dos tipos “Mortes Violentas” e “Homicídio de Homens” tem maior força de associação entre os suspensos. O “Homicídio de Mulheres” tem forte associação em todos os grupos (suspensos e não-suspensos). Desta forma, a hipótese H4 (a exclusão escolar não está associada às famílias residirem em localidades com maior índice de violência) foi rejeitada.

A compreensão da interação entre fatores sociais é examinada via análise bivariada de PIB per capita e violência, mostrando como eles influenciam as taxas de exclusão educacional em cada região, isoladamente. A violência mostrou relações moderadas a fortes com a suspensão de benefício, se destacando na população suspensa com relação mais forte para todos os tipos de violência. Também foram evidenciadas associações mais fortes em SERRP onde o PIB per capita é mais elevado e a violência é mais prevalente. Ao separar as regiões para análise das associações, foi possível diagnosticar que o todos os tipos de violência são mais significativamente associados aos altos valores de PIB per capita em SERRP, com V de Cramer muito mais altos do que aqueles encontrados em NÃO-SERRP. Desta forma, a hipótese H2 (a exclusão escolar não está associada às famílias residirem em localidades com maiores PIB per capita municipal) foi rejeitada.

A associação entre perfil familiar e exclusão escolar é compreendida pela análise bivariada, composta por características da infraestrutura domiciliar e por características pessoais dos membros da família do beneficiário. A análise revela que fatores estruturais, tais como escoamento sanitário do tipo fossa rudimentar afeta mais fortemente os suspensos, em ambas as regiões. A raça-cor aparece com maior grau de força na categoria “branca” seguido pela categoria “parda”. Também atuam de forma moderada para o aumento das faltas escolares a iluminação do domicílio do tipo comunitária, o material

predominante nas paredes do tipo taipa não revestida, o abastecimento de água do tipo cisterna, a quantidade de cômodos da casa inferior a 3 somente quando suspenso e quantidade de dormitórios ≤ 1 . Esta análise permite rejeitar a H3 (a exclusão escolar não está associada às famílias possuírem infraestrutura domiciliar mais precária ou terem membros em situação mais vulnerável), pois o modelo revela que domicílios menos robustos e com membros brancos ou pardos são mais determinantes em regiões com taxas de suspensão mais elevadas.

O estudo dos fatores socioeconômicos foi detalhado por monoparentalidade a fim de avaliar o comportamento das variáveis de PIB per capita e violência com a exclusão escolar, avaliando os efeitos para duas macrorregiões SERRP e NÃO-SERRP, separadamente, cada macrorregião agrupando um conjunto de estados segundo os padrões de faltas encontrados na bases de dados. Foi possível constatar que as famílias monoparentais suspensas são mais afetadas pela violência quando associadas à maiores PIB per capita. No grupo de suspenso, as associações são mais fortes para todos os tipos de violência, em especial nas áreas mais ricas. Aqui, SERRP também se destaca com graus de associação maiores em todos os modelos de relações. Isso permitiu identificar que, para os grupos monoparentais residentes de regiões mais desenvolvidas, as condições socioeconômicas locais são mais determinantes para a suspensão, esclarecendo como a monoparentalidade afeta a suspensão a medida que mais variáveis são acrescentadas ao modelo.

Por fim, o perfil familiar é analisado novamente, agora sob a ótica da monoparentalidade, mostrando que a relação entre seus componentes e a suspensão aparece mais forte no grupo monoparental. Neste modelo, foi possível entender melhor a característica pessoal raça-cor branca que se mostra mais fortemente associada aos não-monoparentais. A infraestrutura domiciliar de escoamento sanitário do tipo fossa rudimentar dos domicílios suspensos apresenta maior grau de associação em famílias monoparentais. Adicionalmente, foi possível constatar que as famílias monoparentais também apresentam um V de Cramer maior em condições de moradia mais precárias como quando a iluminação do domicílio é do tipo comunitária, o material predominante nas paredes é do tipo taipa não revestida, além de ser o único grupo associado ao tipo de abastecimento de água via poço ou nascente. A quantidade de cômodos da casa inferior a 3 afeta de forma significativa apenas o grupo monoparental e quantidade de dormitórios ≤ 1 **não** afeta os não-monoparentais não-suspensos. Os dados revelam que as infra-estruturas mais precarizadas são mais associadas às regiões com taxas de suspensão mais elevadas com destaque nas famílias monoparentais, sugerindo uma ligação direta entre a monoparentalidade com infraestruturas domiciliares mais pobres e maiores incidências de exclusão educacional.

As Figuras 5.1a, 5.1b, 5.2a,5.2b, 5.3a, 5.3b,5.4a e 5.4b apresentam um resumo sobre aspectos da exclusão escolar que decorrem em suspensão de benefício vistos ao longo desse

trabalho.

VARIÁVEIS COM GRANDE ASSOCIAÇÃO COM A SUSPENSÃO				
CARACTERÍSTICAS PESSOAIS DOS MEMBROS DA FAMÍLIA	SERRP		NÃO-SERRP	
	SUSPENSO	NÃO-SUSPENSO	SUSPENSO	NÃO-SUSPENSO
Raça-Cor (Branca; Preta; Amarela; Parda; Indígena;)	Branco	Branco	Branco	Branco
INFRAESTRUTURA DOMICILIAR				
Escoamento Sanitário (Rede coletora de esgoto ou pluvial; Fossa séptica; Fossa rudimentar; Vala a céu aberto; Direto do público um rio, lago ou mar; Outra forma)	Fossa rudimentar	Rede coletora de esgoto	Fossa rudimentar	Rede coletora de esgoto
VIOLÊNCIA				
Taxa de Suicídio (Média Municipal)	[3,20 ; 3,28]	PEQUENO	[2,87 ; 3,41]	PEQUENO
Taxa de Morte Violenta por Causa Indeterminada	[8,29 ; 9,05]	MODERADO	[4,15 ; 5,81]	MODERADO
Taxa de Homicídio de Mulheres	[1,39 ; 1,49]	[6,59 ; 33,13]	[2,17 ; 3,21]	[2,17 ; 3,21]
Taxa de Homicídio de Homens	[84,27 ; 151,45]	MODERADO	[104,80 ; 212,14]	MODERADO

(a) Fatores Fortemente Associados à Exclusão Escolar dos Beneficiários PBF

VARIÁVEIS COM MODERADA ASSOCIAÇÃO COM A SUSPENSÃO				
INFRAESTRUTURA DOMICILIAR	SERRP		NÃO-SERRP	
	SUSPENSO	NÃO-SUSPENSO	SUSPENSO	NÃO-SUSPENSO
Iluminação (Elétrica com medidor próprio; Elétrica com medidor comunitário; Elétrica sem medidor; Óleo ou querosene ou gás; Vela; Outra forma)	Elétrica com medidor comunitário			
Material Paredes Domicílio (Alvenaria/tijolo com revestimento; Alvenaria/tijolo sem revestimento; Madeira aparelhada; Taipa revestida; Taipa não revestida; Madeira aproveitada; Palha; Outro Material)	Taipa não revestida	Taipa não revestida	Taipa não revestida	Taipa não revestida
Abastecimento de Água (Rede geral de distribuição; Poço ou nascente; Cisterna; Outra forma)	Cisterna	Cisterna	Cisterna	Cisterna
Quantidade de Cômodos no Domicílio (4 faixas EF)	Menor ou igual a 3 cômodos	PEQUENO	Menor ou igual a 3 cômodos	PEQUENO
Quantidade de Dormitórios no Domicílio (4 faixas EF)	Menor ou igual a 1 dormitório			
VIOLÊNCIA				
Taxa de Morte Violenta por Causa Indeterminada	GRANDE	[4,15 ; 5,81]	GRANDE	[8,29 ; 9,05]
Taxa de Homicídio de Homens	GRANDE	[27,90 ; 44,76]	GRANDE	[62,51 ; 76,07]

(b) Fatores Moderadamente Associados à Exclusão Escolar dos Beneficiários PBF

Figura 5.1: Relação de Efeitos de Características Pessoais de Membros da Família, Infra-estrutura Familiar e Violência

ATRIBUTO	VARIÁVEIS COM GRANDE ASSOCIAÇÃO COM A SUSPENSÃO							
	SERRP				NÃO-SERRP			
	SUSPENSO		NÃO-SUSPENSO		SUSPENSO		NÃO-SUSPENSO	
PERFIL DOS MEMBROS DA FAMÍLIA	Monoparental	Não-Monoparental	Monoparental	Não-Monoparental	Monoparental	Não-Monoparental	Monoparental	Não-Monoparental
Raça-Cor (Branca; Preta; Amarela; Parda; Indígena;)	Branco	Branco	Branco	Branco	Branco	Branco	Branco	Branco
INFRAESTRUTURA DOMICILIAR								
Escoamento Sanitário (Rede coletora de esgoto ou pluvial; Fossa séptica; Fossa rudimentar; Vala a céu aberto; Direto para um rio, lago ou mar; Outra forma)	Fossa Rudimentar	Fossa Rudimentar	Rede Coletora de Esgoto	Rede Coletora de Esgoto	Fossa Rudimentar	Fossa Rudimentar	Rede Coletora de Esgoto	Rede Coletora de Esgoto
VIOLÊNCIA								
Taxa de Suicídio (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	[3,20; 3,28]	MEDIO	PEQUENO	PEQUENO	[2,87; 3,41]	MEDIO	PEQUENO	PEQUENO
Taxa de Morte Violenta por Causa Indeterminada (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	[8,29; 9,05]	[8,29;9,05]	[9,05; 10,15]	MEDIO	[4,15; 5,81]	[4,15; 5,81]	[5,81; 7,67]	MEDIO
Taxa de Homicídio de Mulheres (Média Municipal - 9 faixas NÃO-SERRP e 10 faixas SERRP)	[1,39; 1,49]	[6,59;33,13]	[6,59; 33,13]	[6,59; 33,13]	[2,17; 3,21]	[2,17; 3,21]	[2,17; 3,21]	[2,17; 3,21]
Taxa de Homicídio de Homens (Média Municipal - 8 faixas NÃO-SERRP e 8 faixas SERRP)	[84,27; 151,45]	[84,27;151,45]	MEDIO	[27,90;44,76]	[104,80; 212,14]	[104,80; 212,14]	MEDIO	[62,51; 76,07]

(a) Fatores Fortemente Associados à Exclusão Escolar dos Beneficiários PBF

ATRIBUTO	VARIÁVEIS COM MODERADA ASSOCIAÇÃO COM A SUSPENSÃO							
	SERRP				NÃO-SERRP			
	SUSPENSO		NÃO-SUSPENSO		SUSPENSO		NÃO-SUSPENSO	
PERFIL DOS MEMBROS DA FAMÍLIA	Monoparental	Não-Monoparental	Monoparental	Não-Monoparental	Monoparental	Não-Monoparental	Monoparental	Não-Monoparental
INFRAESTRUTURA DOMICILIAR								
Iluminação (Elétrica com medidor próprio; Elétrica com medidor comunitário; Elétrica sem medidor; Óleo ou querosene ou gás; Vela; Outra forma)	Elétrica com medidor comunitário	Elétrica com medidor comunitário	Elétrica com medidor comunitário	Elétrica com medidor comunitário	Elétrica com medidor comunitário	Elétrica com medidor comunitário	Elétrica com medidor comunitário	Elétrica com medidor comunitário
Material Paredes Domicílio (Alvenaria/tijolo com revestimento; Alvenaria/tijolo sem revestimento; Madeira aparelhada; Taipa revestida; Taipa não revestida; Madeira aproveitada; Palha; Outro Material)	Taipa não revestida	Taipa não revestida	Taipa não revestida	Taipa não revestida	Taipa não revestida	Taipa não revestida	Taipa não revestida	Taipa não revestida
Abastecimento de Água (Rede geral de distribuição; Poço ou nascente; Cisterna; Outra forma)	Cisterna	Cisterna	Poço ou nascente	Cisterna	Cisterna	Cisterna	Poço ou nascente	Cisterna
Quantidade de Cômodos no Domicílio (4 faixas EF)	Menor ou igual a 3 cômodos	PEQUENO	Menor ou igual a 3 cômodos	PEQUENO	Menor ou igual a 3 cômodos	PEQUENO	Menor ou igual a 3 cômodos	PEQUENO
Quantidade de Dormitórios no Domicílio (4 faixas EF)	Menor ou igual a 1 dormitório	Menor ou igual a 1 dormitório	Menor ou igual a 1 dormitório	PEQUENO	Menor ou igual a 1 dormitório	Menor ou igual a 1 dormitório	Menor ou igual a 1 dormitório	PEQUENO
VIOLÊNCIA								
Taxa de Suicídio (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	GRANDE	[3,20;3,28]	PEQUENO	PEQUENO	GRANDE	[2,87; 3,41]	PEQUENO	PEQUENO
Taxa de Morte Violenta por Causa Indeterminada (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	GRANDE	GRANDE	GRANDE	[8,29; 9,05]	GRANDE	GRANDE	GRANDE	[4,15; 5,81]
Taxa de Homicídio de Homens (Média Municipal - 8 faixas NÃO-SERRP e 8 faixas SERRP)	GRANDE	GRANDE	Faixa de Violência 5	GRANDE	GRANDE	GRANDE	Faixa de Violência 5	GRANDE

(b) Fatores Moderadamente Associados à Exclusão Escolar dos Beneficiários PBF

Figura 5.2: Relação de Efeitos de Características Pessoais de Membros da Família, Infra-estrutura Familiar e Violência Controlado para Monoparentalidade

VIOLÊNCIA	VARIÁVEIS COM GRANDE ASSOCIAÇÃO COM A SUSPENSÃO					
	NÃO-SERRP					
	SUSPENSO			NÃO-SUSPENSO		
	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3
Taxa de Suicídio	[0,00 ; 1,97]	[0,00 ; 1,97]	[0,00 ; 1,97]	MODERADO	MODERADO	MODERADO
Taxa de Morte Violenta por Causa Indeterminada	[0,00 ; 1,01]	[8,29 ; 9,05]	[12,17 ; 33,47]	[12,17 ; 33,47]	[8,29 ; 9,05]	[12,17 ; 33,47]
Taxa de Homicídio de Mulheres	[0,00 ; 2,17]	[6,36 ; 7,35]	[0,00 ; 2,17]	[0,00 ; 2,17]	[9,00 ; 69,55]	[0,00 ; 2,17]
Taxa de Homicídio de Homens	[104,80 ; 212,14]	[18,60 ; 29,35]	[104,80 ; 212,14]	[104,80 ; 212,14]	[62,51 ; 76,07]	[104,80 ; 212,14]
VIOLÊNCIA	SERRP					
	SUSPENSO			NÃO-SUSPENSO		
	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3
Taxa de Suicídio	[0,00 ; 2,68]	[3,20 ; 3,28]	[0,00 ; 2,68]	[0,00 ; 2,68]	[3,20 ; 3,28]	[3,20 ; 3,28]
Taxa de Morte Violenta por Causa Indeterminada	[4,15 ; 5,81]	[5,81 ; 7,67]	[5,81 ; 7,67]	[4,15 ; 5,81]	[5,81 ; 7,67]	[5,81 ; 7,67]
Taxa de Homicídio de Mulheres	[1,49 ; 1,93]	[1,49 ; 1,93]	[1,49 ; 1,93]	[1,49 ; 1,93]	[1,93 ; 2,64]	[1,49 ; 1,93]
Taxa de Homicídio de Homens	[69,55 ; 84,27]	[17,78 ; 22,61]	[17,78 ; 22,61]	[69,55 ; 84,27]	[69,55 ; 84,27]	[17,78 ; 22,61]

(a) Fatores Fortemente Associados à Exclusão Escolar dos Beneficiários PBF

VIOLÊNCIA	VARIÁVEIS COM MODERADA ASSOCIAÇÃO COM A SUSPENSÃO					
	NÃO-SERRP					
	SUSPENSO			NÃO-SUSPENSO		
	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3
Taxa de Suicídio	GRANDE	GRANDE	GRANDE	[0,00 ; 1,97]	[4,03 ; 4,83]	[0,00 ; 1,97]

(b) Fatores Moderadamente Associados à Exclusão Escolar dos Beneficiários PBF

Figura 5.3: Relação de Efeitos de PIB Per Capita e Violência.

ATRIBUTO PIB E VIOLÊNCIA	VARIÁVEIS COM GRANDE ASSOCIAÇÃO COM A SUSPENSÃO											
	SERRP						NÃO-SERRP					
	SUSPENSO			NÃO-SUSPENSO			SUSPENSO			NÃO-SUSPENSO		
	Monoparental	Não-Monoparental		Monoparental	Não-Monoparental		Monoparental	Não-Monoparental		Monoparental	Não-Monoparental	
PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	
Taxa de Suicídio (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	[0,00 ; 1,97]	[2,87 ; 3,41]	[2,87 ; 3,41]	[2,87 ; 3,41]	[2,87 ; 3,41]	[2,87 ; 3,41]	[0,00 ; 1,97]	[2,87 ; 3,41]	[2,87 ; 3,41]	[0,00 ; 1,97]	[2,87 ; 3,41]	[2,87 ; 3,41]
Taxa de Morte Violenta por Causa Indeterminada (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	[4,15 ; 5,81]	[5,81 ; 7,67]	[5,81 ; 7,67]	[5,81 ; 7,67]	[5,81 ; 7,67]	[5,81 ; 7,67]	[4,15 ; 5,81]	[5,81 ; 7,67]	[5,81 ; 7,67]	[5,81 ; 7,67]	[5,81 ; 7,67]	[5,81 ; 7,67]
Taxa de Homicídio de Mulheres (Média Municipal - 9 faixas NÃO-SERRP)	[3,21 ; 3,95]	[3,21 ; 3,95]	[3,21 ; 3,95]	[3,21 ; 3,95]	[3,21 ; 3,95]	[3,21 ; 3,95]	[3,21 ; 3,95]	[3,95 ; 4,75]	[3,21 ; 3,95]	[3,21 ; 3,95]	[3,95 ; 4,75]	[3,21 ; 3,95]
Taxa de Homicídio de Homens (Média Municipal - 8 faixas NÃO-SERRP e 8 faixas SERRP)	[29,35 ; 49,29]	[29,35 ; 49,29]	[29,35 ; 49,29]	[29,35 ; 49,29]	[29,35 ; 49,29]	[29,35 ; 49,29]	[90,14 ; 104,80]	[90,14 ; 104,80]	[29,35 ; 49,29]	[90,14 ; 104,80]	[90,14 ; 104,80]	[29,35 ; 49,29]
ATRIBUTO PIB E VIOLÊNCIA	VARIÁVEIS COM MODERADA ASSOCIAÇÃO COM A SUSPENSÃO											
	SERRP						NÃO-SERRP					
	SUSPENSO			NÃO-SUSPENSO			SUSPENSO			NÃO-SUSPENSO		
	Monoparental	Não-Monoparental		Monoparental	Não-Monoparental		Monoparental	Não-Monoparental		Monoparental	Não-Monoparental	
PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	
Taxa de Suicídio (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	[0,00 ; 2,68]	[4,49 ; 5,17]	[4,49 ; 5,17]	[0,00 ; 2,68]	[4,49 ; 5,17]	[4,49 ; 5,17]	[0,00 ; 2,68]	[3,65 ; 4,49]	[0,00 ; 2,68]	MEDIO	MEDIO	MEDIO
Taxa de Morte Violenta por Causa Indeterminada (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	[0,00 ; 1,01]	[8,29 ; 9,05]	[4,78 ; 6,26]	[0,00 ; 1,01]	[1,01 ; 1,84]	[0,00 ; 1,01]	[0,00 ; 1,01]	[8,29 ; 9,05]	[12,17 ; 33,47]	MEDIO	MEDIO	MEDIO
Taxa de Homicídio de Mulheres (Média Municipal - 10 faixas SERRP)	[0,00 ; 1,39]	[5,48 ; 6,59]	[0,00 ; 1,39]	[0,00 ; 1,39]	[3,57 ; 4,31]	[0,00 ; 1,39]	[0,00 ; 1,39]	[6,59 ; 33,13]	[0,00 ; 1,39]	[0,00 ; 1,39]	[6,59 ; 33,13]	[0,00 ; 1,39]
Taxa de Homicídio de Homens (Média Municipal - 8 faixas NÃO-SERRP e 8 faixas SERRP)	[84,27 ; 151,45]	[27,90 ; 44,76]	[84,27 ; 151,45]	[84,27 ; 151,45]	[27,90 ; 44,76]	[84,27 ; 151,45]	[84,27 ; 151,45]	[27,90 ; 44,76]	[84,27 ; 151,45]	[84,27 ; 151,45]	[13,73 ; 17,78]	[84,27 ; 151,45]

(a) Fatores Fortemente Associados à Exclusão Escolar dos Beneficiários PBF

ATRIBUTO PIB E VIOLÊNCIA	VARIÁVEIS COM MODERADA ASSOCIAÇÃO COM A SUSPENSÃO											
	NÃO-SERRP											
	SUSPENSO						NÃO-SUSPENSO					
	Monoparental	Não-Monoparental		Monoparental	Não-Monoparental		Monoparental	Não-Monoparental		Monoparental	Não-Monoparental	
PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	PIB 1	PIB 2	PIB 3	
Taxa de Suicídio (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	[0,00 ; 1,97]	[4,03 ; 4,83]	[0,00 ; 1,97]
Taxa de Morte Violenta por Causa Indeterminada (Média Municipal - 10 faixas SERRP e NÃO-SERRP)	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	GRANDE	[0,00 ; 0,29]	[0,29 ; 0,90]	[10,49 ; 91,37]

(b) Fatores Moderadamente Associados à Exclusão Escolar dos Beneficiários PBF

Figura 5.4: Relação de Efeitos de PIB Per Capita Violência Controlado para Monoparentalidade

5.2 Trabalhos Futuros

Com base nos resultados desta pesquisa, que analisou dados históricos de famílias beneficiárias do Programa Bolsa Família (PBF) ao longo de seis anos (2014 a 2019), identificando as variáveis mais associadas à suspensão devido ao grande número de faltas escolares, sugere-se como desdobramento futuro a aplicação dessas variáveis em modelos de análise multivariada e preditiva. A consideração da dimensão temporal dos dados oferece oportunidades adicionais para aprofundar a compreensão dos fatores que influenciam a suspensão dos beneficiários e possibilita o desenvolvimento de estratégias mais assertivas de gestão do programa e políticas públicas direcionadas.

Um das abordagens é a construção de modelos de regressão logística multivariada que incorporem a variável temporal, permitindo quantificar o impacto individual de cada variável associada à suspensão ao longo do período estudado. Essa abordagem possibilita identificar tendências e mudanças nos fatores de risco, como monoparentalidade, condições de infraestrutura domiciliar, taxas de violência e PIB per capita municipal, oferecendo *insights* sobre como essas variáveis influenciam o risco de exclusão escolar ao longo do tempo. Além disso, modelos de classificação, como Random Forest, Gradient Boosting ou Redes Neurais, podem ser aplicados para prever com maior precisão quais famílias estão em maior risco de suspensão, criando ferramentas preditivas para intervenções preventivas. O modelo Decision Tree, em especial, apresenta a vantagem de ser altamente interpretável, permitindo identificar os critérios de decisão mais relevantes de forma clara e intuitiva, o que pode ser extremamente útil para otimizar a orientação aos gestores do programa.

Dado o caráter longitudinal da base de dados, modelos de classificação temporal, como análises de séries temporais ou modelos de sobrevivência (survival analysis), podem ser utilizados para prever com maior precisão quais famílias estão em maior risco de suspensão em diferentes momentos. Isso permite o desenvolvimento de ferramentas preditivas que considerem a evolução das condições socioeconômicas e familiares, auxiliando na implementação de intervenções preventivas mais oportunas.

Ademais, explorar métodos estatísticos e computacionais que capturam relações não lineares e interações entre variáveis é outro caminho promissor. Exemplos incluem a aplicação de modelos bayesianos, que permitem incorporar incertezas nas análises, ou métodos de agrupamento (*clustering*), que podem identificar perfis ocultos entre os beneficiários suspensos e traçar estratégias diferenciadas de atendimento.

Por fim, estudos futuros podem expandir o escopo para incluir variáveis e contextos socioeconômicos que mudaram após 2019, analisando como fatores como flutuações do PIB per capita, mudanças em políticas públicas e novas dinâmicas de violência municipal afetam a suspensão de beneficiários, assim como investigar como esses eventos influenciam as taxas de suspensão em diferentes regiões (SERRP e NÃO-SERRP) e estratos popula-

cionais, permitindo a elaboração de políticas mais adaptadas às necessidades locais e o aperfeiçoamento das políticas de combate à exclusão escolar e à melhoria da eficácia do Programa Bolsa Família.

Referências

- [1] M. d. C. M. Senna, L. Burlandy, G. L. Monnerat, V. Schottz, and R. Magalhães, “Programa bolsa família: nova institucionalidade no campo da política social brasileira?” *Revista Katálisis*, vol. 10, no. 1, pp. 86–94, 2007. 5
- [2] “Long-term impacts of conditional cash transfers: Review of the evidence,” vol. 34, 6, 10
- [3] A. Banerji and U. Gentilini, “Social safety nets: Lessons from global evidence and practice,” in *Bank of Namibia’s Annual Symposium on Social Safety Nets in Namibia, Windhoek, September*, vol. 26, 2013. 9
- [4] A. Pires, “Afinal, para que servem as condicionalidades em educação do programa bolsa família?” *Ensaio: avaliação e políticas públicas em educação*, vol. 21, no. 80, pp. 513–531, 2013. 10
- [5] J. Das, Q.-T. Do, and B. Özler, “Reassessing conditional cash transfer programs,” *The World Bank Research Observer*, vol. 20, no. 1, pp. 57–80, 2005. 10
- [6] A. de Brauw, D. O. Gilligan, J. Hoddinott, and S. Roy, “The impact of bolsa família on schooling,” *World Development*, vol. 70, pp. 303–316, 2015. 10
- [7] L. H. Paiva, F. V. Soares, F. Cireno, I. A. V. Viana, and A. C. Duran, “The effects of conditionality monitoring on educational outcomes: Evidence from brazil’s bolsa família programme,” Discussion Paper, Tech. Rep., 2016. 11
- [8] E. F. d. L. Amaral and V. d. P. Monteiro, “Avaliação de impacto das condicionalidades de educação do programa bolsa família (2005 e 2009),” *Dados*, vol. 56, pp. 531–570, 2013. 11, 12
- [9] S. A. Reynolds, “Brazil’s bolsa familia: Does it work for adolescents and do they work less for it?” *Economics of Education Review*, vol. 46, pp. 23–38, 2015. 12
- [10] F. Cireno, J. Silva, and R. Proença, “Condicionalidades, desempenho e percurso escolar de beneficiários do programa bolsa família,” *Programa Bolsa Família: uma década de inclusão e cidadania. Brasília: Ipea*, pp. 297–304, 2013. 12
- [11] L. F. B. d. OLIVEIRA and S. S. D. Soares, “O impacto do programa bolsa família sobre a repetência: resultados a partir do cadastro único, projeto frequência e censo escolar,” 2013. 12

- [12] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Taylor & Francis, 2013. [Online]. Available: <https://books.google.com.br/books?id=cIJH0lR33bgC> 13, 17, 19, 20, 21, 25, 26, 68
- [13] P. C. F. Assis. Janilson P.; Sousa, Roberto P.; Linhares, *Testes de Hipóteses Estatísticas*. EdUFERSA, 2020. [Online]. Available: <https://livraria.ufersa.edu.br/testes-de-hipoteses-estatisticas> 14, 18, 49
- [14] P. R. Freedman, D. and R. Purves, *Statistics*. W. W. Norton & Company, New York, 2007, vol. 4. 15, 16, 17
- [15] M. F. Triola, *Elementary Statistics*. Pearson, Boston, 2018, vol. 13. 15
- [16] L. Hogben, *Statistical Theory: The Relationship of Probability, Credibility and Error*. George Allen & Unwin, London, 1957. 15, 16
- [17] D. C. Howell, *Statistical Methods for Psychology*. Wadsworth, Belmont, 2012. 15, 16
- [18] W. G. Cochran, *Sampling Techniques*. John Wiley & Sons, New York., 1977, vol. 3. 16
- [19] K. L. Delucchi, “The use and misuse of chi-square: Lewis and burke revisited.” *Psychological Bulletin*, vol. 94, no. 1, p. 166, 1983. 17
- [20] “Biometrika Tables for Statisticians,” *Journal of the Royal Statistical Society Series D: The Statistician*, vol. 10, no. 1, pp. 52–53, 12 2018. [Online]. Available: <https://doi.org/10.2307/2987487> 18
- [21] H. BLALOCK, *SOCIAL STATISTICS SECOND EDITION*, 1972. [Online]. Available: <https://books.google.com.br/books?id=4wsmwPxU1toC> 21, 26
- [22] W. L. Hays, *Statistics*. New York,: Harcourt Brace College Publishers, 1994. 21, 25, 26
- [23] A. Edwards, “Experimental design in psychological research, rev,” 1960. 21
- [24] J. P. Guilford, “Fundamental statistics in psychology and education,” 1950. 25
- [25] “Data mining: Data mining concepts and techniques,” in *2013 International Conference on Machine Intelligence and Research Advancement*. 27
- [26] P. Chapman, *CRISP-DM 1.0: Step-by-step Data Mining Guide*. SPSS, 2000. [Online]. Available: <https://books.google.com.br/books?id=po7FtgAACAAJ> 27, 29, 32
- [27] U. M. Fayyad and K. B. Irani, “Multi-interval discretization of continuous-valued attributes for classification learning,” in *Ijcai*, vol. 93, no. 2. Citeseer, 1993, pp. 1022–1029. 28, 30, 46

- [28] R. Dash, R. L. Paramguru, and R. Dash, “Comparative analysis of supervised and unsupervised discretization techniques,” *International Journal of Advances in Science and Technology*, vol. 2, no. 3, pp. 29–37, 2011. 29, 31
- [29] S. García, J. Luengo, J. A. Sáez, V. López, and F. Herrera, “A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 734–750, 2013. 29
- [30] M. HACIBEYOĞLU and M. H. IBRAHIM, “Comparison of the effect of unsupervised and supervised discretization methods on classification process,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 4, no. Special Issue-1, pp. 105–108, 2016. 29
- [31] F. J. Ruiz, C. Angulo, and N. Agell, “Idd: A supervised interval distance-based method for discretization,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 9, pp. 1230–1238, 2008. 29
- [32] K. Ho and P. Scott, “Zeta: a global method for discretization of continuous variables,” in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 1997, pp. 191–194. 29
- [33] M. ML, “The chi-square test of independence,” vol. 329(2). PMC3900058, 2013. 29, 45
- [34] G. Agre and S. Peev, “On supervised and unsupervised discretization,” *Cybernetics and information technologies*, vol. 2, no. 2, pp. 43–57, 2002. 29, 46, 47
- [35] R. Wirth and J. Hipp, “Crisp-dm: Towards a standard process model for data mining,” in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 2000, pp. 29–39. 34
- [36] A. Kadadi, R. Agrawal, C. Nyamful, and R. Atiq, “Challenges of data integration and interoperability in big data,” in *2014 IEEE International Conference on Big Data (Big Data)*, 2014, pp. 38–40. 37
- [37] P. Bussab, WO; Morettin, *Estatística Básica*. Editora Saraiva, 2003. 48
- [38] L. Castro, Jorge Abrahão de e Modesto, *Bolsa Família 2003-2010: Avanços e Desafios*. Instituto de Pesquisa Econômica Aplicada (Ipea), 2010. [Online]. Available: <https://repositorio.ipea.gov.br/handle/11058/3248> 56
- [39] V. M. Campos and G. S. L. d. Resende, “A desestruturação familiar e o adolescente em conflito com a lei: Pontos e contrapontos,” *Revista Eletrônica do Curso de Direito da UFSM*, vol. 11, no. 1, p. 365–390, jun. 2016. [Online]. Available: <https://periodicos.ufsm.br/revistadireito/article/view/21935> 57
- [40] A. M. Liebetrau, *Measures of association*. Sage, 1983, vol. 32. 67

Anexo I

Indicadores de Faltas

Tabela I.1: Distribuição Nacional do Público Suspenso e do Público Beneficiário

Estado	Beneficiários (%)	Beneficiários Suspendos (%)
Acre	0,78	0,18
Alagoas	2,73	0,92
Amapá	0,61	0,07
Amazonas	3,31	2,41
Bahia	10,58	5,78
Ceará	6,48	3,59
Distrito Federal	0,80	0,46
Espírito Santo	1,58	2,46
Goiás	2,87	1,40
Maranhão	6,71	1,14
Mato Grosso	1,55	0,98
Mato Grosso do Sul	1,19	1,06
Minas Gerais	8,44	8,24
Pará	7,00	1,69
Paraíba	2,90	1,32
Paraná	3,40	5,66
Pernambuco	6,76	3,84
Piauí	2,67	0,36
Rio de Janeiro	6,16	10,20
Rio Grande do Norte	2,19	1,04
Rio Grande do Sul	3,26	5,33
Rondônia	0,89	0,42
Roraima	0,45	0,13
Santa Catarina	1,33	0,91
São Paulo	12,63	39,01
Sergipe	1,64	1,04
Tocantins	1,09	0,37
Total geral	100,00	100,00

Tabela I.2: Estatística Descritiva da Discretização por Entropia MDL das Taxas de Suicídio

				N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo
		Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total
Taxa Média Municipal de Suicídio de 2014 a 2019	NAO-SERP	1	0,00	1,97	6.210.637	98,8%	73.041	1,2%	6.283.678
		2	1,97	2,87	6.259.152	98,3%	107.059	1,7%	6.366.211
		3	2,87	3,41	6.217.029	98,5%	92.857	1,5%	6.309.886
		4	3,41	4,03	6.225.433	98,4%	99.575	1,6%	6.325.008
		5	4,03	4,83	6.249.924	98,7%	79.310	1,3%	6.329.234
		6	4,83	5,49	6.186.441	97,7%	143.514	2,3%	6.329.955
		7	5,49	6,26	6.212.299	98,5%	94.980	1,5%	6.307.279
		8	6,26	7,47	6.257.843	98,6%	90.999	1,4%	6.348.842
		9	7,47	9,48	6.224.714	98,4%	98.805	1,6%	6.323.519
		10	9,48	44,10	6.223.681	98,4%	102.850	1,6%	6.326.531
		Total			62.267.153	98,4%	982.990	1,6%	63.250.143
		N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo		
		Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total
	SERRP	1	0,00	2,68	1.798.105	95,8%	78.218	4,2%	1.876.323
		2	2,68	3,20	1.100.599	93,3%	79.667	6,7%	1.180.266
		3	3,20	3,28	2.409.329	89,7%	276.900	10,3%	2.686.229
		4	3,28	3,65	1.777.012	93,0%	134.060	7,0%	1.911.072
		5	3,65	4,49	1.812.458	93,4%	127.989	6,6%	1.940.447
		6	4,49	5,17	1.631.813	91,7%	147.008	8,3%	1.778.821
		7	5,17	6,16	1.920.055	92,8%	149.202	7,2%	2.069.257
		8	6,16	7,31	1.761.572	92,1%	151.866	7,9%	1.913.438
		9	7,31	10,01	1.777.619	92,2%	149.424	7,8%	1.927.043
		10	10,01	59,08	1.826.123	94,7%	101.862	5,3%	1.927.985
		Total			17.814.685	92,7%	1.396.196	7,3%	19.210.881

Tabela I.3: Estatística Descritiva da Discretização por Entropia MDL das taxas de Morte Indeterminada

			N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo	
Taxa Média Municipal de Morte Indeterminada de 2014 a 2019	NAO-SERP		63.250.143	0,00	91,37	2.150	10	0,115	
		Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total
		1	0,00	0,29	6.153.275	98,9%	66.523	1,1%	6.219.798
		2	0,29	0,90	6.312.554	98,4%	101.661	1,6%	6.414.215
		3	0,90	1,48	6.255.358	98,7%	80.683	1,3%	6.336.041
		4	1,48	2,13	5.865.248	98,7%	80.058	1,3%	5.945.306
		5	2,13	2,97	6.492.842	98,6%	95.164	1,4%	6.588.006
		6	2,97	4,15	6.354.284	98,6%	90.146	1,4%	6.444.430
		7	4,15	5,81	6.197.980	98,4%	97.786	1,6%	6.295.766
		8	5,81	7,67	6.229.121	98,0%	124.868	2,0%	6.353.989
		9	7,67	10,49	5.164.540	98,2%	96.875	1,8%	5.261.415
10	10,49	91,37	7.241.951	98,0%	149.226	2,0%	7.391.177		
	Total		62.267.153	98,4%	982.990	1,6%	63.250.143		
			N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo	
			19.210.881	0,00	33,47	998	10	0,374	
	Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total	
	1	0,00	1,01	1.802.388	94,0%	114.257	6,0%	1.916.645	
	2	1,01	1,84	1.787.891	93,3%	127.410	6,7%	1.915.301	
	3	1,84	3,32	1.795.187	93,1%	132.124	6,9%	1.927.311	
	4	3,32	4,78	1.776.800	92,3%	148.099	7,7%	1.924.899	
	5	4,78	6,26	1.668.716	92,5%	134.916	7,5%	1.803.632	
	6	6,26	8,29	1.881.646	92,8%	146.792	7,2%	2.028.438	
	7	8,29	9,05	699.223	95,1%	35.650	4,9%	734.873	
	8	9,05	10,15	2.762.140	90,1%	304.588	9,9%	3.066.728	
	9	10,15	12,17	891.965	92,4%	72.991	7,6%	964.956	
	10	12,17	33,47	2.748.729	93,9%	179.369	6,1%	2.928.098	
		Total		17.814.685	92,7%	1.396.196	7,3%	19.210.881	

Tabela I.4: Estatística Descritiva da Discretização por Entropia MDL das Taxas de Homicídio de Mulheres

			N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo	
Taxa Média Municipal de Homicídio de Mulheres de 2014 a 2019	NAO-SERP		63.250.143	0,00	44,10	3.080	10	0,115	
		Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total
		1	0,00	2,17	12.318.434	98,6%	169.723	1,4%	12.488.157
		2	2,17	3,21	6.370.069	98,3%	108.026	1,7%	6.478.095
		3	3,21	3,95	6.211.210	98,3%	109.360	1,7%	6.320.570
		4	3,95	4,75	6.261.572	98,8%	76.614	1,2%	6.338.186
		5	4,75	5,55	6.175.750	98,5%	94.381	1,5%	6.270.131
		6	5,55	6,36	6.070.745	98,2%	114.043	1,8%	6.184.788
		7	6,36	7,35	6.388.999	98,1%	123.996	1,9%	6.512.995
		8	7,35	9,00	6.222.812	98,5%	97.528	1,5%	6.320.340
		9	9,00	69,55	6.247.562	98,6%	89.319	1,4%	6.336.881
	Total		62.267.153	98,4%	982.990	1,6%	63.250.143		
			N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo	
			19.210.881	0,00	59,08	1.512	10	0,373	
	Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total	
	1	0,00	1,39	1.768.373	93,0%	133.957	7,0%	1.902.330	
	2	1,39	1,49	194.827	88,8%	24.504	11,2%	219.331	
	3	1,49	1,93	3.206.626	90,1%	353.182	9,9%	3.559.808	
	4	1,93	2,64	1.822.510	91,0%	179.821	9,0%	2.002.331	
	5	2,64	3,30	1.760.055	91,8%	157.038	8,2%	1.917.093	
	6	3,30	3,57	1.785.852	93,0%	135.297	7,0%	1.921.149	
	7	3,57	4,31	1.765.820	94,6%	100.458	5,4%	1.866.278	
	8	4,31	5,48	1.860.802	95,3%	91.783	4,7%	1.952.585	
	9	5,48	6,59	1.802.514	94,5%	105.179	5,5%	1.907.693	
	10	6,59	33,13	1.847.306	94,1%	114.977	5,9%	1.962.283	
		Total		17.814.685	92,7%	1.396.196	7,3%	19.210.881	

Tabela I.5: Estatística Descritiva da Discretização por Entropia MDL das Taxas de Homicídio de Homens

Taxa Média Municipal de Homicídio de Homens de 2014 a 2019			N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo
			63.250.143	0,00	212,14	3.568	8	0,116
	Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total
	1	0,00	18,60	6.213.553	98,4%	103.844	1,6%	6.317.397
	2	18,60	29,35	6.245.769	98,6%	86.181	1,4%	6.331.950
	3	29,35	49,29	12.453.026	98,4%	197.249	1,6%	12.650.275
	4	49,29	62,51	6.220.491	98,7%	84.651	1,3%	6.305.142
	5	62,51	76,07	6.105.664	98,5%	89.906	1,5%	6.195.570
	6	76,07	90,14	6.364.137	98,4%	105.137	1,6%	6.469.274
	7	90,14	104,80	5.737.192	98,1%	110.817	1,9%	5.848.009
8	104,80	212,14	12.927.321	98,4%	205.205	1,6%	13.132.526	
Total			62.267.153	98,4%	982.990	1,6%	63.250.143	
		N	Mínimo	Máximo	Número de Valores Distintos	Número de categorias	Entropia de modelo	
		19.210.881	0,00	151,45	1.523	8	0,116	
Bin	Inferior	Superior	Não	%Não	Sim	%Sim	Total	
1	0,00	13,73	1.753.231	91,4%	164.645	8,6%	1.917.876	
2	13,73	17,78	1.123.442	91,1%	109.484	8,9%	1.232.926	
3	17,78	22,61	4.046.411	90,6%	419.782	9,4%	4.466.193	
4	22,61	27,90	1.814.929	91,8%	161.155	8,2%	1.976.084	
5	27,90	44,76	3.583.458	93,5%	247.680	6,5%	3.831.138	
6	44,76	69,55	1.787.671	94,3%	107.251	5,7%	1.894.922	
7	69,55	84,27	1.833.067	96,1%	74.168	3,9%	1.907.235	
8	84,27	151,45	1.872.476	94,4%	112.031	5,6%	1.984.507	
Total			17.814.685	92,7%	1.396.196	7,3%	19.210.881	

Tabela I.6: Resumo da Discretização do Valor Médio de PIB Per Capital Municipal no Período de 2014 a 2019

Macrorregião	Bin	Terminal Inferior	Terminal Superior	Não Suspenso	Suspenso	Total	Entropia
NÃO-SERRP	1	3,85	9,19	20895838	197137	21092975	0,114
	2	9,19	16,95	20789107	275338	21064445	
	3	16,95	304,77	20620373	510939	21131312	
SERRP	1	7,63	26,09	6030870	344497	6375367	0,374
	2	26,09	49,68	5902541	519695	6422236	
	3	49,68	473,08	5881578	532044	6413622	

Figura I.1: Crosstab de Suicídio e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.

Crosstab NÃO-SERRP													MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE SUICÍDIO	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR	
			N	%	X2	N	%	X2	N	%	X2				
NÃO-SUSPENSO	NÃO	1	1811577	15,1%	206.816,62	1034103	10,0%	6.617,62	396579	5,2%	224.666,70	0,136	1.111.735,7650	0,000	
		2	1148622	9,6%	7.419,73	899910	8,7%	334,49	606826	7,9%	7.469,65				
		3	1091751	9,1%	6.267,52	812701	7,9%	4.577,35	631071	8,2%	411,50				
		4	976393	8,2%	3.168,37	701103	6,8%	41.870,07	911924	11,9%	95.114,56				
		5	970852	8,1%	37.613,68	1274659	12,3%	62.020,79	714822	9,3%	2.228,00				
		6	767374	6,4%	89.685,91	873994	8,4%	3.777,57	1060142	13,9%	198.865,50				
		7	973354	8,1%	49.399,37	1224951	11,8%	27.427,02	854825	11,2%	7.268,10				
		8	1327197	11,1%	765,06	1067390	10,3%	2.610,62	851426	11,1%	617,59				
		9	1438826	12,0%	2.754,58	1248857	12,1%	2.706,21	762556	10,0%	15.910,52				
		10	1453746	12,2%	1.479,08	1214003	11,7%	18,76	859841	11,2%	1.853,20				
	SIM	1	1520831	17,1%	604.123,93	987329	9,5%	840,89	460218	3,5%	448.862,73	0,183	2.155.662,4450	0,000	
		2	1012287	11,4%	375,38	1144404	11,0%	336,72	1447103	11,2%	0,17				
		3	883892	9,9%	16.788,88	1159589	11,1%	742,64	1638025	12,6%	17.370,68				
		4	804298	9,0%	38.957,57	877993	8,4%	74.903,55	1953722	15,1%	167.321,14				
		5	853693	9,6%	3.065,55	1367366	13,1%	87.344,73	1068532	8,2%	48.073,51				
		6	566982	6,4%	161.048,42	912933	8,7%	40.250,62	2005016	15,5%	262.621,24				
		7	657979	7,4%	51.871,14	1173824	11,2%	23.014,23	1327366	10,2%	2.766,80				
		8	904395	10,2%	6.692,18	905631	8,7%	4.660,47	1201804	9,3%	42,69				
		9	817042	9,2%	3.615,06	964137	9,2%	5.135,49	993296	7,7%	13.019,28				
		10	880055	9,9%	25.332,10	943154	9,0%	5.982,38	872882	6,7%	40.502,28				
SUSPENSO	NÃO	1	17114	15,7%	5.739,62	12950	9,7%	111,37	7336	4,1%	4.637,00	0,174	25.502,8080	0,000	
		2	10121	9,3%	1,80	11343	8,5%	59,63	17179	9,6%	31,38				
		3	10776	9,9%	462,31	11493	8,6%	58,56	11639	6,5%	544,51				
		4	8260	7,6%	70,71	9240	6,9%	300,81	17558	9,8%	461,34				
		5	9103	8,4%	8,03	13789	10,4%	832,90	11300	6,3%	731,65				
		6	6981	6,4%	3.445,79	12600	9,5%	1.128,53	34212	19,1%	5.571,13				
		7	8359	7,7%	642,19	14942	11,2%	164,04	19336	10,8%	75,94				
		8	11126	10,2%	0,07	12476	9,4%	96,20	19556	10,9%	74,78				
		9	12923	11,9%	24,51	15713	11,8%	24,16	19237	10,7%	65,41				
		10	14138	13,0%	0,00	18441	13,9%	79,67	22136	12,3%	58,77				
	SIM	1	14042	16,0%	12.852,78	11405	8,0%	623,30	10194	3,1%	5.585,33	0,201	45.378,4100	0,000	
		2	9642	11,0%	105,37	14396	10,1%	499,29	44378	13,4%	397,14				
		3	9591	10,9%	14,70	16292	11,4%	122,53	33066	10,0%	85,15				
		4	6029	6,9%	1.637,03	13055	9,2%	664,09	45433	13,7%	1.422,83				
		5	10355	11,8%	1.539,06	16058	11,3%	1.870,01	18705	5,6%	2.355,98				
		6	5834	6,6%	4.793,93	14873	10,4%	2.720,15	69014	20,8%	4.876,09				
		7	5995	6,8%	587,96	15035	10,6%	236,19	31313	9,4%	5,83				
		8	8103	9,2%	51,03	11503	8,1%	31,81	28235	8,5%	0,00				
		9	9133	10,4%	170,16	15180	10,7%	400,15	26619	8,0%	393,05				
		10	9143	10,4%	345,07	14529	10,2%	445,22	24463	7,4%	547,20				

Crosstab SERRP												MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE SUICÍDIO	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	NÃO	1	496249	19,0%	291.142,30	77582	3,3%	86.649,54	33309	1,9%	99.078,40	0,498	3.348.422,1700	
		2	126676	4,9%	3.102,54	177360	7,5%	13.460,67	78959	4,5%	4.470,38			
		3	4131	0,2%	246.082,86	7954	0,3%	215.733,64	645436	36,6%	1.304.784,82			
		4	68450	2,6%	64.460,69	68378	2,9%	51.478,19	314432	17,8%	327.247,93			
		5	198979	7,6%	18.140,57	282933	11,9%	6.010,43	213187	12,1%	5.457,21			
		6	274006	10,5%	3.033,87	255024	10,7%	4.170,99	108769	6,2%	20.143,32			
		7	258561	9,9%	10.392,39	389987	16,4%	36.621,16	168199	9,5%	9.610,76			
		8	222330	8,5%	8.629,11	384167	16,2%	77.211,42	93379	5,3%	43.850,80			
		9	428256	16,4%	37.549,27	348847	14,7%	11.891,10	47362	2,7%	131.198,40			
		10	530065	20,3%	63.925,39	380526	16,0%	4.486,01	59604	3,4%	148.408,00			
	SIM	1	1050539	30,7%	1.264.291,41	113205	3,2%	187.008,72	27221	0,7%	390.302,84	0,603	8.054.255,0240	
		2	197394	5,8%	2.700,59	400383	11,3%	128.751,09	119827	2,9%	81.098,33			
		3	2739	0,1%	536.168,30	10296	0,3%	538.051,55	1738773	42,2%	1.813.166,61			
		4	74420	2,2%	274.573,75	112740	3,2%	227.226,71	1138592	27,6%	844.473,85			
		5	286433	8,4%	10.088,91	492323	13,9%	52.020,22	338603	8,2%	14.294,93			
		6	417464	12,2%	39.463,92	353942	10,0%	4.335,03	222608	5,4%	58.587,77			
		7	278551	8,1%	11.478,58	541716	15,3%	102.633,71	283041	6,9%	39.555,49			
		8	197495	5,8%	52.091,47	726538	20,6%	444.988,19	137663	3,3%	167.634,57			
		9	476418	13,9%	112.049,79	423983	12,0%	47.493,92	52753	1,3%	256.942,02			
		10	441714	12,9%	118.482,56	354353	10,0%	24.339,73	59861	1,5%	209.960,45			
SUSPENSO	NÃO	1	21070	14,4%	20.635,40	4897	2,4%	3.470,20	1629	1,1%	5.427,34	0,525	275.037,2760	
		2	7014	4,8%	62,19	11837	5,9%	160,94	7329	4,9%	46,85			
		3	113	0,1%	21.333,59	644	0,3%	28.198,43	72484	48,0%	113.802,40			
		4	4032	2,8%	3.077,48	7464	3,7%	2.271,06	20488	13,6%	12.002,03			
		5	10300	7,0%	903,79	23126	11,5%	937,52	13579	9,0%	31,95			
		6	16436	11,2%	0,00	30543	15,2%	2.910,23	8832	5,8%	3.869,21			
		7	17227	11,8%	22,16	33235	16,6%	3.190,41	10199	6,8%	3.651,69			
		8	13000	8,9%	856,74	34294	17,1%	5.598,55	9756	6,5%	3.289,62			
		9	32124	21,9%	7.855,14	31301	15,6%	714,84	3460	2,3%	13.952,50			
		10	25272	17,2%	6.703,67	23053	11,5%	254,13	3256	2,2%	9.807,24			
	SIM	1	41850	21,1%	84.475,86	7192	2,3%	6.483,99	1580	0,4%	18.430,93	0,597	640.970,6500	
		2	11837	6,0%	0,23	29909	9,4%	6.246,27	11741	3,1%	5.283,46			
		3	85	0,0%	44.704,20	890	0,3%	70.620,42	202684	53,2%	156.524,61			
		4	4861	2,5%	13.819,93	14126	4,4%	13.530,11	83089	21,8%	36.556,49			
		5	16315	8,2%	131,01	41175	12,9%	5.332,72	23494	6,2%	3.433,46			
		6	25498	12,9%	1.453,17	47562	14,9%	7.077,40	18137	4,8%	10.916,08			
		7	21688	11,0%	243,37	48146	15,1%	8.834,74	18707	4,9%	9.463,55			
		8	13747	6,9%	2.443,38	67381	21,1%	33.656,53	13688	3,6%	17.504,62			
		9	37799	19,1%	21.150,09	40488	12,7%	4.237,56	4252	1,1%	27.026,74			
		10	24229	12,2%	15.608,63	22392	7,0%	1.143,09	3660	1,0%	14.638,04			

Figura I.2: Crosstab de Suicídio e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.

Crosstab NÃO-SERRP												MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE MORTE INDETERMINADA	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	NÃO	1	1992864	16,7%	224.250,97	1044186	10,1%	29.555,97	538227	7,0%	153.755,52	0,148	1.318.150,8170	0,000
		2	1276045	10,7%	4.761,98	787555	7,6%	60.866,19	943732	12,3%	40.282,30			
		3	1203283	10,1%	0,52	1066356	10,3%	560,63	746808	9,8%	709,66			
		4	1182951	9,9%	4.273,30	923096	8,9%	1.750,69	684632	8,9%	1.093,20			
		5	1262838	10,6%	625,10	1289117	12,5%	26.310,71	682876	8,9%	24.782,71			
		6	1387422	11,6%	13.057,81	980041	9,5%	11.071,86	787066	10,3%	419,30			
		7	1275954	10,7%	168,06	1284080	12,4%	33.864,70	600006	7,8%	53.026,64			
		8	912888	7,6%	29.226,86	974182	9,4%	917,46	847343	11,1%	31.870,22			
		9	789101	6,6%	28.420,14	997531	9,6%	35.848,40	602669	7,9%	89,50			
		10	676346	5,7%	199.682,39	1005527	9,7%	16,66	1216653	15,9%	306.891,37			
	SIM	1	1337489	15,0%	553.713,13	764259	7,3%	5.644,29	476250	3,7%	301.519,71	0,205	2.727.204,6470	0,000
		2	974789	11,0%	4.508,31	872572	8,4%	35.677,23	1457861	11,2%	12.954,56			
		3	829439	9,3%	4.447,70	1160139	11,1%	12.379,42	1249333	9,6%	1.985,55			
		4	957392	10,8%	14.343,57	887599	8,5%	11.235,46	1229578	9,5%	17,10			
		5	1002355	11,3%	12.199,76	1188338	11,4%	17.532,03	1067318	8,2%	44.223,34			
		6	1040206	11,7%	28.513,72	851437	8,2%	32.129,95	1308112	10,1%	436,88			
		7	928506	10,4%	9.988,68	1367464	13,1%	151.865,97	741970	5,7%	186.971,00			
		8	652354	7,3%	100.167,09	1028363	9,9%	8.964,31	1813991	14,0%	120.514,36			
		9	628191	7,1%	24.362,95	1198559	11,5%	101.742,37	948489	7,3%	24.595,36			
		10	550733	6,2%	348.747,04	1117630	10,7%	58.095,59	2675062	20,6%	497.728,20			
SUSPENSO	NÃO	1	16539	15,2%	4.644,95	12305	9,3%	9,93	9053	5,0%	3.113,61	0,188	29.686,6270	0,000
		2	8991	8,3%	456,48	9261	7,0%	1.464,39	25309	14,1%	2.458,31			
		3	10176	9,3%	151,03	8083	6,1%	774,53	16602	9,2%	206,87			
		4	11374	10,4%	670,23	9596	7,2%	156,55	13575	7,6%	88,28			
		5	15511	14,2%	1.308,37	17230	13,0%	655,30	12194	6,8%	2.520,99			
		6	10835	9,9%	2,59	11739	8,8%	127,60	18707	10,4%	71,72			
		7	13094	12,0%	92,32	18487	13,9%	974,10	15005	8,4%	1.179,87			
		8	7342	6,7%	1.744,17	15615	11,7%	81,22	23078	12,9%	613,70			
		9	8339	7,7%	507,02	15228	11,5%	373,36	17697	9,9%	0,82			
		10	6700	6,2%	3.074,04	15443	11,6%	13,71	28269	15,7%	2.150,57			
	SIM	1	11340	12,9%	10.511,56	9650	6,8%	791,01	7636	2,3%	5.072,50	0,236	62.598,7470	0,000
		2	7690	8,8%	215,63	8564	6,0%	2.577,07	41846	12,6%	1.666,93			
		3	7930	9,0%	80,77	9418	6,6%	414,67	28474	8,6%	75,99			
		4	11040	12,6%	2.157,19	9818	6,9%	255,32	24655	7,4%	180,73			
		5	13091	14,9%	3.483,89	19884	14,0%	4.021,53	17254	5,2%	5.176,68			
		6	9354	10,6%	381,96	10698	7,5%	229,42	28813	8,7%	0,02			
		7	9347	10,6%	222,99	20400	14,3%	4.248,56	21453	6,5%	2.540,50			
		8	5756	6,6%	3.508,03	16531	11,6%	594,79	56546	17,1%	2.160,30			
		9	6784	7,7%	422,20	17056	12,0%	622,89	31771	9,6%	33,36			
		10	5535	6,3%	6.371,57	20307	14,3%	895,24	72972	22,0%	3.685,46			

Figura I.3: Crosstab de Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.

Crosstab SERRP													MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE MORTE INDETERMINADA	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR	
			N	%	X2	N	%	X2	N	%	X2				
NÃO-SUSPENSO	NÃO	1	473572	18,2%	35.790,93	372693	15,7%	6.203,00	84773	4,8%	103.354,72	0,402	2.175.835,4320	0,000	
		2	226057	8,7%	11.854,28	408538	17,2%	87.094,93	100012	5,7%	44.090,22				
		3	288335	11,1%	1.224,47	325978	13,7%	7.542,77	181471	10,3%	3.387,61				
		4	221546	8,5%	17.222,32	426815	18,0%	96.957,57	108060	6,1%	40.663,25				
		5	316464	12,1%	9.776,27	289985	12,2%	9.692,90	80131	4,5%	54.986,27				
		6	367718	14,1%	16.938,40	259536	10,9%	410,96	140256	8,0%	18.165,84				
		7	194324	7,5%	135.802,96	22341	0,9%	40.078,67	6121	0,3%	46.637,30				
		8	83807	3,2%	168.952,08	77245	3,3%	152.283,49	651452	37,0%	907.670,29				
		9	135766	5,2%	1.373,93	56925	2,4%	26.872,29	124793	7,1%	21.056,81				
		10	300114	11,5%	1.789,94	132702	5,6%	57.043,72	285567	16,2%	50.917,23				
	SIM	1	392244	11,5%	56.006,59	393000	11,1%	47.796,85	86106	2,1%	174.826,40	0,499	5.503.458,4080	0,000	
		2	227018	6,6%	29.878,24	671796	19,0%	336.272,25	154470	3,8%	143.804,21				
		3	305115	8,9%	48,85	406427	11,5%	24.212,91	287861	7,0%	18.952,80				
		4	208544	6,1%	36.250,54	660510	18,7%	345.473,05	151325	3,7%	137.282,84				
		5	421562	12,3%	45.776,35	457034	12,9%	66.176,11	103540	2,5%	187.642,80				
		6	408333	11,9%	11.838,28	479211	13,6%	43.318,82	226592	5,5%	85.178,49				
		7	449933	13,1%	621.711,02	21388	0,6%	112.117,81	5116	0,1%	167.163,62				
		8	82760	2,4%	448.640,24	112508	3,2%	416.869,05	1754368	42,6%	1.459.965,10				
		9	234395	6,8%	18.147,14	77261	2,2%	61.209,61	262825	6,4%	11.280,94				
		10	693263	20,3%	6.836,56	250344	7,1%	243.388,63	1086739	26,4%	145.392,29				
SUSPENSO	NÃO	1	24720	16,9%	4.411,76	24773	12,4%	293,85	5720	3,8%	7.257,00	0,453	203.945,3070	0,000	
		2	15712	10,7%	41,65	31608	15,8%	6.155,84	3379	2,2%	9.358,65				
		3	17645	12,0%	92,52	26355	13,2%	684,11	11758	7,8%	1.568,68				
		4	13894	9,5%	803,30	37599	18,8%	7.499,73	8504	5,6%	5.160,46				
		5	19351	13,2%	746,20	28487	14,2%	2.090,91	6198	4,1%	6.334,31				
		6	21538	14,7%	1.686,07	23650	11,8%	84,52	10174	6,7%	2.605,73				
		7	9150	6,2%	9.446,68	2048	1,0%	1.504,28	506	0,3%	2.609,27				
		8	5074	3,5%	15.988,55	7002	3,5%	21.752,75	73233	48,5%	86.718,50				
		9	6719	4,6%	64,72	5409	2,7%	2.201,68	13051	8,6%	3.841,35				
		10	12785	8,7%	11,18	13463	6,7%	1.144,58	18489	12,2%	1.786,46				
	SIM	1	21404	10,8%	5.416,29	31143	9,8%	4.914,84	6497	1,7%	13.738,67	0,532	508.678,5810	0,000	
		2	18520	9,4%	154,80	52789	16,5%	23.890,03	5402	1,4%	22.634,80				
		3	21946	11,1%	1.557,68	36111	11,3%	2.962,34	18309	4,8%	6.125,36				
		4	14559	7,4%	1.213,39	60906	19,1%	27.960,87	12637	3,3%	16.373,13				
		5	25808	13,0%	3.579,58	47074	14,7%	11.681,69	7998	2,1%	20.179,00				
		6	27901	14,1%	2.985,57	46083	14,4%	5.678,58	17446	4,6%	11.741,32				
		7	21615	10,9%	50.595,71	1860	0,6%	5.197,94	471	0,1%	9.238,12				
		8	5359	2,7%	38.192,13	9299	2,9%	60.452,96	204621	53,7%	133.886,91				
		9	13813	7,0%	1.020,07	6305	2,0%	6.723,68	27694	7,3%	2.708,15				
		10	26984	13,6%	242,24	27691	8,7%	8.495,70	79957	21,0%	9.137,02				

Figura I.4: Crosstab de Morte Violenta por Causa Indeterminada e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.

Crosstab NÃO-SERRP												MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE HOMICÍDIO DE MULHERES	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR
			N	%	X2	N	%	X2	N	%	X2			
NÃO-SUSPENSO	NÃO	1	4482730	37,5%	823.581,61	2090821	20,2%	78.015,16	765320	10,0%	655.764,22	0,241	3.468.513,8800	0,000
		2	1556565	13,0%	44.646,66	1101234	10,6%	1.163,12	634843	8,3%	50.410,12			
		3	1302509	10,9%	6.875,39	1001738	9,7%	2.076,40	730177	9,5%	2.567,38			
		4	1098407	9,2%	2.474,33	1158481	11,2%	26.179,32	628576	8,2%	15.880,85			
		5	934979	7,8%	28.948,13	1035414	10,0%	5.176,88	821913	10,7%	16.650,92			
		6	774302	6,5%	41.004,06	843093	8,1%	0,01	823076	10,8%	64.162,35			
		7	609634	5,1%	197.093,05	824815	8,0%	10.826,21	1242478	16,2%	457.147,63			
		8	536103	4,5%	235.306,65	918105	8,9%	906,62	1120908	14,7%	326.607,28			
		9	664463	5,6%	216.833,02	1377970	13,3%	133.508,83	882721	11,5%	24.707,65			
	SIM	1	2910341	32,7%	1.724.651,61	1386511	13,3%	30.673,83	682711	5,3%	866.619,88	0,282	5.123.356,1830	0,000
		2	1228155	13,8%	170.480,57	954459	9,1%	1.585,49	894813	6,9%	93.858,12			
		3	966050	10,9%	9.403,67	1021647	9,8%	20,72	1189089	9,2%	5.815,33			
		4	893680	10,0%	1.437,21	1250686	12,0%	23.482,98	1231742	9,5%	11.249,42			
		5	661702	7,4%	78.523,48	1204965	11,5%	11.465,34	1516777	11,7%	18.524,86			
		6	725478	8,2%	75.493,44	1061802	10,2%	10.497,60	1842994	14,2%	102.115,23			
		7	576841	6,5%	194.456,37	878908	8,4%	85.538,01	2256323	17,4%	394.031,82			
		8	462341	5,2%	293.074,52	1012374	9,7%	23.387,84	2172981	16,8%	343.062,55			
		9	476866	5,4%	210.120,31	1665008	16,0%	326.204,99	1180534	9,1%	17.580,99			
SUSPENSO	NÃO	1	41405	38,0%	11.376,73	33440	25,1%	368,30	20560	11,5%	9.920,33	0,240	48.598,7020	0,000
		2	14441	13,3%	263,51	14699	11,1%	32,65	19682	11,0%	59,69			
		3	8257	7,6%	1.012,60	14582	11,0%	6,12	22428	12,5%	513,35			
		4	9717	8,9%	158,01	12871	9,7%	562,75	10512	5,9%	912,68			
		5	10968	10,1%	188,05	11847	8,9%	0,78	14419	8,0%	130,95			
		6	6284	5,8%	1.340,76	10264	7,7%	287,18	21888	12,2%	1.858,33			
		7	5645	5,2%	3.596,02	9385	7,1%	2.109,88	32540	18,1%	7.438,69			
		8	5698	5,2%	1.423,75	9444	7,1%	339,62	21020	11,7%	2.047,91			
		9	6486	6,0%	1.339,04	16455	12,4%	1.304,34	16440	9,2%	6,68			
	SIM	1	28275	32,2%	23.835,30	24547	17,2%	1.732,99	21496	6,5%	11.400,76	0,256	73.592,6930	0,000
		2	11851	13,5%	723,23	14663	10,3%	7,74	32690	9,9%	144,59			
		3	6990	8,0%	920,19	16646	11,7%	10,01	40457	12,2%	183,48			
		4	8982	10,2%	694,25	16437	11,5%	2.653,64	18095	5,5%	2.239,62			
		5	8897	10,1%	0,22	12649	8,9%	232,10	35601	10,7%	104,51			
		6	6447	7,3%	2.448,77	14290	10,0%	1.238,10	54870	16,6%	2.355,97			
		7	5877	6,7%	3.091,77	10923	7,7%	3.682,37	59626	18,0%	4.678,12			
		8	5200	5,9%	2.017,37	11784	8,3%	912,75	44382	13,4%	1.842,58			
		9	5348	6,1%	777,73	20387	14,3%	4.723,38	24203	7,3%	941,18			

Figura I.5: Crosstab de Homicídio de Mulheres e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.

Crosstab SERRP													MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE HOMICÍDIO DE MULHERES	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR	
			N	%	X2	N	%	X2	N	%	X2				
NÃO-SUSPENSO	NÃO	1	511010	19,6%	68.187,00	319052	13,4%	56,70	88814	5,0%	95.404,85	0,444	2.653.697,9500	0,000	
		2	13987	0,5%	8.340,96	60239	2,5%	40.694,64	2666	0,2%	15.121,08				
		3	83118	3,2%	229.340,19	118628	5,0%	147.012,93	773797	43,9%	1.055.445,86				
		4	217753	8,4%	15.891,41	458388	19,3%	152.697,76	60973	3,5%	90.029,22				
		5	197381	7,6%	23.311,58	390587	16,5%	75.059,72	130542	7,4%	17.466,33				
		6	101850	3,9%	32.880,76	90681	3,8%	31.653,48	268912	15,3%	182.310,19				
		7	313608	12,0%	20.848,49	264661	11,2%	8.776,33	48806	2,8%	80.836,46				
		8	405811	15,6%	59.580,04	182290	7,7%	19.425,75	128899	7,3%	18.274,37				
		9	289355	11,1%	2.189,22	218328	9,2%	2.197,56	178229	10,1%	6,35				
		10	473830	18,2%	75.218,49	269904	11,4%	1.420,32	80998	4,6%	84.019,92				
	SIM	1	392807	11,5%	64.497,74	323428	9,2%	10.224,18	133262	3,2%	105.704,84	0,538	6.418.647,2950	0,000	
		2	13216	0,4%	14.821,75	101773	2,9%	109.550,25	2946	0,1%	38.180,95				
		3	91198	2,7%	519.477,94	175074	5,0%	404.187,39	1964811	47,7%	1.551.443,60				
		4	267701	7,8%	13.733,03	731614	20,7%	429.728,05	86081	2,1%	249.986,44				
		5	240672	7,0%	20.554,33	560691	15,9%	157.471,30	240182	5,8%	55.996,85				
		6	158917	4,6%	153.326,92	110375	3,1%	230.309,29	1055117	25,6%	641.936,62				
		7	672546	19,6%	291.683,60	385103	10,9%	1.343,69	81096	2,0%	276.976,90				
		8	751079	21,9%	446.640,75	224357	6,4%	53.962,51	168366	4,1%	155.411,06				
		9	373876	10,9%	2.375,84	438798	12,4%	19.278,72	303928	7,4%	29.916,76				
		10	461155	13,5%	66.491,63	478266	13,6%	71.138,61	83153	2,0%	232.295,73				
SUSPENSO	NÃO	1	29420	20,1%	6.834,09	24348	12,2%	14,57	8237	5,5%	5.936,91	0,501	249.733,6040	0,000	
		2	2109	1,4%	147,55	6946	3,5%	2.716,06	272	0,2%	2.310,48				
		3	4057	2,8%	22.366,43	11618	5,8%	20.979,53	86006	57,0%	98.721,86				
		4	13600	9,3%	2.461,44	52523	26,2%	20.584,88	4356	2,9%	13.547,92				
		5	15917	10,9%	152,76	32646	16,3%	3.116,36	11074	7,3%	2.717,55				
		6	6692	4,6%	1.127,55	10018	5,0%	1.014,11	17466	11,6%	4.867,52				
		7	12782	8,7%	239,35	22263	11,1%	3.235,72	2829	1,9%	6.523,78				
		8	16393	11,2%	3.572,64	11371	5,7%	531,35	7298	4,8%	1.045,60				
		9	17492	11,9%	3.365,45	12391	6,2%	614,53	8573	5,7%	817,94				
		10	28126	19,2%	12.774,56	16270	8,1%	641,48	4901	3,2%	6.753,65				
	SIM	1	26794	13,5%	7.549,41	30000	9,4%	765,64	15158	4,0%	7.734,76	0,559	561.823,2850	0,000	
		2	1613	0,8%	896,11	13264	4,2%	11.479,62	300	0,1%	5.852,31				
		3	5576	2,8%	44.824,56	18668	5,8%	55.957,04	227257	59,6%	136.245,98				
		4	16997	8,6%	2.089,61	85879	26,9%	56.871,57	6466	1,7%	34.354,02				
		5	24143	12,2%	335,11	53196	16,7%	9.966,41	20062	5,3%	10.935,97				
		6	11031	5,6%	5.680,22	12068	3,8%	15.858,80	78022	20,5%	28.760,69				
		7	24542	12,4%	8.383,93	32869	10,3%	5.073,75	5173	1,4%	17.211,10				
		8	33031	16,7%	33.734,61	15182	4,8%	1.229,70	8508	2,2%	10.054,29				
		9	24160	12,2%	6.085,00	28178	8,8%	839,36	14385	3,8%	6.845,65				
		10	30022	15,2%	16.708,73	29957	9,4%	1.872,33	5701	1,5%	17.627,02				

Figura I.6: Crosstab de Homicídio de Mulheres e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.

Crosstab NÃO-SERRP													MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE HOMICÍDIO DE HOMENS	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR	
			N	%	X2	N	%	X2	N	%	X2				
NÃO-SUSPENSO	NÃO	1	2241652	18,7%	333.312,27	954472	0,0922	102.441,59	631.813,00	0,0826	122.182,51	0,290	5.046.875,7630	0,000	
		2	2187588	18,3%	373.296,42	877585	0,0848	114.182,19	571.100,00	0,0747	137.537,80				
		3	3356185	28,1%	169.733,21	2010176	0,1942	41.624,37	1.351.422,00	0,1767	77.171,28				
		4	1520612	12,7%	27.063,63	1402552	0,1355	54.543,65	410.827,00	0,0537	227.879,28				
		5	1217982	10,2%	70,42	1342999	0,1297	74.191,05	513.592,00	0,0671	93.853,38				
		6	568648	4,8%	183.027,37	1059670	0,1024	45.506,35	865.609,00	0,1132	82.237,05				
		7	456679	3,8%	231.362,54	861837	0,0833	5.789,09	979.708,00	0,1281	263.075,96				
		8	410346	3,4%	1.099.106,45	1842380	0,178	42.879,99	2.325.941,00	0,304	1.144.807,90				
	SIM	1	1246619	14,0%	528.299,52	582483	0,0558	45.953,38	556.514,00	0,0429	168.004,55	0,315	6.408.363,7630	0,000	
		2	1478011	16,6%	801.203,18	590579	0,0566	75.580,80	540.906,00	0,0417	244.989,72				
		3	2388862	26,8%	413.734,11	1456592	0,1396	84.713,42	1.889.789,00	0,1457	73.879,02				
		4	1175972	13,2%	182.163,11	1169939	0,1121	60.468,60	540.589,00	0,0417	329.716,25				
		5	1029642	11,6%	45.278,59	1284883	0,1231	95.432,81	716.566,00	0,0553	205.596,89				
		6	536666	6,0%	263.136,36	1444964	0,1385	30.318,47	1.888.580,00	0,1456	72.249,20				
		7	618259	6,9%	114.443,11	1155229	0,1107	1.764,31	1.665.480,00	0,1284	58.853,28				
		8	427423	4,8%	1.524.939,36	2751691	0,2637	1.107,94	5.169.540,00	0,3986	986.537,78				
SUSPENSO	NÃO	1	18850	17,3%	909,28	19024	0,1431	15,53	20.707,00	0,1154	722,52	0,243	49.624,4540	0,000	
		2	20981	19,3%	6.940,35	10528	0,0792	1.100,30	14.517,00	0,0809	1.320,54				
		3	28364	26,0%	1.013,72	26419	0,1987	178,09	36.088,00	0,2011	177,24				
		4	15223	14,0%	1.891,39	16332	0,1228	802,32	9.925,00	0,0553	3.393,88				
		5	11136	10,2%	33,40	18189	0,1368	2.193,97	11.468,00	0,0639	2.008,84				
		6	5532	5,1%	1.530,99	12059	0,0907	42,37	18.420,00	0,1026	618,77				
		7	4613	4,2%	3.556,59	10934	0,0822	391,18	26.300,00	0,1465	4.029,40				
		8	4202	3,9%	9.631,94	19502	0,1466	75,81	42.064,00	0,2344	7.046,04				
	SIM	1	11617	13,2%	2.904,67	12492	0,0878	90,93	21.154,00	0,0638	1.155,97	0,254	72.445,9460	0,000	
		2	15034	17,1%	12.191,10	8869	0,0623	167,92	16.252,00	0,049	2.338,68				
		3	23560	26,8%	2.874,43	20896	0,1468	1.363,43	61.922,00	0,1868	11,62				
		4	11761	13,4%	3.711,26	16841	0,1183	3.182,23	14.569,00	0,044	4.669,69				
		5	11034	12,6%	1.460,54	19372	0,1361	3.853,65	18.707,00	0,0564	3.643,18				
		6	5167	5,9%	2.949,67	17963	0,1262	11,30	45.996,00	0,1388	663,69				
		7	5391	6,1%	2.702,01	15727	0,1105	175,54	47.852,00	0,1444	1.256,52				
		8	4303	4,9%	14.058,32	30166	0,212	756,59	104.968,00	0,3167	6.253,01				

Figura I.7: Crosstab de Homicídio de Homens e Média Categorizada de PIB Per Capita em Família Monoparentais de NÃO-SERRP.

Crosstab SERRP													MEDIDAS		
VARIÁVEIS	MONOPARENTAL?	FAIXA DE HOMICÍDIO DE HOMENS	PIB FAIXA 1			PIB FAIXA 2			PIB FAIXA 3			V DE CRAMER	X2 TOTAL	P-VALOR	
			N	%	X2	N	%	X2	N	%	X2				
NÃO-SUSPENSO	NÃO	1	385067	14,8%	7.727,92	352050	0,1484	7.550,65	127.183,00	0,0722	43.157,06	0,260	910.393,3130	0,000	
		2	211919	8,1%	523,00	238460	0,1005	16.473,71	71.054,00	0,0403	31.234,22				
		3	320098	12,3%	79.882,96	352000	0,1484	33.017,53	685.104,00	0,3887	307.579,26				
		4	280852	10,8%	265,94	334171	0,1408	18.933,45	133.909,00	0,076	19.547,22				
		5	483558	18,5%	725,04	431656	0,1819	1.444,74	384.553,00	0,2182	5.906,16				
		6	259752	10,0%	3.888,10	359591	0,1515	32.036,95	139.691,00	0,0793	17.377,97				
		7	375115	14,4%	96.887,39	55190	0,0233	110.795,96	156.316,00	0,0887	57,68				
		8	291342	11,2%	13.898,65	249640	0,1052	6.238,90	64.826,00	0,0368	55.242,87				
	SIM	1	316004	9,2%	6.164,03	379047	0,1074	32.296,71	193.880,00	0,0471	56.611,13	0,393	3.424.717,5560	0,000	
		2	185133	5,4%	5,36	312485	0,0885	75.751,56	104.391,00	0,0253	63.839,50				
		3	355216	10,4%	272.785,37	548970	0,1555	110.882,79	1.785.023,00	0,4334	615.253,13				
		4	334081	9,8%	61,19	506950	0,1436	82.190,33	224.966,00	0,0546	74.263,94				
		5	500907	14,6%	59.620,54	565437	0,1602	36.304,87	1.217.347,00	0,2955	159.181,16				
		6	281552	8,2%	4.185,96	512105	0,1451	103.457,93	234.980,00	0,057	57.007,78				
		7	942227	27,5%	804.593,86	68497	0,0194	272.164,66	235.722,00	0,0572	112.094,94				
		8	508047	14,8%	34.603,34	635988	0,1802	133.513,72	122.633,00	0,0298	257.883,73				
SUSPENSO	NÃO	1	27448	18,7%	1.404,06	35400	0,1767	994,93	11.560,00	0,0766	5.366,08	0,307	93.690,0880	0,000	
		2	12808	8,7%	50,36	25936	0,1294	2.853,81	7.583,00	0,0502	2.975,40				
		3	20155	13,7%	8.241,62	31686	0,1581	7.737,90	76.598,00	0,5072	36.395,49				
		4	16902	11,5%	97,96	35022	0,1748	4.082,21	10.037,00	0,0665	4.076,79				
		5	25894	17,7%	108,10	31126	0,1553	127,65	25.445,00	0,1685	7,68				
		6	13573	9,3%	40,80	21851	0,109	1.045,67	8.227,00	0,0545	1.896,06				
		7	13429	9,2%	4.736,16	4229	0,0211	3.514,13	7.746,00	0,0513	0,23				
		8	16379	11,2%	3.433,98	15144	0,0756	59,97	3.816,00	0,0253	4.443,08				
	SIM	1	25539	12,9%	1.609,10	43508	0,1363	4.075,88	21.190,00	0,0556	7.629,79	0,395	280.219,2820	0,000	
		2	12973	6,6%	63,90	38511	0,1206	11.492,54	11.673,00	0,0306	8.532,00				
		3	26445	13,4%	22.198,34	57036	0,1787	20.898,17	207.862,00	0,5455	57.457,81				
		4	22879	11,6%	47,85	59252	0,1856	16.328,61	17.063,00	0,0448	14.872,63				
		5	32105	16,2%	507,52	43148	0,1351	4.131,74	89.962,00	0,2361	5.636,13				
		6	17349	8,8%	793,88	31055	0,0973	3.157,57	15.196,00	0,0399	5.146,98				
		7	33259	16,8%	47.176,89	4564	0,0143	9.406,66	10.941,00	0,0287	4.591,14				
		8	27360	13,8%	6.476,91	42187	0,1321	8.174,05	7.145,00	0,0188	19.813,16				

Figura I.8: Crosstab de Homicídio de Homens e Média Categorizada de PIB Per Capita em Família Monoparentais de SERRP.