



Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

**ANÁLISE DE AGRUPAMENTOS
APLICADA AOS DADOS DE VIOLÊNCIA
NO BRASIL**

Julia Mariana Pereira Nascimento

Uberlândia-MG
2019

Julia Mariana Pereira Nascimento

**ANÁLISE DE AGRUPAMENTOS
APLICADA AOS DADOS DE VIOLÊNCIA
NO BRASIL**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Prof^a Dr^a Priscila Neves Faria

**Uberlândia-MG
2019**



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Profª Drª Priscila Neves Faria

Prof. Dr. Lúcio Borges de Araújo

Profª Drª Patrícia Ferreira Paranaíba

**Uberlândia-MG
2019**

AGRADECIMENTOS

Por trás deste trabalho está o apoio de muitas pessoas e sem as quais nada disto teria sido possível.

Agradeço, primeiramente, a Deus, que permitiu que tudo isso acontecesse ao longo de minha vida, não somente nestes anos como universitária, mas em todos os momentos pela força e saúde.

À minha mãe Luciana Nascimento, ao meu pai Marcos Antônio e ao meu irmão Marcos Diego, pelo amor, incentivo, apoio incondicional, por todo o esforço investido na minha educação e por acreditarem que eu seria capaz de superar os obstáculos que a vida me apresentou.

Meus agradecimentos sinceros aos meus familiares e amigos que me ajudaram nesta jornada e de verdade eu quero que saibam que reconheço tudo que fizeram por mim, a perseverança que incutiram no meu pensamento para não desistir, o conforto de saber que nunca estarei só e que me ajudaram a relaxar nos momentos difíceis, que me mostraram que um pouco de diversão também é necessária e me colocaram para cima quando mais precisei.

Agradeço a esta universidade, ao curso de Estatística e seu corpo docente, por me proporcionar conhecimento, não apenas racional, mas manifestação do caráter afetividade da educação no processo de formação profissional, por tanto que se dedicaram mim, não somente por terem me ensinado, mas por terem me feito aprender. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos.

À minha orientadora, Priscila Neves Faria, pelo sua dedicação e paciência durante o projeto. Seus conhecimentos fizeram grande diferença no resultado final deste trabalho.

A todos que estiveram presentes em momentos de minha graduação, deixo aqui o meu muito obrigada e minhas mais sinceras gratidões.

RESUMO

Os estudos sobre violência nos estados brasileiros possibilitam o monitoramento de características e tendências comportamentais no país, além de permitirem a identificação de perfis diferentes em uma mesma região e perfis iguais entre regiões distintas. Diante disso, o presente trabalho teve como objetivo analisar os dados de violência coletados, aplicando a metodologia de Estatística Descritiva na análise de variáveis relacionadas a diferentes tipos de agressores e características específicas das vítimas, e a metodologia Multivariada que foi aplicada na análise de agrupamentos feita em três etapas, considerando faixa etária, grau de escolaridade e etnia das vítimas. Com esse estudo, foi possível obter resultados significativos, sendo um deles, o fato de que, em todas as regiões brasileiras, o sexo feminino obteve duas vezes ou mais registros do que o sexo masculino, de acordo com os registros de violência do ano de 2016. Outro resultado significativo foi a descoberta de algumas mudanças em relação aos grupos estabelecidos, sendo que esse fato ocorreu quando foram consideradas variáveis diferentes na análise de agrupamentos.

Palavras-chave: Clusters, Comparativo, Estados brasileiros.

ABSTRACT

The studies regarding violence in Brazilian states make it possible to monitor behavioral characteristics and tendencies in the country; they also allow the identification of different profiles in the same region and identical profiles in distinct regions. Therefore, the present paperwork had the objective of analyzing collected violence data, applying the Descriptive Statistics Method in the analysis of variables related to different kinds of aggressors and specified characteristics of the victims, and the Multivariate Method applied in the cluster analysis that it was done in three phases, considering age range, educational level and ethnicity. With this study, it was possible to obtain significant results, one of them being the fact that, in all of the Brazilian regions, the feminine sex had two times more registers than the masculine sex, according to the 2016 violence data. Another significant result was the discovery of some changes relating the established groups, as this fact occurred when different variables were considered in the cluster analysis.

Keywords: Clusters, Comparative, Brazilian states.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	III
1 Introdução	1
2 Metodologia	3
2.1 Conjunto de dados	3
2.2 Distância euclidiana	3
2.3 Distância Euclidiana Padronizada	4
2.4 Distância de Mahalanobis	4
2.5 Análise de Agrupamentos	4
2.6 Coeficiente de Correlação Cofenética	7
2.7 Número de clusters da partição final	7
3 Resultados	9
3.1 Análise Descritiva	9
3.2 Análise Agrupamento	12
4 Conclusões	21
Referências Bibliográficas	23

LISTA DE FIGURAS

2.1	Modelo de um dendrograma e a diferença entre Clusters Hierárquicos Aglomerativo e Divisivo.	5
2.2	Etapas para realização da Análise de Agrupamentos.	6
3.1	Número de vítimas por sexo e por região.	9
3.2	Número de vítimas de violência por grau de escolaridade e região.	10
3.3	Número de vítimas de violência por faixa etária e por região.	10
3.4	Número de vítima por etnia e por região.	11
3.5	Dendrograma obtido pelo método de agrupamento hierárquico da Ligação Mediana considerando a faixas etária dos indivíduos que sofreram algum tipo de violência no ano de 2016.	13
3.6	Dendrograma obtido pelo método de agrupamento hierárquico do Centroide considerando a etnia dos indivíduos que sofreram algum tipo de violência no ano de 2016.	15
3.7	Dendrograma obtido pelo método de agrupamento hierárquico da Ligação Mediana considerando a escolaridade dos indivíduos que sofreram algum tipo de violência no ano de 2016.	17

LISTA DE TABELAS

3.1	Estatísticas descritivas dos tipos de agressores.	12
3.2	CCC para os métodos e distâncias considerados: Análise da faixa etária.	12
3.3	Grupos formados no Dendrograma da Figura 3.5	13
3.4	CCC para os métodos e distâncias considerados: Análise da etnia.	15
3.5	Grupos formados no Dendrograma da Figura 3.6	16
3.6	CCC para os métodos e distâncias considerados: Análise da escolaridade.	17
3.7	Grupos formados no Dendrograma da Figura 3.7	18

1. INTRODUÇÃO

A violência é definida como uma ação que envolve força física ou verbal, no caso deste trabalho, de uma pessoa contra outra. Fenômeno este que vem sendo abordado e estabelecido em diferentes debates e posicionamentos políticos, éticos e filosóficos em suas mais diversas vertentes. Sendo que o mesmo foi observado como a quarta causa de morte na população em geral, segundo Estudo Global da Carga de Doenças de 2015 [25].

Todos os diversos gêneros de violência propiciam problemas graves à saúde mental [32]. Os estudos apontam o estresse pós-traumático, a depressão, o crescimento dos índices de suicídio, a ansiedade, além de maiores tendências de dependência química [7] [34] [35].

No que diz respeito à violência associada ao grande problema dos estupros no país, o Atlas da Violência [10] trás dados sobre esse fenômeno, em que 68% dos registros, no sistema de saúde, se referem a estupro de menores e onde quase um terço dos agressores das crianças (até 13 anos) são amigos e conhecidos da vítima e outros 30% são familiares mais próximos como pais, mães, padrastos e irmãos.

Outra questão apontada por [10] é a desigualdade das mortes violentas por etnia, que tem se acentuado na última década, quando a taxa de homicídios de indivíduos não negros diminuiu 6,8%, ao passo que a taxa de vitimização da população negra aumentou 23,1%.

Dados de 1991 a 2000 apresentados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e pelo Departamento de Informática do Sistema Único de Saúde do Brasil (DATASUS) [38] mostram números alarmantes, como os que homens de 15 a 19 anos têm taxa de mortalidade 6,3 vezes maior que as mulheres.

Dados de 2016 divulgados pelo Ministério da Justiça e Cidadania em relação à violência contra idosos mostraram números alarmantes. Nos quatro primeiros meses de 2016, o Disque 100 (canal da Ouvidoria Nacional de Direitos Humanos) registrou 12.454 denúncias de violência contra idosos [28]. Sendo que 77% das denúncias recebidas foram por negligência, seguido por violência psicológica com 51%, enquanto o abuso financeiro e econômico ou a violência patrimonial foi responsável por 38% e violência física e maus tratos ficaram com 26% das denúncias.

Pesquisa realizada pelo site Vagas.com e divulgada pela British Broadcasting Corporation (BBC) Brasil [4] constatou que de 4.975 profissionais brasileiros entrevistados, 52% afirmou ter sido vítima de assédio moral ou sexual. E 34% dos que declararam não ter passado por isso, afirmaram já ter presenciado algum episódio desse tipo de violência.

Analisando os dados das ocorrências de violência relacionadas à saúde no ano de 2016 regis-

trados em todo o país, é possível perceber um composto de eventos que merecem maior atenção. Um estudo a respeito dos índices de violência no Brasil, permite a disponibilidade de evidências acerca de sua intervenção na saúde pública, possibilitando a discussão sobre a necessidade ou não de novos conhecimentos, disponibilidade e incorporação de novos pensamentos e ações na sociedade contemporânea.

O presente estudo teve como objetivo analisar dados referentes à violência, por meio da análise de agrupamentos, para gerar informação a respeito das características das vítimas nos 27 estados brasileiros no ano de 2016 .

2. METODOLOGIA

2.1 CONJUNTO DE DADOS

Os dados do presente estudo referem-se às ocorrências de violência doméstica, sexual e/ou outras violências por um determinado conjunto de variáveis relacionadas aos tipos de agressores e às características específicas das vítimas no ano de 2016 nos estados brasileiros, com o objetivo de estudar doze categorias de agressores e quatro características específicas das vítimas. Os dados foram obtidos do Departamento de Informações e Informática do Sistema Único de Saúde (DATASUS).

O banco de dados utilizado, obtido no site do DATASUS, considera vinte e sete estados para a análise: os estados do Brasil – Acre, Alagoas, Amapá, Amazonas, Bahia, Ceará, Distrito Federal, Espírito Santo, Goiás, Maranhão, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Pará, Paraíba, Paraná, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Norte, Rio Grande do Sul, Rondônia, Roraima, Santa Catarina, São Paulo, Sergipe, Tocantins.

Foram avaliadas dezesseis variáveis para análise de violência no Brasil, as variáveis são: Cônjuge (X1), Cuidador (X2), Desconhecido (X3), Escolaridade (X4), Ex-cônjuge (X5), Ex-namorado (a) (X6), faixa etária (X7), Filho (a) (X8), Madrasta (X9), Mãe (X10), Namorado (a) (X11), Padrasto (X12), Pai (X13) e Patrão/chefe (X14), Raça (X15) e sexo (X16).

Para a execução das análises de estatística descritiva, utilizou-se o software Excel e, para a análise multivariada, foi desenvolvido rotinas no software R.

2.2 DISTÂNCIA EUCLIDIANA

A distância euclidiana, sem dúvida, é a mais comum e utilizada quando se trabalha com medidas de distâncias. Segundo [11] e [27], ela calcula a distância entre dois indivíduos a partir da raiz quadrada do somatório dos quadrados das diferenças entre os valores de i e j para todas as variáveis ($k=1,2,...,p$), sendo definida pela seguinte fórmula:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

2.3 DISTÂNCIA EUCLIDIANA PADRONIZADA

A distância euclidiana padronizada nada mais é do que a distância euclidiana sendo trabalhada com as variáveis padronizadas por $z_i = \frac{x_i - \bar{x}}{s_i}$, uma vez que as variáveis neste estudo não possuem a mesma escala ou unidade. Feita a transformação, a distância euclidiana passa a ser dada por [6]:

$$d(A, B) = \left[\sum_{n=1}^p (z_i(A) - z_i(B))^2 \right]^{\frac{1}{2}} \quad (2.2)$$

A expressão acima pode ser escrita da seguinte forma em notação vetorial:

$$d(A, B) = \left[(x(A) - x(B))' D^{-1} (x(A) - x(B)) \right]^{\frac{1}{2}}, \quad (2.3)$$

em que D é uma matriz diagonal, tendo como componentes as variâncias s_i^2 , isto é, $D = \text{diagonal}(s_1^2, s_2^2, \dots, s_p^2)$.

2.4 DISTÂNCIA DE MAHALANOBIS

Segundo [12], a distância de Mahalanobis, considera a variabilidade de cada unidade amostral, sendo recomendada para dados provenientes de delineamento experimentais, e, principalmente, quando as variáveis são correlacionadas. De acordo com que foi dito pelo [12] por considerar a variabilidade, esta distância utiliza $\sum$, que é a matriz de covariância entre as variáveis.

A distância de Mahalanobis, entre os indivíduos i e j , é estimada segundo [31] por:

$$D_{IJ}^2 = (\bar{X}_i - \bar{X}_j)' \Sigma^{-1} (\bar{X}_i - \bar{X}_j) \quad (2.4)$$

2.5 ANÁLISE DE AGRUPAMENTOS

De acordo com [20], a Análise de Agrupamentos é uma técnica analítica para encontrar subgrupos significativos de indivíduos ou objetos e para [22] corresponde a uma técnica exploratória que busca reconhecer a estrutura de agrupamentos, fornecer hipóteses acerca de associações e identificar outliers. Especificamente, o objetivo é classificar uma amostra de entidades (indivíduos ou objetos) em um número pequeno de grupos mutuamente exclusivos, onde elementos

dentro de um mesmo cluster (grupo) sejam semelhantes em relação a essas variáveis e elementos em clusters diferentes sejam distintos entre si.

Segundo [27] Análise de Agrupamentos (Análise de Clusters) pode ser feita por meio de métodos Hierárquicos (aglomerativos ou divisivos) e os métodos Não-Hierárquicos. O primeiro consiste em uma série de sucessivos agrupamentos ou sucessivas divisões dos indivíduos, onde são agregados ou desagregados até a formação do gráfico, em forma de uma árvore de classificação, dendrograma, que apresenta uma síntese objetiva dos resultados e sendo que cada ramo representa um indivíduo. Já os Não-Hierárquicos necessitam, a priori, da informação do número de grupos existentes para iniciar o processo. De acordo com [3], a ideia central destes métodos Não-Hierárquicos é escolher uma partição inicial dos indivíduos e, em seguida, alterar os membros dos grupos para obter-se a melhor partição.

Os métodos Hierárquicos Aglomerativos considera que, no início do processo, cada elemento do conjunto de dados observado é considerado um grupo isolado, assim, tem-se n grupos [27]. Em cada estágio do algoritmo, um grupo ou elemento é ligado a outro de acordo com sua similaridade, até o último passo, onde é formado um grupo único com todos os elementos. Já os métodos divisivos trabalham na direção oposta dos métodos Aglomerativos, ou seja, um grupo inicial contendo todos os elementos é dividido em dois subgrupos, de tal forma que os elementos em um subgrupo estejam distantes dos elementos do outro subgrupo. Estes métodos também podem ser úteis na seleção de variáveis mais importantes para caracterizar um determinado tipo de situação.

[2] definem que, ao contrário dos métodos Aglomerativos Hierárquicos, os métodos Divisivos não foram muito usados nem examinados pelo fato de não serem tão compreendidos e relatam também que, no ano de 1973, de todos os seus artigos publicados utilizando análise de agrupamento, dois terços utilizam método Hierárquico Aglomerativo por ser de mais fácil compreensão e simples ligação e, segundo [24], estes métodos são pouco mencionados na literatura, pois exigem uma maior capacidade computacional.

Na Figura abaixo é possível identificar a diferença entre os métodos.

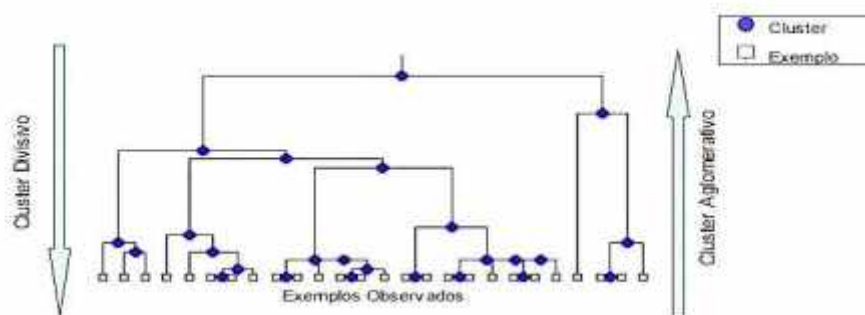


Figura 2.1: Modelo de um dendrograma e a diferença entre Clusters Hierárquicos Aglomerativo e Divisivo.

Fonte:[18]

Existem vários métodos Aglomerativos e, de acordo com [3], a maioria aparentam ser formu-

lações alternativas de três grandes conceitos de agrupamento aglomerativo: métodos de Ligação (single linkage, complete linkage, average linkage, median linkage), método do Centroide; método de soma de erros quadráticos ou variância (método de Ward). Segundo[14], esses são considerados os principais métodos de agrupamentos.

Com isso, tem-se que a estrutura básica da aplicação da Análise de Agrupamentos Hierárquico Aglomerativo pode ser definida de acordo com os passos a seguir representados na Figura 2.2.



Figura 2.2: Etapas para realização da Análise de Agrupamentos.

Fonte: elaborada pelo autor

Estes passos são dependentes e, algumas vezes, será fundamental voltar a passos anteriores para corrigir e melhorar certos passos posteriores. Neste trabalho serão utilizados os métodos hierárquicos aglomerativos que através do Coeficiente de Correlação Cofenética irá ser escolhido o melhor, e obteve que os utilizados até o fim da análise foram o método de Centroide e o da Mediana (median linkage).

O método de Centroide foi proposto por [36] e, neste método, a distância entre dois grupos é definida como a distância entre os seus centroides, que é o ponto médio no espaço multidimensional, logo, calcula-se a distância entre dois grupos como a diferença entre as suas médias, para todas as variáveis. Segundo[27], em cada etapa do agrupamento é necessário voltar aos dados originais para o cálculo da matriz de distâncias, sendo agrupados os conglomerados que apresentam a menor distância. Considerando dois conglomerados C_1 e C_2 e seus respectivos vetores de médias amostral $\overline{X}_1$ e $\overline{X}_2$, a distância é definida por $d(C_1, C_2) = (\overline{X}_1 - \overline{X}_2)'(\overline{X}_1 - \overline{X}_2)$. Vale lembrar que os métodos de Centroide e de Ward são apropriados apenas para dados quantitativos, já que utilizam como base a comparação dos vetores de médias.

Já o método da Mediana, teve com base na metodologia de [29] e [19], um algoritmo particular do método Centroide, com a diferença de que, ao agrupar os dois grupos A e B, os seus centroides, $\overline{X}_A$ e $\overline{X}_B$, recebem pesos iguais antes de produzirem o centroide do novo grupo resul-

tante do agrupamento. O novo centroide $\bar{X}$ fica a meio dos centroides dos grupos aglomerados, isto é, $\bar{X} = \frac{\bar{X}_A + \bar{X}_B}{2}$. Para [5], o método da Mediana é uma modificação do método do centroide para a independência da distância do tamanho do grupo. Deve-se perceber que a mediana mencionada não representa a mediana estatística, mas, sim, a mediana de um triângulo, isto é, um segmento de reta que une um vértice de um triângulo ao ponto médio do lado oposto [30].

2.6 COEFICIENTE DE CORRELAÇÃO COFENÉTICA

O Coeficiente de Correlação Cofenética (CCC) é uma medida de validação utilizada, principalmente, nos métodos Hierárquicos. [37] criaram o CCC utilizando a ideia de correlação de Pearson. O Coeficiente de Correlação Cofenética (CCC) efetua medidas do grau de ajuste entre a Matriz Fenética (matriz de distancias) com a Matriz Cofenética (matriz obtida por meio do Dendrograma). Tal correlação foi calculada de acordo com [6]:

$$r_{cof} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})(s_{ij} - \bar{s})}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})^2} \sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (s_{ij} - \bar{s})^2}}, \quad (2.5)$$

em que, c_{ij} : valor de similaridade entre os indivíduos i e j , obtidos a partir da matriz cofenética; s_{ij} : valor de similaridade entre os indivíduos i e j , obtidos a partir da matriz de similaridade;

$$\bar{c} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_{ij} \quad e \quad \bar{s} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n s_{ij} \quad (2.6)$$

O menor grau de distorção será refletido pelo maior coeficiente cofenético, que é dado em porcentagem, sendo que valores próximos à unidade indicam melhor representação [5] [13] [9]. [33] recomendam que o CCC seja maior do que 70%, caso isso não ocorra, o ideal é partir para outro método Hierárquico.

2.7 NÚMERO DE CLUSTERS DA PARTIÇÃO FINAL

De acordo com [27], determinar o número final de grupos de acordo com o conjunto de dados analisado é uma questão de grande importância em como deve-se proceder e, para [15], definir o número ótimo de grupos na última etapa dos métodos hierárquicos é, para os pesquisadores, uma tarefa bastante difícil.

Para [5], o número de grupos pode ser definido, a priori, através de um conhecimento dos dados, conhecimentos e opiniões dos analistas e pesquisadores ou pode ser definido, a posteriori, com base nos resultados da análise e [23] relataram que a utilização do método subjetivo, baseado no exame visual do dendrograma, pode gerar dificuldades ao número de grupos gerados quanto à tomada de decisão.

De acordo com [17], os dois melhores desempenhos no estudo de [26], estudo no qual eles testaram um total de 30 medidas de validação de agrupamento pretendendo definir o número ideal de grupos para cada uma das medidas usadas no processo de agrupamento hierárquico, utilizando dados artificiais com números de grupos conhecidos (intervalo de 2 a 5 grupos), foram as técnicas Pseudo F, introduzida por [8], e Pseudo T^2 , introduzido [16]. E [39], destaca que o índice Pseudo F é o mais útil para identificar número de grupos, sendo que este será o utilizado neste trabalho.

[8] definiram que, para cada passo do agrupamento, o cálculo da estatística Pseudo F seja calculado por meio da seguinte equação:

$$F = \frac{SSB/(g^* - 1)}{SSR/(n - g^*)} = \left(\frac{n - g^*}{g^* - 1} \right) \frac{R^2}{1 - R^2} \quad (2.7)$$

em que SSB é a soma de quadrados total entre os g grupos da partição, sendo que $SSB = \sum_{i=1}^{g^*} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})$ onde SSR é a soma de quadrados total dentro dos grupos da partição (Soma de Quadrados Residual):

$$SSR = \sum_{i=1}^{g^*} SS_i = \sum_{i=1}^{g^*} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)' (X_{ij} - \bar{X}_i) \quad (2.8)$$

R^2 é o coeficiente da partição, dado por:

$$R^2 = \frac{SSB}{SST_c} \quad (2.9)$$

SST_c é a soma de quadrados total corrigida para a média global em cada variável, expresso por $SST_c = \sum_{i=1}^{g^*} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})' (X_{ij} - \bar{X})$ em que g é o número de grupos relacionado com a partição do respectivo estágio de agrupamento; n é o número de elementos amostrais; $X'_{ij} = (X_{i1j} X_{i2j} \dots X_{ipj})$ é o vetor de medidas observadas para o j -ésimo elemento amostral do i -ésimo grupo; $\bar{X}'_i = (X_{i1} X_{i2} \dots X_{ip})$ é o vetor de médias do i -ésimo grupo; e $\bar{X}' = (\bar{X}_1 \bar{X}_2 \dots \bar{X}_p)$ é o vetor de médias global, sem levar em conta qualquer posição, em que $\bar{X}_{.1} = \frac{1}{n} \sum_{i=1}^{g^*} \sum_{j=1}^{n_i} X_{ij}$, $l=1,2,\dots,p$.

Com isso, tem-se que a ideia por trás deste critério é como se, em cada passo do algoritmo de agrupamento, estivesse sendo feito um Teste F de Análise de Variância para a comparação dos vetores de médias dos grupos que foram formados naquele passo. Com base nesses estudos e nas análises aplicadas aos dados, será utilizado a estatística Pseudo F para determinar a quantidade de grupos em cada etapa de análise de agrupamentos e considerando a quantidade mínima de 2 grupos e a máxima de 6.

3. RESULTADOS

3.1 ANÁLISE DESCRITIVA

Primeiramente, foi realizada uma análise das características entre pessoas que sofrem ou sofreram algum tipo de violência, abrangendo, graficamente, todas as regiões do Brasil, separando-se pelas categorias sexo, faixa etária, etnia e grau de escolaridade. Na Figura 3.1, foram comparados por sexo os registros do número de vítimas de violência em cada região do Brasil.

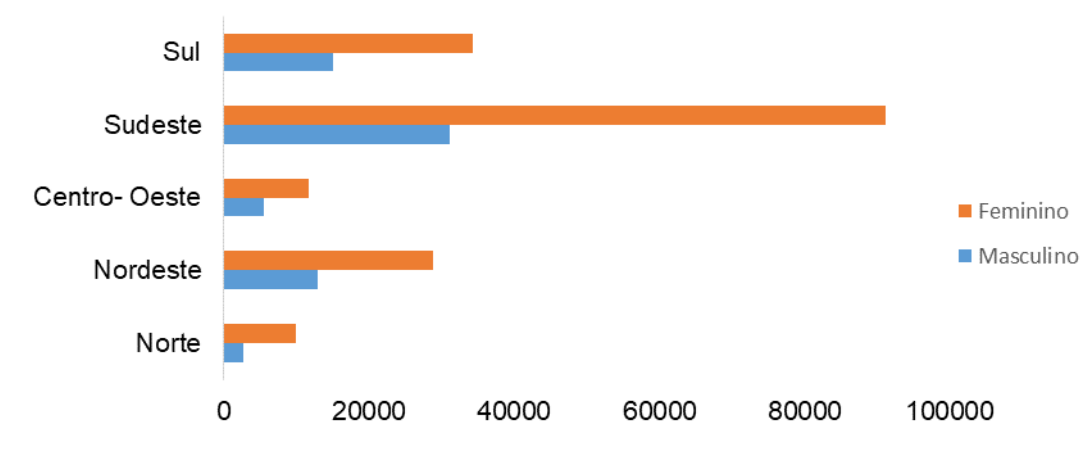


Figura 3.1: Número de vítimas por sexo e por região.

De acordo com os dados utilizados neste trabalho, foi obtido que, em todas as regiões, o número de vítimas de violência prevalece maior no sexo feminino, superando em duas vezes ou mais as ocorrências no sexo masculino.

A partir das estimativas populacionais dos municípios brasileiros e para as Unidades da Federação em 01/07/2016, dados disponibilizados pelo site do IBGE [21], obteve-se que o Sul tinha uma população de 29,44 milhões de habitantes, o Nordeste de 56,91 milhões, o Sudeste sendo a região mais populosa do Brasil e, apesar da população do Nordeste ser aproximadamente 1,92 vezes a população do Sul, teve maior registro de violência no Sul, tendo o Sudeste um número discrepante das outras regiões para registro de violência no ano de 2016.

Na Figura 3.2, tem-se o gráfico que contém o número de vítimas de agressão de acordo com o grau de escolaridade em cada região.

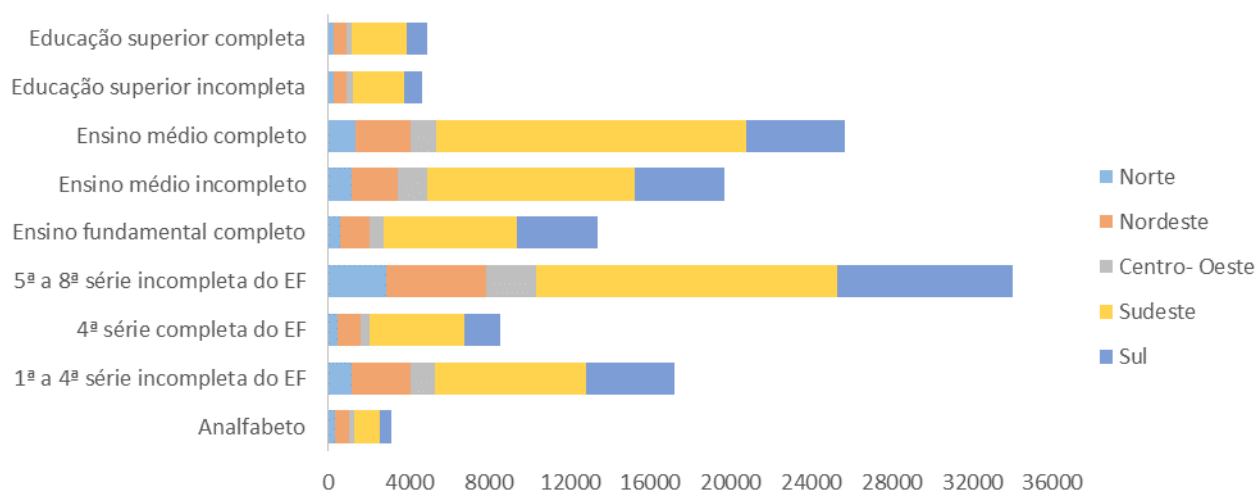


Figura 3.2: Número de vítimas de violência por grau de escolaridade e região.

Para realizar gráfico (Figura 3.2), foram consideradas apenas respostas válidas, desconsiderando os ignorados ou em branco. Com esta análise, foi possível observar que na categoria de 5ª à 8ª série incompleta do ensino fundamental, para todos os estados, tiveram maiores registros quando comparado às demais categorias para esta variável, isto se difere apenas para a região Sudeste, que possui um número maior para Ensino Médio completo e, posteriormente, para 5ª à 8ª série incompleta do ensino fundamental.

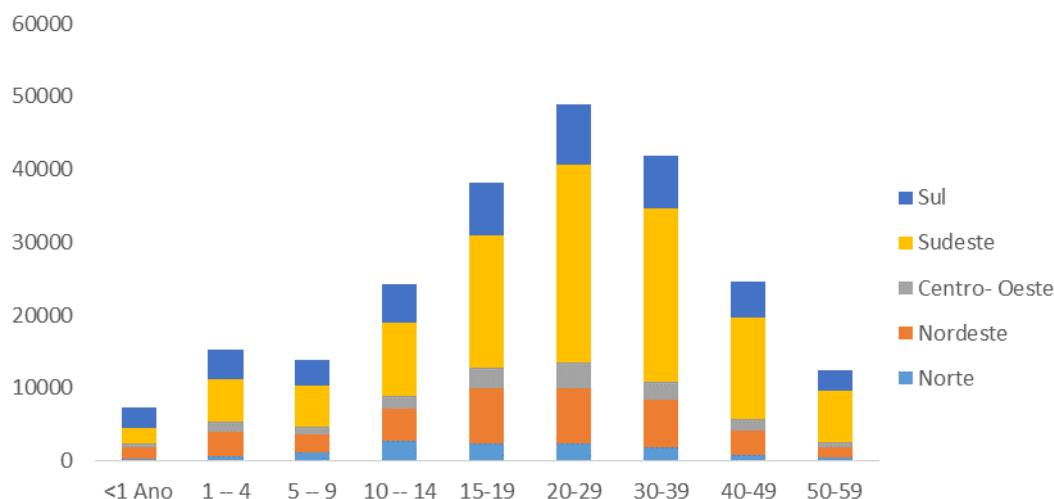


Figura 3.3: Número de vítimas de violência por faixa etária e por região.

Na Figura 3.3, tem-se o número de vítimas de agressão de acordo com a faixa etária em cada região e para a construção foi levado em conta apenas as vítimas que responderam em qual faixa etária pertenciam e, de acordo com esses dados, obteve-se que, praticamente, em

todas as regiões, o maior número de vítimas de agressão está na faixa de 20 a 29 anos, diferindo apenas na região Norte que tem o maior número entre 10 e 14 anos.

A partir da faixa etária de 10 a 14 anos, há um crescimento significativo de vítimas até a faixa etária de 20 a 29 anos e, logo após, inicia-se um decréscimo no decorrer das outras faixas etárias (Figura 3.3).

No gráfico da Figura 3.4, tem-se o número de vítimas de acordo com a sua etnia dentro de cada região. Considerando apenas as vítimas que declararam sua etnia, observa-se que nas regiões Norte, Nordeste e Centro- Oeste ocorreu o maior número de vítimas de etnia parda. Já nas demais regiões, na etnia branca.

Em quase todas as regiões, os menores números estão disponíveis para as etnias amarela e indígena, mudando apenas para a região Norte que obteve números menores para as etnias amarela e preta (Figura 3.4).

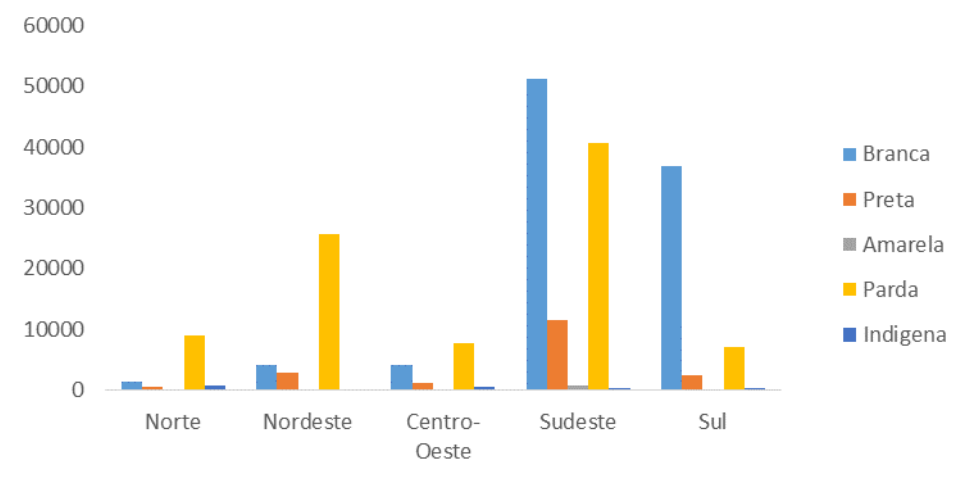


Figura 3.4: Número de vítima por etnia e por região.

Após realizar as análises para as variáveis das características das vítimas, realizou-se uma análise descritiva para as doze variáveis de agressões considerando os dados dos 27 estados do Brasil (Tabela 3.1). De acordo com a Tabela 3.1, é possível verificar que a maioria das variáveis possui um valor elevado para o desvio padrão e, com isso, tem-se que, quanto maior, menos homogêneos são os dados, observa-se isto em questão dos valores de um estado para o outro se diferirem bastante e verifica-se isso através dos valores mínimos e máximos de cada variável. Sendo que um dos valores mais discrepantes se encontra no variável cônjuge, já que enquanto em um estado o valor foi de 6 ocorrências, em outro foi de 9654.

De acordo com esta análise preliminar encontrou-se também que as maiores médias estão nas variáveis cônjuge e mãe e que, também, são as que possuem a maior variação de um estado para o outro.

Tabela 3.1: Estatísticas descritivas dos tipos de agressores.

	Mínimo	Máximo	Média	Desvio padrão	Variância
Ex-Cônjuge	11	4096	521,96	852,17	726201,27
Cônjuge	6	9654	1304,54	2179,08	4748374,70
Cuidador	1	221	45,88	59,86	3583,28
Desconhecido	5	6082	845,68	1255,63	1576614,30
Ex-namorado(a)	1	1228	154,78	268,99	72356,33
Filho(a)	6	2411	336,19	549,73	302199,31
Namorado(a)	1	208	29,00	45,89	2105,74
Padrasto	3	1339	188,50	280,24	78535,15
Madrasta	3	4604	726,68	1088,48	1184797,86
Mãe	6	5347	996,54	1433,70	2055509,37
Pai	3	4604	726,68	1088,48	1184797,86
Patrao/chefe	1	173	20,83	36,92	1363,42

3.2 ANÁLISE AGRUPAMENTO

Após a realização das análises descritivas e gráficas das variáveis que estão sendo estudadas, foi aplicada a análise de agrupamento em três análises, sendo a primeira considerando as faixas etárias dos indivíduos, abaixo de 1 ano, de 1 a 4 anos, de 5 a 9 anos, de 10 a 14 anos, de 15 a 19 anos, de 20 a 29 anos, de 30 a 39 anos, de 40 a 49 anos, de 50 a 59 anos e maior de 60 anos, que já sofreram algum tipo de violência; a segunda, considerando a etnia dos mesmos; e a terceira, sua escolaridade.

Estudando os dados e suas características para esta primeira análise, a distância euclidiana padronizada foi a escolhida, devido ao fato dos dados não estarem na mesma escala e, a fim de selecionar o melhor método hierárquico de forma a verificar se este está adequado para os dados, utilizou-se o CCC, que avalia a consistência dos agrupamentos formados pelos métodos, isto é, avalia se o método hierárquico utilizado está adequado aos dados. Quanto maior o coeficiente, melhor com isso tem-se que o método escolhido é o Ligação Mediana com CCC de 96,3% (Tabela 3.2).

Tabela 3.2: CCC para os métodos e distâncias considerados: Análise da faixa etária.

	Distância Euclidiana	Distância Euclidiana Padronizada	Distância de Mahalanobis
Ligação Simples	0,948	0,956	0,898
Ligação Completa	0,955	0,897	0,901
Ligação Média	0,955	0,908	0,921
Centroide	0,955	0,939	0,918
Ligação Mediana	0,953	0,963	0,780
Ward	0,946	0,889	0,946

O gráfico dendrograma foi gerado (Figura 3.5) e, assim, realizado o corte bem definido através do Pseudo F, que indicou 6 grupos (Tabela 3.3).

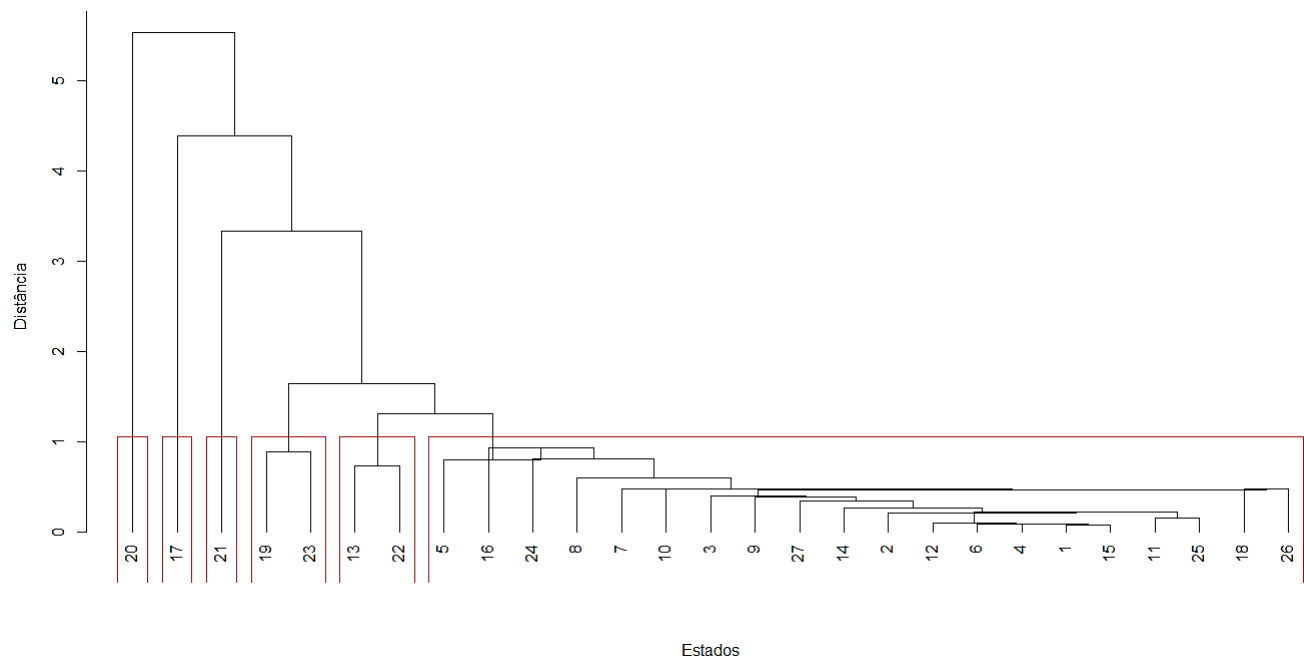


Figura 3.5: Dendrograma obtido pelo método de agrupamento hierárquico da Ligação Mediana considerando a faixas etária dos indivíduos que sofreram algum tipo de violência no ano de 2016.

Após verificar os grupos que se formaram, percebe-se que em 3 deles, Grupo 1, Grupo 2 e Grupo 3, permanecem estados totalmente isolados. Os Grupos 4 e 5, são formados por apenas dois estados pertencentes a diferentes regiões. Já o Grupo 6 é o maior e se comporta com uma mistura de estados de todas as regiões do Brasil (Tabela 3.3).

Tabela 3.3: Grupos formados no Dendrograma da Figura 3.5

Grupo 1	São Paulo
Grupo 2	Minas Gerais
Grupo 3	Paraná
Grupo 4	Rio de Janeiro, Rio Grande do Sul
Grupo 5	Pernambuco, Santa Catarina
Grupo 6	Pará, Bahia, Mato Grosso do Sul, Maranhão, Tocantins, Ceará, Amazonas, Piauí, Distrito Federal, Alagoas, Acre, Paraíba, Amapá, Roraima, Rondônia, Sergipe, Rio Grande do Norte, Mato Grosso, Espírito Santo, Goiás

De acordo com [21], o Brasil possui mais de 200 milhões de habitantes, sendo que os seis estados mais populosos, respectivamente, São Paulo, Minas Gerais, Rio de Janeiro, Bahia, Rio Grande do Sul e Paraná, cinco pertencem aos Grupos 1, 2, 3 e 4. Apesar de serem os estados mais populosos, temos que a diferença mínima entre as populações é de aproximadamente 23

milhões de habitantes e que estes mesmos cinco estados foram os que, de acordo com os dados de registros de violência por meio das faixas etárias no ano de 2016, obtiveram os maiores registros sendo, respectivamente, São Paulo, Minas Gerais, Paraná, Rio de Janeiro e Rio Grande do Sul, com a diferença mínima entre eles de 15.777 registros, e os únicos onde encontrou-se em todas as faixas etárias, números maiores do que a média dos dados.

Apesar desses cinco estados terem características semelhantes, o que fez permanecerem em grupos distintos está no quesito população, uma vez que São Paulo é o estado mais populoso, com os maiores registros e as maiores médias, formando o Grupo 1. Em seguida, tem-se Minas Gerais, que, apesar de estar em segundo, pois possui população, número de registros e médias em cada faixa etária menores, comparando-se com São Paulo, formando o Grupo 2.

O que causou uma mudança em relação aos grupos é que, apesar do estado do Paraná ser o sexto estado mais populoso, foi o terceiro com o maior registro de violência e, por isso, permanece em um grupo exclusivo, o Grupo 3.

Os estados Rio de Janeiro e Rio Grande do Sul, tendo populações e números de violência no ano de 2016 próximos, formam o Grupo 4.

Já no Grupo 5 tem-se dois estados, Pernambuco e Santa Catarina, pertencentes a regiões diferentes e sendo respectivamente a sétima e a décima primeira população brasileira, apesar dessas diferenças, em relação aos dados que estão sendo analisados, eles possuem números de registros semelhantes neste ano de 2016, com os maiores números de registros após os estados do Grupo 1, 2, 3 e 4, e, também, características semelhantes dentre as faixas etárias.

O grupo 6 é o maior grupo, com vinte estados pertencentes a todas regiões brasileiras, todos os estados da região Norte e Centro-Oeste, a maioria dos estados da região Nordeste (oito estados), e apenas um estado da região Sul e Sudeste. Dentre os estados pertencentes a este grupo estão inclusos os que possuem as seis menores populações do Brasil e estados que estão entre os dez mais populosos, mas apesar da diferença entre as populações e regiões, neste grupo estão os estados com os menores registros de violência de acordo com as faixas etárias no ano de 2016 e dentre eles quinze, que corresponde a 75% deste grupo, tem todas os registros de violência abaixo da média de acordo com as faixas etárias consideradas.

Considerando o registro de violência de acordo com as faixas etárias, os Grupos 1, 2, 3 e 4 têm aproximadamente 64% dos registros de violência enquanto os Grupos 5 e 6, que possuem mais estados, têm aproximadamente 36%.

Na realização da segunda análise, foram considerados os registros de violência de acordo com as vítimas que declaram pertencer à alguma etnia, sendo estas: branca, preta, amarela, parda ou indígena. Para verificar-se qual o melhor método hierárquico, foi-se utilizado o CCC novamente e, para certificar que a melhor distancia seria a euclidiana padronizada, em questão, novamente, de os dados não terem a mesma escala (Tabela 3.4).

Tabela 3.4: CCC para os métodos e distâncias considerados: Análise da etnia.

	Distância Euclidiana	Distância Euclidiana Padronizada	Distância de Mahalanobis
Ligação Simples	0,969	0,948	0,901
Ligação Completa	0,942	0,961	0,885
Ligação Média	0,973	0,972	0,927
Centroide	0,969	0,974	0,920
Ligação Mediana	0,971	0,960	0,903
Ward	0,795	0,959	0,829

Conforme as informações fornecidas na Tabela 3.4, quando se utiliza a distância euclidiana padronizada, obtém-se o melhor CCC no método de Centroide, que é de 97,4%, sendo considerado uma excelente correlação. Após as análises, encontrou-se o seguinte dendrograma (Figura 3.6) e nele feito o corte de acordo com a estatística Pseudo F, encontrando os seguintes grupos (Tabela 3.5).

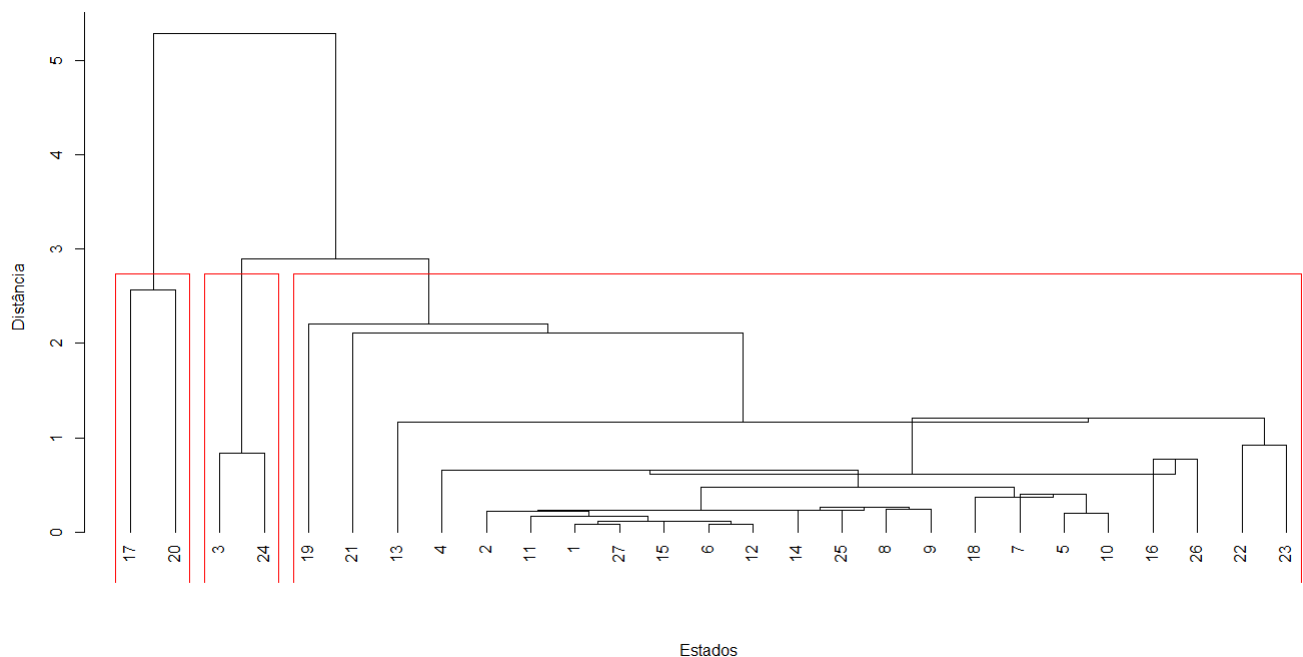


Figura 3.6: Dendrograma obtido pelo método de agrupamento hierárquico do Centroide considerando a etnia dos indivíduos que sofreram algum tipo de violência no ano de 2016.

No Grupo 1, tem-se os dois estados mais populosos do Brasil, pertencentes à mesma região e também são os que tiveram os maiores números de registro de violência de acordo com os dados de 2016, considerando a variável etnia. Dentre os 27 estados pertencentes na análise, 3 deles, que correspondem a 11,11%, foram os únicos que obtiveram registros acima da média em todas as etnias consideradas neste estudo, sendo que Minas Gerais e São Paulo estão dentre eles. Apenas este grupo corresponde a 40,4% do registro total, de acordo com a etnia (Tabela 3.5).

Nesta segunda análise, percebe-se que houve mudanças em relação a quantidade de grupos da etapa anterior, sendo que, também, houve alterações dos estados pertencentes em cada grupo, onde nenhum estado permaneceu isolado em algum grupo e, com isso, percebe-se que os dados são menos heterogêneos.

Tabela 3.5: Grupos formados no Dendrograma da Figura 3.6

Grupo 1	Minas Gerais, São Paulo
Grupo 2	Amazonas, Mato Grosso do Sul
Grupo 3	Rio De Janeiro, Paraná, Pernambuco, Roraima, Acre, Rio Grande do Norte, Rondônia, Distrito Federal, Sergipe, Amapá, Paraíba, Alagoas, Mato Grosso, Maranhão, Piauí, Espírito Santo, Tocantins, Pará, Ceará, Bahia, Goiás, Santa Catarina, Rio Grande do Sul

Já o Grupo 2 foi formado por dois estados de regiões diferentes, sendo que Amazonas possui mais de 1 milhão de habitantes a mais do que o Mato Grosso do Sul e, mesmo Manaus possuindo uma população maior, Mato Grosso do Sul obteve um maior número de registros, sendo, aproximadamente, 2000 a mais. Apesar de todas essas diferenças, estes dois estados possuem uma característica única: foram os estados com maior número de registro na etnia indígena, sendo Manaus com o maior número, seguido por Mato Grosso do Sul, posteriormente. Os valores destes estados chegaram a ultrapassar a média em 5,7 vezes ou mais e, dentro dessa etnia, eles correspondem a 45,33% dos dados, sendo que no contexto geral, eles representam 4,34%.

No Grupo 3 é onde encontra-se o maior número de estados, 23, sendo que nele, estão presentes o terceiro estado mais populoso do Brasil, o menos populoso e, também, tem-se todos os estados das regiões nordeste, sudeste, 8 estados da região norte e 2 estados da região centro-oeste e sudeste. Sendo que, neste grupo, 83,33% deles obtiveram 0,1 ou 2 valores acima da média de acordo com as etnias e, apesar de ser um grupo que possui 85,19% dos estados, os seus registros representam 55,27%.

Na terceira análise, considerou-se os registros de violência de acordo com as vítimas que declaram a sua escolaridade, sendo elas: Analfabeto, 1^a a 4^a série incompleta do Ensino Fundamental (EF), 4^a série completa do EF, 5^a a 8^a série incompleta do EF, ensino fundamental completo, Ensino médio incompleto, Ensino médio completo, Educação superior incompleta, Educação superior completa e Não se aplica. Nesta etapa, também se utilizou a distância euclidiana padronizada, pois a maioria dos dados não possui a mesma escala, para certificar qual o melhor método Hierárquico de análise de agrupamento, utilizou-se novamente o CCC e, com isso, foi produzida a Tabela 3.6 onde se compara os CCC em todos os métodos utilizando as três distâncias mencionadas neste trabalho.

Como se nota na Tabela 3.6 o método que mais se adapta aos dados é o de Ligação Mediana, tendo um CCC igual a 97,8% sendo considerado uma excelente correlação cofenética. Após a

Tabela 3.6: CCC para os métodos e distâncias considerados: Análise da escolaridade.

	Distância Euclidiana	Distância Euclidiana Padronizada	Distância de Mahalanobis
Ligação Simples	0,969	0,977	0,913
Ligação Completa	0,963	0,922	0,747
Ligação Média	0,972	0,965	0,939
Centroide	0,968	0,944	0,941
Ligação Mediana	0,960	0,978	0,922
Ward	0,880	0,919	0,790

escolha mais adequada da distância e método a se utilizar, foi gerado o gráfico dendrograma (Figura 3.7).

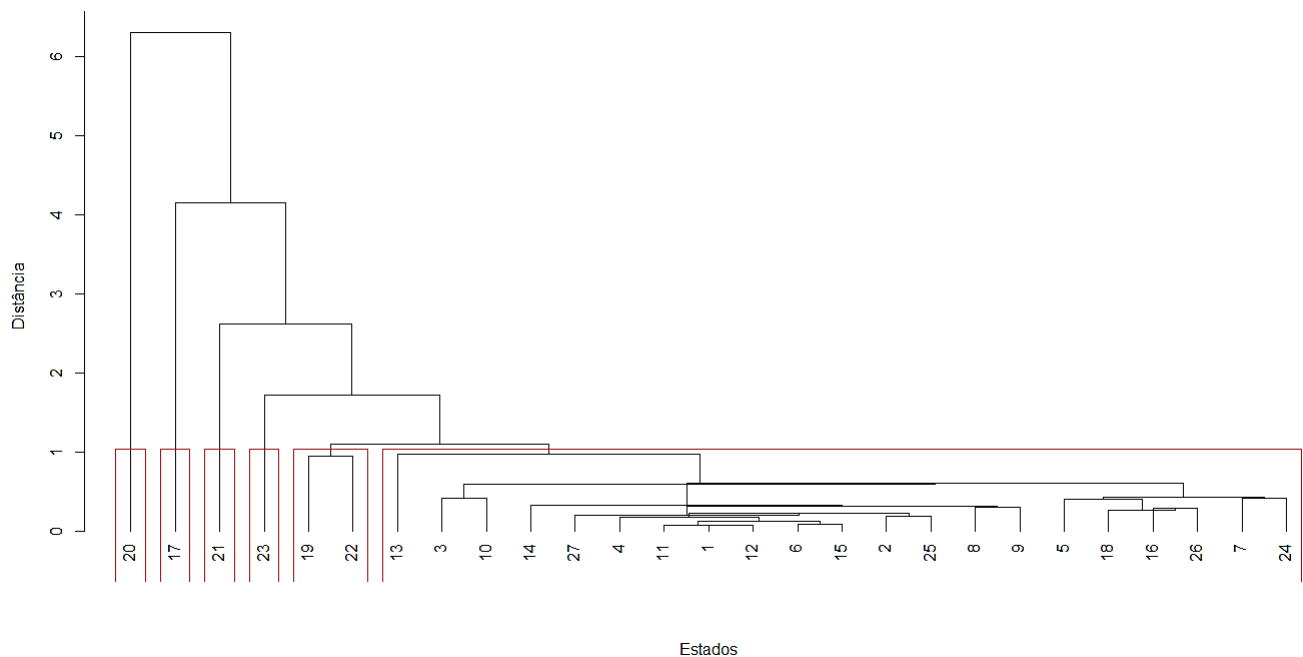


Figura 3.7: Dendrograma obtido pelo método de agrupamento hierárquico da Ligação Mediana considerando a escolaridade dos indivíduos que sofreram algum tipo de violência no ano de 2016.

No dendrograma foi realizado o corte definido de acordo com a estatística Pseudo F, sendo encontrados os seguintes grupos (Tabela 3.7).

Tabela 3.7: Grupos formados no Dendrograma da Figura 3.7

Grupo 1	São Paulo
Grupo 2	Minas Gerais
Grupo 3	Paraná
Grupo 4	Rio Grande do Sul
Grupo 5	Rio de Janeiro, Santa Catarina
Grupo 6	Pernambuco, Amazonas, Ceará, Alagoas, Distrito Federal, Roraima, Rio Grande do Norte, Rondônia, Paraíba, Amapá, Sergipe, Acre, Mato Grosso, Maranhão, Piauí, Pará, Espírito Santo, Bahia, Goiás, Tocantins, Mato Grosso do Sul

Nesta etapa obteve-se um número maior de grupos, comparado com a etapa anterior, e a mesma quantidade de grupos da primeira etapa, sendo que houve mudanças em relação aos estados pertencentes a cada grupo. Os estados São Paulo, Minas Gerais e Paraná permaneceram da mesma maneira, cada um em um grupo isolado, sendo os grupos 1, 2 e 3. E, também, uma grande mudança que ocorreu é que um novo estado ficou isolado, Rio Grande do Sul.

Dentre todos os estados, os únicos que possuem valores acima da média para todos os níveis de escolaridade foram os pertencentes aos grupos 1, 2, 3 e 4, que são respectivamente os estados, São Paulo, Minas Gerais, Paraná e Rio Grande do Sul, sendo estados pertencentes à região sudeste e sul do Brasil. Estes, também, são os estados com os maiores registros de ocorrência de vítimas, sendo que a escolaridade e o número de ocorrências desses 4 estados correspondem a 58,56% das ocorrências nesta variável no ano de 2016. Isso faz com que estes estados não pertençam ao mesmo grupo por possuírem números de registros totalmente discrepantes, já que a diferença mínima entre eles é de aproximadamente 4800.

No grupo 5, tem-se Rio de Janeiro e Santa Catarina, que são estados pertencentes a regiões diferentes e possuem características diferentes. O Rio de Janeiro é pertencente à região Sudeste do Brasil, que é a região mais populosa, enquanto Santa Catarina, pertence à região Sul e é o menor estado desta região. Nesta etapa, eles foram os únicos dois estados que tiveram registros acima da média para todas as variáveis, exceto para o nível analfabeto que, de acordo com o [1], os estados do Rio de Janeiro e Santa Catarina foram os que obtiveram os menores percentuais de analfabetismo, respectivamente, 2,5% e 2,6%. E, em relação ao número de registros, eles ficam atrás apenas dos estados pertencentes aos grupos 1, 2, 3 e 4.

No Grupo 6 é onde se encontra o maior número de estados, apesar de se encontrar 21 dos 27 estados do Brasil, o registro de ocorrência deles representa apenas, aproximadamente, 30,15%, sendo que, no Grupo 5 tem-se apenas dois estados, Rio de Janeiro e Santa Catarina, com, aproximadamente, 11,29% dos registros. Com isso, tem-se que o Grupo 6, tem 10 vezes mais o número de estados comparando-se ao Grupo 5 e tem apenas, aproximadamente, 2,67 vezes a mais de registros.

Este grupo só não possui estado pertencente à região sul, dentre os estados pertencentes,

tem-se todos da região norte, nordeste e centro-oeste. Neste grupo, encontra-se grande diversidade em relação a populações e características entre as regiões, contudo, apesar das diferenças, o que faz com que eles pertençam ao mesmo grupo é o fato de que eles possuem números de registros próximos em relação aos níveis de escolaridade das vítimas de violência no ano de 2016. Quando se realiza uma análise mais detalhada em relação aos dados que estão disponíveis, aproximadamente, 71,42% dos estados pertencentes, possuem os registros, referentes aos dez níveis de escolaridade, abaixo da média.

4. CONCLUSÕES

Realizar a análise descrita foi importante a fim de possibilitar conhecer determinadas características gerais dos dados. Já a análise de agrupamento foi uma ferramenta importante por permitir classificar os indivíduos com base na observação de semelhanças de acordo com as variáveis nas quais está sendo utilizada.

Neste estudo, foi possível avaliar o comportamento dos estados brasileiros de acordo com registro de violência no ano de 2016, disponível pelo DATASUS. Realizando-se três análises de agrupamentos, considerando as variáveis Faixa Etária, Etnia e Escolaridade. Dessa maneira, de acordo com o Pseudo F, foram identificadas quantidades de grupos diferentes, sendo que, na primeira e terceira etapa, foram 6 grupos e, na terceira, 3 grupos. Em cada etapa foram identificadas mudanças em relação aos indivíduos dos grupos, sendo que, em todos os grupos, foi possível identificar características específicas de acordo com a variável em estudo.

Uma semelhança entre as três etapas foi que os estados de Minas Gerais e São Paulo, quando não estavam em grupos isolados, estavam no mesmo grupo, já que, no decorrer das análises, foram os estados com os maiores números de registros de violência e com todos os valores acima da média, para todas as variáveis em estudos.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Agência IBGE Notícias: *Analfabetismo cai em 2017, mas segue acima da meta para 2015*, 2018. <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/21255-analfabetismo-cai-em-2017-mas-segue-acima-da-meta-para-2015>, acessado em 20/09/2019.
- [2] Andenderfer, M. S. e Blashfiel, R. K.: *Series: Quantitative Applications in the Socila Sciences*. Califórnia: University of Iowa, 1984.
- [3] Anderberg, M. R.: *Cluster analysis for applications*. New York: Academic Press, 1973.
- [4] Barifouse, R.: *Metade dos brasileiros já sofreu assédio no trabalho, aponta pesquisa.*, 2015. https://www.bbc.com/portuguese/noticias/2015/06/150610_assedio_trabalho_pesquisa_rb, acessado em 13/05/2019.
- [5] Barroso, L. P. e Artes, R.: *Análise multivariada*. Lavras: UFLA, 2003.
- [6] Bussab, W. O., Miazaki, .S. e Andrade, D. F.: *Introdução à análise de agrupamentos*. São Paulo: Associação Brasileira de Estatística, 1990.
- [7] Caldwell, J. E., Swan, S. C. e Woodbrown, V. D.: *Gender differences in intimate partner violence outcomes*. Psychology of Violence, 2(1):42–57, 2012.
- [8] Calinski, T. e Harabasz, J.: *A Dendrite Method for Cluster Analysis*. Communications in Statistics, 3(1):1–27, 1974.
- [9] Cargnelutti, F. A., Ribeiro, N. D. e Burin, C.: *Cálculo com Geometria Analítica*, vol. 45. Pesq. agropec. bras. [online], 2010. ISSN 0100- 204X.
- [10] Cerqueira, D. *et al.*: *Atlas da violência*. Brasília: Ipea, 2018.
- [11] Cormack, R. M.: *A review of classification*. Journal of the Royal Statistical Society. Series A (General), 134(3):321–367, 1971.
- [12] Cruz, C. D.: *Aplicação de algumas técnicas multivariadas no melhoramento de plantas*. Tese de Doutorado, 1990.
- [13] Cruz, C. D. e Carneiro, P. C. S.: *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: UFV, 2003.

- [14] Cruz, C.D., Ferreira, F.M. e Personi, L.A.: *Biometria aplicada ao estudo da diversidade genética*. Visconde do Rio Branco: Suprema., 2011.
- [15] Dias, A.: *Seleção multivariada e identidade de modelos não lineares para o crescimento e acúmulo de nutrientes em frutos de mangueira*. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Coordenação do Programa de Pós-Graduação de Estatística e Experimentação Agropecuária, 2014.
- [16] Duda, R.O. e Hart, P.E.: *Pattern classification and scene analysis*. New York: Wiley, 1973.
- [17] Everitt, B.S., Landau, S., Leese, M. e Stahl, D.: *Cluster Analysis*. Chichester: Wiley, 2011.
- [18] Freitas, J.A.B.: *Análise de Cluster da Lisozima*. Relatório Científico apresentado ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por conta da conclusão de projeto de Iniciação Científica., 2006.
- [19] Gama, M.d.P.: *Bases de análise de agrupamentos (“Cluster Analysis”)*. Dissertação (Mestrado em Estatística e Métodos Quantitativos), 1980.
- [20] Hair, J.F. et al.: *Analise Multivariada de Dados*. New Jersey: Prentice Hall, 3ª ed., 2009.
- [21] IBGE: *Estimativas populacionais para os municípios e para as Unidades da Federação brasileiros em 01.07.2016*, 2016. https://ww2.ibge.gov.br/home/estatistica/populacao/estimativa2016/estimativa_dou.shtm, acessado em 23/06/2019.
- [22] Johnson, R. e Wichern, D.: *Applied multivariate statistical analysis*. São Paulo: Bookman, 6ª ed., 1992.
- [23] Karasawa, M. et al.: *Aplicação de métodos de agrupamento na quantificação da divergência genética entre acessos de tomateiro*. Horticultura Brasileira 23, pp. 1000–1005, 2005.
- [24] Kaufman, L. e Rousseeuw, P. J.: *Finding groups in data: an introduction to cluster analysis*. New York: Wiley, 1990.
- [25] Malta, D.C. et al.: *Fatores associados ao diabetes autorreferido segundo a Pesquisa Nacional de Saúde*. Rev. Saúde Pública, 51, supl. 1, 12s, 2013.
- [26] Milligan, G.W. e Cooper, M.C.: *An examination of procedures for determining the number of clusters in a data set*. Psychometrika, 50(2):159–179, 1985.
- [27] Mingoti, S.A.: *Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada*. Belo Horizonte: UFMG, 2005.

- [28] Ministério da Mulher, da Família e dos Direitos Humanos: *Dados do Disque 100 mostram que mais de 80% dos casos de violência contra idosos acontece dentro de casa*, 2016. <https://www.mdh.gov.br/sdh/noticias/2016/junho/dados-do-disque-100-mostram-quemais-de-80-dos-casos-de-violencia-contraidosos-acontece-dentro-de-casa>, acessado em 13/05/2019.
- [29] Orlóci, L.: *Multivariate analysis in vegetational research*. The Hague: Dr. W. Junk B. V. Publishers, 2^a ed., 1978.
- [30] Quintal, G.: *Análise de clusters aplicada ao Sucesso/Insucesso em Matemática*. Dissertação de mestrado em Matemática, 2006.
- [31] Rao, C.R.: *An advanced statistical method in biometric research*. New York: Ed. John Wiley e Sons, 1952.
- [32] Razera, J., Cenci, C.M.B. e Falcke, D.: *Violência Doméstica e Transgeracionalidade: Um Estudo de Caso*. Revista de Psicologia da IMED, 6(1):47–51, 2014.
- [33] Rohlf, F.J. e Fisher, D.L.: *Test for hierarchical structure in random data sets.*, vol. 17. Washington: Systematic Zoology, 1968.
- [34] Sillito, C.: *Gendered physical e emotional health consequences of situational couple violence for heterosexual married e cohabiting couples*. Feminist Criminology, 7(4):255–281, 2012.
- [35] Silva, A.C.L.G. da, Coelho, E.B.S. e Njaine, K.: *Violência conjugal: as controvérsias no relato dos parceiros íntimos em inquéritos policiais*. Revista Ciência Saúde Coletiva, 19(4):24–38, 2014.
- [36] Sokal, R.R. e Michner, C.D.: *A statistical method for evaluating systematic relationships*. Bulletin of the Society University of Kansas, (38):109–1438, 1958.
- [37] Sokal, R.R. e Rohlf, F.J.: *A The comparison of dendrograms by objective methods*. Taxon, p. 33–40, 1962.
- [38] Souza, E.R. de.: *Masculinidade e violência no Brasil: contribuições para a reflexão no campo da saúde*. Revista Ciência e Saúde Coletiva, 10(1):59–70, 2005.
- [39] Timm, N.H.: *Applied multivariate analysis*. Ney York: Springer, 2002.