



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA**

FERNANDA DE OLIVEIRA LIMA

**UM INIMIGO SILENCIOSO: ÓBITOS POR SUICÍDIO NO ESTADO DA PARAÍBA NO
PERÍODO DE 2015-2019**

CAMPINA GRANDE - PB

2022

FERNANDA DE OLIVEIRA LIMA

**UM INIMIGO SILENCIOSO: ÓBITOS POR SUICÍDIO NO ESTADO DA PARAÍBA NO
PERÍODO DE 2015-2019**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Orientador: Prof. Dr. Edwirde Luiz Silva Camêlo

CAMPINA GRANDE - PB

2022

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

L732i Lima, Fernanda de Oliveira.
Um inimigo silencioso [manuscrito] : óbitos por suicídio no estado da Paraíba no período de 2015-2019 / Fernanda de Oliveira Lima. - 2022.
41 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2022.

"Orientação : Prof. Dr. Edwirde Luiz Silva Camêlo ,
Coordenação do Curso de Estatística - CCT."

1. Mortalidade. 2. Análise de correspondência. 3. Perfil epidemiológico. 4. Suicídio. I. Título

21. ed. CDD 519.5

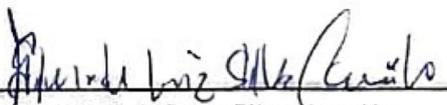
FERNANDA DE OLIVEIRA LIMA

UM INIMIGO SILENCIOSO: ÓBITOS POR SUICÍDIO NO ESTADO DA PARAÍBA NO
PERÍODO DE 2015-2019

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Trabalho aprovado em 02 de Dezembro de 2022.

BANCA EXAMINADORA

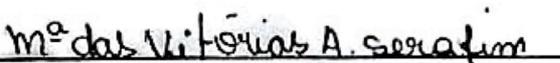


Prof. Dr. Edwirde Luiz Silva Camêlo
(Orientador)

Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba (UEPB)



Profa. Ma. Maria das Vitórias Alexandre
Serafim

Universidade Estadual da Paraíba (UEPB)

À Deus, por mais essa conquista.

À minha família, que não mediram esforços para que eu conseguisse concluir o curso.

AGRADECIMENTOS

Primeiramente agradeço a Deus por me manter forte e me permitir viver mais essa conquista.

Em especial aos meus pais, por nunca medirem esforços para me educar, me ensinarem a ser persistente e não desistir dos meus sonhos, vocês são minha base!

Meus irmãos por serem meus companheiros, confidentes e a melhor parte de mim.

Ao meu namorado, por não soltar minhas mãos, me apoiar em todas as escolhas que faço, me acalmar, não deixar que eu desistisse e me ajudar a seguir sempre e por me presentear com Hellon, que me ajudou muito a desacelerar nos últimos tempos.

Ao meu professor Orientador Edwirde, que sempre foi muito atencioso e me auxiliou muito para realização deste trabalho.

Aos professores Tiago e Vitória, por aceitarem participar da banca examinadora do meu trabalho, vocês são excelentes profissionais e tenho muito orgulho de ser aluna de vocês.

Aos meus amigos Anderson, Bruna, Eduardo, Emanuela, Karolina, Marina, Suziane, Thaynã e Wellington, cada um de vocês participaram de uma etapa da minha vida e sempre serão muito importantes para mim.

Todos da minha família que confiaram em mim, me deram oportunidades e contribuíram com essa conquista.

Agradeço à Universidade Estadual da Paraíba por me acolher tão bem ao longo desses cinco anos e por me permitir a experiência de conviver e assistir aulas dos professores do Departamento de Estatística.

"Tudo tem o seu tempo determinado, e há tempo para todo propósito debaixo do céu."
(Eclesiastes 3:1)

RESUMO

A mortalidade por suicídio é um grande problema para população e continua crescendo na Paraíba, temos variação em função do gênero e cor, sendo assim, esse trabalho tem como objetivo analisar o perfil epidemiológico dos casos de suicídio na Paraíba. No período de 2015 a 2019, coletou-se dados do Sistema de Informação sobre Mortalidade (SIM) e foi analisado segundo as categorias de lesões autoinfligidas intencionalmente (X60 a X84) da Classificação Internacional de Doenças (CID-10), utilizando as variáveis gênero, raça, faixa etária, estado civil, escolaridade e local de residência (por microrregiões e mesorregiões). Foram registradas 1.140 mortes, sendo principalmente de pessoas com idade ativa de trabalho de 20 a 59 anos (73,4%), com maior predominância do sexo masculino (78,0%), com ligeiro predomínio na faixa etária de 30 a 39 anos (20,0%), solteiros (mais de 25%), de cor parda (mais de 60%) com 4 a 7 anos de estudos (6,3%), o principal meio utilizado foi a categoria X70 (enforcamento, estrangulamento e sufocação) com 83,0%. A Análise de Correspondência mostra o comportamento associativo das categorias de duas variáveis qualitativas por vez, onde foi feita a análise referente à faixa etária e ao CID-10, foram sugeridos 4 grupos para as categorias da variável CID-10 e foi observado que a variável idade, com categorias faixas etárias 20 a 29 anos está mais associada com as categorias X61, X62, X64, X70, X71, X77 e X84. No estado da Paraíba, as taxas de suicídio são mais elevadas nas microrregiões de João Pessoa e Campina Grande, portanto, são necessárias estratégias de promoção, prevenção e acompanhamento no âmbito da ação social e dos cuidados de saúde mental, com importantes variações regionais.

Palavras-chaves: Mortalidade. Análise de Correspondência. Perfil epidemiológico.

ABSTRACT

Suicide mortality is a major problem for the population and continues to grow in Paraíba, we have variation according to gender and color, so this work aims to objective to analyze the epidemiological profile of suicide cases in Paraíba/Brazil, in the period from 2015 to 2019, data were collected from the Mortality Information System (SIM) and was analyzed according to the categories of intentionally self-inflicted injuries (X60 to X84) of the International Classification of Diseases (ICD-10), using the variables gender, race, age age, marital status, education, and place of residence (by microregions and mesoregions). 1,140 deaths were recorded, being mainly of people with working age from 20 to 59 years old (73.4%), with a greater predominance of males (78.0%), with a slight predominance in the age group from 30 to 39 years (20.0%), singles (more than 25%), of brown color (more than 60%) with 4 to 7 years of studies (6.3%); the main means used was category X70 (hanging, strangulation and suffocation) with 83.0%. Correspondence Analysis shows the associative behavior of the categories of two qualitative variables at a time, where the analysis of the age group and the ICD-10, groups were suggested for the ICD-10 variable categories, and it was observed that the category of the variable Age Group 20 to 29 years is most associated with the categories X61, X62, X64, X70, X71, X77 and X84. In the state of Paraíba, suicide rates are higher in the micro-regions of João Pessoa and Campina Grande, therefore, promotion, prevention and monitoring strategies are needed in the scope of social action and mental health care, with important regional variations.

Keywords: Mortality. Correspondence Analysis. Epidemiological profile.

LISTA DE ILUSTRAÇÕES

Figura 1	– Óbitos por suicídio por ano segundo o gênero. Paraíba, Brasil, 2015-2019.	27
Figura 2	– Óbitos por suicídio e gênero nas microrregiões de acordo com as mesorregiões. Paraíba, Brasil, 2015-2019.	28
Figura 3	– Porcentagem de óbitos por suicídio, raça e estado civil de acordo com o gênero. Paraíba, Brasil, 2015-2019.	29
Figura 4	– Óbito por suicídio por gênero e ano de acordo com a faixa etária. Paraíba, Brasil, 2015-2019.	30
Figura 5	– Distribuição de óbitos por suicídio nas microrregiões. Paraíba, Brasil, 2015-2019.	31
Figura 6	– Dispersão de CID-10 com Faixa etária	32
Figura 7	– Dispersão com a contribuição variável faixa etária	33
Figura 8	– Associação dos CID-10 com todas as variáveis faixa etária	33
Figura 9	– Clusters baseados em coordenadas de Análise de Correspondência no espaço de dimensionalidade 2.	34
Figura 10	– Agrupamento hierárquico	35
Figura 11	– Teste de Malinvaud para a significância das dimensões da Análise de Correspondência	35
Figura 12	– Valores de correlação de cada variável faixa etária com a dimensão 2	36
Figura 13	– Dispersão da inércia	36

LISTA DE TABELAS

Tabela 1 – Tabela de Contingência	17
Tabela 2 – Óbitos por suicídio de acordo com as lesões autoprovocadas intencionalmente CID-10 (X60 a X78) e gênero. Paraíba, Brasil, 2015-2019.	26
Tabela 3 – Continuação da tabela: Óbitos por suicídio de acordo com as lesões auto- provocadas intencionalmente CID-10 (X79 a X84) e gênero. Paraíba, Brasil, 2015-2019.	27

SUMÁRIO

1	INTRODUÇÃO	11
2	METODOLOGIA	13
2.1	Análise de Correspondência	13
2.1.1	<i>Principais Conceitos da Análise de Correspondência</i>	15
2.1.2	<i>Desenvolvimento Algébrico da ANACOR</i>	17
2.1.2.1	<i>Tabela de Contingência</i>	17
2.1.2.2	<i>Matriz de Correspondência</i>	17
2.1.2.3	<i>Vetores (Massa)</i>	19
2.1.2.4	<i>Matriz Diagonal</i>	20
2.1.2.5	<i>Posto (Rank)</i>	21
2.1.2.6	<i>Inércia</i>	21
2.1.3	<i>Distância</i>	23
3	APLICAÇÃO	25
3.1	Materiais	25
3.2	Análise Exploratória	25
3.3	Análise de Correspondência	31
4	CONCLUSÃO	37
	REFERÊNCIAS	38

1 INTRODUÇÃO

O suicídio constitui-se em um dos mais antigos temas com relação à saúde dos indivíduos e à forma como são afetados pelas sociedades e coletividades em que vivem. Em termos históricos, sua relevância no plano social pode ser identificada desde a Grécia antiga. Em tempos atuais, desde o século XVIII, tem sido tratado como fenômeno social e segundo perspectivas sociológicas, históricas, econômicas e filosóficas (RIBEIRO; MOREIRA, 2018).

A definição para o suicídio pode ser entendida como o resultado de comportamentos autoinfligidos que têm como intenção a própria morte (KAIRI; LEO, 2014). Essa forma de violência autoinfligida resulta da interação de múltiplos fatores: psicológicos, biológicos, socio-culturais e ambientais como mostra (CEVS, 2018). Entretanto, na maioria das vezes, as tentativas e mortes por suicídio estão ligados a problemas de saúde mental Cavanagh et al. (2003) que, se diagnosticados e tratados apropriadamente, podem levar a uma diminuição das ocorrências. A depressão, por exemplo, em diferentes idades apresenta uma grande relação com as tentativas de suicídio (CHACHAMOVICH et al., 2009; CAVALCANTE; MINAYO; MANGAS, 2013; TEISMANN et al., 2018; ROCA et al., 2019).

No mundo, mais de 800.000 pessoas cometem suicídio todos os anos, representando uma morte a cada 40 segundos. É a segunda principal causa de morte na faixa etária de 15 a 29 anos OMS (2018), assim, este fenômeno tornou-se um grave problema de saúde pública em muitos países e não ocorrendo apenas em países de alta renda, pois é um fenômeno em todas as regiões do mundo. Em 2019, 70% desses casos aconteceram em países com vulnerabilidade social (WHO, 2019).

No Brasil, em 2016 o número de suicídios aumentou 2,3% em 1 ano comparado ao ano 2015, ocorreu 1 caso a cada 46 minutos, e tornou-se a quarta causa de morte entre jovens de 15 a 29 anos (BRASIL, 2017). Embora o governo estime que o número seja ainda maior devido à subnotificação dos registros. No *ranking* de países com maior incidência de suicídios, o Brasil ocupava a oitava posição ultrapassando o número de 12 mil casos anuais de mortes autoprovocadas intencionalmente (OMS, 2018).

No Nordeste, as notificações de suicídio na região aparecem de forma preocupante, entre 2003 e 2013, os estados da Paraíba, Piauí e Sergipe, mais que triplicaram seus quantitativos (CECCON et al., 2014). De acordo com Silva et al. (2021), a região Nordeste apresentou crescimento significativo da mortalidade por suicídio em 2008 a 2018.

De acordo com o Sistema de Informação sobre Mortalidade da Secretaria Estadual de Saúde (SIM/SES) a Paraíba registrou um caso de suicídio, a cada 34 horas. O dado foi divulgado pelo Ministério Público da Paraíba, (BRASIL, 2019). Conforme estudo realizado por Lavor et al. (2020), em que analisou dados de 2013 a 2017, observou que o número de mortes causadas por autoeliminação voluntária no estado vem aumentando a cada ano.

A análise de correspondência segundo Whitlark e Smith (2001) é uma técnica que apresenta as associações entre um conjunto de variáveis categóricas em um mapa perceptual e

pela análise de Greenacre (1984), ela mostrar as linhas e colunas de uma matriz de dados em uma tabela de contingência de dupla entrada, como pontos em um espaço multidimensional de vetores e por Olariaga e Hernández (2000), sendo I e J dois conjuntos finitos, seu produto cartesiano é $I \times J$, para definir uma correspondência entre I e J , consistem em associar a cada elemento (i,j) de $I \times J$ um número não negativo $K(i,j)$. Assim, a correspondência é representada por meio de uma tabela retangular.

Considerando o suicídio um importante problema de saúde pública ainda pouco estudado no estado da Paraíba, o presente estudo tem como objetivo analisar o perfil epidemiológico dos casos de suicídio e as associações das variáveis na Paraíba, de 2015 a 2019, por meio de análise exploratória dos dados, análise de correspondência e utilização do pacote CAinterprTools.

2 METODOLOGIA

2.1 Análise de Correspondência

Segundo Whitlark e Smith (2001), a análise de correspondência é uma técnica que mostra as associações entre um conjunto de variáveis categóricas não métricas em um mapa perceptual, possibilitando, assim, um exame visual de qualquer padrão ou estrutura nos dados. Como mostram Batista, Escuder e Pereira (2004), a análise de correspondência é uma técnica de representação gráfica em projeção plana das relações multidimensionais das distâncias Qui-quadrado (χ^2) entre as categorias das variáveis estudadas. As categorias que possuem localização próxima na projeção plana, têm relação mais forte do que categorias separadas por distâncias maiores, e uma categoria (representada como um ponto no mapa perceptual), pode ser analisada em separado e caracterizada segundo a proximidade das projeções de todas as outras categorias sobre uma reta que ligue seu ponto característico à origem dos eixos do plano de projeção.

Quando categorias de uma variável se encontram em posições próximas no mapa da análise de correspondência, mostra que elas podem ser consideradas iguais no que tange à distribuição de massas do total das observações realizadas. Os primeiros artigos que mostraram a noção de análise de correspondência inicia em 1930, a derivação algébrica da Análise de Correspondência (ANACOR) é mostrada por Hirschfeld (1935), que desenvolveu a formulação da relação entre linhas e colunas de uma tabela de contingência de dupla entrada. Porém, segundo Beh (2004), esta técnica teve origem do analista francês Jean-Paul Benzécri no ano de 1960 e representa uma aplicação de análise multivariada para a exibição de linhas e colunas de uma matriz de dados como pontos em um espaço dimensional que seja qualitativo (GREENACRE, 1984).

A análise de correspondência é uma técnica multivariada que vem se popularizando para a redução dimensional e o mapeamento perceptual. Entende-se por mapa perceptual a representação visual das percepções de objetos de um indivíduo em duas ou mais dimensões e, normalmente, esse mapa tem níveis opostos de dimensões nos extremos dos eixos X e Y. A ANACOR é uma das técnicas recentemente desenvolvidas para analisar relações não lineares de dados com respostas categóricas e tem como objetivo principal agrupar variáveis altamente associadas, e tendo como consequência a redução do número de variáveis preditoras do modelo e a representação das relações entre as categorias das variáveis em um mapa perceptual, conforme é descrito por (HAIR, 2005).

A técnica reside no fato de fornecer um meio para examinar as relações entre as variáveis em linha ou em colunas individualmente, e entre as variáveis em linha e em coluna conjuntamente, o que implica que seja possível comparar associações entre as categorias das variáveis que foram utilizadas. A análise de correspondência é uma técnica que é utilizada para mostrar as linhas e colunas de uma matriz de dados em uma tabela de contingência de dupla entrada, como pontos em um espaço multidimensional de vetores, de acordo com (GREENACRE, 1984).

O método consiste em duas etapas, referentes ao cálculo da medida de associação e à criação do mapa perceptual. A Análise de Correspondência usa o teste Qui-quadrado (χ^2) para deixar os valores das frequências padronizados e formar a base para as associações. A partir de uma tabela de contingência, as frequências esperadas e o valor do (χ^2) são calculados para cada célula, considerando-se as diferenças entre as frequências observadas e as esperadas. Assim, com as medidas padronizadas da associação, a ANACOR gera uma medida em distância e cria projeções ortogonais, as quais as categorias podem ser alocadas, de forma a representar o grau de associação dado pelas distâncias (χ^2) em um espaço dimensional.

Segundo Pestana e Gageiro (2005), é recomendado inicialmente a realização do teste (χ^2) para verificar a existência de dependência entre as duas variáveis, e para avaliação da adequação da aplicação da ANACOR.

É indicado um roteiro para a realização da ANACOR, por meio do autovalor (*eigenvalue*) e das inércias parciais e acumuladas de cada dimensão, julga-se a pertinência de considerá-la para o modelo. Para cada dimensão, o quadrado de cada autovalor é chamado de inércia das dimensões e mede a importância de cada dimensão. O quociente entre a inércia de cada dimensão e a inércia total dá a proporção da variância explicada pela dimensão.

O número máximo de dimensões (eixos nos gráficos) que pode ser estimado é um a menos do que o menor número entre a quantidade de linhas ou de colunas. Em uma tabela de contingência com 3 colunas e 5 linhas, o número máximo de dimensões será [mín(linha,coluna) -1]. Depois de determinada a dimensionalidade, os resultados podem ser examinados em uma representação gráfica, chamada de mapa perceptual.

A análise desse gráfico se faz pelas relações de proximidade geométrica e por projeções em dimensões que podem ser identificadas a partir de pontos no plano, as categorias mais explicativas das dimensões são as que apresentam maior inércia por dimensão e que também se situam mais afastadas da origem (0,0).

Segundo Batista, Escuder e Pereira (2004), para possuir uma boa interpretação dos dados, devemos entender que o plano de análise desta técnica tem natureza descritiva, não comportando inferências de causa e efeito. O teste Qui-quadrado (χ^2) e a análise de resíduos comparam o distanciamento entre as observações realizadas e esperadas por simples aleatoriedade. A análise de correspondência oferece informações de contraste entre relações de categorias de variáveis contingenciadas, então, uma relação mais forte entre duas categorias em comparação com outras relações não pressupõe efeitos de uma sobre a outra.

É mostrado por Hair (2005) que a análise de correspondência é uma técnica de interdependência a qual vem se tornando cada vez mais popular para a redução dimensional e o mapeamento perceptual. Sua aplicação mais direta é representar a "correspondência" de categorias de variáveis, em particular aquelas medidas em escalas qualitativas.

Assim, Deville e Saporta (1983) acredita que a análise de correspondência é provavelmente o mais popular método multidimensional de análise de dados na França, e para Benzécri e sua escola, a ANACOR tornou-se sinônimo de análise de dados. Para os autores, as razões para

o sucesso são muitas, e destaca-se o poder da representação gráfica.

Autores como Benzécri, Volle e Foucart, acreditam no aprofundamento matemático para que se alcance considerável competência na aplicação prática. Entretanto, os desenvolvimentos matemáticos que esses autores apresentam (cálculo tensorial, espaço dual) são inacessíveis à maioria dos pesquisadores das Ciências Sociais que os aplicam.

Segundo Olariaga e Hernández (2000) e sendo I e J dois conjuntos finitos, seu produto cartesiano é $I \times J$. Para definir uma correspondência entre I e J , devemos consistir em associar a cada elemento (i,j) de $I \times J$ um número não negativo $K(i,j)$. Assim, a correspondência é representada por meio de uma tabela retangular.

Didaticamente, o mais clássico é lidar com uma tabela de contingência, onde está a chave, ela pode ser aplicada a qualquer tabela de números positivos. Se todos os valores $K(i,j)$ forem inteiros, trata-se de uma correspondência estatística, pois os números indicam quantas vezes o elemento (i,j) ocorre, onde essas tabelas são conhecidas como tabelas de contingência. Se o $K(i,j)$ são números entre 0 e 1 e, conseqüentemente, probabilidades, trata-se de uma correspondência probabilística. É possível analisar tabelas de medidas, tabelas de intensidade, em que $K(i,j)$ é uma nota de mérito de i no aspecto j , tabelas de preferência, em que cada i ordena os elementos da tabela de J , tabela de Burt.

Desmembrar uma tabela suficientemente grande, que seja inacessível à inspeção visual, em outras tabelas de fácil compreensão, é um dos problemas mais comuns que surgem. Pretende-se também representar a informação fundamental de uma grande tabela em poucas dimensões e se possível em duas, pois isso dá origem a uma representação gráfica plana e assim, acessível e ilustrativa. Outra opção seria transformar uma informação sem métricas em uma sob fatores ortogonais e com métricas euclidianas para que, seja possível classificar os assuntos. Caso não seja estabelecido um objetivo muito específico, a análise de correspondência solicita uma classificação baseada nas coordenadas fatoriais, e assim, uma redução da tabela original.

De acordo com Fávero et al. (2009), esses são alguns exemplos de aplicações:

- Associação de tipos de empresas com porte.
- Associação de *ratings* de crédito concedidos, como, por exemplo, o da *Standard & Poor's*, e empresas provenientes de diversos setores da economia.
- Relacionamento dos segmentos de atuação com características qualitativas.
- Avaliação de marcas, produtos ou serviços em função de atributos e benefícios representados de forma conjunta.
- Estudo de posicionamento de organizações com os respectivos portes.

2.1.1 Principais Conceitos da Análise de Correspondência

A ANACOR pode ser utilizada como um teste complementar ao Qui-quadrado, sendo necessária que a hipótese nula de igualdade de frequências seja rejeitada, ou seja, é preciso

rejeitar a hipótese de que as variáveis são independentes ou combinam-se de forma aleatória. É uma técnica de interesse quando possuímos muitas categorias em linha ou em coluna, pois representa graficamente a relação entre elas, por meio de um mapa perceptual, onde as categorias mais semelhantes ficam dispostas próximas umas das outras.

Na utilização dessa técnica, "massa" representa uma medida que indica a influência de um objeto com base em sua frequência marginal, ou seja, representa o quanto do total de cada categoria está presente no estudo (vertical e horizontal). A massa afeta o centroide, que é uma média ponderada do perfil da linha e da coluna. Os pontos com maior massa puxam o centroide linha (coluna) fortemente para sua posição (FÁVERO et al., 2009).

Com o estudo de Pestana e Gageiro (2005), entendemos que podemos obter as coordenadas de cada categoria no gráfico pelos escores, que são quantificações ótimas atribuídas às linhas e colunas de cada dimensão. A primeira dimensão é a que mais explica a variação existente nos dados, sendo a segunda dimensão a seguinte e, assim, sucessivamente. Os escores nas dimensões representam as coordenadas dos pontos linhas e dos pontos colunas na representação gráfica e dependem do método de normalização que é escolhido.

O número de dimensões que forem selecionadas está relacionado ao nível geral de variância explicada desejada e na explicação adicional dada pelo acréscimo de outra dimensão. Segundo Hair (2005), “uma regra prática é que as dimensões com inércia (autovalores) maior que 0,2 devem ser incluídas na análise”.

O *singular value* (valor singular, autovalor *eigenvalue*) corresponde ao coeficiente da correlação R de Pearson entre os escores das duas variáveis (em linha e em coluna), ou seja, mostra variância total explicada em cada dimensão ou também restituída pelos eixos. Para cada dimensão, a inércia corresponde ao quadrado do *eigenvalue*, ou valor singular, sendo os maiores valores referentes às categorias dominantes naquela dimensão.

Em termos de distância, pode-se interpretar a ANACOR dependendo do método de normalização utilizado. A normalização é usada para distribuir a inércia pelos escores em linha, assim como em coluna. Segundo Fávero et al. (2009), os métodos usuais de normalização são:

- Simétrica (*Symmetrical*): a inércia é distribuída simetricamente pelos escores das linhas e das colunas. É usado quando se busca examinar as diferenças ou similaridades entre as duas variáveis;
- Principal (*Principal*): utilizado quando há interesse nas semelhanças ou diferenças entre as categorias de cada variável por si só, e não nas diferenças entre as duas variáveis (não avalia a relação conjunta);
- Em Linhas (*Row principal*): usado quando se busca verificar como as categorias da variável em linha variam entre si. Ou seja, maximiza as distâncias entre as categorias das linhas;
- Em Colunas (*Column principal*): usado quando se busca verificar como as categorias da variável em coluna variam entre si. Maximiza as distâncias entre as categorias das colunas.

As normalizações simétrica e principal distribuem a inércia simetricamente entre as categorias em linha e em coluna, não maximizando as distâncias, como ocorre nas normalizações *Row principal* e *Column principal* (PESTANA; GAGEIRO, 2005). A escolha da normalização afeta diretamente o cálculo dos escores linha e coluna e das variâncias.

2.1.2 Desenvolvimento Algébrico da ANACOR

2.1.2.1 Tabela de Contingência

Uma tabela de contingência $I \times J$ (conhecida como tabela de correspondência, tabela cruzada, *contingency table*, *correspondence table*, *cross-tabulation*) e pode ser representada de acordo com a Tabela 1 (FÁVERO et al., 2009):

Tabela 1 – Tabela de Contingência

Variável	1	2	...	J	Total
1	n_{11}	n_{12}	...	n_{1J}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2J}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
I	n_{I1}	n_{I2}	...	n_{IJ}	$n_{I.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.J}$	$n_{..=n}$

Fonte: Produzida por Luiz Paulo Fávero (2009).

A tabela de contingência apresentada possui $1 \dots I$ categorias de linhas da variável A e $1 \dots J$ categorias de colunas da variável B. Assim, n_{ij} representa o número de elementos que pertencem à categoria i da variável A e à categoria j da variável B.

2.1.2.2 Matriz de Correspondência

Em seguida, é calculada a matriz de correspondência, que corresponde à matriz P de dimensão $I \times J$ das frequências relativas (proporções). Cada elemento será dividido por n , a soma dos elementos de P passa a ser igual a 1 e, neste caso, temos (FÁVERO et al., 2009):

$$p_{ij} = \frac{n_{ij}}{n}, \quad P = \frac{1}{n}(N).$$

Com o estudo de Olariaga e Hernández (2000), K é definido por:

$$(i, j) \rightarrow K(i, j).$$

Como uma correspondência entre os elementos i e j dos conjuntos I e J, e em que os elementos $K(i, j)$ são números não negativos.

$$K_{IJ} = \{K(i, j), i \in I, j \in J\}.$$

A distribuição marginal por linhas é: $K(i) = \sum\{K(i, j), j \in J\}$.

A distribuição marginal por colunas é $K(j) = \sum\{K(i, j), i \in I\}$.

A soma das frequências de todas as células é:

$$K = \sum\{K(i, j), i \in I, j \in J\} = \sum\{K(i), i \in I\} = \sum\{K(j), j \in J\}.$$

Usando os valores $K(i, j) / K(i)$, $i \in I$ como coordenadas da linha i , devido a:

$$\sum \left\{ \frac{K(i, j)}{K(i)}, j \in J \right\} = 1.$$

Cada linha tem a possibilidade de ser geometricamente representável, já que:

$$\sum \left\{ \frac{K(i, j)}{K(i)}, j \in J \right\}.$$

Então, será um ponto do hiperplano da equação:

$$\sum \left\{ \frac{K(i, j)}{K(i)}, j \in J \right\} = 1.$$

Podemos raciocinar com as colunas no sistema de coordenadas definido pelas linhas. Podemos utilizar esses perfis de quocientes e usar a notação:

$$f_j^i = \frac{K(i, j)}{K(i)}.$$

$$f_i^j = \frac{K(i, j)}{K(j)}.$$

Portanto, um perfil de linha é $f_j^i = \{f_j^i, j \in J\}$.

Um perfil de coluna é $f_i^j = \{f_i^j, i \in I\}$.

Cada perfil é uma distribuição condicional, onde f_j^i é considerada a distribuição de J condicional em i , e f_i^j é a distribuição de I condicional em j . Os perfis serão as coordenadas dos pontos; assim, a matriz de dados que é objeto direto de análise não será K_{IJ} , mas sim em frequências relativas, que irá se chamar F , onde elemento genérico é designado por f_{ij} .

As frequências relativas serão:

$$f_{ij} = K(i, j)/K \text{ e } f_{IJ} = \{f_{ij}, i \in I, j \in J\}.$$

Assim, verificamos que:

$$f_j^i = \frac{f_{ij}}{f_i} \text{ com } f_i \neq 0.$$

$$f_i^j = \frac{f_{ij}}{f_j} \text{ com } f_j \neq 0.$$

2.1.2.3 Vetores (Massa)

É necessário calcular a matriz de perfil das linhas (r) e das colunas (c). Primeiramente, é preciso que sejam obtidos os vetores de frequências relativas marginais, que também são denominados de massa, para cada célula da tabela de contingência (FÁVERO et al., 2009).

$$r_i = \sum_{j=1}^J p_{ij} = \sum_{j=1}^J \frac{n_{ij}}{n}.$$

$$c_j = \sum_{i=1}^I p_{ij} = \sum_{i=1}^I \frac{n_{ij}}{n}.$$

A análise de correspondências simples será caracterizada pela ponderação que os denominadores das expressões dos perfis implicam.

As massas dos pontos de linha e coluna são $K(i)$ e $K(j)$, ou em termos relativos ou probabilísticos representados por:

$$f_i = \frac{K(i)}{K} \forall i \in I.$$

$$f_j = \frac{K(j)}{K} \forall j \in J.$$

Assim, os sistemas de massa são:

$$f_I = \{f_i, i \in I\}.$$

$$f_J = \{f_j, j \in J\}.$$

Então, as duas nuvens de pontos são:

$$N(I) = \{(f_j^i, f_i), i \in I\}.$$

$$N(J) = \{(f_i^j, f_j), j \in J\}.$$

Os baricentros ou centros de gravidade serão representados respectivamente:

$$N(I) : \sum \{f_i \times f_j^i, i \in I\} = f_J \equiv i_g.$$

$$N(J) : \sum \{f_j \times f_i^j, j \in J\} = f_I \equiv j_g.$$

Assim, o centro de gravidade da nuvem $N(I)$ ou o perfil médio das linhas i será o sistema de massas do conjunto J e o da nuvem $N(J)$ é o sistema de massas de o conjunto I .

2.1.2.4 Matriz Diagonal

A matriz de perfil das linhas (*row*) é composta da matriz diagonal dos elementos r_i :

$$D_r = \begin{vmatrix} r_1 & 0 & 0 & 0 \\ 0 & r_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & r_J \end{vmatrix}.$$

E a matriz de perfil das colunas (*columns*) é dada pela matriz diagonal dos elementos c_j por:

$$D_c = \begin{vmatrix} c_1 & 0 & 0 & 0 \\ 0 & c_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & c_I \end{vmatrix}.$$

Para a decomposição ser feita em coordenadas principais, seguimos o que foi proposto por Mingoti (2005). Considerando a matriz \tilde{P} , dada por:

$$\tilde{P} = P - rc'.$$

Essa matriz possui entrada do tipo $p_{ij} - \frac{n_i}{n} \times \frac{n_j}{n}$ e, busca-se comparar as proporções observadas com as esperadas, com o objetivo de verificar a independência das variáveis A e B .

2.1.2.5 Posto (Rank)

O posto (*rank*) \tilde{P} é dado por K , sendo $K = \min(I-1, J-1)$ (FÁVERO *et al.*, 2009).

A matriz \tilde{P} pode ser decomposta em seus valores singulares:

$$\tilde{P} = A \Lambda B'$$

Sendo decomposta, temos que as colunas da matriz A definem as coordenadas dos eixos para os pontos representados no perfil coluna de P . Similarmente, as colunas da matriz B' definem as coordenadas dos eixos para os pontos representados no perfil linha de P .

Em que A e B são matrizes de dimensão $i \times k$ e $j \times k$, respectivamente, sendo:

$$A = D_r^{1/2} U_{I \times K}.$$

$$B = D_c^{1/2} V_{J \times K}.$$

$E \Lambda$ é a matriz $K \times K$ que possui os autovetores da matriz \tilde{P} . U e V são matrizes ortogonais. Utilizando a decomposição em valores singulares, a matriz \tilde{P} pode ser expressa da seguinte maneira:

$$\tilde{P} = P - rc' = \sum_{i=1}^k \hat{\lambda}_i \tilde{a}_i \tilde{b}_i.$$

Em que \tilde{a}_i denota a i -ésima coluna da matriz A e \tilde{b}_i representa a i -ésima coluna da matriz B . $K = \text{posto}(\tilde{P}) = \min(I-1, J-1)$. (MINGOTI, 2007).

As coordenadas principais das linhas da matriz \tilde{P} são definidas como:

$$Y_{i \times k} = D_r^{-1} A_{ik} \Lambda_{k \times k}.$$

E as coordenadas principais das colunas da matriz \tilde{P} são:

$$Z_{j \times k} = D_c^{-1} B_{jk} \Lambda_{k \times k}.$$

Mingoti (2007) explica que as primeiras duas coordenadas das linhas e das colunas são as que contêm os pares que melhor representam os dados, em termos de associação entre as variáveis A e B , já que estão diretamente relacionados aos maiores autovalores da matriz \tilde{P} .

2.1.2.6 Inércia

O conceito de inércia é um ponto importante, segundo Johnson e Wichern (2007), a inércia total é definida como a soma dos quadrados de todos os valores singulares ou autovalores não nulos. Assim, temos:

$$\sum_{i=1}^k \lambda_i^2.$$

Em que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ são os elementos não nulos da matriz Λ . A inércia de cada dimensão e a inércia total possuem uma relação que resulta na proporção de variância explicada pela dimensão. Assim:

$$\frac{\lambda_i^2}{\sum_{i=1}^k \lambda_i^2}.$$

Sendo assim, a inércia é igual ao quadrado do valor singular ou autovalor, para cada dimensão.

É mostrado por Johnson e Wichern (2007) que a lógica da ANACOR está relacionada à análise estatística do Qui-quadrado em uma tabela de contingência.

Sobre o processo da ANACOR, Fávero, Belfiore e Fouto (2006) mostram que o método consiste em duas etapas básicas, referentes ao cálculo da medida de associação e à criação do mapa perceptual. A ANACOR usa o teste Qui-quadrado (χ^2) para fazer a padronização dos valores das frequências e formar a base para as associações. A partir de uma tabela de contingência, calculam-se as frequências esperadas e o valor do χ^2 para cada célula, considerando as diferenças entre as frequências observadas e as esperadas. Assim, com as medidas padronizadas da associação, a ANACOR cria uma medida em distância métrica e cria projeções ortogonais sobre as quais as categorias podem ser alocadas, de forma a representar o grau de associação dado pelas distâncias χ^2 em um espaço dimensional.

A estatística Qui-quadrado (χ^2) para a mensuração do grau de associação entre linhas e colunas em uma tabela de contingência é dada por:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

Em que $O_{ij} = n_{ij}$ representa o número de observações reais na célula (i,j), e $E_{ij} = np_{ij}$ representa o número esperado de observações na célula (i,j), sendo a hipótese nula o fato de que as duas variáveis A e B são independentes, então, temos que não há associação entre suas categorias.

Assim, pode-se reescrever a equação da inércia, por meio da substituição da expressão da estatística Qui-quadrado:

$$\sum_{i=1}^k \lambda_i^2 = \frac{\chi^2}{n} = \frac{1}{n} \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

Para encontrar os eixos principais, ou inércia máxima de N(I) e N(J), onde a inércia ou variância, é a soma ponderada da massa de cada ponto vezes sua distância do ponto de referência ao quadrado, como representado abaixo:

$$Inercia_{i_g}[N(I)] = \sum\{f_i \times d^2(i, i_g), i \in I\} = \sum\left\{(f_{ij} - f_i f_j)^2 \frac{1}{f_i f_j}, i \in I, j \in J\right\} = \frac{\chi^2}{K}.$$

$$Inercia_{j_g}[N(J)] = \sum\{f_j \times d^2(j, j_g), j \in J\} = \sum\left\{(f_{ij} - f_i f_j)^2 \frac{1}{f_i f_j}, i \in I, j \in J\right\} = \frac{\chi^2}{K}.$$

Essas são as inércias ou variâncias das nuvens $N(I)$ e $N(J)$ relacionadas aos baricentros correspondentes i_g e j_g . O valor que normalmente avalia a distância entre a distribuição empírica e a esperada sob condição de independência em uma tabela de contingência é o χ^2 .

Os valores das inércias em relação à origem são (OLARIAGA; HERNÁNDEZ, 2000):

$$Inercia_o[N(I)] = \sum\{f_i \times d^2(i, O), i \in I\} =$$

$$\sum\left\{f_i (f_j^i)^2 \frac{1}{f_j}, i \in I, j \in J\right\} = \sum\left\{\frac{f_{ij}^2}{f_i f_j}, i \in I, j \in J\right\}.$$

$$Inercia_o[N(J)] = \sum\{f_j \times d^2(j, O), j \in J\} =$$

$$\sum\left\{f_j (f_i^j)^2 \frac{1}{f_i}, i \in I, j \in J\right\} = \sum\left\{\frac{f_{ij}^2}{f_i f_j}, i \in I, j \in J\right\}.$$

2.1.3 Distância

Para entender a relação entre $N(I)$ e $N(J)$, as distâncias a seguir serão definidas:

entre elementos de $N(I)$, no espaço \mathbb{R}^p :

$$d^2(i, i') = \sum\left\{\left(f_j^i - f_j^{i'}\right)^2 \frac{1}{f_j}, j \in J\right\} \text{ com } i, i' \in I,$$

entre os elementos de $N(J)$, no espaço \mathbb{R}^n :

$$d^2(j, j') = \sum\left\{\left(f_i^j - f_i^{j'}\right)^2 \frac{1}{f_i}, i \in I\right\} \text{ com } j, j' \in J.$$

Dividir por f_j atribui a todas as colunas a mesma importância, ou ainda, tomar como referência o perfil médio f_j . O mesmo pode ser aplicado ao denominador f_i , em termos de balanceamento das linhas, onde f_i é o perfil médio de referência para as colunas.

A distância euclidiana entre as linhas i e i' que é obtida em $K(i, j)$ reflete apenas a diferença entre as frequências marginais $K(i)$ e $K(i')$, assim, existiria o manifesto de um efeito de tamanho. A distância euclidiana entre os perfis de i e i' reflete sua semelhança. Essa distância, com a ponderação por colunas, origina a uma distância que tem a propriedade de equivalência distribucional e oferece a possibilidade de estabelecer relações de transição.

Essa distância, chamada χ^2 , tem a propriedade de equivalência distribucional, em que a distância entre as linhas não se altera caso as colunas j e j' com perfis semelhantes forem mescladas. Essa propriedade é notável contra a arbitrariedade da codificação, garantindo assim a robustez, já que a informação não é obtida pela decomposição de uma classe em subclasses homogêneas, e também não é perdida pela fusão de classes homogêneas em outra.

3 APLICAÇÃO

3.1 Materiais

Neste trabalho, realizou-se um estudo descritivo, investigando-se o perfil das pessoas que cometeram suicídio, por gênero, raça, faixa etária, estado civil, escolaridade e local de residência (por microrregiões e mesorregiões) e também uma análise de correspondência com as categorias das variáveis CID-10 e faixa etária para observar a associação entre elas, os dados são do período de 2015 a 2019 no estado da Paraíba.

Os dados utilizados neste trabalho foram coletados das estatísticas de óbitos causadas por lesões autoprovocadas intencionalmente de acordo com as causas X60 a X84 da 10ª Classificação Internacional de Doenças (CID-10) pelo Sistema de Informações de Mortalidade (SIM) em março de 2021. Posteriormente, foram analisados no programa estatístico R Team (2021), sendo os pacotes utilizados ggplot2 Wickham (2016) para criação de gráficos baseado na gramática de gráfico, *leaflet* Cheng et al. (2021) que usa uma biblioteca JavaScript para criação de mapas interativos e o CAinterprTools Alberti (2020) para criação dos gráficos de análise de correspondência. Com o pacote CAinterprTools, foram utilizadas as funções: caPlot, caScatter, caCluster, table.collapse, malinvaud, cols.corr e sig.dim.perm. Na Paraíba, o número de óbitos por suicídio no período de 2015 a 2019 foi de 1.140 mortes. A Tabela 2 e Tabela 3 apresentam o quantitativo e percentual por tipo de lesão autoprovocadas intencionalmente (o suicídio) de acordo com o gênero.

3.2 Análise Exploratória

Conforme a Tabela 2 e Tabela 3, temos que 78% dos óbitos por suicídio correspondem ao gênero masculino, com média de 177,8 óbitos por ano, 22,0% são do gênero femininos com média de 50,2. Estes resultados estão em conformidade com os apontados para o Brasil segundo Machado e Santos (2015), outros estados e no Distrito Federal Macente e Zandonade (2011), Bando et al. (2012), Sehnem e Palosqui (2014), Carmo et al. (2018), Baére e Zanello (2018), Beringuel et al. (2020) e também conforme a literatura internacional (ROMERO-PIMENTEL et al., 2018; NUNES, 2018; KOOTBODIEN et al., 2020).

Majoritariamente o tipo de lesão que possui mais frequência entre os óbitos por suicídio foi CID-10 (X70), ou seja, 73,33% ocorrem por enforcamento, estrangulamento e sufocação, sendo 696 (83%) homens e 140 (17%) mulheres. Em segundo lugar, o tipo de lesão mais comum são as por pesticidas (X68) com 9,74% (59% para homens e 41% para mulheres). Entre os suicídios auto cometidos por homens, também é recorrente o disparo de arma de fogo (X72 e X74), com 82% e 90%, respectivamente.

Tabela 2 – Óbitos por suicídio de acordo com as lesões autoprovocadas intencionalmente CID-10 (X60 a X78) e gênero. Paraíba, Brasil, 2015-2019.

Categoria CID-10	Masc.(%)	Fem.(%)	Total(%)
Auto-intoxicação por e exposição, intencional			
Analgésicos, antipiréticos e anti-reumáticos, não-opiáceos (X60)	0 (0)	1 (100)	1 (0,09)
Drogas anticonvulsivantes [antiepilépticos] sedativos, hipnóticos, antiparkinsonianos e psicotrópicos não classificados em outra parte (X61)	3 (60)	2 (40)	5 (0,44)
Narcóticos e psicodislépticos [alucinógenos] não classificados em outra parte (X62)	1 (100)	0 (0)	1 (0,09)
Outras substâncias farmacológicas de ação sobre o sistema nervoso autônomo (X63)	1 (20)	4 (80)	5 (0,44)
Outras drogas, medicamentos e substâncias biológicas e às não especificadas (X64)	13 (39)	20 (61)	33 (2,89)
Auto-intoxicação			
Voluntária por álcool (X65)	9 (100)	0 (0)	9 (0,79)
Intencional por solventes orgânicos, hidrocarbonetos halogenados e seus vapores (X66)	2 (100)	0 (0)	2 (0,18)
Intencional por outros gases e vapores (X67)	0 (0)	1 (100)	1 (0,09)
Por e exposição, intencional, a pesticidas (X68)	65 (59)	46 (41)	111 (9,74)
Por e exposição, intencional, a outros produtos químicos e substâncias nocivas não especificadas (X69)	10 (59)	7 (41)	17 (1,49)
Lesão autoprovocada intencionalmente			
Enforcamento, estrangulamento e sufocação (X70)	696 (83)	140 (17)	836 (73,33)
Afogamento e submersão (X71)	2 (100)	0 (0)	2 (0,18)
Disparo de arma de fogo de mão (X72)	18 (82)	4 (18)	22 (1,93)
Disparo de espingarda, carabina, ou arma de fogo de maior calibre (X73)	1 (50)	1 (50)	2 (0,18)
Disparo de outra arma de fogo e de arma de fogo não especificada (X74)	26 (90)	3 (10)	29 (2,54)
Pela fumaça, pelo fogo e por chamas (X76)	4 (57)	3 (43)	7 (0,61)
Vapor de água, gases ou objetos (X77)	1 (100)	0 (0)	1 (0,09)
Objeto cortante ou penetrante (X78)	5 (83)	1 (17)	6 (0,53)

Obs.: As categorias Dispositivos explosivos (X75) e outros meios especificados (X83) não registraram óbitos.

Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Tabela 3 – Continuação da tabela: Óbitos por suicídio de acordo com as lesões autoprovocadas intencionalmente CID-10 (X79 a X84) e gênero. Paraíba, Brasil, 2015-2019.

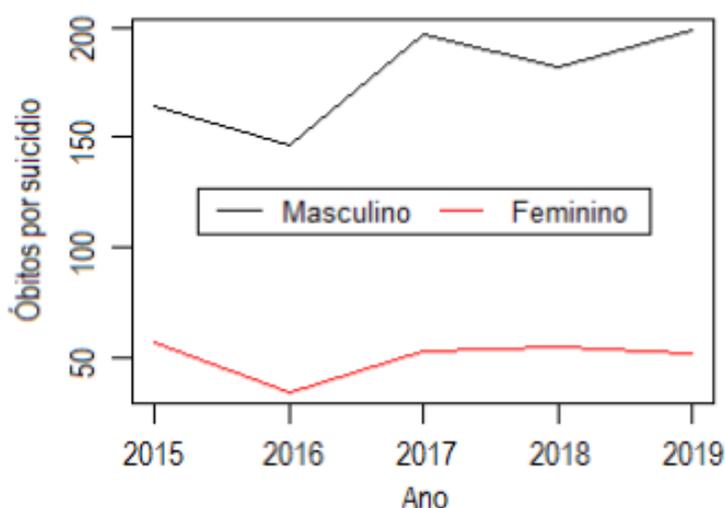
Categoria CID-10	Masc.(%)	Fem.(%)	Total(%)
Lesão autoprovocada intencionalmente			
Objeto contundente (X79)	0 (0)	1 (100)	1 (0,09)
Precipitação de um lugar elevado (X80)	29 (69)	13 (31)	42 (3,68)
Precipitação ou permanência diante de um objeto em movimento (X81)	1 (33)	2 (67)	3 (0,26)
Impacto de um veículo a motor (X82)	1 (50)	1 (50)	2 (0,18)
Meios não especificados (X84)	1 (50)	1 (50)	2 (0,18)
Total	889 (78)	251 (22)	1140 (100)

Obs.: As categorias Dispositivos explosivos (X75) e outros meios especificados (X83) não registraram óbitos.

Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

No tocante ao número de óbitos por suicídio para o período de 2015 a 2019 na Figura 1, nota-se que no ano de 2016 houve uma diminuição em ambos os gêneros e nos anos seguintes voltaram a crescer, principalmente entre o sexo masculino.

Figura 1 – Óbitos por suicídio por ano segundo o gênero. Paraíba, Brasil, 2015-2019.



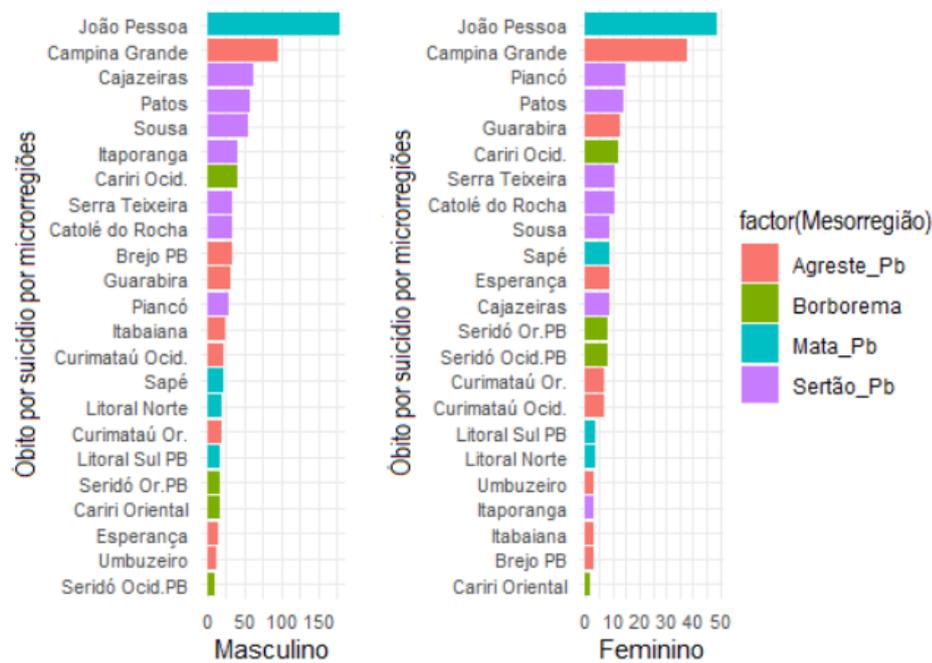
Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Na Figura 2, temos a distribuição da mortalidade por suicídio e gênero nas microrregiões de acordo com as mesorregiões.

A Figura 2 mostra que as microrregiões de João Pessoa (20,0%) e Campina Grande (11,8%) lideram os casos de óbitos por suicídio no estado da Paraíba, ao contrário de Umbuzeiro,

que apresenta o menor número (1,3%). Para os gêneros masculino e feminino, João Pessoa apresenta respectivamente 20,1% e 19,5%, enquanto Campina Grande 10,8% e 15,1%, respectivamente. Seridó Ocidental Paraibano apresenta o menor quantitativo masculino (1,0%) de óbitos por suicídio e Cariri Oriental para o feminino com 0,8%. Ao considerar as mesorregiões, Sertão Paraibano apresenta maior percentual de óbitos por suicídio com 33,6%, Agreste Paraibano (29,8), Mata Paraibana (26,6%) e a mesorregião da Borborema por último, com 9,9% dos casos.

Figura 2 – Óbitos por suicídio e gênero nas microrregiões de acordo com as mesorregiões. Paraíba, Brasil, 2015-2019.



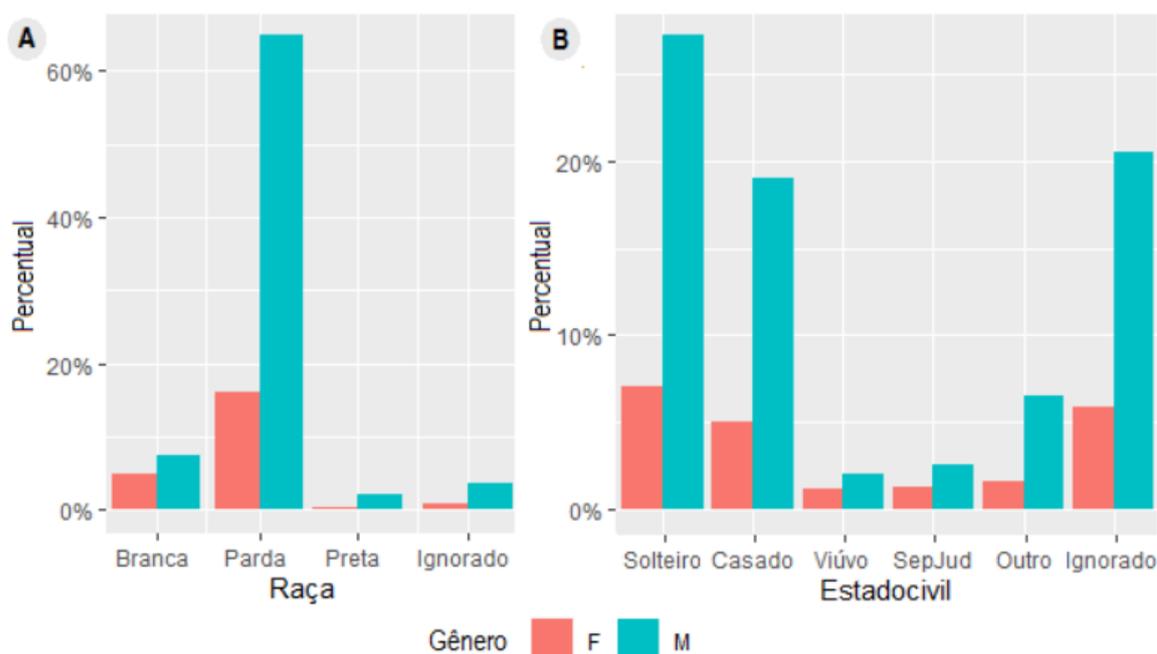
Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

A Figura 3 apresenta o percentual casos de óbitos por suicídio por raça, estado civil de acordo com o gênero.

Conforme com a Figura 3 (A) e Figura 3 (B), os óbitos por suicídio ocorreram com predominância em homens pardos (+ de 60%) e solteiros (+ de 25%), também nota-se o mesmo perfil para as mulheres.

Estudos nacionais Machado e Santos (2015), Franck, Monteiro e Limberger (2020), Silva et al. (2021) e estrangeiros Fisher, Overholser e Dieter (2015), Romero-Pimentel et al. (2018), Nunes (2018), Kootbodien et al. (2020) também revelam maiores ocorrências de mortes por suicídio entre homens.

Figura 3 – Porcentagem de óbitos por suicídio, raça e estado civil de acordo com o gênero. Paraíba, Brasil, 2015-2019.



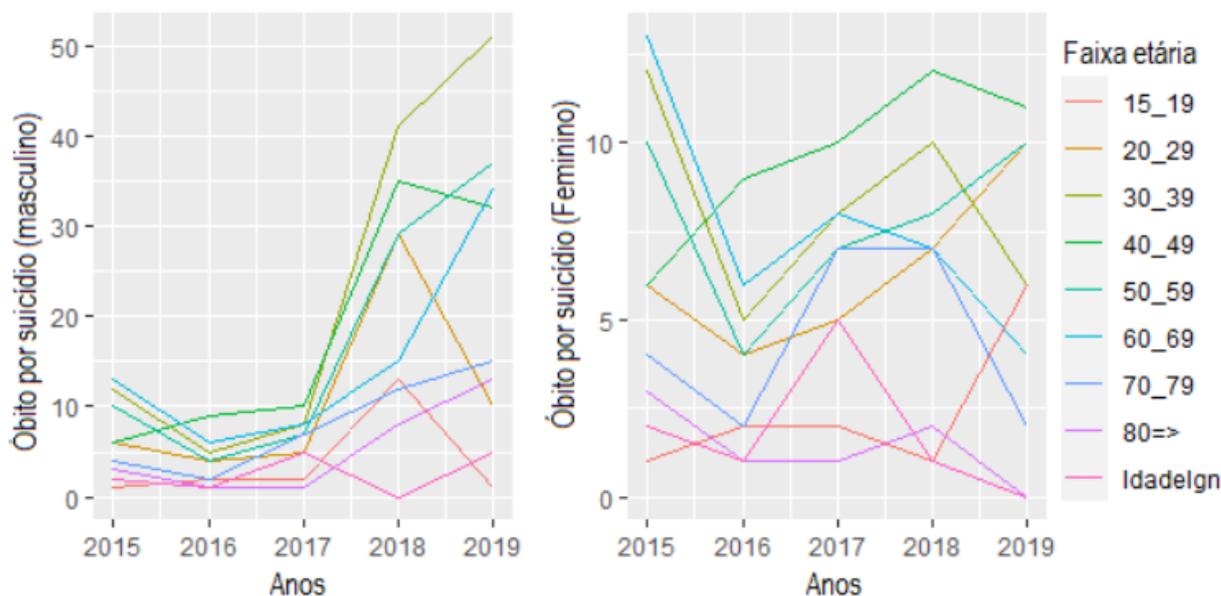
Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Um estudo no Ceará realizado por Lemos, Jorge e Linard (2021) analisou o perfil dos casos de suicídio no período de 2012 a 2019 e verificou que há uma prevalência no sexo masculino, em indivíduos pardos, solteiro, com idade compreendida entre 30 a 39 anos e escolaridade variando de 4 a 7 anos estudados. Em relação á faixa etária também corrobora com o achado neste estudo, conforme observa-se na Figura 4.

Conforme observa-se na Figura 4, a mortalidade por suicídio ocorre principalmente em pessoas com idade ativa de trabalho de 20 a 59 anos (73,4%). O gênero masculino com faixa etária de 30-39 anos (20,0%) apresenta o maior aumento contínuo de suicídio desde o ano de 2017, em seguida aparece os de 50-59 anos (17,0%) e 60-69 (10,0%) em menor número. Enquanto jovens de 20-29 anos (17,8%) observou-se o maior declínio no número de morte autoprovocada seguido dos que tinham idade entre de 15-19 anos (AGUIAR; CAMÊLO; LIMA, 2021).

Por outro lado, nota-se um aumento de suicídio entre o gênero feminino com faixa etária de 15 a 19 anos e na idade de 20 a 29 mais que dobrou entre os anos de 2016 a 2019; esse aumento também é significativo entre mulheres de 50-59 anos. Em relação as mulheres de 40-49 anos são elas detentoras dos maiores números de suicídio desde o ano de 2016 no estado, com uma leve redução em 2019 e mulheres de 70-79 anos tiveram a maior redução de mortalidade por suicídio de 2018 para 2019.

Figura 4 – Óbito por suicídio por gênero e ano de acordo com a faixa etária. Paraíba, Brasil, 2015-2019.



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Segundo Aguiar, Camêlo e Lima (2021), no que diz respeito aos anos de estudos, os que não tinham nenhuma instrução (3,1%), de 1 a 3 anos de estudos (4,2%), de 4 a 7 anos (6,4%), de 8 a 11 anos (4,7%), de 12 ou mais anos (4,2%) e por fim, escolaridade ignorada apresentou elevadíssimo percentual de 77,4%. Em homens, a mortalidade por suicídio ocorreu mais entre os que tinham de 4 a 7 anos de estudo (6,3%) e para as mulheres predominou entre as que tinham 12 ou mais anos de escolaridade. Outros estudos também se deparam com percentuais consideráveis de dados ignorados para o quesito escolaridade, tanto a nível nacional com 31% Mata, Daltro e Ponde (2020) como também para o Nordeste com o mesmo percentual de 31% (SILVA et al., 2021).

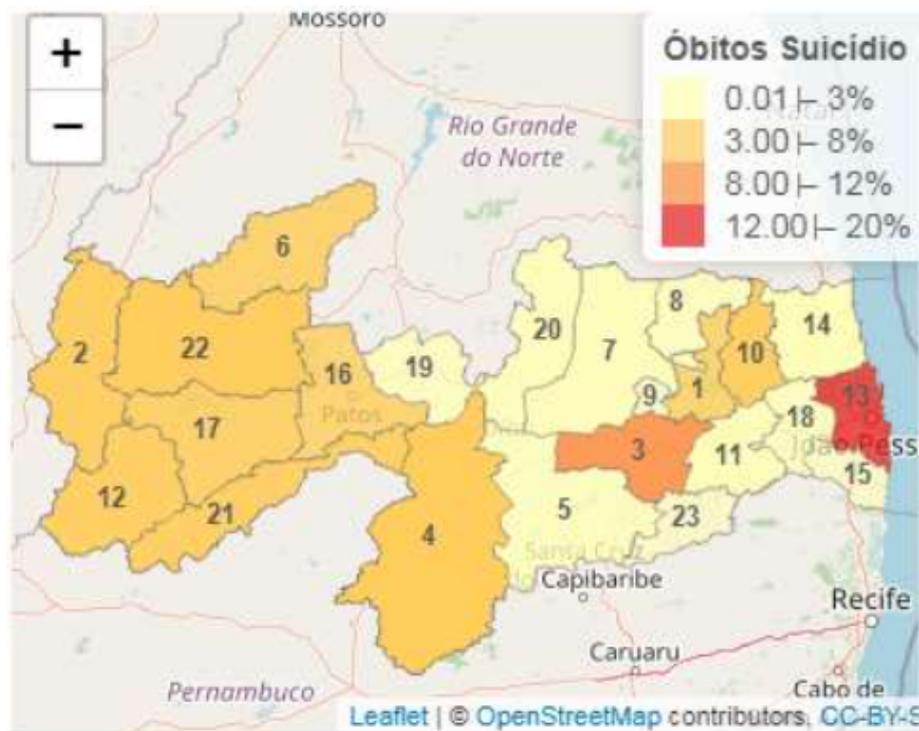
Na Figura 5, temos a distribuição percentual da mortalidade por suicídio nas microrregiões.

Observa-se na Figura 5 que os óbitos por suicídios no estado da Paraíba estão mais concentrados nas microrregiões de João Pessoa (12 a 20%) e em seguida aparece Campina Grande (8 a 12%). A cidade de João Pessoa, como é o município mais populoso do estado da Paraíba sozinha concentrou (14,1%). Franck, Monteiro e Limberger (2020) estudaram a mortalidade no Rio Grande do Sul no período de 2017 e 2018 e também identificou que a capital Porto Alegre como sendo responsável pelo maior número de vítimas do estado.

Na Figura 5 também nota-se que todas as microrregiões pertencentes à mesorregião do Sertão Paraibano (2 = Cajazeiras, 6 = Catolé do Rocha, 12 = Itaporanga, 16 = Patos, 17 = Piancó, Serra do Teixeira e 22 = Sousa), a microrregião do Cariri Ocidental (4), Brejo Paraibano (1) e Guarabira (10) detém de 3 a 8% cada uma delas. E as demais microrregiões (5, 7, 8, 9, 11, 14, 15, 18, 19, 20 e 23) apresentam os menores percentuais (0,01 a 3%) (AGUIAR; CAMÊLO;

LIMA, 2021).

Figura 5 – Distribuição de óbitos por suicídio nas microrregiões. Paraíba, Brasil, 2015-2019.



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Obs.: Os números indicam as microrregiões:

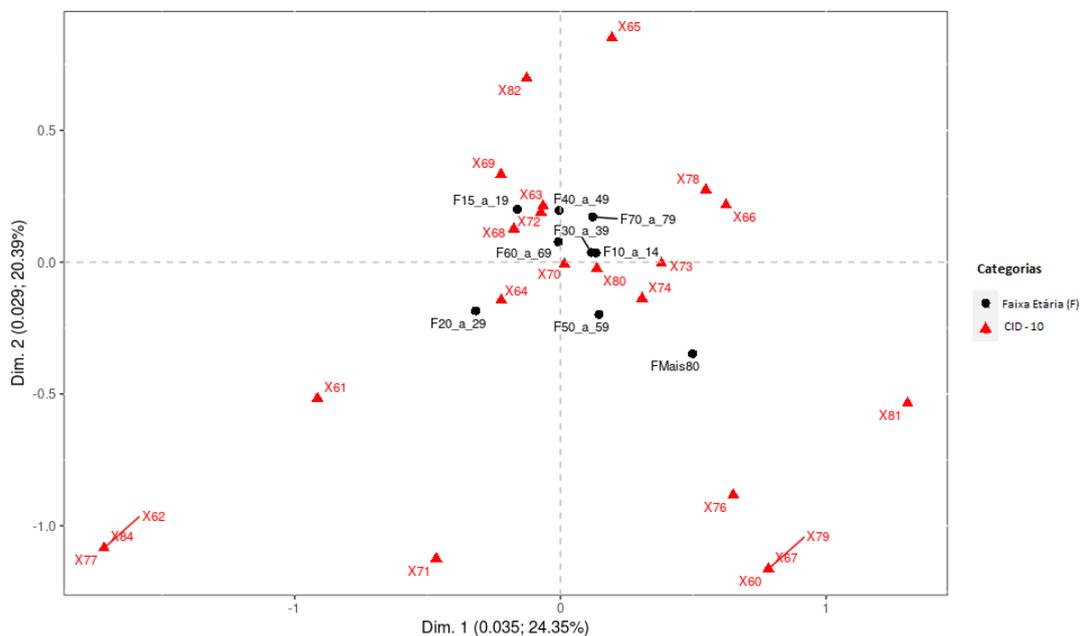
- | | | |
|------------------------|-------------------|-----------------------|
| 1. Brejo Paraibano | 9. Esperança | 17. Piancó |
| 2. Cajazeiras | 10. Guarabira | 18. Sapé |
| 3. Campina Grande | 11. Itabaiana | 19. Seridó Ocidental |
| 4. Cariri Ocidental | 12. Itaporanga | 20. Seridó Oriental |
| 5. Cariri Oriental | 13. João Pessoa | 21. Serra do Teixeira |
| 6. Catolé do Rocha | 14. Litoral Norte | 22. Sousa |
| 7. Curimataú Ocidental | 15. Litoral Sul | 23. Umbuzeiro |
| 8. Curimataú Oriental | 16. Patos | |

3.3 Análise de Correspondência

A Figura 6 exibe o gráfico de dispersão da CA, com os rótulos das linhas indicando qual CID-10 tem maior contribuição para a definição das dimensões selecionadas. Os rótulos de colunas informam as faixas-etárias e a correlação com as dimensões selecionadas. E como

mostra Batista, Escuder e Pereira (2004), as categorias que possuem localização próxima na projeção plana, são as que possuem uma relação forte e se correspondem mais que as categorias separadas por distâncias maiores. Observa-se que a categoria 20 a 29 anos da variável faixa etária está mais associada com as categorias X61, X62, X64, X70, X71, X77 e X84, a categoria 50 a 59 anos com a X73, X74 e X80, enquanto a categoria 15 a 19 anos, está mais próxima das categorias da variável CID-10: X63, X68, X69 e X72, a categoria X65 que é referente a autointoxicação voluntária por álcool é a mais isolada em relação às demais.

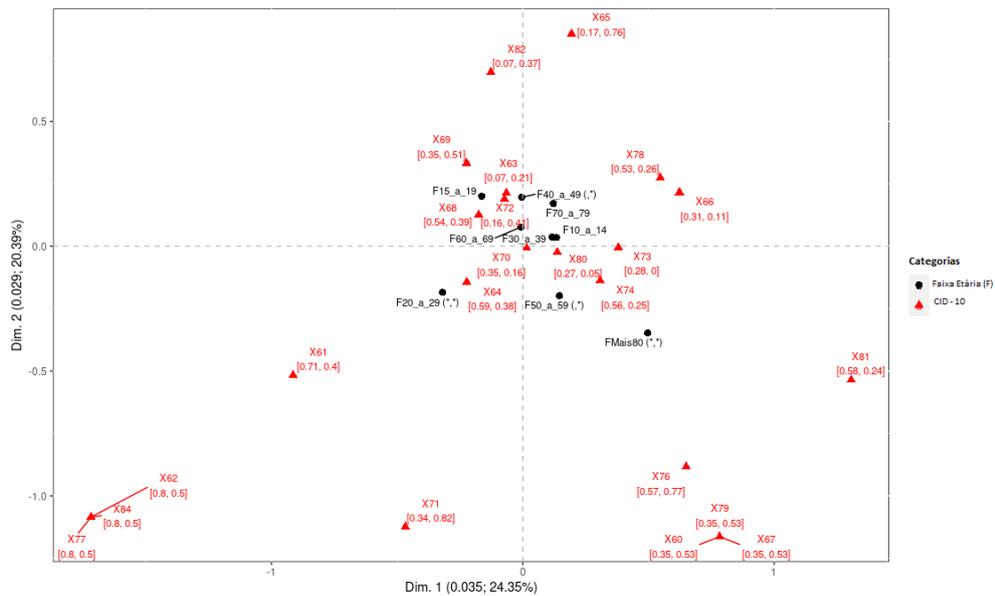
Figura 6 – Dispersão de CID-10 com Faixa etária



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Na Análise de Correspondência, exploramos o comportamento associativo das categorias de duas variáveis qualitativas, para explicar a associação entre as categorias das variáveis em um espaço multidimensional, segundo a Figura 7, em relação a variável CID-10, temos que para categoria X61 a primeira dimensão explica 0,71 e a segunda dimensão explica 0,4, onde teve um percentual alto de explicação para primeira dimensão, já que esses valores variam entre 0 e 1. A X76 mostrou ser significativa para as duas dimensões, o mesmo acontece para X84 e X68, já em relação às faixas etárias, temos que a categoria de 20 a 29 anos e mais de 80 anos são significativas para as duas dimensões, por outro lado, as categorias de 10 a 14 anos, 15 a 19, 30 a 39, 60 a 69 e 70 a 79 não são bem explicadas por nenhuma das duas dimensões.

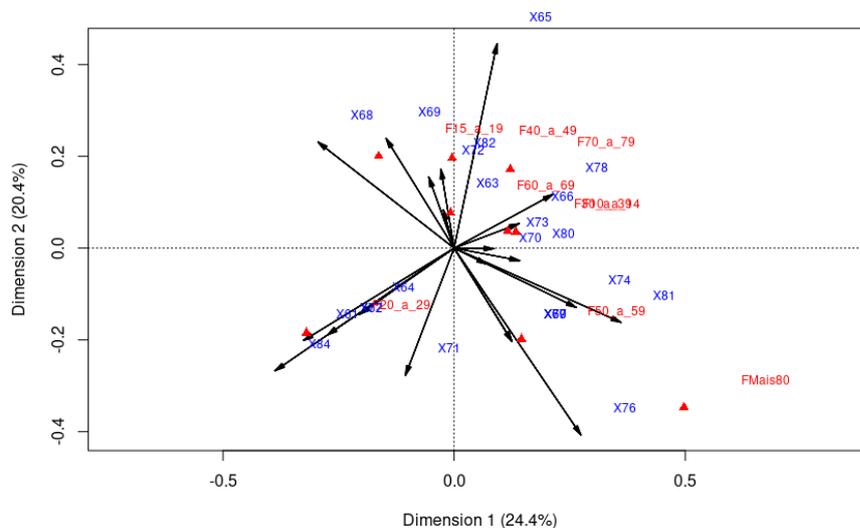
Figura 7 – Dispersão com a contribuição variável faixa etária



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Como comenta Greenacre (1984), a análise de correspondência mostrar as linhas e colunas de uma matriz de dados em uma tabela de contingência de dupla entrada, como pontos em um espaço multidimensional de vetores, na Figura 8 são apresentados vetores de categorias de linha e vetores de categorias de colunas, o vetor próximo de 0 tem pouca influência e quanto mais distante do 0 maior é a influência. A categoria X76 tem muita influência no eixo da segunda componente, já que o vetor indica a direção e a intensidade e a X65 no eixo 2 na segunda coordenada.

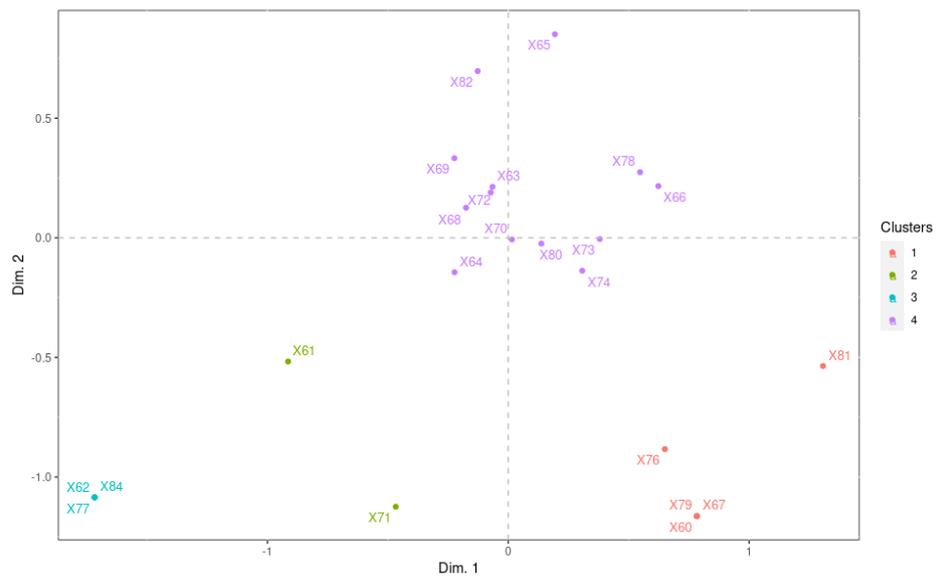
Figura 8 – Associação dos CID-10 com todas as variáveis faixa etária



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

A primeira dimensão é a que mais explica a variação existente nos dados, sendo a segunda dimensão a seguinte e, assim, sucessivamente, com o estudo de Pestana e Gageiro (2005), assim, utilizando 2 dimensões (X,Y) e fazendo a análise de agrupamento com linha e coluna em cluster, foram considerados 4 clusters em relação às categorias da variável CID-10, em que as categorias X60, X67, X76, X79 e X81 são agrupado no cluster 1, as categorias X61, X71 no cluster 2, as X62, X77 e X84 no 3 e as demais categorias no cluster 4, como mostra a Figura 9.

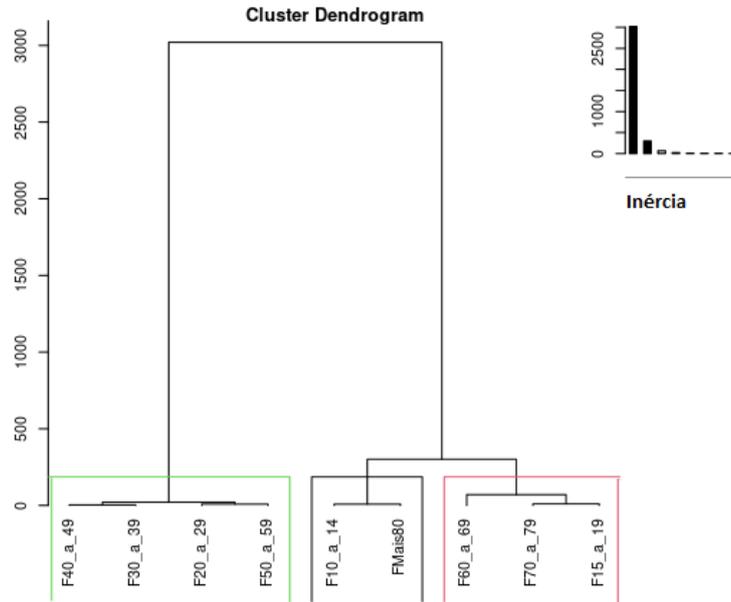
Figura 9 – Clusters baseados em coordenadas de Análise de Correspondência no espaço de dimensionalidade 2.



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Por Hair (2005), a Análise de Correspondência tem como objetivo, agrupar variáveis altamente associadas, assim, foram agrupadas as categorias da variável faixas etárias que são altamente associadas, sendo criados 3 grupos, o com as categorias 20 a 29 anos, 30 a 39, 40 a 49 e 50 a 59, o grupo com 10 a 14 e mais de 80 anos e um com 15 a 19 anos, 60 a 69 anos e 70 a 79, vemos também pelo gráfico ao lado que mostra a inércia respectivas para as dimensões e observamos que os dados são melhor explicados por no máximo duas dimensões.

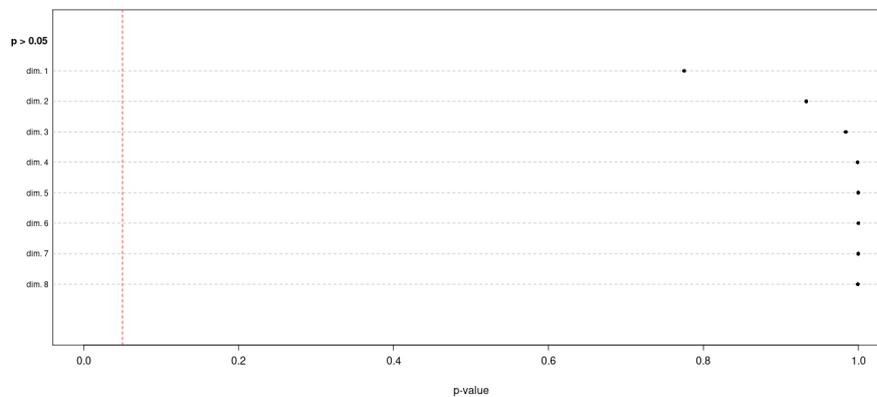
Figura 10 – Agrupamento hierárquico



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Para as 8 possíveis dimensões [mín(linha,coluna) - 1], é mostrado com a Figura 11 o p-valor para cada uma das dimensões. Podemos observar que para todas as dimensões temos um p-valor próximo de 1, ou seja, não significativas. Porém, vamos prosseguir a análise para fins didáticos.

Figura 11 – Teste de Malinvaud para a significância das dimensões da Análise de Correspondência

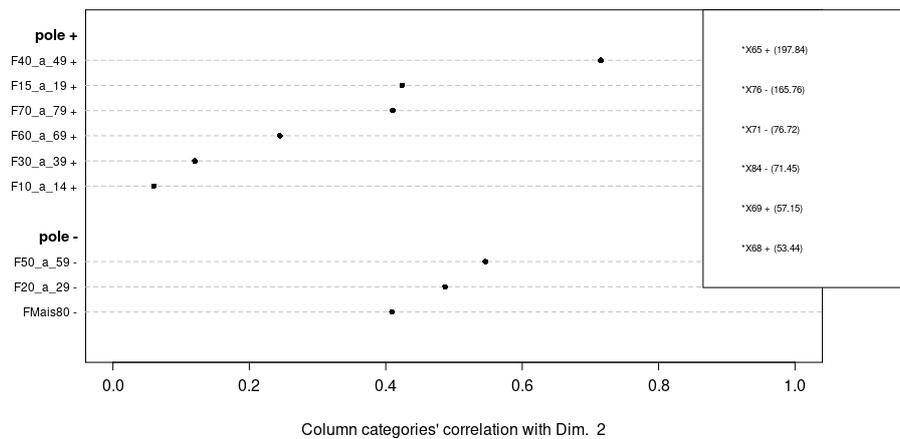


Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

A função utilizada permite calcular a correlação $\sqrt{\cos^2}$ das categorias idades com a

dimensão selecionada, os sinais + e - indicam se houve contribuição positiva ou negativa. A contribuição varia de 0 a 1, quanto mais próximo de 1, maior a contribuição. Observa-se que a categoria 40 a 49 anos da variável faixa etária, apresentou uma maior contribuição positiva na dimensão 2, enquanto as categorias de 20 a 29, 50 a 59 anos e mais de 80 anos, estão correlacionadas negativamente com a dimensão 2. Ao lado, temos apenas as categorias de CID-10 que possuem grande contribuição para a definição da dimensão 2.

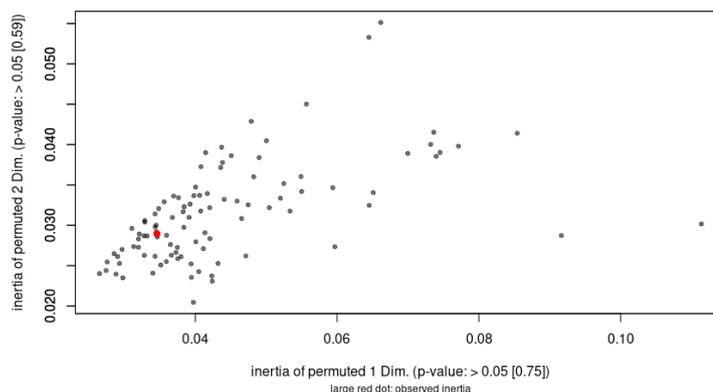
Figura 12 – Valores de correlação de cada variável faixa etária com a dimensão 2



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

Segundo Johnson e Wichern (2007), a inércia total é definida como a soma dos quadrados de todos os valores singulares ou autovalores não nulos. O máximo de dimensões que podemos estimar será 8. Considerando a Figura 13 e pela análise de Hair (2005), "uma regra prática é que as dimensões com inércia (autovalores) maior que 0,2 devem ser incluídas na análise", sendo assim, duas dimensões explicam bem os dados.

Figura 13 – Dispersão da inércia



Fonte: Elaboração própria a partir de dados do SIM/DATASUS/MS.

4 CONCLUSÃO

O suicídio é um problema de saúde pública tanto no Brasil como em todo o mundo e esta realidade tem contribuído para o aumento significativo das taxas de mortalidade nas unidades federativas. Com este estudo, conclui-se que a mortalidade por suicídio no estado da Paraíba apresenta tendência de crescimento e é marcada por cor e gênero. O perfil é majoritariamente definido por homens, pardos, solteiros, em idade ativa de trabalho com 20 a 59 anos, com predominância na faixa etária de 30 a 39 anos, 4 a 7 anos de estudos e residentes nas duas principais microrregiões de João Pessoa e Campina Grande, onde se localiza as duas cidades mais populosas do estado, João Pessoa e Campina Grande.

Na utilização da Análise de Correspondência e o pacote CAinterprTools, a categoria 20 a 29 anos da variável faixa etária foi mais associada com as categorias X61, X62, X64, X70, X71, X77 e X84, utilizando 2 dimensões (X,Y) e fazendo a análise de agrupamento com linha e coluna em cluster, foram considerados 4 clusters em relação às categorias da variável CID-10, enquanto para as categorias da variável faixa etária, foram criados 3 grupos, foi observado também que o máximo de dimensões que podemos estimar foram 8 [mín(linha,coluna) - 1 e duas dimensões explicam melhor os dados.

Como trabalhos futuros pretende-se aplicar o estudo com as categorias das demais variáveis, para observar a relação dos óbitos por suicídio com elas e expandir este estudo para os demais estados do Brasil.

REFERÊNCIAS

- AGUIAR, D. C.; CAMÊLO, E. L. S.; LIMA, F. de O. Um inimigo silencioso: óbitos por suicídio no estado da paraíba no período de 2015-2019. *Research, Society and Development*, v. 10, n. 10, p. e196101018575–e196101018575, 2021. Citado 3 vezes nas páginas 29, 30 e 31.
- ALBERTI, G. *Graphical Aid in Correspondence Analysis Interpretation and Significance Testings*. [S.l.], 2020. R package version 1.1.0. Disponível em: <<https://CRAN.R-project.org/package=CAinterprTools>>. Citado na página 25.
- BAÉRE, F. d.; ZANELLO, V. O gênero no comportamento suicida: Uma leitura epidemiológica dos dados do distrito federal. Programa de Pós-graduação em Psicologia e do Programa de Pós-graduação em . . . , 2018. Citado na página 25.
- BANDO, D. H. et al. Suicide rates and trends in são paulo, brazil, according to gender, age and demographic aspects: a joinpoint regression analysis. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 34, p. 286–293, 2012. Citado na página 25.
- BATISTA, L. E.; ESCUDER, M. M. L.; PEREIRA, J. C. R. A cor da morte: causas de óbito segundo características de raça no estado de são paulo, 1999 a 2001. *Revista de saúde Pública*, SciELO Brasil, v. 38, p. 630–636, 2004. Citado 3 vezes nas páginas 13, 14 e 32.
- BEH, E. J. Simple correspondence analysis: a bibliographic review. *International Statistical Review*, Wiley Online Library, v. 72, n. 2, p. 257–284, 2004. Citado na página 13.
- BERINGUEL, B. M. et al. Mortalidade por suicídio no estado de pernambuco, brasil (1996-2015). *Revista Brasileira de Enfermagem*, SciELO Brasil, v. 73, 2020. Citado na página 25.
- BRASIL. Perfil epidemiológico das tentativas e óbitos por suicídio no brasil ea rede de atenção à saúde. *Bol Epidemiol*, v. 48, n. 30, p. 1–14, 2017. Citado na página 11.
- BRASIL, M. P. F. *MPF e órgãos parceiros destacam importânciada prevenção ao suicídio na PB. Ministério Público Federal*. [S.l.], 2019. Citado na página 11.
- CARMO, É. A. et al. Tendência temporal da mortalidade por suicídio no estado da bahia. *Cogitare Enfermagem*, v. 23, n. 1, 2018. Citado na página 25.
- CAVALCANTE, F. G.; MINAYO, M. C. d. S.; MANGAS, R. M. d. N. Diferentes faces da depressão no suicídio em idosos. *Ciência & Saúde Coletiva*, SciELO Public Health, v. 18, p. 2985–2994, 2013. Citado na página 11.
- CAVANAGH, J. T. et al. Psychological autopsy studies of suicide: a systematic review. *Psychological medicine*, Cambridge University Press, v. 33, n. 3, p. 395–405, 2003. Citado na página 11.
- CECCON, R. F. et al. Suicídio e trabalho em metrópoles brasileiras: um estudo ecológico. *Ciência & Saúde Coletiva*, SciELO Brasil, v. 19, p. 2225–2234, 2014. Citado na página 11.
- CEVS, C. E. de V. S. *Boletim de Vigilância Epidemiológica de Suicídio e Tentativa de Suicídio*. Rio Grande do Sul, 2018. Citado na página 11.

CHACHAMOVICH, E. et al. Which are the recent clinical findings regarding the association between depression and suicide? *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 31, p. S18–S25, 2009. Citado na página 11.

CHENG, J. et al. *Create interactive web maps with the JavaScript 'Leaflet' library. R package version 2.0. 4.1.* 2021. Citado na página 25.

DEVILLE, J.-C.; SAPORTA, G. Correspondence analysis, with an extension towards nominal time series. *Journal of econometrics*, Elsevier, v. 22, n. 1-2, p. 169–189, 1983. Citado na página 14.

FÁVERO, L. P. et al. *Análise de dados: modelagem multivariada para tomada de decisões.* [S.l.]: Rio de Janeiro: Elsevier, 2009. Citado 5 vezes nas páginas 15, 16, 17, 19 e 21.

FÁVERO, L. P. L.; BELFIORE, P. P.; FOUTO, N. M. M. D. Escolha de meios de pagamento por populações de média e baixa renda: uma abordagem sob a perspectiva da análise fatorial e de correspondência. *Revista de Economia e Administração*, v. 5, n. 2, 2006. Citado na página 22.

FISHER, L. B.; OVERHOLSER, J. C.; DIETER, L. Methods of committing suicide among 2,347 people in ohio. *Death studies*, Taylor & Francis, v. 39, n. 1, p. 39–43, 2015. Citado na página 28.

FRANCK, M. C.; MONTEIRO, M. G.; LIMBERGER, R. P. Suicide mortality in rio grande do sul, brazil: a cross-sectional analysis of cases, 2017-2018. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v. 29, p. e2019512, 2020. Citado 2 vezes nas páginas 28 e 30.

GREENACRE, M. J. Theory and applications of correspondence analysis. London (UK) Academic Press, 1984. Citado 3 vezes nas páginas 12, 13 e 33.

HAIR, J. Jf; anderson, re; tatham, rl; black, wc análise multivariada de dados. *Tradução: Adonai Schlup Sant'Anna e Anselmo Chaves Neto*, v. 5, 2005. Citado 5 vezes nas páginas 13, 14, 16, 34 e 36.

HIRSCHFELD, H. O. A connection between correlation and contingency. In: CAMBRIDGE UNIVERSITY PRESS. *Mathematical Proceedings of the Cambridge Philosophical Society.* [S.l.], 1935. v. 31, n. 4, p. 520–524. Citado na página 13.

JOHNSON, R. A.; WICHERN, D. W. Applied multivariate statistical analysis. 6th. *New Jersey, US: Pearson Prentice Hall*, 2007. Citado 3 vezes nas páginas 21, 22 e 36.

KAIRI, K.; LEO, D. D. Regions with the highest suicide rates for children and adolescents—some observations. *J Child Adolesc*, 2014. Citado na página 11.

KOOTBODIEN, T. et al. Trends in suicide mortality in south africa, 1997 to 2016. *International journal of environmental research and public health*, MDPI, v. 17, n. 6, p. 1850, 2020. Citado 2 vezes nas páginas 25 e 28.

LAVOR, M. d. L. S. S. et al. Sazonalidade e tentativas de suicídio: comparativo entre a paraíba, região nordeste e brasil. *Brazilian Journal of Health Review*, v. 3, n. 3, p. 3960–3970, 2020. Citado na página 11.

LE MOS, A. M.; JORGE, M. S. B.; LINARD, C. F. B. M. Perfil do suicídio em um estado do nordeste brasileiro. *Research, Society and Development*, v. 10, n. 2, p. e45410212598–e45410212598, 2021. Citado na página 29.

- MACENTE, L. B.; ZANDONADE, E. Estudo da série histórica de mortalidade por suicídio no espírito santo (de 1980 a 2006). *Jornal brasileiro de psiquiatria*, SciELO Brasil, v. 60, p. 151–157, 2011. Citado na página 25.
- MACHADO, D. B.; SANTOS, D. N. d. Suicídio no brasil, de 2000 a 2012. *Jornal Brasileiro de Psiquiatria*, SciELO Brasil, v. 64, p. 45–54, 2015. Citado 2 vezes nas páginas 25 e 28.
- MATA, K. C. R. da; DALTRO, M. R.; PONDE, M. P. Perfil epidemiológico de mortalidade por suicídio no brasil entre 2006 e 2015. *Revista Psicologia, Diversidade e Saúde*, v. 9, n. 1, p. 74–87, 2020. Citado na página 30.
- MINGOTI, S. A. Análise de dados através de métodos estatística multivariada: uma abordagem aplicada. In: *Análise de dados através de métodos estatística multivariada: uma abordagem aplicada*. [S.l.: s.n.], 2007. p. 295–295. Citado na página 21.
- NUNES, A. M. Suicídio em portugal: um retrato do país. *Jornal Brasileiro de Psiquiatria*, SciELO Brasil, v. 67, p. 25–33, 2018. Citado 2 vezes nas páginas 25 e 28.
- OLARIAGA, L. J.; HERNÁNEZ, L. L. Análisis de correspondencias. *Análisis de correspondencias*, La Muralla, p. 0–0, 2000. Citado 4 vezes nas páginas 12, 15, 17 e 23.
- OMS, O. M. de S. *Datos y cifras sobre el suicidio: infografía*. [S.l.], 2018. Citado na página 11.
- PESTANA, M.; GAGEIRO, J. Análise dos dados para ciências sociais: A complementaridade do spss lisboa. *Portugal: Edições Silabo*, 2005. Citado 4 vezes nas páginas 14, 16, 17 e 34.
- RIBEIRO, J. M.; MOREIRA, M. R. Uma abordagem sobre o suicídio de adolescentes e jovens no brasil. *Ciência & Saúde Coletiva*, SciELO Brasil, v. 23, p. 2821–2834, 2018. Citado na página 11.
- ROCA, M. et al. Suicidal risk and executive functions in major depressive disorder: a study protocol. *BMC psychiatry*, BioMed Central, v. 19, n. 1, p. 1–6, 2019. Citado na página 11.
- ROMERO-PIMENTEL, A. L. et al. Demographic and clinical characteristics of completed suicides in mexico city 2014–2015. *Frontiers in psychiatry*, Frontiers Media SA, v. 9, p. 402, 2018. Citado 2 vezes nas páginas 25 e 28.
- SEHNEM, S. B.; PALOSQUI, V. Características epidemiológicas do suicídio no estado de santa catarina. *Fractal: Revista de Psicologia*, SciELO Brasil, v. 26, p. 365–378, 2014. Citado na página 25.
- SILVA, I. G. da et al. Diferenciais de gênero na mortalidade por suicídio. *Rev Rene*, Universidade Federal do Ceará, v. 22, p. 47, 2021. Citado 3 vezes nas páginas 11, 28 e 30.
- TEAM, R. C. R. *A language and environment for statistical computing (R Versin 4.0.3, R Foundation for Statical Computing, Vienna, Austria, 2020)*. 2021. Disponível em: <<https://www.r-project.org/>>. Acesso em: Março de 2021. Citado na página 25.
- TEISMANN, T. et al. Positive mental health moderates the association between depression and suicide ideation: A longitudinal study. *International Journal of Clinical and Health Psychology*, Elsevier, v. 18, n. 1, p. 1–7, 2018. Citado na página 11.

WHITLARK, D. B.; SMITH, S. M. Using correspondence analysis to map relationships. *Marketing Research*, v. 13, n. 3, 2001. Citado 2 vezes nas páginas 11 e 13.

WHO, W. H. O. *Suicide in the world -Global Health Estimates*. [S.l.], 2019. Citado na página 11.

WICKHAM, H. Data analysis. In: *ggplot2*. [S.l.]: Springer, 2016. p. 189–201. Citado na página 25.