



**Universidade de Brasília
Departamento de Estatística**

**DETERMINANTES SOCIAIS DA SAÚDE
Mortalidade Adulta na Área Metropolitana de Brasília**

Luana Gonçalves Barreto

Trabalho de Conclusão de Curso apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2021**

Luana Gonçalves Barreto

DETERMINANTES SOCIAIS DA SAÚDE
Mortalidade Adulta na Área Metropolitana de Brasília

Orientador(a): Ana Maria Nogaes Vasconcelos.
Coorientador(a): Valeria Maria Rodrigues Fachine.

Trabalho de Conclusão de Curso apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2021

Agradecimento

À Deus, pela minha vida, por estar comigo na minha caminhada e por tudo que ele colocou no meu caminho.

À minha família e, em especial, dedico este trabalho à minha mãe, Dona Isabel, pela minha trajetória e pela pessoa que sou. Obrigada por me apoiar em todas as minhas decisões e ser a pessoa que mais admiro e respeito.

À professora Ana Maria Nogales Vasconcelos e à mentora Valeria Maria Rodrigues Fachine, pelas correções e ensinamentos.

Aos meus amigos de faculdade Flávia, Felipe e Hellen que me fizeram companhia nas rotinas de faculdade tornando mais leves os momentos livres.

Ao meu amigo Jhony pelo apoio recebido que foi de fundamental importância com palavras e atitudes de apoio nesse período de quarentena.

E a todos aqueles que contribuíram, de alguma forma, para a realização deste trabalho.

Resumo

Tendo como base conceitual os Determinantes Sociais da Saúde, o presente trabalho teve como objetivo caracterizar a mortalidade adulta na Área Metropolitana de Brasília. Para tanto, buscou-se analisar a associação entre o perfil socioeconômico dos falecidos e as causas básicas de morte. Foram utilizados dados sobre óbitos entre 30 e 69 anos do Sistema de Informações sobre Mortalidade (SIM) entre 2011 e 2020, para o Distrito Federal e 12 municípios goianos que fazem parte da Área Metropolitana de Brasília. Foi aplicada a técnica de Análise de Correspondência Múltipla (ACM) para avaliar similaridade entre as categorias das variáveis selecionadas. Além disso, identificaram-se associações entre as variáveis com base em dois eixos principais, o primeiro diretamente relacionado aos aspectos socioeconômicos, e o segundo às características individuais dos falecidos. Foram identificadas as seguintes associações: as causas externas estão associadas ao gênero masculino; as categorias de baixa escolaridade estão associadas à baixa renda no primeiro eixo; e as Regiões Administrativas de alta renda têm como causas de morte neoplasias e doenças do coração, ocorridas em idades mais avançadas. Com o uso da técnica de ACM foi possível associar os perfis com as causas de mortalidade, reforçando a importância que tais associações sejam levadas em consideração para o direcionamento de políticas de saúde para a população adulta.

PALAVRAS CHAVES: Área Metropolitana de Brasília, Correspondência Múltipla, Determinantes Sociais da Saúde, Mortalidade.

Sumário

1 Introdução	1
2 Objetivos	2
2.1 Objetivo Geral	2
2.2 Objetivos Específicos.	2
3 Analise bibliográfica	3
3.1 Determinantes Sociais da Saúde - DSS.	3
3.1.1 Gênero	4
3.1.2 Escolaridade	5
3.1.3 Raça/Cor	5
3.1.4 Estado Civil	6
4 Metodologia	8
4.1 Fonte de Dados	8
4.1.1 Construção das variáveis de interesse dos Determinantes Sociais	8
4.2 Grupos da Área Metropolitana de Brasília (AMB)	10
4.3 Causas Básicas	11
4.4 Análise de Correspondência	14
4.4.1 Qui-quadrado e inércia	14
4.4.2 Redução dimensional	18
4.4.3 Elementos suplementares	19
4.4.4 Análise de Correspondência Múltipla	20
5 Análises e Resultados	24
5.1 Análise exploratória	24
5.2 Qui-quadrado	26
5.3 Análise de Correspondência Múltiplas	27
6 Considerações Finais	36
7 Referências	37
8 Apêndice	40

Lista de Tabelas

1	Informações sobre a nomenclatura das variáveis	10
2	Capítulos CID-10	12
3	Grupos de causas CID-10	13
4	Banco de dados com N observações e ($Q > 2$) variáveis categóricas.	21
5	Codificação binária das categorias das variáveis originais	21
6	Caraterística do óbito. AMB, 2011-2020	26
7	Tabela Qui-Quadrado (χ^2)	27
8	Principais inércias e porcentagem da matriz indicadora das médias	28
9	Contribuições das categorias para os 5 primeiros eixos da matriz indicadora das médias.	29

Lista de Figuras

1	Modelo de Dahlgreen e Whitehead,1990	4
2	Área Metropolitana de Brasília	10
3	Representação gráfica de um perfil a_i em um espaço multidimensional . . .	18
4	Perfis em um espaço multidimensional	18
5	Gráfico de sedimentação	28
6	Contribuição no eixo 1	30
7	Contribuição no eixo 1	30
8	Nuvem de variáveis (Dim 1-2)	31
9	Nuvem de categorias (Dim 1-2)-ACM	32
10	Nuvem de categorias ativas separadas (Dim 1-2) -ACM	32
11	Nuvem de categorias suplementares separadas (Dim 1-2) -ACM	33
12	Contribuição dos indivíduos por variável	35

1 Introdução

Os estudos sobre Determinantes Sociais da Saúde (DSS) tiveram grande destaque ao longo do século XXI, principalmente na compreensão sobre iniquidades em saúde, relacionando as condições sociais em que as pessoas crescem, vivem e trabalham. A compreensão dos DSS pode auxiliar nas análises das relações entre o acesso a serviços de saúde, a qualidade de vida das pessoas, e como essas relações tornam-se mais suscetíveis a doenças e, por consequência, aumentam o risco de morte em um determinado grupo populacional.

Uma das camadas dos DSS abordam as características individuais como, sexo, idade, escolaridade, raça/cor. Segundo Chor e Lima (2005), ao contrário de outras áreas como educação, trabalho e justiça, estudos sobre as condições de saúde segundo raça/cor ainda são escassos no Brasil. Tendo em vista, que os dados sobre etnia foram coletados a partir dos anos 1990 nos principais sistemas de registros contínuos Sistema de Informações sobre Mortalidade (SIM) e Sistema de Informações sobre Nascimentos (Sinasc), no entanto a melhoria da qualidade dessa informação é recente.

No presente estudo, delimitou-se o período de 2011 a 2020 da população de interesse em idade adulta de 30 a 69 anos, onde encontra-se a maior parte da força de trabalho brasileira. A região a ser estudada será a Área Metropolitana de Brasília (AMB) que consiste no território do Distrito Federal e mais 12 municípios goianos, localizados na Periferia Metropolitana. Os dados foram extraídos do SIM com base nas informações das secretarias de saúde de cada UF.

Nesse sentido, na análise da mortalidade é importante relacionar as causas de morte com características individuais, por isso para melhor compreensão das relações entre as variáveis utilizou-se a técnica de Análise de Correspondência Múltipla.

O SIM possui variáveis limitadas para construir um perfil dos DSS, então o perfil estudado trará as seguintes variáveis ativas: gênero, raça/cor, escolaridade, faixa etária, grupos de regiões da AMB segundo renda e estado civil, vale ressaltar que a variável que identifica o trabalho não foi considerada, pois contava com muitas informações vazias; E as variáveis suplementares são as de causas básicas e o período analisado.

No trabalho, foi utilizado o programa estatístico R, que é um software livre, para a organização e análise de dados. O R contém várias ferramentas que permitem: gerenciamento de arquivos, análise estatística, manipulação de banco de dados, criação de gráficos, criação de relatórios e entre outros. As análises foram realizadas por meio do software R versão 3.6.3 de 2020, aplicando os pacotes FactoMineR (Husson, et,2020) e factoextra (Kassambara e Mundt, 2020)

2 Objetivos

2.1 Objetivo Geral

Analisar o perfil da mortalidade adulta de 30 a 69 anos segundo alguns Determinantes Sociais da Saúde (DSS) na Área Metropolitana de Brasília (AMB) no período de 2011 a 2020.

2.2 Objetivos Específicos

- Avaliar a qualidade dos dados do Sistema de Informações sobre Mortalidade (SIM) para os dados que se referem às localidades que formam a AMB;
- Analisar mudanças no perfil da mortalidade segundo causas de morte no período;
- Analisar a intensidade da mortalidade adulta por idade, sexo e causas segundo raça/cor na AMB;
- Compreender mais profundamente a técnica de Análise de Correspondências Múltiplas ao relacionar com variáveis categóricas de interesse utilizados nesse estudo.

3 Análise bibliográfica

3.1 Determinantes Sociais da Saúde - DSS

Pellegrin e Buss (2011), relata que distribuição da saúde e da doença, nas populações, não é aleatória e obedece à estratificação socioeconômica. No século XIX alguns autores como Villermé (1782-1863) na França, e Engels (1820-1895) e Chadwick (1800-1890) na Inglaterra, notaram uma nítida associação entre a pobreza e a alta mortalidade.

Em 1948 com a criação da Organização Mundial de Saúde (OMS), define-se que a saúde é um estado completo de bem-estar físico, mental e social, e não exclusivamente pela ausência de enfermidade, neste momento é confirmado a importância dos Determinantes Sociais de Saúde (DSS). Segundo a OMS, os DSS estão relacionados a maneira em que um indivíduo vive e trabalha ou, ainda, ao perfil em que sua vida transcorre. Sendo assim, problemas de saúde poderiam ser desencadeados por fatores sociais, econômicos, culturais, ecológicos, comportamentais e étnico-raciais. E, igualmente, as condições de moradia, saneamento, alimentação, emprego, renda e escolaridade.

Em 2005, a OMS funda a Comissão de Determinantes Sociais da Saúde (CDSS), onde considera-se que existe dois agrupamento para o DSS: os **estruturais** que geram estratificação social, como renda, educação, gênero, raça e sexualidade; e os **intermediários** que referem-se à estratificação social e determinam as diferenças na exposição a vulnerabilidade que afetam a saúde, tais como condições de vida e trabalho, acesso e atenção a saúde.

Nessa conjuntura, destaca-se a importância da integração do setor da saúde aos demais setores da sociedade no combate às iniquidades da saúde. Em 2006 o Brasil funda a Comissão Nacional dos Determinantes Sociais da Saúde (CNDSS) escolhendo o modelo de Dahlgren e Whitehead (1991), que divide os DSS em diferentes camadas segundo o nível de abrangência:

- Proximais, vinculados aos comportamentos;
- Intermediários, relacionados às condições de vida e trabalho;
- Distais, referentes à macroestrutura econômica, social e cultural.



Figura 1: Modelo de Dahlgren e Whitehead, 1990

Pellegrini e Buss, 2011, p. 595

Existem nos determinantes uma relação de desigualdades no acesso aos serviços de saúde. Além das condições socioeconômicas relacionadas às condições de saúde já mencionadas, a desigualdade de acesso aos serviços também está relacionada à configuração da política nacional de saúde, ou seja, o que constitui o princípio da política, da organização, dos serviços e do relacionamento entre clientes e profissionais de saúde.

3.1.1 Gênero

Nesse estudo trabalhará com variável gênero, que enfatiza as relações entre homens e mulheres. Que desenvolve o conceito que o gênero tem uma influência importante na divisão do trabalho, na distribuição de recursos e na distribuição de benefícios, assim pode-se afirmar que ele desempenha um papel de um determinante estrutural. O que define que cada gênero estará suscetível a diferentes riscos.

Barata (2009), diz que as taxas de mortalidade masculina desde o nascimento até terceira idade são maiores em comparação com a feminina. Apenas em uma sociedade onde a assistência à gravidez e ao parto é altamente instável, a taxa de mortalidade entre as mulheres em certas idades é maior. Esse fato é atribuído às diferenças nos fatores e situações de risco ao longo da vida, sendo as diferenças geralmente maiores para os homens em ambientes de trabalho insalubres ou relacionadas a comportamentos prejudiciais como consumo excessivo de álcool, cigarro, drogas e outras situações de violência.

Há uma tendência maior dos homens serem vítimas de homicídio, não limita-se ao fato de serem mais vulneráveis à situações de perigo, mas também ao consumo de álcool e drogas ilícitas. No entanto, pode haver outros aspectos que estão diretamente relacionados à construção do papel do homem na sociedade, ou seja, um sentimento

primitivo de provação de masculinidade usando da violência, como relata Barata (2009). Por outro lado, a vitimização das mulheres tem relação a construção de papéis femininos na sociedade como objeto de posse do sexo masculino. Um exemplo típico desta situação é o homicídio de um parceiro, por vingança, traição, e ou como resultado do aumento da violência doméstica.

Após o ajuste para variáveis socioeconômicas (como escolaridade e renda), a diferença entre as condições de saúde masculina e feminina tornou-se mais pronunciada, indicando que determinantes estruturais não explicam as diferenças encontradas. Da mesma forma, a raça/cor e a faixa de idade não mudam essa relação, como relata Barata (2009).

3.1.2 Escolaridade

De acordo com o Artigo 21 da Lei de Diretrizes e Bases da Educação Nacional (Lei n o 9.394/96) a educação escolar é composta por **“I - educação básica, formada pela educação infantil, ensino fundamental e ensino médio; II - educação superior.”**

De acordo com o Instituto de Pesquisa Econômica Aplicada (IPEA), educação está dividida em cinco níveis: o primeiro nível inclui pessoas com ensino fundamental incompleto; o segundo nível abrange pessoas com ensino fundamental completo e ensino médio inacabado; o terceiro nível compõe de pessoas com ensino médio completo ou superior incompleto; e o quarto nível contém pessoas com ensino superior completo. Enquanto que o quinto nível engloba pessoas com pós-graduação, embora ainda em andamento.

Como o nível de escolaridade tem uma relação importante com o estilo de vida e os comportamentos relacionados à saúde de cada indivíduo, os estudos sobre desigualdade em saúde costumam usar a variável de escolaridade como um indicador das condições sociais. Assim Souza (2012), menciona que sujeitos com maior escolaridade formal têm maior autocuidado, participação em planos de prevenção e capacidade para solucionar problemas relacionados à saúde. Além disso, a educação é um indicador estável para adultos, fácil de medir e eficaz, desse modo essa variável se torna importante para os estudos de DSS.

3.1.3 Raça/Cor

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), as pesquisas sobre cor ou raça da população brasileira são autodeclaratórias. Ou seja, as pessoas respondem sobre sua percepção de raça/cor com base nas seguintes opções: branco, preto, pardo, indígena ou amarelo.

As desigualdades sociais relacionadas aos grupos étnicos têm uma decisão dupla de acordo com Barata (2009): uma é o status social que esses grupos ocupam na sociedade e a outra trata da aceitação ou rejeição que podem ter diante do grupo majoritário. No estudo da desigualdade social em saúde, a desigualdade racial ou étnica é geralmente atribuída às diferentes condições socioeconômicas ou valores culturais causados pela baixa penetração social desses grupos na sociedade. No entanto, mesmo depois de controlar os efeitos dessas variáveis, as diferenças ainda indicam os efeitos independentes de pertencer a uma determinada raça ou grupo étnico sobre a saúde.

Barata (2009) diz que nos estudos epidemiológicos, pelo menos três aspectos dessa relação são geralmente subestimados: o impacto do acúmulo de desvantagens ao longo da vida, a concentração de minorias étnicas na pobreza em áreas residenciais em degradação urbana e o impacto de fundo de ambientes sociais prejudiciais – ou seja, impacto de viver em uma sociedade considerada racista. No Brasil, embora nunca tenha existido um sistema de segregação social sólido, a população negra sofre sistematicamente com maior desvantagem social. Sendo uma população mais concentrada nas regiões mais pobres do país, com um nível de desenvolvimento humano inferior ao da população em geral e seu acesso a serviços básicos de saúde, educação e trabalho bastante reduzido. Fato que acaba por prejudicar seu desenvolvimento econômico, alimentando uma roda de pobreza para as demais gerações.

O baixo nível de renda per capita dos indivíduos e suas famílias restringe a liberdade pessoal e social dos sujeitos e cria um ambiente de fraqueza, cansaço e causador de doença (Brasil, 2007).

De acordo com Barata (2009), alguns problemas de saúde e comportamentos são mais vulneráveis à discriminação e racismo: transtornos mentais, pressão alta, baixo peso e nascimento prematuro, doenças cardíacas, diabetes e obesidade, alcoolismo, uso de drogas e tabagismo. Os efeitos do racismo e da discriminação na saúde podem ser agudos ou crônicos e podem atuar por meio do mecanismo fisiopatológico do estresse.

Barata (2009) aponta a experiência de como a exposição ao racismo e ao comportamento discriminatório pode atuar como um estressor agudo, enquanto viver em uma sociedade racista pode atuar como um estressor crônico. Essa influência pode ser alterada pela adaptação a um ambiente estressante de estilo e expectativas pessoais, estrutura da comunidade, fatores históricos e políticos de resposta e ao racismo.

3.1.4 Estado Civil

O estado civil é o termo que se refere ao estado em que o cidadão está em relação ao matrimônio e suas relações sociais afetivas. Na lei brasileira se diferem em: solteiro,

casado, separado, divorciado e viúvo, tendo também menção à união estável.

De acordo com o Ministério Público (MP), a pessoa que se classifica como casado deve possuir um casamento religioso e/ou civil. Estabelecendo a união plena, com base nos direitos e deveres realizado em um cartório ou em uma igreja. União estável, de acordo com o MP, é a relação de pessoas com convivência pública e duradora com o objetivo de construção familiar.

Solteiro refere-se à pessoa que nunca se casou legalmente; separado, a pessoa que quebra o vínculo jurídico do casamento e não poderá se casar até o fim do processo de divórcio e assim estará divorciado. Viúvo é a pessoa que tinha um casamento legal que fora interrompido pelo morte de seu cônjuge.

4 Metodologia

4.1 Fonte de Dados

O presente estudo foi realizado com dados do Sistema de Informações sobre Mortalidade - SIM -, disponíveis na plataforma do Datasus (Departamento e Informática do Sistema Único de Saúde do Brasil). Esse sistema contém um conjunto de ações relativas à coleta, codificação, processamento de dados, fluxo, consolidação e divulgação sobre os óbitos ocorrido no país (Portaria N° 116). O SIM foi criado pelo Ministério da Saúde em 1975 a fim de coletar os dados sobre óbitos no Brasil de forma contínua a partir de um instrumento padronizado “Declaração de Óbito (DO)”, que deve ser assinado por médico, o qual se responsabiliza por todas os dados declarados. Os dados de 2020 para GO foram fornecidos pela Secretaria de Estado da Saúde de Goiás. A Secretaria de Estado da Saúde do Distrito Federal também disponibilizou os dados para todo o período com informações detalhadas o que permitiu a identificação da RA de residência do falecido.

A variável causa básica de morte é codificada seguindo a Classificação Internacional de Doença (CID, atualmente na 10^a versão). Na CID 10 têm-se os códigos relativos às doenças, sintomas, aspectos, anomalias, circunstâncias sociais e causas externas. O SIM traz, além das causas de morte, dados sobre características do indivíduo e circunstâncias do óbito. Para óbitos fetais e menores de um ano, são coletados dados importantes para análise dos DSS para esses grupos específicos. O SIM constitui uma importante base de dados com informações detalhadas para a análise da mortalidade no Brasil. A longa série histórica (desde 1979) está disponível no portal do Datasus. A partir dos dados do SIM, é possível realizar análises comparativas nacional e internacionalmente.

4.1.1 Construção das variáveis de interesse dos Determinantes Sociais

Para a construção da variável raça/cor, foram agrupadas a cor branca e a cor amarela, por essa ter um numero relativamente baixo assim como também melhor se assemelhar com esse grupo, Já as informações sobre a mortalidade indígena foram retiradas por representarem um valor muito baixo, sendo um total de 36 mortes de indígenas registrado em 10 anos.

Na variável “idade” a informação está descrita em anos completos. Por ser uma variável quantitativa, será usada nesse trabalho para a correspondência, de modo que as variáveis qualitativas sejam transformadas em faixas etárias separadas em períodos de 10 anos. E assim, as quatro faixas etárias usadas no estudo serão: 30 a 39 , 40 a 49, 50 a 60 e 60 a 69 anos.

Outra variável importante a ser mencionada é a variável correspondente a causas de morte que terão o padrão da CID10, que conta com inúmeras doenças que serão fatores importantes para a construção do comportamento das tipicidade do óbito em diferentes grupos.

Por fim, conta-se com mais uma variável referente ao ano do óbito no intervalo de 2011 a 2020.

Variável	Descrição	Codificação	Tipo
IDADE: Faixa Etária	Valores numéricos	30-39 40-49 49-59 60-69	Qualitativa Ordinal
ESC: Escolaridade	1: Nenhuma 2: 1 a 3 anos 3: 4 a 7 anos 4: 8 a 11 anos 5: 12 e mais 9: Ignorado	Nenhum Fundamental 1 Fundamental 2 Médio Superior	Qualitativa Ordinal
ESTCIVIL: Estado Civil	1: Solteiro 2: Casado 3: Viúvo 4: Separado judicialmente 5: União Estavel 9: Ignorado	Solteiro Casado Viúvo Separado União	Qualitativa Nominal
RACACOR: Raça/Cor	1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena NA's	Branca Preta Parda	Qualitativa Nominal
CODMUNRES Municípios de residência do falecido da AMB	Grupos de 1 a 5: Municípios e RA's de residência do falecido agrupados NA's	GP1 GP2 GP3 GP4 GP5	Qualitativa Nominal
SEXO: Sexo	0: Ignorado 1: Masculino 2: Feminino	Masculino Feminino	Qualitativa Nominal

próximos ao DF. Os municípios goianos que compõem a AMB são: Águas Lindas de Goiás, Alexânia, Cidade Ocidental, Cocalzinho de Goiás, Cristalina, Formosa, Luziânia, Novo Gama, Padre Bernardo, Planaltina, Santo Antônio do Descoberto e Valparaíso de Goiás.

Para melhor visualização, optou-se pelo agrupamento dos municípios usados pela CODEPLAN de acordo com características de cada região administrativa (RA), dividindo-as em quatro grupos, de acordo com a renda, incluindo o grupo dos municípios de Goiás, ficando assim:

- Grupo 1 (alta renda): Brasília, Plano Piloto, Asa Norte, Asa Sul, Jardim Botânico, Lago Norte, Lago Sul, Park Way e Sudoeste/Octogonal
- Grupo 2 (média alta renda): Águas Claras, Candangolândia, Cruzeiro, Gama, Guará, Núcleo Bandeirante, Sobradinho, Sobradinho II, Taguatinga e Vicente Pires
- Grupo 3 (média baixa renda): Brazlândia, Ceilândia, Planaltina - DF, Riacho Fundo, Riacho Fundo II, SIA, Samambaia, Santa Maria e São Sebastião
- Grupo 4 (baixa renda): Fercal, Itapoã, Paranoá, Recanto das Emas, SCIA e Varjão
- Grupo 5 (Periferia Metropolitana de Brasília - PMB): Águas Lindas de Goiás, Alexânia, Cidade Ocidental, Cocalzinho de Goiás, Cristalina, Formosa, Luziânia, Novo Gama, Padre Bernardo, Planaltina-GO, Santo Antônio do Descoberto e Valparaíso de Goiás.

O estudo dessa variável é importante pelo fato de estar relacionada à renda e qualidade de vida em diferentes áreas da AMB, uma vez que renda é um importante medida no que se diz respeito aos DSS, uma vez associadas a outras variáveis pode aumentar ou diminuir a qualidade de vida.

4.3 Causas Básicas

A causa básica final é obtida após investigação do óbito e, ainda de acordo com a OMS, "causa básica da morte é (a) a doença ou lesão que iniciou uma sucessão de eventos e que termina com morte ou (b) nos casos de acidentes ou violências, as circunstâncias dos mesmos". Utilizaram os capítulos do CID 10 para estudar esta variável em relação ao comportamento dos DSS nos diversos eixos, chegando a 20 classes sendo divididas por capítulos, subcapítulos e por combinações de capítulos com poucas frequência para melhor explicação.

Capítulo	Descrição	Cód. CID-10
I	Algumas doenças infecciosas e parasitárias	A00-B99
II	Neoplasmas [tumores]	C00-D48
III	Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários	D50-D89
IV	Doenças endócrinas, nutricionais e metabólicas	E00-E90
V	Transtornos mentais e comportamentais	F00-F99
VI	Doenças do sistema nervoso	G00-G99
VII	Doenças do olho e anexos	H00-H59
VIII	Doenças do ouvido e da apófise mastóide	H60-H95
IX	Doenças do aparelho circulatório	I00-I99
X	Doenças do aparelho respiratório	J00-J99
XI	Doenças do aparelho digestivo	K00-K93
XII	Doenças da pele e do tecido subcutâneo	L00-L99
XIII	Doenças do sistema osteo muscular e do tecido conjuntivo	M00-M99
XIV	Doenças do aparelho geniturinário	N00-N99
XV	Gravidez, parto e puerpério	O00-O99
XVI	Algumas afecções originadas no período perinatal	P00-P96
XVII	Malformações congênitas, deformidades e anomalias cromossômicas	Q00-Q99
XVIII	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte	R00-R99
XIX	Lesões, envenenamentos e algumas outras consequências de causas externas	S00-T98
XX	Causas externas de morbidade e de mortalidade	V01-Y98
XXI	Fatores que influenciam o estado de saúde e o contato com os serviços de saúde	Z00-Z99
**	CID 10 ^a Revisão não disponível ou não preenchido ou inválido	U99, vazio ou inválido

Tabela 2: Capítulos CID-10

Fonte: CID-10

Desses capítulos, separam-se neoplasias em cinco classes (Neoplasias do sistema digestivo, neoplasias do sistema respiratório, neoplasias do sistema respiratório, neoplasia de mama e outras neoplasias), que foram separadas, pois o capítulo possuía um número relativamente maior em relação aos outros grupos. Assim como o capítulo XX foi separado em três causas: óbitos por acidentes, auto-provocados e agressões; o capítulo IX se divide em quatro causas sendo elas: doenças isquêmicas cardíacas, doença cerebrovascular, causas cerebrovascular, causas garbage cardíaca (causas garbage são causas de mortes não diagnosticadas para as quais ações de saúde pública não podem ser determinadas e/ou planejadas) e outras causas cardíacas; o capítulo II de causas infectuosas foi dividido em dois: covid-19 e demais causas infecciosas; os capítulos IV, X, XI não foram divididos, pois tinham um número relativamente grande para realizar as análises; os capítulos V e

VI foram juntados em um só; e os demais capítulos não mencionados foram reunidos em uma classe apenas.

Causas	Codigo CID 10
Covid-19 (Cap. I)	B342
Infecciosas (Cap. I)	A00- B341; B343-B99
Neoplasia de mama (Cap. II)	C50-C509 D059 D486
Neoplasia Sist. Digestivo (Cap. II)	C00-C26;
Neoplasia Sist. Reprodutor (Cap. II)	C51-C63
Neoplasia Sist. Respiratório (Cap. II)	C30-C40
Outras Neoplasias (Cap. II)	C41-D89
Endócrinas (Cap. IV)	E00-E90
Mental & Sist. Nervoso (Cap. V & VI)	F00-G99
Cerebrovascular (Cap. IX)	I601-I609; I620; I64 I67-I69
Isquemia cardíaca (Cap. IX)	I20-I25
CG.cardíaca (Cap. IX)	I10; I46-I517; I950
Doença cerebrovascular (Cap. IX)	I61-I63; I672-I693
Demais cardio circulatória (Cap. IX)	I44-I99
Digestivas (Cap. XI)	K00-K93
Respiratória (Cap. X)	J00-J99
Acidentes (Cap. XX)	V00-X59
Agressões (Cap. XX)	X85-Y98
Auto-provocadas (Cap. XX)	X60-X84
Outras (Cap. faltantes)	D50-D89; H00-H95;L00-T98; Z00-Z99;U99

Tabela 3: Grupos de causas CID-10

4.4 Análise de Correspondência

Desenvolvida por Jean-Paul Benzcri na década de 60, “A análise de correspondência é uma técnica de representação gráfica em projeção plana das relações multidimensionais das distâncias χ^2 entre as categorias das variáveis estudadas” (BATISTA et al., 2004), sendo chamada de método exploratório, pois busca identificar associações entre duas ou mais variáveis categóricas, utilizando gráficos de dispersão para análise visual, além de transformar dados não métricos em métricos criando então, uma tabela de frequência com duas ou mais entradas, e como trata de um valor de contagem não existe valores negativos nesta tabela. A Análise de Correspondência (AC) transforma as linhas e coluna de uma tabela em unidades correspondentes. Sendo a um método estatístico de interdependência, utilizado para estudar as relações entre variáveis qualitativas e seu comportamento associativo entre suas categorias.

A técnica permite que as dimensões originais dos dados sejam reduzidas. Um ponto interessante da AC é que, com base na quantidade de variáveis categóricas, as quantidades geradas para a dimensão podem ser tratadas como variáveis contínuas. Ou seja, essa técnica é muitas vezes utilizada para transformar um grande número de variáveis categóricas em um menor número de variáveis contínuas.

O objetivo da AC é gerar um procedimento gráfico para representar associações e uma tabela de frequência ou contagens utilizando variáveis categóricas ou qualitativas. Dessa maneira, o método de análise se relaciona com a nomeação dos eixos e faz a verificação de como as características se associam no espaço, ou seja, como os dados se inter-relacionam por meio de um mapa percentual.

4.4.1 Qui-quadrado e inércia

A disposição gráfica dos perfis, mede de modo implícito as distâncias entre eles e interpreta as suas posições. Com AC, as distâncias entre os perfis são medidas de forma diferente. Calcula-se o quadrado da diferença entre a frequência observada e a frequência esperada dividindo-o pela frequência esperada das características interessadas para chegar no valor final: a Qui-quadrado (χ^2):

$$\chi^2 = \sum \frac{(\text{observado} - \text{esperado})^2}{\text{esperado}}, \quad (4.4.1)$$

no qual a hipótese de independência segue uma distribuição quadrada $(I - 1)(J - 1)$ graus de liberdade.

A AC está mais interessada na competência do χ^2 de medir a falta de homoge-

neidade, ou seja, medir a heterogeneidade entre as variáveis, do que no teste estatístico de homogeneidade descrito.

Pode-se expressar a inércia com a seguinte forma:

$$Inércia = \sum_i (i - \text{ésimo massa})(\text{distância} \chi^2 i - \text{ésimo perfil da média})^2. \quad (4.4.2)$$

Sendo a inércia alta quando os perfis da linha apresentarem grandes desvios à sua média, ou seja, tais desvios estarem longe da média da inércia; e sendo baixa quando esses desvios se encontrarem perto da sua média.

Em AC a *inércia total*, o valor χ^2/n , onde n é o total da tabela. Sendo este valor uma média da variância total da tabela independente de seu tamanho, podendo expressar a inércia com ϕ^2 . A inércia (total) de uma tabela quantifica a variação existente nos perfis de coluna; e linhas, sendo que cada linha e coluna contribui para a inércia total. Valores altos na inércia podem indicar uma forte associação entre colunas e linhas.

$$\phi^2 = \frac{\chi^2}{n}. \quad (4.4.3)$$

$$\phi^2 = \frac{\chi^2}{n} = \sum_i r_i \|a_i - c_i\|_c^2 = \sum_i r_i \sum_j \left(\frac{p_{ij}}{r_i} - c_j \right)^2 / c_j, \quad (\text{por linha}) \quad (4.4.4)$$

$$\phi^2 = \frac{\chi^2}{n} = \sum_i c_j \|b_i - r\|_r^2 = \sum_i c_j \sum_j \left(\frac{p_{ij}}{c_j} - r_i \right)^2 / r_i. \quad (\text{por coluna}) \quad (4.4.5)$$

Sendo a equação de inércia denotada por (Greenacre, 2017):

- n_{ij} : elemento da tabela de contingencia situado na i -ésima linha e na j -ésima coluna.
- n : ou n , o total da tabela.
- p_{ij} : n_{ij}/n
- r_i : o tamanho da i -ésima linha; simboliza-se o tamanho do vetor por \mathbf{r} .
- c_j : o tamanho da j -ésima coluna; simboliza-se o tamanho do vetor por \mathbf{c} .
- a_i : simboliza-se a categoria da linha
- b_j : simboliza o perfil da coluna

- a_i : simboliza a categoria da linha.

A distância entre χ^2 e o centroide \mathbf{c} pode ser medida como sendo a distância euclidiana ponderada. A distância entre os raiz quadrada da diferença o quadrado do valores esperados e valores observado do perfil, Greenacre (2017). Sendo a distância Euclidiana Ponderada (DEP) definida com a seguinte equação:

$$DEP = \sqrt{\sum_{j=1}^p w_j (x_j - y_j)^2}, \quad (4.4.6)$$

w_j são os valores positivos e x_j e y_j com $j=1, \dots, p$, eles são pontos em um espaço p -dimensional onde essa dimensão é definida por variáveis contínuas e é medida em várias escalas também usadas na Análise de Componentes Principais (ACP), um método relacionado à AC. No ACP, é removido o impacto da escala na variância e divide os dados por seus desvios padrão. s_j das variáveis. Assim substitui x_j e y_j das variáveis por x_j/s_j e y_j/s_j respectivamente. Assim a DEP possui peso $w_j = 1/s_j$, os inversos das variância. Que por definição a χ^2 entre os perfil, sendo os valores igual $w_j = 1/c_j$ são iguais ao inverso dos elementos do perfil médio.

Apesar que que AC os perfis tem a mesma escala de frequências relativas, segue a necessidade de compensar as diferenças entre as variáveis e assim as distâncias χ^2 cumprem o princípio de equivalência distribucional, que garante a estabilidade entre as distâncias das linhas quando dividida pela coluna com elementos similares, ou quando unimos colunas similares.

4.4.1.1 Contribuição

Dependendo do número das variáveis de interesse contidas no banco de dados, a análise gráfica por ser complexa devido ao grande numero de informação. Por causa desse obstaculo são usadas algumas estatísticas para reforçar a análise dos resultados sendo estas estatísticas chamadas de “contribuição” divididas em duas principais: a contribuição relativa e a contribuição absoluta. Elas propiciam identificar quais posições mais contribuem para explicações e formações dos eixos e, assim, conseguem quantificar um valor individual.

4.4.1.1.1 Contribuição Absoluta

Quando um ponto apresenta a contribuição na caracterização do eixo é chamando de contribuição absoluta e assim entre os coeficientes pode-se diferenciar quais pontos devem ser considerado para análise do eixo principal.

O ponto \mathbf{a}_i , representa um perfil no espaço multidimensional, por exemplo o i -ésimo perfil, de tamanho r_i , a uma distância d_j do perfil da linha média \mathbf{c} . Pela equação 4.4.2 a inércia total é igual a $\sum_i r_i d_i^2$. Sendo a coordenada principal simbolizada por \mathbf{a}_i no eixo principal k por f_{ik} . Assim a contribuição absoluta sera:

$$C_{ik}^{(a)} = \sum_i r_i f_{ik}^2 \quad e \quad (4.4.7)$$

onde a soma das contribuições absoluta é igual a 1:

$$\sum_i C_{ik}^{(a)} = 1.$$

4.4.1.1.2 Contribuição Relativa .

Quando um eixo contribui para a caracterização de um ponto é chamada de contribuição relativa. Tendo cada ponto uma importância diferente em determinado eixo. Em relação a contribuição relativa a linha dela está definida por:

$$\frac{f_{ik}^2}{\sum_k f_{ijk}^2}. \quad (4.4.8)$$

Sendo $f_{ik} = \frac{1}{s_k} \sum_j \left(\frac{p_{ij}}{p_i} \right) g_{jk}$ coordena do i -ésimo perfil da linha sobre o k -ésimo eixo principal.

De forma semelhante, pode-se considerar a contribuição relativa do k -ésimo eixo ao j -ésimo perfil coluna, definida como:

$$\frac{g_{jk}^2}{\sum_k g_{jk}^2}. \quad (4.4.9)$$

Sendo $g_{jk} = \frac{1}{s_k} \sum_i \left(\frac{p_{ij}}{p_j} \right) f_{ik}$ coordena do j -ésimo perfil da coluna sobre o k -ésimo eixo principal.

A contribuição relativa ainda pode ser vista como $\cos^2(\theta)$. Considerando a figura 3 onde f_{ik} trata-se da coordenada principal de a_i em junção ao k -ésimo eixo principal; r_i trata-se da massa do i -ésimo perfil da linha a_i ; d_i trata-se da distância χ^2 de a_i ; e o angulo θ_{ik} é formado pelo vetor a_i e o k -ésimo eixo principal relacionado ao centroide \mathbf{c} .

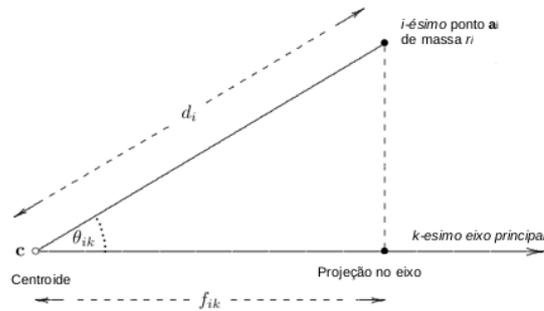


Figura 3: Representação gráfica de um perfil a_i em um espaço multidimensional

Fonte: Tradução Greenacre, (2017)

Assim, a contribuição relativa pode ser descrita da seguinte forma:

$$C_{ik}^{(r)} = \cos^2(\theta) = \frac{f_{ik}^2}{d_i^2}, \quad (4.4.10)$$

onde $d_i^2 = f_{i1}^2 + \dots + f_{ik}^2 = \sum_k f_{ik}^2$. A soma dessa contribuição atinge o valor máximo de 1 e um valor mínimo de 0, sendo perfil linha perpendicular ao eixo principal, e assim conclui-se que a contribuição relativa é um indicador de qualidade para representação de um ponto no eixo.

4.4.2 Redução dimensional

Os dados constituídos por m categorias estão localizados exatamente, em espaços de dimensionalidade $m - 1$. Portanto, dados com mais de quatro categorias estão localizados em espaços de dimensionalidade maior que três que não podem ser observados diretamente.

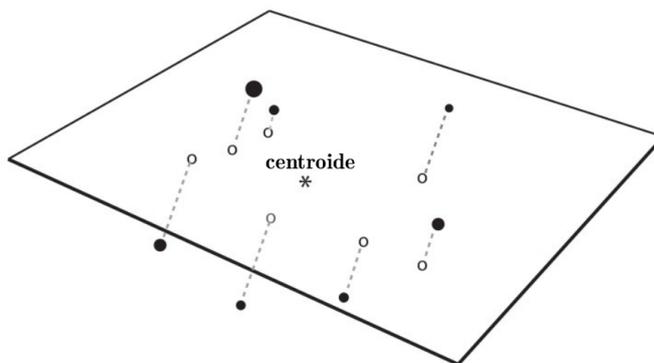


Figura 4: Perfis em um espaço multidimensional

Fonte: Greenacre, (2017)

Para visualizar a posição relativa dos perfis, é necessário reduzir essa dimensionalidade, o que significa perda de informação, uma vez que essas projeções não têm suas verdadeiras posições.

O critério para a seleção de (S) é calcular a proximidade de todos os perfis a ele como uma soma ponderada das distâncias quadradas e minimizá-la de forma que (S) passe pelo centroide dos pontos. Tendo i perfis e S um candidato a subespaço com poucas dimensionalidades no espaço original, referindo-se a não mais que duas ou três dimensões.

Simboliza como $d_i(S)$ a distância χ^2 entre o i -ésimo perfil da massa m_i e S . Calcula-se a proximidade deste perfil para o subespaço como $m_i [d_i(S)]^2$, Calcula-se a proximidade de todos os perfis a S , como a soma destes valores:

$$\text{proximidade a } S = \sum_i m_i [d_i(S)]^2. \quad (4.4.11)$$

Pode-se observar que o subespaço S buscado deve passar pelo centroide do ponto do perfil no espaço multidimensional (Figura 4) e o plano passando pelo espaço, portanto, deve considerar apenas o subespaço que contém o centroide, para obter uma medida única da proximidade de todos os pontos à linha, poderia ser a soma das distâncias de todos os perfis a essa linha fictícia. E assim o plano mais adequado para a direção do quadrado mínimo deve passar pelo centroide dos pontos.

Assim que o (S) ótimo é encontrado, os perfis devem ser projetados sobre ele e suas projeções devem ser observadas como uma aproximação de suas verdadeiras posições no espaço de dimensão superior. Para determinar a discrepância entre a posição exata e a aproximada, utiliza-se a inércia total dos perfis, que representa uma medida da variabilidade total, ou seja, uma medida da extensão geométrica dos pontos a partir de suas posições reais. Desse modo a precisão dos ajustes se expressa pela porcentagem da inércia.

4.4.3 Elementos suplementares

Para a construção de um gráfico de dispersão usado para análise de AC, e baseia-se em pontos dos indivíduos e categorias ativas. Sendo chamado de pontos ativos as linhas e colunas de uma tabela utilizada para análise de correspondência. Sendo estes os pontos que contribuem para a construção dos mapas de dispersão de baixa dimensionalidade.

Os elementos suplementares (pontos passivos) são as linhas ou colunas da tabela que não participaram na construção do mapa, são pontos que estão presentes nos espaços completos dos perfis de linha ou coluna. Pode projetá-los em um mapa de baixa dimensão para interpretar suas posições em relação aos pontos de acesso.

Os pontos suplementares não são adicionados à solução CA, porém pode-se calcular as contribuições dos eixos principais (em termos de cosseno ou correlação quadrada). Com base nessas contribuições, pode-se julgar se os pontos adicionais estão bem representados no mapa.

4.4.4 Análise de Correspondência Múltipla

A Análise de Correspondência Múltipla (ACM) permite avaliar as relações entre as categorias de diversas variáveis não métricas (qualitativas). As variáveis na ACM podem ser consideradas variáveis ativas ou variáveis suplementares. O espaço ativo é utilizado para definir o espaço geométrico no qual o conjunto de dados será avaliado, e faz parte do cálculo inercial e das principais coordenadas das linhas e colunas. Ao contrário, as variáveis suplementares não interferem no cálculo da ACM, mas podem ser usadas para ajudar a explicar os resultados do gráfico de correspondência.

Na ACM não é possível verificar simultaneamente se existe associação entre duas ou mais variáveis para sua análise, pois as estatísticas do teste χ^2 são calculadas apenas com base em uma tabela de contingência bidimensional. Sendo recomendado que seja elaborado um teste χ^2 para cada par de variáveis antes da elaboração de uma ACM. Se caso uma não apresente associação estatisticamente significativa a nenhuma das outras variáveis então recomenda-se que seja excluída do estudo.

Fávero e Belfiore (2017), informa que para explicar o ACM em detalhes, é necessário apresentar o conceito de matriz binária. Seja um banco de dados com N observações e Q variáveis onde $Q > 2$, e com cada variável ($q = 1, \dots, Q$) possuindo k_q categorias, assim o número de categorias envolvidas para este tipo de análise é:

$$J = \sum_{q=1}^Q J_q. \quad (4.4.12)$$

Assim a Tabela 4 apresenta a forma mostra esquematicamente um banco de dados com N observações e Q variáveis com mais de duas variáveis categóricas. Note que com base na Tabela 4 $J_1 = 1$ e $J_2 = 2$ E $J_Q = 4$.

Com base no banco de dados da Tabela 4 sera possível formar um novo banco de dados apenas com variáveis binarias transformado a variável não-métrica (qualitativa) em uma variável métrica (quantitativa) designando-se 0 ou 1 a uma categoria e ou objeto transforma-se em uma variável dicotômica.

Observação	Variável q			
	1	2	...	Q
1	categoria ₂	categoria ₁	...	categoria ₂
⋮	⋮	⋮	⋮	⋮
N	categoria ₃	categoria ₂	...	categoria ₁
Número de Categorias J _q	1	2	...	4

Tabela 4: Banco de dados com N observações e (Q > 2) variáveis categóricas.

A matriz binária $X = [X_1, X_2, \dots, X_Q]$ tem a codificação de zeros e uns das variáveis originais por intermédio da qual pode ser definida a inércia essencial da ACM, onde o calculo depende apenas da número total de variáveis introduzidas na análise e da quantidade de categorias de cada uma delas, em conformidade com as frequências absoluta de todas as categorias.

Observação	Variável 1		Variável 2			...	Variável Q			
	cat ₁	cat _n	cat ₁	...	cat _n	...	cat ₁	...	cat _n	
1	1	...	0	0	...	0	...	0	...	1
2	0	...	1	1	...	0	...	0	...	0
3	0	...	0	0	...	1	...	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
N	0	...	1	1	...	0	...	0	...	1

Tabela 5: Codificação binária das categorias das variáveis originais

Para cada matriz X é composta por uma matriz X_q, onde X pode ser defina com a inércia principal parcial da dimensão fundamental normalmente igual a 1, e a inércia principal total, equivalente a J_q - 1. Assim a inércia principal total corresponde à média das inércias principais totais das matrizes X_q que a constituem, ou seja, deve ser obtida por intermédio da expressão:

$$I_{Total} = \frac{\sum_{q=1}^Q (J_q - 1)}{Q} = \frac{J - Q}{Q} = \frac{\chi^2}{n}. \tag{4.4.13}$$

Mediante a codificação binária, expõe que a matriz X de J - Q dimensões, pode definir as coordenadas de cada uma das categorias das variáveis inseridas na ACM, para assim poder ser montado o mapa percentual. Essas coordenadas são conhecidas como coordenadas padrão.

Na ACM inicialmente usa-se a matriz indicadora e Burt para explorar a relação entre os grupos de variáveis categóricas. Os elementos contidos na matriz $X = [X_1, \dots, X_Q]$ possibilitam criar a matriz Burt (B), que corresponde a uma tabulação cruzada (produto

interno) de todas categorias e variáveis, em outros termos, $B = X'X$, cuja matriz é simétrica. Segue:

$$B = \begin{bmatrix} X'_1X_1 & X'_1X_2 & \cdots & X'_1X_Q \\ X'_2X_1 & X'_2X_2 & \cdots & X'_2X_Q \\ \vdots & \vdots & \ddots & \vdots \\ X'_QX_1 & X'_QX_2 & \cdots & X'_QX_Q \end{bmatrix}_{J \times J}.$$

Na diagonal da matriz $B = \{b_{kh}\}$ estão os totais de cada categoria, além da soma das categorias de cada variável tem que ser igual ao número de observações. Assim mencionada as propriedades da matriz de Burt, considere de que J categorias de k variáveis categóricas, então a matriz de correspondência será igual a versão simples, ou seja, $P = \{p_{kh}\}$, em que $p_{kh} = \frac{b_{kh}}{b_{++}}$ e $b_{++} = \sum_{k=1}^J \sum_{h=1}^J b_{kh}$.

A estrutura de associação na matriz \mathbf{S} é revelada usando o decomposição de valor singular (DVS)

$$S = U\Phi V', \quad \phi_1 \geq \phi_2 \geq \cdots \geq \phi_\infty \quad (4.4.14)$$

onde S é a classificação da matriz \mathbf{S} . E Φ trata-se de uma matriz diagonal de valores singulares $s_k = \sqrt{\lambda_k} : k = 1, \dots, K$, onde λ_k são os autovalores de XX' ou $X'X$, sendo $\lambda_1 \geq \cdots \geq \lambda_k$; e a correspondente matriz de autovetores (\mathbf{V}), de maneira que $v_{kh} > 0$ para o primeiro q. As colunas U são chamadas de vetores singulares a esquerda e a V vetores singulares a esquerda, sendo ortonormais, pois $U'U = V'V = I$.

Segundo Greenacre (2006), a inércia total é medida pela inércia média de todos os blocos fora da diagonal de X, calculada diretamente das próprias tabelas ou ajustando a inércia total de \mathbf{B} removendo as contribuições fixas dos blocos diagonais do seguinte modo:

$$\text{inércia fora da diagonal média} = \frac{Q}{Q-1} \left(\text{inércia}(B) - \frac{J-Q}{Q^2} \right). \quad (4.4.15)$$

Para Greenacre (2006) as partes da inércia são então calculadas a partir das inércias principais $\sum_t \lambda_t^2$ de \mathbf{B} (ou das inércias principais $\sum_t \lambda_t$ de \mathbf{X}) da seguinte forma: para cada $\sum_t \lambda_t \leq 1/Q$, calcule as inércias ajustadas:

$$\lambda_t^{adj} = \left(\frac{Q}{Q-1} \right)^2 \left(\lambda_t - \frac{1}{Q} \right)^2, \quad (4.4.16)$$

na qual é o número de variáveis ativas. Nessa versão ajustada, o total de inércia é:

$$\lambda_t^{adj} = \left(\frac{Q}{Q-1} \right)^2 \sum_t \lambda_t^2 - \frac{(J-Q)}{Q^2}, \quad (4.4.17)$$

e depois expresse-os como porcentagens da Equação 4.4.16, sendo elas recomendadas para aplicação da ACM, pois melhoram os resultados.

O DVS fornece todos os resultados para produzir o gráfico para análise de correspondência. As coordenadas principais e padrão são calculadas por:

$$\text{coordenadas principal das linhas} : F = D_r^{-1/2} U \Phi, \quad (4.4.18)$$

$$\text{coordenadas padrão das linhas} : A = D_r^{-1/2} U, \quad (4.4.19)$$

$$\text{coordenadas principal das colunas} : G = D_c^{-1/2} V \Phi, \quad (4.4.20)$$

$$\text{coordenadas padrão das colunas} : B = D_c^{-1/2} V. \quad (4.4.21)$$

Assim para um mapa bidimensional usaria as duas primeiras colunas das matrizes de coordenadas F e G para um mapa simétrico e se assimétrico as duas primeiras colunas de A e G.

A interpretação dos valores de coordenadas, valores de qualidade e outras estatísticas são semelhantes, mas envolvem toda a inércia associada a toda a matriz de indicadores. O efeito ACM é semelhante ao resultado para as coordenadas da coluna na exibição do indicador de matriz.

Em suma, a ACM faz parte do tópico de redução de dimensionalidade. O objetivo principal da AC é encontrar um dimensão menor onde a amostra possa ser bem representada que maximiza o percentual de explicação da variância dos próprios dados, afim de apurar a existência de padrões nas dimensões.

5 Análises e Resultados

Os dados foram filtrados formando um novo conjunto de dados, contendo apenas as variáveis que serão usadas nesse estudo. Como se tratava de muitos bancos de dados e apesar de pertencerem ao mesmo sistema, foram encontradas diferenças entre eles em relação ao número de variáveis como, por exemplo, a informação do SIM de 2019 não conta mais com CRM, assim como as nomenclaturas de dados que diferem dependendo do ano pesquisado. Os dados do DF foram fornecidos pela Secretaria de Saúde, pois conta com uma variável que pode estratificar as Regiões Administrativas, variável não presente nos dados do SIM. Os dados de Goiás de 2020 foram disponibilizados pela Secretaria de Saúde de Goiás e os demais anos foram retirados do DATASUS. No primeiro momento foi consolidado o banco de dados e identificados quantos dados das variáveis de interesse ficaram sem o preenchimento no DF e Goiás. Foi retirado do banco as linhas com valores vazios e com dados apontados como ignorado, para uma melhor visualização do perfil, uma vez sendo necessário que os dados estejam inteiramente preenchidos.

5.1 Análise exploratória

O banco de dados contém 78.009 informações de mortalidade de 2011 a 2020, porém 67.994 indivíduos são válidos, ou seja, aqueles que foram devidamente preenchidos em todas as variáveis, sendo 62,8% do sexo masculino, com idade entre 30 e 69 anos, com predominância na faixa etária 60 a 69 anos (35,8%). Quanto à situação conjugal, 41% eram solteiros e 5,6% casados. Entre os casos dos óbitos, 25,7% estavam no ensino fundamental 1 e a cor parda corresponde 55,1%. As principais causas de mortalidade foram as agressões (7,97%) seguida por doença isquêmica do coração com (7,75%), como na tabela:

Variáveis	Frequência	Proporção
Sexo		
Masculino	46.369	62,8%
Feminino	27.496	37,2%
Raça		
Branca	26.406	35,7%
Preta	6.710	9,1%
Parda	40.714	55,1%
Faixa Etária		
30-39	11.321	15,3%
40-49	15.186	20,6%
50-59	20.927	28,3%
60-69	26.432	35,8%

Grupo			
	GP1	4.762	6,4%
	GP2	16.663	22,6%
	GP3	24.272	32,9%
	GP4	5.283	7,2%
	GP5	22.886	31,0%
Escolaridade			
	Nenhuma	7.234	9,8%
	Fundamental 1	18.953	25,7%
	Fundamental 2	17.473	23,7%
	Medio	16.710	22,6%
	Superior	8.200	11,1%
Estado Civil			
	Solteiro	30.293	41,0%
	Casado	26..262	35,6%
	Viuvo	4.406	6,0%
	Separado	8.831	12,0%
	Uniao	2.713	3,7%
Ano			
	2011	6.699	9,1%
	2012	6.955	9,4%
	2013	6.759	9,2%
	2014	7.191	9,7%
	2015	7.309	9,9%
	2016	7.317	9,9%
	2017	7.371	10,0%
	2018	7.344	9,9%
	2019	7.469	10,1%
	2020	9.452	12,8%
Causa			
	Acidentes	5.017	7,38%
	Agressões	5.417	7,97%
	Auto-provocada	1.164	1,71%
	Cerebrovascular	2.382	3,50%
	CG. cardíaca	2.261	3,33%
	Covid-19	1.619	2,38%
	D. cerebrovascular	2.302	3,39%
	Demais cardiovascular	3.736	5,49%
	Doenças Digestivas	4.980	7,32%

Doenças Endócrinas	3.691	5,43%
Doenças Infecciosas	3.534	5,20%
Doenças Respiratórias	3.535	5,20%
Isq. cardíaca	5.271	7,75%
Mental & Sist. Nervoso	3.348	4,92%
Neo. de mama	1.745	2,57%
Neo. Sist. Digestivo	4.292	6,31%
Neo. Sist. Reprodutivo	1.810	2,66%
Neo. Sist. Respiratório	2.942	4,33%
Out. Neoplasias	4.544	6,68%
Outras	4.404	6,48%
Total	67994	100%

Tabela 6: Caraterística do óbito. AMB, 2011-2020

5.2 Qui-quadrado

Associações	χ^2	GL	P-valor
Causas x Raça	2.521,1	38	<0,001
Causas x Sexo	7.719,7	19	<0,001
Causas x Escolaridade	2.808,5	76	<0,001
Causas x Idade	14.071	57	<0,001
Causas x Estado Civil	4.106,7	76	<0,001
Causas x Ano	12.600	171	<0,001
Causas x Grupo	2.466,6	76	<0,001
Raça x Escolaridade	4064,4	8	<0,001
Raça x Faixa Etária	1.411,9	6	<0,001
Raça x Sexo	485,81	2	<0,001
Raça x Ano	107,31	18	<0,001
Raça x Grupo	3.900,7	8	<0,001
Raça x Estado Civil	1.425,3	8	<0,001
Escolaridade x Faixa Etária	2.623,1	12	<0,001
Escolaridade x Sexo	279,9	4	<0,001
Escolaridade x Ano	864,3	36	<0,001
Escolaridade x Grupo	1.4412	16	<0,001
Escolaridade x Estado Civil	2.376,5	16	<0,001
Sexo x Faixa Etária	801,81	3	<0,001
Sexo x Ano	12,425	9	0,1904
Sexo x Grupo	114,43	4	<0,001

Sexo x Estado Civil	2.213,8	4	<0,001
Faixa Etária x Ano	294,54	27	<0,001
Faixa Etária x Grupo	1.160,7	12	<0,001
Faixa Etária x Estado Civil	1.0431	12	<0,001
Ano x Grupo	71,419	36	<0,001
Ano x Estado Civil	256,77	36	<0,001
Grupo x Estado Civil	1.132,2	16	<0,001

Tabela 7: Tabela Qui-Quadrado (χ^2)

Usando o Teste Qui-quadrado de *Person* ao nível de significância de 5% foi verificado a associações de cada par de variáveis. Ao olhar para a Tabela 7, como o p-valor é menor que o nível de significância, verificou-se que quase todas as associações aconteceram, ou seja, o número de mortalidade não ocorre de forma independente entre os fatores analisados. Vale ressaltar, que a variável relativa ao ano foi trabalhado como variável categórica. Além disso, sexo e ano apresentaram um p-valor maior que 5% indicando que não há evidências para concluir que sexo e ano estão associados. Contudo, apesar desse par de variável apresentar esta rejeição, optou-se por não retirá-lo posto que tais variáveis não repetem os resultados em outras interações com variáveis diferentes. Também foi levado em consideração que a variável ano é suplementar e não interfere no cálculo das dimensões das variáveis ativas no estudo da ACM.

5.3 Análise de Correspondência Múltiplas

O banco de dados usado tem mais de 80 variáveis e muitas são de sistemas, portanto houve uma filtragem com seleção daquelas que seriam relevantes para o estudo de causas básicas em relação aos determinantes sociais. Com o banco consolidado, manteve-se $Q=6$ variáveis ativas é: $K = 3 + 2 + 5 + 4 + 5 + 5 = 24$ categorias, sendo 3 Raça/Cor (branca, parda e preta), 2 é numero de Sexo, 5 as faixas de Escolaridade, 4 a quantidade de Faixas Etárias, 5 grupos de cidades e 5 Estado Civil. Onde a variância da nuvem de pontos é a mesma da nuvem indivíduo, ou seja é: $(K/Q) - 1 = (24/6) - 1 = 3$, sendo a variância da nuvem de pontos. Sua dimensionalidade é dada por $K - Q = 24 - 6 = 18$, que é o numero máximos de eixos para compreender a variância total da nuvem.

Com base na variância total e na dimensionalidade da nuvem das categorias, pode-se obter a variância média por meio de $\bar{\lambda} = ((K/Q) - 1)/(K - Q) = 3/18 = 0.1667$. Assim, ao conhecer a variância média, foram considerados os primeiros eixos com autovalores acima de $\bar{\lambda}$, que na Tabela 8, são os oito primeiros, que explicam 53,4219% da variância da nuvem de palavras.

Os autovalores correspondem a informação retida em cada dimensão. As di-

mensões são organizadas de forma decrescente e ordenadas de acordo com o valor de variação da explicação na solução. Ou seja na Tabela 8, a dimensão 1 explica a maior parte da variação na solução, seguida pela dimensão 2 e assim por diante.

Dimensão	Autovalores	%	% acumulada
1	0,299	9,964	9,964
2	0,237	7,914	17,878
3	0,195	6,502	24,38
4	0,182	6,062	30,442
5	0,177	5,892	36,334
6	0,174	5,814	42,148
7	0,17	5,674	47,822
8	0,168	5,597	53,419

Tabela 8: Principais inércias e porcentagem da matriz indicadora das médias

Pode-se observar na Tabela 8 que existe a possibilidade de usar mais de 8 dimensões. Porém é improvável que essas dimensões contribuam significativamente para a interpretação da característica da associação entre colunas e linhas. A porcentagem de contribuição dessas variáveis são altas em outras dimensões, por exemplo, a categoria grupo 5 (que representa os municípios de Goiás) da variável grupo de cidades contribui 10,97% (ver Tabela 9) para o eixo 5; a categoria Casado da variável estado civil tem uma

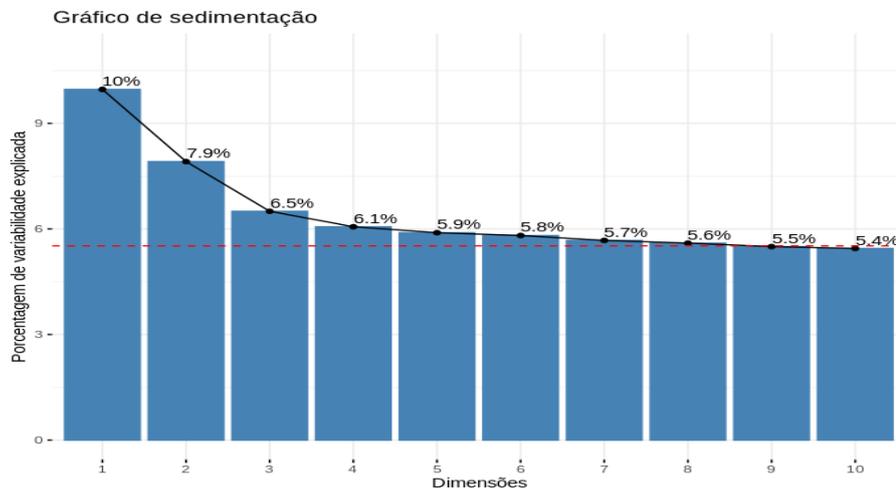


Figura 5: Gráfico de sedimentação

contribuição no eixo 4 de 16,81%; Porém, as variâncias nesses eixos têm valores proporcionalmente menores, 0,195, 0,182 e 0,177, respectivamente no eixo 3, 4 e 5; e dadas as contribuições de outras variáveis nesses eixos, esses percentuais foram insuficientes para refletir uma possível interpretação dessas variáveis nesses eixos.

Para a interpretação de cada eixo, as categorias apresentadas contribuíam su-

periores ou próxima da $cont = 1/K = 1/24 = 0.0,0417 = 4,017\%$. Essas contribuições estão destacadas em negrito na Tabela 9. As categorias consideradas no primeiro eixo responderam por 78,84% de sua variância, enquanto as categorias selecionadas no segundo e terceiro eixos representaram 59,57% e 86,81%, respectivamente.

Categorias	Contribuição em %					Cos2 em %				
	Eixos					Eixos				
	1	2	3	4	5	1	2	3	4	5
Branca	12,62	0,57	0	0,69	0,04	35,3	1,3	0	1,2	0,1
Parda	6,14	0,01	0,1	0,26	0,5	24,5	0	0,3	0,6	1,2
Preta	0,93	1,7	0,59	0,15	4,52	1,8	2,7	0,8	0,2	5,3
Total	19,69	2,28	0,69	1,1	5,06	61,6	4	1	2	6,5
Feminino	1,77	6,12	1,83	13,96	4,09	5	13,9	3,4	24,3	6,9
Masculino	1,05	3,63	1,09	8,28	2,43	5	13,9	3,4	24,3	6,9
Total	2,82	9,75	2,92	22,24	6,52	10,1	27,8	6,8	48,5	13,8
Nenhuma	0,44	10,99	9,86	0,16	5,94	0,9	17,5	12,9	0,2	7
Fund. 1	1,1	5	1,12	9	7,06	2,7	9,8	1,8	13,6	10,4
Fund. 2	3,83	0,85	0,14	0,58	0,22	9,2	1,6	0,2	0,9	0,3
Médio	0,47	2,32	13,01	11,17	6	1,1	4,4	20,1	16,1	8,4
Superior	16,67	8,85	10,67	0,91	1,12	34	14,3	14,2	1,1	1,3
Total	22,51	28,01	34,8	21,82	20,35	47,9	47,6	49,3	31,9	27,5
30-39	5,55	10,28	4,77	4,81	2,16	11,8	17,4	6,6	6,2	2,7
40-49	1,96	2,98	0,03	0,22	2,13	4,4	5,3	0,003	0,3	2,8
50-59	0,14	0,03	4,92	0,68	4,76	0,3	0,001	8	1	7
60-69	5,32	12,88	0,16	1,17	4,34	14,8	28,4	0,3	2	7,1
Total	12,97	26,17	9,88	6,88	13,39	31,3	51,2	15	9,5	19,7
GP1	14,65	5,45	16,19	2,17	3,01	28,1	8,3	20,3	2,5	3,4
GP2	4,62	0,99	7,18	7,65	5,93	10,7	1,8	10,9	10,8	8,1
GP3	0,9	1,11	3,57	0	9,04	2,4	2,4	6,2	0	14,3
GP4	1,06	0,05	0,09	1,58	9,49	2	0,1	0,1	1,9	10,8
GP5	4,6	0,55	5,1	1,31	10,97	11,9	1,1	8,6	2,1	16,8
Total	25,83	8,15	32,13	12,71	38,44	55,2	13,7	46,1	17,3	53,5
Casado	3,19	0,19	4,92	16,81	3,07	9	0,4	9	28,7	5,1
Separado	3,24	0,05	5,08	1,86	3,81	6,6	0,1	6,8	2,3	4,6
Solteiro	8,67	3,05	5,11	5,34	2,26	26,8	7,5	10,3	10	4,1
União	0,06	0,93	1,34	1,78	6,99	0,1	1,4	1,6	2	7,7
Viuvo	1,05	21,44	3,14	9,46	0,13	2	32,5	3,9	11	0,1
Total	16,21	25,66	19,59	35,25	16,26	44,5	41,8	31,6	54,1	21,7

Tabela 9: Contribuições das categorias para os 5 primeiros eixos da matriz indicadora das médias.

Como pode ser lido na Tabela 9, e visto na Figura 6 que as categorias que contribuí para o eixo 1 são cortadas pela linha vermelha sendo elas: em relação a escolaridade foi a Superior; para variável Raça/Cor foram as categorias Parda e Branca; em relações faixas etárias foram 30-39 e 60-69; Grupo de cidades foram o 1 e 2; e o estado civil foi solteiro.

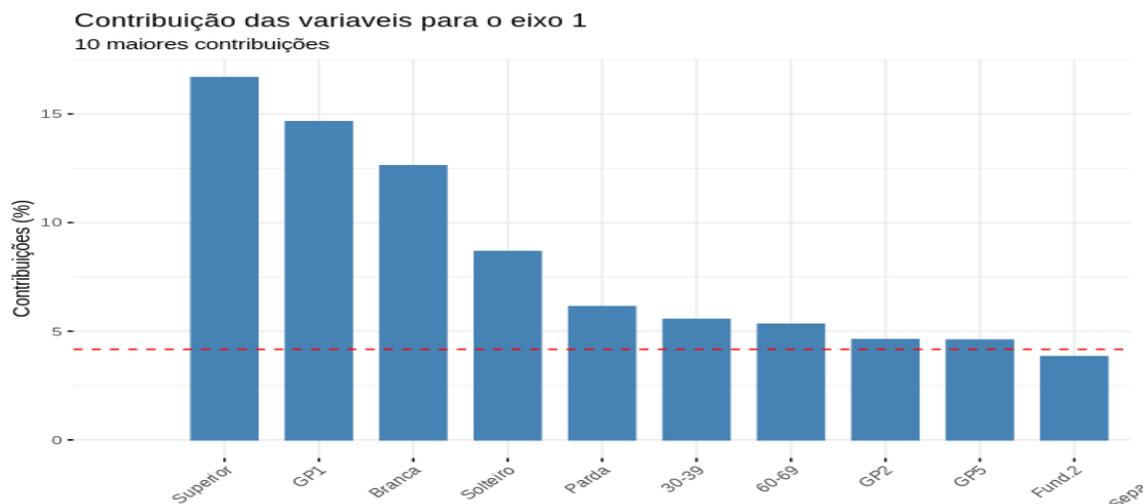


Figura 6: Contribuição no eixo 1

Na Figura 7 observou que as contribuições para o eixo 2 foram: para a variável escolaridade as categorias Nenhuma, Fundamenta 1 e Superior; para faixas de idade 30-39 e 60-69; em relação a variável Gênero o Feminino; e o grupo o 1.

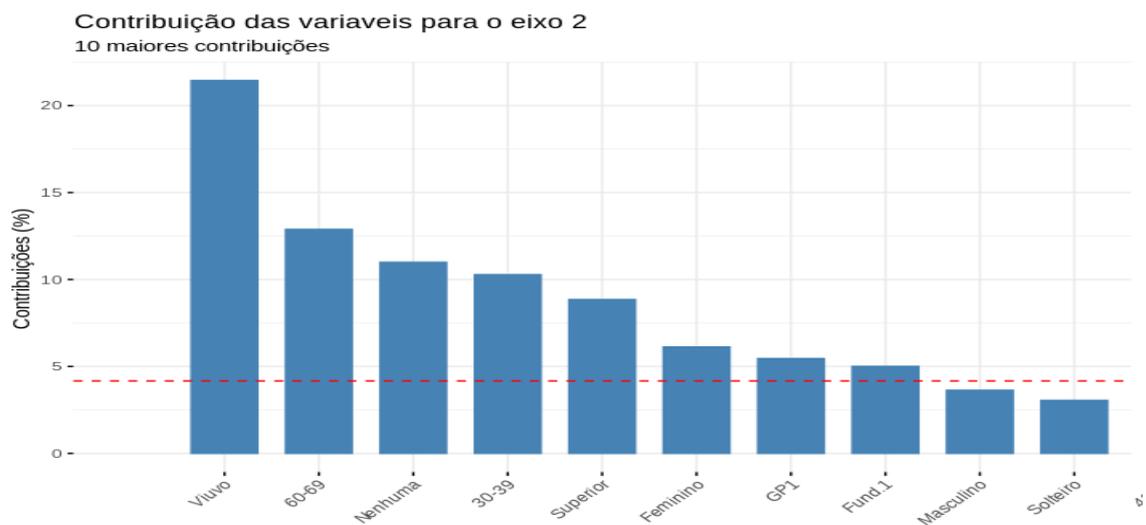


Figura 7: Contribuição no eixo 1

A linha tracejada vermelha no Gráficos 6 e 7 indica o valor médio esperado, se as contribuições fossem uniformes. O cálculo do valor esperado de contribuição.

Pode ser visto que:

- As categorias Branca, Solteiro, Parda, Grupo 2 e 5 são as mais importantes na definição da primeira dimensão.
- As categorias Viúvo, Nenhuma, Feminino e ensino Fundamental 1 contribuem mais para a dimensão 2

Na Figura 8, os pontos representam as variáveis ativas e as variáveis suplementares, na cores vermelha e verde respectivamente. As variáveis raça/cor, Escolaridade e grupo de cidades contribui mais para o eixo 1 com 19,69%, 16,67% e 25,83% respectivamente como relata a Tabela 9. Assim o eixo 2 está relacionado com as variáveis sexo, faixa de idade e estado civil, com 9,75%, 26,17% e 25,66% nessa ordem. Ano e causas de mortalidade são as variáveis suplementar.

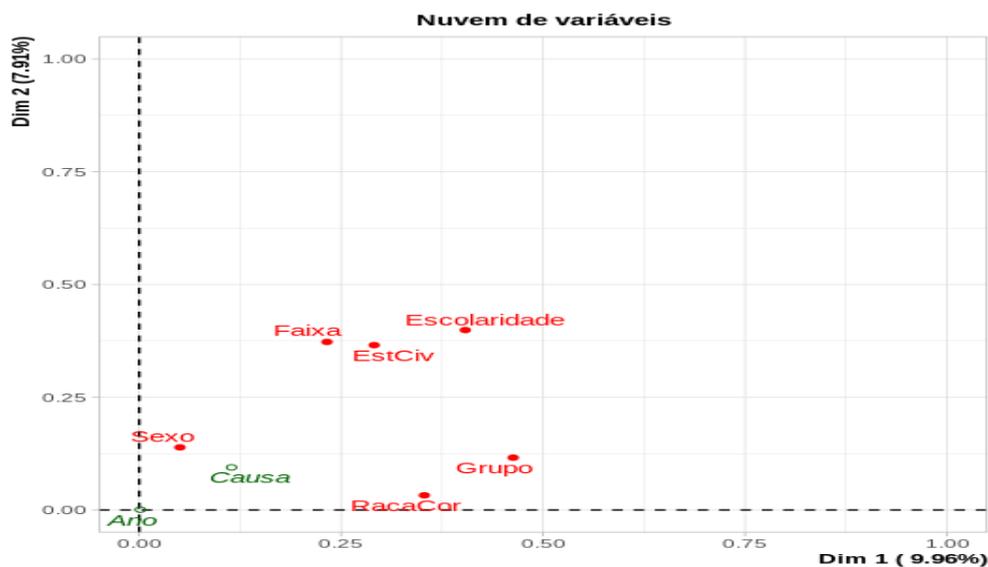


Figura 8: Nuvem de variáveis (Dim 1-2)

Nos Gráficos 9, 10, 11 são apresentados os gráficos de correspondência das categorias para as duas primeiras dimensões. Tendo em consideração tais gráficos, o 10 apresenta ambas as variáveis (ativas e suplementares), sendo as cores vermelho e verde representando as variáveis ativas e suplementares, respectivamente. Ademais, os gráficos 10 e 11 são sobre as variáveis ativas e suplementares separadamente.

Na Figura 9 observando a Tabela 9 demonstra que indicam que quando mais a direita uma características demográfica terá mais causas de mortalidade. Ainda é possível observar com ajuda da Tabela 9 que o eixo 1, as categorias: branca e parda da variável raça cor.

O gráfico de variáveis ativas possui 6 variáveis, que mostram os pontos das categorias sociodemográficas em relação ao eixo da variável suplementar. A escala dessas variáveis suplementares no Gráfico 11 é menor em relação ao Gráfico 10 para melhor visualização das variáveis, posto que as variáveis suplementares estão mais posicionadas entre os pontos $[-0,5 : 0,5]$ tanto do eixo 1 quanto do eixo 2. Pode-se observar que para o primeiro eixo, as categorias do ensino maiores classes, esta relacionado grupo 1 e grupo 2, e categoria feminina estão do lado positivo (lado a direita), enquanto a escolaridade inferior está relacionada aos homens e aos grupos com menores renda e estão do lado

negativo (lado a esquerda) do eixo. Portanto, pode-se concluir que o ensino superior está

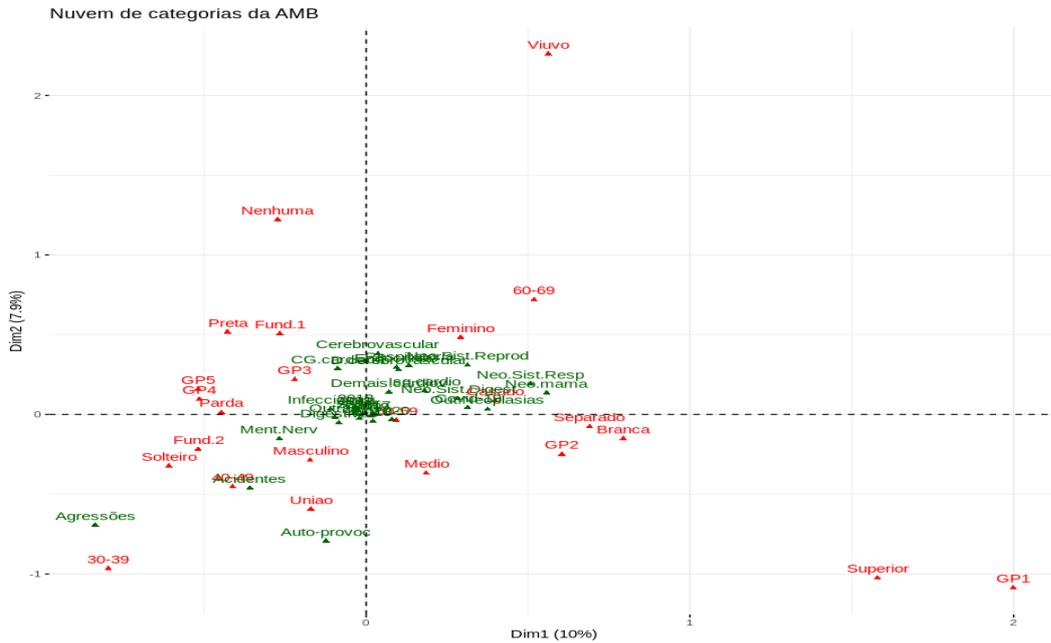


Figura 9: Nuvem de categorias (Dim 1-2)-ACM

relacionado às mulheres, enquanto o nível de escolaridade inferior está relacionado aos homens. Para o segundo eixo, as variáveis sem escolaridade, feminino e faixa etária da terceira idade (60-69), pretos e pardos encontra-se do lado positivo (lado superior), opondo-se ao ensino superior, faixa etária inferior e masculino, encontrado no eixo 2 do lado negativo (lado inferior).

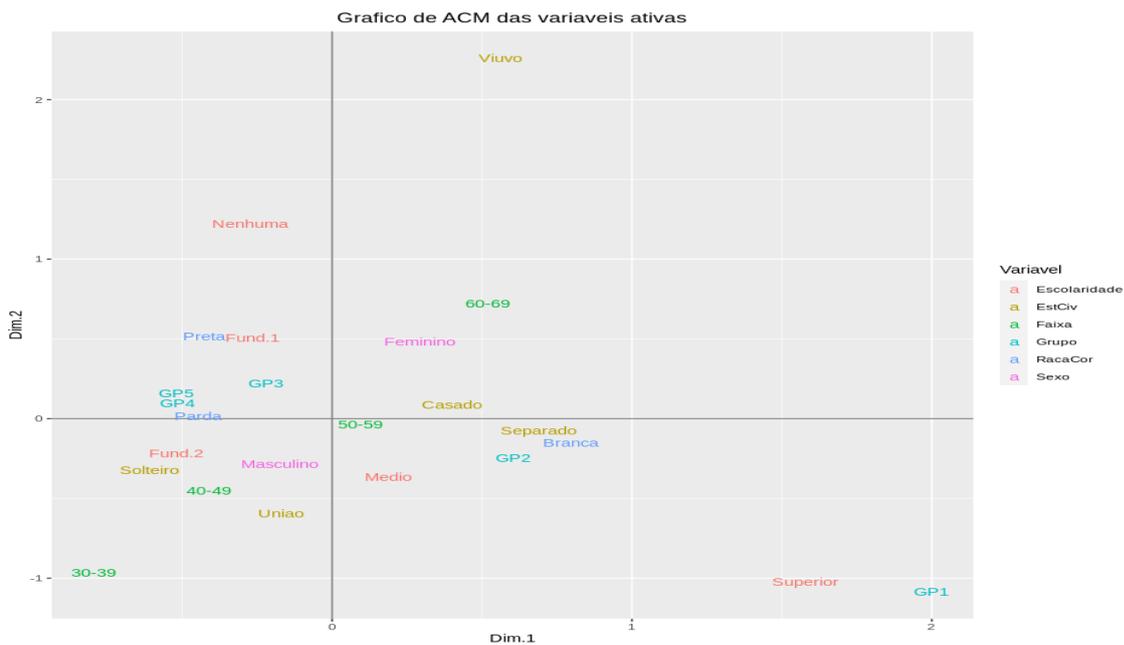


Figura 10: Nuvem de categorias ativas separadas (Dim 1-2) -ACM

Na Figura 11, pode-se observar que o eixo 1 se destaca pelo lado positivo pela presença dos capítulos da CID-10 sendo eles IV, XI e maior parte do capítulo X, que são as neoplasias, doenças endócrinas, doenças respiratória e boa parte das doenças circulatória respectivamente, além da causa básica relacionada à covid-19.

Sobre a categoria Ano, encontram-se ainda do lado direito os anos variando de 2017 a 2020; entende-se que estas doenças estão mais relacionadas a estes anos. Do lado esquerdo do eixo 1, encontram-se os capítulos XX (separado em auto-provocada, acidentes e agressões), VI-V (causas de óbitos por doenças mentais e do sistema nervoso), capítulo II (causas infectuosas exceto a covid-19), capítulo XI (que são as causas de doenças no sistema digestivo), capítulo (causas circulatória causas *garbage*) e mostra também as outras caracterizada pelos capítulos faltantes; em relação ao ano, no eixo esquerdo encontra-se o intervalo de 2011 a 2016.

Em relação ao eixo 2, na parte superior encontra-se as causas do capítulos II, IV, XI e todo o capítulo X, assim como os anos de 2011 a 2015, além de 2017; na parte inferior do eixo, encontram-se os capítulos V, VI, XI, XX e demais capítulos, além dos anos 2016, 2018, 2019 e 2020.

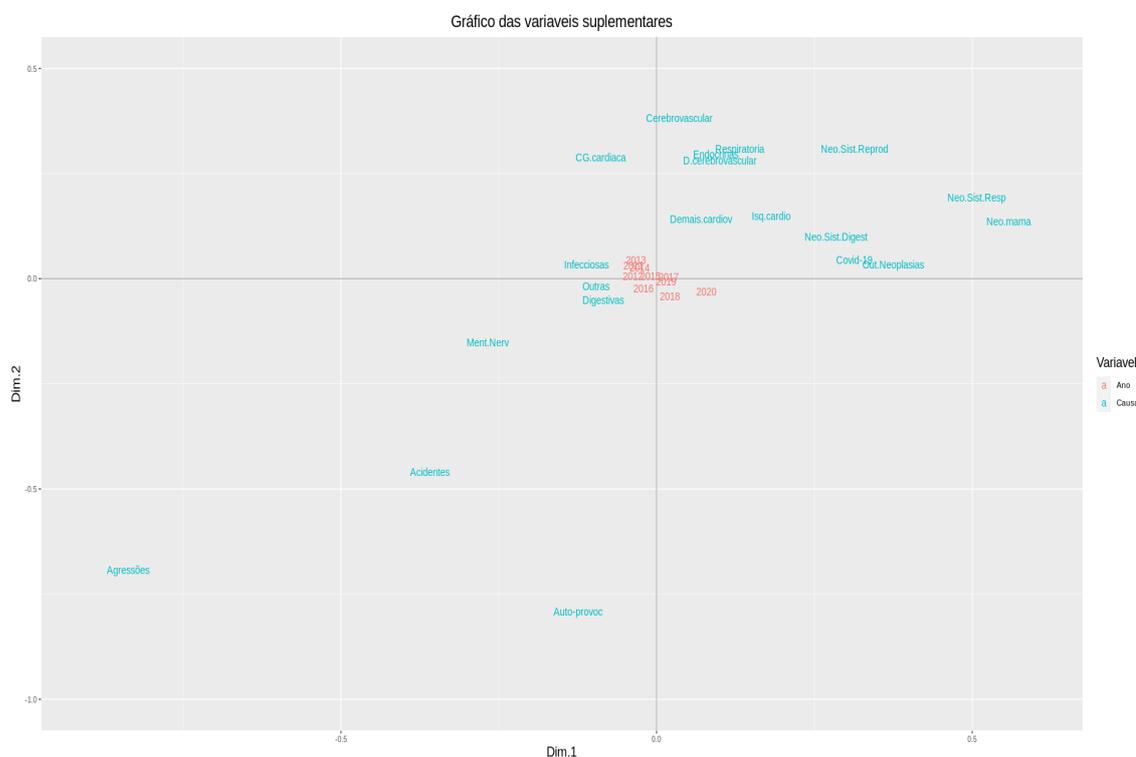


Figura 11: Nuvem de categorias suplementares separadas (Dim 1-2) -ACM

Com base nos gráficos 9, 10 e 11, pode-se identificar que a causas acidentes, agressões e auto-provocadas estão ligadas ao sexo masculino, em ambos os eixos. Por outro lado, as neoplasias apresentam-se mais ligadas ao sexo feminino. Continuando no

eixo 1, percebe-se que as causas externas de morte aparecem mais relacionadas às classes mais baixas e à baixa escolaridade. Mostra que homens de baixa renda já que os grupos de cidades varia pela renda, com pouca escolaridade e sendo eles pardo ou preto esta suscetivos a morrer por causas externas em idades mais baixas.

A contribuição relativa para as variáveis suplementares ficaram abaixo de 0,01 enquanto que a soma de qualidade nos eixos 1 e 2 tem os maiores valores em agressões (0,102), acidentes (0,027). as neoplasias, as neoplasias do sistema respiratório tem 0,013 de contribuição relativa, as causas de mortalidade auto-provocada 0,011 e as demais categorias suplementares tiveram contribuição relativa menor que 0,01.

Pode-se notar que nos perfis da sociedade, que existe uma nítida divisão de causas de mortalidade que poderiam ser evitadas, como o exemplo de causas externas de mortalidade, doenças infecciosas, e mal definidas estão mais ligados aos óbitos que ocorreram em uma população pobre, de baixa escolaridade, parda, negra e jovem, informando o quão vulnerável é essa parcela da população. E assim dizer que pode existir uma distribuição das causas de morte é diferente em distintos perfis. Em quando que no lado direito mostra óbitos mais tardios, com maiores níveis de escolaridade, branca, com maior poder aquisitivo tem causas definidas, indicando que não estão tão vulneráveis a causas que podem ser evitadas.

No Figura 12 pode-se observar a contribuição dos indivíduos em cada variável e sua distribuição por categorias. Em relação à variável Sexo, pode-se notar que a divisão dos indivíduos está mais visível na dimensão 2, onde indivíduos masculinos se encontram nitidamente divididos na parte debaixo desse eixo e os indivíduos do sexo feminino do lado de cima. Assim pode-se dizer que o Gráfico 10 tem relação com o ponto centroide de cada categoria da variável de acordo com a dispersão dos indivíduos com tais características na nuvem. Observa-se também uma dispersão do lado direito para esquerdo, em relação à variável Raça/ Cor, onde nota-se que a cores preta e parda estão mais concentradas do lado esquerdo e ambas chegam ao lado direito, porém sem grandes concentração, com a cor branca sobressaindo do lado direito. Pode-se observar que o indivíduos com menores escolaridade estão na parte esquerda superior, e com maior escolaridade estão na parte direita inferior. Observa-se também que os indivíduos menores faixas de idade (30 a 39 e 40 a 49) estão mais concentrados na parte esquerda inferior do gráfico de dispersão e já os indivíduos com idades maiores (50 a 59 e 60 a 69) estão no lado direito superior.

Em relação aos grupos por regiões administrativas, os grupos 3, 4 e 5 estão os grupos 5, 4 e 3 estão do lado esquerdo, ao mesmo tempo em que os indivíduos dos grupos 1 e 2 estão do lado direito do gráfico de dispersão.

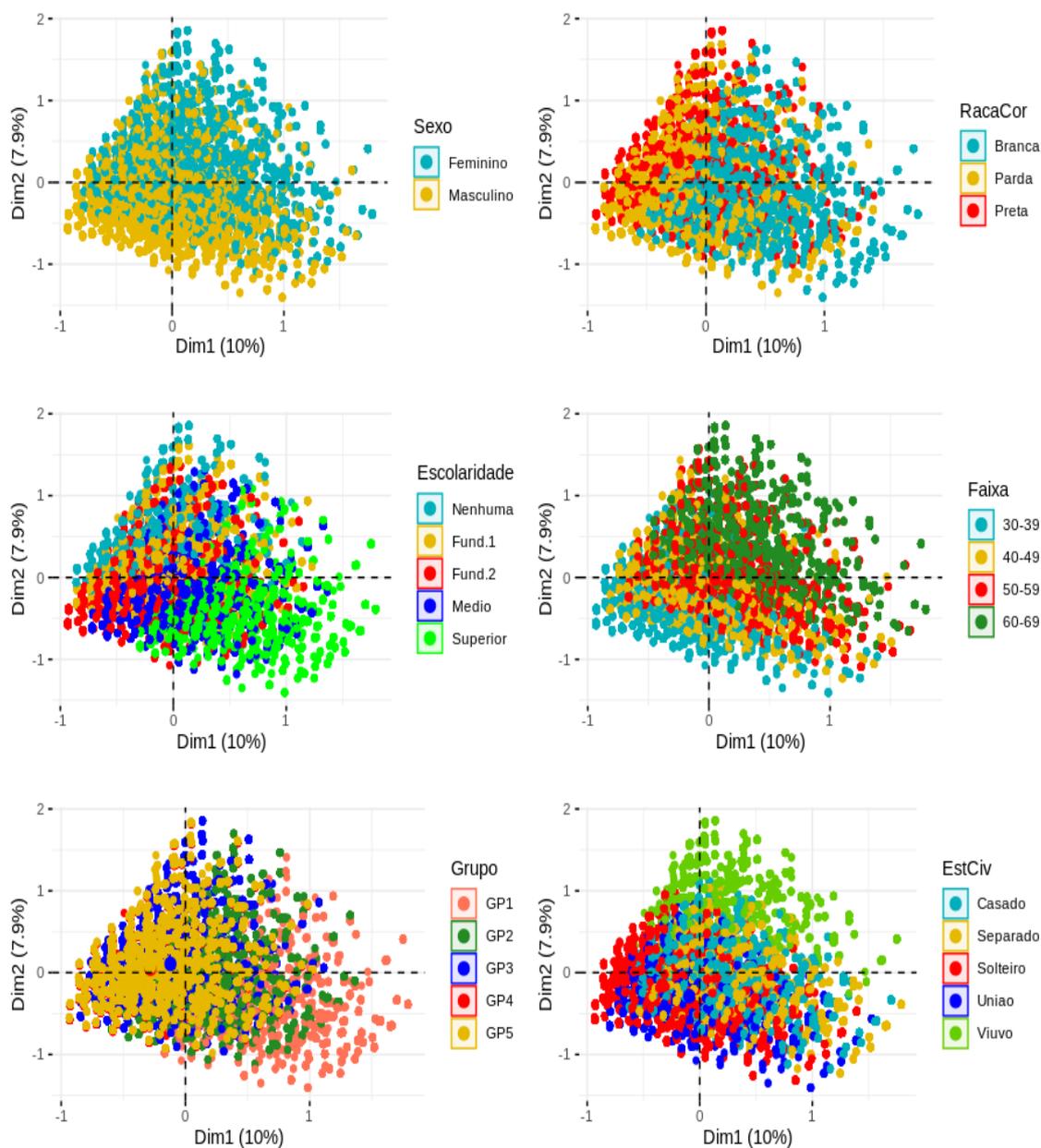


Figura 12: Contribuição dos indivíduos por variável

Já na variável Estado civil, os indivíduos mais concentrados no lado inferior esquerdo são as pessoas solteiras e em união estável; do lado direito, encontram-se os indivíduos casados e separados, sendo o mais afastados os viúvos, concentrados no lado esquerdo superior.

6 Considerações Finais

Neste trabalho, foi apresentado um estudo sobre determinantes sociais da saúde, por meio de variáveis DSS para a explicação das causas de mortes adultas. Usando a Análise de Correspondências como a técnica mais conveniente para o estudo, posto que permite observar as associações vinculadas entre as variáveis.

E assim, em relação a explicação dos eixos, os valores ficaram pequenos, uma alternativa seria agrupar algumas categorias das variáveis ativas, ou adição de mais variáveis relativa aos determinantes sociais de saúde. Foi identificado que o eixo 1 corresponde a questão socioeconômica e o eixo 2 as características do indivíduo.

“Os homens se tornam vítimas de homicídio mais frequentemente do que as mulheres” (Barata, 2009), assim a fundamentação teórica tem como base o pilar que se sustenta a seguinte constatação: as causas principais de óbitos relacionadas ao gênero masculino estarem correlacionada com causas e eventos externos, fato que foi possível notar no eixo 1, marcado também pela diferença entre as escolaridades baixas associadas a baixa renda pelo lado esquerdo desse eixo, além de demonstrar que essa população negra, parda, jovem são também correlacionadas ao lado esquerdo do eixo 1. Enquanto que o lado direito desse eixo 1 mostra uma população com alto poder aquisitivo e suas mortes estão ligadas a causas neoplasias e doenças do coração definidas, e com idades mais tardias.

A análise de mortalidade é usada para avaliar e monitorar serviços de saúde no mundo. Estudar da mortalidade é importante para determinar os tipos de óbitos que ocorrem em determinados perfis sociais e destina verbas e recursos para políticas públicas de enfrentamento às principais doenças que atingem as classes da população, uma vez que os serviços de saúde precisam de dados fidedignos que apoiem as tomadas de decisões. O Brasil conta com um banco completo sobre dados de mortalidade que permite que esses estudos sejam feitos de forma fidedigna. Contudo, apesar de sua importância, existem poucos estudos em mortalidade para diferentes perfis.

Assim, um estudo que busque identificar determinantes sociais de saúde se torna crucial para o entendimento e uso mais eficaz do Sistema de Informação de Mortalidade (SIM), como banco de dados com diversas variáveis sociais que permite o estudo das mesmas.

7 Referências

BARATA, RB. Como e por que as desigualdades sociais fazem mal à saúde [online]. Rio de Janeiro: Editora FIOCRUZ, 2009. Temas em Saúde collection. 120 p. ISBN 978-85-7541-391-3. Available from SciELO Books <http://books.scielo.or>.

BATISTA, L. S. et al. A cor da morte: características de raça no Estado de São Paulo, 1999 a 2001. **Rev Saúde Pública**, São Paulo, p. 630-636, 19 mar. 2004. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102004000500003&lng=pt&tlng=pt. Acesso em: 17 mar. 2021.

BRASIL. Ministério da Saúde. Indicadores e Dados Básicos;[2000]. Disponível em:<http://tabnet.datasus.gov.br/cgi/idb2000/fqa09.htm>. Acesso em: 27 de fev. 2021.

BRASIL. Ministério da Saúde. Política Nacional de Saúde Integral da População Negra. Brasília, fev. 2007.

BRASÍLIA. Plano Distrital de Saúde da SES/DF- 2016/2019: Resolução CSDF Nº 457, abr. 2016.

BUSS, P. M; PELLEGRINI FILHO, A. A Saúde e seus Determinantes Sociais. **PHYSIS: Rev. Saúde Coletiva** Rio de Janeiro, 17(1), p. 77-93, 2007.

CHOR, D.; LIMA, C. R. A. Aspectos epidemiológicos das desigualdades raciais em saúde no Brasil. **Cad. Saúde Pública**, Rio de Janeiro, p. 1586-1594, set/out 2005. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2005000500033. Acesso em: 18 mar. 2021.

CYRINO, Carolina Dutra. Análise Geométrica de Dados através de Análise de Correspondência Múltipla. 2011. Dissertação (Graduação em Estatística) - Universidade Federal de Juiz de Fora, Juiz de Fora, 2011.

COMPANHIA DE PLANEJAMENTO DO DISTRITO FEDERAL. Delimitação do Espaço Metropolitano de Brasília (Área Metropolitana de Brasília). Nota técnica n o 1/2014, p.24. Brasília, 2014.

Dahlgren G, Whitehead M. Policies and Strategies to promote social equity in health. Stoclm: Institute for Future Studies; 1991.

DE SOUZA, Damião Ernane. Determinação Social da Saúde: Associação entre sexo, escolaridade e saúde autorreferida. Orientador: Naomar de Almeida Filho. 2012. 1 f. Tese (Doutor em Saúde Pública) - Universidade Federal da Bahia, [S. l.], 2012.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de Análise de dados: Estatística e Modelagem Multivariada com Excel® , SPSS® e Stata®**. 1. ed. Rio de Janeiro - RJ: ELSEVIER, 2017. ISBN 978-85-352-7087-7.

GREENACRE, Michael. **Correspondence Analysis in Practice**. 3. ed. [S. l.]: CRC Press, 2017. 327 p. ISBN 9781498731775.

GREENACRE, Michael. **Multiple Correspondence Analysis and Related Methods**. ed. [S. l.]: Chapman & Hall/CRC, 2006. 607 p. ISBN 1584886285.

HUSSON, François et al. **Exploratory Multivariate Analysis by Example Using R**. 2ª. ed. Londres: Chapman & Hall/CRC, 2017. ISBN 978-1-1381-9634-6.

HUSSON, François et al. **FactoMineR: Multivariate Exploratory Data Analysis and Data Mining**. R package version 2.4 (2020).

HAIR, J. F. Jr. et. al. Escalonamento Multidimensional e Análise de Correspondência: Análise de correspondência. In: **ANÁLISE Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009. cap. 9, p. 506-534. ISBN 978-85-7780-534-1.

KASSAMBARA, Alboukadel; MUNDT, Fabian. **factoextra: Extract and Visualize the Results of Multivariate Data Analyses**. R package version 1.0.7 (2020).

PELLEGRINI FILHO, A.; BUSS, P. M. O Movimento dos Determinantes Sociais da Saúde no Brasil. In: GALVÃO, L. A. C. et al. Determinantes ambientais e sociais da saúde. Rio de Janeiro: FIOCRUZ, 2011. cap. 27, p. 589-601. ISBN 978-92-75- 73129-1.

PRADO, Mariele Vilela Bernardes. **Métodos de Análise de Correspondência Múltipla: Estudo de caso aplicado à avaliação da qualidade do café**. 2012. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, LAVRAS – MG, 2012.

SOARES, Darli Antônio; ANDRADE, Selma Maffei de; CAMPOS, João José Batista de. Epidemiologia e indicadores de saúde. In: ANDRADE, Selma Maffei de; SOARES, Darli

Antonio; CORDONI JUNIOR, Luiz (Org.). **Bases da saúde coletiva**. Londrina: Ed. UEL, 2001. cap.10, p. 183-210.

8 Apêndice

```
#####
##### MCA #####
#####

rm(list=ls(all=TRUE));library(read.dbc);library(lubridate);library(tidyverse);
library(readxl);library(writexl);library(foreign)
library(dplyr);library("FactoMineR");library("factoextra");library("corrplot");
library(gridExtra)library("graphics");library("amap");library(rgl);library(MASS);
library(ca);library("amap")

dados<-read_xlsx("/home/usuario/Documentos/UnB/1-2021/TCC/Dados/AMB/amb_sem_NA.xlsx")

sapply(dados, function(x) sum(is.na(x)))/nrow(dados)

dados<-dados[,c("capitulao", "Ano", "RacaCor", "Sexo", "Escolaridade", "Faixa", "Grupo",
"EstCiv")]

colnames(dados)<-c("Causa", "Ano", "RacaCor", "Sexo",
"Escolaridade", "Faixa", "Grupo", "EstCiv")

dados<-dados%>%filter(!is.na(RacaCor)& !is.na(Sexo) & !is.na(Escolaridade) &
!is.na(Gruo) & !is.na(EstCiv))
dados<-na.omit(dados)
sapply(dados, function(x) sum(is.na(x)))/nrow(dados)

dados$Escolaridade<-factor(dados$Escolaridade,
      levels = c( "Nenhuma" , "Fund.1" , "Fund.2" , "Medio" , "Superior"))

levels(as.factor(dados$Escolaridade))
```

```
###numera a quantidade de categorias de variaveis no cando de dados

cats = apply(dados[,3:8], 2, function(x) nlevels(as.factor(x)))

cats

sum(cats)

### valor de K

length(cats)

# valor de (K/Q)-1

sum(cats)/length(cats) -1

# Dimensionalidade K-Q

sum(cats)-length(cats)

####valor de lambda

((sum(cats)/length(cats)-1)/(sum(cats)-length(cats)))

### formula do FACTOMINER
library(FactoMineR)
# Aplicar ao mca
mca1 = MCA(dados,quali.sup =c(1,2),graph = F)

##### gráficos individuais

raca<-fviz_mca_ind(mca1,
  label = "none", # hide individual labels
  habillage = "RacaCor", # color by groups
  palette = c("#00AFBB", "#E7B800","red", "blue","green"),
```

```
addEllipses = TRUE, ellipse.type = "confidence",
ggtheme = theme_minimal()) + labs(title = "")
```

```
sexo<-fviz_mca_ind(mca1,
  label = "none", # hide individual labels
  habillage = "Sexo", # color by groups
  palette = c("#00AFBB", "#E7B800","red", "blue","green"),
  addEllipses = TRUE, ellipse.type = "confidence",
  ggtheme = theme_minimal()) + labs(title = "")
```

```
esc<-fviz_mca_ind(mca1,
  label = "none", # hide individual labels
  habillage = "Escolaridade", # color by groups
  palette = c("#00AFBB", "#E7B800","red", "blue","green"),
  addEllipses = TRUE, ellipse.type = "confidence",
  ggtheme = theme_minimal()) + labs(title = " ")
```

```
faixa<-fviz_mca_ind(mca1,
  label = "none", # hide individual labels
  habillage = "Faixa", # color by groups
  palette = c("#00AFBB", "#E7B800","red", "#228B22","green"),
  addEllipses = TRUE, ellipse.type = "confidence",
  ggtheme = theme_minimal()) + labs(title = " ")
```

```
grup<-fviz_mca_ind(mca1,
  label = "none", # hide individual labels
  habillage = "Grupo", # color by groups
  palette = c("#FF7256", "#228B22", "blue","red","#00AFBB"),
  addEllipses = TRUE, ellipse.type = "confidence",
  ggtheme = theme_minimal()) + labs(title = " ");grup
```

```
estciv<-fviz_mca_ind(mca1,
  label = "none", # hide individual labels
  habillage = "EstCiv", # color by groups
  palette = c("#00AFBB", "#E7B800","red", "blue","#66CD00"),
```

```
addEllipses = TRUE, ellipse.type = "confidence",
ggtheme = theme_minimal()) + labs(title = "")

grid.arrange(sexo,raca, esc,faixa,grup,estciv)

mca1$var$coord
#Um método alternativo para determinar o número de dimensões é olhar para
#um gráfico de Scree Plot,
# que é o gráfico de autovalor / variância ordenado do maior para o menor.

fviz_screepLOT(mca1, addlabels = TRUE, ylim = c(0, 11))+
  geom_hline(yintercept=5.52, linetype=2, color="red")+
  labs(title = "Gráfico de sedimentação", x = "Dimensões",
        y = "Porcentagem de variabilidade explicada")

#### Mapa de fator de variáveis usando o gráfico de base FactoMineR :

plot(mca1, axes= c(1,3),choix = "var")+labs(title = "Nuvem de variáveis")

cos2_atvas<-data.frame(mca1$var$cos2)
rowSums(cos2_atvas)

## Também é possível fazer o mapa de fatores de variáveis usando factoextra :

fviz_mca_var(mca1)+labs(title = "Nuvem de categorias da AMB")

# tabela de autovalores
autovalores<-data.frame(mca1$eig)
colnames(autovalores)<-c("Inercia", "%", "% acumulada")
head(autovalores)

#write_xlsx(autovalores, "/home/usuario/Documentos/UnB/1-2021/TCC/Dados/AMB/auto-in
```

```
# coluna das coordenadas
cor_col<-data.frame(mca1$var$coord)
head(cor_col)

#linha de coordenadas
cor_row<-data.frame(mca1$ind$coord)
head(cor_row)

# Usando o pacote FactoMineR para fazer a análise de correspondência
summary(mca1)

sqrt(mca1$var$cos2)

#plotellipses(mca1, cex=0.2, magnify=12, keepvar=2:7)

# extrair e visualize os resultados das variáveis
var<-get_mca_var(mca1)
var

var$contrib

##
#fviz_mca_var(mca1,axes = c(1,3))+
#labs(title = "Nuvem de categorias da AMB \n Eixo 1 e 3")

# Descrição das duas dimensões
dimdesc(mca1)

ind<-get_mca(mca1,element = c("var","ind","mca.cor","quanti.sup"))

mca1$quali.sup

corde_sup<-data.frame(mca1$quali.sup$coord)
cos2_sup<-data.frame(mca1$quali.sup$cos2)
```

```
eta_sup<-data.frame(mca1$quali.sup$v.test)

ind$cos2
cont<-data.frame(ind$contrib)

write_xlsx(cos2_sup,
"/home/usuario/Documentos/UnB/1-2021/TCC/Dados/AMB/contibuicao sup.xlsx")

write_xlsx(data.frame(levels(as.factor(dados$Causa))),
"/home/usuario/Documentos/UnB/1-2021/TCC/Dados/AMB/varivale causas.xlsx")

#res.desc <- dimdesc(mca1, axes = c(1,2))
# Description of dimension 1
#res.desc[[1]]
# Description of dimension 2
#res.desc[[2]]

### A proporção das variações retidas pelas diferentes dimensões (eixos) pode
#ser extraída usando a função get_eigenvalue() [ pacote factoextra ] como segue:
eig.val <- get_eigenvalue(mca1)
head(eig.val)

table(dados$EstCiv)

mca1_vars_df = data.frame(mca1$var$coord, Variable = rep(names(cats), cats))
mca1_obs_df = data.frame(mca1$ind$coord)
head(mca1_obs_df)

##### GRAFICO PARA AS VARIÁVEIS SUPLEMENTAR

cats2 = apply(dados[,1:2], 2, function(x) nlevels(as.factor(x)))

cats2

mca1_sup<-data.frame(mca1$quali.sup$coord,Variable=rep(names(cats2), cats2))
```

```
# grafico das variaveis categoricas
ggplot(data = mca1_sup, aes(x = Dim.1, y = Dim.2, label = rownames(mca1_sup))) +
  geom_hline(yintercept = 0, colour = "gray70") + geom_vline(xintercept = 0,
colour = "gray70") + geom_text(aes(colour = Variable)) +
  ggtitle("Gráfico das variaveis suplementares - Eixo 1 e 2")+
  ylim(-0.5,0.5)+xlim(-0.5,0.6)+
  theme(axis.text=element_text(size=8),title=element_text(size=14))+
  scale_colour_discrete(name = "Variavel") +
  theme(plot.title = element_text(hjust = 0.5))
```

```
mca1_sup_df<-data
```

```
#O código R abaixo representa categorias de variáveis qualitativas
#(variáveis ativas e suplementares):
```

```
fviz_mca_var(mca1, repel = TRUE, ggtheme= theme_minimal())
```

```
fviz_screplot(mca1, addlabels = TRUE, ylim = c(0, 10))
```

```
fviz_mca_var(mca1, choice = "mca.cor", repel = TRUE)
```

```
# contribuição das contribuições
```

```
fviz_contrib(mca1, choice = "var", axes = 2, top = 10)
```

```
## resultado para os individuos
```

```
fviz_mca_ind(mca1, col.ind = "cos2", gradient.cols=c("00AFBB", "#E7B800", "#FC4E07"), re
```

```
# descrição da dimensão
```

```
#A função dimdesc()[em FactoMineR] pode ser usada para identificar as variáveis mais
#correlacionadas com uma determinada dimensão:
```

```
#res.desc <- dimdesc(mca1, axes = c(1,2))

# Descrição da dimensão 1
#res.desc[[1]]

# Descrição da dimensão 2
#res.desc[[2]]

# grafico das variaveis categoricas
ggplot(data = mca1_vars_df, aes(x= Dim.1,y=Dim.2,label=rownames(mca1_vars_df)))+
geom_hline(yintercept = 0, colour = "gray50") + geom_vline(xintercept = 0,
colour="gray50")+geom_text(aes(colour=Variable))+ggtitle("Grafico de ACM
das variaveis ativas") + scale_colour_discrete(name = "Variavel") +
  theme(plot.title = element_text(hjust = 0.5))

ggplot(data=mca1_vars_df,aes(x=Dim.1,y=Dim.2,label=rownames(mca1_vars_df))) +
geom_hline(yintercept = 0, colour = "gray50") + geom_vline(xintercept = 0,
colour = "gray50") + geom_text(aes(colour = Variable)) + ggtitle("Grafico de ACM
das variaveis ativas Dimensão 1 e 2")+ scale_colour_discrete(name = "Variavel")
```