

Universidade Federal de Uberlândia

Instituto de Ciências Exatas e Naturais do Pontal

Curso de Matemática

Trabalho de Conclusão de Curso

**Análise de componentes principais
aplicada a dados de mortalidade por
câncer no Brasil**

por

Ana Beatriz Atique[†]

Bacharelado em Matemática - Ituiutaba - MG

Orientador: Profa. Dra. Franciella Marques da Costa

Análise de componentes principais aplicada a dados de mortalidade por câncer no Brasil

Este exemplar corresponde à redação final da Monografia devidamente corrigida e defendida por **Ana Beatriz Atique** e aprovada pela comissão julgadora.

Ituiutaba, 01 de abril de 2022.

Profa. Dra. Franciella Marques da Costa

Banca examinadora:

Profa. Dra. Franciella Marques da Costa.

Profa. Dra. Gabriella de Freitas Alves.

Profa. Dra. Kátia Gomes Facure Giaretta.

Monografia apresentada ao Instituto de Ciências Exatas e Naturais do Pontal, UFU como requisito parcial para obtenção do título de **Bacharelado em Matemática**.

A minha Orientadora
Franciella

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter sempre me guiado em minhas decisões e nunca ter faltado comigo.

Aos meus pais, Ana Lúcia e Marcelo, e irmão, Mateus, pelo incentivo, apoio e confiança durante minha jornada acadêmica longe de casa.

Agradeço a minha tia Luciana, por ter me dado abrigo e apoio durante os anos de graduação em Ituiutaba.

Agradeço a minha amiga Sara Viana, por mesmo distante sempre se fazer presente em minha vida e dia a dia.

Agradeço a República UFUracão, todos os moradores e agregados, por todos os conselhos e estarem sempre presentes para me ouvir.

Obrigada a minha orientadora Franciella Marques por ter confiado e acreditado no meu potencial para a realização deste trabalho.

A Universidade Federal de Uberlândia, ao curso de Matemática, aos docentes, diretores, administração e coordenação que sempre estiveram presentes para auxiliar e proporcionar o melhor dos ambientes para a realização deste trabalho.

A todos que direta, ou indiretamente, participaram dessa jornada, um muito obrigada.

RESUMO

Sendo uma das principais causas de morte no mundo, estudos relacionados ao câncer são de extrema importância. O objetivo deste trabalho é analisar dados de mortalidade por câncer nos 26 estados e no Distrito Federal por meio da análise de componentes principais. Utilizou-se variáveis referentes as taxas brutas de mortalidade por 15 tipos de câncer. Para realizar as análises estatísticas utilizou-se o software R juntamente com os pacotes ellipse, factoextra, FactoMineR, MVN, plotrix e psych. Inicialmente foi realizada a análise estatística descritiva, em seguida testou-se a normalidade multivariada dos dados por meio do teste de Mardia. Para verificar se os dados são correlacionados aplicou-se o teste de esfericidade de Bartlett e por fim aplicou-se as técnicas de análise de componentes principais. A análise de componentes principais para os dados de mortalidade por câncer mostrou-se eficiente, pois reduziu a quantidade inicial de 15 variáveis para quatro componentes principais que explicam aproximadamente 70% de toda variação dos dados.

Palavras-chave: Componentes principais, Estatística multivariada, Mortalidade.

LISTA DE FIGURAS

1.1	Desenvolvimento do câncer.	3
3.1	<i>Boxplot</i>	15
3.2	<i>QQ-plot</i> multivariado.	16
3.3	Elipses de correlação.	18
3.4	Gráfico de dispersão e coeficiente de correlação entre as variáveis.	19
3.5	<i>Screeplot</i>	21
3.6	Porcentagem da variância total explicada por cada componente.	21
3.7	Gráficos de barras contendo a importância das variáveis no primeiro, segundo, terceiro e quarto componente principal.	23
3.8	Gráficos da correlação entre CP1, CP2 e as variáveis.	25
3.9	Biplot.	27

LISTA DE TABELAS

3.1	Análise descritiva das variáveis.	14
3.2	Correlação entre as variáveis.	17
3.3	Componentes principais (CP), autovalores, porcentagem da variância total explicada pelo componente e a porcentagem acumulada da explicação da variância total.	20
3.4	Correlação entre os CP1, CP2, CP3, CP4 e as variáveis	24
3.5	Escores dos 4 primeiros componentes	26

CONTEÚDO

Agradecimentos	i
Resumo	ii
Lista de Figuras	iii
Lista de Tabelas	iv
Introdução	1
Objetivo	2
Estrutura dos Tópicos Apresentados	2
1 Referencial teórico	3
1.1 Câncer	3
1.2 Teste de esfericidade de Bartlett	5
1.3 Componentes Principais	6
2 Materiais e Métodos	11
3 Resultados e discussão	13
4 Conclusão	28
Referências	29
Anexo	32

INTRODUÇÃO

Estudos relacionados ao câncer são de extrema importância. Segundo a Organização Pan-Americana da Saúde (OPAS), em 2008 foram notificadas 1,2 milhão de mortes por câncer na América. No mesmo ano, foram registradas 9,6 milhões de mortes por câncer no mundo (OPAS, 2020). Sendo uma das principais causas de mortes no mundo. No Brasil, o Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) registrou aproximadamente 232 mil mortes no ano de 2019 (INCA, 2021).

A estatística dispõe de diversos métodos utilizados para a análise e interpretação de dados em diferentes áreas, inclusive áreas médicas. Dentre esses métodos está a estatística multivariada que compreende um conjunto de técnicas aplicadas para analisar de forma simultânea diversas variáveis em uma mesma unidade amostral. A análise de componentes principais é uma ferramenta da estatística multivariada que visa simplificar a estrutura do conjunto de dados.

A análise de componentes principais foi utilizada por Martins (2018) no estudo da morbidade hospitalar do SUS, Machado (2004) realizou uma análise das causas múltiplas de mortalidade infantil no estado de São Paulo no ano de 1999, já Santos (2014) utilizou a técnica para analisar fatores associados à mortalidade dos idosos com as condições de vida da população do nordeste do Brasil.

Os estudos estatísticos acerca da mortalidade são de grande importância em demografia e na área de saúde pois permitem compreender o impacto na vida e saúde da população contribuindo com as formulações de políticas públicas de prevenção e tratamento da doença.

Objetivo

Este trabalho teve como objetivo analisar dados de mortalidade por câncer nos 26 estados e no Distrito Federal por meio da análise de componentes principais.

Estrutura dos tópicos apresentados

O trabalho está organizado da seguinte maneira:

- Na **introdução** foi fornecido informações importantes em relação ao câncer e o objetivo do trabalho.
- No **referencial teórico** será discutido a teoria referente ao tópico análise de componentes principais.
- No **material e métodos** descreve-se o conjunto de dados, enumera-se os passos para a análise e apresenta-se o *software* utilizado nas análises estatísticas.
- Nos **resultados e discussão** será apresentado e discutido as análises estatísticas realizadas.
- **Conclusão.**
- **Referências.**
- No **anexo** encontra-se a rotina do *software* R utilizada no trabalho.

CAPÍTULO 1

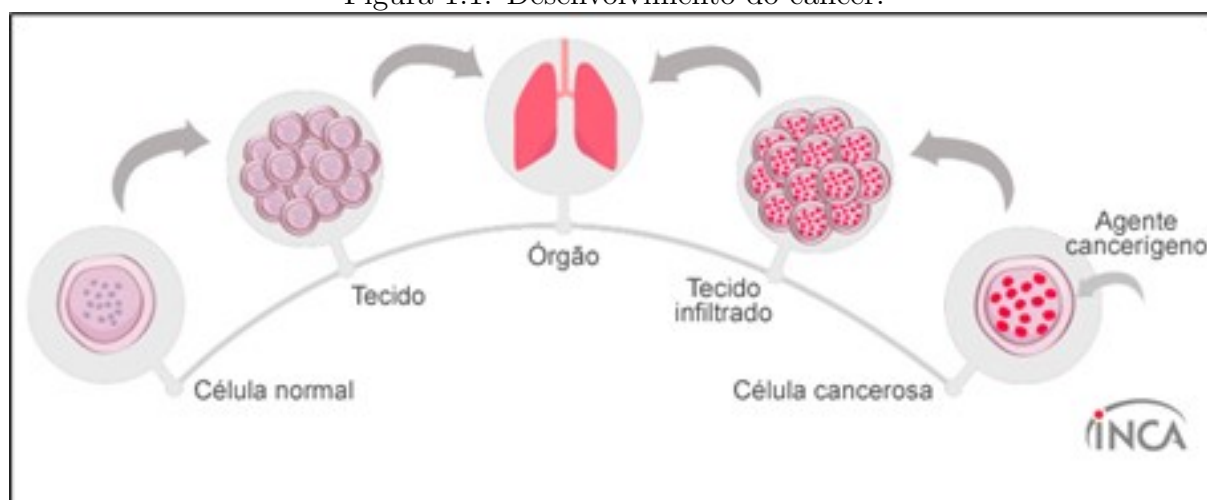
REFERENCIAL TEÓRICO

1.1 Câncer

Câncer é o nome genérico para diferentes tipos de doenças malignas, que têm em comum o crescimento descontrolado e anormal das células, podendo atingir tecidos próximos ou órgãos a distância (INCA, 2020).

A Figura 1.1 mostra o desenvolvimento do câncer.

Figura 1.1: Desenvolvimento do câncer.



Fonte: INCA, 2020.

A Oncologia é a especialidade médica responsável por tratar o câncer e oncologista é o profissional que trata vários tipos de tumores. O câncer é uma doença de alta complexidade, sendo necessário o acompanhamento de uma equipe multidisciplinar, coordenada pela Oncologia, uma vez que podem surgir outras doenças durante o tratamento e a medicação utilizada pode provocar diversos efeitos colaterais. O trabalho da equipe multidisciplinar é essencial para cuidar da qualidade de vida dos pacientes (ICB, 2018).

Segundo o INCA (2020), o câncer pode ser causado por fatores externos (presentes no meio ambiente) e internos (como hormônios, condições imunológicas e mutações genéticas). Estes fatores podem interagir de várias maneiras, tendo como consequência o surgimento do câncer. Sendo que os fatores externos são responsáveis por 80% a 90% dos casos de câncer.

O INCA (2020) fornece uma lista com 12 dicas para prevenir o câncer, são elas:

- Não fumar;
- Manter uma alimentação saudável;
- Manter o peso corporal adequado;
- Praticar atividades físicas;
- Amamentar;
- Fazer o exame preventivo do câncer do colo do útero a cada três anos (mulheres com idade entre 25 e 64 anos);
- Vacinar contra o HPV os meninos com idade entre 11 e 14 anos e as meninas de 9 a 14 anos;
- Vacinar contra a hepatite B;
- Evitar o consumo de álcool;
- Evitar ingerir carne processada;
- Evitar a exposição ao sol no período de 10h a 16h;
- Evitar a exposição a agentes cancerígenos no trabalho.

De acordo com o INCA (2020), o paciente com câncer possui alguns direitos especiais garantidos na legislação, tais como auxílio-doença, tratamento fora do domicílio e saque do Fundo de Garantia por Tempo de Serviço (FGTS).

1.2 Teste de esfericidade de Bartlett

A obtenção das componentes principais não exige normalidade multivariada. Mas é necessário que as variáveis sejam correlacionadas. Se não houver correlação entre as variáveis, ou seja, a matriz de correlação for uma matriz identidade, a análise de componentes principais vai obter, em alguma ordem, as variáveis originais. Nessa situação não é apropriado utilizar a análise de componentes principais (MINGOTI, 2020).

Segundo Mingoti (2020), quando as variáveis seguem uma distribuição normal multivariada, pode-se utilizar o teste de esfericidade de Bartlett para verificar se a matriz de correlação é uma matriz identidade.

O teste de esfericidade de Bartlett descrito a seguir está de acordo com o exposto em Lattin, Carroll e Green (2011).

As hipóteses do teste são:

$$H_0 : P_{pp} = I_{pp}$$

$$H_1 : P_{pp} \neq I_{pp}$$

de forma que P_{pp} é a matriz de correlação populacional e I_{pp} é a matriz identidade. A estatística de teste é apresentada a seguir:

$$\chi^2 = - \left[(n-1) - \frac{(2p+5)}{6} \right] \ln |R|$$

em que $\ln |R|$ é o logaritmo natural do determinante da matriz de correlação, p é o número de variáveis e n é o tamanho amostral. De acordo com Ferreira (2008), a estatística de teste, sob H_0 verdadeira, tem distribuição assintótica qui-quadrado com $\frac{p^2-p}{2}$ graus de liberdade.

A matriz de correlação amostral R_{pp} é definida por,

$$R_{pp} = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1p} \\ R_{21} & R_{22} & \dots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p1} & R_{p2} & \dots & R_{pp} \end{bmatrix}$$

em que o coeficiente de correlação amostral de Pearson entre a i -ésima e a j -ésima variáveis é dado por,

$$R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$$

sendo a variância amostral da i -ésima variável definida por

$$S_{ij} = S_{ii} = \frac{\sum_{l=1}^n (X_{il} - \bar{X}_i)^2}{n-1}, \text{ para } i = j,$$

e a covariância amostral entre a i -ésima e j -ésima variáveis dada por

$$S_{ij} = \frac{\sum_{l=1}^n (X_{il} - \bar{X}_i)(X_{jl} - \bar{X}_j)}{n-1}, \text{ para } i \neq j.$$

Também pode-se calcular o coeficiente de correlação utilizando a fórmula disponibilizada em Triola (2017),

$$R_{ij} = \frac{n(\sum X_i X_j) - (\sum X_i)(\sum X_j)}{\sqrt{n(\sum X_i^2) - (\sum X_i)^2} \sqrt{n(\sum X_j^2) - (\sum X_j)^2}}$$

Existem diversos testes que podem ser utilizados na análise de normalidade multivariada, dentre eles os testes de hipótese propostos por Mardia, que são baseados nos coeficientes de assimetria e curtose da distribuição normal multivariada (MINGOTI, 2020).

1.3 Componentes Principais

A análise de componentes principais é uma técnica da estatística multivariada que visa simplificar a estrutura do conjunto de dados em estudo. Seu objetivo principal é o de explicar a estrutura de variância e covariância de um vetor aleatório, composto de p variáveis aleatórias, através da construção de combinações lineares das variáveis originais (MINGOTI, 2020).

A teoria a seguir está de acordo com o exposto em Mingoti (2020). Considere uma amostra aleatória de n elementos, em que, para cada elemento da amostra, foram observadas p variáveis de interesse. Dessa forma, a matriz de dados X de ordem $n \times p$ é dada por:

$$X_{n \times p} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix}$$

onde o primeiro índice é referente à variável e o segundo índice ao elemento amostral.

A matriz de covariâncias amostrais é definida por:

$$S_{pxp} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix}$$

em que a variância amostral da i -ésima variável é dada por

$$S_{ij} = S_{ii} = \frac{1}{n-1} \sum_{l=1}^n (X_{il} - \bar{X}_i)^2, \text{ para } i = j$$

e a covariância amostral da i -ésima e j -ésima variáveis é definida por

$$S_{ij} = \frac{1}{n-1} \sum_{l=1}^n (X_{il} - \bar{X}_i)(X_{jl} - \bar{X}_j), \text{ para } i \neq j$$

Os autovalores $\hat{\lambda}_i$, onde $i = 1, 2, \dots, p$, da matriz S_{pxp} , são obtidos através da equação característica dada por,

$$|S_{pxp} - \hat{\lambda}I_{pxp}| = 0.$$

Para cada autovalor $\hat{\lambda}_i$, encontra-se um autovetor $\hat{v}_i$ não nulo, dado por,

$$S_{pxp}\hat{v}_{px1} = \hat{\lambda}\hat{v}_{px1}$$

de forma que,

$$\hat{v}_i = \begin{bmatrix} \hat{v}_{i1} \\ \hat{v}_{i2} \\ \vdots \\ \hat{v}_{ip} \end{bmatrix}$$

Para obter os autovetores normalizados $\hat{e}_i$, com $i = 1, 2, \dots, p$, é necessário dividir cada componente do vetor $\hat{v}_i$ pelo seu comprimento que é dado por,

$$\|\hat{v}_i\| = \sqrt{(\hat{v}_{i1})^2 + (\hat{v}_{i2})^2 + \dots + (\hat{v}_{ip})^2}$$

Segundo Mingoti (2020), o j -ésimo componente principal amostral estimado da matriz de covariâncias é definido por,

$$\hat{Y}_j = \hat{e}_{j1}X_1 + \hat{e}_{j2}X_2 + \dots + \hat{e}_{jp}X_p, \text{ com } j = 1, 2, \dots, p.$$

A variância total explicada pelo j -ésimo componente amostral é dado por,

$$\frac{Var[\hat{Y}_j]}{\text{Variância Total Estimada de } X} = \frac{\hat{\lambda}_j}{\sum_{i=1}^p \hat{\lambda}_i}$$

em que $Var[\hat{Y}_j]$ é a variância estimada do componente $\hat{Y}_j$.

A correlação estimada entre a variável aleatória X_i , $i = 1, 2, \dots, p$, e o j -ésimo componente é dada por,

$$r_{\hat{Y}_j, X_i} = \frac{\hat{e}_{ji} \sqrt{\hat{\lambda}_j}}{\sqrt{S_{ii}}}$$

onde S_{ii} é a variância amostral da variável aleatória X_i .

A obtenção dos componentes principais de um conjunto de dados pode ser feita através da matriz de covariâncias ou matriz de correlação. Os componentes principais obtidos via matriz de covariâncias tem pouca utilidade para os casos onde há uma discrepância muito acentuada entre as variáveis, isso ocorre devido a influência que as variáveis de maior variância causam nos componentes principais.

Nos casos em que existe essa discrepância entre as variâncias observadas é necessário que seja feita uma transformação nos dados, de modo a equilibrar os valores de variância. Uma das transformações mais utilizadas é a padronização de cada variável pela sua média e desvio padrão. Obter os componentes principais através da matriz de covariâncias das variáveis padronizadas é equivalente a utilizar as variáveis originais e matriz de correlação (MINGOTI, 2020).

A matriz de correlação amostral é definida por:

$$R_{pp} = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1p} \\ R_{21} & R_{22} & \dots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p1} & R_{p2} & \dots & R_{pp} \end{bmatrix}$$

em que os coeficientes de correlação amostral de Pearson entre as i -ésimas e j -ésimas variáveis é dado por

$$R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$$

Os autovalores $\hat{\lambda}_i$, $i = 1, 2, \dots, p$, da matriz R_{pp} , são obtidos resolvendo a equação característica dada por,

$$|R_{pp} - \hat{\lambda}I_{pp}| = 0.$$

Para cada autovalor $\hat{\lambda}_i$, existe um autovetor associado $\hat{v}_i$, não nulo, dado por

$$R_{pxp}\hat{v}_{px1} = \hat{\lambda}\hat{v}_{px1}$$

de forma que,

$$\hat{v}_i = \begin{bmatrix} \hat{v}_{i1} \\ \hat{v}_{i2} \\ \vdots \\ \hat{v}_{ip} \end{bmatrix}$$

Para obter os autovetores normalizados $\hat{e}_i$, com $i = 1, 2, \dots, p$, é necessário dividir cada componente do vetor $\hat{v}_i$ pelo seu comprimento que é dado por,

$$\|\hat{v}_i\| = \sqrt{(\hat{v}_{i1})^2 + (\hat{v}_{i2})^2 + \dots + (\hat{v}_{ip})^2}$$

O j -ésimo componente principal da matriz de correlação é definido por,

$$\hat{Y}_j = \hat{e}_{j1}Z_1 + \hat{e}_{j2}Z_2 + \dots + \hat{e}_{jp}Z_p, \text{ com } j = 1, 2, \dots, p.$$

em que Z_i é a variável padronizada e é dada por,

$$Z_i = \frac{(X_i - \mu_i)}{\sigma_i}$$

sendo $E(X_i) = \mu_i$ e $Var(X_i) = \sigma_i^2$, $i = 1, 2, \dots, p$.

A correlação entre o j -ésimo componente principal da matriz de correlação $\hat{Y}_j$ e a variável padronizada Z_i é definida por,

$$r_{\hat{Y}_j, Z_i} = \hat{e}_{ji}\sqrt{\hat{\lambda}_j}$$

A variância total explicada pelo j -ésimo componente principal amostral obtida pela matriz de correlação é dada por,

$$\frac{Var[\hat{Y}_j]}{\text{Variância Total Estimada de Z}} = \frac{\hat{\lambda}_j}{\sum_{i=1}^p \hat{\lambda}_i} = \frac{\hat{\lambda}_j}{p}$$

onde $Var[\hat{Y}_j]$ é a variância estimada do componente principal $\hat{Y}_j$, $j = 1, 2, \dots, p$.

Como um dos objetivos da análise de componentes principais é a redução da dimensão do espaço amostral sem perda de informação, após obter os componentes é necessário decidir quantos serão retidos. Segundo Lattin, Carroll e Green (2011), pode-se utilizar de alguns critérios de escolha para a análise, na significância operacional das soluções e na

possibilidade de interpretação dos componentes principais, ou seja, no processo de seleção do número de componentes a serem mantidos.

Dentre os principais critérios, tem-se o gráfico scree, ou *scree plot*, proposto por Cattell, onde são plotados, em ordem decrescente, a variância explicada por cada componente principal. Buscamos com esse modelo o ponto em que os autovalores $\hat{\lambda}_i$ tendem a se estabilizar, uma vez que esse é o ponto a partir do qual os autovalores se aproximam de zero (MINGOTI, 2020).

A regra proposta por Kaiser recomenda manter no sistema somente os componentes com autovalores que sejam iguais ou excedam a unidade, isto é, $\hat{\lambda}_i \geq 1$. Este critério é adotado quando se trabalha com variáveis padronizadas, ou seja, quando a matriz de correlação é utilizada para extrair os componentes principais. Outro critério é reter os K primeiros componentes capazes de explicar pelo menos 70% da variação total da amostra (FERREIRA, 2008).

Segundo Mingoti (2020) para fazer o uso dos K componentes principais considerados relevantes na análise é necessário calcular os escores dos componentes, que são os valores numéricos de cada elemento.

CAPÍTULO 2

MATERIAIS E MÉTODOS

Para a realização deste trabalho utilizou-se os dados referentes as taxas brutas de mortalidade por câncer, por 100.000 homens e mulheres, nas 27 unidades federativas do Brasil, no ano de 2019.

Os dados foram retirados do site <https://mortalidade.inca.gov.br/MortalidadeWeb/pages/Modelo06/consultar.xhtml#panelResultado>.

Para a análise estatística foram consideradas 15 variáveis referentes a taxa bruta de mortalidade por câncer, por 100.000 homens e mulheres nos 26 estados brasileiros e no Distrito Federal. As variáveis utilizadas foram:

- X1 - Taxa bruta de mortalidade por câncer de bexiga.
- X2 - Taxa bruta de mortalidade por câncer de brônquios e pulmão.
- X3 - Taxa bruta de mortalidade por câncer de colo do útero.
- X4 - Taxa bruta de mortalidade por câncer de esôfago.
- X5 - Taxa bruta de mortalidade por câncer de estômago.
- X6 - Taxa bruta de mortalidade por câncer de fígado e vias biliares intra-hepáticas.
- X7 - Taxa bruta de mortalidade por câncer de intestino delgado.
- X8 - Taxa bruta de mortalidade por câncer de laringe.

- X9 - Taxa bruta de mortalidade por leucemia linfoide.
- X10 - Taxa bruta de mortalidade por câncer de mama.
- X11 - Taxa bruta de mortalidade por câncer de pâncreas.
- X12 - Taxa bruta de mortalidade por meloma maligno da pele.
- X13 - Taxa bruta de mortalidade por câncer de próstata.
- X14 - Taxa bruta de mortalidade por câncer de rim.
- X15 - Taxa bruta de mortalidade por câncer de vesícula biliar.

Para a análise dos dados foi utilizada a seguinte metodologia:

- Análise descritiva dos dados;
- Teste de Mardia para normalidade multivariada, ao nível de significância de 1%;
- Teste de esfericidade de Bartlett, ao nível de significância de 1%;
- Análise de componentes principais.

As análises estatísticas foram realizadas através do *software* R (R CORE TEAM, 2021), juntamente com os pacotes *ellipse* (MURDOCH; CHOW, 2020), *factoextra* (KASSAMBARA; MUNDT, 2020), *FactoMineR* (LÊ; JOSSE; HUSSON, 2008), *MVN* (KORKMAZ; GOKSULUK; ZARARSIZ, 2014), *plotrix* (LEMON, 2006) e *psych* (REVELLE, 2021). A rotina com os comandos do *software* R utilizados na análise dos dados consta no anexo.

CAPÍTULO 3

RESULTADOS E DISCUSSÃO

Inicialmente foi realizada uma análise estatística descritiva das variáveis em estudo. Foi obtida a média, o desvio padrão, o valor mínimo, o valor máximo, o primeiro quartil, a mediana e o terceiro quartil para cada variável, conforme Tabela 3.1

Tabela 3.1: Análise descritiva das variáveis.

Variáveis	Média	Desvio Padrão	Mínimo	Máximo	Primeiro Quartil	Mediana	Terceiro Quartil
X1	0,74	0,38	0,06	1,57	0,43	0,74	1,06
X2	5,85	2,12	3,13	12,54	4,28	5,29	7,28
X3	3,93	1,22	2,12	7,32	3,30	3,68	4,22
X4	1,82	1,28	0,25	5,27	0,81	1,38	2,70
X5	3,22	1,25	1,37	5,28	2,05	3,10	4,06
X6	2,28	0,89	0,00	4,48	1,85	2,40	2,74
X7	0,26	0,09	0,12	0,46	0,20	0,24	0,33
X8	1,02	0,87	0,12	2,62	0,27	0,93	1,67
X9	0,53	0,22	0,25	1,38	0,40	0,51	0,58
X10	3,77	4,10	0,05	13,76	0,09	3,26	7,17
X11	2,21	1,08	0,01	5,49	1,54	2,10	2,60
X12	0,43	0,35	0,11	1,64	0,24	0,36	0,51
X13	7,26	1,62	4,17	10,46	6,08	7,62	8,37
X14	0,70	0,32	0,12	1,45	0,46	0,75	0,87
X15	0,35	0,21	0,07	1,04	0,22	0,32	0,44

O diagrama em caixa ou *boxplot* das variáveis em estudo é apresentado na Figura 3.1. Observa-se na Figura 3.1, que a variável X10 possui grande dispersão e as variáveis X7, X9, X12 e X15 possuem pequena variabilidade. As variáveis X2, X3, X6, X9, X11, X12 e X15 possuem *outliers*.

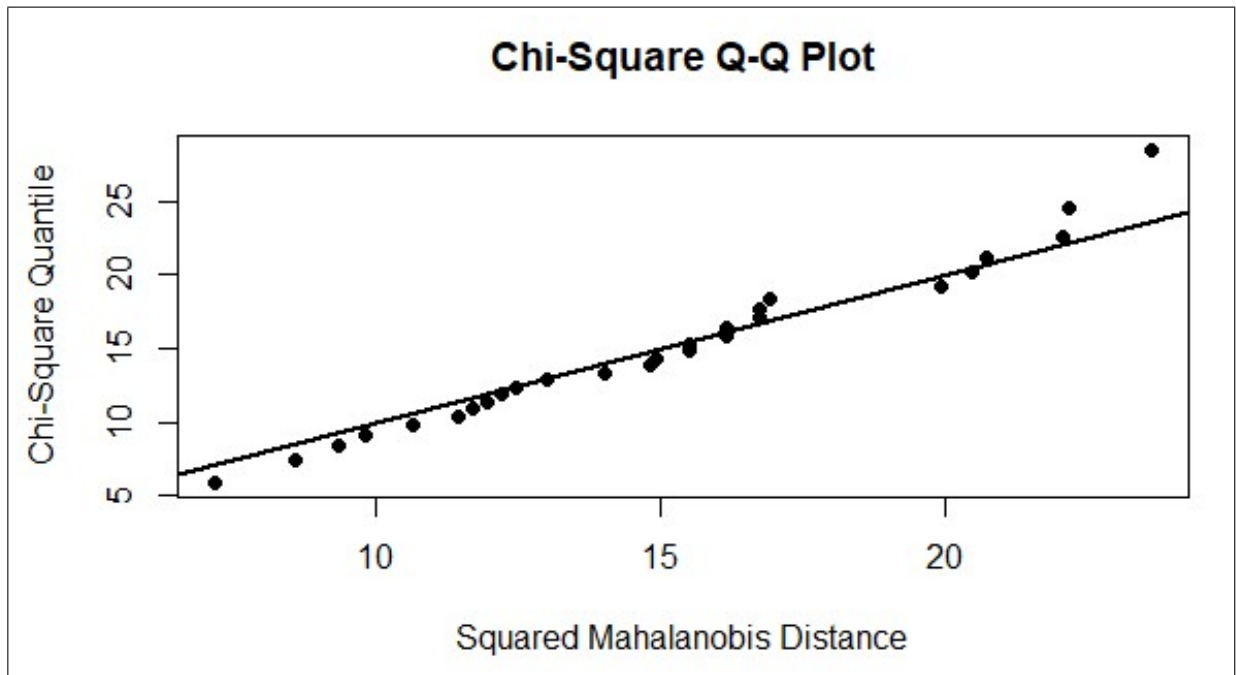


Figura 3.2: *QQ-plot* multivariado.

Com o objetivo de confirmar o resultado, visto que a análise gráfica é bastante subjetiva, foi aplicado o teste de Mardia. Com a aplicação do teste multivariado Mardia (assimetria), obteve-se estatística de teste igual a 683,63 e $\text{valor} - p = 0,46$. A aplicação do teste multivariado Mardia (curtose), forneceu estatística de teste igual a $-1,28$ e $\text{valor} - p = 0,20$. Dessa forma, ao nível de significância de 1%, conclui-se que as variáveis em estudo seguem uma distribuição normal multivariada.

Em seguida foi obtida as correlações, apresentadas na Tabela 3.2, entre as variáveis em análise.

Tabela 3.2: Correlação entre as variáveis.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
X1	1,00	0,67	-0,39	0,37	0,48	0,04	0,42	0,12	0,13	0,06	0,41	0,26	0,56	0,71	-0,37
X2		1,00	-0,40	0,34	0,26	0,18	0,52	0,26	0,15	0,23	0,65	0,71	0,67	0,77	-0,21
X3			1,00	-0,31	0,07	0,17	-0,42	-0,26	0,31	-0,31	-0,24	-0,32	-0,53	-0,28	0,08
X4				1,00	0,06	-0,31	0,16	0,22	0,11	0,11	0,17	0,16	0,48	0,38	0,24
X5					1,00	-0,16	0,26	0,00	0,21	0,00	0,01	-0,08	0,04	0,57	-0,26
X6						1,00	-0,19	0,02	0,05	-0,17	0,47	0,30	0,21	-0,11	-0,17
X7							1,00	0,31	0,25	0,19	0,30	0,51	0,53	0,41	-0,05
X8								1,00	0,23	-0,15	0,24	0,46	0,23	0,26	0,29
X9									1,00	-0,10	0,24	0,18	0,09	0,18	0,11
X10										1,00	-0,11	-0,13	0,27	-0,03	-0,09
X11											1,00	0,73	0,60	0,55	0,04
X12												1,00	0,48	0,51	0,09
X13													1,00	0,47	-0,11
X14														1,00	-0,21
X15															1,00

As elipses de correlação, Figura 3.3, permitem verificar a força e a direção das associações entre as variáveis. As cores são proporcionais a força da correlação entre as variáveis. Dessa forma, tem-se que as correlações próximas de 1 e -1 são representadas por elipses mais escuras e as correlações próximas de zero se aproximam do branco. A correlação igual a 1 é representada por um segmento de reta e quanto mais próximo de zero for a correlação mais próximo de um círculo. A elipse cuja reta que contém o seu eixo maior for crescente indica correlação positiva e a elipse cuja reta que contém o seu eixo maior for decrescente indica correlação negativa. Observa-se que as variáveis X14 e X2 apresentam o maior coeficiente de correlação, sendo igual a 0,77, já as variáveis X5 e X8 não são correlacionadas. Também não existe correlação entre as variáveis X5 e X10.

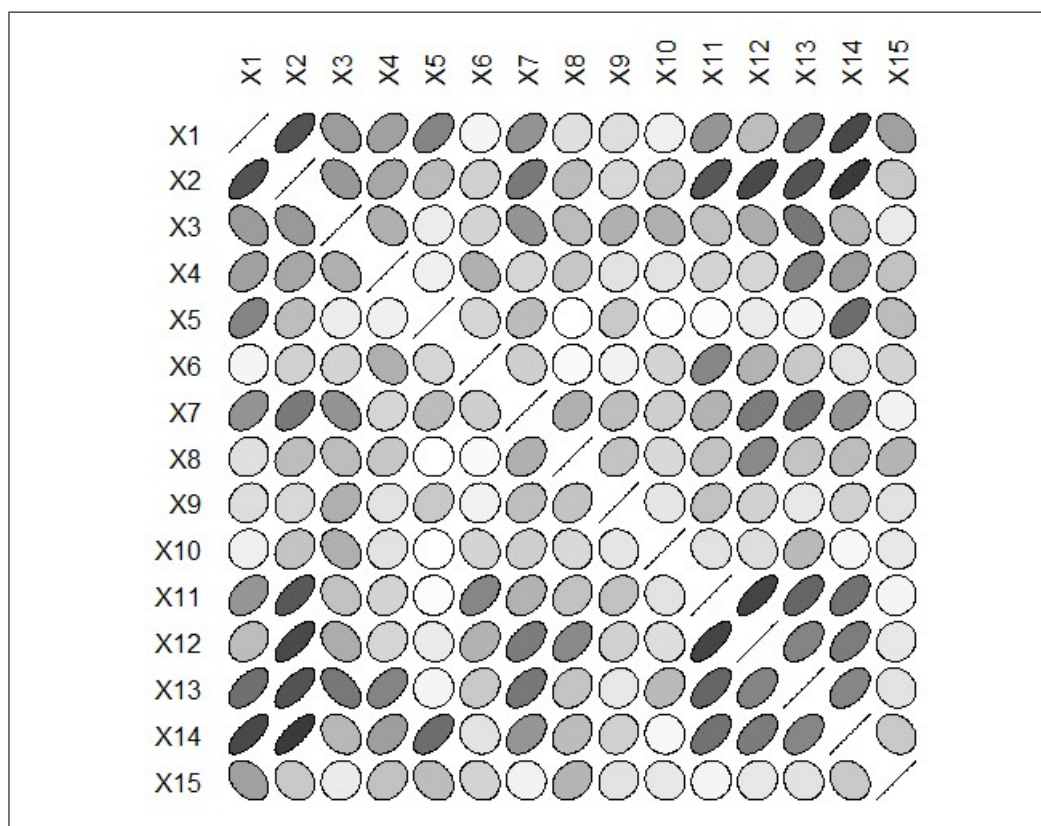


Figura 3.3: Elipses de correlação.

A Figura 3.4 além de apresentar o coeficiente de correlação, também apresenta o gráfico de dispersão entre as variáveis. A Tabela 3.2, a Figura 3.3 e a Figura 3.4 fornecem informações relevantes em relação a correlação entre as variáveis em estudo.

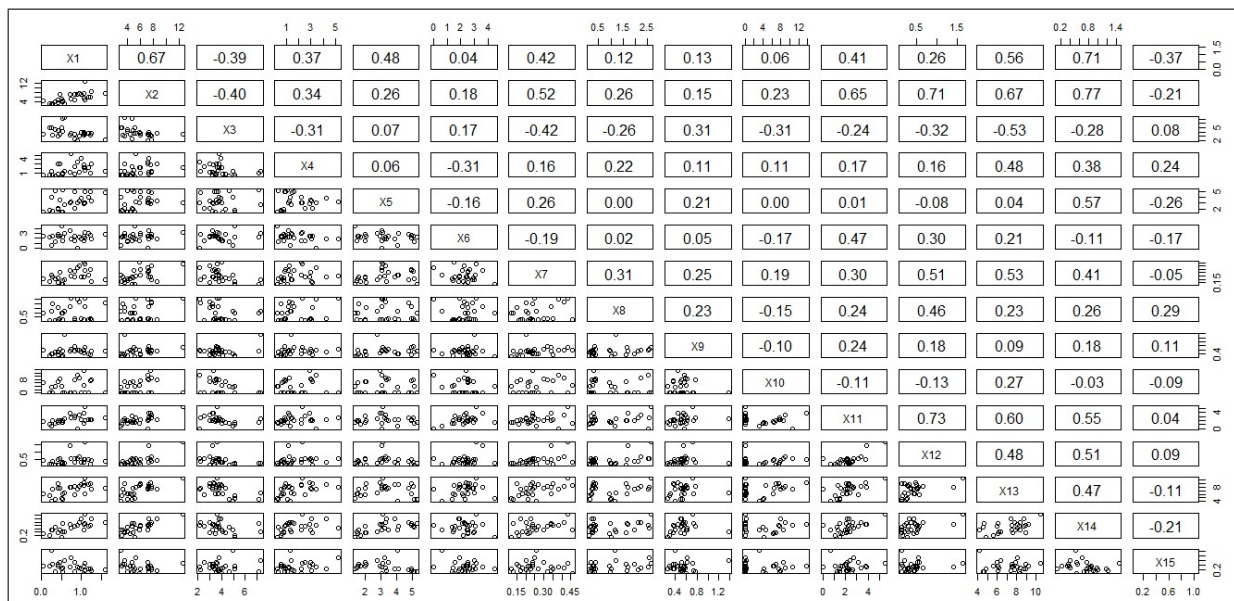


Figura 3.4: Gráfico de dispersão e coeficiente de correlação entre as variáveis.

Para verificar se as variáveis em estudo estão correlacionadas de alguma forma aplicou-se o teste de esfericidade de Bartlett. Obteve-se $\chi^2 = 218,52$ e $\text{valor}-p = 5,665238 \times 10^{-10}$, logo pode-se concluir que as variáveis não são mutuamente independentes, ao nível de significância de 1%.

Após verificar que as variáveis são correlacionadas, realizou-se a análise de componentes principais. Observa-se na Tabela 3.3 que os quatro primeiros componentes explicam aproximadamente 69,80% da variação total dos dados, sendo aproximadamente 33,80% do primeiro componente principal (CP1), 13,42% do CP2, 11,90% do CP3 e 10,68% do CP4.

Tabela 3.3: Componentes principais (CP), autovalores, porcentagem da variância total explicada pelo componente e a porcentagem acumulada da explicação da variância total.

Componentes Principais	Autovalor	Porcentagem	Porcentagem acumulada
CP1	5,07	33,80	33,80
CP2	2,01	13,42	47,22
CP3	1,78	11,90	59,12
CP4	1,60	10,68	69,80
CP5	0,99	6,58	76,38
CP6	0,93	6,23	82,61
CP7	0,62	4,12	86,73
CP8	0,56	3,74	90,47
CP9	0,45	2,99	93,45
CP10	0,33	2,20	95,65
CP11	0,24	1,58	97,23
CP12	0,20	1,32	98,55
CP13	0,11	0,75	99,30
CP14	0,06	0,40	99,70
CP15	0,04	0,30	100,00

Observa-se na Figura 3.5 o *Screeplot* e na Figura 3.6 o gráfico da porcentagem da variância total explicada por cada componente.

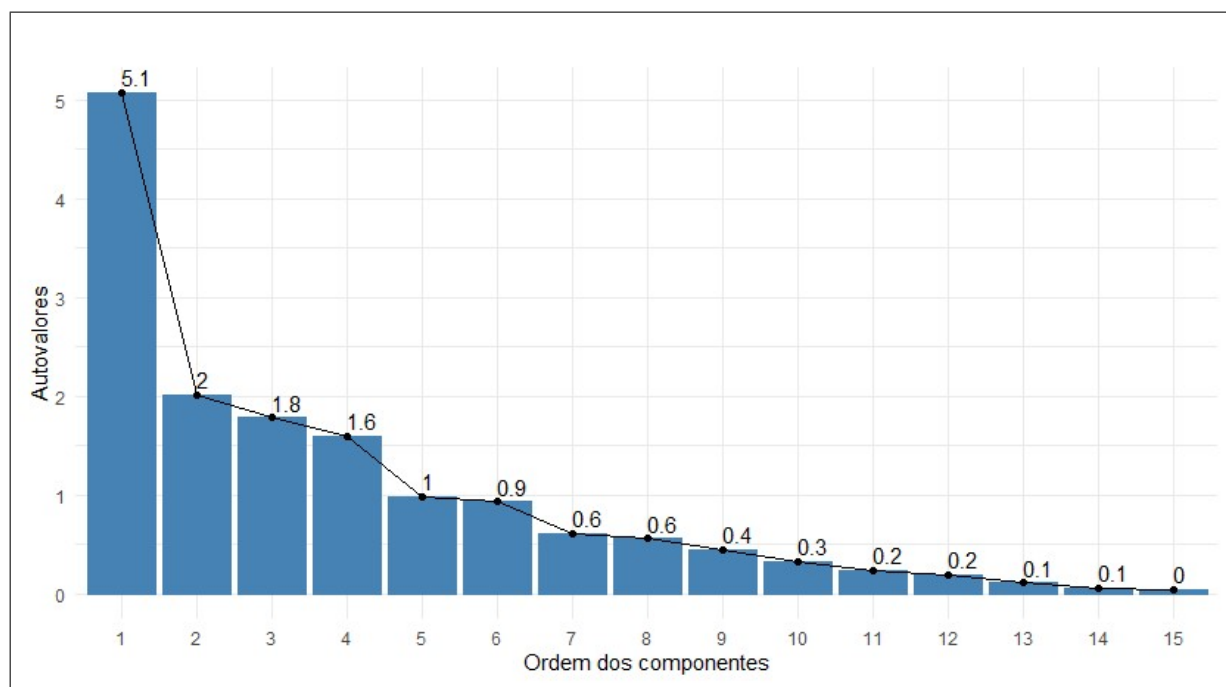
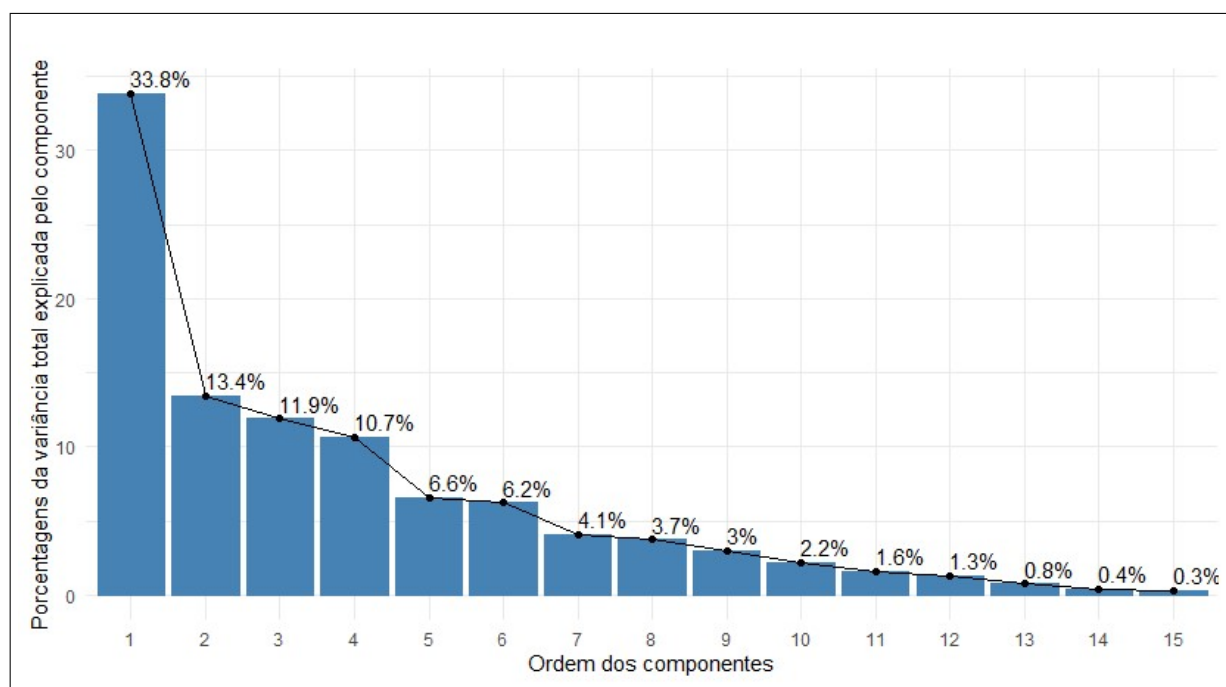
Figura 3.5: *Screeplot*

Figura 3.6: Porcentagem da variância total explicada por cada componente.

Para definir o número de componentes que serão retidos, foram utilizados os critérios discutidos no referencial teórico. Com base na avaliação dos autovalores (Tabela 3.3 e Figura 3.5) e da porcentagem acumulada da explicação da variância total (Figura 3.6) concluiu-se que os CP1, CP2, CP3 e CP4 resumem de maneira adequada a variação dos dados, visto que explicam aproximadamente 70% da variação total e estão associados aos autovalores maiores que 1. Dessa forma, apresentam-se as equações para os quatro primeiros componentes principais:

$$CP1 = 0,34Z_1 + 0,40Z_2 - 0,24Z_3 + 0,20Z_4 + 0,14Z_5 + 0,05Z_6 + 0,30Z_7 + 0,18Z_8 + 0,10Z_9 + 0,06Z_{10} + 0,32Z_{11} + 0,32Z_{12} + 0,36Z_{13} + 0,36Z_{14} - 0,06Z_{15}$$

$$CP2 = -0,24Z_1 - 0,01Z_2 + 0,19Z_3 - 0,14Z_4 - 0,34Z_5 + 0,44Z_6 - 0,09Z_7 + 0,26Z_8 + 0,19Z_9 - 0,34Z_{10} + 0,35Z_{11} + 0,36Z_{12} - 0,13Z_{14} + 0,29Z_{15}$$

$$CP3 = 0,21Z_1 + 0,08Z_2 + 0,37Z_3 - 0,34Z_4 + 0,39Z_5 + 0,33Z_6 - 0,12Z_7 - 0,25Z_8 + 0,17Z_9 - 0,24Z_{10} + 0,10Z_{11} - 0,05Z_{12} - 0,13Z_{13} + 0,19Z_{14} - 0,46Z_{15}$$

$$CP4 = 0,01Z_1 - 0,10Z_2 + 0,29Z_3 + 0,22Z_4 + 0,35Z_5 - 0,38Z_6 + 0,11Z_7 + 0,28Z_8 + 0,47Z_9 - 0,28Z_{10} - 0,12Z_{11} - 0,05Z_{12} - 0,22Z_{13} + 0,19Z_{14} + 0,33Z_{15}$$

É importante conhecer quais são as variáveis que mais contribuíram para a variabilidade dos quatro primeiros componentes. A Figura 3.7 mostra os gráficos de barras contendo a importância das variáveis no primeiro, segundo, terceiro e quarto componente principal.

A variável X2 é a variável de maior importância no componente 1 e as variáveis X6, X10 e X15 possuem menor importância no CP1. No CP2 a X6 é a variável de maior importância, e as variáveis X2 e X13 tem coeficientes próximos de zero, não tendo relevância neste componente.

A componente principal 1 (CP1) representa uma comparação entre a variável mortalidade por câncer de colo de útero (X3) e as demais variáveis X1, X2, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13 e X14, exceto a variável X15 que possui coeficiente muito pequeno. Isso significa que um estado que possui escore negativo para essa componente apresenta taxa maior de câncer de colo de útero em comparação com os outros tipos de câncer.

A CP2 é uma comparação entre as variáveis X1, X4, X5, X10, X14 e as variáveis X3, X6, X8, X9, X11, X12, X15, visto que os coeficientes relacionados as variáveis X2, X7 e X13 são bem próximos de zero.

A CP3 representa uma comparação entre as variáveis X1, X3, X5, X6, X11, X14 e as variáveis X4, X7, X8, X9, X10, X13, X15, visto que os coeficientes relacionados as variáveis X2 e X12 são bem próximos de zero.

A componente principal 4 (CP4) representa uma comparação entre a variável X2, X6, X10, X11, X13 e as variáveis X3, X4, X5, X7, X8, X9, X14, X15, já que o coeficiente relacionado as variáveis X1 e X12 é muito pequeno.

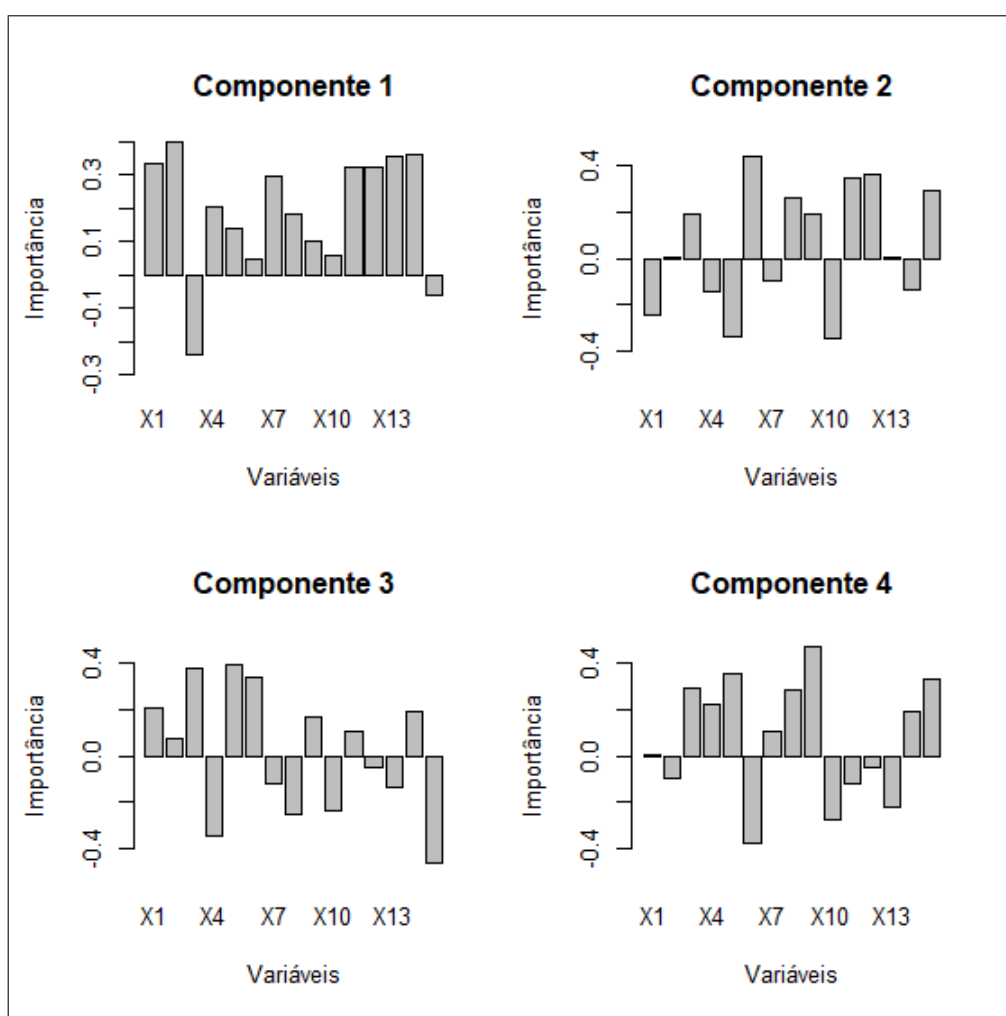


Figura 3.7: Gráficos de barras contendo a importância das variáveis no primeiro, segundo, terceiro e quarto componente principal.

A Tabela 3.4 e a Figura 3.8 apresentam as correlações entre os quatro primeiros componentes e as variáveis em análise. A variável X2 é altamente correlacionada com a CP1, sendo a variável de maior importância na componente. Existe uma correlação negativa entre as variáveis X3, X15 e CP1, sendo que a correlação entre X15 e CP1 é baixa. A correlação entre X2, X13 e CP2 é próximo de zero, indicando que essas são as variáveis de menor importância CP2.

Tabela 3.4: Correlação entre os CP1, CP2, CP3, CP4 e as variáveis

Variáveis	CP1	CP2	CP3	CP4
X1	0,76	-0,34	0,28	0,01
X2	0,90	-0,01	0,10	-0,12
X3	-0,54	0,27	0,50	0,36
X4	0,46	-0,20	-0,45	0,28
X5	0,31	-0,48	0,52	0,44
X6	0,11	0,62	0,45	-0,48
X7	0,66	-0,13	-0,16	0,13
X8	0,41	0,37	-0,33	0,36
X9	0,23	0,26	0,23	0,60
X10	0,13	-0,48	-0,31	-0,35
X11	0,73	0,49	0,14	-0,15
X12	0,73	0,51	-0,07	-0,06
X13	0,80	-0,01	-0,18	-0,27
X14	0,82	-0,19	0,25	0,24
X15	-0,14	0,41	-0,61	0,42

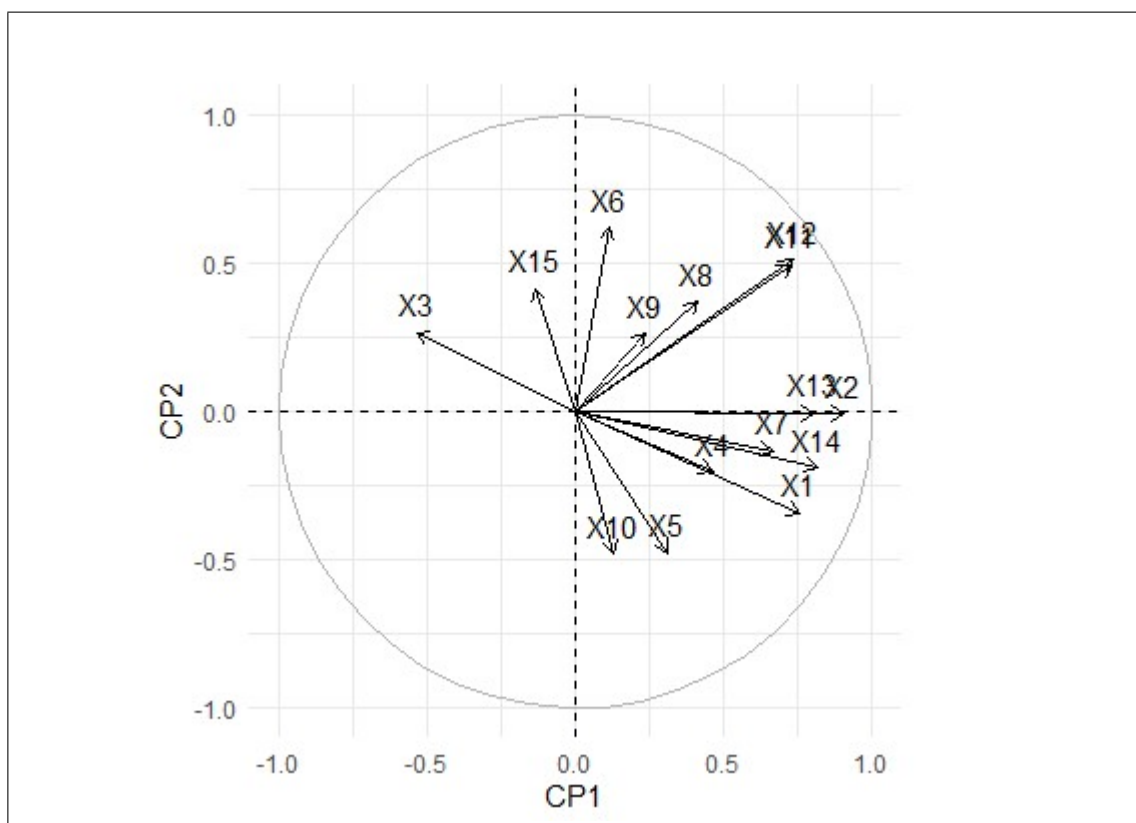


Figura 3.8: Gráficos da correlação entre CP1, CP2 e as variáveis.

Os escores das quatro primeiras componentes principais podem ser visualizados na Tabela 3.5 e no Biplot, Figura 3.9. Observa-se no Biplot que no Acre, Alagoas, Amapá, Amazonas, Bahia, Distrito Federal, Maranhão, Mato Grosso, Pará, Piauí, Roraima e Sergipe prevalece a mortalidade por câncer de colo de útero em relação a todos os demais.

Tabela 3.5: Escores dos 4 primeiros componentes

Codificação dos estados	Estados	Escores da CP1	Escores da CP2	Escores da CP3	Escores da CP4
1	Acre	-3,09	1,66	0,86	-2,35
2	Alagoas	-2,16	0,98	-0,91	-1,02
3	Amapá	-3,17	-0,29	-1,92	2,64
4	Amazonas	-2,09	-0,22	2,95	1,60
5	Bahia	-1,12	0,13	-2,08	-1,08
6	Ceará	0,36	-0,16	-0,96	-0,22
7	Distrito Federal	-0,14	0,31	-0,94	0,05
8	Espírito Santo	1,01	1,04	-2,21	2,06
9	Goiás	1,43	-1,77	0,63	-0,39
10	Maranhão	-3,84	0,51	0,01	-0,82
11	Mato Grosso	-1,43	1,38	-1,15	-0,52
12	Mato Grosso do Sul	1,44	-1,24	0,46	0,48
13	Minas Gerais	2,18	-0,48	1,27	0,06
14	Paraná	2,58	-0,07	-0,43	-0,31
15	Paraíba	1,07	-1,79	0,94	-1,00
16	Pará	-2,42	-1,67	1,53	0,61
17	Pernambuco	0,88	-0,37	1,07	-0,72
18	Piauí	-1,55	0,63	-0,22	-1,37
19	Rio Grande do Norte	1,03	-1,16	1,46	-0,27
20	Rio Grande do Sul	6,39	3,13	0,15	0,43
21	Rio de Janeiro	1,88	-0,88	-0,69	-1,62
22	Rondônia	0,76	-0,57	1,30	1,53
23	Roraima	-1,94	2,60	1,45	2,89
24	Santa Catarina	2,49	2,59	0,49	-0,77
25	Sergipe	-2,34	0,11	-0,23	-0,94
26	São Paulo	1,27	-2,56	-2,90	1,20
27	Tocantins	0,51	-1,83	0,08	-0,15

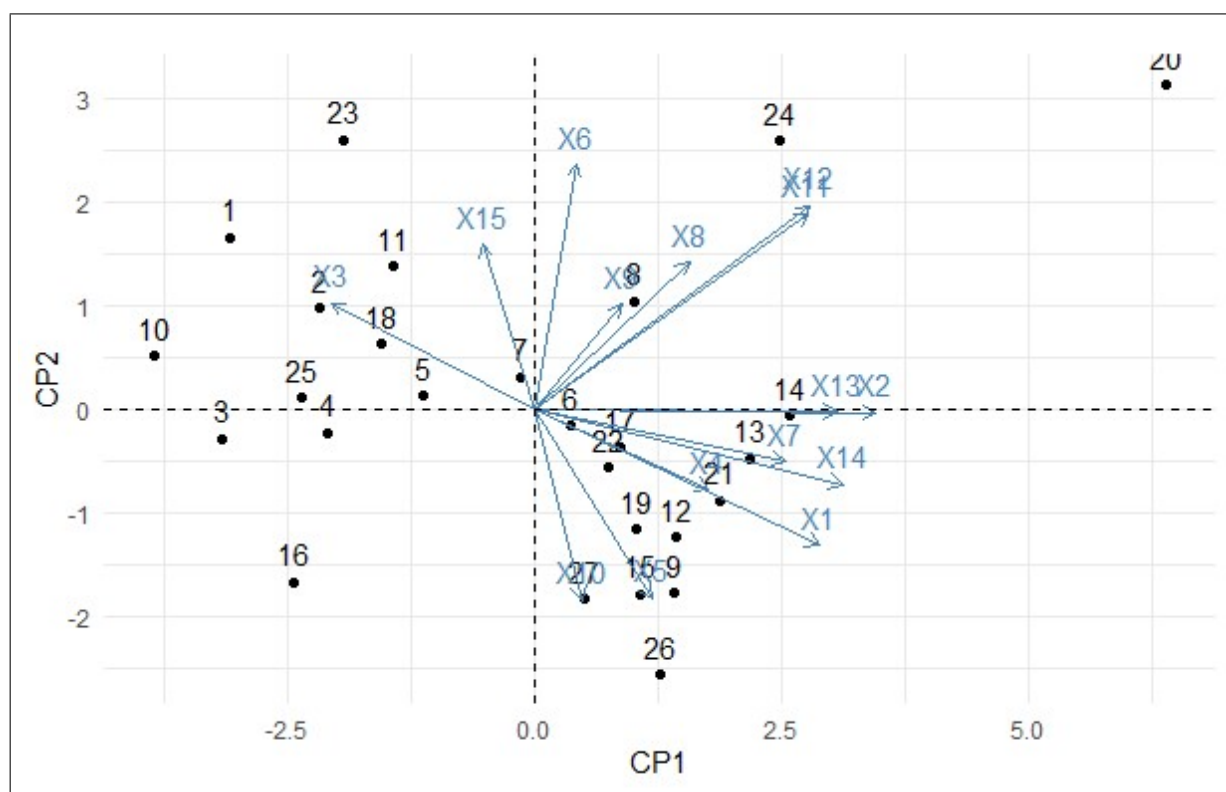


Figura 3.9: Biplot.

CAPÍTULO 4

CONCLUSÃO

A análise de componentes principais para os dados de mortalidade por câncer nos estados brasileiros em 2019 mostrou-se eficiente, pois reduziu a quantidade inicial de 15 variáveis para quatro componentes principais que explicam aproximadamente 70% de toda variação dos dados. O trabalho forneceu informações importantes, principalmente em relação ao câncer de colo de útero, que podem contribuir para o direcionamento de ações preventivas em relação a esse tipo câncer.

REFERÊNCIAS

- [1] FERREIRA, D. F. **Estatística multivariada**. Lavras: Editora UFLA, 2008.
- [2] INSTITUTO DE CÂNCER DE BRASÍLIA. **Oncologia: entenda como a especialidade conduz o tratamento do câncer**. Brasília: ICB, 2018. Disponível em: <https://institutodecancer.com.br/oncologia-entenda-como-especialidade-conduz-o-tratamento-cancer/> Acesso em: 14 fev. 2022
- [3] INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Atlas On-line de Mortalidade. Representação espacial das taxas brutas ou ajustadas por idade pela população mundial de mortalidade por câncer, por 100.000, segundo sexo, nas Unidades da Federação, Regionais de Saúde ou Municípios, por período selecionado**. Rio de Janeiro: INCA, 2022. Disponível em: <https://mortalidade.inca.gov.br/MortalidadeWeb/pages/Modelo06/consultar.xhtml#panelResultado> Acesso em: 7 jan. 2022.
- [4] INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Estatísticas de câncer**. Rio de Janeiro: INCA, 2021. Disponível em: <https://www.inca.gov.br/numeros-de-cancer> Acesso em: 14 fev. 2022.

- [5] INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **O que é câncer?** Rio de Janeiro: INCA, 2020. Disponível em: <https://www.inca.gov.br/o-que-e-cancer> Acesso em: 14 fev. 2022
- [6] KASSAMBARA, A.; MUNDT, F. R package version 1.0.7. **factoextra: Extract and Visualize the Results of Multivariate Data Analyses**. [S.I], 2020. Disponível em: <https://CRAN.R-project.org/package=factoextra>
- [7] KORKMAZ, S.; GOKSULUK, D.; ZARARSIZ, G. MVM: An R Package for Assessing Multivariate Normality. **The R Journal**, v. 6, n. 2, p. 151-162, 2014. Disponível em: <https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf>
- [8] LATTIN, J.; CARROLL, J. D.; GREEN, P. E. **Análise de dados multivariados**. São Paulo: Cengage Learning, 2011.
- [9] LÊ, S.; JOSSE, J.; HUSSON, F. FactoMineR: A Package for Multivariate Analysis. **Journal of Statistical Software**, v. 25, n. 1, p. 1-18, 2008.
- [10] LEMON, J. Plotrix: a package in the red light district of R. **R-News**, v. 6, n. 4, p. 8-12, 2006.
- [11] MACHADO, C. J. Mortalidade infantil no Estado de São Paulo, 1999: uma análise das causas múltiplas de morte a partir de componentes principais. **Revista Brasileira de Epidemiologia**, v. 7, n. 1, p. 11-21, 2004.
- [12] MARTINS, A. A. **Estudo da morbidade hospitalar do SUS pela Análise de Componentes Principais**. 2018. 21 f. Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Federal de Uberlândia, Uberlândia, 2018. Disponível em: <https://repositorio.ufu.br/handle/123456789/24111> Acesso em: 15 jan. 2022
- [13] MINGOTI, S. A. **Análise de dados através de métodos estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2020.
- [14] MURDOCH, M.; CHOW, E. D. R package version 0.4.2. **ellipse: Functions for Drawing Ellipses and Ellipse-Like Confidence Regions**. [S.I], 2020. Disponível em: <https://CRAN.R-project.org/package=ellipse>

- [15] ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. **Câncer**. Disponível em: <https://www.paho.org/pt/topicos/cancer> Acesso em: 14 fev. 2022
- [16] R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2021. Disponível em: <https://www.R-project.org/>
- [17] REVELLE, W. R package version 2.1.9. **psych: Procedures for Psychological, Psychometric, and Personality Research**. Evanston, Illinois: Northwestern University, 2021. Disponível em: <https://CRAN.R-project.org/package=psych>
- [18] SANTOS, J. P. **Fatores associados à mortalidade dos idosos com as condições de vida da população do Nordeste do Brasil**. 2014. 114 f. Tese (Doutorado em Ciências da Saúde) - Universidade Federal do Rio Grande do Norte, Natal, 2014. Disponível em: <https://repositorio.ufrn.br/handle/123456789/13358> Acesso em: 14 fev. 2022
- [19] TRIOLA, M. F. Introdução à Estatística: Atualização da Tecnologia. 12^a. Ed. Rio de Janeiro: LTC, 2017.

ANEXO

Rotina de comandos do *software* R.

#instalando os pacotes

```
install.packages("MVN",dependencies=TRUE)
install.packages("plotrix",dependencies=TRUE)
install.packages("ellipse",dependencies=TRUE)
install.packages("psych",dependencies=TRUE)
install.packages("FactoMineR",dependencies=TRUE)
install.packages("factoextra",dependencies=TRUE)
```

#referencias

```
citation()
citation("MVN")
citation("plotrix")
citation("ellipse")
citation("psych")
citation("FactoMineR")
citation("factoextra")
```

#lendo os dados

```

dados <-read.table("dadoscancer.csv",header=T, sep = ";")
dados
dados <- dados_cancer[, -1] #retira coluna das UF
dados

#analise descritiva dos dados
summary(dados)
round(sd(dados$X1), digits=2)
round(sd(dados$X2), digits=2)
round(sd(dados$X3), digits=2)
round(sd(dados$X4), digits=2)
round(sd(dados$X5), digits=2)
round(sd(dados$X6), digits=2)
round(sd(dados$X7), digits=2)
round(sd(dados$X8), digits=2)
round(sd(dados$X9), digits=2)
round(sd(dados$X10), digits=2)
round(sd(dados$X11), digits=2)
round(sd(dados$X12), digits=2)
round(sd(dados$X13), digits=2)
round(sd(dados$X14), digits=2)
round(sd(dados$X15), digits=2)

boxplot(dados, col="gray")

#teste de normalidade

library(MVN) #baixando o pacote

qqplot <- mvn(dados, multivariatePlot = "qq")

teste_normalidade <- mvn(data = dados, mvnTest = "mardia")
teste_normalidade

```

```

#obtendo a matriz de correlacao

matriz_correlacao <- round(cor(dados), 2)
matriz_correlacao

#grafico para representar a matriz de correlacao

#funcao disponivel em help(pairs), porem com algumas
modificacoes

panel.cor <- function(x, y, digits = 2, prefix = "",
cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- round(cor(x, y), 2)
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = 1.5)
}

pairs(dados, upper.panel = panel.cor)

#grafico para representar a matriz de correlacao

library(plotrix)
library(ellipse)

mcores <- color.scale(1-abs(matriz_correlacao))
plotcorr(matriz_correlacao, col=mcores)

```

```

#teste de esfericidade de Bartlett

library(psych)

cortest.bartlett(matriz_correlacao,n=27)

# Analise de componentes principais

library(FactoMineR)

acp.cor <- PCA(dados, scale.unit = T, graph = FALSE, ncp=15)

round(acp.cor$eig,2) #autovalores
round(acp.cor$svd$V,2) #autovetores

par(mfrow = c(2,2))

#grafico de barras contendo a importancia das variaveis no
primeiro componente

barplot(acp.cor$svd$V[,1], beside=TRUE,ylim = c(-0.3, 0.4),
ylab="Importancia",xlab="Variaveis", names.arg = c("X1","X2",
"X3","X4","X5","X6","X7","X8","X9","X10","X11","X12","X13",
"X14","X15"),main="Componente 1")

#grafico de barras contendo a importancia das variaveis no
segundo componente

barplot(acp.cor$svd$V[,2], beside=TRUE,ylim = c(-0.5, 0.5),
ylab="Importancia",xlab="Variaveis", names.arg = c("X1","X2",
"X3","X4","X5","X6","X7","X8","X9","X10","X11","X12","X13",
"X14","X15"),main="Componente 2")

```

```

#grafico de barras contendo a importancia das variaveis no
terceiro componente

barplot(acp.cor$svd$V[,2], beside=TRUE, ylim=c(-0.5, 0.5),
ylab="Importancia", xlab="Variaveis", names.arg=c("X1", "X2",
"X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10", X11, "X12", "X13",
"X14", "X15"), main="Componente 3")

#grafico de barras contendo a importancia das variaveis no
terceiro componentes

barplot(acp.cor$svd$V[,2], beside=TRUE, ylim=c(-0.5, 0.5),
ylab="Importancia", xlab="Variaveis", names.arg=c("X1", "X2",
"X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10", X11, "X12", "X13",
"X14", "X15"), main="Componente 4")

library(factoextra)

#visualizacao dos autovalores

fviz_eig(acp.cor, choice="eigenvalue", ncp=15, axes=1, title="",
xlab="Ordem dos componentes", ylab="Autovalores", addlabels=TRUE)
+theme_minimal()

#porcentagem da variacao explicada pelo componente

fviz_screplot(acp.cor, ncp=15, title="", xlab="Ordem dos
componentes", ylab="Porcentagens da variancia total explicada
pelo componente", addlabels=TRUE)+theme_minimal()

#correlacao entre a variavel e o componente principal

```

```
round(acp.cor$var$cor,2)

fviz_pca_var(acp.cor, title="", xlab="CP1", ylab="CP2")
+theme_minimal()

#obtencao_dos_escores

round(acp.cor$ind$coord,2)

#biplot

fviz_pca_biplot(acp.cor, title="", xlab="CP1", ylab="CP2")
+theme_minimal()
```