



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Bacharelado em Estatística

**ESTUDO DA MORBIDADE
HOSPITALAR DO SUS PELA ANÁLISE
DE COMPONENTES PRINCIPAIS**

Andriely Antunes Martins

Uberlândia-MG

2018

Andriely Antunes Martins

**ESTUDO DA MORBIDADE
HOSPITALAR DO SUS PELA ANÁLISE
DE COMPONENTES PRINCIPAIS**

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

Orientadora: Mirian Fernandes Carvalho Araújo

**Uberlândia-MG
2018**



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

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

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Mirian Fernandes Carvalho Araújo

Priscila Neves Faria

Patrícia Ferreira Paranaíba

**Uberlândia-MG
2018**

AGRADECIMENTOS

A Deus desde o primeiro momento em que fui abençoada ao ser aprovada no Vestibular. Obrigada por me transmitir força, foco e fé que me acompanharam ao longo desses anos e que não me permitiram desistir. Serei eternamente grata à Deus por todas as bênçãos sobre a minha família e por proporcionar tranquilidade aos corações daqueles que acompanharam a minha trajetória acadêmica.

A Universidade Federal de Uberlândia que me deu a oportunidade de cursar Estatística nesta renomada instituição. Obrigada por proporcionar um ambiente saudável para todos os alunos, além de estimular a criatividade, a interação e a participação nas atividades acadêmicas. Sou grata a todo corpo docente, à direção e a administração dessa instituição.

A todos os mestres do curso, que compartilharam seus conhecimentos em sala de aula e acompanharam a minha jornada enquanto universitária. Sou grata especialmente aos professores Mirian Fernandes Carvalho Araújo e Lúcio Borges de Araújo, que foram meus orientadores, no qual me auxiliaram em toda a parte do meu trabalho, pela organização dos dados, nas pesquisas e na revisão da minha monografia.

Agradeço aos meus pais, que me proporcionaram a melhor educação e lutaram para que eu estivesse concluindo mais essa etapa da minha vida. Sei o quanto vocês se doaram para a realização desse sonho. Ao meu irmão, tios, avós e primos que entenderam a minha ausência, acompanharam a minha dedicação e torceram por mim. Sou grata ao meu namorado, que me apoiou desde o primeiro dia de aula, e em todos os momentos perante as dificuldades soube me dar forças para vencer mais essa etapa da minha vida. E claro que não podia deixar de agradecer aos meus amigos e colegas que torceram muito por mim.

RESUMO

Os estudos estatísticos da mortalidade são de extrema importância em demografia e na área da saúde, pois permitirão compreender a magnitude desse fenômeno, bem como o impacto na vida e saúde da população, de modo a auxiliar formulações de políticas públicas e de promoção/prevenção. O principal objetivo deste trabalho foi destacar os principais casos de homicídios para cada cidade pesquisada, em relação à morbidade hospitalar de grupos de causas da CID10. Diante disso, utilizou a técnica de Análise de Componentes Principais e o Teste U de Mann-Whitney, para verificar se houve diferença entre cidades grandes com mais de 30 mil habitantes, com as cidades pequenas com menos de 30 mil habitantes em relação as causas de internações hospitalares do SUS. Como resultado, obteve quatro componentes principais no estudo, e pelo teste t aceita-se a hipótese nula de que não há diferença entre os grupos tratados em relação as causas de internações hospitalares.

Palavras-chave: CID10, Mortalidade, Saúde, Multivariada.

ABSTRACT

Statistical studies of mortality are extremely important in demography and in the health area, since they will allow the understanding of the magnitude of this phenomenon, as well as the impact on the life and health of the population, in order to support formulations of public policies and promotion / prevention. The main objective of this study was to highlight the main homicide cases for each city surveyed, in relation to the hospital morbidity of cause groups of CID10. In this context, the principal components analysis and the Mann-Whitney U test were used to to verify if there was a difference between large cities with more than 30 thousand inhabitants, with small cities with less than 30 thousand inhabitants in relation to the hospital admission causes of SUS. As a result, we obtained four main components in the study, and the t-test accepts the null hypothesis that there is no difference between the treated groups regarding the causes of hospital admissions.

Keywords: CID10, Mortality, Health, Multivariate.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	III
1 Introdução	1
2 Justificativa	3
3 Objetivos	5
4 Metodologia	7
5 Resultados	11
6 Conclusões	19
Referências Bibliográficas	21

LISTA DE FIGURAS

4.1	Morbidade Hospitalar do SUS CID-10.	7
5.1	Boxplot das causas de internações hospitalares nas cidades do Triângulo Mineiro do ano 2009.	12
5.2	O Screeplot dos autovalores dos componentes principais.	14
5.3	Biplot $CP\ 1$ x $CP\ 2$ do percentual das causas de internações hospitalares.	16

LISTA DE TABELAS

5.1	Análise descritiva das causas de internações nas cidades do Triângulo Mineiro em porcentagem no ano de 2009.	12
5.2	Componentes Principais (CP's), autovalores (λ_i) e porcentagem da variância explicada e proporção acumulada (%) pelos componentes.	13
5.3	Coefficientes de ponderação das características com os quatro primeiros componentes principais.	14
5.4	Interpretação de cada componente.	16
5.5	Teste de Normalidade.	17
5.6	Teste de Levene para igualdade de variâncias e teste t de Student para igualdade de médias.	18

1. INTRODUÇÃO

O Triângulo Mineiro é uma das regiões mais ricas do estado de Minas Gerais, no sudeste do Brasil, com a economia voltada a distribuição. As principais indústrias ali instaladas relacionam-se aos setores de processamento de alimentos e de madeira, de açúcar e álcool, fumo e de fertilizantes. Nos últimos anos o Triângulo Mineiro é a região que mais tem recebido investimentos e mais empregos tem gerado [1]. Minas Gerais tem dez regiões de planejamento, das quais duas, Triângulo Mineiro e a do Alto Paranaíba, formam a Mesorregião do Triângulo Mineiro e Alto Paranaíba, definida pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O total de municípios dessa mesorregião é de 66, sendo sete microrregiões. Está situado entre os rios Grande e Paranaíba, formadores do rio Paraná. Uberlândia, Araguari, Ituiutaba e Araxá são as principais cidades do Triângulo Mineiro.

O Produto Interno Bruto (PIB) do Triângulo Mineiro registrado em 2009 era de 42,897 bilhões de reais. Se compararmos a economia do Triângulo Mineiro com a dos estados do Brasil, o Triângulo Mineiro possui um PIB pouco menor que o do Maranhão, com seus 45,256 bilhões, e do Mato Grosso do Sul, com seus 43,514 bilhões. Minas Gerais, está atrás apenas da Mesorregião Metropolitana de Belo Horizonte. O Índice de Desenvolvimento Humano (IDH) do Triângulo Mineiro era de 0,885. É considerado elevado pelo Programa das Nações Unidas para o Desenvolvimento (PNUD), sendo o segundo melhor do estado, atrás somente da Mesorregião Metropolitana de Belo Horizonte, com 0,911.

Os serviços de saúde da mesorregião do Triângulo Mineiro e Alto Paranaíba tem destaque para os municípios pólos de Uberlândia, Uberaba e Patos de Minas, que são os principais responsáveis por atender a demanda gerada pelo fluxo de indivíduos que se dirigem a esses locais, visto o atendimento médico oferecido pelos mesmos. De acordo com o IBGE, a mesorregião possuía em 1999, 2,8 médicos por mil habitantes e 2,1 especialistas por mil habitantes. Os cinco principais municípios da mesorregião que se destacam por oferecer mais de cem tipos de serviços de saúde são os municípios de Uberlândia e Uberaba que possuem mais de cem estabelecimentos de saúde, enquanto que os municípios de Patos de Minas, Araguari, Ituiutaba e Araxá possuem de 31 a 100 estabelecimentos de saúde [23].

Para o estudo de conjunto de causas capazes de produzir uma doença, é preciso adotar um padrão de classificação de diagnósticos hospitalares. Sabe-se que as classificações são abstrações da realidade e, portanto, possuem limitações. A Classificação Internacional de Doenças (CID) vem sendo utilizada e aplicada há décadas para que leitores e pesquisadores, independente da nacionalidade, possam discutir tendo como base um código previamente conhecido e aceito. As

causas externas vêm se configurando como uma importante causa de mortalidade e morbidade tanto nos países desenvolvidos como nos países em desenvolvimento. Estimou-se que, no ano 2000, em todo o mundo, mais de 1,6 milhões de pessoas morreram como resultado de violência e que menos de 10% dessas mortes ocorreu em países de renda alta [15].

As internações pelo Sistema Único de Saúde (SUS) por causas externas, embora tenham menor tempo médio de permanência do que as internações por causas naturais, tem custo-dia maior e seu gasto médio é também mais elevado [18]. O impacto econômico dos acidentes de trânsito nas regiões urbanas brasileiras foi estudado pelo Instituto de Pesquisa Econômica Aplicada (IPEA) em 2001, no qual foram levados em consideração custos diversificados, com predomínio dos custos com perda de produção, danos a veículos e atendimento médico hospitalar. Os acidentes com vítimas tiveram custo muito maior, apesar da quantidade bem menor de ocorrências, em relação aos acidentes sem vítimas [13].

O Ministério da Saúde implantou a CID-10 para morbidade ambulatorial e hospitalar, a partir de janeiro de 1998, através da portaria n. 1.311 de 12 de setembro de 1997 [2]. Na mesma época, foi implantado no Sistema de Informações Hospitalares do SUS (SIH-SUS), através da Portaria n. 142, de 13 de novembro de 1997, o uso da CID-10 com o registro no campo “diagnóstico principal” do código referente à natureza da lesão provocada pela causa externa que motivou a internação, e no campo “diagnóstico secundário”, do código referente a causa externa que motivou a internação [3]. Desta forma, foi estabelecido o marco legal para a formação de uma base de dados de abrangência nacional sobre internações pelo SUS por causas externas.

Existem poucos estudos sobre a confiabilidade dos dados de internação hospitalar [17]. Entre essas pesquisas, [19] realizou análise de fidedignidade dos dados estatísticos hospitalares, anterior a existência da autorização da internação hospitalar (AIH), disponíveis na Secretaria de Estado da Saúde em 1974, no Vale do Paraíba, utilizando o Boletim CAH 101 - Movimento de Pacientes Saídos. [22] estudaram a confiabilidade dos dados das AIHs na cidade dos dados do campo diagnóstico principal, em nível de capítulo e de agrupamento da Classificação Internacional de Doenças, em Maringá, 1992. Outros estudos realizados com a finalidade de avaliar a qualidade das informações no SIH-SUS foram o de [10], que avaliou a qualidade de informações sobre infarto agudo do miocárdio e o de [12]. Um estudo sobre a qualidade dos dados de tipo de vítima de acidente de transporte foi realizado em Londrina, em 1996 [4].

É fácil entender a classificação de doenças como instrumento estatístico para uso em pesquisas. Assim, como exemplo, na descrição e análise da mortalidade por causas ou o mesmo para diagnósticos de internação hospitalar ou assistência ambulatorial, a Classificação Internacional de Doenças é o instrumento utilizado e por todos conhecida.

Desta maneira, a aplicação da técnica multivariada para a análise das internações hospitalares relacionada aos grupos de causas de doenças, é de suma importância, para extrair mais informações que possam identificar os principais casos de homicídios nas cidades do Triângulo Mineiro – MG.

2. JUSTIFICATIVA

No Século XVIII, particularmente durante sua segunda metade, havia um grande interesse na obtenção de um instrumento estatístico que sistematizasse as causas de morte e que fosse de uso internacional principalmente para possibilitar comparações. Após muitas tentativas conseguiu-se, em 1893, um acordo internacional tendo sido proposta e aprovada uma "Classificação de Causas de Morte" que passou a ser usada por vários países. Aprovou-se, também, uma recomendação para que houvessem revisões periódicas e, no caso, uma periodicidade foi estabelecida como sendo decenal.

Os estudos estatísticos da mortalidade são de extrema importância em demografia e na área da saúde, pois permitirão compreender a magnitude desse fenômeno, bem como o impacto na vida e saúde da população, de modo a auxiliar formulações de políticas públicas e de promoção/prevenção.

Diante disso, este trabalho tem por objetivo retratar a importância de se utilizar a técnica multivariada de Componentes Principais para destacar os principais casos de homicídios, para cada cidade pesquisada, em relação à morbidade hospitalar de grupos de causas da CID10.

3. OBJETIVOS

- Estudar a técnica multivariada de componentes principais.
- Identificar nas cidades do Triângulo Mineiro a classificação de internação hospitalar de doenças semelhantes.
- Identificar, comparar e desenvolver pesquisas sobre os principais grupos de causas de internação hospitalar.

4. METODOLOGIA

Para a realização deste trabalho utilizou-se dados do Sistema de Informações Hospitalares do Sistema Único de Saúde – SIH/SUS, através do site do Departamento de Informática do SUS – DATASUS [9]. As variáveis em estudo são os percentuais das internações hospitalares por grupo de causas, no qual denota-se por Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde – CID10. Os dados são de todas as cidades do Triângulo Mineiro em Minas Gerais, relativos ao ano de 2009.

Na Figura 4.1 apresenta-se todos as categorias que constituem no grupo CID-10 com as respectivas descrições e códigos de cada capítulo:

Capítulo	Códigos	Título
I	A00-B99	Algumas doenças infecciosas e parasitárias.
II	C00-D48	Neoplasmas (tumores).
III	D50-D89	Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários.
IV	E00-E90	Doenças endócrinas, nutricionais e metabólicas.
V	F00-F99	Transtornos mentais e comportamentais.
VI	G00-G99	Doenças do sistema nervoso.
VII	H00-H59	Doenças do olho e anexos.
VIII	H60-H95	Doenças do ouvido e da apófise mastoide.
IX	I00-I99	Doenças do aparelho circulatório.
X	J00-J99	Doenças do aparelho respiratório.
XI	K00-K93	Doenças do aparelho digestivo.
XII	L00-L99	Doenças da pele e do tecido subcutâneo.
XIII	M00-M99	Doenças do sistema osteomuscular e do tecido conjuntivo.
XIV	N00-N99	Doenças do aparelho geniturinário.
XV	O00-O99	Gravidez, parto e puerpério.
XVI	P00-P96	Algumas afecções originadas no período perinatal.
XVII	Q00-Q99	Malformações congênitas, deformidades e anomalias cromossômicas.
XVIII	R00-R99	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte
XIX	S00-T98	Lesões, envenenamentos e algumas outras consequências de causas externas.
XX	V01-Y98	Causas externas de morbidade e de mortalidade.
XXI	Z00-Z99	Fatores que influenciam o estado de saúde e o contato com os serviços de saúde.
XXII	U00-U99	Códigos para propósitos especiais.

Figura 4.1: Morbidade Hospitalar do SUS CID-10.

Primeiramente realizou-se um estudo descritivo, para caracterizar o comportamento dos homicídios no Triângulo Mineiro, quanto a necessidade de internações. Na sequência aplicou-se a técnica estatística multivariada de análise de componentes principais, no qual permite

comparar os diferentes perfis entre as cidades da região em que está sendo estudada. Os procedimentos estatísticos executados neste trabalho foram realizados com o auxílio do software SPSS.

O método de análise de componentes principais, é utilizado com o objetivo de resumir o padrão de correlação entre as variáveis em estudo, e chegar a um conjunto de variáveis que sejam não correlacionadas entre si, obtendo-se geração de índices e agrupamento de indivíduos, segundo [7].

A análise de componentes principais está relacionada com a explicação da estrutura de covariância por meio de poucas combinações lineares das variáveis originais em estudo, ou seja, procura explicar a estrutura de variância-covariância da matriz de dados através de combinações lineares não correlacionadas das p variáveis originais [5].

Este método consiste em transformar um conjunto de p variáveis originais $X_1, X_2, \dots, X_p$ em um novo conjunto de variáveis, $Y_1, Y_2, \dots, Y_p$ denominados componentes principais. De modo que Y_1 explica a maior parcela da variabilidade total dos dados, Y_2 explica a segunda maior parcela e, assim sucessivamente, até que toda a variância tenha sido explicada [21].

Em [20] define-se o primeiro componente principal (Y_1) de um conjunto de p variáveis, $X_1, X_2, \dots, X_p$, contidas no vetor $X' = (X_1, X_2, \dots, X_p)$ como a combinação linear:

$$Y_1 = b_{11}X_1 + b_{21}X_2 + \dots + b_{p1}X_p = b'_1X$$

cujos coeficientes b_{i1} são elementos do vetor característico b_1 , associado a maior raiz característica (λ_1) da matriz de covariância amostral, S , das variáveis X_i 's. Os autovalores (ou raízes características) ordenados, ou seja, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, são as variâncias amostrais dos componentes principais

Os componentes principais dependem somente da matriz de covariâncias (S) ou da matriz de correlação (R). Na maioria das aplicações tem sido utilizado a matriz de correlação (R), pois, as variáveis que apresentam valores em escalas diferentes, devem ser padronizadas. As variáveis padronizadas, $Z_1, \dots, Z_p$, são representadas da seguinte forma:

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}}; Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}}; \dots; Z_p = \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}}$$

A contribuição de cada componente na explicação da variação total, é representada por:

$$Contr(Y_i) = \frac{Var(Y_i)}{\sum_{i=1}^p Var(Y_i)} = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i}; \text{ em que } i = 1, \dots, p.$$

Segundo [16], pode-se utilizar diversos critérios de escolha a serem utilizados na análise, na possibilidade de interpretação e na significância operacional das soluções dos componentes principais, ou seja, no processo de seleção do número de componentes principais a serem retidos.

Dentre os principais critérios, tem-se o Gráfico Scree, no qual foi proposta por [8], onde envolve desenhar a variância explicada por cada componente principal, na ordem do maior para o menor. Então, busca-se um “ângulo” na curva – isto é, um ponto após o qual os autovalores remanescentes declinam de modo aproximadamente linear – e retêm-se somente

aqueles componentes que estão acima do ângulo no gráfico.

Quando o Gráfico de Scree não oferece diagnóstico, a Regra de Kaiser pode ser conveniente. O critério foi proposto por [14], no qual recomendou a retenção dos componentes principais com autovalores que excedam a unidade (se tratando de variáveis padronizadas), esse critério corresponde à exclusão de componentes com autovalores inferiores a um, segundo [6].

De acordo com [11], para verificar a adequação da análise de componentes principais, tem-se que examinar a matriz de correlações do teste de esfericidade de Bartlett, cujo requisito é que as variáveis apresentem normalidade multivariada. Esse teste é utilizado para testar a hipótese de que a matriz de correlações pode ser a matriz identidade com determinante igual a um. Se a hipótese nula for rejeitada, significa que há indícios de que existem correlações significativas entre as variáveis originais.

Com o objetivo de transformar os coeficientes dos componentes principais em uma estrutura simplificada e de fácil interpretação, aplica-se o método de rotação de fatores ortogonais. Sendo assim, o critério Varimax proposto por Kaiser em 1958, reduz o número de variáveis que têm altas cargas em um fator, cuja finalidade é maximizar a variação entre os pesos de cada componente principal.

5. RESULTADOS

Os dados referentes à distribuição percentual das internações hospitalares por grupo de causas, foram retiradas em [9] no Caderno do Estado de Minas Gerais, no qual retirou-se as informações somente das 66 cidades do Triângulo Mineiro, e as variáveis em estudo foram as 21 causas de internações hospitalares, porém, retirou-se a variável *XX* - Causas externas de morbidade e mortalidade, por apresentar valores nulos no banco de dados, restando portanto 20 variáveis para a análise¹.

Na Tabela 5.1 são apresentados os valores mínimo e máximo, a média e o desvio padrão das variáveis em estudo no período avaliado. Observa-se nos resultados obtidos na Tabela 5.1 e na Figura 5.1, que a internação por gravidez e por doenças respiratórias apresentam os maiores desvios padrão, iguais a 7,473 e 6,714, respectivamente. Isso significa que essas variáveis são as de maior variabilidade em relação as suas respectivas médias, e esse fato já é uma indicação de que essas variáveis poderão ter uma maior influência nos componentes principais, quando comparadas com as demais.

¹As análises foram realizadas com o auxílio do software SPSS.

Tabela 5.1: Análise descritiva das causas de internações nas cidades do Triângulo Mineiro em porcentagem no ano de 2009.

Causas	N	Mínimo	Máximo	Média	Desvio Padrão
I	66	1,4	28,1	6,017	4,180
II	66	1,4	14,9	5,438	3,016
III	66	0	4,9	0,964	0,863
IV	66	0	16,6	2,565	2,397
V	66	0,1	8,4	2,927	2,072
VI	66	0	5,9	1,274	0,921
VII	66	0	1	0,245	0,274
VIII	66	0	1,4	0,162	0,274
IX	66	3,6	28,3	12,547	4,581
X	66	4,2	36	15,385	6,714
XI	66	4,5	18,8	9,806	2,938
XII	66	0	2,8	0,955	0,677
XIII	66	0	3,5	1,267	0,828
XIV	66	0,6	17,6	7,477	3,217
XV	66	3	42,3	19,355	7,473
XVI	66	0	6,8	1,561	1,232
XVII	66	0	2,2	0,591	0,557
XVIII	66	0	5,6	1,267	1,062
XIX	66	4,1	18,9	9,147	2,951
XXI	66	0	5,8	1,014	1,210
Total	1320	22,9	269,8	99,962	47,435

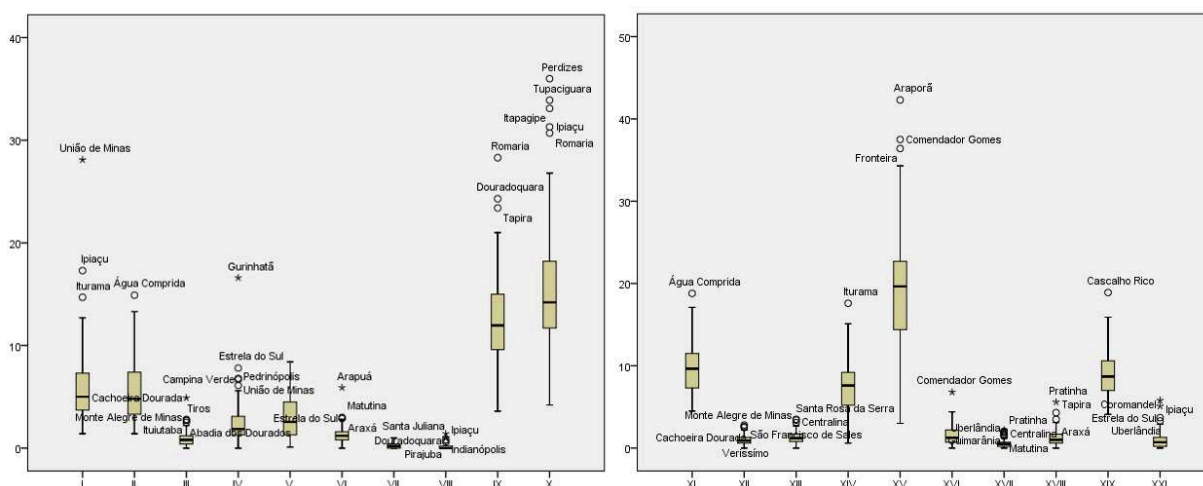


Figura 5.1: Boxplot das causas de internações hospitalares nas cidades do Triângulo Mineiro do ano 2009.

Com base nos resultados obtidos pela técnica dos componentes principais, os respectivos autovalores e porcentagens da variância explicada por cada um estão apresentados na Tabela 5.2. Os quatro primeiros componentes principais foram responsáveis por 82,080% da variação total sobre as internações hospitalares, em que o primeiro componente (CP1) foi responsável por 44,414%, o segundo (CP2) por 17,129%, o terceiro (CP3) por 12,066%, e o quarto (CP4) por 8,471% das variações dos dados.

Nesses resultados, os primeiros quatro componentes principais têm autovalores maiores do que 1. O gráfico Scree (Figura 5.2) mostra que os autovalores começam a formar uma linha reta após o quarto componente principal. Se 82,080% é uma quantidade adequada de variação explicada nos dados, usa-se os primeiros quatro componentes principais.

Tabela 5.2: Componentes Principais (CP's), autovalores (λ_i) e porcentagem da variância explicada e proporção acumulada (%) pelos componentes.

Componente Principal	Autovalores	Proporção (%)	Proporção Acumulada (%)
CP1	85,941	44,414	44,414
CP2	33,145	17,129	61,543
CP3	23,348	12,066	73,61
CP4	16,391	8,471	82,08
CP5	9,028	4,666	86,746
CP6	6,307	3,26	90,006
CP7	5,56	2,873	92,879
CP8	4,132	2,135	95,014
CP9	3,683	1,904	96,918
CP10	1,796	0,928	97,846
CP11	1,249	0,646	98,492
CP12	0,862	0,445	98,937
CP13	0,703	0,363	99,301
CP14	0,438	0,226	99,527
CP15	0,386	0,199	99,726
CP16	0,243	0,126	99,852
CP17	0,179	0,092	99,944
CP18	0,07	0,036	99,98
CP19	0,036	0,019	99,999
CP20	0,001	0,001	100

Com a seleção de quatro componentes principais, a redução da dimensão de 20 variáveis originais para quatro componentes principais é bastante razoável. Portanto decidiu-se utilizar unicamente os quatro primeiros componentes principais para a composição das equações 5.1, 5.2, 5.3 e 5.4. De acordo com a Tabela 5.3 a seguir, tem-se que a equação para cada componente segue da seguinte maneira:

$$CP\ 1 = -0,813X_{10} + 0,893X_{15} \quad (5.1)$$

$$CP\ 2 = -0,508X_{10} + 0,573X_{19} \quad (5.2)$$

$$CP\ 3 = -0,764X_9 \quad (5.3)$$

$$CP\ 4 = 0,600X_2 \quad (5.4)$$

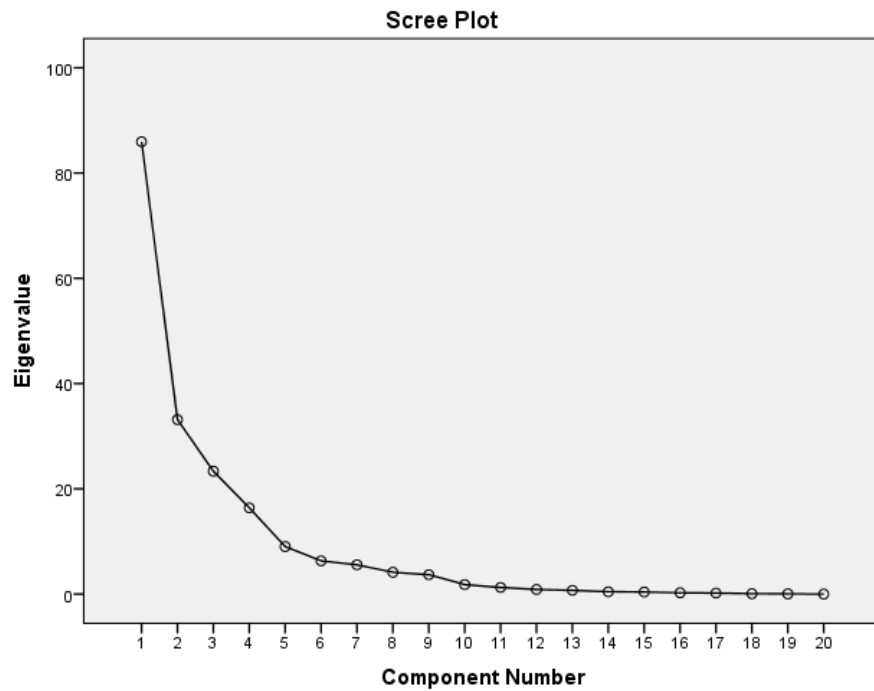


Figura 5.2: O Screeplot dos autovalores dos componentes principais.

Tabela 5.3: Coeficientes de ponderação das características com os quatro primeiros componentes principais.

Causas de Interações	CP1	CP2	CP3	CP4
I	-0,597	-0,205	0,455	-0,467
II	0,340	0,317	0,170	0,600
III	-0,045	0,155	-0,031	-0,027
IV	-0,227	-0,049	0,162	-0,355
V	0,230	0,329	-0,077	-0,020
VI	-0,011	0,079	0,124	0,259
VII	0,156	0,121	0,191	0,288
VIII	-0,029	-0,043	-0,199	-0,161
IX	-0,253	0,427	-0,764	-0,388
X	-0,813	-0,508	-0,190	0,200
XI	0,122	0,439	0,066	0,303
XII	0,248	0,246	0,104	0,308
XIII	0,331	0,070	0,192	0,224
XIV	-0,287	0,212	0,680	-0,289
XV	0,893	-0,427	-0,089	-0,088
XVI	0,395	-0,404	-0,091	0,056
XVII	0,160	0,132	0,266	0,365
XVIII	0,155	0,257	-0,079	-0,061
XIX	0,348	0,573	-0,031	0,393
XXI	-0,118	0,009	0,202	0,088

Para o primeiro componente principal, de acordo com o a equação (5.1) e a Tabela 5.4, teve um destaque positivo para a variável XV, e negativo para a variável X. Portanto, conclui-se

que, quanto menor for o score da $CP1$, maior será a quantidade de casos de internação devido a doenças do aparelho respiratório, e quanto maior for o score da $CP1$, maior será quantidade de casos de internação por gravidez, sendo assim pode ser denominado para o primeiro componente, morbidade por doenças respiratórias e por gravidez.

No segundo componente principal, conforme a equação (5.2) e a Tabela 5.4, teve destaque positivamente para a variável XIX, e negativamente para a variável X. Portanto pode-se concluir que, quanto menor for o score da $CP2$, maior será o número de casos de internação por doenças respiratórias, e quanto maior for o score da $CP2$, maior será o número de casos por causas externas envolvendo lesões e envenenamento, desta forma, para o segundo componente denomina-se por morbidade em doenças respiratórias e por lesões, envenenamentos e algumas outras consequências de causas externas.

Já no terceiro componente principal, pela equação (5.3) e a Tabela 5.4, houve um destaque negativo para a variável IX. Então, sabe-se que quanto menor for o score da $CP3$, maior será a quantidade de casos para doenças do aparelho circulatório, logo, pode ser denominado para o terceiro componente, morbidade em doenças por aparelho circulatório.

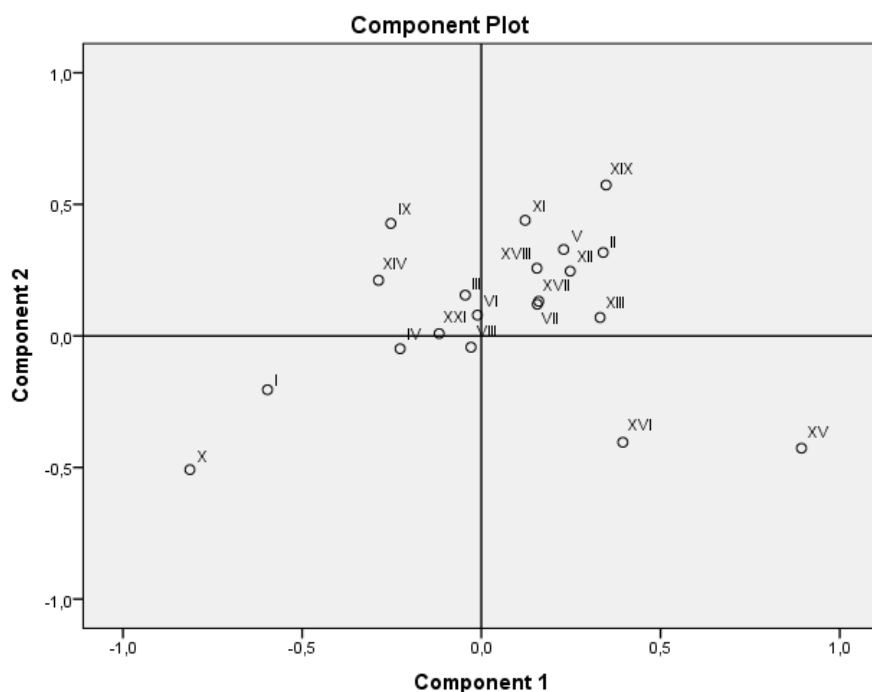
Por fim, no quarto componente principal, conforme a equação (5.4) e a Tabela 5.4, houve um destaque positivo para a variável II. Sendo assim, quanto maior for o score para a $CP4$, maior será a quantidade de casos por neoplasmas, então, denomina-se para o quarto componente, morbidade por doenças de tumores.

As variáveis que não entraram nas quatro componentes principais não contribuíram para a explicação da variabilidade total, pois explicaram muito pouco para a análise, portanto não foram consideradas.

De acordo com a Figura 5.3, apresenta-se o Gráfico Biplot da $CP1$ x $CP2$ em que, as duas primeiras componentes principais explicam cerca de 61,5% da variabilidade total sobre os percentuais das causas de internações hospitalares do SUS das cidades do Triângulo Mineiro. Pode-se concluir pelo gráfico e pelas equações (5.1) e (5.2) que, a variável XV – Gravidez, parto e puerpério, apresenta-se o maior número em porcentagem de causas de internações hospitalares pela $CP1$, e a variável XIX – Lesões, envenenamentos e algumas outras consequências de causas externas, apresenta-se o maior número em porcentagem de causas de internações hospitalares pela $CP2$. A variável X – Doenças do aparelho respiratório, obteve o menor número de casos de internações pela $CP1$ e a $CP2$.

Tabela 5.4: Interpretação de cada componente.

	CP1	CP2	CP3	CP4
Quanto menor for o score para a CP, maior será a quantidade de casos de internação por	Doenças do aparelho respiratório	Doenças do aparelho respiratório	Doenças do aparelho circulatório	
Quanto maior for o score para a CP, maior será a quantidade de casos de internação por	Gravidez, parto e puerpério	Lesões, envenenamentos e algumas outras consequências de causas externas		Neoplasmas [tumores]
Denomina-se por morbidade em	Doenças respiratórias e por gravidez	Doenças respiratórias e por lesões, envenenamentos e algumas outras consequências de causas externas	Doenças do aparelho circulatório	Tumores

Figura 5.3: Biplot $CP\ 1$ x $CP\ 2$ do percentual das causas de internações hospitalares.

Foi aplicado o Teste de Levene para averiguar a homogeneidade das variâncias, e o Teste t de Student para fazer a comparação de dois grupos, ou seja, denotou-se por grupo 1, cidades

grandes com mais de 30 mil habitantes localizadas no Triângulo Mineiro, nos quais são Uberlândia, Uberaba, Patos de Minas, Araguari, Ituiutaba, Araxá, Patrocínio, Frutal, Monte Carmelo, Iturama, São Gotardo e Carmo do Paranaíba. E as outras cidades localizadas no Triângulo Mineiro foram denotadas por grupo 2, das quais, 54 cidades pequenas possuem menos de 30 mil habitantes.

Na Tabela 5.5 apresenta o teste de normalidade dos dados, em que obteve valores de p-valor grande para cada componente principal, ou seja, foram maiores que 5% de significância, logo pela hipótese nula do teste os dados seguem normalidade, portanto segue um modelo paramétrico.

$$\begin{cases} H_0 : \text{Os dados seguem uma distribuição normal} \\ H_1 : \text{Os dados não seguem uma distribuição normal} \end{cases}$$

Tabela 5.5: Teste de Normalidade.

	CP1	CP2	CP3	CP4
Absoluto	0,352	0,185	0,380	0,241
Positivo	0,157	0,185	0,148	0,213
Negativo	-0,352	-0,148	-0,380	-0,241
Kolmogorov-Smirnov Z	1,102	0,580	1,190	0,754
Asymp. Sig. (Bilateral)	0,176	0,889	0,118	0,620

Conforme a Tabela 5.6, aplicou-se o Teste de Levene, teste paramétrico aplicado para duas amostras independentes em que permite averiguar a homogeneidade das variâncias, pode-se concluir que as variâncias são diferentes nos dois grupos tratados, uma vez que a significância associada ao teste é inferior a 0,05. Não assumindo a homogeneidade das variâncias, utilizou-se o Teste t de Student, em que obteve valores de p-valor maiores que 5% de significância, logo não é rejeitada a hipótese nula de que não há diferença entre os grupos tratados.

Portanto, não há diferença entre as cidades grandes com as cidades pequenas em relação as causas de internações hospitalares do SUS.

$$\begin{cases} H_0 : \text{Não há diferença entre os grupos} \\ H_1 : \text{Há diferença entre os grupos} \end{cases}$$

Tabela 5.6: Teste de Levene para igualdade de variâncias e teste t de Student para igualdade de médias.

		Teste de Levene para Igualdade de Variâncias		Teste t para igualdade de médias				
		F	Sig.	t	df	Sig. (Bilateral)	Média das diferenças	Desvio padrão
CP1	Assumindo variâncias iguais	4.582	0.036	-0.258	64.000	0.798	-0.874	3.391
	Assumindo variâncias diferentes			-0.399	36.864	0.692	-0.874	2.188
CP2	Assumindo variâncias iguais	2.792	0.100	0.695	64.000	0.489	1.000	1.437
	Assumindo variâncias diferentes			1.039	33.290	0.306	1.000	0.962
CP3	Assumindo variâncias iguais	4.193	0.045	-0.620	64.000	0.538	-0.695	1.122
	Assumindo variâncias diferentes			-0.960	36.870	0.343	-0.695	0.724
CP4	Assumindo variâncias iguais	5.154	0.027	-0.500	64.000	0.619	-0.291	0.581
	Assumindo variâncias diferentes			-0.692	27.447	0.495	-0.291	0.420

6. CONCLUSÕES

De acordo com os resultados, a Análise de Componentes Principais aplicada foi efetiva pelo fato de ter reduzido a quantidade de variáveis de 20 para quatro componentes que explicam parte satisfatória dos dados originais. A redução se dá pelo fato de dezesseis variáveis possuírem correlação entre si ou pouca variabilidade, desta forma atribuindo redundância aos dados analisados.

O principal objetivo da análise foi cumprido: redução do número de variáveis sem perda significativa de informação sobre a variabilidade do conjunto de variáveis originais permitindo a interpretação dos dados de forma simplificada.

Pelo Teste t de Student, não houve diferença significativa nas causas de internações hospitalares entre os grupos tratados, ou seja, entre as cidades grandes com mais de 30 mil habitantes com as cidades pequenas com menos de 30 mil habitantes.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] *Notícia original inacessível em 2012-4-1: «Investimentos em Minas». Governo de Minas Gerais, Portal.*
- [2] *Ministério da Saúde. Portaria n. 1311, de 12 de setembro de 1997. Define a competência janeiro de 1998, para que a CID-10 vigore, em todo o território nacional, em morbidade hospitalar e ambulatorial.* Diário Oficial da União, Brasília, 178^a ed., 1997a.
- [3] *Ministério da Saúde. Dispõe sobre o preenchimento de Autorização de Internação Hospitalar - AIH, em casos com quadro compatível com causas externas.* Diário Oficial da União. Diário Oficial da União, Brasília, 222^a ed., 1997b.
- [4] ANDRADE SM, M. J. M.: *Acidentes de transporte terrestre em cidade da Região Sul do Brasil: avaliação da cobertura e qualidade dos dados*, vol. 17. Cad. Saúde Pública, 6^a ed., 2001.
- [5] Anselmo, P.: *Componentes Principais*. <https://docs.ufpr.br/~soniaisoldi/ce090/5COMPONENTESPRINCIPAIS.pdf>, acessado em 28/05/2018.
- [6] ARTES, R.: *Aspectos estatísticos da análise fatorial de escalas de avaliação*, vol. 25. Revista de Psiquiatria Clínica, São Paulo, 1998.
- [7] BARROSO, L.P.; ARTES, R.: *Análise multivariada*. Lavras: UFLA, 2003.
- [8] CATTELL, R. B.: *The scree test for the number of factors*, vol. 1. Multivariate Behavioral Research, 1996.
- [9] DATASUS: *Departamento de informática do SUS*. <https://http://datasus.saude.gov.br/>, acessado em 28/05/2018.
- [10] Escosteguy CC, Portela MC, M. R. V. M.: *O sistema de informações hospitalares e a assistência ao infarto agudo do miocárdio*, vol. 36. Rev. Saúde Pública, 4^a ed., 2002.
- [11] FAVERO, L. P.; BELFIORE, P. S. F. L. C. B. L.: *Análise de dados – modelagem multivariada para tomada de decisão*. Elsevier Ltda, Rio de Janeiro, 8^a ed., 2009.
- [12] Fernandes W, Laurenti R, B. C. C. M.: *Accuracy of the information about diagnosis in the discharges in public hospital in Brasil. In: Meeting of Heads of WHO Collaborating Centres for the Classification of Diseases*. Genebra, Rio de Janeiro, 2000.

- [13] [IPEA]: *Instituto de Pesquisa Econômica Aplicada. Impactos sociais e econômicos dos acidentes de trânsito nas aglomerações urbanas*. Brasília (DF), 2003.
- [14] KAISER, H. F.: *The application of electronic computers to factor analysis*. In *Symposium on the Application of Computers to Psychological Problems*. American Psychological Association, 1959.
- [15] KRUG EG, DAHLBERG LL, M. J. Z. A. L. R. e.: *World report on violence and health*. Geneva: World Health Organization, 2002.
- [16] LATTIN, J., C. J. G. P.: *Análise de dados multivariados*. Cengage Learning, São Paulo, 2011.
- [17] Laurenti R, Mello Jorge MHP, G. S.: *A confiabilidade dos dados de mortalidade e morbidade por doenças crônicas não-transmissíveis*, vol. 9. Ciênc. Saúde Coletiva, 4^a ed., 2004.
- [18] Mello Jorge MHP, K. M.: *Gastos governamentais do SUS com internações hospitalares por causas externas: análise no Estado de São Paulo*, vol. 7. Rev. Bras Epidemiol, 2^a ed., 2004.
- [19] ML, L.: *Análise da fidedignidade dos dados estatísticos hospitalares disponíveis na Secretaria de Estado da Saúde de São Paulo*, vol. 12. Rev. Saúde Pública, 2^a ed., 1978.
- [20] MORRISON, D.: *Multivariate statistical methods*. Singapore: McGraw Hill, 2^a ed., 1976.
- [21] STRAPASSON, E.; VENCovsky, R. B. L.: *Seleção de Descritores na Caracterização de Germoplasma de Paspalum sp. por meio de Componentes Principais*, vol. 29. Revista Brasileira Zootecnia, 2^a ed., LATTIN 2011.
- [22] Veras CMT, M. M.: *A confiabilidade dos dados nos formulários de autorização de internação hospitalar (AIH)*, vol. 10. Cad Saúde Pública, Rio de Janeiro, 3^a ed., 1994.
- [23] Wikipedia: *Mesorregião do Triângulo Mineiro e Alto Paranaíba*. https://pt.wikipedia.org/wiki/Mesorregi%C3%A3o_do_Tri%C3%A2ngulo_Mineiro_e_Alto_Parana%C3%ADba, acessado em 28/08/2018.