

ESTUDO DA MORBIDADE HOSPITALAR DO SUS PELA ANÁLISE DE COMPONENTES PRINCIPAIS

Andriely Antunes Martins¹

Priscila Neves Faria²

Lúcio Borges de Araújo³

Mirian Fernandes Carvalho Araújo⁴

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 óbitos 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 às 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.

Resumen. Estudio de morbilidad hospitalaria en sus mediante análisis de principales componentes. Los estudios estadísticos de mortalidad son extremadamente importantes en demografía y salud, ya que nos permitirán comprender la magnitud de este fenómeno, así como el impacto en la vida y la salud de la población, para ayudar a formular políticas públicas y promoción / prevención. El objetivo principal de este estudio fue destacar los principales casos de homicidios para cada ciudad encuestada, en relación con la morbilidad hospitalaria de los grupos causantes de ICD10. Por lo tanto, utilizamos la técnica de Análisis de Componentes Principales y la Prueba U de Mann-Whitney para verificar si había una diferencia entre las grandes ciudades con más de 30 mil habitantes y las ciudades pequeñas con menos de 30 mil habitantes en relación con las causas de Admisiones hospitalarias del SUS. Como resultado, obtuvo cuatro componentes principales en el estudio, y la prueba t acepta la hipótesis nula de que no hay diferencia entre los grupos tratados con respecto a las causas de las hospitalizaciones.

Palabras clave: ICD10; Mortalidad; Salud; Multivariante.

Abstract. Study of hospital morbidity in sus by analysis of main components. The statistical studies of mortality are extremely important in demography and health, as they will allow us to understand the magnitude of this phenomenon, as well as the impact on the life and health of the population, in order to help formulate public policies and promotion / prevention. The main objective of this study was to highlight the main cases of homicides for each city surveyed, in relation to the hospital morbidity of ICD cause groups 10. Therefore, we used the Principal Component Analysis Technique and the Mann-

¹ Graduada em Estatística pela Universidade Federal de Uberlândia. E-mail: andriely_95@hotmail.com.

² Graduada em Licenciatura Plena em Matemática e Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Viçosa e Doutora em Estatística e Experimentação Agronômica pela ESALQ/USP. Docente da Faculdade de Matemática da Universidade de Uberlândia. E-mail: priscilaneves@ufu.br.

³ Graduado em Matemática pela Universidade Federal de Uberlândia, Mestre e Doutor em Estatística e Experimentação Agronômica pela ESALQ/USP. Docente da Faculdade de Matemática da Universidade de Uberlândia. E-mail: lucio.araujo@ufu.br.

⁴ Graduada em Matemática e Especialista em Estatística Aplicada pela Universidade Federal de Uberlândia, Mestre e Doutora em Estatística e Experimentação Agronômica pela ESALQ/USP. Docente da Faculdade de Matemática da Universidade Federal de Uberlândia. E-mail: mirian@ufu.br.

Whitney U Test to verify if there was a difference between large cities with more than 30 thousand inhabitants, and small cities with less than 30 thousand inhabitants in relation to the causes of SUS hospital admissions. As a result, it 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 hospitalizations.

Keywords: ICD10; Mortality; Causes; Multivariate.

1 Introdução

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. 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. As causas externas vêm se configurando como importantes causas 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 ocorreram em países de renda alta (KRUG; DAHLBERG, 2002).

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 (JORGE; KOIZUMI, 2004). 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 (IPEA, 2003).

O Ministério da Saúde implantou a Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID10) para morbidade ambulatorial e hospitalar, a partir de janeiro de 1998, através da portaria n. 1.311 de 12 de setembro de 1997 (MINISTÉRIO DA SAÚDE, 1997a). 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 à causa externa que motivou a internação (MINISTÉRIO DA SAÚDE, 1997b). 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 (LAURENTI et al., 2004). Entre essas pesquisas, 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 (LEBRÃO, 1978). 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 (MORRISON, 1976).

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 óbitos nas cidades do Triângulo Mineiro – MG.

Diante disso, este trabalho teve por objetivo aplicar a técnica multivariada de componentes principais para destacar os principais casos de óbitos, para cada cidade pesquisada, em relação à morbidade hospitalar de grupos de causas da CID10, identificar nas cidades do Triângulo Mineiro a classificação de internação hospitalar de doenças semelhantes e promover ideias para o desenvolvimento de pesquisas sobre os principais grupos de causas de internação hospitalar.

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 (DATA/SUS). 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), cujas categorias e respectivas descrições e códigos de cada capítulo estão no Quadro 1. Os dados são de todas as cidades do Triângulo Mineiro em Minas Gerais, relativos ao ano de 2009.

Quadro 1 - Morbidade Hospitalar do SUS CID10

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 hematopoiéticos 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.

Fonte: Elaborado pelos autores (2021).

Primeiramente, realizou-se um estudo descritivo para caracterizar o comportamento dos óbitos no Triângulo Mineiro, quanto à 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. Optou-se pela análise de componentes principais, pois com esta técnica é possível reduzir o número de variáveis em estudo, tornando a interpretação mais fácil. Os procedimentos estatísticos executados neste trabalho foram realizados com o auxílio do software SPSS.

2 Análise de Componentes Principais

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 (BARROSO; ARTES, 2003). Ela 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.

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 (STRAPASSON; VENCOVSKY, 2011).

Morrison (1976) define 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:

$$Y_1 = b_{11}X_1 + b_{21}X_2 + \dots + b_{p1}X_p = b_1'X$$

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 utilizada 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:

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

Podem-se utilizar diversos critérios de escolha na análise para possibilitar a interpretação e na significância operacional das soluções dos componentes principais, ou seja, no processo de seleção do número de componentes principais a serem retidos (LATTIN, 2011). Dentre os principais critérios, tem-se o Gráfico *Scree plot*, que envolve desenhar a variância explicada por cada componente principal, na ordem do maior para o menor (CATTEL, 1996). Então, busca-se um “ângulo” na curva – isto é, um ponto após o qual os autovalores remanescentes declinam de modo aproximadamente linear – e são retidos somente aqueles componentes que estão acima do ângulo no gráfico.

Quando o Gráfico *Scree plot* não oferece diagnóstico, a Regra de Kaiser pode ser conveniente. O critério proposto por (KAISER, 1959) 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 (ARTES, 1998).

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 (FAVERO; BELFIORE, 2009). 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.

3 Resultados

Os dados referentes à distribuição percentual das internações hospitalares por grupo de causas foram obtidos pelo DATASUS 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 o estudo.

Na Tabela 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 1 e no Gráfico 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 às 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.

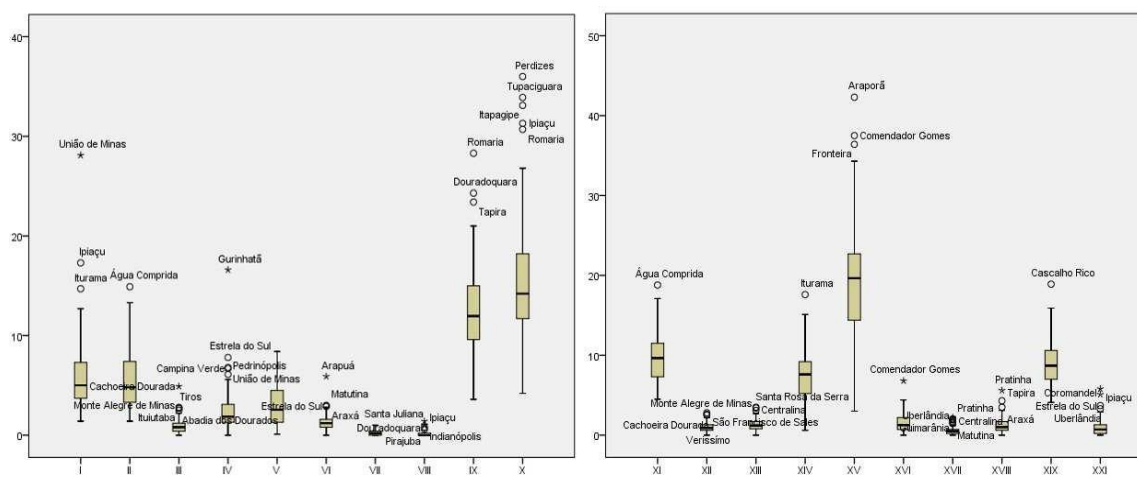
Tabela 1 - Análise descritiva das causas de internações nas cidades do Triângulo Mineiro em percentagem no ano de 2009

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	0,0	4,9	0,964	0,863
IV	66	0,0	16,6	2,565	2,397
V	66	0,1	8,4	2,927	2,072
VI	66	0,0	5,9	1,274	0,921
VII	66	0,0	1,0	0,245	0,274
VIII	66	0,0	1,4	0,162	0,274

IX	66	3,6	28,3	12,547	4,581
X	66	4,2	36,0	15,385	6,714
XI	66	4,5	18,8	9,806	2,938
XII	66	0,0	2,8	0,955	0,677
XIII	66	0,0	3,5	1,267	0,828
XIV	66	0,6	17,6	7,477	3,217
XV	66	3,0	42,3	19,355	7,473
XVI	66	0,0	6,8	1,561	1,232
XVII	66	0,0	2,2	0,591	0,557
XVIII	66	0,0	5,6	1,267	1,062
XIX	66	4,1	18,9	9,147	2,951
XXI	66	0,0	5,8	1,014	1,210
Total	1320	22,9	269,8	99,962	47,435

Fonte: Elaborado pelos autores (2021).

Gráfico 1 – Boxplot das causas de internações hospitalares nas cidades do Triângulo Mineiro em 2009



Fonte: Elaborado pelos autores (2021).

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 2. Os quatro primeiros componentes principais foram responsáveis por 82,08% 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.

Tabela 2 - Componentes principais (CP's), autovalores (λ_i) e porcentagem da variância explicada e proporção acumulada (%) pelos componentes.

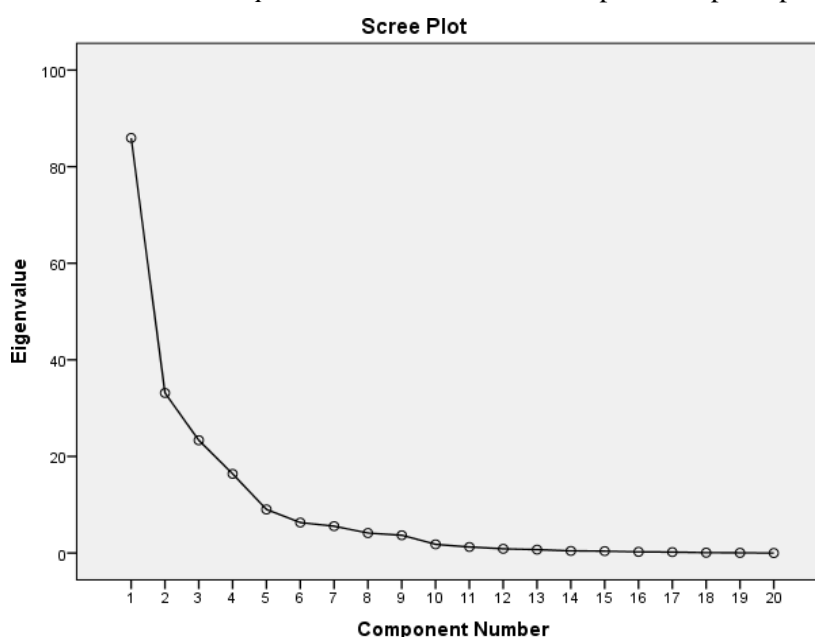
Componente principal	Autovalores	Proporção (%)	Proporção Acumulada (%)
CP1	85,941	44,414	44,414
CP2	33,145	17,129	61,543

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

Fonte: Elaborado pelos autores (2021).

Nesses resultados, os primeiros quatro componentes principais têm autovalores maiores do que 1. O gráfico *Scree plot* (Gráfico 2) mostra que os autovalores começam a formar uma linha reta após o quarto componente principal. Considerando que 82,08% é uma quantidade adequada de variação explicada nos dados, usa-se os primeiros quatro componentes principais.

Gráfico 2 - O *Scree plot* dos autovalores dos componentes principais



Fonte: Elaborado pelos autores (2021).

Com a seleção de quatro componentes principais, a redução da dimensão de 20 variáveis para quatro 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 3, tem-se que a equação para cada componente segue da seguinte maneira:

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

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

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

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

Tabela 3 - Coeficientes de ponderação das características com os quatro primeiros componentes principais.

Causas de Internações	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

Fonte: Elaborado pelos autores (2021).

Para o primeiro componente principal, de acordo com o a equação (5.1) e o Quadro 2, houve 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.

Quadro 2 - 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	
	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

Fonte: Elaborado pelos autores (2021).

No segundo componente principal, conforme a equação (5.2) e o Quadro 2, houve destaque positivo para a variável XIX e negativo 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.

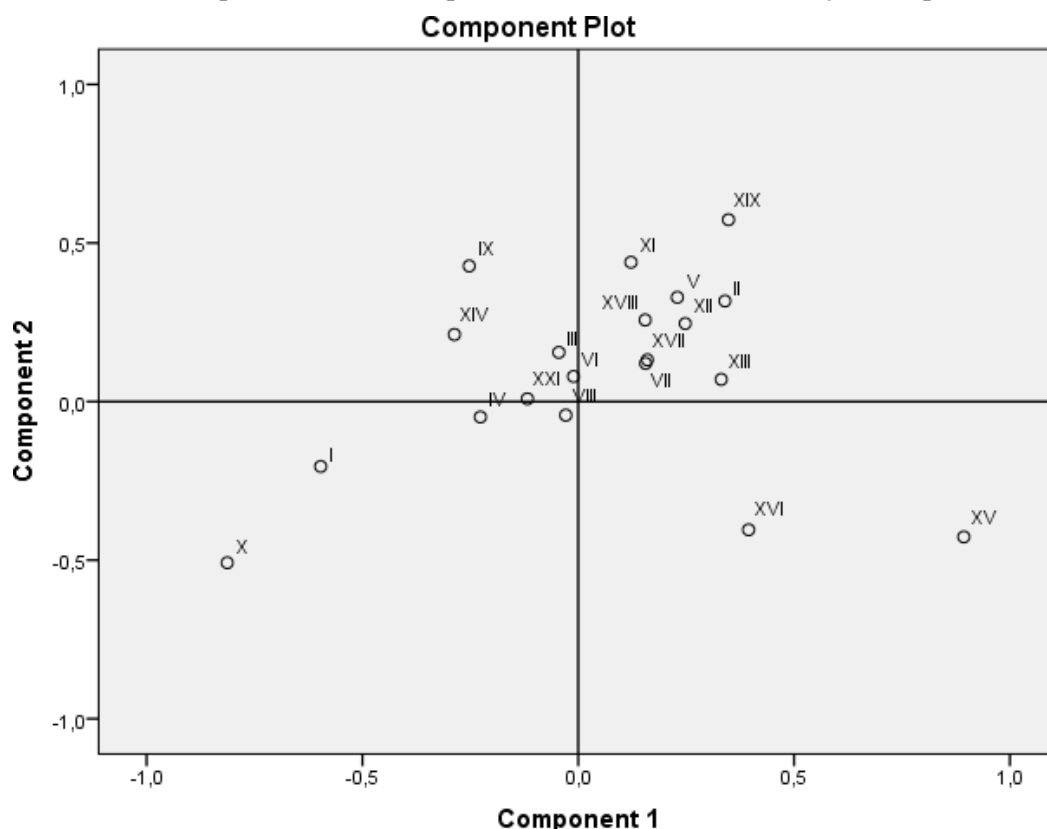
Já no terceiro componente principal, pela equação (5.3) e o Quadro 2, houve um destaque negativo para a variável IX. Então, sabe-se que quanto menor for o score da CP3, maior será a quantidade de casos para doenças do aparelho circulatório, logo, pode ser denominado para o terceiro componente, morbidade em doenças por aparelho circulatório.

Por fim, no quarto componente principal, conforme a equação (5.4) e o Quadro 2, 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.

No Gráfico 3, apresenta-se o 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 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 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.

Gráfico 3 - Biplot CP1 x CP2 do percentual das causas de internações hospitalares



Fonte: Elaborado pelos autores (2021).

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. Denotou-se por grupo 1, as cidades grandes com mais de 30 mil habitantes localizadas no Triângulo Mineiro (Uberlândia, Uberaba, Patos de Minas, Araguari, Ituiutaba, Araxá, Patrocínio, Frutal, Monte Carmelo, Iturama, São Gotardo e Carmo do Paranaíba). Já as outras cidades localizadas no Triângulo Mineiro foram denotadas por grupo 2, das quais, 54 pequenas cidades com menos de 30 mil habitantes.

A Tabela 4 apresenta o teste de normalidade dos dados, em que obteve altos valores *p* para cada componente principal, ou seja, foram maiores que 5% de significância; logo pela hipótese nula do teste, os dados seguem normalidade; um modelo paramétrico, portanto.

Tabela 4 - 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
Valor p	0,176	0,889	0,118	0,620

Fonte: Elaborado pelos autores (2021).

Foi aplicado o teste de Levene, teste paramétrico aplicado para duas amostras independentes, que permite averiguar a homogeneidade das variâncias. A conclusão foi que as variâncias são diferentes nos dois grupos tratados, uma vez que a significância associada ao teste é inferior a 5%. Não assumindo a homogeneidade das variâncias, utilizou-se o teste t de Student, em que obteve valores p 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 às causas de internações hospitalares do SUS.

4 Considerações finais

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. Com isso sugere-se aos gestores municipais que estejam atentos às morbidades hospitalares do SUS CID-10, pois independente do porte da cidade, os índices são os mesmos. Com isso, uma proposta é que ocorra investimentos em serviços de atenção primária e secundária à saúde.

5 Referências

ARTES, R. Aspectos estatísticos da análise fatorial de escalas de avaliação. **Revista de Psiquiatria Clínica**, vol. 25(5). São Paulo, 1998, p. 223-228.

BARROSO, L.P.; ARTES, R.: **Análise multivariada**. Lavras: UFLA, 2003.

FAVERO, L. P.; BELFIORE, P. S. F. L. C. B. L.: **Análise de dados** – modelagem multivariada para tomada de decisão, 8^a ed. Rio de Janeiro: Elsevier, 2009.

IPEA - Instituto de Pesquisa Econômica Aplicada. **Impactos sociais e econômicos dos acidentes de trânsito nas aglomerações urbanas**. Brasília: IPEA, 2003.

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.

KRUG, E.G.; DAHLBERG L.L. **World report on violence and health**. Geneva: World Health Organization, 2002.

LATTIN, J., C. J. G. P. **Análise de dados multivariados**. São Paulo: Cengage Learning, 2011.

LAURENTI, R. et al. A confiabilidade dos dados de mortalidade e morbidade por doenças crônicas não-transmissíveis. **Ciênc. Saúde Coletiva**, vol. 9, 2004, p. 909-920.

JORGE, M. H. P. M.; KOIZUMI, M. S. Gastos governamentais do SUS com internações hospitalares por causas externas: análise no Estado de São Paulo. **Rev. Bras. Epidemiol**, Vol. 07, N^o. 2, 2004, p. 228-238.

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.

_____. Dispõe sobre o preenchimento de Autorização de Internação Hospitalar - AIH, em casos com quadro compatível com causas externas. Brasília: **Diário Oficial da União**, 222^a ed., 1997b.

LEBRÃO, M. Análise da fidedignidade dos dados estatísticos hospitalares disponíveis na Secretaria de Estado da Saúde de São Paulo, **Rev. Saúde Pública**, vol. 12, 1978, p. 234-249.

MORRISON, D.: **Multivariate statistical methods**. Singapore: McGraw Hill, 2a ed., 1976.

STRAPASSON, E.; VENCovsky, R; BATISTA, L. A. R. Seleção de Descritores na Caracterização de Germoplasma de *Paspalum* sp. por meio de Componentes Principais. **Revista Brasileira Zootecnia**, vol. 29 (2), 2000, p. 273-381.