

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA

**ANÁLISE DA MORTALIDADE POR CAUSAS EXTERNAS NO BRASIL NO
PERÍODO DE 2009 A 2012 POR MEIO DE TÉCNICAS MULTIVARIADAS**

HEVERTON RODRIGUES FERNANDES

**Trabalho de conclusão de curso
apresentado ao curso de Estatística da
Universidade Federal de Uberlândia
(UFU) como parte dos requisitos para a
conclusão do Curso de Bacharelado em
Estatística.**

UBERLÂNDIA - MG

2015

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA

**ANÁLISE DA MORTALIDADE POR CAUSAS EXTERNAS NO BRASIL NO
PERÍODO DE 2009 A 2012 POR MEIO DE TÉCNICAS MULTIVARIADAS**

HEVERTON RODRIGUES FERNANDES

Orientador:

Prof. Dr. LÚCIO BORGES DE ARAÚJO

**Trabalho de conclusão de curso
apresentado ao curso de Estatística da
Universidade Federal de Uberlândia
(UFU) como parte dos requisitos para a
conclusão do Curso de Bacharelado em
Estatística.**

UBERLÂNDIA - MG

2015

HEVERTON RODRIGUES FERNANDES

Monografia defendida em 11 de Dezembro de 2015 e aprovada pela seguinte Banca
Examinadora:

Prof. Dr. Lúcio Borges de Araújo
Faculdade de Matemática – UFU

Prof. Dr. Janser Moura Pereira
Faculdade de Matemática – UFU

Profa. Dra. Priscila Neves Faria
Faculdade de Matemática – UFU

Agradecimentos

A Deus por sempre ter guiado meus passos.

Aos familiares e amigos pelo incentivo dado ao longo de todo o curso.

Aos professores pelos conhecimentos transmitidos.

Ao professor doutor Lúcio Borges de Araújo pelo apoio, orientação e ensinamentos.

A Maria Ivonete Ramos pelo apoio acadêmico e também pelo carinho, respeito e amizade.

Aos meus colegas de turma pelas experiências trocadas e pelas lições apreendidas com cada um.

Resumo

As mortalidades por causas externas atualmente representam um grande problema de saúde pública no Brasil e no mundo. Grande parte dos acidentes e os atos violentos que as compõem são previsíveis e preveníveis, motivo pelo qual são passíveis de serem enfrentados. Desse modo, o presente trabalho teve por objetivo descrever e analisar os óbitos por causas externas no Brasil, no período de 2009 a 2012, através de técnicas multivariadas. A Análise de *Cluster* foi utilizada para comparar a similaridade entre os estados brasileiros com relação as variáveis em estudo e a Análise de Componentes Principais para verificar as associações existentes entre as variáveis, entre os estados e entre as variáveis e os estados. Ambas análises se mostraram eficazes para o estudo das mortalidades por causas externas no Brasil.

Palavras-chave: Mortalidade, Causas externas, Biplot, Cluster

Abstract

The mortality from external causes currently represents a major public health problem in Brazil and worldwide. Most of the accidents and violent acts that compose it are predictable and preventable. Those situations are, therefore, likely to be confronted. Thus, this study aimed to describe and analyze deaths from external causes in Brazil, from 2009 to 2012, through multivariate techniques. Cluster's analysis was used to compare the similarity between the Brazilian states regarding the variables in study and the principal components analysis to verify the associations between variables among the states and between the variables and states. Both analyzes were shown to be effective in the study of mortality from external causes in Brazil.

Keywords: Mortality, External causes, Biplot, Cluster

Sumário

1. INTRODUÇÃO.....	7
2. METODOLOGIA.....	8
3. RESULTADOS E DISCUSSÃO	16
4. CONCLUSÃO.....	29
5. REFERÊNCIAS	30

1. INTRODUÇÃO

As Causas Externas estão incluídas na Classificação Internacional de Doenças (CID) e designam uma extensão de eventos classificados em acidentais e intencionais, abrangendo os agravos à saúde que causam morbidade e mortalidade. Incluem-se neste grupo: as lesões provocadas por quedas, afogamentos, episódios no transporte, homicídios, agressões, envenenamentos, suicídios, queimadura e ocorrências provocadas por eventos ambientais (COCCO; LOPEZ, 2010).

De acordo com Jorge et al. (2007) as causas externas atualmente representam um dos maiores problema de saúde pública uma vez que sua mortalidade e morbidade são elevadas, abrangem faixas etárias jovens e representam elevados custos para o setor público. O conjunto de acidentes e violências que as compõem são previsíveis e preveníveis, razão pela qual podem e devem ser afrontados.

No Brasil, o crescimento das causas externas, que a partir de 1980, passam a ocupar o segundo lugar entre as causas de morte no país, têm se configurado como grande problema de saúde pública. No ano 2000, ocorreram 118.367 mortes por causas externas, representando 12,5% do total de mortes. Este problema não ocorre somente no Brasil, mas também em outros países do mundo (GAWRYSZEWSKI et al., 2004) .

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 disto, o presente trabalho teve como objetivo contribuir para os estudos sobre mortalidade por causas externas no Brasil através de análise comparativa dos padrões e tendências das mortes por causas externas nos estados brasileiros por meio de técnicas multivariadas.

2. METODOLOGIA

Foram utilizados os bancos de dados do Sistema de Informações de Mortalidade (SIM) composto de informações provenientes das declarações de óbito. As variáveis em estudo foram as taxas relativas à causa básica da morte: acidentes de transporte terrestres (TT), homicídios (TH), suicídios (TS), causa de intencionalidade indeterminada (TI) e demais causas (TD) relativos aos anos de 2009, 2010, 2011 e 2012.

Primeiramente realizou-se um estudo descritivo para caracterizar o comportamento das mortes por causas externas no Brasil. Na sequência, foram aplicadas as técnicas estatísticas multivariadas (Análise de *Cluster* Hierárquica e Análise de Componentes Principais) que permitiram comparar os diferentes perfis entre os estados brasileiros. Os procedimentos estatísticos executados neste trabalho foram realizados com o auxílio do software R.

2.1 Análise de *Cluster*

A Análise de *Cluster* é um método multivariado usado para classificar objetos ou indivíduos em grupos relativamente homogêneos chamados *clusters*. Os objetos, em cada *cluster*, tendem a ser semelhante entre si, mas diferentes de objetos em outros *clusters* não havendo qualquer informação, a priori, sobre a estrutura ou comportamento dos dados (MALHOTRA, 2001).

A Análise de *Cluster* Hierárquica interliga as observações por suas associações, produzindo um dendrograma onde as observações semelhantes são agrupadas entre si. No dendrograma quanto menor a distância entre os pontos, maior a semelhança entre as observações (NETO e MOITA, 1998).

Na maioria dos métodos de Análise de *Cluster* deve-se estabelecer uma medida de similaridade ou dissimilaridade, normalmente expressa como uma função distância ou métrica. Estas medidas são utilizadas na análise de cluster de forma a determinar a

distância entre elementos. Essa distância, é normalmente representada na forma de uma matriz denominada matriz de similaridade (DONI, 2004).

Existem várias medidas que podem ser utilizadas como medidas de similaridade ou dissimilaridade, das quais podemos citar: distância Euclidiana, distância de Manhattan, distância de Mahalanobis, entre outras.

2.1.1 Medidas de distâncias

Segundo Cormack (1971) a distância Euclidiana entre dois casos (i e j) é a raiz quadrada do somatório dos quadrados das diferenças entre os valores de i e j para todas as variáveis ($v = 1, 2, \dots, p$), isto é:

$$d_{ij} = \sqrt{\sum_{v=1}^p (X_{iv} - X_{jv})^2}$$

em que X_{iv} representa a característica do indivíduo i ; X_{jv} representa a característica do indivíduo j ; p é o número de parcelas na amostra; e v é o número de indivíduos na amostra.

Conforme Quintal (2006) a distância de Manhattan é menos afetada por *outliers* do que a distância Euclidiana e de mais fácil interpretação. Esta distância mede a distância numa configuração retilínea, numa configuração de cidade (os quarteirões). Dados dois objetos i e j , a distância de Manhattan define-se por:

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|$$

De acordo com Quintal (2006), a distância de Mahalanobis além de reduzir a dependência das unidades de medição, reduz também a correlação entre variáveis. A distância de Mahalanobis entre os grupos i e j é usualmente estimada segundo Rao (1952) por:

$$D_{ij}^2 = (\bar{X}_i - \bar{X}_j)' \cdot \Sigma^{-1} \cdot (\bar{X}_i - \bar{X}_j)$$

em que $\bar{X}_i$ é o vetor de médias do i 'ésimo grupo; $\bar{X}_j$ é o vetor de médias do j 'ésimo grupo; e Σ é a estimativa combinada da matriz da covariância/variância dentro dos grupos.

Após a construção da matriz de similaridade existem vários métodos para o agrupamento hierárquico. Os mais comuns são: Ligação Simples, Ligação Completa, Centroides, Ligação Média, Ligação Mediana e método Ward.

2.1.2 Método da Ligação Simples

No método da Ligação Simples, a similaridade entre dois conglomerados é definida pelos dois elementos mais parecidos entre si (SNEATH, 1957). Por meio desse método, a distância entre uma observação k e um grupo formado pelas observações i e j é dada por:

$$D_{(ij)k}^2 = \min (D_{ik}^2; D_{jk}^2)$$

em que $D_{(ij)k}^2$ é a distância entre o grupo ij e a observação k ; e $\min (D_{ik}^2; D_{jk}^2)$ é a menor distância entre os grupos de observações ik e jk (AMARAL JÚNIOR e THIÉBAUT, 1999).

Segundo Anderberg (1973), algumas características desse método são: grupos muito próximos podem não ser identificados; permite detectar grupos de formas não-elípticas; apresenta pouca tolerância a ruído uma vez que tem tendência a incorporar os ruídos em um grupo já existente; apresenta bons resultados tanto para distâncias Euclidianas quanto para outras distâncias; e tem tendência a formar longas cadeias (encadeamento).

2.1.3 Método da Ligação completa

O agrupamento por Ligação Completa é exatamente o oposto do Método da Ligação Simples. Nesse caso, os elementos são agrupados considerando a distância máxima (ou similaridade mínima) (DILLON e GOLDSTEIN, 1984). O algoritmo é iniciado encontrando a menor distância $D = \{d_{ik}\}$ e agrupando os objetos correspondentes (U e V) para formar o grupo (UV). No passo 3 do algoritmo, as distâncias entre (UV) e qualquer outro grupo W são calculadas por (JOHNSON e WICHERN, 2002):

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\}$$

Este método tem tendência para encontrar *clusters* compactos compostos de objetos muito semelhantes entre si. Quando um objeto é acrescentado a um grupo, a distância do novo grupo aos restantes aumenta ou então fica inalterada. O método de ligação completa tende a formar grupos pequenos que depois serão aglutinados para formar grupos maiores (QUINTAL, 2006).

2.1.4 Ligação Média

A ligação média entre *clusters*, também conhecida por UPGMA (Unweighted Pair-Group Method using the Average approach), é um método não-ponderado de agrupamento aos pares, utilizando médias aritméticas das medidas de dissimilaridade, que evita caracterizar a dissimilaridade por valores extremos (máximo ou mínimo) (CRUZ & CARNEIRO, 2003). Este método trata a distância entre dois conglomerados como a média das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois conglomerados que estão sendo comparados. Portanto, se o conglomerado C_1 tem n_1 elementos e o conglomerado C_2 tem n_2 elementos, a distância entre eles será definida por (MINGOTI, 2013):

$$d(C_1, C_2) = \sum_{l \in C_1} \sum_{k \in C_2} \left(\frac{1}{n_1 n_2} \right) d(X_l, X_k)$$

A grande vantagem deste critério é tornear as consequências da existência de valores extremos e considerar toda a informação dos grupos. Só que, dependendo do

tipo de *cluster* que se espera obter, esta propriedade também pode ser vista como uma desvantagem (QUINTAL, 2006).

2.1.5 Método do Centroide

No método do Centroide (ou UPGMC), a distância entre dois grupos é definida como sendo a distância entre os vetores de médias, também chamados de centroides, dos grupos que estão sendo comparados. Logo, considerando dois conglomerados C_1 e C_2 e seus respectivos vetores de médias amostral $\overline{X}_1$ e $\overline{X}_2$, a distância entre C_1 e C_2 é definida por:

$$d(C_1, C_2) = (\overline{X}_1 - \overline{X}_2)'(\overline{X}_1 - \overline{X}_2)$$

que é a distância Euclidiana ao quadrado entre os vetores de médias amostral $\overline{X}_1$ e $\overline{X}_2$. O método do centroide também pode ser usado com a distância Euclidiana usual entre os vetores de médias. Em cada passo do algoritmo do agrupamento, os conglomerados que apresentam o menor valor de distância são agrupados (MINGOTI, 2013).

De acordo com Mingoti (2013), o método do centroide é direto e simples, porém, para fazer o agrupamento, é necessário em cada passo voltar-se aos dados originais para o cálculo da matriz de distâncias, exigindo um tempo computacional maior comparado com outros métodos. O método do centroide não pode ser usado em situações nas quais se dispõe apenas da matriz de distâncias entre os n elementos amostrais.

2.1.6 Ligação mediana ou WPGMC

O método da distância mediana, também conhecido por WPGMC (weighted pair-group method using the centroid approach), é semelhante ao método do centroide, com a diferença de que ao aglutinar os dois grupos A e B, os seus centroides, $\overline{X}_A$ e $\overline{X}_B$, recebem pesos iguais antes de produzirem o centroide do novo *cluster* resultante da aglutinação. O novo centroide $\overline{X}$ fica a meio dos centroides dos grupos aglutinados,

$$\overline{X} = \frac{(\overline{X}_A + \overline{X}_B)}{2}$$

com o objetivo de evitar que o grupo com maior número de objetos absorva o grupo com menor número de objetos. Repare que a mediana referida não corresponde à mediana estatística mas sim à mediana de um triângulo, isto é, um segmento de reta que une um vértice de um triângulo ao ponto médio do lado oposto (QUINTAL, 2006).

2.1.7 Critério de Ward

O método de agrupamento proposto por Ward (1963) é fundamentado na mudança de variação entre os grupos e dentro dos grupos que estão sendo formados em cada passo do agrupamento. Cada elemento é considerado como um único conglomerado, e em cada passo do algoritmo de agrupamento é calculado a soma de quadrados dentro de cada conglomerado. Esta soma é dada por:

$$SS_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)' (X_{ij} - \bar{X}_i)$$

em que n_i é o número de elementos no conglomerado C_i quando se está no passo k do processo de agrupamento, X_{ij} é o vetor de observações do j – ésimo elemento amostral que pertence ao i – ésimo conglomerado, $\bar{X}_i$ é o centroide do conglomerado C_i , e SS_i representa a soma de quadrados correspondente ao conglomerado C_i . No passo k , a soma de quadrados total dentro dos grupos é definida como:

$$SSR = \sum_{i=1}^{g_k} SS_i$$

onde g_k é o número de grupos existentes quando se está no passo k .

A distância entre os conglomerados C_l e C_i é, então, definida como:

$$d(C_l, C_i) = \left[\frac{n_l n_i}{n_l + n_i} \right] (\bar{X}_l - \bar{X}_i)' (\bar{X}_l - \bar{X}_i)$$

que é a soma de quadrados entre os *clusters* C_l e C_i . Em cada passo do algoritmo de agrupamento, os dois conglomerados que minimizam a distância são combinados.

Tem-se que a medida de distância nada mais é do que a diferença entre o valor de SSR depois e antes de se combinar os conglomerados C_l e C_i em um único

conglomerado. Portanto, a cada passo do conglomerado, o método de Ward combina os dois conglomerados que resultam no menor valor de *SSR* (MINGOTI, 2013).

De acordo com Mingoti (2005) os diferentes métodos de agrupamento e combinações entre as medidas de similaridade ou dissimilaridade levam a padrões de agrupamento distintos. Deste modo, deve-se utilizar vários métodos e comparar os resultados, para que a análise dos dados seja realizada pela técnica mais adequada.

Conforme Valentin (2000) um método é mais adequado que outro quando o dendograma provê uma imagem menos distorcida da realidade. É possível avaliar o grau de deformação provocado pela construção do dendograma obtendo-se o Coeficiente de Correlação Cofenético.

O Coeficiente de Correlação Cofenético é obtido calculando-se o Coeficiente de Correlação Linear de Pearson entre os elementos da matriz de dissimilaridade obtida a partir dos dados originais) e os elementos da matriz cofenética obtida a partir do dendrograma. O menor grau de distorção, será refletido pelo maior coeficiente cofenético sendo que valores próximos à unidade indicam melhor representação (BARROSO e ARTES, 2003; CRUZ e CARNEIRO, 2003; CARGNELUTTI et al., 2010).

2.2 Análise de Componentes principais

A Análise de Componentes Principais é uma técnica multivariada que consiste em transformar um conjunto de p variáveis originais $X_1, X_2, \dots, X_p$, pertencentes a n indivíduos ou populações, em um novo conjunto de variáveis, $Y_1, Y_2, \dots, Y_p$ de dimensão equivalente, denominados componentes principais. Cada componente principal é uma combinação linear das variáveis originais, construído de maneira a explicar o máximo da variabilidade total dessas variáveis originais e não correlacionadas entre si (STRAPASSON et al., 2000).

Segundo Morrison (1976), o primeiro componente principal (Y_1) de um conjunto de p variáveis, $X_1, X_2, \dots, X_p$, contidas no vetor $X' = (X_1, X_2, \dots, X_p)$ é definido 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 à maior raiz característica (λ_1) da matriz de covariância amostral, S , das variáveis X_i 's. O segundo componente principal (Y_2) estará associado à segunda maior raiz característica (λ_2) e, assim sucessivamente, até que toda a variância tenha sido explicada. 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. O primeiro componente também é definido como o de maior importância, uma vez que retém a maior parte da variação total encontrada nos dados originais. A técnica envolve a matriz de covariância amostral S ou a matriz de correlação.

A técnica dos componentes principais tem a vantagem adicional de avaliar a importância de cada caractere estudado sobre a variação total disponível entre os acessos avaliados, possibilitando o descarte dos caracteres menos discriminantes, por já estarem correlacionados com outras variáveis ou pela sua invariância (DAHER et al., 1997).

O Biplot é um método desenvolvido por Gabriel (1971) que permite representar graficamente resultados de Análise de Componentes Principais. No Biplot é possível visualizar as relações existentes entre variáveis, entre observações e entre variáveis e observações.

3. RESULTADOS E DISCUSSÃO

Na Tabela 1 são apresentados os valores mínimo e máximo, a média e a mediana, os 1º e 3º quartis, a variância e o desvio padrão das variáveis em estudo no período avaliado. Analisando os resultados obtidos, nota-se que a variável, Taxa de homicídio, apresenta os maiores desvios padrão, isto indica que essa variável provavelmente terá uma maior influência na construção dos agrupamentos e nos componentes principais comparado com as demais variáveis.

Observa-se que as taxas médias permaneceram estáveis ao longo destes 4 anos, no entanto houve um aumento considerável na Taxa máxima de homicídios comparando os anos de 2009 e 2011(59,3 e 71,46 respectivamente), ambas taxas correspondente ao estado de Alagoas.

Além disso, verifica-se que taxa de homicídios seguido da taxa de acidentes de transporte terrestres se configuram como as principais causas de mortalidade por causas externas no Brasil.

Tabela 1: Estatística descritiva das taxas de mortalidade por causas externas nos anos de 2009 a 2012.

2009	TT	TH	TS	TI	TD
Min.	11,10	12,20	2,00	0,40	10,70
1QT.	17,25	22,25	4,20	1,90	12,30
Mediana	20,00	30,70	4,90	3,10	13,50
Média	22,47	30,42	5,22	4,72	13,99
3QT.	29,30	34,20	6,10	5,15	14,70
Max.	35,60	59,30	10,20	22,60	20,50
Variância	50,81	127,82	3,87	25,00	5,92
Desvio Padrão	7,13	11,31	1,97	5,00	2,43
2010					
Min.	13,70	13,20	2,50	0,40	9,90
1QT.	19,45	23,35	4,25	1,65	12,50
Mediana	22,50	31,80	5,20	2,70	14,20
Média	25,49	31,34	5,23	3,51	14,27
3QT.	31,55	37,00	6,00	4,85	15,70
Max.	39,20	66,90	9,70	9,10	22,00
Variância	51,70	143,23	3,10	5,74	8,36

Desvio Padrão	7,19	11,97	1,76	2,40	2,89
2011					
Min.	14,50	12,80	2,70	0,30	10,20
1QT.	20,50	22,95	4,05	1,50	12,70
Mediana	22,80	32,10	5,40	2,50	14,40
Média	25,44	31,25	5,40	3,56	14,73
3QT.	31,85	36,95	6,45	4,95	15,70
Max.	37,60	71,40	9,70	10,50	25,40
Variância	44,06	145,23	3,29	7,93	9,36
Desvio Padrão	6,64	12,05	1,81	2,82	3,06
2012					
Min.	12,90	12,90	2,90	0,40	10,40
1QT.	19,95	27,00	4,15	1,60	13,25
Mediana	26,20	34,50	5,20	3,00	14,80
Média	26,32	33,64	5,53	3,78	14,97
3QT.	32,20	40,70	6,45	5,50	16,45
Max.	41,80	64,60	11,00	12,50	23,00
Variância	54,64	122,32	3,87	8,68	7,76
Desvio Padrão	7,39	11,06	1,97	2,95	2,78

A Análise de *Cluster* foi utilizada para comparar a similaridade entre os estados brasileiros com relação às taxas de mortalidade por causas externas. Para a determinação da melhor metodologia, foram avaliados os valores dos Coeficientes de Correlação Cofenética obtidos para cada combinação de diferentes medidas de dissimilaridade e métodos de agrupamento (Tabela 2). O critério de agrupamento escolhido foi o Método de Ligação Média obtido pela matriz de distância Euclidiana por apresentar maior coeficiente de correlação cofenética e, portanto, ser este o mais coerente na formação dos agrupamentos.

Tabela 2: Comparação dos métodos por meio do Coeficiente de Correlação Cofenética.

Métodos de agrupamento					
Distância Euclidiana	Ligação Simples	Ligação Completa	Centroide	Ligação Média	Ward
2009	0,765365	0,740863	0,788751	0,823019	0,497219
2010	0,728214	0,735527	0,759837	0,776289	0,54071
2011	0,788595	0,789149	0,800971	0,850498	0,477796
2012	0,709276	0,729164	0,693488	0,747374	0,530159
Distância de Manhattan					
2009	0,688897	0,740667	0,709928	0,782828	0,521061
2010	0,702905	0,584988	0,670623	0,76769	0,610091
2011	0,729156	0,705033	0,738409	0,791504	0,479956
2012	0,63663	0,628492	0,656081	0,734743	0,502039
Distância de Mahalanobis					
2009	0,686796	0,569173	0,762705	0,690161	0,444642
2010	0,643087	0,474285	0,703281	0,702206	0,410145
2011	0,729326	0,501241	0,787696	0,753433	0,425378
2012	0,625401	0,40094	0,689533	0,694409	0,358797

Os dendogramas das Figuras 1 a 4 mostram, nos eixos das ordenadas, a distância Euclidiana multidimensional entre os grupos de unidades da federação (distância considerando as 5 variáveis usadas). Nos eixos das abscissas, os dendogramas mostram as Unidades da Federação e os grupos a que elas pertencem.

Analisando-se as Figuras, verifica-se que em cada ano em estudo houve a formação de 4 grupos distintos e comparando-as, observa-se que em cada ano analisado obteve-se um grupo formado apenas pelo estado de Alagoas, exceto no ano de 2009 (Figura 1) em que o estado do Espírito Santo foi agrupado juntamente ao estado de Alagoas. Observa-se também que no período analisado os estados de Alagoas e Espírito Santo foram as que obtiveram as maiores taxas de homicídio. Constata-se que o grupo 2 foi formado pelos estados do Pará, Pernambuco, Bahia e Rio de Janeiro. Já o grupos 3 é constituído pelos estados de Santa Catarina e Piauí. Os demais estados se encontram agrupados no grupo 1.

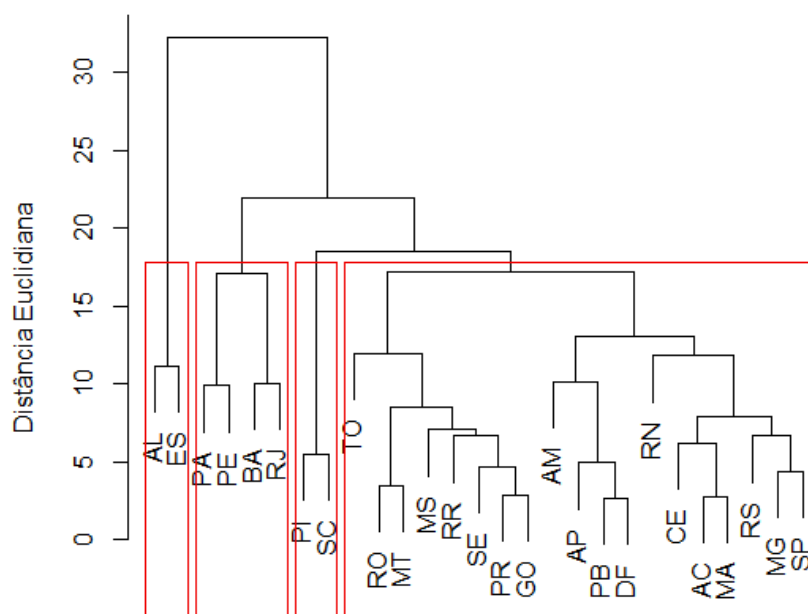


Figura 1: Dendograma relativo ao ano de 2009.

Conforme a Tabela 3, o grupo 2 caracteriza-se por estados com taxas de óbitos em acidentes de transporte terrestres, suicídio e demais causas relativamente baixas, no entanto possui taxa média de homicídio alta e é o grupo com maior taxa média de causas não identificadas. O grupo 3 formado pelos estados de Santa Catarina e Piauí foi caracterizado por taxa média de homicídio baixa.

Tabela 3: Estatística descritiva dos agrupamentos relativa ao ano de 2009.

2009	Grupo	n	Min.	Média	Max.	Desvio Padrão
TT	1	19	11,1	23,2	35,6	7,4
	2	4	12,7	15,3	20	3,2
	3	2	25,5	27,9	30,4	3,5
	4	2	20,8	24,1	27,3	4,6
TH	1	19	15,8	27,6	35,8	6
	2	4	33,5	39	45	4,9
	3	2	12,2	12,8	13,4	0,8
	4	2	56,9	58,1	59,3	1,7
TS	1	19	2,5	5,6	10,2	1,7
	2	4	2	2,7	3,7	0,7
	3	2	6,6	7,5	8,5	1,3
	4	2	3,5	3,9	4,3	0,6
TI	1	19	0,4	3,8	14	3,1
	2	4	2,3	11,7	22,6	8,9

	3	2	1,9	2,4	2,9	0,7
	4	2	0,9	2,3	3,7	2
TD	1	19	10,7	14,1	18,4	2,1
	2	4	10,9	13,2	15,9	2,4
	3	2	11,8	11,9	11,9	0,1
	4	2	12,3	16,4	20,5	5,8

Com relação ao ano de 2010 (Figura 2) verifica-se que o grupo 2 é constituído por 8 estados: Acre, Maranhão, Piauí, Rio Grande do Norte, Minas Gerais, São Paulo, Santa Catarina e Rio Grande do Sul. Os demais estados foram agrupados no grupo 1, exceto o estado de Alagoas e Espírito Santo que constitui um grupo cada estado (grupo 3 e grupo 4, respectivamente).

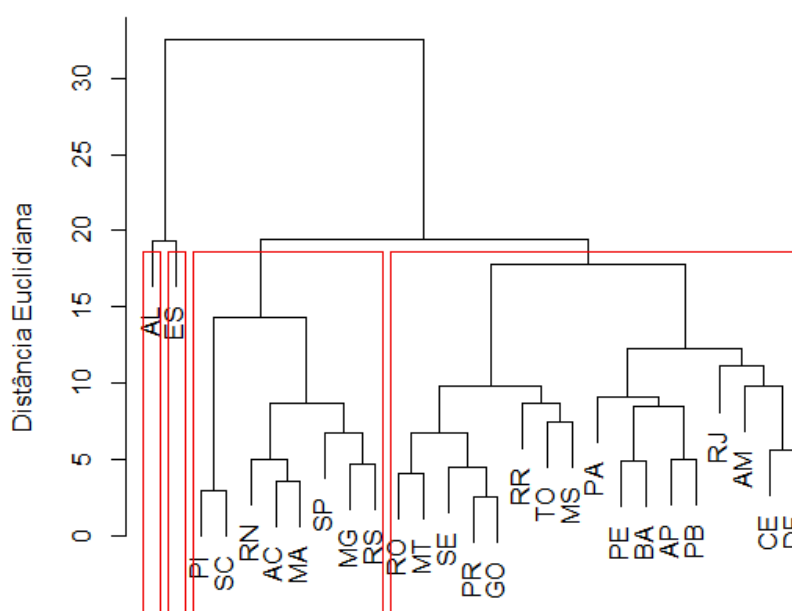


Figura 2: Dendrograma relativo ao ano de 2010

Com relação a taxa de óbitos por acidentes de transporte terrestre e demais causas, observa-se que não houve muita variação comparando os grupos 1,2 e 3 (Tabela 4), no entanto o grupo 4, formado pelo estado do Espírito Santo, obteve taxas de óbitos em acidentes de transporte terrestres e demais causas alta (31,8 e 22,0 respectivamente). Os estados de Alagoas (grupo 3) e Espírito Santo (grupo 4) foram as que obtiveram

maiores taxas de homicídio, em contrapartida o grupo 2 foi constituído de estados com as menores taxas de homicídio.

Tabela 4: Estatística descritiva dos agrupamentos relativa ao ano de 2010.

2010	Grupo	n	Min.	Média	Max.	Desvio Padrão
TT	1	17	13,7	26,5	39,2	8,1
	2	8	17,9	22,6	31,3	5,1
	3	1	-	25,1	-	-
	4	1	-	31,8	-	-
TH	1	17	23,6	34	46,4	5,8
	2	8	13,2	18,8	25,6	4,8
	3	1	-	66,9	-	-
	4	1	-	51	-	-
TS	1	17	2,5	5,1	7,7	1,5
	2	8	3,2	6	9,7	2,2
	3	1	-	2,7	-	-
	4	1	-	4,6	-	-
TI	1	17	0,4	3,7	9,1	2,6
	2	8	1,5	3,4	6,4	2,1
	3	1	-	0,7	-	-
	4	1	-	3	-	-
TD	1	17	9,9	14,6	20,6	2,7
	2	8	10,7	12,7	15	1,6
	3	1	-	13,7	-	-
	4	1	-	22	-	-

Em 2011 nota-se que o resultado foi semelhante ao encontrado no ano de 2010 (Figura 3). Entretanto, o estado do Rio Grande no Norte que em 2010 encontrava-se no grupo 2 em 2011 agrupou-se ao grupo 1. Por sua vez, o estado de Roraima, antes pertencente ao grupo 1, encontrou-se agrupado no grupo 2. Os demais estados pertenceram ao grupo 1, exceto o Estado do Rio de Janeiro que correspondeu a um grupo isolado (grupo 4).

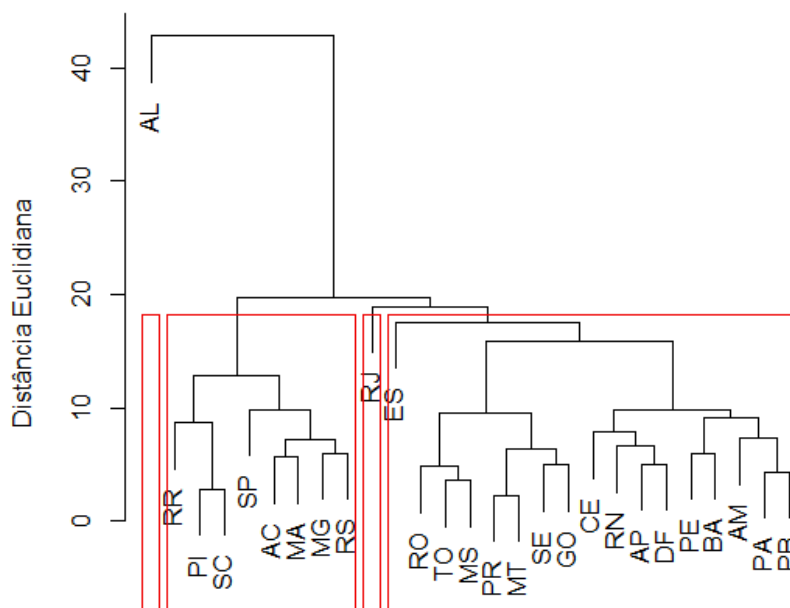


Figura 3: Dendrograma relativo ao ano de 2011.

Pela Tabela 5 verifica-se que o estado do Rio de Janeiro (grupo 4) teve como característica principal altas taxas de mortes por causas indeterminadas e demais causas externas. O grupo 2 caracteriza-se por taxas de homicídios baixas e o grupo 3 por taxas de homicídios moderadas. Com relação às variáveis taxas de óbitos no transporte terrestres e demais causas, observa-se que não houve muita variação entre os grupos.

Tabela 5: Estatística descritiva relativa aos agrupamentos do ano de 2011.

2011	Grupo	n	Min.	Média	Max.	Desvio Padrão
TT	1	17	14,5	26,2	37,6	7,3
	2	8	18,2	24,8	32,1	5,4
	3	1	-	26,4	-	-
	4	1	-	17,2	-	-
TH	1	17	25,8	35,0	47,1	5,6
	2	8	12,8	18,5	23,9	4,3
	3	1	-	71,4	-	-
	4	1	-	29,7	-	-
TS	1	17	2,9	5,1	8,5	1,4
	2	8	3,3	6,6	9,7	2,0
	3	1	-	3,3	-	-

	4	1	-	2,7	-	-
TI	1	17	0,7	3,4	10,5	2,7
	2	8	1,4	3,4	7,3	2,2
	3	1	-	0,3	-	-
	4	1	-	10,5	-	-
TD	1	17	12,0	14,8	19,8	2,3
	2	8	10,2	13,5	15,4	2,0
	3	1	-	12,2	-	-
	4	1	-	25,4	-	-

Na Figura 4 nota-se no ano de 2012 que o grupo 2 permaneceu semelhante ao ano de 2011, exceto pelo estado de Roraima que se agrupou juntamente com os estados de Tocantins, Paraná, Mato Grosso do Sul e Mato Grosso no grupo 1. Os demais estados encontraram-se agrupados no grupo 3, exceto o estado de Alagoas que ficou no grupo 4.

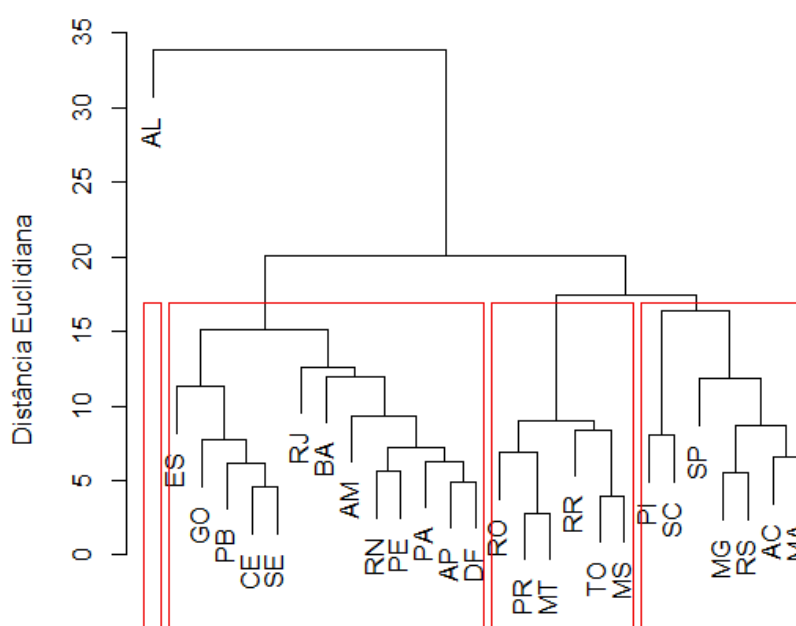


Figura 4: Dendrograma relativo ao ano de 2012

De acordo com a Tabela 6, averigua-se que o grupo 1 possui a maior taxa média de mortes por acidentes de transporte terrestres com relação aos demais grupos e possui taxa média de homicídio regular. Com relação às taxas médias de homicídio observa-se que o grupo 4 (formado pelo estado de Alagoas) seguido do grupo 3 obtiveram as maiores taxas.

Tabela 6: Estatística descritiva dos agrupamentos relativa ao ano de 2012

2012	Grupo	n	Min	Média	Máx	Desvio Padrão
TT	1	6	31,3	35,2	41,8	3,8
	2	7	17,2	24,6	35,8	6,6
	3	13	12,9	23,1	33,4	6,3
	4	1	-	26,4	-	-
TH	1	6	26,7	30,9	34,5	3,3
	2	7	12,9	20,6	27,4	5,6
	3	13	29,4	39,5	46,6	4,9
	4	1	-	64,6	-	-
TS	1	6	4,6	6,6	8,4	1,4
	2	7	3,1	6,7	11	2,6
	3	13	2,9	4,6	6,5	1,2
	4	1	-	3,4	--	-
TI	1	6	1,6	3,2	6,6	1,9
	2	7	0,7	3,1	5,7	2
	3	13	0,7	4,6	12,5	3,6
	4	1	-	0,4	-	-
TD	1	6	12,6	16,2	18	1,9
	2	7	10,4	13,7	16,5	2,1
	3	13	12	15,3	23	3,2
	4	1	-	11,5	-	-

Nas Figuras 5 a 8 é possível visualizar por meio do *Biplot* a relação entre os estados e as variáveis em estudo que estão mais associadas. O primeiro componente (PC1) está representado pelo eixo das abscissas, logo os estados que se localizam no primeiro e quarto quadrantes são os mais associados ao primeiro componente, isto é, são os estados que apresenta coordenada positiva em relação ao eixo das abscissas, enquanto que os estados localizados no segundo e terceiro quadrantes são os menos associados a esse componente. Quanto mais próximos os estados estão dos vetores representando as variáveis, maior é a associação do estado com aquela variável.

No ano de 2009 (Figura 5) observa-se que há relação positiva entre as variáveis TT com TS e entre TD com TH. As variáveis TT e TS apresentam relação negativa com TI. Os estados MS, MT e RO apresentam altas taxas de TT e TD. Os estados de AL e PE estão relacionados positivamente à TH. Os estados de SC, RS e PI estão relacionados positivamente com TS. Os estados de RJ, PA e BA estão relacionados positivamente a TI. Os demais estados não estão associados a elevadas taxas de

mortalidade. Vale ressaltar que SP, MG, AC, MA e RN apresentam baixas taxas de TT e TD.

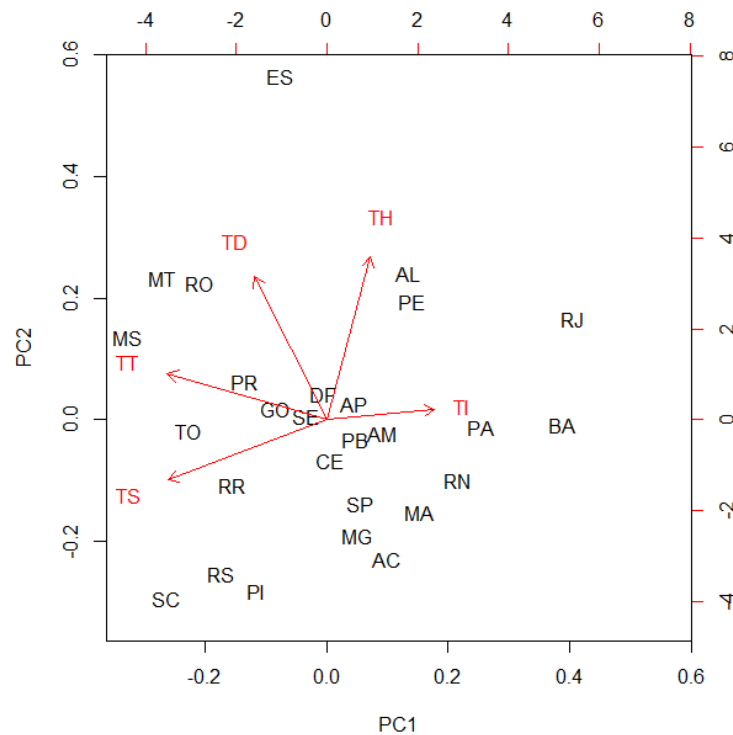


Figura 5: Biplot dos estados em relação às taxas de mortalidade por causas externas relativas ao ano de 2009

No ano de 2010 (Figura 6), verifica-se que as variáveis TS e TT e as variáveis TD, TI e TH apresentam correlação positiva. Além disso, TH apresenta uma correlação negativa com TS e TT. Nota-se também que os estados SC, RS, RR e PI estão relacionados positivamente com TS. Os estados MS, TO, MT e RO estão relacionados positivamente com TT. Os estados ES e RJ estão relacionados positivamente a TD, TI e TH. Já os estados PE, BA, e AL estão associados positivamente a TH. Os estados AC, MA, PB, AP e PA apresentam baixas taxas de TS, TT, TD e TI.

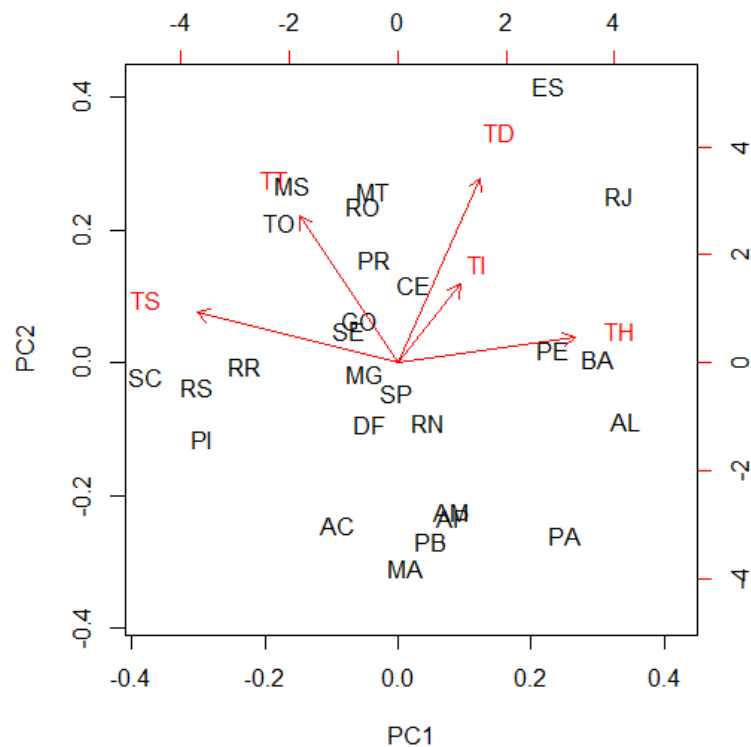


Figura 6: Biplot dos estados em relação às taxas de mortalidade por causas externas relativas ao ano de 2010

Com relação ao ano de 2011 (Figura 7) é possível observar que a variável TT está relacionado positivamente com TS e estas duas taxas estão relacionadas negativamente com TH. Já as variáveis TD e TI apresentam uma alta correlação positiva. Nota-se também que os estados PI, SC, TO, RR, MS e RS estão relacionados positivamente a TT e TS. Os estados AL, PA, PB estão relacionados positivamente a TH. Os estados RJ, MG, SP, CE, BA e PE estão relacionados positivamente a TD e TI. Além disso, PE e BA apresentam uma correlação negativa com TS e TT.

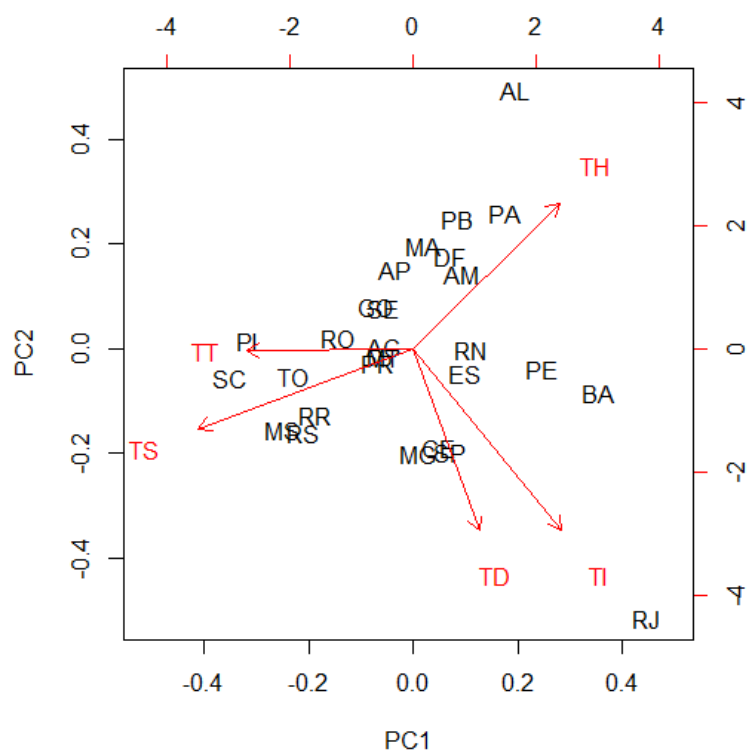


Figura 7: Biplot dos estados em relação às taxas de mortalidade por causas externas relativas ao ano de 2011

Para o ano de 2012 (Figura 8) tem-se uma correlação positiva entre as variáveis TS e TT e também entre TI e TH, mas estes dois grupos de causas de mortes apresentam uma correlação negativa entre eles. Observa-se que os estados SC, RS, PI, MS e RR estão relacionados positivamente a TS. A TT está associada positivamente aos seguintes estados: GO, RO, MT, PR e TO. Já os estados PA, ES, PB, SE e AL apresentam altos valores de TH. Em relação a TI têm-se associados os estados de RJ, BA, RN, SP e MG.

4. CONCLUSÃO

As Análises de *Cluster* e Análise de Componentes Principais se mostraram eficazes no estudo das mortalidades por causas externas no Brasil. Na Análise de *Cluster* a variável taxa de homicídio teve uma maior influência na construção dos agrupamentos. Deste modo foi possível comparar a similaridade entre os estados brasileiros com relação às taxas de mortalidades principalmente no que diz respeito ao homicídio, que no período estudado se configurou como a principal causa de mortalidade por causas externas. Foi possível identificar que os estados de Alagoas e Espírito Santo obtiveram as maiores taxas de homicídio. Por meio da Análise de Componentes Principais e construção dos Biplots foi possível identificar os padrões das mortes por causas externas nos estados brasileiros. Pode-se concluir que os estados: Mato Grosso, Mato Grosso do Sul, Rondônia e Tocantins estiveram relacionados com altas taxas de mortes por acidentes de transporte terrestre. Com relação ao homicídio os estados de Alagoas, Pará, Espírito Santo, Pernambuco e Paraíba foram as que obtiveram maiores taxas no período analisado. As taxas de suicídio estiveram associadas a dois estados do sul do país (Santa Catarina e Rio Grande do Sul) juntamente com os estados de Roraima e Piauí. Rio de Janeiro e Bahia estiveram associados à taxa de mortalidade não identificada. Pode-se observar que em todo o período analisado as taxas de suicídio estiveram relacionadas positivamente com as taxas de óbitos em acidentes terrestres. Pretende-se em trabalhos futuros avaliar como os atributos sociais se assemelham e relacionam com os diferentes tipos de mortalidade por causas externas.

5. REFERÊNCIAS

- ANDERBERG, M. R. *Cluster analysis for applications*. New York: Academic Press, 1973.
- BARROSO, L.P.; ARTES, R. *Análise multivariada*. Lavras: UFLA, 2003. 151p.
- BUSSAB, W. de O.; MIAZAKI, S. E.; ANDRADE, D. F. **Introdução à análise de agrupamento**. In: 9. Simpósio Brasileiro de Probabilidade e Estatística. São Paulo: 1990.
- CARGNELUTTI, F.A. RIBEIRO, N. D. BURIN, C. Consistência do padrão de agrupamento de cultivares de feijão conforme medidas de dissimilaridade e métodos de agrupamento. *Pesq. agropec. bras.* [online]. 2010, vol.45, n.3, pp. 236-243. ISSN 0100-204X.
- COCCO, M. LOPES, M. J. M. Morbidade por Causas Externas em Adolescentes de uma Região do Município de Porto Alegre. *Revista Eletrônica de Enfermagem – UFG*.2010. Disponível em: <http://www.revistas.ufg.br/index.php/fen/article/view/9527/6594>. Acesso em 09 jul. 2015.
- CRUZ, C.D.; CARNEIRO, P.C.S. *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: UFV, 2003. 585p.
- DAHER, R.F.; et al. Seleção de caracteres morfológicos em capim-elefante (*Pennisetum purpureum* Schum.). *Revista Brasileira de Zootecnia*, v.26, n.2, p. 247-259, 1997.
- DILLON; W. R.; GOLDSTEIN, M. **Multivariate Analysis – Methods and Applications**, John Wiley & Sons, Estados Unidos, 1984.
- DONI, M. V. *Análise de cluster: métodos hierárquicos e de Particionamento*. São Paulo: Universidade Presbiteriana Mackenzie, São Paulo, 2004.
- GABRIEL, K.R. The biplot graphic display of matrices with application to principal componente analysis. *Biometrika*, v.58, n.3, p.453-467, 1971. Disponível em: <<http://biomet.oxfordjournals.org/content/58/3/453.abstract>>. Acesso em: 13 nov . 2015. doi: 10.1093/biomet/58.3.453.

GAWRYSEWSKI VP, KOIZUMI MS, MELLO-JORGE MHP. As causas externas no Brasil no ano 2000: comparando a mortalidade e a morbidade. *Cad Saúde Publica*. 2004;20(4):995-1003. Disponível em: <http://www.scielo.br/pdf/csp/v20n4/14.pdf>. Acesso em 08 jul. 2015.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**, Prentice Hall, New Jersey, 2002.

JORGE, M., KOIZUMI, M., TONO, V.. Causas Externas: O que são, como afetam o setor saúde, sua medida e alguns subsídios para a sua prevenção.. *Revista Saúde - UnG*, América do Norte, 1, aug. 2007. Disponível em: <http://revistas.ung.br/index.php/saude/article/view/67>. Acesso em: 08 Jul. 2015.

MALHOTRA, N. *Pesquisa de Marketing: uma orientação aplicada*. Porto Alegre: Bookman, 2001.

MINGOTI, S. A. *Análise de dados através de métodos de Estatística Multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG, 2005. 295p.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2013. 297p.

MORRISON, D.F. *Multivariate statistical methods*. 2.ed. Singapore: McGraw Hill. 1976, 415p.

NETO, J.M.N. MOITA, G.C. Uma introdução à análise exploratória de dados multivariados. *Química Nova*. 1998; 21(4):467-469.

QUINTAL, G. **Análise de clusters aplicada ao Sucesso/Insucesso em Matemática**. Dissertação de mestrado em Matemática, Universidade da Madeira, Funchal. 2006.

RAO, C. R.; **An advanced statistical method in biometric research**. New York, Ed. John Wiley e Sons, p.390, 1952.

SNEATH, P. H. A. Prejudice in bacterial classification. **J. gen. Microbiol.** 17, ii. 1957.

STRAPASSON, E.; VENCOSKY, R.; BATISTA, L. A. R. Seleção de Descritores na Caracterização de Germoplasma de *Paspalum* sp. por meio de Componentes Principais. *Rev. bras. zootec.*, 29(2):373-381, 2000.

VALENTIN, J. L. *Ecologia numérica: uma introdução à análise multivariada de dados ecológicos*. Rio de Janeiro: Interciência, 2000.

VALLI, M. Análise de *Cluster*. **Augusto Guzzo Revista Acadêmica** (São Paulo), v. 04, p. 77, 2002.

WARD, J. H. Jr. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, Alexandria, v.58, p.236-244, 1963.