

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Carolina Dutra Cyrino

**Análise Geométrica de Dados através de Análise de
Correspondência Múltipla**

Juiz de Fora
2011

Carolina Dutra Cyrino

*Análise Geométrica de Dados através de Análise de
Correspondência Múltipla*

Monografia apresentada ao Curso de Estatística da
Universidade Federal de Juiz de Fora, como requi-
sito para a obtenção do grau de Bacharel em Es-
tatística.

Orientador: Ronaldo Rocha Bastos

Doutor em Urban and Regional Planning - Liverpool University

Juiz de Fora

2011

Cyrino, Carolina

Análise Geométrica de Dados através de Análise de Correspondência Múltipla / Carolina Cyrino - 2011

41.p

1.Análise de Correspondência Múltipla. I.Título.

CDU N/A

Carolina Dutra Cyrino

*Análise Geométrica de Dados através de Análise de
Correspondência Múltipla*

Monografia apresentada ao Curso de Estatística da
Universidade Federal de Juiz de Fora, como requi-
sito para a obtenção do grau de Bacharel em Es-
tatística.

Aprovado em 08 de julho de 2011

BANCA EXAMINADORA

Ronaldo Rocha Bastos

Doutor em Urban and Regional Planning - Liverpool University

Marcel de Toledo Vieira

Doutor em Estatística - University of Southampton

Augusto Carvalho Souza

Mestre em Estudos Populacionais e Pesquisas Sociais - Escola Nacional de Ciências Estatísticas

À minha família.

Resumo

Neste trabalho aplicamos a técnica estatística multivariada, denominada Análise de Correspondência Múltipla (ACM), seguindo uma abordagem de análise geométrica, à base de dados sobre transplante de córnea, onde as linhas representam as córneas coletadas de doadores após o óbito e as colunas representam as categorias de diferentes variáveis que caracterizam as córneas e os doadores. As informações foram obtidas através de prontuários médicos. A base de dados compreende um conjunto de dados reais, com 264 indivíduos e 8 variáveis, obtido do Banco de Olhos do Hospital João Penido. Como a Análise de Correspondência (AC) é desenvolvida utilizando variáveis categóricas ou categorizadas, e algumas variáveis apresentadas são de natureza numérica, utilizamos um método para agrupá-las em categorias ordinais correspondentes aos quartis, apesar da técnica tratá-las como categorias nominais. O objetivo é mostrar geometricamente as linhas e colunas da tabela de dados em um espaço de baixa dimensão, de modo que a proximidade no espaço indique associações, similaridade ou dissimilaridade das categorias e dos indivíduos. Através dos gráficos da ACM, podemos analisar separadamente as variáveis, as categorias, os indivíduos, e conjuntamente os indivíduos e as categorias. A implementação da ACM foi realizada através do *software* livre R.

Palavras-chave: Análise Multivariada, Análise de Correspondência Múltipla, Análise Geométrica de Dados.

Abstract

In this work we applied the multivariate statistical technique, called Multiple Correspondence Analysis (MCA), following an approach based on geometric analysis of data from a corneal transplantation database, where the lines represent the corneas collected from donors after death and the columns represent the different categories of variables that characterize the corneas and donors. The information was obtained from medical records. The database comprises a set of real data with 264 individuals and 8 variables, obtained from the Eye Bank of João Penido Hospital. As Correspondence Analysis (CA) is developed using categorical variables or continuous variables which have been categorized, and some variables are numerical in nature, we used a method to group them into ordinal categories corresponding to quartiles, although the technique treat them as nominal categories. The goal is to show geometrically the rows and columns of the data table in a low-dimensional space, so that the proximity in space indicating associations, similarity or dissimilarity of categories and individuals. Through the graphs of the ACM, we can analyze separately the variables, classes, individuals, and both individuals and categories. The implementation of the MCA was performed using the free software R.

Keywords: Multivariate Analysis, Multiple Correspondence Analysis, Geometric Data Analysis

Agradecimentos

Agradeço à Deus por me dar forças e iluminar meu caminho.

Aos meus pais, Ana e Aloysio, aos quais devo em grande parte o que eu sou hoje e por estarem ao meu lado me apoiando, dando amor, carinho, compreensão e incentivo. À minha irmã Aline, pela companhia, carinho e pelos momentos de descontração vividos a cada dia.

Ao Iago, obrigada por todo apoio, pelo companheirismo, pela compreensão e paciência e por ter compartilhado bons e maus momentos ao meu lado. À Maria Célia, pelo carinho e por me incluir em suas orações.

À Laura e ao Victor pela amizade, convivência, dedicação demonstrada e pelos estudos até mesmo tarde da noite ou pelo telefone.

Ao Samuel, ao Lu, ao Thiago e ao Thales pelas caronas oferecidas.

À Raquel, à Sarah e à Leiliane pela disposição em ajudar.

Aos demais amigos da faculdade, Priscila, Luís, Roberto, Bruno, Isabela e Marcos pelos momentos agradáveis dentro e fora da universidade.

Às minhas amigas Nathara, Mariana, Raffaella, Raissa, Thamara, Mayara e Camila, pelos momentos de alegria e distração.

Ao professor Ronaldo, pela orientação e aprendizado por esses meses que trabalhamos juntos.

Aos professores Marcel e Augusto por participarem da banca e todos os professores do departamento pelas aulas ministradas e conhecimento repassado.

Enfim a todos que contribuíram de alguma forma para este trabalho. Muito obrigada.

“Entre quantidade e qualidade existe a geometria”.

Brigitte Le Roux & Henry Rouanet

Sumário

Lista de Figuras	8
Lista de Tabelas	9
1 Introdução	10
2 Banco de Dados	12
3 Metodologia	15
3.1 Matriz indicadora	16
3.1.1 Nuvem de pontos	19
3.1.2 Elementos suplementares	20
3.1.3 Ponto médio da nuvem	20
3.1.4 Distância	20
3.1.5 Variância da nuvem de pontos	22
3.1.6 Eixos principais	24
3.1.7 Coordenadas principais	24
3.1.8 Contribuição do ponto para os eixos principais	25
3.1.9 Contribuição relativa	26
3.1.10 Fórmulas de transição	27
4 Resultados e Análises	28
5 Conclusão	37
I Anexo	38

Lista de Figuras

3.1	Matriz indicadora	16
3.2	Exemplo de matriz indicadora	17
3.3	Exemplo de nuvem de pontos: indivíduos classificados a partir de Q variáveis, com um total de K categorias	19
3.4	Distâncias de pontos a um ponto qualquer P , sendo G o ponto médio . . .	22
3.5	Ângulo (θ) formado pelos vetores GM e GM'	26
4.1	Nuvem de indivíduos	30
4.2	Nuvem das variáveis	31
4.3	Nuvem das categorias	33
4.4	Nuvem das categorias e indivíduos	36
I.1	Parecer do Comitê de Ética	38
I.2	Carta Convite	39

Lista de Tabelas

4.1	Variância dos eixos e suas porcentagens, taxas de variância, e taxas modificadas e suas porcentagens	29
4.2	Coordenadas principais e contribuições dos indivíduos	29
4.3	Coordenadas principais e contribuições das categorias	32
4.4	Contribuição relativa	35

1 Introdução

Segundo dados da Associação Brasileira de Transplante de Órgãos (ABTO), o número de transplantes de córnea no Brasil, em geral, cresce a cada ano, e, aparentemente, vem se estabilizando nos últimos 2 anos. O recente crescimento do número de doações de córnea pode estar associado à boa qualidade da córnea, que, por sua vez, pode estar associada à rápida notificação e preservação, idade do doador, causa do óbito, entre outras características (SANO et al, 2010).

Nesta monografia, aplicaremos uma técnica estatística denominada Análise de Correspondência Múltipla (ACM), seguindo a abordagem de análise geométrica de dados, à base de dados sobre transplante de córnea, onde as linhas representam as córneas coletadas de doadores após o óbito e as colunas representam as categorias das diferentes variáveis.

A Análise de Correspondência (AC) se popularizou na década de 60 e 70 através do estatístico francês Jean-Paul Benzécri, primeiramente na França e se estendendo pela Europa, porém teve suas origens em trabalhos do início do século XX feitos por Pearson e Fisher, dois dos mais importantes nomes da estatística (FOX, 2010)

Tratando a ACM pela visão geométrica, podemos compreender melhor a técnica de forma intuitiva sem envolver necessariamente o labirinto de decomposições de matrizes da álgebra linear. Uma característica da ACM é a análise gráfica, através dela podemos verificar similaridades entre as categorias de uma variável, associações entre as categorias de diferentes variáveis, e entre as categorias e os indivíduos. Estudando os indivíduos, podemos verificar as similaridades ou dissimilaridades entre eles em termos de todas as variáveis, ou seja, indivíduos são comparados com base nas categorias nas quais eles estão classificados. As variáveis também podem ser estudadas em termos de categorias. Desta forma, podemos estudar as categorias representando as variáveis e grupos de indivíduos.

Nesta monografia, seguiremos, basicamente, as seguintes etapas para a análise dos dados: (i) preparação da tabela de dados para ACM; (ii) resultados básicos de análise como variância, coordenadas, contribuições e nuvens de pontos correspondentes às variáveis e indivíduos; (iii) decisão de quantos eixos de projeção e suas interpretações

em relação às nuvens; e, por último, (iv) investigação das nuvens.

Sendo assim, no capítulo 2 será descrita a base de dados contendo informações sobre doadores de córnea, que será utilizada como aplicação da metodologia. No capítulo 3 é apresentada a metodologia utilizada, que consiste na Análise de Correspondência Múltipla, enfatizando a análise geométrica de dados. No capítulo 4 os resultados e análises são apresentados através de gráficos e tabelas. E, finalmente, no capítulo 5, apresentaremos as conclusões.

2 Banco de Dados

Um conjunto de dados reais obtido do Banco de Olhos do Hospital Regional João Penido, será utilizado para efeito de aplicação da técnica Análise de Correspondência Múltipla. O parecer favorável do comitê de ética em pesquisa N°004/2010 encontra-se no anexo I. O banco de dados traz informações dos doadores de córnea, que autorizaram a doação em vida, sendo que as informações colhidas consistem no que foi preenchido no prontuário médico de cada doador.

A base inicial era constituída por 622 linhas, onde cada linha representava uma córnea (do olho direito ou do olho esquerdo). Em geral, cada 2 linhas representavam um indivíduo diferente. A partir de informações, tais como, se um indivíduo apresentasse a mesma classificação da qualidade da córnea tanto do olho direito quanto para o olho esquerdo, a córnea do olho direito que seria escolhida para um possível transplante; e, se um indivíduo apresentasse classificações diferentes da córnea do olho direito e do esquerdo, a córnea escolhida para um possível transplante seria a que apresentasse melhor classificação, e considerando somente os possíveis doadores, obtivemos 264 indivíduos (unidades de observação, ou objetos) e 8 variáveis: idade, local da coleta da córnea, gênero do indivíduo, tempo em horas do óbito à enucleação, tempo em horas da enucleação à preservação, tempo em horas do óbito à preservação, classificação da qualidade da córnea e causa do óbito. Os possíveis doadores mencionados correspondem aos indivíduos que tiveram alguma classificação da córnea; aqueles que não obtiveram a classificação, as córneas foram coletadas, porém não chegaram a ser examinadas na lâmpada de fenda pelo oftalmologista, decorrente do doador apresentar contra indicação absoluta ao transplante como tumor cerebral, hepatite C entre outros.

As variáveis local da coleta da córnea, gênero, causa do óbito e classificação da qualidade da córnea são variáveis qualitativas nominais, enquanto as demais são quantitativas. Pelo fato da ACM ser utilizada com variáveis categóricas, fizemos a categorização das mesmas: a variável idade foi categorizada segundo Szaflik, Grabska-Liberek & Brix-Warzecha (2003) onde define-se um grande grupo ≤ 40 anos e os anos posteriores por décadas, pois a cada década após os 40 anos verifica-se uma perda da quantidade de células endoteliais da córnea, que é uma característica que está associada à sua classificação; a

variável tempo em horas do óbito à enucleação foi categorizada segundo resolução nº 27 da Agência Nacional de Vigilância Sanitária (2008); e as variáveis tempo em horas da enucleação à preservação e tempo em horas do óbito à preservação foram categorizadas segundo seus quartis. Apesar da variável causa do óbito ser qualitativa, observamos muitas categorias e algumas categorias com baixa frequência, por isso agrupamos as categorias menos frequentes para uma nova categoria denominada “outras”.

Vale ressaltar que a variável tempo decorrido entre o óbito e a preservação é a soma das variáveis tempo em horas do óbito à enucleação e tempo em horas da enucleação à preservação.

O quadro 2.1 compreende as descrições das variáveis contidas no banco de dados utilizado.

Quadro 2.1 - Descrição das variáveis disponíveis

Variáveis	Descrição	Categorias
I	Idade do indivíduo	I 1: Até 40 anos
		I 2: 41 a 50 anos
		I 3: 51 a 60 anos
		I 4: 61 a 70 anos
		I 5: 71 a 80 anos
		I 6: 80 anos ou mais
L	Local de coleta da córnea	L 1: Juiz de Fora
		L 2: Zona da Mata
G	Gênero do indivíduo	G 1: Masculino
		G 2: Feminino
TE	Tempo em horas do óbito à enucleação	TE 1: Até 6 horas
		TE 2: Mais de 6 horas
TP	Tempo em horas da enucleação à preservação	TP 1: Até 3,79 horas
		TP 2: De 3,80 a 6,50 horas
		TP 3: De 6,51 a 10,54 horas
		TP 4: Mais de 10,54 horas
TT	Tempo em horas do óbito à preservação	TT 1: Até 7,25 horas
		TT 2: De 7,26 a 10,25 horas
		TT 3: De 10,26 a 15 horas
		TT 4: Mais de 15 horas
Q	Classificação da qualidade da córnea	Q 1: Aceitável
		Q 2: Inaceitável
O	Causa do óbito	O 1: Doenças do aparelho circulatório
		O 2: Doenças do aparelho respiratório
		O 3: Neoplasias (tumores)
		O 4: Causas externas de morbidade e mortalidade
		O 5: Sintomas, sinais e achados anormais de exames clínicos e de laboratório não classificados em outra parte
		O 6: Outras

3 Metodologia

A Análise de Correspondência é uma técnica de análise multivariada, desenvolvida para o estudo da relação entre variáveis qualitativas. O objetivo desta técnica é mostrar geometricamente as variáveis, suas categorias e os objetos observados na base de dados em um espaço de baixa dimensão, de modo que a proximidade no espaço indica associação entre as linhas e colunas (GREENACRE, 2010). O uso da AC tem se tornado muito comum devido a sua fácil implementação através de pacotes estatísticos computacionais com inúmeros recursos de análise estatística e gráfica.

Em Análise de Correspondência Simples (ACS), a aplicação é feita em tabelas de contingência, sendo essas, tabelas de duas entradas, onde nas linhas são apresentadas as categorias de uma variável, nas colunas as categorias de outra variável e nas células as frequências observadas de objetos ou indivíduos. A representação gráfica se dá para as categorias das variáveis e por cada variável (GREENACRE, 2007).

Em Análise de Correspondência Múltipla, com a qual trabalharemos, a aplicação é feita em tabelas multidimensionais, onde as linhas representam os objetos observados (no caso, indivíduos) e as colunas as diferentes categorias de diferentes variáveis. Em ACM a representação gráfica pode ser feita para os indivíduos, para as variáveis, para as categorias e para as categorias e indivíduos (LE ROUX & ROUANET, 2010).

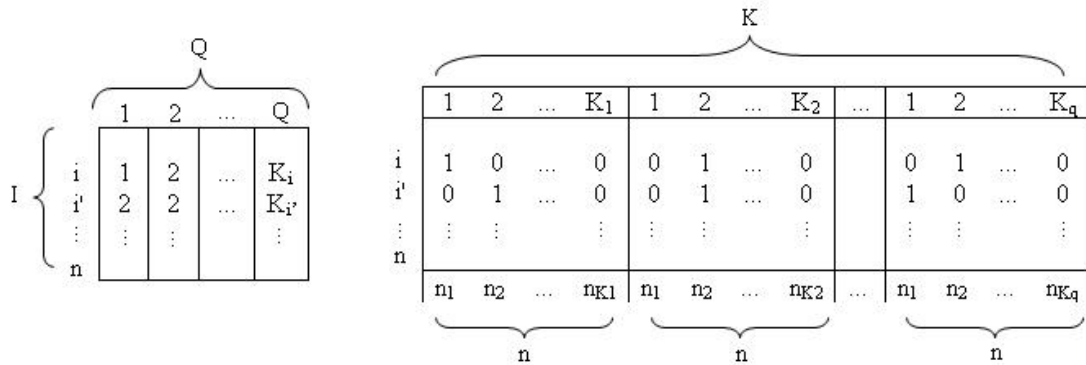
Podemos desenvolver a ACM através da matriz indicadora Z de dimensão $I \times K$, onde I representa as linhas, e K as colunas; e através da matriz quadrada de Burt $B = Z^T Z$. A última, consiste em transformar a matriz retangular em uma matriz quadrada simétrica composta por tabelas de contingência bidimensionais realizando todos os cruzamentos possíveis entre as variáveis envolvidas. A matriz indicadora, a qual utilizaremos, será apresentada na seção seguinte.

Para a implementação da metodologia utilizamos o pacote *FactoMineR* versão 1.14 (HUSSON et al, 2010) através do *software* livre R versão 2.13.0 (2011).

3.1 Matriz indicadora

Para explicarmos a ACM, adotaremos as notações, conceitos, explicações dos mesmos através de uma tabela representada por Indivíduos x Questões baseado em Le Roux & Rouanet (2010). As linhas correspondem aos indivíduos e as colunas às possibilidades de resposta a diferentes questões. As questões são variáveis categorizadas com número finito de categorias denominadas modalidades. Cada indivíduo poderá escolher uma e somente uma categoria por questão, representada por código binário: 1 para a categoria escolhida de uma questão e 0 para as demais categorias dessa mesma questão.

Figura 3.1: Matriz indicadora



Onde: I representa o grupo de n indivíduos; Q o grupo de questões; K_q o número de categorias da questão q ; K o número de total de categorias e K_i representa o padrão de resposta do indivíduo i , que é, o grupo das Q categorias escolhidas pelo indivíduo i . Chamaremos esta matriz indicadora de Z .

Para facilitar a compreensão, a Figura 3.2 representa um exemplo de uma matriz indicadora, onde são considerados dois indivíduos (i e i') e três questões (A, B e C), portanto $Q=3$, sendo duas categorias na primeira questão ($A=\{a1, a2\}$), portanto $K_a = 2$, duas na segunda questão ($B=\{b1, b2\}$), portanto $K_b = 2$, e três categorias na terceira questão ($C=\{c1, c2, c3\}$), portanto $K_c = 3$, temos então um total de $K=7$ categorias. As categorias $a1, b2, c2$ representam o padrão de resposta do indivíduo i , enquanto as categorias $a2, b1, c3$, representam o padrão de resposta do indivíduo i' .

Figura 3.2: Exemplo de matriz indicadora

Fonte: Le Roux & Rouanet (2010)

Linha e coluna marginal

Seja $i = \{1, 2, \dots, n\}$ e $k = \{1, 2, \dots, K\}$:

A linha marginal da matriz indicadora Z , é a soma das observações na k -ésima coluna:

$$z_{.k} = \sum_{i=1}^n z_{ik}, \quad (3.1)$$

que representa o número de indivíduos que escolheram a categoria k .

A coluna marginal de Z é a soma das observações da i -ésima linha:

$$z_{i.} = \sum_{k=1}^K z_{ik}, \quad (3.2)$$

que representa o número de questões do questionário, sendo $z_{i.} = Q \forall i = \{1, 2, \dots, n\}$

O total representado por $z_{..}$ é dado por:

$$z_{..} = \sum_{i=1}^n z_{i.} = \sum_{k=1}^K z_{.k} = nQ, \quad (3.3)$$

que representa o número de indivíduos multiplicado pelo número de questões.

Massa de um elemento

A massa de um elemento $i \in I$ é o quociente do total da i -ésima linha pelo total geral:

$$r_i = \frac{z_{i.}}{z_{..}} = \frac{Q}{nQ} = \frac{1}{n}, \quad (3.4)$$

Podemos notar que a massa depende da quantidade de indivíduos, e é a mesma $\forall i = \{1, 2, \dots, n\}$. A massa de um elemento $k \in K$ é o quociente do total da k -ésima coluna pelo total geral:

$$c_k = \frac{z_{.k}}{z_{..}} = \frac{n_k}{nQ}, \quad (3.5)$$

que corresponde ao quociente do número de indivíduos que escolheram a k -ésima categoria pelo número total de indivíduos multiplicado pelo número de questões.

Perfil

Um perfil do vetor de linha i que tem k elementos k é dado por:

$$r_K^i = \frac{z_{ik}}{z_{i.}} = \frac{z_{ik}}{Q}. \quad (3.6)$$

Os perfis de linhas são uma “recodificação” da matriz indicadora Z , onde todos elementos que não são 0 são iguais a Q^{-1} . Um perfil do vetor de categoria k que tem i elementos é dado por:

$$c_I^k = \frac{z_{ik}}{z_{.k}} = \frac{z_{ik}}{n_k}. \quad (3.7)$$

Os perfis de colunas podem ser considerados como uma primeira quantificação das categorias das variáveis qualitativas, uma vez que os valores são pesos relativos de cada categoria dentro da respectiva variável.

O perfil da linha marginal corresponde às massas dos elementos $k \in K$. E o perfil da coluna marginal corresponde às massas dos elementos $i \in I$.

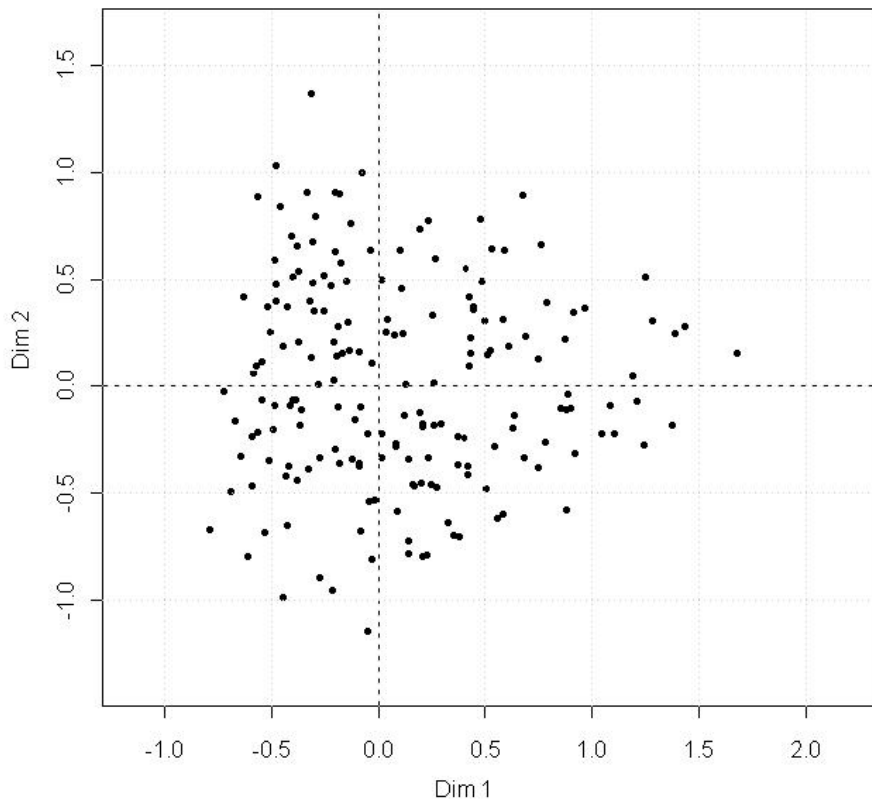
3.1.1 Nuvem de pontos

Segundo Souza (2004), a nuvem de pontos dos indivíduos é formada a partir do conjunto dos perfis de cada linha, cada um associado à sua massa. Da mesma forma, a nuvem de pontos das categorias é formada a partir do conjunto de perfis de cada coluna, cada um associado à sua massa. A interpretação é a mesma para as duas nuvens: perfis próximos e massas próximas resultam em pontos próximos na nuvem. A dimensionalidade da nuvem é L :

$$L \leq K - Q, \quad (3.8)$$

que corresponde o número total de categorias menos o número total de questões. Uma nuvem de pontos pode representar tanto indivíduos quanto variáveis e categorias. A Figura 3.2 mostra especificamente uma nuvem de indivíduos.

Figura 3.3: Exemplo de nuvem de pontos: indivíduos classificados a partir de Q variáveis, com um total de K categorias



3.1.2 Elementos suplementares

A construção das nuvens é baseada em indivíduos e categorias, sendo que estes são chamados indivíduos ativos e categorias ativas, respectivamente. Elementos que distorcem a análise (distorcendo a configuração do gráfico) e representam informações adicionais, podem ser considerados elementos suplementares, os quais são representados graficamente, porém, não fazem parte da solução (BENZÉCRI, 1992; GREENACRE, 1984).

3.1.3 Ponto médio da nuvem

Seja P um ponto qualquer no espaço e $(M^i)_{i=1,2,\dots,n}$ pontos da nuvem, o ponto médio da nuvem, denominado G , pode ser encontrado pelo vetor: $\overrightarrow{PG} = \sum \overrightarrow{PM^i}/n$, que consiste no somatório das distâncias entre o ponto P qualquer e todos os pontos da nuvem dividido pelo número total de pontos. O ponto médio não depende da escolha do ponto P .

Se substituirmos P por G , temos um vetor nulo:

$$\frac{1}{n} \sum \overrightarrow{GM^i} = \vec{0}. \quad (3.9)$$

Com isso, podemos definir G como as médias das coordenadas dos pontos:

$$G = \sum M^i/n. \quad (3.10)$$

3.1.4 Distância

Distância entre indivíduos

Podemos calcular a distância entre dois indivíduos somente quando ambos escolherem diferentes categorias de uma mesma questão, pois quando isso não ocorre, ou seja, quando os indivíduos escolherem a mesma categoria, a distância, denotada por $d_q(i, i')$, passa a ser nula. Logo, na representação gráfica, os dois pontos representados pelos diferentes indivíduos coincidem.

A distância ao quadrado entre indivíduos, que escolheram categorias diferentes, ou seja, o indivíduo i escolheu a categoria k e o indivíduo i' escolheu a categoria k' , é

dada por:

$$d_q^2(i, i') = \frac{1}{f_k} + \frac{1}{f_{k'}}, \quad (3.11)$$

onde f_k representa a frequência relativa de indivíduos que escolheram a categoria k , com $f_k = n_k/n$.

O total da distância ao quadrado entre indivíduos é dado por:

$$d^2(i, i') = \frac{1}{Q} \sum_{q \in Q} d_q^2(i, i'). \quad (3.12)$$

Segundo Le Roux & Rouanet (2010, p. 35-36, tradução nossa), “quanto menor a frequência das diferentes categorias, maior é a distância entre indivíduos”. Sendo assim, o ponto M^i ficará longe do centro, localizando-se na periferia da nuvem.

O peso de um ponto referente a um indivíduo i (M^i) é Q (coluna marginal - expressão 3.2), a soma dos pesos para o total de indivíduos é nQ (expressão 3.3). Em termos de peso relativo, temos $p_i = Q/nQ = 1/n$ (massa de um elemento $k \in K$ - expressão 3.4).

Distância entre categorias

O peso de um ponto referente a uma categoria k (M^k) é n_k (linha marginal - expressão 3.1), a soma dos pesos para cada categoria de uma dada questão é n e para todas o total de categorias é nQ (expressão 3.3). Em termos de peso relativo de uma categoria, temos $p_k = n_k/nQ = f_k/Q$ (massa de um elemento $k \in K$ - expressão 3.5), sendo assim, a soma dos pesos relativos em cada questão é $1/Q$, e para todas as questões é 1.

Seja $n_{kk'}$ o número de indivíduos que escolheram ambas categorias k e k' , a distância ao quadrado entre M^k e $M^{k'}$ é dado por:

$$(M^k M^{k'})^2 = \frac{n_k + n_{k'} - 2n_{kk'}}{n_k n_{k'} / n}. \quad (3.13)$$

Se k e k' são duas diferentes categorias de uma mesma questão, então $n_{kk'} = 0$. Segundo Le Roux & Rouanet (2010, p. 38, tradução nossa), “quanto mais categorias k e

k' forem escolhidas pelo mesmo indivíduo, menor é a distância entre M^k e $M^{k'}$; e, quanto menor a frequência da categoria k , mais o ponto M^k se distancia do centro da nuvem”.

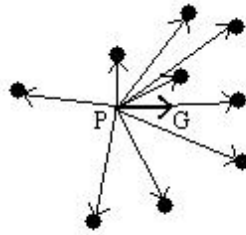
3.1.5 Variância da nuvem de pontos

Uma medida de dispersão da nuvem de pontos em relação à um ponto qualquer é a variância. Se esse ponto for o centro de gravidade (ponto médio da nuvem), a variância se torna mínima. A variância de uma nuvem de pontos é dada por:

$$V_{nuvem} = \frac{1}{n} \sum (PM)^2 - (PG)^2, \quad (3.14)$$

que corresponde à média das distâncias ao quadrado dos pontos da nuvem à um ponto P , menos a distância ao quadrado entre o ponto médio (G) e o ponto P . Podemos ver essa relação através da Figura 3.3.

Figura 3.4: Distâncias de pontos a um ponto qualquer P , sendo G o ponto médio



Sendo o ponto P o ponto médio da nuvem ($P=G$), temos:

$$V_{nuvem} = \frac{1}{n} \sum (GM)^2. \quad (3.15)$$

Variância da nuvem de indivíduos

A distância ao quadrado do ponto M^i ao ponto G é:

$$(GM^i)^2 = \left(\frac{1}{Q} \sum_{k \in K_i} \frac{1}{f_k} \right) - 1, \quad (3.16)$$

onde K_i representa a resposta padrão do indivíduo i , que é, o grupo das Q categorias escolhidas pelo indivíduo i .

Segundo demonstração de Le Roux & Rouanet (2004, p. 186-187):

$$V_{nuvem_i} = \sum_{i \in I} \frac{(GM^i)^2}{n} = \frac{1}{nQ} \sum_{i \in I} \sum_{k \in K} \frac{z_{ik}}{f_k} - \sum_{i \in I} \frac{1}{n} = \frac{1}{nQ} \sum_{k \in K} \frac{n_k}{f_k} - 1 = \frac{K}{Q} - 1. \quad (3.17)$$

Variância da nuvem de categorias

A distância ao quadrado do ponto M^k ao ponto G é:

$$(GM^k)^2 = \frac{1}{f_k} - 1. \quad (3.18)$$

Sendo f_k/Q , o peso relativo p_k do ponto M^k , a variância da categoria k pode ser demonstrada, segundo Husson, Lê & Pagés (2010, p. 137-138), como sendo:

$$V_k = p_k(GM^k)^2 = \frac{f_k}{Q} \left(\frac{1}{f_k} - 1 \right) = \frac{1 - f_k}{Q} = \frac{1}{Q} \left(1 - \frac{n_k}{n} \right) \quad (3.19)$$

Portanto, a variância de todas as K_q categorias da variável q é:

$$V_q = \sum_{k=1}^{K_q} \frac{1}{Q} \left(1 - \frac{n_k}{n} \right). \quad (3.20)$$

Como, $\sum_{k=1}^{K_q} n_k = n$, obtemos:

$$V_q = \frac{K_q - 1}{Q}. \quad (3.21)$$

Concluimos, que a variância associada à todas as categorias, que corresponde à nuvem de categorias, é dada por:

$$V_{nuvem_k} = \sum_{q=1}^Q \frac{K_q - 1}{Q} = \frac{K}{Q} - 1. \quad (3.22)$$

Portanto, a nuvem de categorias tem a mesma variância da nuvem de indivíduos. A frequência de indivíduos em uma certa categoria influencia em sua con-

tribuição: categorias infrequentes contribuem mais para a variância global, tornando assim recomendável, sempre que possível, o agrupamento de categorias com frequência abaixo de 5%; e quanto mais categorias uma questão possui, mais esta contribui para a variância da nuvem (LE ROUX & ROUANET, 2010).

3.1.6 Eixos principais

A projeção de uma nuvem corresponde à projeção ortogonal de seus pontos. Portanto, a variância de uma nuvem ortogonalmente projetada é sempre menor ou igual à variância da nuvem inicial.

A abordagem geométrica quanto à formação dos eixos principais corresponde primeiramente na projeção da nuvem de pontos em eixos arbitrários, através das distâncias entre os pontos. Então, é feita uma rotação de uma reta, que passa pelo ponto médio da nuvem, que minimiza as distâncias perpendiculares dos pontos à reta, e maximiza a variância da nuvem projetada (captando assim maior variabilidade dos pontos). Essa reta é denominada primeiro eixo principal. Para obtenção do segundo eixo principal, determina-se uma nova reta, passando também pelo ponto médio da nuvem (G) e perpendicular ao primeiro eixo principal. Da mesma forma, conseguimos obter o terceiro eixo principal, o quarto eixo principal, e assim por diante.

A variância de cada eixo é chamada variância do eixo 1, variância do eixo 2, ..., variância do eixo ℓ , *ou primeiro autovalor, segundo autovalor, ..., também denominada por* $\lambda_1, \lambda_2, \dots, \lambda_\ell$. A soma dos autovalores é igual à variância da nuvem é dada por: $\sum \lambda_\ell = V_{nuvem}$, com $\ell = 1, 2, \dots, L$.

O primeiro eixo principal oferece o melhor ajuste unidimensional da nuvem. Da mesma forma, o plano gerado pelo eixo 1 e 2, oferece o melhor ajuste bidimensional, e assim por diante.

3.1.7 Coordenadas principais

As coordenadas principais dos pontos definem a nuvem referida aos seus eixos principais. As coordenadas principais dos pontos referentes aos indivíduos M^i relativo ao eixo principal l são denominadas y_l^i e as coordenadas dos pontos referentes às categorias M^k são denominadas y_l^k .

3.1.8 Contribuição do ponto para os eixos principais

A contribuição de um ponto a um determinado eixo, consiste na importância desse ponto ao eixo e é calculada através da multiplicação do peso desse ponto e sua coordenada nesse eixo dividido pela variância do eixo. Sendo p , o peso relativo do ponto, e y a coordenada relativa ao eixo de variância λ , a contribuição do ponto a um eixo é dada por:

$$Ctr = \frac{(py^2)}{\lambda}. \quad (3.23)$$

De (3.23) podemos definir a contribuição do ponto referente aos indivíduos (M^i) como sendo:

$$Ctr_i = \frac{\frac{1}{n}(y^i)^2}{\lambda}. \quad (3.24)$$

Da mesma forma, a contribuição do ponto referente às categorias (M^k) é:

$$Ctr_k = \frac{\frac{f_k}{Q}(y^k)^2}{\lambda}. \quad (3.25)$$

Taxas de variância e taxas modificadas

Em ACM obtemos uma nuvem com alta dimensionalidade, pois vimos que esta depende do número total de categorias e número de variáveis, sendo assim, as taxas de variância dos eixos principais são baixas. A taxa de variância é dada por:

$$\tau_\ell = \frac{\lambda_\ell}{V_{nuvem}} = \frac{\lambda_\ell}{\frac{K}{Q} - 1}. \quad (3.26)$$

Para valorizar os primeiros eixos, é proposto por Benzécri (1992) utilizar taxas modificadas:

$$\tau_{\ell'} = \frac{\lambda'_\ell}{S}, \quad (3.27)$$

onde, $\lambda'_\ell = \left(\frac{Q}{Q-1}\right)^2 (\lambda_\ell - \bar{\lambda})^2$; $\bar{\lambda} = \left(\frac{K}{Q} - 1\right) / (K - Q) = 1/Q$; $S = \sum_{\ell=1}^{\ell_{\max}} \lambda'_\ell$

Vale ressaltar que S corresponde à soma dos autovalores que apresentaram valores acima da média dos autovalores.

Segundo Greenacre & Blasius (2006, p. 151, tradução nossa), “As taxas modificadas indicam o quanto a nuvem se afasta de uma nuvem esférica, ou seja, aquela em que seus autovalores são iguais ao autovalor médio”.

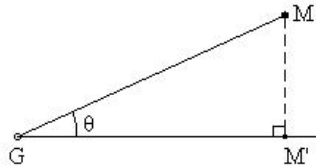
3.1.9 Contribuição relativa

A contribuição relativa consiste na contribuição do eixo para a variância do ponto, e é calculada da seguinte forma:

$$\frac{(GM')^2}{(GM)^2} = \cos^2 \theta. \quad (3.28)$$

A expressão 3.28 pode ser melhor compreendida através da Figura 3.5, que representa o cosseno do ângulo θ , dado pela razão entre a medida do cateto adjacente e a medida da hipotenusa.

Figura 3.5: Ângulo (θ) formado pelos vetores GM e GM'



Essa contribuição auxilia na interpretação de pontos suplementares, pois esses não contribuem para a variância total, por possuírem massa zero. Segundo Greenacre (2007), suas contribuições relativas, que se relacionam com os ângulos entre perfis e eixos e não envolve massa, podem ainda ser interpretadas para diagnosticar quão bem elas são representadas. A qualidade da representação consiste na soma dos cossenos ao quadrado para cada eixo, de forma que a qualidade de representação de um ponto no plano corresponde à soma do cosseno ao quadrado no eixo 1 e o cosseno ao quadrado no eixo 2.

3.1.10 Fórmulas de transição

As duas fórmulas que serão apresentadas a seguir tem como finalidade “ligar” as coordenadas principais referentes aos indivíduos y^i e as coordenadas principais referente às categorias y^k . São elas:

$$y^i = \frac{1}{\sqrt{\lambda}} \sum_{k \in K_i} \frac{y^k}{Q}, e \quad (3.29)$$

$$y^k = \frac{1}{\sqrt{\lambda}} \sum_{i \in I_k} \frac{y^i}{n_k}. \quad (3.30)$$

A primeira fórmula de transição (3.31) permite calcular as coordenadas principais de um indivíduo cujas respostas a perguntas ativas são conhecidas, e localizar este indivíduo na nuvem de categorias. A segunda fórmula de transição (3.32) permite calcular as coordenadas principais de qualquer categoria a qual pertencem um grupo de indivíduos, e localizar esta categoria na nuvem de indivíduos. A partir de (3.32) podemos também obter as coordenadas principais e conseqüentemente plotar as categorias de variáveis suplementares no gráfico, de forma a examinar sua localização.

4 Resultados e Análises

Conforme visto no capítulo 2, a variável tempo do óbito à preservação é a soma das variáveis tempo do óbito à enucleação e tempo da enucleação à preservação. Portanto, consideramos essa variável, correspondente ao tempo total do óbito à preservação, como suplementar, de forma a não participar nos cálculos. Porém, a mesma também será plotada no gráfico. Consideramos as demais variáveis como ativas. Desta forma, nossa base de dados possui $Q = 7$ variáveis ativas e $K = 6 + 2 + 2 + 2 + 4 + 2 + 6 = 24$ categorias, e como visto na seção 3.1.5, que a variância da nuvem de indivíduos é a mesma para a nuvem de categorias, temos portanto que a variância da nuvem de pontos é: $(24/7) - 1 = 2,429$ (expressão encontrada em 3.17). A dimensionalidade da nuvem é $24 - 7 = 17$ (expressão 3.8).

Como visto na seção 3.1.8, as taxas modificadas são calculadas de forma a corrigir a baixa variância em cada eixo devido à alta dimensionalidade. Para tal efeito, calculamos a média dos autovalores $\bar{\lambda} = 1/Q = 1/7 = 0,1428571$ e consideramos todas as variâncias acima deste valor, que foram 8, conforme a tabela 4.1. Calculamos a taxa de variância (expressão 3.26) e finalmente, as taxas modificadas (expressão 3.27).

Notamos que do primeiro autovalor para o segundo há uma queda de 15,23% $((0.2698-0.2287)/0.2698)$; do segundo para o terceiro há uma queda maior de 17,62% $((0.2287-0.1884)/0.2287)$; já do terceiro para o quarto a queda é de 5,52% $((0.1884-0.1780)/0.1884)$. E, analisando a coluna de taxas modificadas, vemos que ocorre uma queda brusca do segundo para o terceiro. Com esses resultados, decidimos interpretar somente os eixos 1 e 2, que são responsáveis por explicar 20,52% da variabilidade dos dados, observando pela coluna da porcentagem da variância; e, essa porcentagem aumenta para 34,83% observando pela coluna da porcentagem das taxas modificadas.

Na figura 4.1 os pontos representam os indivíduos, formando uma nuvem que não apresenta uma tendência de formação de subgrupos. Devido ao grande número de indivíduos citaremos apenas 4: os indivíduos #24, #187, #202 e #207. A tabela 4.2 apresenta as coordenadas e contribuições desses pontos para a inércia dos eixos e mais um ponto para efeito de comparação.

Tabela 4.1: Variância dos eixos e suas porcentagens, taxas de variância, e taxas modificadas e suas porcentagens

Eixo	Variância	% da variância	% acumulado da variância	Taxa de variância	Taxa modificada	% da taxa modificada	% acumulado da taxa modificada
1	0,2698	11,108	11,108	4,5740	0,5804	23,899	23,899
2	0,2287	9,416	20,524	3,8772	0,2654	10,927	34,826
3	0,1884	7,758	28,282	3,1944	0,0748	0,0308	34,8568
4	0,1780	7,328	35,61	3,0175	0,0444	0,0183	34,8751
5	0,1670	6,878	42,488	2,8320	0,0211	0,0087	34,8838
6	0,1585	6,528	49,016	2,6880	0,0089	0,0036	34,8874
7	0,1545	6,360	55,376	2,6188	0,0048	0,0020	34,891
8	0,1456	5,995	61,371	2,4687	0,0003	0,0001	34,8911

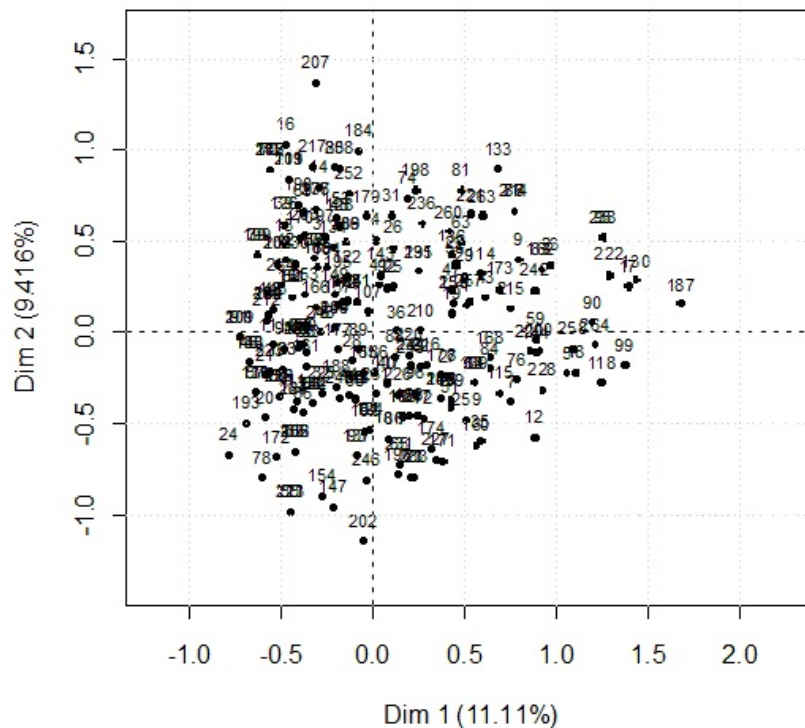
Os 4 pontos escolhidos se localizam nos extremos da nuvem e apresentam as maiores contribuições para a variância do eixo que os discriminam segundo seus lados. Por exemplo, a dimensão 1 conseguiu discriminar maximamente os indivíduos #24 (posicionado no lado esquerdo, com contribuição 0,88) e #187 (posicionado no lado direito, com contribuição 3,944), enquanto a dimensão 2 conseguiu discriminar os indivíduos #202 (posicionado no lado abaixo, com contribuição 2,151) e #207 (posicionado no lado acima, com contribuição 3,098).

Tabela 4.2: Coordenadas principais e contribuições dos indivíduos

Indivíduo	Coordenadas		Contribuição (em %)	
	Eixo 1	Eixo 2	Eixo 1	Eixo 2
12	0,880	-0,575	1,086	0,548
24	-0,792	-0,669	0,880	0,742
187	1,676	0,157	3,944	0,041
202	-0,052	-1,140	0,004	2,151
207	-0,315	1,368	0,139	3,098

Pelo eixo 1, os pontos 24, 202 e 207 se localizam no mesmo lado à esquerda,

Figura 4.1: Nuvem de indivíduos



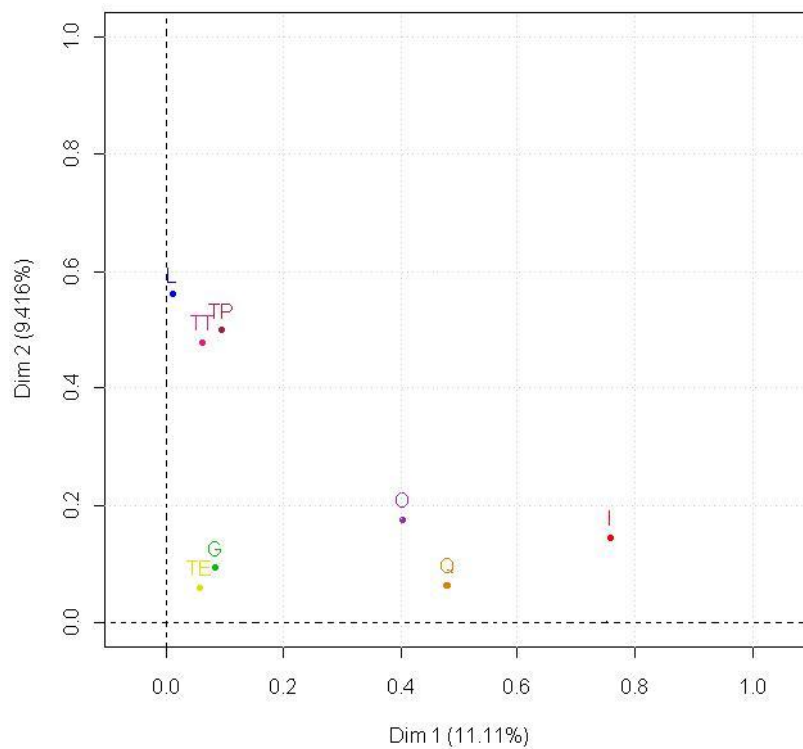
sendo o ponto 24 que mais contribui para variância do eixo 1. Os pontos 12 e 187 se localizam no mesmo lado à direita, e o ponto 187 que mais contribui para a variância do eixo 1 por se localizar mais distante deste em toda a nuvem.

Pelo eixo 2, os pontos 187 e 207 se localizam no mesmo lado acima, e o ponto 207 que mais contribui para a variância do eixo 2 por se localizar mais distante deste em toda a nuvem. Os pontos 12, 24 e 202 se localizam no lado abaixo, e o ponto 202 que mais contribui para a variância do eixo 2.

Na figura 4.2, os pontos representam as variáveis ativas e a variável suplementar. As variáveis idade (I), classificação (Q) e causa do óbito (O) estão relacionadas com o primeiro eixo, essas variáveis contribuem para tal eixo em 40,17%, 25,39% e 21,34% respectivamente (ver tabela 4.3). Já as variáveis procedência (L) e tempo de preservação (TP) estão relacionados com o segundo eixo, essas variáveis contribuem para tal eixo em 35,05% e 31,25% respectivamente (ver tabela 4.3). A variável tempo total representada por TT é a variável suplementar.

Nosso critério para interpretação dos eixos foi selecionar as categorias que contribuíram mais que a média das contribuições ($1/24 = 4,17\%$) ou valores muito próximos. Pela tabela 4.3, que apresenta as coordenadas e contribuições das categorias ativas, as

Figura 4.2: Nuvem das variáveis



categorias selecionadas no eixo 1, contribuem juntas 75,74%, enquanto as categorias selecionadas no eixo 2 contribuem 81,83%.

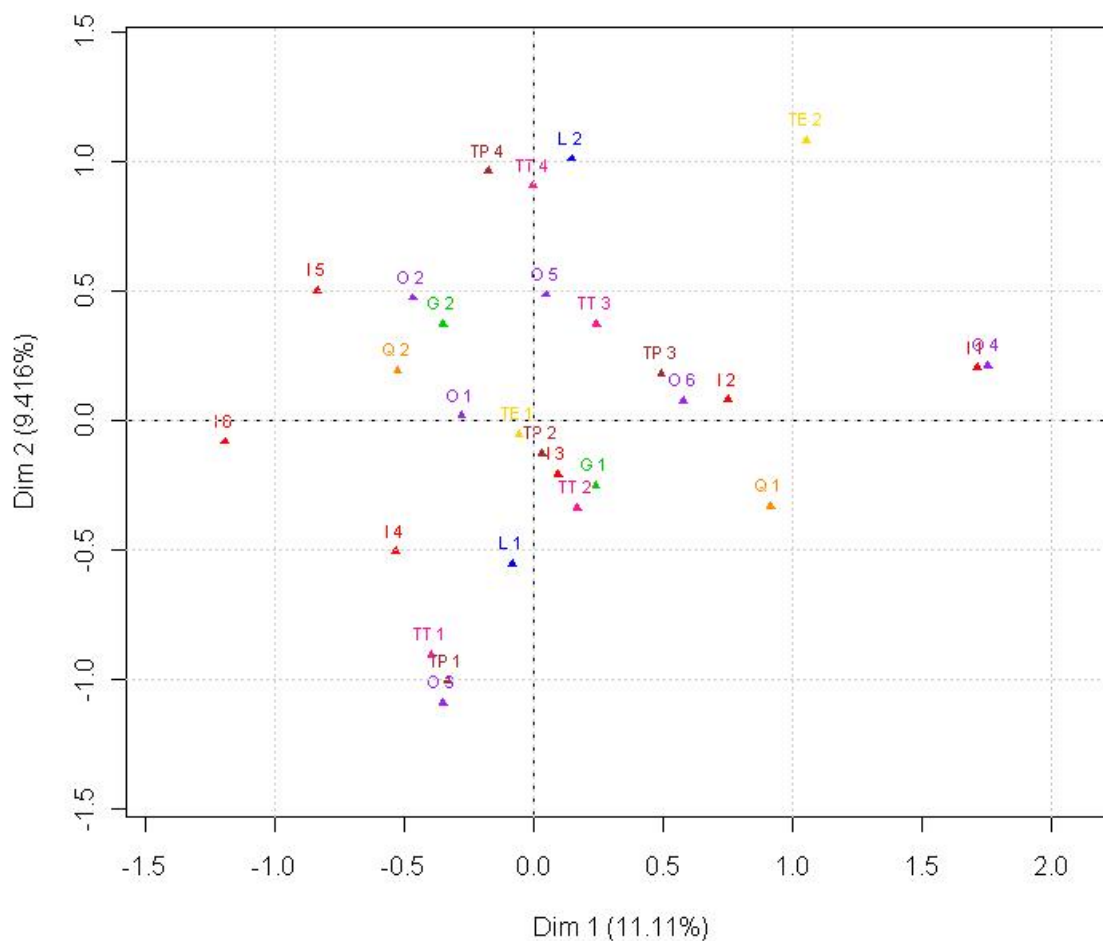
Na figura 4.3, os pontos representam as categorias das variáveis ativas e a variável suplementar. Nesta, observamos, com auxílio das categorias selecionadas na tabela 4.3, que para o eixo 1, as categorias: até 40 anos (I 1) e de 41 a 50 anos (I 2), da variável idade (I); aceitável (Q 1), da variável classificação da córnea (Q); e, causas externas de morbidade e de mortalidade (O 4), da variável causa do óbito (O), estão localizadas no lado à direita, com isso, podemos concluir que indivíduos com até 50 anos, estão relacionados com uma qualidade da córnea boa, de forma a ser aceitável, e com causas de óbito externas de morbidade e de mortalidade. Enquanto as categorias: 71 a 80 anos (I 5), da variável idade (I); e, inaceitável (Q 2), da variável classificação da córnea (Q), se localizam no lado à esquerda, com isso, podemos concluir que indivíduos de mais idade estão relacionados com uma qualidade pior da córnea, de forma a não ser aceitável para transplante.

Já para o eixo 2, as categorias: de 71 a 80 anos (I 5) da variável idade (I); Zona da Mata (L 2), da variável local da coleta da córnea (L); e, 10,55 ou mais (TP 4), da variável tempo em horas, da enucleação à preservação (TP) se localizam no lado

Tabela 4.3: Coordenadas principais e contribuições das categorias

Categorias	Coordenadas		Contribuição (em %)	
	Eixo 1	Eixo 2	Eixo 1	Eixo 2
I 1	1,715	0,201	21, 828	0,355
I 2	0,752	0,077	4, 080	0,051
I 3	0,096	-0,211	0,100	0,567
I 4	-0,533	-0,510	3,700	3, 999
I 5	-0,833	0,501	9, 608	4, 090
I 6	-1,192	-0,085	0,855	0,005
L 1	-0,082	-0,557	0,227	12, 479
L 2	0,148	1,007	0,411	22, 569
G 1	0,242	-0,255	1,827	2,403
G 2	-0,349	0,369	2,638	3,471
TE 1	-0,055	-0,056	0,150	0,186
TE 2	1,053	1,081	2,892	3,595
TP 1	-0,330	-1,009	1,446	15, 913
TP 2	0,034	-0,132	0,016	0,279
TP 3	0,494	0,180	3,085	0,481
TP 4	-0,174	0,959	0,405	14, 581
Q 1	0,916	-0,334	16, 157	2,538
Q 2	-0,523	0,191	9, 233	1,450
O 1	-0,276	0,015	1,991	0,007
O 2	-0,467	0,471	1,529	1,834
O 3	-0,348	-1,093	0,704	8, 195
O 4	1,755	0,208	14, 835	0,246
O 5	0,047	0,483	0,005	0,662
O 6	0,578	0,073	2,277	0,043

Figura 4.3: Nuvem das categorias



acima, com isso, podemos concluir que indivíduos de 71 a 80 anos que faleceram em outra localidade na Zona da Mata, estão relacionados com os maiores tempos da enucleação à preservação de suas córneas. Enquanto as categorias: de 61 a 70 anos (I 4), da variável idade (I); Juiz de Fora (L 1), da variável local da coleta da córnea (L); até 3,79 (TP 1), da variável tempo em horas da enucleação à preservação (TP); e, Neoplasias (tumores) (O 3), da variável causa do óbito (O), se localizam no lado abaixo, com isso, podemos concluir que indivíduos de 61 a 70 anos estão relacionados com falecimento em Juiz de Fora, de neoplasias (tumores), e com os menores tempos do óbito à enucleação de suas córneas.

Em suma, o primeiro eixo opõe-se aos mais jovens e com classificação aceitável da córnea com os mais velhos e com classificação inaceitável da córnea. O segundo eixo opõe-se aos que faleceram em Juiz de Fora e com os menores tempos da enucleação à preservação com os que faleceram em outras localidades na Zona da Mata e com os maiores tempos da enucleação à preservação.

As variáveis gênero e tempo em horas do óbito à enucleação, não foram mencionadas, pois as mesmas contribuem pouco para a variância nos eixos 1 e 2. Isso pode ter acontecido devido ao fato da variável sexo não estar mesmo associada a nenhuma das outras variáveis, talvez possa estar associada a causa de óbito, mas deveria ser feito um estudo para esse questionamento e verificar se o presente estudo não conseguiu apresentar essa associação ou se realmente não existe. Porém, apesar das categorias dessa variável apresentarem autovalores abaixo do critério escolhido, a localização da categoria G 2, gênero feminino, está no mesmo quadrante que as categorias I 5, TP 4, Q 2, O 2, O 1; e a categoria G 1, gênero masculino, está no mesmo quadrante que as categorias I 3, TP 2, Q 1, o que pode apresentar alguma associação mesmo que baixa. Já a não associação da variável tempo do óbito à enucleação com outras variáveis pode estar relacionada à diferença entre as frequências de cada uma de suas categorias, isto é, a categoria TE 1 possui 251 indivíduos e TE 2 possui 13 indivíduos e categorias que apresentam uma grande frequência “puxam” o centro para si. A porcentagem de contribuição dessas variáveis são altas em outras dimensões, por exemplo, a categoria feminino da variável gênero contribui 4,02% para o eixo 3, e a categoria acima de 6,01 da variável tempo em horas do óbito à enucleação contribui 10,56% para o eixo 5. Porém, as variâncias nesses eixos são menores, 0,1884 no eixo 3 e 0,1670 no eixo 5 (ver tabela 4.1), e diante das contribuições das outras variáveis nesses eixos, essas porcentagens não foram suficientes para pensarmos em uma possível interpretação dessas variáveis nesses eixos.

Quanto à variável suplementar, tempo em horas do óbito à preservação, apesar de suas categorias não contribuírem para a variância dos eixos, suas contribuições relativas, que compreendem a contribuição do eixo para o ponto, ainda podem ser interpretadas de forma a verificar como estão representadas. A tabela 4.4 apresenta as contribuições relativas das categorias suplementares para os eixos 1 e 2, e a qualidade em 2 dimensões.

Notamos que os valores para o eixo 2 são maiores quando comparados com o eixo 1, isso indica que a variável é melhor representada pelo eixo 2. Para a categoria TT 1, a qualidade da representação no plano é 36,7%; para a categoria TT 2, 7,36%; para a categoria TT 3, 10,02%; e para a categoria TT 4, 30,38%.

Apesar da ACM considerar categorias nominais, observamos que a ordenação das categorias das variáveis tempo em horas do óbito à enucleação e idade foi mantida. Esta é uma confirmação da adequação da variável para a mensuração ordinal.

Tabela 4.4: Contribuição relativa

Categorias	Contribuição relativa		Qualidade em 2 dimensões
	Eixo 1	Eixo 2	
TT 1	0,0583	0,3087	0,367
TT 2	0,0147	0,0589	0,0736
TT 3	0,0301	0,0701	0,1002
TT 4	0,0000	0,3038	0,3038

A figura 4.4 representa os indivíduos, na cor preta, e as categorias, cada cor indicando uma variável. A posição de um indivíduo reflete a quais categorias o mesmo está associado.

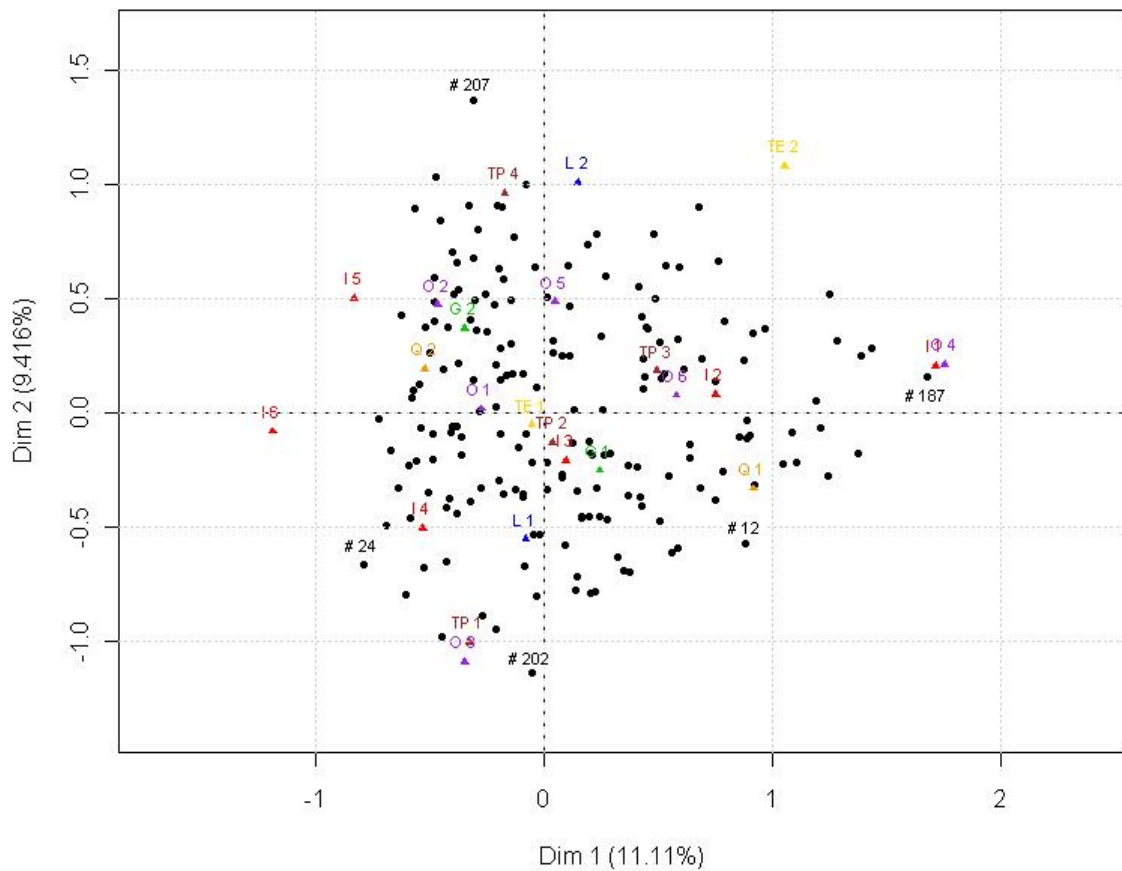
Considerando os 4 indivíduos selecionados, os que estão mais próximos, ainda assim não muito, são os indivíduos #24 e #202. Estes compartilham as categorias: I 4, L 1, TP 1, O 3, o que reflete a distância entre eles, desta forma indivíduos mais próximos compartilham mais categorias. Indivíduos muito distantes não compartilham nenhuma ou muito poucas categorias, como os indivíduos #187 e #207 que compartilham apenas uma categoria, TE 2.

Conforme mencionado na seção 3.1.10, as fórmulas de transição permitem localizar qualquer indivíduo a partir da nuvem de categorias (expressão 3.29), e localizar qualquer categoria a partir da nuvem de indivíduos (expressão 3.30). Faremos, a título de exemplo, a localização do indivíduo #187 na nuvem de categorias e a localização da categoria TE 2 na nuvem de indivíduos.

O indivíduo #187 está associado às seguintes categorias: I 1, L 1, G 1, TE 2, TP 3, Q 1 e O 4, com isso conseguimos calcular o ponto médio dessa nuvem de categorias (expressão 3.11): para o eixo 1, $(1,715 - 0.082 + 0.242 + 1.053 + 0.494 + 0.916 + 1,755)/7 = 0,87$; dividindo esse resultado por $\sqrt{\lambda_1}$, obtemos a coordenada no eixo 1: $0,87/\sqrt{0,2698} = 1,674936$ (ver tabela 4.2). Para o eixo 2, $(0,201 - 0.557 - 0.255 + 1,081 + 0,18 - 0,334 + 0,208)/7 = 0,075$; dividindo esse resultado por $\sqrt{\lambda_2}$, obtemos a coordenada no eixo 2: $0,075/\sqrt{0,2287} = 0,1568297$ (ver tabela 4.2).

Considerando a categoria TE 2, temos 13 indivíduos que pertencem a essa categoria, a partir de suas coordenadas, conseguimos calcular o ponto médio dessa nuvem

Figura 4.4: Nuvem das categorias e indivíduos



de 13 indivíduos: para o eixo 1, $(0,511 + 0,791 + 1,387 + 0,272 + 0,192 - 0,205 + 0,679 + 0,691 - 0,079 + 1,676 + 0,23 - 0,315 + 1,28)/13 = 0,5469231$; dividindo esse resultado por $\sqrt{\lambda_1}$, obtemos a coordenada no eixo 1: $0,5469231/\sqrt{0,2698} = 1,052554$ (ver tabela 4.3). Para o eixo 2, $(0,15 + 0,397 + 0,25 - 0,469 + 0,737 + 0,906 + 0,898 + 0,236 + 0,999 + 0,157 + 0,779 + 1,368 + 0,313)/13 = 0,517$ dividindo esse resultado por $\sqrt{\lambda_2}$, obtemos a coordenada no eixo 2: $0,517/\sqrt{0,2287} = 1,081079$ (ver tabela 4.3).

5 Conclusão

A partir desse estudo, concluímos que a abordagem adotada, que consiste na análise geométrica dos dados através de ACM se deu de forma satisfatória, fazendo com que a metodologia seja melhor compreendida e de forma intuitiva, o que geralmente não ocorre quando abordamos de forma algébrica através de decomposições de matrizes.

Em relação à análise gráfica, a ACM conseguiu discriminar na dimensão 1, os mais jovens e com classificação aceitável da córnea com os mais velhos e com classificação inaceitável da córnea. Isso foi mais uma confirmação de vários estudos relacionados nessa área como pode ser visto em Sano et al (2010), por exemplo; e na dimensão 2, os que faleceram em Juiz de Fora e com os menores tempos da enucleação à preservação com os que faleceram em outras localidades na Zona da Mata e com os maiores tempos da enucleação à preservação. Como o tempo de enucleação à preservação corresponde ao tempo em horas do local onde ocorreu retirada da córnea até o local onde é feita a preservação da mesma (Juiz de Fora), ou seja, corresponde ao tempo de viagem da córnea, o que torna fácil a compreensão da associação realizada pela dimensão 2. Também na análise gráfica, a respeito da variável suplementar, mesmo que a ACM trate as categorias como qualitativas nominais, foi preservada a ordenação das categorias, a qual foi feita através dos quartis.

Nota-se que especificamente a ACM, não é ainda muito difundida, daí a importância de estudos que tratem deste assunto, seja pelo enfoque geométrico ou algébrico.

I Anexo

Figura I.1: Parecer do Comitê de Ética



PARECER Nº 004/2010

Registro CEP/FHEMIG: 004/2010, (este nº deve ser citado nas correspondências referente a este projeto).
 Data: 29/01/2010
 CAAE - 0002.0.287.180-10

Hospital Regional Dr. João Penido – HRJP/FHEMIG.

Pesquisador Responsável: Rafael Vidal Mérula

DECISÃO:

O Comitê de Ética em Pesquisa, (CEP/FHEMIG), aprovou em 28 de Janeiro de 2010, o projeto de pesquisa intitulado: "Efeito da Idade do Doador na Utilização do Tecido Corneano".

Segundo o item VII.13.d da Resolução 196/96, os pesquisadores ficam **OBRIGADOS** a enviar relatórios anualmente.

Assim o relatório parcial deverá ser entregue em **28/01/2011** e/ou o relatório final ao término da pesquisa. O relatório poderá seguir o modelo de acompanhamento disponível na Intranet/FHEMIG.

- Formulário CAAE – Certificado de Apresentação para Apreciação Ética – deverá ser assinado, pelo pesquisador responsável, nos espaços apropriados, e em seguida o NEP deverá encaminhar a este CEP/FHEMIG 02 (duas) vias devidamente assinadas, enquanto a outra deverá ser entregue ao pesquisador responsável.


 Vanderson Assis Romualdo
 Coordenador do Comitê de Ética em Pesquisa

Figura I.2: Carta Convite

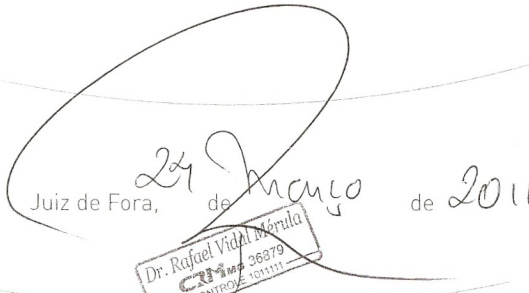
NÚCLEO VISÃO

Dr. Rafael Vidal Mérula - CRM 36.879
Residência em Oftalmologia - Hospital das Clínicas - UFMG
Doutor em Oftalmologia - Faculdade de Medicina - UFMG
Ex-fellow do Hospital de Olhos - Augesnpital - Universidade de Basiléia - Suíça

Venho por meio deste documento convidar o Prof. Ronaldo Rocha Bastos a participar do projeto de pesquisa "Efeito da Idade do Doador na Utilização do Tecido Corneano" coordenado pelo Dr. Rafael Vidal Mérula, no que diz respeito a análise estatística dos dados e interpretação dos resultados.

O referente projeto está registrado pelo CEP/FHEMIG sob o número 004/2010.

Juiz de Fora, de 24 de Junho de 2011


Dr. Rafael Vidal Mérula - CRM 36.879

Rua Marechal Deodoro 566 sl 1403/1404 Centro Juiz de Fora MG cep 36013-001 tel/fax 32 3212 7771 nucleovisao@yahoo.com.br

Referências Bibliográficas

- [1] Associação Brasileira de Transplante de Órgãos. *Estatísticas de Transplantes*. São Paulo: ABTO. Disponível em: <<http://www.abto.org.br/>>. Acesso em: mar. de 2011.
- [2] BENZÉCRI, J. P. *Correspondence Analysis Handbook*. New York: Marcel Dekker, 1992.
- [3] Brasil. Agência Nacional de Vigilância Sanitária. Resolução nº 67 de 30-09-2008. Dispõe sobre o Regulamento Técnico para o Funcionamento de Banco de Tecidos Oculares de origem humana.
- [4] FOX, J. *Series editor's introduction*, p. VIII - IX, em LE ROUX, B; ROUANET, H. *Multiple Correspondence Analysis*. edição ilustrada. v. 163. Londres: SAGE, 2010.
- [5] GREENACRE, M. J. *Theory and Applications of Correspondence Analysis*. London: Academic Press, 1984.
- [6] GREENACRE, M. J.; BLASIUS, J. *Multiple Analysis and Related Methods*. Boca Raton: Chapman & Hall/CRC, 2006.
- [7] GREENACRE, M. J. *Correspondence analysis in practice*. 2 ed. Boca Raton: Chapman & Hall/CRC, 2007.
- [8] GREENACRE, M. J. Correspondence analysis *Computational Statistics - Focus Article*, v. 2, p.613-619, 2010.
- [9] HUSSON, F.; LÊ, S.; PAGÉS, J. *Exploratory Multivariate Analysis by Example Using R*. edição ilustrada. Londres: CRC Press, 2010.
- [10] HUSSON, F.; JOSSE, J.; LE, S.; MAZET, J. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R*. R package version 1.12, 2010. Disponível em: <<http://www.R-project.org>>.
- [11] LE ROUX, B; ROUANET, H. *Geometric Data Analysis - From Correspondence Analysis to Structured Data Analysis*. Dordrecht: Kluwer Academic Publishers, 2004.

- [12] LE ROUX, B; ROUANET, H. *Multiple Correspondence Analysis*. edição ilustrada. v. 163. Londres: SAGE, 2010.
- [13] SOUZA, A. C. *Análise de Correspondência aplicada à ECINF: a diversidade do setor informal urbano no Brasil*. Dissertação de Mestrado Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2004.
- [14] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, 2009. Disponível em <<http://www.R-project.org>>.
- [15] SANO, R. Y.; SANO, F. T.; DANTAS, M. C. N.; LUI, A. C. F.; SANO, M. E.; NETO, A. L. Análise das córneas do Banco de Olhos da Santa Casa de São Paulo utilizadas em transplantes. *Arq Bras Oftalmol.*, v.73, n.8, p. 254-258 ,2010
- [16] SZAFLIK, J.; GRABSKA-LIBEREK, I.; BRIX-WARZECHA, M. The importance of Various Factors Relating to the Morphological Quality of Corneas Used for PKP by the Warsaw Eye Bank from 1996 to 2002. *Annals of transplantation*, v. 8, n. 2, p. 28-33, 2003.