

ESTUDO DE TABELAS DE CONTINGÊNCIA ATRAVÉS DA ANÁLISE DE
CORRESPONDÊNCIA

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sra.

BLANCA ROSA MAQUERA SOSA
e aprovada pela comissão Julgadora.

Campinas, 13 de Julho de 1989



Prof. Dr. Pushpa N. Rathie
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística

So71e

11380/BC

UNICAMP
BIBLIOTECA CENTRAL

**ESTUDO DE TABELAS DE CONTINGÊNCIA ATRAVÉS DA ANÁLISE DE
CORRESPONDÊNCIA**

BLANCA ROSA MAQUERA SOSA

Orientador

Prof. Dr. PUSHPA NARAYAN RATHIE

Dissertação apresentada ao Instituto
de Matemática, Estatística e Ciência
da Computação da Universidade Esta-
dual de Campinas como requisito
parcial para obtenção do título de
Mestre em Estatística.

- 1989 -

Glossário de Abreviações

A -	Arbitrador.	PD -	Preserve-Defer.
AU -	Allocation Unit.	PE -	Processing Element.
BY -	Bypass.	PG -	Preserve-Generate.
CEM -	Condition/Event Modelling Net.	PHT -	Parallel Hash Table.
CU -	Clear Unit.	PEIB -	Processing Element input buffer.
C/E -	Condição/Evento.	PEOB -	Processing Element output buffer.
D -	Distribuidor.	PPIB -	Preprocessor to Processor input buffer.
DD -	Decrement-Defer.	PPOB -	Preprocessor to Processor output buffer.
DU -	Defer Unit.	PSOB -	Processing Element to Switch output buffer.
EA -	Extract-Abort.	PU -	Processing Unit.
ED -	Extract-Defer.	RSIB -	Ring to Switch input buffer.
EW -	Extract-Wait.	RWC -	Alternating Read-Write Controller.
FU -	Function Unit.	SHIB -	Switch to Host input buffer.
GCS -	Global Control Store.	SHOB -	Switch to Host output buffer.
HKG -	Hash Key Generator.	SIE -	Special Instruction Executor.
HMIB -	Hash Unit to Matching Store input buffer.	SRC -	Store Read Controller.
HMOB -	Hash Unit to Matching Store output buffer.	SROB -	Switch to Ring output buffer.
HSIB -	Host to Switch input buffer.	SSU -	Structure Store Unit.
HSOB -	Host to Switch output buffer.	ST -	Segment Table.
ID -	Increment-Defer.	STIB -	Switch to Token Queue smoothing buffer.
L/T -	Lugar/Transição.	STC -	Segment Table Controller.
mf -	Matching Function.	SW -	Switch Unit.
MFDM -	Máquina de Fluxo de Dados de Manchester.	SWC -	Store Write Controller.
MNIB -	Matching Store to Node Store input buffer.	TMIB -	Token Queue to Matching Unit Input buffer.
MNOB -	Matching Store to Node Store output buffer.	TMOB -	Token Queue to Matching Unit output buffer.
MSC -	Matching Store Controller.	TQSB -	Token Queue smoothing buffer.
MSU -	Matching Store Unit.	TQU -	Token Queue Unit.
MVOB -	Matching Store to Overflow output buffer.	TS -	Token Store.
NIE -	Normal Instruction Executor.	UD -	Universo de Discurso.
NDS -	Node Description Store.	VMIB -	Overflow to Matching Store input buffer.
NPIB -	Node Store to Processing Unit input buffer.	VU -	Overflow Unit.
NPOB -	Node Store to Processing Unit output buffer.		
NSC -	Node Store Controller.		
NSU -	Node Store Unit.		

A meus Pais
Arnaldo e Florencia

AGRADEÇO

A Deus pela Proteção e Cuidado.

Ao Prof. Dr. P. N. Rathie, pela orientação segura durante o transcorrer deste trabalho. Particularmente, pela atenção, disponibilidade e estímulos recibidos.

À Universidade Estadual de Campinas, por proporcionar me condições para a conclusão deste trabalho.

À família Pereira Novo, por fazer de seu lar meu segundo lar.

À Elvira pelas sugestões que hoje fazem parte deste trabalho e acima de tudo sua amizade.

Ao C.N.Pq. e C.A.P.E.S. pela ajuda financeira.

Aos meus amigos e companheiros, que direta ou indiretamente, contribuíram para a realização deste trabalho.

INDICE GERAL

INTRODUÇÃO

CAP. I.	INTRODUÇÃO À ANÁLISE DE CORRESPONDÊNCIA.....	Pág. 1
1.1	Espaço Euclidiano Ponderado	
1.2	Matriz de Correspondência, Perfis e Centróide.	
1.3	Distância χ^2 , Inércia e Equivalência Distribucional.	
CAP. II.	ANÁLISE DE CORRESPONDÊNCIA DE UMA TABELA DE CONTINGÊNCIA PELO USO DA D. V. S. G.....	10
2.1	Geração de Subespaços Ótimos.	
2.2	Correspondência entre linhas e colunas.	
2.2.1	Coordenadas dos perfis das linhas.	
2.2.2	Coordenadas dos perfis das colunas.	
2.2.3	Transição das coordenadas.	
	A) Dos perfis das colunas para as filas.	
	B) Dos perfis das filas para as colunas.	
2.3	Decomposição da Inércia e Interpretação dos perfis nos subespaços.	
2.3.1	Contribuição Absoluta.	
2.3.2	Contribuição Relativa.	
2.4	Perfis Suplementares.	
CAP. III.	ESTUDO DA PRODUÇÃO DE GRÃOS NO BRASIL.....	29
3.1	Introdução.	
3.2	Tabelas, Gráficos e Estudos.	
	A) Estudo das Safras.	
	B) Estudo da produção de Grãos.	
	C) Estudo da Correspondência entre a Produção de Grãos e Safras.	
3.3	Conclusões.	

CAP. IV.	UM ESTUDO SOBRE OS VEÍCULOS ENVOLVIDOS EM ACI - DENTES DE TRÂNSITO.....	37
4.1	Introdução.	
4.2	Descrição das Variáveis.	
4.2.1	Lista das variáveis.	
4.2.2	Apresentação das tabelas de dados.	
4.3	Resultados da Análise de Correspondência.	
A)	Estudo da Correspondência entre U.F.	
B)	Estudo da Correpondência entre espécie de veículo.	
C)	Estudo da Correspondência entre espécies de veículo e U.F.	
D)	Estudo da Correspondência entre espécies de veículo e U.F. considerando a Área e o Período como perfis suplementares.	
4.4	Conclusões.	
CAP. V.	ESTUDO DO MÊS DE MORTE, MÊS DE ANIVERSÁRIO E PROFISSÃO.	49
5.1	Introdução.	
5.2	Apresentação dos dados.	
5.3	Tabelas Gráficos e Estudo dos Resultados.	
A)	Estudo do mês de morte relativo ao mês de aniversário segundo a profissão.	
B)	Estudo da relação entre as profissões segundo o mês de morte com respeito ao mês de aniversário.	
C)	Estudo da correspondência entre o mês de morte com respeito ao mês de aniversario e profissão.	
5.4	Conclusões.	
	CONCLUSÕES.....	59
	REFERÊNCIAS BIBLIOGRÁFICAS.....	60
	APÊNDICE.....	63

INTRODUÇÃO

Muitas ciências, tais como ciências biológicas, sociais, agrícolas, resumem seus dados em tabelas cruzadas geralmente referidas como tabelas de contingência. As unidades de amostra em tais circunstâncias são classificadas conforme variáveis categóricas.

As formas conhecidas para análise destas tabelas são os modelos log-lineares que são testados mediante o teste χ^2 , cujo pioneiro é Yule(1903), e o teste χ^2 de independência de Pearson(1900) que é utilizado para testar a independência entre as variáveis categóricas.

Nestes últimos anos, desenvolveram-se técnicas gráficas multidimensionais que nos permitem descobrir a estrutura que encobre uma tabela de contingência; por ser a interpretação de um gráfico muito mais fácil de ser assimilado e interpretado que uma quantidade numérica, a potência desta forma de análise faz-se aplicável a diferentes campos de estudo como genético, psicológico, social, ecológico, criminológico, etc. Uma destas técnicas gráficas multidimensionais é a Análise de Correspondência (Greenacre 1984). Atribui-se a H.O. Hartley a origem matemática desta forma de análise, que publicou um artigo (1935), no qual dá uma formulação algebraica da "correlação" entre as linhas e colunas de uma tabela de contingência, embora Richardson e Kuder(1933) e Horst(1935) independentemente sugerem idéias similares não matemáticas na literatura psicométrica. O termo "análise de correspondência" e a forma geométrica têm origem na França. O primeiro expositor deste método com este título foi J.P. Benzécri (1963), que também guiou um grupo de analistas de dados desde a década do 1960, o qual ganhou experiência prática intensiva em Análise de Correspondência e outras técnicas

multivariadas; os resultados destes estudos podem ser vistos nas revistas "Les cashiers de L'Analyse des données.", onde se observa o processo evolutivo desta análise de dados e aplicação a diferentes campos, inclusive música. Esta técnica é baseada na decomposição de valores singulares generalizada (D.V.S.G.), aproximação de uma matriz de posto menor e distância χ^2 .

Neste trabalho, fazemos um estudo de três problemas através da Análise de Correspondência.

O trabalho conta com cinco capítulos. No capítulo I explicamos os aspectos mais importantes para obter a correspondência entre linhas e colunas de uma Tabela de Contingência (T.C.).

No capítulo II desenvolvemos a Análise de Correspondência pelo uso da D.V.S.G., e suas características.

No capítulo III estudamos o processo evolutivo da Produção de Grãos no Brasil nas safras 1979/80 a 1986/87.

No capítulo IV fazemos um estudo sobre os veículos envolvidos em acidentes de trânsito segundo as Unidades de Federação (U.F) do Brasil durante o ano de 1985, além disso fazemos um estudo global considerando os acidentes de trânsito na Área e Período como perfis suplementares.

No capítulo V, estudamos o mês de morte com respeito ao mês de aniversário de pessoas ilustres segundo a sua profissão.

No apêndice incluímos o programa computacional, feito em linguagem Fortran.

Finalmente, damos as conclusões gerais do uso desta técnica gráfica multidimensional aos problemas estudados.

CAPITULO I

INTRODUÇÃO À ANÁLISE DE CORRESPONDÊNCIA

Neste capítulo introduziremos conceitos nos quais baseia-se uma das técnicas gráficas multidimensionais para a análise de uma tabela de contingência, o qual desenvolve-se sobre um espaço euclidiano ponderado; assim iniciaremos este capítulo definindo norma, distância, ângulo, momento de inércia, de vetores dentro de um espaço euclidiano ponderado; na seção 1.2 definimos a matriz de correspondência, os perfis das filas e colunas da matriz e seus respectivos centróides; na seção 1.3 desenvolvemos a distância χ^2 , o Princípio de Equivalência Distribucional e o momento de inércia ou inércia dos perfis da matriz de correspondência.

1.1. ESPAÇO EUCLIDIANO PONDERADO

Nesta seção introduzimos definições referente aos vetores sobre um espaço euclidiano ponderado.

DEF.1.1.1. Se v_1 e $v_2 \in R^J$, então o Produto Escalar Euclidiano ponderado é definido por :

$$\langle v_1, v_2 \rangle = v_1^T D_q v_2 = \sum_{j=1}^J q_j v_{1j} v_{2j}$$

onde:

- . D_q é a matriz diagonal de ordem $J \times J$ cujos elementos $q_1, \dots, q_J$, são números reais positivos que definem a ponderação relativa assinalada às coordenadas dos vetores.
 - . R^J representa o espaço euclidiano ponderado.
- O símbolo " τ " representa a transposta do vector v_1 .

DEF 1.1.2. A Norma Euclidiana Ponderada de um vetor $v_1 \in R^J$ é definida por:

$$||v_1||^2 = v_1^T D_q v_1 = \sum_{j=1}^J q_j v_{1j}^2$$

DEF 1.1.3. Se $v_1, v_2 \in R^J$ a Distância Euclidiana Ponderada entre v_1 e v_2 é definida como a soma ponderada dos quadrados das diferenças de seus coordenadas, isto é:

$$\begin{aligned} d^2(v_1, v_2) &= ||v_1 - v_2||_{D_q}^2 = (v_1 - v_2)^T D_q (v_1 - v_2) \\ &= \sum_{j=1}^J q_j (v_{1j} - v_{2j})^2 \end{aligned}$$

DEF 1.1.4. Se v_1 e $v_2 \in R^J$, o Ângulo α , entre v_1 e v_2 é definido como:

$$\cos \alpha = v_1^T D_q v_2 / ||v_1|| ||v_2||$$

DEF 1.1.5. Seja $v_1, \dots, v_J \in R^J$ e associado a cada um dos vetores uma massa positiva m_j . A Inércia Total destes pontos com relação a $y \in R^J$ é definida como:

$$In(y) = \sum_{j=1}^J m_j d^2(v_j, y)$$

onde:

$$d^2(v_j, y) = (v_j - y)^T D_q (v_j - y)$$

$In(y)$ é uma medida de dispersão dos pontos em volta de y .

Se $D_q = I$ (Matriz de Identidade), as definições (1.1.1) a (1.1.5) coincidem com as definições usuais do espaço euclidiano.

1.2. MATRIZ DE CORRESPONDÊNCIA, PERFIS E CENTRÓIDE

Antes de apresentarmos a matriz de correspondência, enunciaremos alguns pontos sobre tabelas de contingência que serão úteis.

DEF 1.2.1. Uma *Tabela de Contingência* (T.C) é essencialmente uma amostra de uma população multivariada com as probabilidades e partições de categorias, sujeitas às restrições da distribuição multinomial.

TABELA DE CONTINGÊNCIA DE 2 FATORES

Suponhamos que temos uma amostra de observações independentes e que a cada uma dessas observações associamos uma categoria em cada um dos 2 fatores: linha (I), coluna(J).

Denominaremos:

I : número de categorias do fator linha.

J : número de categorias do fator coluna.

I x J : número de celas da tabela de contingência.

A tabela de contingência pode então ser representada por uma matriz N da seguinte maneira.

$$N = \begin{bmatrix} n_{11} & n_{12} & . & . & n_{1j} & . & n_{1J} \\ . & . & . & . & . & . & . \\ n_{i1} & n_{i2} & . & . & n_{ij} & . & n_{iJ} \\ . & . & . & . & . & . & . \\ n_{I1} & n_{I2} & . & . & n_{Ij} & . & n_{IJ} \end{bmatrix}$$

onde:

n_{ij} = frequência de ocorrência na i-ésima linha e na j-ésima coluna.

Denotaremos:

$n_{i.} = \sum_{j=1}^J n_{ij}$ como a soma dos elementos da i -ésima linha.

$n_{.j} = \sum_{i=1}^I n_{ij}$ como a soma dos elementos da j -ésima coluna.

$n_{..} = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ como o número total de elementos de N .

Uma Matriz de Correspondência P , é definida como a matriz de elementos de N , divididos pelo número total de elementos de N , isto é;

$$P = N/n_{..} = \begin{bmatrix} P_{11} & P_{12} & . & . & . & . & . & P_{1J} \\ P_{21} & P_{22} & . & . & . & . & . & P_{2J} \\ . & . & . & . & . & . & . & . \\ P_{I1} & P_{I2} & . & . & . & . & . & P_{IJ} \end{bmatrix} \quad (1.2.1)$$

Para se encontrar as relações entre linhas e colunas de uma T.C., utiliza-se vetores de frequências relativas, os quais são chamados perfis, além disso associase uma "massa" a cada um destes perfis.

Os perfis das linhas e colunas de uma matriz N são definidos como os vetores das linhas e colunas de N , divididos pela soma de suas coordenadas respectivamente.

Consideremos as linhas de N e P , denotemos por $\underline{r}_i$ o i -ésimo perfil de linha, que define-se como:

$$\underline{r}_i = \begin{bmatrix} r_{i1} \\ . \\ . \\ r_{iJ} \end{bmatrix} \in R^J \text{ onde, } r_{ij} = n_{ij}/n_{i.}$$

A soma dos elementos das linhas de P, são elementos de um novo vetor que é chamado *Vetor de massa dos perfis das linhas* e é definida como:

$$r = P 1_j = \begin{bmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_I \end{bmatrix} \quad (1.2.2)$$

onde:

$r_i = n_{i.}/n_{..}$ é a massa do i-ésimo perfil de linha de P.
 1_j é o vetor de J elementos composta por uns.

A matriz que contém os perfis das linhas denotamos por R o qual é definida da seguinte maneira:

$$R = D_r^{-1} P \quad (1.2.3)$$

onde:

D_r é matriz diagonal positiva cujos elementos são as massas dos perfis das linhas de P.

O símbolo D_r^{-1} representa a inversa da matriz D_r .

Agora, consideremos as colunas de N e P, denotemos por $\underline{c}_j$ o j-ésimo perfil coluna, definido por:

$$\underline{c}_j = \begin{bmatrix} c_{1j} \\ \vdots \\ c_{ij} \\ \vdots \\ c_{Ij} \end{bmatrix} \in R^I, \text{ onde } c_{ij} = n_{ij}/n_{.j}$$

As somas das colunas de P são elementos do vetor de massa dos perfis das colunas de P e é definida como:

$$c = P^T 1_I = \begin{bmatrix} c_1 \\ \vdots \\ c_J \end{bmatrix} \quad (1.2.3)$$

onde:

$c_j = n_{.j}/n_{..}$ é a massa do j -ésimo perfil de coluna de P .
 1_I é o vetor de I elementos composta por uns.

A matriz que contém os perfis das colunas de P se denota por C , a qual é definida como:

$$C = D_c^{-1} P^T \quad (1.2.5)$$

onde:

D_c é matriz diagonal positiva cujos elementos são as massas dos perfis das colunas.

É interessante notar que os perfis das linhas e perfis das colunas definem duas nuvens de pontos no espaço euclidiano ponderado J e I - dimensional respectivamente.

Com estas definições podemos agora definir o centróide dos perfis de P .

Seja $r_1, \dots, r_I$, os perfis das linhas de P , o centróide destes perfis é definido por:

$$c = \sum_{i=1}^I r_i r_i \quad ; \quad \sum_{i=1}^I r_i = 1 \quad (1.2.6)$$

Similarmente, o centróide dos perfis das colunas, $\underline{c}_1, \dots, \underline{c}_c$ é definido por:

$$r = \sum_{j=1}^v c_j \underline{c}_j \quad ; \quad \sum_{j=1}^v c_j = 1 \quad (1.2.7)$$

Estes dois centróides estão relacionados, devido ao fato de que o centróide dos perfis das colunas é igual ao vetor de massa dos perfis das linhas, e o centróide dos perfis das linhas é igual ao vetor de massa dos perfis das colunas de P. (ver Greenacre, 1984:, p.85).

As combinações lineares (1.2.6) e (1.2.7) muitas vezes é chamado de *Baricentro*.

A distância entre os perfis e o centróide é medida, por uma distância euclidiana ponderada chamada distância χ^2 .

1.3. DISTANCIA χ^2 , INÉRCIA E EQUIVALENCIA DISTRIBUCIONAL.

Consideremos dois perfis de linhas $\underline{r}_i$ e $\underline{r}_{i'}$. Pela def. 1.1.3, a distância entre $\underline{r}_i$ e $\underline{r}_{i'}$ no espaço euclidiano ponderado J-dimensional, ponderado na matriz D_c^{-1} , é definida por:

$$\begin{aligned} d^2(\underline{r}_i, \underline{r}_{i'}) &= (\underline{r}_i - \underline{r}_{i'})^T D_c^{-1} (\underline{r}_i - \underline{r}_{i'}) \\ &= \sum_{j=1}^v (r_{ij} - r_{i'j})^2 / c_j \end{aligned} \quad (1.3.1)$$

Escrita esta expressão em função dos dados originais, temos:

$$d^2(\underline{r}_i, \underline{r}_{i'}) = \sum_{j=1}^v n_{..} / n_{.j} (n_{ij} / n_{i.} - n_{i'j} / n_{i'.})^2$$

Pela analogia com a estatística χ^2 , chama-se distância χ^2 (ver J.P. Benzécri, 1973).

Similarmente, se consideramos dois perfis das colunas $\underline{c}_j$ e $\underline{c}_{j'}$, a distância χ^2 entre eles no espaço euclidiano ponderado na matriz D_r^{-1} é definida por:

$$\begin{aligned}
d^2(\underline{c}_j, \underline{c}_{j.}) &= \sum_{i=1}^I n_{i.}/n_{i.} (n_{ij}/n_{.j} - n_{ij}/n_{.j})^2 \\
&= \sum_{i=1}^I (c_{ij} - c_{ij.})^2 / r_i
\end{aligned} \tag{1.3.2}$$

Substituindo em (1.3.1) c_j por r_{ij} temos:

$$\begin{aligned}
d^2(\underline{r}_i, c) &= \sum_{j=1}^J (r_{ij} - c_j)^2 / c_j \\
&= (\underline{r}_i - c)^T D_c^{-1} (\underline{r}_i - c)
\end{aligned} \tag{1.3.3}$$

que é a distância entre o i-ésimo perfil de linha e seu centróide.

Da mesma maneira se em (1.3.2) substituirmos r_i em lugar de c_{ij} , obtemos a distância χ^2 entre o j-ésimo perfil de coluna e seu centróide.

$$\begin{aligned}
d^2(\underline{c}_j, r) &= \sum_{i=1}^I (c_{ij} - r_i)^2 / r_i \\
&= (\underline{c}_j - r)^T D_r^{-1} (\underline{c}_j - r)
\end{aligned} \tag{1.3.4}$$

Utilizam-se estas distâncias χ^2 , para medir o momento de inércia ou inércia dos perfis em relação a seu centróide.

Considerando a def.1.1.5, o momento de inércia dos perfis das linhas In(I) é definido por:

$$\begin{aligned}
\text{In(I)} &= \sum_{i=1}^I r_i (\underline{r}_i - c)^T D_c^{-1} (\underline{r}_i - c) \\
&= \text{Traço}(D_r (R - 1 \ C^T) D_c^{-1} (R - 1 \ C^T))
\end{aligned} \tag{1.3.5}$$

Da mesma maneira define-se o momento de inércia dos perfis das colunas In(J) por:

$$\text{In(J)} = \sum_{j=1}^J c_j (\underline{c}_j - r)^T D_r^{-1} (\underline{c}_j - r)$$

$$= \text{Traço}(D_c(C - 1 r^T) D_r^{-1}(C - 1 r^T)) \quad (1.3.6)$$

As inércias de ambos perfis são iguais, além disso são iguais à estatística χ^2 com $(r-1)(c-1)$ g.l. que testa a independência das linhas e colunas de uma T.C. , dividido pelo número total de elementos da amostra,

$$\text{In}(I) = \text{In}(J) = \chi^2/n.. \quad (1.3.9)$$

(ver Greenacre, 1984, seção 4.1.6).

Agora, talvez nos perguntemos por que se utiliza a distância χ^2 , para medir a distância entre os perfis. Uma das razões é que pela utilização desta distância χ^2 , as inércias totais dos dois conjuntos de perfis com respeito a seus centróides são iguais.

Outra razão para utilizar a distância χ^2 é por causa do pelo chamado "Princípio de Equivalência Distribucional" (J.P.Benzécri et al., 1973), o qual é peculiar nesta forma de análise e em particular no gráfico das coordenadas principais. (seção 2.2). O princípio diz: " Se dois perfis de linhas (colunas) são idênticos, então as duas linhas (colunas) correspondentes da matriz original de dados podem ser substituídas pela soma delas sem afectar a geometria dos perfis das colunas (linhas).".

Escofier (1978) mostra que a distância χ^2 verifica bem esta propriedade.

No capítulo seguinte , desenvolveremos a análise de correspondência de uma T.C. pelo uso da decomposição de valores singulares generalizada (D.V.S.G).

CAPITULO II

ANÁLISE DE CORRESPONDÊNCIA DE UMA TABELA DE CONTINGÊNCIA PELO USO DA D.V.S.G.

O Método gráfico da análise de uma tabela de contingência através da D.V.S.G. é mais conhecida com o nome de análise de correspondência (ver Greenacre (1984), Jolliffe (1986)). Este método nos permite observar as relações entre as variáveis categóricas de uma T.C. sobre subespaços unidimensionais ou bidimensionais, tais que neles esteja maximizada a inércia.

Na seção 2.1 deste capítulo desenvolvemos a D.V.S.G. e a aplicação desta para gerar subespaços; na seção 2.2 a correspondência entre linhas e colunas de uma T.C.; na seção 2.3 Decomposição da Inércia e Interpretação da projeção dos perfis nos subespaços e na seção 2.4 os perfis suplementares.

2.1. GERAÇÃO DE SUBESPAÇOS ÓTIMOS

Definimos o termo subespaço em um sentido diferente ao da matemática usual.

DEF. 2.1.1. O conjunto de vetores $u + \alpha_1 v_1 + \dots + \alpha_k v_k$ é chamado de Subespaço K-dimensional de um espaço J-dimensional, onde u é qualquer vetor J-dimensional fixo, e $v_1, \dots, v_k$, são vetores J-dimensionais linearmente independentes e $\alpha_1, \alpha_2, \dots, \alpha_k$, são números reais.

Nesta definição o vetor fixo u pode ser considerado como o primeiro passo de transferência dos vetores da origem do subespaço ou como uma redefinição, de uma nova origem do espaço, no ponto u .

Em 1901, Karl Pearson publicou um trabalho sobre o ajuste de um sistema de pontos em um multiespaço a uma linha ou um plano utilizando para tal a geometria euclidiana. Este enfoque foi retomado em 1933 por Hotelling, que foi o primeiro a formular a análise por componentes principais que perdura até nossos dias.

O trabalho original de Pearson (1901) se centrava naqueles componentes e combinações lineares de variáveis originais, para as quais a variância não explicada fosse mínima. Estas combinações geram um plano, função das variáveis originais na qual o ajuste do sistema de pontos é "o melhor", por ser mínima a soma das distâncias de cada ponto ao plano de ajuste.

A seguir com este mesmo critério mas utilizando a distância χ^2 , (def.1.1.3 (Greenacre, (1984))), desenvolve-se o ajuste dos pontos no espaço J-dimensional a um subespaço K-dimensional.

Seja S o subespaço de dimensão K, denotemos por y_i aos pontos que pertencem ao espaço euclidiano J-dimensional ponderado por D_q ; $\hat{y}_i$ representa um ponto do subespaço "mais próximo" a y_i , a distância mínima entre estes pontos será denotada por d_i .

Se y_i é ponderada pela massa w_i ($i=1, \dots, I$), então nossa definição de proximidade de todo o conjunto de pontos para o subespaço S é,

$$\psi(S; y_1, y_2, \dots, y_I) = \sum_{i=1}^I w_i d_i^2 \quad (2.1.1)$$

onde:

$$d_i^2 = \|y_i - \hat{y}_i\|_{D_q}^2 = (y_i - \hat{y}_i)^T D_q (y_i - \hat{y}_i)$$

e D_q é matriz diagonal de ordem $J \times J$.

Claramente ψ depende do subespaço S, então temos que encontrar o subespaço S^* tal que minimize a função ψ em (2.1.1)

Pela Def. 2.1.1 pode-se considerar um ponto s , como um subespaço de dimensão zero, logo a função (2.1.1) torna-se

$$\psi(s; y_1, \dots, y_I) = \sum_{i=1}^I w_i (y_i - s)^T D_q (y_i - s) \quad (2.1.2)$$

desde que $\hat{y}_i$ é igual a $s \forall i$.

Derivando (2.1.2) com respeito a s

$$\partial \psi / \partial s = \sum_{i=1}^I w_i 2 D_q (y_i - s) = 0$$

$$s = \sum_{i=1}^I w_i y_i = \bar{y}$$

O centroide $\bar{y}$ é o ponto que minimiza esta função, assim neste sentido, o centroide $\bar{y}$ é o ponto mais próximo a todos os pontos dados $y_1, \dots, y_I$.

Greenacre (1984 p. 44), mostra que qualquer subespaço S ótimo no sentido de minimizar (2.1.1) necessariamente contém o centróide. Com este resultado restringiremos as aproximações $\hat{y}_i$, dos pontos y_i à forma:

$$\hat{y}_i = \bar{y} + \sum_{k=1}^K f_{ik} v_k \quad i = 1, 2, \dots, I$$

onde $v_1, v_2, \dots, v_K$ são vetores bases do subespaço. Logo a função (2.1.1) pode ser escrita como:

$$\psi(S; y_1, \dots, y_I) = \sum_{i=1}^I w_i (y_i - \bar{y} - \sum_k f_{ik} v_k)^T D_q (y_i - \bar{y} - \sum_k f_{ik} v_k) \quad (2.1.3)$$

As variáveis desta função objetivo são os K eixos, $v_1, \dots, v_K$ implicando um total de I vezes K variáveis escalares.

Pela escolha particular da distância χ^2 , a solução da função ψ é consideravelmente simplificada.

A solução completa teórica para a minimização de (2.1.3), que especifica a dimensão de K, está envolta no conceito de D.V.S.G e na aproximação de uma matriz de posto menor que serão dados nos seguintes teoremas respectivamente.

TEOREMA 2.1.1. Sejam duas matrizes simétricas positivas $\theta: I \times I$ e $\phi: J \times J$, então qualquer matriz $A: I \times J$ com posto $K \leq \min(I, J)$ pode ser expressa como:

$$A = N D_{\mu} M^T = \sum_{k=1}^K \mu_k n_k m_k^T \quad (2.1.4)$$

onde $D_{\mu}: K \times K$ é matriz diagonal de elementos positivos μ_k , $N: I \times K$ e $M: J \times K$ são matrizes ortonormais relativamente a θ e ϕ respectivamente, i.é

$$N^T \theta N = M^T \phi M = I \quad (2.1.5)$$

e n_k, m_k são as colunas das matrizes N e M respectivamente.

DEMONSTRAÇÃO:

Para demonstrar este teorema, utiliza-se a Decomposição de Valores Singulares (D.V.S) e propriedades dela quando aplicada às matrizes simétricas (ver Mardía (1979), pág 466-474).

A D.V.S de θ é:

$$\theta = W D_{\alpha} W^T, \text{ então } \theta^{1/2} = W D_{\alpha}^{1/2} W^T.$$

Aplicando a D.V.S á matriz $\theta^{+1/2} A \phi^{+1/2}$, resulta:

$$\theta^{1/2} A \phi^{1/2} = U D_{\mu} V^T \quad (2.1.6)$$

onde:

$$U^T U = V^T V = I \quad (2.1.7)$$

pré-multiplicando por $\theta^{-1/2}$ e pós-multiplicando por $\phi^{-1/2}$ a ambos os membros da igualdade (2.1.6) resulta que :

$$A = \theta^{-1/2} U D_{\mu} V^T \phi^{-1/2}$$

fazendo:

$$N = \theta^{-1/2} U \quad \text{e} \quad M = \phi^{-1/2} V$$

e considerando (2.1.7) obtemos:

$$A = N D_{\mu} M^T$$

e

$$N^T \theta N = M^T \phi M = I. \blacksquare$$

A decomposição (2.1.4) é chamada de decomposição de valores singulares generalizada, os números $\mu_1, \mu_2, \dots, \mu_k$ são os valores singulares generalizados da matriz A e as matrizes N e M são formadas pelos vetores singulares generalizados da matriz A . Eles são bases ortonormais para as colunas e linhas de A no espaço euclidiano I, J - dimensional ponderado nas matrizes θ e ϕ respectivamente. Além disso, as coordenadas das linhas e colunas de A com respeito aos vetores base M e N são $N D_{\mu}$ e $M D_{\mu}$ respectivamente.

Esta decomposição em valores singulares generalizada tem aplicação na aproximação de uma matriz de posto menor, o qual é tratado por Eckart e Young (1936).

Uma matriz $A_k^*: I \times J$ de posto k^* é a melhor aproximação de uma matriz $A: I \times J$ de posto k , $k \geq k^*$, se

$$\|A - A_k^*\|^2 = \inf_{\text{posto}(X)=k^*} \|A - X\|_{\theta\phi}^2$$

onde:

$$\begin{aligned} \|A - X\|_{\theta\phi}^2 &= \sum_{i=1}^I \sum_{j=1}^J \theta_i \phi_j (a_{ij} - x_{ij})^2 \\ &= \sum_{i=1}^I \theta_i (a_i - x_i)^T \phi (a_i - x_i) \end{aligned} \quad (2.1.8)$$

O teorema que enunciaremos a seguir é importante para o nosso problema de aproximação.

TEOREMA 2.1.2. Seja $A: I \times J$ de posto k , e seja

$$A = N D_{\mu} M^T = \sum_{k=1}^K \mu_k n_k m_k^T \quad (2.1.9)$$

sua decomposição em valores singulares generalizados. Então para $k^* \leq k$, a melhor aproximação de posto k^* da matriz A é dado por:

$$A = \sum_{k=1}^{k^*} \mu_k n_k m_k^T$$

A aproximação é única se $\mu_k > \mu_{k+1}$.

A melhor aproximação dada pelo teorema 2.1.2 é no sentido de mínimos quadrados (ver Greenacre 1984, pag. 342 - 344).

Pelo teorema 2.1.2, vemos que a D.V.S.G dá a solução para qualquer subespaço de dimensão k^* . Por exemplo: para $k^*=1$ utilizamos o primeiro vetor singular associado ao seu valor singular e assim sucessivamente podemos adicionar os demais vetores singulares generalizados (eixos principais) associados a seus respectivos valores singulares generalizados.

O quadrado dos valores singulares generalizados dão uma idéia de como uma matriz é representada sobre os eixos principais (vetores singulares generalizados).

A variação total de uma matriz é quantificada pela sua norma quadrado (ver equação (2.1.8))

$$\|A\|_{\theta\phi}^2 = \sum_{i=1}^I \theta_i a_i^T \phi a_i = \text{traço}(\theta A^T \phi A)$$

$$\begin{aligned} \|A\|_{\theta\phi}^2 &= \text{traço}(\theta M D_{\mu} N^T \phi N D_{\mu} M^T) = \text{traço}(\theta M D_{\mu} D_{\mu} M^T) \\ &= \text{traço}(D_{\mu}^2 M^T \theta M) = \sum_{k=1}^K \mu_k^2 \end{aligned}$$

Similarmente a variação da aproximação de A_k^* é :

$$\|A_k^*\|_{\phi}^2 = \sum_{k=1}^K \mu_k^2$$

e a variação inexplicada é:

$$\|A - A_k^*\|_{\phi}^2 = \sum_{k=K+1}^K \mu_k^2.$$

A medida tradicional da qualidade de aproximação de A por A_k^* é a percentagem da soma de quadrados:

$$\tau = 100 \frac{\sum_{k=1}^K \mu_k^2}{\sum_{k=1}^K \mu_k^2}$$

Baseados nestes conceitos desenvolvemos a correspondência entre linhas e colunas de uma T.C.

2.2. CORRESPONDÊNCIA ENTRE LINHAS E COLUNAS DE UMA TABELA DE CONTINGÊNCIA.

A correspondência é consequência da existência de uma relação recíproca entre bases (eixos principais) e coordenadas dos dois espaços euclidianos ponderados que contêm os perfis.

Seja $(P - rc^T): I \times J$, onde P é matriz de correspondência e rc^T é uma matriz cujos elementos estão compostos pelo producto do vetor massa, dos perfis das linhas de P , e o vetor transposto de massa, dos perfis das colunas.

Aplicando o teorema 2.1.1 sobre a matriz $(P - rc^T)$, temos:

$$(P - rc^T) = V D_{\mu} B^T = \sum_{k=1}^K \mu_k v_k b_k^T, \text{ para } K \leq \text{posto}(P - rc^T) \quad (2.2.1)$$

onde:

$$V^T D_r^{-1} V = B^T D_c^{-1} B = I$$

D_μ é matriz diagonal e I é a matriz identidade, cuja dimensão é igual ao posto de $P-rc^T$.

Pelo Teorema 2.1.1, as colunas de B denominam-se eixos principais (vetores singulares generalizados), para o conjunto de I "observações" dado pelas linhas de $P-rc^T$. Similarmente, as colunas de V denominam-se eixos principais para o conjunto de J "observações" dado pelas colunas de $P-rc^T$, supondo-se que o posto de $P-rc^T$ é K ; então, das K colunas de B e V pode-se derivar as coordenadas do perfil das linhas e colunas sobre o espaço K -dimensional respectivamente.

2.2.1. COORDENADAS PRINCIPAIS DOS PERFIS DAS LINHAS, COM RESPEITO AO CENTRÓIDE.

Seja $(R-1c^T):I \times J$ matriz dos perfis das linhas centradas no seu centróide, aplicando a D.V.S.G. a esta matriz temos:

$$(R-1c^T) = L D_\alpha Y^T \quad (2.2.2)$$

onde $L:I \times K$ e $Y:K \times J$ são matrizes ortonormais relativamente a D_r e D_c^{-1} respectivamente, i.é

$$L^T D_r L = Y^T D_c^{-1} Y = I$$

pré-multiplicando a igualdade (2.2.2) por D_r , temos:

$$\begin{aligned} (D_r R - D_r 1 c^T) &= D_r L D_\alpha Y^T \\ (P-rc^T) &= D_r L D_\alpha Y^T \end{aligned} \quad (2.2.3)$$

ortonormalizando as bases de (2.2.3) e considerando (2.2.1), temos:

$$L^T D_r L = Y^T D_c^{-1} Y = I \quad (2.2.4)$$

observe que as colunas de Y são idênticas a B , além disso

$$D_\alpha = D_\mu.$$

pré-multiplicando por D_r^{-1} , pós-multiplicando por D_c^{-1} a igualdade (2.2.3), e considerando (2.2.4) temos que:

$$L D_o = D_r^{-1} (P - r c^T) D_c^{-1} B$$

Esta matriz contém as coordenadas dos perfis com respeito ao centróide em todos os subespaços, a qual denotaremos por F , logo:

$$F = D_r^{-1} (P - r c^T) D_c^{-1} B \quad (2.2.5)$$

LEMA 2.2.1. As coordenadas dos perfis das linhas com respeito a seus eixos principais, estão relacionados com os eixos principais dos perfis das colunas i.e $F = D_r^{-1} V D_\mu$

DEMONSTRAÇÃO:

Para a demonstração consideremos a D.V.S.G. da matriz $(P - r c^T)$ como em (2.2.1), pós-multiplicando por $D_c^{-1} B$ a essa expressão obtemos:

$$(P - r c^T) D_c^{-1} B = V D_\mu B^T D_c^{-1} B, \quad (B^T D_c^{-1} B = I)$$

ou

$$(P - r c^T) D_c^{-1} B = V D_\mu \quad (2.2.6)$$

substituindo (2.2.6) em (2.2.5)

resulta que:

$$F = D_r^{-1} V D_\mu \quad (2.2.7)$$

2.2.2. COORDENADAS PRINCIPAIS DOS PERFIS DAS COLUNAS, COM RESPEITO AO CENTRÓIDE.

Seja $(C - 1r^T): J \times I$ a matriz dos perfis das colunas centradas no seu centróide, então aplicando a D.V.S.G. a esta

matriz obtemos:

$$(C - 1r^T) = W D_{\beta} T^T \quad (2.2.8)$$

onde:

$$W^T D_c W = T^T D_r^{-1} T = I$$

O procedimento para achar as coordenadas é similar a subseção 2.2.1., pré-multiplicando a igualdade (2.2.8) por D_c temos:

$$(D_c C - D_c 1_r^T) = D_c W D_{\beta} T^T$$

ou

$$(P^T - cr^T) = D_c W D_{\beta} T^T \quad (2.2.9)$$

ortonormalizamos as bases de (2.2.9), e considerando (2.2.1) temos que:

$$W^T D_c W = T^T D_r^{-1} T = I \quad (2.2.10)$$

Observe que as colunas de T são idênticas a V e $D_{\beta} = D_{\mu}$, pré-multiplicando por D_c^{-1} , e pós-multiplicando por $D_r^{-1} T$ a igualdade (2.2.9), além disso considerando (2.2.10), temos:

$$W D_{\beta} = D_c^{-1} (P^T - cr^T) D_r^{-1}$$

A matriz que contém as coordenadas dos perfis das linhas denotaremos por G , logo:

$$G = D_c^{-1} (P^T - cr^T) D_r^{-1} V \quad (2.2.11)$$

LEMA 2.2.2. As coordenadas dos perfis das colunas com respeito a seus eixos principais estão relacionados com os eixos principais dos perfis das linhas

$$\text{i.e. } G = D_c^{-1} B D_{\mu} \quad \blacksquare$$

(A demonstração é similar ao lema 2.2.1).

Nestas duas últimas subseções, observamos que as matrizes diagonais, que contém os valores singulares das matrizes $(R-1c^T)$ e $(C-1r^T)$, são iguais. Além disso, há reciprocidade entre os eixos dos dois conjuntos de perfis, os quais somam-se às particularidades da Análise de Correspondência.

A seguir mostraremos que os centróides da projeção dos perfis nos subespaços coincidem com as origens deles.

LEMA 2.2.3. O centróide das coordenadas dos perfis das linhas coincide com a origem dos subespaços, isto é,
 $r^T F = 0^T$

DEMONSTRAÇÃO:

Utilizando (2.2.5), temos

$$\begin{aligned} r^T F &= r^T (R-1c^T) D_c^{-1} B = (r^T R - C^T) D_c^{-1} B \\ &= (c^T - c^T) D_c^{-1} B = 0^T \blacksquare \end{aligned}$$

LEMA 2.2.4. O centróide das coordenadas dos perfis das colunas coincide com a origem dos subespaços, isto é,
 $c^T G = 0^T$.

DEMONSTRAÇÃO:

Considerando (2.2.11), temos:

$$c^T G = c^T (C-1r^T) D_r^{-1} V = (r^T - r^T) D_r^{-1} V = 0^T \blacksquare$$

2.2.3. TRANSIÇÃO DAS COORDENADAS

A Análise de Correspondência permite utilizar os resultados dos perfis das linhas (colunas) para a análise dos perfis das colunas (linhas) através de fórmulas de transição, as quais trataremos nesta subseção. Para isto, mostraremos

primeiramente o seguinte lema.

LEMA 2.2.5. As coordenadas dos perfis das linhas e colunas estão relacionados pela seguinte equação $V^T F = B^T G$.

DEMONSTRAÇÃO:

De (2.2.7) temos:

$$F = D_r^{-1} V D_\mu$$

pré-multiplicando por V^T , e considerando $V^T D_r^{-1} V = I$, temos:

$$V^T F = V^T D_r^{-1} V D_\mu = D_\mu \quad (2.2.12)$$

do lema 2.2.2,

$$G = D_c^{-1} B D_\mu.$$

pré-multiplicando por B^T , e considerando $B^T D_c^{-1} B = I$ temos:

$$B^T G = B^T D_c^{-1} B D_\mu = D_\mu \quad (2.2.13)$$

logo de (2.2.12) e (2.2.13) resulta

$$V^T F = B^T G \quad \blacksquare$$

No que segue mostraremos a transição das coordenadas:

A) DOS PERFIS DAS COLUNAS PARA AS LINHAS

pós-multiplicando G , a (2.2.1) tem-se

$$(P - rc^T)G = V D_\mu B^T G$$

$$PG - rc^T G = V D_\mu B^T G$$

considerando o Lema 2.2.4 e a relação (2.2.13) resulta:

$$P G = V D_{\mu}^2$$

$$P G D_{\mu}^{-1} = V D_{\mu} \quad (2.2.14)$$

Agora, substituindo (2.2.14) em (2.2.7)

obtem-se:

$$F = D_r^{-1} P G D_{\mu}^{-1}$$

por (1.2.4) temos:

$$F = R G D_{\mu}^{-1} \quad (2.2.15)$$

B) DOS PERFIS DAS LINHAS PARA AS COLUNAS

Pós-multiplicando por F á transposta de (2.2.1) tem-se

$$(P^T - r c^T) F = B D_{\mu} V^T F$$

$$P^T F - c r^T F = B D_{\mu} V^T F$$

considerando o lema (2.2.3) e a relação (2.2.12) nesta igualdade temos:

$$P^T F = B D_{\mu}^2$$

$$P^T F D_{\mu}^{-1} = B D_{\mu} \quad (2.2.16)$$

substituindo (2.2.16) e considerando o lema 2.2.2 obtem - se:

$$G = D_c^{-1} P^T F D_{\mu}^{-1}$$

por (1.2.5) temos:

$$G = C F D_{\mu}^{-1} \quad (2.2.17)$$

As fórmulas (2.2.15) e (2.2.17) são conhecidas como *Fórmulas de Transição*. Observe nessas fórmulas que os dois

conjuntos de coordenadas F e G estão relacionados. Este relacionamento mostra que a correspondência entre pontos de diferentes espaços euclidianos ponderados são governados por baricentros.

Assim, a j -ésima linha de G , da fórmula (2.2.17) é igual ao baricentro $c_j^T F$ seguido por uma expansão na escala de $1/\mu_k$ sobre o k -ésimo subespaço. Os coeficientes do baricentro são os elementos do j -ésimo perfil de coluna. Simetricamente a i -ésima linha de F da fórmula (2.2.15) é igual ao baricentro $r_i^T G$, seguida pela mesma expansão, onde os coeficientes do baricentro são os elementos do i -ésimo perfil de linhas.

Para observar graficamente as relações entre as linhas e colunas da matriz N , alocamos as coordenadas F e G nos subespaços de interesse. Mas para interpretar estes gráficos precisamos de uma informação adicional porque estes pontos mostram a projeção dos perfis sobre os subespaços, e, logicamente, este gráfico não mostra quais dos pontos estão perto ou longe do subespaço que estamos considerando. Por isso, na seção seguinte deste capítulo, desenvolveremos esta informação adicional para saber onde a exposição é precisa e onde não é. Isto é análogo para muitas outras áreas de Estatística. Por exemplo, na construção de um modelo para dados, nós estudamos a qualidade do ajuste do modelo, isto é, onde o ajuste do modelo nos dados é satisfatório ou não.

2.3. DECOMPOSIÇÃO DA INÉRCIA E INTERPRETAÇÃO DOS PERFIS NOS SUBESPAÇOS

A inércia total de uma matriz de dados pode ser decomposta sobre os eixos principais, a qual é análoga à decomposição da soma de quadrados na análise de variância.

Nas seguintes tabelas, é apresentada a decomposição da inércia para cada conjunto de pontos que representam às linhas e colunas, respectivamente. Temos, na última linha das tabelas, as

TABELA 2.3.1
DECOMPOSIÇÃO DE INÉRCIA DAS LINHAS DE $(R - 1c^T)$.
EIXOS

	1	2	.	.	.	k	TOTAL
1	$r_1 f_{11}^2$	$r_1 f_{12}^2$	.	.	.	$r_1 f_{1k}^2$	$r_1 \sum f_{1k}^2$
2	$r_2 f_{21}^2$	$r_2 f_{22}^2$	.	.	.	$r_2 f_{2k}^2$	$r_2 \sum f_{2k}^2$
.	.	.	.	.	.	.	.
I	$r_I f_{I1}^2$	$r_I f_{I2}^2$	.	.	.	$r_I f_{Ik}^2$	$r_I \sum f_{Ik}^2$
TOTAL	μ_1^2	μ_2^2	.	.	.	μ_k^2	In(I)

Observação: $r_i f_{ik}^2$ é a Contribuição do i-ésimo perfil de linha para a inércia do k-ésimo eixo. (f_{ik} elementos de F)

TABELA 2.3.2
DECOMPOSIÇÃO DE INÉRCIA DAS LINHAS DE $(C - 1r^T)$.
EIXOS

	1	2	.	.	.	K	TOTAL
1	$c_1 g_{11}^2$	$c_1 g_{12}^2$	.	.	.	$c_1 g_{1k}^2$	$c_1 \sum g_{1k}^2$
2	$c_2 g_{21}^2$	$c_2 g_{22}^2$	.	.	.	$c_2 g_{2k}^2$	$c_2 \sum g_{2k}^2$
.	.	.	.	.	.	.	.
J	$c_J g_{J1}^2$	$c_J g_{J2}^2$	.	.	.	$c_J g_{Jk}^2$	$c_J \sum g_{Jk}^2$
TOTAL	μ_1^2	μ_2^2	.	.	.	μ_k^2	In(J)

Observação: $c_j g_{jk}^2$ é a Contribuição do j-ésimo perfil de coluna para a inércia do k-ésimo eixo. (g_{jk} elementos de G)

inércias decompostas sobre cada eixo , na última coluna das tabelas a soma da inércia total da cada ponto com respeito aos eixos, e na intersecção da última linha com a última coluna, temos a inércia total da matriz de dados.

Para interpretar as relações entre as linhas e colunas de uma T.C. sobre os subespaços gerados pelos eixos obtidos através da análise de correspondência, é necessário analisar dois conjuntos de coeficientes para cada um dos perfis em correspondência:

- 1) A contribuição absoluta, que exprime o aporte de um perfil para a inércia "explicada" em um subespaço.
- 2) A contribuição relativa ou correlação entre um eixo e o perfil, que exprime o aporte de um subespaço à explicação da dispersão do perfil.

2.3.1. CONTRIBUIÇÃO ABSOLUTA

Cada perfil contribui para a inércia decomposta em cada eixo, o qual depende de sua massa e da distância ao quadrado de sua projeção sobre o subespaço e o centróide. Assim, o quociente

$$CTR(i) = r_i f_{ik}^2 / \mu_k^2$$

mede a proporção de contribuição do i-ésimo perfil de linha para a inércia explicada no k-ésimo eixo.

Este quociente é chamado de *Contribuição Absoluta* do i-ésimo perfil linha para o k-ésimo eixo.

Observe que : $\sum_{i=1}^I CTR_k(i) = 1$ para todos os eixos.

Da mesma maneira é definida a contribuição absoluta para o j-ésimo perfil coluna para o k-ésimo eixo por:

$$CTR_k(j) = c_j g_{jk}^2 / \mu^2$$

onde:

$$\cos \theta_k = f_{ik}/d_i \quad \text{e} \quad \cos^2 \theta_k = (f_{ik}/d_i)^2$$

Cada elemento de (2.3.1) nos mostrará a qualidade de representação do perfil r_i pelo k-ésimo eixo. Este valor chama-se *Contribuição Relativa* do k-ésimo eixo para o i-ésimo perfil de linha. Observe que a contribuição relativa não depende das massas.

Da mesma maneira, definimos a contribuição relativa do k-ésimo eixo para o j-ésimo perfil de coluna como:

$$\cos^2 \theta_{kj} = g_{jk} / d_{j\chi}^2,$$

Agora, se nós queremos saber a qualidade de representação do perfil r_i pelo plano gerado pelo l-ésimo eixo e m-ésimo eixo, a qualidade será obtida pela soma de $\cos^2 \theta_l + \cos^2 \theta_m$.

Para finalizar este capítulo, descrevemos o que é um perfil suplementar e a sua utilidade.

2.4. PERFIS SUPLEMENTARES

Os perfis suplementares são de grande interesse na análise de dados. Aplicações destes perfis (ver Hatziparadissis A., 1987, Tekaa, F., Saint Martin, J., Sansonetti, P., Vusllier, F. & Claveirie, J. M., 1988)

Um perfil suplementar é aquele perfil que não participa ativamente nas análises, isto é, que não é utilizado para o cálculo do centróide, inércia total e base dos subespaços, porque a massa associada a ele é zero. Mas as coordenadas deste perfil são calculadas em função das bases e o centróide dos demais perfis ativos logo representadas nos subespaços ótimos. Por exemplo se r_s é um perfil suplementar linha, as coordenadas deste perfil nos

2.3.2. CONTRIBUIÇÃO RELATIVA

A qualidade de um perfil ser representado pelos eixos é obtida pelo $\cos^2 \theta_k$, onde θ_k é o ângulo entre o perfil e a projeção do perfil sobre o k -ésimo eixo que mede a proximidade entre eles. Assim, por exemplo, se a matriz $(P-rc^T)$ é de posto $k=2$, teremos dois subespaços. Observe na figura 2.1, a projeção do perfil r_i sobre os dois eixos é dado pelo $r_i \cos \theta_1$ e $r_i \cos \theta_2$, logo a distância do perfil r_i ao centróide se decompõe sobre seus eixos como $r_i \cos \theta_1$, $r_i \cos \theta_2$, e pelo Teorema de Pitágoras temos que:

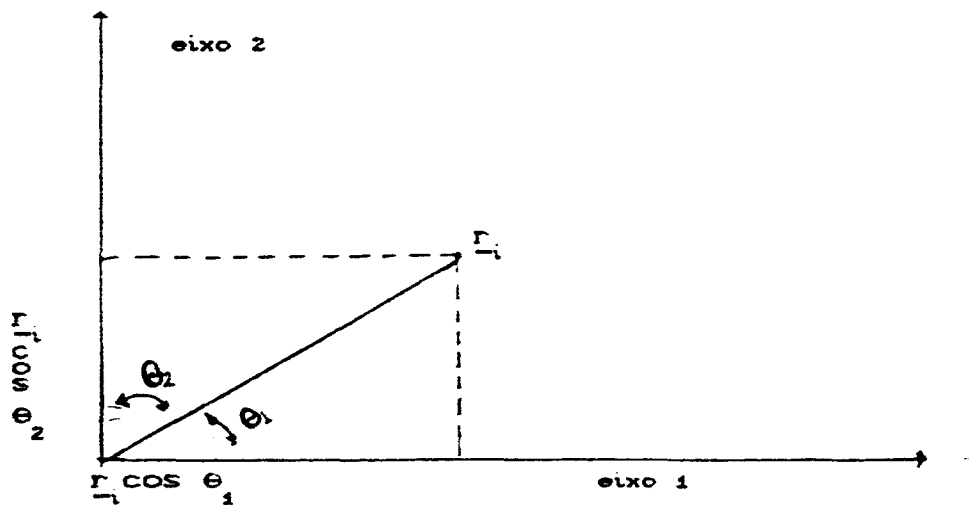
$$\cos^2 \theta_1 + \cos^2 \theta_2 = 1.$$


FIGURA 2.1. POSIÇÃO DO PERFIL r_i SOBRE UM ESPAÇO BI-DIMENSIONAL

Agora, se a matriz $(P-rc^T)$ é de posto $k = k^*$ a posição do perfil no espaço k^* -dimensional é dado por:

$$\sum_{k=1}^{k^*} \cos^2 \theta_k = 1 \quad (2.3.1)$$

subespaços é determinada utilizando a fórmula (2.2.18) i, é

$$F_e = (r_{1e}, r_{2e}, \dots, r_{1e}) G D_{\mu}^{-1}$$

De maneira similar, acha-se as coordenadas dos perfis coluna suplementares c_e , mas utilizando a fórmula (2.2.20).

Os perfis suplementares são utilizados na análise de uma T.C. quando:

- Existem observações aberrantes que são observadas através de uma análise preliminar.
- Temos um caso novo.
- Quando os elementos são de natureza diferente (variáveis quantitativas, qualitativas)

(ver P.Cazes (1982)).

As variáveis cujas categorias são considerados como perfis suplementares chama-se de *Variáveis Inativas* e caso contrario *Variáveis Ativas*.

Aplicamos a teoria desenvolvida neste capítulo a três conjuntos de dados, os quais apresentamos no capítulo III, IV, e V. Os cálculos numéricos foram feitas num Terminal ligado ao VAX/VMS instalado no laboratório de Estatística do IMECC - UNICAMP. Usamos a subrotina FO2WCF da NAG para achar os valores e vetores singulares. O programa principal esta listado no final do trabalho.

CAPITULO III

ESTUDO DA PRODUÇÃO DE GRÃOS NO BRASIL

3.1. INTRODUÇÃO

Nós estudaremos, neste capítulo, a evolução da produção de grãos no Brasil, através da análise de correspondência desenvolvida no capítulo II.

O objetivo deste estudo é mostrar, em forma conjunta, a evolução da produção dos grãos durante as safras 1979/80 a 1986/87.

Os dados para o estudo foram obtidos na fundação Instituto Brasileiro de Geografia e Estatística (IBGE), os quais apresentamos na tabela 3.0.

TABELA 3.0

PRODUÇÃO DE GRÃOS NO BRASIL, SAFRAS 1979/80 A 1986/87
(EM MIL SACAS DE 60KG)

=====					
GRÃOS					
=====					
SAFRAS	ARROZ(A)	FEIJÃO(C)	MILHO(M)	SOJA(S)	TRIGO(T)
=====					
1979/80 (S0)	162930	32800	339530	252600	45030
1980/81 (S1)	137130	39020	351950	250120	36830
1981/82 (S2)	162250	48380	364030	213930	30450
1982/83 (S3)	129030	26350	312180	243030	37280
1983/84 (S4)	150450	43770	352730	259020	33050
1984/85 (S5)	150420	42480	366970	304650	72000
1985/86 (S6)	173420	36980	342350	222250	93970
1986/87 (S7)	173700	33650	446450	280230	98150
=====					

Fonte: Fundação Instituto Brasileiro de Geografia e Estatística.

Nas linhas está representada a variável safra com oito categorias e nas colunas a variável grão com cinco categorias. Utilizaremos as letras que estão dentro do parênteses, nas seguintes seções para representá-las.

3.2. TABELAS, GRÁFICOS E ESTUDOS

O número total de sacas em mil sacas de 60kg. consideradas para o estudo são 6891540.

Primeiramente mostramos na tabela 3.1 a matriz de correspondência e a frequência da cada variável em suas categorias.

TABELA 3.1
MATRIZ DE CORRESPONDÊNCIA E FREQUÊNCIAS RELATIVAS

SAFRA .	GRÃOS					FREQUÊNCIA
	A	F	M	S	T	
S0	0,0236	0,0048	0,0493	0,0367	0,0065	0,1209
S1	0,0199	0,0057	0,0511	0,0363	0,0053	0,1183
S2	0,0235	0,0070	0,0528	0,0310	0,0044	0,1188
S3	0,0187	0,0038	0,0453	0,0353	0,0054	0,1085
S4	0,0218	0,0064	0,0512	0,0376	0,0048	0,1217
S5	0,0218	0,0062	0,0533	0,0442	0,0105	0,1360
S6	0,0252	0,0054	0,0497	0,0323	0,0136	0,1261
S7	0,0252	0,0049	0,0648	0,0407	0,0142	0,1498
FREQUÊNCIA	0,1799	0,0440	0,4174	0,2940	0,0648	1

Na análise de correspondência a última coluna desta tabela representa o vetor massa dos perfis das linhas e o centróide dos perfis das colunas, da mesma forma a última linha da tabela representa o vetor massa dos perfis das colunas e o centróide dos perfis das linhas.

As massas da safra 1986/87 e dos grãos milho e soja são maiores que as outras categorias devido à maior frequência observada nelas.

Em seguida, aplicamos a análise de correspondência à tabela 3.0. Desta aplicação, obtemos a tabela 3.2 na qual apresentamos os valores singulares generalizados e proporção de inércia explicada por cada eixo. Nesta tabela observamos que o primeiro eixo explica o 70,89% da inércia total; considerando os dois primeiros eixos temos uma inércia explicada de quase o 90%. Esta situação parece ser muito vantajosa, pois poderíamos visualizar a estrutura da T.C. em um plano, que explica o 90% da inércia total.

TABELA 3.2

VALORES SINGULARES GENERALIZADOS INÉRCIA E PROPORÇÃO DE INÉRCIA EXPLICADA.

EIXO	VALORES SINGULARES	INÉRCIA	PROP. DE INÉRCIA EXPLICADA	
			ABSOLUTA(%)	ACUMULADA(%)
1	0,1056	0,0112	70,89	70,89
2	0,0551	0,0030	18,99	89,89
3	0,0302	0,0009	5,70	95,58
4	0,0259	0,0007	4,43	100,00

Na tabela 3.3, mostramos as coordenadas dos perfis das variáveis grãos e safras para os três primeiros eixos principais. A projeção dos perfis sobre os planos gerados pelos eixos 1 e 2, e eixos 1 e 3 apresentamos nas figuras 3.1 e 3.2 respectivamente.

Para interpretar os pontos que mostram os gráficos, analisamos a contribuição absoluta da cada perfil para a inércia dos eixos, e contribuição relativa que mostra como os eixos os representam. Estas contribuições expomos nas tabelas 3.4 e 3.5 respectivamente.

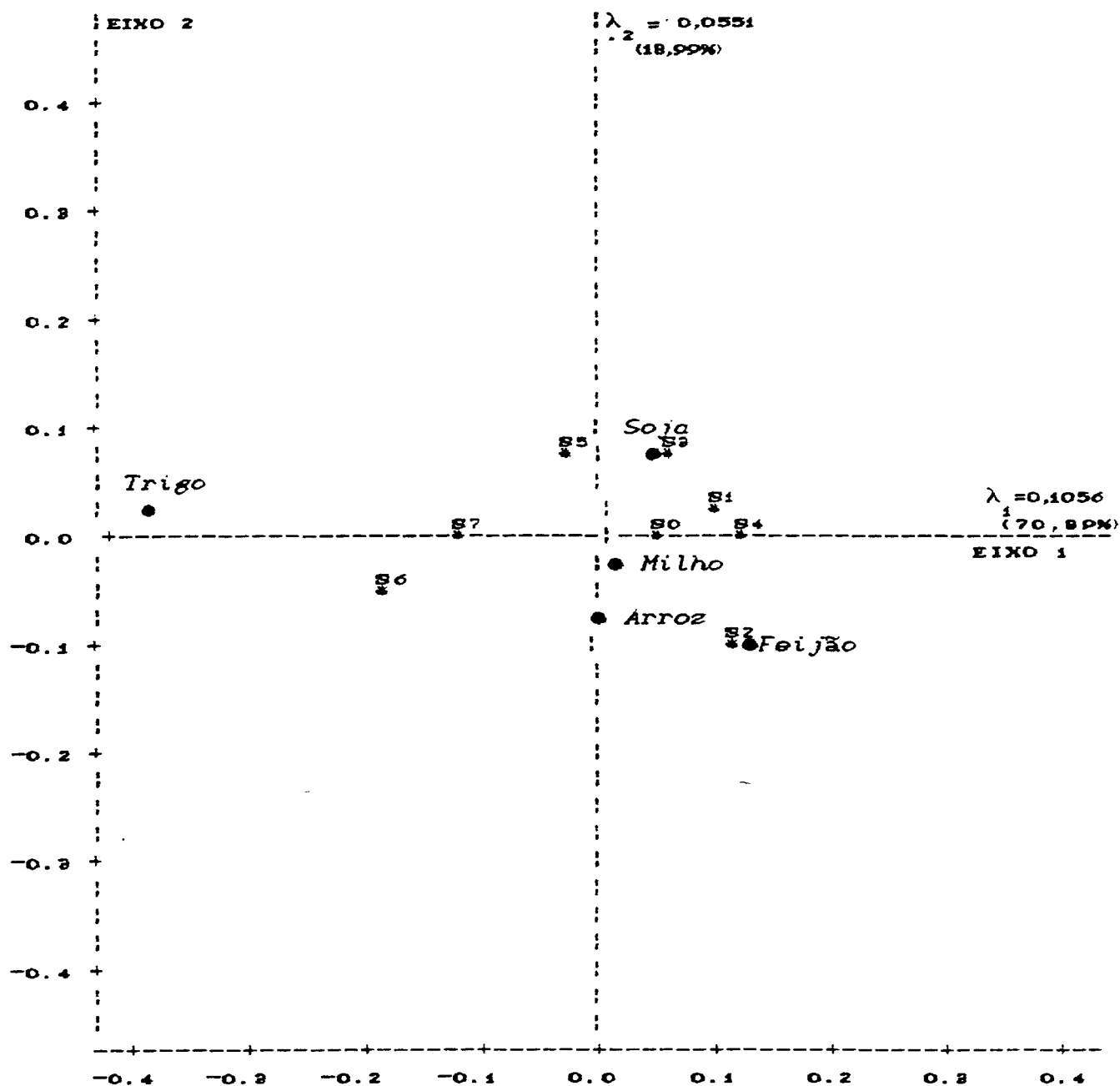


FIGURA 3.1. PROJEÇÃO DOS PONTOS QUE REPRESENTAM AS SAFRAS E OS GRÃOS SOBRE O PLANO GERADO PELOS EIXOS 1 E 2. O SÍMBOLO ●, REPRESENTA OS GRÃOS E O SÍMBOLO * REPRESENTA AS SAFRAS.

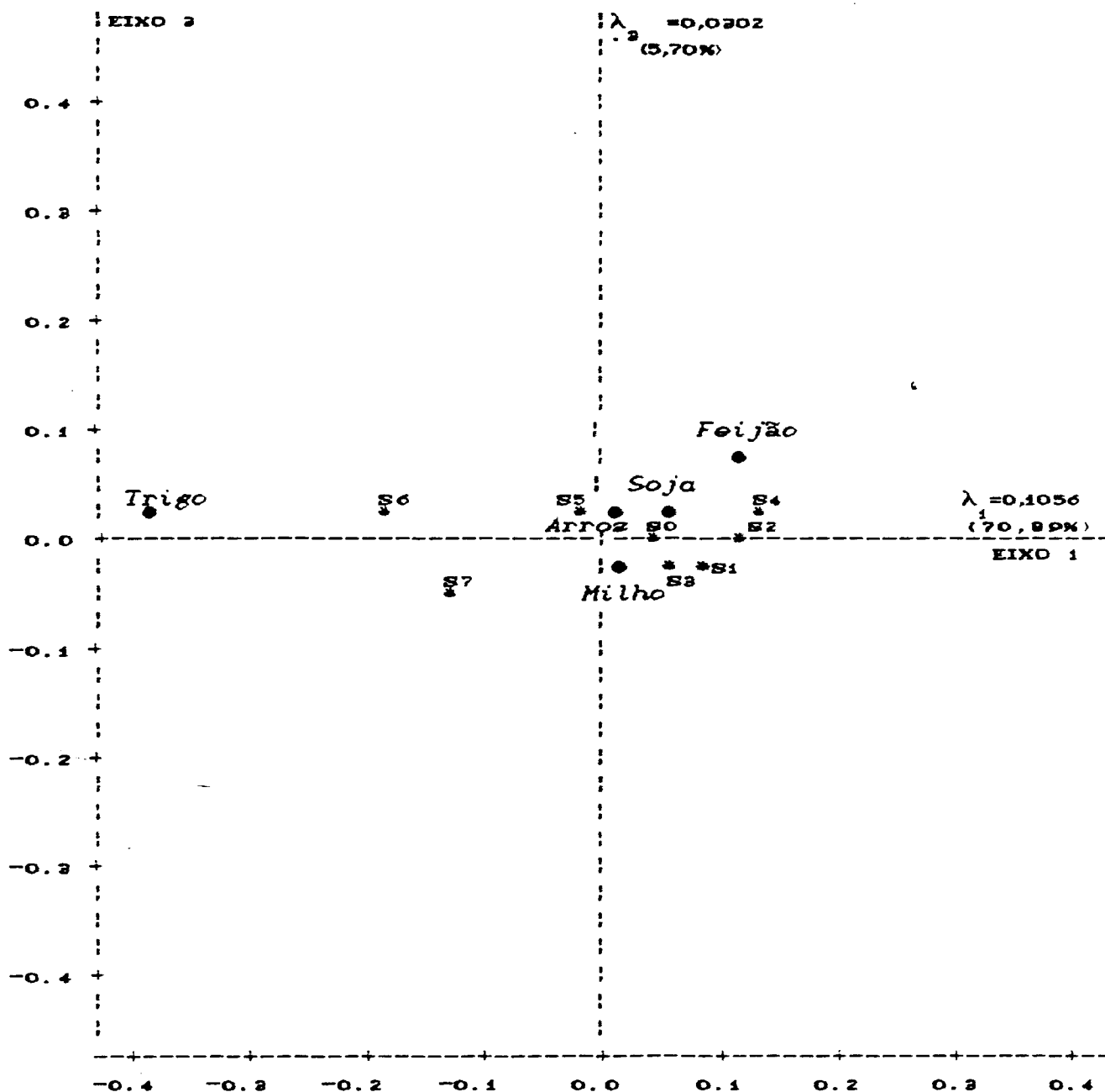


FIGURA 3.2. PROJEÇÃO DOS PONTOS QUE REPRESENTAM OS GRÃOS E SAFRAS SOBRE O PLANO GERADO PELOS EIXOS 1 E 2. O SÍMBOLO ● REPRESENTA OS GRÃOS E O SÍMBOLO * REPRESENTA AS SAFRAS.

TABELA 3.3

COORDENADAS DOS PERFIS DAS LINHAS E COLUNAS, SOBRE OS TRÊS PRIMEIROS EIXOS

SAFRAS	1	2	3
S0	0.0370	0.0030	0.0115
S1	0.0846	0.0158	-0,0162
S2	0,1105	-0,1084	-0,0044
S3	0,0585	0,0636	-0,0186
S4	0,1093	0,0002	0,0148
S5	-0,0357	0,0750	0,0374
S6	-0,1814	-0,0565	0,0350
S7	-0,1304	0,0043	-0,0549
GRÃOS	1	2	3
A	0,0005	-0,0651	0,0245
F	0,1114	-0,0956	0,0790
M	0,0195	-0,0178	-0,0324
S	0,0417	0,0766	0,0164
T	-0,3916	0,0126	0,0128

TABELA 3.4

CONTRIBUIÇÃO ABSOLUTA DOS PERFIS PARA OS TRÊS PRIMEIROS EIXOS

SAFRA	1	2	3
S0	0,0148	0,0004	0,0177
S1	0,0759	0,0098	0,0342
S2	0,1302	0,4598	0,0026
S3	0,0333	0,1446	0,0413
S4	0,1303	0,0000	0,0293
S5	0,0155	0,2520	0,2089
S6	0,3720	0,1325	0,1696
S7	0,2281	0,0009	0,4965
GRÃOS	1	2	3
A	0,0000	0,2515	0,1185
F	0,0490	0,1326	0,3019
M	0,0143	0,0435	0,4814
S	0,0458	0,5690	0,0864
T	0,8910	0,0034	0,0117

Observando a primeira coluna da tabela 3.4 que corresponde à contribuição absoluta dos perfis ao primeiro eixo, vemos que a porcentagem de contribuição das categorias S6 e S7 da variável safra e a categoria trigo da variável grão são maiores, indicando dessa maneira uma boa representação no primeiro eixo. Similarmente, no segundo eixo a porcentagem de contribuição das categorias S2 da variável safra e as categorias soja e arroz da variável grão são mais elevadas, garantindo assim uma boa representação neste eixo.

Geralmente quando as contribuições absolutas de certas categorias são elevadas as contribuições relativas dos subespaços para estas categorias também são, observa-se esta situação na tabela 3.5. Veja que as categorias S1, S4 da variável safra estão bem representadas pelo primeiro eixo sendo a contribuição absoluta delas muito pequenas. O 99% da inércia total da variável trigo está explicada no primeiro eixo.

TABELA 3.5
CONTRIBUIÇÃO RELATIVA ABSOLUTA E ACUMULADA DOS TRÊS PRIMEIROS
EIXOS PARA OS PERFIS.

CATEGÓRIAS	ABSOLUTA			ACUMULADA		
	EIXOS			EIXOS		
	1	2	3	1+2	1+3	1+2+3
S0	0,3263	0,0022	0,0318	0,3285	0,3581	0,3603
S1	0,8832	0,0310	0,0325	0,9142	0,9157	0,9467
S2	0,5049	0,4852	0,0008	0,9901	0,5057	0,9909
S3	0,3892	0,4603	0,0394	0,8495	0,4285	0,8889
S4	0,9747	0,0000	0,0179	0,9747	0,9926	0,9926
S5	0,1381	0,6092	0,1514	0,7474	0,2896	0,8988
S6	0,8796	0,0852	0,0327	0,9648	0,9123	0,9975
S7	0,8452	0,0009	0,1500	0,8461	0,9952	0,9961
A	0,0000	0,6512	0,0920	0,6512	0,0920	0,7432
F	0,3550	0,2615	0,1784	0,6164	0,5334	0,7949
M	0,2099	0,1741	0,5777	0,3840	0,7876	0,9617
S	0,2201	0,7446	0,0339	0,9647	0,2540	0,9986
T	0,9960	0,0010	0,0011	0,9970	0,9970	0,9981

Somando as contribuições relativas dos perfis do primeiro e segundo eixo entre si, observamos que todos os perfis estão bem representados pelo plano gerado por estes primeiros eixos, a exceção do perfil coluna S0 e o perfil linha M.

Devido à qualidade de representação dos perfis pelos planos 1x2 e 1x3, nós estudamos a correspondência entre as safras e grãos baseados nos gráficos 3.1, 3.2 e tabela 3.4 e 3.5.

A) ESTUDO DAS SAFRAS 1979/80 A 1986/87

EIXO 1:

A projeção das safras S6 e S7 no lado esquerdo do eixo e a projeção das safras S1 e S4 no lado direito do eixo produzem uma contraposição entre estas safras, da qual podemos concluir que a porcentagem de produção de grãos nestas safras são diferentes, e da proximidade entre as safras S1 e S4 concluímos que a porcentagem de produção de grãos nestas safras são similares.

EIXO 2:

Este eixo contrapõe as safras S5, S3 com S2. e mostra a aproximação entre as safras S5 e S3.

PLANO 1x2:

Com a elevada porcentagem de inércia explicada 89,99%, observamos neste plano as variações moderadas de amplitude entre as safras. Mas considerando o transcurso do tempo vemos variações maiores de amplitude entre as safras S2 e S3, S4 e S5, e S5 e S6. A posição dos pontos que representam as safras descrevem uma curva de forma parabólica.

PLANO 1x3:

Este plano revela uma nova relação não linear, os pontos descrevem uma curva de forma cúbica. Além disso, mostra uma variação moderada de amplitude entre as safras e considerando a ordem a variação maior de amplitude é dada entre as safras S4 e S5.

B) ESTUDO DA PRODUÇÃO DOS GRÃOS

EIXO 1:

A projeção da produção do grão trigo no extremo esquerdo do eixo mostra o afastamento dos demais grãos.

EIXO 2:

Este eixo contrapõe os grãos arroz e soja.

PLANO 1x2:

Observamos, neste plano, a contraposição global do grão trigo com os grãos arroz, feijão, milho, e soja.

PLANO 1x3:

Este plano mostra o mesmo resultado que o anterior.

C) ESTUDO DA CORRESPONDÊNCIA ENTRE A PRODUÇÃO DE GRÃOS E SAFRAS.

O estudo é baseado simultaneamente sobre os planos 1x2 e plano 1x3.

A produção do grão soja está associada à safra S3 e está mais perto das safras S1, S4, e S5.

A produção do grão feijão está relacionada à safra S2 e também está perto das safras S4 e S1.

A produção do grão arroz está contornado por todas as safras, tendo um ligeiro afastamento a partir da safra S5 em diante.

A produção do grão milho está perto do centróide e está relacionado com a safra S0.

A produção do grão trigo está menos distante da safra S6, aumentando a produção dele à partir da safra S5 em diante.

3.3. CONCLUSÕES

Nos diferentes estudos percebemos um aumento gradual da produção do grão trigo nas safras 1984/85, 1986/87, 1995/86 e um retrocesso na produção dos grãos soja, milho e feijão.

A produção dos grãos arroz e milho foi quase igual em todas as safras.

Existe uma irregularidade na ordem das safras, conseqüentemente na produção de grãos, devendo-se talvez a fatores climatológicos, investimento de capital, desenvolvimento industrial, apoio do governo, etc.

CAPITULO IV

UM ESTUDO SOBRE OS VEÍCULOS ENVOLVIDOS EM ACIDENTES DE TRÂNSITO

4.1. INTRODUÇÃO

Neste capítulo estudamos às espécies de veículos que foram envolvidos nos acidentes de trânsito com vítimas nas Unidades de Federação^o (U.F.) do Brasil durante o ano de 1985; e também estudamos os acidentes de trânsito ocorridos na Área e o Período nas U.F. durante o mesmo ano.

O objetivo principal é determinar que espécie de veículo foi o mais envolvido nos acidentes de trânsito e em que unidades de federação, além disso observar o comportamento dos acidentes de trânsito ocorridos no Período e Área com respeito às anteriores.

4.2. DESCRIÇÃO DAS VARIÁVEIS

Para se analisar este problema, conta-se com quatro variáveis, duas variáveis ativas (Ver seção 2.4), Unidades de Federação com 26 categorias e Espécie de Veículo com 7 categorias, e as duas últimas variáveis Área e Período como inativas (Ver seção 2.4).

4.2.1. LISTA DAS VARIÁVEIS

As letras que estão dentro do parênteses são as abreviações das U.F. e espécie de veículo. Daqui em diante utilizamos estas abreviações para identificar as categorias.

LISTA DAS UNIDADES FEDERATIVAS

REGIAO NORTE

Rondônia (RO)

Acre (AC)

Amazonas (AM)

Roraima (RR)

Pará (PA)

Amapá (AP)

REGIAO NORDESTE

Ceará (CE)

Rio Grande do Norte (RN)

Paraíba (PB)

Pernambuco (PE)

Alagoas (AL)

Sergipe (SE)

Bahia (BA)

Maranhão (MA)

Piauí (PI)

REGIÃO CENTRO - OESTE

Goiás (GO)

Mato Grosso (MT)

Mato Grosso do Sul (MS)

Distrito Federal (DF)

REGIÃO SUDESTE

Minas Gerais (MG)

Espírito Santo (ES)

Rio de Janeiro (RJ)

REGIÃO SUL

Paraná (PR)

Santa Catarina (SC)

Rio Grande do Sul (RS)

LISTA DE VEÍCULOS

Passeio (PAS)

Ônibus (ONI)

Motocicleta (MOT)

Outros (OUT)

Táxi (TAX)

Caminhão (CAM)

Bicicleta (BIC)

LISTA DAS VARIÁVEIS DICOTÔMICAS

ÁREA: Rural e Urbana.

PERÍODO: Dia e Noite.

4.2.2. APRESENTAÇÃO DA TABELA DE DADOS

Os dados apresentados nas tabelas 4.1 e 4.2 foram obtidos do Anuário Estatístico do Brasil, 1986- Ministério de Justiça, Dpto Nacional de Trânsito, Divisão de Pesquisas.

TABELA 4.1

VEÍCULOS ENVOLVIDOS EM ACIDENTES DE TRANSITO COM VÍTIMAS, POR ESPÉCIE DE VEÍCULO, SEGUNDO AS UNIDADES DE FEDERAÇÃO - 1985

ESPECIE DE VEICULOS							
U.F	PAS	TAX	ONI	CAMI	MOTO	BIC	OUT
RO	353	46	29	137	97	49	36
AC	54	12	5	32	33	3	5
AM	583	155	143	81	58	18	79
RR	37	1	1	22	20	7	3
PA	1246	111	159	319	103	43	44
AP	182	29	9	43	44	54	41
MA	698	14	165	315	129	72	244
PI	531	25	79	167	171	136	164
CE	2079	197	319	431	436	112	752
RN	1002	20	59	156	100	13	39
PB	452	31	46	125	63	19	35
PE	2985	240	291	526	377	160	640
AL	732	4	67	133	81	12	63
SE	455	6	63	409	57	35	103
BA	6146	654	1226	970	813	216	329
MG	15180	434	1579	3558	4757	2300	1381
ES	1929	49	161	436	325	103	82
RJ	23868	788	3287	2187	3484	476	2115
SP	94370	1027	6550	11656	26584	7641	8364
PR	12117	166	734	3108	3736	1326	1541
SC	8474	43	266	1503	4218	897	405
RS	17056	603	1244	2474	7190	1460	2453
MS	2698	11	149	463	1147	1136	298
MT	930	6	91	459	164	51	38
GO	3796	82	365	1366	1509	588	1156
DF	5205	140	491	354	641	384	77
FONTE - Ministerio da Justica, Departamento Nacional de Transito, Divisao de Pesquisas.							

TABELA 4.2

ACIDENTES DE TRANSITO REGISTRADO COM VÍTIMAS , POR VÁRIOS ASPECTOS E PESSOAS VITIMADAS, SEGUNDO AS UNIDADES DE FEDERACAO - 1985.

ACIDENTES DE TRANSITO REGISTRADOS COM VÍTIMAS				
UNIDADES FEDERA- TIVAS.	PERIODO		AREA	
	DIA	NOITE	URBANA	RURAL
RO	336	185	426	125
AC	80	24	82	22
AM	627	344	897	74
RR	35	21	42	14
PA	787	431	923	295
AP	150	95	202	43
MA	792	413	706	499
PI	560	280	703	137
CE	1569	1063	2347	285
RN	643	353	780	216
PB	348	163	360	151
PE	2690	1484	3632	542
AL	510	230	557	183
SE	395	226	494	127
BA	5729	2856	7240	1345
MG	15860	9771	20707	4924
ES	351	811	1542	620
RJ	14839	12912	26612	1139
SP	65033	42564	97405	10192
PR	8833	5872	10272	4433
SC	6921	3641	8477	2085
RS	13586	8171	18572	3185
MS	2534	1474	3637	371
MT	697	359	740	316
GO	3654	2171	4417	1408
DF	2725	1862	4531	56

FONTE - Anuario Estatístico do Brasil - 1986.

4.3. RESULTADOS DA ANÁLISE DE CORRESPONDÊNCIA E ESTUDO

Depois de aplicar a análise de correspondência aos dados da tabela 4.1, construímos a tabela 4.2 onde mostramos os valores singulares generalizados e a percentagem total de inércia explicada em cada um dos eixos.

TABELA 4.3

VALORES SINGULARES GENERALIZADOS, INÉRCIA EXPLICADA

EIXO	VALORES SINGULARES	INERÇIA	PROPORÇÃO DE INERÇIA TOTAL	
			ABSOLUTA(%)	ACUMULADA(%)
1	0,1899	0,0361	44,57	44,57
2	0,1402	0,0197	24,32	68,89
3	0,0970	0,0094	11,61	80,50
4	0,0930	0,0085	10,49	90,99
5	0,0764	0,0058	7,16	98,15
6	0,0393	0,0015	1,85	100,00

Nesta tabela observamos que o primeiro eixo explica o 44,57% da inércia total e o plano gerado pelos eixos 1 e 2 explica o 68,89% da inércia total.

Analisamos o conjunto de dados sobre os subespaços gerados pelos quatro primeiros eixos, devido a que a maioria dos perfis têm a maior correlação e contribuição absoluta nestes eixos o qual mostramos nas tabelas 4.4 e 4.5. Nestas tabelas eliminamos o ponto decimal para facilitar a impressão e análise, multiplicando todas as quantidades de natureza relativa por 10000.

Na primeira linha das tabelas introduzimos notações que representam o seguinte:

- COR Contribuição relativa do eixo para o perfil.
- QLT Qualidade do perfil ser representado pelo subespaço de nosso interesse.
- MAS Massa do perfil.
- INER Porcentagem de inércia total do perfil em todos os subespaços.
- FK Coordenada do perfil de linha no k-ésimo eixo.
- GK Coordenada do perfil de coluna no k-ésimo eixo
- CTR Contribuição Absoluta do perfil para a inércia do subespaço.

Estas notações serão também utilizadas nas posteriores tabelas.

TABELA 4.4

DECOMPOSIÇÃO DA INÉRCIA E COORDENADAS DOS PERFIS PARA OS DOIS PRIMEIROS EIXOS.

ESPAÇO DAS LINHAS

CATEG.	QLT	MASS	INER	F1	COR	CTR	F2	COR	CTR
RO	5732	21	78	-1407	688	12	3882	5084	163
AC	3800	4	34	-1491	330	3	4836	3470	44
AM	6988	32	529	-9126	6184	734	3290	804	175
RR	6344	3	12	3138	2526	7	3857	3818	20
PA	6781	58	316	-4244	5927	288	1611	854	76
AP	4012	11	68	-664	91	1	4359	3921	111
MA	6960	47	227	-1480	554	28	5031	6406	600
PI	7466	36	101	698	215	5	4050	7251	302
CE	6416	123	551	-3000	2483	307	3776	3933	893
RN	3859	39	68	-2046	3011	46	-1086	848	24
PB	6597	22	45	-2676	4361	44	1916	2236	41
PE	7482	149	371	-3075	4675	390	2383	2807	429
AL	2939	31	43	-1804	2893	28	-227	46	1
SE	5824	32	394	-420	18	2	7593	5806	42
BA	8259	295	1230	-5238	8113	2243	702	146	74
MG	4952	831	388	725	1391	121	1160	3561	569
ES	1709	88	82	-1099	1600	29	286	109	4
RJ	9351	1031	1406	-3034	8332	2631	-1061	1019	590
SP	9128	4447	538	520	2754	333	-791	6374	1414
PR	7736	647	330	1125	3065	227	1388	4671	634
SC	7245	450	692	2896	6731	1047	-800	514	146
RS	1893	92	454	834	1748	178	240	145	27
MS	4807	168	983	4576	4420	976	1354	387	157
MT	2372	50	250	-225	12	1	3108	2360	243
GO	8480	252	574	1310	931	120	3732	7549	1786
DF	6519	208	334	-1865	2666	200	-2242	3853	531

ESPAÇO DAS COLUNAS

CATEG.	QLT	MASS	INER	G1	COR	CTR	G2	COR	CTR
PAS	8963	5785	762	-490	2251	385	-846	6712	2106
TAX	7286	139	1916	-8351	6260	2695	3381	1026	810
ONI	8574	501	1202	-4082	8563	2312	148	11	6
CAM	6817	895	1555	483	166	58	3060	6651	4260
MOT	7560	1604	1626	2457	7352	2685	-413	208	139
BIC	5794	493	1631	3684	5061	1855	1402	733	493
OUT	4093	583	1308	-254	35	10	2715	4058	2186

* DIA	0	0	-197	315	0	289	701	0
* NOITE	0	0	-589	2027	0	-289	490	0
* RURAL	0	0	1319	558	0	3782	4587	0
* URBANA	0	0	-606	2215	0	-499	1497	0

* Perfis Suplementares

TABELA 4.5
DECOMPOSICÃO DA INÉRCIA E COORDENADAS DOS PERFIS PARA
O TERCEIRO E QUARTO EIXO.

ESPAÇO DAS LINHAS							
CATEGORIA	QLT	F3	COR	CTR	F4	COR	CTR
RO	1896	-2005	1357	90	1264	539	40
AC	572	-1318	258	8	1455	314	10
AM	2329	- 210	3	1	5594	2326	1174
RR	2084	-2471	1567	17	-1420	517	6
PA	2192	-2559	2155	401	336	37	8
AP	4369	- 694	99	6	4549	4270	279
MA	1453	939	233	44	-2205	1230	267
PI	686	485	104	9	1148	582	56
CE	3332	3407	3202	1519	687	130	68
RN	4460	-1213	1058	62	-2175	3402	221
PB	1990	-1555	1472	56	- 923	518	22
PE	1198	1499	1111	355	420	87	31
AL	6554	- 445	176	7	-2679	6378	263
SE	4111	-2744	758	257	-5770	3353	1261
BA	1284	-1435	609	645	1511	675	793
MG	4664	-1269	4261	1422	390	403	149
ES	7546	-1814	4358	307	-1551	3188	249
RJ	221	288	75	91	- 401	146	129
SP	494	154	242	112	- 157	252	129
PR	2037	- 297	213	60	- 867	1824	574
SC	68	- 241	46	28	166	22	15
RS	5948	1327	4423	1730	779	1525	662
MS	3587	-2288	1105	934	3429	2482	2329
MT	7180	-3421	2858	616	-4207	4322	1033
GO	1263	1387	1043	516	- 637	220	121
DF	2585	-1788	2450	705	419	135	43
ESPAÇO DAS COLUNAS							
CATEGORIA	QLT	G3	COR	CTR	G4	COR	CTR
PAS	556	- 96	86	57	- 224	470	342
TAX	1969	- 245	5	9	4678	1964	3595
ONI	197	- 524	141	146	329	56	64
CAM	2855	-1248	1107	1482	-1569	1748	2597
MOT	923	683	569	796	539	354	550
BIC	3078	-1834	1254	1762	2211	1824	2842
OUT	5116	3046	5107	5749	- 125	9	11
* DIA		- 417	1490	0	394	1292	0
* NOITE		- 117	80	0	23	3	0
* RURAL		-1508	729	0	- 887	252	0
* URBANA		- 118	84	0	- 419	1056	0

Perfis suplementares.

As projeções dos pontos que representam as U.F. e os Veículos sobre o plano gerado pelos eixos 1 e 2 e sobre o plano gerado pelos eixos 3 e 4 estão representados nas figuras 4.1 e 4.2.

Para interpretar os pontos nos gráficos, consideramos a CTR média de cada categoria sendo para as U.F. $(10000/26) = 385$ e para os Veículos $(10000/7) = 1429$. Interpretamos aquelas variáveis cujas CTR são maiores a 280 das U.F. e 1300 dos veículos. além disso aquelas cuja COR seja maior o igual a 3100.

AO ESTUDO DA CORRESPONDÊNCIA ENTRE UNIDADES FEDERATIVAS

EIXO 1:

Este eixo mostra o afastamento da U.F. de Amazonas das demais e a contraposição global das U.F. de Bahia, Pará, Pernambuco, Paraíba, Rio de Janeiro e Amazonas com as de Santa Catarina, Mato Grosso do Sul.

EIXO 2:

Contrapõe globalmente as U.F. de Ceará, Rondônia, Piauí, Goiás, Amapá, Roraima, Maranhão, Sergipe com Distrito Federal, São Paulo.

EIXO 3:

Mostra a contraposição das U.F. de Ceará, Rio Grande do Sul com Minas Gerais e Espírito Santo.

EIXO 4:

As U.F. de Rio Grande do Norte, Alagoas, Sergipe, Espírito Santo, Mato Grosso estão melhor representadas neste eixo.

PLANO 1x2:

É interessante notar neste plano duas nuvens de pontos bem diferenciadas composta pelas U.F. das regiões norte, nordeste com regiões centro-oeste, sudeste, e sul, com exceção das U.F. de Rio Grande do Norte, Alagoas, Goiás e Mato Grosso; perto

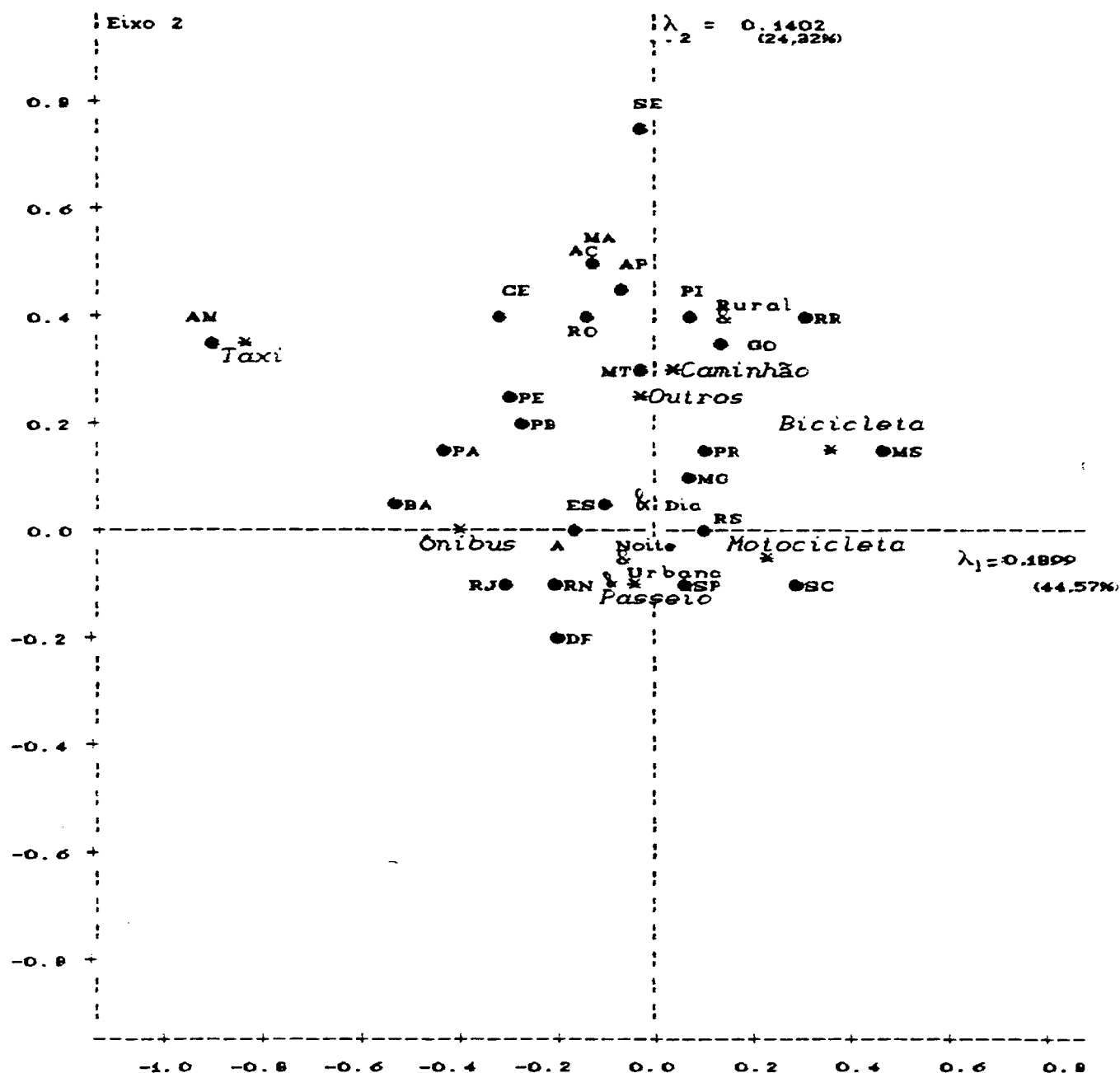


FIGURA 4.1. PROJEÇÃO DOS PONTOS QUE REPRESENTAM AS U.F. E AS ESPÉCIES DE VEÍCULOS SOBRE O PLANO GERADO PELOS EIXOS 1 E 2. O SÍMBOLO ●, REPRESENTA AS U.F E O SÍMBOLO * REPRESENTA AS ESPÉCIES DE VEÍCULO E O SÍMBOLO & REPRESENTA OS PERFIS SUPLEMENTARES.

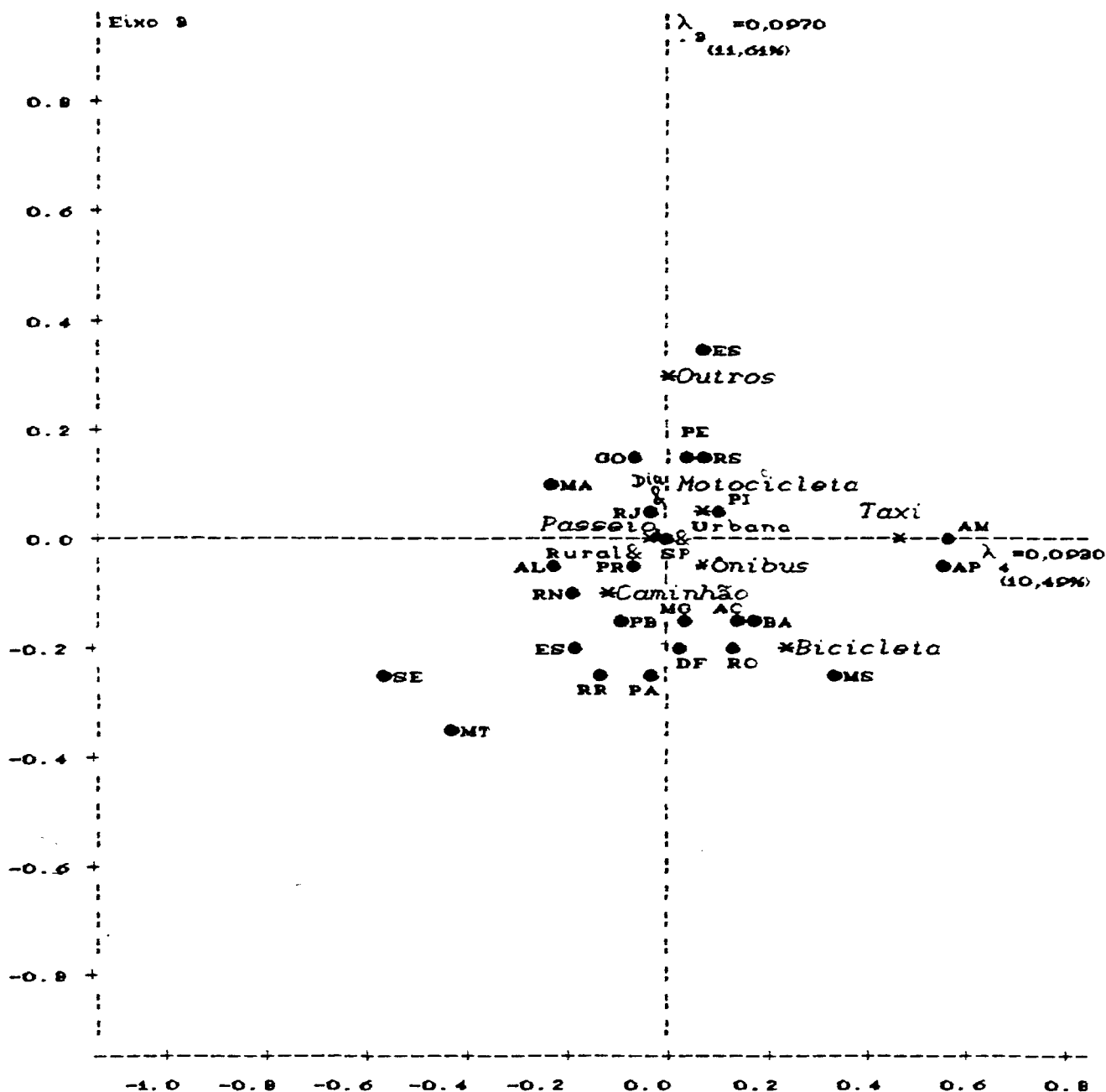


FIGURA 4.2. PROJEÇÃO DOS PONTOS QUE REPRESENTAM AS U.F. E AS ESPÉCIES DE VEÍCULOS SOBRE O PLANO GERADO PELOS EIXOS 3 E 4. O SÍMBOLO ● REPRESENTA AS U.F. E O SÍMBOLO * REPRESENTA AS ESPÉCIES DE VEÍCULO E O SÍMBOLO & REPRESENTA OS PERFIS SUPLEMENTARES

do centróide estão as U.F. de Minas Gerais, Paraná ,e São Paulo; observa-se no terceiro quadrante as U.F. de Rio de Janeiro e Distrito Federal.

PLANO 3x4:

Este plano mostra uma separação global das U.F. de Mato Grosso, Sergipe ,Espírito Santo, Rio Grande do Norte e Alagoas com as de Amapá, Rio Grande do Sul e Ceará; e uma aproximação entre as U.F. de Espírito Santo, Rio Grande do Norte e Alagoas.

B) ESTUDO DA CORRESPONDÊNCIA ENTRE ESPÉCIE DE VEÍCULO

EIXO 1:

Mostra a contraposição entre os veículos Táxi, Ônibus e Motocicleta, Bicicleta.

EIXO 2:

Neste eixo observamos a contraposição dos veículos Passeio com Caminhão e Outros.

EIXO 3:

Estabelece uma proximidade entre o veículo Caminhão e Bicicleta mas a correlação entre estes perfis e o subespaço não é elevada; o veículo Outros, está mais correlacionado com este eixo.

EIXO 4:

Define a contraposição entre os veículos Taxi, Bicicleta e Caminhão, mas a correlação destes perfis, neste eixo não é significativa.

PLANO 1x2:

Observamos dentro deste plano uma aproximação moderada ao centróide das espécies de veículos Ônibus, Caminhão,

Bicicleta, Motocicleta. O veículo Taxi se afasta dos demais veículos e o veículo passeio está perto do centróide.

PLANO 3x4:

Este plano mostra um afastamento entre os veículos Bicicleta e Outros.

C) ESTUDO DA CORRESPONDÊNCIA ENTRE A ESPÉCIE DE VEÍCULO E UNIDADES FEDERATIVAS.

Estudamos a correspondência entre espécie de veículo e U.F. baseados nas figuras 4.1 e 4.2 e nas tabelas 4.3 e 4.4.

PLANO 1x2:

Observamos uma associação do veículo Taxi com a U.F. de Amazonas; as U.F. de Bahia, Paraná, Pernambuco, Paraíba e Rio de Janeiro estão mais próximas ao veículo Ônibus; o veículo Passeio está perto do centróide do plano e está mais associado com as U.F. de São Paulo, Paraná e Minas Gerais; o veículo Outros está associado com as U.F. de Ceará, Piauí, Goiás, Rondônia e Amapá; o veículo Caminhão está associado com as U.F. de Piauí, Goiás e Roraima; A U.F. mais próxima do veículo Motocicleta é Santa Catarina; e o último veículo Bicicleta está mais relacionado com a U.F. de Mato Grosso do Sul.

PLANO 3x4:

No plano 3x4 observa-se uma proximidade entre o veículo Outros com as U.F. de Ceará e Rio Grande do Sul; não podemos tirar outras relações, devido à qualidade de representação dos perfis pelo plano não ser muito elevada.

A seguir, faremos um estudo dos veículos envolvidos em acidentes de trânsito, considerando o Período e Área como perfis suplementares. (tabela 4.2)

D) ESTUDO DA CORRESPONDENCIA ENTRE A ESPÉCIE DE VEÍCULO E UNIDADES DE FEDERAÇÃO CONSIDERANDO A ÁREA E O PERÍODO COMO PERFIS SUPLEMENTARES.

PLANO 1x2:

No que diz respeito aos perfis suplementares, observamos uma proximidade entre o período dia e noite; existe uma maior distância entre a Área rural e urbana.

As U.F. de Rio de Janeiro, Distrito Federal e São Paulo estão próximos do período noturno, zona urbana e veículos Passeio e Ônibus. De maneira semelhante as U.F. de Goiás, Piauí, Roraima, Maranhão, Rondônia , estão perto de zona rural, e veículos Caminhão e Outros.

PLANO 3x4:

A interpretação deste plano é dificultosa, pelo fato da projeção dos perfis suplementares estar perto do centróide. Além disso, a correlação destes perfis com o plano é baixa.

4.4. CONCLUSÕES

Do comportamento das espécies de veículos frente às U.F., vemos neste estudo que:

Os veículos Taxi, Bicicleta, Motocicleta foram envolvidos em maior número nos acidentes de trânsito ocorridos nas U.F. de Amazonas, Mato Grosso do Sul, Santa Catarina respectivamente.

Nas U.F. de São Paulo, Distrito Federal, o veículo que teve maior envolvimento nos acidentes de trânsito foi o de Passeio.

Nos acidentes ocorridos nas U.F. Bahia, Pará, Pernambuco, Paraíba, e Rio de Janeiro, o veículo Ônibus teve maior envolvimento.

O veículo caminhão foi o mais envolvido nos acidentes de trânsito ocorridos nas U.F. de Goiás e Piauí.

Nos acidentes de trânsito das U.F. Minas Gerais e Paraná os veículos que tiveram maior envolvimento foram Bicicleta, Caminhão, Motocicleta e Passeio.

Considerando os perfis suplementares concluímos que não existe muita variação nos acidentes de trânsito ocorridos no dia e noite; mas os acidentes de trânsito ocorridos na zona urbana e rural são mais variados. Existe uma tendência de que os acidentes ocorridos na zona urbana são no período da noite.

Nas U.F. de Rio de Janeiro, Distrito Federal e São Paulo o número de acidentes é maior no período noturno e zona urbana envolvendo o veículo Passeio, nas U.F. de Goiás, Piauí, Roraima, Maranhão, Rondônia, os acidentes de trânsito são maiores na zona rural envolvendo os veículos Caminhão e Outros.

CAPITULO V

ESTUDO DO MÊS DE MORTE, MÊS DE ANIVERSÁRIO E PROFISSÃO

5.1. INTRODUÇÃO

Será que a data de morte de uma pessoa é influenciada por fatores psicológicos, sociais ou outros? Um artigo publicado por Phillips 1976 mostra uma relação entre o mês de morte e mês de aniversário. Para tal resultado considerou uma amostra de pessoas muito famosas que atingiram alto status social e de pessoas pertencentes a famílias da aristocracia dos E.U.A.

O estudo se baseia na porcentagem de mortes relacionadas ao mês do aniversário, mostrando que à medida em que a pessoa é mais famosa, sua morte tende a ocorrer depois do aniversário; visto que, pela popularidade que a pessoa possui, o seu aniversário é geralmente celebrado com publicidade, recebendo uma atenção privilegiada.

Com estes resultados fazemos a seguinte pergunta: será que existe alguma relação entre o mês de morte, mês de aniversário e profissão de uma pessoa?

Neste capítulo estudamos este problema através do método de análise de correspondência.

5.2. APRESENTAÇÃO DOS DADOS

Nós estudamos a data de morte e profissão das pessoas que nasceram no transcurso dos séculos XV ao XIX, e que contribuíram com seus conhecimentos, investigações, arte, e idéias para o desenvolvimento da ciência, da cultura e sociologia humana.

Os dados foram obtidas das seguintes fontes de informações:

- Enciclopédia Gênios da Humanidade, vol 1,2,3
- The History of One Hundred Great Composiers
- Os Pensadores vol 1,2

Pela falta de informação sobre novelistas, poetas, escritores nas fontes, antes citadas, e devido ao nosso interesse em uma amostra nestas áreas, restringimos tal amostra aos Autores Paulistas coletados no "Diccionario de autores paulistas." que obtivemos através de uma amostragem sistemática, considerando os nascidos até o ano de 1900.

Não consideramos aquelas pessoas nascidas no ano de 1900 que não morreram até a data da edição da fonte de informação; com exceção da fonte :The history of one hundred great composiers.

A relação entre o aniversário e morte é estudada através do mês de nascimento e mês de morte. Para o estudo, se diz que uma pessoa morreu no mês de seu aniversário se o mês de sua morte tem o mesmo nome de seu mês de aniversário.

Estudamos o nascimento, morte e profissão de 774 pessoas as quais classificamos segundo duas variáveis.

Primeira variável:

Mês de morte com respeito ao mês de aniversário, com 12 categorias, as quais denotamos por:

1MA, 2MA, 3MA, 4MA, 5MA, 6MA, MMA, 1MD, 2MD, 3MD, 4MD, 5MD, 6MD

onde o número representa o mês de morte com respeito ao mês de aniversário, a letra M representa a palavra mês e a última letra A ou D indica se a morte da pessoa foi antes ou depois do aniversário respectivamente. A categoria MMA representa a morte de uma pessoa no mês de seu aniversário.

Segunda variável:

Profissões com nove categorias

Astrônomo(AST)	Autor(AUT)	Físico(FIS)
Matemático(MAT)	Médico(MED)	Músico(MUS)
Naturalista(NAT)	Pensador(PEN)	Químico(QUI).

Na tabela 5.1 apresentamos os dados.

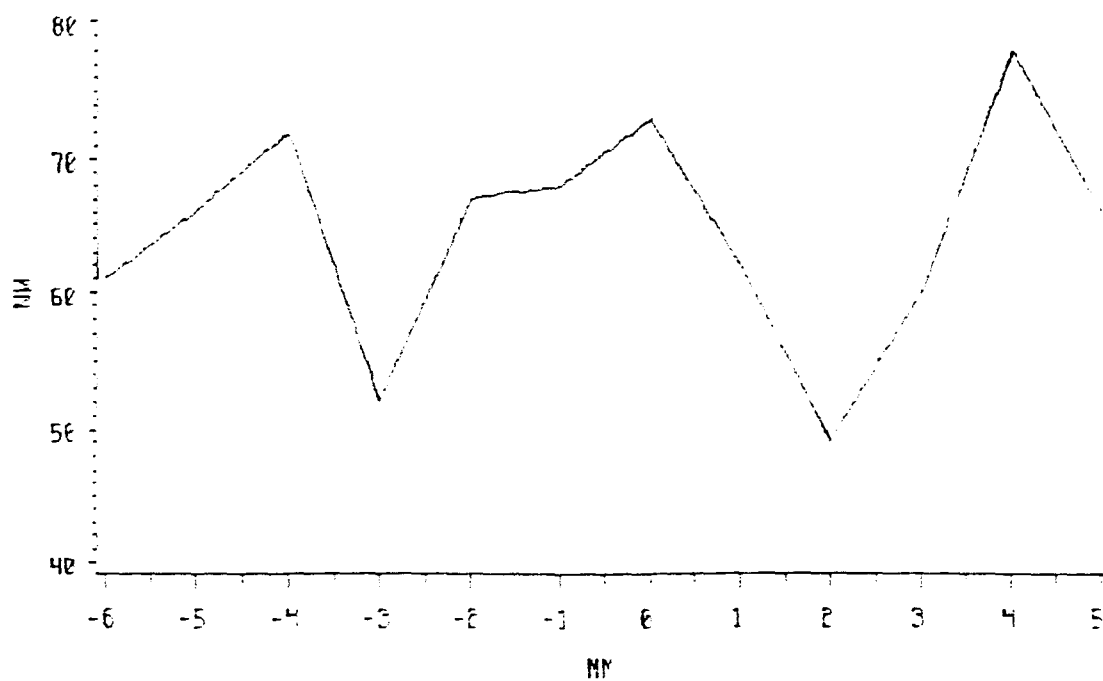
Observe nesta tabela que a porcentagem de morte relativa ao mês de aniversário aumenta a partir do segundo mês antes do aniversário, tendo uma maior porcentagem no mês do aniversário, diminuindo no primeiro e segundo mês subsequentes. Também observamos duas outras frequências elevadas no quarto mês antes do aniversário e o quarto mês depois. Para uma melhor visualização colocamos em um gráfico estes valores, na figura 5.1

TABELA 5.1

=====											
MORTE COM				PROFISSÕES							
=====											
RELACÃO AO	AST	AUT	FIS	MAT	MED	MUS	NAT	PEN	QUI	TOT.	PERC.
=====											
6MA	7	10	11	5	6	10	1	1	10	61	7,88%
5MA	6	11	17	1	4	5	2	6	14	66	8,53%
4MA	10	13	16	3	2	10	3	2	13	72	9,30%
3MA	5	7	9	2	4	11	2	2	10	52	6,72%
2MA	7	10	12	8	5	9	1	4	11	67	8,66%
1MA	11	6	13	5	7	9	1	2	14	68	8,79%
MMA	8	12	10	6	5	6	2	4	20	73	9,45%
1MD	3	7	14	6	8	5	1	3	15	62	8,01%
2MD	7	2	13	5	3	4	3	1	11	49	6,33%
3MD	9	11	4	3	5	10	1	4	13	60	7,75%
4MD	11	8	18	5	12	8	1	1	14	78	10,08%
5MD	6	11	17	5	5	4	0	0	18	66	8,53%
=====											
TOTAL	90	108	154	54	66	91	18	30	163	774	
=====											

Na seguinte seção, mostramos os resultados da aplicação da análise de correspondência aos dados da tabela 5.1.

figura 5.1



Número de mortes antes, durante e depois
do mês de aniversário

5.3. TABELAS, GRÁFICOS E ESTUDO DOS RESULTADOS.

De maneira similar aos capítulos III e IV, para o estudo completo dos dados através da análise de correspondência, precisamos construir as tabelas em que estejam as informações dos subespaços, e a qualidade de representação dos perfis nestes subespaços.

Na tabela 5.2, mostramos os valores singulares e a inércia explicada em cada subespaço.

Observando esta tabela, vemos que o primeiro eixo explica 29,77% da inércia total, o que é relativamente pouco. Devido a este resultado, faremos o estudo dos dados sobre os quatro primeiros eixos. Nas figuras 5.2 e 5.3, apresentamos a projeção dos perfis sobre os planos gerados pelos eixos 1 e 2, e eixos 3 e 4, respectivamente.

TABELA 5.2

VALORES SINGULARES GENERALIZADOS INÉRCIA E PROPORÇÃO
DE INÉRCIA EXPLICADA.

EIXO	VALORES SINGULARES	INÉRCIA	PROP. DE INÉRCIA EXPLICADA	
			ABSOLUTA(%)	ACUMULADA(%)
1	0,1780	0,0317	29,77	29,77
2	0,1501	0,0225	21,18	50,95
3	0,1335	0,0178	16,74	67,69
4	0,1076	0,0116	10,88	78,57
5	0,1005	0,0101	9,49	88,06
6	0,0848	0,0072	6,76	94,82
7	0,0648	0,0042	3,94	98,76
8	0,0365	0,0013	1,25	100,00

Para interpretar os pontos do gráfico, construímos a tabela 5.3 na qual apresentamos as informações dos perfis com respeito aos quatro primeiros eixos. Os valores desta tabela são multiplicados por 1000, as notações do encabeçamento das colunas são descritas no capítulo IV, seção 4.3.

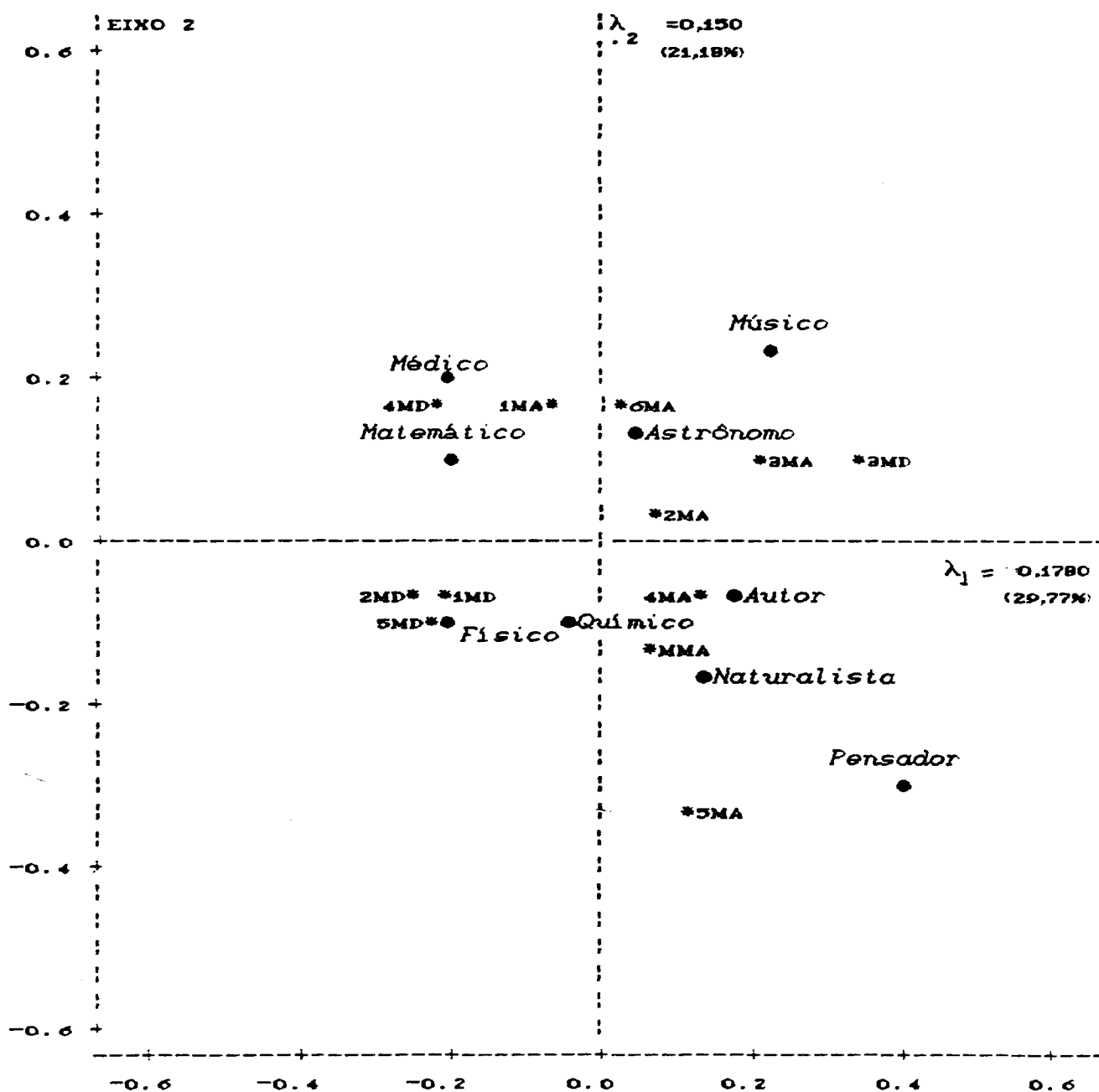


FIGURA 5.2. PROJEÇÃO DOS PERFIS DAS PROFISSÕES E MÊS DE MORTE RESPEITO AO MÊS DE ANIVERSÁRIO SOBRE O PLANO GERADO PELOS EIXOS 1 E 2. O SÍMBOLO ● REPRESENTA A PROFISSÃO E O SÍMBOLO * REPRESENTA O MÊS DE MORTE RESPEITO AO MÊS DE ANIVERSÁRIO.

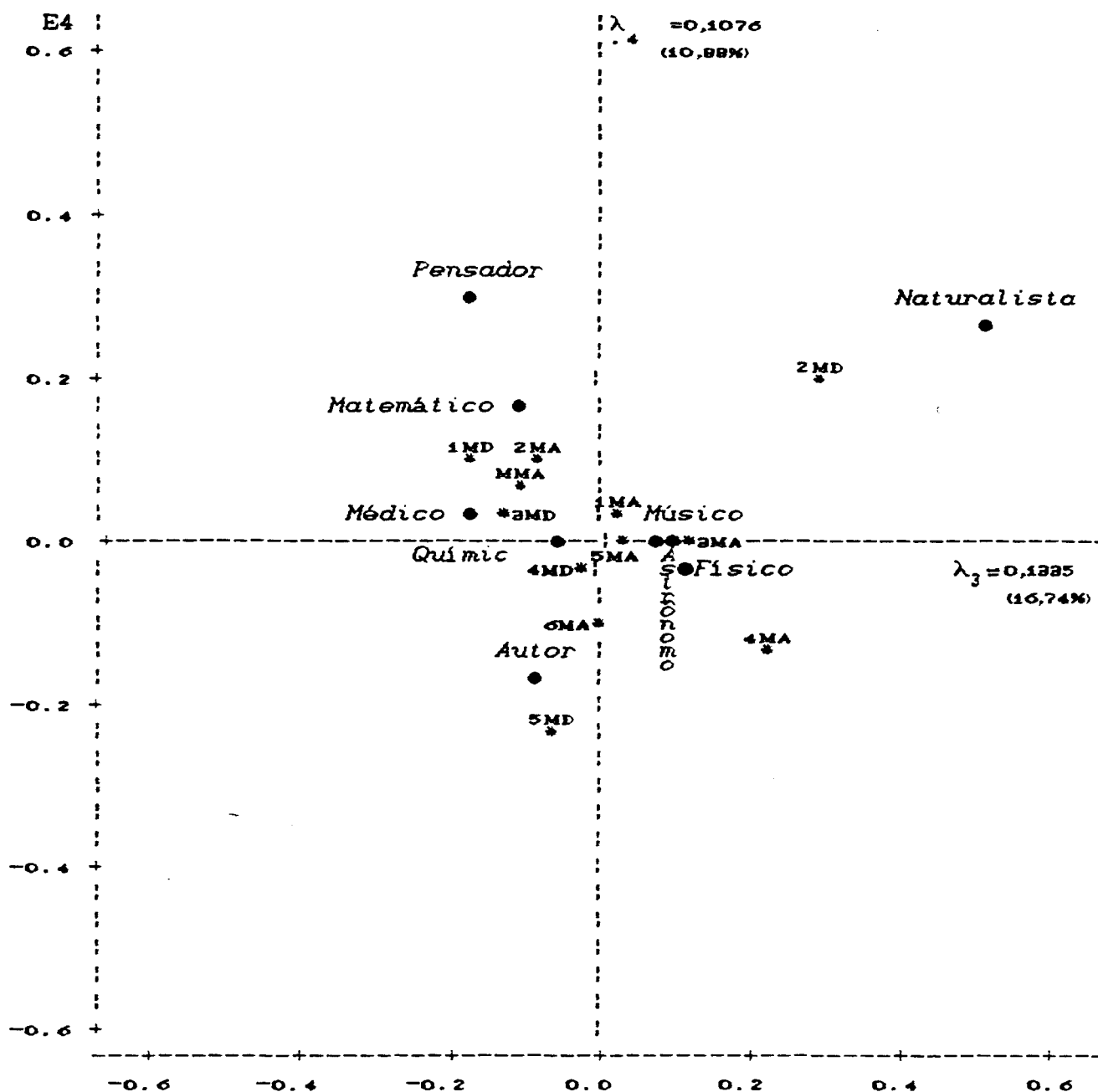


FIGURA 5.3. PROJEÇÃO DOS PERFIS DAS PROFISSÕES E MÊS DE MORTE RESPEITO AO MÊS DE ANIVERSÁRIO SOBRE O PLANO GERADO PELOS EIXOS 3 E 4. O SÍMBOLO ● REPRESENTA A PROFISSÃO E O SÍMBOLO * REPRESENTA O MÊS DE MORTE RESPEITO AO MÊS DE ANIVERSÁRIO.

TABELA 5.3
DECOMPOSIÇÃO DA INÉRCIA E COORDENADAS DOS PERFIS
PARA OS QUATRO PRIMEIROS EIXOS

=====

ESPAÇO DAS LINHAS

=====

CATEG	MASS	INER	F1	COR	CTR	F2	COR	CTR	F3	COR	CTR
6MA	79	40	33	20	3	182	617	116	-9	2	0
5MA	85	132	114	79	35	-336	683	426	20	2	2
4MA	93	83	143	216	60	-59	37	15	227	541	268
3MA	67	69	199	364	84	115	121	39	122	136	56
2MA	87	55	61	55	10	42	26	7	-94	130	43
1MA	88	40	-66	90	12	155	498	93	32	21	5
MMA	94	59	74	83	16	-119	213	60	-118	209	74
1MD	80	77	-191	357	92	-80	63	23	-179	314	144
2MD	63	120	-238	279	113	-50	13	7	293	424	305
3MD	78	116	339	721	282	113	79	44	-131	107	74
4MD	101	99	-216	449	149	165	262	122	-13	2	1
5MD	85	110	-231	398	144	-113	93	48	-77	44	28

=====

ESPAÇO DAS COLUNAS

=====

CATEG	MASS	INER	G1	COR	CTR	G2	COR	CTR	G3	COR	CTR
AST	116	76	43	27	7	136	265	96	118	197	90
AUT	140	107	187	429	154	-78	75	38	-89	98	63
FIS	199	137	-190	494	228	-111	169	110	112	172	141
MAT	70	107	-192	225	91	109	73	37	-118	85	55
MED	85	123	-199	260	107	199	259	150	-184	221	162
MUS	118	143	232	415	200	238	439	297	75	43	37
NAT	23	91	141	48	15	-160	62	27	504	612	332
PEN	39	155	401	380	197	-310	227	166	-189	84	78
QUI	211	61	-43	59	12	-93	282	81	-61	120	44

=====

RESULTADOS GERADOS PELO EIXO 4

=====

ESPAÇO DAS FILAS				ESPAÇO DAS COLUNAS			
CATEG.	F4	COR	CTR	CATEG.	G4	COR	CTR
6MA	-89	146	53	AST	7	1	1
5MA	5	00	00	AUT	-163	326	320
4MA	-121	155	118	FIS	-48	31	39
3MA	-6	00	00	MAT	169	174	172
2MA	104	161	81	MED	39	10	11
1MA	47	46	17	MUS	-11	1	1
MMA	60	55	30	NAT	274	181	151
1MD	94	87	61	PEN	302	215	305
2MD	212	222	246	QUI	-3	00	00
3MD	19	2	3				
4MD	-38	14	12				
5MD	-227	373	378				

=====

À partir desta tabela contruímos as relações entre os perfis, considerando a qualidade de representação dos mesmos pelos subespaços. Baseados nestes resultados fazemos o estudo.

**A) ESTUDO DO MÊS DE MORTE RELATIVO AO MÊS DE ANIVERSÁRIO,
SEGUNDO A PROFISSÃO.**

EIXO 1:

Coluna COR: 3MD = 721 com F1 = 330
4MD = 440 com F1 = -216
5MD = 388 com F1 = -231
1MD = 357 com F1 = -101
3MA = 364 com F1 = 100

EIXO 2:

Coluna COR: 6MA = 617 com F2 = 182
5MA = 683 com F2 = -336
1MA = 408 com F2 = 155

EIXO 3:

Coluna COR: 4MA = 541 com F3 = 277
1MD = 314 com F3 = -170
2MD = 424 com F3 = 203

EIXO 4:

Coluna COR: 2MD = 222 com F4 = 212
5MD = 373 com F4 = -227

No primeiro eixo, observamos a contraposição entre as categorias 3MD, 3MA com 4MD, 5MD, 1MD e uma proximidade entre as categorias 4MD e 5MD; o segundo eixo, mostra uma contraposição entre as categorias 5MA com 6MA, 1MA., além disso, uma proximidade entre as categorias 6MA e 1MA.; no terceiro eixo, observamos a contraposição entre as categorias 1MD com 4MA, 2MD, e proximidade entre 4MA e 2MD.; finalmente o quarto eixo mostra a contraposição entre as categorias 2MD e 5MD.

Das contraposições, concluímos que a data de morte dos personagens com respeito a seu aniversário, são diferentes segundo a sua profissão, e das proximidades, concluímos que a data de morte dos personagens com respeito a seu aniversário são similares, segundo a sua profissão.

PLANO 1 x 2 :

Dentro deste plano, vemos uma maior proximidade entre as categorias 1MD e 5MD com QLT 420 e 481, respectivamente, e 1MA e 6MA com QLT 588 e 637 respectivamente, do qual pode-se concluir que a data de morte com respeito ao aniversário nestas categorias segundo a profissão são similares. A categoria 5MA com QLT 762 afasta-se das demais categorias.

PLANO 3 x 4 :

Observamos o afastamento entre os pontos 2MD, 4MA, 5MD com QLT 646, 696, 416, respectivamente do qual concluímos que a porcentagem de morte nestes meses são diferentes.

B) ESTUDO DA RELAÇÃO ENTRE AS PROFISSÕES, SEGUNDO O MÊS DE MORTE COM RESPEITO AO MÊS DE ANIVERSÁRIO.

Similarmente ao A), primeiro analisamos a correlação de cada perfil com respeito a seu eixo e o valor da coordenada nos eixos.

EIXO 1:

Coluna COR: AUT = 429 com G1 = 187
 FIS = 494 com G1 = -190
 MUS = 415 com G1 = 232
 PEN = 380 com G1 = 401
 *MAT = 225 com G1 = -192
 *MED = 260 com G1 = -199

* Categorias que têm a maior correlação neste eixo com respeito aos demais eixos.

EIXO 2:

Coluna COR: AST = 265 com G2 = 136
MUS = 439 com G2 = 238
QUI = 282 com G2 = -93

EIXO 3:

Coluna COR: NAT = 612 com G3 = 504

EIXO 4:

Coluna COR: AUT = 326 com G4 = -163

No primeiro eixo, observamos a contraposição entre as profissões de Físico, Matemático, Médico com Autor, Músico, Pensador, e a proximidade dentro de cada grupo de profissões. O segundo eixo nos mostra uma contraposição das profissões de Músico, Astrônomo com Químico; no terceiro e quarto eixo, não se observa nenhum tipo de relação.

PLANO 1 x 2:

O plano mostra uma dispersão entre as profissões de Médico, Físico, Músico, Autor e Pensador com QLT 519, 663, 854, 504 e 607 respectivamente, do que podemos concluir que as data de morte nestas profissões com respeito a seu aniversário são diferentes.

PLANO 3 x 4:

Este plano dá uma maior informação sobre as profissões de Naturalista e autor com QLT 793, 424, respectivamente, e mostra o afastamento entre elas do que conclui-se que a data de morte dos profissionais nestas áreas são diferentes.

A seguir estudaremos a correspondência entre a profissão e mês de morte, com respeito ao mês de aniversário.

C) ESTUDO DA CORRESPONDÊNCIA ENTRE O MÊS DE MORTE COM RESPEITO AO MÊS DE ANIVERSÁRIO E PROFISSÃO.

O estudo se baseia nos estudos anteriores e nos gráficos 5.2 e 5.3.

PLANO 1 x 2:

O profissional Músico está perto dos meses 3MA, 3MD, e 6MA; o profissional Médico está associado com meses 4MD e 1MA; enquanto o profissional Físico está mais perto do mês 1MD e 5MD; Já o profissional Autor está mais perto do meses 3MA e 3MD. Finalmente o profissional Pensador aproxima-se do mês 5MA.

PLANO 3 x 4:

Mostra uma associação maior entre o profissional Autor e o mês 5MD, e o profissional Naturalista e o mês 2MD.

Fazendo um estudo conjunto dos planos, observamos que os profissionais Médico e Físico têm sua data de morte com respeito a seu aniversário adiada; o profissional Músico e Autor não tem uma tendência marcante. (morrem tanto antes como depois do mês de aniversário); os pensadores têm uma tendência de morrer antes da data do aniversário e o naturalistas depois da data de aniversário.

No que diz respeito aos profissionais Químico, Matemático e Astrônomo não podemos fazer nenhuma análise pois a QLT deles pelos planos não é elevada; mas pelo fato de existir uma dispersão entre eles podemos concluir que a data de morte dos profissionais nessas áreas são diferentes. A interpretação é ainda mais dificultosa quando relacionadas aos meses de morte e outras profissões.

5.4. CONCLUSÕES

Nos resultados obtidos observa-se que a data de morte com respeito ao aniversário dos profissionais das áreas sob estudo está associada a diferentes datas, não existindo uma tendência de concentração em determinado mês.

Embora os dois primeiros eixos acumulem uma forte inércia, a proximidade entre os valores das inércias nos conduz a pensar que não existe dependência entre o mês de morte e profissão.

CONCLUSÕES GERAIS

Pelos estudos realizados nas Tabelas de Contingência para duas variáveis, concluímos que o método da análise de correspondência tem algumas vantagens em relação aos métodos numéricos:

.No fato de mostrar graficamente as relações entre as categorias das variáveis (linhas e colunas) em forma simultânea num mesmo espaço.

.A síntetização global dos dados, cujos vetores de frequências envolvem valores observados em diferentes intervalos de tempo.

.Uso de variáveis ilustrativas (perfis suplementares) permitindo uma maior informação do relacionamento entre as variáveis originais.

.Uma grande versatilidade enquanto método exploratório de investigação multidimensional.

REFERÊNCIAS BIBLIOGRÁFICAS

- ASIMOV, I - (1980) - Enciplopédia Gênios da Humanidade, Vol I, II, e III, Bloch Editores S.A., Rio de Janeiro, Brasil.
- BENZÉCRI, J.P. - (1973) - L'Analyse des Données, Tome 2: L'Analyse des Correspondences, Dunod (2de Edition 1976).
- BERTIER, P & BOUROCHE, J.M. - (1975) - Analyse des Données Multidimensionnelles, França: Presses Universitaires de França.
- CAZES, P. - (1982) - Note sur les éléments supplémentaires em analyse das correspondences, Les cashiers de L'Analyse de Donnés, Vol VII, Nro.1, Pág.9-23.
- COMISSÃO DO IV CENTENARIO DA CIDADE DE SÃO PAULO - (1954) - Diccionario de Autores Paulistas, São Paulo.
- ECKART, C & YOUNG, G. - (1936) - The aproximation of one matrix by another of lower rank. Psychometrika, 1, 211-218.
- ESCOFIER, B. - (1978) - Analyse Factorielle et distances répondant au principe D'Equivalence Distributionnelle, Vol XXVI, No4, pag.29 - 36.
- EVERITT, B.S. - (1977) - The Analysis of Contingency Tables, New York: John Wiley & Sons.
- FIENBERG, S.E. - (1977) The Analysis of Cross-Classified Categorical Data.
- GREENACRE, M.J. - (1984) - Theory and Applications of Correspondence Analysis, London: Academic Press.
- HATZIPARADISSIS, A. - (1987) - Les orientations de la Recherche Dans L'Espace de la communante Economique européenne, Les cashiers de L'Analyse des Données , Vol XII PP 271-289.

- HIRSCHFELD, H. O. - (1935) - A Connection between correlation and contingency; Cambridge Philosophical Soc. Proc. (Math. Proc) 31, 520-524.
- JOLLIFE, J. F. - (1986) - Principal Component Analysis, New York: Springer-Verlag.
- KAUFMANN, H. - (1943) - The story of one hundred great composers, New York.
- LEBART, L., MORINEU, A. & FÉNELON, J. P. - (1982) Traitement des données statistiques méthodes et programmes, 2da Edition, Bordas, Paris.
- MARDIA, K. V., KENT, J. T. & BIBBY, J. M. - (1979) - Multivariate Analysis, London : Academic Press.
- NAG (Numerical Algorithms Group) - (1981)
- PEARSON, K. - (1900) - On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos., Mag. 50, Nro 5, Pág. 157-175.
- PEARSON, K. - (1901) - On lines and planes of closest fit to a system of points in space, Philosophical Magazine and Journal of Science, Series 6, 2, 559-572.
- PHILLIPS, D. P. - (1970) - Deathday and Birthday: An Unexpected Connection; Guide to the Unknown, USA, edited by TANUR, J. M. & MOSTELLER, F.
- SOUZA, J. C. & OUTROS - (1973) Os Pensadores, vol I, II, Abril Cultural, São Paulo.
- TEKAA, F., SAINT MARTIN, J. de., SANSONETTI, P., VUSLLIER, F. & CLAVEIRIE, J. M. - (1988) - Intérêt de l'antigénémie dans la définition des stades de l'infection par le virus VIH., Les cahiers de L'Analyse des Données, Vol XIII, 1988-pág. 407 - 424.

YULE, G. U. - (1903) - Notes on the theory of association of attributes in statistics. *Biometrika* 2, pág. 121-134.

RICHARSON, M. & KUDER, G. F. - (1933) - Making a Rating Scale that Measures; *Personnel J.* 12, 36-40..

APÊNDICE

PROCESSO COMPUTACIONAL

```

C*****
C      PROGRAMA PRINCIPAL PARA A ANÁLISE DE UMA TABELA DE CONTIN
C      GÊNCIA ATRAVÉS DA ANÁLISE DE CORRESPONDÊNCIA.
C*****
C      NL=NUMERO DE LINHAS
C      NC=NUMERO DE COLUNAS
C      NK=NUMERO DE VALORES SINGULARES
C*****
C      DIMENSION AC(30,20),XNC(30,20)
C      DIMENSION XPC(30,20),RC(30),CC(20),PRCT(30,20)
C      DIMENSION DPCTDC(32,15),PT(15,15),QC(32,15),SV(15),WORK(40)
C      DIMENSION BTC(20,20),FC(30,20),GC(20,20),CTRFC(30,20)
C      DIMENSION DTF(30,20),DTCC(20,20),PFC(30,20),PCC(20,30)
C      DIMENSION SINERFC(30),TINERFC(30),SINERCC(20),TINERCC(20)
C      DIMENSION QDFC(30),QDCC(30),COS2FC(30,20),COS2CC(20,20)
C      DIMENSION CTRCC(20,20),DFC(30),DCC(20),RDFC(30),RDC(20)
C      DOUBLE PRECISION A,XN,XP,R,C,PRCT,DPCTD,PT,Q,SV,WORK
C      DOUBLE PRECISION BT,F,G,DTF,DTCC,PF,PC,DF,DC,RDF,RDC
C      DOUBLE PRECISION SINERF,SINERC,TINERF,TINERC,QDF,QDC,COS2F
C      DOUBLE PRECISION COS2C,CTRF,CTRC
C      INTEGER NL,NC,NROE,I,NRA,NRQ,NRPT,LWORK,J,M,N,IFAIL
C      INTEGER MINMN
C      NL=8
C      NC=5
C      NRA=32
C      NRQ=32
C      NRPT=15
C      LWORK=40
C      IFAIL=0
C
C      LEITURA DOS ELEMENTOS DA TABELA DE CONTINGÊNCIA
C
C      DO 20 I=1,NL
C      READ(10,*) (XNC(I,J),J=1,NC)
20  CONTINUE
C
C      SOMA DOS ELEMENTOS DA AMOSTRA
C
C      NROE=0
C      DO I=1,NL
C      DO J=1,NC
C      NROE=NROE+XNC(I,J)
C      END DO
C      END DO

```

C
C
C
MATRIZ DE CORRESPONDÊNCIA

DO I=1,NL
DO J=1,NC
XPCI,J)=XNCI,J)*1/NROE
TYPE *,XPCI,J)
END DO
END DO

C
C
C
CÁLCULO DAS MARGINAIS DAS LINHAS E COLUNAS DA TABELA

DO I=1,NL
RCI)=0.0
DO J=1,NC
RCI)=RCI)+XPCI,J)
END DO
END DO
DO J=1,NC
CCJ)=0.0
DO I=1,NL
CCJ)=CCJ)+XPCI,J)
END DO
END DO

C
C
C
C
C
CÁLCULO DAS MATRICES QUE CONTEM OS PERFIS DAS:

C
C
C
LINHAS

DO I=1,NL
DO J=1,NC
PFI,J)=(1/RCI))*XPCI,J)
END DO
END DO

C
C
C
COLUNAS

DO I=1,NC
DO J=1,NL
PCI,J)=(1/CCJ))*XPCJ,I)
END DO
END DO

C
C
C
C
C
C
CÁLCULO DA DISTANCIA QUI QUADRADO ENTRE OS PERFIS E SEU CENTRÓIDE:

C
C
C
LINHAS

DO I=1,NL
DFCI)=0.0
DO J=1,NC
DFCI)=DFCI)+(PFI,J)**2)*(1/CCJ))
END DO
QDFCI)=DFCI)-1
RDFCI)=SQRT(DFCI)-1)
END DO

C

C COLUNAS

C
C DO I=1,NC
C DCC(I)=0.0
C DO J=1,NL
C DCC(I)=DCC(I)+(PC(I,J)**2)*(1/R(J))
C END DO
C QDCC(I)=DCC(I)-1
C RDCC(I)=SQRT(DCC(I)-1)
C END DO
C NK=NC-1

C
C CONSTRUÇÃO DA MATRIZ PARA CALCULAR OS VALORES E VETORES
C SINGULARES

C
C DO I=1,NL
C DO J=1,NC
C PRCT(I,J)=XP(I,J)-R(I)*CC(J)
C END DO
C END DO
C DO I=1,NL
C DO J=1,NC
C DPCTD(I,J)=(1/SQRT(R(I)))*PRCT(I,J)*(1/SQRT(CC(J)))
C END DO
C END DO

C
C CHAMADA DA SUBROTINA QUE CÁLCULA OS VALORES E VETORES SIN -
C GULARES

C
C CALL FO2WCFC(NL,NC,NC,DPCTD,NRA,Q,NRQ,SV,PT,NRPT,WORK,LWORK,
C * IFAIL)
C IF (IFAIL.EQ.0) THEN
C TYPE *, 'VALORES SINGULARES'
C DO I=1,NC
C TYPE *,SV(I)
C END DO
C ELSE
C TYPE *, 'IFAIL=',IFAIL
C TYPE *, 'ERROR NO PROGRAMA'
C END IF
C NK=NC-1

C
C CÁLCULO DA INÉRCIA DECOMPOSTA NOS SUBESPACOS GERADOS PELOS
C VETORES SINGULARES.

C
C SITFC=0.0
C DO I=1,NK
C SITFC=SITFC+(SV(I)**2)
C END DO

C
C CÁLCULO DA BASE PARA OS PERFIS DAS COLUNAS

C
C DO I=1,NL
C DO J=1,NC
C AC(I,J)=SQRT(R(I))*Q(I,J)
C END DO
C END DO


```

C      CÁLCULO DA CONTRIBUIÇÃO ABSOLUTA DOS PERFIS PARA A INÉRCIA
C      DE CADA SUBESPAÇO.
C
C      PERFIS DAS LINHAS
C
      DO I=1,NL
        DO J=1,NK
          CTRF(I,J)=DTF(I,J)/(SV(J)**2)
        END DO
      END DO
C
C      PERFIS DAS COLUNAS
C
      DO I=1,NC
        DO J=1,NK
          CTRC(I,J)=DTC(I,J)/(SV(J)**2)
        END DO
      END DO
C
C      CÁLCULO DA CONTRIBUIÇÃO RELATIVA DOS SUBESPACOS PARA OS
C      PERFIS:
C      LINHAS
C
      DO I=1,NL
        DO J=1,NK
          COS2F(I,J)=(CF(I,J)**2)/QDF(I)
        END DO
      END DO
C
C      COLUNAS
C
      DO I=1,NC
        DO J=1,NK
          COS2C(I,J)=(GC(I,J)**2)/QDC(I)
        END DO
      END DO
C
C      IMPRIMINDO OS RESULTADOS
C
      WRITE(4,200)(NROE)
200    FORMAT(/,5X,'NÚMERO TOTAL DE ELEMENTOS',5X,I15)
      WRITE(4,210)
210    FORMAT(1X,26H MATRIZ DE CORRESPONDÊNCIA)
      WRITE(4,240)((XPC(I,J),J=1,NC),I=1,NL)
      WRITE(4,230)
230    FORMAT(21H MARGINAIS DAS LINHAS)
      DO I=1,NL
        WRITE(4,600)(R(I))
      END DO
      WRITE(4,250)
250    FORMAT(23H MARGINAIS DAS COLUNAS)
      DO I=1,NC
        WRITE(4,600)(C(I))
      END DO

```

```

WRITE(4,264)
264  FORMAT(/,5X,19H VALORES SINGULARES)
    DO I=1,NK
    WRITE(4,600)(SV(I))
    END DO
    WRITE(4,268)
268  FORMAT(/,5X,21H BASE PARA OS PERFIS DAS COLUNAS)
    WRITE(4,240)((AC(I,J),J=1,NK),I=1,NL)
    WRITE(4,270)
270  FORMAT(/,5X,19H BASE PARA OS PERFIS DAS LINHAS)
    WRITE(4,240)((BTC(I,J),J=1,NK),I=1,NK)
    WRITE(4,280)
280  FORMAT(/,5X,33H COORDENADAS DOS PERFIS DAS LINHAS)
    WRITE(4,295)((FC(I,J),J=1,NK),I=1,NL)
    WRITE(50,295)((FC(I,J),J=1,NK),I=1,NL)
    WRITE(4,290)
290  FORMAT(/,5X,35H COORDENADAS DOS PERFIS DAS COLUNAS)
    WRITE(4,295)((GC(I,J),J=1,NK),I=1,NK)
    WRITE(50,295)((GC(I,J),J=1,NK),I=1,NK)
295  FORMAT(1X/4F9.4)
    WRITE(4,340)
340  FORMAT(/,5X,35H DECOMPOSIÇÃO DE INÉRCIA PARA AS LINHAS)
    WRITE(4,345)((DTF(I,J),J=1,NK),I=1,NL)
345  FORMAT(1X/4F9.4)
    WRITE(4,350)
350  FORMAT(/,5X,37H DECOMPOSIÇÃO DE INÉRCIA PARA AS COLUNAS)
    WRITE(4,355)((DTCC(I,J),J=1,NK),I=1,NK)
355  FORMAT(1X/4F9.5)
    WRITE(4,357)
357  FORMAT(/,5X,42H CONTRIBUIÇÃO ABSOLUTA DOS PERFIS DAS LINHAS)
    WRITE(4,345)((CTRF(I,J),J=1,NK),I=1,NL)
    WRITE(4,362)
362  FORMAT(/,5X,44H CONTRIBUIÇÃO ABSOLUTA DOS PERFIS DAS COLUNAS)
    WRITE(4,355)((CTRC(I,J),J=1,NK),I=1,NK)
    WRITE(4,366)
366  FORMAT(/,5X,44H CONTRIBUIÇÃO RELATIVA PARA OS PERFIS DAS
* LINHAS)
    WRITE(4,345)((COS2F(I,J),J=1,NK),I=1,NL)
    WRITE(4,368)
368  FORMAT(/,5X,46H CONTRIBUIÇÃO RELATIVA PARA OS PERFIS DAS
* COLUNAS)
    WRITE(4,355)((COS2C(I,J),J=1,NK),I=1,NK)
    WRITE(4,370)
370  FORMAT(/,5X,34H INÉRCIA TOTAL DOS PERFIS DAS LINHAS)
    DO I=1,NL
    WRITE(4,600)(TINERF(I))
    END DO
    WRITE(4,372)
372  FORMAT(/,5X,36H INÉRCIA TOTAL DOS PERFIS DAS COLUNAS)
    DO I=1,NK
    WRITE(4,600)(TINERC(I))
    END DO
    STOP
240  FORMAT(1X/5F9.5)
600  FORMAT(F9.5)
    END.

```