

Introdução à Análise Exploratória de Dados Multivariados

Publicações Matemáticas

**Introdução à Análise Exploratória
de Dados Multivariados**

Pedro J. Fernandez
Provokers do Brasil

Victor Yohai
Universidad de Buenos Aires



Copyright © 2014 by Pedro J. Fernandez e Victor Yohai

Impresso no Brasil / Printed in Brazil

Capa: Noni Geiger / Sérgio R. Vaz

Publicações Matemáticas

- Introdução à Topologia Diferencial – Elon Lages Lima
- Criptografia, Números Primos e Algoritmos – Manoel Lemos
- Introdução à Economia Dinâmica e Mercados Incompletos – Aloísio Araújo
- Conjuntos de Cantor, Dinâmica e Aritmética – Carlos Gustavo Moreira
- Introdução à Economia Matemática – Aloísio Araújo
- The Index Formula for Dirac Operators: an Introduction – Levi Lopes de Lima
- Introduction to Symplectic and Hamiltonian Geometry – Ana Cannas da Silva
- Primos de Mersenne (e outros primos muito grandes) – Carlos Gustavo T. A. Moreira e Nicolau Saldanha
- The Contact Process on Graphs – Márcia Salzano
- Introduction to Toric Varieties – Jean-Paul Brasselet
- Birational Geometry of Foliations – Marco Brunella
- Teoria dos Corpos – Otto Endler
- Introdução à Dinâmica de Aplicações do Tipo Twist – Clodoaldo G. Ragazzo, Mário J. Dias Carneiro e Salvador Addas Zanata
- Elementos de Estatística Computacional usando Plataformas de Software Livre/Gratuito – Alejandro C. Frery e Francisco Cribari-Neto
- Uma Introdução a Soluções de Viscosidade para Equações de Hamilton-Jacobi – Helena J. Nussenzveig Lopes, Milton C. Lopes Filho
- Elements of Analytic Hypoellipticity – Nicholas Hanges
- Variedades Diferenciáveis – Elon Lages Lima
- A Student's Guide to Symplectic Spaces, Grassmannians and Maslov Index – Paolo Piccione e Daniel Victor Tausk
- Métodos Topológicos en el Análisis no Lineal – Pablo Amster
- Tópicos em Combinatória Contemporânea – Carlos Gustavo Moreira e Yoshiharu Kohayakawa
- O Teorema de Poncelet – Marcos Sebastiani
- Aspectos Ergódicos da Teoria dos Números – Alexander Arbieto, Carlos Matheus e C. G. Moreira
- A Survey on Hyperbolicity of Projective Hypersurfaces – Simone Diverio e Erwan Rousseau
- Algebraic Stacks and Moduli of Vector Bundles – Frank Neumann
- O Teorema de Sard e suas Aplicações – Edson Durão Júdice
- Holonomy Groups in Riemannian Geometry – Andrew Clark e Bianca Santoro
- Linear ODEs: an Algebraic Perspective - Letterio Gatto

IMPA - ddic@impa.br

<http://www.impa.br>

ISBN: 978-85-244-0393-4

ÍNDICE

Capítulo 1 – Revisão de Álgebra Linear	5
Capítulo 2 – Análise de Componentes Principais	24
Capítulo 3 – Análise de Correspondência	60
Capítulo 4 – Escalonamento Multivariado	109
Referências –	139

INTRODUÇÃO (1980)

Nesta monografia estão incluídas algumas técnicas de Análise de Dados especialmente úteis para a análise de grandes conjuntos de dados. Estas técnicas são de natureza exploratória e não confirmatória. Todas elas são basicamente técnicas dependentes de um uso adequado de resultados de Álgebra Linear. No Capítulo 1 apresentamos um resumo dos principais resultados de Álgebra Linear que utilizaremos. O Capítulo 2 trata de componentes principais. A versão que apresentamos é especialmente adequada para ser aplicada posteriormente nos Capítulos 3 e 4 nos quais as técnicas de Análise de Correspondências e Escalamento Multivariado são apresentadas. Exemplos de aplicação são incluídos em todos os capítulos.

Rio de Janeiro, junho/1980

Introdução a Edição Digital

Esta versão tem como objetivo fazer disponível a versão original da Monografia sobre algumas Técnicas de Análise de Dados cuja edição original, publicada em 1980 por ocasião do Colóquio de Matemática, está esgotada .

As partes puramente matemáticas e estatísticas foram preservadas de forma idêntica a versão original. Naturalmente com o passar dos anos o software disponível para analisar os dados tem tido um desenvolvimento surpreendente.

Para todos os exemplos temos usado R Statistical Software. R é simultaneamente uma linguagem de programação e um poderoso e importante conjunto de software “open source”.

R pode ser baixado e instalado visitando o site <http://www.r-project.org>. Temos trocado os exemplos originais por outros mais relevantes aos problemas atuais e analisados eles usando diferentes pacotes disponíveis em R.

Os dados fazem parte das bases de dados de diferentes pacotes. Isto facilita a eventual reprodução dos resultados e evita o uso de bases de dados que não podem ser disponibilizadas devido a sua confidencialidade.

Rio de Janeiro, Janeiro 2014

Capítulo 1

Revisão de Álgebra Linear

O presente capítulo tem como objetivo introduzir a notação básica e enunciar os principais resultados de álgebra linear a serem utilizados nos próximos capítulos. Só serão incluídas as demonstrações daqueles resultados que são menos conhecidos ou que pela sua frequente aparição nos capítulos seguintes mereçam um destaque ou ênfase especial.

O leitor deve estar familiarizado com a noção de espaço vetorial, subespaço, dependência e independência linear de vetores, bases, transformações lineares e sua representação matricial fixadas as bases, mudança de base, sistemas lineares, autovalores e autovetores. Strang (1988), Lima (1998) podem ser consultados com proveito para esclarecimento ou aprofundamento desses pontos.

Neste capítulo introduzimos uma versão essencialmente matricial dos mesmos.

Passamos agora à apresentação concisa dos principais temas que são objeto deste capítulo.

Se A indica uma matriz, A' indicará a sua transposta e caso A seja quadrada, $|A|$ indicará o determinante de A . Se $|A| \neq 0$, A^{-1} indicará a matriz inversa da matriz A . Se A tem elementos a_{ij} usaremos a notação $A = (a_{ij})$ ou também $A(i, j) = a_{i,j}$ para indicar que o elemento (i, j) de A é a_{ij} .

Se $x \in \mathbb{R}^m$ tem componentes $\{x_i\}_{i=1,\dots,m}$ indicaremos com $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$ a *norma* de x (ou comprimento de x).

Sejam A e B matrizes particionadas em blocos

$$A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hdashline & \\ A_{21} & A_{22} \end{array} \right) \quad B = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hdashline & \\ B_{21} & B_{22} \end{array} \right)$$

onde A_{ij} e B_{ij} representam matrizes para todo i e j .

As afirmações (ii) e (iii) da Proposição 1 abaixo são válidas sempre que as operações indicadas sejam compatíveis com as dimensões das matrizes.

Proposição 1.

$$(i) \quad A' = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hdashline & \\ A_{21} & A_{22} \end{array} \right)' = \left(\begin{array}{c|c} A'_{11} & A'_{21} \\ \hdashline & \\ A'_{12} & A'_{22} \end{array} \right)$$

$$(ii) \quad AB = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hdashline & \\ A_{21} & A_{22} \end{array} \right) \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hdashline & \\ B_{21} & B_{22} \end{array} \right) =$$

$$= \left(\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hdashline & \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array} \right)$$

$$(iii) \quad A + B = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hdashline & \\ A_{21} & A_{22} \end{array} \right) + \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hdashline & \\ B_{21} & B_{22} \end{array} \right) \\ = \left(\begin{array}{c|c} A_{11} + B_{11} & A_{12} + B_{12} \\ \hdashline & \\ A_{21} + B_{21} & A_{22} + B_{22} \end{array} \right)$$

Seja A uma matriz quadrada tal que

$$A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hdashline & \\ A_{21} & A_{22} \end{array} \right)$$

com A_{11} e A_{22} matrizes quadradas.

Proposição 2.

$$(i) \quad \text{Se } A = \left(\begin{array}{c|c} A_{11} & 0 \\ \hdashline & \\ A_{21} & A_{22} \end{array} \right) \text{ então } |A| = |A_{11}| |A_{22}|$$

$$(ii) \quad \text{Se } |A_{22}| \neq 0, \text{ então } |A| = |A_{22}| |A_{11} - A_{12} A_{22}^{-1} A_{21}|$$

$$(iii) \quad \text{Se } |A_{11}| \neq 0, \text{ então } |A| = |A_{11}| |A_{22} - A_{21} A_{11}^{-1} A_{12}|$$

Demonstração:

- (i) Procedemos por indução na ordem de A . Se A é 2×2 , $A = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix}$, temos que $|A| = a_{11} a_{22}$, que é o resultado procurado para ordem 2.

Suponhamos que o resultado é válido para matrizes de ordem menor ou igual a $n - 1$. Seja A $n \times n$. Temos pelo desenvolvimento de Laplace pela primeira linha que

$$|A| = \sum_{j=1}^n a_{1j}(-1)^{1+j} \tilde{A}_{1j} = \sum_{j=1}^p a_{1j}(-1)^{1+j} \tilde{A}_{1j}$$

onde p é a ordem de A_{11} e \tilde{A}_{1j} é o menor de A correspondente ao elemento a_{1j} . Agora \tilde{A}_{1j} é um determinante $(n - 1) \times (n - 1)$. Por indução temos

$$\tilde{A}_{1j} = (\text{menor de } a_{1j} \text{ em } A_{11}) \cdot |A_{22}|.$$

Portanto

$$\begin{aligned} |A| &= |A_{22}| \sum_{j=1}^p a_{1j}(-1)^{1+j} (\text{menor de } a_{1j} \text{ em } A_{11}) = \\ &= |A_{22}| |A_{11}|. \end{aligned}$$

A soma entre parênteses é igual a $|A_{11}|$ pelo desenvolvimento de Laplace.

$$\begin{aligned} \text{(ii)} \quad \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} &= \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} I & 0 \\ -A_{22}^{-1} A_{21} & I \end{vmatrix} = \\ &= \left| \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{22}^{-1} A_{21} & I \end{pmatrix} \right| = \\ &= |A_{22}| |A_{11} - A_{22}^{-1} A_{22}^{-1} A_{21}|. \end{aligned}$$

A primeira e quarta igualdades utilizam a parte (i) da Proposição 2.

(iii) O Argumento é igual ao desenvolvido para a parte (ii).

Lembremos que o traço de uma matriz quadrada $A = (a_{ij})$ é definido como a soma dos elementos contidos na diagonal de A . Ou seja

$$\text{Tr}(A) = \sum_{i=1}^n a_{ii}.$$

Seja A $n \times p$, e B $p \times n$ duas matrizes. O traço tem a seguinte propriedade de demonstração imediata.

Proposição 3.

$$\text{Tr}(AB) = \text{Tr}(BA).$$

Mais geralmente

$$\text{Tr}(A_1 A_2 \dots A_n) = \text{Tr}(A_n A_1 A_2 \dots A_{n-1}) = \dots = \text{Tr}(A_2 A_3 \dots A_n A_1).$$

Duas situações de aplicação frequente desta propriedade são as seguintes: Seja A ($n \times n$) e x um vetor ($n \times 1$). Então

$$\begin{aligned} \text{Tr}(xx') &= \text{Tr}(x'x) = x'x \\ \text{Tr}(Axx') &= \text{Tr}(x'A x) = x'A x \end{aligned}$$

Indicaremos com \mathbb{R}^m o espaço euclidiano m -dimensional. Se A é uma matriz $m \times m$, A induz uma transformação linear em \mathbb{R}^m da seguinte forma natural

$$T(x) = Ax, \quad x \in \mathbb{R}^m \quad x \text{ } m \times 1$$

Se $e_i = (0, \dots, 1, \dots, 0)'$ indica o i -ésimo vetor da base canônica de \mathbb{R}^m (o 1 está na posição i) temos que $Te_i = a_i$ onde a_i é a coluna i de A e $A = (a_1 : \dots : a_m)$. (Esta notação que será usada frequentemente, descreve A como uma matriz particionada em m blocos de ordem $m \times 1$ correspondentes as m colunas de A). Como $Te_i = a_i$, a representação de T na base canônica de \mathbb{R}^m é precisamente A .

Definição 1. Se existe λ real ($\lambda \in \mathbb{R}^1$) e $x \neq 0$, $x \in \mathbb{R}^m$ tal que $Ax = \lambda x$, então λ é chamado *autovalor* de A e x *autovetor* de A associado a λ . Note-se que $Ax = \lambda x$, é equivalente a $(A - \lambda I)x = 0$. Como $x \neq 0$, para que essa equação tenha uma solução não trivial devemos ter $|A - \lambda I| = 0$. $|A - \lambda I|$ é um polinômio em λ de grau m . Portanto os autovalores são raízes da equação de grau m $|A - \lambda I| = 0$. Se a matriz A é simétrica, resulta que os autovalores são reais. Temos ainda para matrizes simétricas o seguinte resultado fundamental.

Teorema 1 (Teorema Espectral). *Seja A uma matriz simétrica $m \times m$.*

Existe uma matriz ortogonal U $m \times m$ tal que

$$U'AU = \Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_m \end{pmatrix}$$

onde Λ é uma matriz diagonal. Se U_i indica a coluna i da matriz U temos que $AU_i = \lambda_i U_i$. Ou seja U_i é autovetor de A correspondente ao autovalor λ_i .

Note-se que por ser U ortogonal temos que $\|U_i\| = 1$, $i = 1, \dots, m$ e mais geralmene $U_i'U_j = \delta_{ij}$ onde $\delta_{ij} = 1$ se $i = j$ e 0 se $i \neq j$.

Trocando, caso seja necessário, as colunas de U sempre podemos supor que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$.

Os valores λ que aparecem na matriz diagonal Λ , que são os autovalores de A e portanto da equação $|A - \lambda I| = 0$ aparecem tantas vezes como é indicado pela sua multiplicidade.

Note-se também que $U'AU$ é a matriz correspondente a transformação linear determinada por A , na base ortonormal formada pelos autovetores de A . A tese do Teorema 1 é que nessa base a transformação é simples de descrever, consistindo em uma homotetia ao longo de cada novo eixo.

As identidades

$$U'AU = \Lambda$$

$$U'U = I$$

podem ser escritas, utilizando os resultados de produto de matrizes em blocos, da seguinte forma

$$A = U\Lambda U' = (U_1 \vdots \dots \vdots U_m) \Lambda \begin{pmatrix} U_1' \\ \vdots \\ U_m' \end{pmatrix} = \sum_{i=1}^m \lambda_i U_i U_i'$$

$$I = UU' = (U_1 \vdots \dots \vdots U_m) \begin{pmatrix} U_1' \\ \vdots \\ U_m' \end{pmatrix} = \sum_{i=1}^m U_i U_i'$$

Como $\lambda_i = U_i' A U_i$, se A fosse definida não negativa (resp. definida positiva) teríamos $\lambda_i \geq 0$ (resp. $\lambda_i > 0$).

Como o posto de uma matriz (máximo número de linhas ou colunas linearmente independentes) permanece inalterado com produtos com matrizes não singulares temos que o posto de A é igual ao posto de Λ que por sua vez é igual ao número de autovalores diferentes de 0.

A seguinte proposição é uma consequência útil do Teorema Espectral.

Proposição 4. *Seja A simétrica definida não negativa. Então*

(i) *Existe B tal que $A = B'B$.*

(ii) *Existe C tal que $A = C^2$.*

Se A é definida positiva, então C é não singular. C é chamada raiz quadrada de A e é denotada por $A^{1/2}$.

Demonstração. (i) Seja $A = U\Lambda U'$. Indicaremos com $\Lambda^{1/2}$ a matriz $\text{diag}(\lambda_1^{1/2}, \dots, \lambda_m^{1/2})$ ou seja a matriz diagonal cujos elementos diagonais são as raízes quadradas dos elementos de Λ . Temos então $A = U\Lambda^{1/2} \Lambda^{1/2} U'$. Seja $B = \Lambda^{1/2} U'$. Então $A = B'B$. Se A é definida positiva todos os λ_i são positivos e portanto Λ é não singular o que implica B não singular.

(ii) Temos que $A = U\Lambda^{1/2} \Lambda^{1/2} U' = U\Lambda^{1/2} U' U \Lambda^{1/2} U'$.

Chamando $C = U\Lambda^{1/2} U'$ temos $A = C^2$.

Como na parte (i), se A é definida positiva Λ e portanto $\Lambda^{1/2}$ são não singulares, o que implica que C é também não singular.

Teorema 2 (Teorema Espectral Generalizado). *Sejam A e B duas matrizes simétricas $m \times m$ com B definida positiva. Existe uma matriz U tal que*

$$U' A U = \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$$

$$U' B U = I$$

Os valores $\lambda_1, \dots, \lambda_m$ são raízes da equação $|A - \lambda B| = 0$. Se U_i indica a coluna i de U , U_i é autovetor de $B^{-1}A$ correspondente ao autovalor

λ_i ou seja $(B^{-1}A)U = U\Lambda$. Também os λ_i , $1 \leq i \leq m$ são autovalores da matriz AB^{-1} correspondentes aos autovetores $V_i = AU_i$.

Demonstração. Como B é definida positiva existe uma matriz não singular T tal que $B = T'T$. Temos que

$$|A - \lambda B| = |A - \lambda T'T| = |T'| |T'^{-1}AT^{-1} - \lambda I| |T| = |T'| |C - \lambda I| |T|$$

onde $C = T'^{-1}AT^{-1}$.

Portanto a equação $|A - \lambda B| = 0$ é equivalente a $|C - \lambda I| = 0$. Como C é uma matriz simétrica existe pelo Teorema Espectral uma matriz ortogonal P tal que

$$P'CP = \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$$

$$P'P = I$$

Os valores λ são as raízes da equação $|C - \lambda I| = 0$ ou equivalentemente $|A - \lambda B| = 0$. Substituindo-se C por $T'^{-1}AT^{-1}$ temos

$$P'T'^{-1}AT^{-1}P = \Lambda$$

$$P'P = I$$

Seja $U = T^{-1}P$. Temos então $TU = P$ e as equações acima transformam-se em

$$U'AU = \Lambda$$

$$U'T'TU = U'BU = I$$

Estas identidades podem ser escritas da forma

$$AU = U'^{-1}A$$

$$BU = U'^{-1}$$

Portanto $AU = BUA$ e também (multiplicando por B^{-1}) $(B^{-1}A)U = U\Lambda$. Se U_i indica a coluna i de U temos que U_i é autovetor de $B^{-1}A$ correspondente ao autovalor λ_i . Também multiplicando a última expressão por A temos $(AB^{-1})(AU) = (AU)\Lambda$. Portanto AU_i é autovetor de AB^{-1} correspondente ao autovalor λ_i .

Note-se que também podemos escrever

$$\begin{aligned} A &= U'^{-1} \Lambda U^{-1} = (U^{-1})' \Lambda U^{-1} \\ B &= U'^{-1} U^{-1} = (U^{-1})' U^{-1} \end{aligned}$$

Portanto se S_i indica a coluna i de $(U^{-1})'$ temos da mesma forma que para o Teorema Espectral

$$\begin{aligned} A &= \sum_{i=1}^m \lambda_i S_i S_i' \\ B &= \sum_{i=1}^m S_i S_i' \end{aligned}$$

Se V é um espaço vetorial (todos os espaços vetoriais que utilizaremos serão reais) indicaremos com $\langle \cdot, \cdot \rangle$ um produto interno ou seja uma forma bilinear, simétrica e positiva. Todo produto interno pode ser representado através de uma matriz simétrica e definida positiva. Ou seja existe uma matriz Σ simétrica e definida positiva tal que

$$\langle x, y \rangle = x' \Sigma y.$$

Usaremos as vezes a notação $\langle \cdot, \cdot \rangle_{\Sigma}$ para indicar que o produto interno que estamos considerando está determinado pela matriz Σ . Um produto interno induz uma norma

$$\|x\| = (x' \Sigma x)^{1/2}.$$

As vezes usaremos a notação $\|x\|_{\Sigma}$ para enfatizar a dependência sobre Σ . A seguinte desigualdade é bem conhecida.

Proposição 5 (Desigualdade de Cauchy-Schwarz).

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

Com igualdade se e somente se x e y são proporcionais.

Seja $E \subseteq V$ um subespaço vetorial do espaço vetorial V . Seja Σ uma matriz simétrica definida positiva e $a \in V$. Então existe $a^* \in E$ único tal que realiza o

$$\min_{b \in E} \|a - b\|_{\Sigma} = \|a^* - b\|_{\Sigma}$$

a^* é chamado *projeção de a em E* e denotada por $P_\Sigma(a, E) = a^*$.

A seguinte proposição resume os principais fatos relacionados com a projeção de um vetor num subespaço.

Proposição 6. (i) $P_\Sigma(a, E)$ é uma transformação linear idempotente.

(ii) Se E^\perp indica o conjunto dos vetores Σ -ortogonais a todos os vetores de E , resulta que E^\perp é subespaço e todo vetor $a \in V$ escreve-se de forma única como soma de um vetor a^* em E e outro vetor $a^{**} \in E^\perp$. Ou seja

$$a = a^* + a^{**}, \quad a^* = P_\Sigma(a, E) \in E, \quad a^{**} \in E^\perp$$

$$(iii) \quad \|a - a^*\|_\Sigma^2 = \|a\|_\Sigma^2 - \|a^*\|_\Sigma^2$$

$$(iv) \quad \|a^*\|_\Sigma^2 \leq \|a\|_\Sigma^2$$

(v) Seja $\{a_1, \dots, a_p\}$ uma base Σ -ortonormal de E . Então esta base pode ser extendida a uma base $\{a_1, \dots, a_p, a_{p+1}, \dots, a_n\}$ Σ -ortonormal do espaço vetorial V .

Temos também para todo $a \in V$

$$a = \sum_{i=1}^n \langle a, a_i \rangle_\Sigma a_i$$

$$a^* = \sum_{i=1}^p \langle a, a_i \rangle_\Sigma a_i$$

O seguinte teorema terá uma série de importantes aplicações nos capítulos seguintes.

Teorema 3 (Decomposição em valores singulares). *Seja Y uma matriz $n \times m$ de posto r . Seja Σ_1 uma matriz $n \times n$ simétrica definida positiva, e Σ_2 $m \times m$ outra matriz simétrica definida positiva. (Σ_1 e Σ_2 induzem produtos internos em \mathbb{R}^n e \mathbb{R}^m respectivamente). Então existem U $n \times n$ e V $m \times m$ tais que:*

a) Os vetores coluna da matriz U , U_1, \dots, U_n , são Σ_1 -ortonormais.

b) Os vetores coluna da matriz V , V_1, \dots, V_m , são Σ_2 -ortonormais.

$$c) \ U' Y V = \left(\begin{array}{cc|c} \theta_1 & 0 & 0 \\ & \ddots & \\ 0 & \theta_r & \\ \cdots & & \\ 0 & & 0 \end{array} \right) \text{ onde } \theta_1 \geq \theta_2 \geq \cdots \geq \theta_r > 0.$$

Demonstração. A matriz $Y \Sigma_2^{-1} Y'$ é uma matriz $n \times n$ simétrica, definida não negativa e de posto r . Pelo Teorema Espectral (generalizado) existe $U \ n \times n$ tal que

$$U' Y \Sigma_2^{-1} Y' U = \begin{pmatrix} \theta_1^2 & & 0 \\ & \ddots & \\ 0 & & \theta_n^2 \end{pmatrix}$$

$$U' \Sigma_1 U = I$$

onde $\theta_1^2 \geq \theta_2^2 \geq \cdots \geq \theta_r^2 > \theta_{r+1}^2 = \cdots = \theta_n^2 = 0$.

(Note-se que os autovalores de $Y \Sigma_2^{-1} Y'$ são não negativos porque esta matriz é definida não negativa). Se U_i indica a coluna i de U sabemos também que

$$(\Sigma_1^{-1} Y \Sigma_2^{-1} Y') U_i = \theta_i^2 U_i$$

ou seja U_i é autovetor de $\Sigma_1^{-1} Y \Sigma_2^{-1} Y'$ correspondente ao autovetor θ_i^2 . Para $i = 1, 2, \dots, r$ seja

$$V_i = \frac{\Sigma_2^{-1} Y' U_i}{\theta_i}.$$

Temos

$$\begin{aligned} (\Sigma_2^{-1} Y' \Sigma_1^{-1} Y) V_i &= \frac{\Sigma_2^{-1} Y' \Sigma_1^{-1} Y \Sigma_2^{-1} Y' U_i}{\theta_i} = \\ &= \frac{\Sigma_2^{-1} Y' (\Sigma_1^{-1} Y \Sigma_2^{-1} Y') U_i}{\theta_i} = \theta_i \Sigma_2^{-1} Y' U_i = \theta_i^2 V_i \end{aligned}$$

Ou seja V_1, \dots, V_r são autovetores de $\Sigma_2^{-1} Y' \Sigma_1^{-1} Y$ correspondentes aos autovalores $\theta_1^2, \dots, \theta_r^2$.

Temos também

$$\begin{aligned} V'_i, \Sigma_2 V_j &= \frac{U'_i Y \Sigma_2^{-1}}{\theta_i} \Sigma_2 \frac{\Sigma_2^{-1} Y' U_j}{\theta_j} = \frac{U'_i Y \Sigma_2^{-1} Y' U_j}{\theta_i \theta_j} = \\ &= \frac{\delta_{ij} \theta_i^2}{\theta_i \theta_j} = \delta_{ij} \end{aligned}$$

Ou seja V_1, \dots, V_r são Σ_2 -ortonormais.

Para escolher os restantes V_j procedemos da seguinte forma. A matriz $\Sigma_2^{-1} Y' \Sigma_1^{-1} Y$ é uma matriz $m \times m$ de posto r . Portanto como transformação de \mathbb{R}^m em \mathbb{R}^m temos que dimensão $(\text{Núcleo}(\Sigma_2^{-1} Y' \Sigma_1^{-1} Y)) = m - r$. Seja V_{r+1}, \dots, V_m uma base Σ_2 -ortonormal desse núcleo. Vamos verificar primeiro que $Y V_j = 0$ para $j \geq r + 1$. Temos $(\Sigma_2^{-1} Y' \Sigma_1^{-1} Y) V_j = 0$ e portanto como Σ_2 é não singular $(Y' \Sigma_1^{-1} Y) V_j = 0$. Como Σ_1^{-1} é simétrica e definida positiva existe uma matriz P não singular tal que $\Sigma_1^{-1} = P' P$. Temos então

$$(Y' P' P Y) V_j = 0.$$

Multiplicando a esquerda por V'_j

$$0 = V'_j Y' P' P Y V_j = \|P Y V_j\|_I^2.$$

Portanto $P Y V_j = 0$ e como P é não singular $Y V_j = 0$ como queríamos provar. Vamos verificar agora que o conjunto $\{V_1, \dots, V_m\}$ é Σ_2 -ortonormal. Pelo visto até agora é suficiente verificar que se $i \leq r$ e $j > r$

$$V'_i \Sigma_2 V_j = 0.$$

Mas

$$V'_i \Sigma_2 V_j = \frac{U'_i Y \Sigma_2^{-1}}{\theta_i} \Sigma_2 V_j = \frac{U'_i Y V_j}{\theta_i} = 0$$

porque $Y V_j = 0$.

Finalmente vamos verificar que os vetores U_i e V_j satisfazem a identidade da parte c). Ou seja vamos verificar que

$$U'_i Y V_j = \begin{cases} \delta_{ij} \theta_i & i \leq r, \quad j \leq r \\ 0 & i > r \quad \text{ou} \quad j > r \end{cases}$$

Se $j \leq r$

$$U'_i Y V_j = U'_i Y \frac{\Sigma_2^{-1} Y' U_j}{\theta_j} = \frac{\delta_{ij} \theta_i^2}{\theta_j} = \delta_{ij} \theta_i.$$

Se $j > r$ como $Y V_j = 0$ temos

$$U'_i Y V_j = 0.$$

Portanto se V é a matriz que tem como colunas os V_j temos

$$U' Y V = \left(\begin{array}{cc|c} \theta_1 & 0 & 0 \\ & \ddots & \\ 0 & \theta_r & \\ \cdots & & \\ 0 & & 0 \end{array} \right)$$

como queríamos demonstrar.

Corolário 1. Definindo $\theta_i = 0$ para todo $i > r$ temos

(i) $\theta_i V_i = \Sigma_2^{-1} Y' U_i$

(ii) $\theta_i U_i = \Sigma_1^{-1} Y V_i$

Demonstração. (i) é válida para $i \leq r$ por definição.

Para $i > r$ $\theta_i = 0$ ou seja o membro esquerdo é 0. Temos que verificar que o membro direito também é nulo. Σ_2^{-1} é uma matriz simétrica definida positiva e sabemos que

$$U' Y \Sigma_2^{-1} Y' U = \left(\begin{array}{cc|c} \theta_1^2 & 0 & 0 \\ & \ddots & \\ 0 & \theta_r^2 & \\ \cdots & & \\ 0 & & 0 \end{array} \right)$$

Portanto para $i > r$

$$0 = U_i' Y \Sigma_2^{-1} Y' U_i = \|Y' U_i\|_{\Sigma_2^{-1}}^2$$

o que implica que $Y' U_i = 0$ e que o membro direito de (i) é nulo.

Vamos verificar agora (ii). Se $i \leq r$ temos

$$\Sigma_1^{-1} Y V_i = \Sigma_1^{-1} Y \frac{\Sigma_2^{-1} Y' U_i}{\theta_i} = \frac{\theta_i^2 U_i}{\theta_i} = \theta_i U_i.$$

Se $i > r$ vimos na demonstração do Teorema da Decomposição em Valores Singulares que $Y V_i = 0$. Ou seja o membro direito de (ii) é 0 para $i > r$ que é o valor do membro esquerdo.

Corolário 2. *Nas condições do Teorema da D.V.S.*

$$\begin{aligned} \max_{X \in \mathbb{R}^n, \|X\|_{\Sigma_1} = 1} X' Y Z &= \theta_1 \\ Z \in \mathbb{R}^m, \|Z\|_{\Sigma_2} &= 1 \end{aligned}$$

O máximo é atingido se $X = U_1$ e $Z = V_1$.

Demonstração. Sejam U e V tais que

$$U' Y V = \left(\begin{array}{cc|c} \theta_1 & 0 & 0 \\ & \ddots & \\ 0 & \theta_r & \\ \dots & & \\ 0 & & 0 \end{array} \right)$$

$$X = \sum_{i=1}^n c_i U_i \quad \text{e} \quad Z = \sum_{j=1}^m d_j V_j \quad \text{com} \quad \sum_{i=1}^n c_i^2 = \sum_{j=1}^m d_j^2 = 1.$$

Então

$$X' Y Z = \left(\sum_{i=1}^n c_i U_i' \right) Y \left(\sum_{j=1}^m d_j V_j \right) = \sum_{i=1}^r \theta_i c_i d_i.$$

Temos pela desigualdade de Cauchy-Schwartz

$$\begin{aligned} \left| \sum_{i=1}^r \theta_i c_i d_i \right| &\leq \left(\sum_{i=1}^r \theta_i c_i^2 \right)^{1/2} \left(\sum_{i=1}^r \theta_i d_i^2 \right)^{1/2} = \\ &\leq \theta_1^{1/2} \left(\sum_{i=1}^r c_i^2 \right)^{1/2} \theta_1^{1/2} \left(\sum_{i=1}^r d_i^2 \right)^{1/2} \end{aligned}$$

Como $\left(\sum_{i=1}^r c_i^2 \right)^{1/2} \leq 1$ e $\left(\sum_{i=1}^r d_i^2 \right)^{1/2} \leq 1$ temos que $|X' Y Z| = \left| \sum_{i=1}^r \theta_i c_i d_i \right| \leq \theta_1$.

Se $c_1 = d_1 = 1$ e o resto é zero, temos $X = U_1$ $Z = V_1$ e $X' Y Z = \theta_1$ e portanto o máximo é atingido.

Outro resultado importante sobre maximização de formas quadráticas está contido na seguinte proposição.

Proposição 7. *Seja A $n \times n$ simétrica e B $n \times n$ simétrica definida positiva. Seja U tal que*

$$\begin{aligned} U' A U &= \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \\ U' B U &= I \end{aligned}$$

Sejam U_1, U_2, \dots, U_n os vetores coluna de U .

Sejam X_1, \dots, X_p p vetores B -ortogonais.

Então

$$\sup_{X_1, \dots, X_p} \sum_{i=1}^p \frac{X_i' A X_i}{X_i' B X_i} = \lambda_1 + \dots + \lambda_p$$

e o máximo é atingido tomando $X_i = U_i$, $i = 1, \dots, p$.

Demonstração. $\{U_1, \dots, U_n\}$ é uma base ortonormal de \mathbb{R}^n . Seja $X_i = \sum_{\ell=1}^n c_{i\ell} U_\ell$, $i = 1, \dots, p$. Completamos os X_i até obter uma base

B -ortogonal de \mathbb{R}^n . Temos

$$\begin{aligned}\delta_{ij} \|X_i\|_B^2 &= \left\langle X_i, X_j \right\rangle_B = \left\langle \sum_{\ell=1}^n c_{i\ell} U_\ell, \sum_{k=1}^n c_{jk} U_k \right\rangle = \\ &= \sum_{\ell} \sum_k c_{i\ell} c_{jk} \langle U_\ell, U_k \rangle = \sum_{\ell} c_{i\ell} c_{j\ell}.\end{aligned}$$

Se definirmos $\alpha_{ij} = \frac{c_{ij}}{\|X_i\|_B}$ temos que $\sum_{\ell=1}^n \alpha_{i\ell} \alpha_{j\ell} = \delta_{ij}$ ou seja a matriz $A = (\alpha_{ij})$ é ortogonal.

Temos em particular que $\sum_{j=1}^n \alpha_{ij}^2 = 1$ para todo i , $\sum_{i=1}^p \alpha_{ij}^2 \leq 1$ para todo

j e $\sum_{i=1}^p \sum_{j=1}^n \alpha_{ij}^2 = p$.

Agora, a expressão a maximizar é igual a

$$\begin{aligned}&\sum_{i=1}^p \frac{\left(\sum_{\ell=1}^n c_{i\ell} U'_\ell \right) A \left(\sum_{j=1}^n c_{ij} U_j \right)}{X'_i B X_i} = \sum_{i=1}^p \frac{\sum_{j=1}^n c_{ij}^2 \lambda_j}{\|X_i\|_B^2} = \\ &= \sum_{i=1}^p \sum_{j=1}^n \alpha_{ij}^2 \lambda_j = \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^p \alpha_{ij}^2 \right) = \sum_{j=1}^n \lambda_j \gamma_j\end{aligned}$$

onde $\gamma_j = \sum_{i=1}^p \alpha_{ij}^2$ por definição.

Temos pelo provado anteriormente que

$$\gamma_j \geq 0, \quad \gamma_j \leq 1, \quad \sum_{j=1}^n \gamma_j = p.$$

Portanto o máximo de $\sum_{j=1}^n \lambda_j \gamma_j$ se obtém tomando-se $\gamma_1 = \gamma_2 = \dots =$

$\gamma_p = 1$, $\gamma_{p+1} = \dots = \gamma_n = 0$. Dessa forma o máximo fica igual a $\sum_{i=1}^p \lambda_i$.

Para essa escolha de γ 's temos que

$$C = \left(\begin{array}{cc|c} c_{11} & c_{1p} & 0 \\ c_{p1} & c_{pp} & \end{array} \right)$$

Tomando $C = (I : 0)$ atingimos o máximo $\sum_{i=1}^p \lambda_i$. Esta escolha de C equivale a tomar $X_i = U_i$ para $i = 1, \dots, p$. A Proposição 8 generaliza o Corolário 2.

Proposição 8. *Seja Y $n \times m$ de posto r . Sejam U e V as matrizes que aparecem na decomposição de Y em valores singulares*

$$U' Y V = \left(\begin{array}{cc|c} \theta_1 & 0 & 0 \\ & \ddots & \\ 0 & \theta_r & \\ \dots & & \\ 0 & & 0 \end{array} \right) \quad \theta_1 \geq \theta_2 \geq \dots \geq \theta_r > 0$$

Então se $p \leq \min(m, n)$

$$\min_{\substack{X_i \text{ } \Sigma_1\text{-ortonormais} \\ Z_i \text{ } \Sigma_2\text{-ortonormais}}} \sum_{i=1}^p X_i' Y Z_i = \sum_{i=1}^p \theta_i.$$

O máximo é atingido tomando $X_i = U_i$, $i = 1, \dots, p$ e $Z_i = V_i$, $i = 1, \dots, p$ onde U_i e V_i são as colunas de U e V respectivamente.

Demonstração. Sejam U ($n \times n$) e V ($m \times m$) as matrizes que aparecem na decomposição em valores singulares de Y :

$$U' Y V = \left(\begin{array}{cc|c} \theta_1 & 0 & 0 \\ & \ddots & \\ 0 & \theta_r & \\ \dots & & \\ 0 & & 0 \end{array} \right)$$

com $\theta_1 \geq \theta_2 \geq \dots \geq \theta_r > 0$ e $r = \text{posto}(Y)$.

Seja $X_i = \sum_{\ell=1}^n c_{i\ell} U_\ell$ e $Z_i = \sum_{h=1}^m d_{ih} V_h$.

Temos

$$X_i' Y Z_i = \left(\sum_{\ell=1}^n c_{i\ell} U_\ell \right)' Y \left(\sum_{h=1}^m d_{ih} V_h \right) = \sum_{\ell=1}^r c_{i\ell} d_{i\ell} \theta_\ell.$$

Portanto

$$\sum_{i=1}^p X_i' Y Z_i = \sum_{\ell=1}^r \left(\sum_{i=1}^p c_{i\ell} d_{i\ell} \right) \theta_\ell.$$

Seja $\gamma_\ell = \sum_{i=1}^p c_{i\ell} d_{i\ell}$.

Como os X_i (resp. Z_i) são Σ_1 -ortonormais (resp. Σ_2 -ortonormais) podemos sempre supor que são parte de uma base Σ_1 -ortonormal (resp. Σ_2 -ortonormal) de \mathbb{R}^n (resp. de \mathbb{R}^m). Ou seja podemos supor que a matriz dos $c_{i\ell}$ (resp. $d_{i\ell}$) é uma matriz ortogonal $n \times n$ (resp. $m \times m$). Temos agora

$$\begin{aligned} |\gamma_\ell| &= \left| \sum_{i=1}^p c_{i\ell} d_{i\ell} \right| \leq \left(\sum_{i=1}^p c_{i\ell}^2 \right)^{1/2} \left(\sum_{i=1}^p d_{i\ell}^2 \right)^{1/2} \leq \\ &\leq \left(\sum_{i=1}^n c_{i\ell}^2 \right)^{1/2} \left(\sum_{i=1}^m d_{i\ell}^2 \right)^{1/2} \leq 1 \end{aligned}$$

$$\begin{aligned} \left| \sum_{\ell=1}^r \gamma_\ell \right| &= \left| \sum_{\ell=1}^r \sum_{i=1}^p c_{i\ell} d_{i\ell} \right| = \left| \sum_{i=1}^p \sum_{\ell=1}^r c_{i\ell} d_{i\ell} \right| \leq \\ &\leq \sum_{i=1}^p \left| \sum_{\ell=1}^r c_{i\ell} d_{i\ell} \right| \leq \sum_{i=1}^p \left(\sum_{\ell=1}^r c_{i\ell}^2 \right)^{1/2} \left(\sum_{\ell=1}^r d_{i\ell}^2 \right)^{1/2} \leq \\ &\leq \sum_{i=1}^p \left(\sum_{\ell=1}^n c_{i\ell}^2 \right)^{1/2} \left(\sum_{\ell=1}^m d_{i\ell}^2 \right)^{1/2} \leq p. \end{aligned}$$

Estas duas desigualdades implicam que o máximo de $\sum_{\ell=1}^r \gamma_\ell \theta_\gamma$ é obtido tomando $\gamma_\ell = 1$ para $\ell = 1, \dots, p$ e $\gamma_\ell = 0$ para $\ell > p$. (Note-se que p pode ser maior que r). Ou seja o máximo é igual a $\sum_{\ell=1}^p \theta_\ell$.

O máximo é atingido tomando $X_i = U_i$, $i = 1, \dots, p$ e $Z_i = V_i$, $i = 1, \dots, p$.

Capítulo 2

Análise de Componentes Principais

A técnica de Componentes Principais pode ser apresentada por diferentes pontos de vista. Nós começaremos a exposição tomando o ponto de vista de Análise de Dados, isto é, uma técnica de representação de dados que permite reduzir tabelas de grande dimensão e tabelas mais pequenas que podem ser graficadas e sobre as quais a interpretação e descoberta de relações entre os dados seja bem mais simples. Esta redução do número de dados conservando na medida do possível a “informação” contida nos dados originais é feita utilizando certos critérios de otimalidade geométrica e/ou algébrica.

2.1 Notação e descrição do problema

A situação típica na qual a Análise de Componentes Principais é executada é a seguinte:

Sobre cada um dos integrantes de um grupo de indivíduos (objetos, famílias, classes de renda, países, etc.) é levantado ou determinado um certo número de medidas ou variáveis.

Por exemplo sobre cada indivíduo a quantidade de diversas substâncias químicas na urina e no sangue podem ser analisadas.

Ou para cada família é determinado a quantidade de dinheiro gasto em diferentes tipos de despesas (alimentos, transporte, educação, saúde, etc.).

Em qualquer caso o conjunto de dados objeto da análise consiste numa tabela na qual colocaremos em cada linha os valores das variáveis medidas sobre o indivíduo correspondente a essa linha. Mais precisamente seja I o número de indivíduos e J o número de variáveis. A tabela a ser analisada consiste de uma matriz $X = (x_{ij})$ $I \times J$ onde x_{ij} é o valor da variável j no indivíduo i . Vamos nos referir as linhas da matriz como *indivíduos* ou *observações* e as colunas como *variáveis*.

Dado o caráter repetitivo que tem as linhas de X , é conveniente as vezes pensar a tabela como um conjunto de I observações de um vetor aleatório em \mathbb{R}^J .

Indicaremos com x_i o vetor de \mathbb{R}^J contendo os valores das variáveis sobre o indivíduo i . Ou seja $x'_i = (x_{i1}, \dots, x_{iJ})$. Se Σ é uma matriz simétrica definida positiva $\|x_i - x_{i'}\|_{\Sigma}^2$ indicará o quadrado da distância determinada por Σ entre os vetores x_i e $x_{i'}$. Os casos usuais serão $\Sigma = I$ (a matriz identidade e o número de indivíduos serão denotados pelo mesmo símbolo I ; o contexto indicará naturalmente de que se trata) e neste caso

$$\|x_i - x_{i'}\|_{\Sigma}^2 = \sum_{j=1}^J (x_{ij} - x_{i'j})^2$$

ou se $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_J)$ com os λ 's todos positivos e então

$$\|x_i - x_{i'}\|_{\Sigma}^2 = \sum_{j=1}^J \lambda_j (x_{ij} - x_{i'j})^2.$$

No caso geral

$$\|x_i - x_{i'}\|_{\Sigma}^2 = (x_i - x_{i'})' \Sigma (x_i - x_{i'}).$$

Se V é uma variedade linear (transladado de subespaço) indicaremos com $x_i^* = P_{\Sigma}(x_i, V)$ a projeção do vetor x_i na variedade V com relação ao produto interno determinado por Σ . x_i^* satisfaz

$$x_i^* = P_{\Sigma}(x_i, V) \text{ e } \min_{y \in V} \|x_i - y\|_{\Sigma}^2 = \|x_i - x_i^*\|_{\Sigma}^2.$$

Finalmente é importante em diversas situações (por exemplo em Análise de Correspondências) considerar o caso no qual os indivíduos não são todos igualmente importantes ou, em outros termos, não tem

todos o mesmo peso. Introduzimos então números μ_i $i = 1, \dots, I$ não negativos e tais que $\sum_{i=1}^I \mu_i = 1$. Nos casos usuais, $\mu_i = \frac{1}{I}$, $i = 1, \dots, I$.

O problema de Análise de Componentes Principais pode agora ser colocado da seguinte forma:

Dado $p \leq J$, encontrar uma variedade linear de dimensão p tal que

$$\sum_{i=1}^I \mu_i \|x_i - x_i^*\|_{\Sigma}^2$$

seja mínimo.

A condição $\sum_{i=1}^I \mu_i = 1$ não é utilizada em geral. Em algumas Proposições (por exemplo na Proposição 2.2 (vi) mais adiante) só introduziria uma constante no resultado final.

2.2 Redução a subespaços

Se V é uma variedade linear em \mathbb{R}^J , existe $b \in \mathbb{R}^J$ tal que $V = b + E$ onde E é um subespaço.

O seguinte lema é de fácil demonstração

Lema 2.1. *Se $V = b + E$ então*

$$P_{\Sigma}(x, V) = P_{\Sigma}(x - b, E) + b;$$

A seguinte proposição vai reduzir o problema de achar a melhor variedade ao de determinar o melhor subespaço.

Proposição 2.1. *Se $\sum_{i=1}^I \mu_i x_i = 0$ a variedade linear ótima (que é a solução do problema de Componentes Principais) é um subespaço.*

Demonstração. Queremos provar que

$$\sum_{i=1}^I \mu_i \|P_{\Sigma}(x_i, V) - x_i\|_{\Sigma}^2 \geq \sum_{i=1}^I \mu_i \|P_{\Sigma}(x_i, E) - x_i\|_{\Sigma}^2.$$

Ou seja para toda variedade V existe um subespaço E com soma menor e portanto podemos nos retrair a subespaços.

Temos que

$$\begin{aligned} \|P_{\Sigma}(x_i, V) - x_i\|_{\Sigma}^2 &= \|P_{\Sigma}(x_i - b, E) - x_i + b\|_{\Sigma}^2 = \\ &= \|(P_{\Sigma}(x_i, E) - x_i) - (P_{\Sigma}(b, E) - b)\|_{\Sigma}^2 = \\ &= \|P_{\Sigma}(x_i, E) - x_i\|_{\Sigma}^2 + \|P_{\Sigma}(b, E) - b\|_{\Sigma}^2 - \\ &\quad - 2\langle P_{\Sigma}(x_i, E) - x_i, P_{\Sigma}(b, E) - b \rangle. \end{aligned}$$

A primeira igualdade é válida pelo Lema anterior. Multiplicando por μ_i e somando temos

$$\begin{aligned} \sum_{i=1}^I \mu_i \|P_{\Sigma}(x_i, V) - x_i\|_{\Sigma}^2 &= \sum_{i=1}^I \mu_i \|P_{\Sigma}(x_i, E) - x_i\|_{\Sigma}^2 + \\ &+ \sum_{i=1}^I \mu_i \|P_{\Sigma}(b, E) - b\|_{\Sigma}^2 - 2 \sum_{i=1}^I \mu_i \langle P_{\Sigma}(x_i, E) - x_i, P_{\Sigma}(b, E) - b \rangle_{\Sigma}. \end{aligned}$$

Para provar a desigualdade basta provar que o último termo do membro direito é 0 porque o seguinte termo é não negativo. Mas o último termo é igual a

$$2\left\langle P_{\Sigma}\left(\sum_{i=1}^I \mu_i x_i, E\right) - \sum_{i=1}^I \mu_i x_i, P_{\Sigma}(b, E) - b \right\rangle$$

usando as propriedades de bilinearidade de \langle, \rangle e linearidade de P_{Σ} .

Como por hipóteses $\sum_{i=1}^I \mu_i x_i = 0$ o último termo é igual a 0 e a desigualdade fica provada.

A Proposição 2.1 é utilizada da seguinte forma. Para uma matriz X defino $\bar{x} = \sum_{i=1}^I \mu_i x_i$. Seja $y_i = x_i - \bar{x}$. Esta nova matriz com linhas y'_i

satisfaz a restrição $\sum_{i=1}^I \mu_i y_i = 0$. Seja E o espaço vetorial ótimo para esta nova matriz. Então a variedade linear $\bar{x} + E$ é a solução do problema de Componentes Principais original.

De agora em diante vamos supor que $\sum_{i=1}^I \mu_i x_i = 0$ e portanto limitar a procura da solução ótima a subespaços.

2.3 Solução do problema de Componentes Principais

Em lugar de procurar um subespaço E vamos determinar um conjunto de p vetores, W_1, \dots, W_p , Σ -ortonormais, que sejam uma base para E , ou seja $E = [W_1, \dots, W_p]$ onde $[]$ indica o subespaço gerado pelos vetores W_1, \dots, W_p .

Teorema 2.1.

- (i) Seja $C = \sum_{i=1}^I \mu_i x_i x_i'$. Então C é simétrica e definida não negativa.
- (ii) Seja U $J \times J$ tal que

$$U' C U = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_J \end{pmatrix} \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_J$$

$$U' \Sigma^{-1} U = I$$

(Note que como C é definida não negativa os λ 's são não negativos). Seja U_α a coluna α de U e $W_\alpha = \Sigma^{-1} U_\alpha$. Então o conjunto $\{W_1, \dots, W_p\}$ é Σ -ortonormal e o subespaço $E = [W_1, \dots, W_p]$ é ótimo, no sentido que $\sum_{i=1}^I \mu_i \|x_i - x_i^*\|_\Sigma^2$ é mínimo onde $x_i^* = P_\Sigma(x_i, E)$.

Demonstração.

- (i) A verificação é imediata.
- (ii) Como $\|x_i - x_i^*\|_\Sigma^2 = \|x_i\|_\Sigma^2 - \|x_i^*\|_\Sigma^2$, mimizar $\sum_{i=1}^I \mu_i \|x_i - x_i^*\|_\Sigma^2$ é equivalente a maximizar $\sum_{i=1}^I \mu_i \|x_i^*\|_\Sigma^2$.

Seja W_1, \dots, W_p uma base Σ -ortonormal do subespaço E (desconhecido).

Temos

$$\|x_i^*\|_\Sigma^2 = \sum_{\alpha=1}^p (\langle x_i, W_\alpha \rangle_\Sigma)^2 = \sum_{\alpha=1}^p (W_\alpha' \Sigma x_i)(x_i' \Sigma W_\alpha).$$

Portanto

$$\begin{aligned}
 \sum_{i=1}^I \mu_i \|x_i^*\|_{\Sigma}^2 &= \sum_{i=1}^I \mu_i \sum_{\alpha=1}^p (W'_{\alpha} \Sigma x_i)(x'_i \Sigma W_{\alpha}) = \\
 &= \sum_{\alpha=1}^p W'_{\alpha} \Sigma \left(\sum_{i=1}^I \mu_i x_i x'_i \right) \Sigma W_{\alpha} = \sum_{\alpha=1}^p W'_{\alpha} \Sigma C \Sigma W_{\alpha} = \\
 &= \sum_{\alpha=1}^p U'_{\alpha} C U_{\alpha} \quad \text{onde} \quad U_{\alpha} = \Sigma W_{\alpha}
 \end{aligned}$$

Os U_{α} são Σ^{-1} -ortonormais porque em geral se U_{α} são Σ^{-1} -ortonormais $W_{\alpha} = \Sigma^{-1} U_{\alpha}$ são Σ -ortonormais

$$U'_{\alpha} \Sigma^{-1} U'_{\alpha} = W'_{\alpha} \Sigma \Sigma^{-1} \Sigma W'_{\alpha} = \delta_{\alpha\alpha'}.$$

Portanto para encontrar o máximo da expressão $\sum_{\alpha=1}^p U'_{\alpha} C U_{\alpha}$ com C simétrica e o conjunto $\{U_1, \dots, U_p\}$ Σ^{-1} -ortonormal, podemos utilizar a Proposição 7 do Capítulo 1. Por essa Proposição existe uma matriz U $J \times J$ tal que

$$\begin{aligned}
 U' C U &= \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_J \end{pmatrix} \quad \lambda_1 \geq \dots \geq \lambda_J \\
 U' \Sigma^{-1} U &= I
 \end{aligned}$$

O máximo de $\sum_{\alpha=1}^p X'_{\alpha} C X_{\alpha}$ onde $X_{\alpha} \in \mathbb{R}^J$ e os $X_i \Sigma^{-1}$ ortonormais é $\lambda + \dots + \lambda_p$ e este máximo é atingido tomando $X_{\alpha} = U_{\alpha}$ para $\alpha = 1, 2, \dots, p$. Isto termina a demonstração do Teorema 2.1. Note-se que na demonstração acabamos também calculando o valor do máximo de $\sum_{i=1}^I \mu_i \|x_i^*\|_{\Sigma}^2$ que é igual a $\sum_{\alpha=1}^p \lambda_{\alpha}$.

Note-se também que a demonstração indica o melhor subespaço para uma matriz de dados iniciais X , independentemente da hipótese $\sum_{i=1}^I \mu_i x_i = 0$.

Em diversas situações utilizaremos todos os J vetores W_1, \dots, W_J (ou U_1, \dots, U_J) sendo que os p primeiros W_1, \dots, W_p geram o subespaço ótimo de dimensão p .

Indicaremos com $F_\alpha(i)$ a componente do vetor x_i sobre o eixo α . Ou seja

$$x_i = \sum_{\alpha=1}^J F_\alpha(i) W_\alpha$$

e também

$$x_i^* = \sum_{\alpha=1}^p F_\alpha(i) W_\alpha$$

$F_\alpha(i)$ será chamada a α -ésima *componente principal* do indivíduo i . $F_i^{(p)}$ indicará o vetor de \mathbb{R}^p contendo as p componentes principais do vetor x_i . Ou seja

$$F_i^{(p)} = \begin{pmatrix} F_1(i) \\ F_2(i) \\ \vdots \\ F_p(i) \end{pmatrix}$$

A seguinte proposição indica as principais propriedades das componentes principais.

Proposição 2.2.

- (i) $F_\alpha(i) = U'_\alpha x_i$
- (ii) $\|x_i - x_i^*\|_\Sigma^2 = \sum_{\alpha=p+1}^J F_\alpha^2(i)$
- (iii) $\sum_{i=1}^I \mu_i F_\alpha(i) = 0$
 $\sum_{i=1}^I \mu_i F_\alpha(i) F_{\alpha'}(i) = \lambda_\alpha \delta_{\alpha\alpha'}$

Em particular $\sum_{i=1}^I \mu_i F_\alpha(i)^2 = \lambda_\alpha$; ou seja as componentes sobre o eixo α tem média 0 e variância λ_α .

- (iv) $\sum_{i=1}^I \mu_i \|x_i\|_\Sigma^2 = \sum_{j=1}^J \lambda_j$

e

$$\sum_{i=1}^I \mu_i \|x_i - x_i^*\|_{\Sigma}^2 = \sum_{j=p+1}^J \lambda_j$$

$$(v) \quad \|x_i^* - x_{i'}^*\|_{\Sigma}^2 = \|F_i^{(p)} - F_{i'}^{(p)}\|_i^2$$

Ou seja a distância Σ entre x_i^ e $x_{i'}^*$ é igual a distância euclidiana entre os vetores $F_i^{(p)}$ e $F_{i'}^{(p)}$.*

$$(vi) \quad \sum_i \sum_{i'} \mu_i \mu_{i'} \left| \|x_i^* - x_{i'}^*\|_{\Sigma}^2 - \|x_i - x_{i'}\|_{\Sigma}^2 \right| = 2 \sum_{\alpha=p+1}^J \lambda_{\alpha}$$

(vii) *O subespaço gerado por W_1, \dots, W_p minimiza a expressão*

$$\sum_i \sum_{i'} \mu_i \mu_{i'} \left| \|x_i^* - x_{i'}^*\|_{\Sigma}^2 - \|x_i - x_{i'}\|_{\Sigma}^2 \right|$$

Demonstração.

$$(i) \quad F_{\alpha}(i) = \langle W_{\alpha}, x_i \rangle_{\Sigma} = W'_{\alpha} \Sigma x_i = U'_{\alpha} x_i$$

$$(ii) \quad \text{Temos } x_i = \sum_{\alpha=1}^J F_{\alpha}(i) W_{\alpha} \quad \text{e}$$

$$x_i^* = \sum_{\alpha=1}^p F_{\alpha}(i) W_{\alpha}$$

Portanto

$$\|x_i - x_i^*\|_{\Sigma}^2 = \left\| \sum_{\alpha=p+1}^J F_{\alpha}(i) W_{\alpha} \right\|_{\Sigma}^2 = \sum_{\alpha=p+1}^J F_{\alpha}^2(i)$$

porque os W_{α} são Σ -ortonormais.

$$(iii) \quad \sum_{i=1}^I \mu_i F_{\alpha}(i) = \sum_{j=1}^I \mu_j U'_{\alpha} x_j = U'_{\alpha} \left(\sum_{i=1}^I \mu_i x_i \right) = 0$$

$$\text{porque } \sum_{i=1}^I \mu_i x_i = 0.$$

$$\begin{aligned} \sum_{i=1}^I \mu_i F_{\alpha}(i) F_{\alpha'}(i) &= \sum_{i=1}^I \mu_i U'_{\alpha} x_i x'_i U_{\alpha'} = \\ &= U'_{\alpha} \left(\sum_{i=1}^I \mu_i x_i x'_i \right) U_{\alpha'} = U'_{\alpha} C U_{\alpha'} = \lambda_{\alpha} \delta_{\alpha\alpha'} \end{aligned}$$

(iv) Sabemos que $\sum_{i=1}^I \mu_i \|x_i^*\|_\Sigma^2 = \sum_{j=1}^p \lambda_j$.

Se $p = J$, $x_i^* = x_i$, o que prova a primeira igualdade.

Temos também que

$$\begin{aligned} \sum_{i=1}^I \mu_i \|x_i - x_i^*\|_\Sigma^2 &= \sum_{i=1}^I \mu_i \|x_i\|_\Sigma^2 - \sum_{i=1}^I \mu_i \|x_i^*\|_\Sigma^2 = \\ &= \sum_{j=1}^J \lambda_j - \sum_{j=1}^p \lambda_j = \sum_{j=p+1}^J \lambda_j. \end{aligned}$$

$$\begin{aligned} \text{(v)} \quad \|x_i^* - x_{i'}^*\|_\Sigma^2 &= \left\| \sum_{\alpha=1}^p F_\alpha(i) W_\alpha - \sum_{\alpha=1}^p F_\alpha(i') W_\alpha \right\|_\Sigma^2 = \\ &= \left\| \sum_{\alpha=1}^p (F_\alpha(i) - F_\alpha(i')) W_\alpha \right\|_\Sigma^2 = \sum_{\alpha=1}^p (F_\alpha(i) - F_\alpha(i'))^2 = \\ &= \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2. \end{aligned}$$

$$\begin{aligned} \text{(vi)} \quad &\sum_i \sum_{i'} \mu_i \mu_{i'} \left| \|x_i^* - x_{i'}^*\|_\Sigma^2 - \|x_i - x_{i'}\|_\Sigma^2 \right| = \\ &\sum_i \sum_{i'} \mu_i \mu_{i'} (\|x_i - x_{i'}\|_\Sigma^2 - \|x_i^* - x_{i'}^*\|_\Sigma^2) = \\ &\sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i - x_{i'}\|_\Sigma^2 - \sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i^* - x_{i'}^*\|_\Sigma^2 \end{aligned}$$

Vamos calcular agora o segundo membro desta diferença:

$\sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i^* - x_{i'}^*\|_\Sigma^2$. Note-se que o primeiro membro é um caso particular, correspondendo ao caso $p = J$ (e portanto $x_i^* = x_i$). Temos

$$\begin{aligned} \sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i^* - x_{i'}^*\|_\Sigma^2 &= \sum_i \sum_{i'} \mu_i \mu_{i'} \sum_{\alpha=1}^p (F_\alpha(i) - F_\alpha(i'))^2 = \\ &= \sum_i \sum_{i'} \sum_{\alpha} \mu_i \mu_{i'} F_\alpha^2(i) + \sum_i \sum_{i'} \sum_{\alpha} \mu_i \mu_{i'} F_\alpha^2(i') \\ &\quad - 2 \sum_i \sum_{i'} \sum_{\alpha} \mu_i \mu_{i'} F_\alpha(i) F_\alpha(i') = \\ &= \sum_{\alpha=1}^p \left(\sum_{i=1}^I \mu_i F_\alpha^2(i) \right) + \sum_{\alpha=1}^p \left(\sum_{i'=1}^I \mu_{i'} F_\alpha^2(i') \right) \\ &\quad - 2 \sum_{\alpha=1}^p \left(\sum_{i=1}^I \mu_i F_\alpha(i) \right) \left(\sum_{i'=1}^I \mu_{i'} F_\alpha(i') \right) = 2 \sum_{\alpha=1}^p \lambda_\alpha. \end{aligned}$$

A primeira igualdade desta cadeia usa a parte (v), e a última igualdade decorre da parte (iii).

Temos então que

$$\begin{aligned} & \sum_i \sum_{i'} \mu_i \mu_{i'} \left| \|x_i^* - x_{i'}^*\|_\Sigma^2 - \|x_i - x_{i'}\|_\Sigma^2 \right| = \\ & = 2 \sum_{\alpha=1}^J \lambda_\alpha - 2 \sum_{\alpha=1}^p \lambda_\alpha = 2 \sum_{\alpha=p+1}^J \lambda_\alpha \end{aligned}$$

como queríamos provar.

(vii) Temos que

$$\begin{aligned} & \sum_i \sum_{i'} \mu_i \mu_{i'} \left| \|x_i^* - x_{i'}^*\|_\Sigma^2 - \|x_i - x_{i'}\|_\Sigma^2 \right| = \\ & \sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i - x_{i'}\|_\Sigma^2 - \sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i^* - x_{i'}^*\|_\Sigma^2 \end{aligned}$$

Portanto minimizar o membro esquerdo desta igualdade é equivalente a maximizar

$$\sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i^* - x_{i'}^*\|_\Sigma^2.$$

Esta expressão é igual a

$$\begin{aligned} & \sum_i \sum_{i'} \mu_i \mu_{i'} \left\{ \|x_i^*\|_\Sigma^2 + \|x_{i'}^*\|_\Sigma^2 - 2 \langle x_i^*, x_{i'}^* \rangle_\Sigma \right\} = \\ & = 2 \sum_i \mu_i \|x_i^*\|_\Sigma^2 - 2 \left\langle \sum_i \mu_i x_i^*, \sum_{i'} \mu_{i'} x_{i'}^* \right\rangle. \end{aligned}$$

Agora o último termo desta diferença é igual a zero porque $\sum_i \mu_i x_i = 0$ e

$\sum_i \mu_i x_i^* = \sum_i \mu_i P_\Sigma(x_i, E) = P_\Sigma\left(\sum_i \mu_i x_i, E\right) = 0$. Portanto a expressão a maximizar é igual a $2 \sum_i \mu_i \|x_i^*\|_\Sigma^2$ o que é equivalente a minimizar

$\sum_{i=1}^I \mu_i \|x_i - x_i^*\|_\Sigma^2$ como vimos na demonstração da parte (ii) do Teorema

2.1. Também vimos nesse teorema que o mínimo de $\sum_{i=1}^I \mu_i \|x_i - x_i^*\|_\Sigma^2$

obtém-se tomando como subespaço o gerado por W_1, \dots, W_p . Isto prova (vii) e a Proposição 2.2.

Note-se que como consequência das partes (iv) e (vi)

$$\frac{\sum_{i=1}^I \mu_i \|x_i - x_i^*\|_\Sigma^2}{\sum_{i=1}^I \mu_i \|x_i\|_\Sigma^2} = \frac{\sum_i \sum_{i'} \mu_i \mu_{i'} \left| \|x_i^* - x_{i'}^*\|_\Sigma^2 - \|x_i - x_{i'}\|_\Sigma^2 \right|}{\sum_i \sum_{i'} \mu_i \mu_{i'} \|x_i - x_{i'}\|_\Sigma^2} = \frac{\sum_{\alpha=p+1}^J \lambda_\alpha}{\sum_{\alpha=1}^J \lambda_\alpha}.$$

Este último quociente é chamado *erro relativo* e dá uma ideia da qualidade da representação em \mathbb{R}^p . Se quisermos, por exemplo, que o erro relativo seja menor que 10% escolhemos p de forma tal que $\left(\sum_{\alpha=p+1}^J \lambda_\alpha \right) / \left(\sum_{\alpha=1}^J \lambda_\alpha \right) \leq 0,1$. Indicaremos o erro relativo com a notação ER_p ; ou seja

$$ER_p = \frac{\sum_{\alpha=p+1}^J \lambda_\alpha}{\sum_{\alpha=1}^J \lambda_\alpha}.$$

O quociente

$$ER_p = \frac{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \left| \|x_i^* - x_{i'}^*\|_\Sigma^2 - \|x_i - x_{i'}\|_\Sigma^2 \right|}{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \|x_i - x_{i'}\|_\Sigma^2}$$

dá o erro relativo médio entre os quadrados das distâncias entre pares de elementos (indivíduos) quando os x_i são substituídos pelos x_i^* .

A representação dos x_i pelos x_i^* não é muito útil devido a duas razões:

- Os x_i^* estão num subespaço de dimensão menor, mas ainda em \mathbb{R}^J . Portanto não podem ser graficados.
- A distância relevante entre os x_i^* é a dada pela matriz Σ e é difícil de visualizar.

A utilização dos $F_i^{(p)}$ em lugar dos x_i^* elimina esses problemas. Note-se que por (v) temos

$$\|x_i^* - x_{i'}^*\|_\Sigma^2 = \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2.$$

Portanto podemos escrever

$$ER_p = \frac{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \left| \|x_i - x_{i'}\|_{\Sigma}^2 - \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2 \right|}{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \|x_i - x_{i'}\|_{\Sigma}^2}.$$

Os $F_i^{(p)}$ estão num espaço de dimensão menor (usualmente pequeno) p , e as distâncias relevantes entre eles é a distância euclidiana usual dada pela matriz I .

Por exemplo se $p = 2$ e ER_2 é pequeno os $F_i^{(p)}$ serão graficados em um plano e observando a distância usual entre eles teremos uma idéia aproximada da distância entre os x_i correspondentes à matriz Σ . Se $p = 4$ pode-se graficar em um plano as primeiras duas coordenadas do $F_i^{(p)}$ e em outro plano as duas últimas.

ER_p é, porém, uma medida que superestima (pessimista) o erro relativo da representação dos x_i pelos $F_i^{(p)}$. O motivo é o seguinte. Desde o ponto de vista da representação geométrica, como esta depende de uma escala arbitrária, vai ser equivalente representar os $F_i^{(p)}$ ou os $\lambda F_i^{(p)}$ onde λ é um escalar fixo. Podemos então escolher λ de forma tal que minimize

$$\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \left| \|\lambda F_i^{(p)} - \lambda F_{i'}^{(p)}\|_I^2 - \|x_i - x_{i'}\|_{\Sigma}^2 \right|.$$

Indiquemos com λ_0 o valor de λ que minimiza essa expressão. Definimos então o *erro relativo corrigido* (ERC_p) da representação dos x_i pelos $F_i^{(p)}$ como

$$ERC_p = \frac{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \left| \|\lambda F_i^{(p)} - \lambda F_{i'}^{(p)}\|_I^2 - \|x_i - x_{i'}\|_{\Sigma}^2 \right|}{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \|x_i - x_{i'}\|_{\Sigma}^2}.$$

Pode-se provar que λ_0^2 é a medida dos valores $Z_{ii'} = \|x_i - x_{i'}\|_{\Sigma}^2 / \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2$ dando a cada um deles um peso proporcional a $w_{ii'} = \mu_i \mu_{i'} \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2$. De outra forma: sejam g_h , $1 \leq h \leq \frac{I(I-1)}{2}$

os valores de $Z_{ii'}$ ordenados de menor a maior e w_h os pesos correspondentes. Seja h_0 um inteiro tal que

$$\frac{\sum_{h=1}^{h_0} w_h}{\sum_{h=1}^N w_h} \leq 0,5 \quad \text{e} \quad \frac{\sum_{h=1}^{h_0+1} w_h}{\sum_{h=1}^N w_h} > 0,5.$$

Então λ_0^2 é qualquer valor tal que

$$g_{h_0} \leq \lambda_0^2 \leq g_{h_0+1}.$$

Se

$$\frac{\sum_{h=1}^{h_0} w_h}{\sum_{h=1}^N w_h} = 0,5$$

então $\lambda_0^2 = g_{h_0}$.

O valor de ERC_p será menor ou igual que ER_p e $\lambda_0^2 \geq 1$. Esta última desigualdade resulta do fato que $\|F_i^{(p)} - F_{i'}^{(p)}\|_I^2 = \|x_i^* - x_{i'}^*\|^2 \leq \|x_i - x_{i'}\|^2$. ($1 \leq g_1 \leq g_2 \leq \dots \leq g_N$).

Como antes o valor de p pode ser escolhido de forma tal que ERC_p seja menor que uma margem de erro fixado a priori.

Proposição 2.3. $\rho(X_j, F_\alpha) = \frac{\sqrt{\lambda_\alpha} W_\alpha(j)}{S_j}$ onde $\rho(X_j, F_\alpha)$ indica o coeficiente de correlação entre a variável j e o eixo α , e $S_j = \left(\sum_{i=1}^I \mu_i (x_{ij})^2 \right)^{1/2}$.

$$\begin{aligned} \rho(X_j, F_\alpha) &= \frac{\text{cov}(X_j, F_\alpha)}{S_j \sqrt{\lambda_\alpha}} = \frac{\sum_{i=1}^I x_{ij} F_\alpha(i) \mu_i}{S_j \sqrt{\lambda_\alpha}} = \\ &= \frac{\sum_{i=1}^I \mu_i x_{ij} \left(\sum_{j'=1}^J x_{ij'} U_\alpha(j') \right)}{S_j \sqrt{\lambda_\alpha}} = \frac{\sum_{j'=1}^J \left(\sum_{i=1}^I x_{ij'} x_{ij} \mu_i \right) U_\alpha(j')}{S_j \sqrt{\lambda_\alpha}} = \\ &= \frac{\sum_{j'=1}^J C(j, j') U_\alpha(j')}{S_j \sqrt{\lambda_\alpha}} = \frac{C U_\alpha(j)}{S_j \sqrt{\lambda_\alpha}}. \end{aligned}$$

Como $C U_\alpha = \lambda_\alpha \Sigma^{-1} U_\alpha = \lambda_\alpha W_\alpha$ temos que o último quociente é igual a

$$\frac{W_\alpha(j)\lambda_\alpha}{S_j \sqrt{\lambda_\alpha}} = \frac{\sqrt{\lambda_\alpha} W_\alpha(j)}{S_j}$$

como queríamos demonstrar.

2.4 Decomposição em valores singulares da matriz X' . Relação entre as análises em \mathbb{R}^I e \mathbb{R}^J

Lembremos que os vetores U_α , essenciais para o cálculo das componentes principais são obtidos das relações

$$\begin{aligned} U' C U &= \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_J \end{pmatrix} \\ U' \Sigma^{-1} U &= I \end{aligned}$$

onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_J$.

Temos que

$$C = \sum_{i=1}^I \mu_i x_i x_i' = X' D X \text{ onde } D = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_I \end{pmatrix}$$

ou seja D é a matriz diagonal formada pelo peso dos I indivíduos.

Temos então

$$\begin{aligned} U' X' D X U &= \Lambda \\ U' \Sigma^{-1} U &= I \end{aligned}$$

Estas identidades são as que aparecem na demonstração do Teorema da Decomposição em Valores Singulares tomando $\Sigma_1 = \Sigma^{-1}$, $\Sigma_2 = D^{-1}$ e $Y = X'$.

Se V é a matriz $I \times I$ construída naquele teorema com colunas D^{-1} -ortonormais, temos

$$U' X' V = \left(\begin{array}{ccc|c} \sqrt{\lambda_1} & & & \\ & \ddots & & \\ & & \sqrt{\lambda_J} & \\ \hline & & & 0 \end{array} \right)$$

e para $\lambda_\alpha > 0$, $\alpha = 1, \dots, r = \text{posto}(X)$,

$$V_\alpha = \frac{D X U_\alpha}{\sqrt{\lambda_\alpha}}.$$

Como $X U_\alpha = F_\alpha$ temos também

$$V_\alpha(i) = \frac{\mu_i F_\alpha(i)}{\sqrt{\lambda_\alpha}} \quad \alpha = 1, \dots, r = \text{posto}(X).$$

No caso importante $\Sigma = I$ e $D = I$, ou seja o caso de utilizar a distância euclidiana com todos os indivíduos com o mesmo peso, podemos naturalmente pensar nas colunas como indivíduos e nas linhas como variáveis ou observações. As componentes que no caso anterior eram obtidas através de XU , serão agora obtidas de $X'V$. Pela Decomposição em Valores Singulares temos

$$X'V = U(A^{1/2} : 0) = (\sqrt{\lambda_1} U_1, \dots, \sqrt{\lambda_J} U_J : 0).$$

Portanto $G_\alpha(j) = U_\alpha(j) \sqrt{\lambda_\alpha}$.

Note-se que ainda no caso $\sum_{i=1}^I x_i = 0$ (variáveis centradas) a análise de componentes principais sobre as colunas de X determinará valores $G_\alpha(j)$ que não satisfarão $\sum_{j=1}^J G_\alpha(j) = 0$. Os valores $G_\alpha(j)$ são as componentes na base ortonormal do subespaço ótimo e não na melhor variedade linear.

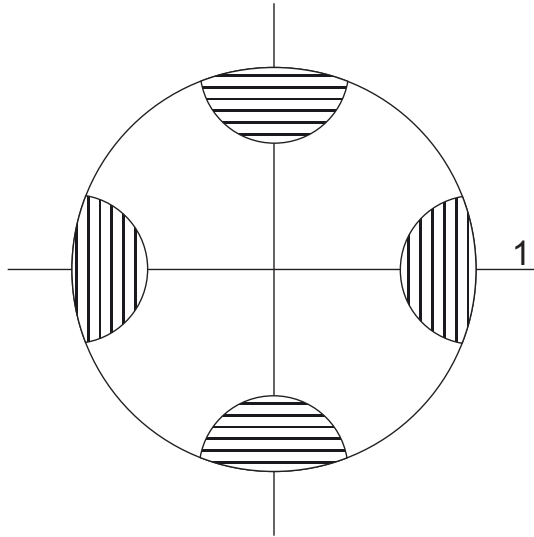
No caso $\Sigma = I$, $D = I$ e com variáveis centradas e padronizadas $\left(S_j^2 = \sum_{i=1}^I (x_{ij})^2 = 1 \right)$ a Proposição 2.3 indica que $\rho(X_j, F_\alpha) = \sqrt{\lambda_\alpha} U_\alpha(j) \sqrt{I}$.

Neste caso temos também que $G_\alpha(j) = U_\alpha(j)\sqrt{\lambda_\alpha}\sqrt{I}$. Ou seja

$$\rho(X_j, F_\alpha) = G_\alpha(j).$$

Portanto a correlação entre a variável e o eixo é igual a coordenada da variável. Este resultado é útil em diversas aplicações para leitura e interpretação das saídas gráficas, contendo simultaneamente a representação dos pontos-indivíduos e pontos-variáveis sobre os dois primeiros eixos fatoriais (duas primeiras componentes principais).

Note-se também que $\sum_{\alpha=1}^p G_\alpha^2(j) \leq 1$. As variáveis para as quais $\sum_{j=1}^p G_\alpha^2(j)$ seja aproximadamente 1 serão variáveis bem ajustadas. Para interpretação eventual dos eixos deveríamos olhar para aquelas variáveis localizadas na parte sombreada do gráfico a seguir.



2.5 Aproximação da matriz X . Taxa de inércia

Vimos na seção anterior que (com $\Sigma_1 = I$ e $\Sigma_2 = I$)

$$U' X' V = (\Lambda^{1/2}; 0)$$

Portanto $X' = U(\Lambda^{1/2}; 0)V'$ e então

$$\begin{aligned} X &= V \begin{pmatrix} \Lambda^{1/2} \\ \dots \\ 0 \end{pmatrix} U' = \left(\sqrt{\lambda_1} V_1, \dots, \sqrt{\lambda_J} V_J \right) \begin{pmatrix} U'_1 \\ \dots \\ \vdots \\ \dots \\ U'_J \end{pmatrix} = \\ &= \sum_{\alpha=1}^J \sqrt{\lambda_\alpha} V_\alpha U'_\alpha. \end{aligned}$$

A matriz X pode ser aproximada pela matriz X^* , obtida somando só os primeiros p termos ou seja

$$X^* = \sum_{\alpha=1}^p \sqrt{\lambda_\alpha} V_\alpha U'_\alpha.$$

A qualidade da aproximação pode ser medida pelo quociente

$$\frac{\sum_{ij} (x_{ij}^*)^2}{\sum_{ij} (x_{ij})^2}$$

onde x_{ij}^* indica o termo (i, j) da matriz X^* . Este quociente é chamado *taxa de inércia* ou *parte da variância explicada pelas primeiras p componentes ou p fatores*.

Uma expressão mais simples pode ser obtida utilizando a seguinte observação. Para uma matriz Z quadrada qualquer $\sum_{ij} (Z_{ij})^2 = \text{Tr}(Z'Z)$.

Para X^* temos

$$X^{*'} X^* = \left(\sum_{\alpha=1}^p \sqrt{\lambda_\alpha} U_\alpha V'_\alpha \right) \left(\sum_{\alpha'=1}^p \sqrt{\lambda_{\alpha'}} V_{\alpha'} U'_{\alpha'} \right) = \sum_{\alpha=1}^p \lambda_\alpha U_\alpha U'_\alpha.$$

Portanto

$$\text{Tr}(X^{*'} X^*) = \sum_{\alpha=1}^p \lambda_{\alpha} \sum_{j=1}^J (U_{\alpha}(j))^2 = \sum_{\alpha=1}^p \lambda_{\alpha}$$

porque $\|U_{\alpha}\|_I^2 = 1$.

Da mesma forma podemos obter $\text{Tr}(X' X) = \sum_{\alpha=1}^J \lambda_{\alpha}$. Portanto a taxa de inércia é igual a

$$\frac{\sum_{\alpha=1}^p \lambda_{\alpha}}{\sum_{\alpha=1}^J \lambda_{\alpha}} = 1 - \frac{\sum_{\alpha=p+1}^J \lambda_{\alpha}}{\sum_{\alpha=1}^J \lambda_{\alpha}} = 1 - \text{erro relativo.}$$

2.6 Aproximação de uma matriz de posto r por uma de posto k , $k \leq r$. Biplot de matrizes

Sejam

$$\{\mu_i\}_{i=1,\dots,n} \quad \text{e} \quad \{\gamma_j\}_{j=1,\dots,m}$$

números positivos e $D_1 = \text{diag}(\mu_1, \dots, \mu_n)$ e $D_2 = \text{diag}(\gamma_1, \dots, \gamma_m)$. Seja Y uma matriz $n \times m$ de posto r . Indicaremos com y_i o vetor correspondente à linha i de Y . Seja $p \leq m$ e consideremos o problema de minimizar

$$\sum_{i=1}^n \mu_i \|y_i - y_i^*\|_{D_2}^2$$

onde os y_i^* pertencem a um subespaço de dimensão p . Sabemos que a solução desse problema obtém-se definindo $C = \sum_{i=1}^n \mu_i y_i y_i'$ ($C = Y' D_1 Y$) e tomando U tal que

$$U' C U = U' Y' D_1 Y U = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$$

$$U' D_2^{-1} U = I$$

Se $W_\alpha = D_2^{-1} U_\alpha$, $[W_1, \dots, W_p]$ é o subespaço solução e os y_i^* são as projeções dos y_i sobre este subespaço. As componentes de y_i^* na base dos $\{W_\alpha\}$ estão dadas por $F_\alpha(i) = y_i' U_\alpha$. Note-se que as equações acima são as que aparecem na decomposição em valores singulares de Y' se tomarmos em \mathbb{R}^m o produto interno determinado por D_2^{-1} e em \mathbb{R}^n o produto interno determinado por D_1^{-1} . Mais precisamente

$$U' Y' V = \left(\begin{array}{ccc|c} \sqrt{\lambda_1} & & & 0 \\ & \ddots & & \\ & & \sqrt{\lambda_J} & \\ \cdots & & & \\ 0 & \cdots & & 0 \end{array} \right)$$

$$U' D_2^{-1} U = I$$

$$V' D_1^{-1} V = I$$

Temos também

$$\sqrt{\lambda_\alpha} U_\alpha = D_2 Y' V_\alpha$$

$$\sqrt{\lambda_\alpha} V_\alpha = D_1 Y U_\alpha$$

Se y^j indica a coluna j de Y e o problema fosse minimizar $\sum_{j=1}^m \gamma_j \|y^j - y^{j*}\|_{D_1}$ o papel anterior de U seria assumido por V . As componentes de y^{j*} seriam $G_{\alpha(j)} = y^{j'} V_\alpha$. Ou de outra forma $G_\alpha = Y' V_\alpha = \sqrt{\lambda_\alpha} D_2^{-1} U_\alpha$.

Proposição 5.1. *Seja Y uma matriz $n \times m$ de posto r . A matriz Z $n \times m$ de posto $p \leq r$ que minimiza*

$$\sum_{i=1}^n \sum_{j=1}^m \mu_i \gamma_j (y_{ij} - Z_{ij})^2$$

é dada por

$$Z = \sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha G'_\alpha.$$

Demonstração. Seja Z_i o vetor correspondente a linha i de Z . Utilizando a notação anterior temos

$$\begin{aligned} & \inf_{\text{posto}(Z) \leq p} \sum_{i=1}^n \sum_{j=1}^m \mu_i \gamma_j (y_{ij} - z_{ij})^2 = \\ & = \inf_{\text{posto}(Z) \leq p} \sum_{i=1}^n \mu_i \|y_i - z_i\|_{D_2}^2 = \\ & = \inf_{\text{subespaços de dimensão} \leq p} \sum_{i=1}^n \mu_i \|y_i - y_i^*\|_{D_2}^2 \end{aligned}$$

onde y_i^* é a projeção de y_i sobre o subespaço. Como aumentando a dimensão do espaço, $\|y_i - y_i^*\|_{D_2}^2$ diminui, temos que a expressão acima é igual a

$$\inf_{\text{subespaço de dimensão } p} \sum_{i=1}^n \mu_i \|y_i - y_i^*\|_{D_2}^2.$$

Temos então um problema de componentes principais. Note-se que neste caso a soma das colunas de Y não é necessariamente igual a zero (ou equivalentemente $\sum_{i=1}^n y_i$ não é zero necessariamente). Calcularemos agora $Z_{kj} = y_k^*(j)$. Sempre utilizando a notação anterior temos

$$\begin{aligned} y_k^* &= \sum_{\alpha=1}^p \langle y_k, W_\alpha \rangle_{D_2} W_\alpha = \sum_{\alpha=1}^p (y_k' D_2 W_\alpha) W_\alpha = \sum_{\alpha=1}^p (y_k' U_\alpha) W_\alpha = \\ &= \sum_{\alpha=1}^p (y_k' U_\alpha) D_2^{-1} U_\alpha. \end{aligned}$$

Como $D_2^{-1} U_\alpha = \frac{G_\alpha}{\sqrt{\lambda_\alpha}}$ e $y_k' U_\alpha = F_\alpha(k)$ temos que a última expressão é igual a

$$\sum_{\alpha=1}^p F_\alpha(k) \frac{G_\alpha}{\sqrt{\lambda_\alpha}}.$$

Portanto $Z_{kj} = y_k^*(j) = \sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha(k) G_\alpha(j)$ ou seja

$$Z = \sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha G_\alpha'$$

como queríamos demonstrar.

Nota: Se $\Lambda_p = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix}$, $F_i^{(p)} = \begin{pmatrix} F_1(i) \\ \vdots \\ F_p(i) \end{pmatrix}$ e

$G_j^{(p)} = \begin{pmatrix} G_1(j) \\ \vdots \\ G_p(j) \end{pmatrix}$ temos que

$$Z_{ij} = \langle F_i^{(p)}, G_j^{(p)} \rangle_{\Lambda_p^{-1/2}} = F_i^{(p)1} \Lambda_p^{-1/2} G_j^{(p)}.$$

De outra forma

$$Z = \begin{pmatrix} F_1^{(p)'} \\ \dots \\ \vdots \\ \dots \\ F_n^{(p)'} \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_p} \end{pmatrix} \left(G_1^{(p)} \mid \dots \mid G_m^{(p)} \right)$$

Corolário 5.1. *Se $p = r$ temos*

$$Y = \sum_{\alpha=1}^r \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha G'_\alpha.$$

Note-se que a matriz $\sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha G'_\alpha$ é igual a

$$\begin{aligned} (F_1 \mid \dots \mid F_p) \begin{pmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_p} \end{pmatrix} \begin{pmatrix} G'_1 \\ \dots \\ \vdots \\ \dots \\ G'_p \end{pmatrix} &= \\ &= \left(\frac{F_1}{\sqrt{\lambda_1}} \mid \dots \mid \right) \begin{pmatrix} G'_1 \\ \dots \\ \vdots \\ \dots \\ G'_p \end{pmatrix} \end{aligned}$$

ou seja é da forma HG' onde H é uma matriz $n \times p$ e G uma matriz $m \times p$.

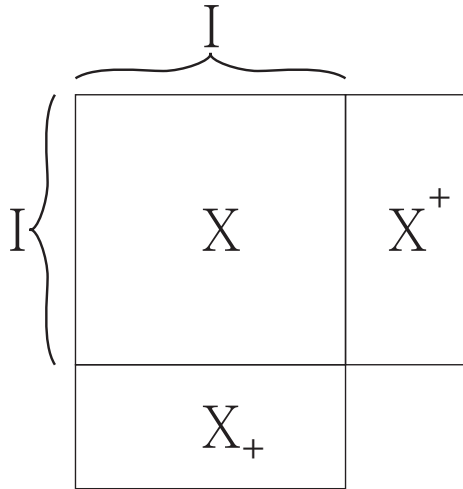
Portanto o problema (problema do Biplot; ver Gabriel (1971)) de achar duas matrizes H e G tais que

$$\sum_{i=1}^n \sum_{j=1}^m \mu_i \gamma_j (y_{ij} - h'_i g_j)^2$$

seja mínimo (h_i (resp. g_j) indica a linha i (resp. j) de H (resp. G)) é também solucionado pela matriz $\sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha G'_\alpha$. Note-se que não há unicidade nas matrizes H e G .

2.7 Indivíduos e variáveis suplementares

É frequente na prática ter informação complementar sobre os indivíduos (observações de variáveis adicionais sobre os I indivíduos) e também observações sobre novos indivíduos. Estas duas situações aparecem representadas na figura a seguir.



A matriz X_+ corresponde a novos indivíduos e a X^+ a novas variáveis sobre os indivíduos originais. Indicaremos com U_α , F_α , λ_α os vetores e

autovalores obtidos da análise de componentes principais sobre a matriz X . Uma análise sobre a matriz $(X : X_+)$ determinaria coordenadas para os novos indivíduos. Se o peso destes indivíduos é relativamente pequeno as suas coordenadas serão aproximadamente dadas por

$$F_\alpha(i) = \sum_{j=1}^J x_{+ij} U_\alpha(j).$$

Esta aproximação evita naturalmente repetir a análise com toda a matriz.

Da mesma forma

$$G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^I \mu_i F_\alpha(i) x_{ij}^+$$

indicará as coordenadas aproximadas das novas variáveis. Os valores exatos seriam os obtidos fazendo a análise sobre a matriz $(X : X^+)$.

Naturalmente os valores x_{+ij} e x_{ij}^+ utilizados para calcular $F_\alpha(i)$ e $G_\alpha(j)$ deverão ser valores centrados (ou centrados e normalizados) se sobre os valores originais foram executadas estas transformações.

2.7 Exemplos.

Os exemplos deste capítulo utilizam a função `prcomp()` e o pacote `BiplotGUI` do R. O primeiro exemplo utiliza dados públicos disponíveis (ver Referência no Exemplo 1) . Tem vários pacotes e funções do R que podem ser utilizados. Por exemplo: `princomp()`, `prcomp()`, e `PCA` do pacote `FactoMineR`.

Exemplo 1. Consumo de Proteínas em Europa.

Os dados indicam o consumo de 9 proteínas (em % sobre o total de proteína) para cada um de 25 países Europeus.

Os dados podem ser encontrados em Hand e outros. (1994).

Os dados são os seguintes:

Country	RedMeat	WhiteMeat	Eggs	Milk	Fish	Cereals	Starch	Nuts	Fr.Veg
Albania	10,1	1,4	0,5	8,9	0,2	42,3	0,6	5,5	1,7
Austria	8,9	14	4,3	19,9	2,1	28	3,6	1,3	4,3
Belgium	13,5	9,3	4,1	17,5	4,5	26,6	5,7	2,1	4
Bulgaria	7,8	6	1,6	8,3	1,2	56,7	1,1	3,7	4,2
Czechoslovakia	9,7	11,4	2,8	12,5	2	34,3	5	1,1	4
Denmark	10,6	10,8	3,7	25	9,9	21,9	4,8	0,7	2,4
E Germany	8,4	11,6	3,7	11,1	5,4	24,6	6,5	0,8	3,6
Finland	9,5	4,9	2,7	33,7	5,8	26,3	5,1	1	1,4
France	18	9,9	3,3	19,5	5,7	28,1	4,8	2,4	6,5
Greece	10,2	3	2,8	17,6	5,9	41,7	2,2	7,8	6,5
Hungary	5,3	12,4	2,9	9,7	0,3	40,1	4	5,4	4,2
Ireland	13,9	10	4,7	25,8	2,2	24	6,2	1,6	2,9
Italy	9	5,1	2,9	13,7	3,4	36,8	2,1	4,3	6,7
Netherlands	9,5	13,6	3,6	23,4	2,5	22,4	4,2	1,8	3,7
Norway	9,4	4,7	2,7	23,3	9,7	23	4,6	1,6	2,7
Poland	6,9	10,2	2,7	19,3	3	36,1	5,9	2	6,6
Portugal	6,2	3,7	1,1	4,9	14,2	27	5,9	4,7	7,9
Romania	6,2	6,3	1,5	11,1	1	49,6	3,1	5,3	2,8
Spain	7,1	3,4	3,1	8,6	7	29,2	5,7	5,9	7,2
Sweden	9,9	7,8	3,5	24,7	7,5	19,5	3,7	1,4	2
Switzerland	13,1	10,1	3,1	23,8	2,3	25,6	2,8	2,4	4,9
UK	17,4	5,7	4,7	20,6	4,3	24,3	4,7	3,4	3,3
USSR	9,3	4,6	2,1	16,6	3	43,6	6,4	3,4	2,9
W Germany	11,4	12,5	4,1	18,8	3,4	18,6	5,2	1,5	3,8
Yugoslavia	4,4	5	1,2	9,5	0,6	55,9	3	5,7	3,2

Em primeiro lugar algumas informações básicas sobre as variáveis. A Figura 2 mostra as médias e variâncias das variáveis. Dada a variação entre estas últimas os dados serão normalizados antes da análise, ou seja usaremos as variáveis padronizadas (média zero e variância 1).

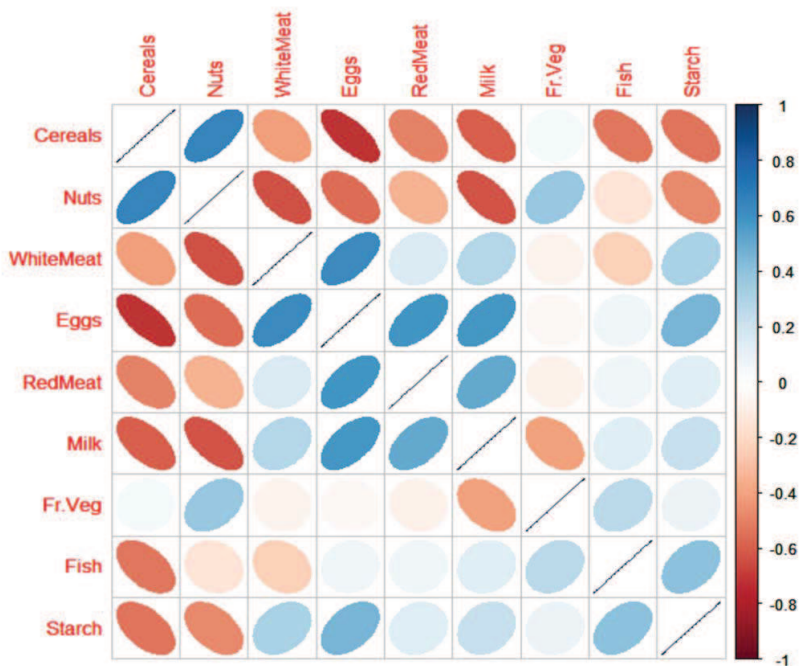
Figura 2. Médias e Varianças das variáveis.

Variavel	Media
RedMeat	9,828
WhiteMeat	7,896
Eggs	2,936
Milk	17,112
Fish	4,284
Cereals	32,248
Starch	4,276
Nuts	3,072
Fr.Veg	4,136

Variavel	Variancia
RedMeat	3,3470783
WhiteMeat	3,6940809
Eggs	1,1176165
Milk	7,1054158
Fish	3,4025334
Cereals	10,974786
Starch	1,6340849
Nuts	1,9856821
Fr.Veg	1,8039032

A Figura 3 mostra a correlação entre as variáveis.

Figura 3



As cores, inclinação e forma das elipses dá uma ideia gráfica rápida das correlações.

Azuis: inclinação positiva, correlação positiva. Marroes: inclinação negativa, correlação negativa.

Forma próxima de um círculo indica uma correlação próxima de zero.

A escala de cores à direita do gráfico dá os valores das correlações.

Temos uma correlação positiva entre as variáveis (consumo de) White-Meat, Eggs, Red Meat e Milk.

O mesmo acontece entre Cereais e Nuts e com menor intensidade entre Starch e Fish.

A Figura 4 indica as cargas das componentes principais (loadings) sobre as 9 componentes.

Note-se que a Primeira Componente (a mais importante) separa claramente as carnes vermelha, branca, ovos e leite, de nozes e cereais. A segunda componente está determinada pelo consumo de peixe e vegetais (e em menor grau amidos).

Figura 4. Cargas da componentes (loadings).

Variavel	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
RedMeat	-0,303	-0,056	-0,298	-0,646	0,322	-0,460	0,150	-0,020	0,246
WhiteMeat	-0,311	-0,237	0,624	0,037	-0,300	-0,121	-0,020	-0,028	0,592
Eggs	-0,427	-0,035	0,182	-0,313	0,079	0,361	-0,443	-0,491	-0,333
Milk	-0,378	-0,185	-0,386	0,003	-0,200	0,618	0,462	0,081	0,178
Fish	-0,136	0,647	-0,321	0,216	-0,290	-0,137	-0,106	-0,449	0,313
Cereals	0,438	-0,233	0,096	0,006	0,238	0,081	0,405	-0,703	0,152
Starch	-0,297	0,353	0,243	0,337	0,736	0,148	0,153	0,115	0,122
Nuts	0,420	0,143	-0,054	-0,330	0,151	0,447	-0,407	0,184	0,518
Fr.Veg	0,110	0,536	0,408	-0,462	-0,234	0,119	0,450	0,092	-0,203

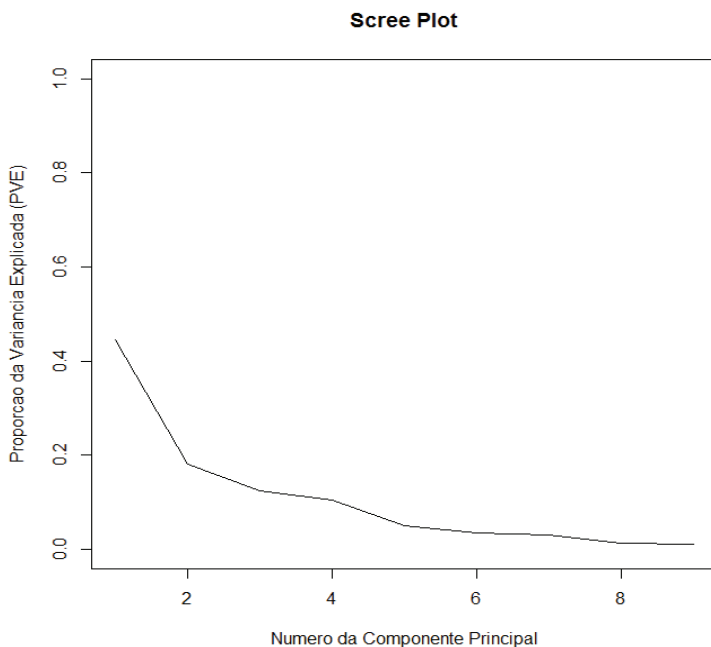
Utilizando as cargas e as variáveis originais (normalizadas) calculamos a posição de cada linha da matriz de dados (pais) sobre cada uma das componentes. Os valores estão indicados na Figura 5.

Figura 5. Componentes (scores).

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Albania	3,485	-1,630	-1,761	-0,230	0,023	-1,034	-0,472	0,762	-0,103
Austria	-1,423	-1,041	1,338	-0,168	-0,933	0,218	-0,181	-0,251	-0,217
Belgium	-1,622	0,159	0,217	-0,521	0,755	-0,290	-0,196	-0,203	-0,033
Bulgaria	3,134	-1,301	0,151	-0,214	-0,485	-0,696	0,465	-0,808	-0,300
Czechoslov	-0,370	-0,603	1,196	0,464	0,257	-0,823	0,315	0,012	-0,149
Denmark	-2,365	0,285	-0,752	0,967	-0,752	-0,170	-0,226	-0,621	0,480
E Germany	-1,422	0,450	1,303	1,136	0,423	-0,648	-0,555	-0,163	-0,260
Finland	-1,564	-0,596	-2,050	1,415	0,037	0,834	0,726	0,226	-0,133
France	-1,488	0,785	0,002	-1,957	0,250	-0,899	0,946	-0,022	0,544
Greece	2,240	1,001	-0,883	-1,794	-0,405	1,144	-0,147	-0,306	0,388
Hungary	1,457	-0,816	1,914	0,217	-0,041	0,539	-0,768	0,146	0,537
Ireland	-2,663	-0,764	-0,020	-0,435	1,014	0,482	-0,029	0,023	-0,079
Italy	1,535	0,399	0,126	-1,222	-0,804	0,214	0,150	-0,080	-0,732
Netherlands	-1,641	-0,912	0,766	0,126	-0,761	0,298	-0,062	0,460	0,261
Norway	-0,975	0,822	-1,704	1,138	-0,415	-0,056	-0,043	-0,107	-0,147
Poland	-0,122	0,532	1,475	0,458	-0,023	0,588	1,261	0,192	-0,221
Portugal	1,706	4,289	0,044	0,894	-0,385	-0,697	0,046	0,205	0,263
Romania	2,757	-1,119	0,070	0,615	0,317	0,131	-0,133	-0,027	0,338
Spain	1,312	2,554	0,515	-0,359	0,516	0,669	-0,597	0,235	-0,477
Sweden	-1,634	-0,207	-1,280	0,734	-0,820	0,044	-0,541	-0,072	-0,108
Switzerland	-0,912	-0,751	-0,154	-1,170	-0,831	-0,090	0,512	0,529	0,067
UK	-1,735	-0,094	-1,153	-1,734	1,084	-0,097	-0,651	-0,239	-0,132
USSR	0,783	-0,111	-0,370	0,928	1,670	0,185	0,574	-0,052	0,092
W German	-2,094	-0,294	0,804	-0,109	-0,068	-0,201	-0,457	0,357	-0,025
Yugoslavia	3,623	-1,038	0,206	0,822	0,378	0,354	0,061	-0,193	0,149

Como vimos o uso do método de Componentes Principais transforma as variáveis originais em variáveis ortogonais e de forma que as primeiras componentes capturam a maior proporção da variância a ser explicada. Para explicar totalmente esta variação deveríamos usar todas as 9 componentes. Em geral se a matriz de dados é $n \times p$ teríamos $\min(n - 1, p)$ componentes principais. Não estamos interessados em todas elas. Só um número pequeno delas para poder analisar e interpretar os dados. Podemos explicar uma grande parte dela usando somente umas poucas componentes (2 ou 3 por exemplo). A proporção explicada por cada componente relativa ao total é usualmente resumida em um grafico chamado de Scree Plot.

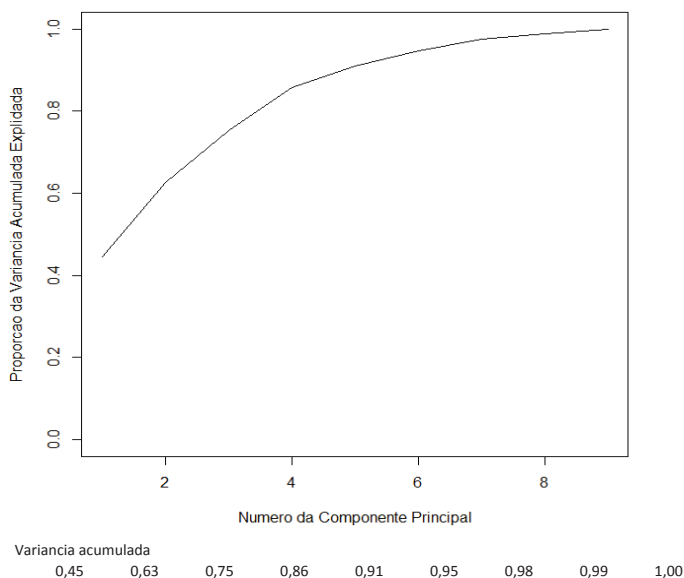
Figura 6. Screeplot.



RedMeat	WhiteMea	Eggs	Milk	Fish	Cereals	Starch	Nuts	Fr.Veg
0,44516	0,181667	0,125324	0,106074	0,051538	0,036126	0,030178	0,012921	0,011012

O Gráfico da Figura 6 mostra que a primeira componente representa mais de 40% da variação total e o Gráfico seguinte da Figura 7 mostra que os dois primeiros representam mais de 60% do total.

Figura 7. Proporção da variância explicada.



Não existe um método formal de determinar o número de componentes. Os procedimentos são relativamente ad-hoc e influenciados pela área de aplicação onde são utilizados. Por exemplo em dados provenientes de Ciências Sociais é comum que os dados contemham muito “ruído” (erro aleatório). Nestes casos curvas como as da Figura 7 crescem lentamente. Um procedimento ad hoc frequentemente utilizado é um sugerido por Cattell (1960) que consiste em selecionar componentes até atingir o “cotovelo” da curva do Scree Plot. Ou seja uma situação na qual a curva já passou pela sua principal descida e está mais ou menos estabilizada. Em nosso caso corresponderia a escolher 2 ou 3 dimensões. Ficaremos com 2 dimensões já que permite uma apresentação gráfica mais simples e direta.

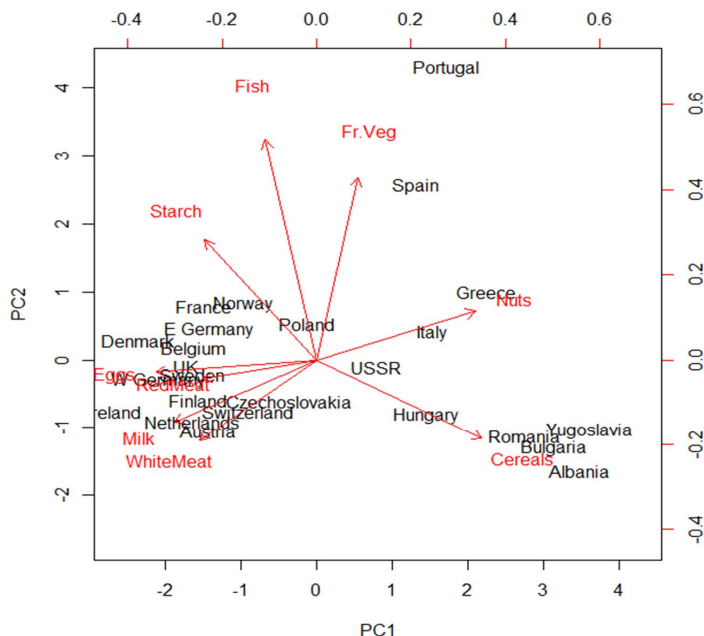
Como sugerido pelos resultados em 2.6, uma representação conveniente é obtida através de um Biplot.

Neste caso as linhas da matriz de dados são representadas como pontos (coordenadas sobre as duas primeiras componentes principais) e as variáveis como direções. Direções aproximadamente horizontais são altamente correlacionadas com a primeira componente. Direções aproxi-

madamente verticais seriam mais correlacionadas com a segunda componente principal.

A Figura 8 mostra o gráfico. Vemos por exemplo que Espanha e Portugal são um grupo de países diferenciado assim como Romênia, Bulgária, Albânia e Yugoslavia.

Figura 8. Biplot tradicional.



Espanha e Portugal mostram que são consumidores de mais peixe, vegetais e amidos em menor grau que outros países. O mesmo acontece com o consumo de cereais e nozes em Albânia, Bulgária, Romênia e Yugoslavia.

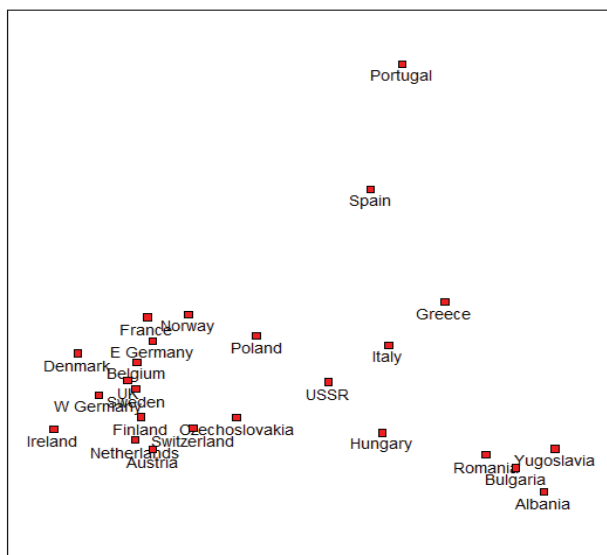
O gráfico sugere também uma segmentação entre os países em função do seu consumo de Proteínas.

Uma versão muito interessante do Biplot pode ser consultada em Gower e outros (2011). Nesta versão os vetores correspondentes aos atributos são tomados como direções e suas escalas mudadas para conseguir uma propriedade interessante: se projetamos perpendicularmente um ponto linha sobre uma direção o valor de sua projeção corresponde aproximadamente ao valor do atributo original. Ou seja poderíamos usar o gráfico

para ter rapidamente uma ideia do valor das variáveis. Naturalmente as interpretações de menor ângulo entre direções indicar maior correlação entre as variáveis e a de direções próximas aos eixos indicar loadings maiores sobre os mesmos são mantidas.

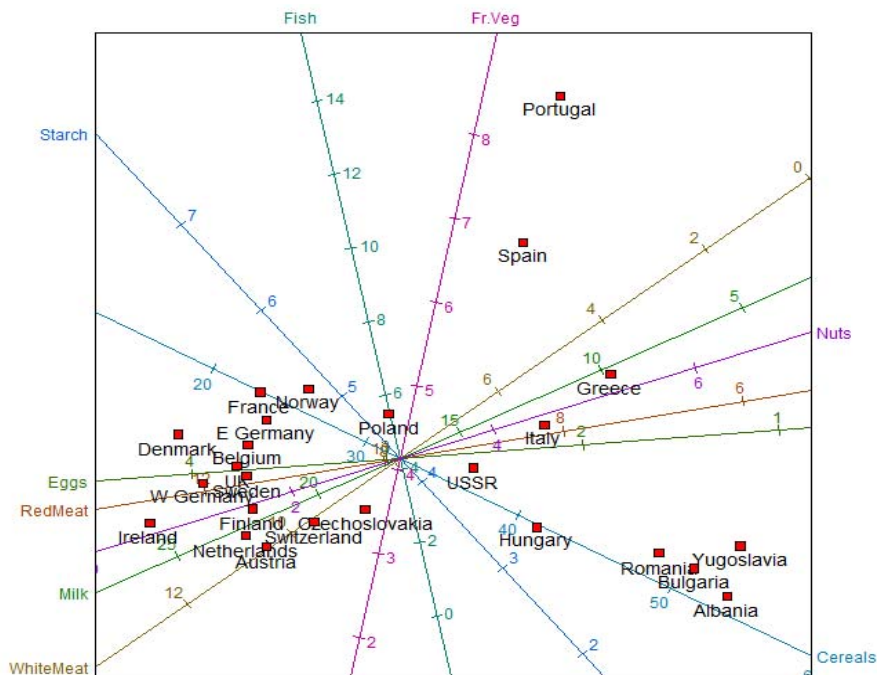
A Figura 9 mostra o gráfico só dos pontos sem as variáveis (e sem os valores das coordenadas que) para ter uma visão “limpa” da posição dos países. Vemos de novo que Espanha e Portugal são um grupo de países diferenciado assim como Romênia, Bulgária, Albânia e Yugoslavia.

Figura 9. Mapa dos países sobre as duas primeiras componentes principais.



A Figura 10 mostra o resultado completo com países (pontos) e variáveis (direções).

Figura 10. Mapa dos países sobre as duas componentes principais. Inclui as direções dos atributos utilizados (Biplot).



Exemplo 2. Dados sobre características de Vinho.

Para este exemplo só vamos indicar os principais quadros. Os dados correspondem a 21 marcas de vinho francês sobre as quais 28 variáveis foram determinadas. Duas destas variáveis são categóricas (fatores): Label (3 níveis: Saumur , Bourgueuil e Chinon) e Soil (4 níveis: Reference, Env1, Env2, Env4) indicando Regiões da França e tipos de solo.

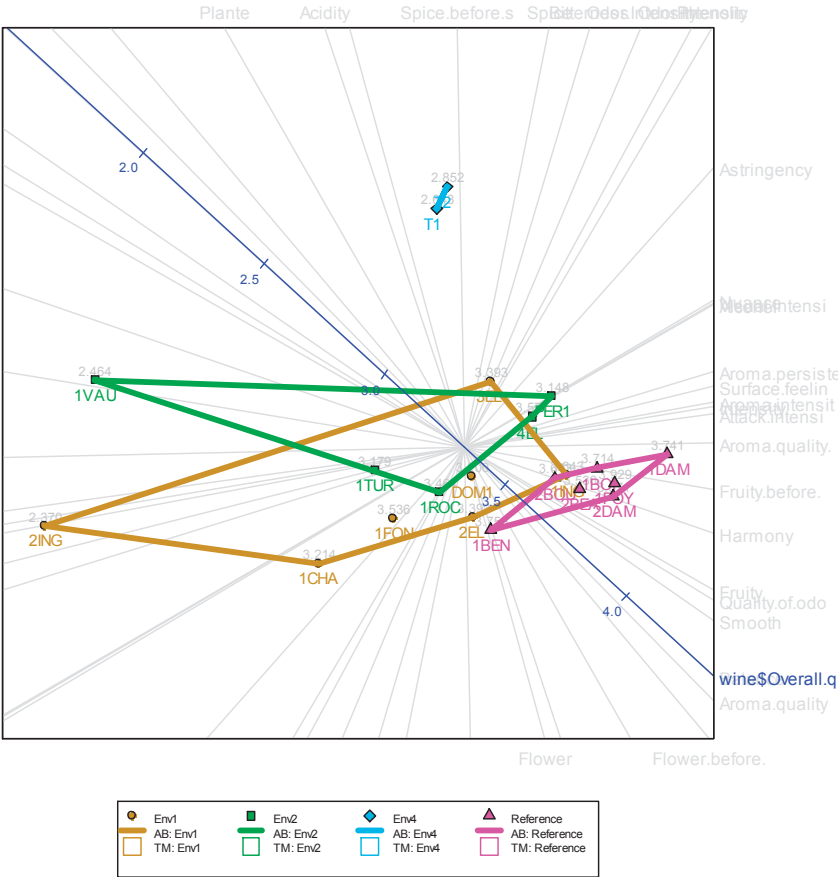
As 28 variáveis numéricas e as 21 Marcas (códigos das marcas) são:

Figura 1

Variaveis		Marcas	
1	Odor.intensity.before.shaking	1	2EL
2	Aroma.quality.before.shaking	2	1CHA
3	Fruity.before.shaking	3	1FON
4	Flower.before.shaking	4	1VAU
5	Spice.before.shaking	5	1DAM
6	Visual.intensity	6	2BOU
7	Nuance	7	1BOI
8	Surface.feeling	8	3EL
9	Odor.intensity	9	DOM1
10	Quality.of.odour	10	1TUR
11	Fruity	11	4EL
12	Flower	12	PER1
13	Spice	13	2DAM
14	Plante	14	1POY
15	Phenolic	15	1ING
16	Aroma.intensity	16	1BEN
17	Aroma.persistency	17	2BEA
18	Aroma.quality	18	1ROC
19	Attack.intensity	19	2ING
20	Acidity	20	T1
21	Astringency	21	T2
22	Alcohol		
23	Balance		
24	Smooth		
25	Bitterness		
26	Intensity		
27	Harmony		
28	Overall.quality		

A Figura 1 indica o posicionamnto das marcas. Só uma variável foi enfatizada no gráfico: a Qualidade Total, para não poluir o grafico com

Figura 3. Posicionamento das Marcas adicionando cápsulas convexas dos grupos de solos. O Solo de Referência é o que tem maior qualidade como um todo com 7 marcas no mesmo.



Capítulo 3

Análise de Correspondência

A técnica de análise de correspondência foi desenvolvida para o tratamento simultâneo das linhas e colunas de uma tabela bi-dimensional, onde as entradas são números não negativos. Como exemplos de tais tabelas, podemos citar: 1.^a) a tabela de contingência onde as linhas e as colunas representam variáveis categóricas e cada entrada o número de observações das variáveis cruzadas, e 2.^a) a tabela de incidência onde as linhas representam, por exemplo, indivíduos e as colunas representam variáveis categóricas, por exemplo, nível de instrução do pai, nível de instrução da mãe, etc.... Cada variável dessas é dividida em classes mutuamente disjuntas, por exemplo, o nível de instrução do pai pode ser superior, colegial, primário completo, etc.... Cada entrada é 1 ou 0, conforme o indivíduo possui ou não uma determinada característica. Outros exemplos serão vistos mais tarde.

Como na técnica de componentes principais, a análise de correspondência pode ser apresentada de diferentes maneiras. (Ver Benzecri (1973), Lebart & Fenelon (1971), Lebart, Morineau & Tabard (1977) e Hill (1974)).

Começaremos a exposição tomando o ponto de vista de Análise de Dados, que vai ser enfatizado, e depois daremos algumas formulações equivalentes. O objetivo é a representação simultânea das linhas e colunas de uma tabela em um espaço de dimensão pequena onde possamos mais facilmente interpretar as relações entre as linhas, entre as colunas e entre as linhas e as colunas. É claro que essa representação simultânea deve ser obtida, como na análise de componentes principais, conservando na

medida do possível a “informação” contida nos dados originais. De fato, o que faremos é transformar os dados de duas maneiras diferentes e sobre cada transformação aplicar uma análise de componentes principais apropriada.

3.1 Notação e descrição do problema

A análise de correspondência é, tipicamente, utilizada para analisar uma tabela de contingência. Essa tabela de contingência pode ter sido obtida, por exemplo, de um grupo de n indivíduos classificados em I classes disjuntas A_1, \dots, A_I de um critério A (por exemplo, município de residência) e em J classes disjuntas B_1, \dots, B_J de um critério B (por exemplo, causa de morte ou nível de renda). Essa tabela pode ser representada da seguinte maneira:

Critério $A \setminus$ Critério B	B_1	\dots	B_j	\dots	B_J	marginal de A
A_1	n_{11}		n_{1j}		n_{1J}	$n_{1\cdot}$
\vdots						
A_i	n_{i1}		n_{ij}		n_{iJ}	$n_{i\cdot}$
\vdots						
A_I	n_{I1}		n_{IJ}		n_{IJ}	$n_{I\cdot}$
marginal de B	$n_{\cdot 1}$		$n_{\cdot j}$		$n_{\cdot J}$	n

onde n_{ij} = número de indivíduos classificados simultaneamente em A_i e B_j ,

$$n_{i\cdot} = \sum_{j=1}^J n_{ij} = \text{número de indivíduos classificados em } A_i,$$

$$n_{\cdot j} = \sum_{i=1}^I n_{ij} = \text{número de indivíduos classificados em } B_j.$$

É claro que temos

$$n = \sum_{i=1}^I \sum_{j=1}^J n_{ij} = \sum_{i=1}^I n_{i\cdot} = \sum_{j=1}^J n_{\cdot j}.$$

Mais geralmente, a análise de correspondência é utilizada numa tabela onde as classes A_i e/ou as classes B_j não precisam ser mutuamente disjuntas, n_{ij} é a entrada correspondente a classe A_i e a classe B_j , com a restrição $n_{ij} \geq 0$, $n_{i\cdot}$, $n_{\cdot j}$ e n são como antes.

Chamemos a essa tabela de N , ie., $N = (n_{ij})_{\substack{1 \leq i \leq I \\ 1 \leq j \leq J}}$.

Mas para facilitar a apresentação da técnica, podemos pensar numa tabela de contingência. A tabela N pode ser normalizada considerando-se a tabela $P = \frac{1}{n} N$, que no caso da tabela de contingência é a tabela de frequência.

Se $p = (p_{ij})_{\substack{1 \leq i \leq I \\ 1 \leq j \leq J}}$, temos então

$$\begin{aligned} p_{ij} &= \frac{n_{ij}}{n}, \\ p_{i\cdot} &= \sum_{j=1}^J p_{ij} = \frac{n_{i\cdot}}{n}, \\ p_{\cdot j} &= \sum_{i=1}^I p_{ij} = \frac{n_{\cdot j}}{n} \quad \text{e} \\ 1 &= \sum_{i=1}^I \sum_{j=1}^J p_{ij} = \sum_{i=1}^I p_{i\cdot} = \sum_{j=1}^J p_{\cdot j}. \end{aligned}$$

Para compararmos, por exemplo, dois municípios em relação ao nível de renda de seus habitantes, não é interessante trabalhar com o número total de habitantes, pois em geral, os municípios têm número de habitantes distintos, podendo ser um município muito maior que o outro. O que é interessante é comparar os “perfis” dos municípios, isto é, comparar as frequências relativas de cada classe de renda nos municípios. Então, se por exemplo, essas frequências relativas são aproximadamente iguais para todas as classes de rendas, esses dois municípios seriam similares.

Representamos então cada classe A_i pelo vetor de frequências condicionais p_i^B no espaço \mathbb{R}^J , isto é,

$$(p_i^B)' = \left(\frac{p_{i1}}{p_{i\cdot}}, \dots, \frac{p_{iJ}}{p_{i\cdot}} \right) = \left(\frac{n_{i1}}{n_{i\cdot}}, \dots, \frac{n_{iJ}}{n_{i\cdot}} \right).$$

Analogamente, representamos cada classe B_j pelo vetor de frequências condicionais p_j^A no espaço \mathbb{R}^I , isto é,

$$(p_j^A)' = \left(\frac{p_{1j}}{p_{\cdot j}}, \dots, \frac{p_{Ij}}{p_{\cdot j}} \right) = \left(\frac{n_{1j}}{n_{\cdot j}}, \dots, \frac{n_{Ij}}{n_{\cdot j}} \right).$$

Ao vetor p_i^B (p_j^A) chamamos de perfil da classe A_i (B_j). Observamos que

$$\sum_{j=1}^J \frac{p_{ij}}{p_{i\cdot}} = 1 \quad \text{e} \quad \sum_{i=1}^I \frac{p_{ij}}{p_{\cdot j}} = 1.$$

Temos então para as classes A_1, \dots, A_I , representações comparáveis no espaço \mathbb{R}^J e para as classes B_1, \dots, B_J representações comparáveis no espaço \mathbb{R}^I .

Para podermos comparar as classes A_1, \dots, A_I , precisamos agora introduzir uma distância no espaço \mathbb{R}^J .

Procuremos então, uma distância quadrática

$$d^2(A_i, A_{i'}) = \|p_i^B - p_{i'}^B\|_{\Sigma^B}^2 = \sum_{j=1}^J \lambda_j \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2$$

que tenha a seguinte propriedade de invariância:

“Se agregamos duas classes B_j e $B_{j'}$ de mesmo perfil, isto é, tais que $p_j^A = p_{j'}^A$, não alteramos a distância entre duas classes quaisquer A_i e $A_{i'}$, isto é, $d^2(A_i, A_{i'})$ permanece inalterada”.

Essa é uma propriedade desejável, pois se duas classes de renda vizinhas tem aproximadamente o mesmo perfil, podemos então agrupar essas duas classes de renda sem alterar a similaridade entre os municípios. Esse fato garante uma certa invariabilidade dos resultados a divisão em classes de renda, que é necessariamente arbitrária.

É claro que λ_j deve depender de B_j e como estamos interessados em poder agregar duas classes B_j e $B_{j'}$ de mesmo perfil, é natural tomarmos λ_j como uma função contínua λ de $p_{\cdot j}$, isto é, $\lambda_j = \lambda(p_{\cdot j})$.

Proposição 3.1. *A distância entre as classes A_i e $A_{i'}$ que satisfaz essa propriedade de invariância, é, a menos de uma homotetia, dada por:*

$$d^2(A_i, A_{i'}) = \|p_i^B - p_{i'}^B\|_{\Sigma^B}^2 = \sum_{j=1}^J \frac{1}{p_{\cdot j}} \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2,$$

isto é, é a distância quadrática em \mathbb{R}^J , definida pela matriz $\Sigma^B = \text{diag} \left(\frac{1}{p_{\cdot 1}}, \dots, \frac{1}{p_{\cdot J}} \right)$.

Demonstração. Sejam j_1 e j_2 dois índices tais que $p_{j_1}^A = p_{j_2}^A$, isto é, $\frac{p_{ij_1}}{p_{\cdot j_1}} = \frac{p_{ij_2}}{p_{\cdot j_2}}$ para todo $i = 1, \dots, I$.

Podemos então agrupar as classes B_{j_1} e B_{j_2} numa nova classe B_{j_0} . Temos então

$$\begin{aligned} n_{ij_0} &= n_{ij_1} + n_{ij_2} \quad \text{para todo } i = 1, \dots, I \quad \text{e} \\ n_{\cdot j_0} &= n_{\cdot j_1} + n_{\cdot j_2}. \end{aligned}$$

Como

$$\frac{n_{ij_1}}{n_{\cdot j_1}} = \frac{n_{ij_2}}{n_{\cdot j_2}} = \frac{n_{ij_1} + n_{ij_2}}{n_{\cdot j_1} + n_{\cdot j_2}} = \frac{n_{ij_0}}{n_{\cdot j_0}},$$

temos

$$(*) \quad \frac{p_{ij_1}}{p_{\cdot j_1}} = \frac{p_{ij_2}}{p_{\cdot j_2}} = \frac{p_{ij_0}}{p_{\cdot j_0}} \quad \text{para todo } i = 1, \dots, I$$

Podemos escrever então:

$$\begin{aligned} d^2(A_i, A_{i'}) &= \sum_{j=1}^J \lambda(p_{\cdot j}) \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2 = \\ &= \lambda(p_{\cdot j_1}) \left(\frac{p_{ij_1}}{p_{i\cdot}} - \frac{p_{i'j_1}}{p_{i'\cdot}} \right)^2 + \lambda(p_{\cdot j_2}) \left(\frac{p_{ij_2}}{p_{i\cdot}} - \frac{p_{i'j_2}}{p_{i'\cdot}} \right)^2 + \\ &\quad + \sum_{\substack{j \neq j_1 \\ j \neq j_2}} \lambda(p_{\cdot j}) \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2. \end{aligned}$$

Para que a propriedade de invariância seja satisfeita, é necessário e suficiente que:

$$\begin{aligned} \lambda(p_{\cdot j_1}) \left(\frac{p_{ij_1}}{p_{i\cdot}} - \frac{p_{i'j_1}}{p_{i'\cdot}} \right)^2 + \lambda(p_{\cdot j_2}) \left(\frac{p_{ij_2}}{p_{i\cdot}} - \frac{p_{i'j_2}}{p_{i'\cdot}} \right)^2 &= \\ &= \lambda(p_{\cdot j_0}) \left(\frac{p_{ij_0}}{p_{i\cdot}} - \frac{p_{i'j_0}}{p_{i'\cdot}} \right)^2, \end{aligned}$$

ou seja que

$$\lambda(p_{\cdot j_1}) p_{\cdot j_1}^2 \left(\frac{p_{ij_1}}{p_{i\cdot} p_{\cdot j_1}} - \frac{p_{i'j_1}}{p_{i'\cdot} p_{\cdot j_1}} \right)^2 + \\ + \lambda(p_{\cdot j_2}) p_{\cdot j_2}^2 \left(\frac{p_{ij_2}}{p_{i\cdot} p_{\cdot j_2}} - \frac{p_{i'j_2}}{p_{i'\cdot} p_{\cdot j_2}} \right)^2 = \lambda(p_{\cdot j_0}) p_{\cdot j_0}^2 \left(\frac{p_{ij_0}}{p_{i\cdot} p_{\cdot j_0}} - \frac{p_{i'j_0}}{p_{i'\cdot} p_{\cdot j_0}} \right)^2.$$

Por (página anterior), as expressões entre parêntesis são iguais. Logo temos a igualdade:

$$\lambda(p_{\cdot j_1}) p_{\cdot j_1}^2 + \lambda(p_{\cdot j_2}) p_{\cdot j_2}^2 = \lambda(p_{\cdot j_0}) p_{\cdot j_0}^2$$

para quaisquer $p_{\cdot j_1} > 0$, $p_{\cdot j_2} > 0$, com $p_{\cdot j_0} = p_{\cdot j_1} + p_{\cdot j_2}$.

O resultado segue então do seguinte lema.

Lema. *Se λ é uma função contínua definida nos reais positivos, satisfazendo a relação*

$$\lambda(a)^2 a^2 + \lambda(b) b^2 = \lambda(a+b)(a+b)^2,$$

então $\lambda(a) = \frac{c}{a}$ onde $c = \lambda(1)$ é uma constante > 0 .

Demonstração. Mostra-se recursivamente que $\lambda(na) = \frac{1}{n} \lambda(a)$ para todo $a > 0$, n inteiro positivo. Segue então que $\lambda(n) = \frac{1}{n} \lambda(1)$ para todo inteiro positivo. Mostra-se depois que $\lambda\left(\frac{m}{n}\right) = \frac{1}{(m/n)} \lambda(1)$ para quaisquer inteiros positivos m e n . E então, segue por continuidade que $\lambda(a) = \frac{1}{a} \lambda(1)$ para todo $a > 0$.

Analogamente, a distância entre duas classes B_j e $B_{j'}$, é dada por

$$d^2(B_j, B_{j'}) = \|p_j^A - p_{j'}^A\|_{\Sigma^A}^2 = \sum_{i=1}^I \frac{1}{p_{i\cdot}} \left(\frac{p_{ij}}{p_{\cdot j}} - \frac{p_{ij'}}{p_{\cdot j'}} \right)^2$$

onde $\Sigma^A = \text{diag} \left(\frac{1}{p_{1\cdot}}, \dots, \frac{1}{p_{I\cdot}} \right)$.

Observação 1. Se as classes B_{j_1} e B_{j_2} têm o mesmo perfil e são agrupadas na classe B_{j_0} , então

$$d^2(B_j, B_{j_0}) = d^2(B_j, B_{j_1}) = d^2(B_j, B_{j_2}) \text{ para todo } j = 1, \dots, J.$$

Esse fato segue diretamente de (*).

Observação 2. Os valores $\lambda_j = \frac{1}{p_{\cdot j}}$ em

$$d^2(A_i, A_{i'}) = \sum_{j=1}^J \frac{1}{p_{\cdot j}} \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2$$

têm a propriedade de atenuar as disparidades causadas pelos diferentes contingentes $n_{\cdot j}$ das classes B_j . Isto é, se $p_{\cdot j_0}$ fosse consideravelmente menos (mais) que os outros $p_{\cdot j}$, $\left(\frac{p_{ij_0}}{p_{i\cdot}} - \frac{p_{i'j_0}}{p_{i'\cdot}} \right)^2$ tenderia a ter um papel muito mais (menos) importante que as demais parcelas na distância euclidiana $\sum_{j=1}^J \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2$.

Lembramos que nosso objetivo é a representação simultânea das classes A_i e B_j em um espaço de dimensão p menor que I e J . Podemos então fazer uma análise de componentes principais para as classes A_1, \dots, A_I , representadas por seus perfis p_1^B, \dots, p_I^B e com a distância acima definida. Como essas classes têm, em geral, contingentes diferentes, é razoável dar-lhes pesos μ_1^A, \dots, μ_I^A iguais às suas frequências, isto é, $\mu_i^A = p_{i\cdot}$ para todo $i = 1, \dots, I$. Fazemos o mesmo com as classes B_1, \dots, B_J , com pesos $\mu_j^B = p_{\cdot j}$ para todo $j = 1, \dots, J$. Na próxima seção, daremos propriedades dessas representações e relações entre elas.

3.2 Resultados

Consideremos as matrizes P^B de dimensão $I \times J$ e P^A de dimensão $J \times I$, cujas linhas são, respectivamente, os perfis das classes A_1, \dots, A_i e os perfis das classes B_1, \dots, B_I , isto é:

$$P^B = \Sigma^A P = \begin{pmatrix} (p_1^B)' \\ \vdots \\ (p_I^B)' \end{pmatrix} \quad \text{e} \quad P^B = \Sigma^B P = \begin{pmatrix} (p_1^A)' \\ \vdots \\ (p_J^A)' \end{pmatrix}$$

Aplicaremos duas análises de componentes principais. A primeira, sobre os vetores p_1^B, \dots, p_I^B pertencentes ao espaço \mathbb{R}^J munido do produto interno definida por Σ^B e com pesos $\mu_i^A = p_{i\cdot}$, $i = 1, \dots, I$. A

segunda, sobre os vetores p_1^A, \dots, p_J^A pertencentes ao espaço \mathbb{R}^I munido do produto interno definido por Σ^A e com pesos $\mu_j^B = p_{\cdot j}$, $j = 1, \dots, J$. Sejam

$$(p^B)' = \left(\sum_{i=1}^I \mu_i^A p_i^B \right)' = (p_{\cdot 1}, \dots, p_{\cdot J}) \quad \text{e}$$

$$(p^A)' = \left(\sum_{j=1}^J \mu_j^B p_j^A \right)' = (p_{1\cdot}, \dots, p_{I\cdot}),$$

os vetores médias ponderadas, respectivamente, dos vetores p_i^B e p_j^A .

Sejam $x_i^B = p_i^B - p^B$ e $x_j^A = p_j^A - p^A$ seus desvios.

Aplicamos agora o Teorema 2.1. Temos então: (i) as matrizes de covariância amostral

$$\begin{aligned} C^B &= \sum_{i=1}^I \mu_i^A x_i^B (x_i^B)' = \sum_{i=1}^I p_{i\cdot} p_i^B (p_i^B)' - p^B (p^B)' = \\ &= (p^B)' (\Sigma^A)^{-1} (p^B) - p^B (p^B)' = P' \Sigma^A P - p^B (p^B)' \end{aligned}$$

e

$$C^A = \sum_{j=1}^J \mu_j^B x_j^A (x_j^A)' = P \Sigma^B P' - p^A (p^A)'$$

e (ii) as matrizes

$$\begin{aligned} U^B &= (u_1^B, \dots, u_J^B), \quad (\Sigma^B)^{-1} - \text{ortonormal e} \\ U^A &= (u_1^A, \dots, u_I^A), \quad (\Sigma^A)^{-1} - \text{ortonormal} \end{aligned}$$

tais que:

$$\Sigma^B C^B U^B = \Lambda^B U^B$$

e

$$\Sigma^A C^A U^A = \Lambda^A U^A$$

onde $\Lambda^B = \text{diag}(\lambda_1^B, \dots, \lambda_J^B)$ e $\Lambda^A = \text{diag}(\lambda_1^A, \dots, \lambda_I^A)$ com $\lambda_1^B \geq \dots \geq \lambda_J^B \geq 0$ e $\lambda_1^A \geq \dots \geq \lambda_I^A \geq 0$.

Ainda pelo Teorema 2.1, $F_\alpha(i) = (u_\alpha^B)' x_i^B$ e $G_\alpha(j) = (u_\alpha^A)' x_j^A$ são, respectivamente, as α -ésimas componentes principais das classes A_i e B_j .

Teorema 3.1.

- (i) O vetor $u_J^B = I_J = (1, \dots, 1)'$ é autovetor de $\Sigma^B C^B$ associado ao autovalor $\lambda_J^B = 0$.

O vetor $u_I^A = 1_A = (1, \dots, 1)'$ é autovetor de $\Sigma^A C^A$ associado ao autovalor $\lambda_I^A = 0$.

- (ii) Os autovetores u_α^B , $\alpha = 1, \dots, J - 1$ satisfazem:

$$(p^B)' u_\alpha^B = \sum_{j=1}^B p_{\cdot j} u_\alpha^B(j) = 0.$$

Os autovetores u_α^A , $\alpha = 1, \dots, I - 1$ satisfazem:

$$(p^A)' u_\alpha^A = \sum_{i=1}^I p_{i \cdot} u_\alpha^A(i) = 0.$$

- (iii) A matriz $H^B = \Sigma^B P' \Sigma^A P$ de termo geral $h_{jj'} = \frac{1}{p_{\cdot j}} \sum_{i=1}^I \frac{p_{ij} p_{ij'}}{p_{i \cdot}}$ tem como autovetores u_α^B , $1 \leq \alpha \leq J$, associados aos autovalores $\lambda_1^B, \dots, \lambda_{J-1}^B$ e $\lambda_J = 1$.

$$A \text{ matriz } H^A = \Sigma^A P \Sigma^B P' \text{ de termo geral } h_{ii'} = \frac{1}{p_{i \cdot}} \sum_{j=1}^J \frac{p_{ij} p_{ij'}}{p_{\cdot j}}$$

tem como autovetores u_α^A , $1 \leq \alpha \leq I$, associados aos autovalores $\lambda_1^A, \dots, \lambda_{I-1}^A$ e $\lambda_I = 1$.

- (iv) A matriz simétrica $S^B = (\Sigma^B)^{1/2} P' \Sigma^A P (\Sigma^B)^{1/2}$ de termo geral $S_{jj'}^B = \sum_{i=1}^I \frac{p_{ij} p_{ij'}}{p_{i \cdot} \sqrt{p_{\cdot j} p_{\cdot j'}}$ tem os mesmos autovalores que a matriz H^B e seus autovetores d_α^B são ortonormais e satisfazem $d_\alpha^B = (\Sigma^B)^{-1/2} u_\alpha^B$, isto é, $d_\alpha^B(j) = \sqrt{p_{\cdot j}} u_\alpha^B(j)$.

A matriz simétrica $S^A = (\Sigma^A)^{1/2} P \Sigma^B P' (\Sigma^A)^{1/2}$ de termo geral $S_{ii'}^A = \sum_{j=1}^J \frac{p_{ij} p_{ij'}}{p_{\cdot j} \sqrt{p_{i \cdot} p_{i' \cdot}}}$ tem os mesmos autovalores que a matriz H^A e seus autovetores d_α^A são ortonormais e satisfazem $d_\alpha^A = (\Sigma^A)^{-1/2} u_\alpha^A$, isto é, $d_\alpha^A(i) = \sqrt{p_{i \cdot}} u_\alpha^A(i)$.

- (v) Os autovalores não nulos de $\Sigma^B C^B$ coincidem com os autovalores não nulos de $\Sigma^A C^A$.

(vi) Se $\lambda_\alpha > 0$, então

$$\begin{aligned} u_\alpha^B &= \frac{1}{\sqrt{\lambda_\alpha}} \Sigma^B P' u_\alpha^A, \text{ isto é, } u_\alpha^B(j), = \\ &= \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^I \frac{p_{ij}}{p_{\cdot j}} u_\alpha^A(i), j = 1, \dots, J. \end{aligned}$$

$$\begin{aligned} u_\alpha^A &= \frac{1}{\sqrt{\lambda_\alpha}} \Sigma^A P u_\alpha^B, \text{ isto é, } u_\alpha^A(i) = \\ &= \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^J \frac{p_{ij}}{p_{i\cdot}} u_\alpha^B(j), i = 1, \dots, I. \end{aligned}$$

(vii) Sejam $F_\alpha = (F_\alpha(1), \dots, F_\alpha(I))'$ e $G_\alpha = (G_\alpha(1), \dots, G_\alpha(J))'$. Então se $\lambda_\alpha > 0$, temos:

$$F_\alpha = \sqrt{\lambda_\alpha} u_\alpha^A$$

e

$$G_\alpha = \sqrt{\lambda_\alpha} u_\alpha^B$$

(viii) Se $\lambda_\alpha > 0$, então

$$F_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \Sigma^A P G_\alpha, \text{ isto é, } F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^J \frac{p_{ij}}{p_{i\cdot}} G_\alpha(j), i = 1, \dots, I,$$

e

$$G_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \Sigma^A P' F_\alpha, \text{ isto é, } G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^I \frac{p_{ij}}{p_{\cdot j}} F_\alpha(i), j = 1, \dots, J.$$

(ix) $\lambda_\alpha \leq 1$ para todo α .

Demonstração. (i) Mostremos que $\Sigma^B C^B 1_J = 0$. Logo precisamos mostrar que a soma em cada linha de $\Sigma^B C^B$ é igual a zero.

Como $\Sigma^B C^B = \Sigma^B P' \Sigma^A P - \Sigma^B p^B (p^B)'$, se $C^b = (c_{jj'}^B)$, temos

$$\begin{aligned} \sum_{j'=1}^J \frac{c_{jj'}}{p_{\cdot j}} &= \sum_{j'=1}^J \frac{1}{p_{\cdot j}} \sum_{i=1}^I \frac{p_{ij} p_{ij'}}{p_{i\cdot} i} - \sum_{j'=1}^J \frac{p_{\cdot j} p_{\cdot j'}}{p_{\cdot j}} = \\ &= \frac{1}{p_{\cdot j}} \sum_{i=1}^I \frac{p_{ij}}{p_{i\cdot}} \sum_{j'=1}^J p_{ij'} - \sum_{j'=1}^J p_{\cdot j'} = 1 - 1 = 0. \end{aligned}$$

(ii) Segue imediatamente do fato de que os u_α^B , são $(\Sigma^B)^{-1}$ -ortonormais e $u_J^B = 1_J$.

(iii) Vimos em (i) que $\Sigma^B C^B = H^B - \Sigma^B p^B (p^B)'$. Logo para $1 \leq \alpha \leq J-1$

$$\lambda_\alpha^B u_\alpha^B = \Sigma^B C^B u_\alpha^B = H^B u_\alpha^B - \Sigma^B p^B (p^B)' u_\alpha^B$$

e o resultado segue de (ii).

Para $\alpha = J$, o resultado segue da demonstração de (i) onde mostramos que $H^B 1_J = 1_J$.

(iv) $H^B u_\alpha^B = \Sigma^B P' \Sigma^A P u_\alpha^B = \lambda_\alpha^B u_\alpha^B$.

Logo

$$(\Sigma^B)^{1/2} P' \Sigma^A P (\Sigma^B)^{1/2} (\Sigma^B)^{-1/2} u_\alpha^B = \lambda_\alpha^B (\Sigma^B)^{-1/2} u_\alpha^B$$

ou seja $S^B d_\alpha^B = \lambda_\alpha^B d_\alpha^B$.

Logo os autovalores de S^B e H^B são os mesmos e $d_\alpha^B = (\Sigma^B)^{-1/2} u_\alpha^B$ são os autovetores de S^B .

Além disso, $(d_\alpha^B)' (d_{\alpha'}^B) = (u_\alpha^B)' (\Sigma^B)^{-1/2} (\Sigma^B)^{-1/2} u_{\alpha'}^B = \delta_{\alpha\alpha'}$ e portanto os $d_\alpha^{B'}$ são ortonormais.

(v) Segue do fato que

$$S^B = (\Sigma^B)^{1/2} P' (\Sigma^A)^{1/2} (\Sigma^A)^{1/2} P (\Sigma^B)^{1/2} = L' L$$

e

$$S^A = (\Sigma^A)^{1/2} P (\Sigma^B)^{1/2} (\Sigma^B)^{1/2} P' (\Sigma^A)^{1/2} = L L'$$

onde $L = (\Sigma^A)^{1/2} P (\Sigma^B)^{1/2}$.

(vi) Como $S^B = L' L$ e $S^A = L L'$ e os d_α^B e os d_α^A , tem norma 1, temos

$$d_\alpha^B = \frac{1}{\sqrt{\lambda_\alpha}} L' d_\alpha^A \quad \text{e} \quad d_\alpha^A = \frac{1}{\sqrt{\lambda_\alpha}} L d_\alpha^B.$$

Logo

$$(\Sigma^B)^{1/2} u_\alpha^B = \frac{1}{\sqrt{\lambda_\alpha}} (\Sigma^B)^{1/2} P' (\Sigma^A)^{1/2} (\Sigma^A)^{-1/2} u_\alpha^A$$

ou seja $u_\alpha^B = \frac{1}{\sqrt{\lambda_\alpha}} \Sigma^B P' u_\alpha^A$.

$$\text{Analogamente } u_{\alpha}^A = \frac{1}{\sqrt{\lambda_{\alpha}}} \Sigma^A P u_{\alpha}^B.$$

(vii) $F_{\alpha}(i) = (x_i^B)' u_{\alpha}^B = (p_i^B - p^B)' u_{\alpha}^B = (p_i^B)' u_{\alpha}^B$ devido a (ii).

Logo $F_{\alpha} = P^B u_{\alpha}^B = \Sigma^A P u_{\alpha}^B = \sqrt{\lambda_{\alpha}} u_{\alpha}^A$ por (vi).

$$\text{Analogamente } G_{\alpha} = \sqrt{\lambda_{\alpha}} u_{\alpha}^B.$$

(viii) Imediato de (vii) e (vi).

(ix) Suponhamos que $\lambda_{\alpha} > 1$. Então $\frac{1}{\sqrt{\lambda_{\alpha}}} < 1$.

Logo por (vi), $\min_i u_{\alpha}^A(i) < u_{\alpha}^B(j) < \max_i u_{\alpha}^A(i)$ para todo j , pois $\sum_{i=1}^I p_{i\cdot} u_{\alpha}^A(i) = 0$ implica $\min_i u_{\alpha}^A(i) < 0 < \max_i u_{\alpha}^A(i)$.

Consequentemente $\min_i u_{\alpha}^A(i) < \min_j u_{\alpha}^B(j) < \max_j u_{\alpha}^B < \max_i u_{\alpha}^A(i)$.

Novamente por (vi) e (ii), temos

$$\min_j u_{\alpha}^B(j) < u_{\alpha}^A(i) < \max_j u_{\alpha}^A(j)$$

o que implica $\min_i u_{\alpha}^A(i) < \min_j u_{\alpha}^B(j) < \max_i u_{\alpha}^A(i)$. Portanto chegamos a uma contradição e λ_{α} tem que ser ≤ 1 .

Observações

a) Na prática, calculamos primeiro, os autovalores e autovetores da matriz S^B (ou S^A), que é uma matriz simétrica. Descartamos, então, o maior autovalor 1 e seu autovetor associado

$$d^B = \left(\frac{1}{\sqrt{p_{\cdot 1}}}, \dots, \frac{1}{\sqrt{p_{\cdot J}}} \right)' \quad \left(d^A = \left(\frac{1}{\sqrt{p_{\cdot i}}}, \dots, \frac{1}{\sqrt{p_{\cdot I}}} \right)' \right).$$

b) O item (viii) nos diz que a α -ésima coordenada da linha i é, a menos de uma dilatação, a média ponderada das α -ésimas coordenadas das colunas, o peso da j -ésima coluna sendo a frequência condicional dessa coluna dado a i -ésima linha, isto é, $p_{ij}/p_{i\cdot}$. E vice-versa.

Assim, se λ_{α} próximo de 1 e se $\frac{p_{ij}}{p_{i\cdot}}$ grande (próximo de 1), a α -ésima coordenada da linha i deverá estar próxima da α -ésima coordenada da coluna j .

Esse fato justifica a representação simultânea das linhas e colunas no mesmo gráfico.

3.3 Formulações Equivalentes

(i) **Método do escore de Fisher para tabelas de contingência.**

Seja $P = (p_{ij})_{\substack{1 \leq i \leq I \\ 1 \leq j \leq J}}$ uma tabela de contingência, $\sum_i \sum_j p_{ij} = 1$. Podemos considerar p_{ij} como uma estimativa da probabilidade de ocorrência do par (i, j) . O método consiste em determinar funções f e g definidas sobre $\{1, \dots, I\}$ e $\{1, \dots, J\}$ respectivamente, tais que tenham correlação máxima.

Para que esse problema tenha solução, é necessário impor restrições às funções f e g . Como é natural, impomos as restrições de que as funções f e g tenha média zero e variância igual a um.

Logo, queremos maximizar $\text{cor}(f, g) = \sum_i \sum_j f(i)g(j)p_{ij}$ sujeito às restrições

$$\begin{aligned} Ef = \sum_i f(i)p_{i\cdot} &= 0, & Eg = \sum_j g(j)p_{\cdot j} &= 0 \\ \text{Var } f = \sum_i f^2(i)p_{i\cdot} &= 1, & \text{Var } g = \sum_j g^2(j)p_{\cdot j} &= 1. \end{aligned}$$

Nesse caso, o método é equivalente a determinar funções f e g satisfazendo as restrições acima e que minimizem $E(f - g)^2 = \sum_i \sum_j (f(i) - g(j))^2 p_{ij}$.

De fato,

$$E(f - g)^2 = \sum_i f^2(i)p_{i\cdot} + \sum_j g^2(j)p_{\cdot j} - 2 \sum_i \sum_j f(i)g(j)p_{ij}.$$

Logo $E(f - g)^2 = 2 - 2 \text{cor}(f, g)$.

Uma vez achados f e g , podemos procurar funções f_2 e g_2 , definidas em $\{1, \dots, I\}$ e $\{1, \dots, J\}$ respectivamente, que tenham correlação máxima, média zero, variância um e sejam não correlacionadas com f e g , isto é, procuramos funções f_1 e f_2 que maximizem $\sum_i \sum_j f_2(i)g_2(j)p_{ij}$

sujeitas às restrições.

$$\begin{aligned}
 E f_2 &= \sum_i f_2(i) p_{i\cdot} = 0, & E g_2 &= \sum_j g_2(j) p_{\cdot j} = 0, \\
 \text{Var } f_2 &= \sum_i f_2^2(i) p_{i\cdot} = 1, & \text{Var } g_2 &= \sum_j g_2^2(j) p_{\cdot j} = 1, \\
 \text{cor}(f, f_2) &= \sum_i f(i) f_2(i) p_{i\cdot} = 0, & \text{cor}(g, g_2) &= \sum_j g(j) g_2(j) p_{\cdot j} = 0, \\
 \text{cor}(f, g_2) &= \sum_i \sum_j f(i) g_2(j) p_{ij} = 0, & \text{cor}(f_2, g) &= \sum_i \sum_j f_2(i) g(j) p_{ij} = 0.
 \end{aligned}$$

e assim por diante.

Os valores $f_\alpha(i)$ e $g_\alpha(j)$ são chamados de α -ésimos escores. Esse método está descrito em Hirschfield (1935), Fisher (1940), Williams (1952), Kendall & Stuart (1961, pg. 569), Benzecri (1973), Lancaster (1969) e Hill (1974). Em particular, Kendall & Stuart (1961) fornece uma justificação teórica baseada na distribuição normal.

As funções f_1, g_1, f_2, g_2 , etc... São usualmente obtidas, utilizando-se, em cada etapa, os multiplicadores de Lagrange.

Mais geralmente, e usando-se notação matricial, procuramos funções f_1, \dots, f_r definidas em $\{1, \dots, I\}$ e funções g_1, \dots, g_r definidas em $\{1, \dots, J\}$ tais que

$$\begin{aligned}
 \text{(a) } E f_\alpha &= \sum_i f_\alpha(i) p_{i\cdot} = 0 & \alpha &= 1, \dots, r \\
 \text{Var } f_\alpha &= \sum_i f_\alpha^2(i) p_{i\cdot} = f_\alpha' (\Sigma^A)^{-1} f_\alpha = 1,
 \end{aligned}$$

onde

$$\begin{aligned}
 f_\alpha &= (f_\alpha(1), \dots, f_\alpha(I))', & \alpha &= 1, \dots, r, \\
 \text{cor}(f_\alpha, f_{\alpha'}) &= \sum_i f_\alpha(i) f_{\alpha'}(i) p_{i\cdot} = f_\alpha (\Sigma^A)^{-1} f_{\alpha'} = 0 & \alpha &\neq \alpha'
 \end{aligned}$$

$$\begin{aligned}
 \text{(b) } E g_\alpha &= \sum_j g_\alpha(j) p_{\cdot j} = 0 & \alpha &= 1, \dots, r \\
 \text{Var } g_\alpha &= \sum_j g_\alpha^2(j) p_{\cdot j} = g_\alpha' (\Sigma^B)^{-1} g_\alpha = 1,
 \end{aligned}$$

onde

$$g_\alpha = (g_\alpha(1), \dots, g_\alpha(J))', \quad \alpha = 1, \dots, r,$$

$$\text{cor}(g_\alpha, g_{\alpha'}) = \sum_j g_\alpha(j) g_{\alpha'}(j) p_{ij} = g_{\alpha'} (\Sigma^B)^{-1} g_\alpha = 0 \quad \alpha \neq \alpha'$$

$$(c) \quad \text{cor}(f_\alpha, g_{\alpha'}) = \sum_i \sum_j f_\alpha(i) g_{\alpha'}(j) p_{ij} = f_\alpha' P g_{\alpha'} = 0 \quad \alpha \neq \alpha'$$

e que maximizem $\sum_{\alpha=1}^r \text{cor}(f_\alpha, g_\alpha) = \sum_{\alpha=1}^r f_\alpha' P g_\alpha$.

Como visto na Proposição 8 do Capítulo 1, a solução desse problema é dada pelos vetores u_α^A e u_α^B , $\alpha = 1, \dots, r$, autovetores das matrizes $\Sigma^A P \Sigma^B P'$ e $\Sigma^B P' \Sigma^A P$, respectivamente, e associados aos autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$. Isto é, $f_\alpha(i) = u_\alpha^A(i)$ e $g_\alpha(j) = u_\alpha^B(j)$, $\alpha = 1, \dots, r$. Temos ainda pelo Teorema 2, que $\text{cor}(f_\alpha, g_\alpha) = \sqrt{\lambda_\alpha}$;

Observação 1. Temos, essencialmente, a mesma solução da análise de correspondência, pois $F_\alpha(i) = \sqrt{\lambda_\alpha} u_\alpha^A(i)$ e $G_\alpha(j) = \sqrt{\lambda_\alpha} u_\alpha^B(j)$, $\alpha = 1, \dots, r$, $i = 1, \dots, I$ e $j = 1, \dots, J$.

Observação 2. O fato de que $\text{cor}(f_\alpha, g_\alpha) = \text{cor}(u_\alpha^A, u_\alpha^B) = \sqrt{\lambda_\alpha}$, nos dá uma outra demonstração de que $\lambda_\alpha \leq 1$.

Observação 3. Os autovetores 1_I e 1_J de $\Sigma^A P \Sigma^B P'$ e $\Sigma^B P' \Sigma^A P$, respectivamente, associados ao maior autovalor 1, não satisfazem a restrição $Ef = Eg = 0$.

Observação 4. O fato de que u_α^A e u_α^B minimizam $E(f_\alpha(i) - g_\alpha(j))^2 = \sum_i \sum_j (f_\alpha(i) - g_\alpha(j))^2 p_{ij}$, nos dá uma outra interpretação a representação por nós obtida. Pois, nesse caso p_{ij} grande em relação aos outros implica $u_\alpha^A(i) \simeq u_\alpha^B(j)$, o que por sua vez implica $F_\alpha(i) \simeq G(j)$. E quanto maior $\sqrt{\lambda_\alpha} = \text{cor}(u_\alpha^A, u_\alpha^B)$, mais verdadeiro é esse fato.

(ii) **Análise de correspondência como um método de escalamento simultâneo.**

Suponhamos que temos uma tabela numérica $N = (n_{ij})_{\substack{1 \leq i \leq I \\ 1 \leq j \leq J}}$ talque $n_{ij} \geq 0$, $n_{i\cdot} > 0$, $n_{\cdot j} > 0$ para todo $1 \leq i \leq I$, $1 \leq j \leq J$. Seja $P = (p_{ij})_{\substack{1 \leq i \leq I \\ 1 \leq j \leq J}}$, $p_{ij} = n_{ij}/n$, onde $n = \sum_i \sum_j n_{ij}$, a tabela normalizada de frequências. Podemos pensar, por exemplo, em uma tabela de contingência.

Consideremos o seguinte problema: Queremos obter escores x_i para as linhas (classes A_i) e escores y_j para as colunas (classes B_j) sobre um mesmo eixo tal que cada escore x_i seja uma média ponderada dos escores y_j , sendo o peso atribuído a y_j igual a importância relativa da coluna j (classe B_j) em relação a linha i (classe A_i), isto é, $p_{ij}/p_{i\cdot}$. Dessa maneira, queremos

$$x_i = \sum_j \frac{p_{ij}}{p_{i\cdot}} y_j.$$

Analogamente, queremos

$$y_j = \sum_i \frac{p_{ij}}{p_{\cdot j}} x_i.$$

Em notação matricial, sendo Σ^A , Σ^B , como antes, $X = (x_1, \dots, x_I)'$ e $Y = (y_1, \dots, y_J)'$, temos

$$X = \Sigma^A P Y$$

e

$$Y = \Sigma^B P' X.$$

É claro que a solução trivial $X = I_I$, e $Y = 1_J$ não é interessante. Em geral, não existem outras soluções do problema da maneira como está apresentado.

Uma modificação possível, é ter os escores proporcionais à média ponderada, isto é,

$$x_i = \beta \sum_j \frac{p_{ij}}{p_{i\cdot}} y_j \quad \text{e} \quad y_j = \beta \sum_i \frac{p_{ij}}{p_{\cdot j}} x_i$$

com β tão próximo de 1 quanto possível, ou em notação matricial

$$X = \beta \Sigma^A P Y \quad \text{e} \quad Y = \beta \Sigma^B P' X,$$

o que equivale a termos

$$X = \beta^2 \Sigma^A P' \Sigma^B P' X \quad \text{e} \quad Y = \beta^2 \Sigma^B P' \Sigma^A P Y.$$

Temos, então, que X deve ser autovetor de $H^A = \Sigma^A P \Sigma^B P'$ e Y deve ser autovetor de $H^B = \Sigma^B P' \Sigma^A P$, ambos associados ao autovalor $1/\beta^2$.

Essa modificação é razoável pois se partirmos de uma solução inicial $X_0 \neq 1_I$, definirmos $Y_0 = \Sigma^B P' X_0$, $X_1 = \Sigma^A P' Y_0$, $Y_1 = \Sigma^B P' X_1$ e iterarmos, teremos:

$$X_n = (\Sigma^A P \Sigma^B P')^n X_0 \quad \text{e} \quad Y_n = (\Sigma^B P' \Sigma^A P)^n Y_0.$$

Sabemos que (ver por exemplo, Morrison (1967)) que se em cada etapa, normalizamos X_n e Y_n , X_n converge a um autovetor de H^A e Y_n converge a um autovetor de H^B .

Ver também Hill (1974) para motivação.

Como vimos antes, H^A e H^B têm os mesmos autovalores e esses são sempre ≤ 1 . Logo, como procuramos β tão próximo de 1 quanto possível, a solução procurada é $X = u_1^A$ e $Y = u_1^B$ associados ao autovalor $\lambda_1 = \frac{1}{\beta^2}$.

Podemos estender de maneira natural esse escalamento em um eixo, a um escalamento em k eixos ortonormais, ortonormalidade essa, definida apropriadamente para as linhas e as colunas, e cuja solução é $X^{(\alpha)} = u_\alpha^A$ e $Y_\alpha^{(\alpha)} = Y^{(\alpha)} = u_\alpha^B$, $\alpha = 1, \dots, k$.

Logo, temos novamente, essencialmente, a solução da análise de correspondência.

3.4 Exemplos de tipos de tabela

Já mencionamos várias vezes a tabela de contingência $A \times B$. Daremos exemplos, agora de outros tipos de tabela que podem ser considerados.

(i) **Tabelas de contingência multivariadas.**

Suponhamos que temos uma tabela de contingência a 3 dimensões $A \times B \times C$. Podemos tratar essa tabela, olhando as tabelas bidimensionais $A \times (B \times C)$, $(A \times B) \times C$ e $(A \times C) \times B$.

Uma outra maneira de tratarmos o problema, quando estamos interessados em explicar, digamos a categoria A , é considerarmos as tabelas $A \times B$ e $A \times C$ justapostas como na figura

	$B_1 \dots B_j$	$C_1 \dots C_K$
A_1		
\vdots		
A_I		

É claro que podemos estender os conceitos acima, a tabelas de dimensão maiores. No primeiro caso, caímos sempre numa tabela de contingência bivariada. Consideremos então o segundo caso em que temos justaposição de tabelas de contingência bivariadas.

Consideremos então, a categoria A com classes $A_i = i = 1, \dots, I$ e categorias $B^{(q)}$, $q = 1, \dots, Q$, a categoria $B^{(q)}$ com J_q classes $B_j^{(q)}$. Formemos a tabela abaixo que é a justaposição das tabelas de contingência. $A \times B^{(q)}$, $q = 1, \dots, Q$.

$A \setminus B$	$B_1^{(1)}$...	$B_{J_1}^{(1)}$	$B_1^{(2)}$...	$B_{J_2}^{(2)}$...	$B_1^{(Q)}$...	$B_{J_Q}^{(Q)}$	
A_1	$n_{11}^{(1)}$		$n_{1J_1}^{(1)}$	$n_{11}^{(2)}$		$n_{1J_2}^{(2)}$		$n_{11}^{(Q)}$		$n_{1J_Q}^{(Q)}$	$Qn_{1.}$
\vdots											
A_I	$n_{I1}^{(1)}$		$n_{IJ_1}^{(1)}$	$n_{I1}^{(2)}$		$n_{IJ_2}^{(2)}$		$n_{I1}^{(Q)}$		$n_{IJ_Q}^{(Q)}$	$Qn_{I.}$
	$n_{.1}^{(1)}$		$n_{.J_1}^{(1)}$	$n_{.1}^{(2)}$		$n_{.J_2}^{(2)}$		$n_{.1}^{(Q)}$		$n_{.J_Q}^{(Q)}$	Qn

Seja

$n_{ij}^{(q)}$ = número de elementos na classe A_i e classe $B_j^{(q)}$,

$n_{i.}$ = número de elementos na classe A_i

$n_{.j}^{(q)}$ = número de elementos na classe $B_j^{(q)}$.

Então, usando-se a notação com o subscrito q para a tabela $A \times B^{(q)}$ e sem o subscrito para a tabela geral, temos:

$$\begin{aligned} p_{i\cdot} &= \frac{Q n_i}{Q n} = \frac{n_{i\cdot}}{n} = p_i^{(q)} && \text{para todo } q = 1, \dots, Q, \\ p_{\cdot j} &= \frac{n_{\cdot j}}{Q n} = \frac{n_{\cdot j'}^{(q)}}{Q n} = \frac{1}{Q} p_{\cdot j'}^{(q)} && \text{para algum } q \text{ e } j', \text{ e} \\ p_{ij} &= \frac{n_{ij}}{Q n} = \frac{n_{ij'}^{(q)}}{Q n} = \frac{1}{Q} p_{ij'}^{(q)} && \text{para algum } q \text{ e } j'. \end{aligned}$$

Logo

$$\begin{aligned} D^2(A_i, A_{i'}) &= \sum_{j=1}^{J_1 + \dots + J_Q} \frac{1}{p_{\cdot j}} \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2 = \\ &= \sum_{q=1}^Q \sum_{j=1}^{J_q} \frac{Q}{p_{\cdot j}^{(q)}} \frac{1}{Q^2} \left(\frac{p_{ij}^{(q)}}{p_{i\cdot}^{(q)}} - \frac{p_{i'j}^{(q)}}{p_{i'\cdot}^{(q)}} \right)^2 = \frac{1}{Q} \sum_{q=1}^Q \left[d^{(q)}(A_i, A_{i'}) \right]^2, \end{aligned}$$

isto é, o quadrado da distância entre as classes A_i e $A_{i'}$ é a média aritmética dos quadrados das distâncias entre as classes A_i e $A_{i'}$ nas tabelas $A \times B^{(q)}$.

Temos também

$$d^2(B_j^{(q)}, B_{j'}^{(q')}) = \sum_{i=1}^I \frac{1}{p_{i\cdot}} \left(\frac{p_{ij}^{(q)}}{p_{\cdot j}^{(q)}} - \frac{p_{ij'}^{(q')}}{p_{\cdot j'}^{(q')}} \right)^2.$$

Logo se $q = q'$; $d^2(B_j^{(q)}, B_{j'}^{(q')}) = \left[d^{(q)}(B_j^{(q)}, B_{j'}^{(q)}) \right]^2$, isto é, a distância entre duas classes na mesma categoria $B^{(q)}$ permanece a mesma. E se duas classes, em categorias diferentes têm perfis parecidos, elas estão próximas.

Vejamos agora, o que acontece com as coordenadas F_α e G_α .

$$G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^I \frac{p_{ij}}{p_{\cdot j}} F_\alpha(i)$$

e

$$\begin{aligned} F_\alpha(i) &= \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^{J_1+\dots+J_Q} \frac{p_{ij}}{p_{i\cdot}} G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{q=1}^Q \sum_{j=1}^{J_q} \frac{1}{Q} \frac{p_{ij}^{(q)}}{p_{i\cdot}} G_\alpha(j) = \\ &= \frac{1}{Q} \sum_{q=1}^Q \left(\frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^{J_q} \frac{p_{ij}^{(q)}}{p_{i\cdot}} G_\alpha(j) \right). \end{aligned}$$

(ii) Tabelas de incidência

Suponhamos I indivíduos A_i e Q critérios ou perguntas $B^{(1)}, \dots, B^{(Q)}$. A pergunta $B^{(q)}$ é composta de J_q itens exclusivos, isto é, cada indivíduo só pode responder a um item da pergunta e tem que responder um deles.

$$\text{Seja } n_{ij}^{(q)} = \begin{cases} 1 & \text{se } A_i \text{ responde } B_j^{(q)} \\ 0 & \text{se } A_i \text{ não responde } B_j^{(q)} \end{cases}$$

Temos então a tabela

$A \setminus B$	$B_1^{(1)}$	\dots	$B_{J_1}^{(1)}$	\dots	$B_1^{(Q)}$	\dots	$B_{J_Q}^{(Q)}$	
A_1	$n_{11}^{(1)}$	\dots	$n_{1J_1}^{(1)}$	\dots	$n_{11}^{(Q)}$	\dots	$n_{1J_Q}^{(Q)}$	Q
\vdots								
A_I	$n_{I1}^{(1)}$		$n_{IJ_1}^{(1)}$		$n_{I1}^{(Q)}$		$n_{IJ_Q}^{(Q)}$	Q
	$n_{\cdot 1}^{(1)}$		$n_{\cdot J_1}^{(1)}$		$n_{\cdot 1}^{(Q)}$		$n_{\cdot J_Q}^{(Q)}$	IQ

Segue então que $p_{ij}^{(q)} = \frac{n_{ij}^{(q)}}{I}$ em cada tabela $A \times B^{(q)}$ e que na tabela geral

$$p_{ij} = \frac{n_{ij}}{IQ} = \frac{1}{Q} \frac{n_{ij}^{(q)}}{I} = \frac{1}{Q} p_{ij}^{(q)} \quad \text{para algum } q \text{ e } j'.$$

Temos também $p_{i\cdot}^{(q)} = p_{i\cdot} = \frac{1}{I}$ e $p_{\cdot j} = \frac{1}{Q} p_{\cdot j}^{(q)}$ para q e j' .

Consideremos a distância entre dois indivíduos.

$$D^2(A_i, A_{i'}) = \sum_{j=1}^{J_1+\dots+J_Q} \frac{1}{p_{\cdot j}} \left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2 =$$

$$= \sum_{q=1}^Q \sum_{j=1}^{J_q} \frac{IQ}{n_{\cdot,j}^{(q)}} \left(\frac{n_{ij}^{(q)} - n_{i'j}^{(q)}}{Q} \right)^2 = \frac{1}{Q} \sum_{q=1}^Q \sum_{j=1}^{J_q} \frac{1}{n_{\cdot,j}^{(q)}} \left(n_{ij}^{(q)} - n_{i'j}^{(q)} \right)^2.$$

Como $n_{ij}^{(q)} = 1$ ou 0 e $n_{i'j}^{(q)} = 1$ somente para um j em $[1, J_q]$, seja $j(i, q)$ tal que $n_{ij(i,q)}^{(q)} = 1$. Então

$$\begin{aligned} d^2(A_i, A_{i'}) &= \frac{I}{Q} \sum_q \left[\frac{1}{n_{\cdot,j(i,q)}^{(q)}} + \frac{1}{n_{\cdot,j(i',q)}^{(q)}} \right] \delta_{ii'}^{(q)} = \\ &= \frac{I}{Q} \sum_q \left[\frac{n_{\cdot,j(i,q)}^{(q)} + n_{\cdot,j(i',q)}^{(q)}}{n_{\cdot,j(i,q)}^{(q)} n_{\cdot,j(i',q)}^{(q)}} \right] \delta_{ii'}^{(q)} \end{aligned}$$

onde

$$\delta_{ii'}^{(q)} = \begin{cases} 1 & \text{se } j(i, q) \neq j(i', q) \\ 0 & \text{caso contrário} \end{cases}$$

No caso $J_q = 2$, $q = 1, \dots, Q$, temos

$$d^2(A_i, A_{i'}) = \frac{I}{Q} \sum_q \frac{I \delta_{ii'}^{(q)}}{n_{\cdot,1}^{(q)} (I - n_{\cdot,1}^{(q)})} = \frac{1}{Q} \sum_q \frac{1}{p_{\cdot,1}^{(q)} (1 - p_{\cdot,1}^{(q)})} \delta_{ii'}^{(q)}.$$

O termo $\frac{1}{p_{\cdot,1}^{(q)} (1 - p_{\cdot,1}^{(q)})} \delta_{ii'}^{(q)}$ mede a distância entre os indivíduos A_i e $A_{i'}$ com respeito a pergunta q . Se a resposta é a mesma, essa distância é nula, se a resposta é diferente, essa distância depende de quão frequentes são cada um dos itens e quanto menos frequente um dos itens de resposta, maior a distância. Observemos que $p_{\cdot,1}^{(q)} (1 - p_{\cdot,1}^{(q)})$ é a variância de uma distribuição binomial com probabilidade $p_{\cdot,1}^{(p)}$ de sucesso.

Finalmente, $d^2(A_i, A_{i'})$ é a média aritmética dos quadrados das distâncias dos indivíduos A_i e $A_{i'}$ com respeito as Q perguntas.

No caso geral, como $B_J^{(q)}$ e $B_{j'}^{(q)}$ são disjuntos, podemos considerar a frequência conjunta das duas classes:

$$\hat{p} \left(B_j^{(q)} \cup B_{j'}^{(q)} \right) = \frac{n_{\cdot,j}^{(q)} + n_{\cdot,j'}^{(q)}}{I} = p_{\cdot,j}^{(q)} + p_{\cdot,j'}^{(q)},$$

e a frequência condicional da classe $B_j^{(q)}$ dada $B_j^{(q)} \cup B_{j'}^{(q)}$:

$$\hat{p}\left(B_j^{(q)}|B_j^{(q)} \cup B_{j'}^{(q)}\right) = \frac{n_{.j}^{(q)}}{n_{.j}^{(q)} + n_{.j'}^{(q)}} = \frac{p_{.j}^{(q)}}{p_{.j}^{(q)} + p_{.j'}^{(q)}}.$$

Temos então

$$d^2(A_i, A_{i'}) = \frac{1}{Q} \sum_q \frac{\delta_{ii'}}{\hat{p}\left(B_{j(i,q)}^{(q)} \cup B_{j'(i',q)}^{(q)}\right) \hat{p}\left(B_{j(i,q)}^{(q)}|B_{j(i,q)}^{(q)} \cup B_{j'(i',q)}^{(q)}\right) \hat{p}\left(B_{j'(i',q)}^{(q)}|B_{j(i,q)}^{(q)} \cup B_{j'(i',q)}^{(q)}\right)}$$

e a interpretação é a mesma que no caso particular. Por outro lado,

$$\begin{aligned} d^2\left(B_j^{(q)}, B_{j'}^{(q')}\right) &= \sum_{i=1}^I \frac{1}{1/I} \left(\frac{n_{ij}^{(q)}}{n_{.j}^{(q)}} - \frac{n_{ij'}^{(q')}}{n_{.j'}^{(q')}} \right)^2 = \\ &= I \sum_{i=1}^I \left(\frac{n_{ij}^{(q)}}{[n_{.j}^{(q)}]^2} + \frac{n_{ij'}^{(q')}}{[n_{.j'}^{(q')}]^2} - 2 \frac{n_{ij}^{(q)} n_{ij'}^{(q')}}{n_{.j}^{(q)} n_{.j'}^{(q')}} \right) = \\ &= I \left(\frac{1}{n_{.j}^{(q)}} + \frac{1}{n_{.j'}^{(q')}} - 2 \frac{\sum_{i=1}^I n_{ij}^{(q)} n_{ij'}^{(q')}}{n_{.j}^{(q)} n_{.j'}^{(q')}} \right) = \\ &= I \frac{n_{.j}^{(q)} + n_{.j'}^{(q')} - 2 \sum_{i=1}^I n_{ij}^{(q)} n_{ij'}^{(q')}}{n_{ij}^{(q)} n_{ij'}^{(q')}}. \end{aligned}$$

Chamando-se $\# B_j^{(q)} = n_{.j}^{(q)}$ = número de indivíduos que responderam o item j da pergunta q , $B_j^{(q)} \cap B_{j'}^{(q')}$ ao conjunto dos indivíduos que responderam simultaneamente o item j da pergunta q e o item j' da pergunta q' , e finalmente denotando-se a diferença simétrica de $B_j^{(q)}$ e $B_{j'}^{(q')}$ por $B_j^{(q)} \triangle B_{j'}^{(q')}$, isto é, o conjunto dos indivíduos que responderam o item j da pergunta q , mas não o item j' da pergunta q' e vice-versa, temos

$$d^2(B_j^{(q)}, B_{j'}^{(q')}) = I \frac{\# B_j^{(q)} + \# B_{j'}^{(q')} - 2\# B_j^{(q)} \cap B_{j'}^{(q')}}{\# B^{(q)} \# B^{(q')}} =$$

$$= I \frac{\#(B_j^{(q)} \triangle B_{j'}^{(q')})}{\#B_j^{(q)} \#B_{j'}^{(q')}} = \frac{\#(B_j^{(q')} \triangle B_{j'}^{(q')})/I}{p_{.j}^{(q)} p_{.j'}^{(q')}}.$$

No caso $q = q'$, $\#(B_j^{(q)} \cap B_{j'}^{(q')}) = 0$ para $j \neq j'$ e temos:

$$\begin{aligned} d^2(B_j^{(q)}, B_{j'}^{(q)}) &= \\ &= \frac{1}{\hat{p}(B_j^{(q)} \cup B_{j'}^{(q)}) \hat{p}(B_j^{(q)} | B_j^{(q)} \cup B_{j'}^{(q)}) \hat{p}(B_{j'}^{(q)} | B_j^{(q)} \cup B_{j'}^{(q)})}. \end{aligned}$$

Com respeito às representações simultâneas, temos:

$$F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^{J_1+\dots+J_Q} \frac{p_{ij}}{p_{i.}} G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{Q} \sum_{q=1}^Q G_\alpha(j(i, q))$$

e

$$G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^I \frac{p_{ij}}{p_{.j}} F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^I \frac{n_{ij}}{n_{.j}} F_\alpha(i).$$

Logo a α -ésima coordenada do i -ésimo indivíduo é a menos de uma dilatação, a média aritmética das α -ésimas coordenadas dos itens por ele respondido e a α -ésima coordenada do item j é a menos de uma dilatação, a média aritmética das α -ésimas coordenadas dos indivíduos que a responderam.

(iii) Tabelas de intensidade

Consideremos I indivíduos A_1, \dots, A_I , por exemplo, estudantes, e sejam $B^{(1)}, \dots, B^{(Q)}$, Q critérios, por exemplo, disciplinas escolares. Suponhamos que as notas em cada disciplina variem de 0 a T , e seja x_{iq}^+ a nota do i -ésimo indivíduo na q -ésima disciplina. Assim como consideramos essa nota de “bondade” x_{iq}^+ , podemos considerar uma nota de “ruindade” $x_{iq}^- = T - x_{iq}^+$ e considerar a seguinte tabela

$A \setminus B$	$B^{(1)+}$	$B^{(1)-}$	\dots	$B^{(Q)+}$	$B^{(Q)-}$	
A_1	x_{11}^+	x_{11}^-		x_{1Q}^+	x_{1Q}^-	QT
\vdots						
A_I	x_{I1}^+	x_{I1}^-		x_{IQ}^+	x_{IQ}^-	QT
	$x_{.1}^+$	$x_{.1}^-$		$x_{.Q}^+$	$x_{.Q}^-$	IQT

Essa tabela é uma generalização da tabela de incidência, caso $J_q = 2$, para todo q , e do princípio de se registrar ao mesmo tempo presença e ausência. Ela tem a propriedade de que os indivíduos têm o mesmo peso, tornando os perfis dos indivíduos comparáveis.

Interpretemos as distâncias.

$$\begin{aligned}
 d^2(A_i, A_{i'}) &= \sum_{q=1}^Q \frac{IQT}{x_{\cdot q}^+} \left[\frac{x_{iq}^+/IQT}{1/I} - \frac{x_{i'q}^+/IQT}{1/I} \right]^2 + \\
 &\quad + \sum_{q=1}^Q \frac{IQT}{x_{\cdot q}^-} \left[\frac{x_{iq}^-/IQT}{1/I} - \frac{x_{i'q}^-/IQT}{1/I} \right]^2 = \\
 &= \frac{I}{QT} \sum_{q=1}^Q \left(\frac{1}{x_{\cdot q}^+} + \frac{1}{x_{\cdot q}^-} \right) (x_{iq}^+ - x_{i'q}^+)^2 = \frac{I^2}{Q} \sum_{q=1}^Q \frac{1}{x_{\cdot q}^+} (IT - x_{\cdot q}^+) (x_{iq}^+ - x_{i'q}^+)^2 = \\
 &= \frac{1}{Q} \sum_{q=1}^Q \frac{1}{\frac{x_{\cdot q}^+}{IT} \left(1 - \frac{x_{\cdot q}^+}{IT} \right)} \left(\frac{x_{iq}^+}{T} - \frac{x_{i'q}^+}{T} \right)^2.
 \end{aligned}$$

De novo, encontramos o inverso da variância de uma binomial ponderando o quadrado da diferença das notas normalizadas dos indivíduos A_i e $A_{i'}$. Observemos que o caso da tabela de incidência é um caso particular do caso acima.

Desse modo, novamente, o quadrado da distância entre os indivíduos A_i e $A_{i'}$ é a média aritmética dos quadrados das distâncias em relação a cada disciplina.

Por outro lado, $d^2(B^{(q)+}, B^{(q')++}) = I \sum_{i=1}^I \left(\frac{x_{iq}^+}{x_{\cdot q}^+} - \frac{x_{iq'}^+}{x_{\cdot q'}^+} \right)^2$, que é a menos de uma constante, o quadrado da distância euclidiana entre os perfis

de $B^{(q)+}$ e $B^{(q')-}$. Analogamente para $(-)$ no lugar de $(+)$.

$$\begin{aligned} d^2(B^{(q)+}, B^{(q)-}) &= I \sum_{i=1}^I \left(\frac{x_{iq}^+}{x_{\cdot q}^+} - \frac{x_{iq}^-}{x_{\cdot q}^-} \right)^2 = \\ &= \frac{IT^2}{[x_{\cdot q}^+(IT - x_{\cdot q}^+)]^2} \sum_{i=1}^I (Ix_{iq}^+ - x_{\cdot q}^+)^2 = \\ &= \frac{1}{\left[\frac{x_{\cdot q}^+}{IT} \left(1 - \frac{x_{\cdot q}^+}{IT} \right) \right]^2} \frac{1}{I} \sum_{i=1}^I \left(\frac{x_{iq}^+}{T} - \frac{x_{\cdot q}^+}{IT} \right)^2. \end{aligned}$$

Assim, o quadrado da distância entre $B^{(q)+}$ e $B^{(q)-}$ é a razão entre a variância amostral das notas normalizadas da disciplina $B^{(q)}$ e o quadrado da variância da binomial com parâmetro $x_{\cdot q}^+/IT$, que é a média das notas normalizadas.

Quanto à representação simultânea temos:

$$G_{\alpha}^{+}(q) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^I \frac{x_{iq}^{+}}{x_{\cdot q}^{+}} F_{\alpha}(i)$$

e

$$G_{\alpha}^{-}(q) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^I \frac{x_{iq}^{-}}{x_{\cdot q}^{-}} F_{\alpha}(i).$$

$$\text{Logo } x_{\cdot q}^{+} G_{\alpha}^{+}(q) + x_{\cdot q}^{-} G_{\alpha}^{-}(q) = \frac{T}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^I F_{\alpha}(i) = 0.$$

Consequentemente, em qualquer eixo, $G_{\alpha}^{+}(q)$ e $G_{\alpha}^{-}(q)$ têm sinais diferentes e são inversamente proporcionais, em módulo, a $x_{\cdot q}^{+}$ e $x_{\cdot q}^{-}$.

Por outro lado,

$$\begin{aligned}
 F_\alpha(i) &= \frac{1}{\sqrt{\lambda_\alpha}} \left[\sum_{q=1}^Q \frac{x_{iq}^+/IQT}{1/I} G_\alpha^+(q) + \sum_{q=1}^Q \frac{x_{iq}^-/IQT}{1/I} G_\alpha^-(q) \right] = \\
 &= \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{QT} \sum_{q=1}^Q \left[x_{iq}^+ G_\alpha^+(q) + x_{iq}^- G_\alpha^-(q) \right] = \\
 &= \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{Q} \sum_{q=1}^Q \frac{I x_{iq}^+ - x_{\cdot q}^+}{IT - x_{\cdot q}^+} G_\alpha^+(q) = \\
 &= \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{Q} \sum_{q=1}^Q \left[\frac{x_{iq}^+ - \frac{x_{\cdot q}^+}{I}}{T - \frac{x_{\cdot q}^+}{I}} \right] G_\alpha^+(q) = \\
 &= \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{Q} \sum_{q=1}^Q \left[\frac{\frac{x_{iq}^+}{T} - \frac{x_{\cdot q}^+}{IT}}{1 - \frac{x_{\cdot q}^+}{IT}} \right] G_\alpha^+(q).
 \end{aligned}$$

3.5 Relação entre a tabela de contingência e a tabela de incidência

Consideremos I indivíduos A_1, \dots, A_I e dois critérios $B^{(1)}$ e $B^{(2)}$, respectivamente com J_1 e J_2 itens exclusivos.

Podemos considerar então a tabela de contingência com J_1 linhas, J_2 colunas e $n_{jj'} = \# B_j^{(1)} \cap B_{j'}^{(2)}$ e a tabela de incidência com I linhas e $J_1 + J_2$ colunas conforme figuras a seguir.

$B^{(1)} \setminus B^{(2)}$	$B_1^{(2)}$	\dots	$B_{J_2}^{(2)}$	
$B_1^{(1)}$	n_{11}		n_{1J_2}	$n_{\cdot 1}$
\vdots				
$B_{J_1}^{(1)}$	$n_{J_1 1}$		$n_{J_1 J_2}$	$n_{J_1 \cdot}$
	$n_{\cdot 1}$		$n_{\cdot J_2}$	n

$A \setminus B$	$B_1^{(1)}$	\dots	$B_{J_1}^{(1)}$	$B_1^{(2)}$	\dots	$B_{J_2}^{(2)}$	
A_1							2
\vdots							\vdots
A_I							2
	$n_{.1}^{(1)}$		$n_{.J_1}^{(1)}$	$n_{.1}^{(2)}$		$n_{.J_2}^{(2)}$	$2I$

Essas duas tabelas estão em correspondência 1-1, a menos de permutação de linhas na tabela de incidência.

Chamemos de B_i a tabela de incidência A por $B^{(i)}$, $i = 1, 2$, e de $B = [B_1; B_2]$ a tabela de incidência considerada obtida pela justaposição das tabelas B_1 e B_2 . Segue que a tabela de contingência é dada por $B'_1 B_2$.

(i) **Relação entre correlação canônica sobre $B[[B_1; B_2]$ e os escores de Fisher sobre $B'_1 B_2$.**

Façamos formalmente uma análise de correlações canônicas considerando os grupos de variáveis formados pelas colunas das tabelas B_1 e B_2 .

Queremos então obter vetores x $J_1 \times 1$ e y $J_2 \times 1$ tais que maximizem a correlação $(B_1 x, B_2 y)$ sujeitos às restrições $EB_1 x = EB_2 y = 0$ e $\text{Var } B_1 x = \text{Var } B_2 y = 1$.

Mas

$$EB_1 x = \frac{1}{I} \sum_{i=1}^I (B_1 x)_i = \frac{1}{I} \sum_{j=1}^{J_1} n_{j.} x_j = \sum_{j=1}^{J_1} p_{j.} x_j = 0$$

e

$$\text{Var } B_1 x = \frac{1}{I} \sum_{i=1}^I (B_1 x)^2 = \frac{1}{I} \sum_{j=1}^{J_1} n_{j.} x_j^2 = \sum_{j=1}^{J_1} p_{j.} x_j^2 = 1.$$

Analogamente

$$EB_2 y = \sum_{j=1}^{J_2} p_{.j} y_j = 0 \text{ e } \text{Var } B_2 y = \sum_{j=1}^{J_2} p_{.j} y_j^2 = 1.$$

E correlação

$$\begin{aligned}(B_1 x, B_2 y) &= \frac{1}{I} \sum_{i=1}^I (B_1 x)_i (B_2 y)_i = \frac{1}{I} \sum_{j=1}^{J_1} \sum_{j'=1}^{J_2} n_{jj'} x_j y_{j'} = \\ &= \sum_{j=1}^{J_1} \sum_{j'=1}^{J_2} p_{jj'} x_j y_{j'}.\end{aligned}$$

Portanto recaímos no problema dos escores de Fisher para a tabela de contingência $B'_1 B_2$. É claro que podemos obter outras variáveis canônicas da maneira usual.

- (ii) Relação entre as análises de correspondência da tabela de contingência $B'_1 B_2$ e da tabela de incidência $B = [B_1 : B_2]$.

Seja $P = \frac{1}{2I} B$, a tabela de incidência normalizada.

$$\begin{aligned}\text{Então } p_{i\cdot} &= \frac{2}{2I} \frac{1}{I}, \quad \Sigma_P^A = \text{diag} \left(\frac{1}{p_{1\cdot}}, \dots, \frac{1}{p_{I\cdot}} \right) = I I_{I \times I} \text{ e } p_{\cdot j}^{(i)} = \frac{n_{\cdot j}^{(i)}}{2I}, \\ i &= 1, 2, \quad \Sigma_P^B = \text{diag} \left(\frac{1}{p_{\cdot 1}^{(1)}}, \dots, \frac{1}{p_{\cdot j_1}^{(1)}}, \frac{1}{p_{\cdot 1}^{(2)}}, \dots, \frac{1}{p_{\cdot j_2}^{(2)}} \right).\end{aligned}$$

Logo

$$\begin{aligned}P_{\cdot j}^{(1)} &= \frac{1}{2} p_{j\cdot} \quad 1 \leq j \leq J_1, \\ P_{\cdot j}^{(2)} &= \frac{1}{2} p_{\cdot j} \quad 1 \leq j \leq J_2\end{aligned}$$

onde $p_{\cdot j}$ e $P_{\cdot j}$ referem-se às proporções na tabela $B'_1 B_2$ e $\Sigma_P^B = 2 \begin{pmatrix} \Sigma^{B_1} & 0 \\ 0 & \Sigma^{B_2} \end{pmatrix}$ onde $\Sigma^{B_1} = \text{diag} \left(\frac{1}{p_{1\cdot}}, \dots, \frac{1}{p_{J_1\cdot}} \right)$ e $\Sigma^{B_2} = \text{diag} \left(\frac{1}{p_{\cdot 1}}, \dots, \frac{1}{p_{\cdot J_2}} \right)$.

Logo

$$\begin{aligned}
 H_P^B &= \Sigma_P^B P' \Sigma_P^A P = \begin{pmatrix} \Sigma^{B_1} & 0 \\ 0 & \Sigma^{B_2} \end{pmatrix} \frac{1}{2I} B' I I_{I \times I} \frac{1}{2I} B = \\
 &= \frac{1}{2I} \begin{pmatrix} \Sigma^{B_1} & 0 \\ 0 & \Sigma^{B_2} \end{pmatrix} \begin{pmatrix} B'_1 B_1 & B'_1 B_2 \\ B'_2 B_2 & B'_2 B_2 \end{pmatrix} = \\
 &= \frac{1}{2} \begin{pmatrix} \Sigma^{B_1} & 0 \\ 0 & \Sigma^{B_2} \end{pmatrix} \begin{pmatrix} (\Sigma^{B_1})^{-1} & P_{B'_1 B_2} \\ P_{B'_1 B_2} & (\Sigma^{B_2})^{-1} \end{pmatrix} = \\
 &= \frac{1}{2} \begin{pmatrix} I_{J_1 \times J_1} & \Sigma^{B_1} P_{B'_1 B_2} \\ \Sigma^{B_2} P_{B'_1 B_2} & I_{J_2 \times J_2} \end{pmatrix} \quad \text{onde} \quad P_{B'_1 B_2} = \frac{1}{I} B'_1 B_2.
 \end{aligned}$$

Mas $H_P^B \begin{pmatrix} a \\ \cdots \\ b \end{pmatrix} = \mu \begin{pmatrix} a \\ \cdots \\ b \end{pmatrix}$ se e somente se

$$\begin{cases} \frac{1}{2} a + \frac{1}{2} \Sigma^{B_1} P_{B'_1 B_2} b = \mu a \\ \frac{1}{2} \Sigma^{B_2} P_{B'_1 B_2} a + \frac{1}{2} b = \mu b \end{cases}$$

se e somente se

$$\begin{cases} \Sigma^{B_1} P_{B'_1 B_2} b = (2\mu - 1)a \\ \Sigma^{B_2} P_{B'_1 B_2} a = (2\mu - 1)b \end{cases}$$

Reconhecemos então a representação simultânea das linhas e colunas da tabela de contingência $B'_1 B_2$. Logo $a_\alpha = u_\alpha^{B_1}$, $b_\alpha = u_\alpha^{B_2}$ e $2\mu_\alpha - 1 = \sqrt{\lambda_\alpha}$ ou seja $\mu_\alpha = \frac{1 + \sqrt{\lambda_\alpha}}{2}$.

A solução $a_\alpha = u_\alpha^{B_1}$, $b_\alpha = -u_\alpha^{B_2}$, $2\mu'_\alpha - 1 = -\sqrt{\lambda_\alpha}$ ou seja $\mu'_\alpha = \frac{1 - \sqrt{\lambda_\alpha}}{2}$ não é interessante, pois temos sempre $\mu'_\alpha \leq \mu_\alpha$ e estamos interessados nos maiores autovalores.

É importante notar que a α -ésima coordenada de $B_j^{(1)}$ na análise de correspondência da tabela de incidência $B = [B_1, B_2]$ é $\sqrt{\mu_\alpha} u_\alpha^{B_1}(j)$, enquanto que na análise da tabela de contingência $B_1' B_2$ é $\sqrt{\lambda_\alpha} u_\alpha^{B_1}(j)$. Da mesma maneira, para $B_j^{(2)}$, temos respectivamente $\sqrt{\mu_\alpha} u_\alpha^{B_2}(j)$ e $\sqrt{\lambda_\alpha} u_\alpha^{B_2}(j)$. Logo temos a mesma representação a menos de uma homotetia.

Na análise de correspondência da tabela de incidência $B = [B_1, B_2]$, temos também a representação dos indivíduos, e a coordenada do indivíduo A_i no α -ésimo eixo é

$$F_\alpha(i) = \frac{1}{\sqrt{\mu_\alpha}} \frac{1}{2} [\sqrt{\mu_\alpha} u_\alpha^{B_1}(j) + \sqrt{\mu_\alpha} u_\alpha^{B_2}(j')] = \frac{1}{2} [u_\alpha^{B_1}(j) + u_\alpha^{B_2}(j')]$$

onde $B_j^{(1)}$ e $B_j^{(2)}$ são os itens escolhidos pelo indivíduo A_i .

3.6 Alguns comentários sobre a tabela de incidência

Seja $B = [B_1; \dots; B_Q]$ uma tabela de incidência formada por I indivíduos A_1, \dots, A_I e Q critérios $B^{(1)}, \dots, B^{(Q)}$, com J_1, \dots, J_q itens exclusivos, respectivamente.

Podemos formar a tabela simétrica

$$B' B = \begin{pmatrix} B_1' B_1 & \dots & B_1' B_Q \\ \vdots & & \vdots \\ B_Q' B_1 & \dots & B_Q' B_Q \end{pmatrix}$$

composta de todas as tabelas 2×2 .

É fácil ver que no caso $Q = 2$, a correspondência entre B , B' e $B_1' B_2$ é 1-1 a menos de permutações nas linhas de B . Mas no caso $Q > 2$, isso não ocorre. Podemos ter tabelas de incidência B e C distintas (isto é, qualquer tabela obtida de B por permutação de linhas é diferente de C), e $B' B = C' C$. Isso é o mesmo que dizer que as distribuições marginais de ordem 2 não determinam as distribuições marginais de ordem maior que 2.

Veremos abaixo, que a análise de correspondência da tabela B é equivalente a uma generalização da análise de correlações canônicas sobre as colunas grupadas por B_1, \dots, B_Q . Antes, mostraremos que a análise de correspondência das tabelas B e $B'B$ são equivalentes.

Para a tabela B ,

$$P_B = \frac{1}{QI} B, \quad \Sigma_B^A = \text{diag} \left(\frac{1}{p_{1\cdot}}, \dots, \frac{1}{p_{I\cdot}} \right) = I I_{I \times I}$$

e

$$\Sigma_B^B = \text{diag} \left(\frac{1}{\tilde{p}_{\cdot 1}^{(1)}}, \dots, \frac{1}{\tilde{p}_{\cdot J_Q}^{(Q)}} \right) = Q \begin{pmatrix} \Sigma^{B_1} & & 0 \\ & \ddots & \\ 0 & & \Sigma^{B_Q} \end{pmatrix}; \quad \tilde{p}_{\cdot j}^{(q)} = \frac{\# B_j^{(q)}}{QI}$$

onde

$$\Sigma^{B_q} = \text{diag} \left(\frac{1}{p_{\cdot j}^{(q)}}, \dots, \frac{1}{p_{\cdot j_q}^{(q)}} \right), \quad p_{\cdot j}^{(q)} = \frac{\# B_j^{(q)}}{I}.$$

Logo

$$H_B^B = \Sigma_B^B P_B' \Sigma_B^A P_B = \Sigma_B^B \frac{1}{QI} B' I I_{I \times I} \frac{1}{QI} B = \frac{1}{Q^2 I} \Sigma_B^B B' B$$

e para a tabela $B'B$,

$$P_{B'B} = \frac{1}{Q^2 I} B' B, \quad \Sigma_{B'B}^A = \Sigma_{B'B}^B = Q \begin{pmatrix} \Sigma^{B_1} & & 0 \\ & \ddots & \\ 0 & & \Sigma^{B_Q} \end{pmatrix} = \Sigma_B^B.$$

Logo

$$H_{B'B}^B = \Sigma_{B'B}^B P_{B'B}' \Sigma_{B',B}^A P_{B'B} = \Sigma_B^B \frac{1}{Q^2 I} B' B \Sigma_B^B \frac{1}{Q^2 I} B' B = H_B^B H_B^B.$$

Consequentemente $H_{B'B}^B$ tem os mesmos autovetores que H_B^B e seus autovalores são os quadrados dos autovalores de H_B^B .

Concluimos, portanto, que as representações dos critérios $B_j^{(q)}$ são as mesmas nas duas análises, a menos de uma homotetia, pois se $(U_\alpha^B)' = ((u_1^B)', \dots, (u_Q^B)'), u_q$ vetor $J_q \times 1$, é o α -ésimo autovetor de H_B^B , correspondente ao autovetor μ_α , então a α -ésima coordenada de $B_j^{(q)}$ será

$\sqrt{\mu_\alpha} u_q^B(j)$ na análise da tabela B e será $\mu_\alpha u_q^B(j)$ na análise da tabela $B'B$.

Pela equivalência da análise de correspondência de uma tabela com os escores de Fisher e usando-se o fato de que $B'B$ é uma matriz simétrica, sabemos que U_1^B maximiza $\frac{1}{Q^2 I} U' B' B U$ em $(U^B)' = ((u_1^B)', \dots, (u_Q^B)')$ sujeito às restrições $EU = 0$ e $\text{Var } U = U'(\Sigma_{B'B}^B)^{-1}U = 1$, ou seja maximiza $\frac{1}{Q^2 I} \sum_{q,q'} u_q' B_q' B_{q'} u_{q'}$ sujeito às restrições

$$U'p. = \frac{1}{Q} \sum_{q=1}^Q u_q' p^{(q)} = 0$$

e

$$\frac{1}{Q} \sum_{q=1}^Q u_q' (\Sigma^{B_q})^{-1} u_q = 1$$

$$\text{onde } p' = \frac{1}{Q} (p^{(1)'}, \dots, p^{(Q)'}) \quad \text{e} \quad p^{(q)'} = (p_{\cdot 1}^{(q)}, \dots, p_{\cdot J_Q}^{(q)}).$$

Mostremos que $(u_q^B)' p^{(q)} = 0$ para todo $q = 1, \dots, Q$.

Seja u_1^A o autovetor de H_B^A associado ao 1º autovalor μ_1 . Lembremos que $(u_1^A)' p^{(A)} = \frac{1}{I} \sum_i u_1^A(i) = 0$.

Temos então:

$$\begin{aligned} (u_q^B)' p^{(q)} &= \sum_{j=1}^{J_q} p_{\cdot j}^{(q)} u_q^B(j) = \sum_{j=1}^{J_q} p_{\cdot j}^{(q)} \frac{1}{\sqrt{\mu_1}} \sum_{i=1}^I \frac{p_{ij}^{(q)}}{p_{\cdot j}^{(q)}} \mu_1(i) = \\ &= \frac{1}{\sqrt{\mu_1}} \sum_{i=1}^I \left(\sum_{j=1}^{J_q} p_{ij}^{(q)} \right) u_1^A(i) = \frac{1}{\sqrt{\mu_1}} \frac{1}{I} \sum_{i=1}^I \mu_1^A(i) = 0. \end{aligned}$$

Logo tanto faz considerarmos a restrição $U'p. = 0$ ou $u_q' p^{(q)} = 0$, $q = 1, \dots, Q$. De fato, a restrição $U'p. = 0$ engloba um maior numero de vetores U , e contém todos os vetores que satisfazem $u_q' p^{(q)} = 0$, $q = 1, \dots, Q$.

Por outro lado, voltando à tabela $B = [B_1 : \dots : B_Q]$ podemos procurar vetores u_1, \dots, u_Q tais que

$$EU_q = u'_q p_{\cdot q} = 0$$

e

$$\frac{1}{Q} \sum_{q=1}^Q \text{Var } u_q = \frac{1}{Q} \sum_{q=1}^Q u'_q (\Sigma^{B_q})^{-1} u_q = 1$$

o que maximizem

$$\sum_{q,q'} \text{cor}(B_q u_q, B_{q'} u_{q'}) = \sum_{q,q'} u'_q B'_q B_{q'} u_{q'}.$$

Como vemos, obtemos a mesma solução que antes. É claro que podemos, de maneira usual, obter outras variáveis canônicas.

Observamos que a generalização de correlações canônicas acima é diferente das apresentadas por Kettering (1971). Uma dessas generalizações, também considerada por Horst (1965), usa a restrição $\text{Var } u_q = u'_q (\Sigma^{B_q})^{-1} u_q = 1$, em vez da restrição $\frac{1}{Q} \sum \text{Var } u_q = 1$.

No caso $Q = 2$, é fácil ver, que essas duas generalizações apresentam a mesma solução.

3.7 Generalização do escore de Fisher para tabelas de contingência justapostas

Vimos na seção 3(i) desse capítulo, a equivalência entre o escore de Fisher e a análise de correspondência para uma tabela de contingência. Como pode ser visto na demonstração dessa equivalência, a única coisa que precisamos é de uma tabela numérica com entradas não negativas e marginais positivas.

Consideremos então, a tabela obtida pela justaposição das tabelas de contingência $A \times B^{(q)}$, $q = 1, \dots, Q$, como na seção 4(i).

Procuramos então funções f e g definidas, respectivamente em $\{1, \dots, I\}$ e $\{1, \dots, J_1 + \dots + J_Q\}$, $g' = (g'_1, \dots, g'_Q)$, g_q assumindo

J_q valores, tais que maximizam

$$\begin{aligned} \text{cor}(f, g) &= \sum_{i,j} f(i)g(j)p_{ij} = \frac{1}{Q} \sum_{q=1}^Q \sum_i \sum_{j=1}^{J_q} f(i)g_q(j)p_{ij}^{(q)} = \\ &= \frac{1}{Q} \sum_{q=1}^Q \text{cor}_q(f, g_q), \end{aligned}$$

onde $\text{cor}(f, g)$ é a “correlação definida pela tabela justaposta e $\text{cor}_q(f, g_q)$ é a correlação definida pela tabela $A \times B^{(q)}$, f e g são normalizadas satisfazendo $Ef = Eg = 0$ e $\text{Var } g = 1$. Sabemos que a solução desse problema são $f = u^A$ e $g = u^B$.

$$\text{Agora } Eu^B = \sum_{j=1}^{J_1+\dots+J_Q} p_{\cdot j} u^B(j) = \frac{1}{Q} \sum_{q=1}^Q \sum_{j=1}^{J_q} p_{\cdot j}^{(q)} u_q^B(j) = 0.$$

$$\text{Se mostramos que } \sum_{j=1}^{J_Q} p_{\cdot j}^{(q)} u_q^B(j) = 0.$$

$$\begin{aligned} \sum_{j=1}^{J_q} p_{\cdot j}^{(q)} u_q^B(j) &= \sum_{j=1}^{J_q} p_{\cdot j}^{(q)} \frac{1}{\sqrt{\lambda_1}} \sum_{i=1}^I \frac{p_{ij}}{p_{\cdot j}} u^A(i) = \\ &= \frac{1}{\sqrt{\lambda_1}} \sum_{i=1}^I \sum_{j=1}^{J_q} p_{\cdot j}^{(q)} \frac{p_{ij}^{(q)}}{p_{\cdot j}^{(q)}} u^A(i) = \frac{1}{\sqrt{\lambda_1}} \sum_{i=1}^I p_{\cdot j} u^A(i) = 0. \end{aligned}$$

Temos também

$$\text{Var } g = g'(\Sigma^B)^{-1} g = \frac{1}{Q} \sum_{q=1}^Q g_q'(\Sigma^{B_q})^{-1} g_q = \frac{1}{Q} \text{Var } g_q,$$

onde $\text{Var}_q g_q$ é a variância de g_q definida pela tabela $A \times B^{(q)}$.

Logo com raciocínio análogo ao da seção anterior, vemos que esse problema é equivalente ao problema de procurar funções f, g_1, \dots, g_Q tais que maximizem $\sum_{q=1}^Q \text{cor}_q(f, g_q)$ e sujeitas às restrições $Ef = E_q g_q =$

$$0, \quad q = 1, \dots, Q, \quad \text{Var } f = 1 \text{ e } \frac{1}{Q} \sum_{q=1}^Q \text{Var}_q g_q = 1.$$

3.8 Interpretação dos resultados

3.8.1 Correlação, contribuição relativa e contribuição absoluta

Uma maneira natural de se interpretar os eixos, é olhar as correlações entre cada linha e coluna com os novos eixos. Lembramos que na análise de correspondência, fazemos duas componentes principais generalizadas.

Pela Proposição 2.3, temos

$$\text{cor}(B_j, F_\alpha) = \frac{(C^B u_\alpha^B)(j)}{\sqrt{\lambda_\alpha} s_j}.$$

Lembramos que

$$\begin{aligned} s_j^2 &= \sum_{i=1}^I p_{i\cdot} \left(\frac{p_{ij}}{p_{i\cdot}} - p_{\cdot j} \right)^2 = \sum_{i=1}^I \frac{1}{p_{i\cdot}} (p_{ij} - p_{i\cdot} p_{\cdot j})^2 = \\ &= p_{\cdot j}^2 \sum_{i=1}^I \frac{1}{p_{i\cdot}} \left(\frac{p_{ij}}{p_{\cdot j}} - p_{i\cdot} \right)^2 = p_{\cdot j}^2 d^2(B_j, \bar{B}) \end{aligned}$$

onde \bar{B} representa o centro de gravidade das variáveis B_j .

Por outro lado, $d^2(B_j, \bar{B}) = \sum_{\alpha} G_{\alpha}^2(j) = \sum_{\alpha} \left[(u_{\alpha}^A)' p_j^A \right]^2$, isto é, $d^2(B_j, \bar{B})$ é a soma dos quadrados das projeções de $B_j - \bar{B}$ sobre os eixos.

Podemos então considerar a razão

$$cr_{\alpha}(j) = \frac{G_{\alpha}^2(j)}{d^2(B_j, \bar{B})} = \frac{\left[(u_{\alpha}^A)' p_j^A \right]^2}{d^2(B_j, \bar{B})} = \frac{p_{\cdot j}^2 \left[(u_{\alpha}^A)' p_j^A \right]^2}{s_j^2},$$

que nos dá a parte do quadrado da distância de B_j ao centro de gravidade explicada pelo eixo α . À $cr_{\alpha}(j)$, Lebart, Morineau & Tabard (1977) chamam de contribuição relativa do fator α à variável B_j .

Lembremos que o quadrado da correlação entre duas variáveis aleatórias pode ser “interpretado como a parte da variância de uma delas explicada pela outra”. Mostraremos que

$$cr_{\alpha}(j) = \text{cor}^2(B_j, F_{\alpha}).$$

Sabemos que $u_\alpha^A = \frac{1}{\sqrt{\lambda_\alpha}} \Sigma^A P u_\alpha^B$.

Logo

$$P' u_\alpha^A = \frac{1}{\sqrt{\lambda_\alpha}} P' \Sigma^A P u_\alpha^B = \frac{1}{\sqrt{\lambda_\alpha}} (C^B + p_{\cdot j}^B (p_{\cdot j}^B)') u_\alpha^B = \frac{1}{\sqrt{\lambda_\alpha}} C^B u_\alpha^B.$$

Portanto,

$$(p_{\cdot j}^A)' u_\alpha^A = \frac{1}{p_{\cdot j}} (P' u_\alpha^A)(j) = \frac{1}{p_{\cdot j}} \frac{1}{\sqrt{\lambda_\alpha}} (C^B u_\alpha^B)(j)$$

e daí segue o resultado,

Outra medida utilizada na interpretação é tirada da relação

$$\lambda_\alpha = \text{Var } G_\alpha = \sum_j p_{\cdot j} G_\alpha^2(j).$$

Consideramos a razão

$$ca_\alpha(j) = \frac{p_{\cdot j} G_\alpha^2(j)}{\lambda_\alpha} = p_{\cdot j} [u_\alpha^B(j)]^2,$$

que nos dá a parte da variância do fator α explicada pela variável B_j . À $ca_\alpha(j)$, Lebart, Morineau & Tabard (1977) chamam de contribuição absoluta da variável B_j ao fator α .

Então, se $ca_\alpha(j)$ é grande, boa parte da variância do fator α é devida à variável B_j , enquanto que se $cr_\alpha(j)$ é grande, boa parte da distância da variável B_j ao centro de gravidade é devido ao fator α .

Expressimos essas medidas de maneira análoga para os A_i 's.

3.8.2 Alguns comentários sobre o valor das distâncias

Para a interpretação dos resultados, é importante saber quão bem representados nos subespaços fatoriais, estão as diversas linhas e colunas da tabela original. Isso, em parte, nos é fornecido pelas contribuições relativas dos eixos fatoriais às linhas e às colunas. Mas, é importante também ter uma ideia das distâncias das linhas e colunas ao centro e o que uma distância grande ou pequena significa.

Se \bar{A} representa o centro de gravidade das linhas A_1, \dots, A_I , \bar{A} é representado pelo vetor $(p_{\cdot 1}, \dots, p_{\cdot J})$ e

$$d^2(A_i, \bar{A}) = \sum_{j=1}^J \frac{1}{p_{\cdot j}} \left(\frac{p_{ij}}{p_{i\cdot}} - p_{\cdot j} \right)^2 = \frac{1}{p_{i\cdot}} \sum_{j=1}^J \frac{p_{ij}^2}{p_{\cdot j}} - 1.$$

Portanto

$$\frac{1}{p_{i\cdot}} \min_j \left\{ \frac{p_{ij}}{p_{\cdot j}}; p_{ij} > 0 \right\} - 1 \leq d^2(A_i, \bar{A}) \leq \frac{1}{p_{i\cdot}} \max_j \left\{ \frac{p_{ij}}{p_{\cdot j}} \right\} - 1.$$

Observemos que $d^2(A_i, \bar{A}) = 0$ se e somente se, $\frac{p_{ij}}{p_{\cdot j}} = p_{i\cdot}$ para todo j .

Logo se $d^2(A, \bar{A})$ pequeno, $\frac{p_{ij}}{p_{\cdot j}} \approx p_{i\cdot}$ para todo j .

Portanto, se $\min_j \left\{ \frac{p_{ij}}{p_{\cdot j}}; p_{ij} > 0 \right\}$ “grande”, isto é, muito maior que $p_{i\cdot}$, então existe associação grande positiva entre A_i e alguns B_j 's. Deverá também existir uma associação grande negativa entre A_i e alguns B_j 's no sentido de que $p_{ij} = 0$ pois se $\frac{p_{ij}}{p_{\cdot j}} \geq p_{i\cdot}$ para todo j e maior para algum j , então

$$\sum_j p_{ij} > \sum_j p_{i\cdot} p_{\cdot j} \Rightarrow p_{i\cdot} > p_{i\cdot}$$

o que nos dá uma contradição.

No caso extremo $\min_j \left\{ \frac{p_{ij}}{p_{\cdot j}}; p_{ij} > 0 \right\}$ “pequeno”, isto é, igual ou somente um pouco maior que $p_{i\cdot}$, então $d^2(A_i, \bar{A})$ pequeno e todos os $\frac{p_{ij}}{p_{\cdot j}}$ são aproximadamente iguais e a classe A_i está próximo do centro \bar{A} .

No caso da matriz de incidência, lembramos que $n_{ij} = 1$ ou 0, $p_{i\cdot} = \frac{Q}{QI} = \frac{1}{I}$ e $p_{\cdot j} = \frac{n_{\cdot j}}{QI}$.

Portanto

$$\min_j \left\{ \frac{n_{ij}}{n_{\cdot j}}; n_{ij} > 0 \right\} = \min_j \left\{ \frac{1}{\#B_j}; n_{ij} = 1 \right\} = \frac{1}{\max_j \{\#B_j; n_{ij} = 1\}}.$$

No caso extremo em que $\max\{\# B_j; n_{ij} = 1\} = 1$, o indivíduo A_i respondeu todas as perguntas de maneira diferente dos outros indivíduos e nesse caso $d^2(a_i, \bar{A}) = I - 1$.

Analogamente

$$\max_j \left\{ \frac{n_{ij}}{n_{.j}} \right\} = \max_j \left\{ \frac{1}{\# B_j}; n_{ij} = 1 \right\} = \frac{1}{\min_j \{\# B_j; n_{ij} = 1\}}.$$

Logo $\max_j \left\{ \frac{n_{ij}}{n_{.j}} \right\}$ pequeno implica $\min_j \{\# B_j; n_{ij} = 1\}$ grande, o que significa que todos os itens respondidos pelo indivíduo i , foram respondidos também por muitos outros indivíduos.

Analogamente, se \bar{B} representa o centro dos B_j 's, então $d^2(B_j, \bar{B}) = \frac{1}{p_{.j}^2} \sum_{i=1}^I \frac{p_{ij}^2}{p_{i.}} - 1$, e temos

$$\frac{1}{p_{.j}} \min_i \left\{ \frac{p_{ij}}{p_{i.}}; p_{ij} > 0 \right\} - 1 \leq d^2(B_j, \bar{B}) \leq \frac{1}{p_{.j}} \max_i \left\{ \frac{p_{ij}}{p_{i.}} \right\} - 1.$$

No caso da matriz de incidência, $p_{i.} = \frac{1}{I}$ é constante e portanto

$$\min_i \left\{ \frac{p_{ij}}{p_{i.}}; p_{ij} > 0 \right\} = \max_i \left\{ \frac{p_{ij}}{p_{i.}} \right\} = \frac{1/QI}{1/I} = \frac{1}{Q}.$$

Logo

$$d^2(B_j, \bar{B}) = \frac{1}{p_{.j} Q} - 1 = \frac{I}{n_{.j}} - 1.$$

Portanto, no caso em que, por exemplo, B_j foi respondido por somente 1 indivíduo, $d^2(B_j, \bar{B}) = I - 1$.

3.8.3 Uso de linhas e colunas suplementares

Como na análise de componentes principais, podemos utilizar linhas e colunas suplementares com o fim de ajudar na interpretação dos eixos, ou com o objetivo de estudar a relação entre esses elementos suplementares e os originais.

Podemos calcular as coordenadas desses elementos suplementares, muito facilmente. Por exemplo, se B_{J+1} é uma coluna suplementar,

$$G_{\alpha}(J+1) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^I \frac{n_{iJ+1}}{n_{\cdot J+1}} F_{\alpha}(i),$$

e se A_{I+1} é uma linha suplementar,

$$F_{\alpha}(I+1) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{j=1}^J \frac{n_{I+1,j}}{n_{(I+1)\cdot}} G_{\alpha}(j).$$

3.8.4 Reconstituição da tabela P

Lembrando o Teorema 3.1, temos $S^B = LL'$, onde $L = (\Sigma^A)^{1/2} P(\Sigma^B)^{1/2} = \sum_{\alpha} \sqrt{\lambda_{\alpha}} d_{\alpha}^A (d_{\alpha}^B)'$.

Logo

$$\begin{aligned} P &= (\Sigma^A)^{-1/2} \left[\sum_{\alpha} \sqrt{\lambda_{\alpha}} d_{\alpha}^A (d_{\alpha}^B)' \right] (\Sigma^B)^{-1/2} = \\ &= (\Sigma^A)^{-1} \left[\sum_{\alpha} \sqrt{\lambda_{\alpha}} u_{\alpha}^A (u_{\alpha}^B)' \right] (\Sigma^B)^{-1}. \end{aligned}$$

Usando-se o fato de que 1_J e 1_I são respectivamente, autovetores de H^B e H^A associados ao autovetor 1, temos

$$p_{ij} = p_{i\cdot} p_{\cdot j} \left[1 + \sum_{\alpha} \sqrt{\lambda_{\alpha}} u_{\alpha}^A(i) u_{\alpha}^B(j) \right]$$

ou seja

$$p_{ij} = p_{i\cdot} p_{\cdot j} \left[1 + \sum_{\alpha} \sqrt{\lambda_{\alpha}} F_{\alpha}(i) G_{\alpha}(j) \right].$$

Desse modo, a expressão

$$\frac{p_{ij}}{p_{i\cdot} p_{\cdot j}} - 1 = \sum_{\alpha} \frac{1}{\sqrt{\lambda_{\alpha}}} F_{\alpha}(i) G_{\alpha}(j)$$

mede a associação entre a linha i e a coluna j .

3.9 Relação com o teste χ^2

Seja uma tabela de contingência $(n_{ij})_{1 \leq i \leq I, 1 \leq j \leq J}$, $\sum_{i,j} n_{ij} = n$. Então temos

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J n \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})^2}{p_{i\cdot} p_{\cdot j}} = n \left(\sum_{i=1}^I \sum_{j=1}^J \frac{p_{ij}^2}{p_{i\cdot} p_{\cdot j}} - 1 \right).$$

$$\text{Mas } \text{Tr } \Sigma^B C^B = \sum_{i=1}^I \sum_{j=1}^J \frac{p_{ij}^2}{p_{i\cdot} p_{\cdot j}} - 1 \quad \text{e} \quad \text{Tr } \Sigma^B S^B = \sum_{i=1}^I \sum_{j=1}^J \frac{p_{ij}^2}{p_{i\cdot} p_{\cdot j}}.$$

Logo $\chi^2 = n \text{Tr } \Sigma^B C^B = n(\text{Tr } S^B - 1)$ ou a medida de associação $\frac{\chi^2}{n} = \text{Tr } \Sigma^B C^B = \text{Tr } S^B - 1$.

Logo se há independência entre as categorias A e B , $p_{ij} \approx p_{i\cdot} p_{\cdot j}$ e $\chi^2 = n \text{Tr } \Sigma^B C^B$ é pequeno.

Lembremos que

$$\text{Tr } \Sigma^B C^B = \sum_i p_{i\cdot} \|x_i^B\|_{\Sigma^B}^1 = \sum_{i \leq i'} p_{i\cdot} p_{i'\cdot} \|x_i^B - x_{i'}^B\|_{\Sigma^B}^2 = \sum_{\alpha=1}^J \lambda_\alpha$$

e portanto se as p primeiras componentes são tais que $\frac{\sum_{\alpha=1}^p \lambda_\alpha}{\sum_{\alpha=1}^J \lambda_\alpha}$ grande, essas p primeiras componentes “explicam” o valor χ^2 .

No caso de tabelas de contingência $A \times B^{(1)}, \dots, A \times B^{(Q)}$ justapostas, temos formalmente

$$\begin{aligned} \chi^2 &= n Q \text{Tr } \Sigma^B C^B = n Q \sum_{i \leq i'} p_{i\cdot} p_{i'\cdot} \|x_i^B - x_{i'}^B\|_{\Sigma^B}^2 = \\ &= n Q \sum_{i \leq i'} p_{i\cdot} p_{i'\cdot} d^2(A_i, A_{i'}). \end{aligned}$$

Mas como visto antes $p_{i\cdot} = p_{i\cdot}^{(q)}$ para $q = 1, \dots, Q$ e

$$d^2(A_i, A_{i'}) = \frac{1}{Q} \sum_{q=1}^Q d^{(q)2}(A_i, A_{i'}).$$

Logo

$$\begin{aligned}
 \chi^2 &= n \sum_{i \leq i'}^Q p_{i.} p_{i' \cdot} \cdot \frac{1}{Q} \sum_{q=1}^Q d^{(q)^2} (A_i, A_{i'}) = \\
 &= n \sum_{q=1}^Q \sum_{i \leq i'} p_{i.}^{(q)} p_{i' \cdot}^{(q)} d^{(q)^2} (A_i, A_{i'}) = n \sum_{q=1}^Q \text{Tr } \Sigma^{B_q} C^{B_q} = \\
 &= n \sum_{q=1}^Q \frac{\chi^2(q)}{n}
 \end{aligned}$$

ou seja

$$\chi^2 = \sum_{q=1}^Q \chi^2(q), \text{ onde } \chi^2(q) \text{ é o valor } \chi^2 \text{ da tabela } A \times B^{(q)}.$$

É claro $\chi^{2(1)}, \dots, \chi^{2(Q)}$ não são independentes.

Também temos

$$\text{Tr } \Sigma^B C^B = \frac{1}{Q} \sum_{q=1}^Q \text{Tr } \Sigma^{B_q} C^{B_q}.$$

Mas $\dim \Sigma^B C^B = J_1 + \dots + J_Q$, logo devemos esperar que a “ex-
plicação” $\frac{\sum_{\alpha=1}^p \lambda_\alpha}{\sum_{j=1}^{J_1+\dots+J_Q} \lambda_\alpha}$ seja menor que no caso de cada tabela $A \times B^{(q)}$,

porque deverá haver um número grande de autovalores pequenos, que de certa maneira, podem ser considerados como “ruído”.

No caso da tabela de incidência $B = [B_1 : \dots : B_Q]$, podemos mostrar resultado análogo, isto é,

$$\chi^2 = \sum_{q=1}^Q \chi^{2(q)} \quad \text{e} \quad \text{Tr } \Sigma^B C^B = \frac{1}{Q} \sum_{q=1}^Q \text{Tr } \Sigma^{B_q} C^{B_q}.$$

Nesse caso, podemos calcular exatamente $\text{Tr } \Sigma^B C^B$. Consideremos o caso $Q = 1$.

Então

$$\text{Tr } \Sigma^{B_q} C^{B_q} = \sum_{i=1}^I \sum_{j=1}^J \frac{p_{ij}^2}{p_{i\cdot} p_{\cdot j}} - 1.$$

Mas

$$p_{ij}^2 = \frac{n_{ij}^2}{I^2} = \frac{n_{ij}}{I^2} \text{ pois } n_{ij} = 1 \text{ ou } 0; p_{i\cdot} = \frac{1}{I} \text{ e } p_{\cdot j} = \frac{1}{I} \sum_{i=1}^I n_{ij}.$$

Logo

$$\text{Tr } \Sigma^B C^B = \sum_{j=1}^J \sum_{i=1}^I \frac{n_{ij}}{p_{\cdot j}} - 1 = J - 1.$$

E no caso geral

$$\text{Tr } \Sigma^B C^B = \frac{1}{Q} \sum_{q=1}^q (J_q - 1) = \frac{J}{Q} - 1.$$

3.10 Comentários e exemplo

Análise de Correspondência é uma técnica de exploração de dados, especialmente útil para visualizar dados coletados através de questionários utilizados rotineiramente em pesquisa na área de ciências sociais.

Os dados consistem em geral em respostas a perguntas cada uma com um número finito de respostas possíveis. Ou seja um conjunto de variáveis categóricas (também chamadas de “qualitativas”) cada uma com um certo número de níveis ou categorias.

Na linguagem do pacote R que usaremos para o exemplo seriam fatores. A parte mais útil para aplicações são as saídas gráficas. Estas saídas deveriam ser as mais “amigáveis” possível. Elas deveriam sugerir rapidamente ideias, interpretações, ou problemas com os dados que estão sendo analisados.

Exemplo.

Os dados deste Exemplo foram analisados utilizando o pacote FactoMineR do R.

Exemplo 1. Mercado de chá.

Os dados para este exemplo são provenientes de uma pesquisa feita com consumidores de Chá. Foram realizadas 300 entrevistas e o tema está relacionado com hábitos de consumo de Chá.

Os dados completos da pesquisa podem ser encontrados em
<http://factominer.free.fr/book/tea.csv>.

Para nosso exemplo só selecionamos algumas variáveis correspondentes às perguntas indicadas abaixo junto com as modalidades/categorias de resposta. (Deixamos a pergunta na lingua original: Inglês).

1. What kind of tea do you drink the most?

“black” “Earl Grey” “green”

2. How do you take your tea?

“alone” “lemon” “milk” “other”

3. How the tea you drink is packed?

“tea bag” “tea bag+unpackaged” “unpackaged”

4. Do you add sugar to your tea.

“No sugar” “sugar”

5. Where do you buy your tea?

“chain store” “chain store+tea shop” “tea shop”

6. Do you drink tea in the afternoon?

“always” “Not always”

A Figura 1 contém as primeiras 15 linhas do arquivo de dados:

#	Tea	How	how	sugar	where	always
1	black	alone	tea bag	sugar	chain store	Not.always
2	black	milk	tea bag	No.sugar	chain store	Not.always
3	Earl Grey	alone	tea bag	No.sugar	chain store	Not.always
4	Earl Grey	alone	tea bag	sugar	chain store	Not.always
5	Earl Grey	alone	tea bag	No.sugar	chain store	always
6	Earl Grey	alone	tea bag	No.sugar	chain store	Not.always
7	Earl Grey	alone	tea bag	No.sugar	chain store	Not.always
8	black	milk	tea bag	No.sugar	chain store	Not.always
9	Earl Grey	milk	tea bag+unpa	No.sugar	chain store+tea shop	Not.always
10	black	alone	tea bag+unpa	No.sugar	chain store+tea shop	Not.always
11	Earl Grey	alone	tea bag	No.sugar	chain store	Not.always
12	Earl Grey	alone	tea bag	sugar	chain store	Not.always
13	Earl Grey	milk	tea bag+unpa	No.sugar	chain store+tea shop	Not.always
14	black	milk	tea bag	sugar	chain store	Not.always
15	black	milk	tea bag	No.sugar	chain store+tea shop	Not.always
					
					
300	black	milk	tea bag	sugar	chain store	Not.always

A Figura 2 contém os valores dos autovalores e as percentagens da variância e seus valores acumulados. Podemos esperar valores menores em geral que no caso de Principal componentes dado que o número total de categorias costuma ser maior que o de variáveis.

Figura 2

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0,28	15,26	15,26
dim 2	0,26	14,06	29,32
dim 3	0,22	12,01	41,33
dim 4	0,19	10,25	51,58
dim 5	0,17	9,21	60,78
dim 6	0,16	8,93	69,71
dim 7	0,15	8,34	78,05
dim 8	0,14	7,55	85,6
dim 9	0,12	6,31	91,91
dim 10	0,09	4,7	96,61
dim 11	0,06	3,39	100

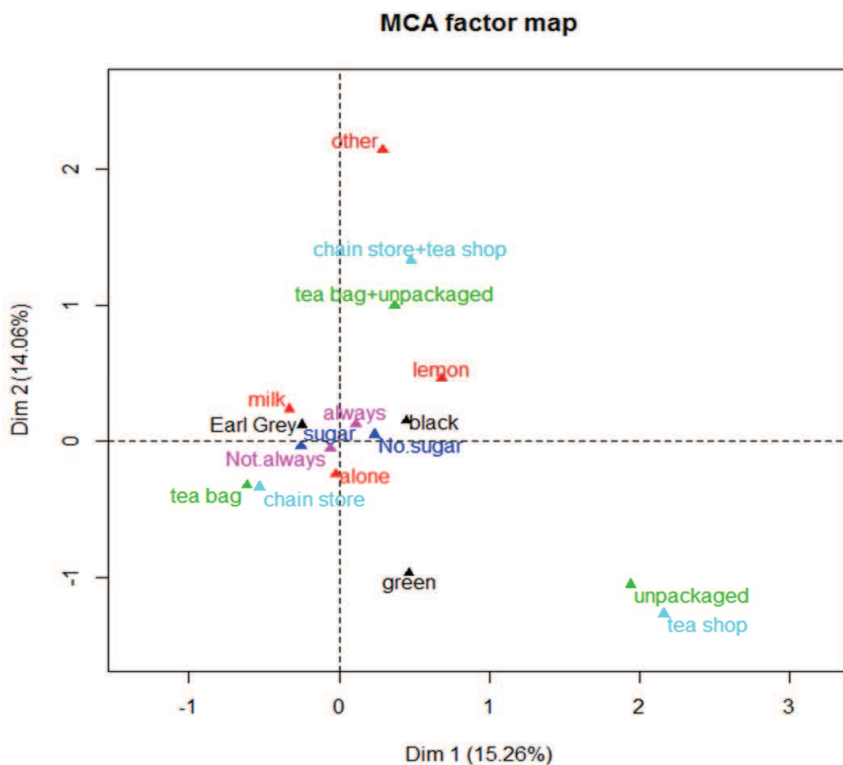
A Figura 3 contém as coordenadas das categorias a serem representadas no Mapa.

Figura 3

	Dim 1	Dim 2
black	0,450	0,140
Earl Grey	-0,250	0,110
green	0,460	-0,970
alone	-0,020	-0,250
lemon	0,680	0,460
milk	-0,330	0,230
other	0,290	2,140
tea bag	-0,620	-0,330
tea bag+unpackaged	0,370	1,000
unpackaged	1,940	-1,060
No.sugar	0,240	0,040
sugar	-0,250	-0,040
chain store	-0,530	-0,340
chain store+tea shop	0,480	1,330
tea shop	2,160	-1,270
always	0,110	0,120
Not.always	-0,060	-0,060

A Figura 4 contém o Mapa com a posição de cada uma das categorias. Cores diferentes indicam variáveis distintas.

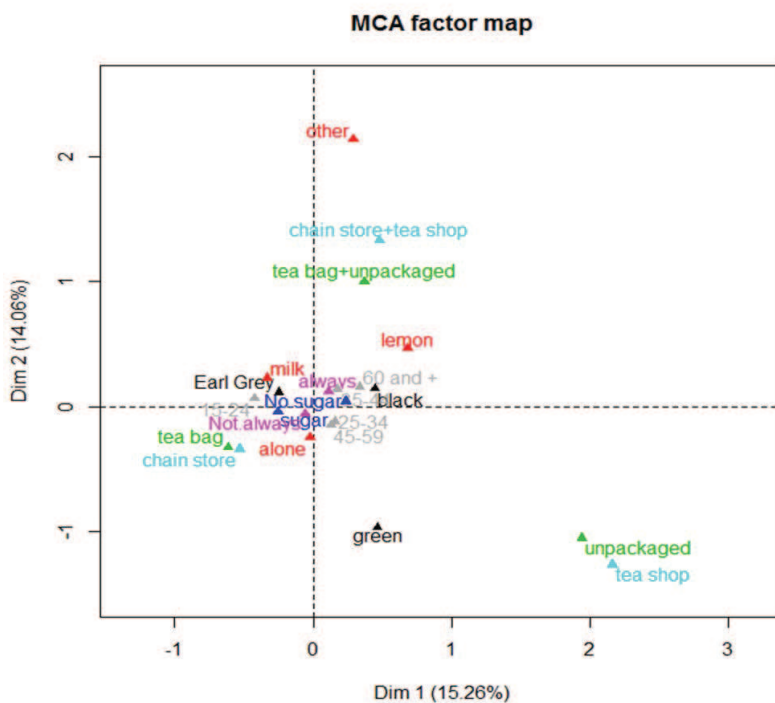
Figura 4



Como foi mencionado é possível agregar no Mapa de base, variáveis suplementares, ou seja variáveis que não entram diretamente no cálculo mas que podem ser projetadas no Mapa da solução. São úteis eventualmente para visualizar algumas possíveis relações entre variáveis.

Na Figura 5 incorporamos a variável Idade (age) ao Mapa. Na representação vemos por exemplo a maior afinidade do grupos de pessoas mais jovens (15-24) pelo consumo em Tea Bugs, com açúcar e a variedade (blend) Earl Gray.

Figura 5

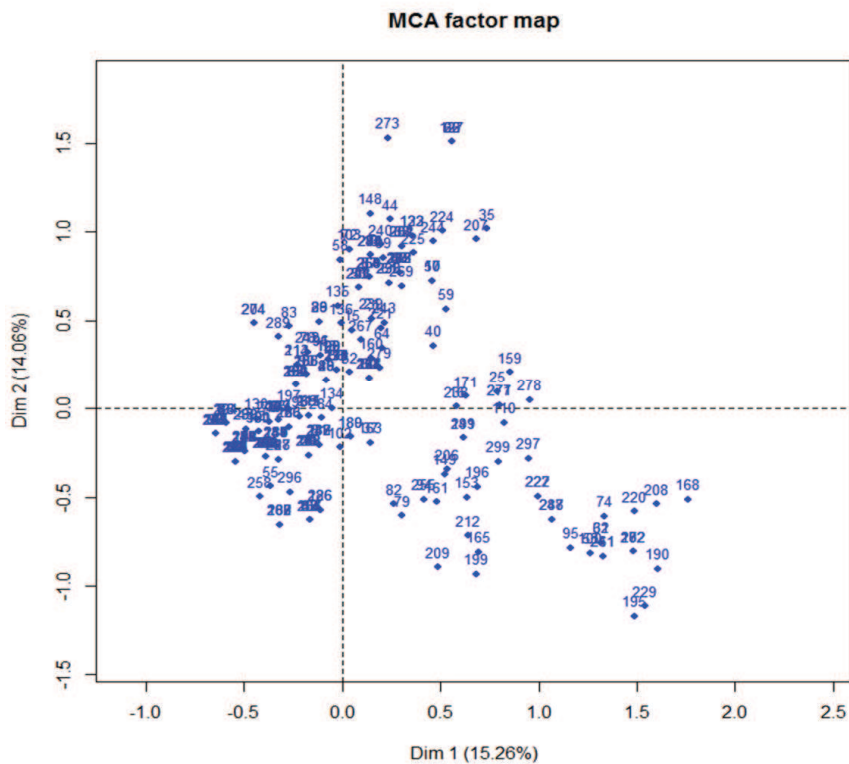


Em algumas circunstâncias, pode também ser de interesse olhar o posicionamento sobre as primeiras dimensões dos Indivíduos (ou seja as linhas do arquivo de dados).

Pode por exemplo sugerir a presença de segmentos, ou grupo de indivíduos com respostas similares.

A Figura 6 mostra o Mapa dos indivíduos neste exemplo. Pareceria sugerir a presença de mais de um segmento de consumidores.

Figura 6



Capítulo 4

Escalonamento Multivariado

4.1 Introdução

As técnicas de escalonamento multivariado permitem representar um conjunto de objetos como pontos de um espaço de dimensão relativamente pequena conhecendo as distâncias entre os objetos ou somente a ordenação dessas distâncias.

As técnicas de escalonamento multivariado foram desenvolvidas fundamentalmente por Shepard (1962) e Kruskal (1964) utilizando algoritmos de cálculo sofisticados. Neste capítulo apresentamos um enfoque alternativo, devido a Benzecri (1973) que dá uma solução aproximada a ótima, utilizando somente técnicas de Álgebra Linear.

Nos referiremos a esta última técnica como Escalonamento Multidimensional clássico ou métrico. Nos referiremos a versão posterior de Shepard (1962) e Kruskal (1964) como Escalonamento Multidimensional não métrico.

Empiricamente os resultados obtidos com este método, diferem geralmente muito pouco das soluções obtidas através do uso da técnica mais sofisticada de Kruskal (ver Benzecri 1973).

4.2

Consideremos um conjunto de I objetos X_i , $1 \leq i \leq I$. Suponhamos que não temos nenhuma representação numérica desses objetos e que somente é conhecida a distância ou “dissimilaridade” entre dois quaisquer deles. A distância entre X_i e $X_{i'}$ será denotada por $d_{ii'}$. Vamos supor que as seguintes relações são satisfeitas

- (i) $d_{ii'} = d_{i'i}$ para todo i, i' .
- (ii) $d_{ii'} \geq 0$ para todo i, i' .
- (iii) $d_{ii'} = 0$ se e somente se $i = i'$.

Seja $D_{ii'} = d_{ii'}^2$.

O problema do escalonamento multivariado consiste em representar cada objeto X_i por um ponto $F_i^{(p)} \in \mathbb{R}^p$ de maneira que $d_{ii'}$ seja aproximadamente $\|F_i^{(p)} - F_{i'}^{(p)}\|_I$.

Vamos supor, como no Capítulo 2, que cada X_i tem um peso μ_i com $\sum_{i=1}^I \mu_i = 1$. Então, o problema do escalonamento ótimo em \mathbb{R}^p , consiste em achar pontos $F_i^{(p)}$ em \mathbb{R}^p tais que

$$(4.1) \quad \sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \left| D_{ii'} - \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2 \right|$$

seja mínimo.

As técnicas desenvolvidas neste capítulo permitem achar soluções aproximadamente ótimas. O primeiro passo consistirá em representar cada X_i por um ponto $x_i \in \mathbb{R}^{I-1}$ de maneira que

$$(4.2) \quad D_{ii'} = \|x_i - x_{i'}\|_{\Sigma^*}^2$$

onde Σ^* será uma matriz não necessariamente definida não negativa.

Nota. Dada uma matriz Σ , $p \times p$, não necessariamente definida não negativa definimos $\langle a, b \rangle_{\Sigma} = a' \Sigma b$ e $\|a\|_{\Sigma}^2 = a' \Sigma a$, onde a e b são

elementos de \mathbb{R}^p . Os pontos x_i que encontraremos em \mathbb{R}^{I-1} satisfarão também a seguinte relação linear

$$(4.3) \quad \sum_{i=1}^I \mu_i x_i = 0$$

ou seja terão média 0.

Provaremos primeiro o seguinte

Lema 4.1. *Para que um conjunto de pontos x_i , $1 \leq i \leq I$, em \mathbb{R}^p que verifiquem (4.3) verifiquem também (4.2) é condição necessária e suficiente que*

$$(4.4) \quad \langle x_i, x_{i'} \rangle_{\Sigma^*} = \frac{1}{2} \left(-D_{ii'} + D_{i\cdot} + D_{i'\cdot} - D_{\cdot\cdot} \right)$$

$$\text{onde } D_{i\cdot} = \sum_{i'=1}^I \mu_{i'} D_{ii'}, \quad D_{\cdot\cdot} = \sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} D_{ii'}.$$

Demonstração. (4.2) é equivalente a

$$D_{ii'} = \|x_i\|_{\Sigma^*}^2 + \|x_{i'}\|_{\Sigma^*}^2 - 2\langle x_i, x_{i'} \rangle_{\Sigma^*}$$

e portanto equivalente a

$$(4.5) \quad \langle x_i, x_{i'} \rangle_{\Sigma^*} = \frac{1}{2} \left(-D_{ii'} + \|x_i\|_{\Sigma^*}^2 + \|x_{i'}\|_{\Sigma^*}^2 \right).$$

Multiplicando (4.5) por $\mu_{i'}$ e somando temos

$$(4.6) \quad \|x_i\|_{\Sigma^*}^2 = D_{i\cdot} - \sum_{i'=1}^I \mu_i \|x_i\|_{\Sigma^*}^2.$$

Multiplicando (4.6) por μ_i e somando temos

$$(4.7) \quad D_{\cdot\cdot} = 2 \sum_{i=1}^I \mu_i \|x_i\|_{\Sigma^*}^2.$$

Substituindo (4.6) e (4.7) em (4.5) temos (4.4). A suficiência decorre facilmente.

O seguinte Lema indica como construir os pontos x_i que satisfazem (4.2) e (4.3).

Lema 4.2. *Seja $x_i = (x_{i,1}, \dots, x_{i,I-1})'$, $1 \leq i \leq I-1$ definido por $x_{i,j} = \delta_{i,j}$. Ou seja x_1, \dots, x_{I-1} é a base canônica de \mathbb{R}^{I-1} . Definimos $x_I = (-\mu_1/\mu_I, \dots, -\mu_{I-1}/\mu_I)'$. Temos*

$$(i) \quad \sum_{i=1}^I \mu_i x_i = 0$$

(ii) *Seja $\Sigma = (\sigma_{ii'})$ a matriz $I \times I$ definida como*

$$(4.8) \quad \sigma_{ii'} = \frac{1}{2} (-D_{ii'} + D_{i\cdot} + D_{i'\cdot} - D_{\cdot\cdot})$$

e Σ^* a matriz $(I-1) \times (I-1)$ formada pelas primeiras $(I-1)$ linhas e $(I-1)$ colunas da matriz Σ . Então (4.2) é válida.

Demonstração.

(i) Decorre imediatamente da definição.

(ii) Pela Lema 4.1 é suficiente varificar (4.4). A verificação é imediata para $i, i' \leq I-1$. Consideremos agora o caso $i \leq I$, $i' = I$. Por (4.3) temos

$$\begin{aligned} \langle x_i, x_I \rangle_{\Sigma^*} &= \left\langle x_{i'} - \sum_{i'=1}^{I-1} (\mu_{i'}/\mu_i) x_{i'} \right\rangle_{\Sigma^*} = \\ &= - \sum_{i'=1}^{I-1} (\mu_{i'}/\mu_I) \langle x_i, x_{i'} \rangle_{\Sigma^*} = \\ &= - \sum_{i'=1}^{I-1} (\mu_{i'}/\mu_I) \sigma_{ii'}. \end{aligned}$$

Como $\sum_{i'=1}^I \mu_{i'} \sigma_{ii'} = 0$ temos que

$$\langle x_i, x_I \rangle = \sigma_{iI} = \frac{1}{2} (-D_{iI} + D_{i\cdot} + D_{I\cdot} - D_{\cdot\cdot})$$

e portanto (4.4) é válida para $i' = I$. Da mesma forma (4.4) pode ser verificado pra $i = i' = I$ o que termina a prova do Lema.

O Lema 4.2 implica que para minimizar (4.1) é suficiente encontrar $F_I^{(p)}$ em \mathbb{R}^p tais que minimizem

$$(4.9) \quad \sum_{i=1}^I \sum_{i'=1}^I \left| \|x_i - x_{i'}\|_{\Sigma^*}^2 - \|F^{(p)} - F_{i'}^{(p)}\|_I^2 \right|.$$

Se a matriz Σ^* fosse definida positiva uma solução aproximada poderia obter-se com componentes principais utilizando a Proposição 2.2 do Capítulo 2. Vamos aplicar esta técnica formalmente com a única condição que a matriz ΣD , onde $D = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_I \end{pmatrix}$, tenha pelo menos p autovalores positivos. Vamos provar que o resultado a ser obtido é bom sempre que os $(I - p)$ autovalores menores da matriz ΣD sejam pequenos em valor absoluto comparados com os p maiores. O seguinte Lema calcula os vetores $D^{(p)}$ da Proposição 2.2 no caso em que Σ^* seja definida positiva.

Lema 4.3. *Sejam x_i , $1 \leq i \leq I$, os vetores do Lema 4.2. A matriz $\Sigma^* = (\sigma_{ii'})$ de $(I - 1) \times (I - 1)$ com $\sigma_{ii'}$ dada por (4.8). Suponhamos que Σ^* é definida positiva. A Proposição 2.2 produz os seguintes resultados*

i) *A matriz C está dada por*

$$(4.10) \quad C = D^* + \phi^*$$

onde $D^* = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_{I-1} \end{pmatrix}$ e ϕ^* é a matriz $(I - 1) \times (I - 1)$ com elemento (i, i') dado por $\varphi_{ii'} = (\mu_i, \mu_{i'}) / \mu_I$.

ii) *Seja $U^* = (U_1^*, \dots, U_{I-1}^*)$ a matriz U da Proposição 2.2. Temos*

$$(4.11) \quad U^{*'} C U = \Lambda$$

$$(4.12) \quad U^{*'} \Sigma^{*-1} U^* = I$$

onde $\Lambda = \begin{pmatrix} \lambda_1^* & & 0 \\ & \ddots & \\ 0 & & \lambda_{I-1}^* \end{pmatrix}$ $\lambda_1^* \geq \lambda_2^* \geq \dots \geq \lambda_{I-1}^*$ e portanto U_α^* é autovetor de Σ^*C correspondente ao autovalor λ_α^* . Temos então

$$(4.13) \quad F_\alpha(i) = U_\alpha^{*'} x_i = U_\alpha^*(i) \quad 1 \leq i \leq I-1$$

$$(4.14) \quad F_\alpha(I) = - \sum_{i=1}^{I-1} (\mu_i / \mu_I) \mu_\alpha^*(i)$$

Demonstração.

- (i) É imediato dado a definição dos x_i .
- (ii) Também é imediato dado a definição dos x_i .

De acordo com o visto no Teorema 2.1, se Σ^* fosse definida positiva uma solução aproximada ao problema de minimizar (4.9) está dada por

$$(4.15) \quad F_i^{(p)} = (F_1(i), \dots, F_p(i))'.$$

O seguinte Teorema indica como pode generalizar-se este resultado ao caso no qual Σ^* não seja definida positiva. Também indica a relação que existe entre os autovalores e autovetores de Σ^*C e os de ΣD , $D = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_I \end{pmatrix}$. Este Teorema é válido sem supor que Σ^* seja definida positiva.

Teorema 4.1.

- (i) A matriz ΣD admite o autovetor $U_I = (1, \dots, 1)'$ correspondente ao autovalor $\lambda_I = 0$.
- (ii) Sejam $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{I-1}$ os outros autovalores de ΣD e os autovetores correspondentes $U_\alpha = (U_\alpha(1), \dots, U_\alpha(I))'$ $1 \leq \alpha \leq I-1$. Os U_α podem ser escolhidos de forma tal que sejam satisfeitas

$$(4.16) \quad \sum_{i=1}^I \mu_i U_\alpha(i) = 0, \quad \alpha = 1, \dots, I-1$$

$$(4.17) \quad \sum_{i=1}^I \mu_i U_{\alpha}(i) U_{\alpha'}(i) = \delta_{\alpha\alpha'}, \quad \alpha, \alpha' = 1, \dots, I-1$$

(iii) $\lambda_{\alpha} = \lambda_{\alpha}^*$ onde os λ_{α}^* são os autovalores de $\Sigma^* C$ do Lema 4.3.

Por outro lado se os U_{α} satisfazem (4.16) e (4.17) e $\lambda_{\alpha} > 0$ teremos

$$U_{\alpha} = \left(U_{\alpha}^* / \sqrt{\lambda_{\alpha}}, - \sum_{i=1}^{I-1} \mu_i / (\mu_I \sqrt{\lambda_{\alpha}}) U_{\alpha}^*(i) \right)$$

onde U_{α}^* é o autovetor de $\Sigma^* C$ correspondente a λ_{α} que satisfaz $U_{\alpha}^* \Sigma^{-1} U_{\alpha} = 1$.

(iv) Sejam $F_{\alpha}(i)$ definidos por (4.13) e (4.14). Se $\lambda_{\alpha} > 0$ temos:

$$F_{\alpha}(i) = \sqrt{\lambda_{\alpha}}(i) U_{\alpha}(i) \quad 1 \leq i \leq I \quad 1 \leq \alpha \leq I-1.$$

Por (4.15) temos

$$(4.18) \quad F_i^{(p)} = \sqrt{\lambda_{\alpha}}(U_1(i), \dots, U_p(i))'$$

(v)

$$(4.19) \quad D_{ii'} = \sum_{\alpha=1}^{I-1} (U_{\alpha}(i) - U_{\alpha}(i'))^2 \lambda_{\alpha}$$

$$(4.20) \quad \sum_{i=1}^I \sum_{i'=1}^I D_{ii'} = 2 \sum_{\alpha=1}^{I-1} \lambda_{\alpha}$$

(vi) Suponhamos que $\lambda_{\alpha} > 0 \quad 1 \leq \alpha \leq p$ e definimos $F_i^{(p)}$ por (4.18). Temos

$$\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \left| D_{ii'} - \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2 \right| \leq 2 \sum_{\alpha=p+1}^{I-1} |\lambda_{\alpha}|$$

com igualdade se todos os λ_i são positivos.

Demonstração. (i) Seja $U_1 = (1, \dots, 1)'$. Por (4.8) temos

$$\begin{aligned} \sum_{i'=1}^I (\Sigma D)_{ii'} U_I(i') &= \frac{1}{2} \sum_{i'=1}^I \mu_{i'} (-d_{ii} + d_{i.} + d_{i'.} - d_{..}) = \\ &= \frac{1}{2} (-d_{i.} + d_{i.} + d_{..} - d_{..}) = 0. \end{aligned}$$

(ii) Pelo Teorema 2 os autovalores de $D\Sigma$ são também $\lambda_1, \dots, \lambda_{I-1}, \lambda_I$. Também pelo Teorema 2 existem autovetores de $D\Sigma$ V_1, \dots, V_I tais que

$$(4.21) \quad V'_\alpha D^{-1} V_\alpha = \delta_{\alpha\alpha'}.$$

Temos também que $U_\alpha = D^{-1} V_\alpha$, $1 \leq \alpha \leq I$ são os autovetores de ΣD . Temos de (4.21) que

$$U'_\alpha D U_\alpha = V'_\alpha D^{-1} D D V_\alpha = \delta_{\alpha\alpha'}$$

e portanto (4.17) é satisfeita. Tomando $\alpha' = I$ temos (4.16).

(iii) Se $\alpha < I$ por (4.16) temos

$$(4.22) \quad U_\alpha(I) = - \sum_{i=1}^{I-1} (\mu_i / \mu_I) U_\alpha(i).$$

Portanto para provar (iii) é suficiente provar que

$$\widehat{U}_\alpha = (U_\alpha(1), \dots, U_\alpha(I-1))'$$

é autovetor de $\Sigma^* C$ correspondente a λ_α e (4.23)

$$(4.23) \quad \widehat{U}_{\alpha'} \Sigma^{*-1} \widehat{U}_\alpha = \frac{1}{\lambda_\alpha}.$$

Como U_α é um autovetor de ΣD correspondente a λ_α temos

$$(4.24) \quad \sum_{i'=1}^{I-1} \sigma_{ii'} \mu_{i'} U_\alpha(i) + \sigma_{iI} \mu_I U_\alpha(I) = \lambda_\alpha U_\alpha(i).$$

De (4.8) resulta

$$(4.25) \quad \sum_{i'=1}^I \mu_{i'} \sigma_{ii'} = 0.$$

Extraindo σ_{iI} de (4.25) e substituindo em (4.24) temos

$$\begin{aligned} \sigma_{iI} \mu_i U_\alpha(i) = & - \sum_{i''=1}^{I-1} \sum_{i'=1}^{I-1} \sigma_{ii'} \mu_{i'} \mu_{i''} U_\alpha(i'') / \mu_i + \\ & + \sum_{i''=1}^{I-1} \sum_{i'=1}^{I-1} \sigma_{ii'} \varphi_{ii'} U_\alpha(i''). \end{aligned}$$

Como por (4.10) $C = D^* + \phi^*$ decorre que \hat{U}_α é autovetor de $\Sigma^* C$ correspondente a λ_α .

Proporemos agora (4.23). Como \hat{U}_α é autovetor de $\Sigma^* C$ temos

$$(4.26) \quad \hat{U}_\alpha \Sigma^{*-1} \hat{U}_\alpha = \frac{\hat{U}_\alpha C \hat{U}_\alpha}{\lambda_\alpha} = \frac{\hat{U}'_\alpha (D^* + \phi^*) \hat{U}_\alpha}{\lambda_\alpha}.$$

Mas

$$\begin{aligned} \hat{U}'_\alpha \phi^* \hat{U}_\alpha = & - \sum_{i=1}^{I-1} \sum_{i'=1}^{I-1} (\mu_i \mu'_i / \mu_I) U_\alpha(i) U_\alpha(i') = \\ = & \mu_I \left(\sum_{i=1}^{I-1} (\mu_i / \mu_I) U_\alpha(i) \right). \end{aligned}$$

Utilizando (4.16) temos

$$\hat{U}'_\alpha \phi^* U_\alpha = \mu_I U_\alpha^2(I)$$

e substituindo em (4.26) obtemos

$$\hat{U}'_\alpha \Sigma^{*-1} U_\alpha = \left(\sum_{i=1}^I \mu_i U_\alpha^2(i) \right) / \lambda_\alpha.$$

Portanto por (4.17) obtemos (4.23).

(iv) É imediato a partir de (iii).

(v) Seja $V = \begin{pmatrix} v_1 & \dots & v_I \end{pmatrix}$ uma matriz $I \times I$ onde V_α $1 \leq \alpha \leq I$ é definido como em (ii). Temos

$$(4.27) \quad V' \Sigma V = \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_I \end{pmatrix}$$

$$(4.28) \quad V' D^{-1} V = I.$$

De (4.27) decorre

$$(4.29) \quad \Sigma = V'^{-1} \Lambda V^{-1}$$

e de (4.28)

$$V^{-1} = V' D^{-1}.$$

Substituindo em (4.29) temos

$$\Sigma = D^{-1} V \Lambda V' D$$

e pelo que vimos em (ii) e como $D^{-1} V = U = (U_1, \dots, U_I)$ temos

$$(4.30) \quad \Sigma = U \Lambda U'.$$

Portanto

$$\sigma_{ii'} = \sum_{\alpha=1}^I \lambda_{\alpha} U_{\alpha}(i) U_{\alpha}(i') = \sum_{\alpha=1}^{I-1} \lambda_{\alpha} U_{\alpha}(i) U_{\alpha}(i').$$

Além do mais por (4.2)

$$\begin{aligned} D_{ii'} &= \|x_i - x_{i'}\|_{\Sigma^*}^2 = \langle x_i, x_i \rangle_{\Sigma^*} + \langle x_{i'}, x_{i'} \rangle_{\Sigma^*} - 2 \langle x_i, x_{i'} \rangle_{\Sigma^*} = \\ &= \sigma_{ii} + \sigma_{i'i} - 2 \sigma_{ii'} = \sum_{\alpha=1}^{I-1} \lambda_{\alpha} U_{\alpha}^2(i) + \sum_{\alpha=1}^{I-1} \lambda_{\alpha} U_{\alpha}^2(i') - \\ &\quad - 2 \sum_{\alpha=1}^{I-1} \lambda_{\alpha} U_{\alpha}(i) U_{\alpha}(i') = \sum_{\alpha=1}^{I-1} \lambda_{\alpha} (U_{\alpha}(i) - U_{\alpha}(i'))^2 \end{aligned}$$

e portanto (4.19) fica provada.

Antes de provar (4.20) provaremos o seguinte: para todo conjunto de números $\gamma_{p+1}, \dots, \gamma_{I-1}$ temos

$$(4.31) \quad \sum_{i'=1}^I \sum_{i=1}^I \sum_{\alpha=p+1}^{I-1} \mu_i \mu_{i'} \gamma_{\alpha} (U_{\alpha}(i) - U_{\alpha}(i'))^2 = 2 \sum_{\alpha=p+1}^{I-1} \gamma_{\alpha}.$$

Para provar (4.30) observamos que

$$\begin{aligned}
 & \sum_{i'=1}^I \sum_{i=1}^I \sum_{\alpha=p+1}^{I-1} \mu_i \mu_{i'} \gamma_\alpha (U_\alpha(i) - U_\alpha(i'))^2 = \\
 &= \sum_{\alpha=p+1}^{I-1} \gamma_\alpha \sum_{i'=1}^I \sum_{i=1}^I \mu_i \mu_{i'} (U_\alpha(i) - U_\alpha(i'))^2 = \\
 &= 2 \sum_{\alpha=p+1}^{I-1} \gamma_\alpha \left[\sum_{i=1}^I \mu_i U_\alpha^2(i) - \sum_{i'=1}^I \sum_{i=1}^I \mu_i \mu_{i'} U_\alpha(i) U_\alpha(i') \right] = \\
 &= 2 \sum_{\alpha=p+1}^{I-1} \gamma_\alpha \left[\sum_{i=1}^I \mu_i U_\alpha^2(i) - \left(\sum_{i=1}^I \mu_i U_\alpha(i) \right)^2 \right]
 \end{aligned}$$

e usando (4.16) e (4.17) obtemos (4.31). Agora de (4.19) e aplicando (4.31) com $p = 0$ e $\gamma_\alpha = \lambda_\alpha$ decorre (4.20).

(vi) De (4.18) e (4.19) decorre que

$$\left| D_{ii'} - \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2 \right| \leq \sum_{\alpha=p+1}^{I-1} |\lambda_\alpha| (U_\alpha(i) - U_\alpha(i'))^2.$$

Aplicando agora (4.30) com $\gamma_\alpha = |\lambda_\alpha|$ decorre (vi).

De acordo com (v) e (vi) do Teorema 4.1 um limite superior do erro relativo $E R_p$ na representação dos objetos x_i pelos pontos $F_i^{(p)}$ em \mathbb{R}^p está dado por

$$(4.32) \quad E R_p = \sum_{i'=1}^I \sum_{i=1}^I \mu_i \mu_{i'} \left| D_{ii'} - \|F_i^{(p)} - F_{i'}^{(p)}\|_I^2 \right| \leq$$

$$\leq \frac{\sum_{i=p+1}^{I-1} |\lambda_i|}{\sum_{i=1}^{I-1} |\lambda_i|} = E R_p^*.$$

Como aconteceu no Capítulo 2 ao estudar Componentes Principais $E R_p$ não leva em consideração que a representação $F_i^{(p)}$ $1 \leq i \leq I$ é

equivalente a $\lambda F_i^{(p)}$ com λ um escalar arbitrário. Portanto um coeficiente mais preciso será o dado por ERC_p definido como

$$ERC_p = \frac{\text{mínimo}_{\lambda} \sum_{i'=1}^I \sum_{i=1}^I \mu_i \mu_{i'} |D_{ii'} - \lambda \|F_i^{(p)} - F_{i'}^{(p)}\|_I|^2|}{\sum_{i'=1}^I \sum_{i=1}^I \mu_i \mu_{i'} D_{ii'}}$$

Pode-se provar que no caso no qual os objetos x_i sejam elementos de um espaço \mathbb{R}^J e se $d_{ii'} = (x_i' \widehat{\Sigma} x_i)^{1/2}$ onde $\widehat{\Sigma}$ é definida positiva $J \times J$, a solução determinada neste capítulo será a mesma que resultaria da aplicação de Componentes Principais aos x_i da mesma forma como foi explicado no Capítulo 2.

Em alguns casos pode não conhecer-se a matriz de dissimilaridades $d_{ii'}$, mas somente a ordem das mesmas quer dizer: é conhecido o par de objetos que encontra-se a distância mínima, o par que vem a continuação em ordem de distâncias, etc., até chegar ao par de objetos que encontrou-se a maior distância. Suponhamos ter I objetos x_i $1 \leq i \leq I$ e que $Q_{ii'}$ indica a ordem da distância entre x_i e $x_{i'}$. Ou seja $Q_{ii'} = 1$ indicará que os objetos x_i e $x_{i'}$ são os mais próximos e $Q_{ii'} = I(I-1)/2$ (existem $I(I-1)/2$ pares de objetos) indicará que x_i e $x_{i'}$ é o par de objetos mais distantes entre si.

Benzecri (1973) propõe o seguinte procedimento heurístico para definir a matriz $D_{ii'}$ a partir de $Q_{ii'}$.

Suponhamos que os x_i possam ser representados em \mathbb{R}^p e que tenham uma distribuição normal multivariada com matriz de covariância I . Então $D_{ii'}/2 = \|x_i - x_{i'}\|_I^2/2$ teria uma distribuição χ^2 com p graus de liberdade. Portanto $d_{ii'}/2$ seria uma amostra (não independente) de tamanho $N = I(I-1)/2$ desta distribuição. Se $Q_{ii'} = k$, $D_{ii'}/2$ seria o percentil empírico k/N desta distribuição e portanto poderia ser aproximado pelo correspondente percentil teórico da distribuição χ^2 com p graus de liberdade. Quer dizer que poderia se reconstruir a matriz $D_{ii'}$ por

$$\frac{D_{ii'}}{2} = \chi_p^2{}^{-1}(k/N) \quad \text{se} \quad Q_{ii'} = k$$

onde χ_p^2 é a função de distribuição de uma χ^2 com p graus de liberdade. Uma vez construídos os $D_{ii'}$ utilizaríamos o procedimento descrito anteriormente.

Como neste caso só interessa conservar a ordem das distâncias e não os valores que são desconhecidos, é conveniente definir uma nova media do erro relativo. Esta medida chamada “Stress” é devida a Kruskal (1964). O Stress correspondente à solução em \mathbb{R}^p estava dada por

$$S_p = \underset{f \in \mathcal{M}}{\text{mínimo}} \frac{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} (\|F_i^{(p)} - F_{i'}^{(p)}\|_I - f(Q_{ii'}))^2}{\sum_{i=1}^I \sum_{i'=1}^I \mu_i \mu_{i'} \|F_i^p - F_{i'}^p\|_I^2}$$

onde \mathcal{M} é o conjunto de funções monotônicas dos naturais nos reais. Para o cálculo da f que realiza o mínimo pode ver-se Kruskal (1964). (Isto é usualmente mencionado como Escalamento Multidimensional não Métrico).

4.2 Notas e comentários sobre MDS

Uma matriz D com elementos d_{ij} satisfazendo as propriedades (i),(ii) e (iii) do parágrafo 2 é chamada de matriz de dissimilaridades.

Se adicionalmente para todos as triplas i, j, k temos $d_{ij} + d_{jk} \geq d_{ik}$ (o que chamaremos de propriedade “métrica” ou desigualdade métrica) a matriz d seria chamada de matriz de distâncias ou simplesmente uma distância entre os elementos do conjunto I .

Exemplos mais frequentes de métricas que são utilizadas sobre subconjuntos I de n pontos em R_p são as métricas euclidianas (Pitágoras), dos quarterões (city- block), Chebychev (ou do máximo) definidas por:

$$\begin{aligned} \text{Euclideana :} \quad d_e &= \left\{ \sum_{i=1}^p (x_i - y_i)^2 \right\}^{1/2} \\ \text{City - Block :} \quad d_{cb} &= \sum_{i=1}^p |x_i - y_i| \\ \text{Chebychev :} \quad d_c &= \max \{ |x_i - y_i|, i = 1, 2, \dots, p \} \end{aligned}$$

As duas primeiras são casos particulares de:

$$\text{Minkowsky :} \quad d_M = \left\{ \sum_{i=1}^p (x_i - y_i)^m \right\}^{1/m} \quad (\text{de ordem } m)$$

Quando os valores que os vetores x_i e y_i tomam são somente 0 ou 1, algumas alternativas específicas e interpretáveis são usadas.

$$\text{Jaccard :} \quad d_J = - \sum_{i=1}^p x_i y_i / \left(\sum_{i=1}^p x_i + \sum_{i=1}^p y_i - \sum_{i=1}^p x_i y_i \right).$$

Coincidência(matching): $\#(i: x_i = y_i)/p$, onde $\#$ indica o cardinal do conjunto.

Uma medida de similaridade que usaremos em exemplo posterior é definida da seguinte forma. Sejam $\{P_k, k = 1, 2, \dots, n\}$ um conjunto de partições finitas de um conjunto finito e fixo I . Definimos para cada par de elementos i e j do conjunto I

$$s_{ij} = \#\{k: i \text{ e } j \text{ pertencem ao mesmo subconjunto da partição } k\}$$

Definimos a dissimilaridade como $d_{ij} = n - s_{ij}$.

Medidas de dissimilaridade podem ser transformadas em medidas de *similaridade* usando diversas transformações como por exemplo $s_{ij} = 1/(1 + d_{ij})$ ou $s_{ij} = c - d_{ij}$ para alguma constante c maior que todos os d_{ij} .

Nem toda dissimilaridade é uma distância ainda que muitas o são. (Ver Gower J.C. 1982).

Não existe uma “melhor” medida de similaridade. Algumas podem ser mais apropriadas que outras para certas situações ou conjunto de dados ou aplicação.

Uma noção importante é a de imersão em um espaço euclidiano.

Uma matriz $D(n \times n)$ de elementos d_{ij} é *Euclidiana* se existem n pontos p_1, p_2, \dots, p_n em algum R_p de forma tal que a distância euclidiana entre p_i e p_j é exatamente d_{ij} .

Neste caso claramente D é uma matriz de distâncias (e naturalmente de dissimilaridades).

Nem todo espaço métrico finito com matriz de distâncias D é imersível em algum espaço Euclidean para algum valor de p .

Seja D uma matriz de dissimilaridades, I a matriz identidade em R^n e $\mathbf{1}$ um vetor de comprimento n com elementos todos iguais a 1 e B a matriz

$$B = (I - \mathbf{1}\mathbf{s}')D(I - \mathbf{s}\mathbf{1}') \text{ onde } s \text{ indica a matriz transposta de } s \text{ e } \mathbf{s}'\mathbf{1} = 1.$$

Teorema: (Condição necessária e suficiente para inmersibilidade). D é imersível em um Espaço Euclidiano se e somente se B é semi-definida positiva para algum s tal que $\mathbf{s}' = \mathbf{1}$ e $\mathbf{s}'D \neq 0$.

Outro resultado interessante para verificar se algumas expressões definem métricas está dado pelo seguinte resultado (ver Deza, Elena (2006)). Seja $f: R \rightarrow R$ uma função duas vezes diferenciável definida para $x \geq 0$ e tal que $f(0) = 0$, $f' > 0$ para todo $x \geq 0$ e $f'' \leq 0$ para todo $x \geq 0$. Neste caso f é côncava e em particular $f(x + y) \leq f(x) + f(y)$.

Dado u espaço métrico (X, d) a função d_f definida por $d_f(x, y) = f(d(x, y))$ é uma métrica sobre X . As métricas são de fato equivalentes (geram a mesma topologia).

Exemplos: αd para $\alpha > 0$, d^α ($0 < \alpha < 1$), $\ln(1 + d)$, $\arcsin d$, $\operatorname{arccosh} d$, e $d/(1 + d)$.

4.3 Exemplos do Capítulo 4

Os exemplos deste capítulo foram rodados usando a função `cmdscale()` do R. Como já foi mencionado, os resultados usando o escalamento não métrico são muito semelhantes. Para o caso não métrico o pacote `smacof` do R é recomendado.

Exemplo 1

Este exemplo contém data de questionários aplicado em estudantes de Ciências Políticas sobre similaridades percebidas entre 12 países: BEL, BRA, CHI, CUB, EGY, FRA, IND, ISR, USA, USS, YUG e ZAI.

Cada entrevistado indicava em uma escala a dissimilaridade entre cada par de países. Os resultados foram depois agregados sobre todos os participantes e uma dissimilaridade média foi calculada. Estes valores então indicados na Figura 1.

Figura 1

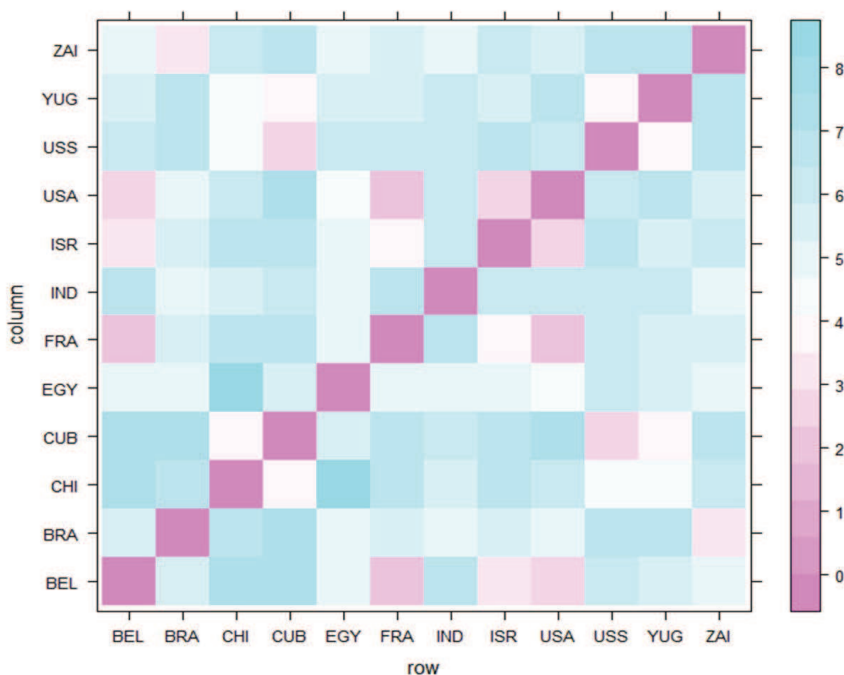
	BEL	BRA	CHI	CUB	EGY	FRA	IND	ISR	USA	USS	YUG	ZAI
BEL		0	5,58	7	7,08	4,83	2,17	6,42	3,42	2,5	6,08	5,25
BRA	5,58		0	6,5	7	5,08	5,75	5	5,5	4,92	6,67	6,83
CHI	7	6,5		0	3,83	8,17	6,67	5,58	6,42	6,25	4,25	4,5
CUB	7,08	7	3,83		0	5,83	6,92	6	6,42	7,33	2,67	3,75
EGY	4,83	5,08	8,17	5,83		0	4,92	4,67	5	4,5	6	5,75
FRA	2,17	5,75	6,67	6,92	4,92		0	6,42	3,92	2,25	6,17	5,42
IND	6,42	5	5,58	6	4,67	6,42		0	6,17	6,33	6,17	6,08
ISR	3,42	5,5	6,42	6,42	5	3,92	6,17		0	2,75	6,92	5,83
USA	2,5	4,92	6,25	7,33	4,5	2,25	6,33	2,75		0	6,17	6,67
USS	6,08	6,67	4,25	2,67	6	6,17	6,17	6,92	6,17		0	3,67
YUG	5,25	6,83	4,5	3,75	5,75	5,42	6,08	5,83	6,67	6,67	3,67	0
ZAI	4,75	3	6,08	6,67	5	5,58	4,83	6,17	5,67	6,5	6,92	0

Esta forma de coletar dissimilaridades é bastante comum em outras áreas como na análise de comportamento de consumidores. Imaginemos que temos um conjunto de marcas em uma categoria (digamos 15 marcas de detergente para roupas) e queremos ter uma ideia do grau de concorrência entre as marcas em termos de como são vistas como substitutas/alternativas de compra entre os consumidores. Seriam marcas que o consumidor poderia comprar uma em lugar da outra. Esta informação é em geral valiosa para os fabricantes porque indica a concorrência mais direta entre marcas.

No Exemplo 2 veremos uma situação deste tipo porém com uma coleta diferente para determinar as dissimilaridades.

A Figura 2 usa os mesmos valores da Figura 1 só que em uma versão gráfica através de um gráfico de níveis (levelplot):

Figura 2



A Figura 3 indica os valores dos autovalores na decomposição espectral.

9 Autovalores positivos e três negativos que indica que a matriz de Dissimilaridades não é Euclidiana.

Figura 3

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
72,50	41,71	25,83	15,60	11,17	9,31	6,03	2,96	0,00	-0,45	-2,78	-7,43

Adicionalmente os primeiros autovalores positivos são muito maiores que os negativos; (ver Mardia e outros (1979)).

Os valores absolutos acumulados dos autovalores estão indicados na Figura 4:

Figura 4

Proporção nas somas em valores absolutos acumuladas

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
0,37	0,58	0,72	0,80	0,85	0,90	0,93	0,95	0,95	0,95	0,96	1,00

Os dois primeiros autovalores indicam uma adequação de 58,3%.

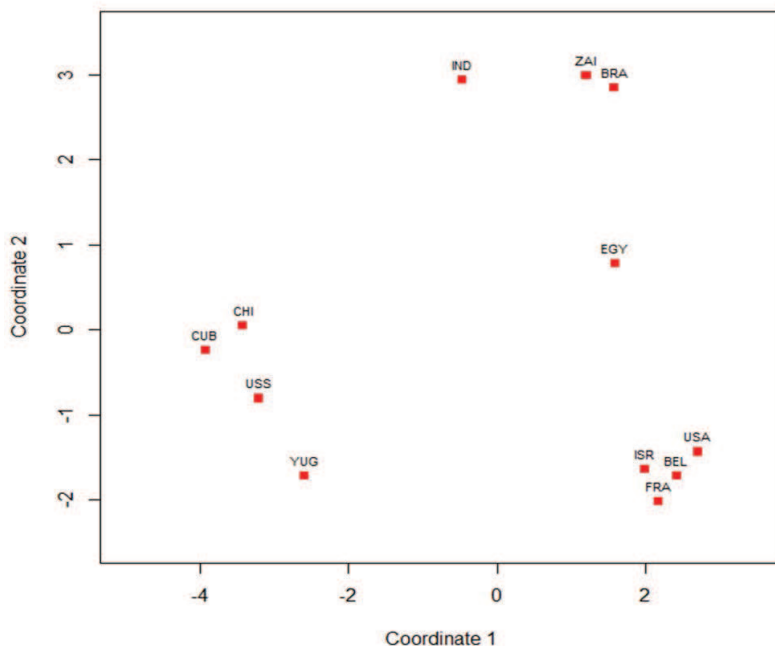
Como as duas primeiras dimensões representam próximo de 60% da variação ficaremos com estas duas dimensões para a representação gráfica.

A Figura 5 contém as coordenadas dos pontos e a Figura 6 a sua representação.

Figura 5

Pais	Dimension 1	Dimension 2
BEL	-5,915148	-2,643098
BRA	-3,712266	5,889597
CHI	7,573524	2,081641
CUB	8,292496	-1,602484
EGY	-2,286907	1,688705
FRA	-5,610125	-3,979386
IND	2,510442	5,640259
ISR	-3,165117	-4,59795
USA	-7,061969	-1,396854
USS	6,592312	-2,807475
YUG	4,495912	-4,729947
ZAI	-1,713152	6,456992

Figura 6



Note-se que os países aparecem agrupados em 3 grandes grupos em termos de suas similaridades com Egito sendo algo diferenciado talvez mais próximo de Brasil e Índia.

Exemplo 2. Como já mencionamos anteriormente imaginemos que temos um conjunto de n consumidores de uma categoria de produtos. Solicitamos a cada entrevistado que forme conjuntos de produtos de forma tal que dois produtos sejam colocados no mesmo grupo se são considerados como eventuais substitutos um do outro.

Sejam $\{P_k, k = 1, 2, \dots, n\}$ a partição dos produtos feita por um consumidor determinado. Esta partição será diferente para cada indivíduo tanto nos subconjuntos que a definem como no seu número. Definimos para cada par de elementos i e j do conjunto de produtos I $s_{ij} = \#$ vezes que i e j pertencem ao mesmo subconjunto/ n ($\#$ indica número). Definimos a dissimilaridade entre os produtos i e j como $d_{ij} = 1 - s_{ij}$. A matriz a seguir indica a matriz de dissimilaridades obtida entre um conjunto de 370 consumidores entre 19 marcas de Margarina.

	DaniHo	DaniMa	Flor	Golf	Cordo	CodoMyB	Manterina	MargaHo	MargaMa	Milhojas	MKTHo	MKTHo	MKTHo	MKTHo	PremHo	PremTran	PremMa	Primicia	PrimHo	PrimMa	Super
DaniHo	0,00	0,40	0,68	0,68	0,58	0,76	0,68	0,51	0,73	0,54	0,42	0,69	0,44	0,50	0,63	0,43	0,46	0,67	0,58	0,67	0,58
DaniMa	0,40	0,00	0,55	0,44	0,82	0,64	0,39	0,79	0,56	0,78	0,67	0,45	0,68	0,71	0,39	0,76	0,67	0,47	0,64	0,67	0,64
Flor	0,68	0,55	0,00	0,39	0,58	0,57	0,55	0,78	0,62	0,52	0,74	0,62	0,61	0,66	0,51	0,71	0,81	0,62	0,51	0,62	0,51
Golf	0,68	0,44	0,39	0,00	0,84	0,55	0,49	0,84	0,68	0,64	0,67	0,47	0,66	0,63	0,42	0,77	0,76	0,51	0,46	0,51	0,46
CordoHo	0,58	0,82	0,58	0,84	0,00	0,26	0,78	0,48	0,60	0,37	0,60	0,81	0,62	0,49	0,77	0,42	0,39	0,64	0,76	0,64	0,76
CodoMyB	0,76	0,64	0,57	0,55	0,26	0,00	0,61	0,68	0,40	0,63	0,77	0,61	0,76	0,67	0,60	0,63	0,60	0,45	0,80	0,45	0,80
Manterina	0,68	0,39	0,55	0,49	0,78	0,61	0,00	0,78	0,55	0,64	0,67	0,42	0,73	0,79	0,49	0,84	0,76	0,54	0,62	0,54	0,62
MargaHo	0,51	0,79	0,78	0,84	0,48	0,68	0,78	0,00	0,47	0,48	0,51	0,82	0,41	0,36	0,70	0,39	0,37	0,70	0,72	0,70	0,72
MargaMa	0,73	0,56	0,62	0,68	0,60	0,40	0,55	0,47	0,00	0,78	0,76	0,60	0,71	0,79	0,45	0,67	0,67	0,32	0,74	0,32	0,74
Milhojas	0,54	0,78	0,52	0,64	0,37	0,63	0,64	0,48	0,78	0,00	0,40	0,64	0,40	0,34	0,73	0,48	0,49	0,72	0,44	0,72	0,44
MKTHo	0,42	0,67	0,74	0,67	0,60	0,77	0,67	0,51	0,76	0,40	0,00	0,42	0,37	0,46	0,66	0,49	0,56	0,86	0,53	0,86	0,53
MKTHo	0,69	0,45	0,62	0,47	0,81	0,61	0,42	0,82	0,60	0,64	0,42	0,00	0,69	0,79	0,42	0,81	0,83	0,63	0,61	0,83	0,61
PremHo	0,44	0,68	0,61	0,66	0,62	0,76	0,73	0,41	0,71	0,40	0,37	0,69	0,00	0,19	0,36	0,45	0,48	0,74	0,58	0,74	0,58
PremTran	0,50	0,71	0,66	0,63	0,49	0,67	0,79	0,36	0,79	0,34	0,46	0,79	0,19	0,00	0,54	0,39	0,40	0,73	0,58	0,73	0,58
PremMa	0,63	0,39	0,51	0,42	0,77	0,60	0,49	0,70	0,45	0,73	0,66	0,42	0,36	0,54	0,00	0,75	0,73	0,45	0,61	0,45	0,61
Primicia	0,43	0,76	0,71	0,77	0,42	0,63	0,84	0,39	0,67	0,48	0,49	0,81	0,45	0,39	0,75	0,00	0,34	0,65	0,66	0,34	0,65
PrimHo	0,46	0,67	0,81	0,76	0,39	0,60	0,76	0,37	0,67	0,49	0,56	0,83	0,48	0,40	0,73	0,34	0,00	0,42	0,67	0,40	0,42
PrimMa	0,67	0,47	0,62	0,51	0,64	0,45	0,54	0,70	0,32	0,72	0,86	0,63	0,74	0,73	0,45	0,65	0,42	0,00	0,66	0,42	0,00
Super	0,58	0,64	0,51	0,46	0,76	0,80	0,62	0,72	0,74	0,44	0,53	0,61	0,58	0,58	0,61	0,66	0,67	0,66	0,66	0,67	0,66

A Figura 2 indica os valores dos autovalores na decomposição espectral.

Autovalores

Figura 2

11 Autovalores positivos o que indica que a matriz de Dissimilaridades não é Euclidiana.

V1	V2	V3	V4	V5	V6	V7	V8	V9
1,32	0,78	0,46	0,35	0,31	0,21	0,19	0,14	0,07

V10	V11	V12	V13	V14	V15	V16	V17	V18	V19
0,04	0,01	0,00	-0,02	-0,02	-0,04	-0,08	-0,11	-0,12	-0,13

Adicionalmente os primeiros autovalores positivos são muito maiores que os negativos. (Ver Mardia e outros (1979)).

Figura 3

Proporção nas somas em valores absolutos acumuladas

V1	V2	V3	V4	V5	V6	V7	V8	V9
0,300932	0,48	0,585	0,665	0,73	0,782	0,825	0,857	0,87

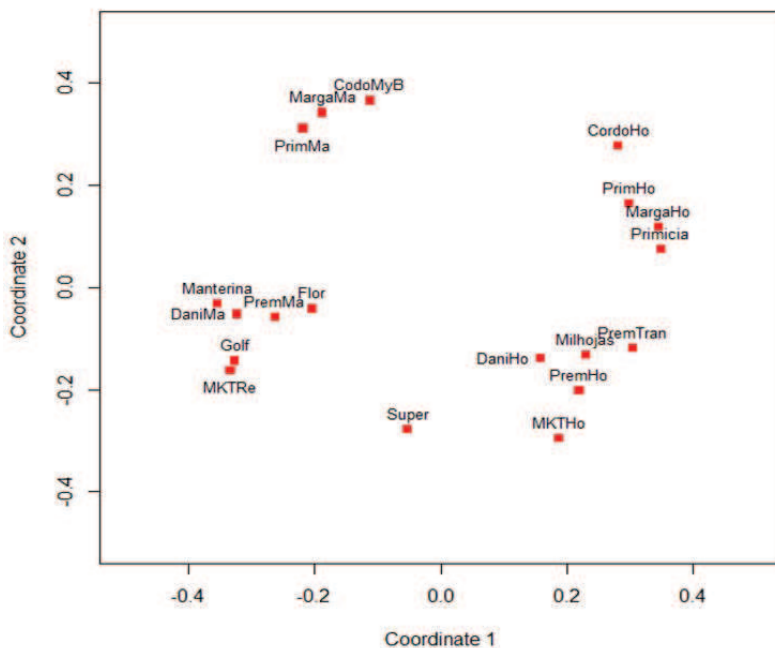
V10	V11	V12	V13	V14	V15	V16	V17	V18	V19
0,881917	0,88	0,884	0,887	0,89	0,902	0,921	0,945	0,97	1

Os dois primeiros autovalores indicam uma adequação de quase 50%
Como as duas primeiras dimensões representam próximo de 50% da variação ficaremos com estas duas dimensões para a representação gráfica.

A Figura 4 contém as coordenadas dos pontos e a Figura 5 a sua representação.

Figura 4

Marcas	Dimensao 1	Dimensao 2
DaniHo	0,158	-0,139
DaniMa	-0,323	-0,053
Flor	-0,204	-0,041
Golf	-0,326	-0,144
CordoHo	0,281	0,277
CodoMyB	-0,112	0,366
Manterina	-0,353	-0,031
MargaHo	0,345	0,119
MargaMa	-0,187	0,342
Milhojas	0,230	-0,132
MKTHo	0,188	-0,296
MKTRe	-0,333	-0,162
PremHo	0,219	-0,201
PremTran	0,304	-0,119
PremMa	-0,262	-0,058
Primicia	0,349	0,075
PrimHo	0,299	0,164
PrimMa	-0,218	0,311
Super	-0,052	-0,278



Note-se que as marcas aparecem agrupadas em 4 grandes grupos em termos de suas similaridades com a marca Super sendo algo diferenciada.

Exemplo 3.

Os dados deste exemplo encontra-se em Gower J. (1996) e no pacote BiplotGUI em R.

Para fazer mobília do tipo Old-Cape em South Africa três variedades de madeira foram utilizadas em diferentes momentos. As espécies foram, *Ocotea kenyensis* (Oken) 7 amostras, *Ocotea bullata* (Obul) 20 amostras, e *Ocotea perosa* (Opor) 10 amostras.

Sobre estas 37 amostras foram medidas 7 variáveis que poderiam caracterizar cada uma das espécies:

Species

Indicador da espécie. Um fator com níveis Obul, Oken e Opor.

VesD

Tangential vessel diameter in micrometres. A numeric vector.

VesL

Vessel element length in micrometres. A numeric vector.

FibL

Fibre length in micrometres. A numeric vector.

RayH

Ray height in micrometres. A numeric vector.

RayW

Ray width in micrometres. A numeric vector.

NumVes

The number of vessels per millimetre squared. A numeric vector.

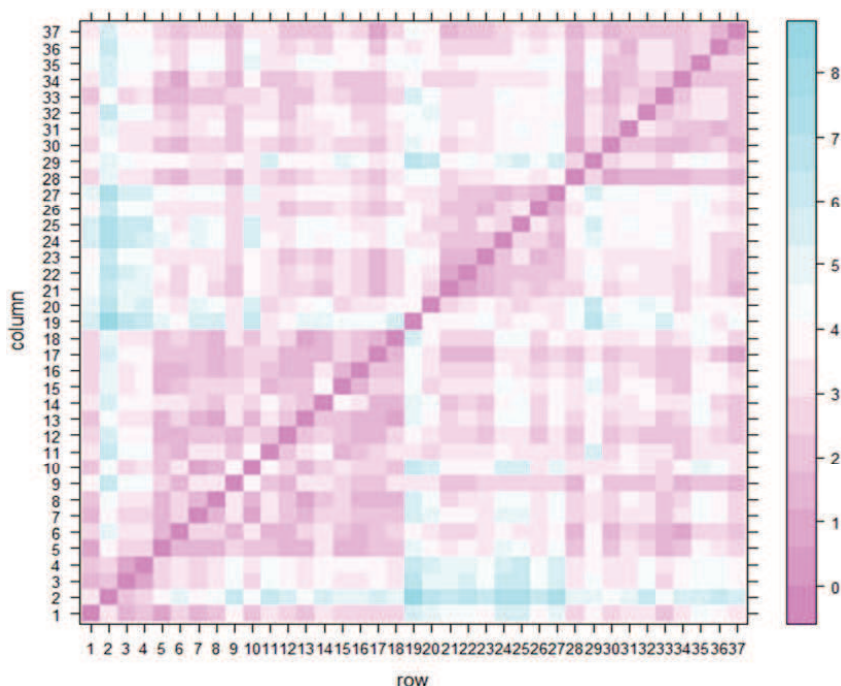
A seguir: **Figura 1. Dados de três espécies de madeira.**

	Species	VesD	VesL	FibL	RayH	RayW	NumVes
1	Obul	79	383	941	333	30	17
2	Obul	78	346	961	223	24	31
3	Obul	82	361	1039	316	27	25
4	Obul	79	324	1048	369	29	26
5	Obul	85	418	1051	347	34	14
6	Obul	111	448	1096	379	40	13
7	Obul	76	320	1130	347	29	13
8	Obul	103	371	1165	326	26	10
9	Obul	129	406	1165	428	44	11
10	Obul	74	281	1175	324	26	11
11	Obul	102	567	1221	395	40	11
12	Obul	95	415	1225	416	38	10
13	Obul	91	372	1234	375	26	11
14	Obul	113	314	1253	466	23	10
15	Obul	93	541	1267	347	34	14
16	Obul	94	437	1271	336	36	10
17	Obul	119	359	1280	412	32	11
18	Obul	104	387	1290	381	22	12
19	Obul	114	569	1369	568	52	11
20	Obul	141	621	1527	419	34	15
21	Oken	147	402	1391	440	32	9
22	Oken	142	393	1468	443	35	6
23	Oken	125	322	1530	459	34	11
24	Oken	156	401	1588	512	42	11
25	Oken	162	502	1591	369	42	8
26	Oken	103	378	1655	441	34	11
27	Oken	126	414	1759	459	42	8
28	Opor	122	346	981	393	40	14
29	Opor	139	133	993	342	33	14
30	Opor	130	368	1005	356	39	16
31	Opor	127	331	1027	473	38	20
32	Opor	112	309	1044	358	47	8
33	Opor	115	352	1048	300	36	14
34	Opor	130	471	1072	409	39	15
35	Opor	153	419	1077	392	48	20
36	Opor	135	370	1104	531	38	15
37	Opor	130	325	1166	428	36	12

Usando as 7 variáveis, calculamos a distância euclidiana usual entre cada par de linhas. Isto nos proporciona a matriz de distâncias (ou dissimilaridades) que será usada para MDS.

A Figura 2 mostra graficamente a matriz de distâncias entre cada par de elementos da amostra (37×37 elementos). O gráfico também sugere a presença de três grupos(ou segmentos) de amostras (que poderiam ou não coincidir com as espécies de *Ocotea* analisadas).

Figura 2. Plot de níveis (levelplot de matriz de distâncias (dissimilaridades)).



Note-se que todos são não negativos porque a matriz de dissimilaridades é neste caso uma matriz de distâncias euclidianas. São tem 6 positivos porque temos 7 variáveis.

Figura 3

V1	V2	V3	V4	V5	V6
102,3307	36,70507	33,60642	19,31997	15,69257	8,345266

As duas primeiras dimensões representam 64 por cento da inércia total, o que é adequado para nossa representação.

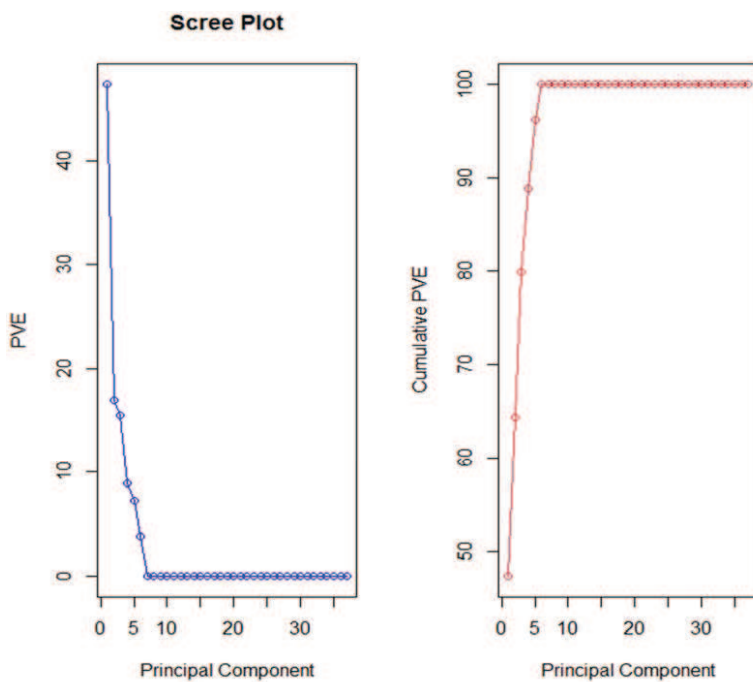
Figura 4. Proporção nas somas em valores absolutos acumulados.

V1	V2	V3	V4	V5	V6
0,473753	0,643684	0,799269	0,888714	0,961365	1

Os dois primeiros autovalores indicam sua adequação de 64%.

A Figura 5 contém o Scree Plot e gráfico com os valores acumulados.

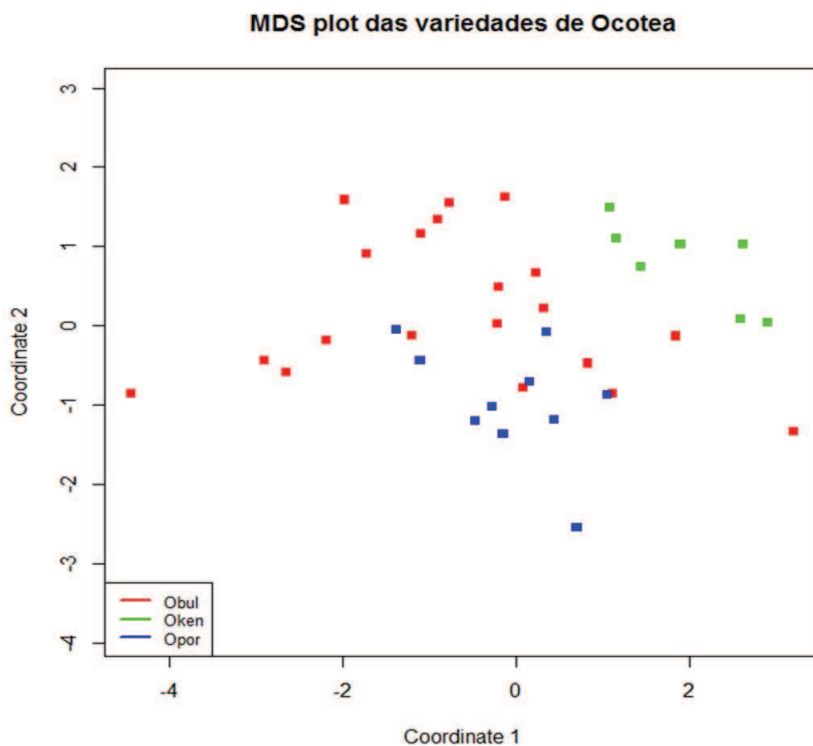
Figura 5



A Figura 6 contém o mapa com 37 pontos(amostras). Os pontos foram coloridos de forma diferente levando em conta a sua espécie.

Note-se que o mapa mostra a presença de 3 grupos/segmentos e que estes de fato correspondem aproximadamente a cada uma das espécies.

Figura 6



REFERÊNCIAS

- Benzeeri, J.P. (1973). L'Analyse des Données, Tome 2: L'Analyse des Correspondences, Dunod, Paris (2a. ed. 1976).
- Cattel, R.B. (1966). The Scree test for the number of factors. *Multivariate Behavioural Research*, 1, 245-276.
- Elena Deza e outros (2006). *Dictionary of Distances*, Elsevier.
- Fischer, R.A. (1940). The Precision of Discriminant Functions. *Ann. Eugen. Lond.* 10, 422-429.
- Gower, J.C. (1982). Euclidean Distance Geometry. *Mathematical Scientist*, 7, 1-14.
- Gower, J.C. & Hand, D.J. (1996). *Biplots*. Chapman & Hall.
- Gower, J.C., Lubbe, S., Le Roux N. (2011). *Understanding Biplots*. Wiley & Sons.
- Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 5m, 3,, p.453.
- Hand, D.J., Daly, F., Lunn, A.D. (1994). *A handbook of small data sets*. London: Chapman & Hill. pag. 297.
- Hill, M.O. (1974). Correspondence Analysis: a Neglected Multivariate Method. *App. Stat* 23, 340-354.
- Hirschfied, H.O. (1935). A Connection Between Correlation and Contingency. *Proc. Camb. Phil. Soc.* 31, 520-524.
- Hoffman, K., Kunze, R. (1961). *Linear Algebra*. Prentice-Hall, Inc. Englewood Cliffs, N.J.
- Horst, P. (1965). *Factor Analysis of Data Matrices*. Holt, Rinchart & Winston, New York.
- Kendall, M.G. & Stuart, A. (1961). *The Advanced Theory of Statistics*. vol. 2. Griffin, London.

- Kettinger, J.R. (1971). Canonical Analysis of Several Sets of Variables. *Biometrika* 58, 433-450.
- Kruskal, J.B. (1964). Multidimensional scating by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, vol. 29, 1-27.
- Kruskal, J.B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, vol. 29, 28-42.
- Lima Elon L. (1998). Álgebra Linear. Coleção Matemática Universitária. IMPA, Rio de Janeiro.
- Lancaster, H.O. (1969). The χ -squared Distribution. Wiley, New York.
- Lebart, L., Fenelon, J.P. (1971). Statistique et Informatique Appliquées. Dunod, Paria (3a. ed. 1975).
- Lebart, L., Morineau, A. & Tabard, N. (1977). Techniques de la Description Statistique. Dunod. Paris.
- Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Chapter 14 of *Multivariate Analysis*, London: Academic Press.
- Morrison, D.F. (1967). Multivariate Statistical Methods. McGraw Hill (2a. ed. 1976).
- Pitombeira, J.C. (1972). Introdução à Álgebra Linear. Ao Livro Técnico S.A. e Editora Universidade de Brasília. Rio de Janeiro.
- Sherpard, R.N. (1962). The analysis of proximities: Muldidimensional scaling with an unknown distance function. *Phychometrika* vol. 27, 125-139, 219-246.