

4. ANÁLISE FACTORIAL DAS CORRESPONDÊNCIAS

TRANSFORMAÇÃO INICIAL DA MATRIZ DE PARTIDA

Seja Q um quadro $n \times p$ de números positivos constituído pelos elementos $K(i,j)$. Este quadro pode ser (como foi referido no capítulo ABORDAGEM INTUITIVA DA ANÁLISE DE DADOS) assimilado a uma tabela de contingência (Fig. 4.1).

	1	j	p	SOMA		
1						$K(i)$		
·								
·								
i							$K(i,j)$	$K(i)$
·								
·						K		
n							$K(j)$	K
SOMA								

Fig. 4.1 - Tabela de contingência.

Embora em Análise das Factorial das Correspondências (AFC) haja uma simetria perfeita entre o estatuto das linhas e colunas da matriz, manter-se-á a designação de indivíduos para as linhas da matriz ($i = 1 \dots n$; $i \in I$) e de propriedades para as colunas ($j = 1 \dots p$; $j \in J$).

Partindo da matriz de termo geral $K(i,j)$, a AFC efectua a transformação particular que a seguir se descreve.

Sendo $K = \sum_i \sum_j K(i,j)$ o número total de ocorrências,

calcula-se $f_{ij} = \frac{K(i,j)}{K}$ (4.1)

obtendo-se assim as frequências relativas de cada elemento da matriz.

Sendo $K(i) = \sum_j K(i, j)$ e $K(j) = \sum_i K(i, j)$ as somas em linha e em coluna, as respectivas

frequências marginais são :

$$f_i = \sum_{j=1}^p f_{ij} = \sum_{j=1}^p \frac{K(i, j)}{K} = \frac{K(i)}{K} \quad (4.2) \text{ frequência marginal do indivíduo } i$$

$$f_j = \sum_{i=1}^n f_{ij} = \sum_{i=1}^n \frac{K(i, j)}{K} = \frac{K(j)}{K} \quad (4.3) \text{ frequência marginal da propriedade } j$$

Evidentemente que se tem $\sum_i f_i = \sum_j f_j = 1$

O conjunto das coordenadas de um indivíduo i em R^p constitui o seu perfil f_j^i .

Uma coordenada de ordem j desse perfil é dada por:

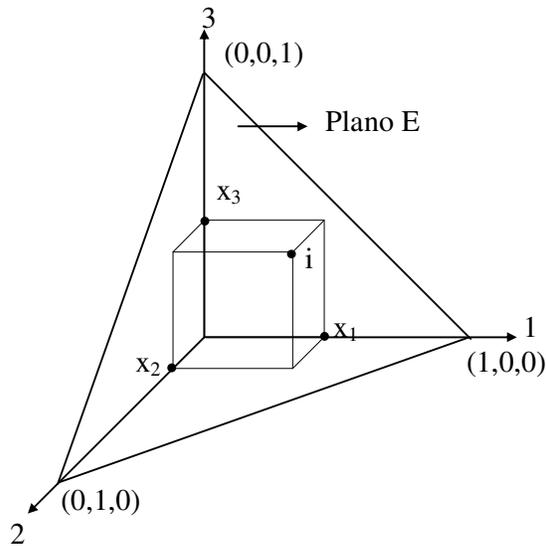
$$f_j^i = \frac{f_{ij}}{f_i} = \frac{\frac{K(i, j)}{K}}{\frac{K(i)}{K}} = \frac{K(i, j)}{K(i)} \quad (4.4)$$

Para tabelas de contingência, as coordenadas dos indivíduos podem ser interpretadas como a probabilidade condicional de ocorrer j , sabendo i .

Verifiquemos que a soma das coordenadas de um indivíduo é unitária :

$$\sum_{j=1}^p f_j^i = \sum_{j=1}^p \frac{f_{ij}}{f_i} = \frac{\sum_{j=1}^p K(i, j)}{K(i)} = \frac{K(i)}{K(i)} = 1 \quad (4.5)$$

Como as coordenadas dos indivíduos têm de respeitar (4.5), estas encontram-se já num espaço de dimensão $p-1$, por força da transformação inicial de dados específica da AFC. Por exemplo, um conjunto I de indivíduos caracterizado por 3 propriedades encontra-se já no plano E da Fig. 4.2.



O indivíduo i situa-se no plano E de equação $x_3 = 1 - x_2 - x_1$.

Fig. 4.2 - Plano que contém os indivíduos caracterizados pelas propriedades 1, 2 e 3.

O conjunto das coordenadas de uma propriedade j em R^n constitui o seu perfil f_j^i .

Uma coordenada de ordem i desse perfil é dada por

$$f_i^j = \frac{f_{ij}}{f_j} = \frac{\frac{K(i,j)}{K}}{\frac{K(j)}{K}} = \frac{K(i,j)}{K(j)} \quad (4.6)$$

Para tabelas de contingência, as coordenadas das propriedades podem ser interpretadas como a probabilidade condicional de ocorrer i , sabendo j .

Também em R^n a soma das coordenadas de uma propriedade é unitária:

$$\sum_{i=1}^n f_i^j = \sum_{i=1}^n \frac{f_{ij}}{f_j} = \frac{\sum_{i=1}^n K(i,j)}{K(j)} = \frac{K(j)}{K(j)} = 1$$

É através desta transformação simétrica das linhas e colunas que se assegura em AFC a representação simultânea dos indivíduos e propriedades no mesmo espaço.

CONSTRUÇÃO DA MATRIZ DE INÉRCIA

Em paralelo com o objectivo da ACP, em AFC pretende-se encontrar os eixos de inércia da nuvem dos indivíduos (ou propriedades) no espaço das propriedades (ou indivíduos). O problema é pois a pesquisa dos vectores próprios de uma matriz de inércia particular, em que as coordenadas dos pontos são dadas pelos respectivos perfis f_j^i ou f_i^j e em que os pesos e distâncias são definidas seguidamente.

Pesos

Peso do indivíduo i :
$$f_i = \frac{K(i)}{K} \quad (\text{frequência marginal do indivíduo } i) \quad (4.7)$$

Peso da propriedade j :
$$f_j = \frac{K(j)}{K} \quad (\text{frequência marginal da propriedade } j) \quad (4.8)$$

Com este sistema de pesos e coordenadas pode calcular-se a coordenada de ordem j do Centro de Gravidade (g_j) da nuvem dos indivíduos $N(I)$:

$$g_j = \sum_{i=1}^n f_i \cdot f_j^i = \sum_{i=1}^n f_i \frac{f_{ij}}{f_i} = \sum_{i=1}^n f_{ij} = \frac{K(j)}{K} = f_j$$

f_i
 \downarrow
 pesos

f_j^i
 \searrow
 coordenadas

O Centro de Gravidade da nuvem dos indivíduos é o perfil da linha marginal do quadro de partida.

Designando esse perfil por f_j , pode escrever-se

$$g_j = f_j \quad (4.9)$$

Simetricamente, vem para o Centro de Gravidade da nuvem das propriedades $N(J)$:

$$g_i = f_i \quad (4.10)$$

A equação(4.10) interpreta-se dizendo que o Centro de Gravidade das propriedades é o perfil da coluna marginal do quadro de partida (a qual coincide com o peso dos indivíduos).

Distância

A distância utilizada em AFC não é a distância euclideana habitual, mas uma distância ponderada* pelo inverso da massa (distância do χ^2). A distância entre 2 indivíduos i e i' em R^p é dada por:

$$d^2(i,i') = \sum_{j=1}^p \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_{i'}} \right)^2 = \sum_{j=1}^p \frac{1}{f_j} (f_j^i - f_j^{i'})^2 \quad (4.11)$$

Simetricamente, a distância entre 2 propriedades j e j' (em R^n) é dada por:

$$d^2(j,j') = \sum_{i=1}^n \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - \frac{f_{i'j'}}{f_{j'}} \right)^2 = \sum_{i=1}^n \frac{1}{f_i} (f_i^j - f_i^{j'})^2 \quad (4.12)$$

A distância do χ^2 , para além da atenuação das coordenadas trazida pela ponderação*, goza ainda da propriedade da equivalência distribucional, a qual permite substituir dois indivíduos de idêntico perfil por um único indivíduo de peso igual à soma dos pesos.

Note-se que a distância do χ^2 se pode converter numa distância euclideana modificando a escala dos eixos. Retomando (4.11), obtém-se

$$d^2(i,i') = \sum_{j=1}^p \left(\frac{f_{ij}}{\sqrt{f_j f_i}} - \frac{f_{i'j}}{\sqrt{f_j f_{i'}}} \right)^2 \quad (4.13)$$

* Esta ponderação permite “equilibrar” a ocorrência de propriedades com diferentes ordens de grandeza – por exemplo, se se tratar de um problema cruzando amostras com teores em diferentes elementos, a distância do χ^2 tenderá a atenuar o peso das variáveis com o teor em SiO₂ (o qual mascará a distância euclideana habitual, visto ser, em geral, muito mais elevado do que o teor em elementos ditos “menores”).

Para trabalhar com a distância euclidiana habitual é preciso fazer a seguinte transformação de coordenadas:

$$\frac{f_{ij}}{f_i} \xrightarrow{\text{curved arrow}} \frac{f_{ij}}{\sqrt{f_j f_i}} \quad R^p \quad (4.14)$$

$$\frac{f_{ij}}{f_j} \xrightarrow{\text{curved arrow}} \frac{f_{ij}}{\sqrt{f_i f_j}} \quad R^n \quad (4.15)$$

Para calcular a matriz de inércia da nuvem é necessário calcular a distância de cada ponto ao Centro de Gravidade .

Em R^p , e dada a transformação de métrica exigida pelo χ^2 , as coordenadas do Centro de Gravidade são agora dadas por :

$$g_j = \sum_{i=1}^n f_i \frac{f_{ij}}{\sqrt{f_j f_i}} = \frac{f_j}{\sqrt{f_j}} = \sqrt{f_j} \quad (4.16)$$

\downarrow pesos \downarrow coordenadas

Inércia

O elemento genérico da matriz de inércia* V escreve-se como o produto dos pesos pelo quadrado das distâncias ao Centro de Gravidade. Usando a transformação exigida pelo χ^2 , vem:

$$v_{jj'} = \sum_{i=1}^n f_i \left(\frac{f_{ij}}{\sqrt{f_j f_i}} - \sqrt{f_j} \right) \left(\frac{f_{ij'}}{\sqrt{f_{j'} f_i}} - \sqrt{f_{j'}} \right) \quad (4.17)$$

\downarrow pesos \downarrow distância ao CG \downarrow distância ao CG

* Note-se que esta matriz de inércia tem um vector próprio trivial porque, pela codificação particular da matriz de partida, o espaço e de dimensão $p-1$ (ou $n-1$), dado que a soma das linhas (ou colunas) é unitária.

Em R^n , o elemento genérico da matriz de inércia W escreve-se:

$$w_{jj'} = \sum_{j=1}^p f_j \left(\frac{f_{ij}}{\sqrt{f_i f_j}} - \sqrt{f_i} \right) \left(\frac{f_{i'j}}{\sqrt{f_{i'} f_j}} - \sqrt{f_{i'}} \right) \quad (4.18)$$

A matriz a diagonalizar para encontrar os valores próprios e vectores próprios (eixos principais de inércia) pode ser V ou W consoante $p < n$ ou $p > n$. Nos eixos principais de inércia são depois projectados os indivíduos e propriedades.

PROJECÇÃO DE INDIVÍDUOS E PROPRIEDADES. PRINCÍPIO BARICÊNTRICO.

A matriz V de termo geral $v_{jj'}$ dado por (4.17) :

$$v_{jj'} = \sum_{i=1}^n f_i \left(\frac{f_{ij}}{\sqrt{f_j f_i}} - \sqrt{f_j} \right) \left(\frac{f_{i'j'}}{\sqrt{f_{j'} f_i}} - \sqrt{f_{j'}} \right)$$

pode ser escrita como o produto de duas matrizes $X^T X$.

De facto, fazendo

$$x_{ij} = \frac{f_{ij} - f_i f_j}{\sqrt{f_i f_j}}$$

x_{ij} é o termo geral de uma matriz X tal que $X^T X = V$, como se pode verificar fazendo o produto x_{ij}

$x_{ij'}$:

$$x_{ij} x_{ij'} = \frac{(f_{ij} - f_i f_j)(f_{i'j'} - f_i f_{j'})}{f_i \sqrt{f_j} \sqrt{f_{j'}}$$

Multiplicando e dividindo por f_i , vem

$$x_{ij} x_{ij'} = f_i \left(\frac{f_{ij} - f_i f_j}{f_i \sqrt{f_j}} \cdot \frac{f_{ij'} - f_i f_{j'}}{f_i \sqrt{f_{j'}}} \right)$$

$$x_{ij} x_{ij'} = f_i \left(\frac{f_{ij}}{f_i \sqrt{f_j}} - \sqrt{f_j} \right) \left(\frac{f_{ij'}}{f_i \sqrt{f_{j'}}} - \sqrt{f_{j'}} \right)$$

o que, depois de somado em n , reproduz (4.17). A matriz V admite um valor próprio nulo correspondente ao vector próprio com coordenadas $\sqrt{f_j}$. De facto, um dado vector próprio é a parte da inércia total da nuvem explicada pelo respectivo valor próprio. Então, se $\sqrt{f_j}$ for um vector próprio, a inércia por ele explicada é nula, porque se trata do Centro de Gravidade.

Como os vectores próprios são ortogonais, e sendo u_α um vector próprio de V diferente de $\sqrt{f_j}$, vem

$$\sum_{j=1}^p u_{\alpha j} \sqrt{f_j} = 0$$

Como $\sqrt{f_j}$ é vector próprio de V correspondente ao valor próprio nulo, qualquer outro vector próprio de V é também vector próprio da matriz de inércia não centrada T de termo geral $t_{jj'}$:

$$t_{jj'} = \sum_{i=1}^n f_i \frac{f_{ij} f_{ij'}}{f_i^2 \sqrt{f_j f_{j'}}} = \sum_{i=1}^n \frac{f_{ij} f_{ij'}}{f_i \sqrt{f_j f_{j'}}} \quad (4.19)$$

Esta matriz T pode ser escrita como produto de duas matrizes $X^{*T} X^*$. O termo geral x_{ij}^* de X^* escreve-se:

$$x_{ij}^* = \frac{f_{ij}}{\sqrt{f_i f_j}} \quad (4.20)$$

T tem agora um vector próprio trivial, correspondente ao valor próprio 1 .

A projecção do indivíduo i em u_α ; (vector próprio não trivial de ordem α de V ou T) pode escrever-se :

$$f'_{i\alpha} = \sum_{j=1}^p \left(\frac{f_{ij}}{f_i \sqrt{f_j}} \right) u_{\alpha j} \quad (4.21)$$

Se a matriz T tiver sido diagonalizada, obtêm-se os respectivos valores próprios e a expressão (4.21) dá a projecção dos indivíduos nos eixos factoriais.

Um raciocínio análogo em R^n , diagonalizando a matriz W e obtendo os respectivos vectores próprios v_α , leva ao cálculo das projecções das propriedades:

$$f'_{j\alpha} = \sum_{i=1}^n \left(\frac{f_{ij}}{f_j \sqrt{f_i}} \right) v_{\alpha i} \quad (4.22)$$

Mas as relações de transição, deduzidas na ANÁLISE GERAL, depois de adaptadas ao formalismo particular da AFC, permitem diagonalizar apenas uma das matrizes (a de menor dimensão).

Sejam então as relações (2.12) :

$$v_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X u_\alpha \quad (4.23)$$

$$u_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X^T v_\alpha \quad (4.24)$$

Fazendo $X = \overset{*}{X}$ (dado por (4.20)), a linha i da relação matricial (4.23) escreve-se:

$$v_{\alpha i} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p \left(\frac{f_{ij}}{\sqrt{f_i f_j}} \right) u_{\alpha j} \quad (4.25)$$

Multiplicando e dividindo por $\sqrt{f_i}$, e rearranjando, escreve-se:

$$\frac{\sqrt{\lambda_\alpha} v_{\alpha i}}{\sqrt{f_i}} = \sum_{j=1}^p \left(\frac{f_{ij}}{f_i \sqrt{f_j}} \right) u_{\alpha j} \quad (4.26)$$

Comparando (4.26) com (4.21), vem:

$$f'_{i\alpha} = \frac{\sqrt{\lambda_\alpha} v_{\alpha i}}{\sqrt{f_i}} \quad (4.27)$$

ou

$$v_{\alpha i} = \frac{f'_{i\alpha} \sqrt{f_i}}{\sqrt{\lambda_\alpha}} \quad (4.28)$$

A linha j da relação (4.24) escreve-se :

$$u_{\alpha j} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \left(\frac{f_{ij}}{\sqrt{f_i f_j}} \right) v_{\alpha i}$$

Comparando com a relação (4.22), obtém-se:

$$f'_{j\alpha} = \frac{\sqrt{\lambda_\alpha} u_{\alpha j}}{\sqrt{f_j}}$$

Substituindo em (4.22) o valor de $v_{\alpha i}$ tirado de (4.28) obtém-se :

$$f'_{j\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \left(\frac{f_{ij}}{f_j} \right) f'_{i\alpha} \quad (4.29)$$

Simetricamente, obtém-se:

$$f'_{i\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p \left(\frac{f_{ij}}{f_i} \right) f'_{j\alpha} \quad (4.30)$$

As relações (2.29) e (4.30) exprimem o PRINCÍPIO BARICÊNTRICO e significam que a projecção de um indivíduo é (a menos do factor de escala $\frac{1}{\sqrt{\lambda_\alpha}}$) o Centro de Gravidade das projecções de todas as propriedades (com pesos dados pelo perfil do indivíduo) e que a projecção de uma propriedade é o Centro de Gravidade das projecções de todos os indivíduos (a menos do factor de escala $\frac{1}{\sqrt{\lambda_\alpha}}$).

É assim possível dar um sentido às projecções simultâneas de indivíduos e propriedades no mesmo espaço. As projecções dos elementos de uma nuvem (por exemplo, os indivíduos em R^p) são o baricentro dos elementos da outra (as propriedades em R^n). Para a interpretação em gráficos planos, o princípio baricêntrico assegura que um indivíduo esteja “rodeado” das propriedades com que ele mais se associa (e vice-versa).

PROJECÇÃO DE ELEMENTOS EM SUPLEMENTAR

Em alguns casos é necessário, após efectuar uma AFC de um certo conjunto de dados (sob a forma de uma matriz $n \times p$), considerar elementos (linhas ou colunas) “suplementares” relativamente à matriz de partida. Os elementos suplementares têm um estatuto diferente dos $n \times p$ elementos principais da matriz (vd. REGRAS DE INTERPRETAÇÃO EM AFC) e não intervêm na sua factorização, pretendendo apenas encontrar-se a posição das projecções em suplementar nos eixos factoriais que resultam da diagonalização da matriz principal.

Para efectuar a projecção em suplementar, basta considerar as relações (4.29) e (4.30) que exprimem o princípio baricêntrico.

Seja pois um indivíduo suplementar de perfil $\frac{f_{i^+j}}{f_{i^+}}$.

As suas coordenadas obtêm-se com base na relação (4.30) :

$$f'_{i^+\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p \left(\frac{f_{i^+j}}{f_{i^+}} \right) f'_{j\alpha} \quad (4.31)$$

onde $f'_{j\alpha}$ são as coordenadas das propriedades relativas à análise em que o indivíduo não entrou (obtidas pela factorização do quadro de partida $n \times p$).

Uma propriedade suplementar de perfil $\frac{f_{ij^+}}{f_{j^+}}$ tem coordenadas que se obtêm da relação (4.29):

$$f'_{j^+\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \left(\frac{f_{ij^+}}{f_{j^+}} \right) f'_{i\alpha} \quad (4.32)$$

RECONSTITUIÇÃO DA MATRIZ DE PARTIDA

Para certas aplicações pode ser interessante dispor de um algoritmo que reconstitua a matriz de partida com base nas projecções dos indivíduos e propriedades nos eixos factoriais. Por outro lado, seleccionando apenas q eixos (com $q < p < n$) e reconstituindo a matriz de partida com base apenas nas projecções nesses q eixos, pode-se avaliar o grau de aproximação que resulta da exclusão dos valores próprios de ordem superior a q (comparando a matriz de partida e a matriz estimada pelas projecções em q eixos através do algoritmo de reconstituição).

Reescrevendo a formula de transição da ANÁLISE GERAL (2.12)

$$v_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X u_\alpha$$

obtém-se

$$X u_\alpha = \sqrt{\lambda_\alpha} v_\alpha$$

Pós multiplicando por u_a^T e somando em α (admitindo que p é o espaço de menor dimensão),

$$X \left\{ \sum_{\alpha=1}^p u_{\alpha} u_{\alpha}^T \right\} = \sum_{\alpha=1}^p \sqrt{\lambda_{\alpha}} v_{\alpha} u_{\alpha}^T \quad (4.33)$$

Como os vectores próprios são ortogonais, $\sum_{\alpha=1}^p u_{\alpha} u_{\alpha}^T$ é a matriz identidade.

Então
$$X = \sum_{\alpha=1}^p \sqrt{\lambda_{\alpha}} v_{\alpha} u_{\alpha}^T \quad (4.34)$$

tomando agora $v_{\alpha} = \frac{1}{\sqrt{\lambda_{\alpha}}} f'_{i\alpha} \sqrt{f_i}$ (relação (4.28)) e a simétrica

$$u_{\alpha} = \frac{1}{\sqrt{\lambda_{\alpha}}} f'_{j\alpha} \sqrt{f_j} \quad (4.35)$$

e substituindo em (4.33), vem

$$X = \sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_{\alpha}}} f'_{i\alpha} f'_{j\alpha} \sqrt{f_i f_j}$$

A matriz X tem termo geral dado por $x_{ij}^* = \frac{f'_{ij}}{\sqrt{f_i f_j}}$ (expressão (4.20)).

Então

$$f_{ij} = f_i f_j \sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_{\alpha}}} f'_{i\alpha} f'_{j\alpha} \quad (4.36)$$

Sabe-se que o primeiro valor próprio da matriz $T = X^T X$ é unitário e correspondente ao vector trivial $\sqrt{f_j}$. Para esse valor próprio, as expressões (4.28) e (4.35) dão $f'_{i1} = f'_{j1} = 1$.

Então (4.36) pode escrever-se

$$f_{ij} = f_i f_j \left\{ 1 + \sum_{\alpha=2}^p \frac{1}{\sqrt{\lambda_\alpha}} f'_{i\alpha} f'_{j\alpha} \right\} \quad (4.37)$$

Retomando a notação inicial do quadro de partida (expressões (4.1) (4.2) e (4.3)):

$$f_{ij} = \frac{K(i,j)}{K}, \quad f_i = \frac{K(i)}{K}, \quad f_j = \frac{K(j)}{K}, \quad \text{obtem-se}$$

$$K(i,j) = \underbrace{\frac{K(i) K(j)}{K}}_A \left\{ 1 + \sum_{\alpha=2}^p \frac{1}{\sqrt{\lambda_\alpha}} f'_{i\alpha} f'_{j\alpha} \right\} \quad (4.38)$$

Para uma tabela de contingência, a expressão (4.38) tem um significado claro: a quantidade A é o valor de $K(i,j)$ admitindo independência (quociente do produto das margens pelo número total de acontecimentos). É pois uma primeira aproximação do elemento $K(i,j)$. As aproximações de ordem superior são dadas pela soma dos produtos das projecções nos sucessivos eixos factoriais da linha e coluna onde o elemento $K(i,j)$ se encontra (a menos da ponderação pela importância de cada eixo, dada pelos sucessivos valores próprios λ_α).

Se q eixos forem suficientes para explicar (com a aproximação desejada) a forma da nuvem, pode-se escrever (4.38) como:

$$K(i,j) = \frac{K(i) K(j)}{K} \left\{ 1 + \sum_{\alpha=2}^q \frac{1}{\sqrt{\lambda_\alpha}} f'_{i\alpha} f'_{j\alpha} \right\}$$

CONTRIBUIÇÕES

Obtidas as projecções dos indivíduos e propriedades nos eixos factoriais, é útil calcular alguns índices de ajuda à interpretação. Estes índices são as contribuições absolutas e relativas.

a) A contribuição absoluta permite a explicação de um eixo factorial a partir dos indivíduos I ou propriedades J - é a contribuição do elemento para a formação do eixo α .

A inércia total explicada pelo eixo é λ_α (porque λ_α é o máximo da Lagrangeana - soma das projecções da inércia total da nuvem em α).

$$\lambda_\alpha = \sum_{i=1}^n f_i f_{i\alpha}'^2$$

\downarrow \searrow
 peso quadrado das distâncias das projecções de i em α

Então uma medida da contribuição do elemento i para a inércia total explicada pelo eixo é a relação entre a parte de inércia atribuível a i e a inércia total da nuvem transportada pelo eixo α .

$$C_{i\alpha}^a = \frac{f_i f_{i\alpha}'^2}{\lambda_\alpha}$$

Do mesmo modo, para as propriedades, vem $C_{j\alpha}^a = \frac{f_j f_{j\alpha}'^2}{\lambda_\alpha}$

b) A contribuição relativa permite a explicação de um elemento i ou j a partir dos eixos α - é a contribuição do eixo α para o elemento i ou j .

O quadrado da distância do ponto i ao CG dado por:

$$\rho^2 = \sum_{\alpha} f_{i\alpha}'^2$$

A contribuição relativa do eixo α para a distância de i ao CG é:

$$C_{\alpha i}^r = \frac{f_{i\alpha}'^2}{\rho^2} = \cos^2 \beta$$

onde β é o ângulo que o eixo α faz com o perfil do indivíduo i .

Do mesmo modo, para as propriedades, vem

$$C_{\alpha j}^r = \frac{f_{j\alpha}^{\prime 2}}{\rho^2}$$

ALGORITMO DA AFC

Sintetiza-se seguidamente o algoritmo a usar para resolver um problema de AFC.

1. Selecção do espaço de menor dimensão (seja R^p - admite-se que há menos propriedades do que indivíduos).
2. Transformação da matriz de partida - cálculo dos perfis dos indivíduos em R^p :

$$\frac{f_{ij}}{f_i} = \frac{K(i, j)}{K(i)}$$

e das propriedades em R^n :

$$\frac{f_{ij}}{f_j} = \frac{K(i, j)}{K(j)}$$

3. Cálculo da matriz de inércia V ($p \times p$) de termo geral $v_{jj'}$:

$$v_{jj'} = \sum_{i=1}^n f_i \left(\frac{f_{ij}}{\sqrt{f_j f_i}} - \sqrt{f_j} \right) \left(\frac{f_{ij'}}{\sqrt{f_j' f_i}} - \sqrt{f_j'} \right)$$

ou da matriz simétrica equivalente T de termo geral $t_{jj'}$:

$$t_{jj'} = \sum_{i=1}^n \frac{f_{ij} f_{ij'}}{f_i \sqrt{f_j f_j'}}$$

4. Diagonalização da matriz de inércia T e obtenção dos seus valores próprios λ_α e vectores próprios u_α .
5. Análise do histograma dos valores próprios e escolha da dimensão do espaço cuja inércia acumulada explique uma percentagem significativa da inércia inicial (em geral 2 ou 3 eixos)
6. Projecção dos indivíduos nos eixos de inércia retidos:

$$f'_{i\alpha} = \sum_{j=1}^p \frac{f_{ij}}{f_i \sqrt{\lambda_\alpha}} u_{\alpha j}$$

7. Projecção das propriedades nos eixos de inércia retidos :

$$f'_{j\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \left(\frac{f_{ij}}{f_j} \right) f'_{i\alpha}$$

8. Cálculo eventual da projecção de elementos em suplementar .
9. Reconstituição eventual da matriz de partida e confirmação da aproximação escolhida (número de valores próprios retidos).
10. Cálculo das contribuições absolutas e relativas.
11. Interpretação com base em valores próprios, projecção dos indivíduos e propriedades nos eixos factoriais, contribuições absolutas e relativas, etc.

REGRAS DE INTERPRETAÇÃO EM AFC

Em resumo, a interpretação dos resultados de uma AFC consiste em atribuir um significado aos eixos de inércia em termos das propriedades (ou indivíduos) que os explicam; num segundo passo, as relações entre indivíduos e propriedades são interpretadas com base no significado conferido aos eixos no primeiro passo.

A primeira decisão a tomar é a escolha do número de eixos de inércia a reter com base na análise do histograma dos valores próprios. Procura-se um compromisso entre a conformidade da nuvem inicial com as suas projecções (% aceitável de inércia explicada) e a simplicidade da interpretação (pequeno número de eixos a interpretar). A reconstituição da matriz de partida pode também dar indicações sobre o número de eixos a reter.

Escolhido um certo número de eixos, interpreta-se cada um deles a partir da contribuição absoluta das propriedades (ou indivíduos) para esses eixos. Cada eixo é explicado pelas propriedades (ou indivíduos) cuja contribuição absoluta é mais importante*. A análise das projecções nos eixos dá o sentido (positivo ou negativo) da relação entre as propriedades (ou indivíduos) que explicam cada eixo. Pode portanto traçar-se gráficos unidimensionais onde se analisa a posição das projecções das propriedades (ou indivíduos) retidos para a explicação de cada eixo.

As contribuições relativas intervêm na análise de propriedades ou indivíduos particulares, para os quais existe um interesse especial. Para esses elementos, pode seleccionar-se quais os eixos que mais contribuem para a explicação do elemento e, como os eixos já estão interpretados, é possível tirar conclusões sobre o elemento particular (e relacioná-lo com os outros, que foram retidos para a interpretação dos eixos).

Interpretados os eixos, segue-se a análise das projecções dos indivíduos e propriedades nos gráficos planos que cruzam todos os pares de eixos retidos. Um par de indivíduos ou propriedades será tanto mais semelhante (a menos de erros de perspectiva que são esclarecidos por outra projecção) quanto mais próximas forem as suas projecções.

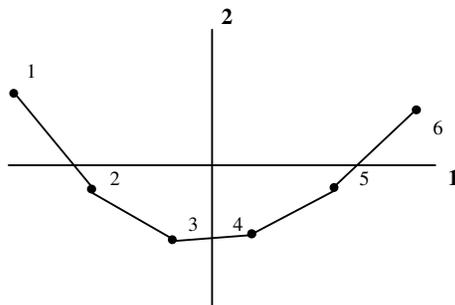
* Como regra prática, pode dizer-se que uma propriedade (ou indivíduo) é retida se a sua contribuição absoluta for superior a $\frac{100}{p}$ (ou $\frac{100}{n}$). Este limite é dado pela distribuição uniforme das contribuições pelas propriedades (ou indivíduos).

Podem assim surgir grupos de indivíduos cuja afinidade é esclarecida pela projecção das propriedades na proximidade desses grupos (ou vice-versa) .

A forma da nuvem, depois de projectada nos eixos factoriais, dá também indicações sobre a estrutura subjacente ao quadro de partida. Procurou-se sintetizar, no Quadro 4.1, o conjunto de formas típicas que é habitual surgir nas projecções obtidas em AFC, bem como a relação dessas formas com determinados aspectos do quadro de partida, depois de este ser rearranjado de acordo com as indicações fornecidas pela interpretação .

Evidentemente que a análise feita no plano 1, 2 tem de ser complementada pelas projecções em eixos de ordem superior, até esgotar os factores retidos.

Em variáveis ordinais (cujas classes sucessivas traduzem um fenómeno interpretável, por exemplo, a sequência das idades) é habitual ligar entre si as projecções das modalidades ordenadas (Fig. 4.3).

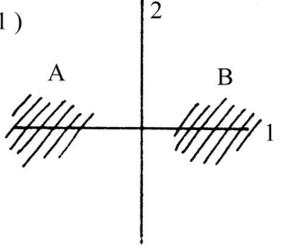
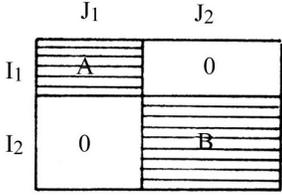
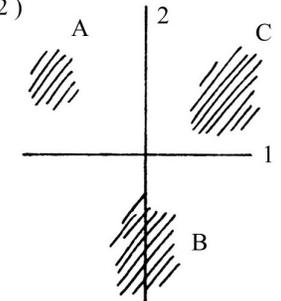
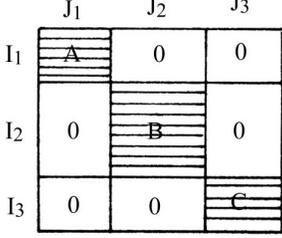
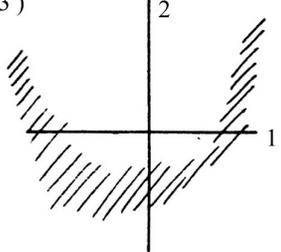
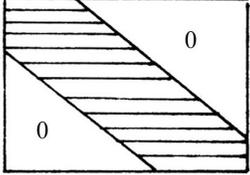
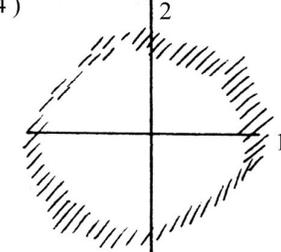
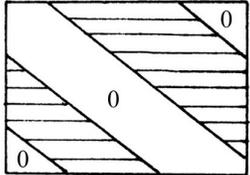
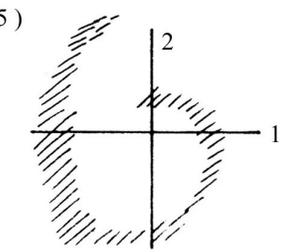
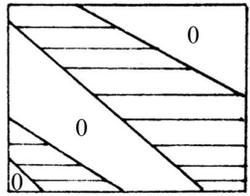


O eixo 1 opõe as classes de ordem baixa. O eixo 2 separa classes extremas das intermédias.

Fig. 4.3 - Projecção de modalidades ordenadas

Em quadros de notas desdobradas (vd. exemplo da Fig. 1.5) em que a soma em linha é constante ou em tabelas de duas modalidades codificadas sob a forma disjuntiva completa (em que essa soma também é constante), a projecção das duas modalidades referentes a cada propriedade alinham-se com a origem (Fig. 4.4). A interpretação dos eixos faz-se com base nas modalidades cuja projecção está mais longe da origem e mais próxima do eixo.

Quadro 4.1 – Forma da nuvem das projecções e sua interpretação.

Forma da Nuvem	Interpretação	Quadro de Partida
<p>(1)</p> 	<p>A nuvem divide-se em 2 grupos claramente opostos pela sua projecção no eixo 1. Rearranjado o quadro de partida, verifica-se que este se pode sectionar em dois blocos A e B rodeados de elementos quase nulos.</p>	
<p>(2)</p> 	<p>A nuvem divide-se em 3 grupos. A interpretação do eixo 2 intervém na distinção entre (A+C) e B. Retomando o quadro de partida, obtém-se os 3 blocos rodeados de elementos quase nulos. Pode prosseguir-se a análise estudando separadamente $I_1 \times J_1$, $I_2 \times J_2$, $I_3 \times J_3$.</p>	
<p>(3)</p> 	<p>A nuvem apresenta uma forma parabólica que indica uma clara ordenação sobre o eixo 1. Ordenando os indivíduos e propriedades segundo as suas projecções no eixo 1, e regressando ao quadro de partida, obtém-se uma estrutura diagonal.</p>	
<p>(4)</p> 	<p>A nuvem apresenta uma estrutura circular que se pode obter de (3) considerando nova parábola invertida sobre a primeira. Cada parábola corresponde a uma estrutura diagonal. O quadro de partida contém pois duas estruturas diagonais e evidencia uma estrutura cíclica.</p>	
<p>(5)</p> 	<p>A nuvem apresenta a forma de uma espiral, correspondente à justaposição de duas parábolas diferentes. O quadro de partida contém duas estruturas “desiguais” do tipo “diagonal assimétrica”.</p>	

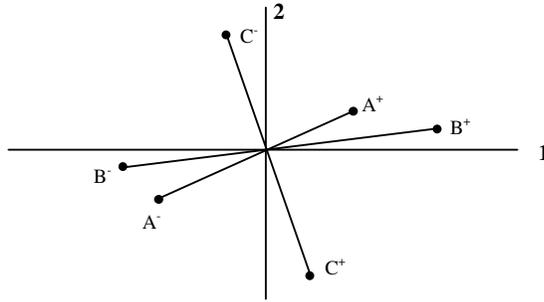


Fig. 4.4 - Projecção das propriedades em quadros de margem constante em linha.

Este tipo de figuras resulta da construção geométrica que a seguir se descreve. Sejam j^+ e j^- as duas modalidades referentes à mesma propriedade de um quadro cuja margem em linha é constante. Então, verificam-se as seguintes relações :

$$K(i) = m q$$

onde m é a soma dos elementos de um par de colunas referentes à mesma propriedade e q é o número de propriedades (o número total de colunas é $p = 2 q$).

$$K = n m q$$

onde n é o número de linhas

$$f_i = \frac{K(i)}{K} = \frac{1}{n}$$

Calculemos a soma dos perfis das colunas j^+ e j^- , ponderadas pelos respectivos pesos:

$$f_{j^+} \frac{f_{ij^+}}{f_{j^+}} + f_{j^-} \frac{f_{ij^-}}{f_{j^-}} = f_{ij^+} + f_{ij^-} = \frac{K(i, j^+) + K(i, j^-)}{K} = \frac{m}{mnq} = \frac{1}{nq} = \frac{1}{q} f_i \quad (4.39)$$

Então f_i é o centro de gravidade das duas colunas j^+ e j^- porque a soma dos seus pesos é $\frac{1}{q}$, como se verifica seguidamente:

$$f_{j^+} + f_{j^-} = \frac{K(j^+) + K(j^-)}{K} = \frac{mn}{mnq} = \frac{1}{q}$$

A relação (4.39) pode escrever-se, de um modo geral:

$$f_{j^+} f_{Ij^+} + f_{j^-} f_{Ij^-} = (f_{j^+} + f_{j^-}) f_I$$

o que pode pôr-se sob a forma:

$$f_{j^+} |f_{Ij^+} - f_I| = f_{j^-} |f_{Ij^-} - f_I| \quad (4.40)$$

A relação (4.40) exprime o “princípio da alavanca” e explica, tomando as projecções nos eixos factoriais, a forma particular da Fig. 4.4 .

* * *

Depois de efectuado um primeiro ensaio de interpretação (com base no conjunto de regras anteriormente expostas), podem surgir conclusões parciais que sugiram nova análise. Tal análise incidirá sobre um novo quadro de partida, obtido do inicial depois de uma reformulação baseada em alguns dos seguintes aspectos:

(i) Eliminação de Indivíduos ou Variáveis pouco Significativas

Se um determinado elemento se projecta na proximidade da origem, esse elemento não contribui para a definição da forma da nuvem e pode portanto ser eliminado (se houver acordo do especialista do domínio em estudo). Também um indivíduo ou propriedade que seja “anómalo” relativamente a uma estrutura com significado interpretável (perturbando, por exemplo, uma das

formas típicas apresentadas no Quadro 4.1) pode ser eliminado, depois de uma análise crítica das suas relações com a estrutura. Para certos objectivos, pode mesmo acontecer que sejam precisamente os elementos anómalos os mais interessantes.

(ii) Projecção de Elementos em Suplementar

Tudo o que foi dito em (i) sobre a eliminação de indivíduos ou propriedades é válido para a sua projecção em suplementar sobre os factores obtidos através da AFC de um quadro de onde foram suprimidos os referidos indivíduos ou propriedades. A vantagem desta atitude, em alternativa à simples eliminação, provém do facto de se ficar a conhecer a posição do elemento relativamente à estrutura que por ele era eventualmente perturbada. Outra aplicação da projecção em suplementar pode resultar de uma segmentação da matriz de partida em blocos de variáveis ou indivíduos com significado afim (essa segmentação pode resultar do primeiro ensaio de interpretação ou do conhecimento, *a priori*, existente sobre os dados). Então, pode ser interessante projectar em suplementar um bloco sobre outro para fazer ressaltar o sistema de relações entre os dois (esse sistema de relações pode “diluir-se” quando os blocos são factorizados simultaneamente). O exemplo típico desta projecção em suplementar é, no tratamento de questionários, a projecção das variáveis de “opinião” sobre as que definem o estatuto socio-económico dos inquiridos.

(iii) Modificação da Codificação das Variáveis

Alterando, por exemplo, os limites que definem cada classe de uma variável ordinal ou nominal (ou de uma variável mensurável transformada em ordinal), podem surgir novas estruturas com maior significado para a interpretação. Por outro lado, modificando o estatuto de certas variáveis

(passando as quantitativas a qualitativas, por exemplo), podem ressaltar certos aspectos de síntese que não eram aparentes após o primeiro ensaio de interpretação.

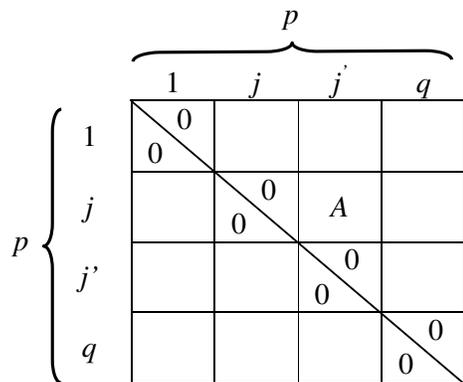
(iv) Reconstrução do Quadro de Partida

Já foi anteriormente referido que um mesmo conjunto de dados, pode ser, organizado segundo diferentes modelos de quadros de partida. Assim, por exemplo, um quadro de descrição lógica (presença-ausência) pode ser transformado numa justaposição de tabelas de contingência (conduzindo em geral, a resultados mais facilmente interpretáveis sob esta última forma).

Seja então um quadro Q ($n \times p$) de descrição lógica codificada sob a forma disjuntiva completa.

Seja q o número de propriedades e n o número de indivíduos.

Fazendo o produto $Q^T Q$ obtém-se uma matriz $p \times p$ que cruza entre si todas as modalidades das q propriedades. Essa matriz tem uma estrutura particular, designada por quadro de Burt (vd. Fig. 4.5) e é constituída por uma justaposição de tabelas de contingência cruzando todas as propriedades entre si.



A matriz de Burt é simétrica. Os blocos jj são matrizes diagonais contendo o número total de indivíduos distribuídos pelas modalidades de j ; os blocos do tipo A são tabelas de contingência cruzando as modalidades de j com j' .

Fig. 4.5 - Matriz de Burt.

Sobre a matriz de Burt pode agora projectar-se em suplementar a matriz Q ($n \times p$) para obter a posição dos indivíduos no espaço das propriedades.

EXEMPLOS DE APLICAÇÃO DA AFC

Apresentam-se neste capítulo alguns exemplos de tratamento de quadros multidimensionais de diferentes tipos através da AFC.

1. Justaposição de Tabelas de Contigência

Dispõe-se dos resultados de 50 ensaios de lixiviação de um minério de U, efectuados nas mesmas condições operacionais.

Construiu-se uma série de 5 tabelas de contingência cruzando a variável Recuperação (1 - teor do resíduo/teor da alimentação) com cada uma das variáveis que caracterizam o ambiente geológico onde a amostra submetida ao ensaio foi colhida (Presença de mineralização expressa, presença de metamorfismo de contacto, litologia, estado de oxidação e teor da alimentação). As variáveis que intervêm no processo não são homogêneas: existem 2 variáveis mensuráveis (recuperação e teor da alimentação), 3 variáveis nominais (presença de mineralização expressa e de metamorfismo, litologia) e uma variável ordinal (estado de oxidação). Para que seja possível tratar este conjunto de variáveis através de qualquer algoritmo de Análise de Dados, é necessário dar-lhe uma forma homogênea, através de codificação apropriada. Neste caso adoptou-se a solução de segmentar as variáveis quantitativas em três modalidades, transformando-as em ordinais. Obtém-se assim o quadro de partida da Fig. 4.6.

		MINERALIZAÇÃO		METAMORFISMO		LITOLOGIA		ESTADO DE OXIDAÇÃO		TEOR DA ALIMENTAÇÃO		
		Sim	Não	Sim	Não	Xisto	Grés	Reduzido	Oxidado	F ₁	F ₂	F ₃
Recuperação	R1											
	R2					A						
	R3											

Fig. 4.6 - Justaposição de tabelas de contigência. O elemento A, por exemplo, é o número de ensaios em que se registou uma recuperação “média” R2 para a amostra de Xisto.

Como a menor dimensão da matriz de partida é 3, 2 eixos explicam 100% da inércia da nuvem. Neste caso, o eixo 1 explica 92% da inércia total. As projecções no plano 1, 2 estão representadas na Fig. 4.7.

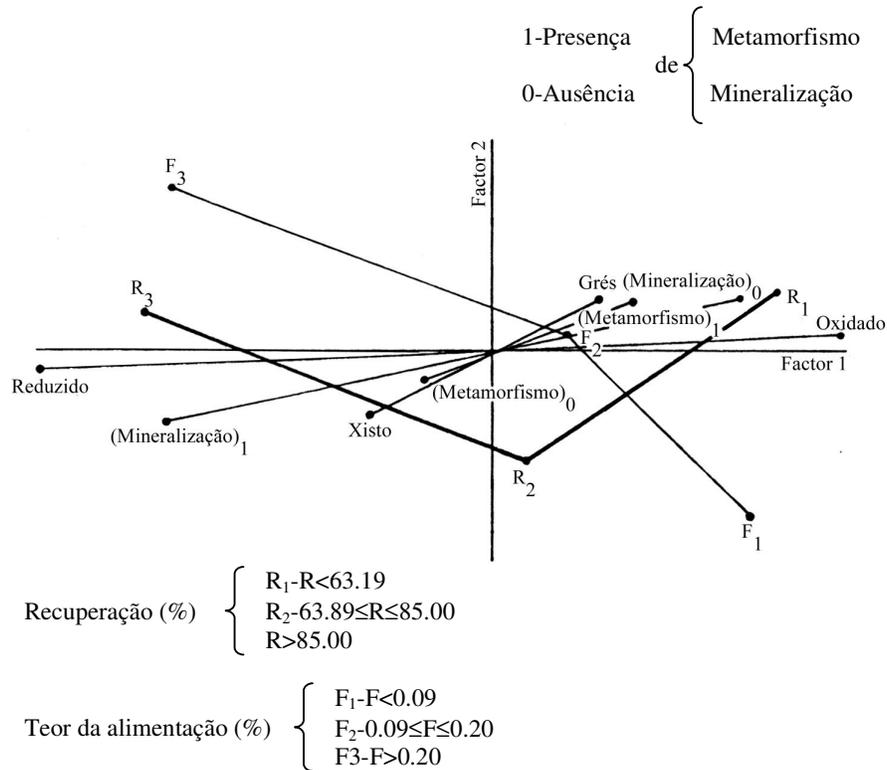


Fig. 4.7 - Projecção das modalidades das variáveis nos dois eixos factoriais resultantes da AFC sobre quadro esquematizado na figura 4.6.

Analisando a Fig. 4.7 pode concluir-se que, ao longo do Factor 1, se distribuem as três modalidades ordenadas da variável Recuperação ($R_1 \rightarrow R_2 \rightarrow R_3$). Então esse factor será interpretado como aquele que discrimina entre uma fácil recuperação (ligada às seguintes modalidades das variáveis geológicas: Alto teor de alimentação, Reduzido, Presença de mineralização, Xisto e Ausência de metamorfismo) e uma difícil recuperação (ligada às seguintes modalidades das variáveis geológicas: Baixo teor de alimentação, Oxidado, Ausência de mineralização, Grés e Presença de metamorfismo).

2. Notas Desdobradas

Retornando ao exemplo dos níveis de literacia em 3 vertentes para 20 países, os dados foram agora codificados em notas desdobradas. Cada uma das vertentes (Documental, Quantitativa, Prosa) foi desdobrada em 3 níveis (o nível (-) engloba a nota 1 e 2; o nível 0 engloba a nota 3 e o nível (+) corresponde às notas 4 e 5). Obtém-se assim a matriz 22×9 contendo as percentagens em cada categoria que se apresenta no Quadro 4.2.

Quadro 4.2 – Dados de partida com as notas desdobradas.

PAÍSES	DOC (-)	DOC (0)	DOC (+)	QUANT (-)	QUANT (0)	QUANT (+)	PROSA (-)	PROSA (0)	PROSA (+)
Canadá	42,8	32,1	25,1	43,0	34,8	22,2	42,2	35,1	22,7
Alemanha	41,6	39,5	18,9	33,3	43,2	23,5	48,6	38,0	13,4
Irlanda	57,0	31,5	11,5	53,1	30,7	16,2	52,4	34,1	13,5
Holanda	35,8	44,2	20,0	35,8	44,3	19,9	40,6	44,1	15,3
Polónia	76,2	18,0	5,8	69,3	23,9	6,8	77,1	19,8	3,1
Suécia	25,1	39,4	35,5	25,2	39,0	35,8	27,9	39,7	32,4
Suíça (francesa)	45,1	38,9	16,0	37,4	42,2	20,4	51,4	38,6	10,0
Suíça (alemã)	47,3	36,6	16,1	40,3	40,7	19,0	55,0	36,1	8,9
EUA	49,6	31,4	19,0	46,2	31,3	22,5	46,5	32,4	21,1
Austrália	44,9	37,7	17,4	43,2	37,7	19,1	44,2	36,9	18,9
Bélgica (Flandres)	39,6	43,2	17,2	39,6	37,8	22,6	46,7	39,0	14,3
Nova Zelândia	50,5	31,9	17,6	49,4	33,4	17,2	45,8	35,0	19,2
Reino Unido	50,4	30,5	19,1	51,0	30,4	18,6	52,1	31,3	16,6
Chile	86,8	11,7	1,5	83,1	14,3	2,6	85,1	13,3	1,6
República Checa	42,3	38,1	19,6	31,1	37,0	31,9	53,8	37,8	8,4
Dinamarca	32,0	42,6	25,4	27,7	43,9	28,4	46,0	47,5	6,5
Filândia	36,8	38,1	25,1	38,2	42,1	19,7	36,7	40,9	22,4
Hungria	67,0	25,0	8,0	52,2	31,7	16,1	76,6	20,8	2,6
Noruega	29,7	40,9	29,4	29,7	42,9	27,4	33,2	49,2	17,6
Portugal	80,2	16,6	3,2	71,8	23,0	5,2	77,1	18,5	4,4
Eslovénia	72,7	22,0	5,3	65,4	26,0	8,6	76,7	20,1	3,2
Suíça (italiana)	48,1	38,3	13,6	45,1	37,9	17,0	54,2	37,5	8,3

Submetendo esta matriz à AFC obtêm-se 6 eixos cujos valores próprios são dados no Quadro 4.3.

Quadro 4.3 – Valores próprios.

EIXO	VALORES PRÓPRIOS
1	0.06955
2	0.01052
3	0.00302
4	0.00100
5	0.00057
6	0.00019

As coordenadas e contribuições dos elementos submetidos à análise encontram-se nos Quadros 4.4 e 4.5 e o cruzando os 2 primeiros eixos é dado na Fig. 4.7.

Quadro 4.4 – Coordenadas.

PAÍSES	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6
Canadá	0.17	0.14	-0.02	0.02	-0.02	-0.01
Alemanha	0.16	-0.08	0.00	-0.04	-0.03	-0.01
Irlanda	-0.09	0.04	0.05	-0.02	0.05	-0.01
Holanda	0.23	-0.06	0.10	-0.01	-0.01	0.01
Polónia	-0.50	0.01	-0.01	0.03	-0.01	-0.01
Suécia	0.53	0.19	-0.11	-0.01	-0.01	0.01
Suíça (francesa)	0.08	-0.11	0.03	-0.02	-0.01	-0.01
Suíça (alemã)	0.03	-0.10	0.02	-0.01	-0.03	-0.01
EUA	0.07	0.13	-0.03	-0.03	0.02	0.00
Austrália	0.11	0.05	0.06	-0.04	0.01	0.00
Bélgica (Flandres)	0.15	-0.05	0.04	-0.04	0.04	0.05
Nova Zelândia	0.03	0.11	0.04	0.00	0.01	-0.01
Reino Unido	-0.01	0.09	-0.02	0.02	0.01	0.02
Chile	-0.71	0.07	-0.02	0.04	0.02	0.02
República Checa	0.15	-0.14	-0.12	-0.02	0.04	0.00
Dinamarca	0.28	-0.20	-0.04	0.07	0.01	0.00
Filândia	0.26	0.09	0.06	0.02	-0.04	-0.01
Hungria	-0.32	-0.10	-0.08	-0.04	-0.04	-0.01
Noruega	0.39	-0.02	0.00	0.07	0.01	-0.02
Portugal	-0.54	0.04	0.01	0.00	-0.01	-0.02
Eslovénia	-0.45	-0.02	-0.01	-0.01	-0.01	0.01
Suíça (italiana)	-0.01	-0.10	0.06	0.00	0.01	0.02
NÍVEL	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6
DOC (-)	-0.32	0.03	-0.01	-0.01	0.00	-0.02
DOC 0	0.25	-0.08	0.05	-0.03	0.02	0.02
DOC (+)	0.48	0.08	-0.07	0.09	-0.03	0.00
QUANT (-)	-0.31	0.08	0.03	0.02	0.02	0.01
QUANT 0	0.20	-0.08	0.04	-0.02	-0.04	-0.01
QUANT (+)	0.39	-0.04	-0.15	-0.03	0.04	0.00
PROSA (-)	-0.28	-0.05	-0.04	-0.01	-0.01	0.01
PROSA 0	0.29	-0.06	0.06	0.04	0.03	-0.01
PROSA (+)	0.47	0.38	0.03	-0.06	-0.01	0.00

Quadro 4.5 – Contribuições.

PAÍSES	CONTRIBUIÇÕES ABSOLUTAS						CONTRIBUIÇÕES RELATIVAS					
	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6
Canadá	1.4	9.0	0.7	2.2	2.8	0.9	0.57	0.40	0.01	0.01	0.01	0.00
Alemanha	1.2	2.5	0.0	6.1	5.9	4.0	0.77	0.17	0.00	0.04	0.02	0.00
Irlanda	0.4	0.7	3.4	1.5	19.3	1.4	0.56	0.10	0.14	0.2	0.15	0.00
Holanda	2.5	1.4	13.8	0.1	0.3	2.17	0.81	0.05	0.14	0.00	0.00	0.00
Polónia	11.7	0.0	0.1	3.3	1.5	1.6	1.00	0.00	0.00	0.00	0.00	0.00
Suécia	13.5	16.2	19.6	0.8	1.3	0.9	0.85	0.11	0.04	0.00	0.00	0.00
Suíça (francesa)	0.3	15.2	1.5	2.3	1.7	3.8	0.32	0.59	0.05	0.02	0.01	0.01
Suíça (alemã)	0.0	4.5	0.4	0.3	5.8	1.2	0.06	0.86	0.02	0.00	0.06	0.00
EUA	0.2	7.5	1.3	3.5	4.0	0.4	0.20	0.71	0.03	0.03	0.02	0.00
Austrália	0.6	1.3	5.2	5.7	0.2	0.4	0.63	0.14	0.17	0.06	0.00	0.00
Bélgica (Flandres)	1.0	1.0	1.9	5.8	11.8	54.6	0.75	0.08	0.04	0.04	0.05	0.08
Nova Zelândia	0.1	5.3	3.0	0.0	0.8	3.2	0.07	0.78	0.13	0.00	0.01	0.01
Reino Unido	0.0	3.7	0.4	1.6	0.4	9.6	0.01	0.88	0.03	0.04	0.01	0.04
Chile	23.6	2.4	0.6	7.4	3.8	6.7	0.98	0.01	0.00	0.00	0.00	0.00
República Checa	1.7	8.8	23.5	1.7	12.2	0.2	0.36	0.34	0.26	0.01	0.03	0.00
Dinamarca	3.6	17.6	2.3	23.2	0.5	0.0	0.62	0.33	0.01	0.04	0.00	0.00
Filândia	3.2	3.3	6.2	1.5	13.1	3.5	0.83	0.09	0.05	0.00	0.02	0.00
Hungria	4.9	4.3	10.6	8.3	12.4	2.1	0.84	0.08	0.06	0.01	0.01	0.00
Noruega	7.0	0.2	0.0	24.5	0.3	10.7	0.96	0.00	0.00	0.03	0.00	0.00
Portugal	13.9	0.7	0.2	0.1	0.6	10.4	0.99	0.01	0.00	0.00	0.00	0.00
Eslovénia	9.6	0.2	0.1	0.1	1.2	2.0	1.00	0.00	0.00	0.00	0.00	0.00
Suíça (italiana)	0.0	4.2	5.2	0.1	1.6	7.1	0.01	0.71	0.25	0.00	0.01	0.02
NÍVEL	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6	Eixo 1	Eixo 2	Eixo 3	Eixo 4	Eixo 5	Eixo 6
DOC (-)	18.2	1.2	0.4	1.1	0.0	25.8	0.99	0.01	0.00	0.00	0.00	0.00
DOC 0	7.1	7.1	9.4	11.1	5.9	35.1	0.85	0.09	0.04	0.01	0.00	0.01
DOC (+)	13.2	3.3	10.0	41.6	11.2	0.2	0.92	0.03	0.02	0.03	0.00	0.00
QUANT (-)	15.1	9.7	4.7	8.4	7.7	11.1	0.92	0.06	0.01	0.01	0.00	0.00
QUANT 0	4.6	7.7	6.5	3.0	38.4	13.7	0.78	0.14	0.03	0.01	0.04	0.00
QUANT (+)	9.8	1.2	46.2	4.6	16.5	0.0	0.85	0.01	0.13	0.00	0.01	0.00
PROSA (-)	14.3	4.8	10.2	1.2	6.2	5.7	0.94	0.03	0.02	0.00	0.00	0.00
PROSA 0	7.9	4.2	11.7	14.1	13.3	8.4	0.88	0.05	0.04	0.02	0.01	0.00
PROSA (+)	9.7	60.6	0.9	15.0	0.8	0.0	0.59	0.40	0.00	0.01	0.00	0.00

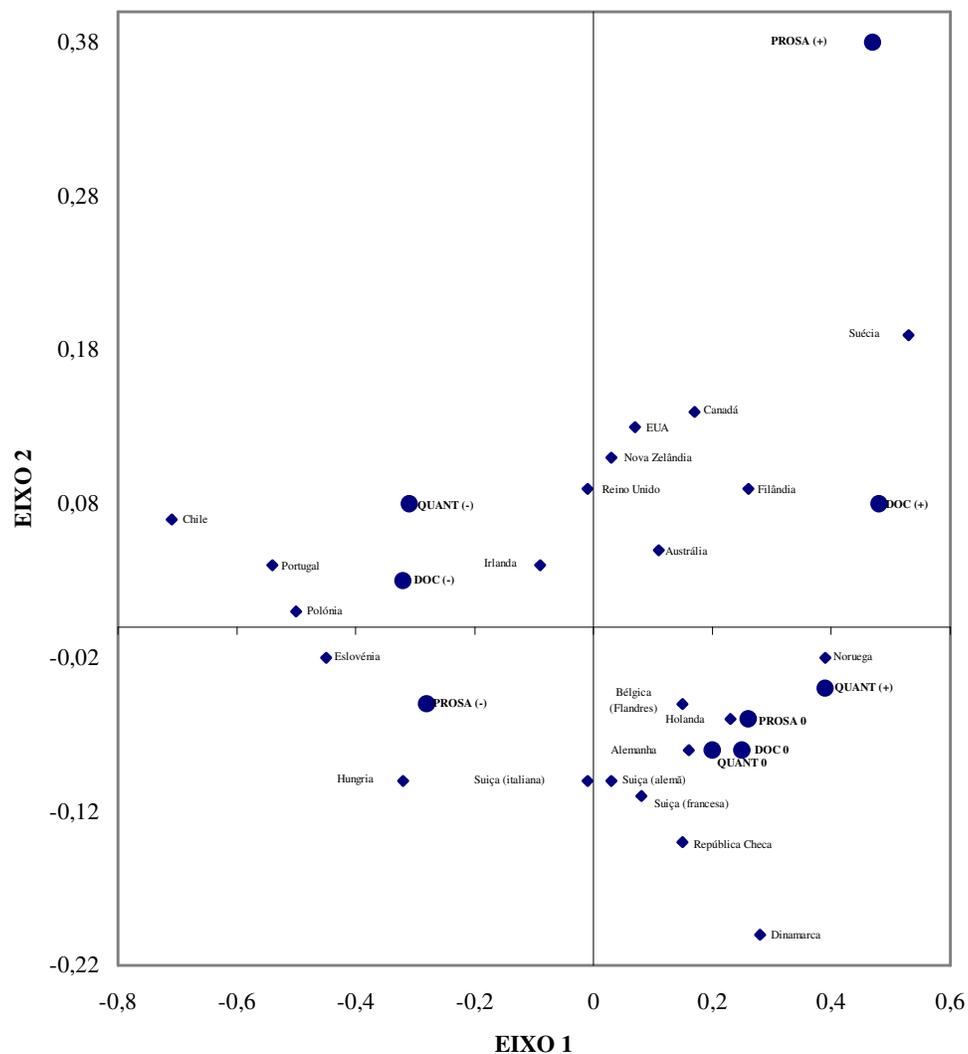


Fig. 4.7 – Projecção dos países e níveis de literacia no plano factorial 1,2.

A interpretação dos resultados é feita com base nas contribuições absolutas. Uma modalidade contribui para um eixo desde que a sua contribuição absoluta seja superior a $\frac{100}{9}$ no caso das vertentes de literacia e $\frac{100}{22}$ no caso dos países.

Assim, o eixo 1 opõe os níveis baixos das 3 vertentes de literacia (QUANT(-), DOC(-), PROSA(-)) ao nível alto da vertente documental (DOC(+)). Os países que se associam aos níveis mais baixos são (por esta ordem):

Chile , Portugal , Polónia , Eslovénia

Os países que se encontram ligados com DOC(+) são a Noruega , Suécia.

Quanto ao eixo 2, ele isola a vertente positiva da literacia textual (PROSA(+)) e o eixo 3 (vd. Fig. 4.8), a vertente da literacia quantitativa (QUANT(+)).

Os países ligados a PROSA(+) são:

Suécia , Canadá , EUA , Nova Zelândia

os quais se opõem a Suíça (francesa) , Dinamarca , República Checa

Os países ligados a QUANT(+) são:

República Checa , Hungria , Suécia

que se opõem a Holanda , Finlândia , Austrália , Suíça (italiana)

Por uma análise mais detalhada dos Quadros 4.4 e 4.5, pode dizer-se ainda que, o eixo 4 isola a Alemanha, ligada com os valores médios de literacia textual (PROSA(0)) e documental (DOC(0)).

Quanto ao eixo 5, que já só explica 0.5% da informação, este pode ser interpretado pela modalidade QUANT(0), que se opõe à Irlanda e Bélgica (vd. Quadro 4.4 e 4.5).

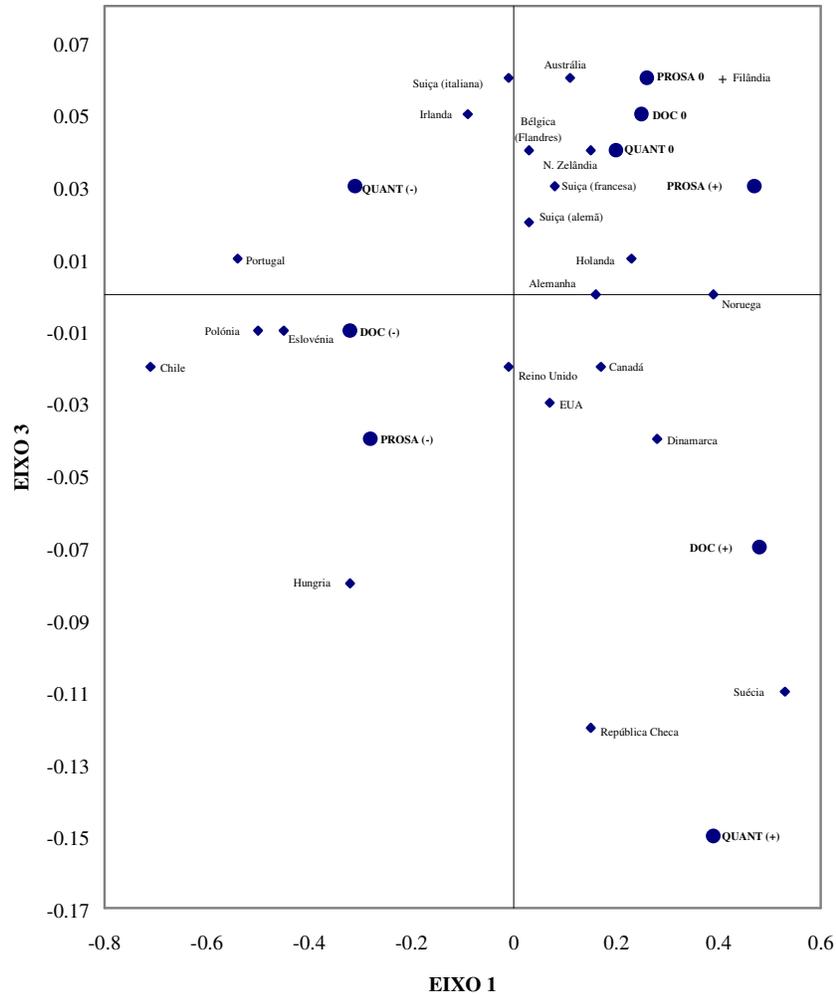


Fig. 4.8 – Projecção dos países e níveis de literacia no plano factorial 1,3.

3. Quadros de Descrição Lógica (Presença-Ausência)

Um inquérito, incidindo sobre 1939 indivíduos, foi efectuado em diferentes regiões do país com o objectivo de avaliar a relação entre o consumo de um determinado produto e a idade dos consumidores. Cada inquérito foi submetido a codificação disjuntiva completa e, após agregação de modalidades, a matriz inicial foi transformada numa matriz de BURT cruzando as regiões e as idades. A essa matriz foram adicionadas, em suplementar, as tabelas de contigência cruzando o consumo do produto com as regiões e idades (Fig. 4.9).

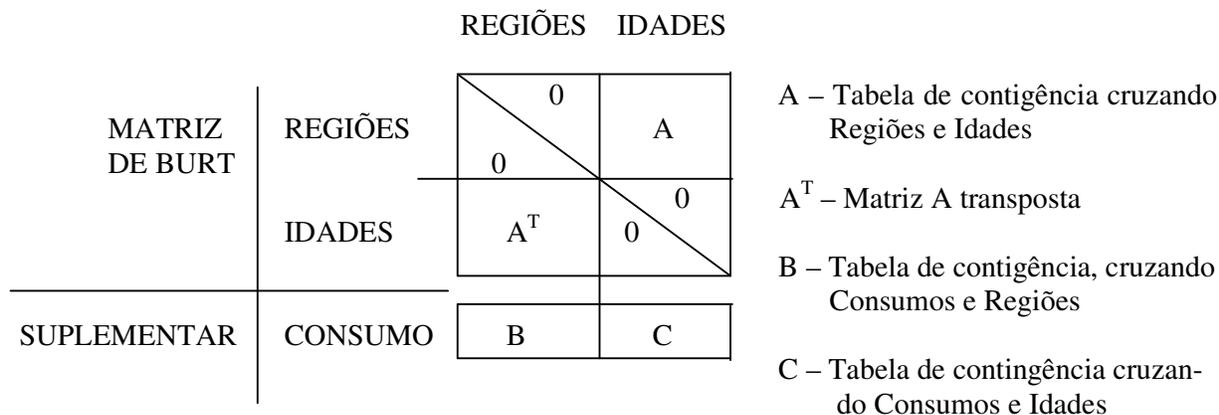


Fig. 4.9 – Modelo dos dados de partida.

A matriz de partida encontra-se no Quadro 4.6.

Quadro 4.6 – Tabela de partida segundo o modelo da Fig. 4.9.

	1	2	3	4	5
1 Lisboa	462	0	0	173	289
2 Porto	0	280	0	101	179
3 Resto do País	0	0	1197	317	880
4 - 35 anos	173	101	317	591	0
5 + 35 anos	289	179	880	0	1348
6 Consome	120	69	240	164	265
7 Não Consome	342	211	957	427	1083

A percentagem de inércia explicada pelos 2 primeiros eixos, encontra-se no Quadro 4.7.

Quadro 4.7 – Inércia explicada pelos 2 primeiros eixos.

EIXO 1	41%
EIXO 2	33%
	74%

Representando as projecções no plano 1,2, obtém-se a Fig. 4.10, da qual se pode concluir que o consumo do produto está ligado aos jovens (- 35 anos) e às regiões urbanas (Lisboa e Porto). É de notar que o eixo 2 separa Lisboa/Porto mas não é influenciado pelo consumo.

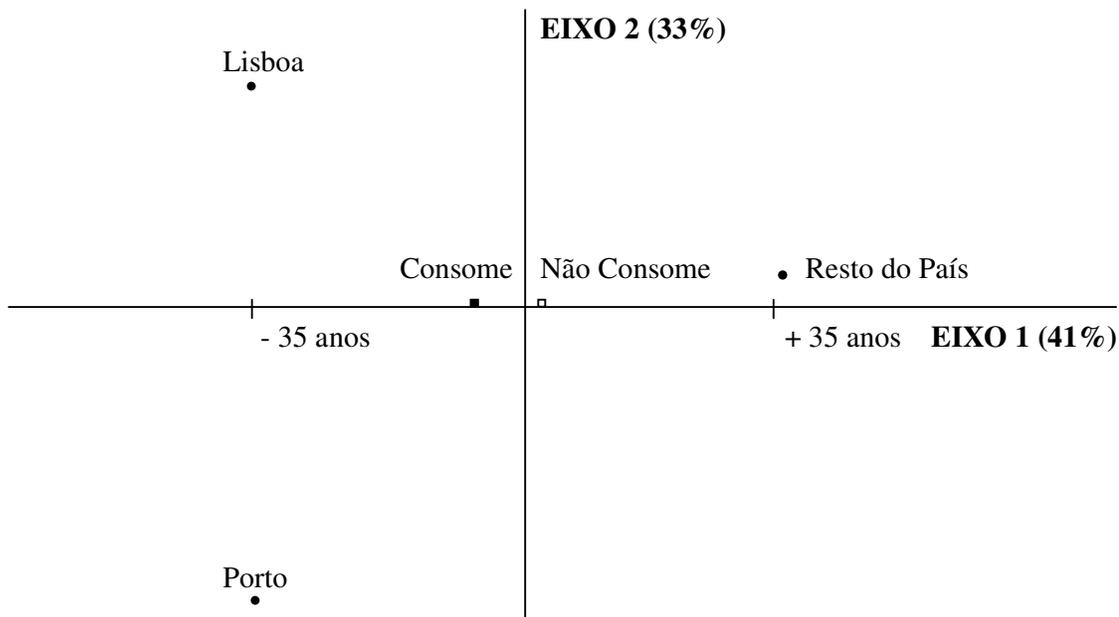


Fig. 4.10 – Projecção no plano 1,2 da matriz de BURT com o consumo em suplementar.