

Avaliação de Riscos, Segurança e Fiabilidade



C. Guedes Soares (Editor)

ÍNDICE

Seminários do Curso Avançado de Avaliação de Riscos, Segurança e Fiabilidade

1. Crise e Gestão de Risco, *José A. Allen Lima* 11
2. Risco Hidrológicos Associado à Produção de Energia em Pequenas Centrais Hidroeléctricas. Casos de Estudo, *Maria Manuela Portela* 27
3. Alta Tensão nas Linhas: O Estranho Caso do Risco das Linhas de Muito Alta Tensão, *Sibila Marques e Maria Luísa Lima* 59
4. Risco e Percepção do Risco: Uma Aproximação Preliminar Unificadora, *Henrique Garcia Pereira e Ana Rita Salgueiro* 85
5. O Risco Industrial – Alguns Conceitos Práticos para a sua Gestão, *João Manuel Coelho* 103
6. Segurança de processos Químicos Industriais, *Maria Cristina Fernandes* 125
7. Selecção de Medidas de Segurança Contra Incêndio: Uma Metodologia, *João Ventura* 145
8. Definindo Estratégias de Manutenção, *António Costa Gonçalves* 157
9. O Planeamento Civil de Emergência no Contexto da Segurança Nacional, *João Vasconcelos Piroto e Isabel Pais* 189

RISCO E PERCEÇÃO DO RISCO: UMA APROXIMAÇÃO PRELIMINAR UNIFICADORA

Henrique Garcia Pereira¹ e Ana Rita Salgueiro²

¹ CERENA, Instituto Superior Técnico,
Lisboa, Portugal
henrique.pereira@ist.utl.pt

² GeoBioTec, Universidade de Aveiro,
Aveiro, Portugal
rita.salgueiro@ist.utl.pt

RESUMO

Propõe-se uma metodologia para tratar conjuntamente o risco objectivo e subjectivo, numa fase preliminar de análise. Essa metodologia apoia-se num tratamento estatístico multidimensional, válido para os dois tipos de risco: na sua estimação, parte-se de um 'histórico', e na sua percepção, de um conjunto de inquéritos por questionário. Com o objectivo de ilustrar a metodologia proposta, centrada na Análise das Correspondências, apresenta-se um estudo de caso, referente ao risco objectivo e subjectivo de derramamento catastrófico de contaminantes contidos numa barragem de estéreis da mina da Panasqueira, onde se explora W e Sn desde o início do século XX. Discutem-se ainda as vantagens, tanto na estimação como na percepção do risco, de utilizar exclusivamente variáveis qualitativas, na construção dos quadros de partida que são o input para a metodologia. Essas vantagens decorrem não só de um melhor aproveitamento de toda a informação disponível – seja qual for a forma em que ela é inicialmente recolhida –, como da unificação dos resultados referentes ao risco objectivo e subjectivo, que podem exprimir-se numa métrica comum.

1 Introdução

A aproximação quantitativa às características da sociedade do risco – tal como foi conceptualizada por Beck (1992) – pode assentar em duas abordagens paralelas, que em geral são desenvolvidas independentemente: a engenharia tende a privilegiar os aspectos objectivos do risco, e a sociologia,

os seus aspectos subjectivos, ligados à percepção do risco (na acepção de Douglas, 1992). Quanto à primeira abordagem, desde que não se disponha de dados fiáveis e detalhados para calcular uma probabilidade de ocorrência de um episódio catastrófico e respectivos danos associados, a única aproximação preliminar possível é ensaiar um método estatístico, baseado em analogias com um 'histórico'; na segunda abordagem, é sempre necessário recorrer a métodos estatísticos para o tratamento de questionários (ou para sintetizar os resultados obtidos a partir de um painel de peritos). Tanto numa abordagem como noutra, o tratamento estatístico sai enriquecido se forem introduzidas variáveis qualitativas capazes de dar conta de aspectos do problema que não podem ser expressos em números reais (resultantes de medidas para os dados objectivos e de médias de pontuações para os dados subjectivos). O método estatístico que permite tratar variáveis qualitativas em ambas as abordagens – abrindo assim o caminho para uma aproximação preliminar unificadora que é válida tanto para o risco objectivo e subjectivo (e que pode levar a resultados integrados) – é a Análise das Correspondências, tal como formulada pela primeira vez em Benzécri (1973) e retomada – do ponto de vista teórico e prático – em Pereira e Sousa (2005).

2 Riscos objectivos e subjectivos

2.1 *Estimação do Risco*

A noção de 'Risco' tem diferentes conotações, dependentes do contexto e do campo de aplicação. A aproximação preliminar aqui apresentada centra-se no domínio dos riscos ambientais associados a acidentes em sistemas industriais que possam ter consequências na saúde humana.

Neste contexto, há sempre duas vertentes do risco: a vertente objectiva, dependente de factores físicos, geológicos ou meteorológicos, e a vertente subjectiva, dependente da percepção humana dos danos associados à possibilidade de ocorrência de uma catástrofe, resultante de falhas tecnológicas.

Quanto ao risco objectivo R , o seu cálculo é feito por:

$$R = p \cdot I \quad (1)$$

Onde p é a probabilidade de ocorrência do desastre e I é uma medida do impacto dos danos associados a essa ocorrência.

Na verdade, o cálculo do risco por (1) apresenta dificuldades de várias ordens. Em primeiro lugar, é necessário estabelecer uma probabilidade p , dependente do conhecimento de uma função densidade de probabilidade que modele a frequência dos acontecimentos ‘catastróficos’. Ora, tais acontecimentos têm uma característica que impede a utilização da estatística clássica (paramétrica e monovariada): são o resultado de fenómenos raros (*outliers*), ligados por sua vez a um feixe complexo de factores, difíceis de individualizar e relacionar. Assim, é pouco fiável o cálculo da probabilidade de ocorrência deste tipo de acontecimentos (excepcionais, globais e singulares), que podem surgir repentinamente em situações complexas e interdependentes, onde qualquer tipo de inferência se torna impraticável pela ausência de regularidades estatísticas observáveis.

Quanto ao outro termo da equação (1) – a medida I do impacto dos danos associados à ocorrência catastrófica –, também a sua estimação apresenta obstáculos difíceis de superar, visto que depende conjuntamente de (pelo menos) dois factores: por um lado, da intensidade do acontecimento ‘catastrófico’, e por outro, da vulnerabilidade do meio que terá de ‘absorver’ esse acontecimento (o que implica o cálculo de uma *dose-response function*). Acresce a esta dificuldade a exigência de que todas as componentes de I sejam expressas nas mesmas unidades (em geral, monetárias).

Em síntese, o risco objectivo é um conceito constituído por atributos múltiplos, ligado a uma incerteza difícil de quantificar. Assim, não é de admirar que o *output* final dos métodos canónicos de estimação do risco objectivo seja um diagrama simplificado como o da Figura 1, onde o valor do risco é expresso apenas por três categorias, resultantes de combinações de intervalos arbitrários nos dois termos da equação (1).

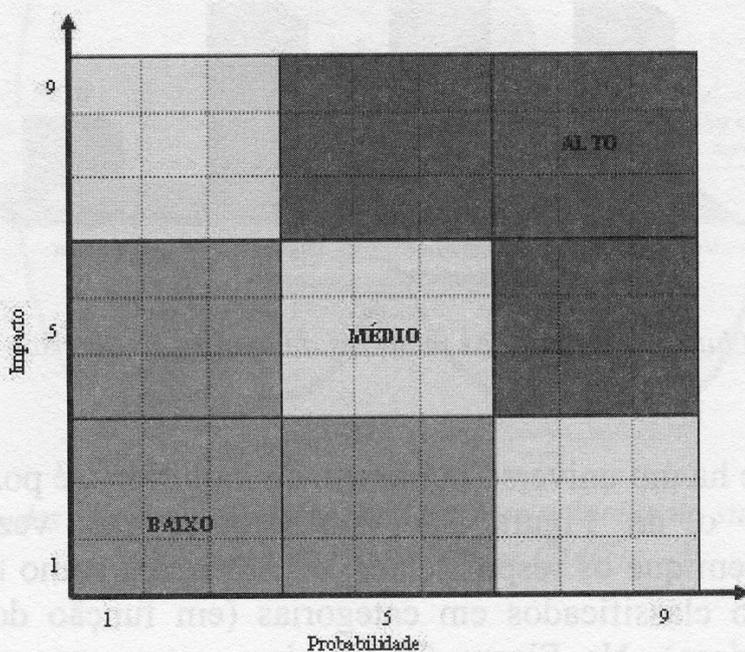


Figura 1 – Output habitual dos métodos de estimação do risco objectivo.

2.2 Percepção do Risco

Através de um sistema de pontuações para diferentes aspectos subjectivos ligados à possibilidade de ocorrência de uma catástrofe (e sua intensidade), pode-se calcular o risco com base em inquéritos por questionário, incidindo sobre a percepção do risco por parte de uma amostra significativa de respondentes.

Essa percepção do risco depende de uma miríade de factores difíceis de sistematizar *à priori*. Tais factores são por vezes contraditórios e de interpretação pouco clara (por exemplo, há estudos que indicam que, paradoxalmente, as pessoas que vivem perto de uma barragem consideram que há menos probabilidade de ela rebentar do que as que vivem longe). Hilson e Murray-Webster (2007) discutem em profundidade a questão da atitude perante o risco, e a sua dependência de factores tecnológicos (concluindo, por exemplo, que o nível de percepção do risco diminui com a intensidade do desenvolvimento tecnológico, mas aumenta com a sua taxa de crescimento).

Apesar desta complexidade, acontece porém que o *output* da grande maioria dos estudos sobre a percepção do risco se resume em geral a histogramas do tipo do da Figura 2, onde surge simplesmente a frequência de respondentes para cada nível de risco (o que revela uma abordagem severamente reducionista).

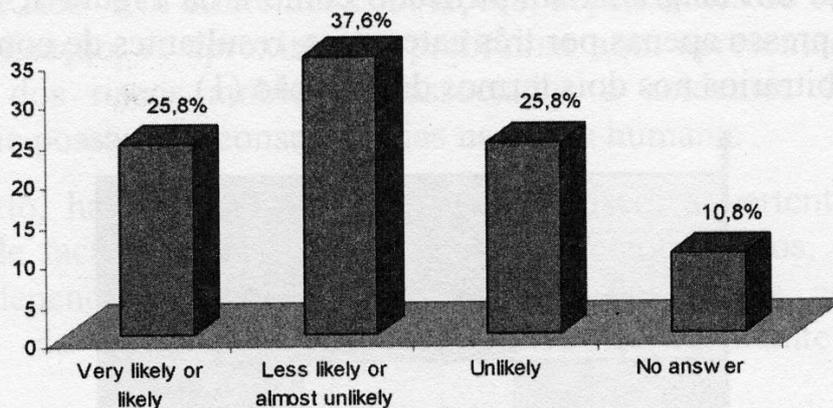


Figura 2 – Output habitual dos métodos de avaliação do risco subjectivo.

No caso em que há um universo numeroso de inquiridos, é possível construir gráficos como o da Figura 3A (sintéticos e por vezes demasiado esquemáticos), em que os respondentes, considerados como uma população homogénea, são classificados em categorias (em função do seu nível de percepção do risco). Na Figura 3B ensaia-se uma conceptualização das

consequências para a gestão do risco decorrentes das categorias básicas estabelecidas pelo inquérito.

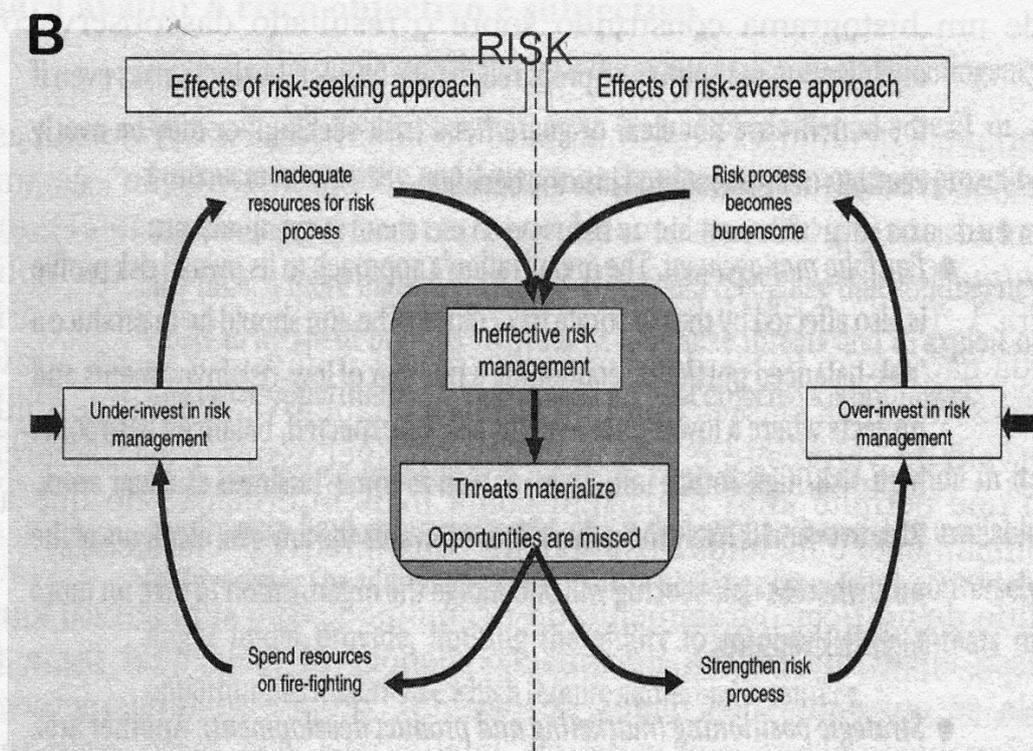
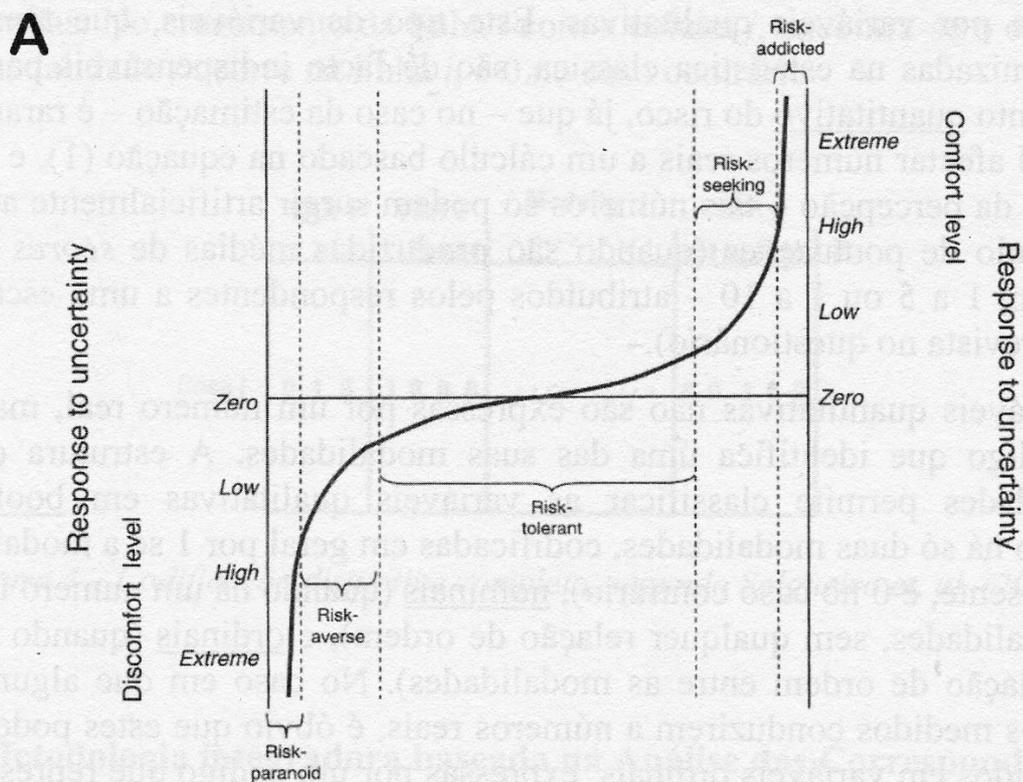


Figura 3 – Atitudes básicas perante o risco e consequências na sua gestão, segundo Hilson e Murray Webster (2007).

3 Importância das variáveis qualitativas no tratamento do risco

Os modelos que permitem abordar a questão do risco, seja ele objectivo ou subjectivo, têm forçosamente de incorporar informação que só pode ser expressa por variáveis qualitativas. Este tipo de variáveis, que têm sido subalternizadas na estatística clássica, são de facto indispensáveis para um tratamento quantitativo do risco, já que – no caso da estimação – é raramente possível afectar números reais a um cálculo baseado na equação (1), e que – no caso da percepção – tais números só podem surgir artificialmente através do cálculo de pontuações (quando são produzidas médias de *scores* – por exemplo, 1 a 5 ou 1 a 10 – atribuídos pelos respondentes a uma escala de valor prevista no questionário).

As variáveis quantitativas não são expressas por um número real, mas por um código que identifica uma das suas modalidades. A estrutura dessas modalidades permite classificar as variáveis qualitativas em booleanas (quando há só duas modalidades, codificadas em geral por 1 se a modalidade está presente, e 0 no caso contrário), nominais (quando há um número inteiro de modalidades, sem qualquer relação de ordem), e ordinais (quando existe uma relação de ordem entre as modalidades). No caso em que alguns dos atributos medidos conduzirem a números reais, é óbvio que estes podem ser convertidos em variáveis ordinais, expressas por um código que representa a classe de um histograma construído sobre o resultado das observações de partida.

Se as variáveis qualitativas estão na base dos questionários onde se baseia o risco subjectivo, já no caso do risco objectivo há a ‘tentação’ de imputar valores reais aos algoritmos necessários para resolver a equação (1), mesmo que o *output* final seja expresso em variáveis ordinais, como se exemplificou na Figura 1. Este tipo de *output* decorre do facto de a maioria das variáveis de partida não serem na verdade números reais (obtidos por vezes através de *expert guesses*), mas classes de uma variável qualitativa ordinal definida por um intervalo de valores (e é sem dúvida preferível assumir esta asserção *ab initio*, o que permite alias tomar em conta mais informação do que aquela que habitualmente é incluída nos modelos ‘quantitativos’ clássicos).

No sentido de unificar o tratamento do risco objectivo e subjectivo, tomando além disso em conta toda a informação disponível (seja qual for a forma como ela se apresenta nos dados de partida), propõe-se uma metodologia baseada exclusivamente em variáveis qualitativas. De facto, este tipo de variáveis é o mais adequado para o tratamento de fenómenos complexos e interdependentes, como é o caso dos diferentes contextos onde o conceito de Risco pode ser analisado.

Apresenta-se na Figura 4 o modelo mais comum para o quadro de partida contendo só variáveis qualitativas, designado por “codificação em disjuntiva completa”. Em linha estão as n unidades estatísticas (indivíduos) onde as q variáveis (desdobradas num número qualquer de modalidades) foram observadas. No cruzamento da linha i com a coluna j , encontra-se o código 1 se a modalidade j ocorre na linha i , e 0 no caso contrário.

	VAR 1			VAR 2				Modality j	VAR q					P
	1	2	3	1	2	3	4	1	2	3	4	5	
Case i	0	1	0	1	0	0	0	0	0	1	0	0	
n														

Figura 4 – Codificação disjuntiva completa, segundo Salgueiro et. al. (2008).

4 Metodologia integradora baseada na Análise das Correspondências para avaliar o risco objectivo e subjectivo

Tanto o Risco objectivo como o subjectivo, desde que as variáveis de partida sejam codificadas segundo o modelo da Figura 4, podem ser tratados com base numa metodologia estatística designada por Análise das Correspondências (ver ANEXO).

De facto, no caso da estimação do risco referente a casos ‘novos’ a partir de um ‘histórico’ contendo n casos ‘análogos’, é possível estabelecer uma sequência de procedimentos apoiados sempre na Análise das Correspondências, tal como está sintetizado na Figura 5. Esta sequência parte de três matrizes em disjuntiva completa: as matrizes A e B referem-se ao ‘histórico’ e C aos casos ‘novos’. A Matriz A contém as variáveis observáveis antes da ocorrência da catástrofe (e que podem estar relacionadas com ela), B as variáveis que dão a dimensão das consequências da catástrofe, e C as variáveis observáveis nos casos novos (onde se pretende estimar o risco). Em vez de aplicar a Análise das Correspondências com objectivos puramente descritivos, a metodologia proposta tira partido da ‘ projecção em suplementar’ para efectuar uma regressão sobre variáveis qualitativas. A projecção em suplementar consiste no posicionamento de variáveis ou indivíduos contidos nas matrizes B e C no espaço factorial resultante da diagonalização da matriz de inércia construída a partir de A, como é exposto detalhadamente em Salgueiro et al. (2008, p. 16).

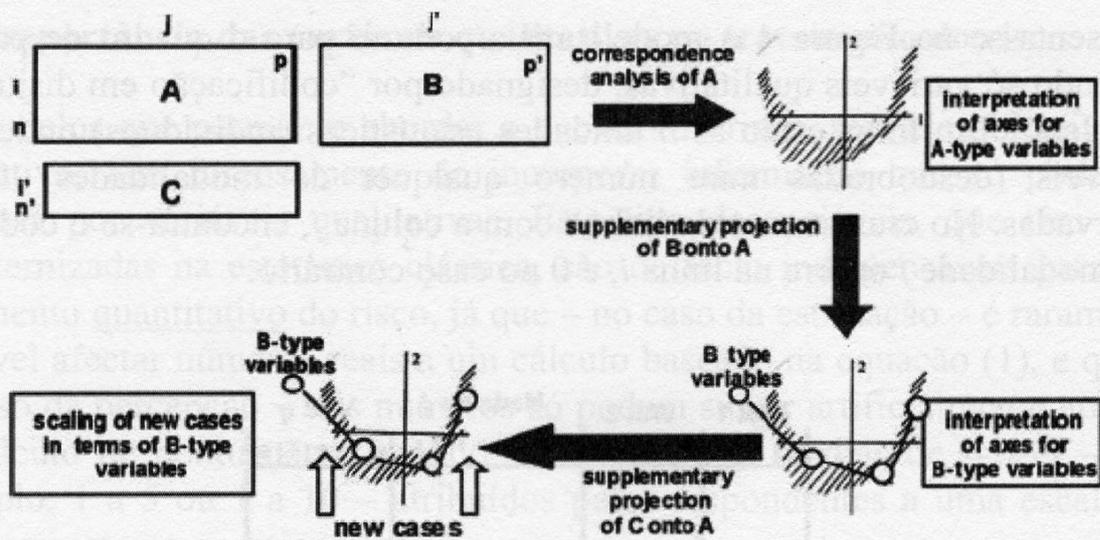


Figura 5 – Esquema da sequência de procedimentos para estimar o risco, segundo Salgueiro et. al. (2008).

Quanto à avaliação do risco subjectivo com base em inquéritos por questionários, é possível também tirar partido da projecção em suplementar em Análise das Correspondências para relacionar a atitude perante o risco com as características sócio-económicas dos respondentes, de acordo com a metodologia descrita em Pereira (1987). Se considerarmos agora a matriz A como descritora das condições que caracterizam os indivíduos sujeitos ao inquérito e B como reveladora da sua atitude em face do risco, a metodologia é, *mutatis metandi*, a que está esquematizada na Figura 5 (podendo ou não existir um conjunto de indivíduos C, cuja atitude em face do risco seja desconhecida).

É então possível conjugar o risco objectivo com o subjectivo através da mesma metodologia, baseada na Análise das Correspondências e sintetizada na Figura 5.

5 Estudo de Caso

No âmbito de um projecto financiado pela UE (e-EcoRisk – A Regional Enterprise Network Decision-Support System for Environmental Risk and Disaster Management of Large-Scale Industrial Spills), a metodologia atrás descrita foi aplicada – entre outros casos na região mediterrânica – à Mina da Panasqueira. Os resultados dessa aplicação específica são resumidos em seguida.

No que diz respeito ao risco objectivo, a situação das barragens de estêreis decorrentes do tratamento do minério explorado na Panasqueira desde o

início do século XX foi caracterizada a partir de um sistema de atributos que foi possível extrair dos registos de um conjunto de desastres ocorridos no passado, reportados sob a forma de uma base de dados em Rico et. al.(2007). Essa Base de Dados Histórica contém 147 casos, cujas variáveis foram estruturadas segundo o modelo da Figura 4, dando origem a matrizes do tipo das que foram denotadas A e B na Figura 5. As variáveis da Matriz A (referentes às observações pré-rotura) são apresentadas na Figura 6, bem como a sua projecção no plano factorial que cruza os factores F1 e F2 resultantes da aplicação da Análise das Correspondências à referida matriz (primeiro passo da metodologia apresentada esquematicamente na Figura 5). Projectando em suplementar as variáveis referentes às diferentes componentes do risco (segundo passo da Figura 5 para a Matriz B), obtém-se a Figura 7, onde é patente que o Risco Baixo se projecta na parte negativa do semi-eixo F1 (*mutatis mutandi* para o Risco Elevado).

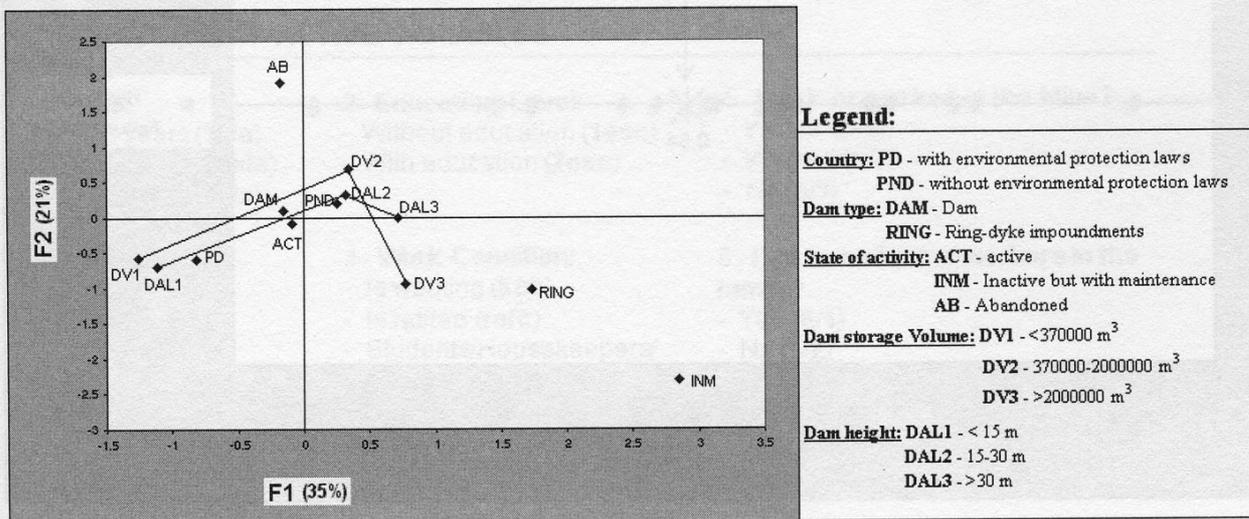


Figura 6 – Primeiro plano factorial da Análise de Correspondências da matriz referente às variáveis pré-rotura contidas na Base de Dados Histórica.

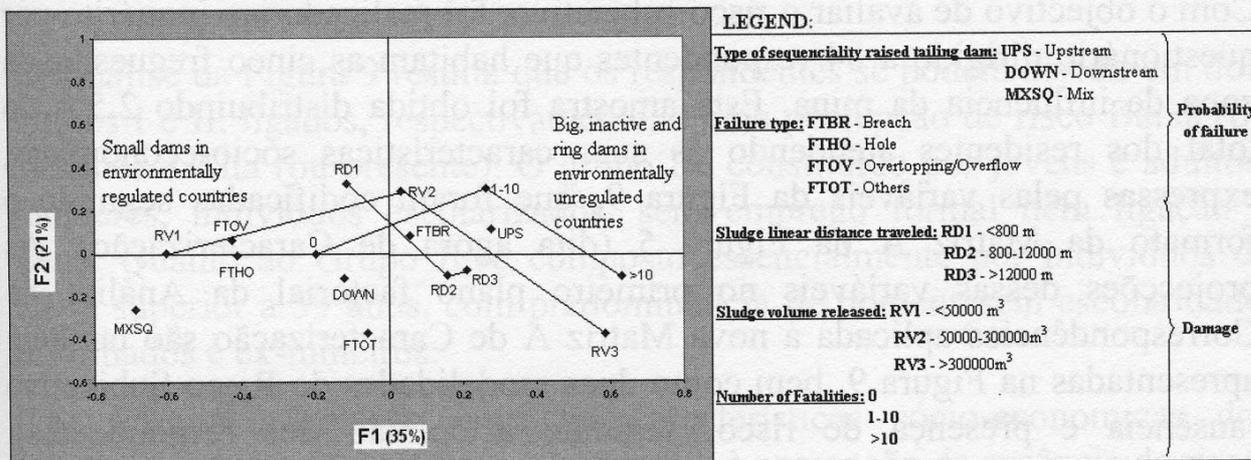


Figura 7 – Projecção em suplementar das variáveis associadas ao risco sobre os eixos factoriais obtidos pela Análise das Correspondências da Matriz A.

Passando então ao terceiro passo da metodologia esquematizada na Figura 5 (e assimilando a Matriz C a um vector contendo as variáveis *ex-ante* referentes à Panasqueira que coincidem com as que estão presentes na Base de Dados Histórica), obtém-se a Figura 8 por projecção em suplementar do referido vector no eixo F1 das Figuras 6 e 7. Na Figura 8, o eixo F1, que expressa quantitativamente o risco objectivo segundo a metodologia proposta, foi standardizado linearmente no intervalo [0,1], para mais fácil comparação do caso da Panasqueira (cujo risco é 0.54), com os análogos da Base de Dados Histórica. Esta standardização permite ainda ensaiar uma equivalência entre a estimação e a percepção do risco.

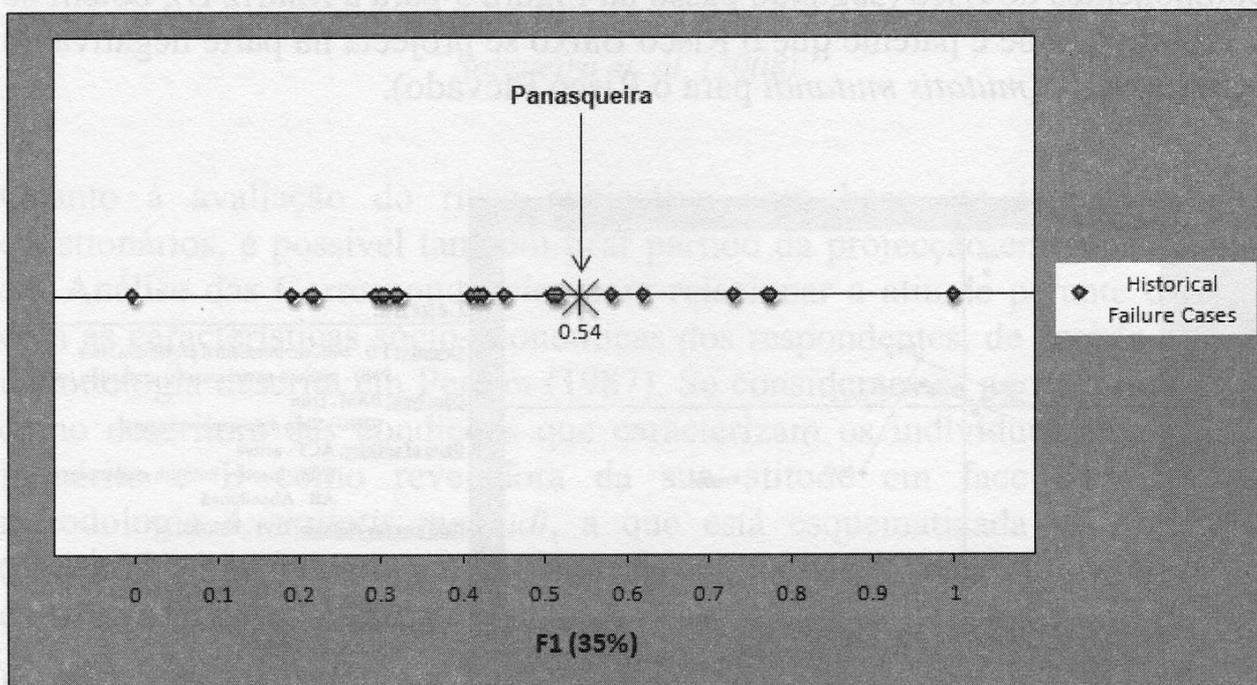
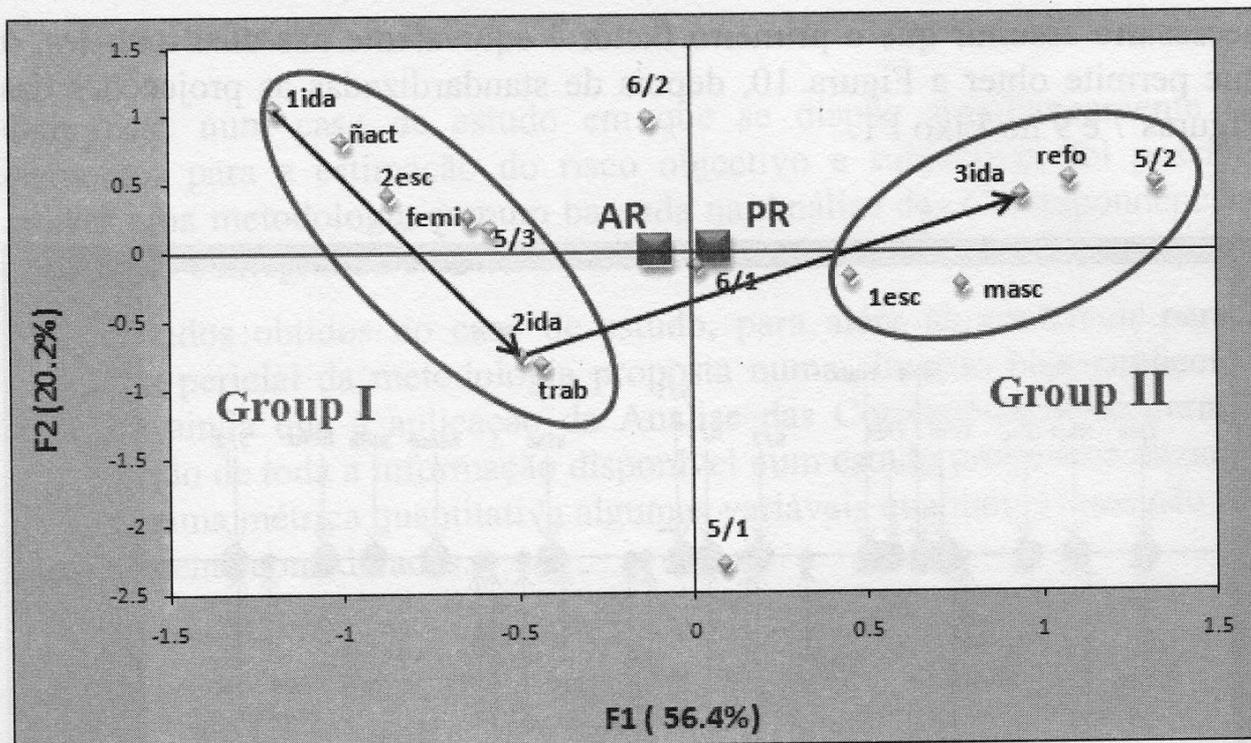


Figura 8 – Posicionamento da Panasqueira numa escala standard de risco objectivo, relativamente aos casos históricos.

Com o objectivo de avaliar o risco subjectivo, foi realizado um inquérito por questionário dirigido a 84 respondentes que habitam as cinco freguesias da zona de influência da mina. Esta amostra foi obtida distribuindo 2.5% do total dos residentes atendendo às suas características sócio-económicas, expressas pelas variáveis da Figura 9, que foram codificadas segundo o formato da Matriz A da Figura 5 (dita agora de Caracterização). As projecções dessas variáveis no primeiro plano factorial da Análise de Correspondências aplicada à nova Matriz A de Caracterização são também apresentadas na Figura 9, bem como duas modalidades do Risco Subjectivo (ausência e presença de risco, segundo a opinião dos respondentes), projectadas em suplementar (e fazendo assim as vezes da Matriz B da Figura 5).



MODALITIES OF ACTIVE VARIABLES

- | | | |
|---|--|---|
| 1 - Age:
- < 30 years (1ida)
- 30-59 years (2ida)
- > 59 years (3ida) | 3 - Education Level:
- Without education (1esc)
- With education (2esc) | 5 - Work or worked at the Mine?
- Works (5/1)
- Worked (5/2)
- No (5/3) |
| 2 - Gender:
- Female (femi)
- Male (masc) | 4 - Work Condition:
- Is working (trab)
- Is retired (refo)
- Students/Housekeepers/
/Unemployed (ñact) | 6 - Has or had mine workers in the family?
- Yes (6/1)
- No (6/2) |

MODALITIES OF SUPPLEMENTARY VARIABLES

AR – Absence of Risk
 PR – Presence of Risk

Figura 9 – Projecção em Suplementar do Risco Subjectivo sobre os factores resultantes da Análise das Correspondências da Matriz de Caracterização.

Da análise da Figura 9 resulta que os respondentes se podem dividir em dois grupos I e II, ligados, respectivamente, a uma percepção de risco Baixa (ou ausente) e Alta (ou presente). O grupo I é constituído por jovens e adultos, mulheres, indivíduos escolarizados, sem emprego formal nem ligação à mina. Quanto ao Grupo II, é composto essencialmente por indivíduos de idade superior a 59 anos, com predominância de homens sem escolaridade, reformados e ex-mineiros.

Para ensaiar a relação entre as características sócio-económicas dos respondentes (em que o Grupo I se associa à percepção de ausência de risco, e o Grupo II, à percepção de presença, segundo a interpretação da Figura 9) e as diferentes componentes do risco objectivo (dadas na Figura 7), é

necessário assumir que o primeiro factor é equivalente nas duas análises, o que permite obter a Figura 10, depois de standardizadas as projecções das Figuras 7 e 9 no Eixo F1.

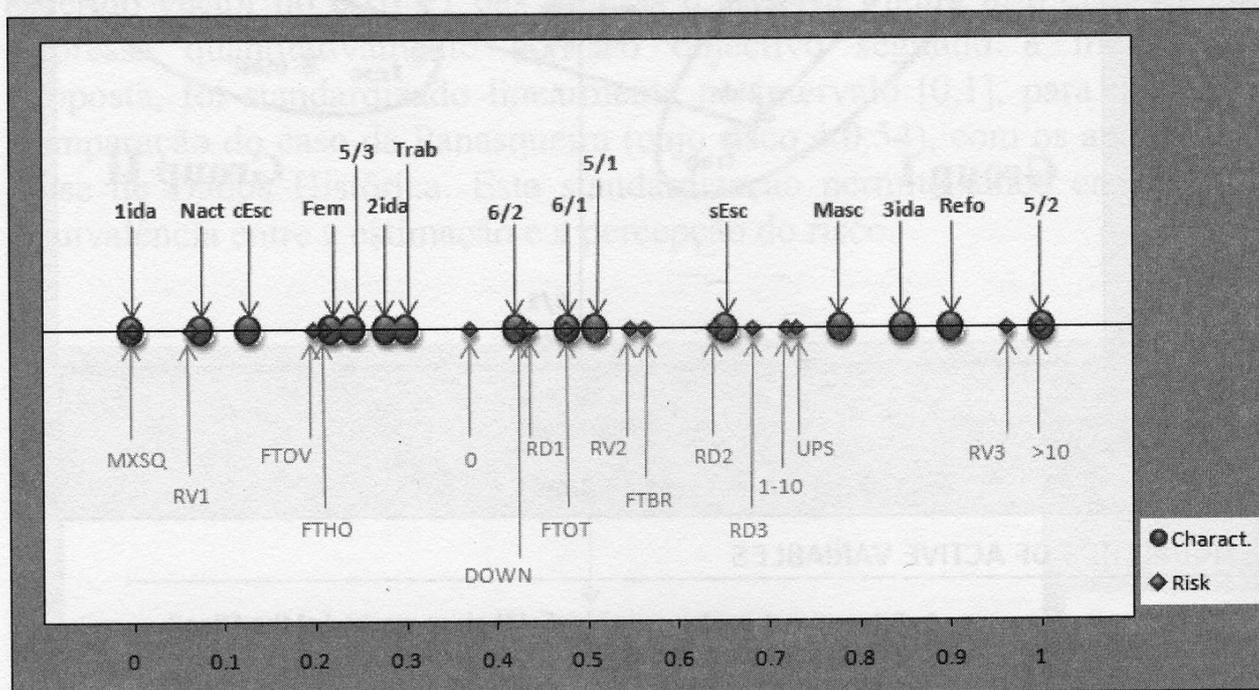


Figura 10 – Relação entre as características sócio-económicas dos respondentes ao inquérito sobre a percepção do risco e os diversos elementos do risco objectivo estimados através da Base de Dados Histórica (legenda nas Figuras 7 e 9).

Da análise da Figura 10 pode concluir-se que o Grupo I tem percepção de um Risco Baixo (<0.4 , na escala de 0 a 1), no que diz respeito às seguintes componentes desse risco: Tipo misto de construção da barragem (MSXQ), Quantidades vertidas inferiores a 5000 m^3 (RV1), Tipo de derramamento por transposição (FTOV) e por ruptura (FTHO), Ausência de consequências mortais (0). Quanto ao grupo II, este tem percepção de um Risco Elevado (>0.6 , na escala de 0 a 1) relativamente às seguintes componentes desse risco: Presença de consequências mortais (>10 e 1-10), Quantidades vertidas superiores a 5000 m^3 (RV2 e RV3); Construção da barragem do tipo ‘a montante’ (UPS), Distâncias lineares atingidas pelo derramamento superiores a 800 m (RD2 e RD3). As outras componentes do risco objectivo não são percebidas como significativas por nenhum dos grupos. De facto, essas componentes projectam-se entre 0.4 e 0.6, o que sugere que o caso da Panasqueira (vd. Figura 8) se encontra numa situação em que qualquer dos grupos é relativamente indiferente quanto à percepção das componentes do risco objectivo aqui consideradas (o que é corroborado também pela projecção das modalidades do risco subjectivo, que estão próximas da origem na Figura 9).

6 Conclusões

Com base num caso de estudo em que se dispõe simultaneamente de elementos para a estimação do risco objectivo e subjectivo, foi possível aplicar uma metodologia comum baseada na Análise das Correspondências, unificando e integrando o tratamento estatístico dos dois tipos de risco.

Os resultados obtidos no caso de estudo, para além de contribuir para a validação pericial da metodologia proposta numa situação bem conhecida, mostram ainda que a aplicação da Análise das Correspondências permite tirar partido de toda a informação disponível num estudo preliminar, fazendo intervir numa métrica quantitativa algumas variáveis qualitativas que não são habitualmente consideradas.

Referências

- Beck (1992), "Risk Society. Towards a new modernity", Sage
- Benzécri, J.-P (1973), "L'Analyse des Données", 2 volumes, Dunod
- Douglas, M. (1992), "Risk and blame. Essays in cultural theory", Routledge.
- Hilson, D., Murray-Webster, R. (2007), "Understanding and Managing Risk Attitude", Gower.
- Pereira, H.G. (1987), "Tratamento Informático de questionários: o ponto de vista da análise factorial das correspondências", *Análise Social*, vol. XXIII, nº 98, pp. 733-746.
- Pereira, H.G., Sousa, A.J. (2005), "Análise de Dados para o tratamento de quadros multidimensionais". Disponível em:
<http://biomonitor.ist.utl.pt/~ajsousa/AnalDadosTratQuadMult.html>
- Salgueiro, A.R.; Rico, M.; Pereira, H.G.; Benito, G. & Díez-Herrero, A. (2008), "Application of Correspondence Analysis to the Assessment of Mine Tailings Dam Breakage Risk in the Mediterranean Region". *Risk Analysis*, Vol. 28(1), pp.13-23
- Rico, M.; Benito, G.; Salgueiro, A.R.; Díez-Herrero, A. & Pereira, H.G. (2008), "Reported tailings dam failures. A review of the European incidents in the worldwide context". *Journal of Hazardous Materials*, nº 152, pp. 846-852.

ANEXO

SÍNTESE DO ALGORITMO DA ANÁLISE DAS CORRESPONDÊNCIAS

Seja D uma matriz em disjuntiva completa como a da Figura 4, onde estão contidas todas as variáveis qualitativas a submeter à Análise das Correspondências. Multiplicando à esquerda a transposta de D por D , é possível obter uma nova matriz quadrada e simétrica de dimensão $p \times p$, designada por matriz de Burt, que contém, na diagonal principal, os histogramas (em frequências absolutas) de cada variável qualitativa considerada, e cujos blocos não diagonais são as tabelas de contingência que cruzam todos os pares de variáveis, de acordo com o modelo da Figura A-1.

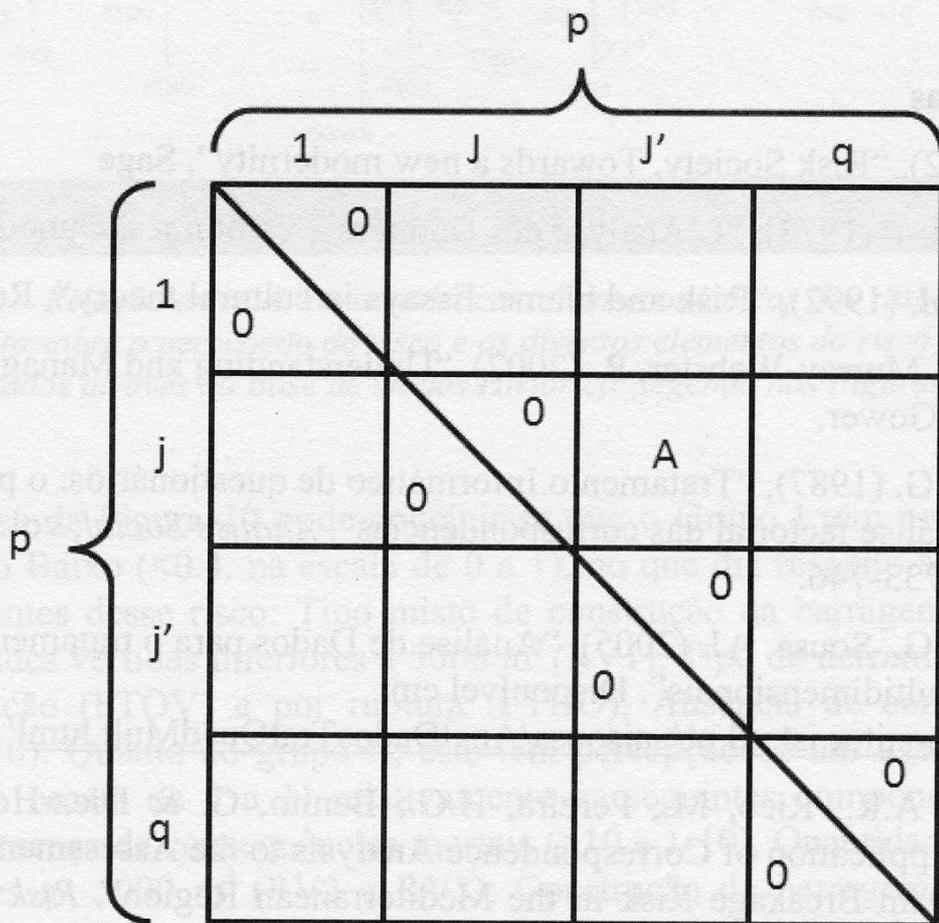


Figura A-1 Matriz de Burt.

Cada tabela de contingência contém os valores absolutos $K(i,j)$ da frequência de co-ocorrências das variáveis i e j , como se mostra na Figura A-2.

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

Figura A-2 Tabela de contingência.

É sobre o modelo da Figura A-2 que o algoritmo da Análise das Correspondências foi desenvolvido (e programado no package Andad, disponível em <http://biomonitor.ist.utl.pt/~ajsousa/Andad.html>). Note-se que tanto a matriz em disjuntiva completa como a de Burt podem ser vistas como uma justaposição de tabelas de contingência.

O primeiro passo do algoritmo consiste em calcular os pesos das linhas e colunas, de acordo com o procedimento descrito na Figura A-3. Esses pesos dão a importância relativa das modalidades das variáveis qualitativas consideradas, vistas como uma nuvem de pontos cujo Centro de Gravidade é também considerado na mesma figura.

ALGORITMO DA AC - pesos e centro de gravidade

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

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

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 $N(I)$:

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

\downarrow \swarrow
 pesos coordenadas

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

Figura A-3 Pesos e Centro de Gravidade em Análise das Correspondências.

Os passos seguintes do algoritmo são dados nas Figura A-4 a A-9.

ALGORITMO DA AC - 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:

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

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

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

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

Então, para tabalhar com a distância euclideana, é necessário fazer a transformação de coordenadas

$$\begin{array}{ccc} \frac{f_{ij}}{f_i} & \curvearrowright & \frac{f_{ij}}{\sqrt{f_j f_i}} & R^p \\ \frac{f_{i'j}}{f_{i'}} & \curvearrowright & \frac{f_{i'j}}{\sqrt{f_j f_{i'}}} & R^p \end{array}$$

Figura A-4 Cálculo da distância em Análise das Correspondências.

Algoritmo da AC - 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:

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

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

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

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

Figura A-5 Cálculo da inércia em Análise das Correspondências.

ALGORITMO DA AC - Diagonalização

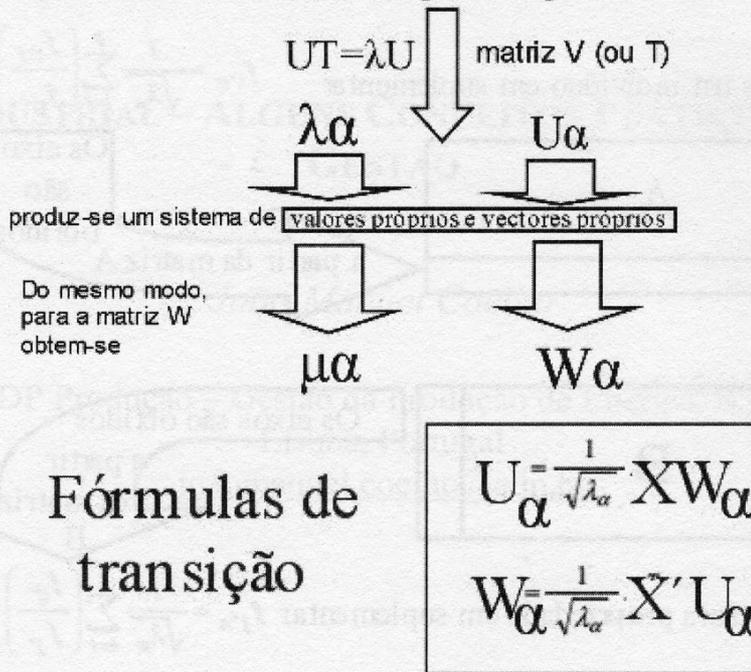


Figura A-6 Diagonalização da Matriz de Inércia em Análise das Correspondências.

ALGORITMO DA AC - projecção de indivíduos e propriedades

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

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

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:

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

Figura A-7 Projecção das linhas e colunas da matriz de dados.

ALGORITMO DA AC - projecção em suplementar

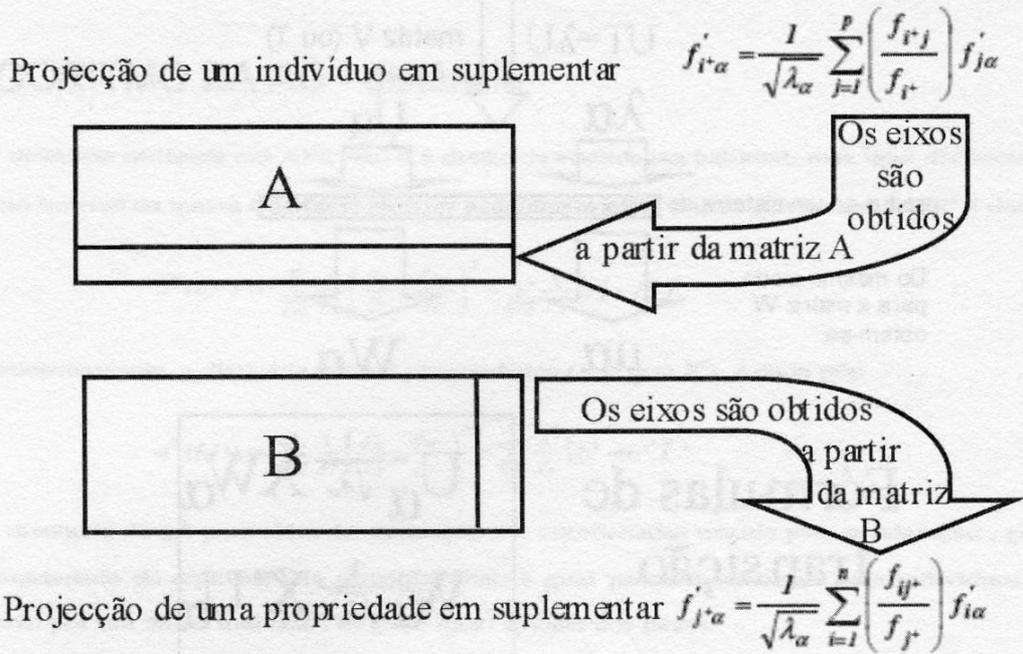


Figura A-8 Projecção em suplementar.

ALGORITMO DA AC - contribuições

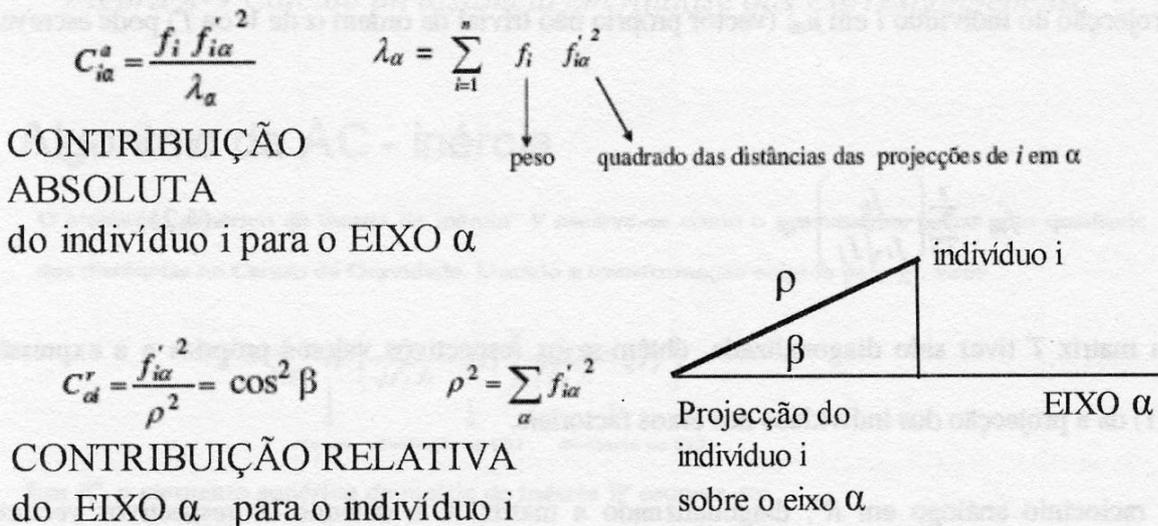
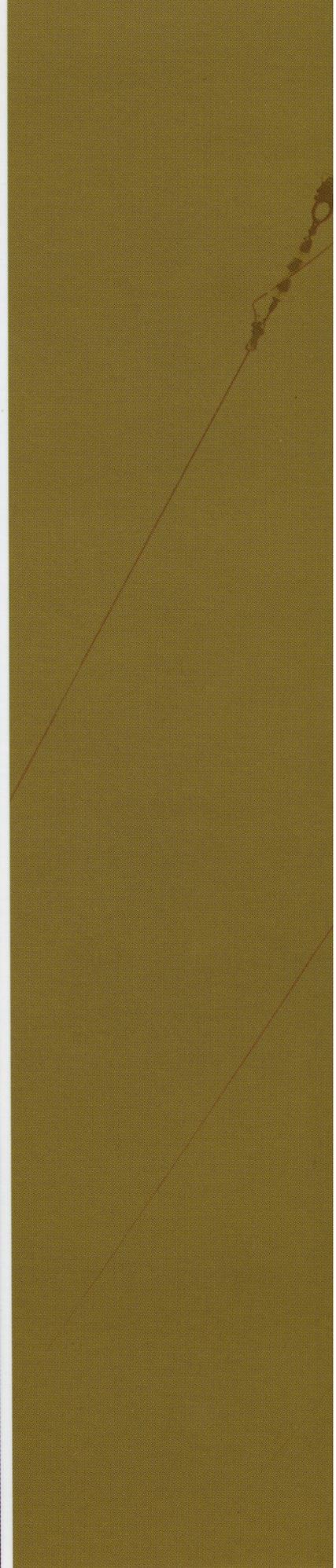


Figura A-9 Medidas de Ajustamento em Análise das Correspondências.

A partir das fórmulas dadas na Figura A-7 e A-8 são produzidos os gráficos que posicionam as linhas e colunas da matriz de partida em planos cruzando os eixos resultantes da diagonalização simbolicamente expressa na Figura A-6. Para interpretar esses eixos em termos dos dados de partida, utilizam-se as medidas de ajustamento definidas na Figura A-9.



ISBN 978-972-689-238-0



9 789726 892380



salamandra