

HELENA CARVALHO

2ª edição
Revista e actualizada

ANÁLISE MULTIVARIADA de DADOS QUALITATIVOS

**Utilização da Análise de
Correspondências Múltiplas
com o SPSS**

EDIÇÕES SÍLABO

Análise Multivariada de Dados Qualitativos

Utilização da Análise de Correspondências Múltiplas com o SPSS

HELENA CARVALHO

2ª Edição

Revista e Actualizada

EDIÇÕES SÍLABO

É expressamente proibido reproduzir, no todo ou em parte, sob qualquer forma ou meio, **nomeadamente fotocópia**, esta obra. As transgressões serão passíveis das penalizações previstas na legislação em vigor.

Visite a Sílabo na rede:
www.silabo.pt

Editor: Manuel Robalo

FICHA TÉCNICA:

Título: Análise Multivariada de Dados Qualitativos – Utilização da ACM com o SPSS

Autora: Helena Carvalho

© Edições Sílabo, Lda.

Capa: Pedro Mota

1ª Edição – Lisboa, Maio de 2008.

2ª Edição – Lisboa, Janeiro de 2017.

Impressão e acabamentos: Cafiessa – Soluções Gráficas, Lda.

Depósito Legal: 419613/16

ISBN: 978-972-618-873-5

EDIÇÕES SÍLABO, LDA.

R. Cidade de Manchester, 2

1170-100 LISBOA

Telf.: 218130345

Fax: 218166719

e-mail: silabo@silabo.pt

www.silabo.pt

ÍNDICE

PREFÁCIO	9
INTRODUÇÃO	13

CAPÍTULO 1

PRESSUPOSTOS

1. A abordagem da estruturalidade: requisitos metodológicos	19
2. Contributo dos métodos multidimensionais: análise de correspondências múltiplas e análise de homogeneidade	21
3. Operacionalização da estruturalidade: alguns exemplos	23

CAPÍTULO 2

A ACM NO CONTEXTO DA *OPTIMAL SCALING*

1. Introdução	35
2. <i>Inputs</i> para a ACM	37
3. Estrutura conceptual	41
3.1. Quantificação de dados qualitativos	42
3.2. Quantificação óptima	50
3.3. Quantificação múltipla	52

CAPÍTULO 3

MEDIDAS DE ANÁLISE DOS RESULTADOS: COMO INTERPRETAR

1. Variância das dimensões: valores próprios e inércia	61
2. Contribuição das variáveis	68
2.1. Medidas de discriminação e contribuições	68
2.2. Leitura gráfica das medidas de discriminação	75
2.3. Selecção das variáveis: interpretação das dimensões	78

3. Quantificações das categorias	95
3.1. Selecção das categorias	95
3.2. Interpretação das dimensões: via categorias	96
4. Tradução gráfica dos resultados: leitura dos planos	106
4.1. Representação das categorias: identificação de configurações	106
4.2. Representação dos objectos	115
5. Síntese do processo interpretativo de uma ACM	122

CAPÍTULO 4

PROCEDIMENTO ACM

1. Gestão das não-respostas numa ACM	127
1.1. Introdução	127
1.2. Estratégias de tratamento das não-respostas via procedimento ACM	129
2. Como realizar uma ACM	134
2.1. Requisitos de codificação	134
2.2. Codificação de categorias decorrentes de uma análise de conteúdo	138
2.3. Inicialização	142
2.4. <i>Outputs</i>	153

CAPÍTULO 5

VARIÁVEIS ACTIVAS VERSUS VARIÁVEIS SUPLEMENTARES

1. Topologia de um espaço simbólico	165
1.1. Indicadores estruturantes das configurações do espaço simbólico: <i>variáveis activas</i>	166
1.2. Eixos de diferenciação dos sistemas de valores e de representações: selecção e interpretação das dimensões	170
1.3. Quadros de valores e de representações, proximidades e oposições: interpretação do plano	180
2. As posições sociais no espaço simbólico	
– <i>variáveis suplementares</i> no espaço das <i>variáveis activas</i>	182
2.1. Distinção entre variáveis activas e variáveis suplementares	183
2.2. As posições sociais no espaço simbólico: contornos de articulação	185

CAPÍTULO 6

DA TOPOLOGIA À TIPOLOGIA ARTICULAÇÃO DA ACM COM A ANÁLISE DE CLUSTERS

1. Segmentação de um espaço de consumo: uma aplicação	193
1.1. Topologia do espaço de consumo: identificação de perfis	194
1.2. Construção da tipologia da ACM à análise de <i>clusters</i>	196
2. Tipologia de culturas	208

CAPÍTULO 7

DEFINIÇÃO DE ÍNDICES VIA ACM

1. Vantagens do uso dos <i>object scores</i>	216
2. Construção do índice: solução unidimensional	218

FICHA TÉCNICA 1

EDIÇÃO DOS GRÁFICOS DA ACM

1. Figuras bidimensionais	230
2. Conversão de gráficos: formato matricial para planos	239
2.1. Gráficos das categorias	239
2.2. Gráficos dos objectos	243

FICHA TÉCNICA 2

REALIZAÇÃO DA ACM COM VARIÁVEIS ACTIVAS E VARIÁVEIS PASSIVAS

250

BIBLIOGRAFIA	260
--------------	-----

PREFÁCIO

*La plus haute fonction des nombres
n'est pas l'énumération, mais l'imagination*

Jean-Paul Benzécri

A Análise Multivariada de Dados, na acepção da escola francesa de Benzécri, desenvolveu-se no interior dos círculos culturais ligados ao estruturalismo, especialmente na sua vertente dedicada à linguística. É significativo que a primeira lição de *Analyse des Données* tenha tido lugar em 1963 no Collège de France, perante os 'intelectuais' parisienses do Quartier Latin, que acolheram com entusiasmo o *nouveau venu* oriundo da Argélia, sublinhando a originalidade da sua teoria (esta fazia a ponte entre um formalismo matemático tão rigoroso como inovador e as aplicações nas 'novas' Ciências Humanas, cujos mentores estavam sedentos de uma qualquer fundamentação empírica para os modelos conceptuais que produziam com extrema prodigalidade).

Foi na crítica do *a priori* e dos métodos visando a «confirmação» de entidades pré-estabelecidas (à maneira das *syntactic structures* de Chomsky) que a abordagem exploratória de Benzécri tomou corpo, com o objectivo de «revelar» as estruturas que subjazem a um determinado conjunto de dados empíricos, fazendo-as emergir da própria análise, segundo o paradigma do método indutivo com raízes em Francis Bacon (a quem, de resto, o matemático francês presta homenagem, na sua importante contribuição para a *Encyclopaedia Universalis*, sob um expressivo título: «NOVUM ORGANUM»).

O método de Análise Multivariada de Dados *par excellence* criado por Benzécri – a Análise das Correspondências – foi desenvolvido com base nas tabelas de contingência, quadros de dupla entrada cruzando duas variáveis qualitativas e contendo as frequências absolutas da co-ocorrência das respectivas modalidades. Para este tipo de dados, antes dos trabalhos de Benzécri, o único tratamento 'quantitativo' disponível era o teste do χ quadrado, que apenas avaliava o grau de independência entre as duas variáveis de partida. Ora a Análise das Correspondências, mesmo na sua versão inicial, permitia estabelecer relações no interior de cada conjunto de modalidades (e entre os dois conjuntos), abrindo assim a porta para a possibilidade de uma certa modelização de variáveis qualitativas, o que constitui um importantíssimo avanço no processamento estatístico de tais variáveis. Mas desse cadinho inicial onde o método fora concebido, cedo (em 1973) o formalismo básico evoluiu naturalmente para a sua versão dita das Correspondências Múltiplas (ACM), que

mais não é do que a sua generalização para o caso de justaposição de tabelas de contingência. Esta nova versão é especialmente adaptada à interpretação de inquéritos por questionário, em que, da codificação preambular em 'disjuntiva completa', se passa à matriz de Burt, contendo todas as combinatórias do cruzamento das variáveis presentes no questionário, sejam elas nominais, ordinais ou quantitativas (divididas em classes).

Com a interpretação sistemática dos diagramas e tabelas decorrentes da Análise das Correspondências Múltiplas (onde é possível associar indivíduos com propriedades, projectados no mesmo gráfico), e particularmente com a « projecção em suplementar » (em que um subconjunto de indivíduos e/ou propriedades é relacionado com os restantes), este método de Análise Multivariada de Dados ganhou um estatuto de grande utilidade em todos os domínios onde se dispunha de uma base de dados amorfa, cujo sistema de interdependências surgia claramente como output do método, depois de uma exigente filtragem. Em particular nas Ciências Humanas cujo principal instrumento de penetração na realidade é o inquérito por questionário, a Análise das Correspondências Múltiplas revelou-se uma ferramenta utilíssima para ultrapassar (com um considerável poder 'explicativo') o procedimento clássico de 'apuramento', baseado em percentagens e cruzamento de variáveis duas a duas. Em especial, tornou possível o estabelecimento das relações entre as perguntas específicas que são objecto do inquérito e as variáveis de caracterização da população dos respondentes (que podem também ser segmentados em grupos, de acordo com diferentes critérios sugeridos pela própria análise).

Apesar de todas estas virtualidades, a ACM encontrou uma imensa resistência por parte dos utilizadores situados no exterior da esfera cultural francófona, em especial no seio da comunidade científica anglo-saxónica, apegada à Análise em Componentes Principais de Hotelling (datada dos anos 30 do século XX, sem qualquer potencialidade de modelização e específica para descrever relações lineares entre variáveis quantitativas). Para o caso de atributos qualitativos, o conjunto de métodos reconhecidos pela comunidade de investigadores acima referida situava-se no domínio do *optimal scaling*, criado pela escola americana de Young e apoiado no colectivo Gifi de Leiden (em que o procedimento era aparentemente inverso do que é adoptado em ACM, já que procurava afectar valores quantitativos às categorias das variáveis qualitativas, em vez de transformar em categorias os números reais).

A resistência à divulgação 'universal' da ACM deve-se – entre outros factores que determinam a incomunicabilidade habitual entre escolas de diferente genealogia e contexto – a um certo 'obscurantismo fundamentalista' que grassava nas 'hostes' da escola francesa, eivada de algum chauvinismo. De facto, para além de um quase 'espírito paroquial' de Benzécri e seus prosélitos (que dificultava o acesso à ACM por parte dos 'não-iniciados'), o software desenvolvido pelos adeptos da Analyse des Données era de difícil acesso e pouco amigável (escrito num Fortran intransponível para outros ambientes de trabalho).

Só em meados da década de 80 do século XX é que a 'metodologia francesa' extravasou do Hexagone, por (algum) mérito de Michael Greenacre, que publicou em inglês o primeiro texto dedicado exclusivamente à Análise das Correspondências (*Theory and Applications of Correspondence Analysis*).

É neste contexto de alguma competição entre as duas ‘escolas’ que se situa a Tese de Doutoramento da autora deste livro – defendida em 1998 no ISCTE sob a minha orientação –, que veio provar a equivalência, tanto formal como operacional, entre os dois grupos de Métodos de Análise Multivariada, ‘resolvendo’ com elegância a controvérsia que se arrastava há anos (e que obstaculizava a permeabilidade e a contaminação virtuosa entre as duas abordagens). Adicionalmente, num capítulo de *Temas e Métodos Quantitativos*, 1, volume editado por Elizabeth Reis e Manuel Alberto Ferreira para a Sílabo, Helena Carvalho divulga para um público mais vasto os resultados da sua investigação, afastando definitivamente os fantasmas da ‘disputa’ entre os partidários da ACM e a escola americana.

Na sequência deste trabalho, vem agora a lume este livro, pioneiro da difusão da ACM entre a comunidade científica portuguesa. Trata-se de uma obra que aborda a ACM nos seus múltiplos aspectos, circunscrevendo-se no entanto a este método (o que é inédito no nosso país).

Nesta obra, explicita-se detalhadamente e com incedível rigor a interpretação dos resultados, salientando o papel das ‘contribuições’ na leitura dos gráficos, o que é a única atitude cientificamente correcta para evitar conclusões apressadas (e muitas vezes falsas) sobre uma realidade complexa, que é muitas vezes excessivamente simplificada pela magia resultante do apelo topológico que descreve de uma forma redutora (mas atractiva) as proximidades e oposições entre as variáveis e indivíduos representados nos diagramas.

Apoiada em exemplos concretos tratados através do SPSS (o *package* mais ‘popular’ entre os investigadores das Ciências Humanas), a autora consegue comunicar brilhantemente com os leitores, transmitindo-lhes – com uma simplicidade rigorosa – a sua notável abordagem da ACM, onde todas as potencialidades desta técnica são exploradas, nomeadamente a gestão das não respostas e a codificação das perguntas abertas, o tema da projecção em suplementar como caminho para uma certa modelização, e os prolongamentos do método como base para o estabelecimento de tipologias e de índices.

Henrique Garcia Pereira

Prof. Catedrático do IST

INTRODUÇÃO

Este livro tem como objectivo principal explorar as potencialidades de utilização de um método de Análise de Dados – Análise de Correspondências Múltiplas também conhecido por Análise de Homogeneidade – particularmente apropriado à abordagem simultânea de múltiplos indicadores e ao tratamento de variáveis qualitativas (extensível a variáveis quantitativas, privilegiadamente categorizadas).

A importância prática e analítica da Análise de Correspondências Múltiplas decorre deste duplo atributo – múltiplas variáveis e de natureza qualitativa – existentes numa boa parte do material empírico da investigação científica e dos estudos aplicados, em domínios muito diversos, desde as ciências sociais (economia, sociologia, psicologia social) às ciências humanas (história, psicologia), desde as ciências empresariais (gestão, marketing) às ciências biológicas e da saúde, entre outros.

A sua leitura poderá, pois, ser útil a todos os investigadores e peritos, estudantes e professores, técnicos e profissionais que necessitem de realizar e/ou interpretar análises multivariadas com variáveis qualitativas (ou tratadas como tal), para assim averiguar as múltiplas associações que se estabelecem entre elas num contexto de interdependência.

Do ponto de vista técnico é possível operacionalizar a Análise de Correspondências Múltiplas no SPSS (*Statistical Package for the Social Sciences*), *software* de estatística com grande difusão, quer no mundo académico, quer no mundo empresarial. Os potenciais utilizadores da Análise de Correspondências Múltiplas – em particular os que tenham de a executar e não, apenas, de interpretar os seus resultados – terão já ou precisarão de adquirir um mínimo de familiaridade com o SPSS.

A estrutura do livro contempla sete capítulos os quais, dado o seu conteúdo, traduzem três objectivos complementares: apresentação dos pressupostos que levam à selecção da Análise de Correspondências Múltiplas como método de análise de dados (Capítulo 1), desenvolvimentos conceptuais e operacionais (capítulos 2, 3 e 4) e, por último, especificações e derivações decorrentes do uso da Análise de Correspondências Múltiplas (capítulos 5, 6 e 7).

À parte as diferenças que assinalam, e que justificam, cada capítulo, existe um aspecto que é comum a todos eles, consistindo na apresentação de exemplos ilustrativos. Convém desde logo esclarecer que apesar dos exemplos privilegiados nos vários capítulos remeterem para objectos de estudo relevantes e recorrentes na investigação em áreas como a sociologia ou a psicologia social, a gestão ou a pesquisa de mercados, bastará substituir o conteúdo temático das variáveis em análise para verificar como a aplicação da Análise de Correspondências Múltiplas é extensível a variadíssimos outros domínios.

O livro começa por uma sistematização das premissas de utilização da Análise de Correspondências Múltiplas (Capítulo 1). A par dessa reflexão apresentam-se as suas principais potencialidades, passando designadamente pela tradução gráfica dos resultados, como ilustram os exemplos que acompanham desde logo essa apresentação. No essencial, pretende-se que este capítulo inicie a familiarização do leitor com Análise de Correspondências Múltiplas.

Os capítulos 2 e 3 apresentam a Análise de Correspondências Múltiplas (ACM). O Capítulo 2 centra-se na estrutura conceptual da ACM, a qual reflecte a sua operacionalização no contexto dos procedimentos de *optimal scaling* para variáveis categorizadas, da responsabilidade da Escola de Leiden. Refira-se que essa exposição mais matemático-formal é acompanhada de forma sistemática pela apresentação de exemplos, os quais ilustram os diferentes conceitos e propriedades em análise.

De seguida são apresentadas as diversas medidas de análise dos resultados de uma ACM (Capítulo 3). São explorados todo o tipo de resultados, quer estatísticos, quer gráficos, disponíveis na estrutura de *outputs* da ACM via SPSS. No sentido de orientar a interpretação das estatísticas e dos gráficos disponibilizados, investe-se na sistematização de algumas propostas, ensaiando com situações concretas de investigação.

Porque este capítulo assinala uma fase essencial (a da interpretação) é apresentado no seu final um quadro-síntese, que se espera que possa ser um instrumento útil de apoio à interpretação de uma ACM.

Concluída a apresentação dos fundamentos da Análise de Correspondências Múltiplas, procede-se à explicação dos procedimentos necessários à sua execução (Capítulo 4). Nesse sentido é feita uma abordagem ao menu da ACM no SPSS (versões 15.0 e 16.0) para o que é necessário ter disponível o módulo *Categories*. Pela sua especificidade este capítulo é particularmente importante para os leitores que pretendam efectuar autonomamente todos os procedimentos da ACM. Este capítulo é completado com uma ficha técnica que inclui alguns elementos técnicos também importantes na interpretação, sobretudo gráfica, dos *outputs*.

Os capítulos seguintes disponibilizam novos desenvolvimentos analíticos, que envolvem especificações e derivações associadas ao uso da Análise de Correspondências Múltiplas. Eles são interessantes para os leitores que pretendam conhecer usos complementares que rentabilizem os resultados da ACM.

No Capítulo 5 é explorada a possibilidade de ser feita uma Análise de Correspondências Múltiplas para a qual faz sentido, por razões de natureza temático-substantiva, afectar às variáveis estatutos diferentes. Está em causa a distinção entre *variáveis activas* (entendam-se as variáveis responsáveis pela definição das dimensões) e *variáveis suplementares* ou ilustrativas.

Nos capítulos 6 e 7 investe-se em novas propostas de análise dos dados que são sustentadas pelos resultados obtidos com a aplicação da ACM.

No Capítulo 6 é apresentado um encadeamento metodológico entre a Análise de Correspondências Múltiplas e a Análise de *Clusters*, que pretende ser uma proposta para construir uma tipologia. Os exemplos apresentados evidenciam como é possível vir a definir uma tipologia, sendo a configuração dos grupos/tipos (que compõem

cada tipologia) conhecida *a priori* com os resultados da aplicação da Análise de Correspondências Múltiplas.

Uma outra proposta de reutilização dos resultados da Análise de Correspondências Múltiplas, em particular das dimensões, traduz-se na definição de *índices* (entenda-se variáveis compósitas) (Capítulo 8). Vai assim demonstrar-se como as quantificações óptimas – designadamente na versão *scores* – podem ser um instrumento de medida de determinados conceitos, designadamente quando é necessário hierarquizar indivíduos (entenda-se, unidades de análise). Note-se que estes três últimos capítulos pressupõem que o leitor esteja familiarizado, quer com os principais conceitos da Análise de Correspondências Múltiplas, quer com a sua interpretação.

Findo esse percurso, o leitor, independentemente das suas áreas de interesse, estará certamente em condições de aproveitar as potencialidades da Análise de Correspondências Múltiplas para tratar múltiplas variáveis qualitativas (ou submetidas previamente a um processo de categorização) e desenvolver análises estruturais da multidimensionalidade.

1

CAPÍTULO

PRESSUPOSTOS

Na elaboração de um plano de pesquisa que preveja uma recolha extensiva de informação empírica, a selecção dos métodos de análise de dados é uma tarefa (entre outras) de importância decisiva, quaisquer que sejam os fenómenos em estudo. É essencial garantir que os instrumentos que venham a ser seleccionados permitam operacionalizar as orientações sistematizadas no modelo de análise, e sejam consistentes com a natureza da informação recolhida.

Serão aqui privilegiadas as situações nas quais é necessário lidar conceptual e metodologicamente com objectos de estudo de configuração complexa. Importa perceber como se pode traduzir a complexidade ao nível da definição do modelo de análise, na perspectiva de vir a reflectir sobre uma estratégia metodológica que, apesar de centrada nos instrumentos de análise dos dados, se afigure susceptível de operar com os requisitos decorrentes dessa complexidade.

1

A ABORDAGEM DA ESTRUTURALIDADE: REQUISITOS METODOLÓGICOS

Existem muitas situações de pesquisa empírica, nas quais o investigador tem por referência modelos de análise que perspectivam, mesmo que não em exclusivo, uma *abordagem multifacetada e relacional* sobre o objecto em estudo – traços da sua configuração complexa. Daí que o entendimento da complexidade de muitos objectos de estudo, independentemente da área científica de referência, conduza ao desenvolvimento de análises baseadas em *abordagens estruturais*.

Um dos requisitos deste tipo de abordagem consiste na identificação dos múltiplos factores, considerados pertinentes para captar a estrutura dos fenómenos em estudo, do que os fenómenos sociais são somente um exemplo. Uma das premissas é, pois, a de que a *complexidade é multidimensional*.

Cada factor detém uma funcionalidade específica na definição da estrutura dos objectos em análise. Apesar de ser preservada a individualidade desses factores, é sobretudo essencial atender à sua interdependência e aos contornos das diferentes combinações que decorrem da sua interacção. Está em causa a aceitação do papel decisivo que também a análise relacional possui na descodificação da complexidade, que caracteriza muitos dos objectos de diferentes áreas de investigação. Daí poder afirmar-se que, em muitos casos, interessa preservar uma *abordagem estrutural da multidimensionalidade*.

Um objecto de estudo que permite ilustrar o que acaba de ser dito é o de *espaço social*. Este assemelha-se a uma matriz de posições, que têm associadas certas condições de existência, socialmente definidas, com configurações específicas.

Cada posição social reflecte uma combinação de múltiplas propriedades, devendo ser definida em função de um «sistema multidimensional de coordenadas» (Bourdieu, 1989:135). Desde propriedades como os recursos económicos, culturais, políticos, a outras como sexo, idade, ou estado civil. Cada uma dessas propriedades detém valor e efeito próprios e, daí, serem analiticamente irreduzíveis. No entanto, a avaliação da sua importância decorre da análise das relações que se estabelecem entre todas.

O espaço das posições sociais é somente um dos exemplos possíveis. São muitos os fenómenos – objecto de investigação em domínios diversificados, desde as ciências biológicas e da saúde até às ciências sociais – cuja configuração complexa leva a que o seu processo de conhecimento perspetive uma *abordagem multifacetada e relacional*. Mais importante do que fazer a análise dos múltiplos factores (componentes essenciais dessa estrutura) enquanto unidades independentes, valoriza-se a análise do sistema de relações que se definem entre eles.

Vai agora analisar-se as implicações dos pressupostos (entenda-se, requisitos) enunciados no que se refere, em particular, à selecção dos *métodos de análise de dados*. De acordo com o exposto, pretende-se interagir com situações de pesquisa empírica nas quais é necessário operacionalizar a multidimensionalidade, o que implica trabalhar com amplos sistemas de indicadores (que se traduzem em variáveis). Nesse sentido, e de entre o conjunto de métodos de análise estatística disponíveis, a escolha começa por se centrar, necessariamente, nos que estão aptos a garantir o desenvolvimento de análises multivariadas.

Convém referir que não se pretende minimizar a importância das análises mais específicas, e que podem implicar o uso de métodos de âmbito univariado e bivariado. Procura-se antes evidenciar que, nos casos em que se considera que os próprios conceitos (definidos no modelo de análise) são multidimensionais, será analiticamente inconsistente apreendê-los através do uso de instrumentos que não foram concebidos para contemplar um número elevado de variáveis. Deste modo, não é possível esperar que os métodos de análise com níveis de aplicação mais restritos – do ponto de vista do número de variáveis que estão aptos a incluir – possam ser um equivalente funcional dos métodos de análise multivariada.

Feita esta advertência, retoma-se a discussão relativa aos métodos de análise de dados susceptíveis de dar resposta aos requisitos explicitados. Para além da necessidade de ter de operar com múltiplos indicadores, pretende-se investigar objectos com configurações relacionais, pelo que uma abordagem estrutural se afigura particularmente adequada.

Reportando ao domínio mais específico das ciências sociais, são exemplo, e para além do já citado espaço das posições sociais, os estudos através dos quais se pretende definir: quadros de valores e de representações, perfis de consumo, estilos de vida, práticas culturais, mapeamento de culturas, segmentos de mercado, entre muitos outros. Em qualquer dos casos, e numa determinada fase, é fundamental a análise das relações entre os múltiplos indicadores seleccionados, na perspectiva de pesquisar as suas configurações. Essas diferentes configurações podem permitir identificar perfis associados, por sua vez, a diferentes grupos de indivíduos que,

coexistindo no mesmo espaço, partilham sistemas distintos de práticas, ou de estilos, ou de atitudes, ou de representações, ou de culturas. Da análise do posicionamento relativo dessas configurações, decorre a possibilidade de se aferir sobre o tipo de relações estabelecidas entre elas.

A acrescentar aos aspectos já referidos – *multidimensionalidade* e *análises estruturais* – destaca-se um terceiro, também decisivo, relacionado com a natureza das variáveis. Em particular, irá centrar-se a atenção nas *variáveis qualitativas*. Estas são variáveis que se traduzem através de categorias e daí também a designação de *variáveis categoriais* (*categorical variables*¹) ou variáveis categorizadas. Uma boa parte dos indicadores usados em algumas pesquisas, designadamente nos exemplos atrás referidos, são traduzidos segundo variáveis qualitativas (ou tratadas como tal).

A sistematização de métodos multidimensionais, aptos a lidar com este tipo de variáveis e a disponibilidade em termos de *software* estatístico, foi permitindo investir em propostas metodológicas que viabilizassem também, a análise quantitativa e multivariada de variáveis qualitativas.

Em síntese, podem então destacar-se três vectores decisivos para a selecção da Análise de Correspondências Múltiplas. São eles: a multidimensionalidade do espaço de análise, a abordagem estrutural dessa multidimensionalidade e a operacionalização de indicadores qualitativos. Acresce reforçar que interessa, antes de mais, preservar a convergência articulada desses vectores.

2

CONTRIBUTO DOS MÉTODOS MULTIDIMENSIONAIS: ANÁLISE DE CORRESPONDÊNCIAS MÚLTIPLAS E ANÁLISE DE HOMOGENEIDADE

O investigador, após identificar os indicadores que pretende usar e decidir que vai submetê-los a uma análise integrada, é desde logo confrontado com a inacessibilidade da leitura da matriz de partida, na qual estão registados os dados resultantes da fase anterior da pesquisa, a da recolha da informação. Para além de ser uma matriz de grandes dimensões (nela dispõem-se n indivíduos caracterizados segundo

(1) Esclareça-se que a tradução de *categorical variables* para *variáveis categoriais* e não *variáveis categóricas* tem por base uma análise da etimologia das duas palavras. Dado o contexto em referência, concluiu-se ser mais adequada a expressão *variáveis categoriais*.

m múltiplas propriedades), a sua análise deve ser realizada na perspectiva de que interessa interpretar o significado do todo, sendo este estruturado à custa da interacção dos múltiplos componentes.

Na sequência do figurino delineado é necessário que na sua análise permaneçam, a jusante, os traços da *multidimensionalidade de tipo estrutural* identificados a montante. Acresce ainda o facto de se admitir que estão em análise *variáveis qualitativas*. Em função destes requisitos existem opções metodológicas (ao nível da análise de dados) cujo ajuste aos mesmos é deveras interessante. Estão assim em referência a *Análise de Correspondências Múltiplas* (ACM) de tradição francesa e *Análise de Homogeneidade* (HOMALS) desenvolvida pela Escola de Leiden.

São várias as diferenças que existem entre a HOMALS e a ACM, do ponto de vista da sua estrutura conceptual e algébrica.¹ Não obstante, quando focalizadas na óptica da utilidade metodológica, que interessará a um investigador preocupado em garantir a combinação articulada dos vectores atrás sistematizados, é consensual a eficácia de ambas, para produzir resultados susceptíveis de lhes responder. Em particular, a *representação gráfica* dos dados recolhidos sobre o fenómeno em estudo é, com toda a propriedade, uma característica decisiva para sustentar a sua aplicação.

Procurar-se-á visualizar através da disposição relativa dos indivíduos ou das categorias, de preferência, em planos, as relações estatísticas delineadas entre as múltiplas variáveis em análise. A partir das configurações definidas pelas categorias das variáveis, espera-se (embora sempre de forma aproximada) conhecer a estrutura que caracteriza o objecto em análise. Por seu turno, ao exibirem diferentes combinações das propriedades envolvidas, as configurações reflectem a presença de grupos de indivíduos (relativamente) homogéneos.

Apesar dos objectos de estudo já referidos terem por enfoque indivíduos que representam pessoas, nada obsta a que sejam usadas como *input* outras unidades de análise. Estas podem ser, por exemplo, empresas, sectores de actividade, concelhos, países, marcas de produtos (objecto habitual em estudos de mercado). Inclusive, as categorias substantivas que derivam de uma análise de conteúdo, podem elas próprias ser submetidas a uma ACM ou a uma HOMALS, com o objectivo de identificar associações privilegiadas.

Seja qual for a natureza das unidades de análise, o que determina a opção pela *Análise de Correspondências Múltiplas* ou pela *Análise de Homogeneidade* é o propósito de realizar uma abordagem relacional sobre as múltiplas variáveis que caracterizam esses indivíduos, podendo vir a ser definidos diferentes *grupos*. A descrição desses grupos pode contemplar dois vectores analíticos:

1. identificação da especificidade das associações entre as categorias das múltiplas variáveis em análise, aferindo-se assim sobre o perfil de cada grupo;

(1) Para uma análise comparativa sobre os dois métodos ver Helena Carvalho (2000), «Homogeneidade e Correspondências Múltiplas: comparação de dois métodos de análise», em Elizabeth Reis e Manuel Alberto Ferreira (eds), *Temas em Métodos Quantitativos*, n.º1, Lisboa, Edições Sílabo.

2. observação do posicionamento relativo das várias configurações. A análise das distâncias entre essas configurações permite detectar a existência de relações de associação ou de oposição.

Mesmo sabendo que os gráficos serão sempre uma representação simplificada da complexidade que caracteriza o objecto de estudo, é possível através deles identificar e desenhar configurações sistémicas, só muito dificilmente legíveis a partir de uma matriz de grandes dimensões. Como se terá ocasião de ilustrar, esses gráficos podem ser, também, um instrumento interessante para ensaiar a construção de modelizações analíticas da realidade em estudo.

Ao ser conhecida a configuração topológica do espaço em análise – objectivo absolutamente transversal qualquer que seja a temática em estudo – pode então vir a usar-se um outro método, que permita que se passe da *topologia* à *tipologia*. Está em causa, e a título de exemplo, a posterior aplicação de um método de agrupamento – como a Análise de *Clusters* – para efectivamente distribuir em tipos/grupos os indivíduos em análise em função dos seus perfis, já delineados a partir dos planos da ACM ou da HOMALS.

Ainda relacionado com as potencialidades das representações gráficas, os novos eixos (dimensões), são também importantes na interpretação. Dada a sua definição compósita – sustentada pelas múltiplas variáveis de *input* na análise – esses eixos permitem identificar o que se poderiam designar como as *dimensões estruturantes* do espaço em análise. Trata-se de aferir se, em função do efeito combinado desses indicadores, existem dimensões com um papel determinante na configuração desse espaço. As dimensões têm no contexto da ACM (ou da HOLMALS) um papel equivalente ao das, por exemplo, componentes principais em uma Análise em Componentes Principais (ACP).

Ao nível da utilização posterior dos resultados da ACM ou da HOMALS, os *factores* (como são designados na ACM de tradição francesa) ou *dimensões* (designação adoptada pela Escola de Leiden) podem estar disponíveis enquanto novas variáveis, agora quantitativas, para outras aplicações.

3

OPERACIONALIZAÇÃO DA ESTRUTURALIDADE: ALGUNS EXEMPLOS

Antes de passar à exposição detalhada da Análise de Correspondências Múltiplas, julga-se ser interessante apresentar, ainda que de forma breve nesta fase, alguns exemplos. Eles vão sobretudo permitir ao leitor captar, em concreto, o que foi

dito sobre as vantagens da tradução gráfica dos resultados, quando é necessário interagir com espaços de análise multifacetados e relacionais, definidos à custa de indicadores (quase exclusivamente) categorizados.

Como proceder então para tornar legível e inteligível uma estrutura complexa com os requisitos que têm vindo a ser explicitados? Existem várias estratégias de decodificação e de cuja complementaridade pode obter-se a descrição e a interpretação do objecto de estudo. Uma aproximação possível passa por, num dado momento, identificar os principais *eixos* que sustentam essa estrutura. Tratar-se de encontrar, a partir da análise empírica, as dimensões mais determinantes na configuração do espaço em estudo. Dessa avaliação é também possível classificar as variáveis de *input* – indicadores de análise – no que respeita ao papel que assumem na definição dessas principais dimensões.

Para ilustrar sugere-se um exemplo cujo objectivo é a definição de *posições sociais*. Na base desse exemplo está um estudo para o qual foi definida uma amostra com cerca de 600 indivíduos, residentes na cidade de Lisboa. Pretendia-se conhecer, de entre vários aspectos, as expectativas desses indivíduos sobre o impacto que viria a ter para Lisboa a realização (na cidade) de um determinado evento.

Para a recolha de informação foi aplicado um inquérito por questionário, no qual figuravam também uma série de indicadores de caracterização sociodemográfica, parte dos quais foram considerados para levar a efeito a descrição das posições sociais dos indivíduos em análise, conforme se descreve no Quadro 1.1.

Veja-se então na Figura 1.1 como se concretiza a afectação das variáveis nas duas primeiras, e mais determinantes dimensões. A disposição dos indicadores permite constatar uma distribuição diferenciada.

A análise das variáveis permite: *i)* detectar as relações privilegiadas que se definem entre elas, em cada uma das (principais) dimensões e, em função dessas especificidades, *ii)* identificar os traços mais determinantes na definição das posições sociais dos indivíduos em observação.

As características de natureza mais sociodemográfica (idade, estado civil e estatuto da casa onde vive) são determinantes na primeira dimensão, enquanto as habilitações só se destacam na definição da segunda dimensão, em parceria com a categoria socioprofissional e o rendimento (Figura 1.1). Como elucida o seu posicionamento diagonal, estes dois indicadores são simultaneamente importantes na composição das duas dimensões.

Ensaçando uma possível classificação, a primeira dimensão destaca-se pela caracterização do *ciclo de vida* dos indivíduos, desde a esfera mais familiar à profissional. Enquanto a segunda tem como referencial a *distribuição de recursos*, designadamente escolares e económicos.

HELENA CARVALHO tem o doutoramento em Sociologia, na área de teoria e método, pelo Instituto Universitário de Lisboa (ISCTE-IUL). É professora auxiliar com agregação do Departamento de Métodos de Pesquisa Social do ISCTE-IUL. Lecciona e coordena diversas unidades curriculares de mestrado: Métodos Avançados de Análise de Dados, Métodos de Análise de Dados. Lecciona e coordena diversas unidades curriculares de doutoramento. Directora da pós-graduação em Análise de Dados em Ciências Sociais (ISCTE-IUL). É investigadora do Centro de Investigação e Estudos de Sociologia do ISCTE (CIES/ISCTE). Tem participado em diversos projectos de investigação desenvolvendo as suas competências em análise de dados. Tem publicado diversos livros e artigos. As suas principais linhas de investigação têm por enfoque a aplicação de métodos de análise multivariada e de interdependência para variáveis categorizadas, métodos de dependência para variáveis categorizadas, modelos de mediação e modelos de moderação com métodos paramétricos e com métodos não-paramétricos (*bootstrap*), modelos longitudinais via análise multinível e modelos de medida e equações estruturais.

Esta obra desenvolve e explora as potencialidades de utilização de um método de Análise de Dados – Análise de Correspondências Múltiplas (ACM), também conhecido por Análise de Homogeneidade (HOMALS) – particularmente apropriado à abordagem simultânea de múltiplos indicadores e ao tratamento de variáveis qualitativas (extensível a variáveis quantitativas, privilegiadamente categorizadas).

A apresentação da teoria é permanentemente acompanhada por ilustrações práticas. É também privilegiado o investimento na interacção do investigador com a gestão dos resultados, sendo apresentados inúmeros exemplos baseados em dados de investigações já realizadas.

A importância prática e analítica da Análise de Correspondências Múltiplas decorre de o duplo atributo *múltiplos indicadores* e *variáveis qualitativas* existir numa boa parte do material empírico da investigação científica e dos estudos aplicados em domínios muito diversos: ciências sociais (economia, sociologia, psicologia social), ciências humanas (história, psicologia), ciências empresariais (gestão, marketing), ciências biológicas e da saúde, entre outros.

Este é pois um livro que pode ser consultado com vantagens por investigadores e peritos, estudantes e professores, técnicos e profissionais das mais diversas áreas que necessitem realizar e/ou interpretar análises multivariadas com variáveis qualitativas ou pretendam consolidar e aprofundar os seus conhecimentos neste método.

