

# *Contribuições dos Métodos de Três-Vias para a Tomada de Decisões na Gestão*

Guilherme José Fresca Mirador de Andrade Castela

Professor Auxiliar

Faculdade de Economia – Universidade do Algarve

Março de 2023

# ÍNDICE

NOTA PRÉVIA.....	iii
<b>1 INTRODUÇÃO.....</b>	<b>6</b>
<b>2 O MÉTODO STATIS.....</b>	<b>10</b>
2.1 Análise da Interestrutura.....	12
2.2 Análise do Compromisso.....	13
2.3 Análise da Intraestrutura.....	14
2.4 Aplicação.....	14
<b>3 A ANÁLISE FATORIAL MÚLTIPLA.....</b>	<b>15</b>
3.1 A Estrutura de Dados.....	15
3.2 A Análise do Consenso.....	16
3.3 A Análise da Intraestrutura.....	17
3.4 A Análise das Trajetórias.....	18
3.5 Aplicação.....	20
<b>4 O MÉTODO DISTATIS.....</b>	<b>21</b>
4.1 As Tarefas de Classificação.....	21
4.2 A Operacionalização do DISTATIS.....	23
4.2.1 <i>Sorting Tasks</i> e Matriz de Ordenação.....	23
4.2.2 Escalas Multidimensionais.....	24
4.3 STATIS.....	26
4.4 Aplicação.....	26
<b>5 O MÉTODO COSTATIS.....</b>	<b>27</b>
5.1 Análise de Co-Inércia.....	27
5.2 A Operacionalização do COSTATIS .....	27
5.3 Aplicação.....	28
<b>REFERÊNCIAS BIBLIOGRÁFICA.....</b>	<b>29</b>
<b>ANEXO 1 .....</b>	<b>35</b>
<b>ANEXO 2 .....</b>	<b>36</b>
<b>ANEXO 3 .....</b>	<b>37</b>
<b>ANEXO 4 .....</b>	<b>38</b>

# Nota Prévia

Esta lição foi elaborada de modo a preencher um dos requisitos de admissão a provas de Agregação na Universidade do Algarve, no ramo das Ciências Económicas e Empresariais, especialidade em Gestão, em conformidade com o estabelecido na alínea a) do n.º 2 do artigo 8.º, do Decreto-Lei n.º 239/2007, de 19 de junho.

A lição que se apresenta encontra-se estruturada em torno de quatro áreas da gestão onde os processos de tomada de decisões têm por base informações complementares resultantes de métodos de Estatística Multivariada, designadamente: a **Gestão Municipal**, a **Gestão de Unidades de Saúde**, a **Gestão de Centros Comerciais** e a **Gestão de Resíduos Urbanos**.

Com efeito, a Estatística Multivariada é uma área da estatística que se concentra em análises de dados, envolvendo múltiplas variáveis simultaneamente e é frequentemente utilizada no processo da tomada de decisão na gestão, tendo em vista auxiliar o desenvolvimento das funções de Planeamento, de Organização, de Direção e de Controlo.

Neste contexto, a importância dos métodos estatísticos de Três-Vias revelam-se particularmente importantes para analisar conjuntos de dados complexos que auxiliam os gestores na tomada de decisões informadas, como por exemplo: compreender as relações entre múltiplas variáveis, identificar padrões e tendências, desenvolver modelos preditivos e promover uma alocação de recursos mais eficiente.

Destacam-se aplicações destes métodos na **Gestão Municipal**, a título exemplificativo, na análise de dados de saúde pública onde se produzem informações, as quais podem facilitar a decisão municipal na identificação de tendências e padrões nos dados, bem como no desenvolvimento de intervenções de saúde pública direcionadas para resolver o problema. Ainda, na análise de padrões de tráfego ao longo do tempo, em diferentes interseções e entre diferentes tipos de veículos onde se obtêm informações que auxiliam a decisão municipal a identificar áreas onde o congestionamento de tráfego é particularmente problemático e, conseqüentemente, desenvolver soluções direcionadas para aliviar o problema.

Os métodos estatísticos de Três-Vias podem ser igualmente importantes na **Gestão de Unidades de Saúde** onde permitem, por exemplo, relacionar dados demográficos de pacientes, diagnósticos médicos e resultados dos tratamentos. Afiguram-se, assim, uma ferramenta essencial para gestores de unidades ou de serviços de saúde que procuram melhorar os resultados dos pacientes, alocar recursos com mais eficiência, monitorizar o desempenho ao longo do tempo e, em síntese, para melhorar a qualidade da saúde.

Os centros comerciais são sistemas complexos que envolvem múltiplas dimensões, destacando-se o número e os tipos de lojas, tráfego de pedestres, receita de vendas, satisfação do cliente e custos operacionais. E, neste âmbito, os métodos de Três-Vias podem ajudar a analisar esses conjuntos de dados complexos e extrair *insights* significativos para apoiar a tomada de decisões na **Gestão de Centros Comerciais**. Por exemplo, na identificação dos fatores subjacentes ao desempenho de um centro comercial, a qual pode incluir fatores como o *mix* de lojas, a atratividade do local, as estratégias de marketing utilizadas e o nível de serviço prestado ao cliente. A utilidade e a importância destes métodos destacam-se na identificação dos indicadores mais relevantes na avaliação de desempenho, na avaliação dos impactos de intervenções específicas de gestão e na comparação de padrões da indústria de centros comerciais.

Nos últimos anos, os métodos de Três-Vias tornaram-se cada vez mais populares na **Gestão de Resíduos Urbanos**. Vários estudos aplicaram estes métodos para analisar dados de aterros sanitários, dados de gestão de resíduos sólidos municipais e dados de compostagem. As aplicações destes métodos nesta área apresentaram várias vantagens, por exemplo, na identificação dos fatores socioeconómicos críticos que afetam a gestão de resíduos urbanos e o seu impacto no meio ambiente e na análise da eficácia de diferentes estratégias de gestão de resíduos urbanos. Em suma, os métodos estatísticos de Três-Vias são ferramentas cruciais que fornecem informações aos gestores de resíduos e formuladores de políticas, sobre os impactos ambientais e económicos de diferentes estratégias, demonstrando elevada utilidade na orientação de tomada de decisões.

Deste modo, a presente Nota Prévía pretende enquadrar o leitor nesta lição que se encontra organizada em cinco capítulos e decorre em torno de quatro áreas da gestão, onde os processos de tomada de decisões informadas são evidenciados com recurso a 4 aplicações provenientes de dissertações de mestrado orientadas pelo candidato. Assim, o capítulo 1 apresenta a Introdução, onde se contextualiza esta temática, seguindo-se o capítulo 2, o qual expõe o método STATIS e mostra uma aplicação na área da Gestão Municipal. Por conseguinte, o capítulo 3 exhibe a Análise Fatorial Múltipla e apresenta uma aplicação na área da Gestão de Unidades de Saúde, o capítulo 4 apresenta o método DISTATIS e mostra uma aplicação na área da Gestão de Centros Comerciais e, por último, o capítulo 5 expõe o método COSTATIS e apresenta uma aplicação na área da Gestão de Resíduos Urbanos.

## 1. INTRODUÇÃO

*"Não há só um método para estudar as coisas."  
Aristóteles (-384 a -322)*

Os métodos de Análise Multivariada (AM) dedicam-se ao estudo de dados obtidos pela medição de mais de uma variável, sobre um conjunto de indivíduos ou objetos, representados numa estrutura de duas vias, denominada matriz ou quadro de dados. Com o desenvolvimento dos recursos informáticos, estes métodos têm demonstrado, nas últimas décadas, a sua eficácia no estudo de grandes quantidades de informação.

Efetivamente, a AM teve início com Pearson (1901) com o estudo de vetores e planos que melhor permitissem o ajustamento de um conjunto de pontos num espaço euclidiano. Posteriormente e, neste seguimento, Spearman (1904) lançou os fundamentos da Análise Fatorial. Porém, apenas mais tardiamente surge a Análise de Componentes Principais com Hotelling (1933), em sequência de desenvolvimentos dos trabalhos de Pearson e de Spearman.

A AM começa então a ser utilizada em diversas áreas, nomeadamente na psicologia, economia e biologia, com o contributo de Benzécri (1976), Harman (1967) e Morrison (1967), os quais contribuem com desenvolvimentos na ligação com a Análise de Componentes Principais, a Análise Fatorial e a Análise Discriminante.

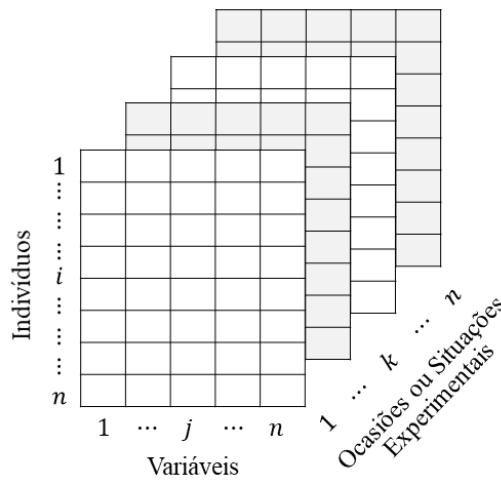
Na realidade, os fenómenos reais são, de um modo geral, de carácter multivariado e traduzem-se, genericamente, de forma dinâmica. Por outro lado, os métodos de análise de dados tradicionais caracterizam, usualmente, tratamentos de cariz estático e não conseguem, na maioria das vezes, captar potenciais dinamismos subjacentes à movimentação simultânea das variáveis envolvidas. De facto, uma compreensão profunda de um determinado fenómeno não se limita à sua perceção num único momento mas sim aquando de determinado período de tempo e/ou ocasiões experimentais.

Nesta linha, a AM introduziu alguma dinâmica no estudo evolutivo de fenómenos, possibilitando, assim, a análise de estruturas multidimensionais de múltiplos conjuntos de dados, dando origem aos denominados métodos de Três-Vias (*Three-Way Methods*), os quais tratam informação organizada em estruturas onde cada dado é representado por três vias:

- a) Uma para identificar os indivíduos que são objeto de estudo;
- b) Outra para as variáveis que se mediram sobre esses indivíduos e,
- c) Uma terceira, para identificar as diversas ocasiões em que essas medições se realizaram.

O termo ocasião pode referir-se, por um lado, a diferentes momentos no tempo ou, por outro lado, a diferentes situações de medição onde cada ocasião define uma matriz de dados (Figura 1.1).

Figura 1.1: Representação de uma estrutura de dados de Três-Vias



Fonte: Elaboração própria

Estas estruturas correspondem, geralmente, a dois tipos de organização de dados:

- i. Observações de diferentes conjuntos de indivíduos, num número fixo de variáveis, em diferentes ocasiões; ou
- ii. Observações dos mesmos indivíduos, em diferentes conjuntos de variáveis, em diferentes ocasiões.

Poder-se-á ainda considerar uma situação intermédia, a qual atribui a designação de dados cúbicos, ou triádico, a observações dos mesmos indivíduos, no mesmo conjunto de variáveis, em diferentes ocasiões.

Contudo, muito embora uma estrutura de dados de três vias (indivíduos, variáveis, ocasiões) se possa considerar como uma sucessão de quadros de duas vias (indivíduos e variáveis), os métodos estatísticos que tradicionalmente analisam estes últimos não são indicados para o tratamento dos de três vias, uma vez que não permitem analisar, em simultâneo, os diversos quadros de dados, dispostos por tempo ou por espaço, com o intuito de detetar as semelhanças e as diferenças observadas nas configurações das trajetórias dos indivíduos e das variáveis. Por outras palavras, os métodos de análise multivariada de dados de duas vias não identificam os padrões da estrutura interna presentes entre os elementos das configurações de dados de três vias. Deste modo, os métodos de análise multivariada de dados de três vias, ao descreverem a interação entre os três modos, geram análises mais completas e, neste contexto, assumem principal relevância ao permitir analisar em simultâneo diversas tabelas de dados, possibilitando uma análise mais robusta e mais aderente à realidade, pelo facto de serem capazes de perceberem os fenómenos e, essencialmente, de captarem as mudanças.

Os principais trabalhos referenciados na bibliografia, sobre este tipo de métodos, evidenciaram diferentes análises provenientes das escolas francesa e anglo-saxónica.

É com a escola francesa que, desde a década de 70, se destacam métodos como CANDECOM (*Canonical Decomposition*) de Carrol e Chang (1970), a Dupla Análise em Componentes Principais (DACP) de Bouroche (1975), o método STATIS (Escoufier, 1973; L'Hermier des Plantes, 1976), o método STATIS DUAL (Lebart, Morineau, e Piron, 1995) e a Análise Fatorial Múltipla (AFM) de Escoufier e Pagès (1985).

Na escola anglo-saxónica, desenvolveram-se numerosos trabalhos sobre a análise conjunta de matrizes de dados, nomeadamente, a família de modelos de Tucker (Tucker 1, 2 e 3) de Tucker (1966), o modelo PARAFAC (*Parallel Factor Analysis*) de Harshman (1970) e os modelos Tuckals 2 e Tuckals 3 de Kroonenberg e Leeuw (1980).

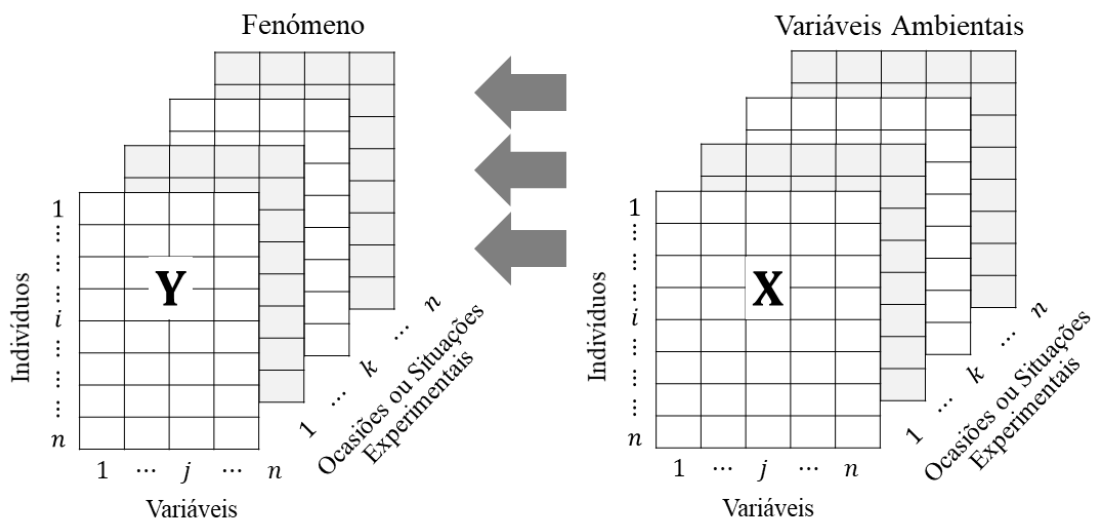
Entre alguns dos estudos que realizaram comparações entre métodos, destacam-se as investigações de Glaçon (1981), Pagès (1996), Kiers (1988,1989), Kroonenberg (1989) e Mures *et al.* (2006). Tais estudos comparativos assumiram extrema relevância como referência para a eleição do método mais apropriado para análise. A título exemplificativo, Kiers (1991) divide os métodos para análise de dados de Três-Vias em

assimétricos e simétricos. Os simétricos, de que são representantes os modelos Tucker e Tuckals, tratam de igual modo as três vias de uma estrutura de dados, enquanto os assimétricos (STATIS e AFM, por exemplo) dão um tratamento diferenciado a uma das vias, em regra, as ocasiões.

Normalmente, as estruturas de dados de Três-Vias incidem numa organização monobloco mais clássica dos dados para observações sobre o mesmo conjunto de indivíduos, nos quais se mediu o mesmo conjunto de variáveis, em diferentes ocasiões.

Contudo, por outras vezes, torna-se necessária a análise de múltiplos conjuntos de dados quando se pretende estudar e interpretar as relações entre as estruturas de duas séries de matrizes de dados sobre os mesmos indivíduos. Esta análise aplica-se, por exemplo, no caso do estudo simultâneo de uma série de matrizes de dados de contagem ou de frequência sobre um determinado fenómeno e, uma série de matrizes de variáveis ambientais que influenciam o fenómeno em estudo. Deste modo, podem existir, para as observações dos mesmos indivíduos, nas mesmas ocasiões, diferentes conjuntos de variáveis em cada série de matrizes. (Figura 3.2).

Figura 1.2: Estrutura multibloco de 2 conjuntos múltiplos de dados com indivíduos fixos



Fonte: Elaboração própria

Desta forma, a análise de dados de Três-Vias, em presença de múltiplos conjuntos de dados, revela, ainda, uma maior utilidade no fornecimento de informações sobre a estabilidade, diversidade, ou mesmo sobre as diferenças ou coincidências das estruturas

comuns em duas séries de matrizes em estudo, onde exista algum tipo de relação. E, neste contexto, foram desenvolvidos dois métodos, o STATICO (Simier *et al.*, 1999) e o COSTATIS (Thioulouse, 2011).

O presente trabalho pretende uma abordagem relativa a quatro métodos de Três-Vias da escola francesa, designadamente o STATIS, AFM, DISTATIS e COSTATIS. Não obstante, é crucial destacar que os três métodos primeiramente mencionados são métodos unibloco, em contraposição ao método COSTATIS, o qual é multibloco.

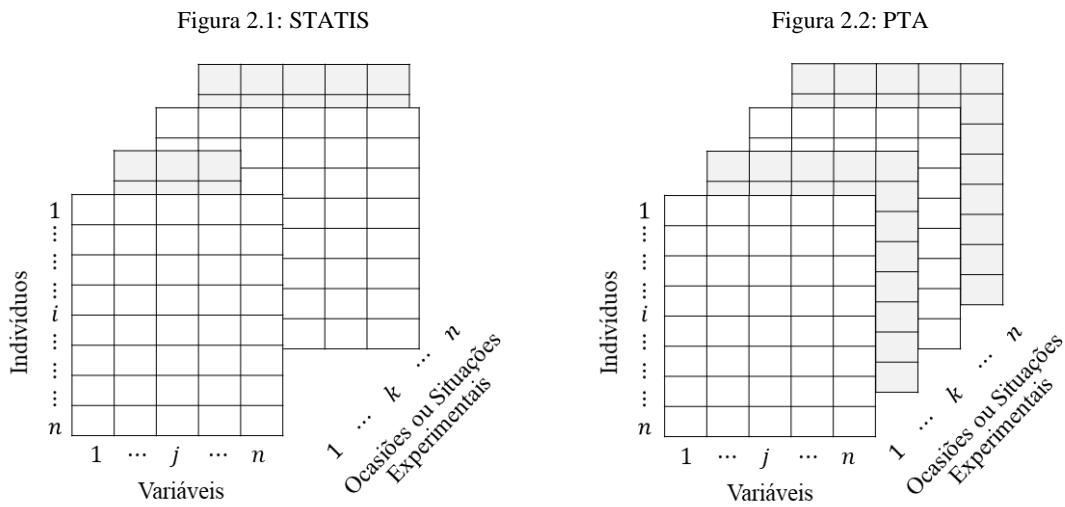
## **2. O MÉTODO STATIS**

O método STATIS (*Structuration de Tableaux À Trois Indices de la Statistique*) foi, de um ponto de vista primordial, introduzido por Escoufier (1973) e L'Hermier des Plantes (1976), no Laboratório de Probabilidades e Estatística da Universidade de Montpellier II, por volta de 1976 e, posteriormente desenvolvida por Lavit (1988b).

O STATIS é um método de análise exploratória de dados, o qual permite detetar uma estrutura comum em várias tabelas de dados obtidas durante um determinado período de tempo, ou no decorrer de várias situações experimentais, por forma a identificar as diferenças e as semelhanças ao longo dessas ocasiões. O STATIS pode ser ainda entendido como uma extensão da Análise de Componentes Principais (ACP) em presença de tabelas de dados relativos a variáveis recolhidas para as mesmas observações (Shinkareva *et al.*, 2014; Boumaza *et al.*, 2015; Rivadeneira *et al.*, 2016). Ainda, o método permite a extração de informações a partir de dados multidimensionais recolhidos em situações ou instantes diferentes.

O STATIS baseia-se na álgebra linear e, em particular, em espaços vetoriais euclidianos, pretendendo comparar configurações de indivíduos ou de variáveis, em diferentes horizontes temporais ou espaciais, para detetar uma estrutura comum, estável e representativa de todas as tabelas de dados. Assim, os dados de tipo quantitativo, apresentam-se, normalmente, através de  $k$  tabelas de dados, recolhidos em diferentes circunstâncias temporais ou espaciais, denominadas “ocasiões” ou “estudos”, sobre os mesmos  $n$  indivíduos, mas em que as  $p$  ou  $q$  variáveis podem diferir ao longo da série (Figura 2.1). Neste âmbito, a Análise Triádica Parcial (PTA) de Jaffrenou (1978)

distingue-se pela particularidade de ser utilizada quando todas as tabelas de dados contêm informação sobre os mesmos indivíduos e sobre as mesmas variáveis (Figura 2.2).



Fonte: Elaboração própria

Estes dois métodos de análise de dados de Três-Vias operacionalizam-se em três etapas: a análise da interestrutura, a análise do compromisso e a análise da intraestrutura, isto é:

- Etapa 1:** A Análise da Interestrutura consiste numa comparação global das tabelas de dados, por forma a identificar semelhanças e/ou diferenças, mas sem, contudo, identificar os indivíduos e/ou as variáveis responsáveis;
- Etapa 2:** A representação da interestrutura num plano fatorial permite aferir a possibilidade de se estabelecer uma matriz Compromisso, a qual, como o próprio nome indica, é uma matriz caraterizadora de uma estrutura comum à série original e representativa de todas as tabelas de dados;
- Etapa 3:** A definição do compromisso, por sua vez, possibilita a representação euclidiana das “posições” dos diversos indivíduos. Por outras palavras, mediante a deteção das associações entre as variáveis e os indivíduos torna-se possível interpretar as “posições” dos indivíduos no espaço-compromisso, no que concerne às semelhanças/diferenças entre as variáveis originais, conduzindo finalmente à identificação dos responsáveis na Análise da Intraestrutura. Por fim, a partir do espaço-compromisso traçam-se as trajetórias que descrevem o comportamento evolutivo de cada indivíduo ou de cada variável.

## 2.1 A Análise da Interestrutura

Em termos factuais, o STATIS permite, não só captar as trajetórias dos indivíduos e das variáveis ao longo do tempo, do espaço ou de situações experimentais, identificando os momentos em que as mudanças mais significativas ocorreram, mas também permite identificar as variáveis e os indivíduos responsáveis, calculando as suas contribuições para as alterações totais.

Na análise da interestrutura, os coeficientes  $R_V$ , entre as matrizes  $S_n$ , são incluídos na matriz  $R_V$ , a qual é estudada mediante uma ACP não-centrada e fornece um conjunto de pesos ótimos para combinar as matrizes  $S_n$ . Esta etapa analisa a similaridade entre as  $N$  matrizes  $S_n$ . Para isso, produz-se uma matriz  $N \times N$ , designada por matriz  $R_V$  (sujeitos), denotada por  $C$  cujo  $n, n'$  – ésimo elemento é o coeficiente  $R_V$  entre  $S_n$ , e  $S_{n'}$ . Este coeficiente  $R_V$  é calculado como:

$$R_V = (S_n, S_{n'}) = \frac{\text{trace}(S_n^T S_{n'})}{\sqrt{\text{trace}(S_n^T S_n) \times \text{trace}(S_{n'}^T S_{n'})}} \quad 2.1$$

O coeficiente  $R_V$  (Escoufier, 1973; Abdi, 2007b) é um cosseno ao quadrado entre matrizes (semi-definidas positivas) e a sua interpretação é semelhante a um coeficiente de correlação ao quadrado (o coeficiente  $R_V$  também pode ser usado para quantificar a similaridade). O coeficiente  $R_V$  varia entre 1 e 0 e indica a quantidade de informação que é partilhada entre duas matrizes. A análise da estrutura de similitude dos sujeitos é obtida a partir de uma Decomposição em Valores Singulares (DVS) de  $C$  (Abdi e Valentin, 2007d):

$$C = P \Theta P^T \text{ onde } P^T P = I \text{ e } \Theta \text{ é diagonal} \quad 2.2$$

Tal corresponde a uma ACP não centralizada de  $C$ . O mapa dos sujeitos, também chamado de mapa  $R_V$  (interestrutura) é obtido através da representação dos sujeitos que estão contidos na matriz  $G$ , a qual é calculada por:

$$G = P \Theta^{\frac{1}{2}} \quad 2.3$$

A primeira coluna de  $P$ , que representa o primeiro vetor próprio de  $C$  é designada por  $p_1$ . Como o coeficiente  $R_V$  é um cosseno ao quadrado, todos os elementos de  $p_1$  têm o mesmo sinal (tal é uma consequência do teorema de Perron-Frobenius (Lancaster e Tismenestsky, 1985) e esses elementos são escolhidos, por convenção, como positivos. O  $n$  – ésimo elemento do primeiro vetor próprio reflete quanto da O  $n$  – ésima matriz tem em comum com as outras matrizes: quanto maior o valor do O  $n$  – ésimo elemento, mais a O  $n$  – ésima matriz partilha informações com as outras matrizes.

## 2.2 A Análise do Compromisso

Portanto, os elementos de  $p_1$  podem ser usados para calcular um conjunto ideal de pesos para combinar as  $N$  matrizes num compromisso. O peso de uma matriz será proporcional à quantidade de informação comum transmitida por essa matriz. Especificamente, esses pesos são obtidos re-escalando o primeiro vetor próprio, de tal forma que a soma dos pesos seja igual a um. Formalmente, se denotarmos por  $\alpha$  o vetor de peso  $N \times 1$  cujos elementos são denotados  $\alpha_n$ , então:

$$\alpha = \frac{p_1}{p_1^T \mathbf{1}} \text{ onde } \mathbf{1} \text{ é um vetor } N \times 1 \text{ de } 1\text{'s} \quad 2.4$$

Por outro lado, na 2ª Etapa (Análise do Compromisso), o cálculo da matriz de compromisso é realizado através da soma ponderada das matrizes  $S_n$ . Na terceira etapa do DISTATIS combinam-se as  $N$  matrizes  $S_n$  numa matriz compromisso  $S_n$ , denominada de  $S_+$ , a qual é calculada como sendo a média ponderada das matrizes  $S_n$ , utilizando os elementos  $\alpha$  como pesos:

$$S_+ = \sum_{n=1}^N \alpha_n S_n \quad 2.5$$

Na quarta e última etapa do PTA, a matriz  $S_+$  é decomposta em valores e vetores próprios:

$$S_+ = V \Lambda V^T \text{ onde } V^T V = I \quad 2.6$$

As  $K$  categorias podem agora ser representadas como  $K$  pontos num mapa fatorial, ou seja, as pontuações dos fatores das  $K$  categorias são obtidas como:

$$F_+ = V \Lambda^{\frac{1}{2}} = S_+ \times V \Lambda^{-\frac{1}{2}} \quad 4.7$$

### **2.3 A Análise da Intraestrutura**

Equivalentemente, a matriz compromisso  $V \Lambda^{-\frac{1}{2}}$  é uma matriz de projeção, que também pode ser usada para projetar cada uma das matrizes  $S_n$  nos fatores do compromisso:

$$F_n = S_n \times V \Lambda^{-\frac{1}{2}} \quad 2.8$$

De facto, a configuração do compromisso descreve uma combinação linear das configurações dos  $N$  assuntos. As configurações observadas podem, simplesmente, ser projetadas (por uma transformação linear) no compromisso, e a distribuição das observações pode ser comparada ao compromisso.

De facto, o método STATIS permite, não só captar as trajetórias dos indivíduos e das variáveis ao longo do tempo, ocasiões ou condições experimentais identificando as estabilidades e as divergências mais significativas, como também reconhecer as variáveis e os indivíduos responsáveis por esses comportamentos, calculando as suas contribuições.

### **2.4 Aplicação**

Dissertação de Mestrado em Gestão Empresarial

**Filipe Lage**, Guilherme Castela e Nelson Silva (2016)

*ASSIMETRIAS NO CRESCIMENTO E NO DESENVOLVIMENTO ECONÓMICO: Um diagnóstico tridimensional da região do Algarve*

Resumo (ANEXO 1)

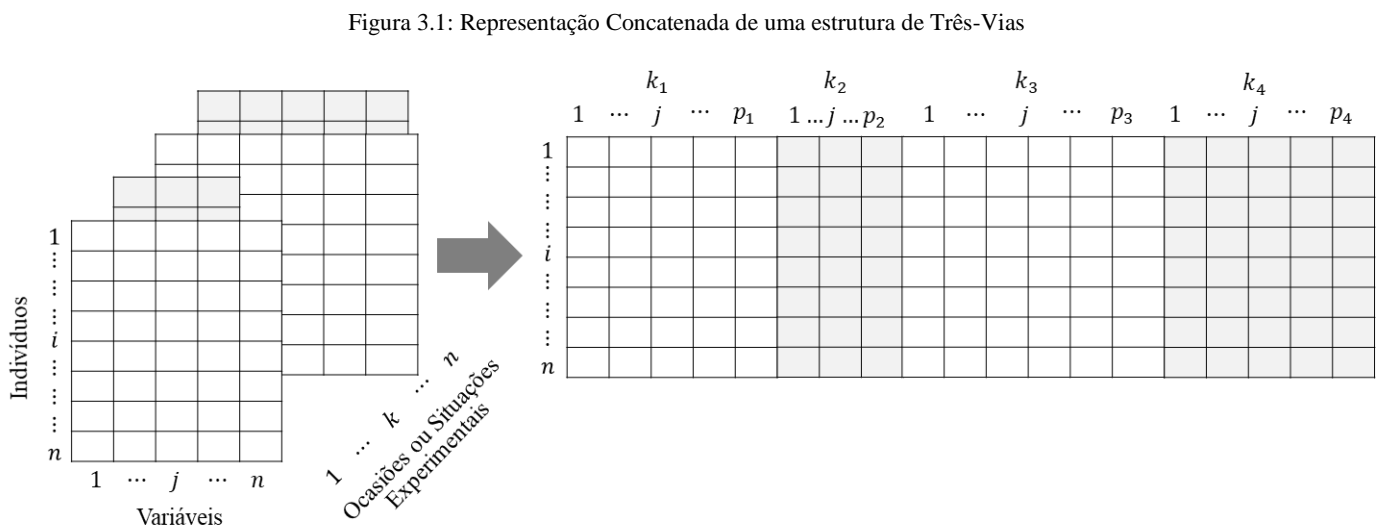
### 3. A ANÁLISE FATORIAL MÚLTIPLA

A Análise Fatorial Múltipla (AFM) foi desenvolvida por Escofier e Pagès (1985) e aplica-se ao tratamento simultâneo de uma sucessão de matrizes de dados que contêm os mesmos indivíduos. Estes são caracterizados por terem conjuntos de variáveis iguais ou diferentes, tanto quantitativas como qualitativas. Deste modo, trata-se de um método de análise exploratória de dados multivariados, baseado na álgebra linear e em espaços vetoriais euclidianos, que procura estabelecer uma estrutura comum e representativa de todas as matrizes que compõem a estrutura de dados.

Os dados de tipo quantitativo, são, então, apresentados em  $K$  matrizes, respeitantes a diferentes condições temporais, espaciais ou outras, e são denominadas por estudos, sobre os mesmos indivíduos, podendo as variáveis ser diferentes ao longo de cada matriz.

#### 3.1 A Estrutura de Dados

Para a estrutura de dados, é fulcral começar por representar um conjunto estrutural de dados ao qual seja possível aplicar uma AFM (Figura 3.1), em que temos  $k$  matrizes de dados ( $X_k$ ), onde  $n$  representa o número de indivíduos em cada matriz e  $p_k$  o número de variáveis no  $k$ -ésimo conjunto.



Fonte: Elaboração própria

Na AFM, um *estudo* é um trio estatístico  $(X_k, Q_k, D)$ , onde  $(X_k)_{n \times p_k}$  com  $k = 1, \dots, K$  representa a matriz de dados associada à  $k^{\text{ésima}}$  ocasião ou situação, o  $n$  é o número total de indivíduos e  $p_k$  consiste no número de variáveis no  $k^{\text{ésimo}}$  conjunto de dados.

A operacionalização da AFM encontra fundamento nos conceitos de Interestrutura, Consenso e Intraestrutura e considera duas etapas essenciais:

**Etapa 1:** Na primeira etapa, realiza-se uma ACP para cada um dos  $K$  grupos de variáveis.

O objetivo consiste na obtenção do primeiro valor próprio de cada análise ( $\lambda_1^k$ ), atendendo à sua utilização como fator de ponderação na etapa seguinte, bem como na avaliação de eventuais efeitos da dimensionalidade de cada grupo na análise que se segue;

**Etapa 2:** Na segunda etapa realiza-se uma ACP normada para a tabela global  $(K_{1\dots k})$ , a qual pondera cada tabela pelo inverso do primeiro valor próprio obtido na etapa anterior, isto, com o objetivo de equilibrar a influência dos  $K$  grupos de variáveis. Seguidamente, todas as  $K$  matrizes ponderadas são justapostas e, finalmente, realiza-se nova ACP sobre a matriz global correspondente. Esta abordagem dá origem à obtenção de um referencial comum chamado “espaço consenso” o qual permite o estudo e a análise da evolução ou dinamismo dos diferentes indivíduos e variáveis, implicando a obtenção de representações gráficas que facilitam a análise dessa evolução.

### 3.2 A Análise do Consenso

Por via desta etapa pretende-se a produção de uma matriz que resuma a informação das várias configurações do estudo, denominada “matriz consenso”.

Se a análise da interestrutura permitir concluir sobre a existência de uma estrutura de indivíduos que se apresente como comum às várias matrizes de dados, é possível proceder-se à construção da matriz consenso. A presente matriz consiste numa combinação linear dos diversos objetos representativos  $W_k$ , ponderados mediante  $\alpha_k$ .

$$W = \sum_{k=1}^K \alpha_k W_k \quad 3.1$$

As ponderações mencionadas representam as coordenadas do primeiro vetor próprio resultante da ACP, sobre cada um dos grupos de variáveis, e a utilização do primeiro valor próprio de cada análise como fator de ponderação para equilibrar a influência desses mesmos grupos de variáveis.

$$\alpha_k = \frac{1}{\lambda_1^k} \quad 3.2$$

No encadeamento do referido, o consenso  $W$  resulta da justaposição de todas as sub-tabelas assim ponderadas. Obtida a matriz consenso pretende-se, seguidamente, obter uma imagem euclidiana referente ao consenso dos indivíduos em estudo, ou seja, uma representação das posições assumidas por esses indivíduos.

Uma ACP sobre a matriz consenso  $W$ , permitirá obter  $B_1, \dots, B_n$ , ou seja, os pontos que representam os indivíduos na imagem euclidiana. As coordenadas destes pontos sobre o  $l^{\text{ésimo}}$  eixo são as coordenadas do vetor  $\sqrt{\theta_l} q_l$ , sendo  $q_l$  o  $l^{\text{ésimo}}$  vetor próprio associado ao  $l^{\text{ésimo}}$  maior valor próprio,  $\theta_l$ , da matriz  $W$ .

$$B_i = \sqrt{\theta_l} q_l = \frac{1}{\sqrt{\theta_l}} W D q_l \quad 3.3$$

Existindo evidência de uma estrutura de indivíduos comum às matrizes de dados, restringe-se a representação da imagem euclidiana do consenso aos dois ou três primeiros eixos verificados, segundo a percentagem de inércia explicada por estes.

### 3.3 A Análise da Intraestrutura

A análise das correlações das variáveis com os eixos do plano possibilita a interpretação das posições dos indivíduos na imagem euclidiana do consenso.

A correlação da variável  $p_k$  com o  $k^{\text{ésimo}}$  eixo é dada pela seguinte expressão:

$$\text{corr}(p_k, U) = X'_k D U_W \quad 3.4$$

Estas correlações resumem-se num gráfico em que a variável  $p_k$  é representada por um ponto no  $k^{\text{ésimo}}$  eixo, em que cuja coordenada é igual à correlação entre a variável e o eixo.

Então, são detetadas as associações entre as variáveis originais e as ocasiões ou situações experimentais. Estas associações permitirão interpretar as posições relativas dos indivíduos no espaço consenso, no que respeita às diferenças observadas entre as variáveis originais na imagem euclidiana do consenso.

### 3.4 A Análise das Trajetórias

Para visualizar graficamente os indivíduos responsáveis pelos desvios entre os estudos, representam-se as diferentes posições dos indivíduos para cada objeto da imagem euclidiana do consenso, diga-se, as suas trajetórias. Desta forma, obtém-se uma representação de  $nK$  pontos –  $n$  trajetórias, cada uma com  $K$  pontos.

Cada trajetória refere-se à deslocação de um indivíduo ao longo do horizonte temporal, delineando a evolução de cada um ao longo dos eixos do consenso. É possível uma melhor compreensão desta evolução pela sua associação com as variáveis que mais contribuíram para a formação de cada um dos eixos, ou, por outras palavras, pela correlação das variáveis com cada um dos eixos. A representação das trajetórias, na imagem euclidiana do consenso, consiste em representar as  $K$  nuvens de indivíduos.

Para o cálculo da coordenada de cada um dos pontos referentes ao indivíduo  $i$ , num determinado quadro  $k$ ,  $B_i^{(1)}, \dots, B_i^{(k)}$ , é utilizada uma expressão semelhante à expressão (3.3), mas agora calculada com base em cada um dos objetos representativos dos quadros de dados:

$$B_i^{(k)} = \frac{1}{\sqrt{\theta_l}} W_k D q_l \quad 3.5$$

O indivíduo consenso  $B_i$  é o centro de gravidade dos pontos  $B_i^{(1)}, \dots, B_i^{(k)}$  ponderados pelos coeficientes  $\alpha_1, \dots, \alpha_k$ , propriedade esta que se mantém ao longo da projeção (Saporta, 1990). Então, as trajetórias definem a mudança na posição de um indivíduo, ao longo do

período temporal ou condições/situações experimentais, tendo como base um indivíduo médio representado no consenso.

Quando os estudos não correspondem a um horizonte temporal, mas a situações experimentais, realiza-se o cálculo da distância entre cada indivíduo no estudo  $k$  e a sua posição consenso, baricentro da posição desse indivíduo para o conjunto de  $K$  estudos.

$$d_{B_i, B_i^{(k)}}^2 = \left\| (B_i) - (B_i)^{(k)} \right\|^2 \quad 3.6$$

Em síntese, a análise dos resultados da AFM tem por base as seguintes cinco fases fundamentais:

1. A análise do consenso traduz-se em resumir as  $K$  matrizes numa só. O espaço criado pelo consenso possibilita um resumo global para o conjunto dos  $K$  grupos;
2. A interpretação dos eixos da intraestrutura de acordo com as variáveis, atendendo às posições “consenso” dos indivíduos correspondentes às suas posições médias ao longo das  $K$  tabelas. A análise da intraestrutura envolve o conceito de trajetória ou distância em relação ao indivíduo médio. Tal conceito tem origem em estudos em que as tabelas são geradas temporalmente e em que se procura descrição da evolução ao longo do tempo do fenómeno em análise. No entanto, o conceito também é aplicável e interpretável a tabelas geradas a partir de diferentes situações experimentais, pugnando pelo conhecimento da evolução de cada um dos elementos que compõem as tabelas com mais detalhe;
3. O estudo da interestrutura e interpretação das posições das tabelas de dados, segundo o qual o objetivo da análise da interestrutura, igualmente designada por Análise Global, é comparar as  $K$  tabelas entre si e reconhecer grupos que contribuem para a estabilidade e para a divergência da realidade apresentada. Para efetuar tais comparações deve, não só, ser tido em conta o conceito de proximidade entre tabelas, como também procurar uma representação gráfica sobre a qual se possa interpretar a proximidade entre dois pontos, enquanto correspondência a duas tabelas similares, no sentido da distância entre eles. Através da análise da interestrutura, tornam-se claras as proximidades entre as tabelas de dados;

4. A interpretação das posições do consenso e das coordenadas parciais dos indivíduos;
5. As representações gráficas associadas à AFM têm por base a aplicação da ACP, o que permite projetar num espaço de reduzida dimensão os indivíduos do consenso, as nuvens de pontos parciais, as variáveis e grupos de variáveis resultantes da aplicação da AFM.

Em resumo, a AFM permite não só captar as influências sobre os indivíduos e as variáveis ao longo de diferentes situações experimentais ou ocasiões temporais, identificando os grupos de variáveis que mais contribuem para a coerência ou divergência do consenso, mas permite também identificar as variáveis e os indivíduos responsáveis por essas coerências ou incoerências da realidade apresentada.

### **3.5 Aplicação**

Dissertação de Mestrado em Gestão Empresarial

**Rogério Guerreiro**, Guilherme Castela e Nelson Silva (2022)

*INFLUÊNCIA DOS INDICADORES DE AVALIAÇÃO DE DESEMPENHO NA ATIVIDADE  
DOS CENTROS COMERCIAIS*

Resumo (ANEXO 2)

#### 4. O MÉTODO DISTATIS

O presente método afigura-se como uma generalização do Escalonamento Multidimensional (MDS), desenvolvido por Togerson (1958), e tem como objetivo primordial analisar várias matrizes de distâncias, calculadas sobre o mesmo conjunto de objetos, numa única representação gráfica. O DISTATIS é igualmente uma metodologia de Três-Vias que, embora sofisticada, é bastante versátil na análise das relações entre a estrutura de uma série de matrizes de dados, como um todo. Também é um método exploratório de análise de dados adequado a matrizes de distâncias de Três-Vias e pode ser utilizado numa variedade de situações aplicadas a tarefas de classificação.

Surgindo como um método cada vez mais utilizado para analisar tarefas de classificação (Lahne *et al.*, 2016; Schwartz *et al.*, 2015), o método avalia, primeiramente, a semelhança entre matrizes de distância e, a partir desta análise, é calculada uma matriz compromisso na qual as matrizes de distância originais são projetadas (Abdi *et al.*, 2005; Abdi *et al.*, 2007; Ballester *et al.*, 2009; Abdi *et al.*, 2012; Abdi *et al.*, 2013; Schwartz *et al.*, 2015; Lahne *et al.*, 2016; Gómez-Corona, 2017).

Destarte, trata-se de um novo método de Três-Vias que permite a análise simultânea de várias matrizes de distâncias. Este método combina o MDS clássico com o STATIS (Escoufier, 1976; Schlich, 1996; Abdi e Valentin 2007b; Santosa *et al.*, 2010; Varela e Ares 2014; Schwartz *et al.*, 2015) e tem a vantagem adicional de permitir a análise sensorial dos participantes de uma forma individual (Abdi, 2007; Varela e Ares 2014; Lahne *et al.*, 2016).

##### 4.1 As Tarefas de Classificação

Sendo o DISTATIS uma metodologia que assenta sobretudo na análise de tarefas de classificação suportadas através de matrizes de distâncias, o conceito de classificação torna-se, portanto, de crucial interesse. Uma tarefa de classificação (*sorting task*) traduz um método simples com elevado potencial de aplicação. Assim, trata-se de um processo de carácter cognitivo natural para recolher dados que sejam similares, relacionados entre si e, que não requerem uma resposta quantitativa. São usualmente dados da vida quotidiana que representam objetos, conceitos e produtos/serviços com características comuns, os quais são agrupados com a finalidade de obtenção de informações relevantes com um

esforço intelectual mínimo (Chollet *et al.*, 2011; Vidal *et al.*, 2014; Varela e Ares, 2014; Morrison *et al.*, 2017; Blanchard *et al.*, 2017; Clicerri *et al.*, 2017).

As tarefas de classificação remontam a 1935 (Hulin e Katz, 1935) e conheceram o seu início, nos anos 70, na área da psicologia (Hulin e Katz, 1935, Coxon, 1999), campo científico onde continuam a ser utilizadas frequentemente (Coxon, 1999; Nielson *et al.*, 2018; Gevaux e Petty, 2018). As *sorting tasks* foram usadas pela primeira vez na avaliação sensorial, no início dos anos 90, para investigar a estrutura perceptiva dos odores (Lawless, 1989; Lawless e Glatter, 1990; MacRae *et al.*, 1992; Stevens e O'Connell, 1996; Chrea *et al.*, 2005) e, na classificação de produtos alimentares por Lawless *et al.* (1995). Atualmente, é frequentemente utilizado na área do marketing exploratório para estudos de marca, de posicionamento e de categorização dos produtos (Chollet *et al.*, 2011; Blanchard *et al.*, 2017).

Neste tipo de tarefas de classificação, é solicitado a avaliadores (peritos, especialistas ou simples consumidores) uma classificação de estímulos tendo por base a similaridade percebida. Este processo é de extrema importância na identificação e na priorização de ideias (Chollet *et al.*, 2011; Laakso *et al.*, 2013; Vidal *et al.*, 2014; Morrison *et al.*, 2017; Nielson, *et al.*, 2018). De acordo com (Blanchard *et al.*, 2017), no domínio sensorial, estas tarefas de classificação são um modo eficiente e económico de se obterem informações relevantes sobre as diferenças sensoriais entre produtos e serviços, possibilitando, assim, determinar a vida útil e o desenvolvimento de produtos, serviços e processos.

Uma tarefa de classificação é executada numa única sessão. Todos os produtos/serviços são apresentados simultaneamente e exibidos aleatoriamente numa tabela, com uma ordem diferente para cada avaliador. A estes é solicitado que olhem primeiro, avaliem depois ou cheirem e/ou saboreiem todos os produtos (dependendo dos objetivos do estudo) e, em seguida, é-lhes pedido que classifiquem os produtos em grupos mutuamente exclusivos, tendo por base as semelhanças percebidas. Os avaliadores podem usar os critérios que desejarem para classificar os estímulos, e são livres para criarem quantos grupos quiserem e colocarem quantos produtos quiserem em cada grupo. A tarefa de classificação pode ser interrompida neste ponto, ou pode continuar com uma etapa de descrição onde os avaliadores são instados a descrever cada grupo de produtos (Lawless *et al.*, 1995; Tang e Heymann, 1999; Faye *et al.*, 2004, 2006; Saint-Eve *et al.*, 2004; Cartier *et al.*, 2006; Faye *et al.*, 2006; Blancher *et al.*, 2017; Lelievre *et al.*, 2008; Santosa *et al.*, 2010; Chollet *et al.*, 2011; Ballester *et al.*, 2013; Hopfer e Heymann, 2014; Bucher

*et al.*, 2016; Honoré-Chedozeau *et al.*, 2017; Guyot *et al.*, 2017; Morrison *et al.*, 2017; Nielson *et al.*, 2018; Gevaux e Petty, 2018).

Por último, a única grande desvantagem da análise MDS de tarefas de classificação, de acordo com Lawless *et al.* (1995), Abdi (2007), Varela e Ares (2014), Lahne *et al.* (2016) e Clicerri *et al.* (2017), é que as informações sobre os avaliadores são perdidas porque os dados individuais são agrupados por forma a obter-se uma matriz de similaridade. Como consequência, as diferenças individuais estão ocultas e a representação média pode ter pouca semelhança com cada uma das representações individuais do avaliador/especialista. Desta forma, não é possível visualizar as diferenças sensoriais e as percepções de cada participante. A variação individual no processo de categorização favorece uma representação espacial menos influenciada pelos avaliadores, porque estes comportam-se de formas distintas e muitas vezes únicas (Varela e Ares 2014; Clicerri *et al.*, 2017).

Uma possível solução para ultrapassar esta limitação seria usar o DISTATIS (Abdi *et al.*, 2005; Hopfer e Heymann., 2014; Schwartz *et al.*, 2015; Lahne *et al.*, 2016; Clicerri *et al.*, 2017; Gómez-Corona *et al.*, 2017; Lahne *et al.*, 2018).

## **4.2 A Operacionalização do DISTATIS**

Tradicionalmente existem três etapas envolvidas num procedimento DISTATIS:

**Etapa 1:** *Sorting Tasks* e Matriz de Ordenação

**Etapa 2:** Escalas Multidimensionais;

**Etapa 3:** STATIS.

### **4.2.1 *Sorting Tasks* e Matriz de Ordenação**

Tradicionalmente as experiências que utilizam *sorting tasks* são normalmente analisadas com recurso a técnicas de AMD, tais como o MDS (Togerson 1958; Abdi, 2007; Hopfer e Heymann, 2014; Ballester, 2013; Lahne *et al.*, 2016; Guyot *et al.*, 2017; Morrison *et al.*, 2017; Gómez-Corona *et al.*, 2017). E, neste âmbito, o procedimento é considerado como sendo uma etapa preliminar para uma aplicação do método DISTATIS.

Assim, é solicitado aos avaliadores uma classificação de estímulos tendo por base a similaridade percebida. Executada numa única sessão, esta tarefa de classificação pretende registar a ordenação dos avaliadores.

Na verdade, é gerada uma matriz de similaridade (a Matriz de Ordenação  $L_{[t]}$ ) através do cálculo do número de vezes que cada par de estímulos é classificado num mesmo grupo. E, quando se utiliza um método estatístico como o MDS por exemplo, essa matriz é analisada mediante um mapa fatorial desses estímulos, os quais são representados por pontos que são posicionados por forma a que as distâncias entre pares de pontos reflitam tão bem quanto possível as semelhanças e as diferenças entre os pares de estímulos (Varela e Ares, 2014; Lahne *et al.*, 2016; Morrison *et al.*, 2017).

#### 4.2.2 Escalas Multidimensionais

No DISTATIS, o *input* para o MDS é constituído por uma matriz simétrica de  $K \times K$  cujas entradas representam as distâncias (euclidianas) entre os  $K$  elementos de um determinado conjunto de interesse. No MDS, a matriz de distâncias é transformada num mapa fatorial no qual as  $K$  categorias são representadas através de pontos posicionados de tal forma que as distâncias entre elas apresentam a melhor aproximação das distâncias originais entre as categorias (Abdi *et al.*, 2005; Abdi *et al.*, 2007; Abdi *et al.*, 2012; Abdi *et al.*, 2014).

Assim, a partir de cada matriz de ordenação ( $L_{[t]}$ ), é criada uma matriz de Co-ocorrência ( $R_{[t]}$ ), onde o valor 1 significa que o objeto representado em linha e a respetiva classificação, representada em coluna, foram colocados no mesmo grupo.

$$R_{[t]} = L_{[t]} L'_{[t]} \quad 4.1$$

Por sua vez, de cada matriz de Co-ocorrência é derivada uma matriz de distância, na qual um valor de 0 implica que o objeto representado (em linha) e o respetivo grupo de pertença (em coluna) foram colocadas no mesmo grupo e um valor de 1 significa que os dois não foram agrupados juntos.

$$D_{[t]} = 1 - R_{[t]} \quad 4.2$$

Por outro lado, a matriz de massas representa um vetor de massa composto de números positivos cuja soma é igual a 1, onde, na maioria dos casos, todas as observações são de igual importância e cada elemento de  $m$  é igual a  $m_i = 1/I$ .

$$\begin{matrix} \mathbf{1} & m^T \\ I \times I & 1 \times I \end{matrix} \quad 4.3$$

A matriz  $S_n$  de  $K \times K$  é calculada por pré e pós-multiplicação da matriz de distância por uma matriz de centragem ( $\Xi$ ), definida como:

$$\Xi = I - \mathbf{1} m^T \quad 4.4$$

Se as matrizes de distâncias são expressas com diferentes escalas, as matrizes  $S_n$  têm de ser normalizadas. Tal é executado, dividindo as entradas de cada matriz de produtos cruzados,  $S_n$  pelo seu primeiro valor próprio. As matrizes normalizadas têm o seu primeiro valor próprio igual à unidade. Esta normalização garante que nenhuma matriz domina as primeiras dimensões da análise (quando várias matrizes são concatenadas, as matrizes com os maiores primeiros valores próprios irão dominar as primeiras dimensões da matriz comum (Escofier e Pagès, 1990).

Onde  $I$  é a matriz de identidade (conforme,  $K \times K$ ),  $m$  é um vetor  $K \times 1$  de massas (isto é,  $m_k \geq 0$  e  $\sum m_k = 1$ , frequentemente utiliza-se  $m_k = \frac{1}{k}$ ) e  $\mathbf{1}$  um vetor coluna (isto é,  $K \times 1$ ) de 1's. Assim, ocorre a transformação do conjunto de matrizes de distâncias num conjunto de matrizes de produtos cruzados ( $S_n$ ). Tal como no MDS, o primeiro passo do DISTATIS é transformar cada matriz de distância  $D_{[t]}$ , numa matriz de somas de produtos cruzados  $S_n$  (essa matriz é análoga a uma matriz de variância-covariância). Formalmente, a matriz  $S_n$  é obtida por:

$$S_n = -\frac{1}{2} \Xi D_n \Xi^T \quad 4.5$$

Seria, portanto, a partir desta matriz de produtos cruzados que, mediante uma Decomposição em Valores Singulares (DVS), usualmente via ACP, este procedimento de recolha de informação permite revelar a estrutura espacial de um determinado produto/serviço, sensação ou problema (Chollet *et al.*, 2011; Hopfer e Heymann, 2014;

Varela e Ares, 2014; Morrison *et al.*, 2017; Blanchard *et al.*, 2017 e Guyot *et al.*, 2017). Mais ainda, os padrões detetados nos resultados assim obtidos, são comparáveis àqueles obtidos através de outros métodos descritivos, (Tang e Heymann, 1999; Faye *et al.*, 2004; Saint Eve *et al.*, 2004; Faye *et al.*, 2006) e parecem ser mais fiáveis e robustos (Falahee e MacRae, 1997; Morrison *et al.*, 2017).

### **4.3 STATIS**

Nesta fase é aplicado o STATIS, mais propriamente uma PTA, de acordo com a descrição efetuada na secção 2. Deste modo, são desenvolvidas as três etapas do método: a Análise da Interestrutura, a Análise do Compromisso e a Análise da Intraestrutura.

### **4.4 Aplicação**

Mestrado em Gestão de Unidades de Saúde

**Alda Pereira**, Guilherme Castela e Nelson Silva (2018)

*O SERVIÇO DE URGÊNCIA DA UNIDADE DE FARO DO CENTRO HOSPITALAR  
UNIVERSITÁRIO DO ALGARVE: Um contributo do método DISTATIS para classificar e  
hierarquizar os principais problemas e desafios*

Resumo (ANEXO 3)

## 5. O MÉTODO COSTATIS

O COSTATIS é um método exploratório simétrico para analisar as relações entre as estruturas de duas séries de matrizes de dados, como um todo.

Na sua gênese, estão duas abordagens da análise de dados de Três-Vias, destacando-se a da escola francesa, onde se destacam métodos como a Dupla Análise em Componentes Principais (DACP) de Bouroche (1975), o método STATIS de Escoufier (1973); L'Hermier des Plantes (1976); Lavit, 1988; Lavit *et al.* (1994) e a Análise Fatorial Múltipla (AFM) de Escoufier e Pagès (1985).

Nesta sequência, o método COSTATIS (Thioulouse, 2011) é suportado pela aplicação de dois procedimentos, a Análise Parcial Triádica (PTA) (Jaffrenou, 1978) e a Análise de Co-Inércia (COIA) (Dolédec e Chessel, 1994; Dray *et al.*, 2003). Deste modo, o COSTATIS beneficia das vantagens dos dois métodos, analisa a estabilidade de cada uma das estruturas, em períodos temporais, espaços ou ocasiões diferentes através da PTA e, através de uma COIA, mede os níveis de influência entre ambas estruturas.

### 5.1 Análise de Co-Inércia

A COIA é um procedimento estatístico de acoplamento de informação entre duas matrizes de dados. Trata-se, de igual forma, de uma metodologia exploratória com a finalidade estudar a inter-relação entre duas matrizes de dados, como por exemplo, a medição das discrepâncias ou das coincidências entre as duas fontes de informação, que se supõem relacionadas.

### 5.2 A Operacionalização do COSTATIS

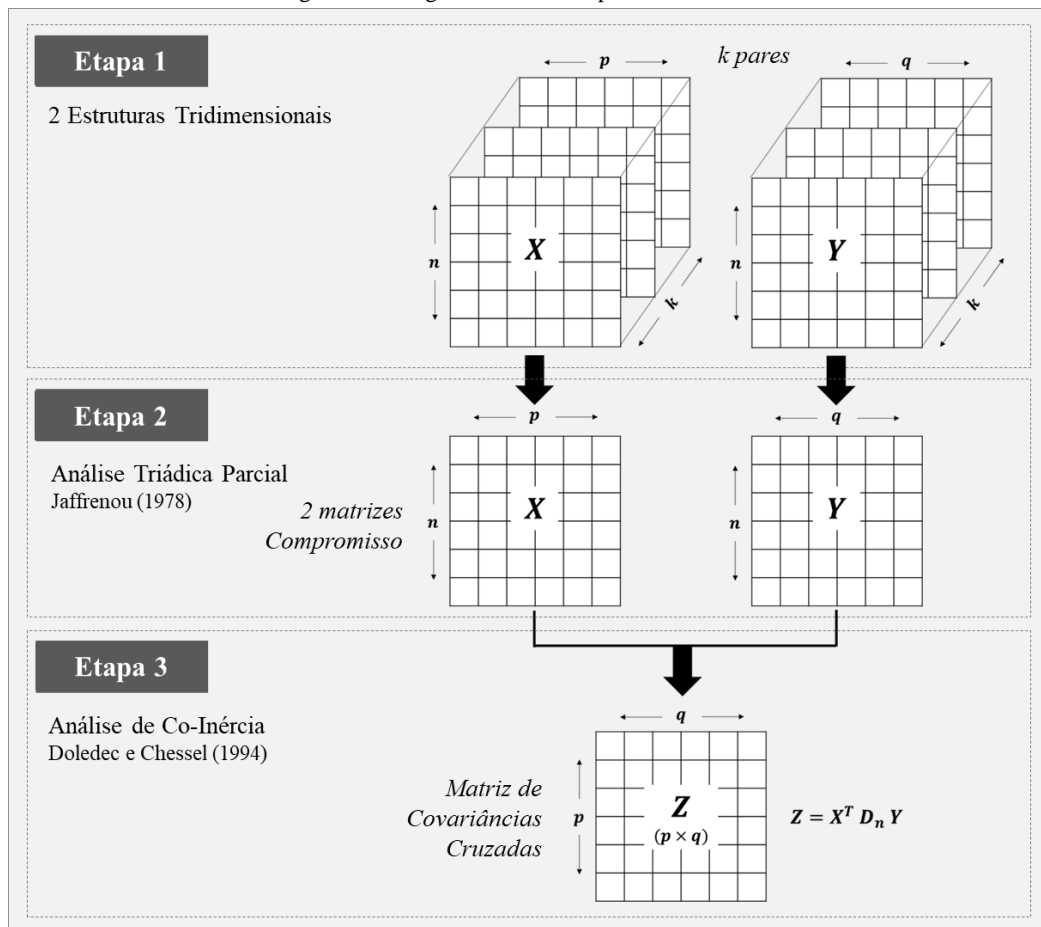
De uma forma sintética, o COSTATIS pode ser descrito como uma COIA de duas matrizes compromisso, obtidas mediante uma PTA para cada uma de duas séries de matrizes de dados, separadamente. Assim, o COSTATIS convencionou três etapas, como descreve a Figura 5.1.

**Etapa 1:** consiste em preparar, individualmente, duas estruturas tridimensionais de dados ( $X$  e  $Y$ ), com as mesmas  $p$  ou  $q$  variáveis, para todas as  $k$  repetições e com os mesmos  $n$  indivíduos.

**Etapa 2:** tem a finalidade de utilizar, simultaneamente, duas PTA para calcular duas matrizes compromisso, relativamente a  $X$  e  $Y$ , com o objetivo de analisar a estabilidade nas duas fontes de informação.

**Etapa 3:** por último, as duas matrizes compromisso são “acopladas” mediante uma COIA, a qual fornece uma imagem média da co-estrutura existente. Isto é, através da matriz de covariâncias cruzadas ( $Z$ ) analisa-se a inter-relação entre esses dois compromissos.

Figura 5.1: Diagrama de decisão para o COSTATIS



Fonte: Elaboração própria

### 5.3 Aplicação

Mestrado em Gestão Empresarial

Ana Custódio, Guilherme Castela e Nelson Silva (2021)

*GESTÃO MUNICIPAL DE RESÍDUOS URBANOS: um diagnóstico para a região do Algarve*

Resumo (ANEXO 4)

## REFERÊNCIAS BIBLIOGRÁFICAS

- Abdi, H., Valentin, D., O’Toole, A. J. e Edelman, B. (2005). DISTATIS: The analysis of multiple distance matrices. In *Proceedings of the IEEE Computer Society: International conference on computer vision and pattern recognition*, San Diego, CA, USA. 42–47.
- Abdi, H. (2007). RV Coefficient and Congruence Coefficient. *Encyclopedia of Measurement and Statistics*, 1–10.
- Abdi, H. (2007b). Metric Multidimensional Scaling (MDS): Analyzing Distance Matrices. *Encyclopedia of Measurement and Statistics*, 1–13.
- Abdi, H. e Valentin, D. (2007b). *DISTATIS: Encyclopedia of measurement and statistics*. N. Salkind (Ed.), Sage, Thousand Oaks.
- Abdi, H., e Valentin, D. (2007d). How to analyze multiple distance matrices. *Encyclopedia of Measurement and Statistics*, 1–15.
- Abdi, H., Williams, L. J., Valentin, D., e Bennani-Dosse, M. (2012). STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*. 4(2), 124–167.
- Abdi, H., Williams, L. J., e Valentin, D. (2013). Multiple factor analysis: Principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(2), 1-31.
- Ballester, J., Abdi, H., Langlois, J., Peyron, D., e Valentin, D. (2009). The odor of colors: Can wine experts and novices distinguish the odors of white, red, and rosé wines? *Chemosensory Perception*, 2(4), 203–213.
- Ballester, J., Mihnea, M., Peyron, D., e Valentin, D. (2013). Exploring minerality of Burgundy Chardonnay wines: A sensory approach with wine experts and trained panellists. *Australian Journal of Grape and Wine Research*, 19(2), 140–152.
- Benzécri, J. P. (1976). *L’Analyse des Données*, 2nd Ed, Dunod, Taxonomie.
- Blanchard, S. J., Aloise, D., e DeSarbo, W. S. (2017). Extracting Summary Piles from Sorting Task Data. *Journal of Marketing Research*, 54(3), 398–414.
- Bouroche, J. M. (1975). Analyse des données ternaires: la double Analyse en Composantes Principales. Thèse de 3e cycle, Université de Paris.

- Bucher, T., Collins, C., Diem, S., e Siegrist, M. (2016). Adolescents' perception of the healthiness of snacks. *Food Quality and Preference*, 50, 94–101.
- Carrol, J. e Chang, J. (1970). Analysis of Individual Differences in Multidimensional Scaling Via an n-Way Generalization of Eckart-Young Descomposition. *Psychometrika*, 35, 283-319.
- Cartier, R., Rytz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., e Martin, N. (2006). Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*, 17, 562–571.
- Chollet, S., Lelièvre, M., Abdi, H., e Valentin, D. (2011). Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask. *Food Quality and Preference*, 22(6), 507–520.
- Chrea, C., Valentin, D., Sulmont-Rosse, C., Ly, M.H., Nguyen, D., e Abdi, H. (2005). Semantic, typicality and odor representation: A cross-cultural study. *Chemical Senses*, 30, 37–49.
- Clicerì, D., Dinnella, C., Depezay, L., Morizet, D., Giboreau, A., Appleton, K. M., Hartwell H., e Monteleone, E. (2017). Exploring salient dimensions in a free sorting task: A cross-country study within the elderly population. *Food Quality and Preference*, 60, 19–30.
- Coxon, A.P.M. (1999). *Sorting Data: Collection and Analysis*. Thousand Oaks, CA: Sage Publications LTD. London. 7(127) Disponível em: <https://books.google.pt/books?id=W8VyAwAAQBAJ&printsec=copyright&hl=pt-PT#v=onepage&q&f=false>.
- Dolédec, S. e Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater biology*, 31(3), 277-294.
- Dray, S.; Chessel, D. e Thioulouse, J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology*, 84(11), 3078-3089.
- Escoufier, Y. (1973). Le Traitement des Variables Vectorielle. *Biometrics*, 29, 751-760.
- Escoufier, Y. (1976). Operators Related to a Data Matrix, in Barra et al. (ed.) *Recent Developments in Statistics*, Amsterdam, North-Holland, 125-131.
- Escoufier, B. e Pagès J. (1985). Mise en oeuvre de l'AFM pour les tableaux numériques, qualitatifs ou mixtes. Publication interne de l'IRISA, 429.

- Falahee, M. e MacRae, A.W. (1997). Perceptual variation among drinking waters: The reliability of sorting and ranking data for multidimensional scaling. *Food Quality and Preference*, 8, 389–394.
- Faye, P., Bremaud, D., Durand-Daubin, D., Courcoux, P., Giboreau, A., e Nicod, A. (2004). Perceptive free sorting and verbalization tasks with naive subjects: An alternative to descriptive mappings. *Food Quality and Preference*, 15, 781–791.
- Faye, P., Bremaud, D., Teillet, E., Courcoux, P., Giboreau, A., e Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference*, 17, 604–614.
- Gevaux, N. S., e Petty, S. (2018). Maximising resilience resources for mental healthcare staff. *Mental Health Review Journal*, 23(1), 37–53.
- Glaçon, F. (1981). L'Analyse Conjointe de Plusieurs Matrices de Données. *Comparaison de Différentes Méthodes*. Thèse de troisième cycle, Université de Grenoble.
- Gómez-Corona, C., Valentin, D., Escalona-Buendía, H. B., e Chollet, S. (2017). The role of gender and product consumption in the mental representation of industrial and craft beers: An exploratory study with Mexican consumers. *Food Quality and Preference*, 60, 31–39.
- Guyot, P., Houix, O., Misdariis, N., Susini, P., Pinquier, J., e André-Obrecht, R. (2017). Identification of categories of liquid sounds. *The Journal of the Acoustical Society of America*, 142(2), 878–889.
- Harman, H. H. (1967). *Modern Factor Analysis*, 2nd Ed, Chicago, University of Chicago Press.
- Harshman R.A. (1970). Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-mode factor analysis. *UCLA Working Papers in Phonetics*, v.16, 1-84.
- Honoré-Chedozeau, C., Lelièvre-Desmas, M., Ballester, J., Chollet, S., e Valentin, D. (2017). Knowledge representation among assessors through free hierarchical sorting and a semi-directed interview: Exploring Beaujolais wines. *Food Quality and Preference*, 57, 17–31.
- Hopfer, H., e Heymann, H. (2014). Judging wine quality: Do we need experts, consumers or trained panelists? *Food Quality and Preference*, 32(PC), 221–233.

- Hotelling, H. (1933). Analysis of a Complex of Statistical Variables in to Principal Components. *Journal of Educational Psychology*, 24, 417-441 and 498-520.
- Hulin, W.S. e Katz, D. (1935). The Frois-Wittmann pictures of facial expression. *Journal of Experimental Psychology*, 18, 482–498.
- Jaffrenou, P. A. (1978). Sur l’analyse des familles finies de variables vectorielles: bases algébriques et application à la description statistique. Available at: <https://books.google.com.ec/books?id=mc4GOgAACAAJ>.
- Kiers, H. (1988). Comparison of Anglo-Saxon and French Three-Mode Methods. *Statistique et Analyse des Données*, 13, 14-32.
- Kiers, H. (1989). *Three-Way Methods for the Analysis of Qualitative and Quantitative Two-Way Data*, Leiden, DSWO Press, **Leiden University**.
- Kiers, H. (1991). Hierarchical Relations among Three-Way Methods. *Psychometrika*, 56, 449-470.
- Kroonenberg, P. M. e Leeuw, J. (1980). Principal Component Analysis of Threemode Data by Means of Alternating Least-Squares Algorithms. *Psychometrika*, 45, 69-97.
- Kroonenberg, P. (1989). *Three-Mode Principal Component Analysis*. Leiden, DSWO Press, Leiden University.
- L’Hermier des Plantes (1976). Structuration des Tableaux à Trois Indices de la Statistique. Thèse de troisième cycle, Université de Montpellier.
- Laakso, K. (2013). Emergency management: Identifying problem domains in communication. *ISCRAM Conference Proceedings - 10th International Conference on Information Systems for Crisis Response and Management*, 724–729.
- Lahne, J., Collins, T. S., e Heymann, H. (2016). Replication Improves Sorting-Task Results Analyzed by DISTATIS in a Consumer Study of American Bourbon and Rye Whiskeys. *Journal of Food Science*, 81(5), S1263–S1271.
- Lancaster, P. e Tismenestsky, M. (1985). *The Theory of Matrices*. Academic Press, New York (1985).
- Lavit, C. (1988b). Presentation de la Méthode STATIS Permettant L’analyse Conjointe de Plusieurs Tableaux de Donn´ees Quantitatives. *Cachiers de la Recherche D´veloppement*, 18, 49-60.

- Lawless, H.T. (1989). Exploration of fragrances categories and ambiguous odors using multidimensional scaling and cluster analysis. *Chemical Senses*, 14, 349–360.
- Lawless, H.T. e Glatter, S. (1990). Consistency of multidimensional scaling models derived from odor sorting. *Journal of Sensory Studies*, 5, 217–230.
- Lawless, H.T., Sheng, N., e Knoops, S.S.C.P. (1995). Multidimensional scaling of sorting data applied to cheese perception. *Food Quality and Preference*, 6, 91–98.
- Lebart, L., Morineau, A. e Piron, M. (1995). *Statistique Exploratoire Multidimensionnelle*. horizon.documentation.ird.fr.
- Lelievre, M., Chollet, S., Abdi, H., e Valentin, D. (2008). What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, 19, 697–703.
- MacRae, A.W., Rawcliffe, T., Howgate, P., e Geelhoed, E.N. (1992). Patterns of odour similarity among carbonyls and their mixtures. *Chemical Senses*, 17, 119–125.
- Morrison, D. F. (1967). *Multivariate Statistical Methods*, New York, McGraw-Hill.
- Morrison, P., Gluyas, H., e Stomski, N. J. (2017). Structuring educational decisions using the multiple sorting task: An example focusing on international placements in nursing. *Nurse Education in Practice*, 26, 53–58.
- Mures, Q., Vallejo, P. e García, G. (2006). Comparación Empírica de Técnicas Estadísticas para Tablas de Tres Entradas: La construcción en Castilla y León en el periodo 2002-2004. *Pecunia*, 3, 95-140.
- Nielsen, K. T., Klokke, L., Guidetti, S., e Wæhrens, E. E. (2018). Identifying, organizing and prioritizing ideas on how to enhance ADL ability. *Scandinavian Journal of Occupational Therapy*, 8128, 1–12.
- Pagès, J. (1996). Éléments de Comparaison entre l'Analyse Factorielle Multiple et la Méthode STATIS. *Revue de Statistique Appliquée*, 44(4), 81-95.
- Pearson, K. (1901). On Lines and Planes of Closest to Systems of Points in Space. *Philosophical Magazine*, 2(6), 559-572.
- Saint-Eve, A., Paci Kora, E., e Martin, N. (2004). Impact of the olfactory quality and chemical complexity of the flavouring agent on the texture of low fat stirred yogurts assessed by three different sensory methodologies. *Food Quality and Preference*, 15, 655–668.

- Santosa, M., Abdi, H., e Guinard, J. X. (2010). A modified sorting task to investigate consumer perceptions of extra virgin olive oils. *Food Quality and Preference*, 21, 881–892.
- Schwartz, A. L., Kleeck, A. V., Beaton, D., Horne, E., Mackenzie, H. e Abdi, H. (2015). A Read-Aloud Storybook selection System for Prereaders at the Preschool Language Level: A Pilot Study. *Journal of Speech, Language, and Hearing Research*, 58, 1273-1291.
- Schlich, P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. In *Multivariate Analysis of Data in Sensory Science*, 1ª Ed. T. Noes and E. Risvik, eds. Amsterdam, the Netherlands: Elsevier Science. (169), 347.
- Simier, M., Blanc, L., Pellegrin, F. e Nandris, D. (1999). Approche simultanée de K couples de tableaux: Application à l'étude des relations pathologie végétale-environnement, *Revue de statistique appliquée*, 47(1), 31-46.
- Spearman, C. (1904). General Intelligence Objectively Determined and Measured. *American Journal of Psychology*, 15, 201-293.
- Stevens, D.A. e O'Connell, R.J. (1996). Semantic-free scaling of odor quality. *Physiological Behavior*, 60, 211–215.
- Tang, C. e Heymann, H. (1999). Multidimensional sorting, similarity scaling and free choice profiling of grape jellies. *Journal of Sensory Studies*, 17, 493–509.
- Thioulouse, J. (2011). Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. *The Annals of Applied Statistics*, 2300.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley.
- Tucker, L.R. (1966). Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika*, 31, 279-311.
- Varela P. e Ares G., (2014) *Novel Techniques in Sensory Characterization and Consumer Profiling*, 1ª ED, Boca Raton; CRC Press.
- Vidal, L., Cadena, R. S., Antúnez, L., Giménez, A., Varela, P., e Ares, G. (2014). Stability of sample configurations from projective mapping: How many consumers are necessary? *Food Quality and Preference*, 34, 79–87.

## **ANEXO 1**

**Filipe Lage**, Guilherme Castela e Nelson Silva (2016)  
Mestrado em Gestão Empresarial

*ASSIMETRIAS NO CRESCIMENTO E NO DESENVOLVIMENTO ECONÓMICO: Um diagnóstico tridimensional da região do Algarve*

### **RESUMO**

Tradicionalmente, a análise económica faz uso de um conjunto de ferramentas estatísticas para reproduzir e simular os principais mecanismos dos sistemas económicos regionais, nacionais ou internacionais. A fim de compreender a relação entre as variáveis socioeconómicas, usualmente aplicam-se modelos matemáticos para ajudar, em consonância com a teoria económica, o processo de tomada de decisão. No entanto, a teoria macroeconómica não é um campo particularmente consensual de investigação, contendo muitas teorias diversas e conflituantes. Neste contexto, temas relacionados com o crescimento e com o desenvolvimento podem, em alguns aspetos, ilustrar esta realidade. Questões quantitativas relativas ao Produto Interno Bruto, bem como as questões sociais podem, por exemplo, produzir associações que não são facilmente capturadas através dos modelos econométricos. A ocorrência destes factos representa, em nossa opinião, uma oportunidade para a utilização de ferramentas de estatística multivariada, tais como os métodos de Três-Vias. Efetivamente, com o crescente número de áreas de aplicação, a análise de dados de vias múltiplas tem-se tornado atrativa como ferramenta de análise exploratória, particularmente pela utilização de métodos como o STATIS. A finalidade destes métodos assimétricos é comparar várias tabelas de dados e investigar se existe uma estrutura comum entre elas. Um dos principais objetivos deste trabalho é mostrar como os dados socioeconómicos podem ser interpretados a partir de uma matriz complexa que pode ser significativamente melhorada através da aplicação de um método de Três-Vias. Assim, pretendemos com o presente estudo clarificar as dinâmicas estruturais e conjunturais que configuram assimetrias regionais. O recurso a um conjunto de dados socioeconómicos sobre os dezasseis municípios da região do Algarve, durante o período 2009-2012, possibilitou a construção de estruturas tridimensionais de dados que possibilitaram, não apenas detalhar a realidade económica e social do Algarve, como também complementar os métodos tradicionais de análise regional. Pensamos que um diagnóstico socioeconómico mais pormenorizado do Algarve será vantajoso para destacar informação útil para a implementação de futuras políticas regionais de gestão autárquica.

**Palavras-chave:** Algarve, crescimento, desenvolvimento, STATIS.

## **ANEXO 2**

**Rogério Guerreiro**, Guilherme Castela e Nelson Silva (2022)  
Mestrado em Gestão Empresarial

### *INFLUÊNCIA DOS INDICADORES DE AVALIAÇÃO DE DESEMPENHO NA ATIVIDADE DOS CENTROS COMERCIAIS*

#### **RESUMO**

Como outras organizações, os centros comerciais implementam sistemas integrados de gestão com o objetivo de responder melhor às exigências do ambiente externo, mas também com o intuito de melhorar os processos internos e maximizar recursos. Seja para aferir do desempenho dos sistemas de gestão ou dos centros comerciais enquanto organizações no seu todo, terão de existir indicadores de avaliação de desempenho que apurem os resultados das ações e atividades realizadas para se atingir os objetivos propostos. Com a aplicação da Análise Fatorial Múltipla aos resultados dos indicadores de avaliação e desempenho dos sistemas integrados de gestão de seis centros comerciais, geridos por uma mesma entidade gestora, pretendeu-se a compreensão das relações que tais indicadores têm com os centros comerciais, caracterizá-las e, daí, retirar informações úteis para a gestão. Os resultados obtidos permitiram perceber a existência de realidades distintas entre centros comerciais, formando-se grupos comportamentais, mas também entre os departamentos de gestão a que estão associados os indicadores de avaliação de desempenho. Em função da influência dos indicadores de avaliação de desempenho, foi possível desenvolver orientações de melhoria para a gestão dos diferentes grupos de centros comerciais, ajustadas aos diferentes departamentos.

**Palavras-chave:** centro comercial, sistemas de avaliação de desempenho, indicadores de avaliação de desempenho, KPI, estratégia, fatores críticos de sucesso.

### ANEXO 3

**Alda Pereira**, Guilherme Castela e Nelson Silva (2018)  
Mestrado em Gestão de Unidades de Saúde

*O SERVIÇO DE URGÊNCIA DA UNIDADE DE FARO DO CENTRO HOSPITALAR  
UNIVERSITÁRIO DO ALGARVE: Um contributo do método DISTATIS para classificar e  
hierarquizar os principais problemas e desafios*

#### RESUMO

Os serviços de urgência caracterizam-se por ambientes complexos e, por vezes, caóticos. Constituem desafios que são únicos, principalmente nas áreas da gestão financeira e na previsão orçamental, no que concerne ao volume de trabalho, equipamentos e recursos humanos necessários. A presente investigação aborda estas questões, primeiramente através da determinação dos dez problemas e/ou desafios do Serviço de Urgência do Hospital Universitário do Algarve em Faro e, de seguida revela as perceções e importâncias atribuídas por um painel interno de especialistas que intervêm neste processo de gestão. Os dados foram recolhidos em duas fases distintas: i) através da técnica de *Brainstorming*, com os chefes de banco e de turno do SU, tendo em vista opiniões e ideias para a aferição dos dez principais problemas; ii) apresentação das dez principais questões a um painel de especialistas que, individualmente, os classificaram e hierarquizaram. Os presentes dados foram analisados através de uma metodologia estatística multivariada através de uma técnica inovadora, o método DISTATIS. Assim, este afigura-se como um método exploratório de Três-Vias, o qual destaca as relações estáveis entre matrizes de dados estruturados (como um todo) através da generalização do Escalonamento Multidimensional e, que leva em conta os dados individuais. Realiza, ainda, cálculos em matrizes de distâncias individuais. Por sua vez, na avaliação das respostas dos especialistas, destacamos as opções caracterizadas por semelhanças e diferenças e, no respeitante à sua estabilidade ou instabilidade. Os resultados finais forneceram informações sobre a oposição e coerência entre especialistas, identificando temas comuns e problemas relevantes, que podem ajudar no processo de gestão e tomada de decisões da organização.

**Palavras-chave:** DISTATIS, Sorting Task, Gestão na Saúde, Serviços de Urgência, ineficiência.

## **ANEXO 4**

**Ana Custódio**, Guilherme Castela e Nelson Silva (2021)  
Mestrado em Gestão Empresarial

*GESTÃO MUNICIPAL DE RESÍDUOS URBANOS: um diagnóstico para a região do Algarve*

### **RESUMO**

Embora a Gestão Municipal de Resíduos Urbanos seja amplamente abordada a nível nacional, estudos que contribuam para a Gestão de Resíduos Urbanos (GRU), em subdivisões administrativas, ainda são limitados. Esta investigação é um estudo retrospectivo da região do Algarve, para o período 2014-2018, sobre os efeitos da estrutura e da conjuntura regionais na GRU. Ao nível estrutural, analisou-se o modo como a população e o setor do turismo influenciaram a GRU e, conjuntamente, a forma como influenciaram as atividades económicas e os negócios instalados na região. Devido às suas características geográficas e condições edafoclimáticas, a região Algarvia adquiriu o estatuto de principal região turística de Portugal. Com o decorrer do tempo, esta atividade tornou-se cada vez mais especializada e a economia regional cada vez mais direcionada para suprir as necessidades deste setor. Todavia, atividades como a GRU enfrentam desafios que não se colocam noutras regiões, existindo continuamente necessidades de adaptação. Portanto, conhecer comportamentos, tendências e as variáveis que influenciam a GRU, permitirá definir estratégias futuras que possibilitarão a melhoria da *performance* operacional das entidades gestoras, a redução de custos, o cumprimento de metas e a mitigação de impactos ambientais, beneficiando a atratividade da região. Estas circunstâncias manifestam-se cruciais nos processos de tomada de decisão, de tal forma, que se revela cada vez mais necessária a produção de informação estratégica para orientar futuros processos organizacionais, designadamente no que concerne a uma GRU mais eficiente e sustentável. Assim, este trabalho, para os dezasseis municípios do Algarve, utilizou a Análise de Dados de Três-Vias, designadamente o método COSTATIS, para produzir um diagnóstico, para o período 2014-2018, que avaliasse, ao longo do espaço e do tempo, a estabilidade da relação entre a População, o Turismo e a Economia Regional e a GRU. Julgamos que, com esta abordagem, seja possível auxiliar o processo de tomada de decisão e, assim, contribuir, não somente para uma GRU mais eficiente, como também para a sustentabilidade ambiental da região.

**Palavras-Chave:** Gestão de Resíduos Urbanos, População, Turismo, Economia Regional, COSTATIS.