SPATIAL AND ORGANIZATIONAL DYNAMICS

DISCUSSION PAPERS

Quantitative Methods Applied to Social Sciences

Nº 3

JULY, 2010

Inferencia Ecológica para la Caracterización de Abstencionistas: El Caso de Portugal
Ecological Inference for the Characterization of Electoral Abstainers: The Portuguese Case
Eugénia Castela, Maria Purificación Villardón

The Pathological-numeric Codification of Public Hospitals in Portugal:
Implementation of Mechanisms to Support the Assessment Process of Hospital Clinical Records and their Relationship with Funding
A Codificação Numérico-patológica dos Hospitais Públicos em Portugal:
Implementação de Mecanismos de Suporte no Processo de Avaliação dos Registos Clínicos Hospitalares e sua Relação com o Financiamento
Ana Marreiros, Guilherme Castela, Efigénio Rebelo, Maria Purificación Villardón

Um Contributo da Metodologia CHAID para a Detecção de Perfis Criminais:
Desenvolvimentos no Planeamento Estratégico e na Orientação Táctica dos Meios Policiais na Região do Algarve
A CHAID Contribution to the Detection of Criminal Profiles:
Developments in Strategic Planning and Tactical Guidance of Police Resources in the Algarve Region
Bruno Mora, Guilherme Castela

A Evolução da Investigação em Turismo: Uma Nova Plataforma de Conhecimento?
The Evolution of Tourism Research: A New Knowledge Platform?
Iris Lopes, João Albino Silva, Efigénio Rebelo, Guilherme Castela

UAlg
UNIVERSIDADE DO ALGARVE

RESEARCH CENTRE FOR SPATIAL AND ORGANIZATIONAL DYNAMICS

ISBN: 1647-3183

DISCUSSION PAPERS Nº 3:

SPATIAL AND ORGANIZATIONAL DYNAMICS

Quantitative Methods Applied to Social Sciences

Texts by:
Ana Marreiros
Bruno Mora
Efigénio Rebelo
Eugénia Castela
Guilherme Castela
Iris Lopes
João Albino Silva
Maria Purificación Villardón

July 2010

University of Algarve

TECHNICAL INFORMATION

Title - Discussion Papers Nº3: Spatial and Organizational Dynamics. Quantitative Methods Applied to Social Sciences

Authors - Several

Edition:
University of Algarve (www.ualg.pt)
CIEO – Research Center for Spatial and Organizational Dynamics
(Centro de Investigação sobre o Espaço e as Organizações)
Campus de Gambelas, Faculdade de Economia, Edifício 9
8005-139, Faro
cieo@ualg.pt
www.cieo.ualg.pt

Editorial Board:
Ana Paula Barreira
António Alhinho Covas
Maria Teresa Noronha
Thomas Panagopoulos

Page Layout:
Marlene Fernandes

Design and Cover Concept:
Helder Rodrigues
Hugo Pinto

Edition N˚3

ISSN: 1647-3183

Faro, July 2010

TABLE OF CONTENTS

Authors Presentation 4

Inferencia Ecológica para la Caracterización de Abstencionistas: El caso de Portugal 6
Eugénia Castela, Maria Purificación Galindo Villardón

1. Introducción 7
2. Justificación y objetivos de la investigación 9
3. El problema de Inferencia Ecológica (IE) 10
4. El modelo de King 11
5. La necesidad de homogeneidad espacial 15
6. Los métodos BIPLOT como herramienta de búsqueda de homogeneidad espacial 16
7. El método HJ-BIPLOT 16
8. Preparación de los datos y reorganización territorial según los métodos BIPLOT 17
9. Los Coeficientes Obtenidos en la Inferencia Ecológica, por Grupo 21
10. Conclusiones 22

A Codificação Numérico-Patológica dos Hospitais Públicos em Portugal: Implementação de Mecanismos de Suporte no Processo de Avaliação dos Registos Clínicos Hospitalares e sua Relação com o Financiamento 26
Ana Marreiros, Guilherme Castela, Efigénio Rebelo, Purificación Galindo Villardón

1. Introdução e Objectivos 27
2. Contextualização e Pertinência do Estudo 28
3. Metodologia 29
4. Resultados 31
5. Conclusões 36

Um Contributo da Metodologia CHAID para a Detecção de Perfis Criminais: Desenvolvimentos no Planeamento Estratégico e na Orientação Táctica dos Meios Policiais na Região do Algarve 39
Bruno Mora, Guilherme Castela

1. Introdução 40
2. A Prevenção Criminal 40
3. Os Modelos de Policiamento 41
4. O Comando Distrital de Polícia de Faro 42
5. A Metodologia da Investigação 43
6. Resultados 45
7. Conclusões 47

A Evolução da Investigação em Turismo: Uma Nova Plataforma de Conhecimento? 49
Iris Lopes, João Albino Silva, Efigénio Rebelo, Guilherme Castela

1. Introdução 50
2. Objectivos 51
3. Metodologia 51
4. Resultados 52
5. Conclusões 54

Editorial Norms 56

AUTHORS PRESENTATION

Ana Marreiros **(ammarreiros@ualg.pt, ana.marreiros@ismat.pt)**

Ana Marreiros is graduated in Statistics and Operational Investigation by the Faculty of Sciences, University of Lisbon. Has a Master in Statistics and Operational Investigation by the Faculty of Mathematics of the University of Seville and is a Ph.D. candidate in Quantitative Methods Applied to Economics and Management at the University of Algarve. Since 1998, she has been teaching subjects of Quantitative Methods area at the Manuel Teixeira Gomes, Superior Institute and, as Assistant Professor at the University of Algarve, in subjects of Data Analysis. She is Member at the Research Center for Spatial and Organizational Dynamics and her research areas are focused on Multivariate Statistics, Hospital Production and Clinical Codification.

Bruno Mora **(mora.bruno@gmail.com)**

Bruno Mora received a degree in Police Sciences and holds a post-graduate degree in General Management from the University of Algarve. He is completing his PhD thesis at the University of Algarve. His interests are in Computer Sciences and Quantitative Methods Applied to Economics and Management Studies.

Efigénio da Luz Rebelo **(elrebelo@ualg.pt)**

Efigénio Rebelo is Full Professor of Quantitative Methods Applied to Economics and Management Science and Dean of The Faculty of Economics in the University of Algarve.

Eugénia Maria Dores Maia Ferreira Castela **(ecastela@ualg.pt)**

Eugénia Castela has a degree in Applied Mathematics by the University Autónoma de Lisboa and PhD in Statistics by the University of Salamanca, Spain. Since 2000, teaches at the Faculty of Economics, University of Algarve, disciplines in the area of quantitative methods for the degrees of Economics, Business Administration and Sociology as well as in the masters. The main research topics focus on Multivariate Statistic and Ecological Inference.

Guilherme Castela **(gcastela@ualg.pt)**

Guilherme Castela is an Assistant Professor in the Faculty of Economics at the University of the Algarve. He holds a Ph.D. in Economics and Business Sciences and Applied Multivariate Statistics from the University of Salamanca. His main scientific areas of research are Strategic Management, Market Research, Data Analysis and both Qualitative and Quantitative Methods Applied to Economics and Management Studies.

Iris Lopes (igoncal@ualg.pt)

Iris Lopes received a degree in Management and holds a Master in Economics and Business Sciences from the University of Algarve, where she is an Assistant at the Faculty of Economics. She is completing her PhD thesis, at the PhD Program in Tourism of the University of Algarve. Her interests are in Tourism Research, Knowledge Management and Qualitative and Quantitative Methods Applied to Economics and Management Studies.

João Albino Matos da Silva (jsilva@ualg.pt)

João Albino is Full Professor at the University of the Algarve, President of the Scientific Committee of the Faculty of Economics at the University of the Algarve, President of the Tourism and Leisure Research Center of the University of the Algarve and Director of the PhD program in Tourism of the University of the Algarve. His main research areas of interest are: Tourism Planning, Management and Development; Quality and Management of Tourism Destinations; Economics of Tourism and Environment. He has authored books (research based) in Tourism Impact Assessment, Regional Development, Planning and Sustainability in Tourist Destinations and has articles published in International and National In-the-Field Journals. He is a member of Editing Committees.

Maria Purificación Villardón (pgalindo@gugu.usal.es)

Maria Purificación Villardón is, at present, Professor at the University of Salamanca. She has a PhD in Biostatistics from the same University and her areas of research are Biostatistics and Analysis of Multivariate Data.

INFERENCIA ECOLÓGICA PARA LA CARACTERIZACIÓN DE ABSTENCIONISTAS: EL CASO DE PORTUGAL

ECOLOGICAL INFERENCE FOR THE CHARACTERIZATION OF ELECTORAL TURNOUT: THE PORTUGUESE CASE

Eugénia Castela
Maria Purificación Villardón

RESUMEN

Los métodos de Inferencia Ecológica (IE) son métodos estadísticos que permiten estimar las celdas de una tabla de contingencia, cuando solo son conocidos los respectivos totales marginales. A partir del modelo de King (1997) y considerando las elecciones legislativas ocurridas en Portugal en 2002 y 2005, pretendemos encontrar los coeficientes de estabilidad (los ciudadanos mantienen la misma actitud ante el voto en las dos elecciones, es decir, cuando votan o se abstienen, en las dos elecciones, consecutivamente) e inestabilidad electoral (los ciudadanos votan en una de ellas y se abstienen en la otra, independientemente del orden) para todos y cada uno de los municipios de Portugal.
En el caso de Portugal, el método de King no ha producido buenas estimaciones.
Así, con vista a obtener la homogeneidad espacial, en términos de tendencias políticas mayoritarias en las dos elecciones en estudio, nuestra propuesta se basa en una "reorganización" territorial partidista/abstencionista, a partir del método HJ-Biplot (Galindo, 1986).
La "reorganización" territorial, proporcionó 6 grupos de provincias a los cuales se aplica el modelo de King para encontrar el porcentaje de electores que han votado y se han abstenido en las dos elecciones, así como el porcentaje de electores fluctuantes, es decir, los electores que han votado en una de las elecciones y se abstenido en la otra.

Palabras Clave: Inferencia Ecológica, HJ-Biplot, Organización territorial, Elecciones portuguesas.

ABSTRACT

Ecological Inference (IE) is a set of statistical methods that estimate the cells of a contingency table when only the marginal totals are known. Based on King's model (1997) and considering the legislative elections in Portugal between the years 2002 and 2005, we try to find the stability coefficients (the citizens who, keep the same attitude towards voting in both elections, i.e., they opt for vote or abstention in two consecutive elections) and electoral instability (the citizens who vote in one election and opt for abstention on the other, regardless of the order) for every and each of the municipalities in Portugal.
In the Portuguese case, King's method did not give good estimations. Therefore, in order to find spatial homogeneity in terms of the main political tendencies on the elections under study, we propose territorial "reorganization" based on an abstention pattern arising from the HJ-Biplot method (Galindo, 1986).
The territorial "reorganisation" has provided 6 groups of provinces to which King's model was applied in order to find the percentage of electors who voted or chose abstention in both elections, as well as the percentage of floating electors, i.e., the electors who voted in one election and not on the other.

Keywords: Ecological Inference, HJ-Biplot, Territorial Organization, Portuguese Elections
JEL Classification: C0

1. INTRODUCCIÓN

La participación en el proceso electoral y en la vida política en general constituye uno de los elementos esenciales de la vida democrática. Es, sin duda, para la gran mayoría de los ciudadanos, la única forma de participación política ejercida con cierta regularidad. Así, la intervención de los ciudadanos en el Gobierno del estado constituye una preocupación fundamental, siendo la participación política definida como, aquellas acciones de ciudadanos particulares a través de las que se persigue apoyar o influir en el Gobierno y en la política (Nie y Verba, 1975; Milbrath y Goel, 1977).

En este contexto, la abstención electoral es una constante en las democracias pluralistas. Afirmar que existen sociedades democráticas significa, de acuerdo con Lipset (1963) *"… admitir, que éstas pueden existir con diferentes niveles de participación, lo que significa, con diferentes niveles de abstención".* La estrecha relación existente entre participación electoral y democracia ha generado una importante cantidad de trabajos relacionados con las causas de la abstención, es decir, explicar por qué unos electores votan y otros no.

Sin embargo, de acuerdo con Piven y Cloward (1977) y Teixeira (1992), delimitar la participación/abstención electoral, teniendo exclusivamente por base la interpretación del comportamiento de los individuos, puede ocultar otras realidades; en otras palabras, una baja participación electoral es muchas veces considerada negativa para la democracia, sea simplemente intrínseca o porque ponga en peligro la legitimidad del sistema o bien porque implique la ausencia de representación de determinados grupos y políticas diferentes.

Desde un punto de vista empírico, la investigación politológica reciente ha mostrado que el nivel de participación electoral constituye una variable importante para explicar variaciones significativas tanto en los resultados electorales (Nagel, 1987; Rosenstone y Hansen, 1993) como en la naturaleza de las políticas gubernamentales (Meltzer y Richards, 1981; Pampel y Williamson, 1988; Hicks y Swank, 1992).

En los modelos sociológicos, la variación de la abstención se ha atribuido, a su vez, a diferencias entre factores estructurales, tales como la edad, hábitat, clase social y renta económica, que al afectar los recursos personales (y, en parte, el grado en que las decisiones políticas afectan el bienestar individual) de cada elector, modifican la probabilidad de que cada elector vote (Boix y Riba, 2000). Asimismo, la abstención puede ser interpretada como un síntoma de "normalización" en la vida política ya que puede ser considerada como un indicador de la satisfacción implícita del funcionamiento regular de las instituciones democráticas. En esta perspectiva, Magalhães (2001) señaló que *"… elevadas tasas de abstención no son incompatibles con las condiciones de una democracia representativa, siempre que resulten de un contexto de igualdad de oportunidades para la participación y permitan la competición entre elites políticas".*

En otro contexto, de acuerdo con Viegas y Faria (2004), *"… el nivel de abstención logrará ser un buen indicador de la desafectación de los ciudadanos con relación al sistema democrático, independientemente de las concepciones más liberales en desvalorizar este factor, invocando para eso, la experiencia de algunas democracias consolidadas con niveles más bajos de participación electoral, como los EE.UU. o Suiza".*

Frecuentemente y de particular importancia es también el grado de competitividad de una elección y la probabilidad de que ésta implique, o no, un diferente Gobierno mayoritario de partido único. Efectivamente, las elecciones muy competitivas que pueden resultar en cambios al nivel de la política gubernamental, motivarán seguramente tanto los que apoyan, como los opositores del Gobierno, a ejercer su derecho de voto y ciertamente estimularán esfuerzos de movilización de los indecisos. Estos factores, de acuerdo con Franklin (2003) *"… se prenden directamente con el acto electoral, y la participación electoral no es más que la suma de múltiples actos electorales".*

Según Freire y Magalhães (2002), la mayor parte de los individuos que se abstienen en determinadas coyunturas políticas, son electores con un elevado grado de integración social, lo que resulta una gran exigencia en el funcionamiento del sistema político. Estos individuos evalúan sus decisiones electorales de acuerdo con la naturaleza y la importancia de las elecciones, la percepción de la utilidad de su voto, el grado de diferenciación y polarización política entre los diferentes partidos, el desempeño anterior de los partidos, etc (Mayer y Perrineau, 1992; Subileau, 1997; Reto y Sá, 2000; Freire, 2000b, 2001a y 2001b).

El fenómeno de la abstención electoral no ha sido estudiado tanto como aquellos estudios que debaten el acto de votar ya que, de acuerdo con Santo (2006), *"… la abstención no siempre fue considerada un objeto con dignidad científica, pero si ha sido considerada desde una perspectiva de civismo/no civismo asociada a la participación política".*

El problema de la Inferencia Ecológica (IE)

De acuerdo con Langbein y Lichtman (1978), la IE, o la Regresión Ecológica, es un conjunto de técnicas estadísticas, que permiten estimar los valores de las celdas (datos desagregados), de una tabla de contingencia, cuando solamente se encuentran disponibles los totales marginales (datos agregados o datos ecológicos) de las tablas referidas.

En 1919, Ogburn y Goltra presentan la primera versión de la "discrepancia entre géneros". De acuerdo con Gow (1985), este artículo puede ser considerado como la primera publicación de técnicas de regresión en una revista de Ciencia Política, pero la relación entre las estimaciones de los coeficientes agregados y el comportamiento de los individuos, sólo ha sido considerada en los trabajos presentados por Miller (1952) y Goodman (1953).

El interés por la IE surge a partir de Robinson (1950), que redescubre el problema de la IE y advierte sobre la problemática de trabajar con datos agregados para inferir sobre relaciones individuales (la "falacia ecológica"). A partir del artículo de Robinson, la utilización de datos agregados, disminuyó substancialmente en relación a otro tipo de datos.

En los años 50 surgen dos trabajos de gran importancia; uno se debe a Goodman (1953, 1959) y el otro a Duncan y Davis (1953).

Goodman (1959) publica un artículo, donde trabaja esencialmente los procedimientos necesarios para la utilización de la regresión, y así constatar la diferencia entre los valores de los coeficientes de correlación de datos agregados y desagregados. Esta técnica pasó a la historia como Regresión Ecológica o Regresión de Goodman y ha pasado a ser el procedimiento estadístico más utilizado para hacer IE. Actualmente, los métodos estadísticos existentes para tratar este problema (IE) continúan de una manera o de otra, basándose en la Regresión de Goodman.

Duncan y Davis (1953), desarrollan otro método – el Método de los Límites - como respuesta al trabajo presentado por Goodman (1953). El método pretende deducir el intervalo de valores admisibles para las cantidades de interés una vez conocidos los totales marginales. Aunque sea una técnica que no ofrece estimaciones puntuales, sino intervalos posibles, es revelante y de mucho interés dado que, puede reducir significativamente la incertidumbre asociada al problema.

A finales del siglo XX, a partir del trabajo desarrollado por Gary King, presentado como "la solución" para el problema de IE, el interés por este tema ha vuelto a renacer, convirtiéndose en un tema de investigación abierta y sumamente interesante y prometedora. King (1997), ha propuesto una nueva modelización para el problema de estimación de datos desagregados a partir de datos agregados, y de acuerdo con varios ensayos empíricos, esta nueva metodología se presenta robusta ante al sesgo de la agregación.

2. JUSTIFICACIÓN Y OBJETIVOS DE LA INVESTIGACIÓN

Indudablemente, el impacto del fenómeno abstencionista en la dinámica de los regímenes democráticos de hoy, como es el caso de Portugal, es de tal modo polémico e interesante que nos incentiva a conocer, analizar, evaluar y, estimar datos desagregados. Lo haremos considerando los datos de las dos últimas elecciones legislativas disponibles en el momento de la realización de este trabajo: las elecciones de 2002 y las de 2005.

Los datos agregados siguen, todavía, siendo una fuente de información y, en algunos casos, la única. En el caso de los datos electorales, y dado el secreto del voto, constituyen una fuente de información con numerosas ventajas (Calvo y Escolar, 2003).

Por otro lado, Burden (2000) discute la no representatividad de los sondeos de opinión sobre la intención de voto, especialmente, después de una votación.

Wakefield (2004), resalta que cuando se trabaja con datos individuales, cómo son los sondeos, uno de los requisitos más importantes es la representatividad que ellos deben traducir, lo que no siempre ocurre en materia de sondeos políticos ya que los encuestados no siempre responden con sinceridad a la pregunta *¿Cómo votó usted en las últimas elecciones?".*

De acuerdo con Magalhães (2001), *"… en materia electoral, no sólo importa dar un comportamiento diferenciado a la identificación partidaria sino también a los abstencionistas".* Esta diferenciación de los abstencionistas hace inevitable una clasificación: los abstencionistas recurrentes, que corresponden a los individuos que se han abstenido en dos elecciones consecutivas y los abstencionistas ocasionales o fluctuantes, aquellos que habiendo votado en una de las elecciones, se han abstenido en la otra. Pero, teniendo en cuenta el secreto del voto, los listados electorales donde se registran los electores que han votado no están disponibles. Así, nunca será posible conocer si el elector X votó o no, en los dos momentos electorales considerados. Luego la única forma de conocer esa información, será a partir de Métodos de IE y atendiendo a la información que se encuentra disponible después de ocurridas las elecciones.

Aunque la información disponible se encuentre agregada al nivel del municipio, de acuerdo con Anselin (1988), O'Loughlin (2003) sostiene que el problema específico de trabajar con datos agregados con carácter geográfico, requiere un abordaje diferente en el modo de hacerse inferencia estadística.

De acuerdo con Calvo y Escolar (2003) *"… las dificultades para realizar una correcta estimación de datos agrupados geográficamente no son pocas, si bien estas dificultades han llevado a que el problema de la IE sea un área de fuerte desarrollo metodológico…"*.

King (1997) *"… a pesar de la pérdida de información proveniente de un proceso de agregación, utilizar información externa, sigue siendo la única forma de realizar IE, con razonable confianza".*

Asimismo, King (1997) mantiene que *"… si las unidades geográficas presentan características similares, es posible encontrar una constancia, en caso de que las unidades geográficas se agrupen en torno a una moda, aunque sea con una amplia varianza. Ésta es una de las motivaciones para efectuar el modelo basado en la distribución normal bivariante truncada".*

Schuessler (1999) defiende que la IE tiene uno de los más importantes problemas en las Ciencias Sociales. Los más recientes avances en la Inferencia Ecológica no han sido en términos estadísticos puros sino en términos de metodología estadística-determinista, es decir, una combinación entre la información conocida de la población y los métodos estadísticos de estimación.

Según O'Loughlin (2003), la metodología de King (1997) proporciona buenos resultados cuando las unidades geográficas presenten homogeneidad hacia la variable predictora (la abstención en este caso) y si las unidades geográficas presentan una desagregación lo más reducida posible. Manteniéndose estos requisitos y para efectos de desagregación, el autor

aconseja para la realización del procedimiento de estimación, un mínimo de 100 unidades geográficas.

Y, así este trabajo presenta aportaciones metodológicas para conseguir la máxima homogeneidad con respecto al variable objeto de estudio: la abstención.

A partir de los datos agregados, oficialmente disponibles en el Secretariado Técnico dos Assuntos para o Processo Eleitoral en Portugal /Ministério da Administração Interna (STAPE/ MAI), nos planteamos como objetivos los siguientes: 1)- Encontrar una caracterización de los tipos de abstencionistas teniendo como base las elecciones legislativas ocurridas en Portugal en 2002 y 2005, 2)- Evaluar el impacto de la abstención fluctuante en los diferentes perfiles de abstencionistas, a partir del modelo de IE de King, 3)- Estimar el porcentaje de electores recuperables dentro del conjunto de electores fluctuantes, en cada freguesia y, 4)- Estimar el porcentaje de electores fluctuantes ya recuperados, en cada freguesia.

Indudablemente, el impacto del fenómeno abstencionista en la dinámica de los regímenes democráticos de hoy, como es el caso de Portugal, es de tal modo polémico e interesante que nos incentiva a conocer, analizar, evaluar y, estimar datos desagregados. Lo haremos considerando los datos de las dos últimas elecciones legislativas disponibles en el momento de la realización de este trabajo: las elecciones de 2002 y las de 2005.

3. EL PROBLEMA DE INFERENCIA ECOLÓGICA (IE)

La IE es un método estadístico, que permite a partir de datos agregados, o sea, a partir de los totales marginales de una tabla de contingencia, estimar los valores de las celdas que no se encuentran disponibles: esto es, los valores desagregados (King, 1997).

Aunque este problema se aborda en contextos muy diferentes, uno de los campos con mayor aplicación es la Ciencia Política. En cada tabla de contingencia, los atributos en estudio, ya sean en filas o en columnas, pueden ser entendidos como variables dicotómicas, exhaustivas y mutuamente excluyentes.

El objetivo del problema de la IE es entonces, estimar las intersecciones fila-columna que dan origen a los totales marginales de fila y columna, respectivamente, de acuerdo con la Tabla 1.

Tabla 1. Notación para el Problema de IE

	Vota	**N**o Vota	Total
Negros (**b**)	β_i^b	$1-\beta_i^b$	X_i
Blancos (**w**)	β_i^w	$1-\beta_i^w$	$1-X_i$
Total	T_i	$1-T_i$	1

Fuente: King (1997, p.31)

Donde:

β_i^b - proporción de negros inscritos en los cuadernos electorales de la región i, que han votado;

β_i^w - proporción de blancos inscritos en los cuadernos electorales de la región i, que han votado;

$1-\beta_i^b$ - proporción de negros inscritos en los cuadernos electorales de la región i, que no han votado,

$1-\beta_i^w$ - proporción de blancos inscritos en los cuadernos electorales de la región i, que no han votado;

X_i - proporción de negros inscritos en los cuadernos electorales de la región i;

$1-X_i$ - proporción de blancos inscritos en los cuadernos electorales de la región i;

T_i - proporción de individuos inscritos en los cuadernos electorales de la región i, que han votado;

$1-T_i$ - proporción de individuos inscritos en los cuadernos electorales de la región i, que no han votado.

Así la ecuación que define la Identidad Básica, se presenta:

$$T_i = \beta_i^b \cdot X_i + \beta_i^w \cdot (1 - X_i) \qquad (1)$$

En la IE, los valores de T_i y X_i son conocidos y se pretende estimar los valores de los parámetros β_i^b y β_i^w, denominados en la bibliografía por cantidades de interés.

4. EL MODELO DE KING

La propuesta de King (1997) utiliza la información determinística contenida en los límites de Duncan y Davis (1953), y un modelo estadístico que permite realizar una estimación para los parámetros de interés, contenidos en esos límites. Como los límites son conocidos, este procedimiento suma una gran cantidad de información al modelo estadístico. Para ello, se mueve sucesivamente del nivel de distrito al de región y viceversa. Primero, al tratar cada distrito por separado, el método utiliza toda la información disponible, para dar un rango de posibles valores para estas cantidades de interés a nivel de distrito (Balsa, 2002). Así, el modelo estadístico "toma fuerza" de todos los datos de los otros distritos para dar la probable ubicación de cada verdadera cantidad de interés dentro de los conocidos límites determinísticos. Finalmente, se calculan, en base a las estimaciones de distrito, los parámetros regionales.

Consideremos la Tabla 1, de acuerdo con lo indicado anteriormente, el objetivo es encontrar los valores desconocidos de los parámetros β_i^b y β_i^w atendiendo a los totales marginales conocidos X_i y T_i, para todas y cada una de la tablas de contingencia, i = 1,…, p, de un área geográfica, así como para el área geográfica en general. Las cantidades de interés al nivel del área geográfica en general, son representados por β_i^b y β_i^w y no son más que las medias ponderadas provenientes de todos los β_i^b y β_i^w, que componen el área en estudio, de acuerdo con:

$$B^b = \frac{\sum_{i=1}^{p} N_i X_i \beta_i^b}{\sum_{i=1}^{p} N_i X_i} \quad \text{y} \quad B^w = \frac{\sum_{i=1}^{p} N_i (1-X_i) \beta_i^w}{\sum_{i=1}^{p} N_i (1-X_i)} \tag{2}$$

El Método de King, parte de dos aspectos centrales:

1)- Las variables desagregadas no observables β_i^b y β_i^w, son obtenidas a partir de una Distribución Normal Bivariante con Truncamiento (DNBT) sobre el cuadrado unitario $[0,1]\times[0,1] \in \Re^2$, atendiendo a que los valores de β_i^b y β_i^w sean proporciones, al revés de Goodman, que asume que estos parámetros son constantes a lo largo de todas las tablas $i = 1,\ldots,p$. King asume que los parámetros β_i^b y β_i^w, tienen algo en común, o sea, pueden variar, pero son parcialmente dependientes uno del otro (si uno aumenta, el otro tendrá que disminuir, atendiendo a los límites determinísticos);

2)- La Identidad Básica se verifica.

De acuerdo con King, la formulación completa del problema está compuesta por cuatro hipótesis:

H1) Las proporciones totales de las filas, representadas por X_i, son fijas. Esta hipótesis representa un tratamiento asimétrico de la información agregada. En efecto, las probabilidades agregadas de las filas, representadas por X_i (un valor para cada tabla) corresponden a información determinística y sus valores son dados *a priori*. En relación a las proporciones agregadas de las columnas, representadas por T_i (un valor para cada tabla) éstas son tratadas como una variable aleatoria.

H2) Las proporciones desagregadas, representadas por β_i^b y β_i^w siguen una Distribución Normal Bivariante Truncada (DNBT), sobre el cuadrado unitario $[0,1]\times[0,1] \in \Re^2$. Formalmente, la función de densidad de probabilidad viene definida por:

$$P\left(\beta_i^b, \beta_i^w\right) = TN\left(\beta_i^b, \beta_i^w \mid \mu, \Sigma\right) \tag{3}$$

Donde, TN representa el truncamiento de los límites para β_i^b y β_i^w.
El vector que representa las medias (μ) y la matriz de varianzas (Σ) vienen dados por:

$$\mu = \begin{pmatrix} \mu_b \\ \mu_w \end{pmatrix} \text{ con } \mu_b = E\left(\beta_i^b\right) \text{ y } \mu_w = E\left(\beta_i^w\right) \tag{4}$$

$$\Sigma = \begin{bmatrix} \sigma_b^2 & \sigma_{bw} \\ \sigma_{bw} & \sigma_w^2 \end{bmatrix}, \text{ respectivamente.} \tag{5}$$

El objetivo del modelo, es estimar β_i^b y β_i^w para todas y cada una de las p observaciones. Así, los parámetros μ y Σ no tienen un interés directo y son utilizados en análisis para entender los resultados intermedios, hasta llegar a los valores estimados para β_i^b y β_i^w.

La distribución NBT tiene una única moda, que indica donde se encuentran la mayor parte de los valores de β_i^b y β_i^w y también el grado de variación de cada uno de ellos en torno de la moda. La distribución también permite que β_i^b y β_i^w tengan un grado de correlación entre -1 y 1.
Así, la DNBT, con truncamiento sobre el cuadrado unitario $[0,1]\times[0,1]\in\Re^2$, es definida por:

$$f\left(\beta_i^b,\beta_i^w \mid \Psi\right)=\frac{1}{2\pi\sigma_b\sigma_w\sqrt{1-\rho^2}R(\Psi)}\exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}\left\{\frac{\left(\beta_i^b-\mu_b\right)^2}{\sigma_b^2}-2\rho\frac{\left(\beta_i^b-\mu_b\right)}{\sigma_b}\frac{\left(\beta_i^w-\mu_w\right)}{\sigma_w}+\frac{\left(\beta_i^w-\mu_w\right)^2}{\sigma_w^2}\right\}\right] \quad (6)$$

donde:

$\Psi=\left[\mu_b,\mu_w,\sigma_b,\sigma_w,\rho\right]^T$ es el vector de parámetros, y

$R(\Psi)$ es el vector de normalización, que hace que $f\left(\beta_i^b,\beta_i^w \mid \Psi\right)$ se integre en el cuadrado $[0,1]\times[0,1]\in\Re^2$.

Así:

$$R(\psi)=R(\mu,\Sigma)=\int_0^1\int_0^1 N(\beta^b\beta^w \mid \mu,\Sigma)d\beta^b d\beta^w$$

Los valores admisibles para β_i^b y β_i^w, son dados por:

$$f\left(\beta_i^b,\beta_i^w \mid \Psi\right)=\begin{cases}(\beta_i^b,\beta_i^w)^T\in[0,1]\times[0,1]\\ 0\,,o.c.\end{cases}$$

Las medias de la DNBT pueden ser estimadas a partir del valor esperado de T_i dado X_i, o sea $E\left(T_i \mid X_i\right)$, a partir de una regresión. Se parte de discriminar los parámetros de las regiones en el parámetro de área geográfica y un término de error:

$$\beta_i^b=\mu^b+\varepsilon_i^b \text{ y } \beta_i^w=\mu^w+\varepsilon_i^w \quad (7)$$

Donde, los términos ε_i^b y ε_i^w, tienen media cero, porque ellos están definidos como los desvíos de sus correspondientes valores condicionales, o sea, $E\left(\varepsilon_i \mid X_i\right)=0$.
De acuerdo con el procedimiento de Goodman (1959), sustituyendo el segundo miembro de la ecuación que define la Identidad Básica, obtenemos:

$$T_i=\beta_i^b\cdot X_i+\beta_i^w.(1-X_i)$$

$$T_i=\left(\mu_i^b+\varepsilon_i^b\right)X_i+\left(\mu_i^w+\varepsilon_i^w\right)\left(1-X_i\right)$$

$$T_i=\mu^b X_i+\mu^w(1-X_i)+\varepsilon_i \quad (8)$$

con,

$$\varepsilon_i=\varepsilon_i^b X_i+\varepsilon_i^w\left(1-X_i\right) \quad (9)$$

H3) Existe independencia en la media entre β_i^b, β_i^w y X_i, lo que implica:

$$\beta_i^b = \mu_b + \varepsilon_i^b \tag{10}$$

$$\beta_i^w = \mu_w + \varepsilon_i^w \tag{11}$$

Esta hipótesis implica que β_i^b y β_i^w varían aleatoriamente en torno a medias constantes y por tanto son independientes de X_i. En la realidad, las constantes μ_b y μ_w en las ecuaciones 10 y 11 son las medias efectivas de β_i^b y β_i^w lo que es una hipótesis más débil que la hipótesis de que β_i^b y β_i^w sean estocásticamente independientes: $\rho_{bw} = 0$ (Goldberger, 1991, p.61).
Esta hipótesis es equivalente a asumir la inexistencia de sesgo en la agregación, lo que es necesario para obtener estimaciones consistentes para μ y $\sum$, aunque esta hipótesis no siempre sea necesaria para obtener estimaciones precisas para las cantidades de interés (King, 1997, p.94).

H4) $T_i|X_i$ es independiente de $T_j|X_j$, para $i \neq j$.
Esto significa que el comportamiento de los totales columna es independiente entre todas las tablas. Esta hipótesis también es conocida como "independencia espacial" **(Cho, 1998).** De acuerdo con King (1997, p.94) la violación de esta hipótesis no representa grandes consecuencias para el modelo.

A partir de estas cuatro hipótesis, el Modelo de King (1997) es implementado en dos etapas:

1ª etapa: Estimación de los parámetros

Se estima mediante máxima verosimilitud el vector de parámetros $\breve{\Psi} = \left[\breve{\mu}_b, \breve{\mu}_w, \breve{\sigma}_b, \breve{\sigma}_w, \breve{\rho}\right]^T$, correspondientes a una DNB sin truncamiento (notación $\cup$ encima de los parámetros). Para construir la función de verosimilitud, King deduce la distribución de probabilidades para la proporción agregada T_i o sea $T_i|X_i$. Combinando H1) y H2) con $T_i = \beta_i^b X_i + \beta_i^w (1 - X_i)$ (Identidad Básica) es posible deerminar la distribución de T_i. A través de la Identidad Básica se verifica que T_i se relaciona con β_i^b y β_i^w de forma lineal.

2ª etapa: Estimación de las cantidades de interés mediante simulación

King utiliza una rutina especializada, en lenguaje GAUSS, para la maximización de la función de verosimilitud bajo restricciones, a través de métodos iterativos del tipo cuasi-Newton. Esta rutina es una aplicación en lenguaje GAUSS denominada *Constrained Maximum Likelihhod* – CML, desarrollada por Schoenberg (1997a, 1997b).
La estimación de la matriz de varianzas-covarianzas se denotará $V(\hat{\phi})$. Si P es un número grande, los resultados asintóticos permiten utilizar una distribución normal para la *a posteriori* (Tanner, 1996):

$$\boldsymbol{\phi \mid T \approx N(\hat{\phi}; V(\hat{\phi}))} \tag{12}$$

Así, el método parte de la información determinística obtenida a partir del Método de los Límites de Duncan y Davis (1953), que establece cual es el intervalo de valores admisibles, para los parámetros a estimar, es decir:

$$\hat{\beta}_i^b \in \left[\max\left(0, \frac{T_i - (1 - X_i)}{X_i}\right), \min\left(\frac{T_i}{X_i}, 1\right)\right]$$

y (13)

$$\hat{\beta}_i^w \in \left[\max\left(0, \frac{T_i - X_i}{1 - X_i}\right), \min\left(\frac{T_i}{1 - X_i}, 1\right)\right]$$

y tiene como objetivo extraer información determininística a partir de información conocida sobre los datos.

A continuación, el Método de King, tiene en cuenta la Identidad Básica y establece una relación lineal entre las cantidades de interés, definida por:

$$\hat{\beta}_i^w = \frac{T_i}{1 - X_i} - \left(\frac{X_i}{1 - X_i}\right)\hat{\beta}_i^b \quad (14)$$

A partir de esta relación lineal, se pasa a saber que los valores de los parámetros se encuentran en una recta delineada en el cuadrado unitario.

El Método de King se puede aplicar en el programa informático EzI - *A(n Easy) Program for Ecological Inference*, desarrollado por Benoit y King en 2002. Es un programa ejecutable en ambiente *Windows* y no requiere cualquier otro *software* para funcionar. Su versión para *Windows* está disponible gratuitamente en la dirección: http://gking.harvard.edu/.

5. LA NECESIDAD DE HOMOGENEIDAD ESPACIAL

De acuerdo con Ansolabehere y Rivers (1995), cuando se estima las transferencias de votos entre dos elecciones consecutivas, dicotómicas, sin recurso a encuestas, la IE produce buenos resultados en regiones homogéneas en su actitud electoral. Pero, cuando existe heterogeneidad en la actitud electoral, la IE se puede tornar ambigua.

Según Cho (2001), la existencia de constancia a lo largo de las unidades espaciales desagregadas, es decir, la homogeneidad espacial, es una característica deseable para generar buenos resultados para las estimativas de las cantidades de interés del modelo de King.

Un dato espacial puede ser definido como la observación de una variable asociada a una localización del espacio geográfico. Cuando se tienen observaciones georeferenciadas, se deben utilizar herramientas que permitan detectar ciertas características dentro de los datos, como la tendencia, esquemas de asociación, entre otros (Bohórquez y Ceballos, 2008).

Ya la heterogeneidad espacial, propuesta inicialmente por Anselin (1988), se trata de un efecto relacionado con la diferenciación espacial o regional y viene definido por la ausencia de estabilidad en el espacio del comportamiento humano o de otras relaciones en estudio. Esto implica que en los modelos espaciales, las formas funcionales y los parámetros varían con la localización geográfica, no siendo homogéneos para toda la matriz de datos (Chasco, 2004). Este efecto espacial suele estar directamente relacionado con la localización geográfica, el área o cualquier otra característica de las unidades espaciales amostrales (Anselin, 1988; Moreno y Vayá, 2000; Chasco, 2003). Las causas de tipo sociológico que pueden llevar a la heterogeneidad espacial, es que las relaciones en sí pueden ser diferentes a través del espacio, por existir variaciones espaciales en las actitudes o preferencias de la población o existirán asuntos administrativos o políticos o de otros contextos, que producen respuestas diferentes ante el mismo estímulo (Haining, 2003). Es decir, la heterogeneidad espacial surge cuando se trabaja con unidades espaciales (países, regiones, municipios) en las que el fenómeno se

distribuye de maneras distintas sobre el espacio, como es el caso, de la tendencia partidarista en Portugal.

A diferencia del caso de la dependencia espacial, los problemas causados por la heterogeneidad espacial podrán ser resueltos mediante procedimientos multivariantes estándar, como el análisis *cluster* (Chasco, 2004), pero, sin embargo, la estructura inicial de los datos puede llevar a procedimientos más eficientes (Bohórquez y Ceballos, 2008).

Con relación a los resultados electorales en Portugal, Caleiro (2008) reconoce de una manera general, la existencia de regiones geográficas donde la influencia de los principales partidos políticos se torna evidente. Así, un análisis de los resultados electorales exige que se tome en cuenta la componente espacial de los mismos. De acuerdo con su estudio, Caleiro (2008) afirma que: "*la componente espacial no puede ser ignorada en la explicación de los resultados electorales en Portugal*".

6. LOS MÉTODOS BIPLOT COMO HERRAMIENTA DE BÚSQUEDA DE HOMOGENEIDAD ESPACIAL

El éxito de la IE radica, entre otras cosas, en el hecho de conseguir que las unidades básicas de análisis que componen la unidad geográfica de orden superior sean homogéneas con respecto a la variable objeto de estudio.

Por esta razón proponemos como paso previo, a la aplicación del modelo de King, aplicar los métodos BIPLOT a la matriz de datos que continen información sobre el porcentaje de votos a los diferentes partidos políticos, en todas y cada una de las áreas consideradas. De esta manera cada área vendrá representada como un punto en un plano factorial en el que los diferentes partidos votados juegan el papel de variables explicativas. Así, unidades (concelho, freguesia, etc) con la misma tendencia política (en términos mayoritarios, por supuesto) ocuparán posiciones próximas en el plano factorial con máxima capacidad informativa, en relación a las variables de interés (partidos votados y/o abstenciones). Esos *clusters* de puntos próximos forman unidades homogéneas, en relación a las tendencias políticas, y serán la base sobre la que se aplicarán los modelos de IE.

7. EL MÉTODO HJ-BIPLOT

Un HJ-Biplot (Galindo, 1986) para una matriz de datos X que contiene, en columnas, información sobre el porcentaje de votos conseguidos por los diferentes partidos políticos, y en filas la distribución de votos en una unidad geográfica (concelho, por ejemplo), se define como una representación gráfica multivariante mediante marcadores j_1, j_2,…,j_n para las filas (freguesias, concelhos, distritos, etc) y h_1, h_2,…,h_p para las columnas de X, (partidos del espectro político), elegidos de forma que ambos marcadores puedan superponerse en el mismo sistema de referencia con máxima calidad de representación. Generalmente las filas se representan como puntos y las columnas como vectores.

El HJ-Biplot, se basa en la descomposición en valores singulares (DVS) de la matriz de datos. Cualquier matriz real $X_{(nxp)}$ de rango r ($r \leq \min(n, p)$) se puede factorizar como el producto de tres matrices del tal forma:

$$X_{(nxp)} = U_{(nxr)} \Lambda_{(rxr)} V'_{(rxp)} \text{ con } U'U = V'V = I_r \qquad (15)$$

donde: $U_{(nxr)}$ es la matriz de vectores propios de X', $V_{(pxr)}$ es la matriz de vectores propios de $X'X$ y, $\Lambda_{(rxr)}$ es una matriz diagonal de $\lambda_1, \lambda_2, ..., \lambda_r$, correspondientes a los r

valores propios de $\boldsymbol{X}^{'}$o $X^{'}X$.

Los elementos de $X_{(nxp)}$ en 15 vienen dados por:

$$x_{ij} = \sum_{k=1}^{r} \sqrt{\lambda_k}\, u_k\, v_{jk} \quad i = 1,2,...,n, \quad j = 1,2,...,p \tag{16}$$

Así, partiendo de la DVS, la elección de marcadores en la dimensión q para las filas y columnas de la matriz X son:

$$J_{(q)} = U_{(q)}\Lambda_{(q)} \text{ y } H_{(q)} = V_{(q)}\Lambda_{(q)} \tag{17}$$

La calidad de representación para las filas y para las columnas de la matriz de datos X es la misma y las filas y columnas vienen expresadas en coordenadas principales.

8. PREPARACIÓN DE LOS DATOS Y REORGANIZACIÓN TERRITORIAL SEGÚN LOS MÉTODOS BIPLOT

A partir de los datos disponibles por el STAPE/MAI, partimos de una matriz de 308 filas por 16 columnas. En cada fila está representado cada uno de los concelhos de Portugal y en las columnas tenemos las variables con respecto al comportamiento electoral de los portugueses en las referidas elecciones. Las variables elegidas son: la ABSTENCIÓN, los votos BLANCOS y NULOS (conjuntamente), los partidos políticos mayoritarios, descritos por orden alfabético: Bloco de Izquierda (BE); Centro Democrático Social-Partido Popular (CDS-PP); Coalición Democrática Unitaria CDU (PCP-PEV); Partido Popular Democrático/ Partido Social Demócrata (PPD/PSD) y Partido Socialista (PS). Consideramos también, todos los partidos minoritarios conjuntamente, atendiendo a su baja representatividad, y a los que denominamos como OTROS. Los valores de la matriz de datos se presentan como porcentajes, ya que intentamos buscar un comportamiento de votación de acuerdo con la dimensión electoral de los municipios, es decir con el número de electores que pertenecen a los listados electorales[1].

En principio podría pensarse en conseguir datos desagregados para las unidades territoriales oficialmente establecidas, a nivel de distrito, a nivel de concelho, o a nivel de freguesia, por ejemplo; sin embargo las estimaciones con los modelos de IE fracasaron, llegando en algún caso hasta el punto de que el procedimiento informático fracasaba. Sin duda debido a la heterogeneidad partidista de las unidades geográficamente establecidas.

De acuerdo con la bibliografía especializada en el campo de la IE, procederemos en una primera etapa, a una "reorganización" territorial, al nivel del concelho, basada en la tendencia partidarista y/o abstencionista, utilizando el método HJ-BIPLOT[2] (Galindo, 1986).

Intentamos así, reducir la heterogeneidad de los comportamientos electorales y obtener grupos homogéneos en su actitud, teniendo por base la abstención y así obtener la información acerca de los porcentajes de abstencionistas ocasionales, al nivel de la freguesia. Este tratamiento preliminar de las áreas geográficas portuguesas, con un enfoque estadístico-determinístico, pretende detectar concelhos

[1] Para el análisis se consideraron sólo 306 concelhos atendiendo a que en los concelhos de Germil (distrito de Viana do Castelo) y Souto (distrito de Guarda), no había datos para las elecciones de 2005. Los datos que contiene la matriz, objeto de estudio, son los valores porcentuales obtenidos por cada partido político en las elecciones legislativas de 2002 y 2005, convenientemente estandarizados por columnas.

[2] Utilizamos el programa informático Multbiplot de Vicente (2007) para pasar el HJ-BIPLOT. Esto programa se encuentra disponible en la dirección electrónica http://biplot.usal.es.

con características similares con relación a las dos votaciones en estudio, es decir, áreas que se agruparán en torno de una moda qué, en nuestro caso, es representada por los partidos políticos más votados.

La Tabla 1 recoge una cuantificación de la absorción de inercia; es decir, contiene la cantidad de información contenida en los datos que es recogida por los tres primeros ejes factoriales. Los tres primeros ejes explican el 66,79% de la información contenida en los datos, consiguiéndose el 55,92% en el primer plano principal.

Tabla 1. Absorción de Inercia

Valor propio	Inercia	Inercia Acumulada
43.928	39.158	39.158
28.737	16.758	**55.915**
23.15	10.875	**66.79**

Considerando las coordenadas de los puntos que representan a los diferentes concelhos sobre los ejes factoriales, es posible establecer *cluster* de puntos (concelhos) con tendencias partidistas similares y conocer además cuales son sus preferencias partidistas. También es posible conocer la posición de los abstencionistas en el espectro político del País. Como se han representado los datos de dos elecciones consecutivas, el ángulo que forman los vectores que identifican a cada partido político, en cada año, nos permite estudiar la estabilidad temporal para cada partido y el cambio del grupo Abstencionista.

El Eje 1 está altamente correlacionado con los partidos políticos mayoritarios, en cualquiera de las dos elecciones. En el Eje 2 se encuentran las variables relativas a la Abstención y a los Votos Blancos y Nulos, también en cada una de las elecciones consideradas. El Eje3 está caracterizado por el bloque OTROS partidos. Esto era esperable ya que al ser minoritarios tienen menor variabilidad (en términos absolutos) y por eso carga en un eje con menor absorción de inercia. En el resto del análisis, el Eje 3 no es analizado atendiendo a no ser objeto de estudio las votaciones en los partidos minoritarios sin representación en la Asamblea de la República[3]. Los resultados obtenidos se presentan en la Tabla 2.

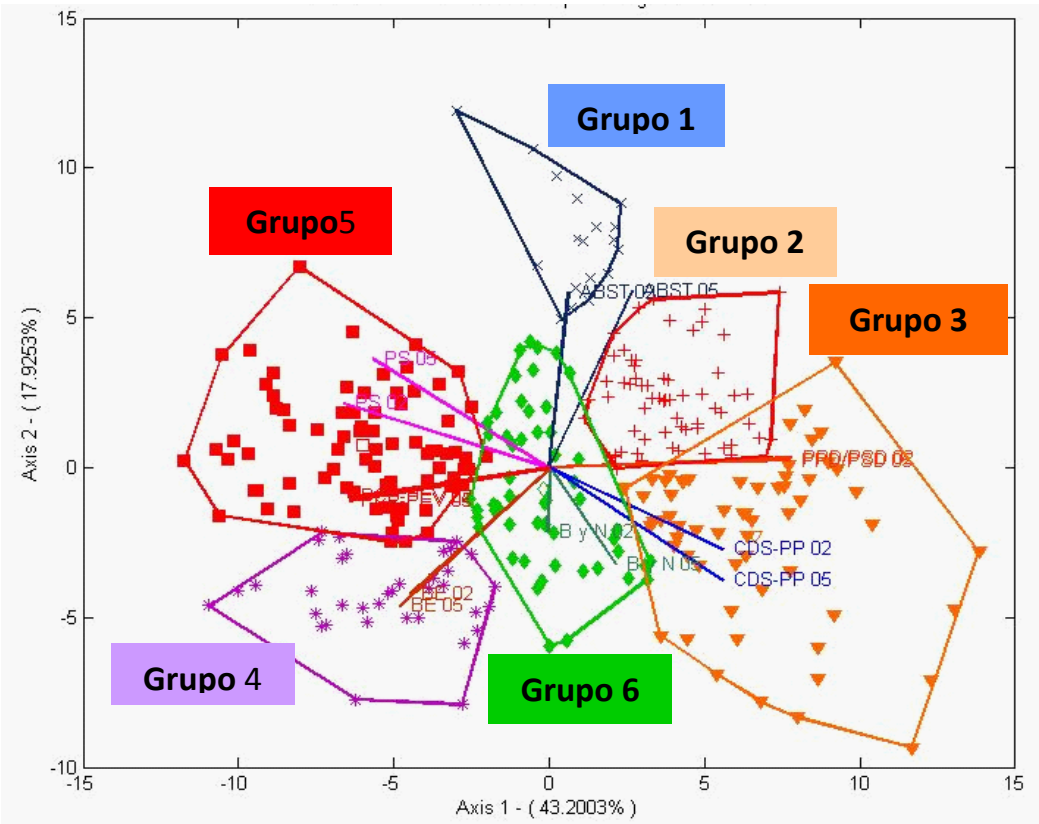
Tabla 2. Contribuciones relativas del factor al elemento

FACTOR	EJE 1	EJE 2	EJE 3
ABST 05	114	**378**	247
B y N 05	66	**208**	3
PPD/PSD 05	**914**	0	0
PS 05	**426**	244	66
BE 05	**373**	233	47
CDS-PP 05	**412**	286	1
PCP-PEV 05	**675**	17	74
OTROS 05	147	130	**348**
ABST 02	5	**397**	288
B y N 02	2	**151**	86
PPD/PSD 02	**918**	2	0
PS 02	**602**	131	109
BE 02	**302**	196	56
CDS-PP 02	**346**	221	23
PCP-PEV 02	**673**	13	71
OTROS 02	289	73	**322**

[3] También eliminamos el municipio Corvo (Islas de los Azores) atendiendo a ser un *outlier* y distorsionar los resultados obtenidos.

A la vista de las contribuciones, el Eje 1, puede ser entendido como la actitud de voto útil y el Eje 2 pasa a presentar la actitud de voto sin utilidad. Estos dos son, pues, los dos gradientes actitudinales recogidos en el estudio. Para visualizar mejor el comportamiento de los resultados electorales, pasamos a presentar la representación bidimensional HJ-BIPLOT de máxima capacidad informativa, en el Gráfico 1.

Gráfico 1. Representación bidimensional HJ-BIPLOT para 6 grupos



Aunque de una manera general, todos los municipios portugueses han cambiado de comportamiento electoral, entendemos de interés percibir si existen comportamientos semejantes y/o diferenciados entre los municipios, que nos permitan caracterizarlos de acuerdo a los resultados electorales de 2002 y 2005.

Así, con relación al Eje 1 – eje de la actitud de voto útil, verificamos que entre los partidos de derecha (CDS-PP) y centro-derecha (PPD/PSD), sí encontramos una correlación positiva entre los votos obtenidos en cualquiera de las elecciones, es decir, una transferencia de votos. También esta relación se verifica entre los partidos de izquierdas (BE y PCP-PEV), así como entre el PCP-PEV y el partido de centro-izquierda PS, aunque los resultados electorales entre el BE y el PS se encuentran poco correlacionados en relación a la elección de 2002 y prácticamente no correlacionados en la elección de 2005.

Observamos en relación al Eje 1 dos actitudes perfectamente distintas, sobre la tendencia política. En la parte izquierda del eje se encuentran los partidos políticos relacionados con la izquierda (PCP-PEV, BE) y centro-izquierda (PS) y en la parte derecha del eje los partidos políticos asociados con la derecha (CDS-PP) y centro-derecha (PPD/PSD).

También se observa que, tanto el PCP-PEV como el BE y el PPD/PSD, tienen las votaciones de 2002 y 2005 superpuestas y perpendiculares al Eje 2. Atendiendo a que el PPD/PSD ha tenido una gran derrota electoral, los vectores superpuestos indican la estabilidad electoral en las dos elecciones, o sea, el comportamiento de los votantes que ha permitido la elección

de los 72 disputados por el partido. En relación a colinealidad y longitud de los vectores correspondientes a las dos votaciones obtenidas por el PCP-PEV, pone de manifiesto la actitud política de los votantes de un partido de izquierdas conservadora, o sea, la fidelización al voto. Además, aunque el BE presente también los dos vectores superpuestos -fidelización de votantes- verificamos que el vector con respecto a la votación de 2005 presenta mayor longitud, lo que se traduce, en un aumento del número de votos obtenidos en la segunda elección.

Pasando a analizar los partidos que presentan un movimiento explícito en el Gráfico 1 empezamos en primer lugar por el PS, el partido que ha aumentado en todos los municipios, el número de votos. Verificamos que el PS se encuentra en el 2º cuadrante del gráfico y presenta un movimiento de la izquierda para la derecha, que atendiendo al valor del coseno en este cuadrante, significa un aumento del número de votos, que se ha traducido en un aumento del 25% de los diputados elegidos, o sea, de 96 diputados en 2002 a 120 diputados en 2005.

El CDS-PP también presenta un movimiento explícito de la primera a la segunda votación. Pero, como se encuentra en el 4º cuadrante y el movimiento es de la derecha para la izquierda, resulta que, atendiendo a los valores del coseno en este cuadrante, ha perdido votos entre la primera y la segunda elección, lo que representó la disminución de 2 diputados elegidos a nivel nacional.

Con relación al Eje 2 - eje de la actitud de voto no útil, es importante señalar, que el movimiento de la abstención a la derecha, en el primer cuadrante del Gráfico 1 ha penalizado los partidos CDS-PP y PPD/PSD, este último en especial. Esta actitud pone de manifiesto, que el aumento de votos elegibles ha beneficiado los partidos influenciados con la izquierda y centro-izquierda y fue utilizado como un factor de penalización sobre los partidos de derecha y centro-derecha. También el movimiento de votos Blancos y Nulos, de la izquierda para la derecha, en el cuarto cuadrante, sugiere una penalización para estos partidos, o sea, algunos de estos votantes han preferido no votar, en vez de cambiar su sentido del voto.

Los grupos encontrados se pueden caracterizar de la seguinte forma:

Grupo 1 – Tendencia abstencionista muy elevada. Este grupo está formado por 17 concelhos, mayoritariamente pertenecientes al Archipiélago de los Azores. Los 17 concelhos representan en su total 187 freguesias. Aunque sea un grupo de tendencia abstencionista muy elevada, presenta una tendencia partidarista de centroizquierda.

Grupo 2 – Actitud abstencionista en la tendencia centroderecha. Grupo formado por 62 concelhos, que en su total representan 1145 freguesias. De los 62 concelhos, 56 se sitúan en el Norte de Portugal, 5 en las islas de los Azores y 1 en la isla de la Madeira. Con respecto al distrito de Bragança referimos que 11 de sus 12 concelhos se encuentran en este grupo, así como 9 concelhos de los 14 del distrito de la Guarda y 13 de los 24 concelhos del distrito de Viseu. Este es un grupo de Tendencia Centroderecha, y alto nivel de abstención.

Grupo 3 – Tendencia de centroderecha y derecha. Este grupo está formado por 61 concelhos, de los cuales 38 se sitúan al Norte, 14 al Centro y 9 en el Archipiélago de la Madeira. Los 61 concelhos representan en su total 876 freguesias.Con respecto al distrito de Aveiro se encuentran en este grupo, 12 de sus 19 concelhos, así como 9 de los 16 concelhos del distrito de Leiria , 10 de los 24 concelhos del distrito de Viseu y 9 de los 11 concelhos del Archipiélago de la Madeira.Es un grupo asociado al centroderecha (PPD/PSD) y derecha (CDS-PP).

Grupo 4 – Tendencia de izquierda tradicional y nueva izquierda. El grupo está formado por 36 concelhos, donde destacamos 8 de los 16 concelhos del distrito de Lisboa (código

11), 6 de los 18 concelhos del distrito de Porto (código 13) y 9 de los 13 concelhos del distrito de Setúbal (Código 15). Los 36 concelhos representan 451 freguesias. Aunque sea un grupo caracterizado por una tendencia de izquierda tradicional (PCP-PEV) es el crecimiento de la nueva izquierda (BE), asociada a los grandes centros urbanos lo que más se destaca en este grupo.

Grupo 5 – Tendencia de centroizquierda e izquierda tradicional. Este grupo está formado por 77 concelhos, donde destacamos que, todos los concelhos de los distritos de Beja y Évora están incluidos. También pertenecen a este grupo, 13 de los 15 concelhos del distrito de Portalegre, 10 de los 21 concelhos del distrito de Santarém y 5 de los 11 concelhos del distrito de Castelo Branco. Los 77 concelhos representan 568 freguesias. Es un grupo de clara tendencia de izquierda tradicionalista (PCP-BE) y centroizquierda (PS).

Grupo 6 – Comportamiento mixto con tendencia centrista. Este grupo está formado por 54 concelhos, donde consideramos que 25 de ellos no se encuentran bien representados en el Plano 1-2. Son elementos que tienen mejor representación en los Planos 1-3 o 2-3. Los 54 concelhos representan 1034 freguesias. Atendiendo a que la información contenida en el Eje 3, donde se encuentran las votaciones de OTROS 02 y OTROS 05, no es objeto de estudio, pasamos a analizar los resultados referentes al Plan 1-2. Es un grupo de tendencias mixtas, pero predomina la tendencia centrista, ya sea de centroizquierda (PS) o de centroderecha (PPD/PSD). En este grupo, se encuentran concelhos de prácticamente todos los distritos, con la excepción de los distritos de Beja, Bragança, Évora, Setúbal y los Archipiélagos de los Azores y Madeira. Dada la tendencia mixta deste grupo, lo hemos dividido en 2 nuevos grupos: el grupo 6A formado por las 469 freguesias con menos de 1000 electores inscritos y el grupo 6B, formado por las 565 freguesias de 1000 o más electores inscritos.

9. LOS COEFICIENTES OBTENIDOS EN LA INFERENCIA ECOLÓGICA, POR GRUPO

A partir del Método de King (1997) para Inferencia Ecológica obtenemos los coeficientes de estabilidad con respecto al voto electoral (votar o abstenerse consecutivamente en dos elecciones), así como los coeficientes de inestabilidad con respecto al voto electoral (votar en una de las elecciones y abstenerse en la otra), por grupo encontrado. Con relación a los coeficientes de inestabilidad electoral, pasamos a diferenciar entre electores ocasionales a recuperar, electores que han votado en 2002 pero que se han abstenido en 2005, y electores ocasionales ya recuperados, aquellos que se han abstenido en 2002 y pasaran a votar en 2005.

Para todos y cada un de los grupos, presentamos en la tabla 3 las estimaciones, B^b y B^w, para un nivel de confianza del 95%, que nos permiten conocer los coeficientes de estabilidad de actitud electoral (votar 2002/votar 2005 y abstener 2002/abstener 2005), así como los coeficientes de inestabilidad de actitud electoral (votar 2002/abstener 2005 y abstener 2002/votar 2005).

Tabla 3. Coeficientes obtenidos en la Inferencia Ecológica

COEFICIENTES	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	
	ABSTENCIÓN	ABSTENCIÓN y PSD	PSD y CDS-PP	CDU y BE	PS y CDU	MIXTO <1000	MIXTO ≥ 1000
Estabilidad Electoral – votantes	0,9471	0,9452	0,9682	0,9816	0,9692	0,9665	0,9833
Inestabilidad electoral-Abstencionistas ocasionales a recuperar	0,0529	0,0548	0,0318	0,0184	0,0308	0,0335	0,0167
Inestabilidad electoral-Abstencionistas ocasionales ya recuperados	0,0667	0,0801	0,0906	0,1264	0,1456	0,0667	0,1087
Estabilidad Electoral – abstencionistas recurrentes	0,9333	0,9199	0,9333	0,8734	0,8544	0,8933	0,8913

10. CONCLUSIONES[4]

Con relación al coeficiente de estabilidad electoral respecto a la actitud de votar en las dos elecciones, sus valores más bajos se encuentran en los grupos afectados a la abstención. Por otro lado, donde se encuentra mayor estabilidad es en el grupo de tendencia mezclada y con número de electores igual o superior a 1000. El Grupo que presenta el segundo valor más elevado es el grupo 4, o sea, el Grupo representativo de la izquierda tradicional.

Con respecto al coeficiente de inestabilidad electoral-abstencionista fluctuante a recuperar-sus valores más altos también se encuentran en los dos grupos afectados a la abstención, grupo1 y 2. Los grupos que presentan los coeficientes de recuperación de abstencionistas más bajos, son nuevamente, el primer el grupo de tendencia mezclada con 1000 o más electores y el grupo de la izquierda tradicional.

Los valores más elevados del coeficiente de inestabilidad electoral-abstencionista fluctuantes ya recuperados- están relacionados con el grupo de izquierda tradicional (grupo 5) y con la izquierda tradicional/moderna (grupo 4). Los valores más bajos de este coeficiente se encuentran en los grupos relativos a la abstención (grupo 1) y al grupo de tendencia mezclada con menos de 1000 electores.

Finalmente, con relación al coeficiente de estabilidad electoral respecto a la actitud de abstenerse en las dos elecciones, los valores más elevados están relacionados con el grupo 1, el grupo de la abstención, y de igual modo con el grupo 3, el grupo de la derecha y centroderecha. El segundo valor más elevado está relacionado con el grupo 2, de la abstención sobre la derecha.

[4] Los resultados obtenidos, al nivel de todas y cada una de las freguesias que componen cada concelho de todos y cada uno de los *Clusters*, bien como toda la metodología desarrollada, se encuentran disponibles en Castela, E.(2009). *Inferencia Ecológica para la caracterización de abstencionistas: el caso de Portugal*. Tesis Doctoral. Departamento de Estadística. Universidad de Salamanca.

BIBLIOGRAFÍA

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic.

Ansolabehere, S. y Rivers, D. (1995). *Bias in Ecological Regression Estimates*. Working Paper. Stanford University.

Balsa, J. (2002). "Ventajas y limitaciones de la metodología de Inferencia Ecológica propuesta por G. King - Aplicaciones al análisis del triunfo de Péron en las elecciones presidenciales argentinas de 1946". *Cinta de Moébio*, 13.

Bohórquez, I. y Ceballos, E. (2008). "Algunos conceptos de la econometría especial y el análisis exploratorio de datos espaciales". *Ecos de Economía*, Medellín, 27, pp.9-34.

Boix, C. y Riba, C. (2000). "Las bases sociales y políticas de la abstención en las elecciones generales españolas: recursos individuales, movilización estratégica e instituciones electorales". *Revista Española de Investigaciones Sociológicas*, 90, pp. 79-128.

Burden, B. C. (2000). "Voter turnout and the national election studies". *Political Analysis*, 8, pp.389-398.

Caleiro, A. (2008). *Para uma visão espacial dos resultados eleitorais em Portugal*. Documento de Trabajo nº 2008/01. Universidad de Évora. Departamento de Economía.

Calvo, E. y Escolar, M. (2003). "The Local Voter: A Geographically Weighted Approach to Ecological Inference." *American Journal of Political Science*, 47, 1, pp.189-204.

Chasco, C. (2003). *Econometría espacial aplicada a la predicción-extrapolación de datos espaciales*. Ed. Comunidad de Madrid. Madrid.

Chasco, C. (2004). "Modelos de heterogeneidad espacial". *Econometrics*. 0411004, EconWPA.

Cho, W. K. T. (1998). "If the Assumption Fits … A Comment on the King Inference Solution". *Political Analysis,* **7** (1), pp. 143-163.

Cho, W. K. T. (2001). "Latent Groups and Cross-Level Inferences". *Electoral Studies*, 20, 2, pp.243-263.

Duncan, O. D. y Davis, B. (1953). "An Alternative to ecological correlation". *American Sociological Review*, 18, pp.665-666.

Franklin, M. N. (2003). "Os enigmas da participação eleitoral". *Análise Social*, XXXVIII (167), pp.321-338.

Freire, A. (2000b). "Clivagens, conjuntura económica e comportamento eleitoral: uma análise das legislativas portuguesas de 1995 com dados agregados", *Sociologia, Problemas e Práticas*, 32, pp.23-54.

Freire, A. (2001a). *Modelos do comportamento eleitoral: Uma Breve Introdução Crítica.* Oeiras: Celta Editora.

Freire, A. (2001b). *Mudança Eleitoral em Portugal Continental: Clivagens, Economia e Voto em Eleições Legislativas, 1983-1999*. Oeiras: Celta Editora.

Freire, A. y Lobo, M. C. (2002). "The Portuguese 2002 Legislative Elections". *West European Politics*,. 25, 4, pp.221-228.

Freire, A. y Magalhães, P. (2002). *A abstenção eleitoral em Portugal*. Colecção Breve. Ciência Política. Lisboa. Imprensa de Ciências Sociais.

Galindo, M. P. (1986). "Una alternativa de representación simultánea: HJ-Biplot". *Qüestiió*, 10 (1), pp.13-23.

Goldberger, A. (1991). *A Course in Econometrics.* Cambridge. Harvard University Press.

Goodman, L. (1953). "Ecological Regressions and the Behavior of Individuals". *American Sociological Review*, 18, pp.663-664.

Goodman, L. (1959). "Some Alternatives to Ecological Correlations". *American Journal of Sociology*, 64, pp.610-624.

Haining, R. (2003). *Spatial Data Analysis, Theory and Practice*. Cambridge, Cambridge University Press, UK.

Hicks, A. M. y Swank, D.H. (1992). "Politics, Institutions and Welfare Spending in Industrialized Democracies, 1960-82," *American Political Science Review*, 86, pp.658-74.

King, G. (1997) *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press.

Langbein, L. I. y Lichtman, A. J. (1978). Ecological Inference. *Series: Quantitative Applications in the Social Sciences*. Sage Publications.

Lipset, S. M. (1963). *El hombre político. Las bases sociales de la política*. EUDEBA, Buenos Aires.

Magalhães, P. (2001). " Desigualdade, desinteresse e desconfiança: a abstenção nas eleições legislativas de 1999". *Análise Social*.. XXXV, 157, Janeiro, pp.1079-1093.

Mayer, N. y Perrineau, P. (1992). *Les Comportements Politiques.* Paris. Armand Colin.

Meltzer, A. H. y Richards, S. F. (1981). "A Rational Theory of the Size of Government," *Journal of Political Economy*, 89, pp.914-927.

Milbrath, L.W. y Goel, M. L. (1977). *Political Participation. How and Why Do People Get Involved in Politics*. Lanham. University Press America.

Miller, G. A. (1952). "Finite Markov Processes in Psychology". *Psychometrica*, 17, pp.149-167.

Moreno, R. y Vayá, E. (2000), *Técnicas económétricas para el tratamiento de datos espaciales: la econometría espacial*. Edicions Universitat de Barcelona, colecció UB 44, manuals.

Nagel, J. H.. (1987). *Participation"* Englewood Cliffs: Prentice-Hall.

Nie, N. H. y Verba, S. (1975). "Political Participation". Greenstein, F. I. y Polsby, N. W. *Handbook of Political Science*, Vol. 4, Reading: Addison-Wesley Publishing Company.

Ogburn, W. F. y Goltra, I. (1919). "How Women Vote: A Study of an Election in Portland, Oregon". *Political Science Quarterly*, 34, pp.413-433.

O'Loughlin, J. (2003). *"Spatial Analysis in Political Geography"*. Agnew,J., Mitchell, K. y Tuathail,G. (eds). *A Companion to Political Geography*. Oxford: Basil Blackwell, pp.30-46.

Pampel, F. y Williamson, J. (1988). "Welfare Spending in Advanced Industrial Democracies 1950-1980". *American Journal of Sociology*, 93, pp.1424-56.

Piven, F. y Cloward, R. (1977). *Poor People's Movements*. Nova Iorque. Vintage Books.

Reto, L. y Sá, J. (2000). "Segmentação e tipologia dos abstencionistas em Portugal". *Revista Portuguesa de Gestão*, III (1), pp.64-74.

Robinson, W. (1950). "Ecological Correlations and the Behavior of Individuals". *American Sociological Review*, 15, pp.351-357.

Rosenstone, S. J. y Hansen, J. M. (1993). M*obilization, Participation, and Democracy in America*. New York: Macmillan.

Santo, P. (2006). *Sociologia Política e Eleitoral – Modelos e Explicações de voto*. Lisboa. Instituto de Ciências Sociais e Políticas.

Schoenberg, R. (1997a). "Constrained Maximum Likelihood". *Computational Economics*,. 10,

3. Springer Netherlands. pp.251-266.

Schoenberg, R. (1997b). *Simulation of Bayesian posterior distributions of parameters of constrained models*. Manuscript.

Schuessler, A. (1999). "Ecological inference". *Proceedings of the National Academy of Sciences USA* ,96, pp.10578-10581.

Subileau, F. (1997). "Lábstentionnisme: apolitisme ou stratégie?". Mayer, N. (org.). *Les Modéles Explicatifs du Vote*. Paris. L'Harmattan, pp.245-267.

Tanner, M.A. (1996). *"Tools for Statistical Inference: Methods for the Exploration of Posterior distributions and likelihood Functions".* Third ed. New York. Springer-Verlag.

Teixeira, R. (1992). *The Disappearing American Voter*. Washington, D. C., Brookings.

Vicente, J. L. (2007). Programa MULTBIPLOT (Version alpha 2.1). Salamanca: Departamento de Estadística, Universidad de Salamanca.

Viegas, J. M. y Faria, S. (2004). " A abstenção nas eleições legislativas de 2002". Freire, A., Lobo, M. C. y Magalhães, P. (orgs.). *Portugal a Votos. As Eleições Legislativas de 2002*. Lisboa. Imprensa de Ciências Sociais, pp.221-259.

Wakefield, J. (2004). "Ecological inference for 2 x 2 tables". *Journal Royal Statistics Society*, A, 167, Part 2, pp.1-42.

A CODIFICAÇÃO NUMÉRICO-PATOLÓGICA DOS HOSPITAIS PÚBLICOS EM PORTUGAL:
IMPLEMENTAÇÃO DE MECANISMOS DE SUPORTE NO PROCESSO DE AVALIAÇÃO DOS REGISTOS CLÍNICOS HOSPITALARES E SUA RELAÇÃO COM O FINANCIAMENTO

THE PATHOLOGICAL-NUMERIC CODIFICATION OF PUBLIC HOSPITALS IN PORTUGAL:
IMPLEMENTATION OF MECHANISMS TO SUPPORT THE ASSESSMENT PROCESS OF HOSPITAL CLINICAL RECORDS AND THEIR RELATIONSHIP WITH FUNDING

Ana Marreiros
Guilherme Castela
Efigénio Rebelo
Maria Purificación Villardón

RESUMO

A informação declarada pelos Hospitais Públicos à Administração Central do Sistema de Saúde, relativamente a episódios com estadia clínica, passa por um processo de agrupamento de Diagnósticos Homogéneos. A codificação "numérico-patológica" assume extrema importância na monitorização dos Contrato-Programa, sendo estes, a concretização documental dos montantes financeiros por linha de produção em cada unidade hospitalar. As variáveis consideradas neste estudo substanciam, em simultâneo, apoio à decisão na gestão hospitalar no que respeita ao controlo/validação dos processos clínicos a codificar, servindo também critérios de avaliação da qualidade dos registos hospitalares. Os objectivos consistem em verificar trajectórias temporais, em três anos consecutivos (2005, 2006 e 2007), dos Hospitais nos indicadores considerados e agrupá-los, de forma homogénea, em grandes grupos assentes numa estrutura multidimensional que ajude a percepcionar perfis de produção hospitalar facturável ao Sistema Nacional de Saúde. Os resultados revelaram a formação de cinco grupos de Hospitais mediante a metodologia HJ-Biplot, os quais foram detectados através de uma Análise Hierárquica de *Clusters* sobre as coordenadas principais da configuração espacial HJ-Biplot.

Palavras-Chave: HJ-Biplot, codificação numérico-patológica.

ABSTRACT

The information stated by Public Hospitals to the Central Administration of the Health System, concerning cases of clinical hospitalization, goes through a process of grouping Homogeneous Diagnoses. The "pathological-numeric" codification is a matter of extreme importance for the monitoring of the Contract-Programs, which are, the documental concretization of the financial amounts by production line in each hospital unit. The considered variables in this study embody, simultaneously, the support for the decision in hospital management concerning the control/validation of the clinical processes to codify, serving also assessment criteria for the quality of hospital records. The objectives consist in checking temporal trajectories of Hospitals, over three straight years, in the considered

indicators and grouping them, homogenously, into big groups set up on a multidimensional structure that helps get the perception of profiles of hospital production chargeable to the National Health System. The results showed the formation of five groups of Hospitals through the HJ-Biplot technique, which were detected by an Analysis of hierarchical Clusters about the main data of the applied HJ-Biplot spatial configuration.

Key-words: HJ-Biplot, Pathological-Numeric Codification.
JEL Classification: I18; C0

1. INTRODUÇÃO E OBJECTIVOS

Actualmente, em Portugal, o modelo de Saúde e o processo de financiamento da rede de cuidados primários e das unidades hospitalares públicas requerem grande reflexão por parte de todos os intervenientes que compõem a articulação orgânica do Sistema Nacional de Saúde (SNS). Gradual e temporalmente a estratégia nacional tende a atribuir maior responsabilidade, de forma individualizada, aos hospitais no que respeita à gestão orçamental accionando cumplicidade directa com as Administrações Regionais de Saúde (ARS) nesta matéria, e em questões de monitorização da produção. Verifica-se então, hoje em dia, que a fracção do orçamento de estado para a saúde é gerida, de forma "regionalizada" e descentralizada, através das ARS onde a responsabilidade das Unidades Hospitalares (UH) é partilhada pelo cumprimento da produção estimada em Contratos-Programa dos hospitais, e pela orientação/acompanhamento dos mesmos através das Administrações Regionais. Este processo, onde se atenua a intervenção económica directa do estado está direccionado para a regulação e supervisão, no qual a figura de prestador de serviços começou a evoluir para uma actuação ao nível da gestão.

A existência de um "elemento *pivot"* neste encadeamento de funcionalidade, concretizado pelas Agências de Contratualização (AC) sedeadas nas ARS, suscita polémica relativamente ao desempenho de papel regulador destes organismos. Antunes (2001) refere algumas responsabilidades que deveriam ser assumidas pelas AC nomeadamente, a arbitragem correcta dos campos de actuação das unidades públicas e privadas regulando a separação das funções de pagador e prestador, proporcionando assim uma situação de concorrência benéfica em que o "utilizador" procurará o melhor serviço em preço e qualidade. O mesmo autor, salienta também que este sistema implica autonomia administrativa e financeira das UH, embora caracterize as funções das AC como subutilizadas e não exploradas em termos funcionais. Defende igualmente que ao estado só competirá, a função de regulador e fiscalizador da acessibilidade e da qualidade.

Não obstante, em Portugal, as despesas do SNS, suportadas pelo estado, garantem que todo o cidadão pertencente ao SNS tem direito aos cuidados, independentemente da sua situação económico-geográfica. Este facto, numa moldura nacional de aumentos progressivos nas despesas com a saúde questiona a sustentabilidade financeira do SNS. Segundo Rodrigues *et al.* (2002), Portugal, numa década (de 1988 a 1998) as despesas na saúde aumentaram substancialmente em termos de percentagem do Produto Interno Bruto (PIB). Este cenário, face ao enquadramento macroeconómico da União Europeia, estendeu-se à maioria dos países[1]. Particularizando o caso Francês cujo crescimento das despesas da saúde foi na ordem dos 14,46%, nos anos considerados, e com históricos de aumentos percentuais constantes entre os anos 50 e 80[2], Béresniak e Duru (1999) apresentaram as sete principais causas para os aumentos dos custos na saúde, nomeadamente: 1)- o avanço tecnológico que accionou

[1] Alemanha, Áustria, Bélgica, Espanha, França Grécia, Holanda, Itália e Reino Unido.
[2] Entre 1950 e 1980, as despesas da saúde aumentaram em média 18% ao ano face a um crescimento de 12% do PIB.

maior exigência e especialização por parte dos médicos, 2)- o aumento da procura de saúde das populações cobertas por seguros de doença, 3)- as mutações da evolução do estilo de vida da população, 4)- o surgimento de flagelos patológicos/sociais (sida, toxicodependência, etc.), 5)- o aumento da demografia médica que ajudou a direccionar uma procura induzida, 6)- uma acentuada "medicalização" dos problemas sociais relacionados à inexistência de estruturas de apoio e acolhimento social e, por fim, 7)- o aumento do desperdício de recursos. Os autores expressaram, de uma forma abstracta, que o aumento dos custos com a saúde é inexorável dado que espelha a própria evolução das sociedades. Nos seus pontos de vista, as políticas de controlo de crescimento das despesas com a saúde são de difícil implementação porque molduram-se em contextos de interesses quer do ponto de vista corporativo, quer do político-económico ou quer de uma perspectiva industrial. Fazendo uma comparação homóloga a Portugal, Escoval e Matos (2009) apontam quatro justificações para o aumento das despesas na saúde, designadamente: 1)- o envelhecimento da população, 2)- o desenvolvimento das tecnologias na saúde, 3)- o aumento na oferta de cuidados e, uma 4)- postura mais exigente/expectante dos cidadãos perante os serviços de saúde prestados.

Neste contexto, e relativamente aos episódios de internamento, existindo uma relação entre os GDH e o montante financeiro de retorno "indirecto", este estudo, atendendo ás variáveis consideradas, tem quatro objectivos, designadamente: 1)- encontrar um instrumento de análise conducente a um processo de formulação e de execução de práticas de codificação médica mais eficaz, favorecendo o sistema de facturação ao S.N.S, 2)- facilitar o diagnóstico da situação dos registos clínicos nacionais e contextualizando-os nos grupos de financiamento de acordo com a legislação em vigor, 3)- encontrar subconjuntos homogéneos de UCH com GDH clínicamente inadequados, que possibilitem a identificação de perfis de produção hospitalar mais concordantes com a actual conjuntura clínico-administrativa e, 4)- convergir para um instrumento que permita desenvolver estratégias futuras mais eficientes, em matéria de financiamento da saúde.

2. CONTEXTUALIZAÇÃO E PERTINÊNCIA DO ESTUDO

A Administração Central dos Serviços de Saúde (ACSS) reúne todos os dados de produção de cada U/CH Pública de Portugal, tanto no que respeita ao total da produção realizada como também à proporção de produção SNS. Esta informação é enviada, mensalmente, à ACSS por todos os H/C H. Também em referência mensal, as ARS, das quais cada UH está afecta, estão munidas dos valores da produção total. As AC, mediante os valores de produção mensais reportados e susceptíveis de facturação ao SNS pelas UH, assumem o papel de monitorização destes valores face ao contratualizado anual. Neste trabalho, apenas foram tratados os dados de produção total referentes a episódios com estadia clínica, ou seja, só foi tratada informação da linha de produção hospitalar internamento. Sublinha-se, neste contexto, que os episódios considerados foram codificados segundo a Classificação Internacional de Doenças – 9ª Revisão (CID – 9-MC) e agrupados em Grupos de Diagnósticos Homogéneos (GDH). Dado que o financiamento desta linha de produção, que consta nos "Contratos-Programa" de cada UH, assenta nestes processos já devidamente codificados, torna-se relevante o acompanhamento deste procedimento de codificação patológica, quer em termos de avaliação da qualidade dos registos clínicos quer num prisma de monitorização/acompanhamento dos Contratos. A existência de relação entre a produção do internamento e respectiva codificação com a parte de facturação dos hospitais ao SNS ajudou na escolha das variáveis consideradas neste estudo. Estas, neste contexto, resultam de uma categorização padronizada pela ACSS e constam dos relatórios de retorno anuais

efectuados por esta entidade, relativamente a cada U/CH. As variáveis em causa resultam do agrupador onde são inseridos os códigos de doença dos processos em cada hospital. O resultante da codificação e a importância de verificação da qualidade do processo de agrupamento em GDH, bem como sua relação "pseudo penalizadora" em termos financeiros conduziu-nos à escolha das seguintes variáveis de estudo: Diagnóstico Principal Inespecífico (DPI), Diagnóstico Principal Inaceitável Clinicamente (DPIC), Diagnóstico Principal de Manifestação, Diagnóstico principal - efeito tardio, Diagnóstico Principal Sintoma (DPS), Diagnóstico Principal Admissão Questionável (DPAQ), Procedimentos cirúrgicos inespecíficos (PCI) e Internamentos prolongados sem complicações assinaladas (IPSCA). Também se consideraram as Complicações com a Gravidez (CG), as Úlceras de Decúbito (UD), a estadia média (EM), os dias de internamento (DI), as readmissões (READ) e os diagnósticos e procedimentos por episódio (DEPIS e PEPIS). Esta investigação reporta ao período temporal entre 2005 e 2007 e a informação tratada contempla 64 Hospitais do Sistema Nacional de Saúde em Portugal.

3. METODOLOGIA

No âmbito da classificação de dados, da redução da dimensionalidade e, sobretudo, na detecção de padrões de interesse, existem varias técnicas de análise multivariada de dados mediante métodos factoriais gráficos. Estes métodos estão descritos inicialmente para as Análises de Componentes Principais ou de Correspondências, mas podem estender-se, de acordo com Galindo e Cuadras (1986), aos métodos BIPLOT. Na realidade, os métodos BIPLOT têm sido fonte de contínuas contribuições para a ciência desde a sua origem com Gabriel (1971) sendo, no entanto, ainda pouco conhecidos no campo da Administração Hospitalar. Um BIPLOT é uma representação gráfica de dados multivariados e, da mesma maneira que um diagrama de dispersão descreve a distribuição conjunta de duas variáveis, um BIPLOT representa três ou mais variáveis, Gabriel e Odoroff (1990). O BIPLOT aproxima a distribuição de uma amostra multivariada a um espaço de dimensão reduzida, normalmente de dimensão dois, e sobrepõe sobre a mesma representação as variáveis sobre as quais a amostra é medida, Gower (1996). As representações das variáveis são vectores e coincidem com as direcções que melhor descrevem as alterações individuais de cada variável. O prefixo "BI" refere-se à sobreposição, na mesma representação, de indivíduos e variáveis. Estes métodos são úteis para descrever graficamente os dados ou para mostrar os resultados de modelos mais formais e, do ponto de vista do utilizador, são importantes porque a sua interpretação é baseada em conceitos geométricos simples, que fazem parte da cultura matemática, nomeadamente: 1)- a similitude entre indivíduos é uma função inversa da distância entre eles, 2)- as longitudes e os ângulos dos vectores que representam as variáveis interpretam-se em termos de variabilidade e covariabilidade, respectivamente, 3)- as relações entre indivíduos e variáveis interpretam-se em termos de produto escalar, isto é, em termos das projecções dos pontos "indivíduo" sobre os vectores "variável". Em suma, o seu objectivo geral é a aproximação de uma matriz de dados, de tal forma que permita a sua descrição ou modelização através de mapas geométricos construídos como projecções de nuvens de pontos-linha e pontos-coluna sobre sub-espaços de ajustamento óptimo, de maneira a que se conservem os padrões fundamentais no que concerne à variação conjunta dos indivíduos e das variáveis. As principais diferenças que os BIPLOT apresentam relativamente às distintas técnicas factoriais dependem das matrizes e tipos de variáveis que se utilizam e, sobretudo, da forma das suas representações gráficas. Normalmente, essas representações são assimétricas no sentido de que não se obtém a mesma qualidade de representação para as linhas e para as colunas da matriz de dados. Assim, quando o propósito é a aproximação

dos elementos da matriz original, os BIPLOT apresentados são óptimos. Além disso, em cada um deles é possível representar com melhor qualidade as características das linhas ou das colunas, quando se pretendem interpretar por separado. No entanto, quando as linhas e as colunas são de igual importância e se deseja interpretar as características de ambas, mantendo certa relação entre as mesmas, são mais úteis as interpretações baseadas em representações simétricas como a Análise Factorial de Correspondências na qual se interpretam as posições das linhas, as posições das colunas e as relações linha-coluna através dos factores, isto é, através de uma interpretação factorial. O problema é que a Análise Factorial de Correspondências está pensada somente para matrizes de frequências, o que não é o presente caso. Assim, torna-se necessário dispor de uma técnica simétrica similar, mas aplicável a qualquer conjunto de dados. Galindo (1986) propõe o que denomina de HJ-BIPLOT e que responde às características descritas nos dados que julgamos vir a estar presentes nesta investigação. Nesta representação particular, a interpretação dos indivíduos e das variáveis é a mesma, sem contudo se impedir que a busca das variáveis que determinam as diferenças entre os indivíduos se realize através de eixos factoriais, isto é, interpretam-se as novas variáveis, combinações lineares das de partida, e interpretam-se as relações das mesmas com as variáveis observadas. A medida da relação entre os eixos da representação HJ-BIPLOT e cada uma das variáveis observadas é denominada por Contribuição Relativa do Factor ao Elemento (variável), que representa a parte da variabilidade de cada uma das variáveis explicada pelo factor e interpreta-se da mesma forma que um coeficiente de determinação na Análise de Regressão. De facto, se os dados estão centrados, representa o coeficiente de determinação da regressão de cada variável sobre o eixo correspondente. Esta contribuição permite-nos saber quais são as variáveis mais directamente relacionadas com cada eixo e, portanto, permite-nos conhecer as variáveis responsáveis pela colocação dos indivíduos sobre as projecções em cada um dos eixos. Como os eixos se constroem para que sejam independentes, a contribuição de cada um deles a cada variável é independente, portanto, é possível calcular a contribuição de um plano através da simples soma das contribuições dos eixos que o formam. Propomos, pois, numa primeira fase, a utilização destas Técnicas de Análise Multivariada, num contexto de redução da dimensionalidade, onde a representação gráfica dos dados, através da metodologia HJ-BIPLOT, permite explicar as relações existentes entre as práticas de codificação clínica entre as diversas U/CH. A justificação assenta na interpretabilidade e na observação de que a grande maioria dos resultados das análises de dados multivariados pode ser expressa por listas de indivíduos, de variáveis ou listas mistas de indivíduos e variáveis. Esta observação empírica tem como consequência a de que todos os resultados expressáveis desta forma podem, naturalmente, ser representados por configurações de marcadores de variáveis e indivíduos em BIPLOT. Assim, este trabalho procura mostrar a utilidade dos métodos BIPLOT, em particular o HJ-BIPLOT, para inspeccionar dados provenientes da produção de internamento hospitalar em termos da codificação em GDH, categorizados como indicadores de avaliação da qualidade dos registos clínicos e, simultaneamente, penalizadores de financiamento. Deste modo, os resultados apurados no âmbito da presente investigação poderão constituir um auxílio nos processos de formulação e de execução de práticas de codificação médica mais eficientes, como também desenvolvem e estimulam sugestões a pesquisa sobre a validade e a fiabilidade do recurso institucionalizado à facturação destes episódios ao SNS. Pretendemos, com esta abordagem, no campo da gestão hospitalar, facilitar o diagnóstico da situação dos registos clínicos nacionais e contextualizar os grupos de financiamento, e, neste sentido, convergir para um instrumento que permita desenvolver estratégias futuras. Resultado este que deve constituir um incentivo a uma melhoria significativa na gestão da saúde.

Numa segunda fase, a Análise *Cluster* engloba diferentes algoritmos de classificação para organizar informações sobre determinadas variáveis e formar grupos homogéneos. Pestana

e Gageiro (2003) definem esta análise como um procedimento que permite detectar grupos homogéneos nos dados, podendo tais grupos ser constituídos por variáveis ou casos. Neste seguimento, a Análise *Cluster* pretende organizar um conjunto de casos em grupos homogéneos, de modo a que os indivíduos (UH) pertencentes a um grupo sejam o mais semelhante possível entre si e simultaneamente diferenciados dos restantes grupos. Esta análise consiste, portanto, numa técnica de interdependência, ou seja, uma vez que não é possível determinar antecipadamente as variáveis dependentes e independentes, este método analisa as relações de interdependência entre o conjunto de variáveis. No âmbito desta análise, destacam-se, tal como referem Pestana e Gageiro (2003), a Análise *Cluster* de Variáveis e a Análise *Cluster* de Casos. A primeira é semelhante à Análise Factorial[3], na medida em que ambos os procedimentos identificam grupos de variavéis relacionadas entre si. Em contrapartida, a Análise *Cluster* de casos é similar à Análise Discriminante[4] ao procurar classificar um conjunto de dados iniciais em grupos ou categorias, utilizando os valores observados das variáveis que se referem ao fenómeno em causa. Existem, assim, dois métodos para formar *Clusters*. Por um lado, a Análise Hierárquica de *Cluster*, que utiliza, quer variáveis, quer casos, e que pressupõe desenvolver uma hierarquia na qual a formação de grupos constitui um processo sequencial que se representa graficamente através de um Dendrograma. Por outras palavras, dada uma população, trata-se de estabelecer uma hierarquia de partes, delimitando um número de subconjuntos de forma que, entre eles, não existam elementos comuns (sejam disjuntos) e que cada subconjunto esteja incluído noutro (hierarquia). Deste modo, o número de grupos identificados depende da sequência considerada. Por outro lado, a Análise Não Hierárquica de *Cluster*, que apenas se aplica a casos e que permite a partição dos indivíduos num número pré-definido de grupos segundo determinado critério, situação que não se adapta ao presente estudo. Acresce ainda, dentro dos procedimentos hierárquicos, os ascendentes ou aglomerativos e os descendentes ou divisivos. Os primeiros utilizam um processo iterativo de procedimentos ascendentes ou aglomerativos de agrupamentos, que começa com tantos grupos como indivíduos. Assim, formam-se grupos entre os indivíduos mais parecidos, segundo um determinado critério[5] e o processo termina com um só grupo, que integra todos os elementos da população. Contrariamente, os procedimentos descendentes ou divisivos, partem de um conjunto populacional que se divide em subconjuntos até chegar ao elemento. Esperamos, pois, encontrar, no fim deste percurso metodológico, grupos de Hospitais Portugueses, com comportamentos homogéneos, estatisticamente relevantes, que permitam avaliar, de acordo com os indicadores da qualidade dos seus registos clínicos, a produção financiada pelo S.N.S., em função da actual conjuntura clínico-administrativa.

4. RESULTADOS

Os resultados revelaram a forma como se agrupam os Hospitais Portugueses em função da qualidade dos seus registos hospitalares e características de produção. Neste contexto, encontramos cinco grupos de Hospitais, bem definidos mediante o procedimento HJ-Biplot, os quais foram detectados através de um *clustering* hierárquico, com recurso ao método de Ward, sobre as coordenadas principais da configuração HJ-Biplot.

[3] Entende-se por análise factorial como um conjunto de técnicas estatísticas que procura explicar a correlação entre as variáveis observáveis, simplificando assim, os dados através da redução do número de variavéis necessárias para os descrever (Pestana e Gageiro, 2003).

[4] A análise discriminante é uma técnica multivariada, que se aplica apenas quando as variavéis independentes são quantitativas e a variável dependente é qualitativa.

[5] De entre os procedimentos ascendentes ou aglomerativos, os mais critérios mais populares são: o vizinho mais próximo ou *Single Linkage*, o vizinho mais afastado ou *Complete Linkage*, o vizinho médio ou *Average Linkage*, o método do centróide e o método de *Ward*. Este último, é um dos métodos mais usuais de agrupamento e utiliza uma análise de variância para avaliar as distâncias entre grupos ao pretender minimizar a soma dos quadrados dos desvios das observações individuais entre grupos. Assim, os grupos vão constituindo de maneira que se produza o menor aumento nas somas dos quadrados. É um critério que permite a formação de grupos o mais homogeneamente possível.

As contribuições do factor ao elemento permitem determinar a parte da variabilidade de cada variável explicada pelo factor. A medida de relação entre os eixos da representação biplot e cada uma das variáveis é a contribuição relativa do factor ao elemento. A tabela 1 apresenta as respectivas contribuições.

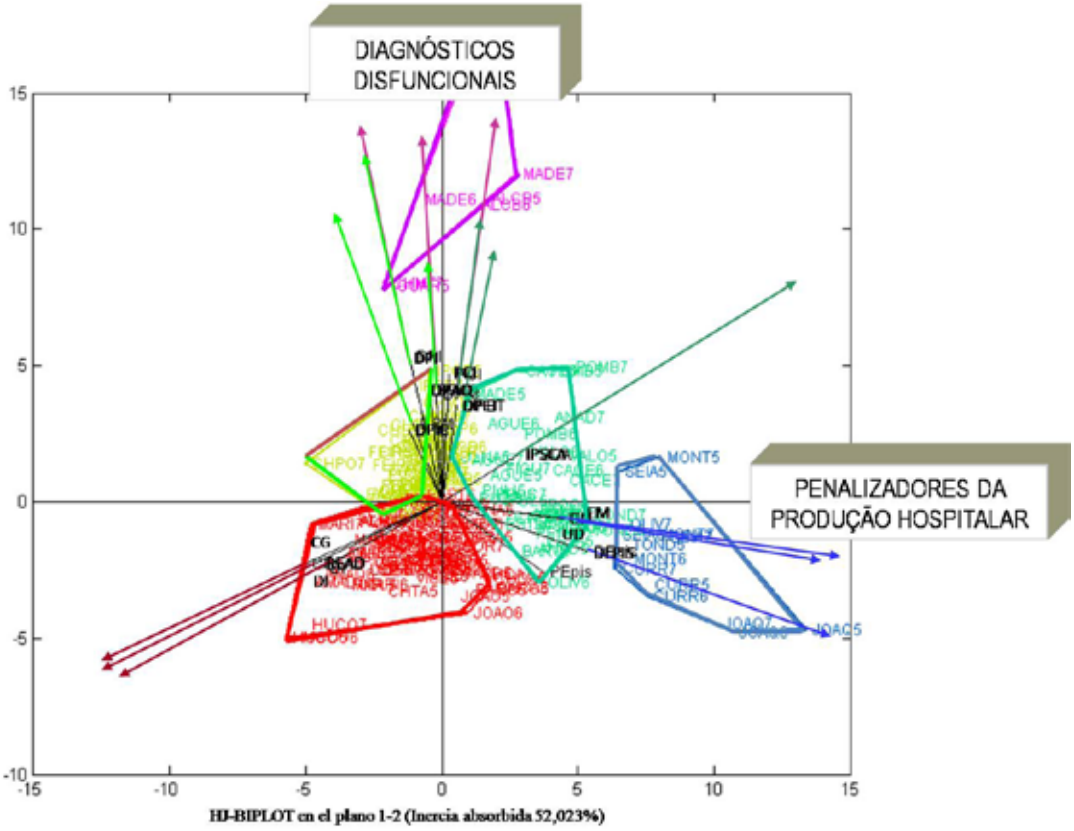
Tabela 1 – Contribuições factor ao elemento

Variáveis	Eixo 1	Eixo 2
Estadia Média - EM	**0,368**	0,08
Dias de Internamento - DI	**0,354**	0,108
Úlceras de Decúbito - UD	**0,354**	0,07
Readmissões - READ	**0,394**	0,091
Complicações com a Gravidez - CG	**0,475**	0,095
Diagnóstico Principal Inespecífico - DPI	0,028	**0,530**
Diagnóstico Principal Inaceitável Clinicamente - DPIC	0,01	**0,290**
Diagnóstico Principal de Manifestação - DPM	0,021	**0,152**
Diagnóstico principal de Efeito Tardio - DPET	0,06	**0,234**
Diagnóstico Principal Sintoma - DPS	0,028	**0,130**
Diagnóstico Principal Admissão Questionável - DPAQ	0,04	**0,317**
Procedimentos Cirúrgicos Inespecíficos - PCI	0,01	**0,404**
Internamentos Prolongados sem Complicações Assinaladas - IPSCA	0,147	0,056
Diagnósticos por Episódio - DEPIS	**0,526**	0,062
Procedimentos por Episódio - PEPIS	**0,259**.	0,123

De acordo com a Figura 1, visualiza-se a formação dos cinco grupos, num espaço bidimensional com uma inércia absorvida de 52,023%, onde se evidencia as variáveis que mais caracterizam cada *cluster*. Os cinco *clusters* de Hospitais diferenciam-se, essencialmente, com base em dois grandes factores latentes: um caracterizado por "Penalizadores da Produção Hospitalar" e representado no 1º eixo, e o outro por "Diagnósticos Disfuncionais" no 2º eixo. Tais designações justificam-se pela formação das contribuições mais elevadas de cada variável pelos factores. Por conseguinte, e com o auxílio da Tabela 1; constata-se que o 1º eixo engloba os indicadores relacionados com estadias clínicas ou factores que influenciam a duração dos internamentos (por exemplo, as UD, as READ e as CG). Relativamente ao 2º eixo, verifica-se que as contribuições relativas com maior peso estão só relacionadas com "más codificações" o que sugere os "Diagnósticos Disfuncionais" com base no seguinte grupo de variáveis: DPI, DPIC, DPM, DPET, DPS, DPAQ e PCI. Sublinha-se o facto destes últimos indicadores intervirem, indirectamente no processo de facturação dos Hospitais ao SNS, constituindo simultaneamente matéria agravante na avaliação da qualidade dos registos clínicos. Através do ângulo formado pelos vectores (variáveis), na configuração Biplot, pode observar-se três grandes grupos de variáveis correlacionadas. Um 1º Grupo constituído pelos

DPI, PCI, DPAQ, DPET e DPIC, um 2º Grupo formado pelas CG, READ e DI e um 3ª Grupo composto pelos DEPIS, EM e UD.

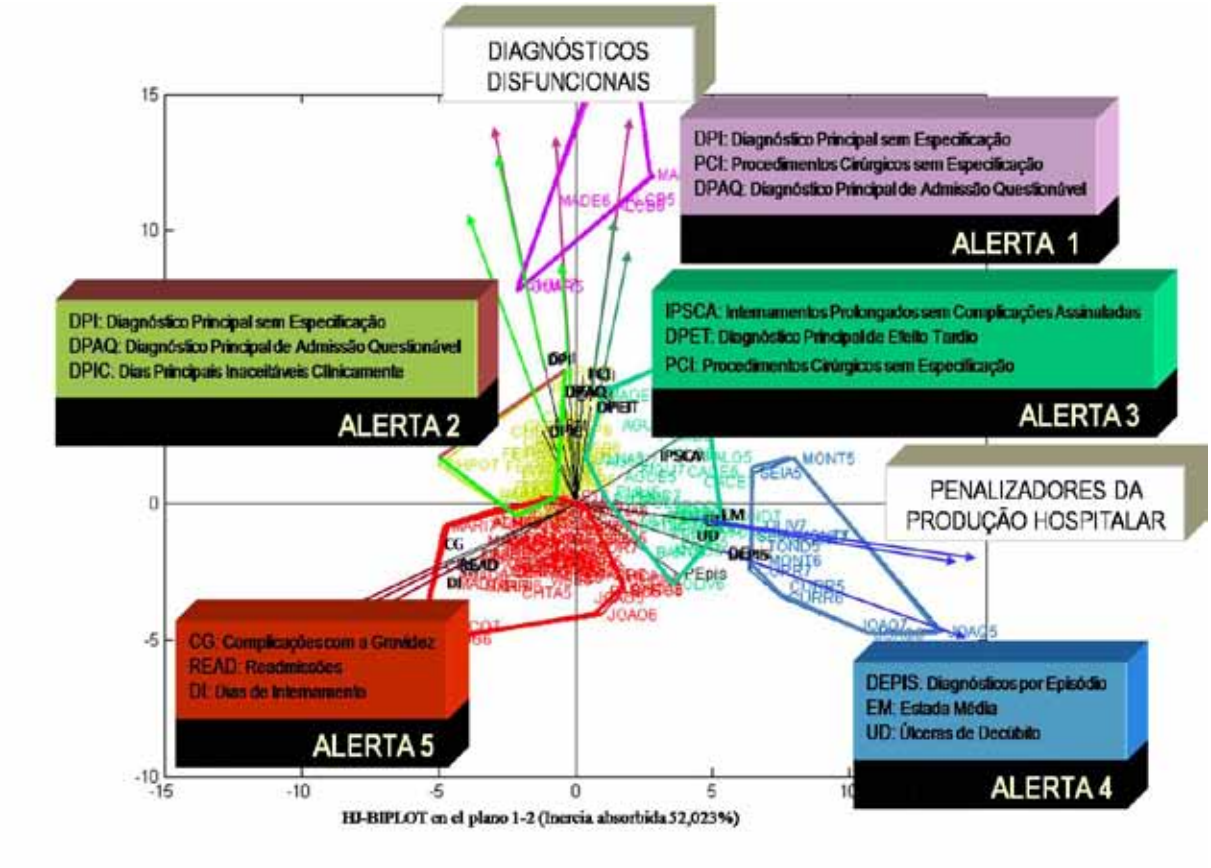
Figura 1 – Configuração HJ-BIPLOT, no plano 1-2, e processo de segmentação



HJ-BIPLOT en el plano 1-2 (Inercia absorbida 52,023%)

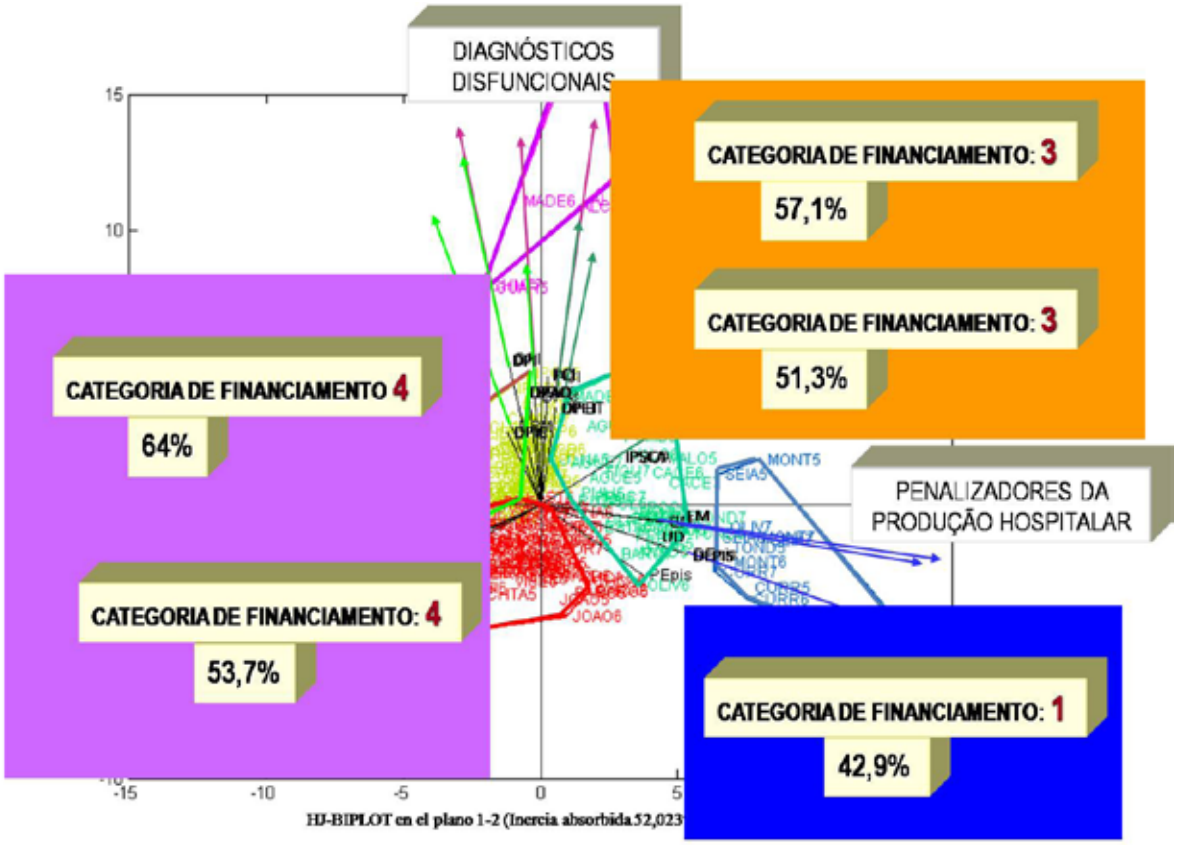
Na Figura 2, pode observar-se, a caracterização de cada *cluster*. As combinações das variáveis que mais contribuíram para a formação dos grupos de hospitais foram designadas por "Alertas". A numeração de cada grupo destas variáveis deve-se somente à identificação do *cluster* de indivíduos a ser considerado (por exemplo: *cluster 1* corresponde à combinação de variáveis Alerta 1). Fixando somente as variáveis com influência nos orçamentos-programa de cada U/CH, que são as mais explicadas pelo factor "Diagnósticos Disfuncionais", à que referir que os *clusters* 1, 2 e 3 são grupos de Hospitais mais associados às "más codificações" e, por conseguinte, menor volume de produção internamento susceptível de facturação ao SNS. Também estes grupos de U/CH serão os mais conotados com menor qualidade de registos clínicos. Relativamente aos *clusters* 4 e 5, sendo caracterizados pelos grupos de variáveis Alerta 4 e Alerta 5, conclui-se que estes hospitais são mais marcados por indicadores de produção hospitalar. O *cluster* 4 sugere que, nestes hospitais, os dias de internamento estão altamente relacionados com as readmissões e as complicações com a gravidez, e o *cluster* 5 de indivíduos onde se evidencia a correlação forte entre demora média, as úlceras de decúbito e o número de diagnósticos por episódio.

Figura 2 – Caracterização dos grupos de U/CH



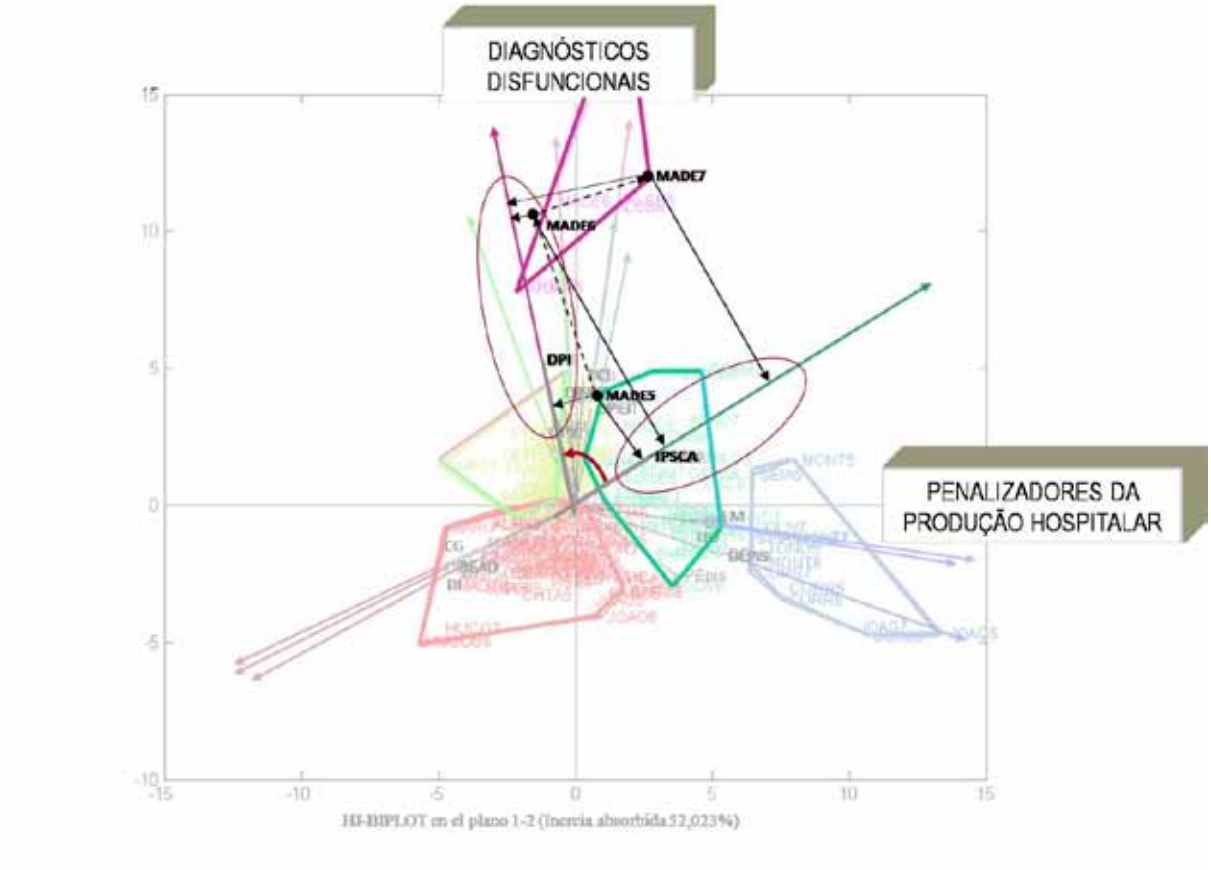
Face ao exposto anteriormente, considerou-se importante saber a que grupos de financiamento estes *clusters* de hospitais, em maior valor percentual, pertenceriam. Assim, por observação da Figura 3, constata-se que os *clusters* 1 e 3 têm 57,1% e 51,3%, respectivamente, de U/CH pertencentes à categoria de financiamento 3. Repare-se que estes grupos de hospitais têm em comum variáveis que mais os caracterizam e que estão relacionadas com as codificações mal executadas. O *cluster* 4 tem a maior proporção de indivíduos afectos ao grupo de financiamento 1 (42,9%), sendo que os *clusters* 2 e 5 pertencem ambos à categoria de financiamento 4. Por observação complementar, note-se que estes dois últimos grupos de hospitais, apesar de estarem na mesma categoria, distinguem substancialmente pelos respectivos grupos de variáveis (alertas 2 e 5) que mais os diferenciam: um é caracterizado, essencialmente, por variáveis de produção hospitalar (*cluster* 5) e o outro definido por variáveis resultantes de uma codificação disforme.

Figura 3 - Caracterização dos grupos de UCH em função do financiamento



Com o objectivo de observar-se as alterações sofridas pelas U/CH influenciadas pelas variáveis de estudo, de forma diferida nos três anos consecutivos considerados, tomou-se, como exemplo das mudanças das coordenadas no plano, uma UH abreviada por MADE. Na Figura 4, observa-se a trajectória sequencial desta UH, cujos pontos referenciados na configuração Biplot identificam o ano pela seguinte terminologia: MADE5, MADE6 e MADE7. Este indivíduo refere-se aos anos de 2005, 2006 e 2007, respectivamente. Neste contexto, a projecção ortogonal de um Hospital sobre um vector que representa um índice permite-nos conhecer o valor tomado por esse indivíduo nessa variável. Assim, pelas projecções ortogonais do MADE nas variáveis que mais contribuem para a caracterização dos *clusters* 1 e 3 (DPI e IPSCA, respectivamente), constata-se que houve uma evolução muito distinta nestas variáveis, ao longo dos três anos. Verificou-se uma significativa variação 2006-2007, traduzida pelo segmento de recta, limitado superior e inferiormente pelas projecções ortogonais de cada ano, na variável IPSCA, cujo comportamento das mesmas projecções na variável DPI foi absolutamente oposto. Ou seja, as variações do mesmo indivíduo, MADE, de 2006 para 2007, nas duas variáveis consideradas, traduziram-se por 60% na IPSCA e de 1,8% na DPI. Relativamente à variação de 2005 a 2006, fixadas as mesmas variáveis, o MADE teve um comportamento definido por um crescimento de 262% na DPI versus 5% na IPSCA.

Figura 4 – Trajectória de uma UH (MADE)



Em suma, estas grandes variações opostas, nos mesmos períodos, nas mesmas variáveis fizeram com que esse indivíduo ficasse a ser mais explicado por outros indicadores cuja consequência traduziu-se numa mudança de *cluster*. Ou seja, o caso particularizado pelo MADE mostrou que, nos três anos de estudo, mudou de *cluster* de hospitais, isto é, passou a ser melhor definido por outra combinação de variáveis, não obstante ter-se mantido no mesmo grupo de Financiamento. No entanto, em termos da qualidade dos registos clínicos, esta mudança, de Alerta 3 para Alerta 1, traduziu a passagem para um estado mais conotado a graves problemas de codificação clínica.

Refere-se que toda a análise levada a cabo neste estudo foi executada no programa MULTIPLOT (Villardón, 2007).

5. CONCLUSÕES

Mediante o estudo exposto, foi encontrado um instrumento de análise conducente a um processo de identificação das práticas de codificação médica praticadas pelas U/CH.

Conseguiu-se agrupar os Hospitais de Portugal em cinco grandes grupos homogéneos, com base em variáveis directas de produção e índices de codificação incorrecta.

Obteve-se um diagnóstico da situação dos Hospitais SNS em função da qualidade dos registos clínicos nacionais e respectiva contextualização nos grupos de financiamento de acordo com a legislação em vigor.

Através do exemplo, MADE, verificou-se que, nos três anos considerados, este Hospital mudou de *cluster* não obstante ter ficado no mesmo grupo de financiamento. Esta mudança

sugere um quadro mais grave em termos da qualidade dos registos clínicos e também de maior penalização em termos de facturação ao SNS.

Em suma, foi possível encontrar subconjuntos homogéneos de U/CH com GDH clínicamente inadequados, que possibilitem a identificação de perfis de produção hospitalar mais concordantes com a actual conjuntura clínico-administrativa.

Conseguir-se-á com este estudo convergir para um instrumento que permita desenvolver estratégias futuras mais eficientes, em matéria de financiamento da saúde, tendo por base a análise e importância da codificação na produção internamento hospitalar.

AGRADECIMENTOS

Este artigo é parcialmente apoiado pela Fundação para a Ciência e a Tecnologia (FCT).

BIBLIOGRAFIA

Antunes, M. (2001). *A Doença da Saúde*. 5ª Edição. Quetzal Editores. Lisboa.

Asenjo, M, Bohigas, L., Trilla, A. e Prat, A. (2001). *Gestión Diaria del Hospital*. 2ª Edição. Masson. Barcelona.

Béresniak, A. e Duru, G. (1999). *Economia da Saúde*. 1ª Edição, Climepsi Editores. Lisboa.

Biscaia, A., Martins, J., Carreira, M., Gonçalves, I., Antunes, A. e Ferrinho, P. (2006). *Cuidados de Saúde Primários em Portugal, Renovar para Novos Sucessos*. 1ª Edição. Padrões Culturais Editora. Lisboa.

Cabral, N., Amador, O. e Martins, G. (Org.). *A Reforma da Saúde: Uma Realidade Iminente?*. Edições Almedina. Coimbra. Colecção Cadernos IDEFF. nº 11.

Campos, A. (2008). *Reformas da Saúde*. *O fio condutor*. Edições Almedina. Coimbra.

Campos, L., Borges, M. e Portugal, R. (2009). *Governação dos Hospitais*. 1ª Edição. Casa das Letras. Alfragide.

Couto, R. e Pedrosa, T. (2003). *Hospital – Gestão Operacional e Sistemas de Garantia de Qualidade*. Medsi Editora Médica e Científica. Rio de Janeiro.

Escoval, A. e Matos, T. (2009). "A Contratualização e Regulação nos Hospitais", in: Campos, L., Borges, M. e Portugal, R. (Org.). *Governação dos Hospitais*. 1ª Edição. Casa das Letras. Alfragide.

Galindo, M.P. (1986) *Una alternativa de representación simultánea: HJ-biplot*. Questíio, 10(1), 13-23.

Galindo, P. e Cuadras, C. (1986). *Una extensión del método Biplot y su relación con otras técnicas*. Publicaciones de Bioestadística y Biomatemática. Universidad de Barcelona 17.

Gabriel, K. (1971). "*The biplot graphic display of matrices with application to principal component analysis*". Biometrika.58 (3). pp 453 - 467.

Gabriel, K. e Odoroff, Ch. (1990). *Biplots in biomedical research*.Statistic in Medicine 9 (1990), pp.469 - 485.

Gower, J. e Hand, D. (1996). *Biplots*. Chapman & Hall, London, UK.

Laverack, G. (2008). *Promoção de Saúde. Poder e Empoderamento*. Lusodidacta. Loures.

López, P., Escribano, J. e Garcia, J. (2005). *Gestión Clínica en Cirurgía*. Arán Ediciones. Madrid.

Malagón-Lodoño, G., Morera, R. e Laverde, G. (2003). *Administración Hospitalaria*. 2ª Edição. Editorial Médica Panamericana. Buenos Aires.

Marques, R. (2005). *Regulação de Serviços Públicos.* 1ª Edição. Edições Sílabo. Lisboa.

Martins, D. (2002). *Custeio Hospitalar por Actividades*. Editora Atlas. São Paulo.

Nunes, R. (2009). *Regulação de Saúde*. 2ª Edição. Vida Económica Editorial. Porto.

Organização Mundial de saúde (2001). *Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde*. *CID – 10*. 10ª Edição. Vol. 2. Editora da Universidade de São Paulo. São Paulo.

Pestana, H., M. e Gageiro., N., J. (1998). *Análise de Dados para Ciências. A Complementariedade do SPSS*. 1ª Edição. Edições Sílabo. Lisboa.

Rego, G. (2008). *Gestão Empresarial dos Serviços Públicos – Uma Aplicação ao Sector da Saúde*. Vida Económica. Porto.

Rodrigues, L., Ginó, A., Sena, C. e Dahlin, K. (2002). *Conhecer os Recursos Humanos do Serviço Nacional de Saúde*. Edições Colibri. Lisboa.

Santana, P. (2005). *Geografias da Saúde e do Desenvolvimento*. Edições Almedina. Coimbra.

Silva, A., Varanda, J. e Nóbrega, S. (2004). *Alquimia da Qualidade na Gestão dos Hospitais*. 1ª Edição, Principia, Publicações Universitárias e Científicas. Cascais.

Vale, L. (2010). "As Taxas Moderadoras e o Financiamento do Serviço Nacional de Saúde: Elementos para uma Perspectiva Constitucional", in: Cabral, N., Amador, O. e Martins, G. (Org.). *A Reforma da Saúde: Uma Realidade Iminente?*. Edições Almedina. Coimbra. Colecção Cadernos IDEFF. nº 11. pp. 101-108.

Villardón, J.L. (2007) *Programa MULTBIPLOT* v.alpha 3.1. Departamento de Estadística. Universidad de Salamanca.

UM CONTRIBUTO DA METODOLOGIA CHAID PARA A DETECÇÃO DE PERFIS CRIMINAIS:
DESENVOLVIMENTOS NO PLANEAMENTO ESTRATÉGICO E NA ORIENTAÇÃO TÁCTICA DOS MEIOS POLICIAIS NA REGIÃO DO ALGARVE

A CHAID CONTRIBUTION TO THE DETECTION OF CRIMINAL PROFILES:
DEVELOPMENTS IN STRATEGIC PLANNING AND TACTICAL GUIDANCE OF POLICE RESOURCES IN THE ALGARVE REGION

Bruno Mora
Guilherme Castela

RESUMO

O conhecimento dos perfis criminais no que concerne aos períodos de ocorrência e caracterização dos suspeitos e das vítimas, é uma valiosa fonte de conhecimento para as forças de segurança tanto numa perspectiva táctica como estratégica. Do ponto de vista mais tradicional de policiamento, esta informação permite alocar os recursos de forma mais eficiente, diminuindo os tempos de resposta e aumentando a produtividade dos departamentos policiais. O objectivo deste trabalho é definir os perfis criminais dos suspeitos e das vítimas de crime no Algarve entre 2008 e 2009 utilizando a metodologia CHAID. O conhecimento dos perfis criminais permitir-nos-á, dentro do actual modelo de policiamento aplicado no Algarve, sugerir adaptações à realidade da criminalidade local, assegurando assim que a polícia estará no sítio certo, na hora certa.

Palavras-chave: Crime, CHAID, Polícia

ABSTRACT

The knowledge of criminal profiles in what relates with the time of occurrence and the characteristics of suspects and victims is a valuable source of information to the police, both in a strategic and tactical perspective.
In a more traditional point of view, this type of information allows for a more efficient allocation of human and material resources, shortening the response times and increasing the productivity of police departments.
The objective of this paper is to define the profiles of suspects and victims of criminal offenses in the Algarve in the period between 2008 and 2009 using the CHAID methodology.
This knowledge provides the necessary information to reorganize the police model in the Algarve, guarantying that the police will be in the right place at the right time, thus increasing productivity and reducing crime rates.

Key-Words: Crime, CHAID, Police
JEL Classification: C00; M00

1. INTRODUÇÃO

O rigor metodológico nas iniciativas de prevenção criminal tem aumentado significativamente nas últimas duas décadas. Em 2004, e no seguimento desta aposta na cientificidade do trabalho policial, foi introduzido na PSP o Sistema Estratégico de Informação, Gestão e Controlo Operacional. O sistema foi desenhado e desenvolvido pela PSP e tem como objectivo permitir o controlo e gestão de toda a actividade operacional diária, incluindo os registos de ocorrências a nível nacional, na áerea da PSP, garantindo um acesso centralizado à informação em tempo real e o cruzamento de dados.

O conhecimento dos perfis criminais no que concerne aos períodos de ocorrência e caracterização dos suspeitos e das vítimas, é uma valiosa fonte de conhecimento para as forças de segurança tanto numa perspectiva táctica, como estratégica.

Do ponto de vista mais tradicional de policiamento, esta informação permite alocar os recursos de forma mais eficiente, diminuindo os tempos de resposta e aumentando a produtividade dos departamentos policiais.

Da mesma forma, em termos estratégicos, a recolha de informações policiais pode ser melhor dirigida se este tipo de informação estiver disponível. Neste sentido, o Policiamento Orientado pelas Informações, internacionalmente designado como *Intelligence-led Policing*, emerge como novo modelo de acção policial destinado a reduzir a criminalidade através da pro-actividade assente na utilização da informação criminal

2. A PREVENÇÃO CRIMINAL

Para falarmos de prevenção criminal torna-se conveniente conhecermos algumas teorias do crime, uma vez que só poderemos delinear estratégias de prevenção se tivermos conhecimento das causas e pressupostos que podem estar por detrás do fenómeno criminal. As teorias do crime assentam num único pressuposto: que a facilidade ou a tentação levam as pessoas ao acto criminal. Este pressuposto encontra-se em cada uma das mais recentes teorias do crime, como a teoria das actividades de rotina e a perspectiva da escolha racional.

2.1 Teoria das Actividades de Rotina

A teoria das actividades de rotina foi proposta por Lawrence Cohen e por Marcus Felson (1979), como resposta à subida da taxa de crime dos Estados Unidos entre 1947 e 1974 e assume que para certo tipo de crimes ocorrerem, tem que existir uma convergência no tempo e no espaço de três elementos: um delinquente motivado, um alvo apropriado e a ausência de um guardião capaz (Cohen e Felson, 1979: 589). Esta teoria toma a presença de um delinquente motivado como garantida e foca-se nos outros elementos. O guardião não é necessariamente um agente policial, mas sim alguém cuja presença ou proximidade pode desencorajar a prática de um crime.

Para que ocorra um crime, o delinquente motivado tem que encontrar um alvo apropriado na ausência de um guardião capaz. Isto significa que o crime pode aumentar sem que existam mais delinquentes, apenas pelo aumento dos alvos apropriados. Significa também que mudanças na vida da comunidade podem levar a que existam mais oportunidades de crime sem que aumente a motivação para a prática de crimes.

2.2. A Perspectiva da Escolha Racional

A perspectiva da escolha racional foca essencialmente o processo de tomada de decisão do delinquente. O principal pressuposto reside no facto de que a prática de um crime é considerada um comportamento propositado, desenvolvido para beneficiar o delinquente de alguma forma. O modelo inicial de escolha formulado para orientar a prevenção situacional

foi, mais recentemente, desenvolvido numa perspectiva de escolha racional (Clarke, 1997: 9). O pressuposto assumido na perspectiva da escolha racional considera que este comportamento propositado é desenvolvido no sentido de satisfazer uma necessidade do delinquente, seja monetária, de estatuto, sexual ou outra, e de que essa satisfação envolve tomar decisões e escolhas (muitas vezes rudimentares), constrangidas pelo facto de que existem limites de tempo, habilidade e disponibilidade de informação relevante.

Finalmente, é reconhecido que o processo de decisão e a informação utilizada varia grandemente de acordo com o tipo de crime a ser praticado. Ignorar estas diferenças e as constingências situacionais directamente associadas, pode reduzir grandemente a possibilidade de intervenção policial.

3. OS MODELOS DE POLICIAMENTO

Os modelos de policiamento têm vindo a ser objecto de transformação e reflexão profunda ao longo dos últimos anos, muito devido à premência de encontrar uma resposta adequada às novas exigências das realidades sociais, culturais e criminais. Evolução que iremos abordar, de forma breve mas objectiva, no que respeita à sua necessidade, concepção e princípios basilares.

3.1 O Modelo Tradicional

O rigor científico e o profissionalismo são as pedras basilares do modelo tradicional de polícia, sendo vistos como a única forma de salvaguardar a polícia das influências políticas e de travar a elevada corrupção policial. A resposta rápida passa a ser o paradigma do modelo tradicional de polícia, passando este modelo a ser caracterizado por um processo em que os meios são deslocados rapidamente para dar resposta aos incidentes, descurando os incidentes conexos ou os problemas que lhe estão subjacentes. As acções policiais são essencialmente reactivas e a recolha e análise de informação está limitada ao próprio incidente.

Desafios às ideias pré-concebidas sobre o valor do policiamento preventivo eram, até há poucos anos, extremamente raros. No entanto, a partir de 1962, à medida que as taxas de criminalidade se incrementavam e se começou a tomar consciência de ocorrências criminais que não eram relatadas à polícia, estes desafios começaram a tornar-se mais comuns.

Foi neste contexto que o Departamento de Polícia de Kansas City, no Missouri, Estados Unidos, através de uma bolsa atribuída pela Police Foundation, levou a cabo o mais extenso estudo científico alguma vez realizado para analisar a eficácia do patrulhamento preventivo do modelo tradicional (Kelling *et al.*, 1974). Os resultados do estudo foram tudo menos o esperado. As três condições experimentais de policiamento não tiveram qualquer efeito aparente nas taxas criminais, no serviço prestado ou na percepção que os cidadãos têm de segurança.

3.2 O Modelo de Policiamento Comunitário

A máxima do modelo de policiamento comunitário é "*the police is the public, and the public is the police*", i.e., a polícia e o público interagem com o objectivo de resolver problemas comuns. A polícia comunitária é um novo conceito que propõe uma nova abordagem filosófica, organizacional e operacional no meio urbano, e cujas ideias base residem no trabalho em parceria entre a polícia e a comunidade, no sentido de uma melhor gestão da criminalidade, da ordem pública e no desenvolvimento de programas de prevenção do crime. Este modelo tem em vista a renovação das relações entre a polícia e a população, de forma a tornar as suas necessidades das prioridades de policiamento. As implicações deste modelo assentam numa mudança dos esforços e objectivos da polícia ao focar-se tanto nos meios como nos

fins. Particularmente, na perspectiva dos meios, a polícia deve utilizar uma grande variedade de ferramentas para além do mero uso da lei criminal.

3.3 O Modelo de Policiamento Orientado para o Problema

O policiamento orientado para o problema (POP) é o resultado da obra de Herman Goldstein (1990) e do trabalho do *Police Executive Research Forum*, conduzido por John E. Eck e William Spelman (1987). A ideia por trás do POP é extremamente simples, sugerindo que o trabalho policial deveria ser, na sua essência, identificar e resolver problemas no seio da comunidade; problemas relevantes esses que normalmente chamam a atenção da polícia como sendo pequenos acontecimentos que perturbam a comunidade. Como Eck e Spelman (1987) sugerem, através deste modelo, o problema, e não a lei, passa a ser a característica que define o trabalho policial. O policiamento orientado para o problema cresceu a partir de preocupações relacionadas com o facto de que a polícia estava mais focada nos seus meios e não nos seus fins, especialmente no impacto das suas intervenções (Goldstein, 1990).

3.4 *Intelligence-led Policing*

O termo *Intelligence-led Policing* surge no início da década de 90, no Reino Unido, mais propriamente na *Kent Police*. Enquanto lógica policial destinada a reduzir a criminalidade através da pró-actividade assente na utilização de informações criminais. A sua concepção teve por base a orientação dos recursos policiais para o desenvolvimento de operações delineadas a partir da recolha e análise de informações criminais destinadas a identificar e deter indivíduos suspeitos de se dedicarem, de forma continuada, à prática de ilícitos. De acordo com Ratcliffe (2008), podemos dizer que *Intelligence-led Policing*:

- É uma filosofia de gestão / modelo de negócio;
- Procura a prevenção e a redição do crime e acabar com a actividade delinquente;
- Utiliza uma estratégia de gestão do topo para a base;
- Utiliza as informações policiais como base de apoio à decisão;
- Foca a actividade operacional nos delinquentes prolíficos.

Aliadas a estas premissas estão implícitos quatro elementos operacionais essenciais (Silva, 2008):

- A definição dos perfis dos delinquentes;
- A gestão de *hot-spots* relacionados com crimes e desordens;
- A investigação das ligações entre séries de crimes e desordens;
- A aplicação de medidas preventivas.

4. O COMANDO DISTRITAL DE POLÍCIA DE FARO

A PSP, enquanto Força de Segurança, foi criada para prosseguir uma das tarefas que constitucionalmente se encontram atribuídas ao Estado, dele fazendo parte integrante, produzindo deste modo um serviço – a segurança interna – em cooperação com outras Forças e Serviços de Segurança. A Lei Orgânica da PSP[1], estabelece que a sua missão é, em geral, assegurar a legalidade democrática, garantir a segurança interna e os direitos dos cidadãos.[2]

A Portaria n.º 340-A/2007, de 30 de Março, do Ministério da Administração Interna, veio reestruturar as áreas de responsabilidade da PSP e da GNR a nível nacional.

Com a reestruturação efectuada em 2007, a PSP passou a ser responsável, no distrito de Faro, por uma área de 265 Km^2 e uma população de 125 715 habitantes, o que corresponde a 5,3% da área geográfica do Algarve e a 29% da população residente.

[1] Aprovada pela Lei n.º 53/2007, de 31 de Agosto.
[2] Art.º 1º, n.º 2, da Lei Orgânica da PSP.

4.1 Organização e Distribuição de Meios

O modelo de policiamento adoptado no Comando Distrital de Polícia de Faro (CD Faro), à semelhança do resto do país, constitui um misto de policiamento tradicional e policiamento de proximidade. O efectivo de cada subunidade é dividido em grupos e, com maiores ou menores orientações operacionais, percorre, a pé ou de viatura, a sua área de jurisdição durante o turno de serviço, respondendo às solicitações dos cidadãos e mantendo uma visibilidade policial. Estes turnos, para o serviço operacional normal, correspondem a seis horas de serviço diário, em regime de rotatividade, designadamente:

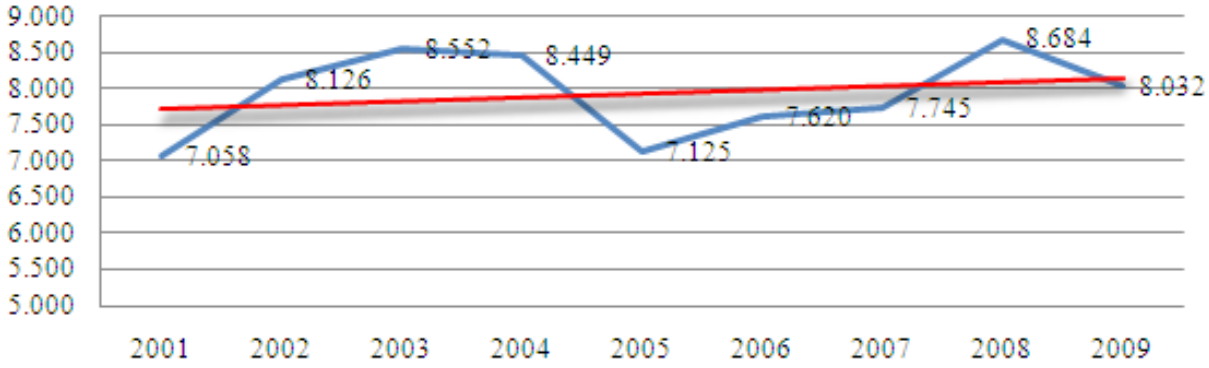
- 02H00 – 08H00;
- 08H00 – 14H00;
- 14H00 – 20H00;
- 20H00 – 02H00.

Apenas o efectivo destacado para outros tipos de policiamento, tem diferentes regimes de turnos, um pouco mais adequados aos horários dos estabelecimentos comerciais e dos estabelecimentos de ensino, ou dos períodos em que se regista maior criminalidade.

4.2 A Criminalidade Denunciada

A Direcção-Geral da Política de Justiça (DGPJ) é a entidade responsável por "*assegurar a recolha, utilização, tratamento e análise da informação estatística da justiça e promover a difusão dos respectivos resultados, no quadro do sistema estatístico nacional*." [3] Atendendo aos dados disponibilizados pela DGPJ, no que concerne à criminalidade denunciada no Distrito de Faro no ano de 2009, podemos verificar que 30,1% dos crimes foram registados na área de responsabilidade da PSP. A evolução da criminalidade denunciada na área do Comando Distrital de Faro entre 2001 e 2009 apresenta uma ligeira tendência de subida, conforme se observa no Gráfico 1.

Gráfico 1 - Evolução da criminalidade na área do CD Faro



Fonte: DGPJ (2010)

5. A METODOLOGIA DA INVESTIGAÇÃO

A análise realizada à criminalidade denunciada na área do CD Faro em 2009, e a distribuição dos meios humanos e materiais existentes, permite constatar que existem grandes discrepâncias. Considerando o atrás exposto relativamente às teorias do crime e aos modelos de policiamento, torna-se necessária uma reformulação do modelo de policiamento utilizado no Comando Distrital de Faro.

[3] Art. 2º, n.º 2, alínea c), do Decreto-Lei n.º 123/2007, de 27 de Abril.

5.1 Recolha e tratamento de dados

Os dados a ser utilizados na análise são os registados no Sistema Estratégico de Informação da PSP (SEI), relativamente a crimes ocorridos na área da PSP do Distrito de Faro, nos anos de 2008 e 2009.[4]

Não sendo exequível uma análise pormenorizada relativamente a todos os tipos de crime[5], decidimos efectuar a análise somente sobre os que respeitam as seguintes condições:

1. Maior peso relativo na criminalidade denunciada;
2. A actuação proactiva da polícia tem impacto na sua diminuição.
   Considerando as condições impostas, seleccionámos para análise os seguintes tipos de crime:

- Furto em veículo motorizado;
- Furto em residência.

5.2 A Metodologia CHAID

A metodologia CHAID é um "*método exploratório para estudar as relações entre uma variável de resposta e um conjunto de variáveis explicativas que podem interagir entre si*" (Pestana e Gageiro, 2009:366). O algoritmo CHAID, desenvolvido por Kass (1980), visa encontrar uma classificação da população em grupos capazes de descrever, da melhor maneira possível, a variável dependente. A metodologia utiliza variáveis categóricas, trabalha com tabelas de contingência e efectua o teste do χ^2 em diferentes etapas do processo, segmentando a amostra de acordo com as categorias do melhor predictor. Para limitar o crescimento da árvore são utilizados vários filtros:

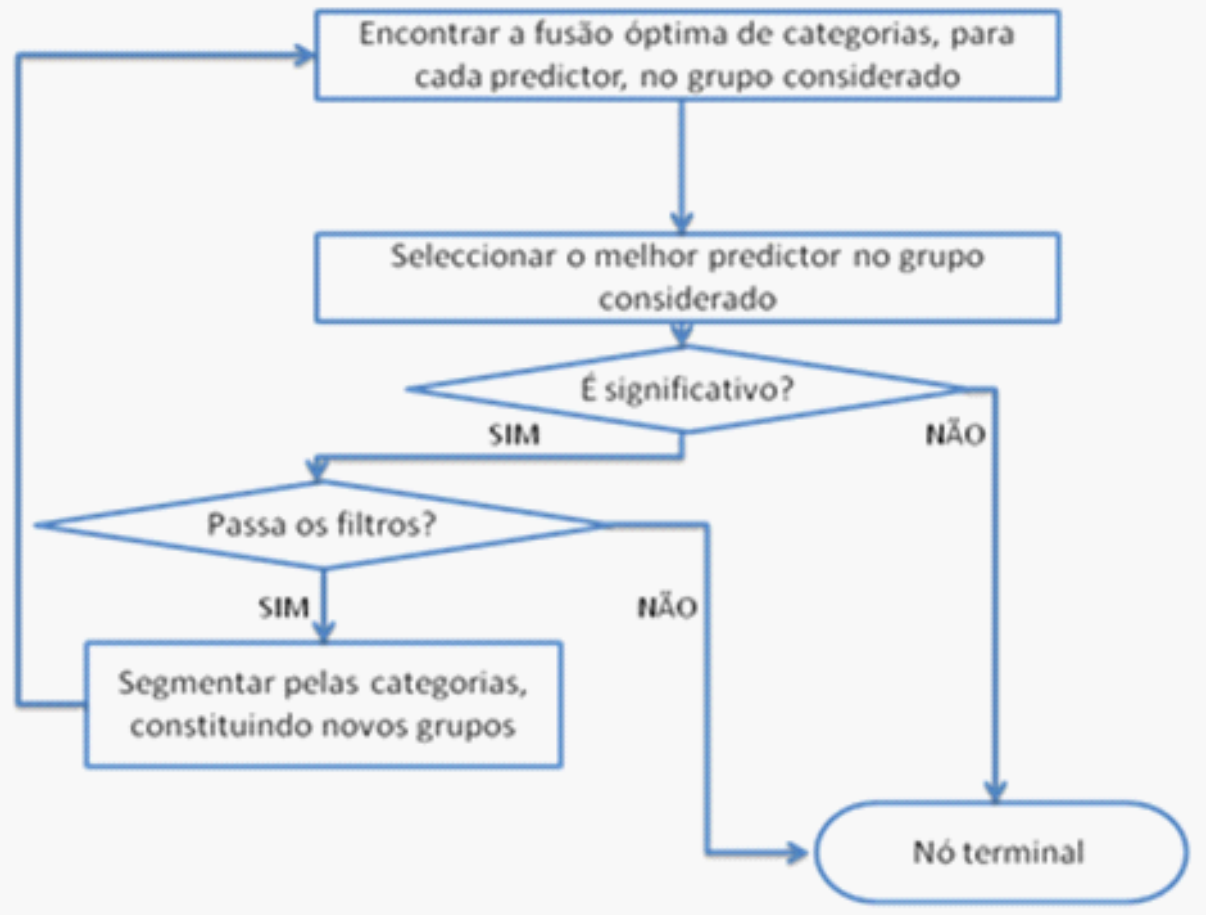
- Significância da categoria: valor do *p-value* acima do qual uma categoria é considerada não significativa e, portanto, é fundida.
- Significância do predictor: valor do *p-value* abaixo do qual uma variável é significativa para ser considerada predictor.
- Dimensão da raíz: número mínimo de elementos que um grupo tem que ter para poder ser segmentado (raíz).
- Dimensão da folha: número mínimo de elementos que um grupo tem que ter para, após a segmentação, ser um grupo autónomo (folha).
- Filtro de nível: número máximo de níveis que a árvore pode ter.

A segmentação termina quando todos os grupos são terminais, i.e., quando não existem mais prdictores significativos ou quando os filtros estabelecidos impedem o crescimento da árvore. A Figura 1 mostra o fluxograma do algoritmo desenvolvido por Kass.

[4] Apesar de o SEI estar em funcionamento desde 2004, apenas em 01 de Janeiro de 2008, a totalidade das subunidades do dispositivo territorial da PSP começaram a realizar as participações criminais no SEI. Também a notação estatística utilizada pela DGPJ apenas foi implementada no SEI a partir dessa data.

[5] De acordo com a tabela de crimes definida pela DGPJ.

Figura 1 – Algoritmo CHAID



Fonte: Própria

6. RESULTADOS

Como variável explicada utilizámos o tipo de crime com duas categorias:

- Furto em veículo motorizado;
- Furto em residência.

Como variáveis explicativas utilizámos:

1. Idade[6], com três categorias:
   - Menos de 16 anos;
   - Entre 16 e 24 anos;
   - Mais de 24 anos.
2. Sexo, com duas categorias:
   - Masculino;
   - Feminino.
3. Horário[7], com quatro categorias:
   - [07h00 – 12h40[;
   - [12h40 – 17h40[;
   - [17h40 – 22h00[;
   - [22h00 – 07h00[.

[6] As categorias utilizadas são as que são definidas pela DGPJ nas estatísticas oficiais.
[7] As categorias do horário foram definidas mediante os quartis dos crimes em análise, tomando como base as 07H00, hora a que se considerou que o dia se inicia.

Como filtros do processo de segmentação, para limitar o crescimento da árvore, utilizámos:

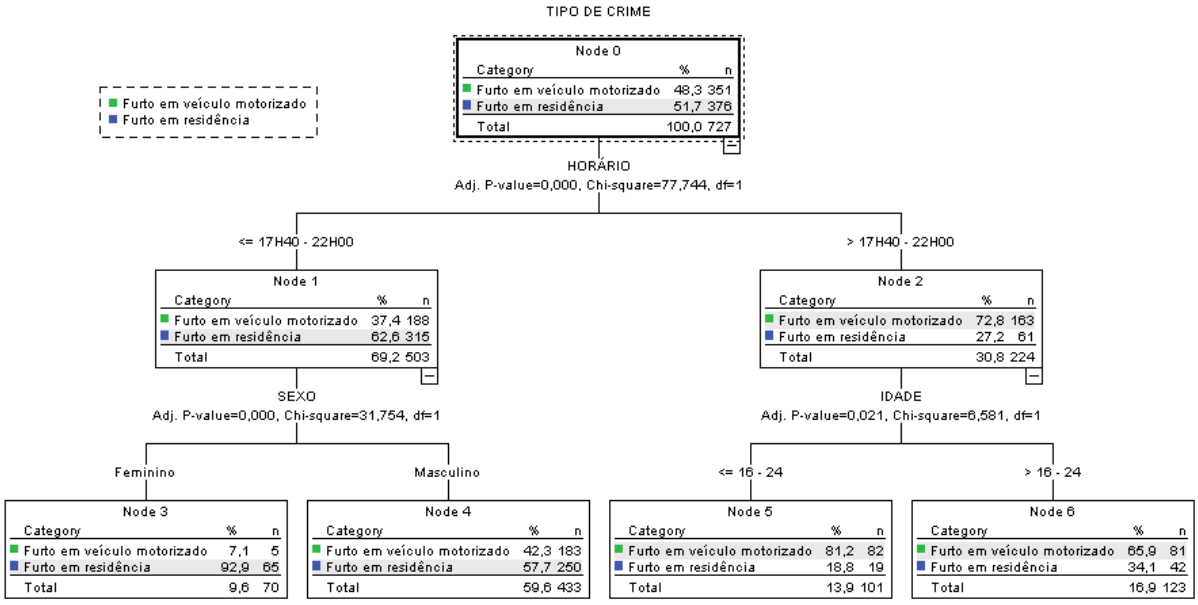
- Significância da categoria: 0,05
- Significância do predictor: 0,05
- Dimensão da raíz: 50
- Dimensão da folha: 25
- Filtro de nível: 3

Considerando que temos como objectivo a definição dos perfis dos suspeitos e das vítimas, foram desenvolvidas duas árvores, partindo de um conjunto de 727 suspeitos e 2612 vítimas.

6.1 Os Perfis dos Suspeitos

A aplicação da metodologia ao conjunto dos suspeitos resultou na árvore representada na Figura 2.

Figura 2 – Árvore dos suspeitos



Fonte: Própria

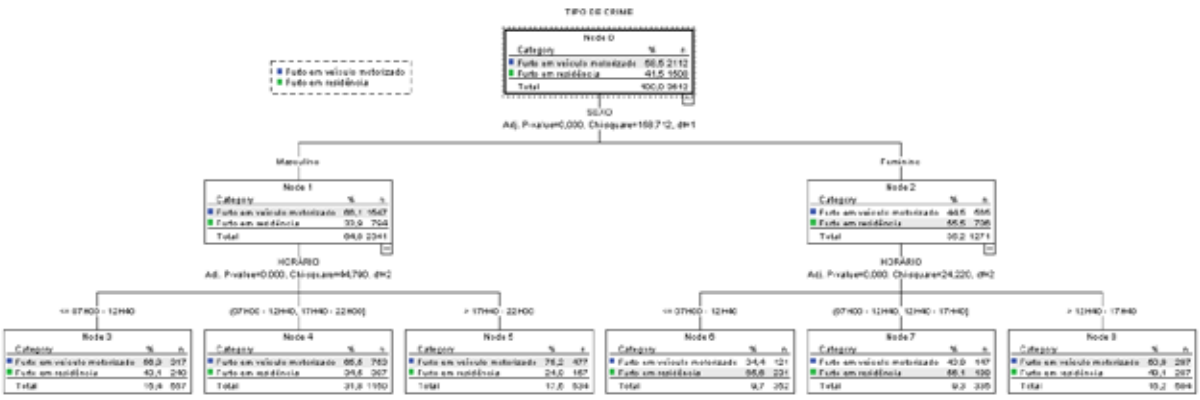
Deste modo, foi possível determinar os seguintes perfis:

- A variável que mais influencia a prática destes crimes é o horário;
- A variável que mais influencia a prática de crimes entre as 07H00 e as 22H00 é o sexo do suspeito;
- 66,5% dos furtos em residência são praticados por homens, entre as 07H00 e as 22H00;
- A variável que mais influencia a prática de crimes entre as 22H00 e as 07H00 é a idade do suspeito;
- 46% dos crimes de furto em veículo motorizado são praticados entre as 22H00 e as 07H00, distribuindo-se a idade dos suspeitos de igual forma abaixo e acima dos 24 anos;
- O risco de má classificação situa-se entre os 30,8% e os 37,8%.

6.2 Os Perfis das Vítimas

A aplicação da metodologia ao conjunto das vítimas resultou na árvore representada na Figura 3.

Figura 3 – Árvore das vítimas



Fonte: Própria

Foi, assim, possível determinar os seguintes perfis:

- A variável que mais influencia a vitimização destes crimes é o sexo, seguida do horário;
- 35,6% das vítimas de furto em veículo motorizado cometido entre as 12H40 e as 22H00 são homens;
- 28% dos furtos em residência são praticados entre as 07H00 e as 17H40, e as vítimas são mulheres;
- 52,5% dos furtos em veículo motorizado com vítimas femininas, são praticados a partir das 17H40;
- O risco de má classificação situa-se entre os 35,7% e os 38,8%.

7. CONCLUSÕES

A metodologia CHAID permitiu analisar as relações entre o tipo de crime e um conjunto de variáveis explicativas (características dos suspeitos e vítimas e horários de cometimento dos crimes). O conhecimento destes perfis criminais possibilita, dentro do actual modelo de policiamento aplicado pela PSP no Algarve, sugerir adaptações à realidade da criminalidade local, assegurando assim que a polícia estará no sítio certo, na hora certa, dissuadindo o cometimento de ilícitos criminais, permitindo ainda lançar campanhas de sensibilização às vítimas, de acordo com os perfis agora definidos. Assim, com base nos dados apurados, sugerimos:

- A reafectação dos elementos policiais pelas subunidades mais críticas;
- A alteração dos horários de serviço;
- A criação de campanhas de sensibilização para as vítimas.

A realização de estudos mais aprofundados, por subunidade e com mais variáveis ou categorizações diferentes, garantirá uma, ainda mais, eficiente alocação de recursos operacionais.

BIBLIOGRAFIA

Belson, W.A. (1959), Matching and Prediction on the Principle of Classification, *Journal of Applied Statistics*, 8, 195-202.

Clarke, R.V. (1997) *Situational Crime Prevention: Successful Case Studies*, 2nd Edition, New York: Criminal Justice Press

Cohen, Lawrence E. and Felson, M. (1979) Social change and crime rate trends: A routine activity approach, *American Sociological Review*, 44, 588-607.

Eck, John, Spelman, William (1987) *Problem-Solving: Problem-Oriented Policing in Newport New,* Washington, D.C.: Police Executive Research Forum.

Felson, Markus, Poulsen, Erika (2003) Simple indicators of crime by time of day, *International Journal of Forecasting*, 19, Elsevier B.V., 595-601

Goldstein, Herman, (1990) Problem Oriented Policing, New York: McGraw-Hill Ltd.

Hawkins, D.M. and Kass, G.V. (1982) Automatic Interaction Detection, D.M. Hawkins (ed.), *Topics in Applied Multivariate Analysis*, Cambridge: Cambridge University Press.

Kelling, George, Pate, Tony, Dieckman, Duane, Brown, Charles (1974) The Kansas City Preventive Patrol Experiment – A Summary Report, Washington: Police Foundation.

Pestana, M.H. e Gageiro, J.N. (2009) Análise Categórica, Árvores de Decisão e Análise de Conteúdo, Lisboa: Lidel.

Ratcliffe, Jerry (2008) Intelligence-Led Policing, Devon: Willan Publishing.

A EVOLUÇÃO DA INVESTIGAÇÃO EM TURISMO: UMA NOVA PLATAFORMA DE CONHECIMENTO?

THE EVOLUTION OF TOURISM RESEARCH: A NEW KNOWLEDGE PLATFORM?

Iris Lopes
João Albino Silva
Efigénio Rebelo
Guilherme Castela

RESUMO

Dada a reputação internacional do *Annals of Tourism Research* como revista académica de referência no campo do Turismo e uma vez que apresenta, desde a sua primeira edição, um índex temático extremamente consistente e abrangente, foi possível a criação de uma base de dados que permitiu observar a evolução do conhecimento em Turismo no período compreendido entre 1973 e 2002.

A partir da identificação das áreas temáticas fundamentais que foram objecto de pesquisa, no período temporal referido, e mediante um processo de redução da dimensionalidade suportado por uma Análise Factorial de Componentes Principais (AFCP), a aplicação da metodologia BiPlot permitiu detectar os temas mais relevantes que, agrupados sob a forma de variáveis compósitas, traduziram as tendências de investigação.

As conclusões apresentadas são consistentes com anteriores análises de conteúdo e com o modelo conceptual de crescimento do conhecimento em Turismo proposto por Jafari (2001): as quatro plataformas de conhecimento em Turismo.

Pretende-se no futuro efectuar análises mais alargadas, de modo a generalizar os resultados ao campo de estudo do Turismo, com a contribuição de mais revistas e incluindo a maior diversidade possível de temas.

Palavras-chave: "Investigação em Turismo", "Métodos BiPlot".

ABSTRACT

By analysing *Annals of Tourism Research* subject index, this study aims to draw conclusions about the evolution/trends of tourism research. Because of *Annals's* reputation as the leading international scholarly journal in the field, and also because of its consistent and comprehensive subject index, it was possible to create a database that allowed to observe the evolution of the field between 1973 and 2002. From the identification of the core subject themes, and through a dimension reduction process, the use of the BiPlot methodology allowed us to identify the most relevant subject themes, which show the main research trends between 1973 and 2002.

The results of this study are consistent with previous studies and the observed trends in subjects revealed a relatively close match to Jafari's (2001) four platforms as an emergent paradigm for tourism knowledge.

For future research similar analyses including not only other tourism journals but also a bigger number of subject themes are needed, to expand the findings to the field as a whole.

Keywords: "Tourism Research", "BiPlot Methods".
JEL classificacion: C49

1. INTRODUÇÃO

Apesar de Franklin e Crang (2001) descreverem a investigação em Turismo como "stale, tired, repetitive, and lifeless", de Page (2005) reiterar a opinião dizendo que «if only 25% of the current tourism outputs were produced, our knowledge base in subject would not be adversely affected», e de Oppermann (2000) a considerar «of limited additional scientific value» (citados por Xiao e Smith, 2008), tem havido nas últimas décadas um interesse crescente em avaliar a sua evolução.

Xiao e Smith (2006) defendem que a evolução do conhecimento em turismo pode ser avaliada sob três perspectivas: sob a perspectiva específica de cada país; sob a perspectiva das disciplinas que contribuem para o estudo do turismo; e, sob a perspectiva das revistas científicas já que, em sua opinião, «academic journals form an important platform on which the evolution of a field can be examined».

Estes autores são peremptórios quanto à importância da monitorização das revistas científicas afirmando «Although general discussions of the nature and evolution of scientific knowledge (Kuhn 1962), paradigmatic and disciplinary debates (Boterill 2001; Leiper 2000; Przeclawski 1993; Tribe 1997), and state-of-the-art tourism research (Dann 2000, 2003; Jafari 1990; Riley and Love 2000) can provide hypotheses regarding the evolution of tourism knowledge, the empirical study of the content of its journals provides verifiable and grounded conclusions about the evolution of research» (Xiao e Smith, 2006).

Também Van Doren e Heit (1973), citados por Xiao e Smith (2006) «noted the importance of monitoring academic journals from time to time in order to recognize trends and to critically appraise their contributions to the advancement of knowledge».

Assim, no campo do turismo podem ser encontrados inúmeros trabalhos que tiveram por base este tipo de análise; a título de exemplo, enumeram-se os seguintes (Xiao e Smith, 2006, 2008):

- Reid and Andereck (1989), Faulkner e Ryan (1999), Riley e Love (2000) e Xiao e Smith (2006), descreveram o uso de técnicas estatísticas em revistas de turismo;
- Pechlaner, Zehrer, Matzler e Abfalter (2004), Sheldon (1990), Jogaratnam, Chon, McCleary, Mena e Yoo (2005), Zhao e Ritchie (2007) discutiram as percepções e o *ranking* de investigadores académicos, instituições e revistas de turismo, no seio da comunidade académica;
- Howey, Savage, Verbeeten e Van Hoof (1999), Van Doren, Koh e McCahill (1994) e Xiao e Smith (2005, 2006) efectuaram análise de citações em revistas de turismo;
- Baloglu e Assante (1999) e Crawford-Welch e McCleary (1992), realizaram observação longitudinal tendo em vista a detecção de alterações nas áreas temáticas e nas técnicas de pesquisa utilizadas (no campo da hospitalidade);
- Swain et al (1998) e Xiao e Smith (2006), documentaram alterações nas áreas temáticas.

Muito embora se possam considerar estes trabalhos como um "inward-looking" já que como comentava McKercher (2005) «few people outside the field consult this literature unless conducting tourism research, and by the same token, those working within the field cite the literature constantly», a verdade é que as revistas científicas são meios de comunicação essenciais para os investigadores e a sua monitorização é essencial para o entendimento da evolução da investigação: «Academic journals are essential communication channels for researchers [...] can also serve as a lens through which the evolution of the field can be viewed» (Xiao e Smith, 2006).

É pois inquestionável que uma das formas encontrada nos últimos anos para avaliar a evolução do conhecimento em turismo se prende com a análise das revistas científicas. Contudo, os estudos anteriores têm-se revelado limitados, quer em termos dos períodos temporais quer em termos do número de revistas analisados, inviabilizando a generalização

de resultados ao domínio em estudo. Assinale-se ainda o facto de a metodologia utilizada se remeter quase exclusivamente a análises de conteúdo, e a contagens de citações e número de publicações por autor/instituições académicas.

2. OBJECTIVOS

Xiao e Smith (2006) realizaram uma análise de conteúdo, para o período de 1973 a 2003, do índice temático do *Annals of Tourism Research*, uma das revistas incontornáveis da área do Turismo dada a sua reputação internacional e também devido ao facto de apresentar, desde a sua primeira publicação, um índice temático extremamente consistente e abrangente. Entre as várias conclusões que apresentam, indicam que o seu trabalho é largamente consistente com anteriores análises de conteúdo (realizadas sobre o *Annals* e/ou sobre outras revistas científicas), e com o modelo conceptual de crescimento do conhecimento em Turismo proposto por Jafari (2001): as quatro plataformas de conhecimento em Turismo.

Os autores reconhecem no entanto as limitações do seu trabalho uma vez que a análise efectuada se baseia numa única revista, o *Annals*, o que obviamente limita a generalização dos resultados encontrados, e sugerem que um desenvolvimento essencial da investigação assentaria na realização de análises similares que incluísse outras revistas de turismo. Esta extensão, nas palavras dos autores, iria permitir uma comparação e/ou validação dos resultados previamente encontrados.

Queremos, num futuro próximo, dar resposta ao desafio lançado por Xiao e Smith. No entanto, considerando que estamos numa fase inicial do nosso projecto de investigação, pretendemos apenas, nesta primeira etapa, testar uma metodologia inovadora nesta área.

Assim, a partir do índex temático do *Annals of Tourism Research*, foi criada uma base de dados que permitiu observar a evolução do conhecimento em Turismo, no período compreendido entre 1973 e 2002, e os resultados foram comparados com os resultados a que chegaram outros autores (nomeadamente Xiao e Smith).

3. METODOLOGIA

Reconhecida a impossibilidade e relevância, nesta fase inicial do nosso projecto de investigação, de trabalhar com os 857 temas principais identificados no índice temático do *Annals* (para além dos 1100 sub-temas e das 573 referências cruzadas), a opção recaíu em onze temas cuja selecção foi sugerida pela revisão bibliográfica: Economia, Alojamento, Desenvolvimento, Metodologia, Atitude, Turismo Alternativo, Autenticidade, Comportamento, Procura, Destino e Marketing. A esta base de dados, trinta anos de publicações nos onze temas escolhidos, foi aplicada, numa primeira fase, a metodologia BiPlot (Gabriel, 1971; Galindo, 1986; Gabriel e Odoroff, 1990) mediante um processo de redução de dimensionalidade, suportado por uma Análise Factorial de Componentes Principais (AFCP). E, numa segunda fase, foi aplicado um procedimento de segmentação hierárquica como forma de validação dos agrupamentos BiPlot detectados anteriormente.

Esta proposta de análise multidimensional, simultaneamente descritiva e exploratória, assenta na interpretabilidade espacial das estruturas Biplot e na observação das relações existentes entre observações (1973-2002), entre variáveis (os 11 temas seleccionados) e entre observações e variáveis. Estes procedimentos têm como consequência o facto de todos os resultados traduzidos desta forma poderem, naturalmente, ser representados por configurações de marcadores de variáveis e observações simultaneamente, expressas em coordenadas principais e com elevada qualidade de representação, num espaço de baixa dimensão. Em particular, foi utilizada a metodologia, GH-Biplot ou CMP_Biplot (*Column*

Metric Biplot) que garante a máxima bondade de ajustamento só para as variáveis. Julgamos assim possível obter, dada a alta fiabilidade dos resultados, uma análise mais adequada, na qual o posicionamento espacial simultâneo das observações e das variáveis permite evidenciar padrões e mudanças no que concerne à variação conjunta das 11 áreas de investigação em turismo para o período temporal em causa e, deste modo, promover uma melhoria da informação conducente ao estudo da evolução do conhecimento em Turismo.

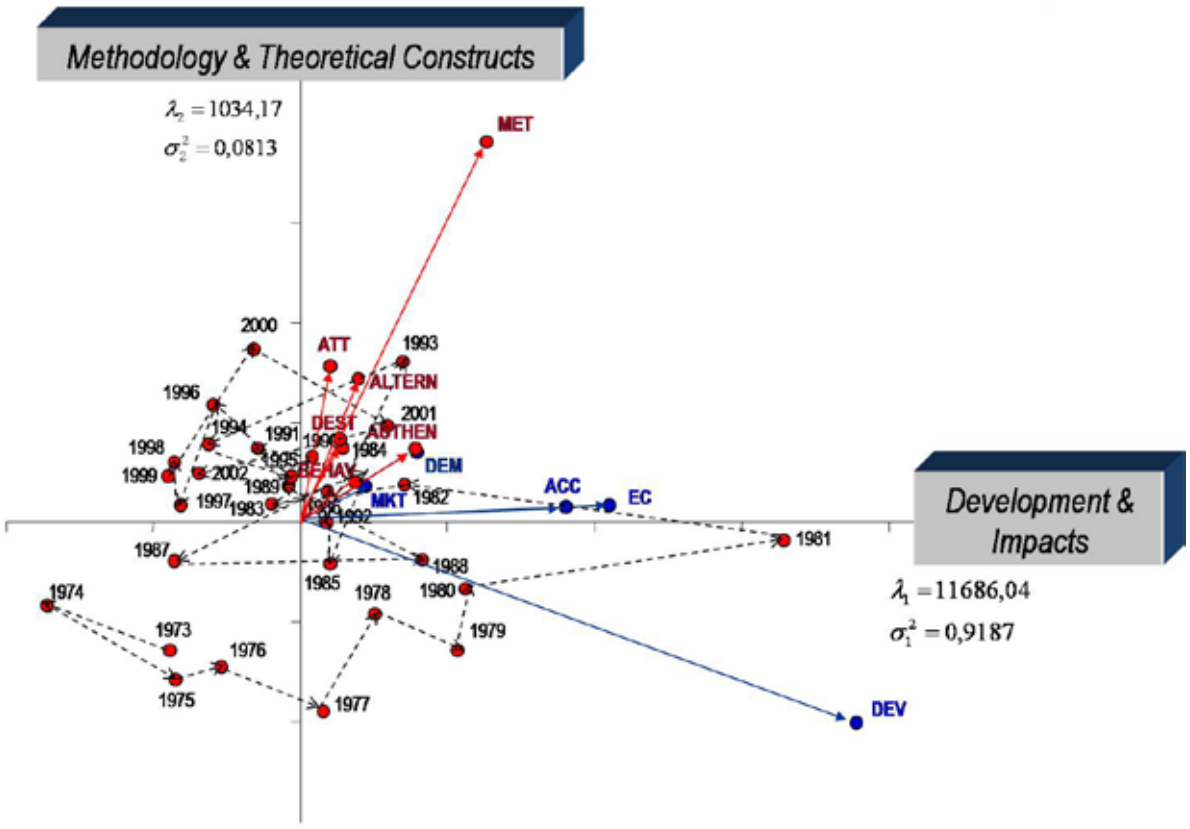
Estas coordenadas GH-Biplot, conduziram-nos a um processo de segmentação hierárquica, tendo por base o método de Ward, que validaram quatro agrupamentos homogéneos de temas principais, os quais possibilitaram detectar os temas mais relevantes que, agrupados sob a forma de variáveis compósitas, traduziram as tendências de investigação.

4. RESULTADOS

A aplicação desta metodologia permitiu chegar a conclusões consistentes com o trabalho de Xiao e Smith (2006), nomeadamente o facto de terem sido encontradas duas dimensões do *Conhecimento em Turismo* (Figura 1):

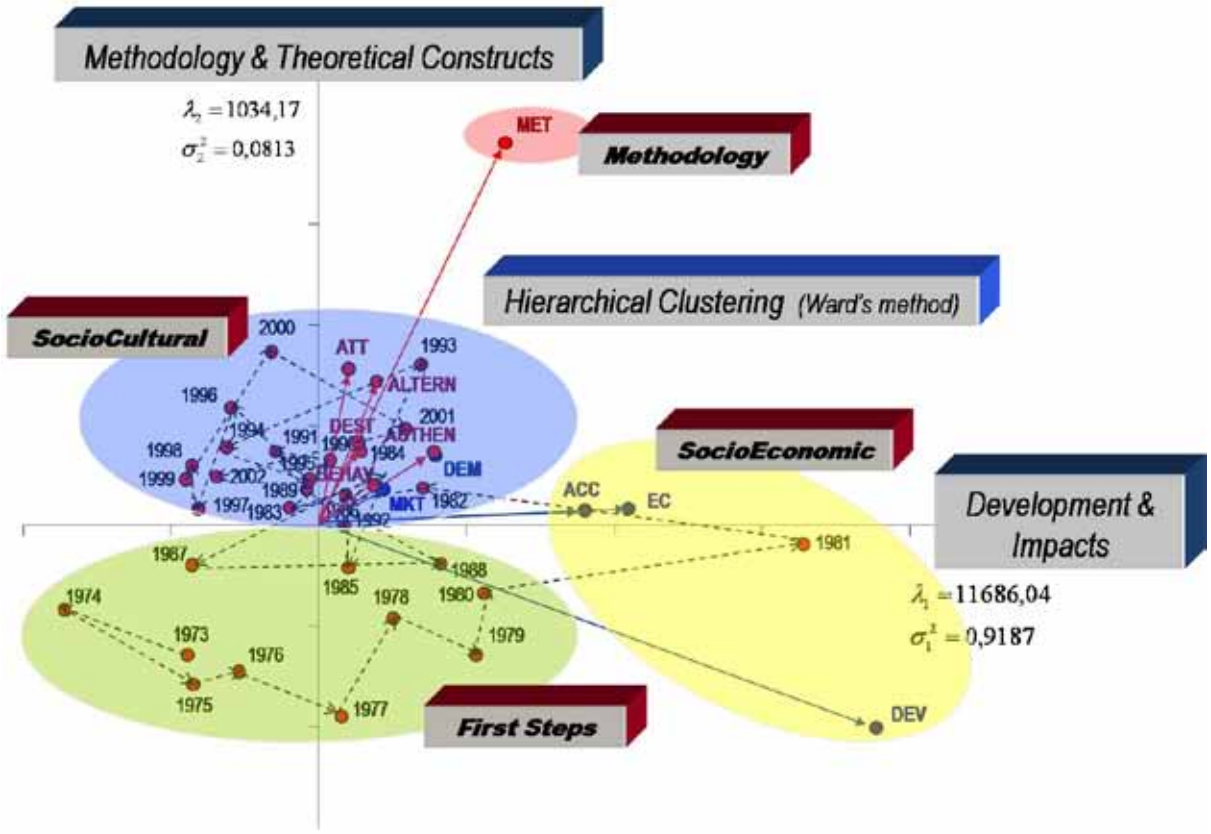
- uma primeira dimensão, *Development and Impacts*, com 92% da capacidade explicativa, à qual estão associados cinco dos onze temas sujeitos a análise: Economia, Alojamento, Desenvolvimento, Procura e Marketing;
- uma segunda dimensão, *Methodology and Theoretical Constructs*, com apenas 8% da capacidade explicativa, a que estão associados os restantes temas: Metodologia, Atitude, Turismo Alternativo, Autenticidade, Comportamento e Destino.

Figura 1: Estrutura multidimensional GH-Biplot



A aplicação dos clusters hierárquicos, permitiu identificar a existência de quatro grupos temáticos diferentes, também estes anteriormente referenciados na bibliografia (Figura 2):

Figura 2: Segmentação hierárquica com coordenadas GH-Biplot



- *First Steps*: "… tourism knowledge has gone through an evolution of formulations, beginning in a somewhat inarticulate form, struggling with definitions and the establishment of basic tenets" (Swain et al,1998).
- *Socio Economic*: "… the focus shifted to management, economics, and socioeconomic perspectives in the early 80s" (Xiao e Smith, 2006).
- *Socio Cultural*: "… the decline in economically-oriented studies has been offset by a rise in studies of sociocultural issues, community development, and environment" (Xiao e Smith, 2006).
- *Methodology*: "… the growth of knowledge around methodology and theoretical constructs is indicative of the efforts of a young field to achieve rigor and to be recognized as a "serious" field of enquiry" (Xiao e Smith, 2006).

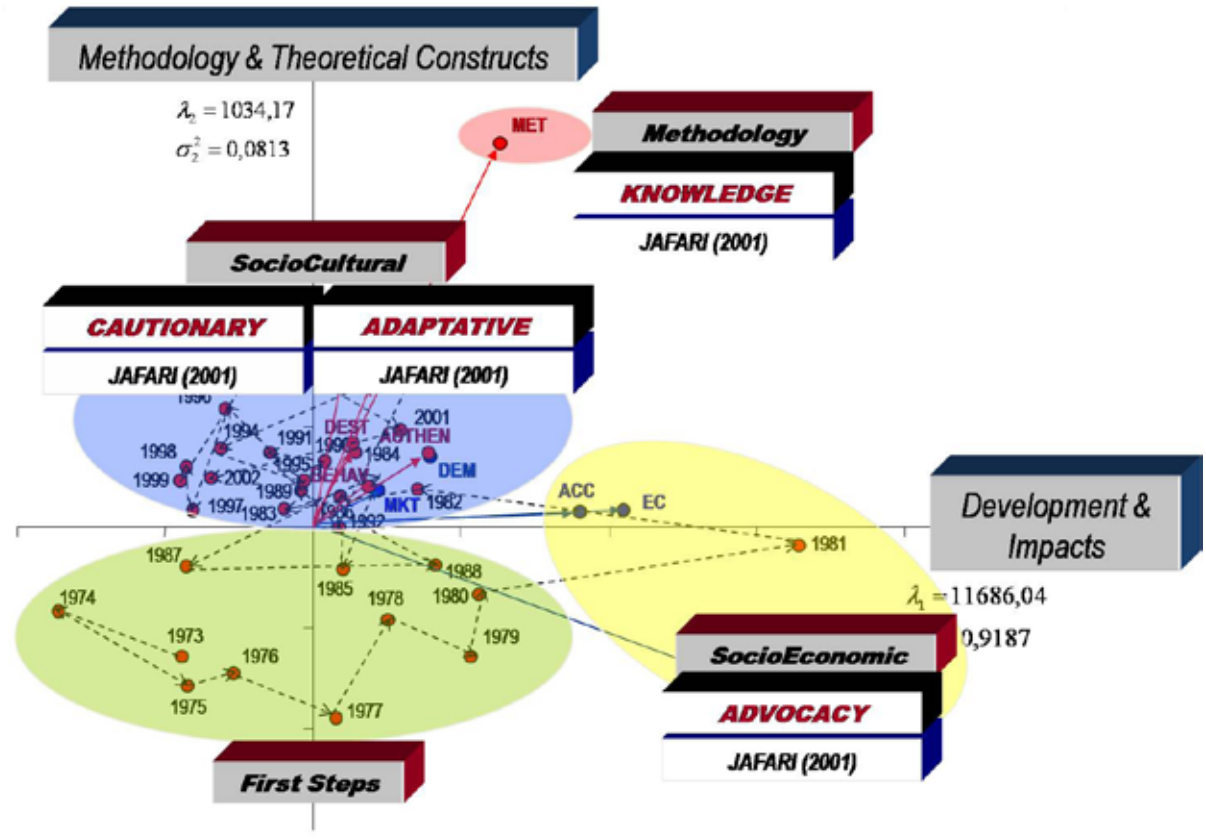
Na Figura 3, pode observar-se que as tendências encontradas ao longo dos 30 anos estão de acordo com as quatro plataformas para o conhecimento do Turismo, observadas por *Jafari (2001),* ou seja,

1) a plataforma de defesa (*advocacy platform*); 2) a de advertência (*cautionary platform*); 3) a de adaptação (*adaptancy platform*); e 4) a plataforma baseada no conhecimento (*knowledge-based platform*). Estas plataformas embora tenha surgido cronologicamente, pela ordem apresentada, não se substituíram entre si, ou seja, actualmente, as quatro coexistem sendo a última a responsável pelo *lançamento académico* do turismo a nível mundial (Jafari, 2005).

A plataforma de defesa surge na década de 60 a promover os benefícios económicos gerados pela actividade turística, em suma, a enaltecer as contribuições da "indústria" para o crescimento e desenvolvimento económicos. O enfoque de advertência surge na década de 70, em oposição à plataforma anterior, articulado maioritariamente por antropólogos e sociólogos que chamaram a atenção para as consequências sócio-culturais, questionando o turismo como motor de desenvolvimento. A plataforma de adaptação tem a sua origem nos primeiros anos da década de 80 e representa uma posição intermédia entre as duas descritas anteriormente, examinando formas alternativas de desenvolvimento potencialmente geradoras de benefícios e, ao mesmo tempo, minimizadoras de custos.

A plataforma baseada no conhecimento inicia-se nos anos 90 e surge a partir da inter-relação e compreensão das limitações das perspectivas anteriores, contemplando o turismo como um todo, o mesmo é dizer, como um sistema, com o objectivo de entender as suas estruturas e funções subjacentes. Em suma, trata de adoptar-se um enfoque holístico para o estudo e análise do turismo, não se limitando à consideração dos seus impactos ou formas, cujo objectivo principal é gerar um corpo de conhecimentos científicos: «a multidisciplinary thrust in nature and scope, nurturing the scientification of tourism» (Jafari, 2005:2).

Figura 3: As 4 plataformas do *Annals Index* versus as 4 plataformas para o conhecimento em Turismo



5. CONCLUSÕES

O presente trabalho permite afirmar que:

- Os resultados encontrados são consistentes com trabalhos anteriores;
- O crescimento de artigos de Metodologia e de Constructos Teóricos, reflectem o

esforço de obtenção de rigor científico;
- A acumulação de conhecimentos em Desenvolvimento e Impactos reflecte a visão do Turismo como um fenómeno multidisciplinar;
- As tendências encontradas ao longo dos 30 anos estão de acordo com as quatro plataformas para o conhecimento do Turismo, observadas por *Jafari (2001).*

Deste modo, ficou comprovado que a metodologia BiPlot é adequada para a detecção das tendências já referenciadas na investigação em Turismo. Consideramos assim que esta é uma alternativa aos métodos de investigação tradicionalmente utilizados pelo que pretendemos, no futuro, utilizá-la em estudos mais abrangentes.

Assim, é nossa intenção efectuar análises mais alargadas, de modo a generalizar os resultados ao campo de estudo do Turismo, com a contribuição de mais revistas e incluindo a maior diversidade possível de temas.

BIBLIOGRAFIA

Gabriel, K.R. (1971), The Biplot Graphic of Matrices with Application to Principal Component Analysis, *Biometrika*, 58, pp. 453-467.

Gabriel, K.R. e C.L Odoroff (1990), Biplots in biomedical research, *Statistics in Medicine*, 9(5), pp. 469-485.

Galindo, M.P. (1986). *Una alternativa de representación simultánea: HJ-biplot*. Questíio, 10 (1), 13-23.

Jafari, J. (2001). The Scientification of Tourism. *In* ***Hosts and Guests Revisited: Tourism Issues of the 21st Century***, V. Smith e M. Brent, eds., Elmsford NY: Cognizant Communication Corporation, 28-41.

Jafari, J. (2005). Bridging Out, Nesting Afield: Powering a new platform. *The Journal of Tourism Studies*, v. 16 (2).

McKercher, B. (2005). A Case for Ranking Tourism Journals. *Tourism Management*, v. 26, 649-651.

Swain, M. e H. Xiao (2003). Annals Index: Volumes 1 through 30 (1973- 2003). *Annals of Tourism Research*, v. 30 (4), 983-1048.

Swain, M., M. Brent e V. Long (1998). Annals and Tourism Evolving: Indexing 25 years of Publication. *Annals of Tourism Research*, v. 25, 991-1014.

Xiao, H. e S.L.J. Smith (2006). The Making of Tourism Research: Insights from a Social Sciences Journal. *Annals of Tourism Research*, v. 33 (2), 490-507.

Xiao, H. e S.L.J. Smith (2008). Knowledge Impact: An Appraisal of Tourism Scholarship. *Annals of Tourism Research*, v. 35 (1), 62-83.

NORMAS EDITORIAIS DOS *DISCUSSION PAPERS*

Com o objectivo de facilitar a publicação dos *Discussion Papers*, solicita-se ao(s) autor(e)s do(s) artigo(s) que, na elaboração dos seus textos, **siga(m) atentamente as normas adiante enunciadas.** Chama-se a atenção de que os textos terão de ser enviados na sua forma definitiva.

Assim, o documento final deve conter as seguintes normas editoriais:

Layout do documento:
Página A4 com 2,5cm na margem direita, esquerda, superior e inferior.

Formato do documento:
Os manuscritos devem ser entregues em formato *Word* com a fonte *Times New Roman* e espaçamento simples. O documento não deve ultrapassar as 25 páginas, sendo obrigatório entregar uma versão em PDF do mesmo documento.

Informação sobre o(s) autor(es):
A informação sobre o(s) autor(es) do artigo deve ser fornecida numa página única, obrigatoriamente em inglês, com um texto até 100 palavras, incluindo a Universidade a que está ligado e o seu endereço de e-mail. Para este texto usar tamanho 11, normal, justificado.

Título:
O título deve ser conciso e informativo em versão portuguesa e inglesa. Usar letras capitais com tamanho 15, negrito e alinhado à esquerda.
Depois do título adicionar o nome(s) do autor(es) em tamanho 11, itálico, alinhado à esquerda.

Resumo:
O resumo deve ter entre 150 a 200 palavras e não deve conter nenhuma abreviatura, sendo obrigatório uma versão em português e outra em inglês em tamanho 11, normal, justificado.
Palavras-chave: indicar até 4 palavras-chave separadas por "," no fim do resumo.
Também deve ser indicada a classificação JEL. Este sistema de classificação é preparado e publicado pelo *Journal of Economic Literature*. Para mais informação consultar www.aeaweb.org/journal/jel_class_system.html.

Formatação do corpo de texto:
Para o corpo de texto comum usar tamanho 11, normal, justificado.
Subtítulos 1º nível - tamanho 11, negrito, letras capitais, alinhado à esquerda;
Subtítulos 2º nível - tamanho 11, negrito, letras normais, alinhado à esquerda;
Subtítulos 3º nível - tamanho 11, itálico, letras normais, alinhado à esquerda.

Índice:
No início do documento deve constar um índice em tamanho 11, normal, alinhado à esquerda.

Abreviações:
As abreviações devem ser escritas por extenso usando parêntesis aquando da sua primeira menção e usadas coerentemente a partir desse ponto.

Notas de rodapé:
As notas de rodapé devem ser usadas para fornecer informação adicional. Não podem conter imagens ou tabelas e devem ter tamanho 8, normal, alinhadas à esquerda, sendo sempre numeradas consecutivamente.Notas de rodapé relativas ao título devem ser indicadas com o símbolo (*).
Não é permitido o uso de notas de fim de texto.

Agradecimentos:
Agradecimentos e dedicatórias devem ser colocados numa secção separada antes do início das referências bibliográficas. Siglas de organizações e afins devem estar escritas por extenso.

Citações:
Ao citar excertos de texto, adicione as referências no fim do mesmo, mencionando apenas o último nome do autor e o ano de publicação da obra entre parêntesis.
Por exemplo: (Flores *et al.*, 1988; Winograd, 1986; Cunha e Cintra, 1996).
Mas se citar o autor dentro do texto, deve apenas mencionar o ano da publicação da obra entre parêntesis.
Ex: Winograd (1986) argumenta …

Entradas bibliográficas:
As entradas bibliográficas devem apenas incluir obras que são mencionadas no texto.
Não usar notas de rodapé ou notas de fim de texto para substituir entradas bibliográficas.
As entradas bibliográficas devem estar ordenadas alfabeticamente pelo último nome do primeiro autor de cada obra.
Para formatar as entradas bibliográficas usar tamanho 11, normal, alinhadas à esquerda.
O nome de todos os autores deve ser indicado, mas também pode ser usada a designação *et al*.
Ex: Pierzynski, G. *et al*. (1994). *Soils and environmental quality*. Lewis Publishers. Florida.

Artigo científico:
Último nome do autor, Primeira inicial. (Ano da publicação). Título do artigo. *Título da obra*. **Volume(Número)**: primeira página-última página.
Ex: Ramírez, P.M., Castro, E. e Ibáñez, J.H. (2001). Reutilização de águas residuais depuradas provenientes da ETAR de Albacete (S.E. Espanha) em campos hortícolas. *Tecnologias do Ambiente*. **44(2)**: 48-51.

Livro:
Último nome do autor, Primeira inicial. (Ano da publicação). *Título do livro*. Informação adicional. Nº da edição, Editora. Cidade da publicação.
Ex: Costa, J. (1995). *Caracterização e constituição do Solo*. 5ª edição, Fundação Calouste Gulbenkian. Lisboa.

Capítulo de livro:
Ex: Silko, L.M. (1991). The man to send rain clouds. Em: W. Brown e A. Ling (eds.), *Imagining America: Stories from the promised land*. Persea. New York.

Documento online:
Último nome do autor, Primeira inicial. (Ano da publicação). *Título do documento*. Acedido em: dia, mês, ano, em: URL.
Ex: Chou, L., McClintock, R., Moretti, F. e Nix, D.H. (1993). *Technology and education: New wine in new bottles – Choosing pasts and imagining educational futures.* Acedido em 24 de Agosto de 2000, no *Web site* da: Columbia University, Institute for Learning Technologies: http://www.ilt.columbia.edu/publications/papers/newwine1.html.

Dissertação de Mestrado ou Doutoramento:
Ex: Tingle, C.C.D. (1985). *Biological control of the glasshouse mealybug using parasitic hymenoptera*. Ph.D. Thesis. Department of Biological Sciences, Wye College, University of London. 375 pp.

Tabelas, Figuras, Gráficos e Quadros:
Todas as tabelas, figuras, gráficos e quadros devem ser numerados com numeração árabe e devem ter um título explicativo antes do seu conteúdo em tamanho 9, negrito e centrado.
A fonte e ano da informação fornecida deve ser colocada abaixo do respectivo corpo, centrado, com tamanho 8, normal.
Para o conteúdo das tabelas e quadros deve ser usado o tamanho 8.
Figuras e gráficos devem ser fornecidos em formato JPEG (imagem).

EDITORIAL NORMS FOR DISCUSSION PAPERS

In order to simplify the editors' task, authors are urged to **adopt the norms listed below** for the publication of the Discussion Papers. Please note that texts should be sent in their final version.

Being so, the final document should have the following editorial norms:

Page layout:
Paper size A4, 2,5cm left, right, bottom and top margins.

Document format:
Manuscripts should be submitted in Word file using font Times New Roman and single line spacing. The document should not have more than 25 pages, and a PDF version of the document must be provided.

Author(s) information:
The author(s) information of the article must be given in english on a single page with a text up to 100 words, including affiliation and e-mail address. Use size 11, regular, justified.

Title:
Should be concise and informative, and must be given in portuguese and english with size 15, bold, left aligned and in capital letters.
After title add author(s) name(s) in size 11, italic, left aligned.

Abstract:
The abstract should have between 150 to 200 words and should not contain any undefined abbreviations.
It is necessary a portuguese and an english version, using size 11, regular, justified.
Keywords: up to 4 keywords separated by "," at the end of the abstract.
An appropriate number of JEL code(s) must be provided. This classification system is prepared and published by the *Journal of Economic Literature*.
For more information about this, please see www.aeaweb.org/journal/jel_class_system.html.

Plain Text body:
For plain text body use size 11, regular, justified.
Subtitles 1st level - size 11, bold, capital letters, left aligned;
Subtitles 2nd level - size 11, bold, regular letters, left aligned;
Subtitles 3rd level - size 11, italic, regular letters, left aligned.

Table of Contents:
A Table of Contents should be provided at the beginning of the manuscript. Use size 11, regular, left aligned.

Abbreviations:
Abbreviations should be defined at first mention and used consistently thereafter.

Footnotes:
Footnotes can be used to give additional information. They should not contain any figures or tables and should be in size 8, regular, left aligned.
Footnotes to the text are numbered consecutively.
Footnotes to the title of the article are given with the reference symbol (*).
Endnotes can not be used.

Acknowledgments:
Acknowledgments of people, grants, funds, and others, should be placed in a separate section before the reference list. The names of funding organizations should be written in full.

Citations:
After quoting a text extract, cite the reference giving only the author's name and publication year in parentheses. Ex: (Flores *et al.*, 1988; Winograd, 1986; Cunha and Cintra, 1996)
But if you are citing the author inside the text, add only the publication year between parentheses. Ex: Winograd (1986) describes …

References list:
The list of references should only include works that are cited in the text.
Do not use footnotes or endnotes as a substitute for a reference list.
Reference list entries should be alphabetized by the last name of the first author of each work.
To format reference list use size 11, regular, left aligned.
Ideally, the names of all authors should be provided, but the usage of *et al.* in long author lists will also be accepted.
Ex: Pierzynski, G. *et al*. (1994). *Soils and environmental quality*. Lewis Publishers. Florida.

Scientific article:
Last name of the author, First initial. (Publication year). Article title. *Title of the Journal or Review*. **Volume(Issue)**: first page-last page.
Ex: Sadiq. M. e Alam, I. (1997). Lead contamination of groundwater in an industrial complex. *Water, Air and Soil Pollution*. **98(2)**: 167-177.

Book:
Last name of the author, First initial. (Publication year). *Book title.* Adicional information. Edition number, Publishing house. Publishing place.
Ex: Costa, J. (1995). *Caracterização e constituição do Solo*. 5th edition, Fundation Calouste Gulbenkian. Lisbon.

Book chapter:
Ex: Silko, L.M. (1991). The man to send rain clouds. In: W. Brown and A. Ling (eds.), *Imagining America: Stories from the promised land*. Persea. New York.

Online document:
Last name of the author, First initial. (Publication year). *Document title*. Accessed in: day, month, year, in: URL.
Ex: Chou, L., McClintock, R., Moretti, F. e Nix, D.H. (1993). *Technology and education: New wine in new bottles – Choosing pasts and imagining educational futures.* Acceded in 24th of August 2000, on the Web site of: Columbia University, Institute for Learning Technologies: http://www.ilt.columbia.edu/publications/papers/newwine1.html.

Dissertation:
Ex: Tingle, C.C.D. (1985). *Biological control of the glasshouse mealybug using parasitic hymenoptera*. Ph.D. Thesis. Department of Biological Sciences, Wye College, University of London. 375 pp.

Tables, Figures, Graphics and Boards:
All tables, figures, graphics and boards are to be numbered using Arabic numerals and should have a title explaining its components above the body, using size 9, bold, centred.
The source and year of the information given in tables, figures, graphics and boards should be included beneath its body, centred, size 8, regular. For tables and boards contents use size 8.
Figures and graphics must be in JPEG format (image).

CIEO

Centro de Investigação sobre o Espaço e as Organizações

Research Centre for Spatial and Organizational Dynamics

AUTORES

Ana Marreiros
Bruno Mora
Efigénio Rebelo
Eugénia Castela
Guilherme Castela
Iris Lopes
João Albino Silva
Maria Purificación Villardón

UAlg
UNIVERSIDADE DO ALGARVE

RESEARCH CENTRE FOR SPATIAL AND ORGANIZATIONAL DYNAMICS

APOIO

cria
Centro Regional para a Inovação do Algarve

FINANCIAMENTO

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR