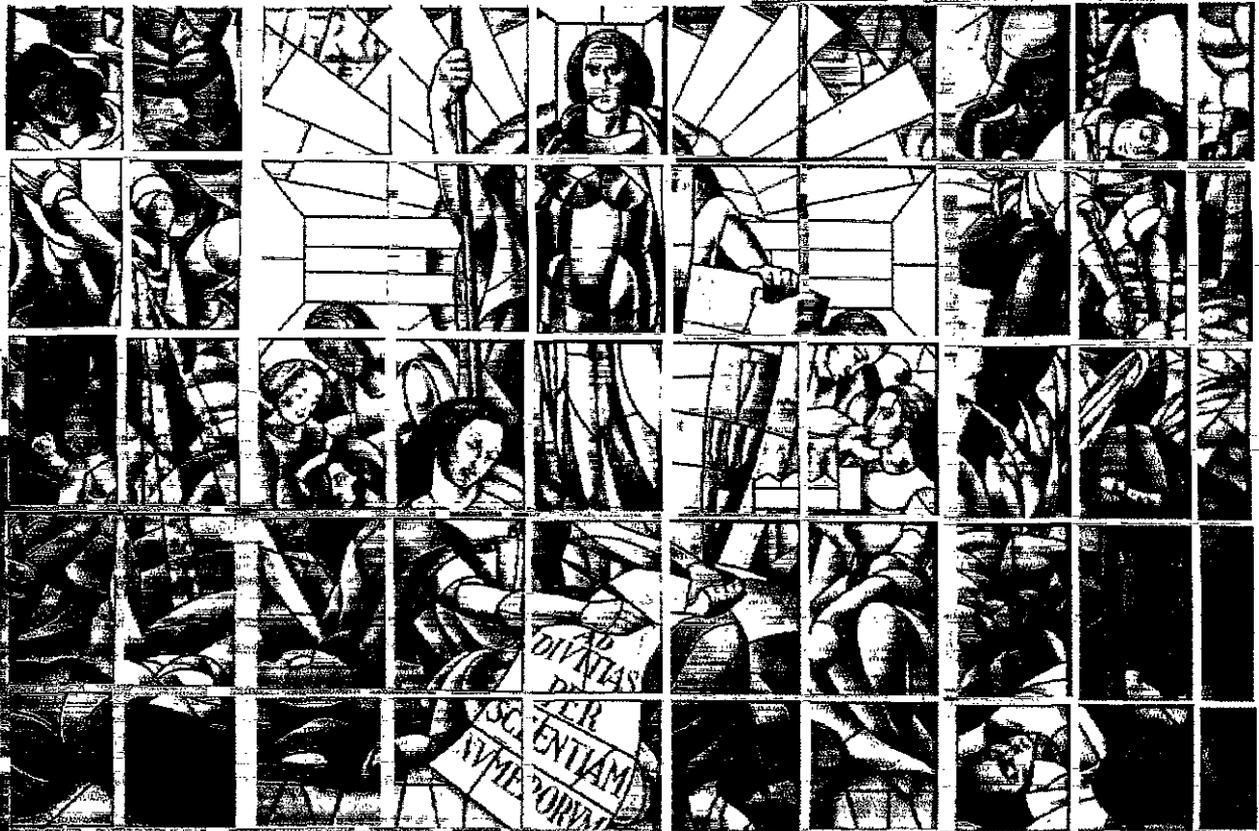




INSTITUTO NACIONAL DE ESTATÍSTICA

PORTUGAL

REVISTA DE ESTATÍSTICA



VOLUME I
1º QUADRIMESTRE 1998

CATALOGAÇÃO RECOMENDADA:

REVISTA DE ESTATÍSTICA. Lisboa, 1996-
Revista de estatística / ed. Instituto Nacional de
Estatística. - Vol. 1, 1º quadrimestre 1996- . -
Lisboa : I.N.E., 1996- . - 30 cm
Quadrimestral
ISSN 0873-4275

FICHA TÉCNICA

- DIRECTOR

- *Adrião Simões Ferreira da Cunha*

- DIRECTOR-ADJUNTO

- *Pedro Jorge Nunes da Silva Dias*

- CONSELHO EDITORIAL

- *Adrião Simões Ferreira da Cunha*
- *António Daniel Correia dos Santos*
- *Dinis Duarte Ferreira Pestana*
- *Francisco José Neto Melro*
- *João António Branco*
- *João Ferreira do Amaral*
- *Óscar Soares Barata*
- *Pedro Jorge Nunes da Silva Dias*
- *Pedro Miguel Girão Nogueira Ramos*
- *Sérgio Manuel Bacelar e Silva*

- SECRETARIADO DE REDACÇÃO

- *Eduarda Liliana Marques Martins*

- EDITOR

- *Instituto Nacional de Estatística*
Av. António José de Almeida, n.º 2
1 000 LISBOA
Telf. (01) 842 61 00
Fax (01) 842 63 64

- CAPA

- *Design de Mário Bouçados sobre o vitral do INE da autoria do pintor*
Abel Manta

- LAYOUT E MAQUETAGEM

- *Mário Bouçadas*

- IMPRESSÃO

- *Instituto Nacional de Estatística*
Secção Artes Gráficas

- TIRAGEM

- *750 exemplares*

- DEPÓSITO LEGAL

- *N.º 99514-96*

PREÇO (IVA 5% incluído)

- N.º avulso 2 310\$00
- Assinatura anual 5 540\$00

© INE, Lisboa, Portugal, 1998 * Reprodução autorizada, excepto para fins comerciais, com indicação da fonte bibliográfica.

CREDITS

- DIRECTOR

- *Adrião Simões Ferreira da Cunha*

- ASSISTANT DIRECTOR

- *Pedro Jorge Nunes da Silva Dias*

- EDITORIAL BOARD

- *Adrião Simões Ferreira da Cunha*
- *António Daniel Correia dos Santos*
- *Dinis Duarte Ferreira Pestana*
- *Francisco José Neto Melro*
- *João António Branco*
- *João Ferreira do Amaral*
- *Óscar Soares Barata*
- *Pedro Jorge Nunes da Silva Dias*
- *Pedro Miguel Girão Nogueira Ramos*
- *Sérgio Manuel Bacelar e Silva*

- EDITORIAL SECRETARIAT

- *Eduarda Liliana Marques Martins*

- PUBLISHER

- *Instituto Nacional de Estatística*
Av. António José de Almeida, n.º 2
1 000 LISBOA
Telf.: 351-(01) 842 61 00
Fax: 351-(01) 842 63 64

- COVER DESIGN

- *Designed by Mário Bouçados on the stain glass window at INE*
by the painter Abel Manta

- LAYOUT AND GRAPHIC DESIGN

- *Mário Bouçadas*

- PRINTING

- *Instituto Nacional de Estatística*
Secção Artes Gráficas

- EDITION

- *750 copies*

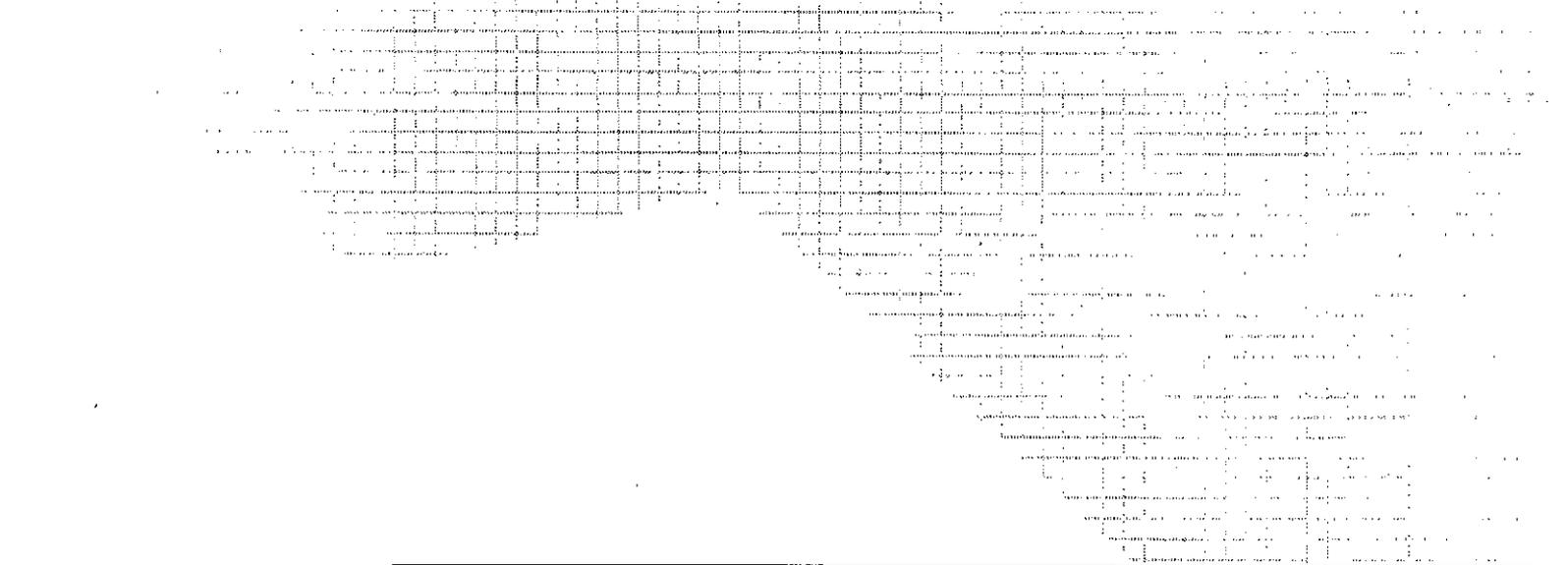
- LEGAL DEPOSIT REGISTRATION

- *N.º 99514-96*

PRICE (VAT 5% included)

Single issue PTE 2,310
Annual subscription PTE 5,540

© INE, Lisbon, Portugal, 1998 * Reproduction authorised except for commercial purposes by indicating the source.



REVISTA DE ESTATÍSTICA

VOLUME 1

1º QUADRIMESTRE DE 1998

ÍNDICE

INDEX

- ARTIGOS

ARTICLES

CLASSIFICATION OF VARIABLES IN DISCRIMINANT ANALYSIS: A NEW APPROACH

CLASSIFICAÇÃO DE VARIÁVEIS NA ANÁLISE DISCRIMINANTE: UMA NOVA ABORDAGEM

Por/By: Paulo Gomes e Adelaide Figueiredo 5

APLICAÇÃO DA EQUAÇÃO DE RICATTI AO ESTUDO DO PERÍODO DE OCUPAÇÃO DO SISTEMA $M | G | \infty$

APPLICATION OF RICATTI EQUATION TO THE BUSY PERIOD STUDY OF THE $M | G | \infty$ SYSTEM

Por/By: Manuel Ferreira 23

TRAJECTÓRIAS TECNOLÓGICAS E FUNÇÕES DE PRODUÇÃO – ASPECTOS METODOLÓGICOS E APLICAÇÃO EMPÍRICA A DOIS PERÍODOS RECENTES DA ECONOMIA PORTUGUESA

TECHNOLOGICAL COURSES AND PRODUCTION FUNCTIONS – METHODOLOGICAL ASPECTS AND EMPIRICAL APPLICATION A TWO RECENT PERIODS OF THE PORTUGUESES ECONOMY

Por/By: José Ramos Pires Manso, PhD 29

AGRUPAMENTOS DE CONCELHOS DE PORTUGAL CONTINENTAL E SUA CARACTERIZAÇÃO

GROUPS OF CONTINENTAL COUNCILS AND ITS PROFILES

Por/By: Ana Brandão, Ana Pires e Jorge Portugal 71

AS SONDAgens E A POPULAÇÃO . O QUE FAZER PARA A RECONCILIAÇÃO

OPINION POLLS AND THE POPULATION – HOW TO RECONCILE THEM

Por/By: Paula Vicente e Elizabeth Reis 95

O MODELO DINÂMICO DE TEIA DE ARANHA E A EXPECTATIVA

COBWED DYNAMIC MODEL AND EXPECTATIVE

Por/By: Valentim Lopes Velasco 111

TRÊS PROBLEMAS DE PARTICIPAÇÃO EM AMOSTRAGEM: ESTRATOS, DOMÍNIOS E GRUPOS	
<i>THREE PROBLEMS OF PARTITION IN SAMPLING: STRATA, DOMAINS AND GROUPS</i>	
Por/By: Osvaldo Caldeira.	141
AS ESTATÍSTICAS AGRÍCOLAS EUROPEIAS NO HORIZONTE 2005	
<i>THE EUROPEAN AGRICULTURAL STATISTICS IN 2005</i>	
Por/By: António de Sousa Macedo.	165
ACTUALIZAÇÃO DOS CADERNOS ELEITORAIS E SUAS CONSEQUÊNCIAS NOS CÍRCULOS DE APURAMENTO	
<i>THE IMPLICATIONS OF AN ACTUALIZATION OF THE PORTUGUESE ELECTORAL BOOKS</i>	
Por/By: Paulo Morais e José António Monteiro.	175
- <i>INFORMAÇÕES</i>	
<i>INFORMATIONS:</i>	
ACTIVIDADES E PROJECTOS IMPORTANTES NO ÂMBITO DO SISTEMA ESTATÍSTICO NACIONAL	
<i>IMPORTANT ACTIVITIES AND PROJECTS IN THE SCOPE OF THE NATIONAL STATISTICAL SYSTEM.</i>	185
CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS	
<i>CONGRESS, SEMINARS AND CONFERENCES.</i>	193
ACÇÕES DESENVOLVIDAS PELO INE NO ÂMBITO DA COOPERAÇÃO BILATERAL E MULTILATERAL	
<i>ACTIONS ACHIEVED BY NSI IN THE SCOPE OF BILATERAL AND MULTILATERAL COOPERATION.</i>	203
FUNDAMENTO, OBJECTO E ÂMBITO DA REVISTA.	
<i>FOUNDATION, SUBJECT MATTER AND SCOPE OF THE REVIEW.</i>	205
NORMAS DE APRESENTAÇÃO DE MANUSCRITOS PARA A REVISTA. . . .	
<i>RULES FOR SUBMITTING MANUSCRIPTS TO THE REVIEW.</i>	207

CLASSIFICATION OF VARIABLES IN DISCRIMINANT
ANALYSIS: A NEW APPROACH

Autores:
Paulo Gomes
e
Adelaide Figueiredo

CLASSIFICATION OF VARIABLES IN DISCRIMINANT ANALYSIS: A
NEW APPROACH

CLASSIFICAÇÃO DE VARIÁVEIS NA ANÁLISE DISCRIMINANTE:
UMA NOVA ABORDAGEM

Autores: Paulo Gomes

- Professor Associado, Universidade Católica Portuguesa
- Director Regional do Instituto Nacional de Estatística

e

Adelaide Figueiredo

- Assistente da Faculdade de Economia da Universidade do Porto

ABSTRACT:

- In this report we develop an approach analogous to the Discriminant Analysis to classify new variables into previously defined groups of variables and give the misclassification probabilities estimates.

KEY WORDS:

- *Discriminant Analysis; Classification; Variables Selection*

RESUMO:

- Neste trabalho procurou desenvolver-se uma abordagem análoga à Análise Discriminante para classificar novas variáveis em grupos de variáveis previamente definidos e obter estimativas das probabilidades de erro de má classificação.

PALAVRAS-CHAVE:

- *Análise Discriminante; Classificação; Selecção de Variáveis*

1. INTRODUCTION

One of the most important questions of Discriminant Analysis tries to solve is the affectation of new individuals into previously defined groups of individuals. We consider the dual problem of the affectation of new variables into previously defined groups of variables. While in Discriminant Analysis, we usually associate the Multivariate Normal distribution to each group of individuals for defining the affectation rules, in our approach we associate the Bingham distribution on the sphere to each group of variables for defining the affectation rules. In our study we suppose that the groups of variables were been previously identified. The identification of those groups can be done using the *k-means method* (Gomes, 1987) or the *E.M. algorithm* of Estimation-Maximization type (Gomes e Figueiredo, 1995). Additionally it's important to give the estimate of the misclassification probability.

In our approach, we suppose that the n individuals are fixed and the p variables, previously normalised, are randomly selected from a population of variables. Then, we associate to each variable a $1 \times n$ random vector $(X_1^j, X_2^j, \dots, X_n^j)$ where X_i^j is a random variable that represents the value of the j^{th} variable for the i^{th} individual. In the classical approach the p variables are fixed and we select randomly the individuals and associate to each individual a $1 \times p$ random vector $(X_i^1, X_i^2, \dots, X_i^p)$ where X_i^j is a random variable that represents the value of the j^{th} variable for the i^{th} individual.

In section 2, we adapt to our approach the classification rules used in Discriminant Analysis, and section 3 applies the rules to the case of a Bingham distribution on the sphere. Some methods to the calculation of the estimates of the misclassification probabilities are described in section 4 and an example is given in section 5. Finally, section 6 contains the conclusion.

2. CLASSIFICATION RULES

Consider our data with n individuals and p variables and suppose that the p variables are divided in two groups. We associate to the groups of variables, two subpopulations G_1 and G_2 . The variables taken from subpopulation G_1 have the density function on the sphere $f_1(\mathbf{x})$ and the variables from subpopulation G_2 have the density function on the sphere $f_2(\mathbf{x})$.

Let S_{n-1} be the surface of the sphere:

$$S_{n-1} = \{\mathbf{x} \in \mathcal{R}^n : \|\mathbf{x}\| = 1\}$$

We divide S_{n-1} into mutually exclusive and exhaustive regions R_1 and R_2 . Variables falling in region R_1 are classified into subpopulation 1 and those falling in region R_2 are classified into subpopulation 2. The *likelihood-ratio rule* is the following:

Assign a variable \mathbf{x} to G_1 if

$$f_1(\mathbf{x}) \geq f_2(\mathbf{x}) \quad (1)$$

and assign \mathbf{x} to G_2 , otherwise.

The classification regions R_1 and R_2 are defined by

$$R_1 = \left\{ \mathbf{x} \in S_{n-1} : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \right\}$$

$$R_2 = \left\{ \mathbf{x} \in S_{n-1} : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < 1 \right\}$$

We now mention a rule based on the Statistical Decision Theory where the misclassification costs and the prior probabilities are considered.

Let π_i be the probability of a given variable belonging to group i , denoted by prior probability, $i = 1, 2$ and $C(i|j)$ the cost of misclassifying a variable from group j into group i , $i, j = 1, 2$. The probability of misclassification is given by

$$P(i|j) = P(\mathbf{X} \in R_i | \mathbf{X} \in G_j) = \int_{R_i} f_j(\mathbf{x}) d\mathbf{x}$$

Then, $\pi_i \cdot P(j|i)$ is the probability of choosing a variable from G_i and wrongly classifying it into G_j , $i, j = 1, 2$.

Denoting by C the total cost of misclassification, we have $E(C) = C(2|1) \cdot \pi_1 \cdot P(2|1) + C(1|2) \cdot \pi_2 \cdot P(1|2)$.

The rule used to minimize $E(C)$, known as *Bayes rule*, is the following:

Assign a variable \mathbf{x} to G_1 if

$$C(2|1) \cdot \pi_1 \cdot f_1(\mathbf{x}) \geq C(1|2) \cdot \pi_2 \cdot f_2(\mathbf{x})$$

and assign \mathbf{x} to G_2 , otherwise.

The classification regions R_1 and R_2 are defined by

$$R_1 = \left\{ \mathbf{x} \in S_{n-1} : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} \right\}$$

$$R_2 = \left\{ \mathbf{x} \in S_{n-1} : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} \right\}$$

If we have equal misclassification costs and equal prior probabilities, the Bayes rule and the likelihood-ratio rule will be the same.

Suppose now, we have k subpopulations G_1, G_2, \dots, G_k with density functions on the sphere $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x})$, respectively. We divide S_{n-1} into R_1, R_2, \dots, R_k mutually exclusive and exhaustive regions. Any given variable is classified into the subpopulation in which region the variable falls. The likelihood-ratio rule is the following:

Assign a variable \mathbf{x} to G_j if

$$f_j(\mathbf{x}) = \max_{i=1, \dots, k} f_i(\mathbf{x})$$

We now consider the Bayes rule.

Let π_i be the prior probability of a given variable belonging to G_i and let $C(j|i)$ be the cost of misclassifying a variable from G_i into G_j . The misclassification probability is given by $P(j|i) = P(X \in R_j | X \in G_i) = \int_{R_j} f_i(\mathbf{x}) d\mathbf{x}$

The total cost of misclassifying a variable belonging to G_i is given by $\sum_{j \neq i}^k P(j|i) \cdot C(j|i)$, and the expected total cost of misclassifying variables belonging to G_i will be $\pi_i \left[\sum_{j \neq i}^k P(j|i) \cdot C(j|i) \right]$. The resulting total expected cost of misclassification, for all subpopulations is defined by

$$E(C) = \sum_{i=1}^k \pi_i \left(\sum_{j \neq i}^k P(j|i) \cdot C(j|i) \right)$$

The rule for minimizing $E(C)$, Bayes rule, is the following:

Assign a variable \mathbf{x} to G_j if

$$\sum_{\substack{i=1 \\ i \neq j}}^k \pi_i \cdot f_i(\mathbf{x}) \cdot C(j|i) < \sum_{\substack{i=1 \\ i \neq l}}^k \pi_i \cdot f_i(\mathbf{x}) \cdot C(l|i) \quad l = 1, \dots, k, \quad l \neq j \quad (2)$$

2.1 POSTERIOR PROBABILITIES

In addition to the affectation group of each variable, we will give more information about each variable if we determine its posterior probability.

The posterior probability of a variable \mathbf{X} belonging to population i is

$$P(X \in G_i | X = \mathbf{x}) = \frac{f_i(\mathbf{x}) \cdot \pi_i}{\sum_{i=1}^k f_i(\mathbf{x}) \cdot \pi_i}, \quad i = 1, \dots, k \quad (3)$$

where $\pi_i = P(X \in G_i)$ is the prior probability of population i .

Based on the posterior probabilities, we can affect a variable \mathbf{x} to the population G_i , for which the posterior probability (3) or $\pi_i \cdot f_i(\mathbf{x})$ is the largest. It is equivalent to using the Bayes rule with equal misclassification costs.

3. APPLICATION OF THE CLASSIFICATION RULES

We use a particular case of the Bingham distribution on the sphere whose density function is given by

$$f(\mathbf{x}) = \left\{ {}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi\right) \right\}^{-1} \cdot \exp(\xi' \mathbf{u} \cdot \mathbf{x} \cdot \mathbf{x} \cdot \mathbf{u}) \quad \mathbf{x} \in S_{n-1}$$

where ${}_1F_1$ is a confluent hypergeometrical function defined by

$${}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi\right) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{n-1}{2}\right)} \int_0^1 e^{\xi t} t^{\frac{1}{2}} (1-t)^{\frac{n-3}{2}} dt$$

This probability distribution denoted by $B_n(\mathbf{u}, \xi)$, has two parameters:

- \mathbf{u} is a directional parameter ($'\mathbf{u}\mathbf{u} = 1$)
- $\xi (> 0)$ - concentration parameter around the directional parameter

If we associate axes to the variables the Bingham distribution will be appropriate for modelling the group of variables because its density function satisfies the antipodal symmetric property $f(\mathbf{x}) = f(-\mathbf{x})$.

The maximum likelihood estimators of the parameters \mathbf{u} and ξ from the Bingham distribution $B_n(\mathbf{u}, \xi)$, based on a random sample of p variables $X = (\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_p)$, were obtained by (Gomes, 1987) and are following:

- $\hat{\mathbf{u}}$ is the eigenvector associated to the largest eigenvalue of the matrix $X'X$, that is

$$X'X \cdot \hat{\mathbf{u}} = \omega \cdot \hat{\mathbf{u}} \quad (4)$$

where ω is the largest eigenvalue of $X'X$. Then, the maximum likelihood estimator for \mathbf{u} and the first principal component associated to the group of variables are coincident.

- $\hat{\xi}$ is the solution of the equation

$$Y(\xi) = \frac{\omega}{p}$$

where

$$Y(\xi) = \frac{d}{d\xi} \ln \left[{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi\right) \right]$$

We refer only the Bayes rule because the rule based on the likelihood ratio is a particular case.

Suppose that the population 1 is Bingham $B_n(\mathbf{u}_1, \xi_1)$ and that the population 2 is Bingham $B_n(\mathbf{u}_2, \xi_2)$.

Let

$$U = \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \frac{\left[{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi_1\right) \right]^{-1} \cdot e^{\xi_1(\mathbf{u}_1 \cdot \mathbf{x})^2}}{\left[{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi_2\right) \right]^{-1} \cdot e^{\xi_2(\mathbf{u}_2 \cdot \mathbf{x})^2}}$$

When the parameters $\mathbf{u}_1, \mathbf{u}_2, \xi_1$ and ξ_2 are unknown, they may be replaced by their maximum likelihood estimates.

The Bayes rule is:

Assign \mathbf{x} to population 1 if

$$\ln U - \ln \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} \geq 0$$

or

$$\mathbf{x} \hat{A} \mathbf{x} + \ln \frac{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_2\right)}{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_1\right)} - \ln \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} \geq 0 \quad (5)$$

where

$$\hat{A} = \left(\sqrt{\hat{\xi}_1} \cdot \hat{\mathbf{u}}_1 + \sqrt{\hat{\xi}_2} \cdot \hat{\mathbf{u}}_2 \right)' \cdot \left(\sqrt{\hat{\xi}_1} \cdot \hat{\mathbf{u}}_1 - \sqrt{\hat{\xi}_2} \cdot \hat{\mathbf{u}}_2 \right)$$

and assign \mathbf{x} to population 2, otherwise.

Let the discriminant function be

$$W_{12} = {}^t \mathbf{x} \hat{A} \mathbf{x} + \ln \frac{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_2\right)}{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_1\right)} - \ln \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} \quad (6)$$

the discriminant regions R_1 and R_2 are given by

$$R_1 = \{x: W_{12} \geq 0\}$$

$$R_2 = \{x: W_{12} < 0\}$$

3.2 MORE THAN TWO GROUPS

Suppose the k subpopulations are Bingham $B_n(\mathbf{u}_i, \xi_i)$, $i = 1, \dots, k$, whose density function is given by

$$f_i(\mathbf{x}) = \left\{ {}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \xi_i\right) \right\}^{-1} \cdot e^{\xi_i({}^t \mathbf{u}_i \mathbf{x})^2} \quad \mathbf{x} \in S_{n-1}$$

The Bayes rule is

Classify \mathbf{x} as a variable from population j if

$$\sum_{\substack{i=1 \\ i \neq j}}^k \pi_i \cdot f_i(\mathbf{x}) \cdot C(j|i) = \min_{l=1, \dots, k} \sum_{i=1}^k \pi_i \cdot f_i(\mathbf{x}) \cdot C(l|i) \quad (7)$$

or

$${}^t \mathbf{x} \hat{A} \mathbf{x} + \ln \frac{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_i\right)}{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_j\right)} - \ln \frac{C(j|i) \cdot \pi_i}{C(i|j) \cdot \pi_j} \geq 0 \quad i = 1, \dots, k, \quad i \neq j \quad (8)$$

where

$$\hat{A} = \left(\sqrt{\hat{\xi}_j} \cdot \hat{\mathbf{u}}_j + \sqrt{\hat{\xi}_i} \cdot \hat{\mathbf{u}}_i \right) \cdot \left(\sqrt{\hat{\xi}_j} \cdot \hat{\mathbf{u}}_j - \sqrt{\hat{\xi}_i} \cdot \hat{\mathbf{u}}_i \right)$$

and $\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\xi}_i, \hat{\xi}_j$ are the maximum likelihood estimators of the parameters.

Let

$$W_{ji} = \mathbf{x}' \hat{\mathbf{A}} \mathbf{x} + \ln \frac{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_i\right)}{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_j\right)} - \ln \frac{C(j|i) \cdot \pi_i}{C(i|j) \cdot \pi_j} \quad (9)$$

be the discriminant function.

The region R_j is given by

$$R_j = \left\{ \mathbf{x} : W_{ji} \geq 0 \quad i = 1, \dots, k, \quad i \neq j \right\}$$

In case of three groups, the discriminant functions are given by

$$W_{12} = \hat{\xi}_1' \mathbf{x} \cdot \hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 \cdot \mathbf{x} - \hat{\xi}_2' \mathbf{x} \cdot \hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 \cdot \mathbf{x} + \ln \frac{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_2\right)}{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_1\right)} - \ln \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1}$$

$$W_{13} = \hat{\xi}_1' \mathbf{x} \cdot \hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 \cdot \mathbf{x} - \hat{\xi}_3' \mathbf{x} \cdot \hat{\mathbf{u}}_3' \hat{\mathbf{u}}_3 \cdot \mathbf{x} + \ln \frac{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_3\right)}{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_1\right)} - \ln \frac{C(1|3) \cdot \pi_3}{C(3|1) \cdot \pi_1}$$

$$W_{23} = \hat{\xi}_2' \mathbf{x} \cdot \hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 \cdot \mathbf{x} - \hat{\xi}_3' \mathbf{x} \cdot \hat{\mathbf{u}}_3' \hat{\mathbf{u}}_3 \cdot \mathbf{x} + \ln \frac{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_3\right)}{{}_1F_1\left(\frac{1}{2}, \frac{n}{2}, \hat{\xi}_2\right)} - \ln \frac{C(2|3) \cdot \pi_3}{C(3|2) \cdot \pi_2}$$

As the condition

$$W_{23} = W_{13} - W_{12} + \ln \frac{C(1|3) \cdot \pi_3}{C(3|1) \cdot \pi_1} - \ln \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} - \ln \frac{C(2|3) \cdot \pi_3}{C(3|2) \cdot \pi_2}$$

holds, it is enough to use only the two functions W_{13} and W_{12} for defining the discriminant regions R_1 , R_2 and R_3 .

$$R_1 = \left\{ \mathbf{x} : W_{12} \geq 0, \quad W_{13} \geq 0 \right\}$$

$$R_2 = \left\{ \mathbf{x} : W_{12} \leq 0, \quad W_{13} \geq W_{12} - \ln \frac{C(1|3) \cdot \pi_3}{C(3|1) \cdot \pi_1} + \ln \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} + \ln \frac{C(2|3) \cdot \pi_3}{C(3|2) \cdot \pi_2} \right\}$$

$$R_3 = \left\{ \mathbf{x} : W_{13} \leq 0, \quad W_{13} \leq W_{12} - \ln \frac{C(1|3) \cdot \pi_3}{C(3|1) \cdot \pi_1} + \ln \frac{C(1|2) \cdot \pi_2}{C(2|1) \cdot \pi_1} + \ln \frac{C(2|3) \cdot \pi_3}{C(3|2) \cdot \pi_2} \right\}$$

4. MISCLASSIFICATION PROBABILITIES

In the case of two groups, the misclassification probabilities are

$$P(2|1) = P(\text{misclassifying a variable from } G_1 \text{ into } G_2) = P(W_{12} < 0 | \mathbf{x} \in G_1)$$

$$P(1|2) = P(\text{misclassifying a variable from } G_2 \text{ into } G_1) = P(W_{12} \geq 0 | \mathbf{x} \in G_2)$$

As the distribution for the random variable W_{12} if $x \in G_1$ or if $x \in G_2$ is unknown, we can not calculate these probabilities, but we can estimate them.

4.1 ESTIMATES OBTAINED BY SIMULATION

If we consider two subpopulations of Bingham, $B_n(\hat{\mathbf{u}}_1, \hat{\xi}_1)$ and $B_n(\hat{\mathbf{u}}_2, \hat{\xi}_2)$, we can simulate an adequate number of variables from each subpopulations. For each simulated variable, we calculated the value for the random variable W_{12} . An estimate of $P(2|1)$ is the proportion of the variables from population 1 for which $W_{12} < 0$ and an estimate of $P(1|2)$ is the proportion of variables from population 2 for which $W_{12} \geq 0$.

4.2 ESTIMATES OF DISCRIMINANT ANALYSIS

In Discriminant Analysis there are three probabilities of correct classification (i.e., three population hit rates):

- *optimal hit rate* obtained when a classification rule based on known parameters is applied to the population.
- *actual hit rate* obtained by applying a rule based on a particular sample to future samples.
- *expected actual hit rate* is the expected proportion of correct classifications over all possible samples.

The actual hit rate is the most common and can be estimated by using an internal analysis or an external analysis (i.e., the Holdout method or the Leave-One-Out Method) or the Bootstrap technique. These methods will be described next.

4.2.1 INTERNAL ANALYSIS

We use samples of each subpopulation to estimate the unknown parameters of the subpopulations, and to determine a rule, then, based on the obtained rule, we classify the units of the same samples. The obtained rates are called *apparent or resubstitution* hit rates.

As the apparent hit rate yields a positively biased estimate, we use an external analysis to validate the classification rule.

4.2.2 EXTERNAL ANALYSIS

Whereas in the internal analysis the units classified are the units used for defining the rule in the external analysis, the rule is determined from one set of units and then used to classify another set of units.

Holdout method

We divide the sample (of each subpopulation) in two subsamples: a training sample and a test sample. We estimate the unknown parameters of the subpopulations based on the training sample and, therefore, obtain the classification rule. Then, with the obtained rule we classify the units of the test sample. The proportion of the test sample units that are correctly classified is a hit rate estimate. This method requires the use of large samples.

Leave-One-Out method (Lachenbruch, P.A. and Mickey, 1968)

We delete one unit from one of the subpopulations samples and, based on the remaining units of that sample, we estimate the unknown parameters of that subpopulation. We also estimate the unknown parameters of the other subpopulations based on the respective samples. Therefore, we obtain a rule that we use for classifying the deleted unit into one of the groups. Afterwards, we delete another unit of the sample and we determine the rule used for classifying the deleted unit. We repeat the process until all units of each sample have been deleted. The proportions of deleted units correctly classified are the hit-rate estimates.

4.2.3 BOOTSTRAP TECHNIQUE (EFRON, 1987)

After having estimated the unknown parameters based on the training samples and having determined the rule, we generate samples by choosing units at random and with replacement from the sample composed by the test samples. For each generated sample, called *bootstrap sample*, we classify its units and calculate the hit rate because we know to which group belongs each variable.

As the hit-rate empirical bootstrap distribution is an approximation to the hit-rate distribution, an estimate of the true hit rate is the average of the hit rates obtained with the bootstrap samples.

In our approach, we use the previous estimates of the misclassification probabilities.

5. EXAMPLE

We considered the data based on the statistic of crime for England and Wales during the period 1950-1963. See appendix.

Using the method of Principal Components, (Ahmad, 1967), has shown that the yearly variation in the number of crimes could be explained by a small number of unrelated factors. They are, mainly, the 1st component, associated with the population growth, and the 2nd component that seems likely to reflect changes in recording practice by the police over the period.

Supposing that the group of variables is composed by two subgroups of variables, each one from a Bingham population and identifying a mixture of two Bingham distributions on the sphere by the *k-means method* or the *E. M. algorithm*, we obtained the following two groups of variables:

Group I - 1,3,6,10,13,17

Group II - 2,4,5,7,8,9,11,12,14,15,16,18

being the variables of group I more associated with changes in the recording of crimes and the variables of group II more associated with the population growth.

After having identified the two groups of variables, we used the likelihood-ratio rule for classifying variables into one of the groups.

Firstly, we determined the estimates of the misclassification probabilities, by simulation. We simulated 1000 variables from each of the Bingham populations $B_{14}(\hat{\mathbf{u}}_1, \hat{\xi}_1)$ and $B_{14}(\hat{\mathbf{u}}_2, \hat{\xi}_2)$ where $\hat{\mathbf{u}}_1$ and $\hat{\xi}_1$ are the maximum likelihood estimators based on the sample of dimension 6 from population 1 and $\hat{\mathbf{u}}_2$ and $\hat{\xi}_2$ the maximum likelihood estimators based on the sample of dimension 12 from population 2. We obtained the estimates:

$$\hat{P}(2|1) = 3.9\%$$

$$\hat{P}(1|2) = 5.2\%$$

Secondly we classified the variables of each sample. The results are shown in the following table:

		Origin group	
		1	2
Affection group	1	5	0
	2	1	12
		6	12

i.e., the apparent estimates of misclassification probabilities where:

$$\hat{P}(2|1) = \frac{1}{6}$$

$$\hat{P}(1|2) = 0$$

and the global hit rate was 94,4%

These estimates gives us just an idea of how the groups of variables were separated. Therefore, we used the *Holdout method* for validating the rule.

In the *Holdout method*, we considered about 30% of the sample of each group for the test sample and about 70% of the sample for the training sample. So, for the test sample, we chose randomly 2 variables from group I, 4 variables from group II and the other 12 variables constitute the training sample which was used for defining

the rule. We considered three test samples to avoid biased sampling. The samples considered are described in the following table:

	Group	Test samples	Training samples
1 st Case	I	13,17	1,3,6,10
	II	11,12,14,15	2,4,5,7,8,9,16,18
2 nd Case	I	6,13	1,3,10,17
	II	4,8,9,11	2,5,7,12,14,15,16,18
3 rd Case	I	1,10	3,6,13,17
	II	2,7,14,15	4,5,8,9,11,12,16,18

The misclassification probabilities estimates in each case, are:

	$\hat{P}(2 1)$	$\hat{P}(1 2)$
1 st Case	0	0
2 nd Case	0.5	0
3 rd Case	0.5	0

We can take the average of these estimates of probabilities as estimates of the misclassification probabilities:

$$\hat{P}(2|1) = 33.3\%$$

$$\hat{P}(1|2) = 0$$

Finally we used also the method Leave-one-out, where we obtained the following misclassification probabilities estimates:

$$\hat{P}(2|1) = 16.7\%$$

$$\hat{P}(1|2) = 0$$

6. CONCLUSION

Our approach is concerned with variables selection under a multivariate statistical analysis, in a data science classification context, where we established, according to a probabilistic model defined in the R^n sphere, a decision rule which enables us to affect a new variable into two or more groups of variables previously defined through the k-means method or E.M. algorithm.

The misclassification error problem was studied by a simulation method, afterwards by using apparent estimates and, finally, based on Holdout and on Leave-one-out methods. The second approach tends to underestimate the misclassification probabilities. Though, using the Leave-one-out method seems to confirm the results

obtained by that second approach. The small number of variables included in the samples suggests the attribution of particular relevance to the estimates of misclassification probabilities provided by simulation.

BIBLIOGRAPHY

- AHMAD, (1967) "An analysis of crimes by the method of Principal Components", *Applied Statistics*, pp.17-39.
- EFRON, (1987) "Better Bootstrap Confidence Intervals", *Journal of the American Statistical Society*, 82 (March), 171-185.
- GOMES, (1987) "Distribution de Bingham sur la n-sphère: Une Nouvelle Approche de l'Analyse Factorielle", *Thèse d'État*, Université de Montpellier.
- Gomes e Figueiredo, (1995) "Identificação de uma mistura de leis de Bingham através dum algoritmo tipo E.M.", *Actas do III Congresso da Sociedade Portuguesa de Estatística*.
- HUBERTY, C., (1994) "Applied Discriminant Analysis" *Wiley Series in Probability and Mathematical Statistics*.
- LACHENBRUCH, P.A. and MICKEY, (1986) "Estimation of error rates in discriminant analysis" *Technometrics*, vol.10, pp.1-11.
- MARDIA, K. (1972) "Statistics of directional data" *Academic Press*.
- WATSON, (1983) "Statistics on spheres". *J. Wiley*.

A APPENDIX

A.1 VARIABLES

- 1- Homicide: Murder, attempted murder, manslaughter, infanticide
- 2- Woundings: Felonious, Malicious, Assault
- 3- Homosexual offences: Buggery and attempts at Indecency between males
- 4- Heterosexual offences: Rape, indecent assault, unlawful intercourse, incest
- 5- Breaking and entering: Sacrilege, burglary, housebreaking, shopbreaking, etc.
- 6- Robbery
- 7- Larceny: Embezzlement, aggravated larceny, etc
- 8- Frauds and false pretences
- 9- Receiving stolen goods
- 10- Malicious injuries to property
- 11- Forgery, etc
- 12- Blackmail
- 13- Assault
- 14- Malicious damage
- 15- Revenue laws
- 16- Intoxicating laws
- 17- Indecent exposure
- 18- Taking motor vehicle without consent

A.2 DATA

ind/var	1	2	3	4	5	6	7	8	9
1	529	5.258	4.416	8.178	92.839	1.021	301.078	25.333	7.586
2	455	5.619	4.876	9.223	95.946	800	355.407	27.216	9.716
3	555	5.98	5.443	9.026	97.941	1.002	341.512	27.051	9.188
4	456	6.187	5.68	10.107	88.607	980	308.578	27.763	7.786
5	487	6.586	6.357	9.279	75.888	812	285.199	26.267	6.468
6	448	7.076	6.644	9.953	74.907	823	295.035	22.966	7.016
7	477	8.433	6.196	10.505	85.768	965	323.561	23.029	7.215
8	491	9.774	6.327	11.9	105.042	1.194	360.985	26.235	8.619
9	453	10.945	5.471	11.823	131.132	1.692	409.388	29.415	10.002
10	434	12.707	5.732	13.864	133.962	1.9	445.888	34.061	10.254
11	492	14.391	5.24	14.304	151.378	2.014	489.258	36.049	11.696
12	459	16.197	5.605	14.376	164.806	2.349	531.430	39.651	13.777
13	504	16.43	4.866	14.788	192.302	2.517	588.566	44.138	15.783
14	510	18.655	5.435	14.722	219.138	2.483	635.627	45.923	17.777

ind/var	10	11	12	13	14	15	16	17	18
1	4.158	3.79	118	20.844	9.447	24.616	49.007	2.786	3.126
2	4.993	3.378	74	19.963	10.359	21.122	55.229	2.739	4.595
3	5.003	4.173	120	19.056	9.108	23.339	55.635	2.598	4.145
4	5.309	4.649	108	17.772	9.278	19.919	55.688	2.639	4.551
5	5.251	4.903	104	17.379	9.176	20.585	57.011	2.587	4.343
6	2.184	4.086	92	17.329	9.46	19.197	57.118	2.607	4.836
7	2.559	4.04	119	16.677	10.997	19.064	63.289	2.311	5.932
8	2.965	4.689	121	17.539	12.817	19.432	71.014	2.31	7.148
9	3.607	5.376	164	17.344	14.289	24.543	69.864	2.371	9.772
10	4.083	5.5987	160	18.047	14.118	26.853	69.751	2.544	11.211
11	4.802	6.59	241	18.801	15.866	31.266	74.336	2.719	12.519
12	5.606	6.924	205	18.525	16.399	29.922	81.753	2.82	13.050
13	6.256	7.816	250	16.449	16.852	34.915	89.709	2.614	14.141
14	6.935	8.634	257	15.918	17.003	40.434	89.149	2.777	22.896

APLICAÇÃO DA EQUAÇÃO DE RICATTI AO ESTUDO
DO PERÍODO DE OCUPAÇÃO DO SISTEMA $M|G|_{\infty}$

Autor:
Manuel A. M. Ferreira

APLICAÇÃO DA EQUAÇÃO DE RICATTI AO ESTUDO DO
PERÍODO DE OCUPAÇÃO DO SISTEMA $M|G|\infty$

APPLICATION OF RICATTI EQUATION TO THE BUSY PERIOD
STUDY OF THE $M|G|\infty$ SYSTEM

Autor: Manuel Alberto Martins Ferreira
Professor Auxiliar do Instituto Superior de Ciências do Trabalho e da Empresa

RESUMO:

- O sistema de fila de espera $M|G|\infty$ pode ser aplicado na modelação de um vasto rol de problemas de âmbito social: doença, desemprego, emigração,... (ver, por exemplo, (Ferreira, 1995 e 1996)). Nestas aplicações assume particular importância o estudo da distribuição da duração do período de ocupação daquele sistema. Mostramos, neste trabalho, que a solução daquele problema pode passar pela resolução de uma equação de Ricatti.

PALAVRAS-CHAVE:

- $M|G|\infty$, Período de ocupação, Equação de Ricatti.

ABSTRACT:

- The $M|G|\infty$ queueing system can be applied in the modelation of many social problems: sickness, unemployment, emigration,... (see, for instance, (Ferreira, 1995 e 1996)). In these situations is very important to study the busy period length distribution of that system. We show, in this work, that the solution of the problem may be in the resolution of a Ricatti's equation.

KEY-WORDS:

- $M|G|\infty$, Busy Period, Ricatti's equation.

VOLUME 1

1º QUADRIMESTRE DE 1998

O estudo do comportamento, como função do tempo, da probabilidade de o sistema $M|G|\infty$ estar vazio, considerando como instante inicial o do começo de um período de ocupação (em que um cliente chega ao sistema encontrando-o vazio) reduz-se ao estudo do sinal de $\frac{g(t)}{1-G(t)} - \lambda G(t)$, $t \geq 0$ (Ferreira, 1996) em que λ é a taxa do processo de Poisson de chegadas, $g(\cdot)$ e $G(\cdot)$ são, respectivamente, a função de densidade de probabilidade e de distribuição do serviço e t é o tempo.

Fazendo $\frac{g(t)}{1-G(t)} - \lambda G(t) = \beta(t)$ ($\beta(\cdot)$ função do tempo qualquer) obtém-se

$$\boxed{\frac{dG(t)}{dt} = -\lambda G^2(t) - (\beta(t) - \lambda)G(t) + \beta(t)} \quad (1)$$

que é uma equação de Ricatti em $G(\cdot)$.

Resolvendo-a, para o que é necessário ter em conta que $G(t)=1$, $t \geq 0$ é uma solução, obtém-se

$$\boxed{G(t) = 1 + \frac{(1-G(0))e^{-\lambda t - \int_0^t \beta(u) du}}{\lambda(1-G(0)) \int_0^t e^{-\lambda w - \int_0^w \beta(u) du} dw - 1}} \quad (2)$$

Entrando com esta colecção de distribuições de serviço em

$$\boxed{\bar{B}(s) = 1 + \lambda^{-1} \left(s - \frac{1}{\int_0^\infty e^{-st - \lambda \int_0^t [1-G(v)] dv} dt} \right)} \quad (3)$$

que é a Transformada de Laplace da duração do período de ocupação do sistema $M|G|\infty$ (Takács, 1962), obtém-se

$$\bar{B}(s) = 1 + \lambda^{-1} \left(s - \frac{1}{\int_0^{\infty} e^{-st} [\lambda(1-G(0)) \int_0^t e^{-\lambda w - \int_0^w \beta(u) du} dw - 1] dt} \right) \quad (4).$$

A sua inversão, com vista à determinação de $b(t)$, função de densidade de probabilidade de duração do período de ocupação é um problema complicado.

Mas, se $\beta(t) = \beta$ (constante) obtém-se

$$G(t) = 1 - \frac{(1-e^{-\rho})(\lambda + \beta)}{\lambda e^{-\rho}(e^{(\lambda+\beta)t} - 1) + \lambda}, \quad t \geq 0, \quad -\lambda \leq \beta \leq \frac{\lambda}{e^{\rho} - 1} \quad (5)$$

e

$$b(t) = \frac{e^{-\rho}(\lambda + \beta) - \beta}{\lambda} \delta(t) + \left(1 - \frac{e^{-\rho}(\lambda + \beta) - \beta}{\lambda} \right) e^{-\rho}(\lambda + \beta) e^{-\rho(\lambda + \beta)t},$$

$$t \geq 0, \quad -\lambda \leq \beta \leq \frac{\lambda}{e^{\rho} - 1} \quad (6).$$

Portanto, se a distribuição de serviço for dada por (5), a distribuição do período de ocupação é dada pela mistura de uma distribuição degenerada na origem com uma exponencial.

REFERÊNCIAS BIBLIOGRÁFICAS

- FERREIRA, M. A. M., "Aplicação do Sistema $M|G|\infty$ ao Estudo do Desemprego numa Certa Actividade", Revista Portuguesa de Gestão, I.S.C.T.E., IV/95.
- FERREIRA, M.A.M., "Comportamento Transiente do Sistema $M|G|\infty$ - Aplicação em Problemas de Doença e de Desemprego", Revista de Estatística, I.N.E., Volume 3 3º Quadrimestre de 1996.
- TAKÁCS, L., "An Introduction to Queueing Theory". Oxford University Press, New York, 1962

TRAJECTÓRIAS TECNOLÓGICAS E FUNÇÕES DE
PRODUÇÃO - *ASPECTOS METODOLÓGICOS E APLICAÇÃO EMPÍRICA A DOIS
PERÍODOS RECENTES DA ECONOMIA PORTUGUESA*

Autor:
José Ramos Pires Manso, PhD

TRAJECTÓRIAS TECNOLÓGICAS E FUNÇÕES DE PRODUÇÃO -
ASPECTOS METODOLÓGICOS E APLICAÇÃO EMPÍRICA A DOIS PERÍODOS
RECENTES DA ECONOMIA PORTUGUESA

TECHNOLOGICAL COURSES AND PRODUCTION FUNCTIONS -
METHODOLOGICAL ASPECTS AND EMPIRICAL APPLICATION A TWO RECENT
PERIODS OF THE PORTUGUESE ECONOMY

Autor: José Ramos Pires Manso, PhD
- Professor Auxiliar da Universidade da Beira Interior (Departamento de
Gestão e Economia)

RESUMO:

- O presente artigo versa a problemática do progresso tecnológico segundo uma metodologia que tem sido seguida noutros países e que requer a utilização da função de produção Translog. Antes de proceder ao desenvolvimento da metodologia o autor revê o “estado da arte” passando em revista a forma como outras funções de produção têm tratado o tema na prossecução dos mesmos fins. Passa em seguida à exposição da metodologia seguida com base na função translog para finalizar com uma aplicação empírica a dois períodos recentes da economia portuguesa, com a apresentação dos resultados obtidos e com a sua interpretação.

PALAVRAS-CHAVE:

- *funções de produção, função custo, translog, progresso tecnológico, economia portuguesa, produção, custo, econometria*

Abstract

- The present paper speaks of the technological progress evolution using a methodological frame that has been tried in other countries and that requires the Translog production function. Before presenting this methodology the author reviews the state of the art of the subject analysing the way that other functions use to achieve the same aims. After this he passes to the exposition of the translog frame and ends with an empirical application of the methodology to two recent periods of the portuguese economy, the presentation of the estimates found and its interpretation.

KEY-WORDS:

- *Production functions, cost function, translog, technological progress, portuguese economy, production, cost, econometric*

INTRODUÇÃO

O presente artigo debruça-se sobre a questão da evolução temporal do progresso tecnológico, ou antes da evolução tecnológica, numa economia, num sector ou num ramo de actividade. Um tema importante directamente relacionado com esta temática consiste no estudo teórico da forma como as diferentes tecnologias tratam a questão da evolução tecnológica, isto é, como é que as funções de produção mais conhecidas - da Cobb-Douglas, à CES e à VES/Translog - tratam da inclusão do progresso técnico. Além disso há a questão do progresso tecnológico incorporado ou não incorporado nos factores de produção, sejam eles a mão-de-obra, o capital ou os inputs intermédios. Estes temas prendem-se com a necessidade de estudar a forma como o progresso tecnológico se tem reflectido na utilização dos *inputs*: têm-se mantido estáveis ao longo do tempo ou tem havido algum enviesamento, e neste caso em que sentido tem evoluído esse enviesamento? Utilizando em termos relativos ou em termos absolutos mais (menos) trabalho e conseqüentemente menos (mais) capital?

Começaremos por estudar as várias teses que sobre esse tema se confrontam, para ver como é que cada uma das três funções atrás referidas tem tratado o progresso tecnológico, por fazer referência a estudos pioneiros que sobre este tema se têm levado a cabo quer no estrangeiro quer em Portugal e aos resultados empíricos conseguidos pelos seus autores e terminaremos com uma breve aplicação empírica a dados referentes a dois períodos recentes da economia portuguesa.

1. EVOLUÇÃO TECNOLÓGICA: AS VÁRIAS TESES EM PRESENÇA

É interessante verificar qual foi o efeito das subidas dos preços dos combustíveis no passado e sobretudo ser capaz de prever o que poderá acontecer se ocorrerem (e isto é inevitável) novas subidas desses preços num futuro mais ou menos próximo. Será que, como aconteceu nos EUA, a evolução do progresso técnico tem sido enviesada, isto é, poupadora de energia, ou, por outras palavras, que os aumentos dos preços da energia têm como consequência distorcer as curvas de procura e de custos? Será que, como previam Jorgenson e Fraumeni em 1981, novos aumentos de combustíveis vão provocar um abrandamento no crescimento da produtividade, com o reflexo negativos que daí podem advir?

Desde Tinbergen (1942) e Solow (1957) que o crescimento da produtividade tem sido associado a alterações tecnológicas e a deslocações ao longo do tempo das funções de produção agregadas.

Mais recentemente a atenção dos técnicos tem-se debruçado, para além destas, sobre outras questões, nomeadamente sobre os efeitos de elevações dos preços energéticos ao nível do factor trabalho e ao nível do crescimento da produtividade (multifactorial).

Hudson e Jorgenson concluíram, a este propósito, que, após o primeiro choque petrolífero, e para os EUA, as elevações de preços dos produtos energéticos fizeram abrandar o crescimento da produtividade devido à complementaridade energia-capital, à substituibilidade energia-trabalho e ao facto de o preço da energia ter induzido alterações no *mix* da produção.

Por sua vez, Berndt e Wood, primeiro em 1975 e depois em 1979, concluíram o mesmo quanto à complementaridade energia-capital e à substituibilidade energia-trabalho para a indústria dos EUA.

Um dos problemas que se tem vindo a discutir é como se repartem os efeitos do progresso técnico sobre os factores de produção, particularmente sobre a sua procura.

Sobre esta temática há diversos pontos de vista ou teses.

Um deles perfilha a tese de que o progresso técnico é do tipo neutro à Hicks, uma tese segundo a qual as procuras de factores são igualmente afectadas por esse efeito, ou, por outras palavras, são afectadas segundo proporções idênticas. Os dois estudos anteriormente referidos partiram do princípio de que o progresso técnico era deste tipo.

Outras hipóteses alternativas, relativamente aos efeitos do progresso técnico sobre a procura dos factores de produção, são a neutra à Harrod, uma representação chamada "*labour-using*" e, em princípio, poupadora dos restantes factores de produção, e a neutra à Solow, uma representação "*capital-using*" (e poupadora dos restantes factores de produção).

Com a consideração dos *inputs* intermédios como factor de produção¹ surgem outras teses alternativas em que, com o incremento do valor acrescentado, apenas são aumentados o capital e o trabalho, não os *inputs* intermédios. Por sua vez, a tese do progresso técnico neutro à Leontief considera taxas diferenciadas de crescimento quer do trabalho quer do capital, enquanto a tese do progresso técnico neutro à Leontief-Hicks considera taxas de crescimento idênticas para os dois factores.

As diferentes teses quanto ao papel do progresso técnico tiveram um papel fulcral no aparecimento e no desenvolvimento de modelos teóricos relativos ao desenvolvimento tecnológico e ao crescimento económico. Contudo, poucos estudos empíricos foram levados a cabo e igualmente poucos resultados metodológicos ou empíricos foram publicados na literatura especializada.

De entre as poucas e honrosas excepções temos, entre outras, as seguintes:

A) Ao nível metodológico:

- Berndt e Jorgenson que elaboraram uma metodologia destinada a testar a especificação da neutralidade à Hicks contra uma representação não-neutral, mais geral;
- Dale W. Jorgenson e Barbara Fraumeni que sugeriram uma outra metodologia alternativa para testar outras formas de progresso técnico, metodologia que assenta na imposição de restrições sobre a função custo como a concavidade global e a monotonicidade;
- Ernest R. Berndt e David O. Wood que propuseram uma metodologia, derivada da proposta por Jorgenson e Fraumeni, que permite testar uma grande variedade de especificações do progresso técnico.

B) Ao nível empírico:

- - O estudo de Berndt e Jorgenson (1973) foi o pioneiro; a metodologia desenvolvida pelos autores foi aplicada por eles próprios a nove sectores industriais dos EUA; trata-se de uma metodologia que registou a adesão de diversos seguidores, nomeadamente: Hans Bingswanger (1974), E. R. Berndt e

¹ São conhecidos alguns estudos em que os *inputs* intermédios são tratados como factor de produção. A título de exemplo cita-se o caso de Kreijger, R. G. - (1980) - "Imports as a Factor of Production: the Substitution Characteristics of Labor, Capital and Material Inputs", Universidade de Amsterdão, Report, AF 70 em que este autor faz uma aplicação aos Países Baixos (Holanda) e o caso de M. J. Vilarés - "Os Bens Intermédios Importados como Factor de Produção - Aplicação à Economia Portuguesa" in A. Sousa e alii (1989) - "Nova Economia Portuguesa - Estudos em Homenagem a António Manuel Pinto Barbosa", Universidade Nova de Lisboa, pp 499-539 - em que é feito um estudo idêntico aplicado a Portugal.

D. O. Wood (1975), John Wills (1979), E. R. Berndt e M. S. Kaled (1979), Alden Toevs (1980), John R. Moroney e John M. Trapani (1981), entre outros.

- - A segunda tese, desenvolvida por Jorgenson e Fraumeni, introduz restrições à função custo como a da concavidade global destinada a garantir o *bom-comportamento* da função custo estimada no espaço constituído pelos preços dos *inputs* (positivos) e não apenas pela informação limitada contida nas amostras, informação em que assentam as estimativas dos parâmetros.

A especificação da concavidade global de Jorgenson e Fraumeni pôs em confronto directo a tese destes autores com uma questão entretanto levantada por John R. Moroney e John M. Trapani, segundo a qual quando o progresso técnico é não neutro a função de custo translog - a função usada no estudo - não é localmente côncava, isto é, não é côncava na região definida pelos seus preços e quantidades históricos. Contudo, uma vez rejeitado o progresso técnico neutro à Hicks, quando se impõem restrições a esse progresso técnico, descobre-se que a função de custo translog estimada é localmente *bem-comportada*.

- - A terceira e última proposta metodológica, a devida a Berndt e Wood, tornou-se necessária porque todos os estudos anteriores e acima referidos se limitaram a apreciar o progresso técnico neutro ou não-neutro à Hicks, nenhum deles tendo estudado outras alternativas como as neutralidades à Harrod, Solow, Leontief ou Leontief-Hicks. Foi isso que fizeram Ernest R. Berndt e David O. Wood em estudos publicados em 1982 e 1985² em que adoptaram a perspectiva de Jorgenson e Fraumeni acerca das restrições relativas à concavidade global que depois aplicaram a diversos tipos de especificações do progresso técnico.

2. O PROGRESSO TECNOLÓGICO E A FORMA COMO TEM SIDO TRATADO PELAS DIFERENTES FUNÇÕES DE PRODUÇÃO

Ao longo dos anos, várias têm sido as metodologias apresentadas para apreciar o progresso tecnológico, metodologias que variam de estrutura produtiva para estrutura produtiva. Vejamos como é que as diferentes tecnologias de produção têm estudado e integrado analiticamente o progresso tecnológico, particularmente as tecnologias de produção do tipo Cobb-Douglas, CES, de Leontief e Translog.

2.1. O PROGRESSO TECNOLÓGICO E A F. P. DE COBB-DOUGLAS

O problema do progresso tecnológico, ou apenas progresso técnico, expressões que aqui usamos com o mesmo sentido, começou por ser levantado por uma análise de Mendershausen em que este punha em causa a questão da constância da função de produção durante longos períodos, hipótese admitida tanto pela dupla Cobb e Douglas na sua análise pioneira, como por outros autores. Aliás, aquele mesmo autor para além de levantar o problema pôs também em causa que o expoente α da função (de Cobb - Douglas) simplificada ($Y = bL^\alpha K^{1-\alpha}$) fosse a elasticidade do produto em relação ao trabalho, resultado que até aí ninguém questionara, com a argumentação de que se se admitisse a evolução das variáveis Y, L e K ao longo do tempo, por exemplo, através das expressões

² Vidé Fericeli e Lésourd (eds) (1985) - "Energie: Modélisation et Econométrie", Economica, Paris, pp. 444-471.

$$Y^* = e^{at}, \quad L^* = e^{bt}, \quad K^* = e^{ct},$$

então aplicando logaritmos, e subtraindo a terceira equação da primeira e a terceira da segunda também viria, após alguns cálculos, $\ln Y^* - \ln K^* = (a-c)t \wedge \ln L^* - \ln K^* = (b-c)t$; dividindo membro a membro estas duas equações poderia afirmar-se que no longo prazo se teria

$$\ln\left(\frac{Y}{K}\right) \approx \frac{a-c}{b-c} \ln\left(\frac{L}{K}\right)$$

razão pela qual, segundo aquele autor, α deveria ser aproximado pelo quociente $\frac{a-c}{b-c}$.

É a partir deste resultado que Mendershausen sustenta que α em vez de representar a elasticidade deveria representar a evolução do progresso tecnológico³.

Desde há muito tempo que é reconhecida a extrema importância do progresso tecnológico na explicação do produto; contudo, a grande dificuldade para o seu estudo foi sempre, e continua a ser, como descobrir qual a melhor forma de introduzir explicitamente, o progresso tecnológico na função de produção para que, a partir daí, ele possa ser convenientemente tratado. Aukrust sugeriu que se introduzisse através da inclusão da variável t como expoente de uma exponencial como se faz na função de produção (1),

$$Y = AL^\alpha K^\beta e^{\alpha t} \quad (1)$$

Tinbergen, um conceituado economista holandês, seguiu a mesma metodologia, propondo para esse efeito a função de produção dinâmica⁴ seguinte

$$Y = A_0 L^a K^b e^{\alpha t} \quad (2)$$

função que no fundo é a dada por (1) e que logaritmicada dá

$$\ln Y = \ln A_0 + a \ln L + b \ln K + \alpha t \quad (3)$$

Como vantagens deste tipo de representação foram-lhe reconhecidas as seguintes:

- I. Possibilidade de estimação directa dos efeitos do tempo sobre o produto;

³ Do inglês "technical change".

⁴ A introdução de t não modifica as propriedades fundamentais da função uma vez que ela continua ainda a satisfazer as condições oportunamente referidas; além disso σ continua a vir igual a 1, pois que a razão $R = \frac{\partial Y}{\partial L} \frac{\partial L}{\partial Y} = \frac{\partial K}{\partial Y} \frac{\partial Y}{\partial K} = \frac{\partial L}{\partial L}$

não se modifica dado o facto de $R = \frac{a A_0 L^{a-1} K^b e^{\alpha t}}{b A_0 L^a K^{b-1} e^{\alpha t}} = \frac{a}{b} \cdot \frac{K}{L}$. (Nesta função K tem o significado que nas duas ou três expressões anteriores tinha a variável C , isto é, representa o capital).

2. Possibilidade de representação algébrica das oscilações temporais da função de produção através do parâmetro α associado à variável t , um “parâmetro de escala das relações entre os factores e o produto”;
3. Possibilidade de as referidas oscilações ou deslocações desta função serem consideradas, para efeitos de análise, como o resultado do progresso tecnológico; e
4. Possibilidade de o coeficiente α ser considerado como um parâmetro de eficácia ou de produtividade global dos factores, eficácia essa devida às modificações de tecnologia.

Se a tecnologia aplicada à produção evoluir regularmente no tempo então mesmo que a quantidade de *inputs* consumidos se mantenha inalterada, o produto pode aumentar ao longo desse mesmo tempo.

Para corroborar o que afirmamos vamos subtrair $\ln L$ aos dois termos de (3)

$$\begin{aligned} \ln Y - \ln L &= \ln A_0 + \alpha t + a \ln L - \ln L + b \ln K \\ \Leftrightarrow \ln \frac{Y}{L} &= \ln y = \ln A_0 + \alpha t + (a - 1) \ln L + b \ln K \end{aligned}$$

onde $y=Y/L$, é a produtividade do factor trabalho (mão-de-obra). Derivando esta expressão em ordem a t vem

$$\frac{dy}{y} = \alpha + (a - 1) \frac{dL}{L} + b \frac{dK}{K}$$

e considerando a taxa de crescimento do capital nula, isto é, $d \ln k = dK/K = 0$, vem

$$\frac{dy}{y} = \alpha + (a - 1) \frac{dL}{L} \quad (4)$$

expressão que significa que a taxa de crescimento da produtividade (ou produto *per capita*), $y=Y/L$, é dada pelo parâmetro α , (parâmetro associado à variável t), pois que o resto são as eventuais economias de escala - no caso de $a > 1$ - ou deseconomias de escala - no caso de $a < 1$ - que acompanham o crescimento do factor trabalho.

Vários estudos empíricos concluíram que 30% a 50% do crescimento do produto é explicado pelo factor tempo ou, por outras palavras, pelo progresso tecnológico.

Os críticos desta forma de representar o progresso tecnológico e das conclusões extraídas a partir deste tipo de representação referem o seguinte:

- Que o crescimento do capital tem mais importância para o crescimento do produto do que mostram alguns estudos feitos naquela base;
- Que o progresso tecnológico não é um *factor exógeno, imaterial, desincorporado e independente dos factores*⁵; e

⁵ cf. A. Sousa, (1970) - “Funções de Produção de Cobb-Douglas na Indústria Transformadora Portuguesa”, Centro de Estudos de Planeamento, nº2. p. 19. Resek e outros são da mesma opinião.

- Que é mais lógico considerar que o progresso tecnológico é veiculado pelos factores de produção e que não pode ser isolado das variações qualitativas e quantitativas destes factores.

Progresso Tecnológico Incorporado Nos Factores

No seguimento do que se acaba de afirmar parece, pois, poder concluir-se que é através da renovação e do acréscimo de capital afecto à produção, isto é, através da renovação de equipamentos mais modernos - e que exigem uma crescente qualificação da mão-de-obra que com eles trabalha ou opera - que se introduzem os progressos da tecnologia no processo produtivo.

É através da substituição escalonada do equipamento, cuja qualidade tecnológica evolui no tempo, que se processa a elevação da produtividade e a modificação das combinações produtivas.

Tanto a evolução temporal do equipamento como a modificação progressiva da composição do stock de capital nos permite admitir:

- Por um lado, que o capital não é homogéneo;
- Por outro, que ele é variável quer qualitativa quer quantitativamente no tempo.

Para prosseguir suponhamos agora,:

1. Que é possível identificar os investimentos realizados em cada momento t ;
2. Que quer o equipamento quer as instalações construídas no momento v beneficiam do progresso tecnológico disponível nesse momento;
3. Que o investimento realizado em v é homogéneo;
4. Que o stock de capital existente em t é um somatório de subconjuntos de bens de produção constituído por diferentes gerações realizadas em v ;
5. Que, ao nível macroeconómico, se pode admitir uma certa substituíbilidade interfactorial.

Nesta ordem de ideias, o investimento realizado em v depende quer do nível tecnológico que ele incorpora quer do seu próprio volume. Se admitirmos que essa tecnologia evolui segundo uma taxa h , então o *input* K_v pode traduzir-se pela expressão

$$K_v = I_v \cdot e^{hv} \quad (5)$$

Representemos por L_v a mão-de-obra utilizada pelo *input* K_v e por Y_v a produção realizada nesse mesmo momento com estes dois factores. A relação entre os *inputs* K e L é dada, segundo a tecnologia de Cobb-Douglas, pelas funções seguintes:

$$Y_v = A \cdot L_v^a \cdot K_v^b = A \cdot L_v^a \cdot I_v^b \cdot e^{\lambda v} \quad (6)$$

com $\lambda = hb$. Admitindo uma taxa de depreciação do capital de δ , num momento t posterior a v , esse capital $I_v(t)$ deprecia-se, através dessa taxa, na sua capacidade de produção em relação a I_v , de acordo com a relação

$$I_v(t) = I_v e^{-\delta(t-v)} = I_v e^{\delta(v-t)} \quad (7)$$

donde, introduzindo (7) na expressão final de (6), vem

$$Y_v(t) = A \cdot L_v^a(t) \cdot I_v^b(t) e^{(\lambda+\delta b)v-b\delta t} \quad (8)$$

A afectação da mão-de-obra às diferentes gerações de capital altera-se com o tempo fruto quer de novos investimentos quer da depreciação dos antigos; este processo é acompanhado da libertação de mão-de-obra excedentária. Contudo, segundo um teorema fundamental atribuído a Fisher, há a garantia que “*não pode haver grandes diferenças entre as produtividades marginais das diversas afectações*”.

Derivando (8) em ordem a L vem a produtividade marginal da mão-de-obra que opera com o capital da geração v

$$\frac{\partial Y_v(t)}{\partial L_v(t)} = a \frac{Y_v(t)}{L_v(t)} \quad (9)$$

donde se deduz, resolvendo primeiro em ordem a L_v e depois em ordem a Y_v , que

$$L_v(t) = M(t) e^{\lambda^* v - \delta^* t} I_v^{\frac{b}{1-a}} \quad (10)$$

$$Y_v(t) = AM^a(t) e^{\lambda^* v - \delta^* t} I_v^{\frac{b}{1-a}} \quad (10')$$

onde

$$M(t) = \left[\frac{\partial Y_v(t)}{\partial L_v(t)} \cdot \frac{1}{aA} \right]^{\frac{1}{a-1}}, \quad \lambda^* = \frac{\lambda + b\delta}{1-a}, \quad \delta^* = \frac{b\delta}{1-a}.$$

Dado o facto de (10) e (10') serem variáveis homogéneas em v então para obter quer o trabalho quer o produto agregados em t vamos integrar (10) e (10') como se indica

$$L(t) = \int_{-\infty}^t L_v(t) dv = M(t) e^{-\delta^* t} \int_{-\infty}^t I_v^{\frac{b}{1-a}} e^{\lambda^* v} dv \quad (11)$$

$$Y(t) = \int_{-\infty}^t Y_v(t) dv = AM^a(t) e^{-\delta^* t} \int_{-\infty}^t I_v^{\frac{b}{1-a}} e^{\lambda^* v} dv \quad (11')$$

(11') pode ainda escrever-se

$$Y(t) = AL^a(t)J^b(t) \quad (12)$$

onde o capital agregado é tecnologicamente homogêneo

$$J(t) = \int_{-\infty}^t I_v(t) e^{bv} dv \quad (13)$$

$J(t)$ é uma variável complexa pois se calcularmos a produtividade marginal do stock de capital K , i. é, se derivarmos $Y(t)$ em ordem a $K(t)$ vem

$$\frac{\partial Y(t)}{\partial K(t)} = ba \cdot L(t)^a \cdot J(t)^{b-1} \frac{dJ(t)}{dK(t)},$$

expressão que mostra que o progresso tecnológico é não neutro uma vez que dividindo a produtividade marginal do trabalho pela do capital vem

$$\frac{\partial Y(t)/\partial L(t)}{\partial Y(t)/\partial K(t)} = \frac{a}{b} \cdot \frac{J(t)}{L(t)} \cdot \frac{dK(t)}{dJ(t)}$$

Sendo o capital medido em termos de *inputs* tecnologicamente homogêneos, então as condições de neutralidade conduzem a

$$\frac{\partial Y(t)}{\partial J(t)} = b \cdot \frac{Y(t)}{J(t)} \Rightarrow R_J = \frac{a}{b} \cdot \frac{J(t)}{L(t)} \Rightarrow \sigma = 1.$$

Este tipo de neutralidade é conhecida como neutralidade à Solow.

Passemos agora à relação marginal de substituição de um equipamento da geração v por outro da geração t . Esta relação depende da taxa a que evolui a tecnologia incorporada

$$-\frac{dI_t(t)}{dI_v(t)} = e^{h(v-t)} \quad (14)$$

dependendo o ritmo de substituição do equipamento antigo da sua vida útil que, por sua vez, é, também, função da tecnologia incorporada.

O tratamento da depreciação por usura física reveste geralmente umas das três situações seguintes:

- O progresso tecnológico é tão intenso que a usura física quase não actua, situação em que se ignora essa usura;
- Através de reparações de equipamentos que melhoram a sua qualidade técnica com reflexos na aceleração da cadência de produção ou nos automatismos;

- O contacto do equipamento antigo com o equipamento novo *contamina* o primeiro sendo esta *contaminação* suficiente, segundo Kaldor, para compensar a depreciação por usura física; esta situação leva a ignorar uma e outra.

Nos dois últimos casos a depreciação por usura física é considerada um meio de difusão dos efeitos do progresso tecnológico sobre o capital mais antigo.

Atendendo à dificuldade prática em identificar analiticamente e em quantificar as três situações, que em grande parte das vezes se produzem simultaneamente, consideram-se por vezes modelos que admitem apenas “*uma taxa de progresso tecnológico que indica as diferenças qualitativas globais entre dois capitais de concepções tecnológicas distanciadas no tempo*”⁶.

Nelson refere uma metodologia para identificar o progresso tecnológico incorporado nos factores, começando por exprimir (14) com valores discretos vem

$$J(t) = \sum_{v=0}^t I_v(t)(1+h)^v \quad (14')$$

$$J(t) = (1+h)^{-\tilde{a}_0} (1+h)^t \sum_{v=0}^t I_v(t)(1+h)^{v-t+\tilde{a}_0}$$

onde \tilde{a}_0 é a idade média do stock de capital existente no início do período ($v=0$).

Substituindo $\sum_{v=0}^t I_v(t) = K(t)$

vem

$$\begin{aligned} J(t) &\approx (1+h)^{-\tilde{a}_0} (1+h)^t K(t) \sum_{v=0}^t \frac{I_v(t)}{K(t)} (1+hv + h\tilde{a}_0 - ht) \\ &\approx (1+h)^{-\tilde{a}_0} (1+h)^t K(t) \left[(1+h\tilde{a}_0) \sum_{v=0}^t \frac{I_v(t)}{K(t)} \right] - h \frac{\sum_{v=0}^t I_v(t)(t-v)}{K(t)} \\ &\approx (1+h)^{-\tilde{a}_0} (1+h)^t K(t) [1+h(\tilde{a}_0 - \tilde{a}_t)] \end{aligned} \quad (15)$$

onde \tilde{a}_t é a idade média dos stocks de capital da geração t , ou seja,

$$\tilde{a}_t = \sum_v \frac{I_v(t)(t-v)}{K(t)}$$

Há um método de cálculo que permite identificar o capital tecnologicamente homogéneo pelas idades médias das gerações de capital que começa por fazer uma

⁶ Cf. Sousa, p. 24.

simplificação, que permite determinar esse capital tecnologicamente homogêneo, quer pela diferença de idades quer pela admissão da hipótese de que essas idades médias praticamente não variam, isto é, que são tais que $1 + h(\tilde{a}_0 - \tilde{a}_t) \approx 1$. Daqui vem

$$\frac{\Delta J}{J} \approx h(1 - \Delta \tilde{a}) + \frac{\Delta K}{K} \quad (16)$$

A partir da expressão (12) pode escrever-se

$$\frac{\Delta Y}{Y} = \frac{\Delta A}{A} + a \frac{\Delta L}{L} + b \frac{\Delta J}{J} \quad (17)$$

expressão onde, por substituição, vamos introduzir (16)

$$\frac{\Delta Y}{Y} = \left[\frac{\Delta A'}{A'} + bh(1 - \Delta \tilde{a}) \right] + a \frac{\Delta L}{L} + a \frac{\Delta L}{L} + b \frac{\Delta K}{K} \quad (17')$$

Esta equação permite reduzir a parte desconhecida do factor residual e decompô-la nas seguintes parcelas:

- bh - parcela que mede os efeitos do progresso tecnológico incorporado no capital;
- $\Delta \tilde{a}$ - parcela que dá os efeitos da aproximação da tecnologia praticada à melhor tecnologia disponível; e
- $\frac{\Delta A'}{A'}$ - parcela que dá o progresso tecnológico não incorporado ou a parte desconhecida do factor residual.

O que se fez para o factor capital poderia fazer-se para o factor trabalho e para o factor energia ou para outros factores que se considerassem.

As primeiras aplicações empíricas da função Cobb-Douglas a Portugal, datam já dos anos 60/70, e ficaram a dever-se a F. Jesus, A. Sousa, J. A. Girão, entre outros. Os dados utilizados foram em geral do tipo cronológico, tendo também havido algumas aplicações que recorreram a dados seccionais (*cross-section*). A questão do progresso tecnológico foi particularmente apreciada por A. Sousa no seu estudo sobre a função de produção de Cobb-Douglas para estudar a indústria transformadora portuguesa⁷.

2.2 O PROGRESSO TECNOLÓGICO E A F. DE PRODUÇÃO CES

A função de produção CES, oportunamente estudada, foi apresentada pela primeira vez em 1961 no decurso de investigações relacionadas, precisamente, com o progresso tecnológico.

As constantes A e B da função de produção CES definida por

$$Y = \gamma \left[AE^{-c} + BK^{*-c} \right]^{-\frac{1}{c}} \quad (18)$$

⁷ A. Sousa (1970) - "Funções de Produção de Cobb-Douglas na Indústria Transformadora Portuguesa", Centro de Estudos de Planeamento, nº 2

são conhecidas como parâmetros de distribuição, de repartição ou de intensidade, sendo a sua soma um indicador de escala. Suponhamos que podemos escrever $A + B = \gamma^{-c}$ tal que

$$A + B = \gamma^{-c} \mu - \gamma^{-c} \mu + \gamma^{-c} = \gamma^{-c} \mu + \gamma^{-c} (1 - \mu).$$

Fazendo $A = \gamma^{-c} \mu$ $B = \gamma^{-c} (1 - \mu)$ e substituindo A e B na expressão da função CES pelos valores acabados de escrever vem a função (18) com o novo aspecto

$$Y = \gamma \left[\mu E^{-c} + (1 - \mu) K^{-c} \right]^{-\frac{1}{c}} \quad (18')$$

onde a expressão entre parêntesis rectos é uma combinação linear convexa de E^{-c} e de K^{-c} .

Os parâmetros c e μ podem estimar-se aplicando técnicas de regressão à expressão (18'). O valor de γ obtém-se destes e do quociente Y/X onde X representa

$$X = \left[\mu E^{-c} + (1 - \mu) K^{-c} \right]^{-\frac{1}{c}} \quad (19)$$

Se Y varia com K e com E então essa variação explica-se pela existência de progresso tecnológico, progresso esse que pode ser medido pela variação de γ . Daí que este parâmetro seja interpretado como um indicador do progresso técnico neutro à Hicks.

Em termos matemáticos cada valor de γ define uma função do tipo (18') e fazendo-se variar γ encontram-se as diferentes funções. Uma série de valores γ pode assim ser considerada função do tempo, razão pela qual se pode escrever $\gamma(t) = A(t)$. Se admitirmos, como é habitual, que o progresso tecnológico cresce a uma taxa constante, por exemplo α , então podemos escrever

$$\gamma_t = \gamma_0 e^{\alpha t} \quad (20)$$

Uma forma de introduzir analítica e explicitamente esse progresso tecnológico na função de produção CES é juntar (20) e (18')⁸ de forma a encontrar

$$Y = \gamma_0 e^{\alpha t} \left[\mu E^{-c} + (1 - \mu) K^{-c} \right]^{-\frac{1}{c}} \quad (21)$$

A expressão (21) é uma função homogénea⁹ de grau 1 e como tal não admite economias de escala pelo emprego de maiores quantidades de *inputs* primários. Contudo, abandonando a hipótese de linearidade a função

⁸ É, aliás, o método seguido por Tinbergen e tantos outros no caso da função de produção de Cobb-Douglas, como já se viu anteriormente.

⁹ Sobre esta noção ver, por exemplo, Novshek (1993), página 104, que refere que f é uma função homogénea de grau k se $f(\lambda x) = \lambda^k f(x)$ para todo o $x \in S$ e $\lambda \in \mathbb{R}$.

$$Y = \gamma \left[\mu E^{-\varepsilon} + (1 - \mu) K^{-\varepsilon} \right]^{-\frac{\varepsilon}{\varepsilon - 1}} \quad (22)$$

é de grau ε indicando $\varepsilon - 1$ os rendimentos de escala associados a uma dada tecnologia.

Para a estimação do grau ε da função e das economias de escala (através da dedução de uma unidade àquele parâmetro) podem pôr-se duas hipóteses:

- Hipótese A): partir de (22), após substituir o segundo factor do lado direito da equação por (19), e escrever

$$\ln Y = \ln \gamma + \varepsilon \ln X,$$

donde sai

$$\varepsilon = \frac{\ln Y - \ln \gamma}{\ln X};$$

- Hipóteses B): Fazer

$$\gamma^{-\frac{\varepsilon}{\varepsilon - 1}} \mu = A \quad \gamma^{-\frac{\varepsilon}{\varepsilon - 1}} (1 - \mu) = B$$

e introduzir em (22) de molde a encontrar

$$Y = \left[A E^{-\varepsilon} + B K^{-\varepsilon} \right]^{-\frac{\varepsilon}{\varepsilon - 1}} \quad (23)$$

Com operações simples (23) pode escrever-se com o duplo aspecto

$$Y = \left\{ A E^{-\varepsilon} \left[1 + \frac{B}{A} \left(\frac{E}{K} \right)^{\varepsilon} \right] \right\}^{-\frac{\varepsilon}{\varepsilon - 1}} = \left\{ B K^{-\varepsilon} \left[1 + \frac{A}{B} \left(\frac{K}{E} \right)^{\varepsilon} \right] \right\}^{-\frac{\varepsilon}{\varepsilon - 1}}$$

Mas aplicando logaritmos a estas duas equações e usando outro resultado já nosso conhecido vem

$$\ln Y = -\frac{\varepsilon}{\varepsilon - 1} \ln A + \varepsilon \ln E - \frac{\varepsilon}{\varepsilon - 1} \ln \left(1 + \frac{\pi_K}{\pi_E} \right) \quad (24)$$

$$\ln Y = -\frac{\varepsilon}{\varepsilon - 1} \ln B + \varepsilon \ln K - \frac{\varepsilon}{\varepsilon - 1} \ln \left(1 + \frac{\pi_E}{\pi_K} \right) \quad (25)$$

onde $\pi_K = B E^{\varepsilon}$, $\pi_E = A K^{\varepsilon}$. Calculando as regressões (24) e (25) vêm os valores procurados; a experiência tem mostrado que as duas regressões podem conduzir a

valores ligeiramente diferentes para os mesmos parâmetros mas que essas diferenças são tão pequenas que se podem considerar aproximadamente iguais.

Sousa (1977, p.12) nota que a expressão (22) requer alguns cuidados quer em termos de tratamento quer em termos de interpretação, atendendo por um lado ao aparecimento de efeitos de escala motivados pela variação do nível dos *inputs* primários e por outro atendendo ao efeito do progresso tecnológico.

Como mostra este mesmo autor a expressão (22) é ainda uma expressão neo-clássica porquanto as condições de primeira ordem são satisfeitas

$$\frac{\partial Y}{\partial E} = \frac{\mu \varepsilon}{\gamma / \varepsilon} Y^{1+\frac{c}{\varepsilon}} E^{-(1+c)} > 0, \quad \frac{\partial Y}{\partial K} = \frac{(1-\mu)\varepsilon}{\gamma / \varepsilon} Y^{1+\frac{c}{\varepsilon}} K^{-(1+c)} > 0 \quad (26)$$

Quanto às condições de segunda ordem pode ser necessário requerer qualificações especiais no caso de $\varepsilon \neq 1$. As segundas derivadas são

$$\frac{\partial^2 Y}{\partial E^2} = \frac{\mu \varepsilon}{\gamma / \varepsilon} Y^{1+\frac{c}{\varepsilon}} E^{-(2+c)} \left\{ [1 - \sigma(1-\varepsilon)] \frac{\eta_E}{\varepsilon} - 1 \right\}, \quad \frac{\partial^2 Y}{\partial K^2} = \frac{(1-\mu)\varepsilon}{\sigma \gamma / \varepsilon} Y^{1+\frac{c}{\varepsilon}} K^{-(2+c)} \cdot \left\{ [1 - \sigma(1-\varepsilon)] \frac{\eta_K}{\varepsilon} - 1 \right\}, \quad \frac{\partial^2 Y}{\partial E \partial K} = \frac{1+\frac{c}{\varepsilon}}{Y} \frac{\partial Y}{\partial E} \frac{\partial Y}{\partial K} = \frac{1-\sigma(1-\varepsilon)}{\varepsilon Y} \frac{\partial Y}{\partial E} \frac{\partial Y}{\partial K} \quad (27)$$

com η_E e η_K as elasticidades do produto em relação aos preços dos factores. Vê-se desta expressão que sendo dada uma determinada tecnologia os valores das derivadas dependem dos valores de σ e ε , consoante o nível de *inputs* do factor. Por esse motivo, se $\varepsilon > 1$, isto é, se houver economias de escala, a redução da produtividade de cada um dos factores pode ser compensada pela verificação desse tipo de economias.

De facto, como σ é sempre positivo o sinal das derivadas de segunda ordem depende dos valores concretos de σ e de ε . Se $\varepsilon = 1$, caso que está associado a homogeneidade linear, a superfície definida por (25) vem côncava. Se $\varepsilon \neq 1$ pode vir ou não.

Partindo de (22) façamos tender para infinito um dos factores, no caso em apreço K, e vejamos o que acontece quer ao produto quer à produtividade média do outro factor, neste caso E, consoante os valores de c:

a) Se $c > 0$ - o produto tende para um limite finito dado por

$$\lim_{K \rightarrow \infty} Y \gamma \mu^{-\frac{\varepsilon}{c}} E^\varepsilon \quad (28)$$

tendendo a produtividade média, com E constante, para o valor também constante

$$\lim_{K \rightarrow \infty} \frac{Y}{E} = \lim_{K \rightarrow \infty} \left[A + \left(\frac{K}{E} \right)^{-c} \right]^{\frac{\varepsilon}{c}} E^{\varepsilon-1} = A E^{\varepsilon-1} \quad (29)$$

b) Se $c < 0$, quer o produto quer a produtividade média tendem para infinito:

$$\lim_{K \rightarrow \infty} Y = \infty, \lim_{K \rightarrow \infty} \frac{Y}{E} = \infty \quad (30)$$

Relativamente ao significado económico destes resultados tem-se:

- a) $c > 0 \Leftrightarrow \sigma < 1$ - a substituíbilidade entre factores é pequena pois estes são parcialmente complementares; nestas condições a constância ou não crescimento de um dos factores limita mais ou menos rapidamente o crescimento do produto provocado pelo crescimento do outro factor;
- b) $c < 0 \Leftrightarrow \sigma > 1$ - a substituíbilidade entre os factores é grande; nestas condições apesar da constância ou não crescimento de um dos factores o acréscimo de consumo do outro faz crescer o produto, porque compensa, por substituição, a escassez do primeiro.

Dinamização da Função CES e o Progresso Tecnológico

A dinamização da função CES torna-se necessária para melhor se poder apreciar a natureza do progresso tecnológico.

Para estudar os efeitos do progresso tecnológico sobre a evolução do produto recorre-se à dinamização dos parâmetros de intensidade ou de distribuição, A e B, com estes expressos em unidades físicas. Esta técnica consiste em fazer variar no tempo os parâmetros A e B, parâmetros que, como já referimos, são considerados como indicadores de eficácia de E e de K.

Representemos A por A^{*-c} e B por B^{*-c} e introduzamos estas parâmetros na função de produção habitual

$$Y = \left[(AE)^{-c} + (BK)^{-c} \right]^{-\frac{1}{c}} \quad (31)$$

onde se retirou logo os asteriscos (*) para não sobrecarregar a notação. Esta nova função continua a satisfazer as condições de primeira e de segunda ordens pois

$$\begin{aligned} \frac{\partial Y}{\partial E} = p_e = A^{-c} \left(\frac{Y}{E} \right)^{1+c} > 0, \quad \frac{\partial Y}{\partial K} = p_k = B^{-c} \left(\frac{Y}{K} \right)^{1+c} > 0, \quad \frac{\partial Y}{\partial A} = p_a = E^{-c} \left(\frac{Y}{A} \right)^{1+c} > 0, \\ \frac{\partial Y}{\partial B} = p_b = K^{-c} \left(\frac{Y}{B} \right)^{1+c} > 0 \\ \frac{\partial^2 Y}{\partial E^2} = - \left[A^c (1+c) \left(\frac{Y}{E} \right)^c \left(\frac{Y - E p_e}{E^2} \right) \right] < 0 \end{aligned}$$

Prova-se ainda, relativamente às segundas derivadas, que, com excepção de uma, todas as outras derivadas são negativas $\left(\frac{\partial^2 Y}{\partial K^2}, \frac{\partial^2 Y}{\partial E^2}, \frac{\partial^2 Y}{\partial A^2}, \frac{\partial^2 Y}{\partial B^2}, \frac{\partial^2 Y}{\partial A \partial B} \right)$, sendo a excepção $\frac{\partial^2 Y}{\partial E \partial B}$ que é positiva.

Admitindo a evolução de A e de B com o tempo, isto é, que $A=A(t)$ e $B=B(t)$, então estes indicam quer o sentido quer a natureza do progresso tecnológico.

Sejam $A(t)=A_0e^{\varepsilon_1 t}$, $B(t)=B_0e^{\varepsilon_2 t}$ com ε_1 e ε_2 as taxas de variação ou de crescimento de A e B, respectivamente. Introduzindo estas expressões em (31) vem

$$Y = \left[(A_0 e^{\varepsilon_1 t} E)^{-c} + (B_0 e^{\varepsilon_2 t} K)^{-c} \right]^{-\frac{1}{c}} \quad (33)$$

Na hipótese de $\varepsilon_1 = \varepsilon_2$ então (33) vem

$$Y = e^{\alpha t} \left[(A_0 E)^{-c} + (B_0 K)^{-c} \right]^{-\frac{1}{c}} \quad (34)$$

equivalendo isso a dinamizar o parâmetro γ de (22). A função fica dinamizada através da introdução do argumento t. Nestas condições diz-se que o progresso tecnológico é neutro à Hicks na medida em que as produtividades dos factores vêm igualmente afectadas, conforme se depreende de (33), o mesmo acontecendo à produtividade global.

Se $\varepsilon_1 \neq \varepsilon_2$ com c constante, então diz-se que a “*evolução da tecnologia não é neutra pois que R se altera ao longo de um eixo factorial definido por k constante*” (cf. Sousa). Sendo variáveis no tempo quer A quer B quer ainda a tecnologia então a taxa marginal de substituição modifica-se definindo os sucessivos R famílias sucessivas de isoquantas homotéticas.

Nestas condições ou as remunerações relativas dos factores se alteram também e K pode permanecer efectivamente constante, ou as remunerações relativas não se modificam suficientemente e K varia conduzindo às combinações produtivas que maximizam o produto ou minimizam o custo.

Relativamente às produtividades marginais dos factores pode dizer-se o seguinte:

- a) Se os factores de produção forem complementares então a produtividade marginal de um factor isoladamente pode aumentar tanto por efeito do progresso tecnológico como por efeito da lei dos rendimentos marginais decrescentes;
- b) Se os factores forem perfeitamente substituíveis a diminuição relativa de um factor não faz aumentar a sua produtividade marginal já que é indiferente consumir maior quantidade desse factor.

Para estudar o caso em que $\varepsilon_1 \neq \varepsilon_2$ vamos partir do quociente das duas primeiras equações de (32) e aplicar-lhe logaritmos

$$d \ln p_e - d \ln p_k = c(d \ln B - d \ln A) + (1+c)(d \ln K - d \ln E) \quad (35)$$

Como a elasticidade de substituição entre os factores homogéneos $\bar{E} = A.E$, $\bar{K} = B.K$ vem $\sigma = \frac{1}{1+c}$ razão pela qual se pode escrever

$$d \ln p_e - d \ln p_k = \sigma = \frac{1-\sigma}{\sigma} c(d \ln B - d \ln A) + \frac{1}{\sigma} (d \ln K - d \ln E)$$

Nesta ordem de ideias a neutralidade do progresso tecnológico pode classificar-se em:

QUADRO Nº 1

Tipo de Progresso tecnológico	Elasticidade	Condição
Poupadora de energia $\Leftrightarrow \frac{dp_e}{p_e} < \frac{dp_k}{p_k}$	$\sigma < 1$	$\frac{dB}{B} < \frac{dA}{A}$
Idem	$\sigma > 1$	$\frac{dB}{B} > \frac{dA}{A}$
Poupadora de capital $\Leftrightarrow \frac{dp_e}{p_e} > \frac{dp_k}{p_k}$	$\sigma < 1$	$\frac{dB}{B} > \frac{dA}{A}$
Idem	$\sigma > 1$	$\frac{dB}{B} < \frac{dA}{A}$
Neutra $\Leftrightarrow \frac{dp_e}{p_e} = \frac{dp_k}{p_k}$	$\sigma < 1$	$\frac{dB}{B} = \frac{dA}{A}$

Algumas aplicações pioneiras desta função à realidade portuguesa foram levadas a cabo por dois conhecidos autores: J. Cravinho, que estimou uma função de produção, sem progresso técnico, do tipo

$$\text{Log}\left(\frac{P}{L}\right) = a + b \text{Log}(w)$$

a várias indústrias portuguesas recorrendo para isso aos dados do inquérito industrial de 1964; e A. Sousa, que em 1977 publicou um trabalho desenvolvido acerca desta função e da sua aplicação à indústria transformadora portuguesa, no primeiro número da revista *Economia*¹⁰.

As conclusões deste último estudo, que teve que superar incompatibilidades de séries cronológicas dividindo-as em dois grupos, foram as seguintes:

A) Com dados relativos ao período 1953-1965¹¹:

- 1º. *Progresso tecnológico*: o progresso tecnológico detectado foi do tipo *labour saving*; o coeficiente de intensidade capitalística cresceu;
- 2º. *Elasticidade de substituição*: a elasticidade de substituição apresentou valor inferior à unidade;
- 3º. *Repartição do produto pelos factores*: concordantemente com o acabado de referir, a parte do trabalho no produto gerado aumentou ligeiramente;
- 4º. *Consequência do aumento do horário efectivo de trabalho*: este elemento actuou como elemento de moderação no crescimento do coeficiente de intensidade capitalística;

¹⁰ A. Sousa (1977) - "Funções de Produção CES na Indústria Transformadora Portuguesa". *Economia*, 1,1.01 77, p 5 a 38

¹¹ cf. obra citada página 30.

5º. *Rendimentos de escala*: os rendimentos de escala estimados foram praticamente insignificantes.

B) Com dados referentes ao período 1960-1970¹²:

- 1º. A *elasticidade de substituição* continuou a ser inferior à unidade;
- 2º. O *progresso tecnológico* continuou também a ser poupador do *input* trabalho;
- 3º. Os dois factores acabados de referir conjuntamente tiveram como reflexo um *aumento da participação do factor trabalho* na repartição do produto interno, admitindo a verificação das hipóteses neo-clássicas;
- 4º. Detectou-se a presença de *economias de escala significativas* embora não espectaculares.

Deixando este breve parêntesis acerca de algumas das primeiras aplicações desta função a Portugal passemos agora à questão teórica que nos preocupa neste momento.

2.3. O PROGRESSO TECNOLÓGICO E A FUNÇÃO DE PRODUÇÃO TRANSLOG

Embora se fale em função de produção translog é habitual trabalhar-se com sua função custo médio translog assim definida:

$$\ln c = \ln \alpha_0 + \sum_{i=1} \alpha_i \ln p_i + \frac{1}{2} \sum_{i=1} \sum_{j=1} \gamma_{ij} \ln p_i \ln p_j + \gamma_Y \ln Y + \frac{1}{2} \gamma_{YY} (\ln Y)^2$$

onde i representa o i -ésimo factor, $i=1(K), 2(L), 3(E), \dots$, p_i é o preço unitário desse mesmo factor, c é o custo médio (ou unitário) e os restantes elementos são os parâmetros do modelo.

A homogeneidade linear da função impõe as seguintes condições sobre os parâmetros:

$$\sum_i \alpha_i = 1 \quad \sum_i \gamma_{ij} = 0 \quad \sum_j \gamma_{ij} = 0 \quad \forall i$$

Esta função desenvolvida por Christensen, Jorgenson e Lau (1973, 1975) foi utilizada com o objectivo que aqui nos preocupa, por Moroney e Trapani (1981), Berndt e Wood (1982,1985), que a usaram para estudar e testar as diferentes especificações do progresso tecnológico, e Marie N. Fagan (1997), entre outros.

Esta função foi já utilizada em Portugal, embora em contextos ligeiramente diferentes, por Vilares¹³ (1989), pelo autor deste texto¹⁴ (Manso, 1991) e ainda por mais alguns (poucos) autores.

¹² cf obra citada pagina 38.

¹³ M. J. Vilares - "Os Bens Intermédios Importados como Factor de Produção. Aplicação à Economia Portuguesa" in A. Sousa e alii (1989) - "Nova Economia Portuguesa - Estudos em Homenagem a António Manuel Pinto Barbosa", Universidade Nova de Lisboa, pp 499-539. Vilares, usou ainda noutro estudo a função de produção de Leontief.

¹⁴ Vidé J. R. Pires Manso (1991) - "A Substituição de Factores na Economia Portuguesa - Uma Aplicação da Função Translog". Tese de Mestrado, ISEG

A função de produção translog é uma função de produção agregada, duplamente diferenciável caracterizada por rendimentos constantes à escala que pode ser definida em termos genéricos e para estudo da evolução tecnológica, acrescida de uma nova variável t, como

$$Y = f(K, L, E, t). \quad (36)$$

onde Y, K, L, E têm os significados então referidos. A inclusão da variável t (tempo) justifica-se pela necessidade de representar, na função de produção e isoladamente, as deslocações temporais extrínsecas do progresso técnico.

Associada com a função (36) e sendo dados os preços dos *inputs* e as quantidades dos *outputs*, existe uma função de custo unitária dual que reflecte a tecnologia da produção

$$c = C/Y = g(p_k, p_l, p_e, t)$$

onde p_k , p_l e p_e , C, e c continuam a representar, respectivamente, os preços dos *inputs*, o custo total e o custo unitário; t representa aqui as deslocações, para baixo, da função custo unitária devida ao progresso técnico considerado isoladamente.

A função custo unitária acabada de referir associada à função translog, é uma função que na forma logaritmizada e acrescentada dos novos parâmetros decorrentes da inclusão da variável t, toma o seguinte aspecto

$$\begin{aligned} \ln c = & \ln \alpha_0 + \gamma_Y \ln Y + \frac{1}{2} \gamma_{YY} (\ln Y)^2 + \sum_{i=1}^3 \alpha_i \ln p_i + \\ & + \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \gamma_{ij} \ln p_i \ln p_j + \sum_i \gamma_{iY} \ln p_i \ln Y + \frac{1}{2} \gamma_{tt} t^2 + \\ & + \sum_{i=1}^3 \gamma_{it} t \ln p_i + \gamma_{Yt} t \ln Y + \gamma_t t \end{aligned} \quad (37)$$

onde γ_{ij} , γ_t , α_i e α_0 (i, j=1(K), 2(L), 3(E)) representam os parâmetros.

A homogeneidade linear¹⁵ impõe as restrições já referidas anteriormente, acrescidas de outras que se deixam em (38)

$$\sum_{i=1}^3 \alpha_i = 1, \sum_{j=1}^3 \gamma_{kj} = 0, \sum_{j=1}^3 \gamma_{lj} = 0, \sum_{j=1}^3 \gamma_{ej} = 0, \sum_{i=1}^3 \gamma_{it} = 0 \quad (38)$$

O aparecimento da última equação fica a dever-se à inclusão da variável independente t.

As equações das quotas dos factores obtêm-se aplicando derivadas logaritmicas à expressão de c; usando o Lema de Shephard, admitindo que o *output* e os preços dos *inputs* são fixos vem:

¹⁵ Sobre esta função ver, entre outros, J. R. P. Manso (1991) - "A Substituição de Factores na Economia Portuguesa - Uma Aplicação da Função Translog", Tese de Mestrado, ISEG, e Vilares (1987).

$$\frac{\partial \ln c}{\partial \ln p_i} = q_i \equiv \frac{p_i X_i}{cY} \quad (39)$$

$$= \alpha_i + \sum_{j=1}^3 \gamma_{ij} \ln p_j + \gamma_{iY} \ln Y + \gamma_{it} t \quad i = K, L, E$$

Para medir o efeito do progresso tecnológico na função de custo unitária derivemos a expressão (37) em relação à variável tempo (t). A equação resultante, (40), quantifica, sob a forma de taxa, a variação (diminuição) do custo:

$$\frac{\partial \ln c}{\partial t} = \dot{c} = \gamma_i + \sum_{i=1}^3 \gamma_{it} \ln p_i + \gamma_{iY} \ln Y + \gamma_{it} t \quad (40)$$

De notar que embora os parâmetros relativos ao enviesamento¹⁶ do progresso técnico γ_{it} ($i = K, L, E$) sejam constantes para o i -ésimo input, a variação do custo por unidade produzida, sob a forma de taxa, $\partial \ln c_t / \partial t = \dot{c}_t / c_t$, é endógena uma vez que os preços dos *inputs* variam ao longo do tempo. Como refere Makota Ohta¹⁷, com rendimentos constantes à escala, o dual da taxa de crescimento do custo, $\partial \ln c / \partial t$, é o negativo do primal da taxa de crescimento da produtividade multifactorial, $\partial \ln y / \partial t$, obtida a partir da função de produção (36). É esta a razão pela qual se pode dizer que (40) relaciona o crescimento da produtividade multifactorial com a evolução dos preços dos *inputs* e com o tempo.

Representando por b_i a taxa de crescimento da quota de custo do i -ésimo factor, pode escrever-se

$$b_i \equiv \frac{\partial q_i}{\partial t} \cdot \frac{1}{q_i} = \frac{\dot{q}_i}{q_i} \quad i = K, L, E \quad (41)$$

Com base neste resultado pode estabelecer-se a seguinte grelha que inter-relaciona os possíveis sinais dos coeficientes b_i (negativos, nulos ou positivos) e o tipo de evolução tecnológica:

- se $b_i < 0 \Leftrightarrow$ o progresso técnico diz-se poupador do *input* i , ou i -poupador;
- se $b_i = 0 \Leftrightarrow$ então o progresso técnico diz-se neutro em relação ao *input* i , ou é i -neutral; e
- se $b_i > 0 \Leftrightarrow$ então o progresso técnico diz-se utilizador do i -ésimo *input* ou i -utilizador.

Mas como

$$\frac{\partial q_i}{\partial t} = \gamma_{it}$$

então aquela taxa de crescimento pode escrever-se

¹⁶ Do inglês "biases", que significa enviesamento, pendor, propensão, inclinação, tendência.

¹⁷ Makota Ohta (1974) - "A Note on the Duality Between Production and Cost Function: Rate of Returns to Scale and Rate of Technical Progress", *Economic Studies Quarterly*, vol 25, 12, pp 63-65.

$$b_i \equiv \gamma_u \cdot \frac{1}{q_i} = \frac{\gamma_{it}}{q_i} \quad i = K, L, E$$

Mas das expressões das quotas dos factores (39) se vê, que como q_i ($i=K,L,E$), a quota de custo é sempre positiva então o sinal de b_i depende do sinal de γ_{it} e, nestas condições, segundo uma metodologia desenvolvida por K. Sato podemos diz-se que o processo tecnológico é:

- a') poupador do *input* i se o parâmetro γ_{it} é negativo.
- b') neutro em relação ao *input* i se γ_{it} é nulo (igual a zero) e
- c') utilizador do *input* i se o parâmetro γ_{it} é positivo.

(Note-se a correspondência entre os valores de b_i e de γ_{it} com $i=K,L,E$.)

De acordo com a abordagem que estamos a ver as expressões (39) podem ainda dar-nos a influência qualitativa do progresso tecnológico através do sinal do parâmetro da variável tempo (t) em cada uma das equações das quotas dos factores.

Como Berndt e Wood¹⁸, achamos que é oportuno dizer que a interpretação do coeficientes γ_{it} requer muito cuidado atendendo, entre outras, às seguintes razões:

1. Em primeiro lugar, dado o facto de os parâmetros γ_{it} serem constantes e como tal não variarem em consequência de alterações verificadas ao nível dos preços relativos dos *inputs*; neste sentido, e consequentemente, os parâmetros γ_{it} não podem ser usados para testar a validade da hipótese da *inovação induzida*, em particular, nos termos do referido por Binswanger¹⁹, quando se põe a hipótese de o comportamento do progresso técnico reflectir quer a escassez relativa dos factores ou *inputs* quer as variações dos preços desses mesmos *inputs*;
2. Em segundo lugar, os parâmetros γ_{it} representam mais o *efeito relativo* do que o *efeito absoluto* das alterações tecnológicas ao nível da procura dos factores de produção. Se, por exemplo, a alteração tecnológica em causa fosse considerada *neutra à Hicks* ($\gamma_{it}=0$ para $i = K, L, E$) então tal situação equivaleria a considerar que as procuras absolutas dos factores de produção seriam reduzidas de uma proporção comum, permanecendo as posições relativas das procuras dos factores inalteradas; acontece o mesmo quando $\gamma_{it}>0$, situação em que se diz que as alterações tecnológicas *são utilizadoras do factor* i ; mas este facto não implica, necessariamente, que a alteração tecnológica acarrete um aumento do valor absoluto da procura do factor i . Num tal caso a redução da procura do i -ésimo factor de produção individualmente pode não ser afectada da mesma forma que a procura dos restantes factores – em consequência do progresso técnico;
3. Em terceiro lugar, uma vez que a equação relativa à taxa de crescimento do custo unitário, (40), indica que este custo é afectado pelos preços dos *inputs* factoriais, então o sinal do parâmetro (γ_{it}) fornece informação de tipo qualitativo acerca do efeito de uma alteração de p_i na variação do custo

¹⁸ Berndt, E. R e Wood, D. O. - "Concavity and the Specification of Technical Progress", in Ferencelli e Lesourd (Editores) (1985) - "Énergie, Modélisation et Économétrie", Économica, Paris, pp 444-471.

¹⁹ V. Hans P. Binswanger (1974) e este mesmo autor com Vernon W. Ruttan (1978).

(crescimento da produtividade multifactorial); quando o progresso técnico ocorre então $\partial \ln c / \partial t$ (taxa de crescimento do custo) vem negativa. Se p_i aumenta e γ_{it} é positivo então o efeito dessa alteração ao nível do preço consiste no abrandamento da taxa de redução do custo unitário. Por sua vez se p_i aumenta e γ_{it} é negativo então a taxa de variação (redução) do custo acelera-se. Por isso quando a alteração tecnológica é *utilizadora do input i* e p_i aumenta, então, *ceteris paribus*, a taxa de crescimento da produtividade multifactorial desce; por sua vez quando a alteração tecnológica é *poupadora do input i* e p_i aumenta, então, *ceteris paribus*, a taxa de crescimento da produtividade multifactorial sobe.

Nos estudos empíricos acerca desta temática tem-se defendido que certas alterações tecnológicas do tipo *factor-augmenting* progridem na função de produção segundo uma exponencial com taxa constante. Visto através da função custo dual, tal alteração técnica "*factor-augmenting*" corresponde a uma diminuição técnica do preço do *input*. Quando uma alteração tecnológica do tipo "*factor-augmenting*" evolui de acordo com uma taxa (exponencial) constante, as unidades do *input* aumentadas X_{jt}^* e as naturais X_{jt} podem ser relacionadas através da relação:

$$X_{jt}^* = X_{jt} \cdot e^{\lambda_j t} \quad (42)$$

onde λ_j é a taxa de crescimento (constante) do *input j* e $t = T + t_0$, onde t_0 identifica um período inicial no tempo.

A relação correspondente entre os dois tipos de preços do "*input diminishing technical change*" pode ser esquematizada assim:

$$p_{jt}^* = p_{jt} \cdot e^{-\lambda_j t} \quad j = 1, 2, \dots \quad (43)$$

onde p_{jt}^* e p_{jt} são preços de factores associados com X_{jt}^* e X_{jt} , respectivamente.

Uma característica importante da especificação *translog* é que ela se adapta facilmente permitindo assim uma abordagem "*factor-augmenting*". Em particular se a função custo *translog* tivesse inicialmente sido escrita usando p_{jt}^* em vez de p_{jt} e se todos os termos envolvendo t tivessem sido desprezados, poder-se-ia ter tomado os logaritmos de (43) como se indica em seguida

$$\ln p_{jt}^* = \ln p_{jt} - \lambda_j t \quad (44)$$

e substituído em (37).

Usando (38) chega-se às seguintes expressões para α_t , γ_{kt} , γ_{lt} , γ_{et} , γ_{mt} e γ_u , respectivamente:

$$\begin{aligned} \alpha_t &= -(\alpha_k \lambda_k + \alpha_l \lambda_l + \alpha_e \lambda_e), & \gamma_{kt} &= -(\gamma_{kk} \lambda_k + \gamma_{kl} \lambda_l + \gamma_{ke} \lambda_e), \\ \gamma_{lt} &= -(\gamma_{kl} \lambda_k + \gamma_{ll} \lambda_l + \gamma_{le} \lambda_e) \\ \gamma_{et} &= -(\gamma_{ke} \lambda_k + \gamma_{le} \lambda_l + \gamma_{ee} \lambda_e), & \gamma_{ut} &= -(\gamma_{kt} \lambda_k + \gamma_{lt} \lambda_l + \gamma_{et} \lambda_e) \end{aligned}$$

Da análise das equações acima podem extrair-se algumas ilacções importantes:

- 1º. Que os coeficientes γ_{it} reflectem simultaneamente quer as taxas de crescimento quer as possibilidades de substituição entre *inputs*. Daí que a relação entre λ_i e γ_{it} seja uma relação complexa. Suponhamos, por exemplo, que λ_i é o maior dos vários parâmetros "input-augmenting". Esse facto não implicaria, no entanto, que a alteração tecnológica fosse *poupadora do input i* ($\gamma_{it}=0$), mas antes, como se vê pelas relações dos γ_{it} , que as taxas relativas *factor-aumentativas* se combinam com os parâmetros de substituição tecnológica γ_{ij} para determinar conjuntamente o efeito "bias" (tendência) refletido por γ_{it} .
- 2º. Quando a tecnologia de produção é Cobb-Douglas, situação que ocorre quando todos os $\gamma_{ij}=0$, e quando a alteração tecnológica é *input-utilizadora*, então a alteração tecnológica é *neutra à Hicks* ($\gamma_{it}=0$). Isto pode constatar-se através do simples exame das expressões dos γ_{it} também. Consequência deste facto é que não é possível identificar as taxas de variação λ_i quando a tecnologia é Cobb-Douglas.
- 3º. Se se estimar apenas as equações das cotas dos factores dadas por (39) deduzidas da função custo translog, e se não for possível identificar as diferentes taxas-acréscimo λ_i , então apenas se pode identificar as diferenças entre taxas de variação ($\lambda_i - \lambda_j$). Para que se possa identificar os λ_i , é necessário estimar os parâmetros, como por exemplo α_t , que aparecem tanto na função custo (37) como na equação relativa à taxa de crescimento do custo (40).
- 4º. Quando a alteração técnica é do tipo "input-using" pode exprimir-se a taxa global de diminuição do custo (crescimento da produtividade multifactorial) como a média ponderada das quotas dos factores e das taxas-acréscimo. Substituindo, por exemplo, as expressões dos γ_{it} em (40) obtém-se:

$$\frac{\partial \ln c}{\partial \bar{\alpha}} = -(q_k \lambda_k + q_l \lambda_l + q_e \lambda_e) \quad (45)$$

resultado que diz que a taxa *endógena* de crescimento do custo é igual à média ponderada das taxas *exógenas* de crescimento dos factores, com os coeficientes de ponderação dados pelas cotas de custo - endógenas também - dos factores.

5º. A teoria do crescimento dos factores permite testar vários tipos de progresso técnico, tendo, aliás, alguns deles sido discutidos na literatura teórica que versa a temática do crescimento económico; esses vários tipos de progresso técnico e as situações em que eles ocorrem deixam-se em seguida:

- Neutralidade à Hicks: quando $\lambda_k = \lambda_l = \lambda_e$
- Neutralidade à Harrod: quando $\lambda_l \neq 0, \lambda_k = \lambda_e = 0$
- Neutralidade à Solow: quando $\lambda_k \neq 0, \lambda_l = \lambda_e = 0$
- Neutralidade à Leontief: quando $\lambda_k, \lambda_l \neq 0, \lambda_e = 0$
- Neutralidade à Leontief/Hicks: quando $\lambda_k = \lambda_l \neq 0, \lambda_e = 0$.

De notar que sob a *neutralidade à Hicks*, todos os λ_j ($j=K,L,E$) são iguais, enquanto sob a *neutralidade à Harrod (Solow)*, $\lambda_l \neq 0$ ($\lambda_k \neq 0$), sendo todos os outros λ_j nulos. Quando os *inputs aumentativos* são apenas os primários – *neutralidade à Leontief* – λ_k e λ_l são não-nulos, mas $\lambda_e = 0$. Finalmente, quando se impõe que as taxas-acréscimo relativas aos valores acrescentados dos inputs primários sejam não-nulas e iguais ($\lambda_k = \lambda_l \neq 0$), com $\lambda_e = 0$, obtém-se a versão condicionada do progresso técnico neutro à Leontief, conhecida por *neutralidade à Leontief-Hicks*. É evidente que é possível especificar outros tipos de progressos técnicos.

Usando a função de custo translog com os factores K, L, E e M, podem estimar-se quer os parâmetros relativos às equações do custo dos factores (39) quer os relativos à equação de diminuição do custo total (40). Empiricamente esta última equação $\partial \ln c / \partial t$ pode ser aproximada, como propõe W. Erwin Diewert (1976), tomando o negativo da taxa de crescimento da produtividade multifactorial e recorrendo a uma aproximação de Tornquist (cf. Ernst R. Berndt e David O. Wood²⁰). Procedendo assim a estimação dos parâmetros pode fazer-se recorrendo a 7 representações alternativas de progresso técnico. Estes sete modelos, juntamente com os correspondentes números de parâmetros livres cuja estimação é necessária, estão referidos no próximo quadro:

QUADRO Nº 2
MODELOS ALTERNATIVOS DE PROGRESSO TÉCNICO

Modelo	Nº de Parâmetros livres
Apenas com a restrição da simetria	14
Progresso técnico <i>factor augmenting</i>	13
Idem com neutralidade à Hicks	10
Idem com neutralidade à Harrod	10
Idem com neutralidade à Solow	10
Idem com neutralidade à Leontief	11
Idem com neutralidade à Leontief-Hicks	10

Voltando novamente à expressão (37) derivemo-la em ordem a $\ln Y$ como se indica de seguida:

$$\frac{\partial \ln c}{\partial \ln Y} = \gamma_Y + \gamma_{YY} \ln Y + \sum_i \gamma_{Yi} \ln p_i + \gamma_{Yt} t$$

²⁰ Obra citada p. 453.

Esta expressão, que varia com o tempo e com os preços dos inputs, permite-nos apreciar a questão das economias de escala, através da análise do sinal da derivada. Assim

- se a derivada for negativa isso traduz a existência de economias de escala;
- se a derivada é positiva então isso traduz a existência de deseconomias de escala; e
- se a derivada é nula então isso traduz a não existência quer de economias (de escala) quer de deseconomias de escala.

3. APLICAÇÃO EMPÍRICA À REALIDADE PORTUGUESA

3.1. PREPARAÇÃO DO MODELO PARA ESTIMAÇÃO E OUTROS PORMENORES

Para ilustrar o que acabamos de dizer, e em particular o que se afirmou relativamente à função de produção *VES Translog* vamos usar dados relativos a um primeiro período de nove de anos (1976-1984) e publicados por M. J. Vilares e a um segundo período de dez anos (1980-1989) recolhidos pelo autor deste texto e apreciar a questão do tipo de progresso (evolução) tecnológico que ocorreu em Portugal nesses mesmos períodos, e a questão da substituíbilidade ou complementaridade entre cada dois pares de *inputs*. Os factores de produção considerados são o capital, o trabalho e os *inputs* intermédios importados, no primeiro período, e o capital, o trabalho e os *inputs* energéticos importados, no segundo período.

A função a utilizar é a função custo (unitária) translog dada por (37) e que passamos a transcrever convenientemente desenvolvida

$$\begin{aligned} \ln c = & \ln \alpha_0 + \gamma_Y \ln Y + \frac{1}{2} \gamma_{YY} (\ln Y)^2 + \alpha_1 \ln p_1 + \alpha_2 \ln p_2 + \alpha_3 \ln p_3 + \\ & + \frac{1}{2} \gamma_{11} \ln^2 p_1 + \gamma_{12} \ln p_1 \ln p_2 + \gamma_{13} \ln p_1 \ln p_3 + \gamma_{21} \ln p_2 \ln p_1 + \\ & + \frac{1}{2} \gamma_{22} \ln^2 p_2 + \gamma_{23} \ln p_2 \ln p_3 + \gamma_{31} \ln p_3 \ln p_1 + \gamma_{32} \ln p_3 \ln p_2 + \\ & + \frac{1}{2} \gamma_{33} \ln^2 p_3 + \gamma_{1Y} \ln p_1 \ln Y + \gamma_{2Y} \ln p_2 \ln Y + \gamma_{3Y} \ln p_3 \ln Y + \\ & + \frac{1}{2} \gamma_{tt} t^2 + \gamma_{1t} t \ln p_1 + \gamma_{2t} t \ln p_2 + \gamma_{3t} t \ln p_3 + \gamma_{Yt} t \ln Y + \gamma_{tt} t \end{aligned}$$

que substituí, com as vantagens já referidas, a função de produção com o mesmo nome e onde γ_{ij} , γ_{it} , α_i e α_0 (onde $i, j=1$ é o capital, 2 é o trabalho e 3 são os *inputs* intermédios importados) representam os parâmetros.

A homogeneidade linear desta função impõe as restrições seguintes aos seus parâmetros

$$\begin{aligned}\alpha_1 + \alpha_2 + \alpha_3 &= 1 \\ \gamma_{11} + \gamma_{12} + \gamma_{13} &= 0 \\ \gamma_{21} + \gamma_{22} + \gamma_{23} &= 0 \\ \gamma_{31} + \gamma_{32} + \gamma_{33} &= 0 \\ \gamma_{1t} + \gamma_{2t} + \gamma_{3t} &= 0\end{aligned}$$

e a homoteticidade da produção impõe ainda que

$$\gamma_{1t} = 0, \quad \gamma_{2t} = 0, \quad \gamma_{3t} = 0.$$

As equações das quotas dos factores obtêm-se aplicando derivadas logarítmicas à expressão de c ; usando o Lema de Shephard, admitindo que o *output* e os preços dos factores são fixos vem

- para o factor 1:

$$\begin{aligned}\frac{\partial \ln c}{\partial \ln p_1} &= q_1 \equiv \frac{p_1 X_1}{cY} \\ &= \alpha_1 + \gamma_{11} \ln p_1 + \gamma_{12} \ln p_2 + \gamma_{13} \ln p_3 + \gamma_{1Y} \ln Y + \gamma_{1t} t\end{aligned}$$

- para o factor 2

$$\begin{aligned}\frac{\partial \ln c}{\partial \ln p_2} &= q_2 \equiv \frac{p_2 X_2}{cY} \\ &= \alpha_2 + \gamma_{21} \ln p_1 + \gamma_{22} \ln p_2 + \gamma_{23} \ln p_3 + \gamma_{2Y} \ln Y + \gamma_{2t} t\end{aligned}$$

- para o factor 3

$$\begin{aligned}\frac{\partial \ln c}{\partial \ln p_3} &= q_3 \equiv \frac{p_3 X_3}{cY} \\ &= \alpha_3 + \gamma_{31} \ln p_1 + \gamma_{32} \ln p_2 + \gamma_{33} \ln p_3 + \gamma_{3Y} \ln Y + \gamma_{3t} t\end{aligned}$$

Como referem E. Berndt e D Wood uma interpretação útil dos parâmetros da translog γ_{ij} está relacionada com a definição de "share elasticity". Como já se viu a primeira derivada (parcial) de $\ln c$ em relação a $\ln p_i$ é a equação da quota ("share") de custo do i -ésimo factor. Por sua vez a derivada da equação da quota de custo em relação a $\ln p_j$ é igual a γ_{ij} o que quer dizer que os parâmetros da função translog

$$\gamma_{ij} = \frac{\partial^2 \ln c}{\partial \ln p_i \partial \ln p_j}$$

nos dão a resposta da i -ésima quota de custo a variações de $\ln p_j$.

As elasticidades parciais de substituição de Allen entre os *inputs* i e j ($i, j = 1, 2, 3$) para a função translog são:

$$\sigma_{ii} = \frac{\gamma_{ii} + q_i^2 - q_i}{q_i^2} \quad i = 1, 2, 3$$

$$\sigma_{ij} = \frac{\gamma_{ij} + q_i q_j}{q_i q_j} \quad i, j = 1, 2, 3 \quad i \neq j$$

Por sua vez a elasticidade procura do *input* i (X_i) em relação ao preço do *input* j definidas como

$$\varepsilon_{ij} = \frac{\partial \ln X_i}{\partial \ln p_j}$$

vem, no caso da função translog, dada por

$$\varepsilon_{ij} = q_j \cdot \sigma_{ij} \quad i, j = 1, 2, 3$$

De notar que quando $\gamma_{ij}=0$ para todo o i, j então a translog converte-se na função Cobb-Douglas, caso em que $\sigma_{ij}=1$ para todo o i, j com $i \neq j$.

Os valores das quotas de custo do capital, do trabalho e dos *inputs* intermédios importados, relativos ao primeiro período, que vão servir de base aos nossos cálculos são²¹ os constantes do anexo 1, enquanto os valores das quotas de custo do capital, do trabalho e dos *inputs* energéticos importados, relativos ao segundo período, se podem ver no anexo 2.

Também os valores dos índices de preços relativos ao primeiro período e *inputs* se podem ver no anexo 1 enquanto os relativos ao segundo período e *inputs* se podem encontrar no anexo 2.

Antes de levarmos a cabo as respectivas estimações convém ter em atenção que a estimação dos parâmetros não é uma questão fácil dado o facto de ser necessário respeitar todas as restrições já antes enunciadas e de as perturbações das equações estarem interrelacionadas. Ora atendendo a essas restrições o sistema de 3 equações que acabamos de escrever converte-se, por integração de algumas dessas restrições, em:

- para o factor capital (factor 1):

$$\hat{q}_1 = \alpha_1 + \gamma_{11} \ln\left(\frac{p_1}{p_3}\right) + \gamma_{12} \ln\left(\frac{p_2}{p_3}\right) + \gamma_{1t} t$$

²¹ Séries referidas por M. J. Vilares e extraídas, para o capital e o trabalho, de Cartaxo e Rosa (com base nas Contas Nacionais do INE), e para os *inputs* intermédios importados dos Anuários do Comércio Externo também do INE.

- para o factor trabalho (factor 2):

$$q_2 = \alpha_2 + \gamma_{21} \ln\left(\frac{p_1}{p_3}\right) + \gamma_{22} \ln\left(\frac{p_1}{p_3}\right) + \gamma_{2t} t$$

- para os *inputs* intermédios importados estimam-se os seus parâmetros indirectamente recorrendo ao sistema de equações seguinte:

$$\alpha_3 = 1 - \alpha_1 - \alpha_2,$$

$$\gamma_{31} = -(\gamma_{11} + \gamma_{21})$$

$$\gamma_{32} = -(\gamma_{12} + \gamma_{22})$$

$$\gamma_{33} = \gamma_{11} + \gamma_{22} + 2\gamma_{12}$$

$$\gamma_{3t} = -(\gamma_{1t} + \gamma_{2t})$$

O efeito do progresso tecnológico na função custo unitária pode analisar-se derivando a expressão $\ln c$ em relação à variável tempo (t):

$$\frac{\partial \ln c}{\partial t} = \dot{c} = \gamma_i + \gamma_{1t} \ln p_1 + \gamma_{2t} \ln p_2 + \gamma_{3t} \ln p_3 + \gamma_{Yt} \ln Y + \gamma_{it} t$$

Costuma dizer-se que esta equação quantifica a taxa de variação (diminuição) do custo.

Por sua vez o efeito do progresso técnico sobre a procura dos factores é dado pela evolução da taxa de crescimento da quota de custo do respectivo factor, ou seja,

- para o factor 1 (capital) por

$$b_1 \equiv \frac{\partial \ln q_1}{\partial t} = \frac{\partial \ln q_1}{q_1}$$

- para o factor 2 (trabalho) por

$$b_2 \equiv \frac{\partial \ln q_2}{\partial t} = \frac{\partial \ln q_2}{q_2}$$

- para o factor 3 (*inputs* intermédios importados) por

$$b_3 \equiv \frac{\partial \ln q_3}{\partial t} = \frac{\partial \ln q_3}{q_3}$$

e como estas taxas se podem escrever como

- para o factor capital

$$b_1 \equiv \frac{\partial q_1}{\partial t} \cdot \frac{1}{q_1}$$

- para o factor trabalho

$$b_2 \equiv \frac{\partial q_2}{\partial t} \cdot \frac{1}{q_2}$$

- e para o factor *inputs* intermédios importados

$$b_3 \equiv \frac{\partial q_3}{\partial t} \cdot \frac{1}{q_3}$$

e derivando as expressões das três quotas de custo em ordem a t vem

$$\frac{\partial q_1}{\partial t} = \gamma_{1t}$$

então vem

- para o factor capital

$$b_1 \equiv \gamma_{1t} \cdot \frac{1}{q_1} = \frac{\gamma_{1t}}{q_1}$$

- para o factor trabalho

$$b_2 \equiv \gamma_{2t} \cdot \frac{1}{q_2} = \frac{\gamma_{2t}}{q_2}$$

- para o factor *inputs* intermédios importados

$$b_3 \equiv \gamma_{3t} \cdot \frac{1}{q_3} = \frac{\gamma_{3t}}{q_3}$$

As variâncias dos parâmetros obtêm-se recorrendo às seguintes expressões

$$\text{var}(\alpha_3) = \text{var}(\alpha_1) + \text{var}(\alpha_2) + 2 \text{cov ar}(\alpha_1, \alpha_2)$$

$$\text{var}(\gamma_{31}) = \text{var}(\gamma_{11}) + \text{var}(\gamma_{21}) + 2 \text{cov ar}(\gamma_{11}, \gamma_{21})$$

$$\text{var}(\gamma_{32}) = \text{var}(\gamma_{12}) + \text{var}(\gamma_{22}) + 2 \text{cov ar}(\gamma_{12}, \gamma_{22})$$

$$\text{var}(\gamma_{33}) = \text{var}(\gamma_{11}) + \text{var}(\gamma_{22}) + 4 \text{var}(\gamma_{12}) + 2 \text{cov ar}(\gamma_{11}, \gamma_{22}) +$$

$$+ 4 \text{cov ar}(\gamma_{11}, \gamma_{12}) + 4 \text{cov ar}(\gamma_{12}, \gamma_{22})$$

$$\text{var}(c_3) = \text{var}(c_1) + \text{var}(c_2) + 2 \text{cov ar}(c_1, c_2)$$

3.2 OS RESULTADOS EMPÍRICOS PROPRIAMENTE DITOS

3.2.1 PRIMEIRO PERÍODO

Levada a cabo a respectiva estimação recorrendo ao método de Zellner para estimação de parâmetros de sistemas de equações aparentemente não relacionadas (SURE) vem, para o primeiro período:

- Para o factor capital (factor 1):

$$\hat{q}_1 = 0,348638995 - 0,16819021 \ln\left(\frac{p_1}{p_3}\right) + 0,193835607 \ln\left(\frac{p_2}{p_3}\right) - 0,019859591t$$

- Para o factor trabalho (factor 2):

$$q_2 = \alpha_2 + \gamma_{21} \ln\left(\frac{p_1}{p_3}\right) + \gamma_{22} \ln\left(\frac{p_2}{p_3}\right) + \gamma_{2t}t$$

e os resultados encontrados foram

$$\hat{q}_2 = 0,488993008 + 0,168497096 \ln\left(\frac{p_1}{p_3}\right) - 0,16819021 \ln\left(\frac{p_2}{p_3}\right) + 0,005905153t$$

- Para obter estimativas dos parâmetros da equação dos *inputs* intermédios importados (factor 3), procedeu-se como já antes se referiu chegando-se aos seguintes valores:

$$\begin{aligned} \hat{\alpha}_3 &= 1 - \hat{\alpha}_1 - \hat{\alpha}_2 = \\ &= 1 - 0,348638995 - 0,488993008 = 0,162367997 \\ \hat{\gamma}_{31} &= -(\hat{\gamma}_{11} + \hat{\gamma}_{21}) = \\ &= -(0,193835607 - 0,168190211) = -0,025645396 \\ \hat{\gamma}_{32} &= -(\hat{\gamma}_{12} + \hat{\gamma}_{22}) = \\ &= 0,168190211 - 0,168497096 = -0,000306885 \\ \hat{\gamma}_{33} &= \hat{\gamma}_{11} + \hat{\gamma}_{22} + 2\hat{\gamma}_{12} = \\ &= 0,193835607 + 0,168497096 - 2 \times 0,168190211 = \\ &= -0,275439608 \\ \hat{\gamma}_{3t} &= -(\hat{\gamma}_{1t} + \hat{\gamma}_{2t}) = -(-0,019859591 + 0,005905153) = \\ &= -0,013954438 \end{aligned}$$

O seguinte quadro dá-nos os valores das estimativas dos parâmetros já antes referidos, os desvios padrões de cada um dos parâmetros e os valores que a estatística t

toma na hipótese nula em que cada um dos parâmetros isoladamente considerado é nulo:

parâmetros estimados	estimativa	desvio padrão	valor da t
α_1	0,348638995	0,008769763	39,75467
c_1	-0,019859591	0,003406871	-5,82928
γ_{11}	0,193835607	0,026137343	7,41604
γ_{12}	-0,168190211	0,025061567	-6,71108
α_2	0,0488993008	0,010974698	44,55640
c_2	0,005905153	0,003817388	1,54691*
γ_{22}	0,168497096	0,033656182	5,00642

Nota: o * identifica valores não significativos ao nível de significância de 5%.

De notar que, de entre os coeficientes estimados directamente, apenas o coeficiente obtido para a variável c_2 é não significativo aos níveis de 5% e 1%.

As expressões das elasticidades parciais de substituição de Allen proporcionaram os seguintes valores:

- Elasticidades médias de substituição próprias:

$$\sigma_{11} = -0,24404535, \sigma_{22} = -0,3893771, \sigma_{33} = -9,01379718$$

- Elasticidades médias de substituição cruzadas:

$$\sigma_{12} = -0,14404219, \sigma_{13} = 0,6477189, \sigma_{23} = 0,99694784.$$

O primeiro grupo de valores mostra uma sensibilidade particular da elasticidade de substituição própria relativa aos *inputs* intermédios importados com um valor absoluto de 9, situação que já não ocorre com tanta intensidade nem com o trabalho nem com o capital com 0,24 e 0,39, respectivamente.

O segundo grupo de valores garante-nos que na economia portuguesa, e durante o período 1980-89, os factores capital e trabalho são complementares enquanto o capital e os *inputs* intermédios importados, por um lado, e o trabalho e os *inputs* intermédios importados, por outro, são substituíveis, que os módulos dessas elasticidades são inferiores à unidade, situação que já ocorreu nos períodos 1953-65 e 1960-70, e que a que tem maior valor absoluto é a elasticidade de substituição entre o *input* capital e os *inputs* intermédios importados.

As expressões das elasticidades procura-preço permitem-nos encontrar os seguintes valores:

- Elasticidades (médias) procura-preço próprias:

$$\varepsilon_{11} = -0,10997001, \varepsilon_{22} = -0,12703558, \varepsilon_{33} = -2,01128162;$$

- Elasticidades (médias) procura-preço cruzadas:

$$\varepsilon_{12}=-0,04699424, \varepsilon_{13}=0,14452789, \varepsilon_{23}=0,22245263,$$

$$\varepsilon_{21}=-0,06490728, \varepsilon_{32}=0,32525756 \text{ e } \varepsilon_{31}=0,29187055.$$

Os resultados das elasticidades procura-preço próprias demonstram que a elevações dos preços dos *inputs* correspondem reduções das suas procuras e em particular no caso dos *inputs* intermédios importados, o que traduz, neste caso, um mercado bastante “price responsive”, confirmando o que já se disse a propósito da elasticidade de substituição própria deste mesmo factor; no caso das elasticidades procura-preço cruzada a elevações dos preços de alguns *inputs* correspondem quer reduções da procura do *input* cruzado - é o caso quando se eleva o preço do capital (trabalho) que se reduz a procura de mão-de-obra (capital), o que se explicará pela complementaridade trabalho-capital - quer aumentos da procura do *input* cruzado - é o caso quando se eleva o preço dos *inputs* intermédios importados, do capital e do trabalho cujos reflexos sobre a procura de trabalho, do capital e dos *inputs* intermédios, respectivamente, são de aumento; o reduzido valor das elasticidades procura-preço cruzadas permitem-nos afirmar que em geral a nossa economia não é muito “price-responsive” dependendo essa capacidade de resposta do binómio de *inputs* em causa.

A qualidade das regressões estimadas pelo método SURE é bastante boa porquanto os coeficientes de determinação da primeira e da segunda equações são, respectivamente, 89% e o 94% (coeficientes de determinação corrigidos de 82% e de 91%).

Analisada a questão da autocorrelação para ambas as equações obteve-se o valor de $d1=1,95281$ para a primeira equação e o valor $d2=2,376909$ para a segunda o que, comparado com os valores da tabela de Durbin-Watson com $k=3$ (variáveis explicativas por equação) e nível de significância de 5% - $d_L=0,455$ e $d_U=2,128$ -, nada nos permite concluir quanto a essa questão uma vez que em qualquer dos casos caem em regiões inconclusivas.

Para analisar o efeito do progresso técnico sobre a procura dos factores, efeito esse dado pela evolução da taxa de crescimento da quota de custo do respectivo factor, vamos substituir, nas expressões respectivas, γ_{it} pelas suas estimativas e concluir em conformidade. Assim:

- Para o factor capital

$$\hat{b}_1 = -\frac{0,019859591}{q_1} < 0$$

Esta expressão, sendo sempre negativa uma vez que a quota de custo do capital é sempre positiva e o numerador é negativo para o período em análise denota um progresso técnico *poupador* do *input* capital, situação que se deverá explicar pelas elevadas taxas de juro em vigor no período.

- Para o factor trabalho

$$\hat{b}_2 = \frac{0,005905153}{q_2} > 0$$

Esta expressão, que é sempre positiva uma vez que a quota de custo do trabalho e o numerador também o são denota um progresso técnico utilizador ou gastador do *input* trabalho, conclusão que representa uma inversão em relação ao que acontecia e nos relata A. de Sousa nos anos 1953-1970. A justificação deste facto poderá ficar a dever-se por um lado à introdução dos contratos a prazo no mercado laboral português o que permitiu um maior recurso a este factor, por outro às já referidas taxas de juro (que tornavam difícil o recurso a novos investimentos mais capital-intensivos) e ainda à carência de divisas (necessárias para pagar as nossas importações quer de equipamentos quer de *inputs* intermédios importados) que na altura muito nos afligia.

- Para o factor identificado como *inputs* intermédios importados vem

$$\hat{b}_3 = -\frac{0,013954438}{q_1} < 0$$

Esta expressão, sendo sempre negativa uma vez que a quota de custo dos *inputs* intermédios é sempre positiva e o numerador é negativo para o período em análise denota um progresso técnico *mais poupador* dos *inputs* intermédios importados, situação a que não será certamente estranho o facto de durante o lapso de tempo em causa as nossas reservas em ouro e divisas (com que as importações necessitam ser pagas) serem escassas, ao encerramento de algumas unidades industriais que assim deixaram de ser consumidoras, à crise energética de finais dos anos setenta princípio dos anos oitenta e também a alguma inactividade do sector produtivo.

3.2.2 SEGUNDO PERÍODO

Fazendo exactamente a mesma coisa para o segundo período em análise e para os factores trabalho, capital e *inputs* energéticos importados vem:

- Para o factor capital:

$$\hat{q}_1 = 30,51880125 - 0,15633682 \ln\left(\frac{p_1}{p_3}\right) + 0,15685240 \ln\left(\frac{p_2}{p_3}\right) - 0,01523002t$$

- Para o factor trabalho:

$$\hat{q}_2 = -48,51777765 + 0,17850074 \ln\left(\frac{p_1}{p_3}\right) - 0,15633682 \ln\left(\frac{p_2}{p_3}\right) + 0,02478316t$$

- Para os *inputs* energéticos importados:

$$\hat{\alpha}_3 = (1 - \hat{\alpha}_1 - \hat{\alpha}_2) = 18,998964$$

$$\hat{\gamma}_{31} = -(\hat{\gamma}_{11} + \hat{\gamma}_{21}) = -0,00051558$$

$$\hat{\gamma}_{32} = -(\hat{\gamma}_{12} + \hat{\gamma}_{22}) = -0,02216392$$

$$\hat{\gamma}_{33} = \hat{\gamma}_{11} + \hat{\gamma}_{22} + 2\hat{\gamma}_{12} = 0,0226795$$

$$\hat{\gamma}_{3t} = -(\hat{\gamma}_{1t} + \hat{\gamma}_{2t}) = -0,00955314.$$

O seguinte quadro dá-nos os valores das estimativas dos parâmetros já antes referidos, os desvios padrões de cada um dos parâmetros e os valores que a estatística *t* toma na hipótese nula em que cada um dos parâmetros isoladamente considerado é nulo:

parâmetros	Estimativa	desvio padrão	valor da <i>t</i>
α_1	30,51880125	43,22769035	0,70600*
c_1	-0,01523002	0,02183182	-0,69761*
γ_{11}	0,15685240	0,16027341	0,97866*
γ_{12}	-0,15633682	0,11599825	-1,34775*
α_2	-48,51777765	32,10322152	-1,51131*
c_2	0,02478316	0,01621394	1,52851*
γ_{22}	0,17850074	0,08692263	2,05356

Nota: o * identifica valores não significativos aos níveis de significância de 5% e 1%.

De notar que, de entre os coeficientes estimados directamente, praticamente todos eles serem não significativos aos níveis de significância habituais (o que retira algum significado às conclusões extraídas a partir dos valores encontrados) se bem que as significâncias globais das regressões iniciais sejam altamente significativas.

As expressões das elasticidades parciais de substituição de Allen proporcionaram os seguintes valores (médios) estimados para este período:

- Elasticidades de substituição próprias:

$$\sigma_{11} = -0,47292827, \sigma_{22} = -0,28108527, \sigma_{33} = -9,1763115$$

- Elasticidades médias de substituição cruzadas:

$$\sigma_{12} = 0,28396422, \sigma_{13} = 0,98102351, \sigma_{23} = 0,29960835.$$

O primeiro grupo de valores mostra uma sensibilidade particular da elasticidade de substituição própria dos *inputs* energéticos importados com um valor absoluto de 9,18, situação que já não ocorre com tanta intensidade com o trabalho e do capital com 0,47 e 0,28, respectivamente.

- O segundo grupo de valores garante-nos que na economia portuguesa, e durante o período 1980-89, os três factores - capital, trabalho e energia - são substituíveis, embora em pequeno grau no caso capital-trabalho e trabalho-

energia (elasticidades entre 0,28 e 0,30), e em grau ligeiramente superior no binómio capital-energia (0,98). Este último valor corrobora as conclusões levadas a cabo por alguns autores segundo os quais estes dois factores são substituíveis na maior parte dos países da Europa; esta conclusão é contrária à por nós próprios extraída em estudo anterior e com exactamente os mesmos dados, dando assim razão a outros autores que afirmam que os valores encontrados para esta elasticidade vem influenciado pelo número de variáveis explicativas que se consideram (neste caso considerou-se a mais a variável t)).

As expressões das elasticidades (médias) procura-preço permitem-nos encontrar os seguintes valores:

- Elasticidades procura-preço próprias:

$$\varepsilon_{11}=-0,20476035, \varepsilon_{22}=-0,14174706, \varepsilon_{33}=-0,57583465;$$

- Elasticidades procura-preço cruzadas:

$$\varepsilon_{12}=0,14319887, \varepsilon_{13}=0,06156148, \varepsilon_{23}=0,01880111,$$

$$\varepsilon_{21}=0,12294594, \varepsilon_{31}=0,15108797 \text{ e } \varepsilon_{32}=0,42474669.$$

Os resultados das elasticidades procura-preço próprias demonstram, como era de esperar, que a elevações (reduções) dos preços dos *inputs* correspondem reduções (elevações) das suas próprias procuras praticamente equivalentes para o trabalho e o capital seguidas de perto pela energia importada; os valores encontrados, próximos de zero, demonstram que a economia portuguesa deste período é pouco "price responsive" no que a esta questão diz respeito.

Os valores das elasticidades procura-preço cruzadas encontrados permitem-nos concluir que:

- a elevações dos preços do factor capital correspondem
 - acréscimos da procura do factor trabalho e
 - acréscimos do factor energia importada;
- a elevações dos preços do factor trabalho correspondem
 - acréscimos da procura do factor capital; e
 - acréscimos da procura do factor energia importada;
- a elevações dos preços do factor energia importada correspondem
 - acréscimos da procura do factor trabalho; e
 - acréscimos da procura do factor capital.

Os reduzidos valores encontrados para estas elasticidades exprimem, mais uma vez, a pouca resposta do nosso mercado a variações dos preços dos factores concorrentes.

A qualidade das regressões estimadas pelo método SURE não é tão boa como na obtida com os dados do primeiro período, se bem que, como já firmamos

anteriormente, sejam significativas em termos estatísticos; os coeficientes de determinação das duas equações vêm iguais a 85,99% e 71,56%.

Analisando a questão da autocorrelação para ambas as equações obteve-se o valor de $d_1=1,43$ para a equação da quota de custo do capital e o valor $d_2=1,63$ para a quota de custo do trabalho, valores que, comparados com os valores críticos fornecidos pela tabela de Durbin-Watson com $k=3$ (variáveis explicativas por equação) e nível de significância de 5% - $d_L=0,525$ e $d_U=2,016$ -, nada nos permitem concluir, quer num caso quer no outro, quanto a essa questão uma vez que em ambos os casos caem em regiões inconclusivas.

Para apreciar as características do progresso técnico ao longo desta década vamos substituir os valores de γ_{it} pelas suas estimativas para obter as estimativas de b_i . Assim:

- Para o factor capital: introduzindo na expressão respectiva o valor estimado do parâmetro γ_{1t} vem

$$\hat{b}_1 = -\frac{0,01523002}{q_1} < 0$$

resultado que traduz, para o período em análise, um progresso técnico relativamente *poupador* do *input* capital. Este facto tem que ver com as ainda elevadas taxas de juro, com ainda relativamente baixo (face às taxas de juro) do factor mão-de-obra, e ainda com algumas dificuldades em termos de disponibilidades em ouro e divisas para custear as nossas importações, dificuldades que já não se verificaram na fase final do período, altura em que já se passou a dispor das ajudas de pré e de adesão à União Europeia.

- Para o factor trabalho: introduzindo nesta expressão o valor estimado do parâmetro γ_{2t} vem

$$\hat{b}_2 = \frac{0,02478316}{q_2} > 0$$

resultado que denota um progresso técnico relativamente *utilizador* do *input* trabalho. A razão de ser deste facto poderá ter que ver com a entrada em vigor da lei dos contratos a prazo, lei que trouxe menor rigidez da legislação ao mercado do trabalho e com os custos horários do trabalho relativamente baixos face aos custos do capital (taxa de juro).

- Para o factor *inputs* intermédios importados: introduzindo nesta expressão o valor estimado do parâmetro γ_{3t} vem

$$\hat{b}_3 = -\frac{0,00955314}{q_3} < 0$$

Esta expressão, negativa, demonstra que o progresso técnico tem sido *poupador* dos *inputs* energéticos importados, situação a que não será certamente estranho o facto de durante o lapso de tempo em causa serem conseguidos sobretudo com uma maior utilização de mão-de-obra e uma melhor utilização quer do equipamento disponível

quer destes mesmos *inputs* energéticos; este período caracterizava-se também por durante uma parte razoável do seu tempo o país ter falta de ouro e divisas com que custear as suas importações quer de capital quer de *inputs* energéticos importados.

Vê-se assim, comparando as conclusões que agora extraímos com as expressas a propósito do primeiro período, que elas no geral são coincidentes, o que não admira face às elevadas taxas de juro activas de ambos os períodos, ao encerramento de diversas unidades industriais que não souberam adaptar-se às novas condições de mercado decorrentes do 25 de Abril - nomeadamente, perda dos mercados das ex-colónias e subida generalizada, logo em seguida, dos salários dos trabalhadores, à rigidez da estrutura produtiva, à escassez de ouro e divisas para pagar as importações, situação que tendeu a atenuar-se na fase final do segundo período com as transferências de verbas da ex-Comunidade Económica Europeia.

De referir ainda que estas conclusões têm que ser tiradas com algum cuidado porquanto as variáveis usadas nos dois períodos não são cem por cento coincidentes, como acontece com o terceiro factor considerado que foi os *inputs* intermédios importados, no primeiro período, e os *inputs* energéticos importados, no segundo; também os índices de preços usados, particularmente no caso deste terceiro factor, mas também no caso do capital, foram calculados de maneira diferente nos dois períodos, o que poderá provocar coeficientes tecnológicos e interpretações diferentes.

REFERÊNCIAS BIBLIOGRÁFICAS

- BERNDT, ERNST R. e WOOD, DAVID O. (1982) - "The Specification and Measurement of Technical Change in the U. S. Manufacturing" in Moroney (edt) "Advances in the Economics of Energy and Resources, Greenwich, Connecticut, Jai Press.v4, pp 199-
- BERNDT, ERNST R. e WOOD, DAVID O. (1985) - "Concavity and the Specification of Technical Change in the U. S. Manufacturing" in Fericeli e Lesourd (eds) "Energie: Modélisation et Econométrie", Actes du Colloque d'Aix-en-Provence, Economica, Paris, pp. 444-471
- BERNDT, ERNST R. e TRIPLETT, JACK E. (eds) (1990) - "Fifty Years of Economic Measurement - The Jubilee of the Conference on Research in Income and Wealth", National Bureau of Economic Research, v. 64
- CHRISTENSEN, LAURITS, R. DALE, W. JORGENSON e LAWRENCE W. LAU (1973) - "Transcendental Logarithmic Production Frontiers", Review of Economics of Statistics, 55(1),February,pp. 28-45
- FAGAN, MARIE N. (1997) - "Resource Depletion and Technical Change on U. S. Crude Oil Findings Costs from 1977 to 1994", The Energy Journal,IAEE,v18.4,pp.91-106
- FORBES, KEVIN e ERNST M. ZAMPELLI (1995) - "Technological Progress, Technical Efficiency and the Search for Oil and Gas in the U. S. Onshore", Proceedings of 18th International Association for Energy Economics, International Conference, July 5, Washington D. C.
- JESUS, FERNANDO (1977) - "Econometria", tomo II, ed. IPC

- JORGENSON, DALE W. (edt) (1995) - "Productivity - Postwar U. S. Economic Growth", v1, MIT Press, MIT, Cambridge, Massachusetts
- JORGENSON, DALE W. (edt) (1995) - "Productivity - International Comparisons of Economic Growth", v2, MIT Press, MIT, Cambridge, Massachusetts
- KREIJGER, R G. (1980) - "Imports as a Factor of Production: The Substitution Characteristics of Labor, Capital and Material Inputs", Univ. de Amsterdão, Report AÉ/70
- MAKOTA OHTA (1974) - "A Note on the Duality Between Production and Cost Function: Rate of Returns to Scale and Rate of Technical Progress", Economics Studies Quarterly, v25,12,pp. 63-65
- MANSO, J. R. PIRES (1991) - "A Substituição Entre Factores na Economia Portuguesa - Uma Aplicação da Função Translog Considerando a Energia um Factor de Produção Autónomo", T. M., ISEG, Lisboa
- MANSO, J. R. PIRES (1996) - "A Dinâmica Estrutural e a Substituição Interfactorial e Interenergética - Metodologia de Análise e Aplicação à Substituição entre Energias em Portugal", T.D., ISEG, Lisboa
- MORONEY, J. R. e J. M. TRAPANI (1981) - "Alternative Models of Substitution and Technical Change in Natural Resource Intensive Industries" in Ernst R. Berndt e Barry C. Field "Modeling and Measuring Natural Resource Substitution ", Cambridge, Massachusetts, MIT Press
- NORSWORTHY e JANG, S. L. (1992) - "Empirical Measurement and Analysis of Productivity and Technological Change - Applications in High-Technology and Services Industries", Contribution to Economic Analysis, North Holland
- SOUSA, ALFREDO DE (1970) - "As Funções de Produção Cobb-Douglas na Indústria Transformadora Portuguesa", Centro de Estudos de Planeamento, n° 2, Lisboa
- SOUSA, ALFREDO DE (1977) - "As Funções de Produção CES na Indústria Transformadora Portuguesa", Economia, v.1, 1, 01/77, Lisboa
- VILARES, M. JOSÉ (1989) - "Os Bens Intermediários Importados Como Factor de Produção - Aplicação à Economia Portuguesa", Sousa *et alii* "Nova Economia Portuguesa - Estudos em Homenagem a António Manuel Pinto Barbosa", Univ. Nova de Lisboa, pp. 499-539

ANEXOQUOTAS DE CUSTO DOS FACTORES DE PRODUÇÃO E ÍNDICES DE PREÇOS
(1976-1984)

Anos	quota capital	quota trabalho	quota inputs intermédios importados	índ.pr. capital	índ.pr. trabalho	índ.preços inputs intermédios importados
1976	0,273307	0,554730	0,171961	64,2435	85,8873	74,0036
1977	0,317957	0,492760	0,189283	100,000	100,000	100,000
1978	0,356069	0,462458	0,181473	157,294	116,566	107,449
1979	0,353535	0,434726	0,211739	216,998	135,244	153,163
1980	0,317730	0,438073	0,244196	251,987	168,150	210,296
1981	0,305590	0,443915	0,250495	315,054	204,627	258,684
1982	0,321413	0,431689	0,246898	425,541	241,327	296,509
1983	0,332693	0,419849	0,247458	662,203	284,336	394,643
1984	0,357923	0,377317	0,264759	854,416	335,765	521,150

QUOTAS DE CUSTO DOS FACTORES DE PRODUÇÃO E ÍNDICES DE PREÇOS
(PERÍODO DE 1980-1989)

Anos	quota capital	quota trabalho	quota inputs energéticos importados	i.p. do capital	i.p. do trabalho	i.p. dos inputs energéticos importados
1980	0,386547	0,532913	0,080540	100,0000	100,0000	100,0000
1981	0,369553	0,545887	0,084760	122,0820	100,7310	133,2267
1982	0,385019	0,532227	0,082754	149,2113	96,0522	193,2125
1983	0,401897	0,519524	0,078579	187,2239	92,3086	249,9644
1984	0,462330	0,459298	0,078372	222,9758	86,9520	338,2018
1985	0,475749	0,452005	0,072246	266,9295	89,7942	406,4321
1986	0,458421	0,498764	0,042815	306,1514	95,7109	364,2857
1987	0,456950	0,504035	0,039015	340,4837	90,9375	296,4463
1988	0,460067	0,510450	0,029483	380,6519	90,5203	297,2636
1989	0,473095	0,487746	0,039159	418,7171	88,9985	342,6794

AGRUPAMENTOS DE CONCELHOS DE PORTUGAL CONTINENTAL E SUA CARACTERIZAÇÃO

Autores:
Ana Brandão,
Ana Pires
e
Jorge Portugal

VOLUME 1

1º QUADRIMESTRE DE 1998

AGRUPAMENTOS DE CONCELHOS DE PORTUGAL CONTINENTAL E SUA CARACTERIZAÇÃO

GROUPS OF CONTINENTAL PORTUGAL COUNCILS AND ITS PROFILES

Autores: Ana Brandão

- Investigadora no Gabinete de Estudos e Planeamento, Instituto Superior Técnico

Ana Pires

- Professora Auxiliar do Departamento de Matemática e Investigadora do Centro de Matemática Aplicada, Instituto Superior Técnico

Jorge Portugal

- Director do Departamento de Estudos de Mercado - SONAE

RESUMO:

- Este trabalho consiste no agrupamento dos concelhos de Portugal Continental por forma a caracterizar o país em termos sociais, económicos e políticos.

Depois de sucessivas etapas aperfeiçoando-se a forma como se introduz a informação no processo de construção dos *clusters* e até o método de agrupamento, utilizam-se critérios numéricos para escolha da melhor solução.

Esta é representada em mapa para melhor percepção dos resultados e interpretada através da caracterização de cada cluster por um perfil baseado nas variáveis escolhidas para efectuar a segmentação. Este permite a identificação de um determinado tipo de pessoas ou famílias que habitam a região.

PALAVRAS-CHAVE:

- *Análise de Clusters, Método das Componentes Principais, critérios de selecção de melhor agrupamento.*

ABSTRACT:

- The purpose of this work was to join together Continental Portugal councils in order to describe the country in social, economical and political aspects.

After several stages in which the way of introducing information into the process of Cluster Analysis was improved, as well as the gathering method itself, numerical criteria for deciding which was the best solution were applied.

This solution was represented on a map in order to have a better perspective of the results and interpreted throughout a profile defined for each cluster and based upon the variables chosen to take part in the segmentation process. It allows the identification of a certain type of persons or families living in the area within the cluster.

KEY WORDS:

- *Cluster Analysis, Principal Components Analysis, criteria on how to select the best cluster.*

1. INTRODUÇÃO

O objectivo do trabalho aqui descrito (Brandão, 1997) foi o de proceder a uma segmentação do nosso país em grupos de unidades mais ou menos homogéneas usando informação relativa à totalidade da população e não apenas a resultante de processos de amostragem. A principal fonte deste tipo de informação é sem dúvida os CENSOS, sendo os seus últimos resultados referentes a 1991.

Trabalhos deste tipo são conhecidos em relação a outros países como sejam os Estados Unidos e a Inglaterra (Ozimek, 1993 e Openshaw, 1995), os quais foram já alvo de estudos aprofundados, mas tanto quanto se sabe não existe nenhuma análise semelhante de carácter global em relação a Portugal, embora tenha sido feito um estudo do género para a Região Norte (Gomes, Bacelar e Saleiro, 1994)

A primeira decisão a tomar ao partir para um trabalho deste tipo refere-se à escolha da unidade de segmentação a utilizar, a qual está condicionada pela mais pequena unidade territorial em relação à qual são divulgados os dados dos CENSOS. Devido a algumas limitações descritas adiante, escolheu-se o concelho como unidade.

Muito interligada à primeira decisão está a que se lhe segue, que compreende a escolha das variáveis características da população que deverão intervir no processo de segmentação e em relação às quais se refere a homogeneidade dos grupos de unidades, ou seja, de concelhos. Essas variáveis devem reflectir um objectivo preestabelecido

Para que este estudo fosse o mais abrangente possível, pretendeu-se caracterizar o país em termos sociais, económicos e políticos, pois as aplicações do mapa resultante desta segmentação são inúmeras, directas ou indirectas. Mediante este objectivo, escolheram-se as variáveis à partida julgadas como bons indicadores daqueles assuntos. Uma descrição completa das variáveis a utilizar encontra-se na secção seguinte.

Os grupos assim obtidos denominados de *clusters*, são passíveis de serem descritos através de um perfil, que, baseado nas variáveis escolhidas a intervir no processo, se distingue dos restantes e permite a identificação, em média, de um determinado tipo de pessoas ou famílias que habitam a região.

Segue-se então uma breve descrição deste estudo feito a Portugal Continental, a apresentação da metodologia utilizada e os resultados obtidos em formato de tabelas a acompanhar o mapa da divisão em *clusters*.

Uma vez que a informação relativa à totalidade da população correspondente aos CENSOS é actualizada de dez em dez anos e vai adquirindo cada vez mais pormenor, este processo pode ser também ele actualizado e consequentemente aperfeiçoado, tanto ao nível técnico como de resultados.

2. METODOLOGIA

Openshaw (1995) fornece a base do que pode ser considerada a melhor “receita” para a classificação de regiões através do uso de informação proveniente dos CENSOS, uma vez que assenta em vários anos de experiência nesta área e que por consequência, toma em linha de conta os problemas que surgem na prática.

Trata-se do seguinte esquema de procedimento:

- 1º passo : Identificação do objectivo da classificação;*
- 2º passo : Escolha da unidade de segmentação dos CENSOS e das variáveis a utilizar;*
- 3º passo: Normalização, ortogonalização e factorização dos dados;*
- 4º passo: Classificação usando a Análise de Clusters;*
- 5º passo: Avaliação e interpretação da segmentação - se possível atribuição de um nome a cada Cluster;*
- 6º passo: Representação dos resultados no mapa.*

É claro que não se trata de um grupo de etapas que uma vez concluídas conduzem de imediato à melhor classificação, ou seja, não se está em presença de um processo directo. A escolha dos métodos estatísticos a utilizar, bem como a das variáveis a fazer intervir no processo, são decisões que apenas podem ser avaliadas aquando da observação dos resultados finais. Trata-se portanto de um processo por tentativas tomando-se, a partir das consequências nos resultados de cada uma delas, a decisão que levará à realização da tentativa seguinte.

2.1. IDENTIFICAÇÃO DO OBJECTIVO DA CLASSIFICAÇÃO

Já foi referido na introdução deste trabalho que o seu objectivo é a obtenção de uma segmentação do nosso país em regiões, devendo estas ser determinadas de modo a que cada território possa ser caracterizado por um perfil que, em "média", defina a população que lá reside.

Como se pretendeu um estudo de carácter abrangente, e também como já foi mencionado na introdução, o objectivo da descrição das regiões resultantes da segmentação recaiu nos temas social, económico e político de Portugal Continental, sendo as variáveis intervenientes escolhidas segundo este objectivo.

2.2. ESCOLHA DA UNIDADE DE SEGMENTAÇÃO DOS CENSOS E DAS VARIÁVEIS A UTILIZAR

A área a ser usada como unidade de segmentação está dependente da que é utilizada na divulgação dos dados dos CENSOS. O Instituto Nacional de Estatística (INE) agrega os dados recolhidos em vários níveis, sendo a unidade das partições mais finas que se publicam as freguesias e os concelhos.

Em relação às freguesias surge uma dificuldade relacionada com o facto de apenas um número muito restrito de variáveis se encontrarem assim segmentadas, não perfazendo uma quantidade aceitável para fazer uma análise com objectivos da natureza dos descritos atrás. Decidiu-se então usar como unidade o concelho.

Nesta fase e por uma questão de limitação de tempo optou-se por utilizar a informação dos CENSOS processada pela MARKTEST (Marktest, 1994) por se encontrar num formato mais conveniente do que o fornecido pelo INE. Infelizmente, e sem explicação aparente, esta empresa não inclui nestes dados os concelhos englobados nas Regiões Autónomas da Madeira e dos Açores. Por este motivo, estes não se encontram na análise realizada, fazendo parte desta apenas os 275 concelhos do Continente.

Qualquer que seja o propósito de uma classificação é certo que não existe um conjunto "correcto" e único de variáveis a usar: diferentes pessoas utilizarão, com certeza, diferentes variáveis com vista a um mesmo objectivo. Convém, no entanto, evitar um grande número de variáveis altamente correlacionadas e tentar arranjar um conjunto representativo de modo a satisfazer fortemente o objectivo da classificação. Como as variáveis provenientes dos CENSOS são em número elevado, foi necessário encontrar um equilíbrio pois um excesso de informação pode trazer dificuldades na interpretação e validação de resultados.

Tendo em atenção os objectivos gerais do estudo, referidos em 2.1, decidiu-se escolher variáveis relacionadas com os seguintes temas:

- **Densidade Populacional**, por ser factor preponderante em qualquer dos três assuntos mencionados;
- **Idade**, social e economicamente discriminativa por excelência;
- **Tipo de Família**, isto é, constituição do agregado familiar no que diz respeito à sua estrutura;
- **Nível de Ensino Atingido**, que indica o mais elevado grau de ensino, completo ou não, que um individuo possui após o *terminus* dos seus estudos;
- **Situação no Trabalho**, que traduz a posição do trabalhador em termos de hierarquia;
- **Desemprego**, também muito importante na avaliação de qualquer um dos três assuntos referidos;
- **Índice de Poder de Compra**, indicador resultante de um estudo (Gabinete de Estudos Regionais da Direcção Regional do Centro do INE, Ramos, 1993) que comporta em si o nível económico, por concelho, dos portugueses;
- **Resultados das Eleições Legislativas de 1995**, que correspondem à informação política mais recente, aquando da data deste estudo;

Após diversas tentativas e análise das respectivas consequências, utilizaram-se 38 variáveis para obter os resultados finais. Uma breve descrição destas encontra-se na Tabela 2.1.

Tabela 2.1: Descrição das variáveis

NOME	DESCRIÇÃO	TIPO	DATA	FONTE
POPAREA	Densidade Populacional	Quociente ²²	1991	INE
IPC	Índice de Poder de Compra	Factor ²³	1993	INE
TXDES_HM	Taxa de Desemprego	Rácio	1991	INE
VOTAÇÃO NAS LEGISLATIVAS DE 1995				
PS95r	Votação do PS	Rácio	1995	STAPE ²⁴
PSD95r	Votação do PSD	Rácio	1995	STAPE
CDS95r	Votação do CDS	Rácio	1995	STAPE
PCP95r	Votação do PCP	Rácio	1995	STAPE

²² Trata-se do quociente entre o total da população e a área a que se refere esse total

²³ (Ramos, 1993)

²⁴ STAPE - Secretariado Técnico dos Assuntos para o Processo Eleitoral - Ministério da Administração Interna

REGIME DE OCUPAÇÃO DE RESIDÊNCIA				
REG_PPr	Regime de Ocupação: Proprietário	Rácio	1991	INE
REG_APr	Regime de Ocupação: Arrendamento	Rácio	1991	INE
REG_OTr	Regime de Ocupação: Outros	Rácio	1991	INE
IDADE				
T0_14r	Indivíduos com <=14 anos	Rácio	1991	INE
T15_24r	Indivíduos entre 15 e 24 anos	Rácio	1991	INE
T25_34r	Indivíduos entre 25 e 34 anos	Rácio	1991	INE
T35_44r	Indivíduos entre 35 e 44 anos	Rácio	1991	INE
T45_54r	Indivíduos entre 45 e 54 anos	Rácio	1991	INE
T55_64r	Indivíduos entre 55 e 64 anos	Rácio	1991	INE
T65_74r	Indivíduos entre 65 e 74 anos	Rácio	1991	INE
T75r	Indivíduos com 75 anos ou mais	Rácio	1991	INE
NÍVEL DE ENSINO ATINGIDO				
ANALFr	Analfabetos com 10 e mais anos	Rácio	1991	INE
FREQr	População a frequentar o ensino	Rácio	1991	INE
E_BASr	Nível de ensino atingido: Básico	Rácio	1991	INE
E_SECr	Nível de ensino atingido: Secundário	Rácio	1991	INE
E_MEDSUPr	Nível de ensino atingido: Médio / Superior	Rácio	1991	INE
SITUAÇÃO NO TRABALHO				
PATRAOr	Patrão	Rácio	1991	INE
TCPPr	Trabalhador por conta Própria	Rácio	1991	INE
TFNPr	Trabalhador Familiar não remunerado	Rácio	1991	INE
TCOr	Trabalhador por conta de outrem	Rácio	1991	INE
COOPERr	Membro de Cooperativa	Rácio	1991	INE
EMPOUTr	Trabalhador em outra situação	Rácio	1991	INE

ESTRUTURA FAMILIAR				
AD1SCRr	1 Adulto Sem Crianças	Rácio	1991	INE
AD1C1MCRr	1 Adulto com 1 ou Mais Crianças	Rácio	1991	INE
AD2SCRr	2 Adultos Sem Crianças	Rácio	1991	INE
AD2C1CRr	2 Adultos com 1 Criança	Rácio	1991	INE
AD2C2MCRr	2 Adultos com 2 ou Mais Crianças	Rácio	1991	INE
AD3MSCRr	3 Adultos ou Mais Adultos Sem Crianças	Rácio	1991	INE
AD3MC1CRr	3 Adultos ou Mais Adultos Com 1 Criança	Rácio	1991	INE
AD3MC2MC Rr	3 Adultos ou mais Adultos Com 2 ou Mais Crianças	Rácio	1991	INE
ESTFAMOU Tr	Estruturas Familiares com Outros Casos	Rácio	1991	INE

2.3. NORMALIZAÇÃO, ORTOGONALIZAÇÃO E FACTORIZAÇÃO DOS DADOS

Embora a maioria das variáveis esteja expressa sob a forma de rácio, a estandardização é indispensável, pois o facto de existirem diferentes escalas, como é o caso da Densidade Populacional e do Índice de Poder de Compra, influenciaria a classificação.

No entanto, a estandardização não era suficiente para que houvesse idêntica importância de todas as variáveis, uma vez que se obtinham classificações pouco equilibradas, no que diz respeito ao número de elementos dos *clusters*, e sem justificação plausível. Procedeu-se então a uma análise mais pormenorizada das variáveis, estudando-se o seu comportamento através dos respectivos histogramas, de modo a avaliar a necessidade de proceder a outras transformações algébricas.

Verificou-se que seria conveniente, pela observação dos gráficos, a aplicação da transformação logarítmica em algumas variáveis visto apresentarem uma concentração excessiva de valores baixos. Com a transformação acima mencionada obter-se-ia uma melhor distinção entre concelhos, pois as diferenças seriam como que vistas à “lupa” nas zonas de maior concentração.

No entanto cair-se-ia num dos problemas iniciais: como apenas algumas delas necessitam a aplicação do logaritmo ter-se-iam variáveis com diferentes escalas. Este problema seria mais flagrante ainda no que toca a variáveis de um mesmo grupo temático. Uma vez que a logaritmização de algumas delas é realmente necessária, a solução adoptada foi a de aplicar a mesma transformação às restantes: a estas variáveis a logaritmização não iria trazer qualquer dificuldade acrescida na distinção de concelhos e ficando todas com a mesma escala. As variáveis às quais foi aplicado o logaritmo passam a ser representadas pela sigla da Tabela 2.1 precedida da letra “L”.

Outra abordagem também levada a cabo foi a redução dos dados determinando outras variáveis, que em menor número que as originais, contivessem em si quase

tanta informação como aquelas. É exactamente esta a função do chamado Método das Componentes Principais (Seber, 1984), componentes essas que não são mais do que combinações lineares das variáveis originais, dependendo apenas da matriz de covariâncias (ou de correlações) entre elas. Pareceu razoável que cada uma dessas componentes dissesse apenas respeito a um grupo de variáveis originais que aborda um mesmo tema. Desta forma, obtiveram-se factores que serviram de base ao agrupamento dos concelhos.

Agruparam-se então as variáveis relacionadas com um mesmo tema e aplicou-se-lhes, o Método das Componentes Principais, procedendo-se à interpretação dos factores que explicam, em conjunto, pelo menos 65 % da variância total original.

1. O primeiro grupo é referente às eleições legislativas de 1995:

PS95r
 PSD95r
 CDS95r
 PCP95r

Factores e interpretação:

$$1^{\circ} \text{ factor: } EL95_F1 = -0.292 \text{ PS95r} + 0.378 \text{ PSD95r} + 0.279 \text{ CDS95r} - 0.321 \text{ PCP95r}$$

contraste entre os partidos ditos mais à direita, CDS e PSD, e aqueles mais à esquerda, PS e PCP.

$$2^{\circ} \text{ factor: } EL95_F2 = -0.792 \text{ PS95r} - 0.181 \text{ PSD95r} + 0.287 \text{ CDS95r} + 0.761 \text{ PCP95r}$$

oposição entre os partidos mais votados, PS e PSD, e os que têm menor representação política, CDS e PCP.

2. O segundo grupo trata das condições do regime de ocupação de residência:

LREG_PPr
 LREG_Arr
 LREG_OTr

Factores e interpretação:

$$1^{\circ} \text{ factor: } LREG_F1 = 0.417 \text{ LREG_ARr} - 0.411 \text{ LREG_PRr} + 0.314 \text{ LREG_OTr}$$

contraste entre proprietário e outros regimes

$$2^{\circ} \text{ factor: } LREG_F2 = -0.380 \text{ LREG_ARr} + 0.456 \text{ LREG_PRr} + 1.103 \text{ LREG_OTr}$$

contraste entre arrendamento e outros regimes

3. Trata-se agora do grupo relacionado com a estrutura etária da população:

LT0_14r
 LT15_24r
 LT25_34r
 LT35_44r
 LT45_54r
 LT55_64r
 LT65_74r
 LT75r

Factores e interpretação:

$$1^{\circ} \text{ factor: } LIDAD_F1 = 0.149LT0_14r + 0.160LT15_24r + \\ 0.161LT25_34r + 0.122LT35_44r - \\ 0.33LT45_54r - 0.171LT55_64r - \\ 0.174LT65_74r - 0.170LT75r$$

contraste entre menos de 45 anos e mais de 45 anos

$$2^{\circ} \text{ factor: } LIDAD_F2 = - 0.245LT0_14r - 0.147LT15_24r + \\ 0.085LT25_34r + 0.422LT35_44r + \\ 0.581LT45_54r + 0.078LT55_64r - \\ 0.057LT65_74r - 0.102LT75r -$$

contraste entre menos de 25 anos juntamente com aqueles que apresentam mais de 65 anos (crianças, adolescentes e idosos), ou seja população dependente, e as idades compreendidas entre os 25 e os 65 anos (população em idade activa)

4. Segue-se o grupo de variáveis referentes ao nível de ensino:

LANALFr
LFREQr
LE_BASr
LE_SECr
LE_MEDSUPr

Factores e interpretação:

$$1^{\circ} \text{ factor: } LEDUC_F1 = - 0.230LANALFr + 0.228LFREQr - \\ 0.227LE_BASr + 0.263LE_SECr + \\ 0.268LE_MEDSUP$$

contraste entre pessoas com baixo nível de ensino (analfabetos e ensino básico) e indivíduos com nível de ensino mais elevado (ensino secundário, médio e superior e também aqueles que frequentam o ensino).

$$2^{\circ} \text{ factor: } LEDUC_F2 = - 0.530LANALFr + 0.498LFREQr + \\ 0.556LE_BASr - 0.273LE_SECr - \\ 0.139LE_MEDSUP$$

contraste entre indivíduos analfabetos conjuntamente com aqueles de nível de ensino médio ou superior (-comum) e pessoas com nível de ensino básico e secundário incluindo ainda aqueles que o frequentam neste momento (+comum)

5. O assunto que se trata a seguir é o trabalho e as variáveis:

LTCOr
LTCPr
LTFNRr
LPATRAOr
LEMPOUTR
LCOOPERR

Factores e interpretação:

$$\text{1º factor: } \text{LTRAB_F1} = - 0.349\text{LTCo}_r + 0.333\text{LTCP}_r + 0.327\text{LTFNR}_r + 0.117\text{LPATRAO}_r - 0.054\text{LEMPOUT}_r - 0.113\text{LCOOPER}_r$$

contraste entre trabalhador por conta de outrem e trabalhador por conta própria ou trabalhador familiar não remunerado

$$\text{2º factor: } \text{LTRAB_F2} = - 0.373\text{LTCo}_r - 0.149\text{LTCP}_r + 0.085\text{LTFNR}_r + 0.613\text{LPATRAO}_r + 0.652\text{LEMPOUT}_r + 0.251\text{LCOOPER}_r$$

contraste entre patrão e trabalhador noutra situação

6. Por último, o grupo de variáveis referentes à estrutura familiar:

LAD1SCR_r
LAD1CIMCR_r
LAD2SCR_r
LAD2CIMCR_r
LM3ADSCR_r
LM3ADCIMCR_r
LESTFAMOUT_r

Factores e interpretação:

$$\text{1º factor: } \text{LCRIAN_F1} = - 0.224\text{LAD1SCR}_r + 0.131\text{LAD1CIMCR}_r - 0.233\text{LAD2SCR}_r + 0.209\text{LAD2CIMCR}_r + 0.170\text{LAD3MSCR}_r + 0.209\text{LAD3MCIMCR}_r + 0.086\text{LESTFOU}_r$$

contraste entre as famílias mais numerosas e as menos numerosas.

$$\text{2º factor: } \text{LCRIAN_F2} = 0.186\text{LAD1SCR}_r + 0.448\text{LAD1CIMCR}_r + 0.135\text{LAD2SCR}_r - 0.154\text{LAD2CIMCR}_r - 0.028\text{LAD3MSCR}_r - 0.113\text{LAD3MCIMCR}_r + 0.863\text{LESTFOU}_r$$

grande peso das estruturas familiares de outros tipos.

2.4. CLASSIFICAÇÃO USANDO A ANÁLISE DE CLUSTERS

Antes de mais, foi necessário decidir qual o tipo de distância que traduz a dissimilaridade entre dois objectos. Por não existir motivo algum que leve à escolha de uma distância específica ao tipo de dados em uso, como acontece por exemplo em estudos de Psicologia, e também por ser a mais comum, escolheu-se a distância Euclidiana para definir a dissimilaridade entre dois objectos; no caso, concelhos.

Numa primeira tentativa e para ter uma ideia do número de *clusters* a esperar na solução final, experimentou-se o Método Hierárquico de construção. No entanto, este não apresentou resultados satisfatórios (Brandão, 1997), o que pode estar relacionado com o facto de não permitir a realocação dos objectos após a sua atribuição a um qualquer *cluster*. Assim, procedeu-se à utilização do Método de Partição denominado K-Médias que por possuir esta propriedade está mais vocacionado a lidar com

conjuntos de dados mais numerosos, sendo para além disso o mais frequentemente aconselhado (Ozimek, 1993, Openshaw, 1995).

A Análise de *Clusters* através do Método das K-Médias (Seber, 1984 e Kaufman, 1990) requer, no seu início, a especificação do número de *clusters* pretendido na solução. Uma vez que o Método Hierárquico não desempenhou esse papel, foi necessário recorrer a diversas tentativas, determinando assim uma gama de valores plausíveis.

Já numa última fase do processo e no que diz respeito às variáveis a incluir no processo, nomeadamente à sua manipulação, optou-se por dois tipos de abordagem envolvendo a Análise de *Clusters*:

- uma que utilizasse todas as variáveis, após eventuais transformações em algumas delas e
- outra que fizesse intervir, em vez de todas as variáveis originais, os factores resultantes da aplicação do Método das Componentes Principais.

Até à chegada à última versão, foram-se eliminando na segunda estratégia, factores que se observava não terem influência na solução final. Foi o caso do segundo factor eleitoral, do segundo factor relativo ao regime de ocupação, ao trabalho e à estrutura familiar.

2.5. AVALIAÇÃO E INTERPRETAÇÃO DA SEGMENTAÇÃO

Para escolher a melhor segmentação utilizaram-se critérios que favorecem a existência de uma grande distância entre *clusters* e uma pequena distância entre os membros de um mesmo grupo (Seber, 1984).

Ao designar-se **W** e **B** como matrizes de variação dentro dos *clusters* e entre os seus centróides, respectivamente, a divisão em *clusters* é tanto melhor, quanto menor for **W** e quanto maior for **B**.

Sendo assim, um critério intuitivo para a construção de *clusters* é o de minimizar **W** ou, o que é equivalente, maximizar **B**.

Apresentam-se de seguida três funções de **W** que permitem, de alguma forma, quantificar o seu "tamanho" e consequentemente, minimizá-lo.

São elas:

- (i) Minimizar $\text{tr } \mathbf{W}$
- (ii) Minimizar $|\mathbf{W}|$
- (iii) Maximizar $\text{tr } (\mathbf{B}\mathbf{W}^{-1})$

Devido a constrangimentos de cálculo não foi possível aplicar o último critério, ficando a decisão que diz respeito à melhor solução a cargo dos dois primeiros.

Esta decisão recaiu sobre a segmentação de Portugal Continental em 8 *clusters*, proveniente da utilização dos factores descritos em 2.3 conjuntamente com as variáveis Índice de Poder de Compra, Densidade Populacional e Taxa de Desemprego, sendo estas três logaritimizadas e estandardizadas.

É no entanto de chamar a atenção que uma vez escolhida a segmentação mais adequada, a caracterização dos *clusters* pode e deve ser feita segundo as variáveis

originais, o que significa que o uso dos factores não corresponde a uma perda de informação.

3. RESULTADOS

A apresentação dos resultados é feita considerando em primeiro lugar a constituição dos *clusters* e em segundo a descrição e interpretação dos respectivos perfis.

3.1. CONSTITUIÇÃO DOS CLUSTERS

A solução final obtida é constituída por oito *clusters* com a seguinte composição:

Cluster 1 - 25 elementos:

Amarante, Amares, Barcelos, Castelo de Paiva, Esposende, Fafe, Feira, Felgueiras, Guimarães, Ílhavo, Lousada, Marco de Canavezes, Mesão Frio, Oliveira de Azeméis, Ovar, Paços de Ferreira, Paredes, Penafiel, Peso da Régua, Póvoa de Lanhoso, Póvoa de Varzim, Santo Tirso, São João da Madeira, Vieira do Minho e Vila do Conde.

Cluster 2 - 42 elementos:

Aguiar da Beira, Alijó, Arcos de Valdevez, Armamar, Arouca, Baião, Boticas, Cabeceiras de Basto, Carrazeda de Ansiães, Castro d'Aire, Celorico de Basto, Cinfães, Macedo de Cavaleiros, Moimenta da Beira, Monção, Mondim de Basto, Murça, Oliveira de Frades, Paredes de Coura, Penalva do Castelo, Penedono, Ponte da Barca, Ponte de Lima, Resende, Ribeira de Pena, Sta. Marta de Penaguião, Santa Comba Dão, São Pedro do Sul, Satão, Sernancelhe, Tábua, Tabuaço, Tarouca, Terras de Bouro, V. Nova de Famalicão, Vagos, Valpaços, Vila Flor, Vila Nova de Paiva, Vila Pouca de Aguiar, Vila Verde e Vouzela.

Cluster 3 - 2 elementos:

Lisboa e Porto.

Cluster 4 - 45 elementos:

Águeda, Albergaria-a-Velha, Alcanena, Alcobaça, Anadia, Ansião, Batalha, Bombarral, Cadaval, Caminha, Cantanhede, Carregal do Sal, Chaves, Estarreja, Gouveia, Lamego, Leiria, Lourinhã, Mangualde, Mira, Miranda do Corvo, Mirandela, Montemor-o-Velho, Mortágua, Murtosa, Nelas, Obidos, Oliveira do Bairro, Oliveira do Hospital, Penacova, Pombal, Porto de Mós, Rio Maior, S. João da Pesqueira, Sabrosa, Seia, Sever do Vouga, Tondela, V. Nova de Cerveira, Vale de Cambra, Valença, Viana do Castelo, Vila Nova de Ourém, Vila Nova de Poiares e Viseu.

Cluster 5 - 56 elementos:

Abrantes, Albufeira, Alcochete, Alenquer, Almeirim, Alpiarça, Arruda dos Vinhos, Azambuja, Beja, Benavente, Borba, Bragança, Caldas da Rainha, Cartaxo, Castelo Branco, Condeixa-a-Nova, Constância, Covilhã, Elvas, Évora, Figueira da Foz, Fundão, Golegã, Guarda, Lagoa, Lagos, Loulé, Lousã, Mafra, Manteigas, Mealhada, Montijo, Nazaré, Olhão, Palmela, Peniche, Portalegre, Portimão, Salvaterra de Magos, Santarém, Santiago do Cacém, Sesimbra, Silves, Sines, Sobral de Monte Agraço, Soure, Tavira, Tomar, Torres Novas, Torres Vedras, V. Nova da Barquinha, V. Real de S. António, Vendas Novas, Vila do Bispo, Vila Real e Vila Viçosa.

Cluster 6 - 42 elementos:

Alcoutim, Alfândega da Fé, Aljezur, Almeida, Alvaiázere, Arganil, Belmonte, Celorico da Beira, Ferreira do Zêzere, Figueira de Castelo Rodrigo, Figueiró dos Vinhos, Fornos de Algodres, Freixo de Espada à Cinta, Góis, Idanha-a-Nova, Mação, Marvão, Meda, Melgaço, Miranda do Douro, Mogadouro, Monchique, Montalegre, Nisa, Oleiros, Pampilhosa da Serra, Pedrogão Grande, Penamacor, Penela, Pinhel, Proença-a-Nova, Sabugal, São Brás de Alportel, Sardoal, Sertã, Torre de Moncorvo, Trancoso, Vila de Rei, Vila Nova de Foz Côa, Vila Velha de Rodão, Vimioso e Vinhais.

Cluster 7 - 23 elementos:

Almada, Amadora, Aveiro, Barreiro, Braga, Cascais, Coimbra, Entroncamento, Espinho, Faro, Gondomar, Loures, Maia, Marinha Grande, Matosinhos, Moita, Oeiras, Seixal, Setúbal, Sintra, Valongo, Vila Franca de Xira e Vila Nova de Gaia.

Cluster 8 - 40 elementos:

Alandroal, Alcácer do Sal, Aljustrel, Almodôvar, Alter do Chão, Alvito, Arraiolos, Arronches, Avis, Barrancos, Campo Maior, Castanheira de Pêra, Castelo de Vide, Castro Marim, Castro Verde, Chamusca, Coruche, Crato, Cuba, Estremoz, Ferreira do Alentejo, Fronteira, Gavião, Grândola, Mértola, Monforte, Montemor-o-Novo, Mora, Moura, Mourão, Odemira, Ourique, Ponte de Sôr, Portel, Redondo, Reguengos de Monsaraz, Serpa, Sousel, Viana do Alentejo e Vidigueira.

A constituição dos *clusters* é também apresentada no mapa da Figura 3.1 que permite ter uma visão global e mais esclarecedora dos resultados obtidos.

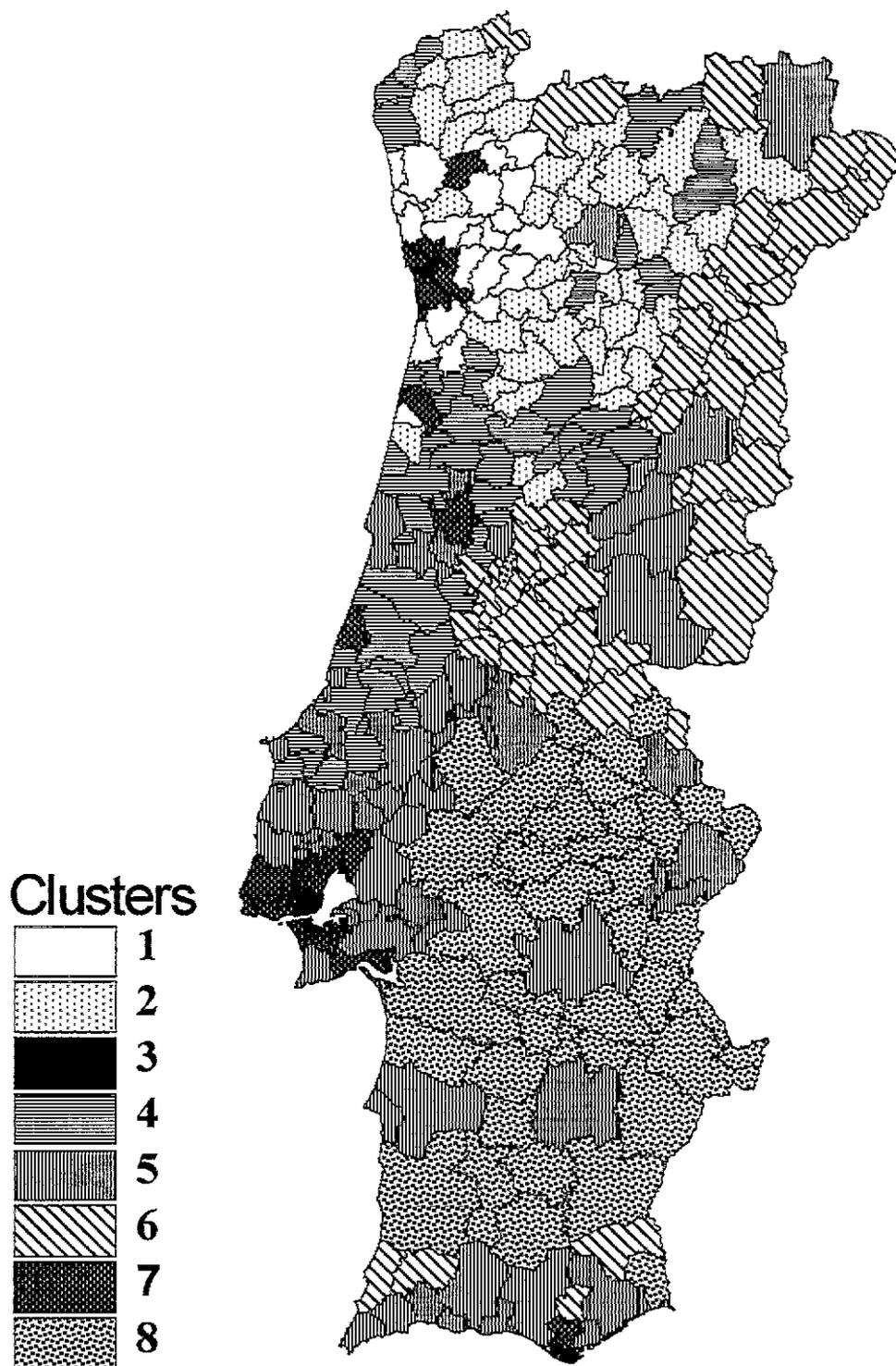


Fig 3.1 Mapa de Portugal Continental com a representação dos 8 clusters

3.2. PERFIS DOS CLUSTERS

Por último faz-se a descrição dos 8 *clusters*, o que corresponde a indicar para cada um e cada variável, das indicadas na tabela 2.1, uma medida de localização. A mediana foi escolhida como medida descritiva em detrimento da média, por ser mais robusta, isto é, menos por eventuais *outliers*.

Obtiveram-se assim as Tabelas 3.1 a 3.7.

Tabela 3.1 - Densidade Populacional em 1991 e Índice de Poder de Compra em 1993

VARIÁVEL (%)	Clusters							
	1	2	3	4	5	6	7	8
POPAREA ²⁵	420.8	58.7	7550.4	106.2	95.7	22.2	1122.3	15.5
IPC ²⁶	45.5	32.2	259.0	61.7	81.1	39.5	112.8	45.0

Tabela 3.2 - Votação nas Eleições Legislativas de 1995

VARIÁVEL (%)	Clusters							
	1	2	3	4	5	6	7	8
CDS95r	5.8	6.8	6.8	6.1	5.1	4.4	5.4	2.6
PCP95r	2.1	1.0	6.3	1.3	6.0	0.9	6.4	17.2
PS95r	31.4	20.6	28.7	24.9	32.8	25.8	32.6	30.3
PSD95r	27.5	28.1	21.6	28.5	17.0	28.1	19.0	12.1

Tabela 3.3 - Regime de Ocupação de Residência de 1991

VARIÁVEL (%)	Clusters							
	1	2	3	4	5	6	7	8
REG_ARr	29.3	7.0	58.9	10.2	20.8	6.3	35.4	15.3
REG_PRR	64.0	88.9	35.6	83.9	71.7	89.6	59.4	74.8
REG_OTr	6.0	3.6	2.8	3.8	5.4	2.9	3.7	9.0

Tabela 3.4 - Estrutura Etária em 1991

VARIÁVEL (%)	Clusters							
	1	2	3	4	5	6	7	8
T0_T14r	24.8	22.2	15.6	20.0	18.2	16.2	20.1	16.8
T15_24r	19.0	16.2	15.7	15.4	14.7	12.8	16.5	13.0
T25_34r	16.3	12.1	13.5	13.1	13.4	10.6	16.0	12.0
T35_44r	12.3	10.2	13.0	12.1	12.8	9.8	15.4	10.6
T45_54r	9.1	10.2	12.4	11.2	12.0	11.0	12.3	11.2
T55_64r	8.6	12.1	13.1	12.0	12.5	14.9	9.7	14.7
T65_74r	6.0	10.1	9.9	9.9	9.6	13.8	6.5	12.5
T75r	3.6	6.7	6.9	6.4	6.2	10.3	3.7	8.5

²⁵ Unidade: nº de habitantes por Km².

²⁶ Por convenção a média do país para o Índice de Poder de Compra é 100. Não possui unidade.

Tabela 3.5 - Nível de ensino Atingido em 1991

VARIÁVEL (%)	Clusters							
	1	2	3	4	5	6	7	8
ANALFr	7.3	13.4	4.8	10.6	11.2	16.9	5.3	18.3
FREQr	15.6	14.6	16.9	14.9	14.9	12.6	18.3	12.7
E_BASr	62.3	60.0	41.4	58.3	51.2	55.8	44.6	52.8
E_SECr	10.0	9.2	21.5	12.9	17.6	10.7	23.0	13.2
E_MEDSUPr	3.0	2.2	15.4	3.5	4.2	2.9	8.0	2.6

Tabela 3.6 - Situação no Trabalho e Taxa de Desemprego em 1991

VARIÁVEL (%)	Clusters							
	1	2	3	4	5	6	7	8
PATRAOr	6.7	6.8	7.5	6.9	6.1	6.5	6.3	5.1
TCOr	79.9	54.2	83.2	68.5	76.1	57.2	82.9	73.3
TCPPr	9.7	29.4	7.4	20.3	15.2	29.8	8.4	17.6
TFNRr	2.1	7.7	0.4	2.9	1.2	4.0	0.6	1.3
COOPERr	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.3
EMPOUTr	0.8	1.0	1.4	0.9	1.2	1.0	1.5	1.4
TXDES_HM	4.2	4.5	7.1	4.7	6.5	5.0	6.6	11.3

Tabela 3.7 - Estrutura Familiar em 1991

VARIÁVEL (%)	Clusters							
	1	2	3	4	5	6	7	8
AD1SCRr	8.0	15.4	20.9	14.0	15.1	19.8	11.9	17.7
AD1C1MCRr	1.2	1.4	1.7	1.3	1.1	0.9	1.7	0.8
AD2SCRr	18.3	25.3	25.3	26.4	28.3	33.1	23.4	32.8
AD2C1CRr	13.6	8.4	8.2	9.9	11.0	7.0	13.9	9.6
AD2C2MCRr	15.9	12.3	6.6	11.6	10.3	8.7	11.5	8.9
AD3MSCRr	21.6	20.4	24.4	21.4	21.6	18.6	24.0	19.4
AD3MC1CRr	12.0	10.0	8.8	10.0	9.4	6.9	10.5	7.5
AD3MC2MCRr	9.1	7.1	4.1	5.6	3.4	3.4	3.7	2.8
ESTFAMOUTr	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0

Uma vez que o processo de identificação e descrição de cada *Cluster* é inevitavelmente subjectivo, tal tarefa deveria ser deixada à responsabilidade de sociólogos ou outros especialistas na matéria. Compete aos estatísticos apresentar ferramentas descritivas suficientes por forma a que a interpretação e avaliação dos clusters seja possível. No entanto o trabalho pareceria incompleto se a nomenclatura não tivesse sido atribuída, sendo também útil para identificação;

A partir destas tabelas e com a ajuda do mapa, fez-se a seguinte caracterização:

Cluster 1: Segunda área envolvente do Porto

- Densidade populacional: média, terceira mais alta;

- Índice de poder de compra: entre os mais baixos (terceira posição);
- Votação nas eleições de 1995: algum equilíbrio entre o PS e PSD mas maioritariamente PS;
- Regime de ocupação de residência dominante: proprietário, apresentando, no entanto, o terceiro menor rácio do país e o terceiro maior em arrendamentos;
- Estrutura etária: predomina a população jovem até aos 25 anos; o rácio da população a partir dos 45 anos é dos mais baixos;
- Nível de ensino atingido: predomina o ensino básico onde atinge o maior rácio do país; Apresenta valores baixos em indivíduos que frequentaram o ensino secundário e médio-superior;
- Situação no trabalho: trabalhador por conta de outrem, sendo o terceiro mais elevado rácio e dos menores para trabalhadores por conta própria;
- Taxa de desemprego: a mais baixa do país;
- Estrutura familiar dominante: possui os maiores rácios do país em famílias com dois ou mais adultos e crianças e apresenta os menores valores para um ou dois adultos sem crianças.

Cluster 2: *Interior médio-norte*

- Densidade populacional: muito baixa (terceiro valor mais baixo);
- Índice de poder de compra: o mais baixo;
- Votação nas eleições de 1995: maioritariamente PSD; atinge o menor rácio de votações no PS e um dos maiores no CDS/PP;
- Regime de ocupação de residência dominante: proprietário, onde apresenta o segundo maior rácio do país tendo também o segundo mais baixo valor de arrendamentos;
- Estrutura etária: predomina a população jovem até aos 25
- Nível de ensino atingido: predomina ensino básico onde atinge o segundo maior rácio; mas em contrapartida, apresenta os menores no ensino secundário e médio-superior e o terceiro maior rácio de analfabetismo;
- Situação no trabalho: obtém o rácio mais baixo de trabalhadores por conta de outrem; e os mais elevados de trabalhador por conta própria e familiar não remunerado;
- Taxa de desemprego: a segunda mais baixa;
- Estrutura familiar dominante: 2 ou 3 adultos sem crianças; apresenta os segundos maiores rácios nas estruturas com duas ou mais crianças;

Cluster 3: *Lisboa e Porto*

- Densidade populacional: a mais elevada;
- Índice de poder de compra: o mais elevado;

- Votação nas eleições de 1995: maioritariamente PS e das maiores votações no CDS/PP;
- Regime de ocupação de residência dominante: arrendamento onde possui o valor mais elevado do país e por acréscimo, o rácio mais baixo em proprietários;
- Estrutura etária: apresenta o rácio mais baixo na classe dos 0 aos 14 anos, o mais elevado na classe dos 45 aos 54 e o terceiro mais elevado nas duas últimas classes;
- Nível de ensino atingido: predomina o ensino básico sendo no entanto o menor rácio do país; atinge o segundo valor mais elevado no ensino secundário e o mais elevado médio superior; apresenta o menor rácio de analfabetismo;
- Situação no trabalho: rácio mais elevado de trabalhadores por conta de outrem, apresentando os valores mais baixos para as duas categorias seguintes;
- Taxa de desemprego: a segunda mais elevada;
- Estrutura familiar dominante: 1, 2 ou 3 adultos sem crianças, apresentando os maiores rácios na primeira e na terceira situações;

Cluster 4: Litoral e Centro

- Densidade populacional: mediana;
- Índice de poder de compra: baixo, mas o quarto mais elevado;
- Votação nas eleições de 1995: algum equilíbrio entre o PS e o PSD mas maioritariamente PSD, em que obteve a maior das votações do país sendo a do PS a segunda menor;
- Regime de ocupação de residência dominante: proprietário, com o terceiro rácio mais elevado e também o terceiro, mas desta vez, mais baixo valor em arrendamentos;
- Estrutura etária: perfil relativamente mediano;
- Nível de ensino atingido: predomina o ensino básico, obtendo o terceiro rácio mais elevado;
- Situação no trabalho: trabalhador por conta de outrem, obtendo no entanto, o terceiro rácio mais baixo do país e o terceiro mais elevado para trabalhadores por conta própria;
- Taxa de desemprego: terceira mais baixa;
- Estrutura familiar dominante: 2 adultos sem crianças: Apresenta valores medianos para quase todo o andamento, excepto para as classes com mais de dois adultos e duas ou mais crianças onde apresenta o terceiro rácio mais elevado;

Cluster 5: Capitais de distrito e concelhos semelhantes

- Densidade populacional: mediana;
- Índice de poder de compra: baixo, no entanto o terceiro mais elevado;

- Votação nas eleições de 1995: maioritariamente PS, atingindo o maior rácio do país para esse partido;
- Regime de ocupação de residência dominante: proprietário, mas com valores medianos em relação a todo o perfil;
- Estrutura etária: pirâmide relativamente equilibrada com valores medianos em relação a todo o perfil;
- Nível de ensino atingido: apresenta valores medianos para todo o andamento;
- Situação no trabalho: trabalhador por conta de outrem, onde obtêm o terceiro valor mais alto do país e por acréscimo o terceiro mais baixo em trabalhador por conta própria;
- Taxa de desemprego: mediana;
- Estrutura familiar dominante: 2 ou 3 adultos sem crianças, obtendo o terceiro maior rácio no primeiro caso e o terceiro no segundo;

Cluster 6: Interior “despovoado”

- Densidade populacional: a segunda mais baixa;
- Índice de poder de compra: o segundo mais baixo;
- Votação nas eleições de 1995: algum equilíbrio entre PS e PSD mas maioritariamente PSD onde obtêm a segunda maior votação do país; Também apresenta o rácio mais baixo para o PCP;
- Regime de ocupação de residência dominante: proprietário e com o rácio mais elevado do país e conseqüentemente o mais baixo em arrendamentos;
- Estrutura etária: os maiores rácios do país na população com mais de 55 anos sendo também o mais predominante; apresenta os menores rácios para as três classes até aos 54 anos;
- Nível de ensino atingido: predomina ensino básico; possui o segundo valor mais elevado de analfabetismo e o menor rácio da população a frequentar o ensino;
- Situação no trabalho: predomina o trabalhador por conta de outrem sendo no entanto o segundo mais baixo; apresenta o valor mais elevado do rácio de trabalhadores por conta própria e o segundo mais elevado de familiares não remunerados;
- Taxa de desemprego: mediana;
- Estrutura familiar dominante: 1, 2 ou 3 adultos sem crianças, obtendo os menores rácios para as restantes variáveis;

Cluster 7: Áreas suburbanas e algo industrializadas

- Densidade populacional: a segunda mais elevada;
- Índice de poder de compra: mediano, mas o segundo mais elevado;
- Votação nas eleições de 1995: maioritariamente PS onde obteve uma das maiores votações do país;

- Regime de ocupação de residência dominante: proprietário, observa-se no entanto o segundo rácio mais elevado de arrendamentos;
- Estrutura etária: população maioritariamente jovem; dos maiores rácios nas classes até aos 54 anos e dos menores a partir dos 55; pirâmide etária estritamente decrescente;
- Nível de ensino atingido: predomina o ensino básico, mas com o segundo mais baixo do país; obtém o rácio mais elevado para o ensino secundário e o segundo mais elevado para o ensino médio-superior;
- Situação no trabalho: trabalhador por conta de outrem, onde apresenta o segundo rácio mais elevado do país e também o segundo, mas mais baixo valor de trabalhadores por conta própria;
- Taxa de desemprego: mediana (terceira mais elevada);
- Estrutura familiar dominante: 3 adultos sem crianças com o valor mais elevado do país para 2 adultos e 1 criança; Apresenta rácios baixos em relação ao resto do país, para as restantes estruturas familiares que não envolvem crianças;

Cluster 8: Alentejo, excepto capitais de distrito

- Densidade populacional: a mais baixa;
- Índice de poder de compra: entre os mais baixos;
- Votação nas eleições de 1995: maioritariamente PS; apresenta o maior rácio de votações no PCP e os menores para o PSD e CDS/PP;
- Regime de ocupação de residência dominante: proprietário, sendo o quarto valor mais elevado em relação ao resto do país;
- Estrutura etária: apresenta dos menores rácios nas classes até aos 44 anos e os segundos maiores para as três classes a partir dos 55 anos.
- Nível de ensino atingido: predomina o ensino básico; mas observa-se o maior valor do rácio de analfabetos;
- Situação no trabalho: são preponderantes os trabalhador por conta de outrem, mas apresentam valores medianos em relação ao resto do país;
- Taxa de desemprego: a mais elevada;
- Estrutura familiar dominante: rácios elevados para as estruturas familiares sem crianças.;

4. CONCLUSÕES

A descrição feita anteriormente reflecte a ideia que todos temos sobre Portugal, mas que sem um estudo deste género, mostramos alguma incapacidade em apontar esta ou aquela característica chave de uma região. Convém lembrar que a ferramenta

estatística utilizada foi a mediana, não se pretendendo uniformizar comportamentos e características de uma região, mas antes conhecê-los em termos médios, ou melhor medianos.

Muito embora uma certa subjectividade estivesse presente nalgumas decisões tomadas ao longo deste trabalho, julga-se que a maior parte acabou por ser fundamentada quer por estudos semelhantes descritos na literatura, quer pela coerência dos resultados obtidos.

A utilidade deste trabalho em estudos de mercado, gestão autárquica ou simplesmente como pura informação, é imediata. Um dos aspectos interessantes que se observa é o facto de cada um dos *clusters* resultantes não formar uma região contígua, isto é, nem todos os concelhos pertencentes a um mesmo grupo possuem fronteiras entre si. Daqui se conclui que a semelhança entre concelhos não se resume à sua posição geográfica mas sobretudo ao tipo de pessoas que lá vivem e às suas condições. Outro assunto para o qual este estudo poderá trazer alguma luz é o da tão falada regionalização.

REFERÊNCIAS BIBLIOGRÁFICAS

- BRANDÃO, A. (1997), *Construção de Perfis Sócio-Económicos de Portugal Continental*, Trabalho Final de Curso da Licenciatura em Matemática Aplicada e Computação, Ramo de Probabilidades e Estatística, Lisboa, IST.
- OPENSHAW, STAN (org.), (1995), *Census Users' Handbook*, Cambridge, GeoInformation International and John Wiley & Sons.
- GOMES, P., BACELAR, S., e SALEIRO, E., (1994), Contributo para a Definição de uma Tipologia Sócio-Económica dos Concelhos da Região Norte, *Revista de Estatísticas e Estudos Regionais do INE - Direcção Regional do Norte*, 5, 6-15
- KAUFMAN, LEONARD e ROUSSEEUW, PETER J. (1990), *Finding Groups in Data - an Introduction to Cluster Analysis*, New York, John Wiley & Sons.
- OZIMEK, JOHN (1993), *Targeting for Success: A Guide to New Techniques for Measurement and Analysis in Database and Direct Response Market*, England, McGraw-Hill, Berkshire.
- SEBER, G. A. F. (1984), *Multivariate Observations*, New York, John Wiley & Sons.
- RAMOS, PEDRO N. (1993), *Estudo sobre o Poder de Compra Concelhio*, Gabinete de Estudos Regionais da Direcção Regional do Centro, INE
- MARKTEST (1994), *Análise do Poder de Compra Regional in Sales Index*

AS SONDAGENS E A POPULAÇÃO - O QUE FAZER
PARA A RECONCILIAÇÃO

Autoras:
Paula Vicente
e
Elizabeth Reis

AS SONDAGENS E A POPULAÇÃO - O QUE FAZER PARA A RECONCILIAÇÃO

OPINION POLLS AND THE POPULATION - HOW TO RECONCILE THEM

Autoras: Paula Vicente

- Assistente no ISCTE e Investigadora da UNIDE - Unidade de Investigação em Ciências Empresariais

e

Elizabeth Reis

- Professora Associada do ISCTE e Investigadora da UNIDE - Unidade de Investigação em Ciências Empresariais

RESUMO:

- Este artigo analisa as razões que levam os portugueses, quando solicitados a colaborar numa sondagem, a recusar fazê-lo, pondo em causa os objectivos do estudo que lhe está subjacente e os próprios resultados. Com base numa consulta a uma amostra de 167 indivíduos, aferem-se, em primeiro lugar, o conhecimento e sensibilidade que os inquiridos têm das sondagens realizadas em Portugal; em seguida, identificam-se os factores que, motivam a recusa, apontando-se soluções para a redução da taxa de recusas; finalmente, identificam-se cinco segmentos de inquiridos, segundo o seu comportamento perante as sondagens: os altruístas, sempre disponíveis para colaborar; os interessados, que quando devidamente estimulados pelo interesse dos temas, não resusam a colaboração; os receosos, a quem a dificuldade de comunicação e o receio de contacto com um estranho, aumentam a probabilidade de se tornarem indisponíveis; os mercantis, indisponíveis e desinteressados à partida, podendo mudar de opinião face a uma recompensa material; e os apáticos, difíceis de caracterizar pela ausência de um comportamento definido.

PALAVRAS-CHAVE:

- *Sondagens, recusas e taxas de resposta, segmentação dos inquiridos, análise de componentes principais, análise de clusters.*

ABSTRACT:

- This article analyses why the Portuguese refuse, when asked, to participate in opinion polls and surveys, thus putting the underlying aims of the study at risk, as well as the actual results. Based on a sample of 167 persons, the respondent's knowledge and awareness of surveys carried out in Portugal was gauged; following this, the main dimensions were extracted which respondent's considered led to their refusal, thus indicating ways to reduce the rate of non-response; finally, five different segments of respondents were identified according to their reactions to

surveys: the unselfish, always willing to co-operate; the interested ones, who do not refuse to co-operate when the subject arouses their interest; the wary, whose difficulty to communicate and fear of contact with strangers increase the likelihood of their being unavailable; the tradesmen, unavailable and disinterested at the start but who might change their mind if given a material reward; the apathetic, who are difficult to characterise due to the lack of clearly defined behaviour.

KEY-WORDS:

- *Opinion polls, surveys, the refusal problem and rates of response, segmentation, principal components analysis, cluster analysis.*

1. INTRODUÇÃO

É mais ou menos frequente ouvirem-se comentários por parte de alguns cidadãos que revelam uma certa desilusão por nunca terem sido seleccionados a colaborar numa sondagem. Num esforço justificativo da sua parte, culpabilizam o azar, ou a falta de sorte, de lhes retirar a oportunidade de participar no “fabrico dos números mágicos”.

Ao mesmo tempo, lamentam-se os profissionais das sondagens por os esquemas de amostragem nem sempre ditarem a selecção de tão voluntariosos participantes, e incluírem alguns outros que, pelas mais diversas razões, acabam por recusar a colaboração. Quando numa sondagem este número é significativo há razões para falar de um problema.

Em qualquer sondagem é difícil, para não dizer impossível, uma taxa de 100% de “sim’s” à colaboração. Carman (1974) sustenta que em qualquer sondagem a taxa de recusas reflecte três aspectos: a eficiência da amostragem, o prestígio da instituição que dirige o estudo e a natureza da tarefa a desempenhar pelos potenciais inquiridos.

Como já referiram Baim (1991) e Meier (1991), o problema das recusas nas sondagens é uma realidade com evolução tendencialmente preocupante para quem depende regra geral, da boa vontade alheia para obter a matéria-prima do seu trabalho - os dados. Apesar de existirem alguns colaboradores entusiastas, não são em número suficiente para contrariar a ideia de que para a maioria das pessoas a participação numa sondagem não constitui um acontecimento de relevo nas suas vidas (Fowler e Mangione, 1993) e portanto estimulante a uma atitude e comportamento mais positivos.

As recusas assumem a forma de problema porque introduzem na sondagem o designado erro de não resposta, contribuindo para afastar os resultados amostrais dos resultados populacionais. As recusas reduzem a dimensão da amostra, facto especialmente visível nas amostras aleatórias, com consequências ao nível da sua representatividade. De facto, como garantir que aqueles que se recusaram a colaborar estão representados na amostra final? Como garantir que as recusas não introduzem enviesamentos nos resultados?

Ainda que o problema das recusas apresente contornos diversos, conforme o método de recolha de informação adoptado na sondagem (entrevista pessoal, telefónica ou questionário por correio), é possível avançar, à luz de estudos já realizados, a identificação de alguns factores comuns explicativos daquele fenómeno.

Tomando por base a tipologia proposta por Struebbe *et al.* (1986), tais factores surgem agrupados em duas grandes ordens. A primeira contempla factores susceptíveis de controlo por parte da empresa que realiza a sondagem - Características do Estudo e Imagem da Empresa que faz a Sondagem junto da Opinião Pública.

(1) **Características do Estudo** - quanto maior é o esforço que o potencial inquirido percebe que está envolvido por participar, mais provável poderá ser a hipótese de recusa. A perspectiva de uma entrevista longa, com perguntas de difícil resposta, sobre um tema pouco interessante, poderão basear a recusa. Também o local e/ou o momento escolhidos para o contacto nem sempre são os mais convenientes ao inquirido.

(2) **Imagem da Empresa que Faz a Sondagem, Junto da Opinião Pública** - este aspecto por vezes passa mais pela ausência de uma imagem. Quantas pessoas serão capazes de mencionar o nome de uma empresa, que em Portugal,

faz sondagens? E destas, quantas não confundirão a empresa que faz a sondagem com aquela que a encomenda e utiliza ou divulga? Até que ponto este facto ajuda o cidadão a ter uma atitude mais positiva quando é abordado para colaborar?

A segunda ordem, inclui factores insusceptíveis de controlo por parte da empresa que realiza a sondagem - Características dos Indivíduos e a Imagem Pública das Sondagens.

(3) **Características dos Indivíduos** - as características dos indivíduos têm também um papel explicativo das recusas numa sondagem. É francamente provável que um indivíduo socialmente inseguro tenda a rejeitar qualquer colaboração perante a abordagem de um estranho - o entrevistador. No entanto, traçar um perfil genérico do recusante é difícil, ainda que alguns estudos já feitos tenham, em casos específicos, apontado caracterizações em termos de sexo, idade, classe social ou ocupação profissional (e.g. Groves, 1989).

(4) **Imagem Pública das Sondagens** - será que as sondagens detêm uma imagem que joga a seu favor? O sentimento de algumas pessoas a respeito das sondagens é de franca descrença. "*Sondagens, a cada um a sua*"; "*As sondagens nunca acertam com os resultados eleitorais*"; "*As sondagens valem o que valem*"; "*Como podem dar as sondagens bons resultados se não me perguntaram a minha opinião?*", são frases comuns da parte dos cidadãos, alguns até com responsabilidade política, escarnecendo das sondagens. Uma incorrecta noção do que é uma sondagem e seus propósitos talvez justifique aquelas afirmações. Identificar a sondagem com uma bola de cristal só pode dar azo a comentários menos favoráveis quando há de facto um afastamento entre os resultados da sondagem e os resultados populacionais.

Na divulgação e interpretação dos resultados das sondagens, têm um papel de responsabilidade os órgãos de comunicação social, que ao produzirem afirmações como "*Se as eleições se realizassem hoje o candidato A obteria 52% dos votos e o candidato B 48%*" ou "*Segundo esta sondagem 60% dos lisboetas apoiam a construção de uma nova ponte sobre o Tejo*" em nada ajudam à formação de uma imagem verídica e credível para as sondagens.

Em 1991, Baim e Meier alargam esta lista propondo mais alguns factores que podem explicar a preocupante baixa na taxa de respostas, em parte pelo factor recusa. Especialmente nas populações residentes em grandes centros urbanos, a densidade populacional, a mobilidade da população, o ritmo acelerado de vida, a taxa crescente de roubos e criminalidade, as preocupações com a privacidade, reduzem a disponibilidade das pessoas para acederem a colaborar com um estranho.

Acrescente-se ainda a questão da desconfiança, em alguns casos motivada pela falta de informação, que faz confundir a sondagem com acções de venda ou mesmo com actividades religiosas. A forma pouco idónea com que a arma da "pequena entrevista" é utilizada para "agarrar" as pessoas é uma adversidade com que as empresas de sondagens se vêem confrontadas.

2. O ESTUDO

As pistas até agora apresentadas são lançadas por estudos realizados noutros países. Será esta também a realidade portuguesa? Que factores explicam a reacção de recusa por parte do inquirido (que muitas vezes não o chega a ser) português?

A resposta a esta interrogação seria idealmente obtida através do contacto com pessoas que recusam a colaboração em sondagens. Tal é, porém, tarefa complicada, pois como conseguir estudar aqueles que não querem ser estudados?

A opção foi então para um estudo por amostragem, tomando como população alvo um conjunto de pessoas que em nosso entender, à partida, soubessem elas próprias apresentar, do seu ponto de vista, razões justificativas da reacção de recusa. A população consistiu de professores e alunos do ensino superior e funcionários administrativos das instituições de ensino envolvidas.

Inquiriram-se 167 indivíduos seleccionados por conveniência.

A recolha da informação foi feita mediante a aplicação de um questionário por auto-resposta, com dois objectivos distintos, mas de alguma forma complementares: (a) identificar factores que, do ponto de vista dos inquiridos, motivam a recusa e encontrar "soluções" para este problema; (b) aferir o conhecimento e sensibilidade que os inquiridos têm das sondagens.

3. OS RESULTADOS

As análises feitas revelaram resultados que, em alguns casos, foram de encontro ao esperado. Há porém outros que merecem alguma reflexão.

3.1. SONDAGEM - O QUE É?

A totalidade dos inquiridos afirmou estar familiarizado com o conceito de sondagem. Porém, quando solicitados a apresentarem o seu conceito de sondagem, constatou-se que apenas uma minoria (cerca de 11%) o conseguiram fazer de modo bem delineado e correcto - Sondagem traduz todas as operações de amostragem que visam o estudo de um universo de elementos a partir de uma amostra (Vicente *et al.*, 1996).

A generalidade das respostas contemplaram definições vagas. Predomina a ideia de que a sondagem é um estudo, de opinião, de mercado ou de um tema. "É um estudo de opinião", "É um estudo para conhecer uma dada realidade" ou "É um estudo de mercado" são frases ilustrativas de 52 respostas obtidas, e que reflectem uma definição de sondagem baseada apenas naquilo que os inquiridos consideram ser a sua finalidade.

O Quadro 1 contém um resumo quantitativo das respostas obtidas.

Quadro 1: Sondagem - Para Quê?

Sondagem - Para Quê?	nº
Estudo de Opinião	32
Estudo de um Tema	10
Estudo de Mercado	4
Previsão	6
Total	52

Outro grupo de respostas apresentou uma definição mais completa indicando, a par da finalidade do estudo, os sujeitos sobre quem ele incide. Encontraram-se referências a amostra, população ou simplesmente a um conjunto de pessoas. Mas foi notório que, para estes inquiridos, não há diferença entre aqueles conceitos.

O número de respostas que conjugaram aqueles dois elementos - Para Quê? e A Quem? consta do Quadro 2.

Quadro 2: Sondagem - Para Quê? e A Quem?

Para Quê?	A Quem?	Amostra	População	Conj. Pessoas
Estudo de Opinião		8	4	6
Estudo de Tema		6	11	6

Foi ainda possível criar uma outra categoria (Sondagem - Como?) para contemplar definições orientadas pelo modo como é feita uma sondagem. Incluíram-se nesta categoria respostas do tipo: "A sondagem é um inquérito" (para alguns um questionário), ou "A sondagem é um estudo quantitativo/estatístico".

O número de respostas contendo uma destas duas ideias foi o seguinte:

Quadro 3: Sondagem - Como?

Sondagem - Como?	n°
Inquérito/Questionário	22
Estudo Quantitativo/Estatístico	8
Total	30

Finalmente, um último grupo de 7 inquiridos que referem nas suas respostas ambos os termos Amostra e População, indiciando que não as consideram equivalentes, mas onde não fica clara a noção de extrapolação dos resultados da amostra à população.

3.2 OBJECTIVO DAS SONDAGENS - QUAL É?

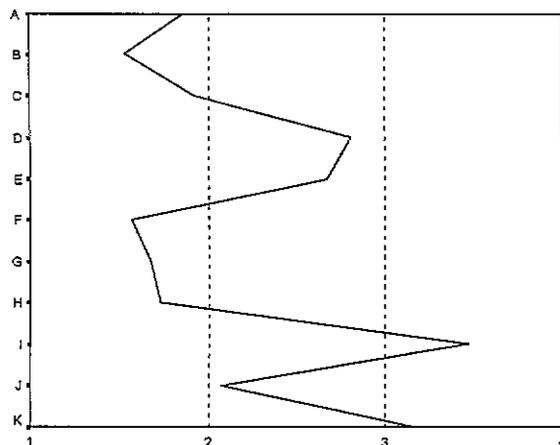
Avaliar a noção que os inquiridos têm dos objectivos de uma sondagem foi também alvo do trabalho feito. Se é certo que as definições de sondagem possibilitaram recolher alguma informação a este nível, pretendia-se esmiuçar um pouco melhor esta questão, pelo que se confrontaram os inquiridos com diversos "objectivos" que uma sondagem pode ter.

Os hipotéticos "objectivos" foram os seguintes:

- A - Projectar qual o candidato presidencial que perderá as eleições
- B - Conhecer a opinião dos cidadãos sobre problemas de ordem social
- C - Prever o comportamento do consumidor
- D - Vender produtos e/ou serviços
- E - Conhecer o comportamento passado das pessoas em determinado domínio de actuação
- F - Conhecer as audiências dos canais de TV
- G - Avaliar a popularidade das figuras públicas
- H - Avaliar a receptividade do mercado a uma nova marca de um produto
- I - Oferecer amostras de produtos
- J - Acertar no partido que vai ganhar as próximas eleições
- K - Fazer promoções de produtos e/ou serviços

Com base na escala (1 - Forte Relação até 4 - Nenhuma Relação) os inquiridos deveriam expressar o grau de relação que, em seu entender, existia entre cada um dos "objectivos" apresentados e o real objectivo de uma sondagem. Os valores médios obtidos para cada "objectivo" foram os seguintes:

**Figura 1 - "Objectivos" de uma Sondagem
(Valores médios)**



Os inquiridos relacionaram a sondagem com o conhecimento de uma determinada realidade (problemas de ordem social, audiências de TV, popularidade de figuras públicas, receptividade do mercado a novos produtos, projecções eleitorais, comportamento do consumidor) e rejeitaram a promoção ou venda de produtos/serviços como finalidades de uma sondagem.

3.3. SONDAGENS EM PORTUGAL - UM PANORAMA

A maioria dos inquiridos (71%) considerou que as sondagens que se fazem actualmente em Portugal são em número razoável. 18% consideraram aquele número baixo; 63% concordaram que se deveriam fazer mais sondagens.

Quanto à credibilidade dos resultados 79% dos inquiridos reconheceram-na como razoável ou mesmo boa, 13% consideraram-na má, e 8% não tem opinião definida.

Quanto à notoriedade reconhecida às empresas/instituições que desenvolvem actividades no domínio das sondagens surgem no topo da lista de mencionadas a Universidade Católica, a Markttest, a SIC, a Norma, o INE e a Metris.

Foi evidente que alguns inquiridos não fazem diferenciação entre as empresas que realizam as sondagens daquelas que divulgam os seus resultados. Esta confusão pode ser justificada, por um lado, pelo facto de os resultados das sondagens serem conhecidos através dos meios de comunicação social, especialmente a televisão e a imprensa, e por outro, por também identificarem a comunicação social como o principal cliente das sondagens. Refira-se ainda que os consumidores e cidadãos em geral são muito pouco mencionados como clientes/beneficiários dos resultados de uma sondagem.

3.4. RECUSAS - PORQUÊ?

A procura de justificações da reacção de recusa, fez-nos incluir no questionário uma questão aberta na qual os inquiridos eram solicitados a, espontaneamente, apontarem três razões potencialmente responsáveis pelas recusas.

De um total de 450 razões apresentadas, a mais referida foi a **Falta de Tempo ou de Disponibilidade para Colaborar** (24% do total de razões).

Seguiu-se a **Falta de Interesse em Participar**, com 16%. Nesta categoria de razões encontraram-se respostas que explicitamente referiam ser aborrecido responder a uma entrevista. A participação foi considerada por outros uma perda de tempo, enquanto o tema desinteressante e a pouca credibilidade das sondagens são outras razões invocadas para a ausência de interesse em colaborar.

Já com algum distanciamento surge a **Desconfiança em Relação à Abordagem** (9%). Algumas pessoas desconfiam do real propósito da entrevista. Associar o entrevistador com um vendedor ou mesmo com um ladrão, é um receio muitas vezes justificado por experiências anteriores negativas.

As **Características da Entrevista**, nomeadamente a duração, o local e o momento em que é feita é uma outra categoria de razões, com 8% de respostas. A duração excessiva, que quase nunca corresponde à inicialmente prometida, os locais seleccionados, alguns com pouca privacidade e os horários escolhidos são motivos de recusa.

Quanto à **Natureza dos Temas**, com 5% de respostas, destacam-se aqueles tópicos de estudo que são íntimos e pessoais e que, por produzirem no inquirido o sentimento de invasão de privacidade, levam também muitas vezes à recusa.

Outras razões menos significativas (com menos de 5% de respostas cada) foram invocadas, como a **Dificuldade de Desempenhar o Papel de Entrevistado** que engloba aspectos como o desconhecimento do tema, e portanto a dificuldade de falar sobre ele ou o receio de não saber responder.

O **Desconhecimento do que é e para que serve uma Sondagem**, a par da **Fraca Garantia de Confidencialidade/Anonimato**, e de **Aspectos que se Prendem com o Entrevistador** são factores menos referidos pelos inquiridos.

Quadro 4: Motivos Para as Recusas

MOTIVOS	Nº
Falta de Tempo / Disponibilidade	107
Falta de Interesse em Participar	74
Desconfiança da Intenção da Abordagem	42
Características da Entrevista	34
Natureza das Questões/Tema	25
Dificuldade de Executar a "Tarefa" Proposta	21
Desconhecimento do Que é uma Sondagem	18
Confidencialidade/Anonimato	15
Aspectos Motivados pelo Entrevistador	15
Outras	99

Em **Outras** ficaram incluídas respostas dispersas por um diverso número de razões, mas onde se podem realçar aquelas que se prendem com as características do próprio inquirido como a timidez, o hábito de recusar o contacto com estranhos ou o não gostar de falar de si ou do tema em causa.

3.5. SENTIMENTOS, EXPECTATIVAS E ATITUDES FACE ÀS SONDAgens

A avaliação das opiniões a este nível foi feita tendo por base um conjunto de 33 afirmações que reflectiam sentimentos, expectativas e motivações face às sondagens. Os inquiridos deveriam expressar o seu grau de concordância com cada uma delas por recurso à escala *1-Discordo Totalmente* até *5-Concordo Totalmente*. Curiosamente, ao serem confrontados com essa listagem, a importância relativa dos factores anteriormente apontados alterou-se.

Os maiores níveis médios de concordância verificaram-se em afirmações relacionadas com o processo de entrevista e o seu contexto: "*A entrevista demora mais do que anunciado*"; "*Os entrevistadores escolhem horas inconvenientes para abordar as pessoas*"; "*A entrevista acaba por ter propósitos de venda*"; "*Tenho medo de abrir a porta a um estranho*"; "*Tenho pouco tempo para perder com entrevistas*".

Um outro grupo de afirmações, onde igualmente se verificaram valores médios de concordância elevados, incluiu afirmações como: "*As sondagens podem fazer melhorar a qualidade de um produto*"; "*As minhas opiniões vão influenciar uma decisão*", que denotam um reconhecimento da importância de colaborar. Esta também é justificada em parte pelo "*Desejo de ajudar o outro*" traduzindo uma postura de altruísmo na colaboração.

Outro grupo de questões reflecte concordância relativamente ao interesse na colaboração: *“Ao colaborar fico a saber mais sobre o tema em questão”*; *“Posso conversar com outra pessoa sobre um tema interessante”*.

Uma Análise de Componentes Principais (ACP) possibilitou um agrupamento das afirmações iniciais em 9 componentes, com um poder explicativo de 54,4% da variância total das respostas dadas.

Quadro 5: Sentimentos, Atitudes e Expectativas nas Sondagens

COMPONENTES PRINCIPAIS	% var. explicada
Indisponibilidade e Falta Interesse	14,7
Dificuldade de Comunicação	8,6
Interesse pelo Tema Abordado	5,8
Desempenho do Entrevistador	5,0
Receio/Medo em Colaborar	4,7
Recompensa	4,4
Reconhecimento da Importância	4,2
Altruismo/Ajuda ao Outro	3,7
Dificuldade em Recusar	3,3

A dimensão **Indisponibilidade e Falta de Interesse em Participar** tem o maior poder explicativo do quadro de sentimentos, atitudes e expectativas da amostra inquirida. *“Tenho pouco tempo para perder com entrevistas”*, *“É aborrecido responder às perguntas”* ou *“Só acedo colaborar se o tema me interessa”* são algumas das afirmações com maior peso nesta componente.

A dimensão **Dificuldade de Comunicação**, explicando 8,6% da variância total, associa-se sobretudo a afirmações que reconhecem no inquirido alguma responsabilidade pela reacção de recusa: *“Raramente se consegue responder às perguntas”*, *“As questões são difíceis de entender”*, *“Muitas vezes não consigo responder por não me lembrar”*.

Opiniões mais unânimes foram registadas em relação às dimensões **Altruismo/Ajuda ao Outro** e **Dificuldade em Recusar**, apresentando portanto um menor poder explicativo da variância total.

3.6. SEGMENTAÇÃO DOS INQUIRIDOS

Perante as atitudes, sentimentos e expectativas demonstrados em relação às sondagens, cuja informação foi sintetizada nas sete anteriores componentes principais, tentou-se uma segmentação dos inquiridos segundo o grau de dissemelhança encontrado nas opiniões dadas. Assim, uma Análise de *Clusters*²⁷ (AC) permitiu encontrar cinco segmentos amostrais com a seguinte caracterização:

- **Altruístas (37%)** - indivíduos disponíveis para colaborar, reconhecem a importância de colaborar numa sondagem e fazem-no fundamentalmente para ajudar o outro.

²⁷ Foram utilizados a distância Euclidiana ao quadrado e o critério de agregação de Ward.

- **Interessados (29%)** - estimulados pelo interesse dos temas, ainda que nalgumas situações se desmotivem perante a dificuldade de comunicação e pelo mau desempenho por parte dos entrevistadores.
- **Receosos (15%)** - caracterizados essencialmente pelo receio em contactar com um estranho a que se junta a dificuldade de comunicação. Além disso não reconhecem importância em colaborar.
- **Mercantis (11%)** - são à partida os mais indisponíveis e desinteressados. Revelam algum receio perante o desconhecido mas valorizam fortemente a recompensa.
- **Apáticos (8%)** - difíceis de caracterizar pela ausência de um traço de comportamento ou sentimental marcante. Talvez a indiferença perante toda a situação de sondagem seja a melhor forma de os definir.

3.7. REDUZIR AS RECUSAS - COMO?

Os motivos de recusa podem ser, como demonstrado, muitos e variados. Mas, em paralelo com o seu conhecimento, devem vir os meios de os minimizar. Foi perguntado aos inquiridos o que se poderia fazer, e obtiveram-se 238 sugestões.

A principal sugestão apresentada constitui um desafio para quem realiza sondagens - **Elucidar as Pessoas da Importância da Sua Colaboração** - mas é talvez o remédio para muitos dos males invocados. Se o inquirido perceber o que está em jogo numa sondagem, e o que poderá advir (para si) de positivo das decisões decorrentes dos resultados de uma sondagem, concerteza teria uma atitude de maior empenho na elaboração de algo que traz benefícios, mais cedo ou mais tarde.

Apesar de o **Entrevistador** não ser apontado como o factor mais potenciador de uma recusa, foi invocado como um aspecto a melhorar em 26% das respostas dadas. Os inquiridos apontaram a necessidade de facilmente identificarem o entrevistador com aquilo que ele é. Houve mesmo propostas de utilização de um uniforme. O profissionalismo e a simpatia destes trabalhadores foram igualmente desejos manifestados. Estas sugestões surgem naturalmente do actual perfil médio do entrevistador português - jovem, estudante, que realiza entrevistas pontualmente e apenas com o objectivo de aumentar o seu rendimento - que de forma alguma encara o trabalho de entrevistador como o seu possível futuro profissional. Corrigir este quadro constitui no mínimo um desafio, uma vez que a desejável formação, adequada e profissional, se torna difícil num grupo que só esporadicamente faz entrevistas e mantém contacto com a empresa empregadora.

Não fazer entrevistas longas foi o principal ponto apresentado na **Melhoria de Factores Associados à Entrevista**.

A **Oferta de Recompensas** materiais, traduzindo o retorno sensível e imediato pelo esforço dispendido teve um peso de 11% no total das sugestões.

Finalmente a opção por **Temas Interessantes** que motivem o inquirido a colaborar foi proposta em 4% do total das sugestões.

Numa altura em que a segmentação de mercados é a frase-chave na definição de uma estratégia de marketing, reconhecendo que nem todos, quer sejam consumidores, utilizadores ou destinatários de um produto, serviço ou mensagem, podem ser tratados da mesma maneira, julgamos ser positivo importar tal conceito para a "resolução" do problema específico das recusas nas sondagens.

Quadro 6: Sugestões para Reduzir as Recusas

SUGESTÕES	Nº
Elucidar as Pessoas da Importância em Colaborar	65
Melhorar Aspectos Ligados com o Entrevistador	63
Melhorar Factores Associados à Entrevista	42
Oferecer Recompensas	25
Escolher Temas Interessantes	10
Outras	33

A AC efectuada revelou que nem todos os (potenciais) inquiridos têm a mesma escala de valores e portanto, a mesma reacção a determinado estímulo. Dar o estímulo certo à pessoa certa é o melhor forma de, pelo menos à partida, assegurar o sucesso de uma estratégia de actuação.

Assim, apostar na informação mais detalhada quanto aos objectivos do estudo, quanto à importância da colaboração e mesmo na justificação do interesse dos temas abordados é a melhor estratégia para os "Altruistas" e para os "Interessados". Os "Receosos" por seu lado, teriam uma mudança positiva de atitude se vissem modificados aspectos relacionados com o desempenho dos entrevistadores. Os "Mercantis" seriam mais facilmente cativados com a melhoria de aspectos ligados com a aplicação do questionário (como o local de realização da entrevista ou a hora escolhida para o efeito), com a oferta de recompensas, não só materiais e com a valorização da importância da sua colaboração. Os "Apáticos" são talvez o grupo mais impenetrável pela dificuldade em perceber o que valorizam.

4. CONCLUSÕES

De uma maneira geral parece existir uma ideia pouco clara e algo incompleta do que é uma sondagem, ainda que a deformação do conceito não seja grave. A sondagem é pensada mais em termos do seu objectivo, e este, marcado pelo tipo de sondagens que a generalidade das pessoas conhece - estudos de populações humanas sobre temáticas sociais, económicas ou políticas.

A falta de tempo e disponibilidade, a desconfiança e o medo, o desagrado pelas características específicas dos estudos (como o tema abordado ou a forma e momento escolhidos para contactar com o inquirido) foram os motivos de recusa que falaram mais alto. *A quem pertence o desafio de mudança deste quadro?*

A dúvida e desconfiança invocadas como razões de recusa têm espaço para existir enquanto não se detém a informação necessária para as fazer dissipar. A amostra de inquiridos parece conhecer razoavelmente o que está implicado numa sondagem. Esta é, porém, uma amostra de indivíduos que podemos considerar privilegiados no que se refere a acesso ao conhecimento e à informação. Mas, e a sociedade civil em geral? Estarão as sondagens justamente divulgadas, sendo-lhes reconhecidas a utilidade e importância que merecem? Talvez não, justificando-se portanto uma intervenção a este nível, podendo aqui caber algum protagonismo a comunicação social pelo seu poder para veicular informação às massas. É igualmente verdade que a familiarização das pessoas com esta forma de estudo, com um progressivo reconhecimento da sua utilidade e importância, passa também pela realização e divulgação de mais sondagens, estando este aumento directamente

dependente da necessidade que os clientes das sondagens têm da sua realização e divulgação.

Minimizar o receio e a desconfiança da população em relação às sondagens poderá também passar por um melhor conhecimento das empresas que as realizam. “Quem é este estranho que me importuna pedindo a minha colaboração?” é a interrogação que desperta receio no espírito da pessoa abordada. Quando o entrevistador se apresenta, a sua imediata e visível identificação com alguém que só quer obter respostas a algumas perguntas, sem intenções secundárias de venda, promoção ou outro tipo de convencimento, será potenciadora de uma atitude cooperativa. A imagem da empresa de sondagens, espelhada, no concreto da entrevista, na figura do entrevistador, deve ser a de uma entidade independente que, movida pela curiosidade da empresa cliente pergunta, indaga, descobre. Se é certo que as empresas de sondagens se podem dar a conhecer pela realização de mais estudos, podem também, através de acções planeadas de comunicação, quer individuais quer concertadas, conseguir alcançar esse objectivo.

A falta de tempo e disponibilidade, por sua vez, tem justificação numa sociedade que é a actual, que vive a um ritmo acelerado, dificultando a tarefa de conseguir captar a atenção das pessoas para algo que, aparentemente e no curto prazo, nada tem para oferecer. O espírito de procura do benefício sensível e imediato a todos os níveis existe também no caso das sondagens. Responder a uma sondagem implica um esforço que se quer recompensado, o quanto antes. Mas onde está a recompensa por participar numa sondagem? Será que a atribuição, em alguns casos, de um pequeno brinde, qual prémio de participação, cumpre essa função? Dificilmente. Cremos que a questão passa por fazer sentir ao inquirido que é ele o beneficiário por colaborar numa sondagem: seja no imediato, ajudando o entrevistador no seu trabalho, seja a médio e longo prazos, através do contributo para o melhoramento dos produtos ou serviços disponíveis para si e para os outros utilizadores ou consumidores.

A pedra de toque de tudo isto parece ser uma premente alteração de mentalidades, nuns casos para desfazer preconceitos e mitos, noutros para aperfeiçoar um conhecimento que a muitos níveis é primário. As Associações Empresariais (em especial do sector de serviços em causa), e especialmente as Escolas e as Universidades, a quem cabe a excelência do ensino, deverão sentir-se chamadas a promover e fornecer mais e melhor (in)formação.

De uma forma ou de outra, responder a uma sondagem é uma oportunidade para o cidadão, como membro responsável de uma comunidade, exercer os seus direitos.

REFERÊNCIAS BIBLIOGRÁFICAS

- BAIM, J. (1991) “Response Rates: A Multinational Perspective”, *Marketing and Research Today*, June, 114-119.
- CARMAN, J. (1974) “Consumer Panels”, *Handbook of Marketing Research*, McGraw-Hill.
- FOWLER, F. J., MANGIONE, T. W. (1990) *Standardized Survey Interviewing*, Sage Publications.
- GROVES, R. (1989) *Survey errors and survey costs*, John Wiley and Sons, NY.
- MEIER, E. (1991) “Response Rate Trends in Britain”, *Marketing and Research Today*, June, 120-123.
- STRUEBBE, J. M., KERNAN, J. B., GROGAN, T. J. (1986) “The Refusal Problem in Telephone Surveys”, *Journal of Advertising Research*, 26, 3, 29-37.
- VICENTE, P., REIS, E., FERRÃO, F. (1996) *Sondagens: A Amostragem como Factor Decisivo de Qualidade*, Ed. Silabo.

VOLUME 1

1º QUADRIMESTRE DE 1998

O MODELO DINÂMICO DE TEIA DE ARANHA E A EXPECTATIVA

Autor:
Valentim Lopes Velasco

O MODELO DINÂMICO DE TEIA DE ARANHA E A EXPECTATIVA

COBWEB DYNAMIC MODEL AND EXPECTATIVE

Autor: Valentim Lopes Velasco

- Técnico superior do Instituto Nacional de Estatística – DRN
- Assistente convidado da Faculdade de Economia do Porto

RESUMO:

- De um modo geral, em microeconomia, o tradicional modelo da procura (D) e da oferta (S) é apresentado em termos estáticos. Por esse facto, o sistema de duas equações de 1º grau que o representa tem uma única solução chamada solução de equilíbrio. Tipicamente o preço de mercado é dado pela intersecção das curvas da procura e da oferta e este processo repete-se indefinidamente - equilíbrio estático. Ora, um simples lance de olhos pela realidade, mostra-nos que as coisas não se passam assim. Este modelo pode ser dinamizado, pelo lado da oferta, fazendo apelo ao conceito matemático diferença finita. Para esse efeito, parte-se do princípio que a procura reage ao preço instantaneamente, i.e., a procura num dado período é função do preço nesse mesmo período, $D_{t+1} = \varphi(P_{t+1})$, enquanto que a oferta reage com o desfasamento de um período, i.e., a oferta num dado período é função do preço no período anterior, $S_{t+1} = \psi(P_t)$. O modelo assim dinamizado pode ter uma menor ou maior adaptação à realidade de acordo com o tipo de expectativa formalizada.

PALAVRAS-CHAVE:

- *Oferta e procura, teia de aranha, trajectória temporal do preço, estabilidade, instabilidade e expectativa.*

ABSTRACT:

- Generally speaking, in microeconomics, the traditional model of supply and demand is presented in static terms. For that very reason, the first-degree equation system, which represents it, has one only solution known as balance solution. Typically the market price is given by the intersection of the supply and demand curves and this process is indefinitely repeated - static equilibrium solution. A quick glance at reality shows us things do not work like this. This model can have certain dynamism, according to the supply, using the mathematical conception finite difference. For this purpose it is assumed that the demand instantaneously reacts to a change in price, that is to say the demand in a certain period is function of the price in that same period, $D_{t+1} = \varphi(P_{t+1})$, whereas the supply reacts with one period difference, that is to say the supply in a certain period is function of price in the previous period, $S_{t+1} = \psi(P_t)$. The dynamic model can have a better adaptation to reality according to the type of formulised prospect.

KEY-WORDS:

- *Supply and demand, cobweb, price temporal trajectory, stability, instability and expectative.*

1. INTRODUÇÃO

As equações funcionais discretas - ordinárias, lineares e com coeficientes constantes - são um instrumento matemático utilizado por vários ramos das ciências para desenvolverem as suas teorias. As ciências sociais em geral e, em particular, a gestão e a economia fazem, frequentemente, apelo a este conceito matemático para dinamizar os seus modelos. Não há, naturalmente, nenhum modelo matemático que tenha uma total adaptação à realidade social. Esta é demasiado complexa para que possa ser explicada por relações matemáticas por mais complicadas que elas sejam. E mesmo quando algumas componentes dessa realidade podem ter um tratamento matemático mais ou menos correcto, há outras cuja complexidade não permite traduzi-las nesses termos. Portanto, a leitura da informação que resulta da sua aplicação deve ser feita com prudência; no dizer de alguém, "há que introduzir, aquando da sua leitura, e correspondente interpretação, a variável bom senso". Quando tais limitações são acauteladas, podemos chegar a algumas conclusões interessantes com a utilização destas "ferramentas".

2. O MODELO DINÂMICO DE TEIA DE ARANHA

Estudo da estabilidade dum equilíbrio de mercado, para um bem perecível e, portanto, não armazenável.

Muitos produtos não podem ser armazenados de um ano para o outro; é o que se passa com a maior parte dos bens de natureza agrícola que são produzidos um ano antes da sua comercialização. As decisões dos produtores quanto às quantidades a produzir são tomadas um período antes da sua venda, ou seja, a oferta corrente depende do preço do ano anterior. Estamos, portanto, em presença dum modelo de mercado sem stock's, o qual nos vai colocar interessantes questões de instabilidade.

Formalização do problema²⁸:

Suponhamos, para simplificar, que as "funções" da procura (D) e da oferta (S) dum dado bem perecível evoluem, ao longo dos diferentes períodos, de acordo com as seguintes equações de primeiro grau:

$$\begin{cases} S_{t+1} = aP_t + b \\ D_{t+1} = -cP_{t+1} + d \end{cases} \quad a, c \in \mathbb{R}^+ \quad (\text{em geral } b < 0 < d) \quad (1.1)$$

Portanto, a oferta no período $t + 1$ depende do preço no período precedente t - um fabricante será tentado a produzir mais se o preço na época anterior se estabeleceu a um nível elevado. A procura depende no período $t + 1$ do preço nesse mesmo período, e o multiplicador "-c" é frequentemente negativo.

Vamos admitir que, em cada período, o mercado determina o preço de tal modo que este torna a procura igual à oferta, ou seja, a procura absorve exactamente as quantidades oferecidas - nenhum produtor fica com output por vender e nenhum consumidor com a procura por satisfazer. Isto significa que a procura no período $t + 1$ é igual à oferta no mesmo período, ou seja:

²⁸ Os parâmetros deste modelo, e dos que se seguem, podem ser estimados através de métodos estatísticos ou econométricos.

$$D_{t+1} = S_{t+1}. \quad (1.2)$$

Resolução e discussão do problema:

Atendendo a (1.2) vem: $P_{t+1} + \frac{a}{c} P_t = \frac{d-b}{c}$. E temos uma equação a diferenças completa, ou não-homogénea, de 1ª ordem, cuja função solução é:

$$P_t = {}^h P_t + {}^p P_t \quad \therefore \quad P_t = \left(P_0 - \frac{d-b}{c+a} \right) \left(-\frac{a}{c} \right)^t + \frac{d-b}{c+a}.$$

Esta função solução tem um preço de equilíbrio dado por $P_0 = \frac{d-b}{c+a} = P_e$, pelo que:

$$P_t = (P_0 - P_e) \left(-\frac{a}{c} \right)^t + P_e$$

P_e representa, economicamente, o preço de equilíbrio, neste caso estático, pois a função solução particular, ${}^p P_t$, é constante. A função solução mostra-nos que se, eventualmente, o preço inicial P_0 é igual a P_e então $P_t = P_e \quad \forall t \in \mathbb{N}_0$ i.e., o preço fixa-se em P_e - desde que não ocorram distúrbios exógenos - e esta situação é designada por equilíbrio estático. A questão de saber se o equilíbrio é estável ou instável prende-se com o estudo da função solução da equação homogénea, ${}^h P_t$.

3. ANÁLISE DA TRAJECTÓRIA TEMPORAL DA EQUAÇÃO DO PREÇO

Normalmente o declive da curva da oferta é positivo e o da curva da procura negativo, pelo que $-\frac{a}{c} < 0$ - pois $a, c \in \mathbb{R}^+$ - logo, a equação geral do preço (P_t) tem uma trajectória temporal alternada - "oscilações impróprias" - em torno do seu valor de equilíbrio. Estas "oscilações impróprias" são convergentes, divergentes ou constantes, de acordo com o valor absoluto dos declives das curvas da oferta e da procura. Assim:

Se $|a| < |c| \Rightarrow \left| \frac{a}{c} \right| < 1$: a equação geral do preço tem uma trajectória temporal alternada com amplitude decrescente e, portanto, convergente.

Se $|a| > |c| \Rightarrow \left| \frac{a}{c} \right| > 1$: a equação geral do preço tem uma trajectória temporal alternada com amplitude crescente e, portanto, divergente.

Se $|a| = |c| \Rightarrow \left| \frac{a}{c} \right| = 1$: a equação geral do preço tem uma trajectória temporal alternada com amplitude constante (divergência finita).

Em jeito de resumo, podemos dizer que, para haver estabilidade, é necessário que a curva da oferta tenha, em valor absoluto, um menor declive que a curva da procura.

Se o declive da primeira é, em valor absoluto, superior ao da segunda, a interacção entre a procura e a oferta produz uma trajectória temporal alternada com amplitude crescente e logo "divergente". O preço afasta-se, ao longo do tempo, do preço de equilíbrio.

Se, pelo contrário, o declive da curva da oferta é, em valor absoluto, inferior ao da curva da procura, a interacção entre a procura e a oferta gera um movimento alternado com amplitude decrescente e, portanto, "convergente". O preço tende, ao longo do tempo, para o preço de equilíbrio.

Se, as duas curvas, têm declives iguais, em valor absoluto, então a interacção entre a oferta e a procura dá origem a uma trajectória temporal alternada com amplitude constante. O preço "oscila", ao longo do tempo, em torno do preço de equilíbrio.

O diagrama abaixo representa um caso possível de movimento alternado (as funções da oferta e da procura foram escritas com o preço como variável independente e representado no eixo das ordenadas²⁹ - como é costume em economia - por esse motivo, os declives devem ser calculados relativamente ao eixo dos Ps. Para este diagrama o declive de (S) é $\text{tg}50^\circ \approx 1,2$, e o declive de (D) é $\text{tg}120^\circ \approx -1,8$).

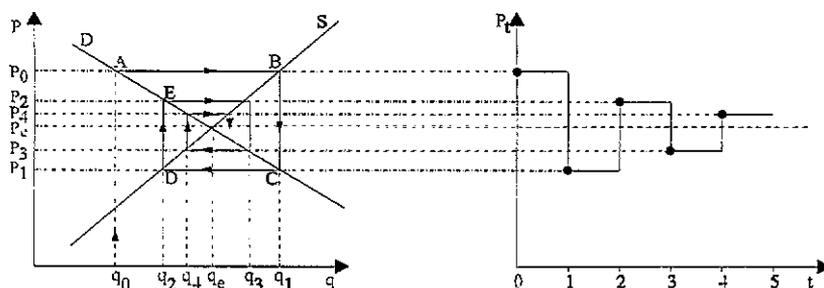


Fig.3.c

Façamos a leitura deste diagrama, recorrendo a uma situação real.

Admita-se que uma produção de concorrência - vamos fazer apelo, v.g., ao caso da criação de aves para abate e conseqüente produção de carne - é posta no mercado, como de costume, de modo a atingir o preço (P), que se obtém deslocando-se na vertical, a partir de uma determinada quantidade (q), até atingir a curva da procura (D).

Agora, o nosso objectivo, é tornar a oferta dinâmica. Para esse efeito vamos supor que os avicultores consideram o preço (P) num dado período e o utilizam para calcular a quantidade (q) que colocarão no mercado no período seguinte: mais pormenorizadamente, se o preço (P) é elevado, aumentarão a criação de aves para abate e, no período seguinte, apresentá-las-ão no mercado. Com efeito, os avicultores têm uma curva da oferta (S), mas a sua actuação é desfasada ligando a quantidade (q) do período seguinte com o preço (P) do período actual - aqui, estamos a considerar o período como o tempo necessário para a criação e abate das aves.

²⁹ Esta opção permite-nos visualizar conjuntamente as curvas da oferta e da procura com a trajectória temporal da equação geral do preço.

Se o preço vigente no mercado fosse aquele que corresponde ao ponto de intersecção da curva da oferta (S) com a curva da procura (D), (q_e , P_e), esta situação representaria um equilíbrio estável, tal como acontece nos modelos estáticos. Em qualquer momento, os avicultores estariam situados na curva (S), a produzir a quantidade (q_e); e os consumidores estariam dispostos a comprar, a esse preço (P_e), exactamente, aquela quantidade que os avicultores haviam decidido comercializar. Por conseguinte, até este momento não há nenhuma diferença.

Vamos admitir, agora, que, inesperadamente, devido a um golpe de canícula, a produção de aves baixa para (q_0). Deslocando-nos para cima até atingirmos a curva da procura no ponto A, obtemos o preço (P_0), mais alto que (P_e), o qual corresponde ao menor volume de output. Mas o processo ainda não terminou uma vez que não estamos, ainda, numa situação de equilíbrio a longo prazo. A explicação reside na resposta à seguinte questão: que quantidade de output produzirão amanhã, os avicultores, a este preço (P_0), mais alto? Movimentar-se-ão para a direita, na sua curva da oferta, criando, no período I, a quantidade correspondente ao ponto B. Como facilmente se verifica, esta quantidade (q_1) é superior a (q_e). Durante quanto tempo será ela vendida no mercado de concorrência? Deslocamo-nos para baixo até à curva da procura e vemos que (P_1) terá de baixar para um valor correspondente a C. Mas o equilíbrio final ainda não foi alcançado. A este preço tão baixo, os avicultores resolvem diminuir o output a apresentar no próximo período, deslocando-se para a esquerda na sua curva da oferta (S) e parando no ponto D. A partir deste ponto, deslocamo-nos para cima, para a curva da procura (D), a fim de atingirmos o preço (P_2) que corresponde ao ponto E.

E assim sucessivamente. Inicialmente (q) é baixo e (P) alto. Mas um (P) alto faz com que, no período seguinte, (q) seja elevado e (P) baixo. Deste modo, tal como um ciclista, ao escalar uma montanha, se inclina demasiadamente para um lado, e depois corrige esse movimento, inclinando-se exageradamente para o outro, assim também o preço de mercado oscila em períodos sucessivos, registando valores ora acima ora abaixo do preço de equilíbrio, traçando uma espécie de teia de aranha³⁰.

Qual vai ser o resultado final? O diagrama que estivemos a analisar foi desenhado com uma curva da oferta com declive, em valor absoluto, inferior ao declive da curva da procura. Deste modo, como se pode ver na Fig. 3.c, as "oscilações impróprias" vão amortecendo gradualmente até que desaparecem por completo: a teia de aranha vai-se enrolando para o centro. Está-se de novo, num equilíbrio no qual se permanecerá até que a próxima perturbação exógena provoque uma nova oscilação, a qual virá, também, a desaparecer gradualmente.

Contudo, numa análise dinâmica, nem todos os pontos de equilíbrio - estático ou dinâmico - são tão estáveis. Os diagramas abaixo mostram-nos outras situações possíveis. Assim, na Fig.3.d a curva da oferta tem, em valor absoluto, maior declive que a curva da procura, de modo que a teia de aranha se expande para fora com "oscilações impróprias" explosivas e o preço tende a aumentar ao longo do tempo. Na Fig.3.e as curvas da oferta e da procura têm, em valor absoluto, o mesmo declive, obtendo-se uma teia de aranha perfeita: conforme a violência da perturbação sofrida pelo mercado, assim o preço oscilará indefinidamente à volta da situação de equilíbrio, sem que as suas "oscilações impróprias" se tornem mais ou menos amplas".

³⁰ E esse aspecto dos diagramas que justifica o nome dado - por Kaldor - ao modelo.

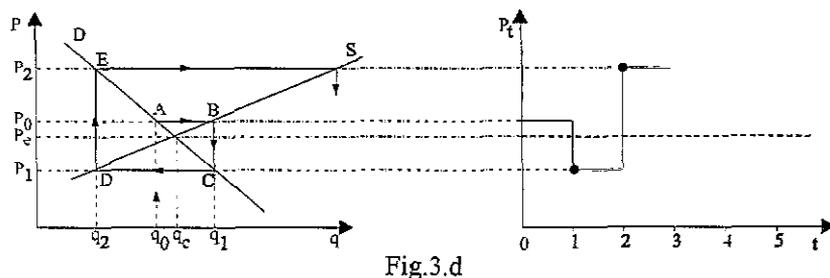


Fig.3.d

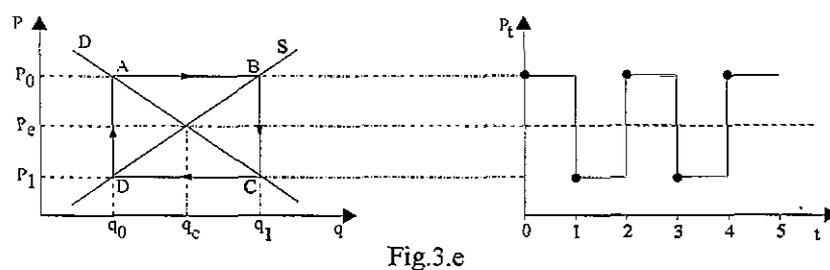


Fig.3.e

Fica assim completo o exame do tradicional modelo de teia de aranha³¹.

Em termos económicos, o elemento irrealista deste modelo reside na hipótese feita a respeito da expectativa. De facto, a não ser no caso singular em que o preço inicial (P_0) é igual ao preço de equilíbrio (P_e), a função solução diz-nos que o preço varia nos diferentes períodos, pelo que a expectativa de que o preço permaneça inalterado ao longo do tempo é incorrecta. Para se proceder à introdução duma expectativa mais de acordo com a realidade, devemos ter em conta que o original modelo de teia de aranha pode ser encarado como um caso particular de um modelo universal, definido da seguinte forma:

$$\begin{cases} S_{t+1} = aP_{t+1}^* + b \\ D_{t+1} = -cP_{t+1} + d \\ D_{t+1} = S_{t+1} \end{cases} \quad a, c \in \mathbb{R}^+ \quad (\text{em geral } b < 0 < d) \quad (1.3)$$

no qual P_{t+1}^* representa o preço "previsto" para o período $t + 1$, i.e., o preço que os produtores no "início" do período - início da produção - pensam que se irá verificar no fim do período - quando o output estiver materializado³².

Vamos analisar três tipos de expectativa, ou seja, três formas possíveis de determinar P_{t+1}^* .

³¹ Estivemos a considerar apenas os casos ditos "normais", ou seja, aqueles em que as curvas da procura e da oferta apresentam, respectivamente, declive negativo e declive positivo.

³² Note-se que a função $S_{t+1} = aP_{t+1}^* + b$ não é descrita como uma curva da oferta habitual. Podemos dizer que ela representa o resultado (colectivo) do comportamento individual dos produtores: se o preço (P_0) é elevado, os produtores, satisfeitos, produzirão no ano seguinte uma quantidade superior (q_1), e o preço cairá então para (P_1); daqui resultará descontentamento, uma produção mais fraca (q_2), um preço (P_2) elevado e assim sucessivamente. Note-se ainda que se S_{t+1} tem um grande declive, em valor absoluto, isso significa que os produtores são muito sensíveis ao preço, durante o período, para escolher o seu próximo nível de produção (se o preço é elevado aumentam muito a sua produção e vice-versa). O declive de S_{t+1} "curva de oferta hipotética" mede, de qualquer modo, a sensibilidade da produção ao preço - com o atraso de um período. Claro que no clássico modelo de teia de aranha, implícita ou explicitamente, $P_{t+1}^* = P_t$.

4. EXPECTATIVA BASEADA NO PREÇO NORMAL

Por preço normal (P_N) entende-se o preço ao qual, de acordo com as expectativas dos produtores, o preço corrente (P_t), mais cedo ou mais tarde, deverá igualar-se. Portanto, se o preço corrente é diferente do preço normal, eles pensam que o primeiro se modificará movendo-se em direcção ao segundo. Uma maneira simples de formalizar este problema é a seguinte:

$$P_{t+1}^* = P_t + \alpha(P_N - P_t) \quad \alpha \in]0, 1[, \quad (1.4)$$

em que o inverso de α representa o número de períodos necessários para ser alcançado o preço normal. V.g.: se $\alpha = 0,5$ então serão necessários dois períodos para alcançar aquele preço. Note-se que se $\alpha = 0$ estamos, novamente, no original modelo de teia de aranha; se $\alpha = 1$ então $P_{t+1}^* = P_N$. Portanto, numa perspectiva económica, um valor de α muito próximo da unidade quer dizer que os produtores estimam que o hiato entre o preço corrente e o preço normal se anulará com celeridade e inversamente, ou seja, um valor muito baixo de α significa que os produtores não esperam que aquele desvio se anule instantaneamente necessitando, para esse efeito, de algum tempo, o qual é, como já foi referido, medido pelo inverso de α .

Se o preço praticado no momento t é inferior ao preço normal, ou seja, se $P_N - P_t > 0$ é esperado um aumento de preço para o próximo período (e inversamente).

Estamos, contudo, perante uma questão delicada e, portanto, de difícil solução que consiste no seguinte: como indicar com exactidão o preço normal? Um caminho simples consiste em admitir, como hipótese, que o preço normal é uma constante igual ao preço de equilíbrio "estático", i.e., $P_N = P_e$. Naturalmente, trata-se duma hipótese pouco consentânea com a realidade, mas que pode ganhar alguma consistência quando, v. g., se durante alguns períodos - antes de se verificarem distúrbios exógenos - o mercado se encontrava em equilíbrio fazendo crer aos produtores que aquele preço não se alteraria.

Posto isto regressemos, de novo, à matemática.

Substituindo em (1. 4) P_N por P_e vem:

$$P_{t+1}^* = P_t + \alpha(P_e - P_t) \quad \alpha \in]0, 1[, \quad (1.5)$$

O modelo (1. 3) é agora:

$$\begin{cases} S_{t+1} = a[P_t + \alpha(P_e - P_t)] + b \\ D_{t+1} = -cP_{t+1} + d \\ D_{t+1} = S_{t+1} \end{cases}$$

pelo que: $P_{t+1} + \left[-\frac{a(1-\alpha)}{c} \right] P_t = \frac{d-b-a\alpha P_e}{c}$.

Trata-se duma equação a diferenças, completa de 1ª ordem cuja função solução é:

$$P_t = (P_0 - P_e) \left[-\frac{a(1-\alpha)}{c} \right]^t + P_e.$$

De acordo com o atrás exposto, o modelo é estável se: $\left| \frac{a(1-\alpha)}{c} \right| < 1$. Se nós compararmos esta condição de estabilidade com a que encontrámos para o primitivo modelo de teia de aranha, verificamos de imediato que, qualquer que seja a situação, se tem sempre: $\left| \frac{a(1-\alpha)}{c} \right| < \left| \frac{a}{c} \right|$. De facto, como $\alpha \in]0,1[\Rightarrow (1-\alpha) \in]0,1[$ e, portanto, $|a(1-\alpha)| < |a|$.

São então legítimas as seguintes conclusões:

- 1ª) Se, com base, no original modelo de teia de aranha a trajectória temporal do preço é alternada com amplitude decrescente e, portanto, convergente, i.e., $\left| \frac{a}{c} \right| < 1$, a introdução deste tipo de expectativa não altera a natureza da trajectória, mas acelera a sua convergência uma vez que: $\left| \frac{a(1-\alpha)}{c} \right| < \left| \frac{a}{c} \right| < 1$ o que faz com que a função solução da correspondente equação homogénea tenda, em valor absoluto, mais rapidamente para zero.
- 2ª) Se, com base, no original modelo de teia de aranha a trajectória temporal do preço é alternada com amplitude constante, i.e., $\left| \frac{a}{c} \right| = 1$, a introdução deste tipo de expectativa altera a natureza da trajectória transformando-a em alternada com amplitude decrescente e, portanto, convergente, i.e., $\left| \frac{a(1-\alpha)}{c} \right| = 1 - \alpha$ e como por hipótese $\alpha \in]0,1[\Rightarrow 1 - \alpha < 1$. Daqui resulta que a função solução da correspondente equação homogénea passe a tender, em valor absoluto, para zero.
- 3ª) Se, com base, no original modelo de teia de aranha a trajectória temporal do preço é alternada com amplitude crescente e, portanto, divergente, i.e., $\left| \frac{a}{c} \right| > 1$, a introdução deste tipo de expectativa conduz a uma das seguinte situações:
- A trajectória divergente passa a convergente.
 - A trajectória divergente passa a uniforme.
 - A trajectória divergente mantém a sua qualidade mas diverge mais lentamente.

Estes resultados mostram-nos que a natureza da nova trajectória depende do peso da divergência e do valor assumido por α . Na realidade, quando a divergência é muito pronunciada e α está próximo da unidade (situação a), a trajectória passa de divergente a convergente, quando a divergência é muito saliente e α não está próximo da unidade (situação c), a trajectória mantém-se divergente, mas diverge mais lentamente.

Estas conclusões permitem afirmar que a introdução da expectativa baseada no preço normal - admitindo, como hipótese simplificadora, que $P_N = P_e$ - torna o modelo mais estável o que representa, sem dúvida, um resultado precioso.

5. EXPECTATIVA AJUSTÁVEL

O original modelo de teia de aranha pode ser ainda aperfeiçoado através da utilização duma expectativa mais adequada à realidade, a chamada "expectativa ajustável". De acordo com esta teoria, a expectativa é examinada minuciosamente em cada período com base no desvio entre o valor observado (P_t) e o valor previsto (P_t^*). A expressão analítica da função que permite expor com precisão este tipo de expectativa é a seguinte:

$$P_{t+1}^* = P_t^* + \gamma(P_t - P_t^*) \quad \gamma \in]0, 1[\quad (1.6)$$

A equação (1. 6) permite concluir o seguinte:

- se o valor observado no período anterior é superior ao valor previsto para esse período, i.e., $P_t - P_t^* > 0$, então o novo valor previsto é corrigido, aumentando, ou seja, $P_{t+1}^* \uparrow$.
- se o valor observado no período anterior é inferior ao valor previsto para esse período, i.e., $P_t - P_t^* < 0$, então o novo valor previsto é corrigido, diminuindo, ou seja, $P_{t+1}^* \downarrow$.
- se o valor observado no período anterior é igual ao valor previsto para esse período, i.e., $P_t - P_t^* = 0$, então o novo valor previsto é igual ao valor esperado para o período precedente, ou seja, $P_{t+1}^* = P_t^*$.

Note-se que, quando $\gamma = 1$ resulta $P_{t+1}^* = P_t$, o que corresponde ao original modelo de teia de aranha. Note-se ainda que a equação (1. 6) não nos fornece, por si só, o valor previsto (P_{t+1}^*) mas, apenas, uma medida da sua variação. Todavia, cálculos simples, dão-nos acesso, através dela, a P_{t+1}^* . De facto, escrevendo a equação (1. 6) na forma:

$$P_{t+1}^* = (1 - \gamma)P_t^* + \gamma P_t \quad (1.7)$$

ou

$$P_t^* = (1 - \gamma)P_{t-1}^* + \gamma P_{t-1} \quad (1.8)$$

Substituindo (1. 8) em (1. 7) vem:

$$P_{t+1}^* = (1 - \gamma)^2 P_{t-1}^* + \gamma(1 - \gamma) P_{t-1} + \gamma P_t \quad (1.9)$$

Substituindo em (1. 9) a expressão de P_{t-1}^* obtida a partir de (1. 8), subtraindo a todos os índices uma unidade, e continuando esse processo, obtemos finalmente:

$$P_{t+1}^* = \gamma P_t + \gamma(1-\gamma)P_{t-1} + \gamma(1-\gamma)^2 P_{t-2} + \dots + \gamma(1-\gamma)^{n-1} P_{t-n+1} + \dots \quad (1.10)$$

ou

$$P_{t+1}^* = \gamma \sum_{n=1}^{\infty} (1-\gamma)^{n-1} P_{t-n+1} \quad (1.10')$$

Este resultado leva-nos a concluir que o preço previsto não é mais do que a média pesada ou ponderada - com os ponderadores a diminuir geometricamente - de todos os preços observados nos períodos anteriores³³.

Temos aqui, no entanto, um problema a resolver. De facto, a utilização da equação (1.10') conduz-nos a equações a diferenças de ordem superior à primeira³⁴. Uma forma de ultrapassar esta situação consiste no seguinte:

$$\text{Como } S_{t+1} = aP_{t+1}^* + b \text{ vem: } S_t = aP_t^* + b \quad \therefore \quad P_t^* = \frac{S_t - b}{a} \text{ e } P_{t+1}^* = \frac{S_{t+1} - b}{a}.$$

Substituindo P_t^* e P_{t+1}^* em (1.7) vem:

$$S_{t+1} = (1-\gamma) S_t + \gamma b + a\gamma P_t \quad (1.11)$$

O modelo (1.3) é agora constituído pelas seguintes equações:

$$\begin{cases} S_{t+1} = (1-\gamma)S_t + \gamma b + a\gamma P_t \\ D_{t+1} = -cP_{t+1} + d \\ D_{t+1} = S_{t+1} \end{cases}$$

Substituindo na 1ª equação S_{t+1} e S_t , respectivamente, por $-cP_{t+1} + d$ e $-cP_t + d$, obtemos:

$$P_{t+1} - \left[1 - \left(1 + \frac{a}{c} \right) \gamma \right] P_t = \frac{(d-b)\gamma}{c}.$$

E temos, novamente, uma equação a diferenças, completa, de 1ª ordem cuja função solução é: $P_t = (P_0 - P_e) \left[1 - \left(1 + \frac{a}{c} \right) \gamma \right]^t + P_e$, na qual $(P_0 - P_e)$ representa o desvio inicial.

A condição para que a equação geral do preço tenha uma trajectória temporal convergente, ou seja, a condição de estabilidade é agora a seguinte:

³³ Dado que $\gamma \in]0, 1[$ a soma dos ponderadores é igual à unidade.

³⁴ Este resultado mostra-nos que a "expectativa adaptável" pode ser utilizada para construir equações a diferenças de ordem superior à primeira. No entanto, as esperanças, de se retirar um claro significado económico das condições de estabilidade, a partir da 3ª ordem, são, infelizmente, muito ténues.

$$\left| 1 - \left(1 + \frac{a}{c} \right) \gamma \right| < 1; \quad -1 < 1 - \left(1 + \frac{a}{c} \right) \gamma < 1, \text{ donde:}$$

$$1 - \frac{2}{\gamma} < -\frac{a}{c} < 1. \quad (1.12)$$

De acordo com o exposto anteriormente, a condição para que a equação geral do preço tivesse uma trajetória temporal convergente, no original modelo de teia de aranha, era $\left| \frac{a}{c} \right| < 1$ ou seja:

$$-1 < \frac{a}{c} < 1. \quad (1.13)$$

A comparação destas duas condições de estabilidade permite tirar a seguinte conclusão:

Dado ser, por hipótese, $\gamma \in]0, 1[$, então $\frac{2}{\gamma} > 2$, pelo que, $1 - \frac{2}{\gamma} < -1$, o que mostra ser, a condição (1.12), menos exigente do que a condição (1.13) no que respeita ao seu lado esquerdo, o qual é, o mais importante, relativamente às funções ditas normais e para as quais $-\frac{a}{c} < 0$.

Portanto, este novo tipo de expectativa torna o modelo ainda mais estável uma vez que, a sua introdução, aumenta a região de estabilidade. De facto, se, v.g., $\gamma = 0,2$, vem neste modelo $-9 < -\frac{a}{c} < 1$, no modelo original $-1 < -\frac{a}{c} < 1$.

1º exemplo.

Vamos ilustrar o exposto com um exercício prático muito simples.

Vamos considerar que as leis de evolução da oferta e da procura, são, em equilíbrio, dadas pelas seguintes funções:

$$\begin{cases} S_{t+1} = 2P_t - 5 \\ D_{t+1} = -P_{t+1} + 10 \end{cases}$$

- Encontre o preço de equilíbrio e verifique se é ou não estável.
- Admita que o preço inicial é $P_0 = 4$ e calcule o valor de P_t até $t = 5$.
- Suponha, agora, que expressão analítica da função oferta é: $S_{t+1} = 2P_{t+1}^* - 5$, em que, P_{t+1}^* é determinado de acordo com a hipótese do "preço normal", i.e., $P_N = P_e$ e $\alpha = \frac{1}{2}$. Qual é, neste caso, a natureza da trajetória temporal do preço?

- d) Admita, finalmente, que o preço é determinado de acordo com a hipótese da expectativa ajustável, para $\gamma = \frac{1}{2}$. Como se comporta, neste novo cenário, a trajectória temporal do preço?

Resolução

- a) Em equilíbrio $D_{t+1} = S_{t+1}$, logo: $P_{t+1} + 2P_t = 15$. Trata-se duma equação a diferenças, não-homogénea, de 1ª ordem.

A função solução da correspondente equação homogénea é: ${}^hP_t = K(-2)^t$.

Uma função solução particular da equação completa é: ${}^pP_t = 5$.

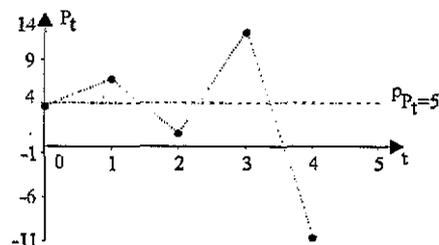
A função solução da equação completa é: $P_t = K(-2)^t + 5$.

O preço de equilíbrio corresponde à função solução particular, da equação não-homogénea, pelo que $P_e = {}^pP_t = 5$.

Uma vez que a base da função exponencial $(-2)^t$ é negativa e superior à unidade, em valor absoluto, o equilíbrio é instável. O preço tem uma trajectória temporal alternada com amplitude crescente, logo divergente. De facto, o declive de (S) é, em valor absoluto, superior ao declive de (D). Ver diagrama abaixo.

- b) Para a condição inicial dada, $P_0 = 4$, vem: $K = -1 \quad \therefore \quad P_t = -(-2)^t + 5$.

t	$P_t = -2P_{t-1} + 15$	$P_t = -(-2)^t + 5$
0	$P_0 = 4$	$P_0 = 4$
1	$P_1 = -2 \times 4 + 15 = 7$	$P_1 = -(-2) + 5 = 7$
2	$P_2 = -2 \times 7 + 15 = 1$	$P_2 = -(-2)^2 + 5 = 1$
3	$P_3 = -2 \times 1 + 15 = 13$	$P_3 = -(-2)^3 + 5 = 13$
4	$P_4 = -2 \times 13 + 15 = -11$	$P_4 = -(-2)^4 + 5 = -11$
5	$P_5 = -2 \times (-11) + 15 = 37$	$P_5 = -(-2)^5 + 5 = 37$



Verifica-se que para $t = 4$ o preço é muito "baixo". Este paradoxo de um preço negativo deve-se ao lado simplista e irrealista das equações de evolução. Estas equações, em particular, não são de facto de 1º grau a não ser numa certa região do plano cartesiano. Na Fig.3.f infra (funções de oferta

³⁵ Note-se que a utilização da equação $P_{t+1} + 2P_t = 15$ ou da sua função solução, $P_t = -(-2)^t + 5$, para aquela condição inicial, conduz, como era de esperar, ao mesmo resultado.

e procura que não são do 1º. grau), a curva da oferta (S) apresenta, no ponto de equilíbrio, maior declive, em valor absoluto, do que a curva da procura (D), a mínima perturbação inicial provocará, no sistema, "oscilações impróprias" de amplitude crescente. Mas nenhuma "explosão" pode desenvolver-se indefinidamente na vida real, de modo que a curvatura das funções conduz, por fim, o sistema ao rectângulo ABCD.

Deste modo, o sistema acaba por oscilar repetidamente sobre o mesmo rectângulo e a amplitude do grau de oscilação será determinada pelas curvas da oferta (S) e da procura (D).

Sempre que factores exógenos originarem nova perturbação no mercado, o sistema tenderá a voltar a este rectângulo, quer a partir do seu interior, quer do seu exterior. E mesmo que o equilíbrio estático (q_e , P_e), fosse, casualmente, restabelecido o abalo mais insignificante afastaria novamente o mercado de um ponto de equilíbrio tão instável.

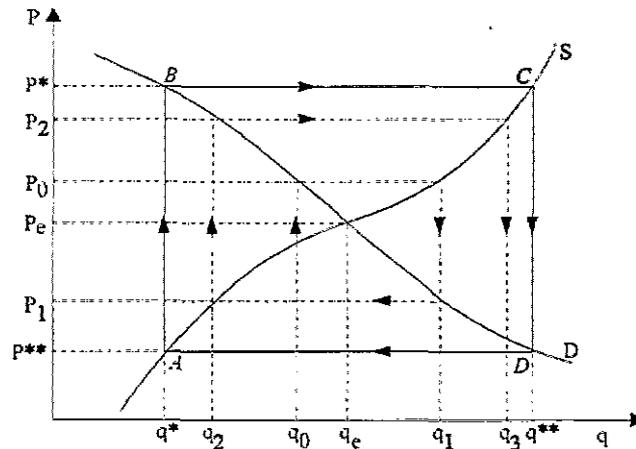


Fig.3.f

c) O sistema é agora:

$$\begin{cases} S_{t+1} = 2P_{t+1}^* - 5 \\ D_{t+1} = -P_{t+1} + 10 \\ P_{t+1}^* = P_t^* + \frac{1}{2}(S - P_t^*) \\ D_{t+1} = S_{t+1} \end{cases}$$

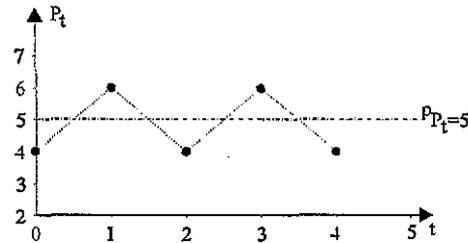
Substituindo P_{t+1}^* em S_{t+1} vem: $S_{t+1} = P_t$. Tendo em conta a condição de equilíbrio obtemos: $P_{t+1} + P_t = 10$. A função solução desta equação a diferenças é:

$$P_t = K(-1)^t + 5.$$

Para a condição inicial dada, $P_0 = 4$, vem: $K = -1 \quad \therefore \quad P_t = -(-1)^t + 5$.

Dada a natureza da expressão analítica da função solução, a trajectória temporal do preço é alternada com amplitude constante. Ver diagrama infra.

Vemos assim que a introdução da hipótese do preço normal veio tornar o modelo mais estável.



d) O sistema é agora:

$$\begin{cases} S_{t+1} = 2P_{t+1}^* - 5 \\ D_{t+1} = -P_{t+1} + 10 \\ P_{t+1}^* = P_t^* + \frac{1}{2}(P_t - P_t^*) \\ D_{t+1} = S_{t+1} \end{cases}$$

Da primeira equação vem: $P_{t+1}^* = \frac{S_{t+1} + 5}{2}$ e $P_t^* = \frac{S_t + 5}{2}$.

Substituindo na 3ª equação vem: $2S_{t+1} - S_t = 2P_t - 5$.

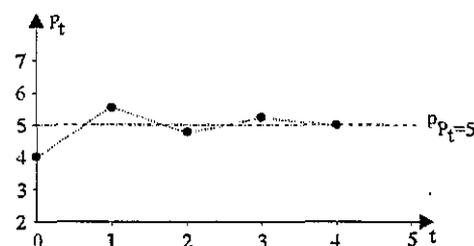
Tendo em conta a 4ª equação e atendendo a que: $D_{t+1} = -P_{t+1} + 10$ e $D_t = -P_t + 10$,

vem: $P_{t+1} + \frac{1}{2}P_t = \frac{15}{2}$.

A função solução desta equação a diferenças é: $P_t = K \left(-\frac{1}{2}\right)^t + 5$.

Para a condição inicial dada, $P_0 = 4$, vem: $K = -1 \quad \therefore \quad P_t = -\left(-\frac{1}{2}\right)^t + 5$.

Verifica-se, de novo, uma melhoria na estabilidade do modelo. De facto, com a introdução da hipótese da "expectativa ajustável", a trajectória temporal do preço torna-se alternada com amplitude decrescente, logo convergente para o preço de equilíbrio estático. Ver diagrama abaixo.



Acabámos de ver que a expectativa subjacente à formulação do original modelo de teia de aranha tem uma fraca adaptação à realidade; vimos também que a introdução das hipóteses do "preço normal" e da "expectativa ajustável" tornavam o modelo mais consentâneo com a realidade.

6. EXPECTATIVA BASEADA NA TENDÊNCIA OBSERVADA

Vamos agora analisar o caso em que a expectativa se exprime, por hipótese, de acordo com uma função cuja expressão analítica é:

$$P_{t+1}^* = P_t + \beta (P_t - P_{t-1}) \quad \beta \in \mathbb{R} - \{0\} \quad (1.14)$$

onde P_{t+1}^* e β representam, respectivamente, o preço previsto para o período $t+1$ e um coeficiente de expectativa³⁶.

Se $\beta > 0$ a expectativa é "extrapolativa", i.e., os produtores prevêem que a tendência observada se manterá no futuro, ou seja, se o preço estava a aumentar (diminuir) ele continuará a aumentar (diminuir). De facto:

$$\text{se } P_t \uparrow, \text{ então } P_{t+1}^* \uparrow, \text{ pois de (1.14) vem: } P_t > P_{t-1} \quad \therefore \quad P_{t+1}^* > P_t.$$

$$\text{se } P_t \downarrow \text{ então } P_{t+1}^* \downarrow, \text{ pois de (1.14) vem: } P_t < P_{t-1} \quad \therefore \quad P_{t+1}^* < P_t.$$

Se $\beta < 0$ a expectativa é tal que os produtores prevêem uma inversão da tendência do preço; se a observação revelou um aumento do preço, prevêem uma diminuição futura; se o preço observado mostrou uma diminuição prevêem um futuro aumento, ou seja:

$$\text{se } P_t \uparrow, \text{ então } P_{t+1}^* \downarrow, \text{ pois de (3.14) vem: } P_t > P_{t-1} \quad \therefore \quad P_{t+1}^* < P_t.$$

$$\text{se } P_t \downarrow \text{ então } P_{t+1}^* \uparrow, \text{ pois de (3.14) vem: } P_t < P_{t-1} \quad \therefore \quad P_{t+1}^* > P_t.$$

Considerando este tipo de expectativa as equações do modelo (1. 3) são as seguintes:

$$\begin{cases} S_{t+1} = a[P_t + \beta(P_t + P_{t-1}) + b] \\ D_{t+1} = -cP_{t+1} + d, \\ D_{t+1} = S_{t+1} \end{cases} \quad (1.15)$$

³⁶ Excluiu-se a possibilidade de $\beta = 0$ porque nesse caso estaríamos, de novo, no original modelo de teia de aranha, ou seja, o preço previsto para o período $t + 1$ seria igual ao preço do período precedente, i. e., $P_{t+1}^* = P_t$. Neste caso, como foi visto, o preço apresenta uma trajetória alternada com amplitude crescente e, portanto, divergente.

Tendo em conta a terceira equação do sistema obtemos:

$$P_{t+1} + \frac{a(1+\beta)}{c} P_t - \frac{a\beta}{c} P_{t-1} = \frac{d-b}{c} \quad (1.16)$$

Trata-se duma equação a diferenças de 2ª ordem não-homogénea, cuja natureza da função solução, ou equação geral do preço, depende do sinal e valor absoluto das raízes da equação característica, ou auxiliar, da correspondente equação homogénea. Vamos, então, calcular a função solução da equação (1.16).

- Cálculo duma função solução particular da equação completa:

Ensaçando a função constante ${}^pP_t = P_e$ obtemos: ${}^pP_{t+1} = {}^pP_t = {}^pP_{t-1} = P_e$, donde:

$$P_e + \frac{a(1+\beta)}{c} P_e - \frac{a\beta}{c} P_e = \frac{d-b}{c} \quad \therefore \quad P_e = \frac{d-b}{c+a}$$

- Cálculo da função solução da correspondente equação homogénea:

A equação característica é:

$$r^2 + \frac{a(1+\beta)}{c} r - \frac{a\beta}{c} = 0. \quad (1.17)$$

A função solução da equação homogénea será:

$${}^hP_t = \begin{cases} K_1 r_1^t + K_2 r_2^t & \text{se } \Delta > 0 \\ (K_1 + K_2 t) r^t & \text{se } \Delta = 0 \\ \rho^t [K_1 \cos \theta t + K_2 \text{sen} \theta t] & \text{se } \Delta < 0 \end{cases}$$

A função solução, ou equação geral do preço, será, então:

$$P_t = \begin{cases} K_1 r_1^t + K_2 r_2^t + P_e & \text{se } \Delta > 0 \\ (K_1 + K_2 t) r^t + P_e & \text{se } \Delta = 0 \\ \rho^t [K_1 \cos \theta t + K_2 \text{sen} \theta t] + P_e & \text{se } \Delta < 0 \end{cases}$$

Onde K_1 e K_2 são duas constantes arbitrárias.

Análise da estabilidade do preço.

As condições necessárias e suficientes de estabilidade, de acordo com o método de Samuelson, são:

$$1^a) \quad 1 + \frac{a(1+\beta)}{c} - \frac{a\beta}{c} = 1 + \frac{a}{c} > 0,$$

$$2^a) \quad 1 - \frac{a(1+\beta)}{c} - \frac{a\beta}{c} = 1 - \frac{a}{c} (1 + 2\beta) > 0, \quad (1.18)$$

$$3^a) \quad 1 + \frac{a\beta}{c} > 0.$$

Vamos analisar estas condições (1.18) para os possíveis valores de β e comparar os resultados com os obtidos quando do estudo do original modelo de teia de aranha.

1º Caso: $\beta > 0$ (estima-se que o preço mantenha a tendência manifestada).

$$1^a) \quad 1 + \frac{a}{c} > 0, \quad 2^a) \quad 1 - \frac{a}{c} (1 + 2\beta) > 0, \quad 3^a) \quad 1 + \frac{a\beta}{c} > 0.$$

Como, por hipótese, a, c e $\beta \in \mathbb{R}^+$, a 1ª e 3ª desigualdades são sempre verdadeiras. Portanto, a desigualdade fundamental é a 2ª. Esta pode ser reescrita na seguinte forma:

$$\frac{a}{c} < \frac{1}{1+2\beta} \quad (1.19)$$

Comparando esta condição de estabilidade com a do original modelo de teia de aranha, a qual é ³⁷:

$$\frac{a}{c} < 1, \quad (1.20)$$

verificamos que esta condição é mais restritiva; na realidade, como $\beta > 0$ temos $\frac{1}{1+2\beta} < 1$. Portanto, a região de estabilidade diminuiu. Tal expectativa, segundo a qual o preço manterá a sua tendência, contém um elemento de instabilidade. Este resultado, para além de intuitivo, é economicamente aceitável uma vez que as expectativas do tipo "extrapolativo" constituem um elemento de instabilidade.

O discriminante da equação característica (1.17) é: $\Delta = \frac{a^2(1+\beta)^2}{c^2} + \frac{4a\beta}{c}$.

Como a, c e $\beta \in \mathbb{R}^+$ Δ é sempre positivo, as suas raízes são reais e distintas. Por outro lado, a sucessão dos sinais dos coeficientes da equação característica (1.17) é (+ + -) permitindo, o teorema de Descartes, concluir que uma raiz é negativa e outra positiva, sendo a negativa superior em valor absoluto. Isto conduz-nos à sobreposição de duas trajectórias, uma alternada e outra monótona, com imposição da primeira.

³⁷ Para funções de oferta e procura "normais", i.e., funções procura com declive negativo e funções oferta com declive positivo.

2º Caso: $\beta < 0$ (estima-se que o preço inverta a tendência manifestada).

$$1^a) 1 + \frac{a}{c} > 0, \quad 2^a) 1 - \frac{a}{c} (1 + 2\beta) > 0, \quad 3^a) 1 + \frac{a\beta}{c} > 0.$$

Como $a, c \in \mathbb{R}^+$, a 1ª desigualdade é sempre verdadeira. Portanto, as desigualdades fundamentais são as outras duas.

A 3ª pode ser reescrita na seguinte forma:

$$1 + \frac{a\beta}{c} > 0 \Rightarrow \frac{a\beta}{c} > -1 \Rightarrow \frac{a(-\beta)}{-c} > -1 \Rightarrow \frac{a}{-c} > \frac{-1}{-\beta} \Rightarrow \frac{a}{-c} > \frac{1}{\beta};$$

$$\frac{a}{c} < \frac{1}{-\beta} \quad (1.21)$$

Para a 2ª devemos considerar dois sub-casos:

se $\beta \leq -\frac{1}{2}$, temos $1 + 2\beta \leq 0$ e a desigualdade é sempre verdadeira, pelo que a única importante é a 3ª.

se $\beta > -\frac{1}{2}$, temos $1 + 2\beta > 0$ e a desigualdade pode ser escrita na forma:

$$\frac{a}{c} < \frac{1}{1+2\beta} \quad (1.22)$$

Importa agora verificar qual das desigualdades é mais restritiva, o que é equivalente a verificar se:

$$\frac{1}{-\beta} \stackrel{?}{=} \frac{1}{1+2\beta}, \quad \text{no intervalo } -\frac{1}{2} < \beta < 0. \quad (1.23)$$

De (1.23) obtemos:

$$1 + 2\beta \stackrel{?}{=} -\beta \quad \therefore \beta \stackrel{?}{=} -\frac{1}{3}, \quad (1.24)$$

logo:

$$\frac{1}{-\beta} \stackrel{?}{=} \frac{1}{1+2\beta} \quad \text{desde que } \beta \stackrel{?}{=} -\frac{1}{3}.$$

Sintetizando os resultados obtidos para $\beta < 0$ temos:

Se $\beta \leq -\frac{1}{3}$, a condição de estabilidade é a (1.21), i.e. : $\frac{a}{c} < \frac{1}{-\beta}$.

Se $-\frac{1}{3} \leq \beta < 0$, a condição de estabilidade é a (1.22), i.e. : $\frac{a}{c} < \frac{1}{1+2\beta}$.

Vamos, agora, tal como fizemos para $\beta > 0$, comparar estas condições de estabilidade com a do original modelo de teia de aranha. As conclusões são as seguintes:

Se $-\frac{1}{3} \leq \beta < 0$, as condições de estabilidade (1.21) e (1.22) são menos restritivas que a condição (1.20). De facto, para b neste intervalo, a condição de estabilidade é a (1.22) que é menos restritiva que a (1.20), uma vez que $\frac{1}{1+2\beta} > 1$, pois $\beta < 0$.

Se $-1 < \beta \leq -\frac{1}{3}$, a condição de estabilidade é a (1.21) e, como $\frac{1}{-\beta} > 1$, é menos restritiva que a (1.20).

Se $\beta = -1$, a condição de estabilidade (1.21) é igual à condição (1.20) pois $\frac{1}{-\beta} = 1$.

Se $\beta < -1$, a condição de estabilidade (1.21) é mais restritiva que a (1.20) pois $\frac{1}{-\beta} < 1$.

Podemos fazer, destes resultados, uma leitura económica. Na realidade, o facto de os produtores preverem uma inversão da tendência do preço é um elemento de estabilidade, desde que essa inversão não seja excessiva. Tal situação é atingida quando o coeficiente de expectativa assume, apenas valores do intervalo $]-1, 0[$. Se este coeficiente pertencer ao intervalo $]-\infty, -1[$ o efeito será exactamente o contrário, i.e., a previsão da inversão da tendência do preço constitui um elemento de instabilidade.

O discriminante da equação característica (1.17) pode apresentar um sinal qualquer, de facto:

$$\text{se } \frac{a}{c} > -\frac{4\beta}{(1+\beta)^2} \Rightarrow \Delta > 0$$

$$\text{se } \frac{a}{c} = -\frac{4\beta}{(1+\beta)^2} \Rightarrow \Delta = 0$$

$$\text{se } \frac{a}{c} < -\frac{4\beta}{(1+\beta)^2} \Rightarrow \Delta < 0$$

Note-se que, quando $\Delta \geq 0$, i.e., quando as raízes são reais, elas são ambas negativas se $\beta \in]-1, 0[$ (de facto, neste caso, a sucessão dos sinais dos coeficientes da equação característica é: + + +); ambas positivas se $\beta \in]-\infty, -1[$ (na realidade, nesta situação, a sucessão dos sinais dos coeficientes da equação característica é: + - +). Quando $\Delta < 0$, i.e., quando as raízes são dois números complexos conjugados, temos, forçosamente, $\beta = -1$; de facto, para este valor do coeficiente de expectativa, a equação característica vem: $r^2 + \frac{a}{c} = 0 \quad \therefore \quad r = \pm i\sqrt{\frac{a}{c}}$. As suas raízes são dois números complexos conjugados puros, dado ser a sua parte real igual a zero. Assim a trajectória temporal pode ser monótona, alternada ou oscilatória. A convergência ou divergência desta trajectória vai ser ditada pelas condições de estabilidade.

Os resultados da análise qualitativa do modelo de Goodwin, para $\beta \in]-\infty, 0[\cup]0, +\infty[$, podem ser representados no plano cartesiano. Para esse efeito vamos considerar as três seguintes funções: $\frac{a}{c} = -\frac{1}{\beta}$, $\frac{a}{c} = -\frac{4\beta}{(1+\beta)^2}$, $\frac{a}{c} = \frac{1}{1+2\beta}$.

Como se verifica, estamos a considerar o quociente entre os declives da curva da oferta e da curva da procura como função do coeficiente de expectativa, β , i.e., $\frac{a}{c} = f(\beta)$. Como, por hipótese, aqueles declives são sempre positivos, do plano cartesiano só nos vai interessar a sua parte positiva, i.e., o primeiro e segundo quadrantes, excluindo os pontos para os quais $\frac{a}{c} = 0$. É esta a razão pela qual, na Fig.3.3, se apresentam, apenas, as porções do traço das diferentes curvas, situadas naquela parte do plano.

As curvas mais carregadas separam a região estável - constituída pelos pontos situados abaixo dessas curvas - da região instável - constituída pelos pontos situados acima delas. A justificação para a utilização de duas curvas, para separar a região estável da instável, deve-se ao facto de, para $\beta \leq -\frac{1}{3}$ a condição de estabilidade, ser $\frac{a}{c} < \frac{1}{-\beta}$, ao passo que, para $\beta \geq -\frac{1}{3}$ a condição de estabilidade é $\frac{a}{c} < \frac{1}{1+2\beta}$. As outras curvas separam, duma forma semelhante, a região das raízes reais da região das raízes complexas. A função $\frac{a}{c} = -\frac{4\beta}{(1+\beta)^2}$ apresenta dois ramos, um para a esquerda e outro para a direita com assíntota vertical $\beta = -1$ (não representada na Fig.3.3). O transformado, ou imagem, de $-\frac{1}{3}$ é, para as três funções, o mesmo, i.e., 3, o que significa ser, $\left(-\frac{1}{3}, 3\right)$ o ponto do plano em que se interceptam.

ANÁLISE DAS DIFERENTES REGIÕES:

Região I

Esta região é instável pois, $\frac{a}{c} > -\frac{4\beta}{(1+\beta)^2}$ e $\frac{a}{c} > \frac{1}{1+2\beta}$, para qualquer dos seus pontos. Além disso $\Delta > 0$, o que implica raízes reais e ambas positivas, pois $\beta \in]-\infty, -1[$. O movimento é monótono e divergente.

Região II

Esta região é instável pois, $\frac{a}{c} > -\frac{4\beta}{(1+\beta)^2}$ e $\frac{a}{c} > \frac{1}{-\beta}$, para qualquer dos seus pontos. Além disso $\Delta < 0$, o que implica serem as raízes dois números complexos conjugados com módulo superior à unidade, donde movimento com oscilações próprias de amplitude crescente e logo divergente.

Região III

Esta região é estável pois, $\frac{a}{c} < -\frac{4\beta}{(1+\beta)^2}$, $\frac{a}{c} > \frac{1}{-\beta}$ e $\frac{a}{c} < \frac{1}{1+2\beta}$, para qualquer dos seus pontos. Além disso $\Delta < 0$, o que implica serem as raízes dois números complexos conjugados com módulo inferior à unidade, donde movimento com oscilações próprias de amplitude decrescente e portanto convergente.

Região IV

Esta região é estável pois, $\frac{a}{c} > -\frac{4\beta}{(1+\beta)^2}$ e $\frac{a}{c} < \frac{1}{1+2\beta}$, para qualquer dos seus pontos. Além disso $\Delta > 0$, o que implica raízes reais e ambas negativas, pois $\beta \in]-1, 0[$, o movimento é alternado com amplitude decrescente e logo convergente.

Região V

Esta região é estável pois, $\frac{a}{c} > -\frac{4\beta}{(1+\beta)^2}$ e $\frac{a}{c} < \frac{1}{1+2\beta}$, para qualquer dos seus pontos. Além disso $\Delta > 0$, o que implica raízes reais uma positiva e outra negativa, pois $\beta \in]0, +\infty[$, sendo a negativa superior à positiva, em valor absoluto, donde movimento alternado com amplitude decrescente e logo convergente, a sobrepor-se a um movimento monótono.

Região VI

Esta região é instável pois, $\frac{a}{c} > -\frac{4\beta}{(1+\beta)^2}$ e $\frac{a}{c} > \frac{1}{1+2\beta}$, para qualquer dos seus pontos. Além disso $\Delta > 0$, o que implica raízes reais ambas negativas, pois $\beta \in]-1, 0[$, o movimento é alternado com amplitude crescente e logo divergente.

Região VII

Esta região é instável pois, $\frac{a}{c} > -\frac{4\beta}{(1+\beta)^2}$ e $\frac{a}{c} > \frac{1}{1+2\beta}$, para qualquer dos seus pontos. Além disso $\Delta > 0$, o que implica raízes reais ambas negativas, pois $\beta \in]0, +\infty[$, sendo a negativa superior à positiva, em valor absoluto, donde um movimento alternado com amplitude crescente e logo divergente, a sobrepor-se a um movimento monótono.

Este estudo só fica completo depois da análise dos pontos das curvas que delimitam as diferentes regiões. Assim:

A) O conjunto de pontos que separa a região I da região II verifica: $\frac{a}{c} = -$

$$\frac{4\beta}{(1+\beta)^2} \text{ e } \frac{a}{c} > \frac{1}{-\beta}. \text{ Trata-se, portanto, de um conjunto instável; além disso } \Delta$$

$= 0$, o que implica uma multiplicidade de duas raízes reais e iguais ambas positivas, pois $\beta \in]-\infty, -1[$; o movimento é monótono e divergente.

B) O conjunto de pontos que separa a região II da região VI verifica: $\frac{a}{c} = -$

$$\frac{4\beta}{(1+\beta)^2} \text{ e } \frac{a}{c} > \frac{1}{-\beta}. \text{ Trata-se, portanto, de um conjunto instável; além disso } \Delta$$

$= 0$, o que implica uma multiplicidade de duas raízes reais e iguais ambas negativas, pois $\beta \in]-1, -\frac{1}{3}[$; o movimento é alternado e divergente.

C) O conjunto de pontos que separa a região V da região VII verifica: $\frac{a}{c} > -$

$$\frac{4\beta}{(1+\beta)^2} \text{ e } \frac{a}{c} = \frac{1}{1+2\beta}. \text{ Trata-se, portanto, de um conjunto instável; além disso,}$$

$\Delta > 0$, o que implica raízes reais uma positiva e outra negativa, sendo a negativa superior à positiva, em valor absoluto, pois $\beta \in]0, +\infty[$; o movimento é alternado com amplitude constante a sobrepor-se a um movimento monótono. Trata-se de um conjunto de pontos que separa estabilidade de instabilidade.

D) O conjunto de pontos que separa a região IV da região VI verifica: $\frac{a}{c} > -$

$$\frac{4\beta}{(1+\beta)^2} \text{ e } \frac{a}{c} = \frac{1}{1+2\beta}. \text{ Trata-se, portanto, de um conjunto instável; além disso}$$

$\Delta > 0$, o que implica raízes reais e ambas negativas, pois $\beta \in]-\frac{1}{3}, 0[$; o movimento é alternado com amplitude constante. Trata-se de um conjunto de pontos que separa estabilidade de instabilidade.

E) O conjunto de pontos que separa a região II da região III verifica: $\frac{a}{c} < -$

$$\frac{4\beta}{(1+\beta)^2} \text{ e } \frac{a}{c} = \frac{1}{-\beta}. \text{ Trata-se, portanto, de um conjunto instável; além disso } \Delta$$

< 0 , o que implica serem as raízes dois números complexos conjugados com módulo igual à unidade. De facto, $A_2 = -\beta \frac{a}{c} = 1$, donde movimento com

oscilações próprias de amplitude constante. Trata-se de um conjunto de pontos que separa estabilidade de instabilidade.

F) O conjunto de pontos que separa a região III da região IV verifica: $\frac{a}{c} = -\frac{4\beta}{(1+\beta)^2}$ e $\frac{a}{c} < \frac{1}{-\beta}$. Trata-se, portanto, de um conjunto estável, além disso $\Delta = 0$, o que implica uma multiplicidade de duas raízes reais e iguais ambas negativas, pois $\beta \in]-\frac{1}{3}, 0[$; o movimento é alternado com amplitude decrescente e logo convergente.

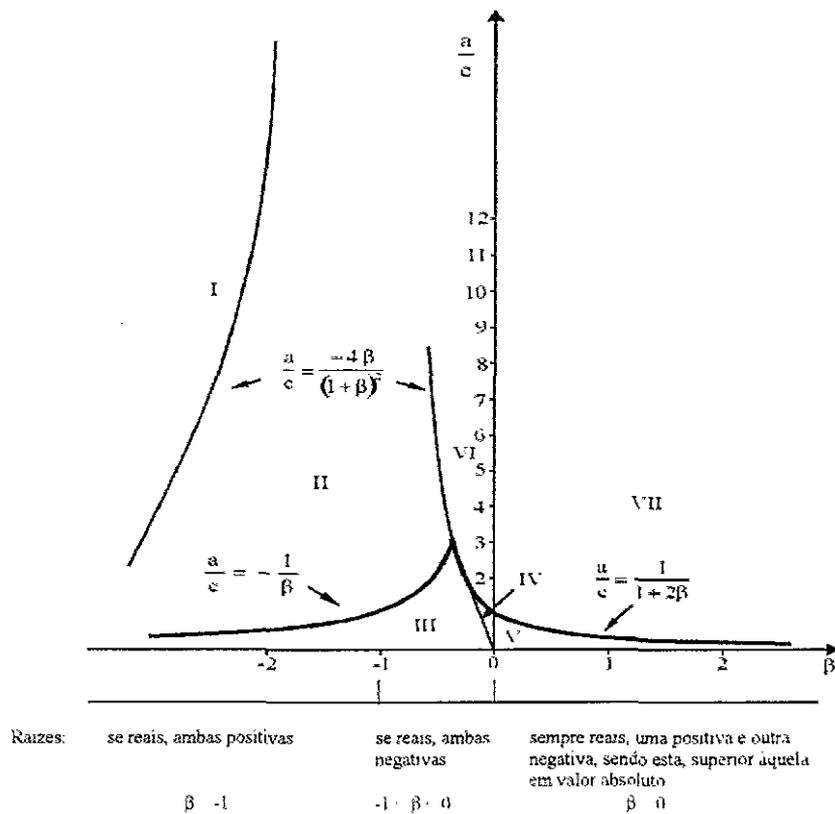


Fig.3.3

Fica assim completada a análise do diagrama da Fig.3.3. Este diagrama ajuda-nos a visualizar que, quando β é menor que zero, é mais provável ocorrerem raízes complexas do que raízes reais, excepto quando β é muito pequeno ou muito grande, em valor absoluto, e o declive da função oferta é muito maior que o da função procura, em valor absoluto.

2º Exemplo

Admita que as leis de evolução da oferta, da procura e do preço, de um dado bem, são as seguintes:

$$\begin{cases} S_{t+1} = 2P_{t+1}^* - 5, \\ D_{t+1} = -P_{t+1} + 10, \\ P_{t+1}^* = P_t + \beta(P_t - P_{t-1}), \\ D_{t+1} = S_{t+1}. \end{cases}$$

- a) Suponha que $\beta = -\frac{1}{3}$ e estude o comportamento da equação geral do preço, ao longo do tempo, determinando, para esse efeito, a sua expressão analítica.
- b) Sabendo que no momento $t = 0$ o preço está em equilíbrio e que no momento $t = 1$, um distúrbio exógeno o altera para o valor 10, determine a localização da função solução no plano cartesiano.

Resposta:

- a) Note-se, antes de mais, que $\beta = -\frac{1}{3}$ e $\frac{a}{c} = 2$, pelo que o par ordenado

$\left(-\frac{1}{3}, 2\right)$ pertence à Região III, que verificámos ser estável (ver Fig.3.3).

As raízes da equação característica são nesta região, para aquele ponto, dois números complexos conjugados e, portanto, o movimento tem oscilações próprias com amplitude decrescente, (amortecido), logo, convergente para o preço de equilíbrio.

Feita esta observação vamos encontrar a expressão analítica da equação geral do preço.

Tendo em conta a quarta equação do sistema vem: $P_{t+1} + \frac{4}{3} P_t + \frac{2}{3} P_{t-1} = 15$.

Trata-se duma equação a diferenças de 2ª ordem completa cuja função solução é:

- Cálculo duma função solução particular da equação não-homogénea.

$$\text{Ensaando } P_t = P_e \text{ vem: } P_e + \frac{4}{3} P_e + \frac{2}{3} P_e = 15 \quad \therefore \quad P_e = 5.$$

Esta função solução particular representa, como sabemos, o preço de equilíbrio. Trata-se de um equilíbrio estático, uma vez que estamos na presença de uma função constante, e estável dado ser, como já concluímos, a função solução ou equação geral do preço, convergente.

- Cálculo da função solução da correspondente equação homogénea:

As raízes da equação característica são: $r_1 = -\frac{2}{3} + i\frac{\sqrt{2}}{3}$ ou $r_2 = -\frac{2}{3} - i\frac{\sqrt{2}}{3}$.

Como:

$$\rho = \sqrt{\frac{2}{3}}$$

$$\cos\theta = \frac{\sqrt{6}}{3} \quad \text{e} \quad \sin\theta = \frac{\sqrt{3}}{3} \quad \therefore \quad \theta = (144,73561)^\circ$$

vem:

$${}^h P_t = \left(\sqrt{\frac{2}{3}}\right)^t [K_1 \cos(144,73561)^\circ t + K_2 \sin(144,73561)^\circ t].$$

A função solução da equação completa é então:

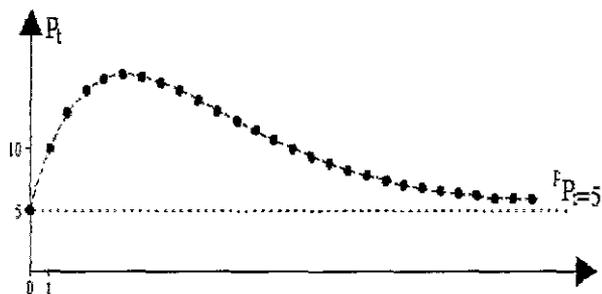
$$P_t = \left(\sqrt{\frac{2}{3}}\right)^t [K_1 \cos(144,73561)^\circ t + K_2 \sin(144,73561)^\circ t] + 5.$$

Dado ser $\rho < 1$, o movimento apresenta, tal como referido supra, oscilações próprias com amplitude decrescente, logo convergente para o preço de equilíbrio. Ver diagrama infra.

Verificamos assim que a introdução deste tipo de expectativa vem, novamente, tornar o modelo mais estável. Naturalmente que esta melhoria não é alheia à escolha de β .

b) Para as condições iniciais dadas, $P_0 = 5$ e $P_1 = 10$, vem: $K_1 = 0$ e $K_2 = \frac{15\sqrt{2}}{2}$ pelo que:

$$P_t = \frac{15\sqrt{2}}{2} \left(\sqrt{\frac{2}{3}}\right)^t \sin(144,73561)^\circ t + 5.$$



REFERÊNCIAS BIBLIOGRÁFICAS

- ALLEN, R. G. D.(1959); *Economía matemática*.
- ALLEN, R. G. D.(1967); *Teoría macroeconómica*.
- BAUMOL, W. J.,(1951); *Economic Dynamics*.
- BAUMOL, W. J.,(1970); *Economic Dynamics*.
- BAUMOL, W. J.,(1948); "Notes on some Dynamics Models", *Economic Journal*.
- CHIANG, ALPHA C.(1982); *Matemática para economistas*.
- GANDOLFO, GIANCARLO (1973); *Metodi di Dinamica Economica*.
- GOLDBERG, SAMUEL (1958); *Introduction to Diference Equations*.
- GOODWIN, R. M.(1947); *Dynamical Coopling With Especial Reference to Makets Having Production Lags*.
- GUELFOND, A. O.(1963); *Calcul des Différences finies*.
- MURATA, Y,(1977); *Mathematics for Stability and Optimization of Economic Systems*.
- KALDOR, N.(1960); *Determinanteness of Static Equilibrium; Nerlove, M.Adaptive Expectations and Cobweb; Phenomena*.
- SAMUELSON, PAUL A.(1981); *Economia*.
- SAMUELSON, PAUL A.(1941a); "Conditions that the Roots of a Polynomial be Less than Unity in Absolute Value".
- WEBER, JEAN E. (1986); *Matemática para Economia e Administração*.
- YAMANE, TARO (1977); *Matemática para economistas*.

TRÊS PROBLEMAS DE PARTIÇÃO EM AMOSTRAGEM:
ESTRATOS, DOMÍNIOS E GRUPOS

VOLUME 1

1º QUADRIMESTRE DE 1998

TRÊS PROBLEMAS DE PARTIÇÃO EM AMOSTRAGEM: ESTRATOS,
DOMÍNIOS E GRUPOS.

THREE PROBLEMS OF PARTITION IN SAMPLING: STRATA,
DOMAINS AND GROUPS.

Autor: Osvaldo Caldeira
Técnico Superior de Estatística do Instituto Nacional de Estatística
Docente do Instituto Superior de Contabilidade e Administração de Lisboa

SUMÁRIO:

- Estratos, Grupos e Domínios representam três variantes distintas da partição de uma população em sub-conjuntos. Na **Amostragem Estratificada**, a população é dividida em sub-populações disjuntas, os estratos. Para cada estrato é concebido um desenho de amostragem. O objectivo é conceber um desenho de amostragem eficiente e simultaneamente prático. O **Modelo de Grupo** associa a cada grupo um ou mais parâmetros caracterizadores. Assim, para um dado desenho de amostragem, a homogeneidade de cada grupo pode ser explorada para se obterem melhores estimativas. Um **Domínio** é qualquer sub-população para qual se solicita, antes ou depois do desenho de amostragem, uma estimativa desagregada. São analisadas algumas das consequências e as suas soluções.

PALAVRAS-CHAVE:

- *População Finita, Estratos, Estimador de Horvitz-Thompson, Super-População, Estimador da Regressão, Model-Assisted, Grupos, Domínios.*

SUMMARY:

- Strata, Groups and Domains represents three types of partitioning of the population into subsets. In stratified sampling, the population is divided into nonoverlapping subpopulations called strata. A sampling design must be specified in each stratum. The objective is to select an efficient, yet practical, sampling design. Group Model have one or more parameters associated with each individual group. For a given sampling design, the homogeneity within groups can be exploited to obtain improved estimates. A domain is any subpopulation for which a separate estimate may be requested, before or after sampling design. We examine some of the issues and their solutions.

KEY-WORDS:

- *Finite Population, Horvitz-Thompson Estimator, Strata, Superpopulation, Regression Estimator, Model-Assisted, Groups, Domains.*

1. INTRODUÇÃO

O conceito de **Estrato** encontra-se ligado à ideia de se usar a informação existente na **Base de Amostragem**³⁸ (População de Referência) para se proceder à sua partição. Desta forma, constituem-se estratos homogêneos, em termos da informação disponível. Para cada um deles é concebido um desenho de amostragem (eventualmente distinto, se se achar conveniente) e da sua aplicação é gerada - em conformidade - uma amostra a observar. Em resultado da exploração desta homogeneidade resulta, em geral, uma estratégia de amostragem mais eficiente. A partição da população em estratos **precede a etapa do desenho** e condiciona a sua concepção. Visa-se, deste modo, obter **uma estratégia de amostragem (desenho+estimação) mais eficiente para a População como um todo**. A técnica encontra uma justificação plena no quadro da amostragem na população finita e preserva uma total autonomia em relação a qualquer modelo subjacente da Super-População. As referências pioneiras sobre a estratificação devem-se a Tschuprow (1923) e a Neyman (1934). Vários autores estudaram outros aspectos da estratificação especialmente considerados em Hansen e Hurwitz (1943) e Wright (1983).

Já a noção de **Grupo**, embora também se encontre ligada a uma partição realizada de acordo com a informação disponível na Base de Amostragem, difere, uma vez que os grupos formados apenas são usados para efeitos de **estimação**. Concretizando, o desenho de amostragem é concebido e realizado, sem atender a qualquer partição da população. Porém, os grupos, assim como os estratos, tiram partido da homogeneidade interna para aumentar a eficiência da estratégia de amostragem. Existem várias razões que podem impedir o uso dessa informação para efeitos de estratificação, das quais destacariamos desde já duas: (1) o facto de haver uma multiplicidade de variáveis auxiliares e de interesse (a recolher), das quais apenas as mais relevantes poderão ser consideradas e/ou (2) o facto de a informação necessária para efeitos da partição só poder ser disponibilizada após a consumação do desenho de amostragem. A "post-estratificação", embora com um sentido menos amplo, é ainda uma aplicação popular dos Grupos para efeitos de estimação. A fundamentação dos estimadores em termos dos Modelos da Super-População (e dos Grupos, em particular) ocupou diversos autores ao longo dos últimos vinte anos. Referências significativas sobre o assunto podem ser vistas em Holt e Smith (1979), Särndal (1982), Isaki e Fuller (1982), Wright (1983) e Särndal, Swensson e Wretman (1989).

Se bem que as motivações de ambos os procedimentos referidos visem a eficiência da estimação, deve dizer-se que o recurso à constituição de Grupos pode introduzir algum enviesamento em amostras de pequena dimensão, o que não acontece na Estratificação. De facto, o uso de Grupos, como técnica de estimação, apoia-se em modelos cujo enviesamento só é passível de correcção em amostras de média ou grande dimensão. Nestas amostras a estimação é assistida pelo modelo, mas enquadra-se ainda nos procedimentos não dependentes do modelo. Contudo, a sua justificação cabal assenta no recurso a modelos da Super-População.

O conceito de **Domínio** encontra-se ligado à necessidade de produzir informação não apenas para a População mas também para os seus Sub-Conjuntos

³⁸ A Base de Amostragem pretende ser uma representação adequada (operacional) da População. Para o efeito, a Base de Amostragem identifica todas as Unidades de Amostragem e dispõe para cada uma delas de informação auxiliar. Em particular, é esta informação auxiliar que é usada para realizar a partição referida. Importa não esquecer que esta será sempre uma operacionalização imperfeita, isto é, onde são inevitáveis os desajustamentos em relação à População.

relevantes, isto é, que derivam de uma partição solicitada pelo utilizador da informação. Os domínios, ao contrário dos estratos ou dos grupos, não exploram a homogeneidade para aumentar a eficiência, mas apenas respondem a solicitações dos utilizadores. Casos há em que a partição solicitada só é realizável mediante o recurso a informação recolhida por amostragem ou, sendo possível, pode ocorrer num momento em que o desenho de amostragem se encontra consumado. Assim, todo o esforço se centra na etapa da estimação, não recorrendo necessariamente à informação da Base de Amostragem, nem para efeitos de concepção do desenho de amostragem, nem para auxiliar a estimação. A estimação em Domínios tem sido uma área de intensa investigação durante a última década. Sobre o tema, referências significativas podem ser vistas em Purcell e Kish (1979), Plateck e outros (1987) e Särndal e Hidiroglou (1989).

A importância destes conceitos não reside, ao contrário do que pode parecer, apenas na necessidade de conferir rigor teórico a um ramo da estatística que necessita de fundamentar os seus procedimentos. O principal benefício dos conceitos apresentados reside no seu poder de “bem formular” e “bem resolver” alguns dos problemas que com frequência se apresentam no desenvolvimento das estratégias de amostragem. Em concreto, ajuda a identificar e equacionar devidamente problemas consoante se trate de promover a eficiência “a priori”, de promover a eficiência “a posteriori” ou apenas de satisfazer a necessidade de informação.

Um exemplo da importância e do uso destes conceitos ocorre em Inquéritos que se repetem no tempo. São frequentes nestes inquéritos as referências a “mudanças de estratos” ao longo do tempo, isto é, algumas unidades de amostragem que se encontravam classificadas num dado estrato no momento em que se realizou a amostragem seriam, no momento da estimação, - sob a mesma regra³⁹ de partição - colocadas num estrato diferente. Diga-se de antemão que, de acordo com o sentido atribuído aos conceitos, não se trata verdadeiramente de estratos. Se o objectivo da reclassificação for o aumento da eficiência da estratégia (através do uso de informação auxiliar disponível), dever-se-á utilizar a designação de grupo; se, pelo contrário, o objectivo for divulgar a informação de acordo com a nova classificação, então a designação de domínio será mais adequada.

Um outro exemplo do interesse destes conceitos prende-se com as várias dificuldades metodológicas que podem impedir a produção de informação estatística regional, algumas das quais relativas à concepção da estratégia de amostragem. Estas últimas são, em muitos casos, removíveis mediante uma utilização adequada das ferramentas apresentadas, em particular, usando com maior propriedade o conceito de domínio. A apresentação do texto tira partido da notação e da sistematização efectuada em Särndal, Swensson e Wretman (1992).

2. ESTRATOS

Na estratificação é realizada uma partição da população, com base em informação auxiliar⁴⁰, dando origem a sub-conjuntos (partes) com algum grau de

³⁹ Com generalidade, diga-se que qualquer alteração da classificação ou apenas das suas regras, ocorrida após a execução do desenho, poderá ser equacionada através do adequado uso destes conceitos.

⁴⁰ A constituição de estratos é um problema anterior ao desenho de amostragem e que não diz respeito exclusivamente à Teoria da Amostragem. Ainda assim colocam-se algumas questões pertinentes, a saber: (1) Quais as variáveis (de entre as disponíveis) a usar, para se realizar a estratificação da população? (2) Sendo a variável de estratificação uma variável quantitativa, quantos estratos (ou quantos intervalos de classe) deverão ser criados através dessa variável? (3) Sabendo o número de estratos a criar, como se definem os limites de cada um dos intervalos de classe? Como referimos, a sua resolução é anterior à amostragem, ou seja, ocorre no universo (base de amostragem) dispondo de informação para todos os elementos do universo (unidades de amostragem).

homogeneidade interna, designados por estratos. O desenho de amostragem é concebido para cada estrato, mas a estimação visa apenas a produção de estimativas para a população como um todo.

A amostragem estratificada é caracterizada do seguinte modo:

1. A população U é previamente particionada em H sub-populações, cada uma delas designada por estrato U_h ($h=1, \dots, H$) de dimensão conhecida $|U_h| = N_h$.
2. A cada estrato h é associado um desenho de amostragem $p_h(\cdot)$, independente dos desenhos associados a todos os outros estratos. A realização do desenho de amostragem dá origem à amostra s_h .

O desenho de amostragem $p(\cdot)$ da população é o resultado dos vários desenhos parcelares. Atendendo à independência dos desenhos parcelares, conclui-se imediatamente que:

$$p(s_1, \dots, s_h, \dots, s_H) = p_1(s_1) \dots p_h(s_h) \dots p_H(s_H) \quad (2.1)$$

A amostra da população s a observar será a reunião das amostras geradas para cada um dos estratos, isto é:

$$s = \bigcup_{h=1}^H s_h \quad \text{com} \quad p(s) = p_1(s_1) \dots p_h(s_h) \dots p_H(s_H) \quad (2.2)$$

Em face da caracterização do desenho de amostragem, resultam imediatas as probabilidades de inclusão⁴¹ na amostra de 1ª e 2ª ordem:

$$\begin{aligned} \pi_k &= \pi_{k/h} \quad k \in U_h \\ \pi_{kl} &= \begin{cases} \pi_{kl/h} & k, l \in U_h \\ \pi_{k/h_1} \pi_{l/h_2} & k \in U_{h_1}, l \in U_{h_2} \end{cases} \end{aligned} \quad (2.3)$$

Um estimador não enviesado de t (total da variável de interesse y) será dado por:

$$\hat{t} = \sum_{h=1}^H \hat{t}_h \quad \text{se} \quad E(\hat{t}_h) = t_h \quad (2.4)$$

onde:

$$t = \sum_{h=1}^H t_h \quad \text{e} \quad t_h = \sum_{k \in U_h} y_k \quad (2.5)$$

⁴¹ Para as probabilidades de inclusão de 1ª e 2ª ordem é usada a notação habitual com o seguinte significado:

$$\begin{aligned} \pi_k &= P(k \in s) & \pi_{k/h} &= P(k \in s_h) \\ \pi_{kl} &= P(k, l \in s) & \pi_{kl/h} &= P(k, l \in s_h) \end{aligned}$$

Sob a hipótese de desenhos de amostragem independentes em cada estrato, poderá a sua variância ser expressa do seguinte modo:

$$V_{ST}[\hat{t}] = \sum_{h=1}^H V_h[\hat{t}_h] \quad (2.6)$$

a estimar através de:

$$\hat{V}_{ST}[\hat{t}] = \sum_{h=1}^H \hat{V}_h[\hat{t}_h] \quad \text{com} \quad E[\hat{V}_h[\hat{t}_h]] = V_h[\hat{t}_h] \quad (2.7)$$

A interpretação da expressão da variância permite concluir que se a fonte de variabilidade no Universo for a variância entre os estratos (inter-estrato), e não no interior de cada estrato (intra-estrato), através da estratificação (devido à independência dos desenhos), limitar-se-á a variabilidade do estimador.

Realizada a estratificação, há que definir uma estratégia de amostragem (desenho de amostragem + estimador) para cada estrato. O estimador natural (não-enviesado), na ausência de informação auxiliar, é o estimador de Horvitz-Thompson (HT) para cada estrato h , isto é:

$$\hat{t}_{\pi h} = \sum_{k \in S_h} \bar{y}_k \quad \bar{y}_k = \frac{y_k}{\pi_{k|h}} \quad (2.8)$$

cuja variância será dada por:

$$V(\hat{t}_{\pi h}) = \sum_{k, l \in U_h} \bar{y}_k \bar{y}_l \overbrace{(\pi_{kl|h} - \pi_{k|h} \pi_{l|h})}^{\Delta_{kl|h}} = \sum_{k, l \in U_h} \Delta_{kl|h} \bar{y}_k \bar{y}_l \quad (2.9)$$

e um seu estimador não-enviesado será:

$$\hat{V}(\hat{t}_{\pi h}) = \sum_{k, l \in S_h} \bar{\Delta}_{kl|h} \bar{y}_k \bar{y}_l \quad \bar{\Delta}_{kl|h} = \frac{\Delta_{kl|h}}{\pi_{kl|h}} = \frac{\bar{\pi}_{kl|h} - \pi_{k|h} \pi_{l|h}}{\pi_{kl|h}} \quad (2.10)$$

O desenho a nível do estrato, entre outros, pode assentar na Amostragem Simples Sem Reposição (SI), de Bernoulli (BE) ou Sistemática (SY). Conforme os casos, o desenho completo (se se adoptar o mesmo desenho para todos os estratos) designar-se-á por Amostragem Estratificada Simples Sem Reposição (STSI), por Amostragem Estratificada de Bernoulli (STBE) ou por Amostragem Estratificada Sistemática (STSY). Eis as expressões derivadas, para o caso do desenho STSI¹²:

¹² Uma vez que para o desenho SI no estrato h temos:

$$\pi_{k|h} = \frac{n_h}{N_h} \quad \text{e} \quad \pi_{kl|h} = \frac{n_h(n_h - 1)}{N_h(N_h - 1)} \quad k \neq l$$

De modo equivalente, para o desenho de BE teríamos:

$$\pi_{k|h} = \pi_h \quad \text{e} \quad \pi_{kl|h} = \pi_h^2 \quad k \neq l$$

$$\hat{t}_\pi = \sum_{h=1}^H N_h \bar{y}_{s_h} \quad \bar{y}_{s_h} = \frac{1}{n_h} \sum_{k \in s_h} y_k \quad (2.11)$$

$$V_{STSI} [\hat{t}_\pi] = \sum_{h=1}^H N_h \frac{1-f_h}{n_h} S_{yU_h}^2 \quad (2.12)$$

com
$$S_{yU_h}^2 = \frac{1}{N_h - 1} \sum_{k \in U_h} [y_k - \bar{y}_{U_h}]^2 \quad \bar{y}_{U_h} = \frac{1}{N_h} \sum_{k \in U_h} y_k$$

$$\hat{V}_{STSI} [\hat{t}_\pi] = \sum_{h=1}^H N_h \frac{1-f_h}{n_h} S_{ys_h}^2 \quad (2.13)$$

com
$$S_{ys_h}^2 = \frac{1}{N_h - 1} \sum_{k \in s_h} [y_k - \bar{y}_{s_h}]^2$$

Um dos problemas mais conhecidos no contexto da estratificação consiste em realizar a repartição óptima da amostra. Concretizando, pretende-se determinar a dimensão da amostra (n) e reparti-la pelos seus diferentes estratos (n_h) de forma a minimizar a variância do estimador da população, respeitando o orçamento estabelecido para a realização do inquérito por amostragem. Alternativamente, o problema da repartição óptima pode ser formulado visando uma dada precisão. Nesta versão, o problema consiste em proceder à repartição da amostra para se atingir a variância desejada ao custo mínimo.

O problema reveste-se de especial interesse (teórico e prático) nos casos em que a variância pode assumir a seguinte forma:

$$V_{ST} [\hat{t}_\pi] = \sum_{h=1}^H \frac{A_h}{n_h} + B = V \quad \text{com } A_h \text{ e } B \text{ não envolvendo } n_h \quad (2.14)$$

Estão nestas condições, entre outros, os desenhos⁴³ STSI e STBE, de uso muito frequente.

⁴³ Nestes casos, temos respectivamente (com n e n_h - amostra esperada):

$$\text{STSI: } A_h = N_h^2 S_{yU_h}^2 \quad B = -\sum_{h=1}^H N_h S_{yU_h}^2$$

$$\text{STBE } A_h = p_h N_h^2 S_{yU_h}^2 \quad B = -\sum_{h=1}^H p_h N_h S_{yU_h}^2 \quad p_h = 1 - \frac{1}{N_h} + c v_{yU_h}^{-2}$$

Admitindo um custo do inquérito uniforme no interior de cada estrato, o custo global da estratégia será dado por uma função de custo linear:

$$C = c_0 + \sum_{h=1}^H n_h c_h \quad c_h > 0 \quad c_0 - \text{custo fixo não dependente de } n_h \quad (2.15)$$

A minimização da variância em n_h , sujeita a um orçamento pré-fixado (C - fixo), é equivalente à minimização de:

$$V^* = V - B = \sum_{h=1}^H \frac{A_h}{n_h} \quad \text{sujeita à restrição} \quad C^* = C - c_0 = \sum_{h=1}^H n_h c_h \quad (2.16)$$

A minimização da função custo em n_h , sujeita à exigência de uma variância pré-fixada (V - fixo), é equivalente à minimização de:

$$C^* = C - c_0 = \sum_{h=1}^H n_h c_h \quad \text{sujeita à restrição} \quad V^* = V - B = \sum_{h=1}^H \frac{A_h}{n_h} \quad (2.17)$$

Qualquer das formulações, (2.16) ou (2.17), é equivalente à minimização do produto V^*C^* . Considerando a desigualdade de Cauchy-Schwartz¹⁴, obtemos a seguinte relação:

$$*C^* = \sum_{h=1}^H \frac{A_h}{n_h} \sum_{h=1}^H n_h c_h \geq \left[\sum_{h=1}^H \left(\frac{A_h}{n_h}\right)^{1/2} (n_h c_h)^{1/2} \right]^2 = \left[\sum_{h=1}^H (A_h c_h)^{1/2} \right]^2 \quad (2.18)$$

Fica assim determinado um minorante do produto V^*C^* , o qual ocorre se e só se:

$$(A_h^{-1} n_h)^{1/2} (n_h c_h)^{1/2} = k \text{ (cte)} \Rightarrow n_h = k (A_h c_h^{-1})^{1/2} \quad (2.19)$$

donde se conclui que o número de indivíduos a observar em cada estrato deverá ser proporcional à raiz quadrada de (A_h / c_h) .

¹⁴A desigualdade de Schwartz afirma que:

$$\sum \alpha_h^2 \sum b_h^2 \geq \left[\sum \alpha_h b_h \right]^2 \quad \text{com a igualdade a verificar-se se e só se } \frac{b_h}{\alpha_h} = \text{cte}$$

Conhecida a restrição orçamental (1ª versão), pode determinar-se o número de indivíduos a observar em cada estrato e a variância induzida (óptima):

$$n_h = \frac{(C - c_0)(A_h c_h^{-1})^{1/2}}{\sum_{h=1}^H (A_h c_h)^{1/2}} \Rightarrow V_{opt} = \frac{1}{C - c_0} \left[\sum_{h=1}^H (A_h c_h)^{1/2} \right]^2 + B \quad (2.20 A)$$

Conhecida a variância desejada (2ª versão), pode determinar-se o número de indivíduos a observar em cada estrato e o custo induzido (óptimo):

$$n_h = \frac{(A_h c_h^{-1})^{1/2}}{V - B} \left[\sum_{h=1}^H (A_h c_h)^{1/2} \right] \Rightarrow C_{opt} = c_0 + \frac{1}{V - B} \left[\sum_{h=1}^H (A_h c_h)^{1/2} \right]^2 \quad (2.20 B)$$

Evidenciou-se a relação entre orçamentos e precisão a atingir. Relacionando (2.19) e (2.20), ficam determinadas as amostras para todas as combinações óptimas de orçamentos e variâncias:

$$n_h = \sqrt{\frac{C - c_0}{V - B}} (A_h c_h^{-1})^{1/2} \quad (2.21)$$

donde se podem tirar as relações:

$$V = \frac{\bar{C} - c_0}{n_h^2} (A_h c_h^{-1}) + B \quad \text{se } \bar{C} - \text{dado} \quad (2.22 A)$$

$$C = \frac{(\bar{V} - B) n_h^2}{(A_h c_h^{-1})} + c_0 \quad \text{se } \bar{V} - \text{dado} \quad (2.22 B)$$

Estas relações permitem determinar o impacto no orçamento (ou na precisão) do crescimento da amostra no estrato h , admitindo que pretendemos permanecer nas combinações óptimas de precisão e orçamento.

O problema de qualquer repartição óptima⁴⁵ reside no facto de os parâmetros A_h serem desconhecidos, uma vez que dependem da variável de interesse. A implementação⁴⁶ em geral assenta no uso de uma variável auxiliar (x), supostamente correlacionada com a variável de interesse (y).

Uma alternativa à repartição óptima, sempre realizável, é a **repartição proporcional** à dimensão dos estratos, isto é:

$$n_h = n \frac{N_h}{\sum_{h=1}^H N_h} \quad (2.23)$$

A **Repartição Proporcional** à dimensão só é óptima "à Neyman" no caso de a variância ser idêntica em todos os estratos.

Importa, neste caso (sempre realizável), avaliar a opção por um desenho STSI por contraposição com um desenho SI. Admitindo idênticas probabilidades de inclusão de 1ª ordem para qualquer dos desenhos em apreciação, a questão fica resolvida pela avaliação do efeito na variância.

$$V_{SI}(\hat{t}_\pi) - V_{STSI}(\hat{t}_\pi) = \frac{N^3}{N-1} \frac{N-n}{Nn} \left[\sum_{h=1}^H W_h (\bar{y}_{U_h} - \bar{y}_U)^2 - \frac{1}{N} \sum_{h=1}^H (1-W_h) S_{yU_h}^2 \right] \quad (2.24)$$

⁴⁵ O problema da repartição óptima deve a sua primeira formulação a Neyman (1934), que pretendia proceder à repartição da amostra de modo a minimizar a Variância. Essa formulação, conhecida como **repartição de Neyman**, e matematicamente equivalente à consideração de um custo uniforme para todos os estratos, isto é, $c_h=c$. Assim, temos:

$$n_h = n \frac{N_h S_{yU_h}}{\sum_{h=1}^H N_h S_{yU_h}}$$

⁴⁶ A variável de interesse (y) pode relacionar-se de formas muito diversas com uma variável auxiliar disponível (x). É habitual o uso destas relações para ultrapassar as dificuldades enunciadas. Por exemplo, para a Repartição de Neyman, de acordo com hipóteses formuladas (e sem preocupações de rigor formal), temos:

Repartição x- Óptima: (realizável)

$$y_k = \alpha x_k \Rightarrow n_h = n \frac{N_h S_{xU_h}}{\sum_{h=1}^H N_h S_{xU_h}}$$

Repartição y- Proporcional: (não-realizável)

$$cv_{yU_h} = \frac{S_{yU_h}}{\bar{y}_{U_h}} = cte \Rightarrow n_h = n \frac{\sum_{k \in U_h} y_k}{\sum_{k \in U} y_k}$$

Repartição x- Proporcional: (realizável)

$$y_k = \alpha x_k \quad e \quad cv_{yU_h} = \frac{S_{yU_h}}{\bar{y}_{U_h}} = cte \Rightarrow n_h = n \frac{\sum_{k \in U_h} x_k}{\sum_{k \in U} x_k}$$

com
$$W_h = \frac{N_h}{N}$$

Esta equação alerta-nos para a possibilidade de a repartição proporcional não conduzir a qualquer ganho de precisão, se $y_{U_h} = y_U \sqrt{W_h}$. Contudo, se a estratificação assentar nalgum “bom senso”, esta possibilidade é, na prática, remota (o que é reforçado pelo facto de a parcela negativa decrescer com a dimensão da população N).

Fica assim colocada a questão de se saber em que circunstâncias se aconselha o uso da estratificação, sobretudo nos casos em que os desenhos de amostragem nos estratos são de probabilidades iguais. A análise da soma dos resíduos quadrados é bastante esclarecedora:

$$SST = SSB + SSW$$

$$(N-1)S_{yU}^2 = \sum_{h=1}^H N_h (\bar{y}_{U_h} - \bar{y}_U)^2 + \sum_{h=1}^H (N_h - 1) S_{yU_h}^2 \quad (2.25)$$

Se SSB , isto é, se a variação inter-estrato absorver a maior parte da variação global SST , então a estratificação proporcional garante um considerável ganho de precisão. No limite, isto é, se for absoluta a homogeneidade no interior dos estratos ($S_{yU_h}=0$), resultará um estimador de variância nula. Em suma, a estratificação usada como processo para reduzir a variância do estimador é aconselhada nos casos em que se presume uma considerável homogeneidade no interior de cada estrato. É interessante notar que SSW , assim como a eficiência dos desenhos STSI ou STBE, depende dos valores de S_{yU_h} .

3. GRUPOS

Na constituição de grupos, ao contrário do que acontece na estratificação, a partição da população não afecta a concepção do desenho de amostragem. A informação auxiliar disponível é usada para, através da formação de grupos homogêneos, se aumentar a eficiência do estimador da população como um todo.

A justificação do uso dos grupos pode ser efectuada por recurso, em geral, aos modelos da super-população e, em particular, aos modelos de regressão. Nestes modelos, o valor das variáveis na população finita são realizações de um processo gerador (dito da super-população). Como já se entende, a homogeneidade num grupo é a homogeneidade do seu modelo, isto é, do seu processo gerador.

O modelo da regressão, que usaremos para justificar os modelos de Grupos, assenta numa relação linear entre as variáveis auxiliares disponíveis (x) e a variável de interesse (y) a observar. O Modelo da Regressão ξ da super-população⁴⁷ pode ser definido do seguinte modo:

1. Os valores da variável de interesse a observar y_1, \dots, y_N são realizações independentes das variáveis aleatórias Y_1, \dots, Y_N .

⁴⁷ Os Modelos da Super-População, assim como os parâmetros envolvidos, são desconhecidos e funcionam como hipóteses de trabalho.

2. Existe uma relação linear entre a variável de interesse e as variáveis auxiliares a observar.

$$E_{\xi}[Y_k] = \sum_{j=1}^J \beta_j x_{jk} = \mathbf{x}'\beta \quad k = 1, \dots, N \quad (3.1a)$$

3. A variância da variável a observar pode ser expressa do seguinte modo:

$$V_{\xi}[Y_k] = \sigma_k^2 \quad k = 1, \dots, N \quad (3.1b)$$

Deste modo, assume-se que o erro de previsão, isto é, o afastamento entre o valor esperado da variável de interesse e a sua realização deve-se a perturbações aleatórias de média nula.

Como se desconhecem os valores de β , há que recorrer a um estimador, por exemplo dos mínimos quadrados. De acordo com as especificações do modelo ξ , temos então:

$$\mathbf{y} = \mathbf{X}'\beta + \mathbf{E} \Rightarrow \mathbf{B} = \underbrace{(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}}_{\mathbf{T}} \underbrace{\mathbf{X}'\Sigma^{-1}\mathbf{y}}_{\mathbf{t}} \Rightarrow E_{\xi}(\mathbf{B}) = \beta \quad (3.2a)$$

com:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \dots & \\ 0 & & \sigma_N^2 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1J} \\ \dots & & \dots \\ x_{N1} & \dots & x_{NJ} \end{bmatrix} \quad (3.2b)$$

O conhecimento da expressão de \mathbf{B} não é suficiente para proceder à sua estimação, pois são desconhecidos os valores de \mathbf{y} na População Finita. Com o desconhecimento dos valores de \mathbf{y} , há que recorrer a ferramentas de amostragem e, em particular, ao estimador de HT, para se obter um estimador de β não-enviesado⁴⁸ e realizável, isto é:

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{t} \Rightarrow \hat{\mathbf{B}} = \hat{\mathbf{T}}^{-1}\hat{\mathbf{t}} \quad (3.3a)$$

onde

$$\mathbf{T} = \mathbf{X}'\Sigma^{-1}\mathbf{X} = \sum_k \frac{\mathbf{x}_k \mathbf{x}_k'}{\sigma_k^2} \Rightarrow \hat{\mathbf{T}} = \sum_{k \in s} \frac{\mathbf{x}_k \mathbf{x}_k'}{\sigma_k^2 \pi_k} \quad (3.3b)$$

$$\mathbf{t} = \mathbf{X}'\Sigma^{-1}\mathbf{y} = \sum_k \frac{\mathbf{x}_k y_k}{\sigma_k^2} \Rightarrow \hat{\mathbf{t}} = \sum_{k \in s} \frac{\mathbf{x}_k y_k}{\sigma_k^2 \pi_k} \quad (3.3c)$$

⁴⁸ Obtivemos um estimador não-enviesado de \mathbf{T} mas não de \mathbf{T}^{-1} . Contudo, é aproximadamente (assimptoticamente) não-enviesado em amostras de dimensão média ou grande. Sempre que necessário, admite-se a não-singularidade das matrizes envolvidas.

Assim, sugere-se y_k^0 para predictor da variável de interesse na População Finita e retiram-se algumas consequências imediatas:

$$y_k^0 = \mathbf{x}'_k \mathbf{B} \quad \forall k \in U \Rightarrow E_k = y_k - y_k^0 \quad \text{e} \quad t_y = t_{y0} + \underbrace{t_E}_{\text{amostragem}} \quad (3.3d)$$

Ainda que fosse possível a estimação por amostragem de t_E , não seria a estimação de t_y , uma vez que \mathbf{B} não é calculável. Sugere-se assim a estimação de \mathbf{B} por amostragem, o que conduz ao predictor y_k^* (realizável) da variável de interesse com as consequências imediatas que se apresentam:

$$y_k^* = \mathbf{x}'_k \hat{\mathbf{B}} \quad \forall k \in U \Rightarrow e_k = y_k - y_k^* \quad \text{e} \quad t_y = t_{y^*} + \underbrace{t_e}_{\text{amostragem}} \quad (3.3e)$$

Fica claro que, com o uso do predictor y_k^* , apenas, o erro de previsão t_e será estimado por amostragem. Usando o estimador HT para a estimação de t_e , obtemos o estimador da regressão⁴⁹ de t_y :

$$\hat{t}_{y,r} = t_{y^*} + \hat{t}_{e\pi} = t_{y^*} + [\hat{t}_{y\pi} - \hat{t}_{y^*\pi}] = \hat{t}_{y\pi} + \sum_{j=1}^J \hat{B}_j [t_{x_j} - \hat{t}_{x_j\pi}] \quad (3.4)$$

A amostragem vem corrigir a previsão do modelo utilizado, acrescentando uma estimativa do seu erro⁵⁰.

Atendendo à expressão do estimador⁵¹, a sua variância (dependente da amostra) será:

$$V[\hat{t}_{y,r}] = V\left[\sum_k y_k^0 + \sum_{k \in s} g_{sk} \check{E}_k\right] = V\left[\sum_{k \in s} g_{sk} \check{E}_k\right] \quad (3.5)$$

⁴⁹ Três expressões equivalentes, de grande interesse em amostragem, são:

$$\begin{aligned} \hat{t}_{y,r} &= \hat{t}_{y\pi} + [\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}]' \hat{\mathbf{B}} = \sum_{k \in s} \check{y}_k + [\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}]' \hat{\mathbf{T}}^{-1} \sum_{k \in s} \frac{\mathbf{x}_k \check{y}_k}{\sigma_k^2} \\ \hat{t}_{y,r} &= \sum_{k \in s} \underbrace{\left[1 + [\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}]' \hat{\mathbf{T}}^{-1} \frac{\mathbf{x}_k}{\sigma_k^2}\right]}_{g_{sk}} \check{y}_k = \sum_{k \in s} g_{sk} \check{y}_k \quad y_k = y_k^0 + E_k \\ \hat{t}_{y,r} &= \sum_{k \in s} g_{sk} \check{y}_k^0 + \sum_{k \in s} g_{sk} \check{E}_k = \sum_k y_k^0 + \sum_{k \in s} g_{sk} \check{E}_k \end{aligned}$$

⁵⁰ A correcção efectuada justifica a classificação de estimador apoiado (model assisted), mas não dependente do modelo (model dependent). Mais ainda, estes estimadores são não-enviesados em termos do desenho de amostragem, embora o seu ganho de eficiência esteja dependente da qualidade do modelo. No caso de alguns modelos, como dos que iremos analisar, o erro de previsão é nulo. Com efeito, mostra-se que:

$$\sigma_k^2 = \lambda' \mathbf{x}_k \Rightarrow \hat{t}_{e\pi} = 0$$

⁵¹ Estimador baseado no modelo, mas não dependente do modelo. Com efeito, mostra-se que é um estimador não-enviesado em termos do desenho de amostragem, embora o ganho de eficiência esteja dependente da qualidade do modelo.

ou aproximadamente:

$$\hat{\mathbf{B}} \cong \mathbf{B} \Rightarrow AV[\hat{t}_{y,r}] = AV[\hat{t}_{c\pi}] = V[\hat{t}_{E\pi}] = \sum_{k,l} \Delta_{kl} \bar{E}_k \bar{E}_l \quad (3.6)$$

que poderá ser estimada através de:

$$\hat{V}[\hat{t}_{y,r}] = \hat{V}\left[\sum_{k \in s} g_{sk} \bar{E}_k\right] = \sum_{k,l \in s} \bar{\Delta}_{kl} (g_{sk} \bar{e}_k)(g_{sl} \bar{e}_l) \quad (3.7)$$

A utilização de Grupos em amostragem pode ser justificada do seguinte modo:

1. A população U é particionada - através da informação auxiliar - em G sub-populações, cada uma delas designada por grupo U_g (com $g=1, \dots, G$) de dimensão conhecida $|U_g| = N_g$.
2. A cada grupo é associado um modelo de regressão distinto. Para o caso, considere-se o Modelo do Ratio para Grupos⁵²:

$$E[y_k] = \beta_g x_k \quad V[y_k] = \sigma_g^2 x_k \quad (3.8)$$

A realização de um desenho de amostragem $p(\cdot)$ sobre a população U gera uma amostra s , que contém elementos dos vários grupos (G sub-populações), isto é:

$$s \cap U_g = s_g \Rightarrow \bigcup_{g=1}^G s_g = s \quad \text{e ainda: } |s_g| = n_{sg} \Rightarrow \sum_{g=1}^G n_{sg} = n_s \quad (3.9)$$

O Modelo do Ratio para Grupos pode ser colocado na forma de um único modelo, isto é:

$$\begin{aligned} E_{\xi}(y_k) &= \mathbf{x}'_k \beta \quad \beta = (\beta_1, \dots, \beta_g, \dots, \beta_G) \\ V_{\xi}(y_k) &= \mathbf{x}'_k \sigma^2 \quad \sigma^2 = (\sigma_{01}^2, \dots, \sigma_{0g}^2, \dots, \sigma_{0G}^2) \end{aligned} \quad (3.10)$$

com:

$$\mathbf{x}_k = (x_{1k}, \dots, x_{gk}, \dots, x_{Gk}) \quad x_{gk} = x_k \text{ se } k \in U_g \text{ e } x_{gk} = 0 \text{ se } k \notin U_g \quad (3.11)$$

⁵²Existe uma vasta possibilidade de modelos lineares para grupos, que podem ser usados em amostragem. O Modelo do Ratio para Grupos assentou na sua vasta e simples aplicação prática, mas está longe de esgotar as suas possibilidades. Em particular, note-se que o Modelo do Ratio é um modelo que usa a mesma variável auxiliar para todos os grupos, o que pode ser restritivo. Outros casos interessantes de estimadores assentes em modelos de grupos são o estimador da regressão post-estratificado e o estimador da regressão separado.

O estimador do parâmetro B_g do Modelo do Ratio para Grupos pode ser derivado do modelo de regressão geral, que conduz ao seguinte resultado:

$$\hat{B}_g = \frac{\sum_{k \in s_g} \tilde{y}_k}{\sum_{k \in s_g} \tilde{x}_k} \quad g = 1, \dots, G \quad g_{sk} = \frac{t_{xg}}{\hat{t}_{xg\pi}} \quad k \in s_g \quad (3.12)$$

O estimador do total t_y , apoiado pelo Modelo de Ratio para Grupos, será então:

$$\hat{t}_{y\pi} = \sum_{g=1}^G t_{xg} \hat{B}_g = \sum_{g=1}^G \frac{t_{xg}}{\hat{t}_{xg\pi}} \hat{t}_{yg\pi} \quad (3.13)$$

que é conhecido como estimador do **ratio post-estratificado** (ou **estimador de ratios separados**). A expressão da Variância deduz-se a partir da Variância do Estimador de Regressão Geral, fazendo as necessárias adaptações.

Um caso particular do Modelo do Ratio para Grupos é o Modelo da Média para Grupos, que se obtém usando como variável auxiliar uma variável de contagem. Assim temos:

$$x_k = 1 \quad \hat{B}_g = \frac{\hat{t}_{yg\pi}}{\hat{N}_g} = \tilde{y}_g \Rightarrow \hat{t}_{y\pi} = \sum_{g=1}^G N_g \tilde{y}_g = \sum_{g=1}^G \frac{N_g}{\hat{N}_g} \hat{t}_{yg\pi} \quad (3.14)$$

conhecido como o **estimador da post-estratificação**. As semelhanças com o estimador da estratificação são óbvias e justificam em grande medida a designação. Aliás, a comparação de uma estratégia assente num desenho SI, reforçada por um estimador da post-estratificação com outra baseada num desenho de amostragem STSI com taxa de amostragem proporcional à dimensão do estrato resulta num ganho insignificante para a eficiência do desenho STSI. Este é mais um argumento a favor da post-estratificação, para os casos em que não foi realizada a estratificação.

O uso dos Modelos de Grupos pode revelar-se como uma alternativa interessante, não apenas por recuperar “a posteriori” a estratificação não realizada, mas também porque o uso destes modelos abre uma excelente possibilidade de tirar partido da informação auxiliar disponível.

Casos há em que, mesmo quando não existe informação para realizar a partição da População é possível usar os Modelos de Grupos. Consideremos o **estimador de ratios separados**, que necessita, para a sua realização, dos seguintes totais:

$$t_{xg} \quad \hat{t}_{xg\pi} \quad \hat{t}_{yg\pi}$$

Se t_{xg} for disponibilizado por uma fonte fiável (por exemplo: fonte administrativa), então o estimador é realizável de acordo com (3.13), desde que se recolha por amostragem a variável auxiliar x , a variável de interesse y e a variável de partição (z) suficientes para a estimação de $\hat{t}_{xg\pi}$ e de $\hat{t}_{yg\pi}$.

4. DOMÍNIOS

Ao contrário da Estratificação ou de um Modelo de Grupos, cujo objectivo é o aumento da eficiência da estratégia de amostragem, a constituição de Domínios visa a produção/divulgação de informação de acordo com uma partição⁵³ solicitada pelo utilizador. A prática de inquéritos por amostragem revela que a procura de estimativas precisas se faz não apenas para a População como um todo, mas também para algumas das suas sub-populações (os Domínios).

A partição em Domínios distingue-se da Estratificação, uma vez que é posterior à realização do desenho de amostragem e distingue-se dos Grupos na medida em que não associa nenhum modelo (de regressão) a cada sub-população gerada.

Tal como acontece quando se concebem estratégias de amostragem da População, nos domínios também se pode aumentar a eficiência por recurso a desenhos de amostragem ou a modelos da super-população. Em particular, os Modelos de Grupos podem ser o suporte para a estimação em Domínios relevantes. Contudo, nestes casos estamos a sair do conceito estrito de Domínio.

Em geral, os domínios são de dimensões modestas, mas não extremamente pequenas. Só neste caso, se pode conceber estimadores eficientes para estes Domínios no quadro da Teoria da Amostragem na População Finita. A estimação em **pequenos domínios** é hoje uma área promissora de investigação mas não será aqui abordada.

A formação dos Domínios pode ser caracterizada do seguinte modo:

1. A população U é particionada - por recurso a condições lógicas - em D sub-populações, cada uma delas designada por domínio U_d (com $d=1, \dots, D$) de dimensão $|U_d| = N_d$, eventualmente desconhecida.
2. Designa-se por dimensão relativa do domínio a quantidade P_d ($Q_d = 1 - P_d$) :

$$P_d = \frac{N_d}{N} \quad \text{com } N_d \text{ eventualmente desconhecido.} \quad (4.1)$$

3. A realização do desenho de amostragem $p(\cdot)$ sobre a população U gera uma amostra s , que contém elementos dos vários Domínios (D sub-populações), isto é:

$$s \cap U_d = s_d \Rightarrow \bigcup_{d=1}^D s_d = s \quad \text{e} \quad |s_d| = n_{sd} \Rightarrow \sum_{d=1}^D n_{sd} = n_s \quad (4.2)$$

onde n_{sd} é aleatório e eventualmente pequeno.

O objectivo é estimar um dos seguintes parâmetros:

$$t_d = \sum_{k \in U_d} y_k \quad \bar{y}_{U_d} = \frac{t_d}{N_d} \quad (4.3)$$

⁵³ Em muitos casos a partição desejada apenas é conhecida em compreensão, isto é, através regras que lhe dão origem.

Um Domínio pode classificar-se (Purcell e Kish - 1979) de acordo com a sua dimensão relativa em Maior ($P_d \geq 0.1$), Menor ($0.01 \leq P_d \leq 0.1$), Mini ($0.0001 \leq P_d \leq 0.01$) e Raro ($P_d \leq 0.0001$). As técnicas abordadas colocam de fora o tratamento de Domínios Minis ou Raros⁵⁴.

Em geral, as dimensões dos domínios são muito variáveis fazendo com que estimativas muito precisas para certos domínios coexistam com estimativas pouco precisas para outros. Havendo necessidade de estimativas para esses domínios, é recomendável uma amostragem em duas fases. A segunda fase, realizada sobre os indivíduos que não figuram na amostra, visa reforçar a amostra em domínios mal cobertos na primeira fase.

A estimação em Domínios, de acordo com o estimador de HT, exige a definição da variável:

$$y_{dk} = \begin{cases} y_k & k \in U_d \\ 0 & k \notin U_d \end{cases} \Rightarrow t_d = \sum_k y_{dk} \Rightarrow \hat{t}_{d\pi} = \sum_k \check{y}_{dk} = \sum_{k \in U_d} \check{y}_k \quad (4.4)$$

cuja variância é:

$$V(\hat{t}_{d\pi}) = \sum_{k,l \in U_d} \Delta_{kl} \check{y}_k \check{y}_l \quad \text{a estimar através de } \hat{V}(\hat{t}_{d\pi}) = \sum_{k,l \in s_d} \check{\Delta}_{kl} \check{y}_k \check{y}_l \quad (4.5)$$

Para se obter o estimador da média, necessitamos de conhecer N_d . Fazendo $y_k = 1$, torna-se possível a sua estimação, isto é:

$$\hat{N}_d = \sum_{k \in U_d} \frac{1}{\pi_k} \quad (4.6)$$

Considerando (4.6), torna-se simples a apresentação de um estimador aproximadamente não enviesado para a média:

$$\hat{y}_{U_d} = \check{y}_{s_d} = \frac{\hat{t}_{d\pi}}{\hat{N}_d} \quad (4.7)$$

⁵⁴ Técnicas especiais são usadas para a estimação em domínios raros ou minis, onde pode dar-se o caso de não existirem observações na amostra. Em particular, as técnicas referidas usam estimadores que assentam a sua validade na adopção de Modelos (válidos) da Super-População.

Nos casos em que N_d é conhecido, pode-se usar um estimador preferível para t (que corrige a dimensão do Domínio⁵⁵):

$$\tilde{t}_d = N_d \frac{\hat{t}_{d\pi}}{\hat{N}_d} \quad (4.8)$$

cuja variância é:

$$AV(\tilde{t}_d) = \sum_{k,l \in U_d} \Delta_{kl} \left(\frac{y_k - \bar{y}_{U_d}}{\pi_k} \right) \left(\frac{y_l - \bar{y}_{U_d}}{\pi_l} \right) \quad (4.9)$$

a estimar através de:

$$\hat{V}(\tilde{t}_d) = \left(\frac{N_d}{\hat{N}_d} \right)^2 \sum_{k,l \in s_d} \tilde{\Delta}_{kl} \left(\frac{y_k - \bar{y}_{s_d}}{\pi_k} \right) \left(\frac{y_l - \bar{y}_{s_d}}{\pi_l} \right) \quad (4.10)$$

Embora este último estimador apresente, em geral, uma maior eficiência superior, é no tratamento de pequenos domínios que as diferenças se acentuam e o seu uso se torna fundamental. Por exemplo, se o desenho de amostragem for SI, o Domínio for pequeno ($P_d \rightarrow 0$) e o coeficiente de variação da variável (cv_{yU_d}) igual a 0.5 (elevada dispersão), a variância do estimador π (4.4) é 5 vezes maior⁵⁵ do que a do estimador proposto (4.8). De facto, o controle da dimensão da amostra é um problema de grande relevância quando se pretende produzir estimativas para pequenos domínios.

Os conceitos de Estratos, Grupos e Domínios desempenham papéis distintos mas não se excluem, isto é, podem contribuir simultaneamente para a realização de uma estratégia de amostragem. Exemplificando, pode pretender-se um estimador para um dado domínio no contexto de um desenho de amostragem estratificado. Deriva-se dos resultados apresentados:

⁵⁵ Este estimador pode ser justificado a partir do Modelo do Ratio. No caso do desenho SI, temos, para os estimadores apresentados:

$$\hat{t}_{d\pi} = \frac{N}{n} \sum_{k \in U_d} y_k \quad \tilde{t}_d = \frac{N_d}{n_{sd}} \sum_{k \in U_d} y_k$$

cujas variâncias são:

$$V_{SI}(\hat{t}_{d\pi}) \cong N^2 \frac{1-f}{n} P_d (S_{yU_d}^2 + Q_d \bar{y}_{U_d}^2) \quad AV_{SI}(\tilde{t}_d) \cong N^2 \frac{1-f}{n} P_d S_{yU_d}^2$$

$$S_{yU_d}^2 = \frac{1}{N_d - 1} \sum_k (y_k - \bar{y}_{U_d})^2$$

⁶⁰ Para o desenho SI, mostra-se que (ver nota anterior):

$$\frac{V(\hat{t}_{d\pi})}{AV(\tilde{t}_d)} \cong 1 + \frac{Q_d}{(cv_{yU_d})^2} \quad cv_{yU_d} = \frac{S_{yU_d}}{\bar{y}_{U_d}}$$

o que nos conduz ao resultado apresentado. Controlar a dimensão do domínio "a priori" (na etapa do desenho) ou "a posteriori" (na etapa da estimação, corrigindo o estimador π) são procedimentos importantes para aumentar a eficiência da estratégia.

$$\hat{t}_{d\pi} = \sum_{h=1}^H \frac{N_h}{n_h} \hat{t}_{dh\pi} \quad \hat{t}_{dh\pi} = \sum_{k \in s_d} \tilde{y}_{dk} = \sum_{k \in s_{dh}} \tilde{y}_k \quad s_{dh} = s \cap U_d \cap U_h \quad (4.11)$$

cuja variância será dada por:

$$V(\hat{t}_{d\pi}) = \sum_{h=1}^H \left(\frac{N_h}{n_h} \right)^2 V(\hat{t}_{dh\pi}) \quad (4.12)$$

Descreveu-se na introdução o problema da “transição de estratos” ou, com mais rigor, com a necessidade de produzir a informação para domínios (“novos estratos”) que se formaram num momento posterior à realização do desenho de amostragem (estratificado). É simples reconhecer neste estimador (4.11) a capacidade para tratar as chamadas “transições de estratos”. Por outro lado, torna-se óbvio, que a produção de estimativas para domínios⁵⁷ não contemplados no desenho de amostragem é, em muitos casos, possível. O exemplo clássico é o da produção de estimativas regionais.

Um outro exemplo da vantagem da combinação das várias técnicas apresentadas reside na possibilidade de uso dos modelos de Grupo⁵⁸ para a produção de estimativas para os Domínios relevantes.

Como o modelo é formulado para uma dada população (ou grupo) a sua estimação considera toda a informação útil, isto é, a informação relativa a todos os elementos da população (ou grupo) pertençam estes ou não ao domínio de interesse.

O estimador será então a soma da previsão para o domínio t_{dy^*} , efectuada de acordo com o modelo, acrescida de uma parcela que visa estimar o erro de previsão t_{de} , isto é:

$$\hat{t}_{dy,r} = t_{dy^*} + \hat{t}_{de\pi} \quad \text{se } N_d \text{ for desconhecido} \quad (4.13)$$

O estimador pode ser melhorado corrigindo a estimativa do erro de previsão de acordo com a dimensão conhecida do domínio (N_d), isto é:

$$\hat{t}'_{dy,r} = t_{dy^*} + \frac{N_d}{\hat{N}_d} \hat{t}_{de\pi} \quad (4.14)$$

que é um estimador mais interessante em pequenos domínios, uma vez que - como já referimos - efectua a recomendável correcção da dimensão. Um inconveniente deste estimador é a possibilidade de inconsistência entre as estimativas dos diferentes

⁵⁷ É óbvio que a solução ideal - em termos de eficiência - faria coincidir os domínios e os estratos. Contudo, essa solução tem, por vezes, um custo incomportável. A questão deriva, em grande medida, da dimensão da amostra no domínio. A questão perde a sua gravidade, nos casos em que a dimensão relativa do domínio P_d seja tal que assegure com elevada probabilidade uma dimensão razoável da amostra no domínio.

⁵⁸ Note-se que outros Modelos de Regressão podem ser usados para auxiliar a estimação em Domínios.

domínios e a estimativa da população. Trata-se, por vezes, de uma opção difícil entre a eficiência e a consistência na agregação⁵⁹.

A variância do estimador⁶⁰ será, então:

$$AV(\hat{t}'_{dy,r}) = \sum_{k,l \in U_d} \Delta_{kl} \left(\frac{E_k - \bar{E}_{U_d}}{\pi_k} \right) \left(\frac{E_l - \bar{E}_{U_d}}{\pi_l} \right) \quad (4.15)$$

a estimar através de:

$$\hat{V}[\hat{t}'_{dy,r}] = \hat{V} \left[\sum_{k \in s} g_{dsk} \bar{E}_k \right] = \sum_{k,l \in s} \tilde{\Delta}_{kl} (g_{dsk} \bar{e}_k) (g_{dsl} \bar{e}_l) \quad (4.16)$$

A estimação em domínios com o apoio de um Modelo para Grupos, por exemplo⁶¹ o Modelo da Média para Grupos, pode contribuir para ganhos consideráveis na estimação.

Neste caso, teríamos:

$$\hat{t}'_{dr} = \sum_{g=1}^G N_{dg} \tilde{y}_{\cdot g} + \frac{N_{d\cdot}}{\hat{N}_{d\cdot}} \sum_{g=1}^G \hat{N}_{dg} (\tilde{y}_{dg} - \tilde{y}_{\cdot g}) \quad (4.17)$$

$$g_{dsk} = N_{d\cdot} \left[\frac{z_{dk}}{\hat{N}_{d\cdot}} + \left(\frac{N_{dg}}{N_{d\cdot}} - \frac{\hat{N}_{dg}}{\hat{N}_{d\cdot}} \right) \frac{1}{\hat{N}_{\cdot g}} \right] \quad (4.18)$$

Se os grupos realizarem uma partição do domínio, teremos o estimador da Post-Estratificação para Domínios, isto é:

⁵⁹ A consistência na agregação pretende que da agregação (grosseiramente: soma) das estimativas dos domínios resulte a estimativa da população. Alguns autores atribuem grande importância à referida propriedade.

⁶⁰ A correcção da dimensão efectuada ao estimador da regressão conduz a alterações dos pesos g , isto é:

$$\hat{t}'_{y,dr} = \frac{N_d}{\hat{N}_d} \hat{t}'_{dy,r} + \left[\mathbf{t}_{dx} - \frac{N_d}{\hat{N}_d} \hat{t}'_{dx,r} \right]' \hat{\mathbf{B}} = z_{dk} = \begin{cases} 1 & k \in Ud \\ 0 & k \notin Ud \end{cases}$$

$$\hat{t}'_{y,dr} = \sum_{k \in s} \underbrace{\left[\frac{N_d}{\hat{N}_d} z_{dk} + \left[\mathbf{t}_{dx} - \frac{N_d}{\hat{N}_d} \hat{t}'_{dx,r} \right]' \hat{\mathbf{T}}^{-1} \frac{\mathbf{x}_k}{\sigma_k^2} \right]}_{g_{dsk}} \tilde{y}_k = \sum_{k \in s} g_{dsk} \tilde{y}_k \quad y_k = y_k^0 + E_k$$

Note-se que o somatório faz-se em s .

⁶¹ Outros Modelos para Grupos podem ser usados, como ficou claro do ponto 3, após a realização das modificações devidas.

$$\hat{t}'_{dr} = \sum_{g=1}^G N_{dg} \tilde{y}_{sdg} \quad (4.19)$$

Embora à primeira vista, seja atractiva a ideia de realizar a partição dos Domínios em Grupos, não se devem esquecer algumas dificuldades. Trata-se de usar modelos muito finos, com as vantagens de um tratamento diferenciado mas cuja estimação exige mais informação. A prática mostra que, em geral, há muitos mais domínios que grupos com capacidade explicativa. Por exemplo, Särndal cita os valores $D=400$ e $G=8$, o que conduziria - em caso de partição dos domínios - a um número elevadíssimo de grupos derivados ($D \times G = 3200$), com o risco acrescido de o desenho de amostragem ter criado Grupos vazios.

Um exemplo apontado para o uso combinado de Grupos e Domínios refere-se à produção de estatísticas regionais. Seguramente que a Região é um Domínio sobre o qual se pretende a informação, por exemplo, sobre os rendimentos. Os Grupos servem para formular modelos, isto é, relacionam a variável de interesse com as variáveis explicativas. Neste caso, o modelo pode dizer que o rendimento é função das variáveis idade/sexo. Assim, os rendimentos das regiões podem ser mais bem percebidos se atendermos à composição da população, em termos de idade e sexo. Contudo, pode ser inadequado - por motivos teóricos e práticos - tomar a região como factor explicativo.

5. CONCLUSÕES

Embora estratos, grupos e domínios realizem a partição da População, os seus objectivos são distintos. A Estratificação visa ajudar a conceber desenhos de amostragem mais eficientes. A formação de Grupos, apoiada por Modelos da Super-População adequados, visa o uso criterioso da informação auxiliar e, desse modo, contribui para uma estimação mais eficiente. Os Domínios reflectem apenas pedidos de informação desagregada a considerar. A percepção das diferenças que envolvem estes conceitos é fundamental para identificar a natureza dos problemas que surgem numa estratégia de amostragem e abrem o caminho para a sua resolução.

REFERÊNCIAS BIBLIOGRÁFICAS

- CALDEIRA, O. (1995). Princípios de Inferência em Amostragem Probabilística. *UNL-ISEGI: Lisboa*.
- HANSEN, M.H. HURWITZ, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics* 14.
- HEDAYAT, A.S. SINHA, B.K. (1991). Finite Population Sampling. *New York: Wiley-Interscience*.
- HOLT, D. SMITH, T.M.F. (1979). Post-Stratification. *Journal of the Royal Statistical Society A* 142.
- ISAKI, C.T. FULLER, W.A. (1982). Survey Design under the regression superpopulation model. *Journal of the American Statistical Association* 77.
- NEYMAN, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97.

- PLATEK, R., SINGH, M., RAO, J.N.K., SARNDAL, C.E. (1987). Small Area Statistics: An International Symposium. *New York: Wiley-Interscience.*
- PURCELL, N.J., KISH, L. (1979). Estimation for small domains. *Biometrics* 35.
- SARNDAL, C.E. (1982). Implication of survey design for the regression estimation of linear functions *Journal of Statistical Planning and Inference* 7.
- SARNDAL, C.E. , HIDIRIGLOU, M.A. (1989). Small Domain Estimation: a conditional analysis *Journal of the American Statistical Association* 84.
- SARNDAL, C.E. , SWENSON, B. e WRETMAN, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika* 76.
- SARNDAL, C.E., SWENSON, B. e WRETMAN, J. (1992). Model assisted survey sampling. *New York: Springer-Verlag.*
- TSCHUPROW, A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron* 2.
- WRIGHT, R.L. (1983). Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association* 78.

AS ESTATÍSTICAS AGRÍCOLAS EUROPEIAS NO
HORIZONTE 2005

Autor:
António de Sousa Macedo

THE EUROPEAN AGRICULTURAL STATISTICS IN 2005

Autor: António de Sousa Macedo
Director do Departamento de Estatísticas da Agricultura e Pescas do
Instituto Nacional de Estatística

RESUMO:

- É objectivo deste artigo reflectir sobre as estatísticas agrícolas europeias e quais os seus possíveis desenvolvimentos no futuro. Apresentam-se alguns factores que vão condicionar a evolução do actual sistema de estatísticas agrícolas e o seu impacto em aspectos essenciais em que se baseia o actual sistema. Finalmente, indicam-se algumas das prováveis principais mudanças que o sistema poderá incorporar.

PALAVRAS-CHAVE:

- *Estatísticas agrícolas europeias; política agrícola comum; dados administrativos; campo de observação e unidade estatística; metodologias estatísticas; harmonização das estatísticas agrícolas.*

ABSTRACT:

- The purpose of this article is to reflect on the development of agricultural statistics and its future. We will introduce some aspects that will restrict the evolution of the present agricultural statistics system and the impact it will have on the different sides the present system is based on now. At last, we will outpoint some of the main possible changes that the system may sustain.

KEY-WORDS:

- *European agricultural statistic; common agricultural policy; administrative data; field of observation and statistical unit; statistical methodology; harmonisation on agricultural statistics.*

1 INTRODUÇÃO

A Conferência dos Estados-membros sobre a política agrícola realizada em Stresa em 1958 serviu de base à Comissão para propôr a Política Agrícola Comum (PAC) da Comunidade.

Em 1962 foram adoptadas as primeiras regulamentações de mercados e em 1968 a política de harmonização dos preços agrícolas.

A gestão de mercados agrícolas impôs, desde logo, a necessidade de a Comunidade dispôr de um sistema de estatísticas agrícolas europeias harmonizado, que permitisse fundamentar as políticas e avaliar as suas consequências.

Por isso, os primeiros dados a obter foram os relativos a preços agrícolas, às produções das culturas, aos efectivos animais e à estrutura das explorações agrícolas, desenvolvendo-se o sistema à medida que eram implementadas as diferentes políticas sectoriais.

O primeiro inquérito à estrutura das explorações agrícolas realizou-se em 1966. A partir do início dos anos 70 estabeleceram-se os Regulamentos e as Directivas sobre o leite e produtos lácteos, os efectivos animais e os abates, os animais de capoeira e os ovos, bem como "gentlemen agreements" sobre as estatísticas da produção vegetal, os balanços de aprovisionamento, as contas da agricultura e os preços agrícolas.

Na segunda metade da década de 70 juntaram-se as estatísticas relativas à vinha e as estimativas precoces da evolução do rendimento agrícola, para além do estabelecimento de uma tipologia comunitária das explorações agrícolas.

Os anos 80 foram principalmente de consolidação do sistema, tendo surgido no início da década de 90 os Regulamentos sobre as produções vegetais, em resultado da necessidade de harmonizar e obter dados fiáveis e em tempo oportuno, para melhor seguir as políticas de combate aos excedentes agrícolas entretanto postas em execução.

Assim, o essencial do programa de estatísticas agrícolas foi constituído numa comunidade de 6 países, apenas com alguns ajustamentos decorrentes das sucessivas adesões de outros Estados-membros.

Nos últimos anos, contudo, diversos factores conduziram a alterações em alguns aspectos do programa. O alargamento da Comunidade Europeia, as exigências da nova PAC de 1992 e as restrições financeiras dos Serviços de Estatística dos Estados-membros determinaram a necessidade de racionalizar e flexibilizar o programa estatístico.

2- CONDICIONANTES FUTURAS DO SISTEMA DE ESTATÍSTICAS AGRÍCOLAS

2.1- POLÍTICA AGRÍCOLA COMUM (PAC)

O sistema de estatísticas agrícolas actual já incorpora alguns aspectos que resultaram das alterações introduzidas pela PAC de 1992 e de condicionantes externas, nomeadamente as negociações no âmbito da Organização Mundial do Comércio (OCM).

A nova política agrícola comum originou modificações importantes na agricultura, alterando substancialmente a realidade agrícola. A redução dos preços agrícolas, acompanhada por ajudas directas ao rendimento, tornou os agricultores menos dependentes da produção dos produtos agrícolas.

Po outro lado, os agricultores, com a introdução das medidas agro-ambientais, são chamados a utilizar práticas agrícolas que preservam e melhoram o ambiente, alargando o campo de observação do fenómeno agrícola.

As novas políticas agrícolas e de desenvolvimento rural reforçarão a procura pelos agricultores de novas actividades ligadas à exploração agrícola, nomeadamente o agro-turismo, o lazer e a transformação dos produtos agrícolas na exploração, das quais os agricultores obterão um complemento de rendimento.

O agricultor será, assim, um factor de equilíbrio e de preservação do território.

2.2 - ADESÃO À UNIÃO EUROPEIA DOS PAÍSES DA EUROPA CENTRAL E ORIENTAL (PECO)

A adesão dos PECO à União Europeia terá um impacto substancial nas estatísticas agrícolas, desde logo pelo facto da coexistência de mais de 20 países com sistemas estatísticos com características muito heterogéneas e agriculturas diversificadas. Os sistemas de estatísticas agrícolas destes países estavam orientados para uma agricultura planificada e colectivizada, o que coloca sérios problemas à construção de um novo sistema, que tem de responder a um extenso programa comunitário. Esta nova realidade vai colocar maiores dificuldades à manutenção de estatísticas agrícolas harmonizadas e à disponibilização em tempo oportuno dos dados europeus.

A coordenação do sistema europeu será para o Serviço de Estatísticas da União Europeia (EUROSTAT) um desafio de grande envergadura.

2.3 - DADOS ADMINISTRATIVOS

Alguns países, nomeadamente do norte da Europa têm uma tradição forte na utilização de fontes administrativas.

A implementação da nova PAC tem, igualmente, fomentando o aparecimento de dados administrativos em todos os Estados-membros.

Há, por isso, uma forte pressão no sentido da utilização destes dados, o que põe em causa os procedimentos estatísticos actualmente utilizados e que baseavam a realização dos inquéritos estatísticos comunitários na definição de metodologias comuns obrigatórias para todos os Estados-membros.

Esta pressão resulta não só da existência de sistemas estatísticos nacionais que baseiam fortemente nas fontes administrativas a recolha de informação agrícola, mas também a cada vez maior diminuição dos recursos financeiros atribuídos às estatísticas agrícolas e à necessidade de reduzir a sobrecarga de inquirição que actualmente incide sobre os agricultores.

2.4 - NEGOCIAÇÕES NO ÂMBITO DO ORGANIZAÇÃO MUNDIAL DE COMÉRCIO

Cada vez mais a União Europeia será confrontada no futuro com a necessidade de dispôr de informação que lhe permita negociar com eficácia com os outros países, quer a nível bilateral, quer nas organizações internacionais de comércio.

O conhecimento dos dados estatísticos não se pode limitar apenas aos dados comunitários, mas são igualmente da maior importância os dados relativos aos países concorrentes, para a preparação da correcta defesa dos interesses da agricultura da União Europeia e do bem estar dos agricultores e das populações rurais.

3 - IMPACTO NAS ESTATÍSTICAS AGRÍCOLAS

3.1 - CAMPO DE OBSERVAÇÃO E UNIDADE ESTATÍSTICA

Tradicionalmente as estatísticas agrícolas têm como unidade estatística a “exploração agrícola”, definida como uma unidade de produção que produz produtos vegetais e animais a partir de determinados limites mínimos. O “produto agrícola” é assim o elemento principal da definição.

No entanto, as futuras políticas ligadas ao desenvolvimento rural pretendem dar à agricultura um papel importante na manutenção do espaço rural, preservando-o e possibilitando a continuação da ocupação humana do território. Por isso, a “unidade agrícola” verá ampliadas as suas funções, que deixarão de ser apenas de produção, e que se vão reflectir em novas modalidades de organização de produção e oferta de produtos para o mercado, no sistema económico e social das famílias agrícolas, no reforço do trabalho a tempo parcial, no trabalho ao domicílio, nas actividades de transformação dos produtos agrícolas na exploração e em actividades exteriores por conta própria ou de outrém.

Neste contexto é de questionar se a actual definição de exploração agrícola se adequa às novas realidades.

Também a definição do actual campo de observação das estatísticas agrícolas poderá ficar desajustado face às novas funções atribuídas à agricultura.

No futuro as explorações agrícolas verão aumentadas as suas ligações com as actividades de outros sectores económicos, bem como o aparecimento de novas actividades ligadas à utilização do território, o que coloca a questão do eventual alargamento do campo das estatísticas agrícolas.

Integrar nos objectivos das estatísticas agrícolas a obtenção de informação sobre a floresta, a caça, o ambiente, a agro-indústria, o artesanato, os rendimentos exteriores à actividade agrícola parece ser indispensável no futuro.

A consideração de todas estas diferentes actividades ligadas à exploração agrícola coloca em discussão o próprio conceito de agricultura, dado ser a existência destas actividades que poderá permitir a continuação da actividade agrícola.

3.2 - METODOLOGIAS ESTATÍSTICAS

O alargamento do campo de observação, uma União alargada de países com especificidades agrícolas diversas e recursos diferentes, a melhoria dos novos instrumentos estatísticos e a existência de cada vez mais informação administrativa e melhor organizada, vão condicionar a definição das metodologias e dos métodos de recolha da informação agrícola.

As mudanças mais rápidas na estrutura das explorações agrícolas e a escassez de meios financeiros dificultarão a manutenção de um ficheiro de explorações agrícolas actualizado, o qual tem sido a base mais utilizada para a realização dos inquéritos agrícolas de carácter específico por amostragem.

Nos países candidatos à adesão à União Europeia esta situação é ainda mais delicada, devido às enormes modificações que se vêm verificando na titularidade da posse das terras e na reorganização fundiária, e que seguramente continuará ainda por muito tempo.

A política agrícola comum ao criar a necessidade de os agricultores preencherem múltiplos formulários relativos à exploração agrícola proporciona o aumento de actos administrativos e vai continuar a pressionar os estatísticos para a utilização da informação administrativa. E, se a sua análise tendo em vista uma eventual incorporação nas estatísticas oficiais é natural, já a opção de a usar como alternativa aos actuais inquéritos agrícolas coloca alguns problemas, em virtude das diferenças a nível do campo de observação, das definições ou da sua representatividade.

A utilização da informação administrativa é acima de tudo particularmente importante pelo impacto que tem na actual filosofia do sistema estatísticas agrícolas, que baseia, em grande medida, a comparabilidade das estatísticas entre os Estados-membros na utilização de metodologias comuns.

O reforço do interesse em dispôr de informação ligada à ocupação/utilização das terras e a necessidade de reduzir os inquéritos dirigidos aos agricultores, vai dar maior espaço ao uso de métodos de recolha de informação baseados na fotografia aérea e nas imagens de satélite.

3.3 - HARMONIZAÇÃO DAS ESTATÍSTICAS AGRÍCOLAS

A harmonização das estatísticas agrícolas é uma das principais bases do sistema europeu, tendo sido obtida recorrendo a discussões aprofundadas das metodologias nos Grupos de Trabalho do Comité de Estatísticas Agrícolas e que culminaram em actos jurídicos ou em "gentlemen agreements".

Se numa Comunidade de 6 países as exaustivas e demoradas discussões técnicas permitiram uma harmonização dos métodos e das definições, atingir este objectivo tornou-se cada vez mais difícil à medida que se efectuaram os sucessivos alargamentos até aos actuais 15 países.

A utilização de novas tecnologias estatísticas, o aumento da informação administrativa, a nova realidade agrícola e a escassez de recursos financeiros e humanos, tornará difícil pôr de acordo os Estados-membros na definição de actos jurídicos que imponham a todos as mesmas metodologias.

Dentro de alguns anos, com mais de 20 países, o objectivo de harmonizar utilizando metodologias comuns será uma tarefa complicada, tanto mais que o nível de discussão técnica nos Grupos de Trabalho terá tendência a diminuir.

Para além do objectivo de harmonização das estatísticas, com a adopção de metodologias comuns pretendia-se, igualmente, garantir a qualidade da informação. Assim, uma nova orientação no sentido de harmonizar em função dos objectivos coloca a questão de como manter a qualidade das estatísticas agrícolas europeias.

3.4 - DIFUSÃO DA INFORMAÇÃO

As estatísticas agrícolas são cada vez mais um instrumento para as políticas publicas, em que a qualidade e a oportunidade contribuem para a eficácia da gestão, complexa e exigente, das políticas agrícolas europeias. São, igualmente, indispensáveis a todos os profissionais ligados directa ou indirectamente à agricultura e ao desenvolvimento rural.

Mas não só a informação produzida pela União Europeia. São, também, do maior interesse os dados relativos à agricultura dos países exteriores, nomeadamente os Estados Unidos da América e os outros países concorrentes no mercado mundial de produtos agrícolas.

O enorme avanço tecnológico dos meios de difusão da informação deve ser aproveitado para que a informação agrícola europeia e dos países nossos concorrentes seja colocada, de forma organizada e em tempo oportuno, à disposição de todos os órgãos decisores dos Estados-membros, dos profissionais agrícolas e da sociedade civil.

É, por isso, essencial que todos os Estados-membros tenham a possibilidade de acederem à informação, de modo a estarem em igualdade na discussão das políticas agrícolas comunitárias, evitando-se que os países com recursos financeiros mais escassos estejam em desvantagem ao não a poderem obter com os seus próprios meios.

3.5 - MISSÕES DO EUROSTAT

Para o EUROSTAT poder cumprir no futuro a sua missão de garantir a coordenação, harmonização, qualidade e difusão da informação agrícola europeia é necessário alargar a sua forma de intervenção.

Numa União Europeia com 20-25 países utilizar a definição de metodologias comuns para cumprir o objectivo de harmonizar os inquéritos agrícolas será seguramente cada vez mais difícil.

A passagem a uma harmonização por objectivos implica necessariamente um reforço da coordenação do EUROSTAT no sentido de garantir a qualidade da informação obtida pelos Estados-membros.

O EUROSTAT deverá, igualmente, encarar a possibilidade de se organizarem operações estatísticas puramente europeias, sem representatividade nacional.

A missão da difusão da informação deve, também, ser reforçada. A criação de Bases de Dados, de acesso simples e rápido, deve ser incrementada e a informação

deve ser apresentada de forma mais compreensível para os utilizadores. Além disso, a apresentação dos dados não pode cingir-se aos números. Uma informação estatística comentada é fundamental para que os utilizadores melhor a compreendam, estando os estatísticos bem colocados para proceder a essa primeira valorização dos dados.

4 - PERSPECTIVAS PARA AS ESTATÍSTICAS AGRÍCOLAS EUROPEIAS NO HORIZONTE 2005

I - O alargamento do campo de observação das estatísticas agrícolas é no futuro indispensável, bem como a revisão do conceito de "exploração agrícola", de modo a incorporar as evoluções do desenvolvimento económico nas comunidades rurais.

II - A espacialização da informação agrícola será reforçada para as estatísticas agrícolas. O recurso à realização de inquéritos areolares, com recurso à fotografia aérea e à teledeteção, e a utilização de sistemas de informação geográfica introduzirá modificações substanciais nas metodologias estatísticas.

III - A harmonização das estatísticas agrícolas europeias tenderá a fazer-se, cada vez mais, através da definição de objectivos, utilizando metodologias diferenciadas em função dos sistemas estatísticos dos Estados-membros e das características das suas agriculturas.

IV - Os inquéritos e análises de carácter económico e sociológico, que permitam avaliar o impacto que as políticas agrícolas e de desenvolvimento rural têm nas "famílias rurais" serão desenvolvidos.

V - O papel do EUROSTAT na coordenação, harmonização e difusão das estatísticas agrícolas europeias terá de ser mais activo, não colocando de parte a possibilidade de se recorrer a operações estatísticas específicas com o objectivo de obter dados apenas para a União Europeia, em alternativa ao actual sistema que se baseia no somatório dos resultados nacionais.

ACTUALIZAÇÃO DOS CADERNOS ELEITORAIS E SUAS CONSEQUÊNCIAS NOS CÍRCULOS DE APURAMENTO

Autores:
Paulo Morais
e
José António Monteiro

VOLUME 1

1º QUADRIMESTRE DE 1998

ACTUALIZAÇÃO DOS CADERNOS ELEITORAIS E SUAS
CONSEQUÊNCIAS NOS CÍRCULOS DE APURAMENTO

THE IMPLICATIONS OF AN ACTUALIZATION OF THE PORTUGUESE
ELECTORAL BOOKS

Autores: Paulo Morais .

- Universidade Lusíada do Porto

e

José António Monteiro

- DOMP - Desenvolvimento Organizacional, Marketing e Publicidade

RESUMO:

- Os cadernos eleitorais portugueses estão largamente inflacionados. Este trabalho tenta avaliar a dimensão do número de eleitores “fantasma”, bem como as causas de tal situação. A diferença entre o número de recenseados e o número de cidadãos com mais de dezoito anos (estimativas INE) é de cerca de um milhão de indivíduos; com diferente incidência nos círculos distritais. As principais causas desta situação são as duplicações de inscrição, os cidadãos já falecidos mas que ainda constam dos cadernos eleitorais e alguns emigrantes incorrectamente inscritos. A actualização em curso provocará a perda de representatividade de alguns dos círculos distritais; a nível da representação partidária na Assembleia da República, não se espera qualquer alteração significativa.

PALAVRAS-CHAVE:

- *métodos de escrutínio, recenseamento, sistemas eleitorais.*

ABSTRACT:

- The aim of this work was to predict the implications of an actualization of the electoral books. The difference between the number of registered electors and citizens with more than eighteen years (according to INE) is around one million. The main factors responsible for this situation are: the duplication of an individual registration; citizens already died and still appears in the electoral books; and some emigrants not regularly registered.

The electoral books actualisation will imply some changes on the actual electoral geography. At the level of the district circles, will result in an eventually loss of representation in some districts. There is no significant changes in the representation of each party in the Parliament.

KEY WORDS:

- *electoral systems, registration.*

1. OBJECTIVOS

Os objectivos deste estudo são os de avaliar a real dimensão do tão discutido empolamento dos cadernos eleitorais, das suas consequências a nível regional e distrital, no que concerne ao número de eleitores por círculo; bem assim como antever que consequências resultariam da actualização dos cadernos eleitorais, em termos do número de deputados por círculo. Finalmente, pretende-se antever e simular a composição da Assembleia da República neste novo enquadramento e as consequências que isso comporta em termos de representatividade das diversas forças políticas.

2. INTRODUÇÃO

As múltiplas dúvidas que vêm sendo sucessivamente levantadas pela opinião pública relativamente à fiabilidade dos cadernos eleitorais, levaram a que reflectíssemos e estudássemos este problema tentando comparar os dados do recenseamento eleitoral com os de outras bases de dados disponíveis, nomeadamente as estimativas para a população residente, da responsabilidade do Instituto Nacional de Estatística.

A primeira vez que esta questão é levantada com acuidade é aquando do censo de 1991, levado a cabo pelo INE; constata-se então haver uma discrepância de aproximadamente 600000 eleitores a mais nos cadernos eleitorais. Haveria pois que avaliar em que medida este fenómeno se iria corrigir com o evoluir dos anos ou se, pelo contrário, se agravaria esta diferença. É o que nos propomos avaliar neste estudo: em que medida são diferentes os números apresentados pelos cadernos eleitorais e os censos do INE, relativamente à população com idade de votar, feita esta comparação em 1995. Optamos por utilizar números relativos a 1995, uma vez que neste ano, para além de dispormos das estimativas de população para o território do continente e regiões autónomas que o Instituto Nacional de Estatística apresenta anualmente, decorreram eleições para a Assembleia da República. Podemos pois, neste contexto, simular as consequências (em termos de representação geográfica e partidária) duma actualização do recenseamento, utilizando para tanto quer as estimativas de população em idade de votar, quer os resultados efectivamente obtidos nas eleições.

Das fontes a que recorremos para a obtenção dos dados, relevamos o STAPE (Secretariado Técnico dos Assuntos para o Processo Eleitoral), o INE (Instituto Nacional de Estatística) e o Instituto IDEA (Institute for Democracy and Electoral Assistance). O *software* adoptado para o tratamento estatístico dos dados foi o SPSS e o EXCEL (programa desenvolvido neste ambiente pela DOMP).

3. DISCUSSÃO E RESULTADOS

Numa primeira análise, e por simples comparação dos cadernos eleitorais de 1995 com as estimativas do INE do mesmo ano, relativamente à população com mais de dezoito anos - constata-se uma diferença de cerca de um milhão de "falsos" eleitores (O Quadro I apresenta essas diferenças). Optamos neste estudo por considerar as estimativas do INE, até pelo facto de podermos trabalhar com dados geograficamente mais finos, nomeadamente ao nível do distrito. Esta variação seria tanto maior se os dados utilizados fossem os do Instituto IDEA (Institute for Democracy and Electoral Assistance), que considera para o nosso país como número

de recenseados e de população em idade de votar os números de aproximadamente 8,9 e 7,5 milhões, respectivamente, o que daria uma diferença de cerca de um milhão e meio.

QUADRO I

DISTRITO	REC. 95	ESTIM. INE 95	DIFERENÇA	DIFERENÇA RELATIVA
AVEIRO	559.718	513.990	45.728	8,17%
BEJA	152.031	129.880	22.151	14,57%
BRAGA	635.169	565.330	69.839	11,00%
BRAGANÇA	151.845	120.940	30.905	20,35%
C. BRANCO	203.039	168.940	34.099	16,79%
COIMBRA	382.941	339.340	43.601	11,39%
ÉVORA	151.251	137.270	13.981	9,24%
FARO	308.767	274.860	33.907	10,98%
GUARDA	175.602	145.910	29.692	16,91%
LEIRIA	374.841	337.170	37.671	10,05%
LISBOA	1.874.449	1.644.320	230.129	12,28%
PORTALEGRE	116.764	104.840	11.924	10,21%
PORTO	1.404.949	1.276.960	127.989	9,11%
SANTARÉM	395.852	354.760	41.092	10,38%
SETÚBAL	639.142	576.210	62.932	9,85%
V. CASTELO	225.243	194.090	31.153	13,83%
V. REAL	221.292	179.090	42.202	19,07%
VISEU	352.553	306.430	46.123	13,08%
AÇORES	188.868	170.340	18.528	9,81%
MADEIRA	205.088	188.820	16.268	7,93%
TOTAL	8.719.404	7.729.490	989.914	11,35%

Esta variação de cerca de um milhão de eleitores a mais reflecte-se naturalmente em todos os círculos, mas com maior incidência nos distritos do interior (Bragança, Castelo Branco, Guarda e Vila Real).

Considerando os dados fornecidos pelas estimativas do Instituto Nacional de Estatística relativos ao ano de 1995, a distribuição dos deputados pelos círculos (calculada pelo método de Hondt, tal como previsto no artigo 13 da lei eleitoral para a Assembleia da República - 14/79) viria a ser a que consta do Quadro II.

QUADRO II

DISTRITO	REC. 95 Nº DEP.	ESTIM. 95 Nº DEP.	DIFERENÇA
AVEIRO	14	15	1
BEJA	4	3	-1
BRAGA	16	17	1
BRAGANÇA	4	3	-1
C. BRANCO	5	5	0
COIMBRA	10	10	0
ÉVORA	4	4	0
FARO	8	8	0
GUARDA	4	4	0
LEIRIA	10	10	0
LISBOA	50	50	0
PORTALEGRE	3	3	0
PORTO	37	38	1
SANTARÉM	10	10	0
SETÚBAL	17	17	0
V. CASTELO	6	5	-1
V. REAL	5	5	0
VISEU	9	9	0
AÇORES	5	5	0
MADEIRA	5	5	0
TOTAL	226	226	0

Verifica-se que, a dar como válidos aqueles números, os distritos de Beja, Bragança e Viana do Castelo teriam perdido nas eleições de Outubro de 1995 um deputado a favor dos de Braga, Porto e Aveiro.

Para além das consequências no número de deputados em cada círculo distrital, haveria que considerar as consequências ao nível da distribuição dos deputados entre os vários partidos, para cada círculo. Feita a simulação com esta nova distribuição de deputados por círculo, e utilizando os dados (votações) das eleições legislativas de Outubro de 1995, obtivemos a composição da Assembleia da República, tal como descrito no Quadro III, na coluna "estimativa 95". Como se poderá constatar, no cômputo geral, o Partido Socialista e o Partido Comunista perderiam um deputado, que seriam ganhos pelo Partido Social Democrata e pelo Partido Popular.

QUADRO III

DISTRITO	Nº DEPUTADOS 95	1995				ESTIMATIVA 95	ESTIMATIVA 1995			
		PS	PSD	PP	PCP		PS	PSD	PP	PCP
AVEIRO	14	6	6	2		15	6	7	2	
BEJA	4	2	1		1	3	2	1		0
BRAGA	16	8	7	1		17	8	7	2	
BRAGANÇA	4	2	2			3	1	2		
C. BRANCO	5	3	2			5	3	2		
COIMBRA	10	6	4			10	6	4		
ÉVORA	4	2	1		1	4	2	1		1
FARO	8	5	3			8	5	3		
GUARDA	4	2	2			4	2	2		
LEIRIA	10	4	5	1		10	4	5	1	
LISBOA	50	24	15	5	6	50	24	15	5	6
PORTALEGRE	3	2	1			3	2	1		
PORTO	37	18	14	3	2	38	19	14	3	2
SANTARÉM	10	5	3	1	1	10	5	3	1	1
SETÚBAL	17	9	3	1	4	17	9	3	1	4
V. CASTELO	6	3	3			5	2	3		
V. REAL	5	2	3			5	2	3		
VISEU	9	4	4	1		9	4	4	1	
AÇORES	5	2	3			5	2	3		
MADEIRA	5	2	3			5	2	3		
TOTAL	226	111	85	15	15	226	110	86	16	14
		226					226			

4. CONCLUSÃO

A actual geografia eleitoral é afectada negativamente pela existência de cadernos eleitorais desactualizados, obsoletos. Estes integram duplas inscrições para além de que contêm indevidamente um número significativo de cidadãos emigrantes e até muitos já falecidos.

Este conjunto de incorrecções origina enviesamentos de vários tipos.

Em primeiro lugar, ao nível da proporcionalidade na distribuição de mandatos por círculos, favorecendo distritos que, sendo menos populosos, comportam nos seus cadernos eleitorais um número considerável de eleitores "fantasma". Com a actualização dos cadernos eleitorais, prevê-se que os distritos de Beja, Bragança e Viana do Castelo percam um deputado cada. Este efeito será tanto mais acentuado,

quanto se prevê também a redução do número de deputados na Assembleia da República.

Ao nível dos vários partidos com representação parlamentar, as consequências duma actualização do recenseamento serão reduzidas, uma vez que apenas se prevêem pequenas variações de mais ou menos um deputado por partido, em termos gerais - sendo que estas resultam de variações mais amplas a nível de cada um dos diversos círculos distritais, mas cujos efeitos se contrariam. Podemos pois concluir que na situação presente, e mesmo com cadernos eleitorais desactualizados, se salvaguarda o princípio de proporcionalidade estipulado pelo artigo 288 da Constituição Portuguesa.

AGRADECIMENTOS:

Ao Instituto Nacional de Estatística, na pessoa do Professor Doutor Paulo Gomes e ao STAPE, na pessoa do Senhor Dr. Jorge Miguéis, pela cedência de dados e pela troca de opiniões, que em muito contribuíram para as melhorias introduzidas no presente trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

COTTERET, JEAN MARIE ET ÉMERI, CLAUDE - *Systèmes électoraux*. Paris, Presses Universitaires de France

IDEA (1997), *Voter Turnout from 1945 to 1997. A Global Report*. Estocolmo, IDEA,

MENDES, MARIA DE FÁTIMA ABRANTES e MIGUÉIS, JORGE - *Lei Eleitoral da Assembleia da República actualizada, anotada e comentada*,

OLIVEIRA, TIAGO DE, (1981) - "O Sistema Eleitoral Português como Forma de Representação", *Análise Social*, vol. XVII (65);

STAPE - *Eleição da Assembleia da República 1995 (escrutínio provisório)*.

INFORMAÇÕES

ACTIVIDADES E PROJECTOS IMPORTANTES NO ÂMBITO DO SISTEMA ESTATÍSTICO NACIONAL

IMPORTANT ACTIVITIES AND PROJECTS IN THE SCOPE OF THE NATIONAL STATISTICAL SYSTEM

O processo de planeamento da actividade estatística abrange a definição dos objectivos estratégicos, através das Linhas Gerais da Actividade Estatística Nacional, de periodicidade quadrienal; a concretização destes objectivos e a apresentação das principais acções a desenvolver a médio prazo, no Programa Estatístico de Médio Prazo, com a mesma periodicidade; e a planificação anual das acções a executar, no Plano de Actividades do INE e das Entidades com delegação de competências do INE.

As Linhas Gerais da Actividade Estatística Nacional e respectivas prioridades para 1998-2002 foram aprovadas pelo Conselho Superior de Estatística em Maio de 1997 (Deliberação nº 125/97). O Programa Estatístico de Médio Prazo 1998-2002, que obteve parecer favorável do Conselho Superior de Estatística, em Novembro de 1997 (Deliberação nº 135/97), concretiza os principais objectivos estratégicos e acções de desenvolvimento para os seguintes grandes domínios da actividade estatística:

• Infra-estruturas Estatísticas	Coordenação do Sistema Estatístico Nacional	Contas Nacionais e Regionais Coordenação de Objectivos Coordenação Técnica Coordenação Regional
	Metodologias e Controlo da Qualidade	
	Infraestruturas Tecnológicas, Informacionais e Físicas	
• Produção Estatística	Administrações Públicas e Privadas Agricultura. Produção Animal, Silvicultura e Pescas Ambiente Ciência e Tecnologia Comércio e Outros Serviços Construção e Habitação Empresas Indústria e Energia População, Emprego e Outras Estatísticas Sociais Transportes e Comunicações Turismo e Restauração Consolidação do Sistema de Informação Estatística à escala regional	
• Estudos	Estudos Demográficos e Sociais Estudos Económicos Estudos Regionais	
• Difusão		
• Cooperação		
• Recursos Humanos e Formação		

Na sequência dos dois documentos acima referidos, foi elaborado o Plano de Actividades do INE e das Entidades com delegação de competências do INE para

1998, sobre o qual o Conselho Superior de Estatística emitiu parecer favorável em Novembro de 1997.

1. OBJECTIVOS DO PLANO DE ACTIVIDADES PARA 1998

Tendo em conta as orientações delineadas pelas Linhas Gerais da Actividades Estatística Nacional para 1998-2002, e posteriormente concretizadas no Programa Estatístico de Médio Prazo para o mesmo período, a actividade estatística nacional em 1998 será orientada pelos seguintes objectivos:

- Cumprir as obrigações legais e os compromissos assumidos no quadro do Programa Estatístico Comunitário para 1998-2002;
- Modernizar os métodos de produção e difusão da informação estatística, com a introdução de novos suportes de recolha da informação, de novas ferramentas de software para tratamento estatístico na produção e de novas soluções para difusão, em particular utilizando a internet e tecnologias multimédia;
- Melhorar progressivamente os prazos de disponibilização da produção estatística, estabelecendo padrões de qualidade (prazos-objectivo) adequados à natureza e periodicidade das informações produzidas;
- Melhorar a qualidade da produção e difusão de informação estatística, criando os instrumentos básicos necessários à formalização de um sistema de gestão da qualidade para o INE, e promovendo a avaliação e o controlo da qualidade das estatísticas produzidas;
- Normalizar e intensificar a sincronização dos calendários de produção das contas nacionais e regionais, e continuar os trabalhos visando aplicação integral do SEC 95;
- Promover o aproveitamento dos actos administrativos para fins estatísticos, insistindo na adequação da legislação a este objectivo;
- Criar uma base geográfica de apoio à execução de operações estatísticas, de primordial importância para a realização dos Censos 2001;
- Continuar a avaliação do processo de delegação de competências para a produção de estatísticas oficiais, visando reduzir os custos da produção e evitar duplicações;
- Prosseguir a criação e desenvolvimento de novas infra-estruturas tecnológicas e informacionais de apoio à produção e difusão da informação estatística;
- Consolidar a formação de quadros, executando planos de formação dirigidos aos técnicos superiores e profissionais;
- Criar e/ou consolidar sistemas de informação estatística à escala regional, em estreito diálogo com os principais agentes e instituições regionais.

No sentido de alcançar estes objectivos, o Plano de Actividades para 1998 tem inscritos 345 modelos estatísticos, dos quais 227 da responsabilidade do INE (66%), e 118 de outras entidades intervenientes na produção estatística oficial (34%).

Dos 345 modelos estatísticos, 270 são de 1ª prioridade (78%), e 75 são de 2ª prioridade (22%); dos modelos de 1ª prioridade, 125 são obrigatórios por legislação nacional ou comunitária (36%).

Apresentam-se ainda 21 modelos estatísticos novos (6%), considerando-se os restantes 324 de desenvolvimento corrente.

O Plano de Actividades para 1998 apresenta a caracterização de todos os modelos estatísticos nele inscritos, ventilados por 29 áreas estatísticas, a seguir enumeradas:

- | | |
|---|---|
| ⇒ Administrações Públicas | ⇒ Estatísticas Gerais |
| ⇒ Agricultura, Produção Animal e Silvicultura | ⇒ Formação Profissional |
| ⇒ Ambiente | ⇒ Habitação, Construção e Obras Públicas |
| ⇒ Ciência e Tecnologia | ⇒ Indústria e Energia |
| ⇒ Comércio Internacional | ⇒ Iniciativas de Produção e Estudos Regionais |
| ⇒ Comércio Interno e Outros Serviços | ⇒ Instituições Financeiras e Seguros |
| ⇒ Condições de Vida das Famílias | ⇒ Justiça |
| ⇒ Conjuntura Económica | ⇒ Pesca |
| ⇒ Contas Nacionais e Regionais | ⇒ Preços |
| ⇒ Cultura, Desporto e Recreio | ⇒ Protecção Social |
| ⇒ Deficiência e Reabilitação | ⇒ Relações e Condições do Trabalho |
| ⇒ Demografia | ⇒ Saúde |
| ⇒ Educação | ⇒ Transportes e Comunicações |
| ⇒ Emprego e Salários | ⇒ Turismo e Restauração |
| ⇒ Empresas | |

2. PRINCIPAIS ACÇÕES A DESENVOLVER

2.1 - PRODUÇÃO, ESTUDOS E DIFUSÃO DE INFORMAÇÃO ESTATÍSTICA

2.1.1 - ESTATÍSTICAS ECONÓMICAS

- Dar continuidade ao processo de mais rápida elaboração das Contas Nacionais portuguesas, prosseguindo o objectivo da disponibilização de contas definitivas em Julho do ano t+2 (sendo t o ano de referência dos dados);
- Aprofundar a interligação entre as Contas Nacionais anuais e as trimestrais no quadro do SEC 95, e possibilitar a normalização do calendário de difusão das Contas Nacionais trimestrais;
- Estabilizar os calendários de produção das Contas Regionais, e aprofundar o sistema no contexto do SEC 95;

- Desenvolver os trabalhos que permitam a obtenção directa dos dados da contabilidade pública, numa óptica de contabilidade nacional;
- Aprofundar o processo de diálogo técnico com o Banco de Portugal, visando a adequada integração das contas financeiras no processo de elaboração das Contas Nacionais portuguesas;
- Desenvolver os trabalhos relativos à aplicação do SEC 95 às Contas Económicas da Agricultura e Silvicultura, em harmonização com os correspondentes ramos das Contas Nacionais;
- Apresentar a 1ª edição das Contas do Ambiente - SERIEE;
- Aprofundar os trabalhos de concepção das Contas-Satélite relativas à Educação e ao Trabalho;
- Ajustar as estatísticas das Administrações Públicas às novas exigências do SEC 95; identificar o universo das Administrações Privadas e implementar um processo de recolha de informação estatística neste domínio;
- Divulgar os resultados do Inquérito às Empresas de 1996 e lançar o inquérito relativo a 1997;
- Difundir regularmente os resultados do Sistema de Contas Integradas das Empresas;
- Estabelecer a periodicidade trimestral no Painel de Empresas e estabilizar os calendários de recolha e tratamento da informação;
- Consolidar a metodologia do projecto Demografia das Empresas, do INE;
- Divulgar os resultados do Inquérito à Estrutura das Explorações Agrícolas de 1997;
- Desenvolver um novo inquérito mensal de conjuntura dirigido aos consumidores, permitindo obter a única informação infra-anual sobre a situação económica e financeira das famílias, bem como as suas expectativas sobre a evolução próxima da economia;
- Produzir e divulgar as estatísticas primárias e derivadas do Comércio Internacional; aplicar as alterações decorrentes do projecto SLIM¹ no sistema INTRASTAT; e expandir a utilização de novos suportes na transmissão de dados do Comércio Internacional;
- Normalizar os calendários do Inquérito Anual à Produção Industrial, com disponibilização dos resultados relativos a 1996 e 1997;
- Produzir e divulgar os resultados dos seguintes índices de base 1995:
 - * Índice de Volume de Negócios e Emprego na Indústria;
 - * Índice de Produção Industrial - Indústria e Energia;
 - * Índice de Preços na Produção Industrial;
 - * Índice de Volume de Negócios no Comércio a Retalho;
- Produzir e divulgar resultados do Índice de Produção Industrial - Construção;
- Definir a metodologia de novos projectos, nomeadamente Índice de Novas Encomendas na Indústria e na Construção, e Índice de Preços na Produção - Construção;

¹ SLIM - Simplificação da legislação para o mercado interno.

- Produzir estatísticas de águas minerais engarrafadas e de termalismo;
- Recuperar o calendário do Inquérito aos Transportes Rodoviários de Mercadorias, no âmbito do novo Regulamento do Conselho (em aprovação).

2.1.2 - ESTATÍSTICAS SOCIAIS

- Executar uma nova edição do Inquérito aos Resíduos Industriais e do Inquérito às Eco-Empresas, após reflexão sobre a metodologia e os resultados dos inquéritos relativos a 1995 e 1996;
- Prosseguir a realização de inquéritos às unidades com potencial envolvimento em actividades de C&T nos sectores das Empresas, Ensino Superior, Estado e Instituições Particulares sem fins lucrativos;
- Divulgar o Índice de Preços no Consumidor, Base 1997 = 100 (metodologia e resultados mensais);
- Lançar, no 2.^o semestre de 1998, o Inquérito ao Consumo (contínuo);
- Fazer a exploração longitudinal do Painel Comunitário de Agregados Familiares (resultados das três vagas) para análise da desigualdade, pobreza e exclusão social;
- Continuar os trabalhos relativos à preparação dos Recenseamentos da População e da Habitação 2001, nomeadamente a elaboração e apresentação do respectivo programa;
- Reformular a metodologia dos inquéritos sobre movimentos migratórios e população estrangeira;
- Elaborar estudos com base nos resultados do Inquérito à Fecundidade e Família;
- Actualizar as metodologias do cálculo das estimativas e projecções de população residente e calcular estimativas mensais de população residente;
- Fazer a previsão de variáveis sócio-demográficas, tais como nados-vivos, óbitos, migrações, casamentos e divórcios, com recurso a análise de sucessões cronológicas; construir modelos econométricos explicativos da fecundidade, da mortalidade e das migrações; e estudar a tendência e a sazonalidade de algumas séries temporais relativas a variáveis sócio-demográficas;
- Iniciar o estudo metodológico do cálculo de projecções de população activa, no âmbito de um grupo de trabalho do Eurostat, e integrado no projecto Matriz Demográfica;
- Disponibilizar, no início do ano lectivo, estatísticas preliminares sobre alunos matriculados, pessoal docente e pessoal não docente;
- Iniciar a divulgação da nova série do Inquérito ao Emprego, decorrente das adaptações metodológicas introduzidas em 1997; estudar a ligação à série anterior e difundir uma série longa para as principais variáveis no mercado de trabalho;
- Produzir informação sobre caracterização do emprego e qualificação dos trabalhadores estrangeiros;
- Analisar o comportamento da população com idade superior a 55 anos, no mercado de trabalho;

- Analisar a realidade dos grupos de exclusão social, numa óptica de emprego e de formação;
- Caracterizar, quantitativa e qualitativamente, os desempregados inscritos nos Centros de Emprego;
- Fazer o levantamento e estudo da informação disponível sobre trabalhadores com baixas remunerações;
- Realizar um inquérito à estrutura de formação profissional;
- Prosseguir com a realização do Inquérito à Evolução das Qualificações e Diagnóstico de Necessidades de Formação, na sequência do acordo de colaboração do INE com o Instituto para Inovação da Formação;
- Realizar um Inquérito à Ocupação do Tempo, na sequência do inquérito-piloto já executado;
- Consolidar o SEEPROS, no âmbito das estatísticas do trabalho e da protecção social, designadamente no que se refere à cobertura das diversas funções e análise da possibilidade de recolha de informação sobre as associações sindicais.

2.1.3- INFRA-ESTRUTURAS DE APOIO À PRODUÇÃO

- Insistir nas autorizações de acesso a fontes administrativas, em particular as fiscais, visando aliar a diminuição dos custos de produção estatística à melhoria da respectiva qualidade;
- Consolidar o processo de actualização permanente do ficheiro central de empresas, na sua componente estabelecimentos, com a intervenção das Direcções Regionais do INE;
- Analisar o processo de integração no ficheiro de empresas da componente “grupo de empresas”;
- Iniciar o processo de actualização permanente da “Amostra-mãe”;
- Concluir os trabalhos de criação de um sistema automático de tratamento dos conceitos estatísticos, seu carregamento e disponibilização;
- Concretizar o “Sistema Integrado de Nomenclaturas Económicas”;
- Prosseguir o processo de digitalização da Base Geográfica de Referenciação da Informação, priorizando as necessidades de organização da operação Censos 2001.

2.2- TECNOLOGIAS DE INFORMAÇÃO

- Implementar novos suportes de recolha da informação, em particular pela utilização de EDI (*Electronic Data Interchange*) e ICR (*Intelligent Character Recognition*);
- Intensificar a utilização da Internet e Intranet, enquanto meios privilegiados para difusão de informação;
- Generalizar no INE o Office 97 devido às novas funcionalidades, especialmente as relacionadas com Inter/Intranet;
- Utilizar novo software para arquitecturas cliente-servidor que permita maior produtividade de desenvolvimento;

- Actualizar o parque de servidores e PC's do INE;
- Aumentar a largura de banda nas conexões entre os edifícios do INE em Lisboa;
- Desenvolver no INE projectos aplicativos, nomeadamente para:
 - * Preparação do Recenseamento da População e da Habitação - 2001;
 - * Preparação do Recenseamento Geral da Agricultura - 1999;
 - * Informatização das Contas Nacionais;
 - * Adaptação das aplicações ao EURO e ao novo milénio;
 - * Descentralização das aplicações informáticas no Serviço Regional de Estatística dos Açores e na Direcção Regional de Estatística (Madeira).

2.3 - REGIONALIZAÇÃO

- Continuar a inventariação de fontes de informação estatística regional e dos actos administrativos dos agentes e instituições regionais, com potencial aproveitamento para fins estatísticos;
- Prosseguir a instalação da nova Direcção Regional do Algarve;
- Regionalizar algumas actividades estatísticas, nomeadamente o Inquérito ao Emprego e o Inquérito às Empresas, ao nível NUTS III;
- Realizar os Inventários Municipais de 1998 em todas as regiões do Continente;
- Iniciar a produção de indicadores estatísticos regionais sobre a problemática da habitação;
- Continuar a promover o desenvolvimento das estatísticas transfronteiriças;
- Prosseguir, nas Regiões Autónomas dos Açores e da Madeira, a implementação de projectos comunitários no âmbito da Conférence des Régions Péripheriques Maritimes;
- Elaborar estudos temáticos e monografias caracterizadoras de aspectos sócio-económicos básicos da realidade regional ou infra-regional;
- Realizar acções de divulgação da informação estatística nas escolas, visando formar os potenciais e/ou futuros utilizadores da produção estatística.

CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS

CONGRESS, SEMINARS AND CONFERENCES

Em Portugal:

In Portugal:

1998

□ *04 de Junho

Regressão não Paramétrica, Évora.

Informações: *Salomé Cabral*, Dep. Estatística e Investigação Operacional
(FCUL), Bloco C2, Piso 2 – Campo Grande, 1700 Lisboa.

No Estrangeiro:

Abroad:

1998

□ 13 - 16 de Maio

**INTERFACE' 98 on the interface between Computer Science and
Statistics**, Radisson Metrodome Hotel, Minneapolis, Minnesota, USA.

Informações: *Sanford Weisberg*, School of Statistics, University of Minnesota,
St. Paul, Minnesota 55108-6042, USA; Telf.: 1 612 6258777;
FAX: 1 612 6242719.

E - mail: interface@stat.umn.edu;

or

WWW: <http://www.stat.umn.edu>

□ 20 - 22 de Maio

Conference on Applied Statistics in Ireland, Athlone, Ireland.

Informações: *T. O' Hara*, Elan Cooperation plc., Monksland, Athlone Co.
Westmeath, Ireland; Telf.: 353 902 95127.

E - mail: ohartom@iol.ie

or

A. Dunne, Department of Statistics, University College Dublin,
Belfield, Dublin 4, Ireland; Telf.: 353 1 706 151.

E - mail: Adrian.Dunne@ucd.ie

□ *25 - 29 de Maio

XXXemes Journées de Statistique, Rennes, France.

Informações: ENSAI, École Nationale de la Statistique et de l'Analyse de
l'Information, Campus de Ker Lann, Rue Blaise Pascal, 35170
Bruz, France.

WWW: <http://www.ensai.fr/asu98/>

- *28 de Maio - 27 de Junho
The East-West Centers 29th Summer Seminar on Population, Honolulu, Hawaii.
Informações: Summer Seminar, East-West Center, Program on Population, 1601 East-West Road, Honolulu, Hawaii 96848-1601. Telf.: 808 9447762 or 808 9447410;
 E - mail: sumsem98@ewc.hawaii.edu.
 or
 WWW: <http://www.ewc.hawaii.edu>
- *29 de Maio
Meta-Analysis in Pharmaceutical Practice: Achieving quality and avoiding pitfalls, London, U.K.
Informações: *Anne Joyce*, Ehenry Stewart Conference Studies, Russell House, 28/30 Little Russell Street, London WC1A 2HN, UK; Telf.: 44 171 4043040; Fax: 44 171 4042081;
 E - mail: annej@henrystewart.demon.co.uk
- 31 de Maio – 03 de Junho
Annual Meeting of the Statistical Society of Canada, Sherbrooke, Quebec, Canada
Informações: *E. Monga*, Local Arrangements Chair, Département de Mathématiques et d'Informatique, faculté des Sciences, Université de Sherbrooke, Sherbrooke, Quebec, J1K 2R1, Canada.
 E - mail: ernest.monga@dmi.usherb.ca
- 01 - 03 de Junho
Sixth International Applied Statistics in Industry and Manufacturing Conference, San Diego, California, USA.
Informações: Sixth IASIM conference, P. O. Box 189, Mulvane, KS67110, USA; FAX: 1 316 7890906;
 E - mail: tracy@isai.org;
 or
 WWW: <http://www.isai.org>.
- 03 - 05 de Junho
Spring Research Conference on Statistics in Industry and Technology, Sante Fe, New Mexico, USA.
Informações: *Robert Easterling*, E-mail: rgeaste@sandia.gov;
 or
Bovas Abraham, E - mail: babraham@setosa.waterloo.ca;
 or
Joan Wendelberger, E - mail: joanne@lanl.gov.
- *03 - 05 de Junho
First meeting on Public Statistics, Household Surveys, Buenos Aires, Argentina
Informações: Technical Secretariat of IASI, Balcarce 184 – 2 piso, Oficina 211, 1327 Buenos Aires, Argentina; Fax: 54 1 3495776
 E - mail: efabb@indec.mecon.ar

- 09 - 11 de Junho
21st International Conference on Regional and Urban Statistics (SCORUS/IAOS), Belfast, N. Ireland.
Informações: *D. Bond*, Ulster Business School, University of Ulster at Coleraine, Coleraine BT52 1RT, UK.
E - mail: D.Bond@ulst.ac.uk
or
A. Manninen, Chair of Programme Committee, City of Helsinki Information Management Centre, P. O. Box 303, FIN-00171 Helsinki, Finland.
E - mail: asta.manninen@hel.fi

- 15 - 17 de Junho
7th ISIP/ICCC Conference on Information Networks and Data Communications, Aveiro, Portugal
Informações: *Mrs. Anabela Viegas*, INESC, University of Aveiro, Portugal;
Telf.: 351 34 370503; FAX: 351 34 370545.
E - mail: anabela@inesca.pt

- 16 - 19 de Junho
20th International Conference on Information Technology Interfaces ITI' 98, Pula, Croatia.
Informações: *Damir Kalpic*, Chairman of the International Program Committee, Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia; Telf.: 385 1 129919; FAX: 385 1 6129915.
E - mail: damir.kalpic@fer.hr

- 21 - 26 de Junho
Fifth International Conference on Teaching Statistics (ICOTS V), to be held at Nanyang Technological University, Singapore.
Informações: *Brian Phillips*, School of Mathematical Sciences, Swinburne University of Technology, P. O. Box 218, Hawthorne 3122, Victoria, Australia; Telf.: 61 3 9214 8288; FAX: 61 9819 0821.
E - mail: bphillips@swin.edu.au
or
WWW: <http://www.nie.ac.sg:8000/~wwwmath/icots.html>

- 05 - 11 de Julho
SPA 98, 25th International Conference on Stochastics Processes and Their Applications, Corvallis, USA.
Informações: *E. Waymire*, Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA; FAX: 1 503 737 0515.
E - mail: waymire@math.orst.edu

- 19 - 25 de Julho
VII International Congress of Ecology, Florence, Italy.
Informações: *Prof. G. P. Patil*, Director, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University, 421 Thomas Building, University Park, PA 16802-2112, USA.
E - mail: gpp@stat.psu.edu

- 21 - 24 de Julho
VI Conference of the International Federation of Classification Societies on Data Science, Classification and Related Methods, Rome, Italy.
Informações: IFCS-98, Dpt. Statistica, Probabilità e Statistiche Applicate, Università "La Sapienza" di Roma, Piazzale A. Moro 5 00 185 Roma, Italy; FAX: 39 6 4959241.
 E - mail: ifcs.98@pow2.sta.uniroma1.it
 or
 WWW: <http://www.pow2.sta.uniroma1.it/ifcs98>
- 27 - 31 de Julho
13th International Workshop on Statistical Modelling, New Orleans, Louisiana, USA.
Informações: Brian D. Marx, Department of Experimental Statistics, 161 Ag.Admin Bldg., Louisiana State University, Baton Rouge, LA 70803-5606, USA; Telf.: 1 504 388 8366; FAX: 1 504 388 8344;
 E - mail: brain@stat.lsu.edu
 or
 WWW: <http://www.stat.lsu.edu/iwsm>.
- Agosto de 1998 – Junho de 1999
The Fields Institute for Research in Mathematical Sciences Course Program in Probability and its Applications.
Informações: The Fields Institute for Research in Mathematical Sciences, 222 College Street, Second Floor, Toronto, Ontario M5T 3J1, Canada; Telf.: 416 348 9710; FAX: +416 348 9385.
 E - mail: probability@fields.utoronto.ca
 or
 WWW: <http://www.fields.utoronto.ca/>
- *07 - 10 de Agosto
European Summer School on Markov Chain Monte Carlo Methods, Rebild, Denmark.
Informações: Lisbeth G. Nielsen, Dept. of Math, Aalborg University; Fax: 45 98 158129;
 E - mail: grubbe@math.auc.dk
 or
 WWW: <http://www.maths.nott.ac.uk/hsss/Workshops/summerschool.html>
- 09 - 13 de Agosto
1998 Joint Statistical Meetings, Dallas, Texas, USA.
Informações: ASA, 1429 Duke St., Alexandria, VA 22314-3402, USA; Telf.: 703 684 1221; FAX: 703 684 2037.
 E - mail: meetings@asa.mhs.compuserve.com
- 12 - 18 de Agosto
EMS'98 European Meeting of Statisticians and the 7th Vilnius Conference, Vilnius, Lithuania.
Informações: A. Plikusas, M. Bloznelis, Institute of Mathematics & Informatics, Akademijos 4, Vilnius 2600, Lithuania; FAX: 370 2 729209.
 E - mail: conf@ktl.mii.lt

- *17 de Agosto - 02 de Setembro
XXVIIIth Probability Summer School, Saint Flour (Cantal), France.
Informações: *P. Bernard*, Université Blaise Pascal, Mathématiques Appliquées,
F-63177 Aubiere, Cedex, France; Telf.: 04 73407052 or 04
73407050; Fax: 047 3407064;
E - mail: bernard@ucfma.univ-bpclermon.fr

- *19 - 21 de Agosto
4th ICSA Statistical Conference jointly with the 2nd Statistics Conference of Two Sides of Straits, Kunming, China.
Informações: *Xiao-Li Meng*, Conference Coordinator, Department of Statistics,
University of Chicago, 5734 University Ave, Chicago, IL 60637,
USA; Telf.: 773 7020958; Fax: 7737029810;
E - mail: meng@galton.uchicago.edu.
or
WWW: <http://icsa.orgkunming98>

- *20 - 22 de Agosto
Perspectives in Modern Statistical Inference: Parametrics, Semiparametrics, Nonparametrics. Workshop Satellite to Prague Stochastics '98, Prague, Czech Republic.
Informações: *J. Antoch, J. Jurevckova*, Dept. of Statistics, Charles University,
Sokolovská 83, 186 75 Prague, Czech Republic;
E - mail: JURECKO@KARLIN.MFF.CUNI.CZ
or
WWW: <http://www.utia.cas.cz/stoch98.html>

- 23 - 28 de Agosto
Prague Stochastics '98, a joint session of the Sixth Prague Symposium on Asymptotic Statistics and the 13th Prague Conference on Information Theory, Statistical decision Function, Random processes, Bristol, England.
Informações: *Zuzana Prásková*, Dept. of Statistics, Charles University,
Sokolovská 83, 186 00 Prague, Czech Republic.
E - mail: stoch98@utia.cas.cz

- 23 - 28 de Agosto
COMPSTAT 98: International Symposium on Computational Statistics, Prague, Czech republic.
Informações: *R Payne*, Statistics Dept., IACR-Rothamsted, harpenden, herts
AL5 2JQ, UK; FAX: +44.1582.760981.
E - mail: compstat98@bristol.ac.uk

- 24 - 28 de Agosto
COMPSTAT'98, Bristol, UK.
Informações: COMPSTAT'98, *Professor Roger Payne*, Statistics Department,
IACR-Rothamsted, harpenden, Herts AL5 2JQ, UK; FAX: 44
1582 760981.
E - mail: compstat-98@bristol.ac.uk
or
WWW: <http://www.stats.bris.ac.uk/COMPSTAT/>

- *24 - 28 de Agosto
15th International Congress on Cybernetics, Namur, Belgium.
Informações: Palais des Expositions, avenue Sergent Vriethoff, 2, B-5000
Namur, Belgium, Telf.: 32 81717171; Fax: 32 81717100;
E - mail: Cyb@info.fundp.ac.be

- *24 - 29 de Agosto
The Fields Institute Workshop on Mathematical Physics of Polymers and Percolation.
Informações: The Fields Institute for Research in Mathematical Sciences, 222 College Street, Second Floor, Toronto, Ontario M5T3J1, Canada;
Telf.: 416 348 9710; Fax: 416 348 9385;
E - mail: probability@fields.utoronto.ca
or
WWW: <http://www.math.yorku.ca/Probability/Fields.html>
- *25 - 27 de Agosto
4th Iranian Statistics Conference, Tehran, Iran.
Informações: *Professor M. Reza Meshkani*, Organizing Committee, Dept. of Statistics, School of Mathematical Sciences, Shahid Beheshti University, Evin, Tehran 19839, Iran; Telf./Fax: 98 21 2403133.
- *30 de Agosto - 05 de Setembro
International Colloquium on Mathematics in Gambling, Budapest, Hungary.
Informações: E - mail: jatek@math-inst.hu
or
WWW: <http://www.math-inst.hu/~jatek>
- 31 de Agosto - 04 de Setembro
15th IFIP World Computer Congress and SEC '98 – 14th International Information Security Conference, Vienna e Budapest.
Informações: Oesterreichische Computer Gesellschaft, (Austrian Computer Society OCG), Wollzeile 1-3, A-1010 Vienna, Austria; Telf.: 43 1 5120235; FAX: 43 1 5120235 9.
E - mail: ifip98@ocg.or.at
or
ocg@ocg.or.at
WWW: <http://www.ocg.or.at>
or
John v. Neumann Computer Society NJSZT, Bathori u. 16, H-1054 Budapest, Hungary; Telf.: 36 1 1329349; FAX: 36 1 1318140.
E - mail: ifip98@neumann.hu
or
h13588zub@ella.hu
WWW: <http://www.njszt.iff.hu>
- 01 - 04 de Setembro
Joint IASS/IAOS Conference, Aguascalientes, Mexico.
Informações: *Dennis Trewin*, Australian Bureau of Statistics, P. O. Box 10, Belconnen, ACT 2616, Australia.
E - mail: dennis.trewin@abs.gov.au
or
Denise Lievesley, ESRC Data Archive, University of Essex, CO4 3SQ, UK.
E - mail: denise@essex.ac.uk
- *07 - 10 de Setembro
EUFIT'98, Aachen, Germany.
Informações: WWW: <http://www.mitgmbh.de>

- 07 - 11 de Setembro

1998 Conference of the Royal Statistical Society, to be held at the University of Strathclyde, Glasgow, UK.

Informações: *Professor Eric Renshaw*, Department of Statistics and Modelling Science, University of Strathclyde, Glasgow G1 1XH, UK; FAX: 44 141 5522079.

E - mail: rss98@stams.strath.ac.uk

WWW: <http://www.stams.strath.ac.uk/rss98/index.html>

- 09 - 11 de Setembro

PROLAMAT' 98. Tenth International IFIP TC5 WG-5.2 WG-5.3 Conference. The theme is "The Globalization of Manufacturing in the Digital Communications Era of the 21st Century: Innovation, Agility, and the Virtual Enterprise", Trento, Italy.

Informações: *Mara Gruber*, Laboratorio di Ingegneria Informatica, Via f. Zeni, 8, 38068 Rovereto (TN), Italy; Telf: 39 464 443134; FAX: 39 464 443141.

E - mail: prolamat@lii.unitn.it;

or

WWW: <http://www.lii.unitn.it/prolamat/>

- 14 - 18 de Setembro

"German Statistical Week", Lübeck, Germany.

Informações: Verband Deutscher Städtestatistiker, Bereich Statistik und Wahlen, Schwartzstrasse 73, 46045 Oberhausen, Germany.

- *25 - 26 de Setembro

Workshop on "Uncertainty in the Risk Assessment of Environmental and Occupational Hazards", Bologna, Italy.

Informações: *Dr. J. C. Bailer*, Department of Health Statistics, University of Chicago, 5841 S Maryland Ave, Chicago, Illinois 60637, USA; Telf.: 773834 1242; Fax: 773 7021979;

E - mail: jcbailer@midway.uchicago.edu

or

Dr. A. J. Bailer, Dept. of Math and Stats, Miami University, Oxford, OH 45056-1641, USA; Telf.: 513 5293538; Fax: 513 5291493;

E - mail: ajbailer@muohio.edu

- *05 - 10 de Outubro

The Fields Institute Workshop on Hydrodynamic Limits.

Informações: The Fields Institute for Research in Mathematical Sciences, 222 College Street, Second Floor, Toronto, Ontario M5T 3J1, Canada; Telf.: 416 348 9710; Fax: 416 348 9385;

E - mail: probability@fields.utoronto.ca

or

WWW: <http://www.math.yorku.ca/Probability/Fields.html>

- 22 - 24 de Outubro

International Seminar on Seasonal Adjustment methods (SAM 98), sponsored by Eurostat and ISI, Bucharest, Romania.

Informações: *Mr. R. Depoutot*; telf.: 352 4301 34926; FAX: 352 4301 34149.

- 25 - 28 de Outubro

FRACTAL 98, "Complexity and Fractals in the Sciences", 5th International Multidisciplinary Conference, Valletta, Malta.

Informações: *Dr. Miroslav M. Novak*, School of Physics, Kingston University, Surrey KT1 2EE, UK; Telf.: 44 181 547 7481; FAX: 44 181 547 7562.

E - mail: novak@kingston.ac.uk;

or

WWW: <http://www.kingston.ac.uk/fractal>.

- *25 - 29 de Outubro

The Fields Institute Workshop on Monte Carlo Methods.

Informações: The Fields Institute for Research in Mathematical Sciences, 222 College Street, Second Floor, Toronto, Ontario M5T 3J1, Canada; Telf.: 416 348 9710; Fax: 416 348 9385;

E - mail: probability@fields.utoronto.ca

or

WWW: <http://www.math.yorku.ca/Probability/Fields.html>

- 02 - 04 de Dezembro

ARS' 98 Conference, 3rd Conference on Statistical Computing of the IASC Asian Regional Section, Manila, Philippines.

Informações: *Dr. Romulo A. Virola*, 2/F Midland Buendia Bldg., 403 Sen. Gil J. Puyat Ave., Makati City, Philippines; Telf.: 632 890 9495 / 632 895 2395; FAX: 632 890 9408.

E - mail: nscbsg@mozcom.com.

- *13 - 18 de Dezembro

XIXth International Biometric Conference, Cape Town, South Africa.

Informações: *Professor Christine McLaren*, Department of Mathematics, Moorhead State University; Telf.: 218 2364004; Fax: 218 2362168;

E - mail: mclaren@mhd1.moorhead.msus.edu

or

Janet Riley, E - mail: janet.riley@bbsrc.ac.uk

- 14 - 16 de Dezembro

Seventh International Applied Statistics in Industry and Manufacturing Conference, Singapore

Informações: Seventh IASIM conference, P. O. Box 189, Mulvane, KS67110, USA; FAX: 1 316 7890906;

E - mail: tracy@isai.org;

or

WWW: <http://www.isai.org>.

1999

- *17 - 21 de Maio

Eleventh International Genstat Conference, Poznan, Poland.

Informações: Genstat'99, *Dr. Pawel Krajewski*, Institute of Plant Genetics, Polish Academy of Sciences, Strzeszynska 34, 60-479 Poznan, Poland; Telf.: 48 61 8233511; Fax: 49 61 8233671;

E - mail: pkra@igr.poznan.pl

- Junho

The IX International Symposium on Applied Stochastic Models and Data Analysis (ASMDA' 99), Lisboa, Portugal.

Informações: *Professor Helena Bacelar Nicolau*, Universidade Lisboa;
E - mail: ulfphelb@cc.fc.ul.pt
or
Professor Fernando C. Nicolau, Universidade Lisboa;
E - mail: fan@laminaria.si.fct.unl.pt.

- 06 - 09 de Junho

Annual Meeting of the Statistical Society of Canada, Regina, Saskatchewan, Canada.

Informações: Local Arrangements Chair, R. J. Tomkins, Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, SAS OA2, Canada.
E - mail: jtomkins@max.cc.uregina.ca.

- 12 - 16 de Julho

19th IFIP TC7 Conference on System Modeling and Optimization, Cambridge, UK.

Informações: E - mail: tc7con@amtp.cam.ac.uk

- 08 - 12 de Agosto

1999 Joint Statistical Meetings, Baltimore, Maryland, USA.

Informações: ASA, 1429 Duke St., Alexandria, VA 22314-3402, USA; Telf.: 703-6841221; FAX: 703-684 2037.
E - mail: meetings@asa.mhs.compuserve.com

- 10 - 18 de Agosto

International Statistical Institute, 52nd Biennial Session, Helsinki, Finland.

Informações: ISI Permanent Office, 428 Prinses Beatrixlaan, P. O. Box 950, 2270 AZ Voorburg, The Netherlands.

- 19 - 23 de Agosto

The 6th Tartu Conference on Multivariate Statistics, Satellite meeting to Helsinki ISI Session, Tartu, Estonia.

Informações: *E.-M. Tiit* or *T. Kollo*, Institute of Mathematical Statistics, University of Tartu, J. Liivi 2, EE2400, Tartu, Estonia; Telf.: 37 27 465488 / 37 27 465486; FAX: 37 27 433509.
E - mail: etit@ut.ee or kollo@ut.ec

- *13 - 16 de Setembro

44th Annual Conference of the German Society of Medical Informatics, Biometry and Epidemiology (GMDS), Heidelberg, Germany.

Informações: *Norbert Victor*, Department of Medical Biometry, Institute for Medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany
or
Lutz Edler, Biostatistics Unit, German Cancer Research Center, IM Neuenheimer Feld 280, D-69120 Heidelberg, Germany; Fax: 49 6221 564195
E - mail: GMDS-ISCB99@dkfz-heidelberg.de
or
WWW: <http://www.dkfz-heidelberg.de/biostatistics/GMDS-ISCB99>

- *13 - 17 de Setembro
Heidelberg Congress Week: Joint Conference of GMDS – ISCB 99,
Heidelberg, Germany.
Informações: *Norbert Victor*, Department of Medical Biometry, Institute for
Medical Biometry and Informatics, University of Heidelberg, Im
Neuenheimer Feld 305, D-69120 Heidelberg, Germany
or
Lutz Edler, Biostatistics Unit, German Cancer Research Center,
IM Neuenheimer Feld 280, D-69120 Heidelberg, Germany; Fax:
49 6221 564195
E - mail: GMDS-ISCB99@dkfz-heidelberg.de
or
WWW: [http://www.dkfz-heidelberg.de/biostatistics/GMDS-
ISCB99](http://www.dkfz-heidelberg.de/biostatistics/GMDS-ISCB99)

- *14 - 17 de Setembro
**20th Annual Conference of the International Society of Clinical Biostatistics
(ISCB), Heidelberg, Germany.**
Informações: *Norbert Victor*, Department of Medical Biometry, Institute for
Medical Biometry and Informatics, University of Heidelberg, Im
Neuenheimer Feld 305, D-69120 Heidelberg, Germany
or
Lutz Edler, Biostatistics Unit, German Cancer Research Center,
IM Neuenheimer Feld 280, D-69120 Heidelberg, Germany; Fax:
49 6221 564195
E - mail: GMDS-ISCB99@dkfz-heidelberg.de
or
WWW: [http://www.dkfz-heidelberg.de/biostatistics/GMDS-
ISCB99](http://www.dkfz-heidelberg.de/biostatistics/GMDS-ISCB99)

* - Novas entradas – Denotes new Entries.

**ACÇÕES DESENVOLVIDAS PELO INE NO ÂMBITO DA
COOPERAÇÃO BILATERAL E MULTILATERAL**

**ACTIONS ACHIEVED BY NSI IN THE SCOPE OF BILATERAL AND
MULTILATERAL COOPERATION**

(DE 1 DE JANEIRO A 30 DE ABRIL DE 1998):

a) *Cooperação desenvolvida com os PALOP e Macau:*

Com a presença do Secretário de Estado dos Negócios Estrangeiros e Cooperação, Dr. Luís Amado, teve início em 15 de Abril, a 9ª reunião dos Directores Gerais de Estatística dos PALOP, de Portugal e de Macau, na qual foram discutidos assuntos respeitantes à cooperação bilateral em curso entre Portugal e os Cinco, bem como a possibilidade de serem desenvolvidos novos projectos comuns, nomeadamente nas áreas dos Censos de População, dos Índices de Preços e de planeamento de médio prazo da actividade estatística nacional. De salientar o trabalho de preparação efectuado com vista à Conferência “Cooperação Estatística no quadro da CPLP”, a ter lugar em Junho de 1998, bem como sobre o segundo projecto estatístico do PIR-PALOP, no quadro do 8º FED da Convenção de Lomé IV.

No âmbito do projecto comum sobre Ficheiros de Unidades Estatísticas (Empresas e Estabelecimentos) destaca-se a instalação e teste da aplicação informática para gestão dos ficheiros de empresas no país piloto - Cabo Verde -, e em Angola, S. Tomé e Príncipe e Guiné-Bissau, bem como o fornecimento do respectivo equipamento e material informático. As experiências dos quatro países neste domínio deverão ser objecto de discussão na 3ª reunião do grupo de trabalho do projecto, a realizar em Cabo Verde, em Junho de 1998.

No que se refere ao projecto comum sobre Classificações, Conceitos e Nomenclaturas, salienta-se o avanço dos trabalhos relativos à elaboração das Classificações de Actividades Económicas da Guiné-Bissau e de S. Tomé e Príncipe que se prevê estarem concluídas no decurso do segundo trimestre deste ano. Foram também iniciados os trabalhos para a elaboração da Classificação Nacional de Bens e Serviços de Angola, através da realização de um estágio para um técnico do INE daquele país.

No período em apreço e no âmbito da cooperação bilateral com os PALOP e Macau, foram realizadas as seguintes acções :

CABO VERDE

Em continuidade do projecto de Apoio Institucional ao INE de Cabo Verde foram realizadas duas missões de assistência técnica : uma missão para preparação das Linhas Gerais da Actividade Estatística Nacional e Respectivas Prioridades, incluindo a participação num seminário sobre o tema :A Reforma do Sistema Estatístico Nacional de Cabo Verde de 1996 e as Modernas Tendências da Engenharia dos Sistemas Estatísticos Nacionais em Ambiente de Democracia Multipartidária e de Economia de Mercado, e uma missão de assistência técnica para a preparação do Plano da Actividade Estatística Nacional para 1998 - 2001.

No âmbito do projecto Índice de Preços no Consumidor foi realizado um estágio para um técnico do INE de Cabo Verde.

MOÇAMBIQUE

Teve lugar em 14 de Abril a 7ª reunião da Comissão Coordenadora da Gestão do Acordo de Cooperação Estatística Luso - Moçambicano, na qual se estabeleceu o programa de cooperação até 1999 no qual são de salientar, a continuidade do projecto de Apoio Institucional e o início de um novo projecto no domínio das Estatísticas Regionais.

SÃO TOMÉ E PRÍNCIPE

Foram realizados dois estágios no âmbito dos projectos de Contas Nacionais e de Estatísticas Demográficas e Sociais.

A 20 de Abril realizou-se a 11ª reunião da Comissão Coordenadora da Gestão do Protocolo de Cooperação Estatística entre Portugal e São Tomé e Príncipe, que acordou na actualização do programa bilateral até 1999.

Prevista a continuidade de acções nos domínios da Difusão Estatística, Contas Nacionais, Estatísticas Demográficas e Sociais, Estatísticas do Comércio Externo, Economia Informal, Agricultura e Pescas e Estatísticas de Conjuntura são consideradas como áreas em desenvolvimento, a participação no Recenseamento Geral da População, incluindo a Digitalização da Base Cartográfica, Estatísticas das Empresas e a Informática.

MACAU

Durante o período considerado foram realizadas missões de assistência técnica nos domínios das Contas Territoriais pela Óptica da Produção, Classificações Económicas e Contas Territoriais Trimestrais.

b) Cooperação desenvolvida com os PECO:

No âmbito do Programa PHARE de assistência técnica aos Países da Europa Central e Oriental (PECO), realizaram-se, no período entre 1 de Janeiro e 31 de Abril de 1998, duas acções de cooperação estatística entre o INE e os organismos de estatística da Roménia e Bulgária.

Com o INE da Roménia realizou-se um estágio na área das Contas Trimestrais, no período de 2 a 6 de Fevereiro de 1998, no Gabinete de Estudos - área económica.

A segunda acção consistiu numa missão à Bulgária (realizada no período de 23 a 27 de Fevereiro de 1998), no âmbito do projecto de cooperação PRODCOM e que contou com a participação de dois técnicos do Departamento de Estatística das Empresas (Serviço dos Inquéritos às Empresas e Serviço de Estatísticas da Indústria e Construção), um do Departamento de Coordenação e Integração (Serviço de Metodologia) e também um técnico do Gabinete de Estudos - área económica (Serviço de Inquéritos de Conjuntura).

FUNDAMENTO, OBJECTO E ÂMBITO

O INE, consciente de como uma cultura estatística é essencial para a compreensão da maioria dos fenómenos do mundo actual, e da sua responsabilidade na divulgação do conhecimento estatístico, fazendo-o chegar ao maior número possível de leitores, tendo reconhecido a necessidade de dar um passo nesse sentido, passa a editar quadrimestralmente a presente Revista de Estatística destinada a divulgar:

- a) Numa perspectiva científica, artigos originais sobre temas especializados da estatística, tanto pura como aplicada, bem como sobre estudos e análises nos domínios económico, social e demográfico;
- b) Informações sobre actividades e projectos importantes no âmbito do Sistema Estatístico Nacional;
- c) Informações sobre congressos, seminários, colóquios e conferências de interesse estatístico ou afim;
- d) Informações sobre acções desenvolvidas pelo INE no âmbito da cooperação bilateral e multilateral.

Para tal, são adoptadas as seguintes formas de contribuição para publicação na Revista:

- Quanto aos artigos referidos em a), contribuições da iniciativa dos próprios autores e por convite do Conselho Editorial, pertencentes ou não ao INE;
- Quanto às informações referidas em b), c) e d), contribuições dos departamentos do INE.

As contribuições por iniciativa dos próprios autores serão objecto de avaliação de mérito científico pelo Conselho Editorial, que decidirá ou não pela respectiva publicação.

Para a elaboração e envio das contribuições para publicação na Revista são adoptadas as Normas de Apresentação de Manuscritos que figuram na última página.

Os autores dos artigos publicados, a que se refere a alínea a), receberão uma contribuição financeira paga pelo INE, de montante a fixar por despacho da Direcção mediante proposta do Director da Revista.

OS PONTOS DE VISTA EXPRESSOS PELOS AUTORES DOS ARTIGOS PUBLICADOS NA REVISTA
NÃO REFLECTEM NECESSARIAMENTE A POSIÇÃO OFICIAL DO INE.

FOUNDATION, SUBJECT MATTER AND SCOPE

INE is conscious of how statistical awareness is essential to the understanding of the majority of phenomena in the present world and is aware of its responsibility to disseminate statistical knowledge, making it available to the widest possible range of readers. INE has recognised the need to take a step in that direction and will begin publication of this *Statistical Review* three times yearly, designed to provide the following:

- a) Within a scientific perspective, original articles on specialised areas of statistics, both pure and applied, as well as studies and analyses within the sphere of economics, social issues and demographics;
- b) Information on activities and projects within the scope of the National Statistical System;
- c) Information on congresses, seminars and conferences of a statistical or related nature;
- d) Information on activities developed by INE within the scope of bilateral or multilateral co-operation;

The following approaches for contributing material for publication in the review have been adopted:

- In relation to the articles referred to in section a), contributions are made by the authors themselves and by invitation of the Editorial Committee, whether they are employees of INE or not;
- In relation to the information referred to in section b), c) and d); contributions are from departments of INE.

The Editorial Committee who has sole discretion in deciding whether or not the material will be published will assess the scientific merit of contributions made on the initiative of the authors themselves.

The preparation and delivery of material for publication in the Review are subject to the Rules for Submitting Manuscripts presented on the last page.

The authors of the published articles referred to in section a) will receive pecuniary compensation from INE in an amount to be determined by resolution of the Board on the recommendation of the Director of the Review.

THE VIEWPOINTS EXPRESSED BY THE AUTHORS OF THE ARTICLES PUBLISHED IN THE REVIEW
DO NOT NECESSARILY REFLECT THE OFFICIAL POSITION OF I.N.E.

NORMAS DE APRESENTAÇÃO DE MANUSCRITOS

Nos termos da alínea b) do n.º 3 do Artigo 5.º do Regulamento da *Revista de Estatística* do Instituto Nacional de Estatística, o Conselho Editorial aprovou as seguintes **Normas de Apresentação de Manuscritos**:

1. Os originais dos artigos serão enviados ao Director da Revista pelos respectivos autores, devendo ser escritos em português e não terem sido ainda totalmente publicados, ou estar em processo de edição em qualquer outra publicação.
2. Poderão também ser apresentados artigos escritos em inglês, cabendo ao Director da Revista a decisão sobre a sua aceitação.
3. Quanto à *avaliação do mérito científico* dos artigos:
 - a) Os artigos apresentados por iniciativa dos respectivos autores serão submetidos à avaliação do mérito científico pelo Conselho Editorial, com garantia do anonimato tanto do autor como dos avaliadores;
 - b) Os autores receberão a informação sobre o resultado da avaliação num prazo máximo de trinta e cinco dias, com indicação, nos casos de avaliação positiva, do número da *Revista* em que serão publicados, e nos casos de avaliação negativa com a devolução do artigo apresentado e respectiva *disquette*.
4. Os artigos aceites para publicação na *Revista de Estatística* serão igualmente divulgados no *site* do INE na *Internet*.
5. Os originais, com uma extensão não superior a trinta páginas, serão processados em *Word for Windows*, integralmente a preto e branco, e entregues em suporte papel acompanhado da respectiva *disquette*.
6. Na apresentação dos originais, os autores respeitarão ainda as seguintes normas:
 - 6.1. Quanto à *estrutura*:
 - a) O texto deve ser dactilografado em formato A₄, com utilização do tipo de letra *Times New Roman* - 11, e com as seguintes margens: *top*: 2,5 cm, *bottom*: 2 cm, *left*: 2,5 cm, *right*: 5 cm;
 - b) A primeira página conterá exclusivamente o título do artigo, bem como o nome, morada e telefone do autor, com indicação das funções exercidas e da instituição a que pertence, devendo, no caso de vários autores, ser indicado a quem deverá ser dirigida a correspondência da Revista;
 - c) A segunda página conterá, em português e inglês, unicamente o título e um resumo do artigo, com um máximo de cem palavras, seguido de um parágrafo com indicação de palavras-chave até ao limite de quinze;
 - d) Na terceira página começará o texto do artigo, sendo as suas eventuais secções ou capítulos numeradas sequencialmente;
 - 6.2. Quanto a *referências bibliográficas*:
 - a) Os autores eventualmente citados no texto do artigo serão indicados entre parênteses curvos pelo seu nome seguido da data da respectiva publicação e, se for caso disso, do número de página (p. ex.: Malinvaud, 1989, 23);
 - b) As referências bibliográficas serão listadas, por ordem alfabética dos apelidos dos respectivos autores, imediatamente a seguir ao final do texto, de acordo com a fórmula seguinte:

ANDERSON, C.W., and TURKMAN, K.F, (1995) "Sums and maxima of stationary sequences with heavy tailed distributions", *Sankhya*, Vol. 57, Series A, pp.1-10.
 - 6.3. Quanto à *revisão de provas e publicação*:
 - a) Uma vez aceite o artigo e antes da sua publicação, receberá o autor dois exemplares de provas para revisão, um dos quais será devolvido ao Director da Revista no prazo máximo de uma semana contado da data da sua recepção;
 - b) Serão da responsabilidade dos respectivos autores as consequências de eventuais modificações da versão inicial aceite, bem como de atrasos na revisão das provas, que impossibilitem a publicação no número da Revista previsto, reservando-se o Conselho Editorial o direito de decidir a data da sua publicação futura;
 - c) Uma vez publicado o artigo, o autor receberá vinte exemplares da sua versão impressa e um exemplar do respectivo número da *Revista*.
7. Para informações adicionais contactar o Secretariado de Redacção:

Eduarda Liliana Martins
Instituto Nacional de Estatística
Av.ª. António José de Almeida, n.º. 5 – 9.º.
1 000 Lisbon - Portugal
Tel.: +351 1 842 61 00 (3905) Fax.: +351 1 842 63 66 e-mail: liliana.martins@ine.pt

RULES FOR SUBMITTING MANUSCRIPTS

Within the terms of sub-section a of no. 3 of Article 5 of the regulations of the *Statistical Review* of the National Statistical Institute (INE), the Editorial Committee has approved the following **Rules for Submitting Manuscripts**:

1. The original articles will be sent to the Review Director by the respective authors. They should be written in Portuguese, they should not have already been published in their entirety nor should they be in the process of being published in any other publication.
2. Articles may also be submitted in English to the Review Director who will decide whether to accept them.
3. In relation to the *evaluation of the scientific merit* of the articles:
 - a) The Editorial Committee will assess all articles submitted on the initiative of the respective authors on the basis of their scientific merit. The identity of both the author and the Committee members will be strictly confidential.
 - b) The authors will receive information regarding the results of the evaluation within a maximum period of thirty-five days. If the article is accepted, the Committee will indicate the issue number of the *Review* in which the article will be published. If the article is not accepted, it will be returned along with the respective diskette.
4. The articles accepted for publication in the *Statistical Review* will also be made public on the INE Internet site.
5. The original articles having no more than thirty pages must be processed in *Word for Windows*, completely at black and white, and they will be delivered in hard copy as well as on diskette.
6. With the presentation of the original articles, the authors must also respect the following rules:
 - 6.1 In relation to the *structure*:
 - a) The text shall be printed on A4 format paper utilising the font *Times New Roman* size 11 and with the following margins: top: 2.5 cm, bottom: 2 cm, left: 2,5 cm, right: 5 cm;
 - b) The first page shall contain only the title of the article as well as the name, address and telephone number of the author, indicating the position held and the institution that he/she belongs to. In the case of various authors, it is necessary to indicate the person to whom all correspondence received by the *Review* should be forwarded;
 - c) The second page shall contain only the title and a abstract of the article in Portuguese and English with the maximum of one hundred words followed by a paragraph indicating key words up to the limit of fifteen;
 - d) The third page will begin the text of the article with its respective sections or chapters sequentially numbered;
 - 6.2 Regarding *bibliographical references*:
 - a) Authors who are cited in the text of the article shall be indicated in parentheses with their name followed by the date of the respective publication and, if necessary, the page number (ex.: Malinvaud, 1989, 23);
 - b) All bibliographical references will be listed in alphabetical order by the surnames of the respective authors, immediately following the end of the text, as in the following example:

ANDERSON, C.W., and TURKMAN, K.F., (1995) "Sums and maxim of stationary sequences with heavy tailed distributions", *Sankhya*, Vol. 57, Series A, pp. 1-10.
 - 6.3 Regarding *proof-reading and publication*:
 - a) Once the article is accepted and prior to its publication, the author will receive two copies for review. One of these copies will be returned to the Director of the *Review* within a maximum period of one week from the date of its reception;
 - b) The consequences of subsequent changes to the accepted first version are the responsibility of the respective authors as well as any delays in proof-reading that make its publication in the planned issue of the *Review* impossible. The Editorial Committee reserves the right to decide upon the date for future publication;
 - c) Once the article is published, the author will receive twenty copies of his/her printed version and a copy of the respective issue of the *Review*.
7. For further information kindly contact the Editorial Secretary:

Eduarda Liliana Martins
Instituto Nacional de Estatística
Av.ª. António José de Almeida, n.º. 5 – 9.º.
1 000 Lisbon - Portugal
Tel.: +351 1 842 61 00 (3905) Fax.: +351 1 842 63 66 e-mail: liliana.martins@ine.pt

Nome _____ Data de nascimento: __/__/__

Profissão/Função _____ Instituição/Empresa _____

Telef.: _____ Fax: _____

DESEJO RECEBER OS EXEMPLARES DA REVISTA DE ESTATÍSTICA:

Em casa . Na Instituição/empresa

Morada para envio: _____

Localidade: _____ Código Postal: _____

Autorizo débito no cartão Visa ou Mastercard n.º:

Valor da transacção: 5540\$00 Validade do cartão __/__/__

 Junto cheque n.º _____ à ordem do *INSTITUTO NACIONAL DE ESTATÍSTICA* sobre o Banco _____

Data: __/__/__ Assinatura: _____

OS DADOS RECEBIDOS SERÃO PROCESSADOS AUTOMATICAMENTE E DESTINAM-SE AOS ENVIOS RELACIONADOS COM A SUA ASSINATURA, RESPECTIVAS OPERAÇÕES ADMINISTRATIVAS E ESTATÍSTICAS, E À EVENTUAL APRESENTAÇÃO DE OUTROS PRODUTOS E SERVIÇOS DO INSTITUTO NACIONAL DE ESTATÍSTICA.

Nome _____ Data de nascimento: __/__/__

Profissão/Função _____ Instituição/Empresa _____

Telef.: _____ Fax: _____

DESEJO RECEBER OS EXEMPLARES DA REVISTA DE ESTATÍSTICA:

Em casa . Na Instituição/empresa

Morada para envio: _____

Localidade: _____ Código Postal: _____

Autorizo débito no cartão Visa ou Mastercard n.º:

Valor da transacção: 5540\$00 Validade do cartão __/__/__

 Junto cheque n.º _____ à ordem do *INSTITUTO NACIONAL DE ESTATÍSTICA* sobre o Banco _____

Data: __/__/__ Assinatura: _____

OS DADOS RECEBIDOS SERÃO PROCESSADOS AUTOMATICAMENTE E DESTINAM-SE AOS ENVIOS RELACIONADOS COM A SUA ASSINATURA, RESPECTIVAS OPERAÇÕES ADMINISTRATIVAS E ESTATÍSTICAS, E À EVENTUAL APRESENTAÇÃO DE OUTROS PRODUTOS E SERVIÇOS DO INSTITUTO NACIONAL DE ESTATÍSTICA.



AUTORIZADO PELOS CTT
NO SERVIÇO NACIONAL

RSF
NÃO PRECISA DE SELO

INSTITUTO NACIONAL DE ESTATÍSTICA
SECÇÃO DE DISTRIBUIÇÃO
E VENDA DE PUBLICAÇÕES

Av. António José de Almeida
1000 LISBOA

AUTORIZADO PELOS CTT
NO SERVIÇO NACIONAL

RSF
NÃO PRECISA DE SELO

INSTITUTO NACIONAL DE ESTATÍSTICA
SECÇÃO DE DISTRIBUIÇÃO
E VENDA DE PUBLICAÇÕES

Av. António José de Almeida
1000 LISBOA



