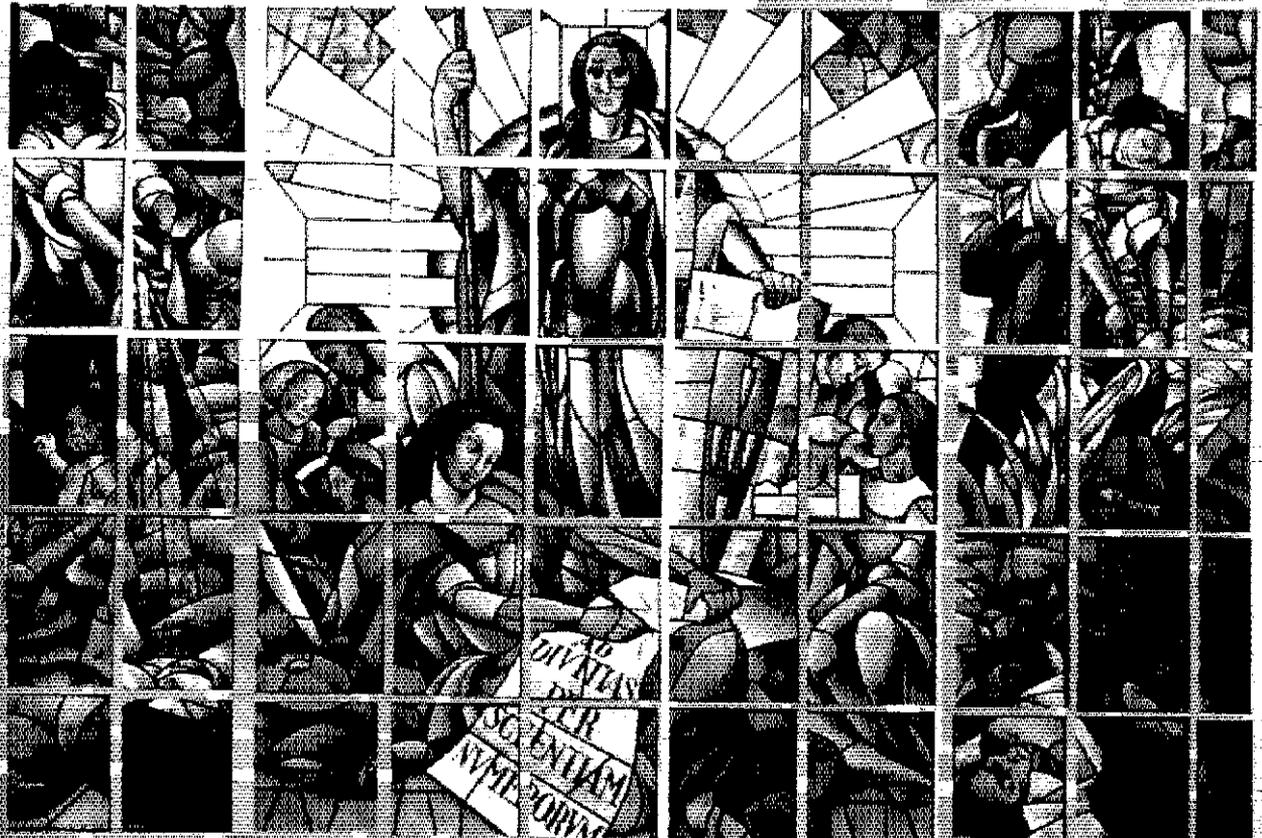




INSTITUTO NACIONAL DE ESTATÍSTICA

PORTUGAL

# REVISTA DE ESTATÍSTICA



## 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*

- *Diris Duarte Ferreira Pestana*

- *Francisco José Neto Melro*

- *João António Branco*

- *João Ferreira do Amaral*

- *Oscar Soares Barata*

- *Pedro Jorge Nunes da Silva Dias*

- *Pedro Miguel Girão Nogueira Ramos*

- *Sérgio Manuel Bancelar 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*

*Tel.: (01) 847 00 50*

*Fax: (01) 847 85 78*

### - 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*

### - TIRAGEM

- *700 exemplares*

### - DEPÓSITO LEGAL

- *N.º 99514/96*

## PREÇO

(IVA 5% incluído)

- N.º avulso 2 270\$00  
- Assinatura anual 5 450\$00

## 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) Num 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.

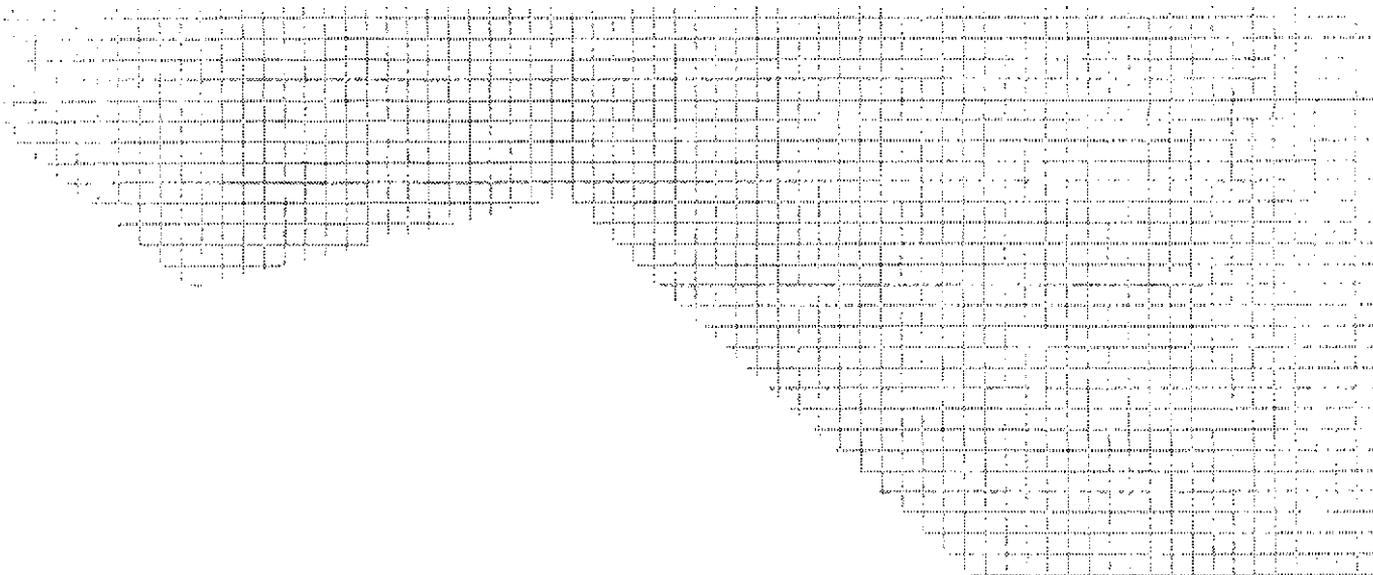
As contribuições por iniciativa dos próprios autores serão objecto de avaliação de mérito científico pelo Conselho Editorial, que decidirá ou não pela respectiva publicação.

Para a elaboração e envio das contribuições para publicação na Revista são adoptadas as Normas de Apresentação de Manuscritos que figuram no verso da contracapa.

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

NAO REFLECTEM NECESSARIAMENTE A POSIÇÃO OFICIAL DO INE.



# REVISTA DE ESTADÍSTICA



## ÍNDICE

## INDEX

-	NÓTULA DO DIRECTOR BRIEF NOTE OF THE DIRECTOR. . . . .	5
-	<i>ARTIGOS</i> <i>ARTICLES:</i>	
•	O EFEITO DE HURST E PROCESSOS COM DEPENDÊNCIA LOCAL PESADA THE HURST EFFECT AND PROCESSES WITH HEAVY LOCAL DEPENDENCE Por/By: Kamil Feridun Turkman . . . . .	7
•	O TEMPO DE PERMANÊNCIA EM REDES DE JACKSON SOJOURN TIME IN JACKSON NETWORKS Por/By: Manuel A. M. Ferreira . . . . .	25
•	A ESTATÍSTICA NA GENÉTICA – UMA REVISÃO BIBLIOGRÁFICA STATISTICS AND GENETICS – A BIBLIOGRAPHICAL REVIEW Por/By: Luzia Gonçalves e Maria Antónia Amaral Turkman. . . . .	45
•	A APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS MULTIVARIADAS NA SEGMENTAÇÃO DA PROCURA TURÍSTICA DO ALGARVE THE TOURISM DEMAND SEGMENTATION OF THE HIGH SEASON IN THE ALGARVE Por/By: Maria Margarida V. de Arrais Viegas. . . . .	63
•	EM TORNO DAS RESPOSTAS TÉCNICAS DOS ESTADISTICISTAS OFICIAIS ÀS DECISÕES POLÍTICAS SOBRE OS CRITÉRIOS DE CONVERGÊNCIA – O CASO PARTICULAR DAS CONTAS NACIONAIS ABOUT THE TECHNICAL ANSWERS FROM OFFICIAL STATISTICIANS TO THE POLITICAL DECISIONS CONCERNING THE CONVERGENCE CRITERIUMS – THE SPECIFIC CASE OF NATIONAL ACCOUNTS Por/By: Ana Maria Leal e Pedro Dias. . . . .	79
•	DETERMINANTES DO PROCESSO INFLACIONISTA ENTRE 1992 E 1997 FACTORS OF THE DISINFLATION PROCESS FROM 1992 TO 1997 Por/By: Francisco Melro. . . . .	99
•	A INTERNET COMO MEIO ALTERNATIVO DE DIFUSÃO DE INFORMAÇÃO ESTATÍSTICA THE USE OF INTERNET AS A MEAN TO DISSEMINATE STATISTICAL INFORMATION Por/By: Carlos Sebastião Afonso Dias. . . . .	115
-	<i>INFORMAÇÕES</i> <i>INFORMATIONS:</i>	
•	ACTIVIDADES E PROJECTOS IMPORTANTES NO ÂMBITO DO SISTEMA ESTATÍSTICO NACIONAL IMPORTANT ACTIVITIES AND PROJECTS IN THE SCOPE OF THE NATIONAL STATISTICAL SYSTEM. . . . .	135
•	CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS CONGRESS, SEMINARS AND CONFERENCES. . . . .	151
•	ACÇÕES DESENVOLVIDAS PELO INE NO ÂMBITO DA COOPERAÇÃO BILATERAL E MULTILATERAL ACTIONS ACHIEVED BY NSI IN THE SCOPE OF BILATERAL AND MULTILATERAL COOPERATION. . . . .	159



---

## NÓTULA DO DIRECTOR

---

---

### BRIEF NOTE OF THE DIRECTOR

---

Verificando-se que ao longo deste primeiro ano e meio de existência da Revista de Estatística, já com cinco volumes publicados, se tem vindo a registar uma melhoria progressiva do seu grau de implantação, tanto no plano nacional como no internacional, visando reforçar e alargar as condições objectivas que têm permitido esta evolução favorável, foi decidido passar a aceitar, em *casos de espécie* a decidir pelo Director da Revista, a publicação de artigos em língua inglesa.

Na verdade, se no passado primeiro o Grego e depois o *Latim* desempenharam o papel de *língua veicular universal* nos domínios científico, literário e diplomático, a que, muito mais próximo de nós, se lhes seguiu o *Francês*, sobretudo nos séculos XVIII e XIX, no nosso século tal papel tem vindo a ser desempenhado de uma maneira ainda mais efectiva pelo *Inglês*.

Não se trata, obviamente, de secundarizar o Português como língua oficial da Revista mas tão somente de materializar consequentemente o reconhecimento de que, para lhe possibilitar uma penetração mais eficaz na comunidade científica internacional, importa permitir que os autores possam escrever para a Revista na língua geralmente utilizada como língua científica veicular preferencial, superando-se deste modo naturais barreiras linguísticas.

Assim, neste número da Revista é publicado pela primeira vez um artigo em Inglês.

No sentido inverso, já a Revista vem sistematicamente praticando a titulação, em Português e em Inglês, dos artigos que publica, bem como a inserção de sumários e *palavras-chave*, em ambas as línguas. O que lança uma ponte de certo precária, mas praticável, entre os autores de língua portuguesa e quantos dominam o Inglês — ao menos como língua veicular.

Ficam os nossos votos e compromisso de diligência no sentido de que o precedente agora aberto venha a ter continuidade.

During this first year and a half of the Statistical Review existence, with five volumes already published, we have been experiencing an improvement on the degree of its acceptance, both in national and international level. So, with the purpose of strengthening and extending the conditions that allowed this favourable evolution, articles in english are now accepted for publishing in particular cases, conditioned to the decision of the Review's Director.

In fact, if in the past, first *Greek* and later *Latin* carried out the role of *universal vehicular language* in scientific, literary and diplomatic spheres, later on, mainly during the 18<sup>th</sup> and 19<sup>th</sup> centuries *French* took their places, and nowadays such a role is performed in a more effective way by the *English* language.

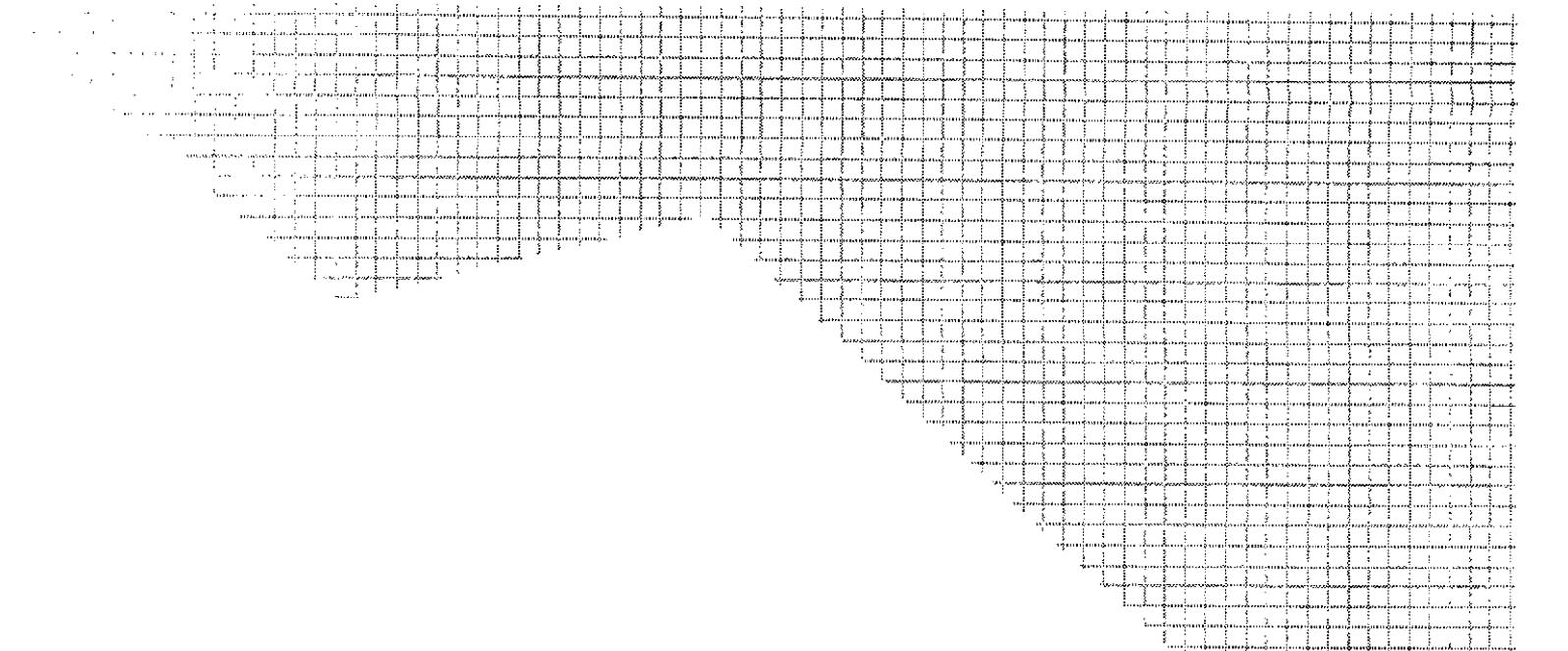
It is not obviously our intention to place portuguese in a second level, as it is the Review's official language, but to accept that the perspective for the authors to write in a language generally adopted as a vehicular scientific one, grants a deeper acknowledgement within the international scientific community and gives them de possibility of exceeding natural linguistic barriers.

Consequently, the first article in English is published in this volume.

The Review already presents its titles, abstracts and key-words in both portuguese and english languages establishing a link, yet delicate but performable, between the portuguese authors and those who prevail English — at least as a vehicular language.

We express our vows assuming the commitment of making all the efforts to assure the continuity of the procedure now undertaken.





# THE HURST EFFECT AND PROCESSES WITH HEAVY LOCAL DEPENDENCE

**Autor:**  
**Kamil Feridun Turkman**



---

## THE HURST EFFECT AND PROCESSES WITH HEAVY LOCAL DEPENDENCE

---

---

## O EFEITO DE HURST E PROCESSOS COM DEPENDÊNCIA LOCAL PESADA

---

Autor: Kamil Feridun Turkman  
Professor Catedrático  
Departamento de Estatística e Investigação Operacional da Faculdade  
de Ciências da Universidade de Lisboa

### *SUMMARY:*

- After more than 40 years, the Hurst effect or phenomenon remains an open problem in stochastic hydrology and has been a driving force in the study of long memory processes such as the fractional Brownian noise and fractionally integrated autoregressive moving average models, and was one of the sources of inspiration for the fractal theory. There has been several explanation for this phenomenon, varying from long range dependence to non-stationarity due to almost negligible trends that may exist in the processes. In this report, we give an example of a process which shows that heavy local dependence may also be an explanation for the Hurst effect.

### *KEY-WORDS:*

- *Hurst effect, long range dependence, fractional Gaussian processes.*

### *SUMÁRIO:*

- Desde há 40 anos que o efeito ou fenómeno de Hurst continua como um problema aberto em hidrologia estocástica. Tem sido, no entanto, uma força impulsionadora no estudo de processos de longa memória e foi ainda uma das fontes de inspiração da teoria dos fractais. Dependência a longo termo e não estacionaridade devido a tendências quase negligíveis que podem existir no processo, são algumas das explicações possíveis para este fenómeno. Neste trabalho damos um exemplo de um processo em que a forte dependência local pode ser uma das explicações para o efeito de Hurst.

### *PALAVRAS-CHAVE:*

- *Efeito de Hurst, dependência a longo termo, processos Gaussianos fraccionários.*



HURST COEFFICIENT AND THE RESCALED ADJUSTED RANGE

Let  $\{X_i\}_{i=1}^n$  be a sequence of random variables, representing for instance, annual river flow into a reservoir over some design span  $n$ , which can be defined as the useful life of a reservoir. Let

$$S_0 = 0, \quad S_t = \sum_{i=1}^t X_i, \quad t = 1, 2, \dots, n.$$

If  $\frac{1}{n}S_n$  represents the ideal release from the reservoir at any year then  $\frac{t}{n}S_n$  is

the ideal amount of water released during the first  $t$  years of the operation and hence

$$S_{t,n} = S_t - \frac{t}{n}S_n$$

is the difference between the cumulative inflow and cumulative outflow over the first  $t$  years, representing the fluctuations in the water content of this ideal reservoir. The adjusted range

$$\begin{aligned} R_n &= \max_{1 \leq t \leq n} S_{t,n} - \min_{1 \leq t \leq n} S_{t,n} \\ &= M_n - m_n, \quad \text{say} \end{aligned}$$

is a measure of the reservoir capacity under the idealized condition that water be withdrawn from the reservoir at an annual rate of  $\frac{1}{n}S_n$  units per year. In fact,  $R_n$  is the minimum reservoir capacity required to satisfy a constant water usage of  $\frac{1}{n}S_n$  without experiencing shortages or spills over the period spanned by the inflow sequence  $\{X_i\}$ . The range  $R_n$  is then rescaled by the sample standard deviation  $s_n$ ,

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

to give a dimensionless reservoir capacity design parameter

$$R_n^* = R_n / s_n,$$

which is called the rescaled adjusted range. Given a useful life span of  $n$  for a dam, it is of importance to design the reservoir capacity as a function of the life span  $n$ . It is clear from the definitions that the adjusted rescaled range will grow with  $n$  and the

Hurst effect or phenomenon addresses the rate of growth in  $R_n^*$  as a function of  $n$ . From a probabilistic point of view, the Hurst effect concerns with the asymptotic stability of  $R_n^*$ , as  $n \rightarrow \infty$ , under certain conditions on the sample record  $X_1, X_2, \dots, X_n$  to be explained later.

The first study on the stability of  $R_n^*$  was carried by Hurst (1951), a hydrologist working on the long term storage capacity of the Aswan dam on the Nile river. Since there are rare historic river flow records exceeding 100 years, Hurst(1951) studied some 690 different annual time series of 75 geophysical variables such as temperature, rainfall, solar spot numbers, atmospheric pressure, tree-ring growth, mud-varve thickness, wheat prices etc. in connection with regions of the Nile river, to come up with the conclusion that

$$R_n^* \sim n^h,$$

where  $h$  is generally known as the Hurst coefficient. Specifically, Hurst reported that the plots of  $\log R_n^*$  versus  $\log n$  (the so called pox diagram) indicated that  $R_n^*$  is asymptotically proportional to  $n^h$  with  $h$  strictly greater than 0.5, with typical values of  $h$  closely scattered around 0.73 with standard deviation of 0.09. An complete account of Hurst's results can be found in Hurst, Black and Simaika (1965).

In order to back up his empirical results, Hurst (1951) used some coin tossing experiments to develop the theoretical asymptotic relationship for the expected value of  $R_n^*$  as

$$E[R_n^*] = 1.2533n^{1/2}. \quad (1)$$

Almost immediately upon learning Hurst's results, Feller (1951) proved rigorously that  $n^{-1/2}R_n^*$  converges to a non-degenerate random variable in distribution, as  $n \rightarrow \infty$ , under the condition that  $\{X_i\}$  is an iid sequence. Now it is known as a consequence of various functional central limit theorems that the result of Feller extends to other stationary sequences with weak dependence. In fact, under quite general standard conditions of stationarity, weak dependence and the condition that the associated iid sequence is in the normal domain of attraction,  $n^{-1/2}R_n^*$  converges to a random variable with mean

$$(\theta\pi/2)^{1/2}, \quad (2)$$

and variance

$$\theta(\pi^2/6 - \pi/2), \quad (3)$$

where

$$\theta = \sum_{i=-\infty}^{\infty} \rho_i,$$

$\rho_i$  being the correlation function of the sequence  $\{X_i\}$ , assumed to be summable. (See for example, Ibrahimov, 1962 and Billingsley, 1968).

The discrepancy between theoretical results which indicate that under fairly reasonable conditions the Hurst coefficient should be  $1/2$  and the empirical evidence shown by Hurst and subsequent scientists which indicate that,  $h > 1/2$ , ( $h \sim 0.73$  precisely) is generally referred to as the Hurst effect or phenomenon. Since then, there has been an intense theoretical interest in explaining the Hurst effect.

Several definitions of the Hurst effect are given:

**Definition 1:** A stochastic process is said to possess a Hurst effect if the sequence  $n^{-h}R_n^*$  converges in distribution to a limit random variable which may be degenerate but not almost surely 0 or  $\infty$  for some  $h > 1/2$ .

**Definition 2:** A stochastic process is said to possess a Hurst effect  $n^{-h}E[R_n^*]$  converges to a constant in  $(0, \infty)$  for some  $h > 1/2$ .

Mandelbrot (1975) showed that the Hurst constant in theory can take any value in  $[0, 1]$  and hence the most general definition of the Hurst effect is given by the following definition.

**Definition 3:** A stochastic process is said to possess a Hurst effect if the sequence  $n^{-h}R_n^*$  converges in distribution to a limit random variable which may be degenerate but not almost surely 0 or  $\infty$  for some  $h \in [0, 1]$ ,  $h \neq 1/2$ .

It is generally accepted that a stochastic process has the Hurst effect if  $h > 1/2$ , since this is connected to long range dependence concept as we will explain in the next section. However, it is of interest to know when  $h$  can take values less than  $1/2$ . In section IV, we will give an example of a process which has  $h$  strictly less than  $1/2$ . This example will show that contrary to long range dependence that produces Hurst constants greater than  $1/2$ , it is the heavy short range dependence that produces Hurst constants less than  $1/2$ .

## 2. POSSIBLE EXPLANATIONS FOR THE HURST EFFECT

The Hurst effect can only be explained by the violation of at least one of the conditions of the central limit theorem. As a consequence, various authors have proposed different explanations related to the violation of these hypothesis, essentially related to the dependence structure of the process, to its stationarity, to the existence of infinite variance and to the penultimate behavior of the limiting random variable.

First stronger dependence arguments were initiated by Feller (1951) who suggested that possible Markovian character of the process may account for the Hurst effect. However, this suggestion short lived and Barnard(1956) showed that the Hurst coefficient is 0.5 for some Markovian processes. Indeed, Mandelbrot and Von Ness (1968) showed that as long as a process is in the domain of attraction of the Brownian motion,  $R_n^*$  grows in the order of  $n^{1/2}$ , irrespective of the dependence structure. Motivated by the possible non-summability of the correlation function in [2], they constructed a class of processes, called fractional Brownian noise, which is capable of exhibiting very long range dependence and which possesses the Hurst effect as a

function of the scale of this long range dependence. (Mandelbrot and Wallis, 1968, 1969, Mandelbrot and Van Ness, 1968).

The fractional Brownian motion of exponent  $H$  ( $0 < H < 1$ ) is a Gaussian random function  $B_H(t)$  with stationary increments such that  $B_H(0) = 0$ ,

$$\text{cov}[B_H(t_1)B_H(t_2)] = \frac{1}{2} [|t_1|^{2H} - |t_1 - t_2|^{2H} - |t_2|^{2H}] \sigma^2,$$

and in particular,  $EB_H^2(t) = t^{2H} \sigma^2$ , where,  $\text{Var}[B_H(1)] = 1$  When  $H = 1/2$ , this process becomes the ordinary Brownian motion. Mandelbrot and Van Ness (1968) calls this process fractional Brownian motion, because  $B_H(t)$  is a fractional integral of the Brownian motion. The fractional Brownian noise with index  $H$  is a process formed from the increments of a stationary fractional Brownian motion. Simple arguments show that the fractional Brownian noise  $X_H(t)$  has the following properties:

- (i)  $X_H(t)$  is strictly stationary,
- (ii)  $S_n = \sum_{t=1}^n X_H(t) = B_H(n)$  which implies that  $\text{Var}(S_n) = n^{2H} \sigma^2$ ,
- (iii)  $\text{cov}[X_H(t), X_H(t+k)] \sim \sigma^2 H(2H-1) k^{2H-2}$ , for large  $k$  and  $H \neq 1/2$ .

Note that the covariances of such a process are not summable and as  $h \rightarrow 1$ ,  $X_h(t)$  is a fully correlated Gaussian process. In fact, Mandelbrot and Van Ness (1968) show that where as  $X_{1/2}(t)$  is strong mixing, the process  $X_h(t)$ , for  $H \neq 1/2$  is not strong mixing.

Starting from a fractional Gaussian noise process with index  $H \neq 1/2$  and defining  $R_n^*$  the usual way, one sees that the partial sum process converges weakly to the fractional Brownian motion and by the continuous mapping theorem, one obtains the Hurst effect for the  $X_H(t)$  process with the Hurst coefficient  $h$  being equal to the exponent  $H$  of the process. Thus, fractional Brownian noise has rightly become the prototype model for the Hurst effect, and the long range dependence is accepted as the probable cause for the appearance of the Hurst phenomenon. However, fractional Gaussian noise is not quite convenient for modeling purposes, since they are defined on continuous times whereas their realizations are on discrete time. Granger and Joyeux (1980) and Hosking (1981) introduced another class of discrete time processes, called fractionally differenced processes which are in the domain of attraction of the fractional Brownian motion and are capable of modeling long range dependence and preserve the Hurst effect. This class of models are more manageable from a point of modeling and generalizes the linear auto regressive integrated moving average (ARIMA) models.

A discrete time process  $X_t$  is said to be fractionally differenced  $ARIMA(p, d, q)$

if

$$\Phi_p(B) \nabla^d X_t = \Theta_q(B) \epsilon_t,$$

where  $\varepsilon_t$  is a white noise process,  $\Phi_p(B)$  and  $\Theta_q(B)$  are respectively autoregressive and moving average polynomials of order  $p$  and  $q$   $d > -1$  is a real number such that

$$\begin{aligned} \nabla^d &= (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\ &= \sum_{j=0}^{\infty} \pi_j B^j, \end{aligned}$$

where,

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0 \leq k \leq j} \frac{k-1-d}{k}.$$

Fractional Brownian noise and fractionally differenced  $ARIMA(p, d, q)$  processes are quite similar. In fact, Geweke and Porter-Hudak (1983) show that  $X_t$  is an  $ARIMA(0, d, 0)$  process iff it is a fractional Brownian noise with  $H = d + 1/2$ .

Although fractionally differenced ARIMA models and fractional Gaussian noise give a very elegant explanation to the Hurst effect and give convenient means of modeling long range dependence, their role is very controversial among hydrologists. Klemes (1981) claims that the existence of long range dependence in geophysical records is not supported neither on physical grounds nor empirical evidence and he argues that there must be other explanations other than long range dependence.

One possible explanation of the Hurst effect is the penultimate behavior of  $R_n^*$ . It is possible to construct mathematical examples in which convergence to the asymptotic growth rate  $n^{1/2}$  is very slow. Therefore the empirical observations for finite sample sizes may give Hurst constants larger than 0.5. Gomide (1978) and Salas et al (1979) show the existence of penultimate behavior in Markovian and  $ARMA(1, 1)$  processes.

Other explanation of the Hurst effect is the non-stationarity of geophysical processes. Hurst (1951) was the first to recognize that non-stationarity may effect the Hurst constant. By simulating series which had sudden changes in the mean (he uses decks of playing cards to simulate fairly complicated processes), he obtained empirical estimates of  $h$  around 0.71. Other simulations from non stationary processes produced empirical series exhibiting Hurst effect (Klemes, 1974, Potter, 1976, Boes and Salas, 1978). However, as Boes and Salas (1978) points out, all the above examples are doubly stochastic processes which are conditionally non-stationary, and in the long run behave like stationary  $ARMA(1, 1)$  processes and hence are in the domain of attraction of the Brownian motion. Hence in principle, these processes should have a Hurst coefficient of 0.5, indicating a possible penultimate behavior.

Contrary to the above examples, Bhattacharya et al (1983) give an example which clearly shows the effect of non-stationarity on the Hurst constant. They consider a sequence  $\{X_n\}$  of the form

$$X_n = Y_n + c(m+n)^\beta,$$

where  $Y_n$  is a stationary, weakly dependent sequence having 0 mean and variance 1.  $m, c$  are positive parameters and  $\beta \leq 0$ . The term  $c(m+n)^\beta$  can be seen as a small monotonic trend, converging to 0 as  $n \rightarrow \infty$ . Bhattacharya et al (1983) show that such processes would exhibit the Hurst effect with the Hurst constant  $h$  depending on the values of  $\beta$  as follows: For  $-1/2 < \beta < 0$ ,  $h = 1 + \beta$ , for  $\beta = 0$ ,  $h = 1$ , and for  $\beta \leq -1/2$ ,  $h = 1/2$ .

This example, shows that almost unnoticeable trends may cause the Hurst effect and are accepted as a plausible reason for geophysical series to possess the Hurst effect.

It is generally accepted that sequences belonging to the stable domain of attraction with  $\alpha < 2$  or belonging to the max-stable domain of attraction with  $\alpha < 2$ , quite often give a good fit to geophysical series and several authors (Moran, 1964, Boes and salas, 1973) proposed to explain the Hurst effect by the possible existence of infinite variance. Soon this approach was dropped, when Mandelbrot (1975) showed that iid sequences with infinite variance have Hurst constant  $h = 0.5$ . However, Ballerini and Boes (1985) exhibit a class of dependent sequences which they call shifting level processes, and then they use this class to construct stationary sequences with infinite second moments, belonging to the stable domain of attraction with index  $\alpha \in (1, 2)$  and possessing a Hurst effect with the Hurst constant  $h = 1/\alpha$ . Ballerini and Boes (1985) further claim that sequences with infinite variances belonging to stable domain of attraction with  $\alpha < 1$  would not possess the Hurst effect and that they would have the Hurst constant  $h = 1/2$  irrespective of dependence structure. In section 4, we exhibit an example of a stationary 1-dependent sequence with infinite variance belonging to stable domain of attraction with  $\alpha < 1$  having a Hurst constant  $h < 1/2$ . In this case, it will be seen that the Hurst effect can be caused by a special type of heavy local dependence.

---

### 3. SOME MATHEMATICS OF THE HURST EFFECT

---

In this section, we give some known limit theorems for weakly and strongly dependent processes which explain the Hurst effect.

#### Theorem 1

Domains of attraction of the Brownian motion. (Mandelbrot, 1975)

- (i) If  $\{X_i\}$  is iid and  $EX^2 < \infty$ , then  $n^{-1/2}R_n^*$  converges to a non-degenerate random variable in distribution.
- (ii) If  $\{X_i\}$  is any sequence belonging to the domain of attraction of the Brownian motion, then  $n^{-1/2}R_n^*$  also converges to a non-degenerate random variable.

#### Theorem 2

General domains of attraction (Mandelbrot, 1975)

- (i) Assume that there are normalizing constants  $b_n$  and  $c_n$  such that the random functions

(ii)

$$U(t, n) = b_n^{-1} \sum_{i=1}^{[tn]} X_i, \quad 0 < t \leq 1, \quad (4)$$

$$V(t, n) = c_n^{-1} \sum_{i=1}^{[tn]} X_i^2, \quad 0 < t \leq 1, \quad (5)$$

converge weakly as  $n \rightarrow \infty$  to  $U(t, \infty)$ ,  $V(t, \infty)$ , not identically equal to 0 or  $\infty$  belonging to the space  $D$  of right continuous real valued functions on  $(0, \infty)$ . Then

$$b_n = n^h L_1(n), \quad c_n = n^{h_2} L_2(n),$$

where  $L_1(n)$ ,  $L_2(n)$  are slowly varying functions and

$$a_n = n^{1/2} \frac{b_n}{c_n} = n^h L(n)$$

with  $h = 1/2 + h_1 - 1/2h_2$  and  $L(n)$  is a slowly varying function.

(iii) If  $h < 1$  or  $h = 1$ ,  $L(n) \rightarrow 0$ , as  $n \rightarrow \infty$  and

$$U(t, \infty) \neq tU(1, \infty),$$

then  $a_n^h R_n^*$  converges to a non-degenerate random variable and the Hurst constant is given by

$$h = \frac{1}{2} + h_1 - \frac{1}{2}h_2.$$

Following theorem is useful in studying the limiting behavior of  $n^h R_n^*$  when the variance of the sequence is finite.

### Theorem 3

(Leiva, 1983, Ballerini and Boes, 1985)

(i) The mapping  $f: D[0,1] \rightarrow R^+$  defined by

$$f(X(\cdot)) = \sup_{0 \leq t \leq 1} (X(t) - tX(1)) - \inf_{0 \leq t \leq 1} (X(t) - tX(1))$$

is continuous with respect to the skorohod metric in  $D[0,1]$  and the usual metric on the line. Hence if  $S_n = \sum_{i=1}^n X_i$  converges weakly with appropriate normalizing constants  $b_n$  to a non-degenerate random variable  $X_\alpha$  then

$$f(X_n) = b_n^{-1} R_n^*$$

converges to  $f(X_\alpha)$ .

- (ii) If  $EX^2 < \infty$  then from ergodic theorem  $S_n$  converges a.s to a positive constant and from Slutsky's theorem  $b_n^{-1}R_n^*$  converges weakly.

---

#### 4. AN EXAMPLE OF A SEQUENCE WITH INFINITE VARIANCE POSSESSING THE HURST EFFECT

---

In this section we give an example of a weakly dependent sequence with infinite variance possessing a Hurst effect with the Hurst constant  $h < 1/2$ . We will also show that this behavior is due to a very special heavy local dependence. Results from Davis and Hsing (1994) indicate that once such special local dependence is excluded, under quite general conditions, stationary sequences with infinite variances would have the Hurst constant  $h = 1/2$ .

**Example** (Anderson and Turkman, 1995)

Let  $Y_i, i = 0, 1, \dots$  be an independent stationary sequence with marginal distribution function

$$F_y(y) = 1 - y^{-\varepsilon}, \quad y \geq 1 \quad (0 < \varepsilon < 1)$$

and let  $p$  denote the function

$$p(y) = \begin{cases} 1 - y^{-\eta} & y \geq 1, (\eta > 0, \varepsilon + \eta < 1), \\ 0 & y < 1. \end{cases}$$

Given  $\{X_i\}$  define conditionally independent indices  $J_i, i = 1, 2, \dots$  by

$$J_i = \begin{cases} 1 & \text{with probability } p(Y_{i-1}) \\ 0 & \text{with probability } 1 - p(Y_{i-1}) \end{cases}$$

and set

$$X_i = \begin{cases} Y_i & \text{if } J_i = 0 \\ -Y_{i-1} & \text{if } J_i = 1, \end{cases}$$

for  $i = 1, 2, \dots$ . Evidently as long as the  $J_i$  take the value 0, the  $X$ -process will coincide with the  $Y$ -process, and this is likely to be the case as long as the  $Y$ -values remain not too large. A large  $Y$ , on the other hand, is likely to result in the next  $J$  taking the value 1, and the corresponding  $X$  being the negative of its predecessor. Thus in the  $X$  process there is a tendency for large positive values to be immediately cancelled out.

The  $X$ -process is clearly stationary and 1-dependent, and its marginal distribution function  $F_X$  has the properties, as  $x \rightarrow \infty$ ,

$$1 - F_X(x) = \varepsilon (2_\varepsilon + \eta)^{-1} x^{-\varepsilon},$$

$$F_X(-x) \sim x^{-\varepsilon}$$

so that the joint limit distribution of sum and maximum  $(\tilde{S}_n, \tilde{M}_n)$  for the equivalent independent sequence would be given by  $(U, V)$  where

$$E(e^{itU} \chi(V \leq v)) = W_\alpha(t, p) \exp \int_v^\infty e^{ikw} dw^{-\alpha}, \quad (6)$$

with  $\alpha = \varepsilon$  and  $p = \varepsilon(2_\varepsilon + \eta)^{-1}$ , where  $\chi(A)$  denotes the indicator function of the event  $A$ ,  $W_\alpha(t, p)$  is the characteristic function of a stable law of index  $\alpha$  and parameter  $p$  and  $k$  is a constant depending on  $\alpha$  and  $p$  (Chow and Teugels, 1978). We note that the components in this case would be dependent.

The behavior of the  $X$ -process can be analyzed by noting that the process regenerates after each negative  $X_i$ . Moreover, conditionally on the time  $K$  between two successive regenerations being equal to  $k$ , the  $k-1$   $X$ -values after the regeneration are independent, the first  $k-2$  having conditional probability density function

$$h(x) \propto x^{-\varepsilon-\eta-1}, \quad x \geq 1,$$

and the  $k-1$ th having conditional probability density

$$g(x) \propto (1 - x^{-\eta})x^{-\varepsilon-1}, \quad x \geq 1,$$

whilst the  $k$ th is identically the negative of the  $k-1$ th. The distribution of  $K$  is geometric with

$$P(K = k) = \varepsilon \eta^{k-2} (\varepsilon + \eta)^{-k+1}, \quad k = 2, 3, \dots$$

If  $S^*$  and  $M^*$  denote respectively the sum and maximum over a regenerative life it follows that

$$E(e^{itS^*} \chi[M^* \leq v] | K = k) = \Phi^{k-2}(t, v) G(v),$$

where

$$\Phi(t, v) = \int_1^v e^{itx} h(x) dx.$$

And

$$G(v) = \int_1^v g(x) dx$$

Unconditionally, therefore

$$E\left(e^{itS^*} \chi\left[M^* \leq v\right]\right) = \varepsilon G(v) / (\varepsilon + \eta - \eta\Phi(t, v))$$

and so, with

$$a_n = n^{1/(\varepsilon+\eta)}, \quad c_n = n^{1/\varepsilon},$$

$$\lim_{n \rightarrow \infty} E\left(e^{it \sum_{j=1}^n S_j^*/a_n} \chi\left[\max_{1 \leq j \leq n} M_j^* \leq c_n v\right]\right) = W_{\varepsilon+\eta}^{\eta/\varepsilon}(t, 1) \Phi_{\varepsilon}^{1+\eta/\varepsilon}(v),$$

where  $\{S_j^*\}$  and  $\{M_j^*\}$  are independent copies of  $S^*$  and  $M^*$  respectively.

A standard argument in which  $S_n$  and  $M_n$  are approximated by the sum and maximum up to the last regeneration before  $n$  now shows that  $(S_n/a_n, M_n/c_n)$  has the same limit distribution, and so

$$\lim_{n \rightarrow \infty} E\left(e^{itS_n/a_n} \chi\left[M_n/c_n \leq v\right]\right) = W_{\varepsilon+\eta}^{\eta/\varepsilon}(t, 1) \Phi_{\varepsilon}^{1+\eta/\varepsilon}(v).$$

For the  $X$ -process therefore the limit distribution has independent components. Hence due to the specific cancellation of large positive values by large negative values, the sum of this sequence converges to a non-degenerate random variable with normalization

$a_n = n^{-1/(\varepsilon+\eta)}$ , contrary to the associated iid sequence which would require normalization  $a_n^* = n^{-1/\varepsilon}$ .

A standard argument using the above regenerative properties can now be applied to show that the time normalized sums

$$n^{-1/(\varepsilon+\eta)} S_{[nt]} = n^{-1/(\varepsilon+\eta)} \sum_{i=1}^{[nt]} X_i, \quad 0 < t \leq 1$$

converges to a stable process with index  $\alpha$ .

Now, once the weak convergence of the process

$$n^{-2/\varepsilon} \sum_{i=1}^{[nt]} X_i^2, \quad 0 < t \leq 1 \tag{7}$$

is established, then Theorem 2 can be applied to show that  $a_n^{-1} R_n^*$  converges to a non-degenerate random variable with  $a_n = n^h$ , where

$$\begin{aligned} h &= \frac{1}{2} + 1/(\varepsilon + \eta) - 1/\varepsilon \\ &= \frac{1}{2} - \frac{\eta}{\varepsilon(\varepsilon + \eta)}. \end{aligned}$$

Note that  $0 < \eta/(\varepsilon + \eta) < 1/2$  and hence the  $X$ -process of the example has a Hurst constant strictly less than  $1/2$ .

Now consider the sequence  $X_n^2$ . This sequence would not have the cancellation of the large negative and positive values and with similar arguments given for the  $X$ -process, (8) can be established. However, following result due to Davis and Hsing (1994) establishes (8) in a much more general set up:

Let  $X_n$  be a stationary sequence of random variables with regularly varying tail probabilities satisfying

$$P[|X_1| > x] = X^{-\alpha} L(x),$$

where  $\alpha \in (0, 2)$ ,  $L(x)$  is slowly varying and

$$\frac{P[X_1 > x]}{P[|X_1| > x]} \rightarrow p, \quad \frac{P[X_1 < x]}{P[|X_1| > x]} \rightarrow q,$$

as  $x \rightarrow \infty$  with  $0 \leq p \leq 1$  and  $p + q = 1$ . Let  $\alpha_n$  be the  $1 - n^{-1}$  quantile of the distribution of  $|X_1|$ . Then

$$\alpha_n = n^{1/\alpha} L(n).$$

#### Theorem 4

Assume that  $X_n$  satisfies the mild mixing condition  $A(\{\alpha_n\})$  condition (Davis and Hsing, 1994) and further assume that

- (i) For some finite positive constant  $\gamma$ ,

$$k_n P\left[\max_{1 \leq j \leq r_n} |X_k| > \alpha_n x\right] \rightarrow \gamma x^{-\alpha}, \quad x > 0$$

for some sequences  $k_n, r_n$  such that  $r_n \rightarrow \infty, r_n/n \rightarrow 0$ , as  $n \rightarrow \infty$  and  $k_n = \lfloor n/r_n \rfloor$ .

- (ii) For some probability measure  $Q$ ,

$$P\left[\sum_{j=1}^{r_n} \varepsilon_{(\max_{1 \leq k \leq r_n} |X_k|)} \in / \max_{1 \leq k \leq r_n} |X_k| > \alpha_n x\right] \rightarrow Q, \quad x > 0,$$

where

$$\varepsilon_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Then the point process

$$N_n = \sum_{k=1}^{\infty} \mathcal{E}_{(k/n, a_n^{-1} X_k)}$$

converges weakly to a point process  $N$  on  $(0, 1] \times R - \{0\}$ , where

$$N = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \mathcal{E}_{(t_k, P_{ij} Q_{ij})},$$

where,  $\sum_{k=1}^{\infty} \mathcal{E}_{(t_k, P_i)}$  is a Poisson processes on  $(0, 1] \times R - \{0\}$ , with intensity  $\nu(dt \times dx) = dt \times \gamma \alpha x^{-\alpha-1} dx$ , and  $\sum_{j=1}^{\infty} \mathcal{E}_{Q_{ij}}$  are mutually independent, identical point processes with distribution  $Q$ .

Note that  $X^2$ -sequence satisfies all the conditions of Theorem 4 with  $\alpha_n = n^{2/\varepsilon}$ .

Let

$$\hat{T}: M_p((0, \infty] \times R - \{0\}) \rightarrow R$$

be a map defined on the space of all positive counting measures  $(0, \infty) \times R - \{0\}$  such that

$$\hat{T}\left(\sum_{k=1}^{\infty} \mathcal{E}_{(k/n, a_n^{-1} X_k^2)}\right) = \sum_{k/n \leq t} a_n^{-1} X_k^2$$

When  $0 < \alpha < 1$  the map  $\hat{T}$  is a.s. continuous with respect to the distribution of the point process

$$\sum_{k=1}^{\infty} \mathcal{E}_{(k/n, a_n^{-1} X_k^2)}$$

(Davis and Hsing, 1994 and Resnick, 1986) and hence by the continuous mapping theorem,

$$\begin{aligned} n^{-2/\varepsilon} \sum_{i=1}^{\lfloor nt \rfloor} X_i^2 &= \hat{T}\left(\sum_{k=1}^{\infty} \mathcal{E}_{(k/n, a_n^{-1} X_k^2)}\right) \\ &\rightarrow \hat{T}(N), \end{aligned}$$

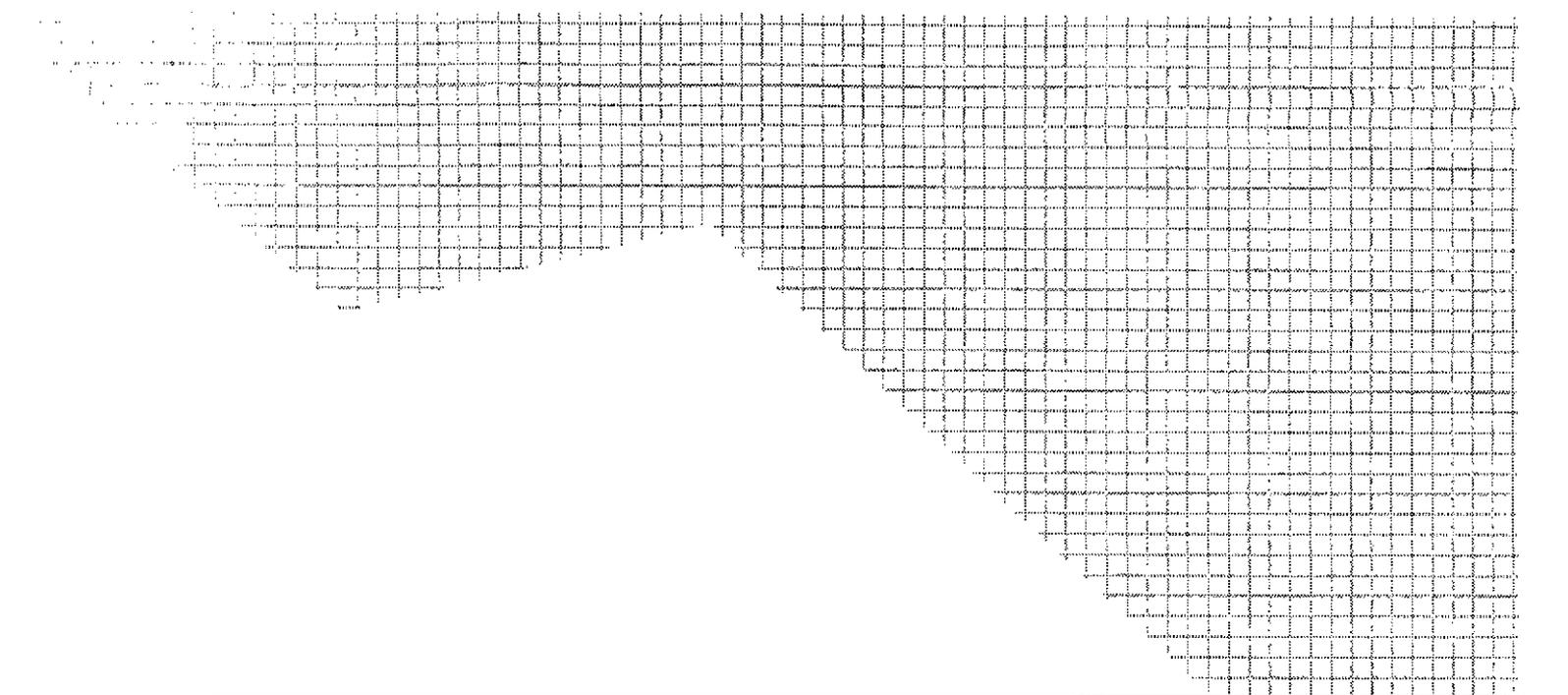
which is a well defined stable process.

Davis and Hsing (1994) result shows that once this specific form of cancellation of large terms is excluded, under quite general conditions, sequences with infinite variances belonging to a stable  $\alpha$ -domain of attraction can not possess the Hurst effect, agreeing with the Ballerini and Boes (1985) conjecture.

## REFERÊNCIAS BIBLIOGRÁFICAS

- 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.
- BALLERINI, R., and BOES, D.C., (1985) "Hurst behavior of shifting level processes" *Water Resour. Res.*, 21(11) 1642-1648
- BARNARD, G.A., (1956) "Discussion of Hurst", *Proc. Inst. Civ. Eng.*, 5, 552-553
- BHATTACHARYA, R.N., GUPTA, V.K., and WAYMIRE, E., (1983) "The Hurst effect under trends", *J. Appl. Probab.*, 20, 649-662
- BILLINGSLEY, P. (1968) *Convergence of Probability measures*. J. Wiley, New York.
- BOES, D.C., and SALAS, J.D., (1978) "Non stationarity of the mean and the Hurst phenomenon", *Water Resour. Res.*, 14(1), 135-143
- BOES, D.C., and SALAS, J.D., (1973) "On the expected range and expected adjusted range of partial sums of exchangeable random variables", *J. Appl. Probab.*, 10, 671-677
- CHOW, T.L. and TEUGELS, J.L., (1978) "The sum and the maximum of i.i.d random variables". In *Proc. Second Prague Symp. Asymptotic Statistics*, 81-92.
- DAVIS, R.A. and HSING, T., (1994) "Point processes and Partial sum convergence for weakly dependent random variables with infinite variance. Technical research report".
- FELLER, W., (1951) "The asymptotic distribution of the range of sums of independent random variables". *Ann. Math. Statist.* 22, 427-432
- GEWEKE, J., and PORTER-HUDAK, S., (1983) "The estimation and application of Long Memory Time series Models", *J. Time Ser. Anal.*, 4, 221-238
- GOMIDE, F.L.S. (1978) "Markovian inputs and the Hurst phenomenon". *J. Hydrol.*, 37, 23-45
- GRANGER, C.W.J., and JOYEUX, R., (1980) "An introduction to long memory time series models and fractional differencing." *J. Time Ser. Anal.*, 1(1), 15-29
- HOSKING, J.R.M., (1981) "Fractional differencing", *Biometrika*, 68, 165-176
- HURST, H.E., BLACK, R.P., SIMAIKA, Y.M., (1965) "Long-term Storage: An experimental study", *Constable, London*
- HURST, H.E., (1951) "Long-term storage capacity of Reservoirs". *Trans. Amer. Soc. Civil Eng.*, 116, 770-808.
- IBRAHIMOV, I.A., (1962) "Some limit theorems for stationary processes", *Theor. Probl. Appl.*, 7, 349-382
- KLEMES, V., SRIKANTHAN, R., and MCMAHON, T.A., (1981) "Long-memory flow models in reservoir analysis: What is their practical value?" *Water Resour. Res.*, 17(3) 737-751
- KLEMES, V. (1974) "The Hurst phenomenon: A puzzle?" *Water Resour. Res.*, 10(4), 675-688
- LEIVA, R. (1983) "Properties, convergence and range behavior of shifting level processes", Ph.D. dissertation, Dept. of Stat., Colo. State Univ., Fort Collins.

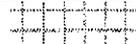
- LLOYD, E.H., (1967) "Stochastic reservoir theory", in *Advances in Hydro-science*, V.T. Chow (ed), 4, 281-339, Academic, San Diego, Calif.
- MANDELBROT, B.B. and WALLIS J., (1968) "Noah, Joseph, and operational hydrology". *Water Resour. Res.*, 4(5) 909-918.
- MANDELBROT, B.B. and WALLIS, J., (1969) "Robustness of rescaled range R/S in measurement of non cyclic long run statistical dependence"., *Water Resour. Res.*, 5(5) 967-988
- MANDELBROT, B.B., (1975) "Limit theorems on the self-normalized range for weakly and strongly dependent processes". *Z. Wahrschein. Gebiete*, 31, 271-285
- MANDELBROT, B.B. and VAN NESS, J.W., (1968) "Fractional Brownian motions, Fractional Gaussian noises and applications" *SIAM Rev.*, 10(4), 422-437
- MORAN, P.A.P., (1964) "On the range of cumulative sums", *Ann. Inst. Statist. Math.*, 16, 109-112
- POTTER, K.W., (1976) "Evidence for non stationarity as a physical explanation of the Hurst phenomenon", *Water Resour. Res.* 12(5), 1047-1052
- RESNICK, S., (1986) "oint processes, regular variation and weak convergence", *Adv. Appl. Prob.*, 18, 66-138
- SALAS, J.D., BOES, D.C., YEVJEVICH, V. and PEGRAM, G.G.S., (1979) "Hurst phenomenon as a pre asymptotic behavior", *J. Hydrol.*, 44, 1-15



# O TEMPO DE PERMANÊNCIA EM REDES DE JACKSON

**Autor:**  
**Manuel A. M. Ferreira**





---

---

## O TEMPO DE PERMANÊNCIA EM REDES DE JACKSON

---

---

---

---

### SOJOURN TIME IN JACKSON NETWORKS

---

---

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

**SUMÁRIO:**

- As redes de filas de espera de Jackson têm muitas aplicações práticas, nomeadamente na modelação de redes computacionais e de telecomunicações. E, como é evidente, o tempo que um cliente passa num daqueles sistemas, o tempo de permanência, é uma medida importante da sua “performance”. Neste trabalho inventariamos os resultados teóricos e práticos que existem actualmente sobre a sua distribuição.

**PALAVRAS-CHAVE:**

- *Redes de Jackson, Tempo de permanência, “randomisation procedure”.*

**SUMMARY:**

- Jackson queueing networks have a lot of practical applications, namely in the modelation of computation and telecommunication networks. And, evidently, the time that one customer spends in one of those systems, its sojourn time, is an important measure of its performance. In this work we collect the theoretic and practical known results that we have about its distribution.

**KEY-WORDS:**

- *Jackson networks, sojourn time, randomisation procedure.*



## 1. INTRODUÇÃO

Pretendemos, neste trabalho, apresentar alguns resultados e problemas que se põem no estudo do tempo de permanência em redes de filas de espera de Jackson. Estas redes têm muitas aplicações, nomeadamente na modelação de redes computacionais e de telecomunicações. E o tempo de permanência de um cliente, num sistema deste tipo, é evidentemente um elemento importante a ter em conta na avaliação da sua performance.

Na secção 2. Descreve-se sucintamente o modelo de rede com que vamos trabalhar, dão-se resultados gerais sobre o tempo de permanência em redes de Jackson e identificam-se alguns problemas. A exposição é suportada por três exemplos clássicos de redes de Jackson, muito usados pelos autores que estudam este assunto:

- Fila de espera  $M|M|1$  com realimentação instantânea de Bernoulli,
- Série de filas de espera simples,
- Rede de Jackson acíclica com 3 nós.

O objectivo principal da secção 3. é a apresentação da fórmula (13) que, em alguns casos, permite o cálculo exacto de momentos do tempo de permanência.

Na secção 4. É dado um método de cálculo numérico para a função de distribuição do tempo de permanência e momentos de qualquer ordem, aplicável a qualquer rede de Jackson. Este método pode aliás aplicar-se a distribuições de tempos de passagem em processos de Markov em tempo discreto, sendo o tempo de permanência em redes de Jackson um caso particular.

Finalmente, dá-se uma pequena lista de referências bibliográficas.

## 2. RESULTADOS GERAIS. EXEMPLOS

Neste trabalho vamos estudar tempos de permanência numa classe de redes de filas de espera markovianas, introduzidas inicialmente por Jackson (Jackson, 1957, 1963). Designam-se por redes de Jackson. Admitem uma única classe de clientes e são constituídas por  $J$  nós numerados  $1, 2, \dots, J$ . É costume pôr

$$U = \{1, 2, \dots, J\} \quad (1).$$

Em cada nó há:

- um único servidor,
- uma disciplina de fila "first-come-first-served"(FCFS),
- uma capacidade de espera infinita.

São redes abertas em que

- o processo de chegadas exógenas ao nó  $j$  é um processo de Poisson com taxa  $\nu_j$ ,  $j = 1, 2, \dots, J$ ,
- os processos de chegadas exógenas aos vários nós são independentes,

- os tempos de serviço no nó  $j$  são independentes e identicamente distribuídos, sendo a distribuição comum exponencial com parâmetro  $\mu_j$ ,  $j=1,2,\dots,J$ ,
- os tempos de serviço no nó  $j$  são independentes dos tempos de serviço nos outros nós,  $j=1,2,\dots,J$ ,
- após completar um serviço no nó  $j$ , um cliente é imediatamente e independentemente conduzido para o nó  $\ell$  com probabilidade  $p_{j\ell}$ , ou abandona a rede com probabilidade  $q_j$ , sendo

$$q_j = 1 - \sum_{\ell=1}^J p_{j\ell}, \quad j=1,2,\dots,J \quad (2).$$

Designa-se por  $P$  a matriz dos  $p_{j\ell}$ , isto é:  
 $P = [p_{j\ell}], \quad j=1,2,\dots,J, \quad \ell=1,2,\dots,J$ .

A taxa total de chegadas (chegadas exógenas e endógenas) ao nó  $j$  é dada por

$$\theta_j = \nu_j + \sum_{\ell=1}^J \theta_\ell p_{\ell j}, \quad j=1,2,\dots,J \quad (3).$$

As equações (3) designam-se por equações de tráfego da rede.

O estado da rede no instante  $t$  é dado por

$$N(t) = [N_1(t), \dots, N_J(t)] \quad (4)$$

sendo  $N_j(t)$  o número de clientes no nó  $j$  no instante  $t$ ,  $j=1,2,\dots,J$ .

Seja,

$$\rho_j = \frac{\theta_j}{\mu_j}, \quad j=1,2,\dots,J \quad (5).$$

Supondo que  $\rho_j < 1$ ,  $j=1,2,\dots,J$  o processo  $N = \{N(t)\}$  tem distribuição estacionária (ou de equilíbrio) dada por (ver, por exemplo, (Disney e König, 1985))

$$\prod (n_1, n_2, \dots, n_J) = \prod_{j=1}^J (1 - \rho_j) \rho_j^{n_j}, \quad n_j \geq 0, \quad j=1,2,\dots,J \quad (6).$$

(6) é também a distribuição estacionária da versão "embedded" de  $N$  imediatamente antes dos instantes de chegadas (endógenas ou exógenas), excluindo o cliente que chega.

O nosso estudo incidirá sobre tempos de permanência em redes de Jackson de clientes típicos quando o processo de estado “visto pelos clientes que chegam” está em equilíbrio. Designaremos por  $S$  o tempo de permanência na rede, isto é, o tempo que decorre entre a chegada à rede e a partida de um desses clientes. Se ele atravessar os nós  $1, 2, \dots, \ell$ , sendo  $S_j$  o seu tempo de permanência no nó  $j$ ,  $j = 1, 2, \dots, J$ , então  $S = S_1 + S_2 + \dots + S_\ell$ .

Veremos seguidamente alguns exemplos de redes de Jackson importantes para o estudo do tempo de permanência.

#### FILA DE ESPERA M|M|1 COM REALIMENTAÇÃO INSTANTÂNEA DE BERNOULLI

É uma rede com um único nó. Portanto,  $J = 1$ ,  $p_{11} = p$ ,  $q_1 = 1 - p$  e  $\theta = \frac{\nu}{1 - p}$  pondo  $\theta = \theta_1$  e  $\nu = \nu_1$ .

Se o cliente é servido  $k$  vezes o seu tempo de permanência é dado por

$$S = (t_1^0 - t_1^a) + (t_2^0 - t_2^i) + \dots + (t_{k-1}^0 - t_{k-1}^i) + (t_k^d - t_k^i) \quad (7)$$

em que

- $t_\ell^0 - t_\ell^i$  é o tempo que o cliente leva a passar pelo sistema de serviço pela  $\ell$ -ésima vez, dado pela diferença entre o instante da  $\ell$ -ésima saída do servidor e o da  $\ell$ -ésima junção à fila de espera,  $\ell = 2, 3, \dots, k - 1$ ,
- $t_1^0 - t_1^a$  é o tempo que o cliente leva a passar pelo sistema de serviço pela primeira vez, dado pela diferença entre o instante da primeira saída do servidor e o da chegada,
- $t_k^d - t_k^i$  é o tempo que o cliente leva a passar pelo sistema de serviço pela última vez, dado pela diferença entre o instante da partida da rede e o da  $k$ -ésima junção à fila de espera.

Mostra-se que (Disney e König, 1985)  $\{t_k^0 - t_k^i : k = 2, 3, \dots\}$  não é uma sucessão de variáveis aleatórias independentes. Assim, não se pode recorrer aos argumentos usuais para somar variáveis aleatórias independentes na determinação do tempo de permanência.

No entanto, consegue-se determinar uma expressão para  $P\{S \leq t\}$  que requer o cálculo das probabilidades de transição em  $k$  passos do processo markoviano de renascimento retardado  $\{N(t^i - 0), (t_k^0 - t_k^i) : k = 1, 2, \dots\}$ , condicionado ao número de vezes que o cliente volta à fila de espera. Designando a matriz dessas probabilidades de transição por  $Q_i^{(k)}(t)$  tem-se (ver ainda (Disney e König, 1985))

$$P\{S \leq t\} = \sum_{k=1}^{\infty} \prod_{i=1}^k Q_i^{(k)}(t) p (1 - p)^{k-1} V \quad (8)$$

em que  $\Pi$  é a distribuição estacionária do comprimento de fila “embedded”  
 $N^i = \{N(t_m^i): m = 1, 2, \dots, k\}$  é o número de vezes que o cliente passa pelo servidor e  
 $V$  é um vector cujas coordenadas valem todas 1.

Bem mais fácil é mostrar que (Lemoine, 1987)

$$E[S] = \frac{1}{\mu(1-p) - \nu} \quad (9)$$

### SÉRIES DE FILAS DE ESPERA SIMPLES

$$p_{j\ell} = \begin{cases} 1 & \text{se } \ell = j+1, j=1, 2, \dots, J \\ 0 & \text{nos outros casos} \end{cases} \quad \nu_1 = \nu, \nu_j = 0, j=2, \dots, J \text{ e } \theta_j = \nu, j=1, 2, \dots, J.$$

Todos os fluxos de clientes, nesta rede, em equilíbrio, são processos de Poisson.

Designando  $S_j$  e  $W_j$  os tempos de permanência e de espera, respectivamente, do cliente no nó  $j$ ,  $j = 1, 2, \dots, J$ , tem-se

$$S_j = W_j + X_j \quad (10)$$

em que  $X_j$  é o tempo do cliente no nó  $j$ . Então, em equilíbrio (Disney e König, 1985)

- os tempos de permanência nos diversos nós são variáveis aleatórias independentes,
- o tempo de permanência no nó  $j$  é uma variável aleatória exponencialmente distribuída com parâmetro  $\mu_j - \nu$ ,  $j = 1, 2, \dots, J$ ,
- os tempos de espera são variáveis aleatórias dependentes.

### REDE DE JACKSON ACÍCLICA COM 3 NÓS

É uma rede com 3 nós.  $p_{12} = p$ ,  $p_{13} = 1 - p$  e  $p_{23} = 1$ .  $p_{j\ell} = 0$  nos outros casos.  $\nu_1 = \nu$  e  $\nu_j = 0$ ,  $j = 2, 3$ .  $\theta_1 = \nu$ ,  $\theta_2 = p\nu$  e  $\theta_3 = \nu$ . Todos os fluxos de clientes, nesta rede, em equilíbrio, são processos de Poisson.

Para esta rede, em equilíbrio (Disney e König, 1985)

- o tempo de permanência no nó  $j$  é uma variável aleatória exponencialmente distribuída com parâmetro  $\mu_j - \theta_j$ ,  $j = 1, 2, 3$ ,
- $S_1$  e  $S_2$  são variáveis aleatórias independentes bem como  $S_2$  e  $S_3$ ,
- $S_1$  e  $S_3$  são variáveis aleatórias dependentes.

Há duas trajectórias alternativas para um cliente se deslocar do nó 1 para o nó 3. E um cliente que siga pelo nó 2 pode ser ultrapassado por outro que vá directamente do nó 1 para o nó 3. Deste modo, um cliente, ao chegar ao nó 3, pode encontrar lá um outro que estava atrás dele no nó 1 ou que até ainda nem tivesse chegado quando ele estava lá.

Estes clientes “ultrapassadores” (“overtaking”) podem atrasar um dado cliente, quando chega ao nó 3, por um tempo mais longo do que se não estivessem presentes. O número destes clientes depende, em parte, do número de clientes que chegam enquanto o cliente que se está a seguir está no nó 1, em parte devido à suposição de uma disciplina FCFS. Em consequência, o tempo que o cliente espera no nó 3 depende de quanto ele esperou no nó 1.

Através de simulação computacional (ver (Kiessler, Melamed, Yadin e Foley, 1988)) conseguiu-se mostrar que a correlação entre  $S_1$  e  $S_3$  para os clientes que seguem a trajectória que integra os nós 1, 2 e 3 é, em muitos casos, bastante pequena. Por outro lado verificou-se também que a função de distribuição simulada está muito próxima da função de distribuição que se obteria para  $S$  se  $S_1$ ,  $S_2$  e  $S_3$  fossem independentes. Segundo (Disney e König, 1985) parece que  $S_1$  e  $S_3$  são positivamente correlacionadas embora apenas ligeiramente.

(Disney e König, 1985) afirmam que foi determinada a distribuição exacta para  $S$ , sendo o resultado bastante complicado e requerendo um conhecimento do processo de partida de uma fila de espera  $M|M|1$  não estacionária com comprimento de fila inicial não nulo.

Vejamos os seguintes conceitos:

- diz-se que uma rede tem realimentação se um cliente pode voltar ao mesmo nó após completar o seu serviço, imediatamente ou em algum instante futuro,
- uma rede sem realimentação diz-se **acíclica**.

O primeiro exemplo apresentado nesta secção é uma rede com realimentação, como é evidente. Os outros, são redes acíclicas. Para redes de Jackson acíclicas tem-se o seguinte resultado (ver, por exemplo, (Kiessler, Melamed, Yadin e Foley, 1983)), de que já vimos algumas particularizações no segundo e no terceiro exemplos:

- suponhamos que um cliente segue a trajectória  $\gamma$  numa rede de Jackson acíclica com um único servidor em cada nó. Se o nó  $j$  pertence à trajectória  $\gamma$ , então  $S_j$  é tal que

$$P_j \{S_j \leq t \mid \text{a trajectória seguida é } \gamma\} = 1 - e^{-(\mu_j - \theta_j)t}, \quad t \geq 0 \quad (11).$$

E, se o nó  $\ell$  for o próximo nó a ser visitado depois do nó  $j$ ,  $S_j$  e  $S_\ell$  são variáveis aleatórias independentes.

O tempo de permanência tem um interesse prático evidente e, recentemente, tem sido estudado intensivamente.

Depois do que acabámos de ver é fácil aceitar que o problema do cálculo de distribuições do tempo de permanência em redes de filas de espera é um dos mais difíceis da teoria destas redes. De facto as soluções analíticas são a excepção e não a

regra e, quando existem, são bastante intratáveis. E a maior parte dos trabalhos conhecidos apenas apresenta resultados sobre distribuições do tempo de permanência para um único cliente em trajectórias sem “overtaking”, com disciplinas FCFS nos nós. Parece não haver resultados sobre distribuições conjuntas dos tempos de permanência de vários clientes.

Por outro lado, são mais numerosos os resultados negativos (não independência, por exemplo) que os positivos.

Exceptuam-se deste panorama sombrio os resultados sobre valores médios e grandezas com eles relacionadas (acessíveis através da Fórmula de Little).

Decorre dos exemplos vistos nesta secção que as dificuldades do cálculo de tempos de permanência em redes de Jackson ocorrem quando é possível darem-se os fenómenos que designámos por

- Realimentação e
- “Overtaking”.

No primeiro caso, o facto de o fluxo de clientes na trajectória de realimentação não ser um processo de Poisson, torna tudo mais complicado.

No segundo caso, existem dependências entre os tempos de permanência de um cliente nos vários nós, simultaneamente complicadas e subtis, que tornam o tempo de permanência total de cálculo difícil mesmo que os tempos de permanência em cada nó sejam fáceis de calcular.

### 3. EQUAÇÕES DE FLUXO DA REDE

O objectivo desta secção é a apresentação das chamadas “Equações de fluxo da rede” para as redes de Jackson, que introduzimos em 2., que permitem a dedução de fórmulas para o cálculo dos momentos de qualquer ordem do tempo de permanência, eficientes em certos casos.

Seguindo o trabalho de (Lemoine, 1987) seja  $\tau_j$  um instante de chegada (endógena ou exógena) ao nó  $j$  e  $\tau_j + T_j$  o instante da partida da rede do cliente que chegou em  $\tau_j$ ,  $j = 1, 2, \dots, J$ . Portanto,

- $T_j$  é o tempo de permanência restante, na rede, para a chegada ao nó  $j$  no instante  $\tau_j$ ,  $j = 1, 2, \dots, J$ .

Seja  $h_j$  a transformada de Laplace da distribuição de  $T_j$ ,  $j = 1, 2, \dots, J$ . Como  $N$  é um processo de Markov forte, e o processo de estado da rede “visto pelas chegadas” está em equilíbrio, a distribuição de  $T_j$ ,  $j = 1, 2, \dots, J$  e a sua transformada de Laplace são univocamente determinadas.

Tratando o tempo de permanência como o tempo de vida de um processo de Markov  $\mathcal{G}$  (como veremos em 4.) consegue-se mostrar que as transformadas

$h_j, j = 1, 2, \dots, J$  verificam um sistema de equações designadas por equações de fluxo da rede. Mais concretamente, prova-se que (ver (Lemoine, 1987))

- sendo  $H_j$  a distribuição de probabilidade com transformada de Laplace  $h_j$ , há uma distribuição de probabilidade com transformada de Laplace  $q_j$  tal que

$$h_j(s) + \frac{sg_j(s)}{(\mu_j - \theta_j)} = q_j + \sum_{\ell=1}^J p_{j\ell} h_\ell(s) \quad (12)$$

para  $s \geq 0$  e  $j = 1, \dots, J$ .

Em redes de Jackson sem "overtaking" as transformadas  $h_j$  e  $g_j$  são idênticas para cada  $j$ . Dadas  $h_j, j = 1, 2, \dots, J$  as transformadas  $g_j, j = 1, 2, \dots, J$  são univocamente determinadas por (12). O converso também se verifica visto que  $I - P$  (em que  $I$  é a matriz identidade) é invertível.

A partir de (12), derivando sucessivamente, (Lemoine, 1987) mostrou que

- Para  $j = 1, \dots, J$  e  $r = 1, 2, \dots$

$$E[T_j^r] = r!(\mu_j - \theta_j)^{-r} + \sum_{\ell=1}^J p_{j\ell} E[T_\ell^r] + \sum_{\ell=1}^J p_{j\ell} \sum_{m=1}^{r-1} \frac{r!}{m!(r-m)!} \mu_j^{-m} E\left[T_\ell^{r-m} \prod_{n=1}^m (N_j(T_\ell) + m)\right] \quad (13).$$

Para  $r = 1$  (13) toma a forma matricial

$$E[T_j] = (I - P)^{-1} (\mu_j - \theta_j)^{-1} \quad (14).$$

Usando (14) pode obter-se a fórmula (9) e

- Para a série de filas de espera simples

$$E[T_j] = \sum_{\ell=1}^J (\mu_\ell - \nu)^{-1}, \quad j = 1, 2, \dots, J \quad (15).$$

- Para a rede de Jackson acíclica com 3 nós

$$\begin{aligned}
 E[T_1] &= \frac{1}{\mu_1 - \nu} + p \frac{1}{\mu_2 - p\nu} + \frac{1}{\mu_3 - \nu} \\
 E[T_2] &= \frac{1}{\mu_2 - p\nu} + \frac{1}{\mu_3 - \nu} \\
 E[T_3] &= \frac{1}{\mu_3 - \nu}
 \end{aligned}
 \tag{16}$$

Note-se que para os exemplos vistos em 2.  $E[S] = E[T_i]$  e os resultados agora obtidos estão de acordo com os então referidos.

Para  $r = 2$  (13) toma a forma

$$E[T_j^2] = 2(\mu_j - \theta_j)^{-2} + \sum_{\ell=1}^j p_{j\ell} E[T_\ell^2] + 2\mu_j^{-1} \sum_{\ell=1}^j p_{j\ell} E[T_\ell (N_j(\tau_\ell^-) + 1)], j=1,2,\dots,J \tag{17}$$

A relação (17) define um sistema de  $J$  equações e  $J^2 + J$  incógnitas. Em geral, quando  $r \geq 2$ , os termos com produtos que envolvem as variáveis  $T_\ell$  e  $N_j(\tau_\ell^-)$  impedem o cálculo exacto dos momentos de ordem  $r$  dos tempos de permanência; há incógnitas a mais e equações a menos. Nesses casos são necessárias outras equações independentes para complementar (17) de modo a ser possível obter soluções exactas.

Quando qualquer par de nós da rede está ligado por no máximo, uma trajectória orientada e  $p_{jj} = 0, j=1,2,\dots,J, T_\ell$  e  $N_j(\tau_\ell^-)$  são independentes para  $j \neq \ell$ . O cálculo de  $E[T_j(N_j(\tau_j^-) + 1)]$  é irrelevante visto que  $p_{jj} = 0, j=1,2,\dots,J$ . Neste caso (13) torna-se numa fórmula recursiva compacta que permite o cálculo dos momentos de qualquer ordem dos tempos de permanência,  $T_j, j=1,2,\dots,J$ . Por exemplo, atendendo a que, então,

$$E[N_j(\tau_j^-)] = \frac{\theta_j}{\mu_j - \theta_j}, j=1,2,\dots,J \tag{18}$$

(17) toma a forma

$$E[T_j^2] = 2(\mu_j - \theta_j)^{-2} + \sum_{\ell=1}^j p_{j\ell} E[T_\ell^2] + 2(\mu_j - \theta_j)^{-1} \sum_{\ell=1}^j p_{j\ell} E[T_\ell], j=1,2,\dots,J \tag{19}$$

Aplicando (19) à serie de filas de espera simples tem-se

$$\begin{aligned} E[T_j^2] &= 2(\mu_j - \nu)^{-2} + E[T_{\ell+1}^2] + 2(\mu_j - \nu)^{-1} \cdot E[T_{\ell+1}], j=1,2,\dots,J-1 \\ E[T_j^2] &= 2(\mu_j - \nu)^{-2} \end{aligned} \quad (20).$$

(20) em conjunto com (15) permite concluir que

$$VAR[T_j] = \sum_{\ell=j}^J (\mu_{\ell} - \nu)^{-2}, j=1,2,\dots,J \quad (21),$$

como é evidente, sendo  $VAR[T_1] = VAR[S]$ .

Sejam agora redes de Jackson que não verificam aquelas condições. (Lemoine, 1987) sugere a identificação de famílias apropriadas de martingalas em  $N$  como um processo de determinar equações independentes para complementar (13). Aplicando este procedimento à fila de espera M|M|1 com realimentação instantânea de Bernoulli obteve, após raciocínios complexos e cálculos bastante laboriosos,

$$E[T_j^2] = \frac{2}{(1-p)\mu - \nu} E\left[ N(\tau^-) T = \frac{2((1-p)\mu + \nu)}{((1-p)\mu - \nu)^3} \right] \quad (22).$$

Como é evidente, neste caso, (17) toma a forma

$$E[T^2] - \frac{2p}{(1-p)\mu} E[N(\tau^-) T] = \frac{2((1-p)\mu - p\nu)}{(1-p)\mu((1-p)\mu - \nu)^2} \quad (23).$$

Usando (9), (22) e (23) (note-se que (22) e (23) são independentes) obtém-se

$$E[T^2] = \frac{2(1-p^2)\mu}{((1-p)\mu - \nu)^2 \cdot ((1-p^2)\mu - \nu p)} \quad (24),$$

$$E[N(\tau^-) T] = \frac{\nu[(1-p)\mu(2+p) - \nu p]}{((1-p)\mu - \nu)^2 \cdot ((1-p^2)\mu - \nu p)} \quad (25)$$

$$\boxed{VAR[T] = \frac{1}{((1-p)\mu - \nu)^2} \frac{(1-p^2)\mu + \nu p}{(1-p^2)\mu - \nu p}} \quad (26)$$

e

$$\boxed{COV[N(\tau^-), T] = \frac{\nu(1-p)\mu}{(1-p^2)\mu - \nu p}} \quad (27).$$

Já referimos o principal óbice ao uso de (13): em geral há poucas equações e muitas incógnitas. E os casos em que (13) se revela eficiente

- $r = 1$  para qualquer rede,
- $r$  qualquer para redes em que qualquer par de nós está ligado por, no máximo, uma trajectória orientada e  $p_j = 0, j = 1, 2, \dots, J$ ,

podem ser tratados por outros métodos. Note-se que a rede de Jackson acíclica com 3 nós não se pode enquadrar neste último caso porque os nós 1 e 3 estão ligados por 2 trajectórias orientadas. Por outro lado, o uso de martingalas como sugere (Lemoine, 1987) não parece ser muito eficiente, tendo em vista a complexidade que revestiu a sua aplicação no caso de uma rede tão simples como a fila de espera M|M|1 com realimentação instantânea de Bernoulli.

---

#### 4. CÁLCULO NUMÉRICO DE DISTRIBUIÇÕES E MOMENTOS DO TEMPO DE PERMANÊNCIA

---

Vamos descrever nesta secção um método geral, cuja chave é o procedimento designado, na literatura de língua inglesa, por “randomisation procedure”, para aproximar distribuições de tempos de primeira passagem (“first-passage times”) em processos de Markov em tempo directo, sendo os tempos de permanência em sistemas de fila de espera um caso particular.

Seja  $\mathfrak{X} = \{X(t) : t \geq 0\}$  um processo de Markov regular, em tempo contínuo, com um espaço de estados contável  $E$  e um gerador infinitesimal limitado matricial  $Q$ .

Os elementos de  $Q$  designam-se por  $Q(x, y), x, y \in E$ .  $Q(x) = \sum_{y \in E - \{x\}} Q(x, y), x \in E$  são os valores absolutos dos elementos da diagonal de  $Q$ .

$\psi_t$  designa o vector probabilidade de estado de  $X(t)$ :

$$\boxed{\psi_t(x) = P\{X(t) = x\}, x \in E} \quad (28).$$

$X$  modela a evolução de um sistema de fila de espera durante a permanência de um dado (marcado) cliente nele.

Os estados de  $E$  têm duas componentes principais:

- i) o estado do sistema de fila de espera,
- ii) a localização do cliente marcado.

Sejam,

- $A$  o subconjunto de estados que descreve o sistema até à partida do cliente marcado, e
- o subconjunto de estados que descreve o sistema depois da partida desse cliente.

Como é evidente

- $\{A, B\}$  é uma partição de  $E$ ,
- Se  $T$  é o tempo que o processo  $\aleph$  passa em  $A$  até atingir , pela primeira vez,  $T$  é precisamente o tempo de permanência do cliente marcado na rede.

Supõe-se que  $\aleph$  permanecerá em , com probabilidade 1, após o ter atingido pela primeira vez. De facto, como a evolução do sistema após a partida do cliente marcado é irrelevante, podemos supor que é um conjunto fechado. Isto é, o processo  $\aleph$  não pode voltar para  $A$  a partir de .

A quantidade de interesse é a função de distribuição de  $T$ ,  $\tau(t)$ . Note-se que

$$\tau(t) = P\{T \leq t\} = P\{X(t) \in B\} = 1 - P\{X(t) \in A\}, t \geq 0 \quad (29)$$

visto que as hipóteses apresentadas garantem que  $\{T \leq t\} = \{X(t) \in B\}$ .

Conclui-se de (29) que

- o problema do cálculo de  $\tau(t)$  é equivalente ao do cálculo da distribuição transeunte de  $X(t)$  em  $A$ .

É então necessário calcular o vector  $\psi_t, t \geq 0$ . Sendo  $P_t, t \geq 0$ , as matrizes de transição de  $\aleph$ ,

$$\psi_t = \psi_0 P_t, t \geq 0 \quad (30)$$

$P_t$  pode pôr-se na forma

$$P_t = \exp(Qt) = \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n, t \geq 0 \quad (31)$$

O "randomisation procedure" consiste em usar em (31) uma representação equivalente (ver Çinlar (1975)):

$$P_t = \exp(-\alpha t) \exp \left[ \alpha t \left( I + \frac{1}{\alpha} Q \right) \right] = \exp(-\alpha t) \sum_{n=0}^{\infty} \frac{\alpha^n t^n}{n!} R^n \quad (32)$$

em que

$$R = I + \frac{1}{\alpha} Q \quad (33)$$

é designada por “randomised matrix” na literatura de língua inglesa,

- $I$  é a matriz identidade, e
- $\alpha$  é um majorante positivo de todos os  $Q(x)$ ,  $x \in E$ .

Note-se que (ver (Melamed e Yadin, 1984 e 1984 a)).

- embora a equação (32) pareça mais complexa que (31), goza, de facto, de propriedades computacionais mais favoráveis. A mais importante é que  $R$  é uma matriz estocástica enquanto que  $Q$  não. Consequentemente, enquanto que a computação usando (32) é estável, usando (31) não é,
- o “randomization procedure” tem um significado probabilístico interessante, útil para determinar extremos para  $\tau(t)$ . De facto, sendo  $R$  uma matriz estocástica, define um processo de Markov em tempo discreto

$$\mathfrak{Y} = \{Y_n : n = 0, 1, \dots\} \quad (34)$$

contanto que tomemos  $Y_0 = X(0)$ . Com este procedimento, a relação entre os processos  $\mathfrak{N}$  e  $\mathfrak{Y}$  é bastante simples, como vamos ver.

Vamos prolongar o processo em tempo discreto a um processo de Markov em tempo contínuo tal que

- i) os intervalos de tempo entre os saltos são variáveis aleatórias exponenciais i.i.d. com média  $\frac{1}{\alpha}$ ,
- ii) os saltos são comandados por  $R$ .

Mostra-se que (Melamed e Yadin, 1984) o processo resultante é precisamente o processo original  $\mathfrak{N}$ ; todavia, sempre que se tiver uma sucessão de saltos de  $\mathfrak{Y}$  do estado  $x \in E$  para ele próprio, isso será notado em  $\mathfrak{N}$  como uma longa permanência no estado  $x$ .

Assim, o “randomisation procedure” pode ser interpretado como um polvilhamento do processo  $\mathfrak{N}$  com saltos aleatórios “falsos” entre os saltos “verdadeiros”. O processo resultante, que designaremos por  $\tilde{\mathfrak{N}}$ , em que os saltos “falsos” são visíveis, tem a mesma estrutura probabilística que  $\mathfrak{N}$  mas com uma vantagem:

— A sucessão dos instantes de salto em  $\tilde{N}$  (“falsos” e “verdadeiros”) constitui agora um processo de Poisson, o que não é, geralmente, o caso de  $N$ .

Note-se que  $Y_n$  é o estado de  $\tilde{N}$  no instante do  $n$ -ésimo salto (“falso” ou “verdadeiro”).

Suponhamos que  $\tilde{N}$  atinge o conjunto  $B$  no seu  $n$ -ésimo salto. Em consequência o tempo de permanência de  $\tilde{N}$  (e, portanto, também de  $N$ ) em  $A$  é a soma de  $n$  variáveis aleatórias independentes exponenciais com média  $\frac{1}{\alpha}$ . Isto é, o tempo de permanência tem uma distribuição de Erlang de ordem  $n$  e parâmetro  $\alpha$ . Vamos designar a sua função de distribuição por  $E_{n,\alpha}(t)$ .

Seja  $h(n)$  a probabilidade de que  $\tilde{N}$  atinja  $B$  no seu  $n$ -ésimo salto. Seja também  $\phi_n$  o vector probabilidade de estado de  $Y_n$ :

$$\phi_n = \psi_0 R^n \quad (35).$$

As quantidades  $h(n)$  são dadas pelas duas fórmulas equivalentes:

$$h(n) = \begin{cases} \sum_{x \in B} \phi_0(x), & n = 0 \\ \sum_{x \in A} \sum_{y \in B} \phi_{n-1}(x) R(x, y), & n > 0 \end{cases} \quad (36)$$

ou

$$h(n) = \begin{cases} 1 - \sum_{x \in A} \phi_0(x), & n = 0 \\ \sum_{x \in A} \phi_{n-1}(x) - \sum_{x \in A} \phi_n(x), & n > 0 \end{cases} \quad (37).$$

Dadas as probabilidades  $h(n)$  e, notando que  $\sum_{n=0}^{\infty} h(n) = 1$ , tem-se

$$\tau(t) = \sum_{n=0}^{\infty} h(n) E_{n,\alpha}(t) = 1 - \sum_{n=0}^{\infty} h(n) \bar{E}_{n,\alpha}(t), \quad t \geq 0 \quad (38)$$

$$E[T^m] = \frac{1}{\alpha^m} \sum_{n=0}^{\infty} n(n+1) \dots (n+m-1) h(n), \quad m = 1, 2, \dots \quad (39).$$

(39) para  $m = 1$  toma a forma

$$E[T] = \frac{1}{\alpha} E[H] \quad (40)$$

sendo  $H$  o número de saltos de  $\aleph$  até atingir  $B$ . (40) é a fórmula de Little neste contexto.

A equação (38) permite obter extremos simples para  $\tau(t)$  que podem, em princípio, tornar-se arbitrariamente próximos. A equação (39) permite obter minorantes de  $E[T^k]$ , em princípio tão próximos de  $E[T^k]$  quanto quisermos. Assim, dado um inteiro  $k \geq 0$  qualquer tem-se

$$L_k(t) \leq \tau(t) \leq U_k(t) \quad (41)$$

em que

$$\begin{aligned} L_k(t) &= \sum_{n=0}^k h(n) E_{n,\alpha}(t), \quad t \geq 0 \\ U_k(t) &= 1 - \sum_{n=0}^k h(n) \bar{E}_{n,\alpha}(t), \quad t \geq 0 \end{aligned} \quad (42)$$

e

$$E[T^m]_{\ell,k} \leq E[T^m], \quad m = 1, 2, \dots \quad (43)$$

em que

$$E[T^m]_{\ell,k} = \frac{1}{\alpha^m} \sum_{n=0}^k n(n+1)\dots(n+m-1)h(n), \quad m = 1, 2, \dots \quad (44)$$

Prova-se facilmente que

— Se, para algum  $\varepsilon > 0$ ,  $k$  for escolhido de acordo com a regra

$$k = \min \left\{ n \geq 0 : \sum_{i=0}^n h(i) \geq 1 - \varepsilon \right\} = k(\varepsilon) \quad (45)$$

ou, de modo equivalente,

$$J = \min \left\{ n \geq 0 : \sum_{x \in A} \phi_n(x) \leq \varepsilon \right\} = J(\varepsilon) \quad (46)$$

$$|\tau(t) - L_{J(\varepsilon)}(t)| \leq \varepsilon \text{ e } |\tau(t) - U_{J(\varepsilon)}(t)| \leq \varepsilon, \text{ uniformemente em } t \geq 0.$$

O principal óbice à aplicação do método descrito (que, em princípio, resolveria todos os problemas de cálculo relativos a distribuições dos tempos de permanência) reside na dificuldade do cálculo dos  $h(n)$ . De facto, para isso é necessário calcular os vectores  $\phi_n$  mas apenas no subconjunto  $A$  do espaço de estados. Quando o espaço de estados  $E$  é finito (como acontece no caso das redes de filas de espera fechadas) ambos  $h(n)$  e  $\phi_n$  podem, em princípio, ser calculados exactamente (à parte os erros devidos aos arredondamentos).

Na prática o espaço de estados é com frequência infinito ou, embora finito, proibitivamente grande. Nestas situações torna-se necessário truncar  $E$ . Em consequência, é introduzido um novo nível de aproximação já que os  $h(n)$ ,  $\phi_n$ , etc. têm que ser também aproximados agora.

De facto, o que se consegue obter são minorantes dos  $h(n)$  porque a truncagem de  $E$  traduz-se em perda de probabilidade (Melamed e Yadin, 1984 a). Assim, com estes valores aproximados de  $h(n)$ , (41) e (43) continuam a ser válidas mas

- Perde-se a propriedade da convergência uniforme vista atrás e
- As regras análogas a (45) e (46) não são equivalentes. A gerada por (45) pode até ser viável e, na prática usa-se, apenas, a gerada por (46) (Melamed e Yadin, 1984 a).

Usando este método (Kiessler, Melamed, Yadin e Foley, 1988) conseguiram mostrar que, numa rede de Jackson acíclica com 3 nós, a função de distribuição do tempo de permanência total para um cliente que siga a trajectória que integra os nós 1, 2 e 3 não é a que se obtém considerando que  $S_1$ ,  $S_2$ , e  $S_3$  são independentes embora esta, que designaremos por  $F(t)$ , seja uma boa aproximação daquela (como referimos em 2.). De facto, em certos casos particulares, verificaram que não era verdade que

$$F^l(t) \leq F(t) \leq F^u(t), t \geq 0 \quad (47)$$

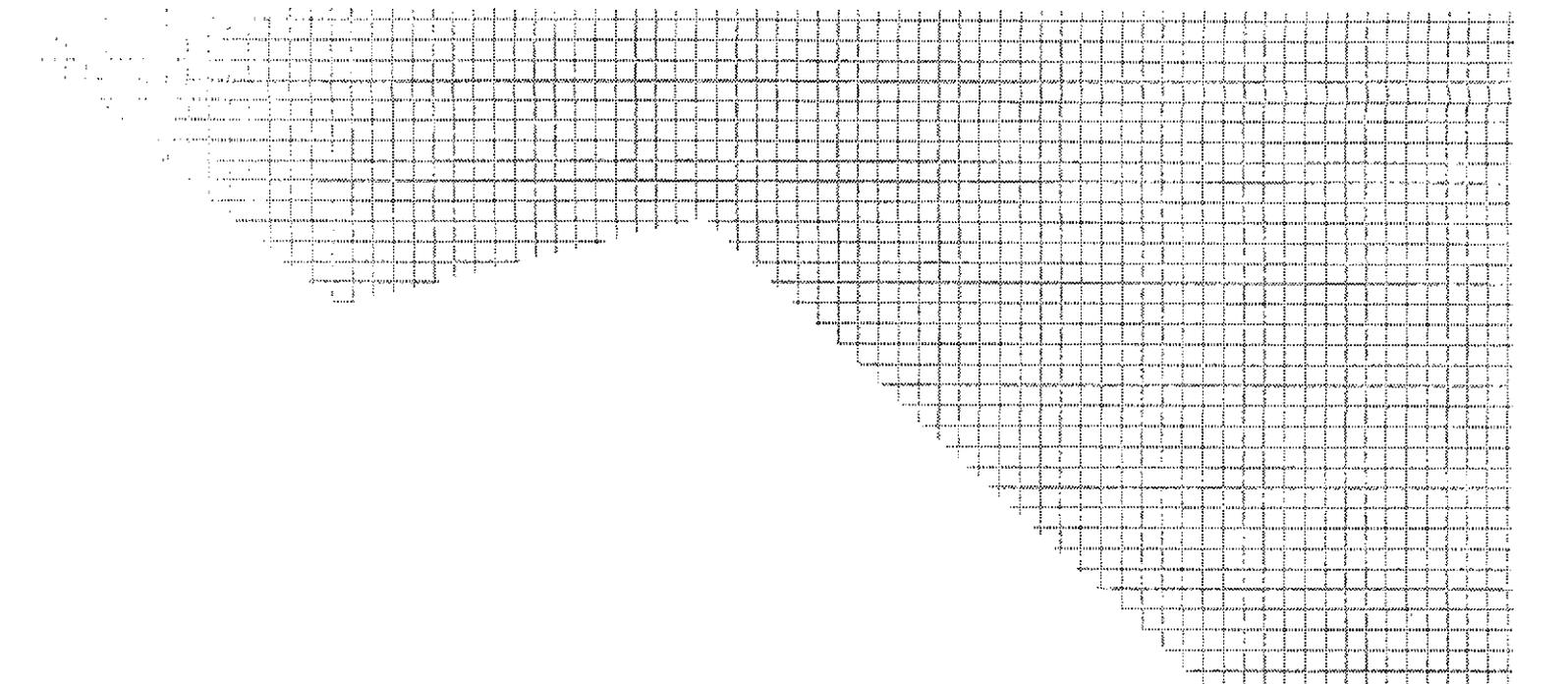
sendo  $F^l(t)$  e  $F^u(t)$  o minorante e o majorante, respectivamente, da função de distribuição do tempo de permanência daquele cliente, obtidos pelo método acabado de descrever.

Esta conclusão é importante porque, apesar da dependência de  $S_1$  e  $S_3$ ,  $F(t)$  podia ser a função de distribuição de  $S$ . De facto, (Feller, 1966) apresenta um exemplo de variáveis dependentes cuja soma tem a mesma distribuição que se as variáveis aleatórias fossem independentes.

Finalmente, note-se que a fórmula (39), aparentemente nova, parece ser mais eficiente que (13), embora apenas permita obter minorantes dos momentos, porque o seu campo de aplicação efectiva é incomparavelmente maior.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ÇINLAR, E., "**Introduction to Stochastic Processes**", Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
- DISNEY, R.L. e KOÖNIG, D., "**Queueing networks: a survey of their random processes**", Siam Review, 3 (1985), 335-403.
- FELLER, W., "**An Introduction to Probability Theory and its Applications**", Vol. II, John Wiley & Sons, New York, 1966.
- JACKSON, J. R., "**Network of waiting lines**", Operations Research, 5 (1957), 518-521.
- JACKSON, J. R., "**Jobshop-like queueing systems**", Management Science, 10 (1963), 131-142.
- KIESSLER, P. C.; MELAMED, B.; YADIN, M. e FOLEY, R. D., "**Analysis of a three node queueing network**", Queueing Systems, 3 (1988), 53-72.
- LEMOINE, A. J., "**On sojourn time in Jackson networks of queues**", J. Appl. Prob., 24 (1987), 495-510.
- MELAMED, B. e YADIN, M., "**Randomisation Procedures in the computation of cumulative-time distributions over discrete state Markov processes**", Operations Research, 4-32 (1984), 926-944.
- MELAMED, B. e YADIN, M., "**Numerical computation of sojourn-time distributions in queueing networks**", Journal of the Association for Computing Machinery, 4-31 (1984 a), 839-854.



# A ESTATÍSTICA NA GENÉTICA - UMA REVISÃO BIBLIOGRÁFICA

**Autoras:**  
**Luzia Gonçalves**  
**e**  
**Maria Antónia Amaral Turkman**



**A ESTATÍSTICA NA GENÉTICA - UMA REVISÃO BIBLIOGRÁFICA**

**STATISTICS AND GENETICS - A BIBLIOGRAPHICAL REVIEW**

Autores: Luzia Gonçalves\*  
Assistente Estagiária no Instituto Superior de Psicologia Aplicada,  
C.E.A.U.L.  
e  
Maria Antónia Amaral Turkman\*  
Professora Catedrática da Faculdade de Ciências da Universidade de Lisboa -  
Departamento de Estatística e Investigação Operacional, C.E.A.U.L.

*SUMÁRIO:*

- Nas experiências realizadas por Mendel e Morgan estavam implícitas técnicas estatísticas simples mas suficientes para enunciar alguns princípios básicos da Genética. Hardy e Weinberg, em 1908, demonstraram independentemente o Princípio de Equilíbrio, contribuindo, assim, para a implantação da Matemática na Genética, principalmente, das Populações. Com a descoberta do DNA, a partir de 1944, as experiências diversificaram-se, com vista a atingir objectivos cada vez mais auspiciosos, embora os princípios básicos da Genética se tenham mantido. Como a Estatística também evoluiu de forma acelerada com o desenvolvimento informático, principalmente a partir de meados deste século, as metodologias estatísticas usadas são mais complexas e diversas.

Sem entrar em pormenores metodológicos ou técnicos, é nosso objectivo neste trabalho, focar algumas referências bibliográficas elucidativas da presença da Estatística em algumas áreas de aplicação da Genética e na sua evolução.

*PALAVRAS-CHAVE:*

- *Equilíbrio de Hardy-Weinberg; Fragmentos de DNA; Cartografia genética; Análise forense.*

*SUMMARY:*

- Very simple statistical techniques were implicit in the experiments realized by Mendel and Morgan. They were, however, simple enough to enunciate some basic principles of Genetics. In 1908 Hardy and Weinberg enunciated, independently, a principle known as the *Hardy-Weinberg equilibrium*, which played an important role in bringing Mathematics in to Genetics. The discovery of the DNA had an extraordinary impact in the development of Genetics. Experiments were diversified and more ambitious objectives have been sought and achieved. Meanwhile important advancements have also been achieved in the field of Statistics and more complex and diverse techniques are today applied in Genetics.  
In this paper we will review some important applications of Statistics in Genetics.

*KEY-WORDS:*

- *Hardy-Weinberg equilibrium; DNA fragments; Mapping genetic; Forensic analysis.*

\* Trabalho parcialmente subsidiado pelo JNICT/PRAXIS XXI/FEDER



## 1. INTRODUÇÃO

A Estatística tem acompanhado a Genética desde os seus primórdios. Foi nos finais do século passado que Mendel estabeleceu os alicerces da Genética. A dedução dos princípios básicos, posteriormente transformados em leis, da transmissão de caracteres hereditários de geração em geração, tiveram como suporte a repetição de experiências com a ervilha-de-cheiro, bem direccionadas para atingir o objectivo inicial, e a interpretação matemática dos resultados dessas (publicados em 1866).

A análise estatística posterior dos dados de Mendel, parece revelar que os dados obtidos experimentalmente se ajustam demasiado bem às suas teorias. Fisher, como iremos ver em 2.1, contesta alguns resultados de Mendel.

O desvendar da transmissão, localização e natureza dos caracteres hereditários associa-se na curta história da Genética, principalmente a Mendel, Morgan, Avery, Waston e Crick. Morgan, entre 1909 e inícios da década de 40, conseguiu desvendar o mistério da localização, estabelecendo que os genes são responsáveis pela hereditariedade e se localizam nos cromossomas. Juntamente com os seus colaboradores, usando a frequência estatística de determinados acontecimentos deduziu algumas distâncias entre genes, mesmo não conhecendo a sua composição e natureza, inaugurando a cartografia genética que tenta mostrar a disposição dos genes ao longo dos cromossomas (ponto 2.3.1).

Em 1944, Avery associa o ácido desoxirribonucleico - DNA - à hereditariedade e desvenda a sua natureza. O gene passa a ser uma entidade física, correspondendo a um fragmento de DNA. Waston e Crick, em 1953, estabeleceram a estrutura da molécula de DNA. Esta é formada por duas cadeias enroladas em hélice, constituídas por sequências de nucleótidos (um nucleótido contém uma base química - adenina (A), guanina (G), timina (T) ou citosina(C) - açúcar e ácido fosfórico).

A partir daqui todos os esforços se encaminham para descodificar as mensagens hereditárias "escritas" com este alfabeto de quatro letras A, G, C e T. O código genético do homem, por exemplo, tem cerca de três mil milhões de letras (bases químicas). A análise do DNA assemelha-se à análise linguística. Jones (1995) refere que *"A genética, é, em si mesma, uma linguagem, um conjunto de instruções herdadas, transmitidas de geração em geração. Tem um vocabulário (os próprios genes), uma gramática (a forma como a informação herdada está disposta) e uma literatura (os milhares de instruções para compor um ser humano)"*.

A aprendizagem de uma língua faz-se por etapas tal como a descodificação das mensagens hereditárias. Por vezes conhece-se a grafia de uma palavra (código genético) mas desconhece-se o seu(s) significado(s) (função de um gene). Porém, o significado de uma palavra numa frase está relacionado com a sua localização. A gramática é essencial para a compreensão da língua tal como a cartografia é essencial para a Genética.

Com a descoberta do DNA dá-se o conseqüente desenvolvimento de técnicas para a sua análise principalmente a partir dos anos 60. A Biologia Molecular encarregou-se de observar o DNA a ponto de permitir medições em termos de pares de bases (bp) ou milhares de pares de bases (kb). Hoje sabe-se que determinados genes têm um determinado número de pares de bases ou que dois genes distam um do outro milhares de pares de bases. A cartografia genética, inaugurada por Morgan, pode ser agora mais exacta. Actualmente aspira-se cartografar o genoma (conjunto de cromossomas) de alguns organismos até ao pormenor máximo que é descrevê-lo base por base.

Os mapas mais em voga na actualidade são os chamados mapas físicos que serão resumidamente abordados em 2.3.3. Para os elaborar, o genoma ou os cromossomas de um organismo são cortados através de enzimas de restrição em sequências de bases específicas, resultando um determinado número de fragmentos de DNA cuja ordem se perdeu; é então necessário restituir essa ordem. A intervenção da Estatística nesta construção pode ser útil.

O avanço da Genética é surpreendente e são muitas as aplicações que hoje se fazem da Genética e em particular da análise do DNA em áreas como a: medicina, agricultura, farmácia, história, identificação da paternidade e também na análise forense que, nos últimos 10 anos, introduziu no esclarecimento de alguns crimes uma poderosa marca de singularidade pessoal: as "impressões digitais" genéticas. Em 2.3.2, 2.4 e 2.5 exemplificam-se algumas aplicações da Genética nessas áreas e onde a Estatística é fundamental como é o caso da análise forense.

Nestas últimas décadas intensificou-se a ligação Estatística-Genética. São inúmeros os artigos publicados em revistas relacionadas com a Genética ou Probabilidades e Estatística. Por exemplo: a *Science*, a *Genomics* ou a *American Journal of Human Genetics*, além das últimas descobertas da Genética descrevem também, em bastantes casos, as técnicas estatísticas usadas no tratamento de dados genéticos. Por outro lado as revistas: *Biometrics*, *Statistical Science*, *The Annals of Applied Probability* e *Journal of the American Statistical Association* foram "invadidas" nos últimos anos por artigos, com mais rigor teórico, sobre metodologias estatísticas úteis na Genética.

Este trabalho, sem entrar em pormenores metodológicos ou técnicos, foca algumas referências bibliográficas elucidativas da presença da Estatística em algumas áreas de aplicação da Genética e na sua evolução.

---

## 2. A ESTATÍSTICA NA EVOLUÇÃO DA GENÉTICA E NA ANÁLISE DO DNA

---

---

### 2.1. FISHER E AS EXPERIÊNCIAS DE MENDEL

---

O delineamento das experiências de Mendel permitiu-lhe uma interpretação estatística muito simples dos resultados, mas suficiente para enunciar alguns princípios básicos da hereditariedade. Com base num grande número de cruzamentos de ervilhas Mendel "probabilizou" os diferentes genótipos<sup>1</sup>, fenótipos<sup>2</sup> e a presença de genes que estão associados à transmissão da hereditariedade de geração em geração.

Por volta de 1936 Fisher, um dos grandes impulsionadores do desenvolvimento da Estatística neste século, analisou os dados originais de Mendel usando o conhecido teste de Qui-Quadrado. Sem pôr em causa as leis de Mendel, notou uma grande proximidade entre os valores observados e esperados, tendo em conta as proporções teóricas postuladas, o que conduziu a valores muito baixos para a estatística de teste. Isto levou-o a concluir que haveria alguma manipulação ou omissão de alguns dados (Fisher, 1936). Em particular, na segunda geração ao tentar inferir o genótipo dos dominantes (AA ou Aa) Fisher contesta as proporções apresentadas, uma vez que Mendel de entre 100 plantas que apresentavam o carácter dominante (A) efectuou a

---

<sup>1</sup> O genótipo representa a constituição genética de um indivíduo para um dado carácter.

<sup>2</sup> O fenótipo é a expressão do genótipo, ou seja, é a aparência do indivíduo.

auto-fertilização e observou 10 plantas da 3ª geração; se todas exibiam o carácter dominante, classificava a planta progenitora da 2ª geração como homozigótica (AA) e no caso de alguma manifestar o carácter recessivo (a) classificava-a como heterozigótica (Aa).

Por exemplo para a forma das vagens: lisa (A) ou com constrictões (a), concluiu que das 100 plantas com vagens lisas, 71 eram heterozigóticas e 29 homozigóticas (Weir, 1990). Mendel espera que 2/3 das plantas sejam heterozigóticas e 1/3 homozigóticas. Fisher discorda dessas proporções, nesta experiência particular, pois a probabilidade ( $p_{Aa}$ ) de uma planta, seleccionada aleatoriamente, na 2ª geração com vagens lisas ser classificada como heterozigótica deve ser calculada atendendo também à probabilidade da classificação como heterozigótica ser correcta,  $(1-0.75^{10})$ , pelo que:

$$p_{Aa} = \frac{2}{3}(1 - 0.75^{10}) = 0.6291 \text{ e } p_{AA} = 0.3709.$$

Contudo, apesar destas serem as verdadeiras proporções, os dados obtidos são mais condizentes com as proporções 2/3 e 1/3. Para a estatística de teste, que tem distribuição Qui-Quadrado com 1 grau de liberdade, sob a hipótese de se ter 2:1, obtém-se o valor 0.8437, enquanto que para as proporções de Fisher tem-se 2.8049, e os valores-p são respectivamente 0.3583 e 0.0940. Se em vez de 10 auto-fertilizações se tivessem considerado 30, por exemplo, as proporções de Fisher seriam praticamente as de Mendel. Weir (1990) apresenta os valores observados das estatísticas de testes para alguns dados originais de Mendel e para outros dados actuais.

## 2.2 A IMPORTÂNCIA DO PRINCÍPIO DE HARDY-WEINBERG

A teoria da evolução defendida por Darwin, em 1859 no livro "*A Origem das Espécies*", renova-se com o interesse pelas frequências genotípicas e genéticas numa determinada população. O estudo evolutivo de uma população ao longo do tempo, ou a sua caracterização genética em determinados momentos, desde cedo impuseram a presença imprescindível da Matemática na Genética, em particular na Genética das Populações. Jones (1995) refere que "*A genética das populações está <<infestada>> de matemática, muita da qual incompreensível mesmo para os geneticistas da população, mas inevitável*".

Em 1908, o matemático inglês G. H. Hardy e o físico alemão G. Weinberg, demonstraram independentemente, o conhecido **Princípio de Equilíbrio de Hardy-Weinberg** que se pode enunciar da seguinte forma: numa população em que os cruzamentos ocorram ao acaso e na ausência de mutações, selecção natural e migração, as frequências genéticas e genotípicas associadas a um determinado *locus*<sup>3</sup> permanecem constantes de geração em geração.

As frequências genotípicas determinam as genéticas. Considerando os alelos<sup>4</sup> A e a os três genótipos possíveis são AA, Aa e aa. Sob a validade deste princípio, a frequência conjunta de um par de alelos é o produto das frequências alélicas; assim, existe a relação entre as frequências genotípicas e genéticas:

<sup>3</sup> *locus* (plural *loci*) significa o posicionamento de um gene (ou outra sequência de DNA) num cromossoma.

<sup>4</sup> Alelos são diferentes formas do mesmo gene.

$$\begin{cases} p_{AA} = p_A^2 \\ p_{Aa} = 2p_A p_a \\ p_{aa} = p_a^2 \end{cases}$$

Se numa população, num dado momento, se verificar um excesso de homozigóticos ou heterozigóticos ocorre um desequilíbrio. Para averiguar o equilíbrio/desequilíbrio de populações foram desenvolvidos bastantes estudos estatísticos. Tamarin (1982) refere um estudo geométrico, efectuado em 1926 pelo estatístico Bruno De Finetti, que permite verificar se uma ou mais populações estão ou não em equilíbrio, tendo em conta os génotipos determinados por dois alelos (A, a) através de um triângulo - Triângulo de De Finetti.

Weir (1990) descreve alguns testes estatísticos para averiguar a validade do equilíbrio de Hardy-Weinberg, por vezes também designados por testes de independência. Ultimamente, com o uso dos fragmentos de DNA (medidos em pares de bases (pb) ou kilobases (kb), 1 kb = 1000 bp), na análise forense e na identificação da paternidade surgiram novos testes e adaptações dos antigos (Guo & Thompson, 1992; Devlin *et al*, 1990; Geisser & Johnson, 1991; Devlin & Risch, 1993; Weir, 1993).

A necessidade de estimar as frequências genotípicas e genéticas que, obviamente, dependem da verificação ou não deste princípio, contribuiu para a implantação da amostragem e para o desenvolvimento de métodos estatísticos de estimação na Genética. Principalmente quando os génotipos são resultantes da combinação de  $m$  alelos múltiplos ( $m \geq 2$ ), a estimação das frequências dos  $m(m+1)/2$  génotipos possíveis, envolve fórmulas bastante extensas pelo que se optou por não as reproduzir. Weir (1990) e Elandt-Johnson (1970), entre outros, exploram com pormenor este assunto.

---

### 2.3. A CARTOGRAFIA GENÉTICA

---

#### 2.3.1. O CONTRIBUTO DAS INVESTIGAÇÕES DE MORGAN

---

Os primeiros registos da organização dos genes nos cromossomas devem-se a Morgan e seus colaboradores que, entre 1909 até por volta de 1940, estudaram activamente a hereditariedade de caracteres da mosca-da-fruta (*Drosophila melanogaster*).

As exaustivas experiências de cruzamentos de milhares e milhares de moscas-da-fruta, revelaram que todos os seus genes podiam ser dispostos em grupos, consoante fossem ou não transmitidos segundo as leis de Mendel. Pares de caracteres que não obedeciam ao esquema mendeliano passavam de geração em geração sempre ligados. Segundo Morgan, os genes que os controlavam estavam no mesmo cromossoma e são designados por **genes ligados**. O tipo de herança, correspondente, designa-se por **ligação factorial** ou *linkage*.

Nalguns caracteres investigados a ligação revelou-se apenas parcial, nalgumas experiências com as moscas-da-fruta apareceram proporções fenotípicas em determinadas gerações menos habituais. Morgan, atribuiu essas percentagens irregulares a alguns acontecimentos que poderão ocorrer na formação dos gâmetas (espermatozóide e óvulo). Os cromossomas homólogos emparelham e antes de se separarem podem "cruzar-se" e trocar segmentos idênticos. Estes processos ficaram

conhecidos por **recombinação**. Em particular, o fenómeno de emparelhamento seguido por uma troca de material passou a ser designado por *crossing-over*.

Quanto mais afastados estejam dois genes num dado cromossoma maior será a probabilidade de ocorrer *crossing-over* entre eles. A frequência de *crossing-over* é expressa em termos de **percentagem de recombinação ou distância genética entre dois loci**. A unidade, arbitrária, da distância genética é chamada: **unidade de recombinação, unidade de mapa, ou centimorgan (cM)** e corresponde, por convenção, a 1% de recombinação.

Com base nas excepções às leis de Mendel iniciou-se a elaboração dos primeiros **mapas genéticos** ou **mapas de linkage** que tentam mostrar a disposição e a distância dos genes para cada cromossoma em vários seres vivos. Como dependem de inúmeros cruzamentos, nem sempre possíveis nalgumas espécies, essa cartografia genética avançou devagar, principalmente no Homem, o alvo das maiores atenções.

Este tipo de mapas, apesar de actualmente os geneticistas os continuarem a elaborar, foram ultrapassados pelos **mapas físicos** que surgiram com a descoberta do DNA e que não dependem de uma observação indirecta da expressão dos genes mas de uma manipulação directa do material genético, como veremos em 2.3.3.

### 2.3.2. LOCALIZAÇÃO DE GENES COM INTERESSE NA MEDICINA

A localização de genes, principalmente os responsáveis por doenças hereditárias, mereceu desde sempre a maior atenção. Os primeiros passos deram-se com as experiências de Morgan, mas no homem os progressos foram lentos. A recombinação de genes, baseada na observação de características fenotípicas fáceis de identificar e distinguir, que apareciam em determinadas gerações resultantes de inúmeros cruzamentos, não tinha aplicabilidade directa no homem. No homem a estratégia é oposta, recua-se aos ascendentes, usando as árvores genealógicas de famílias com casos de doenças ou anomalias suspeitas de serem hereditárias. Actualmente as árvores genealógicas continuam a usar-se e a recombinação também, mas com um complemento poderoso: a análise de DNA.

A recombinação estende-se agora, por exemplo, a fragmentos de DNA não-codificante que não produzem proteínas e que não se manifestam na aparência de um indivíduo. Esse tipo de fragmentos são marcadores genéticos determinantes para a análise forense e identificação da paternidade.

Embora de uma forma abstracta, as sequências de DNA correspondentes a alguns genes foram localizadas antes da grande expansão da Biologia Molecular, porém, o DNA não-codificante só se descobriu quando foi possível estudar a hereditariedade a nível molecular.

A Estatística esteve presente neste âmbito, logo no começo, com a estimação da taxa de recombinação, usada para estabelecer distâncias entre genes, expressas em percentagem ou cM. Mas, ao que parece 1 cM corresponde aproximadamente a 1.000.000 pares de bases de DNA (Weir, 1990; Navidi e Arnheim, 1994), pelo que continua útil conhecê-la, pelo menos para uma primeira abordagem, na elaboração dos mapas actuais (Bishop *et al.*, 1983).

Na medicina, a análise de *linkage*, é explorada para determinar a posição num cromossoma de um gene responsável por uma doença em relação a um ou mais genes de localização conhecida (marcadores). Consideremos o *locus* de uma doença com os alelos **A** e **a** e um segundo *locus* (marcador) com os alelos **B** e **b**. Se pais com o genótipo **ab/AB** (esta representação significa que num cromossoma situam-se os alelos

a e b e no seu homólogo A e B) deram origem a um descendente com o genótipo aB/Ab ocorreu uma recombinação entre os *loci* da doença e do marcador.

Seja,  $\theta$ , a taxa de recombinação entre dois *loci* referentes ao gene de uma doença e a um marcador. Se estes estão em cromossomas diferentes têm uma transmissão independente e  $\theta=1/2$ ; se estão no mesmo cromossoma, são mais frequentemente transmitidos em bloco e  $0 \leq \theta < 1/2$ . Caso o gene da doença e o marcador estejam próximos,  $\theta$  deve aproximar-se de 0; se estão afastados deve estar próximo de 1/2. Diversos testes estatísticos têm sido desenvolvidos para testar:

$$H_0 : \theta = \frac{1}{2} \text{ versus } H_1 : 0 \leq \theta < \frac{1}{2}$$

tendo em conta dados de famílias. Lemdani e Pons (1995) e Shoukri e Lathrop (1993) deduziram vários testes estatísticos, segundo uma abordagem clássica, para dados heterogêncos, ou seja, a amostra das famílias (com o mesmo número de descendentes ou não) apresenta uma proporção,  $\lambda$ , de famílias onde os *loci* da doença e do marcador estão ligados e nas restantes  $(1-\lambda)$  não estão ligados. Esta situação é apropriada para dados de uma mesma doença mas causada por mutações em diferentes *loci*. Risch (1988) segue uma abordagem bayesiana, exemplificando a aplicação dos testes apresentados com dados de famílias, para averiguar a ligação de três *loci* de DNA obtidos por RFLP e o *locus* do síndrome do X-frágil (caracteriza-se por deficiências mentais congénitas).

Thompson (1994) estuda a recombinação entre mais de dois *loci* e, dada a complexidade dos modelos propostos, recorre aos Métodos de Monte Carlo via Cadeias de Markov, em particular a amostragem de Gibbs e o algoritmo de Metropolis-Hasting. Estes métodos de simulação têm sido usados para a análise de árvores genealógicas associadas a famílias com casos de doenças hereditárias, como por exemplo o síndrome de Werner's (Thompson, 1994) e a fibrose cística (Geyer & Thompson, 1993).

Na epidemiologia os métodos tradicionais mais frequentes assumem que os indivíduos não estão relacionados. Porém esta hipótese não é realística ao estudar famílias, pois os indivíduos dentro delas estão relacionados. Assim, ao incluir a componente genética de algumas doenças (como os tumores malignos), até há pouco tempo consideradas resultantes de causas ambientais, houve necessidade de alterar os métodos habituais nesses estudos.

Thomas e Gauderman (1996) usam a amostragem de Gibbs para estimar parâmetros de modelos probabilísticos complexos que incluem uma componente genética (relação entre fenótipos e genótipos e diversos genes dentro de cada família) além do conjunto de variáveis usuais em estudos epidemiológicos (idade, sexo do indivíduo, alimentação, etc.), e onde a matriz de dados corresponde a árvores genealógicas. Os referidos autores dão diversas referências bibliográficas de trabalhos que incluem estes métodos de simulação recentemente (re)descobertos na Estatística e que são a solução para ultrapassar os pesados cálculos associados a modelos complexos que surgem com frequência também na Genética.

---

### 2.3.3. MAPAS FÍSICOS

Os fragmentos de DNA podem ser medidos, com intervenção estatística, em termos de pares de bases (bp) ou milhares de pares de bases (kb). Esses fragmentos podem ser obtidos com **enzimas de restrição** que cortam o DNA de um organismo em sequências de bases específicas, vulgarmente com 4 a 8 pares de bases. Estas enzimas

cortam o genoma ou uma sub-região deste (cromossoma ou gene) num menor ou maior número de segmentos consoante a menor ou maior frequência dessas sequências.

Os comprimentos dos fragmentos de DNA obtidos por acção das enzimas de restrição são inferidos depois de serem submetidos a electroforese em gel (que permite a sua separação segundo o seu tamanho). Este gel pode ser fotografado com iluminação ultra violeta, resultando uma imagem típica. Para determinar, de forma aproximada, os comprimentos dos fragmentos evidenciados na fotografia são aplicados no mesmo gel fragmentos de DNA com massa molecular conhecida, isto é, **marcadores de massa molecular** ou **fragmentos standard**. A estes correspondem as bandas laterais na fotografia que são comparadas com as restantes que contêm os fragmentos cujo tamanho se pretende identificar. Uma estimativa do comprimento de um fragmento pode obter-se comparando a sua posição com a dos marcadores mais próximos. Por vezes para os propósitos práticos este método é satisfatório. Mas em geral, traça-se uma recta a partir do logaritmo dos comprimentos dos fragmentos *standard* ( $l_i$ ) em função da distância ( $d_i$ ) percorrida por cada um deles no gel, num determinado período de tempo:

$$d_i = a + b \cdot \ln(l_i)$$

onde  $a$  é a ordenada na origem e  $b$  o declive da recta que é negativo porque os fragmentos que migram mais, são os que têm menores dimensões. Conhecida a distância migrada, no gel, de um fragmento qualquer de DNA o seu comprimento é estimado, com base nessa recta.

Outros métodos, embora nenhum seja totalmente satisfatório, têm sido sugeridos (Weir, 1990, Nelson e Speed, 1994).

Uma das etapas fundamentais para a obtenção de uma cartografia dos cromossomas (e subsequente localização de genes e outras sequências de DNA) é a construção de um **mapa físico**. Este é uma representação ordenada de fragmentos de DNA, resultantes da acção de enzimas de restrição.

Além destes fragmentos de restrição podem-se procurar outras sequências de DNA recorrendo a **sondas**. Estas são sequências de DNA marcadas, por exemplo, radioactivamente e com as quais é possível detectar sequências iguais ou semelhantes num conjunto de fragmentos de restrição. Por este processo, designado por **hibridação**, é possível localizar fragmentos uns em relação aos outros e/ou confirmar a sua posição num cromossoma.

Nos últimos anos, assistiu-se a um desenvolvimento de técnicas estatísticas auxiliares na construção de mapas físicos. Nelson e Speed (1994) propõem uma abordagem estatística bayesiana auxiliar na identificação de sobreposições entre dois fragmentos de DNA. Inicialmente partem de condições de aplicação prática menos usuais (por exemplo, que os fragmentos em comparação têm a mesma dimensão), mas que são convenientes para deduzir um modelo matemático simples. Este modelo dá origem a outros que contemplam condições práticas mais realistas. Gonçalves e Zé-Zé (1997) aplicam tal abordagem na construção do mapa físico da bactéria *Oenococcus oeni*, utilizada comercialmente na indústria vinícola.

O mapa físico ideal é aquele que descreve o genoma base por base. Embora para os maiores genomas se exija um grande esforço e talvez não seja de importância primordial, recomenda-se esse nível de resolução para determinados genes, áreas cromossomais ou genómicas de maior interesse. A sequenciação completa de muitos genomas pequenos, como sejam alguns vírus, e alguns maiores como sejam o da bactéria *Escherichia coli* e da levedura *Saccharomyces cerevisiae* já foi efectuada. O "Projecto Genoma Humano", em curso desde há alguns anos com uma duração prevista

de 15 anos, visa a sequenciação completa dos três mil milhões de bases de DNA do Homem.

Também se pode construir o mapa a partir de fragmentos de DNA "armazenados" em clones. Uma colecção de clones representativos do DNA de um organismo constitui uma **biblioteca genómica**. Os clones deverão conter fragmentos diferentes do genoma, com sobreposições parciais, e o seu número deve ser tal que exista uma elevada probabilidade de qualquer porção do genoma estar representada. A primeira questão que se levanta é quantos clones se devem seleccionar de forma a cobrir quase na totalidade o genoma do organismo. A escolha desse número, e de outros aspectos, baseia-se em modelos probabilísticos (Fu *et al.*, 1992, Port *et al.*, 1995, Gonçalves, 1997).

No início esperava-se que os mapas de *linkage* e os mapas físicos fossem semelhantes mas com o decorrer das investigações verificou-se que eram muito diferentes. Quer se trate de um mapa de *linkage* ou de um mapa físico é o tempo dedicado à investigação e o desenvolvimento das técnicas usadas que determinam o seu aperfeiçoamento.

---

#### 2.4. ANÁLISE FORENSE

---

Na análise forense são procurados marcadores genéticos de DNA em células recolhidas no local do crime e numa amostra de células retiradas do suspeito de ter cometido o crime. Os laboratórios de Biologia Forense usam geralmente 3 a 5 *loci* VNTR - fragmentos de DNA não-codificante que correspondem a repetições sucessivas da mesma base ou de sequências de bases - que diferem de indivíduo para indivíduo, excepto para os gémeos verdadeiros. Encontra-se frequentemente referência aos *loci*: D1S7, D2S44, D4S139, D10S28 e D17S79.

Ultimamente surgiram diversas abordagens estatísticas e desenvolvimento de modelos probabilísticos próprios na Genética com finalidades jurídicas. Aitken (1996) expõe alguns dos métodos estatísticos e probabilísticos desenvolvidos nos últimos anos.

Do tratamento laboratorial conveniente do material genético resultam as "impressões digitais" genéticas, que aparecem sob a forma de autorradiografia. O comprimento (geralmente em kb) dos fragmentos de DNA não-codificante são medidos com um determinado erro que varia de laboratório para laboratório.

Se os cientistas forenses declaram que há diferenças entre as "impressões digitais" genéticas do suspeito e da amostra recolhida no local do crime, através das autorradiografias correspondentes, não é necessária a intervenção da Estatística. Caso contrário, se não são "detectadas diferenças" recorre-se à Estatística para se calcular a proporção de perfis "semelhantes" à amostra do local do crime, numa população de referência. Estas populações de referência devem ser suficientemente bem estudadas quanto a esses *loci* e a um conjunto de variáveis relacionadas com a composição genética.

Até aqui a Estatística e os modelos probabilísticos podem intervir no conhecimento dos erros de medição, no conhecimento da genética das populações, mas os tribunais que aceitam as "impressões digitais" genéticas esperam um "resumo numérico" que contribua para a sua decisão jurídica.

Considerem-se, por exemplo, os *loci*  $L_1, L_2, L_3, L_4$  e  $L_5$  cada um com vários alelos na população. Cada indivíduo possui apenas dois alelos ( $A_{i1}$  e  $A_{i2}$ ,  $i=1, \dots, 5$ ) situados nos cromossomas homólogos, sendo um herdado da mãe e outro do pai.

Suponhamos que, tanto na amostra de DNA do suspeito, como na do local do crime, se tinham observado os seguintes alelos em cada *loci*:

<i>Locus</i>	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$
Alelos	$A_{11}, A_{11}$	$A_{21}, A_{22}$	$A_{32}, A_{32}$	$A_{41}, A_{42}$	$A_{52}, A_{52}$

Como se constatou a igualdade de perfis, a Estatística é necessária para estimar a probabilidade de uma pessoa, seleccionada aleatoriamente, da população de referência ter a mesma composição genética do suspeito e da amostra do crime. Admitindo o equilíbrio de Hardy-Weinberg entre e dentro dos *loci*, podemos sugerir:

$$p_{11}^2 \cdot (2p_{21} \cdot p_{22}) \cdot p_{32}^2 \cdot (2p_{41} \cdot p_{42}) \cdot p_{52}^2$$

onde  $p_{ij}$  é a proporção do alelo  $A_{ij}$  na população de referência para o *loci*  $L_i$ ,  $i=1, 2, \dots, 5$  e  $j=1, 2$ .

Nos primeiros tempos de utilização das "impressões digitais" genéticas, esta estimativa da probabilidade era apresentada ao juiz, mas frequentemente, com duas interpretações consoante as conveniências da defesa ou da acusação (Aitken, 1996). Como os *loci* VNTR apresentam uma diversidade de alelos muito grande numa população, superior aos grupos sanguíneos que são usados há mais tempo em questões relacionadas com a investigação da paternidade e na análise forense, a probabilidade de um outro indivíduo ter as mesmas "impressões digitais" genéticas (excepto nos gémeos verdadeiros) é muito pequena.

O cálculo desta probabilidade tendo em conta a independência entre e dentro dos *loci* (embora se possa testar como já foi referido atrás), a sua interpretação e a escolha da população de referência, entre outros problemas, instalaram muitas dúvidas, suscitando uma intervenção crescente dos estatísticos nesta área (Aitken, 1996; Roeder, 1994).

Como a população que está em causa nestas áreas é a humana, pelo menos uma das condições para que se verifique o equilíbrio de Hardy-Weinberg falha: os cruzamentos não são aleatórios. Mas em populações numerosas, a falha desta e das outras condições (selecção natural, mutações, ...) pode não afectar muito o equilíbrio. Esses cruzamentos não aleatórios podem determinar subpopulações dentro de uma população geral que não devem ser ignoradas nos tratamentos estatísticos.

Na aplicação pioneira dirigida por Jeffreys, num caso de violação de duas jovens em Inglaterra, a população de referência era constituída por cerca de 200 jovens voluntários do sexo masculino (Bader, 1987), o próprio FBI norte-americano no início tinha uma população de referência constituída por funcionários policiais (Jones, 1995). Actualmente há uma grande preocupação com a escolha, dimensão e conhecimento da estrutura genética destas populações e certamente as amostras por conveniência estão a diminuir. Por exemplo, existem bases de dados com indivíduos caucasianos, hispânicos, africanos, japoneses, etc. (Herrin, 1993; Berry, 1991; Roeder, 1994). A existência de subpopulações tem merecido uma atenção especial (Roeder, 1994), havendo um controlo, por exemplo, de eventuais excessos de indivíduos homocigóticos ou heterocigóticos, consanguinidade, etc.

Para um determinado *locus* pode surgir outra complicação. Pode ser difícil distinguir na autorradiografia uma situação em que o indivíduo é homocigótico, doutra

- **coalescência** - em que os dois alelos têm comprimentos similares ou iguais (Devlin *et al.*, 1992).

O interesse estatístico por estas áreas fez com que se desenvolvessem modelos probabilísticos que têm em conta os erros de medição no peso molecular dos fragmentos de DNA, a estrutura da população, coalescência, testemunhas oculares, etc (Berry, 1991; Devlin *et al.*, 1992; Roeder, 1994; Aitken, 1996).

Actualmente em vez do cálculo da referida probabilidade, estão a tornar-se mais frequentes os testes de hipóteses, principalmente, segundo uma abordagem bayesiana. Em face dos dados das "impressões genéticas" do suspeito e da amostra de DNA recolhida no local do crime sujeitas às limitações anteriormente referidas (erros na medição, coalescência, ...) e a testemunhas oculares, por exemplo, a Estatística indica qual das hipóteses é mais evidente ou plausível: as amostras de DNA do local do crime e do suspeito provêm do mesmo indivíduo contra a hipótese de serem provenientes de indivíduos diferentes.

Esta metodologia também se usa na investigação da paternidade e mais raramente (mas também acontece) da maternidade. Neste caso comparam-se as "impressões digitais" genéticas da criança, mãe e pretense(s) pai(s). Com base nos dados dessa comparação avalia-se qual das hipóteses é mais evidente: o pretense pai é o verdadeiro pai da criança contra a negação desta (Aitken, 1996; Curnow & Wheeler, 1993).

Os problemas de natureza jurídica, genética e estatística que se têm levantado ainda não permitem confiar totalmente nas "impressões digitais" genéticas, mas com o empenho actual a credibilidade deste método aumentará num futuro próximo.

## 2.5. OUTROS ESTUDOS QUE ENVOLVEM SEQUÊNCIAS DE DNA

Curiosamente os modelos de regressão logística tão usados na epidemiologia são também usados na Genética. Amfoh *et al.* (1994) aplicam a regressão logística para analisar a frequência de triplos de bases de DNA, principalmente os que correspondem a aminoácidos, no genoma mitocondrial humano, tendo em conta algumas variáveis explicativas como, por exemplo, o número de mutações de uma base para que o tripleto se transforme num sinal de paragem (isto é, sequências de DNA de três bases que sinalizam o fim da síntese proteica).

Quando se analisam curtas sequências de DNA, diversos estudos mostram que não há independência estatística. Ou seja, a probabilidade de encontrar uma sequência ATG:

$$P(ATG) \neq P(A).P(T).P(G)$$

pelo que se deve ter em conta que:

$$P(ATG) = P(A).P(T/A).P(G/AT).$$

As associações entre bases adjacentes podem induzir associações entre bases mais distantes pelo que se deve saber até onde se estendem essas associações. As Cadeias de Markov (principalmente de 1ª ordem e 2ª ordem) têm sido amplamente usadas em estudos que envolvem, por exemplo, a estimação da frequência de aparecimento de curtas sequências de bases de DNA numa determinada região genómica (Weir, 1990; Prum *et al.*, 1995; Amfoh *et al.*, 1994).

Na evolução de espécies ao longo do tempo os modelos Markovianos (caso contínuo) têm sido usados para descrever as alterações ocorridas no DNA que podem dar origem a espécies diferentes (Kelly, 1994). Quanto mais próximas estiverem duas espécies mais semelhante é o seu DNA, pelo que na evolução usam-se algumas medidas descritas, por exemplo, em Weir (1990) para avaliar essa proximidade ou distanciamento a partir de um ancestral comum.

A pesquisa de sequências de DNA e a semelhança entre duas, é também explorada por estatísticos e informáticos quando se recorre aos bancos informáticos onde se registam as descobertas genómicas (Waterman & Vingron, 1994). Não é só no processo de produção de informação que a Estatística, a Investigação Operacional e a Informática intervêm, mas também, no acesso - que se quer rápido - à enorme quantidade de sequências genómicas, já conhecidas.

### 3. CONSIDERAÇÕES FINAIS

Embora a Estatística tenha acompanhado a Genética desde os seus primórdios, a ligação Estatística – Genética intensificou-se nas últimas décadas. O grande desenvolvimento informático verificado a par do desenvolvimento de novas técnicas estatísticas, não é estranho a este facto. Manter a Estatística a passo com a Genética nem sempre tem sido fácil. A variedade de investigações, associadas principalmente à análise de DNA, e conseqüentemente a quantidade de informação produzida é tal que se torna difícil acompanhá-las. Como tal, a Estatística presente na Genética actual não pode dar modelos definitivos, pelo que se exige um acompanhamento atento e permanente desta rápida evolução da Genética. Essa “desactualização” não implica uma ruptura total com as metodologias passadas, estas são em muitos casos apenas reformuladas.

As aplicações da Genética na Medicina, na análise forense ou identificação da paternidade, por exemplo, têm contribuído para o desenvolvimento de técnicas estatísticas, talvez, impensáveis ainda há dez anos atrás.

Cada vez mais a Genética precisa da Estatística, mas o contrário também é verdade. A Estatística tem muito a beneficiar com esta ligação e o futuro promete ser auspicioso.

### AGRADECIMENTOS

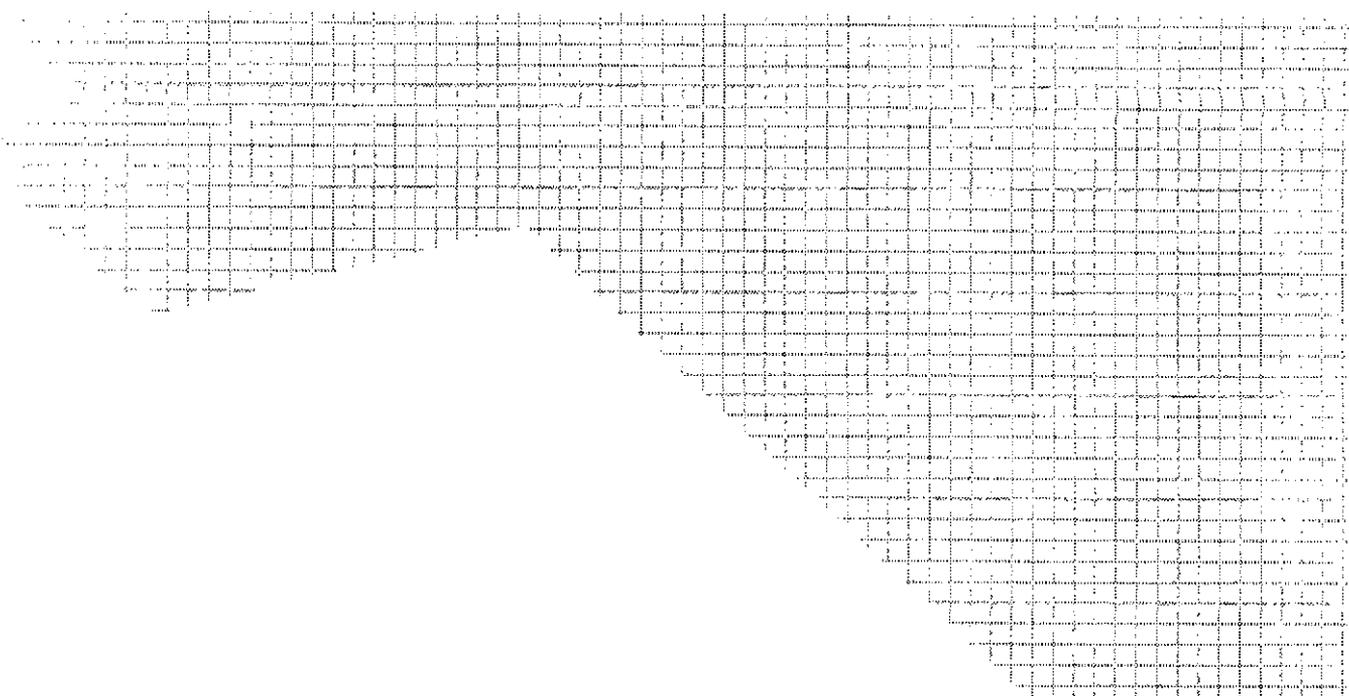
Agradecemos à Dr<sup>a</sup> Líbia Zé-Zé, à Prof<sup>a</sup> Doutora Helena Paveia e ao Prof<sup>o</sup> Doutor Rogério Tenreiro, do Departamento de Biologia Vegetal, FCUL, todo o apoio prestado.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AITKEN, C. (1996), *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons.
- AMOFOH, K. K., SHAW, R. e BONNEY, G. E. (1994), "The use of logistic models for the analysis of codon frequencies of DNA sequences in terms of explanatory variables", *Biometrics*, 50, 1054-1063.
- BADER, J.-M. (1987), "Sherlock Holmes découvre la Génétique dans une histoire de viols", *Science & Vie*, 834, 12-17, 156.
- BERRY, D. A. (1991), "Inferences using DNA profiling in forensic identification and paternity cases", *Statistical Science*, 6, 175-205.
- BISHOP, T. D., CANNINGS, C., SKOLNICK e WILLIAMSON, J. A. (1983), "The number of polymorphic DNA clones required to map the human genome", In *Statistical Analysis of DNA Sequence Data*, (ed. B. S. Weir), 181-200. Marcel Dekker. Nova Iorque.
- CURNOW, R.N. e WHEELER, S. (1993), "Probabilities of incorrect decisions in paternity case using multilocus deoxyribonucleic acid probes", *Journal of the Royal Statistical Society, Series A*, 156, 207-223.
- DEVLIN, B., RISCH, N. (1993), "Physical properties of VNTR data, and their impact on a test of allelic independence", *American Journal of Human Genetics*, 53, 324-329.
- DEVLIN, B., RISCH, N. e ROEDER, K. (1990), "No excess of homozygosity at loci used for DNA fingerprinting", *Science*, 249, 1416-1420.
- DEVLIN, B., RISCH, N. e ROEDER, K. (1992), "Forensic inference from DNA fingerprints", *Journal of the American Statistical Association*, Vol. 87, Nº 418, 337-350.
- ELANDT-JOHNSON, R. (1970), *Probability Models and Statistical Methods in Genetics*, John Wiley & Sons, Inc. Nova Iorque.
- FISHER, R. A. (1936), "Has Mendel's work been rediscovered?" *Annals of Science*, 1, 115-137.
- FU, Y.-X., TIMBERLAKE, W. E. e ARNOLD, J. (1992), "On the design of genome mapping experiments using short synthetic oligonucleotides", *Biometrics*, 48, 337-359.
- GEISSER, S. e JOHNSON, W. (1991), "Testing Hardy-Weinberg equilibrium on allelic data from VNTR loci", *Technical Report*, 569, 1-9.
- GEYER, C.J. e THOMPSON, E.A. (1993), "Annealing Markov Chain Monte Carlo with applications to pedigree analysis", *Technical Report*, 589, School of Statistics University of Minnesota.
- GONÇALVES, L. (1997), *A Estatística na Genética e em Particular na Análise do DNA*. (Dissertação submetida a provas para obter o grau de mestre em Probabilidades e Estatística), Faculdade de Ciências da Universidade de Lisboa.
- GONÇALVES, L., ZÉ-ZÉ, L. (1997), "A Contribuição da Metodologia Bayesiana na Construção de Mapas Físicos", *Nota 9/97*, Centro de Estatística e Aplicações da Universidade de Lisboa.

- GUO, S. W. e THOMPSON, E.A. (1992), "Performing the exact test of Hardy-Weinberg proportion for multiple alleles", *Biometrics*, 48, 361-372.
- HERRIN Jr., G. (1993), "Probability of matching RFLP patterns from unrelated individuals", *American Journal of Human Genetics*, 52, 491-497.
- JONES, S. (1995), *A Linguagem dos Genes*, Difusão Cultural (Tradução de Isabel Mafra)
- KELLY, C. (1994), "A test of Markovian model of DNA evolution", *Biometrics*, 50, 653-664.
- LEMDANI, M. e PONS, O. (1995), "Tests for genetic linkage and homogeneity", *Biometrics*, 51, 1033-1041.
- NAVIDI, W. e ARNHEIM, N. (1994), "Analysis of genetic data from the polymerase chain reaction", *Statistical Science*, Vol. 9, 3, 320-333.
- NELSON, D. O. e SPEED, T. P. (1994), "Statistical issues in constructing high resolution physical maps", *Statistical Science*, Vol. 9, Nº 3, 334-354.
- PORT, E., SUN, F., MARTIN, D. e WATERMAN, M. S. (1995), "Genomics mapping by end-characterized random clones: a mathematical analysis", *Genomics*, 26, 84-100.
- PRUM, B., RODOLPHE, F. e TURCKHEIM, E. (1995), "Finding words with unexpected frequencies in deoxyribonucleic acid sequences", *Journal of the Royal Statistical Society, Series B*, Vol. 57, Nº1, 205-220.
- RISCH, N. (1988), "A New statistical test for linkage heterogeneity", *American Journal of Human Genetics*, 42, 353-364.
- ROEDER, K. (1994), "DNA Fingerprinting: a review of the controversy", *Statistical Science*, 9, 222-278.
- SHOUKRI, M.M. e LATHROP, G.M. (1993), "Statistical testing of genetic linkage under heterogeneity", *Biometrics*, 49, 151-161.
- TAMARIN, R. H. (1982), *Principles of Genetics*, Willard Grand Press. Massachussts.
- THOMAS, D.C. e GAUDERMAN (1996), "Gibbs sampling methods in genetics", In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson e D. J. Spiegelhalter), 420-440. London: Chapman & Hall.
- THOMPSON, E. A. (1994), "Monte Carlo likelihood in genetic mapping", *Statistical Science*, Vol. 9, Nº3, 355-366.
- WATERMAN, M. e VINGRON, M. (1994), "Sequence Comparison significance and Poisson Approximation", *Statistical Science*, Vol. 9, Nº3, 367-381.
- WEIR, B.S. (1990), *Genetic Data Analysis - Methods for Discrete Population Genetic Data*, Sinauer Associates, Inc. Publishers. Sunderland.
- WEIR, B.S. (1993), "Independence tests for VNTR alleles defined as quantile bins", *American Journal of Human Genetics*, 53, 1107-1113.





# APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS MULTIVARIADAS NA SEGMENTAÇÃO DA PROCURA TURÍSTICA DO ALGARVE

**Autora:**  
Maria Margarida Viçoso de Arrais Viegas

**VOLUME 2**

**2º QUADRIMESTRE DE 1997**

---

## APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS MULTIVARIADAS NA SEGMENTAÇÃO DA PROCURA TURÍSTICA DO ALGARVE

---

---

## THE TOURIST DEMAND SEGMENTATION OF THE HIGH SEASON IN THE ALGARVE

---

Autora: Maria Margarida Viçoso de Arrais Viégas  
Equiparada a Professora Adjunta, Núcleo de Métodos Quantitativos, Escola  
Superior de Gestão, Hotelaria e Turismo - Universidade do Algarve

### SUMÁRIO:

- O objectivo geral deste trabalho é a identificação de diferentes segmentos de turistas que visitam o Algarve na época alta. É utilizada uma segmentação por benefícios em que as variáveis critério são os factores de escolha do Algarve como destino turístico.

O objectivo geral do trabalho engloba três objectivos de natureza específica:

- (i) Identificação de uma estrutura de dimensões de escolha através da utilização de técnicas de análise de componentes principais;
- (ii) Identificação de diferentes segmentos de turistas, com base nas dimensões de escolha, através de técnicas de análise de clusters;
- (iii) Identificação das dimensões que mais discriminam entre os vários segmentos.

O método de amostragem utilizado foi a amostragem por quotas e o instrumento de recolha um questionário administrado por entrevista directa na sala de embarque do Aeroporto Internacional de Faro, em 1995.

Foram identificadas sete dimensões de escolha do Algarve como destino turístico e três segmentos de turistas com base nessas dimensões. Os segmentos são descritos com base nas variáveis que, através de testes F ou  $\chi^2$  de Pearson, apresentam diferenças significativas entre eles. Os resultados da análise discriminante indicam a existência de três dimensões de escolha importantes na discriminação entre os segmentos de procura turística.

### PALAVRAS-CHAVE:

- *Investigação em Turismo, Marketing Turístico, Segmentação por Benefícios, Segmentação de Mercado.*

### SUMMARY:

- The main objectives of this research is the identification of different segments of tourists who visit de Algarve during the high season. Benefit segmentation is used.

There are three specific issues at hand:

- (i) Structure identification of choice dimensions by using principal components analysis;

- (ii) Tourist segments identification based on choice dimensions by using cluster analysis;
- (iii) Dimensions which differentiate the most amongst segments are identified by discriminant analysis.

Quota sampling was the procedure used and the data collection instrument was a questionnaire made by direct interview at Faro International Airport in 1995.

Seven choice dimensions and three tourists segments of the Algarve as a summer tourist destination were identified. The segments are described using the variables which present significant differences (analysis of variance and Pearson's  $\chi^2$  test were performed) amongst them. Discriminant analysis results show the existence of three important choice dimensions in the discrimination amongst segments.

*KEY-WORDS:*

- *Tourism Research, Tourism Marketing, Benefit Segmentation, Tourism Market Segmentation.*

## 1. INTRODUÇÃO

Embora jovem, quando comparado com sectores tradicionais como a agricultura, as pescas, o comércio, a construção civil ou a indústria transformadora, o turismo tem sido, nas últimas quatro décadas, uma das actividades económicas com maior dinâmica de crescimento, sendo actualmente reconhecido como uma poderosa força económica e social. Esta rápida evolução, associada à complexidade e diversidade inerentes ao fenómeno turístico, teve o duplo efeito de criar enormes possibilidades de investigação e de originar um défice em termos de investigação existente. Isto é particularmente verdade no caso do Algarve onde, apesar de ser o turismo a mais importante actividade económica, a investigação de natureza estatística até agora desenvolvida não vai além da identificação do perfil do turista, perfil esse definido em termos de características demográficas, despesas efectuadas e alojamento escolhido.

Este estudo pretende chegar a resultados que, para além da caracterização da procura em termos do perfil do turista, possam indicar novos caminhos a seguir, nomeadamente ao nível de novas estratégias de marketing turístico a implementar. Isto porque a segmentação da procura turística, tema deste trabalho, é uma das condições necessárias à formulação deste tipo de estratégias, das quais o Algarve está assumidamente carenciado.

É neste enquadramento que surge o presente estudo, tendo como objectivo a identificação e caracterização de segmentos de mercado e utilizando como critério de segmentação os benefícios do produto. O objecto de estudo é a procura turística do produto Algarve na época alta (meses de Julho a Setembro).

## 2. SEGMENTAÇÃO - REVISÃO DA LITERATURA

A segmentação constitui um método para estudar as características do mercado, identificando oportunidades e colocando o desafio em termos da correcta afectação de recursos para alcançar os fins estabelecidos. Dibb (1991) define três elementos básicos da segmentação:

- Segmentação - selecção das variáveis de segmentação, identificação e caracterização dos segmentos;
- *Targeting* - Selecção dos segmentos a atingir (segmentos alvo);
- Posicionamento - Tradução das necessidades dos consumidores num *mix* tangível de produto, preço, promoção e nível de serviço, de forma a que os segmentos alvo percepcionem o produto como aquele que melhor satisfaz as suas necessidades e desejos.

A revisão da literatura permitiu identificar vários critérios (variáveis) utilizados no processo de segmentação:

- Demográficos (Weaver e McCleary 1984);
- Geográficos (Andreck e Caldwell 1994);
- Psicográficos (Waryszac e Kim 1994);

- . Comportamentais:
  - . Atributos/Benefícios procurados (Carmichael 1996, Shoemaker 1994)
  - . Frequência de utilização (Andereck e Caldwell 1994)
  - . Ocasão de utilização (Shoemaker 1989)

A pesquisa mais recente em segmentação tem incidido menos sobre as tradicionais variáveis geográficas e sócio-demográficas (segmentação *a priori*) e mais sobre as variáveis psicográficas e comportamentais (segmentação *a posteriori*). A maioria dos estudos atrás referidos utilizam técnicas de análise multivariada, nomeadamente análise factorial e análise de clusters, para identificação dos segmentos de mercado.

### 3. METODOLOGIA

O estudo teve como universo todos os turistas com idade superior a 18 anos - ou inferior desde que viajassem não acompanhados por adultos - que utilizaram o Aeroporto Internacional de Faro entre Julho e Setembro de 1995 para regressar aos locais de origem.

Os dados foram recolhidos pelo método de entrevista directa realizada por alunos da Escola Superior de Gestão, Hotelaria e Turismo da Universidade do Algarve, na sala de embarque do Aeroporto. O método de amostragem foi a amostragem por quotas, tendo estas sido definidas por país de residência de acordo com o plano de vôos do Aeroporto para o período em análise.

O questionário, redigido em português, inglês, francês e alemão, permitiu recolher a seguinte informação:

- . Dados Pessoais - idade, habilitações literárias, actividade profissional e país de residência.
- . Visita ao Algarve - experiência de visita, motivo da viagem, importância dos atributos de escolha do destino, noites de permanência, tipo, local e modo de reserva do alojamento e despesas realizadas.
- . Opinião sobre as férias: classificação das férias por atributo de escolha e intenção de regressar ao Algarve.

Os atributos de escolha utilizados resultaram da pesquisa bibliográfica, tendo por base os trabalhos de Shoemaker (1994), Hu e Ritchie (1993) e Phelps (1986).

A análise dos dados repartiu-se por três fases:

- 1ª Fase - identificação das dimensões de escolha do Algarve, a partir dos 28 atributos apresentados no questionário<sup>5</sup>, através do recurso a técnicas de análise factorial.
- 2ª Fase - identificação dos segmentos de mercado, determinados pelas dimensões de escolha, através da utilização de técnicas de análise de clusters.

<sup>5</sup> Os 28 atributos foram apresentados numa escala tipo-Lickert, de 1=nada importante a 10=muito importante

3<sup>a</sup> Fase - identificação das dimensões de escolha que mais diferenciam os segmentos de mercado utilizando técnicas de análise discriminante.

#### 4. RESULTADOS OBTIDOS

##### 1<sup>a</sup> FASE - IDENTIFICAÇÃO DAS DIMENSÕES DE ESCOLHA

Nesta fase foi utilizada a Análise em Componentes Principais (ACP). A ACP é um método estatístico multivariado que permite transformar um conjunto inicial de variáveis, correlacionadas entre si, num outro conjunto de variáveis não correlacionadas, as chamadas componentes principais, que resultam de combinações lineares do conjunto inicial. O número de componentes/dimensões a extrair resultou da aplicação do critério de Kaiser que exige que estas apresentem valores próprios iguais ou superiores à unidade<sup>6</sup>.

Quadro 1: Componentes extraídas segundo o critério de Kaiser

Componentes	Valores próprios	Variância relativa (%)	Variância rel. ac. (%)
1	7.423	26.510	26.510
2	2.498	8.920	35.430
3	1.750	6.249	41.679
4	1.593	5.689	47.368
5	1.370	4.893	52.261
6	1.231	4.398	56.659
7	1.001	3.574	60.233

As componentes são extraídas por ordem decrescente de importância, assim a 1<sup>a</sup> componente é sem dúvida a mais importante, representando cerca de 27% da variância total. As 7 componentes seleccionadas por este critério representam cerca de 60% da variância total, o que segundo Hair e Anderson (1992) é uma solução satisfatória em estudos na área das Ciências ditas "não duras".

Nos quadros seguintes apresentam-se as contribuições das várias variáveis para as 7 componentes retidas<sup>7</sup>. As contribuições representam a correlação entre cada variável original e cada um dos factores e permitem interpretar a sua natureza. Para o efeito consideraram-se apenas as contribuições significantes (entre 0.4 e 0.5) e muito significantes (> 0.5), baseando-se nestas o nome a atribuir a cada componente.

<sup>6</sup> Os valores próprios representam a quantidade de variância contida por cada componente.

<sup>7</sup> As contribuições constantes dos quadros são já as obtidas depois de aplicado o método VARIMAX de rotação.

**Quadro 2: Dimensões de escolha do Algarve**

<b>Dimensão Animação (Férias Activas)</b>	
Atributos de escolha	Contribuições
Entretenimento e vida nocturna	.784
Desp. e oportunidades de recreio	.687
Praias	.581
Reputação	.568
Compras	.483
Gastronomia	.471
Qual. e disp. do alojamento	.432
<b>Dimensão Cultural</b>	
Atributos de escolha	Contribuições
Actividades culturais	.845
Atracções históricas	.842
Tipicidade	.679
Paisagem	.627
Aquisição de conhecimentos	.600
Festivais e eventos especiais	.584
<b>Dimensão Familiar</b>	
Atributos de escolha	Contribuições
Local onde vive familiar/amigo	.809
Conhecer alguém que já esteve no Algarve	.615
Bom local para crianças	.430
<b>Dimensão Lazer (Férias Passivas)</b>	
Atributos de escolha	Contribuições
Descanso e relax	.719
Clima	.713
Paisagem	.418
Praias	.412
<b>Dimensão Ambiental</b>	
Atributos de escolha	Contribuições
Baixa criminalidade	.801
Pouco trânsito	.741
Ar puro	.660
Infraestruturas urbanas	.656
Hospitalidade	.576
Nível de preços	.469
Qualidade e disp. do alojamento	.402
<b>Dimensão Acessibilidade/Novidade</b>	
Atributos de escolha	Contribuições
Acessibilidade	.657
Local onde nunca se esteve	.643
Nível de preços	.552
Barreira linguística	.480
Conhecer alguém que já esteve no Algarve	.451
<b>Dimensão Golfe</b>	
Atributos de escolha	Contribuições
Golfe	.723

**2ª FASE - IDENTIFICAÇÃO DOS SEGMENTOS DE MERCADO**

A identificação dos segmentos de mercado foi realizada através da Análise de Clusters, a qual consiste num conjunto de procedimentos estatísticos multivariados que reorganiza uma amostra de entidades em grupos relativamente homogéneos.

O dendograma resultante da aplicação do método hierárquico de Ward a uma sub-amostra de 300 indivíduos, limitação imposta pelo programa Statistica, sugeriu a existência de 3 a 8 clusters. De seguida foram analisadas através do método *k-means* as soluções de 2 a 10 clusters (Shoemaker, 1994). Finalmente, de acordo com os critérios heurísticos de escolha da melhor solução (sobreposição mínima entre clusters, dimensões aproximadas e em número que justifique a consideração de diferentes estratégias de marketing), a opção recaiu sobre 3 clusters, constituídos respectivamente por 305, 302 e 285 casos. A solução foi validada através do teste de significância a variáveis externas (Aldenderfer e Blashfield, 1985).

De referir que apesar da amostra inicial ser constituída por 1345 indivíduos, foi reduzida a 892 pelo facto dos restantes não terem indicado a importância de todos os 28 atributos de escolha. Esta sub-amostra tem uma maior presença relativa de turistas com níveis de formação mais elevados e de estudantes, em detrimento do pessoal administrativo e similares.

Os três segmentos (clusters) identificados apresentam diferenças significantes<sup>8</sup> em relação às seguintes variáveis:

- Dados Pessoais - idade, habilitações literárias, actividade profissional e país de residência.
- Visita ao Algarve - atributos de escolha do destino e tipo de alojamento.
- Opinião sobre as Férias - classificação das férias por atributo de escolha.

**Quadro 3: Dados Pessoais**

Variáveis Sócio-Demográficas	Total	Seg. I	Seg. II	Seg. III
Idade média	31.4	34.0	33.3	26.6
Habilitações Literárias				
. Ensino superior	62.0%	69.6%	69.3%	46.5%
. Secundário completo	32.3%	32.2%	26.3%	48.2%
. Secundário incompleto	4.0%	2.6%	2.6%	5.3%
. Básico	1.7%	1.8%	1.8%	0.0%
Actividade Profissional				
. Prof. científicas, técnicas, artísticas e similares	34.7%	37.2%	38.0%	28.5%
. Estudantes	23.2%	13.8%	12.1%	45.0%
. Pessoal adm. e similares	11.6%	14.2%	11.8%	8.3%
. Directores e quadros sup. adm.	11.5%	14.2%	13.3%	6.4%
. Pers. dos serviços de protecção e segurança, pessoais e domésticos	5.9%	7.3%	6.2%	3.9%
. Pers. do comércio e vendedores	5.4%	5.3%	6.7%	4.4%
. Domésticas	3.7%	4.4%	6.2%	0.5%
. Pers. das indústrias extractiva e transformadora	2.8%	3.2%	3.6%	1.5%
. Reformados	0.9%	0.4%	2.1%	0.5%
. Agricultores	0.3%	0.0%	0.0%	1.0%
País de residência (6 maiores origens)				
. Reino Unido	43.1%	56.8%	36.8%	35.2%
. Alemanha	22.2%	12.9%	24.1%	30.3%
. Holanda	9.1%	7.6%	9.7%	10.2%
. Áustria	5.2%	2.6%	5.7%	7.4%
. Irlanda	4.3%	5.6%	5.0%	2.1%
. Bélgica	4.0%	4.6%	4.0%	3.2%

<sup>8</sup> As diferenças foram identificadas através da ANOVA e da análise da  $\chi^2$ . Todos os testes foram realizados para um nível de significância de 5%.

Os segmentos I e II apresentam uma grande semelhança em termos de idade, habilitações literárias e actividade profissional. As diferenças ocorrem ao nível do país de residência verificando-se uma maior presença relativa de britânicos e irlandeses no segmento I e de alemães, austríacos e holandeses no segmento II. O segmento III caracteriza-se por uma maior presença de jovens, nomeadamente estudantes, com uma maior presença relativa de alemães, austríacos e holandeses.

**Quadro 4: Tipo de Alojamento**

Tipo de Alojamento	Total	Segmento I	Segmento II	Segmento III
Hotel	21.26%	15.72%	25.68%	22.50%
Apartamento turístico	20.00%	24.41%	18.58%	16.78%
Casa/apartº s/ resid.	14.17%	20.07%	12.50%	9.64%
Aparthotel	13.03%	9.70%	14.86%	14.64%
Casa própria	7.31%	8.36%	6.42%	7.14%
Aldeamento	6.97%	10.37%	5.10%	5.36%
Casa de familiar/amigo	4.46%	1.67%	4.05%	7.86%
Pensão	3.66%	4.00%	3.38%	3.57%
Casa/apartº c/ resid.	3.54%	1.67%	3.38%	5.71%
Campismo	2.40%	2.00%	2.36%	2.86%
Vários	1.71%	0.67%	2.03%	2.50%
Outros	1.49%	1.34%	1.69%	1.43%

O segmento I revela uma maior apetência pelo tipo de alojamento *self-catering* (apartamento turístico, casa/apartamento sem residentes e aldeamento). Os segmentos II e III apresentam um padrão de escolha de alojamento mais próximo, sendo de destacar uma maior preferência pelo hotel por parte do segmento II e pela casa de familiares/amigos e casa/apartamento com residentes por parte do segmento III, o que se revela consistente com as características sócio-demográficas atrás referidas.

Quadro 5: Atributos de escolha do Algarve

Atributos de escolha	Total	Seg. I	Seg. II	Seg. III	F (ANOVA)	Teste de Scheffé (1)
Descanso e relax	8.31	8.54	8.71	7.64	23.01	3*1,2
Praias	8.28	7.53	8.37	8.99	33.33	1*2,3 e 2*3
Clima	8.10	7.87	8.43	7.98	5.68	2*1,3
Gastronomia	7.75	7.29	8.28	7.67	14.61	2*1,3
Nível dos preços	6.81	6.97	7.43	5.99	26.45	3*1,2
Ar puro	6.77	6.03	7.79	6.48	38.31	2*1,3
Disponibilidade do aloj.	6.71	6.31	7.16	6.64	7.89	1*2
Hospitalidade	6.56	5.30	7.48	6.93	60.01	1*2,3 e 2*3
Baixa criminalidade	6.35	5.89	7.56	6.35	46.58	2*1,3
Paisagem	6.30	5.09	7.64	6.17	87.08	1*2,3 e 2*3
Entretenimento e vida nocturna	6.26	4.89	6.28	7.71	80.79	1*2,3 e 2*3
Acessibilidade	6.17	6.49	6.81	5.15	31.90	3*1,2
Desporto e op. de recreio	5.88	4.80	6.33	6.56	39.09	1*2,3
Reputação	5.68	4.98	6.52	5.55	22.07	2*1,3
Tipicidade	5.67	3.91	7.16	5.95	143.48	1*2,3 e 2*3
Aquisição de conhecimentos	5.66	3.86	6.69	6.50	119.02	1*2,3
Compras	5.65	5.12	6.70	5.10	35.34	2*1,3
Local onde nunca se esteve	5.64	5.13	6.19	5.62	6.83	1*2
Infraestruturas urbanas	4.91	4.43	6.11	4.14	53.01	2*1,3
Pouco trânsito	4.85	4.35	6.22	3.94	69.88	2*1,3
Conhecer alguém que já esteve no Algarve	4.68	4.10	5.07	4.89	7.07	1*2,3
Barreira linguística	4.33	4.33	4.80	3.83	8.02	2*3
Atrações históricas	4.07	2.47	6.43	3.28	343.60	1*2,3 e 2*3
Bom local para crianças	4.04	4.06	5.05	2.96	29.19	1*2,3 e 2*3
Actividades culturais	3.88	2.12	6.17	3.34	380.33	1*2,3 e 2*3
Festivais e eventos especiais	3.61	1.98	5.34	3.54	193.44	1*2,3 e 2*3
Local onde vive familiar/amigo	3.03	2.11	3.67	3.22	21.32	1*2,3
Golfe	2.87	2.69	4.28	1.59	66.22	1*2,3 e 2*3

(1) A média do segmento à esquerda do \* é significativamente diferente da média do(s) segmento(s) à direita do \*.

Descanso e relax, praias, clima e gastronomia são os atributos de escolha mais importantes para todos os segmentos. Entretenimento e vida nocturna constitui a única excepção a figurar no topo das preferências do segmento III.

Para a classificação das férias recorreu-se a 24 dos 28 atributos de escolha, adicionados da apreciação sobre o Algarve no geral (ver Quadro 6).

Quadro 6: Classificação das férias

Atributos de classificação	Total	Seg. I	Seg. II	Seg. III	F (ANOVA)	Teste de Scheffé (1)
Praias	8.68	8.43	8.84	8.77	5.67	1*2,3
Clima	8.67	8.62	8.86	8.52	3.86	2*3
Descanso e relax	8.55	8.63	8.80	8.21	9.08	3*1,2
Algarve no geral	8.32	8.22	8.46	8.26	-----	-----
Gastronomia	8.16	8.09	8.27	8.10	-----	-----
Baixa criminalidade	7.62	7.52	7.96	7.36	6.71	2*1,3
Ar puro	7.61	7.43	8.01	7.37	9.24	2*1,3
Acessibilidade	7.36	7.54	7.56	6.94	8.85	3*1,2
Hospitalidade	7.29	6.79	7.73	7.37	12.23	2*1,3
Paisagem	7.27	6.83	7.75	7.22	14.62	2*1,3
Entretenimento e vida nocturna	7.18	6.48	7.21	7.88	27.78	2*1,3 1*3
Preços	7.17	7.27	7.52	6.69	11.94	2*1,3
Disponibilidade e qualidade do alojamento	7.11	6.97	7.44	6.90	5.62	2*1,3
Bom local para crianças	7.02	7.33	7.24	6.42	10.83	3*1,2
Desporto e op. de recreio	6.91	6.64	7.21	6.89	4.57	1*2
Compras	6.81	6.68	7.37	6.36	16.83	2*1,3
Barreira linguística	6.68	6.77	6.69	6.57	-----	-----
Tipicidade	6.57	5.94	7.18	6.58	25.30	1*2,3 e 2*3
Aquisição de conhecimento	6.13	5.19	6.73	6.46	36.89	1*2,3
Golfe	5.82	6.06	6.58	4.70	20.48	3*1,2
Infraestruturas urbanas	5.77	5.34	6.39	5.55	19.11	2*1,3
Pouco trânsito	5.55	5.24	6.20	5.18	17.33	2*1,3
Atrações históricas	5.22	4.40	6.37	4.78	68.12	2*1,3
Actividades culturais	5.09	4.33	6.03	4.84	45.41	1*2,3 e 2*3
Festivais	4.97	4.27	5.74	4.86	27.17	1*2,3 e 2*3

(1) A média do segmento à esquerda do \* é significativamente diferente da média do(s) segmento(s) à direita do \*.

Praias, clima, descanso e relax, Algarve no geral e gastronomia são os atributos de classificação das férias mais valorizados por todos os segmentos, verificando-se uma elevada consistência entre importância e desempenho pois tratam-se simultaneamente dos atributos mais relevantes para a escolha do destino (ver Quadro 5).

### 3ª FASE: DISCRIMINAÇÃO ENTRE OS SEGMENTOS

Nesta fase pretendeu-se averiguar quais as dimensões de escolha do destino Algarve que mais diferenciam entre os segmentos identificados, para o que foi utilizado a Análise Discriminante (AD). O objectivo básico da AD é descobrir quais as características (variáveis independentes métricas) que melhor diferenciam entre os vários grupos de indivíduos, constituindo estes grupos as diferentes categorias da variável endógena (variável dependente não métrica).

Estatisticamente, a análise discriminante envolve a derivação de uma combinação linear<sup>9</sup> de duas ou mais variáveis independentes que melhor discriminam entre grupos definidos à priori.

No presente caso foram estimadas, pelo método *stepwise*, duas funções discriminantes, estatisticamente significantes na discriminação entre grupos.

**Quadro 7: Coeficientes Discriminantes**

Variáveis: dimensões de escolha	Função 1	Função 2
Cultural	-.985	.288
Golfe	-.549	-.777
Animação	-.153	.748
Novidade acessível	-.008	-.553
Qualidade ambiental	-.390	-.328
Familiar	-.152	.359
Lazer	.033	-.259

Os coeficientes discriminantes representam a contribuição relativa da variável para o poder discriminante da função. O sinal algébrico apenas indica o sentido dessa contribuição. Verifica-se que as variáveis que mais contribuem para o poder discriminantes da 1ª função<sup>10</sup> são as dimensões cultural e golfe, ambas no sentido negativo, para a 2ª função<sup>11</sup> são as dimensões golfe, também no sentido negativo e animação no sentido positivo. A análise dos coeficientes discriminantes não nos permite no entanto concluir entre que grupos as respectivas funções discriminam, pelo que terão ainda de ser analisadas as médias dos grupos nas duas funções.

**Quadro 8: Médias dos Segmentos nas Funções Discriminantes**

Segmentos	Função 1	Função 2
Segmento I	1.192	-1.194
Segmento II	-1.729	-.336
Segmento III	.556	1.634

<sup>9</sup> Cada combinação linear constitui uma função discriminante na forma geral:  $Z = W_1X_1 + W_2X_2 + \dots + W_nX_n$ ; em que Z é o score discriminante,  $W_i$  os pesos ou coeficientes discriminantes e  $X_i$  as variáveis independentes.

<sup>10</sup> Coeficiente de correlação canónica = 0.785

<sup>11</sup> Coeficiente de correlação canónica = 0.762

A 1ª função discrimina principalmente entre o segmento II e os restantes, dado o valor da média para este segmento ser razoavelmente diferente dos restantes e a 2ª função entre o segmento III e os restantes.

Considerando o sinal dos coeficientes discriminantes com maior valor absoluto e as médias dos segmentos nas duas funções podemos concluir que:

- Segmento I - distingue-se por uma menor importância atribuída às dimensões Cultural e Animação.
- Segmento II - distingue-se por uma maior importância atribuída à dimensão Cultural.
- Segmento III - distingue-se por uma maior importância atribuída à dimensão Animação e uma menor importância atribuída ao Golfe.

## 5. CONCLUSÕES

Os benefícios procurados pelos turistas residentes no estrangeiro que passam férias no Algarve, no período de Julho a Setembro, podem classificar-se segundo uma estrutura de 7 componentes/dimensões independentes: Animação (Férias Activas), Cultural, Familiar, Lazer (Férias Passivas), Ambiental, Acessibilidade/Novidade e Golfe.

As referidas dimensões, identificadas a partir de 28 atributos de escolha, conduziram à identificação dos seguintes 3 segmentos de mercado:

- Segmento I: maioritariamente constituído por britânicos e com maior presença relativa de irlandeses. Classe etária com maior frequência relativa, a dos 40 aos 50 anos. Preferência pelo alojamento self-catering.
- Segmento II: reduz-se a presença de britânicos e irlandeses em favor dos alemães, holandeses e austríacos. A classe etária com maior frequência relativa é a classe dos 30 aos 40 anos. Maior preferência pelo alojamento em hotel.
- Segmento III: Maior preferência relativa de alemães, holandeses e austríacos, na sua maioria estudantes com idades inferiores a 30 anos. Maior preferência relativa pelo alojamento em casa de familiares/amigos e casa/apartamento com residentes.

Os segmentos revelaram-se homogêneos em termos do local de permanência no Algarve, da duração da estadia, dos gastos realizados e da intenção futura de regressar.

Os atributos mais importantes para a escolha do destino Algarve, comuns a todos os segmentos, são o descanso e relax, as praias, o clima e a gastronomia. A única excepção é constituída pelo atributo entretenimento e vida nocturna para o segmento mais jovem.

A dimensão Lazer (Férias Passivas) concentra a maioria dos atributos mais importantes pelo que constitui a principal dimensão de escolha para todos os segmentos. Estes no entanto diferenciam-se pela importância atribuída a outras dimensões:

O segmento I diferencia-se pela menor importância atribuída às dimensões Cultural e Animação;

O segmento II diferencia-se pela maior importância atribuída à dimensão Cultural;

O segmento III diferencia-se pela maior importância atribuída à dimensão Animação e pela menor importância atribuída ao Golfe.

Globalmente poderá afirmar-se que a principal conclusão a reter do estudo aqui apresentado é a de que a ideia de "turista médio", ou "turista sol e praia" como muitas vezes é designado o turista de época alta no Algarve, não é consubstanciada pelos resultados aqui alcançados. Pelo contrário, a análise estatística efectuada indica a existência de diferentes segmentos, com diferentes razões de escolha deste destino e com diferentes características.

Ressalvando o facto deste ser o primeiro estudo efectuado nesta área e a imprescindibilidade da confirmação destes resultados em estudos posteriores, poderá no entanto afirmar-se que, numa perspectiva de marketing, os resultados encontrados indicam que a ênfase quantitativa em mais turistas, mais dormidas, mais receitas, traduzida em práticas de marketing indiferenciado, não será provavelmente a melhor forma de afirmar o Algarve como o destino turístico competitivo que se pretende que seja. Nem sequer na época alta.

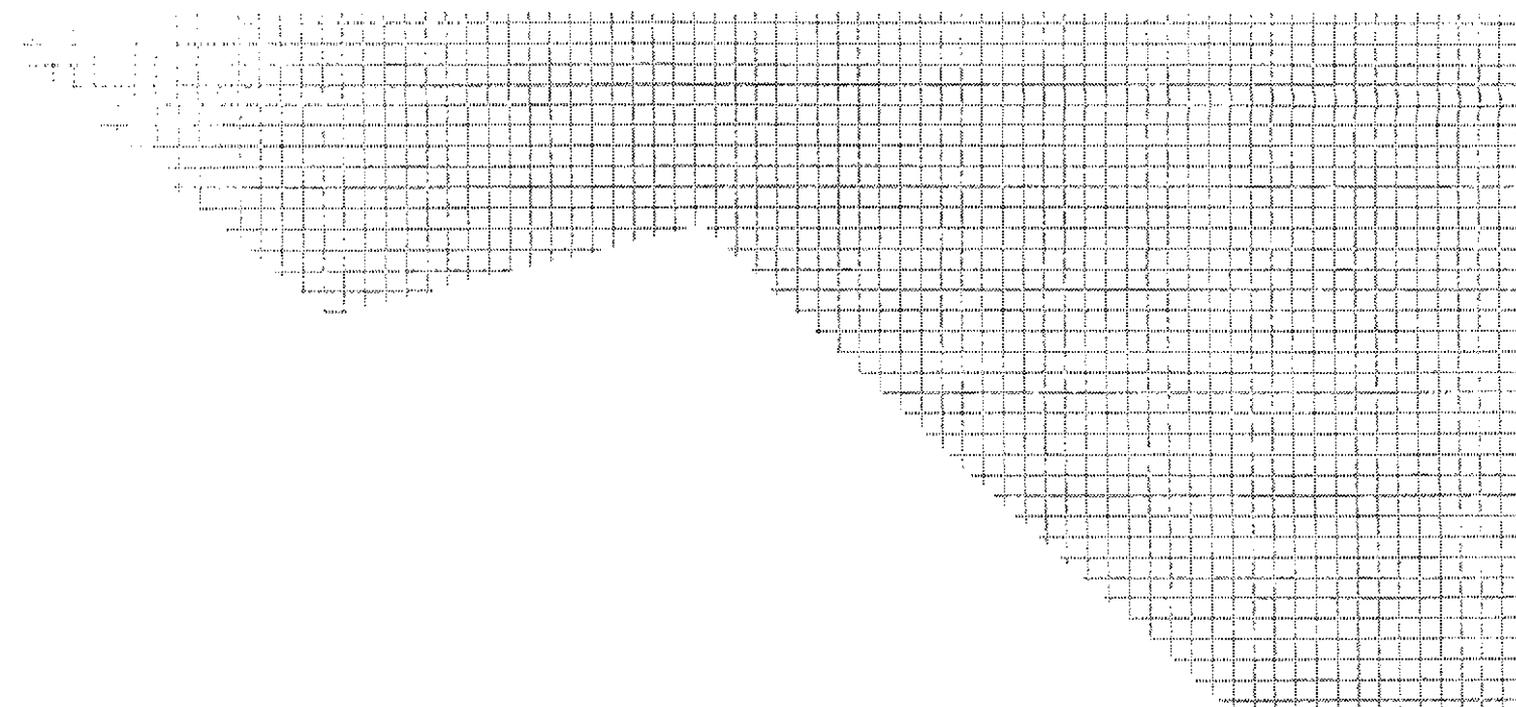
---

#### REFERÊNCIAS BIBLIOGRÁFICAS

---

- ABELL, D. E HAMMOND, J. (1979), *Strategic Marketing Planning: Problems and Analytical Approaches*, Nova Iorque, Englewood Cliffs: Prentice Hall.
- ALDENDERFER, M. E BLASHFIELD (1985), *Cluster Analysis*, Beverly Hills, CA: Sage.
- ANDERECK, K. E CALDWELL, L. (1994), "Variable Selection in Tourism Marketing Segmentation Models", *Journal of Travel Research*, (Fall), 40-46
- BAILEY, K. (1974), "Cluster Analysis" in David R. Heise (ed.), *Sociological Methodology*, São Francisco, CA: Jossey-Bass Publishers.
- BIGGADYKE, E. (1981), "The Contribution of Marketing to Strategic Management", *Academy of Management Review*, vol.6, nº4.
- BRYANT, B. E MORRISON, A. (1980), "Travel Market Segmentation and the Implementation of Market Strategies", *Journal of Travel Research*, (Winter), 2-8.
- CALANTONE, R. E JOHAR, J. (1984), "Seasonal Segmentation of the Tourism Market using a Benefit Segmentation", *Journal of Travel Research*, (Fall), 14-24.
- CALANTONE, R., SCHEWE, C. E ALLEN, T. (1980), "Targeting Specific Advertising Messages at Tourists Segments", *Tourism Marketing and Management*, Washington D.C.: George Washington University, 133-147.
- CARMICHAEL, B. (1996), "Conjoint Analysis of Downhill Skiers used to Improve Data Collection for Market Segmentation", *Journal of Travel and Tourism Research*, Vol. 5 (3), 187-206.

- CHATFIELD, C. (1991), *Introduction to Multivariate Analysis*, Cambridge, G.B.: University Press.
- CRAUSER, HARVATOPOULOS E SARNIN (1986), *Guide Pratique d'Analyse des Données*, Les Editions d'Organisation.
- DIBB, S. E SIMKIN, L. (1991), "Targeting, Segmenting and Positioning", *International Journal of Retail & Distribution Management*, vol.19, n°3, 5.
- HAIR, ANDERSON, TATHAM E BLACK (1992), *Multivariate Data Analysis With Readings*, Nova Iorque, Macmillan Publishing Company.
- KIM, J. E MUELLER, C. (1978), *Introduction to Factor Analysis*, Beverly Hills, CA:Sage.
- LAWSON, R. (1980), "Discriminant Analysis-An Aid to Market Segment Description", *European Journal of Marketing*, vol.14, n°7, 387-396.
- MANZANEC, J. (1984), "How to Detect Travel Market Segments: A Clustering Approach", *Journal of Travel Research*, (Summer), 17-21.
- MARASCUJLO, L. E LEVIN, J. (1990), *Multivariate Statistics in the Social Sciences: A Researcher's Guide*, Brooks/Cole Publishing Company.
- MORRISON, D. (1976), *Multivariate Statistical Methods*, McGraw-Hill.
- PRENTICE, R. (1989), "Market targeting" in Witt, S. e Moutinho, L. (eds) *Tourism Marketing and Management Handbook*, London, Prentice Hall, 247-251.
- REIS, E. E MOREIRA, R. (1993), *Pesquisa de Mercado*, Edições Sílabo.
- SAUNDERS, J. (1980), "Cluster Analysis for Market Segmentation", *European Journal of Marketing*, vol.14, n°7, 422-435.
- SHOEMAKER, S. (1989), "Segmenting the Senior Pleasure Travel Market", *Journal of Travel Research*, (Winter), 14-21.
- SHOEMAKER, S. (1994), "Segmenting the U.S. Travel Market according to Benefits Realized", *Journal of Travel Research*, (Winter), 8-21.
- VAN VEEN, W. E VERHALLEN, T. (1986), "Vacation Market Segmentation", *Annals of Tourism Research*, Vol. 13 (1),37-58.
- WARYZAK, R. E KIM, H. (1994), "Psychographic Segmentation of Tourists as a Predictor of their Travel Behavior", *Journal of Hospitality & Leisure Marketing*, Vol. 2 (4), 5-19.
- WEAVER, P. E MCCLEARY, K. (1984), "A Market Segmentation Study to Determine the Appropriate Ad/Model Format for Travel Advertising", *Journal of Travel Research*, Vol. XXIII (1), 12-16.
- WOODSIDE, A. E JACOBS, L. (1985), "Step Two in Benefit Segmentation: Learning the Benefits Realized by Major Travel Markets", *Journal of Travel Research*, (Summer), 7-11.



EM TORNO DAS RESPOSTAS TÉCNICAS DOS  
ESTADÍSTICAS OFICIAIS ÀS DECISÕES POLÍTICAS  
SOBRE OS CRITÉRIOS DE CONVERGÊNCIA -  
O CASO PARTICULAR DAS CONTAS NACIONAIS

**Autores:**  
**Ana Maria Leal**  
**Pedro Dias**



---

---

**EM TORNO DAS RESPOSTAS TÉCNICAS DOS ESTADÍSTICAS  
OFICIAIS ÀS DECISÕES POLÍTICAS SOBRE OS CRITÉRIOS DE  
CONVERGÊNCIA - O CASO PARTICULAR DAS CONTAS NACIONAIS**

---

---

---

---

**ABOUT THE TECHNICAL ANSWERS FROM OFFICIAL  
STATISTICIANS TO THE POLITICAL DECISIONS CONCERNING THE  
CONVERGENCE CRITERIUMS - THE SPECIFIC CASE OF NATIONAL  
ACCOUNTS**

---

---

Autores: Ana Maria Leal

Directora-Adjunta do Departamento de Coordenação e Integração do INE

e

Pedro Dias

Director do Departamento de Coordenação e Integração do INE e Secretário  
do Conselho Superior de Estatística.

*SUMÁRIO:*

- O presente texto visa apresentar, na sua primeira parte, a forma como no quadro comunitário europeu recente têm vindo a ser utilizados os dados de Contas Nacionais, em particular no contexto da determinação dos recursos próprios (recurso IVA e 4º recurso comunitário) e do controlo e orientação da política monetária europeia (determinação dos rácios do défice e dívida pública no PIB). Em particular, é discutida a racionalidade dos prazos de aplicação aos dados reportados, das metodologias previstas no SEC 95 ou no SEC 79.

É apresentada, em seguida, uma descrição das mais recentes decisões de carácter técnico produzidas pelo Eurostat, decorrentes da necessidade de interpretação, à luz do SEC 79, relativas a certas matérias sobre as quais este é omissivo ou pouco claro.

Finalmente, apresenta-se um resumo das decisões que maior impacto tiveram para Portugal.

*PALAVRAS-CHAVE:*

- *SEC 79 (Sistema Europeu de Contas 79); SEC 95; défice público; dívida pública.*

*SUMMARY:*

- The purpose of this text, in its first part, is to present how the data of national accounts is been used in the European community framework, in special for the own resources calculation (VAT resource and fourth community resource) and also for monitoring and guiding European monetary policy (determination of public debt and public net borrowing in percentage of GDP)

It is also discussed the rationality of the implementation delays for data reported under ESA 95 or ESA 79 methodologies.

The text continues presenting a description of the more recent technical decisions of Eurostat, in consequence of the necessity of interpretation, under ESA 79 rules, of certain cases about which ESA 79 is blank or not clear.

Finally, it is presented a short reference of the most important decisions for the Portuguese case.

*KEY-WORDS:*

- *ESA 79 (European System of Accounts); ESA 95; public net borrowing; public debt.*



## A - INTRODUÇÃO

A recente polémica a propósito das decisões, tomadas pelo EUROSTAT, quanto aos aspectos estatísticos de determinação do défice orçamental e da dívida pública, “mortiferamente” designadas pela comunicação social como “contabilidade criativa”, tornam útil, pelo menos para conforto dos estaticistas, procurar precisar alguns dos aspectos principais que determinam e justificam esta situação.

Uma breve cronologia de acontecimentos é, também neste caso, um elemento útil para a compreensão deste processo, incluindo as razões de algumas perplexidades que os economistas mais atentos foram consolidando.

Atentemos então nas seguintes datas:

- utilização do PNB na determinação do 4.<sup>o</sup> recurso próprio comunitário, a partir de 1988, com base no SEC79 - Directiva do Conselho de **Fevereiro de 1989**;
- aprovação do Sistema de Contas Nacionais (SCN 1993) - Comissão de Estatísticas das Nações Unidas em **Fevereiro de 1993**;
- utilização do SEC79 no procedimento relativo aos défices excessivos - Regulamento do Conselho de **Novembro de 1993**;
- aprovação do SEC95 - Regulamento do Conselho de **Junho de 1996** (publicado em Novembro do mesmo ano).

A relevância da alteração que se produziu com o início da utilização do PNB na determinação do quarto recurso próprio comunitário, é um dos aspectos mais salientes do processo de adopção formal da contabilidade nacional para fundamentação das decisões políticas ao nível europeu.

Passou-se de um “acordo de cavalheiros”, base da anterior utilização do quadro conceptual de referência que o SEC constitui para elaboração das contas nacionais, para uma obrigatoriedade assente em instrumentos jurídicos comunitários tradicionais.

À data da assinatura do protocolo sobre o procedimento relativo aos défices excessivos anexo ao Tratado da União Europeia (Maastricht, Fevereiro de 1992), já o agora denominado SEC95 se encontrava em fase avançada de elaboração em articulação com o SCN93, sendo unanimemente reconhecida, pelos contabilistas nacionais da Europa, a inadequação da anterior versão do SEC à realidade, em particular por não contemplar de forma explícita ou sequer implícita, diversas situações que a evolução económica das duas últimas décadas tinha tornado muito relevantes e que se repercutiam directamente na determinação das grandezas mais sensíveis para verificação do cumprimento dos denominados “critérios de convergência”, ou seja, no défice orçamental e na dívida pública.

Independentemente dos aspectos técnicos do SEC 95, sobre os quais foi possível estabelecer um amplo acordo, na base dos princípios já aprovados ao nível mundial com o Sistema de Contas Nacionais de 1993, ocorreram sucessivas dificuldades na aprovação do Regulamento do Conselho relativo ao SEC95, ligadas às questões da sua utilização no processo de decisão política.

Em particular, colocou-se a questão de fazer coincidir os dados da contabilidade nacional utilizados para a determinação do 4.<sup>o</sup> recurso próprio comunitário, e subsidiariamente para a determinação do recurso IVA, com os usados para a determinação do défice orçamental e da dívida pública, o que era incompatível com a simultaneidade de utilização de dois sistemas de contas nacionais.

Sendo, embora, unanimemente reconhecida a não actualidade do SEC79 e, portanto, a sua fragilidade conceptual face ao SEC95, não foi possível fazer prevalecer uma decisão que pudesse conduzir à adopção deste último como base para a elaboração das contas nacionais, a utilizar, no mais curto prazo, na determinação dos "rácios" relativos ao défice orçamental e à dívida pública.

As dificuldades dos países em disponibilizar em prazo adequado aqueles objectivos, contas nacionais segundo o SEC95 e, acima de tudo, a constatação da inadequação de utilizar diferentes dados para a determinação dos recursos próprios e dos "critérios de convergência", os primeiros segundo o SEC79 e os segundos de acordo com o SEC95, aliados à inoportunidade política de proceder à alteração das regras de determinação dos recursos próprios comunitários, determinaram a decisão de recorrer aos dados das contas nacionais, segundo o SEC79, para aqueles dois objectivos.

A transcrição, ordenada, de alguns dos considerandos do Regulamento do Conselho relativo ao Sistema Europeu de Contas Nacionais e Regionais na Comunidade (Reg. (CEE) n° 2223/96 de 25/6/96), facilita a compreensão do que atrás se refere:

"(1) Considerando que a concretização e o acompanhamento da união económica e monetária necessitam de **informações comparáveis, actualizadas e fiáveis** sobre a estrutura e a evolução da situação económica de cada país e/ou região;"

"(2) Considerando que a Comissão deve contribuir para a gestão da união económica e monetária e, nomeadamente, apresentar um relatório ao Conselho sobre os **progressos alcançados pelos Estados-membros no cumprimento das suas obrigações para a realização da união económica e monetária;**"

"(3) Considerando que as contas económicas representam um **instrumento fundamental para a análise da situação económica de um país e/ou região, desde que sejam elaboradas com base em princípios únicos e não diversamente interpretáveis;**"

"(4) Considerando que a Comissão deve utilizar **agregados das contas nacionais para os cálculos comunitários administrativos e, em especial, orçamentais;**

(...)

"(7) Considerando que a Comissão de estatística das Nações Unidas adoptou, em Fevereiro de 1993, o novo Sistema de contabilidade nacional (SCN) a fim de garantir, em todos os países membros das Nações Unidas, a **comparabilidade dos resultados a nível mundial;**"

(...)

"(10) Considerando que **deve ser instaurado um sistema europeu de contas tendo em vista as necessidades da união económica e monetária, o qual deverá ser utilizado para a elaboração das contas nacionais e regionais previstas em actos comunitários;**

"(11) Considerando que os resultados das contas e quadros de todos os Estados-membros elaborados de acordo com o sistema instaurado pelo presente regulamento deverão ser postos à disposição dos utilizadores pela Comissão em datas precisas, **especialmente no que se refere ao acompanhamento da convergência económica a fim de assegurar a mais estreita coordenação das políticas económicas dos Estados-membros;**"

Do conjunto de considerandos transcritos, pareceria decorrer uma conclusão que pudesse conduzir à adopção do SEC95 - recorde-se que o ano base para as contas nacionais segundo este sistema será, em todos os Estados-membros, exactamente 1995 - no mais curto prazo possível ou, pelo menos, a expressa adopção das regras do SEC95 para a interpretação de lacunas existentes no SEC79 com significativo impacto na determinação do conjunto de indicadores relevantes para o processo de decisão política a nível europeu e nacional.

Contudo, tal não foi o caso.

Embora, na prática, o recurso ao SEC95 para colmatar lacunas do SEC79 e facilitar a apreciação de casos concretos, seja um dos caminhos que tem vindo a ser prosseguido ao nível dos Grupos de trabalho técnicos que, com a participação de peritos de todos os Estados membros, apoiam a Comissão - representada pelo EUROSTAT - no processo de interpretação do SEC79, não deixa de ser verdade que, nem sempre, ao novo Sistema é reservado um papel compatível com o consenso técnico conseguido no respectivo processo de discussão autonomizado da questão da utilização dos dados das contas nacionais no processo de decisão política.

O último considerando do Regulamento que, de seguida, também se transcreve, quase permitiria abolir os 20 anteriores sem retirar coerência à parte mais conclusiva deste instrumento jurídico comunitário:

“(21) Considerando que a Directiva 89/130/CEE, EURATOM do Conselho, de 13 de Fevereiro de 1989, relativa à harmonização da determinação do produto nacional bruto a preços de mercado (PNBpm), estabelece que a comparabilidade do PNBpm é garantida pelo respeito das definições e das regras de contabilização do Sistema europeu de contas integradas e que o Regulamento (CEE, EURATOM) n.º 1553/89 do Conselho, de 29 de Maio de 1989, relativo ao regime uniforme e definitivo de cobrança dos recursos próprios provenientes do imposto sobre o valor acrescentado, prevê que, para o cálculo da taxa média ponderada do IVA, a repartição das operações tributáveis é determinada por meio das contas nacionais, elaboradas de acordo com o Sistema europeu de contas económicas integradas e que convém, para esses actos, assim como no contexto do Regulamento (CE) n.º 3605/93 do Conselho, de 22 de Novembro de 1993, relativo à aplicação do protocolo relativo aos défices excessivos, da Decisão 94/728/CE, Euratom do Conselho, de 31 de Outubro de 1994, relativa ao sistema de recursos próprios das Comunidades Europeias e da Decisão 94/729/CE do Conselho, de 31 de Outubro de 1994, relativa à disciplina orçamental, prever um período transitório para aplicação do sistema instaurado pelo presente regulamento.”

São pois adoptadas um conjunto de disposições que, de facto, mantém em vigor o SEC79, cujos dados continuam a ser utilizados em todos os processos de formação das mais importantes decisões políticas entrando, entretanto, em vigor o SEC95 - com amplas derrogações quanto ao calendário de aplicação - e cuja única previsão de utilização é feita por exclusão, ou seja, admitindo que a disposição fixando que as notificações relativas aos défices orçamentais excessivos serão, até 1 de Setembro de 1999, feitas de acordo com o SEC79, implica a utilização dos dados produzidos segundo o SEC95 na notificação seguinte.

Este contexto determinou, ao longo dos últimos dois anos, um processo de análise de diversas questões ou de natureza puramente metodológica, essencialmente decorrentes de lacunas no SEC79, ou de aplicação concreta de princípios insuficientemente desenvolvidos naquele sistema.

Muitas das questões suscitadas pelos diferentes países não se colocavam antes da definição dos denominados “critérios de convergência” pela simples razão de que do ponto de vista da determinação dos principais macro agregados da contabilidade nacional - PIB e PNB - a sua influência era nula ou irrelevante.

Não sendo, até recentemente, variáveis como o défice orçamental e a dívida pública objecto de utilização para determinação de rácios de grande relevância para o processo de decisão política, natural se tornava que a procura da máxima harmonização na elaboração das contas nacionais dos diferentes países se concentrasse em domínios de maior relevância na determinação dos macro agregados da contabilidade nacional.

Como queremos poder demonstrar pela abordagem que faremos de aspectos técnicos relevantes do processo de harmonização do tratamento de diversas situações ao nível da contabilidade nacional, os estatísticos europeus salvaguardaram devidamente a sua autonomia e independência técnicas.

A ocorrência de situações que vieram determinar, nalguns casos, alterações importantes na determinação do défice orçamental e da dívida pública nada tiveram a ver com o processo de interpretação concertada de disposições do SEC79, ou de suprimento de lacunas daquele Sistema, mas antes com uma maior atenção do poder político quanto à relevância da abordagem das contas públicas na óptica da contabilidade nacional.

O processo de harmonização das contas nacionais entre os diferentes países europeus progrediu significativamente sem prejuízo do reconhecimento de que muitas questões importantes e talvez de maior impacto sobre os indicadores em causa não puderam ser concretizadas.

O reforço da harmonização que foi conseguido, não obstante a sua relevância face ao curto período em que decorreu, não legitima que não permaneçam reservas de monta quanto à comparabilidade dos dados em causa, o que continua a retirar conteúdo técnico à rigidez, politicamente assumida, quanto à exacta verificação dos rácios definidos, em particular o défice orçamental.

---

## **B - DECISÕES TÉCNICAS VISANDO A HARMONIZAÇÃO**

---

---

### **B.1 - ASPECTOS GERAIS**

---

Como já se referiu, as questões debatidas tiveram principalmente duas origens:

- **questões de natureza metodológica** suscitadas pela interpretação do SEC 79, em especial resultantes de aspectos menos claros ou, ainda, decorrentes do facto dos procedimentos de elaboração do “Reporte dos Défices Excessivos” assentarem na apresentação de dados sobre fluxos e stocks, quando o SEC 79 apenas aborda fluxos. Desta limitação de âmbito do SEC 79 surgem alguns problemas de coerência quando se pretendem conjugar regras deste sistema com as disposições do Regulamento n.º 3605/93, relativo aos procedimentos dos défices excessivos.
- **questões suscitadas por determinadas decisões dos Estados-membros**, no âmbito da tomada de medidas tendentes a reduzir os rácios do défice e da dívida no PIB, para as quais houve que proceder à interpretação do SEC 79. Em muitos casos, houve que analisar essa nova interpretação do SEC 79 e procurar uma resposta harmonizada que aumentasse a comparabilidade entre

os diferentes países. Em alguns casos muito particulares, após a constatação de que o SEC 79 não continha a resposta para determinada situação, houve que recorrer ao SEC 95. De qualquer modo, as decisões tomadas a partir do caso concreto de um país tornavam-se automaticamente aplicáveis a todos os casos semelhantes.

Todas as decisões, antes de formalmente tomadas e comunicadas aos Estados-membros para aplicação, foram objecto de um amplo processo de discussão técnica e, finalmente, de consulta escrita junto dos membros do Comité das Estatísticas Monetárias, Financeiras e da Balança de Pagamentos (CMFB).

## B.2 - TEMAS DEBATIDOS E DECISÕES TOMADAS

### 1. DISTINÇÃO MERCANTIL/NÃO MERCANTIL

A distinção entre actividade mercantil e não mercantil é de fundamental importância dadas as consequências na composição do universo do sector das Administrações Públicas.

A decisão relativa à distinção entre actividades mercantis e não mercantis estipula que devem ser seguidos os critérios do SEC 79, os quais, se baseiam em grande parte na classificação da actividade principal da unidade em termos de NACE-CLIO<sup>12</sup>.

De acordo com o SEC 79, as unidades que desenvolvem actividades consideradas **não mercantis**, de acordo com a NACE-CLIO, devem integrar o Sector das Administrações Públicas ou das Administrações Privadas, sendo a inclusão em um ou outro dos sectores baseada na natureza dos seus recursos.

São classificadas nas **Administrações Públicas** as unidades institucionais cuja actividade principal seja não mercantil e cujos recursos principais provenham de pagamentos obrigatórios realizados por outras unidades.

Por outro lado, são classificadas nas **Administrações Privadas** as unidades cuja actividade principal seja não mercantil e cujos recursos principais (excepto os resultantes de eventuais vendas) provenham de contribuições voluntárias efectuadas pelas Famílias ou de rendimentos de propriedade.

O SEC 79 determina que a produção de bens e de alguns serviços previamente especificados é, por definição, uma actividade mercantil. Por outro lado, existe um outro conjunto de serviços, cuja produção é, por convenção, considerada uma actividade **não mercantil**. Finalmente, a produção dos restantes serviços será considerada uma actividade mercantil se **mais de 50%** dos recursos correntes forem provenientes da venda dos respectivos produtos e não mercantil em caso contrário.

A decisão especifica que, em caso de dúvida sobre a definição de “vendas” a utilizar no critério dos 50%, devem ser seguidas as orientações do § 3.33 do SEC 95, isto é:

“ as “vendas” abrangem as vendas, excluindo os impostos sobre os produtos mas incluindo todos os pagamentos efectuados pelas administrações públicas ou pelas instituições da União Europeia e concedidas a toda a espécie de produtor neste tipo de actividade, isto é, incluem-se todos os pagamentos ligados ao volume ou ao valor da produção, mas excluem-se os pagamentos efectuados para cobrir um défice geral” (SEC 1995, § 3.33, a).

<sup>12</sup> NACE-CLIO - Nomenclatura de Actividades das Comunidades Europeias para o estabelecimento dos quadros de entradas-saídas.

## 2. ASSUNÇÕES, ANULAÇÕES E CANCELAMENTOS DE DÍVIDAS

O SEC 79 apenas refere estas operações de uma forma muito geral no seu § 4112 g), considerando-as como outras transferências de capital.

A necessidade de tratar de diferente forma alguns casos particulares já havia sido abordada no âmbito dos trabalhos preparatórios para a elaboração do SEC 95. Este, no seu § 5.16 indica que “...A operação de contrapartida de um assunção ou anulação de dívida é classificada na categoria de “transferências de capital” (D9) e é registada na conta de capital”.

Contudo, são de considerar os seguintes casos particulares de assunções ou anulações de dívidas:

- a) se a operação se faz entre uma quase-sociedade e o seu proprietário, será de registar em operação financeira (Acções e outras participações).
- b) se a operação se faz com uma empresa pública que desaparece enquanto unidade institucional no Sistema, não será de registar qualquer operação financeira, considerando-se como outras variações no volume de activos (Outros fluxos).
- c) se a operação se faz em benefício de uma empresa pública no âmbito de um processo de privatização a concretizar a curto prazo, então a contabilização deve fazer-se em operações financeiras (Acções e outras participações).
- d) a anulação total ou parcial de crédito mal parado pelos credores e o cancelamento unilateral de um passivo pelo devedor não são de classificar em operações financeiras, sendo de considerar como outras variações no volume de activos (Outros fluxos). A recusa de pagamento por parte do devedor não é reconhecida pelo Sistema.

Os casos particulares previstos de a) a d) não afectam o défice.

Este é um dos casos em que as indicações do SEC 79 não são suficientemente claras para permitir dar resposta a este tipo de questões, pelo facto de não serem contempladas as contas de património e, nessa medida, não se encontrarem previstas opções que se traduzem em variações patrimoniais. Por isso, foi decidido, relativamente a esta matéria, seguir as orientações do SEC 95 (§ 5.16)

## 3. TRATAMENTO DAS RECEITAS DA VENDA DIRECTA OU INDIRECTA DE ACÇÕES E OUTRAS PARTICIPAÇÕES

A análise desta situação e a respectiva decisão foram, no contexto em causa, da maior relevância pois visavam clarificar o tratamento a atribuir às receitas de privatizações. Foram analisados os casos seguintes:

- o produto da venda por uma unidade das Administrações Públicas relativo a acções ou participações que detenha; neste caso, falar-se-á de “venda directa”.
- o pagamento a uma unidade das Administrações Públicas efectuado por uma empresa em que ela participe ou sobre a qual exerça controlo, resultante do produto da venda de um conjunto de acções e participações noutra empresa; neste caso, falar-se-á de “venda indirecta”.

A decisão tomada consistiu em considerar que a venda directa ou indirecta de acções deve ser registada como operação financeira sempre que se trate de um caso de

privatização importante e claramente identificado. Nestes termos, a obtenção de receitas de privatizações não tem uma influência directa positiva no défice.

#### **4. MOMENTO DE REGISTO DAS OPERAÇÕES DE DISTRIBUIÇÃO: CASOS DE DECISÕES DE TRIBUNAL COM EFEITO RETROACTIVO**

A questão foi suscitada pela existência na Irlanda de uma decisão judicial desfavorável ao Estado irlandês e pela qual foram reconhecidos direitos às famílias em matéria de prestações sociais.

Esta situação decorre do facto de, em certas situações, as sentenças judiciais produzirem efeitos retroactivos, colocando-se a questão relativamente ao período no qual deverão ser registados os eventuais compromissos daí resultantes: o período no qual o compromisso foi originado, o período no qual a decisão é produzida ou o momento em que o pagamento é efectuado?

Para responder a esta questão há que ter em consideração que o SEC 79 indica no seu § 708 que “ O momento de registo das operações de repartição é o momento em que os pagamentos se tornam exigíveis, ou seja, quando surge a obrigatoriedade de pagar. Porém, para certas operações é difícil precisar o momento em que um montante se torna exigível, considerando-se então o momento em que se efectua o pagamento em numerário ou, na sua falta, o momento do fornecimento efectivo das prestações”.

Há ainda que ter em consideração outros elementos adicionais para decidir quanto ao momento de registo de um compromisso:

- deverá existir um reconhecimento do compromisso por parte do devedor
- deverá existir uma primeira estimativa do valor desse compromisso

Assim, sempre que exista uma disputa relativamente a direitos / compromissos entre duas partes, poderá ser exigido o recurso a tribunais, que irão decidir quanto à existência ou não do compromisso e qual o seu valor, decisão que deve ser vista como aplicável a ambas as partes a partir do momento em que não seja mais possível o acesso ao recurso. Nestes casos, foi decidido que o compromisso deve ser registado no período em que ficou indiscutivelmente estabelecido, isto é, no momento da decisão do tribunal.

#### **5. OPERAÇÕES ENTRE O ESTADO E AS EMPRESAS PÚBLICAS RELACIONADAS COM O FINANCIAMENTO DE REGIMES DE PENSÕES**

A análise desta questão foi suscitada a partir da existência de uma transferência para o Estado francês de montantes associados ao sistema de pensões de funcionários da France Telecom. Contudo, através desta decisão procurou-se dar uma resposta mais ampla, que pudesse aplicar-se a outras situações de transferência de valores para o Estado, em simultâneo com a transferência de responsabilidades relativamente ao pagamento futuro de pensões.<sup>13</sup>

Assim foram analisadas duas situações tipo:

- sistemas que operam em regime de repartição
- sistemas que operam em regime de capitalização

<sup>13</sup>Importa referir que esta decisão por ter sido motivada pelo caso da France Telecom se refere a situações em que a empresa é responsável pelo pagamento de pensões aos seus próprios funcionários, o que é um caso bastante comum em diversos países relativamente a empresas públicas.

Para fundamentar a decisão, tomou-se em consideração o facto de o SEC 79 reconhecer direitos (activos) das Famílias sobre as reservas dos sistemas que operam em regime de capitalização (a variação dessas reservas é considerada como parte da poupança das Famílias).

Assim, no caso de ocorrer uma transferência de valores de uma empresa pública para o Estado acompanhada da transferência de responsabilidades com pensões, tal montante deve ser sempre registado em transferências de capital, o que provoca um efeito positivo no défice. Contudo, nos casos de sistemas que operam em regime de capitalização há que registar uma outra transferência de capital de igual montante do Estado para a empresa pública, que traduz a assunção do passivo da empresa (activo das Famílias) relativo aos encargos futuros com pensões. Neste caso (regime de capitalização), o efeito final no défice é nulo.

## 6. JUROS

O momento de registo dos juros segundo o SEC 79 é o momento do seu vencimento. Contudo, a aplicação desta regra torna-se algo difícil quando se trata de títulos não convencionais emitidos com prémio ou desconto e cujo período de vida é superior a um ano. Recorde-se que o SEC 79 não cobre de uma forma adequada os rendimentos de todos os tipos de títulos actualmente existentes, em consequência da inovação financeira ocorrida desde a época em que foi elaborado.

As primeiras análises e decisões incidiram sobre casos de interpretação relativamente fácil:

- **Juros de obrigações de cupão zero**<sup>14</sup> - foi decidido que a diferença entre o preço de emissão e o preço de amortização deve ser considerado como juro no momento da amortização
- **Juros capitalizados de títulos e outros instrumentos financeiros cobertos pelo SEC 79**<sup>15</sup> - foi decidido que:
  - o juro deve ser registado separadamente do capital
  - o registo dos juros deve ser feito no momento do seu pagamento, não sendo necessário reparti-los por mais de um período.
  - a dívida deve ser registada excluindo o valor dos juros corridos e não pagos (registo bruto de acordo com o indicado no Regulamento n.º 3605/93).
- **Juros de obrigações fungíveis**<sup>16</sup> (cupões corridos) - Atendendo à existência de tranches diferentes com o mesmo cupão, o mais provável é que o preço de emissão suba relativamente às últimas tranches, para permitir a mesma taxa de remuneração aos investidores. Esta diferença, em geral, não é elevada.

Nas contas das Administrações Públicas os cupões corridos devem ser contabilizados como débitos a curto prazo na rubrica Desfasamentos contabilísticos não afectando, por isso, o défice.

<sup>14</sup> As obrigações de cupão zero são obrigações emitidas sem o pagamento expresso de juros, sendo a remuneração do capital efectuada através da diferença entre o valor de emissão (mais baixo) e o valor de amortização (mais alto).

<sup>15</sup> Os juros capitalizados são aqueles que não são pagos no final do período em que são devidos, adicionando-se assim ao capital.

<sup>16</sup> As obrigações fungíveis são obrigações emitidas sob determinadas condições previamente fixadas, mas para as quais a linha de emissão se faz através de várias tranches, todas elas com as mesmas taxas e a mesma data de pagamento dos cupões.

Os restantes casos, por se revestirem de maior complexidade, foram de resolução menos imediata. Para fundamentar as decisões foram analisadas as referências relevantes existentes no SEC 79 em relação ao momento de registo dos juros, que são as seguintes:

- §708 - Momento de registo dos juros efectivos - “momento do seu vencimento. Se os juros se reportam a vários períodos contabilísticos não é necessário reparti-los pelos diferentes períodos”
- § 706 Avaliação das operações financeiras- Indica o tratamento a efectuar à diferença entre o preço de emissão e o valor nominal para títulos considerados “convencionais”, isto é aqueles para os quais essa diferença é pequena:
  - a) Para títulos de curto prazo, a diferença entre o preço de emissão e o valor nominal deve ser considerada como juros a registar no momento da emissão (referência idêntica no § 539, nota (1) sobre os Títulos de curto prazo). Esta diferença tem, assim, um impacto no défice.
  - b) Para títulos de médio e longo prazo a diferença entre o preço de emissão e o valor nominal não deve ser considerada como juros mas sim como ganhos e perdas.

Houve então que definir curto / médio e longo prazo<sup>17</sup> e títulos convencionais<sup>18</sup>. Após essa definição foram analisados alguns tipos de títulos:

- **Obrigações de desconto profundo** <sup>19</sup> - Para este tipo de títulos a diferença entre o valor de emissão e o valor nominal deve ser considerada como juros a registar no momento da amortização do título (decisão consistente com a dos títulos de cupão zero).
- **Obrigações indexadas** - Foram considerados os casos de indexação a um índice de preços ou a um activo financeiro (como, por exemplo, divisas ou ouro).

No caso de indexação a um índice de preços o aumento de valor do capital resultante da evolução do índice deve ser registado como juros no momento da amortização do título.

No caso de indexação a um activo financeiro o aumento do valor do capital deve ser registado em ganhos e perdas, à semelhança do previsto para as emissões de títulos em moeda estrangeira.

- **Juros de obrigações lineares** <sup>20</sup>- As obrigações lineares são, à semelhança das obrigações fungíveis, emitidas em várias tranches a partir de uma mesma linha, isto é com a mesma taxa de juro nominal e um plano de amortização idêntico para o pagamento dos cupões e para a amortização do título.

Estas obrigações podem, no entanto, ser emitidas ao longo de vários anos. A evolução da taxa de juro de mercado pode determinar o aparecimento de prémios ou descontos quando se compara o preço a que são colocadas com o preço inicial de

<sup>17</sup> Curto prazo é aqui considerado um período até 12 meses.

<sup>18</sup> Títulos convencionais são aqueles para os quais a diferença entre o valor de emissão e o valor nominal é reduzida ( decorrente do § 706).

<sup>19</sup> As obrigações de desconto profundo são obrigações emitidas abaixo do valor nominal, remuneradas com taxas de juro igualmente inferiores às do mercado. O Eurostat decidiu precisar esta definição nos termos seguintes: Obrigações de desconto profundo são aquelas para as quais o cupão nominal representa mais de 50% do rendimento total no limite da maturidade (rendimento calculado por referência ao valor de emissão).

<sup>20</sup> As obrigações lineares são obrigações emitidas em tranches em momentos diferentes do tempo, por vezes ao longo de vários anos, sem alteração da data de pagamento do cupão.

emissão. Esses prémios ou descontos podem ser significativos, visto que a emissão pode desenrolar-se ao longo de vários anos. O problema que se coloca é como classificar essa diferença entre o valor nominal e o valor de emissão no momento de colocação de uma nova tranche.

A decisão distinguiu as tranches emitidas durante os 12 meses iniciais das emissões em períodos posteriores.

Assim, para as tranches emitidas nos 12 meses iniciais a diferença entre o valor nominal e o preço de emissão deve ser considerada em ganhos e perdas, não afectando por isso o défice.

Para cada uma das tranches emitidas para além de 12 meses após a inicial, a diferença entre o valor nominal e o preço de emissão (desconto ou prémio), deve ser registada em juros, com consequências no défice.

#### 7. *SWAPS* DE TAXAS DE JURO E *SWAPS* DE DIVISAS

Nos casos de *swaps* de taxas de juro, apenas os fluxos líquidos de juros devem ser registados. Assim, a influência no défice limitar-se-á à diferença entre os fluxos de juros que as duas partes envolvidas acordaram trocar.

Nos *swaps* de divisas, de acordo com o art. 5º do Regulamento 3605/93, a dívida externa deve ser registada anualmente de acordo com as taxas de câmbio de mercado e não com a taxa de câmbio do contrato *swap*. Neste sentido, estes acordos não influenciam a dívida para efeitos dos procedimentos dos défices excessivos.

#### 8. VENDAS DE OURO PELOS BANCOS CENTRAIS

O tratamento contabilístico das vendas do ouro deve ser realizado na mesma linha da venda directa de acções e outras obrigações, isto é, como operação financeira. Nesse sentido, as eventuais receitas da venda de ouro pelos Bancos Centrais transferidas para o Estado não podem ser utilizadas para beneficiar os défices.

#### 9. DEFINIÇÃO DE ADIANTAMENTOS FINANCEIROS

Foi analisado, neste contexto, o caso do “Eurotax” italiano. Em Itália os empregados acumulam um montante de poupança em fundos de salários, registados no balanço das empresas, sendo os saldos entregues aos empregados quando estes abandonam as empresas. São devidos impostos sobre esses Fundos na altura do seu levantamento. Em 1996 o governo italiano determinou que durante 1997 deveriam ser pagos 5% dos montantes acumulados por esses Fundos. Esta disposição legal tem como consequência antecipar o momento em que parte do correspondente imposto é devido, visto que o restante continua a ser devido no momento em que os empregados abandonam a empresa.

Foi considerado que, este imposto não deve ser considerado um adiantamento financeiro, mas sim um novo imposto devido em 1997, sendo pois, de registar como tal nesse ano. Esta decisão tem um impacto positivo no défice italiano em 1997.

#### 10. *LEASING* FINANCEIRO

O SEC 79 não estabelece qualquer distinção de tratamento entre a actividade de *leasing* operacional e de *leasing* financeiro. Alguns países, entre os quais Portugal,

vinham a tratar de forma diferente o *leasing* financeiro, enquanto operação financeira, reconhecendo a sua função equivalente à actividade de concessão de crédito. Contudo, da análise efectuada conclui-se que à luz do SEC 79, e contrariamente ao SEC 95, não deveria ser feita qualquer distinção entre estes dois tipos de locação.

Assim, o *leasing* financeiro deve ser registado como produção de serviços não financeiros. Os activos adquiridos em regime de locação financeira devem ser registados em FBCF do locador e o montante total das rendas pagas deve constituir consumo intermédio do locatário e produção do locador

#### **11. CLASSIFICAÇÃO DAS ENTIDADES NACIONAIS ACTUANDO EM NOME DA UE (FEOGA, ETC)**

A questão prende-se com a dupla natureza da actividade frequentemente desenvolvida pelos organismos de regulamentação de mercado actuando em nome da UE. Na verdade, essa intervenção pode, em alguns casos, ser feita através da compra / venda no mercado de determinados produtos ou através da distribuição de subsídios. A primeira, normalmente será uma actividade mercantil e a segunda, uma actividade não mercantil.

Se as contas das unidades institucionais que realizam as actividades de regulamentação de mercado e distribuição de subsídios, não permitirem isolar cada uma dessas actividades de modo a constituir ramos de actividade separados, então as unidades deverão ser classificadas globalmente nas Administrações Públicas se a percentagem dos custos suportados com a intervenção no mercado relativamente aos custos totais, for inferior a 80%.

Daqui decorre que, consoante o modelo organizacional adoptado para estas unidades institucionais pelos diferentes países, assim integram ou não o sector das Administrações Públicas.

#### **12. FUNDOS DE PENSÕES**

Foram analisadas as condições segundo as quais alguns fundos de pensões funcionando em sistema de capitalização devem ou não ser classificados nas Administrações Públicas (Segurança Social).

Os critérios a ter em conta para inclusão nas Administrações Públicas são os seguintes:

- as unidades em análise devem poder ser consideradas unidades institucionais na medida em que disponham de um conjunto completo de contas e de autonomia de decisão.
- as prestações concedidas não tenham em conta a exposição individual ao risco, isto é, sejam sistemas de pensões construídos na base de um equilíbrio financeiro colectivo.

#### **13. DÍVIDAS DE EMPRESAS PÚBLICAS COM GARANTIA PELO ESTADO**

A dívida de empresas públicas garantida pelo Estado deve integrar directamente a dívida pública se forem verificadas as seguintes condições:

- A lei que autoriza a emissão de dívida indique claramente que a obrigação de pagamento pertence ao Estado.
- O orçamento de Estado especifique anualmente os montantes a pagar.

- Os pagamentos (capital e juros) associados a esta dívida serem sistematicamente assegurados pelo Estado.

#### 14. FINANCIAMENTO E EXPLORAÇÃO DE INFRA-ESTRUTURAS PÚBLICAS PELO SECTOR DAS SOCIEDADES <sup>21</sup>

Em anos recentes os governos têm recorrido a novos métodos de financiamento das vultuosas despesas de construção de infra-estruturas públicas. Esses métodos, envolvem frequentemente o sector privado, associando-o, em graus diferenciados de acordo com cada caso concreto, às fases de financiamento, construção ou exploração das referidas infra-estruturas. Os casos conhecidos vão desde a situação mais simples em que a entidade privada é responsável pelo financiamento e construção, tomando o Estado posse imediata logo que a infra-estrutura se encontra construída, a situações mais complexas envolvendo a concessão do activo a construir ou ainda de outros já existentes. Existem ainda algumas variantes, em particular, ser o próprio Estado a assegurar o pagamento, total ou parcial, dos serviços prestados.

Foram identificados 6 casos-tipo que se apresentam em seguida em simultâneo com o método de registo decidido:

- 1) **O Governo encarrega uma entidade privada das fases de financiamento e construção das infra-estruturas públicas e toma posse da infra-estrutura imediatamente após a construção. Como exemplo foram referidas situações ocorridas na Alemanha relativamente a financiamentos de infra-estruturas de transportes.**
  - A solução apontada foi no sentido de registar de imediato um valor de FBCF no Estado como contrapartida da assunção de um passivo. O passivo irá sendo abatido à medida que o Estado for efectuando os respectivos pagamentos.
  - Efeitos: no momento inicial agrava o défice e a dívida do Estado. A dívida irá sendo reduzida à medida que são efectuados os pagamentos.
- 2) **A entidade privada encarrega-se das fases de financiamento e construção e é-lhe atribuída a concessão por toda a vida útil do activo. O Estado não efectua quaisquer pagamentos. O caso examinado foi o da ponte entre a Dinamarca e a Suécia.**
  - A solução apontada foi de não efectuar qualquer registo nas contas do Estado em nenhum dos períodos. O activo deve constar do património da empresa concessionária que recebe igualmente as portagens.
  - Efeitos: os efeitos nas contas do Estado são nulos em qualquer dos períodos.
- 3) **A entidade privada encarrega-se das fases de financiamento e construção e é-lhe atribuída a concessão por toda a vida útil do activo. O Estado, que consome serviços produzidos a partir da referida infra-estrutura, efectua as aquisições de serviços correspondentes. Os exemplos analisados referem-se à parte das infra-estruturas públicas no Reino Unido**

<sup>21</sup> Os casos em análise abrangem a construção de infra-estruturas públicas efectuada pelo sector das "Sociedades e quase-sociedades não financeiras" (definição SEC79) o qual inclui empresas públicas (com capital total ou maioritariamente público), quase-sociedades públicas e sociedades privadas. No presente texto utilizar-se-á o termo de "sociedades" como forma abreviada de "Sociedades e quase-sociedades não financeiras".

efectuadas a partir de financiamentos privados (“PFI - Private Financing Initiatives”) e nas quais não existe direito de reversão do activo.

- A solução adoptada é idêntica ao caso 2), havendo, no entanto, que registar em consumo intermédio do ano nas contas do Estado as respectivas despesas de consumo.
  - Efeitos: neutralidade em termos de défice e dívida no que respeita às despesas de construção e manutenção; contudo, neste caso o défice é agravado anualmente pelo valor da aquisição de serviços efectuada pelo Estado.
- 4) A construção da infra-estrutura encontra-se associada a uma concessão que se estende por um período inferior à vida útil do activo, findo o qual se encontra previsto o direito de reversão para o Estado. Este não efectua quaisquer pagamentos pela utilização dos serviços. Exemplo referido, a construção, seguida de concessão à Lusoponte, da Ponte Vasco da Gama.
- A solução proposta é considerar o investimento inicial no sector das Sociedades, não se efectuando quaisquer registos na conta do Estado; no final do período de concessão a ponte entra no Balanço do Estado pelo valor residual, registando-se nesse momento como FBCF do Estado em contrapartida de um registo em operação não financeira de um montante idêntico.
  - Efeitos: O efeito nas contas do Estado é nulo, no período inicial, tanto no défice como na dívida, visto que nenhum registo é efectuado. O efeito é igualmente nulo no final da concessão, visto que se regista um fluxo em recursos e outro, de igual valor, em empregos. Esta situação reflecte o facto de não ter sido efectuado qualquer pagamento efectivo do Estado à empresa.
- 5) A construção da infra-estrutura encontra-se associada a uma concessão que se estende, igualmente por um período inferior à vida útil do activo, mas o Estado efectua aquisições de serviços, à semelhança do caso 2). Os exemplos analisados foram as “PFI - Private Financing Initiatives” do Reino Unido em que existe direito de reversão.
- A solução proposta é semelhante ao caso 4); contudo, neste caso há que registar aquisições de serviços anuais por parte do Estado.
  - Efeitos: Nulos, tanto no défice como na dívida, no que respeita às consequências do registo de FBCF e à semelhança do caso 4). Os pagamentos relativos à aquisição de serviços realizados anualmente, serão de registar em consumo intermédio agravando o défice desses anos.
- 6) Concessão de um activo existente associada ou não à concessão de uma infra-estrutura a construir. Como exemplo foi analisado o caso da concessão em exploração da ponte 25 de Abril.

Este caso apresentou-se bastante complexo, visto não existirem fluxos de pagamento efectivos que traduzam a remuneração que o Estado é suposto receber ao ceder um activo para exploração a um terceiro. A contrapartida encontra-se incluída no valor da nova ponte Vasco da Gama no momento da reversão, em particular permitindo encurtar o período de concessão e possibilitar ao Estado tomar posse mais cedo do activo em questão.

A solução proposta para o caso concreto da ponte 25 de Abril, tendo igualmente presente que a sua concessão se insere na operação de construção da nova ponte, foi de

considerar igualmente a sua neutralidade no défice. Em particular, não existindo qualquer pagamento efectivo esta concessão não poderá de forma alguma beneficiar o défice durante o período em que vigore. A solução admitida é de não fazer qualquer registo em nenhum dos períodos.

**Efeitos:** Nulos, simultaneamente na dívida e no défice, tanto no início como no fim do período de concessão.

Este caso tal como se apresenta é neutro para as contas do Estado. Contudo, podem existir algumas particularidades a partir deste caso-tipo. Por exemplo, se não existisse uma concessão associada como no caso das duas pontes portuguesas, era de esperar que o Estado recebesse rendas efectivas pela cedência do direito de exploração. Esses pagamentos seriam de registar em rendas, beneficiando o défice nos anos abrangidos.

Podem também imaginar-se que a concessão de um activo existente possa ser seguida de um consumo dos respectivos serviços por parte do Estado. Nesse caso, os défices do Estado seriam agravados anualmente pelo valor das aquisições efectuadas.

#### 18. TRATAMENTO DOS SEGUROS À EXPORTAÇÃO COM A GARANTIA DO ESTADO

Esta questão foi suscitada pelo facto de, a determinada altura, terem existido opiniões distintas por parte de Portugal e do Eurostat sobre o registo dos pagamentos do Estado português relativos às garantias de crédito à exportação, actividade exercida pela COSEC e que pode beneficiar da garantia do Estado português.

De uma forma geral, o seguro de crédito à exportação assume características diferenciadas da actividade seguradora normal, situação que lhe permite, em regra, beneficiar da garantia do Estado. Esta situação decorre do elevado risco associado a este tipo de seguro, claramente incomportável se tivesse de ser totalmente integrado no valor do prémio. A situação particular de cada país leva a que, em alguns casos, a unidade que exerce essa actividade seja classificada no sector segurador enquanto noutros casos se encontra nas Administrações Públicas.

Estes organismos recebem prémios por parte dos exportadores e pagam indemnizações sempre que se verifiquem as situações enquadráveis no âmbito dos riscos segurados, normalmente falhas de pagamento do importador

A existência de garantia do Estado traduz-se na transferência para o Estado desses contratos, sendo este, nessa situação, o responsável pelo pagamento das indemnizações ao exportador. Trata-se, então, de decidir a forma de contabilização desse pagamento - operação financeira ou operação não financeira.

A decisão tomada foi no seguinte sentido:

- de uma forma geral, todas as operações associadas a essa actividade devem ser consideradas operações não financeiras.
- a cobertura pelo Estado das perdas verificadas em função dos riscos garantidos deve ser registada nas contas não financeiras, com impacto no défice das Administrações Públicas.

- reconhecendo a situação diferenciada que pode apresentar-se em cada um dos Estados-membros, admitiu-se a possibilidade de adaptar este quadro geral a cada uma das situações particulares.

---

#### C- DECISÕES COM MAIOR IMPACTO NAS CONTAS DAS ADMINISTRAÇÕES PÚBLICAS PORTUGUESAS E JÁ REPERCUTIDAS NO REPORTE DE MARÇO DE 1997

---

As decisões do Eurostat tiveram algum impacto sobre os dados calculados por Portugal, seja pela modificação que houve que introduzir em procedimentos já utilizados seja pelos tratamentos a aplicar em novas situações.

Entre as decisões tomadas a que provocou maior impacto nos valores dos défices e da dívida em Portugal foi a relacionada com o registo dos juros. Referimo-nos a modificações já repercutidas no reporte de 1 de Março de 1997 e que implicaram inclusivamente correcções a dados de anos anteriores.

As decisões mais importantes, cujo impacto é obviamente diferenciado, foram as seguintes:

- a) Juros capitalizados, com influência na dívida devido aos certificados de aforro e no défice em consequência das obrigações de capitalização automática (OCAs)
- b) Financiamento e exploração de infra-estruturas públicas pelo sector das sociedades
- c) Tratamento do seguro à exportação com garantia do Estado

Duas destas decisões [b) e c)] foram produzidas a partir da análise de casos portugueses:

- construção e concessão da ponte Vasco da Gama e concessão da ponte 25 de Abril
- seguro de crédito à exportação realizado pela COSEC com garantia do Estado

As recomendações produzidas foram no sentido da confirmação do tratamento que Portugal já previa para estes dois casos, pelo que não irão implicar modificações nos resultados que têm vindo a ser divulgados.

No quadro seguinte apresenta-se a comparação dos valores incluídos no reporte de 1 de Setembro de 1996 e os de 1 de Março de 1997, após as correcções decorrentes das decisões já notificadas à data do último reporte:

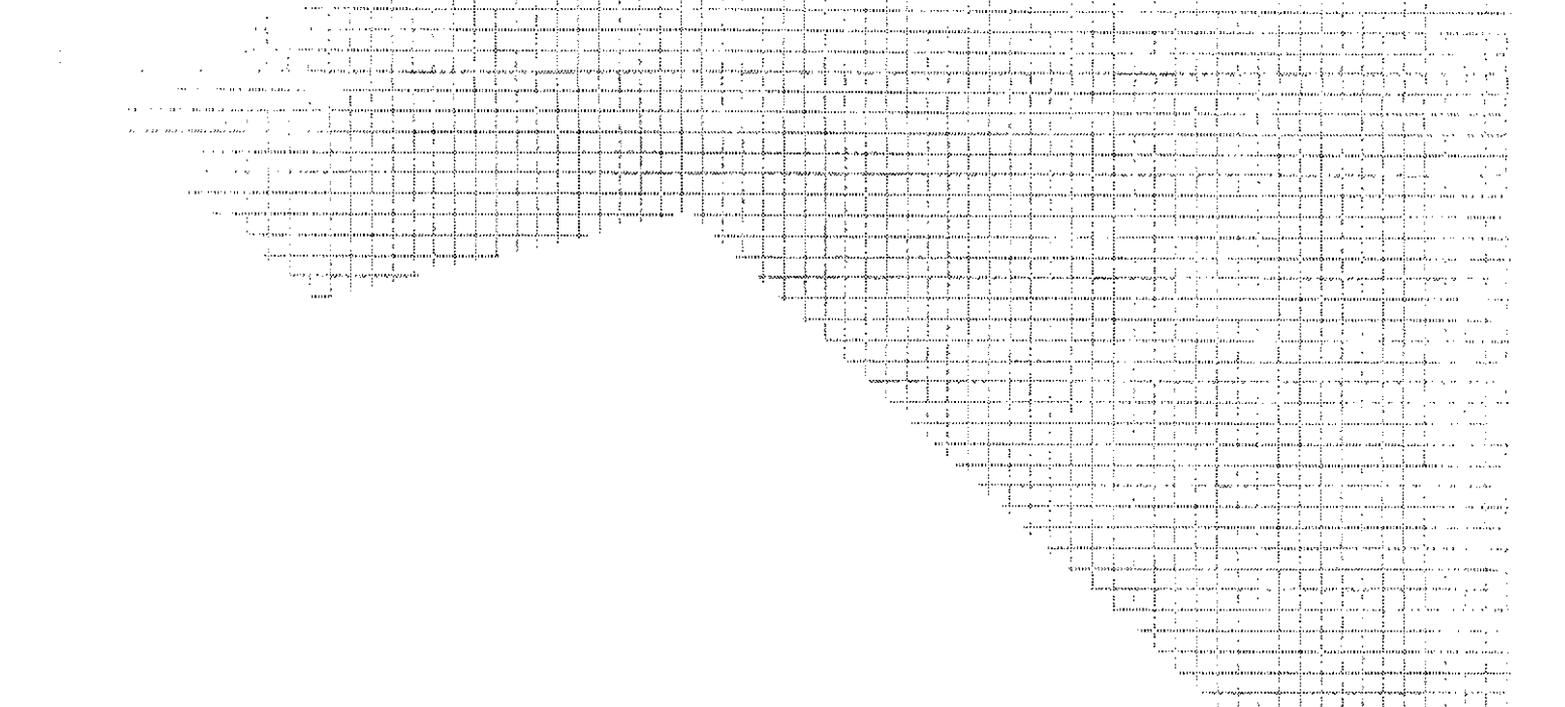
**Efeitos no défice e dívida pública das decisões notificadas até 1 de Março de 1997**

*Unidade: 1 000 milhões de esc.*

	1993	1994	1995	1996	1997
<b>Dados incluídos no reporte de Setembro 1996</b>					
Défice	-894.9	-817.6	-764.3	-654.5	-493.8
Dívida Pública	9166.5	10048.6	11036.8	11667.1	12162.9
<b>Dados incluídos no reporte de Março 1997</b>					
Défice (1)	-822.4	-881.8	-896.5	-670.6	-520.0
Dívida Pública (2)	8497.1	9331.0	10333.8	10870.3	11309.1
<b>Diferença (situação em Março 97-Setembro 96)</b>					
Défice	72.5	-64.2	-132.2	16.1	-26.2
Dívida Pública	-669.4	-717.6	-703.0	-796.8	-853.8

(1)- Excluindo a capitalização anual das OCA e incluindo os juros das OCA pagas na amortização.

(2)- Excluindo juros capitalizados acumulados das OCA e Certificados de Aforro.



# DETERMINANTES DO PROCESSO INFLACIONISTA ENTRE 1992 E 1997

**Autor:**  
**Francisco Melro**



---

---

FACTORS OF THE DISINFLATION PROCESS FROM 1992 TO 1997

---

---

Autor: Francisco Melro

Chefe do Serviço de Análise Económica de Conjuntura do Gabinete de Estudos  
do Instituto Nacional de Estatística.

*SUMÁRIO:*

- Neste texto apresentam-se as conclusões de um trabalho realizado com o objectivo de identificar as principais determinantes do processo de desinflação em Portugal desde o início de 1992. Da utilização do método das componentes principais resultam 3 vectores que justificam cerca de 78 por cento da variância da variação homóloga mensal dos índices de preços no consumidor dos 75 grupos de produtos que integram o índice de preços geral (excluindo habitação). Do estudo da primeira componente principal, que identifica a tendência de fundo dos preços dos diferentes grupos de produtos, conclui-se que a desinflação que se seguiu a 1992 foi essencialmente determinada pela recessão de 1993, mas que tanto a melhoria da produtividade como a abertura do mercado nacional contribuíram também para a descida verificada. O estudo das segunda e terceira componentes sugere a ocorrência de outras oscilações ainda importantes da inflação, em torno dessa tendência de fundo, associadas com evoluções irregulares muito significativas dos preços de alguns produtos e com a flutuação cambial, atribuindo-se a restante variância da inflação a comportamentos idiossincráticos dos preços dos diferentes grupos de produtos. Este trabalho propõe adicionalmente uma metodologia de cálculo da inflação subjacente, a partir da informação resultante da aplicação do método das componentes principais.

*PALAVRAS-CHAVE:*

- *Inflação, inflação subjacente, componentes principais.*

*SUMMARY:*

- This paper analyses the Portuguese disinflation process from the beginnings of 1992 until the beginnings of 1997. In this period, the annual rate change of consumer prices fell significantly, from 12% to 2%. We use principal components methodology to identify the independent variables that justify the variance of annual evolution of 75 indices consumer prices of groups of products. Statistical correlation between the principal components and the 75 indices and between the principal components and the different macroeconomics variables were used to study the first components. The first component justifies 60% of the among variance and the first three components justify about 78% of that variance. The first component identifies the underlying inflation and we suggest a methodology to quantify this vector. The conclusion is that the main contribution to the Portuguese disinflation process was given by the recession of 1993 but that other contributions, such as the productivity growth and the more openness of the national market, were also very important.

*KEY-WORDS:*

- *Portuguese inflation, Underlying inflation, principal components.*

**VOLUME 2**

**2º QUADRIMESTRE DE 1997**

A variação homóloga mensal do índice de preços no consumidor português no território do Continente passou de um valor próximo de 12 por cento no início de 1992 para um valor médio de 2 por cento durante os meses de Abril e Maio de 1997. Esta tendência do índice geral foi igualmente constatada na generalidade dos índices de preços elementares. Para apreender as principais determinantes deste processo de desinflação, foi analisada a evolução no território do Continente português da variação homóloga dos índices de preços no consumidor de 75 grupos de produtos que servem de base ao cálculo do índice de preços geral (sem habitação).

Com este propósito, utilizou-se o método das componentes principais para identificar as principais variáveis independentes responsáveis pela variância do conjunto das variações homólogas dos índices e procedeu-se à interpretação dessas componentes, estudando as suas relações com cada um dos índices e com diferentes variáveis macroeconómicas.

Os resultados da análise das variações homólogas dos 75 índices através do método das componentes principais revelam que as três primeiras componentes são responsáveis por cerca de 78 por cento da variância das variações homólogas dos 75 índices. A primeira destas componentes tem uma forte correlação com a esmagadora maioria das variações dos 75 índices, atingindo em alguns casos 97 por cento, e explicando cerca de 60 por cento da variância do conjunto dos índices. A segunda e a terceira componentes têm uma importância relativa muito próxima, justificando, respectivamente, cerca de 10 por cento e de 8 por cento da variância global.

A primeira componente principal identifica uma variável subjacente, não observada, que capta a tendência de fundo comum à generalidade dos 75 índices. O facto desta variável apresentar um perfil descendente muito pronunciado e explicar uma percentagem tão significativa da variância do conjunto dos índices evidencia, em primeiro lugar, que a evolução dos preços esteve ao longo deste período sujeita a choques muito intensos e, em segundo lugar, que os efeitos destes choques foram determinantes na evolução da generalidade dos índices. A análise da primeira componente sugere que tais choques terão sido gerados essencialmente pela recessão, sobretudo até ao final de 1994, mas que também a abertura do mercado nacional e a melhoria da produtividade terão agido no sentido da descida da inflação subjacente.

Assim, o trabalho realizado revela que a primeira componente apresenta um perfil tendencial, fortemente correlacionado com o nível do desemprego e com a evolução nominal dos salários. A segunda componente, apesar de apresentar um perfil próximo do ciclo do consumo, está fortemente correlacionada com alguns dos grupos de produtos que neste período apresentaram evoluções irregulares muito acentuadas. A terceira componente aparece correlacionada, de uma forma positiva ou negativa, com preços de produtos caracterizados por um comportamento muito irregular mas também parece estar influenciada com a evolução cambial. As restantes componentes, para além da terceira, surgem associadas a comportamentos idiossincráticos dos preços dos diferentes grupos de produtos.

Adicionalmente, apresenta-se também neste texto uma proposta de metodologia para o cálculo da inflação subjacente, tendo por base a primeira componente principal e a sua relação com os preços dos diferentes grupos de produtos.

Embora as três primeiras componentes justifiquem cerca de 78 por cento da variância da variação homóloga do conjunto dos 75 índices, a sua contribuição para a variância da variação homóloga mensal do índice de preços no consumidor geral

(excluindo habitação) é de 97 por cento, o que significa que a variância global dos 75 índices não explicada pelas três primeiras componentes tem muito fraca incidência na variância da variação do índice global, podendo, por isso, ser ignorada.

## **2. OBJECTIVOS**

Com este trabalho pretende-se identificar e interpretar as principais determinantes da inflação em Portugal entre Julho de 1992 e Maio de 1997. Durante este período, a variação homóloga mensal do índice de preços no consumidor geral (excluindo rendas e despesas com a conservação da habitação) passou de 11,8 por cento para 2,1 por cento. Pretende-se sobretudo identificar as causas das oscilações de preços com impacto significativo não só no índice geral mas também na generalidade dos índices elementares, não sendo estudadas algumas oscilações que, embora sendo importantes para alguns índices elementares, possuem um impacto reduzido ou momentâneo no índice de preços global e uma fraca correlação com a generalidade dos índices elementares. A evolução da inflação é aqui medida pela variação homóloga mensal do índice de preços no consumidor.

A escolha da variação homóloga mensal para medida da inflação, em vez da evolução em cadeia mensal, é justificada pela existência de uma importante componente sazonal, embora relativamente instável, nos preços de diversos bens e pelo facto da estrutura sazonal ser diferente de produto para produto. A utilização das variações em cadeia mensais tornaria muito mais difícil a identificação dos comovimentos dos diferentes índices elementares.

Como base de trabalho foram utilizados os índices de preços dos 75 grupos de produtos constantes no quadro 1. Esta decomposição do índice de preços geral é assumida como suficiente para os objectivos pretendidos.

Pretende-se como objectivo complementar obter uma metodologia de cálculo da variável subjacente, não observada, que identifica a tendência de fundo da inflação.

## **3. ANÁLISE DA EVOLUÇÃO DOS ÍNDICES DE PREÇOS ATRAVÉS DE COMPONENTES PRINCIPAIS**

Os índices de preços no consumidor resultam de um inquérito mensal aos preços de bens de consumo com maior peso na despesa do consumidor médio nacional. A agregação ponderada dos índices de preços elementares dá origem a índices de preços de grupos e classes de produtos e ao índice de preços geral, utilizando-se como ponderador a importância relativa de cada um desses bens na despesa do consumidor médio. Nos índices de preços calculados a partir de Dezembro de 1991, utilizou-se como estrutura da despesa a resultante do Inquérito ao Orçamento e Despesa das Famílias (IOF) realizado em 1988/1989. Dada a mudança de estrutura então operada no cálculo dos índices, no presente trabalho foi apenas considerada a evolução dos índices de preços desde Dezembro de 1991. Este período é, de resto, adequado aos objectivos pretendidos, que são os de determinar as causas da desinflação em Portugal desde o início de 1992.

A identificação dos comovimentos dos 75 índices pode ser conseguida recorrendo ao método das componentes principais.

Numa primeira fase, este método normaliza as evoluções dos diferentes índices, de forma a impedir que os índices com flutuações mais acentuadas possuam uma maior influência na determinação das componentes. De seguida, com base nas correlações detectadas entre estas variáveis normalizadas, este método fornece-nos um conjunto de vectores ortogonais que explicam a variância do conjunto das variáveis iniciais. Estas componentes são hierarquizadas, de acordo com a sua capacidade explicativa dessa variância. A primeira componente é, por isso, a que apresenta maior capacidade explicativa e corresponde à tendência de fundo, não observada, subjacente ao conjunto das variáveis analisadas. A primeira componente é semelhante à variável subjacente, não observada, obtida por outros métodos aplicados ao mesmo conjunto de variáveis, nomeadamente pelo Filtro de Kalman.

Para além da identificação e da hierarquização das componentes e da respectiva contribuição para a variância global do conjunto das variáveis iniciais, este método fornece-nos também a correlação entre as componentes e essas variáveis.

A aplicação deste método (ver quadro 1) mostra que a primeira componente justifica 59,4 por cento da variância das variações homólogas mensais dos 75 índices de preços no consumidor durante o período em análise. As três primeiras componentes explicam 77,9 por cento dessa variância e as cinco primeiras componentes 87,7 por cento, o que revela que o essencial dos comovimentos de um conjunto tão numeroso de índices pode ser resumido num número excepcionalmente reduzido de vectores independentes. A análise das correlações entre as componentes e as variáveis de base evidencia que apenas as três primeiras componentes apresentam uma correlação significativa com um número razoável das variáveis iniciais, pelo que a análise das restantes componentes não se enquadra nos objectivos do presente trabalho. Como adiante se verá, as três primeiras componentes são também suficientes para explicar cerca de 97 por cento da variância da variação homóloga do índice de preços geral (excluindo a habitação), reforçando a opção assumida.

#### QUADRO 1

VARIÁVEIS ANALISADAS (75): ÍNDICES DE PREÇOS NO CONSUMIDOR - VARIACÕES HOMÓLOGAS MENSAIS

##### PRINCIPAL COMPONENTS

COMPONENT	NAME	ITERATIONS	EIGENVALUE	CUMULATIVE R-SQUARED
1	P1	10	44.577	0.594
2	P2	58	7.872	0.699
3	P3	48	5.981	0.779
4	P4	102	3.928	0.831
5	P5	32	3.394	0.877

##### FACTOR LOADINGS

VARIÁVEIS	P1	P2	P3	P4	P5
Pão e Produtos de Padaria	0.833	0.425	0.034	0.140	0.103
Legumes	-0.445	-0.292	-0.101	0.313	-0.504
Peixe Fresco e Frigorificado	-0.090	0.486	-0.579	0.487	0.274
Pastelaria	0.969	0.097	-0.032	-0.056	-0.129
Alim.Preparado ou parte p/cons.casa	0.954	0.185	-0.074	0.142	-0.117

(Continua)

**QUADRO 1 (Continuação)**

**FACTOR LOADINGS**

<i>VARIÁVEIS</i>	P1	P2	P3	P4	P5
Aliment.consum.fora de casa	0.925	0.149	-0.118	-0.176	0.129
Serviços de confecção	0.826	0.223	-0.013	0.380	-0.075
Limpeza Vestuário em Lavandaria	0.966	0.181	-0.030	0.028	0.117
Reparação de Calçado	0.963	0.050	0.034	-0.045	0.098
Água	0.685	0.192	0.040	-0.523	-0.185
Electricidade	0.842	-0.244	-0.066	0.112	0.239
Serv.Rep.Conser.c Manut.B.Duráveis	0.891	-0.092	-0.035	0.079	-0.054
Medicamentos	0.959	0.159	-0.039	-0.018	0.033
Material p/Pensos	0.623	-0.597	0.024	0.005	0.289
Serv.Médicos e Paramédicos	0.921	-0.313	-0.138	0.020	-0.068
Meios Auxil.de Diagnóstico	0.855	-0.248	-0.308	0.159	-0.096
Carburantes e Lubrificantes	-0.429	-0.459	0.287	0.499	0.262
Outras Desp. de Manut. de veículos	0.958	0.074	0.038	0.046	0.174
Outras Despesas	0.853	-0.374	0.201	-0.036	-0.092
Transp.Púb. Terrestres	0.868	-0.229	0.009	0.086	0.117
Outros Trans.Públicos	0.071	0.345	-0.336	0.502	0.342
Comunicações	0.468	-0.093	0.158	0.332	0.391
Ensino	0.610	-0.629	-0.204	-0.263	0.214
Serviços de Reparação	0.670	-0.337	-0.132	0.073	-0.106
Espectáculos	-0.491	0.459	0.325	0.101	-0.324
Desp. e/Ut.Instal.Desp.Recreativas	0.361	-0.497	0.204	0.700	0.133
Taxas Radiof.,out.Serviços	0.952	-0.005	0.113	0.032	-0.065
Desp.compra Livros,jornais,etc.	0.873	0.399	0.109	-0.089	-0.076
Tabaco	0.748	0.146	0.221	-0.254	-0.176
Serviços	0.866	-0.372	-0.132	-0.021	-0.033
Hot.e outr.Serv.Alojamento	0.800	0.521	0.059	0.150	0.088
Viagens Turísticas	-0.542	0.040	-0.109	-0.150	-0.094
Jogos e Apostas	0.098	-0.613	0.602	0.038	0.117
Serviços Diversos	0.763	-0.441	-0.231	-0.117	-0.225
Cereais	0.554	-0.497	-0.185	0.296	-0.480
Farinhas	0.974	0.026	0.183	-0.084	-0.018
Massas Alimentícias	0.972	0.082	0.141	0.042	-0.077
Feculas e Amidos	-0.187	-0.463	-0.716	-0.095	0.261
Leguminosas Secas	-0.674	-0.317	-0.534	0.230	-0.131
Frutas	-0.501	0.133	-0.399	0.492	-0.454
Carnes, Miudezas,Salsich.Anim.Cap.	0.219	0.664	0.054	0.436	-0.178
Peixe Congelado	0.818	0.433	-0.290	0.116	0.039
Moluscos e Crustáceos	-0.132	-0.442	-0.693	-0.279	0.120
Peixe em Conserva	0.907	0.052	-0.114	0.116	-0.273
Peixe Seco ou em Salmoura	0.342	0.359	-0.798	0.039	0.038
Ovos	0.052	-0.117	0.732	0.477	-0.408
Leite,Prod.Lacteos e/exc.mant. e diet.lact.	0.902	0.311	0.053	-0.048	-0.187
Produtos Dietéticos	0.930	0.167	-0.023	-0.132	-0.147
Óleos e Gorduras	-0.321	0.533	-0.215	0.164	-0.602
Açúcar e Mel	0.826	-0.194	-0.145	0.257	-0.282
Geleia e Compotas	0.926	-0.192	-0.253	-0.073	-0.109
Bolachas e Biscoitos	0.921	0.191	-0.097	-0.213	0.005

QUADRO 1 (Continuação)

FACTOR LOADINGS

VARIÁVEIS	P1	P2	P3	P4	P5
					(Continua)
Confeitaria	0.890	-0.067	0.115	-0.149	-0.287
Cacau, café e chá	0.808	-0.057	-0.257	-0.394	-0.207
Diversos: Temperos, Condimentos, etc.	0.921	0.225	-0.010	-0.199	-0.154
Bebidas	-0.410	-0.416	-0.744	-0.081	-0.263
Vestuário de Homem	0.979	-0.098	-0.014	0.072	0.077
Vestuário de Senhora	0.946	0.108	-0.101	0.121	0.019
Vestuário p/Bébé e Criança	0.969	-0.160	-0.001	-0.025	-0.048
Acessórios de Vestuário	0.924	0.310	-0.042	0.074	0.066
Artigos p/Confecção de Vestuário	0.894	0.260	0.004	0.100	0.117
Retrosaria	0.892	0.216	-0.140	-0.235	0.022
Calçado p/Homem	0.942	0.058	-0.045	0.032	0.227
Calçado p/Senhora	0.863	0.423	-0.146	0.000	0.143
Calçado p/Criança	0.844	0.235	0.233	0.137	0.305
Gás	0.404	-0.237	0.238	0.091	0.247
Aquis. Bens Domést. Duráveis	0.922	0.140	0.208	0.046	0.150
Produtos Util. Corrente	0.925	0.224	0.186	-0.130	0.060
Aparelhos e Mat. Terapeutico	0.824	-0.284	0.312	-0.040	-0.277
Aquisição de Veículos	0.733	-0.548	0.320	-0.141	-0.098
Aquisição de Aparelhos Radio, TV, etc	0.779	-0.056	-0.517	0.160	-0.086
Outras Despesas do Fumador	0.661	-0.483	0.339	0.102	-0.346
Artigos não Duráveis	0.963	-0.198	-0.029	-0.104	0.023
Artigos Duráveis	0.790	-0.420	-0.299	0.248	-0.084
Outros Artigos	0.747	-0.383	-0.092	0.377	0.209

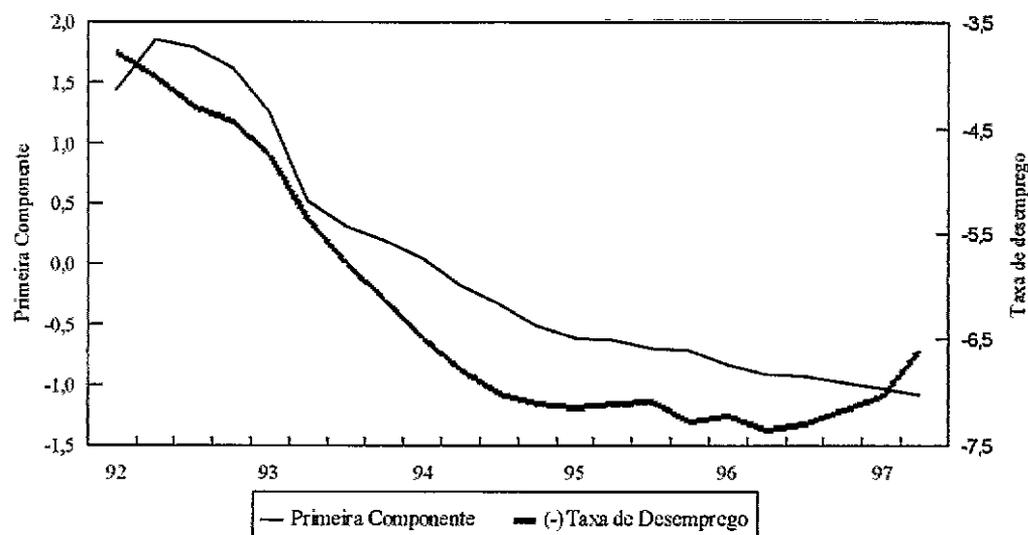
#### 4. A INFLAÇÃO SUBJACENTE

Para além de explicar 59,4 por cento da variância, a primeira componente apresenta uma correlação positiva muito significativa com a esmagadora maioria das 75 variáveis iniciais. Apenas os "legumes", o "peixe fresco e frigorificado", os "carburantes e lubrificantes", os "outros transportes públicos", os "espectáculos", as "viagens turísticas", os "jogos e apostas", as "féculas e amidos", as "leguminosas secas", as "frutas", as "carnes", os "moluscos e crustáceos", os "óleos e gorduras", os "ovos" e as "bebidas" apresentam uma correlação insignificante ou negativa com a primeira componente principal, sendo também reduzido o número de grupos de produtos em que a correlação é inferior a 50 por cento. Estes resultados comprovam que a primeira componente é uma variável não observada com um peso determinante na generalidade dos índices de preços elementares, sendo, assim, perfeitamente adequada para a descrição do comovimento subjacente à generalidade desses índices. Trata-se de uma variável que se identifica com a tendência de fundo da inflação e que, deste modo, pode ser interpretada como um indicador da inflação subjacente.

O perfil acentuadamente descendente evidenciado pela primeira componente principal demonstra que, ao longo deste período, a evolução dos preços esteve sujeita a choques contínuos de grande intensidade.

Entre o início de 1992 e o final de 1994, a primeira componente apresentou uma tendência inversa à da taxa de desemprego e bastante próxima da evolução nominal dos salários contratados (vide gráfico 1). A forte subida do desemprego e a correspondente, e não menos significativa, moderação salarial foram originadas pela recessão de 1993. No entanto, a primeira componente continuou em queda entre o final de 1994 e o final de 1996, período em que a taxa de desemprego e a evolução salarial apresentaram já uma relativa estabilidade.

GRÁFICO 1: PRIMEIRA COMPONENTE E TAXA DE DESEMPREGO



Assim, entre o final de 1994 e o final de 1996, as actualizações salariais terão desempenhado um papel de inércia na evolução da inflação. O facto de, durante este período, a primeira componente ter continuado a descer, ainda que de uma forma bastante mais moderada, evidencia a presença de outros choques para além dos gerados pela recessão de 1993. Assim, a inflação subjacente terá mantido um perfil descendente graças à melhoria da produtividade, sobretudo na produção de bens transaccionáveis, e à maior abertura do mercado nacional. Refira-se que, em média anual, entre o final de 1991 e de 1996, o PIB português registou um crescimento de 1,8 por cento, enquanto o emprego por conta de outrem diminuía 1,2 por cento e o número de horas trabalhadas descia 0,5 por cento. Embora o desemprego tenha baixado durante o primeiro semestre de 1997, as actualizações salariais conheceram uma evolução mais moderada que nos anos anteriores, voltando a impulsionar a descida da inflação.

Assim, a recessão de 1993, a melhoria da produtividade e a abertura dos mercados terão constituído as principais determinantes do processo de desinflação registado entre o início de 1992 e Maio de 1997. A recessão terá dado a contribuição mais decisiva. Recorde-se que a inflação subjacente, medida através da variação homóloga mensal, desceu de cerca de 13 por cento em Junho de 1992 para apenas 4,5 por cento em Dezembro de 1994 (quadro 2). Isto equivale a dizer que a recessão, ao fazer disparar o desemprego e ao impor uma drástica moderação salarial, foi, em última análise, a principal responsável pela descida significativa da inflação subjacente para níveis historicamente muito baixos.

A estimativa da inflação subjacente é feita, tradicionalmente, excluindo do cálculo do índice de preços geral os produtos alimentares não transformados e os produtos energéticos. Este procedimento tem sido correntemente utilizado pelo Banco de Portugal no cálculo da inflação subjacente. A utilização desta via para o cálculo da inflação subjacente baseia-se no pressuposto de que os preços destes produtos são particularmente afectados por choques “não económicos”, resultantes de condições climáticas ou de outros acontecimentos imprevistos, que perturbam a identificação da tendência de fundo da inflação. O método das componentes principais, ao estimar a tendência de fundo subjacente à generalidade dos índices de preços e ao identificar a sua relação com os diferentes índices, permite encontrar vias alternativas para esta estimação.

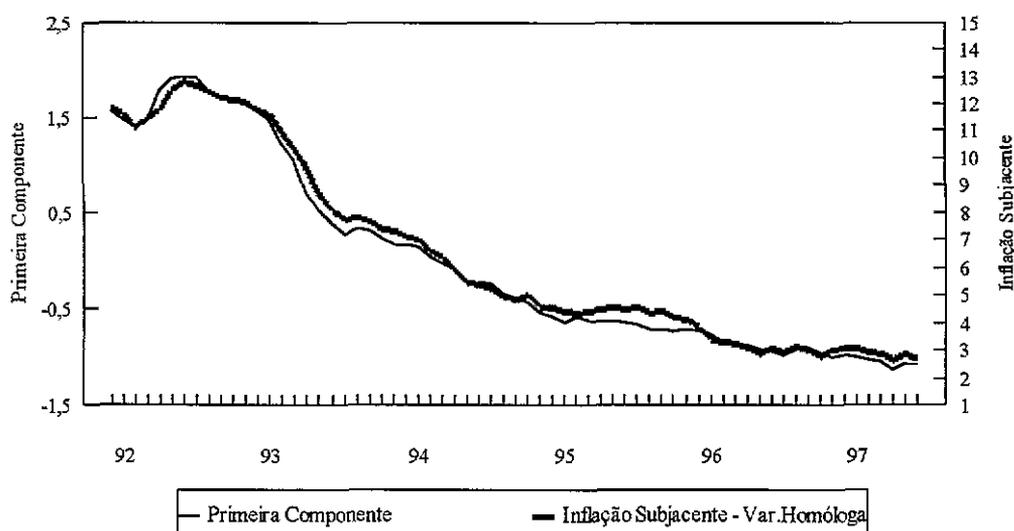
Uma primeira alternativa consistirá na construção de um índice resultante da agregação ponderada dos índices elementares que apresentam uma correlação positiva com a primeira componente principal igual ou superior a 50 por cento, isto é, aqueles índices em que a primeira componente justifica o essencial da sua evolução. A ponderação para cada grupo de produtos seria a mesma que é usada no cálculo do índice de preços no consumidor. A aplicação deste método excluiria índices englobados nas categorias de bens alimentares não transformados e de produtos energéticos mas também alguns outros, casos dos “outros transportes públicos”, das “comunicações”, dos “espectáculos”, da “utilização de instalações recreativas”, das “viagens turísticas” e dos “jogos e apostas”

Uma segunda alternativa consiste em usar a própria primeira componente como indicador da inflação subjacente, através de uma transformação simples, atribuindo-lhe a média e o desvio-padrão da estimativa obtida na primeira alternativa.

Qualquer das alternativas apresenta vantagens e desvantagens.

A primeira alternativa tem algumas vantagens, nomeadamente, a de permitir estimar um índice de preços e proporcionar, por isso, uma informação sobre a sua evolução complementar à da variação homóloga mensal. Tem como principal incon-

GRÁFICO 2: PRIMEIRA COMPONENTE E INFLAÇÃO SUBJACENTE



veniente “contaminar” a determinação da inflação subjacente com os comportamentos idiossincráticos dos diferentes grupos de produtos que entram na sua agregação, nomeadamente daqueles com maior peso na despesa dos consumidores.

A segunda alternativa utiliza apenas a informação relacionada com a tendência de fundo presente no conjunto das variáveis, incluindo nas menos correlacionadas com a primeira componente. Tendo em conta o conceito de inflação subjacente, esta alternativa parece ser mais adequada. No entanto, apresenta o inconveniente de reduzir a avaliação da inflação subjacente à variação homóloga mensal.

Uma alternativa de síntese poderia ser obtida, aplicando a metodologia das componentes principais aos próprios índices de preços dos 75 grupos de produtos e utilizando a primeira componente principal deste procedimento como um indicador do nível dos preços durante o primeiro ano de existência, ou seja, entre Dezembro de 1990 e Novembro de 1991. A quantificação da primeira componente neste período resultaria de uma regressão com o índice obtido na primeira alternativa. A partir de Dezembro de 1991, o índice assumiria os valores resultantes da aplicação da variação homóloga mensal estimada na segunda alternativa.

O confronto gráfico entre a evolução da primeira componente principal e a variação homóloga mensal do índice de preços obtido pela primeira alternativa confirma a grande proximidade entre as duas estimativas da inflação subjacente (vide gráfico 2). As variações homólogas mensais fornecidas pelos dois métodos são apresentadas no quadro 2.

QUADRO 2

INFLAÇÃO SUBJACENTE (VARIAÇÃO HOMÓLOGA)

DATA	AGREGAÇÃO DE ÍNDICES	ESTIMADA - PRIMEIRA COMPONENTE	DATA	AGREGAÇÃO DE ÍNDICES	ESTIMADA - PRIMEIRA COMPONENTE
	11.9	11.9			
92	11.7	11.6	94	7.0	7.0
	11.2	11.3		6.6	6.6
	11.5	11.5		6.3	6.4
	11.8	12.6		5.9	6.1
	12.6	13.0		5.6	5.7
	12.9	13.1		5.5	5.7
	12.8	13.1		5.3	5.7
	12.5	12.6		5.0	5.3
	12.3	12.4		4.9	5.1
	12.2	12.3		5.1	5.0
	12.1	12.2		4.7	4.6
	11.8	11.9		4.6	4.5
93	11.7	11.6	95	4.4	4.3
	11.2	10.7		4.3	4.4
	10.6	10.2		4.4	4.3
	9.8	8.9		4.5	4.3
	8.6	8.3		4.6	4.3
	8.0	7.8		4.5	4.3
	7.8	7.4		4.5	4.2
	7.9	7.7		4.3	4.0
	7.6	7.6		4.4	4.0
	7.4	7.3		4.2	4.0
	7.3	7.1		4.1	4.0
	7.1	7.0		3.8	4.0

(Continua)

QUADRO 2 (Continuação)

DATA	AGREGAÇÃO DE ÍNDICES	ESTIMADA - PRIMEIRA COMPONENTE	DATA	AGREGAÇÃO DE ÍNDICES	ESTIMADA - PRIMEIRA COMPONENTE
96	3.3	3.8	97	3.1	3.0
	3.3	3.5		2.9	3.0
	3.2	3.5		2.9	2.9
	3.0	3.4		2.6	2.6
	2.9	3.2		2.9	2.8
	3.1	3.2		-	-
	2.9	3.0		-	-
	3.1	3.3		-	-
	3.0	3.3		-	-
	2.8	3.2		-	-
	3.0	3.0		-	-
3.1	3.0	-	-		

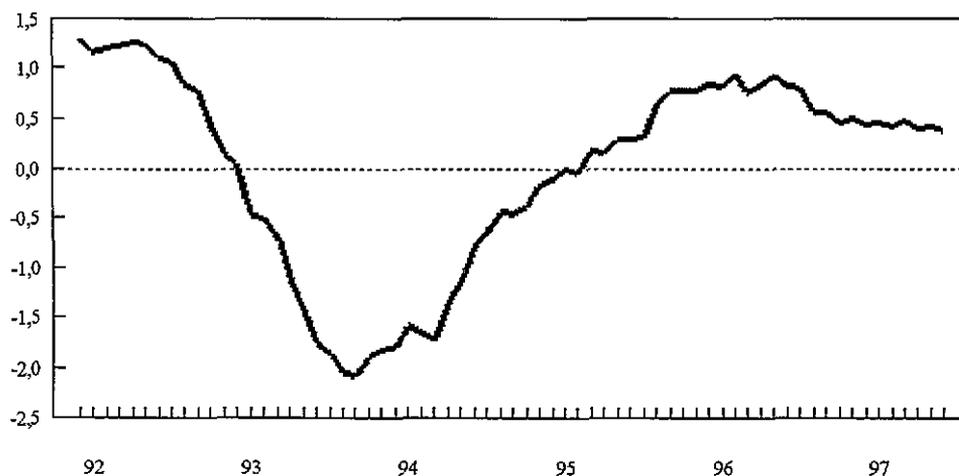
## 6. A INFLUÊNCIA DAS EVOLUÇÕES FORTEMENTE IRREGULARES

A segunda componente principal explica cerca de 10 por cento da variância do conjunto das 75 variáveis. O seu perfil de evolução e a consulta da matriz das correlações sugerem que esta variável poderá estar associada com as evoluções de alguns grupos de produtos que apresentaram flutuações irregulares muito intensas (gráfico 3).

De facto, esta segunda componente aparece correlacionada positivamente com os preços das "carnes" e dos "óleos e gorduras" e apresenta uma correlação negativa também significativa com os "cereais", as "féculas e amidos", os "crustáceos e moluscos" e as "bebidas", produtos cujos preços assumem, por vezes, comportamentos bastante irregulares, e com outros em que essa irregularidade está ligada a uma intervenção pública, casos do "Ensino" e dos "Jogos e Apostas".

O facto desta variável apresentar um perfil próximo da evolução do rendimento real das famílias poderá justificar a correlação positiva, embora relativamente fraca, com alguns bens duradouros.

GRÁFICO 3: SEGUNDA COMPONENTE

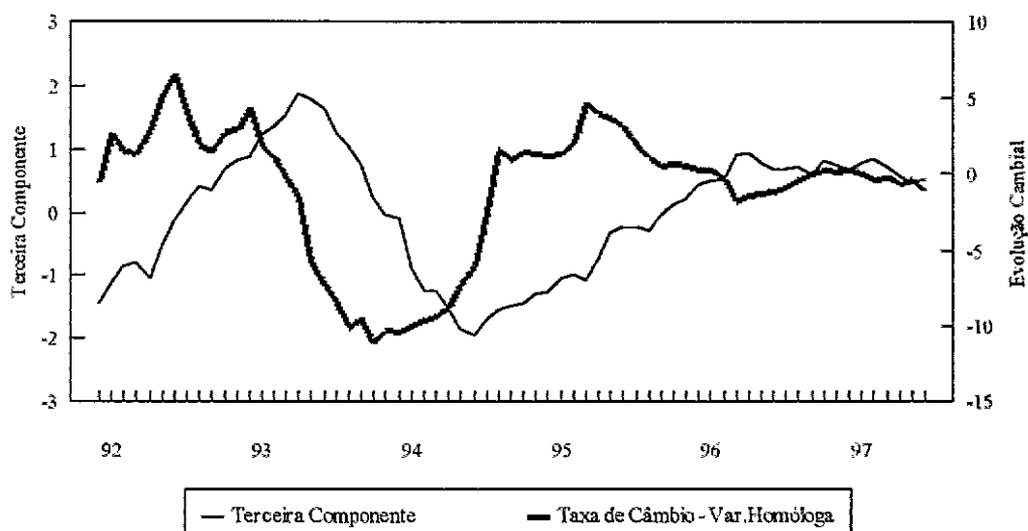


## 7. A INFLUÊNCIA DA EVOLUÇÃO CAMBIAL

A terceira componente evidencia também uma forte correlação positiva com os preços de produtos caracterizados por uma evolução irregular, casos dos "jogos e apostas" e dos "ovos" mas também uma correlação negativa significativa com alguns produtos em que a componente importada é importante e que, em alguns casos, se conjuga com evoluções irregulares, casos do "peixe seco ou em salmoura", dos "moluscos e crustáceos", das "féculas e amidos", das "bebidas" e da "aquisição de aparelhos de rádio e tv".

No caso das flutuações dos preços dos grupos de produtos com uma parcela importada significativa, estas poderiam, eventualmente, estar associadas com a evolução cambial do escudo. O sinal negativo desta correlação está de acordo com o esperado, ou seja, uma subida do escudo tem um efeito oposto sobre os preços dos bens importados. Assim, a terceira componente apresenta um perfil de evolução semelhante ao da variação homóloga da taxa de câmbio efectiva do escudo mas com pontos de viragem atrasados entre 7 e 13 meses (gráfico 4). Este "atraso" poderia resultar do desfaseamento entre o momento da desvalorização (valorização) do escudo e o do seu impacto nos preços de venda no mercado interno.

GRÁFICO 4: TERCEIRA COMPONENTE E EVOLUÇÃO CAMBIAL



Em resumo, as segunda e terceira componentes representariam oscilações das evoluções dos preços dos produtos, em torno da sua tendência de fundo simbolizada pela primeira componente principal, sendo essas oscilações geradas pela evolução cambial e pelo comportamento errático dos preços de alguns produtos.

As quarta e quintas componentes principais são ainda de mais difícil interpretação e a matriz das correlações sugere que estão essencialmente associadas a comportamentos idiossincráticos dos diferentes grupos de produtos durante o período em análise.

## 8. AS PRINCIPAIS DETERMINANTES DA EVOLUÇÃO DO ÍNDICE GERAL

As três primeiras componentes possuem uma correlação próxima de 97 por cento com a variação homóloga do índice de preços no consumidor geral (ver quadro 3), sendo esta correlação próxima de 93 por cento quando se considera apenas a primeira componente principal. Destes resultados conclui-se que a evolução da inflação efectiva entre 1992 e 1997 anos foi muito próxima da registada pela inflação subjacente. Uma fatia ainda significativa da variância restante está associada com a flutuação cambial e com evoluções irregulares de um pequeno grupo de produtos.

A parcela da variância do conjunto dos índices, que servem de base ao índice de preços no consumidor geral (sem habitação), não explicada pelas três primeiras componentes principais tem, no seu conjunto, uma fraca influência na evolução do índice geral.

### QUADRO 3

METHOD OF ESTIMATION = ORDINARY LEAST SQUARES

DEPENDENT VARIABLE: IPC (VARIACÃO HOMÓLOGA)

CURRENT SAMPLE: 1991:12 TO 1997:05

NUMBER OF OBSERVATIONS: 66

MEAN OF DEPENDENT VARIABLE = 5.4

STD. DEV. OF DEPENDENT VAR. = 2.245

SUM OF SQUARED RESIDUALS = 9.930

VARIANCE OF RESIDUALS = .1602

STD. ERROR OF REGRESSION = 0.400

R-SQUARED = 0.970

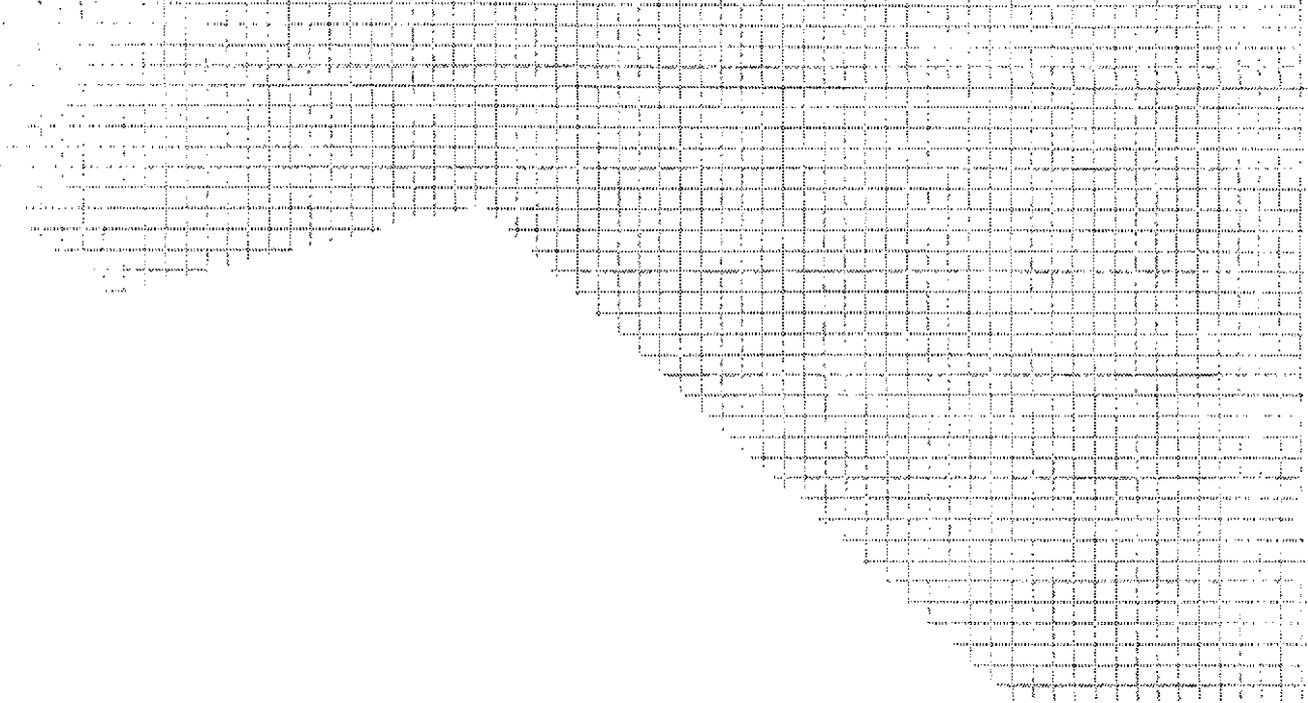
VARIABLE	ESTIMATED COEFFICIENT	STANDARD ERROR	T-STATISTIC
Constante	5.402	0.049	109.664
Primeira Componente	2.187	0.050	44.061
Segunda Componente	-0.132	0.050	-2.649
Terceira Componente	-0.295	0.050	-5.952

### REFERÊNCIAS BIBLIOGRÁFICAS

CHATFIELD, C. E A.J. COLLINS (1991), "Introduction to multivariate analysis", Chapman and Hall.

NASCIMENTO, T. (1990), "Indicadores de Inflação", *Boletim Trimestral* do Banco de Portugal, Dezembro de 1990.





# A INTERNET COMO MEIO ALTERNATIVO DE DIFUSÃO DE INFORMAÇÃO ESTATÍSTICA

**Autor:**  
Carlos Sebastião Afonso Dias



---

## A INTERNET COMO MEIO ALTERNATIVO DE DIFUSÃO DE INFORMAÇÃO ESTATÍSTICA

---

---

## THE USE OF INTERNET AS A MEAN TO DISSEMINATE STATISTICAL INFORMATION

---

Autor: Carlos Sebastião Afonso Dias  
Mestre em Estatística e Gestão de Informação pelo Instituto Superior de  
Estatística e Gestão de Informação da Universidade Nova de Lisboa  
e  
Coordenador de Equipa - Departamento de Sistemas de Informação e  
Informática  
Instituto Nacional de Estatística

### *SUMÁRIO:*

- Nos últimos anos, a questão da difusão da informação estatística oficial vem adquirindo uma extraordinária importância, dada a evolução verificada quer na procura de informação quer nas tecnologias de suporte à sua difusão.

O principal objectivo deste artigo, é o de analisar sob diversos aspectos as potencialidades da Internet, como meio alternativo de difusão de informação estatística.

Como resultado da investigação realizada nos sites dos Institutos Nacionais de Estatística de todo o mundo, apresenta-se um protótipo de um Sistema de Difusão na Internet para o INE.

### *PALAVRAS-CHAVE:*

- *Difusão, Informação, Estatística, Internet.*

### *SUMMARY:*

- In recent years, the issue of dissemination of official statistical information has gained enormous importance given the fact that both the demand for such information and the technological means by which it is supplied have changed dramatically.

The main purpose of this paper is to analyse the potentiality that lies in the use of Internet as a mean to disseminate official statistical information.

The prototype which is presented aims to describe a System of Information Dissemination, based on Internet, to the Portuguese National Statistics Institute (INE). Its development is the result of an investigation undertaken on the sites of the national statistical offices throughout the world.

### *KEY-WORDS:*

- *Dissemination, Information, Statistics, Internet.*



## 1. INTRODUÇÃO

A questão da difusão da informação estatística oficial assume particular importância no momento actual dada a evolução acentuada, quer da procura da informação quer das tecnologias de suporte à sua difusão, experimentadas nos últimos anos. Os Institutos Nacionais de Estatística (INE's), dispendo de um conjunto volumoso de informação estatística sobre as várias áreas sócio-económicas do país, têm forçosamente de acompanhar esta evolução. Daí a necessidade de rever a adequação dos meios e processos utilizados na difusão com o objectivo de se poder tirar partido das inovações tecnológicas, optando-se por alternativas que se oferecem nesse domínio.

A criação da Internet, rede dotada de enorme potencial informativo, e a sua cada vez maior divulgação no mundo inteiro, permite alargar o âmbito do sistema de informação tradicional de forma a ser utilizada, a custos reduzidos, pelo grande público. Ao longo deste artigo, analisam-se as potencialidades da Internet como meio alternativo de difusão de informação e em particular de informação estatística.

Ao analisar o crescimento da Internet verificado em todo o Mundo, apesar de ser uma tecnologia recente e recorrendo às experiências de outros INE's, a evolução aponta no sentido de que cada vez mais utilizadores, também em Portugal, procurarão informação estatística na Internet. Adoptando o INE uma política de difusão orientada para o utilizador, prever as suas necessidades constitui um dos seus maiores desafios. Nesse pressuposto apresenta-se um protótipo de um sistema de difusão de informação estatística que permita dar continuidade ao já existente e evoluir no sentido de que a Internet funcione como um canal de comunicação directo entre o INE e os utilizadores da informação estatística.

## 2. DIFUSÃO DE INFORMAÇÃO ESTATÍSTICA

Será concerteza um lugar-comum afirmar que a informação é cada vez mais considerada como um elemento indispensável à tomada de decisões a todos os níveis. A informação estatística, em particular, podendo ser utilizada de formas diversificadas desde Análises Económicas, Análises de Mercado, Análises Financeiras, Análises de Pessoal, constituirá sem dúvida uma das bases essenciais para a gestão moderna e assumirá, cada vez mais, uma importância estratégica no desenvolvimento económico e social das sociedades.

A informação estatística é produzida fundamentalmente nos organismos oficiais e será a essa informação que nos refriremos ao longo deste artigo. Os INE's, que sendo detentores de vastos repositórios de informação estatística sobre as várias áreas sócio-económicas do país, têm de encontrar formas de a difundir junto dos seus utilizadores.

Em primeiro lugar, é necessário distinguir entre o conceito de difusão de informação e o de disponibilização que lhe está associado. A disponibilidade decorre de uma atitude passiva, isto é, um determinado produto é dado como pronto a utilizar e fica disponível, a aguardar que os utilizadores o solicitem. A difusão, pelo contrário, corresponde a uma atitude activa, pronto o produto a utilizar, o produtor toma a iniciativa de o difundir entre os seus públicos. Constitui actualmente um dos principais desafios dos INE's a adopção de estratégias de difusão orientadas para as necessidades dos utilizadores.

Os INE's difundem a informação sobre diversas formas das quais podemos destacar as publicações, os produtos electrónicos e os serviços.

Actualmente, as **Publicações** constituem ainda o principal meio de difusão de informação estatística. Ao longo dos anos foram evoluindo em termos de conteúdo, incluindo gráficos, comentários e análises a acompanhar as tradicionais tabelas ou quadros estatísticos, facilitando deste modo a sua compreensão.

Tratando-se de meios relativamente incómodos e dispendiosos e respondendo à solicitação dos utilizadores que pretendiam a informação noutros suportes, os difusores disponibilizam a informação estatística em **produtos electrónicos**, acompanhando as evoluções tecnológicas: bandas magnéticas, disquetes, tapes e, recentemente, o CD-ROM (*Compact Disc Read Only Memory*).

Consequência ainda da rapidez da evolução tecnológica, outros serviços foram aparecendo disponíveis ao utilizador como resultado da adesão dos institutos às novas ferramentas de difusão. Podemos referir os **serviços on-line** como o Videotexto, o acesso a bases de dados on-line, e, mais recentemente, a Internet. Este tipo de serviços permitem ao utilizador aceder à informação a partir de casa ou do local de trabalho ininterruptamente

---

### **3. ESTATÍSTICAS NA INTERNET**

---

#### **3.1. VANTAGENS E LIMITAÇÕES DA INTERNET**

---

Actualmente, as novas Tecnologias colocam novos desafios aos difusores de informação estatística, que poderão constituir alternativas aos meios utilizados. A **Internet**, a maior rede mundial de computadores, que interliga milhares de redes locais constituídas por computadores dos mais variados tipos, apresenta vantagens significativas em relação a outros meios de difusão:

- A capacidade multimédia do WWW (*World Wide Web*) permite que este possa ser visto como um meio privilegiado de **promoção**, divulgando os catálogos de produtos e serviços (sem incorrer em despesas como a impressão gráfica, publicidade e portes de correio) como se tratasse de uma "montra virtual" acessível de qualquer parte do mundo. Além dos catálogos de produtos, podem incluir-se formulários de encomenda, possibilitando aos potenciais clientes o seu preenchimento e envio.
- Possibilidade de **melhorar o serviço aos clientes** através do correio electrónico permitindo responder a perguntas, receber comentários e sugestões. A inclusão no WWW de páginas com as perguntas mais frequentes e as suas respostas, para apoio a dúvidas simples e rotineiras é certamente mais eficiente e barato do que a utilização de pessoal de apoio, libertando esse mesmo pessoal para outras tarefas.
- A constituição de uma lista de endereços (*Mailing List*) dos subscritores dos seus serviços facilita a divulgação de novos produtos e serviços.
- Permitir **acessos geograficamente dispersos**. Populações do interior e de zonas afastadas das fontes de informação passam a ter acesso de forma igual à informação, eliminando um importante factor de atraso no desenvolvimento.

- A presença de um servidor na Internet, permite o acesso rápido e barato à informação ininterruptamente 24 horas por dia e 7 dias por semana.
- A informação no WWW pode aparecer em **formatos diversos**, legíveis por qualquer cliente: texto, gráficos, tabelas e imagens.
- Na Internet é possível disponibilizar o acesso a bases de dados de informação estatística permitindo não só a consulta mas também a exportação dos dados para o computador cliente (*download*) em vários tipos de formatos.
- Podem coexistir na Internet bases de dados de informação estatística, Sistemas de Informação Geográfica (SIG), aplicações multimédia e todo um conjunto de outros serviços.

Existem contudo algumas limitações na utilização da Internet, de entre as quais podemos destacar:

- **Limitações a nível de rede** - Não estando as redes preparadas para o crescimento exponencial do número de utilizadores que se tem verificado nos últimos dois anos, o excesso de tráfego de informação provoca por vezes afunilamentos que implicam tempos de espera consideráveis na recepção da informação.
- **Problemas de Segurança** - Os *hackers*, especialistas em crimes informáticos espreitam oportunidades de aceder indevidamente aos sistemas das organizações ou particulares, podendo roubar, destruir ou alterar informação valiosa, o que poderá causar prejuízos irreparáveis.

A apresentação dos dados no WWW podem realizar-se com dois tipos de páginas: estáticas e dinâmicas.

A maior parte da informação estatística disponível no WWW quer sob a forma de tabelas, gráficos ou texto são **páginas estáticas**. Cada página é preparada manualmente, ou usando ferramentas de edição de documentos, em linguagem HTML (*HyperText Markup Language*) e armazenada no servidor de Web, onde estão acessíveis.

As páginas geradas no momento em que o utilizador as acede através de programas específicos (*gateways*) designam-se por **páginas dinâmicas**. Essa possibilidade pode permitir fazer interrogações (*queries*) em tempo real a uma base de dados e visualizar a resposta em páginas HTML. Os programas CGI (*Common Gateway Interface*) permitem fazer o interface entre o servidor WWW e os repositórios de dados (Sistemas de Gestão de Bases de Dados, sistemas de ficheiros e SIG) Os fornecedores de informação podem construir programas CGI que processem os pedidos obtidos dos utilizadores, aceder a uma base de dados e converter a resposta no formato HTML.

---

### 3.2. QUE INFORMAÇÃO ESTATÍSTICA NA INTERNET?

---

Os INE's sempre acompanharam de perto a evolução das Tecnologias da Informação e não constituiu surpresa que em 1994, alguns INE's, como o canadiano, o Bureau of Census (EUA) e o holandês, acompanhando o crescimento da Internet, fossem pioneiros na sua utilização como meio de divulgação de informação estatística. Actualmente os institutos de estatística de cerca de três dezenas de países já disponibilizam informação na Internet.

Poderemos catalogar a informação disponível nos servidores dos INE's em quatro categorias de informação: apresentação, promocional, estatística e adicional.

A **informação de apresentação** inclui a informação de carácter geral normalmente dando a conhecer a natureza do Sistema Estatístico Nacional (SEN), Órgãos do SEN, Ordenamento jurídico/legislação do SEN/INE, enquadramento do INE no SEN, Organização do INE e Locais de atendimento.

A **informação promocional** inclui informação sobre:

- Os produtos de difusão (os catálogos de publicações, a informação disponível em disquetes, tapes, CD-ROM, Videotexto, on-line)
- Comunicados de Imprensa (*Press-releases*), Folhas de Informação Rápida, novidades
- Perguntas mais frequentes (*Frequently Asked Questions*), formulação de perguntas (*Helpline*) e comentários através de correio electrónico.

A **informação estatística** corresponde aos dados estatísticos propriamente ditos:

- Estatísticas básicas - apresentadas sob a forma de tabelas, gráficos, figuras ou cartogramas
- Acesso a Base de dados de Microdados, Macrodados e Meta-informação. Pode incluir SIG.

Na **informação adicional** podemos incluir informação relacionada com as estatísticas:

- Ligação a outros servidores de outros institutos, instituições, universidades, escolas ou grupos de interesse no campo da Estatística
- Eventos/Acontecimentos como *meetings*, feiras, seminários, *workshops*, subordinados a temas relacionados com a Estatística
- Artigos ou documentação técnica sobre tratamento e sobre informação estatística.

Em relação à informação disponível nota-se que a maior parte tem sempre versões bilingues: inglês e a língua de origem. Nalguns casos, somente a informação promocional vem em inglês.

---

#### 4. PROTÓTIPO DE UM SISTEMA DE DIFUSÃO DO INE NA INTERNET

---

---

##### 4.1. RETROSPECTIVA

---

Na comemoração dos seus 60 anos, em 18 de Maio de 1995, o INE disponibiliza as suas primeiras páginas na Internet. Inicialmente alojada num servidor do Instituto Superior Técnico (IST), a informação disponível referia-se essencialmente à apresentação do Sistema Estatístico Nacional e do INE, da sua estrutura organizativa e locais de atendimento.

Após várias mudanças, a partir de Setembro de 1996, a informação é colocada num servidor próprio e atribuído um **domínio próprio** ine.pt.

Em Novembro de 1996 existiam no site WWW do INE (Fig. 1) cerca de 200 páginas de informação possíveis de consultar de qualquer parte do mundo. Além da informação disponível existe um serviço de correio electrónico (*mail*).

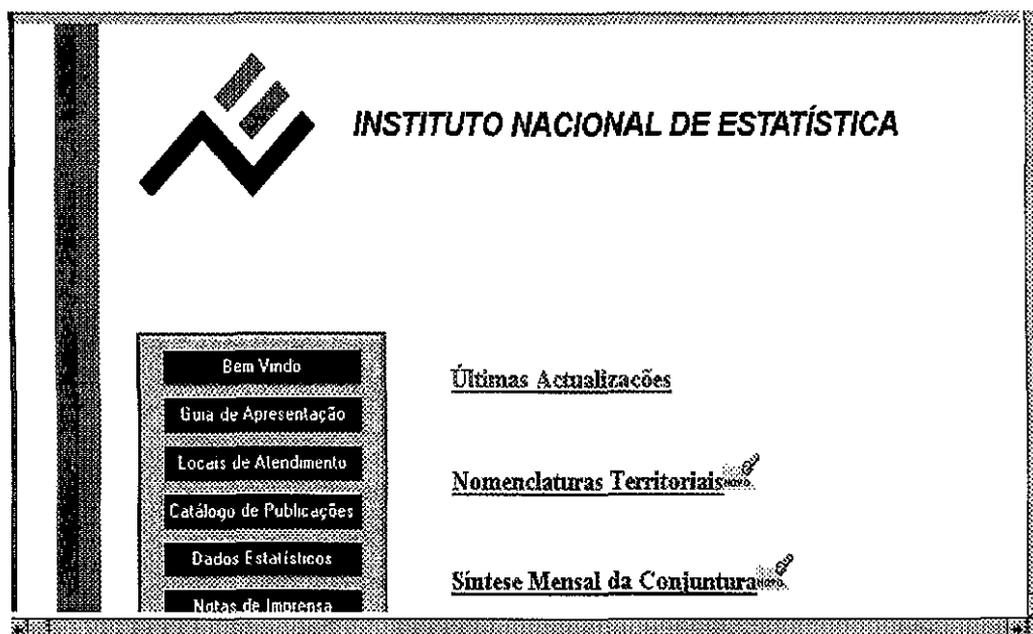


Fig. 1 - Home-page do INE (Nov. 1996) - <http://www.ine.pt>

Analisando as estatísticas de utilização do WWW do INE verificou-se uma média de 30 acessos/dia nos meses de Maio e Junho de 1996. Cerca de 95% dos utilizadores foram entidades portuguesas. O facto de não existir uma versão em Inglês, desencoraja as consultas dos utilizadores estrangeiros.

Não existem estatísticas fidedignas do número de utilizadores da Internet em Portugal. As estimativas, apresentadas em artigos de revistas da especialidade, indicam que o nosso país tem potencialidades para até ao ano 2000 ter cerca de 1 milhão de utilizadores na Internet. Ao analisar o crescimento da Internet verificado em todo o Mundo, apesar de ser uma tecnologia recente, e atendendo às experiências de outros institutos nacionais de estatísticas, a evolução aponta no sentido de que cada vez mais utilizadores, também em Portugal, procurarão informação estatística na Internet.

Adoptando o INE uma política de difusão orientada para o utilizador, prever as suas necessidades constitui um dos seus maiores desafios. Nessa perspectiva, apresenta-se um protótipo de um sistema de difusão de informação estatística que permita dar continuidade ao já existente e evoluir no sentido de que a Internet funcione como um canal de comunicação directo entre o INE e os utilizadores da informação estatística.

O objectivo não será abandonar os outros meios de difusão já existentes, em papel, disquetes, CD-ROM e serviço Videotexto, mas sim encontrar alternativas numa política de difusão integrada.

#### 4.2. OBJECTIVOS

Os objectivos principais do sistema, podem ser agrupados de acordo com dois pontos de vista distintos.

Numa **perspectiva da difusão** da informação pretende-se:

- Disponibilizar aos utilizadores de uma forma orientada a consulta on-line de informação estatística
- Flexibilidade no manuseamento da informação
- Facilidade e rapidez de actualização da informação
- Harmonização e coerência da informação
- Possibilidade de tarifação da informação a disponibilizar
- Permitir troca de mensagens com os utilizadores
- Segurança e restrições de acessos à informação.

Numa **perspectiva técnica** pretende-se criar um sistema informático que respeite as seguintes características:

- Modular, adaptável à cada vez mais rápida obsolescência tecnológica
- Privilegie a utilização de produtos standard, aproveitando de cada um deles as suas melhores características
- Com uma componente Interactiva que utilize um GUI (*Graphical User Interface*), de modo a facilitar o diálogo dos utilizadores com o Sistema
- Manutenção de uma Base de Dados com a informação a disponibilizar
- Capacidade de ligação remota dos utilizadores a um ou mais servidores do INE.

Este sistema de difusão destinar-se-á a todo o tipo de utilizadores de informação estatística. Apesar de não existirem estatísticas dos utilizadores do INE que têm acesso à Internet é possível identificar os mercados potenciais. A decisão do actual Ministro da Investigação Científica de, até ao final do ano, possibilitar o acesso de todas as Universidades e escolas à Internet, antevê um grande número de utilizadores de informação estatística para investigação e estudo. A Internet pode proporcionar uma forma rápida e eficaz de dar resposta ao aumento acentuado da procura de informação estatística por parte das empresas, das instituições governamentais e do público em geral, não só ao nível nacional como internacional.

A implementação do projecto deverá ser faseada no tempo:

Numa **primeira fase** preconiza-se a colocação de todas as publicações que o INE disponibiliza em suporte papel durante o ano em páginas de hipertexto pré-definidas na perspectiva dos quadros publicados.

Na **segunda fase** prevê-se a possibilidade de aceder a bases de dados de séries estatísticas. Com recurso ao CGI como interface entre o servidor WWW e o motor de Base de dados, serão implementadas funcionalidades de pesquisa e geração dinâmica de páginas sob a forma de tabelas, gráficos e cartogramas. A informação estará fisicamente distribuída pelos servidores das Direcções Regionais do INE.

Na óptica do utilizador, as **funcionalidades dos Serviços** a disponibilizar apresentarão as seguintes características:

- Possibilidade de acesso remoto dos utilizadores a um ou vários servidores do INE
- Permitir a troca de mensagens entre os utilizadores e o INE
- Permitir interrogar a Base de Dados através de pesquisas pré-definidas

- Consultar e importar a informação estatística disponível
- Permitir publicitar toda a informação a comercializar pelo INE.

#### 4.3. FASE 1

##### INFORMAÇÃO A DISPONIBILIZAR

Numa primeira fase a informação a disponibilizar incluirá, toda a informação estatística publicada (ou a publicar) durante o ano. A forma de apresentação será uma imagem dos quadros estatísticos publicados em papel, estando também acessível a meta-informação associada. Por meta-informação entende-se “informação sobre informação”, isto é, descrição, metodologia e conceitos associados às variáveis estatísticas, que possibilitem ao utilizador falar a mesma linguagem do estaticista.

##### ESTRUTURAÇÃO DA INFORMAÇÃO

A informação a disponibilizar na primeira fase estará organizada à semelhança dos Anuários Estatísticos publicados pelo INE, isto é, hierarquicamente agrupada por Temas/Áreas Estatísticas e dentro destes por Sub-temas/Sub-áreas que incluirão os quadros estatísticos disponíveis. A navegação será direccionada para a obtenção de um determinado quadro estatístico. Os quadros estatísticos são previamente preparados e desenhados em páginas estáticas em formato tabela.

##### ARQUITECTURA FÍSICA

A solução preconizada para este sistema assenta, na arquitectura cliente-servidor, constituindo a Rede Internet o meio de ligação entre o cliente e o servidor.

O INE dispõe actualmente de 4 Direcções Regionais: Norte (Porto), Centro (Coimbra), Lisboa e Vale do Tejo (Lisboa) e Alentejo (Évora). Na primeira fase as páginas de informação pré-definida ficam alojadas no equipamento de uma Direcção Regional, que já disponha de ligação à Internet.

Por segurança, este equipamento encontra-se isolado da rede interna do INE. As restantes Direcções Regionais dispõem de acessos individuais para aceder à Internet através de PCs também isolados da rede do INE

##### INTERFACE COM O UTILIZADOR

O acesso à informação pretendida pode realizar-se de duas formas:

- Por navegação através da hierarquia seleccionando sucessivamente Tema, Sub-tema e Quadro
- Pesquisa por palavra-chave

Na **pesquisa por palavra-chave**, o utilizador digita a(s) palavra(s)-chave em formulário apropriado. O motor de pesquisa do sistema devolve uma lista dos títulos dos quadros estatísticos em que encontrou a(s) palavra(s) indicadas. O resultado aparece sob a forma de *links* a que o utilizador pode aceder.

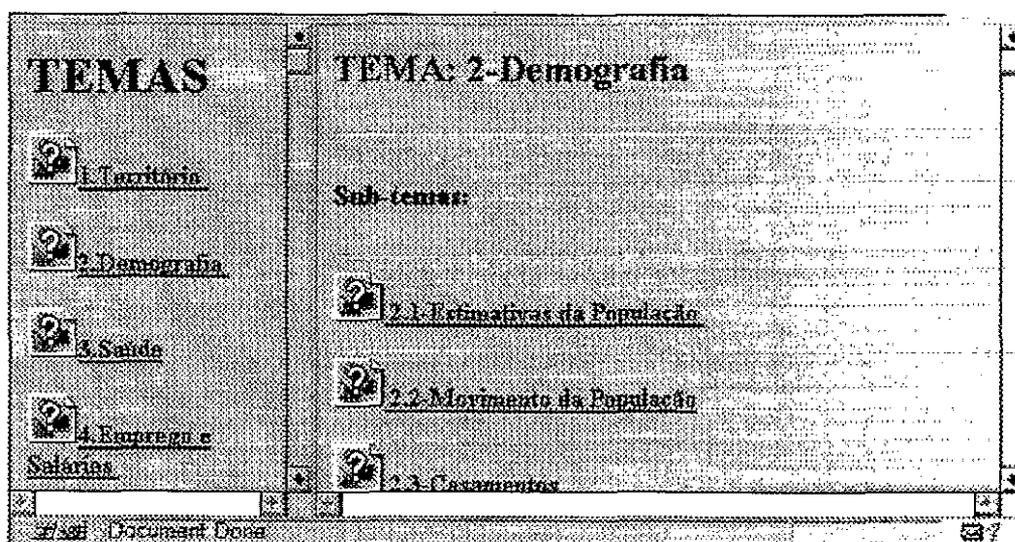


Fig. 2 - Selecção do Tema e Sub-tema Estatístico

A Selecção da Informação faz-se por etapas:

- A selecção da informação inicia-se com a escolha de um Tema ou Área Estatística.
- Após a selecção do Tema o utilizador escolhe o Sub-tema (Fig. 2).
- Após a selecção do Sub-tema pretendido apresenta-se a página de selecção dos Quadros disponíveis. A apresentação do quadro é uma imagem da Publicação editada (Fig. 3)

Estimativa da População Residente segundo Grupos Etários e Sexo em 1994

#020100124.xls

CONCELHOS	Grupos Etários								
	0-14		15-24		25-64		Mais de 64		
	HM	H	HM	H	HM	H	HM	H	
Continente, Açores e Madeira	1 783	912	1 634	826	5 062	2 444	1 431	590	120
Região Norte	693	357	626	316	1 766	853	426	173	140
Mitinho-Lima	45 110	23 110	42 000	20 870	118 600	53 870	43 840	17 360	

Excel Document Done

Fig. 3 - Página de apresentação do Quadro Estatístico seleccionado

Será também permitida a importação das páginas (*download*), em formatos standards (p.e. *Microsoft Excel*), que possibilitam a sua reutilização noutras aplicações. A Meta-informação aparece sempre associada a cada um dos Temas, Sub-temas e Quadros.

#### INFORMAÇÃO A DISPONIBILIZAR

Na fase 2 permitir-se-à o acesso a séries estatísticas de informação agregada, apresentando as seguintes características:

- Desagregação geográfica, pelo menos até ao limite de concelho.
- Na preparação das séries, incluir toda a informação estatística publicada ou não, pelo INE, desde 1990 ( para anos anteriores, se for possível)
- Formas de apresentação: quadros estatísticos, gráficos e cartogramas
- Incluir Meta-informação associada

#### ESTRUTURAÇÃO DA INFORMAÇÃO

A segunda fase reveste-se de alguma complexidade. Se na primeira fase os trabalhos de preparação passavam pela transformação dos formatos dos quadros estatísticos já existentes para tabelas HTML, nesta fase é necessário preparar bases de dados de séries estatísticas.

Nesta fase, ainda que o resultado final da selecção também dê origem a uma tabela, é o utilizador que escolhe o seu conteúdo com base na informação disponível e define a sua apresentação. A procura da informação será feita em função da variável estatística enquanto na fase 1 era em função do quadro.

A informação encontra-se agrupada por Temas/Áreas Estatísticas e dentro destes por Sub-temas/Sub-áreas. Hierarquicamente dependentes existem as Variáveis, Cruzamentos e Categorias.

Pode considerar-se a Variável como um item que pode apresentar um valor e a que corresponde a valência de "campo". Exemplos: População residente, Indivíduos com mais de 10 anos, Índice de preços no consumidor.

Para a mesma variável podem definir-se um ou mais cruzamentos com outras variáveis. Cada uma das variáveis cruzadas tem um conjunto de Categorias.

Por exemplo, para a variável População residente podemos ter os seguintes cruzamentos:

1. População residente por sexo
2. População residente por grupo etário e por estado civil

O Cruzamento 1 só cruza População residente com uma variável (sexo) que tem duas Categorias (Masculino e Feminino). O Cruzamento 2 cruza População residente com 2 outras variáveis (grupo etário e estado civil) em que cada uma delas pode ter várias Categorias.

Existem Variáveis que não têm cruzamentos na sua definição, o que ocorre, na maior parte das vezes, com variáveis que correspondem a índices.

Após a definição pelo utilizador da Variável e Cruzamento a pesquisar são identificados os Níveis Geográficos sobre os quais existe a informação pretendida.

O Nível Geográfico corresponde aos níveis dentro da hierarquia das Áreas Geográficas. Temos duas estruturas hierárquicas possíveis:

- Divisão Administrativa - corresponde à divisão administrativa do País em Continente, Açores e Madeira. Por sua vez essas Zonas estão divididas em Distritos, estes em Concelhos que por sua vez se sub-dividem em Freguesias.
- Nomenclatura das Unidades Territoriais para fins estatísticos (NUTS) - Divisão do território nacional em regiões criada no âmbito da CEE (Decreto-Lei Nº 46/89 de 15 de Fevereiro). A estrutura hierárquica é Nuts I, Nuts II, Nuts III e Concelho. Os concelhos dentro das NUTS, em termos de área física correspondem aos da Divisão Administrativa, a sua codificação é no entanto diferente.

Cada nível geográfico tem uma ou mais unidades geográficas, que poderão ser seleccionadas. Cada unidade corresponde a um "registo". Exemplos: Freguesia de Alvalade, Concelho de Oeiras ou Distrito de Portalegre.

Estão ainda disponíveis os períodos a que a informação se refere: Anual, Semestral, Trimestral ou Mensal. Cada período temporal tem uma ou várias unidades temporais que podem ser seleccionadas. Por exemplo para uma determinada variável seleccionada anualmente pode existir informação referente aos anos 1991, 1992 e seguintes.

Dado que muitas vezes o utilizador não conhece os conceitos e metodologias utilizadas no INE é necessário disponibilizar meta-informação sobre os Temas, Sub-temas, Variáveis, Cruzamentos e Categorias.

#### ARQUITECTURA FÍSICA

Na segunda fase possibilita-se o acesso a bases de dados de informação, que estarão distribuídas por uma ou várias Direcções Regionais por forma a permitir maior rapidez de acessos.

A página de entrada (*Home-page*) para aceder à informação disponível é a mesma para todos os utilizadores. Após a escolha do Tema Estatístico, de forma transparente ao utilizador, os procedimentos seguintes serão realizados no servidor que contém a informação do item seleccionado. A solução técnica para disponibilizar as bases de dados na Internet incluiu a utilização de um CGI e um interface especial a desenvolver. De cada vez que o utilizador requisita dados, o interface transforma o pedido numa *query* para a Base de Dados e gera na resposta uma página em HTML.

#### INTERFACE COM O UTILIZADOR

Nesta fase, uma pesquisa encontra-se dividida em duas etapas: selecção do conteúdo da informação e definição das formas de apresentação.

##### Seleção da Informação

Da mesma forma que na fase I pode procurar-se a informação por navegação ou por palavra-chave. Se a pesquisa por palavra-chave se processa da mesma forma, a navegação tem uma sequência diferente. Nesta fase, pode caracterizar-se a selecção do conteúdo da informação em vários passos:

Definição da Variável e Cruzamento a inquirir que inclui:

- Seleção de Tema Estatístico
- Escolha do Sub-tema
- Escolha da Variável

- Selecção de Cruzamento

Selecção do Tipo Temporal (Mensal, Anual,...) e do Tipo de Área Espacial (Distritos, Concelhos, Freguesias, Nuts,...)

O utilizador vai escolher dentro de cada lista as Categorias ou Unidades que pretende que apareçam no quadro final. Após a selecção o sistema internamente gerou um conjunto de parâmetros com os quais efectuará uma *query* á Base de Dados. Os programas CGI devolvem ao utilizador o resultado em páginas HTML sobre a forma de tabela.

**População residente por Grupo Etário, Estado Civil e Sexo**

Escolha as opções para a Tabela

	CONTEUDO	APRESENTAÇÃO
Unidades Geográficas (7)	Norte Centro LVT Alentejo	Na linha, 1a. quebra <input type="checkbox"/> Por e.o. app.
Grupos Etários (18)	0 a 4 anos 5 a 9 anos 10 a 14 anos 15 a 17 anos	Na linha, 2a. quebra
Estado Civil (6)	Solteiro Casado c/ registo Casado s/ registo Viúvo	Na coluna, 1a. quebra
Sexo (2)	Feminino Masculino	Na coluna, 2a. quebra
Ano (2)	1981 1991	

Fig. 4 - Definição da apresentação da Informação

#### Apresentação da Informação

Na definição da apresentação da informação o utilizador pode optar por tabela, gráfico ou cartograma.

O formato do tipo tabela, apresenta um quadro de dados estatísticos semelhante ao resultado da fase I (Fig. 3), só que gerado dinamicamente a partir das selecções efectuadas. É possível ao utilizador definir as localizações, em linha ou em coluna, das variáveis cruzadas (e respectivas Categorias), das unidades geográficas e temporais e da ordem das quebras

Também será possível apresentar a informação na forma de gráfico e cartograma.

É sempre permitido ao utilizador fazer a importação para o seu próprio equipamento (*download*) da informação visualizada em formatos PDF (*Portable Document Format*), para possibilitar a sua posterior impressão com qualidade, XLS (*Microsoft Excel*), para futura utilização em folhas de cálculo e GIF (*Graphics Interchange Format*), correspondente às imagens dos mapas.

---

#### 4.5. PERSPECTIVAS DE EVOLUÇÃO DO SISTEMA

---

O sistema apresentado pode evoluir em termos do conteúdo de informação e das funcionalidades permitidas.

##### CONTEÚDO DA INFORMAÇÃO

Uma evolução previsível, a curto prazo, em termos do conteúdo da informação será possibilitar uma desagregação da informação a um nível mais detalhado. Em termos das unidades geográficas por exemplo, permitir obter informação ao nível de freguesia, secção e subsecção estatística. É evidente que qualquer que seja o nível de detalhe a atingir devem existir mecanismos que garantam a aplicação do princípio do segredo estatístico.

Outra possibilidade em termos evolutivos, a longo prazo, seria proporcionar a utilizadores privilegiados o acesso, através da Internet, aos micro-dados permitindo-lhes realizar as suas próprias análises e apuramentos. Esta expansão deverá ser limitada e a realizar mediante protocolos, à semelhança do que se verifica em outros institutos como o *Statistics Canada*, com Universidades e Centros de Investigação. A informação a circular entre o INE e as instituições deve estar encriptada de forma a impedir a sua utilização por indivíduos não autorizados.

##### FUNCIONALIDADES

A tradução de toda a informação para inglês poderá permitir que um maior número de utilizadores, a nível internacional acedam ao sistema.

A possibilidade de construção dinâmica de quadros a pedido, permitindo ao utilizador a definição dos seus próprios cruzamentos, não o condicionando a escolhas pré-definidas. O apuramento dos quadros seria realizado dinamicamente no momento do pedido, com recurso aos micro-dados. Antes da apresentação dos resultados ao utilizador a tabela seria submetida a um filtro que tornaria ilegível a informação que não respeitasse o princípio do segredo estatístico. Para a implementação deste sistema será necessário desenvolver (uma vez que actualmente não existe) um filtro suficientemente robusto com essas características. Outro problema adicional que o recurso directo aos micro-dados pode mostrar algumas incoerências da informação, não detectáveis em informação agregada.

Outro aspecto importante em termos evolutivos enquadra-se no âmbito da comercialização. Com o evoluir da tecnologia e da possibilidade de transacções comerciais seguras através da Internet, pode realizar-se a taxação on-line, isto é, o utilizador pagaria em tempo real, o valor correspondente ao volume da informação transferida. O pagamento processar-se-ia de forma automática, através do cartão de crédito ou de um método seguro de transacção, no momento da recepção da informação.

Outra possibilidade adicional no caso dos quadros a pedido seria o fornecimento de um orçamento prévio, possibilitando ao utilizador o conhecimento antecipado do valor a pagar pela satisfação do seu pedido.

A utilização da Internet nos institutos de estatística pode expandir-se para além da difusão de informação estatística. Uma das áreas possíveis é a realização de inquéritos às empresas ou a individuais através da Internet. Da mesma forma que hoje se recebem as respostas em papel ou em disquete, a utilização da Internet possibilitaria um meio mais rápido, barato e eficaz de resposta a inquéritos.

## INFOLINE

Para finalizar, salienta-se que a primeira parte do Sistema proposto pelo autor foi desenvolvida no INE no Projecto INFOLINE - Informação Estatística On-line, estando disponível desde 2 Janeiro de 1997.

Contém actualmente cerca de 5 000 quadros com a informação mais recente, publicada ou disponível para Publicação, com desagregação máxima ao nível de concelho.

Neste momento estão em fase de desenvolvimento e implementação algumas das funcionalidades propostas para a fase 2.

## 5. CONCLUSÕES

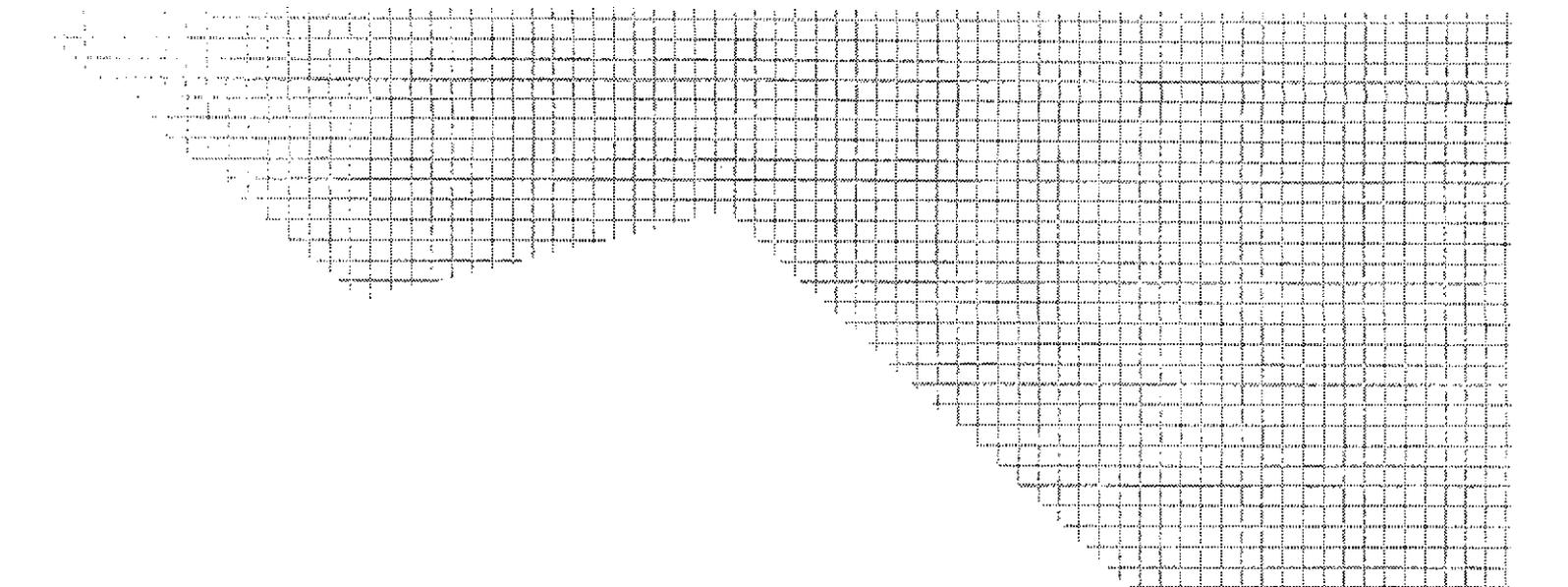
A aposta estratégica em termos de difusão de informação estatística, nos próximos anos, na generalidade dos INEs, passa pela utilização da Internet como meio privilegiado de difusão.

Uma conclusão evidente é a de que a Internet vai desempenhar um papel crucial no desenvolvimento da sociedade da Informação e pode alterar consideravelmente a forma como a informação estatística será utilizada no futuro.

## REFERÊNCIAS BIBLIOGRÁFICAS

Este artigo constitui uma síntese da tese de dissertação de Mestrado do autor, apresentada em Abril de 1997 no ISEGI - Instituto Superior de Estatística e Gestão de Informação, Lisboa.





# INFORMAÇÕES



---

---

## ACTIVIDADES E PROJECTOS IMPORTANTES NO ÂMBITO DO SISTEMA ESTATÍSTICO NACIONAL

---

---

---

---

## IMPORTANT ACTIVITIES AND PROJECTS IN THE SCOPE OF THE NATIONAL STATISTICAL SYSTEM

---

---

---

---

### O INQUÉRITO ÀS EMPRESAS HARMONIZADO - REFORMULAÇÃO METODOLÓGICA

---

---

---

#### INTRODUÇÃO

---

O Inquérito às Empresas/Harmonizado, operação estatística realizada pelo INE desde finais da década de oitenta, foi recentemente objecto de significativa reformulação, tendo por parâmetros essenciais o Regulamento (CE, EUROTOM) n.º 58/97, de 20/12/96, relativo às estatísticas estruturais das empresas, e as necessidades do Sistema Europeu de Contas - SEC 1995.

O presente texto constitui uma pequena síntese da metodologia definida pelo INE neste novo projecto, já em execução para o ano de referência de 1996, que será oportunamente divulgada.

---

#### 1 - APRESENTAÇÃO DO PROJECTO

---

---

##### 1.1 - DEFINIÇÃO DOS OBJECTIVOS

---

O Inquérito às Empresas/Harmonizado tem como principal objectivo estabelecer um quadro comum de recolha, compilação e transmissão de dados sobre a estrutura e actividade das empresas.

Pretende-se disponibilizar informação estatística que permita designadamente analisar:

- A estrutura e evolução da actividade das empresas
- Os factores de produção utilizados e outros elementos que permitam medir a actividade, os resultados e a competitividade das empresas
- O desenvolvimento regional, nacional, comunitário e internacional das empresas
- As relações com os mercados externos
- As características das pequenas e médias empresas
- As particularidades das empresas face às especificidades dos sectores em que desenvolvem a sua actividade

---

## 1. 2 - ÂMBITO E COBERTURA

---

### 1. 2. 1 - DE OBSERVAÇÃO

---

Com esta operação estatística pretende-se observar as empresas portuguesas em actividade, de acordo com o FGUE - Ficheiro Geral de Unidades Estatísticas, do Instituto Nacional de Estatística (INE).

### 1. 2. 2 - GEOGRÁFICO

---

Ao nível geográfico o inquérito tem cobertura nacional - CONTINENTE E REGIÕES AUTÓNOMAS DOS AÇORES E DA MADEIRA, sendo assegurada a representatividade por Regiões NUTS II.

### 1. 2. 3 - TEMPORAL

---

A inquirição é feita, anualmente, no 1º semestre do ano  $n + 1$  sobre o ano  $n$ .

Os resultados serão disponibilizados em duas fases:

- 10 meses após o período de observação - Resultados rápidos (EUROSTAT);
- 18 meses após o período de observação - Resultados globais da operação estatística.

### 1. 2. 4 - DE ACTIVIDADE

---

O projecto abrange as empresas que, de acordo com a *Classificação Portuguesa das Actividades Económicas - CAE - Rev. 2*, se encontram classificadas com actividade principal nas secções:

*Secção A*

Agricultura, produção animal, caça e silvicultura

*Secção B*

Pesca

*Secção C*

Indústrias extractivas

*Secção D*

Indústrias transformadoras

*Secção E*

Produção e distribuição de electricidade, gás e água

*Secção F*

Construção

*Secção G*

Comércio por grosso e a retalho; reparação de veículos automóveis, motociclos e de bens de uso pessoal e doméstico

*Secção H*

Alojamento e restauração

*Secção I*

Transportes, armazenagem e comunicações

*Secção K*

Actividades imobiliárias, alugueres e serviços prestados às empresas

*Secção M*

Educação

*Secção N*

Saúde e acção social

*Secção O*

Outras actividades de serviços colectivos, sociais e pessoais

São também consideradas as empresas que, em termos de CAE - Rev. 2, desenvolvem actividades secundárias no âmbito da secção D - Indústrias transformadoras.

---

**1.3 - UNIDADE ESTATÍSTICA DE OBSERVAÇÃO**

A unidade estatística de observação é a empresa.

A elaboração da resposta, por parte do informador, terá como base os principais documentos contabilísticos da empresa.

---

**1.4 - IDENTIFICAÇÃO DOS PRINCIPAIS UTILIZADORES**

O projecto foi concebido com o objectivo de dar resposta aos seguintes utilizadores:

**Internos**

- Contas Nacionais / Regionais / Intermédias / Satélite
- FGUE - Ficheiro Geral de Unidades Estatísticas
- Departamentos de Produção
- Outras Unidades Orgânicas do INE
- Direcções Regionais do INE
- Serviço Regional de Estatística dos Açores
- Direcção Regional de Estatística da Madeira

**Externos**

*Nacionais*

- Administração Pública Central e Local
- Institutos Públicos

- Empresas
- Representações Diplomáticas
- Estabelecimentos de Ensino
- Professores / Estudantes
- Outros

#### *Internacionais*

- EUROSTAT - Regulamento (CE, EURATOM) Nº 58 / 97, de 1996. 12. 20
- ONU, OCDE, OIT, FAO, UNCTAD, FMI
- Outros

---

### 1.5 - PRINCIPAIS ÁREAS TEMÁTICAS DE INQUIRIÇÃO

---

O inquérito visa fundamentalmente a produção de informação estatística nas seguintes áreas:

#### *Empresas*

Número de empresas e respectiva distribuição por actividade, classes de dimensão e localização geográfica.

#### *Emprego*

Pessoal ao serviço e duração do trabalho.

#### *Custos*

Matérias primas e mercadorias, Fornecimentos e serviços externos, Custos com o pessoal (remunerações, encargos e outros custos com o pessoal) e outros custos da Demonstração dos Resultados.

#### *Proveitos*

Vendas, Prestações de serviços e outros proveitos da Demonstração dos Resultados.

#### *Elementos patrimoniais da empresa*

Balanço - Activo, Capital próprio e Passivo.

#### *Investimento*

Activo bruto - imobilizado incorpóreo e corpóreo - aquisições e desinvestimentos (total e por tipos de bens).

---

### 1.6 - MÉTODO DE RECOLHA

---

O método de recolha utilizado é o envio por *via postal*.

Dependendo da taxa de resposta, da importância relativa das empresas em falta, quer pela sua dimensão, quer pela representatividade na actividade; considerando ainda, a qualidade da resposta e localização geográfica da empresa, o envio por via postal poderá ser complementado com *entrevista telefónica e ou recolha directa*.

## 1.7 - ENQUADRAMENTO LEGAL

O projecto pretende dar resposta ao Regulamento (CE, EURATOM) N° 58 / 97 do Conselho, de 20 de Dezembro de 1996, relativo às estatísticas estruturais das empresas.

Este projecto não permite responder às exigências do Regulamento no que concerne à Secção J - Actividades financeiras, dado que as especificidades contabilísticas das empresas financeiras não se enquadram nas das empresas que desenvolvem outras actividades.

## 1.8 - IMPLICAÇÕES HORIZONTAIS DO PROJECTO

O Departamento de Estatística das Empresas é o responsável pela coordenação do projecto ao nível nacional.

No entanto ao nível da concepção, desenvolvimento, gestão e execução é indispensável o envolvimento de outras Unidades Orgânicas do INE, Direcções Regionais, Serviço Regional de Estatística dos Açores e Direcção Regional de Estatística da Madeira.

### *CONCEPÇÃO, METODOLOGIA E FICHEIROS*

#### **DCI - Departamento de Coordenação e Integração**

##### *Serviço de Coordenação Estatística*

- Controlo e aprovação dos Instrumentos de Notação

##### *Serviço de Ficheiros Gerais de Unidades Estatísticas*

- Criação do Universo de Referência

##### *Serviço de Metodologia Estatística*

- Gestão do Universo de Referência, selecção e gestão da Amostra e do Ficheiro de lançamento
- Apoio técnico na vertente Metodologia Estatística

### *INFORMÁTICA*

#### **DSII - Departamento de Sistemas de Informação e Informática**

- Concepção, desenvolvimento, gestão e manutenção da aplicação informática
- Impressão e personalização dos Instrumentos de Notação
- Apoio técnico na componente informática do projecto

## **GESTÃO E EXECUÇÃO**

### **Direcções Regionais do Norte, Centro e Alentejo**

- Gestão e execução do inquérito em todas as fases, junto das empresas da respectiva região

### **Serviço Regional de Estatística dos Açores**

- Gestão e execução do inquérito em todas as fases, junto das empresas da respectiva região

### **Direcção Regional de Estatística da Madeira**

- Gestão e execução do inquérito em todas as fases, junto das empresas da respectiva região

## **DIFUSÃO**

### **Direcções Regionais do Norte, Centro e Alentejo**

- Reprodução e comercialização das publicações
- Difusão de informação não-publicada

### **Serviço Regional de Estatística dos Açores**

- Reprodução e comercialização das publicações
- Difusão de informação não-publicada

### **Direcção Regional de Estatística da Madeira**

- Reprodução e comercialização das publicações
- Difusão de informação não-publicada

---

## **2 - INSTRUMENTOS BÁSICOS DE SUPORTE**

---

### **2.1 - FICHEIRO**

FGUE - Ficheiro Geral de Unidades Estatísticas do INE, é o instrumento base para constituição do Universo de Referência e selecção da Amostra do Inquérito às Empresas / Harmonizado.

No momento de constituição do Universo o FGUE deve ter a mais recente actualização possível, no que se refere aos dados físicos e económicos reportados ao ano anterior ao período de referência.

---

### **2.2 - NOMENCLATURAS**

O inquérito tem como suporte uma nomenclatura de actividades e uma nomenclatura geográfica:

## ACTIVIDADES

### Classificação Portuguesa das Actividades Económicas (CAE - Rev. 2)

- nível de desagregação foi definido em função das necessidades expressas dos utilizadores.

## REGIONAL

### Nomenclatura das Unidades Territoriais para Fins Estatísticos

- Regiões NUTS II - Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve, Açores e Madeira

---

## 2.3 - CONCEITOS

A concretização do inquérito abrange um conjunto de conceitos, em particular estatísticos, contabilísticos e económicos, tendo-se utilizado como principais documentos de referência o *Manual de Conceitos Estatísticos do INE, os Conceitos e Definições do Regulamento, N.º 58 / 97 (CE, EURATOM)* sobre estatísticas estruturais das empresas, *Regulamento N.º 696 (CEE)* relativo às unidades estatísticas de observação e de análise do sistema produtivo na Comunidade (15 de Março de 1993) e o *Plano Oficial de Contabilidade (POC)*.

---

## 3 - ESPECIFICAÇÕES METODOLÓGICAS

---

### 3.1 - UNIVERSO

O Universo do Inquérito às Empresas / Harmonizado, é constituído a partir do FGUE - Ficheiro Geral de Unidades Estatísticas do INE.

*A constituição e gestão do Universo de Referência do IEH é da responsabilidade do Departamento de Coordenação e Integração (DCI).*

---

#### 3.1.1 - CRITÉRIOS DE CONSTITUIÇÃO

O universo de referência do IEH deve incluir as empresas que, simultaneamente, respeitem o seguinte conjunto de regras:

- Classificação Portuguesa das Actividades Económicas CAE - Rev. 2:

#### *Secção A*

Agricultura, produção animal, caça e silvicultura

#### *Secção B*

Pesca

#### *Secção C*

Indústrias extractivas

*Secção D*

Indústrias transformadoras

*Secção E*

Produção e distribuição de electricidade, gás e água

*Secção F*

Construção

*Secção G*

Comércio por grosso e a retalho; reparação de veículos automóveis, motociclos e de bens de uso pessoal e doméstico

*Secção H*

Alojamento e restauração

*Secção I*

Transportes, armazenagem e comunicações

*Secção K*

Actividades imobiliárias, alugueres e serviços prestados às empresas

*Secção M*

Educação

*Secção N*

Saúde e acção social

*Secção O*

Outras actividades de serviços colectivos, sociais e pessoais

Devem, ainda, ser consideradas as empresas que, em termos de CAE - Rev.2, desenvolvam actividades secundárias no âmbito da secção D - Indústrias transformadoras.

- **Forma jurídica**

- 01 - Empresa pública
- 02 - Empresa estrangeira
- 03 - Empresário em nome individual
- 04 - Sociedade irregular
- 05 - Agrupamento complementar de empresas
- 06 - Sociedade em nome colectivo
- 07 - Sociedade em comandita
- 10 - Sociedade anónima
- 11 - Sociedade anónima sem participação maioritária do sector público
- 12 - Sociedade anónima com participação maioritária do sector público
- 13 - Sociedade anónima com participação maioritária do sector público a 100 %
- 20 - Sociedade por quotas

- 21 - Sociedade por quotas sem participação maioritária do sector público
  - 22 - Sociedade por quotas com participação maioritária do sector público
  - 23 - Sociedade por quotas com participação maioritária do sector público a 100 %
  - 24 - Sociedade unipessoal por quotas
  - 30 - Sociedade cooperativa
  - 32 - Sociedade civil sob forma comercial
  - 34 - Sociedade civil estrangeira sob forma comercial
  - 36 - Estabelecimento individual de responsabilidade limitada
  - 38 - Agrupamento europeu de interesse económico comercial
  - 97 - Entidade equiparada a pessoa colectiva estrangeira
- **Data de constituição da empresa**

DCO  $\leq$  n (ano do inquérito)
  - **Localização da sede**

Incluir todas as empresas com localização da sede no Continente e nas Regiões Autónomas dos Açores e da Madeira.

LOC (DT / CC / FG)  $\in$  [ 010100, 490101 ]
  - **Situação da empresa**

STA = 02 (em actividade)
  - **Pessoal ao serviço**

No universo do IEH devem ser excluídas as empresas que, simultâneamente, apresentem zero “ 0 ” pessoas ao serviço e ausência de volume de vendas “ 0 ”.

### 3.1.2 - DESENHO DE REGISTO

#### *Nível: Empresa*

<b>Atributo</b>	<b>Designação</b>
NPC	Número de Pessoa Colectiva
DD-COD	Código de Distrito
CC-COD	Código de Concelho dentro do Distrito
FG-COD	Código de Freguesia dentro do Concelho
NUTS	Nomenclatura das Unidades Territoriais para Fins Estatísticos
CAE	CAE 73

NACE	CAE - Rev. 2
DT-NACE	Data da Última Actualização da CAE - Rev. 2
ANO	Ano de Constituição da Empresa
MÊS	Mês de Constituição
FJR	Forma Jurídica
STA	Situação da empresa
SIN	Reclassificação preliminar
CPS	Capital Social
PTE	Percentagem de Capital Estrangeiro
PNP	Percentagem de Capital Público
PNV	Percentagem de Capital Privado
NPS	Número de Pessoas ao Serviço
ENPS	Escalão de Pessoal ao Serviço
NST	Número de Estabelecimentos
VVN	Volume de Vendas
EVVN	Escalão de Volume de Vendas
NOM	Nome
MOR	Morada
CTT-COD	Código Postal
CTT-AUX	Código Postal Auxiliar
CTT-DSG	Código de Localidade Postal
TLF-IND	Indicativo do Telefone
TLF	Telefone
FAX-IND	Indicativo do Fax
FAX	Fax
TLX	Telex
D-CAE	Documento da CAE 73
D-NACE	Documento da CAE - Rev. 2
D-NPS	Documento do Número de Pessoas ao Serviço
D-VVN	Documento do Volume de Vendas
D-NST	Documento do Número de Estabelecimentos

*Nível: Estabelecimento*

Atributo	Designação
NPC	Número de Pessoa Colectiva
EST	Número do Estabelecimento

DD-COD	Código de Distrito
CC-COD	Código de Concelho dentro do Distrito
FG-COD	Código de Freguesia dentro do Concelho
NUTS	Nomenclatura das Unidades Territoriais para Fins Estatísticos
CAE	CAE 73
NACE	CAE - Rev. 2
SEDE-EST	Indicação de Sede
STA	Situação
NPS	Número de Pessoas ao Serviço
VVN	Volume de Vendas
MOR	Morada
CTT-COD	Código Postal
CTT-AUX	Código Postal Auxiliar
CTT-DSG	Código de Localidade Postal
TLF-IND	Indicativo do Telefone
TLF	Telefone
FAX-IND	Indicativo do Fax
FAX	Fax
TLX	Telex

*Nível: NACE-SEC (Actividades secundárias)*

Atributo	Designação
NPC	Número de Pessoa Colectiva
NACE-SEC	Actividades Secundárias (CAE - Rev. 2)

### 3. 2 - AMOSTRA

*A selecção e gestão da Amostra do IEH é da responsabilidade do DCI - Serviço de Metodologia Estatística.*

#### 3. 2. 1 - VARIÁVEIS DE ESTRATIFICAÇÃO

A Amostra deve ser definida tendo em conta as seguintes variáveis:

- FJR** - Forma jurídica
- NPS** - Número de pessoas ao serviço
- NUTS** - Nomenclatura das Unidades Territoriais para Fins Estatísticos

3. 2. 2 - ESCALÕES DE ESTRATIFICAÇÃO

- Escalões de forma jurídica

**EFJR = 1 - Empresas do sector público**

Corresponde às unidades estatísticas do FGUE classificadas com as formas jurídicas: 01, 12, 13, 22 e 23.

**EFJR = 2 - Empresas privadas**

Corresponde às unidades estatísticas do FGUE classificadas com as formas jurídicas: 02, 04, 05, 06, 07, 10, 11, 20, 21, 24, 30, 32, 34, 36, 38 e 97.

**EFJR = 3 - Empresário em nome individual**

Corresponde às unidades estatísticas do FGUE classificadas com a forma jurídica 03.

- ENPS - Escalões de pessoal ao serviço

0	0	peessoas ao serviço
1	1 a 9	peessoas ao serviço
2	10 a 19	peessoas ao serviço
3	20 a 49	peessoas ao serviço
4	50 a 99	peessoas ao serviço
5	100 a 199	peessoas ao serviço
6	200 a 499	peessoas ao serviço
7	500 e mais	peessoas ao serviço

- ENUT - Escalões de NUTS II

101 -	Norte
102 -	Centro
103 -	Lisboa e Vale do Tejo
104 -	Alentejo
105 -	Algarve
201 -	Açores
301 -	Madeira

- Escalões de Classificação Portuguesa das Actividades Económicas / CAE  
- Rev. 2 (ECAE)

Só devem ser consideradas no Universo e Amostra do IEH as empresas que no FGUE se encontrem classificadas ao nível máximo de desagregação da CAE - Rev. 2.

As empresas insuficientemente reclassificadas (CAE's inválidas) não entram no Universo do IEH.

- EVVN - Escalões de volume de vendas

EVVN = 1    VVN ≤ 30 000 mil escudos

EVVN = 2    VVN > 30 000 mil escudos

A variável Volume de vendas deve ser utilizada apenas para assegurar uma maior homogeneidade entre as empresas dos diversos estratos do Universo.

### 3.3 - TIPO DE INQUIRição

O inquérito será realizado por amostragem e de forma exaustiva, de acordo com os seguintes parâmetros:

**Amostragem** - Unidades estatísticas com menos de 100 pessoas ao serviço -  
NPS < 100

**Exaustivo** - Unidades estatísticas com 100 e mais pessoas ao serviço -  
NPS ≥ 100

### 4 - INSTRUMENTOS DE NOTACÃO

Na concepção dos Instrumentos de Notação estiveram presentes, nomeadamente, os seguintes aspectos:

- Necessidades expressas pelos utilizadores;
- Ajustamento da sua dimensão à capacidade de resposta das empresas;
- Identificação com o Plano Oficial de Contabilidade e Modelos Fiscais.

Com o inquérito pretende-se a inquirição em três perspectivas distintas:

- *Empresa*
- *Estabelecimento*
- *Actividade*

O inquérito é composto por um *Modelo Comum* - Modelos A e B - a enviar às empresas em função do número de pessoas ao serviço, de acordo com a seguinte distribuição:

- *Modelo A* - Unidades Estatísticas com EFJR = 1,2 e 3 e ENPS ≥ 3
- *Modelo B* - Unidades Estatísticas com EFJR = 1,2 e 3 e ENPS ≤ 2

O inquérito tem uma componente adicional de *Anexos Específicos*, de acordo com os respectivos sectores de actividade.

---

## 5 - PRODUTOS RESULTANTES DO IEH

---

Os resultados do IEH têm como objectivo corresponder às necessidades dos utilizadores de dados desta operação estatística.

Pretende-se, também, dotar as Direcções Regionais do INE e as Regiões Autónomas de uma componente parametrizada de obtenção de produtos, que lhes permita dar resposta à procura desta informação estatística nas respectivas Regiões.

---

### 5.1 - DADOS DE REGISTO (MICRO-DADOS SUBMETIDOS A CONTROLO DE QUALIDADE)

---

A disponibilização de dados de registo destina-se fundamentalmente a:

- Actualização do FGUE - Ficheiro Geral de Unidades Estatísticas
- Contas Nacionais / Regionais / Intermédias / Satélite
- Outros utilizadores

---

#### 5.1.1 - FGUE - FICHEIRO GERAL DE UNIDADES ESTATÍSTICAS

---

O FGUE será actualizado, com base nos dados do IEH em três vertentes:

*a) Empresas*

- **Dados físicos**
  - Dados gerais de identificação, caracterização e situação
- **Dados económicos**
  - Pessoal ao serviço e Volume de vendas

*b) Estabelecimentos*

- **Dados físicos**
  - Dados gerais de identificação, caracterização e situação
- **Dados económicos**
  - Pessoal ao serviço e Volume de vendas

*c) Actividades secundárias*

- Identificação das actividades secundárias da empresas

### 5.1.2 - CONTAS NACIONAIS / REGIONAIS / INTERMÉDIAS / SATÉLITE

Para as Contas Nacionais / Regionais / Intermédias / Satélite será preparada uma componente de disponibilização de dados de registo, por forma a que estes utilizadores possam ter acesso a toda a informação recolhida e registada, independentemente da sua situação em termos de apuramento.

Esta componente terá integrada a respectiva validação, por forma a que possam ser identificadas as empresas com erros de aviso e erros fatais, bem como os respectivos mapas de erro.

Em princípio esta informação destina-se a servir de suporte à preparação de outras aplicações informáticas, a desenvolver pelos respectivos utilizadores, conforme o interesse manifestado no decorrer do processo.

### 5.1.3 - OUTROS UTILIZADORES

Serão disponibilizados os seguintes produtos:

- VR - Dados por Número de Pessoa Colectiva (NPC)
- ESP - Dados do quadro dos Estabelecimentos
- ACP - Dados das Actividades Secundárias

### 5.2 - DADOS POR ESTRATO

A especialização de dados tem início na imputação de não-respostas e construir-se-á na obtenção das variáveis apuráveis, calculadas por estrato.

Obtém-se como produto intermédio um “ Output “ fixo de dados resposta, por NPC, com identificação das empresas submetidas a tratamento de não-respostas.

Esta fase termina com a produção de um “ Output “ fixo com dados por estrato, tipo QQ.

### 5.3 - APURAMENTOS

O desenvolvimento e realização dos apuramentos tem como pressuposto uma perspectiva integrada, ao nível nacional e regional, e parte do princípio de que todas as variáveis questionadas são apuradas.

Assim, pretende-se que a especialização de resultados cubra integralmente os Modelos A e B - Módulo Comum e respectivos Anexos Específicos.

#### 5.3.1 - QUADROS PRÉ-PROGRAMADOS

O DSII assegurará a programação de quadros, em número reduzido, com as principais variáveis, associando-lhe o cálculo dos erros de amostragem, definido pelo Serviço de Metodologia Estatística.

---

### 5.3.2 - QUADROS PROGRAMADOS EM SAS

Estes quadros destinam-se às publicações do IEH e a dar resposta a pedidos de utilizadores.

---

### 5.3.3 - EUROSTAT

Em conformidade com as disposições do Regulamento (CE, EURATOM) Nº 58 / 97 estão, neste momento, definidos os seguintes domínios de informação:

- Módulo Comum para as Estatísticas Estruturais Anuais
- Módulo Pormenorizado para as Estatísticas Estruturais da Indústria
- Módulo Pormenorizado para as Estatísticas Estruturais do Comércio
- Módulo Pormenorizado para as Estatísticas Estruturais da Construção

---

### 5.4 - APURAMENTOS ESPECIAIS

Apuramentos, não previstos de forma regular, cuja concepção e realização decorrerá de necessidades específicas expressas pelos utilizadores, sendo assegurada a sua realização mediante análise casuística da informação estatística em causa.

No Estrangeiro:

*Abroad:*

1997

- 07 - 12 de Setembro  
**SPRUCE-IV International Conference on Statistical Aspects of Health and the Environment**, Enschede, Holanda.  
Informações: *A. Stein*, ITC, PO Box 6, 7500 AA Enschede, The Netherlands;  
Telf.: +31.53.4874222; FAX: +31.53.4873399.  
E - mail: spruce@itc.nl
- 08 - 12 de Setembro  
**Implementação do Novo Sistema de Contas Nacionais das Nações Unidas**, Maputo, Moçambique.  
Informações: *Dr. Valeriano Levene*, Vice-Presidente do INE-Moçambique;  
Telf.: 258 1 49 09 30; FAX: 258 1 49 09 30.
- 10 - 12 de Setembro  
**IV International Meeting of Multidimensional Data Analysis NGUS'97**, Bilbao, Espanha.  
Informações: Facultad de CC.EE. y Empresariales, Universidad del País Vasco, Avda. Lehendakari Aguirre 83, 48015 Bilbao, Spain; FAX: +34.4.4797554. E - mail: ngus@bs.ehu.es
- 14 - 20 de Setembro  
**Workshop on Resampling Methods**, Montréal, Canadá.  
Informações: Centre de recherches mathématiques, Univ. de Montréal, C.P. 6128, Succ. Centre-ville, Montréal (Québec) H3C 377 Canada; FAX: +51.4.3432254. E - mail: pelletl@crm.umontreal.ca
- 15 - 16 de Setembro  
**3rd International Meeting on Statistical Methods in Biopharmacy**, to be held in Paris, France.  
Informações: *B. Scherrer*; Telf.: 33.1.40967511; FAX: 33.1.40967591.  
E - mail: Scherrer-b-m@magic.fr
- 15 - 18 de Setembro  
**42nd Annual Conference of the German Society of Medical Informatics, Biometry and Epidemiology (GMDS)**, to be held at the University of Ulm, Germany.  
Informações: Tagungssekretariat 42, Jahrestagung GMDS, *Gisela Büchele*, Abteilung Biometrie und Med. Dokumentation, Universität Ulm, D-89070 Ulm, Germany; Telf.: 49-731-5026905; FAX: 49-731-5026902.  
E - mail: gmds97@medizin.uni-ulm.de  
or  
Internet: <http://www.uni-ulm.de/uni/fak/medizin/biodok/gmds.html>

- 15 - 19 de Setembro  
**Sixth ECAS Course on Time Series Analysis**, El Escorial - Madrid, Espanha.  
Informações: *D. Pena*, Dept. of Statistics and Econometrics, Universidad Carlos III de Madrid, Madrid 126 GETAFE 28903, Spain; FAX: +34.1.6249849. E - mail: [dpena@est-econ.uc3m.es](mailto:dpena@est-econ.uc3m.es)
  
- 16 - 19 de Setembro  
**First International Conference on Mathematical Methods in Reliability, MMR'97**, Bucareste, Roménia.  
Informações: *P. Ulmeanu*, Politehnica Univ. of Bucharest, Faculty of Power Eng., MMR'97 Int. Conf., 313 Spaiul Independentei, R77206 Bucharest, Romania; FAX: +40.1.3123161.  
E - mail: [mmr@ccel.fiab.pub.ro](mailto:mmr@ccel.fiab.pub.ro)
  
- 22 - 26 de Setembro  
**German Statistical Week**, Bielefeld, Germany.  
Informações: Verband Deutscher Stadttestatistiker, Amt fur Statistik und Wahlen, Schwartzstrasse 73, 46045 Oberhausen, Germany.
  
- 26 - 27 de Setembro  
**Fourth Workshop on Case Studies of Bayesian Statistics in Science and Technology**, Pittsburg, Estados Unidos.  
Informações: *R. Kass*, Dept. of Statistics, Carnegie Mellon Univ., Pittsburg, PA 15213-3890, USA; FAX: +412.2687828.  
E - mail: [kass@stat.cmu.edu](mailto:kass@stat.cmu.edu)
  
- 29 de Setembro - 01 de Outubro  
**Analise de Conjuntura nas Economias de Mercado**, Maputo, Moambique.  
Informações: *Dr. Valeriano Levene*, Vice-Presidente do INE-Moambique; Telf.: 258 1 49 09 30; FAX: 258 1 49 09 30.
  
- 16 - 18 de Outubro  
**IFIP WG 10.5 Advanced Research Working Conference on Correct Hardware Design and Verification Methods (CHARME '97)**, Montreal, Quebec, Canada.  
Informações: IFIP Secretariat, Hofstrasse 3, A-2361 Laxenburg, Austria; Telf.: 43 2236 73616; FAX: 43 2236 736169.  
WWW: <http://www.ifip.or.at/>
  
- 09 - 15 de Novembro  
**Workshop on Empirical Bayes and Likelihood Inference**, Montreal, Canada.  
Informações: Centre de recherches mathematiques, Univ. de Montreal, C.P. 6128, Succ. Centre-ville, Montreal (Quebec) H3C 377 Canada; FAX: +51.4.3432254. E - mail: [pelletl@crm.umontreal.ca](mailto:pelletl@crm.umontreal.ca)
  
- 18 - 21 de Novembro  
**IFIP TC6, Joint International Conference on Formal Description Techniques for Distributed Systems and Communication Protocols and Protocol Specification, Testing and Verification**, to be held in Osaka, Japan.  
Informações: IFIP Secretariat, Hofstrasse 3, A-2361 Laxenburg, Austria; Telf.: 43 2236 73616; FAX: 43 2236 736169.  
WWW: <http://www.ifip.or.at/ftp:ftp.ifip.or.at>

- 08 - 10 de Dezembro  
**Meeting on "Territory Cuttings: Scientific and Politic Approaches"**, to be held in Entretiens Jacques Cartier, Lyon, France.  
Informações: *G. Desplanques*, INSEE, Regional Direction for Rhone-Alpes Region, 165, rue Garibaldi, 69003 Lyon, France.  
 Telf.: 33 4 78632506; FAX: 33 4 78632511.
  
- 15 - 17 de Dezembro  
**International Conference on Comparative Analysis of Entrepriase (Micro)Data**, Bergamo, Italy.  
Informações: CAED Secretary, Dipartimento di Matematica, Statistica, Informatica et Applicazioni, Universita di Bargamo, Piazza Rosate 2, 24129 Bergamo, Italy; Telf.: 39 35 277536; FAX: 39 35 249598.  
 E - mail: CAED@ibguniv.unibg.it  
 or  
 WWW: <http://www.unibg.it/dmsia/caed/caedhome.html>
  
- 26 - 28 de Dezembro  
**3rd Triennial Calcutta Symposium on Probability and Statistics**, to be held in Calcutta, India.  
Informações: *Kalyan Das*, Department of Statistics, Calcutta University, 35, Ballygunge Circular Road, Calcutta 700 019, India; FAX: 91-033-241-3222.
  
- 29 de Dezembro - 01 de Janeiro  
**International Conference on "Recent Advances in Statistics and Probability"**, Calcutta, India.  
Entidades Promotoras: *Indian Statistical Institute*.  
Informações: Ayanendranath Basu, Computer Science Unit, Indian Statistical Institute, 203 B. T. Road, Calcutta 700 035, India.  
 FAX: 91 33 5566680, 91 33 5566925.  
 E - mail: isibern@isical.ernet.in  
 or  
*R. L. Karandikar*, E - mail: isibern@isid.ernet.in  
 or  
*Mohan Delampady*, E - mail: isibern@isibang.ernet.in

## 1998

- Janeiro  
**7ème Colloque de l'Association de Comptabilité Nationale**, Paris, França.  
Informações: *Michel Boeda*, Organizing Committee Chair, D201, INSEE, 18 Bd. Aldolphe Pinard, 75675 Paris Cedex 14, France.  
 FAX: 33 1 41176849.
  
- 08 - 10 de Janeiro  
**International Conference on Stochastic Processes and Their Applications**, to be held at Anna University, Madras, India.  
Informações: *Dr. A. Vijaykumar*, Department of Mathematics, Anna University Chennai, (Madras) 600 025, India; FAX: 91 44 2350397.  
 E - mail: annalib@simetm.ernet.in

- 09 - 13 de Fevereiro  
**ProbaStat '98, International Conference on Probability and Mathematical Statistics**, Smolenice Castle – Bratislava, Eslováquia.  
Informações: ProbaStat '98, Mathematical Institute of the SAS, Stefanikova 49, SK-84173 Bratislava, Slovakia; FAX: +427.397316.  
E - mail: probstat@savba.sk
- 12 - 17 de Fevereiro  
**AAAS Annual Meeting – Sesquicentennial Celebration of the American Association for the Advancement of Science**, Philadelphia, Pennsylvania, Estados Unidos.  
Informações: AAAS Meetings Office, 1333 H St., NW, Washington D.C. 200005, USA; Telf.: 202 326 6450; FAX: 202 289 4021.
- 18 - 20 de Março  
**Agricultural Statistics 2 000: An International Conference on Agricultural Statistics – Data Needs and Methodology**, Washington, D.C., Estados Unidos.  
Informações: *Fred Vogel*, National Agricultural Statistics Service, U. S. Department of Agriculture, Room 5801-So. Buld., 1400 Independence Ave SW, Washington, D.C. 20250-2000, USA; FAX: 202 690 1311.  
E - mail: fvogel@nass.usda.gov
- 24 - 27 de Março  
**Munich Stochastic Days 1998**, Neubiberg/Munich, Germany.  
Informações: Muenchener Stochastik-Tage 1998, Univ. Prof. Dr. K. Marti, UniBw Muenchen, Fak. LRT, D-8557 Neubiberg, Munich.  
E - mail: Stoch.tage.98@unibw-muenchen.de
- 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
- 09 - 11 de Junho  
**21<sup>st</sup> 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

- 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](mailto:bphillips@swin.edu.au)  
or  
WWW: <http://www.nie.ac.sg:8000/~wwwmath/icots.html>
- 05 - 11 de Julho  
**SPA 98, 25<sup>th</sup> 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](mailto: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](mailto: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](mailto:ifcs.98@pow2.sta.uniroma1.it)  
or  
WWW: <http://www.pow2.sta.uniroma1.it/ifcs98>
- 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](mailto:probability@fields.utoronto.ca)  
or  
WWW: <http://www.fields.utoronto.ca/>
- 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](mailto:meetings@asa.mhs.compuserve.com)

- 12 - 18 de Agosto  
**EMS'98 European Meeting of Statisticians and the 7<sup>th</sup> 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](mailto:conf@ktl.mii.lt)
  
- 23 - 28 de Agosto  
**Prague Stochastics '98, a joint session of the Sixth Prague Symposium on Asymptotic Statistics and the 13<sup>th</sup> Prague Conference on Information Theory, Statistical decision Function, Random processes**, Prague, Czech republic.  
Informações: *Zuzana Prásková*, Dept. of Statistics, Charles University, Sokolovská 83, 186 00 Prague, Czech Republic.  
 E - mail: [stoch98@utia.cas.cz](mailto:stoch98@utia.cas.cz)
  
- 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](mailto:compstat-98@bristol.ac.uk)  
 or  
 WWW: <http://www.stats.bris.ac.uk/COMPSTAT/>
  
- 31 de Agosto - 04 de Setembro  
**15<sup>th</sup> IFIP World Computer Congress and SEC '98 - 14<sup>th</sup> 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](mailto:ifip98@ocg.or.at)  
 or  
[ocg@ocg.or.at](mailto: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](mailto:ifip98@neumann.hu)  
 or  
[h13588zub@ella.hu](mailto: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](mailto:dennis.trewin@abs.gov.au)  
 or  
*Denise Lievesley*, ESRC Data Archive, University of Essex, CO4 3SQ, UK.  
 E - mail: [denise@essex.ac.uk](mailto:denise@essex.ac.uk)

- 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](mailto:rss98@stams.strath.ac.uk)  
WWW: <http://www.stams.strath.ac.uk/rss98/index.html>
- 14 - 18 de Setembro  
**“German Statistical Week”**, Lübeck, Germany.  
Informações: Verband Deutscher Städtestatistiker, Bereich Statistik und Wahlen, Schwartzstrasse 73, 46045 Oberhausen, Germany.

### 1999

- 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](mailto:meetings@asa.mhs.compuserve.com)
- 19 – 23 de Agosto  
**The 6<sup>th</sup> 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: [etiit@ut.ee](mailto:etiit@ut.ee)  
or  
[kollo@ut.ee](mailto:kollo@ut.ee)



---

---

## 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 MAIO A 31 DE AGOSTO DE 1997):

a) *Cooperação desenvolvida com os PALOP e Macau*

A mobilização da população de Moçambique para o II Recenseamento Geral da População e Habitação, que se realiza entre 1 e 17 de Agosto de 1997, começou na televisão e rádio a 1 de Julho, através da transmissão na RTP Internacional, Televisão Nacional de Moçambique e RDP África, dos spots publicitários produzidos pela empresa moçambicana GOLO, com o apoio financeiro do Instituto da Cooperação Portuguesa.

Com Angola realizaram-se duas missões de identificação, nos domínios das Estatísticas Sociais e Estatísticas de Conjuntura, que permitiram efectuar o levantamento das necessidades e propor um programa de trabalho conjunto. Tiveram também lugar dois estágios no INE, um destinado a dois técnicos do Índice de Preços no Consumidor, e o outro a três técnicos do Departamento de Estatísticas do Comércio Externo, todos do INE Angolano.

Ocorreu a primeira missão ao novo Instituto Nacional de Estatística de Cabo Verde, anterior Direcção Geral de Estatística, que visou o levantamento da situação com vista à definição da arquitectura informática e do sistema de informação para o novo Instituto.

Com S. Tomé e Príncipe realizaram-se quatro acções de cooperação nas áreas das Estatísticas Regionais, Comércio Externo, Estatísticas da Pesca e Agricultura e de Conjuntura. O INE participou também num Seminário de Sensibilização sobre Estatísticas do Comércio Externo, realizado em São Tomé e Príncipe, no âmbito da implantação do Sistema Eurotrace (financiado pela União Europeia) e no Seminário sobre Informação Estatística e Gestão do Desenvolvimento, destinado aos quadros superiores políticos e administrativos do país.

Os projectos comuns com os PALOP sobre Estatísticas da Indústria e Construção e sobre as Classificações Económicas prosseguiram o seu curso normal, o primeiro através da realização de missões técnicas a Angola e Guiné-Bissau no domínio dos inquéritos estruturais às empresas industriais e de construção, as quais permitiram constatar francos progressos nestes dois países ao nível da reformulação da metodologia aplicável aos inquéritos.

Quanto às Classificações Económicas, Angola já dispõe de uma Classificação de Actividades Económicas, a publicar brevemente, e Cabo Verde tem a sua praticamente concluída. Prevê-se que os restantes países disponham das respectivas Classificações de Actividades Económicas até final deste ano.

Finalmente, salienta-se a admissão do INE na Comissão Interministerial para a Cooperação, órgão sectorial de apoio ao Governo, que tem por competência, nomeadamente a "... coordenação da execução dos programas e projectos de iniciativa pública".

No âmbito da cooperação entre o INE e a DSEC Macau, realizou-se no período de 18 de Abril a 10 de Maio uma missão à DSEC do responsável pelo Núcleo de Nomenclaturas, Dr. Saraiva Aguiar, para acompanhamento dos trabalhos no âmbito da Revisão da Classificação de Actividades de Macau (CAM).

#### ***b.1) Cooperação Estatística com os Países da Europa Central e Oriental (PECO) - Programa PHARE***

No âmbito do Programa PHARE de Assistência Técnica aos Países da Europa Central e Oriental, realizou-se na semana de 14 a 18 de Julho de 1997 uma missão ao INE da Bulgária com o objectivo de avaliar as acções de cooperação já realizadas e a definição de um programa de cooperação para 1998 entre os dois Institutos. Esta missão incluiu também um contacto com duas das Direcções Regionais Búlgaras, com o objectivo de se alargar o âmbito da cooperação à área das Estatísticas Regionais.

#### ***b.2) Cooperação Mediterrânica - Programa MEDSTAT***

Teve lugar em Malta, de 24 a 26 de Junho, a primeira reunião do Comité Director do Programa MEDSTAT. Participaram nesta reunião a Comissão Europeia e vários representantes dos Sistemas Estatísticos da União Europeia, incluindo Portugal, e dos Países Mediterrânicos. O Comité Director discutiu e aprovou os trabalhos a realizar nos próximos anos, ao nível dos sectores prioritários do MEDSTAT, e analisou novos sectores de actividades a incluir neste programa de cooperação regional.

#### ***b.3) Cooperação com os países da ex-URSS e Mongólia - Programa TACIS***

No âmbito do Programa TACIS de assistência técnica aos países da ex-URSS e Mongólia, Portugal está a desenvolver um projecto de cooperação na área da Difusão de Estatísticas Oficiais com o SSO (State Statistical Office) da Mongólia. Este projecto, dirigido pelo INE, está a ser concretizado pela DSEC-Macau, tendo-se já realizado uma missão à Mongólia, no período entre 2 a 9 de Maio último, e uma visita à DSEC-Macau da Directora do SSO da Mongólia, Mrs. Davaasuren e de dois técnicos da área da Difusão.

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 nº:    

Valor da transacção: 5.450\$00 Validade do cartão \_\_/\_\_/\_\_

 Junto cheque nº \_\_\_\_\_ 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 nº:    

Valor da transacção: 5.450\$00 Validade do cartão \_\_/\_\_/\_\_

 Junto cheque nº \_\_\_\_\_ 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



## NORMAS DE APRESENTAÇÃO DE MANUSCRITOS

Nos termos da alínea b) do n.º 3 do Artigo 5º 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 para publicação na *Revista de Estatística* 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. Os originais, com uma extensão não superior a vinte páginas, serão processados em *Word for Windows* e entregues em suporte papel acompanhado da respectiva *disquette*.
3. Na apresentação dos originais, os autores respeitarão ainda as seguintes normas:
  - 3.1. 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 sumário do artigo, com um máximo de cem palavras, seguido de um parágrafo com indicação de três a seis palavras-chave;
    - d) Na terceira página começará o texto do artigo, sendo as suas eventuais secções ou capítulos numeradas sequencialmente;
  - 3.2. 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:

BESSON, JEAN-LOUIS (1992b), "Les statistiques: vraies ou fausses?" in Besson, Jean-Louis (org.), *La Cité des chiffres ou l'illusion des statistiques*, Paris, Autrement, 26-62.
  - 3.3. 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*.
  - 3.4. 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*.

