



ISSN 0873-4275

INSTITUTO NACIONAL DE ESTATÍSTICA

PORTUGAL

REVISTA DE ESTATÍSTICA

STATISTICAL REVIEW



VOLUME I
1º QUADRIMESTRE 2002

REVISTA DE ESTATÍSTICA

STATISTICAL REVIEW

VOLUME I

1º QUADRIMESTRE DE 2002

ÍNDICE

INDEX

- MENSAGEM DO DIRECTOR SOBRE O NOVO FORMATO DA REVISTA <i>DIRECTOR'S STATEMENT ON THE NEW JOURNAL FORMAT</i>	5
- ARTIGOS <i>ARTICLES</i>	13
SERÁ QUE AS EMPRESAS PORTUGUESAS APRESENTAM VANTAGEM COMPETITIVA? <i>DO PORTUGUESE COMPANIES HAVE CMPETITIVE ADVANTAGES?</i> Por/By: <i>Manuela Magalhães Hill e Andrew B. Hill, PhD</i>	15
QUASI-RANDOM VERSUS PSEUDO-RANDOM NUMBERS TO APPROXIMATE POSTERIOR DISTRIBUTIONS NÚMEROS QUASI-ALEATÓRIOS VERSUS PSEUDO-ALEATÓRIOS PARA ESTIMAR DISTRIBUIÇÕES POSTERIORES Por/By: <i>Carlos Javier Pérez Sánchez, Jacinto Martín Jiménez e José Carlos Rojano Martín</i>	31
ANÁLISE DE VARIÂNCIA E ANÁLISE DE REGRESSÃO COM VARIÁVEIS DUMMY: MAIS SEMELHANÇAS DO QUE DIFERENÇAS <i>ANOVA AND REGRESSION WITH DUMMY VARIABLES: MORE SIMILARITIES THAN DIFFERENCES</i> 7Por/By: <i>Patrícia Oom do Valle e Efigénio Rebelo</i>	47
EASYCARE: UM SISTEMA DE AVALIAÇÃO DE IDOSOS (QUALIDADES PSICOMÉTRICAS) <i>EASY-CARE: ELDERY ASSESSMENT SYSTEM</i> 7Por/By: <i>Liliana Sousa, Helena Galante e Daniela Figueiredo</i>	87
- INFORMAÇÕES <i>INFORMATIONS</i>	103
CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS <i>CONGRESS, SEMINARS AND CONFERENCES</i>	105
FUNDAMENTO, OBJECTO E ÂMBITO DA REVISTA <i>FOUNDATION, SUBJECT MATTER AND SCOPE OF THE REVIEW</i>	119
NORMAS DE APRESENTAÇÃO DE ORIGINAIS PARA A REVISTA <i>RULES FOR SUBMITTING ORIGINALS TO THE REVIEW</i>	121

**MENSAGEM DO DIRECTOR
SOBRE O NOVO FORMATO DA REVISTA**

**DIRECTOR'S STATEMENT
ON THE NEW JOURNAL FORMAT**

VOLUME I

1º QUADRIMESTRE DE 2002

MENSAGEM DO DIRECTOR SOBRE O NOVO FORMATO DA REVISTA

Como os leitores terão certamente presente a Revista de Estatística começou a ser editada pelo Instituto Nacional de Estatística (INE) em 1996, há já seis anos, 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 respectivo conhecimento numa base estatística, fazendo-o chegar ao maior número possível de leitores, verificando-se que presentemente a Revista tem 334 assinantes.

Editada quadrimestralmente, enquanto publicação de pendor científico a Revista tem visado divulgar artigos originais sobre: investigação e análise estatística, pura e aplicada; e estudos e análises nos domínios económico, social e demográfico, pretendendo constituir um veículo prestigiado e, portanto, procurado pelos autores de trabalhos para os tornar amplamente conhecidos, não só pela comunidade dos académicos, mas também, em geral, pelos utilizadores dos métodos estatísticos e dos próprios utilizadores da informação estatística oficial nacional e comunitária.

Sendo uma publicação aberta a autores externos ao INE, nacionais e estrangeiros, com um Conselho Editorial de índole nacional, em que metade dos membros são professores universitários, intentou-se que os trabalhos publicados pudessem significar, para além do seu valor intrínseco, um patamar mais elevado de salutar intercâmbio científico e pessoal que se pretendeu fomentar.

Visando o sempre desejado desenvolvimento da Revista, reconhecendo-se, realisticamente, que a língua portuguesa constituía uma barreira tanto ao conhecimento internacional dos artigos publicados como ao surgimento de colaborações de autores estrangeiros, há três anos passou-se a aceitar a publicação de artigos na língua Inglesa, tendo-se logo fixado o objectivo de caminhar a médio prazo, de uma forma decidida e mais consistente, para a internacionalização da Revista.

Neste sentido, foi agora decidido proceder à efectiva internacionalização da Revista no próximo ano (2003), sendo oportuno salientar desde já os seguintes elementos caracterizadores do seu novo formato:

- O título passa a ser REVSTAT - Statistical Journal.
- A política editorial passa a ser centrada na publicação de artigos de investigação de alto nível nos domínios das Probabilidades e Estatística, com ênfase na originalidade e importância da investigação, sendo todos os artigos objecto de avaliação científica por pelo menos duas pessoas, uma do corpo de Editores e outra externa.
- A língua de trabalho passa a ser exclusivamente o Inglês.
- A periodicidade da edição continuará a ser quadrimestral, Março, Julho e Setembro, sendo o primeiro número publicado no final de Março de 2003.
- Serão publicados em média 6 artigos por cada número.
- **Corpo de Editores**, de índole internacional, será constituído por:
 - **Editor-Chefe**, Prof.^a *M. Ivette Gomes*, Faculdade de Ciências da Universidade de Lisboa, Portugal

- **Co-Editor**, Prof.^a *M. Antónia Amaral Turkman*, Faculdade de Ciências da Universidade de Lisboa, Portugal
- **Editor Executivo**, Dr. *Ferreira da Cunha*, Consultor da Direcção do Instituto Nacional de Estatística, Portugal, e Director da actual Revista desde a sua criação.
- **Editores Associados:**
 - Prof. *Barry Arnold*, Universidade da Califórnia, Riverside, EUA.
 - Prof.^a *H. Bacelar Nicolau*, Faculdade de Psicologia da Universidade de Lisboa, Portugal
 - Prof.^a *Susie Bayarri*, Universidade de Valência, Espanha
 - Prof. *João Branco*, Instituto Superior Técnico de Lisboa, Portugal
 - Prof.^a *M. Lucília Carvalho*, Faculdade de Ciências da Universidade de Lisboa, Portugal
 - Sir *David Cox*, Universidade de Oxford, Reino Unido
 - Prof. *Edwin Diday*, Universidade de Paris IX Dauphine, França
 - Prof. *Dani Gamerman*, Universidade Federal do Rio de Janeiro, Brasil
 - Prof.^a *Marie Husková*, Universidade Carlos de Praga, República Checa
 - Prof. *Isaac Meilijson*, Universidade de Tel-Aviv, Israel
 - Prof.^a *Nazaré Mendes-Lopes*, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Portugal
 - Prof. *António Pacheco*, Instituto Superior Técnico de Lisboa, Portugal
 - Prof. *Dinis Pestana*, Faculdade de Ciências da Universidade de Lisboa, Portugal
 - Prof. *Ludger Rüschendorf*, Instituto de Matemática Estocástica de Freiburg, Alemanha
 - Prof. *Gilbert Saporta*, Conservatório Nacional das Artes e Ofícios, Paris, França
 - Prof. *Jef Teugels*, Universidade Católica de Lovaina, Bélgica
 - Prof. *Radu Theodorescu*, Universidade de Laval, Sainte-Foy, QC, Canadá
- Os artigos deverão ser escritos em Inglês e poderão ser submetidos de três maneiras:
 - Enviando 3 cópias em papel ao Editor Executivo, que estará em contacto permanente com o Editor-Chefe e o Co-Editor,
 - Enviando 2 cópias em papel ao Editor Executivo e uma cópia a um dos dois Editores ou dos Editores Associados, cuja opinião o autor queira que seja tomada em linha de conta,
 - Enviando 1 cópia em papel ao Editor Executivo, conjuntamente com o correspondente ficheiro PDF ou Postscript para o E-mail: revstat@fc.ul.pt.
- Os manuscritos devem ser escritos numa só página de folha A4 em duplo espaço, com uma margem clara do lado esquerdo, e não podem ter mais de 30 páginas. A primeira página deverá incluir o nome, instituição e endereço do(s) autor(es), e um resumo com o máximo de 100 palavras, seguido das palavras-chave num máximo de 6, e a classificação do assunto

AMS 2000. Os autores deverão escrever a versão final dos artigos utilizando o LaTeX, no estilo da REVSTAT. Este estilo (REVSTAT.sty), exemplos de ficheiros (REVSTAT.tex) e instruções adicionais aos autores poderão ser obtidas em princípio no próximo mês de Setembro através do endereço web do INE: <http://www.ine.pt/revstat>.

- A submissão de artigos significa que contém um trabalho original ainda não publicado nem em vias de ser publicado sob qualquer forma algures.

Neste novo contexto, **os autores que queiram submeter artigos para publicação na REVSTAT poderão começar a fazê-lo a partir do dia 1 de Outubro deste ano.**

Informações complementares sobre esta iniciativa, direccionadas aos potenciais autores de artigos, serão dadas no próximo número da Revista de Estatística a ser editado na 1^a quinzena de Setembro próximo.

Irá ser também desenvolvida uma campanha de publicidade de que se destaca:

- (a) Publicação de informação relevante no Boletim da Sociedade Portuguesa de Estatística;
- (b) Envio por e-mail de informação relevante aos sócios da Associação Portuguesa de Classificação e Análise de Dados;
- (c) Elaboração de um folheto promocional em Inglês para distribuição em congressos e seminários;
- (d) Envio por e-mail de informação relevante em Inglês aos sócios da Sociedade Bernoulli do Instituto Internacional de Estatística;
- (e) Publicação de informação relevante em Inglês numa entrada específica da REVSTAT no website do INE.

DIRECTOR'S STATEMENT ON THE NEW JOURNAL FORMAT

As readers will certainly recall, the *Revista de Estatística* began to be published by the National Statistical Institute (INE) in 1996, six years ago now, aware of how vital a statistical culture is to understanding most phenomena in the present-day world, and its responsibility in disseminating statistical knowledge, ensuring that it reaches the greatest possible number of readers, of which 334 currently subscribe to it.

Coming out every four months as a science-oriented publication, the goal of the *Revista de Estatística* has been to publish original articles on: statistical research and analysis, both theoretical and applied, and studies and analyses in the economic, social and demographic domains, intended to constitute a prestigious medium and, therefore, sought by authors to make their work more widely known not only to the community of scholars, but also, generally speaking, to users of statistical methods as well as to the users of official national and Community statistical information.

Since the publication is open to authors outside INE, both domestic and foreign, with an Editorial Board, national in nature and on which half of the members are university professors, an attempt was made to ensure that published articles could denote, in addition to their intrinsic value, a higher level of beneficial scientific and personal exchange, which it aimed to develop.

Without ever losing sight of the desired aim of developing the *Revista de Estatística*, and recognising the fact that, realistically, the Portuguese language hindered both international access to the articles published and the contribution of foreign authors, three years ago it was accepted that the articles would be published in English, with decisive and more consistent steps being taken from the outset to ensure the medium-term goal of internationalising the Journal.

As a consequence, it has now been decided that the *Revista de Estatística* will be internationalised next year (2003), and being opportune to point out the features characterising its new format:

- The title will now be REVSTAT – Statistical Journal.
- The editorial policy will now focus on publishing research articles at the highest level in the domains of Probability and Statistics, with emphasis on the originality and importance of the research. All articles will undergo scientific evaluation by at least two persons, one from the Editorial Board and another, external.
- The only working language allowed will be English.
- The edition will continue to be published every four months, in March, July and September, with the first issue coming out at the end of March 2003.
- On average, 6 articles will be published per issue.
- The **Editorial Board**, international in nature, will comprise:
 - **Editor-in-Chief**, Prof. *M. Ivette Gomes*, Faculty of Science, University of Lisbon, Portugal

- **Co-Editor**, Prof. *M. Antónia Amaral Turkman*, Faculty of Science, University of Lisbon, Portugal
- **Executive Editor**, Dr. *Ferreira da Cunha*, Consultant to the Board of Directors of the National Statistical Institute, Portugal, and Director of the *Revista de Estatística* since its inception.
- **Associate Editors**:
 - Prof. Barry *Arnold*, University of California, Riverside, USA.
 - Prof. H. Bacelar Nicolau, Faculty of Psychology, University of Lisbon, Portugal
 - Prof. Susie Bayarri, University of Valencia, Spain
 - Prof. João Branco, Higher Technical Institute of Lisbon, Portugal
 - Prof. M. Lucília Carvalho, Faculty of Science, University of Lisbon, Portugal
 - Sir David Cox, University of Oxford, United Kingdom
 - Prof. Edwin Diday, University of Paris IX Dauphine, France
 - Prof. Dani Gamerman, Federal University of Rio de Janeiro, Brazil
 - Prof. Marie Husková, Charles University of Prague, Czech Republic
 - Prof. Isaac Meilijson, University of Tel Aviv, Israel
 - Prof. Nazaré Mendes-Lopes, Faculty of Science and Technology, University of Coimbra, Portugal
 - Prof. António Pacheco, Higher Technical Institute of Lisbon, Portugal
 - Prof. Dinis Pestana, Faculty of Science, University of Lisbon, Portugal
 - Prof. Ludger Rüschendorf, Institute of Stochastic Mathematics, Freiburg, Germany
 - Prof. Gilbert Saporta, National Conservatory of Arts and Crafts, Paris, France
 - Prof. Jef Teugels, Catholic University of Leuven, Belgium
 - Prof. Radu Theodorescu, University of Laval, Sainte-Foy, QC, Canada
- The articles must be written in English and may be submitted in three different ways:
 - By sending three paper copies to the Executive Editor who will be in permanent contact with the Editor-in-Chief and the Co-Editor,
 - By sending two paper copies to the Executive Editor and one copy to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account.
 - By sending a paper copy to the Executive Editor, together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.
- Manuscripts must be submitted using A4-size paper, double-spacing, and a large left margin and must not contain more than 30 pages. The first page must include the name, institution and address of the author(s) and a summary of not more than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification. The authors must write the final version of the articles using LaTeX, in REVSTAT style.

This style (REVSTAT.sty), examples of files (REVSTAT.tex) as well as additional instructions for the authors may be obtained, in principle, in September through the following INE Web address: <http://www.ine.pt/revstat>.

- The submission of an article means that it contains original work that has not been nor is about to be published elsewhere in any form.

Within this new context, **authors wishing to submit articles for publication in REVSTAT may begin to do so from 1 October 2002**.

Further information on this initiative, intended for the potential authors of articles, will be provided in the next issue of the Revista de Estatística, to be published during the first fortnight of September 2002.

A publicity campaign will be also developed including:

- (a) The publication of relevant information in the Bulletin of the Portuguese Statistical Society;
- (b) An e-mailing relevant information to members of the Portuguese Association of Classification and Data Analysis;
- (c) The creation of a promotional leaflet in English to be distributed at congresses and seminars;
- (d) An e-mailing relevant information in English to members of the Bernoulli Society of the International Statistical Institute;
- (e) The publication of relevant information in English in a specific entry of the REVSTAT in the INE's website.

ARTIGOS

VOLUME I

1º QUADRIMESTRE DE 2002

Será que as empresas PORTUGUESAS apresentam vantagem competitiva?

Autores:
Manuela Magalhães Hill
e
Andrew B. Hill, PhD.

VOLUME I

1º QUADRIMESTRE DE 2002

SERÁ QUE AS EMPRESAS PORTUGUESAS APRESENTAM VANTAGEM COMPETITIVA?¹

DO PORTUGUESE COMPANIES HAVE COMPETITIVE ADVANTAGES?

Autores: Manuela Magalhães Hill

- Professora Associada do Instituto Superior de Ciências do Trabalho e da Empresa

Andrew B. Hill, PhD.

- Investigador do DINAMIA/ISCTE

RESUMO:

- Este artigo procura investigar a relação entre competitividade empresarial e dois factores, os mercados onde as empresas operam, ou seja, se exportam os seus produtos ou não, e se estas vendem produtos com marca própria ou não. As empresas foram classificadas em quatro grupos: exportam/marca própria, não exportam/marca própria, exportam/não marca própria e não exportam/não marca própria. Os resultados revelam que cada grupo apresenta a sua estratégia competitiva específica.

PALAVRAS-CHAVE:

- *Competitividade, mercados, industria transformadora, áreas metropolitanas de Lisboa e Porto, MANOVA.*

ABSTRACT:

- This article reports on investigation into the relationship between company competitiveness and two factors – whether or not companies export their products and whether they sell their own brand. Companies were classified into one of four groups; export/own brand, no export/own brand, export/no own brand, no export/no own brand. Statistical analyses of variables found in a previous study related to competitiveness showed that the four groups used different strategies to achieve competitiveness.

KEY-WORDS:

- *Competitiveness, markets, manufacturing industry, metropolitan areas of Lisbon and Oporto, MANOVA.*

¹ O estudo aqui desenvolvido tem como ponto de partida a pesquisa realizada no DINÂMIA – Centro de Estudos sobre a Mudança Socioeconómica, do ISCTE, no âmbito de um projecto mais abrangente sobre *Competitividade e exclusão social nas áreas metropolitanas de Lisboa e Porto*, subsidiado pela FCT, programa PRAXIS, e levado a cabo por três instituições, DINÂMIA e CIES do ISCTE e Instituto de Sociologia da Faculdade de letras da Universidade do Porto.

VOLUME I

1º QUADRIMESTRE DE 2002

1. INTRODUÇÃO

As últimas décadas têm sido palco de profundas transformações nos sistemas económicos e uma rápida evolução na utilização de novas tecnologias sendo que as empresas têm de enfrentar um mercado cada vez mais global. De acordo com Perez (1988) estas transformações têm contidas «janelas de oportunidade» para os países menos desenvolvidos por traduzirem alterações nos processos tecnológicos e formas de conhecimento das empresas, sendo que os aspectos culturais e sociais são muito importantes para a adaptação do país a essas transformações.

O papel das tecnologias tem vindo a aumentar significativamente nas economias dos países mais desenvolvidos. Ao falar de novas tecnologias, nomeadamente de automação industrial, pensa-se em duas consequências à sua implementação, uma positiva e outra negativa: as novas tecnologias significam progresso, modernidade e elevação do nível de qualificação dos trabalhadores, mas esse progresso faz-se à custa do aumento de desemprego e da desqualificação de uma grande parte da mão-de-obra. Claro está que as maiores ou menores repercussões decorrem da natureza das próprias novas tecnologias.

Segundo Child (1985) a introdução de novas tecnologias tem subjacente quatro motivos ou intenções estratégicas: redução de custos (mão-de-obra ou stocks, por exemplo), maior flexibilidade (por exemplo, adaptação rápida a mudanças na produção), melhoria na qualidade do produto ou serviço (como por exemplo, ajustamento a especificações do cliente) e maior integração e controlo (nomeadamente por parte dos níveis superiores de gestão). A maior ou menor acentuação destes motivos poderá ter efeitos positivos ou negativos. A maior flexibilidade e a melhoria na qualidade do produto inserem-se claramente no tipo de estratégias que poderá originar aumento de desemprego e desqualificação da mão-de-obra.

Mas a boa aceitação da introdução de novas tecnologias vai depender, como refere Adler (1987), de maneira crucial, dos comportamentos dos trabalhadores, das suas capacidades para resolver problemas, da qualidade do trabalho de equipa e do desejo de aumentar as suas aptidões à medida que a automação evolui.

A competitividade de uma empresa está assim associada não só a factores-preço (custos de mão-de-obra, matérias primas, produtos, por exemplo) mas cada vez mais a factores não-preço (qualidade dos seus produtos e serviços). Segundo a OCDE (1992), é a eficácia das estruturas produtivas, através da qualidade dos serviços oferecidos às unidades de produção, da envolvente científica e tecnológica e da qualificação da mão-de-obra que explica a competitividade das empresas.

O presente artigo parte dos resultados apresentados em Hill (2001) e desenvolve uma nova perspectiva sobre a problemática da competitividade empresarial, tendo em atenção os mercados onde as empresas operam e se vendem produtos com marca própria ou não.

2. O MÉTODO

A amostra

O estudo teve como população alvo as empresas da indústria transformadora portuguesa com mais de 10 trabalhadores localizadas nas áreas metropolitanas de Lisboa e Porto e constantes da base BELEM (Base de estabelecimentos e empresas) de 1998 do INE. A amostra seleccionada é constituída por 678 empresas e representativa em termos de dimensão da empresa, sector de actividade (CAE a 2 dígitos) e área metropolitana.

Os dados

Os dados foram recolhidos em 1999 por meio de dois questionários: um deles contendo 85 questões referentes a seis áreas: caracterização da empresa (incluindo informação sobre o tipo de tecnologias utilizadas), actividades de I&D, gestão dos recursos humanos, organização da empresa e conteúdo do trabalho, e relações profissionais; o outro contendo 15 questões dirigidas especificamente ao empresário ou gestor da empresa.

O indicador de competitividade

- O indicador de competitividade aqui utilizado é o «cash flow em 1998 por trabalhador»².

Outros indicadores

- Indicador de inovação
- Indicador de Intensidade das Tecnologias de Informação
- Indicador de Mudança Organizacional
- Indicador de Mudança na Gestão dos Recursos Humanos

Estes indicadores são a primeira componente resultante da aplicação da análise de componentes principais não linear a variáveis medidas em escalas ordinais dicotómicas.

Os factores de competitividade

A aplicação de modelos de regressão logística ao conjunto das empresas da amostra permitiu seleccionar cinco factores de competitividade. Estes factores conseguiram classificar correctamente 72,8% das empresas em dois grupos (grupo das empresas mais competitivas e grupo das empresas menos competitivas), o valor de Kappa para a classificação global foi 0,648 ($P < 0,0001$) e o coeficiente de validade preditiva do modelo foi $\phi = 0,457$ ($P < 0,0001$) distinguir as empresas mais

² Ver justificação da sua utilização em Hill (2001)

competitivas das menos competitivas, e foram objecto de uma reflexão e análise em Hill (2001).

FACTORES DE COMPETITIVIDADE
Indicador de Intensidade das tecnologias de informação
Indicador de Inovação
Utilização da técnica de planeamento e gestão de produção MRP
Percentagem dos custos salariais em relação aos custos totais de exploração
A empresa costuma recorrer a outros serviços da universidade para além do recrutamento

No entanto, as análises sobre as quais os resultados se basearam não tiveram em atenção os mercados onde as empresas estavam a operar. Parece-nos assim razoável pensar que a natureza desses mercados, bem como a pressão das forças competitivas dentro deles, podem influenciar a estratégia utilizada pelas empresas para se tornarem mais competitivas.

É óbvio que, dada a diversidade de sectores na amostra, bem como a grande variação de actividades em cada um dos sectores, não é possível trabalhar com mercados muito específicos. Deste modo, qualquer tentativa para analisar estratégias de competitividade tendo em consideração mercados diversificados deve considerar o termo «mercados» em sentido lato. Talvez os dois mercados mais óbvios que se enquadram nesta restrição são o mercado interno e o mercado externo. Na nossa amostra, 53% das empresas exportam todos ou parte dos seus produtos e 47% vendem apenas para o mercado interno.

Mas onde, ou seja, o tipo de mercados onde a empresa coloca os seus produtos talvez não seja o único factor que influencie a estratégia adoptada pela empresa. O que a empresa vende pode também influenciar a sua estratégia. Contudo, a diversidade de produtos vendidos pelas empresas leva a que a distinção adoptada entre os diferentes tipos de produtos produzidos pelas empresas seja «vende marca própria/ não vende marca própria». Na nossa amostra, 54,9% das empresas vendem produtos com marca própria e 45,1% não.

Usando a definição de mercados e a distinção entre tipos de produtos indicada, fomos definir quatro grupos de empresas:

GRUPO 1. Não vende produtos com marca própria e não exporta.

GRUPO 2. Não vende produtos com marca própria mas exporta.

GRUPO 3. Vende produtos com marca própria e não exporta.

GRUPO 4. Não vende produtos com marca própria e exporta.

Analisemos então quais as características das empresas mais e menos competitivas em cada um dos grupos, para determinar se diferentes circunstâncias

comerciais sugerem diferentes aproximações estratégicas em termos de competitividade.

Na comparação dos quatro grupos usámos duas novas variáveis, para além de todas as variáveis encontradas como factores de competitividade:

- A variável dicotómica «a empresa realiza actividades de I&D» ou não.
- Uma combinação das duas variáveis referentes a aumentos salariais individualizados que apareceram como factores de competitividade específicos de cada área metropolitana: para a área metropolitana de Lisboa – (a) *por mérito /desempenho* e para a área metropolitana do Porto – (b) *tomando em consideração o contributo para os objectivos da empresa*. Estas variáveis foram combinadas para formar uma única variável «aumentos salariais individualizados (a ou b)».

Além disso, o indicador «inovação» foi substituído por uma nova variável mais específica «Inovação no produto nos 5 anos anteriores», com a seguinte escala com quatro níveis:

1. Não houve produtos novos nem melhoramento nos produtos.
2. Não houve produtos novos mas houve melhoramento nos produtos.
3. Houve produtos novos mas não houve melhoramento nos produtos.
4. Houve produtos novos e houve melhoramento nos produtos.

O indicador «intensidade das tecnologias de informação» não foi utilizado. Em vez disso, os 12 tipos de tecnologias de informação foram sujeitos a uma análise factorial por máxima verosimilhança, a qual produziu quatro factores. Os detalhes técnicos da análise factorial são apresentados de seguida. Só 4 das 66 correlações entre os 12 tipos de tecnologias de informação não apresentaram significância inferior a 0,05. O valor de KMO foi 0,790. Embora a bondade de ajustamento para a solução com quatro factores não pareça ser adequada em termos do teste convencional usado pela análise de máxima verosimilhança ($\chi^2 = 38,967$, g.l. = 24, $P = 0,028$), só 3% das correlações residuais foram superiores a 0,05, pelo que se concluiu que os quatro factores oferecem uma solução adequada. Os quatro factores explicam 41,3% da variância total das variáveis.

A solução factorial inicial foi sujeita tanto a rotação ortogonal como a rotação oblíqua (ver Quadros 1 e 2, onde se excluíram os pesos inferiores a 0,25). O Quadro 3 apresenta as correlações entre os factores na solução oblíqua.

QUADRO 1 – SOLUÇÃO VARIMAX

Rotated Factor Matrix^a

	Factor			
	1	2	3	4
Computadores pessoais				
computador				
médio/grande porte		,266	,328	
Componentes				
robotizados			,397	
Centros de maquinagem				,361
Equipamentos CNC				,911
Equipamento CAD	,430			,296
Equipamento CAM	,852			
Equipamento CAD/CAM	,588			
Armazém automatizado				
Gestão da produção				
assistida por				
computador		,280	,637	
Rede de comunicação				
entre postos de trabalho		,970		
Controlo automatizado de				
secções de fabris			,498	

Extraction Method: Maximum Likelihood.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

QUADRO 2 – SOLUÇÃO OBLIMIN

Pattern Matrix^a

	Factor			
	1	2	3	4
Computadores pessoais				
computador				,300
médio/grande porte				
Componentes				,371
robotizados				
Centros de maquinagem		,350		
Equipamentos CNC		,963		
Equipamento CAD			,381	
Equipamento CAM			,942	
Equipamento CAD/CAM			,618	
Armazém automatizado				,270
Gestão da produção				
assistida por				
computador				,619
Rede de comunicação				
entre postos de trabalho	,996			
Controlo automatizado de				
secções de fabris				,509

Extraction Method: Maximum Likelihood.

Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 9 iterations.

QUADRO 3 – CORRELAÇÕES ENTRE OS FACTORES

Factor Correlation Matrix

Factor	1	2	3	4
1	1,000	4,542E-02	,289	,383
2	4,542E-02	1,000	,449	,305
3	,289	,449	1,000	,429
4	,383	,305	,429	1,000

Extraction Method: Maximum Likelihood.

Rotation Method: Oblimin with Kaiser Normalization.

A solução varimax não atribuiu peso apreciável à variável «Armazém automatizado» pelo que, tendo também em consideração que se espera que os factores de tecnologia de informação possam estar correlacionados, se optou pela solução oblíqua e se calcularam os valores desta solução para cada empresa.

Vamos designar os quatro factores oblimin por INFOTEC(1), INFOTEC(2), INFOTEC(3) e INFOTEC(4).

Para analisar a associação entre competitividade e as variáveis não métricas foi utilizado o teste do qui-quadrado e para comparar os grupos das empresas mais e menos competitivas usou-se a MANOVA. Cada uma das variáveis que define cada um dos factores INFOTEC(1), (2), (3) e (4) foi também objecto da análise por meio do qui-quadrado para fornecer informação adicional sobre a relação entre tecnologias de informação e competitividade. Estas análises foram efectuadas separadamente para cada um dos quatro grupos de empresas.

3. RESULTADOS

O Quadro 4 resume os resultados obtidos. Em cada um dos quatro grupos, não se encontram diferenças significativas entre empresas mais e menos competitivas em termos de dimensão.

Vejamos quais as características das empresas mais competitivas em cada um dos grupos.

As células em branco no Quadro 4 indicam que as empresas mais competitivas não diferem significativamente das menos competitivas nas variáveis que definem as células. As células que se encontram marcadas a cinzento indicam que existe diferença significativa mas fraca ($0,05 < P < 0,10$) entre as empresas mais e menos competitivas. As células que contêm valores numéricos mas sem o cinzento, indicam que a diferença é significativa ($P < 0,05$).

GRUPO 1. Não vende produtos com marca própria e não exporta

Estas são em geral empresas de pequena ou média dimensão, sendo que aproximadamente dois terços (65,7%) pertencem a sete sectores (CAE_2 – sectores 15, 17, 20, 22, 26, 28, 31). Muito embora seja de esperar que sejam as empresas mais competitivas as que realizam mais actividades de I&D e desenvolvem mais inovação no produto, isto não aparece fortemente relacionado com competitividade. A maior parte das empresas deste grupo não possuem tecnologias de informação, muito embora a posse de rede de comunicação entre postos de trabalho e a posse de equipamento CAD surjam associadas com competitividade. A competitividade não está associada com a automação da produção mas apenas com novas formas de planeamento, controlo e gestão da produção (possivelmente ajudada pela rede de comunicação entre postos de trabalho). Neste grupo, a competitividade está bastante menos fortemente associada com uma política de custos salariais reduzidos em relação aos custos totais de exploração do que nos outros três grupos.

GRUPO 2. Não vende produtos com marca própria mas exporta

Em média, as empresas deste grupo são de média ou grande dimensão. Dois terços (68,3%) das empresas pertencem a nove sectores (CAE_2 – sectores 15, 17, 18, 24, 26, 27, 28, 34, 36). As empresas mais competitivas são, de certo modo, semelhantes às suas congéneres do grupo 1, com um maior cometimento com actividades de I&D e uma aproximação mais sofisticada ao planeamento, controlo e gestão da produção, incluindo o uso da técnica MRP. Com excepção de equipamento para controlo automatizado das secções fabris, que é utilizado por pouco de menos de um quinto das empresas mais competitivas, as tecnologias de informação não estão significativamente relacionadas com competitividade. Em contraste, tem-se que baixos custos salariais/custos totais estão fortemente relacionados com competitividade.

GRUPO 3. Vende produtos com marca própria e não exporta

As empresas deste grupo são, em geral, de pequena ou média dimensão. 70% das empresas pertencem a oito sectores (CAE_2 – sectores 15, 22, 24, 26, 29, 30, 33, 36). Embora estas empresas não difiram das do grupo 1 em termos de dimensão, o facto de venderem produtos com marca própria para o mercado interno parece estar associado com uma mais moderna aproximação à competitividade. É de estranhar que não se verifique um maior cometimento com actividades de I&D (aproximadamente 20% das empresas, tanto muito como pouco competitivas, desenvolvem actividades de I&D).

Ao contrário do que acontece com as suas congéneres nos outros grupos, as empresas mais competitivas têm maior tendência para usar aumentos salariais individuais baseados no mérito/desempenho ou na contribuição para os objectivos da empresa.

QUADRO 4

B = BAIXA competitividade A = ALTA competitividade

VARIÁVEL	Marca = 0 Export = 0		Marca = 0 Export = 1		Marca = 1 Export = 0		Marca = 1 Export = 1	
	B	A	B	A	B	A	B	A
Dimensão da empresa (Nº de trabalhadores) Valor médio	72	61	148	149	73	79	227	335
I & D	2,8	11,8	9,1	36,8			36,1	49,5
Aumentos salariais individualizados (mérito/desempenho ou contributo p/ objectivos)					47,2	68,9		
INOVAÇÃO NO PRODUTO								
Produtos melhorados								
Produtos novos								
Novos & melhorados							33,3	42,2
Técnica de planeamento das necessidades da produção MRP			15,9	31,6			22,2	53,2
Novos procedimentos de planeamento, controlo e gestão da produção	18,7	38,2	31,8	50,0			34,7	51,4
INFOTEC (1)	A > B				A > B		A > B	
Rede de comunicação entre postos de trabalho	16,9	44,1			30,6	57,8	51,4	70,6
INFOTEC (2)								
Centro de maquinaria CNC					2,8	13,3	8,3	17,4
INFOTEC (3).							A > B	
CAD		17,6					25,0	34,9
CAM,	4,2							
CAD/CAM							5,6	18,3
INFOTEC (4).					A > B		A > B	
Computador médio /grande porte,					33,3	53,3	50,0	70,6
Componentes robotizados							12,5	27,5
Armazém automatizado					2,8	17,8		
Gestão assistida por computador							33,3	56,9
controlo automatizado de secções fabris			4,5	18,4	0,0	13,3	12,5	33,9
% CUSTOS SALARIAIS / CUSTOS TOTAIS (Valor médio)	38,7	29,6	34,6	22,0	34,2	21,5	34,4	20,7

Isto quase nos leva a crer que, em vez de desenvolverem actividades de I&D, as empresas usam este tipo de aumentos

salariais como um meio de, através da motivação dos trabalhadores, encontrar novas ideias para a inovação do produto. A competitividade parece basear-se numa maior utilização das tecnologias de informação, em particular, computadores de grande porte e redes de comunicação entre postos de trabalho. Mas a percentagem de empresas que usam formas mais sofisticadas de tecnologias de informação associadas com a automação e o controlo automático da produção é em geral pequena. Embora neste grupo a competitividade esteja ligada com a posse de tecnologias de informação mais sofisticadas, parece que maior competitividade não se baseia firmemente numa melhoria na eficiência na produção (nem na inovação no produto). Novos processos de planeamento, controlo e gestão da produção, e o uso da técnica MRP, não revelam estarem associadas significativamente com competitividade, no entanto baixos custos salariais/custos totais sim.

GRUPO 4. Não vende produtos com marca própria e exporta

Em média, as empresas deste grupo são empresas de grande dimensão, embora não haja relação entre dimensão e competitividade. 70% destas empresas pertencem a sete sectores (CAE - sectores 15, 17, 24, 28, 29, 31, 36). A competitividade das empresas baseia-se em I&D, inovação no produto, eficiência na produção através do uso da técnica MRP e em novos procedimentos de planeamento. Nas empresas mais competitivas nota-se a presença significativa de vários tipos de tecnologias de informação: equipamentos especializados de utilização flexível tais como CAD e CAD/CAM, equipamento associado com automação ou controlo automático da produção, computadores de grande porte e rede de comunicação entre postos de trabalho. Tal como nos outros grupos, as empresas competitivas também apresentam rácios custos salariais/custos totais significativamente mais baixos.

4. CONCLUSÃO

Tem-se assim que a estratégia competitiva das empresas que não vendem produtos com marca própria e não exportam se baseia mais na eficiência da produção que num investimento forte em I&D ou na inovação do produto. Isto é bastante compreensível pois é provável que o vender para o mercado interno produtos cujas marcas pertencem a outras empresas não traz grande incentivo para o desenvolvimento de actividades de I&D ou para tentar inovar o produto. No entanto, a estratégia competitiva parece ser relativamente «tradicional» pois faz pouco uso de tecnologia de informação mais sofisticada para melhorar a eficiência na produção, não

se observando um uso significativo da técnica MRP no planeamento das necessidades de produção.

A estratégia das empresas que exportam mas não vendem produtos com marca própria baseia-se em dois pontos:

1. Baixos custos de mão de obra em Portugal. (O facto de comparativamente serem poucas as empresas que dispõem de tecnologia sofisticada sugere que o investimento em tecnologia de informação não é responsável pelos baixos rácios custos salariais/custos totais verificados).
2. Eficiência na produção – sem, contudo, haver um uso apreciável de tecnologia de informação.

Contrariamente ao que acontece com as empresas do grupo 4, que têm de proteger os seus próprios produtos nos mercados internacionais, as empresas do grupo 2 parecem não estar a competir em termos de inovação no produto, embora mostrem um certo cometimento com as actividades de I&D e possam muito bem estar numa fase de transição para competir através da inovação no produto. Tal transição é importante porque é improvável que a actual estratégia possa ser mantida no médio ou longo prazo. Primeiro, porque o provável alargamento da U.E. aos países de leste onde a mão de obra é mais barata, pode fazer desaparecer a actual vantagem concorrencial em termos custos da mão de obra. Em segundo lugar, é natural que, sem uma transição para um maior uso das tecnologias de informação para melhorar a eficiência da produção, a actual eficiência tenda gradualmente a desaparecer.

O grupo 3 é o mais estranho e o mais difícil de entender. Por um lado, parece que as empresas que vendem produtos com marca própria no mercado interno ainda não formularam uma estratégia coerente para promover competitividade. As mais competitivas já começaram a usar as tecnologias de informação associadas com a automação e o controlo automático da produção, embora em pequena escala, mas não desenvolvem mais actividades de I&D que as menos competitivas, não apresentam inovação no produto significativa e parecem não basear a competitividade numa produção mais eficiente. Como o rácio custos salariais/custos totais é em média muito mais baixo nas empresas mais competitivas, somos levados a concluir que a competitividade das empresas que vendem as suas próprias marcas exclusivamente no mercado interno pode ser conseguida através de baixos preços dos produtos.

As empresas competitivas que vendem produtos com marca própria e exportam, assemelham-se mais às dos países mais avançados tecnologicamente. Talvez porque necessitem de proteger os seus produtos da competição a que são sujeitos nos mercados internacionais, estas empresas parecem ter adoptado uma estratégia moderna, multifacetada e integrada para poderem ser competitivas, a qual se baseia nas actividades de I&D, na inovação no produto e na eficiência na produção. Estas empresas podem ter também a vantagem de terem baixos custos de mão de obra, embora não seja claro que seja por esta razão que têm baixo rácio custos salariais/custos totais. É possível que os custos totais destas empresas sejam muito maiores como resultado do investimento em I&D e em tecnologias de informação.

BIBLIOGRAFIA

- ADLER, P. S. (1987) "Automation et qualifications. Nouvelle orientation" *Sociologie du travail*, nº 3, p. 299.
- CHILD, J. (1985) *Organization – a Guide to Problems and Practice*. Harper and Row, London.
- HERZBERG, F. MAUSNER, B. & SNYDERMAN, B. (1959) *The Motivation to Work*. John Wiley & Sons, New York.
- HILL, M. M. (2001) "Determinantes da Competitividade Empresarial nas Áreas metropolitanas de Lisboa e Porto". *Revista de Estatística*. INE, vol 1, pp. 7-25.
- OCDE (1992) *La Technologie et l'Économie – les Relations Déterminantes*. Paris. OCDE.
- PEREZ, C. (1988) "New technologies and development" em FREEMAN, C. e LUNVALL, B. *Small Countries Facing Technological Revolution*. Londres, Printer Publishers.
- PFEFFER, J. (1998) "Six Dangerous Myths about pay", *Harvard Business Review*, Maio-Junho, pp. 109-119.
- RAPPAPORT, A. (1999) "New Thinking on how to Link Executive Pay with Performance", *Harvard Business Review*, Março-Abril, pp. 91-101.
- SALAVISA LANÇA, I. S. (2000) *A Indústria Portuguesa. Especialização Internacional e Competitividade*. Lisboa, Celta.
- SIMON, H. (1991) Organizations and markets. *Journal of Economic Perspectives*, vol. 5, Spring, 25 – 44.
- GHOSHAL, S., BARTLETT, C. & MORAN, P. (2000) Um novo Manifesto para a gestão. *Revista Portuguesa de Gestão*, Inverno, 15, 1, Inverno, 18 – 29.
- SALAVISA LANÇA, I. & Hill, A.B. (1999) Patterns of technological innovation in intermediate countries – An alternative index and some conclusions. Paper presented at the EAEPE Conference, Prague, Czech Republic.

VOLUME I

1º QUADRIMESTRE DE 2002

Quasi-random versus pseudo-random numbers to approximate posterior distributions

Autores:
Carlos Javier Pérez Sánchez
Jacinto Martín Jiménez
e
José Carlos Rojano Martín

QUASI-RANDOM VERSUS PSEUDO-RANDOM NUMBERS TO APPROXIMATE POSTERIOR DISTRIBUTIONS

NÚMEROS QUASI-ALEATÓRIOS VERSUS PSEUDO-ALEATÓRIOS PARA ESTIMAR DISTRIBUIÇÕES POSTERIORES

Autores: Carlos Javier Pérez Sánchez

- Assistant Professor - University of Extremadura - Faculty of Veterinary -
Department of Mathematics

Jacinto Martín Jiménez

- Associate Professor - University of Extremadura - Technical School -
Department of Mathematics

José Carlos Rojano Martín

- Associate Professor - University of Málaga - Faculty of Sciences -
Department of Statistics

ABSTRACT:

- The main computational problem in Bayesian Inference is computing the posterior distribution. More concretely, computing expectations of functions with respect to the posterior distributions.

In this work we review some stochastic simulation methods for one-dimensional posterior distributions. We propose adapting these methods by using quasi-random numbers as a way to obtain better approximations to the target distribution and therefore an improvement in the estimations of posterior characteristics. We compare with the pseudo-random case.

Finally, we illustrate by an example the extension to multidimensional problems.

KEY-WORDS:

- *Bayesian Inference, Pseudo-random numbers, Quasi-random numbers, Simulation, F-discrepancy, Low-discrepancy point sets and sequences, Inverse transform method, Weighted Bootstrap, Ratio-of-uniform method.*

RESUMO:

- O principal problema computacional na Inferência Bayesiana é o cálculo da distribuição posterior. Mais concreto, o cálculo de expectativas de funções com respeito às distribuições posteriores.

Neste trabalho nós revisamos alguns métodos estatísticos de simulação para distribuições posteriores unidimensionais. Nós propomos a adaptação destes métodos mediante o uso de números quasi-aleatórios como forma de obter melhores aproximações à distribuição objetivo e, consequentemente, uma melhora

nas estimações das características posteriores. Nós realizamos comparações com o caso pseudo-aleatório.

Por último, nós ilustramos com um exemplo a extensão a problemas multidimensionais.

PALAVRAS-CHAVE:

- *Inferência Bayesiana, números pseudo-aleatórios, números quasi-aleatórios, Simulação, F-discrepância, Sucessões e conjuntos de pontos de baixa discrepância, método da transformação inversa, Bootstrap ponderado, método do quociente de uniformes.*

1. INTRODUCTION

One of the most important problems in Bayesian analysis is computing the posterior distributions. As Berger (2000) points out, most Bayesian computations are focused on the calculation of posterior expectations, which are mainly integrals. Let π be a prior distribution, l the likelihood function, and x the observation from an experiment. Then, the posterior distribution is:

$$(1) \quad \pi(\theta | x) = \frac{l(\theta | x)\pi(\theta)}{\int_{\Theta} l(\theta | x)\pi(\theta)d\theta} = \frac{l(\theta | x)\pi(\theta)}{m(x)}$$

and there are several methods to compute $I_g = \int_{\Theta} g(\theta)\pi(\theta|x)d\theta$ for some functions $g(\theta)$.

The main problem is that I_g is a quotient of integrals. If we try to approximate both of them, we can not bound the error (Evans and Swartz, 2000). This problem has been solved by using the well-known Markov Chain Monte Carlo (MCMC) methods, see Tierney (1994). The basic idea of these methods lies in building (under regularity conditions) a Markov chain whose stationary (or invariant) distribution is the target posterior distribution. Unfortunately, we have two sources of errors: the one committed by the discretization to estimate the integral and the error based on the approximation to the posterior distribution. However, very often, these methods have been proved to be useful and they have been fundamental in the application and development of bayesian statistical models.

In this work we try to apply and introduce different methods to solve the problem. We are interested in generating directly from the posterior distribution. We consider using quasi-Monte Carlo (QMC) methods to reduce the approximation error in the integral. The main disadvantage of QMC is the dependence among the generated values which has prevented a systematic use in MCMC. Liao (1998) developed a quasi-random Gibbs algorithm in which a randomly permuted quasi-random sequence is used in place of a sequence of pseudo-random deviates. This algorithm was proved about 4-25 times as efficient as the pseudo-random algorithm. No rigorous theoretical justification has been found. We will use generation methods not influenced by this fact.

We review some stochastic simulation methods for one-dimensional posterior distributions. Furthermore, we improve some of the methods. Section 2 introduces QMC methods. Section 3 considers two simulation methods, inverse transform method and weighted Bootstrap. In Section 4 we adapt the ratio of uniform method and illustrate the differences between non-uniform pseudo random and quasi random numbers. Finally, we use QMC method to study a multidimensional integration problem.

2. QUASI-MONTE CARLO METHODS

Monte Carlo (MC) methods are statistical sampling techniques that have been extensively applied to approximate integrals and other purposes (Robert and Casella, 1999). QMC methods can be considered as deterministic versions of those. In the integration problem, an advantage of QMC methods is that they provide deterministic error bounds, as given by the Koksma-Hlawka inequality, with a better order than the Monte Carlo's, which is probabilistic (see Niederreiter, 1992). QMC methods emphasise "uniformity" of the points, instead of "randomness" of them. The concept of discrepancy, which measures the uniformity of a set of points, is crucial to these methods. The so-called low-discrepancy point sets and sequences are used. They are known as quasi-random point sets and sequences and their elements quasi-random numbers. The more general concept of F -discrepancy, a measure of the representation of a set of points with respect to a distribution F , was suggested by Wang and Fang (1990).

In the one-dimensional case we can find a set of points with the lowest discrepancy. In multi-dimensional settings we have to choose among a wide range of low-discrepancy point sets and sequences available. Empirical comparisons of various low-discrepancy point sets and sequences that have been reported in the literature are somewhat inconclusive. Fox (1986), Bratley and Fox (1988) and Bratley, Fox and Niederreiter (1994) compared empirically accuracy and speed of various low-discrepancy sequence generators.

We compare the results obtained by using pseudo-random and different kinds of quasi-random numbers applying different methods. The low-discrepancy point sets are used to compute non-uniform quasi-random values, which yield better approximations to the target distribution: the posterior distribution in the Bayesian case. These improved approximations are reflected by the attainment of better estimations of posterior expectations.

3. TWO ADAPTED STOCHASTIC SIMULATION METHODS

Traditionally, the *inverse transform method* has been used to generate random variates from non-uniform distributions. Fang and Wang (1994) use this method to obtain representative points of probability distributions. This method is only applied in particular cases because we must know the normalisation constant $m(x)$ and the inverse function of the distribution function. However, it has the desirable property of maintaining the F -discrepancy through the transformation. If we choose the points uniformly scattered in $[0,1]$ with the lowest discrepancy, this property allows us to ensure that the values obtained have the lowest F -discrepancy.

Example 1: Beta(1,3) distribution. This is an easy example which shows us how low-discrepancy point sets can be used with inverse transform method to produce better estimations of the mean. The mean value is 0.25. In Figure 1 we can see the histograms of the 1000 values generated by this method using pseudo-random and quasi-random numbers. The density is also represented for convenience.

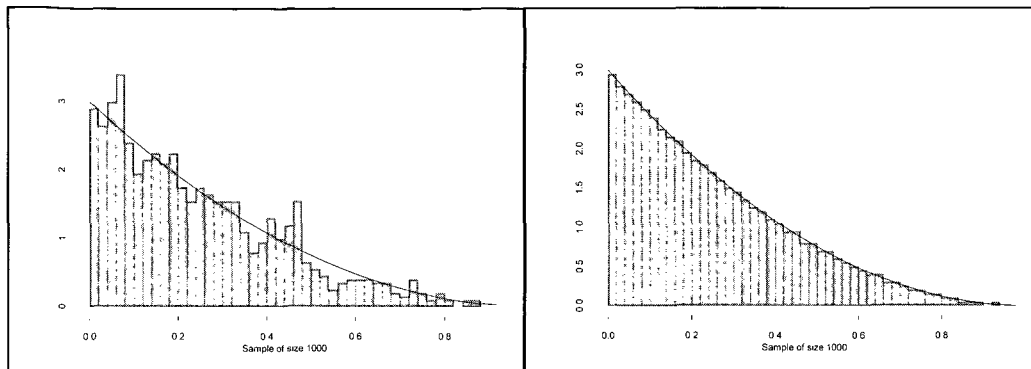


FIGURE 1. HISTOGRAMS FOR PSEUDO-RANDOM (LEFT) AND QUASI-RANDOM (RIGHT) NUMBERS.

In Figure 2 we can see the estimations of the mean for several sample sizes.

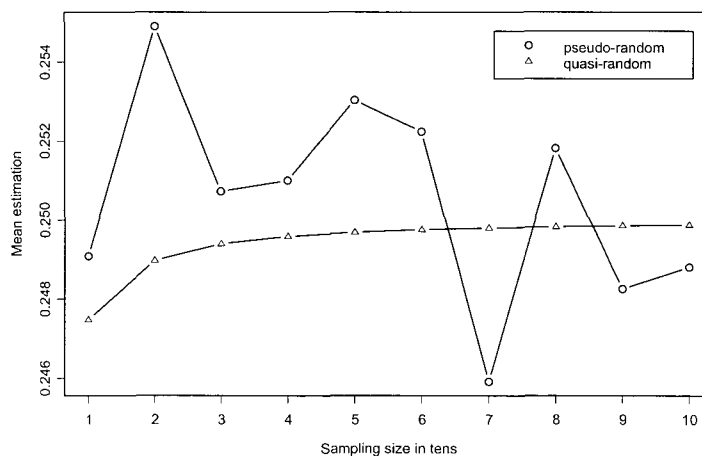


FIGURE 2. ESTIMATIONS OF THE MEAN FOR DIFFERENT SAMPLE SIZES.

We can see that quasi-random approximation is more stable and yields better estimations of the target characteristic. Similar results are obtained for other distributions as Weibull, Cauchy and Pareto.

When we apply this method directly to posterior distributions we can estimate characteristics of them, for example, the posterior means. However, we need an expression for the posterior distribution.

Another method we consider is the *weighted Bootstrap*. It was introduced by Smith and Gelfand (1992). This method is similar to the Sampling Importance Resampling (SIR) algorithm (Rubin, 1988). It is an approximation algorithm that is useful when we can not calculate the inverse of the distribution function in order to apply the preceding procedure or when the bound M in the rejection method is not readily available.

This technique requires sampling from a discrete distribution (with N points) with weights. The variates obtained by a particular weighted random sampling are approximately distributed according to the target distribution with the approximation “improving” as N increases. In order to build the discrete distribution we propose using quasi-random numbers.

When applying this method in order to generate from posterior distributions we don’t need to compute $m(x)$ in (1). Remember that this is a very important fact because in many problems it is impossible to compute analytically $m(x)$. We illustrate the idea of using QMC with an example:

Example 2: Consider a prior $Beta(1,10)$. Assume that the likelihood is $Binomial(20,\theta)$. Then we know that the posterior distribution is $Beta(x+1, 20+10-x)$. Assume also that we have observed the experiment and the result is $x=2$. So the posterior mean is $3/31$.

We apply this method with $N=500$ and different sample sizes. As a weighted random sampling is involved, the improvement in only one estimation can not be ensured. So, for a fixed sampling size, we consider 1000 estimations and compute the mean of the absolute errors.

In Figure 3 we can see them for different sample sizes. We can observe that the means of the absolute errors obtained by using quasi-random numbers (Δ) are lower than those of pseudo-random’s (o). Changing the scale allows us a better visualisation of the pseudo-random case and we can see the non-stability of the means of the absolute errors (Figure 4).

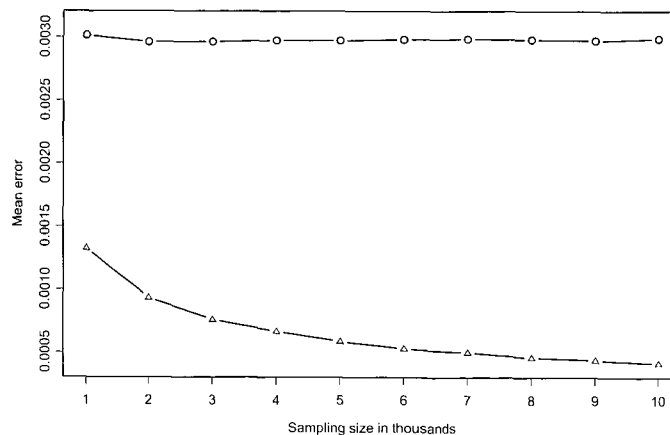


FIGURE 3. MEANS OF THE ABSOLUTE ERRORS IN 1000 ESTIMATIONS.

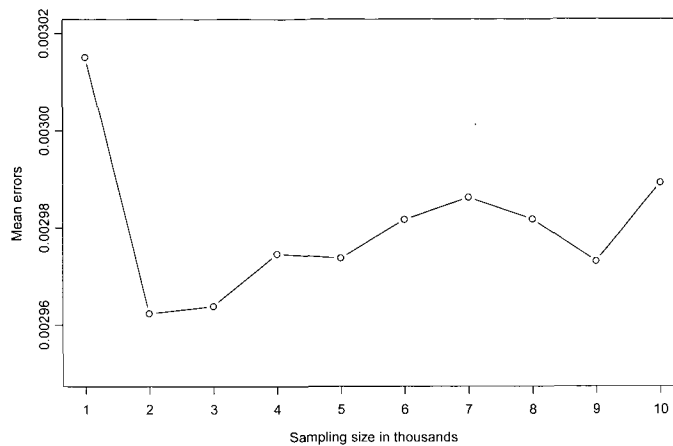


FIGURE 4. PSEUDO-RANDOM CASE CHANGING THE SCALE.

Figure 5 shows the histograms of the means of the absolute errors for a sample of size 5.000. We see a better behaviour of the adapted method in relation to the original.

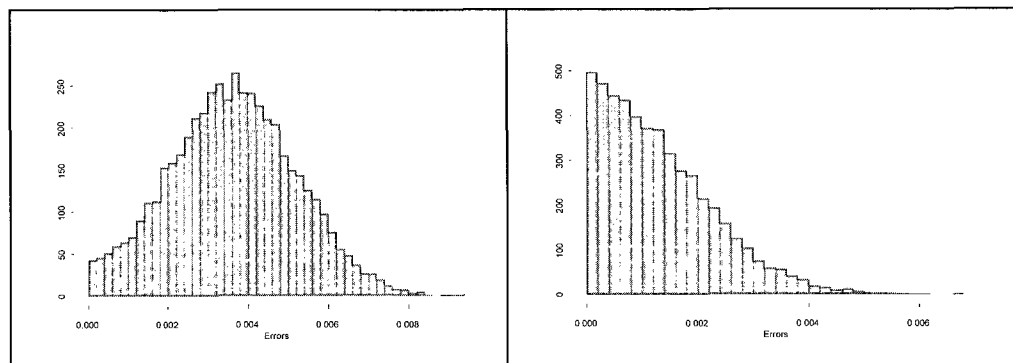


FIGURE 5. HISTOGRAMS OF THE ERRORS USING PSEUDO-RANDOM (LEFT) AND QUASI-RANDOM NUMBERS (RIGHT).

4. RATIO-OF-UNIFORM DEVIATES METHOD

The *ratio of uniform deviates* method is based on the following result from Kinderman and Monahan (1977):

Theorem: Let $C_h = \{(u, v): 0 \leq u \leq (h(v/u))^{1/2}\}$ for any nonnegative function h such that $0 < \int h < \infty$. Then C_h has finite volume and if we can generate (U, V) uniformly over C_h , then $X = V/U$ has density function $f = h / \int h$.

Barbu (1982, 1987), Vaduva (1985), Stefanescu and Vaduva (1987) and Wakefield et al (1991) improved the method by using suitable transformations of the ratio in order to increase the algorithm efficiency. Generalisations to multidimensional case have been proposed too. For a complete review see Barabesi (1993).

This method yields simple algorithms easy to understand and computationally efficient for a wide range of distributions. It is very useful in Bayesian Inference because we do not need to know the normalisation constant.

In order to apply this method we have to generate uniformly over the region C_h . Except in particular cases, we have to use accept-reject techniques to obtain the values required. Sometimes the probability of acceptance is very low. This fact forces us to apply techniques, when possible, to allow an increase in efficiency. We propose generating uniformly using a method based on adaptative sampling (Thompson, 1992).

We also propose using quasi-random numbers in order to generate uniformly on C_h . In examples 3 and 4 we compare the results obtained using pseudo-random numbers and different types of quasi-random numbers. We can see the superiority of the latter ones in terms of F -discrepancy and errors of the estimations.

Example 3: Cauchy distribution. In this problem, the target region to generate points uniformly is the unit circle.

In order to illustrate the method we use Good Lattice Points (GLP), Halton sequences (Halton), Hammersley sets (Hammersley) and pseudo-random numbers (Pseudo). The sample sizes are chosen so that the GLP sets are uniformly scattered. Comparison with arbitrary sample sizes can not be done for GLP.

Table 1 shows the F -discrepancy of the values obtained using this method for different sample sizes and sets. The best F -discrepancies are in bold type.

Sample sizes	GLP	Halton	Hammersley	Pseudo
55	0.0276	0.0320	0.0409	0.1058
144	0.0209	0.0121	0.0148	0.0731
610	0.0033	0.0048	0.0056	0.0327
987	0.0015	0.0024	0.0041	0.0336
1597	0.0009	0.0016	0.0025	0.0213
6765	0.0002	0.0004	0.0009	0.0144
17711	0.0001	0.0002	0.0004	0.0062

TABLE 1. **F-DISCREPANCY.**

The left side of Figure 6 shows, from top to bottom, the graphs of 987 points generated in the unit circle using GLP, Halton, Hammersley and Pseudo-random numbers, respectively. Close to each of them we can see the histograms obtained by using the ratio of uniform method. The range in the histograms has been established

between -10 and 10 for a better visualisation. The GLP set is so uniform that the values generated by this method are symmetrical.

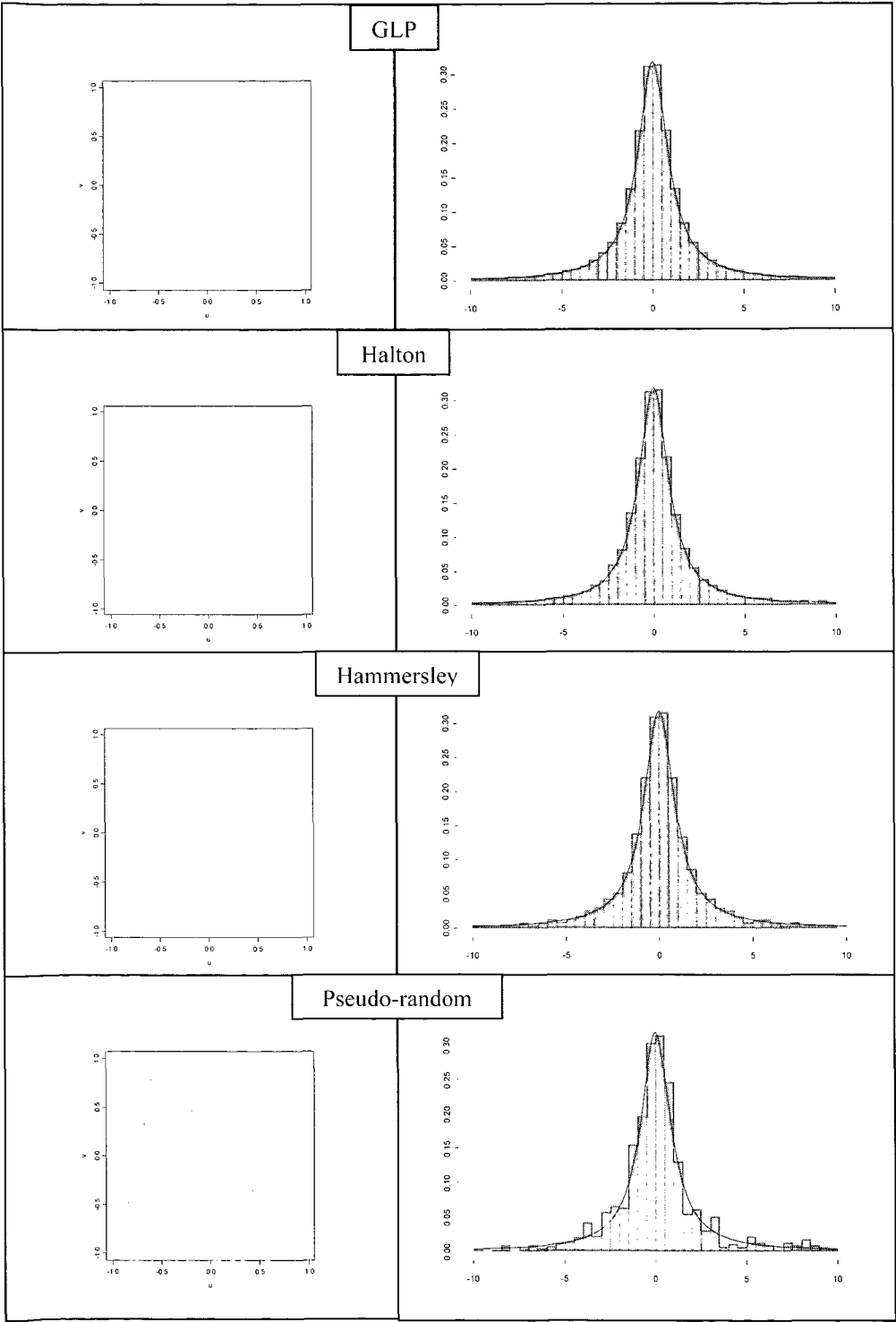


FIGURE 6. GENERATED POINTS (LEFT) AND HISTOGRAMS (RIGHT)

Example 4: We come back to example 2. Analogous results for the actual region and the histograms are obtained. Now, we focus on the estimation of the posterior mean.

In order to compare results, we use 100 estimations for each sample size in the pseudo-random case and consider the mean of the absolute errors. For the quasi-random case, we have only one possible finite sequence. Table 2 shows the punctual estimations and the confidence intervals for several sample sizes.

Sample size	Punctual estimations	Confidence intervals
987	$1.86 \cdot 10^{-3}$	$(1.62 \cdot 10^{-3}, 2.09 \cdot 10^{-3})$
1597	$1.39 \cdot 10^{-3}$	$(1.20 \cdot 10^{-3}, 1.58 \cdot 10^{-3})$
2584	$9.72 \cdot 10^{-4}$	$(8.46 \cdot 10^{-4}, 1.10 \cdot 10^{-3})$
4181	$8.32 \cdot 10^{-4}$	$(7.41 \cdot 10^{-4}, 9.24 \cdot 10^{-4})$
6765	$6.93 \cdot 10^{-4}$	$(6.02 \cdot 10^{-4}, 7.84 \cdot 10^{-4})$
10946	$4.92 \cdot 10^{-4}$	$(4.35 \cdot 10^{-4}, 5.50 \cdot 10^{-4})$
17711	$4.41 \cdot 10^{-4}$	$(3.88 \cdot 10^{-4}, 4.94 \cdot 10^{-4})$

TABLE 2. PUNCTUAL ESTIMATIONS AND CONFIDENCE INTERVALS.

Figure 7 shows the errors obtained for GLP, Halton and Hammersly. In order to get a good visualisation we have eliminated the results obtained from the pseudo-random case. The errors on Figure 7 can be compared with the ones in table 2. We can see that the errors obtained using quasi-random numbers are lower than the ones obtained for pseudo-random numbers. In general, the accuracy depends on the kinds of quasi-random numbers we use.

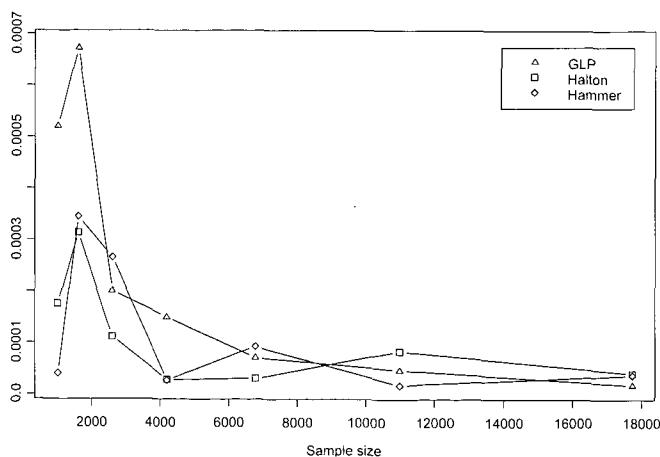


FIGURE 7. ERRORS USING QUASI-RANDOM NUMBERS.

5. A MULTIDIMENSIONAL INTEGRATION EXAMPLE

We expose the use of quasi-random numbers in the following multidimensional problem. This is an easy independent variable problem. However, we don't generate quasi-random values for each variable independently, because this would yield the same values for each variable.

Example 5: A multidimensional integral. Suppose that the posterior distribution is the random vector $\mathbf{x} \sim N_8(\mathbf{0}, \mathbf{I}_8)$, the standard normal distribution in R^8 . We want to evaluate the probability of $[0,1]^8$, i.e. $g = I_{[0,1]^8}$:

$$p = \int_0^1 \int_0^1 \dots \int_0^1 (2\pi)^{-4} \exp\left\{(-1/2)(x_1^2 + x_2^2 + \dots + x_8^2)\right\} dx_1 dx_2 \dots dx_8$$

We estimate p (true valor with 14 decimal positions is 0.00018430867518) by:

$$\hat{p} = \frac{1}{n} \sum_{k=1}^n f(\mathbf{x}_k)$$

where $\{\mathbf{x}_k, k=1, \dots, n\}$ is a set of points in $[0,1]^8$.

Figures 8 and 9 represent the absolute errors in the estimation of p for different point sets and sample sizes.

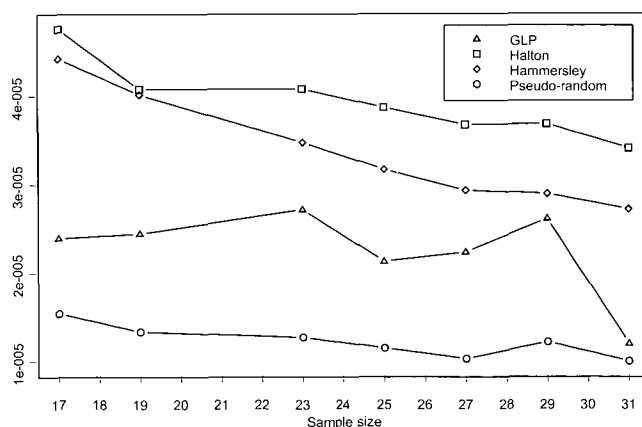


FIGURE 8. VERY SMALL SAMPLE SIZES.

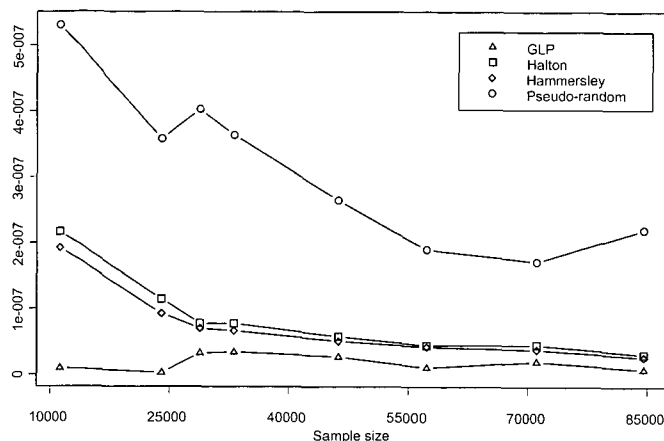


FIGURE 9. MODERATE SAMPLE SIZES.

Table 3 shows the confidence intervals for the mean of the absolute errors for the pseudo-random case.

Sample size	Confidence Intervals	Sample size	Confidence Intervals
17	$(1.32 \cdot 10^{-5}, 1.76 \cdot 10^{-5})$	11215	$(4.44 \cdot 10^{-7}, 1.22 \cdot 10^{-6})$
19	$(1.11 \cdot 10^{-5}, 1.58 \cdot 10^{-5})$	24041	$(1.81 \cdot 10^{-7}, 6.18 \cdot 10^{-7})$
23	$(1.09 \cdot 10^{-5}, 1.46 \cdot 10^{-5})$	28832	$(1.48 \cdot 10^{-7}, 5.36 \cdot 10^{-7})$
25	$(9.75 \cdot 10^{-6}, 1.33 \cdot 10^{-5})$	33139	$(2.11 \cdot 10^{-7}, 5.16 \cdot 10^{-7})$
27	$(8.70 \cdot 10^{-6}, 1.19 \cdot 10^{-5})$	46213	$(9.79 \cdot 10^{-8}, 4.32 \cdot 10^{-7})$
29	$(1.03 \cdot 10^{-5}, 1.41 \cdot 10^{-5})$	57091	$(4.47 \cdot 10^{-8}, 3.36 \cdot 10^{-7})$
31	$(8.49 \cdot 10^{-6}, 1.16 \cdot 10^{-5})$	71053	$(8.52 \cdot 10^{-8}, 2.58 \cdot 10^{-7})$
		84523	$(1.40 \cdot 10^{-8}, 3.02 \cdot 10^{-7})$

TABLE 3. CONFIDENCE INTERVALS

When the sample size is as small as the ones used in this problem, the results obtained using pseudo-random numbers (using the mean of the absolute errors) are better than the results obtained using quasi-random numbers. The opposite occurs when we have moderate sample sizes. For the quasi-random numbers, GLP produce the best estimations for the sample sizes studied followed by Hammersley and finally the Halton sets. When the sample size is high the results are in the same direction. For example, when we use 957.838 points the errors are $3.81 \cdot 10^{-10}$ for GLP, $3.91 \cdot 10^{-9}$ for Halton, $3.22 \cdot 10^{-9}$ for Hammersley and $9.42 \cdot 10^{-8}$ for the mean of the absolute errors in 100 estimations using pseudo-random numbers. For this sample size the confidence interval is $(4.47 \cdot 10^{-8}, 1.44 \cdot 10^{-7})$. When the dimension is higher other kinds of low-discrepancy sets and sequences such as Faure's, Niederreiter's or Sobol's must be used (Niederreiter (1992), Fox (1986) and Bratley and Fox (1988))

6. CONCLUSION

We have illustrated the idea of using QMC methods in Bayesian Inference, mainly in one-dimensional problems. The QMC approximations may improve the results obtained using pseudo-random numbers in terms of F -discrepancy and error in the estimation of the target characteristics.

We also propose using QMC methods in multidimensional setting obtaining promising results.

ACKNOWLEDGMENTS

This work has been supported by the grant number IPR00A075 from the Junta de Extremadura. It was completed while Carlos J. Pérez Sánchez was visiting the University of Málaga.

REFERENCES

- BERGER, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *J. Amer. Stat. Assoc.*, **95**, 452, 1269-1276.
- BARABESI, L. (1993). Random variate generation by using the ratio-of-uniforms method. Università degli Studi di Siena. Dipartimento di Metodi Quantitativi. Collana di Pubblicazioni-Serie Ricerca-Monografie, n.1.
- BARBU, G. (1982). On computer generation of a random variable by transformations of uniforms variables. *Bull. Math. Soc. Sci. Math. Romania*, **26**, 129-139.
- BARBU, G. (1987). A new fast method for computer generation of gamma and beta random variables by transformations of uniform variables. *Statistics*, **18**, 453-464.
- BRATLEY, P. and FOX, B. L. (1988) ALGORITHM 659: Implementing Sobol's quasirandom sequence generator. *ACM Trans. on Mathematical Software*, **14**, 1, 88-100.
- BRATLEY, P., FOX, B. L. and NIEDERREITER, H. (1994). Algorithm 738: Programs to generate Niederreiter's low-discrepancy sequences. *ACM Trans. on Mathematical Software*, **20**, 494-495.
- EVANS, M. and SWARTZ, T. (2000) Approximating Integrals via Monte Carlo and Deterministics Methods. Oxford University Press.

- FANG, K. T. and WANG, Y. (1994). Number-theoretic Methods in Statistics. Chapman and Hall, London.
- FOX, B. L. (1986) ALGORITHM 647: Implementation and relative efficiency of quasirandom sequence generators. *ACM Trans. on Mathematical Software*, **12**, 4, 363-376.
- KINDERMAN, A. J. and MONAHAN, J. F. (1977). Computer generation of random variables using the ratio of uniform deviates. *ACM Trans. on Mathematical Software*, **3**, 3, 257-260.
- LIAO, J. G. (1998). Variance Reducction in Gibbs Sampler Using Quasi Random Numbers. *Journal of Computational and Graphical Statistics*, **7**, 3, 253-266.
- NIEDERREITER, H. (1992). Random number generation and quasi-Monte Carlo methods. Society for Industrial and Applied Mathematics.
- ROBERT, C. P. and CASELLA, G. (1999) Monte Carlo Statistical Methods. Springer Verlag.
- RUBIN, D. G. (1988). Using the SIR Algorithm to Simulate Posterior Distributions. In Bayesian Statistics 3, eds. J. M. Bernardo, M. H. Degroot, D. V. Lindley and A. F. M. Smith. Oxford University Press, 395-402.
- SMITH, A. F. M. and GELFAND, A. E. (1992). Bayesian Statistics Without Tears: A Sampling-Resampling Perspective. *The American Statistician*, **46**, 2, 84-88.
- STEFANESCU, S. and VADUVA, I. (1987). On computer generation of random vectors by transformation of uniformly distributed vectors. *Computing*, **39**, 191-189.
- THOMPSON, S. K. (1992). Sampling. Wiley.
- TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**, 4, 1701-1762.
- VADUVA, I. (1985). Computer generation of random vectors based on the transformation of uniformly distributed vectors. *Proceedings 7th Conference of Probability Theory, Brasov, VNU Scientific Press, Utrech*, 589-598.
- WAKEFIELD, J. C., GELFAND, A. E. and SMITH, A. F. M. (1991). Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing*, **1**, 129-133.
- WANG, Y and FANG, K. T. (1990). Number theoretic methods in applied statistics. *Chinese Ann. Math. Ser. B*, **11**, 41-55.

Análise de Variância e Análise de Regressão com Variáveis Dummy: Mais Semelhanças do que Diferenças

Autores:
Patrícia Oom do Valle
e
Efigénio Rebelo

ANÁLISE DE VARIÂNCIA E ANÁLISE DE REGRESSÃO COM VARIÁVEIS DUMMY: MAIS SEMELHANÇAS DO QUE DIFERENÇAS

ANOVA AND REGRESSION WITH DUMMY VARIABLES: MORE SIMILARITIES THAN DIFFERENCES

Autores: Patrícia Oom do Valle

- Assistente da Faculdade de Economia da Universidade do Algarve - Área Científica de Métodos Quantitativos

Efigénio Rebelo

- Professor Associado da Faculdade de Economia da Universidade do Algarve - Área Científica de Métodos Quantitativos

RESUMO:

- Um modelo de regressão que apenas incluía regressores dummy é por vezes classificado de “modelo de análise de variância” sem que o desenvolvimento teórico e matemático que fundamenta tal designação seja apresentado. A finalidade deste trabalho é esclarecer em que medida estas técnicas se relacionam e demonstrar que ambos os modelos descrevem a mesma realidade e conduzem a testes estatísticos similares.

PALAVRAS-CHAVE:

- *Regressão linear, Variáveis dummy, ANOVA com classificação dupla.*

ABSTRACT:

- When all the regressors of a linear regression model are dummy variables, the model is often classified as “ANOVA model”, without the presentation of the theoretical and mathematical developments that sustain such designation. The purpose of the current study is to show how these techniques are related and prove mathematically that both models describe the same reality and produce similar statistical tests.

KEY-WORDS:

- *Linear regression, Dummy variables, Two-way ANOVA.*

VOLUME I

1º QUADRIMESTRE DE 2002

1. INTRODUÇÃO

Embora os regressores dummy surjam como uma forma de permitir a introdução de factores explicativos de natureza qualitativa em modelos de regressão linear - sendo esta a sua aplicação mais importante e divulgada - a sua utilidade estende-se a domínios mais vastos e complexos, dos quais se destaca a realização de experiências de Análise de Variância com classificação simples ou com classificação dupla.

Porém, o estudo deste contexto de aplicabilidade dos regressores dummy não tem sido suficientemente divulgado nos manuais da especialidade. Na verdade, os manuais de Estatística limitam-se a apresentar uma análise pormenorizada das técnicas de Análise de Variância, sem mencionar que o mesmo tipo de experiências podem ser realizadas mediante o ajustamento de um modelo de regressão com regressores dummy. Similarmente, a generalidade dos manuais de Econometria refere, quanto muito, que um modelo de regressão linear que apenas inclua variáveis dummy pode ser entendido como um modelo de análise de variância. A ausência sistemática de um desenvolvimento teórico do relacionamento entre as duas técnicas, e a inexistência das demonstrações matemáticas que sustentam este tipo de afirmações, dificulta bastante a sua compreensão.

Assim, o objectivo deste trabalho é apresentar o desenvolvimento teórico e matemático necessário ao entendimento da relação que se estabelece entre a técnica de Análise de Variância com classificação dupla e o ajustamento de um modelo econométrico com regressores dummy.

2. ANÁLISE DE VARIÂNCIA COM CLASSIFICAÇÃO DUPLA

Utilizando a definição proposta por Sandy (1990, p.474), a análise da variância com classificação dupla traduz-se num “...teste à igualdade das médias de duas ou mais populações, utilizado sempre que se suspeita que dois factores influenciam a variável de interesse”.

O seguinte exemplo ilustra uma experiência em que a análise de variância com classificação dupla encontra campo de aplicação. Assim, imagine-se que um determinado investigador, suspeitando que a produtividade de um indivíduo é influenciada pelo horário de trabalho em que desempenha a sua actividade bem como pelo escalão etário a que pertence, selecciona uma amostra aleatória de n trabalhadores, $1/3$ com idade compreendida em cada um dos escalões etários: “menos de 35 anos”, “entre 36 e 50 anos” e “mais de 50 anos”. Também de forma aleatória, o investigador coloca metade dos n trabalhadores a exercer funções em cada um dos horários de trabalho (de tal forma que em cada um dos horários opere o mesmo número de indivíduos com idade compreendida em cada nível etário). Designe-se por

r e s , respectivamente, o número de categorias de cada um dos factores explicativos. No final de um determinado período temporal (dia, semana, etc.), o investigador regista o número de unidades produzidas em média por hora por cada trabalhador e, com base nesse conjunto de valores, pretende dar resposta às seguintes questões:

- i) A produtividade esperada varia com o horário de trabalho em que as várias funções são realizadas?
- ii) O escalão etário em que se insere um trabalhador influencia a sua produtividade esperada?
- iii) O facto de um indivíduo se inserir em simultâneo num determinado escalão etário e ter exercido funções num certo horário de trabalho provoca um efeito adicional significativo na produtividade diferente do que resultaria da sua simples pertença a esse escalão e a esse horário?

Nesta situação, a variável resposta consubstancia-se no número de unidades produzidas por hora de um determinado bem e os dois factores explicativos, “horário de trabalho” e “escalão etário”, possuem, respectivamente, dois e três níveis ou categorias (isto é, $r = 2$ e $s = 3$): “horário diurno” e “horário nocturno”, no que respeita ao primeiro factor e “menos de 35 anos”, “entre 36 e 50 anos” e “mais de 50 anos”, no que concerne ao segundo factor. Trata-se, portanto, de um modelo de efeitos fixos uma vez que os níveis explicitamente identificados de cada um dos factores constituem todos os níveis de interesse da análise³. É importante salientar que o caso particular de análise de variância que se desenvolve nesta secção assenta na hipótese de que um número igual de observações (designado por m) é seleccionado de cada combinação diferente de categorias dos dois factores explicativos. Aliás, esta assunção está bem patente na ilustração apresentada acima, uma vez que se está a admitir que, na experiência que se pretende conduzir, participam em igual número indivíduos com idade compreendida em cada um dos três níveis etários, para além de metade destes indivíduos ser colocada a trabalhar em cada um dos dois horários de trabalho. Refira-se que esta hipótese caracteriza a grande generalidade das experiências de análise de variância com classificação dupla descritas nos manuais de estatística.

Observe-se também que, embora se esteja perante o caso particular em que $r = 2$ e $s = 3$, apresenta-se nesta secção fórmulas aplicáveis a um número qualquer de níveis em cada factor.

Para prosseguir na análise, importa definir previamente o seguinte conjunto de parâmetros

μ_{ij} = média da variável resposta à (i,j) ésima combinação dos dois factores

³ De acordo com a natureza dos efeitos, pode-se distinguir três tipos de experiências de aplicação de análise de variância com classificação dupla:

- i) Modelos de efeitos fixos: são aqueles em que os r níveis do primeiro factor e os s níveis do segundo esgotam todos os níveis considerados de interesse pelo investigador;
- ii) Modelos de efeitos aleatórios: são aqueles em que s categorias de ambos os factores constituem amostras aleatórias de grupos maiores de categorias de interesse;
- iii) Modelos de efeitos mistos: são aqueles em que os níveis de um dos factores verificam as características de um modelo de efeitos fixos e os níveis do outro factor verificam as características de um modelo de efeitos aleatórios.

$$\mu_{i..} = \frac{\sum_{j=1}^s \mu_{ij.}}{s} = \text{média da variável resposta ao } i\text{-ésimo nível do primeiro factor}$$

$$\mu_{.j.} = \frac{\sum_{i=1}^r \mu_{ij.}}{r} = \text{média da variável resposta ao } j\text{-ésimo nível do segundo factor}$$

$$\mu = \frac{\sum_{i=1}^r \sum_{j=1}^s \mu_{ij.}}{rs} = \text{média global da variável resposta}$$

bem como as correspondentes estimadores

$$\bar{y}_{ij.} = \text{média amostral da variável resposta à } (i,j) \text{ ésima combinação dos dois factores}$$

$$\bar{y}_{i..} = \frac{\sum_{j=1}^s \sum_{k=1}^m y_{ijk}}{s} = \text{média amostral da variável resposta ao } i\text{-ésimo nível do primeiro factor}$$

$$\bar{y}_{.j.} = \frac{\sum_{i=1}^r \sum_{k=1}^m y_{ijk}}{r} = \text{média amostral da variável resposta ao } j\text{-ésimo nível do segundo factor}$$

$$\bar{y}_{...} = \frac{\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m y_{ijk}}{mrs} = \text{média amostral global da variável resposta}$$

em que y_{ijk} é a k -ésima observação da variável resposta ao i -ésimo nível do primeiro factor e j -ésimo nível do segundo factor.

No modelo de análise de variância que se apresenta nesta secção, assume-se explicitamente que os dois factores considerados podem interagir. Em termos gerais, isto significa que a combinação de qualquer nível i do primeiro factor com qualquer nível j do segundo factor produz eventualmente um efeito na variável resposta diferente do que resulta da soma do efeito separado de cada um desses níveis. Formalmente, a hipótese de existência de interacções implica que a média da (i,j) ésima população se possa expressar como

$$(1) \quad \mu_{ij.} = \mu + \alpha_i + \gamma_j + \xi_{ij} \quad (i = 1, 2, \dots, r; j = 1, 2, \dots, s)$$

em que

$\alpha_i = \mu_{i..} - \mu =$ efeito na variável resposta do i-ésimo nível do primeiro factor ($i = 1, 2, \dots, r$)

$\gamma_j = \mu_{.j.} - \mu =$ efeito na variável resposta do j-ésimo nível do segundo factor ($j = 1, 2, \dots, s$)

$\xi_{ij} = \mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu$ representa o efeito da interacção entre o nível i do primeiro factor e o nível j do segundo factor.

e_{ijk} = componente residual aleatória

tornando claro que se está a assumir que a intensidade com que um nível de um dos factores influencia a variável de interesse depende, também, do nível presente do outro factor.

O modelo de análise de variância correspondente a esta situação,

$$(2) \quad y_{ijk} = \mu + \alpha_i + \gamma_j + \xi_{ij} + e_{ijk}$$

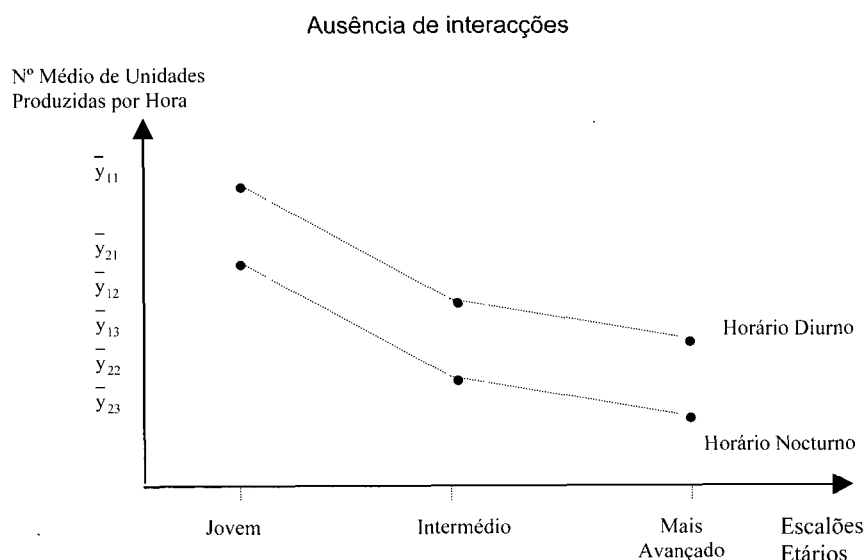
evidencia precisamente a ideia de que cada observação resulta da adição de cinco componentes: o efeito médio global (μ), o efeito do i-ésimo nível do primeiro factor (α_i), o efeito do j-ésimo nível do segundo factor (γ_j), o efeito da interacção entre os dois factores (ξ_{ij}) e o efeito de fontes não identificáveis (e_{ijk}).

A aplicação deste tipo de análise de variância pressupõe a verificação do conjunto de pressupostos que passamos a explicitar:

- i) Cada uma das $r \times s$ populações envolvidas na experiência segue uma distribuição normal;
- ii) As $r \times s$ populações têm a mesma variância, σ^2 ;
- iii) As amostras extraídas de cada população são aleatórias e independentes;
- iv) Os efeitos dos níveis de cada factor interagem entre si.

Milton et al. (1990, p.530) introduzem o conceito de interacção entre os níveis dos dois factores explicativos de uma forma bastante sugestiva e esclarecedora ao defini-lo como “...uma medida de paralelismo”. De facto, e recorrendo à ilustração apresentada, se os dois factores “horário de trabalho” e “escalonamento etário” não interagirem, o gráfico representativo dos valores médios de variável resposta (número de unidades produzidas em média por hora) associados às seis combinações dos seus níveis (diurno/jovem, diurno/idade intermédia, diurno/idade mais avançada, nocturno/jovem, nocturno/idade intermédia e nocturno/idade mais avançada) têm a seguinte configuração:

FIGURA 1. CONFIGURAÇÃO DAS DIFERENÇAS DE PRODUTIVIDADE MÉDIA.



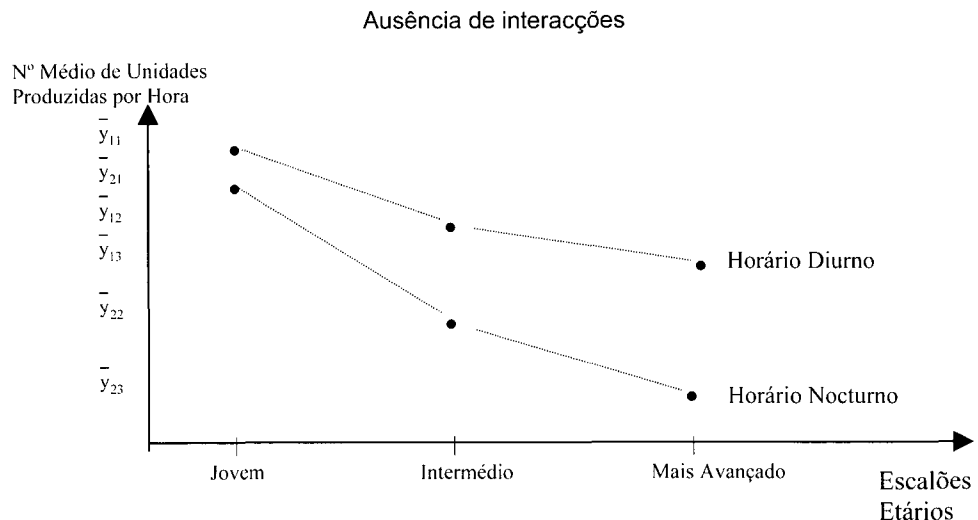
Como ilustra a figura 1, a ausência de interações origina diferenças de produtividade entre os dois horários de trabalho de magnitude constante qualquer que seja o escalão etário a que pertençam os trabalhadores, isto é,

$$\bar{y}_{11} - \bar{y}_{21} = \bar{y}_{12} - \bar{y}_{22} = \bar{y}_{13} - \bar{y}_{23}.$$

As “linhas” representativas de cada um dos horários de trabalho são, portanto, paralelas. Se um conjunto de relações desta natureza caracterizar de facto o fenómeno em análise, o investigador deverá orientar a sua acção para verificar se em cada um dos horários de trabalho a diferença registada entre as médias é ou não significativa e, simultaneamente, se as diferenças apuradas entre as médias inerentes aos três escalões etários são ou não relevantes. Se a análise de variância permitir comprovar a significância destas diferenças, o investigador conclui que a produtividade varia com o horário de trabalho em que o trabalhador desempenha as suas funções (independentemente do escalão etário a que pertença) e também que o grupo etário em que se insere um indivíduo influencia a sua produtividade (qualquer que seja o horário em que exerça as suas tarefas). Neste caso, por exemplo, qualquer uma das diferenças $\bar{y}_{11} - \bar{y}_{21}$, $\bar{y}_{12} - \bar{y}_{22}$ ou $\bar{y}_{13} - \bar{y}_{23}$, indica a diferença de produtividade associada ao exercício das várias funções em horários diferentes.

Pelo contrário, se existirem interações entre os níveis dos dois factores, o gráfico ilustrativo das seis médias ($\bar{y}_{ij}$, com $i = 1,2$; $j = 1,2,3$) pode ser o seguinte:

FIGURA 2. CONFIGURAÇÃO DAS DIFERENÇAS DE PRODUTIVIDADE MÉDIA.



Como evidencia a figura 2, na presença de interações entre os factores explicativos, as diferenças de produtividade entre os dois horários dependem do escalão etário a que pertencem os trabalhadores, ou seja,

$$\bar{y}_{11} - \bar{y}_{21} \neq \bar{y}_{12} - \bar{y}_{22} \neq \bar{y}_{13} - \bar{y}_{23}.$$

Por outras palavras, o investigador não pode concluir acerca das diferenças de produtividade existentes entre os dois horários de trabalho sem atender ao escalão etário em que se inserem os indivíduos, e nem retirar nenhuma conclusão no que concerne às diferenças de produtividade associadas ao desempenho da actividade produtiva por indivíduos pertencentes a diferentes níveis etários, sem ter em atenção o horário de trabalho em que a actividade é realizada.

Como se demonstra seguidamente, todas estas situações podem ser testadas através da análise de variância com classificação dupla.

Milton e al. (1990) sugerem a seguinte sequência de hipóteses a ensaiar no âmbito da análise de variância com classificação dupla e interações:

- i) Testar a ausência de interações significativas entre os dois factores. A hipótese nula a considerar é:

$$H_0 : \zeta_{ij} = 0, \forall (i, j) \quad (i=1,2,\dots,r ; j = 1,2,\dots,s)$$

- ii) No caso da hipótese de ausência de interações não ser rejeitada, testar os “efeitos principais”, isto é, testar que os níveis de cada um dos factores não têm efeitos diferentes na variável resposta. As hipóteses nulas nesta situação são:

$$H_0^1 : \mu_{1..} = \mu_{2..} = \dots = \mu_{r..} \quad \text{ou} \quad H_0^1 : \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$$

e

$$H_0^2 : \mu_{.1} = \mu_{.2} = \dots = \mu_{.s} \quad \text{ou} \quad H_0^2 : \gamma_1 = \gamma_2 = \dots = \gamma_s = 0$$

- iii) Se a hipótese de ausência de interações for rejeitada, cada nível do primeiro factor não tem um comportamento uniforme nos vários níveis do segundo factor pelo que é importante comparar cada combinação de níveis realizando $r + s$ análises de variância com classificação simples para testar as seguintes hipóteses:

$$H_0^3 : \mu_{i1.} = \mu_{i2.} = \dots = \mu_{is.} \quad i = 1, 2, \dots, r$$

$$H_0^3 : \mu_{1j.} = \mu_{2j.} = \dots = \mu_{rj.} \quad j = 1, 2, \dots, s.$$

No que se relaciona com a ilustração que se tem vindo a acompanhar, e em conformidade com a análise efectuada a propósito das figuras 2.2.1 e 2.2.2, estas hipóteses têm a seguinte interpretação:

- i) $H_0 : \zeta_{ij} = 0, \forall (i, j) \quad (i=1,2; j=1,2,3)$: o facto de um indivíduo se inserir em simultâneo num determinado escalão etário e num certo horário de trabalho não provoca um efeito adicional significativo na produtividade para além do que resultaria da sua simples pertença a esse escalão e a esse horário;
- ii) $H_0^1 : \mu_{1..} = \mu_{2..}$ ou $H_0^1 : \alpha_1 = \alpha_2 = 0$: não existe uma diferença significativa entre os efeitos provocados pelos dois níveis do factor explicativo “horário de trabalho” na variável resposta ou, dito de outra forma, a produtividade média de um trabalhador do horário diurno não difere significativamente da produtividade média de um trabalhador do horário nocturno (independentemente do escalão etário a que pertencem).

$H_0^2 : \mu_{.1} = \mu_{.2} = \mu_{.3}$ ou $H_0^2 : \gamma_1 = \gamma_2 = \gamma_3 = 0$: interpretação idêntica à atribuída a H_0^1 considerando o factor explicativo escalão etário e os seus três níveis.

- iii) $H_0^{3,1} : \mu_{11.} = \mu_{12.} = \mu_{13.}$: não existem diferenças de produtividade média entre os três escalões etários quando a actividade é realizada no horário diurno.

$H_0^{3,2} : \mu_{21.} = \mu_{22.} = \mu_{23.}$: não existem diferenças de produtividade média entre os três escalões etários quando a actividade é realizada no horário nocturno.

$H_0^{3,3} : \mu_{11.} = \mu_{21.}$: não existem diferenças de produtividade média entre os trabalhadores dos dois horários de trabalho com menos de 35 anos

$H_0^{3,4} : \mu_{12.} = \mu_{22.}$: não existem diferenças de produtividade média entre os trabalhadores dos dois horários de trabalho com idade compreendida entre os 36 e os 50 anos.

$H_0^{3,5}: \mu_{13..} = \mu_{23..}$: não existem diferenças de produtividade média entre os trabalhadores dos dois horários de trabalho com mais de 50 anos

A decomposição da Soma dos quadrados total (SQT) com base na qual são definidas as estatísticas que permitem realizar os testes definidos em i) e ii) é a seguinte:

Soma dos Quadrados Total (SQT) = Soma dos Quadrados Explicada pelo Primeiro Factor (SQE1) + Soma dos Quadrados Explicada pelo Segundo Factor (SQE2) + Soma dos Quadrados da Interação (SQI) + Soma dos Quadrados Residual (SQR)

ou, tendo em atenção as expressões matemáticas que descrevem cada uma das somas⁴,

$$(3) \quad \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{...})^2 = sm \sum_{i=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2 + rm \sum_{j=1}^s (\bar{y}_{.j.} - \bar{y}_{...})^2 + m \sum_{i=1}^r \sum_{j=1}^s (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2$$

Os rácios F que permitem testar cada uma das hipóteses referidas em i) e ii) definem-se, respectivamente, como:

$$(4) \quad F_1 = \frac{(m-1)rs}{(r-1)(s-1)} \cdot \frac{m \sum_{i=1}^r \sum_{j=1}^s (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2}{\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2} \sim F[(r-1)(s-1), (n-1)rs] \quad \left| \begin{array}{l} H_0: \zeta_{ij} = 0 \end{array} \right.$$

$$(5) \quad F_2 = \frac{(m-1)rs}{r-1} \cdot \frac{sm \sum_{i=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2}{\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2} \sim F[(r-1), (n-1)rs] \quad \left| \begin{array}{l} H_0^1: \mu_{1..} = \mu_{2..} = \dots = \mu_{r..} \end{array} \right.$$

$$(6) \quad F_3 = \frac{(m-1)rs}{s-1} \cdot \frac{rm \sum_{j=1}^s (\bar{y}_{.j.} - \bar{y}_{...})^2}{\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2} \sim F[(s-1), (n-1)rs] \quad \left| \begin{array}{l} H_0^2: \mu_{.1.} = \mu_{.2.} = \dots = \mu_{.s.} \end{array} \right.$$

⁴ Demonstração em apêndice.

A regra de decisão para o teste à hipótese enunciada em i) é:

$$\text{Rejeitar } H_0 \text{ se } F > F_{\alpha}[(r-1)(s-1), (m-1)rs]$$

$$\text{Não Rejeitar } H_0 \text{ se } F \leq F_{\alpha}[(r-1)(s-1), (m-1)rs]$$

e, para a primeira hipótese referida em ii):

$$\text{Rejeitar } H_0^1 \text{ se } F > F_{\alpha}[(r-1), (m-1)rs]$$

$$\text{Não Rejeitar } H_0^1 \text{ se } F \leq F_{\alpha}[(r-1), (m-1)rs]$$

e, para a última hipótese referida em ii):

$$\text{Rejeitar } H_0^2 \text{ se } F > F_{\alpha}[(s-1), (m-1)rs]$$

$$\text{Não Rejeitar } H_0^2 \text{ se } F \leq F_{\alpha}[(s-1), (m-1)rs]$$

O quadro ANOVA correspondente a esta situação é o seguinte:

QUADRO 1. QUADRO ANOVA: ANÁLISE DE VARIÂNCIA COM CLASSIFICAÇÃO DUPLA COM INTERACÇÕES (M OBSERVAÇÕES POR CADA COMBINAÇÃO DIFERENTE DOS DOIS FACTORES)

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrados Médios	Rácio F
Factor Explicativo 1	$SQE_1 = sm \sum_{i=1}^r (\bar{y}_{i..} - \bar{y}_{...})^2$	$r - 1$	$QME_1 = \frac{SQE_1}{r - 1}$	$F_1 = \frac{QME_1}{QMR}$
Factor Explicativo 2	$SQE_2 = rm \sum_{j=1}^s (\bar{y}_{.j.} - \bar{y}_{...})^2$	$s - 1$	$QME_2 = \frac{SQE_2}{s - 1}$	$F_2 = \frac{QME_2}{QMR}$
Interacção	$SQI = m \sum_{i=1}^r \sum_{j=1}^s (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	$(r-1)(s-1)$	$QMI = \frac{SQI}{(r-1)(s-1)}$	$F_3 = \frac{QMI}{QMR}$
Resíduos	$SQR = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2$	$(n-1)rs$	$QMR = \frac{SQR}{(n-1)rs}$	-----
Total	$SQT = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{...})^2$	$mrs - 1$	-----	-----

Cada uma das hipóteses explicitadas em iii) é testável através da realização de análises de variância com classificação simples como a que foi desenvolvida na subsecção anterior.

3. REGRESSÃO COM DUAS VARIÁVEIS NOMINAIS E ANOVA COM CLASSIFICAÇÃO DUPLA

Nesta secção procura-se demonstrar a relação existente entre um modelo de regressão com duas variáveis nominais e os resultados produzidos por um modelo de análise de variância com classificação dupla.

A relação entre a variável dependente “número de unidades produzidas por hora” e as duas variáveis explicativas nominais “horário de trabalho” e “escalão etário” pode ser expressa através do seguinte modelo de regressão linear

$$(7) y_i = \beta_0 + \beta_1 H_{1i} + \beta_2 I_{1i} + \beta_3 I_{2i} + \beta_4 (H_{1i} I_{1i}) + \beta_5 (H_{1i} I_{2i}) + u_i ; i = 1, 2, \dots, n$$

em que H_1 , I_1 e I_2 são regressores dummy, definidos da seguinte forma:

$$H_{1i} = \begin{cases} 1, & \text{se o } i\text{-ésimo trabalhador opera no horário diurno} \\ 0, & \text{se o } i\text{-ésimo trabalhador opera no horário nocturno} \end{cases}$$

$$I_{1i} = \begin{cases} 1, & \text{se o } i\text{-ésimo trabalhador tem menos de 35 anos} \\ 0, & \text{caso contrário} \end{cases}$$

$$I_{2i} = \begin{cases} 1, & \text{se o } i\text{-ésimo trabalhador tem entre 36 e 50 anos} \\ 0, & \text{caso contrário} \end{cases}$$

É importante notar que se está a assumir que quando I_1 e I_2 assumem simultaneamente o valor 0 a observação diz respeito a um indivíduo com idade superior a 50 anos. $H_{1i}I_{1i}$ e $H_{1i}I_{2i}$ são regressores dummy de interacção definidos como:

$$H_{1i}I_{1i} = \begin{cases} 1, & \text{se o } i\text{-ésimo trabalhador opera no horário diurno e tem menos de 35 anos} \\ 0, & \text{caso contrário} \end{cases}$$

$$H_{1i}I_{2i} = \begin{cases} 1, & \text{se o } i\text{-ésimo trabalhador opera no horário diurno e tem entre 36 e 50 anos} \\ 0, & \text{caso contrário} \end{cases}$$

Estas variáveis de interacção permitem, por um lado, que o efeito do horário de trabalho na produtividade de um indivíduo varie com o grupo etário em que se insere e que, por outro, a diferença de produtividade entre quaisquer dois níveis etários dependa igualmente do horário em que a actividade é desempenhada.

Note que, em extensão, o modelo (7) pode escrever-se como:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1_{11} & 1_{11} & 1_{11} & 0_{11} & 1_{11} & 0_{11} \\ 1_{12} & 1_{12} & 0_{12} & 1_{12} & 0_{12} & 1_{12} \\ 1_{13} & 1_{13} & 0_{13} & 0_{13} & 0_{13} & 0_{13} \\ 1_{21} & 0_{21} & 1_{21} & 0_{21} & 0_{21} & 0_{21} \\ 1_{22} & 0_{22} & 0_{22} & 1_{22} & 0_{22} & 0_{22} \\ 1_{23} & 0_{23} & 0_{23} & 0_{23} & 0_{23} & 0_{23} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{23} \end{bmatrix}$$

em que:

y_{ij} : vector coluna das n_{ij} observações da variável dependente relativas à categoria i da primeira variável explicativa e à categoria j da segunda variável ($i = 1,2 ; j = 1,2,3$)

1_{ij} : vector coluna de 1's com n_{ij} elementos ($i = 1,2 ; j = 1,2,3$)

0_{ij} : é um vector coluna de 0's com n_{ij} elementos ($i = 1,2 ; j = 1,2,3$)

u_{ij} : é o vector das variáveis residuais com n_{ij} elementos ($i = 1,2 ; j = 1,2,3$).

Os valores médios da variável dependente, correspondentes às diferentes combinações de valores das variáveis dummy, são agora dados por:

$$(8) \quad E(y_i | H_{1i} = 1, I_{1i} = 1, I_{2i} = 0) = \beta_0 + \beta_1 + \beta_2 + \beta_4$$

$$(9) \quad E(y_i | H_{1i} = 1, I_{1i} = 0, I_{2i} = 1) = \beta_0 + \beta_1 + \beta_3 + \beta_5$$

$$(10) \quad E(y_i | H_{1i} = 1, I_{1i} = 0, I_{2i} = 0) = \beta_0 + \beta_1$$

$$(11) \quad E(y_i | H_{1i} = 0, I_{1i} = 1, I_{2i} = 0) = \beta_0 + \beta_2$$

$$(12) \quad E(y_i | H_{1i} = 0, I_{1i} = 0, I_{2i} = 1) = \beta_0 + \beta_3$$

$$(13) \quad E(y_i | H_{1i} = 0, I_{1i} = 0, I_{2i} = 0) = \beta_0$$

ou, de forma abreviada,

$$(14) \quad \beta_0 + \beta_1 + \beta_2 + \beta_4 = \mu_{11}.$$

$$(15) \quad \beta_0 + \beta_1 + \beta_3 + \beta_5 = \mu_{12}.$$

$$(16) \quad \beta_0 + \beta_1 = \mu_{13}.$$

$$(17) \quad \beta_0 + \beta_2 = \mu_{21}.$$

$$(18) \quad \beta_0 + \beta_3 = \mu_{22}.$$

$$(19) \quad \beta_0 = \mu_{23}.$$

A assunção de que um igual número de observações é recolhido de cada combinação das duas variáveis explicativas permite que se expresse a equação de regressão (7) de uma forma ligeiramente diferente. Seja (20) a equação de regressão adaptada a esta particularidade:

$$(20) \quad y_{ijk} = \beta_0 + \beta_1 H_{1k} + \beta_2 I_{1k} + \beta_3 I_{2k} + \beta_4 (H_{1k} I_{1k}) + \beta_5 (H_{1k} I_{2k}) + u_{ijk}$$

onde y_{ijk} representa o número de unidades produzidas por hora pelo k -ésimo trabalhador ($k = 1, 2, \dots, m$) no horário i ($i = 1, 2$) com idade compreendida no escalão etário j ($j = 1, 2, 3$) e u_{ijk} representa o desvio associado a cada observação.

Pretende-se agora demonstrar que todas as hipóteses formuladas no contexto da análise de variância (mais especificamente, na sequência proposta por Milton et al. (1990)) podem ser testadas através da realização de determinados testes de significância aos parâmetros do modelo (20). Mostrar-se-á também que as estatísticas que permitem conduzir estes testes são exactamente as apresentadas em (4) a (6), a propósito da descrição da técnica estatística de análise de variância com classificação dupla.

Para estabelecer a equivalência, entre análise de variância com classificação dupla e a análise de regressão que considere efeitos de interacção entre as suas duas variáveis explicativas nominais, é necessário determinar a expressão que define a Soma dos Quadrados dos Resíduos (SQR) do modelo (20). Assim, através de alguma manipulação algébrica é possível demonstrar que este vector se define como:

$$(21) \quad \hat{\mathbf{u}} = \begin{bmatrix} y_{11} - \bar{y}_{11.} \\ y_{12} - \bar{y}_{12.} \\ y_{13} - \bar{y}_{13.} \\ y_{21} - \bar{y}_{21.} \\ y_{22} - \bar{y}_{22.} \\ y_{23} - \bar{y}_{23.} \end{bmatrix}$$

Assim, o resíduo associado a cada observação pode expressar-se da seguinte forma

$$(22) \quad \hat{u}_{ijk} = y_{ijk} - \bar{y}_{ij.} \quad (i = 1, 2; j = 1, 2, 3; k = 1, 2, \dots, m)$$

pelo que:

$$(23) \quad SQR = \hat{\mathbf{u}}' \hat{\mathbf{u}} = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2.$$

À hipótese de ausência de interações entre os dois factores explicativos, formulada a propósito da análise de variância, corresponde a hipótese nula $H_0: \beta_4 = \beta_5 = 0$ no contexto do modelo de regressão (20). A estatística que permite conduzir este teste é a seguinte

$$(24) \quad F = \frac{\left(\hat{\mathbf{u}}_r' \hat{\mathbf{u}}_r - \hat{\mathbf{u}}' \hat{\mathbf{u}} \right) / g}{\hat{\mathbf{u}}' \hat{\mathbf{u}} / (n - k)} \sim F_{(g, n-k)} \quad \left| \begin{array}{l} H_0: \beta_4 = \beta_5 = 0 \end{array} \right.$$

em que $\hat{\mathbf{u}}' \hat{\mathbf{u}}$ é a SQR do modelo (20) e $\hat{\mathbf{u}}_r' \hat{\mathbf{u}}_r$ o SQR do mesmo modelo uma vez imposta a restrição $\beta_4 = \beta_5 = 0$. Note que a imposição desta restrição transforma o modelo (20) num modelo sem efeitos de interacção. Adicionalmente, a hipótese de que um número idêntico de observações é seleccionado de cada combinação das duas variáveis nominais permite que se expresse o modelo:

$$(25) \quad y_{ijk} = \beta_0 + \beta_1 H_{1k} + \beta_2 I_{1k} + \beta_3 I_{2k} + u_{ijk}$$

Observe que, num modelo assim formulado, o vector de estimadores dos mínimos quadrados ordinários (EMQO) é dado por⁵:

$$(26) \quad \hat{\beta} = \begin{bmatrix} \bar{y}_{2..} + \bar{y}_{.3.} - \bar{y}_{...} \\ \bar{y}_{1..} - \bar{y}_{2..} \\ \bar{y}_{.1.} - \bar{y}_{.3.} \\ \bar{y}_{.2.} - \bar{y}_{.3.} \end{bmatrix}.$$

Em resultado de (26),

$$(27) \quad \hat{u}_r = y - \hat{y} = y - X\hat{\beta} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{y}_{2..} + \bar{y}_{.3.} - \bar{y}_{...} \\ \bar{y}_{1..} - \bar{y}_{2..} \\ \bar{y}_{.1.} - \bar{y}_{.3.} \\ \bar{y}_{.2.} - \bar{y}_{.3.} \end{bmatrix} = \begin{bmatrix} y_{11} - (\bar{y}_{1..} + \bar{y}_{.1.} - \bar{y}_{...}) \\ y_{12} - (\bar{y}_{1..} + \bar{y}_{.2.} - \bar{y}_{...}) \\ y_{13} - (\bar{y}_{1..} + \bar{y}_{.3.} - \bar{y}_{...}) \\ y_{21} - (\bar{y}_{2..} + \bar{y}_{.1.} - \bar{y}_{...}) \\ y_{22} - (\bar{y}_{2..} + \bar{y}_{.2.} - \bar{y}_{...}) \\ y_{23} - (\bar{y}_{2..} + \bar{y}_{.3.} - \bar{y}_{...}) \end{bmatrix}$$

em que 1 é um vector coluna de 1's e 0 um vector coluna de 0's, ambos com m elementos. Portanto,

$$(28) \quad \hat{u}_{r(ijk)} = y_{ijk} - (\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}) \quad (i = 1, 2; j = 1, 2, 3; k = 1, 2, \dots, m)$$

pelo que:

$$(29) \quad \hat{u}_r' \hat{u}_r = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2.$$

Assim, atendendo a que $g = 2$ e substituindo (23) e (29) na estatística F (24), tem-se

$$(30) \quad F = \frac{6m-6}{2} \cdot \frac{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2}{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m (y_{ijk} - \bar{y}_{ij.})^2} \sim F(2, 6m-6) \quad \left| \begin{array}{l} H_0: \beta_4 = \beta_5 = 0 \end{array} \right|$$

⁵ Demontração em apêndice.

ou, simplificando o numerador através da seguinte igualdade⁶

$$(31) \quad \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2$$

$$= m \sum_{i=1}^2 \sum_{j=1}^3 \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2$$

$$(32) \quad F = \frac{6m-6}{2} \cdot \frac{m \sum_{i=1}^2 \sum_{j=1}^3 \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2}{\sum_{k=1}^m \sum_{i=1}^2 \sum_{j=1}^3 \left(y_{ijk} - \bar{y}_{ij.} \right)^2} \sim F(2, 6m-6) \quad \left| \begin{array}{l} H_0: \beta_4 = \beta_5 = 0 \end{array} \right.$$

que não é mais do que a estatística que apresentámos em (4) para o caso particular em que $r=2$ e $s=3$ e atendendo a que $H_0: \beta_4 = \beta_5 = 0 \Leftrightarrow H_0: \zeta_{ij} = 0$. Ficou assim demonstrado em que medida a hipótese i) formulada na estratégia de Milton et al. (1990) pode ser testada através da análise de regressão com variáveis dummy. Destaque-se também a equivalência entre a estatística necessária à condução do teste nos dois contextos.

Por sua vez, a hipótese de ausência de efeitos significativos provocados pelos níveis do primeiro factor, que definimos no âmbito da análise de variância (hipótese ii) $H_0^1: \mu_{1..} = \mu_{2..}$ da sequência proposta por Milton et al.), encontra contrapartida no modelo de regressão (20) na hipótese nula $H_0: \beta_5 + 3\beta_1 + \beta_4 = 0$. Esta conclusão é facilmente compreensível se atender à definição de $\mu_{i..} = \sum_{j=1}^s \mu_{ij.} / s$ na secção 2. Assim, uma vez que no caso particular em análise, $s = 3$, então

$$H_0: \mu_{1..} = \mu_{2..} \Leftrightarrow \frac{\mu_{11.} + \mu_{12.} + \mu_{13.}}{3} = \frac{\mu_{21.} + \mu_{22.} + \mu_{23.}}{3}$$

$$\Leftrightarrow \mu_{11.} + \mu_{12.} + \mu_{13.} - \mu_{21.} - \mu_{22.} - \mu_{23.} = 0$$

ou ainda,

$$(\beta_0 + \beta_1 + \beta_2 + \beta_4) + (\beta_0 + \beta_1 + \beta_3 + \beta_5) + (\beta_0 + \beta_1) - (\beta_0 + \beta_2) - (\beta_0 + \beta_3) - \beta_0 = 0$$

$$\Leftrightarrow \beta_5 + 3\beta_1 + \beta_4 = 0$$

tendo em consideração os resultados que apresentámos em (14) a (19).

⁶ Demonstração em apêndice.

A estatística que permite ensaiar esta hipótese, usando o modelo de regressão (20), é a igual à indicada em (24) para testar a ausência de efeitos de interacção entre as variáveis nominais. Neste caso específico, $g = 1$, uma vez que apenas se pretende

testar uma hipótese, e $\hat{u}_r^* \hat{u}_r$ é o SQR do modelo (20) após a sua reparametrização com a referida restrição. É simples verificar que substituindo β_5 por $-3\beta_1 - \beta_4$ no modelo (20) este se transforma em

$$(33) y_{ijk} = \beta_0 + \beta_1 [H_{1k} - 3(H_{1k} I_{2k})] + \beta_2 I_{1k} + \beta_3 I_{2k} + \beta_4 [(H_{1k} I_{1k}) - (H_{1k} I_{2k})] + u_{ijk}$$

donde provem o seguinte vector de EMPO.⁷

$$(34) \quad \hat{\beta} = \begin{bmatrix} (15\bar{y}_{...} - 2\bar{y}_{11.} - 2\bar{y}_{12.} - 2\bar{y}_{13.} - 3\bar{y}_{21.} - 3\bar{y}_{22.})/3 \\ (-12\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + 4\bar{y}_{13.} + 3\bar{y}_{21.} + 3\bar{y}_{22.})/3 \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + 2\bar{y}_{21.} + \bar{y}_{22.} \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + \bar{y}_{21.} + 2\bar{y}_{22.} \\ 6\bar{y}_{...} - \bar{y}_{12.} - 2\bar{y}_{13.} - 2\bar{y}_{21.} - \bar{y}_{22.} \end{bmatrix}$$

e o vector de resíduos

$$(35) \quad \hat{u} = \begin{bmatrix} y_{11} - \bar{y}_{...} - \bar{y}_{11.} + \bar{y}_{1..} \\ y_{12} - \bar{y}_{...} - \bar{y}_{12.} + \bar{y}_{1..} \\ y_{13} - \bar{y}_{...} - \bar{y}_{13.} + \bar{y}_{1..} \\ y_{21} - \bar{y}_{...} - \bar{y}_{21.} + \bar{y}_{2..} \\ y_{22} - \bar{y}_{...} - \bar{y}_{22.} + \bar{y}_{2..} \\ y_{23} - \bar{y}_{...} - \bar{y}_{23.} + \bar{y}_{2..} \end{bmatrix}$$

Sendo o resíduo associado a cada observação dado por

$$(36) \quad \hat{u}_{ijk} = y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{i..} \quad (i = 1, 2; j = 1, 2, 3; k = 1, 2, \dots, m)$$

então,

⁷ Demonstração em apêndice.

$$(37) \quad \hat{u}_r \hat{u}_r = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{i..} \right)^2$$

o que significa que o rácio F (24) pode escrever-se como

$$(38) \quad F = \frac{6m-6}{1} \cdot \frac{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{i..} \right)^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2}{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2} \sim F(1, 6m-6) \quad \left| \begin{array}{l} H_0: \beta_5 + 3\beta_1 + \beta_4 = 0 \end{array} \right.$$

ou, após a simplificação do numerador tendo em atenção a igualdade⁸

$$(39) \quad \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{i..} \right)^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2 \\ = 3m \sum_{i=1}^2 \left(\bar{y}_{i..} - \bar{y}_{...} \right)^2,$$

da seguinte forma:

$$(40) \quad F = \frac{6m-6}{1} \cdot \frac{3m \sum_{i=1}^2 \left(\bar{y}_{i..} - \bar{y}_{...} \right)^2}{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2} \sim F(1, 6m-6) \quad \left| \begin{array}{l} H_0: \beta_5 + 3\beta_1 + \beta_4 = 0 \end{array} \right.$$

Esta estatística identifica-se com a apresentada em (5) após a concretização dos valores de r e s por 2 e 3, respectivamente e, tendo em atenção a equivalência entre as hipótese $H_0: \beta_5 + 3\beta_1 + \beta_4 = 0$ e $H_0: \mu_{1..} = \mu_{2..}$. Mais uma vez ficou demonstrada a equivalência entre a estatística a utilizar para realizar o teste no âmbito da análise de variância e no âmbito da análise de regressão com regressores dummy.

Incidindo agora a análise na hipótese de ausência de efeitos significativos induzidos pelo segundo factor (hipótese $H_0: \mu_{1.} = \mu_{2.} = \mu_{3.}$ na sequência sugerida por Milton et al.), refira-se que $H_0: \begin{cases} \beta_4 + 2\beta_2 = 0 \\ \beta_5 + 2\beta_3 = 0 \end{cases}$ constitui a hipótese que lhe é equivalente no contexto do modelo de regressão (20). Para justificar esta afirmação,

⁸ Demonstração em apêndice.

será utilizado o mesmo procedimento explicado anteriormente. Com efeito, uma vez que $\mu_{.j.} = \sum_{i=1}^r \mu_{ij.} / r$ e, na presente situação, $r = 2$,

$$H_0 : \mu_{.1.} = \mu_{.2.} = \mu_{.3.} \Leftrightarrow \frac{\mu_{11.} + \mu_{21.}}{2} = \frac{\mu_{12.} + \mu_{22.}}{2} = \frac{\mu_{13.} + \mu_{23.}}{2}$$

ou, atendendo às fórmulas apresentadas de (14) a (19)

$$\Leftrightarrow \begin{cases} (\beta_0 + \beta_1 + \beta_2 + \beta_4) + (\beta_0 + \beta_2) = (\beta_0 + \beta_1) + \beta_0 \\ (\beta_0 + \beta_1 + \beta_3 + \beta_5) + (\beta_0 + \beta_3) = (\beta_0 + \beta_1) + \beta_0 \end{cases} \Leftrightarrow \begin{cases} \beta_4 + 2\beta_2 = 0 \\ \beta_5 + 2\beta_3 = 0 \end{cases}$$

Reparametrizando o modelo (20), introduzindo-lhe este par de restrições, vem

$$(41) \quad y_{ijk} = \beta_0 + \beta_1 H_{1k} + \beta_2 [I_{1k} - 2(H_{1k} I_{1k})] + \beta_3 [I_{2k} - 2(H_{1k} I_{2k})] + u_{ijk}$$

cujo vector de EMQO e vector de resíduos são dados por⁹

$$(42) \quad \hat{\beta} = \begin{bmatrix} (24\bar{y}_{...} - 3\bar{y}_{11.} - 3\bar{y}_{12.} - 6\bar{y}_{13.} - 3\bar{y}_{21.} - 3\bar{y}_{22.}) / 6 \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + 2\bar{y}_{13.} + \bar{y}_{21.} + \bar{y}_{22.} \\ (-6\bar{y}_{...} + \bar{y}_{12.} + 2\bar{y}_{13.} + 2\bar{y}_{21.} + \bar{y}_{22.}) / 2 \\ (-6\bar{y}_{...} + \bar{y}_{11.} + 2\bar{y}_{13.} + \bar{y}_{21.} + 2\bar{y}_{22.}) / 2 \end{bmatrix}$$

e

$$(43) \quad \hat{u} = \begin{bmatrix} y_{11} - \bar{y}_{...} - \bar{y}_{11.} + \bar{y}_{.1.} \\ y_{12} - \bar{y}_{...} - \bar{y}_{12.} + \bar{y}_{.2.} \\ y_{13} - \bar{y}_{...} - \bar{y}_{13.} + \bar{y}_{.3.} \\ y_{21} - \bar{y}_{...} - \bar{y}_{21.} + \bar{y}_{.1.} \\ y_{22} - \bar{y}_{...} - \bar{y}_{22.} + \bar{y}_{.2.} \\ y_{23} - \bar{y}_{...} - \bar{y}_{23.} + \bar{y}_{.3.} \end{bmatrix}$$

respectivamente.

⁹ Demonstração em apêndice.

Mediante a expressão que define o resíduo associado a cada observação,

$$(44) \quad \hat{u}_{ijk} = y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{.j.} \quad (i = 1, 2; j = 1, 2, 3; k = 1, 2, \dots, m)$$

pode concluir-se que

$$(45) \quad \hat{u}_r' \hat{u}_r = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{.j.} \right)^2$$

e, consequentemente, o rácio F (24) pode escrever-se como:

$$(46) \quad F = \frac{6m-6}{2} \cdot \frac{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{.j.} \right)^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2}{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2} \sim F(2, 6m-6) \quad \left| \begin{array}{l} H_0: \begin{cases} \beta_4 + 2\beta_2 = 0 \\ \beta_5 + 2\beta_3 = 0 \end{cases} \end{array} \right.$$

Por outro lado, a igualdade

$$(47) \quad \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{.j.} \right)^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2 = 2m \sum_{j=1}^3 \left(\bar{y}_{.j.} - \bar{y}_{...} \right)^2$$

permite que se expresse o rácio F (46) da seguinte forma¹⁰

$$(48) \quad F = \frac{6m-6}{2} \cdot \frac{2m \sum_{j=1}^3 \left(\bar{y}_{.j.} - \bar{y}_{...} \right)^2}{\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2} \sim F(2, 6m-6) \quad \left| \begin{array}{l} H_0: \begin{cases} \beta_4 + 2\beta_2 = 0 \\ \beta_5 + 2\beta_3 = 0 \end{cases} \end{array} \right.$$

o que equivale à estatística apresentada em (6), para $r = 2$ e $s = 3$, no contexto da análise de variância com classificação dupla.

Por último, no que respeita ao terceiro conjunto de hipóteses a ensaiar, referidas na parte final do desenvolvimento teórico da técnica análise de variância com classificação dupla, é possível demonstrar a sua correspondência com as hipóteses

¹⁰ Demonstração em apêndice.

$$H_0^1 : \beta_2 + \beta_4 = \beta_3 + \beta_5 = 0$$

$$H_0^2 : \beta_2 = \beta_3 = 0$$

$$H_0^3 : \beta_1 + \beta_4 = 0$$

$$H_0^4 : \beta_1 + \beta_5 = 0$$

$$H_0^5 : \beta_1 = 0$$

no contexto da análise de regressão.

4. CONSIDERAÇÕES FINAIS

Este estudo demonstra que os objectivos subjacentes à aplicação da técnica de análise de variância com classificação dupla são semelhantes ao que suportam a especificação de um modelo de regressão com duas variáveis explicativas nominais. A par de objectivos similares mostrou-se igualmente que ambas as técnicas conduzem a testes estatísticos exactamente iguais.

A análise de regressão com regressores dummy apresenta, porém, algumas vantagens relativamente à análise de variância. Assim, e em primeiro lugar, através da especificação de modelos de regressão linear é possível ensaiar, com relativa facilidade, uma grande diversidade de hipóteses. A análise de variância com classificação dupla apenas se encontra prontamente preparada para permitir testar um número muito restrito destas hipóteses. Por outro lado, numa aplicação de análise de regressão, é irrelevante em termos de cálculos se o número de observações em cada combinação de categorias das variáveis explicativas é ou não igual. No âmbito de uma situação de aplicação de análise de variância com classificação dupla, a existência de um número diferente de observações em cada combinação de níveis exige fórmulas especiais de cálculo que raramente aparecem apresentadas nos manuais de estatística (inclusivamente nos que abordam estes temas com mais detalhe) e que nem todos os programas informáticos dispõem. Finalmente, a análise de regressão com regressores dummy pode adaptar-se a um número qualquer de variáveis nominais enquanto que os programas informáticos disponíveis para análise de variância não permitem um número muito grande de factores explicativos.

BIBLIOGRAFIA

- CHOW, G. C. (1960), "Tests of Equality Between Sets of Coefficients in two Linear Regressions", *Econometrica*, Vol. 28(3), pp. 591-605.
- DARLINGTON, R. B. (1990), *Regression and Linear Models*. New York (McGraw-Hill).
- GREENE, W. H. (1993), *Econometric Analysis*. New York (Macmillan Publishing Company).
- GRIFFITHS, W. E., CARTER H. e JUDGE, G. (1992), *Learning and Practicing Econometrics*. New York (John Wiley & Sons, Inc.).
- GUJARATI, D. (1970), "Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Note", *The American Statistician*, Vol. 24(1), pp. 50-52.
- GUJARATI, D. (1970), "Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Generalisation", *The American Statistician*, Vol. 24(5), pp. 18-21.
- HARDY, M. A. (1993), *Regression With Dummy Variables*. Newbury Park (Sage Publications, Inc.).
- HAYS, W. L. (1994), *Statistics*. Philadelphia (Harcourt Brace College Publishers).
- KENNEDY, P. (1986), "Interpreting Dummy Variables", *The Review of Economics and Statistics*, Vol. 68(1), pp. 174-175.
- KOOYMAN, M. A. (1976), *Dummy Variables in Econometrics*. Netherlands (Tilburg University Press).
- KULLBACK, S. e ROSENBLATT H. M. (1957), "On the Analysis of Multiple Regression in k Categories", *Biometrika*, 67-83.
- LONG, J. S. e MIETHE, T. D. (1988), *The Statistic Comparison of Groups*. Newbury Park (Sage Publications, Inc.).
- MADDALA, G. S. (1992), *Introduction to Econometrics*. New York (Macmillan Publishing Company).
- MILTON, J. S. e ARNOLD, J. C. (1990), *Introduction to Probability and Statistics, Principles and Applications for Engineering and the Computing Sciences*. New York (McGraw-Hill).
- POKORNI, M. (1992), *An Introduction to Econometrics*. Oxford (Blackwell).
- SANDY, R. (1990), *Statistics for Business and Economics*. New York (McGraw-Hill).
- SEARLE, S. R. e UDELL J. G. (1970), "The Use of Regression on Dummy Variables in Management Research", *MANAGEMENT SCIENCE*, VOL. 16(6), pp. 397-409.
- SILVERTRE, A. (1992), *Análise de variância*. Lisboa (Edições Cedin).

- STEWART, J. (1991), *Econometrics*. Cambridge (Philip Allan).
- SUITS, D. B. (1957), "Use of Dummy Variables in Regression Equations". *Journal of the American Statistical Association*, Vol. 52(280), pp. 548-551.
- SUITS, D. B. (1984), "Dummy Variables: Mechanics V. Interpretation", *The Review of Economics & Statistics*, Vol. 66, pp. 177-180.
- SWAN, C. e CASSIDY H. J. (1975), "One-Time Dummies: Zero-One Dummy Variables Which Take on Only One Non-Zero Value", *Journal of Economics and Business*, Vol. 27(1), pp. 268-271.
- THOMAS, R. L. (1985), *Introductory Econometrics*. London and New York (Longman).
- VALLE, P. O. (1995), *Regressores Dummy e sua Aplicabilidade no Modelo Clássico de Regressão Linear*. Tese de Mestrado.
- WONNACOTT, T. J. e WONNACOTT, R. J. (1990). *Introductory Statistics for Business and Economics*. New York (John Wiley & Sons).

APÊNDICE

Neste apêndice apresenta-se algumas demonstrações que suportam a análise precedente.

Demonstração da Igualdade (3)

Adicionando e subtraindo $\bar{y}_{i..}, \bar{y}_{.j.}, \bar{y}_{...}$ e $\bar{y}_{ij}$ à expressão que se encontra entre parêntesis no somatório que se encontra no primeiro membro, vem

$$\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m (y_{ijk} - \bar{y}_{...})^2 =$$

$$= \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left[\left(\bar{y}_{i..} - \bar{y}_{...} \right) + \left(\bar{y}_{.j.} - \bar{y}_{...} \right) + \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right) + \left(y_{ijk} - \bar{y}_{ij} \right) \right]^2$$

e, desenvolvendo o quadrado da soma:

$$(49) \quad = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left[\left(\bar{y}_{i..} - \bar{y}_{...} \right)^2 + \left(\bar{y}_{i..} - \bar{y}_{...} \right) \left(\bar{y}_{.j.} - \bar{y}_{...} \right) + \left(\bar{y}_{i..} - \bar{y}_{...} \right) \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right) + \right.$$

$$+ \left(\bar{y}_{i..} - \bar{y}_{...} \right) \left(y_{ijk} - \bar{y}_{ij} \right) + \left(\bar{y}_{.j.} - \bar{y}_{...} \right) \left(\bar{y}_{i..} - \bar{y}_{...} \right) + \left(\bar{y}_{.j.} - \bar{y}_{...} \right)^2 +$$

$$+ \left(\bar{y}_{.j.} - \bar{y}_{...} \right) \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right) + \left(\bar{y}_{.j.} - \bar{y}_{...} \right) \left(y_{ijk} - \bar{y}_{ij} \right) +$$

$$+ \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right) \left(\bar{y}_{i..} - \bar{y}_{...} \right) + \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right) \left(\bar{y}_{.j.} - \bar{y}_{...} \right) +$$

$$+ \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2 + \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right) \left(y_{ijk} - \bar{y}_{ij} \right) +$$

$$+ \left(y_{ijk} - \bar{y}_{ij} \right) \left(\bar{y}_{i..} - \bar{y}_{...} \right) + \left(y_{ijk} - \bar{y}_{ij} \right) \left(\bar{y}_{.j.} - \bar{y}_{...} \right) +$$

$$\left. + \left(y_{ijk} - \bar{y}_{ij} \right) \left(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right) + \left(y_{ijk} - \bar{y}_{ij} \right)^2 \right]$$

Isolando os vários somatórios em (49), vem

$$= \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(\bar{y}_{i..} - \bar{y}_{...} \right)^2 + \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(\bar{y}_{.j.} - \bar{y}_{...} \right)^2$$

$$+ \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2 + \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij} \right)^2 +$$

$$\begin{aligned}
& + 2 \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(\bar{y}_{i..} - \bar{\bar{y}}_{...} \right) \left(\bar{y}_{.j.} - \bar{\bar{y}}_{...} \right) + 2 \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(\bar{y}_{.j.} - \bar{\bar{y}}_{...} \right) \left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{\bar{y}}_{...} \right) + \\
& + 2 \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{\bar{y}}_{...} \right) \left(\bar{y}_{.j.} - \bar{\bar{y}}_{...} \right) + \\
& + 2 \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(\bar{y}_{i..} - \bar{\bar{y}}_{...} \right) \left(y_{ijk} - \bar{y}_{ij..} \right) + 2 \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(\bar{y}_{.j.} - \bar{\bar{y}}_{...} \right) \left(y_{ijk} - \bar{y}_{ij..} \right) + \\
& + 2 \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij..} \right) \left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{\bar{y}}_{...} \right) \\
& = sm \sum_{i=1}^r \left(\bar{y}_{i..} - \bar{\bar{y}}_{...} \right)^2 + rm \sum_{j=1}^s \left(\bar{y}_{.j.} - \bar{\bar{y}}_{...} \right)^2 + m \sum_{i=1}^r \sum_{j=1}^s \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{\bar{y}}_{...} \right)^2 + \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij..} \right)^2
\end{aligned}$$

uma vez que as restantes igualdades são nulas.

Demonstração da Igualdade (26)

Atendendo a que o regressor dummy H_1 identifica com o valor 1 a primeira categoria da variável nominal “horário de trabalho”, I_1 a primeira categoria da variável “escalão etário” e I_2 a segunda categoria desta última variável, pode escrever-se o modelo (25),

$$y_{ijk} = \beta_0 + \beta_1 H_{ak} + \beta_2 I_{1k} + \beta_3 I_{2k} + u_{ijk}$$

na forma matricial

$$(50) \quad y = X\beta + u ; u \sim N(0, \sigma^2 I_n)$$

e, em extensão, como:

$$(51) \quad \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{23} \end{bmatrix}$$

onde:

y_{ij} : vector coluna das m observações da variável dependente relativas à categoria i da primeira variável explicativa e à categoria j da segunda variável ($i = 1, 2$; $j = 1, 2, 3$)

1: vector coluna de 1's com m elementos

0: é um vector coluna de 0's com m elementos

u_{ij} : é o vector dos desvios ($i = 1, 2$; $j = 1, 2, 3$)

No caso em análise, as operações entre matrizes/vectores necessárias à aplicação da igualdade $\hat{\beta} = (X'X)^{-1} X'y$ são:

$$(52) X'X = \begin{bmatrix} 1' & 1' & 1' & 1' & 1' & 1' \\ 1' & 1' & 1' & 0' & 0' & 0' \\ 1' & 0' & 0' & 1' & 0' & 0' \\ 0' & 1' & 0' & 0' & 1' & 0' \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 6m & 3m & 2m & 2m \\ 3m & 2m & m & m \\ 2m & m & 2m & 0 \\ 2m & m & 0 & 2m \end{bmatrix}$$

$$(53) X'y = \begin{bmatrix} \sum_{k=1}^m y_{11k} + \sum_{k=1}^m y_{12k} + \sum_{k=1}^m y_{13k} + \sum_{k=1}^m y_{21k} + \sum_{k=1}^m y_{22k} + \sum_{k=1}^m y_{23k} \\ \sum_{k=1}^m y_{11k} + \sum_{k=1}^m y_{12k} + \sum_{k=1}^m y_{13k} \\ \sum_{k=1}^m y_{11k} + \sum_{k=1}^m y_{21k} \\ \sum_{k=1}^m y_{12k} + \sum_{k=1}^m y_{22k} \end{bmatrix} = \begin{bmatrix} 6\bar{y}_{...} \\ 3\bar{y}_{1..} \\ 2\bar{y}_{.1.} \\ 2\bar{y}_{.2.} \end{bmatrix}$$

$$(54) (X'X)^{-1} = \frac{1}{m} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{3} & \frac{2}{3} & 0 & 0 \\ -\frac{1}{2} & 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 \end{bmatrix}$$

A multiplicação de (14) por (13) produz:

$$(55) \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 4\bar{y}_{...} - \bar{y}_{1..} - \bar{y}_{.1.} - \bar{y}_{.2.} \\ -2\bar{y}_{...} + 2\bar{y}_{1..} \\ -3\bar{y}_{...} + 2\bar{y}_{.1.} + \bar{y}_{.2.} \\ -3\bar{y}_{...} + \bar{y}_{.1.} + 2\bar{y}_{.2.} \end{bmatrix}$$

ou, tendo em atenção que

$$(56) \bar{y}_{...} = \frac{2m\bar{y}_{.1.} + 2m\bar{y}_{.2.} + 2m\bar{y}_{.3.}}{6m} = \frac{\bar{y}_{.1.} + \bar{y}_{.2.} + \bar{y}_{.3.}}{3}$$

ou

$$(57) \bar{y}_{...} = \frac{3m\bar{y}_{1..} + 3m\bar{y}_{2..}}{6m} = \frac{\bar{y}_{1..} + \bar{y}_{2..}}{2}$$

e substituindo estas igualdades em (55), esta pode expressar-se alternativamente como:

$$(58) \quad \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \bar{y}_{2..} - \bar{y}_{.3.} - \bar{y}_{...} \\ \bar{y}_{1..} - \bar{y}_{2..} \\ \bar{y}_{.1.} - \bar{y}_{.3.} \\ \bar{y}_{.2.} - \bar{y}_{.3.} \end{bmatrix}$$

Demonstração da Igualdade (31)

Agregando num só somatório triplo os dois somatórios do primeiro membro,

$$\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left[\left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2 - \left(y_{ijk} - \bar{y}_{ij.} \right)^2 \right]$$

e, após simplificações e eliminação do somatório triplo, vem:

$$(59) \quad \sum_{i=1}^2 \sum_{j=1}^3 \left(-2\bar{y}_{i..} \sum_{k=1}^m y_{ijk} - 2\bar{y}_{.j.} \sum_{k=1}^m y_{ijk} + 2\bar{y}_{...} \sum_{k=1}^m y_{ijk} + \bar{y}_{i..}^2 m + 2\bar{y}_{i..} \bar{y}_{.j.} m - \right. \\ \left. - 2\bar{y}_{...} \bar{y}_{i..} m + \bar{y}_{.j.}^2 m - 2\bar{y}_{.j.} \bar{y}_{...} m + \bar{y}_{...}^2 m - \bar{y}_{ij.}^2 m + 2\bar{y}_{ij.} \sum_{k=1}^m y_{ijk} \right)$$

Dado que $\sum_{k=1}^m y_{ijk} = m\bar{y}_{ij.}$, (59) pode expressar-se como

$$m \sum_{i=1}^2 \sum_{j=1}^3 \left(-2\bar{y}_{i..} \bar{y}_{ij.} - 2\bar{y}_{.j.} \bar{y}_{ij.} + 2\bar{y}_{...} \bar{y}_{ij.} + \bar{y}_{i..}^2 + 2\bar{y}_{i..} \bar{y}_{.j.} - 2\bar{y}_{...} \bar{y}_{i..} + \bar{y}_{.j.}^2 - 2\bar{y}_{.j.} \bar{y}_{...} + \right. \\ \left. + \bar{y}_{...}^2 - \bar{y}_{ij.}^2 + 2\bar{y}_{ij.} \bar{y}_{ij.} \right)$$

e, após simplificações,

$$m \sum_{i=1}^2 \sum_{j=1}^3 \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \right)^2$$

como se pretendia mostrar.

Demonstração das Igualdades (34) e (35)

A forma como se definem as variáveis dummy H_1 , I_1 e I_2 , permite representar o modelo (33)

$$y_{ijk} = \beta_0 + \beta_1 [H_{1k} - 3(H_{1k} I_{2k})] + \beta_2 I_{1k} + \beta_3 I_{2k} + \beta_4 [(H_{1k} I_{1k}) - (H_{1k} I_{2k})] + u_{ijk}$$

$$; i = 1, 2 ; j = 1, 2, 3 ; k = 1, 2, \dots, n_i ; u_{ijk} \sim \text{IIN}(0, \sigma^2)$$

da seguinte forma:

$$(60) \quad \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & -2 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{23} \end{bmatrix}$$

donde provêm os seguintes resultados:

$$(61) \quad X'X = \begin{bmatrix} 1' & 1' & 1' & 1' & 1' & 1' \\ 1' & -2' & 1' & 0' & 0' & 0' \\ 1' & 0' & 0' & 1' & 0' & 0' \\ 0' & 1' & 0' & 0' & 1' & 0' \\ 1' & -1' & 0' & 0' & 0' & 0' \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & -2 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 6m & 0 & 2m & 2m & 0 \\ 0 & 6m & m & -2m & 3m \\ 2m & m & 2m & 0 & m \\ 2m & -2m & 0 & 2m & -m \\ 0 & 3m & m & m & 2m \end{bmatrix}$$

$$(62) \quad X'y = \begin{bmatrix} 1' & 1' & 1' & 1' & 1' & 1' \\ 1' & -2' & 1' & 0' & 0' & 0' \\ 1' & 0' & 0' & 1' & 0' & 0' \\ 0' & 1' & 0' & 0' & 1' & 0' \\ 1' & -1' & 0' & 0' & 0' & 0' \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 6m\bar{y}_{...} \\ m(\bar{y}_{11.} - 2\bar{y}_{12.} + \bar{y}_{13.}) \\ m(\bar{y}_{11.} + \bar{y}_{21.}) \\ m(\bar{y}_{12.} + \bar{y}_{22.}) \\ m(\bar{y}_{11.} - \bar{y}_{12.}) \end{bmatrix}$$

$$(63) \quad (X'X)^{-1} = \frac{1}{m} \begin{bmatrix} \frac{5}{6} & -\frac{4}{3} & -1 & -1 & 1 \\ -\frac{4}{3} & \frac{4}{3} & 1 & 1 & -2 \\ -1 & 1 & 2 & 1 & -2 \\ -1 & 1 & 1 & 2 & -1 \\ 1 & -2 & -2 & -1 & 4 \end{bmatrix}$$

e, após simplificações, o vector de EMQO com a seguinte estrutura:

$$(64) \quad \hat{\beta} = \begin{bmatrix} (15\bar{y}_{...} - 2\bar{y}_{11.} - 2\bar{y}_{12.} - 2\bar{y}_{13.} - 3\bar{y}_{21.} - 3\bar{y}_{22.})/3 \\ (-12\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + 4\bar{y}_{13.} + 3\bar{y}_{21.} + 3\bar{y}_{22.})/3 \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + 2\bar{y}_{21.} + \bar{y}_{22.} \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + \bar{y}_{21.} + 2\bar{y}_{22.} \\ 6\bar{y}_{...} - \bar{y}_{12.} - 2\bar{y}_{13.} - 2\bar{y}_{21.} - \bar{y}_{22.} \end{bmatrix}$$

O vector de resíduos da estimação do modelo será dado por:

$$\begin{bmatrix} \hat{u}_{11} \\ \hat{u}_{12} \\ \hat{u}_{13} \\ \hat{u}_{21} \\ \hat{u}_{22} \\ \hat{u}_{23} \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & -2 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} (15\bar{y}_{...} - 2\bar{y}_{11.} - 2\bar{y}_{12.} - 2\bar{y}_{13.} - 3\bar{y}_{21.} - 3\bar{y}_{22.})/3 \\ (-12\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + 4\bar{y}_{13.} + 3\bar{y}_{21.} + 3\bar{y}_{22.})/3 \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + 2\bar{y}_{21.} + \bar{y}_{22.} \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + \bar{y}_{21.} + 2\bar{y}_{22.} \\ 6\bar{y}_{...} - \bar{y}_{12.} - 2\bar{y}_{13.} - 2\bar{y}_{21.} - \bar{y}_{22.} \end{bmatrix}$$

(65)

$$= \begin{bmatrix} y_{11} - \bar{y}_{...} - \bar{y}_{11.} + \bar{y}_{1..} \\ y_{12} - \bar{y}_{...} - \bar{y}_{12.} + \bar{y}_{1..} \\ y_{13} - \bar{y}_{...} - \bar{y}_{13.} + \bar{y}_{1..} \\ y_{21} - \bar{y}_{...} - \bar{y}_{21.} + \bar{y}_{2..} \\ y_{22} - \bar{y}_{...} - \bar{y}_{22.} + \bar{y}_{2..} \\ y_{23} - \bar{y}_{...} - \bar{y}_{23.} + \bar{y}_{2..} \end{bmatrix}$$

Demonstração da Igualdade (39)

A diferença de somatórios triplos

$$\sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{i...} - \bar{y}_{.ij} + \bar{y}_{i..} \right)^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2$$

pode escrever-se como

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m & \left(y_{ijk}^2 - y_{ijk} \bar{y}_{i...} - y_{ijk} \bar{y}_{.ij} + y_{ijk} \bar{y}_{i..} - \bar{y}_{i...} y_{ijk} + y_{i...}^2 + \bar{y}_{i...} \bar{y}_{.ij} - \bar{y}_{i...} \bar{y}_{i..} - \bar{y}_{.ij} y_{ijk} + \right. \\ & \left. + \bar{y}_{.ij} y_{i...} + \bar{y}_{.ij}^2 - \bar{y}_{i..} \bar{y}_{.ij} + y_{ijk} \bar{y}_{i..} - \bar{y}_{i..} \bar{y}_{i...} - \bar{y}_{i..} \bar{y}_{.ij} + \bar{y}_{i..}^2 - y_{ijk}^2 - \bar{y}_{ij.}^2 + 2y_{ijk} \bar{y}_{ij.} \right) \end{aligned}$$

e, eliminando o somatório triplo,

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^3 & \left(-\bar{y}_{i...} \sum_{k=1}^m y_{ijk} - \bar{y}_{.ij} \sum_{k=1}^m y_{ijk} + \bar{y}_{i..} \sum_{k=1}^m y_{ijk} - \bar{y}_{i...} \sum_{k=1}^m y_{ijk} + y_{i...}^2 m + \bar{y}_{i...} \bar{y}_{.ij} m - \right. \\ (66) \quad & \left. - \bar{y}_{i...} \bar{y}_{i..} m - \bar{y}_{.ij} \sum_{k=1}^m y_{ijk} + \bar{y}_{ij.} \sum_{k=1}^m y_{ijk} + \bar{y}_{ij.} \bar{y}_{i...} m - \bar{y}_{i..} \bar{y}_{ij.} m \right. \\ & \left. + \bar{y}_{i..} \sum_{k=1}^m y_{ijk} - \bar{y}_{i..} \bar{y}_{i...} m - \bar{y}_{i..} \bar{y}_{.ij} m + \bar{y}_{i..}^2 m + 2\bar{y}_{ij.} \sum_{k=1}^m y_{ijk} \right) \end{aligned}$$

Atendendo, mais uma vez, ao facto de $\sum_{k=1}^m y_{ijk} = m\bar{y}_{ij.}$, a expressão (66) pode simplificar-se como:

$$\begin{aligned} m \sum_{i=1}^2 \sum_{j=1}^3 & \left(-\bar{y}_{i...} \bar{y}_{ij.} - \bar{y}_{ij.}^2 + \bar{y}_{i..} \bar{y}_{ij.} - \bar{y}_{i...} \bar{y}_{ij.} + y_{i...}^2 + \bar{y}_{i...} \bar{y}_{ij.} - \bar{y}_{i...} \bar{y}_{i..} - \bar{y}_{ij.}^2 + \bar{y}_{ij.} \bar{y}_{i...} - \right. \\ & \left. - \bar{y}_{i..} \bar{y}_{ij.} + \bar{y}_{i..} \bar{y}_{ij.} - \bar{y}_{i..} \bar{y}_{i...} - \bar{y}_{i..} \bar{y}_{ij.} + \bar{y}_{i..}^2 + 2\bar{y}_{ij.}^2 \right) \end{aligned}$$

ou ainda

$$m \sum_{i=1}^2 \sum_{j=1}^3 \left(\bar{y}_{i...}^2 - 2\bar{y}_{i...} \bar{y}_{i..} + \bar{y}_{i..}^2 \right)$$

e, finalmente,

$$3m \sum_{i=1}^2 \left(\bar{y}_{i...} - \bar{y}_{...} \right)^2 .$$

Demonstração das Igualdades (42) e (43)

A tradução matricial do modelo (41)

$$y_{ijk} = \beta_o + \beta_1 H_{1k} + \beta_2 [I_{1k} - 2(H_{1k} I_{1k})] + \beta_3 [I_{2k} - 2(H_{1k} I_{2k})] + u_{ijk}$$

$$; i = 1, 2 ; j = 1, 2, 3 ; k = 1, 2, \dots, n_i ; u_{ijk} \sim \text{IIN}(0, \sigma^2)$$

pode ser representada em extensão como:

$$(67) \quad \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_o \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{23} \end{bmatrix}$$

donde resulta:

$$(68) \quad X'X = \begin{bmatrix} 1' & 1' & 1' & 1' & 1' & 1' \\ 1' & 1' & 1' & 0' & 0' & 0' \\ -1' & 0' & 0' & 1' & 0' & 0' \\ 0' & -1' & 0' & 0' & 1' & 0' \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 6m & 3m & 0 & 0 \\ 3m & 3m & -m & -m \\ 0 & -m & 2m & 0 \\ 0 & -m & 0 & 2m \end{bmatrix}$$

$$(69) \quad X'y = \begin{bmatrix} 1' & 1' & 1' & 1' & 1' & 1' \\ 1' & 1' & 1' & 0' & 0' & 0' \\ -1' & 0' & 0' & 1' & 0' & 0' \\ 0' & -1' & 0' & 0' & 1' & 0' \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 6m\bar{y} \\ m(\bar{y}_{11} - 2\bar{y}_{12} + \bar{y}_{13}) \\ m(-\bar{y}_{11} + \bar{y}_{21}) \\ m(-\bar{y}_{12} + \bar{y}_{22}) \end{bmatrix}$$

$$(70) \quad (X'X)^{-1} = \begin{bmatrix} 2/3 & -1 & -1/2 & -1/2 \\ -1 & 2 & 1 & 1 \\ -1/2 & 1 & 1 & 1/2 \\ -1/2 & 1 & 1/2 & 1 \end{bmatrix}$$

O vector EMQO é então dado por:

$$(71) \quad \hat{\beta} = \begin{bmatrix} (24\bar{y}_{...} - 3\bar{y}_{11.} - 3\bar{y}_{12.} - 6\bar{y}_{13.} - 3\bar{y}_{21.} - 3\bar{y}_{22.})/6 \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + 2\bar{y}_{13.} + \bar{y}_{21.} + \bar{y}_{22.} \\ (-6\bar{y}_{...} + \bar{y}_{12.} + 2\bar{y}_{13.} + 2\bar{y}_{21.} + \bar{y}_{22.})/2 \\ (-6\bar{y}_{...} + \bar{y}_{11.} + 2\bar{y}_{13.} + \bar{y}_{21.} + 2\bar{y}_{22.})/2 \end{bmatrix}$$

O vector de resíduos define-se como:

$$\begin{bmatrix} \hat{u}_{11} \\ \hat{u}_{12} \\ \hat{u}_{13} \\ \hat{u}_{21} \\ \hat{u}_{22} \\ \hat{u}_{23} \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} - \begin{bmatrix} 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} (24\bar{y}_{...} - 3\bar{y}_{11.} - 3\bar{y}_{12.} - 6\bar{y}_{13.} - 3\bar{y}_{21.} - 3\bar{y}_{22.})/6 \\ -6\bar{y}_{...} + \bar{y}_{11.} + \bar{y}_{12.} + 2\bar{y}_{13.} + \bar{y}_{21.} + \bar{y}_{22.} \\ (-6\bar{y}_{...} + \bar{y}_{12.} + 2\bar{y}_{13.} + 2\bar{y}_{21.} + \bar{y}_{22.})/2 \\ (-6\bar{y}_{...} + \bar{y}_{11.} + 2\bar{y}_{13.} + \bar{y}_{21.} + 2\bar{y}_{22.})/2 \end{bmatrix}$$

$$(72) \quad = \begin{bmatrix} y_{11} - \bar{y}_{...} - \bar{y}_{11.} + \bar{y}_{.1.} \\ y_{12} - \bar{y}_{...} - \bar{y}_{12.} + \bar{y}_{.2.} \\ y_{13} - \bar{y}_{...} - \bar{y}_{13.} + \bar{y}_{.3.} \\ y_{21} - \bar{y}_{...} - \bar{y}_{21.} + \bar{y}_{.1.} \\ y_{22} - \bar{y}_{...} - \bar{y}_{22.} + \bar{y}_{.2.} \\ y_{23} - \bar{y}_{...} - \bar{y}_{23.} + \bar{y}_{.3.} \end{bmatrix}$$

Demonstração da Igualdade (47)

A forma de demonstrar esta igualdade é a mesma que se utilizou para mostrar a igualdade (39). Assim,

$$\begin{aligned}
 & \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{...} - \bar{y}_{ij.} + \bar{y}_{.j.} \right)^2 - \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk} - \bar{y}_{ij.} \right)^2 = \\
 & = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^m \left(y_{ijk}^2 - y_{ijk} \bar{y}_{...} - y_{ijk} \bar{y}_{ij.} + y_{ijk} \bar{y}_{.j.} - \bar{y}_{...} y_{ijk} + \bar{y}_{...}^2 + \bar{y}_{...} \bar{y}_{ij.} - \bar{y}_{...} \bar{y}_{.j.} - \bar{y}_{ij.} y_{ijk} + \right. \\
 & \quad \left. + \bar{y}_{ij.} \bar{y}_{...} + \bar{y}_{ij.}^2 - \bar{y}_{.j.} \bar{y}_{ij.} + y_{ijk} \bar{y}_{.j.} - \bar{y}_{.j.} \bar{y}_{...} - \bar{y}_{.j.} \bar{y}_{ij.} + \bar{y}_{.j.}^2 - y_{ijk}^2 - \bar{y}_{ij.}^2 + 2 y_{ijk} \bar{y}_{ij.} \right) \\
 & = \sum_{i=1}^2 \sum_{j=1}^3 \left(-\bar{y}_{...} \sum_{k=1}^m y_{ijk} - \bar{y}_{ij.} \sum_{k=1}^m y_{ijk} + \bar{y}_{.j.} \sum_{k=1}^m y_{ijk} - \bar{y}_{...} \sum_{k=1}^m y_{ijk} + \bar{y}_{...}^2 m + \bar{y}_{...} \bar{y}_{ij.} m - \right. \\
 & \quad (73) \quad \left. - \bar{y}_{...} \bar{y}_{.j.} m + \bar{y}_{ij.} \sum_{k=1}^m y_{ijk} + \bar{y}_{ij.} \bar{y}_{...} m - \bar{y}_{.j.} \bar{y}_{ij.} m + \bar{y}_{.j.} \sum_{k=1}^m y_{ijk} - \bar{y}_{.j.} \bar{y}_{...} m - \right. \\
 & \quad \left. - \bar{y}_{.j.} \bar{y}_{ij.} m + \bar{y}_{.j.}^2 m + 2 \bar{y}_{ij.} \sum_{k=1}^m y_{ijk} \right)
 \end{aligned}$$

Novamente, a igualdade $\sum_{k=1}^m y_{ijk} = m \bar{y}_{ij.}$, permite que se expresse (73) como

$$\begin{aligned}
 & m \sum_{i=1}^2 \sum_{j=1}^3 \left(-\bar{y}_{...} \bar{y}_{ij.} - \bar{y}_{ij.}^2 + \bar{y}_{.j.} \bar{y}_{ij.} - \bar{y}_{...} \bar{y}_{ij.} + \bar{y}_{...}^2 + \bar{y}_{...} \bar{y}_{ij.} - \bar{y}_{...} \bar{y}_{.j.} - \bar{y}_{ij.}^2 + \bar{y}_{ij.} \bar{y}_{...} - \right. \\
 & \quad \left. - \bar{y}_{.j.} \bar{y}_{ij.} + \bar{y}_{.j.} \bar{y}_{ij.} - \bar{y}_{.j.} \bar{y}_{...} - \bar{y}_{.j.} \bar{y}_{ij.} + \bar{y}_{.j.}^2 + 2 \bar{y}_{ij.} \right)
 \end{aligned}$$

e, após simplificações,

$$\begin{aligned}
 & = m \sum_{i=1}^2 \sum_{j=1}^3 \left(\bar{y}_{...}^2 - 2 \bar{y}_{...} \bar{y}_{.j.} + \bar{y}_{.j.}^2 \right) \\
 & = 2m \sum_{j=1}^3 \left(\bar{y}_{.j.} - \bar{y}_{...} \right)^2.
 \end{aligned}$$

EASYcare: um Sistema de Avaliação de Idosos (qualidades psicométricas)

Autores:
Liliana Sousa
Helena Galante
e
Daniela Figueiredo

EASYCARE: UM SISTEMA DE AVALIAÇÃO DE IDOSOS (QUALIDADES PSICOMÉTRICAS)

EASY-CARE: ELDERLY ASSESSMENT SYSTEM

Autores: Liliana Sousa

- Psicóloga, Professora Auxiliar do Departamento de Ciências da Educação da Universidade de Aveiro

Helena Galante

- Licenciada em Matemática Aplicada e Computação, bolsista de iniciação à investigação no Departamento de Ciências da Educação da Universidade de Aveiro

Daniela Figueiredo

- Licenciada em Ciências da Educação, bolsista de iniciação à investigação no Departamento de Ciências da Educação da Universidade de Aveiro

RESUMO:

- Este estudo tem como objectivo principal caracterizar a qualidade de vida e bem estar dos idosos (com 75 ou mais anos). O instrumento utilizado na recolha dos dados foi o EASYcare (Sistema de Avaliação dos Idosos), tendo-se inquirido uma amostra de 516 sujeitos. A análise estatística baseou-se na Análise em Componentes Principais e Análise de *Clusters*. Os resultados revelaram que o EASYcare possui boas qualidades psicométricas (α de Cronbach = 0,86) e permitiram encontrar quatro factores: competências cognitivas, saúde e qualidade de vida, mobilidade e actividades de vida diária. A análise de clusters identificou quatro grupos de idosos: autónomos, autónomos com problemas cognitivos, dependentes com problemas cognitivos ligeiros e dependentes com graves problemas cognitivos. Verifica-se que 52,5% dos idosos é autónomo e apenas 19,7% se inclui nos dois grupos de dependentes.

PALAVRAS-CHAVE:

- *Idosos; Qualidade de Vida e Bem Estar; Análise de Componentes Principais; Análise de Clusters.*

ABSTRACT:

- The present investigation aims to characterize the older people (aged over 75 years old) quality of life and well being. The EASYcare (Elderly Assessment System) was used to collect data. The sample comprises 516 subjects. The statistical analysis was based on Principal Components Analysis and Cluster Analysis. EASYcare's reveal good psychometric proprieties (Cronbach α = 0,86). The principal component analysis identified 4 factors: cognitive competencies, health and quality of life, motor competencies and daily living autonomy. The cluster analysis shows 4 groups of elder people: autonomous, autonomous with cognitive

impairment, dependent with moderate cognitive impairment and dependent with severe cognitive impairment. 52,5% of the subjects fills in the independent group and only 19,7% are included within the two groups of dependent people.

KEY-WORDS:

- *Elderly; Quality of Life and Well Being; Principal Component Analysis; Cluster Analysis.*

1. INTRODUÇÃO

O envelhecimento da população na Europa, em geral, e em Portugal, particularmente, é um facto que tem sido por demais referido e é do conhecimento da sociedade. Importa pois desenvolver meios para melhor atender às dificuldades do crescente grupo de idosos. Parece-nos que a obtenção de dados de caracterização da qualidade de vida e bem estar dos idosos pode ser fundamental para dinamizar medidas adequadas a esta população.

Nesta pesquisa recolhemos dados junto de uma amostra de 516 sujeitos, aos quais aplicámos o EASYcare (Sistema de Avaliação de Idosos). Trata-se de uma escala desenvolvida com o intuito de caracterizar a qualidade de vida e bem-estar da população idosa, com 75 anos ou mais. O instrumento utilizado é a versão portuguesa do EASYcare (Sousa & Figueiredo, 2000; Figueiredo & Sousa, 2000) que avalia a percepção dos idosos em relação às suas capacidades (ser capaz ...), não indicando as competências (saber-fazer). Trata-se de um instrumento de avaliação multi-dimensional, desenvolvido para avaliar as necessidades dos idosos a nível social e de saúde. Neste instrumento pontuações mais elevadas significam maior incapacidade.

Este instrumento foi inicialmente desenvolvido no âmbito de um projecto financiado pela União Europeia (European Prototype for Integrated Care – EPIC). A construção da escala realizou-se a partir de outros instrumentos que avaliavam apenas algumas dimensões consideradas pertinentes. A inovação do EASYcare centra-se no facto de integrar, numa só escala, itens relativos às várias dimensões da qualidade de vida e bem-estar do idoso.

Esta escala pretende ajudar os técnicos a melhorar os cuidados que prestam ao idoso. Todavia os dados obtidos podem, ainda, ser utilizados para avaliar problemas diversificados na população, necessidades e objectivos, que auxiliam a gestão dos serviços, a investigação e a formulação de políticas.

O EASYcare está, de momento, disponível em 16 países da União Europeia e na Polónia. Apesar de neste momento ainda não existirem dados passíveis de comparação entre os 17 Países envolvidos no projecto, dentro em breve essa possibilidade existirá, permitindo um conhecimento abrangente das necessidades dos idosos a nível europeu.

2. OBJECTIVOS E METODOLOGIA

O principal objectivo deste estudo é caracterizar a qualidade de vida e bem-estar dos idosos, com 75 anos ou mais. Pretende-se, também, conhecer melhor a relação entre as variáveis envolvidas no EASYcare e, desta forma, avaliar as suas qualidades psicométricas.

Este instrumento efectua a caracterização dos idosos com base nas seguintes 26 variáveis: visão (“vê bem?”); audição (“ouve bem?”); mastigação (“tem dificuldade em mastigar a comida?”); fala (“sente dificuldade em fazer-se compreender, devido a problemas com a fala/linguagem?”); saúde (“considera que a sua saúde é: excelente, ... fraca?”); solidão (“sente-se sozinho: nunca, ... sempre?”); habitação (“a sua habitação é: excelente, ... fraca?”); realizar o trabalho doméstico; preparar as próprias refeições; ir às compras; administrar o próprio dinheiro; usar o telefone; tomar os medicamentos; sair de casa e caminhar na rua; movimentar-se dentro de casa; subir e descer escadas; deslocar-se da cama para a cadeira; utilizar a sanita; usar a banheira ou chuveiro; cuidar da aparência pessoal; vestir-se; alimentar-se; incontinência urinária; incontinência fecal; depressão; diminuição cognitiva.

Os dados foram recolhidos por enfermeiros e técnicos de serviço social, entre Maio e Setembro de 2000. Os técnicos tiveram formação para aplicarem a escala de acordo com as normas preconizadas e de forma estandardizada. Esta formação realizou-se com base no GATT (Geriatric Assessment Technological Training) (Sousa & Figueiredo, 2000).

A amostra considerada é constituída por 516 sujeitos, apesar de a escala ter sido administrada a 544 idosos. A eliminação de 28 inquiridos deveu-se ao facto de eles apenas terem dado informação relativa aos dados de identificação. Este facto revela dois dos limites do EASYcare: não pode ser aplicado a sujeitos muito dependentes, incapazes de se expressar; não pode ser administrado a idosos com baixo nível cognitivo.

Passamos, então, a caracterizar a amostra considerada nesta análise. Assim, os sujeitos provêm de 5 Distritos: Porto (18%), Coimbra (34,5%), Viana do Castelo (9,7%), Viseu (15,9%), Santarém (1%), Lisboa (4,5%) e Setúbal (1%).

Os inquiridos são 55,8% do sexo feminino e 44,2% do sexo masculino. Quanto à distribuição da amostra por escalões etários verifica-se, naturalmente, a diminuição do número de elementos da amostra com o aumento da idade (Quadro I).

QUADRO I – DISTRIBUIÇÃO DA AMOSTRA POR ESCALÕES ETÁRIOS

Escalão Etário (anos)	Amostra (%)
75 - 79	56,7
80 - 84	27,2
85 - 89	12,0
90 - 94	3,5
95 - 98	0,6

Relativamente à qualificação académica a maioria dos inquiridos tem como habilitações literárias a 4ª classe, para a época em que, provavelmente, ingressaram na actividade profissional, em geral, teriam habilitações dentro da média (Quadro II).

QUADRO II – QUALIFICAÇÃO ACADÉMICA		
Qualificação Académica	Amostra (%)	nº de indivíduos
Não sabe ler nem escrever	13,6	70
Sabe ler e escrever mas sem diploma	6,6	34
1º CEB/4ª Classe	46,3	239
2º CEB/6º	5,8	30
3º CEB/9º ano/5º ano liceal ou equivalente	15,7	81
Ensino secundário ou equivalente	10,7	55
Ensino superior	0,8	4
Não - respostas	0,3	3

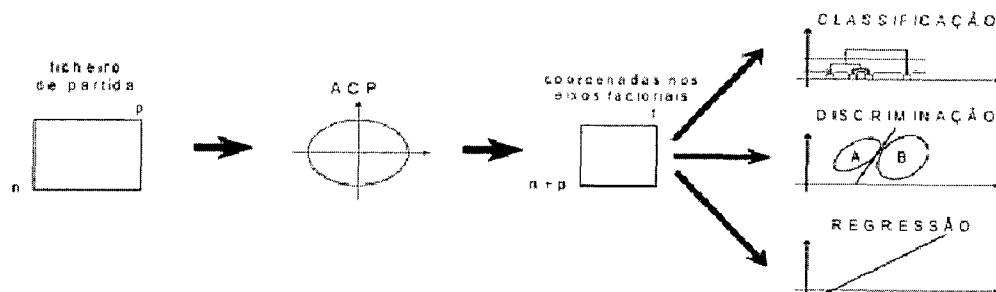
Quanto às profissões exercidas pelos sujeitos da amostra, verifica-se que 33,5% eram operários, 28,1% são domésticas e 20,2% empregados de escritório, comércio e serviços; todas as outras categorias são residuais.

3. ANÁLISE ESTATÍSTICA E RESULTADOS

3.1. EASYCARE: QUALIDADES PSICOMÉTRICAS E FACTORES QUE AVALIA

A existência de muitos parâmetros (variáveis) numa análise dificulta o estudo e compreensão de um sistema. A análise em componentes principais (ACP) é uma das técnicas de análise exploratória de dados que permite definir grupos de variáveis correlacionadas. Calculando as correlações entre as variáveis poder-se-ão encontrar modelos explicativos mais simples através de regressões, classificações ou discriminações (como se esquematiza na Figura 1).

FIGURA 1. ANÁLISE EM COMPONENTES PRINCIPAIS

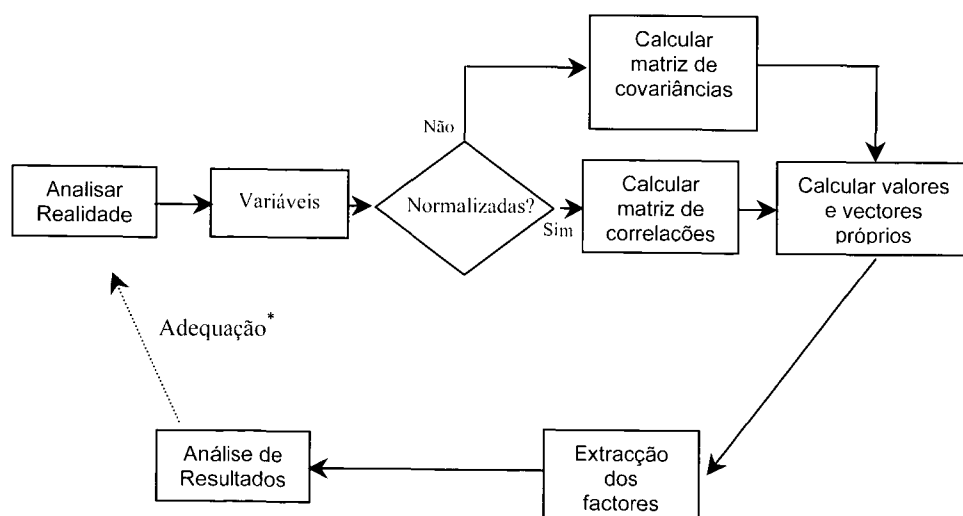


Adaptado de CVRM – Centro de Geo-Sistemas do Instituto Superior Técnico (Janeiro 2001)

O objectivo da ACP é a obtenção de um pequeno número de combinações lineares (componentes principais) de um conjunto de variáveis originais. Trata-se de um dos métodos mais simples de análise multidimensional, pois não tem grandes pressupostos estatísticos e espera-se somente que as variáveis originais sejam correlacionadas (quanto mais correlacionadas, melhor é o resultado da análise). Basicamente, o procedimento nesta técnica é partir de um conjunto de m variáveis e encontrar combinações lineares de modo a serem produzidos índices não correlacionados. Este não correlacionamento é importante pois significa que descrevem diferentes “dimensões” do conjunto dos dados originais (Manly, 1994). Foram estas as motivações para a aplicação da ACP à análise do EASYcare.

Resumidamente, a ACP deve desenvolver-se de acordo com as fases esquematizadas na Figura 2.

FIGURA 2. FASES DA ACP



* Depois da obtenção dos resultados há um processo de adequação com a realidade, se os resultados obtidos se adequam a análise termina, caso contrário o procedimento repete-se até se atingir a adequação.

A análise estatística do EASYcare foi elaborada com o auxílio do *software STATISTICA*.

No caso em estudo, as variáveis estão moderadamente correlacionadas, os valores variam entre 0,00 e 0,79. Tendo em conta os resultados obtidos, este facto em nada prejudicou a análise, até porque um “olhar” cuidadoso das correlações mostra que:

- as variáveis com correlações mais baixas realmente não têm associação, por exemplo, ouvir bem e ser capaz de tomar a medicação apresentam uma correlação de 0,00, o que parece óbvio;
- as variáveis que envolvem capacidades próximas estão bem correlacionadas, por exemplo, usar a sanita e deslocar-se da cama para a cadeira, obtêm uma correlação elevada (0,79) o que é lógico, uma vez que os movimentos envolvidos na realização das tarefas são idênticos.

As 26 variáveis, acima descritas, foram submetidas a ACP, optando-se pela eliminação de 6 (visão, audição, mastigação, fala, cuidar da aparência pessoal e incontinência urinária) por não se identificarem com nenhum dos factores. Extraíram-se 4 factores (componentes principais) que explicam 62,2% da variância (Quadro III).

QUADRO III. ACP – FACTORES, VALORES PRÓPRIOS E VARIÂNCIAS

Factores	Valores próprios	% Variância total	Valores próprios acumulados	% Variância total acumulada
1	8,5	42,5	8,5	42,5
2	1,6	8,1	10,1	50,6
3	1,2	6,1	11,3	56,7
4	1,1	5,5	12,4	62,2

Como era de esperar, o primeiro factor é aquele que explica mais variância (42,5%), os restantes factores apresentam valores muito idênticos (entre 5,5% e 8,1%). Numa primeira fase procedeu-se ao cálculo das contribuições de cada variável para cada factor sem qualquer rotação. Nesta análise destacaram-se dois grupos distintos, mas devido à pouca clareza na interpretação dos resultados desta primeira análise procedeu-se ao cálculo das contribuições de cada variável para cada factor com rotação varimax* (normalizada) e obtiveram-se os seguintes resultados (quadro IV):

* Rotação Varimax é uma rotação ortogonal, mantém o ângulo de 90° entre os eixos, e os factores mantêm-se independentes. Esta rotação tenta maximizar a variância nos novos eixos.

QUADRO IV – CONTRIBUIÇÕES DAS VARIÁVEIS PARA CADA FACTOR

Variáveis	Factor 1	Factor 2	Factor 3	Factor 4
	Actividades de vida diária	Saúde e qualidade de vida	Competências cognitivas	Mobilidade
Saúde	0,23	0,59	0,04	0,14
Solidão	0,09	0,66	0,2	-0,01
Habitação	-0,11	0,54	-0,11	0,08
Doméstico	0,8	0,13	0,19	0,2
Preparar refeições	0,78	0,03	0,18	0,15
Ir às compras	0,75	0,17	0,21	0,23
Gerir o dinheiro	0,53	0,08	0,6	0,28
Usar o telefone	0,33	0,10	0,64	0,28
Tomar os medicamentos	0,47	0,03	0,64	0,23
Sair de casa e andar na rua	0,68	0,12	0,12	0,50
Movimentar-se dentro de casa	0,54	0,09	0,05	0,64
Subir e descer escadas	0,43	0,29	0,07	0,54
Sair da cama	0,35	0,06	0,17	0,80
Usar a sanita	0,24	0,01	0,32	0,79
Usar o chuveiro	0,66	0,14	0,31	0,32
Vestir-se	0,35	-0,02	0,32	0,7
Alimentar-se	0,12	0,03	0,52	0,43
Incontinência fecal	0,01	0,13	0,37	0,54
Depressão	0,2	0,73	0,2	-0,02
Diminuição cognitiva	0,09	0,12	0,71	0,11
Variância explicada	4,25	1,84	2,65	3,68
Prp total	0,21	0,09	0,13	0,18

Procedeu-se à interpretação das contribuições de cada variável para cada factor, o que levou aos seguintes agrupamentos:

- Factor 1 chamou-se “Actividades de Vida Diária”, já que compreende as variáveis relativa à independência do idoso no seu quotidiano, tais como: fazer o trabalho doméstico, preparar refeições, ir às compras, andar na rua e usar o chuveiro.
- Factor 2 denominou-se “Saúde e Qualidade de Vida” por abranger as variáveis saúde, solidão, habitação e depressão, as quais estão relacionadas com o estado de saúde em geral e com as componentes que favorecem o bem estar do idoso.
- Factor 3 designou-se por “Competências Cognitivas” uma vez que inclui as variáveis gerir o dinheiro, usar o telefone, tomar os medicamentos,

alimentar-se e diminuição cognitiva, tarefas que exigem ou envolvem a capacidade de raciocínio

- Factor 4 indicou-se como “Mobilidade” por envolver as variáveis movimentar-se dentro de casa, subir e descer escadas, sair da cama, usar a sanita, vestir-se e incontinência fecal que exigem claramente capacidades motoras para a sua realização.

Por fim, e de modo a estudarmos a coerência e a uniformidade das respostas do inquérito, ou seja, a consistência interna da análise, foram calculados os α 's de Cronbach (Quadro V). Estes apresentam valores bastante satisfatórios para os grupos de Actividades de Vida Diária e Mobilidade. Para os outros dois factores (Competências Cognitivas e da Saúde e Qualidade de Vida) os α 's de Cronbach apresentam valores globais moderados. No entanto, o mesmo não se verifica com os itens de cada um destes grupos, na medida em que, apresentam valores significativamente elevados (cf. Quadro V). O valor do α para a escala global é bastante bom (0,92).

QUADRO V - CONSISTÊNCIA INTERNA (COEFICIENTE DE CRONBACH - α)

Variável	α 's de Cronbach
Factor 1 – Actividades vida diária	0,87
Fazer o trabalho doméstico	0,84
Preparar as refeições	0,84
Ir às compras	0,84
Sair de casa e caminhar na rua	0,84
Usar o chuveiro	0,83
Factor 2 – Saúde e qualidade de vida	0,57
Saúde	0,85
Solidão	0,85
Habitação	0,86
Depressão	0,85
Factor 3 – Competências cognitivas	0,45
Gerir o dinheiro	0,84
Usar o telefone	0,85
Tomar os medicamentos	0,85
Alimentar-se	0,85
Diminuição Cognitiva	0,91
Factor 4 - Mobilidade	0,87
Movimentar-se dentro de casa	0,84
Subir e descer escadas	0,85
Deslocar-se da cama para a cadeira	0,84
Usar a sanita	0,84
Vestir-se	0,84
Incontinência fecal	0,85
Escala global	0,86
α normalizado	0,92

3.2. GRUPOS DE IDOSOS DE ACORDO COM AS (IN)CAPACIDADES

Pretendeu-se estudar grupos de idosos de acordo com as incapacidades. Para tal, foi necessário recorrer a uma análise que permitisse a classificação de dados. A metodologia considerada consistiu na análise de *clusters* (análise classificatória). Trata-se de uma técnica de análise exploratória de dados para resolver problemas de classificação.

Em geral, a análise de *clusters* pretende agrupar um determinado número de dados em classes de elementos que apresentam semelhanças entre si. Uma das possíveis abordagens para a análise de *clusters* consiste na resolução de um problema de optimização, em que se pretende maximizar a semelhança intra grupos e minimizar a semelhança inter grupos.

Há vários tipos de análise de *clusters*, nomeadamente hierárquicos e particionais (não hierárquicos). O procedimento do método hierárquico é juntar sucessivamente pequenos *clusters* formando grandes *clusters*, ou dividindo grandes *clusters*. O resultado final do algoritmo é uma árvore de *clusters* chamada dendrograma, que mostra como os *clusters* estão relacionados. Cortando o dendrograma ao nível desejado o *cluster* de dados da separação do grupo é obtido.

O método particional por sua vez produz uma classificação particionando o conjunto de dados, dando origem a grupos não sobrepostos e cujas relações entre eles são não hierárquicas. Ao contrário dos métodos hierárquicos, os métodos que realizam partições dos dados não requerem que a alocação de um objecto num *cluster* seja definitiva. Esta técnica de partição dos dados é baseada na ideia de otimizar algum critério pré-definido. O uso destas técnicas de partição usualmente assume que o número final de *clusters* seja conhecido à partida, apesar de haver alguns métodos que permitem que esse número varie ao longo da análise (Dillon, 1984).

Para o estudo em causa adoptou-se pelo método particional K-means. O objectivo deste método é encontrar uma partição de dados tal que minimize o quadrado da distância euclidiana ao centro do *cluster*, procurando assim grupos homogéneos.

Os membros de cada classe (*cluster*) são reorganizados através de um processo iterativo. Através análise de *clusters* encontraram-se grupos de idosos, de acordo com os seus níveis de (in)capacidade nos 4 factores (cf. Quadro VI). Optámos por constituir 4 *clusters* pois foi a solução que se mostrou mais interessante (relembramos que neste instrumentos valores mais elevados significam maior incapacidade).

QUADRO VI - GRUPOS DE IDOSOS DE ACORDO COM AS INCAPACIDADES (MÉDIAS)

Factores (média)	Grupo 1 (N=41) Dependentes com graves problemas cognitivos	Grupo 2 (N=135) Autónomos com problemas cognitivos	Grupo 3 (N=271) Autónomos	Grupo 4 (N=61) Dependentes com problemas cognitivos ligeiros
Actividades de vida diária (1,2)	3,4 (++)	0,7 (-)	0,5 (--)	3,9 (++)
Saúde e qualidade de vida (2,5)	2,8 (+)	2,6 (=)	2,4 (=)	3,0 (+)
Competências cognitivas (1,7)	6,0 (++)	2,5 (+)	0,5 (--)	2,0 (+)
Mobilidade (0,7)	2,8 (++)	0,3 (-)	0,2 (-)	2,4 (+)

Nota: os sinais utilizados (--), (-), (=), (+) e (++) pretendem compara cada média, com a média da amostra global.

O primeiro elemento a destacar é que o grupo dos “autónomos” inclui 52,5% do total de sujeitos da amostra estudada. Trata-se de um grupo que não demonstra incapacidades em nenhuma das áreas consideradas.

O grupo designado como “autónomos com problemas cognitivos” abrange 26,2% dos elementos da amostra. Trata-se de um núcleo que se apresenta bem ao nível da motor, autónomo, percepcionando a sua saúde e qualidade de vida medianamente, mas com dificuldades em tarefas que envolvam esforços de natureza cognitiva.

O grupo dos “dependentes com problemas cognitivos ligeiros”, inclui 11,8% da amostra. Os sujeitos aqui incluídos apresentam incapacidade ligeira na área cognitiva e dificuldades a nível da mobilidade, actividades de vida diária e percepção de qualidade de vida e saúde.

O grupo dos “dependentes com graves problemas cognitivos”, envolve apenas 7,9% dos sujeitos. Neste *cluster* encontramos idosos com elevado grau de dependência a nível motor e de autonomia, os quais percepcionam a sua saúde e qualidade de vida como fraca e, principalmente, têm já uma considerável diminuição cognitiva.

Tentámos, por fim, perceber se nestes grupos de idosos se podia distinguir alguma diferença ou tendência em termos do género, idade e residência. Para tal procedemos ao cálculo das frequências esperadas e comparámos com as frequências observadas.

Em relação ao género verifica-se (Quadro - VII) que no grupo dos “autónomos” há menos mulheres e mais homens, relativamente ao esperado. Isto pode indicar uma tendência para que as mulheres nesta idade apresentem um índice superior de limitações.

QUADRO VII – GRUPOS DE IDOSOS *VERSUS* GÉNERO

		Feminino	Masculino
Dependentes com graves problemas cognitivos	Frequência esperadas	23	18
	Frequência observadas	31	10
Autónomos com problemas cognitivos	Frequência esperadas	75	60
	Frequência observadas	85	50
Autónomos	Frequência esperadas	151	120
	Frequência observadas	132	139
Dependentes com problemas cognitivos ligeiros	Frequência esperadas	34	27
	Frequência observadas	35	26

Os resultados relativos à idade não apresentam surpresas, apenas se percebe que o grupo dos autónomos predomina no escalão etário mais baixo (75 – 79 anos). No escalão etário dos 85-89 anos nota-se a diminuição clara dos sujeitos autónomos e aumento dos dependentes. No que concerne à variável que identifica “com quem vive o idoso” nota-se que os idosos sozinhos são os mais autónomos.

4. DISCUSSÃO E CONCLUSÕES

Um primeiro dado a realçar diz respeito ao facto do EASYcare ter apresentado boas qualidades psicométricas, já que a consistência interna obtida foi boa.

Os resultados obtidos modificaram em alguns pontos a nossa percepção dos aspectos avaliados pela escala, apresentamos alguns casos:

- A equipa envolvida na construção da escala apenas previu um item relativo às competências cognitivas (diminuição cognitiva). A partir desta análise verificou-se a associação com outras quatro variáveis (gerir o dinheiro, usar o telefone, tomar os medicamentos e alimentar-se).
- Gerir o dinheiro, usar o telefone e tomar os medicamentos estavam, inicialmente, associados com a autonomia pessoal, mas, na verdade, uma análise mais cuidada permite entender que estas tarefas exigem alguma capacidade de abstracção.
- Outro aspecto de destaque é a separação entre competências motoras e actividades da vida diária, também não prevista inicialmente com estes contornos. As actividades incluídas nas competências motoras incluem

aspectos dependentes da motricidade mas de carácter mais impessoal. Os aspectos incluídos no factor actividades da vida diária são aqueles nos quais a incapacidade é sentida de forma mais intensa pois coloca o idoso na dependência de outros para a realização de tarefas mais íntimas.

No que concerne aos grupos de idosos identificados parece-nos interessante o facto dos problemas cognitivos surgirem, consistentemente, como um elemento diferenciador dos grupos. Isto leva-nos a colocar a hipótese de que um dos aspectos a ter sempre presente quando se intervém junto da população idosa é a diminuição das competências cognitivas.

Um dos resultados mais “agradáveis” é aquele que nos mostra que 52,5% dos idosos inquiridos se encontram no grupo dos autónomos, apresentando raras limitações. É óbvio que se mostram um pouco descontentes com o seu estado de saúde e qualidade de vida, provavelmente, por compararem com aquilo que se passava há uns anos atrás. Também é interessante que apenas 19,7% dos inquiridos se incluem nos dois grupos de dependentes.

Menos “agradável” é a constatação de que 34,1% estão inseridos em grupos com tendência para diminuição cognitiva (moderada ou grave).

Em síntese, poderemos dizer que o panorama da qualidade de vida e bem estar dos idosos é bastante boa, sendo que se tem de prestar particular atenção à diminuição das competências cognitivas.

BIBLIOGRAFIA

- DILLON, W.; GOLDSTEIN, M. (1984). *Multivariate Analysis and Methods and Applications*. New York: John Wiley & Sons.
- KRZANOWSKI, W.; MARRIOT, F. (1995). *Multivariate Analysis, Part 2, Classification, covariance structures and repeated measurements*. New York: Halsted Press.
- MANLY, B. (1994). *Multivariate Statistical Methods. A primer*. London: Chapman & Hall.
- SOUSA, L.; FIGUEIREDO, D. (2000). Facilitar os cuidados aos idosos: uma escala de avaliação da qualidade de vida e bem estar. *Psychologica*, 25, 19-24.
- FIGUEIREDO, D.; SOUSA, L. (2001). EASYcare: um instrumento de avaliação da qualidade de vida e bem estar do idoso. *Revista Geriatria*, nº 130. Janeiro.
- FIGUEIREDO, D.; SOUSA, L. (2000). *Guia de utilização do EASYcare, alguns aspectos fundamentais a considerar*. Aveiro: Universidade de Aveiro (documento policopiado elaborado no âmbito do Projecto GATT – Geriatric Assessment Technological Training)

BIBLIOGRAFIA RETIRADA DA INTERNET

CVRM – Centro de Geo-Sistemas do Instituto Superior Técnico (2000). Curso de Especialização - Análise Espacial de Dados para as Ciências Sociais e do Ambiente. Retrieved January, 25, 2001 from the World Wide Web: <http://alfa.ist.utl.pt/~cvrm/cursos/analise-dados/index.html>

INFORMAÇÕES

VOLUME I

1º QUADRIMESTRE DE 2002

CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS

CONGRESS, SEMINARS AND CONFERENCES

2002

□ 26-29 May

Annual Meeting of the Statistical Society of Canada, Hamilton, Ontario, Canada.

Informações: Local Arrangements Chair: Peter Macdonald Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada; Phone: (905) 525-9140 x 23423 Fax: (905) 522-0935. Program Chair: Bruce Smith Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, B3H 3J5, Canada; Phone: (902) 494-2257 Fax: (902) 494-5130.

E-mail: pdmamac@mcmail.cis.mcmaster.ca

bsmith@mathstat.dal.ca

□ 27-29 May

Work Session on Statistical Data Editing of the Conference of European Statisticians of the UN, to be held in Helsinki, Finland.

Informações: Mrs. Jana Meliskova, UNU/ECE Statistical Division, tel. +41 22 917 41 50; fax +41 22 917 00 40.

E-mail: jana.meliskova@unece.org

URL: www.unece.org/stats/

□ June

2nd ICM-EARCOME (East Asia Regional Conference on Mathematics Education) to be in Singapore.

Informações: EARCOME 2002, Division of Mathematics, National Institute of Education, 469 Bukit Timah Road, Singapore 259756, Republic of Singapore.

E-mail: earcome2@nie.edu.sg

□ 2-6 June

Seventh Valencia International Meeting on Bayesian Statistics, to be held at the Playa de las Americas, Tenerife, Canary Islands, Spain.

Information: URL: www.uv.es/valencia7

www.stat.duke.edu/valencia7

□ 5-9 June

Hawaii International Conference on Statistics and Related Fields, to be held at the Sheraton Waikiki Hotel, Honolulu, Hawaii, USA.

Informações: Hawaii International Conference on Statistics, 2440 Campus Road #517, Honolulu, HI, 96822, USA. Tel. (808) 223-1748, fax (808) 947-2420.

E-mail: statistics@hicstatistics.org

URL: www.hicstatistics.org

□ 9-13 June

Nordstat 2002 - The 19th Nordic Conference on Mathematical Statistics, to be held in Stockholm, Sweden.

Information: URL: www.math.kth.se/nordstat/

□ 12-15 June

23rd SCORUS conference "Statistics for the Cities of Tomorrow", to be held in Lisbon, Portugal.

Informações: Local Organising Committee, tel: +351 21 842 62 90, fax: +351 21 842 63 56.

E-mail: alberto.pina@ine.pt.

URL: www.ine.pt/novidades/semin/scorus.html

□ 12-15 June

Summer School on Sensitivity Analysis of Model Output, to be held at the University of Venice, Italy. The goal of the summer school is to make the participants aware of the state-of-the-art methods in sensitivity analysis, to stimulate the flow of new ideas, and to suggest new research directions. Most of the methodological advances in the field of sensitivity analysis will be covered.

Informações: Stefano Tarantola, Ph.D. Technological and Economic Risk Management Unit, Institute for the Protection and Security of the Citizen, JRC - Ispra Site, TP361, I-21020 Ispra (VA), Italy Tel: +39-0332-789928 or Sec: +39-0332-789686 Fax: +39-0332-785733

URL: <http://www.jrc.cec.eu.int/uasa/evt-samo2002.asp>

□ 17-20 June

MMR 2002, Third International Conference on Mathematical Methods in Reliability, to be held at the Norwegian University of Science and Technology, Trondheim, Norway.

Informações: Professor Bo Lindqvist, Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway. Tel.: +47-73 59 35 20 - Fax: +47-73 59 35 24.

E-mail: mmr2002@math.ntnu.no

URL: <http://www.math.ntnu.no/mmr2002/>

- 18-22 June

TIES 2002: Annual Conference of The International Environmetrics Society, to be held in the Faculty of Engineering in Genoa, Italy.

Information: URL: www2.stat.unibo.it/ties2002

- 22-25 June

C.Warren Neel Conference on Statistical Data Mining & Knowledge Discovery, to be held at the Hyatt Regency Knoxville, Tennessee, USA.

Informações: Conference Coordinator: Derek M Norton, Department of Statistics, 321 Stokely Management Ctr, The University of Tennessee, Knoxville, TN 37996-0532, USA. Phone: (865) 974-1646 Fax: (865) 974-2490.

E-mail: stat_dmconference2002@utk.edu

URL: <http://stat.bus.utk.edu/conference/>

- 23-26 June

International Symposium on Forecasting, ISF2002, to be held in Trinity College, Dublin, Ireland.

Informações: John Haslett

E-mail: John.Haslett@tcd.ie

URL: <http://www.isf2002.org/>

- 23-29 June

The 8th International Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius, Lithuania.

Informações: Professor Vytautas Statulevicius, Institute of Mathematics and Informatics, Akademijos str. 4, 2600 Vilnius, Lithuania.

E-mail: conf@ktl.mii.lt

E-mail: pdmac@mcmail.cis.mcmaster.ca

E-mail: conf@ktl.mii.lt

- 24-27 June

ITI 2002 (24th International Conference Information Technology Interfaces), to be held in Cavtat, Dubrovnik, Croatia.

Informações: URL: <http://iti.scre.hr/>

- 24-27 June

BIOSTAT 2002- 7th School of Biometrics, to be held in Cavtat, Dubrovnik, Croatia.

Informações: URL: <http://iti.scre.hr/>

□ 27-28 June

The seventh conference on "Messung der Teuerung (Measurement of Inflation)" to be held in Bremen in the building of the Chamber of Industry and Commerce. To be conducted in German language with some English lectures.

Informações: Prof. Dr. Eckart Elsner; Statistisches Landesamt Berlin; 10306 Berlin; Tel. +49 (30) 90 21 34 88; Fax +49 (30) 90 21 25 61;
E-mail: e.elsner@statistik-berlin.de
URL: <http://www.statistik-berlin.de/>

□ 1-3 July

The second ISI meeting on Professional Ethics, to be held in Berlin in the building of Deutscher Städtetag. The conference language is English.

Information: Prof. Dr. Eckart Elsner; Statistisches Landesamt Berlin; 10306 Berlin Tel. +49 (30) 90 21 34 88; Fax +49 (30) 90 21 25 61
E-mail: e.elsner@statistik-berlin.de
URL: <http://www.statistik-berlin.de/>

□ 1-5 July

28th Conference on Stochastic Processes and their Applications, to be held in Melbourne, Australia.

Information: Postal address: SPA28: Bronwen Hewitt, Conference Management, Old Physics Building, The University of Melbourne, VIC 3010, Australia. Fax: (+61-3) 8344 6122, tph: (+61-3) 8344 6389.
E-mail: bhewitt@unimelb.edu.au
URL: <http://www.spa28.ms.unimelb.edu.au/>

□ 1-6 July

Workshop on Optimal Stopping and Stopping Games, organized by Banach Center at Mathematical Research and Conference Center, Bedlewo, Poland.

Information: URL: neyman.im.pwr.wroc.pl/oss2002/
E-mail:

□ 1-6 July

2nd International Conference on the Teaching of Mathematics at the Undergraduate Level, to be held in Crete, Greece.

Information: Dept of Math/Computer Science, Capital University, USA.
E-mail: ivakalis@capital.edu
URL: www.math.uoc.gr/~ictm2

□ 1-7 July

Workshop on Optimal Stopping and Stochastic Games, to be held at the Mathematical Research and Conference Center in Bedlewo, Poland. The workshop will be devoted to the optimal stopping of stochastic processes and stochastic games. It has important applications both in real life and other fields of probability theory and mathematical statistics.

Information: Krzysztof Szajowski.

E-mail: szajow@im.pwr.wroc.pl

URL: <http://neyman.im.pwr.wroc.pl/oss2002>

□ 2-5 July

MCQT'02 - First Madrid Conference on Queueing Theory, to be held at the Department of Statistics and OR, Faculty of Mathematics, University Complutense of Madrid, Spain.

Information: Jesus R. Artalejo.

E-mail: mc_qt@mat.ucm.es

URL: <http://www.mat.ucm.es/deptos/es/mcqt/conf.html>.

□ 7-10 July

The 2002 Taipei International Statistical Symposium and Bernoulli Society EAPR Conference, to be held in the Activity Center, Academia Sinica, Taipei, Taiwan, China.

Information: E-mail: 2002symp@stat.sinica.edu.tw

URL: www.stat.sinica.edu.tw/2002symp

□ 7-11 July

16th Australian Statistical Conference, to be held at the National Convention Centre, Canberra, Australia.

Information: E-mail: asc16@con-sol.com

URL: www.statsoc.org.au/asc16.html

□ 7-12 July

The Sixth International Conference on Teaching Statistics (ICOTS6), to be held in Durban, South Africa.

Information: Maria-Gabriella Ottaviani - IPC Chair; Brian Phillips - International Organizer; , Dani Ben-Zvi - IPC Scientific Secretary.

E-mail: mariagabriella.ottaviani@uniroma1.it;

bphillips@swin.edu.au;

dani.ben-zvi@weizmann.ac.il.

URL: <http://icots.itikzn.co.za/>.

□ 8-12 July

17th International Workshop on Statistical Modelling: Statistical Modelling in Society, to be held in Chania-Crete, Greece.

Information: E-mail: d.stasinopoulos@unl.ac.uk

URL: www.unl.ac.uk/iwsm/

- 14-18 July
Sixth Multi-Conference on Systemics, Cybernetics and Informatics, to be held in Orlando, Florida, USA.
Information: General Chair, Professor Nagib Callaos.
 E-mail: ncallaos@callaos.com,
 ncallaos@aol.com
 callaos@telcel.net.ve
 URL: www.iiis.org/sci2002/
- 15-19 July
Current Advances and Trends in Nonparametric Statistics, to be held on Crete, Greece.
Informações: Michael G. Akritas and Dimitris N. Politis IMS Representative:
 Michael G. Akritas,
 E-mail: mga@stat.psu.edu
 URL: <http://www.stat.psu.edu/~npconf/>
- 21-26 July
IBC 2002 - International Biometric Conference 2002, to be held at the University of Freiburg, Germany.
Information: Chair: Robert Curnow; Chair Local Organizing Committee: Martin Schumacher.
 E-mail: r.n.curnow@reading.ac.uk
 ms@imbi.uni-freiburg.de
 URL: <http://www.ibc2002.uni-freiburg.de/>
- 21-26 July
PME 26, to be held in Norwich, UK.
Information: Anne Cockburn.
 E-mail: pme26.chair@uea.ac.uk
 URL: www.uea.ac.uk/edu/pme26
- 22-24 July
26th Annual Conference of the Gesellschaft für Klassifikation (GfKI), to be held at the University of Mannheim, Germany.
Informações: local organizer Prof. Dr. Martin Schader.
 URL: <http://www.gfki.de/gfki2002>
- 24-26 July
SSDBM 2002 (14th International Conference on Scientific and Statistical Database Management) to be held at Edinburgh University, United Kingdom.
Information: E-mail: ssdbm2002@ed.ac.uk
 URL: <http://www.ssdbm2002.ed.ac.uk/>

- 28 July – 2 August

IMS Annual Meeting/Fourth International Probability Symposium, to be held in Banff, Canada.

Informações: IMS Program Chair Tom DiCiccio, Cornell, Symposium Chair: Tom Kurtz, U. Wisconsin, IMS Local Chair: Subhash Lele, U. Alberta.
E-mail: tjd9@cornell.edu
Kurtz@math.wisc.edu
slele@ualberta.ca

- 4-9 August

Fourth International Conference on Statistical Data Analysis based on the L_1 -Norm and Related Methods - to be held at the University of Neuchâtel, Switzerland.

Information: Prof. Yadolah Dodge, Conference Organizer Statistics Group, Case Postale 1825, CH-2002 Neuchatel. Phone +41 32 718 13 80 Fax +41 32 718 13 81.
E-mail: Yadolah.Dodge@unine.ch

- 11-15 August

Joint Statistical Meetings, New York, Hilton and Sheraton New York. Sponsored by ASA, ENAR, WNAR, IMS, and SCC.

Informações: ASA, 1429 Duke St., Alexandria, VA 22314-3415;
Tel. (703) 684-1221;
Email meetings@amstat.org

- 14-17 August

Perspectives in Modern Statistical Inference II, (a satellite to the 24th European Meeting of Statisticians, August 19-23,2002) to be held at Masaryk University, Brno, Czech Republic.

Information: E-mail: jurecko@karlin.mff.cuni.cz.
URL: www.math.muni.cz/workshop_2002

- 15-17 August

Symposium on Stochastics and Applications (SSA) to be held at the National University of Singapore.

Informações: E-mail: ssa@math.nus.edu.sg
URL: <http://www.math.nus.edu.sg/ssa>

- 17-23 August

Baltic-Nordic Conference on Survey Sampling, to be held in Ammarnäs, Sweden.

Information: E-mail: banocoss@matstat.umu.se
URL: <http://www.matstat.umu.se/banocoss>

- 19-23 August

24th European Meeting of Statisticians, Prague, Czech Republic.

Informações: Martin Janzura, Institute of Information Theory and Automation,
POB 18, 182 08 Praha 8, Czech Republic.

Tel: 420 2 6605 2572.

Fax: 420 2 688 4903.

Email: janzura@utia.cas.cz

- 24-28 August

Compstat2002 to be held in Berlin, Germany.

E-mail: info@compstat2002.de, website <http://www.compstat2002.de>

Informações: E-mail: info@compstat2002.de

URL: <http://www.compstat2002.de>

- 25-28 August

International Conference on Improving Surveys (ICIS-2002), to be held at the University of Copenhagen.

Information: International Conference Services, P.O. box 41, Strandvejen 171,
DK-2900 Hellerup, Copenhagen, Denmark. Telephone: +45 3946 0500, Fax +45 3946 0515.

E-mail: ICIS2002@ics.dk

- 26-28 August

Sixth International Statistics Conference, to be held at the Tarbiat Modarres University, Tehran, Iran.

Information: Dr. M. Mohammadzadeh, Dept. of Statistics, Tarbiat Modarres Univ., P.O.Box 14155-175 Tehran Iran.

E-mail: isc6@modares.ac.ir

URL: www.modares.ac.ir/isc6/

- 27-29 August

IAOS Conference - Official Statistics and the New Economy, to be held in London, UK.

Information: URL: <http://www.statistics.gov.uk/iaoslondon2002/registration.asp>

URL: www.singstat.gov.sg/IAOS/uk.html

- 29-31 August

The Eleventh International Workshop on Matrices and Statistics, in Celebration of George P. H. Styan's 65th Birthday, to be held at Lyngby, Denmark.

Information: URL: <http://www.imm.dtu.dk/matrix02/>

□ 30-31 August

ISI Briefing Seminar for Chief Statisticians, to be held at the offices of the ONS, London, U.K.

Information: ISI Permanent Office, 428 Prinses Beatrixlaan, P.O. Box 950, 2270 AZ, Voorburg, The Netherlands, Tel: +31 70 3375737, Fax: +31 70 3860025.

E-mail: isi@cbs.nl

URL: <http://www.cbs.nl/isi/NLet021-10.htm>

□ 30 August-2 September

Fifth International Conference on Forensic Statistics, ICFS5, to be held in Isola di San Servolo, Venice, Italy.

Information: Julia Mortera, Dipartimento di Economia, Università di Roma Tre, Via Ostiense, 139, 00154 Roma. Tel: +39 - 06-5737-4206, fax +39 - 06-5737-4093.

E-mail: icfs5@eco.uniroma3.it

URL: icfs5.eco.uniroma3.it

□ 2-6 September

RSS 2002 Conference to be held at the University of Plymouth, Plymouth, England.

Information: The 2002 Conference of the Royal Statistical Society (4-6 September) will be preceded by short courses (2-3 September).

E-mail: J.Stander@plymouth.ac.uk

□ 09-13 September

ISCB 23 - 23rd Meeting of the International Society for Clinical Biostatistics, to be held in Dijon, France.

Information: David W. Warne

E-mail: david_w_warne@bluewin.ch

harbajan.chadha-boreham@serono.com

URL: <http://www.iscb-homepage.org/>

□ 09-14 September

XIVe Ecole Européenne d'Eté en E.D.A., IUT de Carcassonne, France.

Information: Le Guen Monique.

E-mail: leguen@univ-paris1.fr

Eugen.horber@politic.unige.ch

URL: <http://www.unige.ch/ses/sococ/mirage/eeda.html>

□ 15-19 September

Annual Meeting of the International Association for Mathematical Geology, IAMG 2002, to be held at the Freie Universität Berlin, Germany. Email:

iamg2002@zedat.fu-berlin.de Website: <http://www.fu-berlin.de/iamg2002/>

Information: E-mail: iamg2002@zedat.fu-berlin.de

URL: <http://www.fu-berlin.de/iamg2002/>

- 16-18 September

KOMAT'2002: International Workshop on Intelligent Knowledge Management Techniques (I-KOMAT'2002) to be held in conjunction with: KES'2002 (Sixth International Conference on Knowledge-Based Intelligent Information & Engineering Systems), Podere d'Ombriano, Crema, Italy.

Information: www-mugc.cc.monash.edu.au/~maum/ikomat-02.htm

Information: URL: www-mugc.cc.monash.edu.au/~maum/ikomat-02.htm

- 20-22 September

2nd Conference in Actuarial Science and Finance, to be held in Samos, Greece.

Information: Website: www.stat.ucl.ac.be/Samos2002/

Information: URL: www.stat.ucl.ac.be/Samos2002/

- 20-25 September

"The Humanistic Renaissance in Mathematics Education", to be held at the Citta del Mare Hotel, near the city of Palermo, Sicily, Italy.

Email arogerson@vsg.edu.au

Information: E-mail: arogerson@vsg.edu.au

- 24-27 September

Third International ICSC Symposium on Engineering of Intelligent Systems and the ISMC 2002 Workshop Workshop on Information Systems for Mass Customization, to be held at the University of Malaga, Malaga, Spain.

Information: ICSC- NAISO The Netherlands (Operating Division), P.O. Box 1091, 3360 BB Slidrecht, The Netherlands. Phone: +31-184-496999, Fax: +31-184-421065.

E-mail: eis2002@ITStransnational.com

planning@icsc.ab.ca

URL: www.icsc-naiso.org/conferences/eis2002/index.html

- 25-28 September

The 10th Annual Conference of the Portuguese Statistical Society (SPE) will be held in Porto, Portugal.

Information: E-mail: spe2002@fep.up.pt

URL: <http://www.fep.up.pt/spe2002>

- 16-19 October

26th CIRET Conference on "Business Surveys, Business Cycle Indicators and Consumer Surveys", to be held in Taipei/Taiwan, China.

Information: URL: www.ciret.org/

- 17-18 October

Third francophone Conference on sampling, in Grenoble and Autrans (France).

Information: Benoît Riandey

E-mail: riandey@ined.fr

URL: sondages2002.upmf-grenoble.fr/

- ❑ 21-23 October
International clinical trials symposium, to be held in Sydney, Australia.
Information: E-mail: trials@icms.com.au
URL: www.ctc.usyd.edu.au/4news/Symposium2002/ICTS_home.htm

- ❑ 31 October-03 November
Euroworkshop on Statistical Modelling - Model Building and Evaluation, to take place in Schloss Hohenried, Bernried, Munich, Germany.
Information: Herwig Friedl.
E-mail: friedl@stat.tu-graz.ac.at
goeran@stats.gla.ac.uk
URL: <http://www.stat.uni-muenchen.de/euroworkshop/2002.html>

- ❑ 06-08 November
XIX International Methodology Symposium "Modelling Survey Data for Social and Economic Research" to be held in Ottawa, Canada.
Information: E-mail: Symposium2002@statcan.ca
URL: <http://www.statcan.ca/english/services/smnrs.htm>
URL: http://www.statcan.ca/francais/services/smnrs_f.htm

- ❑ 14-17 November
International Conference on Questionnaire Development, Evaluation, and Testing, Methods, to be held in Charleston, South Carolina, USA.
Information: Jennifer Rothgeb
E-mail: jennifer.m.rothgeb@census.gov
or
URL: <http://www.jpsm.umd.edu/qdet>

- ❑ 20-23 November
International Conference on Information and Communication Technologies in Education, to be held at the Universidad de Extremadura, Badajoz, Spain.
Information: Antonio Mendez Vilas.
E-mail: amvilas@unex.es
URL: <http://www.formatex.org/congreso.html>

- ❑ 09-13 December
9th Annual Biopharmaceutical Applied Statistics Symposium, Savannah, Georgia, United States.
Information: MCV/VCU & GASOU Biostatistics Depts.
E-mail: KEPeace@gasou.edu
URL: <http://views.vcu.edu/bis/bass/>

- 21-24 December

Eighth Islamic Countries Conference on Statistical Science, to be organised by the Islamic Society of Statistical Sciences (ISOSS) in collaboration with the University of Bahrain. The conference will take place at the University of Bahrain.

Information: Organizing Committee, College of Business Administration, Department of Management & Marketing, University of Bahrain, P.O. Box # 32038, Sakhir, Bahrain. Tel: +973-449408/449413/449400, Fax: +973-449476, OR Mr. Akhlaq Ahmad, Executive Director, Islamic Society of Statistical Sciences, RB/2, 2nd Floor, Awami Complex, Usman Block, New Garden Town, Lahore-54600 (Pakistan).

E-mail: <mailto:akrammoh@internic.uob.bh>
drmunir@brain.net.pk

- 28-30 December

International Conference on "Ranking and Selection, Multiple Comparisons, Reliability, and Their Applications". Tentative Venue: Hotel Savera, Chennai, Tamilnadu, India.

Organizers: bala@mcmail.cis.mcmaster.ca; NKannan@utsa.edu; H. N. Nagaraja, Ohio State University, <mailto:hnn@stat.ohio-state.edu>

Information: N. Balakrishnan, McMaster University; N. Kannan, University of Texas at San Antonio; H. N. Nagaraja, Ohio State University.

E-mail: bala@mcmail.cis.mcmaster.ca
NKannan@utsa.edu
<mailto:hnn@stat.ohio-state.edu>

2003

- 02-04 January

International Conference on 'Statistics in industry and business' to be held in Cochin, India.

Information: Bovas Abraham, IIQP Univ of Waterloo, Waterloo, Ontario N2L 3G1, Canada; Tele: 519-888-4593, Fax: 519-746-5524; Dr.P.G.Sankaran Dept. of Statistics Cochin University of Science & Technology Cochin-682022; Tel.no: 0484-555893(Office), 0484-741693(Res.).

E-mail: babraham@uwaterloo.ca
pgsankaran@cusat.ac.in,
sankaranpg@yahoo.com

□ 08-10 January

The seventh conference on "Statistics and youth assistance in big cities", the forth SCORUS Berlin meeting to be held in "Haus am Rupenhorn".

Information: Prof. Dr. Eckart Elsner; Statistisches Landesamt Berlin; 10306 Berlin; Tel. +49 (30) 90 21 34 88; Fax +49 (30) 90 21 25 61
E-mail: e.elsner@statistik-berlin.de
URL: <http://www.statistik-berlin.de/>

□ 08-11 June

Annual Meeting of the Statistical Society of Canada, Halifax, Nova Scotia, Canada.

Information: Chris Field Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, B3H 3J5, Canada; Phone: (902) 494-3339 or (902) 494-2572 Fax: (902) 494-5130.
E-mail: field@mscs.dal.ca

□ 09-12 June

International Conference on "Advances in Statistical Inferential Methods" (ASIM) Almaty, Republic of Kazakhstan.

Information: E-mail: voinov@kimep.kz
URL: www.kimep.kz/research/asim

□ 24-27 June

International NAISO Symposium on Information Technologies in Environmental Engineering (ITEE'2003), to be held at the Technical University of Gdansk, Poland.

Information: E-mail: itee2003@global-conf.org
URL: <http://www.icsc-naiso.org/conferences/itee2003/index.html>

□ 24-27 June

First World Congress on Information Technology in Environmental Engineering ITEE 2003, Technical University of Gdansk, Poland.

Information: URL: <http://www.icsc-naiso.org/conferences/itee2003/index.html>

□ 14-18 July

International Conference on Robust Statistics 2003 (ICORS 2003), to be held at the University of Antwerp, Belgium.

Information: Stefan Van Aelst
E-mail: statis@uia.ua.ac.be
URL: <http://win-www.uia.ac.be/u/statis/lcors03.htm>

□ 20-24 July

ISCB 24 - 24th Meeting of the International Society for Clinical Biostatistics, to be held in London, UK.

Information: E-mail: diana.elbourne@lshtm.ac.uk
URL: <http://www.iscb-homepage.org/>

□ 13-20 August

International Statistical Institute, 54th Biennial Session (includes meetings of the Bernoulli Society, The Intern. Assoc. for Statistical Computing, The Intern. Assoc. of Survey Statisticians, The Intern. Assoc. for Official Statistics and The Interna. Assoc. for Statistical Education), to be held in Berlin, Germany.

Informações: ISI Permanent Office, Prinses Beatrixlaan 428,
P.O. Box 950, 2270 AZ Voorburg, The Netherlands.
Tel.: +31-70-337-5737;
Fax: +31-70-386-0025;
E-mail: isi@cbs.nl
or visit the Session website at <http://www.isi-2003.de>

□ 10-12 December

World Summit on Information Society, to be held in Geneva, Switzerland.

Information: Alain Clerc, Director, Civil-Society Division - Executive Secretariat,
World Summit on the Information Society.
E-mail: info@geneva2003.org
URL: <http://www.geneva2003.org/>

2004

□ 20-24 July

ISCB 25 - 25th Meeting of the International Society for Clinical Biostatistics, to be held in Leiden, Netherlands.

Information: E-mail: jcvanhouwelingen@lumc.nl
URL: <http://www.iscb-homepage.org/>

2005

□ 05-12 April

International Statistical Institute, 55th Biennial Session (includes meetings of the Bernoulli Society, the International Association for Statistical Computing, the International Association of Survey Statisticians, the International Association for Official Statistics and the International Association for Statistical Education), to be held in Sydney, Australia.

Information: ISI Permanent Office, Prinses Beatrixlaan 428, P.O. Box 950, 2270 AZ Voorburg, The Netherlands. Tel.: +31-70-337-5737; Fax: +31-70-386-0025.
E-mail: isi@cbs.nl

FUNDAMENTO, OBJECTO E ÂMBITO DA REVISTA

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, passou a editar quadrimestralmente a presente *Revista de Estatística* destinada a divulgar:

- a) Numa perspectiva científica, artigos originais sobre temas especializados da estatística, tanto pura como aplicada, bem como sobre estudos e análises nos domínios económico, social e demográfico;
- b) Informações sobre actividades e projectos importantes do Sistema Estatístico Nacional;
- c) Informações sobre acções desenvolvidas pelo INE no âmbito da cooperação.
- d) Informações sobre congressos, seminários, colóquios e conferências de interesse estatístico ou afim;

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

Para a elaboração e envio das contribuições de artigos para publicação na Revista são adoptadas as *Normas de Apresentação de Originais* que figuram na última página.

Os autores dos artigos publicados, a que se refere a alínea a), receberão uma contribuição financeira paga pelo INE, de montante a fixar por despacho da Direcção mediante proposta do Director da Revista.

OS PONTOS DE VISTA EXPRESSOS PELOS AUTORES DOS ARTIGOS PUBLICADOS NA REVISTA
NÃO REFLECTEM NECESSARIAMENTE A POSIÇÃO OFICIAL DO INE.

FOUNDATION, SUBJECT MATTER AND SCOPE OF THE REVIEW

INE is conscious of how statistical awareness is essential to the understanding of the majority of phenomena in the present world and is aware of its responsibility to disseminate statistical knowledge, making it available to the widest possible range of readers. INE has recognised the need to take a step in that direction and will begin publication of this *Statistical Review* three times yearly, designed to provide the following:

- a) Within a scientific perspective, original articles on specialised areas of statistics, both pure and applied, as well as studies and analyses within the sphere of economics, social issues and demographics;
- b) Information on activities and projects of the National Statistical System;
- c) Information on activities developed by INE within the scope of co-operation;
- d) Information on congresses, seminars and conferences of a statistical or related nature;

The following approaches for contributing material for publication in the review have been adopted:

- In relation to the articles referred to in section a), contributions are made by the authors themselves and by invitation of the Editorial Committee, whether they are employees of INE or not;
- In relation to the information referred to in section b), c) and d); contributions are from departments of INE.

The Editorial Committee who has sole discretion in deciding whether or not the material will be published will assess the scientific merit of contributions made on the initiative of the authors themselves.

The preparation and delivery of material for publication in the Review are subject to the *Rules for Submitting Originals* presented on the last page. The authors of the published articles referred to in section a) will receive pecuniary compensation from INE in an amount to be determined by resolution of the Board on the recommendation of the Director of the Review.

THE VIEWPOINTS EXPRESSED BY THE AUTHORS OF THE ARTICLES PUBLISHED IN THE REVIEW
DO NOT NECESSARILY REFLECT THE OFFICIAL POSITION OF I.N.E.

NORMAS DE APRESENTAÇÃO DE ORIGINAIS

Nos termos do *Regulamento da Revista de Estatística*, o Conselho Editorial aprovou as seguintes **Normas de Apresentação de Originais**:

1. Os originais dos artigos serão enviados ao Director da Revista pelos respectivos autores, devendo ser escritos em *português* e não terem sido ainda totalmente publicados, ou estar em processo de edição em outra publicação.
2. Poderão também ser apresentados artigos escritos em *inglês*, cabendo ao Director da Revista a decisão sobre a sua aceitação.
3. Quanto à *avaliação do mérito científico* dos artigos:
 - a) Os artigos apresentados por *iniciativa* dos respectivos autores serão submetidos à avaliação do mérito científico pelo Conselho Editorial, com garantia do anonimato tanto do autor como dos avaliadores;
 - b) Os autores receberão a informação sobre o resultado da avaliação num prazo máximo de trinta 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 original apresentado.
4. Os artigos aceites para publicação na *Revista de Estatística* serão igualmente divulgados no *site* do INE na *Internet*.
5. Os originais, com uma extensão não superior a trinta páginas, serão processados em *Word for Windows*, integralmente a preto e branco, com indicação do(s) software(s) adicional(ais) eventualmente utilizado(s) na produção do documento original, e entregues em suporte papel acompanhado da respectiva *disquette*, ou enviados por E-mail para o seguinte endereço: liliana.martins@ine.pt
6. Na apresentação dos originais, os autores respeitarão ainda as seguintes normas:
 - 6.1. Quanto à *estrutura*:
 - a) O texto deve ser processado em formato *A₄*, com utilização do tipo de letra *Times New Roman 11*, espaçamento *at least 12*, e com as seguintes margens: *top*: 4 cm, *bottom*: 3 cm, *left*: 2,5 cm, *right*: 5 cm, *header*: 1,25cm, *footer*: 1,25cm;
 - b) A primeira página conterá exclusivamente o título do artigo, bem como o nome, morada e telefone, fax e E-mail do autor, com indicação das funções exercidas e da instituição a que pertence, devendo, no caso de vários autores, ser indicado a quem deverá ser dirigida a correspondência da Revista;
 - c) A segunda página conterá, em português e inglês, unicamente o título e um *resumo* do artigo, com um máximo de 100 palavras,

seguido de um parágrafo com indicação de *palavras-chave* até ao limite de 15;

- d) Na terceira página começará o texto do artigo, sendo as suas eventuais secções ou capítulos numeradas sequencialmente;

6.2. Quanto a *referências bibliográficas*:

- a) Os autores eventualmente citados no texto do artigo serão indicados entre parênteses curvos pelo seu nome seguido da data da respectiva publicação e, se for caso disso, do número de página (p. ex.: Malinvaud, 1989, 23);
- b) As Referências Bibliográficas serão listadas, por ordem alfabética dos apelidos dos respectivos autores, imediatamente a seguir ao final do texto, de acordo com a fórmula seguinte:

GREENE, W. H., “*Econometric Analysis*”, Prentice-Hall, New Jersey, 1993.

6.3. Quanto à *revisão de provas e publicação*:

- a) Uma vez aceite o artigo e antes da sua publicação, receberá o autor provas para revisão, as quais serão devolvidas 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 Director o direito de decidir a data da sua publicação futura;
- c) Uma vez publicado o artigo, o autor receberá vinte exemplares da sua versão impressa e um exemplar do respectivo número da *Revista*.

7. Para *informações adicionais* contactar o Secretariado de Redacção:

Eduarda Liliana Martins
Instituto Nacional de Estatística
Av.^a. António José de Almeida, n.º 5 – 9.º.
1000-043 Lisboa - Portugal

☐ Tel.: +351 21 842 62 05

☐ Fax.: +351 21 842 63 84

☐ e-mail: liliana.martins@ine.pt

RULES FOR SUBMITTING ORIGINALS

Within the terms of the *Regulation of the Statistical Review*, the Editorial Committee has approved the following **Rules for Submitting Originals**:

1. The original articles will be sent to the Review Director by the respective authors. They should be written in *Portuguese*, they should not have already been published in their entirety nor should they be in the process of being published in any other publication.
2. Articles may also be submitted in *English* to the Review's Director who will decide whether to accept them.
3. In relation to the *evaluation of the scientific merit* of the articles:
 - a) The Editorial Committee will assess the articles submitted on the initiative of the authors on the basis of their scientific merit. The identity of both the author and the Committee members will be strictly confidential;
 - b) The authors will receive information regarding the results of the evaluation of scientific merit within a maximum period of 30 days. If the article is accepted, the Committee will indicate the issue number of the *Review* in which the article will be published. If the article is not accepted, the original will be returned to the author.
4. The articles accepted for publication in the *Statistical Review* will also be made public on the Internet site of the INE.
5. The original articles having no more than thirty pages must be processed in *Word for Windows*, completely at black and white, with the information on the additional(s) software(s) eventually used in the production of the original document, and they will be delivered in hard copy as well as on diskette, or sent by E-mail to: liliana.martins@ine.pt
6. With the presentation of the original articles, the authors must also respect the following rules:
 - 6.1 In relation to the *structure*:
 - a) The text shall be printed on A4 format paper utilising the font *Times New Roman* size 11, spacing at least 12, and with the margins: *top* 4cm, *bottom* 3cm, *left* 2,5cm, *right* 5cm, *header* 1,25cm, *footer* 1,25cm;
 - b) The first page shall contain only the title of the article as well as the name, address and telephone, fax and E-mail number of the author, indicating the position held and the institution that he/she belongs to. In the case of various authors, it is necessary to indicate the person to whom all correspondence received should be forwarded;

- c) The second page shall contain in *Portuguese* and *English* only the *title* and an *abstract* of the article with the maximum of 100 words followed by a paragraph indicating *key words* up to the limit of 15;
- d) The third page will begin the text of the article with its respective sections or chapters sequentially numbered;

6.2 Regarding *Bibliographical References*:

- a) Authors who are cited in the text of the article shall be indicated in parentheses with their name followed by the date of the respective publication and, if necessary, the page number (ex.: Malinvaud, 1989, 23);
- b) All bibliographical references will be listed in alphabetical order by the surnames of the respective authors, immediately following the end of the text, as in the following example:

GREENE, W. H., "*Econometric Analysis*", Prentice-Hall, New Jersey, 1993.

6.3 Regarding *proof-reading and publication*:

- a) Once the article is accepted and prior to its publication, the author will receive a copy for review. These copy will be returned to the Director of the Review within a maximum period of one week from the date of its reception;
- b) The consequences of subsequent changes to the accepted first version are the responsibility of the respective authors as well as any delays in proof-reading that make its publication in the planned issue of the Review impossible. The Director reserves the right to decide upon the date for future publication;
- c) Once the article is published, the author will receive twenty copies of his/her printed version and a copy of the respective issue of the *Review*.

7. For *further information* kindly contact the Editorial Secretary:

Eduarda Liliana Martins
Instituto Nacional de Estatística
Av.^a. António José de Almeida, n.º. 5 – 9.º.
1000-043 Lisbon - Portugal

- ☐ Tel.: +351 1 21 842 62 05
- ☐ Fax.: +351 1 21 842 63 84
- ☐ e-mail: liliana.martins@ine.pt