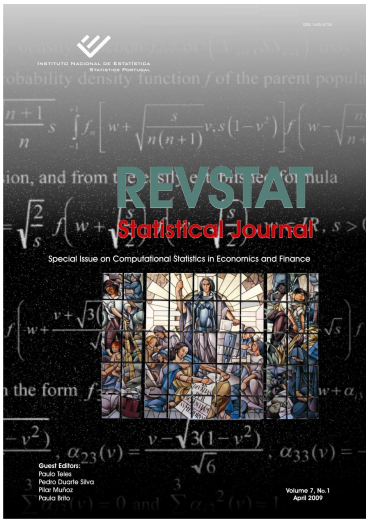
07 April, 2011

Multithemes

Volume 9, Number 1, March 2011

REVSTAT- STATISTICAL JOURNAL

REVSTAT-STATISTICAL JOURNAL, Volume 9, No. 1 – March 2011



In 2003 the National Statistical Institute launched the scientific statistical journal **REVSTAT-STATISTICAL JOURNAL**, published in English two times a year, with a prestigious international Editorial Board, which came to substitute the *Revista de Estatística* [Statistical Review], published in Portuguese between 1996 and 2002.

The aim of the Editorial Board of **REVSTAT** is to publish articles of high scientific content, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems, covering all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

REVSTAT is proud to be chosen by the scientific community to publish their research results.

The **REVSTAT** is covered by the following abstracting/indexing services: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic*.

For more information about **REVSTAT**, namely on-line articles, subscription of the publication, and submission of papers, please visit the link of the Statistics Portugal's website: http://www.ine.pt/revstat/inicio.html

This Volume of **REVSTAT: Volume 9, No. 1 - March 2011**, is about **"Statistical Modelling: Challenges in Health"** and includes six articles. Their abstracts are presented below:

A COMPARATIVE GENOMICS APPROACH TO THE IDENTIFICATION OF QTL CANDIDATE GENES
Authors: *Howsun Jow, Richard J. Boys* and *Darren J. Wilkinson*.

Despite rapid advances in sequencing technology, many commercially relevant species remain unsequenced, and many that are sequenced have very poorly annotated genomes. There is therefore still considerable interest in

using comparative approaches to exploit information from well-characterised model organisms in order to better understand related species. This paper develops a statistical method for automating part of a comparative genomics bioinformatic pipeline for the identification of genes and genomic regions in a model organism associated with a QTL region in an unsequenced species. A non-parametric Bayesian statistical model is used for characterising the density of a large number of BLAST hits across a model species genome. The method is illustrated using a test problem demonstrating that markers associated with Bovine hemoglobin can be automatically mapped to a region of the human genome containing human hemoglobin genes. Consequently, by exploiting the (relatively) high quality of genome annotation for model organisms and humans it is possible to quickly identify candidate genes in those well-characterised genomes relevant to the quantitative trait of interest.

OPTIMAL DYNAMIC TREATMENT METHODS

Authors: *Robin Henderson, Phil Ansell* and *Deyadeen Alshibani.*

This paper reviews and develops methods for implementing in practice recent ideas in the field of optimal dynamic treatment allocation. Given longitudinal sequences of observational data on health status and treatment selection for a cohort of patients, the aim is to determine a regime, or decision rule, which can be used to select treatment in order to optimise some final response or outcome. The approach to this problem that has been taken in the causal inference literature is shown to be extendable to problems in the field of stochastic optimisation. New diagnostic techniques to aid in model assessment are developed, and an application in anticoagulation is presented.

MISSING DATA IN REGRESSION MODELS FOR NON-COMMENSURATE MULTIPLE OUTCOMES

Authors: *Armando Teixeira-Pinto* and *Sharon-Lise Normand*.

Biomedical research often involves the measurement of multiple outcomes in different scales (continuous, binary and ordinal). A common approach for the analysis of such data is to ignore the potential correlation among the outcomes and model each outcome separately. This can lead not only to loss of efficiency but also to biased estimates in the presence of missing data. We address the problem of missing data in the context of multiple non-commensurate outcomes. The consequences of missing data when using likelihood and quasi-likelihood methods are described, and an extension of these methods to the situation of missing observations in the outcomes is proposed. Two real data examples illustrate the methodology.

A REVIEW ON JOINT MODELLING OF LONGITUDINAL MEASUREMENTS AND TIME-TO-EVENT

Author: *Inês Sousa*.

In longitudinal studies subjects are measured for one or more response variable, over time. Although the underlying evolution of such response variables is continuous in time, in practice the measurements are observed at discrete time points. In longitudinal clinical trials it is also common to observe relevant events, generating time-to-event data. If both types of data are available, we might be interested in the association between the two processes, longitudinal and time-to-event. Commonly, when death is considered the event, the observation sequence of longitudinal measurements is terminated by the event process. When the two observed processes are related, the analysis of the data set should be suited to the specific objectives. We distinguish three situations: if the interest is to analyse the longitudinal outcome response variable with drop-out at the time-to-event; to analyse time-to-event, whilst exploiting correlation with a noisy version of a time-varying risk factor; or to analyse the relationship between the two processes. Joint models assume a full distribution for the joint distribution of longitudinal and time-to-event processes, which includes a description of the relation between the two processes.

INFERENCE FOR NON-MARKOV MULTI-STATE MODELS: AN OVERVIEW

Author: *Luís Meira-Machado*.

In longitudinal studies of disease, patients can experience several events across a follow-up period. Analysis of such studies can be successfully performed by multi-state models. This paper considers nonparametric and semiparametric estimation of important targets in multi-state modeling, such as the transition probabilities and bivariate distribution function (for sequentially ordered events). These estimators are shown to be consistent even for data which is non-Markov. We illustrate the methods on two data sets.

MIXTURES OF FACTOR MODELS FOR MULTIVARIATE DISEASE RATES

Authors: *T.C. Bailey* and *P.J. Hewson*.

A range of different approaches have been suggested for the multivariate modelling of the geographical distribution of different but potentially related diseases. We suggest an addition to these methods which incorporates a discrete mixture of latent factors, as opposed to using CAR or MCAR random effect formulations. Our proposal provides for a potentially richer range of dependency structures than those encompassed in previously used models in that it is capable of representing an enhanced range of correlation structures between diseases at the same time as implicitly allowing for less restrictive spatial correlation structures between geographical units. We illustrate results of using the model on data taken from cancer registries on four carcinomas in some 300 UK geographical areas.