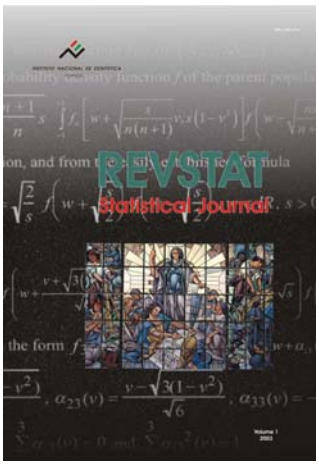
07th March, 2006

Estatísticas Gerais

REVSTAT- STATISTICAL JOURNAL

REVSTAT-STATISTICAL JOURNAL, Volume 4, No. 1 – March 2006



I

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 hopes to become a place where scientists may feel proud of publishing their research results changing the character of the previous *Revista de Estatística* from a national to an international scientific journal.

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

This Volume of **REVSTAT: Volume 4, No. 1 - March 2006**, is about **"Statistical Applications in Bioinformatics"** and includes four articles. Their abstracts are presented below:

QUALITY CONTROL AND LOW - LEVEL STATISTICAL ANALYSIS OF ILLUMINA BEADARRAYS

Authors: *Mark J. Dunning, Natalie P. Thorne, Isabelle Camilier, Michael L. Smith* and *Simon Tavaré*

The Illumina BeadArray™ platform is a novel microarray technology based on randomly assembled arrays of beads. Each bead on the array carries copies of a single gene-specific probe with, on average, about 30 replicates of each bead type on an array. Given the encouraging results regarding the reproducibility of BeadArray™ data and high profile studies already being carried out using the BeadArray™ technology, there is likely to be an increase in the volume of BeadArray™ data available. A major advantage of BeadArray™ technology is the high degree of replication of beads of a given type. However, current analysis methods give summarised information for each bead type as output rather than information for each individual bead on the array. The *beadarray* R package is able to recreate individual bead information for arrays using raw images as input. Here, we use a particular experiment to illustrate the image processing steps used by Illumina and corresponding methods available in *beadarray*. Our investigations into BeadArray™ data have demonstrated a high degree of reproducibility both within and between arrays. However, we identified some aspects of the low-level analysis that could be improved.

NETWORK MOTIFS: MEAN AND VARIANCE FOR THE COUNT

Authors: *C. Matias, S. Schbath, E. Birmelé, J.-J. Daudin* and *S. Robin*

Network motifs are at the core of modern studies on biological networks, trying to encompass global features such as small-world or scale-free properties. Detection of significant motifs may be based on two different approaches: either a comparison with randomized networks (requiring the simulation of a large number of networks), or the comparison with expected quantities in some well-chosen probabilistic model. This second approach has been investigated here. We first provide a simple and efficient probabilistic model for the distribution of the edges in undirected networks. Then, we give exact formulas for the expectation and the variance of the number of occurrences of a motif. Generalization to directed networks is discussed in the conclusion.

INFERRING GENE DEPENDENCY NETWORKS FROM GENOMIC LONGITUDINAL DATA: A FUNCTIONAL DATA APPROACH

Authors: *Rainer Opgen-Rhein* and *Korbinian Strimmer*

A key aim of systems biology is to unravel the regulatory interactions among genes and gene products in a cell. Here we investigate a graphical model that treats the observed gene expression over time as realizations of random curves. This approach is centered around an estimator of dynamical pairwise correlation that takes account of the functional nature of the observed data. This allows to extend the graphical Gaussian modelling framework from i.i.d. data to analyze longitudinal genomic data. The new method is illustrated by analyzing highly replicated data from a genome experiment concerning the expression response of human T-cells to PMA and ionomicin treatment.

STATISTICAL EVALUATION OF METHODS FOR THE ANALYSIS OF DYNAMIC PROTEIN EXPRESSION DATA FROM A TUMOR STUDY

Authors: *Klaus Jung, Ali Gannoun, Barbara Sitek, Ognjan Apostolov, Alexander Schramm, Helmut E. Meyer, Kai Stühler* and *Wolfgang Urfer*

In this article, we analyze time dependent protein expression data obtained from a proteome study of a neuroblastoma cell line. Neuroblastoma are common solid tumors which occur in early childhood. The expression data was obtained by difference gel electrophoresis (DIGE). It is known that the clinical outcome of neuroblastoma depends on the activation of different neurotrophin receptors by their ligands. Here, we are looking for proteome changes resulting from the activation of Tyrosine Kinase (TrkA) receptors by their ligand NGF (nerve growth factor). Before analyzing the data by longitudinal data analysis we do data preprocessing and apply a method for the imputation of missing values.