

# A survey of regression modelling techniques for analysing microarray data

Fernando Henrique Ferraz Pereira da Rosa  
Júlia Maria Pavan Soler

Department of Statistics, University of São Paulo, Brazil

## Abstract

Studies in genetics involving microarray experiments allow simultaneous comparison and quantification of gene expression on a large scale. Many sources of variation play an important role on the analysis of data coming from such experiments.

A proper experimental design and subsequent development of suitable models for analysis are essential steps in order to assure the identification of significant genes, helping researchers distinguish between variations which are due to actual biological changes from random noise [3].

In this work, we evaluate some of the various regression modelling strategies proposed in the literature, such as the use of mixed-effects models [4], [1] and the issue of variable selection in microarray gene expression profiling [2].

We apply the models studied to data from a study of the Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of São Paulo Medical School (InCor-USP) based on strains of rat, which aims to identify genes that have a regulatory role on the mechanism of hypertension.

## References

- [1] M. Kathleen Kerr and Gary A. Churchill. Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, (7), 2001.
- [2] Wentian Li and Yaning Yang. *Methods of Microarray Data Analysis*, chapter How many genes are needed for a discriminant microarray data analysis, pages 137-150. Kluwer Academic, 2002.
- [3] Janet Wittes and Herman P. Friedman. Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. *J. Natl. Cancer Inst.*, 91:400-401, 1999.
- [4] Russell D. Wolfinger, Greg Gibson, Elizabeth D. Wolfinger, Lee Bennett, Hisham Hamadeh, Pierre Bushel, Cynthia Afshari, and Richard S. Paules. Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, 8(6):625-637, 2001.