

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, June 21, 2005
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Cyr Emile M'LAN, Ph.D.
Assistant Professor
University of Connecticut
215 Glenbrook Road, U-4120
Storrs, CT, 06269

**ESTIMATING GLOBAL AND GENE-SPECIFIC PARAMETERS IN
UNBALANCED MULTIFACTORIAL ANOVA MODEL OF MOICROARRAY
DATA**

The ability to measure thousands of mRNA transcript expressions simultaneously using high-throughput genomic technology has revolutionized the field of Genetics. In our study, RNA from a collection of Lymphoblastoid Cell Lines were hybridized onto Affymetrix genechip arrays. The goal of this study is to discover the genes that are highly variable in the human population subgroups. An ANOVA model was performed and the effect of the Population subgroups were isolated and tested after adjusting for confounding effects such as ChipLot, Operator, and Gender. Here the complexity of our analysis lies in the fact that the covariates Gender and Population subgroups have gene-specific effects while the covariates ChipLot and Operator have a global effect (that is, they are common to all genes). In addition, our microarray data is highly unbalanced, therefore no results in the microarray literature is of any help.

In this talk we discuss how multifactorial analysis of variance containing both global and gene-specific parameters can be carried out efficiently in spite of the large size of microarray data. We first derive an analytical form of the solutions of the normal equations and use these solutions to suggest a low-cost two-stage analysis. Our procedure can be viewed as an extension of the work by Kerr et al. (2000) for balanced (orthogonal) designs. We also review permutation tests for both balanced and unbalanced ANOVA designs. All these results are applied to our unbalanced microarray data. We also discuss how to get around of the computational complexities in computing ten of thousands of empirical p-values efficiently based on large permutation size (more than 100,000).