

GRAND ROUNDS

Sponsored by: The Division of Biostatistics

Miguel Marino, PhD

Assistant Professor

Department of Family Medicine, OHSU

Thursday, April 18, 2012

12 PM-1 PM

Campus Services Building Room 679

**Factor analysis when the number of random variables
exceeds the sample size**

ABSTRACT

Factor analysis which studies correlation matrices is an effective means of data reduction whose inference on the correlation matrix typically requires the number of random variables, p , to be relatively small and the sample size, n , to be approaching infinity.

Contemporary data collection for biomedical studies, instrument development, disease surveillance and genetics typically produce data matrices where the number of variables is comparable to or often larger than the sample size. When $p > n$, the use of existing factor analysis methods to study the correlation matrix are limited. We formalize a test statistic to perform inference on the structure of the correlation matrix when $p > n$. We develop an approach based on group sequential theory to estimate the number of underlying factors to be extracted. The motivation for this work comes from studying the correlation matrix of log annual US cancer mortality rate change for $p=59$ cancer types from 1969 to 2008 ($n=39$).