

# Analyzing High Dimensional Data with Fewer Observations

Muni S. Srivastava

*Department of Statistics, University of Toronto*

email: [srivasta@utstat.toronto.edu](mailto:srivasta@utstat.toronto.edu)

## Abstract

In this talk, I consider the problem of analyzing datasets obtained on several characteristics ‘ $m$ ’ of an individual, such as in DNA microarrays in which such characteristics are typically in several thousands. The number, however, of subjects ‘ $N$ ’ in which these observations can be taken is typically small, often less than 50. Standard multivariate theory requires that ‘ $m$ ’ be substantially smaller than ‘ $N$ ’, and very little theory is available to analyze the case when ‘ $N$ ’ is less than ‘ $m$ ’. In the last decade or so, results have been obtained for this case also. I will present some of these results. Using these results, I will analyze two datasets from microarrays. Both datasets require the comparison of two groups, the so-called two-sample test. Often, it is required to find the components (or genes) that may have caused the rejection of the equality of two mean vectors. An alternative to the currently used False Discovery rate (FDR) will be presented.