

# Variable Selection in Multivariate Multiple Regression

Anita Brobbey

*Supervisor:* Dr. Asokan Mulayath Variyath

Department of Mathematics and Statistics, Memorial University

August 27, 2015

HH 3017

10:00 am

## **Abstract**

In many applied science or public health studies, researchers are interested in modeling the relationship between response variable(s) and explanatory variables (independent variables). In some studies, the number of explanatory variables can be considerably large due to an addition of interaction effects of covariates. If there are more than one response of interest, then the joint model for all responses result in high dimension of covariates. Despite the large number of covariates, some of them have no influence on the response variable(s). Thus, parameter estimation and variable selection are two important problems in multivariate regression analysis. Selecting a smaller number of important variables results in a simpler, interpretable model and reduce variability. In this seminar, we address the variable selection problem in multivariate multiple regression models. We propose penalized generalized estimating equation (GEE) approach to utilize the potential correlation among the responses. We conducted series of simulations to investigate the performance of our proposed approach. Our simulation showed that SCAD with BIC tuning criteria works well in selecting important variables. The estimates of  $\beta$  are unbiased (Liang and Zeger, 1986) regardless of the choice of correlation structure. However, estimates obtained from the unstructured working correlation have reduced standard errors. We also applied our method to concrete slump test data to investigate variable selection in continuous and binary multi-response framework.