

Speaker: Shannon Carroll

Title: *General Linear Model for Predicting Violent Crimes in the United States*

#### Abstract

A wide variety of general linear models can be created from a single dataset by using different statistical analysis methods. A collection of data on communities throughout the United States was created, including 147 attributes such as: percent of the population considered urban, median family income, and several variables involving law enforcement. These variables can be used to predict the number of violent crimes in a community. After splitting the data into testing and training set, forwards, backwards and best subset selection processes on BIC, AIC and CP criterion were used to create different linear models. By viewing residuals from both test and train sets, r squared, and mean square prediction error values, the better models were used to continue analyzing. Several methods were used to correct effects from outlying variables, test the transformation of the predicted variable and adjust for multicollinearity. The most accurate model used a combination of 14 variables to predict violent crimes in a city.