

## STA 426/526 Midterm Project – Classification

This project requires you do the following. Remember this is primarily about demonstrating what we've learned through homework 5.

- Import the given CSV file into JMP
- Address any data cleaning (missing data, etc.).
  - This step may include a little bit of variable transformation, but keep it simple.
- Select a training subset of 90% of the data for analysis and modeling.
  - Keep the remaining 10% for your validation set.
- Identify an interesting grouping of the data using unsupervised classification.
  - Associate your grouping with one or more of the available categorical variables.
  - Note that you likely will not have an exact match to anything.
- Investigate supervised learning using your new grouping variable.
  - Verify multivariate normality.
  - Reduce the number of variables as appropriate.
  -
- Create a JMP script to score a new data set.
- Validate your work with your validation subset.
- Report on your process and what you've learned using the following outline.
  - Overview
  - Data cleaning and sub-setting
  - Unsupervised classification investigation
  - Selection of your classifier variable(s)
  - Supervised classification investigation
  - Validation
  - Results