Math 327 - Data Analysis Project #2 Checklist

Title and Abstract, *5 pts*

___ Does the title give an accurate preview of what the project is about? (i.e. Is it informative, specific and precise?)

___ One paragraph stating the data, problem, and/or questions that are being addressed. Are the main points described clearly and succinctly?

Data Characteristics, *5 pts*

___ Numerical and/or graphical summaries of the individual variables, or just the key variables, such as those selected for the final model, if you have many variables.

Initial Model, 10 pts

___ Describe the results of your initial model (first-order model with all predictor variables)

___ Use plots and/or tables, along with summary text, as needed, to describe the results and next steps

Model Selection, *10 pts*

___ Use stepwise regression, direction='both', with either the AIC or BIC criterion, or both.

___ Evaluate the impact/benefit of adding 2-way interaction effects among the significant predictors using stepwise regression. Interpret the significant interaction effects that remain in the model.

___ State and interpret your final model; provide model parameters, standard errors, and p-values. If using logistic regression, also include anti-logged parameter values. Confidence intervals for key results (parameters, and/or predictions) are recommended.

Model Diagnostics, *10 pts*

___ Plot of deviance residuals vs. fitted values (fitted probabilities), with Lowess smooth line, and interpretation

___ P-values for goodness-of-fit test and likelihood ratio test, with interpretation

___ Results of influence diagnostics analysis (Cook's distance, hat values, etc.) and interpretation. If there are a few points with high influence or leverage, consider re-fitting the model without those data points and determine how much the results change.

Conclusion, *10 pts*

___ Describe the overall conclusions of your analysis. Include an ROC plot, and show the AUC, along with sensitivity, specificity, positive predictive value, and negative predictive value for the cutoff with the highest percentage of correctly classified observations.

___ Does your analysis raise any questions that can't be answered from the current data set? If so, what are they?

The criteria above will be applied to the R Markdown file for a total of 50 points and to the poster for a total of 50 points. The R markdown file should have all of your R code and results with just enough text for a

reader who is not familiar with the data to know what you did, what the results were, and what you concluded. The poster should have a subset of those results that fit within the space of the poster and tell as complete a story as possible about what you did, what you found, and what you concluded. As an option, the poster can be about your Project #1 results instead of Project #2.