**Logit Analysis Template**

*For this question, you will need to find a dataset (one we have not yet examined in class) that has a binary response variable, Yi, and four predictors, X1, X2 , X3 and X4. Two of the predictors ought to be continuous. The predictors ought to be plausibly related to the response variable. Your sample size ought to be sufficiently large to proceed with maximum likelihood estimation. Having compiled the data in R, please do the following:*

a. Briefly describe the data (including summary statistics for relevant variables) and explain the hypotheses with respect to the predictors.

b. Estimate both a logit and a probit model with four predictors. Present the results in a table with labeling appropriate for a social science research paper.

c. Calculate the P(Yi=1) for all levels of your categorical predictors, setting other predictors at their median values. For continuous predictors, calculate the P(Yi=1) at the 25th and 75th percentiles, setting other predictors at their median values. Report the associated standard errors or confidence intervals for these point estimates.

For the logit model, please do the following:

d. Calculate the odds ratio for each categorical predictor and the associated confidence intervals. What is the percent change in the odds associated with changing from the lowest level of the predictor (e.g., X1=0) to the highest level (e.g., X1=1).

e. For at least one continuous predictor, plot the predicted probability, i, over the range of values for the predictor, setting other variables at their median values. Include upper and lower bounds for i based on the 95% confidence interval.

f. For one continuous predictor, suggest a possible interaction between it and a categorical predictor. Estimate a logit model that includes this interaction. Is there an interaction effect? Justify your answer with an appropriate statistical test or figure.

g. Compare the predictive power of the logit model with the interaction to the logit model without the interaction. What is the proportional reduction in error associated with each? Is the difference in PRE values statistically significant?

h. Plot the ROC curve for one of these models. Briefly characterize the predictive power of the model.