

Instructions: Show work or attach R code used to perform calculations (or any other technology used). Be sure to answer all parts of each problem as completely as possible, and attach work to this cover sheet with a staple.

1. The file **325homework4data.xlsx** has data on beer preferences and several variables for prediction. (The first column is not a variable.)
 - a. Build a logistic regression model for this data to predict beer preferences.
 - b. How good is your model? How many of the predictions did your model get correct?
 - c. Which variables appear to be the most important in the classification?
 - d. What is the probability that a new customer who is Male, Married, Age 47 and with Income of \$42,000 will prefer regular beer?
2. Using the same data in the previous problem. Perform transformations on age and income: squares, square roots, log, reciprocals and an interaction term. Create a logistic model using these terms, either in addition to or in place of the original variables. Are you able to improve the model? Create a set of residual plots for each model to check model assumptions.
3. Assess the data set for potential problems with classification. Is the proportion of one group much larger than the other? Can you improve the predictions by adjusting the classification break to something other than 50% probability?