Homework 6

Each part of the problems 5 points
Due on Blackboard by 3pm on Friday October 23.
JWHT is 'An Introduction to Statistical Learning' by James, Witten, Hastie and Tibshirani.

1. A psychologist conducted a study to examine the nature of relation, if any, between an employee's emotional stability (X) and and the employee's ability to perform a task in a group (Y). Emotional stability was measured by a written test, for which the higher score the greater the emotional stability. Ability to perform a task in a group was evaluated by the supervisor ($Y=1$ if able, $Y=0$ if not). The results for the 27 employees are in the file `performance.csv`. The following questions all relate to this dataset.

   (a) Partition the employees into two groups: one group with employees with the emotional stability below the median level, and the other with the emotional stability above the the median level. Report the counts of the partitioned table in the 2x2 table. Compare the proportions of employees able to perform the task between the groups, and interpret the results.

   (b) Partition the employees into four groups, according to the quartiles of emotional stability, with roughly equal number of subjects. Report the counts of the partitioned table in the 4x2 table. Use Pearson χ^2 test to test for association between X and Y. Compare your conclusions to the conclusions in (a).

   (c) Fit a simple logistic regression, which expresses the association between the original values of X and Y. State the model and the assumptions. Report the parameter estimates.

   (d) Plot the scatterplot of the data. Overlay a smooth approximation with a lowess curve, and the fitted logistic regression. Does the model fit well?

   (e) Use the logistic regression to test for association between X and Y. Interpret the results. Compare your conclusions to the conclusions in (a) and (b).

   (f) Obtain $exp(\hat{\beta}_1)$ and it's 95% confidence interval. Explain this number.

   (g) What is the estimated probability that employees with emotional stability test score of 550 will be able to perform a task in a group?

   (h) Use predicted probability cutoff of 0.5 to classify the employee into two groups: those able to perform the task, and those not able. Report and interpret the confusion matrix.

   (i) Produce and interpret the ROC curve for this model.

2. In R, define a vector X using `x <- seq(from=-10, to=200, length=500)`.

(a) Calculate and plot the $P\{Y = 1\}$ for the corresponding simple logistic regression when $\beta_0 = -25$ and $\beta_1 = 0.2$. Overlay the plot for another logistic regression, which has a stronger positive association between X and Y. Interpret the pattern.

(b) Plot and interpret $P\{Y = 1\}$ for a logistic regression with two predictors, X and X^2, where the coefficients $\beta_0 = -25$ and $\beta_1 = 0.2$ as above, and the coefficient of X^2 $\beta_2 = -2$. Interpret the pattern.