STAT-S 432: Homework 4

Due March 29, 2019 by 11:59 PM

Instructions: You must submit this homework by pushing a file named "hw4.Rmd" file to your team's repo. Note that that is the **only** file you will be allowed to push. Commit early and often.

Data named n90_pol.csv

The data set `n90_pol.csv` contains information on 90 university students who participated in a psychological experiment designed to look for relationships between the size of different regions of the brain and political views. The variables `amygdala`and `acc` indicate the volume of two particular brain regions known to be involved in emotions and decision-making, the amygdala and the anterior cingulate cortex (ACC); more exactly, these are residuals from the predicted volume, after adjusting for height, sex, and similar anatomical variables. The variable orientation gives the subjects' scores on a five-point scale from 1 (very conservative) to 5 (very liberal). `orientation` is an ordinal but not a metric variable, so scores of 1 and 2 are not necessarily as far apart as scores of 2 and 3.

1. Creating a binary response variable. (5 points)
   - Create a vector, `conservative`, which is 1 when the subject's `orientation` is less than or equal to 2, and 0 otherwise.
   - Explain why the cut-off was put at an `orientation` score of 2 as opposed to some other cut-off.
   - Use code to check that your `conservative` vector has the proper values without manually examining all 90 entries.
   - Add the `conservative` vector that you created to the existing data frame.
2. Perform exploratory data analysis on the data. (7 points)
   - Examine the univariate distributions of the varaibles. Describe the distributions.
   - Examine bivaraite relations. This should include the following:
     - how the ACC and amygdala volumes change across the different orientation scores.
     - how the ACC and amygdala volumes relate to each other.
     - how the distribution of both ACC and amygdala brain volumes change whether someone is classified as a "conservative" or not.
   - Create a scatter plot of `amygdala` versus `acc`. Use shape and color to show whether an individual is a "conservative" or not on the scatterplot. Explain how the amygdala and ACC volumes may be related to the chance that a randomly selected student is a "conservative" or not.
3. Fit a logisitic regression model of `conservative` on `amygdala` and `acc`. (8 points)
   - Provide a kable-table of the coefficients, their standard errors, test-statistics, and p-values.
   - Summarize the table including a an interpretation of how the predictor varaibles affect the *odd ratio* (not the log-odds ratio).
   - Based on your table, is the model an effective way of classifying a randomly chosen student as a "conservative" or not. (Ignore the fact that we would have to get their brain volumes. . .)
4. We want to use the logistic regression model to classify individuals as conservative or not. (5 points)
   - Find predictions for each subject. Use the probability 0.5 as a cut-off for your classification.
   - What fraction of subjects are mis-classified?
   - What fraction of subjects would be mis-classified by "predicting" that none of them are conservative?
   - Based on this information, is this an effective model or not?

Rubric

Scoring will be done according to the following criteria.

Each question or bullet point has been addressed. Questions are answered accurately. Any plot or table produced is explained and relationships are described accurately.

If document does not knit (assuming required packages are installed) -5

Problem 1 (5 points)

Problem 2 (7 points)

Problem 3 (8 points)

Problem 4 (5 points)