

STA321 Week #6 Assignment

Cheng Peng

This weeks' assignment focuses on the simple linear regression model. Please follow the instructions to pick an appropriate data set for the following three assignments (combining the three individual assignments to complete project #2)

1. Find a data set that contains at least **two** categorical variables, **three** numerical variables, and a **binary response variable**. **Please go to D2L Discussion Board to post the information of your data set so your classmates will use your data for the assignment and project #2 (on logistic regression).**
2. The number of observations is at least 150. This number of observations is slightly larger than the one required in project #1 since we will introduce a machine learning algorithm using the logistic regression model.
3. If you have a data set with no binary response variable but you really want to use it for your assignments, you can dichotomize the response variable in a meaningful way. For example, you have a data set with several predictor variables that are potentially associated with the response variable GPA. You can dichotomize the continuous GPA in the following

$\text{bin.gpa} = 1$ if $\text{GPA} < 2.75$ $\text{bin.gpa} = 0$ if $\text{GPA} \geq 2.75$

This dichotomization is meaningful since most graduate schools use 2.75 as the admission cut-off of GPA.

4. Choose one numerical predictor variable or a **binary** predictor variable to fit a simple logistic regression model as I did in the case study in my class note as your assignment for this is week.

To be more specific, you need to provide the following key components in your analysis report.

- (a) Describe your data set and the variables.
- (b) Formulate a practically meaningful analytic question.
- (c) Perform exploratory data analysis using a graphical or a numerical approach.
- (d) Build a simple logistic regression model.
- (e) Interpret the regression coefficients from the practical perspective (odds ratio).
- (f) study the behavior of the success probability (probability curve and the rate of change in success probability).