Homework 4

Make a copy of this document or answer in a separate document.
Before you share make sure permissions are updated on this doc as well as the notebook you submit.

1. *(1 point) Visit section 3.3.3 Potential Problems and choose **one** of the problems to read about. Define the problem and the approach described in the book to address the problem.*
   Non-linearity of the data is when a linear regression model assumes that there is a straight line relationship between predictors and the response. This is a problem because if the true relationship is not linear, then virtually all of the conclusions that were drawn from the fit are suspect and prediction accuracy can be significantly reduced. Residual plots are a useful tool to identify non-linearity. If you can spot a pattern like a u-shape, then there is a problem. To fix this, use non-linear transformations of the predictors such as log X, X^2, or square root X in the model.

2. *(1 point) ISLP 3.7 problem 3.*
   a. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.
   b. $137.2k predicted starting salary
   c. False because scale needs to be considered instead of just magnitude. To test if a variable is significant, statistical significance tests should be performed. For example, using hypothesis tests and confidence intervals. Finding the p-value can show significance or not.

3. *(1 point) Read section 4.3.4 and explain how confounding predictors can lead to seemingly paradoxical behavior when comparing simple logistic models (with only one predictor) with multiple logistic models (which include both variables).*
   Confounding predictors can lead to seemingly paradoxical behavior when comparing simple logistic models with multiple logistic models because performing regressions involving a single predictor when other predictors may be relevant may result in stark differences. Since a simple regression model does not address the other predictors' effects on the outcome, it captures the combined effect of the predictors. This changes with a multiple logistic model because this model adjusts for the cofounder. This can lead to the paradoxical flip where the primary predictor's coefficient can flip because the confounding effect is now controlled.

4. *(1 point) Making Predictions. Table 4.3 shows the coefficients for several predictors to predict whether an individual will Default on a loan. Use the coefficients to make predictions for the **probability** of default of the following individuals. Also report the **odds** and **log(odds)** of default.*
   a. *Not a student, a balance of $1,400, and an income of $30,000*
      Log-odds = -10.8690+0.0057(1400)+0.0030(30)+(-0.6468)(0) = -2.799
      Odds = 0.0609
      Probability = 0.0574 or 5.74%

b. *A student with a balance of $3000, and an income of $19,000*
   Log-odds = 10.8690+0.0057(3000)+0.0030(19)+(-0.6468)(1) = 5.6412
   Odds = 281.8007
   Probability = 0.9965 or 99.65%

Problems 5 - 9 are in Google Colab : here
Attach your Google Colab notebook link here :