Exercise 7.1 Logistic Regression

Logistic Regression on Binary Y:

1. Re-run the Rscript passexam.R on passexam2.csv dataset. What is the cause of the error? Explain.
2. Execute logistic regression on default.csv dataset to predict default:
    a. Verify the baseline reference level for default.
    b. Which variables are statistically insignificant?
    c. Keeping only statistically significant variables, show the confusion matrix.
    d. Using set.seed(2) with 70-30 train-test splt, and keeping only statistically significant variables, show the trainset confusion matrix and testset confusion matrix.
    e. An analyst commented that AvgBal is a weak predictor of Default. Do you agree? Explain.

Questions for Research Paper [Freitas et. al. (2012)] Reading:

1. How are outliers determined? Why is this impt?
2. What is the difference between adjusted Odds Ratio and unadjusted Odds Ratio?
3. How did Freitas et. al. (2012)] identify high risk factors? *Hint: See their Table 1.*

Logistic Regression on Multi-category Y:

1. Set Service Rating = Neutral as the baseline reference level for Rating, in rating.csv dataset.
2. Develop Logistic regression to predict Rating using the multinom() function from Rpackage nnet. Which variables are statistically significant.
3. What is the model predicted service rating for each of the case in the dataset?
4. Show the confusion matrix.