STA 712 Challenge Assignment 2: Logistic Regression in the Wild

Due: Friday, September 30, 12:00pm (noon) on Canvas

Instructions:

- Submit your work as a single PDF. For this assignment, you may include written work by scanning it and incorporating it into the PDF. Include all R code needed to reproduce your results in your submission.
- You are welcome to work with others on this assignment, but you must submit your own work.

Introduction

Logistic regression is a popular choice for modeling a binary response, and is used in many applications. The purpose of this assignment is to see how logistic regression is used by practitioners in real research projects. In this assignment, you will find a research paper which applies logistic regression to real data. You will then write a short review of the paper, describing what the authors did, what their results were, and whether there were any issues in their methods. Finally, you will attempt to reproduce their results, using their data.

Questions

1. Find a research paper in the sciences (biology, physics, chemistry, etc.) or the social sciences (psychology, sociology, etc.) which uses logistic regression as a *main* analysis method to address a research question. Make sure to choose a paper for which the data is publicly available; part of this assignment will be to reproduce the original analysis.
2. Write a short peer review of your chosen paper. The purpose of this review is to describe how the authors used logistic regression in their analysis. Your review should answer the following questions:
   (a) What research question are the authors trying to answer?
   (b) What data do the authors collect to answer their research question?
   (c) What statistical methods do the authors use to analyze their data; in particular, how do the authors use logistic regression to help answer the question?
   (d) Does their use of logistic regression appear reasonable, or are their any issues with their analysis?
   (e) Is their anything you would suggest to improve the paper?
3. Using the data associated with the paper, fit the same logistic regression model as the authors. (If the paper fits several logistic regression models, you may choose one). If the paper does not provide sufficient detail to reproduce their model, explain what additional detail you would need, then fit your best guess at their logistic regression model.
4. Assess the logistic regression assumptions using model diagnostics from class (empirical logit plots, quantile residual plots, leverage and Cook's distance, and variance inflation factors).
   (a) Are there any issues with the fitted model? If so, do your best to address these issues (e.g. with variable transformations), and report your final model.

(b) What diagnostics (if any) did the researchers use to assess the logistic regression assumptions in their paper?

5. Does your logistic regression model support the researchers' conclusions in the original paper? Carry out any additional analysis necessary to answer the research question (e.g., hypothesis testing), and summarize your findings.