DSC 241 - Homework 7

Problem 1. Find the dataset **Placekick.csv** in the **Datasets** subfolder. Use this dataset to build a logistic regression model to estimate the probability of success for a placekick. Here is the data dictionary:

- **week**: Week of the season
- **distance**: Distance of the placekick in yards
- **change**: Lead-change (1) vs. non-lead-change (0) placekicks
- **elap30**: Number of minutes remaining before the end of the half
- **PAT**: Type of placekick, where a PAT attempt is a 1 and a field goal attempt is a 0
- **type**: Outdoor (1) vs. dome (0) placekicks
- **field**: Grass (1) vs. artificial turf (0) placekicks
- **wind**: Windy conditions (1) vs. nonwindy conditions (0)
- **good**: Successful (1) vs. failed (0) placekicks; this is our response variable

a. Fit a logistic regression model with **good** as response and **distance** as predictor. Interpret the fitted model coefficients and visualize the model fit.

b. Now consider all predictors. Apply the forward selection algorithm to compute the forward selection path from 'intercept only' to 'full model' and chooses the model on that path that minimizes the AIC.

c. Consider the model selected by the forward selection algorithm. Compute the decision boundary when the decision threshold for the probability of success is 0.5.

Problem 2

a. Write a function **bootGLM(x, y, B=1000)** that resamples observations and returns standard errors for each of the predictor variables (when the others are present in the model) in a logistic model.

b. Consider the model selected by the forward selection algorithm from Problem 1(b). Apply your **bootGLM**, and compare with the standard errors returned by the summary function.