

STOR 664: Homework 5
Due Date: Thursday, November 10, 2022

These are all exercises from Chapters 10 and 11 of Faraway's *Linear Models with R* (second edition). To load the datasets, first install the `faraway` package and then type `library(faraway)` followed by `data(dataset)` where "dataset" is the name of the data you're loading. Also note that after installing `dataset`, you can use the help command `?dataset` to get more information about the background of the data. None of the problems should require R expertise beyond what we have covered in class, but if you do have Faraway's book, you may well find other ideas for R routines that are useful to linear regression.

1. Use the `prostate` data with `lpsa` as the response and the other variables as predictors. Implement the following variable selection methods to determine the "best" model:
 - (a) Backward elimination
 - (b) AIC
 - (c) Adjusted R^2
 - (d) Mallows' C_p
2. Fit a linear model to the `stackloss` data with `stack.loss` as the response and the other variables as predictors. Simplify the model if possible. Check the model for outliers and influential points. Now return to the full model, determine whether there are any outliers or influential points, eliminate them and then repeat the variable selection procedures. Write a summary of your conclusions.
3. Use the `seatos` data with `hipcenter` as the response.
 - (a) Fit a model with all eight predictors. Comment on the effect of leg length on the response.
 - (b) Compute a 95% prediction interval for the mean value of the predictors.
 - (c) Use AIC to select a model. Now interpret the effect of leg lengths and compute the prediction interval. Compare the conclusions from the two models.
4. This is Problem 4, page 181, from Faraway:

Load the dataset `library(faraway)` and `data(fat)`. The response variable is percentage of body fat, `siri`, and the predictors are all the other variables except `brozek` and `density`. Remove every tenth observation from the dataset to use as a test sample. Use the remaining data as a training sample to build the following models:

- (a) Linear regression with all predictors
- (b) Linear regression with variables selected using AIC
- (c) Principal component regression
- (d) Partial least squares regression

(e) Ridge regression

(f) LASSO

Use the models you find to predict the response in the test sample. Based on these conclusions, write a short report summarizing the relative performances of methods (a) through (f) on this dataset.