Statistics 550 Homework assignment 4

Problem 1 is problem 12.3 (note two errors) in the text.

2. For the linked automobile data, fit a regression model to predict miles per gallon (mpg) from volume (vol), horsepower (hp), speed (sp), and weight (wt). Check residual-by-predicted, residual-by-covariate, and residual-by-hat plots as well as a normal plot of the residuals. Comment on the validity of the model given these plots. Now use the Box-Cox method to find an optimal transformation of y. Also check a partial regression plot of the constructed variable for the Box-Cox transformation. Perform a new regression on the transformed dependent variable and check the same plots as above. Is the new model better?

3. Exercise 13.8 (Hint: Use the result that for two nonsingular k x k matrices A and B, the matrix AB is nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$)

4. For the linked baseball data set, fit a multiple regression model to predict salary in 1987 (salary87) from at bats, hits, homeruns, runs scored, runs batted in, and walks in 1986 (ab86 h86 hr86 r86 rbi86 w86) and years of major league service (years). Calculate variance-inflation factors and perform a principal components analysis of the covariates to assess collinearity.

5. For the data above from problem 4, use stepwise selection and best subsets regression methods (with MSE, AIC, and BIC) to investigate models for predicting the response variable. Use the same covariates as in the problem above. Which model(s) is(are) best?

6. Use a ridge regression estimator for the regression model from problem 2. Pick a value for the ridge constant d by looking at the ridge trace figure - how does your estimate of $\boldsymbol{\beta}$ compare to the estimate from problem 2?

7. Give me an update on your progress for finding a topic for your class talk at the end of the semester.