DSA211 Statistical Learning with R

Homework 4

Use R functions and data file (GPAGMAT.csv) to solve the following problems:

1. The director of graduate studies at a large college of business has the objective of predicting the grade point average (GPA) of students in an MBA program. The director begins by using the Graduate Management Admission Test (GMAT) score. A sample of 80 students who have completed two years in the program is randomly selected and stored in GPAGMAT.csv.

   (a) Construct a scatter plot with the estimated regression equation.

   (b) Find the Y-intercept and slope of the regression equation and the adjusted r^2.

   (c) Perform a residual analysis on your results and evaluate whether the assumptions of regression have been seriously violated.

   (d) At the 0.05 level of significance, is there evidence of a linear relationship between GMAT score and GPA? What are the values of test statistic and p-value?

   (e) Construct a 90% confidence interval estimate of the population slope.

   (f) Construct 95% confidence interval estimates of the means GPA of students with GMAT scores of 600, 630, 660, or 690.

   (g) Construct 99% prediction interval of the GPA for an individual with GMAT scores of 600, 630, 660, or 690.

-END-