HUDM 5126 Linear Models and Regression Analysis Homework 3*

Yifei Dong

Sep 16 2020

1 Grade Point Average

The question is adapted from Q3.3 Refer to Grade Point Average data in Problem 1.19.

(a) Prepare a boxplot of the ACT scores (X variable). Ate there any noteworthy features on the plot?

(b) Prepare a histogram of the residuals. What information does the plot provide?

(c) Plot the residuals against the fitted values $\widehat{Y}$. What are your findings about departures from the regression assumptions?

(d) Prepare a normal probability plot of the residuals. Test the reasonableness of the normality assumption with the KS test using $\alpha = 0.05$. What do you conclude?

(e) Conduct the BP test to determine if the variance varies with the level of X. Use $\alpha = 0.05$. State your conclusion. Does your conclusion support your preliminary findings in part c)?

2 Per capita earnings

A sociologist employed linear regression model (2.1) to relate per capita earnings (Y) to average number of years of schooling (X) for 12 cities. The fitted values $\widehat{Y}_i$ and the semistudentized residuals e_i^* follow.

(a) Plot the semistudentized residuals against the fitted values. What does the plot suggest?

(b) How many semistudentized residuals are outside ± 1 standard deviation? Approximately how many would you expect to see if the normal error model is appropriate?

3 R and RStudio

Read the data frame from the file HW3.RData by double-clicking on it and opening with RStudio.

*This homework is written in LaTeX.