

Economics 144: Homework 1
Winter 2019, UCLA
Instructor: Dr. Rojas

Due Date: January 22, 2019

1. (40%) For this problem we follow the study by Lalonde (1986) on earnings related to various factors such as training, race, gender, etc. The data are available directly from the DAAG package (the data file is `nsw74psid1`). Your task will be to fit a multiple regression model of the form $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots$, where \hat{y} = real earnings in 1978 (`re78`), and for the predictors, you will need to decide which ones to keep. The complete assignment needs to be typed, include all the plots, and the R source code as well.
 - (a) Plot a histogram (include the respective density curve) of each variable and discuss its properties.
 - (b) Estimate your full regression model $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots$ and show your results (e.g., the output from the ‘summary’ command in R). Discuss these results.
 - (c) Compute the Mallows CP statistic for all the plausible models and choose only one model. Discuss why you chose this model. For the next questions, only use your selected model from this part.
 - (d) Plot the residuals vs. the fitted values.
 - (e) Plot and discuss the VIF plot.
 - (f) Plot and discuss the correlation graph (use the `corrplot` library)
 - (g) Plot the Cook’s Distance values. Are there any outliers? If so, discuss what you would do with them.
 - (h) Plot a histogram of the residuals and discuss the results.
 - (i) Plot the QQ Normal Plot and discuss the results.
 - (j) Plot the observed vs. predicted values, overlay a Lowess smoother, and discuss the results.
2. (15%) Problem 2.2 (i.e., Chapter 2, Problem 2) from Textbook^a.
3. (15%) Problem 3.1 (i.e., Chapter 3, Problem 1) from Textbook^a.
4. (15%) Problem 3.3 (i.e., Chapter 3, Problem 3) from Textbook^a.
5. (15%) Problem 4.3 (i.e., Chapter 4, Problem 3) from Textbook^a.