Spring 2020 Quantitative Analysis
Chung-Ming Kuan

Problem Set 2

Due: 3/16

Part One: Hand-Written Exercise

1. Verify the statement in slide 16, Lecture 2. That is, let $\hat{r}_{i,1}$ be the OLS residual of regressing x_1 on the constant one and $x_2, ..., x_k$. Show that $\sum_{i=1}^n \hat{r}_{i,1} x_{i,1} = \sum_{i=1}^n \hat{r}_{i,1}^2$. (25 points)

2. Suppose $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{u}_i$. Under Classical Assumptions, discuss whether the OLS estimator $\hat{\beta}_1$ is biased in each of the following scenarios. Justify your answers with brief explanation. (25 points)

   (a) $\text{Var}(y_i)$ is no longer a constant, but a function of x_i instead.

   (b) $\mathbb{E}(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$

   (c) $\mathbb{E}(y_i) = \beta_0 + \beta_1 x_{i1}$

3. For the multiple linear regression, the data matrix noted by $\mathbf{X}$ is below:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 1 & 8 \\ 2 & 4 & 5 & 7 \\ 3 & 6 & 2 & 9 \\ 4 & 8 & 2 & 2 \end{bmatrix}$$

   For this data matrix, can you calculate the OLS estimators? Why or why not? Please give a brief explanation. (25 points)

4. Suppose there is a multiple linear regression model without intercept term, that is, $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i$. Please show that $SST \neq SSR + SSE$, which means R^2 may not be between 0 and 1. (25 points)

Part Two: Computer Exercise

1. (100 point)
   Please load the data set "mtcars" from R using the code `data(mtcars)`. The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models).

   (a) Please show the data for the automobile "Duster 360".

(b) Please show the `qsec` (1/4 mile time) for all the automobile.

(c) Please show the data with `cyl` (number of cylinders) = 6.

(d) Please list the automobiles with `mpg` (miles/gallon) > 15, `vs` (Engine) = 1, and `hp` (horsepower) between 50 and 150.

(e) Suppose we have the following model:

$$\texttt{drat}_i = \beta_0 + \beta_1 \texttt{wt}_i + \beta_2 \texttt{hp}_i + \beta_3 \texttt{qsec}_i + \beta_4 \texttt{vs}_i + u_i,$$

and we wish to test the hypothesis $H_0 : \beta_1 = \beta_3 = 0$. Please construct the constrained and unconstrained model, obtain R^2_{ur} and R^2_r and construct the F statistic.

(f) Instead of R^2_{ur} and R^2_r, please obtain SSR_{ur} and SSR_r and recalculate the F statistic. Verify that it's identical to (e).

(g) Use the function `linearHypothesis()` to directly obtain the F statistic. Verify that it's identical to (e).