PO 7005: Assignment 1

Thomas Chadefaux

NOTE: Always justify your answer. Show R code when relevant. Circle or highlight answer. Scanned version of (neatly) handwritten answers are fine.
Late submissions will NOT be accepted, as I will circulate the answer key immediately.

Question A (10%)

Consider the following vector $\mathbf{e}$ and matrix $\mathbf{W}$:

$$\mathbf{e} = \begin{pmatrix} -0.91 \\ -0.3 \\ -0.34 \\ 0.38 \end{pmatrix} \qquad \mathbf{W} = \begin{pmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.16 & 0 & 0 \\ 0 & 0 & 0.29 & 0 \\ 0 & 0 & 0 & 0.35 \end{pmatrix}$$

Write down the result of (2% each):

1. $\sum_{i=1}^{4} e_i$ and $\sum_{i=1}^{4} w_{ii}$ and $\sum_{i=1}^{4}\sum_{j=1}^{4} w_{ij}$
2. $\mathbf{e}'\mathbf{e}$
3. $\mathbf{e}\mathbf{e}'$
4. $\mathbf{e}'\mathbf{W}\mathbf{e}$
5. $\mathbf{W}'\mathbf{W}$

Question B (20%)

Consider the following matrices:

$$\mathbf{X} = \begin{pmatrix} 1 & 0.3 & -0.5 \\ 1 & 0.52 & -1.3 \\ 1 & -0.14 & 2.1 \\ 1 & -1.95 & 1.2 \\ 1 & 0.3 & 2.1 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 2 \\ 4 \\ 0 \\ -1 \\ 0.3 \end{pmatrix} \qquad \beta = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

Calculate (4% each):

1. $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}'\mathbf{X}$
2. Use R to find $(\mathbf{X}'\mathbf{X})^{-1}$ and $|\mathbf{X}'\mathbf{X}|$. Show your code and results. What are the dimensions of $\mathbf{X}'\mathbf{X}$?
3. Produce R code to calculate $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$. Show your code and results
4. In R, plot y against X_2 (the second column of X). Estimate the model $y = \beta_0 + \beta_1 X_2 + \varepsilon$ using OLS and plot the resulting estimated line. What is the equation corresponding to that line? What are its coefficients? Label the line in R, return your plot properly labeled.
5. Calculate the residuals $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$. What is the sum of squared residuals?

Question C (30%)

Import data.csv in R (data.csv is available online in the homework folder). Explain your answers to (5% each):

1. Using OLS, regress y1 on x1. Regress y2 on x2. Regress y3 on x3. Regress y4 on x4. Report your estimated coefficients for each.
2. Plot yi against xi for $i \in \{1, 2, 3, 4\}$. Add the estimated regression line to the plot. Write down the regression equation with the estimated coefficients.
3. Generate 11 observations drawn from a random normal distribution and store them in a variable called rnv1. Regress y1 on x1 and rnv1. How has the coefficient associated with x1 changed?
4. Repeat the previous step 1,000 times. Collect the coefficients associated with x1. What is their mean? Plot their distribution using a histogram overlaid with a density plot.
5. Does your result change if you draw rnv1 from a normal distribution with mean 10 instead? What if you draw it from a chi-square distribution with 2 degrees of freedom? What if you multiply x1 by 3?
6. Create a variable $x_5 = \begin{pmatrix} 3 & 3 & 3 & \dots & 3 \end{pmatrix}$. Regress y1 on x1 and x5. What happens? Why? What if instead you define $x_5 = 2x_1$? What if $x_5 = x_1^2$? Be sure to explain and provide R code.

Question D (40%)

Suppose your model is

$$y_i = \beta_0 + \varepsilon_i \tag{1}$$

(5% each):

1. write (1) in matrix form. Be sure to specify the matrices clearly before you state the formulation.

2. What is the least square estimate of β_0?

   Suppose now that your model is

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i \tag{2}$$

3. write (2) in matrix form. Be sure to specify the matrices clearly before you state the formulation.

4. What is the least square estimate of β_0? Of β_1?

5. Show that the least squares normal equations imply that (i) $\sum_i e_i = 0$ and (ii) $\sum_i x_i e_i = 0$

6. What is the interpretation of b_0? Of b_1?

7. Suppose now that you are considering two models :

$$y_i = \alpha_0 + \alpha_1 x_1 + u_i \tag{3}$$
$$y_i = \beta_0 + \beta_1 (x_1 + 3) + v_i \tag{4}$$

   Explain intuitively why $a_0 \neq b_0$, but that $a_1 = b_1$, where a_0 and a_1 are the OLS estimates of α_0 and α_1 respectively (and correspondingly with b_i and β_i).

8. Suppose now that your are interested in a modified dependent variable, namely $z_i = 3y_i$. Your model is

$$z_i = \beta_0 + \beta_1 x_1 + \varepsilon_i \tag{5}$$

   How do your estimates of β_0 and β_1 in this model differ from those in model (2)? Why?