Econometrics I, ECO341A, Summer 2023
Homework I (100 points)

Instructor: M.A. Rahman

Deadline: 4:00 pm, June 7, 2023.

Please read the instructions carefully and follow them while writing answers.

- *Solutions to homework should be written in A4 size loose sheets. If you are not comfortable writing on white sheets, please ask for biology paper in Tarun Book Store.*
- *Questions should be answered in order as they appear in the homework. Every new question should begin in a new page. Please number all the pages of your homework solution.*
- *Please leave a margin of one inch from top and one inch from left. Staple the sheets on the top-left.*
- *For Matlab assignments (if any), please answer the questions. Please do not dump the codes as an answer, they should be in the appendix.*
- *Please write your name and names of your group members on the first page of your answer script.*

---

1. **($3 \times 8 = 24$ points) Distribution, Moments and MGF's**: Write down the pdf's and **derive** the mean, variance and MGF's of the following distributions.

(a) Logistic distribution

(b) Chi-square distribution

(c) Laplace distribution

(d) Student-t distribution with ν degrees of freedom

(e) Gamma distribution (*Statistical Inference* by Casella and Berger (henceforth, CB), page 99)

(f) Beta distribution (CB, page 106)

(g) $\mathrm{F}(\nu_1, \nu_2)$ distribution. Does the MGF exists?

(h) log-normal distribution. Does the MGF exists?

2. **(3+2+2+1+1+2+2+2+2+3 = 20 points)**. The data in Table 1, for a consumer finance company operating in six cities, the number of competing loan companies operating in the city (X) and the number per thousand of the company's loan made in that city that are currently delinquent (Y). Assuming a simple linear regression model:

$$y_i = \beta_1 + x_i\beta_2 + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2),$$

and define as usual $y = (y_1, \cdots, y_n)'$, $X = [1_{6\times 1},\ x_i']$. Based on the above setting answer the following.

Table 1: Six Cities Data

	1	2	3	4	5	6
x_i	4	1	2	3	3	4
y_i	16	5	10	15	13	22

(a) Using the matrix method, find $y'y$, $X'X$ and $X'y$.

(b) Vector of estimated regression coefficients $\hat{\beta}$. Interpret $\hat{\beta}_1$ and $\hat{\beta}_2$ in the context of the problem.

(c) Vector of residuals $\hat{\epsilon}$.

(d) Sum of squares due to regression (SSR).

(e) Sum of squares due to errors (SSE).

(f) Covariance matrix of $\hat{\beta}$, i.e., $Cov(\hat{\beta})$.

(g) Report the standard error of $\hat{\beta}_1$ and $\hat{\beta}_2$.

(h) Find the projection matrix P and residual generator matrix M.

(i) Find the standard error of regression $\hat{\sigma}$ and R^2. Comment on R^2.

(j) Point estimate of $E(Y_h)$ and $\hat{\sigma}(y_h)$, when $x_h = 4$.

Note: The coefficient of determination $R^2 = \frac{\hat{\beta}'X'X\hat{\beta} - n\bar{y}^2}{y'y - n\bar{y}^2} = \frac{\hat{\beta}'X'M_0X\hat{\beta}}{y'M_0y}$, where $M_0 = [I_n - \iota(\iota'\iota)^{-1}\iota']$, I_n is an identity matrix of dimension $n \times n$ and ι is a column vector of ones with size $n \times 1$.

3. **(3+3+2+4+8 = 20 points.)** Consider the linear regression model,

$$y = X\beta + \epsilon, \tag{1}$$

where as usual y and ϵ are of dimension $n \times 1$, X is $n \times k$ and β is $k \times 1$. However, X does **not** contain a column of 1s i.e., there is **no** intercept in the model. Let M_0 be the deviation-from-mean matrix of size n as defined in Question 2. Assume that you pre-multiply the model (1) by M_0. Based on this setting, answer the following.

(a) Show the form of the resulting OLS estimator and call it $\tilde{\beta}$. (*Hint*: Define $\tilde{y} = M_0 y$ and $\tilde{X} = M_0 X$).

(b) Do you get the same result if you only de-mean your data X, but not y?

(c) What if you only de-mean your outcome variable y, but not X? Call this estimator $\ddot{\beta}$.

(d) Now let the error term be correlated with the data X such that $E(\epsilon|X) = \gamma = [\gamma_1, \gamma_2, \ldots, \gamma_n]' \neq 0$. Assume you de-mean X and run OLS. Under the assumption of the structure of the bias γ, $\tilde{\beta}$, is the OLS estimator $\tilde{\beta}$ a biased estimator of β?

(e) Now assume, for the true model in (1), that $E(\epsilon|X) = \iota\gamma$, where ι is a column of 1's and γ is a fixed-valued scalar. Show that under these new assumptions, the original OLS estimator $\hat{\beta} = (X'X)^{-1}(X'y)$ is still biased, but the deviation from the mean estimator $\tilde{\beta}$ becomes an unbiased estimator for β.

4. **(10 points)** For a given sample of size n, express the coefficient of determination (R^2) as a function of the mean squared error (MSE) defined as $MSE = \hat{\epsilon}'\hat{\epsilon}/n$. Show that R^2 has to increase whenever MSE decreases.
(*Hint*: Make use of the matrix $M_0 = [I_n - \iota(\iota'\iota)^{-1}\iota']$, where I_n is an identity matrix of dimension $n \times n$ and ι is a column vector of ones with size $n \times 1$.)

5. **(3+3+3+3+3+3+5+3=26 points)** Consider the data in the file 'time.xlsx' and the model (not relevant for Parts (a) and (b)),

$$y_t = C_t\beta_{11} + D_t\beta_{21} + x_{t2}\beta_2 + x_{t3}\beta_3 + e_t \tag{2}$$

where $C_t = 1 - D_t$ and

$$D_t = \begin{cases} 1 & \text{if year} = 1939, \cdots, 1945 \\ 0 & \text{otherwise} \end{cases}$$

(a) Regress y_1 on intercept x_2 and x_3. Find the regression coefficients, standard errors, R^2 and standard error of regression.

(b) Regress y_3 on intercept x_2 and x_3. Find the regression coefficients, standard errors, R^2 and standard error of regression.

(c) Regress y_1 on C_t, D_t, x_2 and x_3. Estimate the variance-covariance matrix. Find the regression coefficients, standard errors, R^2 and standard error of regression. Interpret the coefficient of C_t and D_t. Is there any problem if we include an intercept term in this regression. Explain.

(d) Using estimates from part (c), find var ($\hat{\delta}$) where $\hat{\delta} = \hat{\beta}_{21} - \hat{\beta}_{11}$.

(e) Regress y_1 on D_t, x_2 and x_3 and include an intercept. Interpret the coefficient of D_t. Find the regression coefficients, standard errors, R^2 and standard error of regression.

(f) Create two variables $x_{t4} = x_{t3}D_t$ and $x_{t5} = x_{t3}C_t$. Regress y_2 on intercept, D_t, x_2, x_3 and x_4. Find the regression coefficients, standard errors, R^2 and standard error of regression.

(g) Regress y_2 on C_t, D_t, x_2, x_4 and x_5. Estimate the variance-covariance matrix. Find the regression coefficients, standard errors, R^2 and standard error of regression. Why don't we include x_3 in this model? Test the hypothesis $\beta_4 = \beta_5 = 0$.

(h) Create another variable $x_{t6} = x_{t2}D_t$ and regress y_3 on intercept, D_t, x_2, x_3, x_4 and x_6. Find the regression coefficients, standard errors, R^2 and standard error of regression.