Department of Inter Disciplinary Studies,
Faculty of Engineering,
University of Jaffna, Sri Lanka
MC 3020 - Assignment 04

30 minutes **14 - 07 - 2023**

Important instructions:

- **Answer all the questions (1-2).**
- **If it is determined that you have violated any policies during this exam, you will receive a score of zero for this assignment, without any exceptions or considerations.**
- **If you have any doubt as to the interpretation of the wording of a question, make your own decision, but clearly state in the script.**

---

1. Suppose you are working as a civil engineer and you are tasked with predicting the compressive strength of concrete based on the age of the concrete sample. You collected data from various concrete samples and measured their age (in days) and compressive strength (in megapascals). The dataset is as follows:
   Age (days): 7, 14, 21, 28, 35, 42, 49, 56
   Compressive Strength (MPa): 15, 20, 23, 27, 30, 33, 37, 40

   (a) Plot the data points on a scatter plot.
   (b) Calculate the equation of the least squares regression line for predicting compressive strength based on the age of the concrete.
   (c) Interpret the slope and intercept coefficients of the regression line in the context of the problem.
   (d) Use the regression line to predict the compressive strength of a concrete sample that is 30 days old.
   (e) Calculate the coefficient of determination (R-squared) for the regression line and interpret its meaning.
   (f) Perform a hypothesis test to determine whether the regression line is statistically significant at a 90% confidence level. State your hypotheses, the test statistic, and the conclusion clearly.
   (g) Discuss any assumptions or limitations of the linear regression model in this context.

2. Suppose you are conducting an experiment to study the relationship between the power consumption of a household appliance and various factors such as voltage (V), current (A), and operating time (hours). You collected data from multiple households and obtained the following dataset:

Household	Voltage (V)	Current (A)	Operating Time (hours)	Power Consumption (W)
1	220	5	4	800
2	240	6	5	960
3	230	4	3	690
4	220	5	6	1080
5	250	7	4	1400
6	230	6	5	1380

   To answer this question using the output of the given R programs (without necessarily employing any formulas),

   (a) Find the multiple linear regression line to determine the relationship between the power consumption (dependent variable) and the voltage, current, and operating time (independent variables).
   (b) Interpret the coefficients of the regression equation in the context of the problem.
   (c) Use the regression equation to predict the power consumption for a household with the following specifications: voltage = 235 V, current = 5.5 A, and operating time = 4 hours.
   (d) Calculate the coefficient of determination (R^2) and interpret its meaning.
   (e) Perform a hypothesis test to determine whether the regression equation is statistically significant at a 95% confidence level. State your hypotheses, the test statistic, and the conclusion.

```
> # Summary of the model
> summary(model)

Call:
lm(formula = PowerConsumption ~ Voltage + Current + OperatingTime,
    data = data)

Residuals:
      1       2       3       4       5       6
-151.55 -205.49   75.77   53.94   53.94  173.38

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    412.606   3658.272   0.113    0.920
Voltage         -4.113     17.775  -0.231    0.839
Current        258.944    199.670   1.297    0.324
OperatingTime   37.254    139.850   0.266    0.815

Residual standard error: 231.1 on 2 degrees of freedom
Multiple R-squared:  0.753,	Adjusted R-squared:  0.3825
F-statistic: 2.033 on 3 and 2 DF,  p-value: 0.3466

> 
```

Figure 1: R output for question 2 of Assignment 4

(f) Perform the hypothesis test for testing the significance of the coefficient of the voltage variable and state the conclusion.

Formula sheet:

1. Linear regression coefficient estimation formulas,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

2. Standard error (S_e) is given by

$$S_e = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-2}}$$

3. Correlation (r) is given by

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)\left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right)}}$$

4. Test statistic value when testing the hypothesis for correlation is given by,

$$T = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}, \quad \text{d.f} = n-2$$

5. Test statistic value when testing the hypothesis for slope coefficient is given by,

$$T = \frac{\hat{\beta}_1 - 0}{\frac{Se}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}}}, \quad \text{d.f} = n-2$$

6. Test statistic value when testing the hypothesis for intercept coefficient is given by,

$$T = \frac{\hat{\beta}_0 - 0}{Se\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}}}, \quad \text{d.f} = n-2$$