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

40 minutes **29 - 07 - 2024**

Answer all the questions (1-2).

1. You are a mechanical engineer working in the automotive industry, tasked with optimizing the performance of internal combustion engines. A key performance metric you are focusing on is the thermal efficiency of the engines, which is influenced by the compression ratio - a critical design parameter. To better understand this relationship and aid in the design of more efficient engines, you collected data from various engine models. The data includes the compression ratio of the engines and their corresponding thermal efficiency percentages, as shown below:

   Compression Ratio: 8.5, 9.6, 10.1, 11.2, 12.1, 13
   Thermal Efficiency (%): 31, 35, 42, 45, 56, 55

   (a) Create a scatter plot of the data points, with the compression ratio on the x-axis and thermal efficiency on the y-axis.

   (b) Determine the equation of the least squares regression line for predicting thermal efficiency based on the compression ratio. Explain the meaning of the slope and intercept coefficients in the context of the problem.

   (c) Use the regression line equation to predict the thermal efficiency for an engine with a compression ratio of 10.5.

   (d) Calculate the correlation (r) between compression ratio and thermal efficiency. Interpret the correlation value.

   (e) Conduct a hypothesis test at a 95% confidence level to determine whether the regression line is statistically significant. State the hypotheses, test statistic, and conclusion.(Hint: Standard error (S_e) = 2.8521)

   (f) Discuss any assumptions or limitations of the linear regression model in this context.

2. As a civil engineering student working on final year project at the Ariviyal Nagar Weather Station, you are tasked with predicting the soil moisture content at a 30cm depth based on environmental factors such as temperature, wind speed, solar radiation, and overall moisture levels. The datasets you have collected is as follows:

Temperature (°C)	Wind Speed (km/h)	Solar Radiation (W/m²)	Moisture Level (%)	Soil Moisture Content at 30cm Depth (%)
25	10	1200	90	15
27	15	1250	85	13
30	20	1300	60	11
32	25	1350	45	20
34	30	1200	70	19
36	35	1150	35	17
38	40	1100	30	15
40	45	1050	25	14

   To utilize this data for predictive analysis and better manage water resources, you need to answer the following questions based on the given R output:

   (a) Determine the equation for predicting soil moisture content at a 30cm depth based on temperature, wind speed, solar radiation, and moisture level.

(b) Explain the meaning of each coefficient in the context of the problem. Describe how changes in each environmental factor affect the soil moisture content.

(c) Use the regression plane equation to predict the soil moisture content at a 30cm depth for the following conditions: temperature of 29°C, wind speed of 22 km/h, solar radiation of 1325 W/m², and moisture level of 65%.

(d) Determine the R-squared value for the regression plane and interpret its meaning in terms of how well the model explains the variability in soil moisture content.

(e) Conduct a test to determine the significance of the temperature variable in predicting soil moisture content. State the hypotheses, the test statistic, and the conclusion clearly.

```
Call:
lm(formula = Soil_Moisture_Content_30cm_Depth ~ Temperature +
    Wind_Speed + Solar_Radiation + Moisture_Level)

Residuals:
      1       2       3       4       5       6       7       8
 2.5945 -2.5945 -4.1716  2.5945  1.3905  1.9925 -0.2711 -1.5348

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.02736  141.28341   0.000    1.000
Temperature      -1.70149    8.16152  -0.208    0.848
Wind_Speed        1.12139    3.66935   0.306    0.780
Solar_Radiation   0.02925    0.03641   0.804    0.480
Moisture_Level    0.09552    0.16211   0.589    0.597

Residual standard error: 3.913 on 3 degrees of freedom
Multiple R-squared:  0.2823,	Adjusted R-squared:  -0.6745
F-statistic: 0.2951 on 4 and 3 DF,  p-value: 0.8654
```

Figure 1: R output for question 2 of Assignment 4

Formula sheet:

1. Linear regression coefficient estimation formulas,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

2. Correlation (r) is given by

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

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

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

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

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

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

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