STAT 3232– Data Analysis & Preparation of Statistical Reports
Tutorial #07

01. You are provided with a data set containing information about the students' exam scores based on the number of daily hours they spend studying.

Hours	Score
2.5	21
3.5	24
4	30
5.1	36
6	40
7	45
8.1	50
9	55
10	60
11	65
3.5	24
9.8	60
4.5	34
5	35
4	32

a) Identify the independent and dependent variables.

b) Calculate the correlation between these variables and interpret.

c) Obtain the scatter plot for variables and comment on it.

d) Fit the simple linear regression model.

e) Obtain a summary of the model and interpret the model output with suitable hypothesis.

f) Interpret the model coefficients.

g) Predict the score of the student when study hour is 8.8.

02. Consider the given dataset called "Solar Thermal Energy". The dataset represents total heat flux (kwatts) of solar thermal and its five effecting variables.

Y: Total heat flux (kwatts)

X1: Insolation (watts/ m^2)

X2: Position of focal point in east direction (inches)

X3: Position of focal point in south direction (inches)

X4: Position of focal point in north direction (inches)

a) Read the data set called "Solar Thermal Energy" into R.
b) Find the correlation between all the variables and interpret.
c) Plot the scatter plot diagram for all the variables and comment on it.
d) Fit the multiple linear regression model for the given data set.
e) Obtain the summary of multiple linear regression model and comment the significant of each variable.
f) Check the assumptions.

Submit on or before 28th of May 2024 at 4.00 p.m.

Note that your commands should be written in R editor. Both commands and outputs should be copied into a word file and upload to the LMS as a pdf document.