Logo

Description automatically generated

**LAB**

SEPT 2023

TEB2043

Data Science

*Lab 4*

|  |  |  |  |
| --- | --- | --- | --- |
| **NO.** | **NAME** | **STUDENT ID** | **PROGRAM (IT / IS / CS / BM)** |
| **1.** | **CHENG PIN JIE** | **21000548** | **CS** |

*Mdm. Syahirah*

**R Script**

#1 Load data

df <- read.csv("toydata.csv")

#2 Create linear regression model

model <- lm(y~X, data = df)

#2 Find intercept and coefficient

summary(model)

#2 For every unit increase in x, y would be increase by 1.1035

#3 the p-value associated with the regression coefficient for variable X is 0.00193.

#3 the p-value for the regression coefficient suggests that X is significantly associated with the target variable y.

#3 In this output, the p-value (0.00193) is less than 0.05, indicating that X is significantly associated with y.

#4 Predict

model\_predict <- predict(model)

#5 Remove Outlier Point and Create lm

df\_RO <- df[-10, ]

model\_RO <- lm(y~X, data = df\_RO)

summary(model\_RO)

model\_predict\_RO <- predict(model\_RO)

#5 Estimates: from -0.1883 to -1.6180

#5 p-value: from 0.00193 to 4.215e-06

#5 R^2: from 0.7193 to 0.9588

#6 Plot

plot(df)

lines(model\_predict, col= "red")

lines(model\_predict\_RO, col = "blue")

#6 Predicted model that remove outlier (Blue) fitted way better than model that outlier is not remove (red).

