

Practical No 13 Module II

Aim : Performing linear regression analysis using lm() (R).

Output :

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

Console Terminal Background Jobs

R - R4.5.1 - ~

```
> # Practical 13: Linear Regression using lm()
>
> # Install required packages (run once)
> #install.packages("ggplot2")
> #install.packages("readr")
> #install.packages("rlang")
> #install.packages("vctrs")
>
> # Load libraries
> library(ggplot2)
> library(readr)
>
> # Load the CSV file
> student_data <- read_csv("data_science_student_marks.csv")
Reading CSV into data frame: 8 columns
  Column specification:
dlimiter: ","
chr (1): location
dbl (7): student_id, age, sql_marks, excel_marks, python_marks, power_bi_marks, english_marks

i use `spec()` to retrieve the full column specification for this data.
i specify the column types or set `show_col_types = FALSE` to quiet this message.
>
> # View first few rows
> head(student_data)
# A tibble: 6 × 8
student_id location age sql_marks excel_marks python_marks power.bi_marks english_marks
  <dbl> <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1        4 Sydney     24      95       99      87      82      75
2        5 Tokyo      24      99       95      89      86      82
3        6 Berlin     22      72       70      99      79      77
4        7 London     23      97       90      74      72      85
5        8 Tokyo      22      91       71      79      80      75
6        9 Toronto    20      93       88      75      93      72
>
> # Linear Regression Model
> # Predicting python_marks using sql_marks and excel_marks
> model <- lm(python_marks ~ sql_marks + excel_marks, data = student_data)
>
> # Display model summary
> summary(model)

Call:
```

SHETH L.U.J. AND SIR M.V. COLLEGE OF ARTS SCIENCE AND COMMERCE

SUBJECT : R Programming

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

Source

Console Terminal Background Jobs

```
R > # Linear Regression Model
> # Predicting python_marks using sql_marks and excel_marks
> model <- lm(python_marks ~ sql_marks + excel_marks, data = student_data)
>
> # display model summary
> summary(model)

Call:
lm(formula = python_marks ~ sql_marks + excel_marks, data = student_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.8023 -8.1277  0.4503  8.3570 15.1172 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 86.69698   5.55605 15.604 <2e-16 ***
sql_marks   -0.02681   0.04567 -0.587   0.557    
excel_marks  0.01126   0.04548  0.248   0.805    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.893 on 494 degrees of freedom
Multiple R-squared:  0.0008347, Adjusted R-squared:  -0.003211 
F-statistic: 0.2063 on 2 and 494 DF,  p-value: 0.8136

> # scatter plot with regression line (SQL vs Python)
> ggplot(student_data, aes(x = sql_marks, y = python_marks)) +
+   geom_point() +
+   geom_smooth(method = "lm", se = FALSE) +
+   labs(
+     title = "Linear Regression: SQL Marks vs Python Marks",
+     x = "SQL Marks",
+     y = "Python Marks"
+   )
`geom_smooth()` using formula = 'y ~ x'
> |
```

26°C Sunny 11:56 19-01-2026

