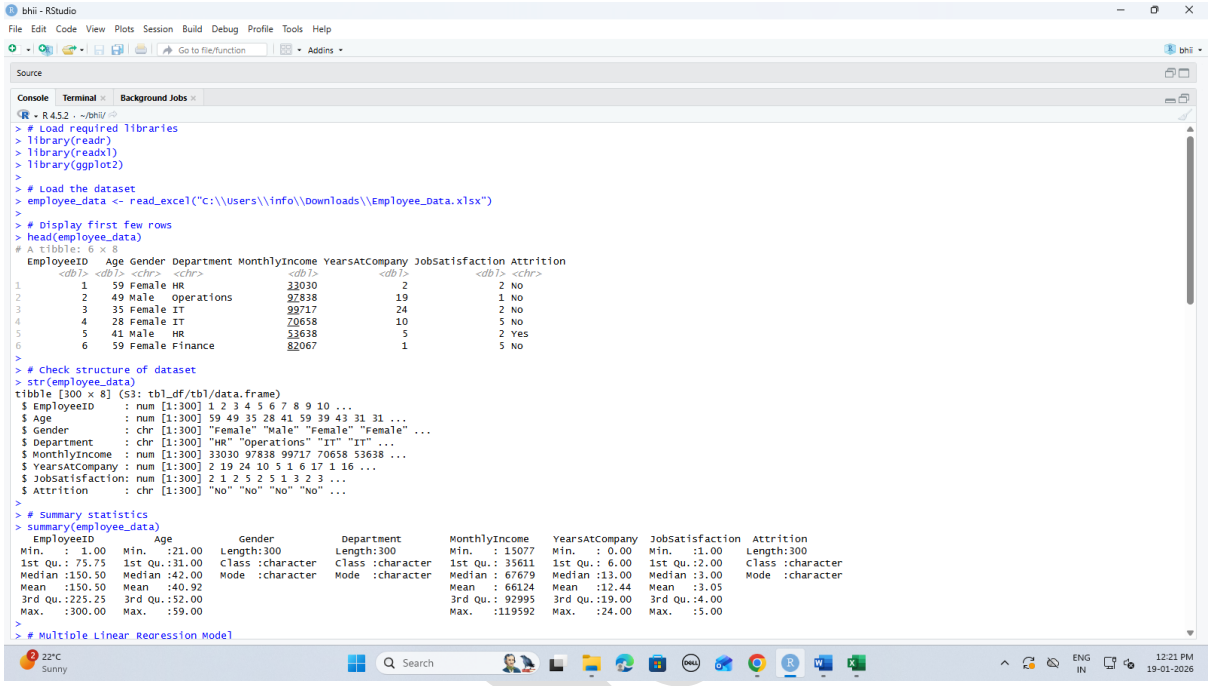
SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R

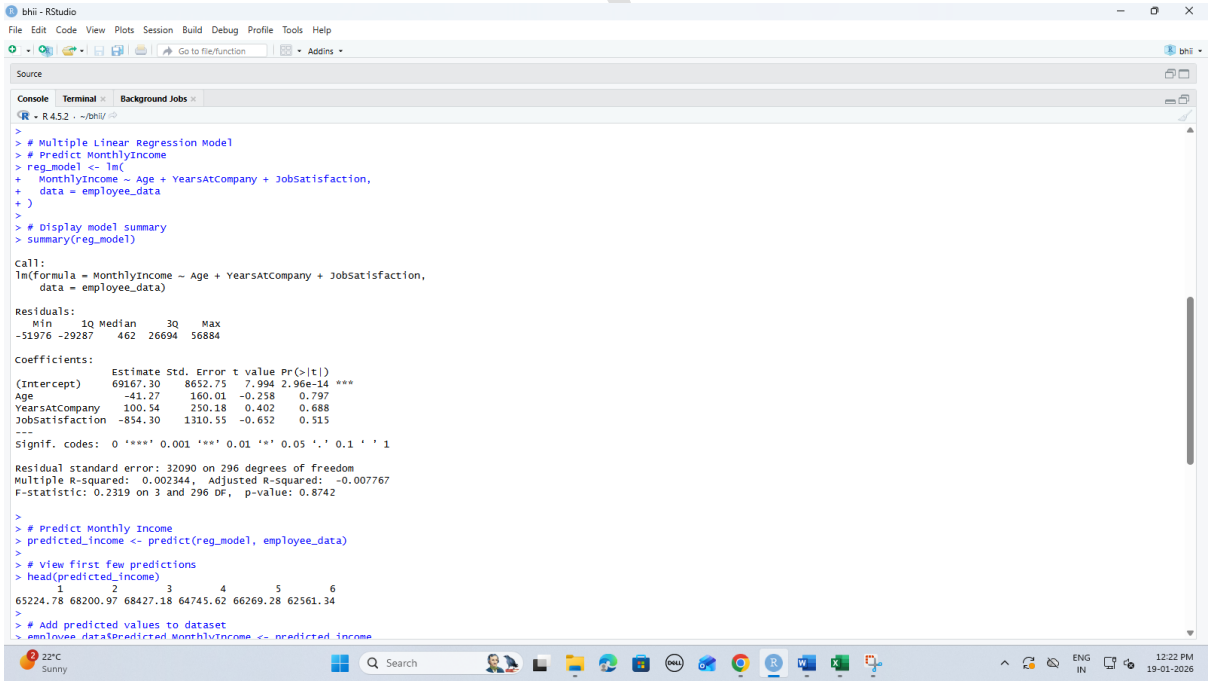
MODULE 2 – PRACTICAL 13

AIM:- Performing linear regression analysis using lm() (R).

OUTPUT:-



```
> # Load required libraries
> library(readr)
> library(readxl)
> library(ggplot2)
>
> # Load the dataset
> employee_data <- read_excel("C:\\Users\\info\\Downloads\\Employee_Data.xlsx")
>
> # Display first few rows
> head(employee_data)
# A tibble: 6 × 8
  EmployeeID   Age Gender Department MonthlyIncome YearsAtCompany JobSatisfaction Attrition
       <dbl> <dbl> <chr>  <chr>              <dbl>          <dbl>           <dbl> <chr>
1          1    59 Female HR                 32030              2               2 No
2          2    49 Male   Operations         97838             19               1 No
3          3    35 Female IT                 99717             24               2 No
4          4    28 Female IT                 70658             10               5 No
5          5    41 Male   HR                 53638              5               2 Yes
6          6    59 Female Finance            82067              1               5 No
>
> # Check structure of dataset
> str(employee_data)
tibble [300 × 8] (S3: tbl_df/tbl/data.frame)
 $ EmployeeID     : num [1:300] 1 2 3 4 5 6 7 8 9 10 ...
 $ Age            : num [1:300] 59 49 35 28 41 59 39 43 31 31 ...
 $ Gender         : chr [1:300] "Female" "Male" "Female" "Female" ...
 $ Department     : chr [1:300] "HR" "Operations" "IT" "IT" ...
 $ MonthlyIncome  : num [1:300] 33030 97838 99717 70658 53638 ...
 $ YearsAtCompany : num [1:300] 2 19 24 10 5 1 6 17 1 16 ...
 $ JobSatisfaction: num [1:300] 2 1 2 5 2 5 1 3 2 3 ...
 $ Attrition      : chr [1:300] "No" "No" "No" "No" ...
>
> # Summary statistics
> summary(employee_data)
   EmployeeID         Age           Gender          Department        MonthlyIncome    YearsAtCompany  JobSatisfaction  Attrition
 Min.   :  1.00   Min.   :21.00   Length:300         Length:300         Min.   : 15077   Min.   : 0.00   Min.   :1.00    Length:300
 1st Qu.: 75.75   1st Qu.:31.00   Class :character   Class :character   1st Qu.: 35611   1st Qu.: 6.00   1st Qu.:2.00    Class :character
 Median :150.50   Median :42.00   Mode  :character   Mode  :character   Median : 67679   Median :13.00   Median :3.00    Mode  :character
 Mean   :150.50   Mean   :40.92                                         Mean   : 66124   Mean   :12.44   Mean   :3.05
 3rd Qu.:225.25   3rd Qu.:52.00                                         3rd Qu.: 92995   3rd Qu.:19.00   3rd Qu.:4.00
 Max.   :300.00   Max.   :59.00                                         Max.   :119592   Max.   :24.00   Max.   :5.00
>
> # Multiple Linear Regression Model
```



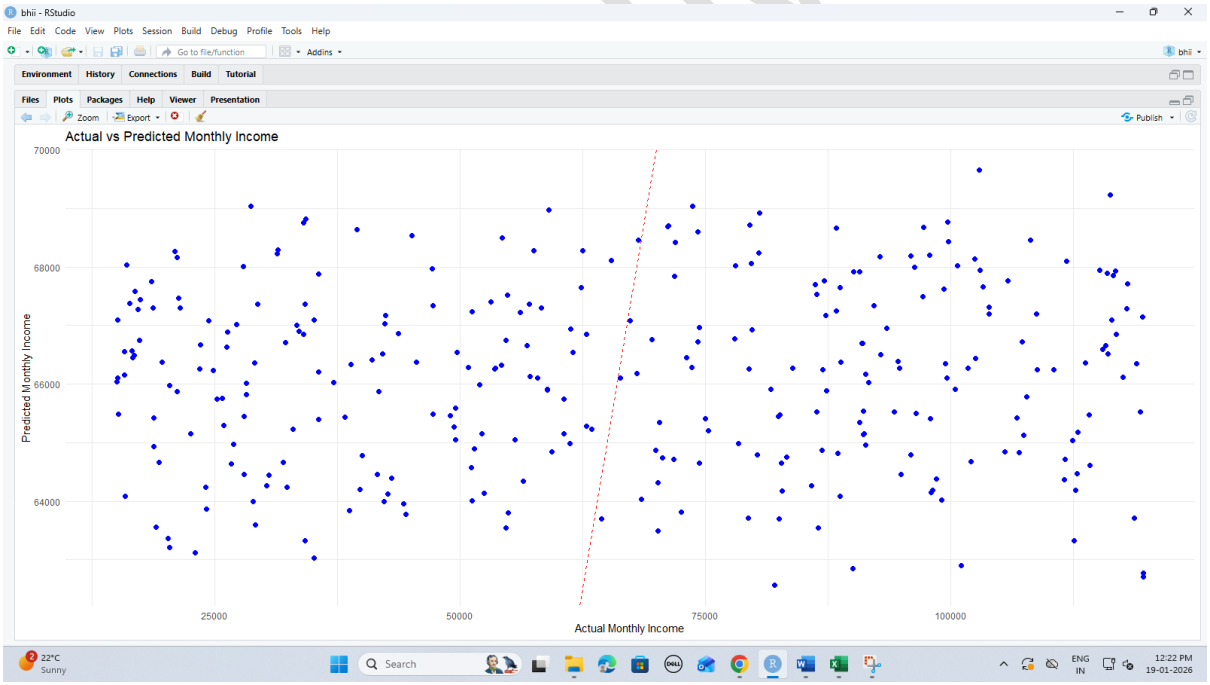
```
>
> # Multiple Linear Regression Model
> # Predict MonthlyIncome
> reg_model <- lm(
+   MonthlyIncome ~ Age + YearsAtCompany + JobSatisfaction,
+   data = employee_data
+ )
>
> # Display model summary
> summary(reg_model)

Call:
lm(formula = MonthlyIncome ~ Age + YearsAtCompany + JobSatisfaction,
    data = employee_data)

Residuals:
   Min     1Q Median     3Q    Max
-51976 -29287    462  26694  56884

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     69167.30    8652.75   7.994 2.96e-14 ***
Age              -41.27     160.01  -0.258    0.797
YearsAtCompany   100.54     250.18   0.402    0.688
JobSatisfaction -854.30    1310.55  -0.652    0.515
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 32090 on 296 degrees of freedom
Multiple R-squared:  0.002344,  Adjusted R-squared:  -0.007767
F-statistic: 0.2319 on 3 and 296 DF,  p-value: 0.8742

>
> # Predict Monthly Income
> predicted_income <- predict(reg_model, employee_data)
>
> # View first few predictions
> head(predicted_income)
       1        2        3        4        5        6
65224.78 68200.97 68427.18 64745.62 66269.28 62561.34
>
> # Add predicted values to dataset
> employee_data$Predicted_MonthlyIncome <- predicted_income
```

SUMEET JITENDRA YADAV
S124

SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R

```
> # Add predicted values to dataset
> employee_data$Predicted_MonthlyIncome <- predicted_income
>
> # View updated dataset
> head(employee_data)
# A tibble: 6 × 9
  EmployeeID   Age Gender Department MonthlyIncome YearsAtCompany JobSatisfaction Attrition Predicted_MonthlyIncome
       <dbl> <dbl> <chr>  <chr>            <dbl>          <dbl>           <dbl> <chr>                        <dbl>
1          1    59 Female HR               33030              2               2 No                           65225.
2          2    49 Male   Operations       97838             19               1 No                           68201.
3          3    35 Female IT               99717             24               2 No                           68427.
4          4    28 Female IT               70658             10               5 No                           64746.
5          5    41 Male   HR               53638              5               2 Yes                          66769.
6          6    59 Female Finance          82067              1               5 No                           62561.
>
> # Residuals & Error
> # Calculate residuals
> residuals_values <- residuals(reg_model)
>
> # View first few residuals
> head(residuals_values)
        1         2         3         4         5         6
-32194.776  29637.028  31289.825   5912.381 -12631.281  19505.657
>
> # Mean Squared Error
> mse <- mean(residuals_values^2)
>
> # Print MSE
> print(mse)
[1] 1015775719
>
> # Actual vs Predicted Regression Plot
> ggplot(employee_data, aes(x = MonthlyIncome, y = Predicted_MonthlyIncome)) +
+   geom_point(color = "blue", size = 2) +
+   geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
+   labs(
+     title = "Actual vs Predicted Monthly Income",
+     x = "Actual Monthly Income",
+     y = "Predicted Monthly Income"
+   ) +
+   theme_minimal()
>
```



SUMEET JITENDRA YADAV

S124