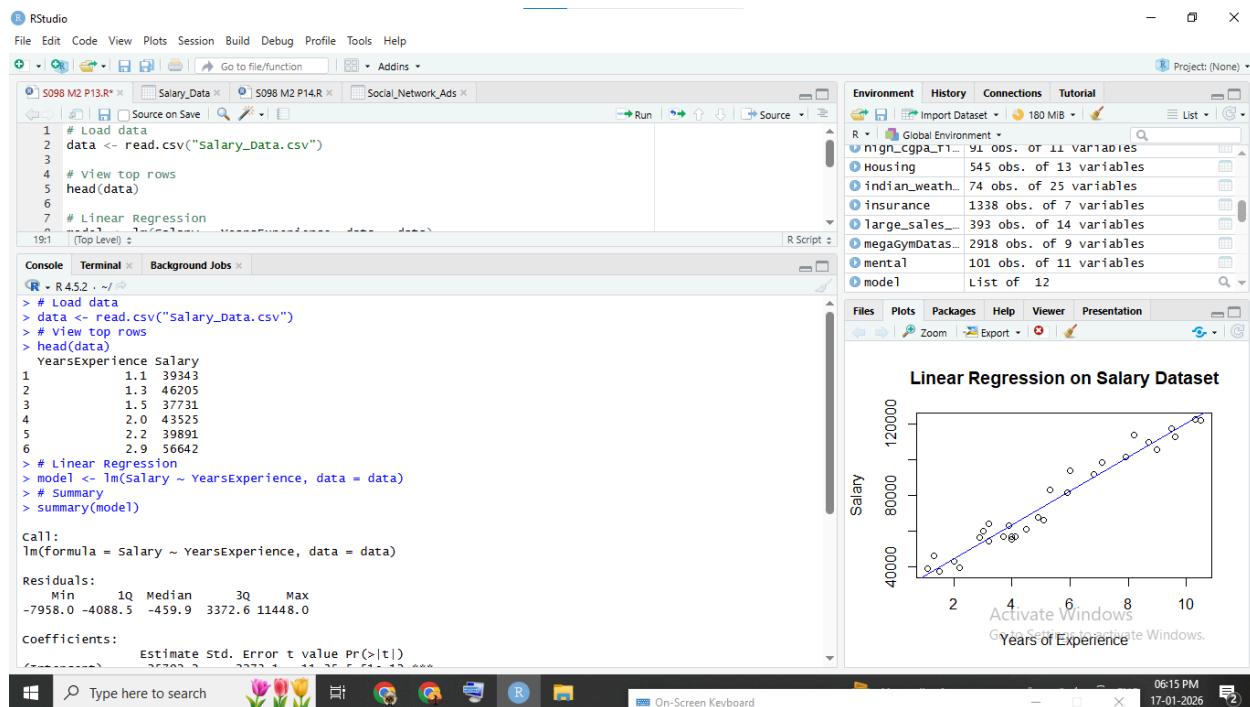


MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

Aim:- 13.Performing linear regression analysis using lm() (R).

Output:-



The screenshot shows the RStudio interface with the following details:

- File Menu:** File Edit Code View Plots Session Build Debug Profile Tools Help
- Console:**

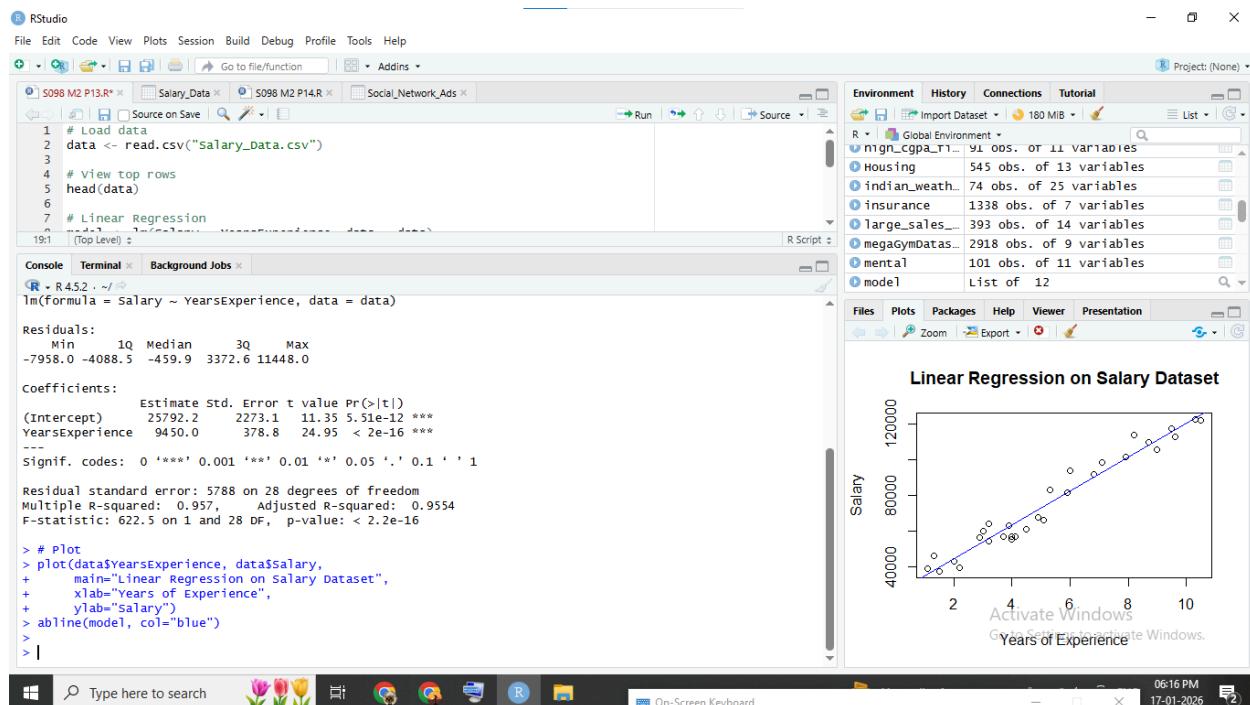
```
R > # Load data
> data <- read.csv("salary_data.csv")
> # View top rows
> head(data)
> # Linear Regression
> model <- lm(Salary ~ YearsExperience, data = data)
> # Summary
> summary(model)

Call:
lm(formula = salary ~ YearsExperience, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-7958.0 -4088.5 -459.9  3372.6 11448.0 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 25792.2   2273.1   11.35 5.51e-12 ***
YearsExperience 9450.0    378.8   24.95 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5788 on 28 degrees of freedom
Multiple R-squared:  0.957, Adjusted R-squared:  0.9554 
F-statistic: 622.5 on 1 and 28 DF, p-value: < 2.2e-16
```
- Environment:** Shows various datasets loaded: sign_cgpa_T1, Housing, indian_weath, insurance, large_sales, megagymDatas, mental, and model.
- Plots:** A scatter plot titled "Linear Regression on Salary Dataset" showing Salary (Y-axis, 40000 to 120000) versus Years of Experience (X-axis, 2 to 10). The data points show a strong positive linear trend, and a blue regression line is drawn through them.
- System Taskbar:** Shows the Windows taskbar with the date (17-01-2026) and time (06:15 PM).



The screenshot shows the RStudio interface with the following details:

- File Menu:** File Edit Code View Plots Session Build Debug Profile Tools Help
- Console:**

```
R > # Load data
> data <- read.csv("salary_data.csv")
> # View top rows
> head(data)
> # Linear Regression
> lm(formula = Salary ~ YearsExperience, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-7958.0 -4088.5 -459.9  3372.6 11448.0 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 25792.2   2273.1   11.35 5.51e-12 ***
YearsExperience 9450.0    378.8   24.95 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5788 on 28 degrees of freedom
Multiple R-squared:  0.957, Adjusted R-squared:  0.9554 
F-statistic: 622.5 on 1 and 28 DF, p-value: < 2.2e-16

> # Plot
> plot(data$YearsExperience, data$Salary,
+       main="Linear Regression on Salary Dataset",
+       xlab="Years of Experience",
+       ylab="Salary")
> abline(model, col="blue")
> |
```
- Environment:** Shows various datasets loaded: sign_cgpa_T1, Housing, indian_weath, insurance, large_sales, megagymDatas, mental, and model.
- Plots:** A scatter plot titled "Linear Regression on Salary Dataset" showing Salary (Y-axis, 40000 to 120000) versus Years of Experience (X-axis, 2 to 10). The data points show a strong positive linear trend, and a blue regression line is drawn through them.
- System Taskbar:** Shows the Windows taskbar with the date (17-01-2026) and time (06:16 PM).

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

14 Performing logistic regression using `glm()` (R).

Output:-

The screenshot shows the RStudio interface with the following details:

- Console Output:**

```

> # Load dataset
> ads <- read.csv("Social_Network_Ads.csv")
> # Take first 10 observations
> ads10 <- ads[1:10, ]
> # Create binary dependent variable
> # Purchased already exists (0 = No, 1 = Yes)
> ads10$Purchased <- as.numeric(ads10$Purchased)
> # Logistic Regression Model
> model_glm <- glm(Purchased ~ Age,
+                      data = ads10,
+                      family = binomial)
> # Model summary
> summary(model_glm)

Call:
glm(formula = Purchased ~ Age, family = binomial, data = ads10)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.9691    8.4331 -1.064   0.288
Age          0.2291    0.2652  0.864   0.388

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6.5017 on 9 degrees of freedom
Residual deviance: 5.5102 on 8 degrees of freedom
AIC: 9.5102

Number of Fisher Scoring iterations: 6

```
- Global Environment:** Shows objects like ads10, anova_result, boston, CarPrice_Assignment, chi_result, cor_matrix, correlation_matrix, and data.
- Plots:** A scatter plot titled "Logistic Regression" showing "Purchased (0/1)" on the y-axis (0.0 to 0.8) versus "Age" on the x-axis (20 to 35). A red curve represents the fitted logistic regression model.
- System Status Bar:** Shows the date (19-01-2026), time (12:44 PM), and system status (ENG).

The screenshot shows the RStudio interface with the following details:

- Console Output:**

```

> # Predicted probabilities
> ads10prob <- predict(model_glm, type = "response")
> # Logistic Regression Graph
> plot(ads10$Age,
+       ads10$Purchased,
+       main = "Logistic Regression",
+       xlab = "Age",
+       ylab = "Purchased (0/1)",
+       pch = 19,
+       col = "blue")
> # Probability curve
> lines(sort(ads10$Age),
+        ads10prob[order(ads10$Age)],
+        col = "red",
+        lwd = 2)
>

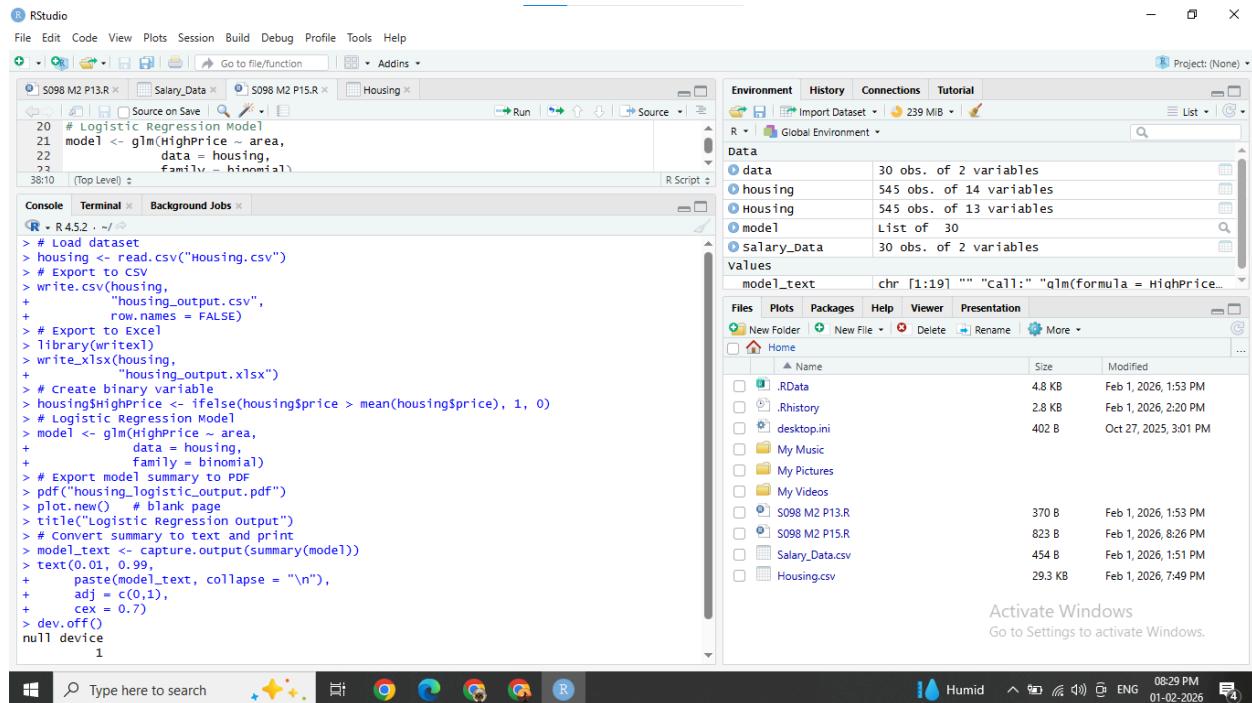
```
- Global Environment:** Shows objects like ads10, anova_result, boston, CarPrice_Assignment, chi_result, cor_matrix, correlation_matrix, and data.
- Plots:** A scatter plot titled "Logistic Regression" showing "Purchased (0/1)" on the y-axis (0.0 to 0.8) versus "Age" on the x-axis (20 to 35). A red curve represents the fitted logistic regression model.
- System Status Bar:** Shows the date (19-01-2026), time (12:45 PM), and system status (ENG).

MVLU COLLEGE

Subject:-Data Analysis with SAS / SPSS /R

15 Exporting results into external files (Excel, CSV, PDF) using write.csv() and writexl (R).

Output:-



```

# Logistic Regression Model
model <- glm(HighPrice ~ area,
              data = housing,
              family = binomial)

# Export to CSV
write.csv(housing,
          "housing_output.csv",
          row.names = FALSE)

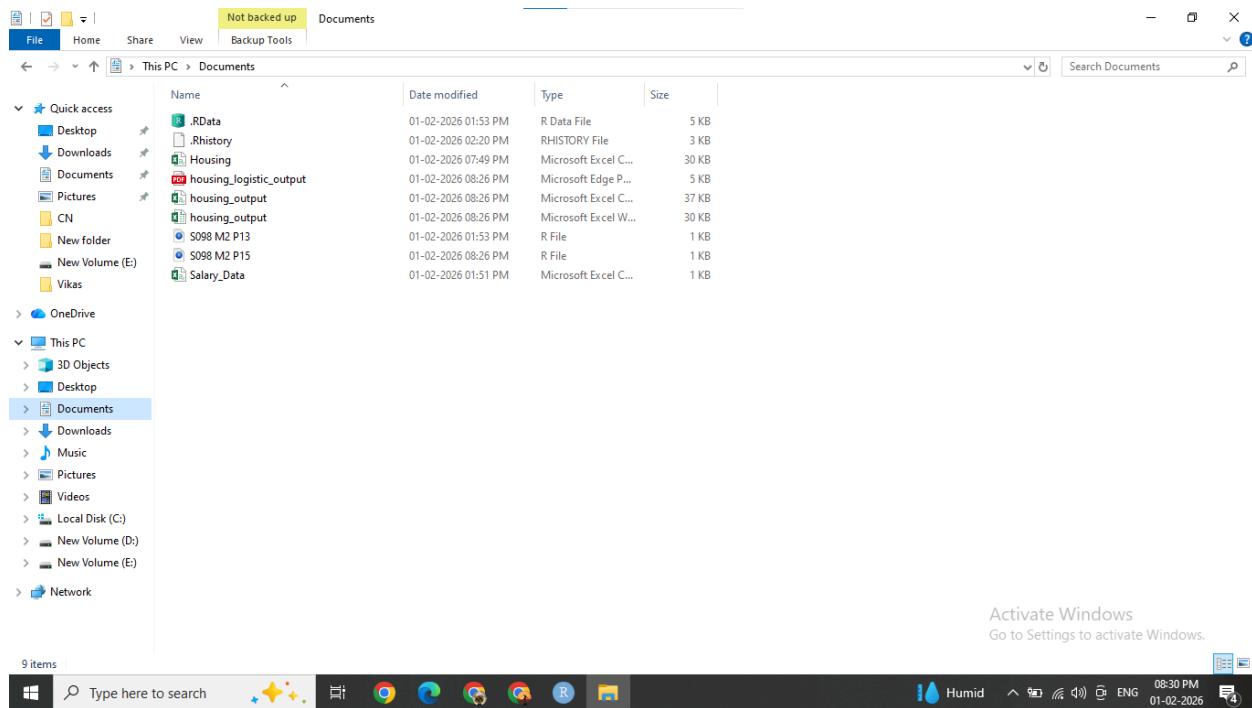
# Export to Excel
library(writexl)
write_xlsx(housing,
           "housing_output.xlsx")

# Create binary variable
housingHighPrice <- ifelse(housing$price > mean(housing$price), 1, 0)

# Logistic Regression Model
model <- glm(highprice ~ area,
             data = housing,
             family = binomial)

# Export model summary to PDF
pdf("housing_logistic_output.pdf")
plot.new() # blank page
title("Logistic Regression output")
# Convert summary to text and print
model_text <- capture.output(summary(model))
text(0.01, 0.99,
     paste(model_text, collapse = "\n"),
     adj = c(0,1),
     cex = 0.7)
dev.off()
null_device
1

```



Vikas Pal
S098