Date 01/03/2024 Expt. No. 5

Expt. Name Page No. 13

EXPERIMENT - 5

Aim:- Perform a comprehensive regression analysis using R, exploring various methods such as linear regression, multiple regression, & polynomial regression.

Introduction:- The aim of this experiment is to perform a comprehensive regression analysis using R, exploring various regression methods including linear regression, multiple regression and polynomial regression. The goal is to understand the application of regression analysis in predicting & modelling relationships between variables.

Software Required:-

1. R Statistical Software
2. R Studio

Relevance of the Experiment:- Regression analysis is a fundamental statistical technique used in various fields for predicting relationship b/w dependent & independent variables. Understanding diff. regression methods helps in selecting the most suitable model for a given dataset. This experiment is relevant for data analysis, statisticians & researchers who need to analyze & interpret relationships in their data.

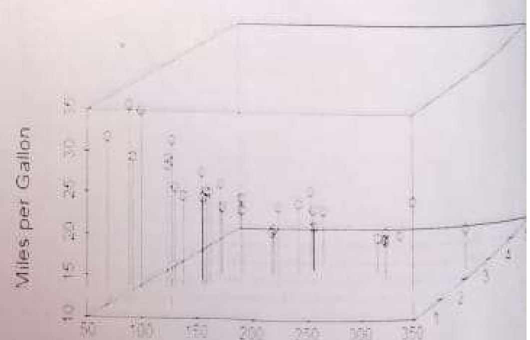
Description:- The experiment involves loading a dataset, exploring its structure & applying various regression techniques in R. The dataset may contain multiple variables, allowing for the exploration of relationships b/w them. Participants will learn how to perform linear regression, multiple regression & polynomial regression, interpret the results & assess the model's performance.

Teacher's Signature:

Aim:- Perform a comprehensive regression analysis using R, exploring various methods such as linear regression, multiple regression, & polynomial regression.

```
# Load the dataset
data(mtcars)

# View the structure of the dataset
str(mtcars)
'data.frame':	32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...

# View the first few rows of the dataset
head(mtcars)
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1

# Load the scatterplot3d package
library(scatterplot3d)
# Perform linear regression
# Example: Predicting mpg (miles per gallon) based on horsepower
linear_model <- lm(mpg ~ hp, data = mtcars)
summary(linear_model)

Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
hp          -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,	Adjusted R-squared:  0.5892
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07

>
> # Perform multiple regression
> # Example: Predicting mpg based on multiple variables
> multiple_model <- lm(mpg ~ hp + wt + cyl, data = mtcars)
> summary(multiple_model)

Call:
lm(formula = mpg ~ hp + wt + cyl, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9290 -1.5598 -0.5311  1.1850  5.8986

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 38.75179    1.78686  21.687  < 2e-16 ***
hp          -0.01804    0.01188  -1.519 0.140015
wt          -3.16697    0.74058  -4.276 0.000199 ***
cyl         -0.94162    0.55092  -1.709 0.098480 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.512 on 28 degrees of freedom
Multiple R-squared:  0.8431,	Adjusted R-squared:  0.8263
F-statistic: 50.17 on 3 and 28 DF,  p-value: 2.184e-11

# Perform polynomial regression
# Example: Predicting mpg based on a quadratic model with horsepower
polynomial_model <- lm(mpg ~ hp + I(hp^2), data = mtcars)
summary(polynomial_model)

Call:
lm(formula = mpg ~ hp + I(hp^2), data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5512 -1.6027 -0.6977  1.5509  8.7213

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.041e+01  2.741e+00  14.744 5.23e-15 ***
hp          -2.133e-01  3.488e-02  -6.115 1.16e-06 ***
I(hp^2)      4.208e-04  9.844e-05   4.275 0.000189 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.077 on 29 degrees of freedom
Multiple R-squared:  0.7561,	Adjusted R-squared:  0.7393
F-statistic: 44.95 on 2 and 29 DF,  p-value: 1.301e-09
```

Multiple Regression



Date
Expt. No.
Expt. Name
Page No. 14

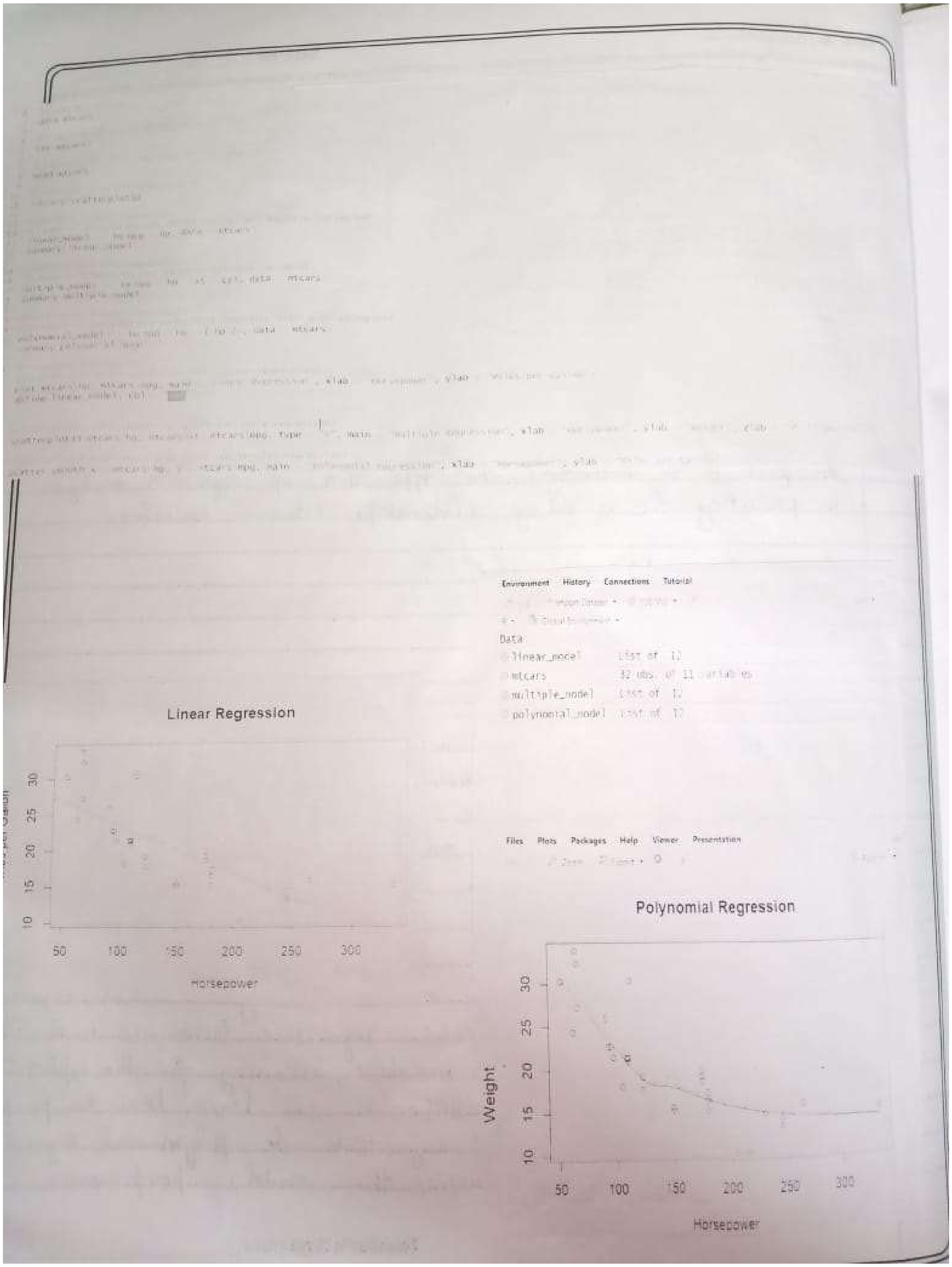
Flowchart / Pseudocode / Algorithm :-

1. Load the dataset into R.
2. Explore the dataset structure & characteristics.
3. Perform simple linear regression to analyze the relationship b/w one dependent and one independent variable.
4. Implement multiple regression to examine the impact of two or more independent variables on the dependent variable.
5. Apply polynomial regression to capture non-linear relationships b/w variables.
6. Evaluate the performance of each regression model using appropriate metrics.
7. Visualize the regression results using plots & charts.

Learning Outcome :-

1. Understanding the principles of linear regression, multiple regression and polynomial regression.
2. Proficiency in implementing regression analysis using R.
3. Interpretation of regression model output & evaluations of model performance.
4. Visualization skills for presenting regression results.
5. Gain insights into selecting appropriate regression models for different types of data.

Teacher's Signature:

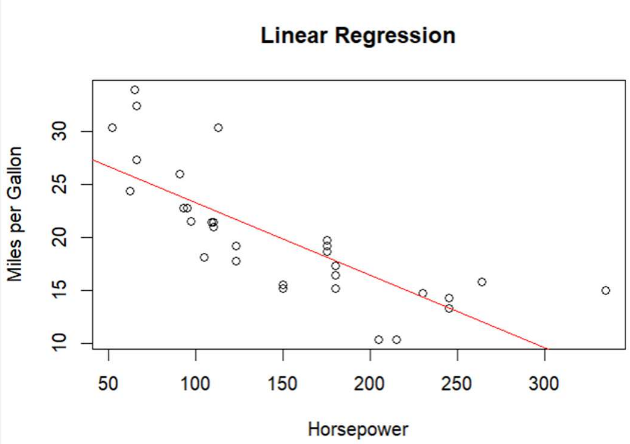


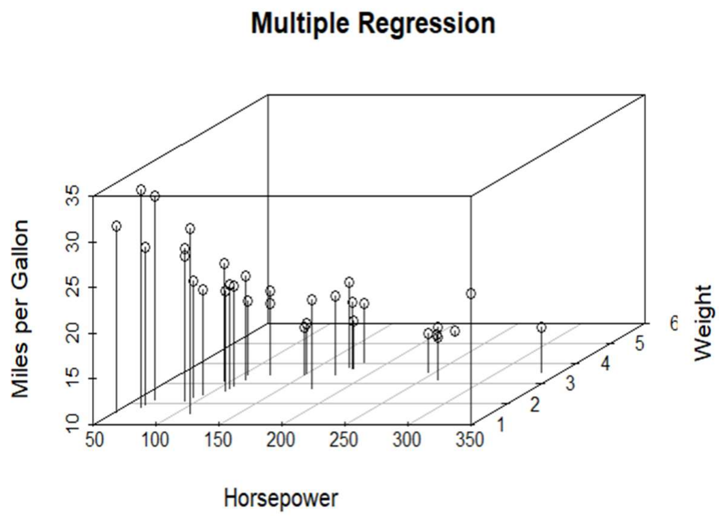
SCREENSHOTS:

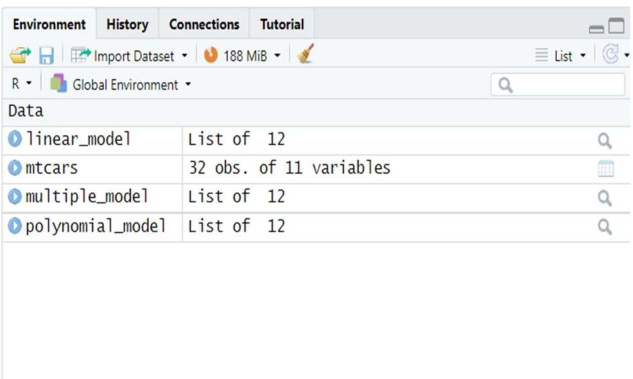
```r
# Load the dataset
data(mtcars)

# View the structure of the dataset
str(mtcars)

# Explore the first few rows of the dataset
head(mtcars)

# Load the scatterplot3d package
library(scatterplot3d)

# Perform linear regression
# Example: Predicting mpg (miles per gallon) based on horsepower
linear_model <- lm(mpg ~ hp, data = mtcars)
summary(linear_model)

# Perform multiple regression
# Example: Predicting mpg based on multiple variables
multiple_model <- lm(mpg ~ hp + wt + cyl, data = mtcars)
summary(multiple_model)

# Perform polynomial regression
# Example: Predicting mpg based on a quadratic model with horsepower
polynomial_model <- lm(mpg ~ hp + I(hp^2), data = mtcars)
summary(polynomial_model)

# Visualize the results
# Scatter plot with regression line (linear regression)
plot(mtcars$hp, mtcars$mpg, main = "Linear Regression", xlab = "Horsepower", ylab = "Miles per Gallon")
abline(linear_model, col = "red")


# Scatter plot with regression plane (multiple regression)
scatterplot3d(mtcars$hp, mtcars$wt, mtcars$mpg, type = "h", main = "Multiple Regression", xlab = "Horsepower", ylab = "Weight", zlab = "Miles per Ga

# Scatter plot with regression curve (polynomial regression)
scatter.smooth(x = mtcars$hp, y = mtcars$mpg, main = "Polynomial Regression", xlab = "Horsepower", ylab = "Miles per Gallon")
```

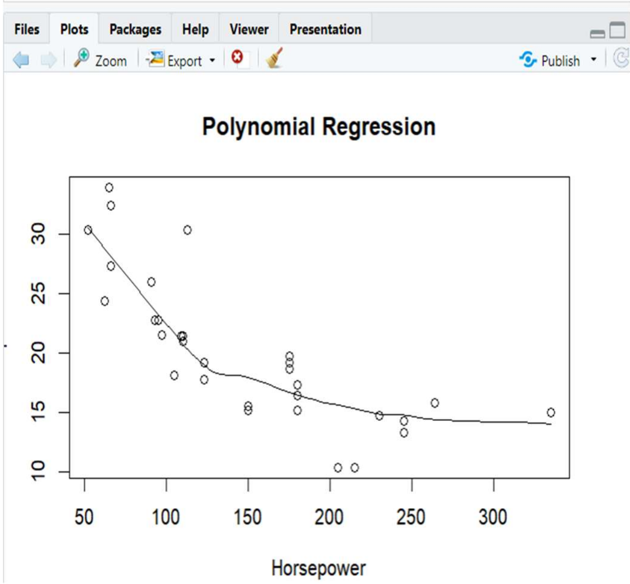
```
> # Load the dataset
> data(mtcars)
>
> # View the structure of the dataset
> str(mtcars)
'data.frame':	32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
>
> # Explore the first few rows of the dataset
> head(mtcars)
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
>
> # Load the scatterplot3d package
> library(scatterplot3d)
> # Perform linear regression
> # Example: Predicting mpg (miles per gallon) based on horsepower
> linear_model <- lm(mpg ~ hp, data = mtcars)
> summary(linear_model)

Call:
lm(formula = mpg ~ hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7121 -2.1122 -0.8854  1.5819  8.2360
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
hp          -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,	Adjusted R-squared:  0.5892 
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07

> 
> # Perform multiple regression
> # Example: Predicting mpg based on multiple variables
> multiple_model <- lm(mpg ~ hp + wt + cyl, data = mtcars)
> summary(multiple_model)

Call:
lm(formula = mpg ~ hp + wt + cyl, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9290 -1.5598 -0.5311  1.1850  5.8986 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 38.75179    1.78686  21.687  < 2e-16 ***
hp          -0.01804    0.01188  -1.519 0.140015    
wt          -3.16697    0.74058  -4.276 0.000199 ***
cyl         -0.94162    0.55092  -1.709 0.098480 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.512 on 28 degrees of freedom
Multiple R-squared:  0.8431,	Adjusted R-squared:  0.8263 
F-statistic: 50.17 on 3 and 28 DF,  p-value: 2.184e-11

> # Perform polynomial regression
> # Example: Predicting mpg based on a quadratic model with horsepower
> polynomial_model <- lm(mpg ~ hp + I(hp^2), data = mtcars)
> summary(polynomial_model)

Call:
lm(formula = mpg ~ hp + I(hp^2), data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5512 -1.6027 -0.6977  1.5509  8.7213 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.041e+01  2.741e+00  14.744 5.23e-15 ***
hp          -2.133e-01  3.488e-02  -6.115 1.16e-06 ***
I(hp^2)      4.208e-04  9.844e-05   4.275 0.000189 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.077 on 29 degrees of freedom
Multiple R-squared:  0.7561,	Adjusted R-squared:  0.7393 
F-statistic: 44.95 on 2 and 29 DF,  p-value: 1.301e-09
```




Multiple Regression
Miles per Gallon
10
15
20
25
30
35
50
100
150
200
250
300
350
1
2
3
4
5
6
Weight
Horsepower


Environment
History
Connections
Tutorial
Import Dataset
188 MiB
List
R
Global Environment
Data
linear_model
List of 12
mtcars
32 obs. of 11 variables
multiple_model
List of 12
polynomial_model
List of 12


Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
Polynomial Regression
10
15
20
25
30
50
100
150
200
250
300
Horsepower