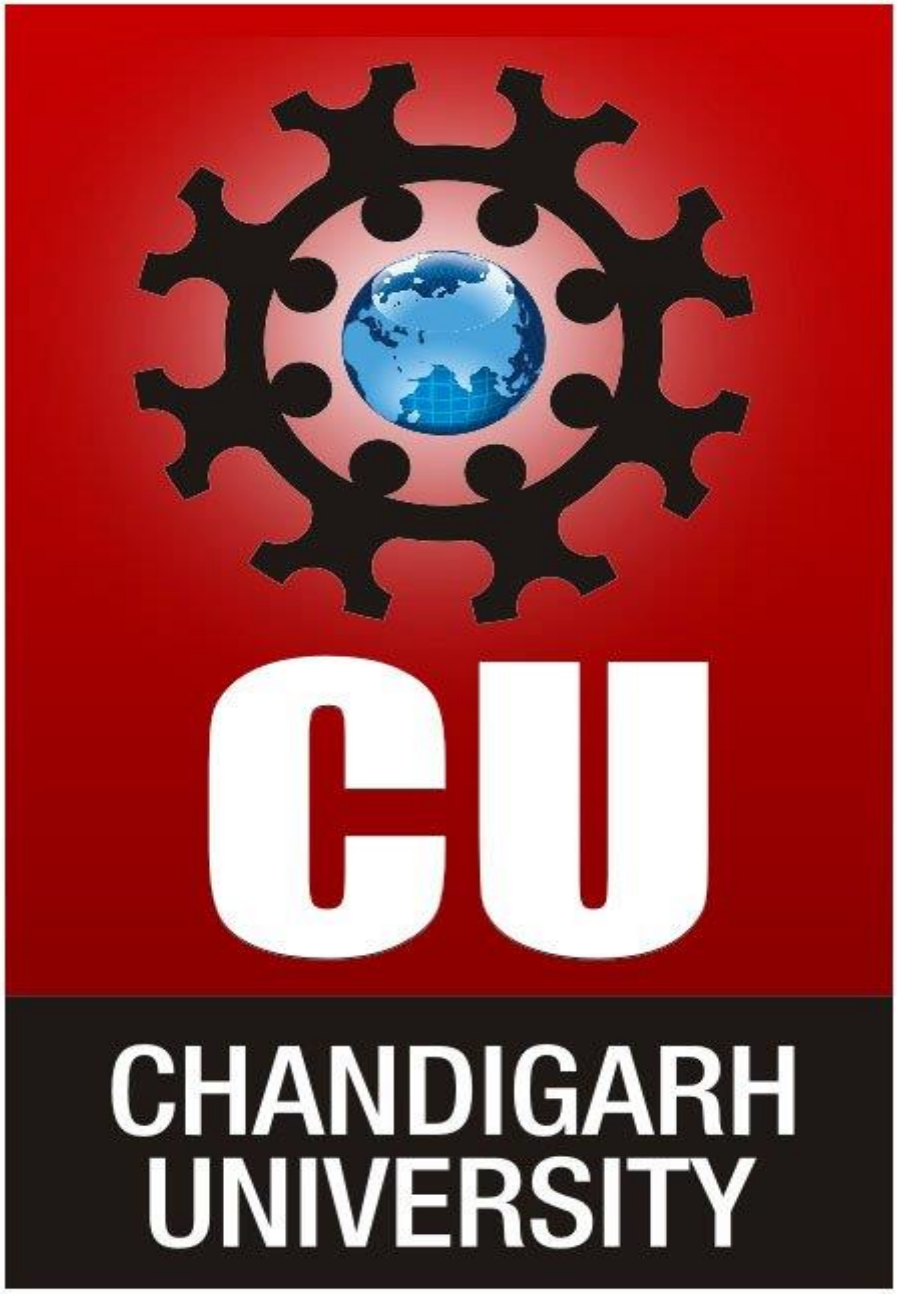
CHANDIGARH UNIVERSITY, GHAURAN MOHALI



SUBJECT NAME- PROGRAMMING IN R

Subject code-24CAP-614

Submitted By:
Tanisha Jain

UID- 24MCI10047

Name- Tanisha Jain
UID – 24MCI10047
Branch - MCA(AI&ML)
Date of Submission- 25 October 2024

Subject Name- Programming in R
Semester- 1st

Subject Code-24CAH-614
Section/Group- 1/B

Project-1

Project Name:

Collect a dataset where you have a dependent variable and several independent variables. Perform linear regression, polynomial regression, or other types of regression analysis using R to predict the dependent variable based on the independent variables. Evaluate the model's performance and interpret the results.

AIM:

The aim of this project is to **predict the miles per gallon (mpg)** of various cars using multiple independent variables such as horsepower (hp), weight (wt), and displacement (disp) from the mtcars dataset. We will assess the performance of both linear regression and polynomial regression models to determine which provides a better fit for the data and offers more accurate predictions.

Methodology:

 Data Collection: Utilize the mtcars dataset, which contains information about different car models, including their mpg, horsepower, weight, and displacement.

 Data Preparation:

- Explore and preprocess the dataset to understand the relationships between variables.
- Identify the dependent variable (mpg) and independent variables (hp, wt, disp).

 Model Development:

- Fit a linear regression model to predict mpg based on hp, wt, and disp.
- Fit a polynomial regression model to include nonlinear relationships by incorporating higher-order terms of the independent variables.

 Model Evaluation:

- Calculate and compare the R-squared values for both models to assess their explanatory.

Task to be done

 Data **Preparation**:

- Load the mtcars dataset.
- Explore the dataset to understand its structure and summary statistics.
- Identify the dependent variable (mpg) and independent variables (hp, wt, disp).

 Model Development:

- **Task 1**: Fit a linear regression model:
  - Use lm() function in R to create a linear regression model with mpg as the dependent variable and hp, wt, and disp as independent variables.
  - Obtain and interpret the summary of the model.
  -
  - **Task 2**: Fit a polynomial regression model:
  - Create a polynomial regression model by adding polynomial terms for hp, wt, and disp using poly() function.
  - Obtain and interpret the summary of this polynomial regression model.
  - **Model Evaluation**:
  - **Task 3**: Compare the performance of both models:
  - Calculate the R-squared value for both the linear and polynomial models.
  - Create residual plots for both models to assess model accuracy and check for assumptions.
  - Results **Interpretation**:
  - **Task 4**: **Analyse** the results:
  - Interpret the coefficients from both models to understand the relationship between the independent variables and mpg.
  - Discuss the advantages and disadvantages of each model based on R-squared values and residual analysis.

Coding/process

1. **Loading the Dataset**

 # Load the mtcars dataset

data(mtcars)

 # Explore the dataset

head(mtcars)

summary(mtcars)

2. **Fit a linear regression model**

 # Linear regression: Predict mpg based on hp, wt, and disp

linear_model <- lm(mpg ~ hp + wt + disp, data = mtcars)

 # Summary of the linear regression model

summary(linear_model)

3. **Fit a polymonial regression**

 # Polynomial regression: Adding quadratic terms of hp, wt, and disp

poly_model <- lm(mpg ~ poly(hp, 2) + poly(wt, 2) + poly(disp, 2), data = mtcars)

 # Summary of the polynomial regression model

summary(poly_model)

4. **Evaluate the model**

 # Check the R-squared values for both models
cat("R-squared of Linear Model: ", summary(linear_model)$r.squared, "\n")
cat("R-squared of Polynomial Model: ", summary(poly_model)$r.squared, "\n")

 # Residual plots for both models
par(mfrow = c(2, 2))

plot(linear_model, which = 1:4) # Residual plots for linear model
plot(poly_model, which = 1:4) # Residual plots for polynomial model

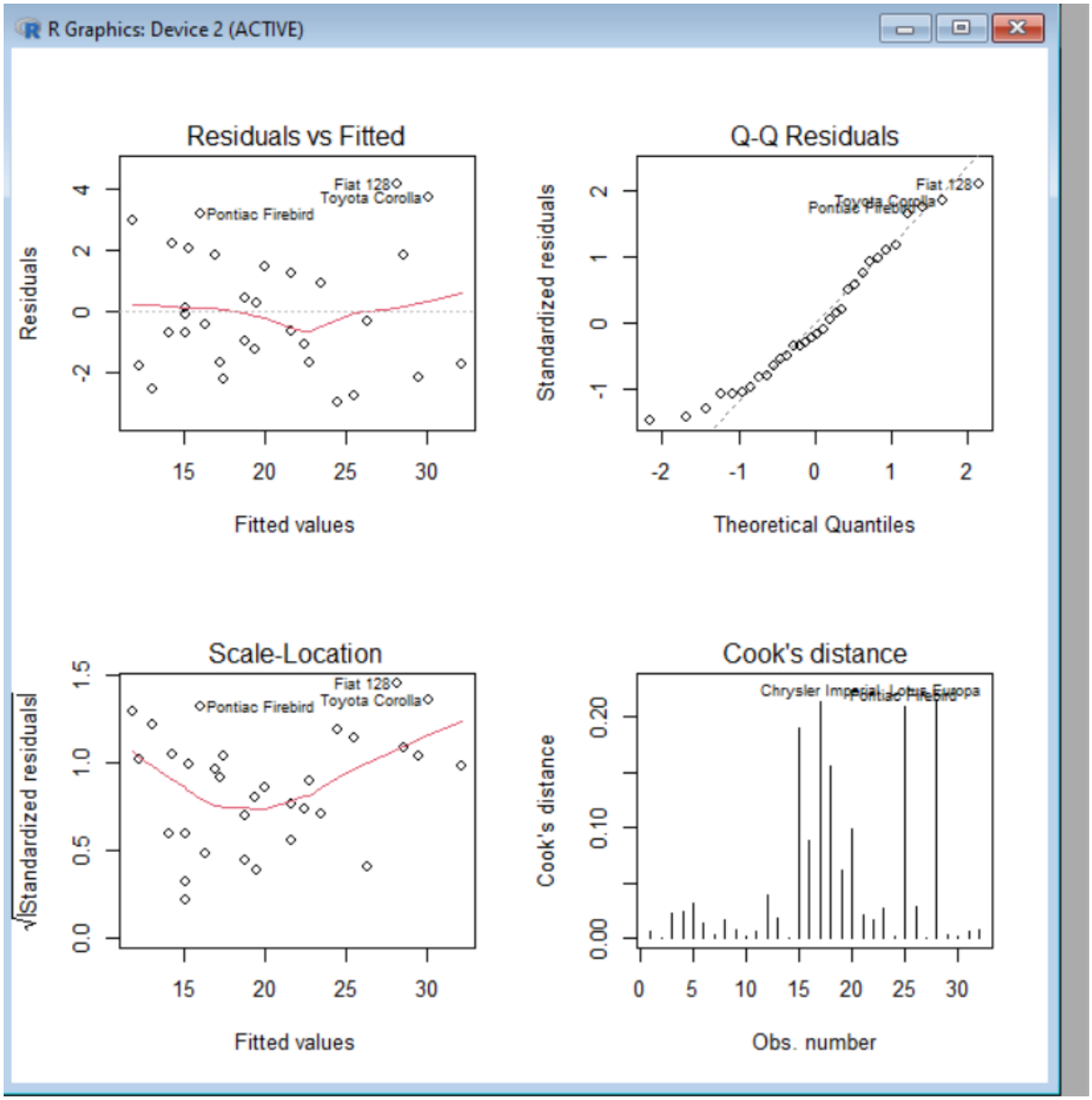
Screenshots

```
> # Load the mtcars dataset
> data(mtcars)
>
> # Explore 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
> summary(mtcars)
      mpg             cyl             disp             hp
 Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
 Median :19.20   Median :6.000   Median :196.3   Median :123.0
 Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
 Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
      drat             wt             qsec             vs
 Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
 Median :3.695   Median :3.325   Median :17.71   Median :0.0000
 Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
 Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
       am              gear            carb
 Min.   :0.0000   Min.   :3.000   Min.   :1.000
 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
 Median :0.0000   Median :4.000   Median :2.000
 Mean   :0.4062   Mean   :3.688   Mean   :2.812
 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
 Max.   :1.0000   Max.   :5.000   Max.   :8.000
> # Linear regression: Predict mpg based on hp, wt, and disp
> linear_model <- lm(mpg ~ hp + wt + disp, data = mtcars)
>
> # Summary of the linear regression model
> summary(linear_model)

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

Residuals:
   Min     1Q Median     3Q    Max
-3.891 -1.640 -0.172  1.061  5.861

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.105505   2.110815  17.579  < 2e-16 ***
hp          -0.031157   0.011436  -2.724  0.01097 *
wt          -3.800891   1.066191  -3.565  0.00133 **
disp        -0.000937   0.010350  -0.091  0.92851
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.639 on 28 degrees of freedom
Multiple R-squared:  0.8268,    Adjusted R-squared:  0.8083
F-statistic: 44.57 on 3 and 28 DF,  p-value: 8.65e-11
```

```
> # Polynomial regression: Adding quadratic terms of hp, wt, and disp
> poly_model <- lm(mpg ~ poly(hp, 2) + poly(wt, 2) + poly(disp, 2), data = mtcars)
>
> # Summary of the polynomial regression model
> summary(poly_model)

Call:
lm(formula = mpg ~ poly(hp, 2) + poly(wt, 2) + poly(disp, 2),
    data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0130 -1.7033 -0.3958  1.5250  4.1655

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     20.0906     0.3884  51.726  < 2e-16 ***
poly(hp, 2)1   -12.5898     4.4925  -2.802 0.009656 **
poly(hp, 2)2     5.1411     2.8273   1.818 0.081005 .
poly(wt, 2)1   -19.1181     4.8764  -3.921 0.000607 ***
poly(wt, 2)2     3.4661     4.6467   0.746 0.462663
poly(disp, 2)1   0.2439     6.4565   0.038 0.970167
poly(disp, 2)2   3.2988     4.8744   0.677 0.504777
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.197 on 25 degrees of freedom
Multiple R-squared:  0.8928,    Adjusted R-squared:  0.8671
F-statistic: 34.71 on 6 and 25 DF,  p-value: 5.972e-11
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.197 on 25 degrees of freedom
Multiple R-squared:  0.8928,    Adjusted R-squared:  0.8671
F-statistic: 34.71 on 6 and 25 DF,  p-value: 5.972e-11

> # Check the R-squared values for both models
> cat("R-squared of Linear Model: ", summary(linear_model)$r.squared, "\n")
R-squared of Linear Model:  0.8268361
> cat("R-squared of Polynomial Model: ", summary(poly_model)$r.squared, "\n")
R-squared of Polynomial Model:  0.8928246
>
> # Residual plots for both models
> par(mfrow = c(2, 2))
> plot(linear_model, which = 1:4)  # Residual plots for linear model
> plot(poly_model, which = 1:4)    # Residual plots for polynomial model
>
```

Output



Learning Outcomes:

1. Understanding regression analysis
2. Model development
3. Communication of R results
4. Practical R skills