Project Name :
Mobile Price Prediction using Multiple Regression

Sir Parashurambhau College

Subject : Statistics

Project Name : Mobile Price Prediction using Multiple Regression

Teacher Name: Dr. Susmita Bhambure Mam

Student Name :

1. Shreyash Vivek Atre(SP04240042)
2. Aditya Sambhaji Dhas(SP04240033)
3. Prakash Mahadev Ovhal(SP04240036)
4. Chinmay Vilas Pugaonkar(SP04240043)
5. Owais Fayaz Malik(SP04240053)
6. Ritesh Deepak Waghmode(SP04240031)

➢ Table of Contents:

❑ Introduction:

The increasing variety of smartphones makes pricing analysis essential for consumers and manufacturers. This project builds a **regression model** to predict mobile prices using **RAM and Battery Capacity** as key factors.

❖ Objective:

- Analyze how **RAM & Battery** impact pricing.
- Develop a **predictive model** using **R programming**.
- Evaluate performance with **R^2, Adjusted R^2, and Error Metrics**.
- Visualize insights through **plots & residual**

❑ Data Source:

our dataset is sourced from **Kaggle**, specifically from the dataset **"Mobile Prices 2023"**. It contains information on various smartphone specifications, including **RAM, Battery, Processor, Ratings, and Price**.

For this project, we focus on **RAM and Battery Capacity** as independent variables, with **Price** as the dependent variable. The dataset enables us to analyze pricing patterns and build a **multiple regression model** to predict mobile phone prices based on these key specifications.

Link: Mobile Prices 2023

❑ Statistical Techniques:

This project applies **Multiple Linear Regression (MLR)** to analyze the relationship between mobile phone prices and key specifications like **RAM** and **Battery Capacity**.

- **Techniques Implemented:**

1. **Multiple Regression Model**: Predicts price based on independent variables.
2. **R^2 & Adjusted R^2**: Measures model accuracy and explanatory power.
3. **Residual Analysis**: Evaluates model fit by examining prediction errors.
4. **Correlation Coefficients**: Determines relationships between variables.
5. **Error Metrics (SSE, MRSS, Residual SD)**: Assesses model performance. These statistical techniques help in understanding pricing trends and improving prediction accuracy.

❑ Explanation:

- This project applies **Multiple Linear Regression (MLR)** to predict mobile phone prices based on **RAM and Battery Capacity**. The dataset, sourced from **Kaggle**, contains various smartphone specifications. We extract **RAM and Battery** as independent variables and **Price** as the dependent variable.
- The model is built using **R programming**, where we fit a regression equation:

$$Price=a+b(RAM)+c(Battery)$$

- The model's effectiveness is evaluated using **R^2, Adjusted R^2, Sum of Squared Errors (SSE), and Residual Standard Deviation**. Correlation analysis helps determine the strength of relationships between variables.
- Data visualization, including **scatter plots and residual analysis**, ensures model accuracy. If **R^2 is low**, it suggests additional factors like **Processor, Brand, and User Ratings** influence pricing.

❑ Input :

The input for this project consists of a **dataset sourced from Kaggle** (Mobile Prices 2023), containing various smartphone specifications.

- **Key Inputs Used:**

1. **Price (Dependent Variable):** The target value to be predicted.
2. **RAM (Independent Variable):** Amount of Random Access Memory in GB.
3. **Battery Capacity (Independent Variable):** Battery size in mAh.

- **Data Preprocessing:**

1. The dataset is **imported** into R using the readxl package.
2. Missing values and inconsistencies are checked.
3. Relevant columns (**RAM, Battery, Price**) are extracted for model training.
4. These inputs are then used in a **Multiple Linear Regression Model** to analyze and predict mobile phone prices based on specifications.

❑ R Code :

```
library(readxl)

# Load the dataset
data = read_excel("C:/Users/shreyash/OneDrive/Desktop/Projects/mobile prices/updated_mobile_prices - Copy.xlsx")

# Define independent and dependent variables
x1 = data$Price  # Dependent variable
x2 = data$RAM    # Independent variable 1
x3 = data$Battery  # Independent variable 2

# Check lengths
n1 = length(x1)
n2 = length(x2)
n3 = length(x3)

# Build multiple regression model
mur = lm(x1 ~ x2 + x3)

# Extract coefficients
cr = coefficients(mur)
mcr = matrix(cr)
a = mcr[1,1]  # Intercept
b = mcr[2,1]  # Coefficient for RAM
c = mcr[3,1]  # Coefficient for Battery
```

```
# Predicted values
x1est = fitted(mur)
meanx1 = mean(x1)

# Sum of squared errors
sum1 = sum((x1 - x1est)^2)
sum2 = sum((x1 - meanx1)^2)

# Mean residual sum of squares
mrss = sum1 / n1
# Residual standard deviation
residual_sd = sqrt(mrss)

# Create dataframe with actual and estimated prices
d1 = data.frame("Actual Price" = x1, "RAM" = x2, "Battery" = x3, "Estimated Price" = x1est)

# Coefficient of determination (R²)
cod = summary(mur)$r.squared

# Adjusted R²
adj_r2 = summary(mur)$adj.r.squared

# Correlation coefficients
r12 = cor(x1, x2)
r13 = cor(x1, x3)
r23 = cor(x2, x3)
```

```
# Multiple correlation coefficient
ne1 = r12^2 + r13^2 - 2 * r12 * r13 * r23
de1 = 1 - r23^2
R1.23 = sqrt(ne1 / de1)

# Rename columns
colnames(d1) = c("Actual_Price", "RAM", "Battery", "Estimated_Price")

# Scatter plot for RAM vs. Price
plot(x2, x1, main = "Price vs. RAM", xlab = "RAM", ylab = "Price", col = "blue", pch = 19)
abline(lm(x1 ~ x2), col = "red", lwd = 2)

# Residuals vs. Fitted Plot
plot(fitted(mur), residuals(mur), main = "Residuals vs. Fitted", xlab = "Fitted Values", ylab =
"Residuals", col = "purple", pch = 19)
abline(h = 0, col = "black", lwd = 2)

# Histogram of residuals
hist(residuals(mur), main = "Histogram of Residuals", xlab = "Residuals", col = "orange", breaks =
20)

# Model summary
summary(mur)
```

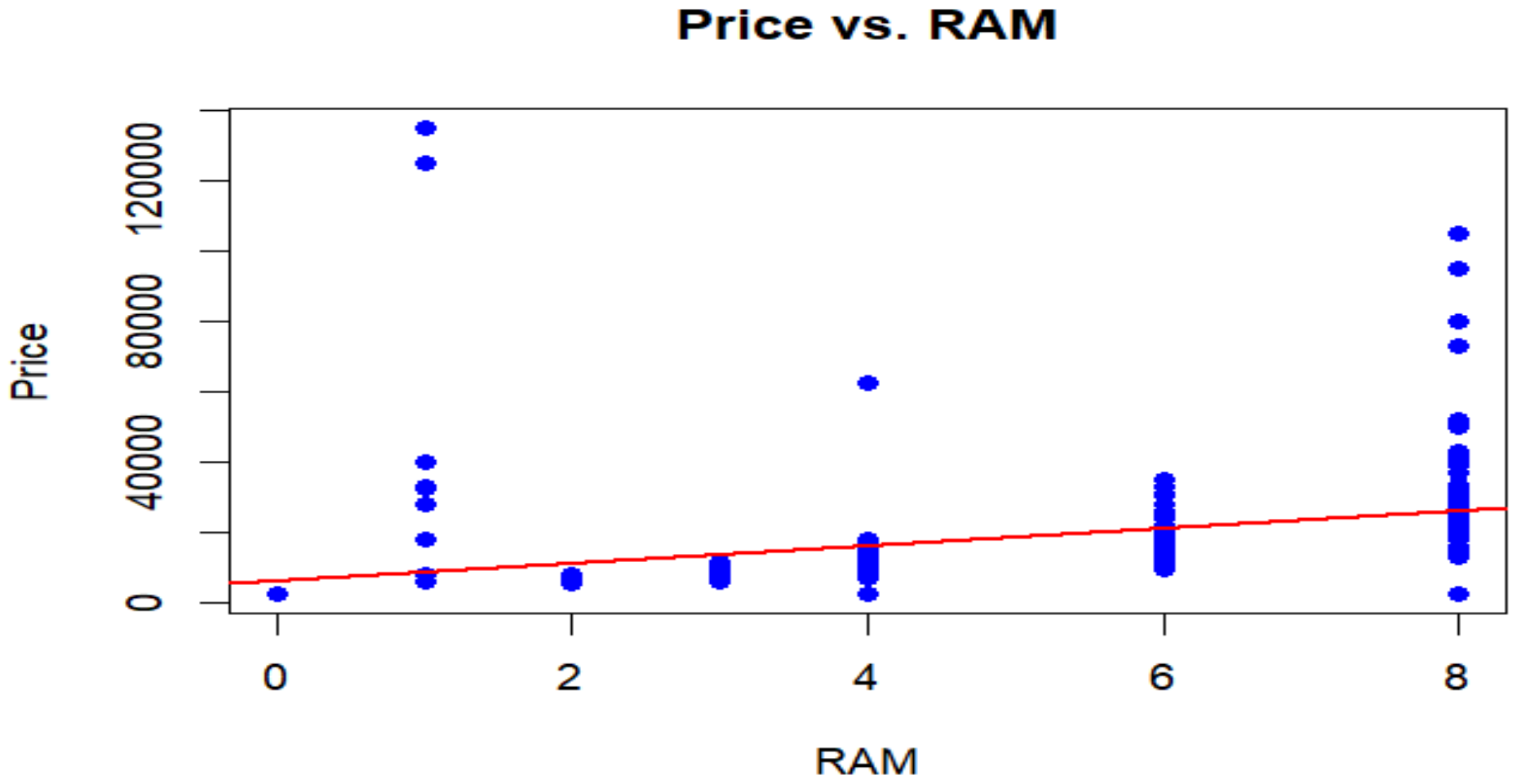
```
# Model Evaluation
cat("\n================ Model Evaluation ================\n")
cat("Multiple Regression Model: Price on RAM and Battery:\n")
cat("Price =", a, "+", b, "* RAM +", c, "* Battery\n")
cat("Mean Price:", meanx1, "\n")
cat("Sum of Squared Errors (SSE):", sum1, "\n")
cat("Total Sum of Squares (TSS):", sum2, "\n")
cat("Mean Residual Sum of Squares (MRSS):", mrss, "\n")
cat("Residual Standard Deviation:", residual_sd, "\n")
cat("Coefficient of Determination (R²):", cod, "\n")
cat("Adjusted R²:", adj_r2, "\n")
cat("Multiple Correlation Coefficient R1.23:", R1.23, "\n")

# Display if the model fit is poor
if (cod < 0.5) {
  cat("\nConclusion: The given dataset does not fit well.\n")
  cat("The R² value is low, indicating that RAM and Battery alone do not explain the variation in Price effectively.\n")
  cat("Consider adding more variables such as Processor type, Brand, Number of Ratings, etc., for better accuracy.\n")
} else {
  cat("\nConclusion: The model is a good fit with a reasonable R² value.\n")
}
```

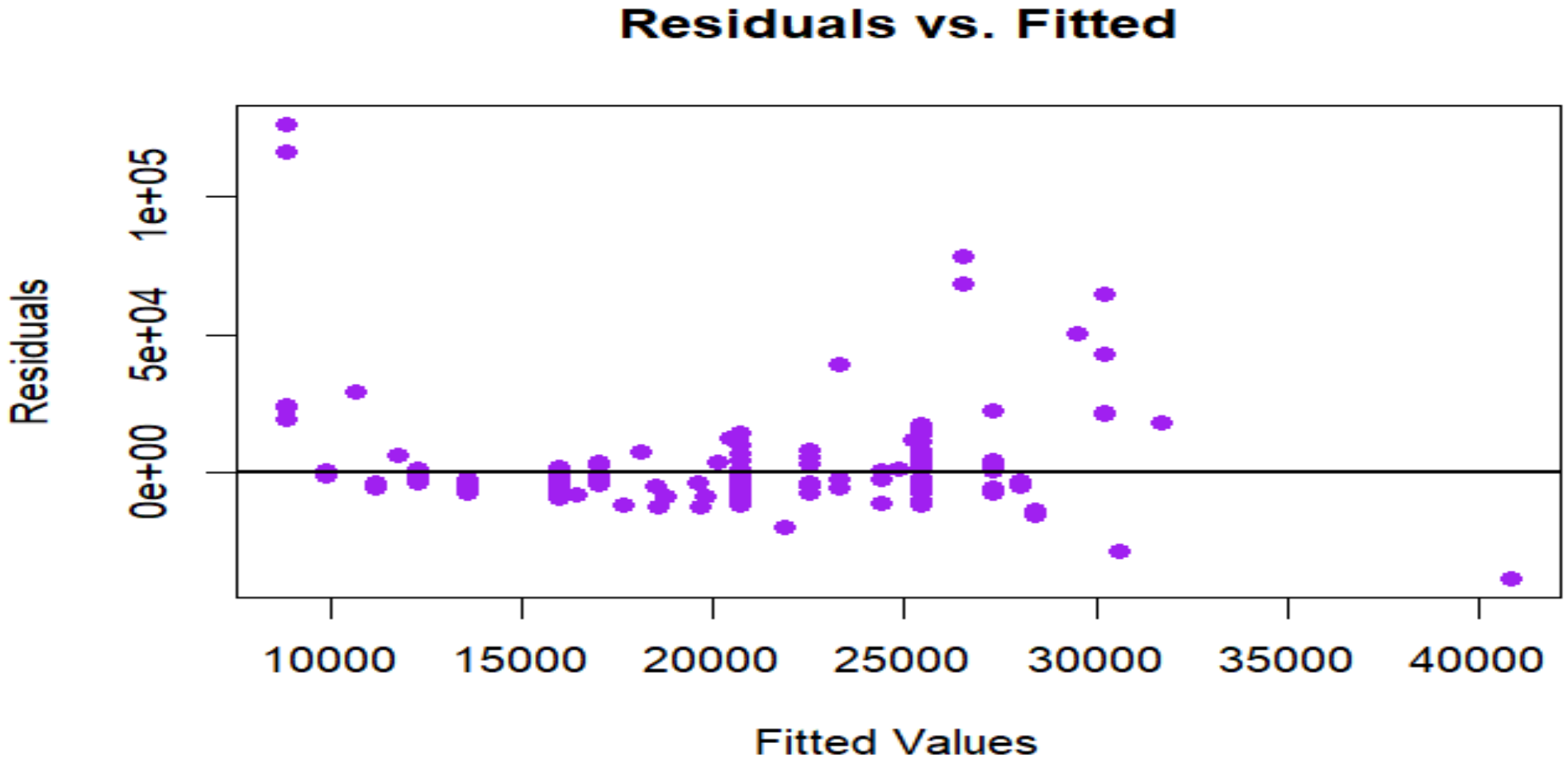
❑ Output :

Multiple Regression Model:
Price on RAM and Battery:
Price = 24812.56 + 2371.849 * RAM + -3.669905 * Battery
Mean Price: 19318.82
Sum of Squared Errors (SSE): 127442941453
Total Sum of Squares (TSS): 142279778165
Mean Residual Sum of Squares (MRSS): 255396676
Residual Standard Deviation: 15981.14
Coefficient of Determination (R^2): 0.1042793
Adjusted R^2: 0.1006675
Multiple Correlation Coefficient R1.23: 0.3229231
Conclusion: In the given dataset, the coefficient of determination (R^2) is 0.1042793, which is very low. Therefore, the fitted equation is not suitable for making accurate predictions on this dataset.

❑ Output(Graphs) :



❑ Output(Graphs) :



❑ Conclusion :

The **Multiple Regression Model** for predicting mobile phone prices using **RAM and Battery Capacity** shows a weak fit. The **R^2 value (0.1043)** is low, indicating that these two factors alone do not significantly explain price variations.

- Key observations:

1. The **regression equation:**

$$\textbf{Price} = 24812.56 + 2371.849 \times \textbf{RAM} - 3.669905 \times \textbf{Battery}$$

1. The **Residual Standard Deviation (15981.14)** is high, suggesting considerable variation in predicted prices.
2. The **Multiple Correlation Coefficient (0.3229)** further confirms weak relationships between the independent variables and price.

- **Final Verdict:**
  This model does not accurately predict mobile prices. To improve accuracy, additional factors such as **Processor Type, Brand, User Ratings, and Storage Capacity** should be considered in future models.

Thank you!