Introduction

In this assignment, we will use linear regression to model the relationship between income levels and happiness measures for individuals. You will work with the dataset income.data.csv, which contains data on individuals' income levels and their self-reported happiness scores.

Your tasks will involve loading the dataset into R, fitting a linear regression model, and evaluating the model's performance. This will allow you to understand the influence of income on happiness and explore basic concepts of regression analysis.

Please submit only your code. You don't have to submit your interpretations, but you're encouraged to work on them in your own time.

Dataset Description

The dataset `income.data.csv` contains two columns:

- `Income`: The annual income of the individuals (in USD).
- `Happiness`: A self-reported measure of happiness on a scale from 1 to 10.

The dataset contains information on several individuals, and you can load it into R using the following command:

```
# Load the dataset
data <- read.csv("income.data.csv")
```

The first few rows of the dataset look like this:

Income (USD)	Happiness
2	5.1
3.5	6.0
5	6.5
7.5	7.0
10	7.5

Problem 1: Simple Linear Regression Model

In this problem, you are tasked with fitting a simple linear regression model to predict happiness based on income.

- Load the data and fit a linear regression model using `Income` as the independent variable and `Happiness` as the dependent variable:

  ```
  model1 <- lm(Happiness ~ Income, data = data)
  ```

- Use the `summary()` function to view the results of the regression:

  ```
  summary(model1)
  ```

- Plot the data points and the regression line:

Answer the following questions:

- What is the estimated regression equation (Happiness $= \beta_0 + \beta_1 \times$ Income)?
- What is the value of R^2, and what does it indicate about the strength of the relationship between income and happiness?
- Interpret the coefficient β_1 (the slope of the regression line). What does it tell you about the relationship between income and happiness?

Problem 2: Model Evaluation

Now that you have created the linear regression model, it's important to evaluate its performance.

- Calculate the residuals of the model:

```
residuals1 <- residuals(model1)
```

- Plot the residuals to check for homoscedasticity (constant variance of residuals):

```
plot(residuals1, main = "Residuals of Simple Linear Regression")
```

- Calculate the residual sum of squares (RSS) for the model:

```
RSS1 <- sum(residuals1)
```

- Calculate the Mean Squared Error (MSE) for the model:

```
MSE1 <- mean(residuals1)
```

Answer the following questions:

- What do the residuals tell you about the model fit? Are there any patterns that indicate the model might not be a good fit?
- What is the value of the residual sum of squares (RSS), and what does it indicate about the fit of the model?
- What is the Mean Squared Error (MSE), and how can it be used to evaluate the model's prediction accuracy?