



Bansilal Ramnath Agarwal Charitable Trust's
Vishwakarma Institute of Information
Technology

**Department of
Artificial Intelligence and Data
Science**

Name: Siddhesh Dilip Khairnar

Class: SY

Division: B

Roll No: 272028

Semester: IV

Academic Year: 2022-2023

Subject Name & Code: ES22201AD: Probability and Statistics

Title of Assignment: Linear Regression in R

Date of Performance: 27-04-2023

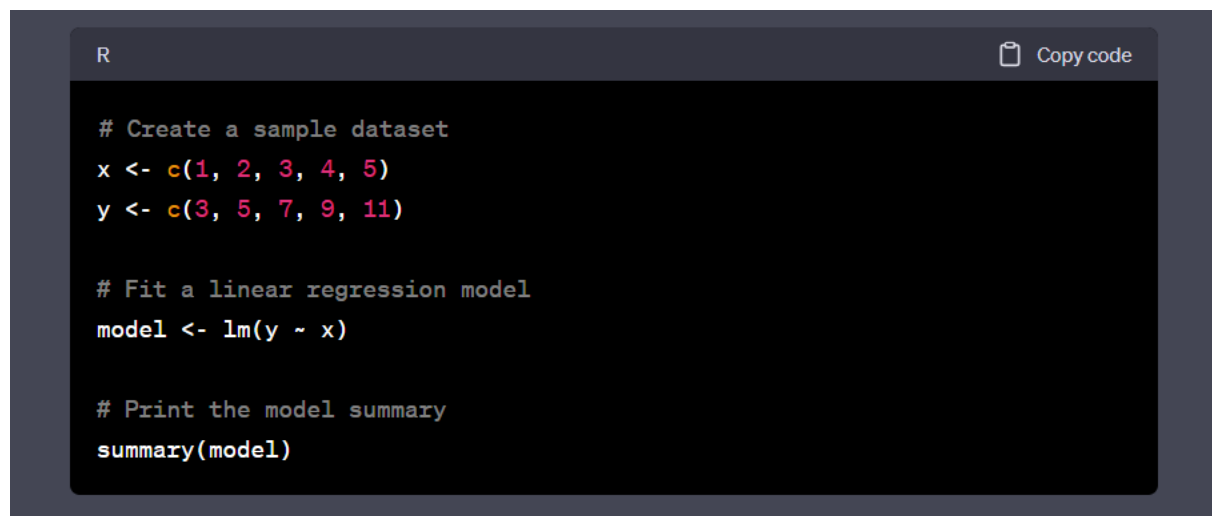
Date of Submission: 27-04-2023

ASSIGNMENT NO. 8

Background information:

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In other words, it is used to find the linear relationship between two or more variables by fitting a linear equation to the observed data.

In R, the `lm()` function is used to perform linear regression. Here is an example of how to perform a simple linear regression in R:

A screenshot of an R console window with a dark background. The title bar shows 'R' on the left and a 'Copy code' button on the right. The console contains the following R code:

```
# Create a sample dataset
x <- c(1, 2, 3, 4, 5)
y <- c(3, 5, 7, 9, 11)

# Fit a linear regression model
model <- lm(y ~ x)

# Print the model summary
summary(model)
```

In this example, we have created a simple dataset with two variables, `x` and `y`. We then fit a linear regression model to the data using the `lm()` function, with `y` as the dependent variable and `x` as the independent variable. The `summary()` function is used to print out a summary of the model, which includes information such as the intercept, the slope of the line, and the R-squared value.

Here are some other important terms to understand when working with linear regression:

- **Coefficients:** These are the values that represent the slope and intercept of the line in a linear regression equation. In R, you can access the coefficients of a linear regression model using the `coefficients()` function.
- **Residuals:** These are the differences between the predicted values of the dependent variable and the actual values. In R, you can access the residuals of a linear regression model using the `resid()` function.
- **R-squared value:** This is a measure of how well the linear regression model fits the data. It represents the proportion of the variance in the dependent variable that is explained by the independent variable(s). In R, you can access the R-squared value of a linear regression model using the `summary()` function.

Here is an example of how to use these functions to analyze a linear regression model:

```
R                                                                    Copy code

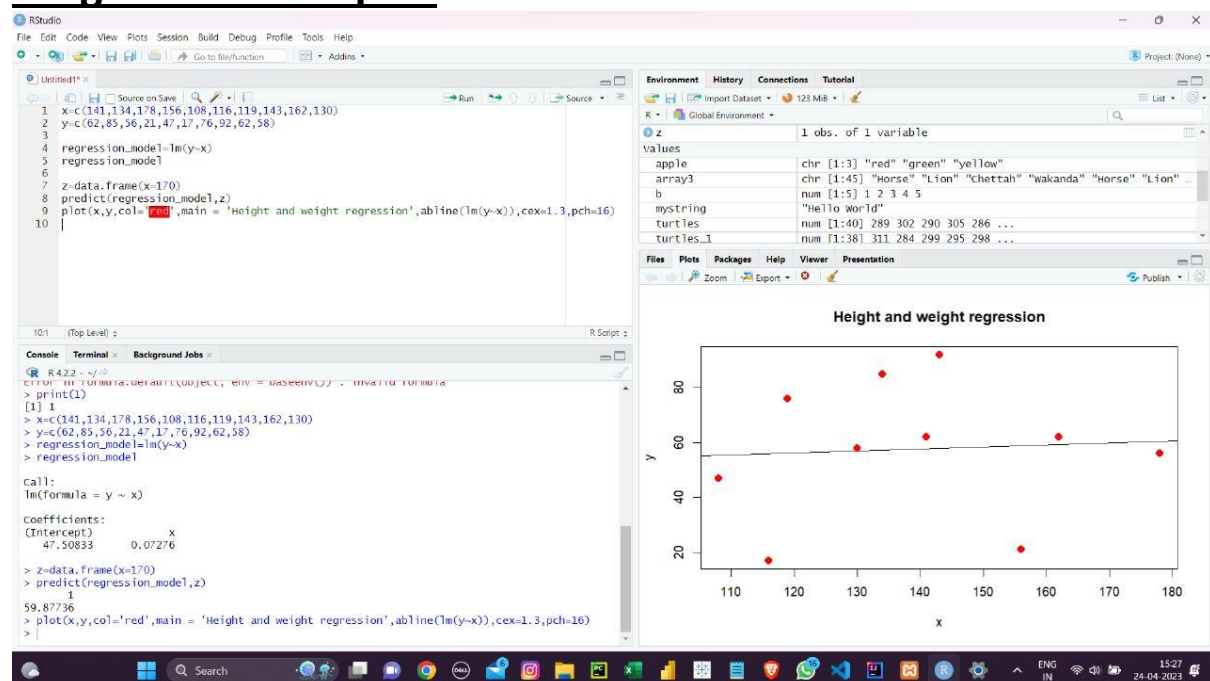
# Access the coefficients of the model
coefficients(model)

# Access the residuals of the model
resid(model)

# Access the R-squared value of the model
summary(model)$r.squared
```

In addition to simple linear regression, R can also be used to perform multiple linear regression (with more than one independent variable), polynomial regression (to model non-linear relationships), and logistic regression (to model binary outcomes). The process for fitting these models is similar to that of simple linear regression, but the specific functions and arguments used may differ.

Program and Output:



Conclusion: We have successfully implemented linear regression in RStudio.