Experiment No. 2

AIM: To write the implementation of linear regression in Python/R.

Objective: To understand the use of simple linear regression techniques by implementing user define dataset and importing dataset

Description:

Regression analysis is a very widely used statistical tool to establish a relationship model between two variables. One of these variables is called a predictor variable whose value is gathered through experiments. The other variable is called response variable whose value is derived from the predictor variable. In Linear Regression these two variables are related through an equation, where the exponent (power) of both these variables is 1. Mathematically a linear relationship represents a straight line when plotted as a graph. A non-linear relationship where the exponent of any variable is not equal to 1 creates a curve.

The general mathematical equation for a linear regression is −

y = ax + b

Following is the description of the parameters used −

· y is the response variable.

· x is the predictor variable.

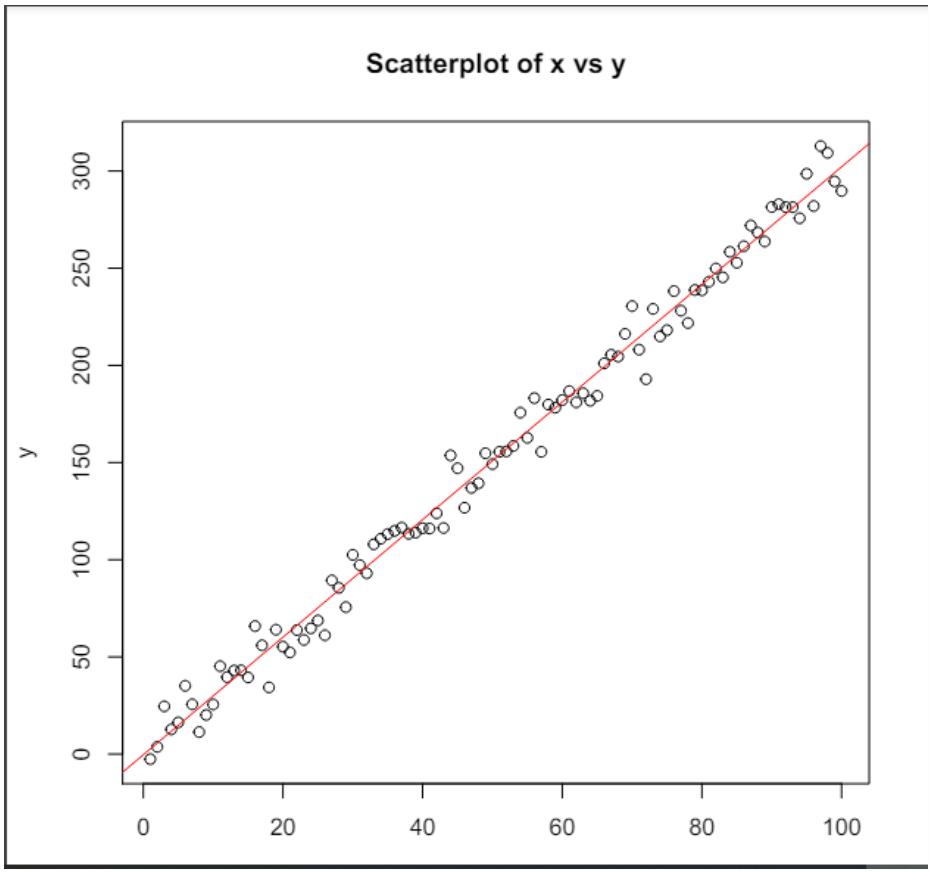
· a and b are constants which are called the coefficients.

Code:

```
# Generate random data

set.seed(123)  # for reproducibility

x <- 1:100

y <- 3 * x + rnorm(100, mean = 0, sd = 10)  # y = 3*x + noise
```

```
# Plot the data
plot(x, y, main = "Scatterplot of x vs y", xlab = "x", ylab = "y")

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

# Print summary of the model
summary(model)

# Plot the regression line
abline(model, col = "red")

# Predict using the model
new_x <- 101:110  # new values of x for prediction
predicted_y <- predict(model, data.frame(x = new_x))

# Print predicted values
print(predicted_y)
```

OUTPUT:



Conclusion:

1. Function used for linear regression in Python/R is _____ (Function_name (Parameters))

the function used for linear regression in Python is `linear_model.LinearRegression()` from the scikit-learn library, while in R, it's `lm()` from the base stats package. Both functions are widely used for fitting linear regression models to data, with Python's scikit-learn providing a comprehensive machine learning toolkit and R offering extensive statistical analysis capabilities.

These functions allow researchers, analysts, and data scientists to easily perform linear regression analysis and extract insights from their datasets.

2. Explain use of numpy library.

the NumPy library is indispensable for numerical computing tasks in Python, offering efficient array operations, mathematical functions, linear algebra capabilities, random number generation, and powerful indexing and slicing functionalities. Its ease of use and performance make it a go-to choice for data manipulation, scientific computing, and machine learning applications.