Name of Student: Pushkar Sane		
Roll Number: 45		**Lab Assignment Number: 8**
Title of Lab Assignment: Implementation of Normal and Binomial Distribution, Univariate and Bivariate analysis.		
DOP: 26-10-2023		**DOS: 27-10-2023**
CO Mapped: **CO6**	**PO Mapped:** **PO1, PO2, PO3, PO4, PO5, PO7, PO12, PSO1, PSO2**	**Signature:**

Practical No. 8

Aim: Implementation of Normal and Binomial Distribution, Univariate and Bivariate analysis.

Description:

1. **Normal distribution:**

   Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a "bell curve".

   It is generally observed that data distribution is normal when there is a random collection of data from independent sources. The graph produced after plotting the value of the variable on the x-axis and the count of the value on the y-axis is a bell-shaped curve graph. The graph signifies that the peak point is the mean of the data set and half of the values of the data set lie on the left side of the mean and the other half lies on the right part of the mean telling about the distribution of the values. The graph is a symmetric distribution.

   In R, there are 4 built-in functions to generate normal distribution:

   - dnorm()
     dnorm(x, mean, sd)
   - pnorm()
     pnorm(x, mean, sd)
   - qnorm()
     qnorm(p, mean, sd)
   - rnorm()
     rnorm(n, mean, sd)

2. **Binomial Distribution:**

   Binomial distribution is a common discrete distribution used in statistics, as opposed to a continuous distribution, such as normal distribution. This is because binomial distribution only counts two states, typically represented as 1 (for a success) or 0 (for a failure), given a number of trials in the data.

Binomial distribution in R is a probability distribution used in statistics. The binomial distribution is a discrete distribution and has only two outcomes i.e. success or failure. All its trials are independent, the probability of success remains the same and the previous outcome does not affect the next outcome. The outcomes from different trials are independent. Binomial distribution helps us to find the individual probabilities as well as cumulative probabilities over a certain range.

It is also used in many real-life scenarios such as in determining whether a particular lottery ticket has won or not, whether a drug is able to cure a person or not, it can be used to determine the number of heads or tails in a finite number of tosses, for analyzing the outcome of a die, etc.

We have four functions for handling binomial distribution in R namely:

- dbinom()
  dbinom(k, n, p)
- pbinom()
  pbinom(k, n, p)
  where n is total number of trials, p is probability of success, k is the value at which the probability has to be found out.
- qbinom()
  qbinom(P, n, p)
  Where P is the probability, n is the total number of trials and p is the probability of success..3
- rbinom()
  rbinom(n, N, p)

dbinom() Function:

This function is used to find probability at a particular value for a data that follows binomial distribution i.e. it finds:

P(X = k)

Syntax:

dbinom(k, n, p)

3. **Univariate Analysis:**

Univariate analysis explores each variable in a data set, separately. It looks at the range of values, as well as the central tendency of the values. It describes the pattern of response to the variable. It describes each variable on its own. Descriptive statistics describe and summarize data.

This type of data consists of only one variable. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it. The example of a univariate data can be height.

Suppose that the heights of seven students of a class is recorded(figure 1), there is only one variable that is height and it is not dealing with any cause or relationship. The description of patterns found in this type of data can be made by drawing conclusions using central tendency measures (mean, median and mode), dispersion or spread of data (range, minimum, maximum, quartiles, variance and standard deviation) and by using frequency distribution tables, histograms, pie charts, frequency polygon and bar charts.

4. **Bivariate Analysis:**

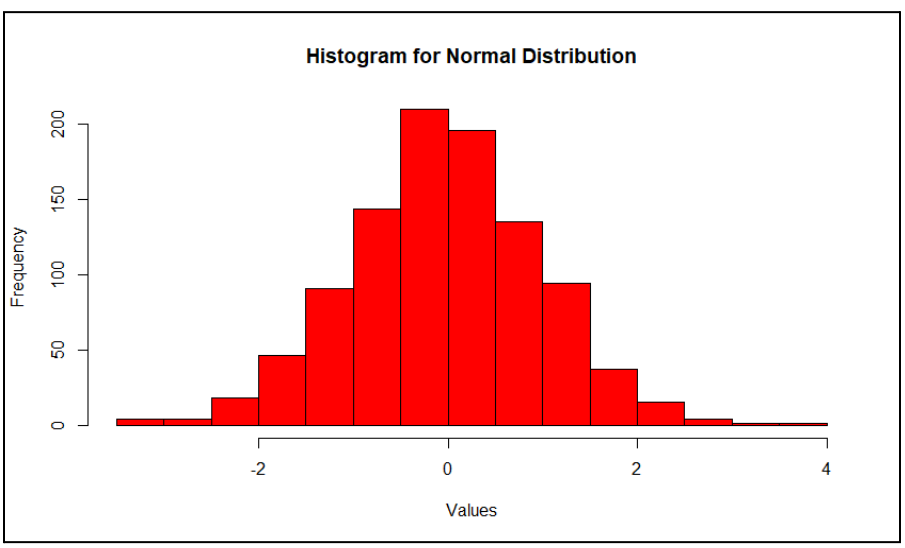
This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables. Examples of bivariate data can be temperature and ice cream sales in summer season.Suppose the temperature and ice cream sales are the two variables of a bivariate data. Here, the relationship is visible from the table that temperature and sales are directly proportional to each other and thus related because as the temperature increases, the sales also increase. Thus bivariate data analysis involves comparisons, relationships, causes and explanations. These variables are often plotted on the X and Y axis on the graph for better understanding of data and one of these variables is independent while the other is dependent.

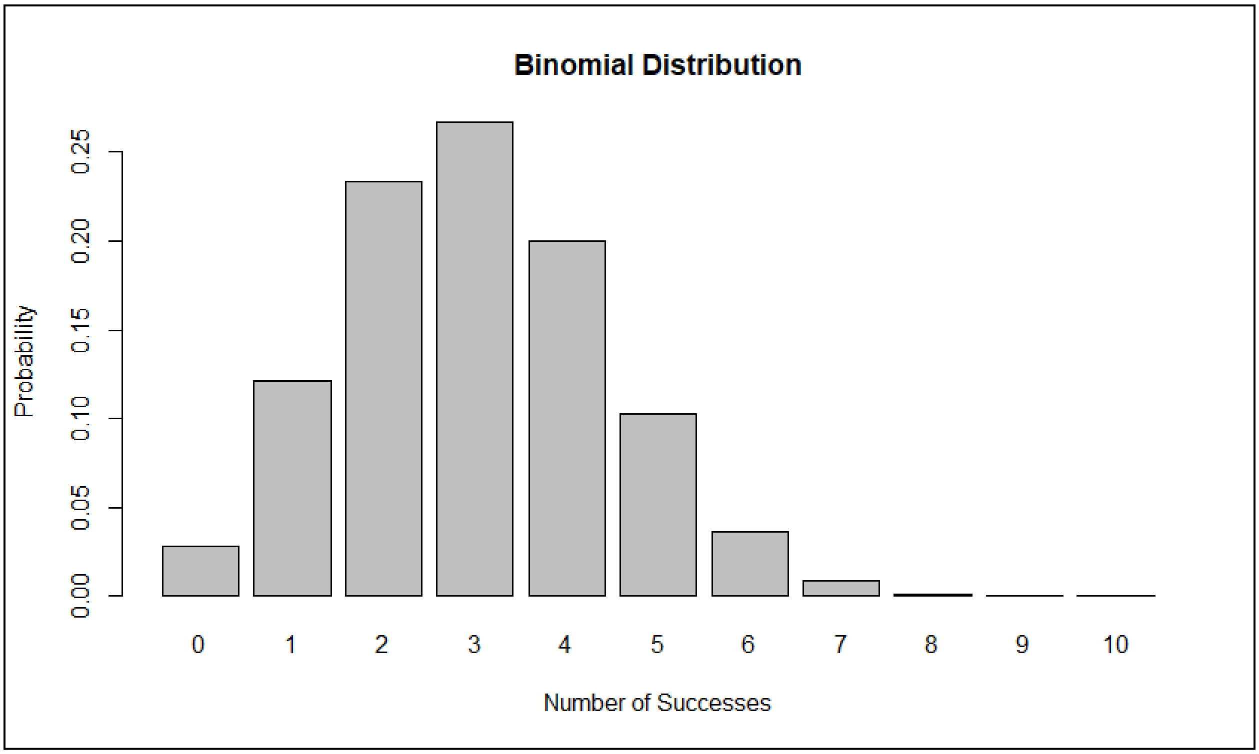
Code (Script):

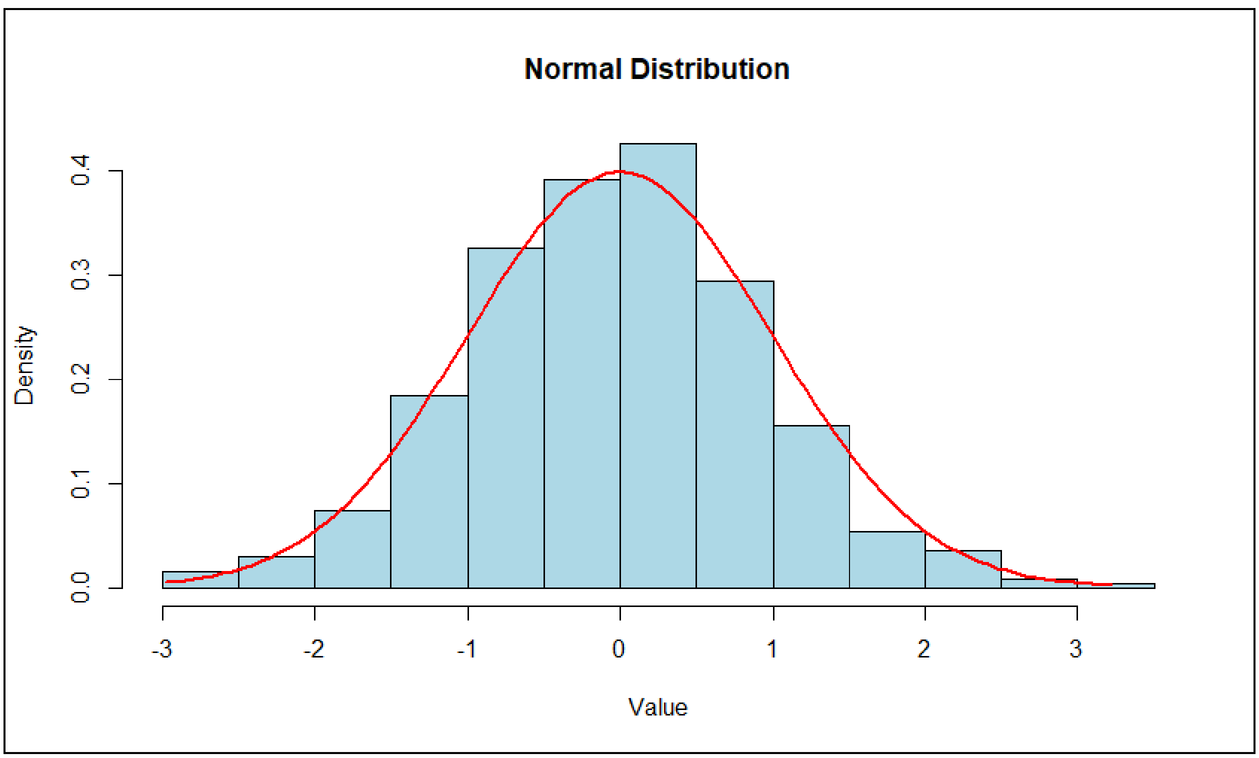
```
# Genereating Random data for normal distribution.
sample<- rnorm(100, mean=0, sd=1)

#calculating summary statistics.
mean_data<-mean(sample)
sd_data<- sd

# Creating a Histogram for data visualization.
hist(data, main="Histogram for Normal Distribution", xlab = "Values", ylab =
    "Frequency", col = "Red")

# Generate random data from a normal distribution
mean_value <- 0
sd_value <- 1
n <- 1000
data <- rnorm(n, mean = mean_value, sd = sd_value)

# Create a histogram
hist(data, main = "Normal Distribution", xlab = "Value", prob = TRUE, col = "lightblue")

# Overlay the probability density function (PDF)
x <- seq(min(data), max(data), length = 100)
y <- dnorm(x, mean = mean_value, sd = sd_value)
lines(x, y, col = "red", lwd = 2)

#Binomial Distribution
# Parameters for the binomial distribution
n_trials <- 10
probability_success <- 0.3
x_values <- 0:n_trials

# Compute the probability mass function
pmf <- dbinom(x_values, size = n_trials, prob = probability_success)
```

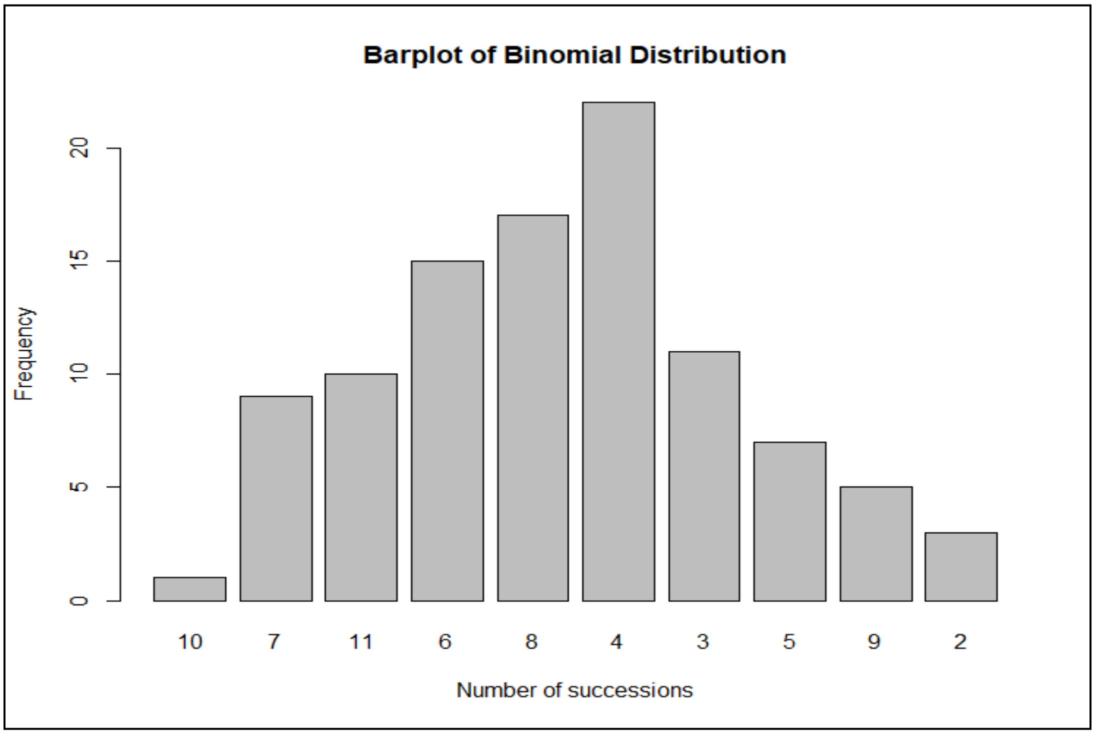
```
# Create a bar plot
barplot(pmf, names.arg = x_values, main = "Binomial Distribution", xlab = "Number of
Successes", ylab = "Probability")

# Generating data set for Binomil distribution.
data<- rbinom(100, size = 20, prob = 0.3)

#summary statistics.
mean_val<- mean(data)
sd_val<- sd(data)

# Visualiazation for the PMF of binomial distribution.
barplot(table(data), names.arg = unique(data), main="Barplot of Binomial Distribution",
    xlab="Number of successions", ylab="Frequency")

# Generating Random data for 2 variables in Normal distribution
x<-rnorm(100, mean=0, sd=1)
y<- rnorm(100, mean = 0, sd=3)

# Creating a scatterplot for Bivariate analysis.
plot(x,y, main="Scatter Plot for Bivariate Analysis", xlab = "X", ylab="Y")
```

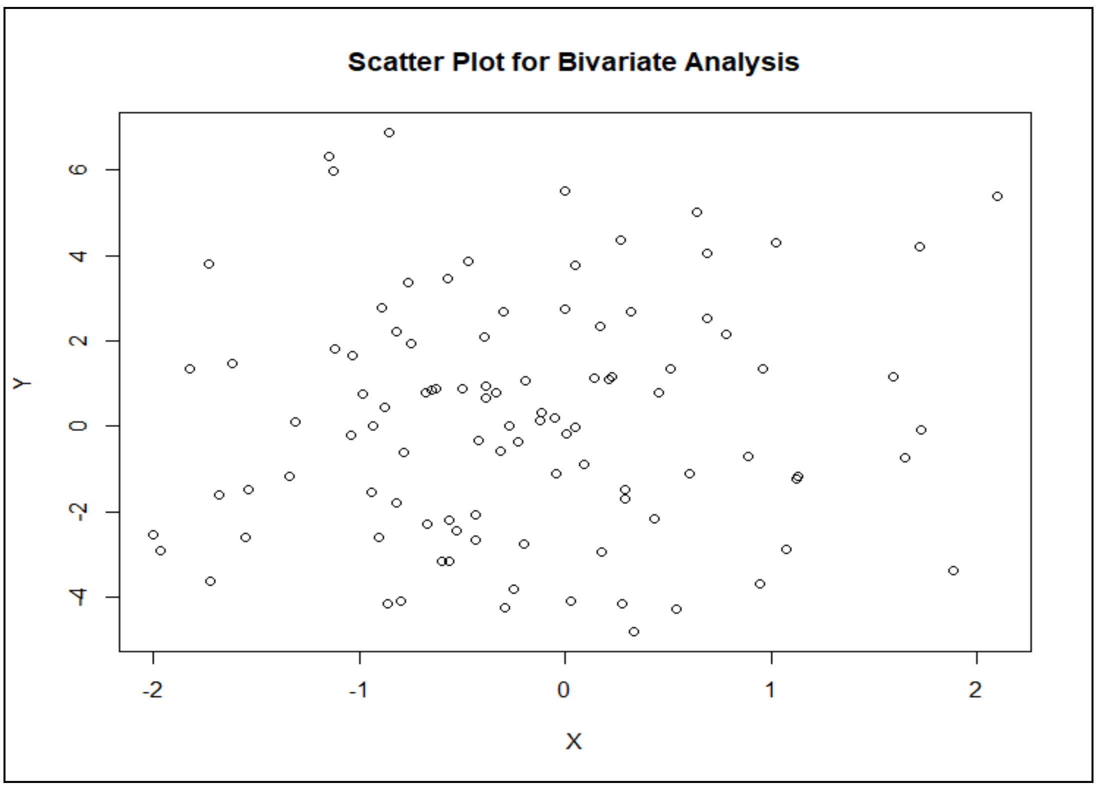
Output:




Binomial Distribution
Probability
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7
8
9
10
Number of Successes


Normal Distribution
Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
Value





Conclusion: Demonstrated the implementation of Normal and Binomial Distribution, Univariate and Bivariate analysis.