**MD2201: Data Science**

**Name of the student: Aaditya Patil Roll No. 58**

**Div: CS - C Batch: B3**

**Date of performance:**

# Experiment No.2

# Title: Laboratory on Statistics

**Aim:** To find the probability for Binomial distribution and Normal distribution and verify the Normal approximation of Binomial distribution.

**Software used:** Programming language R.

**Code Statement:**

Write a single R code to answer the following questions.

Data Set: Travelled Abroad

1. Find out the % of Indians in the sample who have travelled abroad using the data source.
2. Treating this value as ‘p’, calculate the following probabilities –
   1. What is the probability that in a randomly chosen sample of 10 persons, no one has travelled abroad?
   2. What is the probability that in a randomly chosen sample of 10 persons, exactly one has travelled abroad?
   3. What is the probability that in a randomly chosen sample of 10 persons, exactly two persons have travelled abroad?
   4. What is the probability that in a randomly chosen sample of 10 persons, exactly three persons have travelled abroad?
   5. What is the probability that in a randomly chosen sample of 10 persons, exactly four persons have travelled abroad?
   6. What is the probability that in a randomly chosen sample of 10 persons, exactly five persons have travelled abroad.
   7. What is the probability that in a randomly chosen sample of 10 persons, exactly six persons have travelled abroad?
   8. What is the probability that in a randomly chosen sample of 10 persons, exactly seven persons have travelled abroad?
   9. What is the probability that in a randomly chosen sample of 10 persons, exactly eight persons have travelled abroad?
   10. What is the probability that in a randomly chosen sample of 10 persons, exactly nine persons have travelled abroad?
   11. What is the probability that in a randomly chosen sample of 10 persons, all 10 persons have travelled abroad?
3. Plot the probability values as a Table / Bar graph/plot and interpret plot.
4. What is the probability that in the randomly chosen sample of 100 persons at least 59 have travelled abroad?

Hint: Expected to perform Normal approximation for the binary distribution.

**Code:**

f<-read.csv("travelledAbroad\_csv.csv")

#1. Find out the % of Indians in the sample who have travelled abroad using the data source

per<-100\*(sum (f$Travelledabroad =="Y" ))/150

cat("Percentage of people travelled abroad :",per)

#2. Treating this value as ‘p’, calculate the following probabilities –

p<-per/100

cat("\n\nProbalility of success :",p)

#a. What is the probability that in a randomly chosen sample of 10 persons, no one has travelled abroad?

k<-0:10

d1<-dbinom(k,10,p)

cat("\n\nprobability for n = 10 by binomial distribution:",d1)

#3. Plot the probability values as a Table / Bar graph/plot and interpret plot.

plot(k,d1,type ="l")

#4. What is the probability that in the randomly chosen sample of 100 persons at least 59 have travelled abroad?

#n = 100 in this case

k1<-59:100

d2<-dbinom(k1,100,p)

pbd1<-sum(d2)

cat("\n\nProbability per n =100 by binomial distribution :",pbd1)

d3<-dbinom(0:100,100,p)

plot(0:100,d3,type="l")

#for finding probability considering it as a normal distribution

#z = (observed value - mean value)/standard deviation

m<-100\*p

sd1<-sqrt(100\*p\*(1-p))

cat("\n\nmean is :",m)

cat("\nstd dev is :",sd1)

pnorm(59,m,sd1)

pnd1<-pnorm(59,m,sd1, lower.tail = F)

cat("\n\nProbability for n = 100 by normal distribution :",pnd1)

# Results: Display the output obtained on R console for all the cases. Also add the plots which you obtained. Give proper title to the plots as per the condition.

# 

# 

# 

# Conclusion:

With the help of the normal distribution and the binomial distribution, we were able to display two charts and determine the likelihood that at least 59 out of every 100 people have visited another country.

According to the plots, 6 out of a sample of 10 people have travelled abroad, and 65 out of 100 people have done so as well.