



MD2201: Data Science

Name of the student: Bhavin Patil

Roll No. 78

Div: D

Batch: B-3

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 –
 - a. What is the probability that in a randomly chosen sample of 10 persons, no one has travelled abroad?
 - b. What is the probability that in a randomly chosen sample of 10 persons, exactly one has travelled abroad?
 - c. What is the probability that in a randomly chosen sample of 10 persons, exactly two persons have travelled abroad?
 - d. What is the probability that in a randomly chosen sample of 10 persons, exactly three persons have travelled abroad?
 - e. What is the probability that in a randomly chosen sample of 10 persons, exactly four persons have travelled abroad?
 - f. What is the probability that in a randomly chosen sample of 10 persons, exactly five persons have travelled abroad.
 - g. What is the probability that in a randomly chosen sample of 10 persons, exactly six persons have travelled abroad?
 - h. What is the probability that in a randomly chosen sample of 10 persons, exactly seven persons have travelled abroad?
 - i. What is the probability that in a randomly chosen sample of 10 persons, exactly eight persons have travelled abroad?
 - j. What is the probability that in a randomly chosen sample of 10 persons, exactly nine persons have travelled abroad?
 - k. 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:

```
# Binomial distribution
f<-read.csv("travelled abroad_csv.csv")
p <- sum(f$Travelledabroad == "Y") / nrow(f)
per <- p * 100
d1 <- dbinom(0:10,10,p)

cat("\n\nProbability p : ", p)
cat("\n\nPercentage of ppl travelled abroad : ", per)
cat("\n\nProbability value from 0 to 10 :\n ",d1)

k <- 0:10
plot(k,d1, type="l")

# to check success and failure satisfaction (success failure rule):
#np>=10
#n(1-p)>=10
#if it is satisfied then normal distribution can be applied
#n is number of samples

# Q.4
d2 <- dbinom(0:100,100,p)
pbd <- sum(dbinom(59:100,100,p))
cat("\n\nProbability for k>59 with Binomial Distribution: ",pbd)
k1<-0:100
plot(k1,d2,type="l")

# Normal Distribution
# mu = np
# sigma = sqrt(np(1-p))
m<- 100*p
sd <- sqrt(100*p*(1-p))
cat("\n\nMean: ", m)
cat("\n\nStandard Deviation: ",sd)
p <- pnorm(59,m,sd,lower.tail = F)
cat("\n\nProbability for k> 59 by Normal Distribution: ",p)
```

Results:

Probability p : 0.56

Percentage of ppl travelled abroad : 56

Probability value from 0 to 10 :

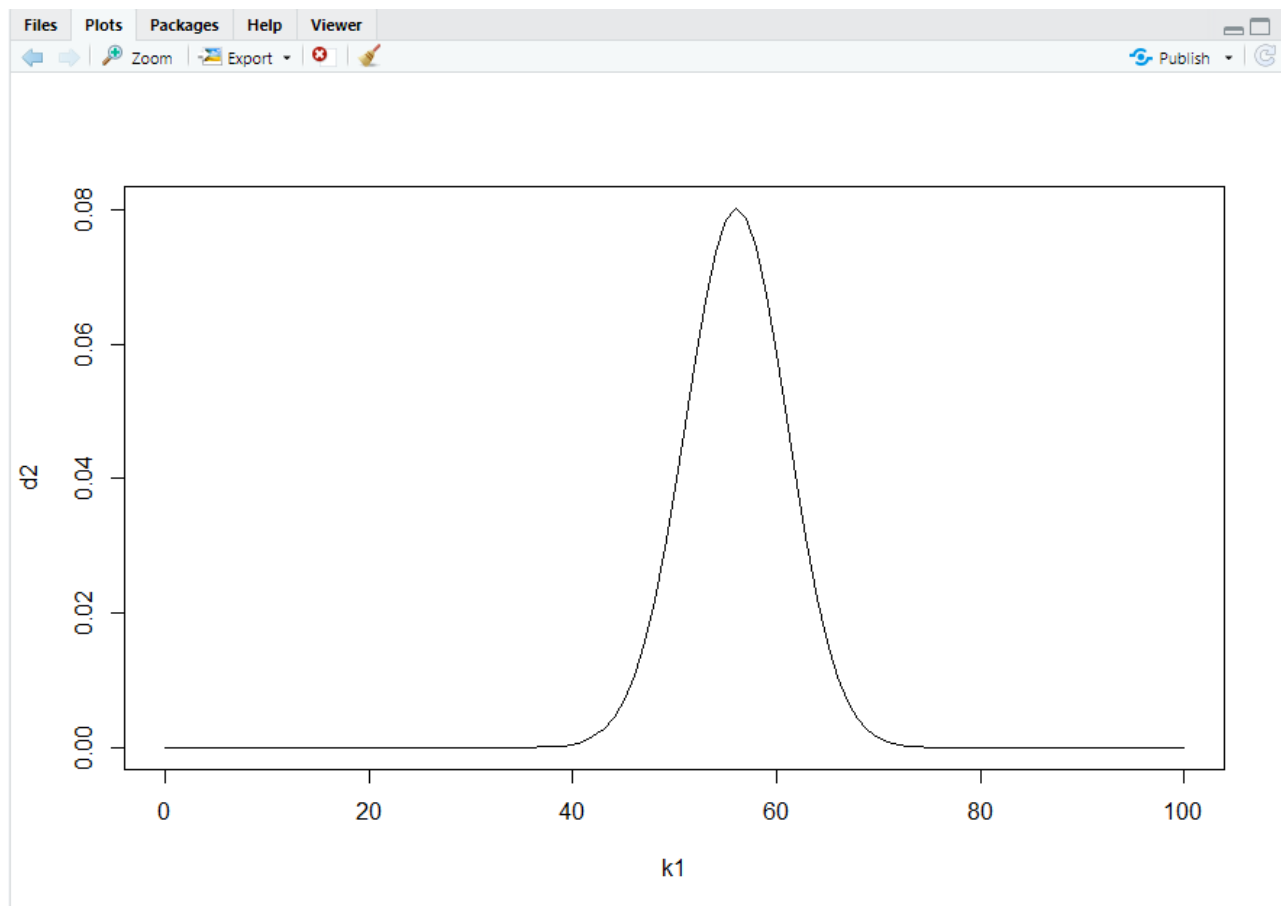
0.0002719736 0.003461482 0.01982485 0.06728435 0.1498606 0.228878 0.2427494 0.176545
0.08426012 0.02383115 0.003033055

Probability for $k > 59$ with Binomial Distribution: 0.3084356

Mean: 56

Standard Deviation: 4.963869

Probability for $k > 59$ by Normal Distribution: 0.2727998



Conclusion: Using the proper commands we have successfully find the probability for Binomial distribution and Normal distribution and verified the Normal approximation of Binomial distribution.