Lab Session 4:Normal Disribution

Chapter 3:Normal Distribution

When plotting a graph with the value of the quantitative variable in the horizontal axis and the count of the values in the vertical axis if we get a bell shape curve, the distribution is known as Normal distribution. The center of the curve represents the mean of the data set. In the graph, fifty percent of values lie to the left of the mean, and the other fifty percent lie to the right of the graph.

R has four in built functions to generate normal distribution. They are described below.

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

Following is the description of the parameters used in above functions −

- **x** is a vector of numbers.
- **p** is a vector of probabilities.
- **n** is number of observations(sample size).
- **mean** is the mean value of the sample data. It's default value is zero.
- **sd** is the standard deviation. It's default value is 1.

dnorm()

This function gives height of the probability distribution at each point for a given mean and standard deviation.

```
#Add a normal curve to the histogram
x <- mtcars$mpg
h<-hist(x, breaks=10, col="red", xlab="Miles Per Gallon",
  main="Histogram with Normal Curve")
xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
```

pnorm()

This function gives the probability of a normally distributed random number to be less that the value of a given number. It is also called "Cumulative Distribution Function".

```
# x<=15
p<-pnorm(15,mean=mean(x),sd=sd(x))

To find P(X ≤ 900):
```

```
p<-pnorm(900, mean=1010, sd=218)
```

Exercise: Find P(X>1000)

qnorm()

This function takes the probability value and gives a number whose cumulative value matches the probability value

```
# p=0.2
x<-qnorm(0.2,mean=mean(x),sd=sd(x))

To find the pth percentile of the distribution of Y:
qnorm(0.3, mean=1010, sd=218)
```

Exercise: Find the first quartile and median of Y

rnorm()

This function is used to generate random numbers whose distribution is normal. It takes the sample size as input and generates that many random numbers. We draw a histogram to show the distribution of the generated numbers.

```
# Create a sample of 50 numbers having standard normal distribution.
y <- rnorm(50)

#To generate 30 random values that are distributed Normal with mean 15 and standard deviation 3:
rnorm(30,15,3)
```

Exercise: Generate 100 random values that are distributed Normal with Mean 50 and standard deviation 10 and Plot the histogram for this sample.