Experiment 11:

Demonstrate the following statistical distribution functions in R:

a. Normal Distribution

b. Binomial Distribution

c. Poisson Distribution

d. Chi-Square Distribution

- Aim: To understand and use various statistical distributions of data

Exp. 11: Statistical distribution functions in R

- A probability distribution describes how the values of a random variable is distributed
- For example, the collection of all possible outcomes of a sequence of coin tossing is known to follow the **Binomial distribution**
- The means of sufficiently large samples of a data population are known to resemble the **Normal distribution**
- These distributions can be used to make statistical inferences on the entire data population as a whole

Exp. 11: Statistical distribution functions in R

- R has functions available for most of the famous statistical distributions.
- Prefix the name as follows:
  - With d for the density or probability mass function (pmf)
  - With p for the cumulative distribution function (cdf)
  - With q for quantiles
  - With r for random number generation

Exp. 11: Statistical distribution functions in R

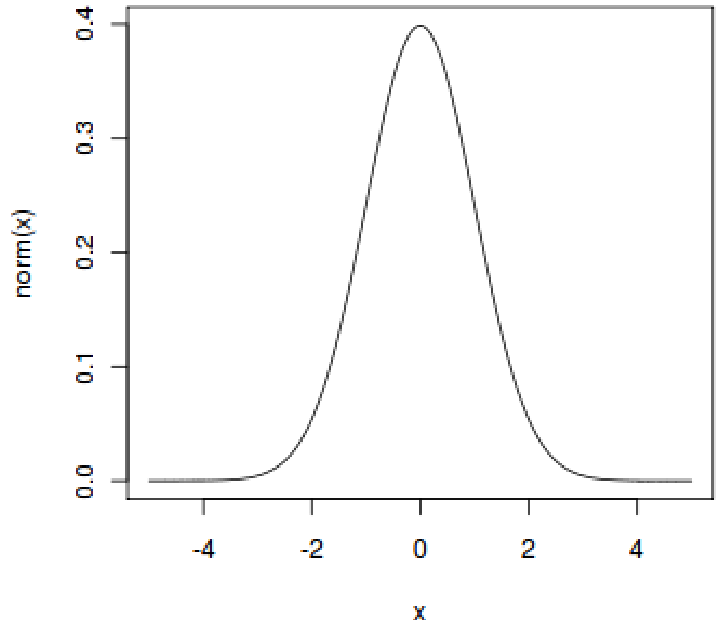
Distribution	Density/pmf	cdf	Quantiles	Random Numbers
Normal	dnorm()	pnorm()	qnorm()	rnorm()
Binomial	dbinom()	pbinom()	qbinom()	rbinom()
Poisson	dpois()	ppois()	qpois()	rpois()
Chi-square	dchisq()	pchisq()	qchisq()	rchisq()

Exp. 11 a: Normal distribution in R

- **Normal Distribution** is a probability function used in statistics that tells about how the data values are distributed
- It is the most important probability distribution function used in statistics because of its advantages in real case scenarios
- For example, the height of the population, shoe size, IQ level, rolling a dice, and many more

Exp. 11 a: Normal distribution in R

- 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 x-axis and count of the value on y-axis is bell-shaped curve graph



Exp. 11 a: Normal distribution in R

- The graph signifies that the peak point is the mean of the data set and half of the values of data set lie on the left side of the mean and other half lies on the right part of the mean, telling about the distribution of the values
- The graph is symmetric distribution

Exp. 11 a: Normal distribution in R

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

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

– *x,q represents the vector of quantiles*

– *p is vector of probabilities*

– *n is number of observations*

– *mean(x) represents the mean of data set* ***x****. (Default value:0)*

Exp. 11 a: Normal distribution in R

– ***sd** represents the standard deviation of data set **x** (Default value:1)*

$$= \sqrt{\frac{\sum_{i=1}^{n}(x_i - mean)^2}{n}}$$

Exp. 11 a: Normal distribution in R

dnorm(x, mean, sd)

- **dnorm()** function in R programming measures density function of distribution. In statistics, it is measured by below formula-

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$

- where, **μ** is mean and **σ** is standard deviation.

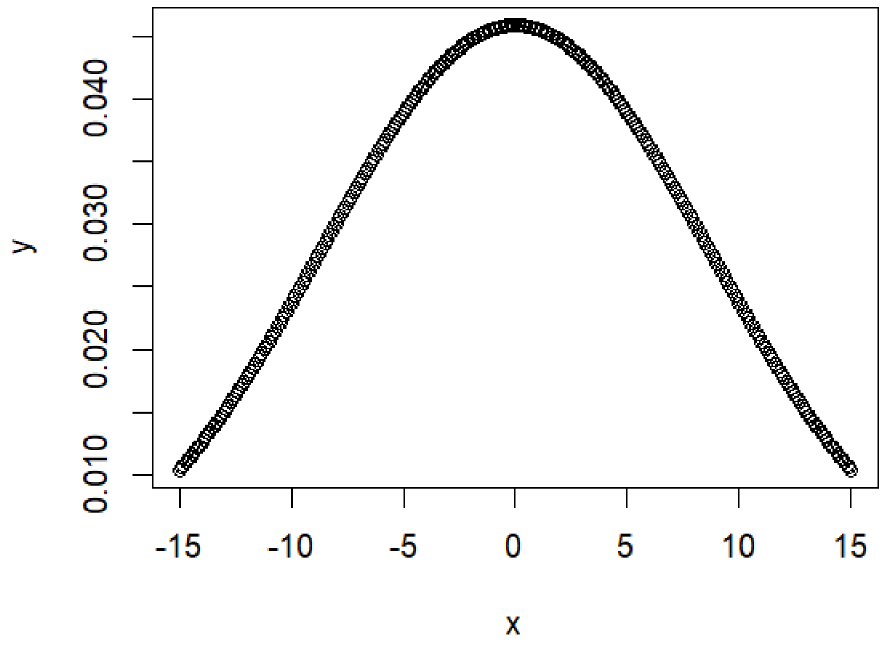
Exp. 11 a: Normal distribution in R

dnorm(x, mean, sd)

```
> x = seq(-15, 15, by=0.1)

> y = dnorm(x, mean(x), sd(x))

> plot(x,y)
```



Exp. 11 a: Normal distribution in R

pnorm(q, mean, sd)

- **pnorm()** function is the cumulative distribution function which measures the probability that a random number takes a value less than or equal to x, i.e., in statistics it is given by-

$$F_X(x) = Pr[X \le x] = \alpha$$

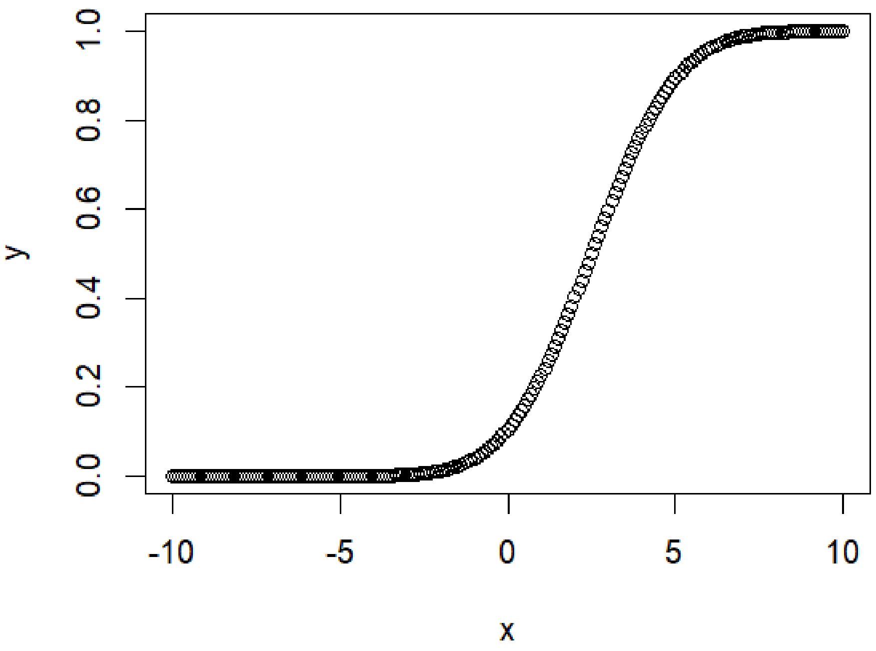
Exp. 11 a: Normal distribution in R

pnorm(q, mean, sd)

```
> x <- seq(-10, 10, by=0.1)

> y <- pnorm(x, mean = 2.5, sd = 2)

> plot(x,y)
```



Exp. 11 a: Normal distribution in R

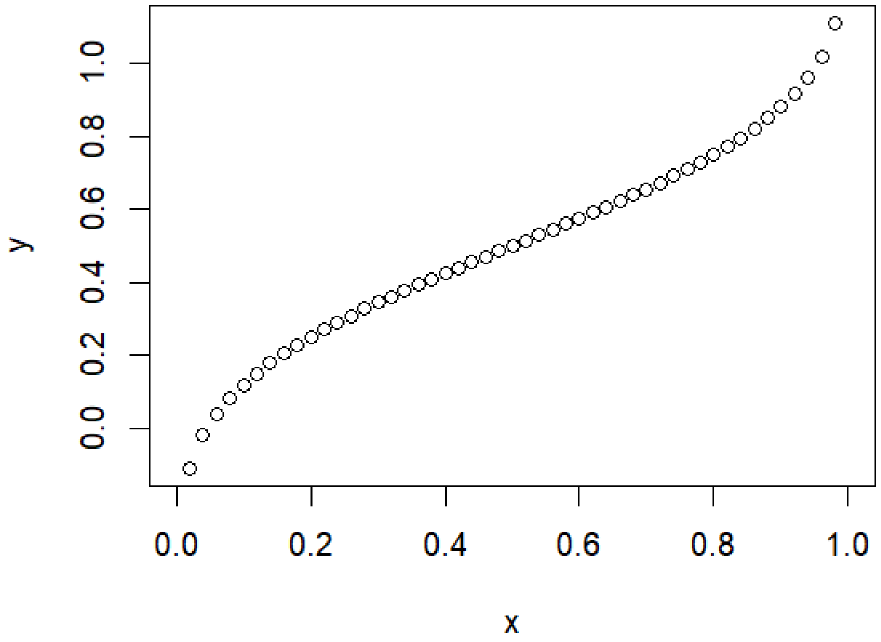
qnorm(p, mean, sd)

- This function takes the probability value and gives a number whose cumulative value matches the probability value
- It is useful in finding the percentiles of a normal distribution

Exp. 11 a: Normal distribution in R

qnorm(p, mean, sd)

```
> x <- seq(0, 1, by = 0.02)
> y <- qnorm(x, mean(x), sd(x))
> plot(x, y)
```



Exp. 11 a: Normal distribution in R

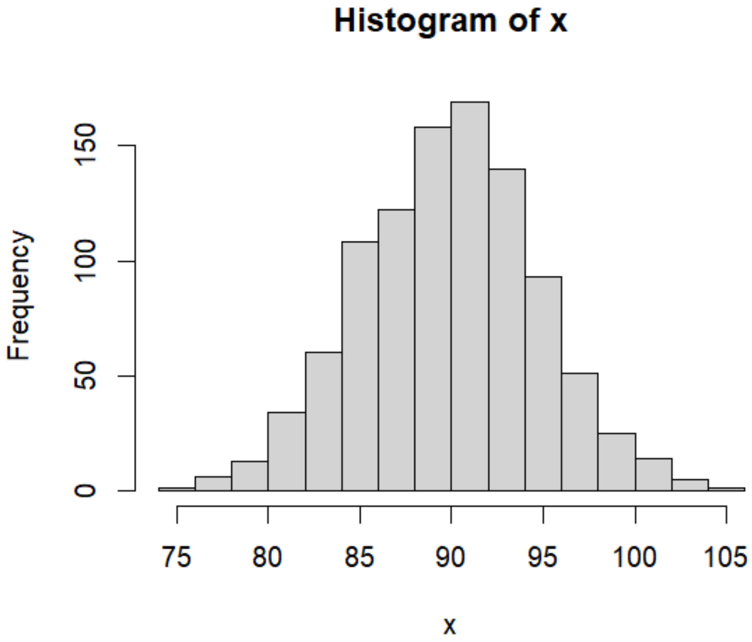
rnorm(n, mean, sd)

- **rnorm()** function in R programming is used to generate a vector of random numbers which are normally distributed

 # Create a vector of 1000 random numbers with mean=90 and sd=5

> x <- rnorm(1000, mean=90, sd=5)

> hist(x)



Exp. 11 b: Binomial distribution in R

- The binomial distribution is a discrete distribution and has only two outcomes i.e. success or failure
- The previous outcome does not affect the next outcome
- It is 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, for analyzing the outcome of a dice, etc.

Exp. 11 b: Binomial distribution in R

- In statistics, it is measured by below formula:

$P(X = k) = nC_r p^r q^{n-r}$, where r = 0, 1, 2, 3…, n

p is the probability of success

q is the probability of failure

$p + q = 1$

Exp. 11 b: Binomial distribution in R

We have four functions for handling binomial distribution in R:

- dbinom(x, size, prob)
- pbinom(q, size, prob)
- qbinom(p, size, prob)
- rbinom(n, size, prob)
  - *x, q represents the vector of quantiles*
  - *p is vector of probabilities*
  - *n is number of observations*
  - *size is number of trials (zero or more)*
  - *prob is probability of success on each trial*

Exp. 11 b: Binomial distribution in R

- **dbinom(x, size, prob)** -This function is used to find probability at a particular value for a data that follows binomial distribution i.e. it finds:

  $$P(X = x)$$

  Eg.1.

  # Probability of getting 3 when a dice is thrown 13 times

  >a = dbinom(3, size = 13, prob = 1 / 6)

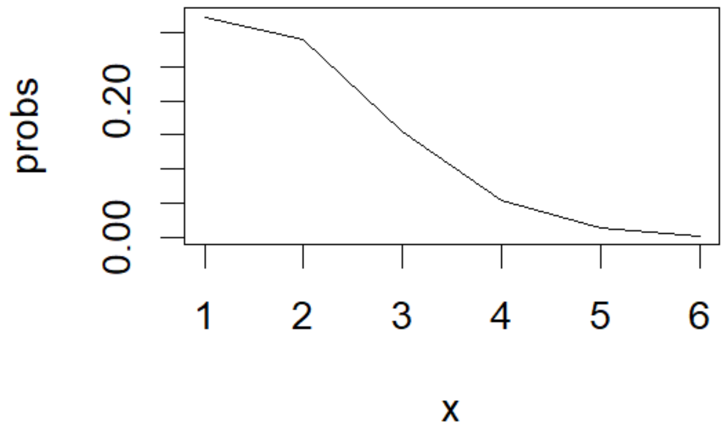
  >print(a)

Exp. 11 b: Binomial distribution in R

- **dbinom(x, size, prob)**

Eg.2.

```
> x=c(1:6)
> probs <- dbinom(x, 10, 1 / 6)
> print(probs)
> plot(x, probs, type = "l")
```



Exp. 11 b: Binomial distribution in R

- **pbinom(q, size, prob)** -It is used to find the cumulative probability of a data following binomial distribution till a given value i.e. it finds

  P(X <= q)

Probability of getting 26 or less heads from 51 tosses of a coin.

```
> x <- pbinom(26,51,0.5)
> print(x)
```

Exp. 11 b: Binomial distribution in R

- **qbinom(p, size, prob)** -This function takes the probability value and gives a number whose cumulative value matches the probability value

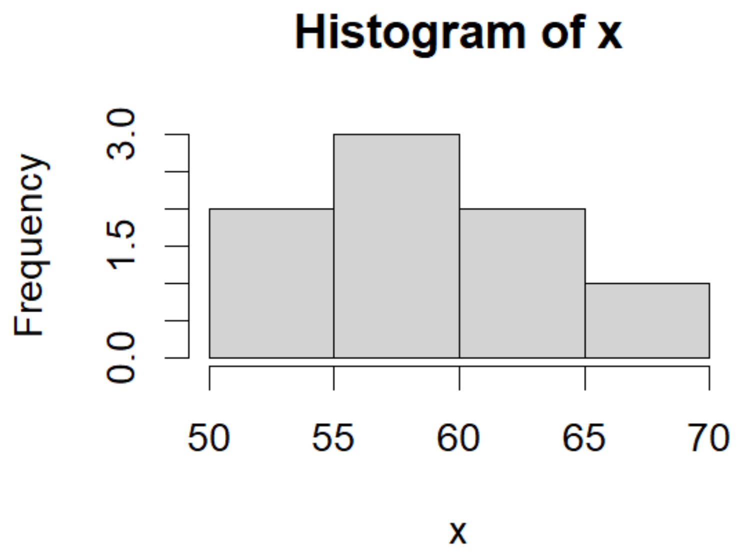
How many heads will have a probability of 0.25 will come out when a coin is tossed 51 times

```
> x <- qbinom(0.25,51,1/2)
> print(x)
```

Exp. 11 b: Binomial distribution in R

- **rbinom(n, size, prob)**-This function generates required number of random values of given probability from a given sample

```
# Find 8 random values from a sample of 150 with probability of 0.4
> x <- rbinom(8,150,0.4)
> print(x)
> hist(x)
```



Exp. 11 c: Poisson distribution in R

- **Poisson distribution**: It is the probability distribution of independent event occurrences in an interval
- If λ is the mean occurrence per interval, then the probability of having *x* occurrences within a given interval is:

$$f(x)=\frac{\lambda^{x}e^{-\lambda}}{x!} \qquad \text{where } x=0,1,2,3,\ldots$$

Exp. 11 c: Poisson distribution in R

- There are four Poisson distribution functions available in R:
  - dpois(x, lambda)
  - ppois(q, lambda)
  - qpois(p, lambda)
  - rpois(n, lambda)

Exp. 11 c: Poisson distribution in R

- **dpois(x,lambda)**– It calculates the probability of a random variable that is available within a certain range

- If there are five cars crossing a bridge per minute on average, find the probability of having 3 cars crossing the bridge in a particular minute

```
> y=dpois(3, lambda = 5)
> print(y)
```

Exp. 11 c: Poisson distribution in R

- **ppois(q, lambda)** –It calculates the probability of a random variable that will be equal to or less than a number

- If there are twelve cars crossing a bridge per minute on average, find the probability of having upto 16 cars crossing the bridge in a particular minute

```
> c=ppois(16,lambda=12)
> print(c)
```

Exp. 11 c: Poisson distribution in R

- **qpois(p, lambda)-** It allows obtaining the corresponding Poisson quantiles for a set of probabilities

- For instance, the quantile 0.75 of a Poisson distribution for $\lambda = 10$:

> **h=qpois**(0.75, lambda = 10)

> print(h) #12

Exp. 11 c: Poisson distribution in R

- **rpois(n,lambda)-**To draw n observations from a Poisson distribution

- To obtain 10 random observations from a Poisson distribution with mean = 4

```
> k=rpois(10, lambda = 4)
> print(k)                              #10 values generated
```

Exp. 11 d: Chi-Square distribution in R

- **Chi-Square distribution** - It is the distribution of the sum of squared standard normal deviates
- A standard normal deviate is a random sample from the standard normal distribution.
- The degrees of freedom of the distribution is equal to the number of standard normal deviates being summed
- The Chi Square distribution is very important because many test statistics are approximately distributed as Chi Square
- Formula : $V = X_1^2 + X_2^2 + \ldots + X_m^2 \sim \chi_{(m)}^2$ where m is degrees of freedom

Exp. 11 d: Chi-Square distribution in R

- **dchisq(x, df)**
- **pchisq(q, df)**
- **qchisq(p, df)**
- **rchisq(n, df)**

where

x, q -vector of quantiles

p -vector of probabilities

n -number of observations

df-degrees of freedom (non-negative, but can be non-integer).

Exp. 11 d: Chi-Square distribution in R

- **dchisq(x, df)**

```
> df = 6
> a=dchisq(4,df)
> print(a)                                    # prints 1 value

> vec <- 1:4
> a=dchisq(vec,df)
> print(a)                                    # prints 4 values
```

Exp. 11 d: Chi-Square distribution in R

- **pchisq(q, df)**

```
df = 5
k=pchisq(5, df)                    # Prob. for <=5
print(k)
```

Exp. 11 d: Chi-Square distribution in R

- **qchisq(p, df)**

```
> qchisq(0.95, df=7)          #7 degrees of freedom
[1] 14.067
```

The 95th percentile of the Chi-Squared distribution with 7 degrees of freedom is 14.067

Exp. 11 d: Chi-Square distribution in R

- **rchisq(n, df)**

  > x=rchisq(10,df=5)

  > print(x) #print 10 values

  > hist(x)