Binomial Distribution

The **binomial distribution** is a discrete probability distribution. It describes the outcome of *n* independent trials in an experiment. Each trial is assumed to have only two outcomes, either success or failure. If the probability of a successful trial is *p*, then the probability of having *x* successful outcomes in an experiment of *n* independent trials is as follows:-

Probability formula:- $P(x) = \binom{n}{x} p^x q^{n-x}$

Syntax:-

```
dbinom(x, size, prob, log = FALSE)
pbinom(x, size, prob, lower.tail = TRUE, log.p = FALSE)
qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
rbinom(n, size, prob)
```

Value:-

dbinom() gives the density
pbinom() gives the distribution function
qbinom() gives the quantile function
rbinom() generates random deviates.

Problem:-

In the past few years, outsourcing overseas has become more frequently used than ever before by U.S. companies. However, outsourcing is not without problems. A recent survey by Purchasing indicates that **20%** of the companies that outsource overseas use a consultant. Suppose **15** companies that outsource overseas are randomly selected.

a. What is the probability that **exactly 5** companies that outsource overseas use a consultant?

b. What is the probability that **none** of the companies that outsource overseas use a consultant?

c. What is the probability that **less than 10** companies that outsource overseas use a consultant?

d. What is the probability that **10 or fewer(at most 10)** of the companies that outsource overseas use a consultant?

e. What is the probability that **more than 6** companies that outsource overseas use a consultant?

f. What is the probability that **6 or more(at least 6)** companies that outsource overseas use a consultant?

g. What is the probability that **between 4 and 7 (not inclusive)** companies that outsource overseas use a consultant?

h. What is the probability that **between 4 and 7 (inclusive)** companies that outsource overseas use a consultant?

i. How many companies that outsource overseas are randomly selected if it has a probability of **0.25(25th percentile)** of total no. of **15** companies.

j. Find **8** random values from a sample of **15** with probability of **0.20**

k. Construct a **graph** for this binomial distribution.

Solution:-

a. **dbinom(x=5, size=15, p=0.20)** #density function is used to calculate the probability of single value
Output:- `0.1031823`

b. **dbinom(x=0, size=15, p=0.20)** #density function is used to calculate the probability of single value
Output:- `0.03518437`

c. **pbinom(9, size=15, p=0.20)** #probability function is used to calculate the cumulative probability from 0 to 9 (x < 10)
Output:- `0.9998868`

d. **pbinom(10, size=15, p=0.20)** #probability function is used to calculate the cumulative probability from 0 to 10 (x ≤ 10)
Output:- `0.9999875`

e. **pbinom(6, size=15, p=0.20, lower.tail = FALSE)** #probability function is used to calculate the cumulative probability from 7 to 15 (x > 6)
#lower.tail:- logical; if TRUE (default), probabilities are $P[X \le x]$, otherwise, $P[X > x]$.
Alternatively:- **1 − pbinom(6, size=15, p=0.20)**
Output:- `0.01805881`

f. **pbinom(5, size=15, p=0.20, lower.tail = FALSE)** #probability function is used to calculate the cumulative probability from 6 to 15 (x ≥ 6)
#lower.tail:- logical; if TRUE (default), probabilities are $P[X \le x]$, otherwise, $P[X > x]$.
Alternatively:- **1 − pbinom(5, size=15, p=0.20)**
Output:- `0.06105143`

g. **sum(dbinom(5:6, size=15, p=0.20))** #sum of density function is used to calculate the probability between 4 and 7 (not inclusive) i.e., p(4 < x < 7)
Output:- `0.1461749`

h. **sum(dbinom(4:7, size=15, p=0.20))** #sum of density function is used to calculate the probability between 4 and 7 (inclusive) i.e., p(4 ≤ x ≤ 7)
Output:- `0.3475981`

i. **qbinom(0.25, size = 15, p=0.20)** # The quantile is defined as the smallest value x such that $F(x) \ge p$, where F is the distribution function.
Output:- `2`

j. **rbinom(8, size=15, p=0.20)** #This function generates required number of random values of given probability from a given sample.

Output:- 5 3 6 5 1 2 3 4 (it generates randomly any numbers less than size, the output is different each & every time)

k. Graph:-

x = seq(0, 15, by=1) #Create a sample from 0 to 15 numbers which are incremented by 1.

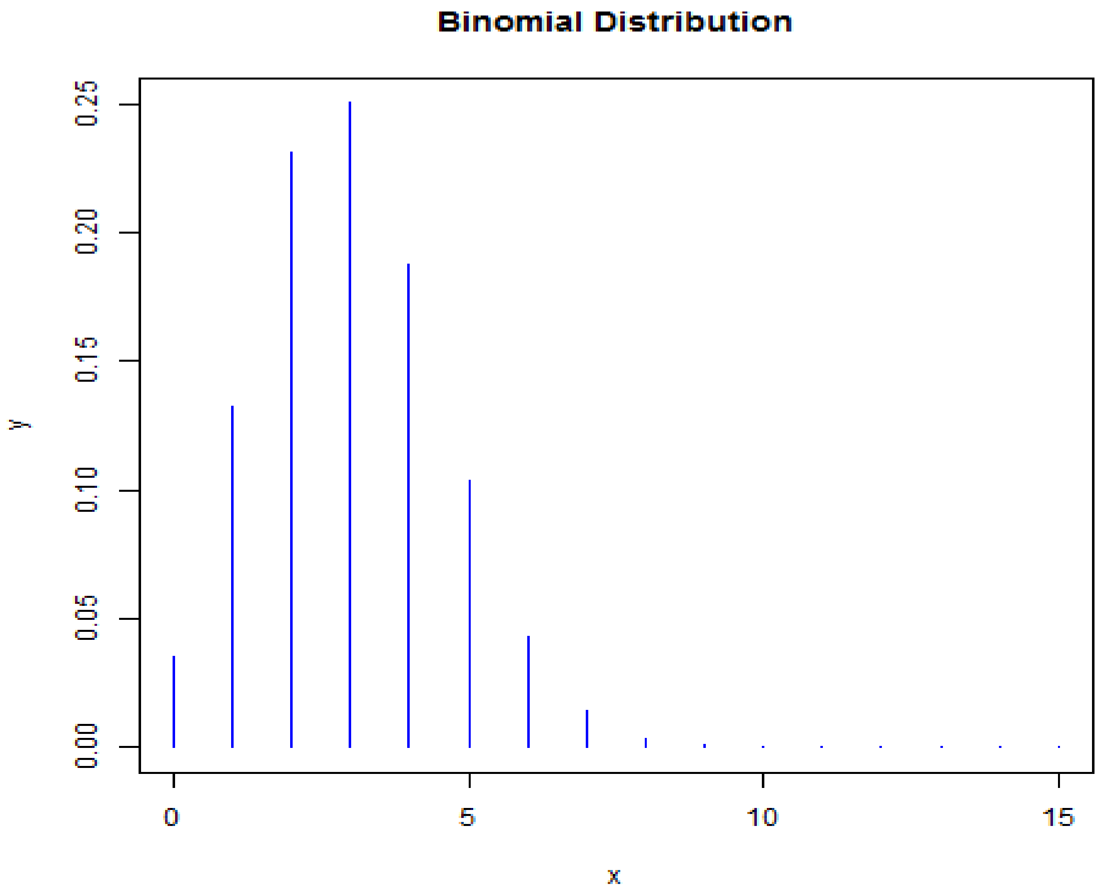
y = dbinom(x, size=15, p=0.20) #Create the binomial distribution.

png(file = "dbinom.png") #Give the chart file a name.

plot(x, y, main ="Binomial Distribution", col="blue", type="h") #Plot the graph for this sample.

dev.off() #Save the file.

Output:-



Poisson Distribution

The **Poisson distribution** 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:-

Probability formula:- $P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

Syntax:-

dpois(x, lambda, log = FALSE)
ppois(x, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)

Value:-

dpois() gives the (log) density
ppois() gives the (log) distribution function
qpois() gives the quantile function
rpois() generates random deviates.

Problem:-

The **average** number of annual trips per family to amusement parks in the United States is **Poisson** distributed, with a **mean** of **0.6 trips per year**. What is the probability of randomly selecting an American family and finding the following?

a. The family did **not** make a trip to an amusement park **last year**.
b. The family took **exactly 1** trip to an amusement park **last year**.
c. The family took **less than 3** trips to amusement parks **last year**.
d. The family took **3 or fewer(at most 3)** trips to amusement parks **last year**.
e. The family took **more than 2** trips to amusement parks **last year**.
f. The family took **2 or more(at least 2)** trips to amusement parks **last year**.
g. The family took **between 1 and 5 (not inclusive)** trips to amusement parks **last year**.
h. The family took **between 1 and 5 (inclusive)** trips to amusement parks **last year**.
i. How many number of annual trips per family to amusement parks in the United States are organized if it has a probability of **0.65(65th percentile)**?
j. Find **6** random values for this given poisson distribution.
k. Construct a **graph** for this poisson distribution.

Solution:-

a. **dpois(x=0, lambda=0.6)** #density function is used to calculate the probability of single value
Output:- `0.5488116`

b. **dpois(x=1, lambda=0.6)** #density function is used to calculate the probability of single value
Output:- `0.329287`

c. **ppois(2, lambda=0.6)** #probability function is used to calculate the cumulative probability from 0 to 2 ($x < 3$)
Output:- `0.9768847`

d. **ppois(3, lambda=0.6)** #probability function is used to calculate the cumulative probability from 0 to 3 ($x \leq 3$)
Output:- `0.9966419`

e. **ppois(2, lambda=0.6, lower.tail = FALSE)** #probability function is used to calculate the cumulative probability from 3 to ……… ($x > 2$)
#lower.tail:- logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.
Alternatively:- **1 − ppois(2, lambda=0.6)**
Output:- `0.02311529`

f. **ppois(1, lambda=0.6, lower.tail = FALSE)** #probability function is used to calculate the cumulative probability from 2 to ……. ($x \geq 2$)
#lower.tail:- logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.
Alternatively:- **1 − ppois(1, lambda=0.6)**
Output:- `0.1219014`

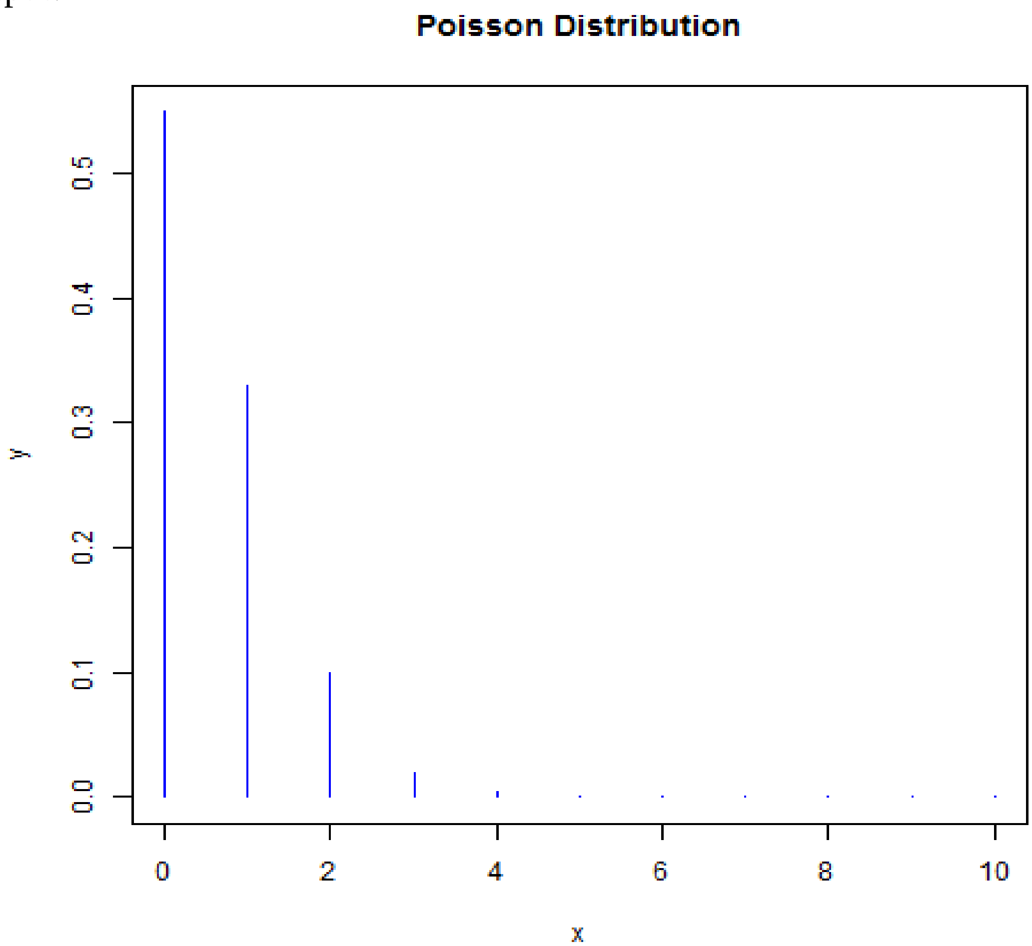
g. **sum(dpois(2:4, lambda=0.6))** #sum of density function is used to calculate the probability between 1 and 5 (not inclusive) i.e., $p(1 < x < 5)$
Output:- `0.1215069`

h. **sum(dpois(1:5, lambda=0.6))** #sum of density function is used to calculate the probability between 1 and 5 (inclusive) i.e., $p(1 \leq x \leq 5)$
Output:- `0.4511495`

i. **qpois(0.65, lambda=0.6)** **#** The quantile is defined as the smallest value x such that $F(x) \geq p$, where F is the distribution function.
Output:- `1`

j. **rpois(6, lambda=0.6)** #This function generates required number of random val ues of given probability from a given sample.
Output:- `1 0 0 0 1 3` (it generates randomly any numbers, the output is different each & every time)

k. **Graph:-**
x = seq(0, 10, by=1) #Create a sample from 0 to 10 numbers which are incremented by 1.
y = dpois(x, lambda=0.6) #Create the poisson distribution.
png(file = "dpois.png") #Give the chart file a name.
plot(x, y, main ="Poisson Distribution", col="blue", type="h") #Plot the

graph for this sample.
dev.off() #Save the file.
Output:-


Normal Distribution

The **normal distribution** is defined by the following probability density function, where μ is the population mean and σ^2 is the variance.

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

Syntax:-

dnorm(x, mean, sd, log = FALSE)
pnorm(x, mean, sd, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean, sd, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean, sd)

Value:-

dnorm() gives the density
pnorm() gives the distribution function
qnorm() gives the quantile function
rnorm() generates random deviates.

Problem:-

Tompkins Associates reports that the **mean** clear height for a Class A warehouse in the United States is **22** feet. Suppose clear heights are **normally distributed** and that the **standard deviation** is **4** feet. A Class A warehouse in the United States is randomly selected.

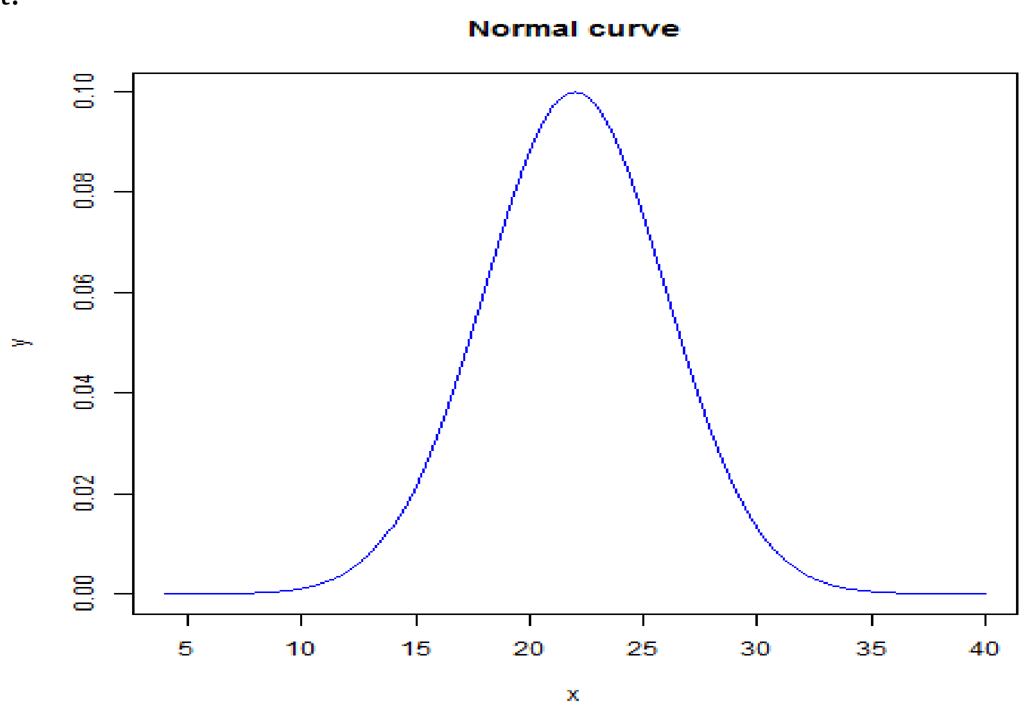
a. What is the probability that the clear height is **exactly 14** feet?
b. What is the probability that the clear height is **less than 18(18 or fewer, at most 18)** feet?
c. What is the probability that the clear height is **greater than 26(26 or more, at least 26)** feet?
d. What is the probability that the clear height is **between 14** and **30** feet?
e. How many number of Class A warehouse in the United States are reported if it has a probability of **0.45(45th percentile)**?
f. Find **4** random values for this given normal distribution.
g. Construct a **graph** for this normal distribution.

Solution:-

a. dnorm(x=14, mean=22, sd=4) #density function is used to calculate the probability of single value
Output:- `0.01349774`

b. pnorm(18, mean=22, sd=4) #probability function is used to calculate the cumulative probability from 18 to less values……. ($x < 18$, $x \le 18$)
Output:- `0.1586553`

c. **pnorm(26, mean=22, sd=4, lower.tail=FALSE)** #probability function is used to calculate the cumulative probability from 26 to more values...($x > 26$, $x \geq 26$) #lower.tail:- logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.
   Alternatively:- **1 − pnorm(26, mean=22, sd=4)**
   Output:- `0.1586553`
d. **pnorm(30, mean=22, sd=4) – pnorm(14, mean=22, sd=4)** #sum of density function is used to calculate the probability between 14 and 30. i.e., $p(14 \leq x \leq 30)$
   Output:- `0.9544997`
e. **qnorm(0.45, mean=22, sd=4)** # The quantile is defined as the smallest value *x* such that $F(x) \geq p$, where *F* is the distribution function.
   Output:- `21.49735`
f. **rnorm(4, mean=22, sd=4)** #This function generates required number of random values of given probability from a given sample.
   Output:- `14.79942 20.64600 15.17985 24.14162` (it generates randomly any numbers, the output is different each & every time)
g. Graph:-
   x = seq(4, 40, by=0.1) # Create a sequence of numbers between 10 and 35 incrementing by 0.1
   y = dnorm(x, mean=22, sd=4) #Create the normal distribution.
   png(file = "dnorm.png") #Give the chart file a name.
   plot(x, y, main ="Normal curve", col="blue", type = "l") #Plot the graph for this sample.
   dev.off() #Save the file.
   Output:-



Exponential Distribution

The **exponential distribution** describes the arrival time of a randomly recurring independent event sequence. If μ is the mean waiting time for the next event recurrence, its probability density function is:-

Density formula:- $$f(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu} & when\ x \geq 0 \\ 0 & when\ x < 0 \end{cases}$$

Probability formula:- $P(x \leq x_0) = 1 - e^{-x_0/\mu}$

Syntax:-

```
dexp(x, rate, log = FALSE)
pexp(x, rate, lower.tail = TRUE, log.p = FALSE)
qexp(p, rate, lower.tail = TRUE, log.p = FALSE)
rexp(n, rate)
```

Value:-
dexp() gives the density
pexp() gives the distribution function
qexp() gives the quantile function
rexp() generates random deviates.

Problem:-
The time between arrivals of vehicles at a particular intersection follows an exponential probability distribution with a **mean** of **12 seconds**.

a. What is the probability that the arrival time is **exactly 10** seconds?
b. What is the probability that the arrival time between vehicles is **less than 12 (12 or fewer, at most 12) seconds**?
c. What is the probability that the arrival time between vehicles is **greater than 15(15 or more, at least 15) seconds**?
d. What is the probability of **between 10** and **15 seconds** between vehicle arrivals?
e. How many arrivals of vehicles at a particular intersection if it has a probability of **0.65(65^{th} percentile)**?
f. Find **5** random values for this given exponential distribution.
g. Construct a **graph** for this exponential distribution.

Solution:-

a. **dexp(x=10, rate = 1/12)** #density function is used to calculate the probability of single value
Output:- `0.03621652`

b. **pexp(12, rate = 1/12)** #probability function is used to calculate the cumulative probability from 12 to less values……. ($x < 12$, $x \leq 12$)
Output:- `0.6321206`

c. **pexp(15, rate = 1/12, lower.tail=FALSE)** #probability function is used to calculate the cumulative probability from 15 to more values…...($x > 15$, $x \geq 15$)
#lower.tail:- logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.
Alternatively:- **1 − pexp(15, rate = 1/12)**
Output:- `0.2865048`

d. **pexp(15, rate = 1/12) − pexp(10, rate = 1/12)** #difference between two probability function is used to calculate the probability between 10 and 15. i.e., $p(10 \leq x \leq 15)$
Output:- `0.1480934`

e. **qexp(0.65, rate = 1/12)** # The quantile is defined as the smallest value *x* such that $F(x) \geq p$, where *F* is the distribution function.
Output:- `12.59787`

f. **rexp(5, rate = 1/12)** #This function generates required number of random values of given probability from a given sample.
Output:- `5.5241299 10.8964906 27.1702649  0.4774654  1.6370954` (it generates randomly any numbers, the output is different each & every time)

g. Graph:-
x = seq(0, 20, by=1) # Create a sequence of numbers from 0 to 20 incrementing by 1
y = dexp(x, rate = 1/12) #Create the exponential distribution.
png(file = "dexp.png") #Give the chart file a name.
plot(x, y, main ="Exponential Curve", col="blue", type = "l") #Plot the graph for this sample.
dev.off() #Save the file.
Output:-