ASSIGNMENT 1
Probability and Statistics

Name: Pranay Malhotra
Roll Number: 102203521
Batch: 3CO35

Question 1:
Create a vector c = [5,10,15,20,25,30] and write a program which returns the maximum and minimum of this vector.

Output:
```
> c <-  c(5,10,15,20,25,30)
> min(c)
[1] 5
> max(c)
[1] 30
```

Question 2:
Write a program in R to find factorial of a number by taking input from user. Please print error message if the input number is negative.

Output:
```
> myFact <- function(x){
+   if(x<0){
+     print("Error... Cant compute for negative values")
+     return(NULL)
+   }
+
+   if(x==0 | x==1){
+     return(1)
+   } else{
+     return(x * myFact(x-1))
+   }
+ }
>
> myFact(6)
[1] 720
```

Question 3:

Write a program to write first n terms of a Fibonacci sequence. You may take n as an input from the user.

Output:

```
> myFib <- function(x){
+   if(x==1){
+     return(0)
+   }else if(x==2){
+     return(1)
+   }
+
+   return(myFib(x-1) + myFib(x-2))
+ }
>
> myFib(5)
[1] 3
```

Question 4:

Write an R program to make a simple calculator which can add, subtract, multiply and divide.

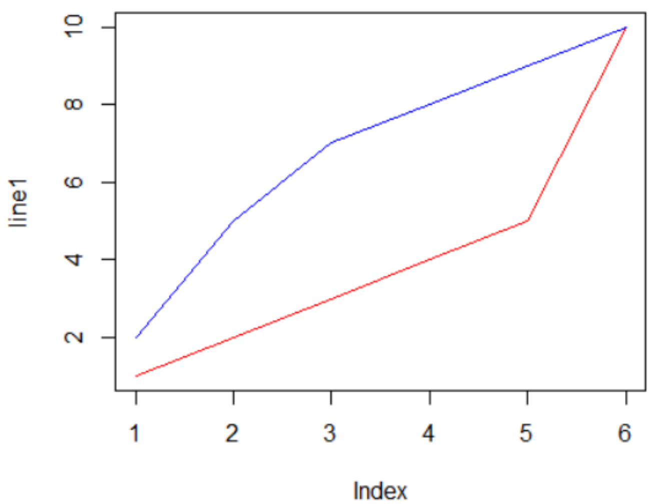
Output:

```
> myAdd <- function(a, b){
+   return(a+b)
+ }
> mySub <- function(a, b){
+   return(a-b)
+ }
> myMul <- function(a, b){
+   return(a*b)
+ }
> myDiv <- function(a, b){
+   return(a/b)
+ }
> myAdd(2, 5)
[1] 7
> mySub(10, 8)
[1] 2
> myMul(2, 6)
[1] 12
> myDiv(10, 2)
[1] 5
```

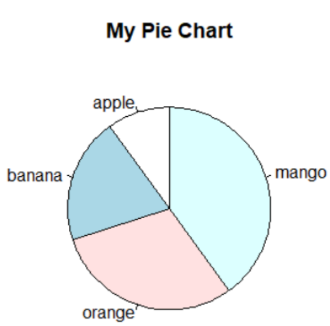
Question 5:

Explore plot, pie, barplot etc. (the plotting options) which are built-in functions in R.

Output:



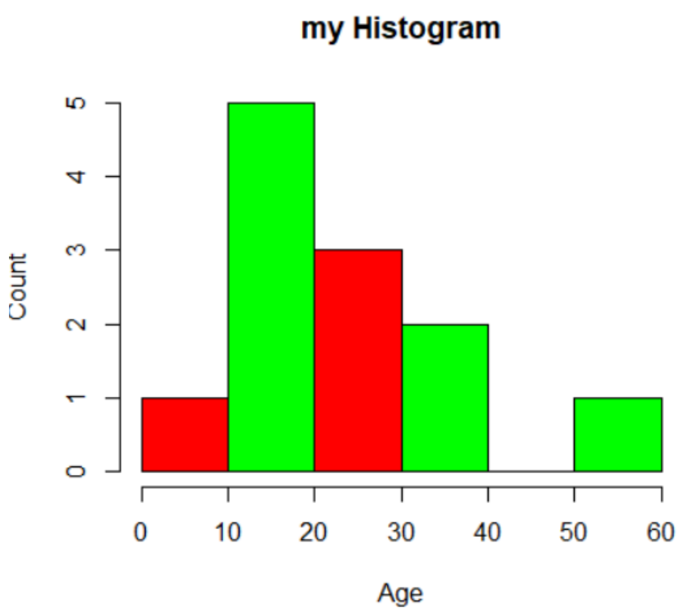

```
x <- c(1, 2.5, 3)
y <- c(10, 20, 22)
plot(x, y)
plot(x, y, type = 'l', main = "My first graph", xlab = "age", ylab = "height", col="red")
plot(x, y, cex=1.5, pch=2)

plot(x, y, type = "l", lwd=3, lty=2, col="blue")

line1 <- c(1,2,3,4,5,10)
line2 <- c(2,5,7,8,9,10)

plot(line1, type="l", col="red")
lines(line2, type="l", col="blue")
```



PIE :

```
x <- c(10, 20, 30, 40)
label <- c("apple", "banana", "orange", "mango")
pie(x, main = "My Pie Chart", labels = label)
pie(x, main = "My Pie Chart", labels = label, init.angle = 90)
```

Histo :




```
v <- c(19, 23, 11, 5, 16, 21, 32, 14, 19, 27, 39, 55)
sort(v)
1]  5 11 14 16 19 19 21 23 27 32 39 55
hist(v, main = "my Histogram", xlab = "Age", ylab = "Count", col = c("red", "green"))
```

ASSIGNMENT 2
Probability and Statistics

Name: Pranay Malhotra
Roll Number: 102203521
Batch: 3CO35

Question 1:

(a) Suppose there is a chest of coins with 20 gold, 30 silver and 50 bronze coins. You randomly draw 10 coins from this chest. Write an R code which will give us the sample space for this experiment. (use of sample(): an in-built function in R)

OUTPUT:

```
gold <- rep("Gold", 20)
silver <- rep("Silver", 30)
bronze <- rep("Bronze", 50)
total <- c(gold, silver, bronze)

draw_sample <- sample(total, size=10, replace = FALSE)
draw_sample
[1] "Gold"   "Silver" "Silver" "Silver" "Gold"   "Bronze" "Silver" "Bronze" "Silver" "Silver"

draw_sample <- sample(total, size=10, replace = TRUE)
draw_sample
[1] "Gold"   "Gold"   "Gold"   "Bronze" "Silver" "Bronze" "Gold"   "Bronze" "Bronze" "Bronze"
```

(b) In a surgical procedure, the chances of success and failure are 90% and 10% respectively. Generate a sample space for the next 10 surgical procedures performed. (use of prob(): an in-built function in R)

Output:

```
operation_sample <- sample(outcomes, size=10, replace = TRUE, prob = probabilities)
operation_sample
1] "Success" "Success" "Success" "Success" "Success" "Success" "Success" "Success" "Success" "Success"
```

Question 2:

A room has n people, and each has an equal chance of being born on any of the 365 days of the year. (For simplicity, we'll ignore leap years). What is the probability

that two people in the room have the same birthday?
(a) Use an R simulation to estimate this for various n.
Output:

(b) Find the smallest value of n for which the probability of a match is greater than .5.

Output:

Question 3:

Write an R function for computing conditional probability. Call this function to do the following problem:
suppose the probability of the weather being cloudy is 40%. Also suppose the probability of rain on a given day is 20% and that the probability of clouds on a rainy day is 85%. If it's cloudy outside on a given day, what is the probability that it will rain that day?

Output:

```
> P_Cloudy_given_Rain <- 0.85
> P_Rain <- 0.20
> P_Cloudy <- 0.40
> P_Rain_given_Cloudy <- myFun2(P_Cloudy, P_Rain, P_Cloudy_given_Rain)
> cat("Probability is: ", P_Rain_given_Cloudy)
Probability is:  0.425
```

Question 4:

The iris dataset is a built-in dataset in R that contains measurements on 4 different attributes (in centimeters) for 150 flowers from 3 different species. Load this dataset and do the following:
(a) Print first few rows of this dataset.
(b) Find the structure of this dataset.
(c) Find the range of the data regarding the sepal length of flowers.

(d) Find the mean of the sepal length.
(e) Find the median of the sepal length.
(f) Find the first and the third quartiles and hence the interquartile range.
(g) Find the standard deviation and variance.
(h) Try doing the above exercises for sepal.width, petal.length and petal.width.
(i) Use the built-in function summary on the dataset Iris.

Output:

```
> data(iris)
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
>
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
>
> length(iris$Sepal.Length)
[1] 150
> range(iris$Sepal.Length)
[1] 4.3 7.9
>
> mean(iris$Sepal.Length)
[1] 5.843333
>
> median(iris$Sepal.Length)
[1] 5.8
>
> q <- quantile(iris$Sepal.Length)
> q
  0%  25%  50%  75% 100% 
 4.3  5.1  5.8  6.4  7.9 
> q1 <- q[2]
> q3 <- q[4]
>
> iqr <- IQR(iris$Sepal.Length)
> cat("First quartile: ", q1)
First quartile:  5.1> cat("Third quartile: ", q3)
Third quartile:  6.4> cat("Inter Quartile Range: ", iqr)
Inter Quartile Range:  1.3>
```

```
> v <- var(iris$Sepal.Length)
> v
[1] 0.6856935
>
> s <- sd(iris$Sepal.Length)
> s
[1] 0.8280661
>
> summary(iris)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species  
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa    :50  
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50  
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50  
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199                  
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800                  
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500                  
>
```

Question 4:R does not have a standard in-built function to calculate mode. So we create a user function to calculate mode of a data set in R. This function takes the vector as input and gives the mode value as output.

Output:

```
> myMode <- function(x){
+   uniquex <- unique(x)
+   value_counts <- tabulate(match(x, uniquex))
+   mode_value <- uniquex[which(value_counts == max(value_counts))]
+
+   return(mode_value)
+ }
>
> x <- c(1, 1, 1, 2, 7, 9, 3, 4, 5)
> myMode(x)
[1] 1
```

ASSIGNMENT 3
Probability and Statistics

Name: Pranay Malhotra

Roll Number: 102203521

Batch: 3CO35

Question 1:

Roll 12 dice simultaneously, and let X denotes the number of 6's that appear. Calculate the probability of getting 7, 8 or 9, 6's using R. (Try using the function pbinom; If we set S = {get a 6 on one roll}, P(S) = 1/6 and the rolls constitute Bernoulli trials; thus X ~ binom(size=12, prob=1/6) and we are looking for P(7 ≤ X ≤ 9).

OUTPUT:

```
> diff(pbinom(c(6,9), size=12, prob=1/6, lower.tail = TRUE))
[1] 0.001291758
```

Question 2:

Assume that the test scores of a college entrance exam fits a normal distribution. Furthermore, the mean test score is 72, and the standard deviation is 15.2. What is the percentage of students scoring 84 or more in the exam?

OUTPUT:

```
> pnorm(84, mean=72, sd= 15.2, lower.tail=FALSE)
[1] 0.2149176
```

Question 3:

On the average, five cars arrive at a particular car wash every hour. Let X count the number of cars that arrive from 10AM to 11AM, then X ~Poisson(λ = 5). What is probability that no car arrives during this time. Next, suppose the car wash above is in operation from 8AM to 6PM, and we let Y be the number of customers that appear in this period. Since this period covers a total of 10 hours, we get that Y ~ Poisson(λ = 5×10 = 50). What is the probability that there are between 48 and 50 customers, inclusive?

OUTPUT:

```
> lambda_X <- 5
> prob_X_0 <- dpois(0, lambda_X)
> cat("Probability that no car arrives from 10 AM to 11 AM:", prob_X_0, "\n")
Probability that no car arrives from 10 AM to 11 AM: 0.006737947
>
> # Part 2: Probability that there are between 48 and 50 customers inclusive
> lambda_Y <- 50
> prob_Y_48_50 <- ppois(50, lambda_Y) - ppois(47, lambda_Y)
> cat("Probability that there are between 48 and 50 customers:", prob_Y_48_50, "\n")
Probability that there are between 48 and 50 customers: 0.1678485
```

Question 4:

Suppose in a certain shipment of 250 Pentium processors there are 17 defective processors. A quality control consultant randomly collects 5 processors for inspection to determine whether or not they are defective. Let X denote the number of defectives in the sample. Find the probability of exactly 3 defectives in the sample, that is, find P(X = 3).

OUTPUT:

```
> N <- 250    # Total number of processors
> K <- 17     # Number of defective processors
> n <- 5      # Sample size
> X <- 3      # Number of defectives we want in the sample
>
> # Probability of exactly 3 defectives in the sample
> prob_X_3 <- dhyper(X, K, N-K, n)
> prob_X_3
[1] 0.002351153
```

Question 5:

A recent national study showed that approximately 44.7% of college students have used Wikipedia as a source in at least one of their term papers. Let X equal the number of students in a random sample of size n = 31 who have used Wikipedia as a source. (a) How is X distributed? (b) Sketch the probability mass function. (c) Sketch the cumulative distribution function. (d) Find mean, variance and standard deviation of X.

OUTPUT:

```
> n <- 31
> p <- 0.447
>
> # Possible values of X
> x_values <- 0:n
>
> # PMF values
> pmf_values <- dbinom(x_values, n, p)
>
> # Plot the PMF
> plot(x_values, pmf_values, type="h", lwd=2, col="blue",
+      main="Probability Mass Function of X",
+      xlab="Number of Students Using Wikipedia (X)",
+      ylab="Probability")
> points(x_values, pmf_values, pch=16, col="red")
>
> # CDF values
> cdf_values <- pbinom(x_values, n, p)
>
> # Plot the CDF
> plot(x_values, cdf_values, type="s", lwd=2, col="blue",
+      main="Cumulative Distribution Function of X",
+      xlab="Number of Students Using Wikipedia (X)",
+      ylab="Cumulative Probability")
> points(x_values, cdf_values, pch=16, col="red")
>
> # Mean
> mean_X <- n * p
>
> # Variance
> var_X <- n * p * (1 - p)
>
> # Standard Deviation
> sd_X <- sqrt(var_X)
>
> cat("Mean:", mean_X, "\n")
Mean: 13.857
> cat("Variance:", var_X, "\n")
Variance: 7.662921
> cat("Standard Deviation:", sd_X, "\n")
Standard Deviation: 2.768198
```

ASSIGNMENT 4
Probability and Statistics

Name: Pranay Malhotra

Roll Number: 102203521

Batch: 3CO35

Question 1:

The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given

x	0	1	2	3	4
$p(x)$	0.41	0.37	0.16	0.05	0.01

Find the average number of imperfections per 10 meters of this fabric. (Try functions sum( ), weighted.mean( ), c(a %*% b) to find expected value/mean

OUTPUT:

```
> x <- c(0, 1, 2, 3, 4)
> px <- c(0.41, 0.37, 0.16, 0.05, 0.01)
> expval <- sum(x * px)
> #OR
> expval <- weighted.mean(x, px)
> #OR
> expval <- c(x %*% px)
> cat("avg value: ", expval)
avg value:  0.88
```

Question 2:

The time T, in days, required for the completion of a contracted project is a random variable with probability density function f(t) = 0.1 e(-0.1t) for t > 0 and 0 otherwise. Find the expected value of T. Use function integrate( ) to find the expected value of continuous random variable T.

OUTPUT:

```
> f <- function(t){
+   t*0.1*exp(-0.1*t)
+ }
>
> expval <- integrate(f, lower = 0, upper = Inf)
> print(expval$value)
[1] 10
```

Question 3:

A bookstore purchases three copies of a book at $6.00 each and sells them for $12.00 each. Unsold copies are returned for $2.00 each. Let X = {number of copies sold} and Y = {net revenue}. If the probability mass function of X is

x	0	1	2	3
$p(x)$	0.1	0.2	0.2	0.5

Find the expected value of Y.

OUTPUT:

```
> x <- c(0, 1, 2, 3)
> px <- c(0.1, 0.2, 0.2, 0.5)
> y <- 12*x - 18 + (3-x)*2
>
> expval <- sum(y* px)
> print(expval)
[1] 9
```

Question 4:

Find the first and second moments about the origin of the random variable X with probability density function f(x) = 0.5e -|x| , 1 < x < 10 and 0 otherwise. Further use the results to find Mean and Variance. (kth moment = E(Xk), Mean = first moment and Variance = second moment – Mean2 .

OUTPUT:

```
> f1 <- function(x){
+   x * 0.5 * exp(-abs(x))
+ }
>
> m1 <- integrate(f1, lower = 1, upper = 10)
> cat("M1: ", m1$value)
M1:  0.3676297>
> f2 <- function(x){
+   (x^2) * 0.5 * exp(-abs(x))
+ }
>
> m2 <- integrate(f2, lower = 1, upper = 10)
> cat("M2: ", m2$value)
M2:  0.9169292> cat("mean:", m1$value)
mean: 0.3676297> cat("variance:", m2$value - (m1$value ^ 2))
variance: 0.7817776
>
```

Question 5:

Let X be a geometric random variable with probability

$$f(x) = \frac{3}{4}\left(\frac{1}{4}\right)^{x-1}, x = 1,2,3,\ldots$$

Write a function to find the probability distribution of the random variable Y = X2 and find probability of Y for X = 3. Further, use it to find the expected value and variance of Y for X = 1,2,3,4,5.

OUTPUT:

```
> x <- c(1, 2, 3, 4, 5)
> y <- x^2
> fy <- function(y){
+     (3/4) * ((1/4) ^ (sqrt(y)-1))
+ }
> py <- fy(y)
> m1 <- sum(y * py)
> m2 <- sum(y*y*py)
> var <- m2 - (m1^2)
> print(m1)
[1] 2.182617
> print(var)
[1] 7.614112
```