INSTITUTE:

VITAP

APPLIED STATISTICS

LAB REPORT

ASSIGNED BY

Dr. Peeyush Singh sir

MAJJIGA JASWANTH 20BCD7171

LAB REPORT

ASSIGNMENT-1

QUESTION-1:

Exercise 2.1. The weights of five people before and after a diet programme are given in the table.

Before	78	72	78	79	105
After	67	65	79	70	93

Read the 'before' and 'after' values into two different vectors called `before` and `after`. Use R to evaluate the amount of weight lost for each participant. What is the average amount of weight lost?

```
> timestamp()
##------ Fri Jul 16 19:57:53 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> before<- c(78,72,78,79,105)
> before
[1]  78  72  78  79 105
> after <- c(67,65,79,70,93)
> after
[1] 67 65 79 70 93
> diff <- c(before-after)
> diff
[1] 11  7 -1  9 12
> weight <- mean(diff)
> weight
[1] 7.6
```

Question-2:

***Exercise 2.2.** How would you write a function equivalent to `sum((x - mean(x))^2)` in a language like C or Java?

Some useful vectors can be created quickly with R. The colon operator is used to generate integer sequences

```
> timestamp()
##------ Fri Jul 16 20:22:27 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x <- c(1,2,3,4,5)
> x
[1] 1 2 3 4 5
> sum <- mean((x-mean(x))^2)
> sum
[1] 2
>  sum((x-mean(x))^2)
[1] 10
```

Question-3:

Exercise 2.3. Create the following vectors in R using seq() and rep().

(i) $1, 1.5, 2, 2.5, \ldots, 12$

(ii) $1, 8, 27, 64, \ldots, 1000.$

(iii) $1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \ldots, -\frac{1}{100}.$

(iv) $1, 0, 3, 0, 5, 0, 7, \ldots, 0, 49.$

(v) $1, 3, 6, 10, 15, \ldots, \sum_{i=1}^{n} i, \ldots, 210$ [look up ?cumsum].

(vi) * $1, 2, 2, 3, 3, 3, 4, \ldots, 9, \underbrace{10, \ldots, 10}_{10 \text{ times}}$. [Hint: type ?seq, and read about the times argument.]

```
> timestamp()
##------ Fri Jul 16 20:33:24 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> seq(from=1,to=12,by=0.5)
 [1]  1.0  1.5  2.0  2.5  3.0  3.5  4.0  4.5  5.0  5.5  6.0  6.5  7.0  7.5  8.0
[16]  8.5  9.0  9.5 10.0 10.5 11.0 11.5 12.0
> seq(from=1,to=10,by=1)
 [1]  1  2  3  4  5  6  7  8  9 10
> seq(from=1,to=10,by=1)^rep(3)
 [1]    1    8   27   64  125  216  343  512  729 1000
> (-1)^rep(0:1,length.out=100)/seq(1,100,by=1)
  [1]  1.00000000 -0.50000000  0.33333333 -0.25000000  0.20000000 -0.16666667
  [7]  0.14285714 -0.12500000  0.11111111 -0.10000000  0.09090909 -0.08333333
 [13]  0.07692308 -0.07142857  0.06666667 -0.06250000  0.05882353 -0.05555556
 [19]  0.05263158 -0.05000000  0.04761905 -0.04545455  0.04347826 -0.04166667
 [25]  0.04000000 -0.03846154  0.03703704 -0.03571429  0.03448276 -0.03333333
 [31]  0.03225806 -0.03125000  0.03030303 -0.02941176  0.02857143 -0.02777778
 [37]  0.02702703 -0.02631579  0.02564103 -0.02500000  0.02439024 -0.02380952
 [43]  0.02325581 -0.02272727  0.02222222 -0.02173913  0.02127660 -0.02083333
 [49]  0.02040816 -0.02000000  0.01960784 -0.01923077  0.01886792 -0.01851852
 [55]  0.01818182 -0.01785714  0.01754386 -0.01724138  0.01694915 -0.01666667
 [61]  0.01639344 -0.01612903  0.01587302 -0.01562500  0.01538462 -0.01515152
 [67]  0.01492537 -0.01470588  0.01449275 -0.01428571  0.01408451 -0.01388889
 [73]  0.01369863 -0.01351351  0.01333333 -0.01315789  0.01298701 -0.01282051
 [79]  0.01265823 -0.01250000  0.01234568 -0.01219512  0.01204819 -0.01190476
 [85]  0.01176471 -0.01162791  0.01149425 -0.01136364  0.01123596 -0.01111111
 [91]  0.01098901 -0.01086957  0.01075269 -0.01063830  0.01052632 -0.01041667
 [97]  0.01030928 -0.01020408  0.01010101 -0.01000000
> seq(from=1,to=49,by=1)*rep(1:0,length.out=49)
 [1]  1  0  3  0  5  0  7  0  9  0 11  0 13  0 15  0 17  0 19  0 21  0 23  0 25  0
[27] 27  0 29  0 31  0 33  0 35  0 37  0 39  0 41  0 43  0 45  0 47  0 49
> cumsum(seq(from=1,to=20,by=1))
 [1]   1   3   6  10  15  21  28  36  45  55  66  78  91 105 120 136 153 171 190
[20] 210
> rep(seq(from=1,to=10,by=1),seq(from=1,to=10,by=1))
 [1]  1  2  2  3  3  3  4  4  4  4  5  5  5  5  5  6  6  6  6  6  6  7  7  7  7  7
[27]  7  7  8  8  8  8  8  8  8  8  9  9  9  9  9  9  9  9  9 10 10 10 10 10 10 10
[53] 10 10 10
```

Question-4:

Exercise 2.4. The ith term in the Taylor expansion of $\log(1+x)$ is $(-1)^{i+1}x^i/i$. Create a vector containing the first 100 terms for $x = 0.5$. *[Write out the first few entries by hand if that helps.]*

Let

$$r_n(x) = \log(1+x) - \sum_{i=1}^{n} \frac{(-1)^{i+1}x^i}{i}.$$

Evaluate $r_n(1)$ for $n = 10, 100, 1000, \ldots, 10^6$.

```
> timestamp()
##------ Fri Jul 16 20:39:11 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> i=1:100
> x=0.5
> log=(((-1)^(i+1))*((x^i)/i))
> log
  [1]  5.000000e-01 -1.250000e-01  4.166667e-02 -1.562500e-02  6.250000e-03
  [6] -2.604167e-03  1.116071e-03 -4.882812e-04  2.170139e-04 -9.765625e-05
 [11]  4.438920e-05 -2.034505e-05  9.390024e-06 -4.359654e-06  2.034505e-06
 [16] -9.536743e-07  4.487879e-07 -2.119276e-07  1.003868e-07 -4.768372e-08
 [21]  2.270653e-08 -1.083721e-08  5.183013e-09 -2.483527e-09  1.192093e-09
 [26] -5.731216e-10  2.759474e-10 -1.330461e-10  6.422914e-11 -3.104409e-11
 [31]  1.502133e-11 -7.275958e-12  3.527737e-12 -1.711990e-12  8.315380e-13
 [36] -4.042199e-13  1.966475e-13 -9.573628e-14  4.664075e-14 -2.273737e-14
 [41]  1.109140e-14 -5.413659e-15  2.643880e-15 -1.291896e-15  6.315935e-16
 [46] -3.089316e-16  1.511793e-16 -7.401487e-17  3.625218e-17 -1.776357e-17
 [51]  8.707632e-18 -4.270089e-18  2.094760e-18 -1.027984e-18  5.046468e-19
 [56] -2.478176e-19  1.217350e-19 -5.981805e-20  2.940209e-20 -1.445603e-20
 [61]  7.109522e-21 -3.497426e-21  1.720956e-21 -8.470329e-22  4.170008e-22
 [66] -2.053413e-22  1.011383e-22 -4.982547e-23  2.455168e-23 -1.210047e-23
 [71]  5.965021e-24 -2.941087e-24  1.450399e-24 -7.153994e-25  3.529304e-25
 [76] -1.741433e-25  8.594084e-26 -4.241952e-26  2.094128e-26 -1.033976e-26
 [81]  5.106053e-27 -2.521892e-27  1.245754e-27 -6.154618e-28  3.041105e-28
 [86] -1.502872e-28  7.427987e-29 -3.671789e-29  1.815266e-29 -8.975484e-30
 [91]  4.438426e-30 -2.195091e-30  1.085744e-30 -5.370968e-31  2.657216e-31
 [96] -1.314768e-31  6.506069e-32 -3.219840e-32  1.593658e-32 -7.888609e-33
```

For n=10

```
> i=1:100
> x=1
> log=(((-1)^(i+1))*((x^i)/i))
> log=sum(log)
> log
[1] 0.6881722
> n=1:10
> a=(((-1)^(n+1))*((x^n)/n))
> a
 [1]  1.0000000 -0.5000000  0.3333333 -0.2500000  0.2000000 -0.1666667  0.1428571
 [8] -0.1250000  0.1111111 -0.1000000
> a=sum(a)
>
> r=log-a
> r
[1] 0.04253726
```

For n=100, n=1000

```
> n=1:100
> a=(((-1)^(n+1))*((x^n)/n))
> a=sum(a)
> r=log-a
> r
[1] 0
> n=1:1000
>  a=(((-1)^(n+1))*((x^n)/n))
> a=sum(a)
> r=log-a
> r
[1] -0.004475251
```

For n=10000, n=100000

```
> n=1:10000
>  a=(((-1)^(n+1))*((x^n)/n))
> a=sum(a)
> r=log-a
> r
[1] -0.004925004
> n=1:100000
>  a=(((-1)^(n+1))*((x^n)/n))
> a=sum(a)
> r=log-a
> r
[1] -0.004970001
```

For n=1000000

```
> n=1:1000000
>  a=(((-1)^(n+1))*((x^n)/n))
> a=sum(a)
> r=log-a
> r
[1] -0.004974501
```

Assignment-2

Question-1:

Q1. Calculate the mean for the following distribution:

x	5	6	7	8	9
f	4	8	14	11	3

solution:

x	f	$f \cdot x$
5	4	20
6	8	48
7	14	98
8	11	88
9	3	27
	$\Sigma f = 40$	$\Sigma f x = 281$

$$\text{Mean} = \frac{\Sigma f x}{\Sigma f} = \frac{281}{4} = 7.025$$

```
> timestamp()
##------ Fri Jul 16 21:06:28 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x <- c(5,6,7,8,9)
> x
[1] 5 6 7 8 9
> f <- c(4,8,14,11,3)
> f
[1]  4  8 14 11  3
> m <- f*x
> m<- sum(m)
> m
[1] 281
>

> f <- sum(f)
> f
[1] 40
> z <- m/f
> z
[1] 7.025
> 
```

Question-2:

Q2. The ages of 40 students are given in the following table:

Age (in years)	12	13	14	15	16	17	18
Frequency	2	4	6	9	8	7	4

Find the arithmetic mean.
solution:

Age in years (x_i)	Frequency (f_i)	$f_i \cdot x_i$
12	2	24
13	4	52
14	6	84
15	9	135
16	8	128
17	7	119
18	4	72
Total (Σ)	40	614

$$\bar{x} = \frac{\Sigma f_i \cdot x_i}{\Sigma f_i} = \frac{614}{40} = 15.35$$

```
> timestamp()
##------ Fri Jul 16 21:09:39 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x <- c(12,13,14,15,16,17,18)
> x
[1] 12 13 14 15 16 17 18
> f <- c(2,4,6,9,8,7,4)
> f
[1] 2 4 6 9 8 7 4
> z<-f*x
> z
[1]  24  52  84 135 128 119  72
> m <- (sum(z)/sum(f))
> m
[1] 15.35
>
```

Question-3:

Q3. Find the mean of the following data:

x	19	21	23	25	27	29	31
f	13	15	16	18	16	15	13

solution:

x	f	$f \cdot x$
19	13	247
21	15	315
23	16	368
25	18	450
27	16	432
29	15	435
31	13	403
	$\Sigma f = 106$	$\Sigma fx = 2620$

$$\text{Mean, } \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{2680}{106} = 25$$

```
> timestamp()
##------ Fri Jul 16 21:12:56 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x <- c(19,21,23,25,27,29,31)
> x
[1] 19 21 23 25 27 29 31
> f<-c(13,15,16,18,16,15,13)
> f
[1] 13 15 16 18 16 15 13
> z <- f*x
> z
[1] 247 315 368 450 432 435 403
> m <- sum(z)/sum(f)
> m
[1] 25
> 
```

Question-4:

Q4. The table below represents Mathematics test scores and frequency for each score.

Scores (x)	Frequency (f)
13	5
17	6
20	4
25	10

(a) Determine the median
(b) Determine the mean
solution:
(a) $\Sigma f = 25$
i.e. there are 25 scores. To determine the median, find the position of the median by adding the frequencies until you reach the position of the median.
Median lies in position 13, hence median =20
(b) mean

$$\text{(b) mean} = \frac{5(13)+6(17)+4(20)+10(23)}{25} = \frac{477}{25} \cong 19$$

```
> timestamp()
##------ Fri Jul 16 21:15:45 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x <- c(13,17,20,25)
> x
[1] 13 17 20 25
> f <- c(5,6,4,10)
> f
[1]  5  6  4 10
> n<-length(f)/2
> n
[1] 2
> me=((x[n]+x[n+1])/2)
> me
[1] 18.5
> z <- f*x
> z
[1]  65 102  80 250
> m = (sum(z)/sum(f))
>
> m
[1] 19.88
```

Question-5:

Enter the elements {5,12,7,2,6,45,11,3,63} and store in a variable x

- Display x values
- Find the number of elements of x

- Find the 5th, 8th elements
- Find the minimum element of x.
- Find the maximum element of x.

Enter the data {1, 2, ... ,19,20} in a variable x.

- Find the 3rd element in the data list.
- Find 3rd to 5th element in the data list.
- Find 2nd, 5th, 6th, and 12th element in the list.
- Print the data as {20, 19, ..., 2, 1} without again entering the data.

```
> timestamp()
##------ Fri Jul 16 21:18:47 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x<-c(5,12,7,2,6,45,11,3,63)
> x
[1]  5 12  7  2  6 45 11  3 63
> length(x)
[1] 9
> x[5]
[1] 6
> x[8]
[1] 3
> max(x)
[1] 63
> min(x)
[1] 2
> x<-1:20
> x
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
> length(x)
[1] 20
> x[3]
[1] 3
> x[3:5]
[1] 3 4 5
> x[c(2,5,6,12)]
[1]  2  5  6 12
> rev(x)
 [1] 20 19 18 17 16 15 14 13 12 11 10  9  8  7  6  5  4  3  2  1
```

Question-6:

Reading a data file and working with it:

a) Read the file first and store it in a.

b) How many rows are there in this table? How many columns are there?

c) How to find the number of rows and number of columns by a single command?

d) What are the variables in the data file?

e) If the file is very large, naturally we cannot simply type 'a', because it will cover the entire screen and we won't be able to understand anything. So how to see the top or bottom few lines in this file?

f) If the number of columns is too large, again we may face the same problem. So how to see the first 5 rows and first 3 columns?

g) How to get 1st, 3rd, 6th, and 10th row and 2nd, 4th, and 5th column?

h) How to get values in a specific row or a column?

```
> timestamp()
##------ Fri Jul 16 21:24:20 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> empid=c(1:14)
> age=c(39,34,31,35,32,33,37,39,37,31,30,42,45,36)
> sex=c(0,1,0,1,0,1,0,1,0,1,0,1,0,1)
> maritalstatus=c(0,1,0,1,0,1,0,1,0,0,1,1,1,1)
> status=c(1,2,1,2,1,2,1,2,1,2,1,2,1,2)
> empinfo=data.frame(empid,age,sex,status,maritalstatus)
> empinfo
   empid age sex status maritalstatus
1      1  39   0      1             0
2      2  34   1      2             1
3      3  31   0      1             0
4      4  35   1      2             1
5      5  32   0      1             0
6      6  33   1      2             1
7      7  37   0      1             0
8      8  39   1      2             1
9      9  37   0      1             0
10    10  31   1      2             0
11    11  30   0      1             1
12    12  42   1      2             1
13    13  45   0      1             1
14    14  36   1      2             1
```

```
> empinfo$sex=factor(empinfo$sex,labels=c("female","male"))
> empinfo$sex=factor(empinfo$sex,labels=c("staff","faculty"))
```

```
> empinfo$maritalstatus=factor(empinfo$maritalstatus,labels=c("single","married"))
> nrow(empinfo)
[1] 14
> ncol(empinfo)
[1] 5
> dim(empinfo)
[1] 14  5
> is(empinfo)
[1] "data.frame" "list"       "oldClass"   "vector"
> head(empinfo)
  empid age     sex status maritalstatus
1     1  39   staff      1        single
2     2  34 faculty      2       married
3     3  31   staff      1        single
4     4  35 faculty      2       married
5     5  32   staff      1        single
6     6  33 faculty      2       married
> head(empinfo,4)
  empid age     sex status maritalstatus
1     1  39   staff      1        single
2     2  34 faculty      2       married
3     3  31   staff      1        single
4     4  35 faculty      2       married
> tail(empinfo,4)
   empid age     sex status maritalstatus
11    11  30   staff      1       married
12    12  42 faculty      2       married
13    13  45   staff      1       married
14    14  36 faculty      2       married
> empinfo[c(1:5),c(1:3)]
  empid age     sex
1     1  39   staff
2     2  34 faculty
3     3  31   staff
4     4  35 faculty
5     5  32   staff
> empinfo[c(3,6,10),c(2,4,5)]
   age status maritalstatus
3   31      1        single
6   33      2       married
10  31      2        single
```

```
> empinfo[3]
      sex
1   staff
2 faculty
3   staff
4 faculty
5   staff
6 faculty
7   staff
8 faculty
9   staff
10 faculty
11   staff
12 faculty
13   staff
14 faculty
> empinfo[3,]
  empid age   sex status maritalstatus
3     3  31 staff      1        single
```

Assignment-3

Question-1:

Example1: Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers, and only one of them is correct. Now if a student attempts to answer every question at random:

i) Find the probability of having four or less correct answers P(x<=4).

ii) Find P(x=3)

iii) Find P(2<x<4)

iv) Generate 3 random numbers

v) Find P(x>=4)

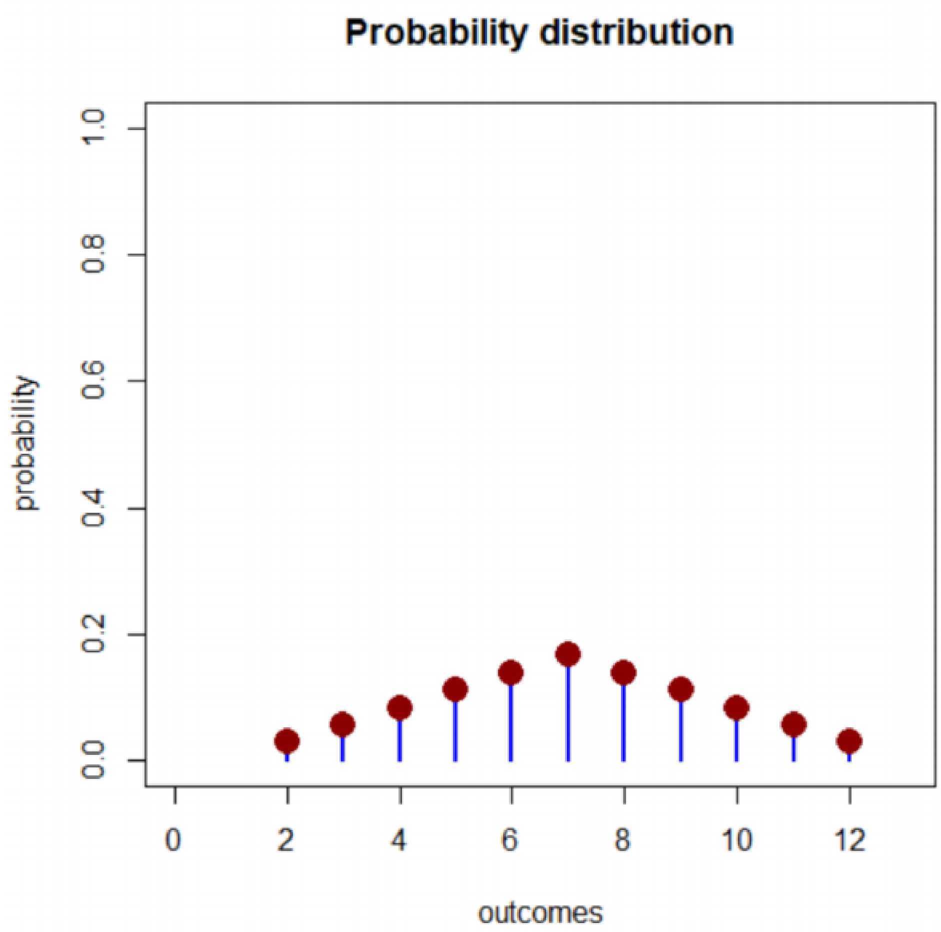
vi) What is the quantile of median?

```
> timestamp()
##------ Fri Jul 16 21:44:12 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> pbinom(4,size=12,prob=0.2)
[1] 0.9274445
> dbinom(3,size=12,prob=0.2)
[1] 0.2362232
> pbinom(3,size=12,prob=0.2)- pbinom(2,size=12,prob=0.2)
[1] 0.2362232
> rbinom(3,size=12,prob=0.2)
[1] 3 2 1
> 1-pbinom(4,size=12,prob=0.2)
[1] 0.0725555
> qbinom(0.5,size=12,prob=0.2)
[1] 2
> 1-pbinom(3,size=12,prob=0.2)
[1] 0.2054311
> rbinom(3,size=12,prob=0.2)
[1] 5 3 3
>
```

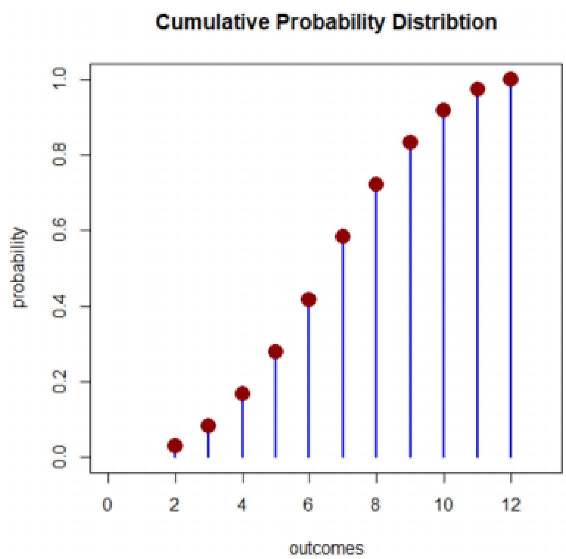
Question-2:

Example: 2 Sketch the probability function and cumulative distribution function for two fair dice thrown together . Here random variable is defined as X = Sum of two number of two dice outcome appears

```
> timestamp()
##------ Fri Jul 16 22:04:51 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> y=c(1:6,5:1)*(1/36)
> x=2:12
> plot(x,y,type="h",xlim=c(0,13),lwd=2,col="blue",main="probality distribution",xlab="outcomes",ylab="probability")
> points(x,y,pch=16,cex=2,col="dark red")
```




```
> z=c(1:6,5:1)*(1/36)
> y=cumsum(z)
> x=2:12
> plot(x,y,type="h",xlim=c(0,13),ylim=c(0,1),lwd=2,col="blue",main="cummilative probality distribution",xlab="outcomes",ylab="probability")
> points(x,y,pch=16,cex=2,col="dark red")
```



Assignment-4

Question-1:

1. A popular cold-remedy was tested for it's efficacy. In a sample of 150 people who took the remedy upon getting a cold, 117 (78%) had no symptoms one week later. In a sample of 125 people who took the placebo upon getting a cold, 90 (75%) had no symptoms one week later. The table summarizes this information.

Group	*#who are symptom Free after one*	*Total # in group (n)*	*Proportion $\hat{p} = x/n$*
Remedy	*117*	*150*	*0.78*
Placebo	*90*	*120*	*0.75*

The Test: Test the claim that the proportion of all remedy users who are symptom-
Free a fter one week is greater than the proportion for placebo users. Test this claim
at the 0.05 significance level.

$$H_0 : p_A \geq p_B$$

$$H_a : p_A < p_B$$

```
> timestamp()
##------ Fri Jul 16 22:15:49 2021 ------##
> getwd
function ()
.Internal(getwd())
<bytecode: 0x09a0cb10>
<environment: namespace:base>
> timestamp()
##------ Fri Jul 16 22:15:59 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> res <- prop.test(x=c(117,90),n=c(150,125))
> res

        2-sample test for equality of proportions with continuity correction

data:  c(117, 90) out of c(150, 125)
X-squared = 1.0161, df = 1, p-value = 0.3135
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.05024171  0.17024171
sample estimates:
prop 1 prop 2
  0.78   0.72

> res$p.value
[1] 0.3134515
```

Null hypothesis is accepted

Question-2:

Problem Statement:

A survey claims that 9 out of 10 doctors recommend aspirin for their patients with headaches. To test this claim, a random sample of 100 doctors is obtained. Of these 100 doctors, 82 indicate that they recommend aspirin. Is this claim accurate? Use alpha = 0.05.

```
> timestamp()
##------ Fri Jul 16 22:18:13 2021 ------##
> getwd
function ()
.Internal(getwd())
<bytecode: 0x09a0cb10>
<environment: namespace:base>
> timestamp()
##------ Fri Jul 16 22:18:21 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> res <- prop.test(x=82,n=100,p=0.9,correct=false)
Error in prop.test(x = 82, n = 100, p = 0.9, correct = false) :
  object 'false' not found
> res <- prop.test(x=82,n=100,p=0.9,correct=FALSE)
> res

	1-sample proportions test without continuity correction

data:  82 out of 100, null probability 0.9
X-squared = 7.1111, df = 1, p-value = 0.007661
alternative hypothesis: true p is not equal to 0.9
95 percent confidence interval:
 0.7333264 0.8829977
sample estimates:
   p
0.82

> res$p.value
[1] 0.007660761
```

Null hypothesis is rejected

Assignment-5

Question-1:

1. An outbreak of salmonella related illness was attributed to ice produced at a certain factory. Scientists measured the level of Salmonella in 9 randomly sampled batches ice crean.The levels (in MPN/g) were:

0.593 0.142 0.329 0.691 0.231 0.793 0.519 0.392 0.418

Is there evidence that the mean level pf Salmonella in ice cream greater than 0.3 MPN/g?

H0 =0.3(population mean)

Ha=population mean>0.3

```
> timestamp()
##------ Fri Jul 16 22:30:37 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
```

```
> x = c(0.593,0.142,0.329,0.691,0.231,0.793,0.519,0.392,0.418)
> res <- t.test(x,alternative="greater",mu=0.3)
> res

        One Sample t-test

data:  x
t = 2.2051, df = 8, p-value = 0.02927
alternative hypothesis: true mean is greater than 0.3
95 percent confidence interval:
 0.3245133       Inf
sample estimates:
mean of x
0.4564444
```

Question-2:

2. Suppose that 10 volunteers have taken an intelligence test; here are the results obtained. The average score of the entire population is 75 in the same test. Is there any significant difference (with a significance level of 95%) between the sample and population means, assuming that the variance of the population is not known. Scores: 65, 78, 88, 55, 48, 95, 66, 57, 79, 81

The average score of entire population = 75 = H0

The average score of entire population ≠ 75 = Ha

```
> timestamp()
##------ Fri Jul 16 22:44:22 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> a=c(175,168,168,190,156,181,182,175,174,179)
> a
 [1] 175 168 168 190 156 181 182 175 174 179
> b=c(185,169,173,173,188,186,175,174,179,180)
> b
 [1] 185 169 173 173 188 186 175 174 179 180
> t.test(a,b)

        Welch Two Sample t-test

data:  a and b
t = -0.94737, df = 15.981, p-value = 0.3576
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.008795   4.208795
sample estimates:
mean of x mean of y 
    174.8     178.2 
```

Assignment-6

Question-1:

1. (Goodness of fit test) Suppose you flip two coins 100 times. The results are 20 HH, 27 HT, 30 TH, and 23 TT. Are the coins fair? Test at a 5% significance level.

```
> timestamp()
##------ Fri Jul 16 22:50:01 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> r<-c(20,57,23)
> r
[1] 20 57 23
> res<-chisq.test(r,p=c(0.25,0.50,0.25))
> res

        Chi-squared test for given probabilities

data:  r
X-squared = 2.14, df = 2, p-value = 0.343
```

Null hypothesis-accepted

Question-2:

2. (Goodness of fit test) One study indicates that the number of televisions that American families have is distributed (this is the given distribution for the American population) as in the table.

No. of Television	Percent
0	10
1	16
2	55
3	11
4	8

The table contains expected (E) percents. A random sample of 600 families in the far western United States resulted in the data in this table.

Number of Televisions	Frequency
0	66
1	119
2	340
3	60
4	15
-	Total=600

The table contains observed (O) frequency values. At the 1% significance level, does it appear that the distribution number of televisions of far western United States families is different from the distribution for the American population as a whole?

```
> timestamp()
##------ Fri Jul 16 22:52:05 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> r<-c(66,119,340,60,15)
> r
[1]  66 119 340  60  15
> res<-chisq.test(r,p=c(0.1,0.16,0.55,0.11,0.08))
> res

        Chi-squared test for given probabilities

data:  r
X-squared = 29.646, df = 4, p-value = 5.776e-06
```

Null hypothesis is rejected

Question-3:

3. Nadir is testing an octahedral die to see if it is biased. The resultsare given in the table below.

Score	1	2	3	4	5	6	7	8
Frequency	7	10	11	9	12	10	14	7

Test the hypothesis that the die is fair.

```
> timestamp()
##------ Fri Jul 16 22:56:29 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> res<-chisq.test(r,p=c(1/8,1/8,1/8,1/8,1/8,1/8,1/8,1/8))
> res

        Chi-squared test for given probabilities

data:  r
X-squared = 4, df = 7, p-value = 0.7798
```

Null hypothesis is accepted

Question-4:

4. (F-test) Two college instructors are interested in whether or not there is any variation in the way theygrade math exams. They each grade the same set of 30 exams. The first instructors grades have avariance of 52.3. The second instructors grades have a variance of 89.9. Test the claim that the first instructors variance is smaller. (In most colleges, it is desirable forthe variances of exam grades to be nearly the same among instructors.) The level of significance is10%.

```
> timestamp()
##------ Fri Jul 16 22:58:03 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x=52.3
> y=89.9
> f=x/y
> f
[1] 0.5817575
> p=0.1
> qf(p,29,29,lower.tail=FALSE)
[1] 1.6199
> pf(f,29,29,lower,tail=TRUE)
Error in pf(f, 29, 29, lower, tail = TRUE) :
  unused argument (tail = TRUE)
> pf(f,29,29,lower.tail=TRUE)
[1] 0.07529685
> 
```

Question-5:

5. (F-test) One of the quality measures of blood glucose meter strips is the consistency of the test results on the same sample of blood. The consistency is measured by the variance of the readings in repeated testing. Suppose two types of strips, A and B, are compared for their respective consistencies. We arbitrarily label the population of Type A strips Population 1 and the population of Type B strips Population 2. Suppose 15 Type A strips were tested with blood drops from a well-shaken vial and 20 Type B strips were tested with the blood from the same vial. The results are summarized in below Table. Assume the glucose readings using Type A strips follow a normal distribution with variance σ_1^2 and those using Type B strips follow a normal distribution with variance with σ_2^2. Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude that the consistencies of the two types of strips are different.

Strip Type	Sample size	Sample Variance
A	$n_1 = 16$	2.09
B	$n_2 = 21$	1.10

```
[1] 0.07329685
> timestamp()
##------ Fri Jul 16 23:01:19 2021 ------##
> getwd()
[1] "C:/Users/91995/OneDrive - vitap.ac.in/Documents"
> x=2.09
> y=1.10
> f=x/y
> f
[1] 1.9
> qf(p=0.05,15,20,lower.tail=FALSE)
[1] 2.203274
> qf(p=0.05,15,20,lower.tail=TRUE)
[1] 0.4296391
```