Practical no. 1.

R Software:-

1. R is a Software for Statistical analysis and data computing.
2. It is an effective data handling software and outcome storage is possible.
3. It is capable of graphical display.
4. It is free Software.

Q1. Solve the following:-

1. $4+6+8\div 2-5$

> 4+6+8/2-5

> 9

2. $2^2+1-31+\sqrt{45}$

> 2^2 + 1 - 31 + Sqrt(45)

> 13.7082

3. $5^3+7\times 5\times 8+46\div 5$

> 5^3 + 7 * 8 * 5 +46 /5

> 414.2

4. $\sqrt{4^2+5\times 3+7\div 6}$

> Sqrt(4^2 + 5*3 + 7/6)

> 5.671567

5. roundoff 46 ÷ 7 + 9 X 8
>round (46 / 7+9*8)
> 79

Q2. Solve the following :-

① >c (2,3,5,7) * 2
> 4 , 6 10 14

② >c(2,3,5,7) * c (2, 3)
> 4 9 10 21

③ x(2,3,5,7) * c (2,3,6,2)
> 4 9 30 14

④ >c (1,6,2,3) * c (-2,-3,-4,-1)
[1] -2 -18 -8 -3

⑤ >c (2,3,5,7)^2
[1] 4 9 25 49

⑥ >c (4,6,8,9,4,5)^c (1,2, 3)
[1] 4 36 512 9 16 125

⑦ >c (6,2,7,5) / c (4,5)
[1] 1.50 0.40 1.75 1.00

Q3. Solve the following :-

1] > x = 20
> y = 30
> z = 2
> x ^ 2 + y ^ 3 + z
[1] 27402

2] $\sqrt{x^2+y^z}$
> sqrt (x ^ 2 + y)
> [1] 20.73644

3] $x^z + y^z$
> x ^ z + y ^ z
[1] 1300

Q4. Solve :-

(i) $\begin{bmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{bmatrix}$

> x = matrix (nrow=4, ncol=2, data= (1:8))
> x

	[,1]	[,2]
[1,]	1	5
[2,]	2	6
[3,]	3	7
[4,]	4	8

(ii) $x = \begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 3 \end{bmatrix}$ $y = \begin{bmatrix} 10 & -5 & 7 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$

> x = matrix (nrow=3, ncol=3, data = c (4,7,9,-2,0,-6,6,7,3))
> y = matrix (nrow = 3, ncol = 3, data = c(10,12,15,-5,-4,-6,7,9,6))

Q5. Prints of marks of computer science students are as follows:-

Data: 59, 20, 35, 24, 46, 56, 55, 45, 27, 22, 47, 58, 54, 40, 50, 32, 36, 29, 35, 39

```
> x = c (given data)
> length (x)
> [1] 20
> breaks = seq (20, 60, 5)
> a = cut (x, breaks, right = FALSE)
> b = table (a)
> c = transform (b)
> c
     x        freq
1  [20, 25)    3
2  [25, 30)    2
3  [30, 35)    1
4  [35, 40)    4
5  [40, 45)    1
6  [45, 50)    3
7  [50, 55)    2
8  [55, 60)    4
```


Probability distribution.

Q1. check whether the followings are p.m.f or not.

(i)

x	$p(x)$
0	0.1
1	0.2
2	-0.5
3	0.4
4	0.3
5	0.5

Since $p(2) = -0.5$,
can't be a probability mass function.
Since in p.m.f $p(x) \geq 0 \ \forall x$

(ii)

x	1	2	3	4	5
$p(x)$	0.2	0.2	0.3	0.2	0.2

It can't be a p.m.f,
as in p.m.f
$\Sigma p(x) = 1$

(iii)

x	10	20	30	40	50
$p(x)$	0.2	0.2	0.35	0.15	0.1

```
> prob=c(0.2,0.2,0.35,0.15,0.1)
> sum(prob)
[1]1.
```

As $\Sigma p(x) = 1$,
it is a p.m.f

Q2.
(i) Find c.d.f for the following p.m.f and sketch the graph.

x	10	20	30	40	50
$p(x)$	0.2	0.2	0.35	0.15	0.1

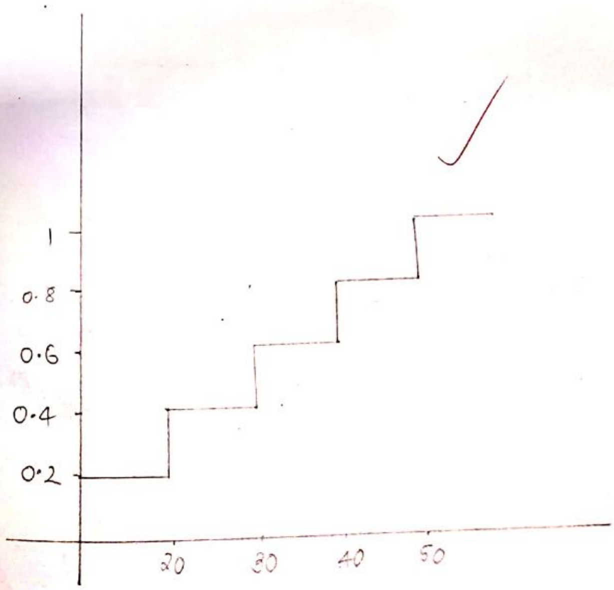
$$F(X) = \begin{cases} 0 & x < 10 \\ 0.2 & 10 \leq x < 20 \\ 0.4 & 20 \leq x < 30 \\ 0.75 & 30 \leq x < 40 \\ 0.95 & 40 \leq x < 50 \\ 1.0 & x \geq 50 \end{cases}$$

Q2.
(i)

```
> prob=c(0.2,0.2,0.35,0.15,0.1)
> sum(prob)
[1]1
```

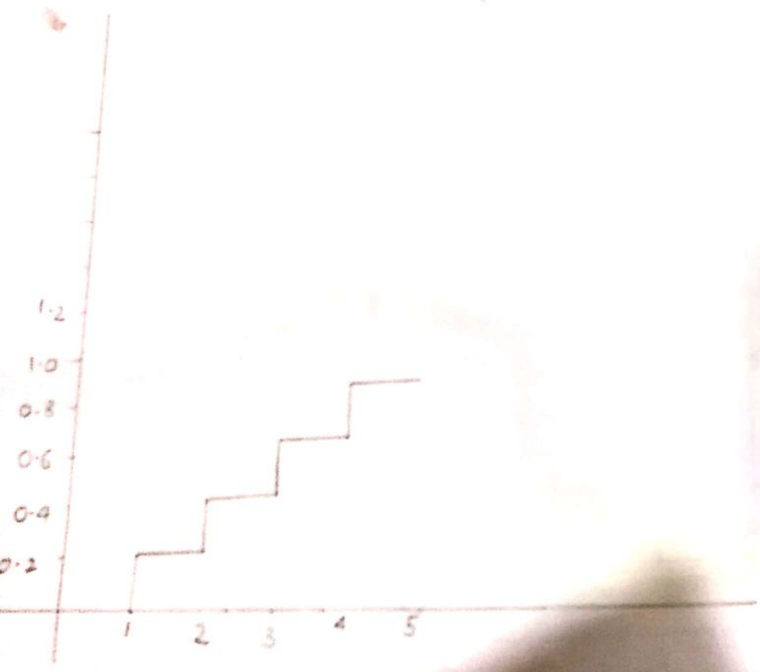
As $\Sigma p(x) = 1$,
it is a p.m.f

```
> cumsum(prob)
[1] 0.20   0.04   0.75   0.90   1.00
> x = c(10,20,30,40,50)
> plot (x, cumsum(prob),"s")
```



3.

```
> prob = c(0.15, 0.25, 0.1, 0.2, 0.2, 0.1)
> sum(prob)
[1] 1
> cumsum(prob)
[1] 0.15  0.40  0.50  0.70  0.90  1.00
> x = c(1, 2, 3, 4, 5, 6)
> plot(x, cumsum(prob), "s")
```



(ii)

x	1	2	3	4	5	6
$p(x)$	0.15	0.25	0.1	0.2	0.2	0.1

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.15 & 1 \le x < 2 \\ 0.40 & 2 \le x < 3 \\ 0.50 & 3 \le x < 4 \\ 0.70 & 4 \le x < 5 \\ 0.90 & 5 \le x < 6 \\ 1.00 & x \ge 6 \end{cases}$$

3

(Q3) Check whether the following is p.d.f or not.

(i) $f(x) = 3-2x \quad ; \quad 0 \leq x \leq 1$

$= \int_0^1 f(x)\,dx$

$= \int_0^1 (3-2x)\,dx$

$= \int_0^1 3\,dx - \int_0^1 2x\,dx$

$= [3x - x^2]_0^1$

$= 2.$

(ii) $f(x) = 3x^2 \quad 0 < x < 1$

$= \int_0^1 f(x)\,dx$

$= \int_0^1 3x^2$

$= \left. \frac{3x^3}{3} \right]_0^1$

$= x^3$

$= 1$

①
> dbinom (10, 100, 0.1)
>[1] 0.1318653

②
i) > dbinom (4, 12, 1/5)
[1] 0.1328756

ii)
> pbinom (4, 12, 1/5)
[1] 0.9274445

iii)
> 1- pbinom (5, 12, 1/5)
[1] 0.01940528

③
> dbinom (0:5, 5, 0.1)
[1] 0.59049 0.32805 0.07290 0.00810 0.00045 0.0000

④
i) > dbinom (5, 12, 0.25)
[1] 0.1032414

ii)
> pbinom (5, 12, 0.25)
[1] 0.9455978

iii) > 1- pbinom (7, 12, 0.25)
[1] 0.00278151

(V) > dbinom (6, 12, 0.25)
[1] 0.04014945

PM

Binomial distribution

1. Find the probability of 10 Success in 100 trials with $p=0.1$

2. Suppose there are 12 MCQ's, each question has 5 options out of which 1 is correct, find the probability
① of having exactly 4 correct answers.
② atmost 4 correct answers
③ More than 5 correct answers.

3. Find the complete distribution when $n=5$ and $p=0.1$

4. $n=12$
$p=0.25$
Find the following probabilities
① $p(X=5)$
② $p(x \leq 5)$
③ $p(x>7)$
④ $p(5<x<7)$

(5) The probability of a Saleoman making a Sell to a customer is 0.15.
Find the probability of
(i) No sells out of 10 customer.
(ii) More than 3 sells out of 20 customers.
(iii)

(6) A Saleoman as a 20% probability of making a sale to a customer. out of 30 customers what minimum number of sales he can make with 88% probability.

(7) X follow binomial distribution
$n = 10$
$p = 0.3$
plot the graph of p.m.f and C.b.f.

(5)
i) >dbinom (0,10,0.15)
[1] 0.1968744

ii) > 1-pbinom (3,20,0.15)
[1] 0.3522748

(6)
> qbinom (0.88, 30, 0.2)
> 6

(7)
>prob = dbinom (0: n, 10, 0.3)
>cumprob = pbinom (0:n, 10, 0.3)
>d = data.frame ("x values" = x, "probability" = prob)
> print(d)

	x.values	probability.
		0.0282
1	0	0.1210
2	2	0.2334
3	2	0.2668
4 5	3	0.2001
6	4	0.1029
7	5	0.370
8	6	
9	7	0:009
10	8	0.001
11	9	0. 0.001
	10	0. 000005

```
>p3 =1-pnorm(14,12,3)
>p3
[1] 0.2524925

>rnorm(5,12,3)
[1]8.825718  13.798805  11.058683  15.467502  13.78143
```

2. X follow normal distribution with $\mu = 10$, $\sigma = 2$

Find:

(i) $P(x \leq 7)$

(ii) $P(5 \leq x < 12)$

(iii) $P(x > 12)$

(iv) generate 10 observation

(v) Find K such that probability $P(x < K) = 0.4$

```
>pnorm(7,10,2)
[1] 0.0668072

>pnorm(12,10,2)-pnorm(5,10,2)
[1] 0.8351351

>1-pnorm(12,10,2)
[1]0.1586553

>rnorm(10,10,2)
[1] 10.806022  11.238265  8.444346  8.047978  6.380056
    13.063895  11.512423  11.553602  13.010855  11.221799

>qnorm(0.4,10,2)
[1]9.493306
```


3. Generate 5 random numbers from normal distribution with $\mu = 15$, $\sigma = 4$. Find sample mean, median, S.D and print it.

4. X follows normal i.e $X \sim N(30,100)$
$\mu = 30$, $\sigma = 10$. Find

i) $P(x \leq 40)$

ii) $P(x > 35)$

iii) $P(25 < x < 35)$

iv) Find K such that $P(x < K) = 0.6$

(4) code:-

```
>pnorm(40,30,10)
[1]0.8413447

>1-pnorm(35,30,10)
[1]0.3085375

>pnorm(35,30,10)-pnorm(25,30,10)
[1]0.3829249

>qnorm(0.6,30,10)
[1]32.53347
```

(3) code:-

```
>x=rnorm(5,15,4)
>x
[1]14.29548  16.93377  14.20567  19.73212  11.87906

>am=mean(x)
>am
[1]15.30922

>me=median(x)
>me
[1]14.29548
```

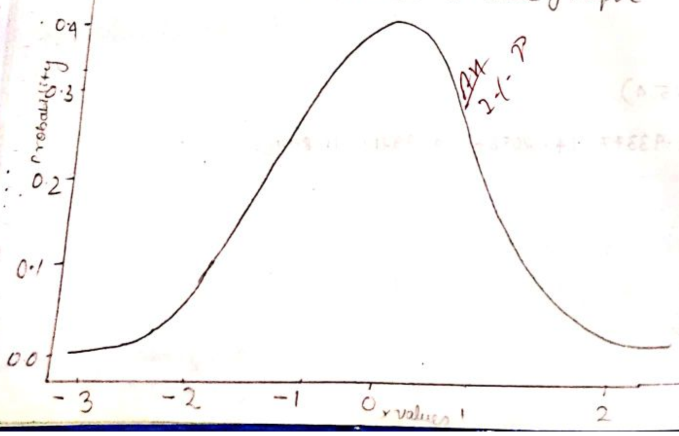
Scanned by CamScanner

8

```
> n = 5
> variance = (n-1) * var (x)/n
> variance
[1] 6.967641
> sd = sqrt (variance)
> sd
[1] 2.639629
> cat ("Sample mean is", am)
Sample mean is 15.30922
> cat ("Sample median is ", me)
Sample median is 14.29548
> cat ("Sample S.D =", sd)
Sample S.D = 2.639629
```

5. Plot Standard normal graph.

```
> x = seq(-3, 3, by = 0.1)
> y = dnorm(x)
> plot (x, y, xlab = "x values", ylab = "probability", main = "Standard graph")
```



Practical no.5. 65

Normal and t - test :

Q1. Test the hypothesis $H_0 : \mu = 15$ against $H_1 : \mu \neq 15$. Random sample of size 400 is drawn and it is calculate Sample mean is 14 and standard deviation is ~~5~~ 3 test the hypothesis at 5 % level of significance.

```
> mo = 15
> mx = 14
> sd = 3
> n = 400
> zcal = (mx - mo)/(sd/sqrt(n))
> zcal
> [1] - 6.666667
> cat (" calculated value of z is =", zcal)
> calculated value of z is - 6.666667
> pvalue = 2 * (1 - pnorm (abs(zcal)))
> pvalue
> [1] 2.616796 e -11
```

Since pvalue is ~~more~~ less than 0.05,

∴ we accept the value of $H_1 : \mu \neq 15$ and reject the value of $H_0 : \mu = 15$.

Q2. Test the hypothesis $H_0: \mu = 10$ against $H_1: \mu \neq 10$. Random saple of size 400 is draw with sample mean 10.2 and standard deviation 2.25. Test the hypothesis at 5% level of significance.

```
> mo = 10
> mx = 10.2
> sd = 2.25
> n = 400
> zcal = (mo - mx) / (sd / sqrt(n))
> zcal =
[1] -1.777778
> pvalue = 2 * (1 - pnor (abs (zcal)))
> pvalue
[1] 0.07544036
```

Since the pvalue is more than 0.05, we accept the value of $H_0: \mu = 10$.

Q3. Test the hypothesis H_0: μ proportion of smokers in a college is 0.2. A sample is collected and is calculated sample proportion = 0.125: Test the hypothesis at 5% level of significance. Sample size is 400.

```
> P = 0.2
> p = 0.125
> n = 400
> q = 1 - P
> z cal = (p - P) / (sqrt (P * q / n))
> zcal
[1] -3.75
```



```
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 0.0001768396
```

Since the pvalue is less than 0.05, we do not accept the value.

Q4. Last year the farmers lost 20% of their crops, a random sample of 60 fields are collected and found that 9 fields crops are insect polluted. Test the hypothesis at 1% level of significance

```
> P = 0.2
> p = 9/60
> n = 60
> q = 1 - P
> Zcal = (p - P) / (sqrt (P * q / n))
> zcal
[1] -0.9682458
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 0.3329216
```

AM 16.01.20

Since the pvalue is more than 0.01, we accept the value

Q5. Test the hypothesis $H_0: \mu = 12.5$, from the following sample at 5% level of significance.

```
> m x = mean (X)
> variance = (n-1) * var(x)/n
> sd = sqrt (variance)
> mo = 12.5
> t = (m x - mo) / (sd / (sqrt(n)))
> p value = 2 * (1 - pnorm (abs (t)))
> pvalue
[1] 0
```

Since the pvalue is less than 0.05, we do not accept the value.

Scanned by CamScanner

Practical no. 6.

Large Sample test.

1) Let the population mean (the amount spent per customer in a restaurant) is 250 a sample of 100 customers selected the sample mean is calculated as 275 and S.D 30, test the hypothesis that population mean is 250 or not at 5% level of significance.

2) In a random sample of 1000 students it is found that 750 use blue pen. Test the hypothesis that the population proportion is 0.8 at 1% level of significance.

1. Ho : $\mu = 275$ against $H_1 : \mu \neq 275$

```
> mx = 275
> mo = 250
> sd = 30
> n = 100
> zcal = (mx-mo)/(sd/sqrt(n))
> zcal
[1] 8.3333
> pvalue = 2*(1- pnorm(zcal))
> pvalue
[1] 0
```

pvalue = 0 < 0.05

we reject Ho at 5% level of significance.

2.
```
>P=0.8
>Q=1-P
>p = 750/1000
>n =1000
>Zcal = ((p-P) / sqrt (P*Q/n))
[1] -3-952847
>pvalue = (2* (1-pnorm (abs (zcal))
> 7.72268e-05
```
pvalue = 7.72268e-05 < 0.01

we do not accept it.

③ Two random sample of Size 1000 and 2000 are drawn from 2 population with the Same Standard deviation 2.5. The Sample means are 67.5 and 68 respectively. Test the hypothesis $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ at 5% level of Significance.

④ A Study of noise level in 2 hospital is given below. Test the claim that the two hospitals are same level of noise at 1% level of Significance.

	Hospital A	Hospital B
Size	84	34
Mean	61.2	59.4
S.D	7.9	7.5

⑤ In a Sample of 600 Students in a college 400 use blue ink in another college from a sample of 900 students 450 use blue ink test the hypothesis that the proportion of students using blue ink in two colleges are equal or not at 1% level of significance.

3. $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$
```
>n1 = 1000
>n2 = 2000
>mx1 = 67.5
>mx2 = 68
>sd1 = 2.5
>sd2 = 2.5
>Zcal = (mx1-mx2)/Sqrt ((sd^2/n1) + (sd2^2/n2))
>Zcal
[1] -5.163978
>pvalue = (2*(1-pnorm (abs (zcal))))
>pvalue
[1] 2.417564e-07
```
pvalue = 2.417564e-07 < 0.05

we reject it.

4.
>n1 = 84
>n2 = 34
>mx1 = 61.2
>mx2 = 59.4
>sd1 = 7.9
>sd2 = 7.5
>zcal = (mx1 - mx2)/sqrt((sd1^2/n1) + (sd2^2/n2))
>zcal
[1] 1.162528
>pval = (2*(1-pnorm(abs(zcal))))
>pval
[1] 0.245021
pval = 0.245021 > 0.1
we accept it.

5. $H_0: P_1 = P_2$ against $H_1: \mu_1 \neq \mu_2$
>n1 = 600
>n2 = 900
>p1 = 400/600
>p2 = 450/900
>p = (n1*p1 + n2*p2)/(n1+n2)
>q = 1-p
>zcal = (p1-p2)/sqrt(p*q*(1/n1 + 1/n2))
>zcal
[1] ~~0.01332648~~ 6.381534
> pvalue = (2*(1-pnorm(abs(zcal))))
>pvalue
[1] 1.753222e-10
pvalue = 1.753222e-10 < 0.1
∴ we do not accept it.

For sample size
6. n1 = 200, n2 = 200
p1 = 44/200
p2 = 30/200
test at 5% level of significance.

$H_0: P_1 = P_2$ vs. $H_1: P_1 \neq P_2$

>n1 = 200
>n2 = 200
>p1 = 44/200
>p2 = 30/200
>zcal = (p1-p2)/sqrt((p*q)*(1/n1+1/n2))
>zcal
[1] 1.412613
>pval = (2*(1-pnorm(abs(zcal))))
>pval
[1] 0.1577696.
pval = 0.1577696 < 0.5
∴ we reject Ho

Anu
23.11.20

Practical no. 7.

Small sample test.

1. The marks of 10 students are given by 63, 63, 66, 67, 68, 69, 70, 70, 71, 72. Test the hypothesis that the sample comes from opposition that the sample with average 66.

$H_0 . \mu = 66$

```
> x = c(63,63,66,67,68,69,70,70,71,72)
> t.test(x)
        One Sample t-test
data: x
t = 68.319, df=9, p-value = 1.558e-13
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
65.65171   70.14829
Sample estimates:
mean of x
   67.9
```

As $1.558e^{-13} < 0.05$, we reject it.

```
> pvalue = 1.558e-13
> if (pvalue>0.05) {cat("accept Ho")} else {cat("reject Ho")}
reject Ho>
```


2. Two groups of students scored the following marks. Test the hypothesis that there is no significant difference between the two groups.

Group1 - 18, 22, 21, 17, 20, 17, 23, 20, 22, 21.
Group2 - 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Ho - there is no difference between the two grps..

```
> x = c(18,22,21,17,20,17,23,20,22,21)
> y = c(16,20,14,21,20,18,13,15,17,21)
> t.test(x,y)
        Welch Two Sample t-test
data: x and y
t = 2.2573, df = 16.376, p-value = 0.03798
alternative hypothesis: true difference in mean is not equal to 0.
95 percent confidence interval:
 0.1628205    5.0371795
Sample estimates:
mean of x mean of y
   20.1      17.5
> pvalue = 0.03798
> if (pvalue>0.05) {cat("accept Ho")}else{cat("reject Ho")}
reject Ho
```

3. The sales data of 6 shops before and after a special campaign given below
before - 53, 28, 31, 48, 50, 42
After - 58, 29, 30, 55, 56, 45
Test the hypothesis that the campaign is effective or not.
Ho there is no significant difference of sales before and after the campaign.

```
> x=c (53, 28,31, 48,50, 42)
> y=c (58,29, 30,55,56, 45)
> t. test (x,y, paired = T, altertiative ="greater")
        Paired  t-test

data: x and y
t= -2.7815, df=5, p-value =0.9806
alternative hypothesis: true difference in means is greater than 0.
95 percent confidence interval:
 -6.035547     Inf
Sample estimates:
mean of the differences
        -3.5

> pvalue =0.9806
> if (pvalue>0.05) { cat ("accept Ho")} else {cat("reject Ho")}
accept Ho
```


4. Two medicines are applied to two grps of patients respectively.
Grp1 - 10, 12, 13, 11, 14
Grp2 - 8, 9, 12, 14, 15, 10, 9
Is there any significant difference between the two elements.

5. The following are the weight before and after a diet program. Is the diet program effective.
Before - 120, 125, 115, 130, 123, 119
After - 100, 114, 95, 90, 115, 99.

4.
```
>x=c (10, 12,13, 11, 14)
>y=c (8,9,12, 14,15,10,9)
> t. test (x,y)
> pvalue =0.4406
> if (pvalue > 0.05) {cat ("accept Ho")} else {cat ("reject Ho")}
accept Ho
```

5 Ho there's no significant differences between before and after.
```
> x = c (120, 125, 115, 130, 123, 119)
>y = c (100, 114, 95, 90, 115, 99)
> t. test (x,y, paired = T, alternative = "less")
>pvalue =0.9963
> if (pvalue >0.05) {cat ("accept Ho")} else {cat ("reject Ho")}
accept Ho
```

AM
6-2-20

Practical no. 8.

Large and Small Sample test :

Q1. $H_0 : \mu = 55$, $H_1 : \mu \neq 55$.

```
>n=100
>mx=52
>m0=55
>sd=7
>zcal=(mx-m0)/(sd/sqrt(n))
>zcal
>-4.285714
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.82153e-05
```

As pvalue < 0.05 we reject.

Q2. $H_0 : P = 0.5$, $H_1 : p \neq 0.5$

```
>P=0.5
>p=350/700
>n=700
>Q=1-P
>zcal=((p-P)/sqrt(P*Q/n))
>zcal
[1] 0
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1
```

As pvalue > 0.01, we accept.

Q3. $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$

```
>n1=1000
>n2=1500
>p1=0.02
>p2=0.01
>p=(n1*p1+n2*p2)/(n1+n2)
>q=1-p
>zcal=((p1-p2)/sqrt((p*q)*(1/n1+1/n2)))
>zcal
[1] 2.084842
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.03708364
```

As pvalue < 0.05, we reject.

Q4. $H_0 : \mu = 99$, $H_1 : \mu \neq 99$.

```
>mx=100
>m0=99
>n=400
>var=64
>sd=sqrt(var)
>zcal=(mx-m0)/(sd/sqrt(n))
>zcal
[1] 2.5
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.01241933
```

Q5. Ho: $\mu = 66$

```
> x = c(63,63,68,69,71,71,72)
> t.test(x)
```

One sample t-test

data: x

t = 47.94, df = 6, p-value = 5.522e-09

alternative hypothesis: true mean is not equal to 0.

95 percent confidence interval:

64.66479 71.62092

Sample estimates:

mean of x

68.14286

As pvalue < 0.01 we reject.

Q7. Ho: $\mu = 1200$, $H_1: \mu \neq 1200$

```
> n = 100
> mx = 1150
> mo = 1200
> sd = 125
> Zcal(mx - mo)/((sd/sqrt(n)))
> Zcal
[1] -4
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 6.334248e-05
```

As pvalue < 0.01, we reject.

Q6. Ho: $\sigma_1 = \sigma_2$, Ho: $\sigma_1 \neq \sigma_2$

```
> x = c(66,67,75,76,82,88,84,90,92)
> y = c(64,66,74,78,82,85,87,92,95,93,97)
> var.test(x,y)
```

F test to compare two variance

data: x and y

F = 0.70686, num df = 8, denom df = 10, p-value = 0.6359.

alternative hypothesis: true ratio of variance is not equal to 1

95 percent confidence interval:

0.1833662 3.0360393

Sample estimates:

ratio of variances

0.7068567

As pvalue > 0.05, we accept.

Q8. Ho: $p_1 = p_2$, $H_1: p_1 \neq p_2$

AM 13/2/20

```
> n1 = 200
> n2 = 300
> p1 = 44/200
> p2 = 56/300
> p = (n1*p1 + n2*p2)/(n1+n2)
> q = 1-p
> zcal = ((p1-p2)/(sqrt((p*q)*(1/n1+1/n2))))
> Zcal
[1] 0.9128709
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.3613104 > 0.01, we accept.
```

35

Practical no. 9.

Topic : Chi square test and ANOVA

① use the following data to test whether the condition of the home and the condition of child are independent or not.

condition of home.

Clean	Dirty		
70	50	Clean	condition of child.
80	20	Fairly clean	
35	45	Dirty	

Sol^n:

Ho: condition of home and child are independent

Ho +

x = scan (70,80,35,60,20,45)

m = 3

n = 2

y = matrix (x, nrow = m, ncol = n)

chisq. test(y)

Pearson's chi-squared test

data = y

x-squared = 25.646 , df = 2, p-value = 2.698 e-06

Ans since pvalue < 0.01, we don't accept Ho.

71

② Test Ho that vaccination and disease are independent or not.

vaccine

All	Not all		
70	46	All	Disease.
35	37	Not all	

Ho = vacc and disease are independent

x = c(70, 35, 46, 37)

m = 2

n = 2

y = matrix. (x, nrow = m, ncol = n)

chisq. test(y)

Pearson's chisq with yate's cont. corr.

data = y

x = sq = 2.0275 , df = 1, pvalue = 0.1545

since > 0.01, we accept Ho.

Perform a ANOVA for the following data

Type	Observations
A	50,52
B	53, 55, 53
C	60, 58, 57, 56
D	52, 54, 54, 55

Soln:

Ho : Means are equal for A,B,C,D

```
>x1=c(50,52)
>x2=c(53,55,53)
>x3=c(60,58,57,56)
>x4=c(52,54,54,55)
>d=stack(list(b1=x1, b2=x2, b3=x3, b4=x4))
>names(b)
>d
```

	values	ind
1	50	b1
2	52	b1
3	53	b2
4	55	b2
5	53	b2
6	60	b3
7	58	b3
8	57	b3
9	56	b3
10	52	b4
11	54	b4
12	54	b4

5

```
> names (d)
[1] "values" "ind"
> oneway.test(values ~ ind), data=d, var.equal =T)
  One way analysis of means
data: values and ind
F= 11.735, num df = 3, denom df=9, p-value = 0.00183
> anova = aov (values ~ ind, data=d)
> Summary (anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	3	71.06	23.688	11.73	0.00183 **
Residuals	9	18.17	2.019		

\- - -

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

④ Following gives the life of a tyres of 4 brands.

Type	life
A	20, 23, 18, 17, 22, 24
B	~~23~~ 19, 15, 17, 20, 16, 17
C	~~18~~ 21, 19, 22, 17, 20
D	15, 14, 16, 18, 14, 15

Ho: The avg. life of A, B, C, D are equal.

```
> x1 = c(20, 23, 18, 22, 24)
> x2 = c(19,15, 17, 20, 16, 17)
> x3 = c(21, 19, 22, 17, 20)
> x4 = c( 21, 19, 22, 17, 20)
> d = stack (list (b1 = x1, b2 = x2, b3 = x3, b4 = x4))
```



```
> names (d)
> oneway.test(values ~ ind, data = d, var.equal = T)
  oneway analysis of means
data: values and ind
F = 7.4403, num df = 3, denom df = 19, p-value = 0.001719
```

⑤

```
> x = read.csv ("/users/ritesh/Desktop/marks.csv")
> print (x)
```

	Statistics	Maths
	40	60
1	45	48
2	42	47
3	15	20
4	37	25
5	36	27
6	45	57
7	59	58
8	20	25
9	27	27

```
> am = mean (x $ Statistics)
> am
[1] 36.6
> mean (x $ maths)
[1] 39.4
> n = length (x $ statistics)
> n
[1] 10
> sqrt ((n-1) * var(x $ statistics)/n)
[1] 12.32234
> sqrt ((n-1) * var(x $ maths)/n)
[1] 15.2
> cor (x $ Statistics, x $ maths)
[1] 0.8150581
```

AM 20.12.20

Practical no. 10.

Topic: Non-parametric test.

Following are the amounts of sulphur oxide emitted by some industry in twenty days apply sign test to test hypothesis that the population median is 21.5 at 5% level of significance.
17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 25, 29, 25.

H_0: population median in 21.5 against H_1: population median less than ~~greater than~~ 21.5

```
> x = c( . . . . . . . . )
> me = 21.5
> sp = length (x [x > me])
> sn = length (x [x < me])
> n = sp + sn
> n
[1] 20
> pv = pbinom (sp, n, 0.5)
> pv
[1] 0.4119015
```

As pvalue i.e 0.4119015 > 0.05, we accept.

If the alternative H_1 ~~is~~ me ≠ or me < then pv = pbinom(sp, n, 0.5)

and if me > then pv = pbinom (sn, n, 0.5)

As pv ≯ 0.05, we accept.

Q2. Following is a data of 10 observations. Apply sign test to test the hypothesis that the population median is 625 against the alternative it is more than 625.
612, 619, 631, 628, 643, 640, 655, 649, 670, 66..
Ho: population median is 625 against H1: population is > 62..

```
> x = c(       )
> me = 625
> sp = length (x [x>me])
> sn = length (x [x<me])
> n = sp+sn
> n
[1] 10
> pv = pbinom (sn, n, 0.5)
> pv
[1] 0.0546875
```

As pv > 0.05 we accept.

Q3. The following are the values of a sample. Test the hypothesis that the population median is 60 against the alternative it is more that 60 at 5% LOS, using wilcoxon signed Rank test.
63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 69, 48, 66, 72, 63, 87, 69.
Ho: population median is 60 against H1: population > 60

```
> x = c( . . . . . )
> length(x)
[1] 20.
```



```
> wilcox.test (x, alter="greater", mu=60)
```

wilcoxon signed rank test with continuity correction

data : x
V = 145, p-value = 0.02298
As pvalue < 0.05, we reject.

Note: If the alternative is less than alter="less" and if the alternative is not ≠ then alter="two sided"

Q4. using WSRT, test the population median is 12 or less than 12.
15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26
Ho: population median is 12 against H1: population < 12.

```
> x = c(       )
> wilcox.test (x, alter="less", mu=12)
```

wilcoxon signed rank test with continuity correction:

data : x
V = 66, p-value = 0.9986
As pvalue > 0.05, we accept.

Q5. The weights of students before and after they stopped smoking are given below. using WSRT, test that there is no significant change.

weights before	weights after
65	72
75	74
75	72
62	66
72	73

H_0: Before and after there is no change against

H_1: there is change in their weight.

```
> x = c(65,75,75,62,72)
> y = c(72,74,72,66,73)
> d = x - y
> wilcox.test(d, alter = "two.sided", mu=0)
```

wilcoxon signed rank test with continuity correction

data: d

V = 4.5, p-value = 0.4982.

As p-value > ~~0.4482~~ 0.05, we accept.

AM 27.02