Sem - 2

Practical no: 1

- **Basics of R software:**

1) R is a software for data analysis and statistical computing.
2) It is a software by which effective data handling and outcome storage is possible.
3) It is capable of graphical display.
4) It's a free software.

1) $2^2 + |-5| + 4 \times 5 + 6 \div 5$

→ > 2 ^ 2 + abs (-5) + 4 * 5 + 6/5

[1] 30.2

2) $x = 20$, $y = 2x$, $z = x + y$, $\sqrt{z}$

→ > x = 20

> y = 2*x

> z = x+y

> sqrt(z)

[1] 7.7459

3) $x = 10$, $y = 15$, $z = 5$

a) $x+y+z$ b) xyz c) $\sqrt{xyz}$ d) round ($\sqrt{xyz}$)

→ > x = 10

> y = 15

> z = 5

> x+y+z

[1] 30

> x * y * z

[1] 750

39

```
>sqrt(x*y*z)
[1] 27.38613
>round(sqrt(x*y*z))
[1] 27
```

• Vector: A vector in R software is denoted by the syntax (c):

1) (2,3,5,7)^2
```
→ >x=c(2,3,5,7)
>x^2
[1] 4 9 25 49
```

2) c(2,3,5,7)^c(2,3)
```
→ >a=c(2,3,5,7)
>b=c(2,3)
>a^b
[1] 4 27 25 49
```

3) c(2,3,5,7,9,11)^c(2,3)
```
→ >a=c(2,3,5,7,9,11)
>b=c(2,3)
>a^b
[1] 4 27 25 343 81 1331
```

4) c(1,2,3,4,5,6)^c(2,3,4)
```
→ >a=c(1,2,3,4,5,6)
>b=c(2,3,4)
>a^b
[1] 1 8 81 16 125 1296
```

40

5) c(2,3,5,7) * 3
```
→ >a=c(2,5,7,9,11)    or >c(2,3,5,7)*3
>a*3                     [1] 6 15 21 27.
[1] 6 15 21 27
```

6) c(2,3,5,7) × c(-2,-3,-5,-7)
```
→ >c(2,3,5,7) * c(-2,-3,-5,-7)
[1] -4 -9 -25 -49
```

7) c(5,6,7,8) +10
```
→ > c(5,6,7,8) +10
[1] 15 16 17 18.
```

8) c(5,6,7,8) + c(-2,-3,-1,0)
```
→ > c(5,6,7,8) + c(-2,-3,-1,0)
[1] 3.0 ~~1.5~~ 3 6 8
```

9) c(2,3,5,7) ÷ 2
```
→ > c(2,3,5,7)/2
[1] 1.0 1.5 2.5 3.5
```

• Sum, product :

Q. Find the sum, product, square root, for the following values.
4, 9, 2, 5, 7, 8, 3, 6, 15, 12, 10, 9, 8, 13, 14.

```
→ >x=c(4,9,2....14)
> length(x)
[1] 15
>y= sum(x)
>y
[1] 125
> b= prod(x)
>b
[1] 8.559323e+12
```

Scanned with CamSc

```
>sqrt(y)
[1] 11.18034
>sqrt(b)
[1] 2925632
```

Q. Find sum, product, square root, maximum & minimum values in the following.

```
→ >x=c(2,8,9,11,10,7,6)^2
>x
[1] 4 64 81 121 100 49 64 36
>sum(x)
[1] 455
>prod(x)
[1] 442597478400
>max(x)
[1] 121
>min(x)
[1] 4
```

• Matrix:

Q. $\begin{bmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{bmatrix}$

```
→ >x<-matrix(nrow=4, ncol=2, data=c(1,2,3,4,
        5,6,7,8))
>x
        [,1]  [,2]
[1,]      1     5
[2,]      2     6
[3,]      3     7
[4,]      4     8
```

Q. $X = \begin{bmatrix} 4 & 7 & 8 \\ 5 & 8 & 0 \\ 6 & 9 & 2 \end{bmatrix}$ $y = \begin{bmatrix} 6 & 11 & 9 \\ 4 & 12 & 7 \\ 5 & 8 & 4 \end{bmatrix}$

a) x+y   b) x*2   c) y*3, d) x*y

```
→ >x<-matrix(nrow=3, ncol=3, data=c(4,5,6,7,
   8,9,4, 0,2))
>y<-matrix(nrow=3, ncol=3, data=c(6,4,5,11,
   12,8,9,7,4))
>x
        [,1]   [,2]   [,3]
[1,]      4      7      8
[2,]      5      8      0
[3,]      6      9      2
>y
        [,1]   [,2]   [,3]
[1,]      6     11      9
[2,]      4     12      7
[3,]      5      8      4
>x+y
        [,1]   [,2]   [,3]
[1,]     10     18     13
[2,]      9     20      7
[3,]     11     17      6
>x*2
        [,1]   [,2]   [,3]
[1,]      8     14      8
[2,]     10     16      0
[3,]     12     18      4
```

> y * 3

	[,1]	[,2]	[,3]
[1,]	18	33	27
[2,]	12	36	21
[3,]	15	24	12

> x * y

	[,1]	[,2]	[,3]
[1,]	24	77	36
[2,]	20	96	0
[3,]	30	72	8

X

A.
2/1/19

Practical no: 2

Binomial Distribution:

n = Total no of trials

p = P(success)

q = P(failure)

x = No of successes out of n.

$P(x) = {}^{n}C_{x}\, p^{x} \cdot q^{n-x}$

$E(X) = np$ $V(X) = npq$.

Q1. Toss a coin 10 times with P(H) = 0.6. Let X be the no of heads. Find the probability of

① 7 heads ② 4 heads ③ Atmost 4 heads ④ Atleast six heads ⑤ No heads ⑥ All heads

Also find exception and variance.

```
> n = 10
> p = 0.6
> q = 0.4
> dbinom(7,10,0.6)
[1] 0.2149908
> dbinom(4,10,0.6)
[1] 0.1114767
> pbinom(4,10,0.6)
[1] 0.1662386
> 1-pbinom(6,10,0.6)
[1] 0.3822806
> dbinom(0,10,0.6)
[1] 0.0001048576
```

> dbinom(10,10,0.6)
[1] 0.006046618

>E=n*p

>E
[1] 6

>V=n*p*q

>V
[1] 2.4

Q.2 Suppose there are 12 MCQ in an English question paper. Each question have five answers only one of them is correct. Find the probability i) 4 correct answer 2) Atmost 4 correct answer 3) Atleast 3 correct answer.

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

Q.4. Find the probability of exactly 10 successes out of 100 trials with $p=0.1$

Q.5. X follows binomial distribution with $n=12$ and $p=0.25$. Find
1) $P(X \le 5)$ 2) $P(X > 7)$ 3) $P(5 < X < 7)$

Q.6. There are 10 members in a committee. Probability of any member attending a meeting is 0.9. What is the probability that 7 or more members will be present in a meeting.

Q.7. A salesman has a 20% probability of making a sell to a customer. On a typical day he will meet 30 customers. What minimum number of sales he will make with 88% probability

Q.8. For $n=10$ and $p=0.6$. Find the binomial probabilities and plot the graphs of pmf and cdf.

Note:
1) $P(X=x)$ = dbinom(x, n, p)
2) Probability of atmost x values
$P(X \le x)$ = pbinom(x, n, p)
3) Probability of atleast x values
$P(X \ge x)$ = 1- pbinom(x, n, p)
4) If x is unknown and the probability is given as p_1, to find x
qbinom(p1, n, p)

Answers.

Q.2.
> n=12
> p=1/5
> dbinom(4,n,p)
[1] 0.1328756
> pbinom(4,n,p)
[1] 0.9274445
> 1-pbinom(2,n,p)
[1] 0.4416543

Q.3.
```
> n=5
> p=0.1
> dbinom(0,n,p)
[1] 0.59049
> dbinom(1,n,p)
[1] 0.32805
> dbinom(2,n,p)
[1] 0.0729
> dbinom(3,n,p)
[1] 0.0081
> dbinom(4,n,p)
[1] 0.00045
> dbinom(5,n,p)
[1] 1e-05
```

Q.4.
```
> n=100
> p=0.1
> dbinom(10,n,p)
[1] 0.1318653
```

Q.5.
```
> n=12
> p=0.25
> pbinom(5,n,p)
[1] 0.9455978
> 1-pbinom(7,n,p)
[1] 0.00278151
> dbinom(6,n,p)
[1] 0.04014945
```

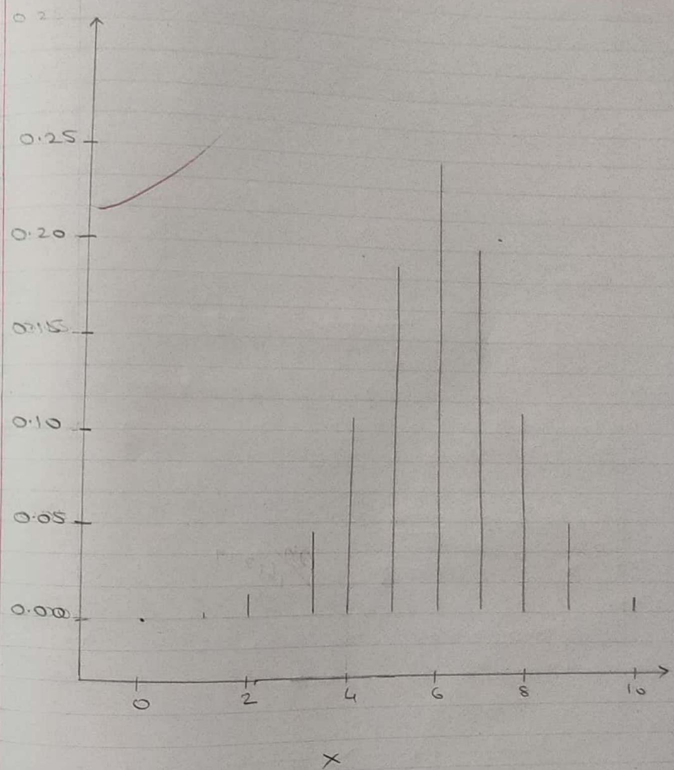
Q.6.
```
> n=10
> p=0.9
> 1-pbinom(6,n,p)
[1] 0.9872048
```

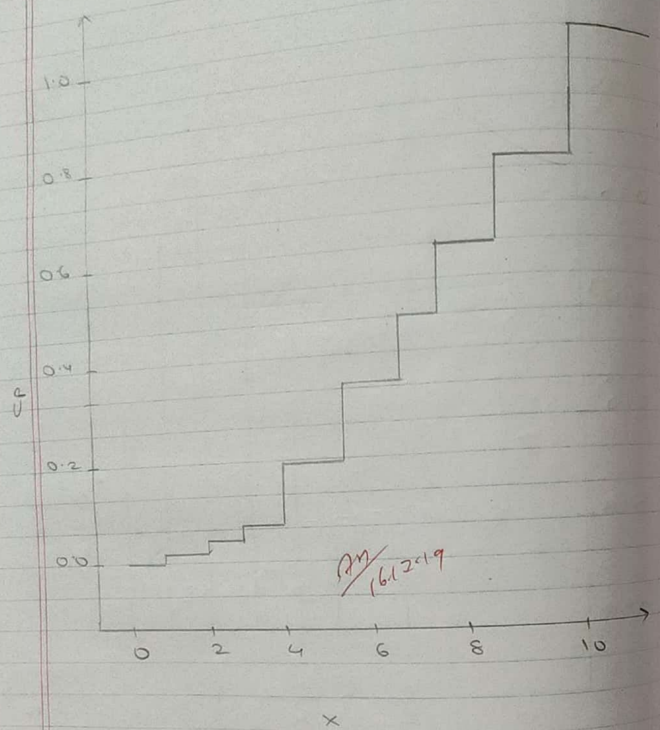
Q.7.
```
> n=30
> p=0.2
> qbinom(0.88,n,p)
[1] 9
```

Q.8.
```
> n=10
> p=0.6
> x=0:n
> bp=dbinom(x,n,p)
> d=data.frame("x.values"=x,"probability"=bp)
> d
```

	x.values	probability
1	0	0.0001048576
2	1	0.0015728640
3	2	0.0106168320
4	3	0.0424673280
5	4	0.1114767360
6	5	0.2006581248
7	6	0.2508226560
8	7	0.2149908480
9	8	0.1209323520
10	9	0.0403107840
11	10	0.0060466176

```
>plot(x, bp, "h")
>cp=pbinom(x, n, p)
>plot(x, cp, "s")
```





Practical no: 3

Q. check the following are pmf (probability mass function) or not:

1)

x	1	2	3	4	5
P(x)	0.2	0.5	-0.5	0.4	0.4

→ since the probability is negative
∴ It is not a pmf.

2)

x	10	20	30	40	50
P(x)	0.3	0.2	0.3	0.1	0.1

→ The condition for pmf is:
① $0 \le P(x) \le 1$ — ①
② $\Sigma Px = 1$ — ②

```
>prob = c(0.3, 0.2, 0.3, 0.1, 0.1)
>sum (prob)
[1] 1
```

since both the conditions are satisfied it is a pmf.

3)

x	0	1	2	3	4
P(x)	0.4	0.2	0.3	0.2	0.1

→ The condition to check pmf are:
① $0 \le P(x) \le 1$
② $\Sigma P(x) = 1$

```
> prob= c( 0.4, 0.2, 0.3, 0.2, 0.1)
> sum (prob)
[1] 1.2
```

Here the second condition does not satisfy
∴ It is not a pmf.

Q.2. Following is a pmf of X. Find mean & Variance.

x	1	2	3	4	5
P(x)	0.1	0.15	0.2	0.3	0.25

x	P(x)	x.P(x)	x^2 P(x)
1	0.1	0.1	0.1
2	0.15	0.3	0.6
3	0.2	0.6	1.8
4	0.3	1.2	4.8
5	0.25	1.25	6.25
		3.45	13.55

→ >x= c(1,2,3,4,5)
>prob = c(0.1, 0.15, 0.2, 0.3, 0.25)
>a=x*prob
>mean=sum(a)
>mean
[1] 3.45
>b=(x^2)*prob
>b
[1] 0.10 0.60 1.80 4.80 6.25
>var = sum(b) -mean ^2
>var
[1] 1.6475

Q.3. Following is a pmf of X. Find mean and variance.

x	5	10	15	20	25
P(x)	0.1	0.3	0.2	0.25	0.15

→ >x=c(5, 10, 15, 20, 25)
>prob = c(0.1, 0.3, 0.2, 0.25, 0.15)
>a = x * prob
>a
[1] 0.50 3.00 3.00 5.00 3.75
>mean=(sum(a)
>mean
[1] 35.25
>b = (x^2) * prob
>b
[1] 2.50 30.00 45.00 100.00 93.75
>var = sum(b) - mean ^2
>var
[1] 38.6875

Q.4. Find cdf of the following pmf and draw the graph of cdf.

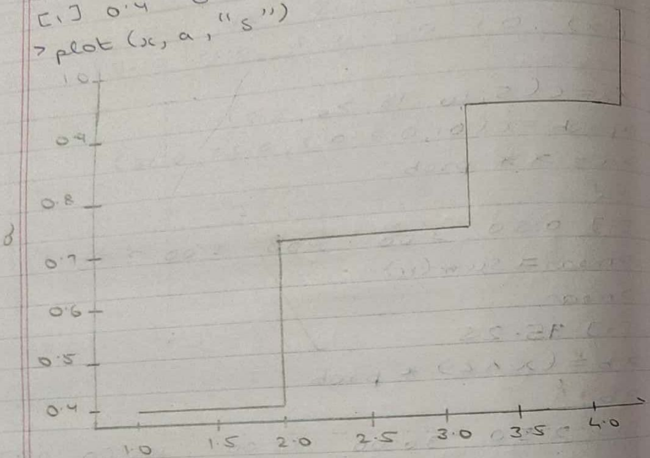
1)

x	1	2	3	4
P(x)	0.4	0.3	0.2	0.1

→ >x=c(1, 2, 3, 4)
>prob=c(0.4, 0.3, 0.2, 0.1)

Scanned with CamSc

```
> a = cumsum(prob)
> a
[1] 0.4   0.7   0.9   1.0
> plot(x, a, "s")
```

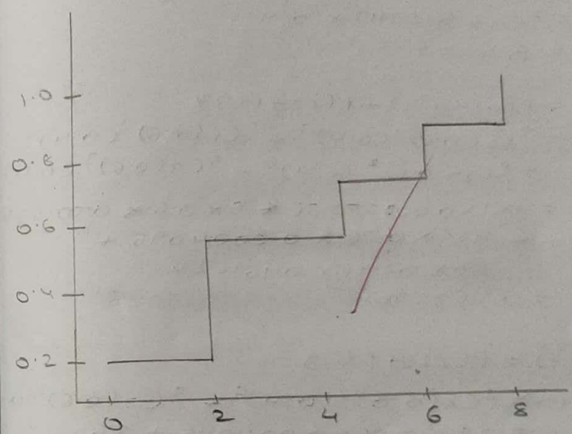


$F(x) = 0$	$x \leq 1$
$= 0.4$	$1 \leq x < 2$
$= 0.7$	$2 \leq x < 3$
$= 0.9$	$3 \leq x < 4$
$= 1.0$	$x \geq 4$

2)

x	0	2	4	6	8
$P(x)$	0.2	0.3	0.2	0.2	0.1


```
> x = c(0, 2, 4, 6, 8)
> prob = c(0.2, 0.3, 0.2, 0.2, 0.1)
> a = cumsum(prob)
> a
[1] 0.2   0.5   0.7   0.9   1.0
> plot(x, a, "s")
```



$F(x) = 0$	$x \leq 0$
$= 0.2$	$0 \leq x \leq 2$
$= 0.5$	$2 \leq x < 4$
$= 0.7$	$4 \leq x < 6$
$= 0.9$	$6 \leq x < 8$
$= 1.0$	$x \geq 8$

Practical no: 4

Practise:

Q. X follows binomial distribution with $p = 0.6$, $q = 0.4$. Find
a) $P(X=7)$ b) $P(X \le 3)$ c)

$\rightarrow P(X=7) = \binom{8}{7} (0.6)^7 (0.4)^{8-7}$
$= {}^8C_7 \times 0.2799 \times 0.4$
$= 0.8957$

$P(X \le 3) = P(0) + P(1) + P(2) + P(3)$
$= {}^8C_0 (0.6)^0 (0.4)^8 + {}^8C_1 (0.6)^1 (0.4)^7$
$+ {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^3 (0.4)^5$
$= 1 \times 1 \times 0.00065536 * 8 \times 0.6 * 0.0016384$
$+ 428 \times 0.36 \times 0.0004096 +$
$56 \times 0.216 \times 0.101024$
$= 0.1736704$

$P(X = 2 \text{ or } 3) = P(2) + P(3)$
$= {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^3 (0.4)^5$
$= 28 \times 0.36 \times 0.004096 + 56 \times$
0.216×0.01024
$= 0.04128768 + 0.12386304$
$= 0.16515012$

——— X ———

Practical no: 5

Normal distribution:

1) Normal distribution is an example of continuous probability distribution.
$X \sim N(\mu, \sigma^2)$

a) $P(X = x) =$ dnorm (x, μ, σ)
b) $P(X \le x) =$ pnorm (x, μ, σ)
c) $P(X > x) = 1 -$ pnorm (x, μ, σ)
d) To find the value of K so that ~~find the~~ the command is qnorm (p, μ, σ)
K $P(X \le K) = p$
qnorm (p, μ, σ)
e) To generate a random sample of size n,
rnorm (n, μ, σ)

Q.1. A random variable $X \sim N(10, 2)$. Find
1) $P(X \le 7)$ 2) $P(X > 12)$ 3) $P(5 \le X \le 12)$
4) $P(X < K) = 0.4$ with $\mu = 10$, $\sigma = 2$

Q.2. $X \sim N(100, 36)$ $\sigma = \text{sqrt}(36)$
1) $P(X \le 110)$ 2) $P(X > 105)$ 3) $P(X \le 92)$
4) $P(95 \le X \le 110)$ 5) $P(X < K) = 0.9$

Q.3. Generate 10 random sample and find the sample mean, median and variance, standard deviation.

Scanned with CamSc

49

Q.1. >m=10
>s=2
>p1= pnorm (7,m,s)
> cat ("P(X<=7) is =", p1)
P(X<=7) is = 0.0668072 >
>p2= 1- pnorm (12,m,s)
> cat ("P(X>12) is =", p2)
P(X>12) is = 0.1586553 >
> p3 = pnorm (12,m,s) - pnorm(5,m,s)
> cat ("P(5<=X<=12) is =", p3)
P(5<=X<=12) is = 0.8351351.
>k= qnorm (0.4, m,s)
> cat ("P(X<K)=0.4, K is =", k)
P(X<K) =0.4, K is = 9.493306 >

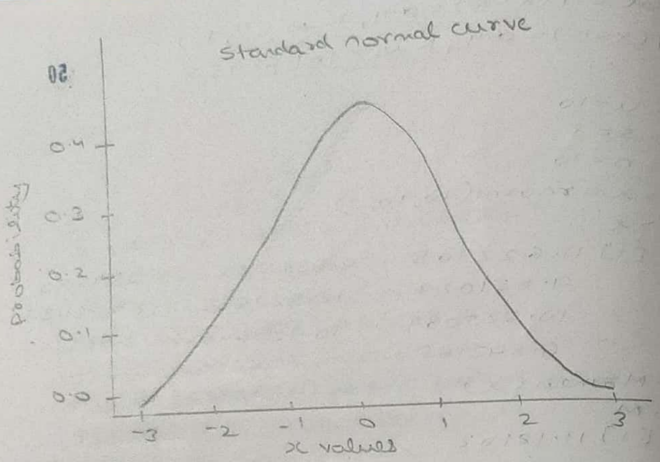
Q.2 >m=100
>s= sqrt (36)
>p1 = pnorm (110,m,s)
> cat ("P(X<=110) is =", p1)
P(X<=110) is = 0.9522096 >
> p2 = 1- pnorm(105,m,s)
>cat ("P(X>105) is =", p2)
P(X>105) is = 0.2023284>
> p3 = pnorm (92,m,s)
> cat ("P(X<=92) is =", p3)
P(X<=92) is = ~~0.0912~~ 0.09121122>
>p4= pnorm (110, m,s) - pnorm (95,m,s)
>cat ("P(95<=X<=110) is =", p4)
P(95<=X<=110) is = 0.7498813>

> k = qnorm(0.9, m,s)
> cat ("P(X<K) =0.9, K is =", K)
P(X<K) = 0.9, K is = 107.6893 >

Q.3 > u=10
> s = 3
> n=10
> x = rnorm(10,10,3)
> x
[1] 11.622705 9.686588 18.259680
9.051027 12.327658 13.706428
10.825069 10.727646 8.761215
6.842165
> M=mean (x)
> M
[1] 11.18102
> m = median (x)
> m
[1] 10.77636
> v = (n-1) * (var (x) /n)
> v
[1] 8.977855
> sd = sqrt (v)
> sd
[1] 2.996307.

Q.4. Plot the standard normal curve.
> x = seq (-3, 3, by =0.1)
> y = dnorm (x)
> plot (x, y, xlab = "x values", ylab = "probability", main = "standard normal curve")

Scanned with CamSc



Q.5. X~N(50, 100)
Find ① P(X ≤ 60) ② P(X > 65)
③ P(45 ≤ X ≤ 60)

→
```
> u=50
> s=sqrt(100)
> s
 [1] 10
> p1=pnorm(60,u,s)
> p1
 [1] 0.8413447
> p2 = 1-pnorm(65,u,s)
> p2
 [1] 0.668072
> p3=pnorm(60,u,s) - pnorm(45,u,s)
> p3
 [1] 0.5328072
```

18/1/25

———X———

Practical no: 6

Topic: Z and t distribution sums.

Q.1. Test the hypothesis Ho: μ = 20 against H1: μ ≠ 20 a sample of size 400 is selected and the sample mean is 20.2. and a standard deviation 2.25. Test that 5% level of significance.

→
```
> m0 = 20
> mx = 20.2
> sd = 2.25
> n = 400
> zcal = (mx - m0)/ (sd/sqrt(n))
> zcal
 [1] 1.777778
> cat ("z calculated is=", zcal)
 z calculated is= 1.777778>
> pvalue = 2 * (1-pnorm(abs(zcal)))
> pvalue
 [1] 0.07544036
```
since pvalue is more than 0.05, we accept Ho. μ = 20.

Q.2. we want to test the hypothesis Ho: μ = 250 against Ho: μ ≠ 250 a sample of size 100 has a mean of 275 and sd 30. Test the hypothesis at 5% level of significance.

```
>mo=250
>mx=275
>sd=30
>n=100
>zcal = (mx-mo)/(sd/sqrt(n))
>zcal
[1] 8.333333
>cat(" z calculated is = ", zcal)
z calculated is = 8.333333 >
>pvalue= 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.
```

Since pvalue is less than 0.05 we reject Ho.

Q.3. We want to test the hypothesis Ho: P=0.2 against H1: P ≠ 0.2 (P= population proportion) A sample of 400 is selected and the sample proportion is calculated 0.125. Test the hypothesis at 1% level of significance.

```
>P=0.2
>Q=1-P
>p=0.125
>n=400
>zcal= (p-P)/sqrt(P*Q/n))
>zcal
[1] -3.75
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.0001768346
```


Since pvalue is less than 0.05 we reject Ho.

Q.4. In a big city 325 men out of 600 men where found to be self-employed. Thus this information support the conclusion that exactly that half of the men are self employed

```
>P=0.5
>n=600
>p=325/600
>Q=1-P
>zcal=(p-P)/sqrt(P*Q/n))
>zcal
[1] 2.041241
>pvalue= 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.04122683
```

Since pvalue is less than 0.05 we reject Ho.

Q.5. Test the hypothesis of Ho: μ=50 against H: μ≠50. A sample of 30 is collected -
50, 49, 52, 44, 45, 48, 46, 45, 49, 45, 40, 47, 55, 54, 46, 58, 47, 44, 59, 60, 61, 41, 52, 44, 55, 56, 46, 45, 48, 49

→ >mo=50
>x=c(50,49,52....... 48,49)
>n= length(x)   >n   [1] 30
>mx=mean(x)   >mx   [1] 49.3333
>variance=(n-1)* var(x)/n   [1] 30.45556
>sd= sqrt(variance)   >sd   [1] 5.563772
>zcal=(mx-mo)/sd/(sqrt(n))
>zcal
[1] -0.6562964
>pvalue= 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.5116234

since pvalue is more than 0.05 we accept Ho: μ=50.

———X———

Practical no: 7

Topic: Large sample test.

Q.1. Two random sample of size 1000 & 2000 are drawn from two population with a standard deviation 2 & 3 respectively. Test the hypothesis that the two population means are equal or not at 5% level of significance. Sample means are 67 and 68 respectively.

Ho: = ?

→ >n1= 1000
>n2 = 2000
>mx1= 67
>mx2 = 68
>sd1= 2
>sd2 =3
>zcal=(mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2/n2))
>cat("z calculated is =", zcal)
z calculated is =-10.84652
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0

Since pvalue is less than 0.05 we reject Ho: μ1=μ2

Q.2. A study of noise level in 2 hospitals is done following data is calculated. First sample size 64, first sample mean 61.2, First standard deviation 7.9. Second sample size =34, Second sample mean = 59.4, second standard deviation = 7.8. Test Ho. $\mu_1 = \mu_2$ at 1% level of significance.

```
→ >n1=84
>n2=34
>mx1= 61.2
>mx2=59.4
> sd1= 7.9
>sd2= 7.8
>zcal= (mx1- mx2)/ sqrt ((sd1^2/n1) +
    (sd2^2/n2))
>cat ("z calculated is =", zcal )
z calculated is = 1.131717>
>pvalue = 2 * ( 1- pnorm( abs(zcal)))
> pvalue
[1] 0.258006.
```

since pvalue is greater than 0.01, we accept Ho $\mu_1 = \mu_2$

Q.3. From each of two population of oranges the following samples are collected. Test whether the proportion of bad oranges are equal or not. First sample size = 250 second sample size = 200; No of bad in the first sample = 44 and second sample = 30.

Ho: $P_1 = P_2$ against $H_1: P_1 \neq P_2$


```
→ >n1= 250
>n2= 200
>p1= 34/250
>p2= 30/200
>P=(n1 * p1 + n2 * p2)/(n1 + n2)
>Q = 1 - P
> zcal = (p1- p2)/ sqrt (P * Q * (1/n1+1/n2))
> zcal
[1] 0.7393581
> pvalue = 2 * (1-pnorm( abs(zcal)))
>pvalue
[1] 0.4596896.
```

since P value is greater than 0.05, we

Q.4. Random sample of 400 males and 600 females when asked whether they want the ATM nearby. 200 males and 390 females were in favor of the proposal. Test the hypothesis ~~that~~ the proportion of males and females favouring the proposal are equal or not at 5% level of significance.

Q.5. ~~Following are two~~ Following are the two independent sample from the two population Test ~~the~~ equality of two means at 5% level of significance.

n1 = 74, ~~77~~, ~~~~ = 79, 77, 76, 82, 72, 75, 78, 77, 78, 76, 76, 74.

n2 = 72, 76, 74, 70, 70, 78, 70, 72, 75, 79, 79, 74, 75, 78, 72, 74, 80

4. >n1=400
>n2=600
>p1= 200/400
>p2 =390/600
>P= (n1*p1+n2*p2)/(n1+n2)
>Q=1-P
>zcal =(p1-p2)/sqrt(P*Q*(1/n1+1/n2))
>zcal
[1]-4.724751
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.303972e-06
since pvalue is less than 0.05 we reject Ho.p1=p2
5. >x1=c(74,77,79,77.....74)
>n1= length(x1)   [1] 14
>mx1 = mean(x1)
>variance1= (n1-1)*var(x1)/n1 [1] 0.4508
>sd1= sqrt(variance1) [1] 0.6714
>x2=c(72,76,...74,80)
>n1= length(x2)
>mx2 = mean(x2)
>variance2=(n2-1)*var(x2)/n2 [1] 0.6163
>sd2=sqrt(variance2) [1] 0.7850
>tcal = (mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2/n
>tcal
[1] 1.5781
>t.test(x1,x2)
pvalue= 0.1387
>0.05 is accepted

27.01.20

———X———

Practical no: 8

Topic: Small sample test.

Q.1. The random sample of 15 observations are given by
80, 100, 110, 105, 122, 70, 120, 110, 101, 88, 83, 95, 89, 107, 125. Do these data support the assumption that the population mean is 100. $H_0: \mu = 100$

>x=c(80,100, 110, ... 107, 125)
>length(x)
[1] 15
>t.test(x)
One Sample t-test
data: x
t = 24.029, df=14, p-value = 8.819e-13
alternative hypothesis: true mean is not equal to 0 95 percent confidence interval:
91.37775   109.28892
sample estimates:
mean of x:
100.3333
since pvalue is less than 0.05 we reject Ho. $\mu_0 = 100$ at 5% level of significance

Q.2 Two groups of 10 students scored the following marks
group 1 – 18, 22, 21, 17, 20, 17, 23, 20, 22, 21.

22

group 2 - 16, 20, 14, 21, 20, 18, 13, 15, 17, 21.
Test the hypothesis that there is no significant difference between the scores at 1% level of significance. $H_0: \mu_1 = \mu_2$

```
> x = c( 18, 22, 21, . . . . 22, 21)
> length (x)
[1] 10
> y = c( 16, 20, 14, . . . . 17, 21)
> length (y)
[1] 10
> t.test (x, y)
        welch Two sample t-test
data : x and y
t = 2.2573, df = 16.376, p-value = 0.03792
alternative hypothesis: true difference in means 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
```

Since pvalue is more than 0.01 we ~~reject~~ accept H_0 at 1%

Q.3 Two types of medicines are used on 5 and 7 patients for reducing their weight. The decrease in the weight after using the medicines are given below.

Med A = 10, 12, 13, 11, 14
Med B = 8, 9, 12, 14, 15, 10, 9
Is there a significant difference in the efficiency of the medicines. $H_0: \mu_1 = \mu_2$

```
> x = c(10, 12, 13, 11, 14)
> y = c(8, 9, 12, 14, 15, 10, 9)
> length(x)
[1] 5
> length(y)
[1] 7
> t.test(x, y)
        welch Two sample t-test
data : x and y
t = 0.50384, df = 9.7594, p-value = 0.4406
alternative hypothesis : true difference in means is not equal to 0  95 percent confidence interval :
 -1.781171   3.781171
sample estimates :
mean of x   mean of y
   12          11
```

Since pvalue is more than 0.05 we accept $H_0: \mu_1 = \mu_0$ at 1% level of significance

Q.4 The weight reducing diet program is conducted & the observations are noted for 10 participants. Test whether the program is effective or not.

Before - 120, 125, 115, 130, 123, 119, 122, 127, 122, 118.

After - 111, 114, 107, 120, 115, 112, 112, 120, 119, 112.

H0. There is no significant difference in weight against H1. the diet program reduced weight.

→ > x = c(120, 125, ... 118)
> length(x)
[1] 10
> y = c(111, 114, ... 112)
> length(y)
[1] 10.
> t.test(x, y, paired = T, alternative = "less")

paired t-test

data: x and y
t = 17, df = 9, pvalue = ~~3.787e-08~~ 1
alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval; -Inf 9.416556
~~7.368921 9.631079~~
sample estimates:
mean of the differences
8.5

Since pvalue is greater tha 0.05 we accept H0: ~~H0~~ = the diet program reduced weight

Q.5. Sample A = 66, 67, 75, 76, 82, 84, 88, 90, 92.
Sample B = 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97.
Test the population mean are equal or not.

Q.6 The following are marks before & after are training program. Test the program is effective whether or not.
before - 71, 72, 74, 69, 70, 74, 76, 70, 73, 75.
after - 74, 77, 74, 73, 79, 76, 82, 72, 75, 78.

Q.5. > x = c(66, 67, 75, ... 92)
> length(x)
[1] 9
> y = c(64, 66, 74, ... 97)
> length(y)
[1] 11
> t.test(x, y)
Welch Two sample t-test
data: x and y
t = -0.63968, df = 17.974, p-value = 0.5304
alternative hypothesis: true difference in means is not equal to 0 95 percent
Confidence interval:
-12.853992 6.853992
Sample estimates:
mean of x ~~and~~ mean of y
80 83

since pvalue is greater than 0.05 we accept x and y at 1% level of significance.

Q.6: >x = c(74, 72, ... 75)
>y = c(74, 77, ... 78)
>t.test (x, y, paired =T, alternative = "greater")

Paired t-test

data: x and y
t = -4.4691 , df = 9, p-value = 0.9992
alternative hypothesis : true difference in means is greater than 0 95 percent confidence interval :
-5.076639 Inf
Sample estimates :
mean of the difference
-3.6

Since the pvalue is greater than 0.05 we accept Ho: μo = increase in marks at 5% level of significance.

AM 03.02.20

Practical no: 9

Topic: Large and small sample test.

Q.1. The arithmetic mean of a sample of 100 items from a large population is 52. If the standard deviation is 7. Test the hypothesis that the population mean is 55 against the alternative it is more than 55 at 5% LOS.

Q.2. In a big city 350 out of 700 males are found to be smokers. Thus these information supports that exactly half of the males in the city are smokers? Test that 1% LOS.

Q.3. Thousand articles from a factory A are found to have two 2% defectives. 1500 articles from a second factory B are found to have 1% defectives. Test that 5% LOS that the two factories are similar or not.

Q.4. A sample of size 400 was drawn at a ~~cont~~ sample mean is 99. Test at 5% LOS ~~of the~~ that the sample comes from a population with mean 100 and variance 64.

Q.5. The flower strings are selected and the heights are found to be (centimetres) 63, 63, 68, 69, 71, 71, 72. Test the hypothesis that the mean height is 66 or not at 1% LOS.

Q.6. Two random samples were drawn from two normal populations and their values are:
A - 66, 67, 75, 76, 82, 84, 88, 90, 92.
B - 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97.
Test the whether the populations have the same variance at 5% LOS.

Solutions

Q.1. Ho. μ = 0.05

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

Since pvalue is less than 0.05 we reject

Q.2. Ho. μ = 0.01

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



```
[1] 1
```

Pvalue is greater than 0.01 we accept ho.P = 0.5

Q.3 ho: P1 = P2 against h1 = P1≠P2

```
> n1= 1000
> n2=1500
> p1= 0.02
> p2 = 0.01
> p = (n1* p1+n2 * p2)/(n1+n2)
> p
[1] 0.014
> q=1-p
> q
[1] 0.986
> zcal = (p1+p2)/sqrt(p*q*(1/n1+1/n2))
> zcal
[1] 2.084842
> pvalue = 2* (1-pnorm (abs (zcal)))
> pvalue
[1] 0.037083 64.
```

Pvalue is less than 0.05 we reject ho: P1 = P2

Q.6.

```
> x= c(66, 67, ... 92)
> y = c(64, 66, .... 97)
> f = var. test(x, y)
> f
```

f test to compare two variances
data: a and b

F = 0.70686, num df = 8, denom df = 10,
p-value = 0.6359
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.1833662 3.0360393
sample estimates:
ratio of variances
0.7068567.
pvalue is greater than 0.05 we reject Ho: μ_0 =
The population have same variance

Ho: $\mu_1 = \mu_2$

Q.5. > x = c(63, 63 . . . . 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
pvalue is less than 0.01 we reject Ho: $\mu = \mu_1$

Ho: μ = 03.

Q.4. > n = 400
> mx = 99
> mx1 = 99
> mx2 = 100

> var = 64
> sd = sqrt(var)
> zcal = (mx1 - mx2) / sqrt(sd ^ 2 / n)
> zcal
[1] - 2.5
> pval = 2 * (1 - pnorm(abs(zcal)))
> pval
[1] 0.01241933
pvalue is less than 0.05 we reject H: μ = 0.5

AM
10.2-20

Practical no: 10

Topic: ~~ANOVA~~ & chi-square

Q.1. use the following data to test whether the cleanliness of home & cleanliness of child independent or not.

		Cleanliness of home Clean	Dirty
cleanliness of child	Clean	70	50
	Fairly clean	80	20
	Dirty	35	45

→ Ho: Cleanliness of child and cleanliness of home is independent.

```
> x=c(70,80,35,50,20,45)
> m=3
> n= 2
> y=matrix(x,nrow=m, ncol=n)
> y
      [,1]  [,2]
[1,]   70    50
[2,]   80    20
[3,]   35    45
> pv=chisq.test(y)
> pv
        Pearson's chi-squared test
data: y
X-squared = 25.646 , df = 2, p-value = 2.698e-06
```

Since p-value is less than 0.05, we reject Ho: CC & CH are independent.

Q.2. use the following data to find a vaccination on a particular disease are independent or not.

		Disease Affected	Not affected
Vac	Given	20	30
	Not given	25	35

Ho: Disease and vaccination is independent.

```
→ > x = c(20,25,30,35)
> m= 2
> n= 2
> y= matrix(x, nrow=m, ncol=n)
> y
       [,1]     [,2]
[1,]    20       30
[2,]    25       35
> pv = chisq.test(y)
> pv
        Pearson's chi-squared test
data: y
X-squared = 0, df = 1, p-value = 1
```

Since p-value is more than 0.05, we accept Ho: Vaccination & diseases are independent.

Q.3. Perform ANOVA for the following data:

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

H0: The means of the varities are equal

```
> 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 (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)
> anova
Call :
   aov (formula = values ~ ind, data = d)
Terms:
                  ind        Residuals
Sum of squares    71.06410   18.16667
Deg of Freedom    3          9
Residual standard error: 1.420746
Estimated effects may be unbalanced.
```

Since p-value is less than 0.05, we reject H0: The means of the varities are equal.

Q.4. The following data gives life of tyre of 4 brands

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

Test the hypothesis that the average life of 4 brands are same.

H0: average life of 4 brands of tyre are same.

```
> x1 = c( 20, 23, 18, 17, 18, 22, 24)
> x2 = c( 19, 15, 17, 20, 16, 17)
> x3 = c( 21, 19, 22, 17, 20)
> x4 = c( 15, 14, 16, 18, 14, 16)
> d = stack ( list (b1 = x1, b2 = x2, b3 = x3,
        b4 = x4))
> names (d)
[1] "values"  "ind"
```

```
>oneway.test(values ~ ind, data = d, var.equal = T)

        One-way analysis of means

data: values and ind
F = ~~4.0364~~ 6.58445, num df = 3, denom df = 20,
p-value = 0.002349
>anova = aov(values ~ ind, data = d)
>~~aov~~ anova
Call:
   aov(formula = values~ind, data = d)
Terms:
```

	ind	Residuals
Sum of squares	91.4381	89.0619
Deg of Freedom	3	20

Residual standard error: 2.110236

Estimated effects may be unbalanced.

Since p-value is less than 0.05 we reject H0: The average life of 4 types of tyres are same.

Q.5. One thousand students of a college are graded according to their IQ & the economic condition of their home. Check that is their any association between IQ & economic condition of their home.

		IQ	
		high	low
Economic condition	high	460	140
	medium	330	200
	low	240	160

H0: IQ and economic condition are independent.

```
>x = c(460, 330, 240, 140, 200, 160)
>m = 3
>n = 2
>y = matrix(x, nrow = m, ncol = n)
>y
        [,1]    [,2]
[1,]    460     140
[2,]    330     200
[3,]    240     160
>pv = chisq.test(y)
>pv

        pearson's chi-squared test

data: y
X-squared = 39.726, df = 2, p-value = 2.364e-09
```

Since p-value is less than 0.05 we reject H0: EC & IQ.

AM 17.7.20

Practical no: 11

Topic: Non-parametric test.

Q.1. Following are the amount's of sulphur oxide emitted by industries in 20 days. Apply sign test to test the hypothesis that the population median is 21.5.
17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26.
Ho: pop median is 21.5

→
```
>x=c(17,15,20,........24,26)
>length (x)
[1] 20
>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
```
Since p-value more than 0.05, we accept Ho at 5% level of significance.

Q.2. Following are the 10 observations 612, 619, 631, 628, 643, 640, 655, 649, 670, 663.

Apply sign test to test the hypothesis that the population median is 625 against the alternative it is greater than 625 at 1% LOS.

Ho: pop median is 625.

Note: If a alternative is greater than pv = pbinom(sn, n, 0.5)

→
```
>x=c(612, 619, 631, 628, 643, 640, 655, 649, 670, 663)
>length (x)
[1] 10
>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
```
Since pvalue is more than 0.05 we accept Ho at 5% LOS.

Q.3. Ten observations are:
36, 32, 21, 30, 24, 25, 20, 22, 20, 18.
using sign test to test hypothesis that the population median is 25 against the alternative it is less than 25 at 5% LOS.

>x = c(36, 32, ... 20, 18)
>length(x)
[1] 10
>me = 25
>sp = length(x[x>me])
>sn = length(x[x<me])
>n = sp+sn
>n
[1] 9
>pv = pbinom(sp, n, 0.5)
>pv
[1] 0.2539063

Since p-value is more than 0.05 we accept Ho at 5% LOS.

Q.4. The following are some measurements 63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 65. Using wilcoxon signed ~~sig~~ rank test. Test the hypothesis that the population median is 60 against the alternative it is greater than 60 at 5% LOS.
Ho: pop median is 60.

>x = c(63, 65, ... 39, 72, 65)
> wilcox.test(x, alt="greater", mu=60)
wilcoxon signed rank test with continuity correction
data: x
V = 68, p-value = 0.06186
alternative hypothesis: true location is greater than 60

since p-value is more than 0.05 we accept Ho at 5% LOS.

Q.5. 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26. We want to test the hypothesis that the population median is 20 against the alternative it is greater than 20 at 5% LOS.

>x = c(15, 17, ... 24, 26)
>length(x)
[1] 12
>wilcox.test(x, alt="less", mu=20)
wilcoxon signed rank test with continuity correction
data: x
V = 48.5, p-value = 0.9232
alternative hypothesis: true location is less than 20.
Since p-value is more than 0.05 we accept Ho at 5% LOS.

Q.6. 20, 25, 27, 30, 18. Test the hypothesis that the population median is 25 against the alternative it is not 25.

>x = c(20, 25, 27, 30, 18)
>length(x)
[1] 5
>wilcox.test(x, alt="two.sided", mu=25)

wilcoxon signed rank test with continuity correction

data : x

v = 3.5 , p-value = 0.7127

alternative hypothesis : true location is not equal to 25

since p-value is more than 0.05 we accept Ho at 5% LOS.

9.3.20