★★ INDEX ★★

IInd Semister.

Source code:-

Q.1.

1) (4+6+8)/2-5

>>> 9

2) 2^2 + abs(-3) + sqrt(45)

>>> 13.7080

3) 5^3 + 7x5x8 + 46/5

>>> 414.2

4) sqrt(4^2 + 5x3 + 7/6)

>>> 5.671567

5) round (46/7 + 9x8)

>>> 7.9.

Q.2.

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) c(2,3,5,7) * c(2,3,6,2)

>>> 4 9 30 14

4) c(1,6,2,3) * (-2,-3,-4,-1)

>>> -2 -18 -8 -3

5) c(2,3,5,7) ^ 2

>>> ~~4 36 512 9 16 125~~

>>> 4.92549

6) c(4,6,8,9,4,5) ^ c(1,2,3)

>>> 4 36 512 9 16 125

7) c(6,2,7,5) / c(4,5)

>>> 1.50 0.40 1.75 1.00

PRACTICAL:-1

Basic of R Software.

* R is a Software for Statical analysis and data computy.
* It is an effective data handling Software and outcome Storage is possible.
* It is capable of graphical display.
* It's a free Software

Q1. Solve the following:-

1) $4+6+8 \div 2 - 5$.
2) $2^2 + |-3| + \sqrt{45}$
3) $5^3 + 7\times 5\times 8 + 46 \div 5$
4) $\sqrt{4^2 + 5\times 3 + 7/6}$
5) round of $(46 \div 7 + 9\times 8)$

Q2 Vector Calculation:-

1) c(2,3,5,7) * 2.
2) c(2,3,5,7) * c(2,3)
3) c(2,3,5,7) * c(2,3,6,2)
4) c(1,6,2,3) * c(-2,-3,-4,-1)
5) c(2,3,5,7) * ^ 2.
6) c(4,6,8,9,4,5) ^ c(1,2,3)
7) c(6,2,7,5) / c(4,5)

Scanned with CamScanner

Q.3.

1) $x = 20$
$y = 30$
$z = 2$, find (i) $x^2 + y^2 + z$
(ii) $\sqrt{x^2 + y}$
(iii) $x^2 + y^2$.

Q.4. Draw Matrix.

→ data = 1,2,3,4,5,6,7,8.

Q.5 Find X+Y and 2X+3Y when X

$$X = \begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 9 \end{bmatrix}$$

$$Y = \begin{bmatrix} 10 & -5 & 7 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$$

Q.3.

1)
→ x = 20
y = 30
z = 2

(i) x^2 + y^2 + z
>>> 27402

(ii) sqrt(x^2 + y)
>>> 20.73644

(iii) x^2 + y^2
>>> 1300.

Q.4.
→ x = Matrix (nrow = 4, ncol = 2, data = c(1,2,3,4,5,6,7

>>> x
>>>

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

Q.5.
→ x = Matrix (nrow = 3, ncol = 3, data = c(4,7,9,-2,0,
6,7,3))

>>> x
>>>

	[,1]	[,2]	[,3]
[1,]	4	-2	6
[2,]	7	0	7
[3,]	9	-5	9

Scanned with CamScanner

```
y=matrix(nrow=3,ncol=3,data=c(10,12,15,-5,-4,-6,7,
   9,5))
>>> y
>>>      [,1] [,2] [,3]
   [1,]  10   -5    7
   [2,]  12   -4    9
   [3,]  15   -6    5
>>> x+y
>>>      [,1] [,2] [,3]
   [1,]  14   -7   13
   [2,]  19   -4   16
   [3,]  24  -11    8
>>> 2*x+3*y
>>>      [,1] [,2] [,3]
   [1,]  38  -19   33
   [2,]  50  -12   41
   [3,]  63  -18   21
```

Q.6

```
->>x=c(59,20,35,24,46,56,55,45,27,22,47,58,54,40,
   50,32,36,29,35,39)
>>> breaks=seq(20,60,5)
>>> a=cut(x,breaks,right=FALSE)
>>> b=table(a)
>>> c=transform(b)
>>> c
        a      Freq
>>> [20,25)    3
    [25,30)    2
    [30,35)    1
    [35,40)    4
    [40,45)    0
    [45,50)    3
    [50,55)    0
    [55,60)    4
```


Q.6. Mark of Statistic of computer Science student of A-Batch out 60.

-> Marks = 59, 20, 35, 24, 46, 56, 55, 45, 27, 22, 47, 58, 54, 40, 50, 32, 36, 29, 35, 39.

A-m
5.12.19

11

PRACTICAL-02.

Probability Distribution.

Q.1 Check whether the following pmfs:-

1) x = 0 1 2 3 4 5
P(x) = 0.1 0.2 -0.5 0.4 0.3 0.5

→ ∵ P(2) = -0.5, can't be a probability mass function.
∴ PMF ≥ 0.

2) x = 1 2 3 4 5
P(x) = 0.2 0.2 0.3 0.2 0.2.

→ It cannot be a PMF, as in PMF,
$\Sigma P(x) = 1$.

3) x = 10 20 30 40 50
P(x) = 0.2 0.2 0.35 0.15 0.1

→ It is pmf, because $\Sigma P(x) = 1$.

Q.2

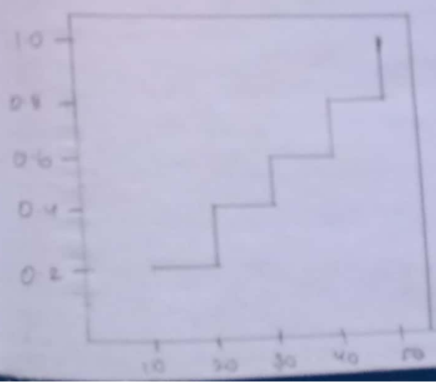
1) Find cdf for the following pmf and sketch the graph.
x = 10 20 30 40 50
P(x) = 0.2 0.2 0.35 0.15 0.1

1)
```
>>> Prob = c(0.2, 0.2, 0.3, 0.2, 0.2)
>>> Sum(Prob)
[1] 1.1
```

2)
```
> Prob = c(0.2, 0.2, 0.35, 0.15, 0.1)
> sum(Prob)
[1] 1
```

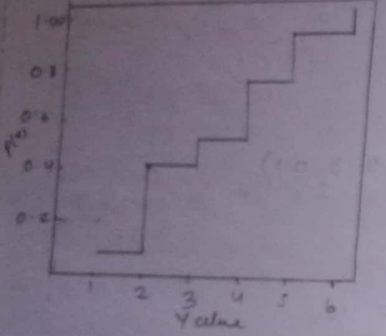
Q.2

1)
```
> Prob = c(0.2, 0.2, 0.35, 0.15, 0.1)
> Sum(Prob)
[1] 1
> cumsum(Prob)
[1] 0.20 0.40 0.75 0.90 1.00
> x = c(10, 20, 30, 40, 50)
> Plot(x, cumsum(Prob), "s", xlab="Value", ylab="P(x)")
```



2)

```
→ > 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 01.00
> x = c(1,2,3,4,5,6)
> Plot(x, cumsum(Prob), "S", xlab = "value", ylab = "P(x)"
```



$F(x) = 0 \quad x < 10$

$= 0.2 \quad 10 \leq x < 20$

$= 0.4 \quad 20 \leq x < 30$

$= 0.75 \quad 30 \leq x < 40$

$= 0.90 \quad 40 \leq x < 50$

$= 1.00 \quad x \geq 50.$

2) Find cdf for the following pmf and sketch the graph:-

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

→ $F(x) = 0 \quad x < 1$

$= 0.15 \quad 1 \leq x < 2$

$= 0.40 \quad 2 \leq x < 3$

$= 0.50 \quad 3 \leq x < 4$

$= 0.70 \quad 4 \leq x < 5$

$= 0.90 \quad 5 \leq x < 6$

$= 1.00 \quad x \geq 6$

Q.3. Check whether the following is p.d.f or not :-

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

→ $F(x) = \int_0^1 3 - 2x$

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

$= [3(1) - (1)^2]$

$= 2.$

∴ It is not a pdf, because it is not equa

ii) $F(x) = 3x^2 \quad ; \quad 0 \leq x \leq 1$

$\rightarrow F(x) = \int_0^1 3x^2$

$= [\cancel{3}x^3/\cancel{3}]_0^1$

$= [x^3]_0^1$

$= [(1)^3]$

$= 1$

$\therefore$ It is pdf because the $\underline{F(x) = 1}$.

i) > dbinom=(x,n,p)

ii) > pbinom=(x,n,p)

iii) > 1- pbinom(x,n,p)

iv) > qbinom(A,n,p)

Q.1.

→ > dbinom(10,100,0.1)

[1] 0.1318653

Q.2

→ i) > dbinom (4,12,1/5)

[1] 0.1328756

ii) > pbinom (4,12,1/5)

[1] 0.927445

iii) > 1- pbinom(5,12,1/5)

[1] 0.01940528

Q.3

→ > dbinom (0:5, 5, 0.1)

[1] 0.59049 0.32805 0.07290 0.00810 0.00045 0.00001.

Q.4.

→ (i) > dbinom (5, 12, 0.25)

[1] 0.1032414

(ii) > pbinom (5,12, 0.25)

[1] 0.9455974

(iii) > 1-pbinom (5,12,0.25)

[1] 0.0227815[1].

(iv) > dbinom (6,12, 0.25)

[1] 0.04014945

PM

Practical :- 0.3

Binomial distribution.

Q.

i) $P(X = x)$

ii) $P(X \leq x)$

iii) $P(X > x)$

(iv) if x is unknow,

$p_1 = P(X \leq x)$, ~~qb~~

Q.1 Find the probability of exactly 10 success in 100 trials with p = 0.1

Q.2 Suppose there are 12 MCQ. each question ~~one~~ has 5 option out of which one is correct. find the probability of having

(i) exactly 4 correct Answer.

(ii) at most 4 correct Answer.

(iii) more than 5 correct Answer.

Q.3 find the complete distribution when n=5 and p

Q.4 n=12, p=0.25, find the following probability distribution

(i) $P(X=5)$

(ii) $P(X \leq 5)$

(iii) $P(X > 5)$

(iv) $P(5 < X < 7)$.

Q.5 Probability of a sales man making a sale to a customer 0.15. Find the probability of
(i) No sales out of 10 customer. i.e x = 0.
(ii) More than 3 sales out of 20 customer.

Q.6 A sales man has a 20% probability of making a sale to a customer. Out of 30 customer what minimum number of sale he can make with 88% probability? (p1, n, p).

Q.7 X follows binomial distribution with n=10, p=0.3. plot the graph of pmf and cdf

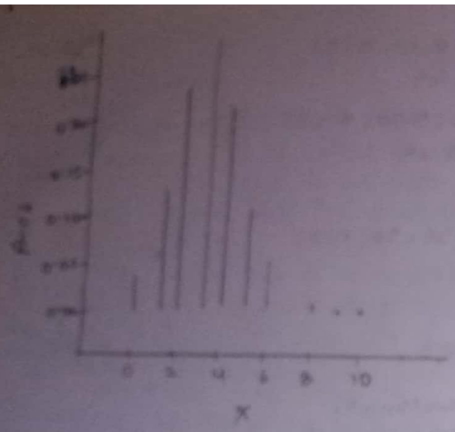
Q.5
→ (i) > dbinom(0, 10, 0.15)
[1] 0.196874
(ii) > 1- pbinom(3, 20, 0.15)
[1] 0.3522746.

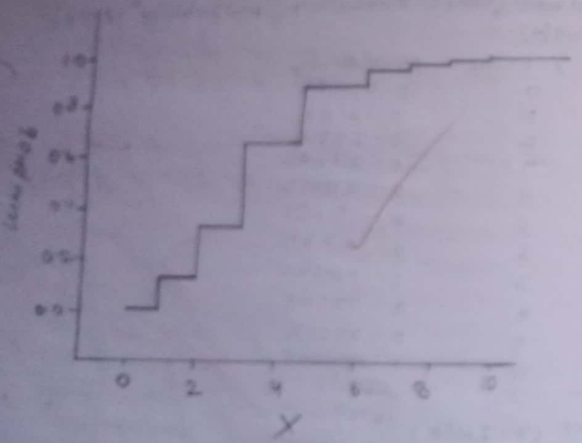
Q.6
→ > qbinom(0.88, 30, 0.2)
[1] 9

Q.7
→ > n=10; p=0.3
> x=0:n
> prob= dbinom(x, n, p)
> cumprob= pbinom(x, n, p)
> d= data.frame("x value"=x, "probability"=prob).
> print(d).

	x.value	probability
1	0	0.02824
2	1	0.12106
3	2	0.23347
4	3	0.26682
5	4	0.20012
6	5	0.10291
7	6	0.03675
8	7	0.009001
9	8	0.00144
10	9	0.00013
11	10	0.000005

> plot(x, prob, "h")
> plot(x, cumprob, "s")





Practical - 04.

Topic: Normal distribution

1) P(X=x) = dnorm (x, μ, σ)
2) P(X≤x) = pnorm (x, μ, σ)
3) P(X>x) = 1 - pnorm (x, μ, σ)
4) To generate random no. from a normal distribution (n random numbers) the R code is
rnorm (n, μ, σ)

Problems:-

Q.1 A random variable X follow normal distribution with mean = μ = 12 and standard deviation = σ = 3. find (i) P(X≤15) (ii) P(10≤X≤13) (iii) P(X>14) (iv) Generate 5 observation (random numbers)

→ (i) μ = 12, σ = 3
p1 = pnorm (15, 12, 3)
[1] 0.8413447
cat("P(x<=15) =", p1)
P(x<=15 = 0.8413447

(ii) P2 = Pnorm (13, 12, 3) - pnorm (10, 12, 3)
[1] 0.3780661
cat("P(10<=x<=13) =", P2)
P(10<=x<=13) = 0.3780661

(iii) P3 = 1 - Pnorm (14, 12, 3)
[1] 0.2524925

cat ("P(x>14)="; P3)
P(x>14) = 0.2524925.

(iv) rnorm (5,12,3)
[1] 9.213365 13.090854 10.035173 6.972853
14.892886.

Q.2. X follows normal distribution with $\mu=10$, $\sigma=12$, find (i) $P(X \leq 7)$ (ii) $P(5<x<12)$ (iii) $P(x>12)$ (iv) Generat 10 observation. (v) find K such that $P(x<K)=0.4$.

→ (i) $\mu=10$, $\sigma=12$.
>P1 = pnorm (7, 10, 12)
[1] 0.0668072

(ii) P2 = pnorm (5,10,2) - pnorm (12,10,2)
[1] 0.0351351

(iii) P3 = 1- pnorm(12,10,2)
[1] 0.1586583.

(iv) rnorm (10, 10, 2)
[1] 7.334054 10.213002 11.319129 10.721121
9.335311 12.591149 8.894187 11.601632
6.324594 12.396941

(v) qnorm (0.4, 10, 2)
[1] 9.493306.

Q.3 Generate five random no from normal distribution with mean=15, S.D=4 find sample mean, median, S.D, print it.

Q.4 X follows / $x \sim N(30,100)$ $\mu=30$, $\sigma=10$ find (i) $P(X \leq 40)$ (ii) $P(X>35)$ (iii) $P(25<X<35)$. (v) find K such that $P(X<K)=0.6$.

Q.4
→ (i) P1 = pnorm (40, 30, 10)
>P1
[1] 0.8413447

(ii) P2 = 1- pnorm (35, 30, 10)
>P2
[1] 0.3085375

(iii) > P3 = pnorm(35,30,10) - pnorm(25,30,10)
> P3
[1] 0.382949

(iv) > ~~p4~~ qnorm (0.6, 30, 10)
[1] 32.53347.

Q.3
→ >x = rnorm (5, 15, 4)
>x
[1] 14.13244 17.49037 18.15842 15.87174 10.78590

>am = mean (x)
>am
[1] 15.28777
>me = median (x)
>me
[1] 15.87174
>variance = (n-1)* var (x)/n
> variance
[1] 7.006958

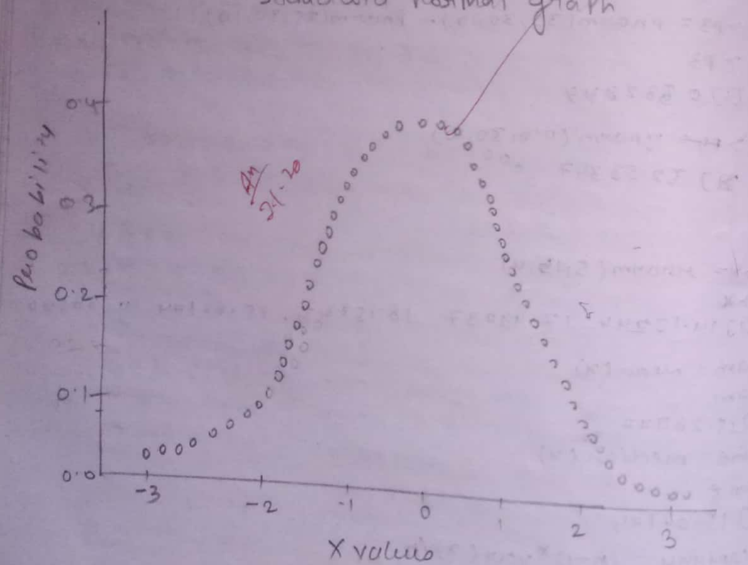
Scanned with CamScanner

```
> sd
[1] 2.647066
> cat ("Sample mean is =", am)
Sample mean is = 15.28777>
> cat (" Sample median is =", me)
Sample median is = 15.87174>
> cat (" Sample sd is =", sd)
Sample sd is = 2.647066>
```

Q.5. Plot the standard normal graph.

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

Standard normal graph



Aim :- Normal and t-test.

Q.1) Test the hypothesis, $\mu = 15$ against $H_1 : \mu \neq 15$
$H_0 : \mu = 15$. random sample of size 400 is drawn and it is calculate the sample mean is 14 and standard deviation is 3. Test the hypothesis at 5% level of significance

-> Formula :- Theory formula | R-Software formula

```
m0 = 15, mx = 14, sd = 3,
n = 400.
zcal = (mx - m0) / (sd / (sqrt(
[1] -6.666667
> cat("Calculate the value of z is =", zcal)
Calculate the value of z is = -6.666667
pvalue = 2 * (1 - pnorm(abs(zcal)))
[1] 2.616796e-11
```

∴ pvalue is less than ~~pvalue~~ 0.05 ~~but~~ we reject $H_0 = 15$

Q.2 Test the $H_0 : \mu = 10$ against $H_1 : \mu_0 \neq 10$ random sample of size 400 is draw standard devia... 2.25 with sample mean 10.2. Test the hypo... at 5% level of significance.

```
-> m0 = 20
mx = 10.2
```

Sd = 2.25
n = 400
zcal = (Mx − Mo) / (sd / (sqrt(n)))
[1] 1.777778
cat ("calculate the value of z is", zcal)
Calculate the value of z is = 1.777778
Pvalue = 2 * (1 − pnorm (abs(zcal)))
Pvalue
[1] 0.07544086
∵ Pvalue more than 0.05
∴ we accept the $H_0 = 10$.

Q.3. Test the hypothesis Ho proportion of smokers in college is 0.2. A sample is collected at It is calculated the sample population at 0.125. Test the hypothesis at 5% level of significant (sample size is 4%)

Proportion

→ P = 0.2
Pp = 0.125
n = 400
Q = 1 − P
Q
[1] 0.8
zcal = (p − P) / (sqrt (P * Q) / n))
zcal
[1] −3.75
Pvalue = 2 * (1 − pnorm (abs (zcal)))
Pvalue
[1] 0.000176834

∴ Pvalue is
∴ we reject the $H_0 = 0.2$.

Q.4. Last year farmers lost 20% of their crops, a random sample of 60 fields are collected and it is found that 9 fields of crops are insect polluted. Test the Hypothesis at 1% level of significance.

→ P = 0.2
p = 9/60
n = 60
Q = 1 − P
Q
[1] 0.8
zcal = (p − P) / (sqrt (P * Q / n))
zcal
[1] −0.9682458
Pvalue = 2 * (1 − pnorm (abs (zcal)))
Pvalue
[1] 0.3329219
∵ Pvalue is ~~less~~ greater than 0.05
∴ we accept the $H_0 = 0.2$.

Q.5. Test the Hypothesis $H_0: \mu = 12.5$ from the following sample at 5% level of significant.

→ x = c(12.25, 11.97, 12.15, 12.08, 12.31, 12.28, 11.94, 11.89, 12.16, 12.04)
n = length (x)
[1] 10
Mx = mean (x)
Mx
12.107
variance = (n−1) * var (x) / n
variance
[1] 0.01521

sd=sqrt(variance)

sd

[1] 0.019521

m0= 12.5

t= (mx-m0)/(sd/sqrt(n))

t

[1] -8.894909

Pvalue= 2*(1-pnorm(abs(t)))

Pvalue

[1] 0

16.1.20

PRACTICAL:-06

AIM:- ~~Large~~ large Sample test.

Q1 Let the population mean [the amount spent by per customer in a restaurant] is 250. A sample of 100 customers selected. The sample mean is calculated is 275 and S.D is 30. Test the hypothesis that the population mean 250 or not at the level of 5% significance.

Q2 In a random sample of one thousand student it is found that 750 use blue pen. Test the hypothesis that the population proportion is 0.8 at 1% level of significance.

Q1.

→ Mx=275, SD=30, n=100

```
>mx= 275
>sd= 30
>n= 100
>m0 = 250
>zcal = (mx-m0)/(sd/(sqrt(n)))
>zcal
[1] 8.333333
cat("Calculate the value of z is =", zcal)
Calculate the value of z is = 8.333333
>pvalue = (1-pnorm(abs(zcal)))*2
> Pvalue
[1] 0
~~cat~~ ("Calculate the value of pvalue is=", pvalue
"Calculate the value of pvalue is 0.
```

Q.2 ~~xp.~~

→ P = 0.8

p = 750/1000

n = 1000

Q = 1-P

```
> zcal = (p-P)/(sqrt(P*Q/n))
> zcal
> -3.952847
cat("calculate zcal =", zcal)
calculate zcal = -3.952847
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 7.72268e-05
cat("calculate pvalue is =", pvalue)
calculate pvalue is = 7.72268e-05
```

Q.3 Two random sample of size 1000 and 2000 are drawn from two population with the same standard deviation 2.5. the sample mean are ~~6.75~~ 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

Q.4 A study of noise level in two hospital given below. Test the claim that of two hospital have same level of noise and 1% of level of significance

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

Q.5: In a sample of 600 student in a college 400 use blue ink. In another college for a sample of 900 student 450 use blue ink. Test the hypothesis that the proportion of student using blue ink in two college are equal or not at 1% level of significance

Q.3 $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

```
→ > n1 = 1000
> n2 = 2000
> mx1 = 67.5
> mx2 = 68
> sd1 = 2.5
> sd2 = 2.5
> zcal = (mx1 - mx2)/sqrt((sd1^2/n1) + (sd2^2/n2))
> zcal
[1] -5.163978
> cat("calculate zcal is =", zcal)
calculate zcal is = -5.163978
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 2.417564e-05
> cat("calculate pvalue is =", pvalue)
calculate pvalue is = 2.417564e-08
```

∴ It's rejected

Q.4

→ $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

$n_1 = 84$

$n_2 = 34$

$mx_1 = 61.2$

$mx_2 = 59.4$

>sd1 = 7.9
>sd2 = 7.5
>zcal = (mx1-mx2) / sqrt((sd1^2/n1) + (sd2^2/n2))
>zcal
[1] 1.1628528
>pvalue = (1-pnorm (abs(zcal)))
>pvalue
[1] 0.2450211
>cat("calculate pvalue =", pvalue)
>calculate pvalue = 0.2450211
∴ It is accept because it is less than 0.05

Q.5

→ H_0: $P_1 = P_2$ against H_0: $P_1 \neq P_2$
>n1 = 600
>n2 = 900
>p1 = 400/600
>p2 = 450/900
>p = (n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.5666667
>q = 1-p
>zcal = (p1-p2) / sqrt(p*q*(1/n1 + 1/n2))
>zcal
> ~~0.01772648~~ 6.381534
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.753222e-10
>cat("calculate pvalue =", pvalue)
[1] 1.753222 e-10
∴ It is rejected it is less than 0.01

6. H_0: P1 = P2 against H_0: P1 ≠ P2
>n1 = 200
>n2 = 200
>p1 = 44/200
>p2 = 30/200
>p = (n1*p1+n2*p2)/(n1+n2)
>p
[1] 1.85
>q = 1-p
>zcal = (p1-p2) / sqrt(p*q(1/n1+1/n2))
>zcal
[1] 1.802741
>pvalue = 2*(1-pnorm (abs(zcal)))
>pvalue
[1] 0.07142888
>cat ("calculate pvalue =", pvalue)
calculate pvalue = 0.07142888.
∴ It is accept because it more than 0.05

23.01.20

22

Practical: 7

Aim :- ~~t~~ Sample Sample Test.

Q.1 The marks of 10 student are given by 63, 63, 66, 67, 68, 69, 70, 70, 71, 72. Test the hypothesis that the Sample comes from population with average marks 66 at 5% level of significance.

→ $H_0: \mu = 66$

\>x = c(. . . . . .)

\>t = test(x)

data: x

t = 68.319, df = 9, p-value = 1.558e-13

alternative hypothesis: true mean is not equal to 0. 5% confidence interval.

65.68171 70.1432.

Sample estimates:

mean of x

67.9.

\> Since value of P is less than 0.05 we reject the $H_0: \mu = 66$.

\> los = 0.05

\> pvalue = 1.558e-13

\> if (pvalue > 0.05) {cat("accept H0")} else {cat("reject H0")}

reject H0.

Q.2 2 groups of student score following marks. Test the hypothesis that there is no significance difference between the 2 groups.

GP1 :- 18, 22, 21, 17, 20, 17, 23, 20, 22, 21

GP2 :- 16, 20, 14, 21, 20, 18, 13, 15, 17, 21. 56

→ Ho there is no difference between two groups.

\> x = c(. . . .)

\> y = c(. . . .)

\> t = test(x, y)

data: x and y.

t = 2.2573, df = 16.376, p value = 0.03798.

Alternate hypothesis: true difference is means is not equal to 0 : ~~true~~ 95 percent confidence interval.

0.1628205 5.0371795

Sample estimates:

mean of x mean of y

20.1 17.5

\> los = 0.05

\> pvalue = 0.03798

\> if (pvalue > 0.05) {cat("accept H0")} else {cat(reject H0")}

reject H0.

Q.3 The Sales of 6 Shop before & After a special campain given below :-

before: 53, 28, 37, 48, 50, 42.

After: 58, 29, 30, 55, 56, 45

Test the hythesis that Campain is effective or not

→ Ho there is no significance difference of Sale before and after campain.

Scanned with CamScanner

> x = c(. . . .)
> y = c(. . . .)
> t.test(x, y, paired = T, alternative = "greater")

paired t-test

data : x and y

t = -2.7815 , df = 5 , p-value = 0.9806

alternative hypothesis : true difference is means is greater than 0.95 percent confidence interval

-6.03545 in f

sample estimate :

mean of differences

-3.5

> los = 0.05

> pvalue = 0.9806

> If (Pvalue > 0.05) {cat ("accept H0")} else {cat ("reject H0")}

reject H0.

Q4. Two medicine are apply to 2 grps of patient respectively.

Grp1 : 10, 12, 13, 11, 14

Grp2 : 8, 9, 12, 14, 15, 10, 9

Is there any significance difference between 2 medicine

→ H0 there is no significance difference between medicine and patient.

> x = c(. . .)

> y = c(. . .)

> t.test(x, y)

> p value = 0.4406

> If (pvalue > 0.05) {cat ("accept H0")} else {cat ("reject H0")}

reject H0

∵ pvalue is less than 0.05 we reject H0.

~~H0 there is no significance difference between weight after and before.~~

Q5 The following are the weight before and after diet program. Is a diet program effective

before : 120, 125, 115, 130, 123, 119

After : 100, 114, 95, 90, 115, 99.

→ H0 there is no significance difference between weight after and before

> x = c(. . . .)

> y = c(. . . .)

> t.test (x, y), paired = T, alternative = "less")

> pvalue = 0.9963

> IF (Pvalue > 0.05) {cat ("accept H0")} else {cat ("reject H0")}

accept H0

∵ pvalue is greater than 0.05 we accept H0

52

Practical: 08

Large and Small test:

Q.1

→ ~~$H_0 : \mu = 5$~~

Q.1 The arithmetic mean of a sample of 100 items from a large population is 52. If the standard 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. Ddoes the information supports that exactly half of the males in the city are smokers ? Test at 1% LOS.

Q.1

→ $H_0 : \mu = 55$

$H_1 : \mu > 55$

```
> n = 100
> MX = 52
> MO = 55
> sd = 7
> zcal = (MX-MO)/(sd / (sqrt(n)))
> zcal
[1] -4.285714
> cat("calculate zcal is = ", zcal)
calculate zcal is = -4.285714.
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 1.82153e-05
> cat("calculate pvalue is = ", pvalue)
calculate pvalue is = 1.82153e-05
```

∵ pvalue is less than 0.05

∴ H_0 is reject at 5% levels of significance.

Q.2

→ $H_0 : P = 0.5$

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

∴ Ho: P is g...
Ho is accepted at 1% level of signif...

Q3.
Ho:
```
->n1= 1000
>n2= 1500
>p1= 0.02
>p2= 0.01
>P= (n1*p1 +n2*p2) / (n1 +n2)
>P
[1] 0.014
>q= 1-P
[1] 0.986
>zcal = (p1 - p2) / (sqrt (P*Q*(1/n1 + 1/n2))
>zcal
[1] 2.02842
>cat ("calculate zcal", zcal)
calculate zcal 2.02842
>pvalue= 2*(1- pnorm(abs(zcal)))
>pvalue
[1] 0.03708364
>cat("calculate pvalue", pvalue)
calculate pvalue 0.03708364
```
∴ Ho is ~~reject~~ less than 0.05
∴ It is accepted at 5% level of significance

Q.4
```
-> Ho : μ: 99
>n= 400
>mx= 99
>Mo= 100
>sd = 8
>zcal = (mx-Mo) / (sd / sqrt(n)))
>zcal
[1] -2.5
```


Q.3 Thousand article from a factory:A are found to have 2% defectives. 1500 articles from a 2nd factory:B are found to have 1% defective. Test at 5% LOS that the two factory are similar are not.

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

Scanned with CamScanner

Q.5. The flower stems are selected and the heights are found to be(cm) 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 2 normal populations and there 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 whether the population have the same variance at 5% LOS.


```
> ~~pvalue~~ cat("calculate zcal", zcal)
calculate zcal = 2.5
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.01241933
> cat("calculate pvalue", pvalue)
calculate pvalue 0.0124193
```

∵ pvalue is less than 0.05

∴ H0 is rejected at 5 level of significance.

Q.5 H0: μ = 66

```
> x = c( . . . . . )
> 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 estimate:
mean of x
 68.14286
> los = 0.01
> if(pvalue > 0.01) {cat("accept (H0)")} else {("rejecte
[1] "Reject H0"
```

Q.6

H0: σ1 = σ2

```
> x = c( . . . . . )
> y = c( . . . . . )
var.test(x, y)
        F test to compare two variances
data: x and y
F = 0.70686, num df = 8, denom df = 10, p-value = 0.63
```

alternate hypothesis: true ratio of variance is not equal to 1.
95 percent confidence interval:
0.1873662   3.0260793
sample estimates:
ratio of variance
0.7060567

Q.6
```
>los = 0.05
>if (pvalue > 0.05) cat({"accept H0"}) else {("reject H0")}
> accept H0
```

Q.7 H0: M = 1200   H1: M ≠ 1200
```
>n = 100
>mx = 1200
>m0 = 1150
>sd = 125
>zcal = (mx-m0)/(sd/sqrt(n))
[1] 4
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 6.334248e-05
```
∴ pvalue less than 0.05
∴ H0 is reject the value at 5% level of significance

Q.8
```
>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
```


7. A company producing light bulbs finds that mean life span of the population of bulbs is 1200 hours with s.d. 125. A sample of 100 bulbs have mean 1150 hours. Test whether the difference between population and sample mean is significantly different?

8. From each of two consignments of apples, a sample of size 200 is drawn and number of bad apples are counted. Test whether the proportion of rotton apples in two assignments are significantly different at 1 % LOS.?

	Sample size	No. of bad apples
Consignment 1	200	44
Consignment 2	300	56

> zcal [1] 0.9128709
~~> pvalue = (f1 - p0) / f1~~

> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.3613101

∴ pvalue is ~~les~~ greater than 0.05
∴ Ho is accepted the value at 5 % level of significance.

PRACTICAL - [illegible]

Aim: Chi Square test & ANOVA
(Analysis of Variance).

Q1. Use the following data to test whether the condition of home & condition of child are independent or not

Cond. Child	Cond. Home Clean	Dirty
Clean	70	50
Fairly clean	80	20
Dirty	35	45

H0: condition of Home & child are 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
```

// They are dependent.
as p-value is less than 0.05 we reject the hypothesis at 5% level of significance

Q2 Test the hypothesis that vaccination & diseases are independent or not

Disease	Vaccine Affected	Not-Affected
Affect	70	46
Not-affected	35	37

H0: Disease & vaccine are independent

```
> x = c(70, 35, 46, 37)
> m = 2
> n = 2
> y = Matrix(x, nrow = m, ncol = n)
> y
       [,1]  [,2]
[1,]    70    46
[2,]    35    37
> pv = chisq.test(y)
> pv
Pearson's chi squared test with Yates
continuity correction
```

data : y
x-squared = 2.0275
df = 2
p-value = 0.1545
∵ p-value is more than 0.05, we accept the hypothesis at 5% level of significance
// They are INDEPENDENT

Q3 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

Ho: The mean's 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(d)
[1] "values" "ind"
> oneway.test(values ~ ind, data = d, var.equal = T)
            One-way analysis of means
data : value and ind
F = 11.735  df = 3, denom df = 9, pvalue = 0.001
```

∵ p-value is less than 0.05 we reject the hypothesis


```
> anova = aov(values ~ ind, data = d)
> summary(anova)
```

	DF	Sum	mean sq	F value
ind	3	71.06	23.687	11.73
Residuals	9	18.17	2.019	

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

Q4 Following data gives a life of tires of 4 brands.

Type	Life
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

Ho: The mean's of A, B, C, D are equal

```
> x1 = c(20, 23, 18, 17, 18, 22, 24)
> x2 = c(B)
> x3 = c(C)
> x4 = c(D)
> d = stack(list(b1 = x1, b2 = x2, b3 = x3, b4 = x4))
> names(d)
[1] "values" "ind"
> oneway.test(value ~ ind, data = d, var.equal = T)
            One-way analysis of means
data : value & ind
F = 6.3445, num df = 3   denom df = 20
p-value = 0.002349.
```

∴ p-value is less than 0.05 we reject the hypothesis

```
>anova = aov(value ~ ind, data = d)
> summary(anova)
            Df    Sum Sq    Mean Sq
ind          3    91.94     30.419
Residuals   20    84.06     4.953
Signif. code : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>x = read.csv("C:/users/admin/desktop/marks.csv")
>x
   Stats   Maths
1   40      60
2   45      48
3   42      47
4   15      20
5   37      25
6   36      27
7   49      57
8   59      58
9   20      25
10  27      27
>am = mean(x$Stats)
>am
[1] 37
>am1 = mean(x$Maths)
```



```
>m1 = median(x$Stats)
>m1
[1] 38.5
>m2 = Median(x$Maths)
[1] 37
>n = length(x$Stats)
>n
[1] 10
>sd = sqrt((n-1)*var(x$Stats)/n)
>sd
[1] 12.64911
>n1 = length(x$Maths)
>n1
[1] 10
>sd1 = sqrt((n-1)*var(x$Maths)/n)
>sd1
[1] 15.2
>cor(x$Stats, x$Maths)
[1] 0.830618
```

27/2

PRACTICAL:10

Topic:- Non-parametric test

Q.1 following are the amount of sulphur oxide emitted by industry in 20 days. Apply sign test to test the 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, 23, 24, 26

Ho: population median is 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

∴ PV is more than 0.05

∴ we accept the hypothesis at 5% level of significance.

[Note: if the Alternative is (H1) median not equal to something H1: me > . or ~~equal to~~ me < ' pv = pbinom(sn, n, 0.5).]

Q.2 following is ... of 10 observation apply sign test, to test the hypothesis that the population median is 625 against the Alternative it is more than 625.

x = 612, 619, 631, 628, 643, 640, 655, 649, 670, 663.

Ho: Median is 625

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

∴ the PV is more than 0.05

∴ the hypothesis is accept at 5% level of significance.

Q.3 Following are value of the ~~factor~~ sample test the hypothesis than the population median is 60 against the Alternative is more than 60 at 5% los using Wilcoxon Signed Rank Test.

x = 63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 69, 48, 66, 72, 63, 87, 69.

Ho: population median = 60.

H1: ... >60

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

wilcoxon signed rank test with continuity correction.

data: x

V = 145, p-value = 0.02298

alternative hypothesis: true location is greater than 60.

Note: If the alternative is "less" and if alternative is not equal to than we prefer "two sided".
alter = "Less". or alter = "two sided".

Q.4. Using WSR test the population median is 12 or less than 12.

x = 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26

H0: population median is 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

alternatives hypothesis: true location is less than 12.

Q.5. The weights of students before and after they stop smoking are given below using WSR test that their is no significance change

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

H0: before and after their is no change

H1: their is change.

> x = c(65, 77, 75, 62, 72)

> y = c(72, 74, 72, 66, 73)

> d = x - y

[1] -7 1 3 -4 -1

wilcoxon signed rank test with continuity correction.

data: d

V = 4.5, p-value = 0.4982

Alternative hypothesis: true location is not equal to 0.

An
27.2.20