PRACTICAL - 1

Basic of R Software

i. R Software for data analysis & statistical Computing

ii. It is a Software by which effective data handling & outcome Storage is possible.

iii. It is Capable of graphical display

iv. It is a free Software

Q.1.

$2^2 + |-5| + 4 \times 5 + 6/5$

```
> 2^2 + abs(-5) + 4* 5+ 6/5
[1] 30.2
```

Q.2.

```
x = 20
y = 2x
z = x+y
√z
```

```
x = 20
y = 2* x
z = x+y
z**0.5
[1] 7.745
```

```
round (2.56)
[1] 3
```

```
round (2.4)
[1] 2
```

Scanned with CamScanner

Q.3.

x = 10
y = 15
z = 5

a) x+y+z
b) xyz
c) sqrt(xyz)
d round of sqrt(xyz)

a x+y+z
[1] 30

b x*y*z
[1] 750

c sqrt(x*y*z)
[1] 27.38613

d round ((x*y*z)**0.5)
[1] 27

Q.4.

A Vector in R Software is denoted by 'c'

c(2, 3, 5, 7)^2
> c(2, 3, 5, 7)^2
> 4 9 25 49

Q.5. c(2,3,5,7)^c(2,3)
> c(2,3,5,7)^c(2,3)
[1] 4 27 25 343

Q.6. c(2,3,5,7,9,11)^c(2,3)
> c(2,3,5,7,9,11)^c(2,3)
[1] 4 27 25 343 81 1331

Q.7.
c(1,2,3,4,5,6)
c(2,3,4)
> c(1,2,3,4,5,6)^c(2,3,4)
> 1 8 81 16 125 1296

Q.8
c(2,4,6,8)
> x=c(2,4,6,8)
x* 3
[1] 6 12 18 24

Q.9.
x=c(2,4,6,8)
y=c(-2,-3,-5,-7)
x*y
[1] -4 -12 -30 -56

Q.10
x=c(2,4,6,8)
x+10
[1] 12 14 16 18

Scanned with CamScanner

Q. 11 (2, 4, 6, 8) + c(-2, -3, -1, 0)

= [1] 0 1 5 8

→ Sum, Product

Q. Find the sum, product, Sqrt of the sum & the product for the following values

= (4, 9, 2, 5, 7, 8, 3, 6, 15, 12, 10, 9, 8, 13, 14)

```
x = c( 4, 9 . . . . . )
y = sum(x)
z = prod(z)
y ^ 0.5
y
[1] 125
z
[1] 5
> y ^ 0.5
[1] 11.18034
> z ^ 0.5
[1] 2925632
```

Q. 13 Find the sum, product, max, min, values of c(2, 8, 9, 11, 10, 7, 6) ^ 2

```
x = c(2, 8 . . . . . ) ^ 2
[1] 4 64 81 121 100 49 36
> sum(x)
[1] 455
> prod(x)
[1] 44 254 747 3400
```



```
> max(x)
[1] 121
> min(x)
[1] 4
```

Q. Matrix

```
x <- matrix(nrow = 3, ncol = 4, c(2, 6, 7, 8, 9, 4, 5, 10, 2, 1, 4, 5))
x
```

	[,1]	[,2]	[,3]	[,4]
[1,]	2	8	5	1
[2,]	6	9	0	4
[3,]	7	4	2	5

Q. Addition of Matrix

```
x = matrix(nrow = 3, ncol = 3, c(4, 5, 6, 7, 8, 9, 4, 0, 2))
y = matrix(nrow = 3, ncol = 3, c(8, 4, 5, 11, 12, 9, 7, 4))
x
```

	[,1]	[,2]	[,3]
[1,]	4	7	4
[2,]	5	8	0
[3,]	6	9	2

Scanned with CamScanner

```
y
     [,1] [,2] [,3]
[1,]    6   11    9
[2,]    4   12    7
[3,]    5    8    4

x * 2
     [,1] [,2] [,3]
[1,]    8   14    8
[2,]   10   16    0
[3,]   12   19    4

y * 2
     [,1] [,2] [,3]
[1,]   12   22   18
[2,]    8   24   14
[3,]   10   16    8

x+y
     [,1] [,2] [,3]
[1,]   10   18   13
[2,]    9   20    7
[3,]   11   17    6

x * y
     [,1] [,2] [,3]
[2,]   24   77   36
[2,]   20   96    0
[3,]   30   72    8
```

Ans 25:11

Binomial Distribution

Practical - 2

n = Total no. of trails
p = P(Success)
q = P(failure)
X = No. of Successes outcome of n

eg:

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

$E(x) = np$

$V(x) = npq$

$n = 10, \; p = 0.6, \; q = 0.4$

$P(x \le x) = P(x \le 3)$

```
> dbinom(x,n,p)
> n * p
> n * p * q
> pbinom(x,n,p)
```

P(x)
E(x)
V(x)
P(X ≤ x)

Q1. Toss a coin 10 times with probability (Head) = 0.6.
Let X be the no of heads
i) Find the Probability of
i) Seven heads
ii) Four heads
iii) atmost 4 heads
iv) Atleast 6 heads
v) No heads
vi) All heads.
also find Expectation & Variance

Ans

```
>n=10, p=0.6, q=0.4, x=7
>dbinom(7,10,0.6)
[1] 0.2149908
>dbinom(4,10,0.6)
[1] 0.1114767
>pbinom(4,10,0.6)
[1] 0.6177194
```

Scanned with CamScanner

> 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
> n * p
[1] 6
> n * p * q
[1] 2.4

Q.2 Suppose there are 12 MCQ in an english question paper.
→ Each Question has 5 answers & only 1 of them is correct find the probability of having.
i) 4 Correct answers
ii) Atmost 4 Correct answers
iii) Atleast 3 Correct answers.

Q.3 Find the Complete binomial distribution when n=5, p=0.1, fix.

Q.4. Find the probability of exactly 10 success out of hundred trails with p=0.1

Q.5. X follows binomial distribution with n = 12 & p = 0.25. Find
i) P(X ≤ 5) ii P(X > 7) iii P(5 < X < 7)

Q.6. There are 10 members in a Committee probability of any number attending a meeting is 0.9. What is the probability, there 7 or more members will be present in a meeting.

Q.7 A Salesman has a 20% probability of making a ~~sale~~ sell to a customer. On a typical day he will meet 30 customers. What minimum no. of sells he will make with 88% probability

Q.8. For n=10 & p=0.6 find the binomial probabilities & plot the graphs of pmf & c.d.f.

Note:-
i) P(X = x) = dbinom(x, n, p)
ii) P(X ≤ x) = probability of atmost x values
= pbinom(x, n, p)
iii) P(X > x) = 1 - pbinom(x, n, p)
iv) If x is known & the probability is given as P_1, to find x = qbinom(P_1, n, p)

A.2 n = 12, p = 1/5, x = 4
i dbinom(x, n, p)
[1] 0.1328756
ii pbinom(x, n, p)
[1] 0.9274445
iii x = 2
1 - pbinom(x, n, p)
[1] 0.4416543

A.3 n = 5, p = 0.1, x = 1
dbinom(x, n, p) i.e dbinom(1, n, p)
[1] 0.32805
dbinom(2, n, p) i.e
[1] 0.0729

Scanned with CamScanner

dbinom(3,n,p)
[1] 0.008
dbinom(4,n,p)
[1]0.00045
dbinom(5,n,p)
[1] 1e-05
dbinom(0,n,p)
[1] 0.59049

A.4. n = 100, p = 0.1, x = 10
dbinom (x,n,p)
[1] 0.1318653

A.5 n=12, p= 0.25 , x=5
pbinom(x,n,p)
[1] 0.9455978
x = 7
1- pbinom(x,n, p)
[1] 0.0027815
x= 6
dbinom (x,n,p)
[1] 0.0401945

A.6 n=10, p=0.9, x=6
1- pbinom (x,n,p)
[1] 0.9872045

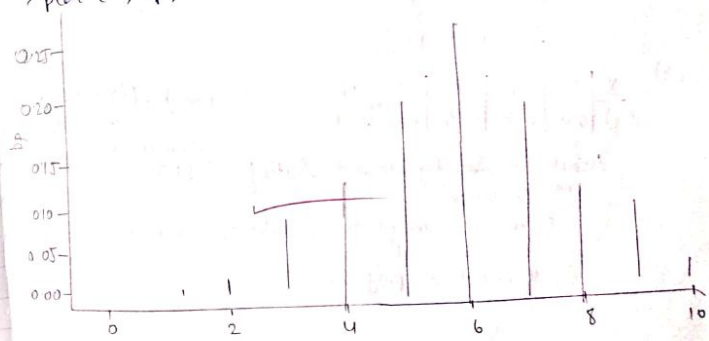
A.7 n= 30, p= 0.2, P1= 0.88
qbinom(p1,n,p)
[1] 9

A.8. n=10
p=0.6
x=0:n
bp=dbinom(x,n,p)
[1]0.0001048576 0.0015728640 0.0106168320 0.0424673280
0.1114767360 0.2006581248 0.2508226560 0.2149908480
0.1209323520 [10] 0.0403107840 0.0060466176

d=data.frame("x-values"=x, "probability"=bp)

	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=pbino(x,n,p) plot.(x, p,'s')

Practical- 5 (Probability distribution)

Q1. Check the following are P.M.f or Not

(1)

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

~~Condit~~ Since Conditions :

i $0 \le p(x) \le 1$

ii $\sum_x p(x_i) = 1$

Since, it doesn't satisfy the first condition

∴ it is not P.M.f

(2)

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

prob = c(0.3, 0.2, 0.3, 0.1, 0.1)

sum(prob)

[1] 1

Hence, it satisfy both the condition

∴ it is pmf.

(3)

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

probab = c(0.4, 0.2, 0.3, 0.2, 0.1)

sum(Probab)

Hence, it ~~satisfy~~ doesn't satisfy [1] 1.2

2nd condition

i.e ~~as~~ & sum of the probability is more than 1

∴ it is not a pmf

Q2. Following is a pmf of (X)

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

Find mean & Variance

X_i	$P(x)$	~~(x)f~~ $x_i P(x)$	$x_i^2 (P(x))$
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

Mean $= E(X) = \sum x P(x) = 3.45$

Var $= V(X) = \sum x^2 (P(x)) - [E(X)]^2$

$= 13.55 - (3.45)^2$

$= 1.6475$

In R software

x = c(1, 2, 3, 4, 5)

prob = c(0.1, 0.15, 0.2, 0.3, 0.25)

a = x * prob

a

[1] 0.10 0.30 0.60 1.20 1.25

Mean = sum(a)

mean

[1] 3.45

Scanned with CamScanner

i?

```
b = (x^2)* prob
var = sum(b) - mean^2
Var
[1] 1.6475
```

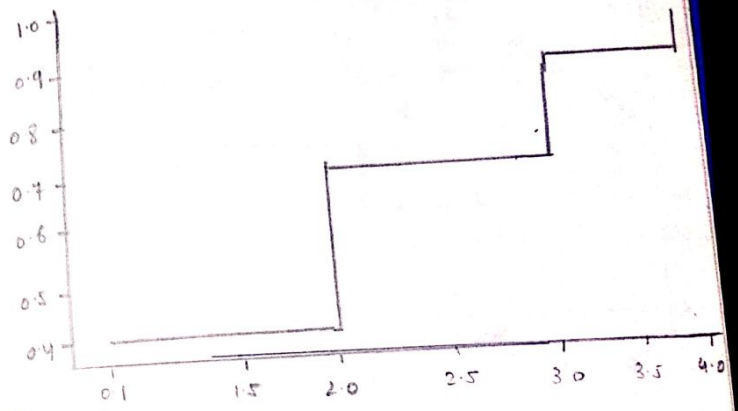
0.3. Find mean & Variance

```
x = c(5, 10, 15, 20, 25)
probab = c(0.1, 0.3, 0.2, 0.25, 0.15)
a = x* probab
a
[1] 0.50   3.00   3.00   5.00   3.75

b = (x^2)* probab
var = sum(b) - mean^2
Var
[1] 38.6875

Mean = sum(a)
Mean
[1] 38.6875
```

0.4. Find C.D.F of the following p.m.f & Draw the graph of C.D.F

(i)

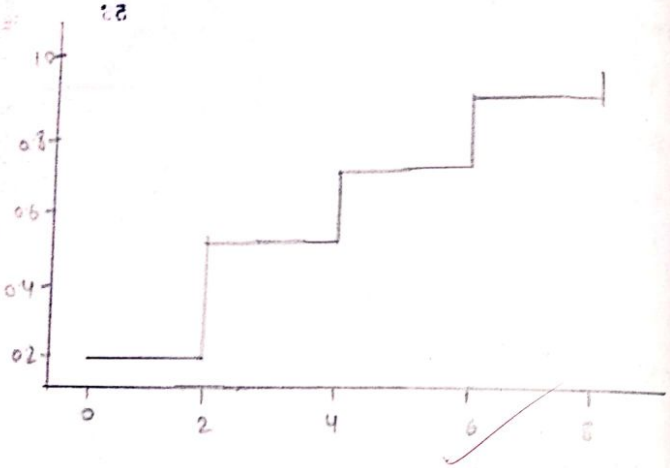
```
x = seq(1, 4, 1)
prob = c(0.4, 0.3, 0.2, 0.1)
a = cumsum(prob)
a
[1] 0.4, 0.7, 0.9, 1.0
> plot(x, a, "s")
```



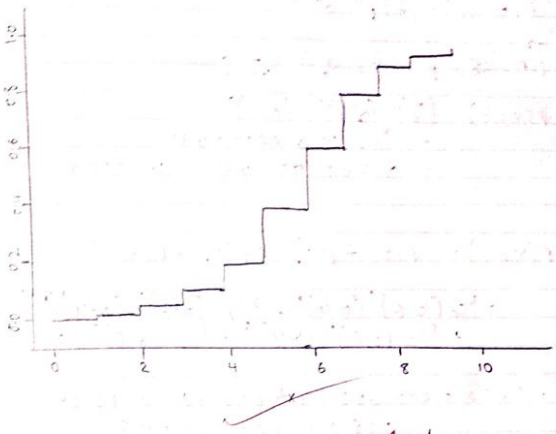
0.4. ii

```
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")
```

Scanned with CamScanner



Asw
9.12.19



Asw
16.12.19

Practical - 4 ~~(Normal distribution)~~

$f(x=x) = \binom{n}{x} p^x q^{n-x}$

$[n = 8,\ p = 0.6,\ q = 0.4]$ given

① $P(x=7) = \binom{8}{7} (0.6)^7 (0.4)^{8-7}$

$= {}^8C_7 \times 0.2799 \times 0.4$

$= 8 \times 0.2799 \times 0.4 = 0.08957$

② $P(X \le 3) = P(0) + P(1) + P(2) + P(3)$

$= {}^8C_0 (0.6)^0 (0.4)^8 + {}^8C_1 (0.6)^1 (0.4)^{7-1}$
$+ {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^2 (0.4)^5$

$= 1 \times 1 \times 0.000655536 + 8 \times 0.6 \times 0.0016384$
$+ 28 \times 0.36 \times 0.004096$
$+ 56 \times 0.26 \times 0.01024$
$= 0.1736704.$

③ $P(x=2 \text{ or } 3) = P(2) + P(3)$

$= {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^3 (0.4)^5$

$= 28 \times 0.36 \times 0.004096 + 56 \times 0.216 \times 0.01024$

$= 0.04128768 + 0.12386304$

$= 0.16515012$

Practical - 5
(Normal distribution)

→ It is an example of Continuous probability distribution

$X \sim (\mu, \sigma^2)$

① $P = (X = x) = \text{dnorm}(x, \mu, \sigma)$

② $P(X \le x) = \text{pnorm}(x, \mu, \sigma)$

③ $P(X > x) = 1 - \text{pnorm}(x, \mu, \sigma)$

④ To find the Value of K

$P(X \le K) = P$

The Command is qnorm(p, μ, σ)

⑤ To generate a random sample of size n

rnorm(n, μ, σ)

Q.1) A random Variable X follows normal distribution with $\mu = 10$, $\sigma = 2$. To find → $P(x \le 7)$

$P(x > 12)$

$P(5 \le x \le 12)$

$P(X < K) = 0.4$

Q.2. $X \sim N(100, 36)$

$\sigma = \text{sqrt}(36)$

Q.3. $P(x \le 110)$

$P(X > 105)$

$P(x \le 92)$

$P(95 \le x \le 110)$

$P(x < n) = 0.9$

Scanned with CamScanner

Q.3 Generate 10 random sample find the sample mean, median & variance & S.D

$\mu = 10, \sigma = 3$

$n = 10$

A.1 p = pnorm (7, 10, 2)

[1] 0.0668072

> cat (" p (x<=7) is = ", p)

q = 1-pnorm (12, 10, 2)

[1] 0.1586553

cat (" p(x>12) is = ", q)

p = pnorm (12, 10, 2) - pnorm (5, 10, 2)

[1] 0.8351351

cat (" p (5 ≤ x ≤ 12) is = ", p)

k = qnorm (0.4, 10, 2)

[1] 9.493306

cat (" p (x<k) is = ", k)

A.2 d = Sqrt (36)

[1] 6

$\mu = 100, \sigma = 6$

p = pnorm(110, 100, 6)

[1] 0.9522096

Scanned with CamScanner

```
q = 1- pnorm(105,100, 6)
q
[1] 0.2023284

p = pnorm(92, 100, 6)
p
[1] 0.9121122

p = pnorm (110, 100, 6) - pnorm (95,100, 6)
p
[1] 0.7495813

k = qnorm (0.9, 100, 6)
k
[1] 107.6893
```

A3 $\mu = 10$, $\sigma = 3$ $n = 10$

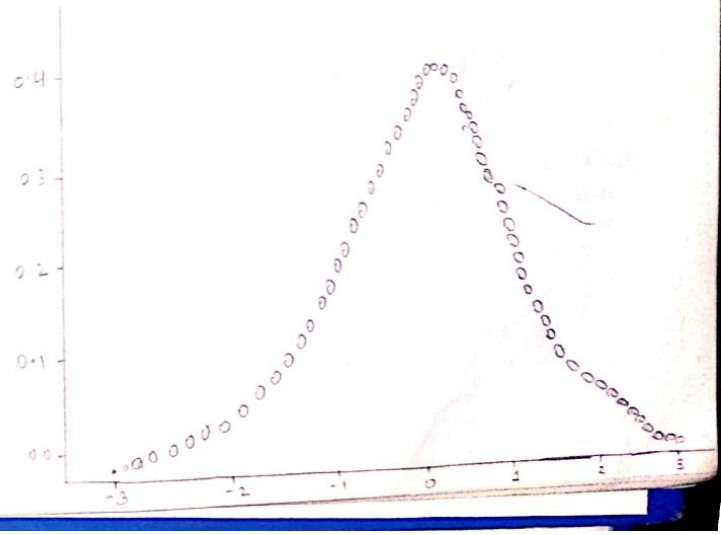
```
x = rnorm(10, 10, 3)
x
[1] 9.975788   9.101868   10.794650   12.497118
    4.763568   7.281763   14.266924   12.234024
    11.959975  14.280874

am = mean (x)
am
[1] 10.71266
```



```
md = median (x)
md
[1] 11.37733

n = 10
Variance = ((n-1) * Var (x))/n
Variance
[1] 8.251349

sd = sqrt (variance)
sd
[1] 2.872516
```

Q.4]

A: `x = seq (-3, 3, by = 0.1)`
`y = dnorm (x)`
`plot (x, y, xlab = " x values", ylab = "probability", main = " Standard normal curve")`



Scanned with CamScanner

Q5 (a) $X \sim N(50, 100)$ $\sigma = 10$

Find i) $P(X \le 60)$

ii) $P(X > 65)$

iii) $P(45 \le x \le 60)$

```
p = pnorm(60, 50, 10)
[1] 0.843447

p = pnorm(60, 50, 10) - pnorm(45, 50, 10)
p
[1] 0.5328072

p = 1 - pnorm(65, 50, 10)
p
[1] 0.0668072
```

Ans 61.20

Practical-6

(z & t distribution Sums)

Q.1. Test the hypothesis $H_0: \mu = 20$ against $H_1: \mu \neq 20$ a sample of size 400 will selected & the sample mean is 20.2 & the standard deviation 2.25.

Test at 5% level of significance

```
> mo = 20; mx = 20.2; sd = 2.25; n = 400
> zcal = (mx - mo) / (sd/sqrt(n))
> zcal
[1] 1.777778
> cat("z calculated is zcal : ", zcal)
z calculated is zcal : 1.777778
> pvalue = 2*(1 - pnorm(abs(zcal)))
> pvalue
[1] 0.07544036
```

Since, the pvalue is more than 0.05, we accept $H_0: \mu = 20$

Q.2. We want to test the hypothesis $H_0: \mu = 250$ against $H_1: \mu \neq 250$. a sample of size 100 has a mean of 275. & s.d = 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
```

Scanned with CamScanner

```
> cat ("z calculated is zcal : ", zcal)
  z calculated is zcal: 8.33333
> pvalue = 2*(1 - pnorm (abs (zcal)))
> p value
[1] 0
```

Since, the pvalue is less than 0.05, we reject the ~~Sample mean~~ $H_0: \mu = 250$

Q.3. We want to test the hypothesis $H_0: P = 0.2$ against $H_1: P \neq 0.2$ (P = population proportion) a sample of 400 ~~be~~ is selected & 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] 0.0001748346
```

Since, the pvalue is less than 0.01, we reject $H_0: P = 0.2$

Q.4. In a big city 325 men out of 600 men, were found to be employed. Does this information support the conclusion that exactly half of the men in the city are self employed?

```
> P= 0.5 ; p=325/600; n=600; Q=1-P
> zcal =(p-P)/(Sqrt (P*Q/n))
> zcal
[1] 2.04124
> pvalue = 2*(1-pnorm (abs(zcal)))
> pvalue
[1] 0.04122
```

Since, the pvalue is less than 0.05, we reject $H_0: P = 0.5$

Q.5 Test the hypothesis $H_0: \mu = 50$ against $H_1: \mu \neq 50$ 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

```
> x = c (Data)
> n = length(x)
> am = mean(x)
> v = var(x)
> Variance =((n-1)*v)/n
> Variance
[1] 30.95556
> sd = Sqrt(Variance)
> sd
[1] 5.563772
> mo= 50; mx=am; sd= 5.563772 ; n=30
> zcal =(mx- mo) / (sd /Sqrt (n))
> zcal
[1] -0.6562964
> pvalue = 2*(1-pnorm (abs (zcal)))
> pvalue
[1] 0.5116334
```

Dr 20/01/20

Since, the pvalue is more than 0 we accept $H_0: \mu = 50$

Practical 7.

Large Sample Test

Q.1.) Two random sample of size 1000 & 2000 are drawn from two population with the S.D (σ) = 2 & 3 respectively. Test the hypothesis that the two population means are equal or not at 5% level of significance. Sample means are = 67 & 68 respectively

Sol^n $H_0: \mu_1 = \mu_2$

against $H_1: \mu_1 \neq \mu_2$

```
>n1=1000
>n2=2000
>mx1= 67
>mx2=68
>sd1=2
>sd2=3
>zcal=(mx1-mx2)/Sqrt((sd^1 /n1)+(sd^2/n2))
>zcal
[1] -10.84652
>pvalue=(2*(1-pnorm(abs(zcal))))
>pvalue
[1] 0
```

Since, the Pvalue is less than ~~0.005~~ 0.05, we reject $H_0: \mu_1 = \mu_2$

Q.2. A study of noise level in two hospital is given. Following data is calculated. Sample size 84, Sample mean = 61.2, S.D1 = 7.9, 2nd Sample size = 34, Sample mean = 59.4, S.D2 = 7.8. Test $H_0: \mu_1 = \mu_2$ 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((sd^2/n1)+(sd2^2/n2))
>zcal
[1] 1.13117
>pvalue =2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.258006
```

Since, pvalue is greater than 0.01, we accept $H_0: \mu_1 = \mu_2$

3. From each of two population of oranges the following samples are collected test, whether the proportion of bad oranges are equal or not, 1st Sample size = 250, 2nd Sample size = 200 No. of bad oranges in the 1st Sample = 44 & 2nd Sample 30

$H_0: P_1 = P_2$ against $H_1: P_1 \neq P_2$

```
>n1=250
>n2=200
>p1=44/250
>p2=30/200
>p=(n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.1644444
>q=1-p
>q
[1] 0.8355556
>zcal=(p1-p2)/sqrt(p*q*(1/n1+1/n2))
>zcal
[1] 0.7233581
```

Scanned with CamScanner

```
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.4596896
```

Since, pvalue is greater than 0.05, we accept the $H_0: P_1 = P_2$

4. Random Sample of 400 males & 600 females, when asked whether they want a ATM near by. 200 males & 390 females were in favour of proposals. Test the hypothesis that the proportion of males & females favouring the proposals are equal or not at 5% LOS

Sol^n

```
>n1 = 400
>n2 = 600
>p1 = 200/400
>p2 = 390/600
>p = (n1*p1 + n2*p2)/(n1+n2)
>p
[1] 0.09
>q 1-p
>zcal = (p1-p2) / sqrt(p*q*(1/n1+1/n2))
>zcal
[1] -4.724731
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.303972e-06
```


Q5. Following are the two independent samples from the two population test equality of two population mean at 5% level of significance

x1 = 74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 76

~~x2 = 77~~

x2 = 72, 76, 74, 70, 70, 78, 70, 72, 75, 79, 74, 74, 75, 78, 72, 74, 80

Sol^n

```
> x1 = c(Data)
> x2 = c(Data)
> n1 = length(x1)
> n2 = length(x2)
> mx1 mean(x1)
> variance = (n1-1) * var(x1)/n1
[1] 0.4508014
> sd1 sqrt(variance)
[1] 0.6714177
> Variance = (n2-1)* var(x2/n2)
[1] 0.4761632
> sd2 = sqrt(variance)
[1] 0.7285031
> tcal = (mx1 - mx2)/sqrt((sd1^2/n1)+(sd2^2/n2))
> tcal
[1] 1.95227
> t.test(x1,x2)
pvalue : 0.1387
```

Nikhil
27.d.20

PRACTICAL-8

(Small Sample test) t-test

Q.1. The Random Sample of 15 observation are given by 80, 100, 110, 105, 122, 70, 120, 110, 101, 88, 83, 95, 89, 107, 125. Do this data Support the assumption that the population mean is 100.

Soln

```
>x=c(Data)
>t.test(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

Hence, p-value is less than 0.005 ∴ We feel rejected the p-value $H_0: \mu = 100$.

Q.2. Write two groups of 10 students Scored the following marks

gp 1 = 18, 22, 21, 17, 20, 17, 23, 20, 22, 21

2 = 16, 20, 14, 21, 20, 18, 23, 15, 19, 21

test the hypothesis that there is no significance difference betn the marks are 1% level of Significance. $H_0: \mu_0 = 100$

Soln

```
>x=c(Data)
>y=c(Data)
>t.test(x,y)
```

welch Two Sample t-test

data: x & y

t = 2.2573 df = 16.376 p-value = 0.03798

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, p-value is greater that 0.01, ∴ We accept $H_0: \mu = 100$.

Q.3. Two types of medicine are used to 5 & 7 patients, for reducing weight. The decrease in the weight after medicine are given below.

Medicine A : 10, 12, 13, 11, 14

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

Is there a Significance different efficiency of the medicine.

$H_0: \mu_1 = \mu_2$.

Soln

```
>x=c(Data)
>y=c(Data)
>t.test(x,y)
```

welch Two Sample t-test

data: x & y

t = 0.80384 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, p-value is greater that 0.01, we accept $H_0: \mu_0 = \mu_1 = 0.01$

Scanned with CamScanner

i.a

(Paired t-test)

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

Before = 120, 125, 115, 130, 133, 119, 122, 127, 128, 118

After- 111, 114, 107, 120, 115, 117, 112, 120, 119, 112

Ho: There is no Significant different in weight against H1: The diet program reduce weight.

Sol^n

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

Paired t-test

data: x and y

t = 8.9019 df = 9 p-value = 1

alternative hypothesis: true difference in means is less than 0

95 percent Confidence interval:

-Inf 11.45627

Sample estimates:

mean of the difference

9.5

Since, p-value is more than 0.05, we accept Ho: the level of Significance.

Q.5. Sample A: 66, 67, 75, 76, 82, 84, 88, 90, 92 66

B: 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97

Test the population mean are equal or not.

Sol^n

```
> x = c (Data)
> y= c (Data)
> 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 mean of y

80 83

Since, p-value is greater than 0.05, we accept level of Significance.

Q.6. The follow are the marks before & after of training program

Test: the program is effective or not

Before: 71, 74, 74, 69, 70, 74, 76, 70, 73, 75

After: 74, 77, 74, 73, 79, 76, 82, 72, 75, 78

Sol^n

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

Paired t-test

data: x and y

t = -4.4691 df = 9 p-value = 0.9992

Scanned with CamScanner

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

∴ Here p-value is more than 0.05

∴ we accept Ho: No's 95 is effective.

Ans 10.2.20

Q.1 The arithmatic 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 does the information supports that exactly half of the males in the city are smokers? Test it at 1% LOS.

Q.3 Thousand article from a factory: A are found to have 2% defectives, 1500 articles from a 2nd factory: B are ~~are~~ found to have 1% defective test at 5% LOS that the two factory are similiar or 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 & variance is 64.

Q.5 The flower stems are selected & the hights are found to be (cm) 63, 63, 68, 69, 71, 71, 72 test the hypothesis that the mean hight is 66 or not at 1% LOS.

Scanned with CamScanner

Q8 Two random samples were drawn from 2 normal populations & these 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.

A11
Soln

```
Ho: μ0 = 0.5
>n=100
>mx = 52
>sd = 30
>mo = 55
>zcal=(mx-mo)/(sd/sqrt(n))
>zcal
[1] -4.285714
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.00001
```

∴ p-value is less ~~that~~ than 0.05, we will reject Ho: μ0 = 0.05

A.2
Soln

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

∴ p-value is greater than 0.01, we accept Ho: P0 = 0.05

A.3
Soln

```
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
>zcal=(p-P)/(sqrt(P*Q*(1/n1+1/n2)))
>zcal
[1] 2.0848
>pvalue =2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.0370
```

∴ P-value is less than 0.05, we reject Ho: P1=P2

Scanned with CamScanner

A.4 : 7

Ans $H_0: \mu_0 = 0.5$

```
> n = 400
> v = 64
> mx = 49
> m0 = 100
> sd = sqrt(v0)
> sd
[1] 8
> zcal = (mx - m0)/(sd/sqrt(n))
> zcal
[1] -2.5
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.0124
```

∴ p-value is less than 0.05 we reject $H_0: \mu_0 = 0.5$

A.5

Ans $H_0: \mu_1 = \mu_2$

```
> x = c(Data)
> t.test(x)
```

One Sample t-test

data: x

$t_0 = 4.94$ df = 6, p-value = 5.522e-09

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

54.66459 71.6354

Sample estimates:

mean of x

63.15

Since, p-value is less that 0.01 ∴ we reject $H_0: \mu_0 = 68$

A.6

Ans H_0 = population is same

```
> x = c(Data)
> y = c(Data)
> j = var.test(x,y)
> j
```

F Test to Compare two Variance

data: x and y

F = 0.70686, num df = 6, 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

Since, p-value is greater than 0.05, ∴ we accept $H_0: \mu_0 =$ The population have same Variance

1.7

PRACTICAL -10

ANOBA & Chi - Square
Analysis of Variance

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

		Clean	Dirty
		Cleanliness of Home	
Cleanliness of child	Clean	~~80~~ 70	50
	Fairly clean	80	20
	Dirty	35	45

Ho: CC & CH are independent

```
> x=c(70,80,35,50,20,45)
> m=3
> n=2
> y<- matrix(x, nrow=m, ncol=n)
     [,1] [,2]
[1,]   70   50
[2,]   80   20
[3,]   35   45
> pv= chisq.test(y)
> pv
```



```
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 ~~depen~~ independent.

Q.2 Use the following data to find a ~~vaccination~~ vaccination & a particular disease are independent or not.

		Disease Affted	Not Affted
Vaccination	Given	20	30
	Not Given	25	35

Ho: Vaccination & disease are 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 with Yates' continuity correction

```
data:  y
X-squared = 0, df = 1, p-value = 1
```

Scanned with CamScanner

Since, p-value is more than 0.05, we accept
Ho: Vaccination & disease 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

Ho: The means of the varieties 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 & 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          3

Residual standard error: 1.420746
Estimated effects may be unbalanced
```

∴ Since, p-value is less than 0.05. we reject
Ho: The means of the varieties are equal.

Q.4. Following data gives lifes of ~~tetr~~ type of four brands

Type	Observations
A	(20, 23, 18, 17, 18, 22, 2
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 for the four ~~B~~ brands are same

Ho: The average life of four type of brands a
Same

Scanned with CamScanner

15

```
> 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))
> d
> names(d)
[1] "values"  "ind"

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

        One-way analysis of means
data:  values and ind
F = 6.8445, num df = 3, denom df = 20, p-value = 0.0023

> anova=aov(values ~ ind, data=d)
> 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 the H0: The average life of four type of brand 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 there any association between IQ & the economic condition of the home are

		IQ High	IQ low
Economic Condition	High	460	140
	Medium	330	200
	low	240	160

H0: EC & IQ 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-0
```

∴ Since, p-value is less than 0.005, we reject H0: EC & IQ are independent.

AM (7,2,22)

Scanned with CamScanner

Pr-11 Non-parametric Test

Q.1 Following are the amounts 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: population median is 21.5

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

Since, p-value is More than 0.05 ∴ we accept Ho: population median is 21.5

Q2. 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 ~~675~~ 625 against the alternative it is greater than ~~Ho~~ 625 at 1% LOS.

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

Ho. Population median is 625

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

∴ Since, p-value is ~~less~~ more than 0.01, ∴ we accept Ho: ~~pop~~ population median is 625.

Scanned with CamScanner

3 10 observations are
36, 32, 21, 30, 24, 25, 20, 22, 20, 18
Using Sign test. Test the hypothesis is 25 against the alternative it is less than 25 at 5% LOS.

Soln

H0: population median is 25
```
>x= c (Data)
> ~~Sp=leng~~ med= 25
> Sp= length (x[x> med])
> Sn= length(x[x< med])
> n= Sp +Sn
> n
[1] 9
> pv = pbinom (Sp, n, 0.5)
> pv
[1] 0.2539063
```

Since, ~~the~~ p-value is ~~accept~~ ~~the~~ more than 0.05 we accept H0: population median is 25

6.4. The following are some measurements - 71
63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 65
Using wilcoxon Signed Rank test, Test the hypothesis that the population median is 60 against the alternative it is greater than 60 at 5% LOS.

Soln
```
> H0: population median is 60
> x= c (Data)
> 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 greater than 0.05, we accept H0: Population median is 60.

6.5.
15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26
Wilcoxon signed rank test ~~with Continuity Corr~~
Test the hypothesis that the population median is ~~less~~ 20 ~~than~~ against the alternative it is less than 20 at 5% LOS

Soln
```
> H0: population median is 20
> x= c(Data)
> wilcox.test(x, alt="less", mu=20)
    Wilcoxon signed rank test with continuity correction
data: x
V= 48.5 p-value= 0.932
alternative hypothesis: true location is less than 20
```
∴ Since, p-value is greater than 0.05, we accept H0: population median is 20.

Scanned with CamScanner

Q6. Data = (20, 25, 27, 30, 18)

Wilcoxon Signed rank test. Test the hypothesis at population median is 25 against the alternative population median is not equal to 25 at 5% los

Sol^n

H0: population median is 25

```
> x = c (Data)
> 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 greater than 0.05, we accept H0: population median is 25

AN
9.370