Exam Seat No.____________

THAKUR COLLEGE OF SCIENCE & COMMERCE

NAAC Accredited with Grade "A" (3rd Cycle)

THAKUR TRUSTS

ISO 9001 : 2015 Certified

Degree College

Computer Journal

CERTIFICATE

SEMESTER 'I' UID No.

Class F.Y.B.Sc Roll No. 1803 Year 2019-2020

This is to certify that the work entered in this journal is the work of Mst. / ~~Ms.~~ Hemant Saw

who has worked for the year 2019-2020 in the Computer Laboratory.

THAKUR COLLEGE OF SCIENCE & COMMERCE Kandivali (E)

Teacher In-Charge

Head of Department

Date :

Examiner

Shot on vivo Z1Pro
Vivo AI camera

Exam Seat No.__________

THAKUR COLLEGE OF SCIENCE & COMMERCE

NAAC Accredited with Grade "A" (3rd Cycle)

THAKUR TRUSTS ®

ISO 9001 : 2015 Certified

Degree College

Computer Journal

CERTIFICATE

SEMESTER II UID No.

Class FYBSc Roll No. 1803 Year 2019-2020

This is to certify that the work entered in this journal is the work of Mst. / ~~Ms.~~ Hemant Sow

who has worked for the year 2019-2020 in the Computer Laboratory.

Teacher In-Charge

Head of Department

Date :

Examiner

Shot on vivo Z1Pro
Vivo AI camera

★★ INDEX ★★

No.	Title	Page No.	Date	Staff Member's Signature
1.	Basics of R Software and frequency	1-6	16/7/19	
2.	Frequency distribution.	7-11	23/7/19	
3.	Graph and diagram.	12-16	30/7/19	
4.	Central tendency of grouped and ungrouped data.	17-21	6/8/19	
5.	Measure of dispersion.	22-26	13/8/19	
6.	Skewness & kurtosis	27-29	27/8/19	
7.	Correlation and regression.	30-30	3/9/19	
8.	Project : Graph & Dispersion.	31-32	17/9/19	
9.	Probability.	33-34	24/9/19	
.				

Shot on vivo Z1Pro
Vivo AI camera

★★ INDEX ★★

Shot on vivo Z1Pro
Vivo AI camera

Date 11/2/19

PRACTICAL-1

BASIC OF R SOFTWARE :-

1) R is a Software for data analysis and statical Computing.

2) It is a Software by which effective data handling and outcome's storage is possible

3) It is capable of graphical display.

4) It's a free Software.

* $2^2 + |-5| + 4\times5 + 6/5$

2 ^ 2 + abs (-5) + 4 * 5 + 6

[1] 30.2

* $x=20$ → x=20

$y=2x$ y = 2 * x

$z= x+y$ z = x+y

$\sqrt{z}$ z ^ 0.5 → 7.745967

* round (2.567)

→ 3

* x=10 , y= 15, z=5 , x+y+z → 30

$\sqrt{xyz}$ → a = (x * y * z) ^ 0.5 → 27.38613.

xyz → 750

round (a) → 27

A vector in r software is denoted by the syntax "C".

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

→ [1] 4 9 25 49

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

→ [1] 4 27 25 343

* c(2,4,6,8) * 3

→ [1] 6 12 18 24

* seq(1,6,1) ^ c(2,3,4)

→ [1] 1 8 81 16 125 1296

* c(2,4,6,8) * c(-2,-3,-5,-7)

→ [1] -4 -12 -30 -56

* c(2,4,6,8) + 10

→ [1] 12 14 16 18

* c(2,4,6,8) + c(-2,-3,-1,0)

→ [1] 0 1 5 8

२६.

Q. Find the Sum, product, square root of sum and product, for the following values.

4,9, 2, 5, 7, 8, 3, 6, 15, 12, 10, 9, 8, 13, 14.

→ x = c(4,9, 2,5, 7,8, 3, 6, 15,12, 10, 9, 8, 12, 14)
y = sum (x)
y
z = prod(x)
z
y ^ 0.5
z ^ 0.5 → 125
8.559323e +12
11.18034
2925632

Q. Find the sum, product, maximum, minimum values of (2,8, 9,11, 10, 7, 6) ^2

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

Matrix (nrow = 3, ncol = 4, seq(1,8,1))

$$\begin{array}{ccc} & [\,,1] & [\,,2] \\ [1,\,] & 1 & 5 \\ [2,\,] & 2 & 6 \\ [3,\,] & 3 & 7 \\ [4,\,] & 4 & 8 \end{array}$$

$$\begin{bmatrix} 2 & 8 & 5 & 1 \\ 6 & 9 & 0 & 4 \\ 7 & 4 & 2 & 5 \end{bmatrix}$$

→ matrix (nrow = 3, nrow = 4, c(2,6,7, 8,9,4,5,0,2, 1,4,5)

$$\begin{bmatrix} 4 & 7 & 4 \\ 5 & 8 & 0 \\ 6 & 9 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 6 & 11 & 9 \\ 4 & 12 & 7 \\ 5 & 8 & 4 \end{bmatrix}$$

→ x = matrix (nrow = 3, ncol = 3, c(4,5,6,7,8,9, 4,0,2)

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

x + y

$$\begin{array}{ccc} 10 & 18 & 13 \\ 9 & 20 & 7 \\ 11 & 17 & 6 \end{array}$$

$x * 2$

8	14	8
10	16	0
12	18	4

$x * y$

24	77	36
20	96	0
30	72	8

$y * 2$

12	22	18
8	24	14
10	16	8

A14
2.12.9

2- BINOMIAL DISTRIBUTION

→ n = Total no. of trials.
→ p = success
→ q = failure
→ X = No. of success outcome of n

→ $P(x) = {}^nC_x \, p^x \, q^{n-x}$

→ $E(X) = np$

→ $V(X) = npq$

→ In R software,

dbinom $(x, n, p) \Rightarrow P(x)$
$n*p \Rightarrow E(X)$
$n*p*q \Rightarrow V(X)$
pbinom $(x, n, p) \Rightarrow P(X \le x)$.

Q.1) Suppose there are 12 MCQ in an english question paper.
Each questions has 5 answer & only one of them is correct. Find probability of
i) 4 correct answer.
ii) Atmost 4 correct answer.
iii) Atleast 3 correct answer.

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

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

iii) 1-pbinom (4,12, 1/2)
[1] 0.4416543

Q.2) find the Complete binomial distribution when n=5 & p=0.1
Soln) pbinom (0,5, 0.1)
[1] 0.59049

pbinom (1,5,0.1)
[1] 0.91854

pbinom (2,5,0.1)
[1] 0.99144

pbinom (3,5,0.1)
[1] 0.99954

pbinom (4 ,5,0.1)
[1] 0.99999

pbinom (5,5, 0.1)
[1] 1

Q.3) Find the probability of exactly 10 successes out of 100 trials with $p = 0.1$.

Soln) dbinom (10, 100, 0.1)

[1] 0.1318653

Q.4) X follows binomial distribution with $n = 12$ & $p = 0.25$ find the i) $P(X \le 25)$ ii) $P(X \ge 7)$ iii) $P(5 < X < 7)$

Soln) > pbinom (5, 12, 0.25)

[1] 0.94755

> 1 - pbinom (7, 12, 0.25)

[1] 0.0027

> pbinom (6, 12, 0.25)

[1] 0.985

Q.5) There are 10 members in a Committee probability of any member attending a meeting is 0.9 what is the probability 7 members are attending the meeting?

> pbinom (7, 10, 0.9)

[1] 0.070190

Q.6) A salesman has 20% probability of making a sale to a customer. On a typical day he will meet 30 customers. What minimum no of sale will he make with 88% probability.

> qbinom (0.88, 30, 0.2)

[1] 9

Q7) for $n=10$, $p=0.6$, find the binomial probability that & plot the graph of pmf & cdf.

Soln)
```
> 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	0.00104
2	1	0.0015
3	2	0.0010
4	3	0.42467
5	4	0.11147
6	5	0.2006
7	6	0.250822
8	7	0.21499
9	8	0.12093
10	9	0.04031
11	10	0.00604

AM
10/21

Date / /

PRACTICAL - 3

Q.1) check the following are P.M.F or NOT.

1)

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

Condition for it is :- i) $0 \le P(x) \le 1$

ii) $\sum_{x} P(x_i) = 1$

Since, it doesn't satisfy the given 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 a pmf.

3)

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

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

Hence it don't satisfy the Condition.

∴ It is not a pmf.

Q.2) 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	$P(x)$	$xP(x)$	$x^2 P(x)$
1	0.1	0.1	0.1
2	0.2	0.3	0.6
3	0.3	0.6	1.8
4	0.4	1.2	4.8
5	0.5	1.25	6.25
		= 3.45	= 13.55

$$\text{Mean} = E(X) = \Sigma x P(x) = 3.45$$

$$Var(X) = \Sigma x^2 \cdot p(x) - [E(X)]^2$$

$$= 13.55 - (3.45)^2$$

$$= 1.6475$$

In R Software,

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

Small p → ~~Prob = c (0.1, 0.2, 0.3, 0.4, 0.5)~~

```
prob = c (0.1, 0.15, 0.2, 0.3, 0.25)
a = (x * Prob)
mean = Sum(a)
[1]  3.45
```

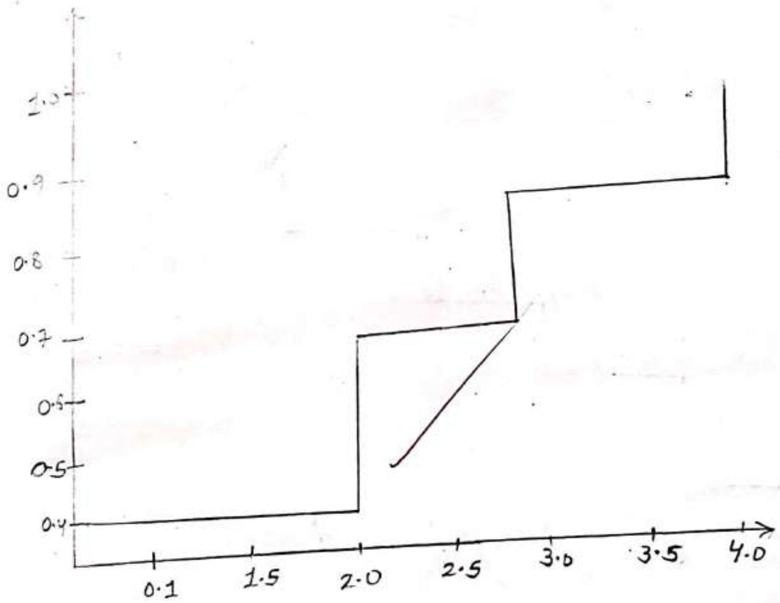
b = (x*2) * prob
var = sum(b) - mean^2
var
[1] 1.6475

Q.3) Find Mean & Variance.
soln) x = c(5,10, 15, 20, 25)
prob = c(0.1, 0.3, 0.2, 0.25, 0.15)
a = x* prob.
a
[1] 0.5 3.0 3.0 5.0 3.75

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

Q.4) Find c.d.f. of the following pmf & Draw the graph of cdf.

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")



ii) x = c (0.2, 4, 6, 8)

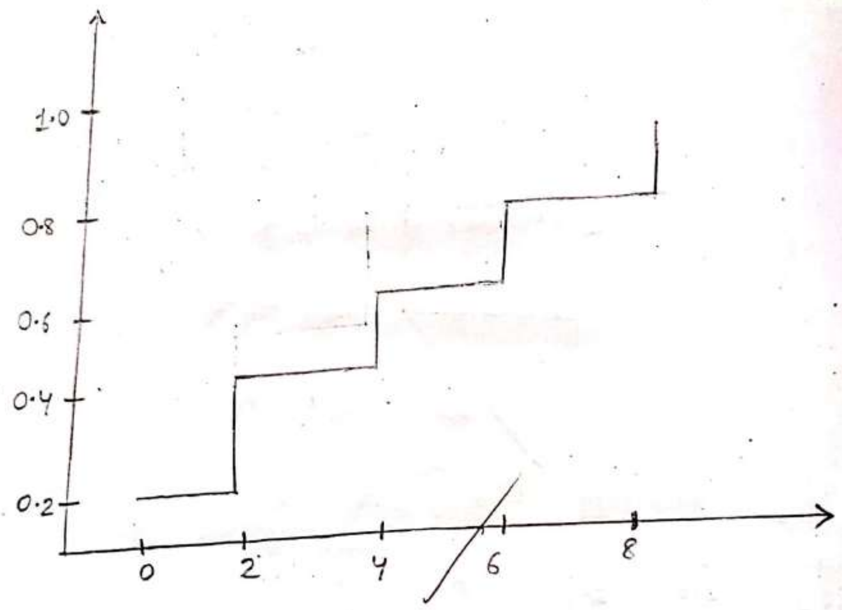
prob = c (0.2, 0.3, 0.2, 0.2, 0.1)

a = cumsum (prob)

a

[1] 0.2 0.5 0.7 0.9 1.0

> plot (x, a, "s")


1.0
0.8
0.6
0.4
0.2
0
2
4
6
8

PRACTICAL - 04

BINOMIAL DISTRIBUTION

Q1) X follows binomial distribution parameters with $n=8$, $p=0.6$. Find $(P(X=2)$, $P(X\leq3)$ & $P(X=2 \text{ or } 3)$

i) $P(X=7) = {}^8C_7 \times (0.6)^7 \times (0.4)^1$
$= 0.0895$

ii) $P(X\leq3) = P(0) + P(1) + P(2) + P(3)$
$= {}^8C_0 (0.6)^0 (0.4)^8 + {}^8C_1 (0.6)^1 (0.4)^7 + {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^3 (0.4)^5$

$= 1 \times 1 + 0.000655 + 8 \times 0.6 \times 0.00163 + 28 \times 0.36 \times 0.004096 + 56 \times 0.26 \times 0.01024$
$= 0.1736704$.

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

$= 0.1651 5012$

PRACTICAL - 05

NORMAL DISTRIBUTION

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

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 find the value of k, so that $P(X \le k) = p$
5) To generate a random sample of size n
rnorm (n, μ, σ).

Q.1) A random variable $X \sim N$ with $\mu = 10$, $\sigma = 2$
find (i) $P(X \le 7)$ ii) $P(X > 12)$
(iii) ~~$P(5 \le 7)$~~ $P(5 \le X \le 12)$
iv) $P(X < k) = 0.4$

Q.2) $X \sim N(100, 36)$
i) $P(X \le 110)$ ii) $P(X > 105)$ iii) $P(X \le 92)$
iv) $P(95 \le X \le 110)$ v) $P(X < k) = 0.9$

Q.3) Generate 10 random sample & find sample mean, median, var, SD.

Q.1) Ans)

> P1 = pnorm (7,10,2)
> cat ("P (X≤ 7) is = ", P1)
[1] P(X ≤ 7) is = 0.0668072
> P2 = 1- pnorm (12,10, 2)
> cat("p (x>12) is ", P2)
[1] p(X>12) is 0.1586553
> P3 = pnorm (12,10,2) - pnorm (5,10, 2)
> cat ("P (5≤X≤ 12) is ' , P3)
[1] P(5 ≤X≤ 12) is 0.8351351
> k= qnorm (0.4,10, 12)
> cat ("P (X<K) = 0.4, k is = ",K)
[1] P (X<K) = 0.4 , k is = 9.493306 .

Q.2) Ans)

> a = pnorm (110, 100,6)
> cat ("P (X ≤110) is", a)
[1] P (X≤ 110) is 0.95220
> b = 1- pnorm (105, 100,6)
> cat (" P(X > 105) is", b)
[1] P(X> 105) is 0.20232
> c = pnorm (92, 100,6)
> cat ("P (X<92) is ", c)
[1] P (X> 92) is 0.0912112
> d = pnorm (110, 100,6) - pnorm (95,100,6)
> cat ("P(95 ≤X≤ 110) =", d)
[1] P(95≤ X≤110) =0.74988

Ans

K= qnorm (0.9, 100,6)

[1] 107.6893

Q.3) Ans)

> X= rnorm (10,10,3)

>X

[1] 6.725493 9.672919 10.278414 11.249729
6.761599 7.422403 16.100297
12.729936 17.730984 9.449516

>am= mean(x)

>am

[1] 10.81013

>Me =Median

>Me

[1] 9.975667

>n=10

>variance = (n-1) * var(x) /n

>variance

[1] 12.746

>sd= sqrt (variance)

>sd

[1] 1.570229

Q.4) Plot the standard Normal curve.

>x =seq (-3, 3,by =0.1)

>y= dnorm (x)

>plot (x,y, xlab = "x values", ylab = "prob",
main= "standard Normal curve")

> P=0.5

>P = 325/600

[1] 0.541

> n=600
> q = 1-P
[1] 0.5
> zcal = (p-P) / (sqrt (P * q/n))
[1] 2.04241
> pvalue = 2* (1 - pnorm (abs (zcal))
[1] 0.04122
~~> pvalue = 2* (1-pnorm (abs (zcal~~
Since, pvalue < 0.5 we reject H_0.

Q.5) Test the hypothesis $H_0: \mu = 50$ against H_1; $\mu \neq 50$. A sample of 30 collected {50, 49, 44, 45, 48, 45, 49, 45, 40, 47, 55, 54, 46, 58, 47, 44, 49, 45, 60, 52, 44, 55, 56, 46, 45, 48, 49 }
> s = (50, 49, 52 .... 48, 49)
> n = length (s)
> MX = mean (s)
[1] 49.33
> M0 = 50
> sd = sd (s)
[1] 5.65
> var = (n-1) * var (s)/n
[1] 639.39
> zcal = (M0-MX) / sd / sqrt(n))
[1] 0.0215
> pvalue = 2* (1-pnorm (abs (zcal)))
[1] 0.982
Since, pvalue > 0.05 we accept H_0.

PRACTICAL-6

Z and t Distribution

Q.1) Test the hypothesis H_0, $\mu = 20$, against H_1; $\mu \neq 20$. A sample of size of 400 is selected & the sample mean is 20.2 & the S.D. 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] 1.77778
>Cat ("zcalculated is = ", zcal)
Zcalculated is = 1.77778
>pvalue = 2* (1-pnorm (abs(zcal)))
[1] 0.07544
Since, pvalue> 0.05, we accept $H_0 : \mu_1 = 20$.

Q.2) We want to test the hypothesis $H_0 : \mu = 250$ against $H_1 : \mu = 250$ A sample of size 100 has mean 275 & SD 30. Test the hypothesis at 5% level of significance.

> M0 = 250
> MX = 275
> sd = 30
> n = 100

Q.3)

46

```
>zcal = (Mx-Mo) /(sd(sqrt (n))
[1] 8.333
>cat ('zcalculated is =',zcal)
zcalculated is = 8.333
>pvalue = 2 * (1-pnorm (abs (zcal))
[1] 0
```

Since, pvalue <0.05 we reject H_0.

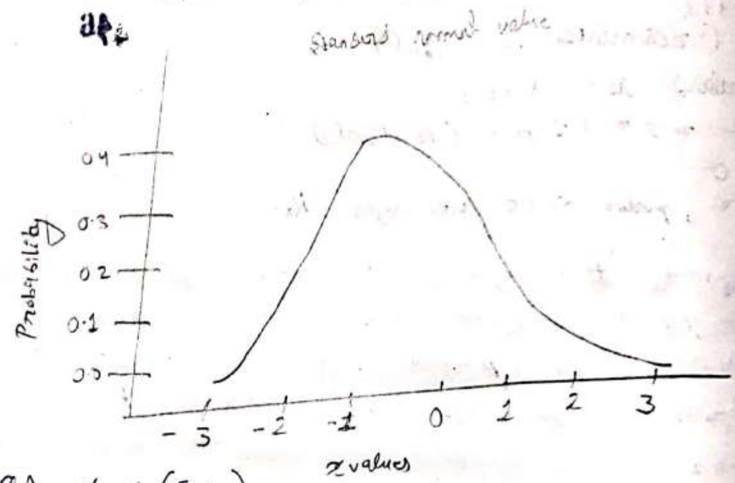
Q.3) We want to hypothesis $H_0: P=0.2$ against $H_1: P \neq 0.2$ (P= population proportion). A sample of 400 is selected & the sample proportion is calculated 0.125. Test the hypothesis at 1% level of significance.

```
>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))
[1] 0.0001
```

Since, pvalue <0.01, we reject H_0.

Q.4) In a big city 325 men out of 600 men were found to be self-employed, this info support the conclusion 1/2 of men are self-employed ?

{Prac - 5 - Q.4 - Graph}



Q.5> $X \sim N(50, 100)$

find i> $P(X \le 60)$

ii> $P(X > 65)$

iii> $P(45 \le 60)$

> P_1 = pnorm (60, 50, 10)

[1] 0.843447

> P_2 = pnorm (60, 50, 10) - pnorm (45, 50, 10)

[1] 0.5328072

> P_3 = 1 - pnorm (65, 50, 10)

[1] 0.0668072

PRACTICAL - 7

LARGE SAMPLE TEST

Q.1) Two random sample of size 100 & 2000 are drawn from two population with the SD (σ) = 2 & 3 respectively. Test the hypothesis that the two population means are equal or not at 5% level of significance. Sample are 67 & 68 respectively.

$H_0 : \mu_1 = \mu_2$

against $H_1 : \mu_1 \neq \mu_2$

n1 = 100
n2 = 2000
mx1 = 67
mx2 = 68
sd1 = 2
sd2 = 3

zcal = (mx1 - mx2) / sqrt ((sd1 ^2/n1) + (sd2 ^ 2/n2))
cat = ("zcalculated is=", zcal)
pvalue = 2 * (1- pnorm (zcal))
pvalue

→ zcalculated is = -10 . 84 652
[1] 0.

→ pvalue [1] 0
Since, pvalue is less than 0.01 we get $H_0 : \mu_1 = \mu_2$

Q.2) A study of noise level in a hospital is done following data is calculated. Sample size 84, mean = 61.2, S.D. = 7.9, Second sample size 34, Mean = 59.4, S.D. 7.8. Test the hypothesis $H_0 : \mu_1 = \mu_2$ at at 1% level of significance

→ > n1 = 84

> n2 = 34

> mx1 = 61.2

> mx2 = 59.2

> sd1 = 7.9

> sd2 = 7.8

> zcal = (mx1 - mx2) / (sqrt (sd1^2/n1 + (sd2^2/n2))

> zcal

[1] 1.31117

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

[1] 0.258001

Since,

pvalue is greater than 0.01, we accept $H_1 : \mu_0 = \mu_2$

AB

Q3) From each of two population of oranges the following samples are collected. Test the whether the proportions of bad oranges are equal or not.

1) samples size = 250
no. of bad oranges = 44

2) sample size = 200
no. of bad oranges = 30

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

```
>n1 = 250
>n2 = 200
> p1 = 44/200
>p2 = 30/200
>p = (n1*p1+n2*p2)/(n1+n2)
> q = 1-p
> q
[1] 0.8355
> zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
> zcal
[1] 0.7393
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.459
```

∴ pvalue is greater than 0.05, we accept

$H_0: P_1 = P_2$

P.4) Random Sample of 400 males & 600 females where asked whether they want the ATM nearly 20 males & 390 females where in favour of the proposal. Test the hypothesis that the proportion of males & females favouring the proposal are equal or not at 5% level of significance.

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

```
> 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))
> pvalue = 2* (1-pnorm (abs (zcal)))
> zcal
[1] -4.724751
> pvalue
[1] 2.303
```

Since, pvalue is less than 0.05, we reject $H_0: P_1 = P_2$.

Shot on vivo Z1Pro
Vivo AI camera

Scanned with CamScanner

Q.5) The following are the 2 independent sample from the population test equality of 2 proportion means at 5% level of significance.

Sample - 74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 76.

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

```
x1 = c( 74, 77, ....... 76)
n1 = length (x1)
mx1 = mean (x1)
variance = (n1 - 1) * var (x1 )/n1
> sd1 = sqrt (variance)
> sd1
[1] 2.51221

x2 = c( 72, 75, 74, 70, 70, 78, 70, 72, 75, 77, 79,
n2 = length (x2)
mx2 = mean (x2)
variance = (n2 -1) * var( x2) /n2
> sd2 = sqrt ( variance)
> sd2
[1] 3.236
> tcal = (mx1 - mx2) /sqrt ((sd1^2/n1) + (sd2^2 /n2))
> tcal
[1] 1.574
> cat ("t calculated is: ", tcal)
[1] t calculated is: 1.574
t.test (x1, x2)
pvalue = 20.1387
```

PRACTICAL- 8

SMALL SAMPLE 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. Mean is 100 (support the assumption that population)

$H_0: \mu = 100$

```
> x = c( 80, 100, 110, ......... 125 )
> length(x)
[1] 15
> t.test( x )
```

$t = 24.029$, $df = 14$, pvalue $= 8.819e^{-13}$

Since, the pvalue is less than 0.5, we accept the $H_0: \mu = 100$ at 5% level of significance.

Q.3) Two group of 10 students the following 50 marks.

group 1 - 18, 22, 21, 17, 20, 23, 17, 20, 22, 21

group 2 - 16, 20, 14, 21, 20, 18, 13, 15, 17, 21.

Test the hypothesis that there is no significant difference between the marks at 1% level of significance

$H_0: \mu = 100$.

> x = c (18, 22, ... 22, 21)

> y = c (16, 20, ..., 15, 17, 21)

> t-test (x, y)

t = 2.2573 , df = 16.376, pvalue = 0.03798

a percent confidence interval

0.1628205 5.0371795

mean of x mean of y

20.1 17.5

> The pvalue is greater than 0.01, we accept H_0 at 1% level of significance.

Shot on vivo Z1Pro
Vivo AI camera

Q.3) Two types of Medicines are used on 5 & 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 2 - (8, 9, 12, 14, 15, 10, 9)

Is there a significant difference in the efficiency of the Medicines.

```
>Med A = c( 10,12 ..... 14)
>Med 2 = c( 8,9,12, ....., 10,9)
> t.test (MedA, Med B)
```

t = 0.80354 , df = 9.7594 , pvalue = 0.4406

95 percent Confidence interval

-1.7811 3.78117

Since, pvalue is greater than 0.05, we accept $H_0: \mu_1 = \mu_2$ at 1% level of significance.

The weight reducing diet program is Conducted & observation are noted for 10 participants test whether the program's effective or not.

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

After - C(111, 114, 107, 120, 112, 115, 112, 120, 112, 119, 120)

H_0: There is no significant difference in weight.

against H_1 - the diet program reduce weight.

> x = c (120, 125, 115, 130, 123, 119, 122, 127, 128, 118)

> y = c (111, 114, ..., 112)

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

t = 17, df = 9, pvalue = 1

95 = percent Confidence Interval :

- inf 9.416551

∴ The pvalue is greater than 0.05, we accept

H_0 : μ_0 = The diet program reduce weight.

Q.5) Sample 1 = 66, 67, 75, 76, 82, 84, 90, 92) ← Before
Sample 2 = (64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97) ← After
Test the population mean are equal or not.

$H_0 : \mu_1 = \mu_2$

```
> a = c( 66, 67, ..... 92)
> b = c( 64, 66, 82, 78, 74, ..... 97)
> t-test ( a, b)
```

t = −0.6399, df = 17.994, pvalue = 0.5304

95 percent confidence interval

−12.8539 6.85399

mean of x   mean of y
80   83

Since, the pvalue is greater than 0.05, we accept $H_0 : \mu_1 = \mu_2$ at 5% level of significance

Q.17 The following are the marks before & after 52
of training program. Test the program is effective or not.

Before → = (71, 74, 69, 70, 74, 76, 70, 73, 75)
After → = (74, 77, 74, 73, 79, 76, 82, 72, 75, 78)

H_0 : No significant difference in marks.
H_1 : The test program increased the marks.

> x = c(71, 72, ..., 75)
> y = c(73, 77, ..., 78)
> t.test (x,y, paired = T, alternative = "greater")

t = -4.4691 ; df = 9, pvalue = 0.9992

95 percent Confidence Interval
-5.076639 Inf

mean of the differences
-3.6

∴ The pvalue is greater than 0.05, we accept H_0 : No increase in marks at 5% level of significance

AM 07.07.20

PRACTICAL - 9

Topic :- Large and Small test.

1) The arithmetic mean of a sample of 100 items from a large population is 52 if the S.D. is 7. Test the hypothesis that the population mean is 55 against the alternative it is more than 55 at 5% loss (level of significance).

Q.2) In a big city 350 out of 700 males are found to be smoker thus this support that exactly half of the males in the city are smokers? Test 1% LOS.

Q.3) Thousand article from a factory A are found to have 2% defectives. 1500 articals from a second factory B are found to have 1% defective. Test that 5% of LOS that the 2 factory are similar or not.

Q.4) A sample of size 400 was drawn & the sample mean is 99. Test at 5% LOS that the sample comes from a population with ~~mean 100~~ & variance 69?

67

9.5) The flower stems are selected & the heights are found to be (cm) 63, 63, 68, 69, 71, 71, 72. Test the hypothesis that the mean height is 66 not at 1% LOS.

9.6) Two random samples were drawn from two normal population & the values are A :- 66, 67, 75, 76, 83, 84, 88, 89, 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.

Solutions:-

9.1) H_0 $\mu_1 = \mu_2$ against H_1 $\mu_1 \neq \mu_2$

```
> 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
```

pvalue is <0.05 ; we reject.

Q.2) Ans) H_0 $\mu_1 = \mu_2$ against H_1 $\mu_1 \neq \mu_2$

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

Since pvalue > 0.05, we accept H_0.

Q.3) Ans) >n1 = 1000
>n2 = 1500
>p1 = 20/1000
>p2 = 10/1500
>p = (n1 * p1 + n2*p2) / (1/n1 + 1/n2)
>p
[1] 0.014
>zcal = (p1 - p2) / sqrt(p* q* (1/n1 + 1/n2))
>zcal
[1] 2.082731

```
> pvalue = 2* (1- pnorm (abs (zcal)))
> pvalue
[1] 0.03727577
```

Since pvalue < 0.05 → ∴ We reject pvalue.

Q.4) Ans)

```
> n = 900
> MX=99
> MO= 100
> variance = 64
> sd = sqrt (variance)
> sd
[1] 8
> Zcal =(MX-MO) /(sd /sqrt( n))
> Zcal
[1] -2.5
> pvalue = 2* (1-pnorm (abs( zcal))
> pvalue
[1] 0.01241933
```

Since pvalue < 0.05

∴ We reject . H_0 $\mu_1 = \mu_2$

95) H_1 $\mu_1 = \mu_2$ against H_0 $P_1 = P_2$.

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

One sample t-test

data: x

47.5f p df=6, pvalue = 5.522e-09

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

64.66479 71.62092

Sample estimation:

mean of x

68.14286

pvalue is < 0.05, we reject.

Q.6) Ans) H_1 $\mu_1 = \mu_2$ against H_0 $P_1 = P_2$

```
> x = c(16, 67, ........, 50, 92)
> y = c(64, 16, ......., 95, 97)
> F = Var.test(x,y)
> F
```

F test to Compare two variables

data : x & y

F = 0.70686 , num df = 8 , denom df = 10 , pvalue = 0.635

alternative hypothesis : true ratio of variances is not equal

x = 1

95 percent Confidence interval :

0.1833662 3.0360393

Sample estimate

ratio of variance

0.7068567

Since pvalue > 0.05 .

We accept H_1 $\mu_1 = \mu_2$.

PRACTICAL - 10

Topic :- ANOVA & Chi-square.

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

→

Cleanliness of home

cleanliness of child	clean	cleanliness of clean	Dirty
fairly clean		70	50
		80	20
Dirty		35	45

soln) → H_0: Cleanliness of child & cleanliness of home is independent.

```
> x = c(70, 80, 50, 35, 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 = 25.46 , df = 2, pvalue = 2.628e-06
```

Since, pvalue is less than 0.05, we reject H_0: CC & CH are independent.

48

Q.2) Use the following data to find a vaccination on a particular disease are depended or not.

	Disease	
	Affected	Not Affected
Vaccination Given	20	30
Not given	25	35

H_0: Disease & vaccination is ~~two~~ 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 pvalue is greater than 0.05,

Therefore we accept.

H_0: The means of the varieties are equal

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

```
> 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"
> One-way.test (value ~ ind , data = d , var.equal = T)

        One-way analysis of means

data: values & ind
f = 11.735 , num df = 3, denom df = 9 , pvalue = 0.0018
> anova = aov (values ~ ind , data = d)
> anova
Call:
    aov (formula = values ~ ind , data = d)
Terms:
                        ind        Residuals
Sum of square        71.06410       18.16667
Deg. of freedom         3              9
```

(a) Residual standard error: 1.420746
Estimated effects may be unbalanced.

Since pvalue is less than 0.05
we reject
H_0: The mess of the varieties are equal

Q.4) The following data gives life type of 4 branch type of observation.

Type	
A	20, 23, 18, 17, 18, 22, 24
B	19, 15, 12, 17, 20, 16, 17
C	21, 19, 22, 17, 20
D	15, 14, 16, 15, 14, 16

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

H_0: average life of 4 brands of type are same.

```
>x1=c( 20,23,18, 17,18, 22,24)
> x2= c( 19,15,17,20, 10, 17)
> x3=c( 21,19, 22,17)
>x4= c( 15,14, 16,15, 14,16)
> 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: values & ind.

f = 6.845 , num df = 3, denom df = 20, pvalue = 0.002391)

> anova = aov (value ~ ind , data = d)

> anova

Call:

aov (formula = value ~ ind , data = d)

Terms:

	ind	Residual
Sum of square	91.481	89.0819
deg of freedom	3	20

Residual standard error: 2.110236

Estimated effects may be unbalanced.

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 & economic condition of their home.

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

```
> 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 = 2.354e-09
```

Hence, P-value is less than 0.05.

~~we~~ ~~reject~~ We reject.

H_0: Economic & IQ.

17.2

Scanned with CamScanner

PRACTICAL NO. 11

Non - Parametric Test

Q1) Following are the amounts of Sulphur oxide emitted by industries in 20 years. Apply Sign test to test the hypothesis to check if population is 21.5
(17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26).

H_0: population Median is 21.5

```
> x = c(17, 15, 20, 29, 19, 18, 22, 25, 27, ......, 23, 24, 26)
> mx = 21.5
> sp = length (x [x > mx])
> sn = length (x [x < mx])
> n = sp + sn
> n
[1] 20
> pv = pbinom (sp, n, 0.5)
> pv
[1]     0.4119015
```

∵ 0.4119015 > 0.05

→ H_0: population median is 21.5 is accepted.

Q2) Follow 612, 619, 631, 628, 643, 640, 655, 649, 670, 663
Apply sign to test the hypothesis to check if population median is 625 against alternative i.e. H1 greater than 625 at 1% LOS
NOTE :- If alternative is greater.
pv = pbinom (sn, n, 0.5)

H_0: population median is 625
```
>x = c( 612, 619, 631, 628, ....... , 670, 663)
> me = 625
> sp = length ( x[x>me])
> sn = length ( x [x<me])
>n = sp+sn
>n
[1] 10
>pv = pbinom (sn ,0, n, 0.5)
> pv
[1] 0.05487
```
∵ 0.9514 > 0.01
∴ H_0: population median is 625 is accepted.

Q3) Ten observation are (36, 32, 29, 21, 30, 24, 20, 22, 18)
Using sign test check the hypothesis that population median is 25 against it is less than 25 at 5% LOS.

H_0: Population median is 25.
```
>x = c (36, 32, 21, 30, 24, 25, 20, 22, 20, 18)
> 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] 0.253906

∵ 0.253906 > 0.05

∴ H_0 : population Median is 25 is accepted.

Q.4) Following are some measurements - 63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 39, 72, 65. Using Wilcoxon signed Rank Test, test the hypothesis that population Median is 60 against the alternative, it is greater then 60 at 5% LOS

H_0: population Median is 60.

>x = c (63, 65, 69, 60, ..., 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

∵ 0.06186 > 0.05

H_0: Population median is 60 is accepted.

Q.5) 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, Use WSRT to test the hypothesis that population Median is 20; against the alternative is less than 5% LOS

```
> x = c( 15 , 17, 24 , ....... , 12 , 25 , 24)
> wilcox . test (x, alt = "Less" , mu = 20)
```

Wilcoxon signed Rank test with Continuity Correction

data: x

V: 35 , p-value = 0.7127

alternative hypothesis : The location is not equal to 25

$\because$ 0.7127 > 0.05

$\therefore$ H_0 is accepted.

Q.6) 20, 25, 27, 30, 18, . Test the hypothesis that the population Media 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 assigned rank test with Continuity Correction

data: x

V = 3.5 , p-value = 0.7127

alternative hypothesis : the location alternative is not equal to 2

Since p-value is More than 0.05 . We accept H_0 at 5% LOS .