PRATICAL-1.

Basics of R software.

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

Problem 1 :- Solve the following

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

Ans

(1) > 4 + 6 + 8 / 2 - 5
[1] 9

(2) > 2 ^ 2 + abs(-3) + sqrt(45)
[1] 13.7082

(3) > 5 ^ 3 + 7 * 5 * 8 + 46 / 5
[1] 414.2

(4) > sqrt(4 ^ 2 + 5 * 3 + 7 / 6)
[1] 5.671567

(5) round(46 / 7 + 9 * 8)
[1] 79.

Q// (2) problem (2)

(i) c(2, 3, 5, 7) * 2.
(ii) c(2, 3, 5, 7) * c(2, 3)
(iii) c(2, 3, 5, 7) * c(2, 3, 6, 2)
(iv) c(1, 6, 2, 3) * c(-2, -3, -4, -1)
(v) c(2, 3, 5, 7) ^ 2
(vi) c(4, 6, 8, 9, 4, 5) ^ c(1, 2, 3)
(vii) c(6, 2, 7, 5) / c(4, 5)

Ans
(i) c(2, 3, 5, 7) * 2

[1] 4 6 10 14

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

[1] 4 9 10 21

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

[1] 4 9 30 14

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

[1] -2 -18 -8 -3

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

[1] 4 9 25 49

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

[1] 4 36 512 9 16 125

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

[1] 1.50 0.40 1.75 1.00

Problem 3:-

Q/ $x = 20$

$y = 30$

$z = 2$.

find i) $x^2 + y^3 + z$.

ii) $\sqrt{x^2 + y}$

iii) $x^2 + y^2$.

Ans.

```
> x = 20
> y = 30
> z = 2
> x ^ 2 + y ^ 3 + z
[1] 27402
> sqrt (x^2 + y)
[1] 20.73644
> x^2 + y^2
[1] 1300.
```

problem 4 :-

create matrix

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

Scanned with CamScanner

my > 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

problem 5 :-

find $x+y$ and $2x+3y$ where

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

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

Ans.

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

> y <- matrix (nrow= 3, ncol= 3, data = c(10, 12, 15, -5, -4, -6, 7, 9, 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	-28	21

Q/ Problem 6 :- Create table.

a) marks of statistic of CS students

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

56

> x = c(59, 20, 25, 24, 46, 56, 55, 45, 27, 22, 47, 58, 54, 40, 50, 32, 36, 29, 35, 39)

> length (x)

[1] 20

> breaks = seq (20, 60, 5)

> a = cut (x, breaks, right = FALSE)

> b = table (a)

> c = transform (b)

> c.

op

	a	freq
1	[20,25)	3
2	[25, 30)	2
3	[30 , 35)	1
4	[35, 40)	4
5	[40, 45)	1
6	[45, 50)	3
7	[50 , 55)	2
8	[55, 60)	4

Abs
2/8/11

PRATICAL - 2 (PROBABLITY DISTRIBUTION)

Q.1) Check whether the followings are pmf or not.

i)

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

ii)

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

iii)

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

A.(i) Since $P(2) = -0.5$, it cant be a pmf.
since pmf $P(x) \geq 0 \ \forall x$.

ii) → It cannot be a pmf as in pmf
$\sum P(x) \neq 1$.

```
> prob = c(0.2, 0.2, 0.3, 0.2, 0.2)
> sum (prob)
[1] 1.1
```

Ans iii) Since $\sum P(x) = 1$ it is a pmf

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

Q 2)

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

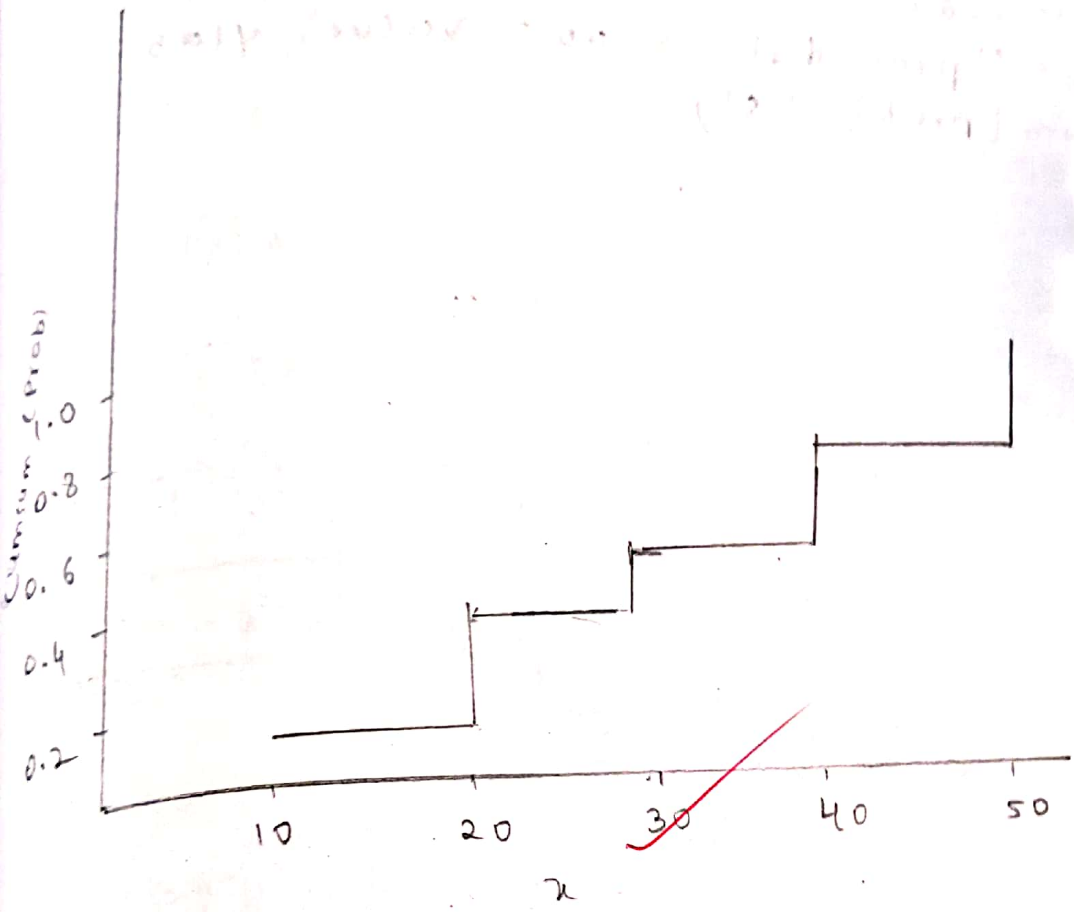
x	10	20	30	40	50.
$P(x)$	0.2	0.2	0.35	0.15	0.1.

Ans

```
> 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
```

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

```
> x = c(10, 20, 30, 40, 50)
> plot (x, cumsum(Prob), "s")
```



x	1	2	3	4	5	6
p(x)	0.15	0.25	0.1	0.2	~~0.3~~ 0.2	0.1

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

sum(prob)

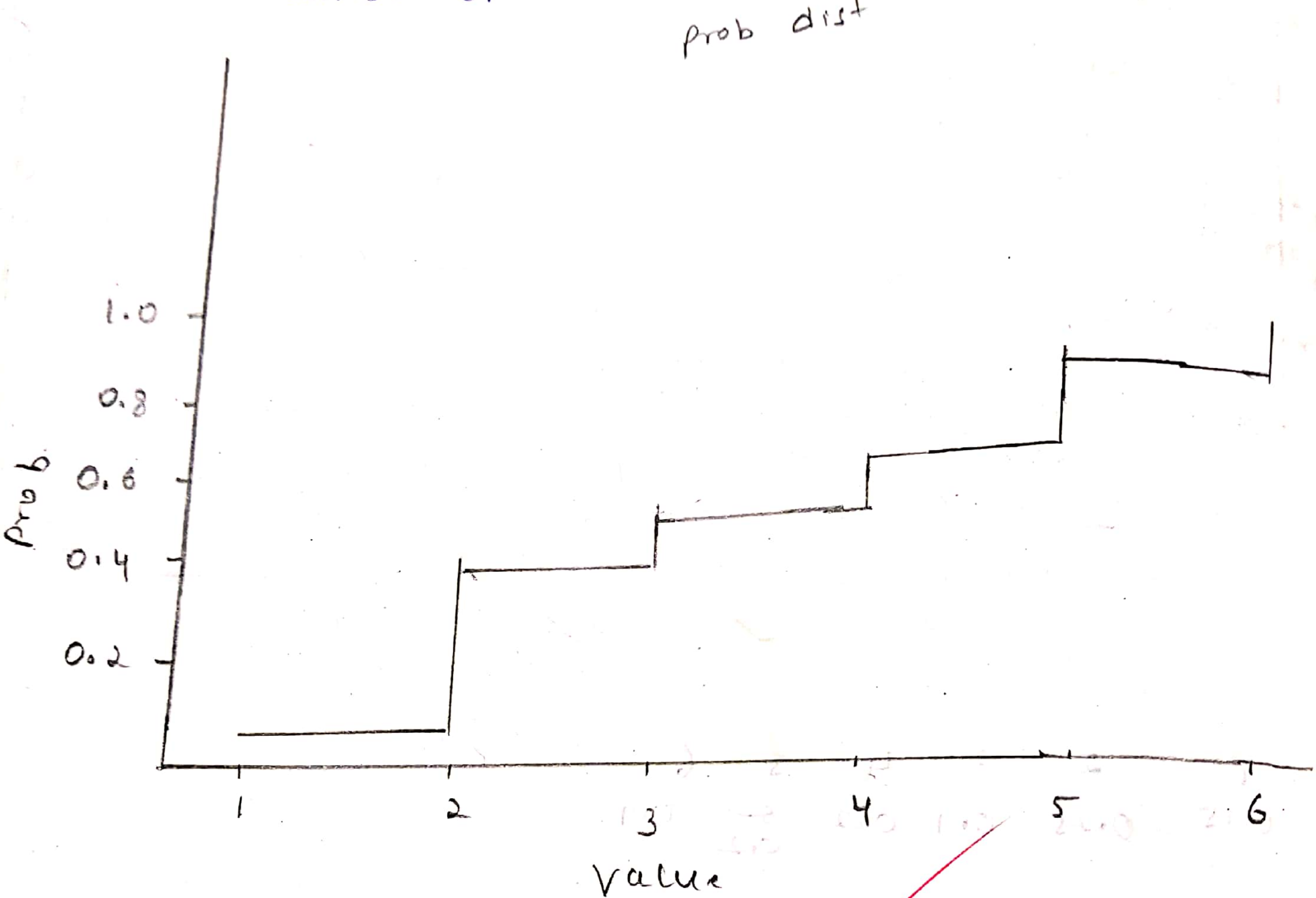
[1] 1

cumsum (prob)

[1] 0.15 0.40 0.50 0.70 0.90 1.00

F(x) = 0	$x < 1$
0.15	$1 \le x < 2$
0.40	$2 \le x < 3$
0.50	$3 \le x < 4$
0.70	$4 \le x < 5$
0.90	$5 \le x < 6$
1.00	$x \ge 6$

x = c(1, 2, 3, 4, 5, 6)
plot(x, main = "prob dist", xlab = "value", ylab = "prob
cumsum(prob), "s").



Check whether the following is pdf or not.

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

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

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

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

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

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

$= 2$

ii) $\int_0^1 3x^2$

$= 3\int_0^1 x^2$

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

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

$= (1-0)$

$= 1$

A4 5/12/19

PRATICAL - 3

Topic :- Binomial distribution.

(i) command:-

(i) $P(X=x) = dbinom(x, n, p)$.
(ii) $P(X \leq x) = Pbinom(x, n, P)$
(iii) $P(X > x) = 1 - pbinom(x, n, p)$
(iv) If x is unknown :
$P_1 = P(X \leq x)$ qbinom (P_1, n, P).

Problem
Q1) find the probability of exactly 10 success in 100 trials with p= 0.1

Q2) Suppose there are 12 mcq. Each question have 5 option out of which one is correct. find the probablity
(i) exactly 4 correct answer.
(ii) Atmost 4 correct answers.
(iii) more than 5 corred answers.

Q3) Find the complete distribution when n=5 and p= 0.1

Q4) n=12 . p= 0.25 . find the following probability.
i) $P(X=5)$
ii) $P(X \leq 5)$
iii) $P(X > 7)$
iv) $P(5 < X < 7)$

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

ans) dbinom (4, 12, 0.2)
[1] 0.1328756

> pbinom (4, 12, 0.2)
ans) [1] 0.9274445

> 1- pbinom (5, 12, 0.2)
[1] 0.01940528

ans 3 dbinom (0:5, 5, 0.1)
[1] 0.59049 0.32805, 0.07290 0.00810, 0.00045 0.00009

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

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

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

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

Ans 5 → > dbinom (0, 10, 0.15)
[1] 0.1968744
> 1 - pbinom (3, 20, 0.15)
[1] 0.3522748

Ans 6 → > qbinom (0.88, 30, 0.2)
[1] 9

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

	x value	probablity
1	0	0.0282
2	1	0.1210
3	2	0.2334
4	3	0.2668
5	4	0.2001
6	5	0.1029
7	6	0.0367
8	7	0.0090
9	8	0.0014
10	9	0.0001
11	10	0.000059.

The probablity of a salesman making a sale to a customer is 0.15. find the probablity of (1) no sells out of 10 customer.
(2) more than 3 sells out of 20 customer.

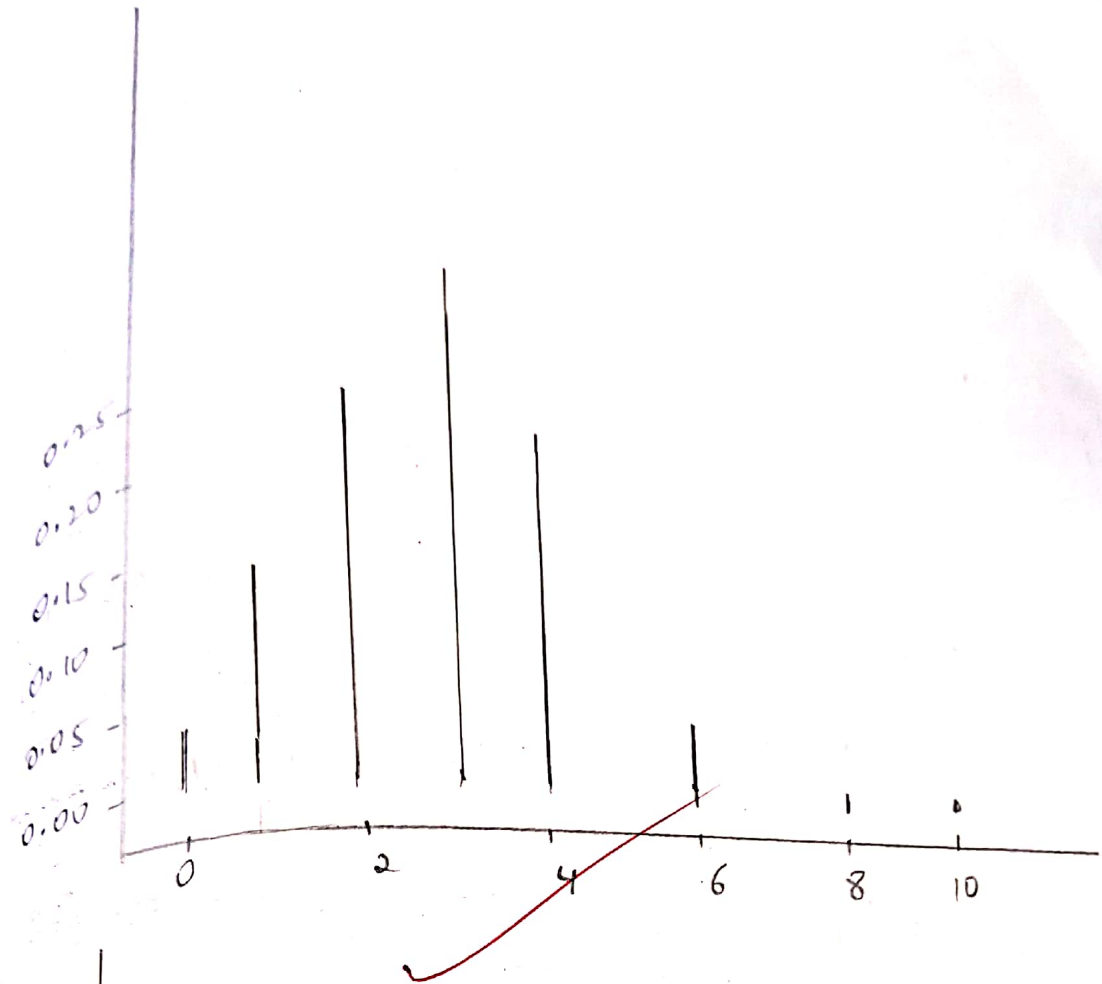
A salesman has a 20% probablity of making a sell to a customer out of 30 customer what minimum no of sells the he can make with 88% probablity.

x follows binomial distribution with $n=10$, $p=0.3$, plot the graph of p.m.f and c.d.f.

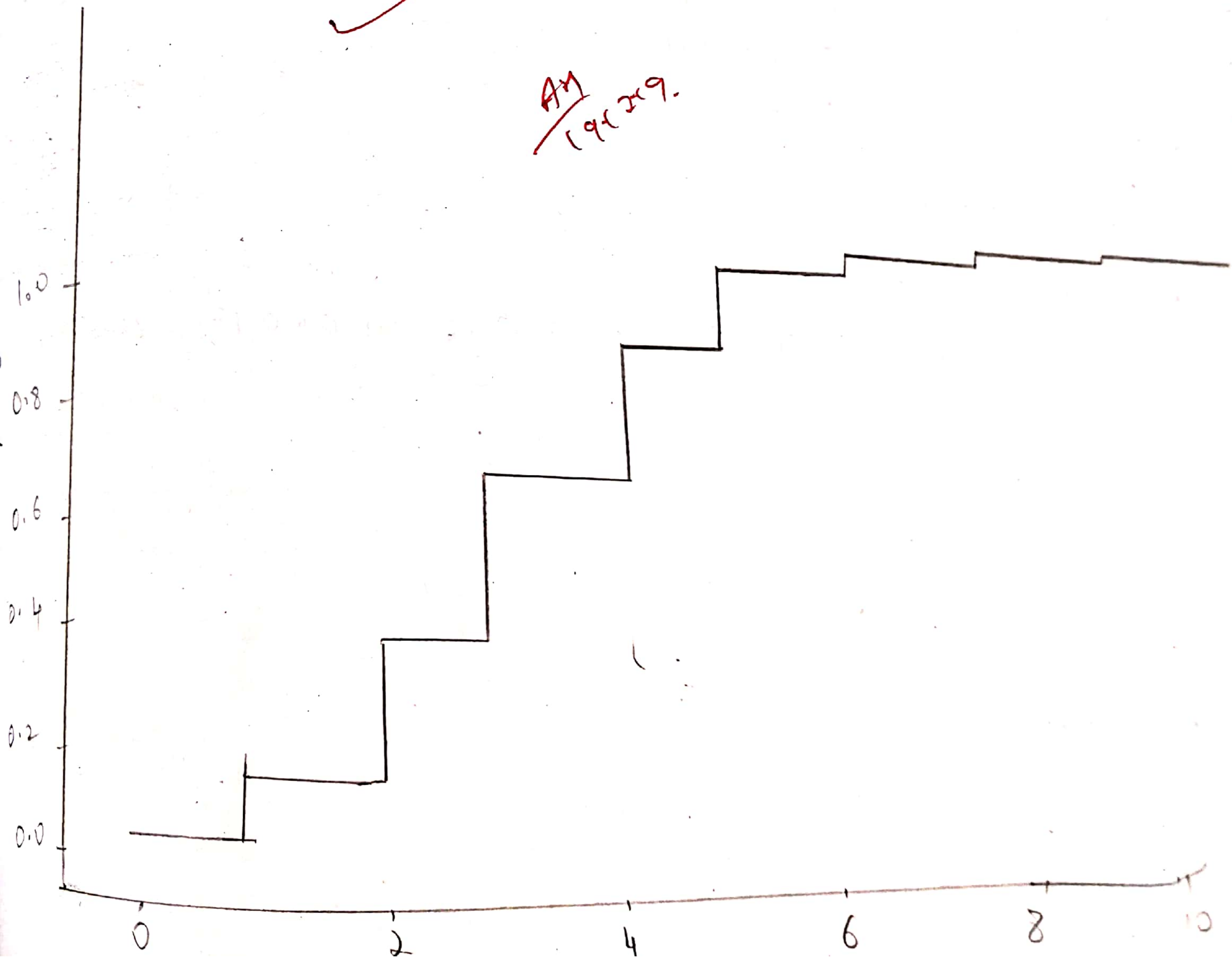
~~Ans 7~~

Ans 7

(ii) plot(x, prob, "h")



AM
19/2/19.



Ans 1)

(i) > P1 = Pnorm (15, 12, 3)

> P1

[1] 0.8413447

> Cat ("P(X <= 15) = ", P1)

P(X <= 15) = 0.8413447

> P2 = Pnorm (13, 12, 3)

> P2

[1] 0.6305587

> cat ("P(X <= 15) = ", P1)

P(X <= 15 = 0.8413447

> P2 = Pnorm (13, 12, 3) - Pnorm (10, 12, 3)

> P2

[1] 0.3780661.

> cat ("P(10 <= X < 13) = ", P2)

P(10 <= X < 13) = 0.3780661

> P3 = 1 - Pnorm (14, 12, 3)

> P3

[1] 0.2524925.

> rnorm (5, 12, 3)

[1] 10.000860   11.26023   13.178796   4.060977   10.041363

PRATICAL 4:-

Topic :- Normal distribution.

$P(X=x) =$ dnorm (x, μ, σ)

$P(X \leq x) =$ pnorm (x, μ, σ)

$P(X > x) = 1 -$ pnorm (x, μ, σ)

To generate random numbers from a normal distribution (in random numbers).

The R code is :-

~~rnoo~~ rnorm (n, μ, σ)

problems :

1) A random variable X follows normal distribution with mean $= \mu = 12$ and S.D $= \sigma = 3$ find

(i) $P(X \leq 15)$

(ii) $P(10 \leq X \leq 13)$

(iii) $P(X > 14)$

(iv) Generate 5 observation. (random numbers)

2) X follows Normal Distribution with $\mu = 10$ $\sigma = 2$, find :- (i) $P(X \leq 7)$ (ii) $P(5 < X < 12)$ (iii) $P(X > 12)$ (iv) Generate 10 observation (v) find k such that probablity $P(X < k) = 0.4$

Q3) Generate 5 random numbers from a normal distribution $\mu = 15$, $\sigma = 4$ find sample mean, median S.D and print it

Q4) X follows $X \sim N(30, 100)$ $\mu = 30$, $\sigma = 10$

i) $P(X \leq 40)$

ii) $P(X > 35)$

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

iv) find k such that $P(X < k) = 0.6$

Q5) plot the standard normal graph.

prs 2

> p1 = pnorm (7, 10, 2)
> p1
[1] 0.0668072

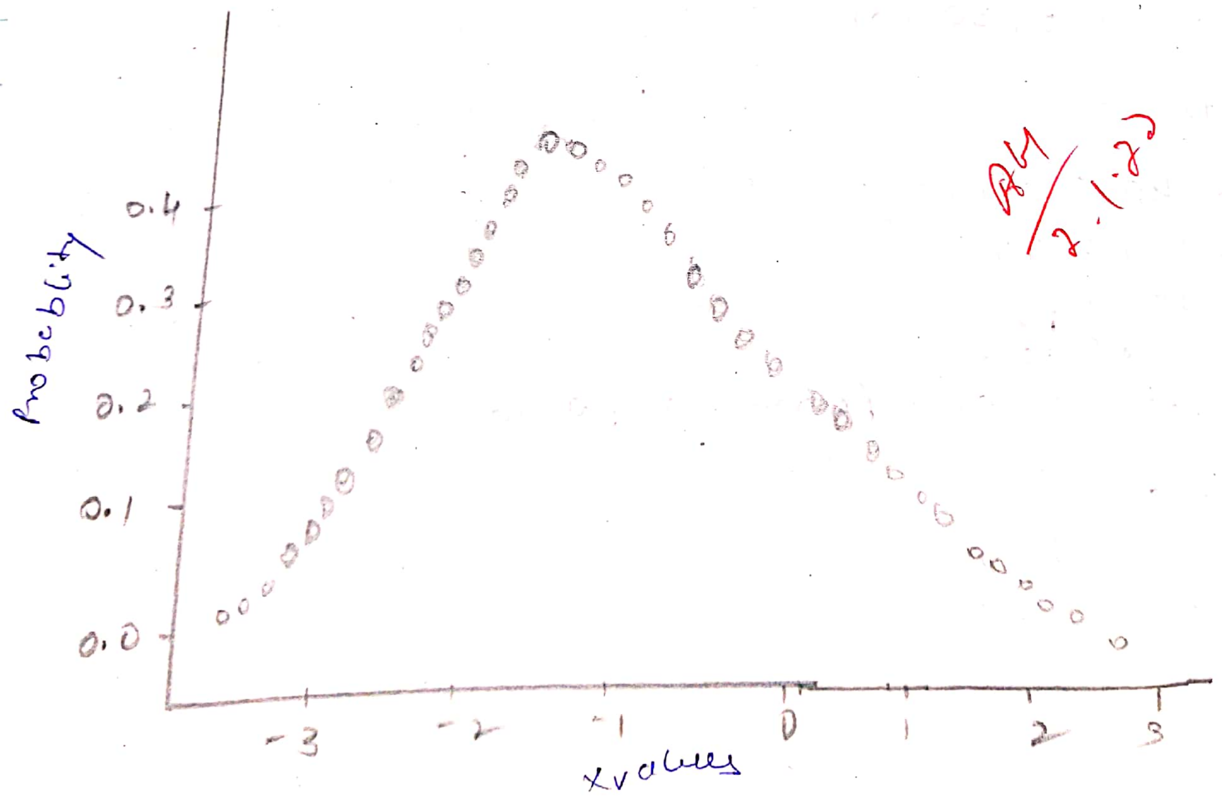
> p2 = pnorm (12, 10, 2) - pnorm (5, 10, 3)
> p2
[1] 0.7935544
> p2 = pnorm(12, 10, 2) - pnorm (5, 10, 2)
> p2
[1] 0.8351351
> p3 = 1 - pnorm (12, 10, 2)
> p3
[1] 0.1586553
> rnorm (10, 10, 2)
[1] 7.214873 6.846673 12.425003 10.724589 11.129223
9.917492 10.114098 10.856805 11.489663
> qnorm (0.4, 10, 2)
[1] 9.493306

prs 4

p1 = pnorm (40, 30, 10)
> p1
[1] 0.8413447
> p2 = 1 - pnorm (35, 30, 10)
> p2
[1] 0.3085375
> cat ("p(x<=40)=", p1)
p(x<=40)= 0.8413447
> cat ("p(x>35=", p2)
p(x>35)= 0.3085375
> p3 = pnorm (35, 30, 10) - pnorm (25, 30, 10)
> p3
[1] 0.3829249
> qnorm (0.6, 30, 10)
[1] 32.53347

Ans 3 → > x = rnorm (5, 15, 4)
> am= mean (x)
> me = median (x)
> n= 5
> variance = (n-1) * var(x) /n
> sd = sqrt (variance)
> x
[1] 14.72018 15.11361 12.39856 11.39886 13.70728
> am
[1] 13.4677
> me
[1] 13.70728
> variance
[1] 1.9516858
> cat ("sample mean is=",am)
Sample mean is = 13.4677 > cat ("sample median is", me)
sample median is 13.70728] cat ("sample sd is", sd)
sample sd is 1.397018)

Ans 5) > x = seq (-3, 3, by = 0.1)
> y= dnorm (x)
> plot (x, y, xlab="xvalues", ylab ="probablity", main = "standard normal graph")



AV 7.1.22

23

PRATICAL-5

Normal and t. Test.

1) Test the hypothesis $H_0: \mu = 15$, $H_1: \mu \neq 15$ where H_0: Null Hypothesis and H_1: Alternative Hypothesis. A random sample mean is 14 and s.d is 3. Test the Hypothesis at 5% level of significance.

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

∴ Pvalue is less than 0.05, we reject $H_0: \mu = 15$

2) Test the Hypothesis $H_0: \mu = 10$ against $H_1: \mu \neq 10$. If random sample of size 400 is drawn with sample mean 10.2 and s.d 2.25. test the Hypothesis at 5% level at significance

⇒ > mo: 10; mx = 10.2; sd = 2.25; n = 400
> zcal = (mx - mo) / (sd / (sqrt (n)))
> zcal.
[1] 1.77778

> cat ("calculated value at z is: ", zcat)
calculated value at z is 1.77778
> pvalue = 2* (1- pnorm (abs (zcal)))
> pvalue
[1] 0.07544036
∴ Pvalue is more than 0.05, we accept Ho: u = 10

Test the Hypothesis (Ho). Proportion of smokers in our college is 0.2. A sample is calculated and it is collected and calculated as 0.125. Test the Hypothesis at 5% level of significance (sample size is 400)

⇒ P = 0.2 ; p = 0.125 ; n = 400 ; q = 1-P
> zcal = (p-P) / (sqrt ~~(P* Q)~~/n (P* Q/n))
> zcal
[1] -3.75
> cat ("calculated value ~~at z~~ is", zcal)
calculated value at ~~z~~ is -3.75
pvalue = 2* (1- pnorm (abs (zcal)))
> pvalue
[1] 0.00017 6834
∴ Pvalue is less than 0.05, we reject Ho: = ~~0.20~~ 0.2

Last year, farmers lost 20% of their crops. A random sample of 60 fields are collected and ~~if~~ it is found that 9 fields are incet polluted. Test the Hypothesis at 1% of significance

33

```
> P = 20/100; p = 9/60; n = 60; Q = 1-P
> zcal = 2* (1- pnorm(abs(zcal)))
[1] -0.9682458
> cat ("calculated value at z is", zcal)
 Calculated value at z is -0.9682458
> pvalue = 2* (1- pnorm(abs(zcal)))
> pvalue
[1] 0.3324216
```

∴ Pvalue is more than 0.05, we accept Ho.

5) Test the Hypothesis Ho: μ=12.5 from the following sample at 5% level of significance.

```
> x = c(12.25, 11.97, 12.15, 12.08, 13.31, 12.28, 11.94, 11.89, 12.16, 12.04)
> n = length (x)
> n
[1] 10
> mo = 12.5
> mo
[1] 12.5
> mx = mean(x)
> mx
> var = (n-1) * var(x)/n
 var
[1] 0.019521
> sd = sqrt (var)
> sd
[1] 0.1397176
> zcal = (mx-mo) /(sd/(sqrt(n)))
> zcal
[1] -8.894909
> cat ("calculated value is", zcal)
  calculated value is -8.894909
> pvalue = 2* (1-pnorm(abs (zcal)))
> pvalue
[1] 0
```

∴ Pvalue is less than 0.05, we reject Ho

PRATICAL - 6 (Large sample test) 57

> mo = 250 ; mx = 275 ; sd = 30 ; n = 100
> zcal = (mx - mo) / (sd / sqrt(n))
> cat ("calculated value is " zcal)
calculated value is 8.333333
> pval = 2 * (1 - pnorm (abs (zcal)))
> pval
[1] 0
> cat ("since, Pvalue is less than 0.05, Hence we reject Ho")
since, Pvalue is less than 0.05, Hence we reject Ho

> P = 0.8 , Q = 1 - P ; p = 80/1000 ; n = 1000
> zcal = (p - P) / (sqrt(P * Q/n))
> cat ("calculated value is", zcal)
calculated value is - 3.452847
> Pval = 2 * (1 - pnorm (abs (zcal)))
> Pval
[1] 7.72268e - 05
> cat ("since, Pvalue is less than 0.05, Hence, we reject Ho at 5% level")
since, Pvalue is less than 0.05, Hence, we reject Ho at 5% level of significance.

63.

→ 3) > 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))
> cat ("calculated value is", zcal)
> calculated value is -5.163978
> Pval = 2* (1- Pnorm (abs (zcal)))
> Pval.
[1] 2.417754e - 07
> cat (" since, Pvalue is less than 0.05, Hence;
we reject Ho at 5% level of significance")
> since pvalue is less than 0.05. Hence, we reject
Ho at 5% level of significance.

→ 4) > n = 84 ; n2 = 34 ; mx1 = 61.2 ; mx2 = 59.4 ; sd1 = 7 ; sd2
> zcal = (mx1 - mx2) / sqrt((sd1^2/n1)+ (sd2^2/n2))
> cat ("calculated value is", zcal)
calculate value is 1.162528.
> pvalue = 2* (1- Pnorm (abs (zcal)))
> Pval.
[1] 0.245024
> cat ("since, Pvalue is greater than 0.05. Hence. we acc
at 1% level of significance")
> ~~cat~~ (since Pvalue is greater than 0.05. Hence we
accept that at 1% level of significance.

Questions (Prac 6)

1) let the population mean (amount spent per customer in a resturant) is 250. A sample of 100 customer is selected. The sample mean is calculated as 275 and s.d is 30. Test the hypothes that population is 250 or not at 5% level of significance.

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

3) 2 random sample of size 1000 + 2000 are drawn from 2 population with the same s.d = 2.5. The sample means are 67.5 + 68 respectively. Test the hypothesis at 5% level of significance.

4) A study of noise level of 2 hospitals is given below Test the claim that the 2 hospitals have same level of noise at 1% level of significance.

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

5) In a sample of 500 students in a college, 400 use blue ink in another college from a sample of 900, 450 use blue ink Test the hypothesis that the proportion of students using blue ink using 2 colleges are equal or not at 1% level of significance.

5) Ho: $P_1 = P_2$ against $H_1: \mu_1 \neq \mu_2$

```
> n1 = 600; n2 = 900; P1 = 400/600 ; P2 = 450/900
> P=(n1 * P1 + n2 * P2) /(n1+n2)
> q = 1-P
> zcal = (P1-P2)/ sqr( P* q* (1/n1 + 1/n2))
> cat ("calculated value is", zcal)
Calculate value is 6.381534.
> P
[1] 0.5666667
> Pval = 2 * ((1- Pnorm(abs (zcal)))
> Pval
[1] 1.753222e-10
> cat (" since, pvalue is less than 0.05 Hence we
reject ~~the~~ Ho at 1% level of significance")
```

since, Pvalue is less than 0.05. Hence we reject Ho at 1% level of significance.

6) For sample size, $n_1 = 200$, $n_2 = 200$, $P_1 = 44/200$, $P_2 = 30/200$
Test at 5% level of significance.

```
> n1 = 200; n2 = 200 , P1 = 44/200; P2 = 30/200
> P (n1 * P1 + n2 * P2) / (n1+n2)
> P
[1] 0.185
> q = 1-P
> q
[1] 0.815
```

> zcal = (p1 - p2) / sqrt (p * q * (1/n1 + 1/n2))
> cat ("calculated value is ", zcal)

calculated value is 1.802741
> pval = 2 * (1 - pnorm (abs(zcal)))
> pval
[1] 0.07142888
> cat ("since, pvalue is greater than 0.05 Hence we accept Ho at 5% level of significance")

since pvalue is greater than 0.05. Hence we accept Ho at 5% level of significance.

PRATICAL- 7

Aim: Small sample spale

1) The marks of 10 students are given by 63, 63, 66, 67, 68, 69, 70, 70, 71, 72. Test the hypothesis that the sample comes from the population with the average 66. $\mu_0 = \mu = 66$.

Ans. Since p-value is less than 0.05 we reject the hypothesis $H_0: \mu = 66$ at 5% level of significance.

code :-

```
> x = c(63, 63, 66, 67, 68, 69, 70, 70, ~~70~~ 71, 72)
> t.test(x)

        One sample t-test

data: x
t = 68.319, df = 9, p-value = 1.558e-13
alternative hypothesis: true mean is not equal to
95 percent confidence interval:
  65.65171    70.14829

sample estimates:
  mean of x
  67.9
```

) Two groups of students score the following marks test the hypothesis that there is no significance difference between the two group
Group 1 :- 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
Group 2 :- 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Ho : There is no difference between the two group

>x <- c(18, 22, 21, 17, 20, 17, 23, 20, 22, 21)
>y <- c(16, 20, 14, 21, 20, 18, 13, 15, 17, 21)
> t.test (x, y)

welch Two sample t-test
data; x and y
t = 2.2573, df = 16.376, P-value = 0.03798
alternative hypothesis; true difference in means is not equal to 0
95 per cent confidence interval:
0.1628205 5.0371795
sample estimates:
mean of x ~~mean~~ of y
20.1 17.5
> Pvalue = 0.0378
> if (Pvalue > 0.05) { cat ("accept Ho") } else { cat ("reject Ho") }
reject ho>

Ans:- Since p-value is less than 0.05 we reject the hypothesis at 5% level of significance.

Q3) Sale data of 6 shop before and after special campaign given below:
Before : 53, 28, 31, 48, 50, 42
After : 58, 29, 30, 55, 56, 45
Test the hypothesis the campaign is effective or not.

Ho: There is no significance difference of sales before and after the campaign

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

Paired t-test.

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

```
> Pvalue = 0.9806
> if (Pvalue > 0.05) {cat ("Accept Ho")} else {cat ("Reject
```

Accept : H.

Two medicines are applied to 2 group of patient
Group1: 10, 12, 13, 11, 14
Group2: 8, 9, 12, 14, 15, 10, 9
is here any significance between two groups

```
a: c(10, 12, 13, 11, 14)
b: c(18, 9, 12, 14, 15, 10, 9)
> t.test (a, b)
```

Welch Two sample t-test

data a and b

t: 0.65591 df = 9.567 , Pvalue = 0.527

alternatives hypothesis ; true difference in mean is not equal to 0 as percent confidence interval

: 1954382 3.534382

sample estimates

mean of x mean of 4..

11.8 11.0,

```
> Pvalue = 0.5273
> if (Pvalue > 0.05) {cat ("Accept Ho")}else
{cat ("RejectH.")}
```

∴ we Accept Ho.

05) Ho: There is no significance difference between before and after

> A = c(120, 125, 115, 130, 123, 119)
> B = c(100, 114, 95, 90, 115, 99)
> t.test (A, B, Pared = T, alternative = "less")

Pared t-test

data: A and B
t = 4.345, df = 5, Pvalue - 0.9963

alternative hypothesis: True difference in mean is less than 0.95 percent confidence interval:
Inf 29.0295

sample estimates:
19.83333

> Pvalue ; 0.9963
> if (Pvalue > 0.05) { cat = ("Accept H0.") else
{ cat ("Reject H0.")}

Accept Ho.

PRATICAL-8.

1) Ho : $\mu = 55$

H1 : $\mu \neq 55$

> n= 100

> mx= 52

> mo = 65

> sd = 7

> zcal = (mx -mo)/(sd / sqrt(n))

> z cal

[1] -4.285714

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

> p value

> 1.82153e-05

Since the pvalue is less than 0.5 ∴ Ho: $\mu = 55$ at 5% level of significance is rejected.

02) Ho: $P = \frac{1}{2} = 1$

> P = 0.5

> Q = 1 - P

> p= 350/700

> n = 700

> zcal = (p- P) / sqrt ((P * Q /n))

> zcal

[1] 0

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

Accept the pvalue at $H_0 : \mu = 1/2$

Q3;

> n1= 1000
> n2 = 1500
> p1 = 0.02
> P2= 0.01
> P= (n1* P1+ n2 * P2) / (n1 + n2)
> Q= 1-P
> zcal = (P1-P2) / sqrt (P* Q * (1/n1+1/n2))
> zcal
[1] 2.084842
> Pvalue= 2* ((1-pnorm (abs (zcal)))
> pvalue
[1] 0.03708364.

since the pvalue is less than 0.05 ∴ we reject the pvalue at 5% level of significance.

4) Ho: u = 100

mx = 99 sd = 8

mo = 100 n = 400

zcal = (mx - mo) / (sd / (sqrt (n)))

zcal

[1] - 2.5

pvalue = 2 * (1 - pnorm(abs (z)))

pvalue

[1] 0.01241933

we reject the value at 5% level of significance.

5) x = c(63, 63, 68, 69, 71, 71, 72)

> t.test (x)

One sample t-test

data: x

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

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

64.66479 71.62092

sample estimates:

mean of x

68.14286

7) Ho = μ = 100

> mx = 1150

> mo = 1200

> sd = 125

> n = 100

> zcal = (mx - mo / (sd / sqrt (n))

> zcal

[1] -4

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

> pvalue

> 6.334248e-05

we reject the value at 5% level of significance

6) x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)

> y = c(64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)

> var.test (x, y)

F test to compare two variancy

data: x and y

f = 0.70686, num df = 8, denom df = 10, p-value: 0.635

alternative hypothesis: true ratio of variances is not equal to 1 95

95 percent confidence interval:

0.1833662 3.036093

sample estimates:

ratio of variances

0.7068567

we accept.

Scanned with CamScanner

8) > n1 = 200
> n2 = 300
> p1 = 44/200
> p2 = 56/300
> p = (n1 * p1 + n2 * p2) / (n1+n2)
> Q = 1 - p
> z cal = (p1 - p2) / sqrt (p * Q * (1/n1 + 1/n2))
> z cal
[1] 0.9128709
pvalue = (2 * (1 - pnorm (abs(z)))
pvalue
[1] 0.3613104
p
0.2

we accept the value at 5% level of significance.

AN
13.1.20

PRATICAL - 9

AIM: Chi square test & ANOVA (Analysis of variance)

Q1) Use the following data to test wether the condition of the home and child are independent or not

	Condition of Home	
	Clean	Dirty
Clean	70	50
Fairly clean	80	20
Dirty	35	45

Ho: condition of home and 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
∴ pvalue is less than 0.05
∴ we reject Ho

1) Test the hypothesis that vaccines and deseases are independent or not.

Vaccines

	Affected	Not Affected
Affected	70	46
Not Affected	35	37

→ Ho = vaccines and desease 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 = 1, p-value = 0.1545
∵ Pvalue is greater than 0.05
∴ We accept Ho.

Q3] Perform ANOVA for following data

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

→ Ho : The mean 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 : values and ind
F = 11.735, num df = 3, denom df = 9, Pvalue = 0.00
> anova = aov(values ~ ind, data = d)
> summary(anova)
            df    sum sq    meansq    F value   Pr(>F)
ind          3    71.06     23.688    11.73     0.001
Residual     9    18.17     2.019
---
signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The following data gives the life of the tyre of 4 ~~other~~ brands

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

H_0 : The average life A, B, C and 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"
> one way.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 = 6.002349
```

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

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

	Df	Sum sq	mean sq	F value	Pr(>F)
ind	3	91.43	30.479	6.845	6.00237**
Residuals	20	89.06	4.453		

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

> x= read. csv ("c:/users/admis/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 $ ~~stats~~ maths)
> am1
(1) 39.4
> m1= median (x $ stat s)
> m1
(1) 38.5
> m = median (x $ maths)
(1) 37
> n= length (x $ stats)
> n
(1) 10

```
> sd = sqrt ((n-1) * var (x $ stat ) / n)
> sd
[1] 12.6491
> n1 = length (x $maths)
> n1
[1] 10
> sd1 = sqrt ((n-1)* var (x $maths)/n )
> sd1
[1] 15.2
> cor (x $stat, x $ maths)
[1] 0.830618
```

PRATICAL- 10

Aim:- Non parametric test

a) following are the amounts of sulphur oxide emmited by industrys in 20 days. Apply sign test. to test the hypothesis that the population median is 21.5 at 5% Los.

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(17, 15, 20, ............ 26)
> me= 21.5
> sp = length (x [x > me])
> sp
[1] 9
> sn = length (x [x < me])
> sn
[1] 11
> n = sp+ sn
> n
[1] 20
> pv = pbinom ( sp, n, 0.5)
> pv
[1] 0.4119015
```

∴ Pr is less than 0.05 we ~~accept~~ reject Ho at 5% level of significance.

If H_1 : me ≠
Or me <
then, Pr = pbinom(sp, n, 0.5)

If H_1 : me > then,
Pr = pbinom (Sn, n, 0.5

Q2) Following is the data of 10 observation, apply sign test. The population median is 625. The alternative H_1 is more than 625.

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

Soln

Ho : Population median is 625
H_1 : me > 625.

```
x = c( 612, 619, 631, 628, 643, 640, 655, 649,
  670, 663)
> me = 625
> sp = length( x[x>me ])
> sn = length (x[x<me])
[1]   n = sp+sn
> n
[1] 10
```

> ~~pv = pbinom (sp, n, 0.5)~~
> ~~pv~~
~~0.9292578~~

(i) > pv = pbinom (sn, n, 0.5)
> pv
0.0546875

(ii) ∴ pv is greater than 0.05 we accept Ho at 5% level of significance

Q3) following are the data of samples. Test the hypothesis that the population median is 60 against the alternatives i.e. more than 60 at 5% level of significance using wilcoxon signed rank test

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

soln : Ho : me = 60 against H1 : me > 60

> x = c (63, 65, 60, 89 . . . . 69)
> wilcox. test (x, 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.
∴ pvalue is less than 0.05 we reject Ho at 5% level of significance

Note :- If ① H1 : me < ~~then~~ alter = "less"

② H1 : me ≠ then alter = "two. sided".

Q4) using WSR test. The population median is 12 or less than 12.

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

Soln:- Ho : me = 12 against H1 : me < 12.

>x = c(15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26)

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

data: x

V = 66, p-value = 0.9986.

alternative hypothesis: true location is less than 12

∴ Pvalue is greater than 0.05 we accept Ho.

Q5] The weights of students before and after the stopped smoking are given below. Using WSR test that there is no significant change

Before	After
65	72
75	74
75	72
62	66
72	73

soln: Ho: Before and after there is no change against H1: There is change.

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

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

>d = x - y

> wilcox.test (d, alter = "two.sided", mu = 0)

data : d

V = 4.5 , P-value = 0.4982

alternative hypothesis: true location is not equal to 0.

∴ Pvalue is greater than 0.05 therefore we accept Ho.

Ay
27-2

Scanned with CamScanner

★★ INDEX ★★

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-cs Roll No. 1734 Year 2019-2020

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

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

A. Ulan
Teacher In-Charge

Head of Department

Date : 27/2/2020

Examiner