18

Practical 1

Aim :- Basics of R Software

1) R is a Software for Startical analysis a
data computing.

2) It is an effective data handling Software :
outcome storage is possible

3) It is capable of graphical display

4) It is a free Software

Q1) Solve the following

1) 4+6 +8÷2 -5

> 4+6+ 8/2 -5

[1] 9

2) $2^2 + |-3| + \sqrt{45}$

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

[1] 13.7082

3) $5^3 + 7 \times 5 \times 8 + 46/5$

> 5 ^3 +7*5*8 +46/5

[1] 414.2

1). $\sqrt{4^2+5\times3+7/6}$

> sqrt (4^2 +5*3 +7/6)

[1] 5.671567

> round off

$46 \div 7 + 9\times 8$

> round(46/ 7+9 *8)

[1] 79

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

[1] 4,6 ,10 ,14

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

[1] 4, 9, 30, 14

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

[1] 4 9 25 49

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

[1] 1.50 0.40 1.75 1.00

3) > x = 20 > y = 30 > z = 2

> x^2 + y^3 +z

[1] 27402

> sqrt (x^2 +y)

[1] 20.73644

> x^2 +y^2

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

[1] 4 9 10 21

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

[1] -2 , -18 , -8 -3

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

[1] 4 36 512 9 16 125

52

Scanned with CamScanner

Q4) > X <- matrix (nrow = 4, ncol = 2, data = c(1,2,3,4,5,6,

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

Q5) Find x+y and 2x+3y where $x = \begin{bmatrix} 4 & -0 & 6 \\ 7 & 2 & 7 \\ 9 & -5 & 3 \end{bmatrix}$

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

> X <- matrix (n row = 3, n col = 3, data = c(4,7,9,-2,
6,7,3))

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

> Y <- matrix(nrow=3, ncol=3, data=c(10,12,15,5,
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

marks of statistics of CS Batch B

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

```
> x = c(data)
> breaks = seq(20, 60, 5)
> a = cut(x, breaks, right = FALSE)
> b = table(a)
> c = transform(b)
> c
```

Scanned with CamScanner

Practical-2

TOPIC:- Probability Distribution

1) Check whether the following are p.m.f or not

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

If the given data is p.m.f then $\Sigma P(x) = 1$

$\therefore P(0) + P(1) + P(2) + P(3) + P(4) + P(5) = P(x)$
$= 0.1 + 0.2 + -0.5 + 0.4 + 0.3 + 0.5$
$= 1.0$

$\therefore$ $P(2) = -0.5$ it can be a probablicity mass function

$\therefore P(x) \geq 0 \;\; \forall\, x$

2)

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

The condition for p.m.f is $\Sigma P(x) = 1$

So

$\Sigma P(x) = P(1) + P(2) + P(3) + P(4) + P(5)$

$= 0.2 + 0.2 + 0.3 + 0.2 + 0.2$

$= 1.1$

$\therefore$ The given data is not a pmf because the $P(x) \neq 1$

3)

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

The condition for pmf is

1) $P(x) \geq 0 \;\forall x$ satisfy

2) $\Sigma P(x) = 1$

$\Sigma P(x) = P(10) + P(20) + P(30) + P(40) + P(50)$

$= 0.2 + 0.2 + 0.35 + 0.15 + 0.1$

$= 1$

$\therefore$ The given data is p.m.f

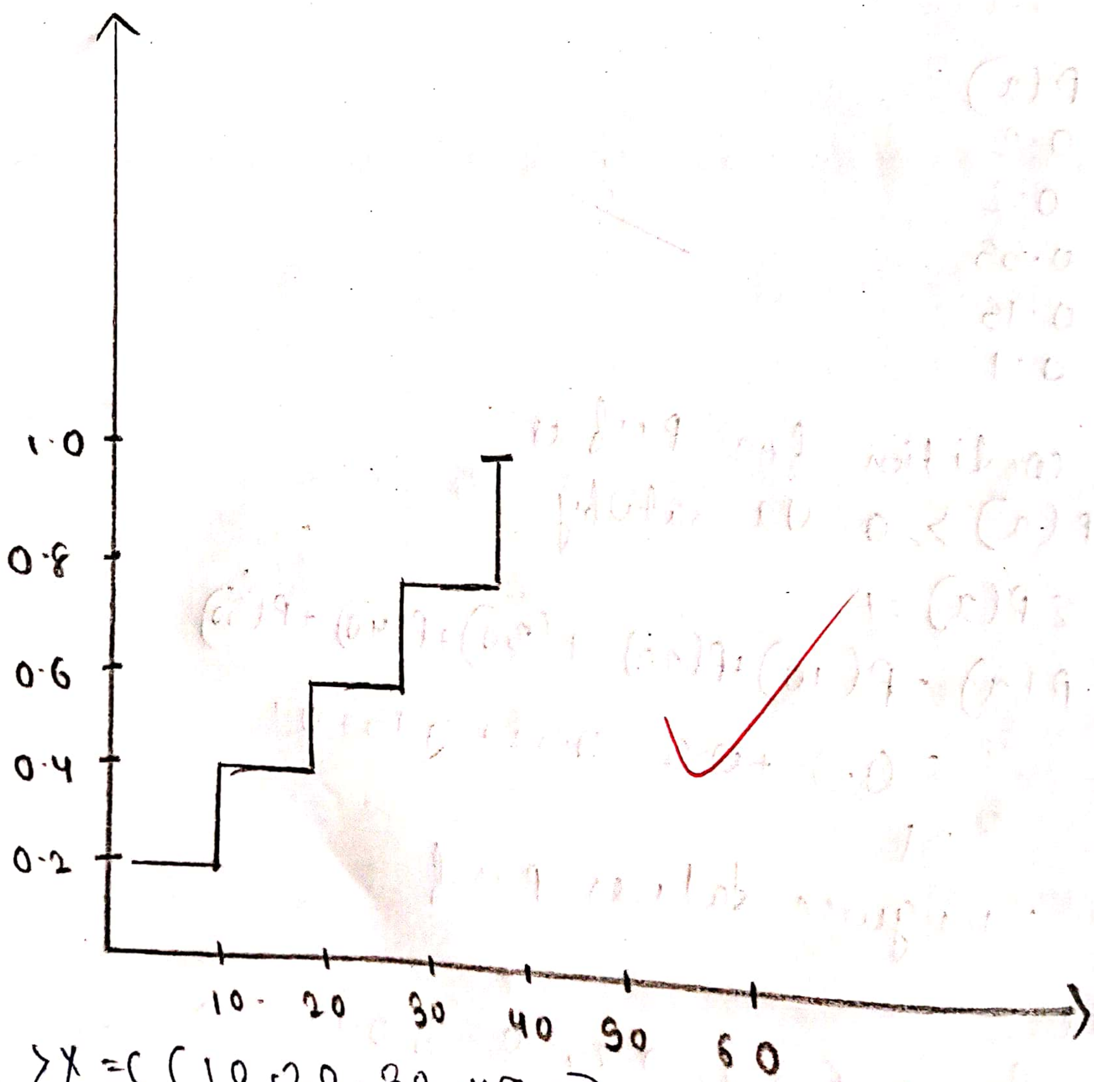
Code

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

Q2) Find the 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

$$F(x) = \begin{cases} 0 & x < 10 \\ 0.2 & 10 \leq x < 20 \\ 0.4 & 20 \leq x < 30 \\ 0.75 & 30 \leq x < 40 \\ 0.90 & 40 \leq x < 50 \\ 1.0 & x \geq 50 \end{cases}$$




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

Q2) Find

x	1	2	3	4	5	6
$P(x)$	0.5	0.25	0.1	0.2	0.2	0.1

$f(x) = 0 \qquad x < 1$

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

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

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

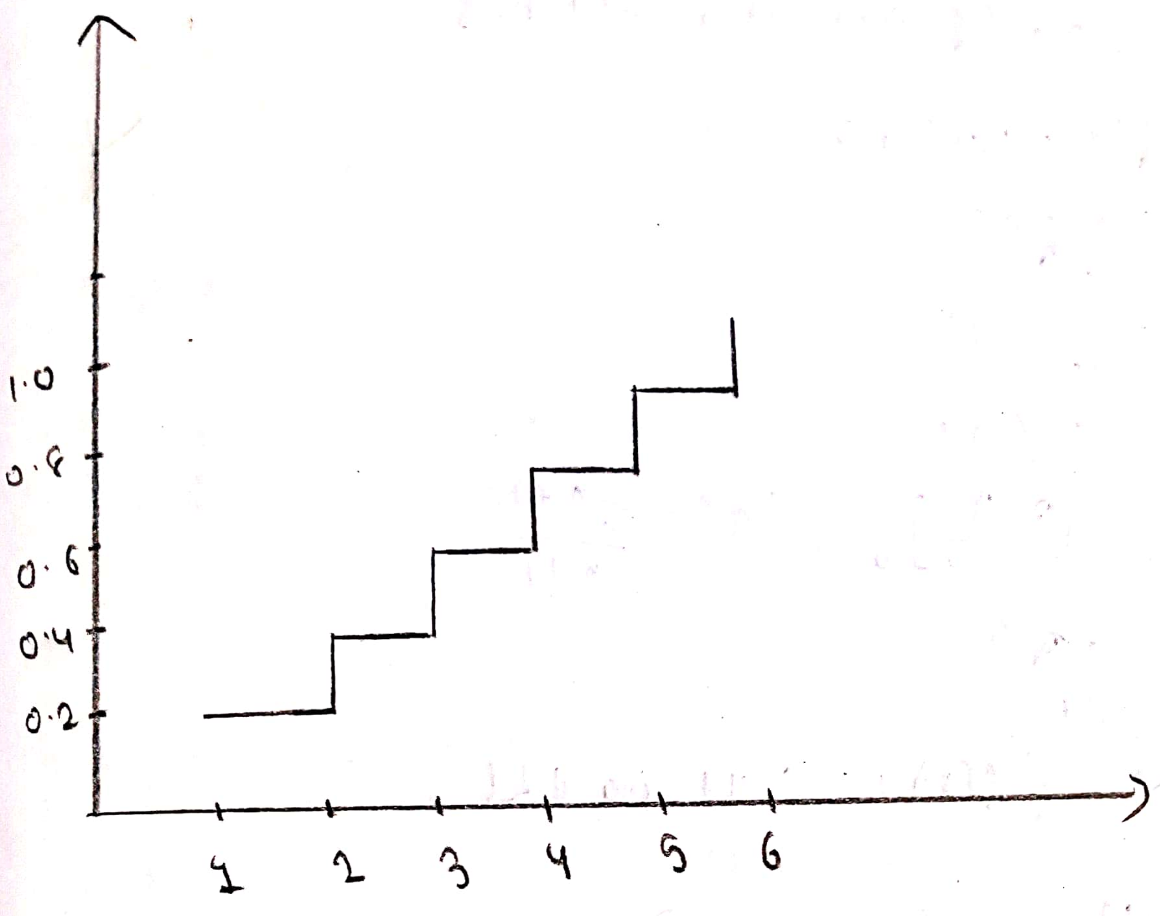
$= 0.70 \qquad 4 \leq x < 5$

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

$= 1.00 \qquad x \geq 6$

```
> 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, 0.10 , 0.90 , 1.00
> x = c (1,2 ,3,4, 5,6)
> plot (x, cumsum ( prob) , "S" xlab = " value"
    ylab = " cumulative prob ability
    main = " CDF graph , " col = " brown")
```

CDF Graph



3) Check that whether the following is pdf or not

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

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

1) $f(x) = 3 - 2x$

$= \int_0^1 f(x)$

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

$= \int_0^0 3\, dx = \int_0^1 2x\, dx$

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

∴ The $\int f(x) = 1$ ∴ It is not a pdf

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

$= \left[3 \frac{x^3}{3}\right]_0^1 \quad \because x^n = \frac{x^{n+1}}{n+1}$

$= x^3$

$= 1$

The $\int_0^1 f(x) = 1$ ∴ It is a pdf

1> x = dbinom (10, 100, 0.1)

> x

[1] 0.1318653

2> i) dbinom (4, 12, 0.2)

[1] 0.1328756

(ii) pbinom (4, 12, 0.2)

[1] 0.4274445

(iii) 1-pbinom (5, 12, 0.2)

[1] 0.01940528

Practical 3

TOPIC :- Binomial distribution

$P(x=x)$ = dbinom (x, n, p)
$P(x \leq x)$ = pbinom (x, n, p)
$P(X > x)$ = 1 - pbinom (x, n, p)
If x is unknown
$P_1 = P(X \leq x)$ = qbinom (P_1, n, p)

1) Find the probablity of exactly 10 success in hundred trials with $P = 0.4$

2) Suppose there are 12 mcq, Each question has 5 options out of which 1 is correct Find the probablity of having exactly 4 correct answers
ii) Atmost 4 correct answers.
iii) more than 5 correct answers.

3) Find the complete distribution when $n=5$ and $P=0.1$

4) $n=12$, $P = 0.25$ find the following probabities
i) $P(X=5)$
ii) $P(X \leq 5)$

5) The probability of a Salesman making a sale to customer 0.15 . Find the prob. of

1) No Sales out of 10 customer
2) More than 3 Sales out of 20 customer.

6) A Salesman has 20% probability out making a sale to customer out of 30 custo what minimum number of Sales he can make with 88% of probability

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

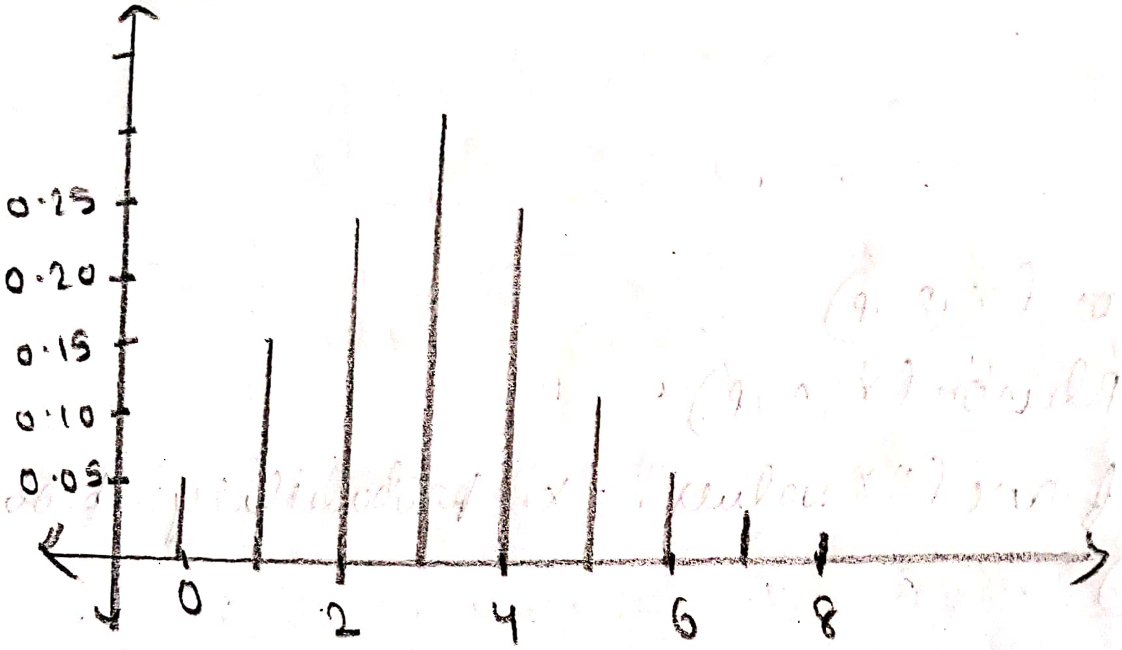
5) =binom (0, 10, 0.15)
(i) 0.196874
> 1 - binom (3, 20, 0.15)

i) qbinom (0.88, 830, 0.7)
[1] 9

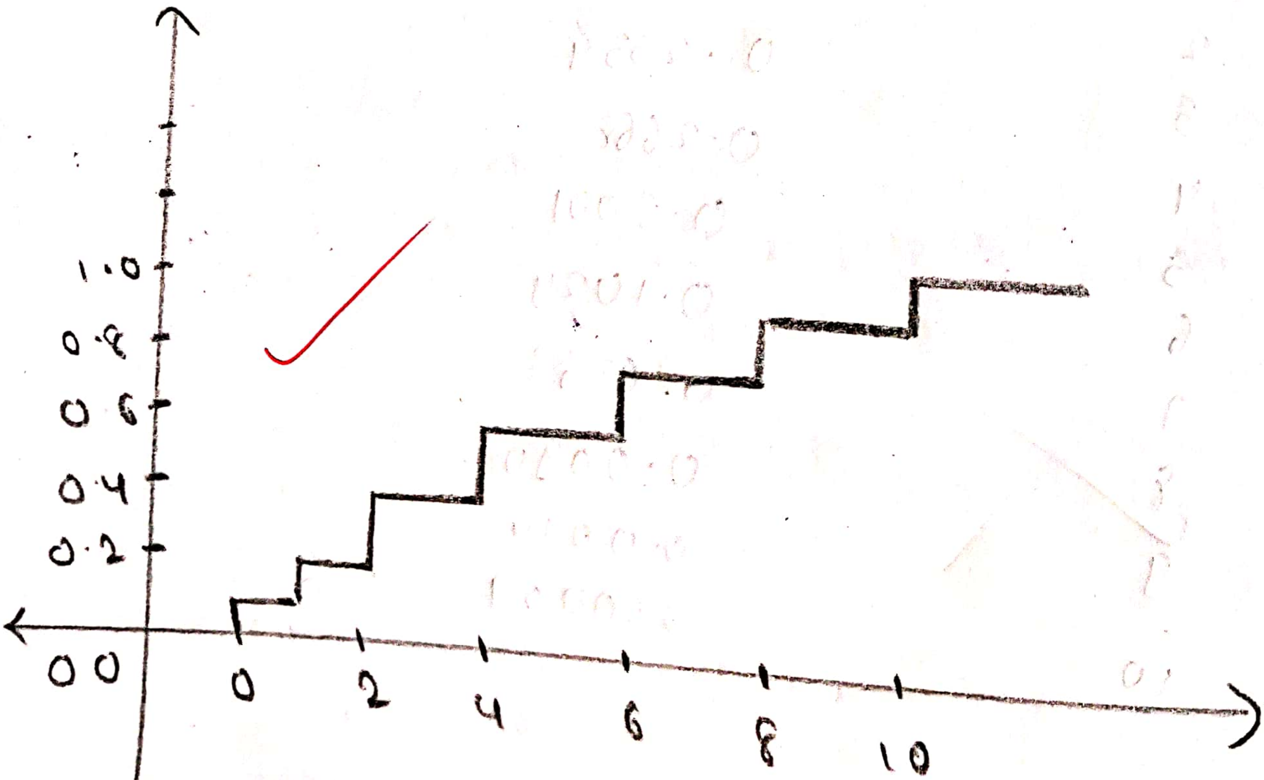
ii)
```
> n=10
> p=0.3
> x=0:n
> prob=dbinom(x,n,p)
> cumprob = pbinom(x,n,p)
> d=data.frame("X values"=x,"probability"=prob
> print(d)
```

	X values	probability
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	

> plot (x, prob, "h")



> plot (x, cumprob, "s")



i) To generate random numbers from a normal distribution (n random variables) the R code is

∴ rnorm (n, μ, σ)

A random variable X follows normal distribution with mean = μ = 12 and S.D = σ = 3. Find

i) P (X ≤ 15)

ii) P(10 ≤ x ≤ 13)

iii) P (X > 14)

CODE:-

```
> P1 = pnorm (15, 12, 3)
> P1
[1] 0.8413447
> cat("P(x <= 15) = ", P1)
```

```
[1] 0.2524925
> cat ("P(X>14) = ", P3)
P(X>14) = 0.2524925
> P4 = rnorm (5,12,3)
> P4
[1] 15.254725 16.548505 11.280515,6.4193
12.2
```

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 $P(X < k) = 0.4$

CODE

```
> a1 = pnorm (7,10,2)
> a1
[1] = 0.668072
> a2 = pnorm (5,10,2) - pnorm (12,10,2)
> a2
[1] = 0.8351351
> a3 = 1 - pnorm (12,10,2)
> a3
[1] 0.1586553
> a4 = rnorm (10,10,2)
> a4
[1] 11.608931   9.928417    12.637741   8.0
    8.721380    9.193726    9.366824    11.70
    9.537584    10.715006
> a5 = qnorm (0.4,10,2)
> a5
[1] 9.493306
```

Q5) Generate 5 random number from a normal distribution $\mu = 15, \sigma = 4$. Find sample mean, median S.D and print it

CODE:

```
> rnorm (15, 15, 4)
[1] 10.7649   7.793249   9.959444   13.349904
    17.509668

> am = mean (x)
> am
[1] 11.87345
> cat ("sample mean is =", am)
sample mean is = 11.87345
> me = median (x)
> me
[1] 11.09965
> SD = sqrt (v)
> SD
[1] 3.33163
> cat ("S.D is =", SD)
S.D is = 3.331613
```

29

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

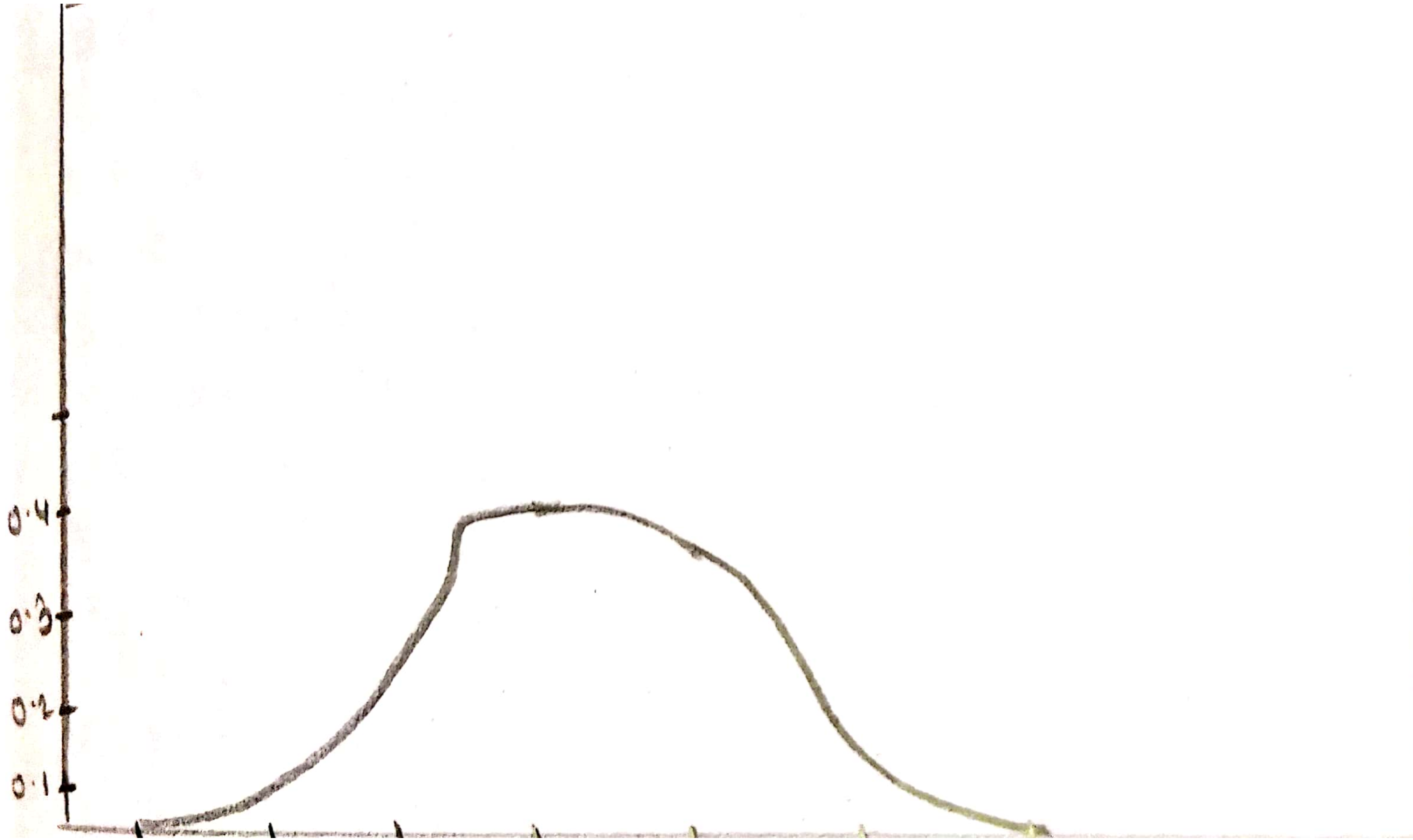
1) $P(x \leq 40)$
2) $P(x > 35)$
3) $P(25 < x < 35)$
4) Find k such that $P(X < k) = 0.6$

```
> F1 = pnorm (40,30,10)
> F1
[1] 0.8413447
> f2 = 1 - pnorm (35,30,10)
> f2
[1] 0.3085375
> f3 = pnorm (25,30,10) - pnorm (35,30,10)
> f3
[1] - 0.3829249
> f4 = qnorm (0.6 ,30,10)
> f4.
[1] 32.53347
```

Q5) Plot the standard normal graph

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

6


0·4
0·3
0·2
0·1

Practical - 5

Topic - Normal and t- test

$H_0 : \mu = 15$ $H_1 : \mu \neq 15$

1) Test the hypothesis

Random sample of size 400 is drawn and it is calculated the sample mean is 14 and S.Dev 3. test the hypothesis at 5% level of significance.

\# 0.05 > Accept the value

\# 0.05 & less then Reject

```
> mo = 15
> mx = 14
> sd = 3
> n = 400
> z cal = (mx - mo) / (sd (csqrt (n))
> z cal
[1] -6.66667
> cat (" calculated value of z is =", z cal)
calculated value of z is = -6.666667
> pvalue = 2 * ( 1- pnorm (abs (z (a))))
> pvalue
[1] 2.616796 e-11
```

∴ The value is less than 0.05 we will reject the value of $H_0 : \mu = 15$

Test the hypothesis H_0 $\mu=10$ against $H_1: \mu \neq 10$
A random sample size of 400 is drawn with sample mean $=10.2$ and $S.D = 2.25$. Test the hypothesis at

```
> m0 = 10
> n = 400
> mx = 10.2
> sd = 2.25
> zcal = (mx - m0) / (sd/(sqrt (n)))
> zcal
[1] 1.77778
> pvalue = 2 * (1 - pnorm(abs(zcal)))
> pvalue
[1] 0.07544036
```

∴ The value p value is greater than 0.05
∴ The value is accepted

Test the hypothesis H0: proportional of smok in college is 0.2
A sample is collected and calculated the sam proportional as 0.125. Test of hypothesis at 5% less of significance (sample size is 900)

```
> P = 0.2
> p = 0.125
> n = 400
> Q = 1 - P
```

```
> zcal = (p-P) /(sqrt(P*Q/n))
> cat ("calculated value of z is - ", zcal)
[1] calculated value of z is= -3.75
> pvalue = 2* (1- pnorm (abs (zcal)))
> pvalue
[1] 0.0001768346 (Reject)
```

1) Last year farmer's lost 20% of their crops. A random sample of 60 fields are collected and it is found that 9 fields crops are insect polluted. Test the hypothesis at 1% level of significance.

```
> P=0.2
> p=9/60
> n=60
> zcal =((p-P)/(sqrt(P*Q/n))
> zcal
[1] -0.9682458
> pvalue = 2* (1- pnorm(abs(zcal)))
> pvalue
[1] 0.3329216
```

∴ The value is 0.1 so values is accepted.

2) Test the hypothesis $H_0: \mu = 12.5$ from the following sample at 5% level of significance.

```
> x=c(12.75, 1197, 12.15, 12.08, 12.31, 12.28, 11.94, 11.89, 12.16,
    12.04)
> n=length(x)
> n
[1] 10
```

```
> mx = mean(x)
> mx
[1] 12.107
> variance = (n-1) * var(x)/n
> variance
[1] 0.019521
> sd = sqrt(variance)
> sd
[1] 0.1397176
> m0 = 12.5
> t = (mx - m0)/(sd/sqrt(n))
> t
[1] -8.894909
> pvalue = 2 * (1- pnorm(abs(t)))
> pvalue
[1] 0
```

∴ The value is less than 0.05 the value is accepted

Practical- 6

AIM : Large Sample Test

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

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

) Solution :

```
> m0 = 250
> mx = 275
> sd = 30
> n = 100
> zcal = (mx - m0)/(sd/(sqrt(n)))
> cat ("calculated value of z is = ", zcal)
[1] calculated value of z is = 8.33333
> pvalue = 2 * (1-pnorm (abs (zcal)))
> pvalue
[1] 0
```

∴ The value is less than 0.05 we will reject the value of $\mu_0 = \mu = 250$

2) Sol^n:

```
> P = 0.8
> Q = 1-P
> p = 750/1000
> n = 1000
> zcal (p-P)/(sqrt(P*Q/n))
> cat("calculated value of z is :", zcal)
[1] calculated value of z is : -3.952847
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 7.72268e-05
```

∴ The value is less 0.01 & reject.

3) To random sample of size 1000 & 2000 are drawn from two population with same SD 2.5. The sample means are 67.5 and 68. Test the hypothesis $\mu_1 = \mu_2$ at 5% OF significance.

4) A study of noise level in 2 hospital is given below test the claim that 2 hospi[tal] have same level of noise at 1% level significance

HOS A	HOS B
84	34
61.2	59.4
79	7.5

In a Sample of 600 Students in clg 400 used Blue ink. In another clg from a Sample of 900 students 450 use blue ink. Test the hypothesis that the proportion of students using blue ink in two colleges are equal or not at 1% Level of significance.

> 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

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

> pvalue

[1] 2.417564e-07 ∴ (Rejected)

n1 = 84

n2 = 34

> mx1 = 61.2

> mx2 = 59.4

> sd1 = 7.9

> sd2 = 7.5

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

> zcal

[1] 1.62528

Scanned with CamScanner

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

[1] 0.2490211

∴ The value is greater than 0.01 we accept H_0 value.

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

> n1 = 600
> n2 = 400
> p1 = 400/600
> p2 = 450/900
> p = (n1 * p1 + n2 * p2) (n1 + n2)
> p
[1] 0.566667
> q = 1-p
> q
[1] 0.433333
> zcal = (p1 - p2) / sqrt (p * q * (1/n1 + 1/n2))
> zcal
[1] 6.381534
> pvalue = 2 * (1-pnorm (abs (zcal)))
> pvalue
[1] 1.753222 e -10

∴ value is less than 0.01 the value is rejected

TOPIC :- Small Sample test

The marks of 10 students are given by 63, 63, 66, 67, 68, 69, 70, 71, 72 test the hypothesis that the sample comes from the population with average 66

$H_0 : \mu = 66$

>x = c (66 , 63, 66, 67, 68, 69, 70, 71)

>t.test (x)

one Sample test

data : X

t = 68.319 , df = 9, pvalue = 1.558e-13

alternative hypothesis

True mean is not equal to 0

95 percent confidence interval

65.65171 70.14829

Sample estimates

mean of x

67.9

The pvalue is less than 0.05 we reject the hypothesis at 5 % level of significance.

2) Two groups of students scored the following marks. Test the hypothesis that there is no significance difference between the 2 groups

GR1 - 18, 22, 21, 17, 20, 17, 23, 20, 22, 21

GR2 :- 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

H_0 : There is not difference b/w the 2 groups

```
> 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 & 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   5.0317795

Sample estimates

mean of x   mean of y

20.1   17.9

```
> P - value = 0.03798
> if (p value > 0.05) {cat ("accept H0")}
else {cat ("reject H0")}
```

reject H_0.

The sales data of 6 shops before & after a special campaign are given below

Before : 53, 28, 31, 48, 50, 42

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

Test the hypothesis that the campaign is effective or not.

Ho : There is no significance difference of sales before & after campaign.

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

paired t test

data : x & y

t = -2.7815 df = 5, pvalue = 0.9806

alternative hypothesis

True difference in means is greater than 0

95 percent confidence interval

-6.035547 Inf

sample estimates

mean of the difference

-3.5

∴ p value is greater than 0.05, we accept the hypothesis at 5% level of significance.

4) Following are the weights before & after the diet program. Is the diet program effective.

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

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

Sol^n :- H_0: There is no significance difference

> x = c (Before)

> y = c (after)

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

paired t-test

data : x & y

t = 4.3458, df = 5, pvalue = 0.9963

alternative hypothesis : true difference in means is less than 0.

95 percent confidence interval

-inf 29.0295

Sample estimates

mean of the difference

19.83333

∴ p value is greater than 0.05 we accept the hypothesis at 5% level of significance.

9

Practical No 8

Topic :- Large and Small Test

1) H0 : μ = 55, H1 = μ ≠ 55
> n = 100
> mx = 52
> m0 = 55
> sd = 7
> zcal = (mx - m0) ((sd / (sqrt (n)))
> zcal
[1] -4.285714
> pvalues = 2 * (1 - pnorm (abs (zcal)))
> p value
[2] 1.82153e-05

As pvalue is less than 0.05 we reject H0 at 5% level of significance.

2) H0 : P = 0.5 against H1 : P ≠ 0.5
> P = 0.5
> q = 1 - P
> n = 700
> zcal = (p - P) / (sqrt (P* q / n))
> zcal
[1] 0
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
Scanned with CamScanner

ea

As pvalue is greater than 0.05 we accept H₀ at 1% level of significance.

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

Sol 3) > $H_1 = 1000$

> $n_2 = 1500$

> $p_1 = 2/1000$

> $p_2 = 1/1500$

> $p = (n1 * p1 + n2 * p2) / (n1 + n2)$

> p

[1] 0.0012

> q = 1 - p

[1] 0.9988

> zcal = (p1 - p2) / sqrt(p * q * (1/n1 + 1/n2))

> zcal

[1] 0.9433752

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

> pvalue

[1] 0.345489

∴ pvalue is greater than 0.05 we accept H₀ and 5% level of significance

Sol

$H_0: \mu = 100$ against $H_1: \mu \neq 100$

> var = 64

> n = 400

> mo = 100

> mx = 90

> sd = sqrt(var)

[1] 8

>

> zcal

zcal (m1-m0)/(sd/sqrt(n))

> zcal

[1] 2.5

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

> pvalue

[1] 0.01241933

Since p value is less than 0.05 we reject Ho at 5% level of significance

5) Ho : $\mu = 66$ against H_1 $\mu \neq 66$

> x = c(63, 63, 66, 69, 71, 71, 72)

> t.test(x)

one sample test

data: x

t = 47.94, df = 6, p-value 5.922e-09

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval

69.66479 71.62092

sample estimates:

mean of x

68.14286

Since p value is less than 0.05 we reject Ho at

Sol^n ⑥ $H_0 \cdot \sigma_1 = \sigma_2$ against $H_1 : \sigma_1 \neq \sigma_2$
07

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

> y.c(64,66,74,78,82,85,87,92,93,95,...

> var.test(x,y)

F test to compare two variances

data : x and y

F: 0.78803 num df · 7, denom df -10 pvalue 0.7...

alternative hypothesis true ratio of variance is not equal to 1

95 percent confidence interval

0.199509 3.751881

Sample estimates

ratio of variance

0.7880285

∴ p-value is greater than 0.05 we accept H_0 at 5% level of significance

7 ⇒ $H_0 : \mu = 1150$ against $H_1 : \mu \neq 1150$

> n=100

> mx= 1150

> m0 = 1200

> zcal = (mx-m0) / (sd / sqrt(n))

> zcal

[1] -4

> pvalue · 2* (1-pnorm(abs (zcal))

p value is less than 0.05 we reject H_0

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

```
> n1=200
> n2=300
> p1 =44/200
> p2 = 56/300
> p =(n1* p1 + n2* p2)/(n1 +n2)
> p
[1] 0.2
> q =1 - p
[1] 0.8
> zcal = (p1 - p2)/(sqrt(p* q* (1/n1 +1/n2)))
> zcal
[1] 0.9128709
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 0.3613104
```

∴ p value is greater than 0.5 we accept H_0 at 1% level of significance

Practical - 9

Topic :- Non-parametric testing of Hypothesis using R-environment

1) The following data represent earings (in dollars) for a random sample of five common stocks listed on the New York Stock exchange. Test whether median earnings is 4 dollars

Data :- 168, 3.35, 2.90, 6.25, 3.24

```
> x <- c( 1.68, 3.35, 2.90, 6.23, 3.24),
> n <- length (x);
> n
> [1] 5
> x > 4;
[1] FALSE FALSE TRUE TRUE
> s <- sum (x>4) ; s;
[1] 1
> binom. Test (s, n, P = 0.5, alternative = "greater");
```

Exact binomial test

data : s and n

number of success - 1, number of trials = 5, p-values

alternative hypothesis : true probability of success is greater than 0.5

95 percent confidence interval

0.01020622   1.0000000

Sample estimates

probability of success

0.2

The Scores of 8 Students in reading before an
after lesson are as follows
Test whether there is effect of reading

Student No	1	2	3	4	5	6	7
Score before	10	15	16	12	09	07	11
Score After	13	16	15	13	09	10	13

CODE:-

```
> b <- c (10, 15, 16, 12, 09, 07, 11, 12);
> a <- c (13, 16, 15, 13, 08, 10, 13, 10);
> D <- b - a ;
> wilcox-test (D, alternative = "greater");
```

Wilcoxon signed rank test with
continuity correction data: D
V = 10.5, p - value = 0.8722
alternative hypothesis: true location is gree
than 0

warning message
In wilcox-test. default (D, alternative = "great
cannot complete exact p - value with ties

∴ p - value is greater than 0.05 we accept it

The diameter of a ball bearing was measured by 6 inspectors each using two different kinds of calipers. The results were. Test whether average ball bearing for

71

Inspector	1	2	3	4	5	6
Caliper 1	0.265	0.268	0.266	0.267	0.269	0.264
Caliper 2	0.263	0.262	0.270	0.261	0.271	0.260

Caliper 1 and caliper 2 are same

CODE:-

```
> x<-c(0.265, 0.268, 0.266, 0.267, 0.269, 0.264)
> y<-c(0.263, 0.262, 0.270, 0.261, 0.271, 0.260);
> wilcox.test (x, y, alternative = "greater")
```

wilcoxon rank sum test

data: X and Y

W = 24, p = 0.197

alternative hypothesis: true location shift is greater than 0

∴ p-value is greater than 0.05 we accept it.

9

Practical-10

Aim :- Chi Square tests & ANOVA (Analysis of variances)

1) Use the following data to test whether the condition of home & condition of child are independent are not

~~cond~~ child	cond Home	
	Clean	Dirty
Clean	70	50
Fairy 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

∵ p value is less than 0.05 we reject the hypothesis at 5% level of significance.

Q) Test the hypothesis that vacation & disease are independent or Not

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

H_0: Disease & vaccine are independent

```
> x = c(70, 39, 46, 37)
> m = 2
> n = 2
> y = matrix(x, nrow = m, ncol = n)
> y
       [,1]   [,2]
[1,]    70     46
[2,]    39     37

> pv = chisq.test(y)
> pv
```

Pearson's chi squared test with Yates' continuity correction

data : Y

X-Square = 2.0275

df = 1

p- value = 0.1549

∴ p- value is more than 0.05 we accept the hypothesis at 5% level of significance

// They are INDEPENDENT

25.

Q3) Perform a ANOVA for the following data

TYPE	OBSERVATION
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] "value" "ind"
oneway.test(value ~ ind, data = d, var.equal = T)
```

One-way analysis of means

data: values and ind.

F = 11.735 df = 3, denom df = 9,

P value = 0.00183

P-value is less than 0.05 we reject the hypothesis.

>anova = anov (values ~ wid, data = d)
>summary (anova)

	Df	Sum	mean Sq	F value	Pr(>F)
wid	3	71.06	23.688	11.73	0.00185 **
Residuals	9	18.17	2.019		

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