INDEX

No.	Title	Page No.	Date	Staff Member's Signature
	Sem. - II			
1.	Basic of R Software	45	28/11/19	A-r
2.	Probability Distribution	48	5/12/19	AY
3.	Binomial Distribution.	50.	19/12/19	AM
4.	Normal Distribution	52	2/1/20	AY
5.	Normal And t - Test	53	17/1/20.	AY
6	Large Sample Test	55	23/1/20.	AY
7.	Small Sample Test	57	6/2/20	AY
8.	Large and Small Sample Test	59	13/2/20	AY
9	Chi - Square Test And ANOVA	61	20/2/20	AY 27.2
10.	Non - Parametric Test	64.	21/2/20	

Scanned with CamScanner

PRACTICAL - I

Basics of R Software

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

Q1. Solve the following :-

1] $4 + 6 + 8 \div 2 - 5$

```
> 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
```

4] $\sqrt{4^2 + 5 \times 3 + 7/6}$

```
> sqrt(4^2 + 5*3 + 7/6)
[1] 5671567
```

5] Round off

46 ÷ 7 + 9 x 8

> round (46 / 7 + 9*8)

[1] 79

Q.II. Solve the following:-

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

[1] 4, 6, 10, 14

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

[1] 4, 9, 10, 21

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

[1] 4, 9, 30, 14

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

[1] 4 9 30 14

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

[1] -2 -18 -8 -3

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

[1] 4 9 25 49

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

[1] 4 36 512 9 16 25

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

[1] 1.50 0.40 1.75 1.00

Q.III. Solve the following:-

1] > x = 20

> y = 30

> z = 2

i] > x^2 + y^3 + z

[1] 27402

>

2] $\sqrt{x^2+y}$

> sqrt (x^2 +y)

[1] 20.73644

3] $x^2 + y^2$

> x^z + y^z

[1] 1300.

Q.IV Solve the following :-

1]

1	5
2	6
3	7
4	8

> x <- matrix (nrow = 4, ncol = 2, data=c(1:8))

> x

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

Q. V (ii) find x+y and 2x+3y such that

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

```
> 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,
> 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. VI :- Points of Marks of Computer Science Students are as follows :-

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

```
> x = c (given data)
> length (x)
[1] 20
> breaks = seq(20, 60, 5)
> a = cut(x, breaks, right = False)
> b = table(a)
> c = transform(b)
> c
        x       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
```

A.V. 5/12/19

PRACTICAL - II

Probability Distribution

Q. Check whether the following functions are p.m.f or not.

i]

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

∵ $P(2) = -0.5$

∴ It cannot be a p.m.f.

∵ In P.M.f, $P(x) > 0 \quad \forall x \in I'$

∴ It cannot be a p.m.f as in p.m.f $\Sigma P(x) = 1$.

ii]

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

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

∵ $\Sigma P(x) \neq 1$

∴ It is not a p.m.f.

iii]

x	1	2	3	4	5
P(x)	0.2	0.2	0.35	0.15	0.1

```
> p = c(0.2, 0.2, 0.35, 0.15, 0.1)
> sum(p)
[1] 1.
```

∴ It satisfies both the condition. it is a p.m.f.

P48

Q.II. 1] Find the c.d.f for the following p.m.f and states the graph.

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

→ F(x) = 0 $x < 10$

= 0.2 $10 \leq x < 20$

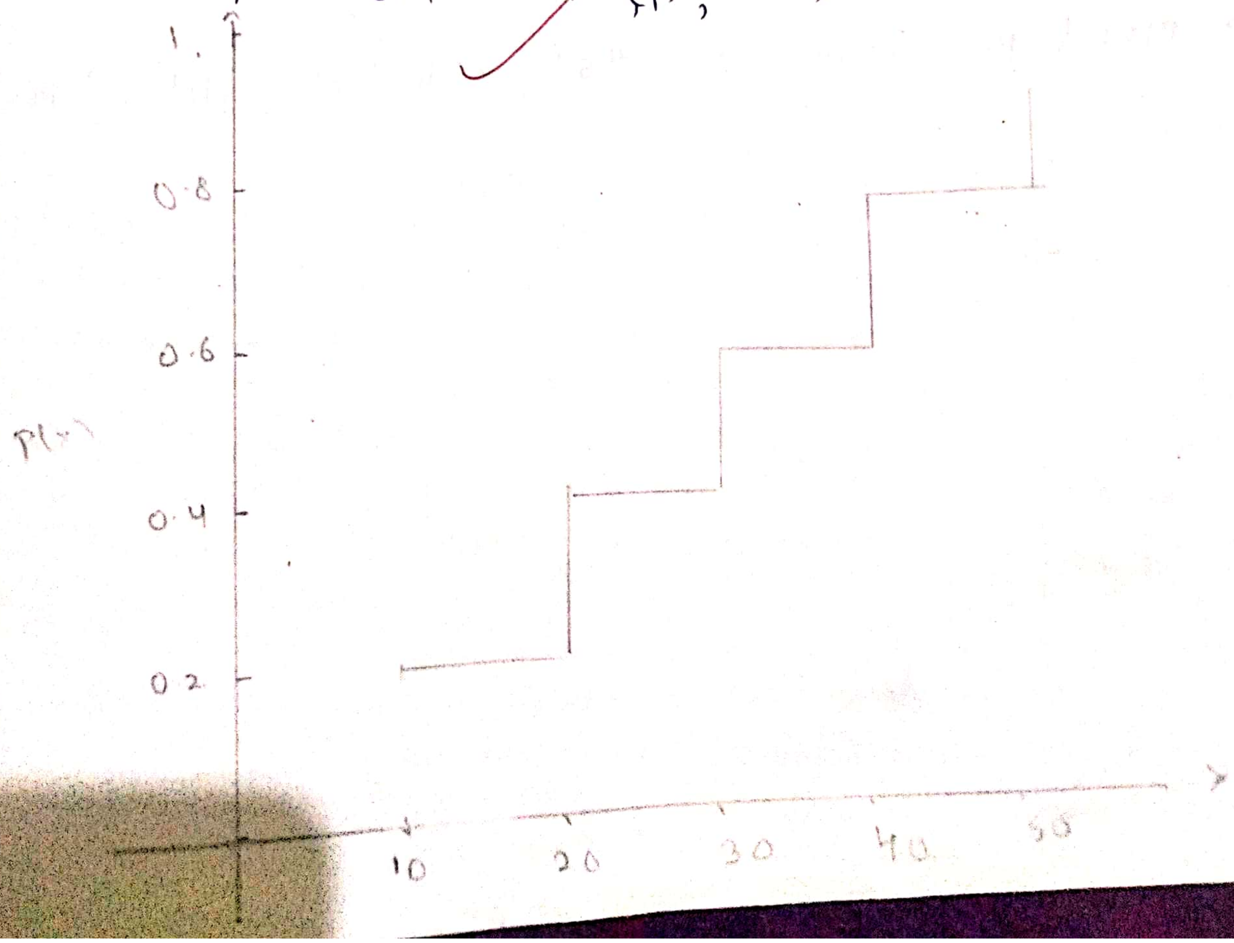
= 0.4 $20 \leq x < 30$

= 0.75 $30 \leq x < 40$

= 0.95 $40 \leq x < 50$

= 1.0 $x \geq 50$

```
> p = c(0.2, 0.2, 0.35, 0.15, 0.1)
> sum(p)
[1] 1
> cumsum(p)
[1] 0.20  0.40  0.75  0.90  1.00
> p1 = c(20, 30, 40, 50)
> plot(p1, cumsum(p), "s")
```



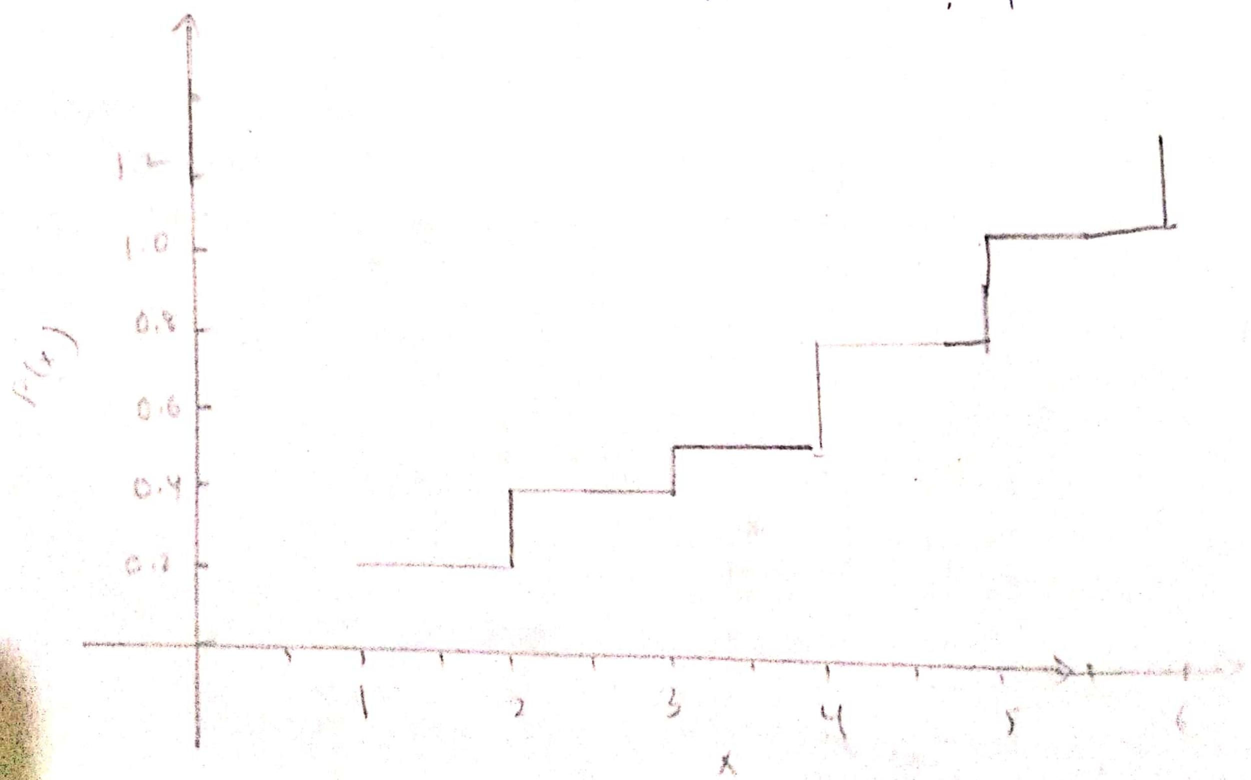
ii]

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

```
> p = c(0.15, 0.25, 0.1, 0.2, 0.2, 0.1)
> sum(p)
[1] 1
> cumsum(p)
[1] 0.15  0.40  0.50  0.70, 0.90, 100
> p1 = c(1,2,3,4,5,6)
```

$$
F(x) = \begin{cases} 0 & x < 1 \\ 0.15 & 1 \le x < 2 \\ 0.40 & 2 \le x < 3 \\ 0.50 & 3 \le x < 4 \\ 0.70 & 4 \le x < 5 \\ 0.90 & 5 \le x < 6 \\ 1.00 & x \ge 6. \end{cases}
$$

```
> plot(p1, cumsum(p), "s", xlab="x", ylab = "P(x)")
```



Scanned with CamScanner

Q. III. Check whether following is p.d.f or not :

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

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

i] $f(x) = 3 - 2x \qquad 0 \leq x \leq 1$

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

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

$= 3[x]_0^1 - \cancel{2}\dfrac{[x^2]_0^1}{\cancel{2}}$

$= 3(1) - 3(0) - (1)^2 - 1(0)$

$= 3 - 1$

$= 2$

ii] $f(x) = 3x^2 \qquad 0 < x < 1$

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

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

$= \cancel{x}\, (1^3) - (1^0)$

$= 1$

AM 5/12/19

Practical - III

Binomial Distribution

i] $P(X=x)$ = dbinom (x, n, p)
ii] $P(X \le x)$ = pbinom (x, n, p)
iii] $P(X > x)$ = 1- binom (x, n, p)
Where x is unknown
iv] $P_1 = P(X \le x)$ = qbinom (P_1, n, p)

Q.i] Find the probability of exactly 10 Success in 100 trials with p = 0.1

Q.ii] Suppose there are 12 mcqs questions each question has 5 Options out of which 1 is correct. Find the probability of having i] exactly 4 correct answers ii] atmost 4 correct answers iii] more than 5 correct answers.

Q.iii] Find the complete distribution when n = 5 and p = 0.1

Q.iv] n = 12, p = 0.25. Find the following probabilities.
i] $P(X=5)$ ii] $P(X \le 5)$ iii] $P(X>7)$
iv] $P(5 < x < 7)$

Q.v] The probability of a salesman making a sale to a customer is 0.15. Find the probability of i] No sales out of 10 customer ii] More than 3 sales out of 20 customers

Q.vi] A salesman has a 20% probability of making a sale to a Customer Out of 30 customer. What minum no. of sales he can make with 88% of Probability?

vii] x follows binomial distribution with $x = 10$, $p = 0.3$. Plot the graph of pmf and c.d.f.

050

Solution :-

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

2. i] > dbinom(4,12,1/5)
[1] 0.1328756

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

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

3] > dbinom(0:5, 5, 0.1)
[1] 0.59049 0.32805 0.07290 0.00810 0.00045 0.00001

4. i] > dbinom(5,12,0.25)
[1] 0.1032414

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

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

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

5] i > dbinom(0, 10, 0.25)
[1] 0.0009765625

ii > 1 - pbinom(3,20,0.5)
[1] 0.9987116.

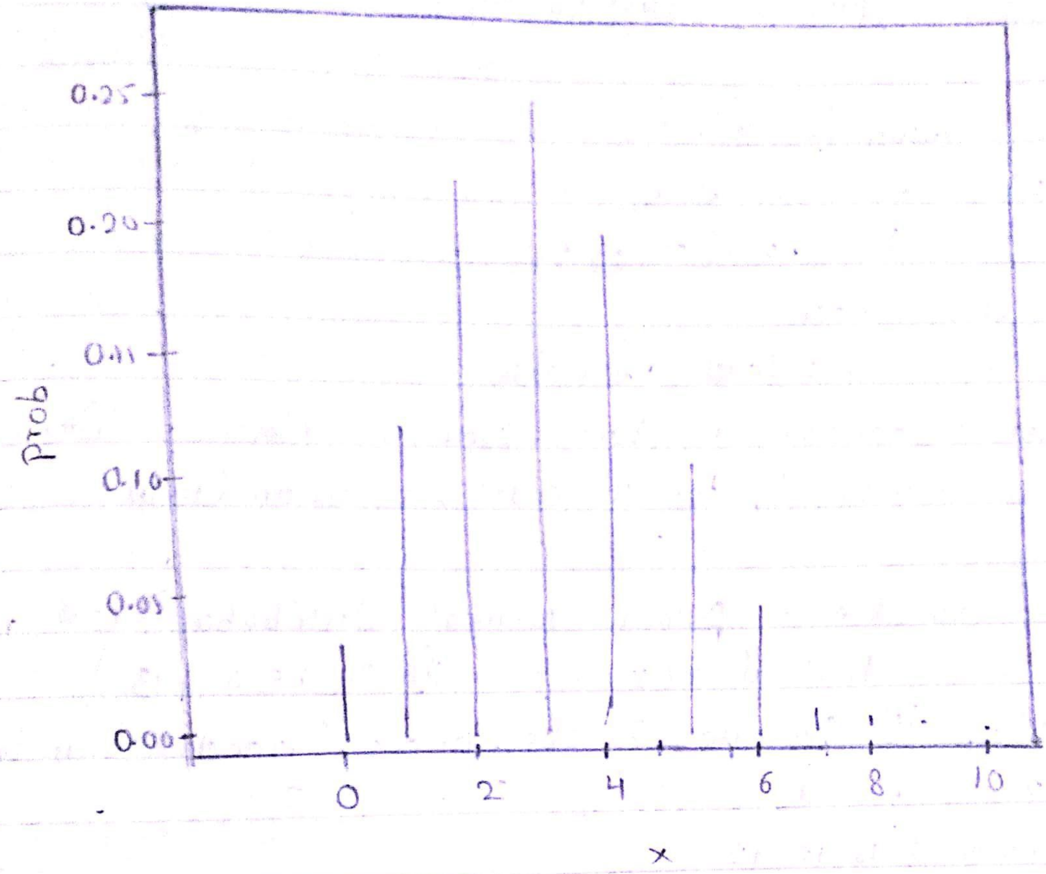
6] > qbinom(0.88, 30, 0.2)
[1] 9.

1] > n = 10
> p = 0.3
> x = 0:n
> x
[1] 0 1 2 3 4 5 6 7 8 9 10
> prob = dbinom (x,n,p)
> prob
[1] 0.282475249 0.121060821 0 0.2334744405 0.2668279320 0.2001209
[6] 0.1029193452 0.367569090 0.0090016920 0.0014467005 0.0001377810
[11] 0.0000059049
> cumprob = pbinom(x,n,p).
> cumprob.
[1] 0.02824752 0.14930835 0.38278279 0.2668279320 0.2001209490
[6] 0.95265101 0.98940792 0.99840961 0.99985631
[10] 0.99999410 1.00000000
> d = data. frame("x Value" = x , "Probability" = prob)
> d.

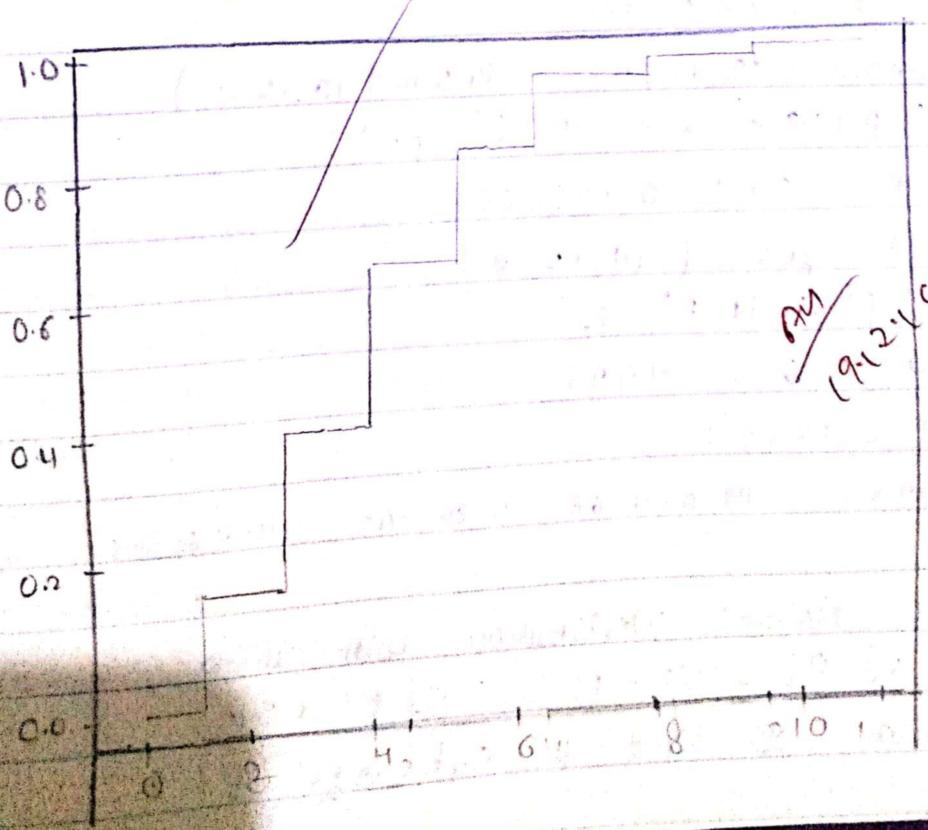
x. Value	Probability
0	0.0282475249
1	0.1210608210
2	0.2334744405
3	0.2668279320
4	0.2001209490
5	0.1029193452
6	0.0367569090
7	0.0090016920
8	0.0014467005
9	0.0001377810
10	0.0000059049.

> plot (x, prob, "h")

Scanned with CamScanner



> plot (x, cumprob, "s")



AY
19.2.19

PRACTICAL - IV

Normal Distribution

i] $P(X=x)$ = dnorm (x, μ, σ)
ii] $P(X \leq x)$ = pnorm(x, μ, σ)
iii] $P(X > x)$ = 1 - pnorm(x, μ, σ)

where μ = Mean
σ = Standard Deviation.

iv] To generate random numbers from a Normal Distribution (in random numbers), the R code is = rnorm(n, μ, σ)

Q.1] A random variable X follows Normal Distribution with mean $\mu = 12$, $\sigma = 3$. Find i] $P(X \leq 15)$ ii] $P(10 \leq x \leq 13)$ iii] $P(X > 14)$ iv] Generate 5 observations (random numbers)

→ $\mu = 12$, $\sigma = 3$.

```
> p1 = pnorm (15, 12, 13)
> cat ('p(x<=15):', p1)
P(x<=15) : 0.8413447
> p2 = pnorm (13, 12, 3) - pnorm (10, 12, 3)
> cat (' p (10 <= x <= 13):', p2)
P (10 <= x <= 13) : 0.3780661
> p3 = 1 - pnorm (14, 12, 3)
> cat ('p(x>14):', p3)
P(x>14) : 0.2524925
> rnorm (5, 12, 3)
[1] 9.007282   14.034168   11.837712   14.460544   11.125136
```

Q.2] X follows Normal Distibution, with $\mu = 10$, $\sigma = 2$. Find. i] $P(X \leq 7)$ ii] $P(5 \leq x \leq 12)$ iii] $P(X > 12)$ iv] Generate 10 observations. v] Find k such that $P(X \leq k) = 0.4$

```
> p1 = pnorm(7,10,2)
> cat('p(x <= 7):', p1)
  p(x <= 7): 0.0668072
> p2 = pnorm(12,10,2) - pnorm(5,12,2)
> cat('p(5 <= x <= 12):', p2)
  p(5 <= X < 12): 0.8351351
> p3 = 1 - pnorm(12,10,2)
> cat('p(x>12):', p3)
  P(x>12): 0.1586553.
> rnorm(10,10,2)
[1] 13.082059   13.997909   14.498445   9.950494   11.048673
[6] 13.554411   11.736256   13.776210   6.919454   11.062674
> qnorm(0.4,10,2)
[1] 9.493306.
```

Q.iii) Generate 5 random numbers from a Normal Distribution with $\mu = 15$ & $\sigma = 4$. Find Sample mean, median, S.D and print it.

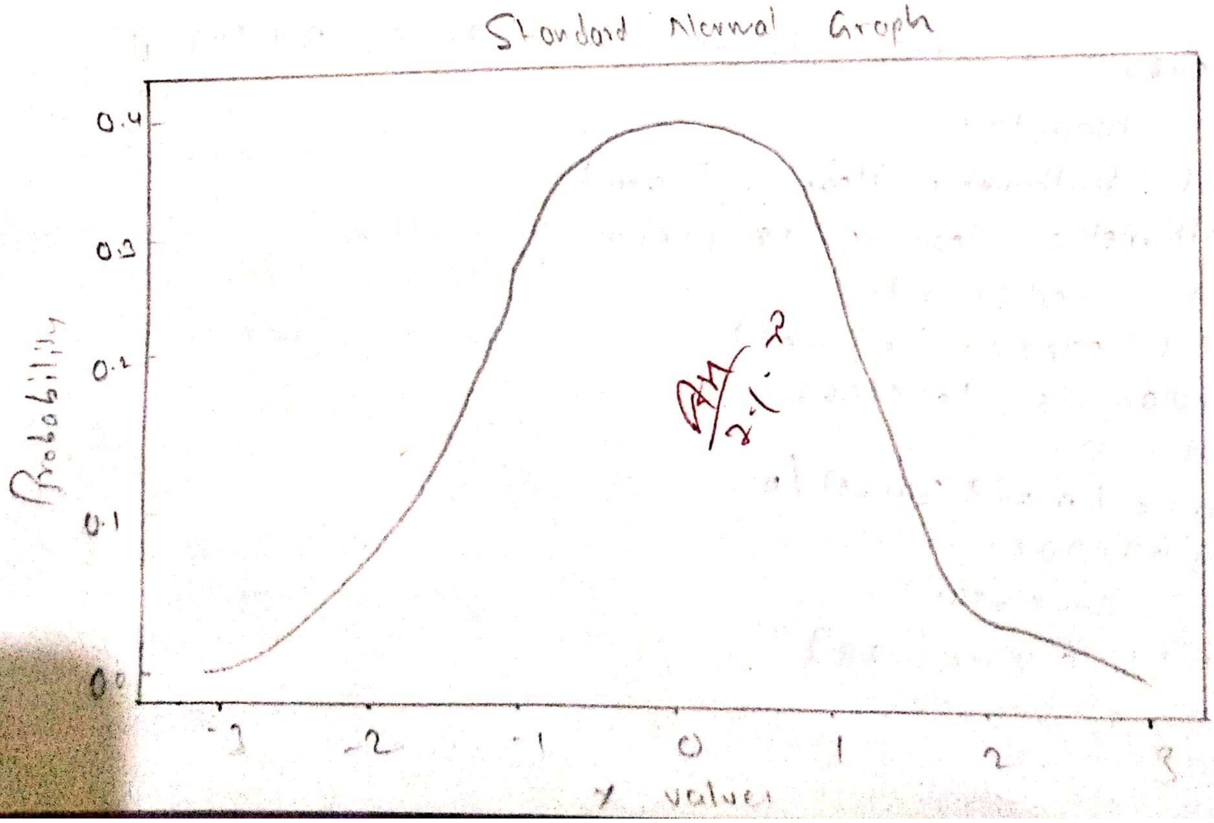
```
> x = rnorm(5,15,4)
> x
[1] 17.94827   15.54493   16.86062.   15.32839   24.94837
> am = mean(x)
> cat('Arithmetic Mean is', am)
  Arithmetic Mean is 18.12612
> med = median(x)
> cat('Median is', med)
  Median is 16.86062
> n = 5
> Var = (n-2) * var(x) / n
, [1] 9.399908
> sd = sqrt(var)
> cat(' S.D is', sd)
  S.D is 3.065927
```

Q.[iv] X Follows Normal Distribution, $X \sim N(30, 100)$, $\mu = 30$, $\sigma = 10$. Find i) $P(X \le 40)$ ii] $P(X > 35)$ iii] $P(5 < X < 35)$ iv] Find k such that $P(X < k) = 0.6$.

```
> p1 = pnorm(40,30,10)
> p1
[1] 0.8413447.
> p2 = 1 - pnorm (35,30,10)
> p2
[1] 0.3085375
> p3 = pnorm (35,30,10) - pnorm (25,30,10)
> p3
[1] 0.3829249
> qnorm(0.6,30,10)
[1] 32.53347.
```

Q.[v] Plot the Standard Normal Graph.

```
> x = seq(-3,3, by =0.1)
> y = dnorm(x)
> plot (" x,y, xlab =" x values", ylab = "Probability" , main =
  " Standard Normal Graph")
```



PRACTICAL - V

Normal And t-Test

1] Tes the Hypothesis $H_0 : \mu = 15$, $H_1 : \mu \neq 15$ where H_0 = Null Hypothesis and H_1 = Alternative Hypothesis. A random sample of Size 400 is drop and it is calculated, the 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 2 is :",zcal)
Calculated Value at 2 is : - 6.666667
> 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 $H_1 : \mu \neq 10$. A random sample of size 400 is drawn with sample mean 10.2 and s.d 2.25, test the Hypothesis at 5 % level of Significance

→
```
> mo: 10; mx = 10.2; sd = 2.25; n = 400
> zcal = (mx -mo) / (sd / (sqrt(n)))
> zcal
[1] 1.777778
```

> cat(" Calculated Value at 2 is : ", zcal)

Calculated Value at 2 is : 1.777778

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

> pvalue

[1] 0.07544036

∵ pvalue is more than 0.05, we accept $H_0: \mu = 10$

3] Test the Hypothesis (H_0), 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))

> zcal

[1] -3.75

> cat(" Calculated Value at 2 is ", zcal)

Calculated Value at 2 is -3.75

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

> pvalue

[1] 0.0001768346.

∵ pvalue is less than 0.05, we reject $H_0: \mu = 0.2$

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

Scanned with CamScanner

```
> 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] 0.3329216
```

∴ pvalue is more than 0.05, we accept H_0.

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

```
> x = c (12.25, 11.97, 12.15, 12.08, 12.31, 12.28, 11.94, 11.89, 12.16, 12.04)
> n = length (x)
> n
[1] 10
> mo = 12.5
> mo
[1] 12.5
> mx = mean (x)
> mx
[1] 12.107
> 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 H_0.

AK 16.1.2

1]
```
> 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, H0")
Since, Pvalue is less than 0.05, Hence, we reject H0
```

2]
```
> P = 0.8; Q = 1-P; p = 750/1000; n = 1000
> zcal = (p-P) / (sqrt(P*Q/n))
> cat("Calculated Value is", zcal )
Calculated Value is -3.952847.
> pval = 2* (1-pnorm(abs(zcal)))
> pval
[1] 7.72268e-05
> cat(" Since, Pvalue is less than 0.05, Hence, we reject H0 at 5% level")
Since, Pvalue is less than 0.05, Hence, we reject H0 at 5%
level of significance.
```

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] 2.417564e - 07
> cat (" Since, Pvalue is less than 0.05. Hence, we reject H0 at
5% level of significance")
> Since, Pvalue is less than 0.05. Hence, we reject H0 at
5% level of significance
```

Scanned with CamScanner

Large . Sample . Test

1] Let the population mean (amount spent per costomer in a restaurant) is 250. A sample of 100 cutomer is selected. The sample mean is calculated as 215 and s.d is 30. Test the hypothesis 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% level 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 links using 2 colleges are equal or not at 1% level of significance.

4] > n = 84; n2 = 34, mx1 = 61.2; mx2 = 59.4; sd1 = 7; sd2 = 7.5

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

> cat(" Calculated Value is ", zcal)

Calculate Value is 1.162528

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

> pval

[1] 0.2450211

> cat(" Since, Pvalue is greater than 0.05. Hence, we accept H₀ at 1 % level of significance")

> cat (Since, Pvalue is greater than 0.05. Hence, we accept H₀ at 1% level of significance.

5] $H_0: P_1 = P_2$ 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) / sqrt (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 H₀ at 1 % level of significance".)

Since, Pvalue is less than 0.05. Hence, we reject H₀ at 1 %. level of significance.

6) For Sample Size, n1 = 200, n2 = 200, p1 = 44/200, p2 = 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 H0
   at 5 % level of significance".)
Since, Pvalue is greater than 0.05. Hence, we accept H0 at
5 % level of significance.
```

AN
27-01-20

PRACTICAL - VII

Small Sample Test

Q1] The marks of 10 students are given by 63, 63, 66, 67, 68, 69, 70, 70, 71, 72. Test the hypothesis that sample comes from population with the average 66.

⇒ $H_0 : \mu = 66$

```
> x = c(63, 63, ... 71, 72)
> n = length(x)
> n
[1] 10
> 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 0
95 percent confidence interval: 65.63171   70.14829
sample estimates :
  mean of x
    67.9
```

∵ pvalue is less than 0.05. Hence, we reject the H_0 at 5 % level of significance.

```
> if (pvalue > 0.05) {cat("Accept H0")} else {cat("Reject H0")}
Reject H0
```

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

Group A : 18, 22, 21, 17, 20, 17, 23, 20, 22, 21

Group B : 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

⇒ H_0 : There is no difference between two 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 and y

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

alternative hypothesis : true difference in means is not equal to 0

0.1628205 5.0371795

sample estimates :

mean of x mean of y

20.1 17.5

> if (pvalue > 0.05) {cat("Accep. H.")} else {cat("Reject H.")}

Reject H_0.

Q.3] Sales data of 6 shop before and after are special campaign given below:-

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

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

Test the hypotheis the campaign is effective or not.

→ H_0 : 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 H0") } else { cat ("Reject H0") }
Accept H0.
```

Q.4] Two medicines are applied to 2 group of patient.
Group 1 : 10, 12, 13, 11, 14
Group 2 : 8, 9, 12, 14, 15, 10, 9
is there any significance difference between two groups.

Q.5] The following weights before and after a diet program
is diet program effective ?
Before : 120, 125, 115, 130, 123, 119
After : 100, 114, 95, 90, 115, 99

H_0 = There is no significace difference between medicines

```
4]. > a = c(10, 12, 13, 11, 14)
> b = c( 8, 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, p-value = 0.5273
alternative hypothesis : true difference in mean is not equal to 0
95 percent confidence interval :
-1.934382      3.534382
sample estimates :
mean of x   mean of y
11.8        11.0.
```

> pvalue = 0.5273
> if (pvalue > 0.05) { cat (" Accept H0") } else { cat ("Reject H0") }
Accept H0.

H0 : There is no significance difference between before and after
⇒ 5] > A = c(120, 125, 115, 130, 123, 119)
> B = c (100, 114, 95, 90, 115, 99)
> t.test (A, B, paired = T, alternative = "less")

Pared t-test

data: A and B.
t = 4.3418, df = 5, p-value = 0.9963
alternative hypothesis : true difference in mean is less than 0
95 - percent confidence interval :
- Inf 29.02295
sample estimates :
mean of the differences
19.83333

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

AX 6-2-20

PRACTICAL - VIII

Large and Small Sample Tests

#1] $H_0 : \mu = 55$. $H_1 : \mu \neq 55$

```
> n = 100
> sd = 7
> m0 = 55
> mx = 52
> zcal = (mx-m0) / (sd / (sqrt(n)))
> zcal
[1]  -4.285714
> pvalue = 2* (1-pnorm(abs(zcal)))
> pvalue
[1]  1.82153 e-05
```

∵ pvalue is less than 0.05

∴ We reject H_0.

#2] $H_0 : P = 0.5$ against $H_1 : P \neq 0.5$

```
> P = 0.5 ; Q = 1-P ; p = 350/700 ; n=100
> zcal = (p-P) / (sqrt (P*Q /n))
> zcal
[1] 0
> pvalue = 2* (1 - pnorm (abs (zcal)))
> pvalue
[1] 1
```

∵ pvalue is greater than 0.05

∴ We accept H_0

⇒3] $H_0 : P_1 = P_2$ against $H_1 : P_1 \neq P_2$

```
> ~~n = 400; mo = 99; mx = 100; var = 64~~
> n1 = 1000; n2 = 1500
> p1 = 2/1000; p2 = 1/1500
> p = (n1 * p1 + n2 * p2) / (n1+n2)
> p
[1] 0.0012
> q = 1-p
> q
[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 ~~less~~ greater than 0.05. Hence, we accept H_0.

⇒4] $H_0 : \mu = 100$ against $H_1 : \mu \neq 100$

```
> n = 400; mx = 99; mo = 100; var = 64
> sd = sqrt(var)
> zcal = (mx-mo) / (sd / (sqrt(n)))
> zcal
[1] -2.5
> pvalue = 2 * (1-pnorm(abs(zcal)))
> pvalue
[1] 0.01241933
```

∵ pvalue is less than 0.05. Hence, we reject H_0.

⇒ 5] $H_0 : \mu = 66$ against $H_1 : \mu \neq 66$ CGU

```
> 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.
```

∴ pvalue is less than 0.05. Hence, We reject H_0.

⇒ 6] $H_0 : \sigma_1 = \sigma_2$ against $H_1 : \sigma_1 \neq \sigma_2$

```
> 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 variances

data: x and y
F = 0.78803, num df = 7, denom df = 10, p-value = 0.7737
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
   0.199509   3.751881
sample estimates:
ratio of variances
   0.7880255
```

∴ pvalue is greater than 0.05. Hence, we accept H_0.

Scanned with CamScanner

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

```
> n = 100; mx = 1150; mo = 1200; sd = 125
> zcal = (mx-mo) /(sd / sqrt(n))
> zcal
[1] -4
> pvalue = 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 6.334248e-05
```

∵ pvalue is ~~greater~~ less than 0.05. Hence, we ~~accept~~ reject H_0

8] $H_0: p_1 = p_2$ 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
```

∴ pvalue is greater than 0.05. Hence, we accept H_0.

AN 13-2-21

FYCS Practical 8
Large and Small Sample tests

Q1 The arithmetic mean of a sample of 100 items from a large population is 52. If the standard is 7, test the hypothesis that the population mean is 55 against the alternative it is more than 55 at 5% LOS.

Q2 In a big city 350 out of 700 males are found to be smokers. Ddoes the information supports that exactly half of the males in the city are smokers ? Test at 1% LOS.

Q3 Thousand article from a factory:A are found to have 2% defectives ,1500 atricles from a 2nd factory:B are found to have 1% defective. Test at 5% LOS that the two factory are similar are not.

Q4. A sample of size 400 was drawn at a sample mean is 99. Test at 5% LOS that the sample comes from a population with mean 100 and variance 64.

Q5. The flower stems are selected and the heights are found to be(cm) 63,63,68,69,71,71,72 test the hypothesis that the mean height is 66 or not at 1% LOS.

Q.6. Two random samples were drawn from 2 normal populations and there values are A-66,67,75,76,82,84,88,90,92 B-64,66,74,78,82,85,87,92,93,95,97. Test whether the population have the same variance at 5% LOS.

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

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

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

PRACTICAL - IX

Chi-Square Test And ANOVA

1] Use the following data to test whether the condition of the home and the condition of child are independent or not.

→

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

→ H_0: 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 H_0.

2] Test the hypothesis that vaccines and diseases are independent or not

		Vaccines	
		Aff.	Not Aff.
Diseases	Aff.	70	46
	Not Aff.	35	37

⇒ H_0 : Vaccines and diseases 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
> chisq.test(y)
        Pearson's Chi-squared test with Yates' Continuity
                  Correction.
data: y
x-squared = 2.0315, df = 1, p-value = 0.1541
```

∵ pvalue is greater than 0.05. ∴ we reject H_0.

3] Perform ANOVA (Analysis of Variance) for following data

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

⇒ H_0: The means 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 = 0.00183

∵ pvalue is less than 0.05. ∴ we reject H_0

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

	Df	Sum sq	Mean sq	F Value	Pr(>F)	
ind	3	71.06	23.688	11.73	0.00183	**
Residuals	9	18.17	2.019			

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

4] The following data gives the life of tyres of brands

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

37

⇒ H_0: The average Life A, B, C & D are equal

```
> 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))
> 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 = 0.002349
> anova = aov(values ~ind, data =d )
> summary(anova)
            Df    Sum Sq   Mean Sq   F Value   Pr(>F)
ind          3    91.44    30.479    6.845     0.00235 **
Residuals   20    89.06    4.453
---
> x = read.csv("C:/Users/Administrator/Desktop/Marks.csv")
> x
     Stats   Maths
1    40      60
2    45      48
3    42      47
4    15      20
5    31      25
6    36      27
7    49      57
8    59      58
9    20      25
10   27      27
```

```
> am1 = mean(x $ stats)
> am 1
[1] 37
> am2 = mean (x $Maths)
> am 2
[1] 39.4
> med 1 = median ( x $ stats )
> med 1
[1] 38.5
> med2 = median (x $Maths)
> med 2
[1]
> n1 = length (x $ stats)
> n1
[1] 10
> n 2 = length ( x $Maths )
> n2
[1] 10
> sd1 = sqrt ( (n1 -1) * var( x $ stats ) /n1 )
> sd 1
[1] 12.64911
> sd 2 = sqrt ( (n2-1) * var ( x $Maths ) /n2 )
> sd 2
[1] 15.20000
> cor ( x $ stats, x $Maths )
[1] 0.830618
```

AM

PRACTICAL - X

Non - Parametric Test

1] Following are the amounts of Sulphur Oxide emitted by the industry 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

→ H_0: Population Median is 21.5

```
> x = c(17, 15, 20, ........., 23, 24, 26)
> me = 21.5
> sp = length(x[x>me])
> sn = length(x[x<me])
> n = sp + sn
> pv = pbinom(sp, n, 0.5)
> pv
[1]  0.4119015
```

∴ pvalue is greater than 0.05. We ~~reject~~ accept H_0.

* If H_1 is me ≠ or me <= then pv = pbinom(sp, n, 0.5)
else pv = pbinom(sn, n, 0.5)

2] Following are the data of 10 observation. Apply sign test to test the hypothesis that the population median 625 against alternative greater than 625

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

→ H_0: Population Median is 625.

```
> x = c(612, 619, 631, ....  670, 663)
> me = 625
> sp = length(x[x>me])
> sn = length(x[x<me])
```

> n = sp + sn
> pv = pbinom (sn, n, 0.5)
> pv

[1] 0.0546875

∴ pvalue is greater than 0.05 . we accept H_0.

3] Following are the data given . Test the hypothesis that population median is $60 against alternative greater than $60 at 5 % LOS using Wilcoxon Sigled Rank Test.

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

H_0 : Population Median is 60

H_1 : Population Median is greater than 60.

> x = c (63, 65, ... 87, 69)

> wilcox. test (x , alter = "greater", me = 60)

Wilcoxon rank sum test with Continuity Correction

data: x and me.

W = 15.5 , pvalue = 0.2037

alternative hypothesis: true location shift is greater than 0

warning message :

In wilcox. test. default (x, alter = "greater", me) :

cannot compute exact pvalue with ties.

∴ pvalue is greater than 0.05. We accept H_0.

4] Following are the data given. Test the hypothesis that population median is 12 against alternative less than 12 using Wilcox Test.

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

=> H_0 : Population Median is 12

H_1 : Population Median is less than 12.

> x = c (15, 17, 24, - - - 24, 26)

> me = 12

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

Wilcoxon rank sum test with continuity correction,

data : x and me

W = 11.5 , pvalue = 0.9463

alternative hypothesis : true location shift is less than 0 .

warning message :

In wilcox . test . default (x, alter = "less", me):

cannot compute exact p-value with ties

∴ p value is greater than 0.05 . We accept H_0 .

5] : The weight of students b/it stopped smoking using Wilcox Test , Test that there is no signi. change

W_B	W_A
65	72
75	74
75	72
62	66
72	73

H_0 : Before and After , There is no significant change

H_1 : There is significant change.

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

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

d = x - y

mu = 0

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

w = 2 , pvalue = 1

alternative hypothesis : true location shift is not equal to 0.

∴ pvalue is greater than 0.05 We accept H_0 .