No.	Title	Page No.	Date	Staff Member's Signature
	Sem - 2			
1	Basics of R	33	25/11/19	AM
2	Binomial Distribution	39	2/12, 16/12	AM
3	Probability distribution	43	9/12/19	AM
			16/12/19	
4	Practice	46	23/12/19	AM
5	Normal Distribution	47	6/1/2020	
6	Z & T distribution sums	44	20/1/2020	AM
7	Large sample test	51	27/1/2020	AM
8	Small Sample Test	53	3/2/2020	AM
9	Large & Small sample Test	56	10/2/2020	
10	ANova & Chi-Square test	58	17/2/2020	AM 9.3
11	Non Parametric Test	61	24/2/2020	

Practical · 1

Basics of R - Software

(1) R is a software of data analysis and statistical computing.
(2) It is a software by which, effective data handling and outcome storage is possible.
(3) It is capable of graphical display.
(4) It is a free software.

Q1] $2^2 + |-5| + 4 \times 5 + 6/5$

→ > 2^2 + abs(-5) + 4*5 + 6/5

[1] 30.2

Q2] x = 20

y = 2x

z = x+y

sqrt(z)

→ > x = 20

> y = 2x

> z = x+y

> sqrt(z)

[1] 7.745967

Q3] x = 10

y = 15

z = 5

80

a) x+y+z
b) xyz
c) √xyz
d) round (sq√xyz)

→ > x = 10
> y = 15
> z = 5
> x + y + z
[1] 27
> a = x * y * z
[1] 2250
~~> sqrt(xyz)~~
> sqrt (a)
[1] 47.43416
> round (sqrt(a))
[1] 47.

Q4] A vector in R - software is denoted by Syntax c.

→ > a = c(2, 3, 5, 7)^2.
> a
[1] 4 9 25 49

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

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

>d= c(1, 2, 3, 4, 5, 6)^c(2, 3, 4)
>d
[1] 2 8 81 16 125 1296

>s= c(2, 9, 5, 8)
>s= c(2, 3, 5, 6)*3
>s
[1] 6 9 15 18

>s= c(2, 3, 5, 6)*c(-2, -3, -5, -7)
>s
[1] -4 -9 -25 -42

>s= c(2, 3, 5, 7) +10
>s
[1] 12 13 15 17

>s= c(2, 3, 5, 7)+c(-2, -3, -1, 0)
>s
[1] 0 0 4 7

>s= c(2, 3, 5, 7)/2
>s
[1] 1.0 1.5 2.5 3.5
```

Q) Find the sum, product, square root of the sum and the product for the following values.
4, 9, 2, 5, 7, 8, 3, 6, 15, 12, 10, 9, 8, 13, 14

→
```
> x = c(4, 9, 2, 5, 7, 8, 3, 6, 15, 12, 10, 9, 8, 13, 14)
> sum(x)
[1] 125
> prod(x)
[1]   8.55932e+12
> sqrt(sum(x))
[1] 11.18034
> sqrt(prod(x))
[1] 2925632
```

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

→
```
> x = c(2, 8, 9, 11, 10, 7, 6)^2
> x
[1]   4   64   81   121   100   49   36
> sum(x)
[1] 455
> prod(x)
[1] 442597478400
> max(x)
[1] 121
> min(x)
[1] 4
```

(i) Matrix

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

```
> x <- matrix (nrow=3, ncol=4, data=c(2,6,7,8,
        9,4,5,0,2,1,4,5))
> x
      [,1]  [,2]  [,3]  [,4]
[1,]    2     8     5     1
[2,]    6     9     0     4
[3,]    7     4     2     5
```

(j) $x = \begin{bmatrix} 4 & 7 & 4 \\ 5 & 8 & 0 \\ 6 & 9 & 2 \end{bmatrix}$, $y = \begin{bmatrix} 6 & 11 & 9 \\ 9 & 12 & 7 \\ 5 & 8 & 4 \end{bmatrix}$

```
> x <- matrix (nrow=3, ncol=3, data=c(4,5,6,7,8,9,4,
        0,2))
> x
      [,1]  [,2]  [,3]
[1,]    4     7     4
[2,]    5     8     0
[3,]    6     9     2

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

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

> z > x+y

> z

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

> x * 2

	[,1]	[,2]	[,3]
[1,]	8	14	8
[2,]	10	16	0
[3,]	12	18	4

> y * 3

	[,1]	[,2]	[,3]
[1,]	18	33	27
[2,]	12	36	21
[3,]	15	24	12

> x * y

	[,1]	[,2]	[,3]
[1,]	24	77	36
[2,]	20	96	0
[3,]	30	72	8

Anu
25.11

080

Practical-2

Binomial Distribution

n = Total number of details
p = P(success)
N = P(failure).
x = No. of success out of n.

$P(x) = {}^{n}C_{x} \; p^{x} N^{n-x}, \; n = 0, 1, \cdots n$

$E(X) = np$

$V(x) = npq$

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

1) Toss a coin 10 times with probability of h = 0.6. let n be the no. of heads.
2) Find the probability of
   (i) 7 heads
   (ii) 4 heads
   (iii) at most 4 heads
   (iv) atleast 6 heads.
   (v) No head.
   (vi) all heads. Also find expectation & variance

>n = 10

>p = 0.6

>q = 0.4.

(i) > dbinom (7, 10, 0.6)

[1] 0.2149908

(ii) > dbinom (4, 10, 0.6)

[1] 0.1114767

(iii) pbinom (4, 10, 0.6)

[1] 0.1662386

(iv) 1- pbinom (6, 10, 0.6)

[1] 0.3822806

(v) No head.

(v) > dbinom (0, 10, 0.6)

[1] 0.0001048576

(vi) All heads.

>. dbinom (10, 10, 0.6)

[1] 0.006046618

```
> x=3
> dbinom(x,n,p)
[1] 0.0081
> x=4
> dbinom(x,n,p)
[1] 0.00045
> x=5
> dbinom(x,n,p)
[1] 1e-05
```

(3)

```
> x=10
> p=0.1
> n=100
> dbinom(x,n,p)
[1] 0.1318653
```

(4) (i) $P(X \leq 5)$

$n=12$
$p=0.25$
$x=5$

```
> pbinom(x,n,p)
[1] 0.9455978
```

(ii)

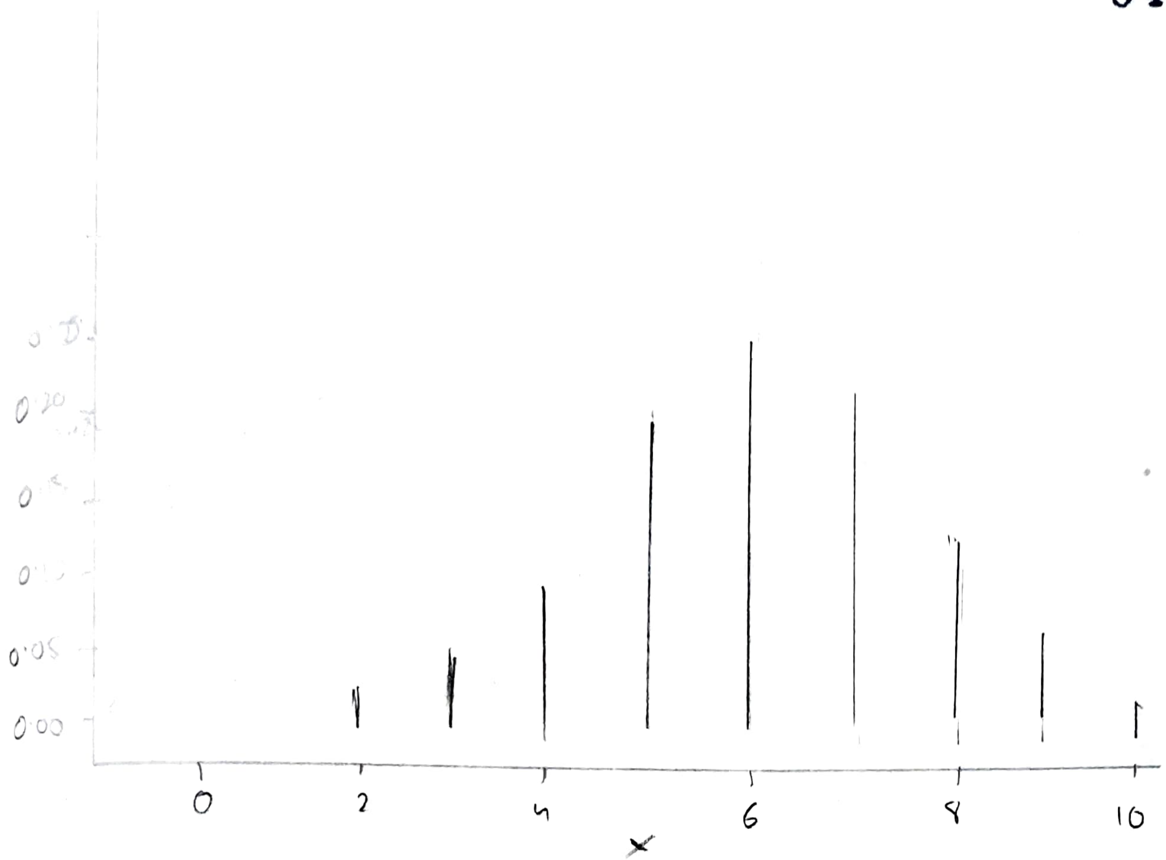
```
> x=7
> 1-pbinom(x,n,p)
[1] 0.00278151
```

(iii)

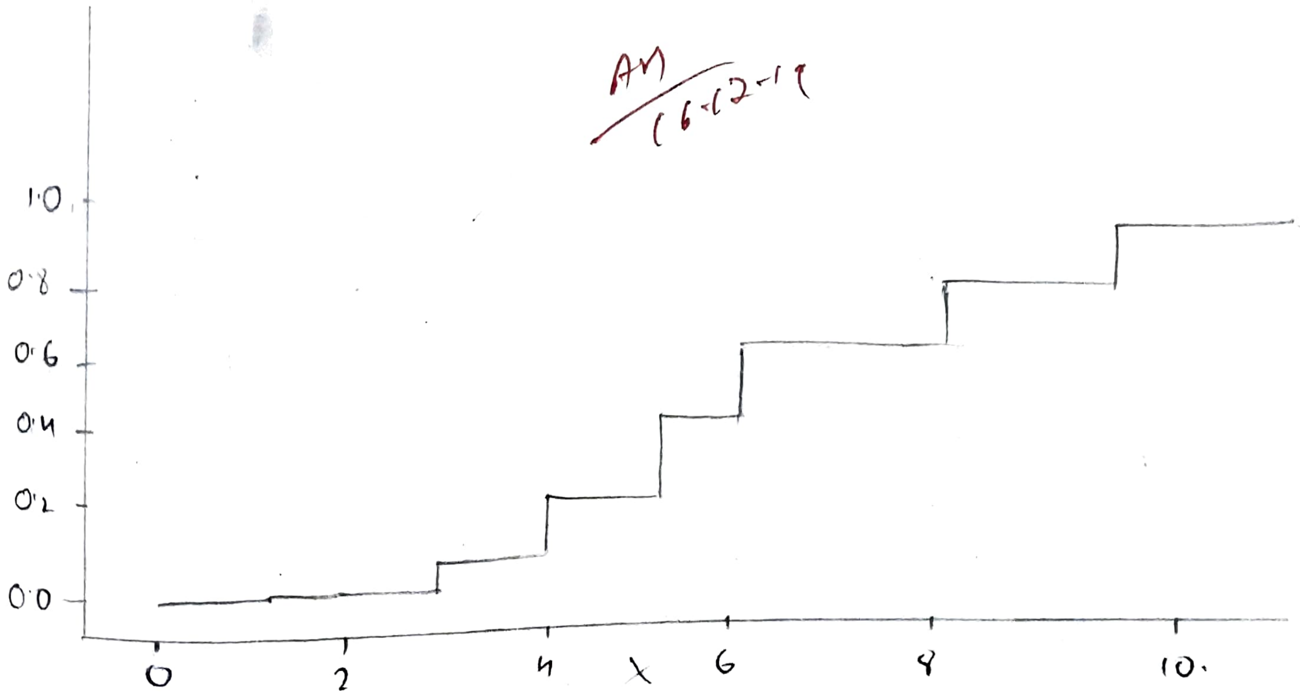
```
> x=6
> dbinom(x,n,p)
[1] 0.04014945
```

```
> n=10
> p=0.9
> x=6
> 1-pbinom(x,n,p)
[1] 0.9872048

> n=30
> p=0.2
> p1=0.88
> qbinom(p1,n,p)
[1] 9

> n=10
> p=0.6    x = 0:n
> bp=binom(x,n,p)
> d=data.frame("x.values"=x, "probability"=bp)
print(d)
   x.values   probability
1      0      0.001048576
2      1      0.0015728640
3      2      0.0106168320
4      3      0.0424673280
5      4      0.1114767360
6      5      0.2006581248
7      6      0.2508226560
8      7      0.2149908480
9      8      0.1209323520
10     9      0.0403107840
11    10      0.0060466176
```

```
> plot (x, bp, "h")
> cp = pbinom (x, n, p)
> plot (x, cp, "s")
```



AM
16-12-19



Practical - 3

i) Check the following P.M.F (probability mass function). or not.

i)

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

ii)

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

iii)

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

→ (i) The conditions to check P.M.F are.

$0 \leq P(x) \leq 1$, ——— (1)

$\sum P(x) = 1$ ——— (2)

As the first condition isn't satisfied, it is not a P.M.F.

(ii) The conditions to check P.M.F are

$0 \leq P(x) \leq 1$ ——— (1).

$\sum P(x) = 1$ ——— (2).

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

Since, both the conditions are satisfied, it is a P.M.F.

(iii) The conditions to check p.m.f are,

$0 \le P(x) \le 1$ ——— ①

$\Sigma P(x) = 1$ ——— ②

> prob = c(0.4, 0.2, 0.3, 0.2, 0.1).

> sum(prob)

[1] 1.2

Here, as the second condition is not satisfied; therefore it is not a p.m.f.

Q] Following is a p.m.f of x. Find, mean & variance of x.

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

→

x	$P(x)$	$x P(x)$	$x^2 P(x)$
1	0.1	0.1	0.1
2	0.15	0.3	0.6
3	0.2	0.6	1.8
4	0.3	1.2	4.8
5	0.25	1.25	6.25
		$\Sigma x P(x) = 3.45$	$\Sigma x^2 P(x) = 13.55$

Mean $= E(x) = \Sigma x\, P(x) = 3.45$

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

$= 13.55 - (3.45)^2$

$= 1.6475$

```
> x = c(1, 2, 3, 4, 5)
> prob = (0.1, 0.15, 0.2, 0.3, 0.25)
> a = x * prob
> mean = sum(a)
> mean
[1] 3.45
> b = (x^2) * prob
> var = sum(b) - mean^2
> var
[1] 1.6475
```

041

2] Find mean and variance of X

X	5	10	15	20	25
P(x)	0.1	0.3	0.2	0.25	0.15

```
-> > x = c(5, 10, 15, 20, 25)
   > prob = c(0.1, 0.3, 0.2, 0.25, 0.15)
   > a = x* prob
   > mean = sum(a)
   > mean
   [1] 15.25
   > b = (x^2) * prob
   > var = sum(b) - mean^2
   > var
   [1] 38.6875
```

Q] Find c.d.f of following p.m.f & draw the graph of c.d.f

(i)

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

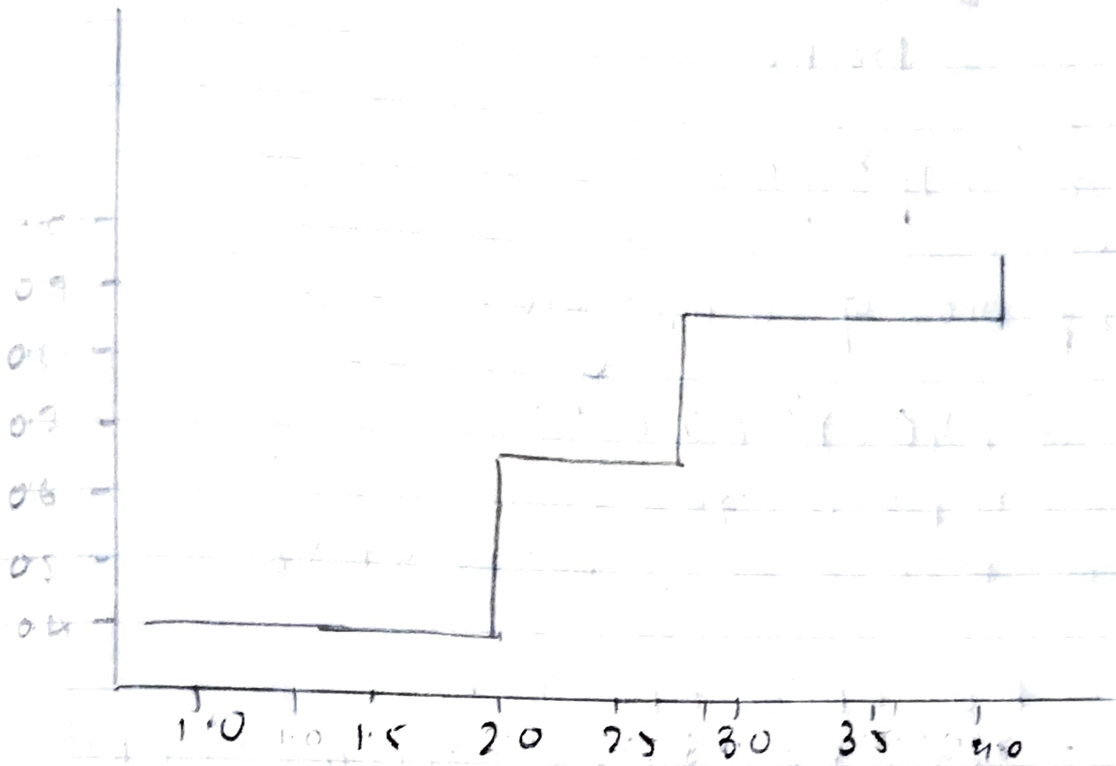
```
→ > x = c(1,2,3,4)
> prob = c(0.4, 0.3, 0.2, 0.1)
> a = cumsum(prob)
> a
[1] 0.4   0.3   0.2   0.1
> plot(a, x, "s")
```

(ii)

X	0	2	4	6	8
P(x)	0.2	0.3	0.2	0.2	0.1

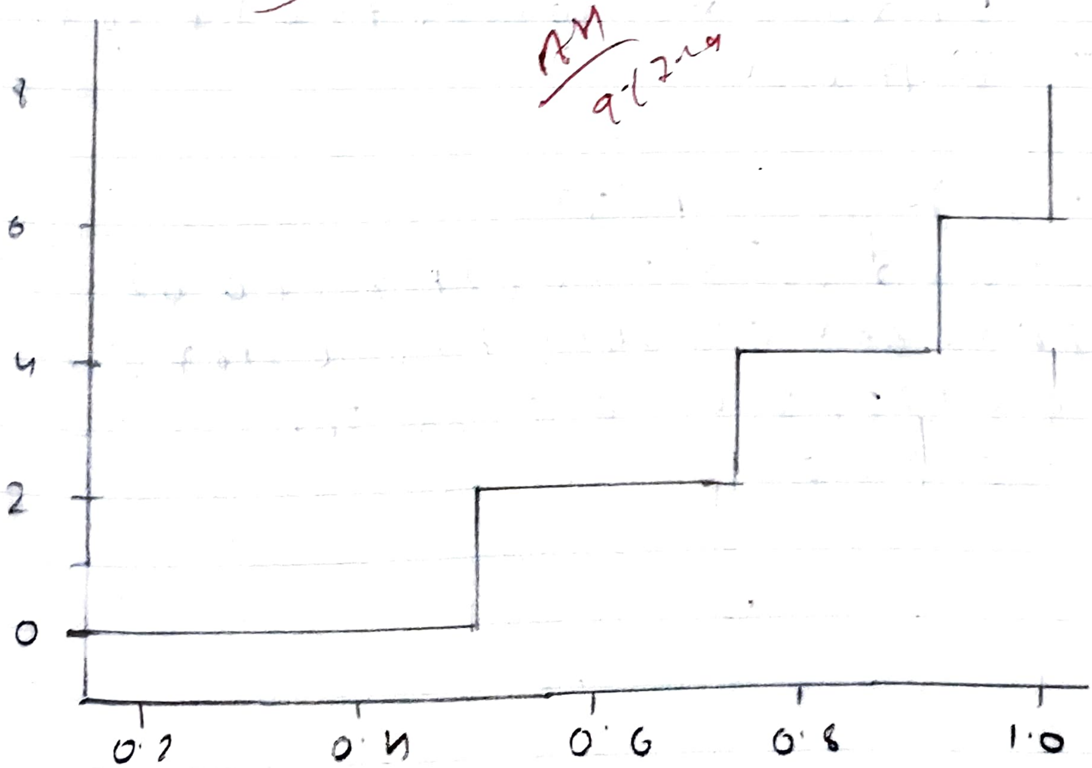
```
→ > x = c(0, 2, 4, 6, 8)
> prob = c(0.4, 0.3, 0.2, 0.1)
> a = cumsum(prob)
> a
[1] 0.4   0.3   0.2   0.1
> plot(a, x, "s")
```

(i)



(ii)

AM
9/7/24



40

Practical - 4

Practice

$P(x=r) = (^nC_r) p^r q^{n-r}$

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

(i) $P(x=7) = (^8C_7)(0.6)^7 (0.4)^{8-7}$

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

$= 8 \cdot 0.2799 \times 0.4 = 0.08956$

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

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

$= 1 \times 1 \times 0.00065536 + 8 \times 0.6 \times 0.0016384 + 28 \times 0.36 \times 0.004096 + 56 \times 0.216 \times 0.01024$

$= 0.1736204$

(iii) $P(x = 2 \& 3) = P(2) + P(3)$

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

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

$= 0.04128768 + 0.12386304$

$= 0.16515072$

Solution:

Q1] (i) $P(X \leq 7)$

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

(ii) $P(X>12)$

```
> 1- pnorm (12, 10, 2)
[1] 0.1586553
```

(iii) $P(5 \leq X \leq 12)$

```
> pnorm(12,10,2) - pnorm(5,10,2).
[1] 0.8351351
```

(iv) $P(X<K)=0.4$

```
> K= qnorm(0.4, 10, 2)
> K
[1] 9.493306
```

Q2] (i)

```
P1 = pnorm (110, 100, 6)
cat ("P(X <= 110) is:", P1)
P(X <= 110) : 0.9522096
```

(ii)

```
P2 = 1- pnorm (105, 100, 6)
cat ("P(X>105) is:", P2)
P(X>105) = 0.2023284
```

(iii)

```
P3 = pnorm (92, 100, 6)
cat ("P (X <= 92) is:", P3)
P(X<=92) = 0.9121122
```

(iv)

```
P4 = pnorm (110, 100, 6) - pnorm (95, 100, 6)
cat ("P(95 <= X <=110) is: ", P4)
P(95 <= X <= 110) = 0.7498813
```

(v)

```
K = qnorm (0.9, 100, 6)
cat ("P (X<K) = 0.9 is: ", K)
P(X<K) = 0.9 = 107.6893.
```

Practical - 5

Normal Distribution:

Normal distribution is example of continuous probability Distribution.

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

(i) $P(X = x) =$ dnorm(x, μ, σ)

(ii) $P(X \le x) =$ pnorm(x, μ, σ)

(iii) $P(X > x) = 1 -$ pnorm(x, μ, σ)

(iv) k $P(X \le k) = p$

qnorm(p, μ, σ)

(v) To generate a random sample of size n:

rnorm(n, μ, σ)

1) A random variable X follows normal distribution with $\mu = 10$, $\sigma = 2$ find

(i) $P(X \le 7)$

(ii) $P(X > 12)$

(iii) $P(5 \le X \le 12)$

(iv) $P(X < k) = 0.4$.

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

$\sigma =$ sqrt(36)

(i) $P(X \le 110)$

(ii) $P(X > 105)$

(iii) $P(X \le 92)$

(iv) $P(95 \le X \le 110)$

(v) $P(X < k) = 0.9$

PO

Q] $X \sim N(10, 3)$
Generate 10 random samples ~~gener~~ find the Sample mean, median, variance & standard deviation.

Q] Plot the standard Normal curve.

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

Q] $X \sim N(50, 100)$
Find (i) $P(X \le 60)$
(ii) $P(x > 65)$
(iii) $P(45 \le x \le 60)$.

→ Solution:

```
P1 = pnorm(60, 50, 10)
cat("P(X ≤ 60) is = ", P1)
```

$P(X \le 60) = 0.841344 7$

```
P2 = 1 - pnorm(65, 50, 10)
cat("P(X > 65) is = ", P2)
```

$P(X > 65) = 0.0668072$.

```
P3 = pnorm(60, 50, 10) - pnorm(45, 50, 10)
cat("P(45 ≤ X ≤ 60) is = ", P3)
```

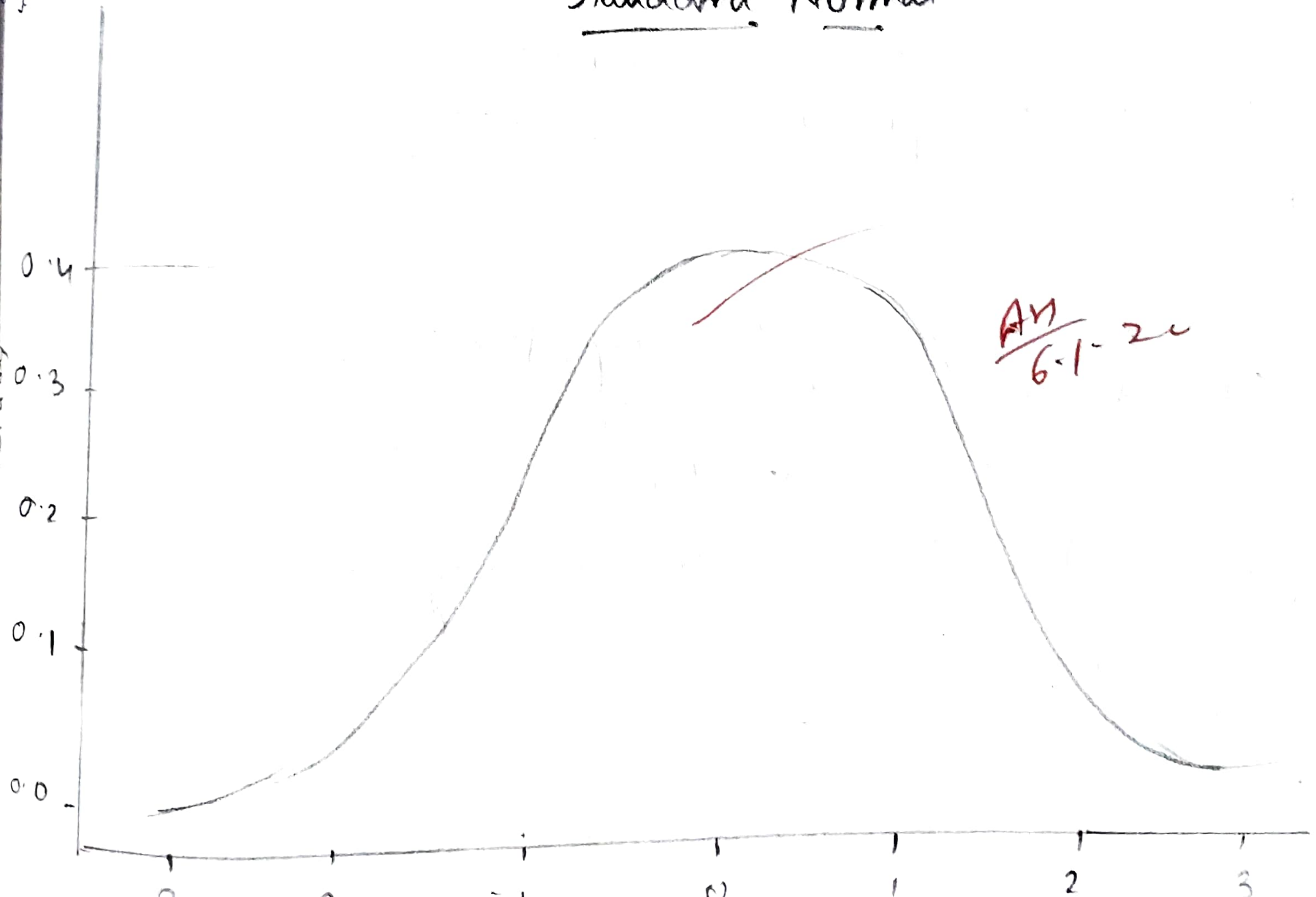
$P(45 \le X \le 60) = 0.5328072$.

048

```
> x=rnorm (10, 10, 3)
> x
[1] 12.666599  8.790523  5.065435  12.282722  14.053965
    11.577426  9.761933  13.339845  13.502146  11.933185
> me = median (x)
> me
[1] 12.05795
> dm = mean (x)
> dm
[1] 11.28738
> n=10
> variance = (n-1)* var(x)/n
> variance
[1] 6.732067
> sd = sqrt (variance)
> sd
[1] 2.594623.
```

[4]

Standard Normal



Practical - 6

Aim: Z and T distribution sums.

Q1] Test the hypothesis $H_0 : \mu = 20$ against $H_1 : \mu \neq 20$. A sample of size 400 withis selected and the sample mean is 20.2 & standard deviation is 2.25. Test at 5% level of significance.

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

Q2] We want to test the hypothesis $H_0, \mu_0 = 250$, $H_1 : n \neq 250$. A sample of size 100 has a mean of 275 & standard deviation of ~~275~~ 30. Test the hypothesis at 5% level of significance.

→
```
> mo = 250
> mx = 275
> sd = 30
> n = 100
```

```
>zcal = (mx-mo)/(sd/sqrt(n))
>zcal
[1] 8.333333
>cat("z calculated is=", zcal)
>calculated is = 8.333333
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0
```

Since, pvalue is less than 0.05, we reject Ho

3] We want to test the hypothesis Ho: 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
```

4] In a big city, 325 men out of 600 men were found to be self employed. Does this information support the conclusion that exactly half of the men in the city are self employed? (P = 0.5, n = 600)

H.AO

→ > P=0.5
> n=600
> p= 365/600
> Q= 1-P (Big)
> zcal= (p-P)/sqrt(P*Q/n))
> zcal
[1] 2.041241
> pvalue= 2* (1-pnorm (abs (zcal)))
> pvalue
[1] 0.04122683
Since, pvalue is less than 0.05, we reject Ho

a5) Test the hypothesis Ho: μ=50 against H1: μo ≠ 50
A sample of 30 is collected. Sample:
50, 49, 52, 44, 45, 48, 46, 45, 49, 45, 48, 47, 55, 54, 46, 58, 47, 44, 59, 60, 61, 41, 52, 44, 55, 56, 46, 45, 48, 49

→ > mo=50
> x=c(50,49,52,44,45,48,46,45,49,45,48,47,55,54,46,58,47,44,59,60,61,41,52,44,55,56,46,45,48,49)
> n= length(x)
> n
[1] 30
> mx = mean(x)
> mx
[1] 49.3333

```
> variance = (n-1)*var(x)/n
  sd = sqrt(variance)
> sd
[1] 1.563772
> zcal = (mx-mu)/sd/sqrt(n)
> zcal
[1] -0.652464
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.5116234
```

Since, pvalue is more than 0.05 we accept $H_0: \mu = 50$

AM
20-01-20

Practical - 1

Topic: Large Sample test

Q1] 2 random sample of size 1000 & 2000 are drawn from 2 population with a standard deviation 2 & 3 respectively. Test the hypothesis that the 2 population means are equal or not at 5% level of significance. The sample means are 67 & 68 respectively

→
```
> n1 = 1000
> n2 = 2000
> mx1 = 67
> mx2 = 68
> sd1 = 2
> sd2 = 3
> zcal = (mx1-mx2)/sqrt((sd1^2/n1) + (sd2^2/n2))
> zcal
[1] -10.84652
> cat("z calculated is=", zcal)
z calculated is = -10.84652
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0
```

Q2] A Study of noise level in 2 hospital is done. Following data is calculated. 1st sample size = 84, 1st sample mean = 61.2, $sd_1 = 7.9$, $n_2 = 34$, $mx_2 = 59.4$, $sd_2 = 7.8$

Test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ against 1% level of significance

> n1 = 84
> n2 = 34
> mx1 = 61.2
> mx2 = 59.4
> 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)))
> pvalue
[1] 0.258066 , we accept H.0 $\mu_1 \neq \mu_2$

Q3) From each of 2 population of oranges, the following samples are collected. Test whether the proportion of bad oranges are equal or not.
1st sample size = 250, 2nd sample size = 200, no. of bad orange = 44 (1st sample), 2nd sample = 30
Ho : $P_1 = P_2$ as $H_1 : P_1 \neq P_2$.

→ > n1 = 250
> n2 = 200
> p1 = 44/250
> p2 = 30/200
> p = (n1 * p1 + n2 * p2) / (n1 + n2)
> Q = 1 - p
> zcal = (p1 - p2) / sqrt (p * q * (1/n1 + 1/n2))
> zcal
[1] 0.164444

```
> Q = 1-P
> Q
[1] 0.8355556
> zcal = (P1-P2) /sqrt (P* Q* (1/n1 + 1/n2))
> zcal
[1] 0.7393581
> pvalue= 2* (1-pnorm (abs(zcal)))
> pvalue
[1] 0.4596896
```

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

(4) Random sample 400 mens & 600 women where asked as whether they loan + atom nearby. 200 males & 390 males. Test the hypothesis that the proportions of males & females following the proposal equal or not at 5% level of significance

(5) The following are the two independent sample from the two population. Test equality of two population means at 5% level of significance.

Sample 1: (74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 76)

Sample 2: (72, 76, 74, 71, 70, 78, 70, 72, 75, 79, 79, 74, 75, 78, 72, 74, 80)

Ho :

```
>n1 = 400
>n2 = 600
>p1 = 200/400
>p2 = 390/400
>p = (n * p1 + n1 * p2) / (n1 + n2)
>q = 1 - p
>zcal = (p1 - p2) / sqrt (p * q * (1/n1 + 1/n2))
>zcal
[1] 4.72475
>pvalue = 2 * (1 - pnorm (abs (zcal)))
>pvalue
[1] 2.303972 e-06
```

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

Ho : $\mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$.

```
>x1 = c (c (74, 77, 74, ---, 76, 70).
>n1 = length (x1)
>mx1 = mean (x1)
>variance = (n1 - 1) * var (x1) / n1
>variance
[1] 0.4508
>sd1 = sqrt (variance)
>sd1
[1] 0.6714
>x2 = c (c (72, 76, 74, ..., 74, 80)
>n2 = length (x2)
>mx2 = mean (x2)
>variance = (n2 - 1) * var (x2) / n2
>variance
[1] 0.6163
>sd2 = sqrt (variance)
>sd2
[1] 0.7850
>tcal = (mx1 - mx2) / sqrt ((sd1^2 / n1) + (sd2^2 / n2))
```

> cat ("tcalculated is", tcal)

[1] tcalculated is : ~~1.5781~~ 1.5751

Since it is greater than 0.05, we accept it.

t.test (x, x₁)

pvalue = 0.1387

Accept

AM

27.01.20

Practical-8

Topic: Small Sample Test

Q1) The random sample of 15observations is given by
80, 100, 110, 105, 122, 70, 120, 140, 101, 88, 83, 95, 89, 107,
125.
Do this data support the assumption that the popula
mean is 100?

Ho: $\mu = 100$

```
> x = c (80, 100, 110, 105, 122, 70, 120, 120, 101, 86, 83, 95, 89, 107,
> length(x)
[1] 15
> mean(x)
[1] 100.3333
> t.test (x)
```

One sample t-test.

data: x
t = 24.029, df = 14, p-value = 8.819 e-13
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
91.37775 109.28892
sample estimates:
mean of x
100.3333

Since, p-value is less than 0.05, we reject it.

Q2) Two groups of 10 students scored the following mar
Group 1: 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
Group 2: 16, 20, 14, 21, 20, 18, 13, 15, 17, 21.

:20

Test the hypothesis that there is no significant difference between the scores at 1% level of significance. $H_0 : \mu_1 = \mu_2$

→ > group A = c(c(18, 22, 21, 17, 20, 17, 23, 20, 22, 21)
> group B = c(16, 20, 14, 21, 20, 18, 13, 15, 17, 21)
> t.test (group A, group B)

Welch Two sample t-test

data : group A and ground B
t = 2.2573, df = 16.376, p-value = 0.03798
alternative hypothesis: true difference in means is not equal to 0.
95 percent confidence interval:
0.1628205 5.0371795.
Sample estimates:
mean of x mean of y
20.1 17.5

Since, p-value is ~~less~~ greater than 0.01% of significance, we accept H0. at 1% significance

Two types of medicines are used for 5 & 7 patients for reducing their weight. The decrease in their weight after medicine are given below.
Medicine A: 10, 12, 13, 11, 14.
Medicine B: 8, 9, 12, 14, 15, 10, 9
Is there a significant diffrent in the effeciency of the medicines?
$H_0 : \mu_1 = \mu_2$

> MedicineA <- c(10, 12, 13, 11, 14)
> MedicineB <- c(8, 9, 12, 14, 15, 10, 9)
> t-test (MedicineA, MedicineB)

Welch Two Sample t-test

data: MedicineA and MedicineB
t = 0.80384, df = 9.7594, p-value = 0.4406
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.781171   3.781171
sample estimates:
mean of x   mean of y
 12   11

Since, p-value is greater, we accept Ho. $U_1 = U_2$ at 1% level of significance

(iv) Since The weight reducing diet program is conducted & the observation is noted for 10 participants. Test whether the program is effective or not.
Before - 120, 125, 115, 130, 123, 119, 122, 127, 128, 118
After - 111, 114, 107, 120, 115, 112, 112, 120, 119, 112.
Ho: there is no significant difference in weight against
H1: The diet program reduce weight.

> x <- c(120, 125, 115, 130, 123, 119, 122, 127, 128, 118)
> y <- c(111, 114, 107, 120, 115, 112, 112, 120, 119, 112)
> t.test (x, y, paired = T, alternative = "less")

Paired t-test

data: x and y
t = 17, df = 9, p-value = 1
alternative hypothesis: true difference in mean is less than 0.
95 percent confidence interval:
 -Inf   9.416556
sample estimates:
mean of the differences:
 8.5

Since, pvalue is greater than 0.05, we accept Ho: $U_0 =$
The diet program reduce weight.

Q.5] ~~Sample~~ A - 66, 67, 75, 76, 82, 84, 88, 90, 92
Sample B 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97
Test ~~the~~ population mean are equal or not.

$H_0: \mu_1 = \mu_2$

→ > x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)
> y = c(64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)
> t.test(x, y)

Welch Two Sample T-Test

data : x and y

t = -0.63968, df = 17.974, p-value = 0.5304

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

95 percent confidence interval:
-12.853992 6.853992.

Sample estimates:

mean of x	mean of y
80	83

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

06) The following are the marks before & after of a training program. Test the program is effective or not.

Before - 71, 72, 74, 69, 70, 74, 76, 70, 73, 75

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

```
> x = c (71, 72, 74, 69, 70, 74, 76, 70, 73, 75)
> y = c (74, 77, 74, 73, 79, 76, 82, 72, 75, 78)
> t.test (x, y, paired = T, alternative = "greater")
        Paired t-test
data: x and y
t = -4.4691, df = 9, p-value = 0.9992
alternative hypothesis: true difference in means is
greater than 0
95 percent confidence interval:
 -5.076639     Inf
sample estimates:
mean of the differences
          -3.6
```

Here Ho: no significant difference, H1: increase in marks

Since, ~~H is~~ pvalue is greater, we accept Ho: Ho = There is uincrease in marks.

AM
03.12.20

020

Practical-9

Topic: Large and small sample test.

Q1] The arithmetic mean of a sample of 100 items from a large population is 52. If the standard deviation is 7, test the hypothesis that the population mean is ~~65~~ 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. Does this information support that exactly half of the males in the city are smokers? Test at 1% LOS

Q3) 1000 articles from a factory A are found to have 2% defectives. 1500 articles from a second factory B are found to have 1% defectives. Test at 5% level of significance that the 2 factories are similar or not.

Q4] 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 64?

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

ii) 2 random samples were drawn from 2 normal population & their 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 populations have the same variance at 5% LOS.

056

Solution:

1) >n=100
>mx=5
>sd=7
>m0=55
>zcal= (mx-m0)/sqrt(sd^2/n)
>zcal
[1] -4.285714.
>pvalue = 2* (1-pnorm(abs(zcal)))
>pvalue
[1] 1.82153e-05
Since p-value is ~~greater~~ less than 0.05, we ~~accept~~ reject it.

2) >n1 = 350
>n2 = 700
>p=0.5
>q=1-p
>P=n1/n2
>p
[1] 0.5
>zcal=(P-p)/sqrt(p*q/n)
>zcal
[1] 0
>pvalue = 2* (1-pnorm(abs(zcal)))
>pvalue
[1] 1
Pvalue is greater than 0.01, we accept H0. P=0.5

③ 1> $n_1 = 1000$
> $n_2 = 1500$
> $P_1 = 0.02$
> $P_2 = 0.01$
> P = (n1*p1 + n2 * p2) /(n1+ n2)
> P
[1] 0.014
> q = 1 - p
> q
[1] 0.986
> zcal = (p1-p2)/sqrt(p*q*(1/n1 + 1/n2)
> zcal
[1] 2.084842
> pvalue = 2* (1-pnorm (abs(zcal)))
> pvalue
[1] 0.03708364

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

) > n = 400
> mx1 = 99
> mx2 = 100
> var = 64
> sd = sqrt(var)
> sd
[1] 8
> zcal = (mx1 - mx2) / sqrt(sd^2/n)
> zcal
[1] -2.5
> pvalue = 2* (1-pnorm (abs(zcal)))
> pvalue
[1] 0.01241933

Pvalue is less than 0.05 we reject $H: \mu = 0.5$

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

One sample t-test

data : x

t = 917.94, df = 6, p-value = 5.522e-09

alternative hypothesis : true mean is not equal to 0

95 percent confidence interval:

64.06479 71.62092

sample estimates:

mean of x

68.74286

Pvalue is less than 0.01, we reject $H_0 : \mu = \mu_1$

```
> x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)
> y = c(64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)
> F = var.test(x, y)
> F
```

F test to compare two variances.

data : x and y

F = 0.70686, num df = 8, denom df = 10, p-value = 0.6359

alternative hypothesis : true ratio of variances is not equal to

95 percent confidence interval:

0.1833662 3.0360393

sample estimates:

ratio of variances

0.7068567

Since, Pvalue is greater than 0.05, we accept Ho. Ho = The population have some variance.

Practical 10

Topic: Anova & Chi-Square.

Q1) Use the following data to test, whether the Cleanliness of home & cleanliness of the child is independent or not

		Clean	Dirty
Cleanliness of child	Clean	70	50
	fairly clean	80	20
	Dirty	35	45

→ Sol^n Ho: Cleanliness of child & Cleanliness of Home 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
```

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

Q] Use the following data to find vaccination of on a particular disease are independent or not.

058

		Disease Affected	Not Affected
vaccination	given	20	30
	Not given	25	35

H0: vaccination & disease are independent

```
> x = c(20, 25, 30, 35)
> m = 2
> n = 2
> y = matrix (x, nrow = m, ncol = n)
> y
      [,1]  [,2]
[1,]   20    30
[2,]   25    35
> pv = chisq.test(y)
> pv
```

Pearson's Chi-squared test with Yates' continuity correction

data: y

X-squared = 0, df = 1, p-value = 1

Since, pvalue is more than 0.05, we accept H0: vaccina & diseases are independent

Q3] Perform ~~anova~~ anova for the following data

Varieties	Observations
A	50, 52.
B	~~53, 55,~~ 53
C	60, 58, 57, 32
D	52, 54, 54, 55

H0: The means of the varieties are equal.

```
>x1=c(80,52)
>x2=c(58,55,53)
>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, p-value = 0.00183
>anova = aov(values~ind, data=d)
>anova.
Call:
   aov(formula = values~ind ,data=d)
Terms:

                            ind        Residuals.
Sum of Squares        71.06410      18.16667
Deg. of Freedom           3              9.
Residual      standard error: 1.420746
Estimated effects may be unbalanced
```

Since pvalue is less than 0.05, we reject Ho: The means of the varieties are equal.

Q4) The following data gives life of tyre of ~~4~~ brands

Type

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)

Test the hypothesis that average life of ...

H_1: ~~Avg~~ Avg life 4 tyres are same

```
> 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 = 0.00~~
 pvalue = 0.002349
> anova = aov( values ~ ind, data = d)
> anova
Call:
   aov( ~~bata~~formula = values ~ ind, data = d)
Terms:
                        ind       Residuals
Sum of squares        91.4381     89.0619
Deg. of freedom          3           20

Residual standard error: 2.110236
Estimated effects may be unbalanced.
```

Since pvalue is less than 0.05, we reject Ho: The average life of 4 tyres of brands are same.

6.20

Q5) 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.

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

H0: IQ and economic condition are independent

```
→ >x = c(460, 330, 240, 140, 200, 160)
> m = 3
> n = 2
> y = matrix(x, nrow = m, ncol = n)
> y
        [,1]   [,2]
[1,]    460    140
[2,]    330    200
[3,]    240    160
> pv = chisq.test(y)
> pv

        Pearson's chi-squared test.
data: y
X-squared = 39.726, df = 2, pvalue = 2.364e-09
```

since pvalue is less than 0.05, we reject H0: a & Ha

Questions:

Q1] Following are the amounts of sulphur oxide emitted by indus... in 20 days. Apply sign test to test the hypothesis that the population median is 21.5

Obs - 17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, ...

Ho: population median is 21.5

```
→ >x = c(17,15,20,29,19,18,22,25,27,9,24,20,17,6,24,14,15,23,24,2...
>median = 21.5
> length(x)
> sp = length(x[x>median])
> sn = length(x[x<median])
> n = sp+sn
> n
[1] 20
> pv = pbinom(sp,n,0.5)
> pv
[1] 0.4119015
```

Since, p-value more than 0.05, we accept Ho at 5% level of significance.

Q2] Following are 10 observations - 612, 619, 631, 628, 643, 640, 655, 649, 670, 663. Apply sign test to test the hypothesis that the population median is 625 against the alternati... is greater than 625 at 1% level of significance.

Note: If the alternative is greater, then pv = pbinom(sn,n,0...

→ Ho: Population median is 625.

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

Practical - 11

Aim: Non Parametric test.

```
>pr= pbinom (sn, n, 0.5)
>pr
[1] 0.0546875
```

Since pvalue is more than 0.01, we accept Ho at 1% level of significance.

Q3] 10 observations are - 36, 32, 21, 30, 24, 25, 20, 22, 20, 18
Using sign test, test the hypothesis that the population median is 25 against the alternative it is less than 25 at 5% level of significance

Ho: population median is 25.

```
> x=c(36, 32, 21, 30, 24, 25, 20, 22, 20, 18)
> length(x)
[1] 10
> median = 25
> sp= length(x[x>median])
> sn= length(x[x<median])
> n =sp+sn
> n
[1] 9
> pr= pbinom (sp, n, 0.5)
> pr
[1] 0.2539063
```

Since pvalue is more than 0.05, we accept Ho at 5% level of significance.

Q4) The following are some measurements - 63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 65. Using wilcoxon signed rank test, test the hypothesis that the population median is 60 against the alternative, it is greater than 60 at 5% level of significance.

Ho: population median is 60.

→ > x = c(63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 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.

Since, pvalue is greater than 0.05, we accept Ho at 5% level of significance.

5) Observations are 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26, use the w.s.r.t to test the hypothesis that the population median is 20 against 5% l.o.s.

Ho: population median is 20.

→ > x = c(15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26)
> wilcox.test (x, alt = "less", mu = 20)

~~warn~~

wilcoxon signed rank test with continuity correction.

062

data: x

V = 48.5 , p-value = 0.9232

alternative hypothesis : true location is less than 20

Since, p-value is greater than 0.05, we accept H0 at 5% level of significance.

(6) 5 observations - 20, 25, 27, 30, 18. Test the hypothesis that the population median is 25 against the alternative it is not 25.

AM
9-3-20