Practical no. 1.

* Basics of R software:
1. R is a software for data analysis and statistical computing.
2. It is a software by which effective data handling and ~~st~~ outcome storage is possible
3. It is capable of graphical display
4. It is a free software.

1.) $2^2 + |-5| + 4 \times 5 + 6 \div 5$

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

[1] 30.2

2.) $x=20$, $y=2x$, $z=x+y$, $\sqrt{z}=?$

> x=20

y= 2*x

z= x+y

sqrt(z)

[1] > 7.7459(-)

3.) x=10, y=15, z=5

Find x+y+z ; x.y.z ; $\sqrt{xyz}$, round ($\sqrt{x}$

> x=10

> y=15

> z=5

> x+y+z

[1] 30.

ii.] > a = x*y*2
> a
[1] 750.

iii.] > sqrt(a)
[1] 27.38613

iv.] > round(a)
[1] 27

Q.4. A vectar in R software is denoted by c

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

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

iii.) > x = c(2, 3, 5, 7, 9)^c(2, 3)
> x
[1] : warning message

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

v.] > x (21, 23, 1, 4) * 3
> x
[1] 63 69 3 31

vi.) > x = c(21, 23, 1, 4) * c(-2, -3, -5, -7)
> x
[1] -42 -69 -5 -28

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

viii.) > x = (c(2, 3, 5, 7)) / 2
> x
[1] 1.0 1.5 2.5 3.5

Q.5.] Find the sum, product, sqroot of sum a product for the given values.

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

> i = c(4, 9, 2, 5, 7, 8, 3, 6, 15, 12, 10, 9, 8, 13, 14,
> s = sum(i)
> p = prod(i)
> s
> p
> sqrt(s)
> sqrt(p)

[1] 125
[1] 11.180
[1] 8.559323 e+12
[1] 2925632

Q.6: Find the sum, product, max, min value

x = c(2, 8, 9, 11, 10, 7, 6)^2.

> sum(x)
[1] 455
> Product(x)
[1] 4425974784

```
>max(x)
[1] 12
>min(x)
[1] 4
x<matrix(nrow=4,ncol=2,data=c(12,2,4,5,8))

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

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

x y

```
x<-matrix(nrow=3,ncol=3,data=c(4,5,6,7,8,9,40,23)
y<-matrix(nrow=3,ncol=3,data=c(6,4,5,11,12,8,9
>x
      [,1]  [,2]  [,3]
[1,]   4     7     4
[2,]   5     8     0
[3,]   6     9     2
>y
      [,1]  [,2]  [,3]
[1,]   6     11    9
[2,]   4     12    2
[3,]   5     8     4
>x*2
      [,1]  [,2]  [,3]
[1,]   8     14    8
[2,]   10    16    0
[3,]   12    18    4
```


Practical - 2

* Binomial distribution

Atleast = 1 - pbinom
Atmost = pbinom

n = Total number of trials
p = P(success)
q = P(failure)
x = numbers of success out of n
$P(x) = nC_x \, p^x \, 2^{n-x}$
$E(x) = np$
$V(x) = np^2$
dbinomial(x, n, p) = P(X = x)
$n * p$
$n * p * q$
pbinom(x, n, p)

* Exercise

Q. Toss a coin 10 times P(heads) = 0.5. Let x be the of heads.
Find the probability.
i.) seven heads
ii) Four heads
iii) Atmost four heads
iv) Atleast six heads.
v.) No Heads
vi.) All heads.
Also find expectation and variance

Practical - no. 3.

Q.1 Check the following are p.m.f ( ) 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

solution :-

(i.) $0 \leq P(x) \leq 1$ (ii.) $\Sigma P(x) = 1$.

(1). Since, it not satisfy the first condition.

Solution;-

(2) Since all the value of $P(x)$ are more than zero and less than one, 1st condition is satisfied.

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

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

∴ Since, both the condition are satisfied it is a p.m.f

(3) Since, all the value of p(x) is between 0 to 1, therefore, condition 1 is satisfied.
But

$$\Sigma P(x) = P(0)+P(1)+P(2)+P(3)+P(4)$$
$$= 0.4+0.2+0.3+0.2+0.1$$
$$= 1.2.$$

It is more than 1
It does not satisfy condition 2
Hence, not ~~proof~~. prob

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

∵ It's not satisfy second condition
∴ It's is not a p.m.f.

Q2. Fallowing is the p.m.f of x

$X = 1, 2, 3, 4, 5$

$P(x) = 0.1, 0.15, 0.2, 0.3, 0.25$

Find tt mean and variance of x.

x	$P(x)$	$xP(x)$	$x^2P(x)$
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	0.5	1.25	6.25
		$\Sigma xP(x) = 3.45$	$\Sigma x^2 P(x) = 13.55$

$$Var = V(x) = \Sigma x^2 p(x) - [E(x)]^2$$
$$= 13.55 - 3.45^2$$
$$= 1.6475$$

```
> x=c(1,2,3,4,5)
> prob=c(0.1,0.15,0.2,0.3,0.25)
> a= x* prob
> a
[1] 0.10. 0.30 0.60 1.20 1.25

>b= (x^2)* prob
>b
[1] 0.10 0.60 1.80 4.80 6.25
```

Name : A...

:10

> mean = sum (a)
> mean
[1] 3.45

> mean = sum (b)
> mean
[1] 13.55

> var = sum (b) - mean ^ 2
> var
[1] 1.6475

Q 3. Find mean and varience 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
> a
[1] 0.5 3.0 3.0 5.0 3.75
> mean = sum (a)
> mean
[1] 15.25
> b = (x ^ 2) * prob
> b
[1] 2.50 30.00 45.00 100.00 93.75
> var = sum (b) - mean ^ 2
> var > 38.6875

Scanned with CamScanner

[1]

[1] 0.2   0.5   0.7   0.9   1.0

> plot(x, a, "s")

[1]

$P(x) = 0, \quad x < 0$

$= 0.2, \quad 0 \le x < 2$

$= 0.5, \quad 2 \le x < 4$

$= 0.7, \quad 4 \le x < 6$

$= 0.9, \quad 6 \le x < 8$

$= 1.0, \quad x \ge 8$

> var

[1] 38.6875

Q4)

(i) Find c.d.f of the following p.m.f and draw the graph of c.d.f

```
> x = c(1, 2, 3, 4)
> prob = c(0.4, 0.3, 0.2, 0.1)

> a = cumsum(prob)
> [1] a = 0.4, 0.7, 0.9, 1.0
> plot(x, a, "s")
[1]
```

$P(x) = 0, \quad x < 0$

$= 0.2, \quad 0 \le x < 2$

$= 0.5, \quad 2 \le x < 4$

$= 0.7, \quad 4 \le x < 6$

$= 0.9, \quad 6 \le x < 6$

$= 1.0, \quad x \ge 8$

$p(x) = 0, \quad x < 1$

$= 0.4, \quad 1 \le x < 2$

$= 0.7, \quad 2 \le x < 3$

$= 0.9, \quad 3 \le x < 4$

$= 1, \quad x \ge 4$

ii) Find c.d.f and draw the graph

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)
> px = c(0.2, 0.3, 0.2, 0.2, 0.1)
> a = cumsum(px)
> a
```

:: 0

Practical - 2.

Binomial distribution

1. Suppose there are twelve MCQs in an English Question paper. Each question has five answers and only one is correct. Find a probability of having
   1.) four correct answers.
   11) almost four correct answers
   111) atleast three correct answers

2. Find the complete distribution where $n = 5$ and $p = 0.1$. 7.

3. Find the probability of exactly ten success out of 100 trials with $p = 0.1$.

4. X follow binomial distribution with $n = 12$ and $p = 0.25$. Find
   1.) $P(X \leq 5)$
   11) $P(X > 7)$
   111) $P(5 < X < 7)$

5. There are 10 members in a commity. Probability of any member attending a meeting is 0.7. What is the

seven and more member present in a meeting.

6. A salesman has a twenty person probability of making a sale to a customer. On a typical day he will meet 30 customers what minimum number of sales he will make with 88% probability.

7. For $n=10$ and $p=0.6$. Find the binomial probability and plot the graph of p.m.f and c.d.f.

Note:-

1) $P(X=x) = dbinom(x, n, p)$

2) $P(X \le x) =$ (probability of atmost value)
$= pbinom(x, n, p)$

3) $P(X > x) \rightarrow$ (probability of atleast)
$= 1 - pbinom(x, n, p)$

4) If x is unknow and probability is given as qbinom (p, n, p)

Solution:-

1] a) n=12
p = 1/5
~~p(x=4)~~.
dbinom (4, 12, 1/5)
[1] 0.1328756

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

iii) 1- pbinom (2, 12, 1/5)
[1] 0.4416543.

2] > n = 5
> p = 0.1
> dbinom (1, n, p)
[1] 0.32805
> dbinom (3, n, p)
[1] 0.0081
> dbinom (5, n, p)
[1] 1e-05
> dbinom (0, n, p)
[1] 0.59049

3] > x = 10
> n = 100
> p = 0.1
> pbinom (10, 100, 0.1)
[1] 0.5831555
> dbinom (10, 100, 0.1)
[1] 0.1318653

4] > n = 12
> p = 0.25
> pbinom (5, n, p)
[1] 0.9455978
> 1 - pbinom (7, n, p)
[1] 0.00278151
> dbinom (6, n, p)
[1] 0.4014945

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

7.] >n=10
>p=0.6
>x=0:n
>bp=dbinom (x, n, p)
>bp
[1] 0.0001048576 0.0015728640 0.0106168320
0.0424673280 0.1114767360 0.2508226560 0.2149908480
[6] 0.2006581248 0.0403107840
0.1209323520
[11] 0.0060466176.

>d=data.frame("x.values=x", "probability"=bp)
>print (d)

	x.x.value.X.	probability
(1)		
1	x.values=X	0.0001048576
2	x.values=X	0.0015728640
3	x.values=X	0.0106168320
4	x.values=X	0.0424673280
5	x.values=X	0.1114767360
6	x.values=X	0.2006581248
7	x.values=X	0.2508226560
8	x.values=X	0.2149908480
9	x.values=X	0.1209323520
10	x.values=X	0.0403107840
11	x.values=X	0.0060466176

It follows binomial distribution parameter.

27.12.19

Scanned with CamScanner

TDO

Practical 4

$P(X=x) = ({}^{n}C_{x}) p^{x} q^{n-x}$

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

(i) $P(x=7) = ({}^{8}C_{7})(0.6)^{7}(0.4)^{0.7}$
$= {}^{8}C_{7} \times 0.2799 * 0.4$
$= 0.02799 \times 0.4$
$= 0.08956.$

(ii) $P(x \leq 3) = P(0) + P(1) + P(2) + P(3)$
$= {}^{8}C_{0}(0.6)^{0}(0.4)^{8} + {}^{8}C_{1}(0.6)^{1}(0.4)^{8-1} + {}^{8}C_{2}(0.6)^{2}.$
$= 1 \times 1 \times 0.000655368 \times 0.6 \times 0.0016384 \times 28 \times 0.36 \times 0.004096 * 56 \times 0.26 \times 0.07204$
$= 0.1736204$

(iii) $P(X = 2 \text{ or } 3) = p(2) + p(3)$
$= {}^{8}C_{2}(0.6)^{2} + 1(0.6)^{2} + {}^{8}C_{3}(0.6)^{3}(0.4)^{3}$
$= 28 \times 0.36 \times 0.004096 + 56 + 0.21670.01024$
$= 0.04128 + 0.12380$
$= 0.16515012.$

Practical - 5

Normal distribution

Normal distribution is example of continuous probability Distribution

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

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

qnorm(p, μ, σ)

(v) To generate a random sample of size n:
rnorm(n, μ, σ).

Q.1] A random variable x follows normal distribution with $\mu = 10$, $\sigma = 2$ find
(i) $P(X \leq 7)$
(ii) $P(x > 12)$
(iii) $P(5 \leq x \leq 12)$
(iv) $P(X < k) = 0.4$

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

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] 0.835135
```
(iv) $p(x < k) = 0.4$
```
> k=qnorm(0.4,10,2)
> k
[1] 9.493306
```

Q2] p1=pnorm(110,100,6)
(i) cat("P(x≤110) is=", p1)
$P(x \leq 110) = 0.9522096$
(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)
$P(x \leq 92) = 0.91211223$
(iv) p4=pnorm(110,100,6)-pnorm(95,100,6)
cat("P(~~x≤ 95~~ ≤x≤110) is=", p4)
$P(95 \leq x \leq 110) = 0.7498813$
(v) k=qnorm(0.9,100,6)
cat("P(x<k)=0.9=", k)
$P(x < k) = 0.9 = 107.6893$

Q] $x \sim N(10, 3)$
Generate 10 random variable 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 = "xvalues", ylab = 'probability', main: 'standard normal curve")

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

→ Solutions :-
P_1 = pnorm(60, 50, 10)
cat ("P(X≤60) is =", P_1)
$P(X \le 60) = 0.8413447$

P_2 = 1 - pnorm(65, 50, 10)
cat ("P(X>65) is =", P_2)
$P(X>65) = 0.668072$.

M

P_3 = pnorm(60, 50, 10) - pnorm(45, 50, 10)
cat ("P(45≤ x<60) is =", P_3)
$P(45 \le X \le 60) = 0.53280$

Practical - 6.

Aim:- Z and T distribution som.

Q.1]. Check the hypothetics

A simple of size = 400. is selected and the sample mean is 20.2 and the standard derivation is 2.25.
Test at 5 person label of significant

→ Ho : μ=20 against H1: μ ≠ 20.

mo = 20; mx = 20.2, sd = 2.25 , n = 400
> Ho = H1
> zcal = (mx - mo) / (sd / sqrt(n)).
> zcal
[1] 1.777778
> cat ("zcalculated is = ", zcal)
> pvalue = 2* (1- pnorm(abs (zcal)))
> pvalue
[1] 0.07547036.

> 0.05
Pvalue is meter than 0.05 we accept
Ho: μ = 0.

80

Q22] We want to test the hypothetic $\mu_0 = 250$, $H_1: \mu \neq 250$. A sample of size 100 has a mean of 275 and standard deviation 0. Test the hypothetic at 5% level of significant.

→ $\mu_0 = \mu_1$

→ > no = 250; nx = 275; sd = 30; n = 100; ~~zcal~~

```
> zcal= (nx-no)/(sd/sqrt(n))
> zcal
[1] 8.33333
> cat("z calculated is =",zcal)
  z calculated is = 8.333333
> pvalue= 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.
```

∴ Since, pvalue is less than 0.05, we reject Ho.

Q3] We want to test the hypothesis ~~Ho~~ Ho: P = 0.2 against H1: P ≠ 0.2, P = population proportion. As sample of 400 is selected the sample proportion is called 0.125. Test the hypothesis at 1% level of significant.

Ho!

→ Ho ~~H1 = H2~~:

```
→ P = 0.2
> Q = 1-P
> P = 0.2
> p = 0.125
> n = 100
> P = 0.2
> n = 400
> zcal = (p-P)/(sqrt(P*Q/n))
> zcal
> [1] -3.75
```

p.value >.

∴ pvalue is less than 0.05 we reject ~~Ho~~.

Arc

Q.4] In a weak city 325 men out of ~~out of~~ 600 men were found to be self employee. Thus this information support the conclusion that exactly half of the men in the city is self. employee.

⇒ $H_0 = H_1$
P = 0.5 : n = 600 ; p = 325/600

~~> p = 325/600~~
> p
~~[1] 0.5416667~~

~~p = 0.~~ -> 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

> zcal=(p-P)/(~~sqrt(P*Q/n)~~)
> zcal
~~[1] 1.613743~~

Q5] Test of hypothesis Ho: μ=50 against H1: μ≠50
A sample of 30 collector - sample=50,49,52,44,45,48,46.45,49,45,40,47,55,54,46,58,47,44,59,60,61,41,52,44,55,56,46,45,48,49.

> ~~Ho=H1~~ Ho H1=H2
> mo=50
> x=c(50,49,52,44,45,48,46,45,49,45,40,47,55,54,46,58,47,44,59,60,61,41,52,44,55,56,46,45,48,49).
> n=length(x)
> n
[1] 30
> mx=mean(x)
> mx
[1] 49.33333
> varience=(n-1)*var(x)/n.
> varience
[1] 30.95556
> sd=sqrt(varience)
> sd
[1] 5.63772
> zcal=(mx-mo)/sd/(sqrt(n))
> zcal
[1] -6.584841.

∴ Since, pvalue is more than 0.05 we a the Ho.

Practical 7

Topic :- Large sample test

Q1. Two random sample of single 1000 and 2000 are drawn from two population unit a standard duration 2 and 3 respectively. Test the hypothesis that two population means equal or not at 5% level of significance. Sample means are 67 and 68 respectively.

→ H_0 $\mu_1 = \mu_2$ against H_1: $\mu_2 \neq \mu_2$.
μ_1 = Population 1st mean
μ_2 = population 2nd mean
mx_1 = mean of sample first
mx_2 = mean of sample second.
sd_1 = standard duration 1st
sd_2 = standard duration 2nd.

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

Q. A study of noise level in two hospital is done follow following data is calculated First sample mean 051 61.2 first standard deviation = 7.9, second sample size = 34 second sample mean = 59.4, second standard deviation = 7.8, Test H_0 $\mu_1 = \mu_2$ at 1% level of significance.

```
> H0 - μ1 = μ2 against H1 μ1 ≠ μ2
> n1 = 84
> n2 = 34
> mx1 = 61.2
> mx2 = 59.4
> sd1 = 7.4
> sd2 = 7.8
> zcal = (mx1 - mx2) / sqrt((sd1^2/n1) + (sd^2/n2))
> zcal
  131.1311
> cat ("z calculated is =", zcal)
  z calculated is = 1.13117
> pvalue = 2* (1-pnorm(abs zcal)))
> pvalue
  [1] 0.258006
```

Since pvalue is greater than 0.01, we accept the $u_1 \neq u_2$

Q.3) From each of two poplution of orange. The following sample are collected. Test whether porortion of Bad orange are equal or not
First sample size = 250 Second sample size = 200
No. of Bad orange in the first sample = 44 and second sample 30.

H_0: $P1 = P2$ as H_1: $P1 \neq P2$.

```
→ > n1 = 250
  > n2 = 200
  > p1 = 44/250.
```

```
>p2 = 30/80
>p = (n1*p1+n2*p2)/(n1+n2)
>q = 1-p
>zcal = (p1-p2)/(sqrt(p*q*(1/n1+1/n2)))
>zcal
[1] 0.739358
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.4546896
```

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

Q) Random sample of 400 males and 600 female were asked whether they want a ATM nearby, 200 males and 290 female were favour of proposal. Test the hypothesis the proportion of males and females favouring the proposal are equal or not at 5% (LOS)

```
> H0: ...
>n1 = 400
>n2 = 600
> p1 = 200/400
> p2 = 390/600
> p = (n1*p1+n2*p2)/(n1+n2)
> Q = 1-p
>zcal = (p1-p2)/(sqrt(p*Q*(1/n1+1/n2)))
>zcal
[1] 4.772475
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.30391e-06
```


Q5) The following are the two independent sample from the two population test equality of two population mean at 5% LOS

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

Sample 2:- 72, 76, 74, 70, 70, 78, 70, 72, 75, 79, 74, 75, 78, 72, 74, 80.

Soln :-

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

```
>x1 = c(79,77,     ,76,76)
>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(72,76,     ,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)
tcal
[1] 1.5751

>cat("t calculated is : ", tcal)
t calculated is : 1.575)

t test(x1,x2) = 1.57
pvalue = 0.1387.

AM

Practical-8

Topic: Small sample test

Q.1]. A random sample of 15 observation is given by 80,100,110,105,122,170,120,110,101,88 83,95,89,107,125.
Due this data support the assumption i.e popular mean = 100.

→ Ho: μ=100
x=c(80,100,110,105,122,70,120,110,101,88,83,
89,107,125)
t.test(x)
One sample t test

data: x
t= 24.029 df=14 p-value=8.819e
alternative hypothesis: true mean is not equal too as
confidence interval:
91.37475 107.28892
Sample estimates:
mean of x : 100.3333
Since, p.value is less than 0.05 we reject.

Q.2] >g1= c(18,22,21,17,20,17,23,20,22,21)
> g2= c(16,20,14,21,20,18,13,15,17,21)
> t.test(g1,g2).
data: g1 and g2
t= 2.2573 df= 16.376, p.value=0.03798
alternative hypothesis: true difference in mean nor equ
as percent confidence interval
0.1628205 5.0371795
Sample estimate:
mean of x mean of y
20.1 17.5

Since, p value is greater than ~~0~~ level of significance
we accept the value at 1% H0

Q3] Two type of medicine are used on two or seven patience for reducing their weight. The decreasing in the weight abo after using the medicines given below.

med A = 10, 12, 13, 11, 14
med B = 8, 9, 12, 14, 15, 10, 9

Is there a significance distance in the efficiency of the medicines.

H0 $\mu_1 = \mu_2$

```
> x = c(10,12,13,11,14)
> y = c(8,9,12,14,15,10,9)
> t-test(x, y)
```

data: x and y.

z = 0.80384, df = 9.7594, p value = 0.4406

alternatives hypothesis: true difference in means is not equal to 95 percent confidence interval:

-1.781171 3.781171

Sample estimates.

mean of x mean of y

12 11

Since, pvalue is more than 0.05 we accept h0. w1 = 20 at 1% LOS.

Q4] The weight reducing diet program is conducted and the observation is noted for ~~test~~ 10 participate 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

H0 there is no significant difference in weight.
H1 the diet program reduce weight.

```
> x = c(120, 125, - - - , 118)
> y = c(111, 114, . . . , 112)
> t.test(x, y, paired = T, alternative = "less")
```

data: x and y.

t = 17, df = 9, pvalue = 1

alternative hypothesis: true difference in means is less than 0 as percent confidence interval:

Inf = 9.416556.

sample estimates

mean of the difference

8.5

Q5] 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 as H0.

```
> x = c(66, 67, . . . 90, 92)
> y = c(64, 66, . . . . 95, 97)
> t-test(x, y)
```

data: x and y
t = -0.63938, df = 17.974, pvalue = 0.5304
alternative hypothesis. True difference in means is
mean equal to 0 as percentage confidence
interval:
-12.853992 6.853992
Sample estimate:-
mean of x mean of y
80 83
Since, pvalue is greater than 0.05 we accept x
a 41% level of significance.

Q6) The following are the make before and after obtai
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
Here H_0: No significance difference H_1: increase in mean
> x = c(71,72, . . . 75)
> y = c(74,77, . . . 78)
> t.test(x,y, paired=T, alternative="greater")
data: x and y
t = -4.4691, df = 9, pvalue = 0.9972
alternative hypothesis: true difference in mean is
greater than 0 as percent confidence interval:
-5.076637 Inf
Sample estimates:-
mean of the difference
-3.6.

M

1st Aim:- Large and Small sample space
The arithmatic mean of a sample of 100 item from a
against an alternative it is more than 55 at 5%
LOS (Level of significant)
large population is 52. It is standard deviation
is 7. Test the hypothesis that the

2. In a big city 350 outoff 700 males are found to be
smokers. Thus, this information support that exactly
half of the males in the city are smokers? Test
at 1% LOS

3. Thousand article from a factory A are found to
have 21 defectives. 500 article from a second
factory B are found to have 11 defective. Test at
5% LOS That the two factory are similar or not

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

5. The flowers steams are selected and the heights a
found to be centimeter (cm) 63, 63, 68, 69, 71,
72. Test the hypothetic that the mean height i
66 or not at 1% LOS

6. Two random ~~two~~ sample where drawn from t
normal population and their value are A = 66, 6
75, 76, 82, 84, 88, 90, 92.
B - 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97
Test whether the population have a same
variance at 5% LOS.

020

Solution:-

$H_0 \cdot \mu_1 = \mu_2$ against

1st

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

p-value is < 0.05 we ~~a~~ rejected

2. $H_0 \cdot \mu_1 = \mu_2$ against H_0 $\mu_1 \neq \mu_2$

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

pvalue > 0.05 we accept

3. $H_0 \cdot \mu_1 = \mu_2$

```
>n1=1000
>n2=1500
>p1=20|1000
>p2=(n1*p1+n2*p2)|(1/n1+1/n2)
>p
[1] 0.014
>zcal=(p1-p2)|sqrt(p*q*(1/n1+1/n2)
>zcal
[1] 2.082731
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.03727577
```

Since, pvalue < 0.05
we reject pvalue

Name

4B. H_0 $\mu_1 = \mu_2$

```
>n=400
>mx=99
>mo=100
>variance=64
>sd = sqrt(variance)
>sd
[1] 8
>zcal=(mx-mo)/(sd/sqrt(n))
>zcal
[1] -2.5
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.01241933
```

Since, pvalue < 0.05
we reject ~~H_0 $\mu_1 = \mu_2$~~.

5 H_0 $\mu_1 = \mu_2$

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

One Sample t test

data = x

t = 47.54, df = 6, p value = 5.522e-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

p value is < 0.05 we reject

(6) H_1 $\mu_1 = \mu_2$ against H_0 $P_1 = P_2$

```
> x=c(66,67, ......,90,92)
> y= c(64,66,       ,95,97)
> F.var.test(x,y)
> F
```

F test to compare two variance

data: x and y

F = 0.70686, num df = 8, denom df = 10, p value = 0.6357

alternative hypothesis: true ratio of variance is not equal to 1

95 percent confidence interval.
0.1833612 3.0360393

Sample estimates
ratio of variance
0.706867

H_1 $\mu_1 = H_0$ against H_0 $P_1 = P_2$

p value > 0.05 we accept H_0 $\mu_1 = \mu_2$

ANOVA & CHI-SQ...

Q1. Use the following data to test whether the cleanliness of home and cleanliness of child is independent or not.

		CH	
		Clean	Dirty
CC	Clean	70	50
	fairly clean	80	20
	Dirty	35	45

Soln: H_0: CC & CH.

```
> 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, p-value is less than 0.05, we reject. Hence, child cleaniness and Home cleaniness are dependent

Q2] Use the following data to find a vacination and a particular disease are independence or not.

		Aff	Not Aff
Vac	Given	20	30
	Not Given	25	35

Soln: H_0: Vac & Dis

```
> 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 Yatas' continuity correction

data= y

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

Since, p value is more than 0.05, we acce Hence, ~~child cleanliness~~ the vaccine and disease independence or each other

Q3 Perform ~~a~~ ANOVA for the following data

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

Soln - H_0: The mean of the varieties are equal

```
> 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 = 3.326, nom df = 3, denom df = 9, p-value = 0.0183

> anova = aov(values~ind, data=d)
> anova
Call:
   aov(formula = values ~ ind, data = d)
Terms
                  ind        Residuals
Sum of squares    79.26923   71.50000062
Deg. of freedom   3          9
Residual standard error: 1.420746
Estimated effects may be unbalanced.
```

Q4) The following data gives life of tired of four ~~balanced~~ branded

Type	Observation
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 hypothetic that the average life of four branded is same.

H_0: The average life of foure branded is 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"
> anova = aov(values~ind, data=d)
> anova
Call:
   aov(formula = values ~ ind, data = d)
Terms:
                  ind        Residuals
Sum of squares    91.4381    89.0619
Deg of freedom    3          20
Residual standard error: 2.110231
Estimated effects may be unbalanced.
```

```
> 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=0.0023
```

Q. One thousand student of a college graded according to the IQ and economic conditional of their home. Check that is there any association between IQ and the economic condition of the home.

		I.Q High	Low
E.C	Hight	460	140
	Pred.	330	200
	low	240	160.

Soln:- Ho: IQ & E.C.

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



```
        Pearson's Chi-squared test
data: y
X-squared= 39.726, df=2, p-value=2.364e-07
```

Since, p-value is less than 0.5, it rejected

An (7.?)

Practical 11

... in 20 days Apply sign test to test the hypothetic the population median is 21.5

17, 15, 50, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 28, 24, 26

Sol

Ho: non median is 21.5

```
> x = c(17,15,20,29,              )
> median = 21.5
> sp = length(x[x>me])
> sn = length(x[x<me])
> n = sp+sn
> n
[1] 20
> pv = pbinom(sp,n,0.5)
> pv
[1] 0.4119015
```

Since, p value is more than ~~0.5~~ 0.05 we accept

Q2] Following the ten question
612, 619, 631, 628, 643, 640, 655, 649, 670, 663

Apply sign test to test the hypothetic that the ~~prob~~ population and median is 625 against the alternative it is greater than 625 at 1% (LOS).

Note: If the alternative is greater then
pv = pbinom(sn,n,0.5)

Ho: non median is 625

```
> x = c( 612, 619, 631, 628, 643,          )
> me = 625
> sp = length(x[x>me])
> sn = length(x[x<me])
> n = sp+sn
> n
[1] 10
> pv = pbinom(sn,n,0.5)
> pv
[1] 0.0546875
```

Q3] Following are the observation
36, 32, 21, 30, 24, 25, 90, 22, 20, 18

Using sign test test the hypothetic that the popular median is 25 against the alternative is less than 25 at 1% LOS.

Ho: non median is 25.

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

Q4 Following are the 12 question
63, 65, 60, 89, 61, 71, 58, 51, 61, 63, 39, 72, 65

Using wilcoxon sign rank test. Test the hypothesis that the population median is 60 against the alternative it is greater than 60 at 5% Los

H0: population median is 60.

> x = c(63, 65, 60, 89, 61, 71, 58, 51, ...)

> wilcox.test(x, alt= "greater", mu= 60)

wilcoxon signed rank test with continuity correction

data: x

V = 56.5, p value = 0.09092

alternative hypothesis true location is greater than 60

Q.5 Following are the observation

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

Use the wilcon sign

> x = c(15, 17, 24, 25, 20, 21, 32, 328, 12, 25, 24, 26

Use the wilcoxon signed rank test

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

> willcox.test(x, alte= "less"), mu= 20)

data x

V = 48.5, p-value: 0.9232

alternative hypothesis: true location is less than 20

Q.6. Following are the five question.
20, 25, 27, 30, 18

Test the hypothetics that the population median is 25 against the hypothetic not the 25?

→> x = c(20, 25, 27, 30, 18)

> wilcox.test(x, alt = "two sided", mu = 25)

wilcon signed. Rank test with continuity correction

data:-

V = 3.5, p-value = 0.7127

alternative hypothesis: true location is not equal to 25.

PM
9.3.2