PRACTICAL No. 1

Aim: Basics of R software.

1) R is a software for stastical analysis and data computing.

2) It is an effective data handling software and outcome storage is possible.

3) It is capable of graphical display.
4) It is a free software.

Q1) Solve the following

1) $4+6+8\div2-5$
> 4+6+8/2 -5
[1] 9

2) $2^2+|-3|+\sqrt{45}$
> 2^2 + abs(-3) + sqrt(45)
[1] 13.7082

3) $5^3+7\times5\times8+46/5$
> 5^3 +7 * 5 * 8 + 46/5
[1] 414.2

4) $\sqrt{4^2+5\times3+7/6}$
> sqrt(4^2 + 5 * 3 + 7/6)
[1] 5.671567

```
5) roundoff
46 ÷ 7+9*8
> round(46/7+9*8)
[1] 79

Q.2) > c(2,3,5,7)*2
[1] 4 6 10 14

> c(2,3,5,7)*c(2,3)
[1] 4 9 10 21

> c(2,3,5,7)*c(2,3,6,2)
[1] 4 9 30 14

> c(1,6,2,3)*c(-2,-3,-4,-1)
[1] -2 -18 -8 -3

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

> c(4,6,8,9,4,5)^c(1,2,3)
[1] 4 36 512 916 2125

> c(6,2,7,9)/c(4,5)
[1] 1.50 0.40 1.75 1.00

Q.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
[1] 1300

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

Q5) find: 8x + y and 2x + 3y where $x = \begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 3 \end{bmatrix}$

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

```
> x <- matrix(nrows=3, ncol=3, data=c(4,7,9,-2,0,-5,6,7,3))
[1] x   [,1]  [,2]  [,3]
[1,]     4    -2     6
[2,]     7     0     7
[3,]     9    -5     3
> y <- matrix(nrows=3, ncol=3, data=c(10,12,15,-5,-4,-6,7,9,5))
[1] y   [,1]  [,2]  [,3]
[1,]    10    -5     7
[2,]    12    -4     9
[3,]    15    -6     5
> x+y
     [,1]  [,2]  [,3]
[1,]  14    -2    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.6) Marks of stastics of CS Batch B.

x = c(60, 20, 35, 24, 46, 56, 55, 45, 27, 22,
47, 58, 54, 40, 50, 32, 36, 29, 35, 39)

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

	c	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

G

F.E.

PRACTICAL No. 2

Topic : Probability distribution.

1) Check whether the followings 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 $\Sigma P(x) = 1$

$\therefore P(0) + P(1) + P(2) + P(3) + P(4) + P(5) \neq P(x)$

$= 0.1 + 0.2 - 0.5 + 0.4 + 0.3 + 0.5$

$= 1.0$

$\because p(2) = -0.5$, it cant be a probability mass function

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

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 p.m.f is

i) $P(x) \geq 0 \ \forall x$ satisfy

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

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

$= 1$

$\therefore$ The given data is pmf.

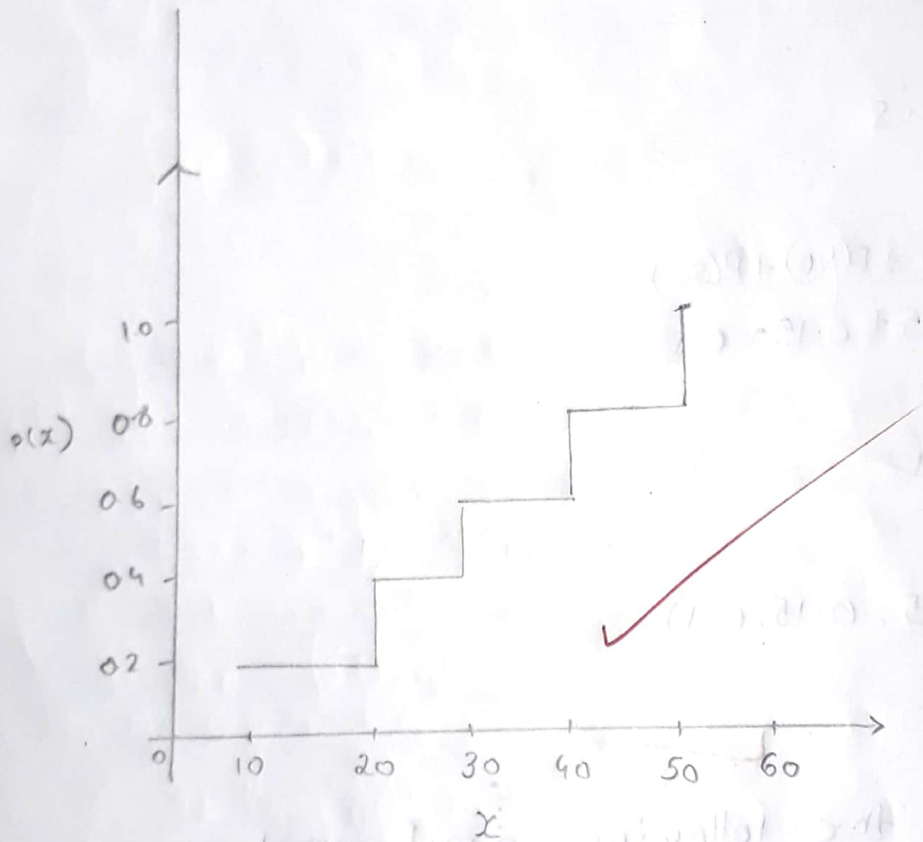
Code:

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

Q.2) 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

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



> x = c(10, 20, 30, 40, 50)

> Plot(x, cumsum(prob), "s")

Q.2) Find

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

F(x) = 0   x < 1

0.15   1 ≤ x < 2

0.40   2 ≤ x < 3

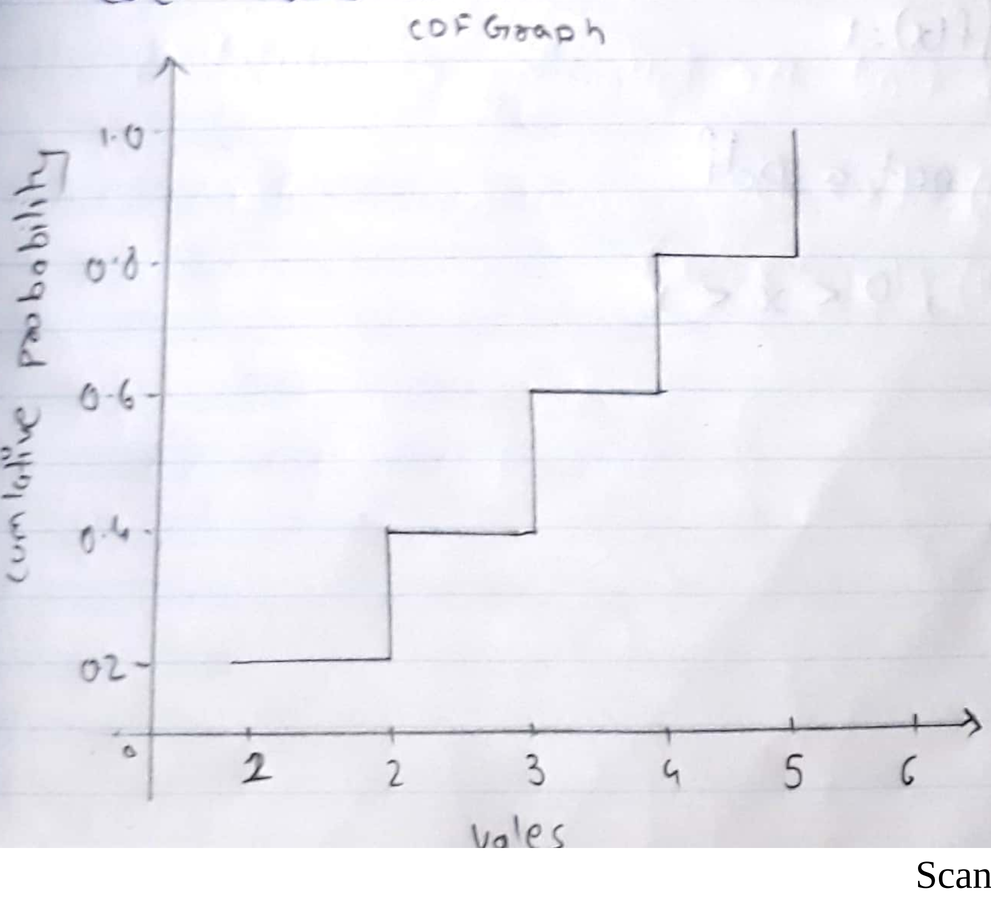
0.50   3 ≤ x < 4

0.70   4 ≤ x < 5

0.90   5 ≤ x < 6

1.00   x ≥ 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 = "Cumlative probability", main = "CDF graph
    col = " brown")
```



Q.3) check that whether the following is p.d.f or not

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

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

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

→ $\int_0^1 f(x)$

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

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

$= \left[3x - x^2\right]_0^1 = 2$

$\because$ The $\int_0^1 f(x) = 1$

$\therefore$ It is not a pdf

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

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

$\because x^n = \frac{x^{n+1}}{n+1}$

$= x^3$

$= 1$

The $\int_0^1 f(x) = 1$

$\therefore$ It is a pdf

6

PRACTICAL NO-

Topic : Binomial distribution

$P(X=x) =$ dbinon (x, n, p)

$P(x<x) =$ Pbinom (x, n, p)

$P(x>x) = 1 -$ Pbinom (x, n, p)

if x is known

$P_1 = P(X \leq x) =$ qbinom (P_1, n, p)

1) find the brobability of exactly 10 Success in hundred trials with $P = 0.1$

2) Suppose there are 12 mcq, each question has 5 options out of which I is correct. Find the probability of having exactly 4 correct answers.

ii) Atmost 4 correct answers

iii) More than 5 correct answers.

3) Find the complete distribution when $n = 5$ & $p = 0.1$

4) $n = 12$, $p = 0.25$, find the following probabilities

i) $P(X=5)$ iii) $P(X>7)$

ii) $P(X \leq 5)$ iv) $P(5<X<7)$

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

2) >dbinom(4, 12, 0.2)
[1] 0.1328756
>pbinom (4, 12, 0.2)
[1] 0.927445
>1- pbinom (5, 15, 0.2)
[1] 0.01940528

3) dbinom (0:5, 5 ,0.1)
0 - 0.59049
1 - 0.32805
2 - 0.07290
3 - 0.00810
4 - 0.00045
5 - 0.00001

4) 1) dbinom (5, 12, 0.25)
[1] 0.1032414
2) pbinom (5, 12, 0.25)
[1] 0.9455978
3) 1-pbinom (7, 12, 0.25)
[1] 0.0027815|
4) dbinom (6, 12, 0.25)
[1] 0.64014945

5) The probability of a Salesman making a sale to customer 0.15 . find the probability of
1) No sales out of 10 customer.
2) More than 3 sales out of 20 customer

6) A Sales man has 20% probability of making a sales to customer out of 30 customers 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

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

qbinom(0.88, 30,0.2)
[1] 9
>n=10
>p=0.3
>x=0:n
>prob= dbinom(x,n,p)
>cumprob= pbinom(x,n,p)
>d= data.frame("xvalues", =x, "probability"= prob)
>print d
```

	xvalues	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	0.0368
8	7	0.0367
9	8	0.0890
10	9	0.0001
11	10	0.0000

PRACTICAL - 4

Aim: Normal distribution.

$p(X=x) = dnorm(n, \mu, \sigma)$

$p(X \leq x) = pnorm(x, \mu, \sigma)$

$p(X > x) = 1 - pnorm(x, \mu, \sigma)$

To generate random number from a normal distribution (in Random numbers)

The r code is $rnorm(n, \mu, \sigma)$

1) A random variable X follows normal distribution with mean $= \mu = 12$ & s.d. $= \sigma = 3$. Find

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

(ii) $p(10 \leq x \leq 13)$

(iii) $p(X > 14)$

(iv) Generate 5 observation (RANDOM NUMBERS)

Ans

(1) P1 = pnorm(15,12,3)

```
>P1
[1] 0.841344 7
>cat ("p (X<=15) = ", p1)
>P(X<=15) : 0.8413447
```

2) P2 = pnorm(13,12,3) - pnorm(10,12,3)

```
>P2
[1] 0.3780661
```

>p(10<=x<=13)= 0.3780661

iii) p3 =1- pnorm(14, 12, 3)
>p3
[1] 0.2524925
>cat("p(x>14) =", p3)
p(x>14): 0.2524925

>rnorm(5, 12,3)
[1] 14.98663 12.51616 13.44904 14.98075 19.73518

2) x follows normal distribution with
$\mu = 10$, $\sigma = 2$, find:?
(i) p(x≤7)
2) p(5<x<12)
3) p(x>12)
4) Generate 10 random observation
v) Find k such that probability p(x<k) = 0.4.

Ans i) p1 = pnorm(7, 10, 2)
p1
[1] 0.0668072
>cat("p(x<=7)=", p1)
[1] p(x<=7)= 0.0668072

ii) p2= pnorm (12,10, 2) - pnorm (5,10,2)
p2
[1] 0.835135
>cat ("p(15<x<12=", p2)
- p(15<x<12)= 0.8351357

iii) p3 = 1- pnorm (12,10,2)
p3
[1] 0.1586553
>cat ("p(p(x>12) = ", p3)
- p(p(x>12)) = 0.1586553

> rnorm (10,10,2)

[1] 11.986324 12.155504 7.130492
10.519359 5.637988 10.959963
7.988735 9.336612 8.165376
13.052133

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

3) Generate 5 random number from a normal distribution with mean = 15 & s.d 4
Find sample mean, median, s.d and print (cat)

4) x~N(30,100), (σ^2, σ=10) find only P1, P2

(no cat command) 41

Ans

>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

5) Ans >x=rnorm(5,15,4)

>x

[1] 16.43602 12.58878 16.91096 17.37383 15.04240

>am=mean(x)

>am

[1]15.67034

>med=median(x)

>me

[1]16.43602

>varience=(n-1)* var(x)/n

>varience

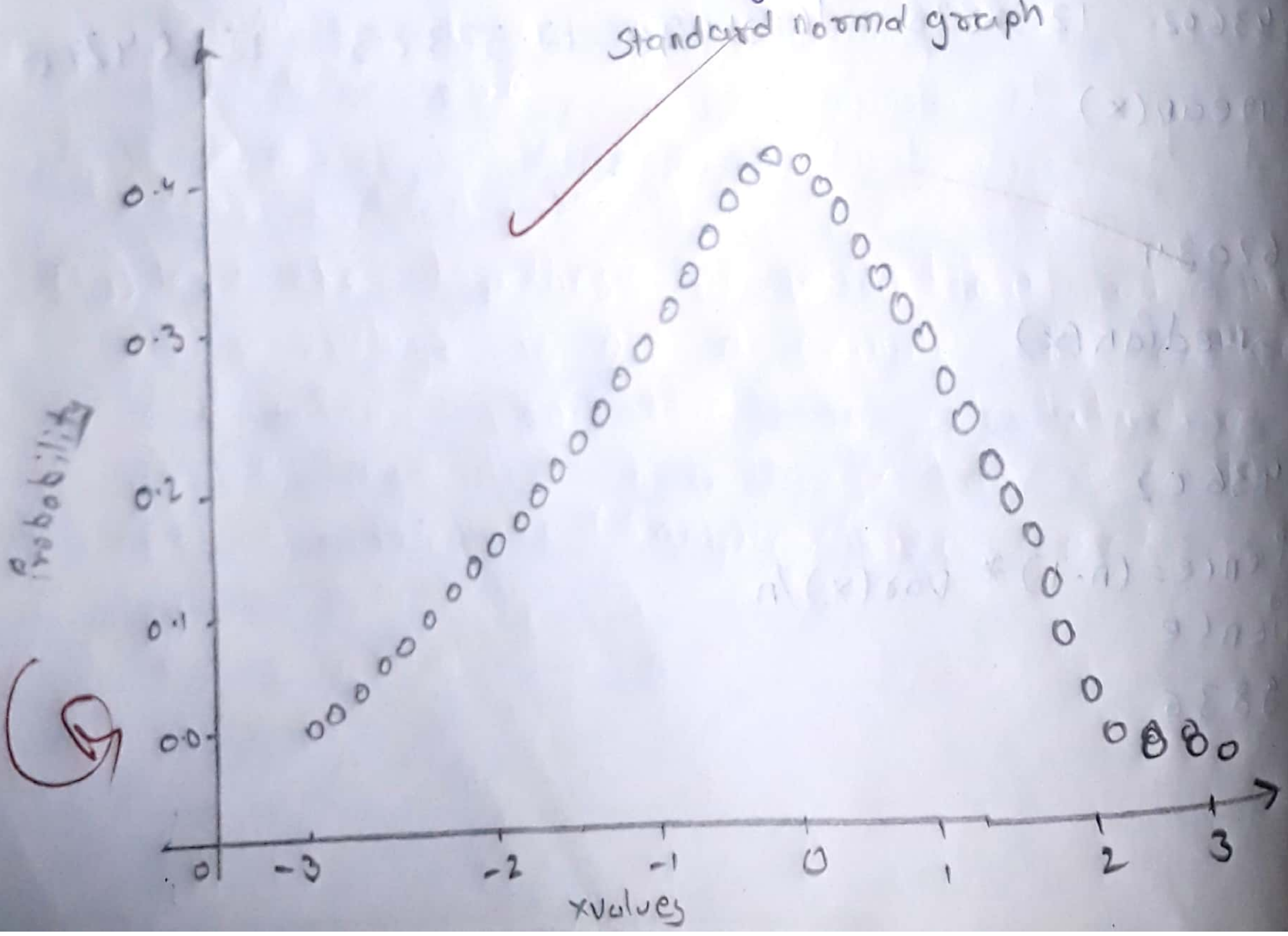
[1] 2.9836

```
>sd= sqrt (variance)
>sd
[1] 1.72731
>cat("sample mean is = ", am)
→ Sample mean is 15.67034
>cat(" sample mean is = ", me)
→ Sample mean is 16.43602
>cat("Sample sd is = ", sd)
→ Sample sd is 1.72731
```

5) Plot the standard, normal graph

Ans

```
x=seq(-3,3,by=0.1)
y<-dnorm(x)
plot(x,y, xlab="xvalues", ylab="probability",
main="standard normal graph")
```



PRACTICAL - 5

Aim: Normal & Z-test

Q1) Test of Hypothesis $H_0 : \mu = 15$, $H_1 : \mu \neq 15$.
Random sample of size 400 is drawn and it is calculated the sample mean is 14. & the standard deviation is 3. Test the hypothesis at 5% level of significance.

m -> mean o -> populatn
σ -> s.d n -> sample size

Soln

```
>mo = 15
>mx = 14
>sd = 3
>n = 400
>zcal = (mx - mo) / (sd / (sqrt (n)))
>zcal
[1] -6.66667
>cat("calculate value at Z is = ,zcal)
-> calculate vale at Z is = -6.666667
>pvalue = 2* (1- pnorm(abs(zcal)))
>pvalue
[1] 2.616796e-11
```

Since p value is less than 0.05 we reject $H_0 . \mu = 15$.

Q.2) Test of hypothesis $H_0: \mu = 10$, $H1: \mu \neq 10$
Random sample of size 400 is drawn wi
sample mean 10.2 & standard deviation
2.25. Test the hypothesis at 5% lev
of significance.

Soln:
```
> mo = 10
> mx = 10.2
> sd = 2.25
> n = 400
> zcal = (mx - mo) / (sd / (sqrt (n)))
> zcal
[1] 1.77778
> cat("calculate value at z is = ", zcal)
-> Calculate value at z is = 1.77778
> pvalue = 2* (1- pnorm (abs (zcal)))
> pvalue
[1] 0.07544036
```

Q.3) Test the hypothesis no porportion of Smokers
in a college is 0.2 a sample is collected
and sample proportion is calculated as
0.125. Test the hypothesis at 5% level
of significance (sample size = 400)
P - population
p - sample

Scanned by CamScanner

Sol: >P=0.2

>p=0.125

>n=400

>Q=1-P

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

>zcal

[1] -3.75

Q) last year farmer's lost 20% of their crops. a random sample of 60 fields are collected. It is found that 9 field crops are insect populated. Test the hypothesis at 1% level of significance.

Sol: >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 value is accepted.

SP

a) Test the hypothesis Ho: U=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
>mx=mean(x)
>mx
[1]12.107
>varience=(n-1)*var(x)/n
>varience
[1] 0.019521
>sd=sqrt(varience)
>sd
[1]0.1397176
>mo =12.5
>t=(mx-mo)/(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

Q1) Let the population mean (the amount spent by customer in a restaurant is 250. A sample of 100 customer is selected. Sample mean is calculated as 275 and SD as 30. Test the hypothesis that population mean is 250 or not at 5% level of significance

Q2) In a random sample of 1000 students it is found that 750 use blue pen. Test the hypothesis that population proportion is 0.8 at 1% level of significance

Q1)

→ $H_0: \mu = 275$ against $H_1: \mu \neq 275$

```
> m0 = 250
> mx = 275
> n = 100
> sd = 30
> zcal = (mx - m0)/(sd/sqrt(n))
> zcal
[1] 8.333333
```

> pvalue= 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 0
> cat("zcal:", zcal)
> zcal: 8.33333
> zcal("pvalue:", pvalue)
> pvalue: 0
∴ pvalue = 0 < 0.05 we reject Ho at 5 % level of significance.

Q.2) Ho : p = 0.8 against H1 : P ≠ 0.8
> P = 0.8
> Q = 1-P
> p = 750/1000
> n = 1000
> zcal = (p-P)/(sqrt(P*Q/n))
> zcal
[1] -3.952847
> pvalue = 2* (1- pnorm(abs(zcal)))
> pvalue
[1] 7.72268e-05
∴ The value is less 0.01, we reject !

Q3) Random sample of size 1000 & 2000 are drawn from two population with same S.d = 2.5. The sample means are 67.5 & 68 resp.. Test the hypothesis

$H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ at 5%
I population

level of Significance.

44

Q4) The study of noise level in two hospital's is given below. Test the claim that the two hospital have same level of noise at one (1%) level of significance

	Hospital (A)	Hospital (B)
Size	84	34
Mean	61.2	59.4
S.D	7.9	7.5

Q5) In A sample of 600 student in a college 400 use blue ink in another college from the sample of 900 students 450 use blue ink. Test the hypothesis that the proportion of students using blue ink in two college are equal or not. At 1% level of significance

Q.3) → $H_0: \mu_1 = \mu_2$ as $H_1: \mu_1 \neq \mu_2$

```
>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
>zcal("zcal:", zcal)
zcal:- 5.163978
>pvalue= 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.417564e-07
>cat("pvalue:", pvalue)
>pvalue: 2.41764e-07
```

∴ Rejected.

Q.4)
```
>n1=84
>n2=34
>mx1=61.2
>mx2=59.4
>sd1= 7.9
>sd2= 7.5
>zcal=(mx1-mx2)/sqrt((sd1^2/n1)+(sd^2/n2))
>zcal
[1]=1.162528
>cat("zcal:", zcal)
zcal:1.162528
```

>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.2450211
∴ The value is greater than 0.01, we accept the value

Q5) >n1=600
>n2=900
>p1=400/600
>p2=450/900
>zcal=(p1-p2)/sqrt(p*q(1/n1+1/n2))
>zcal
[1] 6.381534
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.753222e-10
∴ The value is less than 0.01 the value is rejected

Q6) First sample size H0: p1=p2 as H1: p1≠p2, n1=200, n2=200
P1=44/200, P2=30/200
→ >n1=200, >n2=200, >P1=44/200 >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] 1.802741
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.07142888
∴ Accepted greater than 0.05

PRACTICAL No. 7

Topic: Small Sample test

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

$H_0 : \mu = 66$

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

\>t.test(x)

One Sample t-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

means of x

67.9

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

Two group of students scored the following marks. Test the hypothesis that there is no significant difference between 2 grps

GR1 - 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
GR2 - 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Ho: There is no difference b/w 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 and y
t = 2.2573 , df = 16.376, pvalue = 0.03798
alternative hypothesis:
    True difference in means is not equal
to 0 95 percent confindence interval:
   0.1628205        5.0371795
sample estimate:
 mean of x   mean of y
    20.1        17.5

> Pvalue = 0.03798
> if (pvalue > 0.05) {cat ("accept ho")}
else {cat ("reject ho")}
  reject Ho.
```

PAIRED T-Test

3) 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 affective 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
```

∵ pvalue is greater than 0.05, we accept the hypothesis at 5% level of significance.

Q) Following are the weight 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: Ho: 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, pvalues = 0.9963
alternative hypothesis is: true difference in means is less than 0.
95 percent confidence interval:
-inf 29.0295
sample estimates:
mean of the differences
19.83333

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

5) 2 medicinis are applied to two groups of patient resp.

GR1 = 10, 12, 13, 11, 14
GR2 : 8, 9, 12, 14, 15, 10, 9
Is there any significance difference b/w 2 medicines
H0 : There is no significance difference
>x = c(grp1)
>y = c(grp 2)
>t.test (x, y)
t = 0.80384 , df = 9.7594, pvalue = 0.4406
alternative hypothesis : true difference in means is not equal to 0
95 percent confidence interval
- 0.9698553   4.2981886
Sample estimates :
Meen of x   mean of y
12.0000   10.33333

∵ pvalue is greater than 0.05 we accept the hypothesis at 5% level of significance

PRATICAL 8.

1) Ho: μ = 55 . H_1 : $\mu \neq 55$

>n=100

>mx = 52

>m0 = 55

>sd = 7

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

>zcal

[1] -4.28714

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

>pvalue

[1]1.82153e-05

we reject

2)>P = 350/700   Ho = P = 1/2

>n = 700

>q = 1-P

>z = (p-P)/(sqrt(p* q/n))

>z

[1] 0

>pv= 2* (1-pnorm(abs(z)))

>pv

[1]=1

Accept the pvalue at Ho: p = 1/2

3) >n1=1000

>n2=1500

> p1=0.02

> p2=0.01

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

> p

[1] 0.014

>q=1-p

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

>z

[1]0.003474737

>pv = 2*(1- pnorm(abs(z)))

>pv

>0.03708364

we reject

4) Ho: μ=100

> mx=99 sd=8

>mo=100 n=400

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

>z

[1] -2.5

>pv= 2*(1-pnorm(abs(z)))

>pv

[1] 0.0124193 3

we reject.

Scanned by CamScanner

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

One sample t.test
t = 47.94 , df = 6 pvalue = 5.522e-09
alternative
64.66579   71.62692
mean of x
68.14286

we reject H_0

b) $H_0 : \mu = 100$
> m0 = 1200
> mx = 1150
> z = (mx - m0) / (sd((sqrt(n)))
> z
[1] -4
> pv = 2 * (1- pnorm(abs(z)))
> pv
[1] 6.334248e - 05

we reject

7) $H_0: \sigma_1 = \sigma_2$
> x = c(66,67, 75, 76, 82, 84, 88, 90, 92)
> y = c(64, 66, 74, 78, 82, 87, 92, 93, 95, 97)
> var.test(x,y)
F test to compare two variances
F = 0.70686, num df = 8, denom df = 10, pvalue = 0.6359
95 percent confidence interval
0.1833662 3.086053
sample estimates
ratio of variances
0.7068567
we accept Ho. $P_1 = P_2$.

8)
> n1 = 200
> n2 = 300
> P1 = 44/200
> P2 = 56/300
> p = (n1 * p1 + n2 * p2) /(n1+n2)
> q = 1-p
> q
[1] 0.9128709
> PV = (2*(1-pnorm(abs(z)))
> Pv
[1] 0.3613104
P
0.2
we accept Ho.

PRACTICAL - 9

Chi-Square test & ANOVA

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

→ condition of home

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

→ Ho: 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(x)
> pv
Person's chisquared test
data: y
X-squared = 25.646, df = 2, pvalue = 2.698e-06
```

∴ Ho is rejected, since pv is less than 0.05.

2) Test the hypothesis the vaccination & disease are independent or not.

Diseases	affected	Not affected
Affected	70	46
Not affected	35	37

Ho: Condition of disease & vaccination are independent

```
> x = c(70, 46, 35, 37)
> m = 2
> n = 2
> y = matrix(x, nrow=m, ncol= n).
> y
      [,1]  [,2]
[1]    70    46
[2]    35    37
> Pv = chisq.test(y)
> Pv
X-squared = 2.6275, df = 1, p v = 0.1545
```

Ho: is accepted. Since, pv is more than 0.05

3) 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

→ 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)
        oneway analysis of means
data. values & ind
F = 11.735, num df = 3, denom df = 9, pvalue =

> anova = aov(values ~ ind, data = d)
Summary(anova)
            df    Sum Sq    Mean Sq    fvalue    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 '
```

4) The following data gives the life of the tire of 4 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

- Ho : The average life of 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)
> oneway.test(values~ind, data=d, var.equal=T)
```

one way analysis of means

data:

f = 6.8445, num df = 3, denom df = 20,

pvalue = 0.002349

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

	df	Sumsq	Mean Sq	fvalue	Pr(>f)
ind	3	91.44	30.479	6.845	0.002357
Residuals	20	89.06	4.453		

> How to import a CSV file in R software

> x = read.csv("C:/users/admin/Desktop/stats.csv")

> print(x)

	stats	maths
1	40	60
2	45	43
3	42	47
4	15	20
5	37	25
6	36	27
7	49	57
8	59	58
9	20	25
10	27	27

> am = mean(x$stats)

> am

[1] 37

> n = length(x$stats)

> sd = sqrt((n - 1)*var(x$stats)/n)

> sd

[1] 12.64911

> am = mean(x$maths)

> ~~am~~

[1] 39.4

> sd = sqrt((n - 1)*var(x$maths)/n)

> sd

[1] 15.2

> cor(x$stats ; x$maths)

[1] 0.830618

9