★★ INDEX ★★

Scanned with CamScanner

Practical No. 1

27/11

Title: BASICS OF R-SOFTWARE

R is a software for data analysis and statistical computing.

This software is used for effective data handling and output storage is possible.

It is capable of graphical display.

It is a free software.

Questions:

Q. (i) $2^2+\sqrt{25}+35$

(ii) $2\times5\times3+62\div5+\sqrt{49}$

(iii) $\sqrt{76+4\times2+9\div5}$

(iv) $42+|-10|+7^2+3\times9$

Solution:

(i) $2^2+\sqrt{25}+35$

Command : 2^2 + sqrt (25) + 35

Output : 44

(ii) $2\times5\times3+62\div5+\sqrt{49}$

Command : 2 * 5 * 3 + 62/5 + sqrt (49)

Output : 49.4

(iii) $\sqrt{76+4\times2+9\div5}$

Command : sqrt (76 + 4 * 2 + 9/5)

Output : 9.262829

P.S

(iv) $42 + |-10| + 7^2 + 3 \times 9$

Command : 42 + abs(-10) + 7 ^ 2 + 3 * 9

Output : 128

(v) $x = 20$; $y = 30$

Find $x+y$, $x^2 + y^2$, $\sqrt{y^3 - x^3}$, $|x-y|$

Command : x + y

Output : 50

Command : x ^ 2 + y ^ 2

Output : 1300

Command : sqrt (y ^ 3 - x ^ 3)

Output : 137.8405

Command : abs(x - y)

Output : 10

(vi) c (2, 3, 4, 5) ^ 2

Output : 4 9 16 25

(vii) c (4, 5, 6, 8) * 3

Output : 12 15 18 27

(viii) c(2, 3, 5, 7) * c (-2, -3, -5, -4)

Output : -4 -9 -25 -28

(ix) c (2, 3, 5, 7) * c (8, 9)

Output : 16 27 40 63

(x) c (2, 3, 5, 7) * c(1, 2, 3)

Output: Warning Message. Obj length is not a multiple of shorter Obj length.

(xi) c(1, 2, 3, 4, 5) ^ c(2, 3, 4)

Output: Warning message.

(xii) c(1, 2, 3, 4, 5, 6) ^ c(2,3,4)

Output: 1 8 81 16 125 1296

(xiii) Find the sum, product, maximum, minimum of the values:

5, 8, 6, 7, 9, 10, 15, 5

Command: x = c(5, 8, 6, 7, 9, 10, 15, 5)

len(x)

sum(x)

prod(x)

max(x)

min(x)

Output: 8

65

11340000

15

5

(xiv) $\begin{bmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{bmatrix}$

Command: x ← matrix (nrow = 4, ncol = 2, data = c(1,2,3,4,5,6,7,8))

Output:

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

(XY) $x = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$ $y = \begin{bmatrix} 2 & 4 & 10 \\ -2 & 8 & -11 \\ 10 & 6 & 12 \end{bmatrix}$

Find $x + y$, $x \times y$, $2x + 3y$

command : x<-matrix(nrow=3, ncol=3, data=c(1,2,3,4,5,6,7,8,9)
y<-matrix(nrow=3, ncol=3, data=c(2,-2,10,4,8,6,10,-11,1

x
~~x**y~~
y

Output :

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 4 & 10 \\ -2 & 8 & -11 \\ 10 & 6 & 12 \end{bmatrix}$$

Command : x+y

Output :

3	8	17
0	13	-3
13	12	21

Command : x*y

Output :

2	16	70
-4	40	-88
30	36	108

Command : 2*x + 3*y

Output :

8	20	44
-2	34	-17
36	30	54

(xvi) command : x = c(2,4,6,1,3,5,7,18,16,14,17,19,3,3,2,
5,0,15,9,14,18,10,12)
length(x)
a = table(x)
transform(a)

Output :

> 23

>

x	Freq
0	1
1	1
2	2
3	3
4	1
5	2
6	1
7	1
9	1
10	1
12	1
14	2
15	1
16	1
17	1
18	2
19	1

command: seq(0, 20, 5)
b = cut(x, breaks, right = FALSE)

Scanned with CamScanner

c = table (b)
transform c

Output :

b	Freq
[0, 5)	8
[5,10)	5
[10,15)	4
[15,20)	6

Ans
2/7/1

Practical No. 2

Title: Problems on P.D.f and C.d.f

Q1. Can the following be p.d.f. ?

(i) $f(x) = \begin{cases} 2-x & , 1 \le x \le 2 \\ 0 & \text{o.w.} \end{cases}$

To prove : $\int f(x)\,dx = 1$

$\therefore \int_1^2 (2-x)\,dx$

$\therefore \int_1^2 2\,dx - \int_1^2 x\,dx$

$\therefore [2x]_1^2 - \left[\frac{x^2}{2}\right]_1^2$

$\therefore (4-2) - (2-0.5)$

$\therefore 0.5$

$\neq 1$

$\therefore f(x)$ is not a p.d.f.

(ii) $f(x) = \begin{cases} 3x^2 & , 0 < x < 1 \\ 0 & \text{o.w.} \end{cases}$

To prove : $\int f(x)\,dx = 1$

$\therefore \int_0^1 3x^2$

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

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

$$\therefore \left(\frac{3(1)^3}{\cancel{x}3}\right) - \left(\frac{3(0)^3}{3}\right)$$

$$\therefore \quad 1 - 0$$

$$= \underline{\underline{1}}$$

$\therefore$ It is a p.d.f.

(iii) $f(x) = \begin{cases} \frac{3x}{2}\left(1-\frac{x}{2}\right) & ; \ 0 \le x \le 2 \\ 0 & \text{O.W} \end{cases}$

To prove : $f(x)$ is a p.d.f.

$$f(x) = \frac{3x}{2}\left(1 - \frac{x}{2}\right)$$

$$= \left(\frac{3x}{2} - \frac{3x^2}{4}\right)$$

$$\therefore \int_0^2 f(x) = \int \left(\frac{3x}{2} - \frac{3x^2}{4}\right)$$

$$\therefore \int_0^2 \left[\frac{3x}{2}\right] - \int_0^2 \left[\frac{3x^2}{4}\right]$$

$$\therefore \left[\frac{3x^2}{4}\right]_0^2 - \left[\frac{3x^3}{12}\right]_0^2$$

$$\therefore \left[\frac{3(2)^2}{4}\right] - \left[\frac{3(2)^3}{12}\right]$$

Scanned with CamScanner

$$\frac{\cancel{12}^{3}}{\cancel{4}_{1}} - \frac{2\cancel{4}^{2}}{\cancel{12}_{1}}$$

$$= 1$$

$\therefore$ $f(x)$ is a p.d.f.

QII. Can the following be p.m.f ?

(i)

x	1	2	3	4	5
$P(X)$	0.2	0.3	-0.1	0.5	0.1

$P(3) = -0.1$

Since, one probability is ~~possiti~~ negative, $P(X)$ is not a p.m.f.

(ii)

x	0	1	2	3	4	5
$P(X)$	0.1	0.3	0.2	0.2	0.1	0.1

$$P(X) = 0.1 + 0.3 + 0.2 + 0.2 + 0.1 + 0.1$$
$$= 1.0$$

$\therefore$ Since, $P(X) \geq 0 \;\; \forall x$
and $\Sigma P(X) = 1$

$P(X)$ is a p.m.f.

(iii)

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

$$\Sigma P(x) = 0.2 + 0.3 + 0.3 + 0.2 + 0.2$$
$$= 1.2$$

since $\Sigma P(x) \neq 1$

$P(x)$ is not a p.m.f.

Q III. Find $P(X \leq 2)$, $P(2 \leq X < 4)$, $P(\text{atleast } 4)$, $P(3 < X < 6)$

x	0	1	2	3	4	5	6
$P(X)$	0.1	0.1	0.2	0.2	0.1	0.2	0.1

<u>Soln:</u>

(i) $P(X \leq 2) = P(0) + P(1) + P(2)$
$$= 0.1 + 0.1 + 0.2$$
$$= \underline{\underline{0.4}}$$

(ii) $P(2 \leq X \leq 4) = P(2) + P(3)$
$$= 0.2 + 0.2$$
$$= \underline{\underline{0.4}}$$

(iii) $P(\text{atleast } 4) = P(4) + P(5) + P(6)$
$$= 0.1 + 0.2 + 0.1$$
$$= \underline{\underline{0.4}}$$

(iv) $P(3 < X < 6) = P(4) + P(5)$
$$= 0.1 + 0.2 = 0.3$$

x	0	1	2	3	4	5	6
$P(X)$	0.1	0.1	0.2	0.2	0.1	0.2	0.1

Soln:

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.1, & 0 \le x < 1 \\ 0.2, & 1 \le x < 2 \\ 0.4, & 2 \le x < 3 \\ 0.6, & 3 \le x < 4 \\ 0.7, & 4 \le x < 5 \\ 0.9, & 5 \le x < 6 \\ 1.0, & x \ge 6 \end{cases}$$

(ii)

x	10	12	14	16	18
$P(x)$	0.2	0.35	0.15	0.25	0.1

Soln:

$$F(x) = \begin{cases} 0.2, & x < 10 \\ 0.55, & 10 \le x < 12 \\ 0.70, & 12 \le x < 14 \\ 0.90, & 14 \le x < 16 \\ 1.0, & x \ge 18 \end{cases}$$

PRACTICAL NO. 3

Title: Binomial Distribution and Probability.

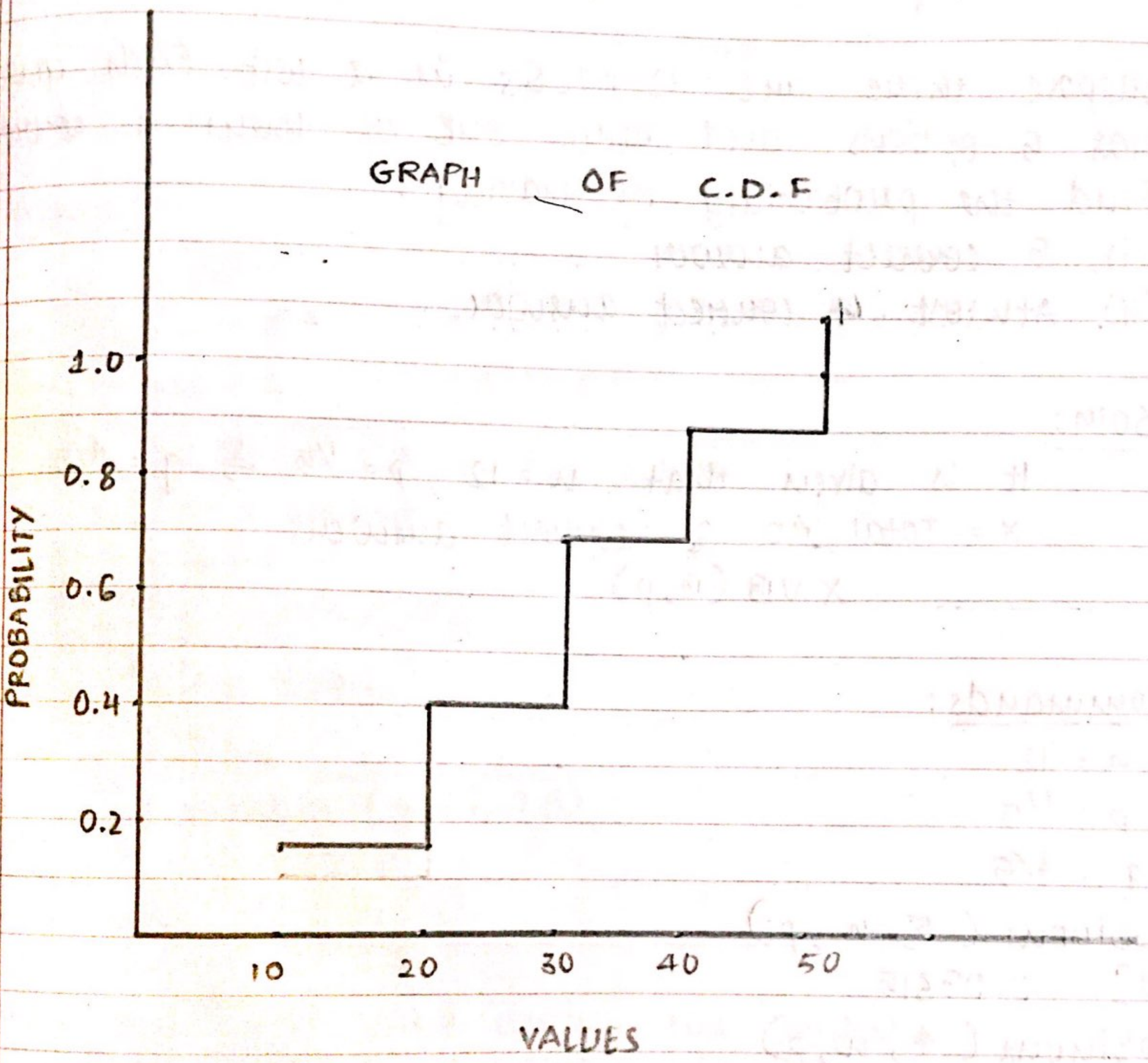
Q. Find the c.d.f. of the following p.d.f. and draw the graph.

x	10	20	30	40	50
$p(x)$	0.15	0.25	0.3	0.2	0.1

Soln:

$$
\begin{aligned}
F(x) &= 0 && \text{if } x < 10 \\
&= 0.15 && \text{if } 10 \le x < 20 \\
&= 0.40 && \text{if } 20 \le x < 30 \\
&= 0.70 && \text{if } 30 \le x < 40 \\
&= 0.90 && \text{if } 40 \le x < 50 \\
&= 1.0 && x \ge 50
\end{aligned}
$$

commands
```
> x = c(10, 20, 30, 40, 50)
> prob = c(0.15, 0.25, 0.3, 0.2, 0.1)
> cumsum (prob)
[1] 0.15, 0.40, 0.70, 0.90, 1.0
> plot (x, probx, xlab = "Values", ylab = "Probability",
    main = "GRAPH OF C.D.F", "s")
```


GRAPH OF C.D.F
PROBABILITY
1.0
0.8
0.6
0.4
0.2
10
20
30
40
50
VALUES

Q. Suppose there are 12 MCQ's in a test. Each question has 5 options and only one of them is correct. Find the probability of having :
(i) 5 correct answer
(ii) Atmost 4 correct answer.

<u>Soln:</u>

It is given that $n = 12$, $p = 1/5$ & $q = 4/5$
X = Total no. of correct answers
$X \sim B(n, p)$

<u>Commands:</u>

```
> n = 12
> p = 1/5
> q = 4/5
```

(i)
```
> dbinom ( 5, n , p)
[1]   0.05315
```

(ii)
```
> pbinom ( 4, n, p)
[1]   0.92744
```

Q. There are 10 members in a committee. The probability of any member attending a meeting is 0.9. Find the probability if :
(i) 7 members attended.
(ii) Atleast 5 members attended.
(iii) Atmost 6 members attended.

<u>Soln:</u>

It is given that $n=10$, $p=0.9$ & $q=0.1$

X = Total no. of members attending.

$X \sim B(n,p)$

<u>Commands:</u>

```
> n = 10
> p = 0.9
> x = 7
> dbinom (x, n, p)
```

(i) [1] 0.05739

```
> x = 5
> 1- pbinom (x, n, p)
```

(ii) [1] 0.99836

```
> x = 6
> pbinom (6, 10, 0.9)
```

(iii) [1] 0.01279

8. Find the c.d.f and draw the draw.

x	0	1	2	3	4	5	6
$p(x)$	0.1	0.1	0.2	0.2	0.1	0.2	0.1

<u>Soln:</u>

$F(x) = 0$ ~~0.1~~   ~~if $x < 0$~~   if $x<0$

$= 0.1$ ~~0.2~~   ~~if $0 < x < 0.1$~~   if $0<x<1$

$= 0.2$ ~~0.4~~   ~~if $1 < x < 0.2$~~   if $1<x<2$

$= 0.4$ ~~0.6~~   ~~if $0.3 < x < 0.2$~~   if $2<x<3$

$= 0.6$ ~~0.7~~   ~~if $x < 0.2$~~   if $3<x<4$

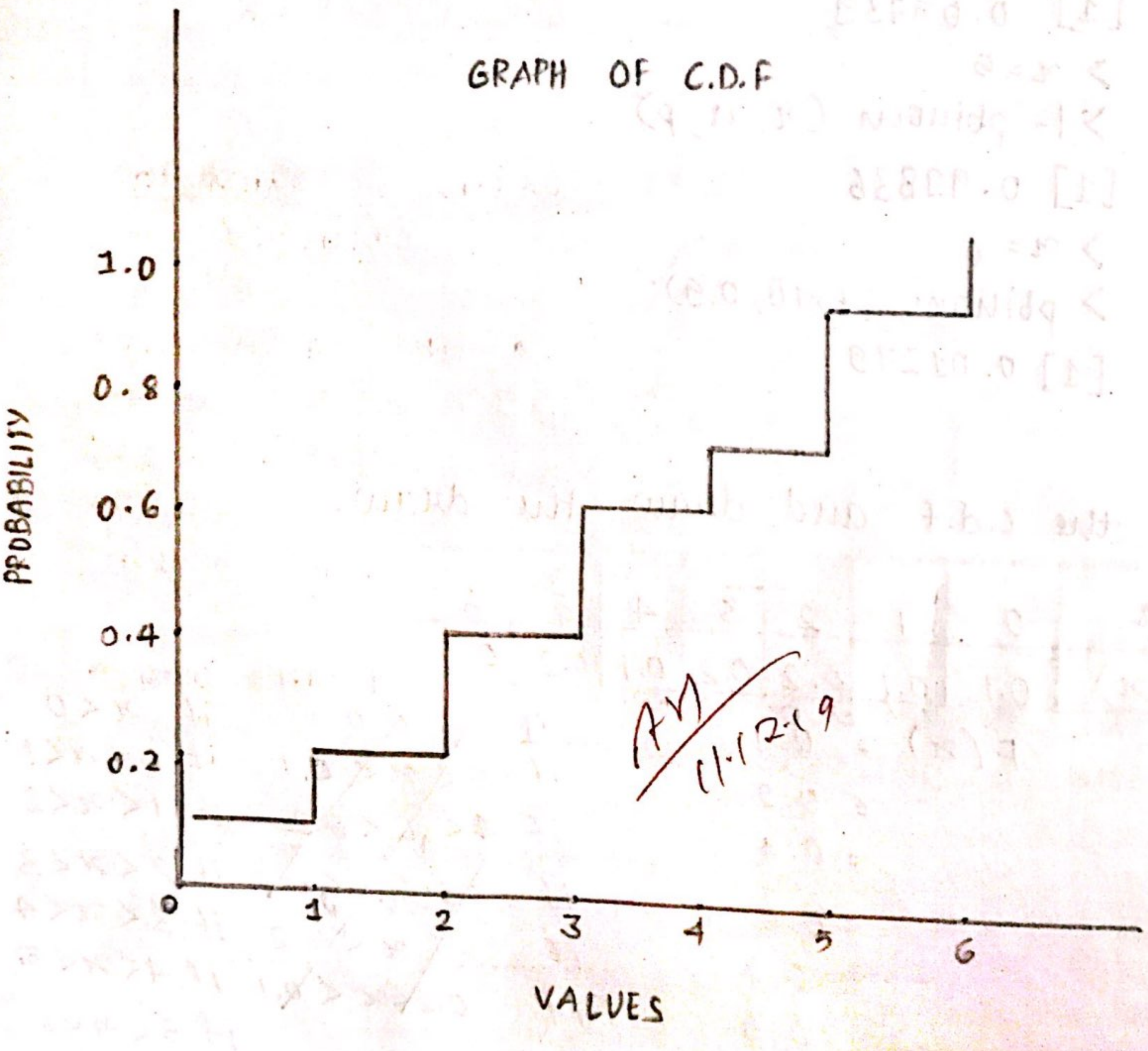
$= 0.7$ ~~0.9~~   ~~if $0.2 < x < 0.1$~~   if $4<x<5$

$= 0.9$ ~~1.0~~   if $5<x<6$

$= 1.0$ if $x<6$

commands:

```
> x = c(0,1,2,3,4,5,6)
> probx = c(0.1,0.1,0.2,0.2,0.1,0.2,0.1)
> cumsum(probx)
[1] 0.1  0.2  0.4  0.6  0.7  0.9  1.0
> plot(x, cumsum(probx), xlab = "VALUES",
      ylab = "PROBABILITY", main = "GRAPH OF C.D.F.", "s")
```



PRACTICAL NO. 4

1. Find the complete binomial distribution $a = 5$ & $p = 0.1$

2. Find the probability of exactly 10 success in 100 trial with $p = 0.1$

3. X follows binomial distribution with $n = 12$, $p = 0.25$ find
   1. $P(x = 5)$
   2. $P(x \leq 5)$
   3. $P(x > 7)$
   4. $P(5 < x < 7)$

4. Probability of salesman makes a sales to the customer is 0.15. Find the probability
   1. No sale for 10 customers
   2. More than ~~8~~ 3 sale in 20 customers

5. A student write 5 mcq. Each question has 4 option out of which 1 is correct. Calculate the probability for atleast 3 correct answers.

6. X follow binomial distribution $n = 10$, $p = 0.4$, plot the graph of p.m.f & c.d.f.

Note:
1. $P(x = x)$, dbinom(x, n, p)
2. $P(x \leq x)$, p binom (x, n, p)
3. $P(x > x)$, 1. pbinom (x, n, p)
4. To find the value of x, for which the probability is given as P1.

37

Solution :

1. $n=5$, $p=0.1$

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

2. $n=100$, $p=0.1$, $x=10$

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

3. i) $n=12$, $p=0.25$, $x=5$

```
> dbinom (5, 12, 0.25)
[1] 0.1032414
```

ii) $n=12$, $p=0.25$, $x=7$

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

iii) $n=12$, $p=0.5$, $x=5$

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

iv) $n=12$, $p=0.25$, $x=6$

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

4. (i) n = 0 , p = 0.15 , x = 0

```
> dbinom (0, 10, 0.15)
[1] 0.1958744
```

(ii) n = 20 , p = 0.15

P(x>3) = 1 - P(x ≤ 3) = 1 - pbinom (3, 20, 0.15)

```
[1] 0.3522748
```

5. n = 5, p = 1/4 , x ≥ 3 or x ≤ 2

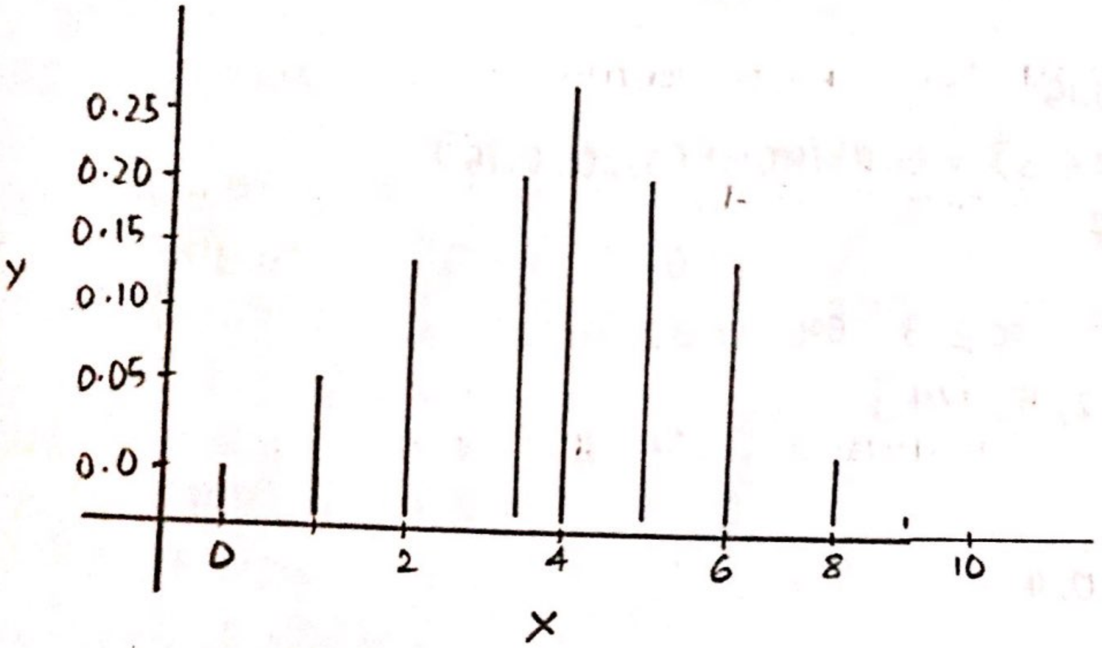
```
> 1-pbinom [ 2, 5, 1/4]
[1] 0.1035156
```

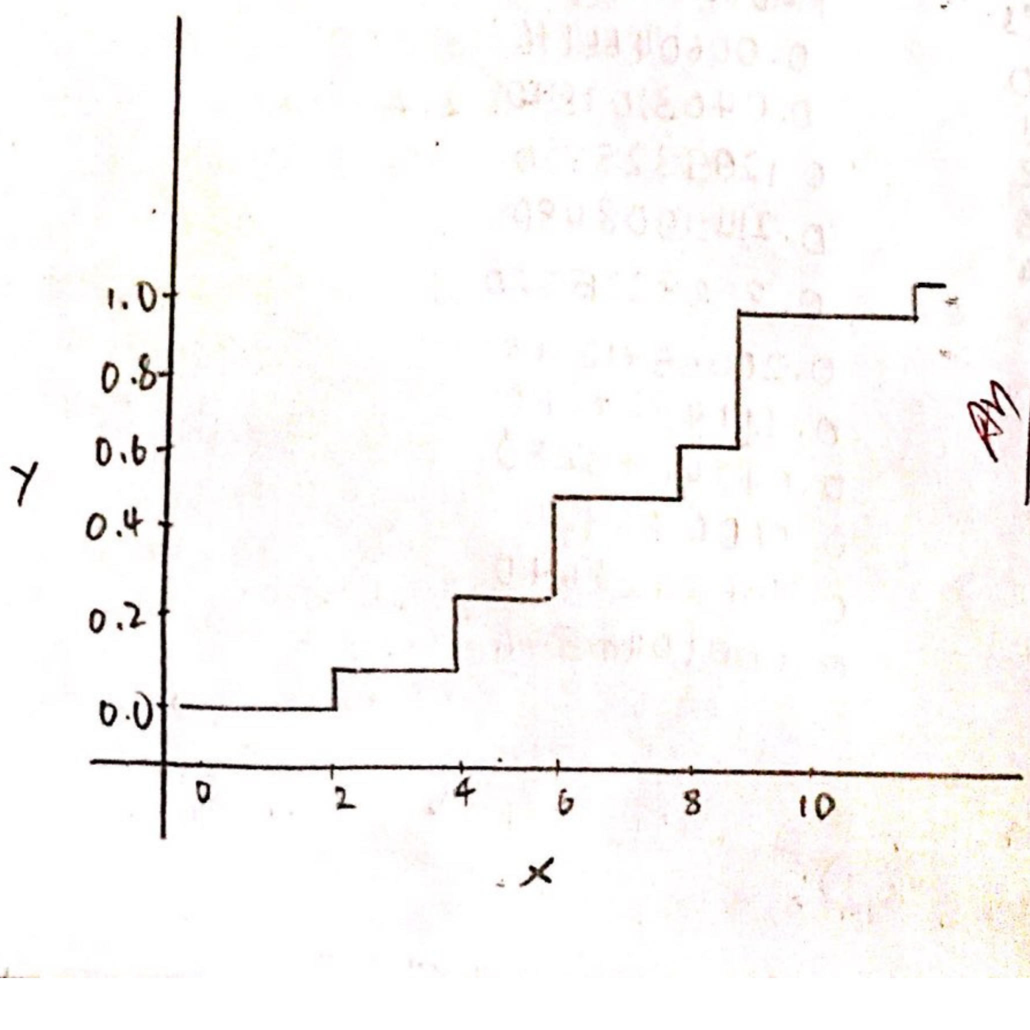
6. n = 10 , p = 0.4

```
x = 0:n
prob = dbinom [ x, n, p]
cumprob = pbinom [x, n, p]
d: data.frame ( "x values": x, "Probability"= prob)
print (d)
```

	x.values	probability
1	0	0.0060466176
2	1	0.0403107840
3	2	0.1209323520
4	3	0.2149908480
5	4	0.2508226560
6	5	0.2006581248
7	6	0.1114767360
8	7	0.0424673280
9	8	0.0106168320
10	9	0.0015728640
11	10	0.0001048576

```
Plot (x, prob , "h")
Plot ( x, cumprob, "s")
```





AM

PRACTICAL NO. 5

Normal Distribution

1. $P[X=x]$ = dnorm (x, μ, σ)

2. $P(X \leq x)$ = pnorm (x, μ, σ)

3. $P(X = x)$ = 1- pnorm (x, μ, σ)

4. $p(x_1 < x < x_2)$ = pnorm $(x2, \mu, \sigma)$

5. To find the value of k so that:
   $P(X \leq k) = P_1$; qnorm $(p1, \mu, \sigma)$

6. To generate 'n' random numbers:
   rnorm (n, μ, σ)

Q1. $X \sim N(\mu = 50,\ \sigma^2 = 100)$

Find: i) $P(X \leq 40)$

ii) $P(42 \leq X \leq 60)$

ii) $P(X > 55)$

iii) $P(X \leq k) = 0.7$; $k = ?$

Q2. $X \sim N(\mu = 100,\ \sigma^2 = 36)$

i) $P(X \leq 110)$

ii) $P(X \leq 95)$

iii) $P(X > 115)$

iv) $P(95 \leq X \leq 105)$

v) $P(X \leq k) = 0.4$; $k = ?$

Q3. Generate 10 random numbers from normal distribution with mean $(\mu) = 60$, standard deviation $(\sigma) = 0.5$. Also calculate the sample mean, median, variance and standard distribution.

Q4. Draw the graph of standard normal distribution.

Solutions:

1. (i) $P(X \leq 40)$

```
> a = pnorm (40, 50, 10)
> cat ("P(x<=40) is: ", a)
[1] P(x<=40) is: 0.1586553
```

(ii)

```
c = pnorm(60,50,10) - pnorm(42,50,10)
> cat ("p(42<x<=60) is: ", c)
[1] 0.6294893
```

(iii) $P(X>55)$

```
> b = 1- pnorm (55, 50, 10)
> cat ("P(X>55) is: ", b)
[1] P(X>55) is: 0.3085375
```

(iv) $P(X \leq k) = 0.7$; $k = ?$

```
> d = qnorm (0.7, 50, 10)
> cat ("P(X<=k)=0.7, k is: ", d)
[1] P(X<=k)=0.7, k is: 55.24401
```

2. (i) $P(X \leq 110)$

```
> a = pnorm (110, 100, 6)
> cat ("P(X<=110) is : ", a)
[1] P(X<=110) is: 0.9522096
```

(ii) $P(X \leq 95)$

```
> b = pnorm (95, 100, 6)
> cat ("P(X<=95) is : ", b)
[1] P(X<=95) is : 0.9522096 0.2023284
```

(iii) $P(X \cancel{\leq} > 115)$

```
> c = 1 - pnorm (115, 100, 6)
> cat ("P(X>115) is : ", c)
[1] P(X>115) is : 0.006209665
```

(iv) P(95 <= X <= 105)

```
> d = pnorm (95, 100, 6) - pnorm (105, 100, 6)
> cat ("P(95<=X<=105) is : ", d)
[1] P(95<=X<=105) is : - 0.5953432
```

(v) $P(X \leq k) = 0.4$, k is ?

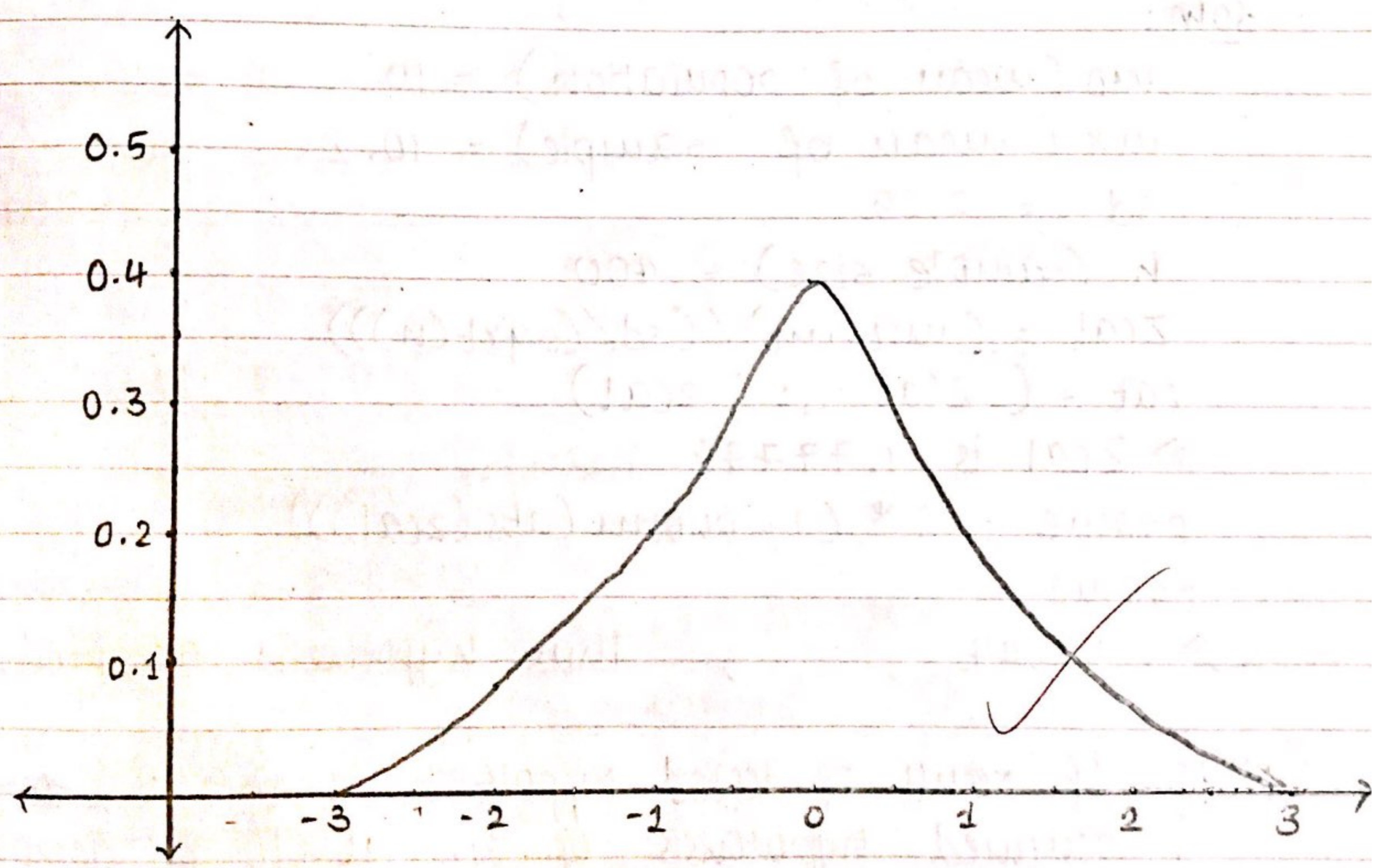
```
e = qnorm (0.4, 100, 6)
cat ("P(X<=k)=0.4 , k is : ", e)
P(X<=k) = 0.4 , k is: 98.47992
```

3.

```
> x = rnorm (10, 60, 5)
> x
[1] 52.4310    59.9626    63.72094    61.2825    61.9102
    51.8524    66.9836    62.5194     61.3292    62.4803
> am = mean(x)
> am
[1] 60.44768

> median = median(x)
> median
[1] 61.6198

> n=10
> variance = (n-1) * var(x)/n
> variance
[1] 20.37886

> sd = sqrt (variance)
> sd
[1] 4.5142
```

4. Draw the graph of standard normal distribution.

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



AN

PRACTICAL NO. 6

TOPIC : ~~SET~~ Z DISTRIBUTION

* **Mean Test :**

Q1. ~~Guess~~ Test the ~~H~~ Hypothesis (H_0) :

(i) $H_0 : \mu = 10$ against $H_1 : \mu \neq 10$.

A sample of size 400 is selected which gives a mean 10.2 and standard deviation 2.25. Test the hypothesis at 5% level of significance.

Soln:

mo (mean of population) = 10
mx (mean of sample) = 10.2
sd = 2.25
n (sample size) = 400
zcal = (mx-mo)/(sd/(sqrt(n)))
cat = ("zcal is: ", zcal)
>> Zcal is 1.77778
pvalue = 2 * (1- pnorm (abs (zcal)))
pvalue
>> 0.07544 , Thus hypothesis accepted.

(NOTE : If result of tested hypothesis is >0.05, then assumed hypothesis of $H_0 : \mu = 10$ is accepted as verified / correct)

Q2. Test the hypothesis (H_0) :

$H_0 : \mu = 75$ against $H_0 : \mu \neq 75$

A sample of size 100 is selected and the sample mean is 80, with sd of 3. Test the hypothesis at 5% of significance.

Soln:

mo = 75
mx = 80
sd = 3
n = 100
zcal = (mx - mo)/(sd/(sqrt(n)))
cat = ("zcal : ", zcal)
pvalue = 2* (1- pnorm (abs (zcal)))
cat("pvalue is : ", pvalue)

Output:

>> zcal : 16.667
>> pvalue is : 0

Thus, hypothesis rejected.

Q3. Test the hypothesis (H_0):
$H_0 : \mu = 25$ against $H_0 : \mu \neq 25$

A sample of 30 is selected. Test the hypothesis at 5% level of significance.

sample(x) 20, 24, 27, 35, 30, 46, 26, 27, 10, 20, 30, 37, 35, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 39, 27, 15, 19, 22, 20, 18

Soln:

mx = mean(x)
>> mx = 26.066
n = length(n)
variance = (n-1) * var(x)/n

sd = sqrt(variance)
>> sd = 7.2798
mo = 25
zcal = (mx-mo)/(sd/sqrt(n))
>> zcal = 0.8025
pvalue = 2 * (1-pnorm(abs(zcal)))
>> pvalue = 0.4223

Thus, hypothesis accepted.

* <u>Proportion Test</u> :

Q1. Experience has shown that 20% students of a college smoke. A sample of 400 students reveal that out of 400 only 50 ~~sp~~ smoke. Test the hypothesis that the experience keeps the correct proportion or not.

<u>Soln</u>:

pp = 0.2 (20%)
q = 1- pp
sp = 50/400
n = 400
zcal = (sp-pp)/(sqrt(pp*q/n)
>> zcal = -3.75
pvalue = 2* (1-pnorm(abs(zcal)))
>> pvalue = 0.00017

Thus, hypothesis is rejected.

Q2. Test the hypothesis : $H_0 : p = 0.5$ against $p \neq 0.5$
A sample of 200 is selected where sp = 0.56 and pp = 0.5. level of significance is 1%.

<u>soln</u>:

pp = 0.5
q = 1 - pp
sp = 0.56
n = 200
zcal = (sp-pp)/(sqrt(pp*q/n)
>> zcal = 1.6970
pvalue = 2 * (1 - pnorm(abs(zcal)))
>> pvalue = 0.08968

Thus, hypothesis is accepted.

AM
22-01-20

RA

PRACTICAL NO. 7

TOPIC : Large Sample Test

Q1. H_0 : A study of noise level in two hospitals is calculated below. Tes the hypothesis (H_0) that the noise level in 2 hospitals are same.

~~No. of Obs.~~	Hospital A	Hospital B
No. of Obs.	84	34
Mean	61	59
S.D	7	8

Soln:

H_0 : The noise levels are same.

```
n1 = 84
n2 = 34
mx = 61
my = 59
sdx = 7
sdy = 8
z = (mx-my)/sqrt((sdx^2/n1)+(sdy^2/n2))
[1] 1.2736
cat(z is: ;z)
pvalue = 2*(1-pnorm(abs(z)))
[1] 0.2027
```

Since pvalue > 0.05, we accept ~~reject~~ H_0 at 5% level of significance.

Q2) H_0 : Two random sample of size 1000 and 2000 are drawn from population with mean 67.5 & ~~44~~ 68 resp. with the standard deviation at 2.5.
Test the hypothesis (H_0) that the mean of two population are equal.

Soln:

H_0 : The ~~noise~~ mean of population are equal

```
n1 = 1000
n2 = 2000
mx = 67.5
my = 68
sdx = 2.5
sdy = 2.5
z=(mx - my)/sqrt((sdx^2/n1)+(sdy^2/n2))
[1] - 5.1639
pvalue = 2*(1-pnorm(abs(z)))
[1] 2.41756e-07
```

since, pvalue < 0.05, we reject the H_0 at 5% level of significance.

Q3. H_0: In FYBSc, 20% of a random sample of 400 students had defective eyesight. In SYBSc, 15.5% of 500 students had the same defect. Is the difference of proportion same?

<u>Soln:</u>

H_0: The proportion of the population are equal.

```
n1 = 400
n2 = 500
p1 = 0.2
p2 = 0.155
p = (n1 * p1 + n2 * p2)/(n1+n2)
[1] 0.175
q = 1-p
z = (p1-p2)/(sqrt(p*q*(1/n1 + 1/n2)
[1] 1.76547
pvalue = 2 * (1-pnorm(abs(z)))
[1] 0.0774
```

Since, pvalue > 0.05, we accept the H_0, at 5% level of ~~[illegible]~~ significance.

Scanned with CamScanner

Q4. H_0 : From each of the box of the apples a sample size of 200 is collected. It is found that there are 44 bad apples in the first sample and 30 bad apples in the second sample.
Test the hypothesis (H_0) that the two boxes are equivalent in terms of no. of bad apples.

<u>Soln:</u>

H_0 : The two boxes are equivalent

```
n1 = 200
n2 = 200
p1 = 44/200
p2 = 30/200
p = (n1 * p1 + n2 * p2)/(n1+n2)
[1] 0.185
q = 1-p
z = (p1-p2)/(sqrt (p * q*(1/n1 + 1/n2)
[1] 1.8027
pvalue = 2 * (1-pnorm (abs (z)))
[1] 0.07142
```

Since, pvalue > 0.05, we accept the H_0, at 5% level of significance.

Q5. H_0 : In a class, out of a sample of 60, mean height is 63.5 inch with a sd of 2.5. In a MCOM class, out of 50 students, mean height is 69.5 inches with sd of 2.5. Test the hypothesis of MA and MCom class are the same.

Soln:

H_0 : mean Heights of two classes is the same.

```
n1 = 60
n2 = 60
mx = 63.5
my = 69.5
sdx = 2.5
sdy = 2.5
z = (mx-my)/sqrt((sdx^2)/n1)+(sdy^2/n2))
[1] 149.1808
pvalue = 2 * (1-pnorm (abs(z)))
[1] 0
```

Since, pvalue < 0.05 we reject the H0, at 5% level of significance.

5.2.20

PRACTICAL NO. 8

TITLE : Small sample Test

Q1. The flower's stems are selected & the heights are found to be 63, 63, 68, 71, 71, 72 cm. Test the hypothesis that the mean height is 66 cm or not, at 1% level of significance.

→ H_0 : Mean = 66 cms

x = c(63, 63, 68, 69, 71, 71, 72)

x

[1] 63 63 68 69 71 71 72

t. test[x]

One sample t-test

data : x

t = 47.94

df = 6

p values = 5.522e-09

alternative hypothesis : True mean is not equal to 0

95% confidence interval

64.66479 71.62092

sample estimates :

mean of x

68.14286

Since, p-value < 0.01, we reject H_0 at 1% level of significance.

Q2. Two random samples were drawn from two different populations

sample 1 : 8, 10, 12, 11, 15, 15, 18, 7

sample 2 : 20, 15, 18, 9, 8, 10, 11, 12

Test the hypothesis that there is no difference between the ~~two~~ two population mean at 5% level of significance.

→ Ho : There's no difference in the population mean.

x = c(8, 10, 12, 11, 15, 15, 18, 7)
y = c(20, 15, 18, 9, 8, 10, 11, 12)
T.test(x, y)

Welch two sample t-test

data : x and y

t = -0.3624 , df = 13.837, p-value = 0.7225

Alternate Hypothesis : True difference is not equal to 0.

95% confidence in interval:

-5.192719 3.692719

sample estimates :

mean of x mean of y

12.125 12.875

∵ p-value > 0.05, we accept hypothesis Ho with 5% level of significance.

Scanned with CamScanner

Q3. Following are the weight of 10 people, before and after a diet program. Test the hypothesis that diet is effective or not.

Before (kg): 100, 125, 95, 96, 98, 112, 115, 104, 109, 110
After (kg): 95, 80, 95, 98, 90, 100, 110, 85, 100, 101

Ho: The diet program isn't effective.
x = c(100, 125, 95, 96, 98, 112, 115, 104, 109, 110)
y = c(95, 80, 95, 98, 90, 100, 110, 85, 100, 101)
t.test (x, y, paired = T, alternative = "less")

Paired t-test

data: x and y
t = 2.6089, df = 9, p-value = 0.9858
alternative hypothesis: True difference in mean is less than 0.

95 % confidence interval:
- INF 18.72908
sample estimates:
mean of the differences
11

∵ p-value > 0.05, we accept the H_0 at 5% level of significance.

SR:

Q. The marks before & after a training program are given below:
Before : 20, 25, 32, 28, 27, 36, 35, 25
After : 30, 35, 32, 37, 37, 40, 40, 23
Test the hypothesis, training program effective or not.

H_0 : Training is not effective.

x = c(20, 25, 32, 28, 27, 36, 35, 25)
y = c(30, 35, 32, 37, 37, 40, 23)
t.test(x, y, paired = t, alternative = "greater")

Paired t-test
data : x and y
t = -3.3859 , df = 7 , pvalue = 0.9942
Alternate Hypothesis : True difference in mean is greater than 0
95% confidence interval
-8.967399 Inf
sample estimate
mean of the differences
-5.75

Since, p-value > 0.05, we accept the H_0 with 5% level of significance

Q. Two random sample were drawn from two normal population & the values are:

A : 66, 67, 76, 82, 84, 88, 90, 92

B : 64, 56, 74, 78, 82, 85, 87, 92, 93, 95, 97

Test whether population have same varianie at 5% level of significance.

H_0: The variances of the two population are equal.

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 variables

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 1.

95% percent confidence interval

0.1833662 3.0360393

sample estimates:

0.7068567

since p-value > 0.05, we accept the H_0 at 5% level of significance.

Q. The arithematic mean of sample of 100 observations is 52. If the s.d is 7, test the hypothesis that the population mean is 55 or not at 5% significance

→ Ho : population mean = 55

n = 100

m2 = 55

sd = 7

m0 = 52

Zcal = (ma-m0)/(sd/sqrt(n))

Zcal

[1] -4.285714

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

pvalue

[1] 1.82153e-05

since, p-value < 0.05, we reject the Ho at 5% level of significance.

Pratical NO. 9

TITLE : Chi square Distribution and ANOVA

Q. use the following data to test whether the cleanliness of a home depends upon the child condition or not.

~~condition of Home~~

~~cond^n. of child | Clean | Fairly Clear | Dirty~~

		Condition of Home	
		Clean	Dirty
Condn. of child	Clean	70	50
	Fairly clean	80	20
	Dirty	35	45

Solution :

H_0 : Condition of home & child are independant.

```
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 < 0.05 we reject hypothesis H_0, at 5% level of significance.

Q. Data below shows relation between the performance between maths and computer of CS students.

		maths		
		HG	MG	LG
comp.	HG	56	71	12
	MG	47	163	38
	LG	14	42	85

H_0: Performance between two subjects are independent
x = c(56, 47, 14, 71, 163, 42, 12, 38, 85)
m = 3
n = 3
y = matrix (x, nrow = m, ncol = 3)
y

	[1]	[2]	[3]
[1]	56	71	12
[2]	47	163	38
[3]	14	42	85

Scanned with CamScanner

pr = chisq . test(y)

pr

Pearson's Chi-squared test

data : y

x-squared = 145.78 , df = 4 , pvalue < 2.2e-16

since, pvalue < 0.05, we reject H_0 at 0.05 ie 5% level of significance.

Q. Perform ANOVA for the following data.

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

H_0 : The means of variety of A, B, C, D 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)
"values" "ind"
```

02

oneway.test (values~ind , data = d, var.equal = T)

One-way analysis of means

data : values and ind

F = 11.735 , num df = 3, p-value = 0.00183

anova = aov(values~ind , data = d)

summary(anova)

	Df	Sum	Mean Sq	F value	Pr(>F)
ind	3	71.06	23.688	11.73	0.00183
Residuals	9	18.17	2.019		

∵ pvalue < 0.05 , we reject hypothesis H_0 , at 5% level of significance.

8. Perform ANOVA.

Types	Observations
A	6, 7, 8
B	4, 6, 5
C	8, 6, 10
D	6, 9, 9

H_0 : means of types A, B, C, D are equal.

```
x1 = c(6,7,8)
x2 = c(4,6,5)
x3 = c(8,6,10)
x4 = c(6,9,9)

d = stack (list(b1=x1, b2=x2, b3=x3, b4=x4))
names(d)
"values"  "ind"
oneway.test (values~ind, data=d, var.equal=T)
```

One-way analysis of means

data : values and ind

F = 2.6667 , num df = 3, denom df = 8

p-value = 0.1189

∴ p-value > 0.05 , we accept the hypothesis at 5% level of significance.

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	3	18	6.00	2.667	0.119
Residuals	8	18	2.25		

8.

x = read.csv("C:/Users/admin/Desktop/marks.csv")
x

	STATS	CALC
1	40	60
2	45	48
3	42	47
4	15	20
5	37	25
6	36	27
7	49	57
8	59	58
9	20	25
10	27	27

PRACTICAL NO. 10

TITLE : Non- Parametric Test

Q.1] Following are the amounts of sulphur oxide emitted by a factory.

17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26.

Apply sign test, to test the hypothesis that the population median is 21.5. against the alternative that it is less than 21.5

H_0 : Population median is 21.5

H_1 : It is less than 21.5

```
x = c(17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20,
      17, 6, 24, 14, 15, 23, 24, 26)
m = 21.5
sp = length (x[x>m])
sn = length( x[x<n])
n = sp+sn
pv = pbinom (sp, n, 0.5)
pv = 0.4119
```

since, p-value > 0.05, we accept the hypothesis at 5% level of significance.

Scanned with CamScanner

[NOTE: If the alternative is greater than median pv = pbinom (sn, n, 0.5)]

Q.2] For the observations : 12, 19, 31, 28, 48, 40, 55, 49, 70, 63

Apply sign test, ~~rule~~ to test hypothesis that population ~~mead~~ median is 25, against the alternative hypothesis that is ~~less~~ greater than 25.

H_0 : Median is 25

H_1 : median is greater than 25.

```
x = c(12, 19, 31, 28, 48, 40, 55, 49, 70, 63)
m = 25
sp = length(x[x>m])
sn = length( x[x<m])
n = sp+sn
pv = pbinom (sn, n, 0.5)
```

pv = 0.0546

since, pvalue is > 0.05, we accept the hypothesis H_0 at 5% level of significance.

Q3]. For the following data : 60, 65, 63, 89, 61, 71, 58, 51, 48, 66.

-Test the hypothesis using wilcoxon's ~~test~~. signed rank for testing the hypothesis that the median is 60, against the alternative, it is greater than 60.

H_0 : median is 60
H_1 : Median is greater than 60

```
x = c(60,65,63,89,61,71,58,51,48,66)
m = 60
```

~~sp = length (x[x>m])~~
~~sn = length (x[x<m])~~
~~n = sp+sn~~
~~pr = pbinom (sn, n, 0.5)~~
~~pr = 0.0~~

Wilcoxon signed rank Test
with continuity correction

data : x
V = 29, p-value = 0.2386
alternative hypothesis : true location is greater than 60

since, p-value > 0.05, we accept the hypothesis H_0 at 5% level of significance

[NOTE : If the alternative is less,
wilcox.test (x, alter= "less", mu = 60)]

If the alternative is not equal to :
wilcox. test (x, alter = "two.sided", m = 60)]

Q4.] Test the hypothesis that median is 12, against the alternative that it is less than 12, using wilcoxon's signed test.
Values : 12, 13, 10, 20, 15, 5, 1, 7, 6,
~~Observ~~ 11, 9, 20

H_0 : Median is 12
H_1 : Median is less than 12

x = c(12, 13, 10, 20, 15, 5, 1, 7, 6, 11, 9, 20)
m = 12
wilcox. test (x, alter = "less", mu = 12)

Wilcoxon signed rank test with continuity correction

data : x
V = 25, p-value = 0.2521
alternative hypothesis : true location less than 12

since, p-value > 0.05 we accept hypothesis H_0 at 5% level of significance.