Practical - 1

Aim: Basics of R-software

(1) R is a software, for statistical analysis, and data computing.
(2) It is an effective data handling software and outcomes storage is possible.
(3) It is capable of graphical display.
(4) It's a free software.

(Q1) Solve the following:-

(1) 4 + 6 + 8 ÷ 2 - 5
• 4 + 6 + 8 / 2 - 5
→ 9

(2) $2^2 + |-3| + \sqrt{45}$
• 2 ^ 2 + abs(-3) + sqrt(45)
→ 13.7082

(3) $5^3 + 7 \times 5 \times 8 + 46/5$
• 5 ^ 3 + 7 * 5 * 8 + 46 / 5
→ 414.2

(4) $\sqrt{4^2 + 5 \times 3 + 7/6}$
→ 5.671

(5) round (46/7) + 9 * 8
→ 79

(Q2)
(1) c(2, 3, 5, 7) * 2
• 4 6 10 14

(2) c(2, 3, 5, 7) * c(2, 3)
• 4 9 10 21

(3) c(2, 3, 5, 7) * c(2, 3, 6, 2)
• 4 9 30 14

(4) c(1, 6, 2, 3) * c(-2, -3, -4, 1)
• -2 -18 -8 -3

(5) c(2, 3, 5, 7) ^ 2
• 4 9 25 49

(6) c(4, 6, 8, 9, 4, 5) ^ c(1, 2, 3)
• 4 36 512 9 16 125

(7) c(6, 2, 7, 5) / c(4, 5)
• 1.50 0.40 1.75 1.00.

Q3)
(1) x = 20, y = 30, z = 2.
Q1) $x^2 + y^3 + z$
• 72402
Q2) sqrt (x ^ 2 + y)
• 20.73644
Q3) x2 + y2
• 1300.

(Q4)
→ x <- matrix(nrow = 4, ncol = 2, data = c(1, 2, 3, 4,
→ x
#

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

Q5) Find x+y & 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 (nrow = 3, ncol = 3, data = c(4,7,9,-2,0,-5,6,7,3))
x     [,1]  [,2]  [,3]
[1,]   10    -5     7
[2,]   12    -4     9
[3,]   15    -6     5
y <- matrix (nrow = 3, ncol = 3, data = c(10,12,15,-5,-4,-6,7,9,5))
x     [,1]  [,2]  [,3]
[1,]    4    -2     6
[2,]    7     0     7
[3,]    9    -5     3

> x + y
      [,1]  [,2]  [,3]
[1,]   14    -7    13
[2,]   19    -4    16
[3,]   24   -11     8
> 2 * x + 3 * y
      [,1]  [,2]  [,3]
[1,]   38   -19    33
[2,]   50   -12    41
[3,]   63   -28    21
```

Q6) Marks of statistics of C.S students

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

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



```
> c
         a        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
```

28/1

Practical - 2

Aim :- Probablity Distribution

(Q1) Check whether the following are pmf or not

1.

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

2.

x	1	2	3	4	5
$p(x)$	0.2	0.2	0.3	0.2	0.2

3.

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

Ans

Q(1). If the given data is pmf then,

$\Sigma . p(x) = 1$

Hence

$\Sigma (P) = P(1) + P(2) + P(3) + P(4) + P(5)$

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

$= 1.0$

Since $P(2) = -0.5$,

It cannot be a probablity mass function as in pmf $P(X) \geq 0$

(2) $\Sigma P(X) = 1$ ... in pmf

(3) $p(x) \geq 0$

(2) $\Sigma p(x) = 1$

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

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

$= 1$

The given data is p.m.f.

code:

```
prob= c(0.2,0.2,0.35,0.15,0.1)
sum(prob).
```

(Q2) Find c.d.f.

(1) Find c.d.f for the foll p.m.f and sketch the graph:-

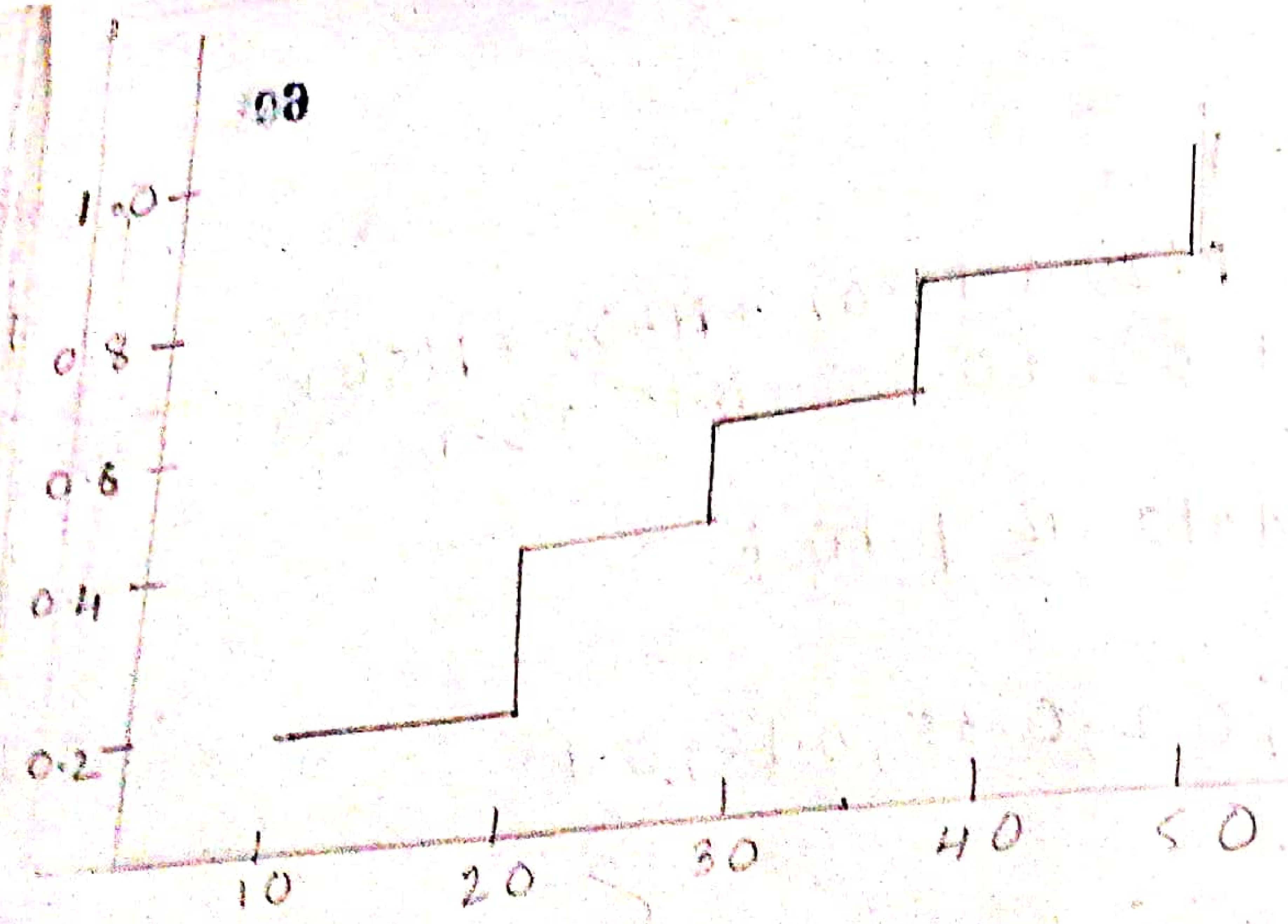
x	10	20	30	40	50
$p(x)$	0.2	0.2	0.35	0.15	0.1

code:

```
cumsum (prob)
```

$\rightarrow F(x) = $

0	$x < 10$
0.2	$10 \leq x < 20$
0.4	$20 \leq x < 30$
0.75	$30 \leq x < 40$
0.95	$40 \leq x < 50$
1.0	$50 \leq x$



(Q2) Find the c.d.f & sketch the graph

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 $1 \le x < 2$

0.40 $2 \le x < 3$

0.50 $3 \le x < 4$

0.70 $4 \le x < 5$

0.90 $5 \le x < 6$

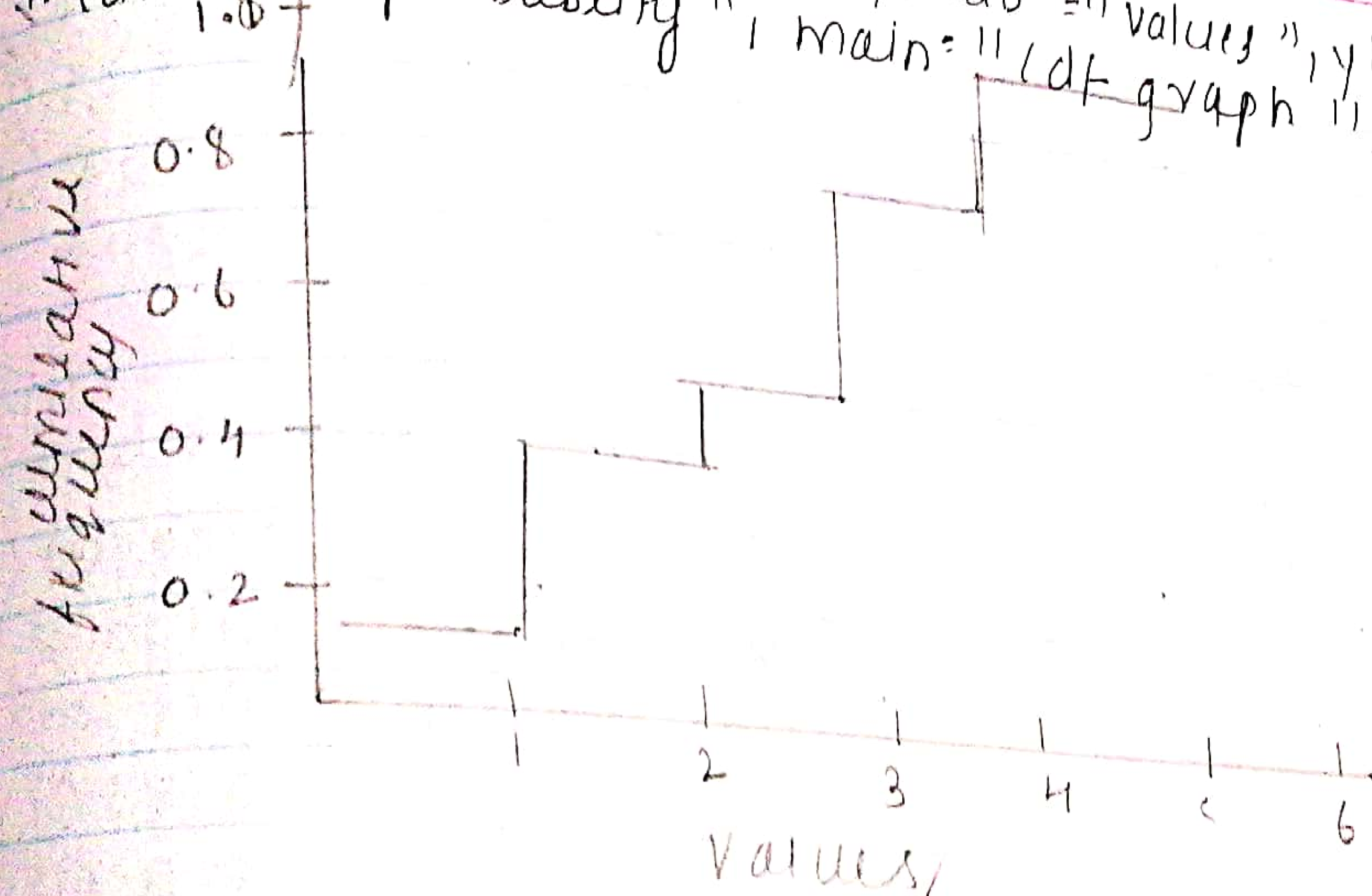
1.00 $x \ge 6$

CODE:

```
>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, 0.70, 0.90, 1.00
x = c (1,2,3,4,5,6)
```



```
plot (x, cumsum(prob), "s", xlab ="values", ylab=
"cumulative probability", main="cdf graph")
```



(Q3)

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

$$\int x^n \, dx = \frac{x^{n+1}}{n+1}$$

$$= \int_0^1 f(x)\, dx$$

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

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

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

:a

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

$\int f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 (x^2)$

$= 3\int_0^1 \left(\frac{x^3}{3}\right)$ $\quad \therefore x^n = \frac{x^{n+1}}{n+1}$

$= x^3$

$= 1$

The $\int_0^1 f(x) = 1$ $\therefore$ It is a pdf

AM

Practical - 3

62

TOPIC: Binomial Distribution

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

(2) $P(X \leq x)$ = pbinom (x, n, p)

(3) $P(X > x)$ = 1 - pbinom (x, n, p)

(4) If x is unknown
$p_1 = P(X \leq x)$ - qbinom (p1, n, p)

Q1) Find the probability of exactly 10 success in 100 trials with $p = 0.1$

Q2) Suppose there are 12 MCQ's. Each question has 5 options out of which one is correct. Find the probability of having

(i) Exactly 4 correct answer

(ii) Atmost 4 correct answer

(iii) More than 5 correct answer

Q3) Find the complete distribution where $n=5$, $p = 0.1$

Q4) $n = 12$, $p = 0.25$. Find the following probability,

1) $P(X = 5)$ 3) $P(X > 7)$

2) $P(X \leq 5)$ 4) $P(5 < X < 7)$

Q5) The probability of a salesman making a sell to a customer is 0.25. Find a probability of

(1) No sales out of 10 customer

(2) More than 3 sales out of 20 customer

(Q6) A salesman has a 20% probability of making a sell to a customer out of 30 customer what minimum number of sales he can make with 88% probability.

(Q7) x follows binomial distribution with n = 10, p = 0.3, Plot the graph of pmf and cdf.

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

(2) > dbinom (4, 12, 0.2)
[1] 0.1328756

> pbinom (4, 12, 0.2)
[1] 0.9274445.

> 1 - pbinom (5, 12, 0.2)
[1] 0.01940528.

(3) > dbinom (0:5, 5, 0.1)

[1] 0.59049 0.32805 0.07290
0.00810 0.00045 0.00001

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

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

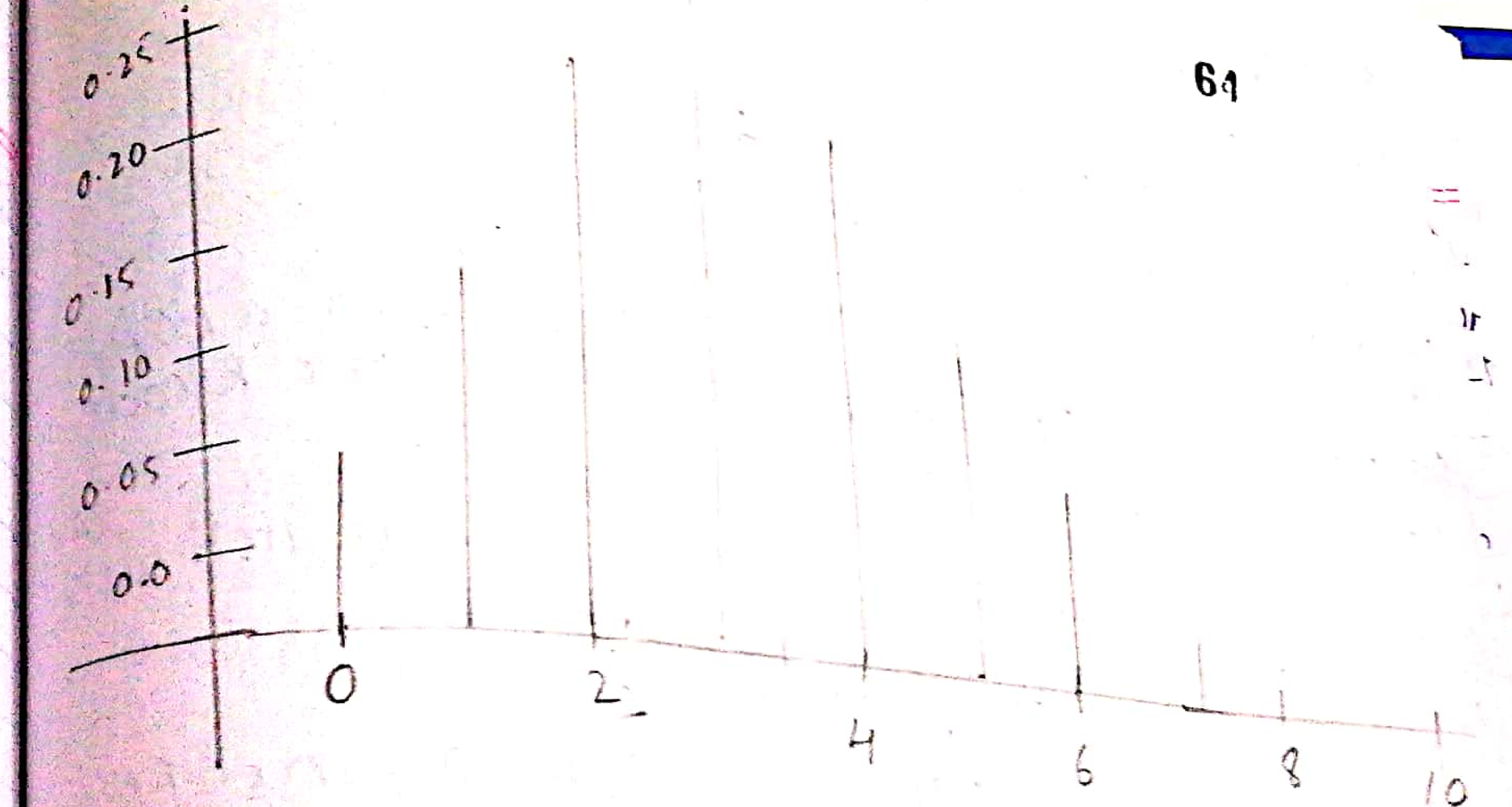
(iii) > 1 - pbinom (7, 12, 0.25)
[1] 0.000 ~~27~~ 82514

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

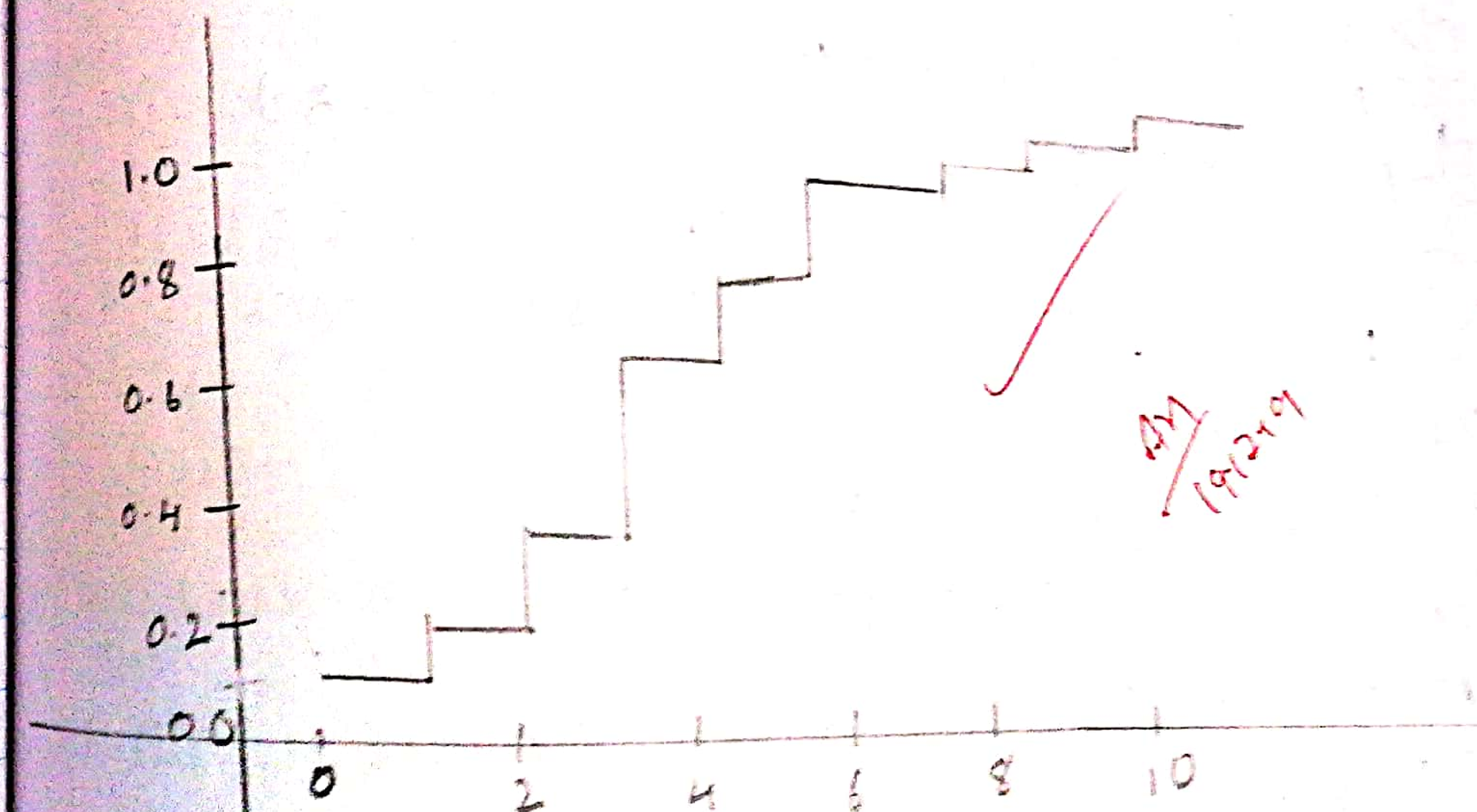
```
(5) (i) > dbinom (0, 10, 0.15)
[1] 0.1968744

(ii) > 1 - pbinom (3, 20, 0.15)
[1] 0.3522748

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

(7) > n = 10
> p = 0.3
> 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	0.0282
2	1	0.1210
3	2	0.2334
4	3	0.2668
5	4	0.2001
6	5	0.1029
7	6	0.367
8	7	0.0090
9	8	0.0014
10	9	0.0001
11	10	0.0000

> plot (, prob , "h")



> plot (x, cumprob, "s")



Practical 4

Aim: Normal Distribution.

Formula:

(1) P(X=x) = dnorm(x, μ, σ)

(2) P(X≤x) = pnorm(x, μ, σ)

(3) P(X>x) = 1 - pnorm(x, μ, σ)

(4) To generate random numbers from a normal distribution (n random number) the R code is = rnorm(n, μ, σ)

Q1 Random variable x follows normal distb with mean = μ = 12 and SD = σ = 3. Find

(1) P(x ≤ 15) (2) P(10 ≤ x ≤ 13) (3) P(x > 14) (4) generate 5 observations (RN)

(1) pnorm(15, 12, 3)

[1] 0.8413447

> cat("P(x<=15) =", p)

[1] P(x<=15) = 0.8413447

> p2 = pnorm(13, 12, 3) - pnorm(10, 12, 3)

> cat("P(10<=x<=13) =", p2)

[1] P(10<=x<=13) = 0.3708066

> p3 = 1 - pnorm(14, 12, 3)

> cat("P(x>14) =", p3)

[1]P(x>14) = 0.2524925

> rnorm(5, 12, 3)

[1] 12.639812 13.041270 7.920397 9.888982 6.965470

x follows normal distribution with μ = 10, σ = 2.
Find
(1) P(x ≤ 7) (2) P(5 < x < 12)
(3) P(x > 12) (4) Generate 10 observations.
(5) Find k such that P(x < k) = 0.4)

> h1 = pnorm(7, 10, 2)

> h1 [1] 0.0668072

> h2 = pnorm(5, 10, 2) - pnorm(12, 10, 2)

> h2 [1] - ~~0~~.8351351

> h3 = 1 - pnorm(12, 10, 2)

> h3 [1] 0.1586553

> h4 = rnorm(10, 10, 2)

> h4

[1] 10.266550 10.596867 10.597316
9.045449 9.978724 5.756808
11.989820 5.833428 8.564714
11.123124

> h5 = qnorm(0.4, 10, 2)

> h5 [1] 9.493306

(3) Generate 5 random variables with mean = 15, SD = 4
Find: (1) sample mean, median, SD, print it

(4) x follows (X~
Find (1) P(x ≤ 40)
(ii) P(25 < x < 35)

aa..

(iv) Find K such that P(x < K) = 0.6

Q4

```
d1 = pnorm(40,30,10)
d1  [1] 0.8413447
d2 = 1-pnorm(35,30,10)
d2    [1] 0.3085375
d3 = pnorm(25,30,10) - pnorm(35,30,
d3    [1] -0.3829249
d4 = qnorm(0.6,30,10)
d4    [1] 32.53347
```

Q3

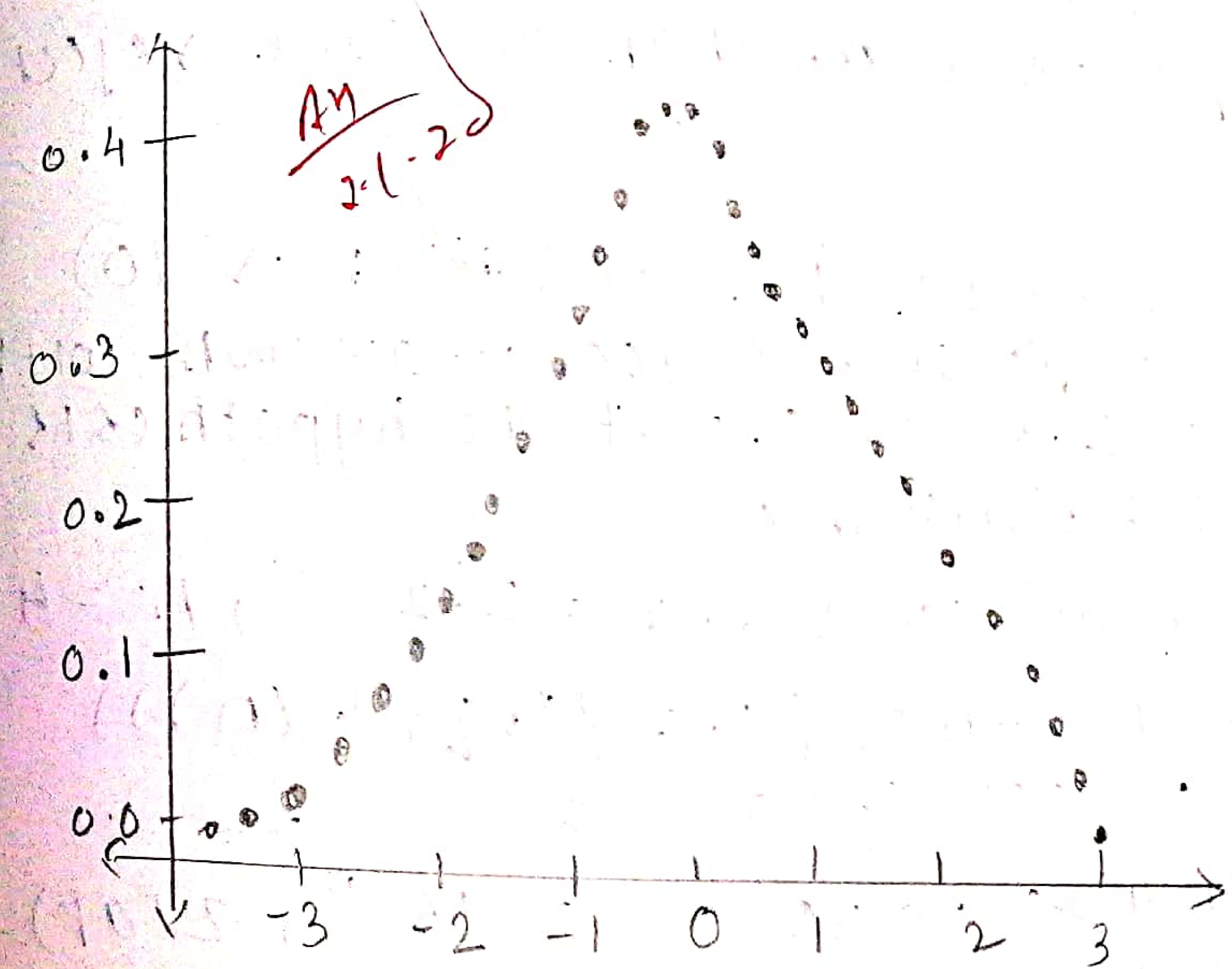
```
x = rnorm(5,15,4)
[1] 12.73337   12.30235  16.62240
    23.24364   13.08828
x = rnorm(5,15,4)
am = mean(x)
am   [1] 15.33276
me = median(x)
me   [1] 14.86556
n = 5
v = (n-1)* var(x)/n
v
2.574805
sd = sqrt(v)
sd
[1] 1.60462
```



```
cat("sample mean", am)
sample mean 15.33276
cat("sample median", me)
sample median 14.86556
cat("sample sd", sd)
sample sd 1.60462
```

Q5) Plot the standard normal graph.

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



An
21-2

Practical - 5

Aim:- Normal and t-test.

• Test the hypothesis (H0. μ = 15), H1 : μ ≠ 15 Random sample of size 400 is drawn and is calculated, the sample mean is 14 by SD - 3. Test the hypothesis at 5 % level of significance

Ans:- mo = 15 , mx = 14 , sd = 3 , n = 400

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

>zcal [1] -6.666667

> cat ("calculate value at z is =", z(al)

> calculate value 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 h0. μ = 15.

• Test the hypothesis (H0. μ = 10), H1(μ : ≠ 10) is drawn with Random sample of size 400 is drawn with mean = 10.2, SD - 2.25. Test the hypothesis at 5 % level of significance.

mo = 10 , mx = 10. 2 , sd = 2.25 , n = 400

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

> zcal = [1] 1.777778

> cat ("calculate value at z is ", zcal)

> calculate value at z is 1.777778

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

> pvalue

[1] 0.07544036

since p value more than 0.05 we accept h0. μ = 10

Test the hypothesis H0 : proportion of smokers in a college is 0.2 A sample is collected and is calculated, as 0.125 Test the hypothesis at 5 % level of significance (sample size - 400).

> P = 0.2 , p = 0.125 , n = 400 , Q = 1 - P

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

> zcal [1] -3.75

since

Last year farmer's lost 20 % of their crop. Random sample of 60 field are collected and it is found that 9 fields crops are insect polluted Test the hypothesis at 1 % Level of significance

P = 0.2 , p = 9 / 60 , n = 60 , Q = 1 - P

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

> pvalue [1] 0.3329216

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

> zcal [1] -0.9682458

Test the hypothesis H0 : μ = 12.5 from the foll sample at 5 % level of significance

~~P = 0.2~~

x = c(12.25 , 11.97 , 12.15 , 12.08 , 12.31 , 12.28 , 11.94 , 11.89 , 12.16 , 12.04)

8

```
n = length(x)
n = 10
mx = mean(x)
mx        [1] 12.107
varience = (n-1) * var(x)/n
varience          [1] 0.01952
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
```

AM 16.1.20

6

Aim:- Large sample test:-

68

Q1) Let the population mean (amount spent by customer in a restaurant) is 250). A sample of 100 cust selected. sample mean is calculated as 275 & SD-30. Test the hypothesis that the pop 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 the population proportion is 0.8 at 1 level of significance.

```
1) mo = 250 , sd = 30
mx = 275 , n=100
zcal = (mx - mo) / (sd / (sqrt(n)))
>zcal
[1] 8.3333
>pvalue = 2 * (1- pnorm(abs(zcal)))
>pvalue    [1] 0
>cat ("zcal :", zcal)
zcal : 8.3333
>cat (" pvalue :", pvalue)
pvalue : 0>

2) P = 0.8 , n = 1000
P = 0.8 , Q = 1- P
p = 750/1000
>zcal = (p - P) / sqrt(P * Q/n)
>zcal       [1] -3.952847
```

>pvalue = 2 * (1-pnorm(abs(zcal)))
>pvalue [1] 7.72268e-05

Q3 2 random sample of size 1000 & 2000 are drawn from 2 population with the same SD-2.5. The sample means are 67.5 & 68 resp. Test the hypothesis $H_0: \mu_1 = \mu_2$ at 5% LOS.

Q4 A study of noise level in 2 hospital is given below. Test the claim that the 2 hospitals have same level of noise at 1% LOS.

	Hosp A	Hosp B
size	84	34
mean	61.2	59.4
SD	7.9	7.5

Q5 In a sample of 600 students in a college, 400 use blue ink in another college from a sample of 900 students, 450 use blue ink. Test the hypothesis that the proportion of students using blue ink in 2 colleges are equal or not at 1% LOS.

Q3: n1 = 1000 , n2 = 2000
mx1 = 67.5, mx2 = 67.5
sd1 = 2.5 , sd2 = 2.5

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

> zcal [1] -5.163978
> cat("zcal:", zcal)
zcal : -5.163978

>pvalue = 2 * (1- pnorm(abs(zcal)))
>pvalue
[1] 2.417564e-07
>cat("pvalue :", pvalue)
pvalue : 2.417564e-07
p value > 0.05 value - accepted.

Q4)
n1 = 84 , n2 = 34
mx1 = 61.2 , mx2 = 59.4
sd1 = 7.9 , sd2 = 7.5
zcal = (mx1 - mx2) / sqrt((sd1^2/n1) + ((sd2^2/n2))
zcal [1] 1.162528
>pvalue = 2 * (1- pnorm(abs(zcal)))
>pvalue [1] 0.245021l

p value > 0.05 value - rejected

Q5.

$H_0: P_1 = P_2$ i.e H_1:
n1 = 600
n2 = 900
p1 = 400/600
p2 = 450/900
p = (n1 * p1 + n2 * p2) / (n1+n2)
p [1] 0.566667
q = 1 - p
>zcal (p1 - p2) / sqrt(p*q * (1/n1+1/n2))
> zcal [1] 6.391534

```
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.753222e-10
```

pvalue > 0.05, value-accepted

Q6

```
>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))
>pvalue=2*(1-pnorm(abs(zcal)))
>zcal
[1] 1.802741
>pvalue
[1] 0.07142888
```

pvalue > 0.05, value-accepted

23.1.20

Aim: small sample test

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

H0: There is no significance difference.

H0: $\mu = 66$

```
x=c(63,63,66,67,68,69,70,70,71,72)
>t.test(x)
One sample t-test
data: x
t=68.319, df=9, p value=1.558e-13
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
65.65171  70.14829
sample estimates
mean of x
67.9
```

p>0.5 ∴ Hence mean is rejected.

```
>if(1.558e-13>0.05){cat("accept H0")}else{cat("reject H0")}
>reject H0.
```

Q2 2 groups of students score the following marks. Test the hypothesis that there is no significant diff between the 2 groups.
grp1 - 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
grp2 - 16, 20, 14, 21, 20, 18, 13, 15, 17, 21
H' - There is no difference between 2 grps
>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 diff in means is not equal to 0
95 percent confidence interval.
0.1628205 5.037195
sample estimates:
mean of x mean of y
20.1 17.5
>if (pvalue > 0.05){cat("accept no")} else {cat("reject no")}
> reject no.

Q3 The sales data of 6 shops before & after.
Before: 53, 28, 31, 48, 50, 42
After: 58, 29, 30, 55, 56, 45
Test the hypothesis is effective or not.
There is no significance difference before and after the sales.
>x = c(53, 28, 31, 48, 50, 42)
>y = c(58, 29, 30, 55, 56, 45)
>t.test(x, y, paired = T, alternative = "greater")

Paired t-test
data: x and y
t = -2.7815, df = 5, p-value = 0.9806
alternative hypothesis: true difference in means is greater than 0.
95 percent confidence interval:
-6.035547 Inf
sample estimates:
mean of the differences
-3.5
>if (0.9806 > 0.05){cat("accept no")} else {cat("reject no")}
>accept number

Q4 Two medicines are applied to two groups of patients resp.
Grp 1:- 10, 12, 13, 11, 14
Grp 2:- 8, 9, 12, 14, 15, 10, 9
Is there any significant diff between 2 medicine?
Ho:
x = c(10, 12, 13, 11, 14)
y = c(8, 9, 12, 14, 15, 10, 9)
Ho:- Ther is no significance difference.
>t.test(x, y))
>p value - 0.4406.
> if (0.4406 > 0.05) cat("accept no") else {cat("reject no")}

> The the value is accepted.

Q5 The foll are the weight before & after the diet program. Is the diet prog effective

before: 120, 125, 115, 130, 123, 119

after: 100, 114, 95, 90, 115, 99

H0: There is no significance difference.

```
> x = c(120, 125, 115, 130, 123, 119)
> y = c(100, 114, 95, 90, 115, 99)
> t.test(x, y, paired = T, alternative = "less")
> data: x and y
t = 4.3458 , df = 5 , pvalue = 0.9963
alternative hypothesis : true differences in
means is less than 0
95 percent confidence interval:
-Inf  29.0295
sample estimates:
mean of the differences
        19.83333
> if (0.9963 > 0.05) { cat ("accept H0") } else
cat ("reject H0") } Q.
> reject H0
```

Practical - 8

Aim: Large and small sample test.

H0: μ = 55

H1: μ ≠ 55

n = 100

mx = 52, m0 = 55, sd = 7.

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

It is smaller the 0.05 ∴ Hence rejects the Ho

H0: μ = 1/2

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

pvalue > 0.05 we accept.

55

H0 : p = P2

(13) n1 = 1000

n2 = 1500

p1 = 20/1000

p2 = 15/1500

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

>p [1] 0.014

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

>z(al [1] 2.081283

>z(al = 1-pnorm(abs(z(al))

> pvalue [1] 0.018704

>pvalue = 2*(1-pnorm(abs(z(al)))

> pvalue

[1] 0.03740801

pvalue < 0.05 we reject

H0: μ = 100

(14) n = 100

mx = 99, m0 = 100

var = 64, sd = 8

>z(al = (mx-m0)/(sd/(sqrt(n)))

>z(al [1] -2.5

>pvalue = 2*(1-pnorm(abs(z(al)))

> pvalue [1] 0.01241933

pvalue < 0.05 we reject.

H0 : x = c(63, 63, 68, 69, 71, 71, 72)

> t.test(x)

One sample t-test

data: x

t = 47.94, df = 6, p-value = 5.522e-09

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

64.66479 71.62092

sample estimates:

mean of x

68.14286

pvalue > 0.05 Hence rejects

H0 : μ = 1200

(7) m0 = 1200, mx = 1150

sd = 125, n = 100

z(al = (mx-m0)/(sd/(sqrt(n)))

z(al [1] -4

pvalue = 2*(1-pnorm(abs(z(al)))

pvalue [1] 6.33429e-05

pvalue $\not<$ 0.05 it rejects.

) H0: σ1 = σ2

x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)

y = c(64, 66, 74, 78, 82, 85, 87, 92, 93, 95

var.test(x, y)

F test to compare two variances

data: x and y
F = 0.70686, num df = 9, denom df = 10, p-value = 0.6359.
alternative hypothesis: true ratio of variances
95 percent confidence interval:
0.1833662 3.0360393
sample estimates:
ratio of variences. 0.7068567.

pvalue > 0.05 we accept.

108) $H_0 : P_1 = P_2$

> n1 = 200 n2 = 300
p1 = 44/200 p2 = 56/300
p = (p1 * p1 - n2 * p2) / (n1 + n2)
q = 1 - p
q [1] 0.2
> z(a) = (p1 - p2) / sqrt (p * q * (1/n1 + 1/n2))
> z(a) [1] 0.9128709
> pvalue = 2 * (1 - pnorm(abs(z(a))))
> pvalue
[1] 0.3613104.
pvalue > 0.05 we accept.

Practical 9

Aim:- chi square and ANOVA

> use the following data to test whether the condition of home & child are independent or not.

	Condition of home clean	dirty
condition child	70	50
dirty child	80	20
dirty child	35	45

H_0: condition of home and child are indepen

Soln:
> 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 H'.

(Q2) Test hypothesis that vaccination and disease are dependent or not

		VACCINE Affected	Not affected
DISEASE	Affected	70	46
	Not Affected	35	37

Disease & vaccine are independent

x = c(70, 35, 46, 37)

m = 2, n = 2

y = matrix(x, nrow = m, ncol = n)

	[,1]	[,2]
[1,]	70	46
[2,]	35	37

pv = chisq.test(y)

pv

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

data: y

x-squared = 2.0275, df = 1, p-value = 0.1545

pvalue > 0.05, hence we reject.

3) Perform a ANOVA for a foll data.

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

H0: The 'mean' are equal of A, B, C, D

x1 = c(50, 52)

x2 = c(53, 55, 53)

x3 = c(60, 58, 57, 56)

x4 = c(52, 54, 54, 55).

>d = stack (list (b1=x1, b2=x2, b3=x3, b4=x4))
>names(d)
[1] "values" "ind"
>oneway.test(values~ind, data=d, var.equal = T)
One-way analysis of means

data: values and ind
F = 11.735, num df=3, denom df=9, p-value=0.01
>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	19.17	2.019		

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

>pvalue > 0.05, Hence we reject.

Q4 The foll data gives the life of the tyres of 4 brands.

Type	Life
a	20, ~~23~~, 18, 17, 19, 22, 24
b	19, 15, 17, 20, 16, 17
c	21, 19, 22, 17, 20
d	15, 14, 16, 19, 14, 16

H0: Average life of tyres are equal.
>x1 = c(20, 23, 18, 17, 19, 22, 24)
>x2 = c(19, 15, 17, 20, 16, 17)
>x3 = c(21, 19, 22, 17, 20)
>x4 = c(15, 14, 16, 19, 14, 16)

> d = stack(list(b1=x1, b2=x2, b3=x3, b4=x4))
> names(d)
[1] "values" "ind")
> one-way analysis of means

data: values and ind
F = 324.68, num df = 3, denom df = 18,
p-value = 7.45e-11.

Q5

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

>print(x)

x & statistics , x & maths)

0618'

AY
20.02.20

78

naths

0

8

17

.0

.5

7

.7

)

tics)

ths)

40r

62'

Practical - 10

(where parent value is not given).

Aim:- Non-parametric test.

Q1 Foll are the amounts of sulphur oxides emitted by some industry in 20 days. Applied sign test to test the hypothesis that the population median is 21.5 at 5% LOS.

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

H0: 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, 26)
>me = 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
```

pvalue > 0.05, Hence we accept the H0

NOTE:- If the alternative is me ≠ or me <

H1: me >

pv = pbinom (sn, n, 0.5)

Q2) Following is the data of 10 observations apply sign test to test the hypothesis that the population median is 625 against the alternative is more than 625.

H0:

```
x = c(612, 619, 631, 628, 643, 640, 655, 649, 670,
      663)
```

H0: Population median is 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
```

> pvalue > 0.05 Hence we accept the val

Q3) Following are the values of a sample. Test the hypothesis that the population median is 60 against the alternative it is more than 60 at 5% LOS, using wilcoxon signal Rank test

H0: Population median is 60.

```
x = c(63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39,
      72, 69, 48, 66, 72, 63, 87, 69)
> wilcox.test (x, alter = "greater", mu = 60)
```

65

wilcoxon signed rank test with continuity correction :

data: x
V = 145 , p-value = 0.02298
alternative hypothesis: true location is greater than 60

p value < 0.05 we reject the H0

NOTE: If the alternative < - altr = "less"
alternative ≠ altr = "two sided"

Q4 using WSRT ; test the population median is 12 or less than 12

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

> willcox.test(x, altr = "less", mu = 12)

Wilcoxon signed rank test with continuity correction :

data : x
V = 16 , p-value = 0.9986
alternative hypothesis : true location is less than 12
pvalue > 0.05 we accept H0 !

Q5 The weights of a students before and after, they stopped smoking. Using WSRT, test that there is no significant change. 80

weights before:	weights after
65	72
75	74
75	72
62	66
72	73

H0: Befor & After there is no change.
H1 : change

x = c(65, 75, 75, 62, 72)
y = c(72, 74, 72, 66, 73)
willcox.test(d, altr = "two.sided", mu = 0)

wilcoxon signed rank test with continuity correction

data : d

V = 4.5 , p-value = 0.4982
alternative hypothesis: true location is not equal to 0.

~~P~~ pV > 0.05 , we accept H0.

AM
27.02