28/11/19

PRACTICAL-01

Topic :- Basics of R software.

- R is a software for statistical analysis and data computing.
- It is an effective data handing software and outcome storage is possible.
- It is capable of graphical display.
- It is a free software.

Q1) Solve the following :-

1) $4 + 6 + 8 \div 2 - 5$

2) $2^2 + |-3| + \sqrt{45}$

3) $5^3 + 7 \times 5 \times 8 + 46/5$

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

5) round of $46 \div 7 + 9 \times 8$

> 4+ 6+ 8/2 -5
[1] 9
> 2^2 + abs(-3) + sqrt (45)
[1] 13.7082
> 5^3 + 7 * 5 * 8 + 46/5
[1] 414.2
> sqrt (4^2 + 5 * 3 + 7/6)
[1] 5.67
> round (46/7 + 9 * 8)
[1] 79.

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

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

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

9) c(6,2,7,5) / c(4,5)
> c(6,2,7,5) / c(4,5)
[1] 1.5 0.4 1.75 1

10) $x=20, y=30, z=2.$
find i) x^2+y^3+z
ii) $\sqrt{x^2+y}$, iii) x^z+y^z.

```
>x=20
>y=30
>z=2
>x^2+y^3+z
[1]27402
>sqrt(x^2+y)
[1]20.73
>x^z+y^z
[1]1300
```

11) >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

12) Find $x+y$ 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(nrow=3, ncol=3, data=c(4,7,9,-2,0,-5,6,7,3))
>y=matrix(nrow=3,ncol=3, data=c(10,12,15,-5,-4,-6,7,9,5))
>x+y
```

	[,1]	[,2]	[,3]
[1,]	14	-7	13
[2,]	19	-4	6
[3,]	24	-11	8

>2*x+3*y

	[,1]	[,2]	[,3]
[1,]	38	-19	33
[2,]	50	-12	41
[3,]	60	-28	21

SEO

13) Marks of Stats of CS division batch-A :- 59,20,35, 24,46,56, 55, 45, 27, 22, 47, 58, 54, 40, 50, 32, 36, 29, 35, 39

```
> x=c(59,20,35,24,46,56,55,45,27,22,47,58,54,40,
      50,32,36,29,35,29)
> breaks = seq(20,60,5)
> a=cut(x,breaks, right=FALSE)
> b = table(a)
> c = transform(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

Arun 28/11

PRACTICAL-02

Aim :- Probability Distribution.

Q1) Check whether the following are pmf or not :-

i)

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

→ Since, $P(2) = -0.5$
Therefore, it cannot be a pmf. as in pmf $P(x) \geq 0 \; \forall x$.

ii)

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

```
→ > prob = c(0.2, 0.2, 0.3, 0.2, 0.2)
  > sum(prob)
  [1] 1.1
```

It cannot be a pmf as in pmf $\sum P(x) = 1$

iii)

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

```
> p = c(0.2, 0.2, 0.35, 0.15, 0.1)
> sum(p)
[1] 1.
```

$\because \sum P(x) = 1$, it is a pmf and $P(x) \geq 0$

Q2) i) Find the cdf for the following pmf and sketch the graph.

i)

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

```
> x = c(10, 20, 30, 40, 50)
> prob= c(0.2, 0.2, 0.35, 0.15, 0.1)
> sum(prob)
[1] 1
> cumsum (prob)
[1] 0.2   0.4   0.75   0.9   1.0
> plot (x, cumsum(prob), "s")
```

$F(x) = 0 \quad x < 10$

$0.2 \quad 10 \le x < 20$

$0.4 \quad 20 \le x < 30$

$0.75 \quad 30 \le x < 40$

$0.9 \quad 40 \le x < 50$

$1.0 \quad x \geq 50.$

ii)

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

```
>prob= c(0.15, 0.25, 0.1, 0.2, 0.2, 0.1)
> cumsum(prob)
>[1] 0.15  0.40  0.5  0.7  0.9  1.
>x= 1:6
```

$F(x) = 0 \quad x < 10$

$0.15 \quad 1 \le x < 2$

$0.4 \quad 2 \le x < 3$

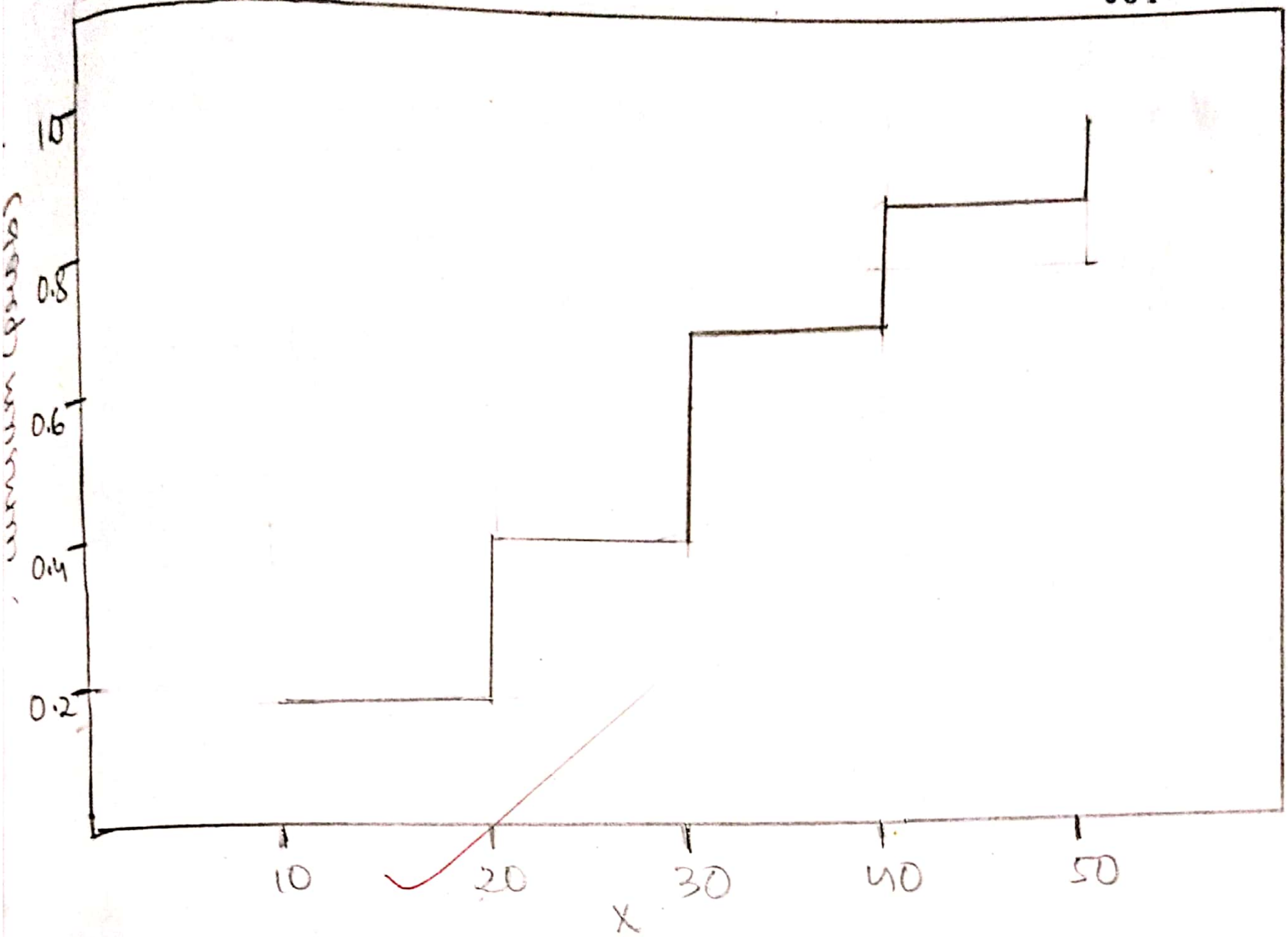
$0.5 \quad 3 \le x < 4$

$0.7 \quad 4 \le x < 5$

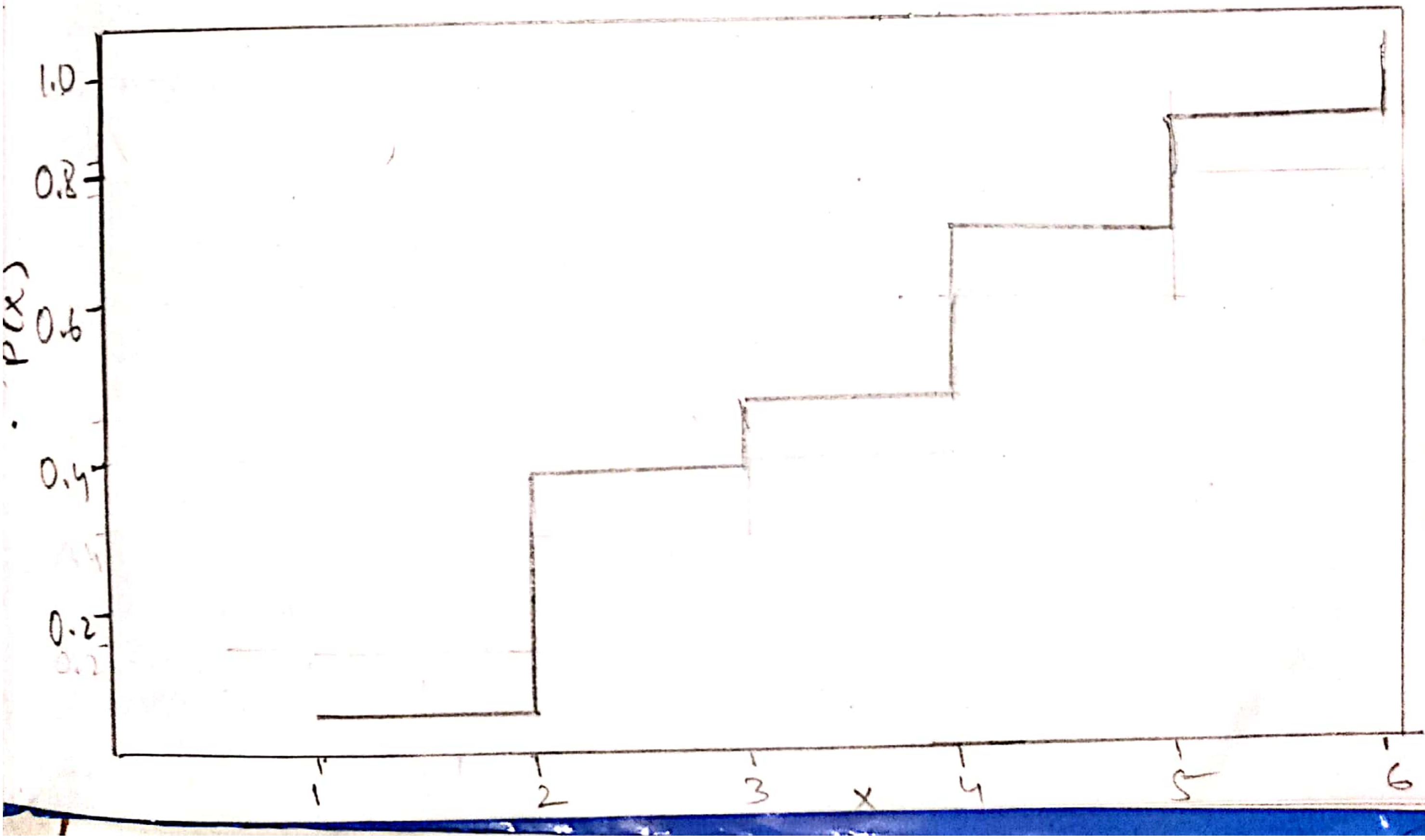
$0.9 \quad 5 \le x < 6$

$1 \quad x \geq 6.$

Q2 i)



Q2) ii) >plot(x, cumsum(prob), "s", xlab="x", ylab=" P(x)")



Q3) Check whether the following is p.d.f or not:-

i) $F(x) = 3 - 2x, \quad 0 \le x \le 1$

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

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

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

$= 3[x]_0^1 - \cancel{2}\left[\frac{x^2}{\cancel{2}}\right]_0^1$

$= 3(1-0) - (1-0)$

$= 3 - 1 = 2 \neq 1$

$\therefore$ It is not a p.d.f

ii) $F(x) = 3x^2\,dx, \quad 0 < x < 1$

$\Rightarrow \int_0^1 F(x)\,dx$

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

An 5/12/19

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

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

$= 1 - 0 = 1$

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

19/12/19

PRACTICAL - 3

Aim :- To study binomial distribution.

i) $P(X = x)$ = dbinom (x, n, p)
ii) $P(X \leq x)$ = pbinom (x, n, p)
iii) $P(X > x)$ = 1 - pbinom (x, n, p)
iv) If x is unknown,
$P_1 = P(X \leq x)$ ---- (given)
qbinom (P_1, n, p)

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

Q2) Suppose there are 12 MCQ. Each questions has 5 options out of which 1 is correct. Find the probability of :-
i) Exactly 4 correct answers.
ii) Atmost 4 correct answers.
iii) More than 5 correct answers.

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

Q4) $n = 12$, $p = 0.25$. Find:-
i) $P(X = 5)$, ii) $P(X \leq 5)$, iii) $P(X > 7)$, iv) $P(5 < X < 7)$

Q1) >dbinom(10,100,0.1]
[1] 0.1318653.

Q2) i)>dbinom (4,12,1/5)
[1] 0.1328756.

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

iii)>1-pbinom(4,12,1/5)
[1] 0.0725555.

Q3) >dbinom(0:5,5,0.1)
[1] 0.59049
0.32805
0.07290
0.00810
0.00045
0.00001.

Q4) 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] 0.0027815.

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

Q5) Probability of salesman making a ~~sale~~ sell to customer 0.15. Find the probability of :-
i) no sells out of 10 customer
ii) more than 3 sells out of 20 customer

Q6) A salesman has a 20% probability of making a sell to a customer. Out of 30 customers, what minimum no. of sells he can make with 88% probability.

Q7) X followed binominal distribution with $n = 10$, $p = 0.3$. Plot the graph of pmf and ~~cdf~~ cvf.

Q8) Answers :-

Q5) i) > dbinom(0,10,0.15)
[1] 0.1968744
ii) 1-pbinom(3,20,0.15)
[1] 0.3522748.

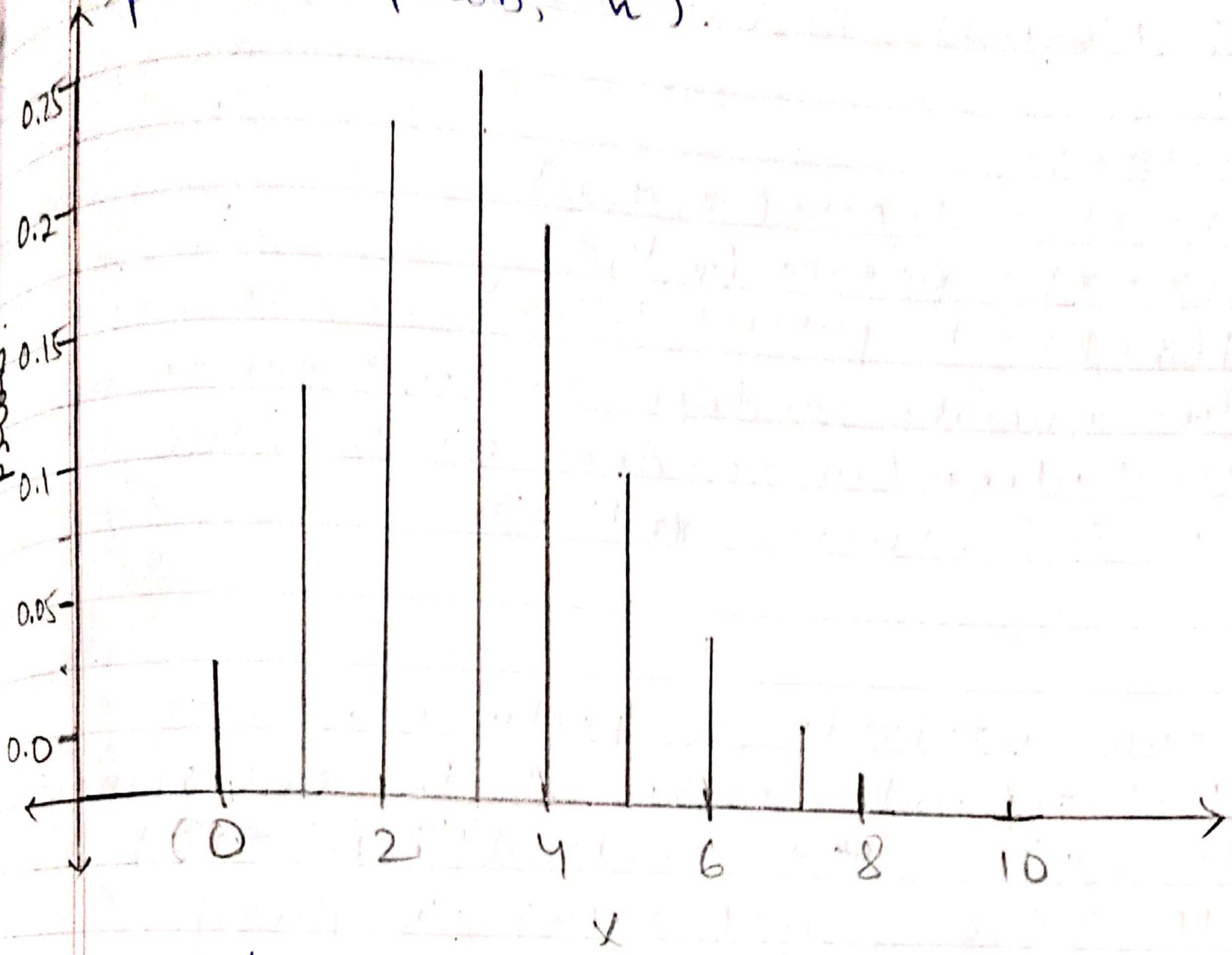
Q6) > qbinom(0.88, 30, 0.2)
[1] 9.

Q7) > 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)

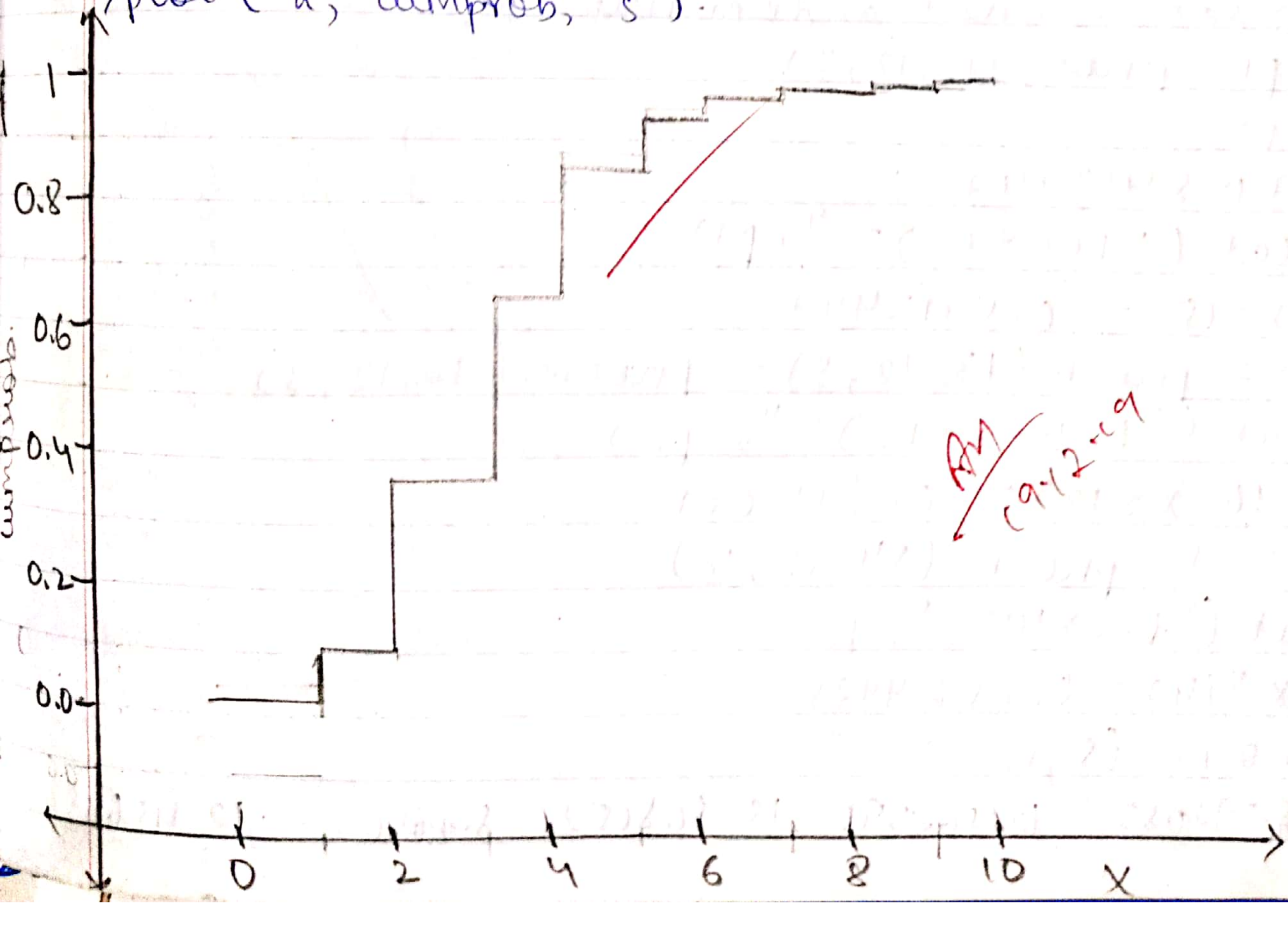
	xvalues	probability
1	0	0.0282475249
2	1	0.1210608210
3	2	0.2334744405
4	3	0.2668279320
5	4	0.2001209490
6	5	0.1029193452
7	6	0.0367569090
8	7	0.0090016925
9	8	0.0014467005
10	9	0.0001377810
11	10	0.0000059049

Scanned with CamScanner

>plot (x, prob, "h").



>plot (x, cumprob, "s").



AM
19-12-19

2/1/2020 087

PRACTICAL-4

Topic :- Normal Distribution

Formulas :-

i) $P(X = x) = dnorm(x, \mu, \sigma)$

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

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

iv) Two generate random number from normal distribution (in random number), the R code is :- $rnorm(n, \mu, \sigma)$

Q1) A random variable x follows normal distribution with mean $= \mu = 12$ and standard deviation $= \sigma = 3$. find i) $P(X \leq 15)$, ii) $P(10 \leq X \leq 13)$, iii) $P(X > 14)$, iv) generic 5 observations (random numbers).

→)

```
>p1= pnorm(15,12,3)
>p1
[1] 0.8413447
>cat ("P(X≤15)=", p1)
P(X≤15)= 0.8413447
>p2 = pnorm(13, 12,3) - pnorm (10,12,3)
>cat ("P(10≤X≤13)=", p2)
P(10≤X≤13)= 0.3780661
>p3 = 1- pnorm (14,12,3)
>cat ("P(X>14)=", p3)
P(X>14)= 0.2524925
>rnorm (5,12,3)
[1] 8.592085  12.862390  15.568052  8.7696  12.456383
```

Scanned with CamScanner

Q2) X follows normal distribution with $\mu = 10$, $\sigma = 2$. Find i) $P(X <= 7)$, ii) $P(5 < X < 12)$, iii) $P(X > 12)$, iv) generate 10 obs; v) find k such that $P(X < k) = 0.4$

038

→ >pnorm(7,10,2)
[1] 0.0668072
>pnorm(12,10,2) - pnorm(5,10,2)
[1] 0.835135
> 1-pnorm(12,10,2)
[1] 0.158655
>rnorm(10,10,2)
[1] 7.91262 8.759731 9.313712 11.896202 10.675737
10.737547 11.728618 9.526577 12.613503 11.2564
>qnorm(0.4,10,2)
[1] 9.493306.

Q3) Generate 5 random no.s from normal distribution $\mu = 15$, $\sigma = 4$. Find sample mean, median, standard deviation and print it.

Q4) X follows normal ($\mu = 30$, $\sigma^2 = 100$). Find i) $P(X \leq 40)$, ii) $P(X > 35)$, iii) $P(25 < X < 35)$, iv) find k such that $P(X < k) = 0.6$.

Answers:-

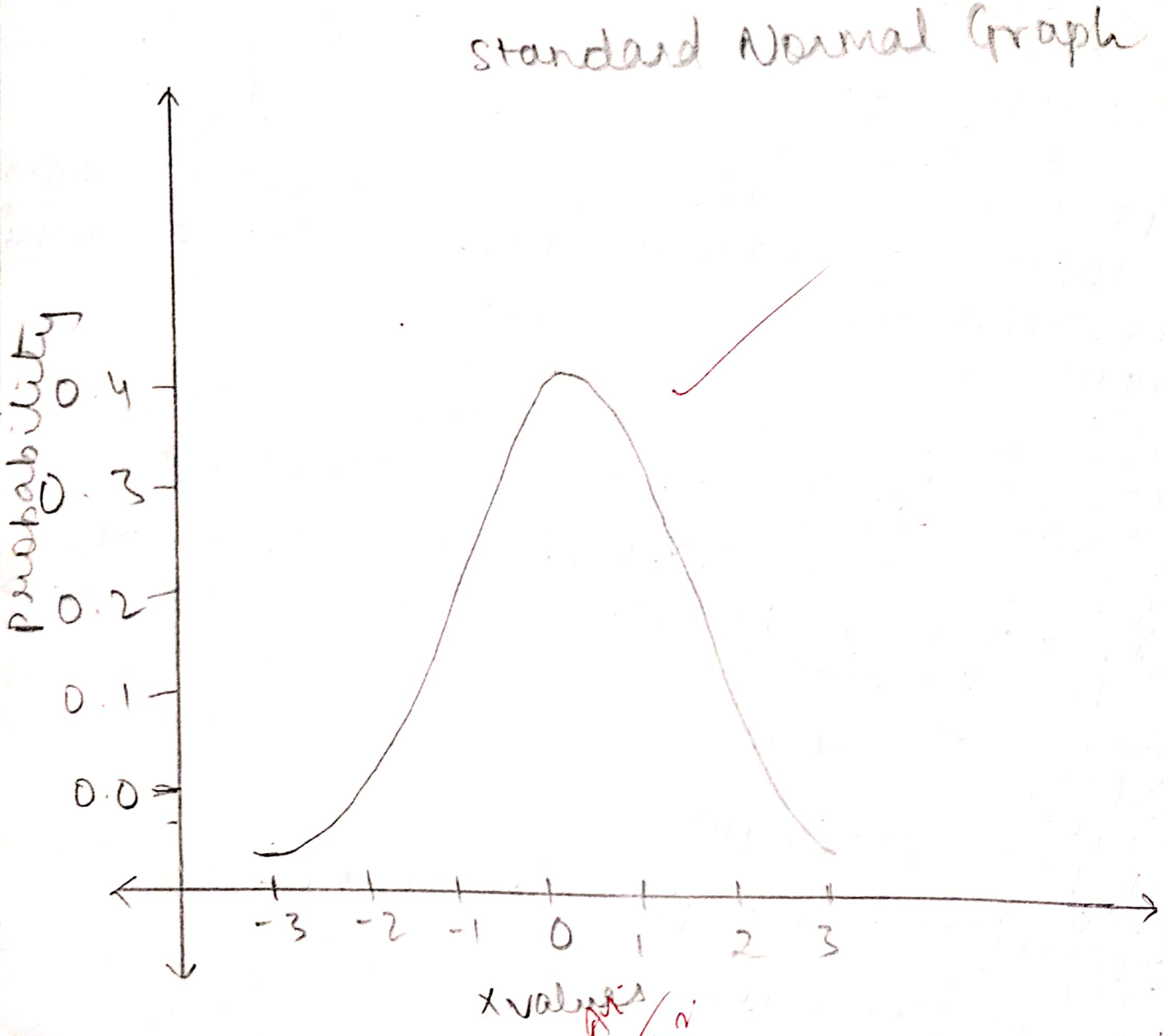
Q4) > pnorm(40,30,10)
[1] 0.8413447
>1- pnorm(35,30,10)
[1] 0.3085375
>pnorm(35,30,10)- pnorm(25,30,10)
[1] 0.3829249
>qnorm(0.6,30,10)
[1] 32.53347

Q3) > rnorm(5,15,4)
[1] 22.32830 12.89411 15.77568 14.22375 23.59594
>x= rnorm(5,15,4)
>x
[1] 14.93497 14.65413 12.48667 18.29222 18.43272
>am=mean(x)
> md=median(x)
>n=5
>variance=(n-1)*var(x)/n
>sd=sqrt(variance).

>cat("Mean=",am)
Mean= 15.76014
>cat("Median=",md)
Median= 14.93497
> cat("Standard Deviation=",sd)
Standard Deviation=2.287965

Q5) Plot 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")
```



AR
2.1.2

PRACTICAL-5

Aim:- Normal & T-Test.

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

→
```
> m=15 ; m1=14 ; sd=3 ; n=400 ;
> zcal = (m1-m) / (sd /(sqrt(n)))
> zcal
[1] -6.666667.
> pvalue = 2 * (1- pnorm(abs(zcal)))
> pvalue
[1] 2.616796e-11
```
∵ The pvalue is less than 0.5, we reject the value $H_0 : \mu = 15$

Q2) Test the hypothesis $H_0 : \mu = 10$ against $H_1 : \mu \neq 10$. Random sample of size 400 is drawn with sample mean 10.2 and standard deviation 2.25. The Test the hypothesis at 5% level of significance.

→
```
> m=10 ; m1= 10.2 ; sd= 2.25 ; n= 400 ;
> zcal = (m1- m) / (sd / (sqrt(n)))
> zcal
[1] 1.777778
> pvalue = 2 * (1- pnorm(abs(zcal)))
> pvalue
[1] 0.07
```

Since the pvalue is greater than 0.05, we accept the value $H_0: \mu = 10$.

3) Test the hypothesis H_0: proportion of smokers in a college is 0.2. A sample is collected and it 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
>pvalue= 2*(1- pnorm(abs(zcal)))
>pvalue
[1] 0.0001768346
```
Since the pvalue is less than 0.05, we reject the value H_0: proportion of smokers as 0.2.

4) Last year farmers lost 20% of their crop. A random sample of 60 ~~fitted~~ field are colletted and it was 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
>zcal= (p-P)/(sqrt(P*Q/n))
>zcal
[1] -0.96
>pvalue= 2*(1 - pnorm(abs(zcal)))
>pvalue
[1] 0.33
```
Since the pvalue is greater that 0.01, we ~~to~~ accept the value of H_0.

5) Test the hypothesis $H_0: \mu = 12.5$ from the following sample at 5% level of significance.

040

```
> 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)
> var = (n-1) * var(x)/n
> sd = sqrt(var)
> m0 = 12.5
> t = (mx - m0) / (sd / sqrt(n))
> t
[1] -8.89
> pvalue = 2 * (1 - pnorm(abs(t)))
> pvalue
[1] 0.
```

Since the pvalue is less than 0.05, value is rejected.

AM 16/1/20

PRACTICAL-06

Topic :- ~~lab~~ large sample test.

Q1) Let the population mean (the amount spent by customer in a restuarant) is 250. A sample of 100 customers selected. The sample mean is collected as ~~250~~ 275 and the standard deviation is 30. Test the hypothesis that the population is 250 at 5% level of significance.

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

Q1) $H_0: \mu = 275$ against $H_1: \mu \neq 275$

```
>sd=30; m=250; m1=275; n=100;
>zcal = (m1-m)/(sd/sqrt(n))
>zcal
[1] 8.33
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0
```

∵ The value is less than 0.05, we ~~fo~~ reject H_0 at 5% level of significance.

Q2) > $H_0: \mu = 0.8$ against $H_1: \mu \neq 0.8$

```
> P=0.8; p=750/1000; n=1000; Q=1-P
>zcal= (p-P) / sqrt (P*Q/n)
```

> zcal
[1] -3.952847
>pvalue = 2 * (1- pnorm(abs(zcal)))
>pvalue
[1] 7.72268e-05

Since, the pvalue is less than 1% level of significance, we reject the hypothesis.

Q3) Two random sample of size 1000 & 2000 are drawn from two population with the ~~sample~~ same sd 2.5. The sample means are 67.5 & 68 respectively. Test the Hypothesis $H_0: \mu_1 = \mu_1$ against $H_0: \mu_1 \neq \mu_1$ at 5% level of significance.

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% level of significance.

	Size	Mean	SD
HOS.A ~~Size~~	84	61.2	7.9
HOS.B	34	59.4	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% level of significance.

Q3) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

```
>n1=1000; n2=2000; m1=67.5; m2=68; sd1=2.5; sd2=2.5;
>zcal = (m1-m2)/sqrt((sd1^2)/n1) + (sd2^2/n2))
>zcal
```

[1] ~~1.162528~~ -5.163978

```
>pvalue = 2* (1 - pnorm(abs(zcal)))
>pvalue
[1] 2.417564e-07.
```

Since, the pvalue is less than 5% level of significance, we reject the hypothesis.

Q4)

```
>n1= 84; n2=34; m1=61.2; m2=59.4; sd1=7.9; sd2=7.5;
>zcal = (m1 - m2)/(sqrt((sd^2/(n1)) + (sd^2/n2)).
>zcal
[1] 1.162528
>pvalue = 2*(1- pnorm(abs(zcal)))
>pvalue
[1] 0.2450211
```

∵ The pvalue is greater than 1% level of significance, we accept the hypothesis.

Q5)

```
>n1= 600; n2= 900; p1= 400/600; p2= 450/900;
>p= (n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.5666667
>q=1-p
>q
[1] 0.433333
>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 pvalue is less than 1% level of significance, we reject the hypothesis

Q6) $H_0: \mu_1 = P_1 = P_2$ against $H_1: \mu_1 = P1 \neq P2$

n1 = 200; n2 = 200; p1 = 44/200; p2 = 30/200;

```
>p = (n1 * p1 + n2 * p2)/(n1 + n2)
>p
[1] 0.185
>q = 1-p
[1] 0.815
>zcal = (p1-p2)/sqrt(p* q* (1/n1 + 1/n2))
>zcal
[1] 1.802741
>pvalue = 2*(1- pnorm(abs(zcal)))
>pvalue
[1] 0.07
```

$\because$ Pvalue is greater than 0.07, we accept the hypothesis.

A.K.
27.1.20

PRACTICA-07

Topic :- 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 the population with the average 66.

→ $H_0 : \mu_0 = 66$

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

>t.test (x)

~~pr.~~ One sample t-test

data: x

t = 68.319, df = 9, pvalue = 1.558e^-13.

alternative hypothesis: true mean is not ~~not~~ equal to 0.

95 percent confidence interval:

65.65171 70.14829

sample estimates:

mean of x

67.9

∵ The pvalue is less than 0.5, we ~~f~~ reject the hypothesis at 5% level of significance.

Q2) Two groups ~~are~~ of student score the following marks. Test the hypothesis that there is no significant difference between the two groups.

Group1 :- 18, 22, 21, 17, 20, 17, 23, 20, 22, 21.

Group2 :- 16, 20, 14, 21, 20, 18, 13, 15, 17, 21.

→ H_0 :- No difference between the two groups.

E.P.O.

```
>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, p-value = 0.03798.

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

95 percent confidence interval:

0.1628205 5.0371795

sample estimates:

Mean of x	mean of y
20.1	17.5

∵ Pvalue is less than 0.5, we reject the hypothesis at 5% level of significance.

Q3) The sales of 6 shops before & after are special campain given below.

Below:- 53, 28, 31, 48, 50, 42,

After:- 58, 29, 30, 55, 56, 45

~~The~~ Test the hypothesis that the campain is effective or not.

→ Ho: ~~In~~ There is no significance difference before and after the campain.

```
>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, 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 differences
-3.5

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

Q4) Two medicines are applied to two groups of patient respectively.
Group1:- 10, 12, 13, 11, 14.
Group2:- 8, 9, 12, 14, 15, 10, 9.
Is there any significance between the two medicines?

Q5) Following are the weights before & after the diet program. Is the diet program effective?
Before: 120, 125, 115, 130, 123, 119.
After:- 100, 114, 95, 90, 115, 99.

Ho: No significant difference
Q4) >x= c(10,12,13,11,14)
>y=c(8,9,12,14,15,10,9).
>t.test (x, y)

Welch Two sample t-test.
data: x and 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:
-1.781171 3.781171
sample estimates:
mean of x mean of y.
12 11.
∴ Pvalue is greater than 0.05, we accept the hypothesis at 5% level of significance.

Scanned with CamScanner

H_0:- diet program ~~effective~~ not effective.

Q5) > x = c(120,125,115,130,123,119)

> y = c(100,114,95,90,115,99).

> t.test(x, y, paired = T, alternative = "less")

Paired t-Test

data: x and y

t = 4.3458, df = 5, pvalue = 0.9963

alternative hypothesis: true difference in means is less than 0.

95 percent confidence interval:

-Inf 29.0295

sample estimates:

mean of the differences

19.83333

A.M

6-7-20

Scanned with CamScanner

Practical 08

Large and Small Sample tests

Q.1 The arithmetic mean of a sample of 100 items from a large population is 52. If the standard is 7, test the hypothesis that the population mean is 55 against the alternative it is more than 55 at 5% LOS.

Q.2 In a big city 350 out of 700 males are found to be smokers. Ddoes the information supports that exactly half of the males in the city are smokers ? Test at 1% LOS.

Q.3 Thousand article from a factory:A are found to have 2% defectives ,1500 atricles from a 2nd factory:B are found to have 1% defective. Test at 5% LOS that the two factory are similar are not.

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

Q.5. The flower stems are selected and the heights are found to be(cm) 63,63,68,69,71,71,72 test the hypothesis that the mean height is 66 or not at 1% LOS.

Q.6. Two random samples were drawn from 2 normal populations and there values are A-66,67,75,76,82,84,88,90,92 B-64,66,74,78,82,85,87,92,93,95,97. Test whether the population have the same variance at 5% LOS.

7. A company producing light bulbs finds that mean life span of the population of bulbs is 1200 hours with s.d. 125. A sample of 100 bulbs have mean 1150 hours. Test whether the difference between population and sample mean is significantly different?

8. From each of two consignments of apples, a sample of size 200 is drawn and number of bad apples are counted. Test whether the proportion of rotton apples in two assignments are significantly different at 1 % LOS.?

	Sample size	No. of bad apples
Consignment 1	200	44
Consignment 2	300	56

PRACTICAL-8

Aim:- Large & small sample tests.

H_0 :- $\mu = 55$

```
>n=100
>m=52
>sd=7
>M1=55
>zcal=(M1-m)(sd/sqrt(n))
>zcal
[1]4.285714
>pvalue=2*(1- pnorm(abs(zcal)))
>pvalue
[1]1.82153e-05
```

∵ Pvalue is less than 0.05 we reject the hypothesis at 5% level of significance

H_0 :- $\mu = 1/2$.

```
>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 is greater than 0.5, we accept the hypothesis at 1% level of significance.

280

Q3) $H_0: P_1 = P_2$

> n1 = 1000

> n2 = 1500
> p1 = 20/1000
> p2 = 15/1500
> p = (n1 * p1 + n2 * p2) / (n1+n2)
> p
[1] 0.014
> zcal = (p1 - p2) / sqrt (p * q * (1/n1 + 1/n2))
> zcal
[1] 2.081283
> pvalue = 2*(1 - pnorm (abs (zcal)))
> pvalue
[1] 0.03740801
∴ Pvalue is less than 0.05, we ~~a~~ reject the hypothesis ~~Ho~~ at 5% level of significance.

Q4) $H_0: \mu = 100$
> n = 400
> m = 99
> m1 = 100
> sd = 64
> zcal = (m - m1) / (sd / sqrt (n))
> zcal
[1] - 2.5
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 0.0124193

∵ Pvalue is less than 0.05, we reject the hypothesis.

H_0: μ = 66

046

Q8) > x = c(63,63,68,69,71,71,72)

> t.test(x)

One sample t-test

data: x

t = 47.94, df = 6, pvalue = 5.22e-09.

alternative hypothesis: true mean is not equal to 0

sample estimates:

mean of x

68.14286

∵ Pvalue is less than 0.01, we reject the hypothesis at 1% level of significance.

Q7) H_0: - μ = 1200

>n = 100

>mx = 1150

>m0 = 1200

>sd = 125

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

>zcal

[1] -4

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

>pvalue

[1] 6.334248e-05

∵ Pvalue is less than 0.05, we reject the hypothesis

H_0: $\sigma_1 = \sigma_2$

Q6) > 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 variances

data: x & y

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

alternative hypothesis: true ratio of variance is ≠ 1.

sample estimates:

ratio of variances

0.7068767.

∵ Pvalue is less than 0.05, we ~~at~~ ~~reject~~ accept the hypothesis at 5% level of significance.

Q8) $H_0: P_1 \neq P_2$

```
> n1=200
> n2=300
> p1=44/200
> p2=56/300
> p=(n1*p1+n2*p2)/(n1+n2)
> p
[1] 0.2
> q=1-p
> zcal=(p1-p2)/sqrt(p*q*(1/n1+1/n2))
> zcal
[1] 0.9128
> pvalue=2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.3613
```

∵ Pvalue is ~~greater~~ more than 0.05, hence it accepts. Ho

AH
13.7.2(

PRACTICAL-09

Topic :- Chi-square tests and ANOVA.

Q1) Use the following data to test whether the condition of the home and child are independent

Cond^n of Child		Cond^n of Home: Clean	Dirty
	Clean	70	50
	Fairly clean	80	20
	Dirty	35	45

→ H_0: Cond^n of child and home are independent

```
>x=c(70,80,35,50,20,45)
>m=3
>n=2
>y=matrix(x,nrow=m,ncol=n)
>y
      [,1]   [,2]
[1,]   70     50
[2,]   80     20
[3,]   35     45
>pv=chisq.test(y)
>pv
        Pearson's Chi-squared test
data: y
X-squared = 25.646, df = 2, pvalue = 2.698e-06
```

∵ Pvalue is less than 0.05, we reject the hypothesis at 5% LOS.

∴ ~~the~~ Cond^n of child and home are dependent.

Q2) Test the hypothesis that vaccine and disease are independent or not

Disease	Vaccine Affected	Not affected
Affected	70	46
Not Affected	35	37

```
>x=c(70,35,46,37)
>m=2
>n=2
>y=matrix(x, ncol=n, nrow=m)
>y
        [,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, pvalue=0.1545
```

∵ Pvalue is greater than 0.05, we accept the hypothesis

∴ Vaccine and disease are independent.

H_0: Vaccine and disease are independent or not.

Q3) Perform a anova for the following data:- 048

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))
> name(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, pvalue = 0.00153
> anova = aov(values ~ ind, data = d)
> summary(anova)
            Df   Sum Sq   Mean Sq   Fvalues   Pr(>F)
ind          3   71.06    23.688    11.73     0.00183 **
Residuals    9   18.17    2.019
---
Signif. codes: 0 '***' 0.01 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.
```

∵ Pvalues is less than 0.05, we reject the hypothesis at 5% LOS

∴ The means are not equal for A, B, C, D.

Scanned with CamScanner

Q4) The following data gives the life of the tyres of 4 ~~gram~~brands

Type	Life
A	20, 23, 18, 17, 18, 22, 24
B	19, 15, 17, 16, 20, 17
C	21, 19, 22, 17, 20
D	15, 14, 16, 18, 14, 16

H_0: The average life of A, B, C, D are equal.

```
>x1= c(20,23,18,17,18,22,24)
>x2= c(19,15,17,16,20,17)
>x3= c(21,19,22,17,20)
>x4= c(15,14,16,18,14,16)
>d=stack(list(b1=x1,b2=x2,b3=x3,b4=x4))
>names(d)
[1] "values" "ind"
>oneway.test(values~ind, data=d, var.equal=T)
        One-way analysis of means
 data: values and ind
 F=6.8445, num df=3, denom df=20, pvalues=0.00
```

∵ Pvalues is less than 0.05, we reject the hypothesis at 5% LOS

∴ The average life of A, B, C, D are not equal.

Q4)
```
>x=read.csv("C:/Users/Admin/Desktop/marks.csv")
>am=mean(x$stats)
>am
[1]37
> median(x$stats)
[1]38.5
>n=length(x$stats)
>sd=sqrt((n-1)*var(x$stats)/n)
[1]12.64911
```

```
>mean (x$Maths)
[1] 39.4
>median (x$Maths)
[1] 37
>sqrt((n-1)*var(x$Maths)/n)
[1] 15.2
>cor (x$Stats, x$Maths)
[1] 0.830618
```

AN
20.02

BAO

PRACTICAL - 10

Topic :- Non-parametric test

Q1) Following are the amounts of sulphur oxide emmited by some industries in 20 days.. Apply 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.

→ H_0: Population Median = 21.5.

```
> x = c(17, 15, - - - - 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 is greater than 0.05, we accept the hypothesis at 5% LOS.

- Note:- If the alternative H_1: Me > then pv = pbinom(sn, n, 0.5)

Q2) ~~Atts~~ following is a data of 10 obs. Apply sign test to test the hypothesis that the population median is 625 against the alternative it is more than 625.

Values:- 612, 619, 631, 628, 643, 640, 655, 649, 670, 663

H_1:-
> H_0:- Population median = 625 against H_1: ≠ 625.

050

```
> x = c(612, - - - , 663)
> me = 625
> sp = length(x[x > me])
> sn = length(x[x < me])
> n = sp + sn
> pv = pbinom(sn, n, 0.5)
> pv
[1] 0.0546875
```

∵ Pvalue is greater than 0.05, we accept the hypothesis at 5% LOS.

Q3) Following are the values of a sample. Test the hypothesis that the population median 60 against the ~~hypothesis~~ alternative it is more than 60. at 5% LOS using Wilcoxon signed Rank test.
Value:- 63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 69, 48, 66, 72, 63, 87, 69.

> H_0: Population median = 60
H_1: Population median > 60.

```
> x = c(63, 65, - - - , 69)
> wilcox.test(x, alter = "greater", mu = 60)
Wilcox signed rank test with continuity

data: x
V = 145, pvalue = 0.02298
alternative hypothesis: true location > 60.
```

∵ Pvalue is less than 0.05, we reject the hypothesis at 5% LOS.

Note: If alternative is less, alter = "less". If alternative is not equal to, alter = "two.sided".

Q4) Using WSRT, test the population median is 12 or less than 12.

Data:- 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26

H_0: Population median = 12

H_1: Population median < 12.

```
> x = c(15,17, . . . . ,26)
> wilcox.test(x, alter="less", mu=12)
            Wilcox signed rank test.
data: x
V = 66, pvalue = 0.9986
```

∵ Pvalue is greater than 0.05, we accept the hypothesis at 5% LOS.

Q5) The ~~waves~~ weight of students before and after they stop smoking are given below. Using WSRT, test that there is no significant change.

Before: 65, 75, 75, 62, 72.

After: 72, 74, 72, 66, 73.

```
> x = c(65, 72, 75, 62, 72)
> y = c(72, 74, 72, 66, 73)
> d = x-y
> wilcox.test(d, alter="two-sided", mu=0)
            Wilcoxon signed Rank test
data: d
V = 4.5, pvalue = 0.4982.
```

∵ Pvalue is greater than 0.05, we accept the hypothesis at 5% LOS.

H_0: There is no difference in weights

H_1: There is difference in weights.

AK 27.01.