PS

Practical - I

Basics of R software.

1) R is a software for stastical analysis & 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.

Q.1] Solve the followings

1) 4 + 6 + 8 ÷ 2 - 5

```
> 4+6+8/2-5
[1] 9
```

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

```
> 2^2 + abs(-3) + sqrt(45)
[1] 13.7082
```

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

```
> 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] round off

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 9 16 125

> c(6,2,7,5) / 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,]	1	5
	[2,]	2	6
	[3,]	3	7
	[4,]	4	8

Q.5] 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,]	4	-2	6
[2,]	7	0	7
[3,]	9	-5	3

> y <- matrix (nrow=3, ncol=3, data =c (10,12,15,-5,-4,-6,7,9,5))

> y

	[,1]	[,2]	[,3]
[1,]	10	-5	7
[2,]	12	-4	9
[3,]	15	6	5

> 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	-25	21

6) Marks of statistics of CS Batch A

x = c(68, 20, 35, 24, 46, 56, 55, 45, 27, 22, 47, 58, 54, 40, 30, 32, 36, 29, 35, 39)

```
> x = c (data)
> breaks = seq (20,60,5)
> a = c (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

Practical - 2

Probability distribution

1) check whether the followings are P.mf 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 $\Sigma P(x) = 1$

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

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

$= 1.0$

$\because P(2) = -0.5$, it can be a Probability mass function

$\therefore P(x) \geq 0 \ \forall, x$

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 $P(x) \neq 1$

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

The condition for pmf is
1) $P(x) \geq 0 \quad \forall x$ satisfy.
2) $\Sigma P(x) = 1$

$\Sigma P(x) = P(10) + P(20) + P(30) + P(40) + P(50)$
$= 0.2 + 0.2 + 0.35 + 0.15 + 0.1$
$= 1$

$\therefore$ The given data is p.m.f.

code :

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

2] find the c.d.f for the following p.m.f & sketch the graph

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

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

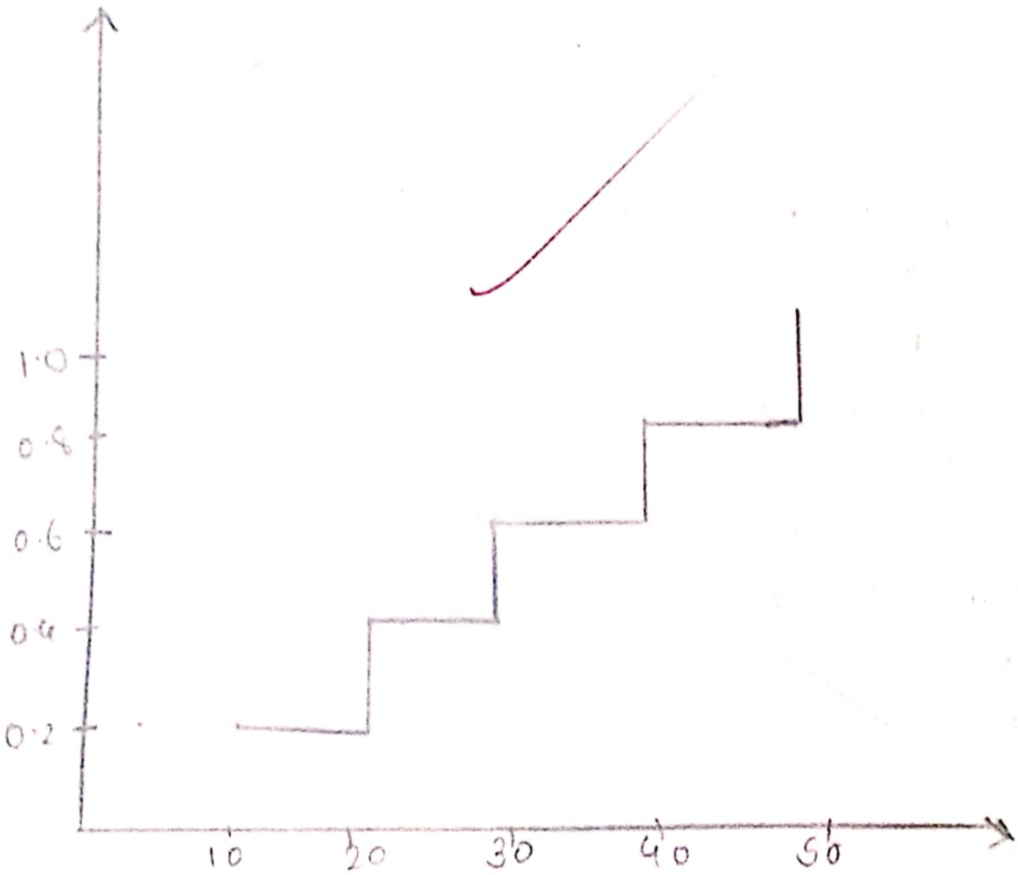
$0.2 \quad 10 \le x < 20$

$0.4 \quad 20 \le x < 30$

$0.75 \quad 30 \le x < 40$

$0.90 \quad 40 \le x < 50$

$1.0 \quad x \ge 50.$




```
> x = c(10,20,30,40,50)
> plot (x, cumsum (prob ,"s")).
```

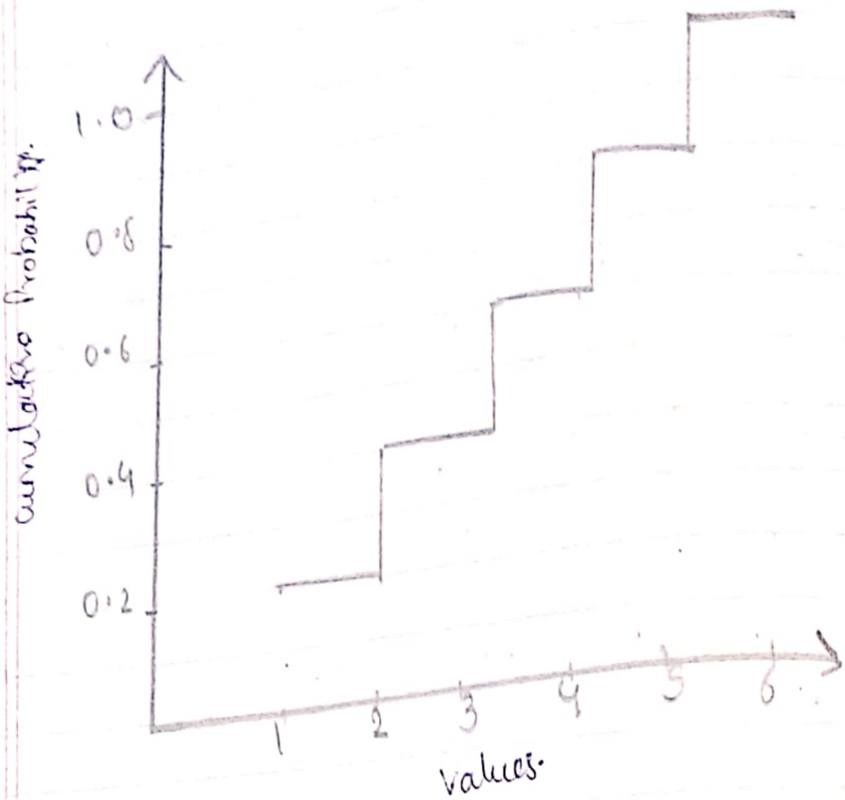
.2) find.

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

$F(x) = 0 \qquad x < 1$
$= 0.15 \qquad 1 \le x < 2$
$= 0.40 \qquad 2 \le x < 3$
$= 0.50 \qquad 3 \le x < 4$
$= 0.70 \qquad 4 \le x < 5$
$= 0.90 \qquad 5 \le x < 6$
$= 1.00 \qquad x \ge 6.$

```
> 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, 0.70, 0.90, 1.00.
> x = c(1,2,3,4,5,6)
> Plot (x, cumsum (Prob, "s", xlab = "value",
  ylab ="Cumulative Probability",
  main = "CDF graph", col= "brown")
```

CDF graph.



3) Check that whether the following is P.d.f or not

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

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

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

$= \int_0^1 f(x)$

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

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

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

$\therefore$ The $\int_0^1 f(x) = 1$ $\therefore$ It is not a pdf.

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

$= \left[3\frac{x^3}{3}\right]_0^1$ $\because x^n = \frac{x^{n+1}}{n+1}$

$= x^3$

$= 1$

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

1 > x = dbinom (10,100,0.1)

> x

[1] 0.1318653.

2 i) dbinom (4,12,0.2)

[1] 0.1328756.

[ii] pbinom ~~(4,12~~,0.2)

[1] 0.~~427~~4445.

(iii) 1 - pbinom (5,12,0.2)

(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] 0.9455928

3) pbinom (7, 12, 0.25)

[1] 0.00298151

4) dbinom (6, 12, 0.25)

[1] 0.04014945

Practical-3

Binomial distribution

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

$P(x \leq x) -$ pbinom (x, n, p)

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

If x is unknown.

$p_1 = P(X \leq x) =$ qbinom (p_1, n, p)

1) Find the Probability 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 1 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 then $n=5$ & $p=0.1$

4) $n=12$, $p=0.25$ find the following probabilities.
i) $P(x=5)$
ii) $P(x \leq 5)$
iii) $P(x>7)$
iv) $P(5 < x < 7)$

5) The Probability of a salesman making a sale to customer 0.15. find the Probability of
i) No sales out of 10 customer
ii) More than 3 sales out of 20 customer

6) A Salesman has 20 % Probability of making a sales to customer out of 30 customer ~~th~~ 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 & c.d.f.

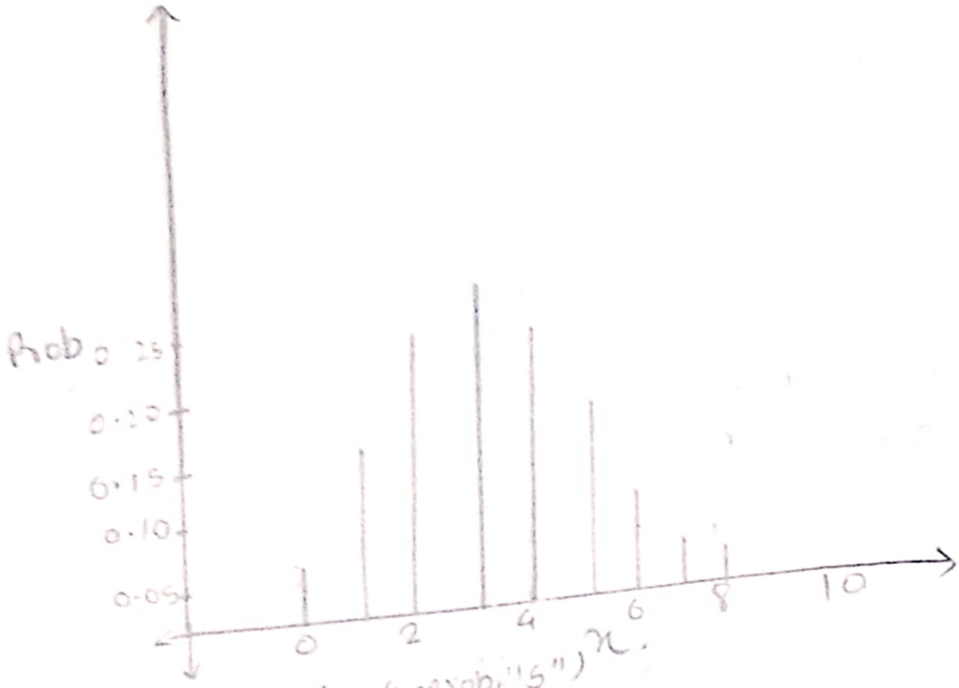
5) > dbinom (0,10, 0.15)
[1] 0.1968744
> 1 - pbinom (3,20,0.15)
[1] 0.3522748

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

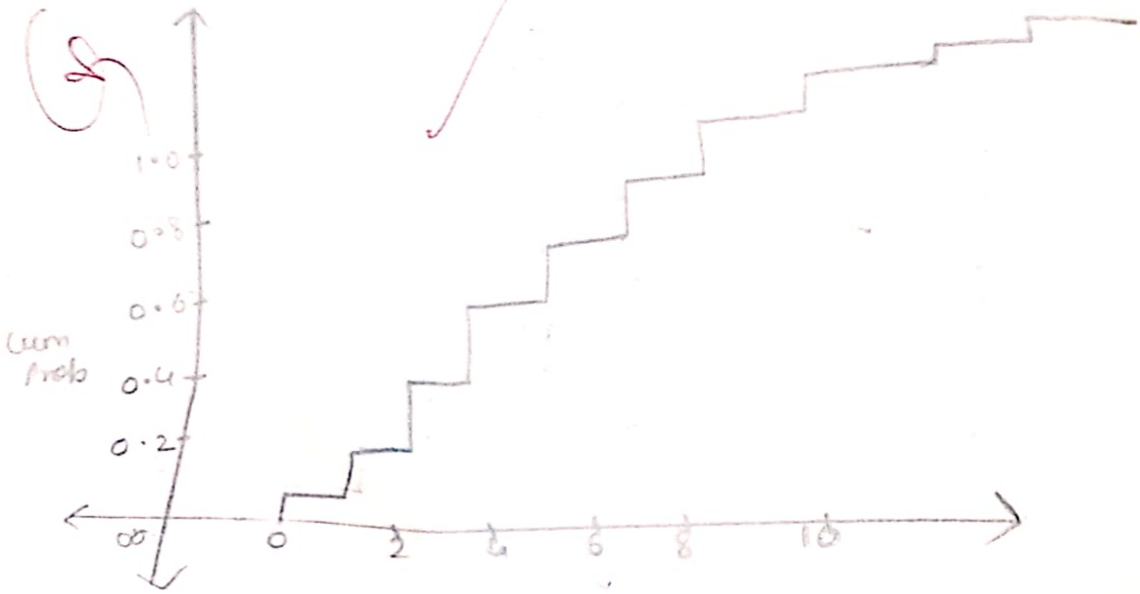
7// > n = 10
p = 0.3
> x = 0 ; n
> Prob = dbinom (x, n, p)
> cumprob = pbinom (x, n, p)
> d = data.frame ("X values" = x, "Probability" = Prob
> Print (d)

	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	0.0367
8	7	0.0090
9	8	0.0014
10	9	0.0001
11	10	0.0000

> Plot (x, ...



> Plot (x, cumprob, "s")



Practical-4.

Normal distribution

\# $P(X = x) =$ dnorm (x, μ, σ)
\# $P(X \le x) =$ pnorm (x, μ, σ)
\# $P(X > x) = 1 -$ pnorm (x, μ, σ)

\# To generate random numbers from a normal distribution (n Random numbers)
the r code is rnorm (n, μ, σ)

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

i) $P(X \le 15)$
ii) $P(10 \le x \le 13)$
iii) $P(X > 14)$
iv) Generate 5 observations (RANDOM NUMBERS)

Ans

i) P1 = pnorm (15, 12, 3)
> P1
[1] 0.8413447
> cat (" P (X <=15) =", P1)
> P (X<=15) = 0.8413447

ii) P2 = pnorm (13, 12, 3) - Pnorm (10, 12, 3)
> P2
→ [1] 0.3780661

> P((10<= X <=13) = 0.378066 1

pii) 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 13.14904 14.98075

19.73518.

2 X follows normal distribution with

$\mu = 10$, $\sigma = 2$ find =?

i) $P(X \le 7)$

ii) $P(5 < X < 12)$

iii) $P(X > 12)$

iv) Generate 10 Random observation

v) find K such that Probability $P(X < K) = 0.$

Ans (i) P1 = Pnorm (7,10,2)

P1

→ [1] 0.0668072

>cat ("P(X<=7) =", P1)

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) = 0.835135

iii) P3 = 1 - Pnorm (12, 10, 2)
P3
→ [1] 0.1586553
> cat ("P (P (X > 12) =", P3)
→ P(P(X > 2) = 0.1586553.

> rnorm (10, 10, 2)
→ [1] 11.986324 12.155504 7.130492
10.514364 5.637988 10.959963
7.986735 9.336612 8.115370
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 & (Print (cat)).

4] X ~N (30, 100) → (σ^2 = 6 = 10) find only P1, P2 .... (no cat command)

Ans

38

```
> P1 = Pnorm (40,30,10)
> P1
→ [1] 0.8413447
> P 2 = 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.
```

B) Ans

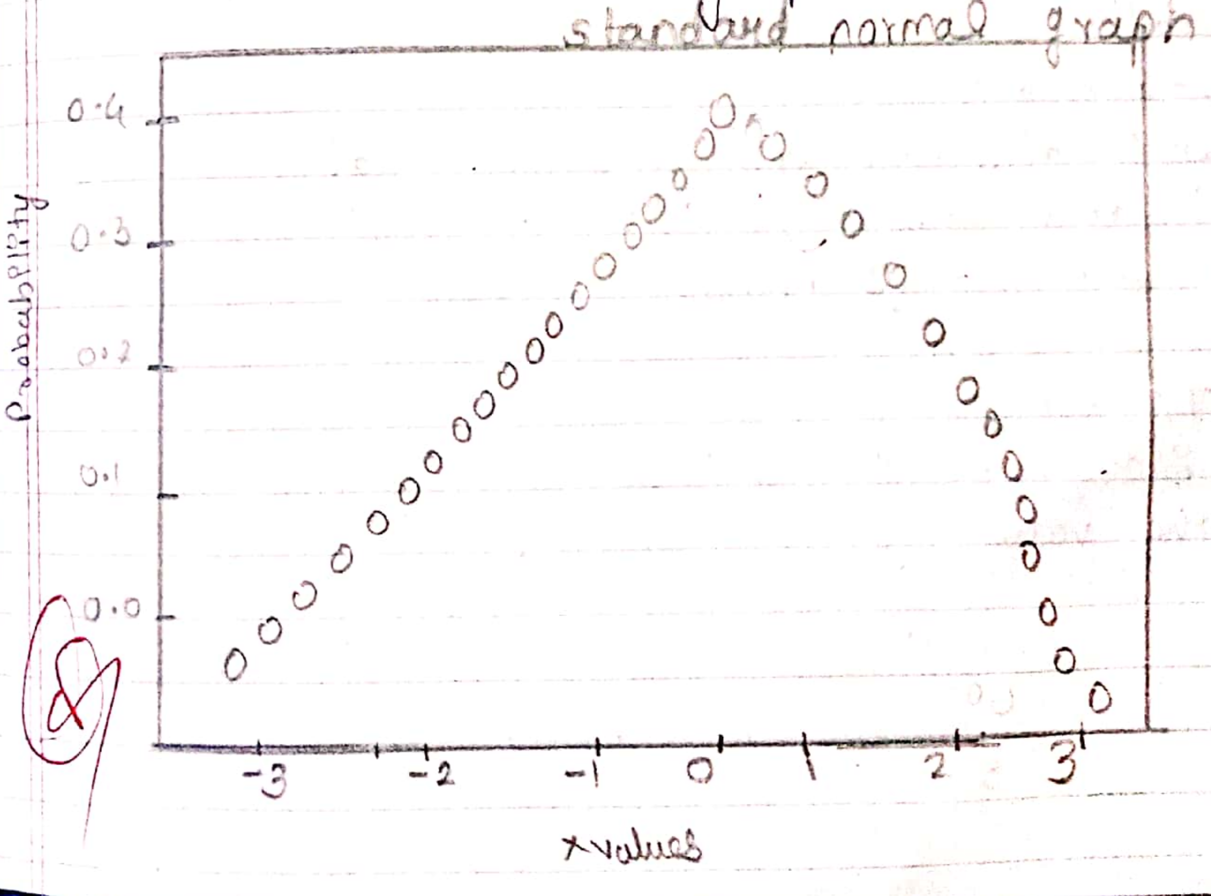
```
> x = rnorm(5,15,4)
> x
→ [1] 16.43602  12.56878  16.91096  17.37383
      15.04210
> am= mean(x)
> am
→ [1] 15.67034
> med= median (x)
> me
→ [1] 16.43602.
> varience = (n-1) * var(x)/n
> varience.
→ [1] 2.9836
```

> sd = sqrt (varience)
> sd
> [1] 1.72731
> cat ("sample mean is = ", am)
> sample mean is = 15.63034
> cat ("sample median is = ", mc)
> sample median 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 = " x values", ylab = " Probability main = " standard normal graph")



Practical-5

Normal & t-test

1] $H_0 : \mu = 15$ $H_1 : \mu \neq 15$

null   alternative

Test the hypothesis

Random sample of size 400 is drawn & it is calculated. the sample mean is 14 & S.D is 3. test the hypothesis at 5 % level of significance

\# 0.05 > accept the value.

\# 0.05 < less than Reject.

```
> mo = 15
> mx = 14
> n = 400
> sd = 3
> zcal = (mx - mo) / (sd / sqrt (n)))
> zcal
[1] -6.666667
> cat (" calculated value of z is =", zcal)
calculated value of z is = -6.666667
> Pvalue = 2 * (1 - Pnorm (abs (zcal)))
> Pvalue
[1] 2.616796e-11
```

∴ The value is less than 0.05 we will reject the value of $H_0 = \mu = 15$.

2] Test the hypothesis H0: U=10 against H1: u ≠ 10. A random sample size of 400 is drawn with sample mean = 10.2 & SD = 2.25 Test the hypothesis at

40

```
>m0=10
>n=400
>mx=10.2
>sd=2.25
>zcal=(mx-m0)/(sd/(sqrt(n)))
>zcal
[1]1.7778
>Pvalue=2*(1-Pnorm(abs(zcal)))
>Pvalue
[1]0.07544036
```

∴ The value Pvalue is greater than 0.05
∴ The value is accepted.

3] Test the hypothesis H0: Proportional is 0.2. A sample is collected & proportional as 0.125 Test the significance (sample size is 400)

```
>P=0.2
>p=0.125
>n=400
>Q=1-P
>zcal=(p-P)(sqrt(P*Q/n))
>cat("calculated values of z is=",
[1] calculated value of z is = -3.75
>Pvalue=2*(1-Pnorm(abs(zcal)))
>Pvalue.
[1]0.0001768346 (Reject)
```

4] last year farmers lost 20% of their crops. A random sample of 60 fields are collected & it is found that 9 field crops are insect polluted. test the hypothesis at 1% level of significance.

```
>P=0.2
>p=9/60
>n=60
>zcal =(p-P))sqrt(P* Q/n))
>zcal.
[1] -0.9682458
>P value =2 * (1-Pnorm (abs (zcal)))
>P value.
[1] 0.3329216
```

∴ The value is 0.1 so value is accepted.

5] Test the hypothesis H0: μ = 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
>variance =(n-1) * var (x)/n
>variance
[1] 0.01952 1
>sd = sqrt (variance)
```

```
>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 values is accepted.

Practical-6

Large sample test

1) Let the Population mean (the amount spent per customer in a restaurant) is 250
A sample of 100 customers selected the sample mean is calculated as 275 & SD 30. test the hypothesis that the population mean is 250 or not on 5 % level of significance.

2) 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.

→ Solution:

```
> m0 = 250
> mx = 275
> sd = 30
> n = 100
> zcal = (mx - m0) / (sd / (sqrt (n)))
> cat ("Calculated value of z is =", zcal )
[1] calculated value of z is = 8.33333
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 0
```

∴ The value is less than 0.05 we will reject the value of H0 = μ = 250.

2) Sol^n

```
> p = 0.5
> q = 1-p
> P = 750/1000
> n = 1000
> zcal (P-p)/(sqrt(p*q/n))
> cat("Calculated value of z is:", zcal)
[1] calculated value of z is: ", -3.952847
> pvalue = 2 * (1 - pnorm(abs(zcal)))
> pvalue
[1] 7.72268e-07
```

∴ the value is less 0.01 we reject!

3] To random sample of size 1000 & 2000 are drawn from two population which same SD 2.5 the sample means are 67.5 & 68 Test the hypothesis $\mu_1 = \mu_2$ at 5% of significance.

4] A study of noise level in 2 hospital in given below test the claim that 2 hospital have same level of noise at 1% level of significance.

HOS A	HOS B
84	34
61.2	59.4
7.9	7.5

5] In a sample of 600 students in clg 400 used ink. In another clg from a sample of 900 student use blue ink. Test the hypothesis that the proportion students using blue ink in two college are equal or at 1% level of significance.

3] Sol

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

4]

```
>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))
>zcal
[1]1.162528
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1]0.245021
```

∴ The value is greater than 0.01 we accept the value.

4) H0 : P1 = P2 against H1 : P1 < P2

```
> n1 = 600
> n2 = 900
> P1 = 400/600
> P2 = 450/900
> P = (n1 * P1 + n2 * P2) (n1 + n2)
> P
[1] 0.566067
> q = 1 - P
> q
[1] 0.4333331
> zcal = (P1 - P2)/sqrt (P*q * (1/n1 + 1/n2))
> zcal
[1] 6.381534
> Pvalue = 2 * (1-Pnorm (abs(zcal)))
> Pvalue
[1] 1.75322e-10
```

∵ Value is less than 0.01 the value is rejected.

5) H0: P1 = P2 as H1: P1 ≠ 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
```

>q
[1] 70.815
>zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
>zcal
[1] 1.802741
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.071428
∴ (Accept ∵ greater than 0.05).

Cd

Practical - 7

Topic: Small sample test

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

Ho: $\mu = 66$

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 estimate

mean of x

67.9

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

2) Two groups of students scored the following marks. Test the hypothesis that there is no significance difference between the 2 groups.

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

H0: there is not difference b/w the 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)
```

Two sample t-test.

data : x & 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 5.0311795
sample estimate:
mean of x mean of y
20.1 17.5

```
> P-value = 0.03798
> if (P value > 0.05) { cat ("accept ho") }
else { cat ("reject ho") }
```

reject H0.

(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 effective or not.

H0: 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 mean is greater than 0
95 Percent confidence interval:
  -6.035547 inf
sample estimates
mean of the differenc
     -3.5
```

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

4) Following are the weights before & the the diet program is the diet program effective:

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

After : 100, 114, 95, 90, 115, 99.

Soln: H0 : There is 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 , Pvalue = 0.9963

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

95 Percent confidence interval:

-inf 29.02915

Sample estimate

mean of the difference.

19.83333 .

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

5) 2 ~~medians~~ is are applied to two groups of patient respectively.

gr1 : 10, 12, 13, 11, 14

gr2 : 8, 9, 12, 14, 15, 10, 9.

Is there any significance difference b/w 2 ~~medians~~. medicines.

H0: There is no significance difference.

> x = c(grp1)

> y = c(grp2)

> t.test (x, y)

data : 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

-0.9648553 4.2981886

sample estimates:

mean of x mean of y.

12.0000 10.3333.

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

Practical - 8

Large & Small Test

1] Ho: u = 55 , H1 : u ≠ 55

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

As pvalue is less than 0.05 we reject ", at 5% level of significance.

2] Ho: P = 0.5 against Ho: P ≠ 0.5

```
> p = 0.5
> q = 1 - p
> n = 200
> zcal = (p - p) / (sqrt (p * q / (n)))
> zcal.
[1] 0
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue.
[1] 1
```

As pvalue is greater than 0.05 we accept Ho at 1% level of significance.

3] H0: P1 = P2 against H1: P1 ≠ P2
```
> n1 = 1000
> n2 = 1500
> p1 = 2/1000
> p2 = 1/1500.
> p = (n1 * p1 + n2 * p2) / (n1 + n2)
> p
[1] 0.0012.
> q = 1-p
[1] 0.9988.
> zcal = (p1-p2) / sqrt (p * q * (1/n1 + 1/n2))
> zcal
[1] 0.9433752
> Pvalue = 2 * (1-pnorm (abs (zcal)))
> Pvalue.
[1] 0.345489.
```

∴ Pvalue is greater than 0.05 we accept H0 & 5% level of significance.

4] H0: μ = 100 against H1: μ ≠ 100
```
> var : 64
> n = 400
> mo = 100
> mx : 900
> sd : sqrt (var)
> sd
[1] 8
> zcal = (mx - mo) / (sd / (sqrt (n)))
> zcal.
```

58.

```
> zcal (m-mo) /(sd/(sqrt(n)))
> zcal
[1] -2.5
> Pvalue 2 * (1-Pnorm (abs(zcal)))
> Pvalue
[1] 0.01241933.
```

Since Pvalue is less than 0.05 we reject H_0 at 5 % level of significance.

5) $H_0: \mu = 66$ against $H_1: \mu \neq 66$.

```
> 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.
```

Since P value is less than 0.05 we reject H_0 at 1 % level of significance.

6] H0: $\sigma_1 = \sigma_2$ against $H_1: \sigma_1 \neq \sigma_2$

```
> x = c(66,67,75,76,82,88,90,92)
> 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 = 0.788803 , num df = 7, denom. df = 10, p-value 0.731
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.199509   3.75188 1
sample estimates:
ratio of variances
        0.7880255
```

p-value is greater than 0.05 we accept H0 at 5% level of significance.

7) H0: $\mu = 1150$ against $H: \mu \neq 1150$.

```
> n = 100
> mx = 1150
> mo = 1200
> sd = 125
> zcal = (mx - mo)/(sd /sqrt(n))
> zcal
[1] -4
> pvalue = 2* (1-pnorm(abs(zcal))
> pvalue
[1] 6.334248e-05
```

8] $H_0: P_1 = P_2$ against $H_1: P_1 \neq P_2$.

```
> n1 = 200
> n2 = 300
> p1 = 44/200
> p2 = 56/300
> p = (n1*p1 + n2*p2)/(n1+n2)
> p
[1] 0.2
> q = 1-p
[1] 0.8
> zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
> zcal
[1] 0.9128709
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.3613104
```

∴ Pvalue is greater than 0.05 we accept H_0 at 1% level of significance.

Practical 9

Topic:- Non-Parametric testing of Hypothesis using R-environment

1) The following data represent earnings (in dollars) for a random sample of five common stocks listed on the New exchange. Test whether median earnings is 4 dollars Data 1.68, 3.35, 2.50, 6.23, 3.24

```
> x <- c(1.68, 3.35, 2.50, 6.23, 3.24);
> n <- length (x);
> n
> [1] 5
> x > 4;
[1] FALSE FALSE FALSE TRUE FALSE
> s <- sum (x > 4); s;
[1] 1
> binom.test (s, n, p = 0.5, alternative = "greater");
        Exact binomial test

data: s and n
number of success = 1, number of trials = 5, p-value = 0.97
alternative hypothesis: true probability of success
 is greater than 0.5;
 95 percent confidence interval:
  0.01020622   1.0000000
  Sample estimates:
  Probability of success 0.2
```

PA

2] The scores of 8 students in reading before and after lesson are as follows. Test whether there is effect of reading.

Student No:	1	2	3	4	5	6	7	8
Score before:	10	15	16	12	09	07	11	12
Score after:	13	16	15	13	09	10	13	10

CODE:

```
> b <- c(10,15,16,12,09,07,11,12);
> a <- c(13,16,15,13,08,10,13,10);
> D <- b-a;
> wilcox-test (D, alternative ="greater");
```

Wilcoxon signed rank test with continuity correction data: D

V = 10.5 , P-value = 0.8722

alternative hypothesis : True location is greater than 0

warning message:

In wilcox.test. (D, alternative = "greater",

Cannot compute exact P-value with ties.

∴ P-value is greater than 0.05 we accept it.

50

3] The diameter of a ball bearing was measured by 6 inspector, each using two different kinds of calipers. The results were, test whether, average ball bearing for.

Inspector :	1	2	3	4	5	6
Caliper 1:	0.265	0.268	0.266	0.267	0.269	0.264.
Caliper 2:	0.263	0.262	0.270	0.261	0.21	0.260.

Caliper 1 & Caliper 2 are same

CODE:

```
> x <- c(0.265, 0.268, 0.266, 0.267, 0.269, 0.264);
> y <- c(0. 263, 0.262, 0.270, 0.261, 0.271, 0.260);
> wilcox.test (x, y, alternative ="greater")
        wilcoxon rank sum test
data: x & y
w = 24, P = 0.197
alternative hypothesis: true location shift is greater than 0.
```

p-value is greater than 0.05 we accept it.

02

Q] An office has three effective type writers A, B & C. firm has kept records in a study of machine usage rate of seven weeks. Machine A was out of machine usage rate of two weeks. It is of interest to find out repairs for two weeks. It is of interest to find out which machine has better usage rate. Analize the following data on usage rates & determine if there is a significant difference in average rate.

A	B	C
12.3	15.7	32.4
15.4	10.8	41.2
10.3	45.0	35.1
08.0	12.3	25.0
14.6	08.2	0.2
–	20.1	18.4
–	26.3	32.5

Code:

```
> x <- c(12.3, 15.4, 10.3, 8.0, 14.6);
> n1 <- length(x);
> n1
[1] 5
> y <- c(15.7, 10.8, 45.0, 12.3, 8.2, 20.1, 26.3);
> n2 <- length(y);
> n2
[1] 7
> z <- c(32.4, 41.2, 35.1, 25.0, 8.2, 18.4, 32.5);
> n3 <- length(z);
> n3
[1] 7
```

> x <-c(x, y, z);
> g <-c (rep (1, n1), rep (2, n2) rep (3, n3));
> Krushkal.test (x, g)

Krushkal - wallis Rank sum test

data : x & g.

Kruskal - wallis chi-squared = 5.217, df = 2, p-value = 0.07365

∴ p-value is greater than 0.05 we accept it.

Practical-10.

Aim:- Chi Square test & ANOVA (Analysis of Variance).

I] Use the following data to test whether the condition of home & condition of home & cor child & independent or not.

Cond child	Cond home	
clean	clean	dirty
fairly clean	70	50
dirty	80	20
	35	45.

H0: 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 (y)

>pv.

Pearson's chi-squared test

data : y

x-squared = 25.646

df = 2

p-value = 2.698e-06

they are dependent

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

2] Test the hypotheses that vacation and disease are independent or not.

disease \ vaccine	affected	not affected
affect	70	46
non-affected	35	37

H0 : disease & vaccine are independent.

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

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

> PV = chisq.test (y)
> PV

Pearson chi-squared test with continuity correction

data : y

X - square = 2.0275
df = 1
P-value = 0.1545

∵ P-value is more than 0.05 we accept the hypothesis at 5 % level of significance.

they are INDEPENDENT.

3) Perform a ANOVA for the following data.

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

HO: 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)
```

one way - analysis of means.

data : ~~values~~ & ind
f = 11.735
df = 3
denom df = 9.
p value = 0.00183

p value is less than 0.05 we reject the hypothesis.

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

	Df	Sum	meansq	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 ' ' 1.