# PERFORMANCE

| Term | Remarks | Staff Member's Signature |
|---|---|---|
| I | Cmple | A.r |
| II | Cmplet | A.rat |

Exam Seat No.________________

# THAKUR COLLEGE OF SCIENCE & COMMERCE

NAAC Accredited with Grade "A" (3rd Cycle)

THAKUR TRUSTS®

ISO 9001 : 2015 Certified

## Degree College

# Computer Journal

# CERTIFICATE

**SEMESTER** II **UID No.** ________________

**Class** FYBSCS **Roll No.** 1738 **Year** 2019-2020

This is to certify that the work entered in this journal is the work of Mst. / ~~Ms.~~ Anakal Suraj Shivkumar

________________________________

who has worked for the year __________ in the Computer Laboratory.

A.Naik

Teacher In-Charge

Head of Department

Date : 27.02.20

**Examiner**

# INDEX

9/11/19

# PRACTICAL - 01

* Basic of R Software :-

- R is a software for statistical analysis and data computing
- It is an effective data handling software and outcome storage is Possiable.
- It is capable of graphical display
- R is a Free software

Q.1 Solve the Following

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

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

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

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

v) Round off
$46 \div 7 + 9 \times 8$

Code :-

1) 4+6+8/2
[1] 14

2) 2^2 + abs.(-3) + sqrt(45)
[1] ~~+4~~ 13.7082

3) 5^3 + 7 * 5 * 8 + 46/5
[1] 414.2

4) sqrt (4^2 + 5 * 3 + 7/6)
[1] 5.671567

5) 46 / 7 + 9 * 8
[1] 78.57143.

Q.2] Solve the Following:

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

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

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

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

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



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

Q.3] Solve the Following

$x=20$, $y=30$, $z=2$

Find, i) $x^2 + y^3 + z$

ii) $\sqrt{x^2+y}$

iii) $x^2 + y^2$

code :-

i) >a = x^2 + y^3 + z
>a
[1] 27402

ii) > b = sqrt( x^2 + y )
>b
[1] 20.73644

iii) >c = x^2 + y^2
>c
[1] 1300

Q4] Convert the data in Matrix Form:

| | |
|---|---|
| 1 | 5 |
| 2 | 6 |
| 3 | 7 |
| 4 | 8 |

Code

```
x= matrix (n row = 4 ,ncol = 2, data = c(1,2,3,4,5,6,7
x
```

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

Q.5] Find x+y and 2x+3y where

$$x=\begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 3 \end{bmatrix} \quad y=\begin{bmatrix} 10 & -5 & 7 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$$

Code :-

```
>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 |



```
> 2*x + 3*y
```

| | [,1] | [,2] | [,3] |
|---|---|---|---|
| [1,] | 38 | -19 | 33 |
| [2,] | 50 | -12 | 41 |
| [3,] | 63 | -28 | 21 |

```
> x + y
```

| | [,1] | [,2] | [,3] |
|---|---|---|---|
| [1,] | 14 | -7 | 13 |
| [2,] | 19 | -4 | 16 |
| [3,] | 24 | -11 | 8 |

Q6] Marks of Statistics of Computer Science student
59, 20, 35, 24, 46, 56, 55, 45, 27, 22, 27, 58, 54, 40, 50, 32, 36, 29, 35, 39

Code :-

```
>x = c(59,20, 35, 24, 46, 56, 55, 45, 27, 22, 27, 58, 54,
     40, 50, 32, 36, 29, 35, 39)

> 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, 50) | 2 |
| 8 | [55, 60) | 4 |

# Topic:- Probablity Distribution

## PRACTICAL - 02

Q Check whether following are p.m.f or Not

1)

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

∴ $P(2) = -0.5$, it can't be a Probablity mass Function, since in p.m.f, $P(x) \geq 0$ for x.

2)

| x | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| P(x) | 0.2 | 0.2 | 0.3 | 0.2 | 0.2 |

It cannot be a p.m.f as in p.m.f
$\Sigma P(x) = 1$

Code:-
```
Prob=c(0.2,0.2,0.3,0.2,0.2)
> sum(Prob)
[1] 1.1
```

3

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

It is a Probablity Mass Function as
In p.m.f $\sum P(x) = 1$

code :- Prob=c (0.2, 0.2, 0.35, 0.15, 0.1)
sum (Prob)
[1] 1

Q.2) a) i) Find c.d.f for following p.m.f and states the graph.

i)

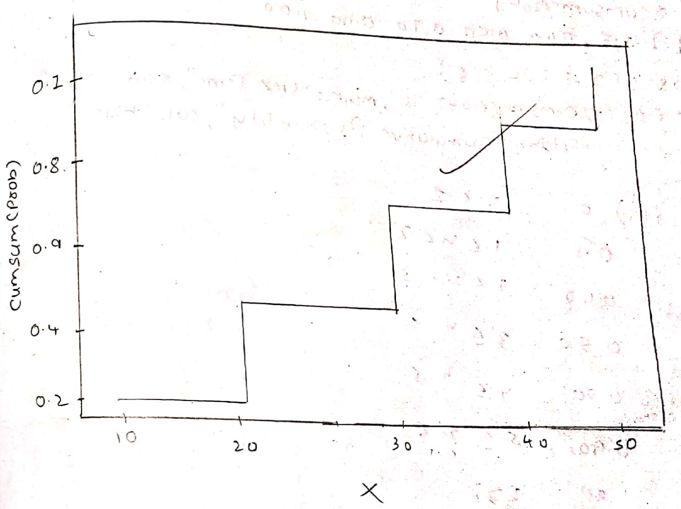
| x | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| P(x) | 0.2 | 0.2 | 0.35 | 0.15 | 0.1 |

Code :-

```
prob=c (0.2, 0.2, 0.35, 0.15, 0.1)
sum(Prob)
[1] 1
cumsum (Prob)
[1] 0.20  0.40  0.75  0.90  1.00
x=c (10, 20, 30, 40, 50)
plot (x, cumsum(Prob), "s")
```



| | |
|---|---|
| f(x) = 0 | $x < 10$ |
| 0.2 | $10 \le x < 20$ |
| 0.4 | $20 \le x < 30$ |
| 0.75 | $30 \le x < 40$ |
| 0.95 | $40 \le x < 50$ |
| 1.0 | $1.0 \; x \ge 50$ |



ii)

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

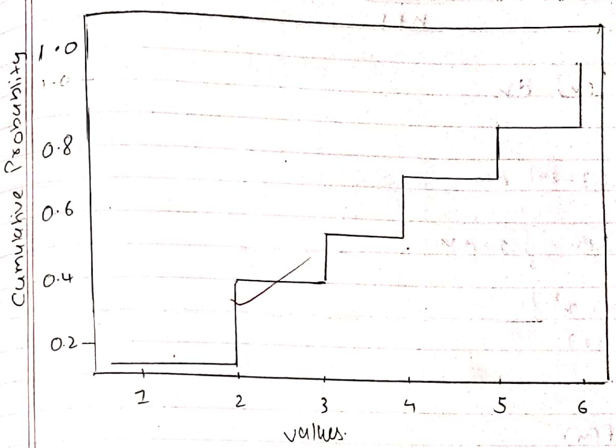
Code:

```
> 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", main="Step Func", xlab="values",
        ylab="Cumulative Probability", col="blue")
```

| | |
|---|---|
| $f(x) = 0$ | $x < 1$ |
| $0.15$ | $1 \le x < 2$ |
| $0.40$ | $2 \le x < 3$ |
| $0.50$ | $3 \le x < 4$ |
| $0.70$ | $4 \le x < 5$ |
| $0.90$ | $5 \le x < 6$ |
| $1.00$ | $x \ge 6$ |



Graph:-




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

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

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

Formula:-

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

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

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

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

$= [3x - x^2]$

$= [3 - 1]$

$= 2$

ii) $\int_0^1 f(x)$

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

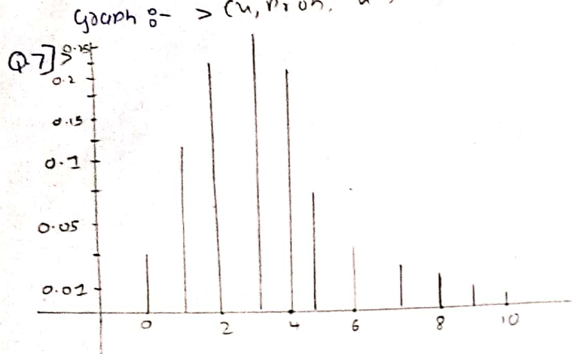
$\left[\frac{3x^3}{3}\right]_0^1$

$= 1$

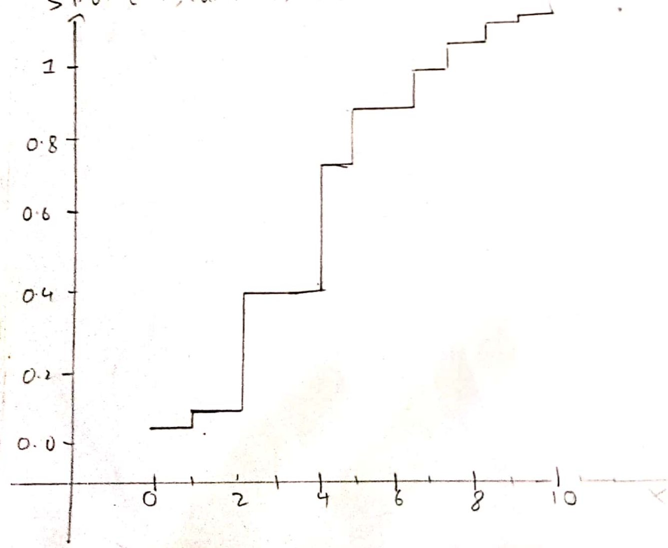
4/2/19

Graph :- > (x, prob, "h")

Q.7] >




> plot ( x, cumprob, "s")






TOPIC :- Binomial Distribution

## PRACTICAL - 03

i) P(X = x) = dbinom ( x, n, p )

→ ~~> dbinom (10, 100, 0.1)~~

~~[1] 0.1318653~~

ii) P(X ≤ x) = Pbinom (x, n, p)

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

iv) If x is unknown

P₁ = P(X ≤ x) . . . . (given)

qbinom ( P₁, n, p )

Q.1) Find the Probability of exactly 10 Success in 100 trials with P = 0.1

→ > dbinom (10, 100, 0.1)

[1] 0.1318653

Q.2) Suppose there are 12 MCQ. Each question has 5 option out of which 1 is correct. Find the Probability of :-

i) Exactly 4 correct answers

ii) Almost 4 correct answers

iii) More than 5 correct answer

→ 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

Q.3] Find the complete distribution when n=5, P=0.1

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

```
[1] 0.59049
    0.32805
    0.07290
    0.00810
    0.00045
    0.00001
```

Q.4) n=12, P = 0.25

Find :- i) p(X=5), ii) p(X≤5), iii) p(X≥7), iv) p(5<X<7)

-> i) >dbinom (5, 12, 0.5)

```
[1] 0.1032414
```

ii) > pbinom (5, 12, 0.5)

```
[1] 0.9455978
```

iii) > 1- pbinom (5, 12, 0.5)

```
[1] 0.0027815 1
```

iv) >dbinom (6, 12, 0.25)

```
[1] 0.04014945.
```

Q5) Probability of Salesman making a 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

-> i) > dbinom (0, 10, 0.15)

```
[1] 0.1968744
```

ii) 1- pbinom (3, 20, 0.15)

```
[1] 0.3522748.
```



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 88% probability.

-> > qbinom (0.88, 30, 0.2)

```
[1] 9
```

Q.7) X followed binomial distribution with n=10 P=0.3

plot the graph of p.m.f. and cdf

-> > 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 | 0.0282475249 |
| 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 | 0.0090016920 |
| 9 | 8 | 0.0014467005 |
| 10 | 9 | 0.0001377810 |
| 11 | 10 | 0.0000059049 |

Source code :-

```
Q.1) > μ=12 , σ = 3
i) > P1 = pnorm(15, 12, 3)
   [1] 0.8413447
ii) > P2 = pnorm(13, 12, 3) - pnorm(10, 12, 3)
   [1] 0.3780661
   cat("P(10<=x<=13) = ", P2)
   P(10<=x<=13) = 0.3780661
iii) > P3 = 1- pnorm(14, 12, 3)
   [1] 0.2524925
   cat("P(x) > 14):", P3)
   P(x >14)= 0.2524925
iv) > rnorm(5, 12, 3)
   [1] 16.380316  13.539877, 13.050514  4.181289

Q.2
i) > P1 = pnorm(7, 10, 2)
   [1] 0.0668072
ii) P2 = pnorm(12, 10, 2) - pnorm(5, 10, 2)
   [1] 0.8351351
iii P3 = 1 - pnorm(12; 10, 2)
   [1] 0.1586553
iv) rnorm(10, 10, 2)
   [1]. 9.856516 10.298521 7.140323 11.081558 11.662138
        11.662138 11.186453 9.982776 9.926634 9.595926
v) qnorm(0.4, 10, 2)
   [1] 9.493306
```

AM



TOPIC :- Normal Distribution

## Practical - 04

$P(X=x)$ = 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 the normal distribution (n random numbers) the R code is

rnorm(n, μ, σ)

Q1 A random variable x follows a normal distribution with mean(μ) = 12 and Standard Deviation (σ) = 3.
Find :- i) $P(x \le 15)$ ii) $P(10 \le x \le 13)$ iii) $P(x > 14)$
iv) Generate 5 Observation (random numbers)

Q.2 x follows normal distribution with μ = 10, σ = 2
Find :- i) $P(x \le 7)$ ii) $P(5 \le X < 12)$ iii) $P(x > 12)$
iv) Generate 10 Observation (random numbers)
v) Find k such that $P(x < k) = 0.4$

Q.3 Generate five rand. numbers from a normal distribution μ = 15, σ = 4
Find :- Sample mean, median, S.D and Print it

Q.4) X Follows normal $X \sim N(\overset{\mu}{30}, \overset{\sigma^2}{100})$
Find :-

i) P(x≤40)
ii) P(x>35)
iii) P(25<x<35)
iv) Find K such that P(x<K) = 0.6

Q.3) i) pnorm(40, 30, 10)
[1] 0.8413447

ii) 1 - pnorm(35, 30, 10)
[1] 0.3085375

iii) pnorm(35, 30, 10) - pnorm(25, 30, 10)
[1] 0.3829249

iv) qnorm(0.6, 30, 10)
[1] 32.53347

Q.4) i) x = rnorm(5, 12, 4)
[1] 5.734772 15.554360 8.101930 13.176196 13.919685

ii) am = mean(x) cat("Sample mean is:", am)
[1] 11.29629 Sample mean is 11.29619

n = 5

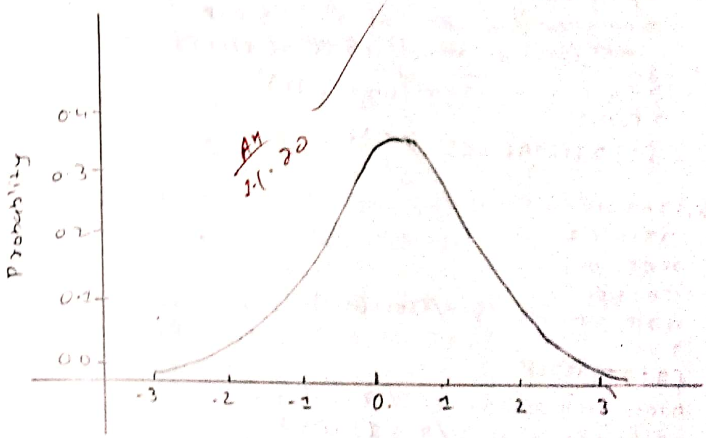
iv) Variance = (n-1)* var(x)/n
variance
[1] 13.93186

v) sd = sqrt(variance) cat("Sample sd is:" sd)
sd Sample sd is: 3.73254)
[1] 3.732541
me = median(x) cat("sample median is:", me)
[1] 13.1702 Sample median is 13.1702



Q5] Plot the standard normal graph
x = seq(-3, 3, by = 0.1)
y = dnorm(x)
plot(x, y, xlab = "x values", ylab = "Probabilities",
main = "~~Standard~~ Standard normal graph")

Standard normal graph



Source code :-

Q.1]
```
>mo=15
>mx=14
>sd=3
>n=400
>zcal=(mx-mo)/(sd/(sqrt(n)))
>zcal
[1]-6.66667
>cat("calculated value of z is:",zcal)
calculated value of z is:-6.66667
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.616796e-11
```

Q.2)
```
>mo=10
>mx=10.2
>sd=2.25
>n=400
>zcal=(mx-mo)/(sd/(sqrt(n)))
>zcal
[1]1.777778
>cat("calculated value of z is:",zcal)
calculated value of z is 1.77778
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1]0.07544036
```



Practical - 05

Q1 Test of Hypothesis :

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

A random sample is drawn of size 400 and it is calculated the sample mean is 14 and standard deviation is 3. Test the hypothesis at 5% level of significance

Since pvalue is less than 0.05 we reject $H_0: \mu = 15$

Q2] Test the hypothesis $H_0: \mu = 0$, $H_0: \mu \neq 10$. A random sample of size 400 is drawn with sample mean 10.2 and sd 2.25. Test the hypothesis at 5% level of significance.

Since pvalue is greater than 0.05 we accept $H_0: \mu = 15$

Q3] Test the hypothesis H0 proportion of smokers in a college is 0.2. A sample is collected and is calculated the sample proportion calculated as 0.125. Test the hypothesis at 5% level of significance (sample size is 400)

Since P value is less than 0.05 we reject H0

Q.4) Last Year farmers ~~tots~~ lost 20% of their crop. A random sample of 60 fill fields are collected and it is found that 9 fields crops are insect polluted. Test the hypothesis of 1% level of significance.

Since P value is greater than 0.05 we accept H0



Q.3)
```
>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
>cat("calculated value of z is:",zcal)
calculated value of z is= -3.75
```

Q.4)
```
>P= 0.2
>p= 9/60
>n= 60
>Q= 1-P
>zcal=(p-P)/(sqrt(P*Q/n))
>zcal
[1] -0.9682458
>> cat("calculated value of z is:",zcal)
calculated value of z is: -0.9682458
>>pvalue= 2*(1-pnorm(abs(zcal)))
>>pvalue
[1] 0.3329216
```

(Q5) > x = c(12.25, 12.97, 12.15, 12.08, 12.31, 11.28, 12.94, 12.89, 12.16, 12.04)

```
> n=length(x)
> n
[1] 10
> mx = mean(x)
> variance = (n-1) * var(x)/n
> sd= sqrt(variance)
> t=(mx-mo)/(sd/sqrt(n))
> t
[1]-8.894909
> pvalue = 2*(1-pnorm(abs(t)))
> pvalue
[1] ~~0.1339216~~ 0
```



Q5 Test the hypothesis H0, $\mu = 12.5$ from the following sample at 5% level of significance.

Since p value is less than 0.05 we reject H0

AN 16/1/20

Topic:- Large Sample Test

23/01/20

Practical - 06

Q.1] Let the population mean (the amount spend per customer in a restaurant is 250. A sample of 100 customers selected. The sample mean is calculated 275 and S.D 30. Test the hypothesis that population mean is 250 or not at 5% level of significance

Q.2] In a random sample of ~~thousand~~ 1000 Student it is found that 750 use blue pen. Test the hypothesis that the population proportion is 0.8 at 1% level of significance.

Code:- HO: M = 275 against H1: M ≠ 275

```
Q.1) > mo = 250
> mx = 275
> sd = 30
> n = 100
> zcal = (mx-mo)/(sd/(sqrt(n)))
> zcal
[1] 8.3333
> cat("calculated value of z is =", zcal)
calculated value of z is = 8.3333
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0
```

Pvalue = 0 < 0.05, we reject HO at 5% level of significance.



HO: P = 0.8 against $H_1$: P ≠ 0.8

```
Q.2.) > P = 0.8
> p = 750
> n = 1000
> zcal = (p-P) / (sqrt(P*Q/n))
> zcal
> [1] ~~5922946~~ -3.952847
> cat("calculated value of z is =", zcal)
calculated value of z is = ~~5922946~~ -3.952847
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 7.72268e-05
```

∴ Pvalue is greater than 0.05 we accept HO

Q.3] Two random sample of size 1000 and 2000 are drawn from two population with the same S.D 2.5. The sample means are 67.5 and 68. respectively. Test the hypothesis HO: $M_1 = M_2$ at 5% level of significance

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

| | Hospital A | Hospital B |
|---|---|---|
| Size | 84 | 34 |
| Mean | 61.2 | 59.4 |
| S.D | 7.9 | 7.5 |

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

Q3] HO : $\mu_1 = \mu_2$ as. H1 : $\mu_1 \neq \mu_2$

```
>n1=1000
>n2=2000
>mx1=~~78.5~~ 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
```

∴ Pvalue = 2.417564e-07 > 0.05, we accept HO : $\mu_1 = \mu_2$ at 5% level of significance.



HO: $P_1 = P_1$ against H1: $\mu_1 \neq \mu_2$

Q4)

```
>n1=84
>n2=34
>mx1=67.2
>mx2=59.4
>sd1=7.9
>sd2=7.5
>zcal=(mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2/n2))
>zcal
[1] 1.62528
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.2450211
```

∴ Pvalue = 0.2450211 > 0.05, we ~~accept~~ reject HO : $\mu_1 = \mu_2$ at 1% level of significance.

Q5) > HO : $P_1 = P_2$ against H1 : $\mu_1 \neq \mu_2$

```
>n1=600
>n2=900
>p1=400/600
>p2=450/600
>p=(n1*p1+n2*p2)/(n1+n2)
>q=1-p
>p=0.566667
~~zcal=2*(1-pnorm(abs(zcal)))~~
Pvalue
>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
pvalue = 1.75322e-10 < 0.05, we reject :
$p_1 = p_2$ at 5% level of significance.

Q.6) $H_0: P_1 = P_2$ against $H_1: M_1 \neq M_2$
> n1= 200
>n2 = 200
>p1 = 44/200
>p2 = 30/200
>p= (n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.185
>q = 1-p
~~>zcal~~
>zcal (p1-p2) / sqrt( p*q*(1/n1+1/n2))
>zcal
[1] 1.802741
>pvalue= 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.07142888
pvalue = 0.0714288, We accept $p_1 = p_2$ at 5% level of significance.

RM 27-01-20



TOPIC:- Small Sample Test
PRACTICAL: 07

Q1] The marks of 10 Students are given as 63, 63, 66, 67, 68, 69, 70, 70, 71, 72. Test the hypothesis that the sample comes from a population of average marks 66.

Solution:-
>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 mean is not equal to 0 95 percent confidence interval: 65.65171 70.14829
sample estimates:
mean of x
67.9

Since p-value is less than 0.05 we reject hypothesis at 5% level of significance.

>los= 0.05
>pvalue= 1.558e-13
>if(pvalue>0.05) {cat ("accept H0")}
else {cat("reject H0")}
reject H0>

Q2] Two groups of students score the following marks. Test the hypothesis that there is no significance difference between the two groups.

group1 = 18, 22, 21, 17, 20, 17, 23, 20, 22, 27

group2 = 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Solution:

H0: There is no difference between two groups.

```
>x=c(18, 22, 21, 17, 20, 17, 23, 20, 22, 27)
>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 = 0.0379
>if(Pvalue>0.05) {cat("accept Ho")}
   else {cat(" reject Ho")}
```

reject Ho)



Since P-value is less than 0.05 we reject hypothesis at 5% level of significance.

Q3] The sales data of six shops before and 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

Solution:- Ho: There is no significant difference of sales before and after the campaign

```
>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 percentage confidence interval

-6.035547 Inf

Sample estimates:

mean of the difference

-3.5

>Pvalue = 0.9806

>if (pvalue >0.05) {cat ("accept H0")}
else {cat ("reject H0")}
accept H0)
Since p-value is greater than 0.05 we accept hypothesis at 5% level of Significance

Q.4] Two median are applied to the two group of patient respectively
Group 1: 10, 12, 13, 11, 14
Group 2: 8, 9, 12, 14, 15, 10, 9
Is there any ~~a~~ significance difrence between two ~~vo~~ medicines
Solution:-
>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, p-value = 0.4406
alternative hypothesis : true difference in means is not equal to 0.
95 percent confidence interval:
-1.781171 3.71171
Sample estimates:
Mean of x and y:
12 11
> pvalue= 0.4406



>if (pvalue >0.05) {cat ("accept H0")}
else {cat ("reject H0")}
accept H0>

Q.5] The following are the weight before and after the diet program effective.
Before: 120, 125, 115, 130, 123, 119.
After: 100, 114, 95, 90, 115, 99
Solution:- H0: There is no significant 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")
Paired t-test
data: x and y
t= 4.3458, df =5 , p-value = 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.
>pvalue = 0.9963

```
> if (pvalue > 0.05) {cat("accept H0")}
  else {cat("reject H0")}
  accept H0>
```

Since the p-value is greater than 0.05, we accept the hypothesis at 5% of significance.

PRACTICAL : 08

## 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 |

Answer :

Q1] >H0 : $\mu$ = 55
>H1 : $\mu \neq$ 55
>n = 100
>mu = 52
>sd = 7
>zcal = (mx-mo)/(sd/(sqrt(n)))
>zcal
[1] -4.285174.
>pvalue= 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.82153e-05
>cat("calculated z value is :" zcal) calculated z value is :-4.285274
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.82153e-05

Since pvalue is less than 0.05 we reject the hypothesis 5% level of significance.

Q2] H0 : $\mu$
>P=0.5
>Q = 1-P
>p=350/700
>n=700
>zcal=(p-P)/sqrt(P*Q/n))
>zcal
[1] 0

>cat("calculated z value is =", zcal) calculated z value is = 0.

>P value = 2*(1-pnorm(abs(zcal)))

> Pvalue.

[1] 1

Since pvalue is greater than 0.05 we accept hypothesis is at 1% level of significance

Q.3] H0: P1 = P2 as H1: P1 ≠ P2

>n1 = 1000

>n2 = 1500

>P1 = 0.02

>P2 = 0.01

>P = (n1*P1 + n2*P2) /(n1 + n2)

>P

[1] 0.014

>q = 1-P

>q

>zcal = (P1-P2)/ sqrt( P*q * (1/n1 + 1/n2))

>zcal

[1] 2.084842



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

>Pvalue

[1] 0.03708364.

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

Q.4] H0 : M = 100

>mx = 99

>mo = 100

>sd = 8

>n = 400

>zcal = (mx-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 the hypothesis at 5% of Significance

Q.5] >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

95 percent confidence interval:

64.66479    71.62092
Sample estimate :
mean of x
68.14286

ho: ~~σ1=σ2~~ σ1=σ2

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 and y
F = 0.70686, num df = 8, denom df = 10, p-value = 0.6359
alternative hypothesis: true ratio of variance is not equal to 1
95 percent confidence interval:
0.183362    3.0360393
sample estimates:
ratio of variances
0.7068567

>pvalue = 0.6359
>if(pvalue>0.05){cat('accept ho')} else{cat("reject ho")}
accept ho

Since p-value is greater than 0.05 we accept hypothesis at 5% level of significance



Q7) Ho: μ=1200
>mx=1150
>mo=1200
>n=100
>sd=125
>zcal=(mx-mo)/(sd/(sqrt(n)))
>zcal
[1] -4
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 6.334248e-05
>cat("calculated z value is :-4)

Since pvalue is less than 0.05 we reject hypothesis.

Q8) ho: p1≠p2
>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] 0.9128709

>cat ("calculated z value is :", zcal)

calculted z value is : 0.9128703

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

> pvalue

[1] 0.3613104

Since pvalue is greater than 0.05

∴ we accept the hypothesis.



PRACTICAL : 09

TOPIC :- Chi squar tests and ANOVA

Q1] Use the Following data to test whether the Condition of the home and contain of child are independent or not

| | Clean | Dirty |
|---|---|---|
| Clean | 70 | 50 |
| Fairly | 80 | 20 |
| Dirty | 35 | 40 |

→ H0: Condition of the Home and Child 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
```

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

Q.2] Test the hypothesis that the vaccination and the disease are independent or not

| | Vaccine Af. | Not Af |
|---|---|---|
| AFF | 70 | 46 |
| N.A | 35 | 37 |

→ Ho: Disease and vaccination are independent

```
>x=c(70,35,46,37)
>m=2
>n=2
>y=matrix(x, nrow=m, ncol=n)
>y
      [,1]  [,2]
[1,]   70    46
[2,]   35    37
>pv=chisq.test(y)
>pv
    Pearson's Chi-Squared test with Yates' continuity
correction
data: y
X-squared = 2.0275, df= 1, p-value= 0.1545
```

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



Q.3] Perform a ANOVA for the following data.

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

→ Ho: Means are equal for A, B, C, D

```
>h1=c(50,52)
>h2=c(53,55,53)
>h3=c(60,50,57,56)
>h4=c(52,54,54,55)
>d=stack(list(b1=h1, b2=h2, b3=h3, b4=h4))
>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=0.00183
>anova=aov(values~ind, data=d)
           Df  Sum Sq  mean Sq  F value  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
```

Since p value is less than 0.05 we reject the hypothesis at 5% level of significance

Q.4) The Following data of the life of the tyre of four brand

| Types | life |
|---|---|
| A | 20, 23, 18, 17, 18, 22, 24 |
| B | 19, 15, 17, 20, 16, 17 |
| C | 21, 19, 22, 17, 20 |
| D | 15, 14, 16, 18, 14, 16 |

→) H0: The average life of tyre is equal A, B, C, D

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

data: values and ind
F = 7.4403 , num df = 3, denom df = 19, p-value=0.001
```

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



Q.5) Import excel file in R software.

```
->) >x= read.csv("C:/Users/admin/Desktop/marks.csv")
>print(x)
     STATS   MATHS
1     40      68
2     45      48
3     42      47
4     25      20
5     37      25
6     36      27
7     49      57
8     59      58
9     20      25
10    27      27

>am = mean(x$STATS)
>am
[1] 37
>me=median(x$STATS)
>me
[1] 38.5
>n= length(x$STATS)
>sd=sqrt((n-1)*var(x$STATS)/n)
>sd
[1] 12.64911
```

```
>am=mean(x$MATHS)
>am
[1] 40.2
>me=median(x$MATHS)
>me=37
>n=length(x$MATHS)
>sd=sqrt((n-1)*var(x$MATHS)/n)
>sd
[1] 16.42437

>cor(x$STATS, x$MATHS)
[1] 0.7802508
```

PRACTICAL : 10

TOPIC :- Non Parametric Test

Q.1] Following are the amount of sulphur oxide emitted by industry in 20 days. Apply sign test to test the hypothesis that the population median is 21.5 at 5% of significance

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...., 26)
> length(x)
[1] 20
> me=21.5
> sp=length(x[x>me])
> sn=length(x[x<me])
> n=sp+sn
> pv=pbinom(sp,n,0.5)
> pv
[1] 0.4119015
```

Since p value is more than ~~0.05~~ 0.05, we accept the hypothesis of 5% of level of significance

Note:- If the alternative is median not equal to ($H_1 \neq me$) ($H_1 < me$) then PV = pbinom (sp, n, 0.5) and if ($H_1 > me$) then pv = binom (sn, n, 0.5).

Q.2] Following is the data of ten observation apply Sign test to test the hypothesis that the population median is 625, begins the alternative it is more than 625.

612, 619, 631, 628, 643, 640,
655, 649, 670, 663

Ho: Population median is 625

```
> u=c(612 ... 663)
> length (u)
[1] 10
> me=625
> sp=length (x[x>me])
> sn=length(u[u<me])
> n=sp+sn
> pv= pbinom(sn, n, 0.5)
> pv
[1] 0.0546875
```

Since pvalue is more than 0.05, we accept the hypothesis of 5% of Significance value



Q3] Following are values of a sample test the hypothesis that a population median is 60 against the alternative it is more than 60 at 5% level of significance using Wilcoxon Signed Ranked Test

63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39
72, 69, 48, 66, 72, 63, 87, 69

→ Ho: population median is 60

```
>x=c(63 .... 69)
>length(u)
[1] 20
>wilcox.test(x,alter="greater", mu=60)
        Wilcoxon signed rank test with continuity
        correction
  data: x
  V= 145 , P-value= 0.02298
  alternative hypothesis: true location is greater than 60
```

Since Pvalue is less than 0.05, we reject the hypothesis of 5% of Significance value

NOTE: If the alternative is less then alter="less" and if it is not equal the alter="two.side"

Q.4) Using W.S.R. Test population median is 12 or less than 12.

15, 17, 24, 25, 20, 21, 32, 28, 13, 25, 24, 26

H0: Population median is 12
H1: Population median is less than 12

```
> x = c(15 ···· 26)
> length(x)
[1] 12
> wilcox.test(x, alter="less", mu12)
```

Wilcoxon signed rank test with continuity correction

data x

V=66, p-value = 0.9986

alternative hypothesis: true location is less than 12.

Since the pvalue is ~~no~~ greater the 0.05 we accept the hypothesis of 5% level of significance



Q.5) The weight of the students before and after they stop smoking below. Using WSRT test that there is no significante chance

| Before (x) | (y) After |
|---|---|
| 65 | 72 |
| 75 | 74 |
| 75 | 72 |
| 62 | 66 |
| 72 | 73 |

H0: Before and after there is no significant change
H1: There is change

```
> x=c(65, 75, 75, 62, 72)
> y=c(72, 74, 72, 66, 73)
> d=x-y
> d
[1] -7 1 3 -4 -1
> wilcox.test(x, alter="two.side", mu=0)
```

Wilcoxon signed rank test with continuity

data x: x

V=15, p-value = 0.05791

alternative hypothesis: true location is not equal to 0

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