KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 4

Lab Code	**: U18MAI4201**
Lab	**: Probability and Statistics**
Course / Branch	**: B.E-CSE,ISE, B.Tech-IT**
Title of the Experiment	**: Applications of Normal Distribution**

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

To predict values and compute probabilities using normal distribution

STEP 2: ACQUISITION

The normal distribution is defined by the following probability density function, where μ is the population mean and σ^2 is the variance.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$

If a random variable X follows the normal distribution, then we write: $X \sim N(\mu, \sigma^2)$
The normal distribution with μ = 0 and σ = 1 is called the standard normal distribution, and is denoted as N(0,1).

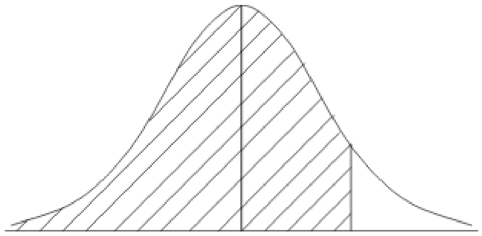
Consider a normal distribution with mean μ and standard deviation σ

R-code for doing the Experiment:

1.	To find $P(X < a) = P(-\infty < X < a)$ **R-code :** pnorm(a, $mean$ = μ, sd = σ)
2.	To find $P(X > a) = P(a < X < \infty)$ R-code: pnorm(a, $mean$ = μ, sd = σ, lower.tail = FALSE)

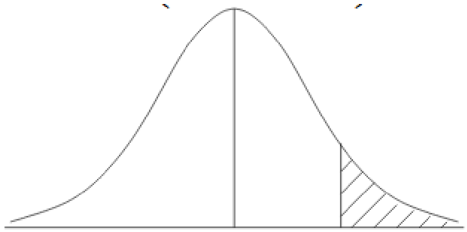
3.	To find $P(a < X < b)$ R-code: pnorm(b, $mean = \mu$, $sd = \sigma$) - pnorm(a, $mean = \mu$, $sd = \sigma$)

To find $P(X < a) = P(-\infty < X < a)$



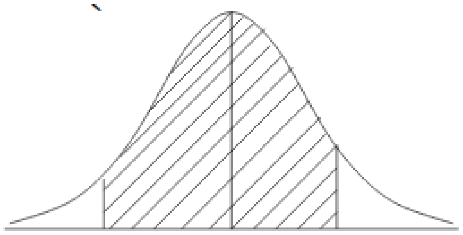
pnorm (a, mean = μ, sd = σ)

To find $P(X > a) = P(a < X < \infty)$



pnorm(a, $mean = \mu$, $sd = \sigma$, lower.tail = FALSE)

To find $P(a < X < b)$



pnorm(b, $mean = \mu$, $sd = \sigma$) - pnorm(a, $mean = \mu$, $sd = \sigma$)

Note:

Use lower.tail=TRUE if you are finding the probability at the lower tail of a confidence interval or if you want to estimate the probability of values no larger than z.

Use lower.tail=FALSE if you aretrying to calculate probability at the upper confidence limit, or you want the probability of values z or larger.

Example

A certain type of storage battery lasts on the average 3.0 years with standard deviation of 0.5 year. Assuming that the battery lives are normally distributed, find the probability that a given battery will last

(i) less than 2.3 years (ii) more than 3.1 years (iii) between 2.5 and 3.5 years

Ans:

(i)

```
pnorm(2.3, mean=3.0, sd=0.5)
 [1] 0.08075666
```

(ii)

```
pnorm(3.1, mean=3.0, sd=0.5, lower.tail=FALSE)
 [1] 0.1586553
```

(iii)

```
pnorm(3.5, mean=3.0, sd=0.5) -   pnorm(2.5, mean=3.0, sd=0.5)
 [1] 0.6826895
```

Task 1

Suppose the heights of men of a certain country are normally distributed with average 68 inches and standard deviation 2.5, find the percentage of men who are

(i) between 66 inches and 71 inches in height

(ii) approximately 6 feet tall (ie, between 71.5 inches and 72.5 inches)

Ans:

(i)

```
pnorm(71, mean=68,sd=2.5)-pnorm(66,mean=68,sd=2.5)
[1] 0.6730749
```

Percentage = 67.30749%

(ii)

```
pnorm(72.5,mean=68,sd=2.5)-pnorm(71.5,mean=68,sd=2.5)
[1] 0.04482634
```

Percentage = 4.482634%

Task 2

The mean yield for one acre plots is 662 kgs with S.D 32. Assuming normal distribution, how many one acre plots in a batch of 1000 plots. Would you expect to yield .

(i) Over 700 kgs

(ii) Below 650 kgs.

(Note: Find the respective probabilities and multiply the probabilities by the number of plots (= 1000) to get the final answers)

Ans:

(i)

```
over=pnorm(700,mean=662,sd=32,lower.tail=FALSE)
print(over*1000)

[1]  117.5152
```

(ii)

```
below=pnorm(650,mean=662,sd=32)

print(below*1000)

[1]  353.8302
```

Task 3

A bore in picking element of a projectile loom part produced is found to have a mean diameter of 2.498 cm. with a SD of 0.012 cm. Determine the percentage of pieces produced you would except to lie within of the drawing limits of 2.5 ± 0.02 **cm.**

Ans:

```
pnorm(2.52,mean=2.498,sd=0.012)-pnorm(2.48,mean=2.498,sd=0.012)
[1]  0.8998163
```

Percentage = 89.998163%

Task 4

An intelligence test is administered to 1000 children. The average score is 42 and S.D is 24. Assuming the test follows normal distribution

i) Find the number of children exceeding the score 60.

ii) Find the number of children with score lying between 20 and 40.

Ans:

(i)

```
pnorm(60,mean=42,sd=24,lower.tail=FALSE)
[1] 0.2266274
```

Number of children exceeding the score 60 = 0.2266274*1000 = 226.6 ≈227

(ii)

```
pnorm(40,mean=42,sd=24)-pnorm(20,mean=42,sd=24)
[1]  0.2871346
```

Number of children with score lying between 20 and 40 = 0.2871346*1000 = 287.1346 ≈287

Task 5

The mean weight of 500 male students in a certain college is 151 *lb* and the standard deviation is 15*lb*. assuming the weights are normally distributed find how many students weight. (i) Between 142 and 155 *lb*. (ii) More than 185 *lb*.

Ans:

(i)

```
one=pnorm(155,mean=151,sd=15)-pnorm(142,mean=151,sd=15)
print(one*500)

[1]  165.442
```

Number of students weigh between 142 and 155 lb ≈166

(ii)

```
two=pnorm(185,mean=151,sd=15,lower.tail=FALSE)
print(two*500)

[1]  5.852649
```

Number of students weigh more than 185 lb ≈6

Task 6

The saving bank account of a customer showed an average balance of Rs.1500 and a standard deviation of Rs.500 .assuming that the account balances are normally distributed.

(i) What percentage of account is over Rs.2000?
(ii) What percentage of account is between Rs.1200 and Rs.1700?

Ans:

(i)

```
pnorm(2000,mean=1500,sd=500,lower.tail=FALSE)
[1]  0.1586553
```

Percentage = 0.1586553*100 = **15.86 %**

(ii)

```
pnorm(1700,mean=1500,sd=500)-pnorm(1200,mean=1500,sd=500)
[1]  0.3811686
```

Percentage = 0.3811686*100 = **38.12 %**

STEP 3: PRACTICE/TESTING

1. **What is the p.d.f. of a normal distribution?**

A continuous random variable X follows normal distribution (or Gaussian distribution) then its p.d.f is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$

The parameters are μ and σ where μ is the mean and σ is the standard deviation of the distribution.

2. **Define standard normal distribution.**

The Normal Distribution with mean = 0 and variance = 1. If X is a normal variate, then $z = \frac{X-\mu}{\sigma}$ is a standard normal variate. The p.d.f of the standard normal variate is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \quad , -\infty < z < \infty$$

Area under the standard normal curve = 1

3. Mention some properties of normal distribution.

1. The graph of the distribution is bell shaped and is called the normal probability curve.
2. The curve is symmetrical about the ordinate at $x = \mu$.
3. x – axis is an asymptote to the curve.
4. For the normal distribution, mean = median = mode.

KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 2

Lab Code	**: U18MAI4201**
Lab	**: Probability and Statistics**
Course / Branch	**: B.E-CSE,ISE, B.Tech-IT**
Title of the Experiment	**: Application of descriptive statistics – Mean, Median, Mode and standard deviation**

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

To find arithmetic mean,median, mode and standard deviation.

STEP 2: ACQUISITION

1. To find the Arithmetic Mean

```
A=c(54,55,53,56,52,52,58,49,50,51)
Mean1=mean(A)
Mean1
[1] 53
```

2. To find the Median

```
A=c(54,55,53,56,52,52,58,49,50,51)
Med=median(A)
Med
[1] 52.5
```

3. To find the mode

# Create the function.

```
mode=function(x){
ux= unique(x)
ux[which.max(tabulate(match(x,ux)))]
}
# Find the mode of the numbers 2,1,2,3,1,2,3,4,1,5,5,3,2,3
x = c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)
# Calculate the mode using the user function.
result= mode(x)
print(result)
```

1. **To find the standard deviation**

```
A=c(54,55,53,56,52,52,58,49,50,51)
Std=sd(A)
Std
Output:
[1] 2.788867
```

Task 1: To find the average set length in a sizing unit

The following set lengths are used in a sizing unit in a factory during a month. Compute the arithmetic mean and median: 1780, 1760, 1690, 1750, 1840, 1920, 1100, 1810, 1050, 1950.

R Code:

```
T1=c(1780, 1760, 1690, 1750, 1840, 1920, 1100, 1810, 1050, 1950)
t1m=mean(T1)
t1md=median(T1)
t1m
t1md
```

Output:

```
> t1m
[1]  1665

> t1md
[1] 1770
```

Task 2: Find the average export of steel in a month from the data given below (in millions of kgs) using mean and median:

Jan'16	105.26
Feb'16	101.05
Mar '16	113.60
Apr'16	105.97
May'16	95.05

Jun'16	93.58
Jul'16	76.21
Aug'16	67.42
Sep'16	77.88
Oct'16	77.97
Nov'16	104.44
Dec'16	174.11

R-Code:

```
T2=c(105.26,101.05,113.60,105.97,95.05,93.58,76.21,67.42,77.88,77.97,104.44,174.11)
t2m=mean(T2)
t2md=median(T2)
t2m
t2md
```

Output:

```
>  t2m

[1]  99.37833


> t2md

[1]  98.05
```

Task 3: To find the average export of raw cotton per year

The following list gives the export quantity of raw cotton (in million kg.) for five consecutive years 2012-2013 to 2016-17: 1945.63, 1864.69, 1093.11, 1297.27, 918.15. Find the mean and median.

R-Code:

```
T3=c(1945.63, 1864.69, 1093.11, 1297.27, 918.15)
```

```
t3m=mean(T3)
t3md=median(T3)
t3m
t3md
```

Output:

```
>  t3m
[1]  1423.77

> t3md
[1]  1297.27
```

To find the Arithmetic mean,median, standard deviation for a frequency distribution

Example

```
d=read.table(header=TRUE,text="Marks          Frequency
+                                        5              15
+                                       15              20
+                                       25              30
+                                       35              20
+                                       45              17
+                                       55              6")
d2= rep(d$Marks, d$Frequency)
multi.fun = function(x) {
c(mean = mean(x), median = median(x), sd = sd(x))
}
multi.fun(d2)
Output:
mean      median    sd
27.03704  25.00000   14.25792
```

Task 4

Find the mean and standard deviation of the frequency distribution:

x:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

R – Code:

```
d=read.table(header=TRUE,text="x              f
                                      1           5
                                      2           9
                                      3           12
```

```
                                        4          17
                                        5          14
                                        6          10
                                        7          6")
t4= rep(d$x, d$f)
multi.fun = function(fr)
{
 c(mean=mean(fr),sd=sd(fr))
}
multi.fun(t4)
```

Output:

```
   mean          sd
4.095890    1.668036
```

Task 5

The following data related to the distance traveled by 520 villagers to buy their weekly requirements.

Miles Traveled: 2 4 6 8 10 13 14 16 18 20
No of Villagers: 38 104140 78 48 42 28 24 16 2
Calculate the arithmetic mean and median.

R – Code:

```
d=read.table(header=TRUE,text="Miles      Villagers
                                 2          38
                                 4          104
                                 6          140
                                 8          78
                                10          48
                                13          42
                                14          28
                                16          24
                                18          16
                                20          2")
t5= rep(d$Miles, d$Villagers)
multi.fun = function(fr)
{
 c(mean=mean(fr),median=median(fr))
}
```

```
multi.fun(t5)
```

Output:

```
    mean       median
7.857692    6.000000
```

Task 6
Calculate the mean and standard deviation for the following:
Size   : 6   7   8   9   10   11   12
Frequency: 3   6   9   13   8   5   4

R – Code:

```
d=read.table(header=TRUE,text="Size          Frequency
                                               6           3
                                               7           6
                                               8           9
                                               9           13
                                              10           8
                                              11           5
                                              12           4")
t6= rep(d$Size, d$Frequency)
multi.fun = function(fr)
{
 c(mean=mean(fr),sd=sd(fr))
}
multi.fun(t6)
```

Output:

```
    mean            sd
9.000000   1.624284
```

Task 7

Find the mean, median and mode for the following data.
14.8, 14.2, 13.8, 13.5, 14.0, 14.2, 14.3, 14.6, 13.9, 14.0, 14.1, 13.2, 13.0, 14.2, 13.5, 13.0, 12.8, 13.9, 14.8, 15.0, 12.8, 13.4, 13.2, 14.0, 13.8, 13.9, 14.0, 14.0, 13.9, 14.8

R – Code:

```
mode=function(x)
{
  ux= unique(x)
  ux[which.max(tabulate(match(x,ux)))]
}
T7=c(14.8,14.2,13.8,13.5,14.0,14.2,14.3,14.6,13.9,14.0,14.1,13.2,13.0,14.2,13.5,13.0,12.8,13.9,14.8,
15.0,12.8,13.4,13.2,14.0,13.8,13.9,14.0,14.0,13.9,14.8)
c(mean=mean(T7),median=median(T7),mode=mode(T7))
```

Output:

```
    mean   median     mode
13.88667 13.95000 14.00000
```

KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 5

Lab Code : **U18MAI4201**

Lab : **Probability and Statistics**

Course / Branch : **B.E-CSE,ISE, B.Tech-IT**

Title of the Experiment : **Applications of Student t-test**

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

1. To apply t-test to test hypothesis about population mean
2. To apply t-test to test hypothesis about two means
3. To apply paired t-test to test hypotheses about means of two dependent samples

STEP 2: ACQUISITION

Student's t – distribution

Student's **t-distribution** has the probability density function given by

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty$$

where ν is the number of degrees of freedom and Γ is the gamma function. This may also be written as

$$f(t) = \frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty$$

Note: (a) The values of $t_\nu(\alpha)$ can be got from the t – table

(b) $t_{\nu}(2\alpha)$gives the critical value of t for a single tail test at α LOS and νd.f

For eg, $t_8(0.05)$ for single tailed test = $t_8(10)$ for two-tailed test = 1.86

Test of Hypothesis about the Population Mean

Test statistic $t = \frac{\overline{x}-\mu}{S/\sqrt{n}}$ follows t – distribution with n-1 degrees of freedom.

where$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$

Null hypothesis H_0 : There is no significant difference between the sample mean $\overline{x}$ and the population mean μ.

If $|t| \leq$ tabulated t, then H_0 is accepted and the difference between $\overline{x}$ and μ is not considered significant.

Assumptions for t – test for population mean

1. The parent population from which the sample is drawn is normal.
2. The sample observations are independent
3. The population standard deviation σ is unknown.

Test of Hypothesis about the difference between two means

To test a hypothesis concerning the difference between the means of two normally distributed populations, when the population variances are unknown, t – test is used.

H_0: The samples have been drawn from populations with same means, ie, $\mu_1 = \mu_2$

Test statistic is $t = \frac{\overline{x}-\overline{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

where$\overline{x} = \frac{\Sigma x}{n_1}$, $\overline{y} = \frac{\Sigma y}{n_2}$,

$$S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i\left(x_i - \overline{x}\right)^2 + \sum_j\left(y_j - \overline{y}\right)^2\right]$$

or $S^2 = \frac{1}{n_1+n_2-2}\left[n_1 s_1^{\ 2} + n_2 s_2^{\ 2}\right]$, where$s_1^{\ 2} = \frac{1}{n_1}\sum_i\left(x_i - \overline{x}\right)^2$, $s_2^{\ 2} = \frac{1}{n_2}\sum_j\left(y_j - \overline{y}\right)^2$

(Note : S^2 is an unbiased estimate of the population variance σ^2)

The test statistic follows t-distribution with $n_1 + n_2 - 2$ degrees of freedom.

If $|t| \leq$ tabulated t, then H$_0$ is accepted and the difference between $\bar{x}$ and μ is not considered significant.

Paired t-test for difference of Means

If the two given samples are dependent, ie, each observation in one sample is associated with a particular observation in the second sample, then we use paired t – test to test whether the means differ significantly or not. Here , both the samples will have same number of units.

The test statistic is

$$t = \frac{\bar{d}}{S/\sqrt{n}} where \ \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i and d_i = x_i - y_i, S^2 = \frac{1}{n-1}\sum\left(d_i - \bar{d}\right)^2$$

$t$follows t – distribution with n-1 d.f. Here n is the number of pairs in the sample

Using R for testing of hypothesis

The R function t.test() can be used to perform both one and two sample t-tests on vectors of data.
The function contains a variety of options and can be called as follows:

t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95)

Here x is a numeric vector of data values and y is an optional numeric vector of data values. If y is excluded, the function performs a one-sample t-test on the data contained in x, if it is included it performs a two-sample t-tests using both x and y.

The option mu provides a number indicating the true value of the mean (or difference in means if you are performing a two sample test) under the null hypothesis. The option alternative is a character string specifying the alternative hypothesis, and must be one of the following: "two.sided" (which is the default), "greater" or "less" depending on whether the alternative hypothesis is that the mean is different than, greater than or less than mu, respectively.

Procedure for doing the Experiment:

1.	To test hypothesis about population mean:
	(a)For a two-tailed test
	x = c(a_1, a_2,..... a_N)
	t.test(x,alternative="two.sided",mu=μ)
	(b) For a one-tailed test
	x = c(a_1, a_2,..... a_N)
	t.test(x,alternative="less"/"greater",mu=μ)

2.	To test hypothesis about two means A= c(a_1, a_2,..... a_m)

	B = c(b_1, b_2,..... b_n)
	t.test(A,B,alternative="two.sided"/"less"/"greater",, var.equal=TRUE)

3.	To use paired t-test
	A= c(a_1, a_2,..... a_m)
	B = c(b_1, b_2,..... b_n)
	t.test(A,B,alternative="greater"/"less"/"two.sided",paired=TRUE)

EXAMPLE – Single mean

Eleven articles produced by a factory were chosen at random and their weights were found to be (in kgs) 63,63,66,67,68,69,70,70,71,71,71 respectively. In the light of the above data, can we assume that the mean weight of the articles produced by the factory is 66 kgs? (Given: the critical value of t for 10 degrees of freedom at 5% LOS is 2.28).

$$NullHypothesis: H_0: \mu = 66$$

$$AlternativeHypothesis: H_1: \mu \neq 66$$

R-code

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

t.test(x,alternative="two.sided",mu=66)
```

Output:

```
One Sample t-test

data:  x

t = 2.3, df = 10, p-value = 0.04425

alternative hypothesis: true mean is not equal to 66

95 percent confidence interval:

66.06533 70.11649

sample estimates:

mean of x

68.09091
```

Conclusion: t -value = 2.3 > 2.228. Hence we reject H_0 and we may conclude that the mean

weight of the articles produced by the factory is not 66

Task 1

Tests made on the breaking strength of 10 pieces of a metal gave the followingresults. 578, 572, 570, 568, 572, 570, 570, 572, 596 and 584 kg.
Test if the mean breaking strength of the wire can be assumed as 577kg.

Null hypothesis: H_0: $\mu = 577$

Alternate hypothesis: H_1: $\mu \neq 577$

R-code

```
 x = c(578,572,570,568,572,570,570,572,596,584)

t.test(x,alternative="two.sided",mu=577)
```

Output:

```
        One Sample t-test


data:  x

t = -0.65408, df = 9, p-value = 0.5294

alternative hypothesis: true mean is not equal to 577

95 percent confidence interval:

 568.9746 581.4254

sample estimates:

mean of x
```

```
575.2
```

Conclusion:

t -value = 0.65408 < 2.262. Hence we accept H_0 and we may conclude that the mean breaking strength of the wire can be assumed as 577kg.

Task 2

The heights of 10 men in a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches?

Null hypothesis H_0: $\mu = 64$

Alternate hypothesis: H_1**:** $\mu > 64$

R-code:

```
x = c(70, 67, 62, 68, 61, 68, 70, 64, 64, 66)

t.test(x,alternative="greater",mu=64)
```

Output :

```
        One Sample t-test

data:  x
t = 2, df = 9, p-value = 0.03828
alternative hypothesis: true mean is greater than 64
95 percent confidence interval:
 64.16689      Inf
sample estimates:
mean of x
      66
```

Conclusion:

t -value = 2 > 1.833 . Hence we reject H_0 and we may conclude that the mean height is greater than 64 inches.

Example 2: Two means

6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group). Their reaction time to a stimulus was measured (in ms).

Placebo group: 91, 87, 99, 77, 88, 91

Treatment group : 101, 110, 103, 93, 99, 104

Can we conclude that the reaction time of the placebo group is less than that of the treatment group? (Required table value of t = 1.1812)

Null hypothesis H_0: $\mu_1 = \mu_2$, ie. the reaction times of the two groups are equal.

Alternate hypothesisH_1: $\mu_1 < \mu_2$ie, the reaction time of the placebo group is less than that of the treatment group

R-code:

```
 Control = c(91, 87, 99, 77, 88, 91)
Treat = c(101, 110, 103, 93, 99, 104)
t.test(Control,Treat,alternative="less", var.equal=TRUE)
```

Output:

```
Two Sample t-test
data:  Control and Treat  t = -3.4456, df = 10, p-value = 0.003136 alternative hypothesis: true difference in means is less than 0
```

Conclusion: t -value =-3.4456 ,$|t|$ =3.4456 > 1.1812. Hence we may conclude that the reaction time of placebo group is less than that of treatment group.

Task 3

Two independent samples are chosen from two schools A and B and common test is given in a subject. The scores of the students are as follows:

School A: 76 68 70 43 94 68 33

School B: 40 48 92 85 70 76 68 22.

Can we conclude that students of school A performed better than students of school B.

Null hypothesisH_0: $\mu_1 = \mu_2$, ie, Students of both schools performed equally well.

Alternate hypothesisH_1: $\mu_1 > \mu_2$ie, Students of school A performed better than students of school B.

R-code:

```
A = c(76,68,70,43,94,68,33)

B = c(40,48,92,85,70,76,68,22)

t.test(A,B,alternative="greater",var.equal=TRUE)
```

Output:

```
            Two Sample t-test

data:  A and B
t = 0.16802, df = 13, p-value = 0.4346
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -18.56956       Inf
sample estimates:
mean of x mean of y
 64.57143  62.62500
```

Conclusion:

t -value = 0.16802 < 1.771. Hence we accept H_0 ,we may conclude that there is no significant difference in the performance of the students of the two schools.

Task 4

Two independent samples of sizes 8 and 7 contained the following values.
Sample 1:19 17 15 21 16 18 16 14
Sample 2:15 14 15 19 15 18 16
Is the difference between the sample means significant?

Null hypothesis H_0: $\mu_1 = \mu_2$, ie, There is no significant difference between the means of the two samples.

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$ ie, There is a significant difference between the means of the two samples.

R-code:

```
Samp1 = c(19,17,15,21,16,18,16,14)

Samp2 = c(15,14,15,19,15,18,16)

t.test(Samp1,Samp2,alternative="two.sided",var.equal=TRUE)
```

Output:

```
        Two Sample t-test


data:  Samp1 and Samp2

t = 0.93095, df = 13, p-value = 0.3688

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

95 percent confidence interval:

 -1.320608  3.320608

sample estimates:

mean of x mean of y

     17        16
```

Conclusion:

t -value = 0.93095 < 2.160. Hence we accept H_0 ,we may conclude that there is no significant difference in the means of the two samples.

Example 3: Paired t-test

A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with either regular or premium gas, decided by a coin toss, and the mileage for that tank was recorded. The mileage was recorded again for the same cars using the other kind of gasoline. The relevant mileages : Regular: 16, 20, 21, 22, 23, 22, 27, 25, 27, 28 Premium :19, 22, 24, 24, 25, 25, 26, 26, 28, 32 . Use a paired t test to determine whether cars get significantly better mileage with premium gas.

Null Hypothesis H_0 :$\mu_1 = \mu_2$, ie, the two types of bulbs are identical regarding length of life.

Alternative Hypothesis: H_1 :$\mu_2 > \mu_1$

```
reg=c(16,20,21,22,23,22,27,25,27,28)

prem=c(19,22,24,24,25,25,26,26,28,32)

t.test(prem,reg,alternative="greater",paired=TRUE)

Paired t-test

data:  prem and reg

t = 4.4721, df = 9, p-value = 0.0007749

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

1.180207     Inf

sample estimates:

mean of the differences

                      2
```

Conclusion: p-value = 0.0007749 < 0.05 Hence we reject H_0 and we may conclude that cars get significantly better mileage with premium gas.

Task 5

The weight gain in pounds under two systems of feeding of calves of 10 pairs of identical twins is given below.

Twin pair	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Weight gain under System A	**43**	**39**	**39**	**42**	**46**	**43**	**38**	**44**	**51**	**43**
Weight gain under System B	**37**	**35**	**34**	**41**	**39**	**37**	**37**	**40**	**48**	**36**

Discuss whether the difference between the two systems of feeding is significant.

Null Hypothesis H_0**:** $\mu_1 = \mu_2$, ie, There is no significant difference between the two systems of feeding.

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$ ie, There is a significant difference between the two systems of feeding.

R-code:

```
SysA=c(43,39,39,42,46,43,38,44,51,43)

SysB=c(37,35,34,41,39,37,37,40,48,36)

t.test(SysA,SysB,alternative="two.sided",paired=TRUE)
```

Output:

```
            Paired t-test


data:  SysA and SysB

t = 6.2644, df = 9, p-value = 0.0001471

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

95 percent confidence interval:

 2.811113 5.988887

sample estimates:

mean of the differences
```

```
        4.4
```

Conclusion:

t -value = 6.2644 > 2.262. Hence we reject H_0 ,we may conclude that there is a significant difference between the two systems of feeding.

Task 6

Ten persons were appointed in the officer cadre in an office. Their performance was noted by giving a test and the marks were recorded out of 100.

Employee	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Before training	**80**	**76**	**92**	**60**	**70**	**56**	**74**	**56**	**70**	**56**
After training	**84**	**70**	**96**	**80**	**70**	**52**	**84**	**72**	**72**	**50**

By applying t test, can it be concluded that the employees have been benefited by the training?

Null hypothesis: H_0**:** $\mu_1 = \mu_2$, ie, The employees have not been benefitted by the training.

Alternate hypothesis H_1: $\mu_1 < \mu_2$ ie, The employees have been benefitted by the training.

R-code:

```
Before=c(80,76,92,60,70,56,74,56,70,56)

After=c(84,70,96,80,70,52,84,72,72,50)

t.test(Before,After,alternative="less",paired=TRUE)
```

Output:

```
        Paired t-test


data:  Before and After
```

```
t = -1.4142, df = 9, p-value = 0.09547

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

    -Inf 1.184826

sample estimates:

mean of the differences

                -4
```

Conclusion:

t -value = 1.4142 < 1.83. Hence we accept H_0 ,we may conclude that the employees have not been benefitted by the training.

STEP 3: PRACTICE/TESTING

1. **Write the test statistic for testing hypothesis about a population mean.**

   T=(x-y)/(s/sqrt(n))

2. **Write the test statistic for testing of hypothesis about the difference between two means .**

   T=(x-y)/s(sqrt(1/n1+1/n2))

3. **Write the test statistic for testing of hypothesis about the difference between means of two dependent samples. (paired t-test)**

   T= d/(s/sqrt(n))

4. **Define level of significance.**

The significance level of an event is the probability that the event could have occurred by chance

KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 6

Lab Code	**: U18MAI4201**
Lab	**: Probability and Statistics**
Course / Branch	**: B.E-CSE,ISE, B.Tech-IT**
Title of the Experiment/experiment	**:Applications of F test**

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

To apply F-test to compare the variances of two samples from normal populations.

STEP 2: ACQUISITION

The null hypothesis is that the ratio of the variances of the populations from which x and y were drawn, or in the data to which the linear models x and y were fitted, is equal to ratio.

Procedure for doing the Experiment:

	R-Code for F-test: var.test(x, y, ratio = 1,alternative = c("two.sided", "less", "greater"),conf.level = 0.95, ...)

Note:

x, y	-	numeric vectors of data values, or fitted linear model objects (inheriting from class "lm").
Ratio	-	the hypothesized ratio of the population variances of x and y.
Alternative	-	a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter.
conf.level	-	confidence level for the returned confidence interval.

In the test statistic, the greater of the two variances S_1^2 and S_2^2 is to be taken in the numerator and v_1 corresponds to the greater variance.

Example:

Two samples of 6 and 7 items respectively have the following values for a variable

Sample 1	**39**	**41**	**42**	**42**	**44**	**40**	
Sample 2	**40**	**42**	**39**	**45**	**38**	**39**	**40**

Do the sample variances differ significantly?

Null Hypothesis: There is no significant difference in sample variances.

Alternative Hypothesis: There is a significant difference in sample variances.

Code:

```
x=c(40,42,39,45,38,39,40)
y=c(39,41,42,42,44,40)
var.test(x, y, ratio = 1,
alternative = c("two.sided"),
conf.level = 0.95)
```

Output:

```
F test to compare two variances
data:  x and y
F = 1.8323, numdf = 6, denomdf = 5, p-value = 0.523
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
 0.2625934 10.9710044
sample estimates:
ratio of variances
        1.832298
```

Critical value of F for (6, 5) d.f. is $F_{0.05} = 4.95$

Conclusion:Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that there is no significant difference in the sample variances.

Task 1:

Two random samples drawn from two normal populations are

Sample 1: 20 16 26 27 23 22 18 24 25 19

Sample 2: 27 33 42 35 32 34 38 28 41 43 30 37

Test whether the populations have the same variances.

Null Hypothesis: H_0: The population have the same variances.

Alternative Hypothesis: H_1: The population have different variances.

R Code:

```
Samp1=c(20,16,26,27,23,22,18,24,25,19)

Samp2=c(27,33,42,35,32,34,38,28,41,43,30,37)

var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Samp1 and Samp2

F = 0.46709, num df = 9, denom df = 11, p-value = 0.2629
```

```
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1301852 1.8272959
sample estimates:
ratio of variances
        0.4670913
```

Critical value of F for (9,11) d.f. is $F_{0.05} = 2.90$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the population have the same variances.

Task 2:

The nicotine content in 2 random samples of tobacco are given below:

Sample 1:	**21**	**24**	**25**	**26**	**27**	
Sample 2:	**22**	**27**	**28**	**30**	**31**	**36**

Test whether the populations have the same variances.

Null Hypothesis: H_0: **The population have the same variances.**

Alternative Hypothesis: H_1: **The population have different variances.**

R Code:

```
Samp1=c(21,24,25,26,27)
Samp2=c(22,27,28,30,31,36)
```

```
var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Samp1 and Samp2

F = 0.24537, num df = 4, denom df = 5, p-value = 0.1981

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.03321253 2.29776367

sample estimates:

ratio of variances

      0.2453704
```

Critical value of F for (4,5) d.f. is $F_{0.05} = 5.19$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the population have the same variances.

Task 3:

2 independent samples of 8 and 7 items have the following values.

Sample 1:	**9**	**11**	**13**	**11**	**15**	**9**	**12**	**14**
Sample 2:	**10**	**12**	**10**	**14**	**9**	**8**	**10**	

Can we conclude that the two samples have drawn from the same normal population.

To test whether the samples come from the same normal population, we have to test for

a. Equality of population means
b. Equality of population variances.

Equality of means is tested using t-test and equality of variances is tested using F-test.

Since t-test assumes $\sigma_1^2 = \sigma_2^2$, we first apply *F*-test and then t-test.

***F*-test:**

Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$

Alternative Hypothesis: H_1: $\sigma_1^2 \neq \sigma_2^2$

R Code:

```
Sample1=c(9,11,13,11,15,9,12,14)

Sample2=c(10,12,10,14,9,8,10)

var.test(Sample1,Sample2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Sample1 and Sample2

F = 1.2108, num df = 7, denom df = 6, p-value = 0.8315

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.2125976 6.1978188
```

```
sample estimates:

ratio of variances

          1.210843
```

Critical value of F for (7,6) d.f. is $F_{0.05} = 4.21$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the population have the same variances.

***t*-test:**

Null Hypothesis: H_0**:** $\mu_1 = \mu_2$,

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$

R Code:

```
Sample1=c(9,11,13,11,15,9,12,14)

Sample2=c(10,12,10,14,9,8,10)

t.test(Sample1,Sample2,alternative="two.sided",var.equal=TRUE)
```

Output:

```
            Two Sample t-test


data:  Sample1 and Sample2

t = 1.2171, df = 13, p-value = 0.2452
```

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

95 percent confidence interval:

 -1.024204  3.667061

sample estimates:

mean of x mean of y

 11.75000  10.42857
```

Conclusion:

t -value = 1.2171 < 2.160 . Hence we accept H_0 ,we may conclude that the population means are same.

Final conclusion:

Since both the null hypothesis are accepted, we may conclude that the given samples have drawn from the same normal population.

Task 4:

Two horses A and B were tested according to the time(in seconds) to run a particular track with the following results:

Horse A: 28 30 32 33 33 29 34

Horse B: 29 30 30 24 27 29

Test whether the two horses have the same running capacity in terms of average and variance of time taken.

F – Test:

Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$ The two horses have the same running capacity in terms of variance of time taken.

Alternative Hypothesis: H_1: $\sigma_1^2 \neq \sigma_2^2$ The two horses does not have the same running capacity in terms of variance of time taken.

R Code:

```
HorseA=c(28,30,32,33,33,29,34)

HorseB=c(29,30,30,24,27,29)

var.test(HorseA,HorseB,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  HorseA and HorseB

F = 0.97604, num df = 6, denom df = 5, p-value = 0.9573

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.1398802 5.8441186

sample estimates:

ratio of variances

      0.9760426
```

Critical value of F for (6,5) d.f. is $F_{0.05} = 4.95$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the two horses have the same running capacity in terms of variance of time taken.

T - Test:

Null Hypothesis: H_0**:** $\mu_1 = \mu_2$, The two horses have the same running capacity in terms of average of time taken.

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$ The two horses does not have the same running capacity in terms of average of time taken.

R Code:

```
HorseA=c(28,30,32,33,33,29,34)

HorseB=c(29,30,30,24,27,29)

t.test(HorseA,HorseB,alternative="two.sided",var.equal=TRUE)
```

Output:

```
        Two Sample t-test


data:  HorseA and HorseB

t = 2.436, df = 11, p-value = 0.03306

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

95 percent confidence interval:

 0.3009233 5.9371719

sample estimates:

mean of x mean of y

 31.28571  28.16667
```

Conclusion:

t -value = 2.436 > 2.201 . Hence we reject H_0 ,we may conclude that the two horses does not have the same running capacity in terms of average of time taken.

Final Conclusion:

In the t-test, the null hypothesis is rejected. So the two horses have the same running capacity only in terms of variance and not in terms of average of time taken.

Task 5:

Two samples are drawn from two normal populations. From the following data test whether the two samples have the same variance at 5% level:

Sample 1:	**60**	**65**	**71**	**74**	**76**	**82**	**85**	**87**		
Sample 2:	**61**	**66**	**67**	**85**	**78**	**63**	**85**	**86**	**88**	**91.**

Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$ the two samples have the same variance.

Alternative Hypothesis: H_1: $\sigma_1^2 \neq \sigma_2^2$ the two samples have different variance.

R Code:

```
Samp1=c(60,65,71,74,76,82,85,87)

Samp2=c(61,66,67,85,78,63,85,86,88,91)

var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Samp1 and Samp2

F = 0.68143, num df = 7, denom df = 9, p-value = 0.6271

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.1623591 3.2866779

sample estimates:
```

ratio of variances

```
0.6814286
```

Critical value of F for (7,9) d.f. is $F_{0.05} = 3.29$

Conclusion:
Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the two samples have the same variance at 5 % level.

STEP 3: PRACTICE/TESTING

1. What is the use of *F*-distribution?

The main use of F distribution is to chech whether two independent samples have been drawn for the same variance or if two independent estimates of the population variance are homogeneous or not, since it is often desirable to compare two variance rather than two averages.

2. State the important properties of *F*-distribution.

1)F- distribution is positively skewed.

2)Value of F lies between 0 and ∞.

3. What is the difference between *F*-test and *t*-test?

t-test is used to test if two sample have the same mean. The assumptions are that they are samples from normal distribution. F-test is used to test if two sample have the same variance.

KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 5

Lab Code : **U18MAI4201**

Lab : **Probability and Statistics**

Course / Branch : **B.E-CSE,ISE, B.Tech-IT**

Title of the Experiment : **Applications of Student t-test**

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

1. To apply t-test to test hypothesis about population mean
2. To apply t-test to test hypothesis about two means
3. To apply paired t-test to test hypotheses about means of two dependent samples

STEP 2: ACQUISITION

Student's t – distribution

Student's **t-distribution** has the probability density function given by

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty$$

where ν is the number of degrees of freedom and Γ is the gamma function. This may also be written as

$$f(t) = \frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty$$

Note: (a) The values of $t_\nu(\alpha)$ can be got from the t – table

(b) $t_{\nu}(2\alpha)$gives the critical value of t for a single tail test at α LOS and νd.f

For eg, $t_8(0.05)$ for single tailed test = $t_8(10)$ for two-tailed test = 1.86

Test of Hypothesis about the Population Mean

Test statistic $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ follows t – distribution with n-1 degrees of freedom.

where$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

Null hypothesis H_0 : There is no significant difference between the sample mean $\bar{x}$ and the population mean μ.

If $|t| \leq$ tabulated t, then H_0 is accepted and the difference between $\bar{x}$ and μ is not considered significant.

Assumptions for t – test for population mean

1. The parent population from which the sample is drawn is normal.
2. The sample observations are independent
3. The population standard deviation σ is unknown.

Test of Hypothesis about the difference between two means

To test a hypothesis concerning the difference between the means of two normally distributed populations, when the population variances are unknown, t – test is used.

H_0: The samples have been drawn from populations with same means, ie, $\mu_1 = \mu_2$

Test statistic is $t = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

where$\bar{x} = \frac{\Sigma x}{n_1}, \bar{y} = \frac{\Sigma y}{n_2},$

$$S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i\left(x_i - \bar{x}\right)^2 + \sum_j\left(y_j - \bar{y}\right)^2\right]$$

or $S^2 = \frac{1}{n_1+n_2-2}\left[n_1 s_1^2 + n_2 s_2^2\right]$, where$s_1^2 = \frac{1}{n_1}\sum_i\left(x_i - \bar{x}\right)^2, s_2^2 = \frac{1}{n_2}\sum_j\left(y_j - \bar{y}\right)^2$

(Note : S^2 is an unbiased estimate of the population variance σ^2)

The test statistic follows t-distribution with $n_1 + n_2$ -2 degrees of freedom.

If $|t| \leq$ tabulated t, then H_0 is accepted and the difference between $\bar{x}$ and μ is not considered significant.

Paired t-test for difference of Means

If the two given samples are dependent, ie, each observation in one sample is associated with a particular observation in the second sample, then we use paired t – test to test whether the means differ significantly or not. Here , both the samples will have same number of units.

The test statistic is

$$t = \frac{\bar{d}}{S/\sqrt{n}} where \ \ \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i and d_i = x_i - y_i, S^2 = \frac{1}{n-1}\sum \left(d_i - \bar{d}\right)^2$$

$t$follows t – distribution with n-1 d.f. Here n is the number of pairs in the sample

Using R for testing of hypothesis

The R function t.test() can be used to perform both one and two sample t-tests on vectors of data.
The function contains a variety of options and can be called as follows:

t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95)

Here x is a numeric vector of data values and y is an optional numeric vector of data values. If y is excluded, the function performs a one-sample t-test on the data contained in x, if it is included it performs a two-sample t-tests using both x and y.

The option mu provides a number indicating the true value of the mean (or difference in means if you are performing a two sample test) under the null hypothesis. The option alternative is a character string specifying the alternative hypothesis, and must be one of the following: "two.sided" (which is the default), "greater" or "less" depending on whether the alternative hypothesis is that the mean is different than, greater than or less than mu, respectively.

Procedure for doing the Experiment:

1.	To test hypothesis about population mean:
	(a)For a two-tailed test
	x = c(a_1, a_2,..... a_N)
	t.test(x,alternative="two.sided",mu=μ)
	(b) For a one-tailed test
	x = c(a_1, a_2,..... a_N)
	t.test(x,alternative="less"/"greater",mu=μ)

2.	To test hypothesis about two means A= c(a_1, a_2,..... a_m)

	B = c(b_1, b_2,..... b_n)
	t.test(A,B,alternative="two.sided"/"less"/"greater",, var.equal=TRUE)

3.	To use paired t-test
	A= c(a_1, a_2,..... a_m)
	B = c(b_1, b_2,..... b_n)
	t.test(A,B,alternative="greater"/"less"/"two.sided",paired=TRUE)

EXAMPLE – Single mean

Eleven articles produced by a factory were chosen at random and their weights were found to be (in kgs) 63,63,66,67,68,69,70,70,71,71,71 respectively. In the light of the above data, can we assume that the mean weight of the articles produced by the factory is 66 kgs? (Given: the critical value of t for 10 degrees of freedom at 5% LOS is 2.28).

$$NullHypothesis: H_0: \mu = 66$$

$$AlternativeHypothesis: H_1: \mu \neq 66$$

R-code

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

t.test(x,alternative="two.sided",mu=66)
```

Output:

```
One Sample t-test

data:  x

t = 2.3, df = 10, p-value = 0.04425

alternative hypothesis: true mean is not equal to 66

95 percent confidence interval:

66.06533 70.11649

sample estimates:

mean of x

68.09091
```

Conclusion: t -value = 2.3 > 2.228. Hence we reject H_0 and we may conclude that the mean

weight of the articles produced by the factory is not 66

Task 1

Tests made on the breaking strength of 10 pieces of a metal gave the followingresults. 578, 572, 570, 568, 572, 570, 570, 572, 596 and 584 kg.
Test if the mean breaking strength of the wire can be assumed as 577kg.

Null hypothesis: H_0: $\mu = 577$

Alternate hypothesis: H_1: $\mu \neq 577$

R-code

```
 x = c(578,572,570,568,572,570,570,572,596,584)

t.test(x,alternative="two.sided",mu=577)
```

Output:

```
        One Sample t-test


data:  x

t = -0.65408, df = 9, p-value = 0.5294

alternative hypothesis: true mean is not equal to 577

95 percent confidence interval:

 568.9746 581.4254

sample estimates:

mean of x
```

```
575.2
```

Conclusion:

t -value = 0.65408 < 2.262. Hence we accept H_0 and we may conclude that the mean breaking strength of the wire can be assumed as 577kg.

Task 2

The heights of 10 men in a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches?

Null hypothesis H_0: $\mu = 64$

Alternate hypothesis: H_1**:** $\mu > 64$

R-code:

```
x = c(70, 67, 62, 68, 61, 68, 70, 64, 64, 66)

t.test(x,alternative="greater",mu=64)
```

Output :

```
        One Sample t-test

data:  x
t = 2, df = 9, p-value = 0.03828
alternative hypothesis: true mean is greater than 64
95 percent confidence interval:
 64.16689      Inf
sample estimates:
mean of x
      66
```

Conclusion:

t -value = 2 > 1.833 . Hence we reject H_0 and we may conclude that the mean height is greater than 64 inches.

Example 2: Two means

6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group). Their reaction time to a stimulus was measured (in ms).

Placebo group: 91, 87, 99, 77, 88, 91

Treatment group : 101, 110, 103, 93, 99, 104

Can we conclude that the reaction time of the placebo group is less than that of the treatment group? (Required table value of t = 1.1812)

Null hypothesis H_0: $\mu_1 = \mu_2$, ie. the reaction times of the two groups are equal.

Alternate hypothesisH_1: $\mu_1 < \mu_2$ie, the reaction time of the placebo group is less than that of the treatment group

R-code:

```
 Control = c(91, 87, 99, 77, 88, 91)
Treat = c(101, 110, 103, 93, 99, 104)
t.test(Control,Treat,alternative="less", var.equal=TRUE)
```

Output:

```
Two Sample t-test
data:  Control and Treat  t = -3.4456, df = 10, p-value = 0.003136 alternative hypothesis: true difference in means is less than 0
```

Conclusion: t -value =-3.4456 ,$|t|$ =3.4456 > 1.1812. Hence we may conclude that the reaction time of placebo group is less than that of treatment group.

Task 3

Two independent samples are chosen from two schools A and B and common test is given in a subject. The scores of the students are as follows:

School A: 76 68 70 43 94 68 33

School B: 40 48 92 85 70 76 68 22.

Can we conclude that students of school A performed better than students of school B.

Null hypothesisH_0: $\mu_1 = \mu_2$, ie, Students of both schools performed equally well.

Alternate hypothesisH_1: $\mu_1 > \mu_2$ie, Students of school A performed better than students of school B.

R-code:

```
A = c(76,68,70,43,94,68,33)

B = c(40,48,92,85,70,76,68,22)

t.test(A,B,alternative="greater",var.equal=TRUE)
```

Output:

```
            Two Sample t-test

data:  A and B
t = 0.16802, df = 13, p-value = 0.4346
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -18.56956       Inf
sample estimates:
mean of x mean of y
 64.57143  62.62500
```

Conclusion:

t -value = 0.16802 < 1.771. Hence we accept H_0 ,we may conclude that there is no significant difference in the performance of the students of the two schools.

Task 4

Two independent samples of sizes 8 and 7 contained the following values.
Sample 1:19 17 15 21 16 18 16 14
Sample 2:15 14 15 19 15 18 16
Is the difference between the sample means significant?

Null hypothesis H_0: $\mu_1 = \mu_2$, ie, There is no significant difference between the means of the two samples.

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$ ie, There is a significant difference between the means of the two samples.

R-code:

```
Samp1 = c(19,17,15,21,16,18,16,14)

Samp2 = c(15,14,15,19,15,18,16)

t.test(Samp1,Samp2,alternative="two.sided",var.equal=TRUE)
```

Output:

```
        Two Sample t-test


data:  Samp1 and Samp2

t = 0.93095, df = 13, p-value = 0.3688

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

95 percent confidence interval:

 -1.320608  3.320608

sample estimates:

mean of x mean of y

     17        16
```

Conclusion:

t -value = 0.93095 < 2.160. Hence we accept H_0 ,we may conclude that there is no significant difference in the means of the two samples.

Example 3: Paired t-test

A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with either regular or premium gas, decided by a coin toss, and the mileage for that tank was recorded. The mileage was recorded again for the same cars using the other kind of gasoline. The relevant mileages : Regular: 16, 20, 21, 22, 23, 22, 27, 25, 27, 28 Premium :19, 22, 24, 24, 25, 25, 26, 26, 28, 32 . Use a paired t test to determine whether cars get significantly better mileage with premium gas.

Null Hypothesis H_0 :$\mu_1 = \mu_2$, ie, the two types of bulbs are identical regarding length of life.

Alternative Hypothesis: H_1 :$\mu_2 > \mu_1$

```
reg=c(16,20,21,22,23,22,27,25,27,28)

prem=c(19,22,24,24,25,25,26,26,28,32)

t.test(prem,reg,alternative="greater",paired=TRUE)
```

```
Paired t-test

data:  prem and reg

t = 4.4721, df = 9, p-value = 0.0007749

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

1.180207     Inf

sample estimates:

mean of the differences

                 2
```

Conclusion: p-value = 0.0007749 < 0.05 Hence we reject H_0 and we may conclude that cars get significantly better mileage with premium gas.

Task 5

The weight gain in pounds under two systems of feeding of calves of 10 pairs of identical twins is given below.

Twin pair	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Weight gain under System A	**43**	**39**	**39**	**42**	**46**	**43**	**38**	**44**	**51**	**43**
Weight gain under System B	**37**	**35**	**34**	**41**	**39**	**37**	**37**	**40**	**48**	**36**

Discuss whether the difference between the two systems of feeding is significant.

Null Hypothesis H_0**:** $\mu_1 = \mu_2$, ie, There is no significant difference between the two systems of feeding.

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$ ie, There is a significant difference between the two systems of feeding.

R-code:

```
SysA=c(43,39,39,42,46,43,38,44,51,43)

SysB=c(37,35,34,41,39,37,37,40,48,36)

t.test(SysA,SysB,alternative="two.sided",paired=TRUE)
```

Output:

```
        Paired t-test


data:  SysA and SysB

t = 6.2644, df = 9, p-value = 0.0001471

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

95 percent confidence interval:

 2.811113 5.988887

sample estimates:

mean of the differences
```

```
        4.4
```

Conclusion:

t -value = 6.2644 > 2.262. Hence we reject H_0 ,we may conclude that there is a significant difference between the two systems of feeding.

Task 6

Ten persons were appointed in the officer cadre in an office. Their performance was noted by giving a test and the marks were recorded out of 100.

Employee	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Before training	**80**	**76**	**92**	**60**	**70**	**56**	**74**	**56**	**70**	**56**
After training	**84**	**70**	**96**	**80**	**70**	**52**	**84**	**72**	**72**	**50**

By applying t test, can it be concluded that the employees have been benefited by the training?

Null hypothesis: H_0**:** $\mu_1 = \mu_2$, ie, The employees have not been benefitted by the training.

Alternate hypothesis H_1: $\mu_1 < \mu_2$ ie, The employees have been benefitted by the training.

R-code:

```
Before=c(80,76,92,60,70,56,74,56,70,56)

After=c(84,70,96,80,70,52,84,72,72,50)

t.test(Before,After,alternative="less",paired=TRUE)
```

Output:

```
        Paired t-test


data:  Before and After
```

```
t = -1.4142, df = 9, p-value = 0.09547
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
    -Inf 1.184826
sample estimates:
mean of the differences
                 -4
```

Conclusion:

t -value = 1.4142 < 1.83. Hence we accept H_0 ,we may conclude that the employees have not been benefitted by the training.

STEP 3: PRACTICE/TESTING

1. **Write the test statistic for testing hypothesis about a population mean.**

   T=(x-y)/(s/sqrt(n))

2. **Write the test statistic for testing of hypothesis about the difference between two means .**

   T=(x-y)/s(sqrt(1/n1+1/n2))

3. **Write the test statistic for testing of hypothesis about the difference between means of two dependent samples. (paired t-test)**

   T= d/(s/sqrt(n))

4. **Define level of significance.**

The significance level of an event is the probability that the event could have occurred by chance

KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 6

Lab Code	**: U18MAI4201**
Lab	**: Probability and Statistics**
Course / Branch	**: B.E-CSE,ISE, B.Tech-IT**
Title of the Experiment/experiment	**:Applications of F test**

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

To apply F-test to compare the variances of two samples from normal populations.

STEP 2: ACQUISITION

The null hypothesis is that the ratio of the variances of the populations from which x and y were drawn, or in the data to which the linear models x and y were fitted, is equal to ratio.

Procedure for doing the Experiment:

	R-Code for F-test: var.test(x, y, ratio = 1,alternative = c("two.sided", "less", "greater"),conf.level = 0.95, ...)

Note:

x, y	-	numeric vectors of data values, or fitted linear model objects (inheriting from class "lm").
Ratio	-	the hypothesized ratio of the population variances of x and y.
Alternative	-	a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter.
conf.level	-	confidence level for the returned confidence interval.

In the test statistic, the greater of the two variances S_1^2 and S_2^2 is to be taken in the numerator and v_1 corresponds to the greater variance.

Example:

Two samples of 6 and 7 items respectively have the following values for a variable

Sample 1	**39**	**41**	**42**	**42**	**44**	**40**	
Sample 2	**40**	**42**	**39**	**45**	**38**	**39**	**40**

Do the sample variances differ significantly?

Null Hypothesis: There is no significant difference in sample variances.

Alternative Hypothesis: There is a significant difference in sample variances.

Code:

```
x=c(40,42,39,45,38,39,40)
y=c(39,41,42,42,44,40)
var.test(x, y, ratio = 1,
alternative = c("two.sided"),
conf.level = 0.95)
```

Output:

```
F test to compare two variances
data:  x and y
F = 1.8323, numdf = 6, denomdf = 5, p-value = 0.523
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
  0.2625934 10.9710044
sample estimates:
ratio of variances
          1.832298
```

Critical value of F for (6, 5) d.f. is $F_{0.05} = 4.95$

Conclusion:Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that there is no significant difference in the sample variances.

Task 1:

Two random samples drawn from two normal populations are

Sample 1: 20 16 26 27 23 22 18 24 25 19

Sample 2: 27 33 42 35 32 34 38 28 41 43 30 37

Test whether the populations have the same variances.

Null Hypothesis: H_0: The population have the same variances.

Alternative Hypothesis: H_1: The population have different variances.

R Code:

```
Samp1=c(20,16,26,27,23,22,18,24,25,19)

Samp2=c(27,33,42,35,32,34,38,28,41,43,30,37)

var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Samp1 and Samp2

F = 0.46709, num df = 9, denom df = 11, p-value = 0.2629
```

```
alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.1301852 1.8272959

sample estimates:

ratio of variances

        0.4670913
```

Critical value of F for (9,11) d.f. is $F_{0.05} = 2.90$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the population have the same variances.

Task 2:

The nicotine content in 2 random samples of tobacco are given below:

Sample 1:	**21**	**24**	**25**	**26**	**27**	
Sample 2:	**22**	**27**	**28**	**30**	**31**	**36**

Test whether the populations have the same variances.

Null Hypothesis: H_0: **The population have the same variances.**

Alternative Hypothesis: H_1: **The population have different variances.**

R Code:

```
Samp1=c(21,24,25,26,27)

Samp2=c(22,27,28,30,31,36)
```

```
var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Samp1 and Samp2
F = 0.24537, num df = 4, denom df = 5, p-value = 0.1981
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.03321253 2.29776367
sample estimates:
ratio of variances
     0.2453704
```

Critical value of F for (4,5) d.f. is $F_{0.05} = 5.19$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the population have the same variances.

Task 3:

2 independent samples of 8 and 7 items have the following values.

Sample 1:	**9**	**11**	**13**	**11**	**15**	**9**	**12**	**14**
Sample 2:	**10**	**12**	**10**	**14**	**9**	**8**	**10**	

Can we conclude that the two samples have drawn from the same normal population.

To test whether the samples come from the same normal population, we have to test for

a. Equality of population means
b. Equality of population variances.

Equality of means is tested using t-test and equality of variances is tested using F-test.

Since t-test assumes $\sigma_1^2 = \sigma_2^2$, we first apply *F*-test and then t-test.

***F*-test:**

Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$

Alternative Hypothesis: H_1: $\sigma_1^2 \neq \sigma_2^2$

R Code:

```
Sample1=c(9,11,13,11,15,9,12,14)

Sample2=c(10,12,10,14,9,8,10)

var.test(Sample1,Sample2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Sample1 and Sample2

F = 1.2108, num df = 7, denom df = 6, p-value = 0.8315

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.2125976 6.1978188
```

```
sample estimates:
ratio of variances
          1.210843
```

Critical value of F for (7,6) d.f. is $F_{0.05} = 4.21$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the population have the same variances.

***t*-test:**

Null Hypothesis: H_0**:** $\mu_1 = \mu_2$,

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$

R Code:

```
Sample1=c(9,11,13,11,15,9,12,14)
Sample2=c(10,12,10,14,9,8,10)
t.test(Sample1,Sample2,alternative="two.sided",var.equal=TRUE)
```

Output:

```
        Two Sample t-test

data:  Sample1 and Sample2
t = 1.2171, df = 13, p-value = 0.2452
```

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

95 percent confidence interval:

 -1.024204  3.667061

sample estimates:

mean of x mean of y

 11.75000  10.42857
```

Conclusion:

t -value = 1.2171 < 2.160 . Hence we accept H_0 ,we may conclude that the population means are same.

Final conclusion:

Since both the null hypothesis are accepted, we may conclude that the given samples have drawn from the same normal population.

Task 4:

Two horses A and B were tested according to the time(in seconds) to run a particular track with the following results:

Horse A: 28 30 32 33 33 29 34

Horse B: 29 30 30 24 27 29

Test whether the two horses have the same running capacity in terms of average and variance of time taken.

F – Test:

Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$ The two horses have the same running capacity in terms of variance of time taken.

Alternative Hypothesis: H_1: $\sigma_1^2 \neq \sigma_2^2$ The two horses does not have the same running capacity in terms of variance of time taken.

R Code:

```
HorseA=c(28,30,32,33,33,29,34)

HorseB=c(29,30,30,24,27,29)

var.test(HorseA,HorseB,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
          F test to compare two variances


data:  HorseA and HorseB

F = 0.97604, num df = 6, denom df = 5, p-value = 0.9573

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.1398802 5.8441186

sample estimates:

ratio of variances

       0.9760426
```

Critical value of F for (6,5) d.f. is $F_{0.05} = 4.95$

Conclusion:

Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the two horses have the same running capacity in terms of variance of time taken.

T - Test:

Null Hypothesis: H_0**:** $\mu_1 = \mu_2$, The two horses have the same running capacity in terms of average of time taken.

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$ The two horses does not have the same running capacity in terms of average of time taken.

R Code:

```
HorseA=c(28,30,32,33,33,29,34)

HorseB=c(29,30,30,24,27,29)

t.test(HorseA,HorseB,alternative="two.sided",var.equal=TRUE)
```

Output:

```
        Two Sample t-test


data:  HorseA and HorseB

t = 2.436, df = 11, p-value = 0.03306

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

95 percent confidence interval:

 0.3009233 5.9371719

sample estimates:

mean of x mean of y

 31.28571  28.16667
```

Conclusion:

t -value = 2.436 > 2.201 . Hence we reject H_0 ,we may conclude that the two horses does not have the same running capacity in terms of average of time taken.

Final Conclusion:

In the t-test, the null hypothesis is rejected. So the two horses have the same running capacity only in terms of variance and not in terms of average of time taken.

Task 5:

Two samples are drawn from two normal populations. From the following data test whether the two samples have the same variance at 5% level:

Sample 1: 60 65 71 74 76 82 85 87

Sample 2: 61 66 67 85 78 63 85 86 88 91.

Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$ the two samples have the same variance.

Alternative Hypothesis: H_1: $\sigma_1^2 \neq \sigma_2^2$ the two samples have different variance.

R Code:

```
Samp1=c(60,65,71,74,76,82,85,87)

Samp2=c(61,66,67,85,78,63,85,86,88,91)

var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)
```

Output:

```
        F test to compare two variances


data:  Samp1 and Samp2

F = 0.68143, num df = 7, denom df = 9, p-value = 0.6271

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.1623591 3.2866779

sample estimates:
```

ratio of variances

0.6814286

Critical value of F for (7,9) d.f. is $F_{0.05} = 3.29$

Conclusion:
Since $F < F_{0.05}$, we accept the null hypothesis and we may conclude that the two samples have the same variance at 5 % level.

STEP 3: PRACTICE/TESTING

1. What is the use of *F*-distribution?

The main use of F distribution is to chech whether two independent samples have been drawn for the same variance or if two independent estimates of the population variance are homogeneous or not, since it is often desirable to compare two variance rather than two averages.

2. State the important properties of *F*-distribution.

1)F- distribution is positively skewed.

2)Value of F lies between 0 and ∞.

3. What is the difference between *F*-test and *t*-test?

t-test is used to test if two sample have the same mean. The assumptions are that they are samples from normal distribution. F-test is used to test if two sample have the same variance.

KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 9

Lab Code **: U18MAI4201**

Lab **: Probability and Statistics**

Course / Branch **: B.E-CSE,ISE, B.Tech-IT**

Title of the Experiment : ANOVA – two way classification

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

To perform analysis of variance for a Randomised Block Design.

STEP 2: ACQUISITION

The data collected from experiments with randomised block design form a two-way classification, classified according to two factors – blocks and treatments. The two-way table has k rows and r columns – ie, N=kr entries.

Consider an agricultural experiment in which we wish to test the effect of k fertilising treatments on the yield of a crop. We divide the plots into r blocks, according to soil fertility, each block containing k plots. The plots in each block will be of homogeneous fertility. I each block, the k treatments are given to the k plots in a random manner in such a way that each treatment occurs only once in each block. The same k treatments are repeated from block to block.

H_{01} : There is no difference in the yield of crop due to treatments

H_{02} : There is no difference in the yield of crop due to blocks

Procedure for doing the Experiment:

Consider a two way table with k rows and r columns

1.	a=c(a_1 , a_2 ,………) (entries entered columnwise)
	f=c("row1","row2","row3","row4","row5")
	k=5
	r=4
	A=gl(k,1,r*k,factor(f))
	A
	B=gl(r,k,k*r)
	B
	av = aov(a ~ A+B)
	summary(av)

Example

The following data represents the number of units of loom crank bushes produced per day turned out by different workers using four different types of machines.

		Machine Type			
		A	**B**	**C**	**D**
	1	**44**	**38**	**47**	**36**
Workers	**2**	**46**	**40**	**52**	**43**
	3	**34**	**36**	**44**	**32**
	4	**43**	**38**	**46**	**33**
	5	**38**	**42**	**49**	**39**

Test whether the 5 men differ with respect to mean productivity and test whether the mean

Productivity is the same for the four different machine types.

R-code:

```
a=c(44,46,34,43,38,38,40,36,38,42,47,52,44,46,49,36,43,32,33,39)
```

```
f=c("w1","w2","w3","w4","w5")

k=5

r=4

worker=gl(k,1,r*k,factor(f))

worker

machine=gl(r,k,k*r)

machine

av = aov(a ~ worker+machine)

summary(av)
```

Output:

```
a=c(44,46,34,43,38,38,40,36,38,42,47,52,44,46,49,36,43,32,33,39)

 f=c("w1","w2","w3","w4","w5")

 k=5

 r=4

worker=gl(k,1,r*k,factor(f))

worker

 [1] w1 w2 w3 w4 w5 w1 w2 w3 w4 w5 w1 w2 w3 w4 w5 w1 w2 w3 w4 w5

Levels: w1 w2 w3 w4 w5

machine=gl(r,k,k*r)

machine

 [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4

Levels: 1 2 3 4

>av = aov(a ~ worker+machine)

>summary(av)

Df Sum Sq Mean Sq F value   Pr(>F)

worker       4  161.5   40.37    6.574  0.00485 **
```

```
machine      3  338.8  112.93  18.388 8.78e-05 ***

Residuals   12   73.7    6.14

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion:

From F-table, $F0.054,12=3.26$

$$F0.053,12=3.49$$

$F1$ = 6.54 >$F0.054,12=3.26$, hence we reject H_{01} and conclude that the 5 workers differ with respect to mean productivity.

$F2$= 18.388 >$F0.053,12=3.49$, hence we reject H_{02} and conclude that the 4 machines differ with respect to mean productivity.

Task 1

A company appoints 4 salesmen A,B,C,D and observes their sales in 3 seasons: summer, winter and monsoon. The figures (in lakhs of Rs.) are given in the following table:

	Salesmen			
Season	**A**	**B**	**C**	**D**
Summer	**45**	**40**	**38**	**37**
Winter	**43**	**41**	**45**	**38**
Monsoon	**39**	**39**	**41**	**41**

Carry out an analysis of variance.

$\mathbf{H_{01}}$: There is no difference between the sales in 3 seasons

$\mathbf{H_{02}}$: There is no difference between the sales of the 4 salesman

R-code:

```
a=c(45,43,39,40,41,39,38,45,41,37,38,41)

f=c("Sum","Win","Mon")

k=3

r=4

season=gl(k,1,r*k,factor(f))

season

salesman=gl(r,k,k*r)

salesman

av = aov(a ~ season+salesman)

summary(av)
```

Output:

```
>season

[1] Sum Win Mon Sum Win Mon Sum Win Mon Sum Win Mon

Levels: Sum Win Mon

>salesman

[1] 1 1 1 2 2 2 3 3 3 4 4 4

Levels: 1 2 3 4

>summary(av)

                Df    Sum Sq    Mean Sq    F value   Pr(>F)

season           2     8.17      4.083      1.8709    0.611
```

```
salesman        3         22.92          7.639          1.000    0.455

Residuals       6        45.83           7.639
```

Conclusion:

From F-table, $F0.056,2=19.3$; $F0.053,6=4.76$

$F1$ = 1.8709 <$F0.056,2=19.3$, hence we accept H_{01} and conclude that there is no difference between the sales in 3 seasons

$F2$= 1.000 <$F0.053,6=4.76$, hence we accept H_{02} and conclude that there is no difference between the sales of the 4 salesman

Task 2

Four different, though supposedly equivalent, forms of a standardized reading achievementtest were given to each of 5 students and the following are the scores which they obtained:

	Student 1	**Student 2**	**Student 3**	**Student 4**	**Student 5**
Form A	**75**	**73**	**59**	**69**	**84**
Form B	**83**	**72**	**56**	**70**	**92**
Form C	**86**	**61**	**53**	**72**	**88**
Form D	**73**	**67**	**62**	**79**	**95**

Perform a two-way analysis of variance to test at the level of significance 0.01 whether it is reasonable to treat the forms as equivalent.

H_{01} : There is no difference between the forms.

H_{02} : There is no difference between the performances of the students.

R-code:

```
a=c(75,83,86,73,73,72,61,67,59,56,53,62,69,70,72,79,84,92,88,95)

f=c("A","B","C","D")

k=4

r=5

form=gl(k,1,r*k,factor(f))

form

student=gl(r,k,k*r)

student

av = aov(a ~ form+student)

summary(av)
```

Output:

```
>form

[1] A B C D A B C D A B C D A B C D A B C D

Levels: A B C D

>student

[1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5

Levels: 1 2 3 4 5

>summary(av)

              Df     Sum Sq    Mean Sq    F value    Pr(>F)

form           3      43.0       14.3      1.9790     0.685

student        4    2326.7      581.7      20.572     2.65e-05 ***

Residuals     12     339.3       28.3
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion:

From F-table, $F0.0112,3=$ 27.05 ; $F0.014,12=$ 5.41

$F1$ = 1.979 <$F0.0112,3=27.05$, hence we accept H_{01} and conclude that there is no difference between the forms.

$F2$= 20.572 >$F0.014,12=5.41$, hence we reject H_{02} and conclude that there is a significant difference between the performances of the students.

Task 3

An experiment was designed to study the performance of different detergents for cleaning fuel injectors. The following ‘cleanness’ readings were obtained with specially designed equipment’s for 12 tanks of gas distributed over 3 different models of engines:

	Engine 1	**Engine 2**	**Engine 3**	**Total**
Detergent A	**45**	**43**	**51**	**139**
Detergent B	**47**	**46**	**52**	**145**
Detergent C	**48**	**50**	**55**	**153**
Detergent D	**42**	**37**	**49**	**128**
Total	**182**	**176**	**207**	**565**

Test at the 0.01 level of significance whether there are differences in the detergents or in the engines.

H_{01} : There is no difference between the performance of the detergents.

H_{02} : There is no difference between the engines.

R-code:

```
a=c(45,47,48,42,43,46,50,37,51,52,55,49)

f=c("A","B","C","D")

k=4

r=3

detergent=gl(k,1,r*k,factor(f))

detergent

engine=gl(r,k,k*r)

engine

av = aov(a ~ detergent+engine)

summary(av)
```

Output:

```
>detergent

        [1] A B C D A B C D A B C D

        Levels: A B C D

>engine

[1] 1 1 1 1 2 2 2 2 3 3 3 3

Levels: 1 2 3

>summary(av)

                Df      Sum Sq      Mean Sq     F value      Pr(>F)

detergent       3       110.92       36.97      11.78       0.00631 **

engine          2       135.17      67.58       21.53       0.00183 **
```

```
Residuals       6         18.83        3.14
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion:

From F-table, $F0.01(3,6)=$ 9.78 ; $F0.012,6=$10.92

$F1$ = 11.78 >$F0.013,6=9.78$, hence we reject H_{01} and conclude that there is a difference between the performance of the detergents.

$F2$= 21.53 >$F0.012,6=10.92$, hence we reject H_{02} and conclude that there is a difference between the engines.

Task 4:

Four experiments determine the moisture content of samples of a powder each observer taking a sample from each of six consignments. The assessments are given below

Observer	**Consignment**					
	1	**2**	**3**	**4**	**5**	**6**
1	**9**	**10**	**9**	**10**	**11**	**11**
2	**12**	**11**	**9**	**11**	**10**	**10**

3	**11**	**10**	**10**	**12**	**11**	**10**
4	**12**	**13**	**11**	**14**	**12**	**10**

Perform an analysis of variance on these data and discuss whether there is any significant difference between consignments or between observers.

H_{01} **:** There is no significant difference between the observers

H_{02} **:** There is no significant difference between the consignments

R-code:

```
a=c(9,12,11,12,10,11,10,13,9,9,10,11,10,11,12,14,11,10,11,12,11,10,10,10)

f=c("1","2","3","4")

k=4

r=6

observer=gl(k,1,r*k,factor(f))

observer

consignment=gl(r,k,k*r)

consignment

av = aov(a ~ observer+consignment)

summary(av)
```

Output:

```
>observer
```

```
[1] 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

Levels: 1 2 3 4

>consignment

[1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5 6 6 6 6

Levels: 1 2 3 4 5 6

>summary(av)

                Df      Sum Sq        Mean Sq       F value     Pr(>F)

observer         3      13.125        4.375         5.000       0.0134 *

consignment      5      9.708         1.942         2.219       0.1064

Residuals       15      13.125        0.875

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion:

From F-table, $F0.053,15=$ 3.29 ; $F0.055,15=$ 2.90

$F1$ = 5.0000 >$F0.053,15=3.29$, hence we reject H_{01} and conclude that there is a significant difference between the observers.

$F2$= 2.219 < $F0.055,15=2.90$, hence we accept H_{02} and conclude that there is no significant difference between the consignments

STEP 3: PRACTICE/TESTING

1. **What is meant by a randomized block design?**

The data collected from experiments with randomised block design form a two-way classification, classified according to two factors – blocks and treatments. The two-way table has *k* rows and *r* columns – i.e., *N*=*kr* entries.

2. **Write the differences between CRD and RBD.**

CRD – One Way Classification ; RBD – Two Way Classification

- RBD is more efficient/accurate than CRD for most types of experimental work.
- RBD is more flexible than CRD since no restrictions are placed on the number of treatments or the number of replications.

3. **Bring out any two advantages of RBD over CRD.**

- This design is more efficient/accurate than CRD, i.e., it has less experimental error.
- This design is more flexible. ie. no restrictions are placed on the number of treatments or the number of replications

2. **When do you apply the analysis of variance technique?**

An ANOVA test is a way to find out if survey or experiment results are significant. They help us to decide whether we should accept or reject the null hypothesis. This technique is used to compare means and the relative variance between them. It is used when three or more populations or samples are to be compared.

KUMARAGURU COLLEGE OF TECHNOLOGY

LABORATORY MANUAL

Experiment Number: 10

Lab Code **: U18MAI4201**

Lab : Probability and Statistics

Course / Branch : B.E-CSE,ISE, B.Tech-IT

Title of the Experiment: Control charts for variables (mean and range chart)

STEP 1: INTRODUCTION

OBJECTIVES OF THE EXPERIMENT

To plot $\bar{X}$- chart and R-chart and comment on the state of control of the process.

STEP 2: ACQUISITION

Statistical Quality Control is a statistical method for finding whether the variation in the quality of the product is due to random causes or assignable causes

Control chart is a graphical device used in statistical quality control for the study and control of the manufacturing process.

There are two types of control charts:

1. Control charts of variables (Mean ($\bar{X}$)and range (R)charts)
2. Control charts of attributes (p-chart, np-chart, c-chart)

The Lower control limit and Upper control limit for mean and range charts

1. $\bar{X}$chart	LCL:$\bar{X}-A2\bar{R}$	UCL: $\bar{X}+A2\bar{R}$
2. R-Chart	LCL: $D3\bar{R}$	UCL: $D4\bar{R}$

Procedure to plot $\bar{X}$ and R charts using RStudio

To install qcc package in RStudio go to the “Tools” menu, select “Install Packages…” and type “qcc” into the packages field being sure to also select “Install Dependencies” and click “Install.”

Load the data from a.csv file with one subgroup per row :

```
my.data = read.csv("my-data.csv",header=FALSE)
```

OR,

Load the data for each subgroup manually:

a1 = c(   )

a2 = c(   )

a3 = c(   )   etc.

If there is more than one subgroup, create a dataframe: my.data = rbind(a1,a2,a3)

Procedure for doing the Experiment:

Suppose the given values are x, y, z, …….

1.	**R code to create dataframe**
	S1=c(a_1 , a_2 ,……)
	S2=c(b_1 , b_2 ,……)
	A= as.data.frame(rbind(S1,S2,……..))
	A

2.	**For X chart:**
	Xbarchart= qcc(data = A,
	type = "xbar",
	sizes = n, # n=number of items in each sample
	title = "X-bar Chart ",
	plot = TRUE)

3.	**For R chart:**
	rchart = qcc(data = A,
	type = "R",
	sizes = n, # n=number of items in each sample
	title = "R Chart",
	plot = TRUE)

Example

The measurements are given below with 5 samples each containing 5 items at equal intervals of time. Construct $\bar{X}$ and R charts and comment on the state of control.

Sample no	Measurements				
1	46	45	44	43	42
2	41	41	44	42	40
3	40	40	42	40	42
4	42	43	43	42	45
5	43	44	47	47	45

#R code to create dataframe

```
S1=c(46,45,44,43,42)

S2=c(41,41,44,42,40)

S3=c(40,40,42,40,42)

S4=c(42,43,43,42,45)

S5=c(43,44,47,47,45)

A= as.data.frame(rbind(S1,S2,S3,S4,S5))

A
```

#**For*X* chart:**

```
Xbarchart= qcc(data = A,

type = "xbar",

sizes = 5,

title = "X-bar Chart ",

plot = TRUE)
```

Output:

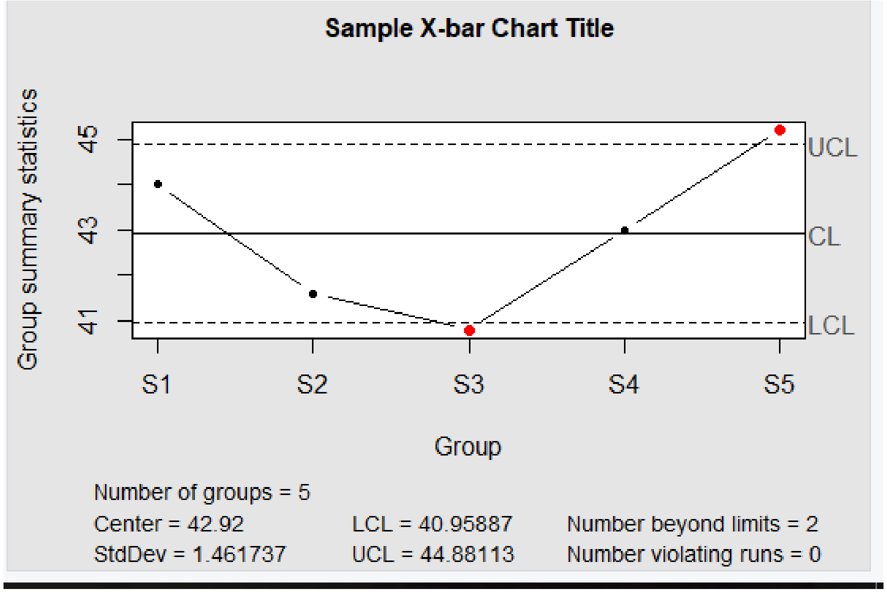
```
  V1 V2 V3 V4 V5

S1 46 45 44 43 42
```

S2 41 41 44 42 40

S3 40 40 42 40 42

S4 42 43 43 42 45

S5 43 44 47 47 45



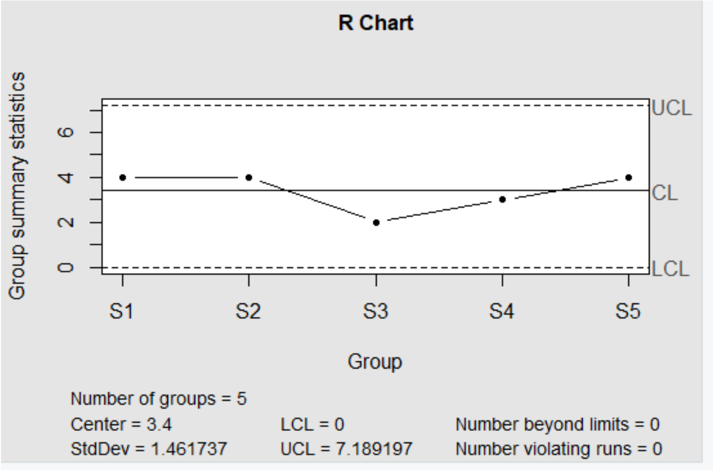
```
Number of groups = 5
Center = 42.92          LCL = 40.95887      Number beyond limits = 2
StdDev = 1.461737       UCL = 44.88113      Number violating runs = 0
```

For R chart:

R-code:

```
rchart = qcc(data = A,

type = "R",

sizes = 5,

title = "R Chart",

plot = TRUE)
```

Output:



```
Number of groups = 5
Center = 3.4              LCL = 0              Number beyond limits = 0
StdDev = 1.461737         UCL = 7.189197       Number violating runs = 0
```

Conclusion:

In $\bar{X}$ chart, two points are beyond the control limits, so as far as sample mean is concerned, the system is out of control.

In R chart, all points are within the control limits, so as far as variability is concerned, the system is under control.

On the whole, the system is out of control.

Task 1

Samples of five ring bobbins each selected from a ring frame for eight shifts have shown following results of count of yarn.

Sample no.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Count of yarn	**27.5**	**27.4**	**25.4**	**28.5**	**28.5**	**28.9**	**28.0**	**28.4**
	28.5	**26.9**	**26.9**	**28.0**	**29.0**	**29.5**	**28.5**	**28.5**
	28	**26.0**	**28.0**	**29.2**	**28.5**	**30.0**	**27.8**	**28.4**
	26.9	**28.7**	**26.7**	**29.0**	**28.5**	**29.4**	**28.0**	**28.0**
	28.6	**29.0**	**28.2**	**28.7**	**28.0**	**28.9**	**28.1**	**28.7**

Draw $\bar{X}$and R chart for the above data and write conclusion about the state of the process.

R-Code:

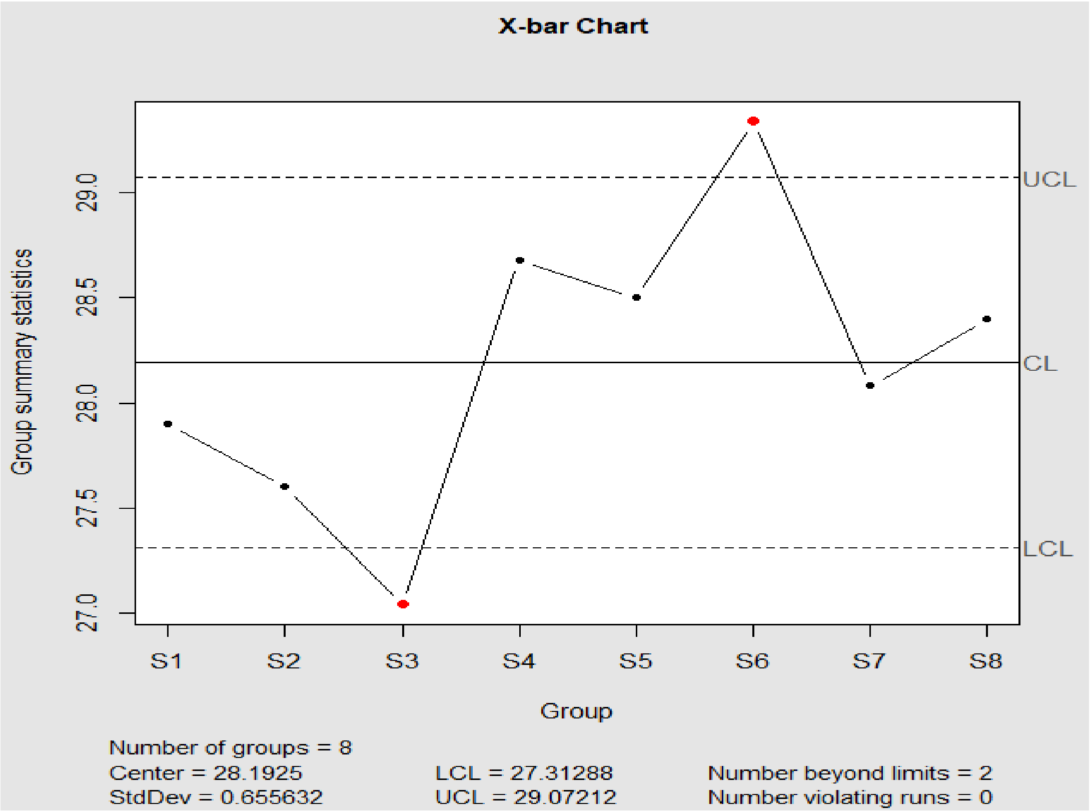
```
S1=c(27.5,28.5,28,26.9,28.6)

S2=c(27.4,26.9,26,28.7,29.0)

S3=c(25.4,26.9,28,26.7,28.2)

S4=c(28.5,28,29.2,29,28.7)

S5=c(28.5,29,28.5,28.5,28)

S6=c(28.9,29.5,30,29.4,28.9)

S7=c(28,28.5,27.8,28,28.1)

S8=c(28.4,28.5,28.4,28,28.7)

A= as.data.frame(rbind(S1,S2,S3,S4,S5,S6,S7,S8))

A
```

Output:

```
> A

     V1   V2   V3   V4   V5

S1 27.5 28.5 28.0 26.9 28.6

S2 27.4 26.9 26.0 28.7 29.0

S3 25.4 26.9 28.0 26.7 28.2

S4 28.5 28.0 29.2 29.0 28.7

S5 28.5 29.0 28.5 28.5 28.0

S6 28.9 29.5 30.0 29.4 28.9

S7 28.0 28.5 27.8 28.0 28.1

S8 28.4 28.5 28.4 28.0 28.7
```

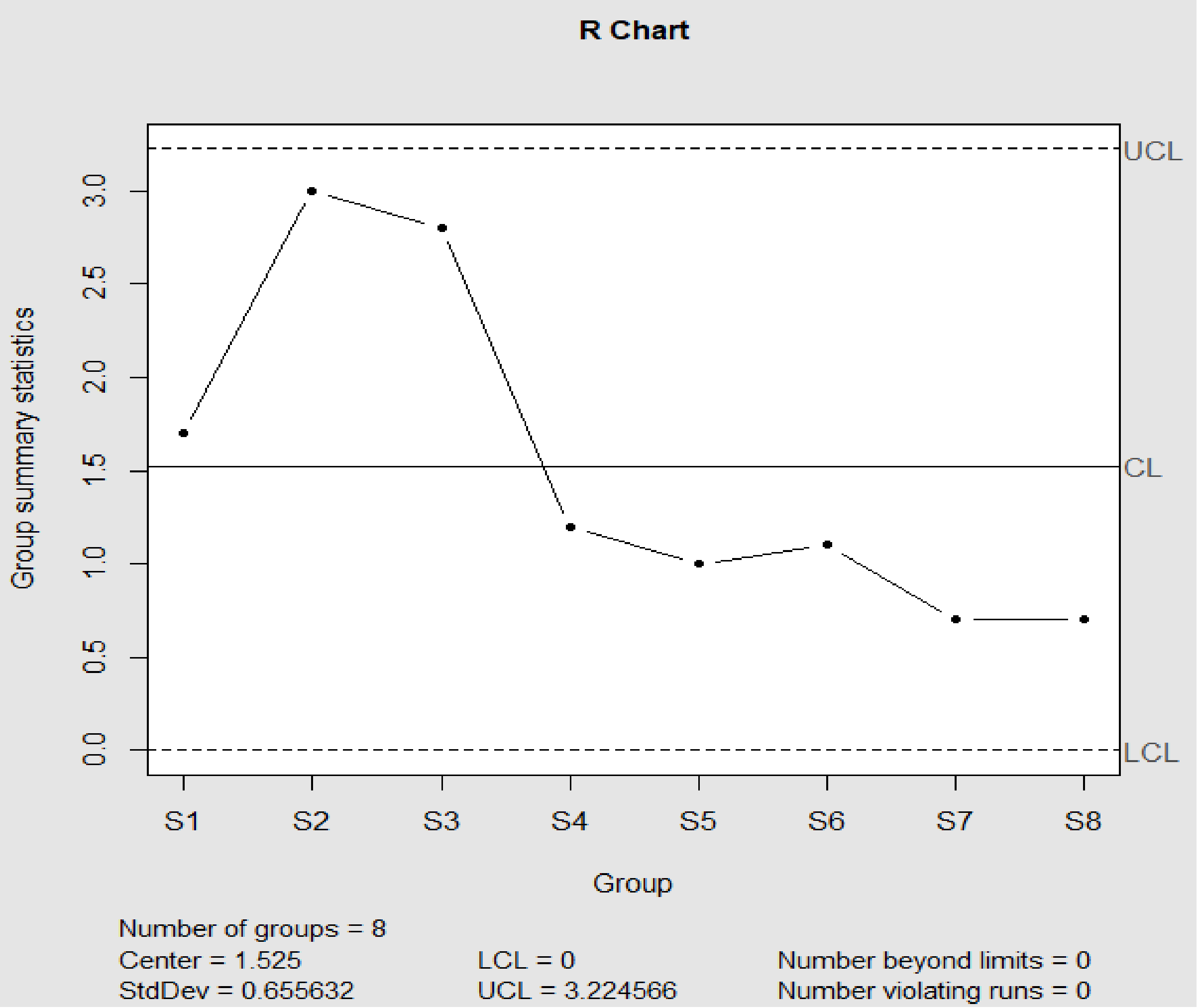
***$\bar{X}$* chart:**

Xbarchart= qcc(data = A,type = "xbar",sizes = 5,title = "X-bar Chart ",plot = TRUE)



R – Chart:

rchart = qcc(data = A,type = "R",sizes = 5,title = "R Chart",plot = TRUE)



Conclusion:

In $\bar{X}$ chart, two points are beyond the control limits, so as far as sample mean is concerned, the system is out of control.

In R chart, all points are within the control limits, so as far as variability is concerned, the system is under control.

On the whole, the system is out of control.

Task 2:

The following data gives the measurements of 10 samples each of size 5, in a production process taken at intervals of 2 hours. Draw the control charts for the mean and range and comment on the state of control:

Sample No.	1	2	3	4	5	6	7	8	9	10
Measurements	47	52	48	49	50	55	50	54	49	53
	49	55	53	49	53	55	51	54	55	50
	50	47	51	49	48	50	53	52	54	54
	44	56	50	53	52	53	46	54	49	47
	45	50	53	45	47	57	50	56	53	51

R-Code:

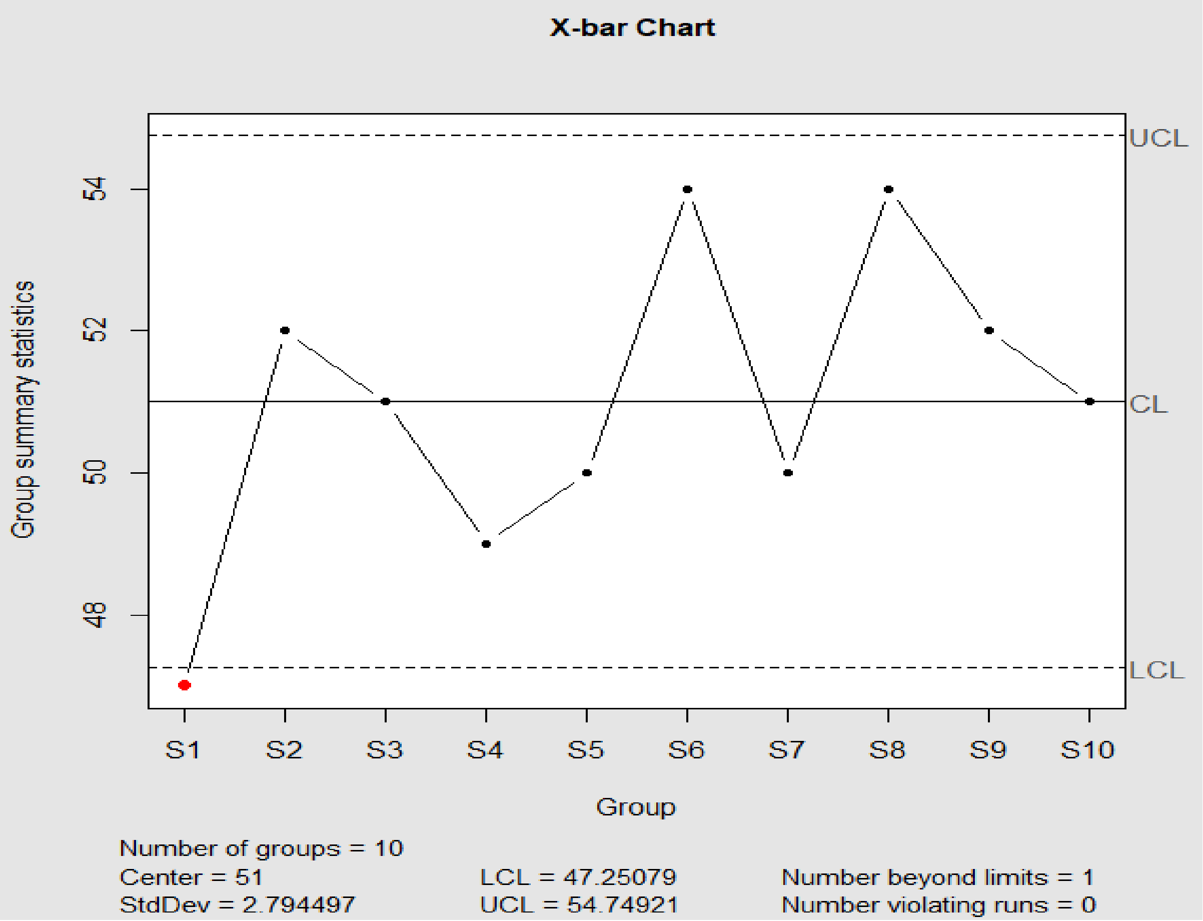
```
S1=c(47,49,50,44,45)

S2=c(52,55,47,56,50)

S3=c(48,53,51,50,53)

S4=c(49,49,49,53,45)

S5=c(50,53,48,52,47)

S6=c(55,55,50,53,57)

S7=c(50,51,53,46,50)

S8=c(54,54,52,54,56)

S9=c(49,55,54,49,53)

S10=c(53,50,54,47,51)

A= as.data.frame(rbind(S1,S2,S3,S4,S5,S6,S7,S8,S9,S10))

A
```

Output:

```
> A
   V1 V2 V3 V4 V5
S1  47 49 50 44 45
S2  52 55 47 56 50
S3  48 53 51 50 53
S4  49 49 49 53 45
S5  50 53 48 52 47
S6  55 55 50 53 57
S7  50 51 53 46 50
S8  54 54 52 54 56
S9  49 55 54 49 53
S10 53 50 54 47 51
```

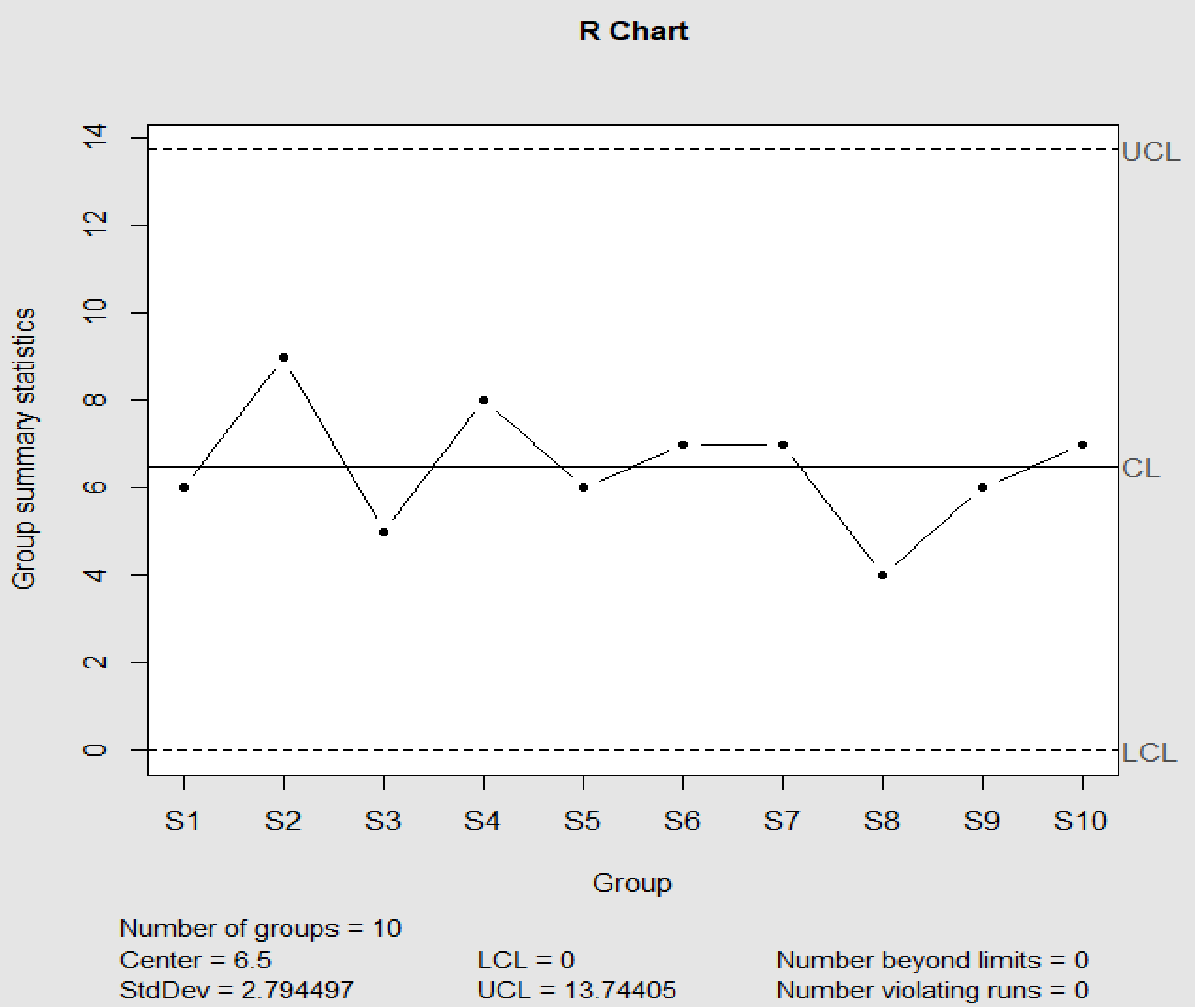
$\bar{X}$ chart:

```
Xbarchart= qcc(data = A,type = "xbar",sizes = 5,title = "X-bar Chart ",plot = TRUE)
```



R – Chart:

```
rchart = qcc(data = A,type = "R",sizes = 5,title = "R Chart",plot = TRUE)
```



Conclusion:

In $\bar{X}$ chart, one point lies beyond the control limits, so as far as sample mean is concerned, the system is out of control.

In R chart, all points are within the control limits, so as far as variability is concerned, the system is under control.

On the whole, the system is out of control.

Task 3:

Plot the mean and range charts for the following data

Rotation Time (msec)

Sample Number	1	2	3	4	5	6
1	**469.92**	**468.67**	**479.76**	**454.38**	**469.58**	**454.46**
2	**457.34**	**454.37**	**475.28**	**453.46**	**480.03**	**480.40**
3	**473.96**	**459.26**	**460.42**	**462.04**	**450.60**	**451.52**
4	**480.06**	**469.86**	**456.42**	**460.63**	**465.66**	**466.99**
5	**467.46**	**476.56**	**474.01**	**465.34**	**475.27**	**462.97**
6	**473.06**	**475.86**	**472.97**	**454.93**	**470.73**	**466.24**
7	**456.27**	**476.37**	**479.50**	**459.86**	**470.73**	**452.35**

R-Code:

```
S1=c(469.92,468.67,479.76,454.38,469.58,454.46)

S2=c(457.34,454.37,475.28,453.46,480.03,480.4)

S3=c(473.96,459.26,460.42,462.04,450.6,451.52)

S4=c(480.06,469.86,456.42,460.63,465.66,466.99)

S5=c(467.46,476.56,474.01,465.34,475.27,462.97)

S6=c(473.06,475.86,472.97,454.93,470.73,466.24)

S7=c(456.27,476.37,479.50,459.86,470.73,452.35)

A= as.data.frame(rbind(S1,S2,S3,S4,S5,S6,S7))

A
```

Output:

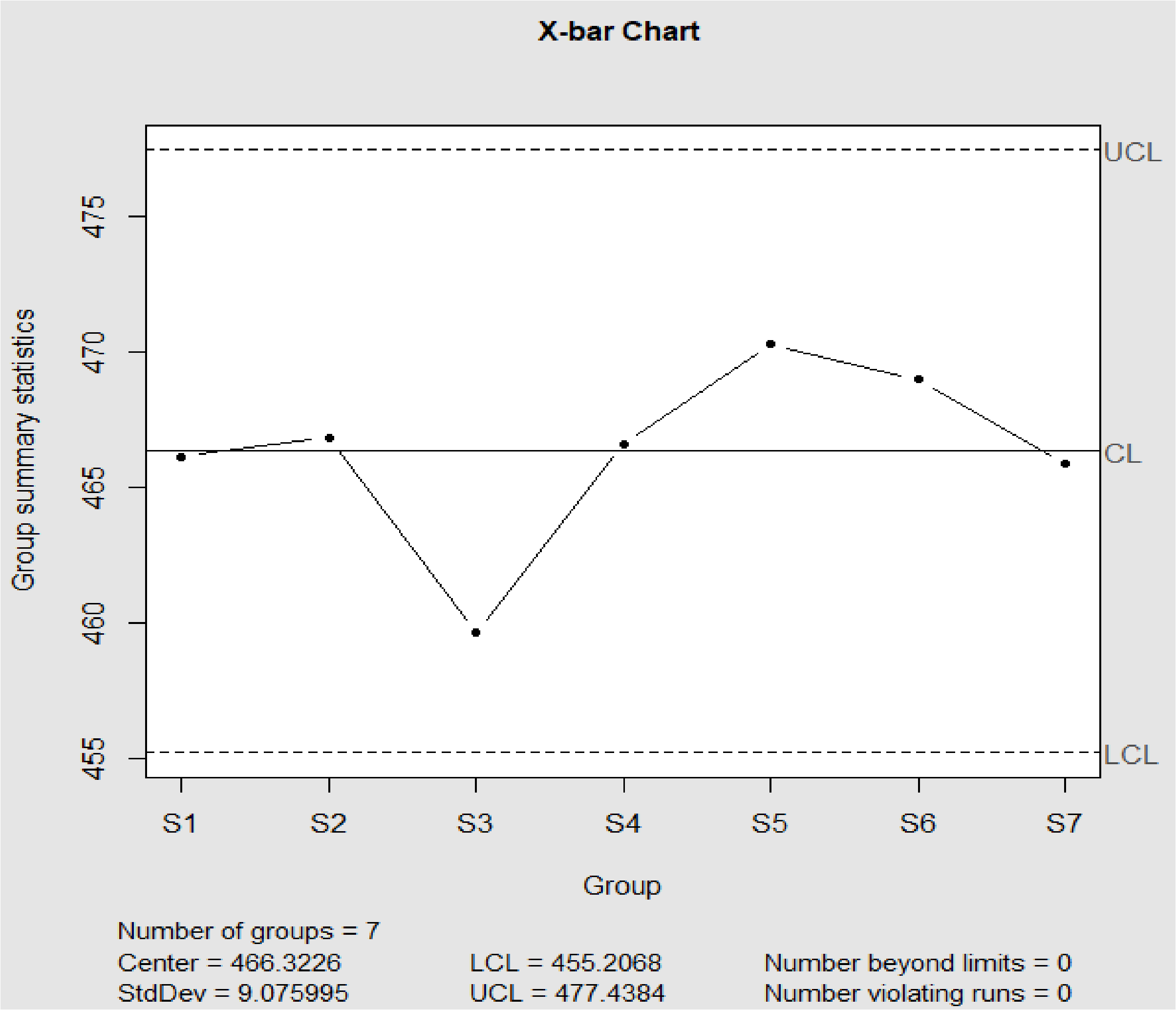
```
> A

      V1     V2     V3     V4     V5     V6

S1 469.92 468.67 479.76 454.38 469.58 454.46

S2 457.34 454.37 475.28 453.46 480.03 480.40

S3 473.96 459.26 460.42 462.04 450.60 451.52

S4 480.06 469.86 456.42 460.63 465.66 466.99

S5 467.46 476.56 474.01 465.34 475.27 462.97
```

S6 473.06 475.86 472.97 454.93 470.73 466.24

S7 456.27 476.37 479.50 459.86 470.73 452.35

$\bar{X}$ chart:

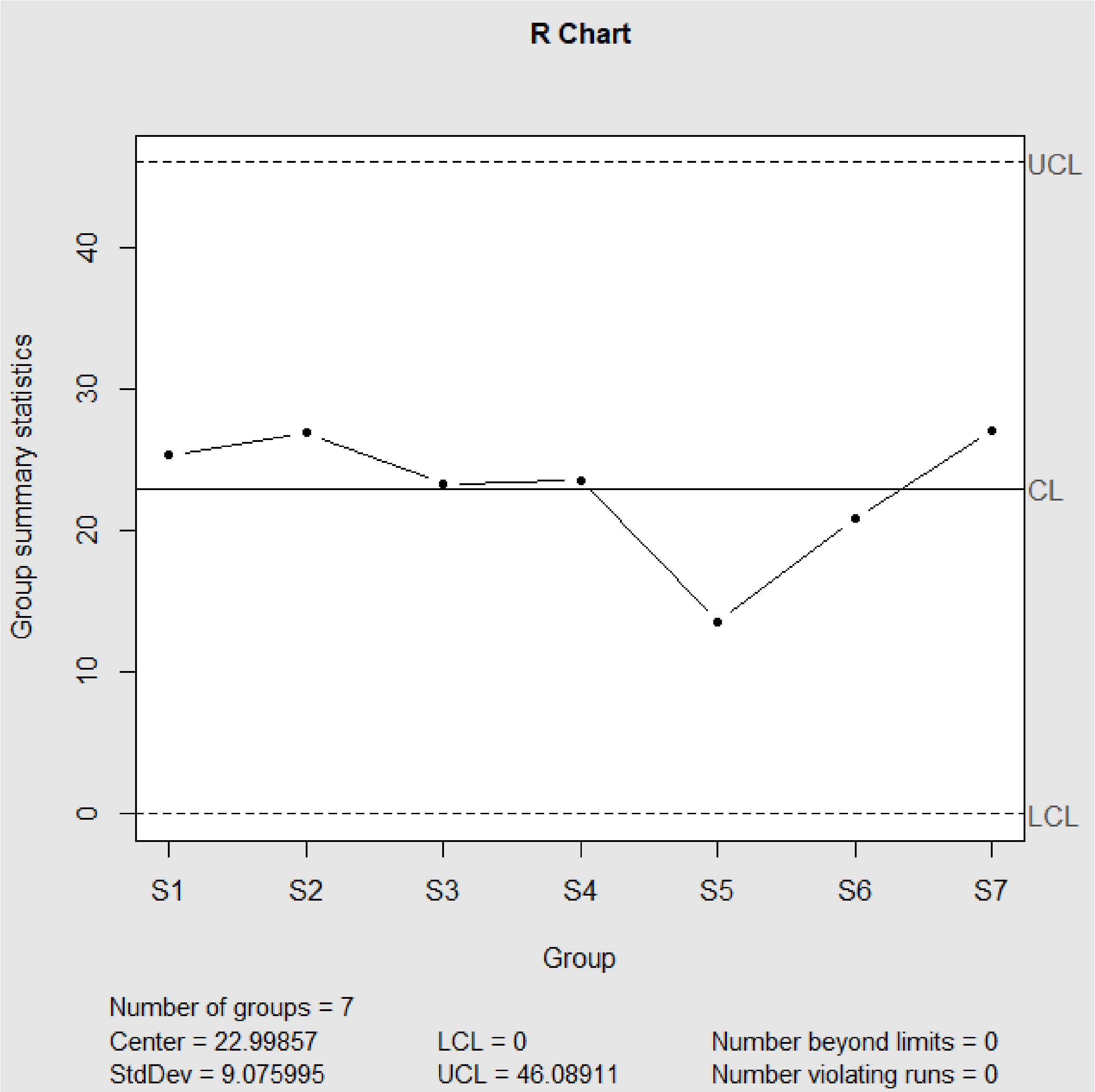
```
Xbarchart= qcc(data = A,type = "xbar",sizes = 5,title = "X-bar Chart ",plot = TRUE)
```



```
Number of groups = 7
Center = 466.3226          LCL = 455.2068          Number beyond limits = 0
StdDev = 9.075995          UCL = 477.4384          Number violating runs = 0
```

R – Chart:

rchart = qcc(data = A,type = "R",sizes = 5,title = "R Chart",plot = TRUE)



```
Number of groups = 7
Center = 22.99857        LCL = 0           Number beyond limits = 0
StdDev = 9.075995        UCL = 46.08911    Number violating runs = 0
```

Conclusion:

In $\bar{X}$ chart, all points are within the control limits, so as far as sample mean is concerned, the system is under control.

In R chart, all points are within the control limits, so as far as variability is concerned, the system is under control.

On the whole, the system is under control

STEP 3: PRACTICE/TESTING

1. **Define Statistical Quality Control.**

- The technique of applying statistical methods based on sampling to establish quality standards and to maintain it in the most economical manner.
- The objective of the SQC is to devise the fact that even after the quality standards have been specified, some variation in quality e statistical methods that isolate assignable variation from random variation and enable us to detect, identify and eliminate the assignable causes of variation.

2. **What are control charts? What are the types of control charts?**

Control Chart is a important statistical tool used for the study and control of repetitive processes. A control chart accepts the normal variation due to chance causes but eliminates entirely the errors due to assignable causes.

Two types of Control Charts:

1. Contorl Charts for Variables - $\bar{X}$ Chart, R Chart
2. Control Charts for Attributes - p Chart, np Chart, c Chart

3. **Write the Lower control limit and Upper control limit for mean and range charts.**

Mean Chart:

Lower Control Limit**:** $\bar{\bar{X}}$ - $A2R$-

Control Limit: $\bar{\bar{X}}$

Upper Control Limit: $\bar{\bar{X}}$+ $A2R$-

R Chart:

Lower Control Limit**:** $D3R$-

Control Limit: R-

Upper Control Limit: $D4R$-