Experiment 13:

Illustrate the Correlation and Covariance analysis using R

- Aim: To analyze data using Correlation and Covariance measures in R

Exp. 13: Correlation and Covariance using R

- When dealing with more than one variable, we need to test their relationships with each other
- Two simple measures
    - Correlation
    - Covariance

Exp. 13: Correlation using R

- **Correlation** is used to evaluate the association between two or more variables
- It is computed as:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

If $r_{xy} = 0$ No Correlation

$r_{xy} > 0$ Positive Correlation

$r_{xy} < 0$ Negative Correlation

- x and y are two vectors of length n
- $\bar{x}$ and $\bar{y}$ corresponds to the means of x and y, respectively.

Exp. 13: Correlation using R

Syntax:

cor(x, y , method)

- **x** is a numeric vector, matrix or data frame
- **y is** NULL (default) or a vector, matrix or data frame with compatible dimensions to x
- **method** ="pearson" or "kendall" or "spearman" #Default:pearson

➢ Correlation range is [-1,+1]

Exp. 13: Correlation using R

```
> pcor=cor(airquality$Wind,airquality$Temp)
> print(pcor)
[1] -0.4579879                          #Negative Correlation

>corw=cor(women$height,women$weight)
>print(corw)
[1] 0.9954948                           # Positive Correlation
```

Exp. 13: Correlation using R

```
> icor=cor(iris[1:4])
> print(icor)                                    #Correlation Matrix
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

Exp. 13: Covariance using R

- **Covariance** of two variables *x* and *y* in a data set measures how the two are linearly related.
- A positive covariance would indicate a positive linear relationship between the variables, and a negative covariance would indicate the opposite.
- Covariance range is [-∞,+∞]

Exp. 13: Covariance using R

- **Covariance** of two variables *x* and *y* in a data set is computed as:

Population Covariance Formula

$$Cov(x,y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N}$$

Sample Covariance

$$Cov(x,y) = \frac{\Sigma(x_i-\bar{x})(y_i-y)}{N-1}$$

Exp. 13: Covariance using R

Syntax:

cov(x, y , method)

- **x** is a numeric vector, matrix or data frame
- **y is** NULL (default) or a vector, matrix or data frame with compatible dimensions to x
- **method** ="pearson" or "kendall" or "spearman" #Default:pearson

Exp. 13: Covariance using R

```
> acov=cov(airquality$Wind,airquality$Temp)
> print(acov)
[1] -15.27214

> wcov=cov(women$height,women$weight)
> print(wcov)
[1] 69
```

Exp. 13: Covariance using R

```
> icov=cov(iris[1:4])
> print(icov)                                    #Covariance Matrix
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    0.6856935  -0.0424340    1.2743154   0.5162707
Sepal.Width    -0.0424340   0.1899794   -0.3296564  -0.1216394
Petal.Length    1.2743154  -0.3296564    3.1162779   1.2956094
Petal.Width     0.5162707  -0.1216394    1.2956094   0.5810063

#Diagonal values are variance of each attribute
```