



# Business Statistics

## Teachers' Guide

### Grade 13

(Implemented from 2018)

Department of Commerce  
Faculty of Science and Technology  
National Institute of Education  
Maharagama

web : [www.nie.lk](http://www.nie.lk)  
Email : [info@nie.lk](mailto:info@nie.lk)

# Business Statistics

Teachers' Guide

Grade 13

Department of Commerce  
Faculty of Science and Technology  
National Institute of Education  
Maharagama

Business Statistics

Grade 13 - Teachers' Guide

© National Institute of Education

First Print 2018

ISBN :

Department of Commerce  
Faculty of Science and Technology  
National Institute of Education

Printed By :

# Content

|   | Page    |
|---|---------|
| Message from the Hon. minister of Education       | iii     |
| Message from the Director General                 | iv      |
| foreword  | v       |
| Curriculum Committee                              | vi-viii |
| Instructions for referring to the Teachers' Guide | ix      |
| Content   | x       |
| Learning Outcomes and Model Activities            | 1-299   |

*Message from the Director General ...*

With the primary objective of realizing the National Educational Goals recommended by the National Education Commission, the then prevalent content based curriculum was modernized, and the first phase of the new competency based curriculum was introduced to the eight year curriculum cycle of the primary and secondary education in Sri Lanka in the year 2007

The second phase of the curriculum cycle thus initiated was introduced to the education system in the year 2015 as a result of a curriculum rationalization process based on research findings and various proposals made by stake holders.

Within this rationalization process the concepts of vertical and horizontal integration have been employed in order to build up competencies of students, from foundation level to higher levels, and to avoid repetition of subject content in various subjects respectively and furthermore, to develop a curriculum that is implementable and student friendly.

The new Teachers' Guides have been introduced with the aim of providing the teachers with necessary guidance for planning lessons, engaging students effectively in the learning teaching process, and to make Teachers' Guides will help teachers to be more effective within the classroom. Further, the present Teachers' Guides have given the necessary freedom for the teachers to select quality inputs and activities in order to improve student competencies. Since the Teachers' Guides do not place greater emphasis on the subject content prescribed for the relevant grades, it is very much necessary to use these guides along with the text books compiled by the Educational Publications Department if, Guides are to be made more effective.

The primary objective of this rationalized new curriculum, the new Teachers' Guides, and the new prescribed texts is to transform the student population into a human resource replete with the skills and competencies required for the world of work, through embarking upon a pattern of education which is more student centered and activity based.

I wish to make use of this opportunity to thank and express my appreciation to the members of the Council and the Academic Affairs Board of the NIE the resource persons who contributed to the compiling of these Teachers' Guides and other parties for their dedication in this matter.

Dr. (Mrs.) Jayanthi Gunasekara  
Director General  
National Institute of Education

## **Guidance and Approval**

Academic Affairs Board  
National Institute of Education

## **Subject Cordinator**

Ms. M. A. Indra Pathmini Perera  
Senior Lecturer,  
National Institute of Education

## **Curriculum Committee**

Ms P H Kusumawathi

Director (Duty Covering),  
Department of Commerce,  
National Institute of Education

Ms. M. A. Indra Pathmini Perera

Senior Lecturer,  
Department of Commerce,  
National Institute of Education

Dr. H. M. L. K. Herath

Senior Lecturer  
Wayamba University of Sri Lanka.

Mr. K. A. Darmasena

Senior Lecturer  
University of Sri Jayawardenepura.

Mr. S. A. C. Stanley Silva

Senior Lecturer  
University of Sri Jayawardenepura.

Ms. M. Kamani Perera

Director Education  
Ministry of Education.

Mr. W. M. P. G. Edirisingha

SLTS 1  
Vishaka Vidyalaya, Colombo 07

Ms. M. L. S. L. Perera

SLTS 11  
Ananda Vidyalaya, Colombo 07

Ms. M. Neranjan

SLTS 1  
Ramanadan Hindu College,  
Colombo 04

Ms. Gayani Arunika Perera

SLTS 11  
Panadura Balika Vidyalaya,  
Panadura.

|                          |   |
|--------------------------|---|
| Mr. Prabakaran           | Senior Lecturer<br>Department of Commerce<br>National Institute of Education    |
| Mr. Ananda Maddumage     | Lecturer<br>Department of Commerce<br>National Institute of Education           |
| Mr. D.L.C.R.Ajith Kumara | Lecturer<br>Department of Commerce<br>National Institute of Education           |
| Mr.S.R.Rathnajeewa       | Assistant Lecturer<br>Department of Commerce<br>National Institute of Education |
| Mr.S.K.Rathnasiri Silva  | Senior Lecturer<br>University of Sri Jayawardanapura                            |

**Writing Panel :**

|                             |  |
|-----------------------------|--|
| Mr. W. M. P. G. Edirisingha | SLTS 1<br>Vishaka Vidyalaya, Colombo 07                          |
| Ms. M. L. S. L. Perera      | SLTS 11<br>Ananda Vidyalaya, Colombo 10                          |
| Mr. W. M. B. Jayasingha     | SLTS 1 (Retired)<br>Nalanda College, Colombo 10                  |
| Ms. M. E. M. Fernando       | SLTS 1 (Retired)<br>St. Joseph Vass Vidyalaya,<br>Wennappuwa.    |
| Ms. K. V. Abrew             | SLTS 1 (Retired)<br>St. Pauls Balika Vidyalaya,<br>Bambalapitiya |
| Mr. C. L. M. Navaz          | ISA, Zonal Education Office,<br>Ibbagamuwa.                      |
| Mr. M. H. M. Buhari         | ISA (Retired)<br>Zonal Education Office, Kegalle.                |

## **Editors Panel :**

Ms. M. A. Indra Pathmini Perera

Senior Lecturer  
Department of Commerce,  
National Institute of Education

Dr. H. M. L. K. Herath

Senior Lecturer  
Wayamba University of Sri Lanka.

Dr. S. Sivarajasingham

Senior Lecturer  
Department of Economics & Statistics,  
University of Peradeniya.

Mr. S. A. C. Stanley Silva

Senior Lecturer  
University of Sri Jayawardenepura.

Mr. K. A. Darmasena

Senior Lecturer  
University of Sri Jayawardenepura.

## **Instructions for Referring to the Teachers' Instructional Manual (Teachers' Guide)**

The Business Statistics syllabus for General Certificate of Education (Advanced Level) is implemented from 2017 onwards after being revised under the national curriculum revising policy which is implemented once in every eight years. The Business Statistics which was introduced as an Advanced Level subject in 1997 was subjected to the first revision on competency based in 2009 and has been listed out under 11 competencies for both grade 12 and grade 13 in this manual. A practical teaching-learning process that can be implemented in the classroom for 38 competency levels from the seventh competency in grade 13 syllabus has been proposed in this manual.

This teacher instruction manual has complied with all the competency levels prescribed for grade 13 Business Statistics syllabus of General Certificate of Education (Advanced Level).

The relevant competency, competency levels, the number of periods allocated for the competency level, the learning outcomes expected to have been achieved at the end of learning the subject matters under the competency level, are contained first, and then the proposed instructions for lesson planning, followed by a guideline to explain the subject matters are also contained in this manual in great detail. Proposed activities for assessment and evaluation are also associated with many competency levels at the end.

The classroom teacher-learning process is expected to be planned in a manner of Business Statistics will be developed parallel to the growth of attitudes, skills and practices of the students. An adequate guidance is expected to be gained for that purpose through this Teacher Instruction Manual.

Every teacher should lead the students for practical learning through planning the lessons for classroom teaching – learning process with reference to the proposed benchmarks under instructions for lesson planning and detailed facts contained in the guidelines to explain the subject matters.

Since Business Statistics is a practically important subject, the lesson plans are expected to be prepared by the teachers expanding the boundaries of the scope of their comprehension, reviewing the updated subject matters simultaneous to the prospective changes that may possibly take place in the business field.

Project Leader

---

---

# Learning Outcomes and Model Activities

---

---

**Competency 7.0 :** Uses Statistical Inference for making business decisions

**Competency Level 7.1 :** Builds up of sampling distributions for statistical inference

**No. of Periods :** 04

**Learning outcomes :**

- Builds up probability distributions for statistics.
- Introduces the probability distribution of a sample statistic as a sampling distribution.
- Distinguishes the population distribution, sample distribution and sampling distribution from one another.
- Computes the mean and variance of a sampling distribution.
- Differentiates the standard deviation and standard deviation of an estimator (standard error).

**Instructions for Lesson Planning :**

- Explain that the height of students in a class of 40 takes various values. Accordingly further explain that the height is a random variable and also that the height of those 40 students is the distribution of that variable. Point out that it is introduced as the population distribution.
- Explain that if 8 students from that class is drawn in random, the height of those eight students takes various values and further that, it is the distribution of a sample (sample distribution)
- Explain that if all the possible samples with size 8 were drawn in random and compute the mean height of all those samples are as  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots$
- Explain that the mean height of the samples is a variable and the probability distribution containing the mean of each and every sample is called the sampling distribution.

Involve the students in following activity.

- I. Consider the population 1, 3, 5 compute the mean  $\mu$  and variance  $\sigma^2$  of that population.

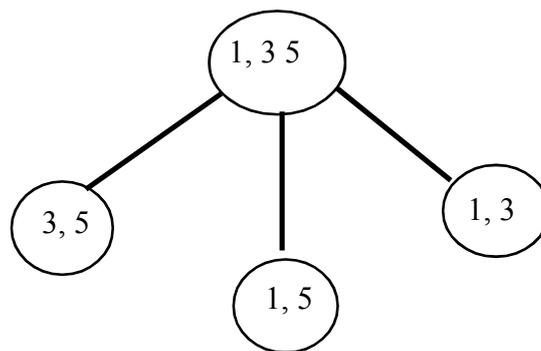
$$\mu = \frac{1+3+5}{3} = \underline{\underline{3}}$$

$$\begin{aligned}\sigma^2 &= \frac{(1-3)^2 + (3-3)^2 + (5-3)^2}{3} \\ &= \frac{4+0+4}{3} = \frac{8}{3} \\ &= \underline{\underline{2.66}}\end{aligned}$$

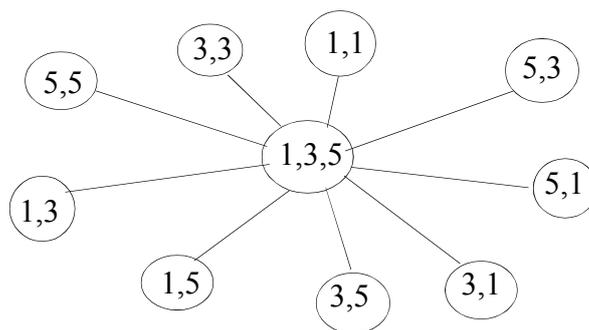
2. Draw out all the possible samples with size 2 from this population separately as.

- without replacement
- With replacement

- Samples without replacement



- Samples with replacement



3. Find the mean of each and every sample drawn with and without replacement separately and construct the probability distribution of sample means for both situations separately.

| Sample | $\bar{x}$ |
|--------|-----------|
| (3, 5) | 4         |
| (1, 3) | 2         |
| (1,5)  | 3         |

Probability distribution of the sample means without replacement.

| Sample mean $\bar{x}$ | $P(\bar{x})$  |
|-----------------------|---------------|
| 2                     | $\frac{1}{3}$ |
| 3                     | $\frac{1}{3}$ |
| 4                     | $\frac{1}{3}$ |

Probability distribution of the sample means with replacement.

| Sample | $\bar{x}$ |
|--------|-----------|
| (3, 5) | 4         |
| (1, 3) | 2         |
| (1,5)  | 3         |
| (5,3)  | 4         |
| (3, 1) | 2         |
| (5, 1) | 3         |
| (1, 1) | 1         |
| (3, 3) | 3         |
| (5, 5) | 5         |

| Sample mean $\bar{x}$ | $P(\bar{x})$  |
|-----------------------|---------------|
| 1                     | $\frac{1}{9}$ |
| 2                     | $\frac{2}{9}$ |
| 3                     | $\frac{3}{9}$ |
| 4                     | $\frac{2}{9}$ |
| 5                     | $\frac{1}{9}$ |

- Find the expected value and variance of each probability distribution you have constructed separately.

For samples without replacement.

| $\bar{x}$ | $P(f)$        | $\bar{x}.P(\bar{x})$ | $\bar{x}^2.P(\bar{x})$ |
|-----------|---------------|----------------------|------------------------|
| 2         | $\frac{1}{3}$ | $\frac{2}{3}$        | $\frac{4}{3}$          |
| 3         | $\frac{1}{3}$ | $\frac{3}{3}$        | $\frac{9}{3}$          |
| 4         | $\frac{1}{3}$ | $\frac{4}{3}$        | $\frac{16}{3}$         |

$$\begin{aligned}
 E(\bar{x}) &= \sum \bar{x}.P(\bar{x}) \\
 &= \frac{9}{3} \\
 &= \underline{\underline{3}}
 \end{aligned}$$

$$\begin{aligned}
 \text{var}(\bar{x}) &= \sum \bar{x}.P(\bar{x}) - [E(\bar{x})]^2 \\
 &= \frac{29}{3} - 3^2 \\
 &= \frac{29}{3} - 9 \\
 &= \frac{29 - 27}{3} \\
 &= \underline{\underline{\frac{2}{3}}}
 \end{aligned}$$

For the samples with replacement.

| $\bar{x}$ | $P(\bar{x})$  | $\bar{x}.P(\bar{x})$ | $\bar{x}^2.P(\bar{x})$ |
|-----------|---------------|----------------------|------------------------|
| 1         | $\frac{1}{9}$ | $\frac{1}{9}$        | $\frac{1}{9}$          |
| 2         | $\frac{2}{9}$ | $\frac{4}{9}$        | $\frac{8}{9}$          |
| 3         | $\frac{3}{9}$ | $\frac{9}{9}$        | $\frac{27}{9}$         |
| 4         | $\frac{2}{9}$ | $\frac{8}{9}$        | $\frac{32}{9}$         |
| 5         | $\frac{1}{9}$ | $\frac{5}{9}$        | $\frac{25}{9}$         |

$$\begin{aligned}
 E(\bar{x}) &= \sum \bar{x}.P(\bar{x}) \\
 &= \frac{27}{9} \\
 &= \underline{\underline{3}}
 \end{aligned}$$

$$\begin{aligned}
 \text{var}(\bar{x}) &= \sum \bar{x}^2.P(\bar{x}) - [E(\bar{x})]^2 \\
 &= \frac{93}{9} - 3^2 \\
 &= \frac{93}{9} - 9 \\
 &= \frac{93 - 81}{9} \\
 &= \underline{\underline{\frac{12}{9}}} = \underline{\underline{\frac{4}{3}}}
 \end{aligned}$$

4. Explain the relationship between the population mean and the mean of the sampling distribution of the sample means as well as the relationship between population variance and the variance of the sampling distribution of the sample means.

- $\mu = 3$                        $E(\bar{x}) = 3$

∴ The means of the sampling distribution of the sample means is equal to the population mean.

$$\sigma^2 = \frac{8}{3} \quad (\text{Population variance})$$

$$\sigma_{\bar{x}}^2 = \frac{2}{3} \quad (\text{For samples without replacement})$$

$$\sigma_{\bar{x}}^2 = \frac{4}{3} \quad (\text{For samples with replacement})$$

- Accordingly point out that the variance of sampling distribution of sample means is less than the population variance.

$$V(\bar{x}) < \sigma^2$$

- $V(\bar{x})$  for samples without replacement  $< V(\bar{x})$  for samples with replacement  $< (\sigma^2)$  the population variance.
- The variance of a sampling without replacement is less than the variance of a sampling with replacement.

$$\frac{2}{3} < \frac{4}{3}$$

- Explain that the sampling without replacement is more appropriate than the sampling with replacement since the precision of the sampling without replacement is greater.
- Explain that there is the following relationship between the population variance and the variance of sampling distribution of sample means with replacement.

- Once the population variance is divided by the sample size, the variance of the sampling distribution of the sample means of the samples with replacement is derived.

$$\frac{\sigma^2}{n} = \frac{\frac{8}{3}}{2} = \frac{8}{3} \times \frac{1}{2} = \underline{\underline{\frac{4}{3}}}$$

Once the sampling is with replacement.

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

- Once the samples without replacement have been drawn from a population, explain that there is the following relationship between the population variance and the variance of the sampling distribution of sample means ( $\bar{x}$ ).

$$\begin{aligned} \sigma_{\bar{x}}^2 &= \frac{\sigma^2}{n} \cdot \left( \frac{N-n}{N-1} \right) \\ &= \frac{8}{3} \times \frac{2}{1} \times \left( \frac{3-2}{3-1} \right) \\ &= \frac{8}{3} \times \frac{1}{2} \times \frac{1}{2} \\ &= \underline{\underline{\frac{2}{3}}} \end{aligned}$$

- Once it is without replacement  $\text{var}(\bar{X}) = \frac{\sigma^2}{n} \cdot \left( \frac{N-n}{N-1} \right)$
- State that the factor  $\left( \frac{N-n}{N-1} \right)$  is known as the Finite Population Correction Factor.
- Point out that when the population is infinite.

(when the sampling fraction  $\frac{n}{N} < 0.05$ ) and when the sampling is with replacement the correction factor is not necessary to be applied.

- Hence point out that when the sampling is without replacement the variance of sampling distribution of sample means can be derived dividing the population variance ( $\sigma^2$ ) by sample size (n) and multiplying by the correction factor  $\left(\frac{N-n}{N-1}\right)$
- Point out that the variance of the sampling distribution is equal to the population variance when the sample size is 1.

$$\frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right) = \text{var}(\bar{x})$$

$$\frac{\sigma^2}{1} \times \left( \frac{N-1}{N-1} \right) = \text{var}(\bar{x})$$

$$\therefore \underline{\underline{\sigma^2 = \text{var}(\bar{x})}}$$

- Compute the population standard deviation and the standard deviation of the sampling distribution with and without replacement separately.
- Population standard deviation.

$$\begin{aligned} (\sigma) &= \sqrt{\sigma^2} \\ &= \frac{8}{3} \\ &= \underline{\underline{1.63}} \end{aligned}$$

- Standard deviation of the sampling distribution when the sampling is without replacement.

- Standard deviation of the sampling distribution when the sampling is with replacement.

$$\begin{aligned} \sigma_{\bar{x}} &= \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \\ &= \sqrt{\frac{8}{3}} \\ &= \underline{\underline{1.15}} \end{aligned}$$

$$\begin{aligned} \sigma_{\bar{x}} &= \sqrt{\frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)} \\ &= \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} \\ &= \sqrt{\frac{8}{3}} \times \left( \frac{3-2}{3-1} \right) \\ &= \underline{\underline{0.82}} \end{aligned}$$

- Inform the students that the standard deviation of the sample distribution is known as the standard error.
- Ascertain that the populations can be very large in practice and once the samples with size 'n' are drawn, the mean, variance and the standard deviation of those sampling distributions can be derived in the same procedure.

**A guideline to explain the subject matters :**

Sample mean ( $\bar{x}$ ) differs from sample to sample. Accordingly,  $\bar{x}$  can be considered as a random variable. Since a probability distribution is always connected with a random variable  $\bar{x}$  also has a probability distribution. This is known as the sampling distribution of  $\bar{x}$ , the sample mean.

If it is mentioned in other words the probability distribution of a sample statistics is known as a sampling distribution.

- Hence the probability distributions of statistics, such as sample means  $\bar{x}$  sample variance  $s^2$  sample standard deviation S and sample proportion (p) are considered as sampling distributions.
- The probability distribution of all the possible values taken by a particular characteristic in connection with the considered population related to the study is known as the population distribution.
- Once a sample drawn from the population related to the study, the probability distribution of various values taken by a considered characteristic studied in that sample in known as a sample distribution.
- Once all the possible same size samples are drawn from a population in random the probability distribution of the values taken by a considered statistic is known as the sampling distribution of that sample statistic.
- The mean (expected value) of the sampling distribution of a particular sample statistic is equal to the relevant population parameter of that statistic.

Example :

| Parameter | Statistic      | Expected value (mean)                           |
|-----------|----------------|---|
| $\mu$     | $\bar{x}$      | $E(\bar{x}) = \mu_{\bar{x}} = \mu$              |
| $\pi$     | $P$            | $E(P) = \mu_P = \pi$                            |
| $\theta$  | $\hat{\theta}$ | $E(\hat{\theta}) = \mu_{\hat{\theta}} = \theta$ |

- The variance of the sampling distribution of a particular statistic is always less than the population variance relevant to that statistic.



$$E(\hat{\theta}) = \theta \qquad \text{var}(\hat{\theta}) < \text{population variance}$$

Example : variance of the sampling distribution.

Since  $\text{var}(\bar{x}) = \frac{\sigma^2}{n}$  or  $\text{var}(\bar{x}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$  that is less than the population variance.

- **Standard deviation**

1. In order to measure the variation of the values of a particular characteristic related to a population or sample the standard deviation is used. Computing the deviations of the variable from its mean, deriving the average value of the squares of those deviations the variance is found and the positive square root of it is the standard deviation.

If the mean of the data set  $x_1, x_2, \dots, x_n$

$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

$$\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

- Population standard deviation is denoted by  $\sigma$  and sample standard deviation by  $S$ .
- Standard error (standard deviation of a statistic)

In order to measure the variation of a particular statistic the standard error is used. The deviation between the population parameter and the values of the relevant estimator is measured using the standard error.

If  $\theta$  is the parameter  $\hat{\theta}$  is the relevant statistic and  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K$  are the possible values taken by  $\hat{\theta}$

$$\text{var } \hat{\theta} = \frac{(\hat{\theta}_1 - \theta)^2 + (\hat{\theta}_2 - \theta)^2 + \dots + (\hat{\theta}_K - \theta)^2}{K}$$

$$\text{Var}(\hat{\theta}) = \frac{\sum_{i=1}^K (\hat{\theta}_i - \theta)^2}{K}$$

$$\text{Standard error} = \sqrt{\frac{\sum_{i=1}^K (\hat{\theta}_i - \theta)^2}{K}}$$

- The standard error of the sampling distribution of a sample statistic is less than the relevant population standard deviation.

Ex: Population standard deviation =  $\sigma$

$$\text{Standard error of } \bar{x} \rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

n – sample size

N population size

**Competency 07 :** Uses Statistical Inference for making business decisions

**Competency Level 7.2 :** Builds the sampling distribution of sample mean for statistical inference.

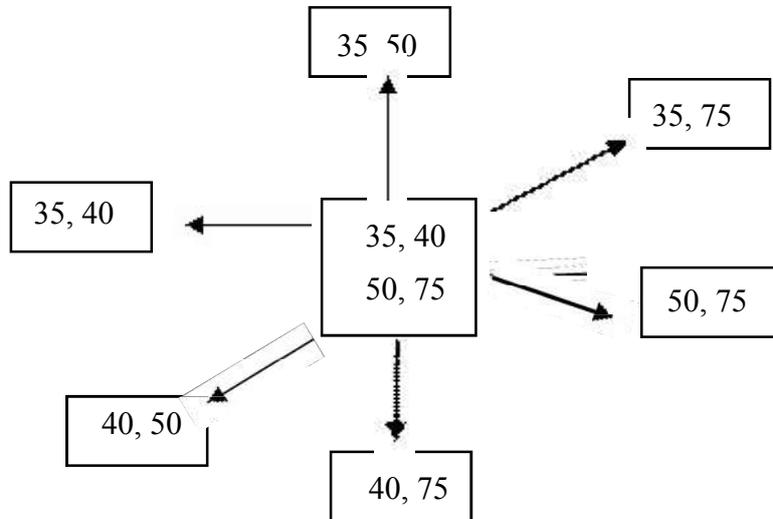
**No. of Periods :** 18

**Learning outcomes:**

- Interprets the sampling distribution of sample means ( $\bar{X}$ )
- Builds the sampling distribution of sample means for large size samples when the population distributes normally with known variance.
- Builds the sampling distribution of sample means for large size samples when the population distributes normally with unknown variance.
- Builds the sampling distribution of sample means for small size samples when the population distributes normally with known variance.
- Builds the sampling distribution of sample means for small size samples when the population distributes normally with unknown variance.
- States the Central Limit Theorem and explains its uses.
- Builds the sampling distribution of sample means for large size samples when the population does not distributes normally with known variance.
- Builds the sampling distribution of sample means for large size samples when the population does not distribute normally with unknown variance.
- Solves problems related to the sampling distribution of sample means ( $\bar{X}$ ).
- Explains that sampling distributions cannot be declared when small size sample are drawn from non-normal populations.

**Instructions for lesson Planning**

- Present the following diagram to the students and discuss the questions associated.



- Marks of four students in an examination are given above. Find the mean of those marks.

$$= \frac{35 + 40 + 50 + 75}{4}$$

$$= \underline{\underline{50}}$$

- How many samples with size 2 can be drawn from this population? 6
- If sample elements are drawn on simple random sampling method, what are the elements that can be contained in each sample?

(35, 40)      (40, 50)  
 (35, 50)      (40, 75)  
 (35, 75)      (50, 75)

- What are the sample means that can be derived from each sample?

$$= 37.5, 42.5, 55, 45, 57.5, 62.5$$

- Ascertain that the values derived for mean ( $\bar{X}$ ) of each sample are different from each other and the probability distribution of  $\bar{X}$  is defined as the sampling distribution of sample means.
- Compute the expected value and variance of the probability distribution of ( $\bar{X}$ ).

| $\bar{x}$ | $P(\bar{x})$  | $\bar{x} \cdot P(\bar{x})$ | $\bar{x}^2 \cdot P(\bar{x})$ |
|-----------|---------------|----------------------------|------------------------------|
| 37.5      | $\frac{1}{6}$ | $\frac{37.5}{6}$           | $\frac{1406.25}{6}$          |
| 42.5      | $\frac{1}{6}$ | $\frac{42.5}{6}$           | $\frac{1806.25}{6}$          |
| 45        | $\frac{1}{6}$ | $\frac{45}{6}$             | $\frac{2025}{6}$             |
| 55        | $\frac{1}{6}$ | $\frac{55}{6}$             | $\frac{3025}{6}$             |
| 57.5      | $\frac{1}{6}$ | $\frac{57.5}{6}$           | $\frac{3306.25}{6}$          |
| 62.5      | $\frac{1}{6}$ | $\frac{62.5}{6}$           | $\frac{3906.25}{6}$          |

$$E(\bar{x}) = \frac{300.0}{6}$$

$$= \underline{\underline{50}}$$

$$\text{var}(\bar{x}) = \sum \bar{x}^2 \cdot P(\bar{x}) - [E(\bar{x})]^2$$

$$= 2579.17 - 2500$$

$$= \underline{\underline{79.17}}$$

- Compute the variance of the population

$$\sigma^2 = \frac{(35-50)^2 + (40-50)^2 + (50-50)^2 + (75-50)^2}{4}$$

$$\sigma^2 = \underline{\underline{237.5}}$$

- Point out that the variance of sampling distribution of sample means also can be derived using population variance.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)$$

$$= \frac{237.5}{2} \left( \frac{4-2}{4-1} \right)$$

$$= \underline{\underline{79.17}}$$

- Explain to the students that whether the population variance is known or unknown, whether the population is normal or not and whether the sample size is large or small they are considered as basis in the construction of the sampling distribution of sample means.
- Consider the following example for building the sampling distribution of the sample means.
- The salaries of the employees in a firm is normally distributed as the mean salary Rs. 25000 and the variance as Rs. 1000. A sample of 100 employees is taken into consideration.
- Discuss the above instance with the students and collect the following details.
- Has the population distributed normally? Yes.
- Is the sample size greater than 30? Yes.
- Point out to the students that the sampling distribution of sample means falls as follows: when the size is large of a sample drawn from a normally distributed population whose variance is known.

$$\bar{x} \sim N \left( \mu, \frac{\sigma^2}{n} \right)$$

$$\bar{x} \sim N \left( 25000, \frac{1000}{100} \right)$$

- Let the students to consider the following instance and hold a discussion on it.
- The salaries of employees in a firm is normally distributed as mean salary of Rs. 25 000 and variance as Rs. 1 000. A sample of 25 elements is taken into consideration.
- Collect the following details holding a discussion with the students in connection with the above mentioned instance.
  - Is the population variance known? yes
  - Has the population distributed normally? yes.
  - Is the sample size greater than 30? no.'
- Point out to the students that the sampling distribution of sample means is as follows, when a small size sample is drawn from a normally distributed population with known variance.

$$\bar{X} \sim N \left( \mu, \frac{\sigma^2}{n} \right)$$

$$\bar{X} \sim N \left( 25000, \frac{1000}{25} \right)$$

- Hence explain that the sampling distribution of sample means is built in the same way, whether the sample size is large or small, when the population distributes normally with known variance.
- Let the students to consider the following instance and hold a discussion with the students.
- The salaries of employees in a firm is normally distributed with mean salary of Rs. 25 000. Once a sample of 100 employees is drawn the standard deviation was Rs. 30.
- Collect the following details discussing the above instance with the students.
  - Is the population variance known? No.
  - Has the population distributed normally? yes
  - Is the sample size large? yes
- Point out to students that the sample variance can be applied for unknown population variance, accordingly that the sampling distribution of sample means is as follows when the size of a sample drawn from a normal population with unknown variance is large.

$$\bar{x} \sim N \left( \mu, \frac{\sigma^2}{n} \right) \qquad \bar{x} \sim N \left( \mu, \frac{S^2}{n} \right)$$

$$\bar{x} \sim N \left( 25000, \frac{30 \times 30}{100} \right)$$

- Hold a discussion with the students to explain the 't' distribution.
  - Explain that when samples of 40 in each are drawn from a population of 50 students who sat for an exam, and once the mean of those samples is computed they can be assumed to be dispersed closer to the population mean.
  - Point out the fact that, if samples of size 5 from 50 students are drawn and computed the mean, those means can be most probably dispersed in a vast range away from the population mean, since there is a possibility of containing the samples with the highest marks as well as the lowest.

- Explain that a symmetrical distribution consisting of a greater dispersion relative to the normal distribution is known as a t-distribution.

- Provide with the following instance to the students and hold a discussion.

Salaries of a business firm is normally distributed as the mean salary is Rs. 25 000/-. Once a sample of 25 employees was taken the variance was computed as Rs. 900.

- Discuss the above mentioned instance with the students and collect the following details.

- Is the population variance known? No
- Is the population distributed normally? Yes
- Is the sample size large? No

- Explain that the sample variance is used since the population variance is unknown and the sampling distribution of  $\bar{X}$  falls in a t-distribution since the sample size is small ( $n < 30$ )

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

- Group the students in the class appropriately and involve them in the following activity.

1. The salaries of a firm is distributed normally with the mean Rs. 50 000 and variance Rs. 2 000. A sample of 50 employees is considered.
2. The weight of the packets of sugar, packeted in a machine has distributed normally with the mean 500g and the variance 50g. A sample of 10 packets is drawn in random.
3. The marks of an examination has distributed normally with the mean 52. Once a sample of 50 students is drawn in random the standard deviation has been computed as 8.
4. The life time of electric appliances is distributed normally as mean is 1 500 hours. The variance in a sample of 10 units has been found as 400 hours.
  - (i) Construct the distribution of population
  - (ii) Construct the sampling distribution of the sample means

#### Solutions for activity-I

1. When considered the salaries of the employees as X.

Population distribution

$$X \sim N (50000, 2000)$$

Sampling distribution of the sample means

$$\bar{X} \sim N\left(50000, \frac{2000}{50}\right)$$

2. When considered the weight of the packets of sugar as X

- Population distribution  $X \sim N(500, 50)$
- Sampling distribution of the sample means

$$\bar{X} \sim N\left(500, \frac{50}{10}\right)$$

3. When considered the marks of the children as Y.

population distribution  $Y \sim N(52, \sigma^2)$

- Population variance is unknown
- Sampling distribution of the sample means

$$\bar{Y} \sim N\left(52, \frac{64}{50}\right)$$

- Sample variance is submitted for population variance ( $\sigma^2$ )

4. When considered the life time of electric appliances as Y.

Population distribution  $Y \sim N(1500, \sigma^2)$

Since the population variance is unknown, sample variance is submitted instead and since the sample size is less than 30 the sampling distribution of  $\bar{Y}$  falls in a 't' distribution.

Then the sampling distribution of sample means

$$\bar{Y} \sim t\left(1500, \frac{400}{10}\right)$$

Activity–02.

- Present the following problems to the students and let them answer first and then hold a discussion.
  1. Name non normal population distributions known to you.
  2. State the parameters in such distributions

3. State the mean and variance of those distributions.
4. Explain how the shape of sampling distribution of sample means can be if the samples are drawn from non normal populations.

Answers for Activity–02 :

- 1 • Binomial distribution                      • Poisson distribution
- 2 • Parameters of a binomial distribution are **n** and **P**

$$X \sim B(n, p)$$

- The parameter of a poisson distribution is  $\lambda$

$$X \sim P(\lambda)$$

3. Of a binomial distribution

Mean             $E(x) = np$

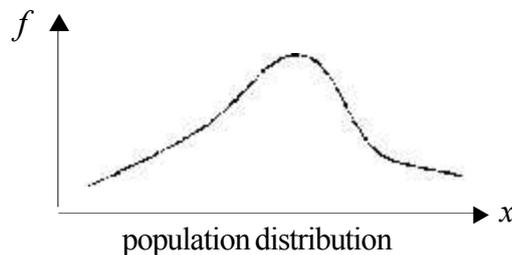
Variance         $Var(x) = npq$

Of a poisson distribution

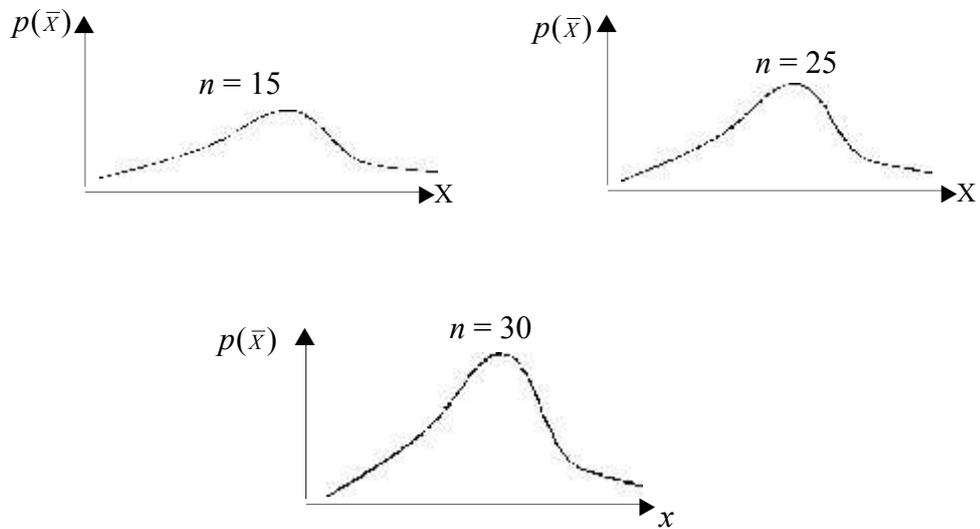
Mean             $E(x) = \lambda$

Variance         $Var(x) = \lambda$

5. Whatever the shape of the population distribution seems to be, if the sample size is large, the sampling distribution of sample means takes a symmetrical shape.



- Ascertain the fact “that the sampling distribution of the sample means  $\bar{X}$  approximately falls in a normal distribution, when the sample size is sufficiently large ( $n \geq 30$ ), whatever the shape of the population distribution seems to be” is stated by the Central Limited Theorem.



sample distributions of sample means

- According to the Central Limited Theorem the sampling distribution of sample means approximately falls normally as follows.

× population characteristic

Distribution of X is unknown ×  ?

If  $n \geq 30$  the sampling distribution of  $\bar{X}$

$$\bar{X} \text{  N \left( \mu, \frac{\sigma^2}{n} \right)}$$

- Explain that once the variance of a non normal population is unknown and when the sample size is sufficiently large

$$\bar{X} \text{  N \left( \mu, \frac{S^2}{n} \right)}$$

- Accordingly the uses of the Central Limit Theorem can be mentioned as follows.
  - Even knowing nothing about the shape of the population, ability of solving the problems considering that the sampling distribution of  $\bar{X}$  approximately falls normally.

- Ascertain the fact that the sampling distribution of the sample means can not be constructed, if the population is non normal and the sample size is small, whether the population variance is known or not.
- Group the students in the class in an appropriate manner and involve them in the following activity.
  1. Marks of the last term test for Statistics in a class has distributed as the mean 60 and variance – 10
    - Random sample of 35 students is considered.
    - A random sample of 15 students is considered.
  2. The monthly expenditure of the families with 4 members living in a particular area distributes with mean Rs. 50 000
    - Once a random sample of 50 families is considered the variance of it has been computed as Rs. 10 000
    - Once a random sample of 20 families is considered the variance has been found as Rs. 15 000
      - a. Construct the population distribution
      - b. Construct the sampling distribution of sample means

**Answers :**

1. When considered the marks as X
  - (a) Population distribution  
The distribution of X is unknown
  - (b) (i) Since  $n > 30$  sampling distribution of  $\bar{X}$  falls as follows.

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\bar{X} \approx N\left(60, \frac{10}{35}\right)$$

- (ii) Since  $n < 30$  and the shape of the population distribution is unknown the sampling distribution of sample means can not be built.
2. When considered the monthly expenses as Y.
  - (a) Population distribution

The distribution of Y is unknown.

- (b) (i) Eventhough the population variance is unknown, the sample variance can be used instead, so that the sampling distribution of sample means in accordance with the Central Limit Theorem is as follows, though the population distribution is unknown.

Distribution of Y is unknown

$$\bar{y} \approx \left( \mu, \frac{S^2}{n} \right)$$

$$\bar{y} \approx \left( 50,000, \frac{10,000}{50} \right)$$

- (ii) Since  $n < 30$  and shape of the population distribution is unknown the sampling distribution of sample means can not be built.

**A Guideline to explaining the subject matters :**

- Once all the possible random samples with the same size are drawn and computed the means of all those samples, the probability distribution of those sample means is known as the sampling distribution of the sample means.
- Once the samle means is denoted by  $\bar{X}$  the mean of the sampling distribution of  $\bar{X}$  is denoted by  $\mu_{\bar{x}}$  Or  $E(\bar{X})$
- The variance of the sampling distribution of sample means  $\bar{X}$  is denoted by  $\sigma_{\bar{x}}^2$  or  $\text{var}(\bar{X})$ .
- The mean of the sampling distribution of the sample means is equal to the population mean  $\mu_{\bar{x}} = \mu$
- The variance of the sampling distribution of sample means can be computed as follows.

$$\sigma_{\bar{x}}^2 = \text{var}(\bar{X}) = \sum \bar{X}^2 \cdot (\bar{X}) - [E(\bar{X})]^2$$

- The variance of the sampling distribution of sample means can be computed using the population variance as follows.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad \text{or} \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \left[ \frac{N-n}{N-1} \right]$$

- $\frac{N-n}{N-1}$  is the finite population correction factor and used in the following instances.

- Once the population is finite and / or the sampling is without replacing.
- But, if the sampling fraction  $\left(\frac{n}{N}\right)$  is less than 0.05 the correction factor is ignored.
- Once the variance of a normal distribution  $(\sigma^2)$  is known the sampling distribution of  $\bar{X}$  falls normally as follows regardless of the sample size.

$$\bar{X} \sim N \left( \mu, \frac{\sigma^2}{n} \right) \quad \text{When the sample size } (n \geq 30)$$

$$\bar{X} \sim N \left( \mu, \frac{\sigma^2}{n} \right) \quad \text{When the sample size } (n < 30)$$

- Once the variance of a normal population is unknown, if the sample size is sufficiently large  $(n \geq 30)$  the sampling distribution of the sample means falls normally as follows.

$$X \sim N (\mu, \sigma^2) \qquad \bar{X} \sim N \left( \mu, \frac{S^2}{n} \right)$$

- Once the variance of a normal distribution is unknown, when the sample size is small  $(n < 30)$  the sampling distribution of sample means falls in a t-distribution whose degree of freedom is  $n-1$

$$\bar{X} \sim t_{n-1} \qquad t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

- Unless the sample size is sufficiently large  $(n < 30)$  standard deviation of the sampling distribution of sample mean  $(\sigma)$  should be estimated using the sample standard deviation and then the sampling distribution  $\bar{X}$  does not further fall normally. At such situations the sampling distribution of  $\bar{X}$  falls in a t-distribution.
- t-distribution consists of following properties.
  - It is also symmetrical distribution as well as the normal distribution.
  - Since the sample standard deviation (S) is applied for population standard deviation  $(\sigma)$  and also since the sample size considered is small the dispersion of t-distribution is greater than that of the normal distribution.
- Expansion of t - distribution can be explained as follows.



- Since the t-distribution depends on the sample size (n) the relevant areas have been tabulated according to (n-1) number of degrees of freedom.
  - t-distribution is applied in following instances.
    - Population should distribute normally.
    - Population variance should be unknown.
    - Sample size should not be sufficiently large ( $n < 30$ ) and the sample variance is applied for population variance.
- The sampling distribution of the sample means associated with a non normal population is approximated to a normal distribution when the sample size  $n \geq 30$  is in accordance with the Central Limit Theorem.
- The Central Limit Theorem means that,
 

“once the size of a sample drawn from any population with mean  $\mu$  and variance  $\sigma^2$  is sufficiently large ( $n \geq 30$ ) the sampling distribution of sample means approximately falls in a normal distribution with mean  $\mu$  and variance  $\frac{\sigma^2}{n}$ ”
- Accordingly when the variance of a non normal population is known and the sample size is sufficiently large the sampling distribution of sample means can be stated as follows.
 
$$\bar{X} \sim N \left( \mu, \frac{\sigma^2}{n} \right)$$
- When the variance of a non normal population is unknown and the sample size is sufficiently large ( $n \geq 30$ ) the sampling distribution of the sample means can be stated as follows.
 
$$\bar{X} \sim N \left( \mu, \frac{S^2}{n} \right)$$
- When the sample size of a non normal population is not sufficiently large ( $n < 30$ ) the Central Limit Theorem can not be applied. Such problems can not be solved using parametric statistical techniques.

### Assessment & Evaluation :

#### Activity 01.

- Employee salaries of a firm distributes normally with mean Rs. 25 000 and standard deviation Rs. 6 000. If a sample of 9 employees is drawn in random.

- (i) State the mean and variance of the sampling distribution of sample means. Hence write the sampling distribution of sample means.
- What is the probability that an employee in the sample getting a salary of Rs. 30 000 or more.
- If the probability of getting a salary less than or equal to Rs. 20 000 to be maintained at 10% level, what should be the sample size?

#### Activity – 2

When considered an assembling process of packaging sugar the average mass of a packet of sugar was 500g. Once a sample of 49 packets was checked the standard deviation of the mass was 10g (the mass of the packets distributes normally)

- (i) Write the mean and variance of the sampling distribution of sample means.
- (ii) State the sampling distribution of sample means
- (iii) Find the probability of the sample mean being less than 499g
- (iv) Find the probability of the sample mean being in the range 495g-510g
- (v) Suppose that the sample size has been increased upto 100
  - (a) Find the probability of the sample mean being less than 499g
  - (b) Find the probability of the sample mean falls in the range 495g to 510g
- (vi) Compare the results of (iii) and (iv) above with the results of (v)
- (vii) Describe the difference of the probabilities observed once the sample size was increased from 49 to 100.

#### Activity-3

The mass of a pack of vegetables processed by a company involved in exporting the fruit distributes normally with the mean 5Kg. Once a sample of 10 packes was drawn in random and checked the standard deviation was 0.75 Kg.

- (i) Find the mean and the variance of the sampling distribution of sample means
- (ii) Write the sampling distribution of sample means

#### Activity-4

The average number of customers coming to a bank in an interval of 30 minutes time in a working day is 4. When considered 36 such intervals of 30 minutes.

- (i) Explain the probability distribution of the population
- (ii) Derive the mean and variance of the sampling distribution of sample means

- (iii) Construct the sampling distribution of the sample means.
- (iv) Find the probability that less than two customers coming in 36 intervals of 30 minutes in each.
- (v) Find the probability that 10 customers or more will be coming in 100 time intervals of 1 hour.
- (vi) State the assumptions made in solving this problem.
- (vii) When considered 10 time intervals of 30 minutes can the sampling distribution of sample means be built?

Appropriately use the following formulae in solving these problems.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \qquad Z = \frac{\bar{X} - \mu}{S / \sqrt{n}} \qquad t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

Answer for Activity – 1

(i)  $\mu_{\bar{X}} = 25000$   $\sigma_{\bar{X}}^2 = \frac{6000 \times 6000}{3 \times 3}$   
 $= 2000^2$

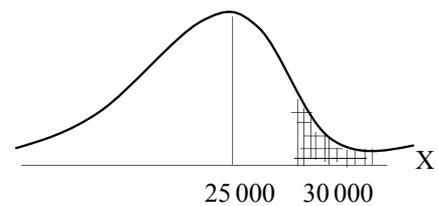
(ii)  $\bar{X} \sim N(25\,000, 2\,000^2)$

(iii) 
$$P(\bar{X} \geq 30000) = \Pr\left(Z \geq \frac{30000 - 25000}{2000}\right)$$

$$= \Pr(Z \geq 2.5)$$

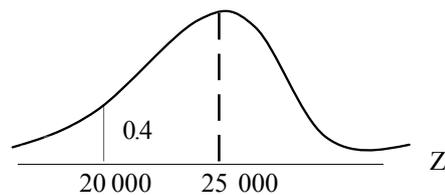
$$= \underline{\underline{0.5000 - 0.4938}}$$

$$= \underline{\underline{0.0062}}$$



(iv)  $\Pr(\bar{X} \leq 20000) = 0.10$   
 $\therefore \Pr(Z \leq -1.28) = 0.10$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$



$$-1.28 = \frac{20000 - 25000}{6000/\sqrt{n}}$$

$$-1.28 = \frac{-5000 \sqrt{n}}{6000}$$

$$-1.28 \times 6 = -5 \sqrt{n}$$

$$\left(\frac{1.28 \times 6}{5}\right) = n$$

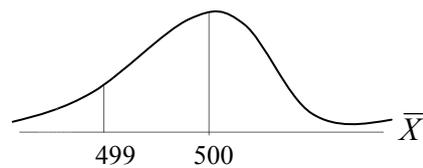
$$\underline{\underline{\therefore n \geq 3}}$$

Answer for Activity-2

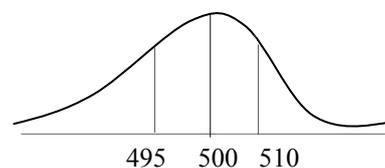
$$(i) \quad \mu_{\bar{X}} = 500 \quad \sigma_{\bar{X}}^2 = \frac{10 \times 10}{49} = 2.04$$

$$(ii) \quad \bar{X} \sim N(500, 2.04)$$

$$(iii) \quad p(\bar{X} < 499) = p\left(Z < \frac{499 - 500}{\sqrt{2.04}}\right) \\ = p(Z < -0.69) \\ = 0.5000 - 0.2549 \\ = 0.2451$$

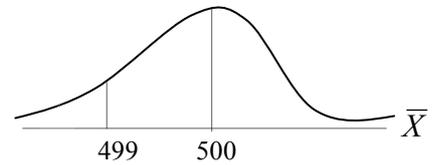


$$(iv) \quad p(495 \leq \bar{X} \leq 510) = p\left(\frac{495 - 500}{\sqrt{2.04}} \leq Z \leq \frac{510 - 500}{\sqrt{2.04}}\right) \\ = p(-3.49 \leq Z \leq 6.99) \\ = 0.4998 + 0.5000 = 9998$$

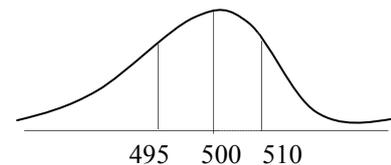


- (v) (a) When the sample size is increased from 49 to 100

$$\begin{aligned} \Pr(\bar{X} < 499) &= \Pr\left(Z < \frac{499 - 500}{10/10}\right) \\ &= \Pr(Z < -1) \\ &= 0.5000 - 0.3413 \\ &= \underline{\underline{0.1587}} \end{aligned}$$



$$\begin{aligned} \text{(b) } \Pr(495 < \bar{X} \leq 510) &= \Pr\left(\frac{495 - 500}{1} \leq Z \leq \frac{510 - 500}{1}\right) \\ &= 0.5000 + 0.5000 \\ &= \underline{\underline{1.0}} \end{aligned}$$



- Accordingly when the sample size is increased from 49 to 100 the probability of the sample mean being less than 499g has decreased from 0.2491 to 0.1587.
- When the sample size is increased from 49 to 100 the probability of sample mean falling between 495g and 510g has increased from 0.9998 to 1.000.
- Hence, when the sample size is increased under the given variance (standard Deviation) the probability of the sample statistic being apart from the parameter decreases and the probability that the sample statistic being centered about the parameter increases.

### Answer for Activity 3

- (i) If the mean of the mass is  $\bar{X}$

$$\mu_{\bar{X}} = 5 \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

Since  $\sigma^2$  is unknown the sample variance  $S^2$  is applied.  $\therefore \sigma_{\bar{X}}^2 = \frac{0.75^2}{10}$

Since the sample size is less than 30 the sample distribution of sample means falls in a t-distribution.

$$\text{(ii) } \bar{X} \sim t\left(5, \frac{0.75^2}{10}\right)$$

Answer for Activity 4 :

If the average number of customers coming in 30 minutes is X

$$(i) \quad X \sim P(\lambda = 4)$$

Since the number of intervals is 36 ( $n > 30$ ) the sampling distribution approximately falls normally in accordance with the Central Limit Theorem.

$$(ii) \quad \mu_{\bar{x}} = 4 \qquad (\sigma_{\bar{x}}^2) = \frac{4}{36}$$

$$(ii) \quad \bar{X} \sim N\left(4, \frac{4}{36}\right)$$

$$\begin{aligned} (iii) \quad \Pr(\bar{X} \leq 2) &= \Pr\left(Z \leq \frac{2-4}{2/6}\right) \\ &= \Pr(Z \leq -6) \\ &= 0.5000 - 0.5000 \\ &= \underline{0} \end{aligned}$$

(iv) Relevant to one hour interval since  $\lambda = 8$  and  $n = 100$

$$\bar{X} \sim N\left(8, \frac{8}{100}\right)$$

$$\begin{aligned} \Pr(\bar{X} \geq 10) &= \Pr\left(Z \geq \frac{10-8}{\sqrt{8/100}}\right) \\ &= \Pr(Z \geq 2.25) \\ &= 0.5 - 0.4878 \\ &= \underline{\underline{0.0122}} \end{aligned}$$

(vi) Population distribution is unknown and since sample size is greater than 30, the sampling distribution is assumed to be approximately fall normally. This is defined as the Central Limit Theorem.

(vii) The Central Limit Theorem is inapplicable, if the sample size is less than 30, when the population distribution is unknown or not normally distributed. Therefore when 10 intervals of 30 minutes are considered the sampling distribution cannot be built

**Competency 07** : Uses Statistical Inference for making business decisions

**Competency Level 7.3** : Builds up of sampling distributions of the difference between two sample means for statistical inference

**No. of Periods** : 04

**Learning outcomes :**

- Explains what the sampling distribution of the difference between two sample means is.
- Builds up the sampling distribution of the difference between two sample means.
- Introduces the mean and variance of the sampling distribution of the difference between two sample means.
- Describes the sampling distribution of the difference between two sample means using large size samples when the populations are normal and the variances are known.
- Describes the sampling distribution of the difference between two sample means using small size samples when the populations are normal and the variances are known.
- Describes the sampling distribution of the difference between two sample means with large size samples when the populations are normal and the variances are unknown.
- Describes the sampling distribution of the difference between two sample means with small size samples when the populations are normal and the variances are unknown.
- Describes the sampling distribution of the difference between two sample means with large size samples when the populations are not normal and the variances are known.
- Describes the sampling distribution of the difference between two sample means for large size samples when the populations are not normal and the variances are unknown.
- Solves the problems using the sampling distributions of the difference between two sample means.
- Makes decisions using the sampling distributions of the difference between two sample means.

**Instructions for Lesson Planning :**

- Display the following two populations on the board.

Population – A

4. 8. 12

Population – B

2. 4. 6

- Inquire the following facts from the students.
  - How many samples with size two can be drawn from each population without replacement?
  - How many samples with size two can be drawn from each population with replacement?

Activity-01

- Let the students copy the following incomplete table and complete it step by step following the instructions given.

| Sample No. | Samples of A | $\bar{X}_A$ | Samples B | $\bar{X}_B$ |
|------------|--------------|-------------|-----------|-------------|
| 1          | .....        | .....       | .....     | .....       |
| 2          | .....        | .....       | .....     | .....       |
| 3          | .....        | .....       | .....     | .....       |

Give the following instructions to complete the table.

- Draw the first sample with size 2 from the population A, find the mean and note it down in the table.
- Draw the first sample with size 2 from population B, find the mean and note it down in the table.
- Proceed the same way and note down the means of the 2<sup>nd</sup> and 3<sup>rd</sup> samples as well in the table.

Solution for Activity-01

| Sample number | Samples of A | $\bar{X}_A$ | Samples of B | $\bar{X}_B$ |
|---------------|--------------|-------------|--------------|-------------|
| 1             | 4, 8         | 6           | 2, 4         | 3           |
| 2             | 4, 12        | 8           | 2, 6         | 4           |
| 3             | 8, 12        | 10          | 4, 6         | 5           |

Activity-2

- Lead the students to compute the following table using the details in Activity-1.

| No | $\bar{X}_A - \bar{X}_B$ |
|----|-------------------------|
| 1  |                         |
| 2  |                         |
| .  |                         |
| .  |                         |
| .  |                         |
| .  |                         |

- Provide with the following instructions to complete the table.
  - Subtract the mean of the first sample taken from B population from the mean of the first sample taken from A. Population and state the answer in the table against No. 1 in  $(\bar{X}_A) - (\bar{X}_B)$  column.
  - Subtract the mean of the second sample taken form B population from the mean of the first sample taken from A population and state the answer in the table against No. 2 in  $(\bar{X}_A) - (\bar{X}_B)$  column.
  - Subtract the mean of the third sample taken from B population from the mean of the first sample taken from A population and state the answer in the table against No. 3 in the  $(\bar{X}_A) - (\bar{X}_B)$  column.
  - In the same way subtract the means of the three samples drawn from B populations, form the mean of the second sample taken from A population and state the corresponding answers in the table against No. 4, 5 and 6 in  $(\bar{X}_A) - (\bar{X}_B)$  Column.

Following the same procedure subtract the means of the three samples drawn from B population from the mean of the third sample taken from A population and state the answer in the table against No. 7, 8 and 9 respectively in  $(\bar{X}_A) - (\bar{X}_B)$  column.

#### Solution – Activity-2

| No | $(\bar{X}_A) - (\bar{X}_B)$ |
|----|-----------------------------|
| 1  | 3                           |
| 2  | 2                           |
| 3  | 1                           |
| 4  | 5                           |
| 5  | 4                           |
| 6  | 3                           |
| 7  | 7                           |
| 8  | 6                           |
| 9  | 5                           |

### Activity-3

Lead the students to tabulate the mean differences derived in the above activity, associated with the corresponding probabilities.

### Solution – Activity – 3

|                               |               |               |               |               |               |               |               |
|-------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $(\bar{X}_A) - (\bar{X}_B)$   | 1             | 2             | 3             | 4             | 5             | 6             | 7             |
| P $(\bar{X}_A) - (\bar{X}_B)$ | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{2}{9}$ | $\frac{1}{9}$ | $\frac{2}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ |

- Explain the students that the sampling distribution of the difference between two sample means is derived when the values of the mean differences associated with corresponding probabilities are tabulated in this manner.

### Activity – 04

- After assuring that the distribution built by the students is the sampling distribution of the difference between two sample means involve the students in the activity providing that with the following guidelines step by step.
  - Find the mean of A population  $\mu_A$
  - Find the mean of B population  $\mu_B$
  - Find the difference between two population means  $\mu_A - \mu_B$
- Following the way that the expected value of a probability distribution is computed as  $E(x) = \sum x.p(x)$  compute the expected value (mean) of the sampling distribution of difference between two sample means  $\mu_{\bar{x}_A - \bar{x}_B}$
- State the relationship between the mean of the sampling distribution of two sample means  $\mu_{\bar{x}_A - \bar{x}_B}$  and the difference between the means of two relevant populations  $(\mu_A - \mu_B)$
- Compute the variance of A population  $\sigma_A^2$
- Compute the variance of B population  $\sigma_B^2$
- In the way that the variance of a probability distribution is calculated, find the variance of the sampling distribution of the difference between two sample means  $\sigma_{\bar{x}_1 - \bar{x}_2}^2$
- Come out with your points of view having compared the variance of the sampling distribution of the difference between two sample means  $\sigma_{\bar{x}_1 - \bar{x}_2}^2$  and the two population variances.

Activity 04 – Solution:

$$\begin{aligned}\mu_A &= \frac{4+8+12}{3} \\ &= \underline{\underline{8}}\end{aligned}$$

$$\begin{aligned}\mu_A - \mu_B &= 8 - 4 \\ &= \underline{\underline{4}}\end{aligned}$$

$$\begin{aligned}\mu_B &= \frac{2+4+6}{3} \\ &= \underline{\underline{4}}\end{aligned}$$

- To compute

$$\mu_{\bar{X}_A - \bar{X}_B}$$

|  |               |               |               |               |                |               |               |
|--|---------------|---------------|---------------|---------------|----------------|---------------|---------------|
| $(\bar{X}_A - \bar{X}_B)$                                | 1             | 2             | 3             | 4             | 5              | 6             | 7             |
| $P(\bar{X}_A - \bar{X}_B)$                               | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{2}{9}$ | $\frac{1}{9}$ | $\frac{2}{9}$  | $\frac{1}{9}$ | $\frac{1}{9}$ |
| $(\bar{X}_A - \bar{X}_B) \cdot P(\bar{X}_A - \bar{X}_B)$ | $\frac{1}{9}$ | $\frac{2}{9}$ | $\frac{6}{9}$ | $\frac{4}{9}$ | $\frac{10}{9}$ | $\frac{6}{9}$ | $\frac{7}{9}$ |

$$\begin{aligned}\therefore E(\bar{X}_A - \bar{X}_B) &= \frac{36}{9} \\ &= \underline{\underline{4}}\end{aligned}$$

$$\therefore \mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B$$

$$\begin{aligned}\sigma_A^2 &= \frac{(4-8)^2 + (8-8)^2 + (12-8)^2}{3} \\ &= \frac{32}{3} \\ &= \underline{\underline{10.67}}\end{aligned}$$

$$\begin{aligned}\sigma_B^2 &= \frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3} \\ &= \frac{8}{3} \\ &= \underline{\underline{2.67}}\end{aligned}$$

- To compute  $Var(\bar{X}_A - \bar{X}_B)$

|  |               |               |                |                |                |                |                |
|--|---------------|---------------|----------------|----------------|----------------|----------------|----------------|
| $(\bar{x}_A - \bar{x}_B)^2$                              | 1             | 4             | 9              | 16             | 25             | 36             | 49             |
| $p(\bar{x}_A - \bar{x}_B)$                               | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{2}{9}$  | $\frac{1}{9}$  | $\frac{2}{9}$  | $\frac{1}{9}$  | $\frac{1}{9}$  |
| $(\bar{x}_A - \bar{x}_B) \cdot p(\bar{x}_A - \bar{x}_B)$ | $\frac{1}{9}$ | $\frac{4}{9}$ | $\frac{18}{9}$ | $\frac{16}{9}$ | $\frac{50}{9}$ | $\frac{36}{9}$ | $\frac{49}{9}$ |

$$Var(\bar{X}_A - \bar{X}_B) = \sum (\bar{X}_A - \bar{X}_B)^2 \cdot P(\bar{X}_A - \bar{X}_B) - [E(\bar{X}_A - \bar{X}_B)]^2$$

$$Var(\bar{X}_A - \bar{X}_B) = \frac{174}{9} - 4^2$$

$$= \underline{\underline{3.33}}$$

Variance of  $(\bar{X}_A - \bar{X}_B)$  using the two population variances is as follows.

$$Var(\bar{X}_A - \bar{X}_B) = \frac{\sigma_A^2}{n_A} \left( \frac{N_A - n_A}{N_A - 1} \right) + \frac{\sigma_B^2}{n_B} \left( \frac{N_B - n_B}{N_B - 1} \right)$$

$$\left( \frac{32}{3} \times \frac{1}{2} \times \frac{1}{2} \right) + \left( \frac{8}{3} \times \frac{1}{2} \times \frac{1}{2} \right)$$

$$\frac{8}{3} + \frac{2}{3} = \frac{10}{3}$$

$$= \underline{\underline{3.33}}$$

- After explaining the mean and variance of the sampling distribution of sample means, explain the shape of the sampling distribution of the sample means at each of the following situations.
- Point out that the sampling distribution of the difference between two sample means, once the populations are normally distributed with known variances is as follows.

$$\bar{X}_A - \bar{X}_B \sim N \left( \mu_A - \mu_B, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} \right)$$

- Once the sampling distribution of the difference between two sample means is built using large samples drawn from normally distributed populations with unknown variances, the sample variances are applied as good estimators for population variances and point out that the sampling distribution falls as mentioned below.

$$\bar{X}_1 - \bar{X}_2 \rightsquigarrow N \left[ \mu_1 - \mu_2, \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]$$

- In the above  $S_1^2$  is used for  $\sigma_1^2$  and  $S_2^2$  is used for  $\sigma_2^2$
- Once the sampling distribution of the difference between two sample means is built using small size samples drawn from normally distributed populations with unknown variances, the sample variances are applied as good estimators for unknown population variances and point out that the sampling distribution falls t - distribution as mentioned below.

$$\bar{X}_1 - \bar{X}_2 \rightsquigarrow t \left[ \mu_1 - \mu_2, \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]$$

- Point out the fact the t-distribution depends on  $n_1 + n_2 - 2$  number of degrees of freedom.
- Once the sampling distribution of the difference between two sample means for large size samples drawn from non-normal populations with known variances is built, explain that the sampling distribution is assumed to fall approximately normally in accordance with the Central Limit Theorem as follows.

$$\bar{X}_1 - \bar{X}_2 \rightsquigarrow N \left[ (\mu_1 - \mu_2), \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right]$$

- Once the sampling distribution of the difference between two sample means for large size samples drawn from non-normal populations with unknown variances is built explain that the sampling distribution is assumed to fall approximately normally in accordance with the Central Limit Theorem since the sample variances are applied as good estimators for unknown population variance as follows.

$$\bar{X}_1 - \bar{X}_2 \rightsquigarrow N \left[ \mu_1 - \mu_2, \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]$$

- In the above  $S_1^2$  is used for  $\sigma_1^2$  and  $S_2^2$  is used for  $\sigma_2^2$

- Assign the following activity for building the sampling distributions of the difference between two sample means using normally distributed populations.

#### Activity-5

- Build up the sampling distribution of the difference between two sample means relevant to each situation given below.
  1. The life time of two batteries in brand A and brand B have distributed normally with the means 1000 h and 800 h respectively, where as the standard deviations 125 h and 100h. A sample of 25 batteries from brand A and a sample of 16 batteries from brand B are taken into consideration.
  2. The weight of boys and girls in a particular school is distributed normally as the means 48Kg and 45 Kg respectively where as the standard deviations are unknown, When considering two samples with 50 boys and 75 girls the standard deviations of them have been derived as 5 Kg and 4 Kg respectively.
  3. The mass of the packets of sugar packeted in two machines are distributed normally with the means 500g and 495g respectively followed by standard deviations as 0.8g and 0.5g. Two samples of size 12 and 15 packets respectively are considered drawn from these two kinds of packets separately.
- The answer for activity-5 is as follows.

1. When considered the life times of the batteries as  $X_A$  and  $X_2$

$$(\bar{X}_A - \bar{X}_B) \sim N\left[(\mu_A - \mu_B), \left(\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\right)\right]$$

$$(\bar{X}_A - \bar{X}_B) \sim N\left[(1000 - 800), \left(\frac{125^2}{25} + \frac{100^2}{16}\right)\right]$$

2. When considered the weight of students X, the boys B and the girls G, the sampling distribution of the difference between two sample means is as follows.

$$(\bar{X}_B - \bar{X}_G) \sim N\left[(\mu_B - \mu_G), \left(\frac{S_B^2}{n_B} + \frac{S_G^2}{n_G}\right)\right]$$

$$(\bar{X}_B - \bar{X}_G) \sim N\left[(48 - 45), \left(\frac{5^2}{50} + \frac{4^2}{75}\right)\right]$$

3. When considered the mass of the sugar contained in packets as X and the two machines as 1 and 2 the sampling distribution of the difference between two sample means is as follows.

$$(\bar{X}_1 - \bar{X}_2) \sim N\left[(\mu_1 - \mu_2), \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)\right]$$

$$(\bar{X}_1 - \bar{X}_2) \sim N\left[(500 - 495), \left(\frac{0.8^2}{12} + \frac{0.5^2}{15}\right)\right]$$

- Assign the following activity to the students for building the sampling distribution of the difference between two sample means using non normal populations.

#### Activity-6

Build the sampling distributions of the difference between two sample means relevant to each situation mentioned below.

1. The number of lorries coming to the filling station A in an hour and the number of lorries coming to the filling station B in an hour fall in two Poisson distributions whose means  $\lambda = 2$  and  $\lambda = 1.5$  respectively. When considered 100 intervals of 1 hour the sampling distribution of the difference between two sample means
2. 2% of the items manufactured in machine A and 1% of the items manufactured in machine B are found to be defective. 20 units in each item are packed in boxes. If two samples of 10 boxes from each item were drawn in random, build up the sampling distribution of the difference between two sample means of the defective units contained in two samples.

The solution for the activity 1 and 2 are as follows.

- I. If the number of Lorries coming to the filling centre is denoted by – X.

$$X_A \sim Po(2)$$

$$X_B \sim Po(1.5)$$

$$\mu_{\bar{X}_A} = 2 \quad \sigma_{\bar{X}_A}^2 = \frac{2}{100} \quad \mu_{\bar{X}_B} = 1.5 \quad \sigma_{\bar{X}_B}^2 = \frac{1.5}{100}$$

$$\therefore (\bar{X}_A - \bar{X}_B) \sim N\left[(2 - 1.5), \left(\frac{2}{100} + \frac{1.5}{100}\right)\right]$$

- II. If the number of defective units contained in a box is denoted by X.

$$X_A \sim B(20, 0.02)$$

$$X_B \sim B(20, 0.01)$$

Since the number of items in 10 samples of 20 in each = 200

$$\therefore X_A \sim N \left[ 4, \frac{3.92}{200} \right] \quad X_B \sim N \left[ 2, \frac{1.98}{200} \right]$$

$$\therefore (\bar{X}_A - \bar{X}_B) \sim N \left[ (4-2), \left( \frac{3.92}{200} + \frac{1.98}{200} \right) \right]$$

### A Guideline to explain the subject matters.

- The probability distribution of the values that can be taken by the difference of the means of all the possible independent samples drawn in random from two populations is known as the sampling distribution of the difference between two sample means.
- If the means of two independent samples drawn in random with the size  $n_1$  and  $n_2$  respectively from the two populations with the size  $N_1$  and  $N_2$  are  $\bar{X}_1$  and  $\bar{X}_2$  respectively, the probability distribution of the possible values taken by  $(\bar{X}_1 - \bar{X}_2)$  is the sampling distribution of the difference between two sample means.
- The mean of the sampling distribution of the difference between two sample means  $\mu_{\bar{X}_1 - \bar{X}_2}$  is equal to the difference of the two population means  $(\mu_1 - \mu_2)$

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

- The variance of the sampling distribution of the difference between two sample means.
  - When the populations are finite

$$\sigma^2_{(\bar{X}_1 - \bar{X}_2)} = \left[ \frac{\sigma_1^2}{n_1} \left( \frac{N_1 - n_1}{N_1 - 1} \right) + \frac{\sigma_2^2}{n_2} \left( \frac{N_2 - n_2}{N_2 - 1} \right) \right]$$

- When the populations are infinite

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

- When  $X$  and  $Y$  are two populations such that  $X_1 \sim N(\mu_1, \sigma_1^2)$  and  $X_2 \sim N(\mu_2, \sigma_2^2)$  the sampling distribution of the difference between two sample means of two samples with size  $n_1$  and  $n_2$   $(\bar{X}_1 - \bar{X}_2)$  falls normally regardless of the fact whether samples are small or large

$$(\bar{X}_1 - \bar{X}_2) \sim N \left[ (\mu_1 - \mu_2), \left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \right]$$

- If the population distributions are not normal, the sampling distribution of the difference between two sample means  $(\bar{X}_1 - \bar{X}_2)$  constructed on large size samples falls approximately normally with the mean  $(\mu_1 - \mu_2)$  and variance  $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

- If the population distributions are normal and the population variances are unknown the sampling distribution of the difference between two sample means of two small size samples  $n_1 < 30$  and  $n_2 < 30$  falls in a t - distribution with  $n_1 + n_2 - 2$  degrees of freedom with the assumption that  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  (common variance).
- Building the sampling distribution of the difference between two sample means

When the populations are normal

|                    |  |   |
|--------------------|--|---|
| $\sigma^2$ Known   | $\left\{ \begin{array}{l} n < 30 \\ n > 30 \end{array} \right\}$ | $(\bar{X} - \bar{Y}) \sim N \left[ (\mu_1 - \mu_2), \left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \right]$ |
|                    |  | $(\bar{X} - \bar{Y}) \sim t \left[ (\mu_1 - \mu_2), \left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right) \right]$           |
| $\sigma^2$ Unknown | $\left\{ \begin{array}{l} n < 30 \\ n > 30 \end{array} \right\}$ | $(\bar{X} - \bar{Y}) \sim N \left[ (\mu_1 - \mu_2), \left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right) \right]$           |

When the populations are non normal

|                    |  |   |
|--------------------|--|---|
| $\sigma^2$ Known   | $\left\{ \begin{array}{l} n > 30 \\ n < 30 \end{array} \right\}$ | $(\bar{X} - \bar{Y}) \sim N \left[ (\mu_1 - \mu_2), \left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \right]$ |
|                    |  | cannot be built   |
| $\sigma^2$ Unknown | $\left\{ \begin{array}{l} n > 30 \\ n < 30 \end{array} \right\}$ | $(\bar{X} - \bar{Y}) \sim N \left[ (\mu_1 - \mu_2), \left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right) \right]$           |
|                    |  | cannot be built   |

### Assessment and Evaluation :

Solve the following problems.

1. The mean and standard deviation of the weight of the boys in a school are 48Kg and 5 Kg respectively. The mean and standard deviation of the weight of the girls in the same school are 45 Kg and 8 Kg respectively.

(a) If the weight of the boys and girls is assumed to be distributed normally.

- (i) Find the probability of the mean weight of 25 boys being 4.5 kg greater than the mean weight of 16 girls.

- (ii) Find the probability of the mean weight of 49 boys being 4.5 kg greater than the mean weight of 64 girls.
- (b) If the weight of these boys and girls is assumed to be non-normal, give your opinion regarding the meaningfulness of the answers in the above section (a).

2. A person who is manufacturing a small radio, purchases the batteries used in them from A and B suppliers. The mean life time of a battery provided by A is 820h and that of B is 790h. The standard deviation of the samples of size 80 drawn from A and the standard deviation of the samples size 90 are 25h and 36h respectively.

- (i) What is the probability that the mean life time of a brand A battery being 35 hours less than that of brand B battery?
- (ii) Find the probability that the difference between the mean life time of brand A battery and brand B battery falling in the range of 10 hours from the difference between two population means?

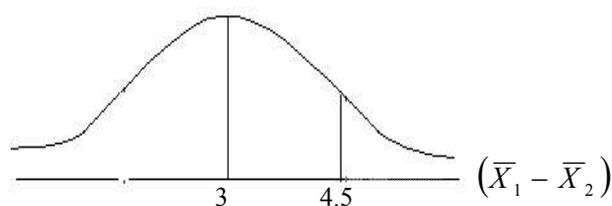
### Answers

- (i). If it is denoted that boys A and girls B the weight-X.

$$\bar{X}_A \sim N \left[ 48, \frac{5^2}{25} \right] \quad \bar{X}_B \sim N \left[ 45, \frac{8^2}{16} \right]$$

$$\therefore (\bar{X}_A - \bar{X}_B) \sim N \left[ (48 - 45), \left( \frac{5^2}{25} + \frac{8^2}{16} \right) \right]$$

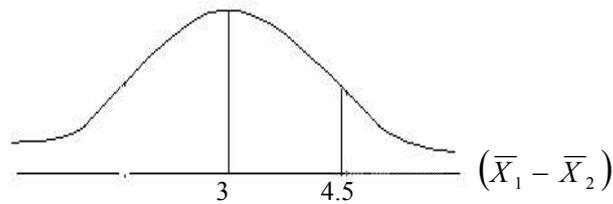
$$\begin{aligned} \Pr(\bar{X}_A - \bar{X}_B) > 4.5 &= \Pr \left[ Z > \frac{4.5 - 3.0}{\sqrt{\frac{5^2}{25} + \frac{8^2}{16}}} \right] \\ &= \Pr(Z - 0.67) = 0.5000 - 0.2486 \\ &= \underline{\underline{0.2514}} \end{aligned}$$



01. (a) (ii)

$$\therefore (\bar{X}_A - \bar{X}_B) \sim N \left[ (48 - 45), \left( \frac{5^2}{49} + \frac{8^2}{64} \right) \right]$$

$$\Pr o(\bar{X}_1 - \bar{X}_2) > 4.5$$



$$= \Pr(Z) > \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \Pr(Z) > \frac{4.5 - 3.0}{\sqrt{\frac{5^2}{49} + \frac{8^2}{64}}}$$

$$P(Z > 1.22) = 0.5000 - 0.3888 \\ = 0.1112$$

(b) If the population distributions are not normal, the answer (a) (i) above is not meaningful, since the sample size is small and there by the sampling distribution of the difference between two sample means are not normal.

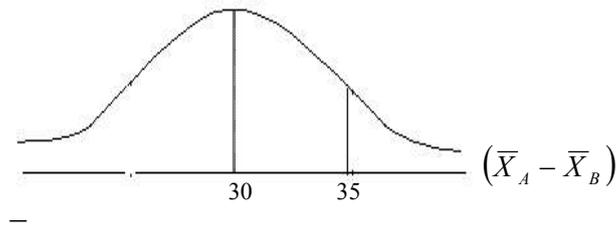
The answer for (a) (ii) above is meaningful, because the sampling distributions of sample means approximately falls normally for large size samples whatever the shape of the populations may be.

(2) (i) Considering the life time of a battery as X

$$\bar{X}_A \sim N \left[ 820, \frac{25^2}{80} \right] \quad \bar{X}_B \sim N \left[ 790, \frac{36^2}{90} \right]$$

$$(\bar{X}_A - \bar{X}_B) \sim N \left[ (820 - 790), \frac{25^2}{80} + \frac{36^2}{90} \right]$$

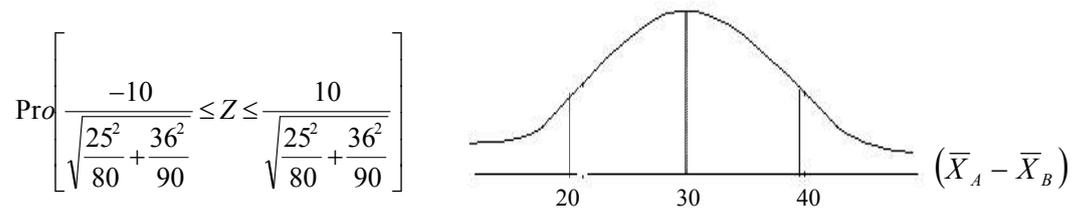
$$\text{Pro}(\bar{X}_A - \bar{X}_B) < 35$$



$$\text{Since } Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

$$\begin{aligned} \text{Pro}(\bar{X}_A - \bar{X}_B) < 35 &= \Pr \left[ Z < \frac{35 - 30}{\sqrt{\frac{25^2}{80} + \frac{36^2}{90}}} \right] \\ &= \Pr[Z < 1.061] \\ &= 0.5000 + 0.3554 \\ &= \underline{\underline{0.8554}} \end{aligned}$$

$$(ii) \quad \text{Pr}[\mu_A - \mu_B - 10 \leq \bar{X}_A - \bar{X}_B \leq \mu_A - \mu_B + 10]$$



$$\text{Pro} \left[ \frac{-10}{\sqrt{\frac{25^2}{80} + \frac{36^2}{90}}} \leq Z \leq \frac{10}{\sqrt{\frac{25^2}{80} + \frac{36^2}{90}}} \right]$$

$$\text{Pro} \left[ \frac{-10}{4.713} \leq Z \leq \frac{10}{4.713} \right]$$

$$\begin{aligned} \text{Pr}[-2.12 \leq Z \leq 2.12] &= 0.4830 + 0.4830 \\ &= \underline{\underline{0.9660}} \end{aligned}$$

**Competency 07** : Uses Statistical Inference for making business decisions

**Competency Level 7.4** : Builds the sampling distribution of sample proportions for statistical inference

**No. of Periods** : 08

**Learning outcomes:**

- Explains the sample proportion and population proportion.
- Describes the Sampling Distribution of Sample Proportions.
- Builds the sampling distribution of sample proportions.
- States the mean and variance of the sampling distribution of sample proportions.
- Describes the sampling distribution of sample proportions when the sample size is large.
- Solves problems using the sampling distribution of sample proportions.
- Makes decisions using the sampling distribution of sample proportions.

**Instructions for lesson planning :**

- Hold a discussion with the students as follows.
  - Suppose that a population consists of three girls named  $F_1, F_2, F_3$  and two boys named  $M_1, M_2$
  - Mention that the factor “female” is an attribute and when the number of elements belonged to that attribute (A) is expressed as a ratio to the total number of population elements N and further that the population proportion is symbolized as

$$\begin{aligned}\pi &= \frac{A}{N} \\ &= \frac{3}{5} \\ &= \underline{\underline{0.6}}\end{aligned}$$

- Point out that once the number of elements belonged to a particular attribute in the sample (**a**) is expressed as a ratio to the total number of sample elements (**n**) that is the sample proportion – **p** and also that the values taken by the sample proportion is denoted by ‘**p**’.

$$p = \frac{a}{n}$$

Activity– 1.

- Raise the following questions to the students related to the above discussed example of the population with 3 girls and 2 boys.
  - Write down all the possible samples that can be drawn per two children in each,
  - Find the proportion of girls in each sample.
  - Construct a chart containing the girls' proportion in each sample and the corresponding probability value.

Solution–Activity-1

- All the possible samples that can be drawn.  
 $(F_1, F_2)$   $(F_1, F_3)$   $(F_1, M_1)$   $(F_1, M_2)$   $(F_2, F_3)$   
 $(F_2, M_1)$   $(F_2, M_2)$   $(F_3, M_1)$   $(F_3, M_2)$   $(M_1, M_2)$
- Girl proportion in each sample –  
 1, 1, 0.5, 0.5, 1, 0.5, 0.5, 0.5, 0.5, 0
- Probability distribution of–P

| $P$ | $P(p)$         |
|-----|----------------|
| 0   | $\frac{1}{10}$ |
| 0.5 | $\frac{6}{10}$ |
| 1   | $\frac{3}{10}$ |

- Explain that this probability distribution is the sampling distribution of sample proportions.

Activity–2

- Find the mean of the sampling distribution of sample proportions constructed in the above activity-I.
- Identify the relationship between that value and the population proportion.
- Compute the variance of the sampling distribution of sample proportions  $\sigma_p^2$  using the following formula.

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n} \left( \frac{N-n}{N-1} \right)$$

- Find the standard deviation of the sampling distribution of sample proportions  $\sigma_p$  as follows.

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n} \left( \frac{N-n}{N-1} \right)}$$

Solution – Activity 2 :

$$\begin{aligned} \mu_p &= \frac{1+1+0.5+0.5+1+0.5+0.5+0.5+0.5+0}{10} \\ &= \underline{\underline{0.6}} \\ \mu_p &= \pi \end{aligned}$$

- Variance

$$\begin{aligned} \sigma_p^2 &= \frac{\pi(1-\pi)}{n} \left( \frac{N-n}{N-1} \right) \\ &= \frac{0.6 \times 0.4}{2} \left( \frac{5-2}{5-1} \right) \\ &= \underline{\underline{0.09}} \end{aligned}$$

Standard deviation

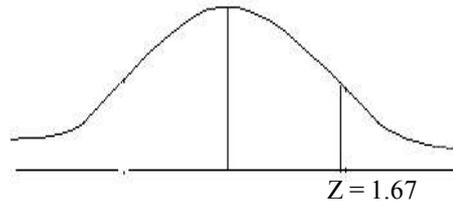
$$\begin{aligned} \sigma_p &= \sqrt{\frac{\pi(1-\pi)}{n} \left( \frac{N-n}{N-1} \right)} \\ &= \sqrt{\frac{0.6 \times 0.4}{2} \left( \frac{5-2}{5-1} \right)} \\ &= 0.3 \end{aligned}$$

Activity 3

- Assign the following activity to the students.  
It has been found that 3% of the units manufactured in a particular machine is defective. In a sample of 800 units, find the probability that there will be,
  - 4% or more
  - 2.5% or more defective units
  - Come out with your idea about the possibility of a particular batch being good enough when the defective proportion in this sample is less than 2.5%.

Solution – Activity 3 :

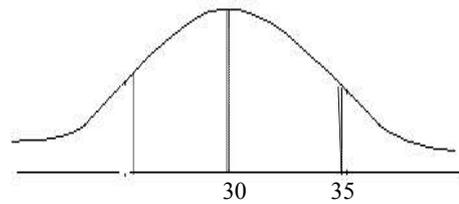
$$\begin{aligned} \text{(i)} \quad \mu_p &= \pi & \sigma_p^2 &= \frac{0.03 \times 0.97}{800} & \sigma_p &= \sqrt{\frac{0.03 \times 0.97}{800}} \\ &= \underline{\underline{0.03}} & &= \underline{\underline{0.000036}} & &= \underline{\underline{0.006}} \\ \Pr(p \geq 0.04) &= \Pr\left(Z \geq \frac{0.04 - 0.03}{0.006}\right) & &= \Pr\left(Z \geq \frac{0.01}{0.006}\right) \\ & & &= \Pr(Z \geq +1.67) \\ & & &= 0.5 - 0.4525 \\ & & &= \underline{\underline{0.0475}} \end{aligned}$$



(ii)

$$Z = \frac{p - \mu_p}{\sigma_p}$$

$$= \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$



$$\Pr(\Pr \geq 0.025) = P\left(Z \geq \frac{0.025 - 0.03}{0.006}\right)$$

$$p(Z \geq -0.83) = 0.5 + 0.2967$$

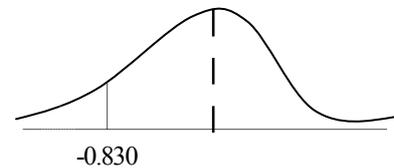
$$= \underline{\underline{0.7967}}$$

(iii)  $\Pr(P \leq 0.025) = \Pr\left(Z \leq \frac{0.025 - 0.03}{0.006}\right)$

$$= \Pr(Z \leq -0.83)$$

$$= 0.5000 - 0.2967$$

$$= \underline{\underline{0.2033}}$$



- Once a sample of 800 units is drawn, it can be determined that the probability of that batch being good is 20%.

**A Guideline to explain the subject matters :**

- Once the number of population elements with a specific attribute (**A**) is expressed as a ratio to the total number of population elements (**N**) the population proportion of that attribute is derived as

$$\pi = \frac{A}{N}$$

- Once the number of sample elements with a specific attribute (**a**) is expressed as a ratio to the total number of sample elements (**n**) the sample proportion is derived as

$$p = \frac{a}{n}$$

- The probability distribution of the values taken by the proportions of all the possible samples with same size that can be drawn from a population is known as the sampling distribution of the sample proportions.
- The mean of the sampling distribution of sample proportions is symbolized as  $\mu_p$  the variance as  $\sigma_p^2$  and the standard deviation as  $\sigma_p$
- The mean of the sampling distribution of sample proportions is equal to the population proportions  $\mu_p = \pi$
- The variance of this distribution when population is infinite

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n}$$

- The variance of this distribution when the population is finite

$$\sigma_p^2 = \frac{\pi(1-\pi)}{n} \left( \frac{N-n}{N-1} \right)$$

- Standard deviation of this distribution when the population is infinite

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- Standard deviation of this distribution when the population is finite

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n} \left( \frac{N-n}{N-1} \right)}$$

- If  $\frac{n}{N} < 0.05$  the application of finite population correction factor is not necessary.
- Once the sample size is large ( $n > 100$ ) the sampling distribution of the sample proportions is approximated to a normal distribution with the above mentioned mean and variance in accordance with the Central Limit Theorem.
- The sampling distribution of the sample means can be symbolically notated as follows.

$$p \sim N(\mu_p, \sigma_p^2)$$

$$p \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

**Competency 7.0** : Uses Statistical Inference for making business decisions

**Competency Level 7.5** : Builds the sampling distribution of the difference between two samples proportions for statistical inference

**No. of Periods** : 08

**Learning outcomes:**

- Gives instances for situations where the difference between two sample proportions is required.
- Builds the sampling distribution of the difference between two sample proportions.
- Derives the mean and variance of the sampling distribution of the difference between two sample proportions.
- Writes down sampling distribution of the difference between two sample proportions.
- Solves the problems related to the sampling distribution of the difference between two sample proportions.

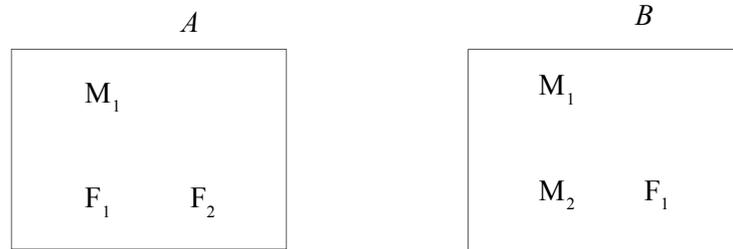
**Instructions for Lesson Planning :**

- Discuss with the students regarding the following situations.
- Firm A declares that 2% of the products made by them are defectives. Firm B declares that 3% of the products made by them are defectives. Regarding the possibility of the products of firm-B being defective more than that of the firm A.
- The company X declares that 85% of the patients to whom their drug was given as treatment for a particular disease were able to escape from this disease. The company Y declares that 78% of the patients to whom their drug was given for the same disease were able to escape from the disease. The possibility of the disease getting cured by using drug X should have been checked over, the possibility of getting that disease cured by using the drug Y.
- It has been detected that 60% of the adults and 70% of the children are fond of eating the food. 'F'. Regarding the preference of adults and children for this food item.
- Hold a discussion highlighting the following facts in connection with each of the above mentioned situations.
  - Introducing the two populations related to each situation.
  - Pointing out the proportional values related to each population.
  - Explaining the need for examining the difference between those population proportions.

Activity – 1.

- Suppose that there are two girls named  $F_1$ ,  $F_2$  and a boy named  $M_1$  in the class A and in class B, two boys named  $M_1$ ,  $M_2$  and a girl named  $F_1$ .
- Write all the possible samples of size 2 from all 3 students in class A.
- Write the girl proportion of each sample drawn from class A.
- Write all the possible samples of size 2 from all 3 students in class – B.
- Write the girl proportion of each sample drawn from class B.
- Write all the possible ways through which the difference between the sample proportions of the sample drawn from A class and the sample proportions of the samples drawn from B class could be derived and write the difference of each pair.
- Note down the various values received for the difference between the sample proportions.
- State the corresponding probability value or each value of the difference of sample proportions.

Solution – Activity – 1.



| Class - A    |             | Class - B    |             |
|--------------|-------------|--------------|-------------|
| Sample       | Proportions | Sample       | Proportions |
| $(M_1, F_1)$ | 0.5         | $(M_1, M_2)$ | 0           |
| $(M_1, F_2)$ | 0.5         | $(M_1, F_1)$ | 0.5         |
| $(F_1, F_2)$ | 1           | $(M_2, F_1)$ | 0.5         |

| Possible ways of the difference between sample proportions to be derived | Difference between sample proportions |
|--|---------------------------------------|
| 0.5 - 0  | 0.5                                   |
| 0.5 - 0.5  | 0                                     |
| 0.5 - 0.5  | 0                                     |
| 0.5 - 0  | 0.5                                   |
| 0.5 - 0.5  | 0                                     |
| 0.5 - 0.5  | 0                                     |
| 1 - 0  | 1                                     |
| 1 - 0.5  | 0.5                                   |
| 1 - 0.5  | 0.5                                   |

| Difference between sample proportions | Probability $pr(p_1 - p_2)$ |
|---------------------------------------|-----------------------------|
| 0                                     | $\frac{4}{9}$               |
| 0.5                                   | $\frac{4}{9}$               |
| 1                                     | $\frac{1}{9}$               |

- Explain that the above distribution is the sampling distribution of the difference between two sample proportions.

#### Activity–2

Involve the students in the following activity based on the above activity – 1.

- Compute the mean of the sampling distribution of the difference between two sample proportions  $(\mu_{p_1 - p_2})$
- Make out the relationship between that value and the difference between two population proportions.
- Compute the variance of the sampling distribution of the difference between two sample proportions  $\sigma_{p_1 - p_2}^2$  using the following formula.

$$\frac{\pi_1(1-\pi_1)}{n_1} \left( \frac{N_1-n_1}{N_1-1} \right) + \frac{\pi_2(1-\pi_2)}{n_2} \left( \frac{N_2-n_2}{N_2-1} \right)$$

- Compute the variance of the sampling distribution of the difference between two sample proportions as the variance of a probability distribution is computed and compare with the value derived through the above formula.

Solution – Activity – 2

- Mean of the sampling distribution of the difference between two sample proportions.

$$\begin{aligned} E(p) &= \left( 0 \times \frac{4}{9} \right) + \left( 0.5 \times \frac{4}{9} \right) + \left( 1 \times \frac{1}{9} \right) \\ &= \underline{\underline{0.3333}} \end{aligned}$$

- Difference between two population proportions.

$$\begin{aligned} \pi_1 &= \frac{2}{3} \\ \pi_2 &= \frac{1}{3} \quad (\pi_1 - \pi_2) \\ &= \underline{\underline{0.3333}} \end{aligned}$$

|                                 |
|---------------------------------|
| $\mu_{p_1-p_2} = \pi_1 - \pi_2$ |
|---------------------------------|

- Variance of the sampling distribution of the difference between two sample proportions.

$$\begin{aligned} &\frac{\pi_1(1-\pi_1)}{n_1} \left( \frac{N_1-n_1}{N_1-1} \right) + \frac{\pi_2(1-\pi_2)}{n_2} \left( \frac{N_2-n_2}{N_2-1} \right) \\ &= \left( \frac{2}{3} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2} \right) + \left( \frac{1}{3} \times \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} \right) \\ &= \underline{\underline{0.1111}} \end{aligned}$$

- That variance as the variance of a probability distribution.

$$\begin{aligned} Var(p_1 - p_2) &= \left( 0^2 \times \frac{4}{9} \right) + \left( 0.5^2 \times \frac{4}{9} \right) + \left( 1^2 \times \frac{1}{9} \right) - \left( \frac{1}{3} \right)^2 \\ &= \underline{\underline{0.1111}} \end{aligned}$$

- Explain that the population correction factor is not applied for infinite populations as well as for samples with replacement.

$$\text{Var}(P_1 - P_2) = \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}$$

### Activity – 3

Involve the students in the following activity.

- Write the sampling distribution of the difference between two sample proportions based on the above activity – 2.

### Solution activity – 3

$$P_1 - P_2 \sim N \left( \mu_{p_1 - p_2}, \sigma_{p_1 - p_2}^2 \right)$$

$$P_1 - P_2 \sim N \left( \pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2} \right)$$

### Activity – 4

Involve the students in following activity.

- It is known that the proportion of recovering from a particular disease by using the drug species A is 0.85 and the proportion of recovering from the same disease by using the drug species B is 0.78. Once a sample of 100 patient using the drug A and a sample of 200 patients using the drug B are drawn out in random and checked up, find the probability that the proportion of recovering from the disease by using drug A being greater than 0.1 to that proportion of the patients using the drug B.

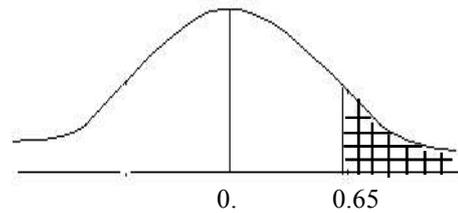
### Solution – Activity – 4

$$\pi_1 = 0.85 \quad n_1 = 100$$

$$\pi_2 = 0.78 \quad n_2 = 200$$

$$\begin{aligned} \pi_1 - \pi_2 &= 0.85 - 0.78 \\ &= \underline{\underline{0.07}} \end{aligned}$$

$$\begin{aligned}
\Pr(\pi_1 - \pi_2 > 0.1) &= P(Z) > \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_2)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} \\
&= P(Z) > \frac{0.1 - 0.07}{\sqrt{\frac{0.85 \times 0.15}{100} + \frac{0.78 + 0.22}{200}}} \\
&= P(Z) > .65 \\
&= 0.5 - 0.2422 \\
&= \underline{\underline{0.2578}}
\end{aligned}$$



The probability that the difference between recovering ratios of the two species of drugs being exceeding 10% is 0.2578.

**A guideline to explain the subject matters :**

- Once the proportions with a particular attribute of two independent samples with the size  $n_1$  and  $n_2$  drawn from two populations are symbolized as  $p_1$  and  $p_2$ , the probability distribution of  $p_1 - p_2$  is known as the sampling distribution of the difference between two sample proportions.
- The mean of sampling distribution of the difference between two sample proportions is symbolized as  $\mu_{p_1-p_2}$  variance as  $\sigma_{p_1-p_2}^2$  and the standard deviation as  $\sigma_{p_1-p_2}$

- Mean  $\mu_{p_1-p_2} = \pi_1 - \pi_2$

- Variance (for infinite populations)  $\sigma_{p_1-p_2}^2 = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$

- Variance (for finite populations)

$$\sigma_{p_1-p_2}^2 = \frac{\pi_1(1-\pi_1)}{n_1} \left( \frac{N_1 - n_1}{N_1 - 1} \right) + \frac{\pi_2(1-\pi_2)}{n_2} \left( \frac{N_2 - n_2}{N_2 - 1} \right)$$

- Standard deviation (for infinite populations)

$$\sigma_{p_1-p_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

- Standard deviation (for finite population)

$$\sigma_{P_1-P_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} \left( \frac{N_1-n_1}{n_1-1} \right) + \frac{\pi_2(1-\pi_2)}{n_2} \left( \frac{N_2-n_2}{n_2-1} \right)}$$

- Once the sample size is large ( $n_1 \geq 100$  and  $n_2 \geq 100$ ) the sampling distribution of the difference between two sample proportions approximately falls normally with the above mentioned mean and variance.
- The sampling distribution of the difference between two sample proportions can be interpreted in following notation.

$$P_1 - P_2 \sim N\left(\mu_{P_1-P_2}, \sigma_{P_1-P_2}^2\right)$$

$$P_1 - P_2 \sim N\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right)$$

**Competency 07** : Uses Statistical Inference for making business decisions

**Competency Level 7.6** : Uses the point estimation for estimating population parameters

**No. of Periods** : 10

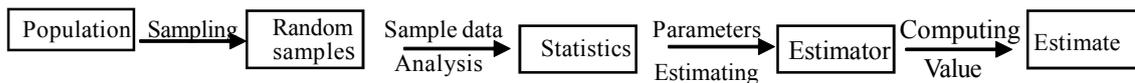
**Learning outcomes:**

- Explains the statistical estimation
- Describes what an estimation is.
- Explains the properties such as unbiasedness, efficiency, consistency and sufficiency that should be possessed by a good point estimator.
- Explains the relationship and the difference between estimator and estimation as well as between estimator and estimate.
- States unbiased estimators for population mean, population proportion and population variance.
- Points out the need of an unbiased estimator with minimum variance for estimation.
- Computes the relative efficiency of an estimator.
- Separates unbiased estimators, efficient estimators and consistent estimators from among few estimators given.
- Points out the need of the variance of estimator being closer to zero, once the sample size is increasing.

**Instructions for lesson planning :**

Activity – 1

Display the following chart regarding the process of statistical estimations.



- Inform the students that the process is indicated by → (arrow head) and the result gained at the end of the process is indicated by  (rectangles) in this flow chart.
- Explain to the students regarding the statistical estimation and the specific terms used in it holding a discussion on the above mentioned flow chart.

Activity – 2

In a small scale business firm where five (5) people are employed, in order to come to a conclusion about their average salary, the average salary is expected to be estimated, using sample data.

- Present the following details and questions to students.
  - Names of five employees and their monthly salaries are as follows.

|                |        |        |          |        |        |
|----------------|--------|--------|----------|--------|--------|
| Name           | Nimal  | Amara  | Thusitha | Sama   | Kumara |
| Monthly salary | 20 000 | 30 000 | 10 000   | 10 000 | 30 000 |

- Write all the possible samples without replacement taking out three names at a time.
- Write their salaries as sample data related to each sample.  
N : B : If the sample data are written in ten thousands, the next calculations may be convenient.
- Find the mean of each sample. Name of those sample means as the statistic -  $\bar{X}$  (function of sample elements)
- Name the unknown parameter regarding this problem.
- Name the estimator.
- Compute the value of the estimator you named using sample data and name it as the “estimate”.

Solutions for activity – 2

- All the possible sample with size – 1 that have been derived.

| Sampling elements          | Monthly salary<br>(Rs. 10 000) | Sample mean<br>$\bar{X}$ |
|----------------------------|--------------------------------|--------------------------|
| 1. Nimal, Amara, Thusitha  | 2, 3, 1                        | 2.00                     |
| 2. Nimal, Amara, Sama      | 2, 3, 1                        | 2.00                     |
| 3. Nimal, Amara, Kumara    | 2, 3, 3                        | 2.67                     |
| 4. Amara, Thusitha, sama   | 3, 1, 1                        | 1.67                     |
| 5. Amara, Sama, Kumara     | 3, 1, 3                        | 2.33                     |
| 6. Thusitha, Sama, Kumara  | 1, 1, 3                        | 1.67                     |
| 7. Nimal, Thusitha, Sama   | 2, 1, 1                        | 1.33                     |
| 8. Nimal, Thusitha, Kumara | 2, 1, 3                        | 2.00                     |
| 9. Amara, Thusitha, Kumara | 3, 1, 3                        | 2.33                     |
| 10. Nimal, Sama, Kumara    | 2, 1, 3                        | 2.00                     |
|                            |                                | 20.00                    |

- Unknown parameter is the average (mean) monthly salary of the five employees in the firm ( $\mu$ )
- Estimator is the sample mean  $\bar{X}$

- Mean of the estimator  $\bar{X}$

$$\bar{X} = 2 + 2 + 2.67 + 1.67 + 2.33 + 1.67 + 1.33 + 2 + 2.33 + 2.$$

$$E(\bar{X}) = \frac{20}{10} = \underline{\underline{2.0}}$$

- The point estimate of the monthly salary of an employee is Rs. 20 000
- Present the following problem to the students.
- Having drawn 3 samples of the employees from the above mentioned firm and the average delay time in reporting to duty has been computed as 12 mts, 8 mts and 10 mts.
  - Compute the point estimate of average delay time in reporting to duty
  - Explain the success or failure of the point estimation in coming to conclusions regarding the population

**Answers :**

1. The point estimate for average delay time in reporting to duty is

$$\begin{aligned} E(\bar{X}) &= \frac{12 + 8 + 10}{3} \\ &= \frac{30}{3} \\ &= \underline{\underline{10}} \end{aligned}$$

2. When conclusions are made for population, problems are arisen regarding the accuracy and precision in estimating with a single value for a large quantity of items.
  - No any idea about the variance of the estimate in not gained.
  - Nothing can be expressed about variations or distributions of population

**Activity 3**

Involve the students in the following activity based on the population considered in the above activity – 2.

1. Write all the possible samples of size 2 drawn from the population of five (5) employees.
2. State the monthly salary related to each sample element as sample data.
3. Find the mean of each sample as  $\bar{X}_1, \bar{X}_2$
4. Compute the expected value of the sampling distribution of those sample means
5. Compute the mean monthly salary  $\mu$  of the population of five employees.
6. Observing the above results, verify that  $E(\bar{X}) = \mu$

- Explain to the students that  $\bar{X}$  is an unbiased estimator for  $(\mu)$

### Activity 3 (Solutions)

| (1) Sample elements | (2) Monthly salary<br>(Rs.10 000) | (3) Sample mean<br>$\bar{X}$ |
|---------------------|-----------------------------------|------------------------------|
| 1. Nimal, Amara     | 2, 3                              | 2.5                          |
| 2. Nimal, Thusitha  | 2, 1                              | 1.5                          |
| 3. Nimal, Sama      | 2, 1                              | 1.5                          |
| 4. Nimal, Kumara    | 2, 3                              | 2.5                          |
| 5. Amara, Thusitha  | 3, 1                              | 2.0                          |
| 6. Amara, Sama      | 3, 1                              | 2.0                          |
| 7. Amara, Kumara    | 3, 3                              | 3.0                          |
| 8. Thusitha, Sama   | 1, 1                              | 1.0                          |
| 9. Thusitha, Kumara | 1, 3                              | 2.0                          |
| 10. Sama, Kumara    | 1, 3                              | 2.0                          |
|                     |                                   | 20.0                         |

$$(4) \quad E(\bar{X}) = \frac{20}{10} = \underline{\underline{2.0}}$$

$$(5) \quad \mu = \frac{2+3+1+1+3}{5} \\ = \frac{10}{5} = \underline{\underline{2.0}}$$

(6)

$\therefore E(\bar{X}) = \mu$

Sample mean  $\bar{X}$  is an unbiased estimator for population mean  $\mu$

### Activity 4

Using the following activity let the students comprehend the unbiasedness of the sample variance ( $S^2$ )

1. Lead the students to compute the variance of monthly salaries of the population of five employees using the relevant formula.

$$\sigma^2 = \frac{\sum(X - \mu)^2}{N}$$

2. Compute the variance of each sample of size 2 consists of monthly salary. Let them

use formula 
$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n}$$

3. Let the students compute the expected value of the sampling distribution of sample variance  $E(S^2)$
4. Raise the following questions to them in connection with the answers received.

(i) Is  $E(S^2) = \sigma^2$

(ii) What is the value of  $\{E(S^2) - \sigma^2\}$ . comment on that value

(iii) Comment on the unbiasedness of the sample variance ( $S^2$ )

- 5; Lead the students to find the variance of monthly salary distribution of five employees population using the formula.

$$\sigma^2 = \frac{\sum(X - \mu)^2}{n - 1}$$

6. Let the students to find the variance of monthly salary in each sample of size 2 using the formula

$$S^2 = \frac{\sum(X - \bar{X})^2}{n - 1}$$

7. Lead them to find the expected value of the sampling distribution of sample variance  $S^2$  found in the above (6).
8. Let them verify the fact that the sample variance  $S^2$  is an unbiased estimator for population variance  $\sigma^2$  reviewing the answers received to the questions from (5) to (7).

**Solutions for Activity 4**

- (i) Population elements 2, 3, 1, 1, 3

Population mean  $\mu = 2.0$

Population variance  $\sigma^2 = \frac{\sum(X - \mu)^2}{N}$

$$\begin{aligned} \therefore \sigma^2 &= \frac{(2-2)^2 + (3-2)^2 + (1-2)^2 + (1-2)^2 + (3-2)^2}{5} \\ &= \frac{0+1+1+1+1}{5} = \frac{4}{5} \\ &= \underline{\underline{0.8}} \end{aligned}$$

(2)

| Sample No. | Sample data monthly salary | Sample mean $\bar{X}$ | Sample variance $S^2 = \frac{\sum(X - \bar{X})^2}{n}$ |
|------------|----------------------------|-----------------------|---|
| 1)         | 2, 3                       | 2.5                   | $\frac{(2 - 2.5)^2 + (3 - 2.5)^2}{2} = 0.25$          |
| (2)        | 2, 1                       | 1.5                   | $\frac{(2 - 1.5)^2 + (1 - 1.5)^2}{2} = 0.25$          |
| (3)        | 2, 1                       | 1.5                   | $\frac{(2 - 1.5)^2 + (1 - 1.5)^2}{2} = 0.25$          |
| (4)        | 2, 3                       | 2.5                   | $\frac{(2 - 2.5)^2 + (3 - 2.5)^2}{2} = 0.25$          |
| (5)        | 3, 1                       | 2.0                   | $\frac{(3 - 2)^2 + (1 - 2)^2}{2} = 1.00$              |
| (6)        | 3, 1                       | 2.0                   | $\frac{(3 - 2)^2 + (1 - 2)^2}{2} = 1.00$              |
| (7)        | 3, 3                       | 3.0                   | $\frac{(3 - 3)^2 + (3 - 3)^2}{2} = 0.00$              |
| (8)        | 1, 1                       | 1.0                   | $\frac{(1 - 1)^2 + (1 - 1)^2}{2} = 0.00$              |
| (9)        | 1, 3                       | 2.0                   | $\frac{(1 - 2)^2 + (3 - 2)^2}{2} = 1.00$              |
| (10)       | 1, 3                       | 2.0                   | $\frac{(1 - 2)^2 + (3 - 2)^2}{2} = 1.00$              |
|            |                            |                       | <u><u>5.00</u></u>                                    |

$$(3) \quad E(S^2) = \frac{5}{10} = \underline{\underline{0.5}}$$

$$(4) \quad (i) \quad E(S^2) \neq \sigma^2$$

$$(ii) \quad \{E(S^2) - \sigma^2\} = 0.5 - 0.8 \\ = \underline{\underline{-0.3}}$$

(iii) The biasedness of the estimator ( $S^2$ ) is  $-0.3$ , hence the sample variance  $S^2$  is not an unbiased estimator for population variance  $\sigma^2$

(5) Population elements 2, 3, 1, 1, 3

Population mean  $\mu = 2.0$

$$\begin{aligned}\text{Population variance } \sigma^2 &= \frac{\sum(X - \mu)^2}{n - 1} \\ &= \frac{0 + 1 + 1 + 1 + 1}{4} \\ &= 1\end{aligned}$$

(6) Sample variances of the ten samples with size 2 in each, drawn from the above mentioned population.

$$\left[ S^2 = \frac{\sum(X - \bar{X})^2}{n - 1} \right]$$

$$(1) \quad \frac{(2 - 2.5)^2 + (3 - 2.5)^2}{1} = 0.5$$

$$(2) \quad \frac{(2 - 1.5)^2 + (1 - 1.5)^2}{1} = 0.5$$

$$(3) \quad \frac{(2 - 1.5)^2 + (1 - 1.5)^2}{1} = 0.5$$

$$(4) \quad \frac{(2 - 2.5)^2 + (3 - 2.5)^2}{1} = 0.5$$

$$(5) \quad \frac{(3 - 2)^2 + (1 - 2)^2}{1} = 2.0$$

$$(6) \quad \frac{(3 - 2)^2 + (1 - 2)^2}{1} = 2.0$$

$$(7) \quad \frac{(3 - 3)^2 + (3 - 3)^2}{1} = 0.0$$

$$(8) \quad \frac{(1 - 1)^2 + (1 - 1)^2}{1} = 0.0$$

$$(9) \quad \frac{(1-2)^2 + (3-2)^2}{1} = 2.0$$

$$(10) \quad \frac{(1-2)^2 + (3-2)^2}{1} = 2.0$$

$$(7) \quad E(S^2) = \frac{10}{10} = 1$$

(8) Since  $\sigma^2 = 1$  and the sample expected value of sample variance  $E(s^2) = 1$ , sample variance  $S^2$  is an unbiased point estimator for population variance  $\sigma^2$

### Activity 5

- Let the students examine the unbiasedness of sample proportions.
  1. Lead them to compute the women proportion of population of 5 employees considered in the activity – 2 above.
  2. Let them write all the possible samples of size 2 in each, that can be drawn from that population.
  3. Lead the students to construct the sampling distribution of women proportions using those sample proportions.
  4. Lead them further to examine whether the sample proportion is an unbiased estimator for population parameter computing the expected value of that sampling distribution.

### Solutions for Activity 5

Employee population of the firm.

(1) Nimal, Amara, Thusitha, Sama, Kumara

No of male employees = 3

No of female employees = 2

Women proportion  $\pi = \frac{2}{5} = \underline{\underline{0.4}}$

| (2) Sample No. | Sample          | (3) Women Proportion |       |
|----------------|-----------------|----------------------|-------|
| (1)            | Nimal, Amara    | 0                    | = 0   |
| (2)            | Nimal, Thusitha | 1/2                  | = 0.5 |
| (3)            | Nimal, Sama     | 1/2                  | = 0.5 |
| (4)            | Nimal, Kumara   | 0                    | = 0.0 |
| (5)            | Amara, Thusitha | 1/2                  | = 0.5 |

|      |                  |     |            |
|------|------------------|-----|------------|
| (6)  | Amara, Sama      | 1/2 | = 0.5      |
| (7)  | Amara, Kumara    | 0   | = 0.0      |
| (8)  | Thusitha, Sama   | 1   | = 1.0      |
| (9)  | Thusitha, Kumara | 1/2 | = 0.5      |
| (10) | Sama, Kumara     | 1/2 | = 0.5      |
|      |                  |     | <u>4.0</u> |

$$(4) \quad E(P) = \frac{4}{10} = 0.4 \quad \text{and} \quad \pi = 0.4$$

$$\therefore E(P) = \pi$$

- Sample proportion 'P' is an unbiased estimator for population proportion  $\pi$

### Activity 6

- Suppose that equal sized samples have been drawn from a finite population distributed normally.
  1. Let the students express the variance of the sampling distribution of the sample mean  $\text{var}(\bar{X})$
  2. Inform that the variance of the sampling distribution of sample median  $\bar{X}_m$  is found as  $\text{var}(X_m) = \frac{\pi\sigma^2}{2n}$
  3. Let the students to examine the estimator with the least variance.
  4. Let the students to determine the more efficient estimator from sample mean  $\bar{X}$  and sample Median  $(X_m)$
  5. Let the students to estimate the efficiency of sample mean  $\bar{X}$  relative to the sample Median  $(X_m)$
  6. Let the students to find the efficiency of the sample Mean  $\bar{X}$  relative to the sample Median  $(X_m)$
  7. Hold a discussion to explain the efficiency of the sample Median  $(X_m)$  relative to the sample Mean  $\bar{X}$

### Solutions for Activity – 6

$$(1) \quad \text{var}(\bar{X}) = \frac{\sigma^2}{n} \qquad (2) \quad \text{var}(X_m) = \frac{\pi\sigma^2}{2n}$$

(3) In  $\text{var}(\bar{X})$ ,  $\sigma^2$  has been multiplied by  $\frac{1}{n}$  Where as in  $\text{var}(X_m)$ ,  $\sigma^2$  has been multiplied by  $\frac{22}{7 \times 2n}$ . Hence the co-efficient of  $\sigma^2$  when  $\text{var}(X_m)$  is computed is  $\frac{1.57}{n}$ .

$$\therefore \frac{\sigma^2}{n} < \frac{\pi\sigma^2}{2n}$$

$$\therefore \text{var}(\bar{X}) < \text{var}(X_m)$$

(4) Sample Mean] ( $\bar{X}$ ) is a more efficient estimator when compared with sample Median ( $X_m$ )

(5) The efficiency of sample mean ( $\bar{X}$ ) relative to the sample Median ( $X_m$ ) is as follows.

$$\begin{aligned} \text{Efficiency} &= \frac{\pi\sigma^2 / 2n}{\sigma^2 / n} = \frac{22}{7} \times \frac{1}{2} \\ &= \underline{\underline{1.57}} \end{aligned}$$

$$(6) \frac{\text{Var}(X_m)}{\text{Var}(\bar{X})} \times 100 = 157\%$$

(7) Relative to the sample median sample mean seems to be efficient at 157%.

#### Activity – 7

1. Guide the students to write the variances of two estimators, sample mean and sample median separately.
2. Inquire about the value of the variances of those estimators, when the sample size (n) is increasing (When  $n \rightarrow \infty$ )
3. Discussing the function of the estimator, when ‘n’ is increasing, explain the consistency of the estimator.

#### Solution – Activity – 7

(1) Variance of the sample mean

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

Variance of the sample median

$$\text{var}(X_m) = \frac{\pi\sigma^2}{2n}$$

(2) When  $n \rightarrow \infty$  ( $n$  is increasing) since the denominator of  $\frac{\sigma^2}{n}$  is increasing, the value of  $\frac{\sigma^2}{n}$  gradually becomes closer to zero.

In the same way, when the sample size is increasing  $n \rightarrow \infty$ , since the denominator of  $\frac{\pi\sigma^2}{2n}$  is also increasing the value of  $\frac{\pi\sigma^2}{2n}$  also gradually becomes closer to zero.

When the variance of an estimator becomes closet to zero that particular estimator goes closer to the parameter once the estimator is centred about the parameter, that estimator becomes consistent. Hence, both the sample mean and sample median are consistent estimators.

#### **A guideline to explain the subject matters :**

- The process of expressing the unknown population parameters using sample statistics (functions of sample elements) derived based on random samples drawn from a population is known as ‘**statistical estimation**’.
- Statistical estimation is of two types
  - Point estimation
  - Interval estimation
- Estimating the unknown population parameter in a single value considering the sample statistic as the estimator, is known as ‘**Point Estimation**’.
- The function of sample elements which are used to estimate the value of the unknown parameter, is known as the ‘estimator’.
- Few estimators used to estimate unknown parameters are as follows.
  - Sample mean -  $(\bar{x})$
  - Sample variance -  $s^2$
  - Sample proportion -  $p$
  - Difference between two sample means -  $(\bar{x}_1 - \bar{x}_2)$
- The value derived by computing the estimator based on sample data is called ‘the estimate’.
- A good point estimator should consist of four properties as follows.
  - Unbiasedness
  - Efficiency

- Consistency
- Sufficiency
- If the expected value of the estimator is equal to the unknown parameter that estimator is an unbiased estimator.
- If the unknown parameter is denoted by  $\theta$  the estimator for that is denoted by  $\hat{\theta}$  Then, if  $E(\hat{\theta}) = \theta$ .

The estimator  $\hat{\theta}$  is considered to be an unbiased estimator for  $\theta$  of an unbiased estimator.

$$\{E(\hat{\theta}) - \theta\} = 0$$

- Examples for an unbiased estimator. Since  $E(\bar{X}) = \mu$  The sample mean  $\bar{X}$  is an unbiased estimator for population mean  $\mu$ .
- Since the Mean ( $\bar{X}$ ) Median (Md) and Mode (Mo) of a normal distribution are equal, Mode and Median are also unbiased estimators.
- When considered the sample proportion as 'P' since  $E(p) = \pi$  Sample proportion (p) is an unbiased estimator for population proportion  $\pi$

- When considered the sample variance ( $S^2$ ) only if ( $S^2$ ) computed as  $\frac{\sum (X_i - \bar{X})^2}{n-1}$  the sample variance ( $S^2$ ) becomes an unbiased estimator for population variance ( $\sigma^2$ ) because at that chance  $E(S^2) = \sigma^2$

- Once equal size samples have been drawn the estimator with the least variance among the few unbiased estimators is known as the efficient estimator.
- Once  $T_1$  and  $T_2$  are two unbiased estimators for  $\theta$  parameter the efficiency of  $T_2$  relative to the efficiency of  $T_1$  is known as the Relative Efficiency of  $T_2$ . Hence the relative efficiency of  $T_2$  is found as follows.

$$Var(T_2 / T_1) = \frac{Var(T_2)}{Var(T_1)}$$

- If the estimator is centred about the unknown parameter, when the sample size (n) is increasing, that's a consistent estimator.
- In estimation of any population with mean  $\mu$  and variance  $\sigma^2$  when the sample size is sufficiently large ( $n \rightarrow \infty$ ) since the variance of  $\bar{X}$  is closer to zero

$\left( Var(\bar{X}) = \frac{\sigma^2}{n} \rightarrow 0 \right)$  the sample mean  $\bar{X}$  is a consistent estimator for population mean  $\mu$ .

- If all the sample data have been used to compute the estimator in order to estimate the unknown parameter, that is known as a sufficient estimator.
- Since all the data are used to compute the sample mean  $\bar{X}$  then  $\bar{X}$  is a sufficient estimator for  $\mu$  and since it can be assumed that all the information about the sample elements are summarized by taking the sample mean.
- When considering the sample mean, the sample median and the sample mode it is seen that the sample mean is a sufficient estimator for  $\mu$ , the sample mean can be considered as the one and only estimator that possesses all the properties: unbiasedness, efficiency, consistency and sufficiency.

**Competency 7.0 :** Uses Statistical Inference for making business decisions

**Competency Level 7.7 :** Uses interval estimation for estimating population parameters

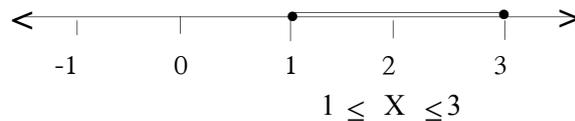
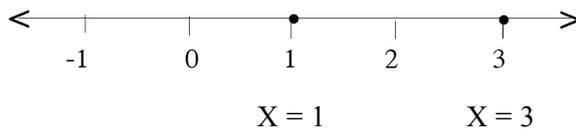
**No. of Periods :** 04

**Learning outcomes:**

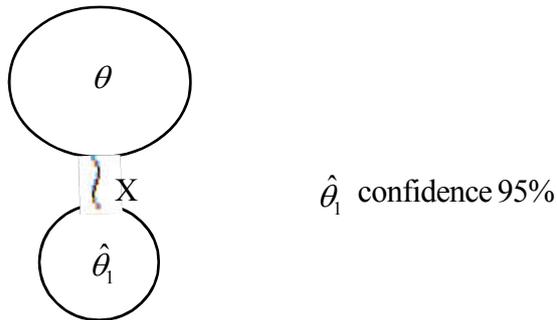
- Explains what the interval estimation is.
- Distinguishes between point estimation and interval estimation.
- Expresses a confident interval for a population parameter related to a given confidence level.
- Interprets ‘a confidence interval’.
- Interprets confidence limits
- Distinguishes between confidence co-efficient and confidence level.
- Introduces the ‘probable error.’
- Analyses the eligibility (appropriateness) of interval estimation rather than point estimation.

**Instructions for Lesson Planning :**

- Produce the following two number lines and the diagram to the class.



- Ensure the facts that,
  - X has been assigned an exact single value in accordance with the first number line, and
  - X can take any value in a given range of values in accordance with the second number line.
- According to the diagram.
- The estimator  $\hat{\theta}$  is used to estimate  $\theta$ , the parameter. The estimate  $\hat{\theta}$  has been derived by analyzing the sample data and hence ensure that  $\theta$  is estimated using that estimate  $\hat{\theta}$  Point out that, this process is just equivalent to derive the solution using a single value on a number line.



$$\mu = \bar{x} \pm t_{\alpha/2; n-1} \cdot \frac{S}{\sqrt{n}}$$

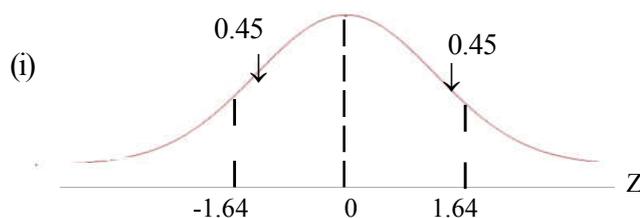
$$\Pr(\theta_{1x} \leq \theta \leq \theta_{1y}) = 95\%$$

- In order to estimate  $\theta$  (the parameter) all that the value of the estimator, confidence level, variance, sample size and population distribution are used. Then, point out that a range of values could be derived for  $\theta$ .
- Explain that this process is just equivalent to derive the solution in a range of values between two points in a number line.
- Ascertain the fact that '95%' is known as the confidence level.
- Point out that  $\hat{\theta}_{1x}$  and  $\hat{\theta}_{1y}$  are called confidence limits.
- Explain that the range containing all possible values between  $\hat{\theta}_{1x}$  and  $\hat{\theta}_{1y}$  is known as the confidence interval.

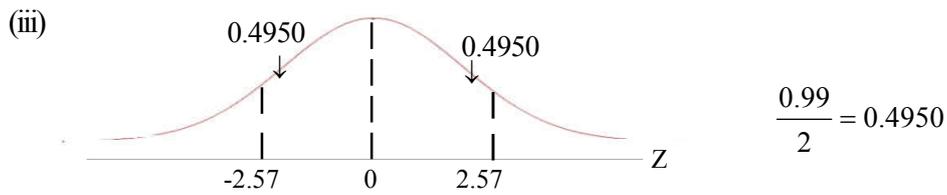
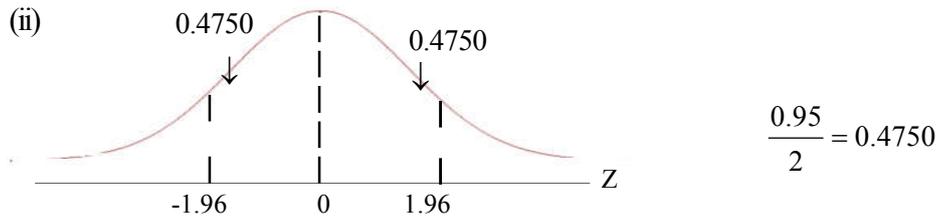
### Activity – 1

- Derive  $Z_{\alpha/2}$  values related to each confidence level given below.
  - for 90% confident level
  - for 95% confident level
  - for 99% confident level

### Solutions

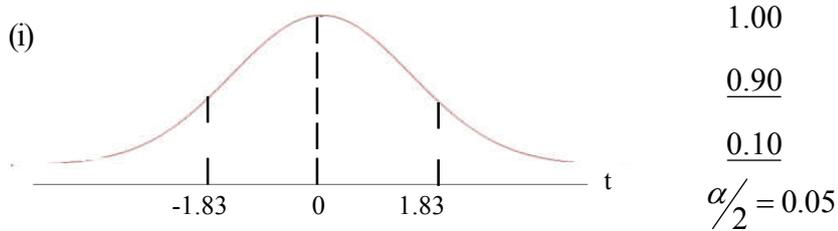


$$\frac{0.90}{2} = 0.4500$$

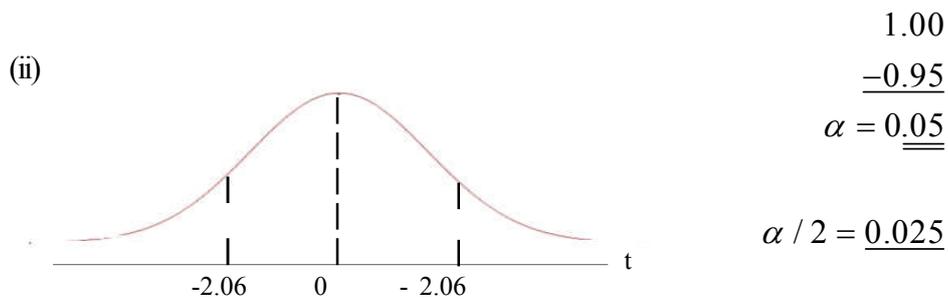


(02) Derive the value of  $t_{\alpha/2; n-1}$  in accordance with each confidence level and the sample size.

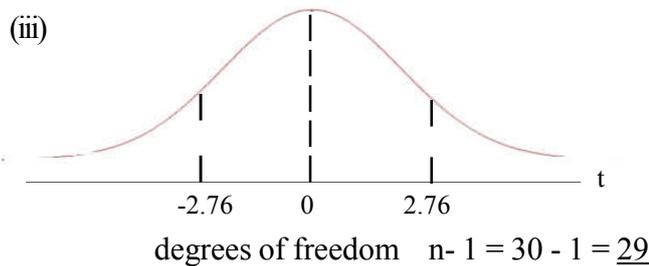
- (i) 90% confident level and  $n = 10$
- (ii) 95% confident level and  $n = 25$
- (iv) 99% confident level and  $n = 30$



degrees of freedom  $n - 1 = 10 - 1 = \underline{9}$



degrees of freedom  $n - 1 = 25 - 1 = \underline{24}$



$$1.00$$

$$\underline{-0.99}$$

$$\alpha = \underline{\underline{0.01}}$$

$$\alpha / 2 = \frac{0.01}{2}$$

$$\underline{\underline{0.005}}$$

### Activity-2

- The average daily production quantity of a factory is needed to have been estimated, using the data collected for 25 days. The average daily production has been computed as 100 units. The factory owner knows in his experience that the variance of daily production is 36. The following expression has been derived, once the daily average production is estimated under 95% confidence level.

$$100 \pm 1.96 \times \frac{6}{5}$$

- Derive the answers for following questions using this expression.
  - Simply the right hand side part of this expression.
  - Derive  $100 - 2.352$  and  $100 + 2.352$ ,  $97.648$  and  $102.352$
  - Derive a possible range of values in which the population mean may contain  
 $\Pr(97.648 \leq \mu \leq 102.352) = 95\%$
  - Explain what is meant by the right hand side value 2.352 of the above expression.  
 This is the probable error.
  - Explain the eligibility of interval estimation rather than point estimation

### Answers :

- Not only the value of estimation (the estimate) but also the sample size, population distribution and confidence level are considered in an interval estimation.
- The confidence of the parameter included in the interval also being contained.

### A Guideline to explain the subject matters:

- Interval estimation means that deriving a range of values considering the point estimator, matching with its standard error, and confidence level with the purpose of estimating a range in which a population parameter is expected to be contained.
- In point estimation, only the value of the estimator is used to estimate the parameter. That means the parameter is estimated with a single numerical value.
- In interval estimation a range of values in which the parameter may be contained is derived and in addition to the value of the estimator.

- The probability distribution
- Confidence level
- Standard error of the estimator are also taken into consideration

As instances :

- When the variance of a normally distributed population is known, the population mean ( $\mu$ ) is estimated using a sample as
- In point estimation  $\mu$  is estimated using the sample means derived from the sample
- In interval estimation
  - Sample mean

- Standard error of  $\bar{X} = \frac{\sigma}{\sqrt{n}}$

- Table value of  $Z_{\alpha/2}$  under the given confidence level

- Hence by subtracting and adding the probable error derived as a product of table value for  $Z_{\alpha/2}$  and the standard error of  $\bar{X}$  from and to the sample mean  $\bar{X}$  the range of values for  $\mu$  is derived and that is called the confidence interval for  $\mu$ .
- The range of values in which the population parameter is expected to fall under a given confidence level, is known as a confidence interval.
- As an example let's suppose that minimum value of the interval derived for  $\theta$  under  $1 - \alpha$  confidence level is  $Y_1$  and the maximum value of that interval is  $Y_2$ . Then the range of values (confidence interval) in which  $\theta$  (parameter) may be contained can be expressed as follows.

$$(Y_1 \leq \theta \leq Y_2) \rightarrow (1 - \alpha)\%$$

- The probabilistic interval in which the values of  $Y_1$ ,  $Y_2$  and all the values in between those two values are contained is the confidence interval.
- The two end values of the confidence interval in which the population parameter is expected to be contained under a given confidence level are called 'confidence limits'.
- $Y_1$  and  $Y_2$  of the above instance are confidence limits.
- The level of probability related to the interval in which the population parameter is expected to be contained is known as the confidence level. That is normally taken as 90%, 95%, 98%, 99%, etc. in general the confidence level is indicated as  $(1 - \alpha) = 100\%$
- Suppose that a confidence interval derived at 95% for population mean  $\mu$  is as follows  $\Pr(50 \leq \mu \leq 55)95\%$

- This implies that, once we have derived all the possible same size samples from the population and constructed a confidence interval at each time, the population mean  $\mu$  can be expected to be covered by the estimated interval at 95% of the times.
- Since only one sample is used practically, falling the parameter in the estimated interval or not is uncertain, and that uncertainty is denoted by  $\alpha$
- Once the probability levels related to the range of values in which the population parameter is expected to be contained are stated as 0.90, 0.95, 0.98 etc. those are known as confidence co-efficient.
- In general the confidence co-efficient is denoted by  $(1 - \alpha)$ 
  - When the confidence co-efficient is 0.95,  $\alpha = 0.05$
  - When the confidence co-efficient is 0.90,  $\alpha = 0.10$
  - When the confidence co-efficient is 0.99,  $\alpha = 0.01$
- A general expression for estimating a population parameter in the case of interval estimation can be mentioned as follows.

|   |
|---|
| $\text{Estimate} \pm \text{table value} \times \text{standard error}$ |
|---|

- The value derived by **table value X standard error** is called the confidence error.
- Eligibility of interval estimation compared with point estimation can be stated as follows.
  - Unlike in point estimation, a confidence level where the parameter falling in the interval also being stated.
  - Not only the value of the estimate, but also the population distribution and confidence error being applied for estimating the parameter.

Competency 7.0 : Uses statistical Inference for making business decisions

Competency Level 7.8 : Uses confidence interval for estimating the population mean

No. of Periods : 12

**Learning outcomes :**

- Gives instances for situations where the population mean  $\mu$  is to be estimated.
- States in an expression how to construct a confidence interval to estimate the population mean  $\mu$ .
- Computes confidence limits for  $\mu$  using that expression.
- Explains the meaning of confidence levels computed for the parameter.
- Writes the sampling distribution of the sample means  $\bar{X}$  for small samples drawn from a population with unknown variance.
- Explains the characteristics of a t-distribution and the situations where t-distribution is applicable.
- Derives a range of values using t-distribution for  $\mu$ .
- Estimates confidence intervals for  $\mu$  based on large samples drawn from a normal population with unknown variance.
- Evaluates the reliability and precision of a confidence interval constructed for  $\mu$ .
- Constructs confidence intervals for the mean  $\mu$  of a non normal population with known variance.
- Constructs confidence intervals for the mean  $\mu$  of a non normal population with unknown variance.

**Instructions for Lesson Planing :**

Display the following expressions on the board.

1. The average length of iron nails is 4.98cm.
2. The average diameter of iron balls is 4.32 mm.
3. The average monthly salary of employees in a firm is Rs 5 000.
4. The number of defective units found in a production line of a particular factory is 10.

Hold a discussion highlighting the following facts.

- Point out that all the iron nails (the entire population) should be examined for finding the mean of iron nails mentioned in the first statement above.
- Point out that all the iron balls should be examined for finding the mean diameter of iron balls mentioned in the second statement.

- Explain further that the salaries of all the employees should be added and divided by the number of employees for finding the mean (average) salary.
- Explain that all the population units should be examined to find out that there are 10 defective units.
- Make the students aware that the task of examining the entire population is considerably difficult and complicated, so that it would be more appropriate practically to estimate population parameters using sample data.
- Accordingly, point out the fact that a confidence interval can be constructed for estimating the population mean  $\mu$  using a sample.
- Point out that the confidence limits for population mean  $\mu$  can be derived using the following formula.

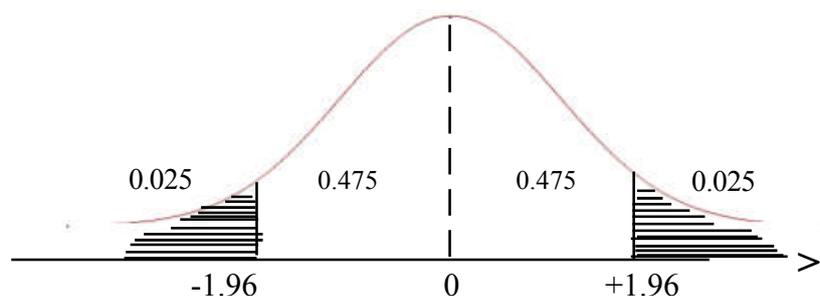
$$\mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

### Activity – 1

- Provide the following information to the students.
  - The average salary of 100 employees working in a firm has been computed as Rs. 25 000/-. The variance of the salaries of all the employees in the firm is Rs. 6 400. If the employee salaries of the firm are normally distributed a 95% confidence interval should have been built.
- Raise the following questions to the students based on the above mentioned details.
  - What is the sample mean of salaries ?
  - What is the tabled value related to 95% confidence level?
  - What is the standard error?
  - What is the probable error?
  - Build the 95% confidence interval.
  - Comment on the above answer you have already derived for (v) above?

### Solution – Activity – I

- Rs. 25 000/=



(ii) Confidence level

$$(1 - \alpha) = 95\%$$

$$\therefore \alpha = 1 - 0.95$$

$$\alpha = \underline{\underline{0.05}}$$

$$(iii) \quad \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{100}} = 8$$

$$(iv) \quad 1.96 \times 8 = 15.68$$

$$(v) \quad \mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$= 25000 \pm 1.96 \times \frac{80}{\sqrt{100}}$$

$$= \Pr(24984.32 \leq \mu \leq 25015.68) \rightarrow 95\%$$

(vi) Once all the possible samples with size 100 are drawn from the employees in this firm and 95% confidence intervals are constructed for population mean  $\mu$ , using the corresponding sample means, 95% of those confidence intervals will cover up the population mean  $\mu$

- Recalling that the sample variance  $S^2$  can be used as a good estimator for population variance  $\sigma^2$  in the absence of it, explain that the confidence intervals for population mean  $\mu$  can be constructed using the following formula, when large size samples are drawn from a normally distributed population of which the variance is unknown.

$$\mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

### Activity 2 :

- Involve the students in the following activity.

The average (mean) life time of a bulb in a random sample of 64 bulbs manufactured in an electric bulb manufacturing firm, has been computed as 226.6h and the standard deviation as 193.5h. Estimate the 99% confidence intervals for the average life time of a bulb manufactured by the firm using this sample.

**Solution for Activity 2 :**

$$\begin{aligned}\mu &= \bar{x} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}} \\ &= 226.6 \pm 2.57 \times \frac{193.5}{\sqrt{64}} \\ &= 226.6 \pm 62.17 \\ &= \text{Pr}(164.43 \leq \mu \leq 288.77) \rightarrow 99\%\end{aligned}$$

- Hence it can be stated that the average life time of a bulb falls in the range 164.43h to 288.77h with a 99% confidence.
- Hold a discussion highlighting the following facts.
  - Explain that the population mean  $\mu$  can be estimated using the 't' distribution, when the small size samples are drawn from a normal population with unknown variance.

$$\mu = \bar{x} \pm t_{\alpha/2; n-1} \cdot \frac{S}{\sqrt{n}}$$

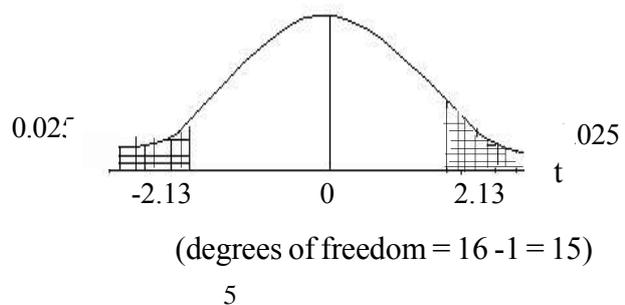
**Activity 3 :**

- Involve the students in the following activity.
  - A sample of 16 iron rods manufactured by a firm was drawn in random and the maximum bearable weight of each was observed. It has been computed  $\bar{x} = 27.3\text{kg}$  and  $S = 1.2\text{kg}$ . If the bearable mass of an iron rod falls normally construct a 95% confidence intervals for the population mean.

**Solution : Activity 3**

- Confidence intervals are computed as follows since the population falls normally, population variance is unknown and the sample size is small.

$$\begin{aligned}\mu &= \bar{X} \pm t_{\alpha/2; n-1} \cdot \frac{S}{\sqrt{n}} \\ &= 27.3 \pm 2.13 \times \frac{1.2}{\sqrt{16}} \\ &= 27.3 \pm 0.64 \\ &= 26.66\text{kg} - 27.94\text{kg} \\ &\underline{\underline{\text{Pr}(26.66 \leq \mu \leq 27.94) \rightarrow 95\%}}\end{aligned}$$



- Hence it can be mentioned that the average bearable weight of an iron rod falls in the range 26.64kg and 27.94kg with 95% confidence.
- If the size of a sample drawn out from a non-normal population is large since the Central Limit Theorem is applicable and following formula can be used for estimating population mean  $\mu$  once the population variance is known.

$$\mu = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

#### Activity – 4

- Provide with the following details and lead the students to construct confidence intervals for population mean.

The variance of the mass of a packet of biscuits produced in a firm has been computed as 36g. The mean mass of a sample of 100 packets of biscuits drawn in random was found as 395g. Compute 95% confidence limits for the mean mass  $\mu$  of a packet of biscuits produced in this firm.

#### Solution – Activity – 4

$$\begin{aligned} \mu &= \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\ &= 395 \pm 1.96 \cdot \frac{6}{\sqrt{100}} \\ &= 395 \pm 1.18 \\ &= 393.82 - 396.18 \\ &= \underline{\underline{\text{Pr}(393.82 \leq \mu \leq 396.18) \rightarrow 95\%}} \end{aligned}$$

- Explain that the Central Limit Theorem is applicable when large size samples are drawn out from a non normal population and the sample variance can be applied as a good estimator in the absence of the population variance.
- Accordingly point out that the following formula can be used for constructing confidence intervals for the population mean  $\mu$  of a non normal population whose variance is unknown using the large size samples.

$$\mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

### Activity – 5

- Provide the following details to the students.

Once a sample of 64 bottles of water filled by a business firm involved in water bottling was examined and the average volume of water contained in a bottle in that sample was computed as 998ml and the variance as 25ml.

Lead the students to construct a 99% confidence interval for the average volume of water containable in a bottle filled by this firm.

### Solution – Activity – 5

$$\begin{aligned}\mu &= \bar{X} \pm Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \\ &= 998 \pm 2.57 \cdot \frac{5}{\sqrt{64}} \\ &= 998 \pm 1.61 \\ &= 996.39ml - 999.61ml \\ &= \underline{\underline{\Pr(996.39 \leq \mu \leq 999.61) \rightarrow 99\%}}\end{aligned}$$

- Emphasize that although the confidence intervals can be constructed for the mean  $\mu$  of a non normal population when the sample size is large whether the population variance is known or not, the confidence intervals cannot be constructed when the sample size is small.
- Involve the students in the following activity to make them aware of the relationship between the width of a confidence interval and the confidence level.

### Activity – 6

Construct the confidence intervals in accordance with the details in Activity – I with 90% and 99% confidence. Compute the width of each confidence interval.

### Solution – Activity – 6

$$\begin{aligned}\mu &= \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} && \text{90\% confidence level} \\ &= 25000 \pm 1.64 \cdot \frac{80}{\sqrt{100}} \times \text{with 90\%} \\ &= 25000 \pm 13.12 \\ &= 24986.88 - 25013.12 \\ &= (24986.88 \leq \mu \leq 25013.12) \rightarrow 90\%\end{aligned}$$

width of the confidence interval Rs. 26.24

99% confidence level

$$\begin{aligned}\mu &= \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\ &= 25000 \pm 2.57 \frac{80}{\sqrt{100}} \\ &= 25000 \pm 20.56 \\ &= 24979.44 - 25020.56 \\ &= (24979.44 \leq \mu \leq 25020.56) \rightarrow 99\%\end{aligned}$$

width of the confidence interval Rs. 41.12

- Accordingly ascertain the fact that the width of the confidence interval increases, when the confidence level increases.
- Explain that the precision of the estimator goes down because it is more deviated from the actual mean, when the width of the confidence intervals is greater.
- Involve the students in the following activity to explain how the width of the confidence interval changes with reaction to the change of the sample size.

### Activity – 7

In accordance with the activity 1 above construct the confidence intervals and find the width only when the sample size changes as follows.

(i)  $n = 36$

(ii)  $n = 64$

### Solution – Activity – 7

(i) When  $n = 36$

$$\begin{aligned}\mu &= \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\ &= 25000 \pm 1.96 \times \frac{80}{\sqrt{36}} \\ &= 24973.87 - 25026.13 \\ &= (24973.87 \leq \mu \leq 25026.13) \rightarrow 95\%\end{aligned}$$

The width of the interval is Rs. 52.26

(ii) When  $n = 64$

$$\begin{aligned} \mu &= \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\ &= 25000 \pm 1.96 \times \frac{80}{\sqrt{64}} \\ &= 24980.40 \leq \mu \leq 25019.60 \\ &= (24980.40 \leq \mu \leq 25019.60) \rightarrow 95\% \end{aligned}$$

The width of the interval is Rs. 39.20

Hence point out the fact that width of the confidence interval decreases when the sample size increases and the precision of the estimator becomes greater.

**A guideline to explain the subject matters :**

- The population mean  $\mu$  can be estimated using a sample. A confidence interval can be constructed for this purpose as,

$$\mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- If  $X_1, X_2, X_3, X_4, \dots, X_n$  is a normally distributed population with a random variable  $X$ , the confidence interval for population mean  $\mu$  under a particular confidence level can be estimated using a sample with the size –  $n$ , in this manner.
- Once a confidence interval has been constructed under a particular confidence level using a sample of size,  $n$  it means that out of all the possible intervals that can be estimated under that confidence level using all the possible samples with size  $n$ , the population mean  $\mu$  is covered by that percentage of intervals stated by the confidence level (95%. 99% ...etc...)
- The formulae that can be applied in each situation to construct the confidence intervals for population mean  $\mu$  are as follows .
- Once the variance is known of a normal population.

$$\left. \begin{array}{l} n > 30 \\ n < 30 \end{array} \right\} \mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- When the population variance of a normal population is unknown.

$$(s^2) \begin{cases} n > 30 & \rightarrow \mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \\ n < 30 & \rightarrow \mu = \bar{X} \pm t_{\alpha/2; n-1} \cdot \frac{S}{\sqrt{n}} \end{cases}$$

- When the population variance of a non-normal population is known.

$$(\sigma^2) \rightarrow n > 30 \rightarrow \mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- When the population variance of a non-normal population is unknown.

$$(S^2) \rightarrow n > 30 \rightarrow \mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

- Estimation of confidence intervals using small size samples drawn from non-normal population with unknown variance is not taken into consideration.
- Once the confidence level is at a greater value keeping sample size (n) and standard deviation held constant, the width of the confidence interval becomes greater. Then the reliability of the confidence interval constructed for  $\mu$  is greater, but the precision is less.
- Confidence level and the precision of the level interval are inversely related.
- Confidence level and the width of the confidence interval are positively related.
- If the sample size is increased keeping the confidence level and the standard deviation held constant, the width of the confidence interval decreases and vice versa.
- Sample size and the width of the confidence interval are inversely related.
- Sample size and the precision of the estimator are positively related.
- In order to estimate a highly reliable and highly precise confidence interval a greater confidence level and a large sample should be used.

**Competency 7.0 :** Uses statistical Inference for making business decisions

**Competency Level 7.9 :** Constructs the confidence intervals for the difference between two population means

**No. of Periods :** 10

**Learning outcomes :**

- Explains the requirement of estimating the difference between the means of two populations.
- Constructs the confidence intervals for the difference between the means of two normal populations with known variances.
- Constructs and interprets the confidence intervals for the means of two non-normal populations with known variances.
- Once the population variances are unknown but equal, the pooled variance is derived using sample variances.
- Computes the confidence intervals for the difference between the means of two normal populations applying the pooled variance using the t-distribution.
- Computes the confidence limits for the difference between means of two normal populations with unknown variances using the t-distribution.
- Constructs the confidence limits for the difference between the means of two non-normal populations with unknown variances, applying the Central Limit Theorem.

**Instructions for Lesson Planning :**

- Present the following statement to the students.  
“A trade union is in need of making an estimation with a 95% confidence for the difference of employee wages in firm A having 1 500 workers and in firm B having 2 300 workers.”
- Hold a discussion with the students regarding the above statement highlighting the following facts.
  - Explain the fact that in order to derive the difference between the employee wages of the two firms A and B, difference of the wages between every employee in the firm A and firm B should be separately computed.
  - Point out that it is not practical and therefore an estimation can be made based on the difference between the wages of two samples of the employees in the two firms A and B.
  - Explain further that such estimations can be made under various confidence levels.

- Explain that, when estimations are made for the difference between the means of two normal populations with known variance, the product of z-value related to the confidence level and the probable error should be adjusted to the difference between two sample means. Provide with the following formula as well.

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

### Activity – 01

- Assign the following activity to the students.
- With the purpose of estimating the difference between the average wages of two firms called X and Y, a sample of 100 employees in firm X and a sample of 125 employees in firm Y were checked up. The average monthly income of those samples drawn from the firms X and Y have been computed as Rs. 55 000 and Rs. 45 000 respectively. Both the populations are normally distributed with the variance 1 000 000 and 703 125 respectively. Estimate 95% confidence intervals for the difference between the mean salaries of two firms and interpret.

### Solution – Activity – 01

$$\begin{aligned} &= (\bar{x} - \bar{y}) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \\ &= (55000 - 45000) \pm 1.96 \times \sqrt{\frac{1000000}{100} + \frac{703125}{125}} \\ &= 10000 \pm 245 \\ &= Rs.9755 - Rs.10245 \\ &= Rs.(9755 \leq \mu_1 - \mu_2 \leq 10245) \rightarrow 95\% \end{aligned}$$

### Activity – 2

Involve the students in following activity.

A wholesaler purchases a particular electric appliance from two companies A and B. The standard deviations of the life time of these appliances are 150h and 125h respectively. Once the average life time was computed taking two samples of 100 units in each they were as 1 500h and 1 450h respectively. The wholeseller is in need of estimating the difference of the average life time of these two appliances with 99% confidence.

- Are the population distributions of the lifetimes of A and B electric appliances known?
- What is the sampling distribution of the difference between the mean lifetimes of A and B electric appliances?

- Find the table value related to 99% confidence level.
- Estimate the difference between the average lifetime of the two populations using the following formula.

$$\mu_{A-B} = \bar{X}_A - \bar{X}_B \pm Z_{\alpha/2} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

- Interpret the answer you derived further.

### Solution – Activity – 2

- Population distributions are unknown.

$$= (\bar{X}_A - \bar{X}_B) \sim N\left(\mu_A - \mu_B, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\right)$$

- (According to Central Limit Theorem)

$$Z_{\alpha/2} = 2.57$$

$$\begin{aligned} \mu_A - \mu_B &= (\bar{X}_A - \bar{X}_B) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \\ &= (1500 - 1450) \pm 2.57 \sqrt{\frac{150^2}{100} + \frac{125^2}{100}} \\ &= 50 \pm 50.19 \\ &= (-0.19 \leq \mu_A - \mu_B \leq 50.19) \rightarrow 99\% \end{aligned}$$

- If all the possible samples with size 100 are drawn from both the populations and the confidence intervals are estimated for  $\mu_1 - \mu_2$  99% of those confidence intervals will cover (contain with) the true  $\mu_1 - \mu_2$

### Activity – 3

- Provide the following problem to the students and involve them in the exercise.

A researcher is in need of being comprehended whether the average expenses of families consist of 5 (five) members in two districts are significant. The variances of the expenses are unknown, but those variances are assumed to be equal. The expenses are known to be distributed normally. Two samples with size 25 in each were drawn in random from both the districts and the following unbiased estimates for mean and variance have been derived.

|           | District - 1 | District - 2 |
|-----------|--------------|--------------|
| $\bar{X}$ | 5 200        | 4 800        |
| $S^2$     | 500          | 600          |

- The fact whether there is a significant difference between the means of the expenses in two districts should have been examined with a confidence level of 95%
  - Considering two districts as 1 and 2 state the population distribution of expenses.
  - Derive the pooled variance using the two sample variances applying in the following formula.

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- Get the square root of this pooled variance.
- Write the sampling distribution of the difference of mean expenses of the two districts.
- Construct the confidence intervals for the difference between the mean expenses of two districts, using the following formula.

$$\mu_1 - \mu_2 = \bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2; n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- Further interpret the answer you derived.

### Solutions - Activity 03

$$(1) \quad X_1 \sim N(\mu_1, \sigma_1^2)$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

$$(2) \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(25-1) \times 500 + (25-1) \times 600}{25+25-2}$$

$$= \frac{12000 + 14400}{48}$$

$$= \underline{\underline{550}}$$

$$(3) \quad = \sqrt{550} = \underline{\underline{23.45}}$$

$$(4) \quad \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1 + n_2 - 2}$$

$$(5) \quad (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2; n_1 + n_2 - 2} Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$400 \pm 1.96 \times 23.45 \sqrt{\frac{1}{25} + \frac{1}{25}}$$

$$= 400 \pm 12.87$$

$$= \underline{\underline{\Pr(387.13 \leq \mu_1 - \mu_2 \leq 412.87) \rightarrow 95\%}}$$

- Once all the possible samples with size 25 are drawn out from both the districts and confidence intervals are constructed true value of  $\mu_1 - \mu_2$  will be included in 95% of such intervals. Since '0' is not included in this interval the means can be said to be significant with 95% confidence.

#### Activity – 4

- Involve the students in the following activity.
- A wholeseller purchases a particular product from the two firms A and B. Once two samples with size 100 are drawn from each firm the mean weight of a sample has been computed as 110g and 98g. He is in need of finding the difference between the population means with 95% confidence.
- Are the population distributions of the weight of the products A and B known?
- What's the sampling distribution of the difference between the weight of A and B products?
- What are the assumptions made for that purpose?
- Estimate the difference between the weight of two populations using the following formula.

$$\mu_A - \mu_B = (\bar{X}_A - \bar{X}_B) \pm Z_{\alpha/2} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}$$

- Further interpret the answer you derived.

**Solution – Activity – 04**

- Distributions of the populations are unknown.

$$(\bar{X}_A - \bar{X}_B) \sim N\left(\mu_A - \mu_B, \frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)$$

- The Central Limit Theorem.

$$\begin{aligned} \mu_A - \mu_B &= (\bar{X}_A - \bar{X}_B) \pm Z_{\alpha/2} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \\ &= 110 - 98 \pm 1.96 \sqrt{\frac{16}{100} + \frac{15.21}{100}} \\ &= 12 \pm 1.96 \sqrt{\frac{31.21}{100}} \\ &= 1.2 \pm 1.96 \times 0.558 \\ &= 12 \pm 1.09 \\ &= \underline{\underline{(10.91 \leq \mu_A - \mu_B \leq 13.09) \rightarrow 95\%}} \end{aligned}$$

- It means that 95% of the intervals constructed for  $\mu_1 - \mu_2$  using all the possible samples with 100 units taken from each firm will include the true value of  $\mu_1 - \mu_2$

**A Guidline to explain the subject matters :**

- Deriving the difference between the means of two populations is practically complicated, so that it should be estimated based on sample means.
- Following formula is used to construct the confidence intervals for the difference between the means of two normal populations with known variances.

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Following formula can be used to construct the confidence intervals for the difference between the means of two non normal populations with known variance based on the Central Limit Theorem.

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Once the populations variances are unknown, but they can be assumed to be equal, the confidence intervals for the difference between the means of two normal populations can be derived using the pooled variance of the sample variances as follows.

$$\mu_1 - \mu_2 = \bar{X}_A - \bar{X}_B \pm t_{n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- The pooled variance  $S_p^2$  is computed as follows.

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- Confidence intervals for the difference between the means of two non-normal populations with unknown variances can be constructed using the following formula supported by the Central limit theorem.

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

**Competency 7.0** : Uses Statistical Inference for making business decisions

**Competency Level 7.10** : Uses confidence intervals for estimating the population proportion

**No. of Periods** : 06

**Learning outcomes :**

- Explains what the estimation of confidence intervals for population proportion is.
- Explains the requirement of constructing confidence intervals for population proportion.
- Computes the confidence interval for population proportion using the normal distribution.
- Interprets a confidence interval constructed for population proportion.
- Solves problem related to confidence intervals for population proportion.

**Instructions for Lesson Planning :**

- Divide the students in the class into three separate groups and assign each of the following situations.
  1. A 95% confidence is needed regarding the percentage of votes expected by the presidential election candidate – A
  2. The Paddy Marketing Board is in need of a 95% assurance regarding the percentage of seedless paddy (thick / empty paddy only with chaff) contained in the paddy harvest in Yala season.
  3. A 99% assurance is expected by the management regarding the proportion of goods manufactured in a firm that has achieved at the expected standard level.
- Raise the following questions to the groups of students in general.
  1. To which population parameter is related with the problem you have received.
  2. What is the unbiased estimator for that parameter?
  3. Mention the steps that you would follow to derive an unbiased estimate for the relevant parameter using that unbiased estimator.
  4. State the standard error of the sampling distribution of the relevant estimator using the standard notation.
  5. Present your findings before the class creatively by one of the members in your group.
- Hold a discussion highlighting the following facts taking the details presented by the groups into consideration.
  - That the estimation of confidence intervals is required for population proportion using the sample elements in such situations.
  - That deriving an estimate using the sample proportion as a point estimator is required in order to construct confidence intervals for the population proportion.

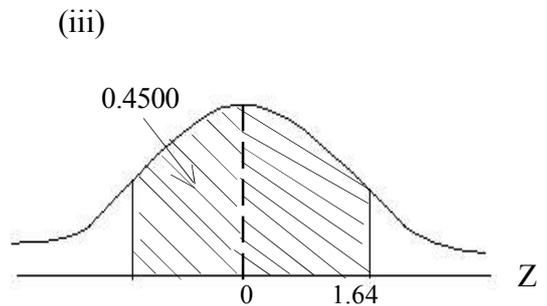
- That  $\sqrt{\frac{\pi(1-\pi)}{n}}$  is the standard error of the sampling distribution of sample proportion.
- That the confidence intervals for population proportion can be estimated using the unbiased estimator, estimate, required confidence level and the standard error of the estimator.
- Mention that the following data have been collected through sample surveys conducted related to the situations assigned on each group.
  1. In a random sample of 10 000 prospective voters registered for voting at presidential election 5 400 have revealed that they would cast their vote for the candidate named A
  2. In a random sample of 5 000 Kg of paddy taken from the paddy harvest in Yala season a quantity of 50 Kg has been found with seedless paddy (with chaff) .
  3. In a random sample of 200 units produced in the factory 20 units have been found that do not meet the expected quality standards.
- Provide the following instructions to the groups.
  - Derive an unbiased estimate for the population proportion considering the sample details related to the situation you have received.
  - Compute the standard error of the estimate.
  - Derive the standard normal value (z -value) relevant to the confidence level at which the estimate should be assured.
  - Derive the margin of error (probable error) as a product of the said z-value and the standard error.
  - Derive the Lower Confidence Level subtracting the margin of error from the unbiased estimate.
  - Derive the Upper Confidence Level adding the margin of error to the unbiased estimate.
  - Write the confidence interval.
  - Make an interpretation on your confidence interval.

**Solution :**

- (1) (i) Sample proportion P can be computed as an unbiased estimate for population proportion

$$\pi = \frac{5400}{10000} = 0.54$$

$$\begin{aligned}
 \text{(ii)} \quad \sigma_p &= \sqrt{\frac{P(1-P)}{n}} \rightarrow \text{for } \sqrt{\frac{\pi(1-\pi)}{n}} \\
 &= \sqrt{\frac{0.54 \times 0.46}{10000}} \\
 &= \underline{\underline{0.005}}
 \end{aligned}$$



$$\begin{aligned}
 \text{(iv)} \quad Z_{\alpha/2} \cdot \sqrt{\frac{P(1-P)}{n}} \\
 = 1.64 \times 0.005 \\
 = \underline{\underline{0.0082}}
 \end{aligned}$$

$$\begin{aligned}
 \text{(v)} \quad &= 0.54 - 0.0082 \\
 &= \underline{\underline{0.5318}}
 \end{aligned}$$

$$\text{(vii)} \quad 0.5318 - 0.5482$$

$$\begin{aligned}
 \text{(vi)} \quad &0.54 + 0.0082 \\
 &= \underline{\underline{0.5482}}
 \end{aligned}$$

$$\therefore \underline{\underline{(0.5318 \leq \pi \leq 0.5482) = 90\%}}$$

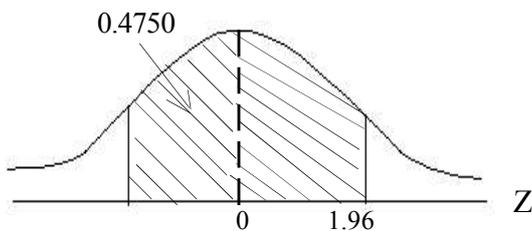
(viii) If all the possible samples are drawn with the size 10 000 units and construct all the confidence intervals using the sample proportion of each sample the population proportion will be covered by 90% of those confidence intervals.

(2)

$$\text{(i)} \quad P = \frac{50}{5000} = \underline{\underline{0.01}}$$

$$\text{(ii)} \quad \sigma_p = \sqrt{\frac{P(1-P)}{n}} \rightarrow \text{for } \sqrt{\frac{\pi(1-\pi)}{n}}$$

(iii)



$$\sigma_p = \sqrt{\frac{0.01 \times 0.99}{5000}} = \underline{\underline{0.0014}}$$

$$\begin{aligned}
 \text{(iv)} \quad Z_{\alpha/2} \cdot \sqrt{\frac{P(1-P)}{n}} \\
 = 1.96 \times 0.0014 \\
 = \underline{\underline{0.0027}}
 \end{aligned}$$

$$\begin{aligned}
 \text{(v)} \quad &0.01 - 0.0027 \\
 &= \underline{\underline{0.0073}}
 \end{aligned}$$

$$(vi) \quad 0.01 + 0.0027$$

$$= \underline{\underline{0.0127}}$$

$$(vii) \quad 0.0073 + 0.0127$$

$$= \therefore \underline{\underline{(0.0073 \leq \pi \leq 0.0127) \rightarrow 95\%}}$$

It is 95% confident that the population proportion will fall in the range 0.73% and 1.27%. That means if all the possible samples with size 5 000Kg are drawn and construct all the possible confidence intervals the population proportion will be covered by 95% of those intervals.

$$3. (i) \quad \frac{10}{200} = 0.05$$

$$(ii) \quad \sigma_p = \sqrt{\frac{P(1-P)}{n}} \text{ for } \sqrt{\frac{\pi(1-\pi)}{n}}$$

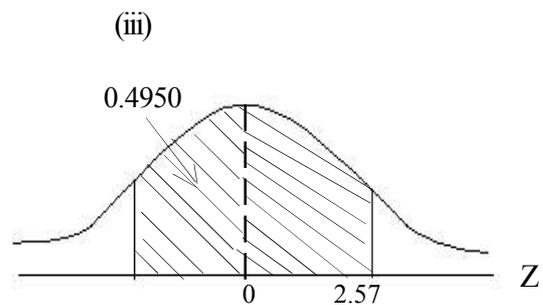
$$= \sqrt{\frac{0.05 \times 0.95}{200}}$$

$$= \underline{\underline{0.015}}$$

$$(iv) \quad Z_{\alpha/2} \cdot \sqrt{\frac{P(1-P)}{n}}$$

$$= 2.57 \times 0.015$$

$$= \underline{\underline{0.039}}$$



$$(v) \quad 0.05 - 0.039 = 0.011$$

$$(vi) \quad 0.05 + 0.039 = 0.089$$

$$(vii) \quad \underline{\underline{0.011 \leq \pi \leq 0.089 \rightarrow 99\%}}$$

(viii) It can be declared with 99% confidence that the proportion of the production that does not meet the expected quality level is between 1.1% and 8.9%.

That means if all the possible samples with the size 200 are drawn in random and constructed all the possible confidence intervals 99% of those intervals will cover up the population proportion  $\pi$

**A guideline to explain the subject matters.**

- Estimating of an interval of values in which the population proportion may possible to fall under a given confidence level is known as construction of confidence intervals for population proportion.
- The unbiased estimator for population proportion  $\pi$  is the sample proportion – P.
- Even though the standard error of the sample proportions is  $\sqrt{\frac{\pi(1-\pi)}{n}}$ , the estimated standard error is given by  $\sqrt{\frac{P(1-P)}{n}}$  because  $\pi$  is estimated by the sample proportion-P.
- Since the confidence intervals for population proportion are derived using large size samples, the sampling proportions is approximated to a normal distribution, so that the standard normal distribution can be used for constructing confidence intervals.
- Confidence intervals for population proportion  $\pi$  can be derived as follows.

$$\pi = p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

**Competency 7.0** : Uses statistical Inference for marking business decisions

**Competency Level 7.11** : Uses confidence intervals for estimating the difference between two population proportions

**No. of Periods** : 06

**Learning outcomes :**

- Points out the requirement of a confidence interval for the difference between two population proportions.
- Constructs the confidence intervals for the difference between two population proportions.
- Solves practical problems using the confidence intervals for the difference between two population proportions.

**Instructions for Lesson Planning :**

- Hold a discussion with students regarding the following matter.
  - A researcher is in need of knowing whether there is a significant difference between the proportions of school drop outs among the students in schooling age in urban and rural areas.
  - Explain the procedure to be followed with this regard considering that you are the researcher.
- Ascertain that two samples including schooling age children should have been selected in random from urban and rural areas,
- Provide with instructions that samples with size 100 or more should be selected.
- Ascertain that the proportion of the children who are not attending school in each sample among schooling age children should be found separately.
- Point out that the difference between two sample proportions should be found.
- Ascertain that the confidence interval should be constructed under an appropriate confidence level using the following formula.

Estimate  $\pm$  table value  $\times$  standard error

$$P_1 - P_2 \pm Z_{\alpha/2} \times \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

- Get some more situations revealed, where the difference between two population proportions should be found.

### Activity – I

- Provide with the following problems to the students and lead them to answer the questions given.
- It has been revealed in a survey that 264 out of 400 male voters and 180 out of 300 female voters are in favour of a particular election candidate.
- A 95% confidence interval for the difference between the proportions of male and female who are in favour of this candidate should be estimated. Answer the following questions with regard to this mater.
  - (i) What is the male proportion in favour of the candidate?
  - (ii) What is the female proportion in favour of the candidate?
  - (iii) What is the difference between male and female proportions in favour of this candidate?
  - (iv) Considering male M and female F derive the variance in accordance with the following formula.

$$\frac{P_m (1 - P_m)}{n_m} + \frac{P_F (1 - P_F)}{n_F}$$

- (v) Derive the z-value relevant to 95% confidence level.
- (vi) Substitute your findings in the following formula.

$$P_M - P_F \pm Z_{\alpha/2} \sqrt{\frac{P_M (1 - P_M)}{n_M} + \frac{P_F (1 - P_F)}{n_F}}$$

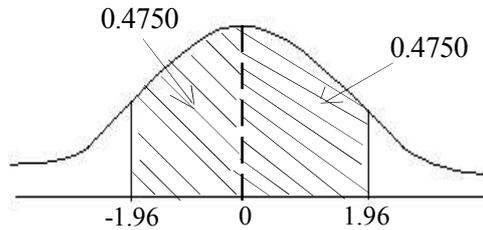
- (vii) State your conclusion regarding the difference between the male and female proportions in favour of this candidate.

### Solution – Activity – I

$$(i) \quad \frac{264}{400} = \underline{\underline{0.66}} \qquad (ii) \quad \frac{180}{300} = \underline{\underline{0.60}} \qquad (iii) \quad 0.66 - 0.60 = \underline{\underline{0.06}}$$

$$(iv) \quad \frac{0.66 \times 0.34}{400} + \frac{0.6 \times 0.4}{400} \\ = 0.000561 + 0.0008 \\ = \underline{\underline{0.001361}}$$

- (v) Value of  $Z_{\alpha/2}$  related to 95% confidence level



$$Z_{\alpha/2} = \underline{\underline{1.96}}$$

$$\begin{aligned}
 \text{(vi)} \quad & P_M - P_F \pm Z_{\alpha/2} \sqrt{\frac{P_M(1-P_M)}{n_M} + \frac{P_F(1-P_F)}{n_F}} \\
 & = 0.06 \pm 1.96 \sqrt{0.001361} \\
 & = 0.06 \pm 0.072 \\
 & = \underline{\underline{-0.012 - 0.132}}
 \end{aligned}$$

- The difference between the proportions of male and female in the population, in favour of the candidate can be said to be between -12% and 13.2% with 98% confidence. Further since this range contains zero (0), there is a possibility for these proportions not being significant as well.

### Activity – 2

- Provide the following problem to the students.  
Once a random sample of 200 nails produced in machine A and a random sample of 200 nails produced in machine B were examined it was found that 15 nails and 7 nails were defective respectively. According to these details can you definitely mention that there is a significant difference in the products of the two machines? (consider a 90% confidence level).

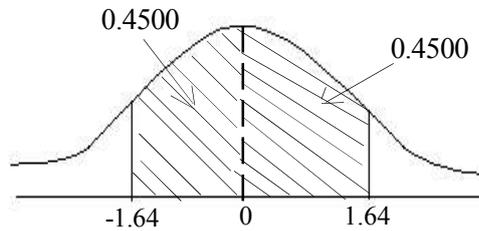
### Solution – Activtiy – 2

$\pi_1 - \pi_2$  Should have been estimated.

$$P_A = \frac{15}{200} = 0.075$$

$$P_B = \frac{7}{200} = 0.035$$

Confidence level = 90%



$$\begin{aligned}
 P_A - P_B \pm Z_{\alpha/2} \sqrt{\frac{P_A(1-P_A)}{n_A} + \frac{P_B(1-P_B)}{n_B}} \\
 = 0.04 \pm 1.64 \times \sqrt{\frac{0.075 \times 0.925}{200} + \frac{0.035 \times 0.965}{200}} \\
 = 0.04 \pm 1.64 \times \sqrt{0.00052} \\
 = 0.04 \pm 1.64 \times 0.023 \\
 = 0.04 \times 0.038 \\
 = 0.002 - 0.078 \\
 = (0.002 \leq \pi_A - \pi_B \leq 0.078) \rightarrow 90\%
 \end{aligned}$$

It can be mentioned that there is a significant difference between the defective proportions of the products with a 90% confidence.

**A guideline to explain the subject matters :**

- It is required to estimate the difference between two population proportions ( $\pi_1 - \pi_2$ ) using the values of the estimators  $P_1$  and  $P_2$  (the sample proportions) derived from the random samples with the size  $n_1$  and  $n_2$  drawn from two independent populations with parameters (proportions)  $\pi_1$  and  $\pi_2$

Ex amples :

- Once a whole seller is in need of getting to know the difference between defective percentages contained in products A and B.
- Once a physician is in need of getting to know whether there is any significance in the percentages of recovering a particular disease applying two alternative treatments.
- Once it is needed to come to know whether there is any significance between the proportions of children and adults who are in favour of a particular programme.

- Since sampling distributions of sample proportions related to the populations whose parameters (proportions) are  $\pi_1$  and  $\pi_2$ , fall in a binomial distribution, the sampling distribution of the difference between two sample proportions can be considered to fall approximately normally for large size samples as  $n_1 \geq 100$  and  $n_2 \geq 100$ . Accordingly the following formula can be used to derive confidence intervals for the difference between two sample proportions. Once the populations are denoted as A and B the confidence intervals of  $(1 - \alpha)$  100% can be derived by

$$P_1 - P_2 \pm Z_{\alpha/2} \times \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

**Competency 8.0** : Uses statistical hypothesis testing for making business decisions

**Competency Level 8.1** : Studies the concepts related to statistical hypothesis testing

**No. of Periods** : 10

**Learning outcomes :**

- Introduces what ‘a hypothesis’ is.
- Explains ‘statistical hypothesis testing’.
- Points out the need of statistical hypothesis testing.
- Interprets the ‘null hypothesis ‘ and ‘ alternative hypothesis’ used in hypothesis testing.
- Differentiates between a simple hypothesis and a composite hypothesis.
- Explains the type I and the type II error.
- Introduces the level of significance (the size) of a hypothesis testing.
- Explains the power of the test and points out how to compute it.
- Introduces the ‘test statistics ‘ and explains how to compute it.
- Explains the single tail test and the two tail test.
- Explains the ‘left tail test’ and the ‘right tail test’?
- Introduces the critical value and points out how to partition the critical region and the acceptance region.
- Explains ‘P-value’ and level of significance.
- Lists out the steps of hypothesis testing.

**Guidelines for Lesson Planning :**

- Hold a discussion with the students as follows.  
“ A Particular bakery owner states that the mass of a loaf of bread produced in his bakery is 450g”  
Suppose that the consumer services authority is in need of examining the truthfulness of this statement.
- Point out the fact that the assumption that the mass of a loaf of bread produced in that factory is 450 g is true or false can be tested using a sample of bread drawn from this bakery.
- Explain to the students that the average (mean) mass of a loaf of bread produced in this bakery is a population parameter and that the assumption which is built regarding that parameter is a statistical hypothesis and further that the rational process testing the truthfulness of that assumption is known as Statistical Hypothesis Testing.

- Explain that the hypothesis which is temporarily accepted to be true for conducting a hypothesis testing is known as the Null hypothesis and in any case of rejecting that Null hypothesis, the second hypothesis built for accepting is known as alternative hypothesis, further explain that the Null hypothesis is denoted by  $H_0$  where as the alternative hypothesis by  $H_1$ 
  - Provide with the following instruction to the student for building a hypothesis stating the value which is equal to the parameter in accordance with the given details.
  - Then read the problem again carefully and pay your attention to build another hypothesis against the Null hypothesis that has been already built.
  - If the null hypothesis is true, build the alternative hypothesis so as to be false.
  - If the null hypothesis is false, build the alternative hypothesis so as to be true.

### Activity – 01

Provide with the following situations to the students and guide them to construct the null hypothesis and alternative hypothesis suit for each situation.

1. The producer states that the average life time of a bulb manufactured in his factory is at least 2 500h.
2. The average quantity of a particular chemical to be contained in a drug is 0.005mg.
3. A bakery owner states that the average weight of a loaf of bread made in his bakery is at least 450 g.
4. A company involved in producing a battery states that the minimum life time of a battery is at least 5 years.
5. A firm involved in manufacturing nails and nuts states that the average diameter of a nut they produce is 5mm.
6. A company states that the average quantity of a particular chemical substance contained in an energetic soft drink bottle is not exceeding 5 ml.
7. A firm involved in manufacturing kids chairs states that the maximum height of a chair manufactured by them is no more than 45 cm.

### Solution Activity – 1

1. (i)  $H_0 : \mu = 2500$  (true)  
 $H_1 : \mu < 2500$  (false)
  - in order to build up of hypothesis in this manner, the given mean value 2 500 is applied to build the null hypothesis.

- If the mean value is equal to 2 500 the fact whether the producer’s claim will be true or falls should be logically considered secondly.
- After  $H_0$  is recognized to be true the hypothesis should be built in a manner that  $H_1$  is false. That means the producer’s claim will be falls when the average life time is less than 2500 h.

2.  $H_0 : \mu = 0.005$   
 $H_i : \mu \neq 0.005$

3.  $H_0 : \mu = 450$   
 $H_i : \mu < 450$

4.  $H_0 : \mu = 5$   
 $H_i : \mu < 0$

5.  $H_0 : \mu = 5mm$   
 $H_i : \mu \neq 5mm$

6.  $H_0 : \mu = 5ml$   
 $H_i : \mu > 5ml$

7.  $H_0 : \mu = 45cm$   
 $H_i : \mu > 545cm$

Involve the students in following activity to make them aware that, even though  $H_1$  hypothesis is build up considering that  $H_0$  is true, there are some situations where  $H_0$  also can be false.

### Activity – 02

- Guide the students to build hypothesis for following situations.
  1. The average life time of a bulb manufactured in a firm was 26 000 h. The firm states that the life time of this bulb is increased after a new filament was introduced.
  2. The average proportion of fats contained in 100g of milk flour produced by a firm where milk flour is manufactured was 12 g. The firm mentions that the proportion of fats contained in milk flour they produce has decreased due to application of modern technology.
  - 3 It has been revealed that the average paddy harvest yielded from a land of 1 hectare in extent was 3 200 Kg in the yala season from long term experience. A new species of pady and some other modern cultivating technologies were applied in the yala season. The fact whether there is any significance in the harvest in last yala season should have been examined.

### Solution – Activity – 1

- |  |   |
|--|---|
| 1. Null hypothesis<br>$H_0 : \mu = 26000$  | Alternative hypothesis<br>$H_1 : \mu > 26000$     |
| 2. Null hypothesis<br>$H_0 : \mu = 12\%$   | Alternative hypothesis<br>$H_1 : \mu < 12\%$      |
| 3. Null hypothesis<br>$H_0 : \mu = 3200Kg$ | Alternative hypothesis<br>$H_1 : \mu \neq 3200Kg$ |

- Point out that in the 1<sup>st</sup> and the 2<sup>nd</sup> problems the direction of the alternative hypothesis is clear since a clue is given in it.
- Further point out that in the 3<sup>rd</sup> problem the fact whether the forces implemented related to the paddy harvest would influence to increase the harvest or to decrease the harvest is not obvious well. Therefore alternative hypothesis has been built up with  $\neq$  sign.

### Activity – 3

- Provide the students with the following situations to make them aware of the simple hypothesis and composite hypothesis.
  1. The average distance that a tyre manufactured in the company A can run is expected to be 2 500km. It is known that this distribution falls normally with standard deviation 100km. The statement of the producer should have been verified by testing a random sample of 50 tyres drawn.
  2. The average distance that a tyre manufactured in the company B can run is expected to be 2 500 Km. The shape of this distribution is unknown and the variance is also unknown. The statement of the producer should have been verified by testing a random sample of 50 tyres drawn.
    - Highlight the differences appearing in these situations as far as possible.
    - Build null hypothesis and alternative hypothesis for both these situations.
- Discuss the following differences between these two situations separately.
  - The variation of the runnable distance of tyre produced in A company is known, but that of B is unknown.
  - The shape of the distribution of the runnable distance of a tyre produced by A company is known but that of B is unknown.
- The hypothesis for these two situations can be built as follows.

$$(1) \quad H_0 : \mu = 2500km$$

$$H_1 : \mu < 2500km$$

$$(2) \quad H_0 : \mu = 2500km$$

$$H_1 : \mu < 2500km$$

- Ascertain that  $H_0 : \mu = 2500km$  related to the factory A, is a simple hypothesis since all the details of that distribution such as variance, shape etc. are revealed.
- Ascertain that  $H_0 : \mu = 2500km$  related to B factory is a composite hypothesis since all the details of the distribution such as variance and the shape are unknown.

#### Activity – 4

- Display the following diagram before the class in order to explain the basic concepts in connection with Hypothesis testing.

| Decision              | Situation          |                     |
|-----------------------|--------------------|---------------------|
|                       | When $H_0$ is true | When $H_0$ is false |
| $H_0$ is rejected     | x                  | √                   |
| $H_0$ is not rejected | √                  | x                   |

- Name the two types of possible errors in a hypothesis testing.

When  $H_0$  is true in accordance with the population, if  $H_0$  is rejected on sample details, it would be an error and that is known as the type I – error.

When  $H_0$  is false in accordance with the population, if  $H_0$  is not rejected on sample details, it would also be an error and that is known as type II - error.

- Explain that the probability of rejecting the true  $H_0$  (probability of committing type – I error) is denoted by  $\alpha$  which is also called the size of the test or the level of significance of the test.
- Explain that the probability of not rejecting  $H_0$  when  $H_0$  is false (probability of type – II error) is denoted by  $\beta$
- Explain that the hypothesis test is conducted so as to minimize the type – II error  $\beta$  as far as possible keeping the type – I error ( $\alpha$ ) at a particular constant level. Point out that  $\alpha$  is determined in small values such as 0.01, 0.02, 0,05 etc.
- Explain the fact that when the probability of occurring type – II error  $\beta$  is minimized, the probability of not occurring type – II error. (making the correct decision -  $1 - \beta$  ) is maximized and that is known as the power of the test.

#### Activity – 5

- In order to introduce some more important concepts related to hypothesis testing provide with the following situation and involve the students in activity.
- An electric bulb manufacturer states that the average life time of a bulb he produces is 2 500h. The life time of a bulb is known to be distributed normally and also experienced that the standard deviation of the life time of bulbs is 600h. Once a random sample of 25 bulbs made by him was tested the average life time was computed as 2 200h.

Whether the statement of the producer is true should have been tested at  $\alpha = 0.05$

1. State the distribution of the lifetime, when considered the life time as X

$$X \sim N(2500, 600^2)$$

2. State the sampling distribution of sample means.

When the sample mean is  $\bar{X}$

$$\bar{X} \sim N\left(2500, \frac{600^2}{25}\right)$$

3. Discussing with students write the null hypothesis and the alternative hypothesis.

$$H_0 : \mu = 2500$$

$$H_1 : \mu < 2500$$

4. Construct a statistic using the details in the sample and population in order to test the null hypothesis.

Since  $X \sim N(\mu, \sigma^2)$  and

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

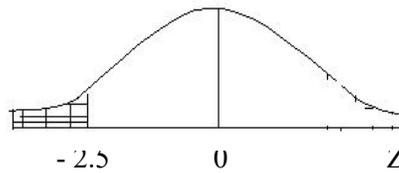
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

- Explain that this is known as the test statistic in hypothesis testing and that the value of test statistic can be computed for the given problem as follows.

$$\begin{aligned} Z &= \frac{2200 - 2500}{\frac{600}{5}} & \mu &= 2500 \\ &= \frac{-300}{120} & \bar{X} &= 2200 \\ &= \underline{\underline{-2.5}} & \sigma &= 600 \\ & & n &= 25 \end{aligned}$$

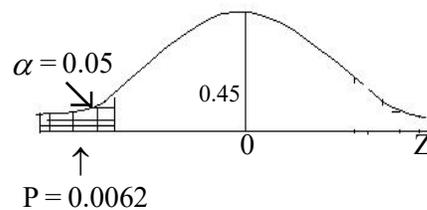
5. Derive  $\Pr(\bar{X} < 2200)$  using the value of test statistic.

$$\begin{aligned} \Pr(\bar{x} < 2200) &= \Pr(z < -2.5) \\ &= 0.5000 - 0.4938 \\ &= \underline{\underline{0.0062}} \end{aligned}$$



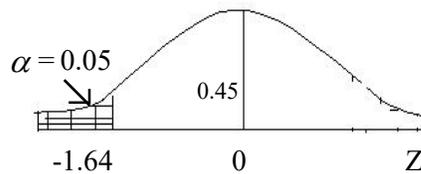
Explain that this is known as the P – value of a hypothesis testing. Further explain that the minimum probability level of  $H_0$  being rejected in accordance with the test statistic is known as the ‘P-value’.

6. Compare the  $\alpha$  Value and P – value computed above.



$\therefore$  P – value  $<$   $\alpha$  Value

7. Find the z-value related to  $\alpha = 0.05$  in the shaded area in following diagram.

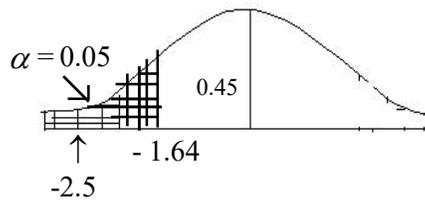


$Z = -1.64$  according to the standard normal distribution. Explain that the derived z-value under a given significance level is called the critical value in hypothesis testing.

- Explain that the shaded area in the above diagram is known as the critical region and the unshaded area is acceptance region.
  - Accordingly explain that the sampling distribution is partitioned in to two regions as critical region and acceptance region in hypothesis testing.
  - The value relevant for  $H_0$  to be rejected are contained in the critical region, whereas the values relevant for  $H_0$  not to be rejected are contained in the acceptance region.
8. Point out that the final conclusion of a hypothesis testing can be achieved in two ways either
- Comparing P-value and  $\alpha$  value or
  - Comparing critical value and the test statistic.
9. Write the possible decision to be made comparing the critical value and the test statistic according to the above problem.

The value of test statistic = -2.5

Critical value = -1.64



- Explain that  $H_0$  is decided to be rejected since the value of test statistic (-2.5) falls in the critical region.

10. Write the possible decision to be made comparing the P-value and the level of significance ( $\alpha$ )

$$P = 0.0062 \approx P = 0.01 \text{ and}$$

$$\alpha = 0.05$$

Since  $P\text{-value} < \alpha$  Value,  $H_0$  is rejected at 0.05 level of significance. Hence, it can be declared that the statement of the producer is not true at  $\alpha = 0.05$

11. Briefly explain the procedure that was followed in testing hypothesis related to the given problem.

Discuss the steps followed in testing hypothesis related to the given problem regarding the life time of electric bulbs.

(i) Construction hypothesis as;

$$H_0 : \mu = 2500$$

$$H_1 : \mu < 2500$$

(ii) Computig the value of test statistic as

$$\begin{aligned} Z &= \frac{2200 - 2500}{\frac{600}{5}} \\ &= \underline{\underline{-2.5}} \end{aligned}$$

(iii). deriving the P-value relevant to the critical region based on the value of test statistic.

$$P = 0.0062 \approx 0.01$$

(iv) Providing the critical value as at the level of significance given.

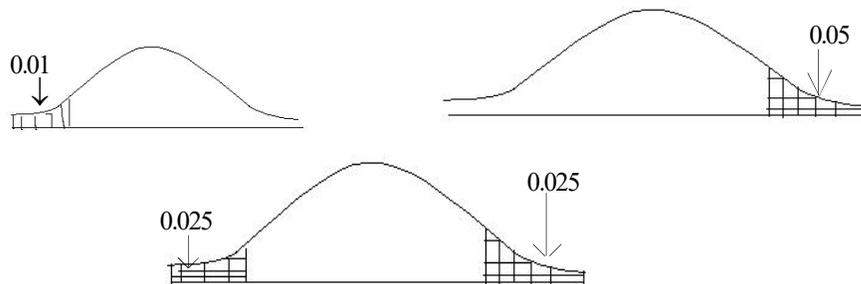
$$\alpha = 0.05 \text{ the critical value } - 1.64$$

5. Expression of the decision and conclusion, following either way.

- (i) comparing P-value and  $\alpha$  value
- (ii) comparing test statistic and critical value.

**Activity – 6**

Present the following diagrams and derive the answers for the questions given below.



Related to each diagram given above.

- (i) Write the probability of occurring type – I error
- (ii) Derive the critical values,
- (iii) State whether the hypothesis testing is a single tail test or a two tail test and if it is a single tail test state whether it is a left tail test or a right tail test.
- (iv) Explain two ways in which the decision of a hypothesis test can be expressed.

1. (i)  $\alpha = 0.01$

(ii)  $\alpha = 0.05$

(iii)  $\frac{\alpha}{2} = 0.025 \therefore \alpha = 0.05$

2. (i)  $Z = -2.33$       (ii)  $Z = 1.64$       (iii)  $Z = \pm 1.96$

3. (i) A single tail test and it is a left tail test.

(ii) a single tail test and it is a right tail test.

(iii) A two tail test

4. (i) Comparing P-value and  $\alpha$  value

(ii) comparing the test statistic and the critical value.

### A Guideline to explain the subject matters :

- The logical assumption which is built considering to be temporarily true regarding the population parameter is called ‘a statistical hypothesis. This can be stated in a statement or symbolically.
- Drawing a random sample from the relevant population and testing whether the statistical hypothesis is true or false in accordance with sample details is the statistical hypothesis testing.
- When the population related to the study
  - is infinite
  - consists of destructive elements in testing. or
  - Once the resource available for the study such as time, trained labour, funds etc... are not adequate.

The population parameters should have been estimated using the details of the sample drawn from the relevant population. The statistical hypothesis testing is a scientific technique which is applicable in such a situation.

- The hypothesized value (assumed value) for the unknown parameter is known as the null hypothesis and symbolized as  $H_0$
- If the null hypothesis is rejected the optimal hypothesis accepted instead is known as the alternative hypothesis and symbolized as  $H_1$
- Using the sample details it is the null hypothesis always rejected or not rejected. If the null hypothesis is decided to be rejected, then only the alternative hypothesis ( $H_1$ ) is accepted. Therefore it is very obvious that the decision of hypothesis testing is always made in connection with the null hypothesis  $H_0$ .

Ex :

The average life time of an electric bulb manufactured by a particular firm has been declared as 2500h.

Once the truthfulness of this statement is tested, the hypothesis can be built as follows.

$$H_0 : \mu = 2500$$

$$H_1 : \mu < 2500$$

- When the null hypothesis  $H_0$  is true if all the details regarding the population are concealed, that is known as a simple hypothesis.

Ex :

Once the variance of a normal population is known as  $\sigma^2 = 25$  the hypothesis whether the population mean  $\mu = 50$  should have been tested. Here  $\mu = 50$  is a simple hypothesis.

- When the null hypothesis  $H_0$  is true unless all the details regarding the population are concealed, that is known as a composite hypothesis.

Ex : Once the variance of a certain population is unknown the hypothesis whether the population mean  $\mu = 50$  should have been tested.

$\therefore H_0 : \mu = 50$  is a composite hypothesis.

- Most probably the null hypothesis for composite hypothesis is expressed as a range of values related to the parameter.

Ex :  $H_0 : \mu < 50$  "  $H_0 : \mu > 50$  or

$H_0 : \mu \neq 50$

- When the hypothesis testing is committed regarding population parameters two types of errors can occur.

**(i) Type – I error**

Rejecting the null hypothesis ( $H_0$ ) using sample details, once  $H_0$  is true in accordance with population details.

The probability of occurring type – I error is denoted by  $\alpha$  and that is also called the size of the test or the level of significance.

**(ii) Type – II error**

Not rejecting the null hypothesis ( $H_0$ ) using sample details once  $H_0$  is not true in accordance with the population details.

The probability of occurring type – II error is denoted by  $\beta$

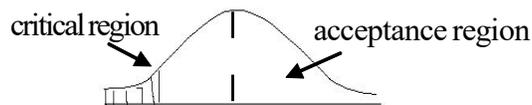
$P_r(\text{not rejecting } H_0 / H_0 \text{ is not (true)}) = \beta$

- In a hypothesis testing keeping the probability of occurring type – I error ( $\alpha$ ) at a particular constant level, the test is committed minimizing the probability of type – II error as far as possible.
- Normally the value of ( $\alpha$ ) is maintained at lower levels such as 0.1, 0.05, 0.01 etc.

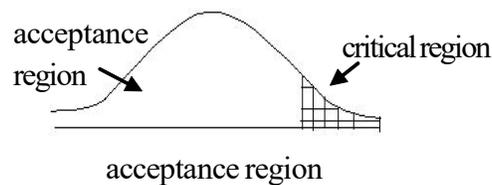
- In order to determine whether to accept the null hypothesis or not, the ‘statistic’ computed using the sample elements for comparing with the critical value is the test statistic.
- The value at the boundary, where the acceptance region and the critical region are partitioned as at any given level of significance is called the critical value.
  - The sampling distribution
  - Size of the test
  - The critical region

Should have been known to determine the critical value.
- Prior to make a decision, the criterion which that decision is made, should have been vividly comprehended.
- Since hypothesis testing is launched based on sample details, the possible space taken by a function of sample elements is bi-partitioned in accordance with a particular specification or a benchmark. In one of these regions  $H_0$  is not rejected.
- The region in which the values of the function of sample elements with relevance to reject the null hypothesis ( $H_0$ ) are contained, is known as the critical region in hypothesis testing.
- The region in which the values of the function of sample elements with relevance not to reject the null hypothesis ( $H_0$ ) are contained, is known as the acceptance region in hypothesis testing.
- The critical region is determined keeping probability of occurring type – I error  $\alpha$  at a small constant level like 0.01, 0.05, 0.10 etc..
- The direction of locating the critical region is determined on the direction led by the alternative hypothesis ( $H_1$ )

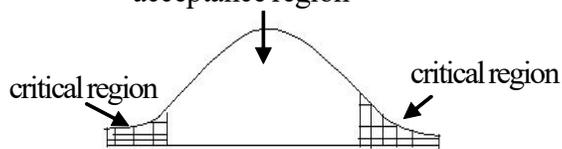
(i) Ex :  $H_0 : \mu = \theta$   
 $H_1 : \mu < \theta$



(ii) Ex :  $H_0 : \mu = \theta$   
 $H_1 : \mu > \theta$

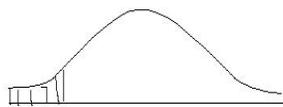


(iii) Ex :  $H_0 : \mu = \theta$   
 $H_1 : \mu \neq \theta$

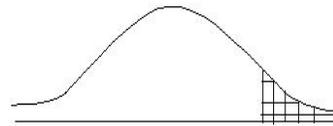


- If the values pointed by the alternative hypothesis are located in a single direction of the value given by the null hypothesis ( $H_0$ ), the hypothesis tests launched with that regard are single tail tests (one tail tests). One tail tests are of two types as left tail tests and right tail tests.
- If the values pointed by the alternative hypothesis are less than the values of the null hypothesis, the tests with regard to that direction are left tail tests and the values pointed out by the alternative hypothesis are greater than the value of the null hypothesis the tests with regard to that direction are right tail tests.

A left tail test

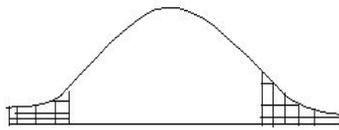


A right tail test



If the values pointed by the alternative hypothesis are located on either side of the value of the null hypothesis the tests launched with that regard are two tail tests.

A two tail test



- The minimum probability level possible for the null hypothesis being rejected in accordance with the value of the test statistic is known as the **p-value**.
- If the **p-value** is greater than the level of significance ( $\alpha$ -value) the null hypothesis  $H_0$  is not rejected.
- If the p-value is less than the level of significance ( $\alpha$ -value) the null hypothesis  $H_0$  is rejected.
- The steps mentioned below should be followed in hypothesis testing.
  - (i) Establishing the null hypothesis ( $H_0$ ) and the alternative hypothesis ( $H_1$ ) in accordance with the given population details.
  - (ii) Drawing a random sample from the relevant population, deriving the value of test statistic using both the relevant sample details and population details and hence computing the p-value.
  - (iii) Deriving the critical value according to the given level of significance.
  - (iv) Making (expressing) the decision and coming to the conclusion comparing the value of test statistic and the critical value or comparing the p-value and the level of significance ( $\alpha$ )

**Competency 8.0** : Uses statistical hypothesis testing for making business decisions

**Competency Level 8.2** : Uses hypothesis tests for the population mean

**No. of Periods** : 10

**Learning outcomes :**

- Explains Hypothesis testing for the mean of a normal population.
- Analyses the test statistic that is applicable in that case.
- Commits the standard normal test for hypothesis testing in connection with the mean of a normal population whose variance is known.
- Substitutes the sample standard deviation (S) for population standard deviation ( $\sigma$ ), in computing the test statistic, when the hypothesis related to the mean of a normal population with unknown variance is tested.
- Applies the t-distribution when hypothesis related to the mean of a normal population with unknown variance is tested using small samples.
- Commits hypothesis testing for the mean of any non normal population using the Central Limit Theorem.

**Instructions for Lesson Planning :**

Present the following three situations to the students.

1. A school student states that the average height of a student in grade 13 in this area is 162 cm.
  2. The manager of a company states that the daily average sales has exceeded 210 units that existed prior to launch the new advertising propaganda.
  3. A consumer accuses that the quantity of liquid contained in a bottle of soda is less than 200ml which is mentioned in the label.
- Hold a discussion with the students highlighting the following facts.
    - Each of the three statements mentioned above comes out with an opinion regarding the mean of a population.
    - The fact whether these opinions presented are true or false can be determined based on the evidence revealed through sample data.
    - For that purpose the hypothesis should be tested using an appropriate level of significance.
    - An estimate should be derived for the estimator –  $\bar{X}$  in order to test a hypothesis regarding the population mean -  $\mu$  . For that purpose sample data are required.
  - Not only that, the standard error of the estimator ( $\bar{X}$ ) also should be computed.  
$$\left( \frac{\sigma}{\sqrt{n}} \right)$$

- Sampling distribution of the sample means (distribution of  $\bar{X}$ ) should be partitioned as the acceptance region and critical region for testing the hypothesis that has been built for population mean  $\mu$ . The level of significance given is useful in this matter.
- It would be convenient to derive the critical value in accordance with the corresponding standard normal distribution (z) related to the sampling distribution of sample means considering the given level of significance.
- In order to compare with the critical value, the corresponding z-value of the value received for the sample mean ( $\bar{X}$ ) should be derived that is known as the 'test statistic'.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- In order to compute the test statistic the mean value declared by the null hypothesis ( $\mu$ ) should be deducted from the sample mean derived and divided by the standard deviation of the sample means.

$$\left( \frac{\sigma}{\sqrt{n}} \right)$$

- After the above discussion involve the students in the activity of testing the hypothesis mentioned in No. 01 at the beginning, leading them step by step.

#### Activity – I

- Lead the students to build the null hypothesis and alternative hypothesis considering that the height of the students in grade 13 in this area (X) distributes normally and assuming that the standard deviation is 5cm.

#### Answer :

$$H_0 : \mu = 162 \text{ cm}$$

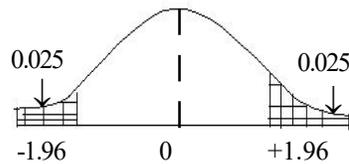
$$H_1 : \mu \neq 162 \text{ cm}$$

(point out that this is the initial step of hypothesis testing)

- Guide the students to compute the mean height of the simple random sample of 9 students drawn from your class considering as the sampling population the grade 13 students in the area which is the target population.
- Guide the students to derive the critical value (cv) required for hypothesis testing using the distribution of z-at 5% level of significance.

(Ascertain the students that  $\left( \frac{\sigma}{\sqrt{2}} \right)$  is required since it is a two tailed test.

Answer :



- Guide the students to state the decisive rule for rejecting or not rejecting  $H_0$  if the test statistic falls in the range  $\pm 1.96$ ,  $H_0$  is not rejected.
- Lead the students to compute the test statistic ( ts) for the sample mean based on sample data using the distribution of  $-z$ .
- Point out that it can be computed as follows, if the sample mean computed by the students is 160.

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{160 - 162}{\frac{5}{\sqrt{9}}} = \underline{\underline{1.2}}$$

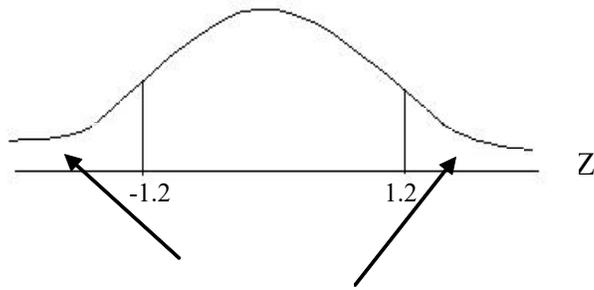
- Let the students to make the decision computing the value of the test statistic with the critical value.  
Since the value of the test statistic (1.2) falls in the acceptance region ( $\pm 1.96$  range)  $H_0$  is not rejected.
- Let the students interpret the statement made by ‘the student’.

**Answer :**

- There is sufficient statistical evidence proposed by the sample data at 5% level of significance to accept the statement made by the student that the average height of a grade 13 student in the area is 162cm.
- Inform the students that the hypothesis testing for population mean also can be conducted using the p-value approach. Explain further that the p-value can be computed using the test statistic.
- Guide the students to compute the p-value.

**Answer :**

- Since the value of the test statistic (ts) is 1.2 and the test is two tailed p-value is computed as follows.



$$P(V) = (0.5000 - 0.3849) \times 2$$

$$= \underline{\underline{0.2302}}$$

- Inquire from the students what the decision should be in accordance with the p-value.

**Answer :**

Since p-value >  $\alpha$  value,  $H_0$  is not rejected.

### Activity – 02

- The manager states that the average daily sales has exceeded 210 units after the newly launched advertising propaganda.
- Once a sample of the sales of 36 days was checked, the mean and the standard deviation have been computed as 225 units and 40 units respectively.
- Point out whether the opinion of the manager is acceptable or not at 1% level of significance.

Compute the p-value of the test.

### Activity 02 : Answer

$$H_0; \mu = 210$$

$$H_1; \mu > 210$$

Critical value  $Z_{\alpha=0.01} = 2.33$

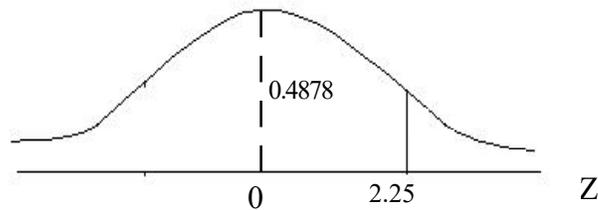
Test statistic  $(t.s); Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{225 - 210}{40/\sqrt{36}} = 2.25$

**Decision :** since  $ts < cv$   $H_0$  is not rejected.

**Conclusion :**

Sufficient evidence for the opinion of the manager to be accepted to be true at 1% level of significance is provided with sample details.

Computing the p-value.



$$\text{p-value} = 0.5000 - 0.4878 = \underline{\underline{0.0122}}$$

Since p-value = 0.0122 > c.v = 0.01  $H_0$  is not rejected.

**Activity - 03 :**

- The quantity of liquid contained in a bottle of soda is mentioned in its label as 200ml, but the customers claim that it does not contain that amount.
- Once a sample of 16 bottles was checked the mean and the standard deviation have been computed as 199ml and 2.5 ml respectively.
- Is the consumers claim acceptable at 0.005 levels of significance?
- Assume that the quantity of liquid contained in the bottle may fall in a normal distribution.

**Answer – Activity – 03**

$$H_0; \mu = 200$$

$$H_1; \mu < 200$$

$$\text{Critical value (CV); } t_{\alpha=0.005, df=16-1} = \underline{\underline{-2.95}}$$

$$\text{Test statistic (ts); } t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{199 - 200}{2.5/\sqrt{16}} = \underline{\underline{-1.6}}$$

Decision : Since t.s. = -1.6 > c.v. -2.95

$H_0$  is not rejected.

Conclusion :

- No sufficient evidence is provided by sample data at  $\alpha = 0.005$  for the claim of the consumer that the quantity of soda contained in a bottle is less than 200ml to be accepted.

### A Guideline to explain the subject matters :

- An opinion about the unknown population mean is known as a hypothesis regarding the population mean.
- Hypothesis associated with the mean of a normal population with known variance should be tested using the standard normal distribution.
- The test statistic used with that regard is

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- Hypothesis associated with the mean of a population whose variance is unknown also should be tested using the standard normal distribution when the sample size is large ( $n \geq 30$ ).
- The test statistic used with that regard is

$$Z = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

- Hypothesis associated with the mean of a normal population whose variance is unknown should be tested using the t-distribution when the sample size is small ( $n < 30$ ).
- The test statistic used with that regard is

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

- Here the distribution of  $\bar{X}$  is  $t_{n-1}$  the number of degrees of freedom in the distribution is given by  $n-1$ .
- A hypothesis testing associated with the population mean can be performed using the p-value approach as well.
- The p-value means the minimum level of probability for the null hypothesis ( $H_0$ ) to be rejected in accordance with the test statistic.
- If the p-value is greater than the level of significance of the test, the null hypothesis won't be rejected.

**Competency 8.0** : Uses statistical hypothesis testing for making business decisions

**Competency Level 8.3** : Uses hypothesis tests for the difference between the means of two populations

**No. of periods** : 10

**Learning Outcomes :**

- Builds up hypothesis for the difference between two population means.
- Examines the evidence for hypothesis built, using test statistics appropriately.
- Derives the critical value with respect of the given details using the normal distribution or t-distribution.
- Makes decisions comparing the value of test statistic with critical value or p-value.

**Instructions for Lesson Planning :**

- Draw the attention of students for following statements.
  - The average height of a boy and a girl in grade 13 is not significant.
  - From the two brands of milk A and B, there is a greater consumer demand for brand A than that for the brand B.
  - When considered the net volume of liquid contained in bottles of soft drinks produced by two companies X and Y, there is an opinion that the net volume of liquid contained in a bottle produced by X company is less than that of in Y company.
- Since it is mentioned that the average height of a boy and a girl in grade 13 is not significant point out that the hypothesis can be built as follows.

Null hypothesis  $H_o; \mu_1 = \mu_2$

Alternative hypothesis  $H_1; \mu_1 \neq \mu_2$

- Since there is an opinion that there is a greater consumer demand exists for the milk of brand A than for brand B, point out that the hypothesis can be built as follows.

Null hypothesis  $H_o; \mu_1 = \mu_2$

Alternative hypothesis  $H_1; \mu_1 > \mu_2$

- Since there is an opinion that the net volume of liquid contained in a bottle of soft drinks produced by X – company is less than that of a same size bottle produced in Y company, point out that the hypothesis can be built as follows.

Null hypothesis  $H_0; \mu_1 = \mu_2$

$$H_1; \mu_1 < \mu_2$$

- In order to test whether the means of two populations are equal or significant, once both the populations distribute normally or not or unknown, point out that the value of test statistic can be computed using the normal distribution as follows, if the size of each sample is separately large (if  $n_1 \geq 30$  and  $n_2 \geq 30$ )
- When the population variances are known

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

When population variances are unknown

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

- Present the following problem to the students and guide them to solve it using a hypothesis testing.

Suppose that the height of men and the height of women of age 18 years distribute normally with standard deviation  $\sigma_1 = 10cm$  and  $\sigma_2 = 10cm$  whether the average height of men of this age is significant to the average height of women of this age is required to be examined. Once samples of 50 men and 40 women were checked, it revealed that the mean height of men is 162.8 cm and that of women is 159.4cm. Test the hypothesis that the height of men and women in this age is not significant at  $\alpha = 0.05$  Level of significance.

**Solution :** Using the critical value,

$$H_0; \mu_1 = \mu_2$$

$$H_1; \mu_1 \neq \mu_2$$

Critical value for the test.

$$X_1 \sim N(\mu_1, \sigma_1^2)$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

$$\therefore (\bar{x}_1 - \bar{x}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

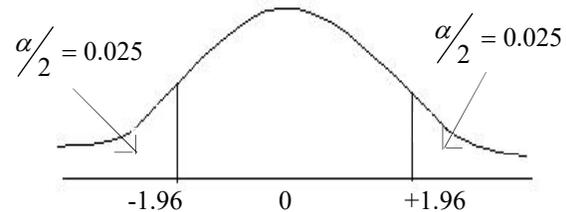
$$n_1 = 50, \quad n_2 = 40 \quad \bar{x}_1 = 162.8\text{cm}, \quad \bar{x}_2 = 159.4\text{cm}, \quad \sigma_1 = 10\text{cm}, \quad \sigma_2 = 12\text{cm}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0.1)$$

$$Z = \frac{162.8 - 159.4}{\sqrt{\frac{10^2}{50} + \frac{12^2}{40}}}$$

$$Z = \frac{3.4}{\sqrt{\frac{10^2}{50} + \frac{12^2}{40}}}$$

$$Z = \underline{\underline{1.436}}$$



### Decision

Since the value of test statistic falls in the acceptance region  $H_0$  is not rejected.

### Conclusion

There is no sufficient evidence at  $\alpha = 0.05$  level to accept that the average height of men and the average height of women of age 18 are significant.

Using P-value approach

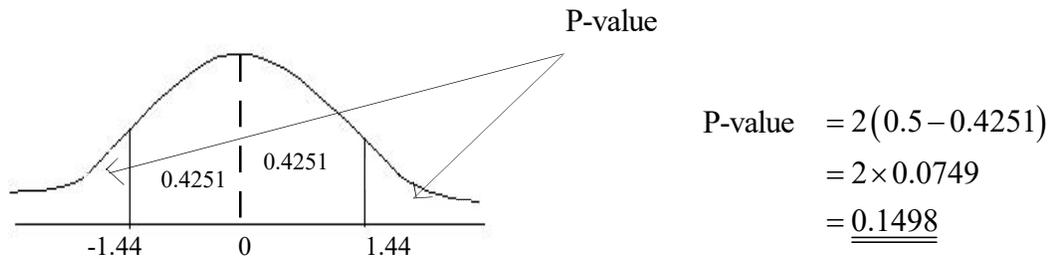
$$H_0; \mu_1 = \mu_2$$

$$H_1; \mu_1 \neq \mu_2$$

The value of test statistic

$$Z = 1.436$$

$$\cong 1.44$$



**Decision :** Since  $P = 0.1498 > \alpha = 0.05$  . $H_0$  is not rejected.

### Conclusion

There is no sufficient statistical evidence at  $\alpha = 0.05$  to accept that the average height of men and the average height of women of age 18 are significant.

- Provide the students with the following problem and guide them to solve it.
- A company involved in manufacturing milk powder has deployed two machines A and B for milk to be packeted. A sample of 100 packets of milk filled in machine A and a sample of 50 packets of milk filled in machine – B were derived and the following measures have been computed.

$$\bar{X}_A = 410g \quad , \quad \bar{X}_B = 396g \quad S_A = 20g, \quad S_B = 50g$$

Test whether more milk powder is filled in packets in machine A than that of in machine B at 5% level of significance.

Building up of hypothesis

$$H_0; \mu_A = \mu_B$$

$$H_1; \mu_A > \mu_B$$

Since  $n_1 = 100$  and  $n_2 = 50$ , in accordance with the Central Limit theorem.

$$\bar{X}_1 - \bar{X}_2 \approx N\left(\mu_A - \mu_B \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\right)$$

Since  $\sigma_1^2$  for  $S_1^2$  and  $\sigma_2^2$  for  $S_2^2$  can be applied as estimators.

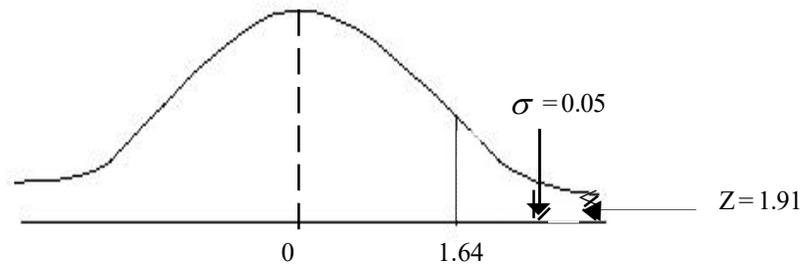
$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim N(0,1)$$

$$Z = \frac{410 - 396}{\sqrt{\frac{20^2}{100} + \frac{50^2}{50}}}$$

$$Z = \frac{14}{\sqrt{54}}$$

$$Z = \underline{\underline{1.91}}$$

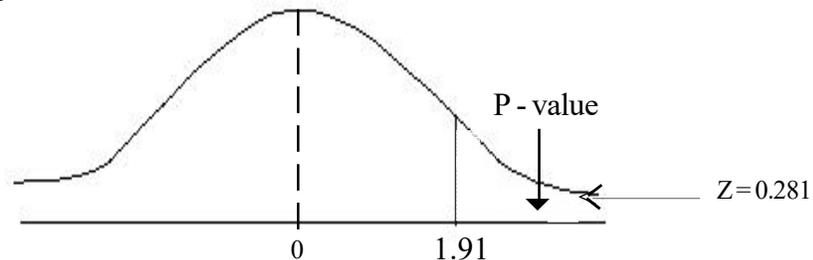
### Critical value



**Decision** : Since the value of the test statistic falls in the critical region,  $H_0$  is rejected.

**Conclusion** : There is sufficient statistical evidence at  $\alpha = 0.05$  level to accept that the average net weight of a packet filled in machinery A is greater than that of a packet filled in machinery B.

### Using P - value approach



**Decision** : Since  $p = 0.0281 < \alpha = 0.05$

$H_0$  is rejected.

**Conclusion** : Same

- Provide the students with the following problem and lead them to solve it.
- A particular company manufactures two brands of batteries which are used in aircrafts. The measures computed using the details derived from a sample of 4 batteries from one brand and a sample of 5 batteries from the other brand regarding their life time are as follows.

$$\bar{X}_1 = 18\ 750h \quad S_1 = 500h$$

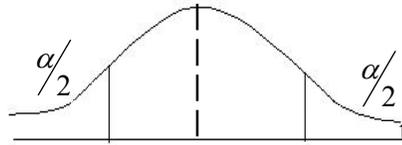
$$\bar{X}_2 = 18\ 480h \quad S_2 = 800h$$

Test whether the mean life time of these two brands of batteries is significant at  $\alpha = 0.05$

**Solution :**

$$H_0; \mu_A = \mu_B$$

$$H_1; \mu_A \neq \mu_B$$



Deriving the critical value using t-distribution.

$$t_{n_1+n_2-2} : \alpha/2 = t_{4+5-2} : 0.02/2$$

$$t_7 : 0.01 = \underline{\underline{2.9998}}$$

- Point out that when the samples drawn from each population is small as  $n_1 < 30$  and  $n_2 < 30$  if the variance of the population is unknown, the normal distribution is inapplicable and then the t-distribution with  $n_1 + n_2 - 2$  degrees of freedom should be used.
- Point out further that the value of test statistic can be computed as follows.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SP \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- $S_p$  is the pooled variance of the variances of two samples and derived as follows.

$$Sp = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \quad \text{is Pooled Standard Deviation}$$

- Lets compute the pooled variance before deriving the test statistics.

$$\begin{aligned}
 Sp^2 &= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \\
 &= \frac{(4 - 1)500^2 + (5 - 1)800^2}{4 + 5 - 2} \\
 &= \frac{3 \times 500 \times 500 + 4 \times 800 \times 800}{7} \\
 &= \frac{750\ 000 + 2\ 560\ 000}{7}
 \end{aligned}$$

$$\begin{aligned}
 S_p^2 &= 472\ 857.14 \\
 \therefore S_p &= \sqrt{472\ 857.14} \\
 &= 687.646 \\
 &= \underline{\underline{687.65}}
 \end{aligned}$$

test Statistic

$$\begin{aligned}
 t &= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\
 t &= \frac{18750 - 18480}{687.65 \sqrt{\frac{1}{4} + \frac{1}{5}}} \\
 t &= \frac{270}{687.65 \sqrt{\frac{9}{20}}} \\
 t &= \frac{270}{687.65 \times 0.671} \\
 t &= \frac{270}{461.41} = 0.585 \\
 t &= \underline{\underline{0.585}}
 \end{aligned}$$

- Decision : Since the value of the tests statistic  $t = 0.585$  is less than the tabled value of t- distribution at  $\alpha = 0.02$   $H_0$  is not rejected.
- Conclusion : There is no sufficient statistical evidence at  $\alpha = 0.02$  to accept that the average life time of these two brands of batteries is significant.

**A guideline to explain the subject matters :**

- There are some practical situations where the accuracy (truthfulness) of some view points expressed regarding the difference of the means of two populations should have been checked.
- The null hypothesis regarding any problem related to the means of two populations is always built as  $H_0 : \mu_1 = \mu_2$   
( $H_0$  : two populations means are not significant)
- The alternative hypothesis can be built either way as following appropriate with the problem as well as the sample data.

$$H_1 : \mu_1 < \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

- If both the populations distribute normally and the variances of both populations  $\sigma_1^2$  and  $\sigma_2^2$  are known, the following formula is used to find the value of the test statistic for hypothesis testing using the normal distribution.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- If the shape of both populations is unknown and/or population variances are unknown, when the size of each sample is large ( $n_1 \geq 30$  and  $n_2 \geq 30$ ) the test statistic can be computed using the following formula applying the normal distribution in accordance with the central limit theorem.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

- If both the populations distribute normally and both population variances are unknown the test statistic should be computed using t-distribution, when small samples are used. The formula to be used is as follows.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Sp-means the pooled standard deviation which is computed as follows.

$$Sp = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \quad t_{n_1+n_2-2}$$

**Competency 8.0** : Uses statistical hypothesis testing for making business decisions

**Competency Level 8.4** : Uses Hypothesis tests for population proportion

**No. of periods** : 04

**Learning Outcomes :**

- Builds up hypothesis for the population proportions.
- Derives the critical value on the level of significance using the normal distribution.
- Computes the value of test statistic using sample data.
- Expresses the decision of the test comparing the test statistic and the level of significance.
- Approaches conclusions regarding the population proportion.

**Instructions for Lesson Planning :**

- Hold a discussion with the students with reference to the following questions.
  - As you suppose which percent of child birth that takes place in Sri Lanka daily may be female?
  - As you suppose which percentage of the teachers employed in Sri Lankan schools may be male teachers ?
  - Which percent of the G.C.E (A/L) qualified students in our country get the chance of getting admitted to local universities for higher education?
- Ascertain that all three situations involve in population proportion.
- Point out that there are official and unofficial opinions expressed on each population proportion.
- Point out that it would be better to make optimal decisions examining the accuracy or inaccuracy of the hypothesis built regarding the population proportion’
- Explain that the same procedure and steps followed in testing hypothesis regarding the population mean are applicable in the same way in testing hypothesis regarding the population proportions as well.
- Point out that the following formula can be applied in computing the test statistic for testing hypothesis regarding the population proportion using the normal distribution.

$$Z = \frac{P - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \sim N(0,1)$$

- Present the following problem to the students.
  - An opinion has been created that a proportion of 0.55 of the child births take place in a day in Sri Lanka is of female. With the purpose of evaluating the verifiability of this opinion a sample of 400 child births was drawn in random and revealed that it contained 216 girls. Test whether this opinion is not true at 5% level of significance.

Solution : Null hypothesis  $H_0 : \pi = 0.55$

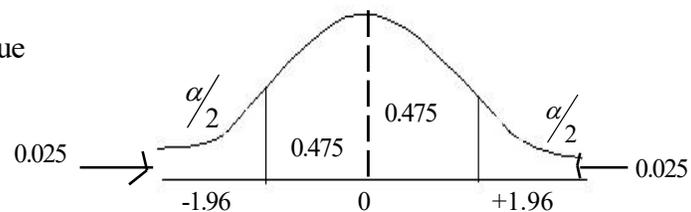
Alternative hypothesis  $H_1 : \pi \neq 0.55$

$n = 400$                        $x = 216$

$$\therefore P = \frac{216}{400} = 0.54$$

$$P \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

Deriving the critical value



Test Statistic

$$Z = \frac{P - \pi}{\frac{\pi(1-\pi)}{n}}$$

$$Z = \frac{0.54 - 0.55}{\sqrt{\frac{0.55 \times 0.45}{400}}}$$

$$Z = \frac{-0.01}{\sqrt{\frac{0.2475}{400}}}$$

$$Z = \frac{-0.01 \times 20}{\sqrt{0.2475}}$$

$$Z = \frac{-0.20}{0.4975} = \underline{\underline{-0.402}}$$

Decision : Since  $Z = -0.402$  is in the acceptance region the null hypothesis is not rejected.

Conclusion : There is no sufficient statistical evidence at 5% level of significance to accept that the female proportion of child births occurred daily is significant from 0.55.

- Group the students appropriately and involve them in the following activity.

### Activity-I

A company manufacturing electric bulbs declared that the proportion of bulbs expired before lighting for 1000 hours is not significant from 0.10. Once a sample of 144 bulbs was drawn in random and checked, it contained 22 bulbs that expired before lighting for 1000 hours. Test whether the statement of this company is not true at

- 5% level of significance
- 2% level of significance
- 1% level of significance

Solution:  $\alpha = 0.05$        $P = \frac{22}{144} = 0.153$        $n = 144$

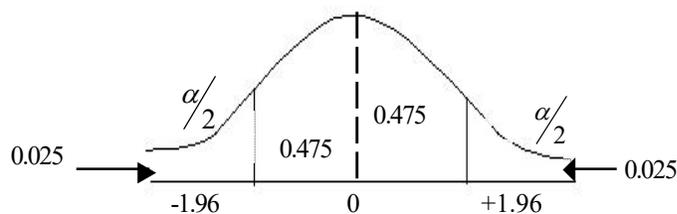
$$P \sim N\left(\pi, \frac{(1-\pi)}{n}\right)$$

Null hypothesis       $H_0 : \pi = 0.1$

Alternative hypothesis       $H_1 : \pi \neq 0.1$

Critical value :

$$\alpha = 0.05$$



Test Statistic

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0.1)$$

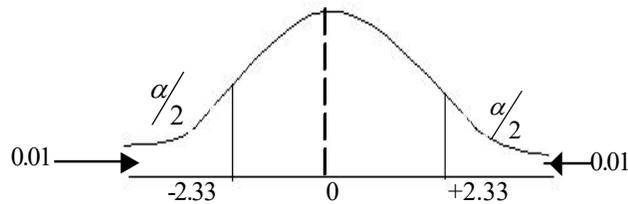
$$z = \frac{0.153 - 0.1}{\sqrt{\frac{0.1 \times 0.9}{144}}} = \frac{0.053 \times 12}{0.3}$$

$$\underline{\underline{Z = 2.12}}$$

Decision : Since the value of test statistic falls in the critical region,  $H_0$  is rejected.

Conclusion : There is sufficient statistical evidence at  $\alpha = 0.05$  to accept that the proportion of bulbs expired before lighting for 1000h is significant from 0.10

Critical value ;  $\alpha = 0.02$

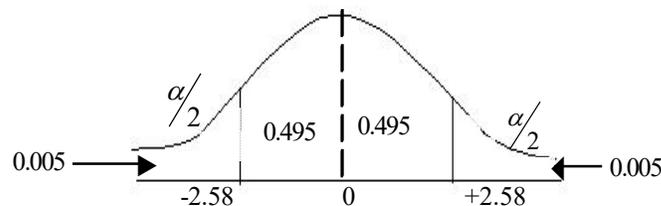


Since the null hypothesis and the alternative hypothesis do not change and the value of test statistic is also the same the decisions of this test can be stated as follows.

Decision : Since the value of the test statistic 2.12 falls in the acceptance region  $H_0$  is not rejected.

Conclusion : There is no sufficient statistical evidence at 2% level to accept that the proportions of the bulbs expired before lighting for 1000h is significant from 0.10.

Critical value :  $\alpha = 0.01$



Since the null hypothesis and alternative hypothesis do not change and the value of the test statistic is the same, the decision and conclusion can be stated as follows.

Decision : Since the value of the test statistic falls in the acceptance region  $H_0$  is not rejected.

Conclusion : There is no sufficient statistical evidence at  $\alpha = 0.01$  to accept that the proportion of bulbs expired before lighting for 1000h is significant from 0.10.

**A guideline to explain the subject matters :**

- Once the proportion of units with a particular attribute contained in a population is unknown, various opinions about that proportion are spread out officially or unofficially at various occasions.
- Once the population distribution is normal, the sampling distribution of the sample proportions also falls normally.
- Unless the population variable distributes normally, it is assumed that the sampling distribution of sample proportions approximately falls normally in accordance with the Central Limit Theorem, once the sample size is large (when  $n \geq 100$ )
- Accordingly the following formula is used for deriving the test statistic in order to test hypothesis regarding the population proportion using the normal distribution.

$$Z = \frac{P - \pi}{\frac{\pi(1 - \pi)}{n}} \sim N(0,1)$$

**Competency 8.0** : Uses statistical hypothesis testing for making business decisions

**Competency Level 8.5** : Uses hypothesis tests for the difference between two population proportions

**No. of periods** : 06

**Learning outcomes :**

- Builds up hypothesis for the difference between two population proportions.
- Derives the critical value on the level of significance using the normal distribution.
- Derives the value of the test statistic using sample statistics.
- Expresses the decision of the test comparing the value of test statistic and the level of significance.

**Instructions for Lesson Planning :**

Present the following flash cards to the class and initiate a discussion.

12% of the employees in A- bank are apprentices

8% of the employees in B- bank are apprentices

6% of the pawned articles in A- bank are not redeemed

5% of the pawned article in B - bank are not redeemed.

- Point out to the students that there are practical situations where decisions have to be made comparing the proportion of a particular attribute in one population with that of another.
- Point out that it would be better to examine whether the assumptions forwarded regarding the difference between the proportions of an attribute in two populations are true or false.
- If both the populations distribute normally or the sample size is sufficiently large ( $n_1 \geq 100$  and  $n_2 \geq 100$ ) point out that the normal distribution with the mean  $\pi_1 - \pi_2$  and variance  $\left( \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2} \right)$  can be used to test the hypothesis regarding the difference between the proportions of a particular attribute of two populations.
- Provide the students with the following problem and guide them to be involved in testing the hypothesis in connection with that problem.

Two samples of house wives in the Matale district and the Colombo district were drawn in random and the details collected regarding the preference for a recently introduced confectionary meal are stated in the following table.

| District | Sample size | Number of housewives used that meal |
|----------|-------------|-------------------------------------|
| Colombo  | 500         | 240                                 |
| Mathale  | 1500        | 600                                 |

Test whether there is a greater demand for this confectionary meal in the Colombo district than that of in Matale at 5% level of significance.

**Solution :**

Hypothesis:

$$H_0 : \pi_1 = \pi_2$$

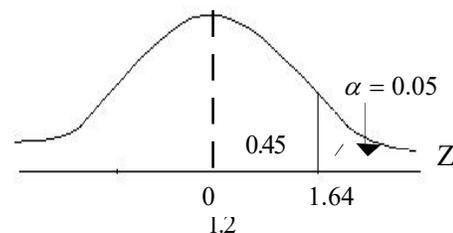
$$H_1 : \pi_1 > \pi_2$$

N. B. : Lets suppose that the proportion of housewives who prefer this confectionary meal in the Colombo district as  $\pi_1$  and that of Matale district as  $\pi_2$

- Deriving the mean of the proportions.
- Deriving the critical value

$$\begin{aligned}\bar{P} &= \frac{X_1 + X_2}{n_1 + n_2} \\ \bar{P} &= \frac{240 + 600}{500 + 1500} \\ \bar{P} &= \frac{840}{2000} \\ \bar{P} &= \underline{\underline{0.42}}\end{aligned}$$

$$\begin{aligned}\bar{q} &= 1 - \bar{p} \\ &= 1 - 0.42 \\ &= \underline{\underline{0.58}}\end{aligned}$$



- The test statistic

$$= \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z = \frac{(0.48 - 0.40) - 0}{\sqrt{0.42 \times 0.58 \left(\frac{1}{500} + \frac{1}{1500}\right)}}$$

$$Z = \frac{0,08}{\sqrt{0.2436 \left(\frac{3+1}{1500}\right)}}$$

$$Z = \frac{0,.08}{\sqrt{0.2436 \times 0.0027}}$$

$$Z = \frac{0.08}{0.0256}$$

$$Z = \underline{\underline{3.125}}$$

Decision : Since the value of the test statistic falls in the critical region,  $H_0$  is rejected.

Conclusion : There is sufficient statistical evidence at 5% level of significance to accept that the proportion of housewives who use this confectionary meal in Colombo district is greater than that of in Matale.

- Separate the students into three groups and present the following problem. Then involve them in activity of testing the hypothesis assigned with.

(i)  $H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 \neq \pi_2$

(iii)  $H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 > \pi_2$

(ii)  $H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 < \pi_2$

- An agricultural research centre has developed two species of chilies. Two samples of seeds including 100 seeds in each were drawn and placed in nurseries separately under the identical physical conditions with the purpose of checking the germination feasibility. Once observed in three days 92 saplings in the 1<sup>st</sup> nursery and 91 saplings in the 2<sup>nd</sup> nursery were observed.

Test whether the germination feasibility of these two species of chilies is significant at 5% level of significance. (Using both the critical value and the p-value approaches)

Solution 01 :

$$H_o : \pi_1 = \pi_2$$

$$H_1 : \pi_1 \neq \pi_2$$

Observed data

$$n_1 = 100 \quad n_2 = 100 \quad x_1 = 92 \quad x_2 = 91$$

$$p_1 - p_2 \sim N\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right)$$

- Test statistic

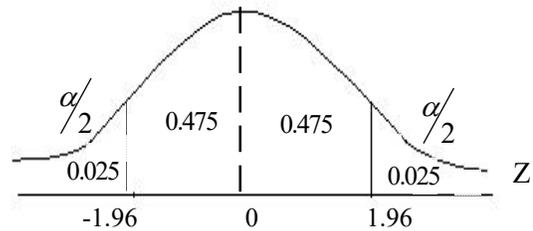
$$Z = \frac{P_1 - P_2}{\sqrt{\bar{p} \bar{q} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z = \frac{(0.92 - 0.91)}{\sqrt{0.915 \times 0.085 \left(\frac{1}{100} + \frac{1}{100}\right)}}$$

$$Z = \frac{0.01}{\sqrt{0.077775 \times 0.02}}$$

$$Z = \underline{\underline{0.2535}}$$

- Critical Value



$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

$$= \frac{92 + 91}{100 + 100}$$

$$= \underline{\underline{0.915}}$$

- Decision

$H_o$  is not rejected.

- Conclusion :

$$\bar{q} = 1 - 0.915$$

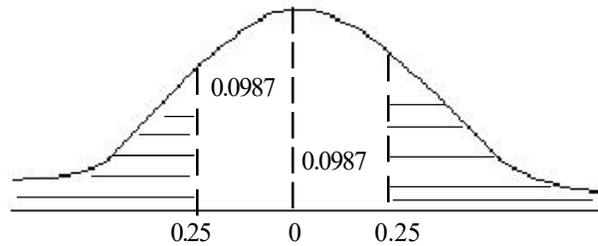
$$= \underline{\underline{0.085}}$$

There are no sufficient statistical evidence at 5% level of significance to accept that the germination feasibly of two species of chilies is significance.

Using p - value approach

- P - value is the minimum probability level of rejecting  $H_o$  in accordance with the value of the test statistic

$$\begin{aligned}
 \text{P-value} &= 1 - 2 \times 0.0987 \\
 &= 1 - 0.1974 \\
 &= \underline{\underline{0.8026}}
 \end{aligned}$$



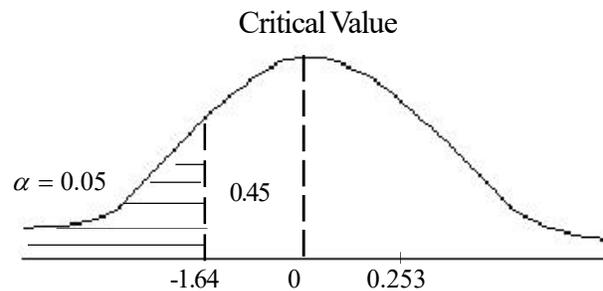
Decision : Since the p-value 0.8026. > critical value  $\alpha = 0.05$   $H_0$  is not rejected.

Conclusion : Same conclusion as approached above.

(ii) Hypothesis

$$H_0 : \pi_1 = \pi_2$$

$$H_1 : \pi_1 < \pi_2$$



Observed data :  $n_1 = 100$      $n_2 = 100$      $x_1 = 92$      $x_2 = 91$

Test statistic  $Z = 0.253$

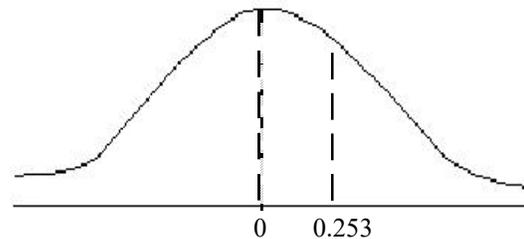
Decision : Since the test statistic is 0.253,  $H_0$  is not rejected.

Conclusion : The same conclusion as approached above

If p-value approach is used.

$$\text{p-value} = 0.5 + 0.0987$$

$$= \underline{\underline{0.5987}}$$



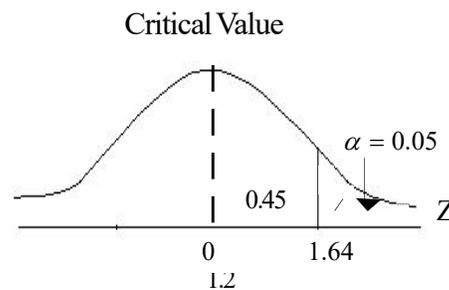
Decision : Since the p-value is > critical value  $H_0$  is not rejected.

Conclusion : The same conclusion as approached above.

(iii) Hypothesis

$$H_0 : \pi_1 = \pi_2$$

$$H_1 : \pi_1 > \pi_2$$



Test statistic  $Z = 0.253$

Decision : Since the test statistic is 0.293,  $H_0$  is not rejected.

Conclusion : The same conclusion approached above

If p-value approach is used

p-value =  $0.5 - 0.0987$

= 0.4013

Decision : Since the p-value is  $>$  critical value  $H_0$  is not rejected.

Conclusion : The same conclusion approached above.

### A Guideline to explain the subject matters :

- There are practical situation where decisions should be made comparing the difference of the proportions of the number of units with a particular attribution in two populations.
- The fact whether the proportion of units with a particular attribute in one population is equal to the proportion of units with the same attribute in another population should be examined. This is known as the hypothesis testing related to the difference between two population proportions.
- Once both the populations are given to be distributed normally or when each sample size is sufficiently large ( $n_1 \geq 100$  and  $n_2 \geq 100$ ) the hypothesis  $\pi_1 < \pi_2$  or  $\pi_1 > \pi_2$  or  $\pi_1 \neq \pi_2$  vs  $\pi_1 = \pi_2$  can be tested using the normal distribution.
- Hence, the sampling distribution of the difference between two sample proportions falls normally as follows.

$$P_1 - P_2 \sim N\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right)$$

- The test statistic for testing hypothesis related to the difference between two population proportions can be derived as follows.

$$Z = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

- $\bar{p}$  is the average (mean) proportion of two samples.

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

$X_1$  – The number of units with the considered attribute drawn from the first population.

$X_2$  - The number of units with the considered attribute drawn from the second population.

- A hypothesis test related to the difference between two population proportions can be performed using p-value approach as well.
- The p-value of a left tail test is equal to the area under the normal curve on left hand side to the test statistic and the p-value of a right tail test is equal to the area under the normal curve on right hand side to the test statistic whereas the p-value of a two tail test is equal to the sum of the area under the normal curve on either side of the test statistics marked on both sides on the curve.

**Competency 8.0 :** Uses statistical hypothesis testing for making business decisions

**Competency Level 8.6 :** Uses Chi-square Testing for making business decision.

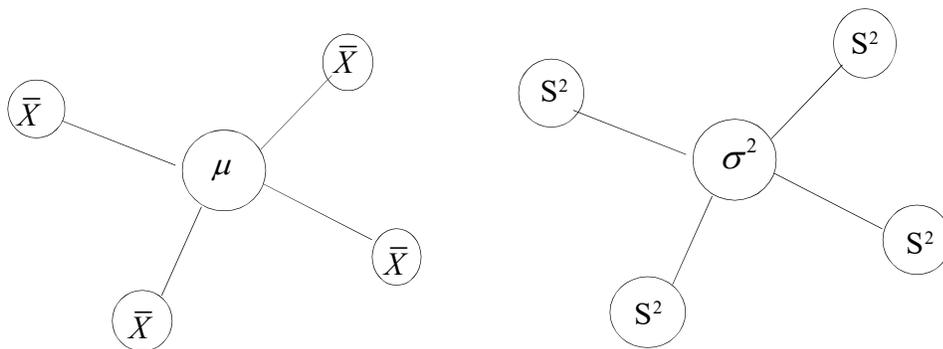
**No. of periods :** 04

**Learning Outcomes :**

- Introduces Chi-square testing.
- Introduces the characteristics of Chi-square distribution.
- Names the instances where Chi-square testing can be applied.
- Tests the homogeneity of an observed distribution.
- Holds a Chi-square test for the independence of two variables.
- Introduces the contingency co-efficient.
- Computes the contingency co-efficient.
- Fits a binomial distribution for observed data.
- Launches a Chi-square test for evaluating the goodness of fit of a binomial distribution.
- Fits a Poisson distribution for observed data.
- Holds a Chi-square testing for evaluating the goodness of fit of that distribution.

**Instructions for Lesson Planning :**

- Display the following diagrams on the board and hold a discussion,



- Inquire from the students how the shape of the sampling distribution of sample means of the samples drawn in equal size from a normal population is ?

- Explain to the students that the sampling distribution of sample variances does not distribute normally, even though the sampling distribution of sample means of the same sizes samples drawn from a normally distributed population with known variance falls normally.
- Express that the sampling distribution of sample variances falls in a chi-square distribution ( $\chi^2$ )
- Discuss the characteristics of a chi-square testing with students.
  - There are no negative (minus) value in  $\chi^2$
  - $\chi^2$  distribution is a positively skewed distribution.
  - Using relevant tables for chi-square distribution the area of the right hand side tail of the curve is derived with reference to degrees of freedom (df) and the  $\alpha$  value given.
- Explain that the  $\chi^2$ -distribution is used when considered regarding a variation between observed data and expected data.
- Discuss with the students regarding the following instances where chi-square distribution is applied.
  - In order to make decisions regarding homogeneity.
  - In order to make decisions regarding independency.
  - In order to make decisions regarding the goodness of fit of the distributions.
- Explain the steps of a hypothesis testing to the students as follows.
  - build up hypothesis related to the relevant problem.
  - compute the value of Chi-square test statistic.
  - derive the table value of  $\chi^2$
  - make the decision comparing the  $\chi^2$  table value and the test statistic computed above.
  - State the conclusion that can be achieved.
  - Point out to the students that the value of  $\chi^2$  is derived as follows.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

- Explain that  $O_i$  – denotes for observed data and  $E_i$ -denotes for expected data.

Activity-I

- Provide the students with the following activity and lead them to get engage in it.
- The details collected to check whether the sales of 5 branches of a company are occurred uniformly are as follows.

| Branch | Sales (units 000) |
|--------|-------------------|
| A      | 40                |
| B      | 60                |
| C      | 47                |
| D      | 40                |
| E      | 63                |

Test whether the sales of the five (5) branches have been distributed equally at  $\alpha = 0.01$  and  $\alpha = 0.05$

Solution – Activity – I

- Guide the students to build the hypothesis as follows depend on the null hypothesis that sales distribute uniformly.

$H_0$ : Sales of five branches distribute uniformly,

$H_1$ : Sales of five branches do not distribute uniformly

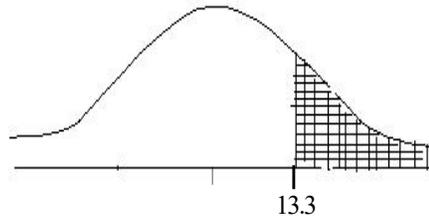
- Compute the test statistic.

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

| $O_i$ | $E_i$ | $(O_i - E_i)$ | $(O_i - E_i)^2$ | $(O_i - E_i)^2 / E_i$ |
|-------|-------|---------------|-----------------|-----------------------|
| 40    | 50    | - 10          | 100             | 2.00                  |
| 60    | 50    | 10            | 100             | 2.00                  |
| 47    | 50    | - 03          | 09              | 0.18                  |
| 40    | 50    | - 10          | 100             | 2.00                  |
| 63    | 50    | 13            | 169             | 3.38                  |
|       |       |               |                 | 9.56                  |

- Since the total sale of five branches was 250 000 units, the sales of one branch should be 50 000 units  $\left(\frac{250}{5} = 50\right)$  and that value is considered as the expected frequency.
- Derive the  $\chi^2$  table value with reference to the degrees of freedom  $k - 1$  and  $\alpha = 0.01$   $K - 1$  and  $\alpha = 0.01$ . That is  $X^2 = 13.3$

$$\chi^2_{4;0.01,4} = 13.3$$



- Decision :  
 $H_0$  is not rejected since the value of the test statistic falls in the acceptance region.
- Conclusion :  
There is sufficient statistical evidence at  $\alpha = 0.01$  to accept that the sales of five branches are dispersed uniformly.
- Let's consider to test the above hypothesis at  $\alpha = 0.05$  there is table value at  $X^2_{4;0.05} = 9.49$ .
- Then the test statistic falls in the critical region. Therefore  $H_0$  is rejected.
- Hence, explain that there is sufficient statistical evidence at  $\alpha = 0.05$  to accept that sales of five branches are significant.
- Explain to the students that a hypothesis which was rejected at  $\alpha = 0.05$  may possible to be accepted at  $\alpha = 0.01$  level of significant.
- Emphasis to the students that the same steps mentioned above should be followed in the same way in testing hypothesis related to independency as well.
- Explain further that the hypothesis should be built dependent on the null hypothesis so that both the variables considered are independent.
- Once  $\chi^2$  test statistic is computed point out to the students that the same formula

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \text{ is applicable.}$$

- Explain how to compute the expected frequencies as follows;

$$E_i = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

- When the table value is read out both the critical value and the degrees of freedom (df) are used.
- Number of degrees of freedom should be computed as follows.

$$(r - 1) (c - 1)$$

r – Number of rows    c – Number of column

#### Activity – 2

- Draw the attention of students to the following problems for explaining the chi-square testing for dependency of a contingency table.

A sample of 200 consumers buying a recently introduced latest brand of soap to the market has been separated from gender and age as follows.

| Age (years)      | Gender |      | Total |
|------------------|--------|------|-------|
|                  | Female | Male |       |
| Less than 25 yrs | 60     | 40   | 100   |
| 25 yrs or more   | 70     | 30   | 100   |
| Total            | 130    | 70   | 200   |

- Test whether the gender and age of these consumers of this brand of soap are independent at 0.05 level of significance.
- Point out that the above table is known as a contingency table and it may consist of a set of observations categorized under two or more variables.

#### Activity – 2 : Solution

- Hypothesis

$H_0$  : Gender and age of the consumers who bought soap are independent from one another.

$H_1$  : Gender and age of the consumers who bought soap are dependent on one another.

- Explain that the expected frequencies are computed as follows.

$$\text{Expected Frequency} = \frac{\text{Column Total} \times \text{Row total}}{\text{Grand total}}$$

| Age (years)  | Gender                                  |  | Total |
|--------------|---|--|-------|
|              | Female                                  | Male                                   |       |
| Less than 25 | 60<br>$\frac{100 \times 130}{200} = 65$ | 40<br>$\frac{100 \times 70}{200} = 35$ | 100   |
| 25 or more   | 70<br>$\frac{100 \times 130}{200} = 65$ | 30<br>$\frac{100 \times 70}{200} = 35$ | 100   |
| Total        | 130                                     | 70                                     | 200   |

Guide the students to compute the chi-square statistic as follows.

| $O_i$ | $E_i$ | $(O_i - E_i)^2$ | $O_i - E_i$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|-------|-------|-----------------|-------------|-----------------------------|
| 60    | 65    | -5              | 25          | $\frac{25}{65} = 0.38$      |
| 40    | 35    | +5              | 25          | $\frac{25}{35} = 0.71$      |
| 70    | 65    | +5              | 25          | $\frac{25}{65} = 0.38$      |
| 30    | 35    | -5              | 25          | $\frac{25}{35} = 0.71$      |
| $X^2$ |       |                 |             | = 2.18                      |

- Point out that the number of degrees of freedom is computed as follows.

$$Df = (r - 1)(c - 1)$$

r = number of rows

c = number of columns

$$df = (2-1)(2-1)$$

$$df = 1 \times 1$$

$$df = \underline{\underline{1}}$$

- Accordingly the number of degrees of freedom related to the above problem is
 
$$d_i = (2-1)(2-1)$$

$$d_i = 1 \times 1 = \underline{\underline{1}}$$
- Table value of  $X_{0.05;1}^2 = 3.84$
- The value of test statistic computed is 2.18
- Provide the students with instructions to test whether the gender and age group related to this problem are independent.
- The table value of  $X_{0.05;1}^2 = 3.84$  and the  $\chi^2$  test statistic is 2.18
- Therefore  $H_0$  is not rejected. Hence the gender and age group are independent to each other.
- Guide the students to compute the contingency co-efficient for the above problem.
- Point out that the contingency co-efficient is a measure involved with the  $\chi^2$  value in order to evaluate the size of the relationship between two types of variables and, that it can be computed as

$$C = \sqrt{\left(\frac{\chi^2}{T + \chi^2}\right)}$$

- Accordingly T – is the total number of observations.

$$C = \sqrt{\frac{2.18}{200 + 2.18}}$$

$$C = \sqrt{\frac{2.18}{202.18}}$$

$$C = \sqrt{0.01078}$$

$$C = \underline{\underline{0.1038}}$$

- Explain the students that the relationship between these two variables is very poor, since the contingency co-efficient is 0.1038.
- Explain that the same steps mentioned above can be followed in testing hypothesis related to the goodness of fit of a probability distribution.

- Explain further that the test is conducted based on the null hypothesis that ‘the probability distribution fit is good enough.’
- Point out that the theoretical frequency is used as the expected frequency in computing the  $\chi^2$  statistic.
- Point out further that the level of significance ( $\alpha$ ) and the degrees of freedom are used in deriving the  $\chi^2$  table value.

$$d.f = k - 1 - m$$

- Inform that ‘m’ indicates the number of parameters estimated using the sample statistics.

### Activity – 3

- Involve the students in following activity to explain how to use the  $\chi^2$  test for examining the goodness of fit of the probability distributions.
- Number of heads received and the respective frequencies, once five coins were thrown together are mentioned below. Fit a binominal distribution for these data and test the goodness of fit at  $\alpha = 0.05$

| No. of Heads | Frequency |
|--------------|-----------|
| 0            | 30        |
| 1            | 152       |
| 2            | 330       |
| 3            | 298       |
| 4            | 164       |
| 5            | 26        |
|              | 1 000     |

### Activity – 3 Solution

- The parameters to be known in order to estimate to fit a binominal distribution are the number of trials – n and the probability of getting success – p.
- Guide the students to compute the mean of the observed frequency distribution.

| $x$ | $f$  | $fX$ |
|-----|------|------|
| 0   | 30   | 0    |
| 1   | 152  | 152  |
| 2   | 330  | 660  |
| 3   | 298  | 894  |
| 4   | 164  | 656  |
| 5   | 26   | 130  |
|     | 1000 | 2492 |

$$\bar{X} = \frac{\sum f \times x}{\sum f}$$

$$\bar{X} = \frac{2492}{1000}$$

$$\bar{X} = \underline{\underline{2.492}}$$

- Recalling that  $\mu = np$  is the mean of a binominal distribution make the above answer equal to  $np$

$$np = 2.492$$

- Since 5 coins are thrown state that the number of trials is 5 and guide the students to solve the above problem.

$$np = 2.492$$

$$5p = 2.492$$

$$p = 0.4984$$

- If  $P = 0.5$  as derived above, inform the students that the relevant probabilities can be derived using the binominal distribution table or the function of binominal distribution.
- Explain that the probability values should be multiplied by the total frequency and rounded off to the nearest whole number and further that the sum of theoretical frequencies should be equal to the total frequency.

| $x$ | $P(X)$ | Theoretical Frequency | Observed Frequency |
|-----|--------|-----------------------|--------------------|
| 0   | 0.0313 | 31                    | 30                 |
| 1   | 0.1563 | 156                   | 152                |
| 2   | 0.3125 | 313                   | 330                |
| 3   | 0.3125 | 313                   | 298                |
| 4   | 0.1563 | 156                   | 164                |
| 5   | 0.0313 | 31                    | 26                 |

- Point out that the test statistic for  $\chi^2$  test is computed as follows using the theoretical frequencies and observed frequencies.

| $O_i$ | $E_i$ | $O_i - E_i$ | $(O_i - E_i)^2$ | $(O_i - E_i)^2 / E_i$ |
|-------|-------|-------------|-----------------|-----------------------|
| 30    | 31    | - 1         | 1               | 0.032                 |
| 152   | 156   | - 4         | 16              | 0.102                 |
| 330   | 313   | 17          | 289             | 0.923                 |
| 298   | 313   | 15          | 225             | 0.719                 |
| 164   | 156   | 08          | 64              | 0.410                 |
| 26    | 31    | 05          | 25              | 0.806                 |

- Point out that the number of degrees of freedom should be taken as  $k - 1 - m$ .
- Since the number of parameters estimated using the sample statistic being 1, point out that the degrees of freedom can be found as follows.

$$\begin{aligned}
 d \cdot f &= k - 1 - m \\
 &= 6 - 1 - 1 \\
 &= \underline{4}
 \end{aligned}$$

- Hence point out that  $\chi^2$  table value should be derived as  $\chi_{0.05;4}^2 = 9.49$
- Since the test statistic is less than the table value  $H_0$  is not rejected.
- Hence, it can be concluded that the theoretical probability distribution (binominal distribution) which has been fit is appropriate at  $\alpha = 0.05$  level of significance.

#### Activity – 4

- The number of subject simultaneous activities for which 56 teachers have been attached are as follows.

| Number of subject Paralel activities | No of teachers |
|--------------------------------------|----------------|
| 0                                    | 30             |
| 1                                    | 12             |
| 2                                    | 08             |
| 3                                    | 03             |
| 4                                    | 02             |
| 5                                    | 01             |
| 6 or more                            | 00             |
|                                      | 56             |

- Fit a Poisson distribution using these data and examine the goodness of fit of that distribution.

$H_o$  : The Poisson distribution fit is sufficiently good.

$H_i$  : The Poisson distribution fit is not sufficiently good.

| $O_i$ | $E_i$ | $O_i - E_i$ | $(O_i - E_i)^2$ | $(O_i - E_i)^2 / E_i$ |
|-------|-------|-------------|-----------------|-----------------------|
| 30    | 23    | 7           | 49              | 2.13                  |
| 12    | 20    | -8          | 64              | 3.20                  |
| 14    | 13    | -1          | 1               | 0.08                  |
|       |       |             |                 | $\chi^2 = 5.41$       |

- Point out that, once the expected frequency is less than 5, those values should be combined to the nearest interval.

The test  $\alpha = 0.05$   $m = 1$  and  $\lambda = 1$

$$f.f = k - 1 - m$$

$$= 3 - 1 - 1$$

$$= \underline{1}$$

$$\chi_{0.05;1}^2 = 3.84$$

- Since the value of test statistics is greater than the table value  $H_o$  is rejected. Therefore the Poisson distribution which has been fit is not good enough at 5% level of significance.

**A Guideline to explain the subject matters.**

- Chi-square distribution is used in making decisive rules for testing hypothesis regarding the population variance. These tests made using chi-square distribution are called Chi-square distribution tests.
- Chi-square distribution consists of the following characteristics.
  - A distribution of the variance of sample data of random samples drawn from normally distributed populations.
  - Chi-square distribution is right skewed.
  - The distribution depends on the number of degrees of freedom, while the number of degrees of freedom is determined on the sample size.
  - Once the sample size is increased the distribution is closer to a normal distribution.

If the observed frequency for k-number of events are  $O_1, O_2, \dots, O_k$  and the expected frequencies are as  $E_1, E_2, \dots, E_k$  In order to check the agreement

between expected frequencies and observed frequencies the test statistic of a chi-square test can be computed as follows.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

- The instances where  $\chi^2$  tests are applicable as to search the independency of two variables, to fit the theoretical probability distributions for a given observed distribution and to check the homogeneity of variables.
- A table which is used to indicate the dependency existing between two or among many variables is known as a contingency table.
- The number of rows and the number of columns of a contingency tables are denoted by 'r' and 'c' respectively. Hence the number of degrees of freedom of a contingency table is found as  $(r - 1) (c - 1)$
- In order to quantify the size of the relationship (conflict) between two attributes the contingency co-efficient can be introduced as a measure which is directly involved in variance. Initially, the contingency co-efficient and the correlative co-efficient are not significant, but the contingency co-efficient can be used for categorical or ranked scale measures as well in addition to interval scale measures.
- Accordingly, the contingency co-efficient can be computed as follows.

$$C = \sqrt{\left(\frac{\chi^2}{T + \chi^2}\right)}$$

Here T is for the total number of observations.

- $\chi^2$  distribution can be applied to test the hypothesis that "there is a significant theoretical distribution for a particular random variable."
- Accordingly the chi-square test can be used to test whether a continuous variable distributes normally as well as whether a discrete random variable has an acceptable theoretical probability distribution.
- The chi-square test which is used to check the goodness of fit can be considered as

$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$  which is built up based on the differences of observed sample frequencies and assumed theoretical frequencies.

- Mainly to approximate a discrete distribution through a chi-square distribution which is of continuous variables, the expected frequency of each interval should be at least five (5)

**Competency 8.0** : Uses statistical hypothesis testing for making business decisions

**Competency Level 8.7** : Uses the techniques of variance analysis for making decisions regarding the equity of the means of more than two populations.

**No. of periods** : 14

**Learning outcomes :**

- Explains the objectives of variance analysis.
- Explains the assumptions for variance analysis.
- Expresses the model of variance analysis.
- Builds up hypothesis regarding the equity of the means of more than two normal populations.
- Computes the variance between the samples and the variance within the samples and then derives the test statistic.
- The test statistic is derived through an ANOVA table
- Introduces the F-distribution.
- Derives the critical value on the level of significance using the F-distribution.
- States the decision of the test comparing the test statistic and the critical value.
- Approaches conclusions regarding the hypothesis.

**Instructons for Lesson Planning :**

- Provide the students with the following situations.
  1. Testing whether four different kinds of fertilizers recommended for a particular crop contributes of giving a similar harvest.
  2. Testing whether the three different techniques of exercises introduced for reducing obesity are equally effective.
  3. Testing whether the motor car batteries manufactured under five (5) different methodologies are of equal life expectation.
- Hold a discussion with the students highlighting the following facts.
  - Point out that the fact whether the means of each situations as mentioned above are equal is required to be examined.
  - State that small samples should be used in such pilot surveys.
  - Explain to the students that, whether the means are equal can be tested through examining the sample variances as well.
  - State further that certain assumptions regarding population distributions may be required.
  - Recall that the sample variances can be used to represent population variances.

- Explain that, for that purpose, the variance between the samples (variance between the treatments) and variance within the samples (experimental errors) can be utilized.
- Point out that since the conclusions can be achieved whether the population means are equal through analysis of sample variances, that technique can be defined as variance analysis.
- Explain that the ratio of the variance between the samples to the variance within the samples is denoted as F-statistic and further that the distribution of the values taken by F is a sampling distribution.
- Explain that a variance analysis can be conducted using F-distribution based on the null hypothesis that “the population means are equal.”
- In variance analysis or in F-test the alternative hypothesis can be built as “at least one of the population means is significant form the other population means.”
- If the variance between the samples is equal to the variance within the samples, then the value of F- is 1. And it is assumed that all the population means are equal with 100%.
- The more population means differ, greater the values of F-statistic.

### Activity

- With the purpose of livestock development three kinds of food items have been introduced as  $R_1$ ,  $R_2$  and  $R_3$ . Each food was given separately for 12 selected pigs as per 4 in each, to test whether the body weight of the animals would increase. After giving this food items for a considerable period. The increased weight of each animal has been observed in kgs and the values deducted per 10 from each observation are given in the following table.

| Pigs | Kind of food |       |       |
|------|--------------|-------|-------|
|      | $R_1$        | $R_2$ | $R_3$ |
| 1    | 3            | 2     | 3     |
| 2    | 4            | 2     | 9     |
| 3    | 5            | 4     | 5     |
| 4    | 2            | 3     | 7     |

- Test whether the effectiveness of three kinds of food items is equal at  $\alpha = 0.05$  and come to your conclusions.

Instructions – 1. State the null hypothesis and the alternative hypothesis for testing the hypothesis.

Solution :

$$H_0 : \mu_{R1} = \mu_{R2} = \mu_{R3}$$

$H_1$  : At least two of the food items are not equally effective.

Instructions 2 : find the degrees of freedom between the samples.

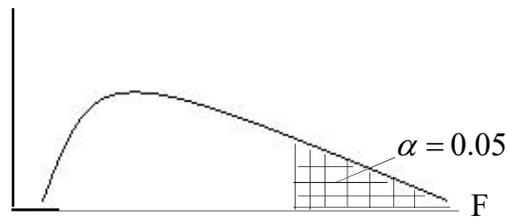
Solution :  $k - 1 = 3 - 1 = 2$

Find the degrees of freedom within the samples.

Solution :  $k(n - 1) = 3(4 - 1) = 9$

Instructions 3 : Derive the critical value of the test using F-distribution table.

Solution :



$$c.v = 4.26$$

Instruction 4 : For computing the value of the test statistic (F) prepare a table containing the data of  $R_1$   $R_2$   $R_3$  the summation of them, the square of those data and their summations.

Solution 4:

| $R_1$ | $R_1^2$ | $R_2$ | $R_2^2$ | $R_3$ | $R_3^2$ |
|-------|---------|-------|---------|-------|---------|
| 3     | 9       | 2     | 4       | 3     | 9       |
| 4     | 16      | 2     | 4       | 9     | 81      |
| 5     | 25      | 4     | 16      | 5     | 25      |
| 2     | 4       | 3     | 9       | 7     | 49      |
| 14    | 54      | 11    | 33      | 24    | 164     |

Instructions 5 : Compute the correction factor  $\frac{T^2}{N}$

Solutio:

$$\begin{aligned} \frac{T^2}{N} &= \frac{(\sum R_1 + \sum R_2 + \sum R_3)^2}{k \times n} \\ &= \frac{(14 + 11 + 24)^2}{3 \times 4} = \frac{2401}{12} \\ &= \underline{\underline{200.08}} \end{aligned}$$

Instruction 6 : Compute the sum of the squares of totals. (SST)

$$\begin{aligned}\text{Solution : } SST &= \left[ \sum R_{i1}^2 + \sum R_{i2}^2 + \sum R_{i3}^2 \right] - \frac{T^2}{N} \\ &= [54 + 33 + 164] - 200.08 \\ &= 251 - 200.08 \\ &= \underline{\underline{50.92}}\end{aligned}$$

Instruction 7 : Compute the sum of the squares between the samples (SSC)

$$\begin{aligned}SSC &= \left[ \frac{(\sum R_{i1})^2}{n_1} + \frac{(\sum R_{i2})^2}{n_2} + \frac{(\sum R_{i3})^2}{n_3} \right] - \frac{T^2}{N} \\ &= \left[ \frac{14 \times 14}{4} + \frac{11 \times 11}{4} + \frac{24 \times 24}{4} \right] - 200.08 \\ &= [49 + 30.25 + 144] - 200.08 \\ &= 223.25 - 200.08 \\ &= 223.25 + 200.08 = \underline{\underline{23.17}}\end{aligned}$$

Instruction 8 : Compute the sum of the squares within the samples (SSE)

$$\begin{aligned}\text{Solution } SSE &= SST - SSC \\ &= 50.92 - 23.17 \\ &= \underline{\underline{27.75}}\end{aligned}$$

Instruction 9 : Compute the test statistic (t.s)

$$\begin{aligned}ts; F &= \frac{SSC / (k - 1)}{SSE / k(n - 1)} \\ &= \frac{23.17 / (3 - 1)}{27.75 / 3(4 - 1)} = \frac{11.58}{3.08} \\ &= \underline{\underline{3.76}}\end{aligned}$$

Instruction 10 : State the decision. Since  $ts < cv$  (since test statistic is less than critical value)  
Ho is not rejected.

Instruction 11 : Describe about the hypothesis and approach at the conclusion.

Solution 11 : There is sufficient statistical evidence to accept that the effectiveness of three kinds of food items is equal at 5% level of significance.

**A Guideline to explain the subject matters :**

- The initial objective of the variance analysis is to test the equality in many population means.

- Since this process is conducted using sample variances, it is known as the variance analysis.
- Following assumptions are made in variance analysis.
  1. that the dependent variable (responding variable) of each population expected to be compared distributes normally.
  2. that the distributions of the dependent variables of each population expected to be compared consists of an equal variance.
- In order to test whether the population means are equal through variance analysis the F-distribution is applied.
- F- distribution is the distribution of the ratio of the variance between small samples to the variance within the sample.
- F- distribution is a sampling distribution and the variance between the samples should be divided by the variance with the samples for deriving the value of F-statistic.

$$F = \frac{\text{Variance between samples}}{\text{Variance within samples}} = \frac{\sigma_{\bar{X}}^2}{\mu_{S_i}^2}$$

- F – distribution is a distribution led by two degrees of freedom. That means the d.f between samples (d. f of numerator) is K-1 and d.f within the samples (d.f of denominator) is K(n-1)

K is the number of treatments (populations) and **n** the sample size.

- F – distribution is a right skewed distribution.
- The null hypothesis of F-test is  
 $H_0; \mu_1 = \mu_2 = \dots \mu_k$
- The alternative hypothesis of the F-test is  
 $H_1$  : at least one of the population means is significant.
- In order to compute the test statistic the application of ANOVA table would be more convenient. That is as follows .

| Source of variance                        | Sum of squares | d. f   | Mean squares               | F - value             |
|---|----------------|--------|----------------------------|-----------------------|
| Between samples<br>(between treatments)   | SSC            | k - 1  | $\frac{SSC}{k-1} = MSC$    | $F = \frac{MSC}{MSE}$ |
| Within the samples<br>(within treatments) | SSE            | k(n-1) | $\frac{SSE}{k(n-1)} = MSE$ |                       |
| Total                                     | SST            | Kn - 1 |                            |                       |

Calculations can be done on the following steps

Step 1 : Correction factor  $n = \frac{T^2}{N}$

$$\frac{T^2}{N} = \frac{(\sum X_1 + \sum X_2 + \dots + \sum_k)^2}{kn}$$

Step 2 : computing the sum of the squares of Total – (SST)

$$SST = (\sum X_1^2 + \sum X_2^2 + \dots + \sum X_k^2) - \frac{T^2}{N}$$

Step 3 : Computing the sum of squares between samples – SSC

$$SSC = \left[ \frac{(\sum R_{i1})^2}{n_1} + \frac{(\sum R_{i2})^2}{n_2} + \dots + \frac{(R_{ik})^2}{n_k} \right] - \frac{T^2}{N}$$

Step 4 : Computing the sum of squares within the samples  $SSE = SST - SSC$

Step 5 : Computing the mean square between the samples

$$MSC = \frac{SSC}{k-1}$$

Step 6 : computing the mean square within the samples.

$$MSE = \frac{SSE}{k(n-1)}$$

Step 7 : Computing the test statistic

$$F = \frac{MSC}{MSE}$$

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.1**: Studies the variations contained in a time dependent variable

**No. of periods** : 08

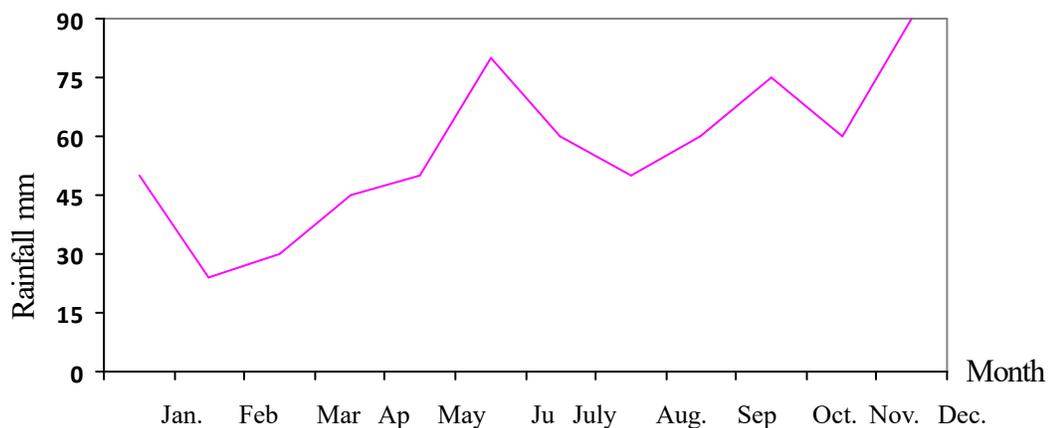
**Learning outcomes :**

- Explains what a time series is
- Interprets the time series
- Plots the function of the time series graphically
- Explains that calendar processing, price processing and population processing should be performed in data processing for time series analysis.
- States the uses of time series analysis
- Introduces the time series components such that the Trend, Seasonal movements, Cyclical movements and Irregular movements
- Provides appropriate instances for each movement

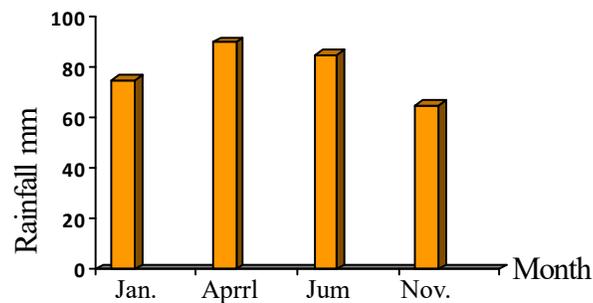
**Instructions for Lesson Planning :**

- Exhibit the following diagrams to the class and hold a discussion with the students.

Monthly rainfall values in Colombo city in the year 2016 (mm)



Months that heavy rainfall received in Kandy city in the year 2016



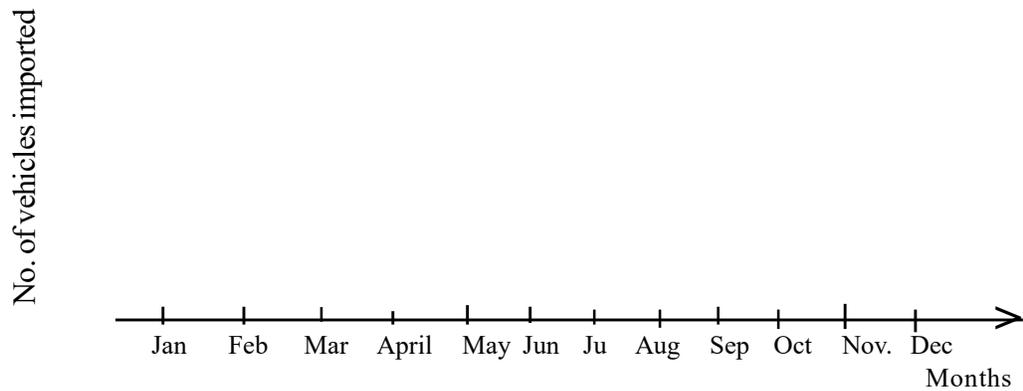
### Sales of meals in Dinusha Grocery

| Period   | Quantity of sugar sold (kg) |
|--|-----------------------------|
| 2015-January first week  | 32                          |
| 2015-February  | 54                          |
| From march 1 <sup>st</sup> to 30 <sup>th</sup> in 2015             | 80                          |
| From July 1 <sup>st</sup> to 30 <sup>th</sup> September, 2015      | 110                         |
| From September 31 <sup>st</sup> to 31 <sup>st</sup> December, 2015 | 90                          |

- Let the students learn that the data represented in the vertical bar chart are not consecutive monthly values.
- Point out that the data contained in the table have been collected for equal time intervals.
- Emphasize that the rain fall data in every month from January to December in 2016 are displayed in the linear graph.
- Point out that the data depicted through the Linear graph are more useful than the other two data sets.
- Point out that, since those data are a set of data collected in equal length and successive time intervals (as a series) it can be introduced as a time series.
- Number of HI-BRID cars, imported by Subagaman Co Ltd which is involved in importing and selling motor vehicles is mentioned in the following table.

| Month       | Jan | Feb | Mar | April | May | June | July | Aug | Sep | Oct | Nov | Dec |
|-------------|-----|-----|-----|-------|-----|------|------|-----|-----|-----|-----|-----|
| No.of cares | 14  | 10  | 18  | 25    | 22  | 18   | 16   | 19  | 21  | 28  | 30  | 26  |

Number of HI-BRID cars, imported by Subagaman Co Ltd – 2016



- Point out that the relevant time unit (Independent variable) is represented on the horizontal axis where as the considered time series variable is on the vertical axis.
- Point out that the origine of a time series graph is not necessary to be zero and also that keeping a false base line is not necessary, once the axes are drawn so as to be intersected by one another.
- Present the following table containing the number of days that a particular trade center had been open in each month and the monthly income of that trade center in the year 2016.

| Month     | No. of days through which the shop was open | Monthly income (Rs.000') |
|-----------|---|--------------------------|
| January   | 27  | 180                      |
| February  | 23  | 125                      |
| March     | 26  | 170                      |
| April     | 20  | 240                      |
| May       | 26  | 160                      |
| June      | 25  | 150                      |
| July      | 26  | 172                      |
| August    | 26  | 168                      |
| September | 24  | 158                      |
| October   | 26  | 185                      |
| November  | 25  | 174                      |
| December  | 26  | 172                      |

- Emphasize that the number of days through which the shop had been kept open is not equal in every month.
- Hence, point out that it would be unfair to compare that the income in February is less than the income in the month of January.
- Point out that processing the data before being analyzed having attached an equal number of days for every month is more appropriate.
- Ensure that the number of days through which the shop had been open is 300.
- Lead the students to find the number of days that the shop had been open in a month supposing that the shop had been open in every month for equal number of days.  

$$\frac{300}{12} = 25$$
- Accordingly guide them to add the income of last two days in January to the income of February.
- Point out that the income of last two days in January is supposed to be Rs. 30 000/- then the total income in February should be Rs. 155 000/-
- Point out that the income of the last day in March and the income of first four days in May should be considered as the income of April.
- State that processing of time series data in this manner is called **calendar processing**.
- Present to the class the following table containing the annual sales income of a company and the corresponding consumer price index.

| Year | Sales Income<br>Rs. Millions | Consumer<br>Price Index |
|------|------------------------------|-------------------------|
| 2008 | 82.0                         | 100                     |
| 2009 | 99.0                         | 110                     |
| 2010 | 122.5                        | 125                     |
| 2011 | 148.4                        | 140                     |
| 2012 | 172.5                        | 150                     |
| 2013 | 196.8                        | 164                     |
| 2014 | 230.4                        | 180                     |
| 2015 | 274.4                        | 196                     |
| 2016 | 342.0                        | 225                     |
| 2017 | 384.0                        | 240                     |

- Point out that the consumer price index has also increased even though the annual sales income has rapidly increased.
- Point out the fact that if the influence of price hikes in economy on increasing of income is removed, the real income can be comprehended.
- Lead the students to deflate each annual income from the corresponding price index.

| Year | Real value of Sales<br>Income Rs. Millions |
|------|--|
| 2008 | 82   |
| 2009 | 90   |
| 2010 | 98   |
| 2011 | 106  |
| 2012 | 115  |
| 2013 | 120  |
| 2014 | 128  |
| 2015 | 140  |
| 2016 | 152  |
| 2017 | 160  |

Real income of 2009

$$= \frac{99}{110} \times 100$$

$$= \underline{\underline{90}}$$

- Point out that deriving the real income, removing the inflationary effects included in time series variables considered is known as **price processing**.
- Discuss with the students that the business income may increase unexpectedly if the number of customers coming to the trade centre increases during a particular week or a month exceeding the normal customers moving about in the area due to any social, cultural, political or religious cause.
- Discuss with the students that the sales income of the school canteen increases in a week, during which many students and parents come to see a medical exhibition held in the school premises.
- Point out that when considered the time series of weekly income of that canteen the income in that particular week is highlighted, reporting a large income and that it would be better to distribute the excess income received in that week beyond the average income among few weeks to either side based on an appropriate way.
- State that processing of data in that manner is known as the **data processing for population changes**. (population processing)

| Week | Income of the canteen<br>Rs. 000' |
|------|-----------------------------------|
| 1    | 20                                |
| 2    | 30                                |
| 3    | 120                               |
| 4    | 23                                |
| 5    | 27                                |

- The excess income earned during the 3<sup>rd</sup> week can be computed as follows.  
Receiving the average income of the four weeks to either direction from that period.

$$\frac{20+30+23+27}{4} = 25$$

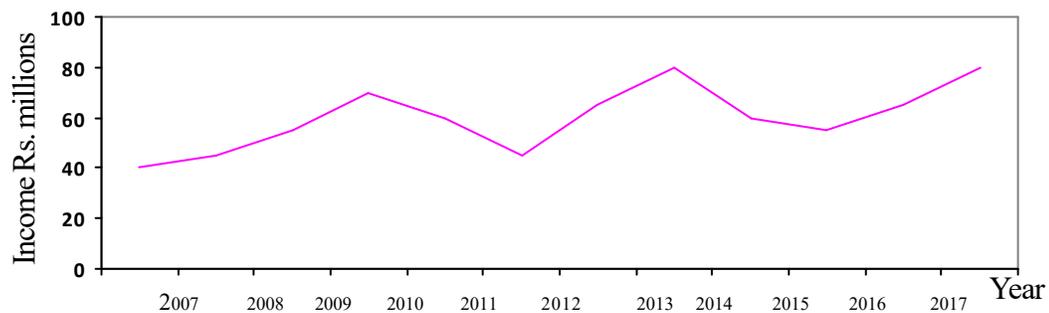
- Assuming that this average income could be the normal income during the 3<sup>rd</sup> week the additional income can be computed as Rs. 120 000 – Rs. 25 000 = Rs. 95 000. Then this additional income of Rs. 95 000 can be distributed equally throughout the five weeks.

Accordingly point out that the weekly income can be derived as follows.

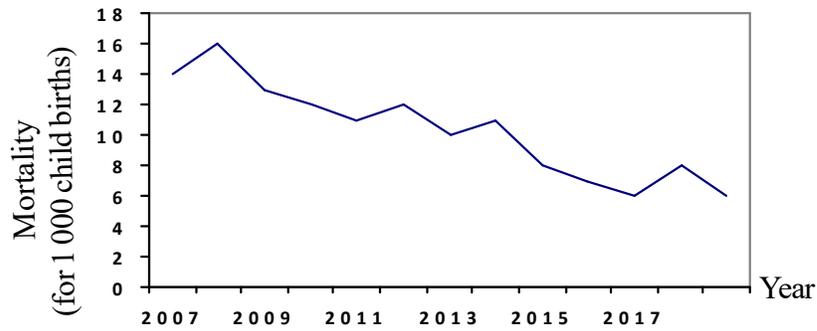
|                      |         |              |
|----------------------|---------|--------------|
| 1 <sup>st</sup> week | 20 + 19 | = Rs. 39 000 |
| 2 <sup>nd</sup> week | 30 + 19 | = Rs. 49 000 |
| 3 <sup>rd</sup> week | 25 + 19 | = Rs. 44 000 |
| 4 <sup>th</sup> week | 23 + 19 | = Rs. 42 000 |
| 5 <sup>th</sup> week | 27 + 19 | = Rs. 46 000 |

- (1) Demonstrate the following diagrams before the class and hold a discussion.

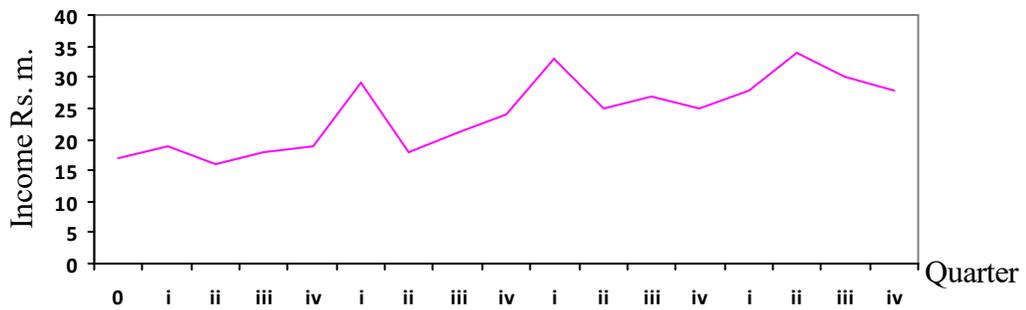
Sales income of ABC company for 12 years



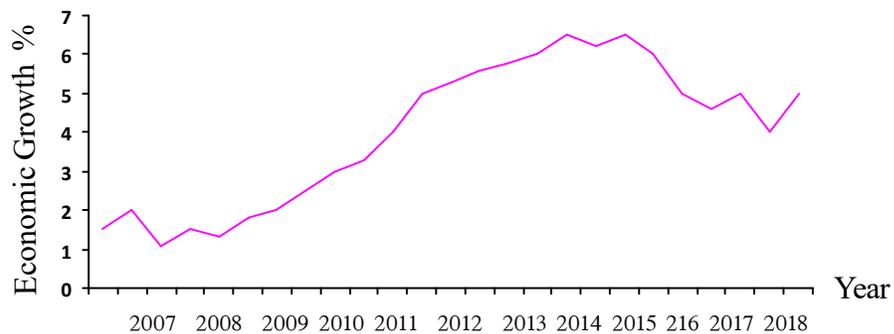
(2) Mortality rate (mother and infant deaths) for 1 000 child births in Mangalagama division (2007-2017)



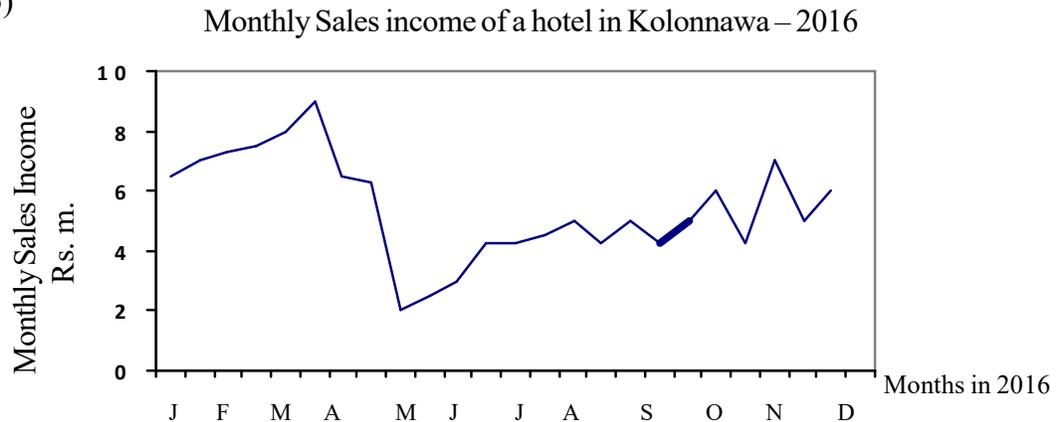
(3) Quarterly income of Suvasala Textile (2007-2017)



(4) Annual Economic Growth Rate of A-country 2007-2018



(5)



- Ascertain that there is an increasing trend in the sales income of ABC company from the year 2007 - 2017 according to the graph No. 1.
- Ascertain that there is a declining trend in mortality rate of child birth from 2007-2017 according to the graph No-2.
- Point out that the direction of the time series movement in the long run despite the short term fluctuations is known as the **long term Trend (Secular Trend)**
- Highlight that the sales income of 'Suvasalu' textile in the second quarter of every year has increased as shown in the graph No-3.
- Such patterns highlighted in the movement of a time series variable repeatedly once in equal length time intervals are called seasonal movements.
- With reference to the graph No 04 ascertain the fact that the period 2010-2014 is a progressive era in the economy of country A showing a greater economic growth.
- Point out that such an oscillation appearing in the long run of a time series variables prevailing for a period of more than one year is known as a **cyclical movement**.
- Point out that the duration from 2010-2014 is a progressive period in the economy of that country and such an experience is caused by the implementation of a good economic policy as well as governing the country by a populist regime
- Ascertain that the income of the hotel in the month of May has sudden dropped down as shown in graph-5.
- Point out that such an unexpected variable is called an **irregular movement** in a time series.
- Point out that the server flood that attacked this area in May 2016 should have caused for this drop out.
- Discuss the uses of time series analysis with the students.

### A guideline to explain the subject matters :

- A set of values taken by a time dependent variable in equal and successive time intervals is known as a 'Time Series.'
- If a time dependent variable denoted by  $y$  takes the values  $y_1, y_2, y_3 \dots y_k$  in equal and successive time intervals  $t_1, t_2, t_3 \dots t_k$  respectively then  $y_1, y_2, y_3 \dots y_k$  is called a time series.
- Since  $Y$  is a time dependent variable  $Y$  is considered as a function of time  $Y = f(t)$ .
- Before a time series being analysed the original data should have been processed as follows.
  1. Calendar processing
  2. Price processing
  3. Population processing
- Once monthly data have been collected for calendar months, since the number of days that a business has been involved in its operations processing of data attaching an equal number of operating days for every month before being analysed is known as **calendar processing**.
- Before an inflationary situation in an economy, the monetary value of the variables such as business income, manufacturing cost etc. is reported in greater values due to continuous price hikes. Then processing of such data, removing the influence of price hikes by deflating the time series using a suitable index number is known as **price processing**.
- The following formula is used to deflate the original time series data.

$$\text{Deflated data} = \frac{\text{Original time series value}}{\text{Corresponding price index}} \times 100$$

- Considering the changes appearing in a time series due to the changes experienced in the movement of the people (consumers of a particular business) in an area during a certain period caused by any social, cultural, political or religious factor, removing those influences on the time series variable, smoothening the data throughout a considerable period to either direction is known as **population processing**.
- All the factors causing for the movement of a time series can be classified as 4 time series components as follows.
  1. Long term trend – T  
(secular trend)

2. Seasonal movements – S
3. Cyclical Movements - C
4. Irregular Movements - I

- The overall direction of a time series variable that has moved in the long run despite the short term fluctuations is known as **long term trend** or **secular trend**.
- The variations take place in a time series variable repeatedly once in an equal length interval during a period of less than one year are known as **seasonal movements**. Such variations can be expected owing to the social and cultural changes in a country. Few examples are as follows.
  - Textile business in our country in April and December
  - The demand for Vesak greeting cards and decorating materials in May
  - Sales of umbrellas in rainy season.
- The oscillations experienced over a period of more than one year in the secular trend of a time series are known as **cyclical movements**.
- Such oscillations can be expected in a time series in practical situations such as an ongoing civil war exercised in the country for a period of more than one year or implementation of populist economic policies under good governance carried out by a newly empowered regime etc.
- A sudden steeper calamity or depression experienced unexpectedly in a time series variable is known as an **irregular movement**.
- A business can enjoy the following uses through time series analysis
  - Ability to evaluate the current situation of the variable through studying the past movement of it
  - Ability to understand various patterns obvious in fluctuations of the variable
  - Ability to understand the influence of each component separately towards the entire movement of the time series variable
  - Get an opportunity to contribute to the future planning
  - Ability to forecast the future trend of the particular time series variable
  - Ability to achieve at optimal conclusions comparing with identical time series in the relevant field or in other firms

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.2** : Studies the models used for time series analysis

**No. of periods** : 08

**Learning outcomes :**

- Names the models used to analyse a time series.
- Introduces the additive model.
- Describes how to separate the components in accordance with the additive model.
- Explains the situations where an additive model is used.
- Introduces the multiplicative model.
- Describes how to separate the components in accordance with the multiplicative model.
- Explains the situations where the multiplicative model is used.
- Explains that there are situations consist of characteristics in both additive and multiplicative models.

**Instructions for Lesson Planning :**

- Depict the following models before the class and hold a discussion.

$$\hat{y} = \beta_0 + \beta_1 x$$
$$y = G + C + I + (x - m)$$
$$A = E + L$$
$$V = I.R$$
$$P(X = x) = {}^n C_x p^x q^{n-x}$$
$$P(A \cup B) = P(A) + P(B) - P(A).P(B)$$
$$P(A \cap B) = P(A) + P(B)$$

- Highlight through a discussion that the various relations existing among /between the terms on the right hand side are used in order to determine the value of the variable given on the left hand side of the above mentioned models.
- Guide the students to separate these models into two categories as the models through which the total value of the variable is derived by adding the terms and as the models through which the total value of the variable is derived by multiplying the terms.
- Point out that such a model can be constructed for the total value of a time series **Y** – using the four time series components that the secular trend – T, seasonal movements – S, cyclical movement – C and irregular movements – I and also that applying such a model is more suitable for analyzing the real movement of a time series variable.

- Point out that a model through which the total value of a time series variable is derived by adding the components in it, would be better named as an additive model and lead the students to build up that model.

$$Y = T + S + C + I$$

- Lead the students to make any given term the subject in this model.

$$\text{Ex: } y - T - C - I = S$$

$$Y - (T + C + I) = S$$

- Inquiring from the students regarding practical situations when an additive model is applicable (Guide to expand their boundaries of concentration in other subjects like Economics and Accounting etc...)

- In Accounting

$$\text{Assets} = \text{Equity} + \text{Liabilities}$$

- In Economics

$$\text{National expenditure} = \text{government expenses} - G$$

$$+ \text{consumer expenses} - C$$

$$+ \text{investments} - I + (\text{exports } (x) - \text{imports } (m))$$

$$Y = G + C + I + (x - m)$$

- Derive the above relations from the students themselves through the discussion held with them.
- Point out that a model through which the total value of a variable is derived as a product of the components in it. It would be better named as a multiplicative model and lead the students to build up that model.

$$\text{Ex: } y = T \cdot S \cdot C \cdot I$$

- Guide the students to make any given term the subject in this model.

$$\boxed{\frac{y}{T.C.S} = I}$$

- Divide the students in the class into four groups and involve them in each activity given below.

Group – 1

- The total value of the variable in a certain year in the time series of the annual income of a company is

$$Y = \text{Rs. } 2\,754\,000$$

Suppose that it has been modeled  $y = T \cdot S \cdot C \cdot I$

If  $S = 1.25$   $C = 1.08$  and  $I = 0.85$

Find the value of T.

Group – 2

The total value of the variable in a certain year in the time series of the annual income of a company is Rs. 2 754 000

Suppose that it has been modeled using the multiplicative model  $y = T.S.C.I$  where.

$S = 1.25$   $I = 0.85$   $T = \text{Rs. } 2\,400\,000/-$  Find the value of C.

Group – 3

The total value of the variable in a certain year in the time series of the annual income of a company is Rs. 2 754 000

Suppose that it has been modeled using the multiplicative model as  $y = T.S.C.I$  where,

$T = \text{Rs. } 2\,400\,000$   $C = 1.08$   $I = 0.85$  find the value of – S

Group – 4

The total value of the variable in a certain year in the time series of the annual income is Rs. 2 754 000.

$S = 1.25$   $C = 1.08$  and  $T = 2\,400\,000$  find the value of – I

Solution - I

$$y = T \cdot S \cdot C \cdot I$$

$$2754 = T \times 1.25 \times 1.08 \times 0.85$$

$$\therefore T = \frac{2754}{1.25 \times 1.08 \times 0.85}$$

$$T = 2400$$

The trend is Rs. 240 000

Solution - II

$$y = T \cdot S \cdot C \cdot I$$

$$2754 = 2400 \times 1.25 \times C \times 0.85$$

$$\therefore C = \frac{2754}{2400 \times 1.25 \times 0.85}$$

$$C = \underline{\underline{1.08}}$$

Solution - III

$$Y = T \cdot S \cdot C \cdot I$$

$$2754 = 2400 \times S \times 1.08 \times 0.85$$

$$\therefore S = \frac{2754}{2400 \times 1.08 \times 0.85}$$

$$S = \underline{\underline{1.25}}$$

Solution - IV

$$Y = T \cdot S \cdot C \cdot I$$

$$2754 = 2400 \times 1.25 \times 1.08 \times I$$

$$\therefore I = \frac{2754}{2400 \times 1.25 \times 1.08}$$

$$I = \underline{\underline{0.85}}$$

- Guide the students to select the models with the characteristics in both the models additive and multiplicative from the list of models given at the beginning.

1.  $\hat{y} = \beta_0 + \beta_1 x$
2.  $P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$

- $Y = T \cdot S \cdot C \cdot I$  is the multiplicative model.
- Here only the trend (T) can be stated in natural numbers (absolutely) but the value of the other three components are stated as index numbers (relatively).
- The value of each component of this model can be derived as follows.

$$S = \frac{y}{T \cdot C \cdot I}$$

- Point out that the models appearing both additive and multiplicative characteristics also can be built up related to various subject fields.

$$T_n = a + (n-1) d$$

(This is the equation used to find the  $n^{\text{th}}$  term in an arithmetic progress in Mathematics)

$$S_n = \frac{n}{2} \{2a + (n-1)d\}$$

(is the equation used to find the sum of 'n' number of terms in an arithmetic progression).

- Inquire about the practical situations where the multiplicative model is used.

Ex: (i) Distance = speed x Time

(ii) Potential difference (voltage) = current x resistance

$$V = IR$$

(iii) Mass = volume x density

$$M = v \times d$$

- Point out that once the influence of each component is independent from one another, the addition model is better to be applied.
- Point out that once the influence of each component depends on one another, the multiplicative model is better to be applied.
- Ascertain that the multiplicative model is more appropriate in the business field, since most of the variables in this field are not moving independently.

#### **Guideline to explain the subject matters :**

- There are two models which are used in time series analysis as;
  - (i) Additive model
  - (ii) Multiplicative model
- Stating that the total value of the time series variable is derived as the sum of the four components such as the Trend (T), Seasonal movements (S) Cyclical movements (C) and Irregular movement (I) is the Additive model.

$$Y = T + S + C + I$$

- The value of one of these components can be derived as follows.

$$Y - (T + C + I) = S$$

- Stating that the total value of the time series variable is derived as the product of the four time series components is the Multiplicative Model.

$$Y = T \cdot S \cdot C \cdot I$$

- The value of any of these components can be derived as follows.

$$S = \frac{Y}{T \cdot C \cdot I}$$

**Competency 9.0 :** Analyses the time dependent variables and forecasts

**Competency level 9.3 :** Uses the Free Hand Method to estimate the trend

**No. of Periods :** 02

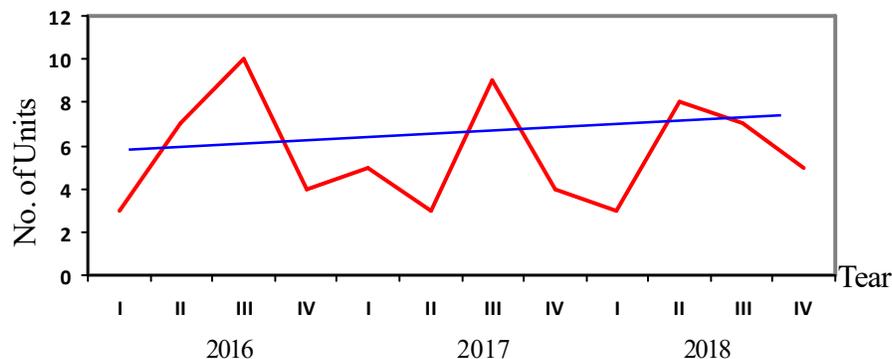
**Learning outcomes :**

- Explains the Free Hand Method as a technique of estimating the trend
- Draws out the trend line on the time series graph for given data
- Explains the merits and demerits of deriving a trend line on free hand method

**Instructions for Lesson Planning :**

- Involve the students in the following activity.
- The number of units produced quarterly by a manufacturing firm is given below.

| Year                  | 2016 |    |     |    | 2017 |    |     |    | 2018 |    |     |    |
|-----------------------|------|----|-----|----|------|----|-----|----|------|----|-----|----|
| quarter               | I    | II | III | IV | I    | II | III | IV | I    | II | III | IV |
| No. of units produced | 3    | 7  | 10  | 4  | 5    | 3  | 9   | 4  | 3    | 8  | 7   | 5  |



- Guide the students to draw a straight line so as to fall the points equally on either side of the line.
- Explain that deriving a straight line so as to fall the points equally on either side of the line with the purpose of getting a rough idea about the trend without using any mathematical basis is known as the free hand method.
- Explain that this free hand method is useful for making quick decisions in a short while to study the movement of a particular variable and further point out the merits and demerits of this method.

### **A Guideline to explain the subject matters :**

- Once a rough idea about the trend is to be derived quickly the trend line is drawn using the free hand method.
- After the time series graph has been constructed, deriving a trend line so as for the points on it to fall on either side of the line as far as possible is known as free hand method.
- Merits of the free hand method.
  - Not to having calculations
  - Ability to have an awareness about the trend quickly and conveniently.
  - Ability of making approximate forecasts
  - A specific skill not being necessary for construction of the line.
- Demerits of the free hand method
  - Being a subjective method
  - The trend line derived not being identical
  - Possible predictions not being reliable

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.4** : Uses the semi average method to estimate the trend

**No. of periods** : 04

**Learning outcomes** :

- interprets the ‘ semi-average’ method.
- Describes how to derive the trend line on semi-average method.
- Derives the semi-average values in a time series consists of an odd number of observations using an appropriate example.
- Derives the semi-average values in a time series consists of an even number of observations.
- Constructs the trend line using the derived semi-average values.
- Constructs the equation of the trend line.
- Makes estimations using the line or equation.
- Points out the merits and demerits of this method.

**Instructions for Lesson Planning :**

- engage the students in the following activity to explain the semi-average method as another method of deriving the trend.

**Activity 01 :**

- The production of the company – X - from 2005 – 2015 (in thousands of units) is mentioned below.

| Year                  | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|-----------------------|------|------|------|------|------|------|------|------|------|------|------|
| No.of prodction units | 1.4  | 1.1  | 1.2  | 1.4  | 2.1  | 2.2  | 1.6  | 1.5  | 2.0  | 2.5  | 3.0  |

- Lead the students to draw the time series graph to represent the data given in the table.
- Lead the students to derive two points (two co-ordinates) to derive the trend using a straight line.
- Provide with instructions to divide the table in to two equal parts and to compute the mean separately for each part.
- Point out that the middle period is ignored in dividing the table when there is an odd number of observations.
- Guide them to plot the received means corresponding to the middle period in each part.

- Lead them to plot the two corresponding mean values on the time series graph and to draw straight line passing through those points.
- Guide them to find the gradient and the intercept of that straight line and derive the equation on the model  $y = \beta_0 + \beta x$ . Hence lead them to derive the trend value of the year 2016.

Solution (1)

| Year | (x) | No. of units produced |
|------|-----|-----------------------|
| 2005 | 1   | 1.4                   |
| 2006 | 2   | 1.1                   |
| 2007 | 3   | 1.2                   |
| 2008 | 4   | 1.4                   |
| 2009 | 5   | 2.1                   |
| 2010 | 6   | 2.2                   |
| 2011 | 7   | 1.6                   |
| 2012 | 8   | 1.5                   |
| 2013 | 9   | 2.0                   |
| 2014 | 10  | 2.5                   |
| 2015 | 11  | 3.0                   |

$\frac{7.2}{5} = 1.44$

$\frac{10.6}{5} = 2.12$

- Since there is an odd number of years in this problem the value for the year 2010 is left out.
- The time series graph and the trend line on semi-average method are as follows.

Number of units produced (000)



Time (years)

- To find the trend value in 2016

$$\hat{y} = \beta_0 + \beta x$$

$$\begin{aligned} \text{Gradient} &= \frac{2.12 - 1.44}{6} \\ &= 11 \end{aligned}$$

$$\begin{aligned} \text{Intercept} \quad \hat{y} &= a + bx \\ 1.44 &= a + 0.11x \\ 1.44 &= a + (0.11 \times 3) \\ a &= \underline{\underline{1.11}} \end{aligned}$$

$$\text{Trend equation} \quad \hat{y} = 1.11 + 0.11x$$

$$\begin{aligned} \text{Forecasting for 2016} \quad \hat{y} &= 1.11 + (0.11 \times 12) \\ \hat{y} &= \underline{\underline{2.43}} \quad (\text{Thosand}) \end{aligned}$$

**Activity 2 :**

- Time series data related to the annual production of a company are given in the following table.

| Year                          | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|-------------------------------|------|------|------|------|------|------|------|------|------|
| production units<br>(Thosand) | 45   | 55   | 63   | 65   | 72   | 84   | 83   | 87   | 90   |

- Derive the trend line using the semi-average method.
- Forecast the trend value for the year 2019 using both the straight line and the equation.

The trend line on semi-average method

No. of units produced

Years

- Explain that there are merits and demerits in semi-average method.

Solution 2 :

| Year | X | Production |
|------|---|------------|
| 2009 | 1 | 45         |
| 2010 | 2 | 55         |
| 2011 | 3 | 63         |
| 2012 | 4 | 65         |
| 2013 | 5 | 72         |
| 2014 | 6 | 84         |
| 2015 | 7 | 83         |
| 2016 | 8 | 87         |
| 2017 | 9 | 90         |

$\left. \begin{array}{l} 2009 \\ 2010 \\ 2011 \\ 2012 \end{array} \right\} \frac{228}{4} = 57$   
 $\left. \begin{array}{l} 2014 \\ 2015 \\ 2016 \\ 2017 \end{array} \right\} \frac{344}{4} = 86$

- Gradient on algebraic method of deriving the trend  $y = a + bx$

$$\begin{array}{ll} \text{Gradient} & \hat{b} = \frac{86 - 57}{5} \\ & \hat{b} = \frac{29}{5} \\ & \hat{b} = \underline{\underline{5.8}} \end{array} \quad \begin{array}{ll} \text{Intercept} & y = a + bx \\ & 57 = a + 5.8 \times 2.5 \\ & 57 = a + 14.5 \\ & 57 - 14.5 = a \\ & a = \underline{\underline{42.5}} \end{array}$$

- Annual production in 2019 according to the trend line

$$\hat{y} = 42.5 + 5.8x$$

$$\hat{y} = 42.5 + 5.8x$$

$$\hat{y} = 42.5 + 5.8 \times 11$$

$$\hat{y} = \underline{\underline{106.3}}$$

**A Guideline to explain the subject matters :**

- The method of deriving a trend line by dividing the time series into two equal parts, computing the mean of each part, plotting those means on the time series graph and joining them with a straight line is known as the semi-average method.
- When there is an odd number of data the value in middle period is ignored.
- Once the time series is separated into two equal parts, when there is a number of data in each part, the mean value does not fall with respect of an exact period, but between two periods.
- Using the trend line derived on semi-average method as well as the algebraic equation the trend value can be forecast.
- Advantages of the semi-average method are'
  - Ability of deriving the trend value conveniently
  - Being an identical line
  - Not being subjective
  - Ability to forecast the trend approximately.

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.5** : Uses the Least Square method to estimate the trend

**No. of periods** : 08

**Learning outcomes :**

- Introduces the Least Square Method.
- Estimates the parameters of the trend line using the least square method.
- Derives the equation of the trend line using the Least Square Method.
- Constructs the trend line on the least square method on the time series graph.
- Forecasts the trend using the trend line or the equation.
- Derives the trend line equation shifting the origine of the trend line.
- Explains the advantages and disadvantages of the Least Square Method.

**Instructions for Lesson Planning :**

- Draw the attention of students for the following pair of equations.

$$\sum y = n\beta_0 + \beta_1 \sum x$$

$$\sum xy = \beta_0 \sum x + \beta_1 \sum x^2$$

- Discuss with the students recalling how they learnt to derive the values of  $\beta_0$  and  $\beta_1$  with substitution of X and Y values in the above pair of normal equations under the regression analysis using the Least Square method.
- Explain that the values of the time unit (x) are assigned in a manner that  $\sum x = 0$  since there are equal lengths and successive time intervals considered in the time series analysis hence point out that the gradient and intercept of the trend line can be derived conveniently.
- Write the two formulae which were used in regression analysis to derive the gradient and intercept on the board.

$$\beta_1 = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

$$\beta_0 = \bar{y} - b\bar{x}$$

- Point out to the students that once  $\sum x = 0$  is substituted in the above formula it is simplified as follows.

$$\begin{aligned}\beta_1 &= \frac{n \sum xy - (0 \times (\sum y))}{n \sum x^2 - (0)^2} \\ &= \frac{n \sum xy}{n \sum x^2} \\ &= \frac{\sum xy}{\sum x^2}\end{aligned}$$

$\therefore$

$$\begin{aligned}\beta_0 &= \bar{y} - b\bar{x} \\ &= \bar{y} - [b \times 0] \\ &= \bar{y}\end{aligned}$$

$$\therefore \hat{B}_0 = \frac{\sum y}{\underline{\underline{n}}}$$

- Assign the following activity to the students and derive the equation of the trend-line discussing with them.

The annual sales income of **xy** company from the year-2007 to 2017 is given in the following table (Rs. millions)

| Year | Sales income<br>Rs. millions - $y$ | $x$ | $xy$  | $x^2$ |
|------|------------------------------------|-----|-------|-------|
| 2007 | 108                                | -5  | -540  | 25    |
| 2008 | 107                                | -4  | -428  | 16    |
| 2009 | 98                                 | -3  | -294  | 9     |
| 2010 | 88                                 | -2  | -176  | 4     |
| 2011 | 86                                 | -1  | -86   | 1     |
| 2012 | 84                                 | 0   | 0     | 0     |
| 2013 | 82                                 | 1   | 82    | 1     |
| 2014 | 73                                 | 2   | 146   | 4     |
| 2015 | 65                                 | 3   | 195   | 9     |
| 2016 | 61                                 | 4   | 244   | 16    |
| 2017 | 50                                 | 5   | 250   | 25    |
|      | 902                                | -   | - 607 | 110   |

- Assign serial values for 'x' as  $\sum x = 0$
- Compute the values of  $\sum xy$  and  $\sum x^2$
- Derive  $\sum xy$  and  $\sum x^2$
- Derive the values using the formulae.

$$\beta_1 = \frac{\sum xy}{\sum x^2} \quad \text{and} \quad \beta_0 = \frac{\sum y}{n}$$

$$\beta_1 = \frac{-607}{110} \quad \beta_0 = \frac{\sum y}{n}$$

$$= -5.5 \quad = \frac{902}{11}$$

$$= \underline{\underline{82}}$$

Derive the trend line equation as  $\hat{y} = 82 - 5.5x$

- Explain that this trend line is known as the Least Square trend line, since the sum of the squares of the errors is minimized.
- Explain that the corresponding value of  $\mathbf{X}$  for the year 2018 is 6, pointing out how the values have been assigned for  $\mathbf{X}$ .
- Hence lead them to derive the trend value for the year 2018.
- Explain that how to derive the trend line equation focusing on the origine to the very middle year in a time series with an odd number of years was exercised in solving the above mentioned problem.
- Point out that when there are even number of data, the origine of  $\mathbf{X}$  can not be focused exactly right to a single year and then it should be focused to the midpoint of an year. Then, further point out that half of an year is taken as one unit of  $\mathbf{X}$  and the trend line should be derived accordingly
- Involve the students in solving the following problem.

$$\hat{y} = 82 - (5.5 \times 6)$$

$$= 82 - 33$$

$$= \underline{\underline{49}}$$

Activity 02 :

- The annual income of 8 years of a particular company is as follows. (in Rs. millions)  
Construct the least square trend line equation and plot the trend line on the time series graph.

| Year                            | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---------------------------------|------|------|------|------|------|------|------|------|
| Annual Income<br>(Rs. millions) | 2    | 3    | 4    | 4    | 5    | 7    | 8    | 10   |

Solution :

| Year | x  | y  | x <sup>2</sup> | xy  |
|------|----|----|----------------|-----|
| 2010 | -7 | 2  | 49             | -14 |
| 2011 | -5 | 3  | 25             | -15 |
| 2012 | -3 | 4  | 9              | -12 |
| 2013 | -1 | 4  | 1              | -4  |
|      | 0  |    |                |     |
| 2014 | 1  | 5  | 1              | 5   |
| 2015 | 3  | 7  | 9              | 21  |
| 2016 | 5  | 8  | 25             | 40  |
| 2017 | 7  | 10 | 49             | 70  |
|      |    | 43 | 168            | 91  |

$$\beta_1 = \frac{\sum xy}{\sum x^2}$$

$$\beta_0 = \frac{\sum y}{n}$$

$$\beta_1 = \frac{91}{168}$$

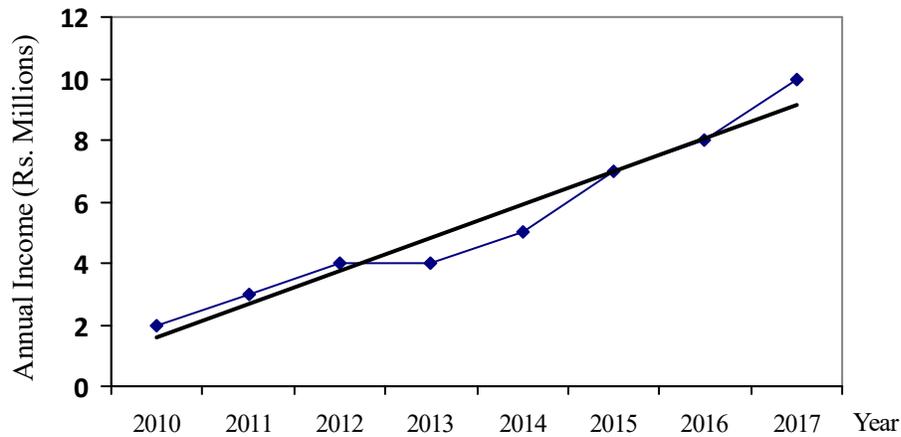
$$\beta_0 = \frac{43}{8}$$

$$\beta_1 = \underline{\underline{0.54}}$$

$$\beta_0 = \underline{\underline{5.375}}$$

The trend equation is

$$\hat{y} = \underline{\underline{5.375 + 0.54x}}$$



$$x = -5$$

$$y = 5.375 + (0.54 \times -5)$$

$$y = \underline{\underline{2.675}}$$

$$x = 5$$

$$y = 5.375 + (0.54 \times 5)$$

$$y = \underline{\underline{8.075}}$$

#### A Guideline to explain the subject matters :

- The trend line of a time series can be fit (constructed) following the least square method applied in regression analysis. Assigning the values for X (equal and consecutive time intervals) in a manner that  $\sum x = 0$   $\beta_0$  and  $\beta_1$  can be estimated.
- Accordingly the trend line can be derived using  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  having computed the intercept and gradient of the line as  $\hat{\beta}_0 = \frac{\sum y}{n}$  and  $\hat{\beta}_1 = \frac{\sum xy}{\sum x^2}$  respectively.
- Once the time series consists of an odd number of observations the origin of the trend line can be focused exactly right against the middle period (a single time unit).
- But, when the time series consists of an even number of observations the origin of X cannot be selected exactly right against a single time unit, and it should be focused in the middle between two periods. Then the trend line can be derived taking a unit of X in half (1/2) of the time unit considered.
- In that manner the method of estimating the trend line minimizing the sum of the squares of the errors  $\sum = (\hat{y}_i - \bar{y})^2$  is known as the Least Square Method.

Merits of the least square method.

- Receiving an identical trend line.
- Ability of forecasting further trend value.
- Receiving a rational trend line based on an advanced mathematical model.

Demerits :

- Calculations being complicated.
- Once the trend line is invalid the forecast values not being accurate.

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.6** : Uses the moving average method to estimate the trend

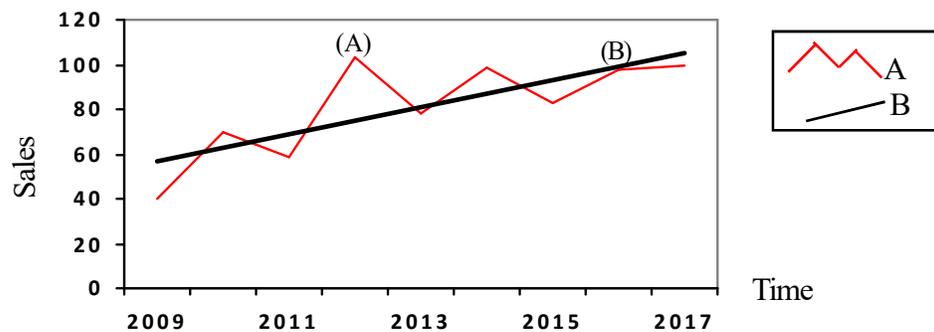
**No. of periods** : 08

**Learning outcomes :**

- Explains the moving averages for time series data.
- Derives moving averages separately when the order is odd and even.
- Explains the need of centered moving averages.
- Removes the seasonal and irregular movements of a time series using moving averages.
- Constructs a trend line using the moving average method for given data.
- Writes the merits and demerits of the moving average method.

**Instructions for Lesson Planning :**

- Hold a discussion highlighting the following facts.
  - Before a house being built, the ground is made a flat surface scattering the oscillations in soil.
  - Before the saplings are rooted on a vegetable bed, the soil in the bed is scattered into a smooth surface.
  - Before paddy saplings are rooted in the paddy field, the soil is stocked well with sufficient water supply and the mud is scattered into a smooth surface.
  - Explain that the moving average method is used to derive a smoothened time series removing the fluctuations in a time series
- Depict the following diagram on the board and point out that there are fluctuations in a time series (A) and draw out a smoothened time series removing those fluctuations using an order (B) in the same graph.



- Involve the students in the following activity.
- Sales of a particular company from the year 2009 to year 2017 are given in the table below.

| Year | Sales (Units - Thousands) |
|------|---------------------------|
| 2009 | 90                        |
| 2010 | 100                       |
| 2011 | 88                        |
| 2012 | 80                        |
| 2013 | 72                        |
| 2014 | 80                        |
| 2015 | 83                        |
| 2016 | 88                        |
| 2017 | 100                       |

- Guide the students to compute the trend on moving average method taking the order 3 and order 5 separately and draw both the trend lines on the same graph.
- Lead the students to compute the moving average of order 3 and then to complete the following table.

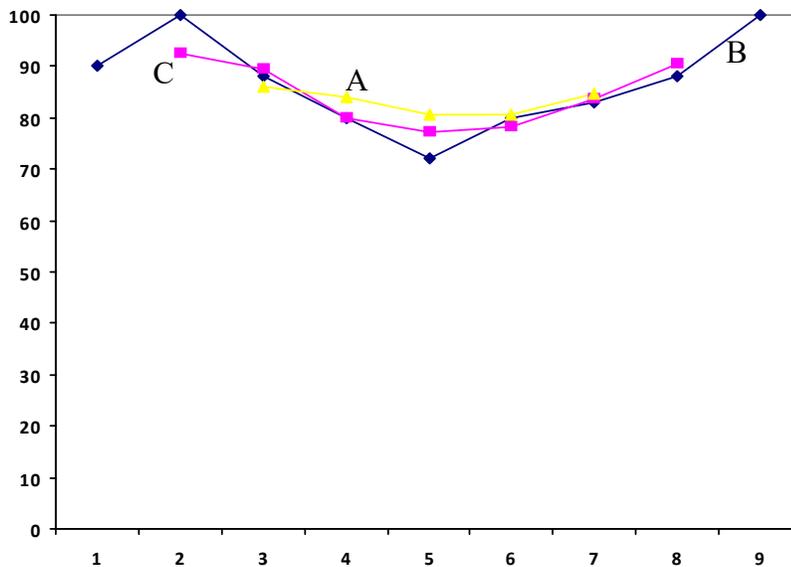
| Year | Sales units | Moving total<br>of 3 years | Moving average<br>of 3 years |
|------|-------------|----------------------------|------------------------------|
| 2000 | 90          | -                          | -                            |
| 2001 | 100         | 278                        | 92.66                        |
| 2002 | 88          | 268                        | 89.33                        |
| 2003 | 80          | 240                        | 80.00                        |
| 2004 | 72          | 232                        | 77.33                        |
| 2005 | 80          | 235                        | 78.33                        |
| 2006 | 83          | 251                        | 83.66                        |
| 2007 | 88          | 271                        | 90.33                        |
| 2008 | 100         | -                          | -                            |

- Lead the students to compute the moving average of order 5 and then to complete the following table.

| Year | Sales Units<br>(Thosands) | Moving total of<br>5 years | Moving averages<br>of 5 years |
|------|---------------------------|----------------------------|-------------------------------|
| 2009 | 90                        | -                          | -                             |
| 2010 | 100                       | -                          | -                             |
| 2011 | 88                        | 430                        | 86.0                          |
| 2012 | 80                        | 420                        | 84.0                          |
| 2013 | 72                        | 403                        | 80.6                          |
| 2014 | 80                        | 403                        | 80.6                          |
| 2015 | 83                        | 423                        | 84.6                          |
| 2016 | 88                        | -                          | -                             |
| 2017 | 100                       | -                          | -                             |

- Guide the students to plot both the trend lines of order 3 and order 5 separately on the same co-ordinate plane. (time series graph)
- Explain that the variations that were not obvious in original time series are appeared on these trend lines and time series has become more smooth as far as possible with greater the order.
- Further point out that the trend line on the moving average method is not sufficient for forecasting.

Graph – A



B - Graph of original data time series

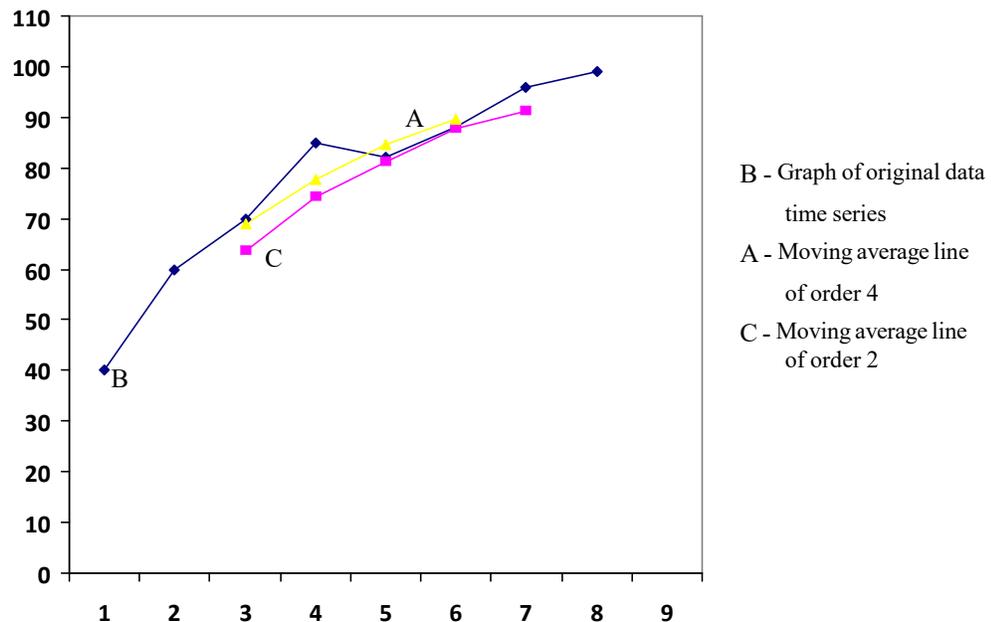
A - Moving average line of order 5

C - Moving average line of order 3

- Explain how to compute the centered moving averages to derive the trend line once the order is an even number using the following problem.
- The business of a particular company from the year 2010 to 2017 is given in the following table. Guide the students to compute the centered moving averages of order – 4.

| Year | Profit | Moving total<br>of 4 years | Moving average<br>of 4 years | Moving total<br>of 2 years | Moving average<br>of 2 years |
|------|--------|----------------------------|------------------------------|----------------------------|------------------------------|
| 2010 | 40     | -                          | -                            | -                          | -                            |
| 2011 | 60     | -                          | -                            | -                          | -                            |
| 2012 | 70     | 255                        | 63.75                        | 138.0                      | 69.0                         |
| 2013 | 85     | 297                        | 74.25                        | 155.5                      | 77.75                        |
| 2014 | 82     | 325                        | 81.25                        | 169.0                      | 84.5                         |
| 2015 | 88     | 351                        | 87.75                        | 179.0                      | 89.5                         |
| 2016 | 96     | 365                        | 91.25                        | -                          | -                            |
| 2017 | 99     | -                          | -                            | -                          | -                            |

- Guide the students to compute the moving total of order 4 and write them straight between two values and then to compute the average of each.
- After that guide the students to get the sum of those two successive moving average values straight between two of them and to get the corresponding centered moving average value dividing by – 2 .



Discuss the merits and demerits of the moving average method.

**A Guideline to explain the subject matters :**

- Once the time series data are given annually / quarterly / monthly/ daily deriving the trend values in moving average or centered moving averages using an appropriate order is known as the moving average method.
- In other words an artificial time series which is derived through the mean of the values of a given period, preceeding period and succeeding period with reference to annual / monthly/ quarterly/ weekly / daily time series data available for a considered time dependent variable is known as the trend line on moving average method.
- Here it is needed the fact that the mean of how many periods is taken (order), so that the ‘order’ is very important when the ‘moving average/ are interpreted.
- The numerical value that indicates the number of periods considered to be added for removing the short term fluctuations is the ‘order’.
- The moving average of order ‘n’ of a time series given as  $y_1, y_2, \dots, y_k$  are as follows.

$$\frac{y_1, y_2, \dots, + y_n}{n}, \frac{y_2 + y_3, \dots, y_{n+1}}{n}, \frac{y_3 + y_4, \dots, y_{n+2}}{n},$$

Hence it can be interpreted as a sequence of means.

- Once the order is an odd value, the moving average value is derived straight against a specific time unit.
- Once the order is an even value the first set of moving averages fall between two time units. By taking the mean of each of those two successive moving averages of centered moving averages can be derived straight right against each successive time unit.
- Although a more smooth trend line is derived once the order increases, the number of trend values lost from either end of the time series is also increasing.
- Since long term moving patterns of a time series variable can be separated from short term fluctuations through smoothening a time series, the moving average method is highly applicable and more successful.
- Merits of moving average method.
  - Being calculable simply and a convenient method.
  - Ability to derive more precise (realistic) trend values since cyclical and irregular movements are abolished.

- A smooth trend line being more practical than a straight linear trend.
- Demerits of the moving average method.
  - Trend values being lost from both ends of the time series according to the order.
  - Determining the 'order' being inconvenient once the cyclical patterns are irregularly arranged.
  - Possibility of appearing new movements which were not obvious in the original time series.

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.7** : Uses the average percentage method to estimate the seasonal indices.

**No. of periods** : 04

**Learning outcomes :**

- Explains the average percentage method for estimating the seasonal indices.
- States the steps of computing the seasonal indices on the average percentage method.
- Computes the seasonal indices on the average percentage method using the given data.
- Writes the merits and demerits of the average percentage method.

**Instructions for Lesson Planning :**

- Recalling the ‘seasonal component’ among the components of time series learnt in competency level 9.1, state that estimating the seasonal components is required in time series analysis.
- Divide the students into three groups and involve them in the following activity giving three attendance registers to each group.
- In order to estimate 5 seasonal indices to the attendance of each week day, draw a sample of four weeks of five working days.
- Compute the total value of attendance of the students in the first week
- Compute the average daily attendance of that week dividing that total value by the number of school days in the week.
- Compute the attendance on Monday as a percentage of the attendance of a day in that week.
- Derive the average percentage value of the attendance on Mondays of all four weeks in this manner and then add those four average percentage values and divide by 4 to derive a common average percentage for Monday.
- When considering the period of the week as the long time and each day as the short time, name the common value derived for Monday as mentioned above as the seasonal index of Monday.
- Compute the seasonal indices for Tuesday, Wednesday, Thursday and Friday in the same manner.
- Compute the total value of those five seasonal indices.
- Check whether the sum of those five indices is equal to 500.
- If the sum of those indices is not 500 divide each index of each day from their total and multiply by 500.

- Name those final values as the index number for each day.
- Comment on the attendance of the students in the class through five days comparing the seasonal indices of the five days in the same class and also comparing the seasonal index of the same day among different classes.
- Hold a discussion with the students highlighting the following facts.
  - Point out that monthly seasonal indices and quarterly seasonal indices can also be computed for the period of one year following the same way that the daily seasonal indices were computed for period of a week.
  - State that computing the seasonal indices following the above mentioned steps is known as the Average percentage Method.
  - Highlight that there are both merits and demerits of computing the seasonal indices on the average percentage method.
  - Compute the seasonal indices discussing with the students step by step involving in the following activity.

**Activity I:**

- The details of quarterly business income of a firm from 2012-2016 are given in the following table (Rs. millions).

| Year | Q <sub>1</sub> | Q <sub>2</sub> | Q <sub>3</sub> | Q <sub>4</sub> |
|------|----------------|----------------|----------------|----------------|
| 2012 | 30             | 40             | 36             | 34             |
| 2013 | 34             | 52             | 50             | 44             |
| 2014 | 40             | 58             | 54             | 48             |
| 2015 | 54             | 76             | 68             | 62             |
| 2016 | 80             | 92             | 86             | 82             |

- Compute the seasonal index for each quarter using the average percentage method.

Solution : Step – 1 Computing the average quarterly sales in each year.

$$2012 - \frac{30 + 40 + 36 + 34}{4} = \frac{140}{4} = 35$$

$$2013 - \frac{34 + 52 + 50 + 44}{4} = \frac{180}{4} = 45$$

$$2014 - \frac{40 + 58 + 54 + 48}{4} = \frac{200}{4} = 50$$

$$2015 - \frac{54+76+68+62}{4} = \frac{260}{4} = 65$$

$$2016 - \frac{80+92+86+82}{4} = \frac{340}{4} = 85$$

Step 2 :

- Expressing all the seasonal data as percentages of quarterly average sales of each year.

|      |                                    |                                     |                                     |                                    |
|------|------------------------------------|-------------------------------------|-------------------------------------|------------------------------------|
| 2012 | $\frac{30}{35} \times 100 = 85.71$ | $\frac{40}{35} \times 100 = 114.29$ | $\frac{36}{35} \times 100 = 102.86$ | $\frac{34}{35} \times 100 = 97.14$ |
| 2013 | $\frac{34}{45} \times 100 = 75.56$ | $\frac{52}{45} \times 100 = 115.56$ | $\frac{50}{45} \times 100 = 111.11$ | $\frac{44}{45} \times 100 = 97.78$ |
| 2014 | $\frac{40}{50} \times 100 = 80.00$ | $\frac{58}{50} \times 100 = 116.00$ | $\frac{54}{50} \times 100 = 108.00$ | $\frac{48}{50} \times 100 = 96.00$ |
| 2015 | $\frac{54}{65} \times 100 = 83.08$ | $\frac{76}{65} \times 100 = 116.92$ | $\frac{68}{65} \times 100 = 104.62$ | $\frac{62}{65} \times 100 = 95.38$ |
| 2016 | $\frac{80}{85} \times 100 = 94.12$ | $\frac{92}{85} \times 100 = 108.24$ | $\frac{86}{85} \times 100 = 101.18$ | $\frac{82}{85} \times 100 = 96.47$ |

Step 3 : • Compute the means of quarterly percentage values.

| Q <sub>1</sub> | Q <sub>2</sub> | Q <sub>3</sub> | Q <sub>4</sub> |
|----------------|----------------|----------------|----------------|
| 85.71          | 114.29         | 102.86         | 97.14          |
| 75.56          | 115.56         | 111.11         | 97.78          |
| 80.00          | 116.00         | 108.00         | 96.00          |
| 83.08          | 116.92         | 104.62         | 95.38          |
| 94.12          | 108.24         | 101.18         | 96.47          |
| <u>418.47</u>  | <u>571.01</u>  | <u>527.77</u>  | <u>482.77</u>  |
| <u>5</u>       | <u>5</u>       | <u>5</u>       | <u>5</u>       |
| 83.69          | 114.20         | 105.55         | 96.55          |

Step 4 : Deriving the seasonal indices of four quarters after making adjustments to bring the sum of quarterly percentage averages to 400

$$\text{Since } 83.69 + 114.20 + 105.55 + 96.55 = 399.99$$

taking the seasonal index of the 1<sup>st</sup> quarter as 83.70 it can be easily and very correctly adjusted equal to 400.

Accordingly

|   |   |               |
|---|---|---------------|
| Seasonal index of the 1 <sup>st</sup> quarter | = | 83.70         |
| Seasonal index of the 2 <sup>nd</sup> quarter | = | 114.20        |
| Seasonal index of the 3 <sup>rd</sup> quarter | = | 105.55        |
| Seasonal index of the 4 <sup>th</sup> quarter | = | 96.55         |
|   | = | <u>400.00</u> |

**A Guideline to explain the subject matters :**

- The steps mentioned below can be followed to compute the quarterly seasonal indices on average percentage method.

Step – 1 : Deriving the average value of quarterly data in each year.

Step – 2 : Computing the quarterly data as a percentage of the quarterly average value of the corresponding year.

Step – 3 : Deriving the average value of the quarterly average percentage values received for different years.

Step – 4 : Making adjustments to bring the sum of averages received in 3<sup>rd</sup> step to 400.

Merits of the average percentage method are as follows.

- This is a very convenient and simple method.
- If only the time series data do not contain long term trends, this method can be recommended as an appropriate method.

Demerits of the average percentage method are as follows.

- The theoretical basement that estimating seasonal indices after removing the trend, cyclical movements and irregular movements has been ignored.
- The indices derived in this method are not recommended for general analysis.

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.8** : Uses the moving average method to estimate seasonal indices.

**No. of periods** : 06

**Learning outcomes :**

- Introduces the moving average method which is used to estimate the seasonal indices.
- Describes the steps of computing seasonal indices on moving average method.
- Computes the seasonal indices using an example on moving average method.
- Describes the advantages and disadvantages of computing seasonal indices on moving average method.

**Instructions for Lesson Planning :**

- Discuss with the students that the trend, cyclical movements and irregular movements also may contain in seasonal indices computed on average percentage method, since it has been pointed out as the major weak point in it, that the seasonal indices not being computed having removed those movements.
- Hence point out the need of deriving the seasonal index removing those components step by step and inform the students that another method called moving average method can be applied to compute the seasonal indices.
- Divide the students in the class into three groups and involve them in the following activity providing with an attendance register to each group collected from three class rooms.
- Derive a sample of four weeks through which the school was held on five days in each in order to estimate only the seasonal indices for each working day in the week.
- List out the daily attendance data of each week vertically down
- Compute the moving average taking the order as 5 and write each average against the corresponding middle day of each set.
- Express the original time series data given, as a percentage of corresponding moving average derived at the above step.
- Separate those percentage values with respect to each day as Monday, Tuesday, Wednesday, Thursday and Friday.
- Compute the means of percentage values with respect of each day.
- Check whether the sum of the computed means is 500.
- If the sum of the means of five days is not 500, divide the mean of each day by the sum of the means of five days and multiply by 500.
- Name the derived values as the seasonal indices of each day.

- Comment on the attendance of the children on five days comparing the five seasonal indices of five days in the same class where as comparing the seasonal index of each day between different classes.
- Involve the students in the following activity to compute the quarterly seasonal indices

**Activity 02 :** You have been provided with following set of quarterly data of a time series.

| Year | Quater - 1 | Quater -2 | Quater - 3 | Quater - 4 |
|------|------------|-----------|------------|------------|
| 2012 | 68         | 62        | 61         | 63         |
| 2013 | 65         | 58        | 66         | 61         |
| 2014 | 68         | 63        | 63         | 67         |
| 2015 | 70         | 59        | 56         | 62         |
| 2016 | 60         | 55        | 51         | 58         |

- Compute the quarterly seasonal indices using the moving average method.

**Solution**

step -1 : Prepare a chart as shown below to compute the original data as a percentage of moving averages.

| Year | Quater | Original Data | Moving total of order-4 | Moving Average | Centred Moving Average | Percentage value |
|------|--------|---------------|-------------------------|----------------|------------------------|------------------|
| 2012 | 1      | 68            | -                       | -              | -                      | -                |
|      | 2      | 62            | -                       | -              | -                      | -                |
|      | 3      | 61            | 254                     | 63.50          | 63.125                 | 96.63            |
|      | 4      | 63            | 251                     | 62.75          | 62.250                 | 101.20           |
| 2013 | 1      | 65            | 247                     | 61.75          | 62.375                 | 104.21           |
|      | 2      | 58            | 252                     | 63.00          | 62.750                 | 92.43            |
|      | 3      | 66            | 250                     | 62.50          | 62.875                 | 104.97           |
|      | 4      | 61            | 253                     | 63.25          | 63.875                 | 95.50            |
| 2014 | 1      | 68            | 258                     | 64.50          | 64.125                 | 106.64           |
|      | 2      | 63            | 255                     | 63.75          | 64.500                 | 97.67            |
|      | 3      | 63            | 261                     | 65.25          | 65.500                 | 96.18            |
|      | 4      | 67            | 263                     | 65.75          | 65.250                 | 102.68           |
| 2015 | 1      | 70            | 259                     | 64.75          | 63.875                 | 109.59           |
|      | 2      | 59            | 252                     | 63.00          | 62.375                 | 94.59            |
|      | 3      | 56            | 247                     | 61.75          | 60.500                 | 92.56            |
|      | 4      | 62            | 237                     | 59.25          | 58.750                 | 105.53           |
| 2016 | 1      | 60            | 263                     | 58.25          | 57.625                 | 104.12           |
|      | 2      | 55            | 228                     | 57.00          | 56.500                 | 97.35            |
|      | 3      | 51            | 224                     | 56.00          | -                      | -                |
|      | 4      | 58            | -                       | -              | -                      | -                |

- Separating the values in percentage value column in the above table quarterly, the quarterly average percentage values can be derived as follows.

|       | percentage values to moving average |                |                |                |
|-------|-------------------------------------|----------------|----------------|----------------|
|       | Q <sub>1</sub>                      | Q <sub>2</sub> | Q <sub>3</sub> | Q <sub>4</sub> |
| 2012  | -                                   | -              | 96.63          | 101.20         |
| 2013  | 104.21                              | 92.43          | 104.97         | 95.50          |
| 2014  | 106.04                              | 97.67          | 96.18          | 102.68         |
| 2015  | 109.59                              | 94.59          | 92.56          | 105.53         |
| 2016  | 104.12                              | 97.35          | -              | -              |
| Total | 423.96                              | 382.04         | 390.34         | 404.91         |
| Mean  | 105.99                              | 95.51          | 97.585         | 101.23         |

- $105.99 + 95.51 + 97.585 + 101.23 = 400.315$   
 rounding off each percentage mean value to the nearest whole number, those can be adjusted equal to 400 as follows.  
 $106 + 95 + 98 + 101 = 400$
- Accordingly point out the fact that these values can be accepted as the seasonal indices of corresponding quarters.

| Quarter        | i   | ii | iii | iv  |
|----------------|-----|----|-----|-----|
| Seasonal index | 106 | 95 | 98  | 101 |

- Once the sum of the means of moving average percentage value is not equal to 400, it can be adjusted right as follows.

The correct seasonal index relevant to the quarter =  $\frac{\text{Mean of the moving average percentage}}{\text{Total value of those means}} \times 400$

Hence :

$$1. \text{ Seasonal index of the 1}^{\text{st}} \text{ quarter} = \frac{105.99}{400.315} \times 400 = 105.9067$$

$$2. \text{ Seasonal index of the 2}^{\text{nd}} \text{ quarter} = \frac{95.51}{400.315} \times 400 = 95.4348$$

$$3. \text{ Seasonal index of the 3}^{\text{rd}} \text{ quarter} = \frac{97.585}{400.315} \times 400 = 97.5082$$

$$4. \text{ Seasonal index of the 4}^{\text{th}} \text{ quarter} = \frac{101.23}{400.315} \times 400 = 101.1503$$

400.0000

Hold a discussion highlighting the following facts.

- Computing the seasonal indices in this manner is known as the Moving Average Method.
- This is a method of deriving the seasonal indices by removing the other components step by step.
- Once the moving averages are calculated the seasonal and irregular movements in time series data are removed and only the trend and cyclical movements remain.
- Once original data are expressed as a percentage of moving averages it can be expected that the trend and cyclical movements are removed while the seasonal and irregular movements remain.
- Once the mean of the percentages are computed only the seasonal movements can be expected to remain removing irregular movements as well.
- There are advantages as well as disadvantages of deriving the seasonal indices on the moving average method.
- Seasonal indices also can be computed for each month using the moving average method.

**A guideline to explain the subject matters :**

- The moving average method can be identified as a method of highlighting only the seasonal movements removing the trend, cyclical and irregular movements contained in time series data step by step.
- Through deriving moving averages or centered moving averages of time series data (y), only the trend and cyclical movements remain, removing seasonal and irregular movements. Once the original data are divided by moving averages only the seasonal and irregular movements remain. Expecting that only the seasonal movements will

remain removing irregular movements as well by computing the mean of corresponding quarterly values is the logical background of the moving average method.

- Once the patterns of seasonal movements of a time series are regular, a better index for evaluating those movements can be expected through moving average method.
- The steps of computing the seasonal indices on moving average method are as follows.
  1. Re-writing the quarterly time series data in each year vertically down in a consecutive order.
  2. Having computed the moving averages of order four, deriving their centered moving averages of order two,
  3. expressing each original time series value ( $y$ ) as a percentage of the corresponding moving average value.
  4. Separating the percentages derived in the above step 3 quarterly, computing the means of the percentages for the four quarters,
- Advantages of using the moving average method for computing seasonal indices are as follows.
  - Ability of computing seasonal fluctuations very accurately using the moving average method.
  - Being the most frequently applied method of constructing seasonal indices.
- Few disadvantages of using the moving average method for computing the seasonal indices are as follows.
  - Not being that convenient and simple like the average percentage method.
  - Percentage moving average values from both ends of the time series being lost.
- When there are extreme values in time series data the moving medians can be used instead of moving averages.

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.9** : Deseasonalizes the data and makes decisions .

**No. of periods** : 08

**Learning outcomes :**

- Introduces ‘deseasonalization’ of data.
- Describes the need of deseasonalizing of time series data.
- Deseasonalises the original time series data using seasonal indices.
- Plots the deseasonalized data on the same time series graph.
- Makes decisions using deseasonalized data.
- Adjusts the influence of seasonal indices to deseasonalised data.

**Instructions for lesson Planning :**

- Present the following dialogue to the class to explain the deseasonalizing of data and involve the students in the activity mentioned below.

Taxi driver : Mudalali ..... today on foot...? Where to.... ?  
Come I’ll give you a lift.

Business man : Oh.. how nice .... Thank you.  
Let’s go then ...

Taxi driver : Where’s the car .... ? any break down ?

Business man : see.... What a pity ? it was shifted by the leasing company  
due to an arrears of few installments.

Taxi driver : What happened Mudalali ... is there any upset in business ?

Business man : Keeping hopes on usual rains in this season. I made a sufficiently lot  
of umbrellas, even using bank loans. I normally follow the same  
procedure in every year as well. Since it’s not raining as expected this  
time all hopes were in vain I’ve been thoroughly disappointed. Even  
the rent may not be affordable.

Taxi driver : We also find the business profitable during rainy season of every year.  
I also took a loan and repaired the vehicle and ..... waiting for .....

Business man : Then you may also be in the same boat as myself.

Taxi driver : It won’t be such a big issue for me as I’ve made an arrangement for a  
sudden drop of rains and if the business goes on lose.

Business man : That’s better. How careful you are .....? Had I also made such vigilant  
arrangement. I wouldn’t have fallen in to a big crisis like this .... These  
seasonal fluctuations are not definite, are they?

- Hold a discussion highlighting the following facts.
  - Since a variable is undergone to temporary fluctuations only on seasonal variations, forecasting the movement of a variable on seasonal variations is not that convenient.
  - Planning of business affairs keeping hopes on seasonal variations is risky.
  - Removal of seasonal influences contained in time series data is required in forecasting, That is known as deseasonalizing of time series data.
  - Time series data can be deseasonalized using seasonal indices.

**Activity 01 :** Provide with the following table to the students.

Sales of Ice-cream on seasonal changes

| Sale          | Heavy rainy season<br>December-March | Summer season<br>June-August |
|---------------|--------------------------------------|------------------------------|
| As usually    | Declines in 60%                      | Inclines in 50%              |
| Sales in 2016 | 800 kg                               | 1 800 kg                     |

- Compute the following, unless the expected changes in weather conditions were experienced in 2016.
  - How could be the Ice cream sales in December – March period?
  - How could be the Ice cream sales in June - August period?

**Solution – Activity 1 :**

- True sales in December – March had been 800 kg.  
 Since this had been the sales after decaling the normal sales by 60% the real sales in this period could have been, unless there were heavy rains is  $\frac{800}{40} \times 100 = 2000kg$
- True sales in June-August had been 1 800 kg.  
 Since this had been the increased sales by 50% beyond the normal sales the real sales that could have been experienced unless the summer had been too severe  $\frac{1800}{150} \times 100 = 1200kg$

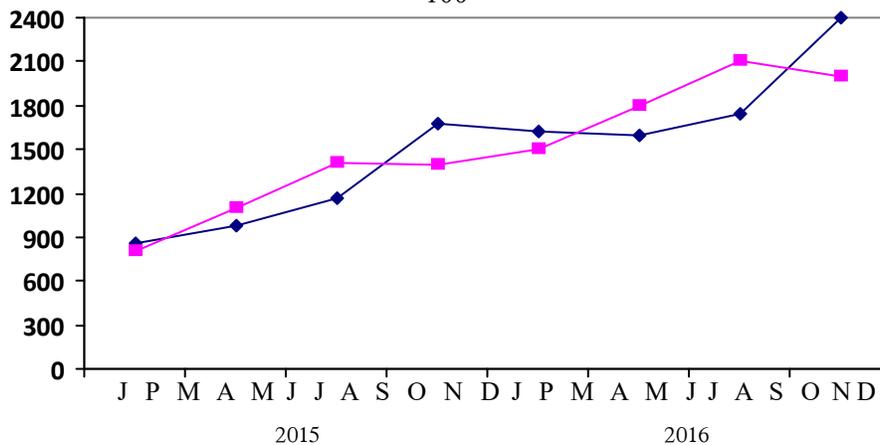


$$2016-1^{\text{st}} \text{ quarter} \quad 1620 \div \frac{108}{100} = 1500$$

$$2^{\text{nd}} \text{ quarter} \quad 1602 \div \frac{89}{100} = 1800$$

$$3^{\text{rd}} \text{ quarter} \quad 1743 \div \frac{83}{100} = 2100$$

$$4^{\text{th}} \text{ quarter} \quad 2400 \div \frac{120}{100} = 2000$$



- Assign the following activity to the students to insert the seasonal influences to deseasonalised data.

### Activity 3 :

- Deseasonalized data related to the number of tourists expected to visit in a particular tourist zone in Sri Lanka in 2016 are as follows.

| QuarterNo.              | No of Tourists |
|-------------------------|----------------|
| 1 <sup>st</sup> quarter | 22             |
| 2 <sup>nd</sup> quarter | 20             |
| 3 <sup>rd</sup> quarter | 16             |
| 4 <sup>th</sup> quarter | 18             |

- If the seasonal indices which were used to deseasonalise data respectively are 80, 110, 90 and 120 name the quarter during which the most number of tourists was reported and the quarter during which the least number of tourists was reported.

### **Solution : Activity 3 :**

Step I - In order to compute the real number of tourists reported in each quarter, multiplication of deseasonalized data by seasonal indices for inserting the seasonal influence in to deseasonalised data.

$$1^{\text{st}} \text{ quarter} \quad 22 \times \frac{80}{100} = 17.6$$

$$2^{\text{nd}} \text{ quarter} \quad 20 \times \frac{110}{100} = 22.0$$

$$3^{\text{rd}} \text{ quarter} \quad 16 \times \frac{90}{100} = 14.4$$

$$4^{\text{th}} \text{ quarter} \quad 18 \times \frac{120}{100} = 21.6$$

Step 2 :

- Identification of quarters with the highest and the lowest number of tourists.
- The quarter with the most number of tourists is the 2<sup>nd</sup> quarter. The quarter with the lowest number of tourists is the 3<sup>rd</sup> quarter.

### **A Guideline to explain the subject matters :**

- It is difficult to forecast using the data with seasonal variations.
- The businessmen should get prepared to face the practical situations that may be arisen due to the absence of expected seasonal influence.
- Therefore the data free from seasonal influence should be derived and the patterns in them also should be studied.
- By deflating the original quarterly or monthly time series data using the corresponding quarterly or monthly seasonal index the data free from seasonal influences could be derived.
- Once the data free from seasonal influences are available, the original data consist of seasonal influences can be regained by multiplying the deseasonalised data from corresponding seasonal indices.

**Competency 9.0** : Analyses the time dependent variables and forecasts

**Competency level 9.10** : Forecasts using the analysis of time series components

**No. of periods** : 06

**Learning outcomes :**

- Describes what forecasting is.
- Forecasts the time series variables using **Long Term Trend** and **Seasonal Indices**.
- Forecasts changing the origin of the Long term Trend line.
- Forecasts converting the annual data into monthly or quarterly data.

**Instructions for Lesson Planning :**

- Ask the students to write the individual marks received by each for Business Statistics at grade 12 - 1<sup>st</sup> term test, grade 12-2<sup>nd</sup> term test, grade 12 -3<sup>rd</sup> term and grade 13 -1<sup>st</sup> term test on a piece of paper in order.
- Then give instructions to them to write the possible marks assumed to be gained at grade 13 -2<sup>nd</sup> term test and grade 13 - 3<sup>rd</sup> term test.
- Point out the fact that logically assuming the possible value of a particular variable to be achieved in future, examining the past movement of that variable in this manner is known as forecasting.
- Present the trend equation of the annual sales of a company  $y = 284 + 1.44 X$  (origin 2014) **Y** - sales income in millions of rupees and I unit of **X** is one year. Lead the students to forecast the trend value in the year 2022.
- Assure that, that value is Rs. 295 520 000/-
- Ascertain that this value has been forecast only using the secular trend.
- Build up an opinion that it would be better to consider the influence of the seasonal component as well as forecasting the value of a time series variable.
- Point out that the cyclical and irregular movements are not frequently occurred, so that considering about those components is not necessary.
- Point out that the total value of the time series variable (**y**) can be forecast using the equation  $Y = (\beta_0 + \beta_1 x) \times \frac{S}{100}$  and involve the students in the following activity.
- $Y = 71 + 0.36 X$  is the trend line constructed using quarterly data. A unit of **X** is one quarter where as a unit of **Y** is one million rupees. The seasonal indices related to the relevant variable have been computed as follows.

| Quarter | Seasonal index |
|---------|----------------|
| I       | 102            |
| II      | 120            |
| III     | 90             |
| IV      | 88             |

- Forecast the value of the variable in the 3<sup>rd</sup> quarter of the year 2020.
- Provide with instructions to list out the quarters in sequential order beginning from the first quarter in 2016 to the third quarter in 2020 and to find the relevant sequential number for 3<sup>rd</sup> quarter in 2020; Hence guide them to forecast the time series value of the 3<sup>rd</sup> quarter in 2020.

Solution :

If the 1<sup>st</sup> quarter in 2016 is 0

The 3<sup>rd</sup> quarter in 2016 is 2

The 3<sup>rd</sup> quarter in 2020 is  $2 + 4 \times 4 = 18$

Accordingly the value in the 3<sup>rd</sup> quarter 2020 is as follows.

$$y = (71 + 0.36x) \times \frac{90}{100}$$

$$y = (71 + 0.36 \times 18) \times 0.9$$

$$y = (71 + 6.48) \times 0.9$$

$$y = 77.48 \times 0.9$$

$$y = 69.732$$

$$\underline{\underline{y = \text{Rs. } 69\,732\,000}}$$

- Point out that the forecasting also can be performed by changing the origin of the long term trend line.
- Explain how to shift the origin of the trend equation  $Y = 284 + 1.44X$  (origin – 2014) up to 2018.

$$y = 284 + 1.44(x + 4)$$

$$y = 284 + 1.44x + 5.76$$

$$y = \underline{\underline{289.76 + 1.44x}} \quad (\text{origin} - 2018)$$

- Point out that once the annual trend line equation is given, hence the monthly trend line equation can be derived.
- Guide the students to shift the origin of the line  $Y = 284 + 1.44x$  (orgine – 2014) to January 2015 and to derive the trend value in September - 2015.
- Point out that the intercept and the gradient of the equation  $y = 284 + 1.44x$  should be divided by 12 and 144 respectively in order to derive the monthly trend line as on 01.07.2014. Then

$$y = \frac{284}{12} + \frac{1.44X}{144}$$

$$Y = 23.67 + 0.01X - 01.07.2014$$

- Point out that when the origin is shifted to January 2015 it would be better to use 15.01.2015 as the base date and guide them to shift the origin up to that date.

$$y = 23.67 + 0.01(x + 6.5)$$

$$y = 23.67 + 0.065 + 0.01x$$

$$y = 23.735 + 0.01X - 15.01.2015$$

- Accordingly the trend value as at 15.09. 2015 should be derived as follows.

$$y = 23.735 + 0.01 \times 8$$

$$y = 23.735 + 0.08$$

$$y = 23.815$$

$$= \underline{\underline{\text{Rs. 23 815 000}}}$$

- Lead the students to check whether the same result could be derived using the first trend equation as well.

$$\text{Then, } y = 23.67 + 0.01x - 01.07.2014$$

Since there are  $14 \frac{1}{2}$  months up to 15.09.2015 from 01.07.2014.

$$y = 23.67 + 0.01x \times 14.5$$

$$y = 23.67 + 0.145$$

$$y = 23.815$$

$$Y = \underline{\underline{\text{Rs. 23 815 000}}}$$

- In the same manner lead the students to derive the quarterly trend line from the annual trend line.

$$y = 284 + 1.44x - 01.07.2014$$

$$y = \frac{284}{4} + \frac{1.44x}{16}$$

$$y = 71 + 0.09x - 01.07.2014$$

- Guide the students to shift the origin of this trend line to the last quarter of 2014 (There are 1 ½ quarters form 01.07.2014 to 15.11.2015).

$$y = 71 + 0.09(x + 1.5)$$

$$y = 71 + 0.135 + 0.09x$$

$$y = 71.135 + 0.09x - (15.11.2014)$$

- Accordingly lead the students to forecast the trend value of the first quarter in 2017). (There are 9 quarters from the ivth quarter, 2014 to 1<sup>st</sup> quarter 2017). Accordingly the trend value of the first quarter in 2017 is as follows.

$$y = 71.135 + 0.09 \times 9$$

$$y = 71.135 + 0.81$$

$$y = 71.945$$

$$\underline{\underline{y = \text{Rs } 71\,945\,000}}$$

- Guide the students to check whether the same result can be derived directly from the origin 01.07.2014.

The number of quarters from 01.07.2014 to 15.02.2017 is 10 ½

$$y = 71 + 0.09 \times 10.5$$

$$y = 71 + 0.945$$

$$y = 71.945$$

$$\underline{\underline{y = \text{Rs. } 71\,945\,000}}$$

**A guideline to explain the subject matters :**

- The final objective of a time series analysis is forecasting.
- Estimating the total value of a considered time series variables as at a given prospective time unit using a constructed time series model is known as forecasting.
- Using a trend line constructed on the model  $y = \beta_0 + \beta_1 x$  the value of y for a particular prospective period can be forecast.
- Forecasting the prospective value of a time series variable can be performed normally using both the secular trend and seasonal component as follows.

$$y = (\beta_0 + \beta_1 x) \frac{S}{100}$$

- The origin or the base period of a constructed trend line can be shifted as required.
- Once the origin is shifted in that manner a new trend equation with a different intercept and the same gradient can be derived.
- Once the annual trend line is given the monthly trend line can be derived as follows.

$$y = \frac{\beta_0}{12} + \frac{\beta_1 x}{144}$$

- Once the annual trend line is given the quarterly trend line can be derived as follows.

$$y = \frac{\beta_0}{4} + \frac{\beta_1 x}{16}$$

- The base date of an annual trend line is completed the mid day of the year, the base date of a monthly trend line is the mid day of the month and the base date of a quarterly trend line is the mid day of the quarter.

**Competency 10.0 :** Uses statistical quality control techniques to make managerial decisions.

**Competency level 10.1:** Studies the quality of a product and the variations caused for the quality

**No. of periods :** 08

**Learning outcomes :**

- Explains what the quality of a product is.
- Explains the random (chance) variations of a product with examples.
- Explains the assignable causes affected for changing the quality of a product with examples.
- Ascertains the need of applying quality control techniques to assure the quality of a product.
- Writes the uses of Statistical Quality Control.
- Introduces statistical techniques applicable for controlling the quality of a good or a service.

**Instructions for Lesson Planning :**

- In order to explain the ‘ quality’ pay the attention of students towards following situations.
  - Related to the uniform of school children the circumstance of the trouser of a pair of trousers being pre determined as 16” (inches).
  - The length of the frock of a girl’s uniform being pre determined up to the midpoint of the knee bone.
  - ISO certificate symbol being displayed in a product.
  - The mass of a packet of curry powder without SLS certificate being mentioned as 500g.
- Hold a discussion with the students regarding the above situations.
  - Explain that the circumference of the trouser of a pair of trousers being 16”, the length of the frock being up to the midpoint of the knee bone are pre-determined standards.
  - Explain that if it is an SLS certificate granted product, it has been ensured to have been made in accordance with the standards determined by Sri Lanka Standards Institution.
- Point out that if it is a product with ISO certificate, it has been produced to meet the requirements to receive the certificate of International Standards.

- Point out the students that the pre-determined standard of the packet of curry powder without SLS certificate mentioned its weight as 500g, is 500g. Point out further that in this situation, even though the SLS certificate hasn't been granted, the consumer gets the chance of checking up the weight of it, since the pre-determined weight is mentioned.
- Point out the students that availability of goods and services that meet the pre-determined standards and specimens is known as the quality of a product.
- Involve the students in following activity to explain that there are two types of factors causes for changing the quality of a product.
- Produce the following reasons that caused for a product not being able to meet the pre-determined standards.

| Causes for not being able to meet to pre-determined standards.   | A | B | C |
|--|---|---|---|
| <ul style="list-style-type: none"> <li>• Machines getting simply heated</li> <li>• Machines not being adjusted (fixed ) properly.</li> <li>• Changing the room temperature</li> <li>• Application of defective raw materials</li> <li>• Machines being depreciated</li> <li>• Changing the humidity</li> </ul> |   |   |   |

- Provide the following intrusions to the students.
  - If it is a random variation put a  $\surd$  in column A and if not a 'x'
  - If the cause of variation can be identified and corrected put a  $\surd$  in column B and if not a 'x'
  - If the error is possible to be corrected naturally without making a correction. Put a  $\surd$  in column 'C' and if not 'x'.
  - Lead the students to separate the causes into two groups taking the causes marked  $\surd$  in column A 'x' in column B and  $\surd$  in column 'C' to one side and the remaining to the other side.

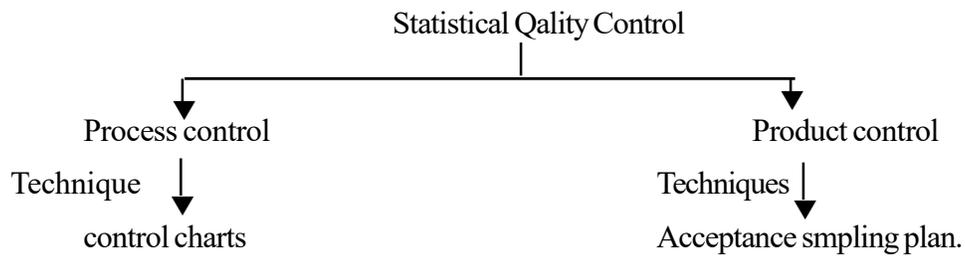
|   |   |
|---|---|
| <ul style="list-style-type: none"> <li>• Machines getting simply heated.</li> <li>• Changing the room temperature.</li> <li>• Changing the humidity.</li> </ul> | <ul style="list-style-type: none"> <li>• Machines not being adjusted (fixed) properly.</li> <li>• Application of defective raw materials</li> <li>• Machines being depreciated</li> </ul> |
|---|---|

- Explain to the students with reference to the above classification that the variances occurred in random identified causes of the variations not being able to correct, and the variations can be expected to be naturally corrected are known as random variations and the other variations as assignable variations.

- Explain to the students that examining whether the raw materials used as input to the process of production and the finished goods exit from the process as the output meet the pre-determined standards is very important, that it is called **“the product control”** and further that the techniques applied in that purpose is known as **“Acceptance Sampling Plan”**.
- Explain that to check whether a product is being made in the process of production to meet the pre-determined standards is known as **process control** and also that the techniques used in this purpose is known as **“the control charts”**.
- Point out the students that the statistical quality control is a joint purpose of the two divisions as process control and product control.
- Discuss with the students and note down the uses of statistical quality control.

**A guideline to explain the subject matters :**

- The pre-determined standard expected in a good or a service is the quality of it.
- There are two types of factors caused for the product being out of the standards.
  1. Random variations (chance variations/allowed variations)
  2. Assignable variations (controllable / preventable /non random variations).
- The variations occurred in random, that the identified errors can not be corrected and those errors can be expected to be corrected naturally are known as random variations.
- The variations that the identified errors can be corrected and until corrected the errors are in the same stage are known as assignable variations.



Uses of statistical quality control,

- Since the errors can be early recognized, wastage of raw materials, labour and time being minimized.
- Ability to enhance the productivity
- Ability to minimize the products being rejected in the market.
- Ability to assure a greater quality level with a lower cost of checking under a panel of trained quality controllers.

**Competency 10.0 :** Uses statistical quality control techniques to make managerial decisions

**Competency level 10.2 :** Uses appropriate methods for variable control

**No. of periods :** 10

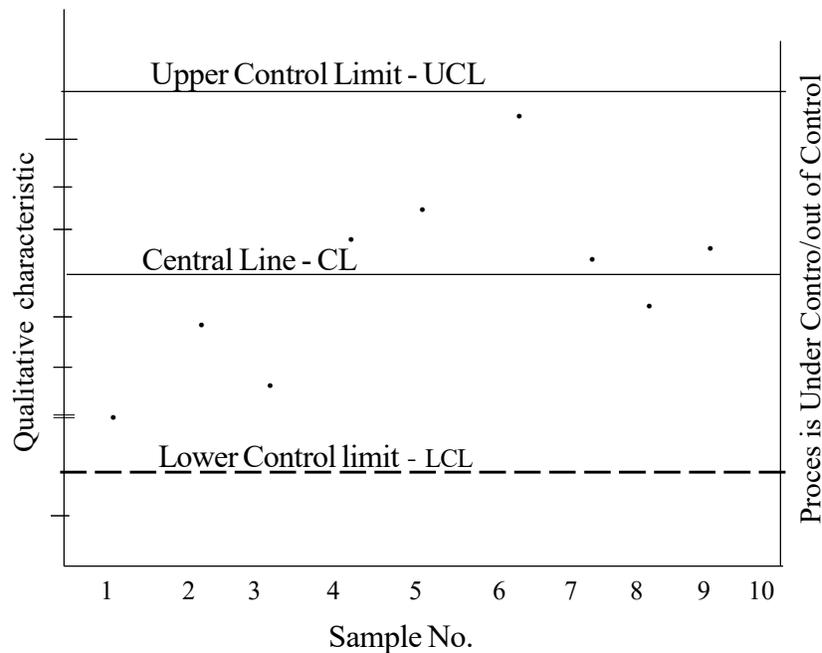
**Learning outcomes :**

- Explains what the process control is.
- Introduces ‘variable control,’
- Explains what is meant by control charts used for variable control.
- Constructs the mean control chart ( $\bar{X}$  - chart) for variable control.
- Constructs the control limits for the mean control chart when the standards are pre-determined
- Constructs the mean control chart when the standards are pre-determined
- Computes the control limits for construction of mean control charts using the relevant formulae once the standards are not pre-determined.
- Constructs  $\bar{X}$  charts once the standards are not pre-determined.
- Introduces the range chart ( R-chart) for process control.
- Computes the control limits for range chart once the standards are pre-determined.
- Constructs the range control chart when the standard are pre-determined.
- Computes the control limits for the range control chart when the standards are not pre-determined.
- Constructs the range control chart, when the standards are not pre-determined.
- Comments on the process of production using the mean control chart and range control chart.
- Explains the uses of control charts.

**Instructions for Lesson Planning :**

- Inquire various stages come across in the process of preparing a meal (rice and curry) at home and note down on the black/white board.
- Discuss with the students and list out the steps to be followed in the entire process of preparing the meal starting from selecting the materials for the meal and until the cooked meals are served on the table, in an orderly manner.

- Ascertain that in order to prepare a delicious and nutritious meal, every step mentioned above should be completed giving priority to cleanliness and being highly vigilant.
- Inquire about any other such production processes and put on the board.
- Inquire how to evaluate the quality of the relevant product (goods / services) considered at each process.
- Point out that the qualitative factor such as net weight, length, width and life time etc... of particular product can be measured and compared with a pre-determined standard and expected specimen.
- Point out that these factors are variables evaluating the quality of a product.
- Point out that the quality control technique that is used to check whether these variations related to the quality of the product are agreed with the relevant standards is known as the control charts.
- Depict the following rough sketch of a control chart before the class.



- Point out that the serial number of each sample tested is represented on horizontal axis whereas the qualitative characteristic like, the weight, length, life time etc. is on vertical axis of the above chart.
- Point out that the expected average quality level is indicated by the Central Line.
- Point out that the Upper Control Limit and the Lower control Limit are established on either side of the central line of a distance of 3 standard deviations.

- Point out that if all the points of sample statistic computed using the measures of the considered characteristic plotted against corresponding sample numbers fall between the control limits, the relevant process is determined to under control.
- Point out that if any value of such sample statistic falls beyond the upper control limit or below the lower control limit that process is determined to be out of control.
- Present the following data to students and involve them in the activity.
- In order to test the net weight of a packet of curry powder packed in machinery 10 samples with 5 packets in each were drawn in random at different occasions in this production process and the observed data are given below.

| Sample No | Packet Number |     |     |     |     |
|-----------|---------------|-----|-----|-----|-----|
|           | 1             | 2   | 3   | 4   | 5   |
| 1         | 240           | 244 | 250 | 249 | 248 |
| 2         | 247           | 246 | 247 | 248 | 251 |
| 3         | 250           | 246 | 245 | 247 | 246 |
| 4         | 251           | 248 | 249 | 250 | 249 |
| 5         | 249           | 249 | 248 | 249 | 248 |
| 6         | 246           | 242 | 247 | 248 | 248 |
| 7         | 244           | 241 | 246 | 249 | 251 |
| 8         | 243           | 244 | 248 | 246 | 245 |
| 9         | 248           | 247 | 247 | 251 | 249 |
| 10        | 251           | 250 | 247 | 249 | 249 |

- Lead the students on following steps for constructing a mean control chart using these observed data.
  1. Compute the mean of each sample
  2. Derive the range of each sample  
Sample Range = Highest observation – Lowest observation
  3. Derive the mean of all the sample means – (the Grand Mean  $\bar{\bar{x}}$ )
  4. Point out that considering this Grand Mean ( $\bar{\bar{x}}$ ) is suitable as the Central Line of the control chart.
  5. Lead to compute the upper control limit as  $\bar{\bar{x}} + A_2\bar{R}$

6. Lead the students to derive the Lower Control Limit using  $\bar{\bar{x}} - A_2 \bar{R}$

N : B : Point out that the value of  $A_2$  factor can be derived reading the relevant line in quality control table.

**Solution :**

Step 1 : Deriving the mean of each sample

| Sample No | Total $\sum_{i=1}^5 X_i$ | $\bar{x}$ | Sample Range - R    |
|-----------|--------------------------|-----------|---------------------|
| 1         | 1231                     | 246.2     | 10                  |
| 2         | 1239                     | 247.8     | 5                   |
| 3         | 1234                     | 246.8     | 5                   |
| 4         | 1247                     | 249.4     | 3                   |
| 5         | 1243                     | 248.6     | 1                   |
| 6         | 1231                     | 246.2     | 6                   |
| 7         | 1231                     | 246.2     | 10                  |
| 8         | 1226                     | 245.2     | 5                   |
| 9         | 1242                     | 248.4     | 4                   |
| 10        | 1246                     | 249.2     | 4                   |
|           | $\sum \bar{x}$           | 2474.0    | $\sum \bar{x} = 53$ |

- Central Line of the control chart  $CL_{\bar{x}} = \bar{\bar{x}} = \frac{\sum \bar{x}}{k}$

- Let's suppose that the number of samples as 'K'

$$\therefore \bar{\bar{x}} = \frac{2474}{10} = \underline{\underline{247.4}}$$

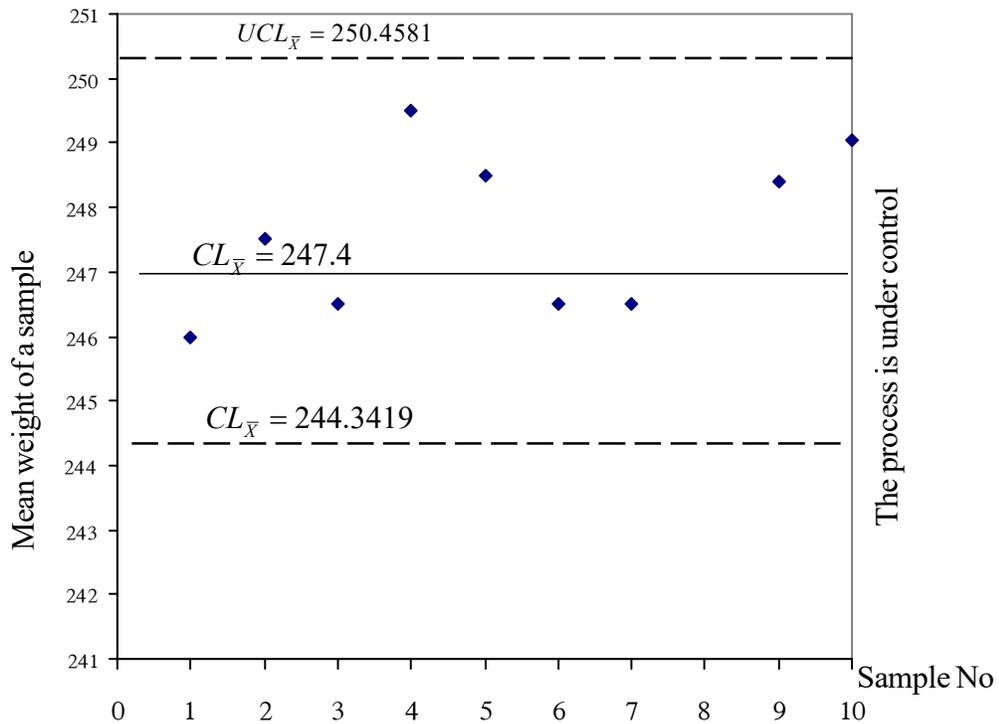
- Upper Control Limit

$$\begin{aligned}
 UCL_{\bar{x}} &= \bar{\bar{x}} + A_2 \bar{R} \\
 &= 247.4 + 0.577 \times \frac{53}{10} \\
 &= 247.4 + 0.577 \times 5.3 \\
 &= 247.4 + 3.0581 \\
 &= \underline{\underline{250.4581}}
 \end{aligned}$$

- Lower Control Limit

$$\begin{aligned}
 LCL_{\bar{x}} &= \bar{\bar{x}} - A_2 \bar{R} \\
 &= 247.4 - 0.577 \times 5.3 \\
 &= 247.4 - 3.0581 \\
 &= \underline{\underline{244.3419}}
 \end{aligned}$$

Mean control Chart



- Hold a discussion with the students to get an approach to construct the control chart when the standards are pre-determined.
- Collect few exercise books in various size from few students and write the length and the breadth of each book mentioned on the cover on the white/black board.
- Highlight that those figures noted down on the board are pre-determined standards expected in a book released out from the process of production.
- Inquire from student about any other products mentioned the standards in this manner.

- Accordingly point out that the net weight of a packet of milk powder and the percentage of ingredients such as calcium, carbohydrate, vitamins, minerals and fats etc. are pre-determined in that manner.
- Hence point out that the mean ( $\mu'$ ) and standard deviation  $\sigma'$  can be pre-determined for a particular variable as standards.
- Inquire from the students how should be the Control Limits for constructing the control chart.
- Highlight through discussion that it would be reasonable to consider the central line as  $\mu'$  which is the pre-determined mean.

- Upper Control Limit as  $UCL_{\bar{x}} = \mu' + \frac{3\sigma'}{\sqrt{n}}$

- Lower Control Limit as  $LCL_{\bar{x}} = \mu' - \frac{3\sigma'}{\sqrt{n}}$

(The value of  $\frac{3}{\sqrt{n}}$  is given in the quality control table of A)

Accordingly point out that Upper Control Limit as  $UCL_{\bar{x}} = \mu' + A\sigma'$

Lower Control Limit as  $LCL_{\bar{x}} = \mu' - A\sigma'$

- Involve the students in the following exercise.

A particular rice mill owner produces packets of rice in 25kg each to the market. Suppose that the standards have been pre-determined as the net weight of a bag of rice is mean  $\mu' = 24.9kg$  and the standard deviation  $\sigma' = 1.5kg$ . Ten samples with 12 bags of rice in each were drawn out from this process and checked. The mean weight of a bag of rice in each sample was as follows.

| Sample No. | Sample mean weight<br>$\bar{x}$ kg |
|------------|------------------------------------|
| 1          | 24.85                              |
| 2          | 24.92                              |
| 3          | 24.76                              |
| 4          | 25.01                              |
| 5          | 24.96                              |
| 6          | 23.82                              |
| 7          | 23.24                              |
| 8          | 24.65                              |
| 9          | 25.73                              |
| 10         | 24.52                              |

Compute the control limits for mean control chart and point out whether the process is under control.

**Solution :** Central Line  $CL_{\bar{x}} = \mu'$

$$= 24.9kg$$

Upper Control Limit  $ICL_{\bar{x}} = \mu' + A\sigma'$

$$= 24.9 + 0.866 \times 1.5$$

$$= 24.9 + 1.299$$

$$= \underline{\underline{26.199}}$$

Lower Control Limit  $LCL_{\bar{x}} = \mu' - A\sigma'$

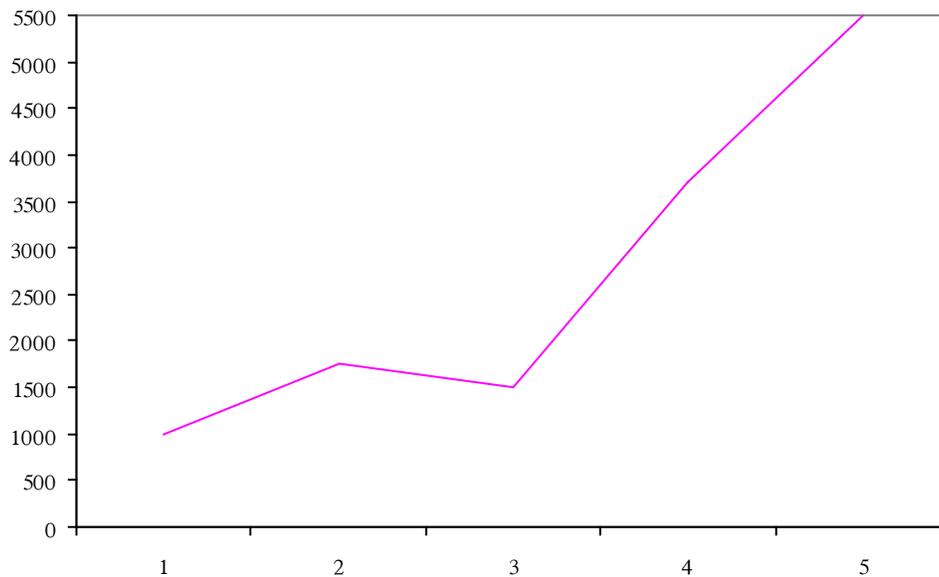
$$= 24.9 - 0.866 \times 1.5$$

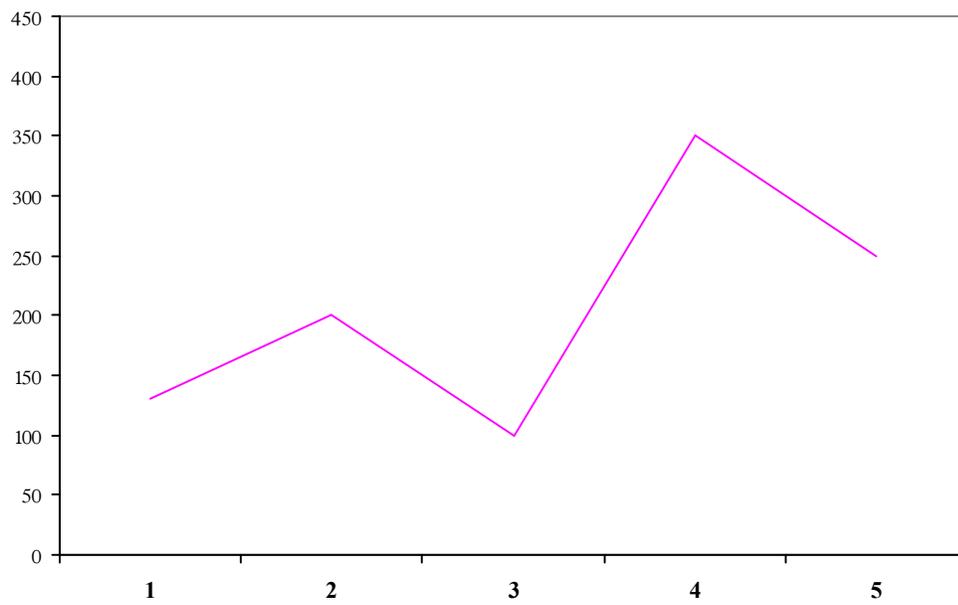
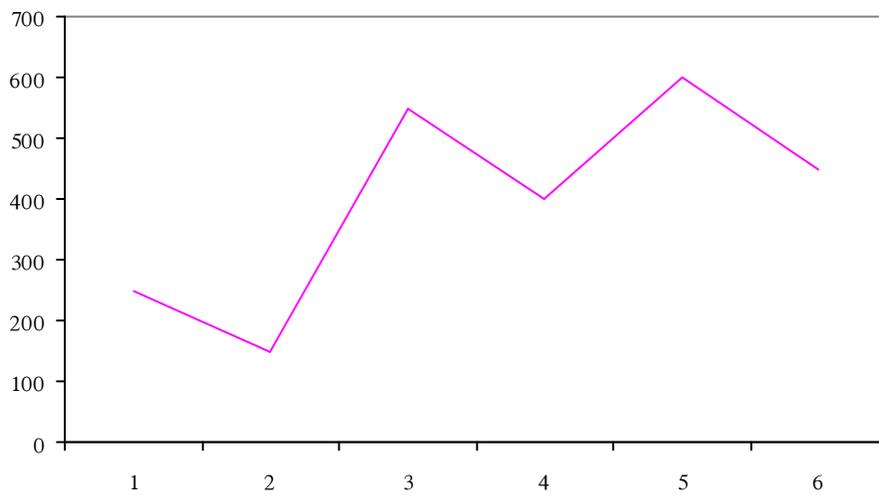
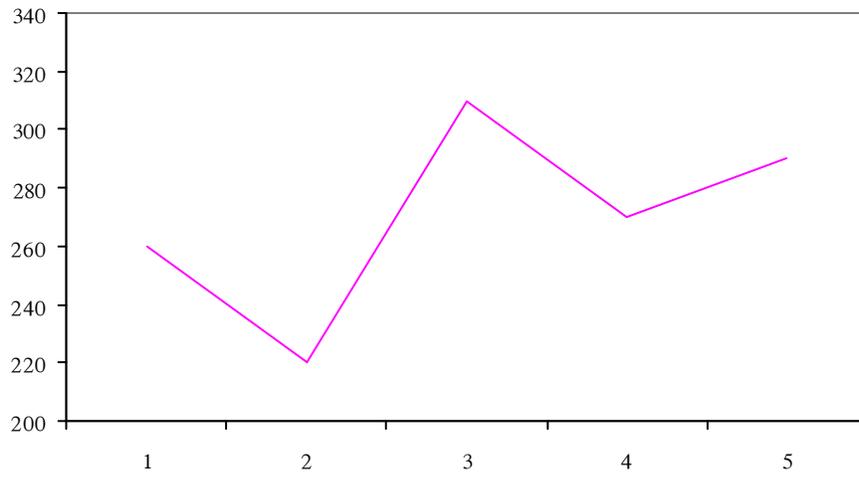
$$= 24.9 - 1.299$$

$$= \underline{\underline{23.601}}$$

Hence point out that this process is out of control, since the mean of sample No. 7  $\rightarrow 23.24$  falls below the Lower Control Limit.

- Display the following graphs before the class and get the highest value and the lower value of each marked on it.





- Mark the difference between those two values using a straight line.
- Highlight through the discussion that constructing control charts using the range of the highest and the lowest values of each sample is also appropriate.
- Recall the ranges and the average range  $\bar{R}$  used in constructing the mean control chart in previous lesson.
- Point out that a control chart can be constructed using those ranges as well.
- Highlight that using the central line of the range chart as  $\bar{R}$ , upper control limit as  $\bar{R} + 3\sigma_R$  and lower control limit as  $\bar{R} - 3\sigma_R$  would be reasonable.
- The central limits of a range chart can be computed using the factors  $D_3$  and  $D_4$  given in the quality control table for convenience of calculating as follows.

$$\text{Upper Control Limit} \quad UCL_R = D_4 \bar{R}$$

$$\text{Lower Control Limit} \quad LCL_R = D_3 \bar{R}$$

$$\text{Central Control Limit} \quad CL_R = \bar{R}$$

- Present the example used in constructing the mean control chart related to the process of packeting curry powder and lead the students to construct a range control chart for that instance.

**Solution :**

| Sample No | Sample Range |
|-----------|--------------|
| 1         | 10           |
| 2         | 5            |
| 3         | 5            |
| 4         | 3            |
| 5         | 1            |
| 6         | 6            |
| 7         | 10           |
| 8         | 5            |
| 9         | 4            |
| 10        | 4            |

- Compute the mean of the ranges

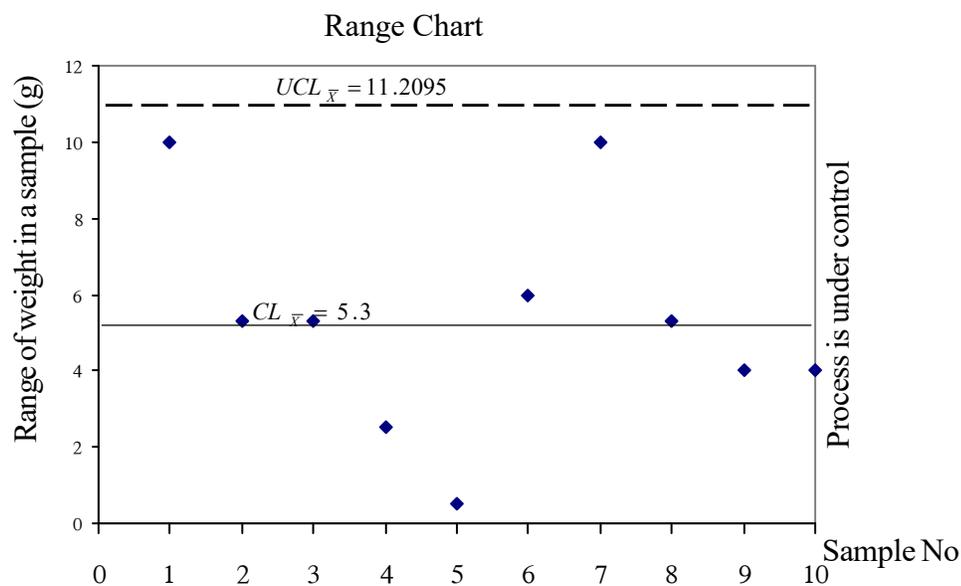
$$\begin{aligned} \bar{R} &= \frac{\sum R}{K} \\ &= \frac{53}{10} \\ &= \underline{\underline{5.3}} \end{aligned}$$

- Upper Control Limit

$$\begin{aligned}
 UCL_R &= D_4 + \bar{R} \\
 &= 2.115 \times 5.3 \\
 &= \underline{\underline{11.2095}}
 \end{aligned}$$

- Lower Control Limit

$$\begin{aligned}
 LCL_R &= D_3 - \bar{R} \\
 &= 0 \times 5.3 \\
 &= \underline{\underline{0}}
 \end{aligned}$$



- Point out that the Control Limits for the range chart can be computed using the following formulae, once the population standard deviation related to the sample ranges of a variable in connection with a particular manufacturing process has been pre-determined as  $\sigma'$

Upper Control Limit  $UCL_R = D_2 \sigma'$

Central Line  $CL_R = d_2 \sigma'$

Lower Control Line  $LCL_R = D_1 \sigma'$

- Point out that the value of the factor  $D_2$ ,  $d_2$  and  $D_1$  can be derived from the quality central table.
  - Assign the following exercise to the students.
  - Suppose that the standard deviation of the distribution of net weight of a packet of biscuits processed in a biscuit manufacturing company has been pre-determined as 2.8 g

- The ranges of the observations in 10 samples of size 10 in each tested are as follows.

| Sample No. | Range |
|------------|-------|
| 1          | 4     |
| 2          | 8     |
| 3          | 3     |
| 4          | 2     |
| 5          | 1     |
| 6          | 10    |
| 7          | 12    |
| 8          | 11    |
| 9          | 8     |
| 10         | 5     |

Compute the Control Limits and point out whether the process is under control.

$$\begin{aligned}
 \text{Control Limits} \quad UCL_R &= D_2 \sigma' \\
 &= 5.469 \times 2.8 \\
 &= \underline{\underline{15.3132}}
 \end{aligned}$$

$$\begin{aligned}
 \text{Central Line} \quad CL_R &= d_2 \sigma' \\
 &= 3.078 \times 2.8 \\
 &= \underline{\underline{8.6184}}
 \end{aligned}$$

$$\begin{aligned}
 \text{Lower Control Limit} \quad LCL_R &= D_1 \sigma' \\
 &= 0.687 \times 2.8 \\
 &= \underline{\underline{1.9236}}
 \end{aligned}$$

**Conclusion :**

- Guide the students to compare the given ranges with the control limits computed. (without constructing the control charts)
- Ascertain the fact that the range value of sample No.5 falls below the Lower Control Limit.
- Point out that it is not reasonable to determine that the process is out of control merely because a single range (observation) falls down the Lower Control Limit.

- Further point out that the range control chart is also as important as the mean control chart.
- If the process is under control in accordance with both these control charts, point out that it would be more favourable.

**A guideline to explain the subject matters :**

- Testing whether the production is carried out in accordance with the expected standard and specimens in connection with a particular production process in a goods and services manufacturing firm, using random samples drawn at various situations from the process under continuous supervision over the function, is known as the process control in statistical quality control.
- The quality of a product is evaluated on any factor or factors such as net weight, length, breadth, height, size and life time of it and such factors are called variables.
- The statistical techniques applied to check whether these variables are functioning in accordance with the expected standards and specimens is known as **Control Charts Analysis**.
- The rectangular shaped chart which is prepared on a co-ordinate plane on which the sample number of the random samples drawn from the production process are represented on horizontal axis where as the qualitative characteristic of the variable considered to be tested the quality of the product on vertical axis; consists of the Central Line for expected quality level and two control limits located at the distance of 3 standard deviations on either side (Lower Control Limit and Upper Control Limit) is known as the control chart.
- There are two types of control charts in process control as control charts for variables and control charts for attributes.
- The Control Charts for variables are again in two types as mean control charts (  $\bar{X}$  - charts) and range control charts (R - charts).
- Control limits for the mean control chart are determined based on the parameters computed using sample data, once the standards for mean control chart haven't been pre-determined.
- Once the means of all the samples  $\bar{X}$  have been computed for the same size samples drawn in random from the production process by now and then, the mean of all those samples (Grand mean  $\bar{\bar{X}}$  ) is computed.
- That Grand mean (  $\bar{\bar{X}}$  ) is considered as the central line of the mean control chart.

- The Upper Control Limit is placed beyond three standard deviations from the central line.

$$UCL_{\bar{X}} = \bar{\bar{X}} + 3\sigma_{\bar{X}}$$

- The Lower Control Limit is placed below the three standard deviations from the central line.

$$LCL_{\bar{X}} = \bar{\bar{X}} - 3\sigma_{\bar{X}}$$

- The value of  $3\sigma_{\bar{X}}$  (Thrice the sample standard deviation) can be derived using  $A_2\bar{R}$  of which  $A_2$  factor is given in the quality control table.  $\bar{R}$  is the mean of the sample ranges (The difference between the highest value and the lowest value of the each sample is the range).

- Accordingly the control charts can be stated as follows :

$$UCL_{\bar{X}} = \bar{\bar{X}} + A_2\bar{R}$$

$$CL_{\bar{X}} = \bar{\bar{x}}$$

$$LCL_{\bar{X}} = \bar{\bar{X}} - A_2\bar{R}$$

- Once the standards for mean of a certain variable as  $\mu'$  and standard deviation as  $\sigma'$  have been pre determined, control limits can be computed for constructing the mean control chart as follows.

Central Line  $CL_{\bar{X}} = \mu'$

Upper Control Limit  $UCL_{\bar{X}} = \mu' + \frac{3\sigma'}{\sqrt{n}}$

Lower Control Limit  $LCL_{\bar{X}} = \mu' - \frac{3\sigma'}{\sqrt{n}}$

- The values of  $\frac{3}{\sqrt{n}}$  have been presented in the quality control table named as A.

Hence the above mentioned central limits can be replaced as follows.

Central Line  $CL_R = \mu'$

Upper Control Limit  $UCL_R = \mu' + A\sigma'$

Lower Control Limit  $LCL_R = \mu' - A\sigma'$

- Another control chart called Range Chart also can be constructed in addition to the mean control chart considering the range of observations in each sample.
- Once the standards of the product haven't been pre-determined control limits for the range chart can be derived as follows.

|                     |  |
|---------------------|--|
| Upper Control Limit | $UCL_R = \bar{R} + \frac{3\sigma_R}{\sqrt{n}}$ |
|---------------------|--|

|              |                  |
|--------------|------------------|
| Central Line | $CL_R = \bar{R}$ |
|--------------|------------------|

|                     |  |
|---------------------|--|
| Lower Control Limit | $LCL_R = \bar{R} - \frac{3\sigma_R}{\sqrt{n}}$ |
|---------------------|--|

- These control limits can be computed conveniently as follows guided by the quality control table.

|                     |                       |
|---------------------|-----------------------|
| Upper Control Limit | $UCL_R = D_4 \bar{R}$ |
|---------------------|-----------------------|

|              |                  |
|--------------|------------------|
| Central Line | $CL_R = \bar{R}$ |
|--------------|------------------|

|                     |                       |
|---------------------|-----------------------|
| Lower Control Limit | $LCL_R = D_3 \bar{R}$ |
|---------------------|-----------------------|

- Once the standards for the standard deviation of sample ranges have been predetermined as  $\sigma'$  the control limits for range chart can be derived as follows.

|                     |                       |
|---------------------|-----------------------|
| Upper Control Limit | $UCL_R = D_2 \sigma'$ |
|---------------------|-----------------------|

|              |                      |
|--------------|----------------------|
| Central Line | $CL_R = d_2 \sigma'$ |
|--------------|----------------------|

|                     |                       |
|---------------------|-----------------------|
| Lower Control Limit | $LCL_R = D_1 \sigma'$ |
|---------------------|-----------------------|

- Both the mean control chart and the range chart are vital in process control for ruling out the variables.
- The mean control chart evaluates how each sample mean is deviated from the grand mean (population mean).
- The range chart evaluates how the variations in observations of the relevant variable are deviated from the average range.
- If the process is under control in accordance with both these control charts that is indeed a vial achievement.
- If a particular production process which is under control in accordance with the mean control chart, seems to be out of control on the range chart, it won't be a satisfactory achievement, since the variations in observations are highly fluctuated.

**Competency 10.0** : Uses statistical quality control techniques to make managerial decisions

**Competency level 10.3** : Uses appropriate techniques for controlling attributes

**No. of periods** : 12

**Learning outcomes :**

- Introduces ‘attributes’
- Provides with instances for attributes from manufacturing field.
- Names the control charts used for controlling attributes.
- Defines ‘**P**’ **chart**
- States the situations where ‘**P**’ **chart** is applicable.
- Constructs a **p-chart** for given data and interpret.
- Defines ‘**np**’ **chart**
- States the situations where ‘**np**’ **chart** is applicable.
- Constructs a **np-chart** for given data and interpret.
- Defines the **c-chart**
- States the situations where **c-chart** is applicable.
- Constructs a **c – chart** for given data and interprets.
- Defines **u - chart**.
- States the situations where **u chart** is applicable.
- Constructs the **u – chart** for given data and interpret.

**Instructions for Lesson Planning :**

- In order to explain attributes control divide the class into 3 groups and distribute packets of 3 different kinds of vegetable seeds to each group. Then provide with following instructions.
  - Observe what is mentioned in the label of the packet of seeds you have got very carefully.
  - Understand the germination percentage of the kind of seeds you have received.
  - One of the members in the group read that germination percentage aloud to be heard by everybody.
- Hold a discussion highlighting the following facts.
  - The quality of vegetable seeds has been expressed based on the germination percentage of them.
  - There is not a measuring unit to evaluate the germination of seeds, so that their quality is determined on the proportion of germination.

- Despite the quality levels that could be evaluated using the measuring units such as height, weight, length and life time etc. there are some other quality levels in production that can not be evaluated using a measuring unit.
- The quality of such products can be evaluated using sensitive factors such as colour, odour sound, taste and smoothness/roughness and those factors are known as attributes.
- In order to test the quality of such products proportion of defectives or proportion of non defectives can be used.
- By ruling out the sample defective proportion or sample defective number in each sample drawn from such a production process, their quality level can be protected.
- There are four types of control charts that can be used with this regard.
  - Defective Proportionate control chart (**p-chart**)
  - Defective Number control chart (**np-chart**)
  - Unit defective number control chart (**c-chart**)
  - Sub section defective number chart (**u – chart**)
- Provide with the following activity to the students to construct a proportionate chart and to evaluate the quality of the process

### Activity

- Electric bulbs are manufactured in a production process. The quality controller draws out and tests a random sample of 100 bulbs daily. The details of 10 days are as follows.

|             |   |   |   |   |   |   |   |   |   |    |
|-------------|---|---|---|---|---|---|---|---|---|----|
| Day         | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| No of bulbs | 3 | 4 | 1 | 3 | 1 | 2 | 4 | 3 | 0 | 4  |

- Compute the control limits to construct a proportionate chart.
- Construct the proportionate control chart and plot the defective proportions of the samples
- Enquire the situation of the quality control of the process

### Solution : Activity

1<sup>st</sup> step : Computing of  $\bar{p}$

$$\begin{aligned}\bar{p} &= \frac{\text{Total number of defective units in all the samples}}{\text{Total number of units in all the samples}} \\ &= \frac{25}{100 \times 10} = \underline{\underline{0.025}}\end{aligned}$$

2<sup>nd</sup> step : Computing the Upper Control Limit

$$\begin{aligned}
 UCL &= \bar{P} + 3\sqrt{\bar{P}(1-\bar{P})/n} \\
 &= 0.025 + 3\sqrt{0.025(1-0.025)/100} \\
 &= 0.025 + 3 \times \sqrt{0.000244} \\
 &= 0.025 + 0.047
 \end{aligned}$$

$$UCL = \underline{\underline{0.072}}$$

3<sup>rd</sup> step : Computing the Lower Control Limit

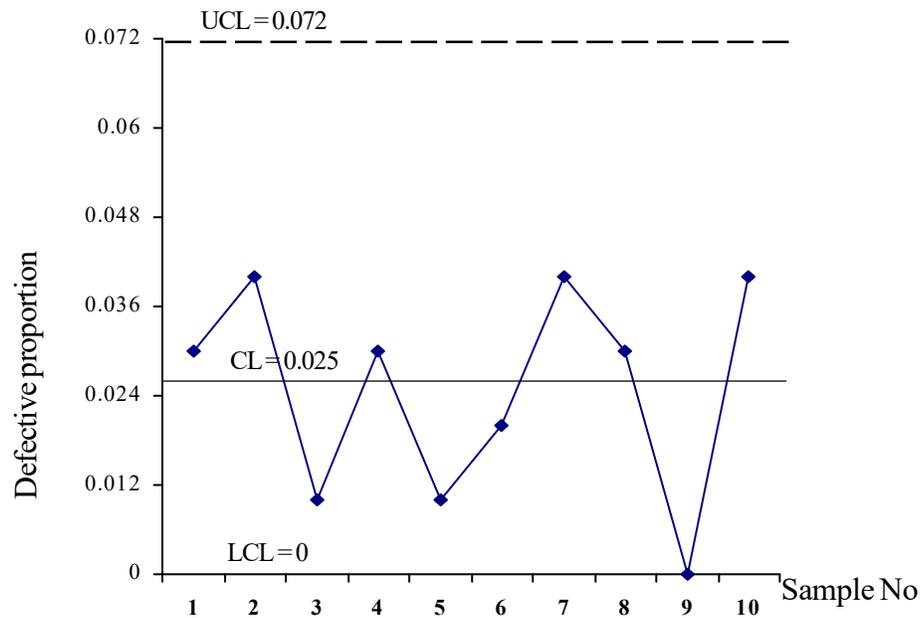
$$\begin{aligned}
 LCL &= \bar{P} - 3\sqrt{\bar{P}(1-\bar{P})/n} \\
 &= 0.025 - 3\sqrt{0.025(1-0.025)/100} \\
 &= 0.025 - 0.047 = -0.022
 \end{aligned}$$

$$LCL = \underline{\underline{0}}$$

4<sup>th</sup> step : Computing the sample proportions

|                   |                 |                 |                 |                 |                 |                 |                 |                 |                 |                 |
|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Sample N0.        | 1               | 2               | 3               | 4               | 5               | 6               | 7               | 8               | 9               | 10              |
| Sample proportion | $\frac{3}{100}$ | $\frac{4}{100}$ | $\frac{1}{100}$ | $\frac{3}{100}$ | $\frac{1}{100}$ | $\frac{2}{100}$ | $\frac{4}{100}$ | $\frac{3}{100}$ | $\frac{0}{100}$ | $\frac{4}{100}$ |
|                   | 0.03            | 0.04            | 0.01            | 0.03            | 0.01            | 0.02            | 0.04            | 0.03            | 0               | 0.04            |

5<sup>th</sup> step : Constructing the proportionate control chart



6<sup>th</sup> step : Interpretation

- The process is under control since all the sample proportions fall between the control Limits.
- Provide the students with the following activity to explain the **np** chart.

**Activity 2 :** A random sample of 50 units is drawn from a production process daily and checked number of defective units found in 10 days is as follows.

| Day                    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------------------|---|---|---|---|---|---|---|---|---|----|
| No. of defective units | 3 | 3 | 4 | 1 | 2 | 1 | 2 | 2 | 4 | 8  |

1. Compute the control limits in order to construct a control chart for controlling the number of defective units in each sample.
2. Build up the **np - chart** by plotting the number of defective units in each sample.
3. Enquire the situation of the control in the process.

**Solution - Activity 2 :**

1<sup>st</sup> step : computing  $\bar{p}$

$$\bar{p} = \frac{\text{Total number of defective in all the samples}}{\text{Total number of units in all the samples}}$$

$$\bar{p} = \frac{30}{50 \times 10} = 0.06$$

**2<sup>nd</sup> step :** Computing of  $n\bar{p}$

$$n\bar{p} = 500 \times 0.06 = 3$$

**3<sup>rd</sup> step :** Computing the Upper Control Limit

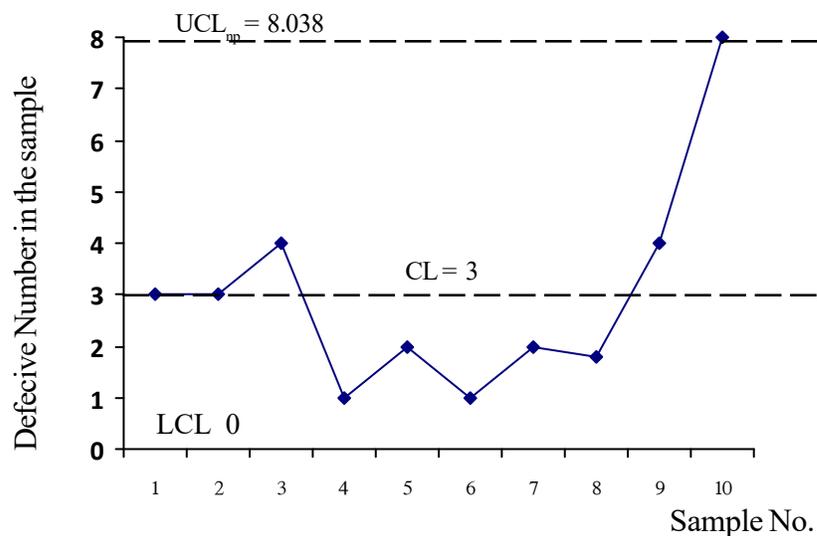
$$\begin{aligned} UCL &= n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})} \\ &= 3 + 3\sqrt{3 \times 0.94} \\ &= 3 + 3\sqrt{2.82} \\ &= 3 + 5.038 \\ &= \underline{\underline{8.038}} \end{aligned}$$

**4<sup>th</sup> step :** Computing the Lower Control Limit

$$\begin{aligned} LCL &= n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})} \\ &= 3 - 3\sqrt{3 \times 0.94} \\ &= 3 - 5.038 \\ &= -2.038 \\ &= \underline{\underline{0}} \end{aligned}$$

∴ The Lower Control Limit is considered as '0' since it is a negative value.

**5<sup>th</sup> step :** Constructing np chart



**6<sup>th</sup> step :** Interpretation of the controlling condition of the process.

Since all the points are fallen between the control limits this process is under control, but it would be better to be vigilant about the process, since the last observation falls very closer to the upper control limit.

- Involve the students in following activity to explain the c-chart

**Activity 3 :**

- Divide the students in the class into 3 groups and provide with 3 different  $A_4$  size type set materials including some errors, with following instructions.
  - Count the number of errors contained in the printed material you received.
  - Any group member may read that number aloud to the class.
  - One of the group members come to the board and write that number.
- Hold a discussion highlighting the following facts.
  - The number of errors (mistakes) found in a type set page can not be measured using a measuring instrument.
  - The quality of the product can be controlled by ruling out the number of errors contained in a unit produced.
  - Controlling the number of errors contained in a unit produced is also considered as controlling of attributes.
  - **C-chart** can be used in order to control the number of errors contained in a unit in the production with the purpose of assuring the quality.
  - How the number of errors contained in each unit are varied from average (mean) number of errors is depicted by c- chart.
  - Such number of errors found in random in a product can be considered to follow a position distribution.
  - The central line of a c-chart is taken as the average number of errors and the upper control limit is taken by adding 3 standard deviation to the mean number of errors.
- After the discussion provide with the following instructions to the students.
  - Find the mean of the number of errors mentioned on the board referring to the type set  $A_4$  pages and name it as C.
  - Find the square root of that mean number of errors ( $\sqrt{C}$ ) reasonify for  $\sqrt{C}$  being considered as the standard deviation.
  - Multiply the received answer by 3 and add to the mean number of errors,
  - Name it as the upper control limit of the c-chart.
  - Observe whether the number of errors contained in each page falls between the control limits and mention your conclusion.
- Provide with the following activity to the students.

**Activity 4 :**

“20 iron plates were drawn in random from a process of producing same size plates and the number of flaws contained in each plate was observed as follows.

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 7 | 6 | 5 | 4 | 3 | 4 | 5 | 7 | 3 | 4 |
| 5 | 6 | 6 | 4 | 4 | 3 | 4 | 3 | 5 | 5 |

- Construct a c - chart and point out whether the number of flaws in a plate is under control or not.

**Solution : Activity 4**

**Step 1 :** Computing the mean of the number of flaws. If the number of flaws in a plate is C and the number of plates is K.

$$\bar{C} = \frac{\sum Ci}{k}$$

$$\bar{C} = \frac{93}{20}$$

$$\bar{C} = \underline{\underline{4.65}}$$

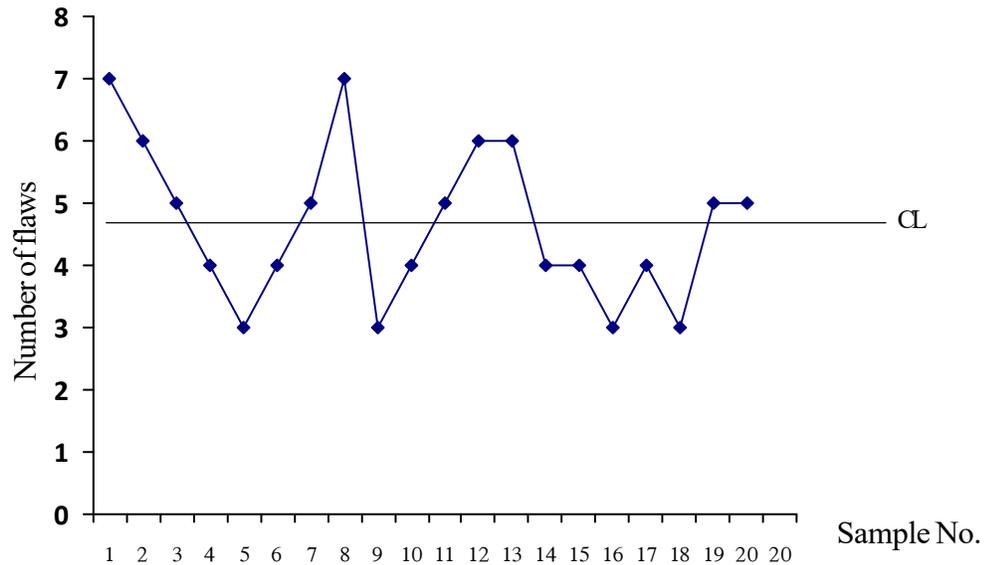
**Step 2 :** Computing the Upper Control Limit

$$\begin{aligned} UCL &= \bar{C} + 3\sqrt{\bar{C}} \\ &= 4.65 + 3\sqrt{4.65} \\ &= 4.65 + 3 \times 2.156 \\ &= 4.65 + 6.47 \\ &= \underline{\underline{11.12}} \end{aligned}$$

**Step3 :** Computing the Lower Control Limit

$$\begin{aligned} LCL &= \bar{C} - 3\sqrt{\bar{C}} \\ &= 4.65 - 3\sqrt{4.65} \\ &= 4.65 - 3 \times 2.156 \\ &= 4.65 - 6.47 \\ &= -1.82 \\ &= 0 \end{aligned}$$

**Step 4:** Constructing the control chart and plotting out the number of flaws in each plate on it.



**Step 5 :** Since all the number of flaws are fallen between the control limits the process is under control.

- Inquire from the students which control chart is suitable for following situations in order to explain the U-chart.
  1. Number of air bubbles contained in a 6' x 4' plastic plate
  2. Number of flaws found in a 1m<sup>2</sup> piece of cloth
  3. Number of flaws found in a shirt of collar size 16"
  4. Number of defects found in a particular kind of motor vehicle
  5. Number of defects found in a building of a building complex constructed by a construction company
- Hold a discussion highlighting the following facts based on the students' responses.
  - that the c-chart is appropriate for controlling the number of flaws in a 6' x 4' plastic plate, since similar type of flaws found in a single unit are needed to be controlled.
  - that the c chart is appropriate for controlling the number of flaws in a 1m<sup>2</sup> piece of cloth, since it is also a single space.
  - that the c-chart is appropriate for controlling the number of flaws found in a shirt of collar size 16" . Since it is a single space.
- C-chart is not appropriate for controlling the number of defects found in a motor vehicle, because a motor vehicle is a large size item which consists of various separate parts (components). There can be found various types of errors in various parts such

as the errors found in the engine, gear system, electric systems in the body etc. The space of each component is not similar. Therefore it seems to be needed a different type of a control chart. In order to control the defects in a produced item in such a situation u-chart is appropriate.

- In order to control the defects found in a building complex also the c-chart is not suitable. A building is also a large item, consists of various components. The roof, walls, the electricity system, bath room and sanitary facilities, air conditioning system and drainage system etc. The u-charts are applied in controlling the errors in such situations as well.
- Once there are immeasurable qualitative characteristics contained in large scale production items, having separated that entire unit in to sub-sections (components) and considering the number of defects found in each sub-section as u, in order to control the defects in that entire unit, the u-chart is constructed.
- Involve the students in the following activity in order to construct the u-chart.

#### Activity – 5

- Number of defects observed in 12 container trucks manufactured in a heavy vehicle manufacturing complex are as follows.

4 3 6 5 4 4 5 6 7 4 8 4

- Compute the average number of defects found in a container truck ( $\bar{u}$ ) considering the number of defect is each as  $U_i$ .
- Derive the control limits to construct the U - chart as follows.

$$UCL_u = \bar{U} + 3\sqrt{\bar{u}}$$

$$CL_u = \bar{U}$$

$$LCL_u = \bar{U} - 3\sqrt{\bar{u}}$$

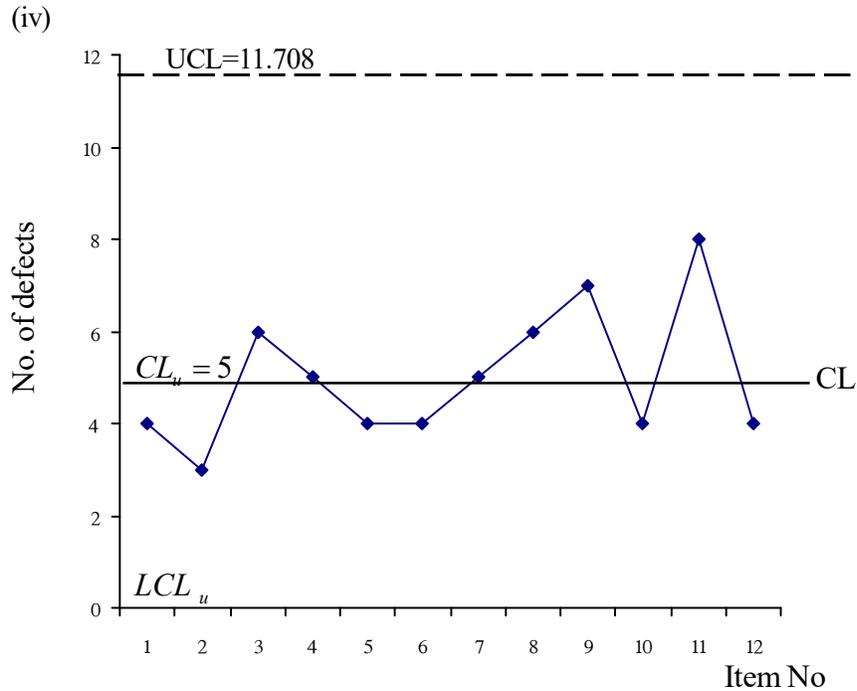
- Draw the control limits for U-chart on a co-ordinate plane and plot the number of defects in each container truck on it.
- State whether the process of manufacturing the container trucks is under control or not.

#### Situation : Activity –5

$$(i) \quad \bar{U} = \frac{\sum_i}{k} = \frac{4+3+6+5+4+4+5+6+7+4+8+4}{12} = 5$$

$$(ii) \quad UCL = \bar{u} + 3\sqrt{\bar{u}} = 5 + 3\sqrt{5} \\ = 5 + 3 \times 2.236 \\ = 11.708$$

$$\begin{aligned}
 \text{(iii) } LCL &= \bar{u} - 3\sqrt{\bar{u}} \\
 &= 5 - 3\sqrt{5} \\
 &= 5 - 3 \times 2.236 \\
 &= -1.708 \\
 &= \underline{\underline{0}}
 \end{aligned}$$



- The process is under control, since all the points are between the control Limits.

**A Guideline to explain the subject matters.**

- Some of the qualitative characteristics in a product are not convenient to measure. Such attributes can be evaluated on the fact whether those are appeared or not in the product.
- Once the products are not in accordance with the specimens, they are categorized as defectives and once the products are in accordance with their specimens they are categorized as non defectives.
- At instances where qualitative characteristics of a product can not be measured using a measuring unit, in order to control their quality, the attribute control can be applied.
- Following control charts can be applied for controlling the attributes.
  - (i) Proportionate Control chart (p chart)
  - (ii) Defective units number control chart (np chart)
  - (iii) Unit defective number control chart (c – chart)

(iv) Parts of the unit defective number control chart (u chart)

- The objective of a proportionate control chart is to check whether the quality of a product has achieved at the pre-determined quality level through controlling the sample defective proportions.
- How the sample defective proportions are varied about the mean defective proportions is depicted by the proportionate control chart (p-chart)
- In order to determine the control limits of the proportionate control chart the mean of the sampling distribution  $\mu_p$  and the standard deviation  $\sigma_p$  are needed.
- Since it is assumed the defective proportions of the sample exceeding 100 units follow

a normal distribution and the standard deviation of the sample proportions is  $\sqrt{\frac{\bar{P}(1-\bar{p})}{n}}$

the control limits for the proportionate chart are determined as follows.

- Centered Line  $CL = \bar{p}$
- Upper Control Limit  $UCL = \bar{p} + 3\sqrt{\frac{\bar{P}(1-\bar{p})}{n}}$
- Lower Control Limit  $LCL = \bar{p} - 3\sqrt{\frac{\bar{P}(1-\bar{p})}{n}}$

- $\bar{p}$  - can be computed using the sample details as follows :

$$\bar{p} = \frac{\text{Total number of defective units in all the samples}}{\text{Total number of observations in all the samples}}$$

- If all the sample proportions fall between the control limits, the process further consists of only the random variations, so that the process can be determined to be under control.
- For the convenience of the quality control officer who is involved in constructing the control charts, the unit defective number control chart (np-chart) can be constructed instead of the proportionate chart in each sample directly as well.
- The mean of the unit defective numbers is  $n\bar{p}$  and the standard deviation is  $3\sqrt{n\bar{p}(1-\bar{p})}$  so that the control limits of the control chart can be computed as follows.

Central Line  $CL = n\bar{p}$

Upper Control Limit  $UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$

Lower Control Limit  $LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$

- $n\bar{p}$  is derived as the product of the mean of the sample defective proportions ( $\bar{p}$ ) and the sample size (n).
- How the number of defectives in each sample is varied about the mean number of defectives in a sample is depicted in the **np-chart**.
- The objective of a **c-chart** is to check the quality of a product through controlling the number of defects (**flaws**) contained in a production unit considered.
- C-charts are applied to control the defective conditions such as the number of flaws found in a roll of cloth, number of errors in a printed page, number of air bubbles found in 1m<sup>2</sup> piece of glass etc.....
- Once the number of defectives is considered as a number of events appeared in a single unit of space the sampling distribution of the total number of defectives can be approximated using the poisson distribution.
- The control limits for c- chart can be derived as follows.

$$\text{Central Line} \quad CL = \bar{c}$$

$$\text{Upper Control Limit} \quad UCL = \bar{c} + 3\sqrt{\bar{c}}$$

$$\text{Lower Control Limit} \quad LCL = \bar{c} - 3\sqrt{\bar{c}}$$

- If the LCL takes a negative value then it is considered LCL = 0
- In order to control over the defects appeared in large production items the **u-chart** can be applied as an attribute control chart.
- Separating such a large item into sub sections (components) and considering the number of defects in each sub section as **u**, the u-chart is constructed.
- The objective of the **u-chart** is to control the quality of a large item by controlling the average number of defects in it.
- Control limits for the chart can be derived as follows.

$$\text{Central Line} \quad CL = \bar{u}$$

$$\text{Upper Control Limit} \quad UCL = \bar{u} + 3\sqrt{\bar{u}}$$

$$\text{Lower Control Limit} \quad LCL = \bar{u} - 3\sqrt{\bar{u}}$$

**Competency 10.0** : Uses statistical quality control techniques to make managerial decisions.

**Competency level 10.4** : Uses appropriate methodologies for product control

**No. of periods** : 10

**Learning outcomes :**

- Introduces ‘Product Control.’
- Explains the Acceptance sampling plan which is used for product control.
- Introduces the single sampling plan.
- Explains how to build up a sampling plan.
- Introduces the Operational Characteristic Curve (OC – Curve)
- Collects the details required for constructing an OC – curve.
- Constructs an OC – curve.
- Interprets the Acceptable Quality Level (AQL)
- Points AQL on OC curve.
- Interprets the producer’s risk.
- Interprets the Lot Tolerance Proportion of Defectives – (LTPD)
- Points LTPD on OC-curve
- Interprets the consumer’s risk.
- Plots the producer’s risk and consumer’s risk on OC – curve.
- Computes AQL, LTPD, producer’s risk and the consumer’s risk.
- Explains the uses of operational characteristic curve.
- Describes the properties of a good acceptance sampling plan.

**Instructions for Lesson Planning :**

- Pay attention of students to the following situations.
  - Testing of paddy samples brought by farmers to be sold at paddy collecting centers of the Paddy Marketing Board.
  - Testing of fuel samples by the Petroleum Corporation for purchasing fuel stocks from foreign companies.

- Hold a discussion highlighting the following facts.
  - That the buyer is interested to ensure the goods are in a condition to meet the agreed quality level before the batches of items are to be bought.
  - Point out that a particular testing method is required to determine which batches of items are good and which are bad.
  - State that testing all the units in a batch is time consuming and expensive.
  - Highlight that many firms make the decision whether to purchase or not a particular batch of items, based on a sample drawn from that whole batch.
  - State that a statistical plan called an acceptance sampling plan can be used to decide whether a batch is accepted or rejected having tested a sample.
  - Point out that basically there are 3 elements in an acceptance sampling plan.
  - Present the following sampling plan to the students and lead them to complete the table given below for constructing an operational characteristic curve.

**Activity 01 :**

**Sampling Plan :**

- A firm purchasing raw materials, makes their purchases in batches containing 2 500 units in each, after testing a random sample of 10 units, if the number of defective units contained in the sample is less than or equal two, the batch will be accepted and if not rejected.
- Complete the following table.
- Use the function of a bi-nominal distribution  ${}^n C_x \cdot p^x \cdot q^{n-x}$  in completing the table.

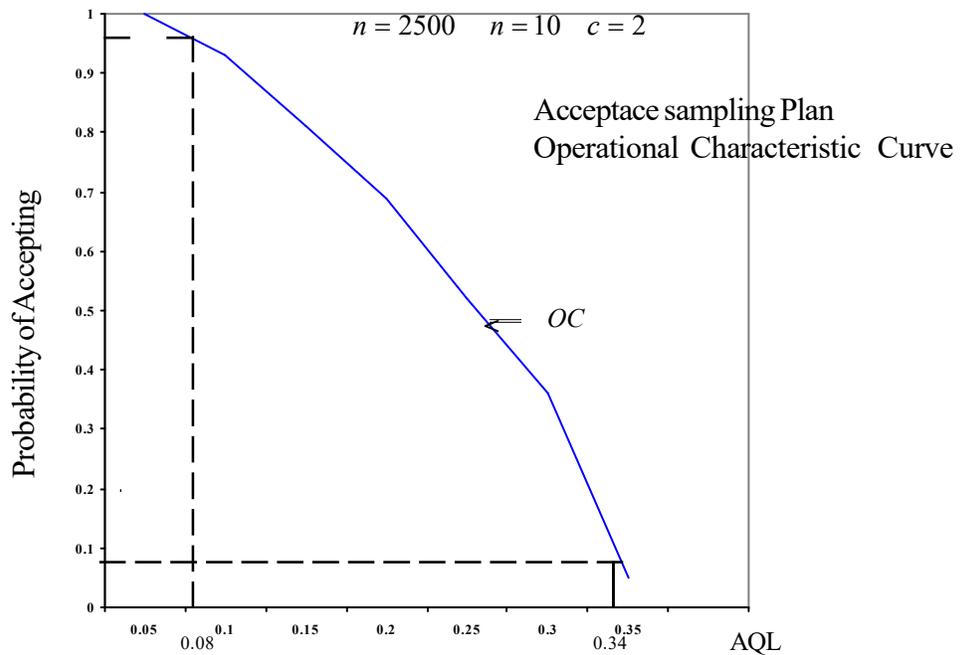
| $N = 2500$ |                      | $n = 10$    |            |            | $c = 2$  |
|------------|----------------------|-------------|------------|------------|--|
| Batch      | Defective Percentage | Probability |            |            | Probability of accepting the batch<br>$P(0) + P(1) + P(2)$ |
|            |                      | $P(x = 0)$  | $P(x = 1)$ | $P(x = 2)$ |  |
| A          | 5%                   | .....       | .....      | .....      | .....  |
| B          | 10%                  | .....       | .....      | .....      | .....  |
| C          | 15%                  | .....       | .....      | .....      | .....  |
| D          | 20%                  | .....       | .....      | .....      | .....  |
| E          | 25%                  | .....       | .....      | .....      | .....  |
| F          | 30%                  | .....       | .....      | .....      | .....  |

- After completing the table provide with following instructions for constructing the OC-curve.
- Use a graph paper.
- Prepare the appropriate scales on the graph paper representing the defective percentage on horizontal axis and the probabilities of accepting on vertical axis for constructing the OC-curve.
- Getting each defective percentage and corresponding acceptance probability values as co-ordinates mark those points on the co-ordinate plane.
- Join all those points using a smooth line.
- Stretch the right end of the smooth line to little bit far, closer to the horizontal axis but not touching it. (Since touching the smooth curve on horizontal axis is not practical expected).
- Name that smooth curve as the OC-curve.
- Depict the OC-curve you have derived to the class.
- Hold a discussion highlighting the following facts.
  - Point out that the graph depicting the relationship between possible defective percentages in each batch and the corresponding probability value of accepting the batch is known as the OC-curve.
  - Point out the shape of the OC-curve depends on the size of the sample in the sampling plan and the acceptance number (-c), but further highlight that the gradient of the OC-curve is highly affected by sample size n.
  - Point out further that steeper the OC-curve greater the consumer's protection.
  - In accordance with the OC-curve constructed by students, guide them to mark the possible maximum proportion of defectives that may be found in a batch which is accepted by consumer at a maximum probability of accepting as 0.95.
  - Point out that, that level is known as Acceptable Quality Level (AQL).
  - In accordance with the OC-curve constructed guide the students to mark the minimum proportion of defectives that may contain in a batch which is accepted by consumers at 0.10 minimum probability level.

- Point out to the students that, that is known as the Rejected Quality Level (RQL) or Lot Tolerance Proportion of Defectives (LTPD).
- Discuss the uses of the OC curve using the curve which AQL and RQL were plotted out.

**Solution - Activity 01 : Completed table**

| $N = 2500$ |                      | $n = 10$    |            |            | $c = 2$  |
|------------|----------------------|-------------|------------|------------|--|
| Batch      | Defective Percentage | Probability |            |            | Probability of accepting the batch<br>$P(0) + P(1) + P(2)$ |
|            |                      | $P(x = 0)$  | $P(x = 1)$ | $P(x = 2)$ |  |
| A          | 5%                   | 0.5987      | 0.3151     | 0.0746     | 0.9884   |
| B          | 10%                  | 0.3487      | 0.3874     | 0.1937     | 0.9298   |
| C          | 15%                  | 0.1969      | 0.3474     | 0.2759     | 0.8202   |
| D          | 20%                  | 0.1074      | 0.2684     | 0.3020     | 0.6778   |
| E          | 25%                  | 0.0563      | 0.1877     | 0.2816     | 0.5256   |
| F          | 30%                  | 0.0282      | 0.1211     | 0.2335     | 0.3828   |



- Involve the students in following activity to compute the producer's risk and consumer's risk.

**Activity 2 :**

- If  $N = 1000$ ,  $n = 10$  and  $c = 1$  separate each batch given below as good batches and bad batches.
  - A: A batch of 1000 units including 50 defective units.
  - B: A batch of 1000 units including 100 defective units
  - C: A batch of 1000 units including 150 defective units.
  - D: A batch of 1000 units including 20% defective units.
- Compute the following table with your findings.

| Batch | Defective Proportions | Good/bad | Probability of Accepting | Probability of Rejecting | Who is a risk possible for? | That risk |
|-------|-----------------------|----------|--------------------------|--------------------------|-----------------------------|-----------|
| A     |                       |          |                          |                          |                             |           |
| B     |                       |          |                          |                          |                             |           |
| C     |                       |          |                          |                          |                             |           |
| D     |                       |          |                          |                          |                             |           |

- Hold a discussion highlighting the following facts.
- That, there is a possibility of rejecting even a good batch by the consumer according to the acceptance sampling plan and that is considered as the type-I error.
- That, there is a possibility of accepting even a bad batch by the consumer according to the acceptance sampling plan and that is considered as type II error.
- That the probability of type I error is known as the producer's risk and the probability of type II error is known as the consumer's risk.

**Solution : Activity 2**

- A batch of which the defective proportion is not exceeding  $c/n$  of the acceptance sampling plan is known as a good batch.

$$\frac{c}{n} = \frac{\text{Acceptance Number}}{\text{Size of the Acceptance Sample}}$$

- In this situation  $\frac{1}{10} = 0.1$
- Accordingly, the batches, whose defective proportions are not exceeding 0.1 are accepted as good batches, where as the batches with defective proportions are exceeding with 0.1 are called bad batches.
- Batch A is good since the defective proportion is  $\frac{50}{1000} = 0.05$
- Batch B is good since the defective proportion is  $\frac{100}{1000} = 0.1$
- Batch C is not good since the defective proportion is  $\frac{150}{1000} = 0.15$
- Batch D is not good since the defective proportion is given as 0.20
- If the number of defective units in the sample is X.

- Probability that the batch A being accepted

$$P(X \leq C) = P(X \leq 1) = P(X = 0) + P(X = 1)$$

Since  $n = 10$  and  $p = 0.05$  using the binomial distribution.

$$P(X \leq 1) = 0.5987 + 0.3151 = 0.9138$$

$$\begin{aligned} \therefore \text{Probability of rejecting the batch A} &= 1 - P(X \leq C) \\ &= 1 - P(X \leq 1) \\ &= 1 - 0.9138 \\ &= \underline{0.0862} \end{aligned}$$

- Probability of accepting the batch B

$$= P(X \leq C) = P(X \leq 1) = P(X = 0) + P(X = 1)$$

Since  $n = 10$  and  $p = 0.1$  using the binomial distribution

$$P(X \leq 1) = 0.3487 + 0.3874 = 0.7361$$

$\therefore$  The probability of the batch B being rejected.

$$\begin{aligned} &= 1 - P(X \leq C) \\ &= 1 - 0.7361 \\ &= \underline{0.2639} \end{aligned}$$

- The probability that the batch C being accepted

$$= P(X \leq C) = P(X \leq 1) = P(X = 0) + P(X = 1)$$

Since  $n = 10$ ,  $p = 0.15$  using the binomial distribution.

$$p(x \leq 1) = 0.1969 + 0.3471 = 0.5440$$

∴ Probability of rejecting the batch C is

$$= 1 - P(X \leq 1)$$

$$= 1 - 0.5440$$

$$= \underline{\underline{0.4560}}$$

- The probability that the batch D being accepted

$$= P(X \leq C) = P(X \leq 1) = P(X = 0) + P(X = 1)$$

Since  $n = 10$  and  $p = 0.20$  using the binomial distribution

$$= P(X \leq 1) = 0.1074 + 0.2684 = \underline{\underline{0.3758}}$$

- Probability that the batch D being rejected.

$$= 1 - P(X \leq 1)$$

$$= 1 - 0.3758$$

$$= \underline{\underline{0.6242}}$$

- Solution completed table.

| Batch | Defective Proportions | Good/bad | Probability of Accepting | Probability of Rejecting | who is a risk possible for ? | That risk |
|-------|-----------------------|----------|--------------------------|--------------------------|------------------------------|-----------|
| A     | 0.05                  | Good     | 0.9138                   | 0.0862                   | Producer                     | 0.0862    |
| B     | 0.10                  | Good     | 0.7361                   | 0.2639                   | Producer                     | 0.2639    |
| C     | 0.15                  | Bad      | 0.5440                   | 0.4560                   | Consumer                     | 0.5440    |
| D     | 0.20                  | Bad      | 0.3758                   | 0.6242                   | Consumer                     | 0.3758    |

- Involve the students in the following activity to study about using the poisson distribution as an approximation to the binomial distribution in acceptance sampling plan.

**Activity 3 :**

1. In accordance with an acceptance sampling plan  $n=40$  and  $C=2$ . Compute the probability of accepting each batch mentioned below.
  - (i) A batch with 1% defectives.
  - (ii) A batch with 2% defectives.
  - (iii) A batch with 5% defectives.
  - (iv) A batch with 6.25% defectives.
2. Find the producer's risk at the 1<sup>st</sup> situation above and the consumer's risk at the iv<sup>th</sup> situation.

**Solution : Activity 3**

(1) (i) Defective proportion of the batch  $= \frac{1}{100} = 0.01$

Since  $n = 40$ ,  $P = 0.01$ ,  $\lambda = np = 40 \times 0.01 = 0.4$

According to the plan, the probability of accepting

$$= P(X \leq C) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0.6703 + 0.2681 + 0.0536$$

$$= \underline{\underline{0.9920}}$$

(ii) Since the defective proportion of the batch is 0.02

$P = 0.02$ ,  $n = 40$ ,  $\lambda = np = 40 \times 0.02 = 0.8$

According to the plan, the probability of accepting

$$= P(X \leq C) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0.4493 + 0.3595 + 0.1438$$

$$= \underline{\underline{0.9526}}$$

(iii) Since the defective proportion of the batch is 0.05

$P = 0.05$ ,  $n = 40$ ,  $\lambda = np = 40 \times 0.05 = 2$

According to the plan, the probability of accepting

$$= P(X \leq C) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0.1353 + 0.2707 + 0.2707$$

$$= \underline{\underline{0.6767}}$$

(iv) Since the defective proportion of the batch is 6.25%.

$$\begin{aligned}
 P &= 0.0625 \quad n = 40 \quad \lambda = np = 40 \times 0.0625 = 2.5 \\
 &= P(X \leq C) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\
 &= 0.0821 + 0.2052 + 0.2565 \\
 &= \underline{\underline{0.5438}}
 \end{aligned}$$

2 (i) According to the plan in the 1<sup>st</sup> situation a good batch is a batch containing a defective proportion of 0.05 or less

$$\left( \frac{c}{n} = \frac{2}{40} = 0.05 \right)$$

Since the defective proportion of the first batch above is 0.01, that is a good batch. The probability of a good batch being rejected by the consumer is the producer's risk.

It is denoted by  $\alpha$

$$\begin{aligned}
 \alpha &= P(X > C) = 1 - P(X \leq C) \\
 &= 1 - P(X > 0) + P(X = 1) + P(X = 2) \\
 &= 1 - 0.6703 + 0.2681 + 0.0536 \\
 &= 1 - 0.9920 \\
 &= \underline{\underline{0.008}}
 \end{aligned}$$

(iv) According to the plan in iv<sup>th</sup> situation batch containing a defective proportion exceeding 0.05 is a bad batch  $\frac{c}{n} = \frac{2}{40} = 0.05$  Since the defective proportion in iv<sup>th</sup> batch is 0.0625, that is a bad batch. The probability of accepting a bad batch is the consumer's risk. It is denoted as  $\beta$

$$\begin{aligned}
 \beta &= P(X \leq C) = P(X = 0) + p(X = 1) + p(X = 2) \\
 &= 0.0821 + 0.2052 + 0.2565 \\
 &= \underline{\underline{0.5438}}
 \end{aligned}$$

**A Guideline to explain the subject matters. :**

- In some situations all the units in the batch can be checked.
- Determining whether a batch is accepted or rejected by checking all the units is known as a 100% inspection. (exhaustive inspection)

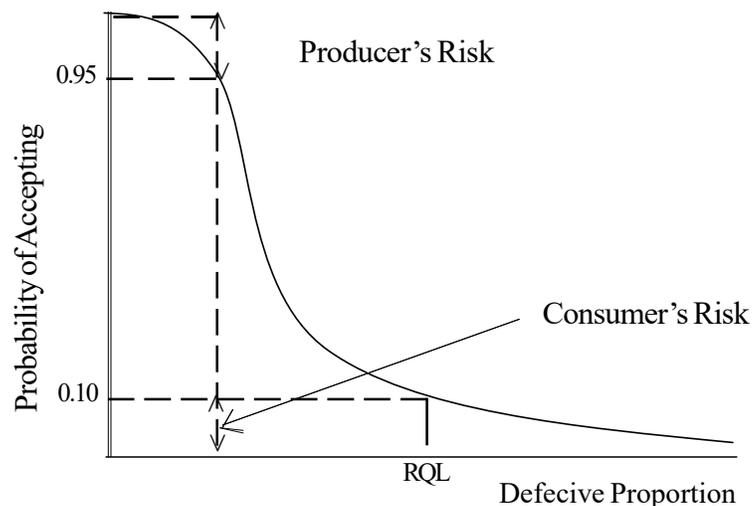
- A 100% inspection is a time consuming and very expensive method.
- 100% inspections is not practical when the test elements are destructive each in testing.
- Testing of samples from each batch is the most frequenting method currently used in industrial field, in making decisions for accepting or rejecting the batches.
- A statistical technique known as an acceptance sampling plan is applied by most of the modern manufacturing firms in order to check whether a batch is accepted or rejected.
- The purchaser (consumer) uses such a plan to ascertain that the goods are in accordance with the pre-determined standards and specimens prior to make purchases.
- Once the raw materials are purchased as well as the finished goods are released to the market, the product control led by the acceptance sampling plan is used.
- Determining whether to accept or reject a batch using one sample drawn from that batch is known as single sampling plan.
- A single sampling plan is implemented as follows.
  1. A sample with 'n' number of units is drawn in random from a batch of N number of units
  2. Testing each and every sample element separately and then, they are separated in to two groups as defective and non-defective.
  3. If the number of defective units in the sample does not exceed the pre-determined acceptance number (C) that batch is accepted.
  4. If the number of defective units in the sample exceeds the pre-determined acceptance number then the batch is rejected.
- There are 3 elements in an acceptance sampling plan.
  - N – size of the batch
  - n – size of the sample
  - C – pre-determined acceptance number
- Here, **N** is the total number of units contained in a batch regarding which an acceptance sampling plan is to be launched, **n** is the number of units selected in random to be checked from a batch of units, **C** - the acceptance number meant for maximum number of defective units allowed to be in the sample for determining to accept the considered batch.
- Once the average defective proportion found in a batch is P and the number of defective units actually found in the sample checked is denoted by X, the probability that the batch being accepted by the consumer is as follows in accordance with the bi-nominal distribution.

$$pr(X \leq C) = \sum_{x=0}^C {}^n C_x P^x (1-p)^{n-x}$$

- If **n** is large and **P** is small (**np < 5**), the Poisson approximation is applied for binominal distribution, so that the probability of accepting the batch in accordance with the Poisson distribution is as follows.

$$P(X \leq C) = \sum_{x=0}^C \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{Here } \lambda = np$$

- The curve that depicts the relationship between each possible defective proportions in the batch and the probability of accepting the batch is known as Operational Characteristic Curve (OC –Curve).
- The bi-nominal probability function can be used for calculating the probabilities of accepting the batch. When the defective proportion of the batch **p** is small and the sample size (**n**) is large (when  $np < 5$ ) probability of accepting can be computed using poisson approximation for binominal distribution.
- Few uses of OC-curve are as follows.
  - Depicting the relationship between quality of a batch and the probability of acceptance.
  - Ability of observing the probability of accepting a batch at a given defective proportion.
  - Assuring the ability of seperating good and bad batches in accordance with an acceptance sampling plan.
  - The probability of rejecting a good batch ( producer’s risk) and the probability of accepting a bad batch (consumer’s risk) can be observerd using the OC curve as follows.



- In accordance with an acceptance sampling plan, the maximum defective proportion in a batch which is accepted with a higher probability level is known as the Acceptance Quality Level (AQL) or else in accordance with an acceptance sampling plan, the maximum proportion of defectives that may contain in a good batch which is accepted by the consumer.
- In accordance with an acceptance sampling plan the minimum defective proportion in a batch which is accepted with a lower probability level is known as the Lot Tolerance Proportion of Defectives or Rejectable Quality Level. (RQL/LTPD) or else in accordance with an acceptance sampling plan the minimum proportion of defectives that may contain in a bad batch rejected by the consumer.
- Since the acceptance sampling is also a technique of making decisions depend on samples errors are possible.
- Rejecting a batch containing a defective proportion less than  $\frac{c}{n}$  is known as type I error in acceptance sampling. The probability of type I error is denoted by  $\alpha$  and, that probability, is called the 'producer's risk'. That means the probability of rejecting a good batch is producer's risk.
- Accepting a batch containing a defective proportion greater than  $\frac{c}{n}$  is known as type II error in acceptance sampling. The probability of type II-error is denoted by  $\beta$  and that probability is called 'consumer's risk'. That means the probability of accepting a bad batch is consumer's risk.

**Competency 11.0 :** Uses index numbers for making business decisions

**Competency Level 11.1 :** Builds up the basis to study index numbers

**No. of Periods :** 02

**Learning Outcome :**

- Defines what an index number is.
- Lists the uses of index numbers.
- States the problems faced in constructing an index number.

**Instructions for Lesson Planning :**

- Present the following situations to the students and hold a discussion raising questions on each.

Situation 01 : Values of colombo consumers price index in last few years.  
150%, 140%, 165%

Situation 02 : Values of import volume indices in Si Lanka in past few years.  
110%, 80%, 225%

Situation 03 : Values of all share price index in Colombo Stock Exchange in past few weeks.  
125%, 105%, 240%

1. Inquire what has been measured by those indices at each time.

- Ascertain that
- The changes in prices of consumer goods and services are evaluated at the first occasion.
- The changes in imported goods and services are examined at the second occasion.

And

- The changes in share prices of the stock exchange (share market) are reviewed at the third occasion.

2. Inquire how they have been computed.

- They have been computed as percentages.

- They have been computed in relevant with a particular period
  - Monthly
  - Annually
  - Weekly

### **Activity – 1**

1. Explain the changes in variable revealed through the values of the indices in 3 situations mentioned above.
2. State few uses of computing the index number you received.
3. State the possible problems that may be arisen in computing the index numbers.

### **Answers for Activity – I**

- Explain that for the first occasion.
    - 150% means that the price has increased by 50%.
    - 140% means that the price has increased by 40%.
    - 165% means that the price has increased by 65%
  - Explain that for the 2<sup>nd</sup> situation
    - 110% means that import volume has increased by 10%
    - 80% means that import volume has decreased by 20%
    - 225% means that import volume has increased by 125%
  - Explain that for the 3<sup>rd</sup> situation
    - 125% means that the share price has increased by 25%.
    - 105% means that the share price has increased by 5%
    - 240% means that share price has increased by 140%.
2. Few uses of computing the index numbers can be highlighted as follows,
    - Ability of being aware of the changes in consumer goods and services
    - Ability of being aware of the changes in the quantity of imported goods and services.
    - Ability of being aware of the changes in share prices in the share market.
  3. Discuss the possible problems may be arisen in computing the index numbers.

### **A Guideline to explain the subject matters :**

- A statistical measure expressed as a ratio or a percentage, used to measure the change or variation of a particular variable is known as an index number.

Ex : These are used to measure the change in factors like prices, salaries and production volume etc.

- The changes in the variable are measured related to time, geographical location or any other characteristic
- Accordingly an index number means that the statistical measure which is used to evaluate the change of a particular variable related to any characteristic like time or the geographical location. This can be presented as a ratio or a percentage.
- Uses of index numbers can be listed out as follows.
  1. Ability of measuring the change in any variable related to time or any other factor.
  2. Ability of comparing the import and export volume of a country in two periods.
  3. Ability of making investment decisions of a company considering the prices of shares exchanged in the share market.
- Uses of a cost of living index are as follows.
  - To measure the changes in cost of living.
  - To compute the real salaries or the real income.
  - To measure the inflation
  - To evaluate the inner value of the money
- Problems arisen in computing the index members can be stated as follows.
  1. Problems arisen in data collection.
  2. Difficulties faced in drawing samples.
  3. Difficulties in choosing the base period.
  4. Difficulties in determining the relevant group of consumers.
  5. Difficulties arisen in selecting the bag of items containing the consumer goods and services to be included in index.
  6. The problem of weighing according to the relative importance of goods and services.

**Competency 11.0 :** Uses index numbers for making business decisions

**Competency Level 11.2 :** Uses simple relative indices for measuring the relative change in a single variable.

**No. of Periods :** 06

**Learning Outcomes :**

- Explains simple relative indices
- Introduces the simple price relative and computes it.
- Introduces the simple quantity relative and computes it
- Introduces the simple value relative and computes it,
- Explains the properties of simple relative indices
- Explains the identical property.
- Explains the time reversal property.
- Explains the factor reversal property.
- Explains the cyclical property.
- Point out the weak points of simple relative indices.

**Instructions for Lesson Planning :**

- Present the following description to the students and solve the problems mentioned through the discussion.

The price of 1kg of sugar in 2010 was Rs. 75/= and it had been increased up to 90/= in 2015 . Further the average monthly sugar consumption was 8 kg and it had fallen down to 6 kg in 2016.

- According to the above description measure ;
  - How the price of sugar changed.
  - How the quantity of sugar consumption changed.
- The change in the price of sugar

$$\frac{\text{Price in 2015}}{\text{Price in 2010}} \times 100 = \frac{90}{75} \times 100$$
$$= 120\%$$

- Price of sugar has increased by 20% in 2015 relative to 2010.

- The change in sugar consumption;

$$\frac{\text{Quantity bought in 2015}}{\text{Quantity bought in 2010}} \times 100 = \frac{6\text{kg}}{8\text{kg}} \times 100 = 75\%$$

The quantity of sugar bought in 2015 has decreased by 25% relative to 2010.

- The change in expenses incurred for sugar consumption.

$$\begin{aligned} \frac{\text{Expenses in 2015}}{\text{Expenses in 2010}} \times 100 &= \frac{90 \times 6}{75 \times 8} \times 100 \\ &= \underline{\underline{90\%}} \end{aligned}$$

The expenses incurred for sugar in 2015 has decreased by 10% relative to 2010.

- Ascertain that the change in price, quantity and value of a single variable can be measured between two periods as mentioned above. Explain that the indices used in that purpose are called simple relative indices.

### Activity 01 :

- The price of 1kg of fish in few years are as follows. Answer the questions given below using that.

| Kind of Fish | 2012       | 2014       | 2016       |
|--------------|------------|------------|------------|
|              | $P_{2012}$ | $P_{2014}$ | $P_{2016}$ |
| Shark        | 200        | 350        | 500        |
| Para         | 150        | 200        | 400        |
| Tuna         | 100        | 300        | 400        |

- (i) Using the formula  $\frac{P_{2016}}{P_{2012}} \times 100$  compute the changes in price of each kind of fish and interpret the answer.

(ii) Derive  $\frac{P_{2016}}{P_{2016}}$

(iii) Derive  $\frac{P_{2016}}{P_{2012}} \times \frac{P_{2012}}{P_{2016}}$

(iv) Derive  $\frac{P_{2016}}{P_{2012}} \times \frac{P_{2012}}{P_{2014}} \times \frac{P_{2014}}{P_{2016}}$

- Hold a discussion using the responses given to the above Activity.

(1). Once the formula  $\frac{P_{2016}}{P_{2012}} \times 100$  is used the change in the price of each kind of fish is as follows.

- Shark  $\frac{500}{200} \times 100 = 250\%$

The price of the Shark has increased by 150% in 2016 relative to 2012.

- Para  $\frac{400}{150} \times 100 = 266.7\%$

The price of the Para has increased by 166.7% in 2016 relative to 2012.

- Tuna  $\frac{400}{100} \times 100 = 400\%$

The price of Tuna has increased by 300% in 2016 relative to 2012.

(2).  $\frac{P_{2016}}{P_{2016}}$

Shark -  $\frac{500}{500} = 1$       Para -  $\frac{400}{400} = 1$       Tuna -  $\frac{400}{400} = 1$

- Make the students aware that the value of simple price relative of a particular period relative to the same period is 1 or 100.

(3). The values derived for  $\frac{P_{2016}}{P_{2012}} \times \frac{P_{2012}}{P_{2016}}$  are as follows.

Shark -  $\frac{500}{200} \times \frac{200}{500} = 1$       Para -  $\frac{400}{150} \times \frac{150}{400} = 1$       Tuna -  $\frac{400}{100} \times \frac{100}{400} = 1$

- Explain that once the price index of a particular period is multiplied by the reciprocal

of the same index, the result derived is 1 or  $\frac{P_{2016}}{P_{2012}} \times \frac{1}{\frac{P_{2016}}{P_{2012}}} = 1$

(4). The value receive for  $\frac{P_{2016}}{P_{2012}} \times \frac{P_{2012}}{P_{2014}} \times \frac{P_{2014}}{P_{2016}}$  are as follows.

For shark  $\frac{500}{200} \times \frac{200}{350} \times \frac{350}{500} = 1$

For Para  $\frac{400}{150} \times \frac{150}{200} \times \frac{200}{400} = 1$

$$\text{For Tuna} \quad \frac{400}{100} \times \frac{100}{300} \times \frac{300}{400} = 1$$

**Activity 2 :**

The unit price and the quantity of three products consumed by a family including five members are as follows.

| Product         | 2010 |    | 2012 |    | 2015 |    |
|-----------------|------|----|------|----|------|----|
|                 | P    | Q  | P    | Q  | P    | Q  |
| Rice - Kg       | 60   | 20 | 75   | 18 | 90   | 15 |
| Eggs - Units    | 08   | 30 | 10   | 25 | 12   | 15 |
| Coconut oli - l | 75   | 02 | 100  | 01 | 150  | 01 |

- Compute separately for each product
  - (a) (i) Simple Price Relative
    - (ii) Simple Quantity Relative
    - (iii) Simple Value Relative for 2015 considering 2012 as the base year
  - (b) Point out that
    - Simple price relative
    - Simple quantity relative
    - Simple value relative

Computed for rice, satisfies the

- (i) Identical property
- (ii) Time reversal property
- (iii) Factor reversal property
- (iv) Cyclical property

**Solution :**

- (a) (i) **Simple Price Relatives**

- **For rice**

$$\frac{P_{2015}}{P_{2012}} \times 100$$

$$\frac{90}{75} \times 100 = 120\%$$

The price of rice has increased in 2015 by 20% relative to 2012.

- **For eggs**

$$\frac{P_{2015}}{P_{2012}} \times 100$$

$$\frac{12}{10} \times 100 = 120\%$$

The price of eggs has increased in 2015 by 20% relative to 2012.

- **For coconut oil**

$$\frac{P_{2015}}{P_{2012}} \times 100$$

$$\frac{150}{100} \times 100 = 150\%$$

The price of coconut oil in 2015 has increased by 50% relative to the year 2012.

## (ii) Simple Quantity Relatives

- **For rice**

$$\frac{q_{2015}}{q_{2012}} \times 100$$

$$\frac{15}{18} \times 100 = 83.3\%$$

The quantity of rice consumed in 2015 has decreased by 16.7% relative to the year 2012

- **For eggs**

$$\frac{q_{2015}}{q_{2012}} \times 100$$

$$\frac{15}{25} \times 100 = 60\%$$

Eggs consumption in 2015 has decreased by 40% relative to the year 2012.

- **For coconut oil**

$$\frac{q_{2015}}{q_{2012}} \times 100$$

$$\frac{01}{01} \times 100 = 100\%$$

No any change in consumption of coconut oil in 2015 relative to the year 2012.

(iii) **Simple Value Relative**

- **For rice**
$$\frac{p_{2015} q_{2015}}{p_{2012} q_{2012}} \times 100$$
$$\frac{90 \times 15}{75 \times 18} \times 100$$
$$100 \%$$

No any change in consumer expenses for rice in 2015 relative to the year 2012.

- **For eggs**
$$\frac{p_{2015} q_{2015}}{p_{2012} q_{2012}} \times 100$$
$$\frac{12 \times 15}{10 \times 25} \times 100$$
$$72 \%$$

The eggs consumption expenses have decreased in 2015 by 28% relative to the year 2012.

- **For coconut oil**
$$\frac{P_{2015} Q_{2015}}{P_{2012} Q_{2012}} \times 100$$
$$\frac{150 \times 1}{100 \times 1} \times 100$$
$$150 \%$$

The coconut oil consumption expenses in 2015 has increased by 50% relative to 2012.

(b) (i) **Identical Property ,**

Simple Price Relative that  $\frac{P_{2015}}{P_{2015}} = 1$  ,  $\frac{90}{90} = 1$

Simple Quantity Relative that  $\frac{q_{2015}}{q_{2015}} = 1$  ,  $\frac{15}{15} = 1$

Simple Value Relative that  $\frac{p_{2015} q_{2015}}{p_{2015} q_{2015}} = 1$   
 $\frac{90 \times 15}{90 \times 15} = 1$

Accordingly all simple relative indices can be said that the identical property is satisfied.

(ii) **Time Reversal Property**

- Simple Price Relative

$$\frac{P_{2015}}{P_{2012}} \times \frac{1}{\frac{P_{2015}}{P_{2012}}} = 1$$

$$\frac{90}{75} \times \frac{1}{\frac{90}{75}} = 1$$

- Simple Quantity Relative

$$\frac{Q_{2015}}{Q_{2012}} \times \frac{1}{\frac{Q_{2015}}{Q_{2012}}} = 1$$

$$\frac{15}{18} \times \frac{1}{\frac{15}{18}} = 1$$

- Simple Value Relative

$$\frac{P_{2015} q_{2015}}{P_{2012} q_{2012}} = \frac{1}{\frac{P_{2015} q_{2015}}{P_{2012} q_{2012}}}$$

$$\frac{90 \times 15}{75 \times 18} \times \frac{1}{\frac{90 \times 15}{75 \times 18}} = 1$$

Accordingly all simple relative indices can be said that the time reversal factor is satisfied.

- **Factor Reversal Property**

$$\frac{p_{2015} q_{2015} / p_{2012} q_{2012}}{p_{2015} / p_{2012}} = \frac{q_{2015}}{q_{2012}} = \frac{\frac{90 \times 15}{75 \times 18}}{\frac{90}{75}} = \frac{15}{18}$$

or

$$\frac{P_{2015} q_{2015} / P_{2012} q_{2012}}{q_{2015} / q_{2012}} = \frac{P_{2015}}{P_{2012}} = \frac{\frac{90 \times 15}{75 \times 18}}{\frac{15}{18}} = \frac{90}{75}$$

- Accordingly simple relative value index satisfies the factor reversal property.

- **Cyclical Property**

- Simple Price Relative

$$\frac{P_{2015}}{P_{2012}} \times \frac{P_{2012}}{P_{2010}} \times \frac{P_{2010}}{P_{2015}} = 1$$

$$\frac{90}{75} \times \frac{75}{60} \times \frac{60}{90} = 1$$

- Simple Quantity Relative

$$\frac{q_{2015}}{q_{2012}} \times \frac{q_{2012}}{q_{2010}} \times \frac{q_{2010}}{q_{2015}} = 1$$

$$\frac{15}{18} \times \frac{18}{20} \times \frac{20}{15} = 1$$

- Simple Value Relative

$$\frac{P_{2015}q_{2015}}{P_{2012}q_{2012}} \times \frac{P_{2012}q_{2012}}{P_{2010}q_{2010}} \times \frac{P_{2010}q_{2010}}{P_{2015}q_{2015}} = 1$$

$$\frac{90 \times 15}{75 \times 18} \times \frac{75 \times 18}{60 \times 20} \times \frac{60 \times 20}{90 \times 15} = 1$$

Accordingly all simple relative indices satisfy the cyclical property .

**A Guideline to explain the subject matters :**

- Indices used for measuring the changes of a particular variable in a considered period relative to a selected based period are known as Simple Relative Indices.
- Simple Relative Indices can be measured in three ways as;
  - (i) Simple Price Relative
  - (ii) Simple Quantity Relative
  - (iii) Simple Value Relative
- Expressing the price of a product in a given period as a ratio or percentage of the price of that product in a base period is known as Simple Price Relative.
- Simple Price Relative can be computed using the following formula.

$$P_{\%} = \frac{P_n}{P_o} \text{ or } \frac{P_n}{P_o} \times 100$$

P – price of the product

n – the considered period

o – the base period

- Expressing the quantity of a product in a considered period as a ratio or percentage of the quantity of that product in a base period is known as Simple Quantity Relative.
- Simple Quantity Relative can be computed using the following formula.

$$Q_{n/o} = \frac{q_n}{q_o} \text{ or } \frac{q_n}{q_o} \times 100$$

Q = quantity of unit of the product

n – considered period

o – base period

- Expressing the value of a product in a considered period as a ratio or percentage of the value of that product in the base period is known as simple Value Relative.
- Simple Value Relative can be computed using the following formula.

$$V_{n/o} = \frac{P_n q_n}{P_o q_o} \text{ or } \frac{P_n q_n}{P_o q_o} \times 100$$

Q = quantity of the product

n – considered period

o – base period

- Following properties are applied in Simple Relative Indices,

#### 1. Identical Property

That means the Simple Price Relative / Simple Quantity Relative / Simple Value Relative index of a particular period relative to the index of the same period is equal to 1 or 100%.

For a, b, c periods

$$\frac{P_a}{P_a} = 1$$

$$\frac{V_a}{V_a} = 1$$

$$\frac{q_a}{q_a} = 1$$

## 2. Time Reversal Property

That means once the index between two periods is multiplied by its reciprocal the result will be 1.

For a and b periods

$$\bullet \quad \frac{P_a}{P_b} \times \frac{P_b}{P_a} = 1 \quad \text{or} \quad \frac{P_a}{P_b} \times \frac{1}{\frac{P_a}{P_b}} = 1$$

$$\bullet \quad \frac{q_a}{q_b} \times \frac{q_b}{q_a} = 1 \quad \text{or} \quad \frac{q_a}{q_b} \times \frac{1}{\frac{q_a}{q_b}} = 1$$

$$\bullet \quad \frac{P_a q_a}{P_b q_b} \times \frac{P_b q_b}{P_a q_a} = 1 \quad \text{or} \quad \frac{P_a q_a}{P_b q_b} \times \frac{1}{\frac{P_a q_a}{P_b q_b}} = 1$$

## 3. Factor Reversal Property

- This means that the product of simple price relative and simple quantity relative is equal to the simple value relative.

$$\text{Simple Quantity Relative} = \frac{\text{Simple Value Relative}}{\text{Simple Price Relative}}$$

$$\text{Simple Price Relative} = \frac{\text{Simple Value Relative}}{\text{Simple Quantity Relative}}$$

for **a** and **b** periods that  $\frac{P_a}{P_b} \times \frac{q_a}{q_b} = \frac{P_a q_a}{P_b q_b}$  or  $\frac{P_a q_a / P_b q_b}{P_a P_b} = \frac{q_a}{q_b}$

- Cyclical Property

If **p** is the price **q** is the quantity and **pq** is the value of the three periods a, b and c that

$$\frac{P_a}{P_b} \times \frac{P_b}{P_c} \times \frac{P_c}{P_a} = 1$$

$$\frac{q_a}{q_b} \times \frac{q_b}{q_c} \times \frac{q_c}{q_a} = 1$$

$$\frac{P_a q_a}{P_b q_b} \times \frac{P_b q_b}{P_c q_c} \times \frac{P_c q_c}{P_a q_a} = 1$$

- Simple price relative indices can be applied for measuring the variation of a single variable in a considered period relative to a selected base period. These are very useful for a firm involved in producing only a single item or rendering only a single service. That is a simple comprehensive and easily calculable measure. But, it consists of following weak points.
- Inability of measuring the variations of price or quantity or value of few items together.  
In practically it would be better to compare the changes in price or quantity or value of many items together using an index, rather than to compare the price or quantity or value of a single item.

**Competency 11.0 :** Uses index numbers for making business decisions

**Competency Level 11.3 :** Uses a single index to measure the relative changes in few variables

**No. of Periods :** 10

**Learning Outcomes :**

- Introduces the simple aggregate indices
- Interprets simple aggregate price index, simple aggregate quantity index and simple aggregate value index.
- Computes the simple aggregate price index, simple aggregate quantity index and simple aggregate value index.
- Describes the uses and limitations of simple aggregate indices.
- Makes decisions using simple aggregate indices.
- Introduces the average indices of simple relatives.
- Classifies and interprets the simple average indices of relatives.
- Computes the average indices of simple price relatives, simple quantity relatives and simple value relatives.
- Describes the uses and limitations of the average indices of simple relatives.
- Makes decisions using average indices of simple relatives.

**Instructions for Lesson Planning :**

- Present the following data related to the price and quantity of few items consumed by a normal family in a week in May and June in 2016.

| Consume item | May             |               | June            |               |
|--------------|-----------------|---------------|-----------------|---------------|
|              | Unit price<br>p | Quantity<br>q | Unit price<br>p | Quantity<br>q |
| Rice (kg)    | 80              | 10            | 90              | 08            |
| Sugar (kg)   | 100             | 03            | 102             | 02            |
| Bread (400g) | 50              | 08            | 60              | 07            |
| Petrol (l)   | 150             | 20            | 120             | 25            |

- Explain to the students that the changes of the price in all the items should be represented in a single index. Inquire from the students about the steps to be followed for that purpose.

- Let them suggest a way to compute the changes in the price of the items in June relative to that of May.

$$\frac{\text{Sum of the prices spent for all items in June}}{\text{Sum of the prices spent for all items in May}} \times 100$$

$$= \frac{372}{380} \times 100$$

$$= 97.9\%$$

- The price of all the items considered has decreased by 2.1% in June compared to the prices in May.
- Explain the students that the way of changing the quantity demanded for three items as well as the way of changing their values also can be expressed in a single index in the same manner.

### Activity – I

- Use the following formula to derive the changes in quantity of goods in June relative to the month of May.

$$\frac{\text{Total number of units of goods bought in June}}{\text{Total number of units of goods bought in May}} \times 100$$

- (i) Use the following formula to derive the changes in the total value of the goods consumed in June relative to the total value of the goods consumed in May.

$$\frac{\text{value of the goods consumed in June}}{\text{value of the goods consumed in May}} \times 100$$

- Discuss the responses given by the students as follows.
- Make the students aware that the changes of price, quantity and value of goods and services between two periods can be computed as a whole with a single index in this manner.
- Involve the students in following activity using the same list of price and quantity of items given above.

### Activity 2 :

- Compute the following using the list of prices and quantity of items you have received.
  - Compute simple price relatives in June compared to May
  - Derive the sum of those price relatives
  - Devide the average relatives by number of items.
  - Get those values as a percentage and explain the changes of the price level.
- Involve in a discussion using the responses derived from students.

**Response :**

$$\begin{aligned} \text{(i) Simple price relative for rice} &= \frac{90}{80} = 1.125 \\ \text{Simple price relative for sugar} &= \frac{102}{100} = 1.02 \\ \text{Simple price relative for bread} &= \frac{60}{50} = 1.200 \\ \text{Simple price relative for petrol} &= \frac{120}{150} = 0.800 \end{aligned}$$

$$\text{(ii) } 1.125 + 1.020 + 1.200 + 0.800 = 4.145$$

$$\text{(iii) } \frac{4.145}{4} = 1.03625$$

$$\text{(iv) } 103.6\%$$

- The price of items has increased in June by 3.6% relative to that of in May.
  - Explain to the students that the index computed in this manner is independent from units.
  - Accordingly explain that a weak point in single aggregate indices that unit changes being ignored, is overcome by using these indices.
  - Explain that the indices computed in this manner are known as the average indices of simple relatives.
- Ascertain the fact that the average indices of simple relatives can be computed using quantity and value also as well as the price.

**Activity 3 :**

- The unit price of the goods consumed by a normal family consists of five members and the quantity of monthly consumption are given in the following table, Answer the questions mentioned below using those details.

| Item             | 2012         |            | 2016         |            |
|------------------|--------------|------------|--------------|------------|
|                  | Unit price p | Quantity q | Unit price p | Quantity q |
| Rice (kg)        | 80           | 20         | 110          | 15         |
| Sugar (kg)       | 75           | 07         | 90           | 05         |
| Eggs (units)     | 10           | 25         | 12           | 20         |
| Milk powder (kg) | 500          | 01         | 600          | 01         |
| Coconut oil (l)  | 125          | 03         | 200          | 02         |

**Compute :**

- (i) The average index of simple price relatives.
  - (ii) The average index of simple quantity relatives.
  - (iii) The average index of simple value relatives for 2016 relative to 2012
- Compute each index as percentage and explain the changes in price, quantity and value.

**Answers :**

- (i). Average index of simple price relatives

$$\begin{aligned}
 &= \frac{\frac{110}{80} + \frac{90}{75} + \frac{12}{10} + \frac{600}{500} + \frac{200}{125}}{5} \\
 &= \frac{1.375 + 1.2 + 1.2 + 1.2 + 1.6}{5} \\
 &= 1.315 \\
 &= \underline{\underline{131.5\%}}
 \end{aligned}$$

- Then price of the goods has increased by 31.5% in the year 2016 relative to the year 2012.

- (ii) The average index of simple quantity relatives.

$$\begin{aligned}
 &= \frac{\frac{15}{20} + \frac{5}{7} + \frac{20}{25} + \frac{1}{1} + \frac{2}{3}}{5} \\
 &= \frac{0.75 + 0.714 + 0.8 + 1 + 0.67}{5} = 0.79 \\
 &= \underline{\underline{79\%}}
 \end{aligned}$$

- The quantity of goods consumed in the year 2016 has decreased by 21% relative to the year 2012.

(iii) The average index of simple value relatives

$$\begin{aligned}
 &= \frac{110 \times 15}{80 \times 20} + \frac{90 \times 5}{75 \times 7} + \frac{12 \times 20}{10 \times 25} + \frac{600 \times 1}{500 \times 1} + \frac{200 \times 2}{125 \times 3} \\
 &= \frac{1650}{1600} + \frac{450}{525} + \frac{240}{250} + \frac{600}{500} + \frac{400}{375} \\
 &= 1.024 \\
 &= \underline{\underline{102.4\%}}
 \end{aligned}$$

- The value of these goods have increased in the year 2016 by 2.4% relative to the year 2012.

**A guideline to explain the subject matters :**

- The single indices computed to measure the changes of few factors in the considered period relative to the base period are known as simple aggregate indices.
- Simple aggregate indices can be measured in three types.
  - Simple aggregate price index
  - Simple aggregate quantity index
  - Simple aggregate value index
- **Simple aggregate price index**
  - The measure derived by expressing the changes in price of 2 or more goods and services in the considered period relative to a base period as a percentage, is known as the Simple Aggregate Price Index.
  - That index can be computed using the following formula.

$$SAPI = \frac{\sum p_n}{\sum p_o} \times 100$$

$\sum P_n$  - Sum of the prices of all the items in considered period

$\sum P_o$  - Sum of the prices of all the items in base period

- **Simple aggregate quantity index.**

The measure derived by expressing the changes in quantity of 2 or more goods and services in the considered period relative to a base period as a percentage is known as the Simple Aggregate Quantity Index.

That index can be computed using the following formula.

$$SAQI = \frac{\sum q_n}{\sum q_o} \times 100$$

$\sum q_n$  - Sum of the Quantities of all the goods and services in the considered period

$\sum q_o$  - Sum of the Quantities of all the goods and services in the base period

- **Simple aggregate value index.**

The measure derived by expressing the changes in value of 2 or more goods and services in the considered period relative to a base period as a percentage is known as the Simple Aggregate Value Index.

$$SAVI = \frac{\sum p_n q_n}{\sum p_o q_o} \times 100$$

$\sum p_n q_n$  - Sum of the values of all the goods and services in the considered period

$\sum p_o q_o$  - Sum of the values of all the goods and services in the base period

- Ability of measuring the change in price or quantity or value in general as a single index is the merit of simple aggregate indices.
- But following weak points are also found in simple aggregate indices.
  1. The relative importance of the goods and services being ignored by these indices.
 

Ex : Sugar and salt are assigned by these indices, the same importance in computing a cost of living index.
  - II. Ignorance of the measuring units of the goods.
 

Ex : Sugar in kilograms, Oil – in liters and Cloth in meters are measured. These units make an influence on the value of the index.
- Having computed simple relatives for each item separately in the considered period relative to a base period and when computed the average value of those relatives, that measure is known as the simple average indices of relatives.
- The average index of simple relatives can be measured in three types as follows.
  - (i) Average index of simple price relatives
  - (ii) Average index of simple quantity relatives
  - (iii) Average index of simple value relatives

- **Average index of simple price relatives**

Once simple price relatives are derived for each item in a set of  $N$  number of goods separately for a given period relative to a base period, when the sum of those price relatives is divided by  $N$ , the average index of simple price relatives is received. That value can be multiplied by 100 and expressed as a percentage. This can be computed using the following formula.

$$AISPR = \frac{\sum \left( \frac{P_n}{P_0} \right)}{N} \times 100$$

- **Average index of simple quantity relatives**

Once simple quantity relatives are derived for each item in a set of  $N$  number of goods separately for a given period relative to a base period, when the sum of those quantity relatives is divided by  $N$ , the average index of simple quantity relatives is received. That value can be multiplied by 100 and expressed as a percentage. This index can be computed using the following formula.

$$AISQR = \frac{\sum \left( \frac{q_n}{q_0} \right)}{N} \times 100$$

- **Average index of simple value relatives**

Once simple value relatives are derived for each item in a set of  $N$  number of goods separately for a given period relative to a base period and when the sum of those value relatives is divided by  $N$ , the average index of simple value relatives is received. This value can be multiplied by 100 and expressed as a percentage.

This index can be computed using the following formula.

$$AISVR = \frac{\sum \left( \frac{P_n q_n}{P_0 q_0} \right)}{N} \times 100$$

- Uses of the average indices of simple relatives are as follows.
  - Once there are few items of goods and services, ability of deriving a single index common for all those
  - Deriving a unit free measure, because the influence of units is removed by deriving the simple relative of each item separately.
- Any way the weak point that the ignorance of relative importance of the goods and services is still contained in these indices as well as in simple aggregate indices.

**Competency 11.0 :** Uses index numbers for making business decisions

**Competency Level 11.4:** Builds up weighted aggregate indices for few variables using weights

**No. of Periods :** 10

**Learning Outcomes :**

- Introduces the weighted aggregate indices
- Points out the uses of considering relative importance of the goods.
- Lists out the weighted aggregate indices.
- Interprets the Laspear price index.
- Computes the Laspear price index and quantity index using the relevant formulae.
- States the properties of the Laspear index.
- Interprets Pasche's price index.
- Computes the Pasche's price index and quantity index using the relevant formulae.
- States the properties of the Pasche index.
- Interpret Marshal Edgeworth index.
- Computes the Marshal Edgeworth price and quantity indices using the relevant formulae.
- Lists out the properties of Marshal Edgeworth index.
- Interprets Fisher ideal price index and quantity index using the relevant formulae.
- Computes the Fisher ideal price index and quantity index using the relevant formulae.
- Lists out the properties of Fisher ideal index.
- Introduces the typical period index.
- Computes the typical period price index and quantity index using the relevant formulae.
- Lists out the properties of typical period indices.

**Instructions for Lesson Planning :**

- Pay attention of students to the following situation.

The price and quantity of few consumer items required by a normal family for consumption for a week related to the prices in two years are as follows.

| Consumer items | 2010      |          | 2016      |          |
|----------------|-----------|----------|-----------|----------|
|                | Price Rs. | Quantity | Price Rs. | Quantity |
| Salt (kg)      | 30        | 500g     | 50        | 500g     |
| Rice (kg)      | 80        | 12 kg    | 90        | 11 kg    |
| Sugar (kg)     | 75        | 03 kg    | 110       | 03 kg    |

- The change in the price of above items for the period of 2016 relative to 2010 should be derived as a percentage.
- Guide the students to build up a formula discussing the following matters.

$$\begin{aligned}
 &= \frac{\text{sum of the price in 2016}}{\text{sum of the price in 2010}} \times 100 \\
 &= \frac{250}{185} \times 100 \\
 &= \underline{\underline{135.14\%}}
 \end{aligned}$$

- The price of the above items has increased in 2016 by 35.14% relative to the year 2010.
  - Explain that the weekly consumption of salt is different from the consumption of rice and sugar and ascertain that the relative importance of each item is different.
  - Ascertain that the quantity of goods consumed also should be considered in order to construct the price index.
  - Make them aware that the price indices can be computed as follows.
- Considering not only the price of the goods but also the quantity of goods.

$$(i) \frac{\sum p_{2016} q_{2016}}{\sum p_{2010} q_{2016}} \times 100$$

$$(ii) \frac{\sum p_{2016} q_{2010}}{\sum p_{2010} q_{2010}} \times 100$$

- Lead the students to compute the price indices using the above formulae.

$$\begin{aligned}
 \text{i)} \quad & \frac{\sum p_{2016} q_{2016}}{\sum p_{2010} q_{2016}} \times 100 \\
 & = \frac{(50 \times 0.5) + (90 \times 11) + (110 \times 3)}{(30 \times 0.5) + (80 \times 11) + (75 \times 3)} \times 100 \\
 & = \frac{1345}{1120} \times 100 \\
 & = \underline{\underline{120.09\%}}
 \end{aligned}$$

- Explain that in this calculation, the changes in the price of few items in 2016 have been computed relative to 2010 as the base year and further that those items have been weighed on the quantity purchased from those goods in 2016.
- Explain that the price of goods has increased by 20.09% in accordance with that calculation.

$$\begin{aligned}
 \text{(ii)} \quad & = \frac{\sum p_{2016} q_{2010}}{\sum p_{2010} q_{2010}} \times 100 \\
 & = \frac{(50 \times 0.5) + (90 \times 12) + (110 \times 3)}{(30 \times 0.5) + (80 \times 12) + (75 \times 3)} \times 100 \\
 & = \frac{1435}{1200} \times 100 \\
 & = \underline{\underline{119.58\%}}
 \end{aligned}$$

- Explain that in this calculation, the changes in the price of few items in 2016 have been computed relative to 2010 as the base year and further that those items have been weighed on the quantities purchased from those goods in 2010.
- Explain that the price of goods has increased by 19.58% in accordance with that calculation.
- Explain that the indices computed in that manner are known as weighted aggregate indices.
- Explain that the weighted aggregate indices are in two types as price indices and quantity indices.
- Explain that when computed price indices the items are weighted on quantities where as in quantity indices the items are weighed on prices.

#### Activity 01 :

- Provide the data set with the following description to the students and involve them in activity.

- Price and quantity of few items consumed monthly by a normal family are mentioned as follows.

| Consumer items  | 2012      |          | 2016      |          |
|-----------------|-----------|----------|-----------|----------|
|                 | Price Rs. | Quantity | Price Rs. | Quantity |
| Rice (kg)       | 70        | 40       | 85        | 35       |
| Sugar (kg)      | 80        | 08       | 100       | 06       |
| Chilies (kg)    | 120       | 01       | 150       | 01       |
| Flour (kg)      | 80        | 03       | 110       | 02       |
| Coconut oil (l) | 150       | 03       | 200       | 02       |

- In order to show the changes in price in 2016 relative to 2012
- Compute the price indices.
  - (a) Weighing on the quantities of 2012 using the formula.

$$\frac{\sum p_{2016} q_{2012}}{\sum p_{2012} q_{2012}} \times 100$$

- (b) Weighing on the quantities of 2016 using the formula.

$$\frac{\sum p_{2016} q_{2016}}{\sum p_{2012} q_{2016}} \times 100$$

and interpret them.

- In order to show the changes in quantity of goods bought in 2016 relative to 2012.
- Compute the quantity indices.
  - (a) Weighing on the price of 2012 using the formula.

$$\frac{\sum q_{2016} p_{2012}}{\sum q_{2012} p_{2012}} \times 100$$

- (b) Weighing on the price of 2016 using the formula

$$\frac{\sum q_{2016} p_{2016}}{\sum q_{2012} p_{2016}} \times 100$$

and interpret them.

**Answers :**

(i) Price Indices

(a) Weighing on the quantities of 2012

$$\begin{aligned} &= \frac{\sum p_{2016} q_{2012}}{\sum p_{2012} q_{2012}} \times 100 \\ &= \frac{(85 \times 40) + (100 \times 8) + (150 \times 1) + (110 \times 3) + (200 \times 3)}{(70 \times 40) + (80 \times 8) + (120 \times 1) + (80 \times 3) + (150 \times 3)} \times 100 \\ &= \left[ \frac{3400 + 800 + 150 + 330 + 600}{2800 + 640 + 120 + 240 + 450} \right] \times 100 \\ &= \frac{5280}{4250} \times 100 \\ &= \underline{\underline{124.24\%}} \end{aligned}$$

The prices of goods have increased in 2016 by 24.24% relative to 2012

(b) Weighing on the quantities of 2016

$$\begin{aligned} &= \frac{\sum p_{2016} q_{2016}}{\sum p_{2012} q_{2016}} \times 100 \\ &= \frac{(85 \times 35) + (100 \times 6) + (150 \times 1) + (110 \times 2) + (200 \times 2)}{(70 \times 35) + (80 \times 6) + (120 \times 1) + (80 \times 2) + (150 \times 2)} \times 100 \\ &= \left[ \frac{2975 + 600 + 150 + 220 + 400}{2450 + 480 + 120 + 160 + 300} \right] \times 100 \\ &= \frac{4345}{3510} \times 100 \\ &= \underline{\underline{123.8\%}} \end{aligned}$$

The prices of goods have increased in 2016 by 23.8% relative to 2012.

(ii) Quantity Indices

• Weighing on the prices of 2012  $\frac{\sum q_{2016} p_{2012}}{\sum q_{2012} p_{2012}} \times 100$

$$\begin{aligned} &= \frac{(70 \times 35) + (80 \times 6) + (120 \times 1) + (80 \times 2) + (150 \times 2)}{(70 \times 40) + (80 \times 8) + (120 \times 1) + (80 \times 3) + (150 \times 3)} \times 100 \\ &= \frac{3510}{4250} \times 100 \\ &= \underline{\underline{82.59\%}} \end{aligned}$$

- The quantity of goods bought in 2016 has decreased by 17.41% relative to 2012.
- Weighing on the prices of 2016.

$$\begin{aligned} & \frac{\sum q_{2016} p_{2016}}{\sum q_{2012} p_{2016}} \times 100 \\ &= \frac{(85 \times 35) + (100 \times 6) + (150 \times 1) + (110 \times 2) + (200 \times 2)}{(85 \times 40) + (100 \times 8) + (150 \times 1) + (110 \times 3) + (200 \times 3)} \times 100 \\ &= \left[ \frac{2975 + 600 + 150 + 220 + 400}{3400 + 800 + 150 + 330 + 600} \right] \times 100 \\ &= \frac{4345}{5280} \times 100 \\ &= \underline{\underline{82.29\%}} \end{aligned}$$

- The quantity of goods bought in 2016 has decreased by 17.71% relative to 2012
- According to the above calculations explain that.
  - Weighing on the base year is Lasper's method and
  - Weighing on the considered year is Pasche's method.

**Activity 2 :**

- Involve the students in the following activity based on the answers of the above activity 1
1. Derive the square root of the product of Laspear price index and Pasche price index

$$\sqrt{\text{Laspear Price Index} \times \text{Pasche price Index}}$$

2. Derive the square root of the product of Laspear quantity index and Pasche quantity index

$$\sqrt{\text{Laspear Quantity Index} \times \text{Pasche Quantity Index}}$$

**Answers :**

1.  $\sqrt{124.24 \times 123.8}$

$$\underline{\underline{124.02}}$$

- The price of the goods has increased by 24.02% in 2016 relative to the year 2012.

2.  $\sqrt{82.59 \times 82.29}$

$$= \underline{\underline{82.44}}$$

- The quantity of goods bought has decreased by 17.56% in 2016 relative to the year 2012.

Explain that computing the geometric mean of Laspears indices and Pasche indices the Fisher ideal indices can be derived.

**Activity 03 :**

- Involve the students in the following activity for computing the indices using the weights depend on both the base year and the considered year.
- Using the table of data related to the activity 1 above and derive the following.

(1)  $q_0 + q_n$

(2)  $P_n (q_0 + q_n)$

(3)  $P_o (q_0 + q_n)$

(4)  $P_o + P_n$

(5)  $q_n (P_o + P_n)$

(6)  $q_o (P_o + P_n)$

(7)  $\frac{\sum P_n (q_o + q_n)}{\sum P_o (q_o + q_n)} \times 100$

(8)  $\frac{\sum q_n (P_o + P_n)}{\sum q_o (P_o + P_n)} \times 100$

- (9) Interpret the values computed for above formulae in (7) and (8).

**Answer :**

පිළිතුර :

| Item        | 2012  |       | 2016  |       |             |                   |                   |             |                   |                   |
|-------------|-------|-------|-------|-------|-------------|-------------------|-------------------|-------------|-------------------|-------------------|
|             | $P_o$ | $q_o$ | $P_n$ | $q_n$ | $q_o + q_n$ | $P_n (q_o + q_n)$ | $P_o (q_o + q_n)$ | $P_o + P_n$ | $q_n (P_o + P_n)$ | $q_o (P_o + P_n)$ |
| Rice (kg)   | 70    | 40    | 85    | 35    | 75          | 6375              | 5250              | 155         | 5425              | 6200              |
| Sugar (kg)  | 80    | 08    | 100   | 06    | 14          | 1400              | 1120              | 180         | 1080              | 1440              |
| Chilli (kg) | 120   | 01    | 150   | 01    | 02          | 300               | 240               | 270         | 270               | 270               |
| Flour (kg)  | 80    | 03    | 110   | 02    | 05          | 550               | 400               | 190         | 380               | 570               |
| Oil         | 150   | 03    | 200   | 02    | 05          | 1000              | 750               | 350         | 700               | 1050              |
|             |       |       |       |       |             | 9625              | 7760              |             | 7855              | 9530              |

$$\frac{\sum P_n (q_o + q_n)}{\sum P_o (q_o + q_n)} \times 100$$

$$= \frac{9625}{7760} \times 100$$

$$= \underline{\underline{124.03}}$$

- The prices of the goods have increased by 24.03% in 2016 relative to the year 2012.

$$\frac{\sum q_n (P_o + P_n)}{\sum q_o (P_o + P_n)} \times 100$$

$$= \frac{7855}{9530} \times 100$$

$$= \underline{\underline{82.42}}$$

- The quantity of goods bought in 2016 has decreased by 17.58% relative to the year 2012.
- Explain that computing the indices weighing on the average of base year price/ quantity and current year price/ quantity is known as the Marshel Edgeworth method.

**Activity 04 :**

Involve the students in the following activity to make them aware that the indices can be computed weighing on the price or quantity of another period rather than weighing on base period or considered period.

Provide with the following table containing the price and quantity of few items consumed by a normal family in a week related to 3 different years and get the questions answered given below.

| Consumer items | 2010           |                | 2012           |                | 2016           |                |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                | p <sub>o</sub> | q <sub>o</sub> | P <sub>t</sub> | q <sub>t</sub> | p <sub>n</sub> | q <sub>n</sub> |
| Bread          | 30             | 10             | 45             | 08             | 50             | 07             |
| Eggs           | 06             | 15             | 08             | 10             | 12             | 08             |
| Rice           | 70             | 10             | 85             | 08             | 90             | 07             |
| Sugar          | 60             | 03             | 80             | 02             | 100            | 02             |

Consider that the

- Base year as 0 (2010)
- Considered year as n (2016)
- Year used for weighing as t (2012)
- Derive the price indices and quantity indices using following formulae.

1. Price index

2. quantity Index

$$\frac{\sum p_n q_t}{\sum p_o q_t} \times 100$$

$$\frac{\sum q_n p_t}{\sum q_o p_t} \times 100$$

**Answers : Price Index**

$$\begin{aligned} \frac{\sum p_n q_t}{\sum p_o q_t} \times 100 &= \frac{(50 \times 8) + (12 \times 10) + (90 \times 8) + (100 \times 2)}{(30 \times 8) + (6 \times 10) + (70 \times 8) + (60 \times 2)} \times 100 \\ &= \left[ \frac{400 + 120 + 720 + 200}{240 + 60 + 560 + 120} \right] \times 100 \\ &= \frac{1440}{980} \times 100 \\ &= \underline{\underline{146.94\%}} \end{aligned}$$

- The prices of goods have increased in 2016 by 46.94% relative to the year 2012

- Quantity Index

$$\begin{aligned} \frac{\sum q_n p_t}{\sum q_o p_t} \times 100 &= \frac{(7 \times 45) + (8 \times 8) + (7 \times 85) + (2 \times 80)}{(10 \times 45) + (15 \times 8) + (10 \times 85) + (3 \times 80)} \times 100 \\ &= \left[ \frac{315 + 64 + 595 + 160}{450 + 120 + 850 + 240} \right] \times 100 \\ &= \frac{1134}{1660} \times 100 \\ &= \underline{\underline{68.31\%}} \end{aligned}$$

- The quantity of goods bought in 2016 has decreased by 31.69% relative to the year 2012.
- Ascertain that the indices computed in this manner using the price / quantity of another period as weights, without weighing on base period or considered period are called **Typical Period Indices**.

**A guideline to explain the subject matters :**

- Two disadvantages in simple aggregate indices were mentioned above. One of them was abolished through the average indices of simple relatives. It was the influence of disregarding the difference measuring units used for different items, being contained in the indices.
- As a treatment for the other weak point that disregarding the relative importance of different items, the weighted aggregate indices have been introduced.
- Weighted aggregate indices means that,
  - the indices computed using an appropriate weighing considering the relative importance of each item.
- Weighted aggregate indices are in two types as price indices and quantity indices.
- Weighted Agregate Price Index

$$P_{W_{n/o}} = \frac{\sum p_n w}{\sum p_o w} \times 100$$

$p_n$  = price of the goods in considered period

$p_o$  = price of the goods in base period

$w$  = assigned weight

- Weighted Agregate quantity Index

$$qw_{n/o} = \frac{\sum q_n w}{\sum q_o w} \times 100$$

$q_n$  - quantity of goods in considered period

$q_o$  - quantity of goods in base period

$w$  - assigned weight

- The use of computing weighted aggregate indices is the ability of deriving indices considering the relative importance of different goods and services.
- Most leading weighted aggregate indices are
  - (i) Laspear indices
  - (ii) Pasche indices
  - (iii) Marshel Edgeworth indices
  - (iv) Fisher ideal indices
  - (v) Typical ideal indices

- **Laspeare Indices**

These are computed weighing the price / quantity of considered period relative to base period on the prices/ quantity of the same items in the base period.

- Laspear Price Index  $Lp_{n/o} = \frac{\sum p_n q_o}{\sum p_o q_o} \times 100$

$P_n$  = price in the considered period

$P_o$  = price in the base period

$q_o$  = quantity in the base period

- Laspear Quantity Index  $Lq_{n/o} = \frac{\sum q_n p_o}{\sum q_o p_o} \times 100$

$q_n$  = Quantity of goods in the considered period

$q_o$  = Quantity of goods in the base period

$p_o$  = Price of goods in the base period

- Properties of Laspear indices

- Ability of computing under a lower cost since they are computed weighing on base period (cost of survey is minimum)
- Convenience in comparisons using the indices, because the base (denominator) is a constant.

- **Pasche Indices**

These are computed weighing on the price/quantity of the goods in considered period relative to based period, on the prices/quantities of the considered period.

Pasche price index is computed as follows.

$$\frac{\sum p_n q_n}{\sum p_0 q_n} \times 100$$

$p_n$  = price in the considered period

$q_n$  = quantity in the considered period

$p_0$  = price in the base period

- Pasche quantity index

$$Pq_{n/o} = \frac{\sum q_n p_n}{\sum q_0 p_n} \times 100$$

$p_n$  = price in the considered period

$q_n$  = quantity of the goods in the considered period

$q_0$  = quantity in the base period

- Properties of Pasche indices.

- Representing the reality through the index since up-to-date data are used.

- **Marshel Edgeworth Indices**

The arithmetic mean of the price / quantity of base period and considered period is applied for Marshel Edgeworth indices in accordance with the typical period method.

Laspear indices are undergone to upward bias (over estimation) due to use of base year weighing, where as the Pasche indices are undergone to downward bias (under estimation) due to use of current year weighing. In order to overcome these weakpoints, Marshel Edgeworth index and Fisher ideal index have been developrd.

- Marshal Edgeworth Price Index

$$MEp_{n/o} = \frac{\sum p_n \frac{1}{2}(q_o + q_n)}{\sum p_o \frac{1}{2}(q_o + q_n)} \times 100$$

$$= \frac{\sum p_n (q_o + q_n)}{\sum p_o (q_o + q_n)} \times 100$$

- Marshal Edgeworth Qaninity Index

$$MEq_{n/o} = \frac{\sum q_n \frac{1}{2}(p_o + p_n)}{\sum q_o \frac{1}{2}(p_o + p_n)} \times 100$$

$$= \frac{\sum q_n (p_o + p_n)}{\sum q_o (p_o + p_n)} \times 100$$

- Properties of Marshal Edgeworth Indices

- Marshal Edgeworth indices satisfy the time reversal test.
- Since both the base period weights and current period weights are used, the index represents the situation in both the periods.
- But Marshal Edgeworth Indices do not satisfy the factor reversal test.
- It is expensive since new data should be collected.

- Fisher Ideal Index

- Fisher ideal index is derived through the geometric mean of the Laspear index and the Pasche index.
- Fisher Price Index

$$FP_{n/o} = \sqrt{\left(\frac{\sum p_n q_o}{\sum p_o q_o}\right) 100 \times \left(\frac{\sum p_n q_n}{\sum p_o q_n}\right) 100}$$

- Fisher Quantity Index

$$FQ_{n/o} = \sqrt{\left(\frac{\sum q_n p_o}{\sum q_o p_o}\right) 100 \times \left(\frac{\sum q_n p_n}{\sum q_o p_n}\right) 100}$$

- Properties of Fisher ideal indices are
  - Being free from both upward bias and downward bias.
  - Satisfying both the time reversal test and the factor reversal test.

- **Typical Period Index**

The indices computed weighing on price or quantity of another period rather than the base period or considered period or weighing on the average price or average quantity of few periods are known as typical period indices.

- Typical period price index

$$TP_{n/o} = \left( \frac{\sum p_n q_t}{\sum p_o q_t} \right) \times 100$$

$p_n$  = price of goods in considered period

$p_o$  = price of goods in base period

$q_t$  = quantity of the goods in typical period

- Typical period quality index

$$Tq_{n/o} = \left( \frac{\sum q_n p_t}{\sum q_o p_t} \right) \times 100$$

$q_n$  = quantity of goods in considered period

$q_o$  = quantity of goods in base period

$p_t$  = price of the goods in typical period

**Competency 11.0** : Uses index numbers for making business decisions

**Competency Level 11.5:** Computes an appropriate consumer price index to measure the cost of living

**No. of Periods** : 04

**Learning outcomes :**

- Introduces the ‘Consumer Price Index’.
- Points out the need of a consumer price index.
- Describes the factors that should be considered in constructing a consumer price index.
- Lists out the steps of constructing a consumer price index.
- Computes a c consumer price index using the given details.
- Describes the uses of a consumer price index.
- Points out the practical problems arisen in constructing a consumer price index.

**Instructions for Lesson Planning :**

- Inquire about the price indices known by students.
- Note down the students responses on the board.
- Lead them to name the consumer price indices out of the indices noted down on the board.
- If consumer price indices are not found among the indices mentioned by the students, note down the following consumer price indices on the board.
  - Colombo Consumer Price Index (CCPI)
  - Sri Lanka Consumer Price Index (SLCI)
  - Greater Colombo Consumer Price Index (GCCPI)
- Inquire from students about the institutional bodies responsible for preparation of those consumer price indices.
- Inquire from students about the categories of consumer items included in each of those indices.
- Point out the students that all the consumer items have been categorized under few headings such that food & beverage, clothing & ornaments, rent, fuel and miscellaneous in Colombo consumer price index.
- Discuss with students about the items included in Sri Lanka consumer price index.
- Discuss with the students about the items included in greater Colombo consumer price index.
- Hold a discussion with the students highlighting the following facts.

- The price value paid for acquisition of goods and services for self satisfaction of a particular family unit is called the final consumption expenditure.
- The consumer price indices are prepared in order to measure the changes in price of goods and services which are used for final consumption.
- When a consumer price index is constructed the consumer items should be weighted in accordance with their relative importance.
- It's necessary to follow various steps in preparing a consumer price index.
- Consumer price indices are applicable in various uses.
- Lead the students step by step in the following activity to explain how to construct a consumer price index.

**Activity 1 :**

- inform that a consumer price index is needed to be prepared for school pupils.
- Inform that the group of community from whom the data to be collected for that purpose is the sample of pupils in this class itself.
- Prepare a list of school items consumed by the pupils on the board discussing with them.
- Separate them in to three categories under books, pens and pencils and miscellaneous.
- Inquire the price of each item from the pupils and write against each item on the board.
- Derive three (3) means of the prices by calaulating the mean peices of each category of items seperately.
- Mention an assumed base year and write the assumed price of each item in that base year on the board.
- Determine the weight of three categories of items as percentage values getting points of view of the students.
- Lead the students to compute the following table using the assumed weight and other details derived in discussion.

| Item           | Price in base year $P_o$ | Price in current year $P_n$ | Simple Price relative $\frac{P_n}{P_o}$ | Weight (W) | Simple Price relaiive X weight |
|----------------|--------------------------|-----------------------------|---|------------|--------------------------------|
| Books          |                          |                             |   |            |                                |
| Pens & Pencils |                          |                             |   |            |                                |
| Miscellaneous  |                          |                             |   |            |                                |

- Guide the students to find the sum of the column **Price Relative X weight** and to deivide that value from the total value of the weights and multiply by 100.
- State that the value derived can be introduced as the consumer price index of the school items consumed by students in the school.

- Assign the students to produce the consumer price index as a formula.
- Let them compare the formula that they produced with the following formula.

$$\text{Consumer Price Index} = \frac{\sum_{i=1}^n [(p_n / p_o) W_i]}{\sum_{i=1}^n W_i}$$

- Provide with the following activity to the students to compute the consumers price index using the given details.

### Activity 2 :

Given below is a set of data collected for calculation of a consumer price index.

| Consumr items         | Average price in the year 2005 | Average price in the year 2016 | Weight |
|-----------------------|--------------------------------|--------------------------------|--------|
| Food and beverage     | 300                            | 500                            | 4      |
| Clothing and garments | 600                            | 1200                           | 2      |
| Fuel & electricity    | 500                            | 800                            | 1      |
| House rent            | 600                            | 600                            | 1      |
| Miscellaneous         | 200                            | 400                            | 2      |

- Compute the consumer price index in 2016 relative to the year 2005
- Interpret your answer.

### Solution Activity - 2

| Item                  | Price Relative<br>( $p_n / p_o$ ) | Price Relative X Weight<br>( $(p_n / p_o)w$ ) |
|-----------------------|-----------------------------------|---|
| Food and beverage     | $(500 / 300) = 1.67$              | $1.67 \times 4 = 6.68$                        |
| Clothing and garments | $(1200 / 600) = 2.00$             | $2.0 \times 2 = 4.00$                         |
| Fuel & electricity    | $(800 / 500) = 1.60$              | $1.60 \times 1 = 1.60$                        |
| House rent            | $(600 / 600) = 1.00$              | $1.00 \times 1 = 1.00$                        |
| Miscellaneous         | $(400 / 200) = 2.00$              | $2.00 \times 2 = 4.00$                        |
| Total                 |                                   | 17.28   |

$$\begin{aligned} \text{consumer price index} &= \frac{\sum (p_n / p_o \times w)}{\sum w} \times 100 = \frac{17.28}{10} \times 100 \\ &= \underline{\underline{172.8}} \end{aligned}$$

The consumer expenditure has increased by 72.8% in 2016 relative to the year 2005.

**A guideline to explain the subject matters :**

- A measure which is used to evaluate the change in the price of a set of consumer items as a percentage relative to a particular period in time is known as a **Consumer Price Index**.
- The consumer price indices are constructed based on the prices paid for acquisition of goods and services directly for the survival of a family unit.
- Measuring the percentage of changes in price, relative to a particular community based on a package of goods with sufficient quantities is occurred through a consumer price index.
- Accordingly, the measure which is used to evaluate the change in prices of a set of consumer goods and services relative to a period of time or a geographical location can be simply introduced as a consumer price index.
- If it is mentioned in other words the index number which is used to measure the changes experienced in the price paid for the goods and services consumed by the consumer.
- Some situations where consumer price indices are usefully applied are given below.
  - To measure the living conditions of a community.
  - To make salary adjustments of the employees to tally with the price level.
  - To evaluate the purchasing power of monetary unit.
  - To prepare salary policies in a firm / in an economy.
- The factors that should be taken in to consideration in construction of a consumer price index are as follows.
  1. Vividly identification of objectives of the index.

The fact whether the index is to measure the changes in price of which consumer items of what sort of community.
  2. Selecting a community

The fact, what's the community from whom the data should be collected.
  3. Selecting a package of items.

Since all the items consumed by the selected community cannot be included to the index a sample of items consumed by majority of the people in the selected group should be chosen.
  4. selecting a suitable base year

Selecting an year as the base year which is more appropriate to compare the prices, is needed. Which year is more suitable for that purpose should be considered. Unless an economically stable year which is free from unusual forces is selected, the comparison will be distorted.

## 5. Determining the weights

Weighing is used in order to represent the relative importance of each item in the index weighing can be performed considering the percentages of consumer expenses through a house hold survey.

## 6. Determining the method of computing the indices.

Various formulae can be used for computing the index. Every consumer price index in Sri Lanka has been computed following the Lapear's method among various techniques such as, Laspear, Pasche and Fisher etc.. considering the convenience in collecting details and computing the index.

Steps of preparing a consumer price index are as follows.

### Step 1 :

Several categories of consumer items are included in a consumer price index

The simple relative of each of those items should be separately computed. i.e. . The simple price relative of food and beverage. The simple price relative of clothing & ornaments etc. ( $P_n/P_o$ )

### Step : 2

Weighing for the simple price relatives of each consumer item.

An appropriate weight should be assigned for each consumer item, based on the percentage of expenses derived on the data collected through a house hold survey or following any other appropriate way.

### Step 3 :

Multiplication of each price relative by the corresponding weights.

$$\left( \frac{P_n}{P_o} \times w \right)$$

### Step 4 :

Dividing the sum of the products of the 3<sup>rd</sup> step above from the sum of the weights and expressing as a percentage.

$$\left( \frac{\sum \frac{P_n}{P_o} . w}{\sum w} \times 100 \right)$$

Few limitations of the index numbers are as follows.

1. The index numbers are only approximate calculations, but not complete values with no any shortcomings.
2. Non sampling errors may occur in following each step of computing the index numbers (possible to be biased)
3. Perhaps the sample may not be a good representation of the population (the package of items or the group of community selected)
4. The accuracy and reliability may be problematic.
5. Computing of index numbers being a very complicated process.

**Competency 11.0** : Uses index numbers for making business decisions

**Competency Level 11.6:** Studies the applications of index numbers

**No. of Periods** : 04

**Learning outcomes :**

- Explains the need of shifting the base year of an index number and re-calculating.
- Shifts the base year of a given series of index numbers and re-calculates.
- Converts the monetary values given in to real values using index numbers.
- Makes decisions using real values.
- Lists out the indices used in Sri Lanka
- Separately describes about each index.

**Instructions for Lesson Planning :**

- Present the following two series of index numbers to the students and hold a discussion inquiring about the facts given below.

Price index of paddy, year - 2000 = 100

| 2010 | 2011 | 2012 | 2013 | 2014 |
|------|------|------|------|------|
| 180  | 200  | 225  | 250  | 300  |

Price index of rice, year - 2005 = 100

| 2010 | 2011 | 2012 | 2013 | 2014 |
|------|------|------|------|------|
| 110  | 130  | 140  | 160  | 200  |

1. When compared the above series of indices do you mean that the price index of rice takes lower values then the price index of paddy?
  2. When considered all the five years, there is a greater increment in price shown in the price of paddy ? in the price of rice?
  3. Compare the growth in the price of paddy in 2014 relative to 2010 and the growth in the price of rice in 2014 relative to 2010.
- Hold a discussion highlighting the following facts based on the responses of students for the above questions 1, 2 and 3.
  - Point out that these comparisons are impossible since the base year of price index of rice is 2005.

- Point out if the base year of both these indices is considered as 2010 the value of both indices in 2010 can be considered as 100.
- With respect to that the value of these indices in other years also can be derived in accordance with the new base year as follows.

The value of the index according to the new base year =  $\frac{\text{Former index of the considered year}}{\text{Former index of the new base year}} \times 100$

- Guide the students to re-compute the values of the price index of paddy and the price index of rice for the period 2010-2014 considering 2010 as the base year and to present the two series of index numbers again.
- Lead the students to answer the above questions 1, 2, 3 comparing the two series of indices derived considering 2010 as the base year.
- Re-calculating the price index of paddy considering 2010 as the base year.

$$2010 = \frac{180}{180} \times 100 = 100$$

$$2011 = \frac{200}{180} \times 100 = 111.11$$

$$2012 = \frac{225}{180} \times 100 = 125$$

$$2013 = \frac{250}{180} \times 100 = 138.89$$

$$2014 = \frac{300}{180} \times 100 = 166.67$$

- Re-calculating the price index of rice considering the year 2010 as the base year.

$$2010 = \frac{110}{110} \times 100 = 100$$

$$2011 = \frac{130}{110} \times 100 = 118.18$$

$$2012 = \frac{140}{110} \times 100 = 127.27$$

$$2013 = \frac{160}{110} \times 100 = 145.45$$

$$2014 = \frac{200}{110} \times 100 = 181.82$$

**Solutions :**

1. When considered both indices on the same base year almost all the values in price index of rice are higher than the corresponding values in the price index of paddy.
2. The increments in price of each index relative to the year 2010 are as follows.

| Year                     | 2011   | 2012   | 2013   | 2014   |
|--------------------------|--------|--------|--------|--------|
| Price index of paddy (%) | +11.11 | +25    | +38.89 | +66.67 |
| Price index of rice %    | +18.18 | +27.27 | +45.45 | +81.82 |

Accordingly a greater increment in every year is reported in the price of rice.

- Involve the students in the following activity to re-calculate the value of a given series of indices shifting the base year.

### Activity – 1

The indices give below have been computed using 2005 as the base year. Re-calculate these indices using 2010 as the base year.

| Year         | 2005 | 2009 | 2010 | 2013 | 2016 |
|--------------|------|------|------|------|------|
| Index number | 100  | 220  | 250  | 300  | 550  |

### Solution for activity – 1

Values of the index number considering the base year as 2010.

$$2005 = \frac{100}{250} \times 100 = 40$$

$$2009 = \frac{220}{250} \times 100 = 88$$

$$2010 = \frac{250}{250} \times 100 = 100$$

$$2013 = \frac{300}{250} \times 100 = 120$$

$$2016 = \frac{550}{250} \times 100 = 220$$

- Produce the following problem to the students in order to explain how to convert the monetary values into real values.
  - Suppose that the price of a loaf of bread in the year 2000 was Rs. 25/- and that of 2015 was Rs. 50/-
  - How many loaves of bread could be bought by a labourer who received a daily wages of Rs. 100/- in the year 2000 with that amount?
  - If the daily wage of a labourer in 2015 was Rs. 200/- how many loaves of bread could he buy with that amount?
  - If the daily wage of that labourer in the year 2000 is stated in number of loaves of bread what should be the value?

- If the daily wage of that labourer in the year 2015 is stated in number of loaves of bread what should be the value?

Hold a discussion highlighting the following facts.

- The daily wage of the labourer has increased by 100% when considered in terms of rupees (monitory value)
- But the number of loaves of bread he could buy has not been changed.
- The number of goods and services could be bought with the monetary units is known as the real value of that monetary unit.
- Accordingly the real wage of the labourer in 2015 is also Rs. 100/= relative to the year 2000.
- Provide the following instructions to the students.
  - Compute the simple price relative of a loaf of bread in 2015 relative to the year 2000.
  - Divide the daily wage received by the labourer in 2015 by the simple price relative of bread.
  - Interpret the result you derived.

Highlight the following facts.

- that the changes in the prices of goods and services are important in order to compute the real values relevant to the monetary values.
- that the real values of monetary values can be computed using price indices.
- that the real values could be derived by dividing the monetary values from the price index (deflating)
- that computing the real values can be utilized in decision making in the business field as well as in real life situations.

**\A guideline to explain the subject matters :**

- Re-calculating the value of an index number using another year as the base period instead of the prevailing base year is known as recalculating the indices shifting the base year.
- Re-calculating the indices in that manner is needed
  1. to compare two series of indices with different base years
  2. to represent the values of indices on a recent base year instead of a prevailing far away base year in the past
- The indices on new base year can be computed as follows.

|  |
|--|
| $\text{Price index on new base year} = \frac{\text{Former index in the considered year}}{\text{Former index of the new base year}} \times 100$ |
|--|

- Removing the influence of price changes including in monetary values is known as conversion of monetary values into real values.
- Conversion of monetary values into real values dividing by price indices is known as deflating.
- The income measured in terms of money without considering the price changes is the monetary income and by deflating that monetary income using the consumer price index the real income can be derived.
- Any monetary values can be converted to the real value as follows.

$$\text{Real value} = \frac{\text{monetary value}}{\text{Price index}} \times 100$$

- Few applications in conversion of monetary value into real value deflating by a price index are mentioned below.
  - Computing the real wage of an employee relevant to his monetary wage.
  - Measuring the purchasing power of the monetary unit (Rupee) annually.
  - Ability of identifying the economic inflation, ability of applying to make salary policies of a firm, guiding the state to make economic policies, to determine the demand for goods and services are very important, because of measuring the purchasing power of the money.
  - Converting the Gross Domestic Product (GDP) on current values into GDP on real value.
- Details of few index numbers which are used in Sri Lanka are mentioned below in brief,
- **National Consumer Price Index (NCPI)**
  - This is currently used price index in Sri Lanka to measure the changes in prices of consumer goods.
  - This index has been constructed using 407 different items organizing under 12 categories with 10 sub divisions.
  - It has weighted as 44.04 for food items and 56.96 for other items.
  - The base year of this index is 2013.
  - The values of this index is monthly computed and declared by the department of Census & Statistics.
  - The data are collected covering all the nine provinces for computing this index.

- **Wholesale Price Index**

- This is the price index which is used to measure the changes in the price of goods sold by the producer in primary market.
- This index has been constructed considering 81 items under 13 categories.
- 67.8% has been weighted for food products in this index.
- 1974 is the base year
- This index is computed by the Central Bank of Sri Lanka.

- **All share Price Index**

- This is the index number which is used to compute the changes in the prices of shares exchanged among the listed companies in Colombo Stock Exchange (Share Market)
- The base year of this index is 1985.

- **S & P – 20 Index**

- This is an index number which is computed by drawing out a sample of 20 companies from the listed companies in Colombo Stock Exchange, on the facts that the Market capitalization, Liquidity and Financial stability of those companies.
- This index number is used to measure the changes in 20 main companies share prices.
- 2012 is the base year of this index and the value of the base year is considered as 1000.