

# 2 Chapter

# Data Processing

A number of statistical techniques are used to analyse the data. There are three such statistical techniques:

1. Measures of Central Tendency
2. Measures of Dispersion
3. Measures of Relationship

## MEASURES OF CENTRAL TENDENCY

The measure of central tendency provides the central value that is an ideal representative of a set of observations. It is the statistical technique used to find out the value which lies in the centre of distributions. This number denoting the central tendency is the representative figure for the entire data set because it is the point about which items have a tendency to cluster.

Measures of central tendency are also known as **statistical averages**. There are a number of the measures of central tendency, such as the mean, median and the mode.

### 1. MEAN

Mean is the simple arithmetic average of the different values of a variable. The mean is the value which is derived by summing all the values and dividing it by the number of observations.

#### Computing Mean from Ungrouped Data:

##### 1. Direct Method

Mean from ungrouped data is calculated by adding the values for each observation and dividing by the sum of all observations.

The mean is calculated using the following formula: **Write Formula**

$$\bar{X} = \frac{\sum x}{N}$$

Where,  
 $\bar{X}$  = Mean  
 $\sum$  = Sum of a series of measures  
 $x$  = A raw score in a series of measures  
 $\sum x$  = The sum of all the measures  
 $N$  = Number of measures

**Example 1:** \_\_\_\_\_

##### 2. Indirect Method

For large number of observations, the indirect method is normally used to compute the mean. It helps in reducing the values of the observations to smaller numbers by subtracting a constant value from them.

For example: in the large number of observations we can reduce these values by selecting 'assumed mean' and subtracting the chosen number from each value. Such an operation is known as coding. The mean is then worked out from these reduced numbers. The following formula is used in computing the mean using indirect method:

$$\bar{X} = A + \frac{\sum d}{N}$$

Where,

$A$  = Subtracted constant

$\sum d$  = Sum of the coded scores

$N$  = Number of individual observations in a series

**Example 2:** \_\_\_\_\_

The mean value is the same when computed either by direct method or indirect method.

### Computing Mean from Grouped Data

The mean is also computed for the grouped data using either direct or indirect method.

#### **1. Direct Method**

When scores are grouped into a frequency distribution, the individual values lose their identity. These values are represented by the midpoints of the class intervals in which they are located. Mean is calculated using the following formula:

$$\bar{X} = \frac{\sum fx}{N}$$

Where:

$\bar{X}$  = Mean

$f$  = Frequencies

$x$  = Midpoints of class intervals

$N$  = Number of observations (it may also be defined as  $\sum f$ )

To find the mean, each midpoint ( $x$ ) is multiplied by the frequency ( $f$ ) and their sum ( $fx$ ) divided by  $N$ .

**Example 3:** \_\_\_\_\_

### Indirect Method

The following formula can be used for the indirect method for grouped data.

$$\bar{x} = A \pm \frac{\sum fd}{N}$$

Where,

$A$  = Midpoint of the assumed mean group

$f$  = Frequency

$d$  = Deviation from the assumed mean group ( $A$ )

$N$  = Sum of cases

$i$  = Interval width

The following steps are involved in computing mean using the direct method:

- (i) Mean is to be assumed in the group. It is preferably assumed from the class as near to the middle of the series as possible. This procedure minimises the magnitude of computation.
- (ii) Find the deviations of midpoint of each class from the midpoint of the assumed mean group.
- (iii) Find the  $fd$  by multiplying values of each  $f$  by its corresponding  $d$ . Then, positive and negative values of  $fd$  are added separately and their absolute difference is found ( $\sum fd$ ). Note that the sign attached to  $fd$  is replaced in the formula following  $A$ , where  $\pm$  is given.

**Example 4:** \_\_\_\_\_

The Indirect mean method will work for both equal and unequal class intervals.

## **2. MEDIAN**

Median is a positional average. It may be defined "as the point in a distribution with an equal number of cases on each side of it". The Median is expressed using symbol  $M$ .

### **Computing Median for Ungrouped Data**

When the scores are ungrouped, these are arranged in ascending or descending order. Median can be found by locating the central observation or value in the arranged series. The following equation is used to compute the median:

$$\text{Value of } \left( \frac{N+1}{2} \right) \text{th item}$$

Computation: Median ( $M$ ) may be calculated in the following steps:

- (i) Arrange the given data in ascending or descending order.
- (ii) Apply the formula for locating the central value in the series.

**Example 5:** \_\_\_\_\_

### **Computing Median for Grouped Data**

When the scores are grouped, we have to find the value of the point where an individual or observation is centrally located in the group. It can be computed using the following formula:

$$M = l + \frac{i}{f} \left( \frac{N}{2} - c \right)$$

Where,

$M$  = Median for grouped data

$l$  = Lower limit of the median class

$i$  = Interval

$f$  = Frequency of the median class

$m = N/2$  = Total number of frequencies or number of observations divided by 2.

$c$  = Cumulative frequency of the pre-median class.

The median is computed in the steps given below:

- (i) Cumulative frequencies (F) are obtained by adding each normal frequency of the successive interval groups.
- (iii) Median number is obtained by  $N/2$ .
- (iv) Count into the cumulative frequency distribution (F) from the top towards bottom until the value next greater than  $N/2$  is reached. Thus the cumulative frequency of the pre median class and actual frequency of the median class is found out.
- (v) The median is then computed by substituting all the values determined in the step 4 in the equation.

**Example 6:** \_\_\_\_\_

### **3. MODE**

The value that occurs most frequently in a distribution is referred to as mode. It is symbolised as Z or  $M_0$ . Mode is a measure that is less widely used compared to mean and median. There can be more than one type mode in a given data set.

#### **Computing Mode for Ungrouped Data**

While computing mode from the given data sets all measures are first arranged in ascending or descending order. It helps in identifying the most frequently occurring measure easily.

- (i) If only one number occurs most times in a series, it is called unimodal.
- (ii) If two numbers are occurring mostly in the distribution, the dataset is called bimodal.
- (iii) If three values have equal and highest frequency, the series is trimodal.
- (iv) Similarly, a recurrence of many measures in a series makes it multimodal.
- (v) However, when there is no measure being repeated in a series it is designated as without mode.

Computation steps:

1. To find the mode the measures are arranged in ascending order.
2. The measure occurring most times in the series is the mode in the given dataset.

**Example 7:** \_\_\_\_\_

#### **Comparison of Mean, Median and Mode**

Normal distribution curve is used to compare the three measures of the central tendency mean, median and mode. The normal curve refers to a frequency distribution. This is often called a bell-shaped curve. Many human traits such as intelligence, personality scores and student achievements have normal distributions. Most of the observations lie on and around the middle value. As one approaches the extreme values, the number of observations reduces in a symmetrical manner.