UNIT 1 MEASURES OF CENTRAL TENDENCY

Structure

1.1 INTRODUCTION

As we know that after the classification and tabulation of data one often finds too much detail for many uses that may be made of information available. We, therefore, need further analysis of the tabulated data to draw inference. In this unit we are going to discuss about measures of central tendencies. For the purpose of analysis, very important and powerful tool is a single average value that represents the entire mass of data.

The term average in Statistics refers to a one figure summary of a distribution. It gives a value around which the distribution is concentrated. For this reason that average is also called the measure of central tendency. For example, suppose Mr. X drives his car at an average speed of 60 km/hr. We get an idea that he drives fast (on Indian roads of course!). To compare the performance of two classes, we can compare the average scores in the same test given to these two classes. Thus, calculation of average condenses a distribution into a single value that is supposed to represent the distribution. This helps both individual assessments of a distribution as well as in comparison with another distribution.

This unit comprises some sections as the Section 1.2 gives the definition of measures of central tendency. The significance and properties of a good measure of central tendency are also described in Sub-sections 1.2.1 and 1.2.2. In sub sequent Sections 1.3, 1.4, 1.5 and 1.6, direct and indirect methods for calculating Arithmetic mean, Weighted mean, Median and mode, respectively are explained with their merits and demerits, whereas in Sections 1.7 and 1.8 methods for calculating Geometric mean and Harmonic mean for ungrouped and grouped data, respectively are explained with their merits and demerits. The concepts and methods of calculating the partition values are described in Section 1.9.

Objectives

After studying this unit, you would be able to

- define an average;
- explain the significance of a measure of central tendency;
- explain the properties of a good average;
- calculate the different types of measures of central tendency;
- describe the merits and demerits different types of measures of central tendency; and
- describe the methods of calculation of partition values.

1.2 MEASURES OF CENTRAL TENDENCY

According to **Professor Bowley**, averages are “statistical constants which enable us to comprehend in a single effort the significance of the whole”. They throw light as to how the values are concentrated in the central part of the distribution.

For this reason as on last page that they are also called the measures of central tendency, an average is a single value which is considered as the most representative for a given set of data. Measures of central tendency show the tendency of some central value around which data tend to cluster.

1.2.1 Significance of the Measure of Central Tendency

The following are two main reasons for studying an average:

1. **To get a single representative**

   Measure of central tendency enables us to get a single value from the mass of data and also provide an idea about the entire data. For example it is impossible to remember the heights measurement of all students in a class. But if the average height is obtained, we get a single value that represents the entire class.

2. **To facilitate comparison**

   Measures of central tendency enable us to compare two or more than two populations by reducing the mass of data in one single figure. The comparison can be made either at the same time or over a period of time. For example, if a subject has been taught in more than two classes so by obtaining the average marks of those classes, comparison can be made.

1.2.2 Properties of a Good Average

The following are the properties of a good measure of average:

1. **It should be simple to understand**

   Since we use the measures of central tendency to simplify the complexity of a data, so an average should be understandable easily otherwise its use is bound to be very limited.

2. **It should be easy to calculate**

   An average not only should be easy to understand but also should be simple to compute, so that it can be used as widely as possible.

3. **It should be rigidly defined**

   A measure of central tendency should be defined properly so that it has an appropriate interpretation. It should also have an algebraic formula so that if different people compute the average from same figures, they get the same answer.

4. **It should be liable for algebraic manipulations**

   A measure of central tendency should be liable for the algebraic manipulations. If there are two sets of data and the individual information is available for both set, then one can be able to find the information regarding the combined set also then something is missing.

5. **It should be least affected by sampling fluctuations**

   We should prefer a tool which has a sampling stability. In other words, if we select 10 different groups of observations from same population and compute the average of each group, then we should expect to get approximately the same values. There may be little difference because of the sampling fluctuation only.

6. **It should be based on all the observations**

   If any measure of central tendency is used to analyse the data, it is desirable that each and every observation is used for its calculation.

7. **It should be possible to calculate even for open-end class intervals**

   A measure of central tendency should able to be calculated for the data with open end classes.

8. **It should not be affected by extremely small or extremely large observations**

   It is assumed that each and every observation influences the value of the average. If one or two very small or very large observations affect the average i.e. either increase or decrease its value largely, then the average cannot be consider as a good average.

1.2.3 Different Measures of Central Tendency

The following are the various measures of central tendency:

1. Arithmetic Mean
2. Weighted Mean
3. Median
4. Mode
5. Geometric Mean
6. Harmonic Mean

1.2.4 Partition Values

1. Quartiles
2. Deciles
3. Percentiles

1.3 ARITHMETIC MEAN

Arithmetic mean (also called mean) is defined as the sum of all the observations divided by the number of observations. Arithmetic mean (AM) may be calculated for the following two types of data:

1. For Ungrouped Data

For ungrouped data, arithmetic mean may be computed by applying any of the following methods:

(1) Direct Method

Mathematically, if $x_1, x_2, \ldots, x_n$ are the n observations then their mean is

$$\overline{X} = \frac{(x_1 + x_2 + x_3 + \ldots + x_n)}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

If f_i is the frequency of x_i (i=1, 2,…, k), the formula for arithmetic mean would be

$$\overline{X} = \frac{(f_1x_1 + f_2x_2 + \ldots + f_kx_k)}{(f_1 + f_2 + \ldots + f_k)}$$

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

(2) Short-cut Method

The arithmetic mean can also be calculated by taking deviations from any arbitrary point "A", in which the formula shall be

$$\overline{X} = A + \frac{\sum_{i=1}^{n} d_i}{n} \qquad \text{where, } d_i = x_i - A$$

If f_i is the frequency of x_i (i=1, 2,…, k), the formula for arithmetic mean would be

$$\overline{X} = A + \frac{\sum_{i=1}^{k} f_i d_i}{\sum_{i=1}^{k} f_i}, \qquad \text{where, } d_i = x_i - A$$

Here, k is the number of distinct observations in the distribution.

Note: Usually the short-cut method is used when data are large.

Example 1: Calculate mean of the weights of five students

54, 56, 70, 45, 50 (in kg)

Solution: If we denote the weight of students by x then mean is obtained by

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Thus, $$\overline{X} = \frac{54+56+70+45+50}{5} = \frac{275}{5} = 55$$

Therefore, average weight of students is 55 kg.

Example 2: Compute arithmetic mean of the weight of students for the given data in Example 1 by using shortcut method.

Solution: For shortcut method, we use following formula

$$\overline{X} = A + \frac{\sum d_i}{n}, \qquad \text{where } d_i = x_i - A$$

If 50 is taken as the assumed value A in the given data in Example 1 then, for the calculation of d_i we prepare following table:

x	**d = x - A**
54	54–50 = 4
56	56–50 = 6
70	70–50 =20
45	45–50 = – 5
50	50–50=0
	$\sum_{i=1}^{n} d_i = 25$

We have A = 50 then,

$$\overline{X} = A + \frac{\sum_{i=1}^{n} d_i}{n} = 50 + \frac{25}{5} = 50 + 5 = 55$$

Example 3: Calculate arithmetic mean for the following data:

x	20	30	40
f	5	6	4

Solution: We have the following frequency distribution:

x	**f**	**fx**
20	5	100
30	6	180
40	4	160
	$\sum_{i=1}^{k} f_i = 15$	$\sum_{i=1}^{k} f_i x_i = 440$

Arithmetic Mean, $\overline{X} = \dfrac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{440}{15} = 29.3$$

E1) Find the arithmetic mean of the following observations:

5, 8, 12, 15, 20, 30.

E2) For the following discrete frequency distribution find arithmetic mean:

Wages (in Rs)	20	25	30	35	40	50
No. of workers	5	8	20	10	5	2

2 For Grouped Data

Direct Method

If f_i is the frequency of x_i (i =1, 2,…, k) where x_i is the mid value of the i^{th} class interval, the formula for arithmetic mean would be

$$\overline{X} = \frac{(f_1x_1 + f_2x_2 + \ldots + f_kx_k)}{(f_1 + f_2 + \ldots + f_k)}$$

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{\sum fx}{\sum f} = \frac{\sum fx}{N},$$

where, $N = f_1 + f_2 + \ldots + f_k$

Short-cut Method

$$\overline{X} = A + \frac{\sum_{i=1}^{k} f_i d_i}{\sum_{i=1}^{k} f_i}, \qquad \text{where, } d_i = x_i - A$$

Here, f_i would be representing the frequency of the i^{th} class, x_i is the mid-value of the i^{th} class and k is the number of classes.

Example 4: For the following data, calculate arithmetic mean using direct method:

Class Interval	0-10	10-20	20-30	30-40	40-50
Frequency	3	5	7	9	4

Solution: We have the following distribution:

Class Interval	Mid Value x	Frequency f	fx
0-10	05	03	15
10-20	15	05	75
20-30	25	07	175
30-40	35	09	315
40-50	45	04	180
		$\sum_{i=1}^{k} f_i = N = 28$	$\sum_{i=1}^{k} f_i x_i = 760$

$$\text{Mean} = \frac{\sum_{i=1}^{k} f_i x_i}{N} = 760/28 = 27.143$$

Now let us solve one exercise.

E3) Find arithmetic mean of the distribution of marks given below:

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	6	9	17	10	8

1.3.1 Properties of Arithmetic Mean

Arithmetic mean fulfills most of the properties of a good average except the last two. It is particularly useful when we are dealing with a sample as it is least affected by sampling fluctuations. It is the most popular average and should always be our first choice unless there is a strong reason for not using it.

Three algebraic properties of mean are given below:

Property 1: Sum of deviations of observations from their mean is zero. Deviation is also called dispersion that will be discussed in detail in Unit 2 of this block.

Proof: We have to prove $\sum(x - \text{mean}) = 0$

The sum of deviations of observations $x_1, x_2, \ldots, x_n$ from their mean is

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = \sum_{i=1}^{n} x_i - n\,\bar{x}$$

$$\Rightarrow \sum_{i=1}^{n} x_i - n\frac{1}{n}\sum_{i=1}^{n} x_i = 0$$

Property 2: Sum of squares of deviations taken from mean is least in comparison to the same taken from any other average.

Proof: We have

$$\sum_{i=1}^{n}\left(x_i - A\right)^2 = \sum_{i=1}^{n}\left(x_i - \bar{x} + \bar{x} - A\right)^2$$

where, A is an assumed mean / Median / Mode

$$\Rightarrow \sum_{i=1}^{n}\left(x_i - A\right)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - A)^2 + 2(\bar{x} - A)\sum_{i=1}^{n}(x_i - \bar{x})$$

$$\Rightarrow \sum_{i=1}^{n}\left(x_i - A\right)^2 - \sum_{i=1}^{n}(x_i - \bar{x})^2 = n(\bar{x} - A)^2 + 0 \qquad \text{(By Property 1)}$$

$$\Rightarrow \sum_{i=1}^{n}(x_i - A)^2 - \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \geq 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \leq \sum_{i=1}^{n}(x_i - A)^2$$

That means the sum of squares of deviations taken from mean is least in comparison to the same taken from any other average.

Property 3: Arithmetic mean is affected by both the change of origin and scale.

Proof: If $u_i = \frac{x_i - a}{h}$,

where, a and h are constant. Then

$$x_i = a + h u_i$$

$$\sum_{i=1}^{n} x_i = na + h\sum_{i=1}^{n} u_i$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i = a + h\frac{1}{n}\sum_{i=1}^{n} u_i$$

$$\overline{X} = a + h\ \overline{U}$$

1.3.2 Merits and Demerits of Arithmetic Mean

Merits of Arithmetic Mean

1. It utilizes all the observations;
2. It is rigidly defined;
3. It is easy to understand and compute; and
4. It can be used for further mathematical treatments.

Demerits of Arithmetic Mean

1. It is badly affected by extremely small or extremely large values;
2. It cannot be calculated for open end class intervals; and
3. It is generally not preferred for highly skewed distributions.

1.4 WEIGHTED MEAN

Weight here refers to the importance of a value in a distribution. A simple logic is that a number is as important in the distribution as the number of times it appears. So, the frequency of a number can also be its weight. But there may be other situations where we have to determine the weight based on some other reasons. For example, the number of innings in which runs were made may be considered as weight because runs (50 or 100 or 200) show their importance. Calculating the weighted mean of scores of several innings of a player, we may take the strength of the opponent (as judged by the proportion of matches lost by a team against the opponent) as the corresponding weight. Higher the proportion stronger would be the opponent and hence more would be the weight. If x_i has a weight w_i, then weighted mean is defined as:

$$\overline{X}_W = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i} \qquad \text{for all } i = 1, 2, 3, \ldots, k.$$

1.5 MEDIAN

Median is that value of the variable which divides the whole distribution into two equal parts. Here, it may be noted that the data should be arranged in ascending or descending order of magnitude. When the number of observations is odd then the median is the middle value of the data. For even number of observations, there will be two middle values. So we take the arithmetic mean of these two middle values. Number of the observations below and above the median, are same. Median is not affected by extremely large or extremely small values (as it corresponds to the middle value) and it is also not affected by open end class intervals. In such situations, it is preferable in comparison to mean. It is also useful when the distribution is skewed (asymmetric). Skewness will be discussed in Unit 4 of this block.

1. Median for Ungrouped Data

Mathematically, if $x_1, x_2, \ldots, x_n$ are the n observations then for obtaining the median first of all we have to arrange these n values either in ascending order or in descending order. When the observations are arranged in ascending or descending order, the middle value gives the median if n is odd. For even number of observations there will be two middle values. So we take the arithmetic mean of these two values.

$$M_d = \left(\frac{n+1}{2}\right)^{th} \text{observation} \;;\; (\text{when } n \text{ is odd})$$

$$M_d = \frac{\left(\frac{n}{2}\right)^{th} \text{observation} + \left(\frac{n}{2}+1\right)^{th} \text{observation}}{2} \;;\; (\text{when } n \text{ is even})$$

Example 5: Find median of following observations:

6, 4, 3, 7, 8

Solution: First we arrange the given data in ascending order as

3, 4, 6, 7, 8

Since, the number of observations i.e. 5, is odd, so median would be the middle value that is 6.

Example 6: Calculate median for the following data:

7, 8, 9, 3, 4, 10

Solution: First we arrange given data in ascending order as

3, 4, 7, 8, 9, 10

Here, Number of observations (n) = 6 (even). So we get the median by

$$M_d = \frac{\left(\frac{n}{2}\right)^{th} \text{observation} + \left(\frac{n}{2}+1\right)^{th} \text{observation}}{2}$$

$$= \frac{\left(\frac{6}{2}\right)^{th} \text{observation} + \left(\frac{6}{2}+1\right)^{th} \text{observation}}{2}$$

$$M_d = \frac{3^{rd} \text{observation} + 4^{th} \text{observation}}{2}$$

$$= \frac{7+8}{2} = \frac{15}{2} = 7.5$$

E4) Find the median of the following values:

(i) 10, 6, 15, 2, 3, 12, 8

(ii) 10, 6, 2, 3, 8, 15, 12, 5

For Ungrouped Data (when frequencies are given)

If x_i are the different value of variable with frequencies f_i then we calculate cumulative frequencies from f_i then median is defined by

$M_d =$ Value of variable corresponding to $\left(\frac{\sum f}{2}\right)^{th} = \left(\frac{N}{2}\right)^{th}$ cumulative frequency.

Note: If N/2 is not the exact cumulative frequency then value of the variable corresponding to next cumulative frequencies is the median.

Example 7: Find Median from the given frequency distribution

x	20	40	60	80
f	7	5	4	3

Solution: First we find cumulative frequency

x	**f**	**c.f.**
20	7	7
40	5	12
60	4	16
80	3	19
	$\sum_{i=1}^{k} f_i = 19$	

M_d = Value of the variable corresponding to the

$\left(\frac{19}{2}\right)^{th}$ cumulative frequency

= Value of the variable corresponding to 9.5 since 9.5 is not among c.f.

So, the next cumulative frequency is 12 and the value of variable against 12 cumulative frequency is 40. So median is 40.

E5) Find the median of the following frequency distribution:

Mid Values	5	10	15	20	25	30	35	40	45	50
No.of students	2	5	10	14	16	20	13	9	7	4

---

2. Median for Grouped Data

For class interval, first we find cumulative frequencies from the given frequencies and use the following formula for calculating the median:

$$\text{Median} = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

where, L = lower class limit of the median class,

N = total frequency,

C = cumulative frequency of the pre-median class,

f = frequency of the median class, and

h = width of the median class.

Median class is the class in which the $(N/2)^{th}$ observation falls. If N/2 is not among any cumulative frequency then next class to the N/2 will be considered as median class.

Example 8: Calculate median for the data given in Example 4.

Solution: We first form a cumulative frequency table (cumulative frequency of a class gives the number of observations less than the upper limit of the class; strictly speaking, this is called cumulative frequency of less than type; we also have cumulative frequency of more than type which gives the number of observations greater than or equal to the lower limit of a class):

Class Interval	Frequency f	Cumulative Frequency (< type)
0-10	3	3
10-20	5	8
20-30	7	15
30-40	9	24
40-50	4	28

$$\sum_{i=1}^{k} f_i = N = 28 \qquad \Rightarrow \qquad \frac{N}{2} = \frac{28}{2} = 14$$

Since 14 is not among the cumulative frequency so the class with next cumulative frequency i.e. 15, which is 20-30, is the median class.

We have L = lower class limit of the median class = 20

N = total frequency = 28

C = cumulative frequency of the pre median class =8

f = frequency of the median class = 7

h = width of median class = 10

Now substituting all these values in the formula of Median

$$\text{Median} = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

$$M_d = 20 + \frac{14-8}{7} \times 10 = 28.57$$

Therefore, median is 28.57.

E6) Find Median for the following frequency distribution:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No.of students	5	10	15	20	12	10	8

E7) Find the missing frequency when median is given as Rs 50.

Expenditure (Rs)	0-20	20-40	40-60	60-80	80-100
No. of families	5	15	30	--	12

1.5.1 Merits and Demerits of Median

Merits of Median

1. It is rigidly defined;
2. It is easy to understand and compute;
3. It is not affected by extremely small or extremely large values; and
4. It can be calculated even for open end classes (like “less than 10” or “50 and above”).

Demerits of Median

1. In case of even number of observations we get only an estimate of the median by taking the mean of the two middle values. We don’t get its exact value;
2. It does not utilize all the observations. The median of 1, 2, 3 is 2. If the observation 3 is replaced by any number higher than or equal to 2 and if the number 1 is replaced by any number lower than or equal to 2, the median value will be unaffected. This means 1 and 3 are not being utilized;

3. It is not amenable to algebraic treatment; and
4. It is affected by sampling fluctuations.

1.6 MODE

Highest frequent observation in the distribution is known as mode. In other words, mode is that observation in a distribution which has the maximum frequency. For example, when we say that the average size of shoes sold in a shop is 7 it is the modal size which is sold most frequently.

For Ungrouped Data

Mathematically, if $x_1, x_2, \ldots, x_n$ are the n observations and if some of the observation are repeated in the data, say x_i is repeated highest times then we can say the x_i would be the mode value.

Example 9: Find mode value for the given data

2, 2, 3, 4, 7, 7, 7, 7, 9, 10, 12, 12

Solution: First we prepare frequency table as

x	2	3	4	7	9	10	12
f	2	1	1	4	1	1	2

This table shows that 7 have the maximum frequency. Thus, mode is 7.

E8) Find the model size for the following items:

4, 7, 6, 5, 4, 7, 8, 3, 7, 2, 7, 6, 1, 2, 5

For Grouped Data:

Data where several classes are given, following formula of the mode is used

$$M_0 = L + \frac{|f_1 - f_0|}{|f_1 - f_0| + |f_1 - f_2|} \times h$$

where, L = lower class limit of the modal class,

f_1 = frequency of the modal class,

f_0 = frequency of the pre-modal class,

f_2 = frequency of the post-modal class, and

h = width of the modal class.

Modal class is that class which has the maximum frequency.

Example 10: For the data given in Example 4, calculate mode.

Solution: Here the frequency distribution is

Class Interval	Frequency
0-10	3
10-20	5
30-40	7
40-50	9
50-60	4

Corresponding to highest frequency 9 model class is 40-50 and we have

$L = 40,\ f_1 = 9,\quad f_o = 7,\ f_2 = 4,\ h = 10$

Applying the formula,

$$\text{Mode} = 40 + \frac{(9-7)}{(2\times 9 - 7 - 4)} \times 10$$

$$= 42.86$$

E9) Calculate mode from the data given in E6)

1.6.1 Relationship between Mean, Median and Mode

For a symmetrical distribution the mean, median and mode coincide. But if the distribution is moderately asymmetrical, there is an empirical relationship between them. The relationship is

$$\text{Mean} - \text{Mode} = 3\ (\text{Mean} - \text{Median})$$

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

Note: Using this formula, we can calculate mean/median/mode if other two of them are known.

E10) In an asymmetrical distribution the mode and mean are 35.4 and 38.6 respectively. Calculate the median.

1.6.2 Merits and Demerits of Mode

Merits of Mode

1. Mode is the easiest average to understand and also easy to calculate;
2. It is not affected by extreme values;
3. It can be calculated for open end classes;
4. As far as the modal class is confirmed the pre-modal class and the post modal class are of equal width; and
5. Mode can be calculated even if the other classes are of unequal width.

Demerits of Mode

1. It is not rigidly defined. A distribution can have more than one mode;
2. It is not utilizing all the observations;
3. It is not amenable to algebraic treatment; and
4. It is greatly affected by sampling fluctuations.

1.7 GEOMETRIC MEAN

The geometric mean (GM) of n observations is defined as the n-th root of the product of the n observations. It is useful for averaging ratios or proportions. It is the ideal average for calculating index numbers (index numbers are economic barometers which reflect the change in prices or commodity consumption in the current period with respect to some base period taken as standard). It fails to give the correct average if an observation is zero or negative.

1. **For Ungrouped Data**

If $x_1, x_2, ..., x_n$ are the n observations of a variable X then their geometric mean is

$$GM = \sqrt[n]{x_1 x_2 ... x_n}$$

$$GM = (x_1 x_2 ... x_n)^{\frac{1}{n}}$$

Taking log of both sides

$$\log GM = \frac{1}{n} \log (x_1 x_2 ... x_n)$$

$$\log GM = \frac{1}{n} (\log x_1 + \log x_2 + ... + \log x_n)$$

$$\Rightarrow GM = \text{Antilog} \left(\frac{1}{n} \sum_{i=1}^{n} \log x_i \right)$$

Example 11: Find geometric mean of 2, 4, 8.

Solution: $GM = (x_1 \, x_2 ... x_n)^{\frac{1}{n}}$

$$GM = (2 \times 4 \times 8)^{\frac{1}{3}} = (64)^{\frac{1}{3}} = 4$$

Thus, geometric mean is 4.

E11) Find the GM of 4, 8, 16.

E12) Calculate GM of 5, 15, 25, 35.

---

2. For Grouped data

If $x_1 x_2, ..., x_k$ are k values (or mid values in case of class intervals) of a variable X with frequencies $f_1, f_2, ..., f_k$ then

$$GM = \left(x_1^{f_1} x_2^{f_2} ... x_k^{f_k} \right)^{\frac{1}{\sum f_i}}$$

$$GM = \left(x_1^{f_1} x_2^{f_2} ... x_k^{f_k} \right)^{\frac{1}{N}}$$

where $N = f_1 + f_2 + ... + f_k$

Taking log of both sides

$$\log GM = \frac{1}{N}\log\left(x_1^{f_1} x_2^{f_2} \ldots x_k^{f_k}\right)$$

$$\log GM = \frac{1}{N}\left(f_1 \log x_1 + f_2 \log x_2 + \ldots + f_k \log x_k\right)$$

$$\Rightarrow GM = \text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{k} f_i \log x_i\right)$$

Example 12: For the data in Example 4, calculate geometric mean.

Solution:

Class	Mid Value (x)	Frequency (f)	log x	f× log x
0-10	5	3	0.6990	2.0970
10-20	15	5	1.1761	5.8805
20-30	25	7	1.3979	9.7853
30-40	35	9	1.5441	13.8969
40-50	45	4	1.6532	6.6128
		N = 28		$\sum f_i \log x_i = 38.2725$

Using the formula

$$GM = \text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{k} f_i \log x_i\right)$$

$$= \text{Antilog}\left(\frac{38.2725}{28}\right)$$

$$= \text{Antilog}\ (1.3669)$$

$$= 23.28$$

E13) Calculate GM of the following distribution

Class	0-10	10-20	20-30	30-40	40-50
Frequency	12	15	25	18	10

1.7.1 Merits and Demerits of Geometric Mean

Merits of Geometric Mean

1. It is rigidly defined;
2. It utilizes all the observations;
3. It is amenable to algebraic treatment (the reader should verify that if GM_1 and GM_2 are Geometric Means of two series-Series 1 of size n and Series 2 of size m respectively, then Geometric Mean of the combined series is given by

   Log GM = (n GM_1 + m GM_2) / (n + m);
4. It gives more weight to small items; and
5. It is not affected greatly by sampling fluctuations.

Demerits of Geometric Mean

1. Difficult to understand and calculate; and
2. It becomes imaginary for an odd number of negative observations and becomes zero or undefined if a single observation is zero.

1.8 HARMONIC MEAN

The harmonic mean (HM) is defined as the reciprocal (inverse) of the arithmetic mean of the reciprocals of the observations of a set.

1. For Ungrouped Data

If $x_1, x_2, ..., x_n$ are the n observations of a variable X, then their harmonic mean is

$$HM = \frac{1}{\frac{1}{n}\left[\frac{1}{x_1}+\frac{1}{x_2}+...+\frac{1}{x_n}\right]}$$

$$HM = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

Example 13: Calculate the Harmonic mean of 1, 3, 5, 7

Solution: Formula for harmonic mean is

$$HM = \frac{1}{\frac{1}{n}\left[\frac{1}{x_1}+\frac{1}{x_2}+...+\frac{1}{x_n}\right]}$$

$$HM = \frac{1}{\frac{1}{4}\left[\frac{1}{1}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}\right]}$$

= 4 / (1.0000 + 0.3333 + 0.2000 + 0.1428)

= 4/1.6761 = 2.39

E14) Calculate the Harmonic Mean of 2, 4, 6.

2. For Grouped Data

If $x_1 x_2, ..., x_k$ are k values (or mid values in case of class intervals) of a variable X with their corresponding frequencies $f_1, f_2, ..., f_k$, then

$$HM = \frac{1}{\frac{1}{N}\left[\frac{f_1}{x_1}+\frac{f_2}{x_2}+...+\frac{f_k}{x_k}\right]}$$

$$HM = \frac{N}{\sum_{i=1}^{k} \frac{f_i}{x_i}} \qquad \text{where, } N = \sum_{i=1}^{k} f_i.$$

When equal distances are travelled at different speeds, the average speed is calculated by the harmonic mean. It cannot be calculated if an observation is zero.

Example 14: For the data given in Example 4, calculate harmonic mean.

Solution:

Class	Mid Value (x)	Frequency (f)	f/x
0-10	5	3	0.600
10-20	15	5	0.330
20-30	25	7	0.280
30-40	35	9	0.257
40-50	45	4	0.088
		N = ∑f = 28	∑f/x = 1.555

Using the formula,

$$HM = \frac{N}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

$$= 28/1.555 = 17.956$$

E15) Calculate harmonic mean for the given data:

Class	0-10	10-20	20-30	30-40	40-50
Frequency	3	5	10	8	4

1.8.1 Merits and Demerits of Harmonic Mean

Merits of Harmonic mean

1. It is rigidly defined;
2. It utilizes all the observations;
3. It is amenable to algebraic treatment; and
4. It gives greater importance to small items.

Demerits of Harmonic Mean

1. Difficult to understand and compute.

1.8.2 Relations between AM, GM and HM

Relation 1: AM ≥ GM ≥ HM

Proof: Let x_1 and x_2 be two real numbers which are non-zero and non negative. Then

$$AM = \frac{x_1 + x_2}{2}$$

$$GM = \sqrt{x_1.x_2}$$

$$HM = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$$

Consider $\left(\sqrt{x_1} - \sqrt{x_2}\right)^2 \geq 0$

$$x_1 + x_2 - 2\sqrt{x_1}x_2 \geq 0$$

$$\frac{x_1 + x_2}{2} \geq \sqrt{x_1 x_2} \quad \text{so} \quad AM \geq GM \qquad \text{... (1)}$$

Again $\left(\frac{1}{\sqrt{x_1}} - \frac{1}{\sqrt{x_2}}\right)^2 \geq 0$

$$\frac{1}{x_1} + \frac{1}{x_2} - \frac{2}{\sqrt{x_1}.\sqrt{x_2}} \geq 0$$

$$\sqrt{x_1 x_2} \geq \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} \quad \text{so} \quad GM \geq HM \qquad \text{... (2)}$$

So by equations (1) & (2)

$$AM \geq GM \geq HM$$

Relation 2: GM = $\sqrt{AM.HM}$

Proof: Let x_1 and x_2 be two real numbers which are non-zero and non negative. Then

$$AM \times HM = \frac{x_1 + x_2}{2} \cdot \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$$

$$= (x_1 + x_2)\frac{x_1 x_2}{(x_1 + x_2)}$$

$$= x_1 x_2$$

$$(GM)^2 = AM \times HM$$

So $\quad GM = \sqrt{AM \times HM}$

1.9 PARTITION VALUES

Partition values are those values of variable which divide the distribution into a certain number of equal parts. Here it may be noted that the data should be arranged in ascending or descending order of magnitude. Commonly used partition values are quartiles, deciles and percentiles. For example, quartiles divide the data into four equal parts. Similarly, deciles and percentiles divide the distribution into ten and hundred equal parts, respectively.

1.9.1 Quartiles

Quartiles divide whole distribution in to four equal parts. There are three quartiles- 1st Quartile denoted as Q_1, 2nd Quartile denoted as Q_2 and 3rd Quartile as Q_3, which divide the whole data in four parts. 1st Quartile contains the ¼ part of data, 2nd Quartile contains ½ of the data and 3rd Quartile contains the ¾ part of data. Here, it may be noted that the data should be arranged in ascending or descending order of magnitude.

For Ungrouped Data

For obtaining the quartiles first of all we have to arrange the data either in ascending order or in descending order of their magnitude. Then, we find the $(N/4)^{th}$, $(N/2)^{th}$ and $(3N/4)^{th}$ placed item in the arranged data for finding out the 1st Quartile (Q_1), 2nd Quartile (Q_2) and 3rd Quartile (Q_3) respectively. The value of the $(N/4)^{th}$ placed item would be the 1st Quartile (Q_1), value of $(N/2)^{th}$ placed item would be the 2nd Quartile (Q_2) and value of $(3N/4)^{th}$ placed item would be the 3rd Quartile (Q_3) of that data.

Mathematically, if $x_1 x_2, ..., x_N$ are N values of a variable X then 1st Quartile (Q_1), 2nd Quartile (Q_2) and 3rd Quartile (Q_3) are defined as:

First Quartile (Q_1) = $\frac{N}{4}$th placed item in the arranged data

Second Quartile (Q_2) = $\frac{N}{2}$th placed item in the arranged data

Third Quartile (Q_3) = $3\left(\frac{N}{4}\right)$th placed item in the arranged data

For Grouped Data

If $x_1, x_2, ..., x_k$ are k values (or mid values in case of class intervals) of a variable X with their corresponding frequencies $f_1, f_2, ..., f_k$, then first of all we form a cumulative frequency distribution. After that we determine the i^{th} quartile class as similar as we do in case of median.

The i^{th} quartile is denoted by Q_i and defined as

$$Q_i = L + \frac{\left(\frac{iN}{4} - C\right)}{f} \times h \qquad \text{for} \quad i = 1, 2, 3$$

where, L = lower class limit of i^{th} quartile class,

h = width of the i^{th} quartile class,

N = total frequency,

C = cumulative frequency of pre i^{th} quartile class, and

f = frequencies of i^{th} quartile class.

"i" denotes i^{th} quartile class. It is the class in which $\left(\frac{i \times N}{4}\right)^{th}$ observation falls in cumulative frequency. It is easy to see that the second quartile (i = 2) is the median.

1.9.2 Deciles

Deciles divide whole distribution in to ten equal parts. There are nine deciles. D_1, D_2,...,D_9 are known as 1^{st} Decile, 2^{nd} Decile,...,9^{th} Decile respectively and i^{th} Decile contains the $(iN/10)^{th}$ part of data. Here, it may be noted that the data should be arranged in ascending or descending order of magnitude.

For Ungrouped Data

For obtaining the deciles first of all we have to arrange the data either in ascending order or in descending order of their magnitude. Then, we find the $(1N/10)^{th}$, $(2N/10)^{th}$,,.., $(9N/10)^{th}$ placed item in the arranged data for finding out the 1^{st} decile (D_1), 2^{nd} decile (D_2), ..., 9^{th} decile (D_9) respectively. The value of the $(N/10)^{th}$ placed item would be the 1^{st} decile, value of $(2N/10)^{th}$ placed item would be the 2^{nd} decile. Similarly, the value of $(9N/10)^{th}$ placed item would be the 9^{th} decile of that data.

Mathematically, if $x_1 x_2,...,x_N$ are N values of a variable X then the i^{th} decile is defined as:

$$i^{th} \text{ Decile } (D_i) = \frac{i\,N}{10} \text{th placed item in the arranged data } (i = 1, 2, 3, \ldots, 9)$$

For Grouped Data

If $x_1, x_2, ..., x_k$ are k values (or mid values in case of class intervals) of a variable X with their corresponding frequencies $f_1, f_2,..., f_k$, then first of all we form a cumulative frequency distribution. After that we determine the i^{th} deciles class as similar as we do in case of quartiles.

The i^{th} decile is denoted by D_i and given by

$$D_i = L + \frac{\left(\frac{iN}{10} - C\right)}{f} \times h \qquad \text{for} \quad i = 1, 2, ..., 9$$

where, L = lower class limit of i^{th} decile class,

h = width of the i^{th} decile class,

N = total frequency,

C = cumulative frequency of pre i^{th} decile class; and

f = frequency of i^{th} decile class.

"i" denotes i^{th} decile class. It is the class in which $\left(\frac{i \times N}{10}\right)^{th}$ observation falls in cumulative frequency. It is easy to see that the fifth quartile (i =5) is the median.

1.9.3 Percentiles

Percentiles divide whole distribution in to 100 equal parts. There are ninety nine percentiles. $P_1, P_2, \ldots, P_{99}$ are known as 1^{st} percentile, 2^{nd} percentile,…,99^{th} percentile and i^{th} percentile contains the $(iN/100)^{th}$ part of data. Here, it may be noted that the data should be arranged in ascending or descending order of magnitude.

For Ungrouped Data

For obtaining the percentiles first of all we have to arrange the data either in ascending order or in descending order of their magnitude. Then, we find the $(1N/100)^{th}$, $(2N/100)^{th}$,..., $(99N/100)^{th}$ placed item in the arranged data for finding out the 1^{st} percentile (P_1) , 2^{nd} percentile (P_2),...,99^{th} percentile (P_{99}) respectively. The value of the $(N/100)^{th}$ placed item would be the 1^{st} percentile, value of $(2N/100)^{th}$ placed item would be the 2^{nd} percentile. Similarly, the value of $(99N/100)^{th}$ placed item would be the 99^{th} percentile of that data.

Mathematically, if $x_1 x_2, ..., x_N$ are N values of a variable X then the i^{th} percentile is defined as:

$$i^{th} \text{ Percentile } (P_i) = \frac{iN}{100} \text{th placed item in the arranged data} (i = 1, 2, \ldots ,99)$$

For Grouped Data

If $x_1, x_2, ..., x_k$ are k values (or mid values in case of class intervals) of a variable X with their corresponding frequencies $f_1, f_2, ..., f_k$, then first of all we form a cumulative frequency distribution. After that we determine the i^{th} percentile class as similar as we do in case of median.

The i^{th} percentile is denoted by P_i and given by

$$P_i = L + \frac{\left(\frac{iN}{100} - C\right)}{f} \times h \qquad \text{for} \quad i = 1, 2, ..., 99$$

where, L = lower limit of i^{th} percentile class,

h = width of the i^{th} percentile class,

N = total frequency,

C = cumulative frequency of pre i^{th} prcentile class; and

f = frequency of i^{th} percentile class.

"i" denotes i^{th} percentile class. It is the class in which $\left(\frac{i \times N}{100}\right)^{th}$ observation falls in cumulative frequency.

It is easy to see that the fiftieth percentile (i =50) is the median.

Example 15: For the data given in Example 4, calculate the first and third quartiles.

Solution: First we find cumulative frequency given in the following cumulative frequency table:

Class Interval	Frequency	Cumulative Frequency (< type)
0-10	3	3
10-20	5	8
20-30	7	15
30-40	9	24
40-50	4	28
	$N=\sum f=28$	

Here, N/4 = 28/4 = 7. The 7th observation falls in the class 10-20. So, this is the first quartile class. 3N/4 = 21th observation falls in class 30-40, so it is the third quartile class.

For first quartile L = 10, f = 5, C = 3, N = 28

$$Q_1 = 10+\frac{(7-3)}{5}\times 10 = 18$$

For third quartile L = 30, f = 9, C = 15

$$Q_3 = 30+\frac{(21-15)}{9}\times 10 = 36.67$$

E16) Calculate the first and third quartiles for the data given in E6)

1.10 SUMMARY

In this unit, we have discussed:

1. How to describe an average;
2. The utility of an average;
3. The properties of a good average;
4. The different types of averages along with their merits and demerits; and
5. The different kinds of partition values.

1.11 SOLUTIONS / ANSWERS

E1) For calculating the arithmetic mean, we add all the observations and divide by 6 as follows:

$$\overline{X} = \frac{\sum x}{n} = \frac{5+8+12+15+20+30}{6} = 15$$

Using short-cut method suppose the assumed mean A = 15.

x	d = x–A
5	–10
8	–7
12	–3
15	0
20	+5
30	+15
	$\sum d_i = 0$

$$\overline{X} = A + \frac{\sum d}{n} \quad = 15 + \frac{0}{6} \quad = 15$$

E2) We have the following frequency distribution:

Wages	No. of Workers	fx
20	5	100
25	8	200
30	20	600
35	10	350
40	5	200
50	2	100
	50	$\sum xf = 1550$

$$\overline{X} = \frac{\sum_{i=1}^{6} f_i x_i}{\sum_{i=1}^{6} f_i} \quad = \frac{1550}{50} = 31$$

Using short cut method with assumed mean A = 30

x	f	d = x - 30	fd
20	5	–10	–50
25	8	–5	–40
30	20	0	0
35	10	5	50
40	5	10	50
50	2	20	40
	N = 50		$\sum f\,d = 50$

$$\overline{X} = A + \frac{\sum_{i=1}^{6} f_i d_i}{N}$$

$$= 30 + \frac{50}{50} = 30 + 1 = 31$$

E3) We have the following frequency distribution:

Marks	No. of Students (f)	Mid Points x	fx
0-10	6	5	30
10-20	9	15	135
20-30	17	25	425
30-40	10	35	350
40-50	8	45	360
	$\sum f_i = 50$		$\sum f_i x_i = 1300$

$$\overline{X} = \frac{\sum_{i=1}^{5} f_i x_i}{\sum_{i=1}^{5} f_i} = \frac{1300}{50} = 26$$

Using short-cut method

Marks	f	x	$d = \frac{x-25}{10}$	f d
0-10	6	5	–2	–12
10-20	9	15	–1	–9
20-30	17	25=A	0	0
30-40	10	35	+1	+10
40-50	8	45	+2	+16
	$\sum f = 50$			$\sum fd = 5$

Now $$\overline{X} = A + \frac{\sum fd}{N} \times h$$

$$\overline{X} = 25 + \frac{5}{50} \times 10$$

$$= 26$$

E4) (1) Here n = 7, after arranging in ascending order, we get

2, 3, 6, 8, 10, 12, 15 and the Median value will be

$$\text{Median} = \text{value of } \left(\frac{n+1}{2}\right)^{th} \text{ item}$$

$$= \text{value of } 4^{th} \text{ item} = 8$$

(2) Here n = 8, so arranging in ascending order we get the values as 2, 3, 5, 6, 8, 10, 12, 15 and therefore

$$M_d = \frac{\left(\frac{n}{2}\right)^{th} \text{observation} + \left(\frac{n}{2}+1\right)^{th} \text{observation}}{2}$$

$$= \frac{4^{th} \text{ Observation} + 5^{th} \text{ Observation}}{2}$$

$$M_d = \frac{6+8}{2} = 7$$

E5)

Marks	5	10	15	20	25	30	35	40	45	50
No. of students	2	5	10	14	16	20	13	9	7	4
Cumulative Frequency	2	7	17	31	47	67	80	89	96	100

Now, Median = Value of the variable corresponding to the $\left(\frac{N}{2}\right)^{th}$ cumulative frequency

= Value of the variable corresponding to the $\left(\frac{100}{2}\right)^{th}$ cumulative frequency

= Value of the variable corresponding to the 50^{th} cumulative frequency,

Since 50 is not among cumulative frequency so the next cumulative frequency is 67 and the value of variable against 67 is 30. Therefore 30 is the median.

E6) First we shall calculate the cumulative frequency distribution

Marks	**f**	**Cumulative Frequency**
0-10	5	5
10-20	10	15
20-30	15	30 = C
30-40	**20 = f**	**50**
40-50	12	62
50-60	10	72
60-70	8	80
	N= 80	

Here $\frac{N}{2} = \frac{80}{2} = 40$,

Since, 40 is not in the cumulative frequency so, the class corresponding to the next cumulative frequency 50 is median class. Thus 30-40 is median class.

$$\text{Median} = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

$$= 30 + \frac{40 - 30}{20} \times 10$$

$$= 35$$

E7) First we shall calculate the cumulative frequency distribution

Class	**f**	**Cumulative Frequency**
0-20	5	5
20-40	15	20
40-60	30	50
60-80	f_4	$50+f_4$
80-100	8	$58+f_4$

Median class is 40-60 since median value is given 50,

$$\text{Median} = L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h$$

$$50 = 40 + \frac{\left(\frac{58+f_4}{2} - 20\right)}{30} \times 20$$

$$50 - 40 = \frac{58 + f_4 - 40}{3}$$

$$18 + f_4 = 30$$

$$f_4 = 30 - 18 = 12$$

E8) First we shall form the frequency distribution

Size	1	2	3	4	5	6	7	8
Frequency	1	2	1	2	2	2	4	1

From the frequency distribution we see that size 7 occurs with maximum frequency of 4 hence mode is 7.

E9) First we shall form the frequency distribution

Marks	f
0-10	5
10-20	10
2-30	15
30-40	20
40-50	12
50-60	10
60-70	8

Here mode class is 30-40 corresponding to the highest frequency 20.

Mode $$M_0 = L + \frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)} \times h$$

$$= 30 + \frac{20-15}{40-15-12} \times 10$$

$$= 30 + \frac{5}{13} \times 10$$

$$= 30 + 3.84 = 33.84$$

E10) We have

Mode = 3 Median – 2 Mean

35.4 = 3 Median – 2 (38.6)

3 Median = 35.4 + 77.2

Median = 112.6/3 = 37.53

E11) We have

$$GM = \sqrt[3]{4.8.16}$$

$$\Rightarrow \sqrt[3]{2^2.2.4.4^2}$$

$$\Rightarrow \sqrt[3]{2^3.4^3}$$

$$\Rightarrow 2 \times 4 = 8$$

E12) First we shall form the frequency distribution

x	log x
5	0.6990
15	1.1761
25	1.3979
35	1.5441
	$\sum \log x = 4.7971$

$$\text{Now GM} = \text{antilog}\left[\frac{\sum \log x}{n}\right]$$

$$= \text{antilog}\left[\frac{4.7971}{4}\right]$$

$$= \text{antilog}\ (1.1993) = 15.82$$

E13) We have the following frequency distribution:

Class	**f**	**x**	**log x**	**flog x**
0-10	12	5	0.6690	8.0280
10-20	15	15	1.1761	17.6415
20-30	25	25	1.3979	34.9475
30-40	18	35	1.5441	27.7938
40-50	10	45	1.6532	16.5320
	80			$\sum f_i \log x_i = 104.9428$

$$\text{GM} = \text{antilog}\left[\frac{\sum f \log x}{N}\right]$$

$$= \text{antilog}\left[\frac{104.9428}{80}\right]$$

$$= \text{antilog}\ (1.3118) = 20.52$$

E14) We have the formulae of the Harmonic mean

$$\text{HM} = \frac{1}{\frac{1}{n}\left[\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}\right]}$$

By putting the given values

$$\text{HM} = \frac{1}{\frac{1}{3}\left[\frac{1}{2} + \frac{1}{4} + \frac{1}{6}\right]}$$

$$= \frac{3 \times 12}{[6 + 3 + 2]}$$

$$= \frac{3 \times 12}{11} = 3.27$$

E15) We have given the following frequency distribution:

Class	f	x	f/x
0-10	3	5	0.600
10-20	5	15	0.333
20-30	10	25	0.400
30-40	8	35	0.225
40-50	4	45	0.088
	N = 30		$\sum \frac{f}{x} = 1.646$

Therefore by putting the values in formulae, we get

$$HM = \frac{N}{\sum_{i=1}^{5} \frac{f_i}{x_i}}$$

$$= 30 / 1.646 = 19.22$$

E16) First we shall calculate the cumulative frequency distribution

Marks	f	Cumulative Frequency
0-10	5	5
10-20	10	15
20-30	15	30
30-40	20	50
40-50	12	62
50-60	10	72
60-70	8	80
	N= 80	

$$\text{Here } \frac{N}{2} = \frac{80}{2} = 40,$$

Here, N/4 = 80/4 = 20. The 20th observation falls in the class 20-30. So, this is the first quartile class. 3N/4 = 3×80/4 = 60th observation falls in class 40-50, so it is the third quartile class.

For first quartile L = 20, f = 15, C = 15, N = 80

$$Q_1 = 20 + \frac{(20-15)}{15} \times 10 = 23.33$$

For third quartile L = 40, f = 12, C = 50

$$Q_3 = 40 + \frac{(60-50)}{12} \times 10 = 48.33$$