4

Measures of Central Tendency

MEANING OF THE MEASURES OF CENTRAL TENDENCY

If we take the achievement scores of the students of a class and arrange them in a frequency distribution, we may sometimes find that there are very few students who either score very high or very low. The marks of most of the students lie somewhere between the highest and the lowest scores of the whole class. This tendency of a group about distribution is named as central tendency and the typical score lying between the extremes and shared by most of the students is referred to as a measure of central tendency. In this way a measure of central tendency as Tate (1955) defines is:

> a sort of average or typical value of the items in the series and its function is to summarize the series in terms of this average value.

The most common measures of central tendency are:

1. Arithmetic mean or mean
2. Median
3. Mode

Each of them, in its own way, can be called a representative of the characteristics of the whole group and thus the performance of the group as a whole can be described by the single value which each of these measures gives. The values of mean, median or mode also help us in comparing two or more groups or frequency distributions in terms of typical or characteristic performance.

Arithmetic Mean *(M)*

This is the simplest but most useful measure of central tendency. It is nothing but the 'average' which we compute in our high school arithmetic and, therefore, can be easily defined as the sum of all the values of the items in a series divided by the number of items. It is represented by the symbol *M*.

Calculation of mean in the case of ungrouped data. Let X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9 and X_{10} be the scores obtained by 10 students on an achievement test. Then the arithmetic mean or mean score of the group of these ten students can be calculated as:

$$M = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + \ldots + X_{10}}{10}$$

The formula for calculating the mean of an ungrouped data is

$$M = \frac{\Sigma X}{N}$$

where ΣX stands for the sum of scores or values of the items and N for the total number of items in a series or group.

Calculation of mean in the case of grouped data (Data in the form of frequency distribution)

General method. In a frequency distribution where all the frequencies, are greater than one, the mean is calculated by the formula:

$$M = \frac{\Sigma fX}{N}$$

where X represents the mid-point of the class interval, f its respective frequency, and N the total of all frequencies.

We can illustrate the use of this formula by taking the frequency distribution given in Table 2.1 of Chapter 2 as follows:

Example 4.1

Scores	*f*	*Mid-point* (*X*)	*fX*
65–69	1	67	67
60–64	3	62	186
55–59	4	57	228
50–54	7	52	364
45–49	9	47	423
40–44	11	42	462
35–39	8	37	296
30–34	4	32	128
25–29	2	27	54
20–24	1	22	22
	$N = 50$		$\Sigma fX = 2230$

$$M = \frac{\Sigma fX}{N} = \frac{2230}{50} = 44.6$$

Short-cut method of computing the mean of grouped data Mean for the grouped data can be computed easily with the help of the following formula:

$$M = A + \frac{\Sigma fx'}{N} \times i$$

where

A = Assumed mean
i = Class interval
f = Respective frequency of the mid-values of the class intervals
N = Total frequency

$x' = \frac{X - A}{i}$ (The quotient obtained after division of the difference between the mid-value of the class and assumed mean by i, the class interval.)

The use of this formula can be easily understood through the following illustration:

Example 4.2: Assumed mean (A) = 42.

Scores	*f*	*X (Mid-point)*	$x' = (X - A)/i$	fx'
65–69	1	67	5	5
60–64	3	62	4	12
55–59	4	57	3	12
50–54	7	52	2	14
45–49	9	47	1	9
40–44	11	42	0	0
35–39	8	37	−1	−8
30–34	4	32	−2	−8
25–29	2	27	−3	−6
20–24	1	22	−4	−4
	N = 50			$\Sigma fx'$ = 26

Computation

$$M = A + \frac{\Sigma fx'}{N} \times i = 42 + \frac{26}{50} \times 5$$

$$= 42 + \frac{26}{10} = 42 + 2.6 = 44.6$$

Median (M_d)

If the items of a series are arranged in ascending or descending order of magnitude, the measure or value of the central item in the series is is termed as *median*. We may thus say that the median of a distribution is the point on the score scale below which half (or 50%) of the scores fall. Thus, median is the score or the value of that central item which divides the series into two equal parts. It should therefore be understood that the central item itself is not the median. It is only the measure or value of the central item that is known as the median. For example, if we arrange in ascending or descending order the marks of 5 students, then the marks obtained by the 3rd student from either side will be termed as the median of the scores of the group of students under consideration.

Computation of median for ungrouped data. The following two situations could arise:

1. *When N (No. of items in a series) is odd.* In this case where N, is odd (not divisible by 2), the median can be computed by the formula M_d = the measure or value of the $(N+1)/2$-th item.

Example 4.3: Let the scores obtained by 7 students on an achievement test be 17, 47, 15, 35, 25, 39, 44. Then, first of all, for calculating median, we have to arrange the scores in ascending or descending order: 15, 17, 25, 35, 39, 44, 47. Here N (= 7) is odd, and therefore, the score of the $(N + 1)/2$-th item or 4th student, i.e. 35 will be the median of the given scores.

2. *When N (No. of items in a series) is even.* In the case where N is even (divisible by 2), the median is determined by the following formula:

$$M_d = \frac{\text{the value of } (N/2)\text{th item} + \text{the value of } [(N/2)+1]\text{th item}}{2}$$

Example 4.4: Let there be a group of 8 students whose scores in a test are, 17, 47, 15, 35, 25, 39, 50, 44. For calculating mean of these scores we proceed as follows:

Arrangement of scores in ascending series: 15, 17, 25, 35, 39, 44, 47, 50

The score of the $(N/2)$th, i.e. 4th student = 35

The score of the $[(N/2) + 1]$th, i.e. 5th student = 39

Then,

$$\text{Median} = \frac{35+39}{2} = 37$$

Computation of median for grouped data (in the form of frequency distribution). If the data is available in the form of a frequency distribution like below, then calculation of median first requires the location of median class.

Example 4.5

Scores	f
65–69	1
60–64	3
55–59	4
50–54	7
45–49	9
40–44	11
35–39	8
30–34	4
25–29	2
20–24	1
	$N = 50$

Actually, as defined earlier, median is the measure or score of the central item. Therefore, it is needed to locate the central item. It may be done through the formulae given earlier in the case of ungrouped data for the odd and even values of N (total frequencies). Here, in the present distribution, N (= 50) is even. Therefore, median will fall somewhere between the scores of 25th and 26th items in the given distribution. In the given frequency distribution table, if we add frequencies either from above or below we may see that the class interval designated as 40–44 is to be labelled as the class where the score representing median will fall.

After estimating the median class, the median of the distribution may be interpolated with the help of following formula:

$$M_d = L + \left[\frac{(N/2) - F}{f}\right] \times i$$

where

L = Exact lower limit of the median class
F = Total of all frequencies before in the median class
f = Frequency of the median class
i = Class interval
N = Total of all the frequencies

By applying the above formula, we can compute the median of the given distribution in the following way:

$$M_d = 39.5 + \frac{(50/2) - 15}{11} \times 5 = 39.5 + \frac{10}{11} \times 5$$

$$= 39.5 + \frac{50}{11} = 39.5 + 4.55 = 44.05$$

Example 4.6: Some Special Situations in the Computation of Median

(a)		(b)		(c)	
Scores	*f*	*Scores*	*f*	*Scores*	*f*
55–59	5	45–49	2	20–21	2
50–54	3	40–44	5	18–19	1
45–49	8	35–39	6	16–17	0
40–44	18	30–34	0	14–15	0
35–39	15	25–29	8	12–13	2
30–34	10	20–24	3	10–11	0
25–29	7	15–19	2	8–9	0
20--24	2			6–7	2
				4–5	1
				2–3	1
				0–1	1
	$N = 68$		$N = 26$		$N = 10$

Let us analyse the medians of the above distributions.

1. We know by definition that median is the point on the score scale below and above which 50% cases lie. Thus the score representing median should be a common score falling in the classes 35–39 and 40–44. This score is nothing but the upper limit of class 35–39 which is also the lower limit of the class 40–44. Therefore, in this case the median is 39.5.
2. In the second distribution, if we try to add the frequencies from below, we see that upto class interval 25–29, 13 cases lie and by adding frequencies from above, we also find that upto the class interval 35–39, 13 cases lie. In this way, the class interval 30–34 divides the distribution into two equal parts below and above which 50% cases lie. It leads us to conclude that median should be the mid-point of the class interval 30–34 and, therefore, 32 is the median of this distribution.
3. In the third case, if we add the frequencies from below, we find that 5 cases lie upto the class interval 6–7. By adding the frequencies from above, we also find that, upto class 12–13, 5 cases lie. The median should then fall mid-way between the two classes 8–9 and 10–11. It should be the common score represented by both these classes. This score is nothing but the upper limit of the class 8–9 and lower limit of the class 10–11 and, therefore, it should be 9.5.

Mode (M_o)

Mode is defined as the size of a variable (say a score) which occurs most frequently. It is the point on the score scale that corresponds to the maximum frequency of the distribution. In any series, it is the value of the item which is most characteristic or common and is usually repeated the maximum number of times.

Computation of mode for ungrouped data. Mode can be easily be computed merely by looking at the data. All that one has to do is to find out the score which is repeated maximum number of times.

Example 4.7: Suppose we have to find out the value of the mode from the following scores of students:

25, 29, 24, 25, 27, 25, 28, 25, 29

Here the score 25 is repeated maximum number of times and thus, value of the mode in this case is 25.

Computation of mode for grouped data. When data is available in the form of frequency distribution, the mode is computed from the following formula:

$$\text{Mode } (M_o) = 3M_d - 2M$$

where M_d is the median and M is the mean of the given distribution. The mean as well as the median of the distribution are first computed and then, with the help of the above formula, mode is computed. For illustration, we can take the distribution given in Table 2.1.

Example 4.8: The mean and median of this distribution have already been computed as 44.6 and 44.05 respectively.

Therefore,

$$M = 44.6, \quad M_d = 44.05$$
$$M_o = 3 \times 44.05 - 2 \times 44.6$$
$$= 132.15 - 89.2 = 42.95$$

Another method for grouped data. Mode can be computed directly from the frequency distribution table without calculating mean and median. For this purpose, we can use the following formula:

$$M_o = L + \frac{f_1}{f_1 + f_{-1}} \times i$$

where

L = Lower limit of the model class (the class in which mode may be supposed to lie)

i = Class interval

f_1 = Frequency of the class adjacent to the modal class for which lower limit is greater than that for the modal class

f_{-1} = Frequency of the class adjacent to the modal class for which the lower limit is less than that for the modal class

Let us illustrate the use of the above formula by taking frequency distribution given in Table 2.1.

Example 4.9:

Scores	*f*
65–69	1
60–64	3
55–59	4
50–54	7
45–49	9
40–44	11
35–39	8
30–34	4
25–29	2
20–24	1
	$N = 50$

$$M_o = L + \frac{f_1}{f_1 + f_{-1}} \times i$$

The crude mode in this distribution may be supposed to lie within the class interval 40–44. Hence,

L = 39.5

f_1 = 9

f_{-1} = 8

Therefore,

$$M_o = 39.5 + \frac{9}{9+8} \times 5 = 39.5 + \frac{45}{17} = 39.5 + 2.65 = 42.15$$

For computing mode in the case of grouped data (frequency distribution), we can use any of these two methods. The first method is more useful in the case when we need the computation of mean and median. Otherwise, the use of the second method is convenient. However, it may also be observed that the mode values computed from these two methods are not identical. The value from the first method is always a little higher than the value obtained through the use of second (direct) method.

Computation of Median and Mode from the Curves of Frequency Distribution

Median. It can be computed directly from the following frequency curves with the help of simple observations and measurements:

From frequency graph. The value on the x-axis at the foot of the ordinate which bisects the area between the frequency graph and the x-axis is the median.

For cumulative frequency graph. A horizontal line is to be drawn from the middle point of the ordinate which represents the total frequency. The foot of the perpendicular from the point of intersection of this line and cumulative frequency graph gives the value of the median.

From cumulative frequency percentage curve or ogive. Here, for obtaining the median, a line parallel to the x-axis is drawn from the highest cumulative frequency shown on the y-axis. Now, from the point where this line cuts the curve, a perpendicular to the x-axis is drawn. The score located on the x-axis through this perpendicular gives the measure of the median.

Mode. It can be computed easily from the following frequency curves:

From frequency graph. For this purpose, a smoothed frequency graph is to be constructed from the given data. The abscissa of the highest point on this curve is taken to be the mode.

From histogram. For this purpose, the highest rectangle of the diagram is selected and at the top, the mid-point is marked. From this point, a perpendicular is drawn on the x-axis to indicate the measure of mode.

From cumulative frequency graph. In this curve, the abscissa of the point where the curve is steepest indicates the measure of the mode.

When to Use the Mean, Median and Mode

Computation of any of the three—mean, median and mode—provides a measure of central tendency. Which of these should be computed in a particular situation is a question that can be raised quite often. This can be answered in light of the characteristics and nature of all these measures.

When to use the mean

1. Mean is the most reliable and accurate measure of the central tendency of a distribution in comparison to median and mode. It has the greatest stability as there are less fluctuations

in the means of the samples drawn from the same population. Therefore, when a reliable and accurate measure of central tendency is needed, we compute the mean for the given data.

2. Mean can be given an algebraic treatment and is better suited to further arithmetical computation. Hence, it can be easily employed for the computation of various statistics like standard deviation, coefficient of correlation, etc. Therefore, when we need to compute more statistics like these, mean is computed for the given data.
3. In computation of the mean, we give equal weightage to every item in the series. Therefore, it is affected by the value of each item in that series. Sometimes there are extreme items which seriously affect the position of the mean. Thus, it is not proper to compute mean for the series that has extreme items and each score carries equal weight in determining the central tendency.

When to use the median

1. The median is the exact mid-point of a series, below and above which 50% of the cases lie. Therefore, when the exact mid-point of the distribution is desired, median is to be computed.
2. The median is not affected by the extreme scores in the series. Therefore, when a series contains extreme scores, the median is perhaps the most representative central measure.
3. In the case of an open end distribution (incomplete distribution "80 and above" or "20 and below"), it is impossible to calculate the mean, and hence the median is the most reliable measure that can be computed.
4. Mean cannot be calculated graphically. But in the case of median, we can compute it graphically. Therefore, when we have suitable graphs like frequency curve, polygon, ogive, and so on, we should try to compute median.
5. Median is specifically useful for the data of the items which cannot be precisely measured in quantities as for e.g. health, culture, honesty, intelligence and so on.

When to use the mode

1. In many cases, the crude mode can be computed by just having a look at the data. It gives the quickest, although approximate measure of central tendency. Therefore, in the cases where a quick and approximate measure of central tendency is all that is desired, we compute mode.

2. Mode is that value of the item which occurs most frequently or is repeated maximum number of times in a given series. Therefore, when we need to know the most often recurring score or value of the item in a series, we compute mode. On account of this characteristic, mode has unique importance in the large scale manufacturing of consumer goods. In finding the sizes of shoes and readymade garments which fit most men or women, the manufacturers make use of the average indicated by mode.
3. Mode can be computed from the histogram and other frequency curves. Therefore, when we already have a graphical representation of the distribution in the form of such figures, it is appropriate to compute mode instead of mean.

SUMMARY

The statistics, mean, median and mode, are known to be the most common measures of central tendency. A measure of central tendency is a sort of average or a typical value of the item in the series or some characteristics of the members of a group. Each of these measures of the central tendency provides a single value to represent the characteristics of the whole group in its own way.

Mean represents the 'average' for an ungrouped data, the sum of the scores divided by the total number of scores gives the value of the mean. The mean, in the case of a grouped data is best computed with the help of a short-cut method using the following formula:

$$M = A + \frac{\Sigma fx'}{N} \times i,$$

where A is the assumed mean,

$$x' = \frac{X - A}{i},$$

X is the mid-point of the class interval, i the class interval, and N the total frequencies of the distribution. Median is the score or value of that central item which divides the series into two equal parts. Hence when N is odd $(N + 1)/2$-th measure, and when N is even, the average of $(N/2)$-th and $[(N/2) + 1]$-th item's measure provides the value of the median. In the case of grouped data, it is computed by the formula

$$M_d = L + \frac{(N/2) - F}{f} \times i,$$

where L represents the lower limit of the median class, F, the total of all frequencies before the median class, and f, the frequency of the median class. The median can also be computed directly from curves like the frequency graph, cumulative frequency graph and ogive.

Mode is defined as the size of the variable which occurs most frequently. Crude mode can be computed by just having a look at the data. In the case of grouped data, it may be computed with the help of the formula, $M_o = 3M_d - 2M$. It can also be computed directly from the histogram and other frequency curves as also with the help of the following formula without first calculating mean and median:

$$M_o = L + \frac{f_1}{f_1 + f_{-1}} \times i$$

where f_1 and f_{-1} are the frequencies of the classes adjacent to the modal class for which lower limits are greater or less than those for the modal class.

Preference for the use of mean, median or mode can be understood from the following tabular matter:

Mean	*Median*	*Mode*
If we have	If we have	If we have
(i) to get a reliable and accurate measure of central tendency,	(i) to get an exact mid-point of the distribution,	(i) to get a quick and approximate measure of central tendency,
(ii) to compute further statistics like standard deviation, coefficient of correlation,	(ii) a series containing extreme scores,	(ii) to know the most often recurring score or value of the item in a series,
(iii) been given a series having no extreme items.	(iii) a distribution with the open end,	(iii) appropriate graphical representation of the data.
	(iv) appropriate graphical representation of data.	

EXERCISES

1. What do you understand by the term, 'measures of central tendency'? Point out the most common measures of central tendency.

2. What is an arithmetic mean? How can it be computed in the case of ungrouped as well as grouped data? Illustrate with the help of hypothetical data.
3. Define median. How can it be computed in the case of ungrouped as well as grouped data? Illustrate with examples.
4. What do you understand by the term mode of a data? Point out the methods of its computation in the case of grouped as well as ungrouped data. Explain the computation process through examples.
5. Define mean, median and mode. Discuss when each one of them should be computed and why.
6. In the situations described below, what measures of central tendency would you like to compute?

   (a) The average achievement of a group.
   (b) The most popular fashion of the day.
   (c) Determining the mid-point of the scores of a group in an entrance examination.
7. Find the mean IQ for the eight students whose individual IQ scores are:

   80, 100, 105, 90, 112, 115, 110, 120.
8. Compute median for the following data:

   (a) 8, 3, 10, 5, 2, 11, 14, 12.
   (b) 72, 74, 77, 53, 58, 63, 66, 82, 89, 69, 71.
9. Find the crude mode:

   (a) 15, 14, 8, 14, 14, 11, 9, 9, 11.
   (b) 3, 3, 4, 4, 4, 5, 7, 7, 9, 12.
10. Compute mean and median from the achievement scores of 25 students as given in the following (taking 2 as class interval):

    72, 75, 77, 67, 72, 81, 78, 65, 86, 83, 67, 82, 76, 76, 59, 70, 83, 71, 62, 72, 72, 61, 67, 68, 64.
11. Find the mean, median and mode for the following sets of scores:

    (a) 24, 18, 19, 12, 23, 20, 21, 22.
    (b) 20, 14, 12, 14, 19, 14, 18, 14.
    (c) 24, 18, 19, 20, 22, 25, 23, 12.
    (d) 9, 14, 8, 13, 10, 10, 11, 12, 10.
12. Compute the mean, median and mode for the following frequency distributions:

(a) Scores	f	(b) Scores	f	(c) Scores	f	(d) Scores	f
70–71	2	120–122	2	45–49	2	135–144	1
68–69	2	117–119	2	40–44	3	124–134	2
66–67	3	114–116	2	35–39	2	115–124	8
64–65	4	111–113	4	30–34	17	105–114	22
62–63	6	108–110	5	25–29	30	95–104	33
60–61	7	105–107	9	20–24	25	85–94	22
58–59	5	102–104	6	15–19	15	75–84	9
56–57	1	99–101	3	10–14	3	65–74	2
54–55	2	96–98	4	5–9	2	55–64	1
52–53	3	93–95	2	0–4	1		
50–51	1	90–92	1				
	N = 36		N = 40		N = 100		N = 100

13. Compute the mean, median and mode for the following frequency distributions:

(a) Scores	f	(b) Scores	f	(c) Scores	f	(d) Scores	f
100–104	1	57–59	1	58–60	1	90–94	1
95–99	2	54–56	1	55–57	1	85–89	4
90–94	1	51–53	5	52–54	1	80–84	2
85–89	6	48–50	9	49–51	2	75–79	8
80–84	7	45–47	5	46–48	2	70–74	9
75–79	3	42–44	8	43–45	3	65–69	14
70–74	2	39–41	10	40–42	3	60–64	6
65–69	1	36–38	6	37–39	7	55–59	6
60–64	2	33–35	4	34–36	8	50–54	4
55–59	4	30–32	7	31–33	7	45–49	3
50–54	0	27–29	0	28–30	5	40–44	3
45–49	1	24–26	1	25–27	4		
				22–24	3		
				19–21	3		
				16–18	2		
	N = 30		N = 57		N = 52		N = 60