BIOSTATISTICS
UNIT – II : MEASURES OF CENTRAL TENDENCY

Dr. KAMLESH BISHT

Academic Consultant

Department of Mathematics

Uttarakhand Open University, Haldwani

Content

- ***Objective.***
- ***Introduction.***
- ***Mean or Arithmetic mean.***
- ***Merits, Demerits and uses of Mean.***
- ***Median.***
- ***Merits, Demerits and uses of Median.***
- ***Mode.***
- ***Merits, Demerits and uses of Mode.***
- ***Summary.***
- ***References.***

OBJECTIVES

After study this unit, you will be able:

- *To know about the measure of central tendency- mean, median and mode.*
- *To know the merits and demerits and uses of these measures.*
- *To know about different methods of measuring mean, median and mode.*
- *To know the situation where which measures is better to use?*
- *To know the advantages of short cut methods of computing mean.*

INTRODUCTION

Measures of central tendency as the name suggests are numerical measurements of the central part of the distribution. Measures of central tendency are also called averages or measures of location because they show the location of the centre of the distribution from which the data were sampled.

According to professor Bowley, averages are , " Statistics constants which enable us to comprehend in a single effort the significance of the whole." In other words, these are numbers tell us where the majority of values in the distribution located. For example the average marks in a distribution of marks of all the students of a class. The averages which are commonly used in biostatistics are as follows:

1. *Mean or arithmetic mean*
2. *Median*
3. *Mode*

MEAN 'OR' ARITHMETIC MEAN

Mean or arithmetic mean of a series of data is the ratio of the sum of the observation to the number of observation. If X_1 , X_2 ,..., X_n are the observation of a series then their arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + ...x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1)$$

And if the corresponding frequencies, $f_1, f_2, ...f_n$ of the variables $x_1, x_2, ...x_n$ are given, then the arithmetic mean is defined as ratio in which the numerator is the sum of products of the variables with their frequencies and denominator is the sum of the frequencies.

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + ...f_n x_n}{\sum f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N} \qquad (2)$$

where, $N = \sum f_i$ = sum of frequencies.

MEAN OF INDIVIDUAL ITEM

Mean of individual items is given by the ratio of the sum of items to the number of items as given in formula (1).

Example 1. Find the arithmetic mean of triglycerides present 10 patients in their blood samples in a hospitalas:

25, 30, 21, 55, 47, 10, 15, 17, 45, 35

Solution. Let $\bar{x}$ be the average triglyceride value and since these are individual items, their mean can be computed by formula

$$\bar{x} = \frac{x_1 + x_2 + \ldots . x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \frac{25+30+21+55+47+10+15+17+45+35}{10} = \frac{300}{10} = 30$$

MEAN IN DIFFERENT FREQUENCY DISTRIBUTION

If $x_1, x_2, \ldots\ldots x_n$ are the observations in a discrete distribution according to some characteristic and $f_1, f_2, \ldots\ldots f_n$ be their corresponding frequencies then the arithmetic mean is given by the formula (2). The computation procedure for mean can be easily understood with the help of the example given below.

Example 2. The distribution of marks of 50 students of B.Sc. class in a botany semester examination is given below. Find the average of marks.

Marks (x)	12	23	25	35	45	15	40
Frequency(f)	3	10	12	10	2	8	5

Solution. Since this is a discrete distribution so the average of marks is given by the formula (2).

For the computation of average marks we prepare the following table:

Table1. For calculation of Mean in Discrete Distribution

Marks (x)	Frequency (f)	f x
12	3	36
23	10	230
25	12	300
35	10	350
45	2	90
15	8	120
40	5	200
Total	$\sum f = 50$	$\sum fx = 1326$

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots\ldots f_n x_n}{\sum f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N} = \frac{1326}{50} = 26.52$$

This is clear that it is not necessary that average will be a number presenting in the data and also it is not an integer value while the marks in integers.

MEAN IN CONTINUOUS DISTRIBUTION

In continuous distribution, there are given class interval and their corresponding frequencies. First of all we find mid values of these classes and treat them as the variable values. Now we apply the formula (2) for the calculation of arithmetic mean.

The procedure will be clear from the following example.

Example 3. For the data given in the below Table 2 on systolic BP of 68 patients, calculate the arithmetic mean.

Table 2.

Systolic BP (mmHG)	Frequency (f)	Systolic BP (mmHG)	Frequency (f)
90-100	3	140-150	11
100-110	5	150-160	9
110-120	7	160-170	6
120-130	10	170-180	2
130-140	15		

Solution. For the calculation of mean we prepare the following table:

Table3. For calculation of Mean in Continuous Distribution

Systolic BP (mmHg)	Frequency (f)	Mid Value (x)	Fx
90-100	3	95	285
100-110	5	105	525
110-120	7	115	805
120-130	10	125	1250
130-140	15	135	2025
140-150	11	145	1595

150-160	9	155	1395
160-170	6	165	990
170-180	2	175	350
Total	$\sum f = 68$		$\sum fx = 9220$

$$\bar{x} == \frac{\sum_{i=1}^{n} f_i x_i}{N} = \frac{9220}{68} = 135.6 mmHg$$

SHORT-CUT METHOD FOR MEAN

For the computation of mean short –cut method is applied when the variable values and their frequencies are large. To make the computations easy we take a middle value in the given values of x as assumed mean and subtract this assumed mean from all the values of x. This assumed mean is also called provisional mean. Then the formula for the calculation of arithmetic mean is given by as follows:

$$\bar{x} = A + \frac{\sum d}{n} \qquad (3)$$

where A= assumed mean, n= number of observations in the given data

d= x- A= deviation of all the variate values from assumed mean A.

Steps of computation for short- cut method:

Step 1. Take any observation (generally, middle value if we arrange the values in ascending or descending order of magnitude) of the individual series as assumed mean A.

Step 2. Find the deviation of the values of variate x from assumed mean A, i.e., calculate the differences d= x- A

Step 3. Find the sum of d and use above formula (3), we find the value of mean.

If the frequencies corresponding to the variate values are given, then we use the formula for mean as follows:

$$\bar{x} = A + \frac{\sum fd}{N} \qquad (4)$$

where, $N = \sum f$ = sum of frequencies. Here we find the product of f and d.

If the data is continuous, we find the mid values as x and then d= x- A. Now apply the above formula (4). The procedure will be clear from the examples as give below.

Example 4. The marks of the 7 students of a class in a test are as given below:

12, 15, 22, 25, 35, 40, 45

Find the mean by short-cut method.

Solution. Let us take assumed mean A=25. Now we prepare the table for the computation of mean as given below:

Table-4 Mean for individual data by short cut method

X	d= x-25
12	-13
15	-10
22	-3
25	0
35	10
40	15
45	20
Total	$\Sigma d = 19$

$$\text{Arithmetic mean } \bar{x} = A + \frac{\sum d}{n} = 25 + \frac{19}{7} = 25 + 2.71 = 27.71$$

Thus the average of marks of the given 7 students of the class is 27.71

EXAMPLE 5 FOR SELF ASSESSMENT OF THE STUDENTS (DO YOURSELF)

Example 5. Ten patients were examined for uric acid test. The operation was performed 1050 times and the frequencies so obtained for different number of patients (x) are shown in the table given below. Compute the arithmetic mean by short- cut method.

x:	0	1	2	3	4	5	6	7	8	9	10
f:	2	8	43	133	207	260	213	120	54	9	1

STEP DEVIATION METHOD OF MEAN

It can be used in grouped data. When all the classes are of equal width (say h), in continuous data and the values of x are at equal interval in discrete grouped data then the we may simplify the calculations by taking d= (x- A)/ h in short-cut method. Now the formula for the calculation of mean becomes.

$$\bar{x} = A + \frac{\sum fd}{N} \times h$$

Here, the symbols have the same meaning as in short-cut method above and h is the gap between the two values of x or class interval.

Example 6. Find the mean by step deviation method for the data of blood pressure of 68 patients as given in the following table.

BP(mmHg) (x)	90	100	110	120	130	140	150	160	170
Frequency (f)	3	5	7	10	15	11	9	6	2

Solution. We take assumed mean A= 130 and here interval between any two values of x is 10, i.e., h= 10. Now prepare the table for the computation of mean.

Table6. Step Deviation Method of Mean in discrete grouped data

BP (mmHg)	Frequency (f)	$d = \frac{x-130}{10}$	fd
90	3	-4	-12
100	5	-3	-15

110	7	-2	-14
120	10	-1	-10
130	15	0	0
140	11	1	11
150	9	2	18
160	6	3	18
170	2	4	8
Total	$\sum f = 68$		$\sum fd = 4$

$$\text{Arithmetic mean } \bar{x} = A + \frac{\sum fd}{N} \times h = 130 + \frac{4}{68} \times 10 = 130 + 0.588 = 130.588 mmHg$$

Thus the average for BP is 130.588mmHg.

WEIGHTED MEAN

In computation of arithmetic mean some items are more important than the others, in such cases the weightage should be given to the items according to their importance. For example if we want to have an idea of the change in cost of living of group of people of a certain locality, then the simple mean of the prices of the commodities consumed by them will not do, since all the commodities are not equally important, e.g., wheat, rice and pulses are more important than cigarettes, tea, confectionery, etc.

If $x_1, x_2, \ldots\ldots x_n$ are the variate values of a distribution and $w_1, w_2, \ldots. w_n$ be their corresponding weights then weighted mean is give by:

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

Example 9. The following table gives the platelets count (in lakh/cmm) from the analysis of the blood samples on five different days in a pathology laboratory. Find the average platelets count per patient.

Day	1	2	3	4	5
Platelates Count (in lakh/cmm) (w)	0.50	0.75	1.00	1.50	2.00
No. of Patient (X)	65	80	95	90	70

Solution. The table for the calculation of weighted mean is given by:

Table 9. Table for Weighted Mean

Platelets count (x)	No. of patients (w)	Wx
0.50	65	32.5
0.75	80	60.0
1.00	95	95.0
1.50	90	135
2.00	70	140
Total	$\sum w = 400$	$\sum wx = 462.5$

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{462.5}{400} = 1.156$$

Thus, the average platelets per patient are 1.156 lakh/cmm.

COMBINED MEAN

If $\bar{x}_1, \bar{x}_2, \ldots\ldots \bar{x}_m$ are the means of m series of sizes $n_1, n_2, \ldots . n_m$ respectively, then their combined arithmetic meaning $\bar{x}$ is given by:

$$\bar{x} = \frac{\sum n_i \bar{x}_i}{\sum n_i}; i = 1,2,\ldots\ldots m$$

Example 10. There are 40 male and 10 female employees in a firm. The mean salary of male employees is Rs.520 and that of female employees Rs. 420. Find the combined average salary of all the employees.

Solution. Here, $n_1 = 40, n_2 = 10, \bar{x}_1 = 520, \bar{x}_2 = 420$

$$\text{Combined mean } \bar{x} = \frac{\sum n_i \bar{x}_i}{\sum n_i} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} = \frac{520 \times 40 + 420 \times 10}{40 + 10} = \frac{25000}{50} = 500$$

Hence, the average salary of all the employees is Rs. 500.

CORRECTED MEAN

Some times there are problems of such type that we used wrong digits while the actual digits were different, then we replace the wrong digits with the correct digits and now we can get the correct mean. The procedure will be clear from the example given below.

Example 11. A student calculates the mean of 20 observations as 25.2. Later on he found that he misread one observation 34 in place of 43, find the correct mean.

Solution. We know the mean of individual series is given by:

$$\bar{x} = \frac{\sum x}{n} or \sum x = n\bar{x} = 20 \times 25.2 = 504$$

But he misread 43 as 34. So the correct total of x= 504-34+43=513.

So correct mean=513/ 20 =25.65

MERITS, DEMERITS AND USES OF MEAN

Merits:

1. Mean is rigidly defined.
2. It can be calculated easily by a non mathematical person also.
3. It is based upon all the observations.
4. Among all the averages, it is affected least by fluctuations of sampling.
5. It is the best measure to compare two or more series.
6. It is easily understandable.
7. It is the most widely used method of central tendency.

Demerits:

1. It is affected much by extreme values.

2. It cannot be calculated in case of open end classes.

3. It cannot be calculated in case of qualitative data such as intelligence, beauty, etc.

`4. In extremely asymmetrical distribution, mean is not a suitable measure of central tendency.

5. It cannot be calculated if any observation is missing.

6. It may lead to wrong conclusions if the details of the data are not given. For example the marks of two students in three successive tests are respectively 30, 40, 50 and 50, 40, 30. We see that average score of both the students is same, we can say that both students are of same level while first is improving and the second is deteriorating.

Uses of Mean:

1. It is very much used in practical situations.

2. A common man uses it for computing his monthly budget.

3. It is very much used in sampling and inference.

4. A businessman uses it for computing per unit profit, output per person, average expenditure and average profit per week or per month, etc.

MEDIAN

Median of a distribution is the middle most value of the variable if the values of the variable are arranged in ascending or descending order of their magnitude. The median divides the observations of the variable in such a way that half of the observations of the variable lie above the median and half below this. Median is thus called a positional average because it locates at the middle of the observations. But if the number of observations is even then after arrangement there will be two middle values and the median will be the average of these two middle values.

MEDIAN IN INDIVIDUAL SERIES

Arrange the data observations, (say n) in ascending or descending order of magnitude. Now there can be two cases:

Case 1. If n is odd then middle most, i.e., (n+1/2)th term value is the median.

Case2. If n is even then there are two middle terms (n/2)th and (n+1/2)th, then median is given by:

$$M_e = \frac{\left[\frac{n}{2}th + \left(\frac{n}{2}+1\right)th\right]term}{2}$$

Example 12. The marks of 9 children in a test exam are: 12, 23, 34, 11, 14, 15, 13, 16, 45.

Find the median of the marks.

Solution. Arrange the given observations in ascending order of magnitude, we get

`11, 12, 13, 14, 15, 16, 23, 34, 45

Here the number of observations n =9, i.e., odd.

So the median is the (9+1)/2 th, i.e., 5^{th} term value, i.e., 15.

Example 13. The number of blood LDL (in mg/dl) present the blood samples of 12 patients are: 5, 19, 42, 11, 50, 30, 21, 0, 22, 52, 36, 27

Find the median of the data.

Solution. On arranging the given observations in ascending order of magnitude, we get,

0, 5, 11, 19, 21, 22, 27, 30, 36, 42, 50, 52

Here number of observations = 12, i.e., even. So median is given by

$$M_e = \frac{\left[\frac{n}{2}th + \left(\frac{n}{2}+1\right)th\right]term}{2} = \frac{\left[\frac{12}{2}th + \left(\frac{12}{2}+1\right)th\right]term}{2}$$

$$= \frac{(6th+7th)term}{2} = \frac{22+27}{2} = \frac{49}{2} = 24.5$$

NOTE: *So median is 24.5 mg/dl which does not belong to the data. So in the case of even number of observation median is not present in the data observation.*

MEDIAN IN DISCRETE FREQUENCY DISTRIBUTION

If $x_1, x_2, ...x_n$ are the observations in a discrete distribution according to some characteristic and $f_1, f_2, ...f_n$ be their corresponding frequencies then for the calculation of median we calculate the cumulative frequencies. The median is calculated with the help of the following steps.

Working steps for median

Step1. Arrange the given values in ascending order of magnitude.

Step2. Find the total of frequencies, called cumulative frequency and denoted by c.f.

Step3. Find $\frac{N}{2}$, where N= $\sum f$.

Step4. Find cumulative frequency just greater than $\frac{N}{2}$. The value of x corresponding to this cumulative frequency is the required median.

Example14. Find the median for the following data.

X	21	15	17	9	5	7	8	10
F	2	5	3	4	5	1	6	12

Solution. For calculating the median we arrange the values of x in ascending order and then prepare the cumulative frequency table as follows:

Table10. Median in Discrete Distribution

x	**f**	**c.f.**
5	5	5
7	1	6

8 9 10 15 17 21	6 4 12 5 3 2	12 16 28 33 36 38
	$N = \sum f = 38$	

Here N/2 = 38/2= 19 and cumulative frequency just greater than 19 is 28. The value of x corresponding to cumulative frequency 28 is 10. So the median of the given data is 10.

MEDIAN IN CONTINUOUS FREQUENCY DISTRIBUTION

When the data is in class interval form, the class corresponding to c.f. just greater than N/2 is called the median class and the median is computed by the following formula:

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

Where L= lower limit of the median class

C= cumulative frequency just before the median class

N= Total of frequency

f = frequency of the median class

h = magnitude of the median class

Example15. The following table gives the distribution of weights of 100 persons. Find the median of this data.

Weight	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90
Frequency	1	3	6	10	15	25	15	10	11	4

Solution. For computing the median we prepare the following table:

Table11. Median in Continuous Distribution

Weight (in kg) (x)	Frequency (f)	Cumulative Frequency (c.f.)
40-45	1	1
45-50	3	4
50-55	6	10
55-60	10	20
60-65	15	35
65-70	25	60
70-75	15	75
75-80	10	85
80-85	11	96
85-90	4	100
Total	$N = \sum f = 100$	

Here N/2 = 10/2= 50 and cumulative frequency just greater than 50 is 60. The class corresponding to cumulative frequency 60 is 65-70. So class 65-70 is the median class. Now median is given by:

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

here L= lower limit of the median class= 65

C= cumulative frequency just before the median class=35

N= Total of frequency =100

f = frequency of the median class=25

h = magnitude of the median class=5

$$M_e = 65 + \frac{(50-35)}{25} \times 5 = 65 + \frac{15}{25} \times 5 = 65 + 3 = 68$$

So median of the given data is 68kg

MERIT, DEMERIT AND USES OF MEDIAN

Merits:

1. Median is rigidly defined.

2. It is not affected at all from extreme values.

3. It is easy to understand and to calculate.

4. In case of individual series data it can be located merely by inspection.

5. It can be calculated in case of open end classes.

6. Its graphical representation is also possible.

7. It can be computed even if the classes are of unequal interval.

8. In case of qualitative data, e.g., beauty, honesty, intelligence, etc. it is the best measure of central tendency.

Demerit:

1. It is not amenable to algebraic treatment.

2. It is a positional average and is based only on the middle term. It does not use all the observations of the data.

3. In case of irregular distribution, it is not a good measure.

4. In case of even number of observations it cannot be determined exactly, it can be estimated only by the average of the two middle terms.

5. In comparison to mean it is affected much by fluctuations of sampling.

Uses:

1. It is a good measure if numerical measurements are not possible.

2. In case of qualitative data where the observations cannot be determined quantitatively, it is the only average.

3. It is generally used in studying the average intelligence or average honesty of a group of people.

MODE

Mode is the most frequent item of the series, i.e., in a given set of observations a item or observation which is repeated maximum number of times an all other observations cluster around this, is called mode. For example, the average height of an Indian male is 5 feet 6 inch; the average size of the shoes of an Indian male is number 7, etc. Mode is also known as norm.

TYPES OF MODE OF DISTRIBUTION

Unimodal: If the data of a distribution has only one mode then the distribution is called unimodal.

Bimodal: If we find that there are two items in a distribution which have the same number of repetitions, then these two items are the modes and the distribution is called bimodal.

Trimodal: Similarly, in a distribution, if there are three such items that they have the same frequency then these three items are called the modes of the distribution and the distribution is called trimodal.

Ill- defined mode: If there exists more than one mode in a distribution, then mode is called ill-defined.

MODE IN INDIVIDUAL DISTRIBUTION

In case individual series mode is the most frequent observation. It is clear from the following example.

Example 16. Find the mode of the series given below:

2, 3, 4, 7, 9, 3, 2, 1, 5, 3, 6, 3, 8, 3

Solution. In the given series the observation 3 is repeated maximum number of times (5) so the mode of the given series is 3.

MODE IN DISCRETE FREQUENCY DISTRIBUTION

In case of discrete frequency distribution, mode is the value of the variable which has the maximum frequency. Consider the following example:

Example 17. Find the mode of the following frequency distribution:

Variable (x)	2	5	7	9	11	25	35	43	52
Frequency (f)	1	3	4	8	25	12	11	10	8

Solution: Here we see that in the given distribution, the variable 1 has the maximum frequency 25. So the mode of this distribution is 11.

MODE IN CONTINUOUS FREQUENCY DISTRIBUTION

In case of grouped continuous frequency distribution the maximum frequency shows that the related class is the modal class and for the computation of mode we use the following formula:

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

Where L= lower limit of the modal class

h= magnitude of the modal class

f_1= frequency of the modal class

f_2= frequency of the class succeeding the modal class

f_0= frequency of the class preceding the modal class

For a moderately asymmetrical distribution the mode can be calculated by a formula given by Karl Pearson as follows:

Mode = 3 Median – 2 Mean

Example 19. Following table shows the blood pressure and the frequency related to it. Find the mode of this distribution.

Table 14.

C.I.	Frequency	C.I.	frequency
70-80	2	110-120	32
80-90	4	120-130	28
90-100	14	130-140	12
100-110	35	140-150	5

Solution. From the table it is clear that maximum frequency is 35 and the related class is the 100-110. So 100-110 is the modal class. Now to compute the mode we use the following formula:

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

Here L= lower limit of the modal class= 100

h= magnitude of the modal class= 10

f_1= frequency of the modal class= 35

f_2= frequency of the class succeeding the modal class= 32

f_0= frequency of the class preceding the modal class= 14

$$\therefore \quad M_o = 100 + \frac{(35-14)}{(70-14-32)} \times 10 = 100 + \frac{210}{24} = 100 + 8.75 = 108.75$$

So mode of the given distribution is 108.75.

MERIT, DEMERIT AND USES OF MODE

Merits:

1. Mode is easy to understand and to calculate.

2. It is not affected by extreme values.

3. It can be determined graphically.

4. In some cases it can be located by inspection only.

5. It can be computed for the distributions of unequal class intervals provided the modal class; the class preceding the modal class and succeeding the modal class are of equal width.

6. It represents the most frequent value of the distribution, practically it is very useful.

Demerits:

1. It is not based upon all the observations.

2. It is not subjected to algebraic treatments, i.e., we cannot compute the combined mode if we have the modes of the two series.

3. In some cases mode is ill defined. In some cases it is not possible to find a clear mode. Some series have two modes and some more than two modes.

4. As compared to mean, mode is affected much by fluctuations of sampling; it is an unstable measure of central tendency.

5. If the modal class or the class preceding or succeeding the modal class are of unequal width, it cannot be determined.

6. There are different formulas for the calculation of mode.

Uses:

1. It is used to find the ideal size; it is very useful in business forecasting.
2. It is very useful in ready-made market, e.g., shoes, shirts, jeans etc.
3. It is very useful in commercial management.

SUMMARY

The study of this chapter provides us the knowledge of central tendency and measures of central tendency. From the study of this chapter we came to know the definitions of the measures of central tendency as mean, median and mode. We studied and learnt different methods of computing mean. We learnt about weighted mean and combined mean. We learnt how we can calculate the mean, median and mode in case of individual series, in case of discrete distribution and continuous distribution. We studied the grouping method of mode. We studied the merits, demerits and uses of mean, median and mode also. From the study of these methods, merits, demerits and uses we came to know the situations where which method is suitable and also which measure is suitable for the particular situation? Over all we learnt a lot about measures of central tendency.

References:

- Mathematical Statistics
  S. C. Gupta, V. K. Kapoor, S.Chand Publicaton

- SLM of UOU, Haldwani
  BIOSTATISTICS. (BSCZO-303)

Dr. Kamlesh Bisht
(Mathematics)
Mob. No.-
8279829875

STAY HOME AND STAY SAFE

COVID-19

THANKS