

Measures of averages - Mean – median – mode –computation of the above statistics for raw and grouped data - merits and demerits - measures of location – percentiles – quartiles - computation of the above statistics for raw and grouped data

In the study of a population with respect to one in which we are interested we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average or a measure of locations. There are five averages. Among them mean, median and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages.

Arithmetic mean or mean

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. It is denoted by the symbol \bar{x} If the variable x assumes n values $x_1, x_2 \dots x_n$ then the mean is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$
$$= \frac{1}{n} \sum_{i=1}^n x_i$$

This formula is for the ungrouped or raw data.

Example 1

Calculate the mean for pH levels of soil 6.8, 6.6, 5.2, 5.6, 5.8

Solution

$$\bar{x} = \frac{6.8 + 6.6 + 5.2 + 5.6 + 5.8}{5} = \frac{30}{5} = 6$$

Grouped Data

The mean for grouped data is obtained from the following formula:

$$\bar{x} = \frac{\sum fx}{n}$$

Where x = the mid-point of individual class

f = the frequency of individual class

n = the sum of the frequencies or total frequencies in a sample.

Short-cut method

$$\bar{x} = A + \frac{\sum fd}{n} \times c$$

Where $d = \frac{x - A}{c}$

A = any value in x

n = total frequency

c = width of the class interval

Example 2

Given the following frequency distribution, calculate the arithmetic mean

Marks : 64 63 62 61 60 59 Number of

Students } 8 18 12 9 7 6

Solution

X	F	Fx	D=x-A	Fd
64	8	512	2	16
63	18	1134	1	18
62	12	744	0	0
61	9	549	-1	-9
60	7	420	-2	-14
59	6	354	-3	-18
	60	3713		-7

Direct method

$$\bar{x} = \frac{\sum fx}{n}$$

$$\bar{x} = \frac{3713}{60} = 61.88$$

Short-cut method

$$\bar{x} = A + \frac{\sum fd}{n} \times c$$

Here A = 62

$$\bar{x} = 62 - \frac{7}{60} \times 1 = 61.88$$

Example 3

For the frequency distribution of seed yield of sesamum given in table, calculate the mean yield per plot.

Yield per plot in(in g)	64.5-84.5	84.5-104.5	104.5-124.5	124.5-144.5
No of plots	3	5	7	20

Solution

Yield (in g)	No of Plots (f)	Mid X	$d = \frac{x - A}{c}$	Fd
64.5-84.5	3	74.5	-1	-3
84.5-104.5	5	94.5	0	0
104.5-124.5	7	114.5	1	7
124.5-144.5	20	134.5	2	40
Total	35			44

A=94.5

The mean yield per plot is

Direct method:

$$\begin{aligned}\bar{x} &= \frac{\sum fx}{n} = \frac{(74.5 \times 3) + (94.5 \times 5) + (114.5 \times 7) + (134.5 \times 20)}{35} \\ &= \frac{4187.5}{35} = 119.64 \text{ gms}\end{aligned}$$

Shortcut method

$$\bar{x} = A + \frac{\sum fd}{n} \times c$$

$$\bar{x} = 94.5 + \frac{44}{35} \times 20 = 119.64 \text{ g}$$

Merits and demerits of Arithmetic mean

Merits

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is possible to calculate even if some of the details of the data are lacking.
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.

Demerits

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be in the study of qualitative phenomena not capable of numerical measurement
i.e. Intelligence, beauty, honesty etc.,
3. It can ignore any single item only at the risk of losing its accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.
6. It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

Median

The median is the middle most item that divides the group into two equal parts, one part comprising all values greater, and the other, all values less than that item.

Ungrouped or Raw data

Arrange the given values in the ascending order. If the number of values are odd, median is the middle value

If the number of values are even, median is the mean of middle two values.

By formula

When n is odd, Median = Md = $\left(\frac{n+1}{2}\right)^{th}$ value

When n is even, Average of $\left(\frac{n}{2}\right)$ and $\left(\frac{n}{2} + 1\right)^{th}$ value

Example 4

If the weights of sorghum ear heads are 45, 60, 48, 100, 65 gms, calculate the median

Solution

Here n = 5

First arrange it in ascending order

45, 48, 60, 65, 100

$$\begin{aligned}\text{Median} &= \left(\frac{n+1}{2}\right)^{th} \text{ value} \\ &= \left(\frac{5+1}{2}\right) = 3^{rd} \text{ value} = 60\end{aligned}$$

Example 5

If the sorghum ear- heads are 5, 48, 60, 65, 65, 100 gms, calculate the median.

Solution

Here n = 6

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

$$\left(\frac{n}{2}\right) = \frac{6}{2} = 3^{rd} \text{ value} = 60 \quad \text{and} \quad \left(\frac{n}{2} + 1\right) = \frac{6}{2} + 1 = 4^{th} \text{ value} = 65$$

$$\text{Median} = \frac{60 + 65}{2} = 62.5 \text{ g}$$

Grouped data

In a grouped distribution, values are associated with frequencies. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution. Whatever may be the type of distribution, cumulative frequencies have to be calculated to know the total number of items.

Cumulative frequency (cf)

Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the previous classes, i.e. adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

Discrete Series

Step1: Find cumulative frequencies.

Step2 : Find $\left(\frac{n}{2} + 1\right)$

Step3: See in the cumulative frequencies the value just greater than $\left(\frac{n}{2} + 1\right)$

Step4: Then the corresponding value of x is median.

Example 6

The following data pertaining to the number of insects per plant. Find median number of insects per plant.

Number of insects per plant (x)	1	2	3	4	5	6	7	8	9	10	11	12
No. of plants(f)	1	3	5	6	10	13	9	5	3	2	2	1

Solution

Form the cumulative frequency table

x	f	cf
1	1	1
2	3	4
3	5	9
4	6	15
5	10	25
6	13	38
7	9	47
8	5	52
9	3	55
10	2	57
11	2	59
12	1	60
	60	

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

Here the number of observations is even. Therefore median = average of $(n/2)$ th item and $(n/2+1)$ th item.

$$= (30^{\text{th}} \text{ item} + 31^{\text{st}} \text{ item}) / 2 = (6+6)/2 = 6$$

Hence the median size is 6 insects per plant.

Continuous Series

The steps given below are followed for the calculation of median in continuous series.

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{n}{2}\right)$

Step3: See in the cumulative frequency the value first greater than $\left(\frac{n}{2}\right)$, Then the corresponding class interval is called the Median class. Then apply the formula

$$\text{Median} = l + \frac{\frac{n}{2} - m}{f} \times c$$

where

l = Lower limit of the medianal class

m = cumulative frequency preceding the medianal class

c = width of the class

f = frequency in the median class.

n = Total frequency.

Example 7

For the frequency distribution of weights of sorghum ear-heads given in table below. Calculate the median.

Weights of ear heads (in g)	No of ear heads (f)	Less than class	Cumulative frequency (m)
60-80	22	<80	22
80-100	38	<100	60
100-120	45	<120	105
120-140	35	<140	140
140-160	24	<160	164
Total	164		

Solution

$$\text{Median} = l + \frac{\frac{n}{2} - m}{f} \times c$$

$$\left(\frac{n}{2}\right) = \left(\frac{164}{2}\right) = 82$$

It lies between 60 and 105. Corresponding to 60 the less than class is 100 and corresponding to 105 the less than class is 120. Therefore the median class is 100-120. Its lower limit is 100.

Here $l = 100$, $n = 164$, $f = 45$, $c = 20$, $m = 60$

$$\text{Median} = 100 + \frac{82 - 60}{45} \times 20 = 109 \quad .78 \text{ gms}$$

Merits of Median

1. Median is not influenced by extreme values because it is a positional average.
2. Median can be calculated in case of distribution with open-end intervals.
3. Median can be located even if the data are incomplete.

Demerits of Median

1. A slight change in the series may bring drastic change in median value.
2. In case of even number of items or continuous series, median is an estimated value other than any value in the series.
3. It is not suitable for further mathematical treatment except its use in calculating mean deviation.
4. It does not take into account all the observations.

Mode

The mode refers to that value in a distribution, which occur most frequently. It is an actual value, which has the highest concentration of items in and around it. It shows the centre of concentration of the frequency in around a given value. Therefore, where the purpose is to know the point of the highest concentration it is preferred. It is, thus, a positional measure.

Its importance is very great in agriculture like to find typical height of a crop variety, maximum source of irrigation in a region, maximum disease prone paddy variety. Thus the mode is an important measure in case of qualitative data.

Computation of the mode

Ungrouped or Raw Data

For ungrouped data or a series of individual observations, mode is often found by mere inspection.

Example 8

Find the mode for the following seed weight

2, 7, 10, 15, 10, 17, 8, 10, 2 gms

\therefore Mode = 10

In some cases the mode may be absent while in some cases there may be more than one mode.

Example 9

(1) 12, 10, 15, 24, 30 (no mode)

(2) 7, 10, 15, 12, 7, 14, 24, 10, 7, 20, 10

the modal values are 7 and 10 as both occur 3 times each.

Grouped Data

For Discrete distribution, see the highest frequency and corresponding value of x is mode.

Example:

Find the mode for the following

Weight of sorghum in gms (x)	No. of ear head(f)
50	4
65	6
75	16
80	8
95	7
100	4

Solution

The maximum frequency is 16. The corresponding x value is 75.

\therefore mode = 75 gms.

Continuous distribution

Locate the highest frequency the class corresponding to that frequency is called the modal class.

Then apply the formula.

$$\text{Mode} = l + \frac{f_s}{f_p + f_s} \times c$$

Where l = lower limit of the model class

f_p = the frequency of the class preceding the model class

f_s = the frequency of the class succeeding the model class

and c = class interval

Example 10

For the frequency distribution of weights of sorghum ear-heads given in table below. Calculate the mode

Weights of ear heads (g)	No of ear heads (f)	
60-80	22	
80-100	38	f_p
100-120	45	f
120-140	35	f_s
140-160	20	
Total	160	

Solution

$$\text{Mode} = l + \frac{f_s}{f_p + f_s} \times c$$

Here $l=100$, $f = 45$, $c = 20$, $m = 60$, $f_p = 38$, $f_s = 35$

$$\begin{aligned} \text{Mode} &= 100 + \frac{35_s}{38 + 35} \times 20 \\ &= 100 + \frac{35_s}{73} \times 20 \\ &= 109.589 \end{aligned}$$

References

Website : <http://eagri.org/>