Graphical Determination of Median

Median can be determined graphically using the following Steps

Draw the less than or more than ogive

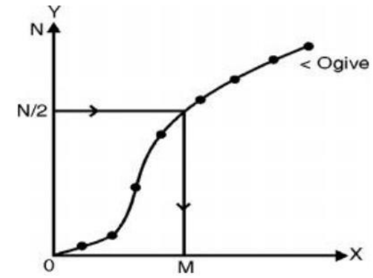
Locate N/2 on the Y axis.

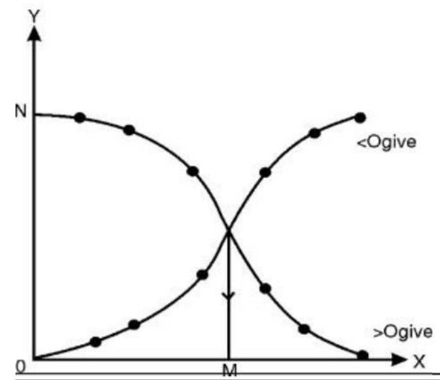
At N/2 draw a perpendicular to the Y axis and extend it to meet the ogive

From the point of intersection drop a perpendicular to the X axis

The point at which the perpendicular meets the X axis will be the median value.

Median can also be determined by drawing the two ogives, simultaneously. Here drop a perpendicular from the point of intersection to the X axis. This perpendicular will meet at the median value.





Merits and Demerits

Merits

It is not unduly affected by extreme items.

It is simple to understand and easy to calculate.

It can be calculated for open end data

It can be determined graphically.

It can be used to deal with qualitative data.

Demerits

It is not rigidly defined. When there are even number of individual observations, median is approximately taken as the mean of the two middle most observations.

It is not based on the magnitude of all the items. It is a positional measure. It is the value of the middle most item.

It cannot be algebraically manipulated. For example, the median of the combined set can not be found from the medians and the sizes of the individual sets alone.

It is difficult to calculate when there are large number of items which are to be arranged in order of magnitude.

It does not have sampling stability. It varies more markedly than A M from sample to sample although all the samples are from one and the same population.

Its use is lesser than that of AM.

Mode

Mode is that value of the variable, which occur maximum number of times in a set of observations. Thus, mode is the value of the variable, which occur most frequently. Usually statements like, 'average student', 'average buyer', 'the typical firm', etc. are referring to mode of the phenomena. Mode is denoted by Z or Mo. For a raw data as well as for a discrete frequency distribution we can locate mode by inspection.

For a frequency distribution mode is defined as the value of the variable having the maximum frequency. For a continuous frequency distribution it can be calculated using the formula given below:

$$Z = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

where	l	:	lower limit of modal class
	Modal class	:	Class having the maximum frequency
	Δ_1	:	difference between the frequency of modal class and that of the premodal class
	Δ_2	:	difference between frequency of the modal class and that of the post modal class
	c	:	class interval

For applying this formula, the class intervals should be (i) of equal size (ii) in ascending order and (iii) in exclusive form.

Example 18

Determine the mode of

420, 395, 342, 444, 551, 395, 425, 417, 395, 401, 390

Solution

Mode = **395**

Example 19

Determine the mode

Size of shoes	: 3	4	5	6	7	8	9
No of pairs sold	: 10	25	32	38	61	47	34

Solution

Mode = Z = 7

Example 20

Calculate mode for the following data

Classes	: 0-9	10 - 19	20-29	30-39	40-49	50-59
f	: 5	10	17	33	22	13

Solution

Classes	f	Atual class
0-9	5	0.5-9.5
10-19	10	9.5-19.5
20-29	17	19.5-29.5
30-39	33	29.5-39.5
40-49	22	39.5-49.5
50-59	13	49.5-59.5

$$Z = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

Modal class is 29.5-39.5

$l = 29.5$

$\Delta_1 = 33 - 17 = 16$

$\Delta_2 = 33 - 22 = 11$, $c = 10$

$$29.5 + \frac{16}{16 + 11} \times 10$$

29.5+5.92 = **35.42**

For a symmetrical or moderately assymmetrical distribution, the empirical relation is

Mean Mode = 3 (Mean Median)

This relation can be used for calculating any one measure, if the remaining two are known.

Example 21

In a moderately assymmetrical distribution Mean is 24.6 and Median 25.1. Find the value of mode.

Solution

We have

Mean Mode = 3(Mean Median)

24.6 Z = 3(24.6 25.1)

24.6 Z = 3(0.5) = 1.5

Z = 24.6 + 1.5 = **26.1**

Example 22

In a moderately assymmetrical distribution Mode is 48.4 and Median 41.6. Find the value of Mean

Solution

We have,

MeanMode = 3(Mean Median)

$\bar{x}$ 48.4 = 3($\bar{x}$ 41.6)

$\bar{x}$ 48.4 = 3 $\bar{x}$ 124.8

3 $\bar{x}$ $\bar{x}$ = 124.8 48.4

2 $\bar{x}$ = 76.4

$\bar{x}$ = 76.4 ÷ 2 = **38.2**

Merits and Demerits

Merits

Mode is not unduly affected by extreme items.

It is simple to understand and easy to calculate

It is the most typical or representative value in the sense that it has the greatest frequency density.

It can be calculated for open-end data.

It can be determined graphically. It is the x-coordinate of the peak of the frequency curve.

It can be found for qualities also. The quality which is observed more often than any other quality is the modal quality.

Demerits

It is not rigidly defined.

It is not based on all the items. It is a positional value.

It cannot be algebraically manipulated. The mode of the combined set cannot be determined as in the case of AM.

Many a time, it is difficult to calculate. Sometimes grouping table and frequency analysis table are to be formed.

It is less stable than the A.M.

Unlike other measures of central tendency, it may not exist for some data. Sometimes there may be two or more modes and so it is said to be ill defined.

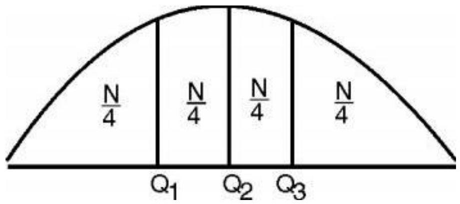
It has very limited use. Modal wage, modal size of shoe, modal size of family, etc., are determined. Consumer preferences are also dealt with.

Partition Values

We have already noted that the total area under a frequency curve is equal to the total frequency. We can divide the distribution or area under a curve into a number of equal parts choosing some points like median. They are generally called *partition values or quantiles*. The important partition values are *quartiles, deciles and percentiles.*

Quartiles

Quartiles are partition values which divide the distribution or area under a frequency curve into 4 equal parts at 3 points namely Q_1, Q_2, and Q_3 . Q_1 is called *first quartile or lower quartile*, Q_2 is called *second quartile, middle quartile or median* and Q_3 is called *third quartile or upper quartile*. In other words Q_1 is the value of the variable such that the number of observations lying below it, is N/4 and above it is 3N/4. Q_2 is the value of the variable such that the number of observations on either side of it is equal to N/2. And Q_3 is the value of the variable such that the number of observations lying below Q_3 is 3N/4 and above Q_3 is N/4.



Deciles and Percentiles

Deciles are partition values which divide the distribution or area under frequency curve into 10 equal parts at 9 points namely D_1, D_2, ..., D_9.

Percentiles are partition values which divide the distribution into 100 equal parts at 99 points namely P_1, P_2, P_3, ... P_{99}. Percentile is a very useful measure in education and psychology. Percentile ranks or scores can also be calculated. Kelly's measure of skewness is based on percentiles.

Calculation of Quartiles

The method of locating quartiles is similar to that method used for finding median. Q_1 is the value of the item at $(n + 1)/4^{th}$ position and Q_3 is the value of the item at $3(n + 1) / 4^{th}$ position when actual values are known. In the case of a frequency distribution Q_1 and Q_3 can be calculated as follows.

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m\right)}{f} \times c$$

where l_1 - lower limit of Q_1 class

Q_1 class - the class in which $N/4^{th}$ item falls

m - cumulative frequency up to Q_1 class

c - class interval

f - frequency of Q_1 class

$$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m\right)}{f} \times c$$

where l_3 - lower limit of Q_3 class

Q_3 class - the class in which $3N/4^{th}$ item falls

m - cumulative frequency up to Q_3 class

c - class interval

f - frequency of Q_3 class

We can combine these three formulae and can be written as

$$Q_i = l_i + \frac{\left(\frac{iN}{4} - m\right)}{f} \times c, \quad i = 1, 2, 3$$

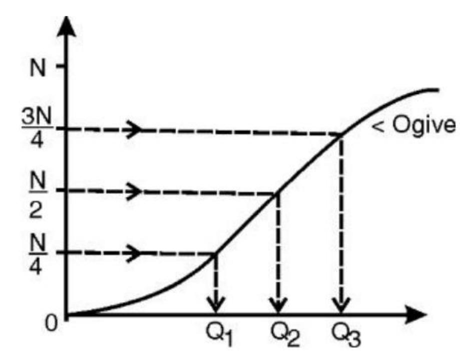
In a similar fashion deciles and percentiles can be calculated as

$$D_i = l_i + \frac{\left(\frac{iN}{10} - m\right)}{f} \times c, \quad i = 1, 2, 3, \ldots 9$$

$$P_i = l_i + \frac{\left(\frac{iN}{100} - m\right)}{f} \times c, \quad i = 1, 2, 3, \ldots, 99$$

Graphical Determination of Quartiles

Quartiles can be determined graphically by drawing the ogives of the given frequency distribution. So draw the less than ogive of the given data. On the Y axis locate N/4, N/2 and 3N/4. At these points draw perpendiculars to the Y axis and extend it to meet the ogive. From the points of intersection drop perpen-diculars to the X axis. The point corresponding to the CF, N/4 is Q_1 corresponding to the CF N/2 is Q_2 and corres-ponding to the CF 3N/4 is Q_3.



Example 23

Find , Q_1, Q_3, D_2, D_9, P_{16}, P_{65} for the following data. 282, 754, 125, 765, 875, 645, 985, 235, 175, 895, 905, 112 and 155.

Solution

Step 1. Arrange the values in ascending order
112, 125, 155, 175, 235, 282, 645, 754, 765, 875,
895, 905 and 985.

Step 2. Position of Q_1 is $\frac{n+1}{4} = \frac{13+1}{4} = \frac{14}{4} = 3.5$

Similarly positions of Q_3, D_2, D_9, P_{16} and P_{65} are 10.5, 2.8, 12.6, 2.24 and 9. 1 respectively.

Step 3.

$$Q_1 = 155 + 0.5(175 - 155) = \mathbf{165}$$

$$Q_3 = 875 + 0.5(895 - 875) = \mathbf{885}$$

$$2 = 125 + 0.8(155 - 125) = \mathbf{149.0}$$

$$D_9 = 905 + 0.6(985 - 905) = \mathbf{953}$$

$$P_{16} = 125 + 0.24(155 - 125) = \mathbf{132.20}$$

$$P_{65} = 765 + 0.1(875 - 765) = \mathbf{776.0}$$

Note

The value of the 12.6-th position (D_9) is obtained as value of 12-th position + 0.6 (value at 13-th position - value at 12-th position)

Example 24

Find Q_1, Q_3, D_4, P_{20} and P_{99} for the data given below.

Mark	:	25	35	40	50	52	53	67	75	80
No of students	:	3	29	32	41	49	54	38	29	27

Solution

Marks	No of students	Cumulative frequency
25	3	3
35	29	32
40	32	64
50	41	105
52	49	154
53	54	208
67	38	246
75	29	275
80	27	302

Step 1. The cumulative frequencies of marks given in ascending order are found

Step 2. The positions of Q_1, Q_3, D_4, P_{20} and P_{99} are found. They are

$$\frac{N+1}{4} = \frac{303}{4} = 75.75$$

$$\frac{3(N+1)}{4} = 3 \times \frac{303}{4} = 227.25$$

$$\frac{4(N+1)}{10} = \frac{40 \times 303}{10} = 121.20$$

$$\frac{20(N+1)}{100} = \frac{20 \times 303}{100} = 60.60$$

$$\frac{99(N+1)}{100} = \frac{99 \times 303}{100} = 299.97$$

Step 3. The marks of students at those positions are found

$$Q_1 = 50 + 0.75(50 - 50) = \mathbf{50} \text{ Marks}$$

$$Q_3 = 67 + 0.25(67 - 67) = \mathbf{67} \text{ Marks}$$

$$D_4 = 52 + 0.20(52 - 52) = \mathbf{52} \text{ Marks}$$

$$P_{20} = 40 + 0.60(40 - 40) = \mathbf{40} \text{ Marks}$$

$$P_{99} = 80 + 0.97(80 - 80) = \mathbf{80} \text{ Marks}$$

Note

Refer the above example to know the method of finding the values of the items whose positions are fractions.

Example 25

Calculate quartiles for the following data

Classes	:	30-35	35-40	40-45	45-50	50-55	55-60	60-65
Freq.	:	10	16	18	27	18	8	3

Solution

Class	f	CF
30-35	10	10
35-40	16	26
40-45	18	44
45-50	27	71
50-55	18	89
55-60	8	97
60-65	3	100
Total	100	

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m\right)c}{f}$$

$$= 35 + \frac{(25 - 10)5}{16}$$

$$35 + \frac{15 \times 5}{16} = 35 + \frac{75}{16}$$

$$= 35 + 4.68 = \mathbf{39.68}$$

$$Q_2 = l_2 + \frac{\left(\frac{N}{2} - m\right)c}{f}$$

$$45 + \frac{(50 - 44)5}{27}$$

$$45 + \frac{6 \times 5}{27}$$

$$45 + \frac{10}{9} = 45 + 1.11 = \mathbf{46.11}$$

$$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m\right)c}{f}$$

$$= 50 + \frac{(75 - 71)5}{18}$$

$$= 50 + \frac{4 \times 5}{18} = 50 + \frac{10}{9} = 50 + 1.11 = \mathbf{51.11}$$

EXERCISES

Multiple Choice Questions

1. Mean is a measure of
   a. location or central value   b. dispersion
   c. correlation   d. none of the above

If a constant value 50 is subtracted from each observation of a set, the mean of the set is:
a. increased by 50   b. decreased by 50
c. is not affected   d. zero

If the grouped data has open end classes, one cannot calculate:
a. median   b. mode   c.mean   d. quartiles

Harmonic mean is better than other means if the data are for:
a. speed or rates   b. heights or lengths
c. binary values like 0 & 1   d. ratio or proportions

Extreme value have no effect on:
a. average   b. median
c. geometric mean   d. harmonic mean

If the A.M. of a set of two observations is 9 and its G.M. is 6. Then the H.M. of the set of observations is:
a. 4   b. $3\sqrt{6}$ c. 3   d. 1.5

The A.M. of two numbers is 6.5 and their G.M. is 6. The two numbers are:
a. 9, 6   b. 9, 5   c. 7, 6   d. 4, 9

8.If the two observations are 10 and 10 then their harmonic mean is:
a. 10   b. 0   c. 5   d. ∞

The median of the variate values 11, 7, 6, 9, 12, 15,, 19 is:
a. 9   b. 12   c. 15   d. 11

10. The second dicile divides the series in the ratio:
    a. 1:1   b. 1:2   c. 1:4   d. 2:5

11. For further algebraic treatment, geometric mean is:
    a. suitable   b. not suitable
    c. sometimes suitable   d. none of the above

12.The percentage of values of a set which is beyond the third quartile is:
a. 100 percent   b. 75 percent
c. 50 percent   d. 25 percent

In a distribution, the value around which the items tend to be most heavily concentrated is called:
a. mean   b. median
c. third quartile   d. mode

14. Sum of the deviations about mean is
    a. zero   b. minimum   c. maximum   d. one

15. The suitable measure of central tendency for qualitative data is:
    a. mode   b. arithmetic mean
    c. geometric mean   d. median

16. The mean of the squares of first eleven natural numbers is:
    a. 46   b. 23   c. 48   d. 42

The percentage of items in a frequency distribution lying between upper and lower quartiles is:
a. 80 percent   b. 40 percent
c. 50 percent   d. 25 percent

Very Short Answer Questions

What is central tendency?

Define Median and mode.

Define harmonic mean

Define partition values

State the properties of AM.

In a class of boys and girls the mean marks of 10 boys is 38 and the mean marks of 20 girls 45. What is the average mark of the class?

23. Define deciles and percentiles.

24 Find the combined mean from the following data.

	Series x	Series y
Arithmetic mean	12	20
No of items	80	60

Short Essay Questions

25 Define mode. How is it calculated. Point out two

Define AM, median and mode and explain their uses

Give the formulae used to calculate the mean, median and mode of a frequency distribution and explain the symbols used in them.

How will you determine three quartiles graphically from a less than ogive?

Three samples of sizes 80, 40 and 30 having means 12.5, 13 and 11 respectively are combined. Find the mean of the combined sample.

Explain the advantages and disadvantages of arithmetic mean as an average.

For finding out the 'typical' value of a series, what measure of central tendency is appropriate?

32 Explain AM and HM. Which one is better? And Why?

Prove that the weighted arithmetic mean of first n natural numbers whose weights are equal to the corresponding number is equal to $(2n+1)/3$

Show that GM of a set of positive observation lies between AM & AM.

What are the essential requisites of a good measure of central tendency? Compare and contrast the commonly employed measures in terms of these requisites.

Discuss the merits and demerits of the various measures of central tendency. Which particular measure is considered the best and why? Illustrate your answer.

. What is the difference between simple and weighted average? Explain the circumstances under which the latter should be used in preference to the former.

Find the average rate of increase in population which in the first decade has increased 12 percent, in the next by 16 per cent, and in third by 21 percent.

39.. A person travels the first mile at 10 km. per hour, the second mile at 8 km. per hour and the third mile at 6 km. per hour. What is his average speed?

Long Essay Questions

Compute the AM, median and mode from the following data

Age last birth day	:	15-19	20-24	25-29	30-34	35-39	40-44
No of persons	:	4	20	38	24	10	

41. Calculate Arithmetic mean, median and mode for the following data.

Ag e	:	55-60	50-55	45-50	40-45	35-40	30-35	25-30	20-25
No of people	:	7	13	15	20	30	33	28	14

Calculate mean, median and mode from the following data

Class	Frequency
Up to 20	52
20-30	161

30-40	254
40-50	167
50-60	78
60-80	64
Over 80	52

Calculate mean, median and mode

Central wage in Rs. :	15	20	25	30	35	40	45
No. of wage earners:	3	25	19	16	4	5	6

(i)Find the missing frequencies in the following distribution given that N = 100 and median of the distribution is 110.

Calculate the arithmetic mean of the completed frequency distribution.

Class	**: 20 - 4 0**	**40 -6 0**	**60 -8 0**	**80- 100**	**100-120**
Frequency	**: 6**	**9**	**-**	**1 4**	**2 0**
Class	**: 120-140**	**140-160**	**160-180**	**180-200**	
Frequency	**: 1 5**	**-**	**8**	**7**	