Chapter 3

Measures of Central Tendency:

Arithmetic Mean

Geometric Mean

Harmonic mean

Median

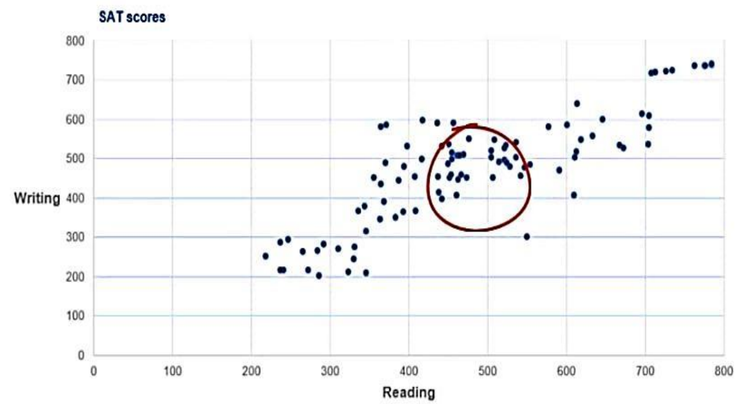
Mode

Measures of Central Tendency

What is Central Tendency?

A measure of central tendency is a summary statistic that represents the center point or typical value of a dataset. These measures indicate where most values in a distribution fall and are also referred to as the central location of a distribution. You can think of it as the tendency of data to cluster around a middle value. **A measure of central tendency (also referred to as measures of Centre or central location) is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or center of its distribution.** In statistics, the most common measures of central tendency are

- AM — Arithmetic mean
- GM — Geometric mean
- HM — Harmonic mean
- Me — Median
- Mo — Mode



Each of these measures calculates the location of the central point using a different method. Choosing the best measure of central tendency depends on the type of data you have. In this post, I explore these measures of central tendency; show you how to calculate them, and how to determine which one is best for your data.

Mean or average or arithmetic mean (AM)?

Mean or average or arithmetic mean (AM) is one of the representative values of data. We can find the mean of observations by dividing the sum of all the observations by the total number of observations.

Mean of raw or ungrouped data: If $x_1, x_2, x_3, \ldots\ldots x_n$ are n observations, then

Arithmetic Mean (AM) = $(x_1 + x_2 + x_3 + \ldots\ldots + x_n)/n = (\sum xi)/n$ $\sum$ *is a Greek letter showing summation*

Example: Weights of 6 boys in a group are 63, 57, 39, 41, 45, 45. Find the mean weight.

Solution: Number of observations = 6   Sum of all the observations = 63 + 57 + 39 + 41 + 45 + 45 = 290

Therefore, arithmetic mean = 290/6 = 48.3

Mean of tabulated or grouped data:

If $x_1, x_2, x_3, x_4, \ldots\ldots x_n$ are n observations, and $f_1, f_2, f_3, f_4, \ldots\ldots f_n$ represent frequency of n observations. Then mean of the tabulated data is given by

$$\text{Arithmetic Mean (AM)} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots\ldots f_n x_n}{f_1 + f_2 + f_3 + \ldots\ldots f_n} = \frac{\sum f_i x_i}{\sum fi} = \frac{\sum f_i x_i}{n}$$

Example: The following table shows the number of plants in 20 houses in a group. Find the mean number of plans per house.

Number of Plants	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	12 - 14
Number of Houses	1	2	2	4	6	2	3

Solution: We have

$\sum fi$ = 1 + 2 + 2 + 4 + 6 + 2 + 3 = 20

$\sum fix_i$ =1 + 6 + 10 + 28 + 54 + 22 + 39 = 160

Therefore, mean = $\sum(f_i x_i)/\sum fi$ = 160/20 = 8 plants

No. of Plants	No. of Houses (f_i)	Mid value (x_i)	$f_i x_i$
0 - 2	1	1	1 × 1 = 1
2 - 4	2	3	2 × 3 = 6
4 - 6	2	5	2 × 5 = 10
6 - 8	4	7	4 × 7 = 28
8 - 10	6	9	6 × 9 = 54
10 - 12	2	11	2 × 11 = 22
12 -14	3	13	3 × 13 = 39
	n = 20		$\sum(f_i x_i)$ = 160

Combined Mean or weighted mean?

A **combined arithmetic mean** or **weighted mean** is a mean of two or more separate groups, and is found by :

1. Calculating the mean of each group
2. Combining the results

Combined mean or weighted mean, $\overline{x}_c = \dfrac{m \cdot \overline{x}_a + n \cdot \overline{x}_b}{m+n}$

$\overline{x}_a$ = the mean of the first set,
m = the number of items in the first set,
$\overline{x}_b$ = the mean of the second set,
n = the number of items in the second set,
$\overline{x}_c$ the combined mean.

Example 1: Suppose you are running a survey on math proficiency (as measured by an achievement test) in kindergarten, and you have results from two different schools.

1. In school 1, 57 kindergarteners were tested and their mean score was 82.
2. In school 2, 23 kindergarteners were tested and their mean score was 63.

Solution 1: The combined mean can be calculated by :

$$\overline{x_c} = [(57 \times 82) + (23 \times 63)]/(57 + 23) \ = \ 76.5$$

Example 2: Now suppose you were running a survey on reading speed, as measured by how long it took 1st graders to read a given block of text. Your results come in for five schools:

School ID	Number of Students	Average Time
School 1	189	83
School 2	46	121
School 3	89	82
School 4	50	147
School 5	12	60

Solution 2:
To calculate the combined mean:

1. Multiply column 2 and column 3 for each row
2. Add up the results from Step 1
3. Divide the sum from Step 2 by the sum of column 2

So, $\overline{x_c} \ = \ (189 \times 83 + 46 \times 121 \ + 89 \times 82 \ + 50 \times 147 + 12 \times 60)/(189 + 46 + 89 + 50 + 12)$ = 94.87

Example 3: Find the weighted mean for the following data.

Solution 3:

	Group – 1	Group – 2
Mean wages	75	60
No. of workers	1000	1500

Given data: $N_1 = 1000$ $N_2 = 1500$

$\overline{x_1} = 75$ & $\overline{x_2} = 60$

$$\text{Group Mean} = \overline{x_{12}} = \frac{N_1\overline{x_1} + N_2\overline{x_2}}{N_1 + N_2}$$

$$= \frac{1000 \times 75 + 1500 \times 60}{1000 + 1500}$$

$= \overline{x_{12}} =$ USD 66

Limitations/demerits of mean

1. It is unduly affected by extreme items.
2. It is sometimes un-realistic.
3. It may leads to confusion.
4. Suitable only for quantitative data (for variables).
5. It can not be located by graphical method or by observations

Merits of arithmetic mean

1. It is simple and easy to compute.
2. It is rigidly defined.
3. It can be used for further calculation.
4. It is based on all observations in the series.
5. It helps for direct comparison.
6. It is more stable measure of central tendency (ideal average).

Geometric Mean (GM)

The **Geometric Mean (GM)** is n-th root of product of quantities of the series. It is observed by multiplying the values of items together and extracting the root of the product corresponding to the number of items. Thus, square root of the products of two items and cube root of the products of the three items are the Geometric Mean. Usually, geometric mean is never larger than arithmetic mean. If there are zeros and negative numbers in the series, the geometric means cannot be used logarithms can be used to find geometric mean to reduce large number and to save time.

In various sectors different types of problems often arise relating to average percentage rate of change over a period of time. In such cases, the arithmetic mean is not an appropriate average to employ, so, that we can use geometric mean in such case. GM are highly useful in the construction of index numbers.

For ungrouped data

$$\text{G.M.} = (x_1 \cdot x_2 \cdot \cdot \cdot x_n)^{1/n}$$

$$\log(\text{GM}) = \frac{1}{n}\log(X_1.X_2......X_n)$$

$$= \frac{1}{n}[\log X_1 + \log X_2 + \cdots\cdots + \log X_n] = \frac{\sum \log X_i}{n}$$

Taking antilog of both sides, we have

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

For grouped data

$$\text{GM} = \left[\underbrace{X_1 \cdot X_1 \cdots\cdots X_1}_{f_1 \text{ times}} \underbrace{X_2 \cdots\cdots X_2}_{f_2 \text{ times}} \cdots\cdots \underbrace{X_n \cdots\cdots X_n}_{f_n \text{ times}}\right]^{\frac{1}{N}}$$

$$= \left[X_1^{f_1} . X_2^{f_2} \ldots\ldots X_n^{f_n}\right]^{\frac{1}{N}}$$

Taking log of both sides, we have

$$\log(\text{GM}) = \frac{1}{N}\left[\log X_1^{f_1} + \log X_2^{f_2} + \cdots\cdots + \log X_n^{f_n}\right]$$

$$= \frac{1}{N}[f_1 \log X_1 + f_2 \log X_2 + \cdots\cdots + f_n \log X_n] = \frac{\sum_{i=1}^{n} f_i \log X_i}{N}$$

Merits of GM

1. It is based on all the observations in the series.
2. It is rigidly defined.
3. It is best suited for percentage averages and ratios.
4. It is less affected by extreme values.
5. It is useful for studying social and economics data.

Demerits of GM

1. It is not simple to understand.
2. It requires computational skill.
3. GM cannot be computed if any of item is zero or negative.
4. It has restricted application.

Example 1: Consider following time series for monthly sales of ABC company for 4months. Find average rate of change per monthly sales.

Month	Sales
I	10000
II	8000
III	12000
IV	15000

Let Base year = 100% sales

Solution 1:

Month	Base year	Sales (Rs)	Increase / decrease %ge	Conversion (x)	log (x)
I	100%	10000	–	–	–
II	– 20%	8000	80	80	1.903
III	+ 50%	12000	130	130	2.113
IV	+ 25%	15000	155	155	2.190
					$\Sigma \log x = 6.206$

$$GM = \text{Antilog}\left[\frac{6.206}{3}\right] = 117.13$$

Average sales = 117.13 – 100 = 17.13%

Geometric Mean for grouped data

1. Find mid value m and take log of m for each mid value.
2. Multiply log m with frequency 'f' of each class to get f log m and sum up to obtain $\Sigma f \log m$.
3. Divide $\Sigma f \log m$ by N and take antilog to get GM.

Example 2: Find out GM for given data below

Solution 2:

Yield of wheat in MT	No. of farms frequency (f)	Mid value 'm'	log m	f log m
1 – 10	3	5.5	0.740	2.220
11 – 20	16	15.5	1.190	19.040
21 – 30	26	25.5	1.406	36.556
31 – 40	31	35.5	1.550	48.050
41 – 50	16	45.5	1.658	26.528
51 – 60	8	55.5	1.744	13.954
	Σf = N = 100		Σ f log m = 146.348	

$$GM = \text{Antilog}\left[\frac{\Sigma f \log m}{N}\right]$$

$$GM = \text{Antilog}\left[\frac{146.348}{100}\right]$$

GM = 29.07

Harmonic Mean (HM)

Harmonic Mean is the total number of items of a value divided by the sum of reciprocal of values of variable. It is a specified average which solves problems involving variables expressed in within 'Time rates' that vary according to time. For example, speed in km/hr, min/day, price/unit. Harmonic Mean (HM) is suitable only when time factor is variable and the act being performed remains constant.

For ungrouped data

If there are n observations $x_1, x_2, \dots \dots\ x_n$, their harmonic mean is defined as, $HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

For grouped data

If there are n observations $x_1, x_2, \dots \dots\ x_n$, occure with respective frequencies $f_1, f_2, \dots \dots .. f_n$, then harmonic mean is defined as,

$$HM = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

$$HM = \frac{N}{\sum \frac{f_i}{X_i}}$$

Merits of HM

1. It is rigidly defined average and its value is always definite.
2. Its value is based on all observation in a given series.
3. It is capable of further algebraic treatment.
4. It is not affected by sampling fluctuations.
5. In problems relating to time and rates, it gives better results as compared to other averages. Harmonic mean gives the best result when distance covered are the same, but speed of coverage varies.

Demerits of HM

1. It is not easily understood and hence its application is ignored.
2. It is not easy to calculate as it involves reciprocal values.
3. It gives undue weights to small items and ignores bigger items. This restricts its use in the analysis of economic data.
4. In case of zero or negative values, it cannot be computed.

Example 1: A man travel by a car for 3 days he covered 480 km each day. On the first day he drives for 10 hrs at the rate of 48 KMPH, on the second day for 12 hrs at the rate of 40 KMPH, and on the 3rd day for 15 hrs @ 32 KMPH. Compute HM.

Solution 1:

x	$1/x$
48	0.0208
40	0.025
32	0.0312
	$\Sigma 1/x$ = **0.0770**

$$HM = \frac{N}{\Sigma 1/x}$$

$$= \frac{3}{0.0770}$$

$$HM = 38.91$$

Example 2: Find the HM for the grouped data given below,

Solution 2:

Class (CI)	**Frequency (f)**	**Mid point (m)**	**Reciprocal** $\left(\frac{1}{m}\right)$	$f\left(\frac{1}{m}\right)$
0 – 10	5	5	0.2	1
10 – 20	15	15	0.0666	0.999
20 – 30	25	25	0.04	1
30 – 40	8	35	0.0285	0.228
40 – 50	7	45	0.0222	0.1554
	Σf = 60		$\Sigma f\left(\frac{1}{m}\right)$ = **3.3824**	

$$HM = \frac{N}{\Sigma f\left(\frac{1}{m}\right)}$$

$$= \frac{60}{3.3824}$$

$$HM = 17.73$$

Median (Me)

Median

Median is the value of that item in a series which divides the array into two equal parts, one consisting of all the values less than it and other consisting of all the values more than it. Median is a positional average. The number of items below it is equal to the number. The number of items below it is equal to the number of items above it. It occupies central position. Thus, Median is defined as the mid value of the variants. If the values are arranged in ascending or descending order of their magnitude, median is the middle value of the number of variant is odd and average of two middle values if the number of variants is even. For example, if 9 students are stand in the order of their heights; the 5th student from either side shall be the one whose height will be Median height of the students group. Thus, median of group is given by an equation,

For ungrouped data $$Median = \frac{N+1}{2}$$

Merits of Median

1. It is simple, easy to compute and understand.
2. It's value is not affected by extreme variables.
3. It is capable for further algebraic treatment.
4. It can be determined by inspection for arrayed data.
5. It can be found graphically also.
6. It indicates the value of middle item.

Example 1:

Odd number 1, 3, 3, **6**, 7, 8, 9

Median = 6

Even number 1, 2, 3, **4**, **5**, 6, 8, 9

Median = (4 + 5) ÷ 2

= 4.5

Demerits of Median

1. It may not be representative value as it ignores extreme values.
2. It can not be determined precisely when its size falls between the two values.
3. It is not useful in cases where large weights are to be given to extreme values.

For grouped data

$$M_e = l + \left(\frac{\frac{n}{2} - cf}{f}\right)h$$

l = Lower limit of the median class
cf = Previous cumulative frequency of median class
f = Frequency of median class
h = Size of the median class
n = Total no of observation or the total of the frequency.

Example 2: Find the Median for the grouped data given below,

CI	Frequency (f)	Cum. frequency (cf)
110 – 120	6	6
120 – 130	25	31
130 – 140	48	79
140 – 150	72	151
150 – 160	116	267 **N/2 class**
160 – 170	60	327
170 – 180	38	365
180 – 190	22	387
190 – 200	3	390
	$\Sigma f = N = 390$	

Solution 2:

Median class,

N/2 = 390/2 = 195

195 lies in the cumulative frequency 267

So, median class is 150-160

l = 150
cf = 151
f = 116
h = 10
n = 390

$$M_e = l + \left(\frac{\frac{n}{2} - cf}{f}\right)h$$

$$= 150 + \frac{10}{116}[195 - 151]$$

$$= 153.8$$

Mode (Mo)

Mode is often said to be that value in a series which occurs most frequently or which has the greatest frequency. But it is not exactly true for every frequency distribution. Rather it is that value around which the items tend to concentrate most heavily.

For ungrouped data

For example, if you are asked to find the mode of the following series: 8, 9, 11, 15, 16, 12, 15, 3, 7, 15, you will find that there are ten observations in the series wherein the figure 15 occurs maximum number of times three. Thus mode is therefore 15.

For ungrouped data

$$Mo = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

l = Lower limit of the modal class
Δ_1 = Difference between frequency of modal class & pre-modal class
Δ_2 = Difference between frequency of modal class & post-modal class
f = Frequency of modal class
h = Size of the modal class

Merits of Mode

1. Easy to understand and simple to calculate.
2. It is not affected by extremely large or small values.
3. It can be located just by inspection in ungrouped data and discrete frequency distribution.
4. It can be useful for qualitative data.
5. It can be computed in an open-end frequency table.
6. It can be located graphically.

Demerits of Mode

1. It is not well defined.
2. It is not based on all the values.
3. It is stable for large values, so not well defined if the data consists of a small number of values.
4. It is not capable of further mathematical treatment.
5. Sometimes the data has one or more than one mode and sometimes the data has no mode at all.

Example 1: Calculate the modal sales of the 100 companies from the following data,

Solution 1:

Sales in Rs(lakhs)	No. of companies
58-60	12
60-62	18
62-64	25
64-66	30
66-68	10
68-70	3
70-72	2

Since the maximum frequency is 30, hence the class 64-66 is the modal class.

$$Mo = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

$$= 64 + \frac{5}{5+20} \times 2$$

$$= 64.4$$

l = 64
Δ_1 = 30-25=5
Δ_2 = 30-10=20
h = 2

Example 2: Calculate the average religion for the following data,

Religion	No. of students
Islam	30
Hindu	15
Buddha	4
Christian	1

Solution 2: As maximum number of students practice Islam among the 50 students in the class, so mode or average religion is Islam.