

Summarizing a raw data set or an organized data set

There are two basic properties of a quantitative data set that are commonly studied. These are central tendency and variability (or dispersion).

Central Tendency:

Quite often it is found that the entries in data set cluster around a central (or middle) value. This behavior of the data set is called the central tendency. The main Challenge is to locate a central value around which the clustering takes place. Three standard methods to measure the location of central tendency are:

- Mean
- Median
- Mode

Variability or Dispersion:

Variability or dispersion of data set means the amount of discrepancies among the data entries. There are several ways to measure dispersion or variability in a data set and these are:

- Range
- Quartile deviation
- Variance
- Standard deviation

Arithmetic Mean or Average

❖ Mean of “ n ” observations (x_1, x_2, \dots, x_n) is given by

$$\bar{x} = \frac{1}{n} \sum x_i .$$

Example: A Monthly income of 10 families in a city is given by:
1600, 1560, 1440, 1530, 1670, 1860, 1750, 1910, 1490, 1800.

Ans:

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{16610}{10} = 1661.$$

❖ In case of the discrete frequency distribution:

If f_i 's are the frequencies of the variable x_i 's then mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i , \quad \text{where } N = \sum_{i=1}^n f_i .$$

Example: The following data represents the marks obtained by 60 students of a class. Obtain the average marks.

Marks	20	30	40	50	60	70
No. of Students	8	12	20	10	6	4

Ans:

Marks	No. of Students (f)	fx
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	$N = 60$	$\Sigma fx = 2460$

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{i=1}^n f_i x_i \\ &= \frac{2460}{60} \\ &= 41\end{aligned}$$

❖ In case of the continuous frequency distribution:

If f_i 's are the frequencies of the variable x_i 's then mean

$$\bar{x} = \frac{\sum(\text{midvalue of each class}) \times f_i}{\sum f_i}.$$

Example: Obtain the Arithmetic mean for the following data.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students	12	18	27	20	17	6

Ans:

Marks	No. of Students (f)	Mid-point (x)	fx
0-10	12	5	60
10-20	18	15	270
20-30	27	25	675
30-40	20	35	700
40-50	17	45	765
50-60	6	55	330

$$\bar{x} = \frac{2800}{100} = 28$$

Calculation of mean by using deviation concept:

Sometime the values of the variable (x) or frequency (f) or both are large. Then, the calculation of mean by previous formulas is quite time-consuming. Hence, to avoid such situations we are calculating mean by taking the deviations of the given values from any arbitrary point “ A ” as explained below:

Discrete frequency distribution: $\bar{x} = A + \frac{1}{N} \sum_{i=1}^n f_i d_i$,

where “ A ” is an arbitrary point, $d_i = x_i - A$ and $N = \sum_{i=1}^n f_i$.

Continuous frequency distribution: $\bar{x} = A + \frac{h}{N} \sum_{i=1}^n f_i d_i$,

where “ A ” is an arbitrary point, “ h ” is the magnitude of class interval and $N = \sum_{i=1}^n f_i$. Here, $d_i = \frac{x_i - A}{h}$, where x_i ’s are the mid value of each class.

Example:

C.I.	0-8	8-16	16-24	24-32	32-40	40-48
Frequency	8	7	16	24	15	7

Ans:

C.I.	Mid-Value	Frequency (f)	$d = \frac{x - A}{h}$	fd
0-8	4	8	-3	-24
8-16	12	7	-2	-14
16-24	20	16	-1	-16
24-32	28	24	0	0
32-40	36	15	1	15
40-48	44	7	2	14
Total		77		-25

$$\bar{X} = A + \frac{h}{N} \sum_{i=1}^n f_i d_i = 28 + \frac{8 \times (-25)}{77} = 25.404$$

Measure of Central Tendency	Merits	Demerits :
a) Arithmetic Mean	<ol style="list-style-type: none"> 1) It is rigidly defined. 2) It is easy to understand and easy to calculate. 3) It is based on all the observations. 4) It is suitable for further mathematical manipulations. 5) Of all the averages, it is affected least by sampling fluctuations. 	<ol style="list-style-type: none"> 1) It is affected very much by extreme values 2) It cannot be used if measurements are qualitative.