

## Measures of Central Tendency

The simplest data analysis tool used to summarise the data set using a central value which is representative of the entire population.

There are three measures of central tendency - mean, median and mode.

### Mean

1. **Arithmetic Mean:** Most common measure of central tendency.

For ungrouped data,  $AM = \sum x / n$

For frequency distribution,  $AM = \sum f x / N$ , where  $N = \sum f$

For simplifying the calculations, shortcut methods can be used by introducing assumed mean (A) and step deviation (h).

$AM = A + (\sum f d' / N)$ , where  $d' = (x - A) / h$

A change in origin (addition / subtraction) or scale (multiplication/ division) of the data set changes the Arithmetic mean by the same amount.

The biggest disadvantage of AM is that it is affected by the presence of extreme values

2. **Geometric Mean:**

For ungrouped data,  $GM = (x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n)^{1/n}$

For frequency distribution,  $GM = (x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n})^{1/N}$ , where  $N = \sum f$

For simplifying the calculation the following formula can be used

$\log (GM) = \sum f \log (x) / N$ , where  $N = \sum f$

3. **Harmonic Mean:**

For ungrouped data,  $HM = n / \sum (1/x)$

For frequency distribution,  $HM = N / \sum (f/x)$ , where  $N = \sum f$

### Median

Median of a distribution in ascending or descending order is that observation which divides the distribution into two equal parts

If number of terms (n) is odd, median is the  $(n + 1)/2$  th term

If number of terms (n) is even, median is the Arithmetic Mean of the  $n/2$  and  $(n/2 + 1)$  th term

For continuous frequency distribution, class containing the  $N/2$  the term is called the median class and median is calculated using the formula

$Median = L + (N/2 - f_c) \times h / f$

where, L is the lower class limit of the median class,  $f_c$  is the cumulative frequency of the class preceding the median class and f is the frequency of median class and I is the width of median class.

## Mode

Mode is the observation in a distribution whose frequency is maximum. In a data set, there can be more than one mode.

For continuous frequency distribution, the class with the maximum frequency is called the modal class. Mode can then be calculated using the formula,

$$\text{Mode} = L + \left( \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times I$$

where, L is the lower class limit of the modal class, f is the frequency of modal class,  $f_0$  is the frequency of the class preceding the modal class and  $f_2$  is the frequency of the class next to the modal class and I is the width of modal class.

## Quartiles and Percentiles

Quartiles divide a data set into 4 equal parts and percentiles divide it into 100 equal parts.

$i^{\text{th}}$  quartile is the  $iN/4$  th term. For continuous distribution it can be found by using the formula

$$i^{\text{th}} \text{ quartile} = L + \left( \frac{iN/4 - f_c}{f} \right) \times I$$

$i^{\text{th}}$  percentile is the  $iN/100$  th term. For continuous distribution it can be found by using the formula

$$i^{\text{th}} \text{ percentile} = L + \left( \frac{iN/100 - f_c}{f} \right) \times I$$