Measure Of Central Tendency

Computer Oriented Numerical and Statistical Methods

Outline

- Introduction
- Some Definitions
- Types Of Average
- Related Positional Measures
- Empirical Relationships Between Average
- Choice Of A Suitable Average

Introduction

- The limit of ingenuity of human mind requires that the entire mass of unwieldy data should not only be compressed in tabular form but its chief characteristics should be represented by a single figure which summarizes and represent he characteristics or general significance of data comprising a set of unequal values.
- This single figure is called an average.
- A concise numerical description also enables us to form a mental image of data and interpret its significance.
- This single value is the point of location around which individual values cluster and therefore called the measure of location.
- Since this single value has a tendency to be somewhere at the centre and within the range of all values it is also known as the measure of central tendency.

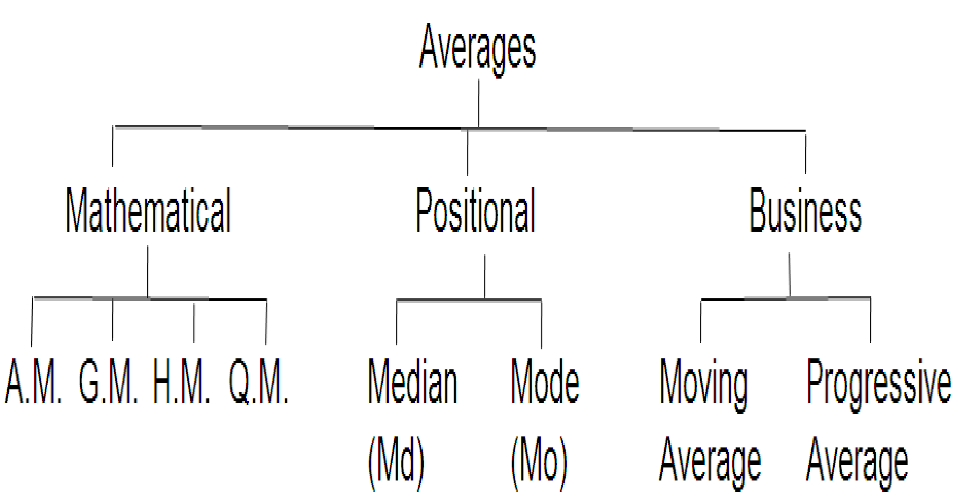
Some Definitions

- Averages are statistical constants which enable us to comprehend in a single effort the significance of the whole. –A.L.Bowely
- An average is a single value selected from a group of values to represent them in some way a value which is supposed to stand for whole group of which it is part, as typical of all the values in the group. – A.E. Waugh
- A measure of central tendency is a typical value around which other figures congregate. – Simpson and Kafea
- An average stands for the whole group of which it forms a part yet represents the whole. – A. E. Waugh
- Averages / Measure of central tendency / measure of location / measure of central location.

Objectives

- To get a single value.
- To facilitate comparison.
- To trace precise relationship.
- To know about the universe from a sample.
- To help in decision-making.

Types of Averages



Arithmetic Mean ($\bar{x}$)

- The arithmetic average or mean usually denoted by $\bar{x}$ of a set of observations is the sum of the values of all observations in a series (Σx) divided by the number of items (N) contained in the series.
- If $x_1 + x_2 + x_3 + ...+ x_n$ are the given n observations then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$Mean = \frac{Sum\ of\ the\ items}{Number\ of\ the\ items} = \frac{\sum X}{N}$$

Arithmetic Mean ($\bar{x}$)

- if the observations $x_1 + x_2 + x_3 + ...+ x_n$ are repeated $f_1 + f_2 + f_3 + ...+ f_n$ times, then we have

$$(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i} \text{ i. e. } \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \dots + f_n x_n}{f_1 + f_2 + f_3 + \dots + f_n}$$

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

- In case of continuous or grouped frequency distribution, the value of x is taken as the mid-point of the corresponding class.

Calculation of Arithmetic Mean (Individual Observations)

- Steps:

1. Add together all the values of the variables X and obtain the total i.e $\sum X$
2. Divide this total by the number of the observations i.e. N

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Arithmetic Mean ($\bar{x}$)

- A.M. is very simple and therefore, commonly used in business and economics.
- Whenever there is a mention of average income, profit, wages, output, the reference is to the arithmetic mean unless there are some qualifying words suggesting some other type of average

Characteristics of Arithmetic Mean

- If the number of items and their A.M. are both known, the aggregate or the sum of items can be obtained by multiplying the average by the number of items i.e.

$$\bar{x} = \frac{\sum X}{N} \text{ or } N\bar{x} = \sum X$$

Characteristics of Arithmetic Mean

- If an observation equal to the mean is excluded, the A.M. of the remaining observations remains unchanged. Similarly, if an observation equal to the mean is added to the series, the average of the total observations remains unchanged.
- If a wrong figure is taken in the computation of A.M., the correction can be made without repeating the entire calculation.

Characteristics of Arithmetic Mean

- The A.M. has two important mathematical properties which provides further mathematical analysis easy and which have made its use more popular than any other type of average.
  - The algebraic sum of the deviations of a given set of individual observations from the A.M. is always zero. Symbolically $\sum(X - \overline{X}) = 0$
  - The sum of squares of deviations of a set of observations is the minimum when deviations are taken from the A.M. symbolically $\sum(X - \overline{X})^2$ is least

Characteristics of Arithmetic Mean

- If each of the value of a variate X is increase (or decreased) by a constant b, the A.M. also increase (or decreased) by the same amount. And when the values of X are multiplied by a constant say k, the A.M. is also multiplied by the same amount k.

Short-cut Method Of Computing A.M.

- Here an assumed or an arbitrary average (indicated by 'A' or 'a') is used as the basis of calculation of deviation from individual values.
- The formula is $\overline{X} = A + \frac{\sum d}{N}$
  - Where A = the assumed mean or arbitrarily selected value
  - d = the deviation of each value from the assumed mean i.e. d = (X – A)
  - ∑d / N is the correction for the difference between the actual average and the assumed average. The ∑d / N = 0 if the assumed average is equal to the actual average.

Short-cut Method Of Computing A.M. For Grouped Data

1. Select the assumed mean A.
2. Find the deviation of each class mid-point from the assumed mean in original units of data, such as Rs. , cm, and so on.
   - Let d = the deviation in original units
   - m = each class mid-point
   - ∴ d = m – A
3. Multiply each deviation d by the frequency in the class to obtain the total deviations of the class or fd
4. Add these products to obtain the total deviations of all items included in the distribution, or ∑fd by the sum of the frequencies (∑f or N) to obtain the correction factor

$$\frac{\sum fd}{\sum f} \text{ or } \frac{\sum fd}{N}$$

Short-cut Method Of Computing A.M. For Grouped Data

5. Add the correction factor to the assumed mean to obtain exact mean of the grouped data.

$$\overline{X} = A + \frac{\sum fd}{\sum f} \quad or \quad \overline{X} = A + \frac{\sum fd}{N}$$

Calculation of A.M. For Discrete Data

- Direct method :
  - The formula for computing mean is

$$\overline{X} = \frac{\sum fx}{\sum f} \quad or \quad \overline{X} = \frac{\sum fx}{N}$$

  - Where f = frequency
  - x = the variable in question
  - N = total number of observations i.e. $\sum f$

Calculation of A.M. For Discrete Data

- Direct method steps:
  1. Multiply the frequency of each row with the variable and obtain the total $\sum fx$
  2. Divide the total obtained by step (1) by the number of observation i.e. total frequency.

Calculation of A.M. For Discrete Data

- Short-cut method :
  - The formula for computing mean is

$$\overline{X} = A + \frac{\sum fd}{\sum f} \quad or \quad \overline{X} = A + \frac{\sum fd}{N}$$

  - Where f = frequency
  - A = assumed mean
  - d = x - A
  - N = total number of observations i.e. $\sum f$

Calculation of A.M. For Discrete Data

- Short - cut method steps:
  1. Take an assumed mean
  2. Take deviations of the variables x from the assumed mean and denote the deviation by d
  3. Multiply the deviation with the respective frequency and obtain the total $\sum fd$
  4. Divide the total obtained by step (3) by the number of observation i.e. total frequency plus the assumed mean.

$$\overline{X} = A + \frac{\sum fd}{\sum f} \quad or \ \overline{X} = A + \frac{\sum fd}{N}$$

Calculation of A.M. For Continuous Data

- Direct method :
  - The formula for computing mean is

$$\overline{X} = \frac{\sum fm}{N}$$

  - Where f = frequency
  - m = mid-point of each class
  - N = total frequency i.e. $\sum f$

Calculation of A.M. For Continuous Data

- Direct method steps:
  1. Obtain the mid-point of each class and denoted it by m.
  2. Multiply these mid-points by the respective frequency of each class and obtain the total $\sum fm$.
  3. Divide the total obtained in step (2) by the sum of the frequency i.e. N

Calculation of A.M. For Continuous Data

- Short-cut method :
  - The formula for computing mean is

$$\overline{X} = A + \frac{\sum fd}{N}$$

  - Where f = frequency
  - m = mid-point of each class
  - A = assumed mean
  - d = deviation of mid-points from the assumed mean i.e. d = m - A
  - N = total frequency i.e. $\sum f$

Calculation of A.M. For Continuous Data

- Short-cut method steps:
  1. Take an assumed mean
  2. Obtain the mid-point of each class and denoted it by m.
  3. From the mid-point of each class deduct the assumed mean.
  4. Multiply these deviations by the respective frequency of each class and obtain the total ∑fd.
  5. Apply the formula $\overline{X} = A + \frac{\sum fd}{N}$

Calculation of A.M. For Continuous Data

- Step- deviation method :
  - In case of grouped or continuous frequency distribution, with class intervals of equal magnitude. The calculation are further simplified by taking d' = d / h = (m – A) / h
  - Where f = frequency
  - m = mid-value of each class
  - A = assumed mean
  - d = deviation of mid-points from the assumed mean i.e. d = m – A
  - h = is the common magnitude of the *class-i*ntervals
  - N = total frequency i.e. ∑f

$$\overline{X} = A + \frac{\sum fd'}{N} * h$$

Calculation of A.M. For Continuous Data

- Step-deviation method steps:
  1. Compute d' = (m – A) / h. Where A being any assumed mean, h is the common magnitude (class-interval) of the class. Algebraic sign + or – are to be taken with deviations.
  2. Multiply d' by the corresponding frequency f to get fd'
  3. Find the sum of the products obtained in step (2) to get ∑fd'.
  4. Divide the sum obtained in step (3) by N, the total frequency.
  5. Add A to the value obtained in step (4)

Advantages of Step-Deviation Method

- All three methods for continuous series gives same answer.
- The direct method though the simplest, involves more calculations when mid-points and frequencies are very large in magnitude. In this case step-deviation method would be far simpler.

Weighted A.M.

- For calculating simple A.M. , we suppose that all the values of size of items in the distribution have equal importance.
- In case some items are more important than others, a simple average computed is not representative of the distribution.
- In such cases proper weight age has to be given to the various items, the weight attached to each item being proportional to the importance of the item in the distribution.
- The term 'weight' stands for the relative importance of different items.

Weighted A.M.

- The formula for the weighted A.M. is given by
  - Direct Method : $$\overline{x_w} = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum wx}{\sum w}$$
  - Short cut method: $$\overline{x_w} = Aw + \frac{\sum wd}{\sum w}$$
    - Where Aw = assumed (weighted mean)
    - $\sum$wd = sum of the product of the deviations from the assumed mean (Aw) multiply by their respective weights.
  - In case of frequency distribution $$\overline{x_w} = Aw + \frac{\sum w(fx)}{\sum w}$$

Weighted A.M.

- Weights may be either actual or arbitrary i.e. estimated.

1. Simple A.M. shall be equal to the weighted A.M. if the weights are equal. $\bar{x} = \overline{x_w}$
2. Simple A.M. shall be less than the weighted A.M. if and only if greater weights are assigned to greater values and smaller weights are assigned to smaller values. $\bar{x} < \overline{x_w} \ \ if \ \ (w_2 - w_1)(x_1 - x_2) < 0$

Weighted A.M.

3. Simple A.M. is greater than the weighted A.M. if and only if smaller weights are attached to the higher values and greater weights are assigned to smaller values. $\bar{x} > \overline{x_w} \ \ if \ \ (w_2 - w_1)(x_1 - x_2) < 0$
4. It may be noted that weighted A.M. is specially useful in problems relating to:
   - Construction of index number
   - Standardized birth and death rates
   - Comparison of results of two or more universities where number of students differ

Geometric Mean

- The G.M. of a series containing N observations is the N^{th} root of the product of the value.
- Ungrouped data : $$G.M. = \sqrt[n]{\text{the product of n value}}$$ $$G.M. = \sqrt[n]{x_1 * x_2 * ... * x_n}$$
- When the number of observations exceed two, the computation are simplified through the use of logarithms, the above formula may be written as

$$\log G.M. = \frac{1}{n}\log(x_1, x_2, ..., x_n)$$
$$\log G.M. = \frac{1}{n}[\log x_1 + \log x_2 + ... + \log x_n]$$

Geometric Mean

$$\log G.M. = \frac{1}{n}\log(x_1, x_2, ..., x_n)$$
$$\log G.M. = \frac{1}{n}[\log x_1 + \log x_2 + ... + \log x_n]$$

- Taking antilog on both sides we have

$$G.M. = \text{Antilog}\ [\frac{1}{n}\sum \log x]$$

- In discrete series $G.M. = \text{Antilog}\ [\frac{\sum f * \log x}{N}]$
- In continuous series $G.M. = \text{Antilog}\ [\frac{\sum f * \log m}{N}]$

Calculation Of G.M. (Individual observations)

$$G.M. = \text{Antilog}\ [\frac{\sum \log x}{N}]$$

- Steps:
  – Take the logarithms of the variable X and obtain the total ∑logX.
  – Divide ∑logX by N and take the antilog of the value so obtained. This gives the value of G.M.

Calculation Of G.M. (Discrete Series)

$$G.M. = \text{Antilog}\ [\frac{\sum f * \log x}{N}]$$

- Steps:
  – Find the logarithms of the variable X
  – Multiply these logarithms with the respective frequencies and obtain the total ∑f*logX.
  – Divide ∑f*logX by total frequency and take the antilog of the value so obtained. This gives the value of G.M.

Calculation Of G.M. (Continuous Series)

$$G.M. = \text{Antilog}\left[\frac{\sum f * \log m}{N}\right]$$

- Steps:
  - Find out the mid-points of the classes and take their logarithms.
  - Multiply these logarithms with the respective frequencies and obtain the total ∑f*logm
  - Divide total obtained in step 2 by the total frequency and take the antilog of the value so obtained. This gives the value of G.M.

Harmonic Mean

- H.M. is defined as the reciprocal of the A.M. of the reciprocal of the given observations.

$$H.M. = \frac{N}{\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}\right)} = \frac{N}{\sum 1/x}$$

Calculation of H.M. (Individual Observations)

$$H.M. = \frac{N}{\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}\right)} = \frac{N}{\sum 1/x}$$

- Where x_1, x_2, …, x_n etc refers to the various items

Calculation of H.M. (Discrete Series)

$$H.M. = \frac{N}{\sum f * \frac{1}{x}} = \frac{N}{\sum \frac{f}{x}}$$

- Steps:
  - Take the reciprocal of the various items of the variable x
  - Multiply the reciprocal by the frequencies and obtain the total $\sum(f * 1/x)$
  - Substitute the values of N and $\sum(f * 1/x)$ in the above formula

Calculation of H.M. (Continuous Series)

$$H.M. = \frac{N}{\sum \frac{f}{m}}$$

- Steps:
  - The reciprocal of the mid values of the class intervals (m) are found
  - Multiply the reciprocal with the respective class frequencies and obtain the total ∑(f * 1/m)
  - The total of the product is divided by the total number of items and the reciprocal of the resultant figure is the H.M. = Reciprocal [(∑(f * 1/m))/N]

Positional Average

- These averages are based on the position of a given observations in a series, arranged in an ascending or a descending order.
- The magnitude or the size of the values does not matter as was in the case of earlier averages.
- Ex. Median and Mode

Median

- The median, as the name suggests, is the middle value of a series arranged in any of the order of the magnitude.
- The middle value in the case of a series with odd number of items can be easily located e.g. the 6th value if the number of items is 11.
- But in case the total number is even, say 10 there will be two middle values, viz, 5th and 6th and in which case the mean of the two middle values shall constitute the median.
- The median is just 50th percentile value below which 50% of the values in the sample fall.

Median

- It splits the observations into two halves.
- The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median -------- L.R. Cannor
- The central values of the distribution, a value such that, greater and smaller values occur with equal frequency, is known as median. ---- Kenney and Keeping

Computation Of Median

- In calculation of median, there are two stages
  - Arrange the data in ascending or descending order of magnitude (Both arrangement gives same answer)
  - The search for the middle item which is indicated by $(N+1)/2$ or $N/2^{th}$ item determined on the basis of the total number of items. The value of this middle item is the median value

Computation Of Median (Ungrouped Data/ Individual Observations)

- In case of ungrouped data, median is the middle value of the series arranged in either ascending or descending order.
- Taking the total items equal to N items, median value (abbreviated as Md) is the value of the $(N+1)/2^{th}$ item.

Computation Of Median (Grouped Data)

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

Computation Of Median (Discrete Series)

1. Arrange the data in ascending or descending order of magnitude.
2. Find out the cumulative frequencies [less than type]
3. Apply the formula : Median = size of (N+1)/2.
4. Now look at the cumulative frequency column and find that total which is either equal to (N+1)/2 or next higher to that and determine the value of the variable corresponding to it. That gives the value of median.

Computation Of Median (Continuous Series)

- The method is the same as shown for the discrete frequency distribution. The two additional points are:

1. In a continuous distribution the value of N/2th item is taken and not of (N+1)/2th item. The reason is that in a continuous distribution only N+1 class limit can make N class division.
2. Having discovered the class in which the median value lies, the exact value has to be interpolated from the relevant class or on the basis of the position of N/2th value in the total frequency concentration in that class. The formula used for this purpose is as follows:

Computation Of Median (Continuous Series)

$$\text{Median} = l_1 + \frac{\frac{N}{2} - C}{f} * h$$

where l_1 = real lower limit of the median class
N/2 = item whose value has to be interpolated
C = cumulative frequency of class preceding the median class
f = frequency of the median class
h = size of the class interval in the median class

Usefulness of Median

- The median is useful for distribution containing open-end intervals.
- It is used when we require a measure of location which is not affected by high or low value item, and when we wish to measure the change in different sets of distribution which move in a similar direction in similar manner.

Mode

- The mode refers to that value in a distribution which occurs most frequently (greatly frequency)
- Mode is the value occurring most frequency in a set of observations and around with other items of the set cluster most densely.
- Its importance is very great in marketing studies where a manager is interested in knowing about the size which has the highest concentration of items. [Ex. Size of shoe]
- It is not affected by extreme values.

Computation Of The Mode (Ungrouped Data/ Individual Observations)

- The data is placed in the form of an array so that items having the same values can be identified and quickly counted, the value of that item which occurs most frequently is the modal value.

Computation Of The Mode (Grouped Data Discrete Series)

- In uni-modal distribution where the highest concentration is in a single discrete value there should not be any difficult in locating modal value or a class-interval containing this value by inspection.
- Some difficulty arise when nearly equal concentrations are found in two or more neighbouring values. There are two ways of dealing with this situations.
  - In a large majority of cases it shall be possible to make a choice of one value by taking the totals of 3 values, the value with highest concentration and it two neighbouring values in case of the competing cases. The central values of the group which yield higher total should be selected.

Computation Of The Mode (Discrete Series)

1. We prepare a grouping table with 6 columns
2. In column I we write down the frequency against the respective items.
3. In column II, the frequency is grouped in twos, starting from the top. Their totals are found out and the highest total is marked.
4. In column III, the frequency is again grouped in two, leaving the first frequency. Highest total is again marked.
5. In column IV, the frequency is grouped in three starting from the top and their totals found out with highest frequency is marked.

Computation Of The Mode (Discrete Series)

6. In column V, the frequencies are again grouped in three, leaving the first frequency. The totals are found out and the highest total is marked.
7. In column VI, leaving the first and second frequency, group is done in threes. After finding their total, the highest total is marked.
8. In analysis table, in order to find out the item which repeats largest number of items, grouping in analyzed.

Computation Of The Mode (Continuous Series Interpolation Formula)

- The exact value of mode in the case of continuous frequency distribution can be obtained by the following formulae:

$$\text{Mode} = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} * h$$

where l_1 = the real lower limit of the modal class
h = magnitude of the modal class
$\Delta_1 = f_1 - f_0$
$\Delta_2 = f_1 - f_2$
f_1 = frequency of the modal class
f_0 = frequency of the class preceding the modal class
f_2 = frequency of the class succeeding the modal class

Computation Of The Mode (Continuous Series Interpolation Formula)

$$\text{Mode} = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} * h$$

the above formula can be written as

$$\text{Mode}(Mo) = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} * h$$

Empirical Relationship Between Average

- Karl Pearson's formula
  - Mean = Mode – 3(Mean – Median)
  - Mean = Mode + 3/2 (Median - Mode) = ½(3Median – Mode)
  - Mode = 3Median – 2 Mean
  - Median = Mode + 2/3(Mean – Mode) = 1/3(2Mean + Mode)
- Unless the values of all items in a distribution are identical the G.M. will be smaller than the A.M. and the H.M. shall be smaller than both G.M. and A.M.
  - A.M. ≥ G.M. ≥ H.M.

Further Partition Of Series

- Just as a median divides the distribution into two parts, there are other positional measures which partition a series into still smaller parts say 4, 10 or 100.
- The value which divides the series into a number of equal parts are called partition value
- The median divides the lower 50% of a set of data from the upper 50% of a set of data.

Quartiles

- The quartiles divide data sets into fourths or four equal parts.
- There are 3 quartiles.
- The 1st quartile, denoted Q_1, divides the bottom 25% the data from the top 75%. Therefore, the 1st quartile is equivalent to the 25th percentile.
- The 2nd quartile divides the bottom 50% of the data from the top 50% of the data, so that the 2nd quartile is equivalent to the 50th percentile, which is equivalent to the median.
- The 3rd quartile divides the bottom 75% of the data from the top 25% of the data, so that the 3rd quartile is equivalent to the 75th percentile.

Deciles

- These are the values which divides the total number of observations into 10 equal parts.
- Obviously there are 9 deciles, D_1, D_2, D_3, ….., D_9
- These are called as first deciles, second deciles etc. Also $D_1 < D_2 < \ldots. < D_9$; $D_5 = Q_2 =$ Median

Percentiles

- The percentile value divide the distribution into 100 parts each containing 1 percent of the observations.
- A division in so many parts is used only when there are a considerable number of observations, at least one thousand, otherwise there would be too few observations in each division to be significant.
- In general, the ***k*th percentile**, denoted P_k , of a set of data divides the lower *k*% of a data set from the upper (100 – *k*) % of a data set.
- In particular $P_{10} = D_1$; $P_{20} = D_2$; $P_{30} = D_3$
  $P_{25} = Q_1$; $P_{50} = Q_2 = D_5$ = Median ; $P_{75} = Q_3$
  $P_1 \leq P_2 \leq \ldots\ldots \leq P_{99}$

Computation Of Partition Values (Discrete Series)

- The quartile, decile and percentile are determined by the same technique which is used in the computation of median.
- For quartile the value of (N+1)/4th item, for deciles, the value of (N+1)/10th item and for percentile the value of (N+1)/100th item is used.
- Now for say 2nd and 3rd quartiles, we have

$$2 * \left(\frac{N+1}{4}\right)^{th} \ item$$

$$3 * \left(\frac{N+1}{4}\right)^{th} \ item$$

Computation Of Partition Values (Discrete Series)

- Similarly for 3rd and 4th deciles, we have

$$3*\left(\frac{N+1}{10}\right)^{th} \text{ item}$$

$$4*\left(\frac{N+1}{10}\right)^{th} \text{ item}$$

- Again say 33rd and 44th percentiles, can be computed using formula

$$33*\left(\frac{N+1}{100}\right)^{th} \text{ item}$$

$$44*\left(\frac{N+1}{100}\right)^{th} \text{ item}$$

Computation Of Partition Values (Continuous Series)

- In case of a continuous series N+1 would be replaced by N. The formula for interpolation will be

$$l_1 + \left(\frac{kN \quad - C}{f}\right) * h$$

  - Where kN is the item number depending on the type of partition sought e.g. it is 3N/10 or 33N/100 for 3rd decile and 33rd percentile, etc.
  - C is the cumulative frequency of the class previous to the class in which 3N/10th or 33N/100th item lies.
  - 'f' is the frequency of the relevant class referred to above.
  - 'h' is the size of class interval of the relevant class referred above. .

Computation Of Quartiles

1. Compute N/4 where N= Σf
2. Find out the cumulative frequency (less than) just greater than N/4.
3. The corresponding class contains Q_1 and the value of Q_1 is obtained by the interpolation formula:

$$l_1 + \left(\frac{\frac{N}{4} - C}{f}\right) * h$$

- Where l_1 is the lower limit of the class containing Q_1
- f is the frequency of the class containing Q_1
- h is the magnitude of the class containing Q_1
- C is the cumulative frequency of the class preceding the

Computation Of Quartiles

- Similarly to compute Q_3 we obtain cumulative frequency (less than) just greater than 3N/4 item, and the value of Q_3 is given by the formula

$$l_1 + \left(\frac{\frac{3N}{4} - C}{f}\right) * h$$

Computation Of Deciles

1. Compute the i[th] decile Di (i = 1, 2, …, 9) we find out cumulative frequency just greater than i * N/10 and the corresponding class contains Di and its value is obtained using the interpolation formula:

$$D_i = l_1 + \left(\frac{\frac{iN}{10} - C}{f}\right) * h; \text{ (I =1,2,…,9)}$$

- Where l_1 is the lower limit of the class containing D_i
- f is the frequency of the class containing D_i
- h is the magnitude of the class containing D_i
- C is the cumulative frequency of the class preceding the class containing D_i.

Computation Of Percentiles

1. Compute the i[th] percentile Pi (i = 1, 2, …, 99) we find out cumulative frequency just greater than i * N/100 and the corresponding class contains Pi and its value is obtained using the interpolation formula:

$$P_i = l_1 + \left(\frac{\frac{iN}{100} - C}{f}\right) * h; \text{ (i =1,2,…,99)}$$

- Where l_1 is the lower limit of the class containing P_i
- f is the frequency of the class containing P_i
- h is the magnitude of the class containing P_i
- C is the cumulative frequency of the class preceding the class containing P_i.

Minal Shah