Central tendency measures

Introduction

The description of statistical data may be quite elaborate or quite brief depending on two factors: the nature of data and the purpose for which the same data have been collected. While describing data statistically or verbally, one must ensure that the description is neither too brief nor too lengthy. The measures of central tendency enable us to compare two or more distributions pertaining to the same time period or within the same distribution over time. It is the most important objective of statistical analysis is to get one single value that describes the characteristics of the entire mass of cumbersome data. Such a value is finding out, which is known as central value to serve our purpose.

- **Arithmetic Mean ($\overline{X}$)**

Sometimes called arithmetic "mean" or "average". Generally, it is considered as the most important of those measures because of its good properties and computations facility.

i) Non-tabulated data

Let $x_1, x_2, x_3, \dots, x_n$ represent a data of X random variable of sample with size (n).

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

ii) Tabulated data

Let $x_1, x_2, x_3, \dots, x_m$ represent a centers of (m) classes in a frequency table, and $f_1, f_2, f_3, \dots, f_m$ that represent the corresponding frequencies to those classes with sample size (n).

So, the arithmetic mean can be obtained-for either equal or not equal class length-by:

$$\overline{X} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i}$$

- **Weighted Arithmetic Mean ($\overline{X}_w$)**

It used when there is a different importance for some components/ values from another. There for, these different magnitudes ought to taken under consideration during arithmetic mean computation. Suppose $x_1, x_2, x_3, \ldots, x_n$, represent sample of size (n), $w_1, w_2, w_3, \ldots, w_n$, represent weights of data.

i) Non-tabulated data

The formula is:

$$\overline{X}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

ii) Tabulated data

$$\overline{X}_w = \frac{\sum_{i=1}^{m} w_i f_i x_i}{\sum_{i=1}^{m} w_i f_i}$$

- **Harmonic mean (H)**

It used to minimize some data type by illustrating reciprocal idea and according to data type.

i) Non-tabulated data

The formula is:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

ii) Tabulated data

The formula is:

$$H = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{x_i}}$$

- **Quadratic mean (Q)**

It is defined as a positive squared root of mean squared values (or classes' center) for the random variable X.

i) Non-tabulated data

The formula is:

$$Q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

ii) Tabulated data

The formula is:

$$Q = \sqrt{\frac{\sum_{i=1}^{m} f_i x_i^2}{\sum_{i=1}^{m} f_i}}$$

- **Mode (M_o)**

It is the most commonly/ frequently value repeated among a set of X values.

i) Non-tabulated data

Let x_1, x_2, x_3, …, x_n represent a dataset of X random variable values of sample with size (n). thus, x_j is one of these values that the most commonly repeated, so x_j will be defined as the mode of this set.

ii) Tabulated data

Let x_1, x_2, x_3, …, x_m represent a centers of (m) classes in a frequency table, and f_1, f_2, f_3, …, f_m that represent the corresponding frequencies to those classes with sample size (n).

1- Discrete variable

The mode here is the modal class center, i.e., the center of the class that correspond to the largest frequency in the frequency distribution.

Remark: in case of there are two/more than two classes correspond the same largest frequency, therefore, it will be said that: the distribution has more than one mode.

2- Continuous variable

Suppose that we have a frequency distribution with (m) classes, and (f_k) represents the largest frequency in this distribution. That means, the modal class is that one which contains the mode value.

While (f_{k-1}) represents the previous frequency of modal class frequency, and (f_{k+1}) the posterior frequency of modal class frequency in another word; $f_{k-1} < f_k < f_{k+1}$.

$$M_o = L_k + \frac{f_k - f_{k-1}}{(f_k - f_{k-1}) + (f_k - f_{k+1})} \, . h_k$$

h_k: length of the modal class.

L_k: the lower limit of the modal class.

- **Median (M_e)**

Defined as: the value that splits the random variable values into two equal parts, i.e., it is x value the makes the number of values before it equals that after it.

i) Non-tabulated data

Let $x_1, x_2, x_3, \ldots, x_n$ represent a dataset of X random variable values of sample with size (n). suppose that x values had been ordered Ascending/Descending fashion, then we can compute the median according to sample size (n) either it is odd/even.

- Odd sample size (n): where the median (after data ordering) $X_{\frac{n+1}{2}}$, i.e., will be x value of $\frac{n+1}{2}$ the order.
- Even sample size (n): where the median will be merely the average of two values that; $X_{\frac{n}{2}}$ and $X_{\frac{n}{2}+1}$,i.e., the average of $(\frac{n}{2})$ th& $(\frac{n}{2}+1)$th values.

ii) Tabulated data (continuous variable)

Suppose that we have a frequency distribution of continuous X random variable with (m) classes, with $f_1, f_2, f_3, \ldots, f_m$ represent the corresponding frequencies. While $F_1, F_2, F_3, \ldots, F_m$ represent the ascending cumulative frequencies. We rely here on the lower limits of the classes. Therefore, median can be computed by finding the median order as: and then obtaining the ascending cumulative frequencies $F_1, F_2, F_3, \ldots, F_m$.

Then, we have to determine the median (k^{th}) class, and finally we will obtain median value as:

$$M_e = L_k + \left[\frac{\sum_{i=1}^{m} f_i}{2} - F_{k-1}\right] . \frac{h_k}{f_k}$$

L_k: lower limit of the median class.

F_{k-1}: previous ascending cumulative frequency corresponding to the median class.

f_k: original frequency corresponding to the median class.

h_k: length of median class.

- <u>**The relationship among some of central tendency measures**</u>

Some central tendency measures are related with each other: mean, mode and median.

1- If the frequency distribution is symmetric, then the following formula will be true: $\overline{X} = M_o = M_e$, where: mean, mode and median are all equal.
2- If the frequency distribution is skewed/ not-symmetric, the following formula is true:

$$\overline{X} - M_e = \frac{(\overline{X} - M_o)}{3}$$

It is a useful relation from several sides, such as: when the mean can't not be obtained for opened frequency distribution table (one side/ both sides), then we can use the following formula:

$$\overline{X} = \frac{3M_e - M_o}{2}$$

$$M_e = \frac{2\overline{X} + M_o}{3}$$

$$M_o = 3M_e - 2\overline{X}$$