CHAPTER-15

MEASURES OF CENTRAL TENDENCY AND DISPERSION.

CENTRAL TENDENCY:

* Central tendency may be defined as the tendency of a given set of observations to cluster around a single central or middle value and the single value that represents the given set of observation is described as a 'measure of central tendency' or location or average.
* It is possible to condense a vast mass of data by a single representative value.

DIFFERENT MEASURES OF CENTRAL TENDENCY:

* Arithmetic mean (AM)
* Median (Me)
* Mode (Mo)
* Geometric mean (GM)
* Harmonic mean (HM)

CRITERIA FOR AN IDEAL MEASURE OF CENTRAL TENDENCY :

* Properly & unambiguously defined.
* Easy to comprehend.
* Simple to compute.
* Based on all the observations.
* Have certain desirable mathematical properties.
* Least affected by the presence of extreme observations.

ARITHMETIC MEAN :

* AM may be defined as the sum of all the observations divided by the number of observations.
* Best and most commonly used measure of central tendency.

MEADIAN :

* Median is a positional average.
* Median is dependent upon the position of the given set of observations.
* It may be defined as the middle most value.
* It does not affected by the presence of extreme observations.

* In case of a grouped frequency, we find median from the cumulative frequency distribution of the variable under consideration.

PARTITION VALUES:

* Quartiles are values dividing a given set of observations into four equal parts.
* Deciles are the values dividing a given set of observations into ten equal parts.
* Percentiles or centiles are the values dividing a given set of observations into 100 equal parts.

MODE:

* Mode may be defined as the value that occurs the maximum number of times.
* Mode is that value which has the maximum concentration of the observations around it.
* The distribution which has more than one mode is called multi-modal distribution.
* The distribution which has two modes is called bi-modal distribution.
* Sometimes, mode is not defined when there is no -modal mark.

* Mode is the most popular measure of central tendency.

GEOMETRIC MEAN:

* GM may be defined as the n^{th} root of the product of the observations.
* GM have limited applications for the computation of average rates & ratios.
* GM can be calculated for set of positive observations.

HARMONIC MEAN:

* HM may be defined as the reciprocal of the AM of the reciprocal of the observation.
* HM have limited application for the computation of average rates & ratios.
* HM can be calculated only for set of non-zero observations.

WEIGHTED AVERAGE: All the observations are not of equal importance.

* When the observations have the hierarchical order of importance, weighted average will be computed which could be either weighted AM or weighted GM or weighted HM.

MEASURES OF CENTRAL TENDENCY	UNGROUPED FREQUENCY DISTRIBUTION	GROUPED FREQUENCY DISTRIBUTION	WEIGHTED AVERAGE
AVERAGE MEAN	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$	$\bar{x} = \frac{\sum f_i x_i}{N}$ where $N = \sum f_i$ $\bar{x} = A + \frac{\sum f_i d_i}{N} \times C$ where $d_i = \frac{x_i - A}{C}$ A = Assumed mean C = Class length.	Weighted AM $= \frac{\sum w_i x_i}{\sum w_i}$
GEOMETRIC MEAN	$G = (x_1 \times x_2 \times x_3 \times \cdots \times x_n)^{1/n}$	$G = (x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \times \cdots \times x_n^{f_n})^{\frac{1}{N}}$ where $N = \sum f_i$	Weighted GM $= \text{Antilog}\left(\frac{\sum w_i \log x_i}{\sum w_i}\right)$
HARMONIC MEAN	$H = \frac{n}{\sum \left(\frac{1}{x_i}\right)}$	$H = \frac{N}{\sum \left(\frac{f_i}{x_i}\right)}$	Weighted HM $= \frac{\sum w_i}{\sum \left(\frac{w_i}{x_i}\right)}$

MEASURE OF CENTRAL TENDENCY	UNGROUPED FREQUENCY DISTRIBUTION	GROUPED FREQUENCY DISTRIBUTION
MEDIAN	Odd : Me = Middle number. Even : Me = Average of two middle numbers.	$Me = l_1 + \left(\frac{N/2 - N_L}{N_u - N_L}\right) \times C$
QUARTILE	$Q = (n+1)p^{th}$ value.	$Q = l_1 + \left(\frac{N_p - N_L}{N_u - N_L}\right) \times C$ where, $N_p = N \times P$
MODE	Value that occurs maximum number of time. Sometimes mode is undefined.	$\text{Mode} = l_1 + \left(\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1}\right) \times C$

PROPERTIES :

ARITHMETIC MEAN	GEOMETRIC MEAN	HARMONIC MEAN	MEDIAN	MODE
If all the observations are constants, say k, then AM is also k.	If all the observations are constants, say $k > 0$, then GM is also k.	If all the observations are constants, say k, then HM is also k.		
Combined AM for 2 groups: $\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$		Combined HM for two groups: $HM = \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$		
If, $y = a + bx$ $\bar{y} = a + b\bar{x}$			If, $y = a + bx$, $y_{me} = a + bx_{me}$	If, $y = a + bx$, $y_{mo} = a + bx_{mo}$
Sum of deviations is 0. i.e., $\Sigma(x_i - \bar{x}) = 0$ $\Sigma f_i(x_i - \bar{x}) = 0$			Sum of absolute deviations is minimum, i.e., $\Sigma\lvert x_i - A\rvert$ is minimum if we choose A as the median.	
	If $z = x.y$; GM of z = (GM of x) $\times$ (GM of y)			
	If $z = x/y$; GM of $z = \frac{\text{GM of } x}{\text{GM of } y}$			
	$\text{Log } G = \frac{1}{n} \Sigma \log x$			

	AM	GM	HM	Me	Mo
RIGIDLY DEFINED	Yes	Yes	Yes	Yes	No
BASED ON ALL OBSERVATIONS	Yes	Yes	Yes	No	No
EASY TO COMPHREHEND	Yes	No	No	Yes	Yes
SIMPLE TO CALCULATE	Yes	No	No	Yes	Yes
MATHEMATICAL PROPERTIES	Yes	Yes	Yes	No	No
AFFECTED BY SAMPLING FLUCTUATIONS.	Yes	-	-	No	Yes.

RELATIONSHIP BETWEEN MEAN, MEDIAN & MODE:

$$\text{Mean} - \text{mode} = 3(\text{Mean} - \text{median})$$

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}.$$

RELATIONSHIP BETWEEN AM, GM, HM:

$$AM \times HM = GM^2$$

$$AM \geq GM \geq HM$$

[Equality sign occurs when all the observations are equal]

DISPERSION:

* Dispersion for a given set of observations may be defined as the amount of deviation of the observations, usally, from an appropriate measures of central tendency.
* Dispersion measures scatterness of a set of observations.

MEASURES OF DISPERSION:

* Absolute measures of dispersion.
  - i) Range
  - ii) Mean deviation
  - iii) Standard deviation
  - iv) Quartile deviation.
* Relative measures of dispersion.
  - i) Coefficient of Range.
  - ii) Coefficient of mean deviation.
  - iii) Coefficient of standard deviation.
  - iv) Coefficient of Quartile deviation.

CHARACTERISTICS FOR AN IDEAL MEASURE OF DISPERSION

* Properly defined.
* Easy to comprehend.
* Simple to compute.

* Based on all the observations.
* Unaffected by sampling fluctuations
* Have certain desirable mathematical properties.

DIFFERENCE BETWEEN ABSOLUTE AND RELATIVE MEASURES OF DISPERSION:

ABSOLUTE MEASURE OF DISPERSION	RELATIVE MEASURE OF DISPERSION
* Dependent on the unit of the variable under consideration.	* Unit free.
* Not used for comparing two or more distributions.	* Helpful in comparing two or more distributions.
* Easy to compute	* Difficult to compute.

RANGE:

* Range may be defined as the difference between largest and smallest of observations.

MEAN DEVIATION :

* Since range is based on only two observations, it is not regarded as an ideal measure of dispersion.
* A better measure of dispersion is provided by mean deviation which, unlike range, is based on all the observations.
* It is based on absolute deviations only.

STANDARD DEVIATION :

* Although mean deviation is an improvement over range so far as a measure of disperson is concerned, mean deviation is difficult to compute and further more, it cannot be treated mathematically.
* The best measure of dispersion is, usually, standard deviation which does not possess the demerits of range & mean deviation.
* It is defined as root mean square deviation.
* Standard deviation is considered for finding a pooled measure of dispersion after combining several groups.
* For any two numbers, SD is always half of the range.

* It is the most widely & commonly used measures of dispersion.
* SD of first n natural numbers is $SD = \sqrt{\frac{n^2-1}{12}}$
* SD of any two numbers is $\frac{|a-b|}{2}$

VARIANCE :

* Square of standard deviation is known as variance.
* To find the more consistent structure, we compare the coefficient of variance of two distributions.
* If $CV_a < CV_b$, then distribution a is more consistent.

QUARTILE / SEMI - INTER QUARTILE DEVIATION :

* Best measure of dispersion for open-end classification.
* It is less affected due to sampling fluctuations.
* Remains unaffected due to change of origin but is affected in the same ratio due to change in scale.
* It is based on the central-fifty-percent of the observations.

BASIS	AM	MEDIAN	MODE	GM	HM
SHIFT OF ORIGIN	Dependent New AM = $\bar{x} \pm q$	Dependent New Me = Me $\pm$ q	Dependent New Mo = Mo $\pm$ q	Dependent Method unknown	Dependent Method unknown
CHANGE OF SCALE	Dependent New AM = $\frac{\bar{x}}{q} \times q$	Dependent New Me = $\frac{Me}{q} \times q$	Dependent New Mo = $\frac{Mo}{q} \times q$	Dependent New GM = $\frac{GM}{q} \times q$	Dependent New HM = $\frac{HM}{q} \times q$