MEASURES OF CENTRAL TENDENCY AND DISPERSION

DEFINITION OF 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 observations is described as a measure of central tendency or, location, or average.

Following are the 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

- It should be properly and unambiguously defined.
- It should be easy to comprehend.
- It should be simple to compute.
- It should be based on all the observations.

MEASURES OF CENTRAL TENDENCY AND DISPERSION

- It should have certain desirable mathematical properties.
- It should be 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. Thus, if a variable x assumes n values x1 , x2 , x3 ,..........xn, then the AM of x, to be denoted by X , is given by :

$$\sum x = \frac{x}{N}$$

Grouped frequency distribution: $\bar{X} = \frac{\sum f_i x_i}{N}$

Properties of AM

- If all the observations assumed by a variable are constants, say k, then the AM is also k.
- the algebraic sum of deviations of a set of observations from their AM is zero.
- AM is affected due to a change of origin and/or scale.

- If there are two groups containing n1 and n2 observations and x 1 and x 2 as the respective arithmetic means, then the combined AM is given by

Combined AM $= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

MEDIAN – PARTITION VALUES

Median is a positional average which means that the value of the median is dependent upon the position of the given set of observations for which the median is wanted $\left(\frac{n+1}{2}\right) th$ obsv

$$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u N_l}\right) \times C$$

Grouped frequency distribution

$$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l}\right) \times C$$

Properties of median

- Ω If x and y are two variables, to be related by y=a+bx for any two constants a and b, then the median of y is given by Y_{me} = a + bx_{me}
- Ω For a set of observations, the sum of absolute deviations is minimum when the deviations are taken from the median. This property states that -|xi –A| is minimum if we choose A as the median.

PARTITION VALUES OR QUARTILES OR FRACTILES

Quartiles are values dividing a given set of observations into four equal parts. So there are three quartiles – first quartile or lower quartile denoted by Q1 , second quartile or median to be denoted by Q2 or Me and third quartile or upper quartile denoted by Q3 . First quartile is the value for which one fourth of the observations are less than or equal to Q1 and the remaining three – fourths observations are more than or equal to Q1.

Prof. Jatin Dembla

Deciles are the values dividing a given set of observation into ten equal parts. Thus, there are nine deciles to be denoted by D1 , D2 , D3 ,.....D9 . D1 is the value for which one-tenth of the given observations are less than or equal to D1 and the remaining nine-tenth observations are greater than or equal to D1 when the observations are arranged in an ascending order of magnitude.

Percentiles or centiles that divide a given set of observations into 100 equal parts. The points of sub-divisions being P1 , P2 ,...........P99. P1 is the value for which one hundredth of the observations are less than or equal to P1 and the remaining ninety-nine hundredths observations are greater than or equal to P1 once the observations are arranged in an ascending order of magnitude.

the pth quartile is given by the (n+1)pth value, where n denotes the total number of observations. p = 1/4, 2/4, 3/4 for Q1, Q2 and Q3 respectively. p=1/10, 2/10,............9/10. For D1 , D2 ,......,D9 respectively and lastly p=1/100, 2/100,....,99/100 for P1 , P2 , P3 ....P99 respectively.

MEASURES OF CENTRAL TENDENCY AND DISPERSION

MODE

set of observations, mode may be defined as the value that occurs the maximum number of times. Thus, mode is that value which has the maximum concentration of the observations around it.

Mode = $l_1 + \left(\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1}\right) \times c$

When it is difficult to compute mode from a grouped frequency distribution, we may consider the following empirical .

Moderately Skewed Distribution.

- Mean – Mode = 3(Mean – Median)
- or Mode = 3 Median – 2 Mean

Symmetric Distribution

Mean = Median = Mode

Prof. Jatin Dembla

MEASURES OF CENTRAL TENDENCY AND DISPERSION

GEOMETRIC MEAN AND HARMONIC MEAN

the geometric mean is defined as the n-th root of the product of the observations. Thus if a variable x assumes n values x1 , x2 , x3 ,………., xn, all the values being positive, then the GM of x is given by

$G = (x1 \times x2 \times x3 \ ……… \times x_n)^{1/n}$

harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observation. So, if a variable x assumes n non-zero values x1 , x2, x3 ,……………,xn, then the HM of x is given by

$H = \frac{n}{\sum(1/x_i)}$

Combined harmonic mean = $\frac{n_1+n_2}{\frac{n_1}{h_1}+\frac{n_2}{h_2}}$

Relation between AM , GM , HM

Observation are difference a≠ b	AM > GM > HM
Observation are Same a = b	AM ≥ GM ≥ HM

MEASURES OF CENTRAL TENDENCY AND DISPERSION

Weighted average

Weighted AM	Weighted GM	Weighted HM
$\frac{\sum w_i x_i}{\sum w_i}$	Ante log $\left(\frac{\sum w_i \log x_i}{\sum w_i}\right)$	$\frac{\sum w_i}{\sum\left(\frac{w_i}{x_i}\right)}$

DEFINITION OF DISPERSION

Dispersion for a given set of observations may be defined as the amount of deviation of the observations, usually, from an appropriate measure of central tendency. Measures of dispersion may be broadly classified into

- **Absolute measures of dispersion**
  - Range
  - Mean Deviation
  - Standard Deviation
  - Quartile Deviation

- **Relative measures of dispersion.**
  - Coefficient of Range.
  - Coefficient of Mean Deviation
  - Coefficient of Variation
  - Coefficient of Quartile Deviation.

Distinction between the absolute and relative measures of dispersion:

- Absolute measures are dependent on the unit of the variable under consideration whereas the relative measures of dispersion are unit free.

UNIT II: DISPERSION

- For comparing two or more distributions, relative measures and not absolute measures of dispersion are considered.
- Compared to absolute measures of dispersion, relative measures of dispersion are difficult to compute and comprehend.

RANGE

Range = L – S

Coff. Range	$\frac{L-S}{L+S} \times 100$

MEAN DEVIATION

mean deviation is defined as the arithmetic mean of the absolute deviations of the observations from an appropriate measure of central tendency.

$$\frac{\sum|x-\bar{x}|}{n}$$

Coff. M.D	$\frac{md}{\bar{x}} \times 100$

Prof. Jatin Dembla

www.Kitest.in

Mean deviation takes its minimum value when the deviations are taken from the median. Also mean deviation remains unchanged due to a change of origin but changes in the same ratio due to a change in scale i.e. if y = a + bx, a and b being constants, MD of y = |b| × MD of x.

STANDARD DEVIATION

The best measure of dispersion is, usually, standard deviation which does not possess the demerits of range and mean deviation. Standard deviation for a given set of observations is defined as the root mean square deviation when the deviations are taken from the AM of the observations.

$$\sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

Coff. (σ^2)	$\frac{SD}{\bar{x}} \times 100$

Sum of square of first natural number $\sqrt{\frac{n^2-1}{12}}$

UNIT II: DISPERSION

QUARTILE DEVIATION

A relative measure of dispersion using quartiles is given by coefficient of quartile deviation which is $\frac{Q_3-Q_1}{2}$

Coff. Q.D.	$\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

Quartile deviation provides the best measure of dispersion for open-end classification. It is also less affected due to sampling fluctuations

Prof. Jatin Dembla