DISPERSION

➢DEFINITION

- The characteristic of scattering of observations is known as dispersion. The amount of scattering of a set of observations can be measured.
- For a set of observation with least dispersion, the mean of the set is a good representative of the entire set of observations. If the dispersion is high, the mean cannot be good representative. It is also to be noted that, for a set if all the observations are same to the mean of that set the dispersion for that set is zero.

➢Various methods to measure Dispersion

1. Range
2. Quartile Deviation (Semi Inter quartile range)
3. Mean Deviation
4. Standard Deviation

In addition to these measures, some situation requires measurement of dispersion graphically. There the method of Lorenz curve is used

➢Properties of good measure of dispersion

1. It should be rigidly defined
2. It should be simple to understand and to calculate
3. It should be based on all the observations
4. It should be capable of further algebraic treatment
5. It should have sampling stability
6. It should not be unduly affected by the extreme values

RANGE

- Range is the difference between the largest and the smallest of the given values
- For raw data,

  Range = Highest observation – Least observation

- For grouped frequency data,

  Range = upper bound of the last class – lower bound of the first class

- Coefficient of range = L–S / L+S where L is the largest observation and S is the smallest

MERITS AND DEMERITS OF RANGE

MERITS

- Range is the simplest measure of dispersion
- Range is very easy to calculate and to understand

DEMERITS

- Range is not based on all the observations
- Range depends only upon the largest and smallest observations
- It is affected largely by extreme observations
- For the observations in grouped frequency form, range cannot be calculated for data with open ended classes

QUARTILE DEVIATION(SEMI-INTER QUARTILE RANGE)

- It is the measure of dispersion
- If n observations given in the form of raw data, arrange the observations in ascending order of magnitude , the observations coming in the $(n/4)^{th}$, $(n/2)^{th}$ and $(3n/4)^{th}$ position are respectively called first, second and third quartiles . That is Q_1 ,Q_2 and Q_3
- Quartile Deviation Q.D $=(Q_3 - Q_1)/2$

MERITS AND DEMERITS OF QUARTILE DEVIATION

MERITS

- It is rigidly defined
- It can be calculated for data with open ended classes

DEMERITS

- Quartile Deviation not considering all the observations
- It is not capable for further algebraic treatment
- It is much affected by fluctuation of sampling

MEAN DEVIATION

- Mean Deviation of a set of observations is the arithmetic mean of the absolute values of deviation of the observations from an average
- It is the scattering of the observations taken from an average
- If $x_1, x_2, x_3, \ldots\ldots, x_r$ are the observations and let 'A' be an average , the mean deviation about 'A' is defined as

$$M.D.(A) = 1/n \sum |x_i - A|$$

Mean deviation about mean $\bar{x}$ of the observation is

$$M.D.(\bar{x}) = 1/n\sum|x_i - \bar{x}|$$

If the observations are given in the form of class and frequency with variable values If x_1, x_2, x_3,……,x_r and corresponding frequencies f_1, f_2,…….,f_r and A is an average then

$$M.D.(A) = 1/n \sum f_i|x_i - A|$$

Then mean deviation about mean x of the observation is

$$M.D.(\bar{x}) = 1/n \sum f_i|x_i - \bar{x}|$$

MERITS AND DEMERITS OF MEAN DEVIATION

- **MERIT**
- Mean Deviation is rigidly defined
- It is based on all the observation
- It is simple to understand and easy to calculate
- It is not much effected by extreme items
- M.D is minimum when it is taken about the median

- **DEMERIT**
- Mean deviation not considering the sign of derivations which make the measure non-algebraic
- It is not capable for further algebraic treatment

STANDARD DEVIATION

- Standard Deviation is defined as the square root of the arithmetic mean of the squares of deviations of the observations from their arithmetic mean
- If x_1, x_2,.....,x_r are the observations with arithmetic mean x , then standard deviations of the observations is defined
- $\text{S.D.} = \sqrt{1/n \sum (x_i - x)}^2$
- The square of the standard deviation is known as **variance**

- If the observations are given in the form of class and frequency with variable values x_1, x_2,……x_r, and corresponding frequencies f_1 ,f_2,…f_r , and with arithmetic mean x , then ,
- S.D. = $\sqrt{\frac{1}{N}\sum_i f_i(x_i-\bar{x})^2}$
- Also $S.D.=\sqrt{\frac{1}{n}\sum_i x_i^2-(\bar{x})^2}$
- For the frequency table form of observations,

$$S.D.=\sqrt{\frac{1}{N}\sum_i f_i x_i^2-(\bar{x})^2}$$

➢Shortcut method of standard deviation

- If the observations are in big statistics we can use the shortcut method to reduce the calculations involved in finding the standard deviation of the set
- Let us transform the x values to u values, so as, $u_i = (x_i - A)/c$, where A and c are the constant value

that is

$$S.D._{(u)} = \sqrt{\frac{1}{N}\sum_i f_i (u_i - \overline{u})^2}$$
$$= \sqrt{\frac{1}{N}\sum_i f_i \left(\frac{x_i - A}{c} - \left(\frac{\overline{x} - A}{c}\right)\right)^2}$$

$$= \sqrt{\frac{1}{N}\sum_i f_i \left(\frac{x_i - \overline{x}}{c}\right)^2}$$

$$= \sqrt{\frac{1}{N}\sum_i f_i \left(\frac{x_i - \overline{x}}{c}\right)^2}$$

$$= \frac{1}{c}\sqrt{\frac{1}{N}\sum_i f_i (x_i - \overline{x})^2}$$

$$S.D._{(u)} = \frac{1}{c} \times S.D._{(x)}$$

$$S.D._{(x)} = c \times S.D._{(u)}$$

➢Properties of standard deviation

1. Standard deviation is not affected by change of origin
2. Standard deviation is affected by change of scale
3. Standard deviation cannot be smaller than mean deviation about mean
4. Combined standard deviation

If one group of n_1 observation have AM $\bar{x}_1$ and SD σ_1 and another group of n_2 observations have an AM $\bar{x}_2$ and SD σ_2 and the SD σ of the two group combined is given by

$$\sigma^2 = \frac{1}{n_i + n_2}\left[n_1 \sigma_1^2 + n_2 \sigma_2^2 + n_1 d_1^2 + n_2 d_2^2\right]$$

where $d_1 = \bar{x}_1 - \bar{x}$ and $d_2 = \bar{x}_2 - \bar{x}$ and $\bar{x}$ is the combined AM

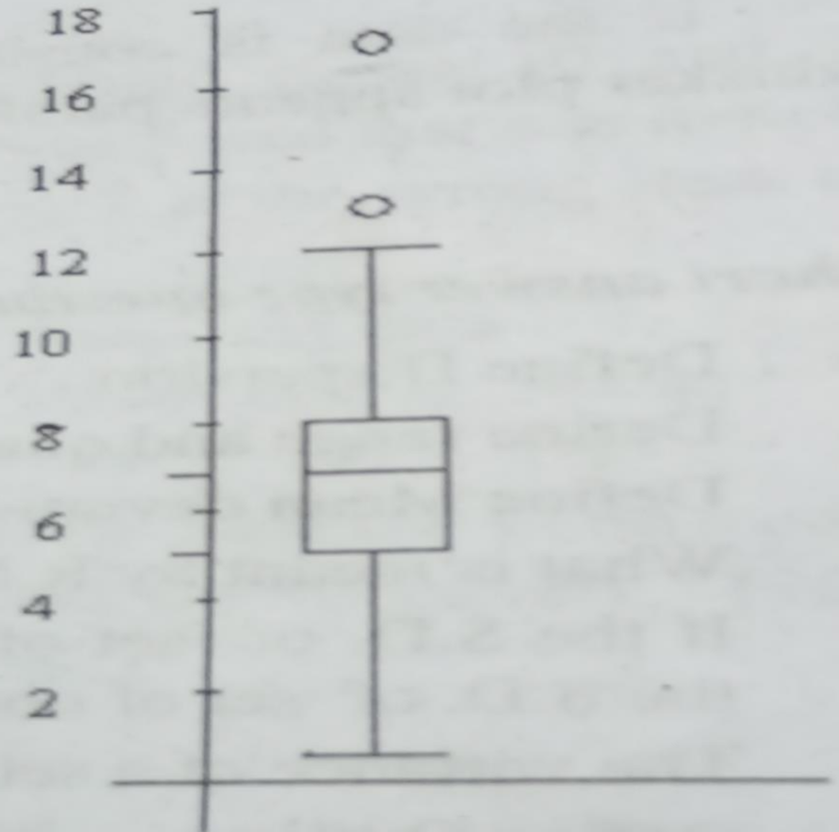
➢Relative measure of dispersion

- Coefficient of variation C.V. = S.D./A.M. X100

- To compare the consistency of two sets of observations, the coefficient of variation is used. The set with less C.V. is more consistent
- Q_3-Q_1/Q_3+Q_1 is the coefficient of quartile deviation
- M.D./A.M. is coefficient of M.D. from mean. Etc.,

Box-Whisker Plot

- A box plot or box-whisker plot is a set of fine summary measures of the set of data; they are (1) median (2) lower quartile (3) upper quartile (4) smallest observation (5) largest observation
- To draw box plot for a set of data, a box is drawn in the graph by considering Q_1 as the bottom and Q_3 as the top of the box. A horizontal line which partition the box is also drawn through the value of Q_2.



Box-whisker plot showing the duration rounded in hours of 45 hospital patients slept following the administration of a certain anesthetic.

- Now we are identify some more values :
  they are

i. H-spread; which is Q_3-Q_1
ii. 1.5x H-spread
iii. Q_3+1.5x H-spread, which is called Upper Inner Fence
iv. Q_1- 1.5x H-spread, which is called Lower Inner Fence
v. Q_3+3x H-spread, which is called Upper Outer Fence
vi. Q_1- 3x H-spread, which is called Lower Outer Fence
vii. Upper Adjacent, which is the largest observation below the upper inner fence
viii. Lower Adjacent, which is the smallest observation above the lower inner fence
ix. Outside values, which are the values beyond an inner fence but not beyond outer fence and
x. Far outside values or extreme values, which are the values beyond an outer fence