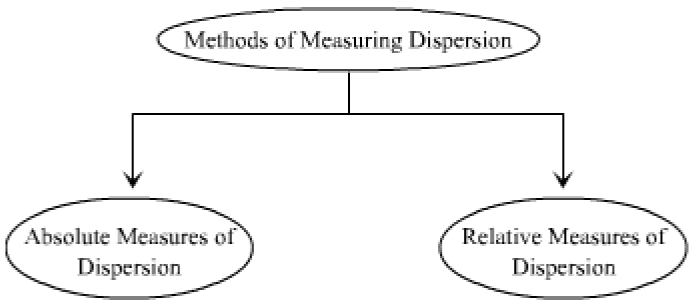
CHAPTER 6

MEASURES OF DISPERSION

- The extent to which different items tend to disperse away from the central tendency is measured by **dispersion**.

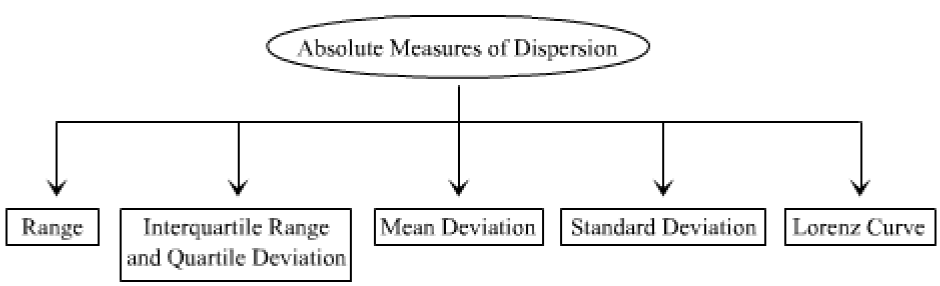
- Types of measures of dispersion



- **Difference between Absolute Measures and Relative Measures**

Absolute Measures	Relative Measures
A dispersion expressed in terms of original unit of series is called absolute dispersion.	A dispersion that expresses the variability of data in relative value or percentage is called relative dispersion.
For example, range, quartile deviation, mean deviation, standard deviation.	For example, coefficient of range, coefficient of quartile deviation, coefficient of mean deviation, and coefficient of variation

- **Absolute Measures**



- **Range** is the difference between the highest value and the lowest value in a series.

$$R = H - L$$

where,

R represents range
H represents the highest value

L represents the lowest value

- **Inter Quartile Range** is the difference between the third quartile (Q_3) and the first quartile (Q_1) of a series.

$$\text{Inter Quartile Range} = Q_3 - Q_1$$

- **Merits of Range**
  - It is rigidly and well-defined.
  - It is easy to calculate and simple to understand.
  - It gives the measures of dispersion in the same units that of the variable. For example, the range of speed is given in kilometres per hour or meters per second.
  - It is not affected by frequencies.

- **Demerits of Range**
  - As it is not based on all the observations, so the change in the maximum and the minimum values can affect the value of Range to a large extent.
  - It cannot be used to calculate Range in the open-ended series (i.e. continuous frequency distribution).
  - Unlike, Mean Deviation and Standard Deviation, Range is independent of the any of the measures of central tendency such as, mean, median and mode.
  - It is affected by the change in the sampling fluctuations.

- **Uses of Range**
  - It is use to study and check quality variation in products.
  - It is use to study fluctuations in various economic variables such as, prices, exchange rates, etc.
  - It is used by the meteorological department to forecast weather conditions.
  - It can also be used to measure variability in of wide number of variables such as, wages, sales, etc.

- **Quartile Deviation (Semi-inter Quartile Range) is half of the Inter Quartile Range**

Formulas to Calculate Quartile Deviation		
For Individual Series		
$QD = \frac{Q_3 - Q_1}{2}$	$Q_1 = \text{Size of} \left(\frac{n+1}{4}\right)^{\text{th}} \text{item}$ n = Number of observations	$Q_3 = \text{Size of} \left(\frac{3(n+1)}{4}\right)^{\text{th}} \text{item}$ n = Number of observations
For Discrete Series		
$QD = \frac{Q_3 - Q_1}{2}$	Locate the Size of $\left(\frac{N+1}{4}\right)^{\text{th}}$ item in the *CF* column and corresponding x value is Q_1 N = Sum of frequencies	Locate the Size of $\left(\frac{3(N+1)}{4}\right)^{\text{th}}$ item in the *CF* column and corresponding x value is Q_3 N = Sum of frequencies

For Continuous Series		
$QD = \frac{Q_3 - Q_1}{2}$	Locate the Size of $\left(\frac{N}{4}\right)^{\text{th}}$ item in CF column and the value of Q_1 will lie in the corresponding class interval. $$Q_1 = l_1 + \frac{\frac{N}{4} - CF}{f} \times i$$ where, l_1= Lower limit of class interval; N = Sum of frequencies; CF = Cumulative frequency of the class preceding the Q_1 class; i = Class interval	Locate the Size of $\left(\frac{3N}{4}\right)^{\text{th}}$ item in CF column and the value of Q_3 will lie in the corresponding class interval. $$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - CF}{f} \times i$$ where, l_1= Lower limit of class interval; N = Sum of frequencies; CF = Cumulative frequency of the class preceding the Q_3 class; i = Class interval

- **Merits of Quartile Deviation**
  - It is easy to calculate and simple to understand.
  - It is less affected by the extreme values in the sample.
  - It does not depend on all the values of the sample.
  - Similar to the Range, it gives the measures of dispersion in the same units that of the variable.
  - It can be calculated in the open-ended series (for continuous series).
  - It is more superior and reliable than the Range.

- **Demerits of Quartile Deviation**
  - It cannot be used for further mathematical treatment.
  - As it considers only first 25% and the last 25 % of the observations, so, it ignores roughly 50% of the observations.
  - It does not provide enough result to conduct meaningful comparisons.
  - It is affected by the change in the sampling fluctuations.

- **Mean deviation** is the arithmetic average of the deviations of all values taken from some average value.

	For Individual Series	For Discrete and Continuous Series
When Mean deviation are taken from **Median** (M)	$MD_M = \frac{\sum \lvert X - M \rvert}{n} = \frac{\sum \lvert D_M \rvert}{n}$	$MD_M = \frac{\sum f \lvert X - M \rvert}{N} = \frac{\sum f \lvert D_M \rvert}{N}$
When Mean deviation are taken from **Arithmetic Mean** $\left(\overline{X}\right)$	$MD_{\overline{X}} = \frac{\sum \lvert X - \overline{X} \rvert}{n} = \frac{\sum \lvert D_{\overline{X}} \rvert}{n}$	$MD_{\overline{X}} = \frac{\sum f \lvert X - \overline{X} \rvert}{N} = \frac{\sum f \lvert D_{\overline{X}} \rvert}{N}$

When Mean deviation are taken from **Mode (*Z*)**	$MD_Z = \frac{\sum\lvert X-Z\rvert}{n} = \frac{\sum\lvert D_Z\rvert}{n}$	$MD_Z = \frac{\sum f\lvert X-Z\rvert}{N} = \frac{\sum f\lvert D_Z\rvert}{N}$

where,

$\sum\lvert D_M\rvert = \sum\lvert X-M\rvert$ = sum of deviations taken from median (M)

$\sum\lvert D_{\overline{X}}\rvert = \sum\lvert X-\overline{X}\rvert$ = sum of deviations taken from mean ($\overline{X}$)

$\sum\lvert D_Z\rvert = \sum\lvert X-Z\rvert$ = sum of deviations taken from mode (Z)

f = frequency

$N = \sum f$ = Sum of frequencies

n = Number of observations

Merits of Mean Deviation

- It is easy to compute and understand.
- It gives us a definite and precise value.
- It is better than range and quartile deviation as, it is affected by all the items of a series.
- It is less affected by the value of the extreme values.

Demerits of Mean Deviation

- It ignores positive and negative signs.
- It cannot be used for further mathematical (or algebraic) treatment.
- It involves a tedious and calculative process, in case the value of mean, median or mode is in fraction.
- It gives unreliable results when mean deviation is calculated from mode. This is because, at times, mode is difficult to calculate.

Standard Deviation is the square root of the arithmetic mean of the square of deviation of item from the mean value.

For Individual Series		
Actual Mean Method	$\sigma = \sqrt{\frac{\Sigma x^2}{n}} = \sqrt{\frac{\Sigma\left(X-\overline{X}\right)^2}{n}}$	x^2 = Sum total of square of deviation $\left(X-\overline{X}\right) = x =$ Deviation from mean σ = Standard Deviation n = Number of observations
Direct Method	$\sigma = \sqrt{\frac{\Sigma X^2}{n} - \left(\overline{X}\right)^2} = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$	$\Sigma X^2 =$ Sum of the square of values of X

Assumed Mean Method/ Short-cut Method	$\sigma = \sqrt{\frac{\Sigma d_x^2}{n} - \left(\frac{\Sigma d_x}{n}\right)^2}$	$d_x = X - A$ $\Sigma d_x^2 =$ Sum total of square of deviation taken from assumed value (A) $\Sigma d_x =$ Sum of deviation taken from assumed value $d_x^2 = (d_x) \times (d_x)$

For Discrete Series		
Actual Mean Method	$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{\Sigma f\left(X - \overline{X}\right)^2}{N}}$	$\left(X - \overline{X}\right) = x =$ Deviation from mean $\sigma =$ Standard Deviation $f =$ frequency $fx^2 = (f) \times (x^2)$ $\overline{X} = \frac{\sum fX}{N}$ $N = \sum f =$ Sum of frequencies
Direct Method	$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - \left(\overline{X}\right)^2} = \sqrt{\frac{\Sigma fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2}$	$\overline{X} = \frac{\sum fX}{N}$
Assumed Mean Method/Short-cut Method	$\sigma = \sqrt{\frac{\Sigma fd_x^2}{N} - \left(\frac{\Sigma fd_x}{N}\right)^2}$	$d_x = X - A$ $\Sigma d_x^2 =$ Sum total of square of deviation taken from assumed value (A) $\Sigma d_x =$ Sum of deviation taken from assumed value $fd_x^2 = (fd_x) \times (d_x)$
Step Deviation Method	$\sigma = \sqrt{\frac{\Sigma fd_x'^2}{N} - \left(\frac{\Sigma fd_x'}{N}\right)^2} \times h$	$d_x' = \frac{X - A}{h}$, h = common factor $\Sigma d_x =$ Sum of deviation taken from assumed value $fd_x'^2 = (fd_x') \times (d_x')$

		$\overline{X} = A + \frac{\sum fd'_x}{N} \times h$ N = Sum of frequencies

For Continuous Series		
Actual Mean Method	$\sigma = \sqrt{\frac{\Sigma fm^2}{N}} = \sqrt{\frac{\Sigma f\left(m-\overline{X}\right)^2}{N}}$	$\left(m-\overline{X}\right)$ = Deviation from mean $\sigma =$ Standard Deviation f = frequency $m =$ mid points of class intervals $\overline{\mathrm{X}} = \frac{\sum fm}{N}$ $N =$ Sum of frequencies
Direct Method	$\sigma = \sqrt{\frac{\Sigma fm^2}{N} - \left(\overline{X}\right)^2} = \sqrt{\frac{\Sigma fm^2}{N} - \left(\frac{\sum fm}{N}\right)^2}$	$\overline{\mathrm{X}} = \frac{\sum fm}{N}$
Assumed Mean Method/Short-cut Method	$\sigma = \sqrt{\frac{\Sigma fd_m{}^2}{N} - \left(\frac{\Sigma fd_m}{N}\right)^2}$	$d_m = m - A$ $\Sigma d_m{}^2 =$ Sum total of square of deviation taken from assumed value (A) Σd_m = Sum of deviation taken from assumed value $fd_m{}^2 = \left(fd_m\right) \times \left(d_m\right)$ $N =$ Sum of frequencies
Step Deviation Method	$\sigma = \sqrt{\frac{\Sigma fd'^2_m}{N} - \left(\frac{\Sigma fd'_m}{N}\right)^2} \times h$	$d'_m = \frac{X-A}{h}$, h = common factor Σd_m = Sum of deviation taken from assumed value $fd'^2_m = \left(fd'_m\right) \times \left(d'_m\right)$ $\overline{X} = A + \frac{\sum fd'_m}{N} \times h$ $N =$ Sum of frequencies

Merits of Standard Deviation

- It is based on all the values of sample.
- It gives us a definite and precise value, therefore, is a well-defined and definite measure of dispersion.
- It depicts a clear picture and is more reliable than Range. This is because it is based on the mean that is an ideal average.
- As, it is based on all the values of the sample, so it is less affected by the change in the sample value.
- Unlike Mean Deviation, Standard Deviation can be used for further mathematical (or algebraic) treatment.

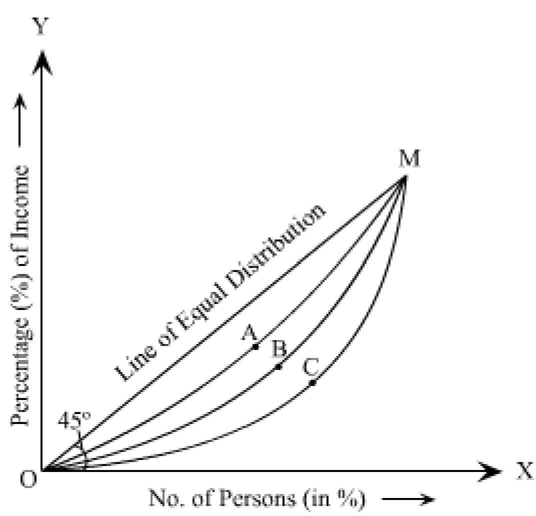
Demerits of Standard Deviation

- It assigns more weights to the extreme values.
- It involves a complex and tedious calculation process. Moreover, it is hard to understand.

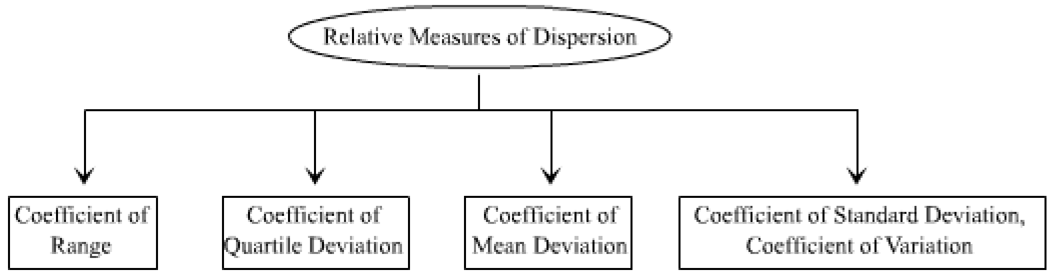
Lorenz Curve- This curve measures the deviation from the actual distribution from the line of equal distribution. In the figure, the line OM denotes equal distribution at an angle of 45°. The curves OAM, OBM and OCM represent the deviation from line in an ascending order.

Applicability of Lorenz Curve- The Lorenz Curve helps us to assess the inequalities in the following economic variables:

- Income
- Wealth
- Population
- Wages
- Profit



Relative Measures of Dispersion



Relative Measure of Range		
Coefficient of Range (For Individual, Discrete and Continuous Series)	$CR = \dfrac{H-L}{H+L}$	CR represents coefficient of range H represents the highest value in the series L represents the lowest value in the series
Relative Measure of Quartile Deviation		
Coefficient of Range (For Individual, Discrete and Continuous Series)	Coefficient of $QD = \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$	Q_3 represents third (upper) quartile Q_1 represents first (lower) quartile
Relative Measure of Mean Deviation		
Coefficient of Mean Deviation (For Individual, Discrete and Continuous Series)	$\dfrac{MD_x}{\overline{X}} = \dfrac{\text{Mean Deviation}}{\text{Arithmetic Mean}}$, if from Mean $\dfrac{MD_M}{M} = \dfrac{\text{Mean Deviation}}{\text{Median}}$, if from Median $\dfrac{MD_Z}{Z} = \dfrac{\text{Mean Deviation}}{\text{Mode}}$, if from Mode	
Relative Measure of Standard Deviation		
Coefficient of Standard Deviation (For Individual, Discrete and Continuous Series)	Coefficient of Standard Deviation $= \dfrac{\sigma}{\overline{X}}$	$\sigma =$ Standard Deviation $\overline{X} =$ Arithmetic Mean

Coefficient of Variation (For Individual, Discrete and Continuous Series)	$\text{CV} = \dfrac{\sigma}{\overline{X}} \times 100$	$\sigma =$ Standard Deviation $\overline{X} =$ Arithmetic Mean

- For comparing the variability or dispersion of two series, we first calculate the C.Vs of each series. The series having higher C.V. is said to be more variable than the other and the series having lower C.V. is said to be more consistent than the other.

- For two series with equal mean values, the series with greater standard deviation (or variance) is more variable or dispersed than the other. Also, the series with lower value of standard deviation (or variance) is said to be more consistent or less scattered than the other.

- **Relationship Between Different Measures of Dispersion**

$$\text{Quartile Deviation} = \frac{2}{3}(\text{Standard Deviation})$$

$$\text{Quartile Deviation} = \frac{5}{6}(\text{Mean Deviation})$$

$$\text{Mean Deviation} = \frac{4}{5}(\text{Standard Deviation})$$

$$6(\text{Standard Deviation}) = 9(\text{Quartile Deviation}) = 7.5(\text{Mean Deviation})$$