Measure of Dispersion:

The dispersion is a word that refers to the variability in the values of units and is defined as "the degree of scatterness and clusterness of the observations of a given data from some central value (measure of central tendency).

Some distributions are said to have more dispersion if the observations are scattered away from its central value which on the other hand, distributions are said to have less dispersion if the observations are clustered arround the central tendency.

The values which are used for measuring the dispersion in a given distribution are called "measure of dispersion".

Types of measure of Dispersion:

1. Absolute measure of dispersion
2. Relative measure of dispersion

1. Absolute measure of dispersion.

Those measures which are used for measuring the absolute dispersion in a distribution are called as absolute measure of dispersion and are classified in four categories:

a) Range
b) Mean deviation
c) Quartile deviation
d) Standard deviation.

Unite of measurement of absolute measure of dispersion is same as that of the observation of given dat

Range :

It is most simplest type of measure of dispersion. It is the difference between maximum and minimum values of the data i.e.,

$$\text{Range} = \text{Maximum value} - \text{Minimum Value}$$
or
$$= X_m - X_0$$

Merits :

i) For ungroup data, it is very easy to calculate.
ii) It is an easy measure to understand and to interpret.
iii) It is very useful in the field of Quality control.

Demerits :

i) It is an unstable measure, because it is defined an extreme values and unusually large value will unnecessary increase the range
ii) It does not depend on all the observations and as such it is not effected by the variation of observation.
iii) It has not so much use in statistics.

b) Quartile Deviation:

This is also called semi-inter quartile range. It is defined as

$$Q \cdot D = \frac{Q_3 - Q_1}{2} = \frac{P_{75} - P_{25}}{2}$$

where,

Q_1 is the lower quartile
Q_3 is the upper quartile
P_{25} is the 25th percentile or percent point
P_{75} is the 75th percentile or percent point

It is the form of reduced range.

Merits:

i) It is better than range
ii) When the central tendency is measured by median the dispersion of the distribution is usually computed by quartile deviation because of its affinity to the median.
iii) It can easily be calculated in group data.

Demerits:

i) It is an absolute measure which is affected by the observations arranged in some order.
ii) It does not provide any information regarding the dispersion of observations lying out side the two quartiles.
iii) It does not depend on all the values.

c) Mean Deviation:

It is defined as the average of the absolute deviation taken from some central value (such as, mean, median and mode). Mathematically,

Ungrouped data   Grouped data

$$M.D_{\bar{x}} = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n} \; ; \; M.D_{\bar{x}} = \frac{\sum_{i=1}^{k} f_i |X_i - \bar{X}|}{\sum_{i=1}^{k} f_i}$$

$$M.D_{Median} = \frac{\sum_{i=1}^{n} |X_i - Median|}{n} \; ; \; M.D_{Median} = \frac{\sum_{i=1}^{k} f_i |X_i - Median|}{\sum_{i=1}^{k} f_i}$$

$$M.D_{Mode} = \frac{\sum_{i=1}^{n} |X_i - Mode|}{n} \; ; \; M.D_{Mode} = \frac{\sum_{i=1}^{k} f_i |X_i - Mode|}{\sum_{i=1}^{k} f_i}$$

Where n is the number of observation.   Where $n = \sum_{i=1}^{k} f_i$ the total frequencies

Merits:

i) It is based on all the observations
ii) It is simple to calculate
iii) It is not affected by extreme values

Demerits:

i) It is not an accurate measure of dispersion.
ii) It is not widely used.
iii) It has limited algebric treatment.

d) Standard Deviation:

It is defined as the positive square root of the average of squared deviations taken from A.M. of the distribution. Mathematically,

Ungrouped data

$$S.D = S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$$

or

$$S = \sqrt{\frac{1}{n}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]}$$

Grouped data

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

$$S = \sqrt{\frac{1}{n}\left[\sum_{i=1}^{k} f_i x_i^2 - \frac{(\sum_{i=1}^{k} f_i x_i)^2}{n}\right]}$$

and $n = \sum_{i=1}^{k} f_i$

Standard deviation is independent of change of origian and scale.

Root mean squared deviation:

It is defined as positive square root of average of squared deviations taken from some arbitrary constant A (say). Mathematically

$$S = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i-A)^2}{n}} \quad ; \quad n = \sum_{i=1}^{k} f_i$$

Relation between S.D and R.M.S.D :

$$S^2 = \sigma^2 + (\bar{X} - A)^2$$

If $\bar{X} = A$, Then

$$S^2 = \sigma^2$$

Merits:

i) It may be manipulated algebraically
ii) It is best measure of dispersion
iii) It is based on all the observations
iv) It is of great importance for Sampling Theory.
v) It is least effected by sampling fluctuations.

Demerits:

i) It is not good measure when the series of data is short.

Variance:

Variance is not an independent measure of dispersion. It is defined as the average of the squared deviation taken from arithmetic mean of the given distribution.

$$Var(X) = \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{(ungrouped data)}$$

$$Var(X) = \sigma^2 = \frac{1}{n} \sum_{i=1}^{k} f_i (X_i - \bar{X})^2 \quad \text{(Grouped data)}$$

Relative Measure of Dispersion:

Those measures which are used for the comparison purposes are called relative measure of dispersion. They have no units of measurements because they are ratios. There are three types of relative measure of dispersions.

a) Coefficient of Quartile Deviation
b) Coefficient of Range
c) Coefficient of Variation

a) Coefficient of Quartile Deviation:

It is defined as the difference between upper quartile and lower quartile divided by the sum of the two quartiles. Mathematically,

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

b) Coefficient of Range:

It is defined by the difference between maximum value and minimum value divided by the sum of maximum and minimum values. i.e.,

$$\text{Coefficient of Range} = \frac{X_m - X_o}{X_m + X_o}$$

Coefficient of Variation:

The standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated. The standard deviation of heights of students cannot be compared with the standard deviation of weights of students, as both are expressed in different units, i.e., heights in metres and weights in kgs. Therefore, the standard deviation converted into a relative measure of dispersion for the purpose of comparison. The relative measure is known as the coefficient of variation.

It is obtained by dividing the standard deviation by the mean and multiplying it by 100. Symbolically,

$$\text{Coefficient of variation (or C.V)} = \frac{\sigma}{\mu} \times 100 \text{ (Population)}$$

or

$$\text{Coefficient of variation (or C.V)} = \frac{s}{\bar{x}} \times 100 \text{ (Sample)}$$

The series or groups of data for which the coefficient of variation is greater, indicates that the group is more variable, less stable, less uniform, less consistent or less homogeneous. If the coefficient of variation is less, it indicates that the group is less variable, more stable, more uniform, more consistent or more homogeneous.

Moments :

Moments of a distribution is designated by the power of the deviation, varies to at end.
when deviation is taken about mean then moments are called as mean moments whereas the deviation about any constant A (when A = 0) then moments are called as Raw moments or moments about origion. and order of the moments designate the power of the deviation.

Raw moments:

$$\mu_r' = \frac{1}{n} \sum f_i (x_i - A)^r \quad \text{(rth moment about constant)}$$

if A = 0, then

$$\mu_r' = \frac{1}{n} \sum f_i x_i^r \quad \text{(rth moment about origion)}$$

$$\mu_0' = \text{Null moment} = 1$$

$$\mu_1' = \frac{1}{n} \sum f_i x_i = \bar{x}$$

$$\mu_2' = \frac{1}{n} \sum f_i x_i^2$$

$$\mu_3' = \frac{1}{n} \sum f_i x_i^3$$

$$\mu_4' = \frac{1}{n} \sum f_i x_i^4$$

and so on.

Moments:

$$\mu_r = \frac{1}{n}\sum f_i (x_i - \bar{x})^r \qquad r\text{th moment about mean.}$$

$r = 0$

$\mu_0 = 1$

$$\mu_1 = \frac{1}{n}\sum f_i (x_i - \bar{x}) = 0$$

$$\mu_2 = \frac{1}{n}\sum f_i (x_i - \bar{x})^2 = \sigma^2 = \text{Variance}$$

$$\mu_3 = \frac{1}{n}\sum f_i (x_i - \bar{x})^3$$

$$\mu_4 = \frac{1}{n}\sum f_i (x_i - \bar{x})^4$$

and so on.

Properties:

i) Effect of change of origion and change of scale on moments.

Proof: by definition

$$\mu_r = \frac{1}{n}\sum f_i (x_i - \bar{x})^r$$

Let $d_i = \dfrac{x_i - A}{h} \Rightarrow x_i = A + h d_i$

$\bar{x} = A + h\bar{d}$

$$\mu_r = \frac{1}{n}\sum f_i \left[(A + h d_i) - (A + h\bar{d})\right]^r$$

$$= \frac{1}{n}\sum f_i \left[h(d_i - \bar{d})\right]^r$$

$$\boxed{\mu_r = \frac{1}{n}\sum f_i (d_i - \bar{d})^r \cdot h^r}$$

$\therefore$ Moments are independent of change of origion but dependent of change of scale.

Relationship between Raw moments and Mean Moments

By the definition of mean moments,

$$\mu_r = \frac{1}{n}\sum f_i (x_i - \bar{x})^r \qquad \because \mu_1' = \bar{x}$$

$$\mu_r = \frac{1}{n}\sum f_i (x_i - \mu_1')^r$$

on expending by binomial theorem we have.

$$\mu_r = \frac{1}{n}\sum f_i \left[x_i^r - \binom{r}{1} x_i^{r-1}\mu_1' + \binom{r}{2} x_i^{r-2}\mu_1'^2 - \binom{r}{3} x_i^{r-3}\mu_1'^3 \right.$$
$$\left. \cdots\cdots + (-1)^r \mu_1'^r \right]$$

$$= \frac{1}{n}\sum f_i x_i^r - \binom{r}{1}\sum f_i x_i^{r-1}\cdot\mu_1' + \binom{r}{2}\sum f_i x_i^{r-2}\mu_1'^2$$
$$- \binom{r}{3}\sum f_i x_i^{r-3}\mu_1'^3 + \cdots\cdots + (-1)^r \mu_1'^r.$$

$$\boxed{\mu_r = \mu_r' - \binom{r}{1}\mu_{r-1}'\mu_1' + \binom{r}{2}\mu_{r-2}'\mu_1'^2 + \cdots\cdots + (-1)^r \mu_1'^r}$$

for $r=1$ $\quad \mu_1 = \mu_1' - \mu_1' = 0$

$r=2$

$$\mu_2 = \mu_2' - 2\mu_1'^2 + \mu_1'^2 = \mu_2' - \mu_1'^2$$

$r=3$ $\quad \mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$

$r=4$ $\quad \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

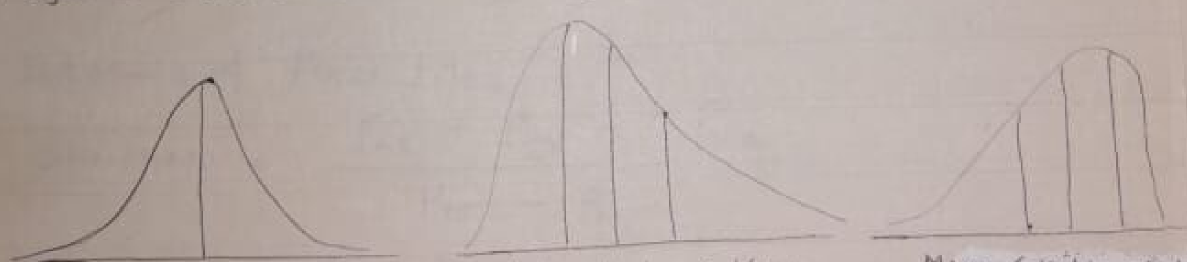
Skewness:

The degree of departure of the frequency curve of a distribution from symmetric nature is called skewness of the distribution.

If the long tail of the frequency curve of a distribution is more alongated towards right then distribution is said to have Positive Skewness. On the other hand if the long tail of the curve of the distribution is moved towards left, the distribution is said to have negative skewness.

In case of symmetric distribution, mean median and mode of the distribution are conside at center point. Where as in case of skewed distribution, they located at different points as shown below

a) Symmetric Distribution   b) Positively Skewed   c) Negatively Skewed



Mean = Median = Mode

Mode < Median < Mean
Mean > Median > Mode

Mean < Median < Mode
Mode > Median > Mean

Measure of Skewness:

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, where β_1 is an absolute measure of skewness

If $\beta_1 = 0$, distribution is symmetric i-e there is no skewness. If $\beta_1 > 0$, then distribution is skewed.

It measures the presence or absence of skewness without mentioning the direction of skewness.

...ative measure of Skewness

1\. According to Karl Pearson

$$\text{Skewness} = \frac{\text{Mean} - \text{Mode}}{S.D}$$

But if mode is not clearly defined in a distribution then mode is replaced by empirical relation among Mean, Median and Mode. i.e.,

$$3(\text{Mean} - \text{Median}) = \text{Mean} - \text{Mode}$$

Hence

(b) $$\text{Skewness} = \frac{3(\text{Mean} - \text{Median})}{S.D} = \frac{3(\text{Mean} - Q_2)}{S.D}$$

2\. According to Bowley

(d) $$\text{Skewness} = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

Providing $0 \leq \text{Skewness} \leq +1$

4\. 3. (f) $$\text{Skewness} = \frac{\sqrt{\beta_1}\,(\beta_2 + 3)}{2[5\beta_2 - 6\beta_1 - 9]}$$

3 4. Interms of Percentiles

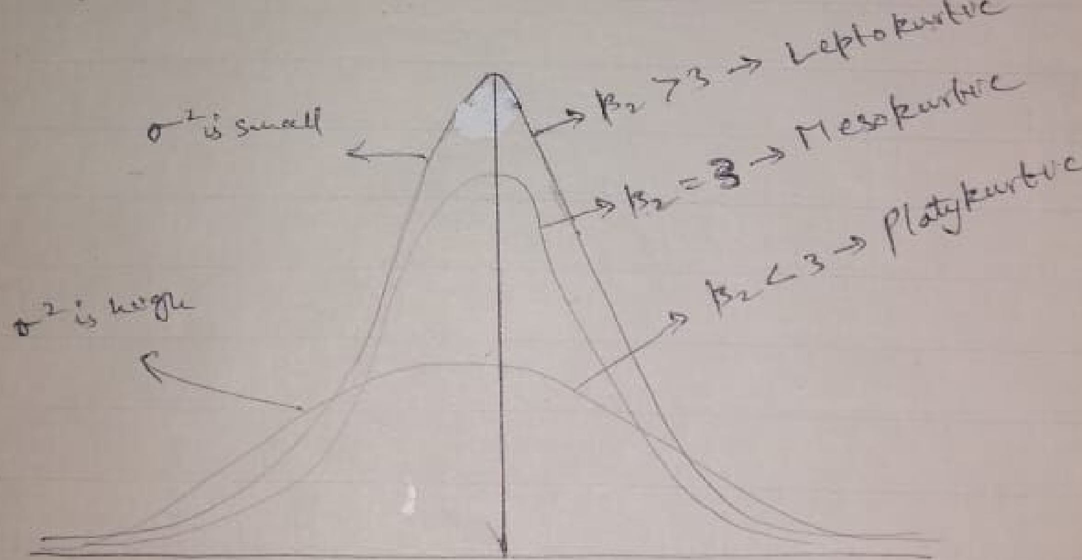
(e) $$\text{Skewness} = \frac{P_{25} + P_{75} - 2P_{50}}{P_{75} - P_{25}}$$

(c) 5. $$\text{Skewness} = \frac{3(\text{Mean} - D_5)}{S.D}$$

5 6. (g) $$\text{Skewness} = \frac{\gamma_1\,(\beta_2 + 3)}{2[5\beta_2 - 6\beta_1 - 9]}$$

Kurtosis :

It is a characteristic of frequency distribution. It measures the flatness and peakness of the curve of a given distribution. There are some distributions whose frequency curve are more peaked. They are said to have leptokurtic. While some distributions can be represented by more flatted curves. They are said to platykurtic. There are some distributions whose curve are neither more flated not more peaked, but they obtain a normal position in between the two such distributions, they are said to be mesokurtic.



Measure of Kurtosis :

β_2 is the measure of Kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

Relative Measure of Kurtosis:

γ_1 and γ_2 are the relative measures of kurtosis. i.e.,

$\gamma_1 = \sqrt{\beta_1}$; where $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

If $\gamma_1 = 0$, this implies that the distribution is symmetrical

$\gamma_2 = \beta_2 - 3$; where $\beta_2 = \frac{\mu_4}{\mu_2^2}$

If $\gamma_2 = 0$, this implies that the distribution is mesokurtic

Show that

a) $\beta_2 \geq 1$ b) $\beta_2 > \beta_1$ X

Proof: we know that

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\Rightarrow \frac{\mu_4}{\mu_2^2} \geq 1$$

$$\mu_4 \geq \mu_2^2$$

$$\because \mu_4 = \frac{\sum f(x-\bar{x})^4}{n} \quad \text{and} \quad \mu_2 = \frac{\sum f(x-\bar{x})^2}{n}$$

$$\Rightarrow \frac{\sum f(x-\bar{x})^4}{n} \geq \left[\frac{\sum f(x-\bar{x})^2}{n}\right]^2$$

$$\frac{\sum f d^2}{n} \geq \left[\frac{\sum f d}{n}\right]^2 \qquad \text{let } d = (x-\bar{x})^2$$

$$\frac{\sum f d^2}{n} - \left(\frac{\sum f d}{n}\right)^2 \geq 0$$

$$\sigma_d^2 \geq 0.$$

It is always true by using the property of variance.

$\therefore \beta_2 \geq 1$

characteristics

ii) $\beta_2 > \beta_1$ X'

P-78

$$\frac{\mu_4}{\mu_2^2} > \frac{\mu_3^2}{\mu_2^3}$$

$$\mu_2 \mu_4 > \mu_3^2$$

$$\left[\frac{\Sigma f(x-\bar{x})^2}{n}\right]\left[\frac{\Sigma f(x-\bar{x})^4}{n}\right] > \left[\frac{\Sigma f(x-\bar{x})^3}{n}\right]^2$$

put $d = x - \bar{x}$

$$\left(\frac{\Sigma f d^2}{n}\right)\left(\frac{\Sigma f d^4}{n}\right) > \frac{(\Sigma f d^3)^2}{n^2}$$

$$\Sigma f d^2 \cdot \Sigma f d^4 > (\Sigma f d^3)^2$$

$$[f_1 d_1^2 + f_2 d_2^2 + \cdots][f_1 d_1^4 + f_2 d_2^4 + \cdots] > [f_1 d_1^3 + f_2 d_2^3 +$$

$$f_1 f_2 d_1^2 d_2^2 [d_1^2 + d_2^2 - 2 d_1 d_2] + \cdots > 0$$

$$f_1 f_2 d_1^2 d_2^2 [d_1 - d_2]^2 + \cdots > 0.$$

which is a positive quantity

$\therefore \beta_2 > \beta_1$ //.

$$[f_1^2 d_1^6 + f_1 f_2 d_1^2 d_2^4 + f_1 f_3 d_3^4 + f_1 f_2 d_1^4 d_2^2 + f_2^2 d_2^6 + \cdots$$

$$\left[-f_1 d_1^3 - f_2 d_2^3 - f_3 d_3^3 - \cdots\right] > -$$

2 March — STAT 301(S) PRACTICAL : Measure of central Tendency & Absolute & Relative Measure of Dispersion and Moments of the Distribution.

Problem:

The following table shows the distribution of weights of 200 students:

Weight (Ibs)	101 – 115	116 – 130	131 – 145	146 – 160	161 – 175	176 – 190	191 – 205
Frequency	2	8	37	44	46	41	10

Weight (Ibs)	206 – 220	221 – 235	236 – 250	251 – 265	Total
Frequency	6	2	3	1	200

1) Calculate an appropriate Measure of central Tendency.

a) Show that M.D (Median) < M.D (Mean)

b) Calculate first four mean moments

c) Calculate the measure of skewness (β_1) and measure of Kurtosis

d) Comment on the behaviour of the distribution.

e) Calculate the Coefficient of variation.

f) Draw a histogram and superimpose on it a frequency curve

HELP:

a) $$\text{M.D (Median)} = \frac{\sum_{i=1}^{k} f_i \left| x_i - \text{Median} \right|}{\sum_{i=1}^{k} f_i} \; ; \quad \text{M.D (Mean)} = \frac{\sum_{i=1}^{k} f_i \left| x_i - \text{Mean} \right|}{\sum_{i=1}^{k} f_i}$$

b) i) Raw Moments: rth moment about origin $= \dfrac{\sum f_i x_i^r}{\sum f_i} = \mu_r'$

ii) Mean Moments: rth mean moment

$$\mu_r = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^r}{\sum_{i=1}^{k} f_i} \; ; \text{ or } \quad \mu_1 = \mu_1' - \mu_1' = 0 \text{ (alway)}$$

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'$$

c) i) Skewness : $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

ii) Kurtosis : $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

iii) $\gamma_1 = \sqrt{\beta_1}$ $\quad \gamma_2 = \beta_2 - 3$

d) Comments

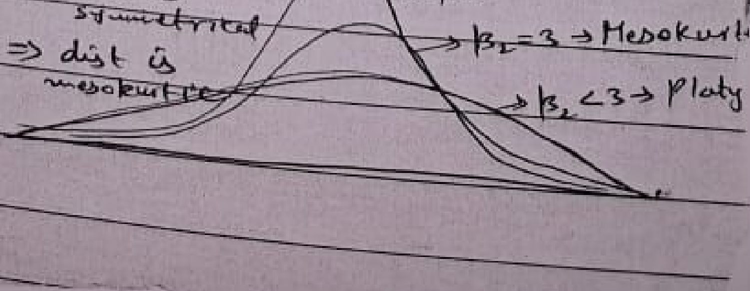
i) –vely skewed, +vely skewed or symmetric

ii)

iii) $\gamma_1 = 0 \Rightarrow$ dist is symmetrical

iv) $\gamma_2 = 0 \Rightarrow$ dist is mesokurtic

Coefficient of Variation $= \dfrac{\sigma}{\mu} \times 100$.



2 March

STAT 301(S) PRACTICAL : Measure of Central Tendency & Absolute & Relative Measure of Dispersion and Moments of the Distribution.

Problem:

The following table shows the distribution of weights of 200 students:

Weight (lbs)	101 – 115	116 – 130	131 – 145	146 – 160	161 – 175	176 – 190	191 – 205
Frequency	2	8	37	44	46	41	10

Weight (lbs)	206 – 220	221 – 235	236 – 250	251 – 265	Total
Frequency	6	2	3	1	200

1) Calculate an appropriate Measure of central Tendency.

a) Show that M.D (Median) < M.D (Mean)

b) Calculate first four mean moments

c) Calculate the measure of skewness (β_1) and measure of kurtosis

d) Comment on the behaviour of the distribution.

e) Calculate the coefficient of variation.

f) Draw a histogram and superimpose on it a frequency curve

HELP:

a) $M.D(\text{Median}) = \dfrac{\sum_{i=1}^{k} f_i |X_i - \text{Median}|}{\sum_{i=1}^{k} f_i}$; $M.D(\text{Mean}) = \dfrac{\sum_{i=1}^{k} f_i |X_i - \text{Mean}|}{\sum_{i=1}^{k} f_i}$

b) i) Raw Moments: rth moment about origin $= \dfrac{\sum f_i X_i^r}{\sum f_i} = \mu'_r$

ii) Mean Moments: rth mean moment

$$\mu_r = \frac{\sum_{i=1}^{k} f_i (X_i - \bar{X})^r}{\sum_{i=1}^{k} f_i} \text{ ; or } \mu_1 = \mu'_1 - \mu'_1 = 0 \text{ (always)}$$

$$\mu_2 = \mu'_2 - \mu'^2_1$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'_1$$

c) i) Skewness : $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

ii) Kurtosis : $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

iii) $\gamma_1 = \sqrt{\beta_1}$ $\quad \gamma_2 = \beta_2 - 3$.

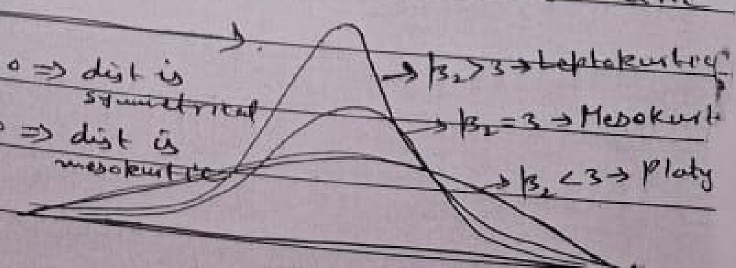
d) Comments

i) –vely skewed, +vely skewed or symmetric

ii)

iii) $\gamma_1 = 0 \Rightarrow$ dist is symmetrical

iv) $\gamma_2 = 0 \Rightarrow$ dist is mesokurtic



e) Coefficient of Variation $= \dfrac{\sigma}{\mu} \times 100$.