LECTURE 24
MOMENTS, SKEWNESS AND KURTOSIS

PROF. SHRUTI

Moments (Definition)

Suppose we have n values of a variables X as X_1, X_2, ………………..X_n. The possible measures of central tendency and dispersion of variable x are mean and variance defined by expression:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; and \; \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$\bar{x}$ is the first moment of X about the origin.

Raw Moments for Ungrouped Data

Definition: If X_1, X_2, ………………..X_n are n values of the variable X, the r^{th} raw moment of X about any point A is defined as-

$$m_r' = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^r \; ; r = 0,1,2 \ldots \ldots ..$$

So, (always)

$$m_0' = \frac{1}{n}\sum(x_i - A)^0 = 1$$

$$m_1' = \frac{1}{n}\sum(x_i - A)^1 = (\bar{x} - A)$$

$$m_2' = \frac{1}{n}\sum(x_i - A)^2$$

$$m_3' = \frac{1}{n}\sum(x_i - A)^3$$

$$m_4' = \frac{1}{n}\sum(x_i - A)^4$$

In particular, if the r^{th} raw data about origin, i.e. for A = 0 is

$$m_r' = \frac{1}{2}\sum x_i^r$$

so that,

$$m_0' = \frac{1}{n}\sum x_i^0 = 1\ (always)$$

$$m_1' = \frac{1}{n}\sum x_i = mean\ of\ the\ distribution$$

$$m_2' = \frac{1}{n}\sum X_i^2$$

$$m_3' = \frac{1}{n}\sum X_i^3$$

$$m_4' = \frac{1}{n}\sum X_i^4$$

Raw Moments for the Grouped Data

If the given values are in the form of a frequency distribution,

Table 1.1

Value of X_i (x)	$X_1, X_2, \ldots\ldots\ldots X_i \ldots\ldots\ldots\ldots X_n$
Frequency	$f_1, f_2, \ldots\ldots\ldots f_i \ldots\ldots\ldots\ldots f_n$

The formula for moments about the point A takes the form

$$m_r' = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^r; \quad r = 0,1,2 \ldots \ldots ..$$

Table 1.2

Class Interval	Mid-point of the class	frequency
	x_1	f_1
	x_2	f_2
	‘	‘
	‘	‘
	‘	‘
	‘	‘
	x_i	f_i

	‘ ‘ ‘ x_n	‘ ‘ ‘ f_n
Total	----	$N=\sum fi$

We shall write it as “ x_i/f_i (I=1,2,……..n) distribution”

Where x_i is class mark of the i^{th} class, or its value of the variable X (Table 1.1), fi is its frequency and N= $\sum fi$ is total frequency. (number of observations)

If A=0, m_r' is r^{th} raw moments about natural origin.

$$m_0' = \frac{1}{n}\sum f_i x_i^0 = \frac{1}{n}\sum f_i = 1\ (always)$$

$$m_1' = \frac{1}{n}\sum f_i x_i = mean\ of\ the\ distribution = \bar{x}$$

$$m_2' = \frac{1}{n}\sum f_i x_i^2$$

$$m_3' = \frac{1}{n}\sum f_i x_i^3$$

$$m_4' = \frac{1}{n}\sum f_i x_i^4$$

Central Moments

If the arbitrary origin of moments of variable X is taken as arithmetic mean i.e. A = $\bar{x}$ the moments are called *central moments*.

Definition: For ungrouped data X_1, X_2, ………………..X_n the r^{th} central moment of variable X is given by

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r; \quad r = 0,1,2 \dots \dots ..$$

If the given values are classified into a frequency distribution, x_i/f_i (i=1,2,……..n), the r^{th} central moment is given by

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r; \quad r = 0,1,2 \dots \dots ..$$

x_i being the mid-value of i^{th} or x^{th} value of the variable (as the case may be) class and f_i its frequency

Evidently, we have

$$m_0 = 1$$

and

$$m_1 = 0\ (always)$$

The second central moment of variable X is variance of the distribution i.e.

$$V\ (x) = m_2 = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2; & (for\ ungrouped\ data) \end{cases}$$

and

$$V\ (x) = m_2 = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} f_i(x_1 - \bar{x})^r; & (for\ ungrouped\ data) \end{cases}$$

Third and fourth central moments for ungrouped and grouped data are-

$$m_3 = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3; & (for\ ungrouped\ data) \end{cases}$$

$$m_3 = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} f_i(x_i - \bar{x})^3; & (for\ grouped\ data) \end{cases}$$

$$m_4 = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n}(x_1 - \bar{x})^4; & (for\ ungrouped\ data) \end{cases}$$

$$m_4 = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} f_i(x_1 - \bar{x})^4; & (for\ grouped\ data) \end{cases}$$

Raw Moments expressed in terms of central moments

Just as central moments can be expressed in terms of moments about an arbitrary origin A, so a moment about an arbitrary origin is expressible in terms of central moments.

$$m'_1 = \bar{x} - A, and$$

$$m'_r = \frac{1}{n}\sum f_i(x_i - A)^r$$

$$= \frac{1}{N}\sum_i f_i(x_i - \bar{x} + \bar{x} - A)^r$$

$$= \frac{1}{N}\sum_i f_i\{(x_i - \bar{x}) + m'_1\}^r$$

$$= \frac{1}{N}\sum_i f_i\left\{(x_i - \bar{x})^r + r_{c_1}(x_i - \bar{x})^{r-1}m'_1 + r_{c_2}(x_i - \bar{x})^{r-2}m_1^2 + m_1^r\right\}$$

$$= \frac{1}{N}\sum_i f_i\,(x_i - \bar{x})^r + r_{c_1}\frac{1}{N}\sum_i f_i(x_i - \bar{x})^{r-1}m'_1 + r_{c_2}\sum_i f_i(x_i - \bar{x})^{r-2}m_1^2 + m_1^r$$

$$= m_r + r_{c_i}m_{r-1}m_1^1 + r_{c_i}m_{r-2}m_1^2 + \cdots \ldots \ldots \ldots .m_1^r$$

In particular,

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

$$\mu_3 = \mu_3 + 3\mu_2\mu_1 + \mu_1^3$$

$$\mu_4 = \mu_4 + 4\mu_3\mu_1 + 6\mu_2/\mu^2 + \mu_1^4$$

These formulae help us to obtain the moments about any point A, if the central moments are known.

Effect of change or origin and scale on central moments

If we change the origin of x on some point A and scale by h. The new variable u is defined $u = \frac{x-A}{h}$ as so that x=A+ hu, $\bar{x}$ =A +h$\bar{u}$ and

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

$$= \frac{1}{n}\sum f_i(hu_i - h\bar{u})^r$$

$$= h^r\frac{1}{N}\sum_i fi(u_i - \bar{u})^r$$

$$= h^r m_r(u)$$

$$where\ m_r^{(4)} = \frac{1}{N}\sum fi(u_i - \bar{u})^r$$

Thus, r^{th} central moment of variable X is h^r times r^{th} central moments of variable U. So we conclude that central moment is unaffected by change of origin but it is affected by change of scale.

Skewness:

Above, it has been seen that the measures of central tendency are deals with the location around which the mostly data or values are placed and the measures of dispersion deals with the variability. However, above both do not give much idea about the distribution. Here skewness is a measure which gives an idea regarding shape of the frequency distribution. When frequency distribution is not symmetrical it is supposed to be skewed.

A distribution is said to be skewed if mean, median and mode are not equal, quartiles are not placed on equidistance from median and also the curve of the distribution be asymmetrical and stretched more to one side than to other side.

Some measures of skewness are:

- *Absolute skewness = Mean – Mode*
- The *Karl Pearson coefficient of measure of Skewness:*

$$S_k = \frac{Mean - Mode}{Standard\ Deviation}$$

If mode is not well defined than

$$S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$$

if

Mean=Median=Mode	$S_k = 0$	Data is symmetrical
$Mean < Mode$ or $Mean < Median$	$S_k < 0$	Data is negatively skewed
$Mean > Mode$ or $Mean > Median$	$S_k > 0$	Data is positively skewed

The limits of Karl Pearson's coefficient of skewness is lies between -3 to 3.

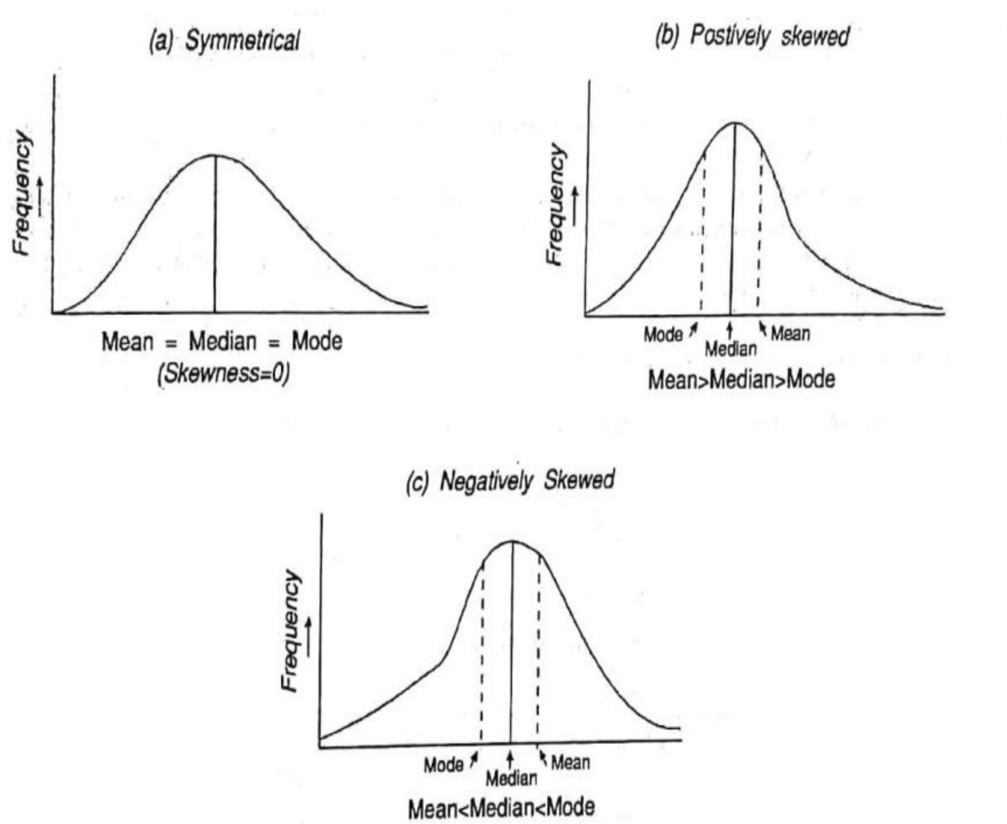
- The *Bowley's coefficient of measure of Skewness:*

It is also famous as Quartile coefficient of skewness.

$$S_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$Q_3 - Q_2 = Q_2 - Q_1$	$S_k = 0$	Data is symmetrical
$Q_3 + Q_1 < 2Q_2$	$S_k < 0$	Data is negatively skewed
$Q_3 + Q_1 > 2Q_2$	$S_k > 0$	Data is positively skewed
$Q_1 = Q_2$	$S_k = +1$	
$Q_2 = Q_3$	$S_k = -1$	

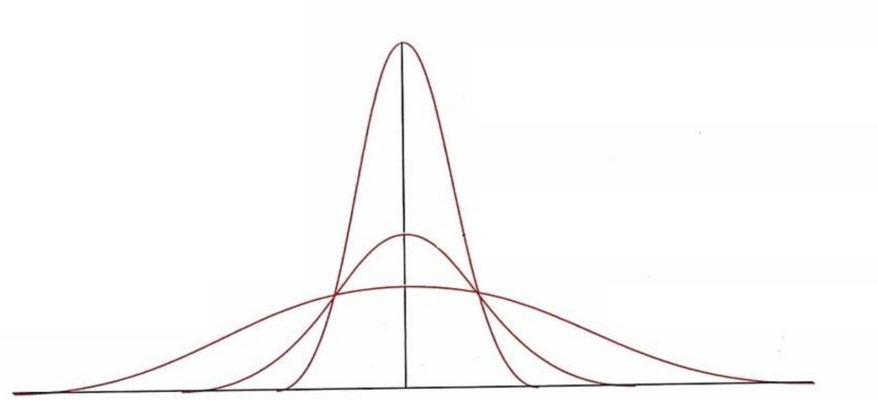
The limit of Bowley's coefficient of skewness is occurs between -1 to 1.



Kurtosis:

It gives an idea regarding the flatness or peakedness of the distribution. The peakedness and flatness are usually taken as relative to a normal bell shaped symmetric curve.

In terms of kurtosis, the normal curve (bell shaped) is called *mesokurtic*. If any curve is additional peaked than normal curve, is identified as *leptokurtic*. And if the curve is more flat than normal, is known as *platykurtic*.



In above figure, the middle curve is mesocurtic curve or ***Normal Bell*** shaped curve. The curve which one is more peaked is leptokurtic; and another is flatter than normal curve, is platykurtic.

The measure of kurtosis is,

$$K_u = \frac{\left\{\frac{(\sum_i^n (x_i - \bar{x})^4)}{n}\right\}}{\left\{\left(\frac{(\sum_i^n (x_i - \bar{x})^2)}{n}\right)^2\right\}}$$

If,

$K_u = 0$	Curve is normal or mesokurtic
$K_u < 0$	Curve is platykurtic
$K_u > 0$	Curve is leptokurtic