Chapter 4

Moments, Skewness and Kurtosis

Moments: The term 'moment' is used in physics and refers to the measure of force which may generate rotation. The possibility of generating such a force depends upon (i) the amount of force needed (ii) the distance from the origin of the point at which the force is applied. The term moment used in statistics is analogous to the term used in physics, where (i) size of class intervals represents the 'force' and (ii) deviation from mid-value of each class from an observation represents the distance.

Moments are popularly used to describe the characteristics of a distribution. According to Karl Pearson, first four moments are sufficient to describe a distribution. There are two kinds of moments. They are known as:

i) Raw moments (about origin and about any arbitrary value)

ii) Central moments or moments about mean or corrected moments

The moment system includes measures like mean, average deviation, standard deviation, skewness, kurtosis and so on.

Mathematically, if $x_1, x_2, \ldots\ldots\ldots\ldots x_n$ be n observations of a variate then the *r*th raw moment (about any arbitrary value or point A) is defined by

$$\mu_r' = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^r \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ \ldots\ldots\ldots\ldots \ (1)$$

*r*th raw moment (about origin) is defined by $\mu_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r$ ……....(2)

The *r*th corrected moment (moment about mean or central moment) is defined by

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r \text{, where } \bar{x} \text{ is arithmetic mean} \qquad \ldots\ldots\ldots\ldots\ldots\ (3)$$

<u>For frequency distribution:</u>

If $x_1, x_2, \ldots\ldots\ldots\ldots x_k$ occur with frequencies $f_1, f_2, \ldots\ldots\ldots\ldots f_k$ respectively then the *r*th raw moment about any arbitrary value A is

$$\mu_r' = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - A)^r \text{, where } n = \sum_{i=1}^{k} f_i \qquad \ldots\ldots\ldots\ldots\ \ldots\ldots\ldots\ldots\ (4)$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

The rth central or corrected moment is defined by

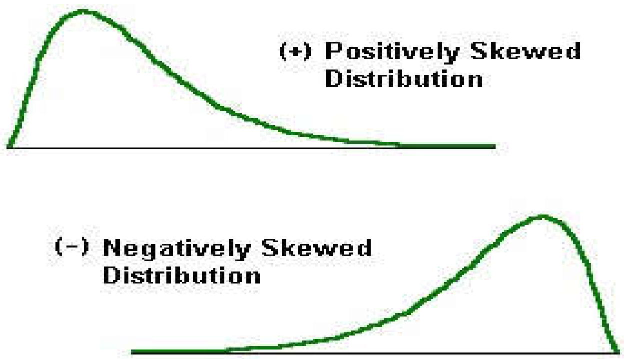
$$\mu_r = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^r \text{ , where } n = \sum_{i=1}^{k} f_i \text{ and } \bar{x} \text{ is the arithmetic mean}$$

Relation between raw moments and central moments:

(i) $\mu_1 = 0$ (ii) $\mu_2 = \mu_2^{\prime} - \mu_1^{\prime 2}$ (iii) $\mu_3 = \mu_3^{\prime} - 3\mu_2^{\prime}\mu_1^{\prime} + 2\mu_1^{\prime 3}$ (iv) $\mu_4 = \mu_4^{\prime} - 4\mu_3^{\prime}\mu_1^{\prime} + 6\mu_2^{\prime}\mu_1^{\prime 2} - 3\mu_1^{\prime 4}$

Skewness:

"A distribution is said to be 'skewed' when the mean and the median fall at different points in the distribution, and the centre of gravity is shifted to one side or the other to left or right." "Skewness is the lack of symmetry. A distribution which is not symmetrical is called a skewed distribution and in such distributions, the Mean, the Median and the Mode will not coincide, but the values are pulled apart. If the curve has a longer tail towards the right, it is said to be positively skewed. If the curve has a longer tail towards the left, it is said to be negatively skewed.

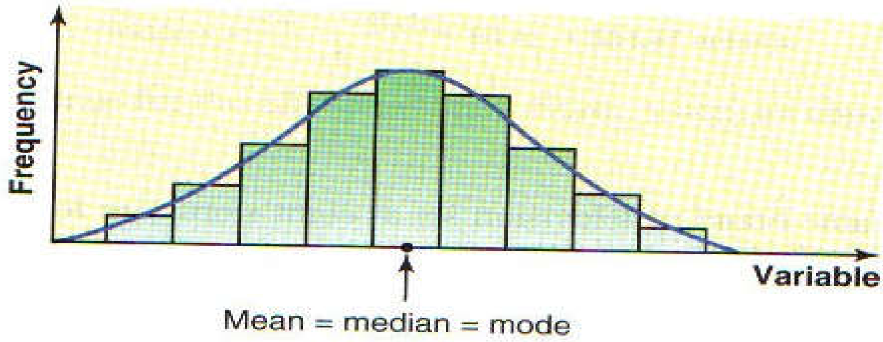


<u>Relationships among the Mean, Median, and Mode</u>

1. For a symmetric histogram and frequency curve with one peak, the values of the mean, median, and mode are identical, and they lie at the center of the distribution.

Figure: Mean, median, and mode for a symmetric histogram and frequency curve.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

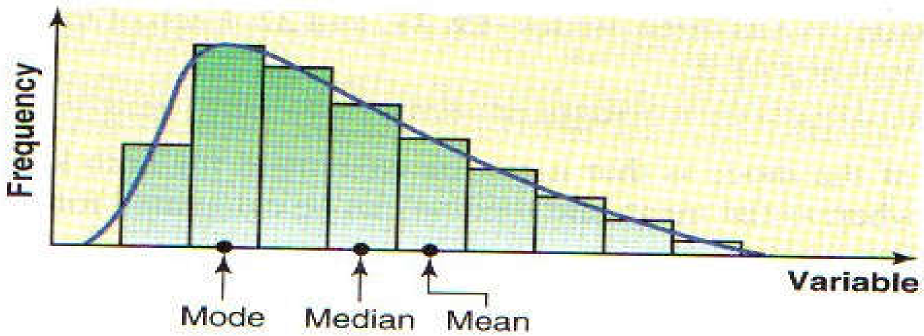


2. For a histogram and a frequency curve skewed to the right (Figure 3.3), the value of the mean is the largest that of the mode are the smallest, and the value of the median lies between these two.
    - Notice that the mode always occurs at the peak point.

The value of the mean is the largest in this case because it is sensitive to outliers that occur in the right tail.

- These outliers pull the mean to the right.

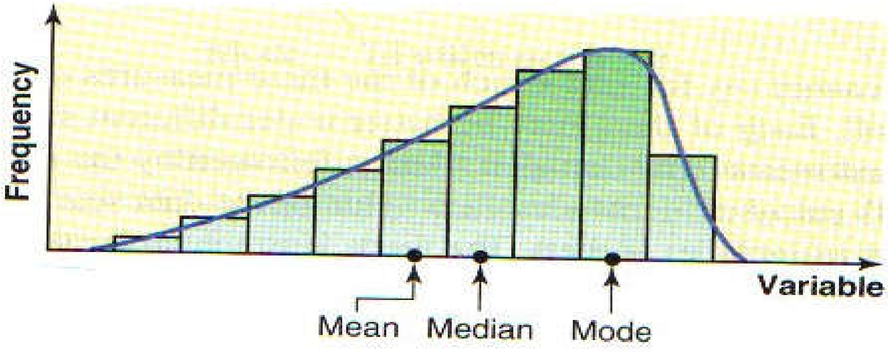
Figure: Mean, median, and mode for a histogram and frequency curve skewed to the right.



3. If a histogram and a distribution curve are skewed to the left (Figure 3.4), the value of the mean is the smallest and that of the mode is the largest, with the value of the median lying between these two.
    - In this case, the outliers in the left tail pull the mean to the left.

Figure: Mean, median, and mode for a histogram and frequency curve skewed to the right.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU



Measure of Skewness:

Karl Pearson's Coefficient pf Skewness $Sk_p = \frac{Mean - Mode}{S\tan dard\ \ Deviation}$

Mode=3Median-2Mean

Karl Pearson's Coefficient of Skewness based on moments is $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

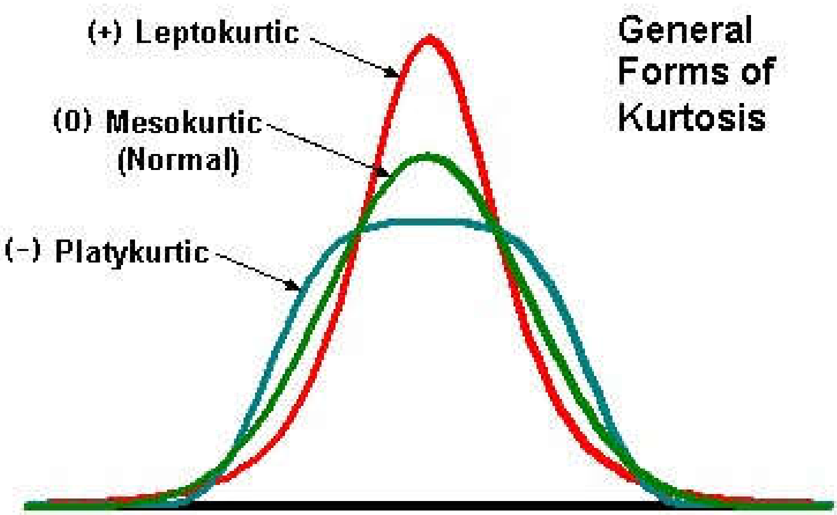
Kurtosis:

"The degree of a kurtosis of a distribution is measured relative to the peakness or a normal."

"A measure of kurtosis indicates the degree to which a curve of a frequency distribution is peaked or flat-topped."

Measures of kurtosis tell us the extent to which a distribution is more peaked or more flat topped than a normal curve. A normal curve which is symmetrical and bell-shaped is designed as Mesokurtic, because it is kurtic in the centre. If a curve is relatively narrower and peaked at the top, it designated as Leptokurtic. If the frequency distribution curve is more flat than normal curve, it is designated as Platykurtic.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU



Measures of Kurtosis:

The measures of kurtosis of a frequency distribution are based on upon the fourth moment about the mean of the distribution. Symbolically,

$$\beta_2 = \frac{\mu_4}{\mu^2{}_2}$$

Where, $\mu_4 = 4^{\text{th}}$ Moment, $\mu_2 = 2^{\text{nd}}$ Moment

If $\beta_2 = 3$, the distribution is said to be normal, and the curve is a normal curve (mesokurtic). If $\beta_2 > 3$, the distribution is said to be more peaked, and the curve is leptokurtic. If $\beta_2 < 3$, the distribution is said to be flat topped , and the curve is platykurtic.

Illustration:
The calculation of Moments about the mean is shown below:

X	(X-$\overline{X}$)	$(X-\overline{X})^2$	$(X-\overline{X})^3$	$(X-\overline{X})^4$
2	-4	16	-64	256
4	-2	4	-8	16
6	0	0	0	0
8	2	4	8	16
10	4	16	64	256
	0	40	0	544

$$\mu_1 = \frac{0}{5},\ \mu_2 = \frac{40}{5} = 8,\ \mu_3 = \frac{0}{5},\ \mu_4 = \frac{544}{5} = 108.8$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Hence, Kurtosis $\beta_2 = \frac{\mu_4}{\mu^2{}_2} = \frac{108.8}{8^2} = \frac{108.8}{64} = 1.7$, hence the distribution is platykurtic.

Example: The following frequency distribution refers to the profits of randomly selected 50 companies of a country:

Profits (in lakhs taka)	No. of companies
70-90	8
90-110	11
110-130	18
130-150	9
150-170	4

Compute first four central moments and hence β_1 and β_2. Also comment on your results.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU