Md- Abdul Alim ①

Moments, Skewness & Kurtosis —5

Moments: Moments are popularly used to describe the characteristics of a distribution. The Greek letter μ (read as mu) is generally used to denote the moments.

For ungrouped data: The r th moment of a variable X about the arithmetic mean $\bar{X}$ is given by

$$\mu_r = \frac{1}{N}\sum\left(X-\bar{X}\right)^r$$

The r th moment of a variable X about any arbitrary point A is given by

$$\mu'_r = \frac{1}{N}\sum\left(X-A\right)^r$$

For grouped data: The r th moment of a variable X about the arithmetic mean $\bar{X}$ is given by

$$\mu_r = \frac{1}{N}\sum f\left(X-\bar{X}\right)^r$$

The r th moment of a variable X about any arbitrary point A is given by

$$\mu'_r = \frac{1}{N}\sum f\left(X-A\right)^r$$

For different values of r, we shall get different moments. Thus if $r=1$, we will get the first moment, if we put $r=2$, we will get second moment and so on.

Moments about mean: Moments about mean is also called central moments.

For ungrouped data

$$\mu_1 = \frac{1}{N}\sum\left(X-\bar{X}\right); \qquad \mu_2 = \frac{1}{N}\sum\left(X-\bar{X}\right)^2$$

$$\mu_3 = \frac{1}{N}\sum\left(X-\bar{X}\right)^3; \qquad \mu_4 = \frac{1}{N}\sum\left(X-\bar{X}\right)^4$$

For grouped data

$$\mu_1 = \frac{1}{N}\sum f\left(X-\bar{X}\right); \qquad \mu_2 = \frac{1}{N}\sum f\left(X-\bar{X}\right)^2$$

$$\mu_3 = \frac{1}{N}\sum f\left(X-\bar{X}\right)^3; \qquad \mu_4 = \frac{1}{N}\sum f\left(X-\bar{X}\right)^4.$$

Note The first moment about the origin tells us about the mean, the second moment about variance, the third moment about Skewness and the fourth moment about the kurtosis.

Moments about Arbitrary point: When actual mean is in fraction, moments are first calculated about an arbitrary point and then converted to moments about the actual mean. When deviations are taken from arbitrary point A, the formulae are

$$\mu'_1 = \frac{1}{N}\sum\left(X-A\right); \qquad \mu'_2 = \frac{1}{N}\sum\left(X-A\right)^2$$

$$\mu'_3 = \frac{1}{N}\sum\left(X-A\right)^3; \qquad \mu'_4 = \frac{1}{N}\sum\left(X-A\right)^4$$

In a frequency distribution, to simplify calculations we can take $d = \frac{X-A}{i}$ or $(X-A) = id$, we get

$$\mu_1' = \frac{1}{N}\sum f(X-A) \quad \text{or} \quad \frac{\sum fd}{N} \times i$$

$$\mu_2' = \frac{1}{N}\sum f(X-A)^2 \quad \text{or} \quad \frac{\sum fd^2}{N} \times i^2$$

$$\mu_3' = \frac{1}{N}\sum f(X-A)^3; \quad \text{or} \quad \frac{\sum fd^3}{N} \times i^3$$

$$\mu_4' = \frac{1}{N}\sum f(X-A)^4 \quad \text{or} \quad \frac{\sum fd^4}{N} \times i^4$$

Finding Central moments from moments about Arbitrary point: With the help of following relationships, moments about an arbitrary point can be converted to moments about mean:

$$\mu_1 = 0$$

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

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

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

Example: You are given the following frequency distribution of the daily earning of employees in a company:

Earnings (in Tk.)	Number of workers	Earnings (in Tk.)	Number of workers
50-70	4	130-150	6
70-90	8	150-170	7
90-110	12	170-190	3
110-130	20		

Calculate the first four moments about the point 120. Convert the result into moments about the mean.

Solution: We know that the moments about any arbitrary point A is

$$\mu_r' = \frac{1}{N}\sum f(X-A)^r$$

Here $A = 120$ and X are the mid-points. To get the first four moments, put $r = 1, 2, 3$ and 4 in the above formula.

Computation of first four moments

Earnings (in Tk.)	Mid-point X	f	$d = \frac{(X-120)}{20}$	fd	fd^2	fd^3	fd^4
50-70	60	4	-3	-12	36	-108	324
70-90	80	8	-2	-16	32	-64	128
90-110	100	12	-1	-12	12	-12	12
110-130	120	20	0	0	0	0	0
130-150	140	6	1	6	6	6	6
150-170	160	7	2	14	28	56	112
170-190	180	3	3	9	27	81	243
Total		60		-11	141	-41	825

Md. Moyazzem Hossain

So moments about the arbitrary point 120 are

$$\mu_1' = \frac{\sum fd}{N} \times i = \frac{-11}{60} \times 20 = -3.6667$$

$$\mu_2' = \frac{\sum fd^2}{N} \times i^2 = \frac{141}{60} \times (20)^2 = 940$$

$$\mu_3' = \frac{\sum fd^3}{N} \times i^3 = \frac{-41}{60} \times (20)^3 = -5466.667$$

$$\mu_4' = \frac{\sum fd^4}{N} \times i^4 = \frac{825}{60} \times (20)^4 = 2200000$$

Moments about mean

We know that $\mu_1 = 0$.

$$\mu_2 = \mu_2' - (\mu_1')^2 = 940 - (-3.667)^2 = 926.5553.$$

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

$$= -5466.6667 - 3(-3.6667)(940) + 2(-3.6667)^2 = 4774.832$$

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

$$= 2200000 - 4(-3.6667)(-5466.6667) + 6(-3.6667)^2(940) - 3(-3.6667)^4$$

$$= 2195107.3$$

Skewness: The term "Skewness" refers to lack of symmetry or departure from symmetry. e. g., when a distribution is not symmetrical (or is asymmetrical) it is called a skewed distribution. The measures of Skewness indicate the difference between the manner in which the observations are distributed in a particular distribution compared with a symmetrical (or normal) distribution.

In a symmetrical distribution the values of mean, median and mode are alike. Hence the two tails are equal length. In a skewed distribution the values differ. If the value of mean is greater than the mode, Skewness is said to be ***positive***. Hence the right tail is longer than the left tail. On the other hand, if the value of mode is greater than mean, Skewness is said to be ***negative***. Hence the left tail is longer than the right tail. The following diagrams would show different Skewness:

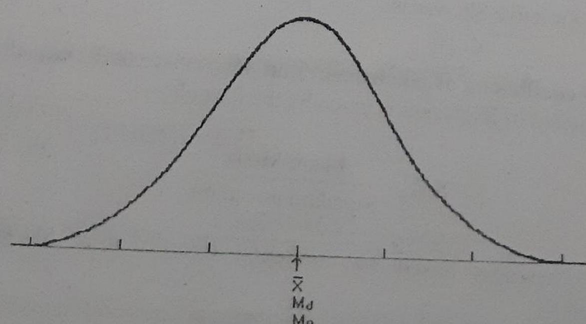


Figure: Symmetrical Distribution

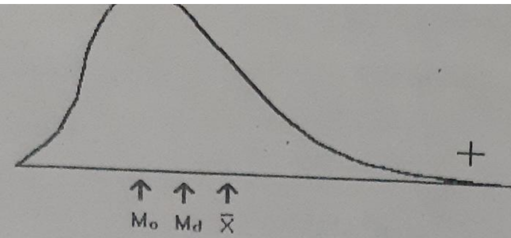


Figure: Positively Skewed Distribution

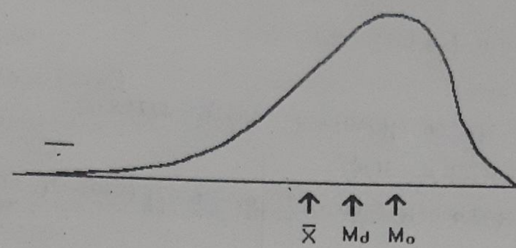


Figure: Negatively Skewed Distribution

Absolute measures of Skewness: Different absolute measures of Skewness are:

1) $S.K = 3(Mean - Median)$

2) $S.K = Mean - Mode$

Coefficients of Skewness: For comparing two series, we do not calculate the absolute measures but we calculate the relative measures called the co-efficient of Skewness which are pure numbers i., e., independent of units of measurement. The following are two important methods of measuring relative Skewness:

1. **Karl Pearson's coefficient of Skewness:** Karl Pearson's coefficient of Skewness or Pearsonian coefficient of Skewness is given by the formula:

$$S.K_p = \frac{\text{Mean-Mode}}{\text{Standard deviation}}$$

2. **Coefficient of Skewness based on moments:** Coefficient of Skewness based on moments is given by

$$C.S.K = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

Where, $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ and $\beta_2 = \frac{\mu_4}{\mu_2^2}$.

Example: The following data relate to the marks of 40 students:

Marks	0-5	5-10	10-15	15-20	20-25	25-30
Number of students	3	5	10	12	6	4

Calculate the coefficient of Skewness.

Solution: We know that, the coefficient of Skewness is defined as

$$S.K_p = \frac{\text{Mean-Mode}}{\text{Standard deviation}}$$

Calculation of coefficient of Skewness

Marks	No. of students (f)	Mid point (x)	fx	x^2	fx^2
0-5	3	2.5	7.5	6.25	18.75
5-10	5	7.5	37.5	56.25	281.25
10-15	10	12.5	125	156.25	1562.5
15-20	12	17.5	210	306.25	3675
20-25	6	22.5	135	506.25	3037.5
25-30	4	27.5	110	756.25	3025
Total	40		625		11600

Calculation of mean: We know that, $\bar{x} = \frac{\sum fx}{N}$.

$$\therefore \bar{x} = \frac{625}{40} = 15.625$$

Calculation of mode: We know that, $Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$.

By inspection mode lies in the class 15-20.

$$\therefore Mode = 15 + \frac{2}{2+6} \times 5 = 16.25$$

Calculation of standard deviation: We know that, $\sigma = \sqrt{\frac{\sum fx^2}{N} - \bar{x}^2}$.

$$\therefore \sigma = \sqrt{\frac{11600}{40} - (15.625)^2} = 6.77.$$

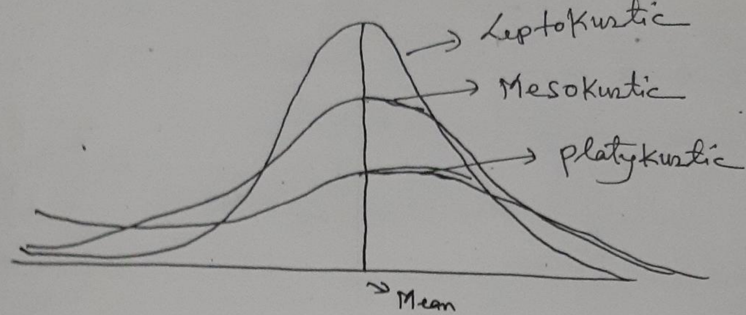
$$\therefore \text{Coefficient of Skewness} = \frac{15.625 - 16.25}{6.77} = -0.0923$$

Since, the coefficient of Skewness is negative so that the distribution of marks is negatively skewed.

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, JU

Kurtosis: In Statistics, kurtosis refers to the degree of flatness or peakedness in the region about the mode of a frequency curve. The degree of kurtosis of a distribution is measured relative to the peakedness of a normal curve.

Karl Pearson, in 1905 introduced three broad patterns of peakedness which are illustrated in the following diagram:



1) If a curve is more peaked than the normal curve, it is called "**leptokurtic**".
2) An intermediate peaked curve which is neither flat topped nor peaked is known as normal or "**mesokurtic**".
3) If a curve is more or flat-topped than the normal curve, it is called "**platykurtic**".

Measures of Kurtosis: Kurtosis is measured by β_2 or γ_2. Where $\beta_2 = \frac{\mu_4}{\mu_2^2}$ and $\gamma_2 = \beta_2 - 3$.

For a normal or **mesokurtic** curve $\beta_2 = 3$ and $\gamma_2 = 0$.

For a **leptokurtic** curve $\beta_2 > 3$ and $\gamma_2 > 0$.

For a **platykurtic** curve $\beta_2 < 3$ and $\gamma_2 < 0$.

Example: From the following data calculate kurtosis.

Age (years)	20-25	25-30	30-35	35-40	40-45	45-50	50-55
Number of employees	8	12	20	25	15	12	8

Solution: We know that, kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$.

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, JU

Calculation of kurtosis

Age (years)	Number of employees (f)	mid point (x)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$	$(x-\bar{x})^4$	$f(x-\bar{x})^4$
20-25	8	22.5	180	217.56	1740.50	47333.44	378667.53
25-30	12	27.5	330	95.06	1140.75	9036.88	108442.55
30-35	20	32.5	650	22.56	451.25	509.07	10181.33
35-40	25	37.5	937.5	0.06	1.56	0	0
40-45	15	42.5	637.5	27.56	413.44	759.69	11395.37
45-50	12	47.5	570	105.06	1260.75	11038.13	132457.55
50-55	8	52.5	420	232.56	1860.50	54085.32	432682.53
Total	100		3725		6868.75		1073826.95

We know that, $\mu_4 = \dfrac{\sum f(x-\bar{x})^4}{N}$ and $\mu_2 = \dfrac{\sum f(x-\bar{x})^2}{N}$.

$$\therefore \mu_4 = \frac{1073826.95}{100} = 10738.27 \text{ and } \mu_2 = \frac{6868.75}{100} = 68.69$$

We have, $\beta_2 = \dfrac{10738.27}{(68.69)^2} = 2.28$.

Since, $\beta_2 < 3$ so the distribution is **platykurtic**.