Chapter 5

Skewness and Kurtosis

Prepared by ***Jannatul Ferdous***, Assistant Professor, Department of Computer Science and Engineering

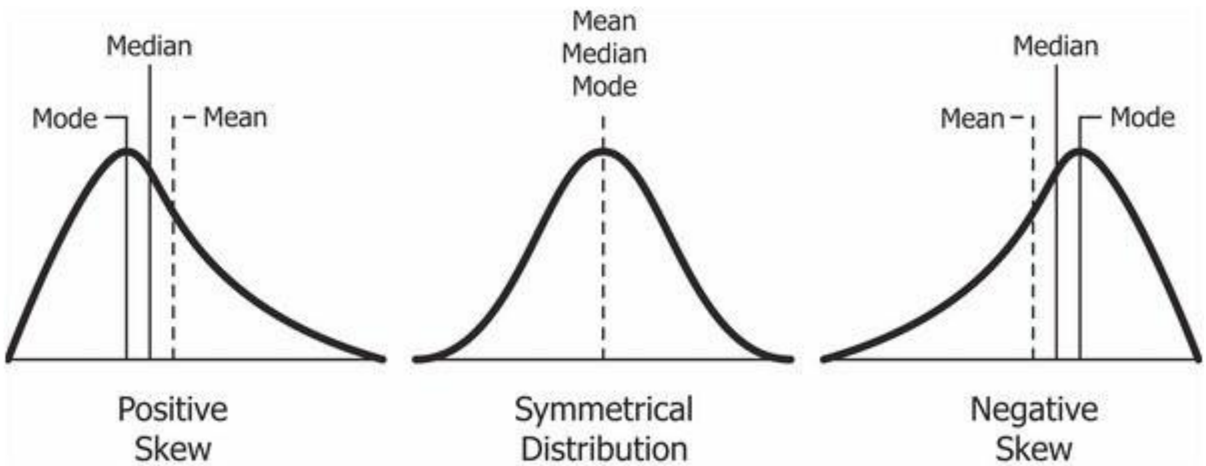
The average and measure of dispersion can describe the distribution but they are not sufficient to describe the nature of the distribution. For this purpose we use other concepts known as Skewness and Kurtosis.

Skewness

Skewness means lack of symmetry. A distribution is said to be symmetrical when the values are uniformly distributed around the mean. For example, the following distribution is symmetrical about its mean 3.

x	:	1	2	3	4	5
Frequency (f)	:	5	9	12	9	5

The symmetrical and skewed distributions are shown by curves as



If Mean = Mode, the skewness is zero.
If Mean > Mode, the skewness is positive.
If Mean < Mode, the skewness is negative.

In a symmetrical distribution the mean, median and mode coincide, that is, mean = median = mode. It is also called as normal distribution.
Several measures are used to express the direction and extent of skewness of dispersion. The important measures are that given by Pearson. The first one is the Coefficient of Skewness:

$$S_k = \frac{3(mean - median)}{standard\ deviation}$$

or,

$$S_k = \frac{mean - mode}{standard\ deviation}$$

S_k= Pearsonian (or Pearson's) coefficient of skewness

For a symmetric distribution S_k = 0. If the distribution is negatively skewed then S_k is negative and if it is positively skewed then S_k is positive. The range for S_k is from -3 to 3.

The other measure uses the β (read 'beta') coefficient which is given by,

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

where, μ_2 and μ_3 are the second and third central moments. The second central moment μ_2 is nothing but the variance. The sample estimate of this coefficient is $b_1 = \sqrt{\frac{m_3^2}{m_2^3}}$ or $b_1 = \frac{m_3}{m_2^{\frac{3}{2}}}$

where m_2 and m_3 are the sample central moments given by

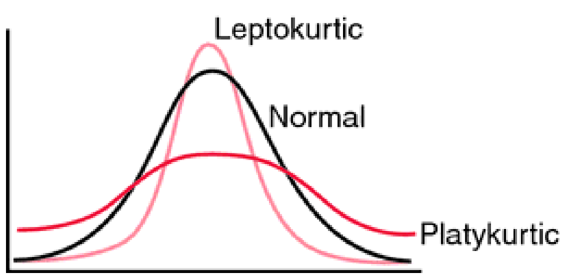
$$m_2 = variance = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} or \frac{\sum_{i=1}^{k} f_i(x_i - \overline{x})^2}{n}$$

$$m_3 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{n} or \frac{\sum_{i=1}^{k} f_i(x_i - \overline{x})^3}{n}$$

For a symmetrical distribution b_1 = 0. Skewness is positive or negative depending upon whether m_3 is positive or negative.

Kurtosis

A measure of the peaked ness or convexity of a curve is known as Kurtosis.



It is clear from the above figure that all the three curves, are symmetrical about the mean. Still they are not of the same type. One has different peak as compared to that of others. Curve 1(black) is known as mesokurtic (normal curve); Curve 2 (pink) is known as leptokurtic (leading curve) and Curve 3 (red) is known as platykurtic (flat curve).

Kurtosis is measured by,

Pearson's coefficient, β_2 (read 'beta - two'). It is giveen by $\beta_2 = \frac{\mu_4}{\mu_2^2}$

The sample estimate of this coefficient is $b_2 = \frac{m_4}{m_2^2}$

where, m_4 is the fourth central moment given by $m_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}$ or $\frac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^4}{n}$

- A normal distribution has kurtosis exactly 3 (excess kurtosis exactly 0). Any distribution with kurtosis ≈ 3 (excess ≈ 0) is called **mesokurtic**.
- A distribution with kurtosis < 3 (excess kurtosis < 0) is called **platykurtic**. Compared to a normal distribution, its tails are shorter and thinner, and often its central peak is lower and broader.
- A distribution with kurtosis > 3 (excess kurtosis > 0) is called **leptokurtic**. Compared to a normal distribution, its tails are longer and fatter, and often its central peak is higher and sharper.

Example 1:

Class Mark, x	Frequency, f	xf	$(x-\bar{x})$	$(x-\bar{x})^2 f$	$(x-\bar{x})^3 f$
61	5	305	-6.45	208.01	-1341.68
64	18	1152	-3.45	214.25	-739.15
67	42	2814	-0.45	8.51	-3.83
70	27	1890	2.55	175.57	447.70
73	8	584	5.55	246.42	1367.63
	∑	6745		852.75	−269.33
		$\bar{x}$ = 67.45		m_2 = 8.5275	m_3 = −2.6933

Solution 1:

The skewness is

$$b_1 = m_3 / m_2^{3/2} = -2.6933 / 8.5275^{3/2} = -0.1082$$

So the distribution is negatively skewed.

Example 2:

The first central moments of a distribution are 0, 16, -36 and 120. Comment on the skewness and of the distribution.

Solution 2:

We are given $m_{1=0,}$ $m_{2=}$ 16, $m_{3=}$ - 36 and $m_{4=}$ 120

Now, we know that, co-efficient of skewness, $b_1 = \frac{m_3}{m_2^{\frac{3}{2}}} = \frac{-36}{16^{\frac{3}{2}}} = \frac{-36}{64}$ = -0.5625

So, the distribution is negatively skewed.

Co-efficient of kurtosis, $b_2 = \frac{m_4}{m_2^2} = \frac{120}{16^2}$ = 0.46875

Since the value of the Co-efficient of kurtosis, b_2 is less than 3, so the distribution is platykurtic.