**Skewness and Kurtosis**

Look at the following frequency distributions and identify the features of the following frequency distributions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| i) | : | 10 | 15 | 20 | 25 | 30 |  |
|  | : | 3 | 7 | 16 | 7 | 3 |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ii) | : | 10 | 15 | 20 | 25 | 30 | 35 |
|  | : | 3 | 7 | 16 | 16 | 7 | 3 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| iii) | : | 5-8 | 9-12 | 13-16 | 17-20 | 21-24 |  |
|  | : | 7 | 18 | 23 | 18 | 7 |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| iv) | : | 5-8 | 9-12 | 13-16 | 17-20 | 21-24 | 25-28 |
|  | : | 7 | 18 | 23 | 23 | 18 | 7 |

It may be noted that in the above distributions, the means are respectively 20, 22.5, 14.5 and 16.5, and in each case, the value of the variable equidistant from the mean have equal frequencies. It can be found that in each case, the median and mode also are equal to the mean.

A frequency distribution of a **discrete variable** is called symmetrical about the value if the frequency of is the same as the frequency of , whatever may be. That is, values of the variable equidistant from have equal frequencies.

**In case of a continuous variable**, the term symmetry should be used in relation to its frequency curve. The frequency curve of a continuous variable is said to be symmetrical about if the frequency-density at is the same as the frequency-density at , whatever may be.

**For any symmetrical distribution mean, median and mode are equal.**

**A distribution which is not symmetrical is called asymmetrical or skew.** In general, frequency distributions are not symmetrical.Some distributions are slightly asymmetrical and some others may be highly asymmetrical.

**Skewness**

The word *skewness* literally denotes *asymmetry* or *lack of symmetry*. By skewness of a frequency distribution we mean the degree of its departure from symmetry. A symmetrical distribution has zero skewness. Skewness may be positive or negative.

The skewness is said to be *positive* if the longer tail of the distribution is towards the higher values of the variable and *negative* if the longer tail is towards the lower values of the variable.

|  |  |
| --- | --- |
| C:\Users\SUSANTA KUMAR GAURI\Desktop\Bell shaped3.jpgFig. 1: Positively skew frequency curve | C:\Users\SUSANTA KUMAR GAURI\Desktop\Bell shaped2.jpg Fig. 2: Negatively skew frequency curve |

**Measures of skewness**

There are different approaches for measuring skewness of a frequency distribution or frequency curve. These are as follows:

1. ***Moment measure***

An important point to be noted in this connection is that all first order central moment () is necessarily zero for any distribution (symmetrical as well as skewed distribution). All other odd-order central moments are also zero for a symmetrical distribution. But, all other odd-order central moments are positive for a positively skewed distribution and negative for a negatively skewed distribution.

Any odd-order central moment except the first order central moment () may, therefore, be considered a measure of the skewness of distribution. The simplest of these measures is .

To make this measure free from the units of the variable, is divided by . Thus we get the measure of skewness as

Skewness,

Here, theoretically can assume any value between , but in practice its value is rarely very high.

Sometimes, is used as a measure of skewness and it is denoted as . Thus

, or

1. ***Pearsons’ first measure***

In a symmetrical distribution, the mean, median and mode (assuming the distribution to be unimodal) coincide. If the distribution is positively skewed, then

and if the distribution is negatively skewed, then

Hence, the difference may be considered as a measure of the skewness of distribution. To make this measure free from the units of the variable, it is divided by the SD. So the measure of skewness becomes

Here, the value of can approximately vary between -3 and 3.

1. ***Pearsons’ second measure***

It is difficult to estimate the mode from a frequency distribution and thus to get the value of . In such cases, the empirical relation, may be used. So, another measure of skewness can be as follows:

The value of can approximately vary between -3 and 3.

1. ***Bowley’s measure***

For a symmetrical distribution, the lower and upper quartiles are equidistant from the median; for a positively skew distribution the lower quartile is nearer the median than the upper quartile is, while for a negatively skew distribution the upper quartile is nearer. Thus, may be taken as a measure of skewness. It is expressed as a pure number on being divided by . Thus, the measure becomes

Here, the value of can approximately vary between -1 and 1.

**Kurtosis**

Kurtosis refers to the degree of peakedness of the frequency curve. Two distributions may have the same average, dispersion and skewness, but one of them may be more peaked (implying high concentration of values near the mode) than other. This characteristic of the frequency distribution is known as kurtosis.

There is only one measure of kurtosis and it is based on moments as follows:

Kurtosis,

Sometimes, is used as a measure of kurtosis and it is denoted as . Thus

, or

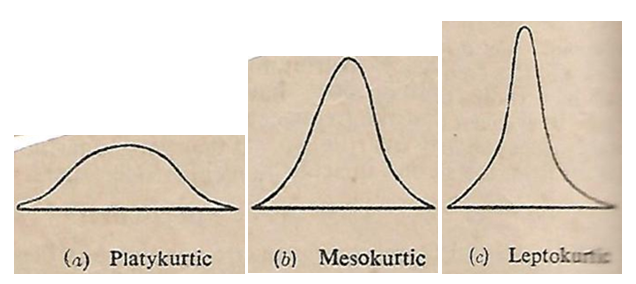
For a normal distribution, this measure () has the value zero. A normal distribution is considered neither very peaked nor flat-top. ***A positive value of*** indicates that the distribution has high concentration of values near the central tendebcy and has high tails, in comparison with a normal distribution with the same standard deviation. ***A negative value of*** *means* that the distribution has low concentration of values near the central tendency and low tails, in comparison with to a normal distribution with the same standard deviation.

**Classification of distributions based on kurtosis value**

**Mesokurtic distribution:** The frequency curve for a mesokurtic distribution is neither very peaked nor flat-top (e.g. normal distribution). For mesokurtic distribution, or .

**Leptokurtic distribution:** The frequency curve for a leptokurtic distribution is relatively high peaked. For leptokurtic distribution, or .

**Platykurtic distribution:** The frequency curve for a platykurtic distribution is relatively flat-topped. For platykurtic distribution, or .



**Fig. 3**: Different type of Kurtosis

**Exercise 1:** A frequency distribution gives the following results: (i) coefficient of variation = 5, (ii) variance = 4, (iii) Pearson’s coefficient of skewness = 0.5. Find the mean and the mode of the distribution. [Ans. 40, 39]

**Exercise 2:** From the data given below, calculate the coefficient of variation: Pearson’s measure of skewness = 0.42; mean = 86; median = 80. [Ans. 50]

**Exercise 3:** Calculate the coefficient of skewness based on quartiles from the following frequency data:

|  |  |  |  |
| --- | --- | --- | --- |
| Observations | Frequency | Observations | Frequency |
| More than 0 | 5474 | More than 60 | 2718 |
| More than 10 | 5426 | More than 70 | 1406 |
| More than 20 | 5259 | More than 80 | 764 |
| More than 30 | 5023 | More than 90 | 370 |
| More than 40 | 4475 | More than 100 | 160 |
| More than 50 | 3712 | More than 110 | 39 |

[Ans. Skewness = -0.16]

**Exercise 4:** If the first quartile is 142 and the semi-interquartile range = 18, find the median (assuming the distribution is symmetrical about mean or median]

[Ans. Median = 160]

**Exercise 5:** Find a suitable measure of skewness for the following distribution:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Annual sales (Rs.’000) | 0-20 | 20-50 | 50-100 | 100-250 | 250-500 | 500-1000 |
| No. of firms | 20 | 50 | 69 | 30 | 25 | 19 |

[Ans. Skewness = 0.55]

**Exercise 6:** The frequency distribution of a large number of balls when classified according to their radius is symmetrical. Show that if we classify the balls according to their volume, the resulting frequency distribution will be skewed.

Ans.: Let , and represent the first, second and third quartiles of the frequency distribution of radius . Then 25% of the balls have radius up to , 50% have radios up to and 75% up to .

It is given that the frequency distribution of radius is symmetrical; hence the first quartile and the third quartile are equidistant from the second quartile on either side.

Let , and ( positive).

If now , and represent the corresponding quartiles of the frequency distribution of volume V, then 25% of the same balls have radius up to , 50% have volume up to and 75% up to .

Again, since the volume V of a ball of radius R is given by , and 25% of te balls have radius upto , the same balls have volume up to . Substituting and for V and R respectively,

Similarly,

Using Bowley’s formula for the frequency distribution of volume V,

This is positive (and not zero), because and are positive.

Thus the frequency distribution of volume is positively skew.