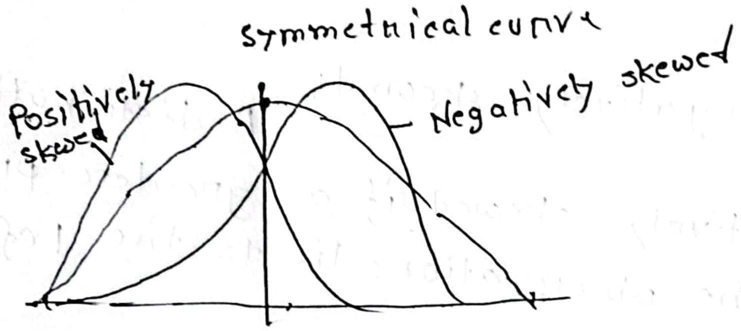
~~Shape of~~ a characteristics of a distribution.

~~Ske~~ the shape of characteristics of a distribution are measured by skewness and kurtosis.

Skewness: Skewness means the lack of symmetry of a distribution. That is, when a distribution is not ~~skewed~~ symmetrical, it is called skewed distribution.

A distribution may be symmetrical, positively skewed or negatively skewed.



(a) Symmetrical distribution:

~~The type~~ o· A distribution is symmetric if left and sides of the distribution are divided at the middle value. In this case, the values of mean, median, mode of the distribution are equal. Normal distribution is an example of symmetrical distribution.

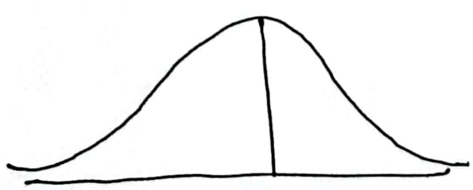
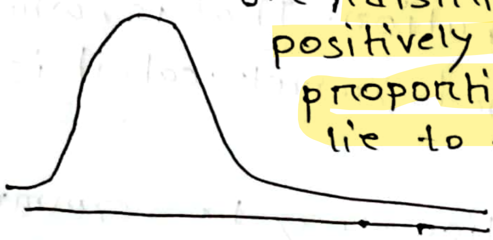


Fig: Symmetrical curve

(b) Positively skewed distribution: In this distribution the long tail to the right indicates the presence of extreme values at the positive end of the distribution. This type of distribution is known as positively skewed.



or A distribution is called positively skewed if a greater proportion of the observations lie to the right side.

Figure: positive skewed curve

(c) Negatively skewed: A distribution is called negatively skewed if a greater proportion of the observations lie to the left sides.

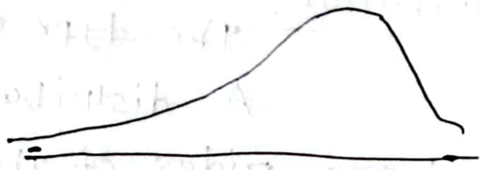


Figure: Negatively skewed.

Kurtosis: kurtosis is defined as the degree of flatness or peakedness of a distribution relative to a normal distribution.

there are three types of kurtosis: they are i) Mesokurtic, (ii) Leptokurtic and (iii) Platykurtic

(i) Mesokurtic: the curve, which neither flat or nor peaked, is called normal curve or mesokurtic. In this curve $\beta_2 = 3$, i.e $\gamma_2 = 0$

(ii) Leptokurtic: If a curve is more peaked than normal curve, it is called laptokurtic. In this case $\beta_2 > 3$, $\gamma_2 > 0$

(iii) Platykurtic: when a curve is less peaked than the normal curve, it is called platykurtic. In this cas $\beta_2 < 3$, and $\gamma_2 < 0$.

the diagram below illustrates the three different types of curves as mentioned above.

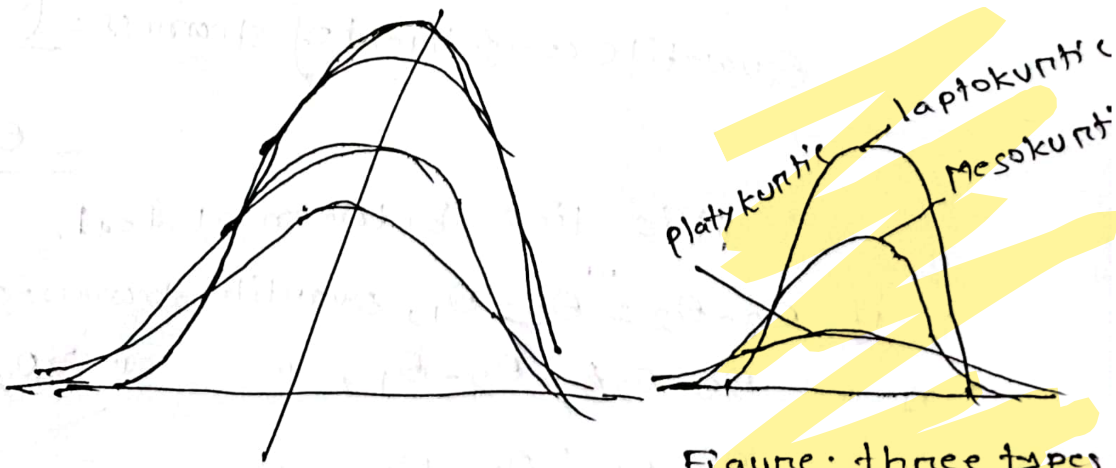


Figure: three types of kurtosis

Some important measures of skewness:

The simplest measure of skewness is the pearson's co-efficient of skewness defined as

$$\text{Pearson's co-efficient of skewness} = \frac{\text{mean} - \text{mode}}{\text{Standard deviation}}$$

if mean > mode, the skew is positive

" " < " , " " negative

" mean = mode, the distribution is symmetrical.

The pearson's co-efficient of skewness assumes the following modified form

$$\text{Pearson co-efficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{Standard deviation}}$$

In terms of the three quartiles Q_1, Q_2 and Q_3, Bowley's Quartile co-efficient of skewness is

$$\text{Quartile co-efficient of skewness} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

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

This lies between -1 to +1.

- if $Q_3 - Q_2 = Q_2 - Q_1$, quartile skewness = 0, the distribution is symmetrical.
- " $Q_3 - Q_2 > Q_2 - Q_1$, " " > 0, the distribution is positively skewed
- " $Q_3 - Q_2 < Q_2 - Q_1$, " " < 0, the distribution is negatively skewed.

Measures of kurtosis: kurtosis is measured by the co-efficient β_2 or its derivation γ_2 given by $\beta_2 = \frac{\mu_4}{\mu_2^2}$ and $\gamma_2 = \beta_2 - 3$.

Here ~~pure~~ β_2 is a pure number and positive

mesokurtic, B2=3
Laptokurtic, B2>3
playtokurtic, B2<3

Moments of a distribution.

Moments: moments are popularly used to describes the characteristics of a distribution.

According to karl pearson, first four moments are sufficient to describes a distribution.

There are two types of moments, they are 1) Raw moment and
2) Central or corrected moment.

But moments are calculated in three different ways. They are

i) Raw moments about origin xi-0
ii) Raw moments about arbitrary value. xi-A
iii) Moments about mean or corrected moments. xi-u

moment about mean are also known as central moment.

The rth Moment about origin:

let $x_1, x_2 \dots x_n$ be n values of a variable, then rth moment about origin, denoted by $V_r = \frac{\Sigma x^r}{n}$. we get first, second, third and fourth raw moments about origin by putting $r = 1, 2, 3, 4$.

1st raw moment $= V_1 = \frac{\Sigma x}{n} = \bar{x}$ = sample mean

2nd " " $= V_2 = \frac{\Sigma x^2}{n}$

3rd " " $= V_3 = \frac{\Sigma x^3}{n}$

4th " " $= V_4 = \frac{\Sigma x^4}{n}$

The rth moment about Arbitrary:

The rth raw moment of a variable X about any arbitrary value A denoted by $\mu'_r = \frac{\Sigma (X-A)^r}{n}$

we get the 1st, 2nd, 3rd and 4th raw moments about arbitrary value A by putting $r = 1, 2, 3, 4$ in the formula μ'_r. that is

$$\mu'_1 = \frac{\Sigma (x_i - A)}{n}$$

$$\mu'_2 = \frac{\Sigma (x_i - A)^2}{n}$$

$$\mu'_3 = \frac{\Sigma (x_i - A)^3}{n}$$

$$\mu'_4 = \frac{\Sigma (x_i - A)^4}{n}$$

The rth corrected or Central moment:

The rth corrected or central moment of the variable x denoted by μ_r is defined as

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

First, second, third and fourth central moments can be obtained by putting $r = 1, 2, 3, 4$ in the formula of μ_r

Such as: $\mu_1 = \frac{\sum(x_i-\bar{x})}{n} = \frac{\sum x_i - n\bar{x}}{n} = \frac{n\bar{x}-n\bar{x}}{n} = 0$

$$\mu_2 = \frac{\sum(x_i-\bar{x})^2}{n} = s^2 = \text{sample Variance}$$

S^2

$$\mu_3 = \frac{\sum(x_i-\bar{x})^3}{n}, \quad \mu_4 = \frac{\sum(x_i-\bar{x})^4}{n}$$

It is noted that first raw moment about origin is arithmatic mean and second corrected moment is the sample variance.

Example: In a firm there are five employees. The number of days absent by these employess for the last years are 24, 27, 30, 31 and 33. Calculate the first four moments about mean, or first corrected moments or first four central moments.

Soln: we consider the observations as population data. the formula for finding the moments about mean are

$$\mu_1 = \frac{\Sigma(x_i-\mu)}{N}, \mu_2 = \frac{\Sigma(x-\mu)^2}{N}, \mu_3 = \frac{\Sigma(x_i-\mu)^3}{N}$$

$$\mu_4 = \frac{\Sigma(x_i-\mu)^4}{N}$$

The mean age $= \mu = \frac{24+27+30+31+33}{5} = 29$ days

Table for calculation

x_i	$x_i-\mu$	$(x_i-\mu)^2$	$(x_i-\mu)^3$	$(x_i-\mu)^4$
24	-5	25	-125	625
27	-2	4	-8	16
30	1	1	1	1
31	2	4	8	16
33	4	16	64	256
total	0	50	-60	914

$$\mu_1 = \frac{\Sigma(x_i-\mu)}{N} = \frac{0}{5} = 0$$

$$\mu_2 = \frac{\Sigma(x_i-\mu)^2}{N} = \frac{50}{5} = 10 = \text{Variance}$$

$$\mu_3 = \frac{\Sigma(x_i-\mu)^3}{N} = \frac{-60}{5} = -12$$

$$\mu_4 = \frac{\Sigma(x_i-\mu)^4}{N} = \frac{914}{5} = 182.8$$