PROBABILIY AND STATISTICS I

LECTURE EIGHT

Measures of shape

Lecturer: Dr. Emily Roche

INTRODUCTION

This lecture will focus on measures of skewness and kurtosis.

Intended learning outcomes

At the end of this lecture, you will be able to distinguish between symmetrical and asymmetrical distributions, compute various coefficients to measure lack of symmetry and kurtosis and distinguish between platykurtic, mesokurtic and leptokurtic distributions.

References

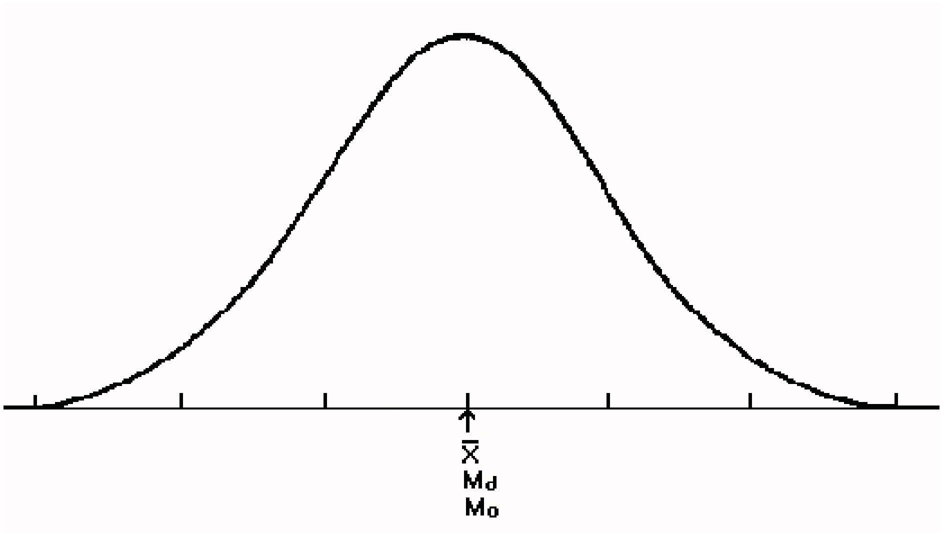
These lecture notes should be supplemented with relevant topics from the book listed in the Bibliography at the end of the lecture and the lecture video recording.

Introduction

Measures of central tendency and dispersion may fail to describe a distribution completely. It is possible to have frequency distributions which differ widely in their nature and composition and yet may have same measures of central tendency and dispersion.
This therefore brings the need to supplement these measures, thus, measures of skewness and kurtosis.

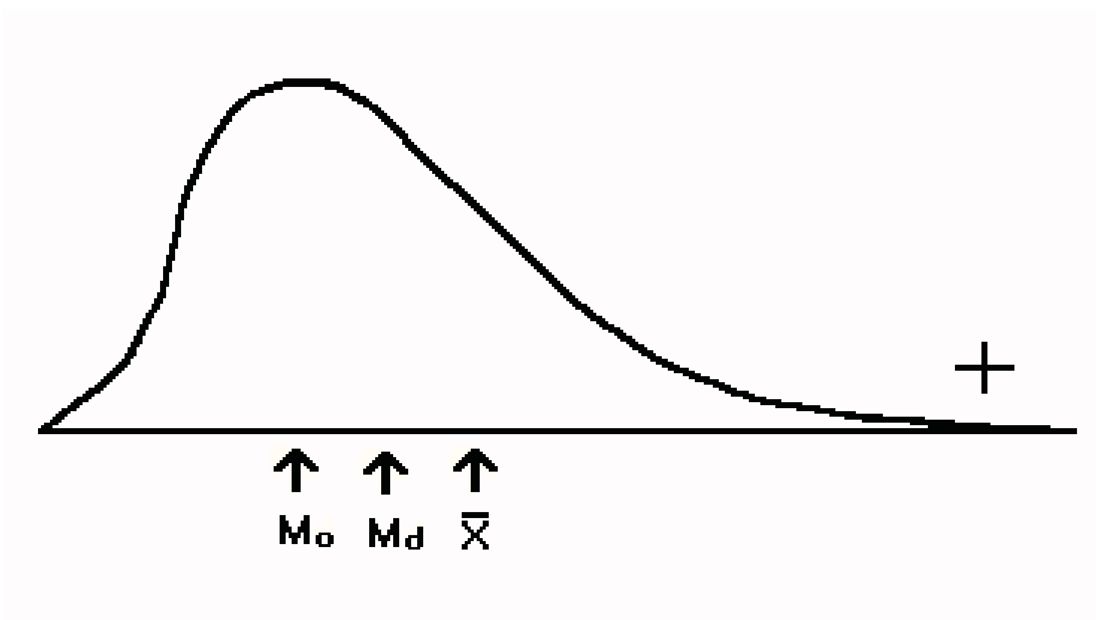
Skewness

This is defined as lack of symmetry. In a symmetrical distribution, the mean, median and mode coincide and the ordinate at mean divides the distribution into two equal parts such that one part is a mirror image of the other. If some observations, of very high or low magnitude, are added to such a distribution, its right or left tail gets elongated.



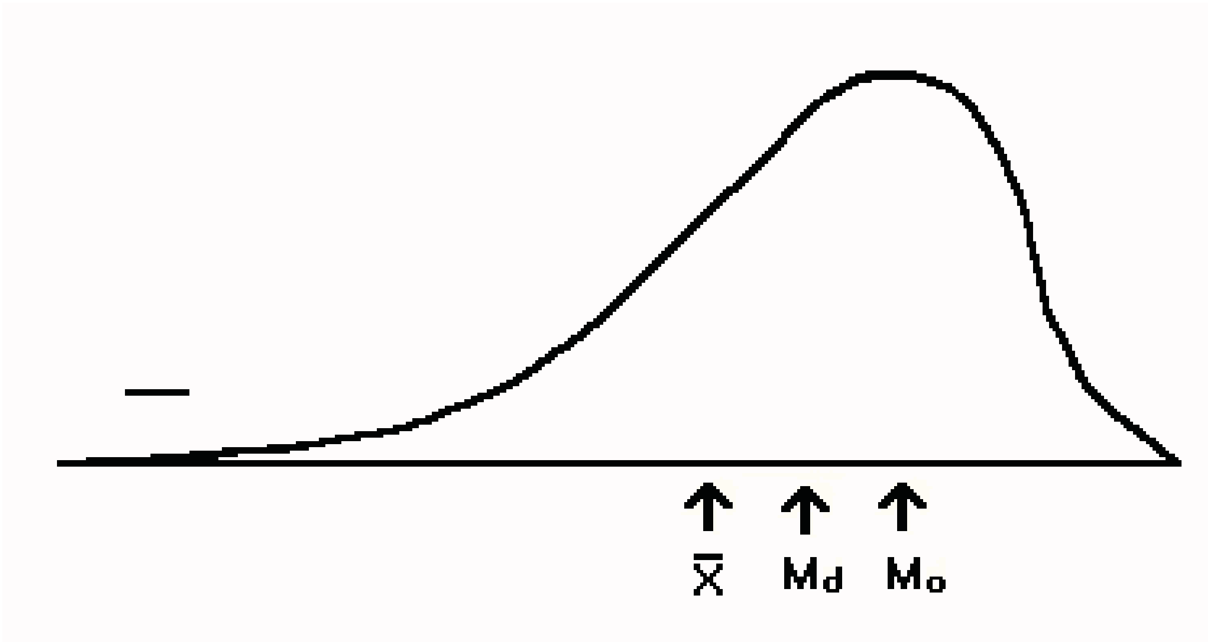
Symmetrical distribution $\bar{x} = M_o = M_d$

These very high or low observations are also known as extreme observations. The presence of extreme observations on the right hand side of a distribution makes it positively skewed or skewed to the right and the three averages, mean, median and mode will no longer be equal.



Positively skewed or skewed to the right $\bar{x} > M_o > M_d$

The presence of extreme observations to the left hand side of a distribution on the other hand makes it negatively skewed or skewed to the left.



Negatively skewed or skewed to the left $\bar{x} < M_o < M_d$

Some of the methods of measuring direction and extent of skewness are:

Karl Pearson's Coefficient of Skewness

This measure is based upon the deviation of mean from mode in a skewed distribution. Since Mean = Mode in a symmetrical distribution, (Mean - Mode) can be taken as an absolute measure of skewness. The absolute measure of skewness for a distribution depends upon the unit of measurement. A relative measure is independent of the units of measurement.

Karl Pearson takes a relative measure of skewness defined as:

Karl Pearson's Coefficient of skewness $(S_k) = \frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}$

The sign of S_k gives the direction ($S_k > 0$ – positively skewed; $S_k < 0$ – negatively skewed) and the magnitude gives the extent of skewness.

It is known that mode is not a unique measure of central tendency. Therefore, if mode is not defined for a distribution the empirical relation between mean, median and mode which states that for a moderately symmetrical distribution Mean − Mode = 3 (Mean − Median) is used to estimate S_k.

Hence Karl Pearson's coefficient of skewness is defined in terms of median as:

$$S_k = \frac{3\,(\text{Mean} - \text{Median})}{\text{Standard deviation}}$$

Example:

Compute the Karl Pearson's coefficient of skewness from the following data:

Height (inches)	58	59	60	61	62	63	64	65
Number of persons	10	18	30	42	35	28	16	8

$Height\ (inches) - x$	$Number\ of\ persons\ (f)$	fx	fx^2
58	10	580	33640
59	18	1062	62658
60	30	1800	108000
61	42	2562	156282
62	35	2170	134540
63	28	1764	111132
64	16	1024	65536
65	8	520	33800
	$\sum f = 187$	$\sum fx = 11482$	$\sum fx^2 = 705588$

$$Mean = \frac{11482}{187} = 61.4; \qquad Mode = 61$$

$$Standard\ deviation = \sqrt{\frac{705588}{187} - \left(\frac{11482}{187}\right)^2} = 1.76$$

$$S_k = \frac{61.4 - 61}{1.76} = 0.227$$

Thus the distribution is positively skewed

Bowley's Measure of Skewness (Quartile coefficient of skewness)

This measure is based on quartiles with the assumption that for a symmetrical distribution, Q_1 and Q_3 are equidistant from the median (M_d).

A relative measure of skewness, known as Bowley's coefficient (SQ), is given by

$$S_Q = \frac{(Q_3 - M_d) - (M_d - Q_1)}{(Q_3 - M_d) + (M_d - Q_1)} = \frac{Q_3 + Q_1 - 2M_d}{Q_3 - Q_1}$$

From the previous example:

Height (inches)	58	59	60	61	62	63	64	65
Number of persons	10	18	30	42	35	28	16	8
Cumulative frequency	10	28	58	100	135	163	179	187

Median is at the 94^{th} position $\therefore$ median $= 61$

Lower quartile is at the 47^{th} position $\therefore Q_1 = 60$

Upper quartile is at the 140^{th} position $\therefore Q_3 = 63$

$$S_Q = \frac{63 + 60 - 2(61)}{63 - 60} = 0.33$$

The distribution is positively skewed.

Kelly's Measure of Skewness (Percentile coefficient of skewness)

Bowley's measure of skewness is based on the middle 50% of the observations because it leaves 25% of the observations on each extreme of the distribution.

As an improvement over Bowley's measure, Kelly suggested a measure based on P_{10} and P_{90} such that only 10% of the observations on each extreme are ignored. Kelly's coefficient of skewness, denoted by S_P, is given by:

$$S_Q = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{50}) + (P_{50} - P_{10})} = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$$

For the data on heights in the previous example,

$$S_Q = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}} = \frac{64 + 59 - 2(61)}{64 - 59} = 0.2$$

Note:

The magnitude of the three coefficients of skewness may not be comparable, but in the absence of skewness, all of them will be equal to zero. The three coefficients will however be consistent in direction for a given data set.

MOMENTS

The r^{th} moment about the mean of a distribution, denoted by μ_r is the mean of the r^{th} power of deviations from the arithmetic mean. It is given by:

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - \bar{X})^r$$

where $r = 0, 1, 2, 3, \ldots$

In particular,

If $r = 0$ we have

$$\mu_0 = \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - \bar{X})^0 = 1$$

For $r = 1$, we have

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - \bar{X})^1 = 0$$

For $r = 2$, we have

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - \bar{X})^2 = \sigma = \text{Variance}$$

For $r = 3$, we have

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - \bar{X})^3$$

For $r = 4$, we have

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{n} f_i(X_i - \bar{X})^4 \ldots$$

These moments are known as central moments.

Moment coefficient of skewness

The moment measure of skewness is based on the assumption that, for a symmetrical distribution, all odd ordered central moments are equal to zero. This is because in a symmetrical distribution, the mean is the balance point so there will be a deviation below the mean which is exactly equal to each deviation above the mean.

Since $\mu_1 = 0$ for every distribution, the lowest odd ordered moment that can provide an absolute measure of skewness is μ_3. A relative coefficient of skewness is given by:

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

Example

Compute the Moment coefficient of skewness (P_k) from the following data.

Marks obtained	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$	$60-70$
Frequency	6	12	22	24	16	12	8

Solution

$$Variance = \frac{\sum f\,(x-\bar{x})^2}{\sum f} = \frac{26000}{100} = 260; \quad Standard\ deviation = \sqrt{260} = 16.12$$

$$Third\ moment\ (\mu_3) = \frac{\sum f\,(x-\bar{x})^3}{\sum f} = \frac{48000}{100} = 480$$

$$Moment\ coefficient\ of\ skewness\ (P_k) = \frac{\mu_3}{\sigma^3} = \frac{480}{\left(\sqrt{260}\right)^3} = 0.11$$

Kurtosis

This is also a measure of the shape of a distribution. Whereas skewness measures the lack of symmetry of the frequency curve of a distribution, kurtosis is a measure of the relative peakedness of its frequency curve. Various frequency curves can be divided into three categories Leptokurtic, Mesokurtic and Platykurtic depending upon the shape of their peak.

A measure of kurtosis is given by **moment coefficient of kurtosis** which is given by the fourth moment divided by the standard deviation raised to power 4.

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

The value of the moment coefficient of kurtosis is 3 for a

mesokurtic curve which is considered as normal. When this coefficient is more than 3, the curve is more peaked than the mesokurtic curve and is referred to as a leptokurtic curve. Similarly, when the coefficient is less than 3, the curve is less peaked than the mesokurtic curve and is termed as a platykurtic curve.

Example

The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75. Examine the skewness and kurtosis of the distribution.

Moment coefficient of skewness is given as

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{0.7}{\left(\sqrt{2.5}\right)^3} = 0.177$$

This implies that the distribution is positively skewed.

Moment coefficient of kurtosis is

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{18.75}{\left(\sqrt{2.5}\right)^4} = 3$$

The curve is mesokurtic.

Bibliography

Gupta, SP (Dr.), (2014). *Statistical methods* (43rd Ed.). Sultan Chand & Sons.

S. C. Gupta and V. K. Kapoor, (2020). *Fundamentals of mathematical Statistics* (12th Ed). Sultan Chand & Sons.