Applied Statistics I

(IMT224β/AMT224β)

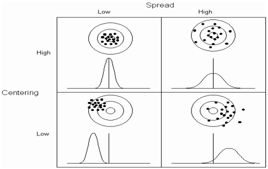
Department of Mathematics
University of Ruhuna

A.W.L. Pubudu Thilan

Measures of skewness

Introduction

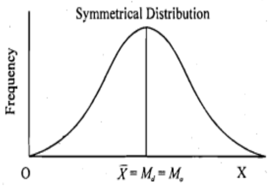
- As mentioned in previous two chapters, we can characterize any set of data by measuring its central tendency, variation, and shape.
- In this chapter we are going to discuss about shape of a given data set.



- The need to study these concepts arises from the fact that the measures of central tendency and dispersion 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 central tendency and dispersion.
- Thus, there is need to supplement the measures of central tendency and dispersion.

Concept of skewness

- The skewness of a distribution is defined as the lack of symmetry.
- In a symmetrical distribution, mean, median, and mode are equal to each other.

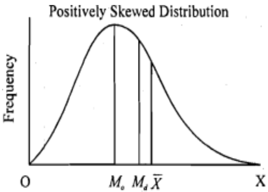


Concept of skewness
Cont...

- The presence of extreme observations on the right hand side of a distribution makes it positively skewed.
- We shall in fact have

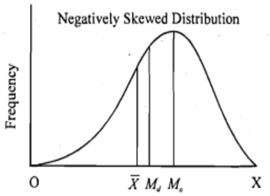
$$\text{Mean} > \text{Median} > \text{Mode}$$

when a distribution is positively skewed.



- On the other hand, the presence of extreme observations to the left hand side of a distribution make it negatively skewed and the relationship between mean, median and mode is:

$$\text{Mean} < \text{Median} < \text{Mode}.$$



Different measures of skewness

The direction and extent of skewness can be measured in various ways. We shall discuss three measures of skewness in this section.

1. Pearson's coefficient of skewness
2. Bowley's coefficient of skewness
3. Kelly's coefficient of skewness

[1] Karl Pearson coefficient of skewness

- The mean, median and mode are not equal in a skewed distribution.
- The Karl Pearson's measure of skewness is based upon the divergence of mean from mode in a skewed distribution.

$$S_{kp_1} = \frac{\text{mean - mode}}{\text{standard deviation}} \quad \text{or} \quad S_{kp_2} = \frac{3(\text{mean - median})}{\text{standard deviation}}$$

Properties of Karl Pearson coefficient of skewness

1. $-1 \leq S_{kp} \leq 1$.
2. $S_{kp} = 0 \Rightarrow$ distribution is symmetrical about mean.
3. $S_{kp} > 0 \Rightarrow$ distribution is skewed to the right.
4. $S_{kp} < 0 \Rightarrow$ distribution is skewed to the left.

Advantage and disadvantage

Advantage

S_{kp} is independent of the scale. Because (mean-mode) and standard deviation have same scale and it will be canceled out when taking the ratio.

Disadvantage

S_{kp} depends on the extreme values.

Example

Calculate Karl Pearson coefficient of skewness of the following data set ($S = 1.7$).

Value (x)	1	2	3	4	5	6	7
Frequency (f)	2	3	4	4	6	4	2

Example
Solution

$$
\begin{aligned}
\text{mean} = \overline{x} &= \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i}^{n} f_i} \\
&= \frac{1 \times 2 + 2 \times 3 + \ldots + 7 \times 2}{25} \\
&= \frac{104}{25} \\
&= 4.16 \\
\text{mode} &= 5 \\
S_{kp} &= \frac{\text{mean-mode}}{\text{standard deviation}} \\
&= \frac{4.16 - 5}{1.7} = -0.4941
\end{aligned}
$$

Since $S_{kp} < 0$ distribution is skewed left.

[2] Bowley's coefficient of skewness

- This measure is based on quartiles.
- For a symmetrical distribution, it is seen that Q_1, and Q_3 are equidistant from median (Q_2).
- Thus $(Q_3 - Q_2) - (Q_2 - Q_1)$ can be taken as an absolute measure of skewness.

$$\begin{aligned} S_{kq} &= \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} \\ &= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} \end{aligned}$$

Properties of Bowley's coefficient of skewness

1. $-1 \leq S_{kq} \leq 1$.
2. $S_{kq} = 0 \Rightarrow$ distribution is symmetrical about mean.
3. $S_{kq} > 0 \Rightarrow$ distribution is skewed to the right.
4. $S_{kq} < 0 \Rightarrow$ distribution is skewed to the left.

Advantage and disadvantage

Advantage

S_{kq} does not depend on extreme values.

Disadvantage

S_{kq} does not utilize the data fully.

Example

The following table shows the distribution of 128 families according to the number of children.

No of children	No of families
0	20
1	15
2	25
3	30
4	18
5	10
6	6
7	3
8 or more	1

Find the Bowley's coefficient of skewness.

Example
Solution

No of children	No of families	Cumulative frequency
0	20	20
1	15	35
2	25	60
3	30	90
4	18	108
5	10	118
6	6	124
7	3	127
8 or more	1	128

Example

Solution⇒Cont...

$$
\begin{aligned}
Q_1 &= \left(\frac{128+1}{4}\right)^{th} \text{ observation} \\
&= (32.25)^{th} \text{ observation} \\
&= 1 \\
Q_2 &= \left(\frac{128+1}{2}\right)^{th} \text{ observation} \\
&= (64.5)^{th} \text{ observation} \\
&= 3 \\
Q_3 &= 3\left(\frac{128+1}{4}\right)^{th} \text{ observation} \\
&= (96.75)^{th} \text{ observation} \\
&= 4
\end{aligned}
$$

Example
Solution⇒Cont...

$$
\begin{aligned}
S_{kq} &= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} \\
&= \frac{4 + 1 - 2 \times 3}{4 - 1} \\
&= -\frac{1}{3} \\
&= -0.333
\end{aligned}
$$

Since $S_{kq} < 0$ distribution is skewed left.

[3] Kelly's 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 has suggested a measure based on P_{10} and, P_{90} so that only 10% of the observations on each extreme are ignored.

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

Thank You