Department of Mathematics
University of Biskra

Descriptive Statistics
Skewness and Kurtosis

1. Introduction

In descriptive statistics, after studying measures of central tendency such as the mean, median, and mode, and measures of dispersion such as the range, variance, and standard deviation, we also study the **shape** of a distribution.

Two important measures of shape are:

- [leftmargin=1.5cm]
- **Skewness**, which describes the asymmetry of a distribution.
- **Kurtosis**, which describes the degree of flatness or peakedness of a distribution.

These two parameters help us understand whether the data are symmetric, whether one tail is longer than the other, and whether the distribution is more peaked or flatter than the normal distribution.

2. Skewness

2.1 Definition

Skewness is a measure of the lack of symmetry of a distribution.

A distribution is said to be:

- [leftmargin=1.5cm]
- **symmetric** if the left and right sides are approximately mirror images,
- **positively skewed** if the right tail is longer,
- **negatively skewed** if the left tail is longer.

2.2 Graphical representation of skewness

a) Symmetric distribution

A symmetric distribution has equal shape on both sides of the center.

$$\text{Skewness} = 0$$

In this case:

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

```
           /\
          /  \
         /    \
        /      \
-------/--------\-------
```

b) Positively skewed distribution

A positively skewed distribution has a longer tail on the right side.

$$\text{Skewness} > 0$$

Usually:

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

```
        /\
       /  \
      /    \
-----/      \-----------
                  ------
```

This occurs when most values are small, but a few very large values pull the mean to the right.

c) Negatively skewed distribution

A negatively skewed distribution has a longer tail on the left side.

$$\text{Skewness} < 0$$

Usually:

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

```
           /\
          /  \
         /    \
----------     \-----
------          \----
```

This occurs when most values are high, but a few small values pull the mean to the left.

2.3 Measures of skewness

A common empirical measure of skewness is **Pearson's coefficient of skewness**.

First Pearson coefficient

$$Sk_1 = \frac{\bar{x} - Mo}{s}$$

where:

- [leftmargin=1.5cm]
- $\bar{x}$ = mean,
- Mo = mode,
- s = standard deviation.

Second Pearson coefficient

When the mode is not easy to determine, we often use:

$$Sk_2 = \frac{3(\bar{x} - Me)}{s}$$

where Me is the median.

Interpretation

- [leftmargin=1.5cm]
- $Sk = 0$: symmetric distribution
- $Sk > 0$: positive skewness
- $Sk < 0$: negative skewness

The greater the absolute value of skewness, the stronger the asymmetry.

3. Kurtosis

3.1 Definition

Kurtosis is a measure of the degree of peakedness or flatness of a distribution relative to the normal distribution.

It indicates whether the distribution is:

- [leftmargin=1.5cm]
- more peaked,
- normal in shape,
- or flatter than the normal distribution.

3.2 Graphical representation of kurtosis

a) Mesokurtic distribution

A mesokurtic distribution has the same kurtosis as the normal distribution.

```
           /\
          /  \
         /    \
--------/------\--------
```

b) Leptokurtic distribution

A leptokurtic distribution is more peaked than the normal distribution.

```
            /\
           /  \
          /    \
         /      \
--------/--------\--------
```

c) Platykurtic distribution

A platykurtic distribution is flatter than the normal distribution.

```
         --------
       /          \
------/            \------
```

3.3 Measure of kurtosis

The coefficient of kurtosis is based on the fourth central moment:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

where:

- [leftmargin=1.5cm]
- μ_4 = fourth central moment,
- μ_2 = second central moment, that is, the variance.

Another common measure is the **excess kurtosis**:

$$\gamma_2 = \beta_2 - 3$$

3.4 Interpretation of kurtosis

For a normal distribution:

$$\beta_2 = 3 \qquad \text{and} \qquad \gamma_2 = 0$$

Thus:

- [leftmargin=1.5cm]
- $\gamma_2 = 0$: mesokurtic
- $\gamma_2 > 0$: leptokurtic
- $\gamma_2 < 0$: platykurtic

4. Central Moments

The central moment of order r is defined by:

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^r$$

where:

- [leftmargin=1.5cm]
- N = total number of observations,
- x_i = observed values,
- $\bar{x}$ = arithmetic mean.

Important moments:

- [leftmargin=1.5cm]
- μ_2: used in variance,
- μ_3: used in skewness,
- μ_4: used in kurtosis.

The Fisher coefficient of skewness is:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

The Fisher coefficient of kurtosis is:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$$

5. Worked Example

Consider the following data set:

$$2,\ 3,\ 4,\ 4,\ 5,\ 6,\ 12$$

Step 1: Mean

$$\bar{x} = \frac{2+3+4+4+5+6+12}{7} = \frac{36}{7} \approx 5.14$$

Step 2: Median

Since there are 7 values, the median is the middle one:

$$Me = 4$$

Step 3: Mode

The most frequent value is 4, so:

$$Mo = 4$$

Step 4: Interpretation

We observe that:

$$Mo = 4, \qquad Me = 4, \qquad \bar{x} \approx 5.14$$

Thus:

$$Mo \leq Me < \bar{x}$$

The mean is greater than the median because the large value 12 pulls the distribution to the right.

So the distribution is **positively skewed**.

```
Values:   2   3   4   4   5   6          12

Most data are near 4 or 5, but one large value (12)
creates a long right tail.
```

Therefore, the data show **right asymmetry**.

6. Another Simple Example of Kurtosis

Suppose we compare two classes of students:

- [leftmargin=1.5cm]
- In the first class, most marks are very close to the average.
- In the second class, marks are more spread out.

The first distribution will be more peaked, so it is **leptokurtic**. The second distribution will be flatter, so it is **platykurtic**.

7. Summary Table

Measure	**Meaning**	**Interpretation**
Skewness	Measures asymmetry	$Sk = 0$: symmetric, $Sk > 0$: right-skewed, $Sk < 0$: left-skewed
Kurtosis	Measures flatness or peakedness	$\gamma_2 = 0$: mesokurtic, $\gamma_2 > 0$: leptokurtic, $\gamma_2 < 0$: platykurtic

8. Conclusion

Skewness and kurtosis are important descriptive measures because they provide information about the shape of a distribution.

- [leftmargin=1.5cm]
- **Skewness** tells us whether the data are symmetric or not.
- **Kurtosis** tells us whether the distribution is more peaked or flatter than the normal distribution.

Together with measures of position and dispersion, they allow a more complete description of statistical data.

9. Short Version to Memorize

Skewness measures the asymmetry of a distribution.

- [leftmargin=1.5cm]
- $Sk = 0$: symmetric
- $Sk > 0$: positively skewed
- $Sk < 0$: negatively skewed

Kurtosis measures the degree of flatness or peakedness.

- [leftmargin=1.5cm]
- $\gamma_2 = 0$: mesokurtic
- $\gamma_2 > 0$: leptokurtic
- $\gamma_2 < 0$: platykurtic