DESCRIPTIVE STATISTICS

Dr Alina Gleska

Institute of Mathematics, PUT

April 21, 2018

Skewness is a measure of the asymmetry of the distribution of a real-valued variable about its mean. The skewness value can be positive or negative, or undefined.

The qualitative interpretation of the skew is complicated and unintuitive. Skew does not refer to the direction the curve appears to be leaning; in fact, the opposite is true. For a unimodal distribution, negative skew indicates that the tail on the left side of the distribution function is longer or fatter than the right side – it does not distinguish these two kinds of shape.

Conversely, positive skew indicates that the tail on the right side is longer or fatter than the left side. In cases where one tail is long but the other tail is fat, skewness does not obey a simple rule. For example, a zero value means that the tails on both sides of the mean balance out overall; this is the case for a symmetric distribution, but is also true for an asymmetric distribution where the asymmetries even out, such as one tail being long but thin, and the other being short but fat. Further, in multimodal distributions and discrete distributions, skewness is also difficult to interpret. Importantly, the skewness does not determine the relationship of mean and median.

Classification:

- positional
  - positional index of skewness W_{pos}
  - positional coefficient of skewness A_Q
- classical
  - central moment of the third order m_3
  - classical coefficient of skewness A_s
- mixed (classical-positional)
  - index of skewness W_{sk}
  - Pearson's coefficient of skewness A_p

THE POSITIONAL INDEX OF SKEWNESS W_{pos} is the sum of the differences between the quartiles and the median:

$$W_{pos} = (Q_3 - Me) - (Me - Q_1) = Q_1 + Q_3 - 2Me.$$

Properties:

- $W_{pos} = 0$, e.g. $Q_3 - Me = Me - Q_1$ - a symmetric distribution,
- $W_{pos} > 0$, e.g. $Q_3 - Me > Me - Q_1$ - a right-skewed distribution,
- $W_{pos} < 0$, e.g. $Q_3 - Me < Me - Q_1$ - a left-skewed distribution.

REMARKS:
1) the positional index of skewness assesses the directions of skewness only for 50% middle observations,
2) the index W_{pos} shows only the direction of asymmetry – for the strength of skewness we need the positional coefficient of skewness.

THE POSITIONAL COEFFICIENT OF SKEWNESS

$$A_Q = \frac{W_{pos}}{R_0} = \frac{Q_1 + Q_3 - 2Me}{Q_3 - Q_1}.$$

defines the direction and the strength of skewness for 50% middle observations (between Q_1 and Q_3).
Usually $A_Q \in [-1, 1]$.

The sign of the coefficient A_Q tells us about the direction of asymmetry, e.g.

- $A_Q = 0$ - a symmetric distribution,
- $A_Q > 0$ - a right-skewed distribution,
- $A_Q < 0$ - a left-skewed distribution,

and its absolute value $|A_Q|$ tells us about the strength of asymmetry:

- $0 - 0,2$ - a very weak asymmetry,
- $0,2 - 0,4$ - a weak asymmetry,
- $0,4 - 0,6$ - a moderate asymmetry,
- $0,6 - 0,8$ - a strong asymmetry,
- more than $0,8$ - a very strong asymmetry,

for 50% middle observations (between Q_1 and Q_3).

THE CENTRAL MOMENT OF THE THIRD ORDER m_3 is calculated by the formula:

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3.$$

It is a member of more general group so called central moments:

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k.$$

So the measure m_2 is the variance, m_3 - the central moment of the third order is the measure of asymmetry, and m_4 - the central moment of the fourth order is the measure of concentration.

THE CLASSICAL COEFFICIENT OF SKEWNESS:

$$A_s = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{s^3}.$$

Usually $A_s \in [-2, 2]$.
A_s is a relative measure of asymmetry so it can be use for comparing two or more distributions.

The sign of the coefficient A_s tells us about the direction of asymmetry, e.g.

- $A_s = 0$ - a symmetric distribution,
- $A_s > 0$ - a right-skewed distribution,
- $A_s < 0$ - a left-skewed distribution,

and its absolute value $|A_s|$ tells us about the strength of asymmetry:

- $0 - 0,4$ - a very weak asymmetry,
- $0,4 - 0,8$ - a weak asymmetry,
- $0,8 - 1,2$ - a moderate asymmetry,
- $1,2 - 1,6$ - a strong asymmetry,
- more than $1,6$ - a very strong asymmetry,

THE INDEX OF SKEWNESS

$$W_{sk} = \overline{x} - Mo.$$

This index defines the direction of asymmetry but it tells nothing about the strength of asymmetry. It is an absolute measure, it has got units so it cannot be used for comparing different distributions.

Properties:

- $W_{sk} = 0$ - a symmetric distribution,
- $W_{sk} > 0$ - a right-skewed distribution,
- $W_{sk} < 0$ - a left-skewed distribution,

THE PEARSON'S COEFFICIENT OF SKEWNESS

$$A_p = \frac{\overline{x} - Mo}{s}$$

assesses both the direction and the strength of asymmetry. Usually $A_p \in [-1, 1]$.

The sign of the coefficient A_p tells us about the direction of asymmetry, e.g.

- $A_p = 0$ - a symmetric distribution,
- $A_p > 0$ - a right-skewed distribution,
- $A_p < 0$ - a left-skewed distribution,

and its absolute value $|A_p|$ tells us about the strength of asymmetry:

- $0 - 0,2$ - a very weak asymmetry,
- $0,2 - 0,4$ - a weak asymmetry,
- $0,4 - 0,6$ - a moderate asymmetry,
- $0,6 - 0,8$ - a strong asymmetry,
- more than $0,8$ - a very strong asymmetry.

REMARKS:
1) A_p is a relative measure of asymmetry so it can be use for comparing two or more distributions;
2) in cases when it is impossible to calculate the arithmetic mean or the mode we cannot calculate the Pearson's coefficient of skewness.

MEASURES OF ASYMMETRY FOR GROUPED DATA

The central moment of the third order m_3 and the classical coefficient of skewness for the data grouped in the categories:

$$m_3 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^3 n_i, \quad A_s = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^3 n_i}{s^3}.$$

The central moment of the third order m_3 and the classical coefficient of skewness for the data grouped in the intervals:

$$m_3 = \frac{1}{n}\sum_{i=1}^{k}(x_i^0 - \overline{x})^3 n_i, \quad A_s = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{k}(x_i^0 - \overline{x})^3 n_i}{s^3}.$$