

Descriptive characteristics – deskriptívne (popisné) štatistiky, resp. charakteristiky

Definition:

Descriptive characteristics are numerical values that express a particular property of the observed statistical attribute. They are divided as follows:

- Measures of location – percentile, quartile
 - Measures of central tendency (centre) – mean, mode, median
 - Measures of variability (spread) – IQR, range, variance, standard deviation, coefficient of variation
 - Measures of skewness – coefficient of skewness
 - Measures of kurtosis (“peakedness”) – coefficient of kurtosis
-

Measures of location – miery polohy

Definition:

A measure of location for a collection of data values means the position of a typical data value along some sort of coordinate axis.

Percentile – percentil

Definition:

The percentile is a number that divides ordered data into hundredths. Percentiles may or may not be part of the data.

Quartile – kvartil

Definition:

The numbers that separate the data into quarters. Quartiles may or may not be part of the data.

Note: The median of the data is the second quartile and the 50th percentile. The first and third quartiles are the 25th and the 75th percentiles, respectively. The zero and fourth quartiles represent the minimum and maximum value in a data set, respectively.

Measures of central tendency (centre) – miery strednej hodnoty

Definition:

A measure of central tendency for a collection of data values is a number that is meant to cover the idea of centralness for the data set.

Mean - priemer

Definition:

The mean of a set of numerical values is the *average* of the set of values. Mean represents a *typical* value or the centre of a set of values.

Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra

Reference: JAISINGH, L.: Statistics for the Utterly Confused

Two general charts related to measures of skewness and kurtosis were taken from the lecture of prof. Zlata Sojková, CSc., Dept. of Statistics and Operations Research, FEM SUA in Nitra

Population mean – priemer základného súboru

Definition:

If the values are from an entire population, then the mean of the values is called a population mean. It is usually denoted by μ (read as "mju").

$$\mu = \frac{\sum x_i}{N} \text{ (simple) or } \mu = \frac{\sum x_i \cdot n_i}{N} \text{ (weighted)}$$

Sample mean – priemer výberového súboru

Definition:

If the values are from a sample, then the mean of the values is called a sample mean. It is denoted by \bar{x} (read as "x-bar").

$$\bar{x} = \frac{\sum x_i}{n} \text{ (simple) or } \bar{x} = \frac{\sum x_i \cdot n_i}{n} \text{ (weighted)}$$

Median - medián

Definition:

The median of a numerical data set is the numerical value *in the middle* when the data set is *arranged in order*.

When the number of values in the data set is *odd*, the median will be the *middle* value in the ordered array. When the number of values in the data set is *even*, the median will be the *average of the two middle values* in the ordered array.

Mode - modus

Definition:

The mode of a numerical data set is the *most frequently occurring value* in the data set.

Note: Unlike the median and the mode, the mean is sensitive to a change in any of the values of the data set (especially outliers, i.e. extreme values in the data set).

Measures of variability (spread) – miery variability

Definition:

A measure of variability for a collection of data values is a number that is meant to convey the idea of spread for the data set.

IQR (interquartile range) – medzikvartilové rozpätie

Definition:

The interquartile range covers the spread of middle 50% of data since it is a difference between the third quartile (Q3) and first quartile (Q1).

Range - rozpätie

Definition:

The range is the difference between the *largest* and *smallest* values in a data set. It is not a very useful measure of variability because it is *affected by outliers*, i.e. large or small values relative to the rest of the data set.

Population variance – rozptyl základného súboru

Definition:

If the values are from an entire population, then the variance called a population variance. It is usually denoted by σ^2 (read as "sigma").

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} \text{ (simple) or } \sigma^2 = \frac{\sum (x_i - \mu)^2 \cdot n_i}{N} \text{ (weighted)}$$

Sample variance – rozptyl výberového súboru

Definition:

The sample variance is an *approximate average* of the squared deviations from the sample mean. That is, the deviations of the data values from the sample mean are first computed, then the values for these deviations are squared, and the approximate average of these square values is found. The average is approximate because we divide not by the sample size but by the sample size minus 1. We do not interpret sample variance because of the *squared measurement units*.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \text{ (simple) or } s^2 = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n - 1} \text{ (weighted)}$$

Population standard deviation – smerodajná (štandardná) odchýlka základného súboru

Definition:

It is the square root of population variance.

$$\sigma = \sqrt{\sigma^2}$$

Note: The variance and the standard deviation are the most common and useful measures of variability. These two measures provide information about how the data vary about the mean.

Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra

Reference: JAISINGH, L.: Statistics for the Utterly Confused

Two general charts related to measures of skewness and kurtosis were taken from the lecture of prof. Zlata Sojková, CSc., Dept. of Statistics and Operations Research, FEM SUA in Nitra

Topic 2 Descriptive characteristics

If the data are clustered around the mean, then the variance and the standard deviation will be somewhat small. If, however, the data are widely scattered about the mean, the variance and the standard deviation will be somewhat large.

Sample standard deviation – smerodajná (štandardná) odchýlka výberového súboru

Definition:

The sample standard deviation is the square root of the sample variance. It is given by:

$$s = \sqrt{s^2}$$

Note: If all of the observations have the same value, the sample standard deviation will be zero. That is, there is no variability in the data set. The variance (standard deviation) is influenced by outliers (very small or very large values) in the data set. The unit for the standard deviation is the same as that for the raw data, so it is preferable to use the standard deviation instead of the variance as a measure of variability.

Sample coefficient of variation – variačný koeficient výberového súboru

Definition:

The sample coefficient of variation is defined as the sample standard deviation divided by the sample mean of the data set. Usually, the result is expressed as a percentage.

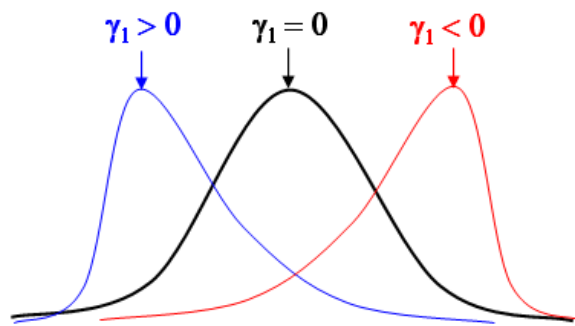
$$CV_s = \frac{s}{\bar{x}} \cdot 100[\%]$$

Note: We use this measure to compare variations for different variables with different units.

Measures of skewness – miery šikmosti (asymetrie)

Definition:

Skewness refers to the shape of a distribution. The coefficient of skewness can take on positive values (positively skewed data), negative values (negatively skewed data) or zero values (data not skewed, a symmetrical distribution).



$$\gamma_1 = \frac{\sum (x_i - \bar{x})^3}{s^3 \cdot n} \text{ (simple) or}$$
$$\gamma_1 = \frac{\sum (x_i - \bar{x})^3 \cdot n_i}{s^3 \cdot n} \text{ (weighted)}$$

Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra

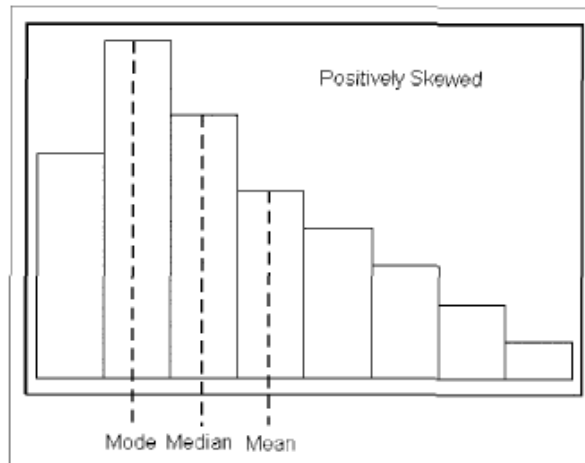
Reference: JAISINGH, L.: Statistics for the Utterly Confused

Two general charts related to measures of skewness and kurtosis were taken from the lecture of prof. Zlata Sojková, CSc., Dept. of Statistics and Operations Research, FEM SUA in Nitra

Positively skewed distribution (left/right-side asymmetry) – pozitívne zošikmené rozdelenie (údaje)

Definition:

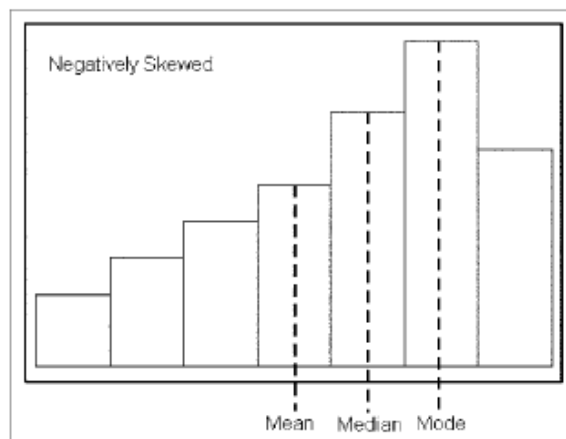
In a positively skewed distribution, most of the data values fall to the *left* of the mean, and the "tail" of the distribution is to the *right*. In addition, the mode is to the left of the median. The coefficient of skewness is *more than zero*.



Negatively skewed distribution (left/right-side asymmetry) - negatívne zošikmené rozdelenie (údaje)

Definition:

In a negatively skewed distribution, most of the data values fall to the *right* of the mean, and the tail of the distribution is to the *left*. In addition, the mode is to the right of the median. The coefficient of skewness is *less than zero*.



Symmetrical distribution – symetrické rozdelenie

Definition:

In a symmetrical (bell-shaped, normal) distribution, the data values are *evenly distributed on both sides of the mean*.

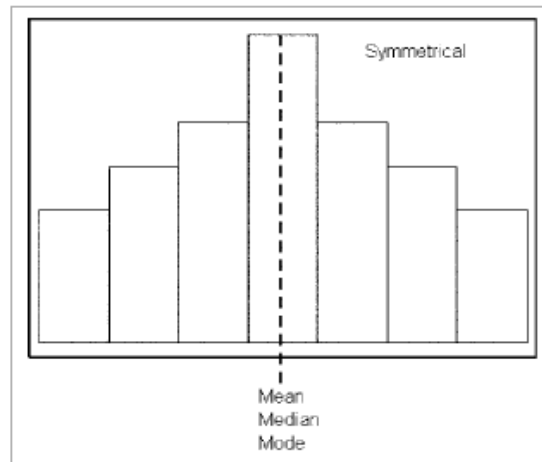
Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra

Reference: JAISINGH, L.: Statistics for the Utterly Confused

Two general charts related to measures of skewness and kurtosis were taken from the lecture of prof. Zlata Sojková, CSc., Dept. of Statistics and Operations Research, FEM SUA in Nitra

Topic 2 Descriptive characteristics

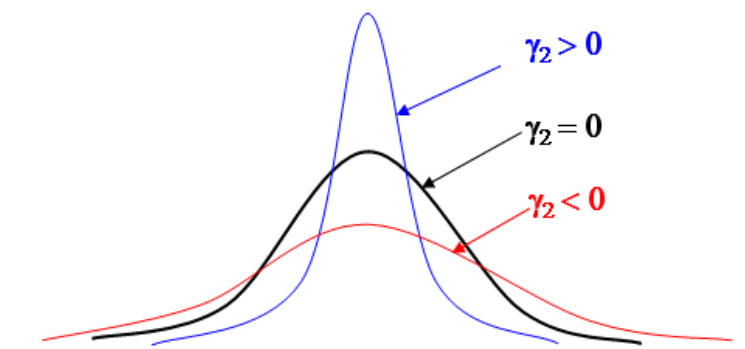
Also, when the distribution is unimodal, the mean, median, and mode are all equal to one another and are located at the center of the distribution.



Measures of kurtosis (“peakedness”) – miery špicatosti

Kurtosis refers to the shape of a distribution (similarly to skewness). The coefficient of kurtosis can take on positive values (peaked distribution), negative values (flat distribution) or zero values (a symmetrical, bell-shaped, normal distribution).

$$\gamma_2 = \frac{\sum (x_i - \bar{x})^4}{s^4 \cdot n} - 3 \text{ (simple) or } \gamma_2 = \frac{\sum (x_i - \bar{x})^4 \cdot n_i}{s^4 \cdot n} - 3 \text{ (weighted)}$$



Flat (platykurtic) distribution – rozdelenie je plochšie ako normálne

Note: Coefficient of kurtosis is less than zero.

Peaked (leptokurtic) distribution – rozdelenie je špicatejšie ako normálne

Note: Coefficient of kurtosis is more than zero.

Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra

Reference: JAISINGH, L.: Statistics for the Utterly Confused

Two general charts related to measures of skewness and kurtosis were taken from the lecture of prof. Zlata Sojková, CSc., Dept. of Statistics and Operations Research, FEM SUA in Nitra