AdTrans project , 2023-1-PL01-KA220-HED-000158917

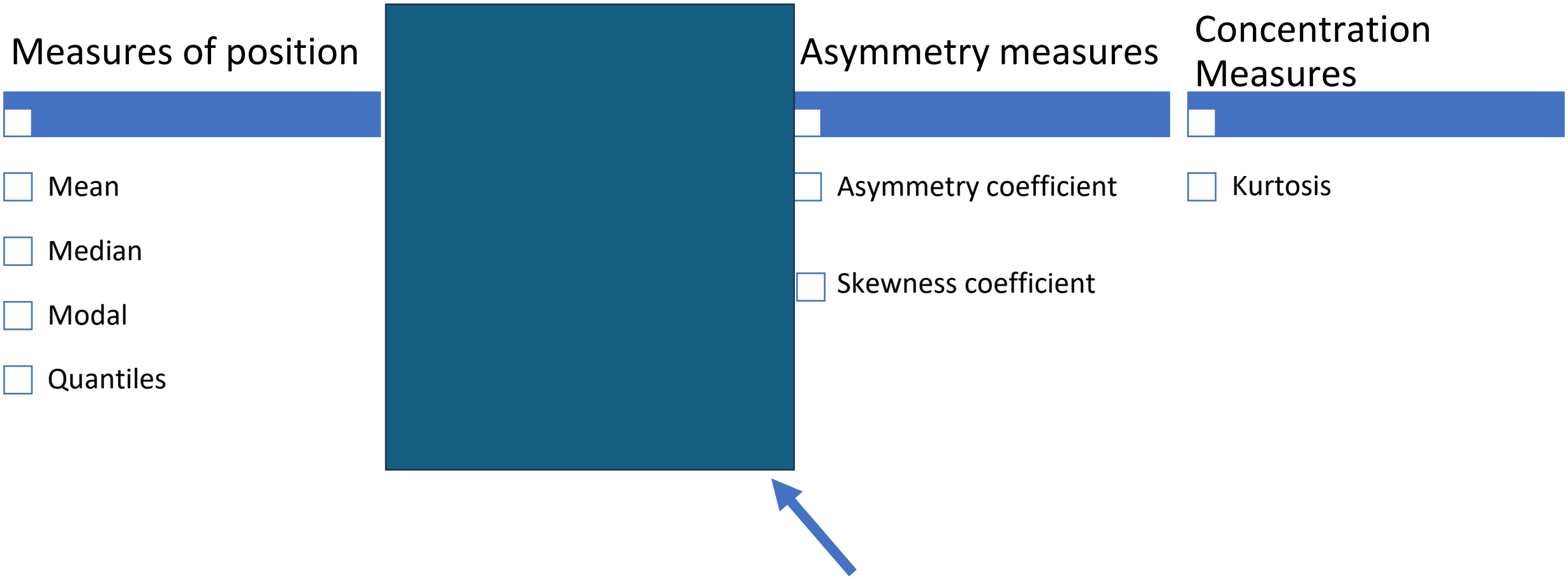
ADTRANS
ADVANCED DATA ANALYSIS FOR TRANSPORTATION AND MOBILITY ACADEMIC TEACHERS

Mathematical Statistics
Measures of Variability, Asymmetry, and Concentration

Anna Gobis

Co-funded by the European Union

Descriptive statistics



Classical measures of variability

Variance

Detailed series: $$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Distribution series: $$s^2 = \frac{1}{n}\sum_{i=1}^{n}(\dot{x}_i - \bar{x})^2 n_i$$

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

Classical measures of variability

Standard deviation:

The standard deviation tells us how widely the values of a given quantity are scattered around its mean.

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

Average deviation:

This measure indicates how much, on average, the results deviate from the mean value; this is the average deviation of the results.

- Detailed series $$d = \frac{1}{n}\sum^{n} |x_i - \bar{x}|$$
- Distribution series $$d = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|\, n_i$$

Classical measures of variability

- **Gap:**

$$R = x_{max} - x_{min}$$

- **Interquartile range = quartile range**

$$IQR = Q_3 - Q_1$$

- **Coefficient of variation:**

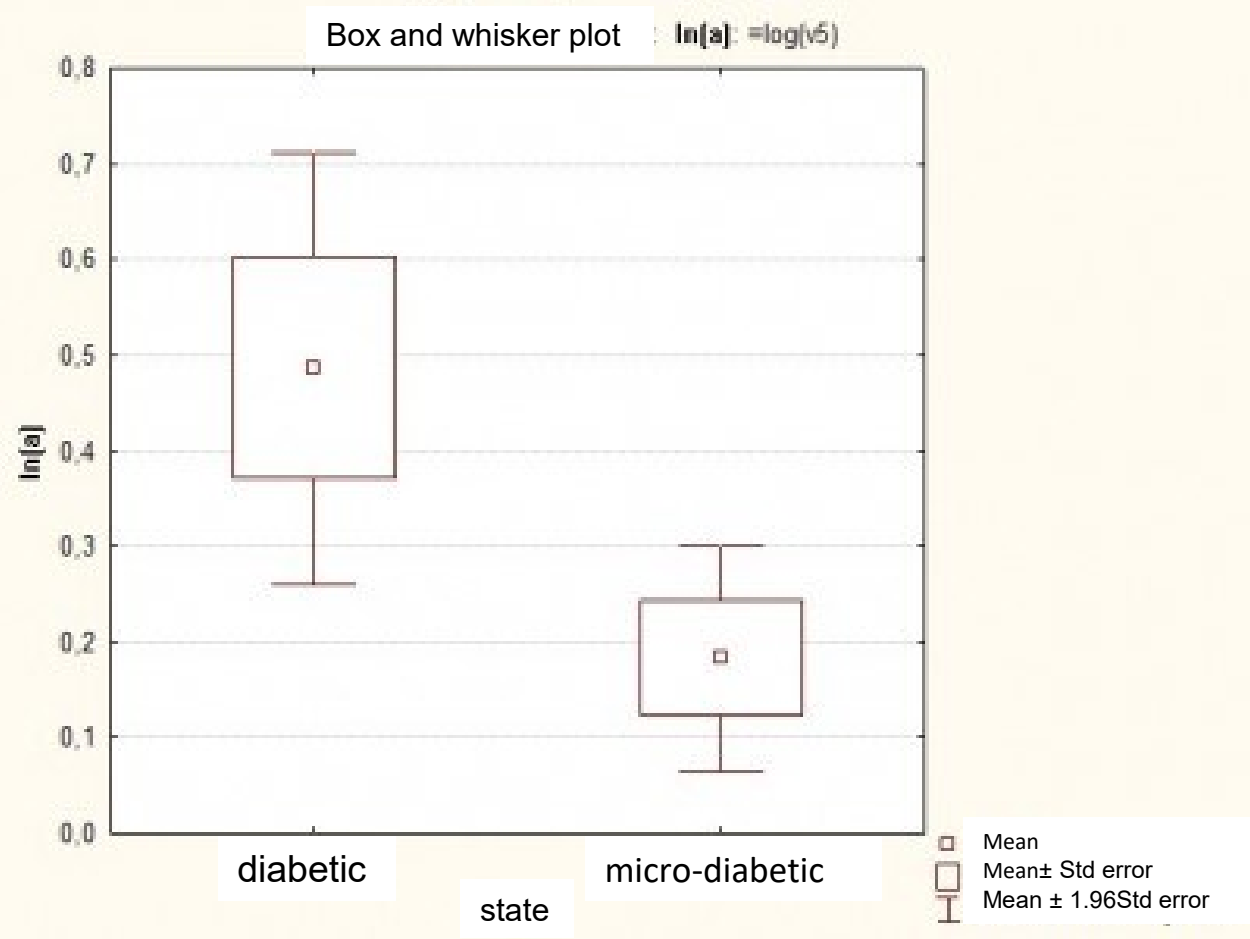
0-20% - weak differentiation of the trait
20-40% - moderate differentiation of the trait
40-60% - strong differentiation of the trait
60% and above - very strong differentiation of the trait

$$V = \frac{s}{\bar{x}} * 100\%$$

- **Typical area of variation**

$$\bar{x} - s \leq x_{typ} \leq \bar{x} + s$$

Box-and-whisker plot



source: statystycy.pl

Various types, e.g.:

- Midpoint – median
- Frame – quartile range (25-75)
- Whiskers - min-max range

Or

- Midpoint - Average
- Frame – standard deviation.
- Whiskers - min-max range

Example 3.1

10 men were randomly selected and asked about their height. The following results were obtained [cm]:

182	170	171	165	179	187	169	168	181	177

Calculate: arithmetic mean and standard deviation for the detailed series. Determine the area of variation of the feature. Present the distribution in the form of a box-and-whisker plot.

Example 3.2

Below are the results from the colloquium of a group of 30 students. The maximum number of points that could be obtained was 10 points.

3, 4, 5, 6, 6, 3, 9, 5, 5, 6, 7, 4, 10, 7, 3, 5, 8, 5, 4, 7, 9, 6, 10, 3, 7, 4, 4, 4, 8, 1

Calculate: arithmetic mean and standard deviation for a point distribution series. Determine the area of variation of the feature. Present the distribution as a box-and-whisker plot

Example 3.3 (data from Example 2.6)

We have data for 50 employees regarding the time it takes to perform a certain task [min]. Calculate the arithmetic mean, variance, standard deviation, range and determine the typical area of variation.

Time [min]	Number of employees
7-8	2
8-9	7
9-10	8
10-11	12
11-12	11
12-13	6
13-14	4

Example 3.4

As part of the assessment, two tests were conducted. In the first, students could obtain 0-40 points, in the second 0-100 points. The average result of the first test was 20 points, the standard deviation was 8 points. In the case of the second test, the average result was 50 points, the standard deviation was 12 points. Compare the coefficients of variation for test no. 1 and test no. 2.

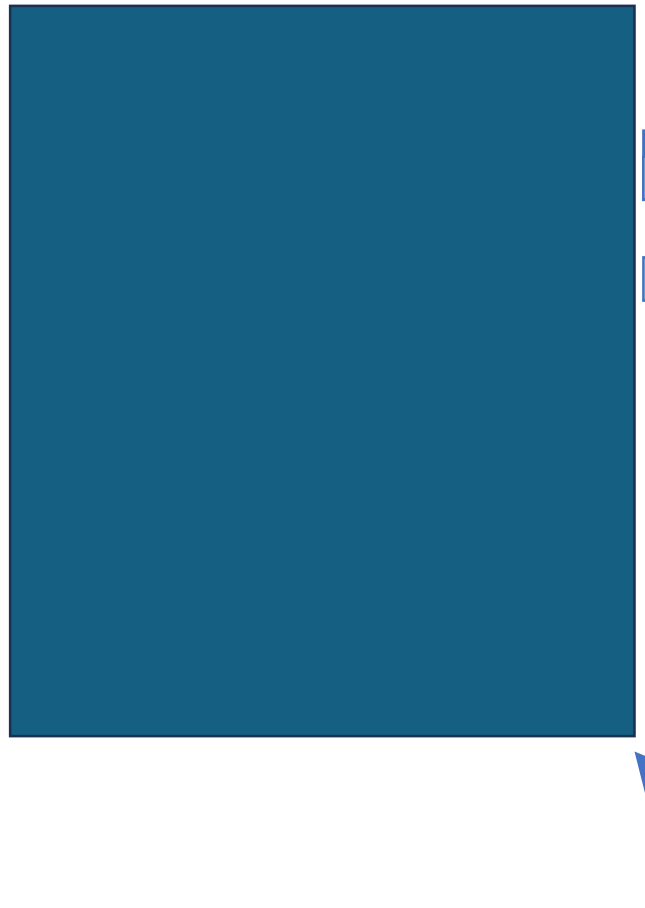
Descriptive statistics

Measures of position

- Mean
- Median
- Modal
- Quantiles

Measures of variability

- Variance
- Standard deviation
- Coefficient of variation



Concentration Measures

- Kurtosis

Measures of asymmetry (skewness)

Absolute

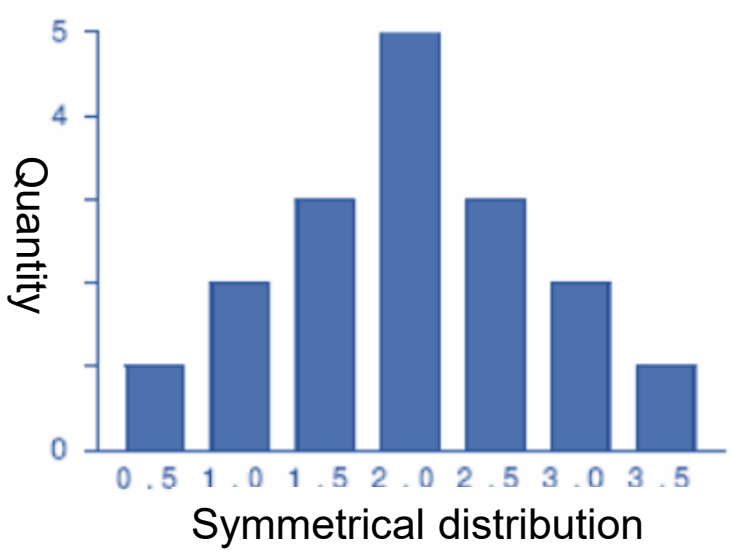
- Asymmetry Index $A = \bar{x} - Mo$
  - Symmetrical distribution $\bar{x} = Me = Mo; \quad A = 0$
  - Right-skewed (positively skewed) distribution $\bar{x} > Me > Mo; \quad A > 0$
  - Left-skewed (negatively skewed) distribution $< Me < Mo; \quad A < 0$

Relative

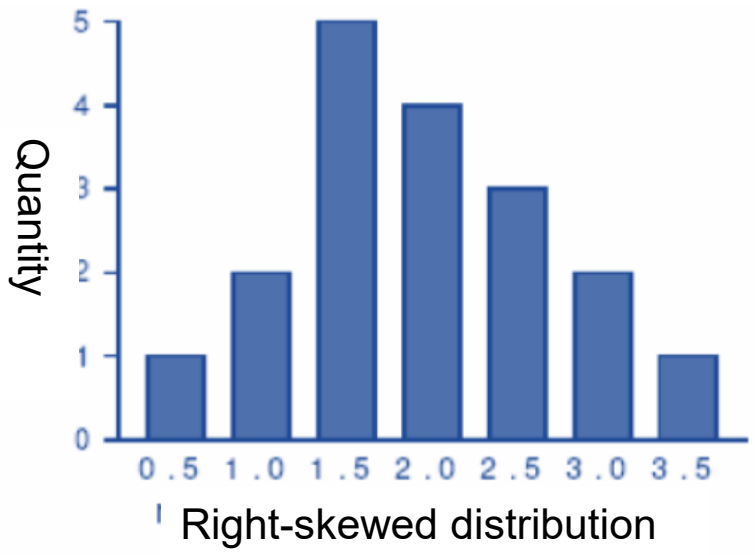
- Asymmetry Coefficient $A_s = \frac{\bar{x} - Mo}{s}$
  - Symmetrical distribution $A_s = 0$
  - Right-skewed (positively skewed) distribution $A_s > 0$
  - Left-skewed (negatively skewed) distribution $A_s < 0$

|Ace|<= 0.3 – weak
Between 0.3 and 0.7 – moderate
Between 0.7 and 0.9 strong
Close to 1 – extreme asymmetry

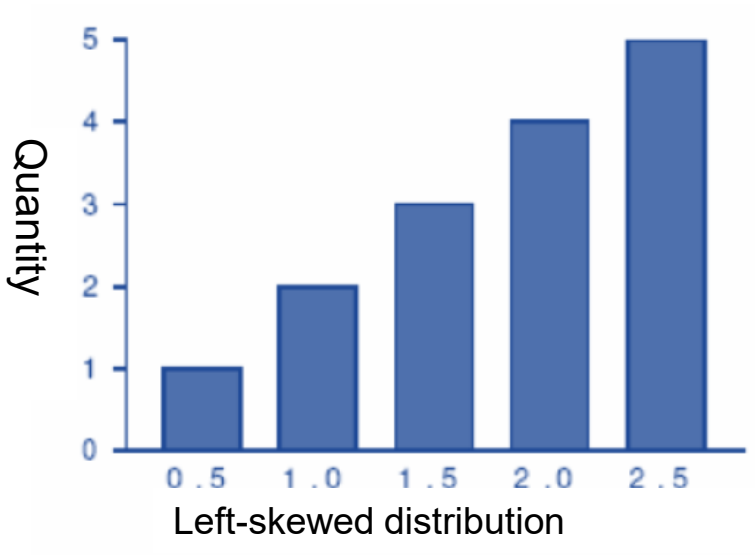
Skewness



Symmetrical distribution



Right-skewed distribution



Left-skewed distribution

Right-sided (also called a **positively skewed distribution**) Mode < Median < Mean

Left-sided (also called **negatively skewed** distribution) Mode > Median > Mean

Descriptive statistics

Measures of position

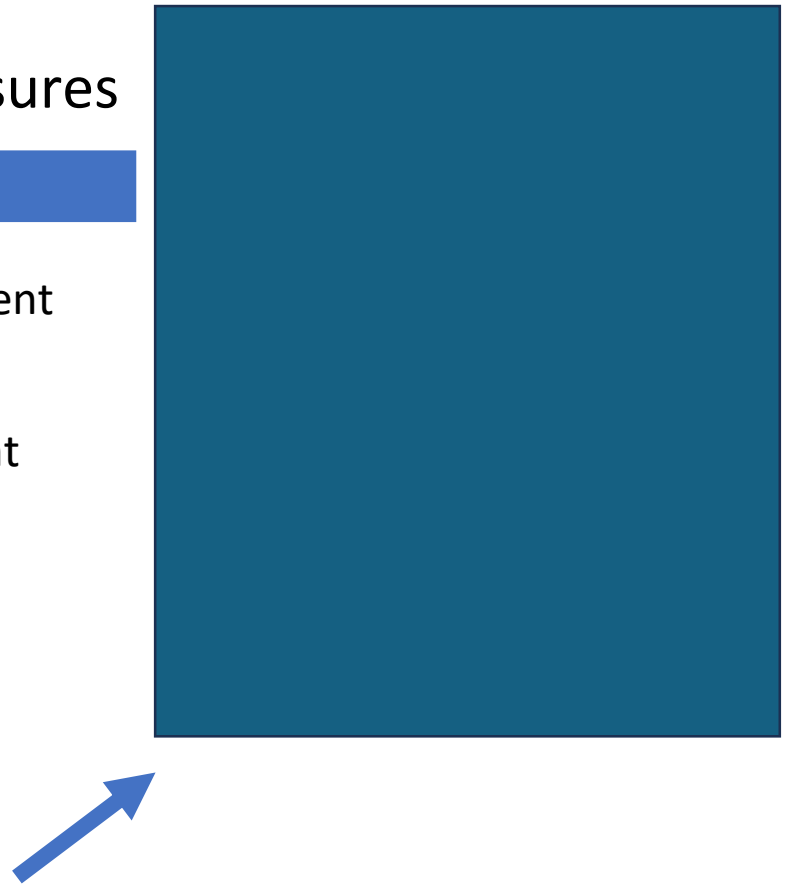
- Mean
- Median
- Modal
- Quantiles

Measures of variability

- Variance
- Standard deviation
- Coefficient of variation

Asymmetry measures

- Asymmetry coefficient
- Skewness coefficient

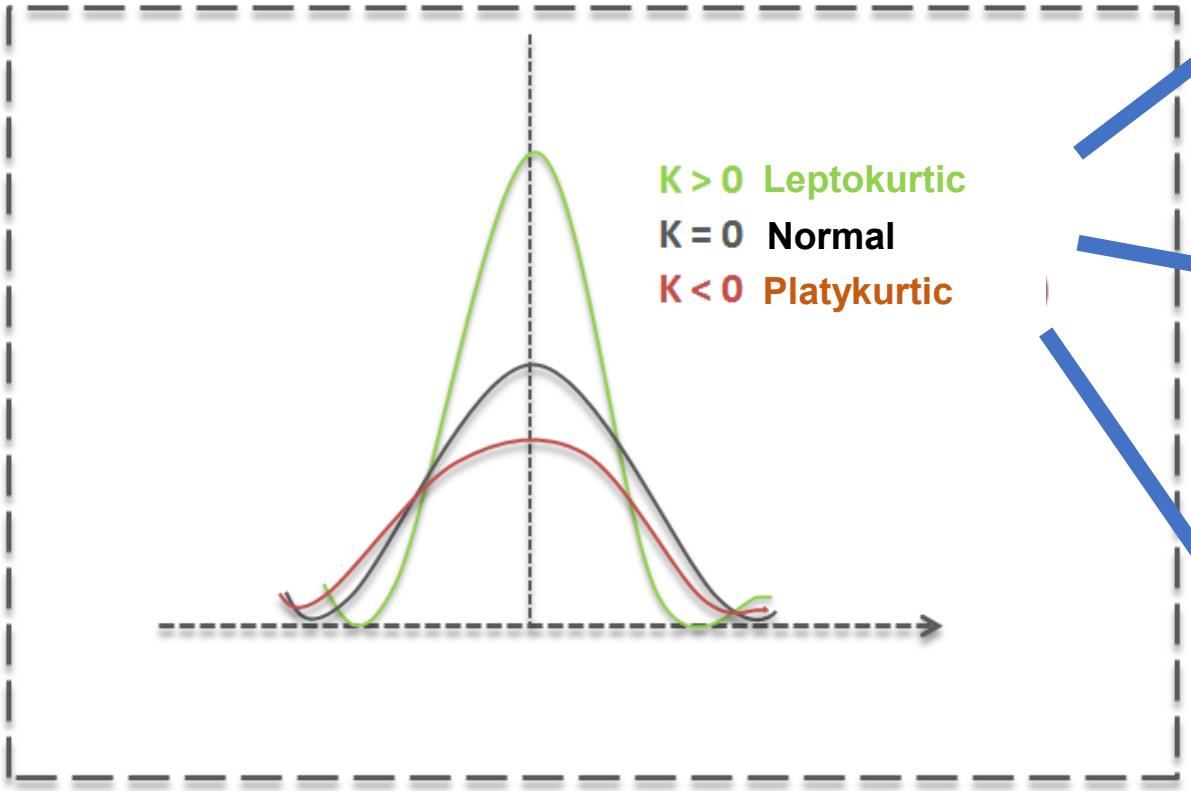


Concentration Measures

- **Kurtosis**
  - Detailed series $$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{s^4} - 3$$
  - Grouped series $$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4 * n_i}{s^4} - 3$$

- Interpretation
  - Leptokurtic distribution: $K > 0$
  - Normal distribution: $K = 0$
  - Platykurtic distribution: $K < 0$

Concentration Measures



the values of the feature we are interested in are concentrated around the mean and there is a greater chance of appearing from the tails

Example of a distribution – normal distribution, about 68% of the sample is concentrated around the mean

the values of the feature we are interested in are scattered around the mean and there is a lower probability of extreme values appearing

source:statisticaopysowa.com

Example 3.5

Calculate the asymmetry and concentration measures for the data from Examples 3.2 and 3.3

Example 3.6 - independent task

The amount of expenditure on bread in PLN, incurred in December 2009 by 200 randomly selected households from Konin, was as follows:

Expenses [PLN]	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Number of farms	12	25	37	62	34	22	8

Specify the population, unit, and statistical variable. Calculate measures of central tendency ($\bar{x}, Me, Mo, Q_1, Q_3$,(dispersion (s, s^2, V, R, R_Q ,(skewness, and concentration. Draw a histogram of frequencies and cumulative frequencies. Determine the variability range relative to the mean. Draw a box-and-whisker plot ($\bar{x}, \sigma, x_{min}, x_{max}$.(
Provide conclusions from the data analysis.

Sources:

- Kot, SM, Jakubowski, J., Sokołowski, A.: Statistics. Second revised edition. Difin Publishing House . Warsaw, 2011.
- Stanisz, A.: Accessible statistics course using STATISTICA PL on examples from medicine. Volume 1. Basic statistics. StatSoft Polska Publishing House. Cracow, 2006.
- Flisikowski , K.: Statistical methods in business. Materials of the AdvancedPhD project . Gdańsk University of Technology, 2015.
- Zimny, A.: Descriptive statistics. Supporting materials for exercises. Second revised edition. State Higher Vocational School in Konin, 2010.