



MULTIVARIATE STATISTICAL ANALYSIS TOOLS FOR BUSINESS DECISIONS

JÓZEF DZIECHCIARZ

WROCLAW UNIVERSITY OF ECONOMICS



**A presentation
a lectures
a questions**

A QUIZ

Table 1.1. Probabilities generated by Benford's Law.

Digit	Position in number			
	1st	2nd	3rd	4th
0	–	0,11968	0,10178	0,10018
1	0,30103	0,11389	0,10138	0,10014
2	0,17609	0,10882	0,10097	0,10010
3	0,12494	0,10433	0,10057	0,10006
4	0,09691	0,10031	0,10018	0,10002
5	0,07918	0,09668	0,09979	0,09998
6	0,06695	0,09337	0,09940	0,09994
7	0,05799	0,09035	0,09902	0,09990
8	0,05115	0,08757	0,09864	0,09986
9	0,04576	0,08500	0,09827	0,09982

Source: [Nigrini 1996, p. 74].

The sampling process

- There are two ways of unit selection:
 - Random selection, where only chance determines which unit from the general population will be drawn. A feature of this selection technique is the ability to apply results for the general population. Characteristics determined while examination of a random sample may be extended on whole population. The error size may be calculated.
 - Non-random selection. In this situation it is not possible to claim, the selected set of units to be representative for whole population. Statistical characteristics, as well as the possibility of carrying out statistical inference about the general population is limited.

The sampling process random; representative

- Determination of the study population
- Determination of the sampling frame (if any)
- Defining the selection (drawing) unit
- The choice of units selection method
- Determination of sample size

The sampling process random; representative

- Selection plan determines the probability of selecting an object to the sample
- Samples for which these probabilities are known, sampling is called random
 - Determination of the sample size (for random selection methods) depends on the size of the population, the variability of the characteristic and accepted level of accuracy
- In contrast to the so-called special purpose or quota samples, where those probabilities can not be determined

The sampling process random; representative

SAMPLING TECHNIQUES

INDIVIDUAL SAMPLING TECHNIQUES

UNLIMITED

LIMITED

With return

Without return

COLLECTIVE SAMPLING TECHNIQUES

ONE STAGE

MULTIPLE STAGE

Strata

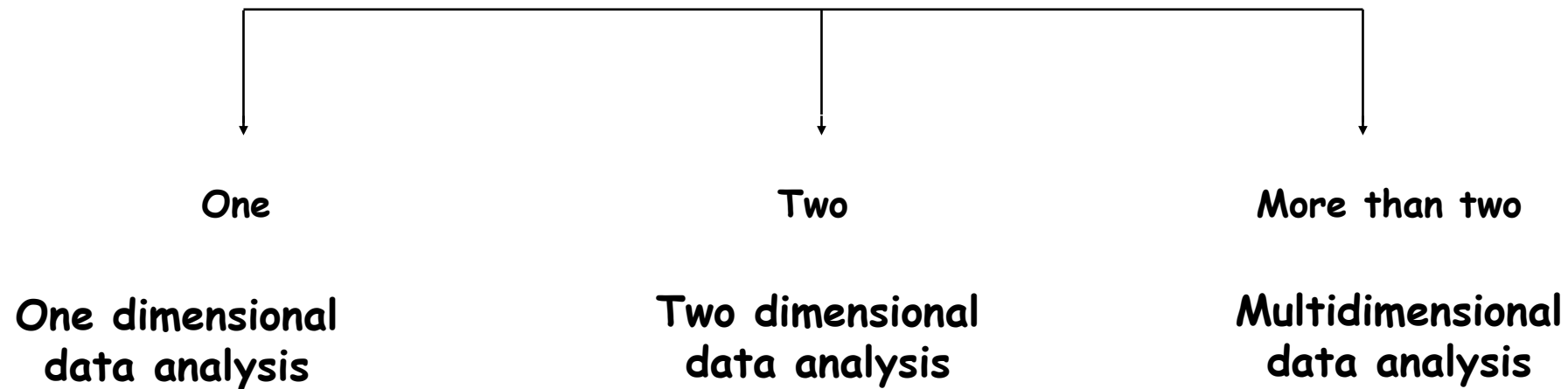
Systematic

Descriptive statistics

- Data analysis
- Interpretation of results

Data analysis

- The number of variables which are the subject of analysis at the same time



Measurement scales

➤ Possible measurement scales

➤ Weak, nonmetric, qualitative

- Nominal (categorical)

- Ordinal (hierarchical)

➤ Strong, metric, quantitative

- Interval

- Quotient (ratio)

- Interval - ratio

Measurement scales

Examples of variables

Scale	Nominal	Ordinal	Interval	Ratio
Examples	Firm profile, colour, car type, profession	Product quality, preference, risk level	Temperature, time, financial result	Sales value, price, advertisement spending, guarantee period, consumer income

Changing the measurement scale

- Possible changes of the measurement scales
 - Only the results of the stronger measurement scale can be transformed to weaker scale
 - Quantitative methods that can be used for the results of weak scale measurement results, can also be used to data obtained from the stronger measurement scale, i.e.
 - the strong measurement results may be coded into weaker scale

Measurement scales

Relationships and arithmetic operations

Scale	Nominal	Ordinal	Interval	Ratio
Acceptable relations	Equality Inequality	The majority The minority	Equality of differences Equality of intervals	Equality of ratios
Permitted arithmetic operations	Event counting (number of equality relations; number of inequality relations)	Event counting (number of equality relations; number of inequality relations; equality, majority, minority)	Addition Subtraction	Multiplication; Division

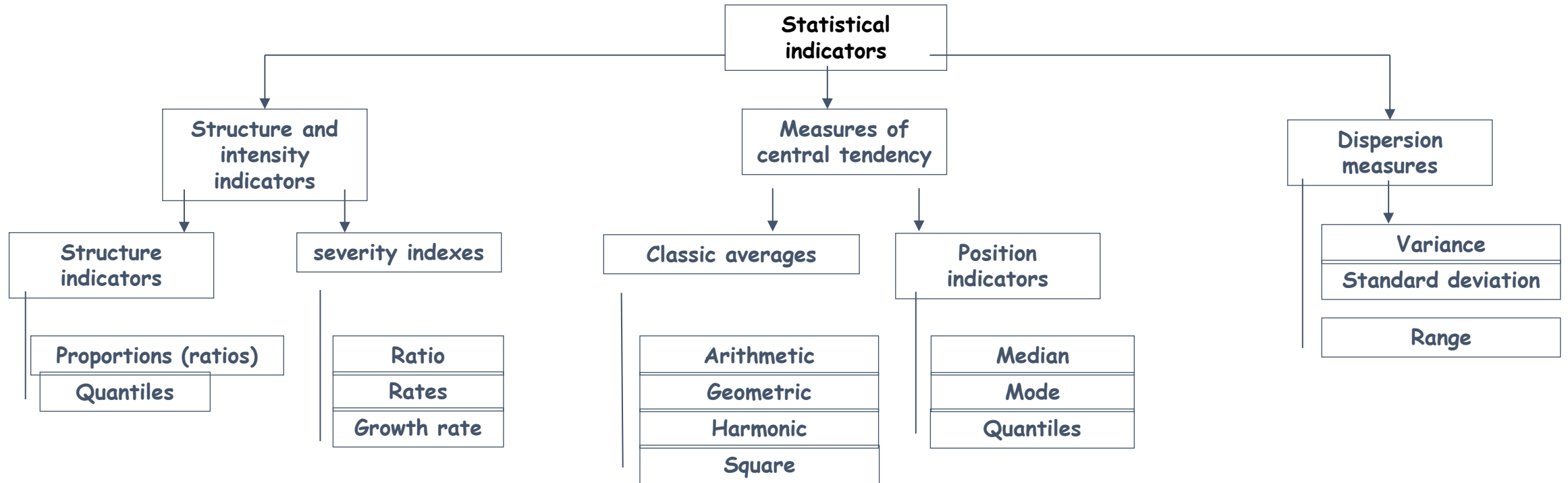
Measurement scales

Coefficients

Scale	Nominal	Ordinal	Interval	Ratio
Measures of position	Mode	Median	Arithmetic mean	Geometric mean Harmonic mean
Dispersion measures	Information measures	Interquartile ranges	Variance, standard deviation	Variation coefficient
Interdependence measures	Pearson measure of statistical dependence	Kendall correlation coefficient τ (tau)	Pearson correlation coefficient	

Statistical analysis of data

Overview of statistical indicators



Statistical analysis of data

One-dimensional

- **Distribution analysis**
 - Average values
 - Measures of dispersion
 - Measures of skewness
 - Measures of concentration
- **Tabulation**
- **Graphic illustration**

Statistical analysis of data

Multi-dimensional

- **Analysis of dependence**
 - Correlation
 - Regression
 - Analysis of interdependence
- **Cluster analysis (Grouping)**
 - Classification
 - Discrimination
- **Linear ordering (hierarchization)**

Statistical analysis of data

➤ Measures of central tendency

➤ Averages

- arithmetical
- geometric
- harmonic

➤ Position indicators

- quantiles
 - Median
 - quartiles
 - quintiles
 - decile
 - percentiles

ARITHMETIC AVERAGE

ARITHMETIC MEAN

- This is the sum of **VALUES** of all the surveyed units divided by number of units
- The average value lies in the middle of all the values
VOLUME (WEIGHT)
- The sum of the deviations from mean value is zero
- The sum of the squared deviations is minimal

Arithmetic mean calculation

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{x} = \frac{\sum_{i=1}^K n_i x_i}{\sum_{i=1}^K n_i} = \frac{\sum_{i=1}^K n_i x_i}{n}$$

$$(xc) = \bar{x}c \quad \overline{(\bar{x} + c)} = \bar{x} + c$$

ARITHMETIC AVERAGE

ARITHMETIC MEAN

- The arithmetic mean is appropriate measure when the population is homogeneous and symmetric
- The weighted arithmetic mean is appropriate measure when the distribution is organised into equally wide intervals which are closed
- The arithmetic mean is very sensitive to outliers and unusual and extreme observations

HARMONIC MEAN

- Like the arithmetic average, also harmonic mean preserves unchanged sum of the values held by the individual units
- The harmonic mean is equal to the arithmetic mean calculated from the reciprocal value of the variable in the population surveyed
- It is appropriate for investigation average ratios, eg. the pace, intensity, etc.

HARMONIC MEAN EXAMPLES

- The average speed of rotation of cash
- The average price of a product (expressed in natural units on the monetary unit)
- The average time required to perform certain actions (eg. Per hour)
- The average speed of vehicles

HARMONIC MEAN CALCULATION FORMULA

$$H_x = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_2}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

GEOMETRIC MEAN

- It is used to study the rate of change and the effects of the substantial differences between objects
- It is used to determine the average part of a whole, for example:
 - The average rate of return (average expressed in percentages)
 - The average interest rate - variable rate

GEOMETRIC MEAN calculation formula

$$G_x = (x_1 x_2 \dots x_n)^{\frac{1}{n}} = \left[\prod_{i=1}^n x_i \right]^{\frac{1}{n}}$$

$$H_x \leq G_x \leq \bar{x}$$

GEOMETRIC MEAN

- The geometric mean has a property in that if at least one value is zero, this average is also zero
- The geometric mean can be a complex number if at least one of the values of the variable is negative
- It follows that **GEOMETRIC MEAN** is suitable only to characterize the distributions having a positive values

Position indicators

- These are the values of specific individuals, chosen because of their position in the population surveyed.
- To determine their position the population has to be ordered.
- The median is the value at which the NUMBER of units having a value not greater than the median is equal to the NUMBER of units having a value not less than the median

Position indicators **MEDIAN**

- In point series, the median calculation is based on the observation whether the number of units is even.
- In the case of an odd number of observations the position (number) of unit, which is the median value, determined from the equation (1)
- In the case of the sample in which the number of units is even, the median is the arithmetic mean of the two values lying in the middle of an ordered series (formula 2)

Position indicators **MEDIAN**
point serie

$$Nr_M = \frac{n + 1}{2}$$

$$M_x = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

Position indicators **MEDIAN** interval serie

$$M_x = x_0 + h \frac{\frac{n+1}{2} - n_s}{n_x}$$

Position indicators QUARTILES

- Quartiles, are the values that in an ordered series of it divides number of units into four equal parts
- As a result of this division we get so called quadrants (quarters)
- The first quartile is the value of the unit, which leads to division of all units into two parts
 - so that $\frac{1}{4}$ units have value lower than first quartile
 - and $\frac{3}{4}$ of units have value not less than first quartile

Position indicators QUARTILES

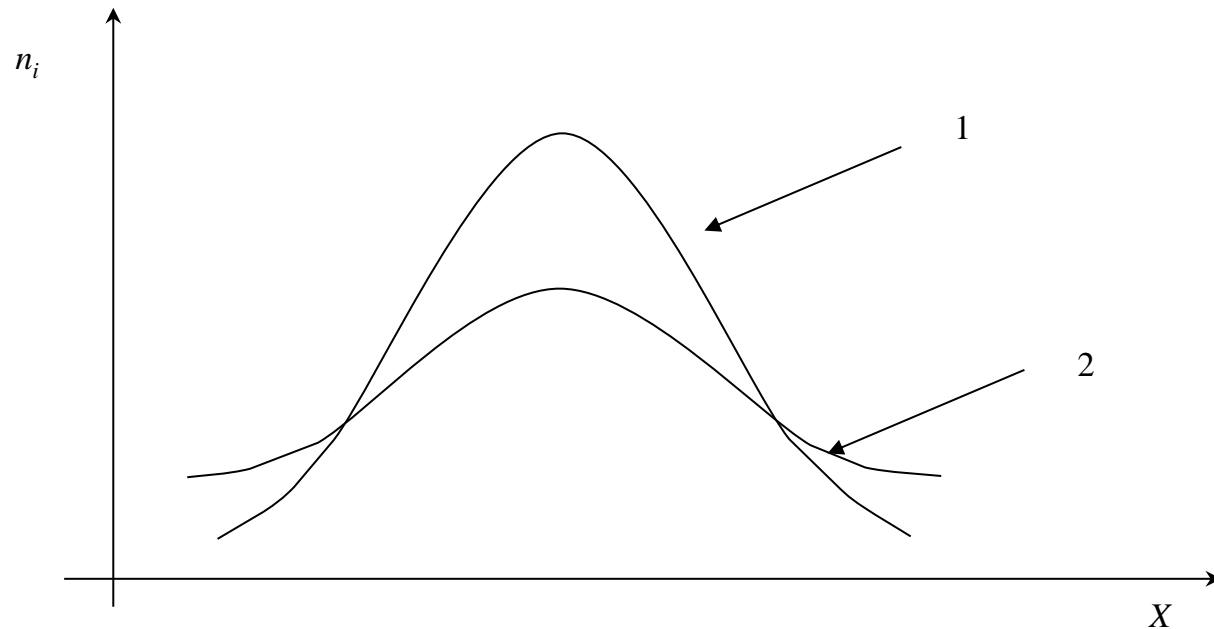
- The first and the third quartile is determined the same way as the median
- The formula for the calculation of the first quartile is as follows

$$Q_1 = x_0 + h \frac{\frac{n}{4} - n_s}{n_x}$$

Measures of variability

- Measures of variation, also known as metrics of dispersion or variability, serve to
 - Describes the degree of differentiation of population individuals by tested variable (characteristics)
- Graphically
 - greater concentration curve corresponding to a more "pointed" with a pronounced maximum (first curve shown)
 - greater dispersion - a curve of more planar shape (the second curve in the drawing)

Measures of variability



Measures of variability

- All measures of variability can be divided on absolute and relative
- Absolute measures answer to the question:
 - if and to what extent individuals differ by values of variable tested measured in natural units of measurement
- Absolute measures include:
 - Range (maximal minus minimal value)
 - Standard deviation and variance,
 - Variation indicators based on arithmetic mean, median and quartile
- Relative measures answer to the question whether the differences are relatively large
 - An example is the coefficient of variation

Relative measures of variability

- The RANGE or so called area of variation is the difference between the highest and the lowest value of the variable
- **Quartile deviation** is the average deviation of quartiles from the median
 - The disadvantage of this measure of volatility is that it is based on only two values, although characteristic, but only two units
- The advantage of quartile deviation indicator is, however, is that value of this measure will not be affected by outliers (non-standard values) understood as less than the first quartile and greater than the third quartile

Relative measures of variability

Quartile deviation

$$Q_x = \frac{(Q_3 - M_x) + (M_x - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$$

Relative measures of variability

- **Average deviation is the average arithmetic mean of the absolute values deviations from the arithmetic mean**
- **For a point series they shall be calculated according to formula (1)**
- **For a frequency distribution, according to formula (2)**

Relative measures of variability

Average deviation

$$D_x = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

$$D_x = \frac{1}{n} \sum_{i=1}^K |\dot{x}_i - \bar{x}| n_i$$

Relative measures of variability

VARIANCE

- The variance is the arithmetic mean of squared deviations from the arithmetic mean
- Calculated for a point series according to the formula (1)
- Since the variance represents a different measurement units than the value of the variable (squared units),
 - often for **variability** measurement the indicator is used in form of the square root of variance
- It is called **standard deviation**
- Standard deviation is measured with the same units as original variable

Relative measures of variability

VARIANCE

$$s_x^2 = \frac{1}{n} \sum_{i=1}^K \left(x_i - \bar{x} \right)^2 n_i$$

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

Relative measures of variability

VARIATION COEFFICIENT

- The coefficient of variation is the ratio of the absolute measure of the variability to the average value of this variable
- Most often used construct is the quotient of the standard deviation to the arithmetic mean
- Similarly, one can calculate the relative deviation for the (any) measurement of position, e.g. taking as a starting point quartile deviation and the value of the median rather than the standard deviation and arithmetic mean

Relative measures of variability

VARIATION COEFFICIENT

$$V_x = \frac{S_x}{\bar{x}}$$

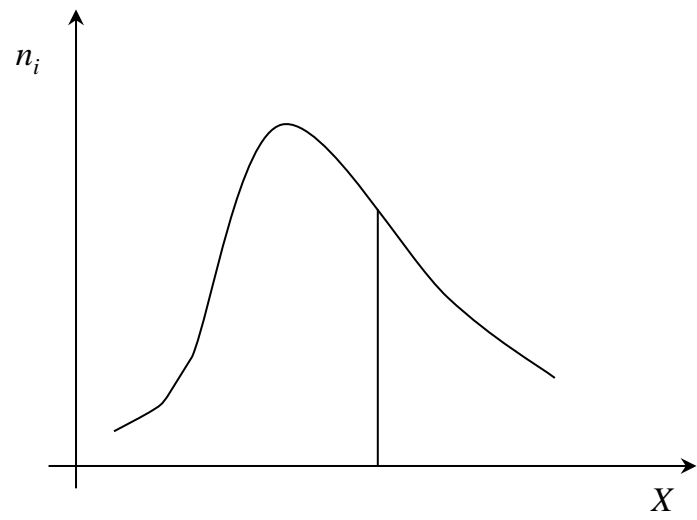
Relative measures of variability

VARIATION COEFFICIENT

- **The coefficient of variation is used for comparisons:**
 - **Several populations in terms of diversity**
 - **The same variable when you want to explore relative intensity of variation**
 - **The same population due to several different variables**

Skewness (Asymmetry) measure

- Classic coefficient of asymmetry (skewness) is a measure of the asymmetry in the distribution of one variable, based on the values of all observations
- It can be calculated also for the variables values ordered in the frequency distribution
- The classic measure of skewness does not inform sufficiently about the absolute power of asymmetry, because it can take values ranging from minus to plus infinity



Left side asymmetry



Right side asymmetry

Calculating asymmetry indicator

$$A_s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s_x^3}$$

$$A_s = \frac{\frac{1}{n} \sum_{i=1}^K \left(\dot{x}_i - \bar{x} \right)^3 n_i}{s_x^3}$$

Skewness (Asymmetry) - Quartile measure

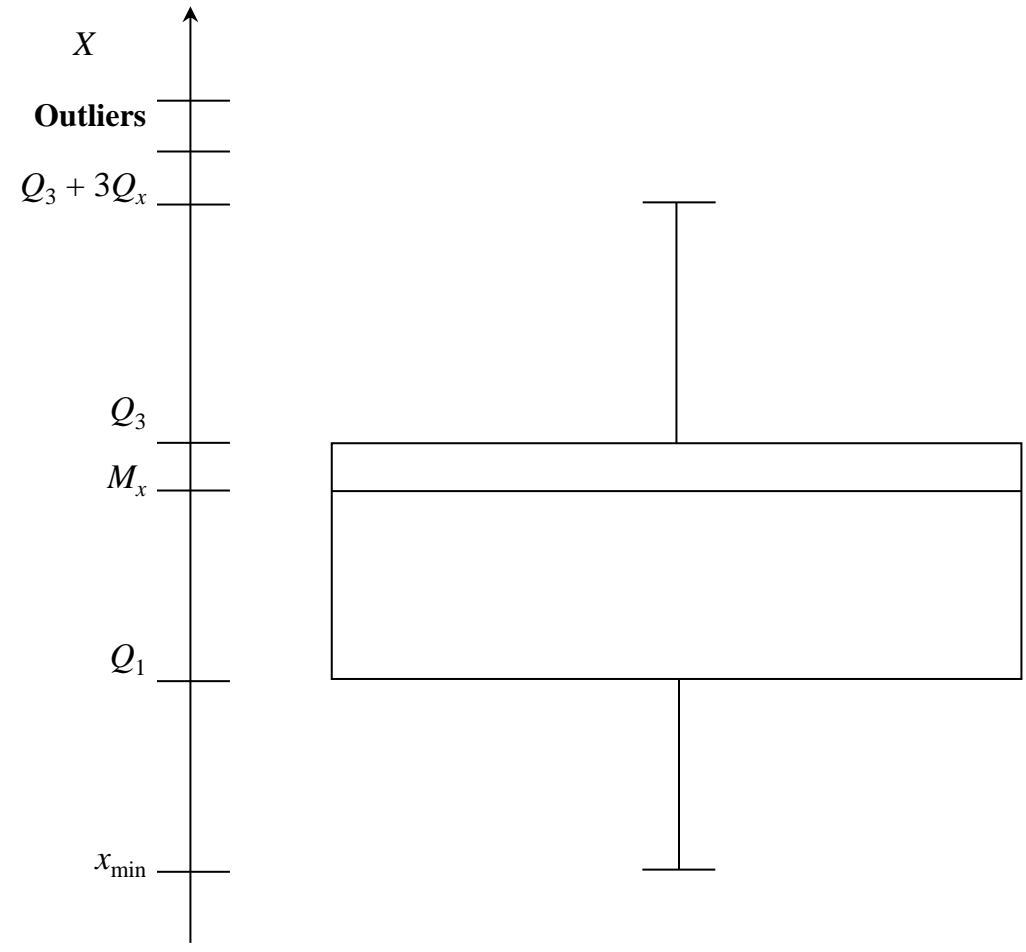
- Another measure of skewness is asymmetry (skewness) coefficient built based on quartile

$$A_Q = \frac{(Q_3 - M_x) - (M_x - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2M_x + Q_1}{Q_3 - Q_1}$$

GRAPHIC PRESENTATION

Box plot

- An important element of the statistical analysis series is the graphic presentation
- The graphic presentation is a concise way of their analysis
- There are many ways to graphically represent the data
- Interesting tool is the box plot



Box plot

Outliers identification

$$x_i < Q_1 - 3Q_x$$

$$x_i > Q_3 + 3Q$$

BOX PLOT

- On the basis of this type of graphical presentation, one can draw conclusions about:
 - The average values (center of the box - the median, the edges of the box - quartiles)
 - variation (the length of a box)
 - skewness (the location of the line dividing the box in relation to the edges of the box)
 - left hand side asymmetry - the dividing line is near the top of the box
 - right hand side asymmetry - the dividing line is closer to the bottom of the box
 - Outliers

Multidimensional comparative analysis

Multidimensional comparative analysis

- Multidimensional comparative analysis examines complex phenomenon, that is, those that are described by at least two variables
- The vector of variables is examined:

$$X = [X_1, X_2, \dots, X_m]$$

- where m - the number of variables ($m > 2$)

Multidimensional comparative analysis

- Raw data are presented in a matrix of observation
 - Where in every row there is an unit
 - In each column are values of individual characteristics (variable)
 - x_{ij} is value of i^{th} unit (object) for j^{th} variable

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

Multidimensional comparative analysis

- Using the methods of multidimensional comparative analysis solves usually the following two issues:
 - Similarity study of units by using a complex phenomenon classification methods
 - The arrangement of the observations (units, objects) with regard to the level of a complex (multivariate) phenomenon description using the methods of linear ordering

Multidimensional comparative analysis

- In practical tasks it is a rule that the variables included in the vector of variables describing units are expressed in different units of measurement
- From a formal point of view, most of the methods of multidimensional comparative analysis can be used only if all variables are expressed in the same units of measurement of similar orders of magnitude
- To achieve this, the normalization of variable's values is necessary

Multidimensional comparative analysis

- ✦ The normalization of variable's values is necessary
- ✦ It aims to:
 - ✦ unification of units of measurement variables,
 - ✦ unify the orders of magnitude of variables
- ✦ The most common way is the standardization.

Multidimensional comparative analysis

- ✦ The most common way is the standardization.
- ✦ This shall be done according to the formula

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (i = 1, 2, \dots, m)$$

where: \bar{x}_j - the arithmetic mean of the variable X_j
 s_j - the standard deviation of the variable X_j
 z_{ij} - standardized variable x_i for the i -th object

Multidimensional comparative analysis

- After standardization of the values of all variables, they become abstract numbers
- Standardization also results in unifying all the variables in respect to their variation and position indicator. This is because standardized variables:
 - The arithmetic mean is ZERO
 - The variance is equal to ONE
 - The standard deviation is equal to ONE

Multidimensional comparative analysis

➤ Another way to normalize variables is unitarisation

$$z_{ij} = \frac{x_{ij} - \min_i x_{ij}}{O_j}$$

Where O_j - range of variable X_j ; i.e.

$$O_j = \max(X_j) - \min(X_j)$$

Multidimensional comparative analysis

- As with the standardization of the unitarisation produces values of all the variables which are unit less,
- Additionally, after unitarisation:
 - all variable values are contained in the interval [0: 1],
 - the arithmetic mean of each variable is equal to:

$$\bar{z}_j = \frac{\bar{x}_j - \min_i x_{ij}}{O_j}$$

Multidimensional comparative analysis

Matrix of standardized observations

- Matrix of standardized observation has the same size as original matrix ($n \times m$)

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1j} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2j} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{i1} & z_{i2} & \dots & z_{ij} & \dots & z_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nj} & \dots & z_{nm} \end{bmatrix}$$

Multidimensional comparative analysis distance measure

- The notion enabling mutual comparison of objects in terms of a complex phenomenon is the multidimensional (taxonomic) distance
- The similarity (dissimilarity) of objects in the multidimensional space is usually measured by DISTANCES

Multidimensional comparative analysis distance measure

➤ Most often used, is so-called Euclidean distance:

$$d_{il} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{lj})^2} \quad (i, l = 1, 2, \dots, n)$$

➤ where:

- d_{il} - the distance between the object i and object l
- z_{ij} - normalized value of the variable X_j in the unit i
- z_{lj} - normalized value of the variable X_j in the object l

Multidimensional comparative analysis distance matrix

- After calculating the distance between each pair of objects distance matrix is obtained
- D - square matrix of dimension $n \times n$; where n - the number of objects:

$$D = \begin{bmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & 0 \end{bmatrix}$$

Multidimensional comparative analysis the task

- The methods of multidimensional comparative analysis, which are discussed later in this section, require standardized data, therefore preliminary stage in these methods is the normalization (standardization or unitarisation).
- Two types of tasks lead to many methods of multidimensional comparative analysis
- We will discuss:
 - Methods **GROUPING** set of observations
 - Classification
 - Discrimination
 - Methods of linear ordering a set of observations

Multidimensional comparative analysis

Classification methods

- Classification method enables partition of a set of n multivariate observations (units, objects) into K disjoint and non-empty subsets, called classes, so that:
 - observations belong to the same class were as similar as possible
 - observations belonging to different classes were also least similar as possible

Multidimensional comparative analysis

Classification methods

- Between observations belonging to individual classes - and therefore also the distance from the observation to the vector of position indicators - should be as low as possible

Multidimensional comparative analysis

Classification methods: METHOD OF GRAVITY CENTRES

- In practice, the method of gravity is used by the algorithm, the idea is as follows:
 - First, determine the number of classes which divide the set of observations
 - Next is an iterative algorithm
 - It starts from the initial screening of the selected set of observations consisting of K classes
 - Then, in each iteration is carried out the following activities:

Multidimensional comparative analysis

Classification methods: METHOD OF GRAVITY CENTRES

- Vectors of averages (centres of gravity) for each class obtained in the previous iteration (or classification initial) are calculated
- n_k - number of objects belonging to the k-th class

$$v_{kj} = \frac{1}{n_k} \sum_{i \in C_k} z_{ij}$$

Multidimensional comparative analysis

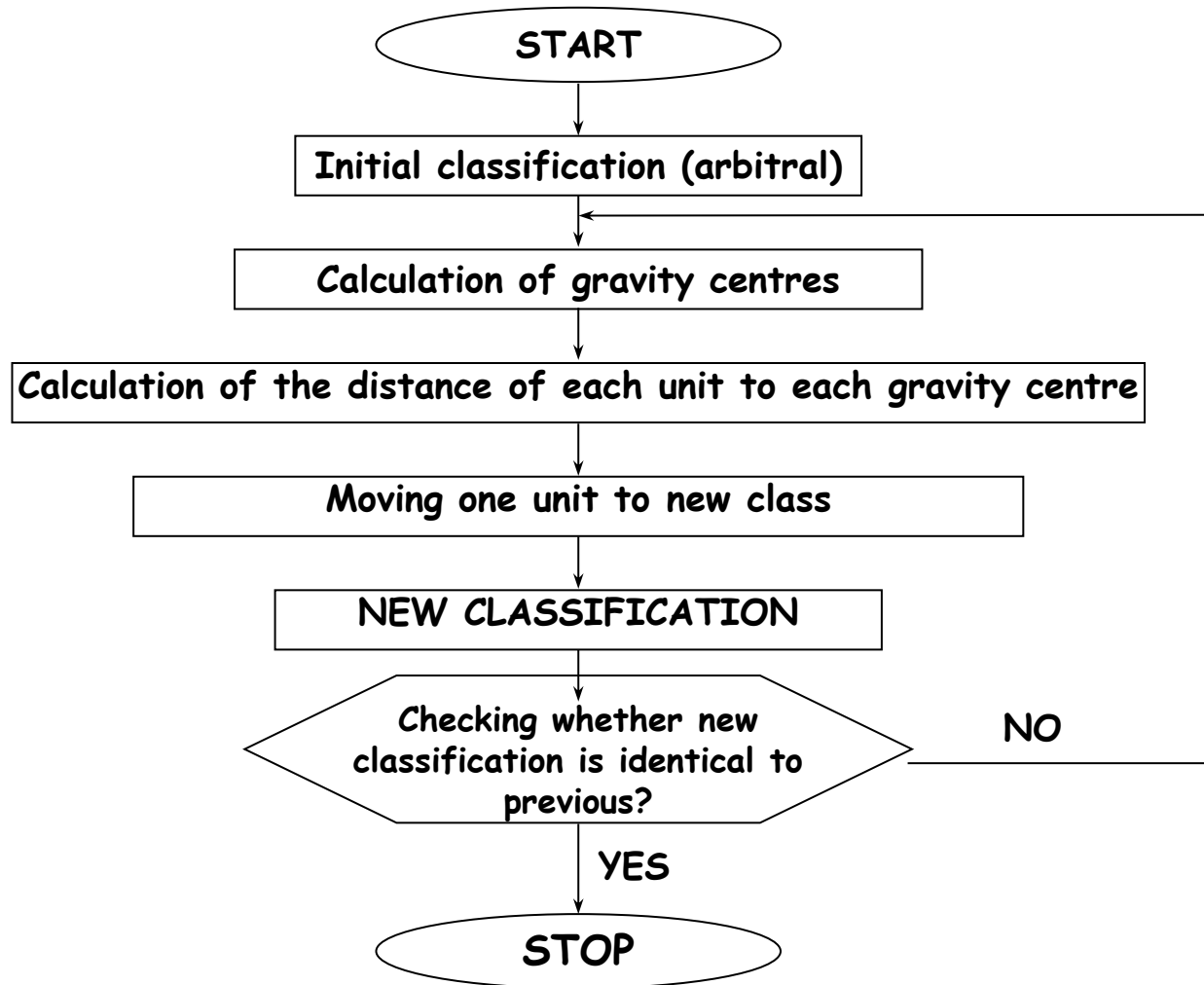
Classification methods: METHOD OF GRAVITY CENTRES

- The distance of each observation to the GRAVITY CENTRE (vector of average values of each variable in each class) is calculated
- A new classification is determined by assigning each observation to the class that is closest to this observation, ie. The class for which the distance from the average vector is the smallest

Multidimensional comparative analysis

Classification methods: METHOD OF GRAVITY CENTRES

- The calculation is continued until the classifications obtained in two successive iterations are the same
- Functional diagram of the method of gravity is the centre of gravity is shown in Figure on next slide



Multidimensional comparative analysis linear ordering

- The specificity of the methods of linear ordering requires that all the features have the same character (interpretation). There are three types of variables (attributes):
 - Stimulants - variables whose value growth is desirable
 - Destimulants - variables whose value reduction is desirable
 - Nominants - variables for which it is recommended that a nominal value or range of values

Multidimensional comparative analysis linear ordering

- Stimulus is variable, where the increase in the value testifies to the increase in the level of a complex phenomenon
- Destimulant and is variable, where the decline in the value testifies to the increase in the level of a complex phenomenon
- Nominants is variable that a particular value (or range of values) is desired, indicates an increase in the level of a complex phenomenon

Multidimensional comparative analysis linear ordering

- For example, to describe a complex phenomenon, which is the socio - economic development, we can choose the following variables
- Stimulants
 - Gross national income per capita
 - Protein consumption per capita
 - The number of telephones per 1000 inhabitants (?)
- Destimulants
 - Infant mortality per 1000 live births
 - The number of inhabitants per one medicine doctor

Multidimensional comparative analysis linear ordering

- To assess the situation of enterprises one can use the following variables:
 - For analysis of sales trends:
 - x1 revenue growth from sales in constant prices, the index chain, previous period = 100%; **Stimulant**
 - For analysis of liquidity:
 - x2 current ratio (A current assets/current liabilities)
nominant wit the recommended range of 1.2 - 2.0,
 - x3 quick ratio (Current assets - inventories/current liabilities)
nominants the recommended range of 1.0 - 1.3,
 - x4 of inventory of finished products in days (finished goods inventories/net sales * 360); **Destimulant,**
 - x5 series of cash on + inventory turnover period of debt collection - liabilities; **Destimulant**

Multidimensional comparative analysis linear ordering

- Destimulants can be easily transformed to the form of stimulants
 - This includes:
 - multiplying all its value by -1 (the standardized data), or
 - by subtracting from large positive number (>4) all the values (the standardized data), or
 - by subtracting from unity all the values (unitarised data)

Multidimensional comparative analysis linear ordering

- The purpose of the methods of linear ordering is to organize the observations in the order from "best" to "worst", the criterion is the order of a complex phenomenon

Multidimensional comparative analysis linear ordering

- At the core of these methods is the idea that there is some overriding criterion which allows the linear arrangement of observation
- Interesting ordering method is so called linear development pattern (Hellwig)
- It requires that the values of the variables are standardized

Multidimensional comparative analysis

linear ordering - normalization - nature of the variables

➤ Stimuli

$$Z_{it} = \frac{X_{it}}{\max_t X_{it}}$$

➤ Nominants with the recommended range of values $[x_{i,\min}, x_{i,\max}]$

$$Z_{it} = \begin{cases} 1 & \text{dla } x_{i,\min} \leq x_{it} \leq x_{i,\max} \\ \frac{x_{it}}{x_{i,\min}} & \text{dla } x_{it} < x_{i,\min} \\ \frac{x_{i,\max}}{x_{it}} & \text{dla } x_{it} > x_{i,\max} \end{cases}$$

➤ Destimuli

$$Z_{it} = \frac{\min_t X_{it}}{X_{it}}$$

Multidimensional comparative analysis linear ordering

- Based on standardised variable values, the composite indicator having the character **STIMULANTS** is build
- Values of the composite indicator (synthetic variable) can be determined as:
 - the sum of the values of the component variables
 - the arithmetic mean value the component variables
 - weighted average values of the variables involved, provided researcher considers that the individual variables carry information with the different weight (importance, significance)

Multidimensional comparative analysis linear ordering - procedure

- Presented method of linear ordering method carry a description of STANDARD DEVELOPMENT MEASURE.
- It is there that the values of the variables are standardized and all have a unitary character stimulant or destimulants
- The algorithm of this method is a three-step:
 - First designated as an abstract observation, the so-called. development pattern of the "best" values for each variable

Multidimensional comparative analysis linear ordering - procedure

- Presented method of linear ordering method carry a description of STANDARD DEVELOPMENT MEASURE.
- It is there that the values of the variables are standardized and all have a unitary character stimulant or destimulants
- The algorithm of this method is a three-step:
 - First designated as an abstract observation, the so-called. development pattern of the "best" values for each variable

Multidimensional comparative analysis linear ordering - definition of standard development measure

$$\mathbf{z}_0 = \left[z_{01} \ z_{02} \ \dots \ z_{0j} \ \dots \ z_{0m} \right]$$

Multidimensional comparative analysis linear ordering - definition of anti development measure

$$\mathbf{Z}_{-0} = \left[z_{-01} \ z_{-02} \ \dots \ z_{-0j} \ \dots \ z_{-0m} \right]$$

Multidimensional comparative analysis linear ordering - procedure

- Then the similarity of observation to abstract the "best" observation by calculating the distance of each observation from the pattern of development is examined

$$d_{i0} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j})^2} \quad (i = 1, 2, \dots, n)$$

- The more similar to the pattern (less distant from it) is the observation, the higher the level of development a complex phenomenon for this observation

Multidimensional comparative analysis linear ordering - procedure

- The final stage is to determine for each observation so. measure development according to the formula

$$m_i = 1 - \frac{d_{i0}}{d_0} \quad (i = 1, 2, \dots, n)$$

- Where

$$d_0 = \sqrt{\sum_{j=1}^m (z_{0j} - z_{-0j})^2}$$

Multidimensional comparative analysis linear ordering - development measure properties

- A measure of development has the following properties:
 - The higher the level of a complex phenomenon, the higher the value of the measure
 - Development of measurement values are included in the interval $[0; 1]$, wherein:
 - measure calculated for the development of the pattern of development is equal to ONE
 - for the anti pattern - ZERO