Peer-reviewed version available at Sustainability 2018, 10, 1637; doi:10.3390/su10051637

Article

Classification of economic regions with regards to selected factors characterizing the construction industry

Bożena Hoła [1] and Tomasz Nowobilski [2,*]

[1] Department of Construction Technology and Management, Faculty of Civil Engineering, Wroclaw University of Science and Technology, 50-370 Wrocław, Poland; bozena.hola@pwr.edu.pl
[2] Department of Construction Technology and Management, Faculty of Civil Engineering, Wroclaw University of Science and Technology, 50-370 Wrocław, Poland; tomasz.nowobilski@pwr.edu.pl
* Correspondence: tomasz.nowobilski@pwr.edu.pl; Tel.: +48-71-320-39-71

Abstract: The article presents the methodology for classifying economic regions with regards to selected factors that characterize a region, such as: the economic structure of the region, and thus the share of individual sectors in the economy; employment; the dynamics of the development of individual sectors expressed as an increase or decrease in production value; the population density in the region and also the level of occupational safety. Cluster analysis, which is a method of multidimensional statistical analysis available in Statistica software, was used to solve the task. The proposed methodology was used to group Polish voivodships with regards to the speed of economic development and occupational safety in the construction industry. Data published by the Central Statistical Office was used for this purpose, such as the value of construction and assembly production, the number of people employed in the construction industry, the population of an individual region and the number of people injured in occupational accidents.

Keywords: economic regions, regional classification, classification methodology, construction industry, cluster analysis, accidents in construction

---

1. Introduction

On the basis of published indicators of economic development [1, 2, 3], it can be stated that individual regions of the world are economically developed to a different extent. An inherent attribute of every economic activity is the phenomenon of the accident rate, and one of the most accident-prone sections of the economy is the construction industry [4]. When analyzing construction statistics, it can be noticed that the values of indicators that characterize the construction industry in developed countries are definitely higher than in developing countries, while the values of indicators characterizing occupational safety in developed countries are significantly lower than those in developing countries. The same differentiation can be observed between the internal economic regions of countries [5].

In scientific and engineering research, the problem of identifying objects with similar characteristics is very common. When carrying out such research, it is essential to properly classify objects that are described by many features into appropriate groups. In this context, research was carried out that aimed to develop a methodology for classifying economic regions in terms of selected factors characterizing the construction industry. One of the methods of multidimensional statistical analysis - cluster analysis, which is available in Statistica software, was used to solve the task. It involves segmenting the data set into subsets in order to distinguish homogeneous objects in an analysed set [6, 7].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2018 doi:10.20944/preprints201804.0371.v1

Peer-reviewed version available at Sustainability 2018, 10, 1637; doi:10.3390/su10051637

The developed methodology, along with an example of its application for classifying Polish voivodships with regards to the factors that describe the construction industry regarding occupational safety and economic aspects, is proposed in this article.

The article was organized in the following way: Chapter 2 presents a review of literature related to the topic of the article and also the justification of undertaking the research topic; Chapter 3 presents and discusses the proposed methodology for the classification of economic regions with regards to selected factors that characterize the construction industry; Chapter 4 contains an example illustrating the application of the developed methodology for the classification of Polish voivodships; and finally, Chapter 5 is a summary of the carried out research.

2. Literature review – application of data classification methods in scientific research

Conducting scientific research requires the processing and analyzing of large amounts of data. In some scientific disciplines, this may complicate or make it impossible to properly investigate phenomena, which may result in situations in which information relevant to researchers remains invisible. In this case, it is necessary to properly organize the observed data into structures, or group it [6, 7]. Methods of multidimensional statistical analysis, including e.g. cluster analysis, are helpful in solving such a research problem. The concept of cluster analysis was introduced in paper [8] and covers various classification algorithms for data exploration. The final effect of the calculations carried out using the above algorithms is the allocation of the analyzed data to appropriate groups in which the individual elements show mutual similarities. It gives an opportunity to capture structures that in the real world create the analyzed data, and to also reduce them to a level that allows them to be properly analyzed.

Data grouping methods are used in many fields of science, including medicine, social sciences, agricultural and technical sciences, as well as in the economy. In the publication of Hartigan [9], many examples of the use of taxononic methods can be found, and for their fundamental application, the author considers the development of the classification of animals and plants that had already been done in the times of Aristotle, and then later by Linnaeus [10]. In addition, the author gives a number of examples of the use of classification methods in: archeology, anthropology, phytosociology, psychology and psychiatry, as well as in other fields. Grouping methods are very often used in medicine, e.g. for DNA code [11] and disease entity [12] analysis. This is due to the fact that in this type of analysis scientists have a lot of interrelated and structured data. However, relationships and structures are not usually directly visible. In economic sciences, classification methods are used, among others for: determining the market structure, determining a product's position on the market, identifying test markets, market segmentation [13], classifying sectors due to financial conditions, classification of the labor market [14] and spatio-temporal analyzes.

One grouping data method is cluster analysis, which is very often used for issues related to the real estate market, among others [15, 16, 17, 18] and technical sciences [19]. It is also a good tool for grouping the areas and regions of a given country. For example, in paper [20], it was used to group Polish voivodships where individual voivodships were characterized by a similar level of transport infrastructure development. Another example of grouping voivodships in Poland is work [21], in which allocation to the appropriate group was made on the basis of similarities in the state of higher education.

Grouping methods are also used for issues related to occupational safety. In paper [22], the authors carried out an analysis of accidents in the construction industry in Hong Kong using cluster analysis, the results of which allowed the identification of the most probable accident situations. A similar study can be found in publication [23], in which an analysis of accidents related to electrical and mechanical works was carried out.

The above literature review shows that data grouping, in order to get information about its similarity, is carried out in all areas of knowledge. An invaluable tool in the study of large amounts of interrelated and structured data are methods of multidimensional statistical analyzes including, among others, cluster analysis. The authors of the article use the method of cluster analysis in research

Peer-reviewed version available at *Sustainability* **2018**, *10*, 1637; doi:10.3390/su10051637

concerning the development of a methodology for grouping economic regions with regards to selected factors that characterize these regions.

3. Proposed research methodology

The subject of the research is the classification of the set of economic regions V with regards to selected factors that characterize the construction industry regarding aspects of economic development and occupational safety.

$$V = \{v; v = 1, \dots, V\}, \tag{1}$$

Based on the literature review, it was stated that the basic factors that characterize economic regions are: the economic structure of a region, thus the share of individual sectors of the economy in the entire economy; the dynamics of development of individual sectors of the economy expressed as an increase or decrease in production value; employment in individual sectors of the economy; population density in a region and also the level of occupational safety in sectors of the economy [24, 25, 26, 27, 28, 29, 30, 31].

Therefore, each economic region can be described by the vector of factors F_v:

$$F_v = [F_1, \dots, F_n, \dots, F_N]; \qquad v = 1, \dots, V, \tag{2}$$

where:

F_n – the factor taken for analysis ($n = 1, 2, 3, \dots, N$).

The set of economic regions V is characterized by the following matrix of factors:

$$F_V = \begin{bmatrix} F_{1,1} & \cdots & F_{n,1} & \cdots & F_{N,1} \\ F_{1,2} & \cdots & F_{n,2} & \cdots & F_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{1,v} & \cdots & F_{n,v} & \cdots & F_{N,v} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{1,V} & \cdots & F_{n,V} & \cdots & F_{N,V} \end{bmatrix} \tag{3}$$

where:

v – the economic region ($v = 1, 2, 3, \dots, V$),

n – the factor taken for analysis ($n = 1, 2, 3, \dots, N$).

In mathematical analyses, knowledge about numerical values of indicators that describe particular factors is important. Thus, the set of economic regions V is characterized by a two-dimensional matrix of indicators, which takes the following form:

$$I_V = \begin{bmatrix} I_{1,1} & \cdots & I_{n,1} & \cdots & I_{N,1} \\ I_{1,2} & \cdots & I_{n,2} & \cdots & I_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{1,v} & \cdots & I_{n,v} & \cdots & I_{N,v} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{1,V} & \cdots & I_{n,V} & \cdots & I_{N,V} \end{bmatrix} \tag{4}$$

The values of indicators adopted for calculations often differ by, e.g. a measuring unit or scale, which may negatively affect the grouping [6]. In order to deal with this, all numerical data should be subjected to standardization, and the choice of the appropriate standardization formula depends on the type of data [32]. In the proposed methodology, a standardization of variables was adopted as one of the standardization methods according to the following formula:

$$P_{n,v} = \frac{I_{n,v} - \overline{I_n}}{\sigma} \tag{5}$$

where:

$P_{n,v}$ – the values of indicators after standardization,

$I_{n,v}$ – the values of unstandardized indicators,

$\overline{I_n}$ – the average value in the analysed set of objects,

σ – the standard deviation of the value of I_n indicator.

Peer-reviewed version available at Sustainability 2018, 10, 1637; doi:10.3390/su10051637

The effect of standardization is the creation of a set of P_V parameters that describe the analyzed set of economic regions F_V, which are written in the form of a two-dimensional matrix. In this matrix, each row contains values of all the parameters that are related to one region, whereas each column contains the data of one parameter for all the regions. This matrix is described by the following formula:

$$P_V = \begin{bmatrix} P_{1,1} & \cdots & P_{n,1} & \cdots & P_{N,1} \\ P_{1,2} & \cdots & P_{n,2} & \cdots & P_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{1,v} & \cdots & P_{n,v} & \cdots & P_{N,v} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{1,V} & \cdots & P_{n,V} & \cdots & P_{N,V} \end{bmatrix} \tag{6}$$

The calculated parameters form the basis for grouping economic regions with the use of cluster analysis. Cluster analysis is one of the methods of data exploration, the main idea of which is to group the analyzed objects in such a way that in a given group there are objects "similar" to each other and also "dissimilar" to objects from other groups [6, 7]. The criterion for assessing the affiliation of an object to a given group is the measure of similarity. The function, which is inverse to the measure of "similarity", and thus the function of " dissimilarity" of objects that is a measure of the distance between them, is used for practical considerations. This means that if the distance between object O_a and O_b is greater than the distance between objects O_a and O_c, and therefore:

$$d(O_a, O_b) > d(O_a, O_c); \quad where: \ a \neq b \neq c; \ a, b, c \in \{v\} \tag{7}$$

then object O_a is more "dissimilar" to object O_b than to object O_c. Consequently, this leads to a situation where objects O_a and O_c can create a cluster because they are more "similar" to each other. There are different distance measures that are used in cluster analysis. Geometric distance in the multidimensional space - the Euclidean distance [6], was used to solve the discussed issue. The general formula of the Euclidean distance takes the following form:

$$d(O_a, O_b) = \sqrt{\sum_n (P_{n,a} - P_{n,b})^2} \tag{8}$$

where:

O_a, O_b – the assessed objects, i.e. economic regions a and b, where $a \neq b$, and $a, b \in \{v\}$,

$P_{n,a}, P_{n,b}$ – the designated parameter values P_n for economic regions a and b, where $a \neq b$, and $a, b \in \{v\}$.

In the analyzed task, objects are the individual economic regions v. For grouping objects, the use of hierarchical and agglomeration grouping techniques is proposed. The agglomeration technique, which is the most often used in research [6], involves the gradual connection of objects, which constitute separate clusters, into new clusters until all objects form one cluster. Each connection of two clusters is called a step. An important issue when determining the appropriate distance between clusters, apart from the choice of the above-mentioned distance measure, is to determine the method of merging objects. Different methods of merging objects were analyzed, including methods in which the distance between specific locations of clusters is determined (e.g. between a given object or center of gravity of a cluster), and also methods that use variance analysis - e.g. the Ward's method [33]. Based on the conducted analyses, it was found that the most unambiguous results are achieved using the Ward method, the main advantage of which is the grouping of objects in a way that allows clusters with a similar number of objects to be formed. This eliminates the so-called effect of chain-linking, and the newly created clusters are characterized by the smallest possible diversity between their individual elements. In the developed methodology, the method of merging objects using the Ward method was adopted.

The result of hierarchical cluster analysis is a tree-shaped graph - the so-called dendrogram. It shows in which step the objects connect with each other. However, it does not give an unambiguous answer for the correct number of clusters. This number depends on the place where the branches of the tree are cut off on the chart. Due to this, an important issue is to correctly determine the place of the cut-off. According to [6], there is no objective rule of how to do it. There are only supportive methods, such as e.g. the method of graphical dendrogram analysis that involves the examination of the distance between successive bonds, the method using the Grabinski meter [34] or the Mojen rule [35]. The developed methodology involves the method of graphical dendrogram analysis.

After selecting the appropriate place of cutting off the branches on the dendrogram, the clusters should be identified, and on their basis the final classification of the analyzed economic regions and their assessment and ranking should be made. The developed methodology for this classification is shown in Figure 1.

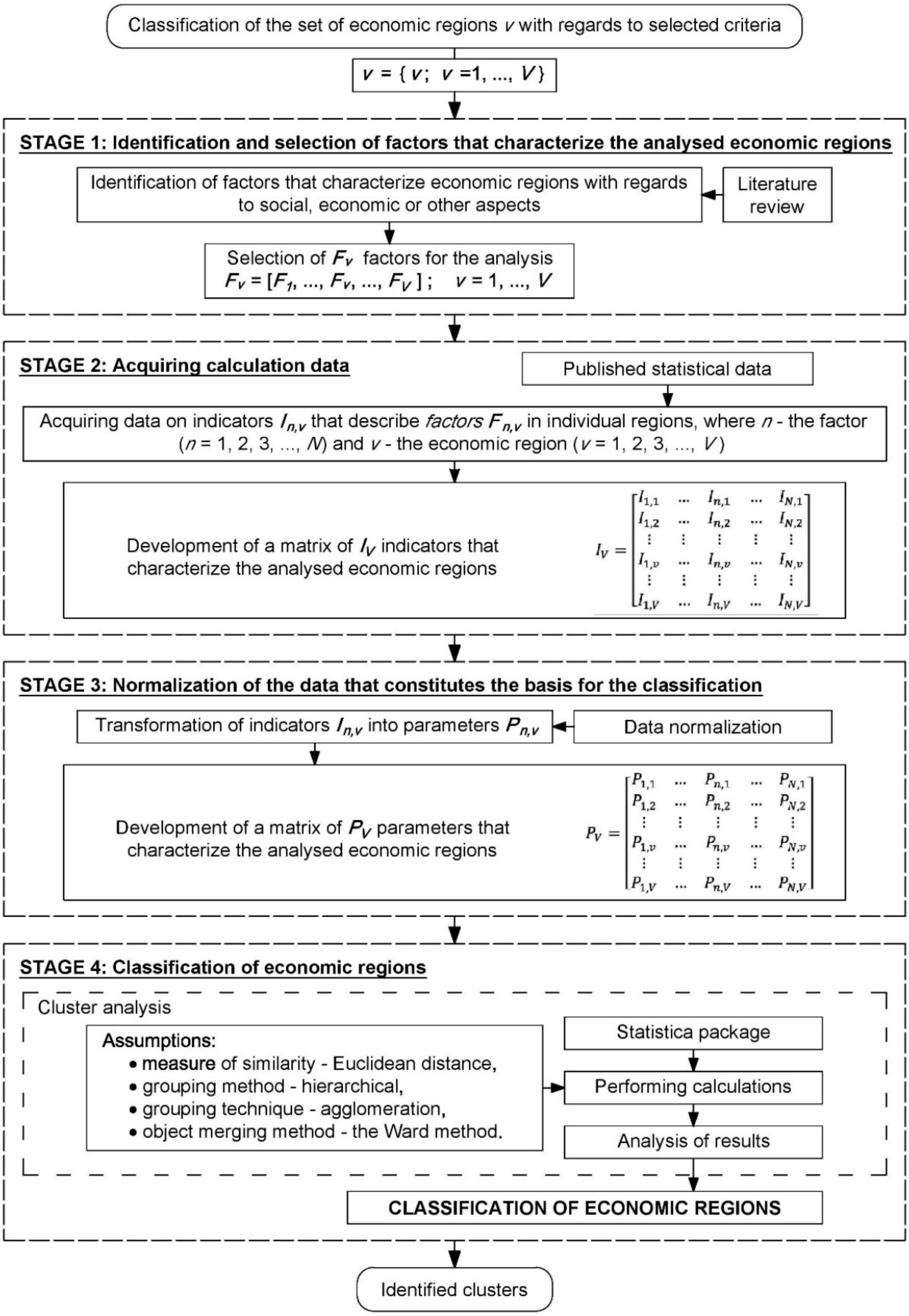

Figure 1. The developed methodology for classifying economic regions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2018 doi:10.20944/preprints201804.0371.v1

Peer-reviewed version available at Sustainability 2018, 10, 1637; doi:10.3390/su10051637

4. Application of the proposed methodology using the classification of Polish voivodships as an example

The proposed methodology was used to group Polish voivodships. The classification was based on data published by the Central Statistical Office [5]. The following indicators were adopted to describe the voivodships: the value of construction and assembly production ($I_{1,y}$), the number of people employed in the construction industry ($I_{2,y}$), the population of a given region ($I_{3,y}$) and the number of people injured in occupational accidents ($I_{4,y}$). The calculations include data for the period from 2008 to 2016 that was obtained for 16 Polish voivodships. Therefore, the number of analyzed indicators for individual voivodships amounted to 36, and some of them are shown in Table 1.

Table 1. Values of selected indicators.

	Economic indicators											
v	**Value of construction and assembly production $I_{1,v,y}$**			**Number of people employed in the construction industry $I_{2,v,y}$**			**Population of a given region $I_{3,v,y}$**			**Number of people injured in occupational accidents $I_{4,v,y}$**		
	2008 $y=1$	…	**2016 $y=9$**	**2008 $y=1$**	…	**2016 $y=9$**	**2008 $y=1$**	…	**2016 $y=9$**	**2008 $y=1$**	…	**2016 $y=9$**
1	14568.1	…	13721.2	68874	…	65039	2877059	…	2903710	892	…	413
2	5288.9	…	6216.6	43952	…	42918	2067918	…	2083927	607	…	269
…	…	…	…	…	…	…	…	…	…	…	…	…
16	7325.1	…	9104.0	39288	…	35483	1692957	…	1708174	362	…	144

where:
y – year ($y = 1, \dots, 9$),
v – voivodship ($v = 1, \dots, 16$).

The obtained values of indicators were subjected to standardization. The effect of this action was the creation of parameters $P_{n,v,y}$ constituting a set of normalized indicators, which were the basis for the classification of voivodships. Some of the data obtained after standardization is presented in Table 2.

Table 2. Values of selected parameters.

	Parameters											
v	**Value of construction and assembly production $P_{1,v,y}$**			**Number of people employed in the construction industry $P_{2,v,y}$**			**Population of a given region $P_{3,v,y}$**			**Number of people injured in occupational accidents $P_{4,v,y}$**		
	2008 $y=1$	…	**2016 $y=9$**	**2008 $y=1$**	…	**2016 $y=9$**	**2008 $y=1$**	…	**2016 $y=9$**	**2008 $y=1$**	…	**2016 $y=9$**
1	0.734	…	0.394	0.481	…	0.291	0.397	…	0.395	0.534	…	0.288
2	-0.537	…	-0.504	-0.264	…	-0.323	-0.254	…	-0.251	-0.110	…	-0.294
…	…	…	…	…	…	…	…	…	…	…	…	…
16	-0.258	…	-0.158	-0.403	…	-0.530	-0.555	…	-0.546	-0.664	…	-0.798

where:
y – year ($y = 1, \dots, 9$),
v – voivodship ($v = 1, \dots, 16$).

After conducting the calculations using Statistica software, the dendrogram shown in Figure 2 was obtained. When analysing the dendrogram, it can be noted that some voivodships very quickly create clear clusters. It was proposed to cut off the branch of the dendrogram at the place marked with a dashed line in Figure 2. With such a cut, four groups of voivodships were obtained, in which each voivodship has a similar level of the development of the construction industry and occupational safety. The identified groups of voivodships are given in Table 3.

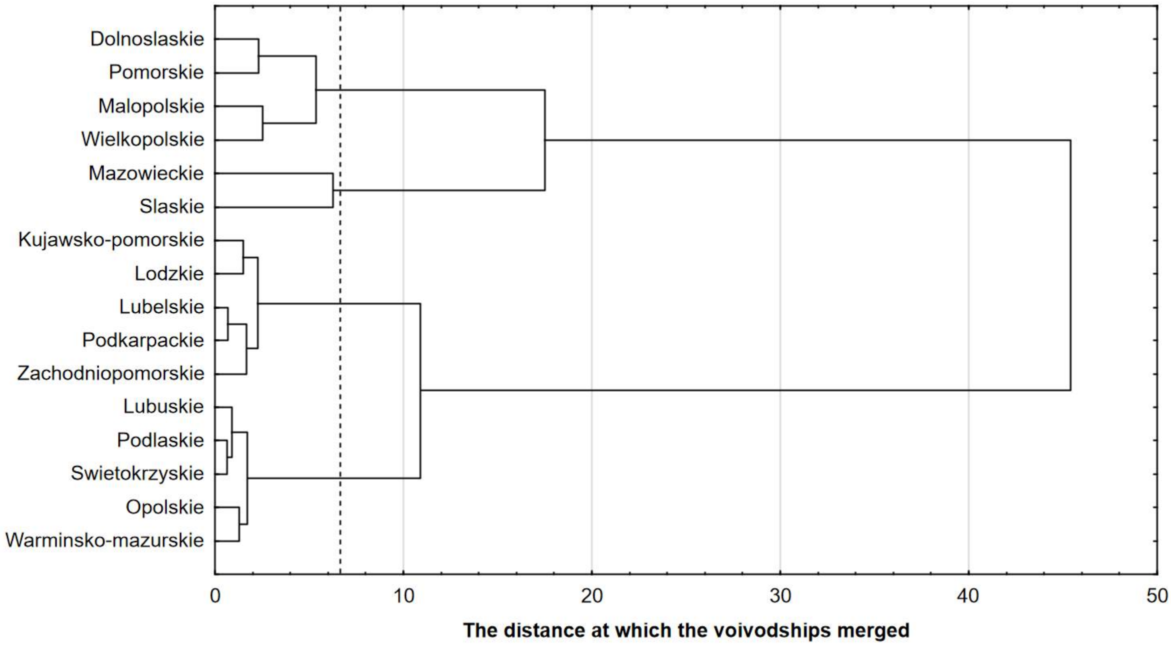

Figure 2. Dendrogram - a graph showing the connection of individual voivodships in the subsequent calculation steps.

Table 3. The obtained groups of voivodships, which are characterized by a similar speed of construction industry development and a similar level of occupational safety.

Cluster	Voivodships	The distance at which the voivodships merged
I	Dolnoslaskie, Pomorskie, Malopolskie, Wielkopolskie	5.38
II	Mazowieckie, Slaskie	6.25
III	Kujawsko-Pomorskie, Lodzkie, Lubelskie, Podkarpackie, Zachodniopomorskie	2.26
IV	Lubuskie, Podlaskie, Swietokrzyskie, Opolskie, Warminsko-Mazurskie	1.73

Based on the analysis of the obtained results, four clusters were distinguished, which include 2 to 5 voivodships. Cluster I consists of the voivodships Dolnoslaskie, Pomorskie, Malopolskie and Wielkopolskie; cluster II is formed from the voivodships Mazowieckie and Slaskie; cluster III consists of the voivodships Kujawsko-Pomorskie, Lodzkie, Lubelskie, Podkarpackie and Zachodniopomorskie; while cluster IV includes the voivodships Lubuskie, Podlaskie, Swietokrzyskie, Opolskie and Warminsko-Mazurskie.

A very fast creation of bonds between voivodships in clusters III and IV was observed. These clusters include the voivodships most similar to each other in terms of construction and assembly production, the number of people injured in occupational accidents, the number of people employed in the construction industry and the population living in the voivodship. The merging distance in the case of cluster III is equal to 2.26, while in the case of cluster IV is equal to 1.73

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 April 2018 doi:10.20944/preprints201804.0371.v1

Peer-reviewed version available at Sustainability 2018, 10, 1637; doi:10.3390/su10051637

The most different voivodships from all of the others are Mazowieckie and Slaskie. They form one cluster, however, the distance at which the merging between them occurred is equal to 6.25 and is more than twice as high as in the case of clusters III and IV. In turn, a comparable level of similarity between pairs of voivodships can be observed in cluster I, namely: Dolnoslaskie and Pomorskie, and also Malopolskie and Wielkopolskie. The merging distance in both pairs is equal to 2.31 and 2.53, respectively. However, the merging distance of these two pairs is equal to 5.38.

5. Summary

The article proposes original and universal methodology for classifying economic regions that are described by selected features. The universality of the methodology lies in the fact that it can be used for grouping other economic objects, e.g. enterprises, and also that other factors characterizing objects can be included in the calculations. The basis of the proposed methodology was one of the methods of multidimensional analysis of statistical data, namely cluster analysis. The developed methodology was used to classify Polish voivodships with regards to factors that characterize the speed of economic development in the construction industry and the level of occupational safety. The basis for the classification was statistical data published by the Central Statistical Office, which characterizes the size of construction production, the number of inhabitants of a voivodship, the number of people employed in the construction industry and the number of people injured in occupational accidents between 2008 and 2016 in the construction industry.

The conducted calculations and analysis of the results allowed the following conclusions to be formulated:

- The qualitative and quantitative structure of statistical data, which was the basis for the classification of voivodships, allowed four distinct clusters consisting from two to five voivodships to be separated.
- The isolated clusters are characterized by different levels of similarity, which is confirmed by the values of the merging distance measure for individual clusters. Cluster ranking with regards to the similarity of the voivodships that form clusters is as follows:
  1. cluster IV consists of the voivodships Lubuskie, Podlaskie, Swietokrzyskie, Opolskie and Warminsko-Mazurskie,
  2. cluster III consists of the voivodships Kujawsko-Pomorskie, Lodzkie, Lubelskie, Podkarpackie and Zachodniopomorskie,
  3. cluster I consists of the voivodships Dolnoslaskie, Pomorskie, Malopolskie and Wielkopolskie,
  4. cluster II consists of the voivodships Mazowieckie and Slaskie.
- The very big similarity between voivodships located in clusters III and IV means that voivodships included in these clusters are characterized by a similar level of construction and assembly production value, occupational safety, the number of people employed in the construction industry and the number of people living in the voivodship.
- The voivodships Mazowieckie and Slaskie are atypical voivodships. They are the most different when compared with the others. Although they form one cluster, the distance at which the merging between them occurs is relatively large when compared to the merging distance in the other clusters.

The proposed methodology can be applied both in the area of scientific research and engineering practice. The results of tests and analyses obtained using this methodology can be the basis for classifying and comparing objects and determining their rankings. The correct classification of objects (which are described by many factors) into groups can be important in determining the characteristics of a given community, making an assessment, or looking for dependencies that apply to this community. In the research conducted by the authors, information about voivodships belonging to the same cluster will be the basis for the construction of multifactorial linear regression models for predicting indicators describing the level of occupational safety in the construction industry. The practical aspect of the proposed methodology is related to the possibility of formulating conclusions that may be important at a higher management level.

Acknowledgments: The article is the result of the implementation by the authors of research project No. 244388 "Model of the assessment of risk of the occurrence of building catastrophes, accidents and dangerous events at workplaces with the use of scaffolding", financed by NCBiR within the framework of the Programme for Applied Research on the basis of contract No. PBS3/A2/19/2015.

Author Contributions: Formal analysis, Tomasz Nowobilski; Methodology, Bożena Hoła and Tomasz Nowobilski; Project administration, Bożena Hoła; Software, Tomasz Nowobilski; Supervision, Bożena Hoła.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Central Statistical Office. Available online: http://stat.gov.pl (accessed on January 2018).
2. U.S. Data and Statistics. Available online: https://www.usa.gov/statistics (accessed on January 2018).
3. Eurostat. Available online: http://ec.europa.eu/eurostat (accessed on January 2018).
4. International Labour Organization, *Safety and Health at Work: A Vision for Sustainable Prevention*, XX World Congress on Safety and Health at Work 2014, Global Forum for Prevention, Frankfurt, Germany, 24–27 August, 2014.
5. Central Statistical Office. *Statistical yearbook 2008-2016*; Department of Statistical Publishing: Warsaw, Poland, 2009-2017.
6. Stanisz, A. *A simple statistic course with the use of STATISTICA PL on the basis of medical examples, Volume 3: Multivariate analysis*; StatSoft, Cracow, Poland, 2007, ISBN: 978-83-88724-19-0.
7. Wierzchoń, S.; Kłopotek, M. *Algorithms of cluster analysis;* Publishing House of WNT, Warsaw, Poland, 2015, ISBN: 978-83-7926-261-8.
8. Tryon, R. C. *Cluster Analysis: Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality*. Edwards Brothers, Ann Arbor. 1939.
9. Hartigan, J. A. *Clustering Algorithms*, New York, Wiley, 1975.
10. Linnaeus, C. *Species plantarum*, Vol. 1, Impensis GC Nauk, C., 1753.
11. Fu, L.; Medico, E. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinformatics* **2017** 8(3), pp. 1-15. DOI: 10.1186/1471-2105-8-3.
12. Woytowicz, E. J.; Rietschel, J. C.; Goodman, R. N.; Conroy, S. S.; Sorkin, J. D.; Whitall, J.; Waller, S. M.; Determining Levels of Upper Extremity Movement Impairment by Applying a Cluster Analysis to the Fugl-Meyer Assessment of the Upper Extremity in Chronic Stroke. *Archives of physical medicine and rehabilitation* **2017** 98(3), pp. 456-462. DOI: 10.1016/j.apmr.2016.06.023.
13. Tkaczynski, A.; Segmentation Using Two-Step Cluster Analysis. *Segmentation in Social Marketing. Springer* **2017** Singapore, pp. 109-125. DOI: 10.1007/978-981-10-1835-0_8.
14. Markowska, M.; Sokołowski, A.; Strahl, D.; Sobolewski, M. Dynamic classification of regions of the European Union at the NUTS 2 level with regards to sensitivity to the economic crisis in the area of the labour market. *HSS* **2015**, vol. XX, 22, pp. 37-50.
15. Napoli, G.; Giuffrida, S.; Valenti, A. Forms and functions of the real estate market of Palermo (Italy). Science and knowledge in the cluster analysis approach. *Green Energy and Technology* **2017**, Issue 9783319496757, pp. 191-202, 2017, DOI: 10.1007/978-3-319-49676-4_14.
16. Wang, W.; Yang, J.; Gong, X. The Regional Real Estate Investment Environment Research Based on Prime Component Analysis: The Case of Shandong. ICCREM 2017: Real Estate and Urbanization - Proceedings of the International Conference on Construction and Real Estate Management 2017, pp. 217-224. DOI: 10.1061/9780784481073.024.
17. Gabrielli, L.; Giuffrida, S.; Trovato, M. R. Gaps and overlaps of urban housing sub-market: Hard clustering and fuzzy clustering approaches. *Green Energy and Technology* **2017** Issue 9783319496757, pp.203-219. DOI: 10.1007/978-3-319-49676-4_15.
18. Wang, N.; Li, H.-M.; Tan, X.; Zhong, X.-R. Development evaluation of real estate industry in China's major cities based on factor and cluster analysis. Xi'an Jianzhu Keji Daxue Xuebao / *Journal of Xi'an University of Architecture and Technology* **2010** Volume 42, Issue 4, Pages 590-594+603. DOI: .

19. Lifshitz, R.; Ostfeld, A. Clustering for analysis of water distribution systems. *Journal of Water Resources Planning and Management* **2018** Volume 144, Issue 5. Article no. 04018016, DOI: 10.1061/(ASCE)WR.1943-5452.0000917.
20. Migała-Warchoł A.; Sobolewski, M. Evaluation of voivodeships diversification in Poland according to transport infrastructure indications. *Quantitative Methods In Economics* **2013** Vol. XIV, No. 2, pp. 89 – 98.
21. Zalewska, E. Application of cluster analysis and methods of linear ordering in the assessment of Polish higher education. *SCIENTIFIC WORKS of Wroclaw University of Economics* **2017**, No 469, Taxonomy 29, Classification and data analysis - theory and applications, pp. 234-242.
22. Chiang, Y.-H.; Wong, F.K.-W.; Liang, S. Fatal Construction Accidents in Hong Kong. *Journal of Construction Engineering and Management* **2018** Volume 144, Issue 3, Article no. 04017121. DOI: 10.1061/(ASCE)CO.1943-7862.0001433.
23. Francis K. W. Wong; Albert P. C. Chan; Andy K. D. Wong; Carol K. H. Hon; Tracy N. Y. Choi, Accidents of Electrical and Mechanical Works for Public Sector Projects in Hong Kong. *International Journal of Environmental Research and Public Health* **2018** 15(3):485. DOI: 10.3390/ijerph15030485.
24. Hoła, A.; Hoła, B.; Szóstak, M. Analysis of the causes and consequences of falls from scaffolding using the Polish construction industry as an example. 3rd International Conference on Innovative Materials, Structures and Technologies: 27-29 September 2017, Riga, Latvia. [Bristol]: IOP Publishing, art. 012050, pp. 1-8. (IOP Conference Series - Materials Science and Engineering, ISSN 1757-899X; vol. 251). DOI: 10.1088/1757-899X/251/1/012050.
25. Hoła, B. Identification and evaluation of processes in a construction enterprise. *Archives of Civil and Mechanical Engineering* **2015** vol. 15, No. 2, pp. 419-426. DOI: 10.1016/j.acme.2014.11.001.
26. Hoła, B.; Sawicki, M.; Skibniewski, M. An IT model of a knowledge map which supports management in small and medium-sized companies using selected polish construction enterprises as an example. *Journal of Civil Engineering and Management* **2015** vol. 21, No. 8, pp. 1014-1026. DOI: 10.3846/13923730.2015.1030865.
27. Hoła, B.; Szóstak, M. Analysis of the state of the accident rate in the construction industry in European Union countries. *Archives of Civil Engineering* **2015** vol. 61, No. 4, pp. 19-34. DOI: .
28. Stępień, T. Identification of factors determining accident rate in construction industry. *Czasopismo Techniczne* **2014** y.111, iss. 1-B, pp. 265-281.
29. Hoła, B.; Nowobilski, T.; Szer, I.; Szer, J. Identification of factors affecting the accident rate in the construction industry. *Procedia Engineering* **2017** Vol. 11. DOI: 10.1016/j.proeng.2017.11.018.
30. Hoła, A.; Sawicki, M.; Szóstak, M. Methodology of Classifying the Causes of Occupational Accidents Involving Construction Scaffolding Using Pareto-Lorenz Analysis. *Applied Sciences* **2018** vol. 8, No. 1, art. 48, pp. 1-11. DOI: .
31. Klempous, R.; Kluwak, K.; Idzikowski, R.; Nowobilski, T.; Zamojski, T. Possibility analysis of danger factors visualization in the construction environment based on Virtual Reality Model. 8th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2017: proceedings, Debrecen, Hungary Danvers, September 11-14, 2017, MA : IEEE, cop. 2017. pp. 363-367. DOI: 10.1109/CogInfoCom.2017.8268271.
32. Jarocka, M. Selection of the normalization formula in the comparative analysis of multi-feature objects. *Economics and Management* **2015** No. 1, pp. 113-126, 2015. DOI: 10.12846/j.em.2015.01.08.
33. Ward, Jr. J. H.; Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **1963** Vol. 58, Issue 301, pp. 236–244.
34. Grabiński, T. *Methods of taxononometry*, Cracow University of Economics; Cracow, Poland, 2007.
35. Mojena, R. Hierarchical grouping methods and stopping rulet: an evaluation. *Computer Journal* **1977** 20, pp. 359-363.