

Classification of Polish voivodeships with regards to the accident rate in the construction industry

Tomasz Nowobilski^{1*}, *Irena Bagińska*², and *Krzysztof Gawron*¹

¹Wrocław University of Science and Technology, Faculty of Civil Engineering, Department of Construction Methods and Managements, Wrocław, Poland

²Wrocław University of Science and Technology, Department of Geotechnics and Hydrotechnics, Wrocław, Poland

Abstract. The article classifies Polish voivodeships into appropriate groups with a similar level of occupational safety in the construction industry. The basis for the adopted classification was statistical data published by the Central Statistical Office regarding population, employment in the construction industry, the value of construction production and the number of occupational accidents. The conducted research allowed a logical and correct, in terms of content, division of the Polish territory to be made, taking into account the aspect of occupational safety in the construction industry. Statistica software and cluster analysis were used to solve the problem.

1 Introduction

It is well known that the construction industry in various regions of Poland is developed to different extents. Differences in the number, size and type of investments mean that individual voivodeships differ from others. This can be seen, among others, on the basis of published statistical data.

The authors of the article conducted research that aimed to classify Polish voivodeships into appropriate groups, which are characterized by a similar level of occupational safety in the construction industry. The basis for the adopted classification was statistical data published by the Central Statistical Office (CSO) regarding population, employment in the construction industry, the value of construction production and the number of occupational accidents. Cluster analysis and Statistica computer software were used to solve the problem. This classification will be the basis for conducting various comparative analyzes and creating models for predicting the values of parameters that characterize various aspects describing the accident rate in the construction industry in particular groups of voivodeships.

2 Research methodology

* Corresponding author: tomasz.nowobilski@pwr.edu.pl

The research methodology presented in Figure 1 is proposed. It consists of four stages, which are described in the following chapters.

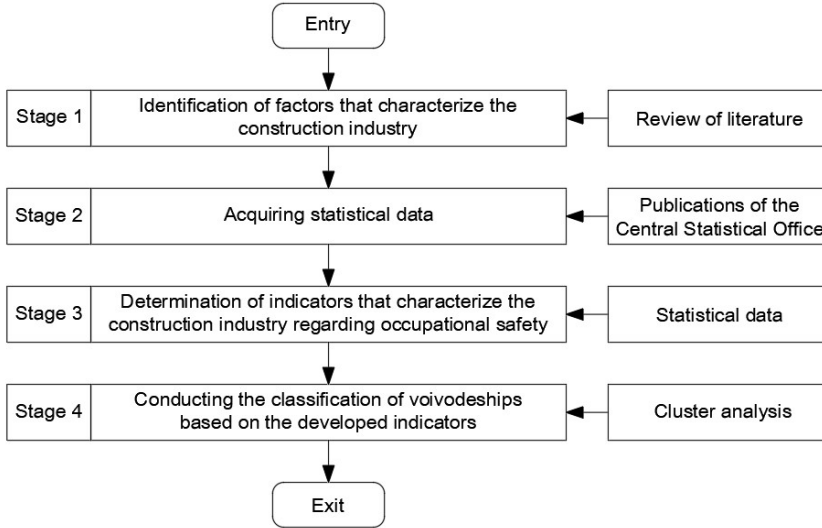


Fig. 1. Methodology of conducted research

2.1 Stage 1 – Identification of factors that characterize the construction industry

The level of development in the construction industry and also occupational safety in this section of the economy are characterized by many factors described, among others, in [1, 2, 3, 4, 5, 6, 7]. The most important of these factors are, among others: the value of construction production, the financial and material potential of construction enterprises, employment and also various aspects of occupational safety. Each of the listed factors is described by a number of sub-factors, which can be obtained, among others, from statistical yearbooks issued by the Central Statistical Office [8, 9].

The authors of the research applied the following factors for the classification of voivodeships with regards to the size of construction production and occupational safety: the value of construction and assembly production (W), the number of people employed in the construction industry (Z), the population of a given voivodeship (L) and the number of people injured in occupational accidents in the construction industry (P).

"Construction and assembly production" [8] should be understood as all kinds of activities carried out in a country, including construction works of an investment nature, renovation and maintenance, as well as temporary construction on construction sites. Construction works are carried out on order with the use of the construction entities' own forces, and the result of such works are building structures and engineering facilities of various functions.

"The population of a voivodeship" means "the total number of people registered for permanent residence in a given administrative unit and residing there, as well as people temporarily staying and registered in that administrative unit for a temporary stay" [10].

"Employees in the construction industry" can be understood as the working people whose actual workplace and type of activity is related to the construction industry. Statistical data concerning employees published by the Central Statistical Office (CSO) contains information about people who perform work that brings them earnings or income, which include, among

others: people employed on the basis of employment contracts, employers and self-employed people [8].

The number of people injured in occupational accidents in the construction industry in a given year and in a specific voivodeship is determined by the Central Statistical Office [8] on the basis of statistical occupational accident cards. It covers all occupational accidents and also accidents treated on an equal footing with them, in particular: fatal accidents, severe accidents and those with lighter health consequences.

2.2 Stage 2 - Acquiring statistical data

The numerical data assumed for the calculations came from statistical yearbooks that were published for each voivodeship by the Central Statistical Office [8, 9]. The research covered the years from 2008 to 2016. A fragment of the data accepted for the calculations for the Lower Silesian voivodeship is presented in Table 1.

Table 1. A fragment of statistical data assumed in the calculations for the Lower Silesia voivodeship.

No.	Factor		Year				
			2008	...	2012	...	2016
1	<i>W</i> – value of construction and assembly production	PLN mln	14568.1	...	13619.3	...	13721.2
2	<i>Z</i> – number of people working in the construction industry	People	68874	...	64181	...	65039
3	<i>L</i> – population		2877059	...	2914362	...	2903710
4	<i>P</i> – people injured in occupational accidents in the construction industry		892	...	679	...	413

The obtained statistical data was used to classify voivodeships using cluster analysis. Cluster analysis is a discipline in multidimensional mathematical statistics that contains methods for the separation of homogeneous subsets of objects from the population [11]. The creation of subsets, otherwise known as clusters, is based on the analysis of variables that characterize objects. In the analysed case, these objects are voivodeships. Each object is characterized by specific indicators (coordinates in multidimensional space), which are created on the basis of collected statistical data.

2.3 Stage 3 - Development of indicators that characterize the construction industry

The following indicators characterizing individual voivodeships were defined on the basis of statistical data regarding the value of construction and assembly production, the number of people employed in construction production, the population of a given voivodeship and the number of people injured in occupational accidents in the construction industry:

- Frequency of accidents [$I_{1,v,y}$] - defining the number of people injured in occupational accidents in the construction industry in relation to the number of employees. It is calculated according to the following formula:

$$I_{1,v,y} = P_{v,y} / Z_{v,y} ; \quad \text{where: } v = 1, \dots, 16; \quad y = 1, \dots, 9 \tag{1}$$

where:

$P_{v,y}$ – people injured in occupational accidents in the construction industry in voivodeship v and in year y ,

$Z_{v,y}$ – number of people working in the construction industry in voivodeship v and in year y .

- Size of the production [$I_{2,v,y}$] - corresponding to the value of construction and assembly production per one person employed in the construction industry. It is calculated according to the following formula:

$$I_{2,v,y} = W_{v,y} / Z_{v,y} ; \quad \text{where: } v = 1, \dots, 16; \quad y = 1, \dots, 9 \quad (2)$$

where:

$W_{v,y}$ – the achieved value of construction and assembly production in voivodeship v and in year y .

- Employment density in the construction industry [$I_{3,v,y}$] - corresponding to the number of inhabitants per one person employed in the construction industry. It is calculated according to the following formula:

$$I_{3,v,y} = L_{v,y} / Z_{v,y} ; \quad \text{where: } v = 1, \dots, 16; \quad y = 1, \dots, 9 \quad (3)$$

where:

$L_{v,y}$ – the population inhabiting voivodeship v in year y .

Each of the above-mentioned indicators were designated for all 16 voivodeships for each year from the analysed period.

2.4 Stage 4 - Mathematical model of the classification of voivodeships

All voivodeships constitute a set of objects $O = \{o_v; v = 1, \dots, 16\}$. In addition, each of them is described by designated indicators:

$$O_v = [I_{1,v,y}, I_{2,v,y}, I_{3,v,y}] ; \quad \text{where: } v = 1, \dots, 16; \quad y = 1, \dots, 9 \quad (4)$$

where:

$I_{1,v,y}, I_{2,v,y}, I_{3,v,y}$ – the determined indicators, which characterize: frequency of accidents, production size and density of employment in the construction industry in voivodeship v and in year y .

Data concerning all voivodeships from the period from 2008 to 2016, which were adopted for the cluster analysis, can be presented in the form of a matrix - Figure 2.

$$O = \begin{bmatrix} I_{1,1,1} & \dots & I_{1,1,y} & \dots & I_{1,1,9} & I_{2,1,1} & \dots & I_{2,1,y} & \dots & I_{2,1,9} & I_{3,1,1} & \dots & I_{3,1,y} & \dots & I_{3,1,9} \\ I_{1,2,1} & \dots & I_{1,2,y} & \dots & I_{1,2,9} & I_{2,2,1} & \dots & I_{2,2,y} & \dots & I_{2,2,9} & I_{3,2,1} & \dots & I_{3,2,y} & \dots & I_{3,2,9} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{1,v,1} & \dots & I_{1,v,y} & \dots & I_{1,v,9} & I_{2,v,1} & \dots & I_{2,v,y} & \dots & I_{2,v,9} & I_{3,v,1} & \dots & I_{3,v,y} & \dots & I_{3,v,9} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{1,16,1} & \dots & I_{1,16,y} & \dots & I_{1,16,9} & I_{2,16,1} & \dots & I_{2,16,y} & \dots & I_{2,16,9} & I_{3,16,1} & \dots & I_{3,16,y} & \dots & I_{3,16,9} \end{bmatrix}$$

Fig. 2. Matrix of analysed data.

Objects that are most "similar" to each other are assigned to a given cluster. The function that is inverse to the measure of "similarity" and thus the function of "non similarity" of the objects, which is a measure of the distance between them, is used for practical considerations. This means that if the distance of object O_j to object O_k is greater than the distance between

objects O_j and O_l : $d(O_j, O_k) > d(O_j, O_l)$, where $j \neq k \neq l$ and $j, k, l \in \{V\}$, then object O_j is more "non similar" to object O_k than to object O_l . Consequently, this leads to a situation where objects O_j and O_l can create a cluster because they are more "similar". In cluster analysis, different measures of this distance are used. Geometric distance in multidimensional space, in other words the Euclidean distance [12], is most often used. The general formula of the Euclidean distance for the analysed data takes the following form:

$$d(O_j, O_k) = [\sum_i \sum_y (I_{i,j,y} - I_{i,k,y})^2]^{1/2}; \quad \text{where: } i = 1, 2, 3; \quad y = 1, \dots, 9 \quad (5)$$

where:

O_j, O_k – the assessed objects (voivodeship j and k , where $j \neq k$ and $j, k \in \{V\}$),

$I_{i,j,y}, I_{i,k,y}$ – the determined values of indicators i in year y for the voivodeship j and k , where $j \neq k$, and $j, k \in \{V\}$).

The aim of the cluster analysis is to group objects in such a way that they form a certain number of clusters [11]. Various methods can be used for grouping objects, among others: the hierarchical method, in which two grouping techniques [11] can be distinguished: agglomeration and division. The effect of these methods is a tree-shaped graph called a dendrogram - Figure 3.

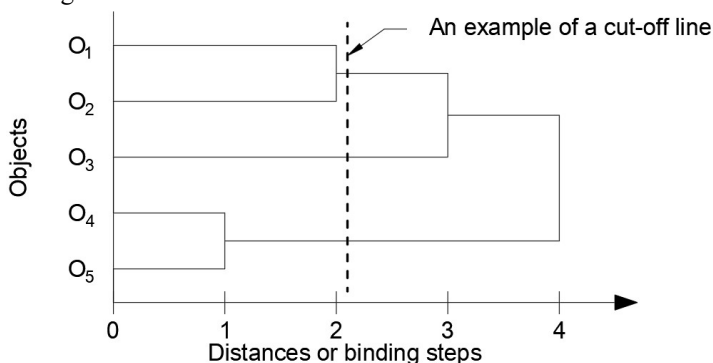


Fig. 3. Dendrogram - a graph resulting from the use of the hierarchical cluster analysis method.

The agglomeration technique, which was used in the research, involves a gradual connection of objects that constitute separate clusters into new clusters until all of the objects form one cluster. In this case, each merging of two clusters is called a step. An important issue when determining the appropriate distance between clusters, in addition to the selection of the above described distance measure is to determine the method of binding objects.

There are many binding methods, among them the method of weighted centres of gravity, in which the distance between two clusters is the distance between the centres of gravity of clusters while taking into account their size i.e. the number of objects. Figure 4 shows an example of the application of this method, where L is the distance between the weighted centres of gravity of clusters 1 and 2, and K is the distance between the weighted centres of gravity of clusters No. 3 (which is a result of merging of clusters No. 1 and 2) and No. 4.

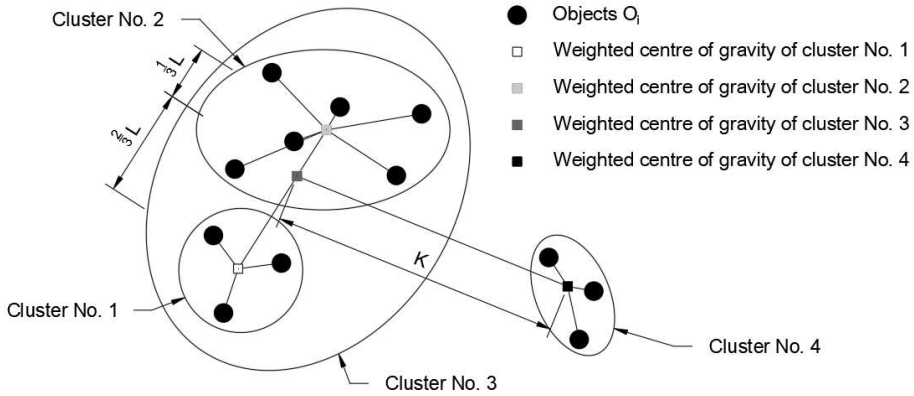


Fig. 4. The binding method - the centre of gravity method.

The result of cluster analysis is the aforementioned dendrogram. It is a graph where at the beginning each object creates a separate cluster. These objects connect with each other as the distance between them increases, and they create at the very end of the graph one cluster. The dendrogram illustrates in which step the objects connect with each other, but it does not give an unambiguous answer as to the right number, which depends on the place of the cut-off in the graph. Therefore, an important task is to properly determine the cut-off place. According to [11], however, there is no objective rule of determining the cut-off place. There are only supporting methods and among them are: the method of graphical dendrogram analysis, which involves the examining of the distance between successive bonds; the method that uses the Grabinski meter [13]; the Mojena rule [14] and others.

3 Application of the developed methodology

The hierarchical method of cluster analysis using the agglomeration technique was used to group voivodeships. The binding of objects was carried out using the method of weighted gravity centres, and the distance between them was determined as the Euclidean distance. Table 2 shows the values of parameters that characterize the selected voivodeships.

Table 2. Values of parameters for the selected voivodeships (objects).

No.	Voivodeships (objects)	Parameters								
		$I_{1,v,y}$			$I_{2,v,y}$			$I_{3,v,y}$		
		2008	...	2016	2008	...	2016	2008	...	2016
1	Dolnoslaskie	12.951		6.350	0.212		0.211	41.773		44.646
2	Kujawsko-pomorskie	13.811		6.268	0.120		0.145	47.049		48.556
3	Lubelskie	11.157	...	6.469	0.148	...	0.165	59.701	...	60.532
...
16	Zachodniopomorskie	9.214		4.058	0.186		0.257	43.091		48.141

The graph, obtained after calculations using the Statistica program, is shown in Figure 5.

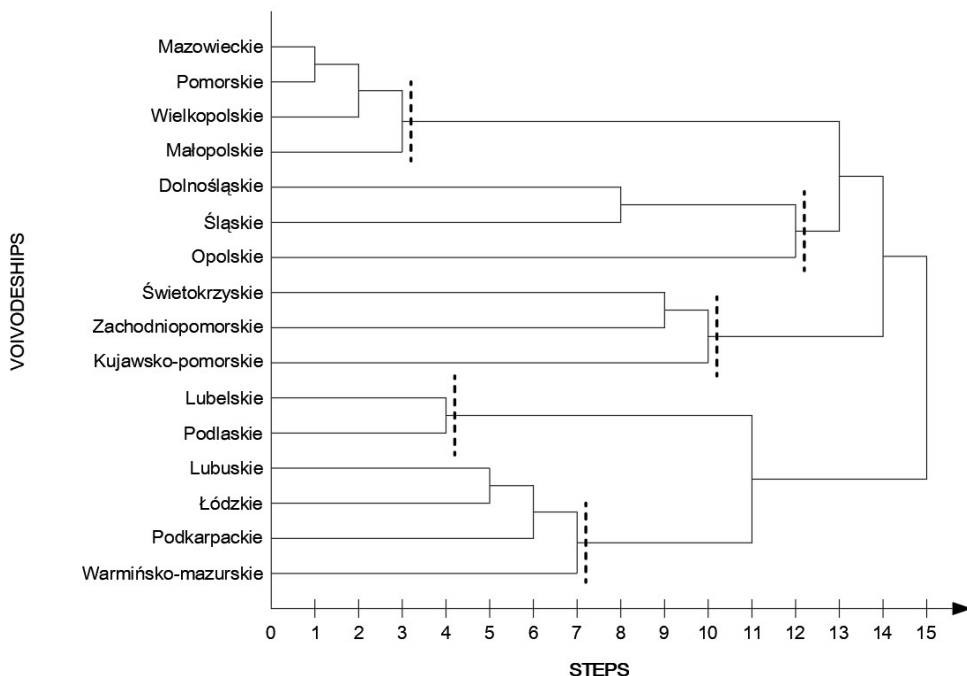


Fig. 5. Dendrogram - a graph showing the connection of individual voivodeships in steps.

When analysing the obtained results presented in Figure 5, it can be seen that some voivodeships very quickly combine into clusters, e.g. Mazowieckie, Pomorskie, Wielkopolskie and Malopolskie. Other voivodeships, however, such as Kujawsko-pomorskie and Opolskie are connected in step 10 and 12. Therefore, it was decided to connect the voivodeships in such a way that in one cluster there were at least two voivodeships. This is indicated in Figure 5 by the cut-off line of the dendrogram branch. In this way, five clusters were obtained, in which each voivodeship has a similar level of occupational safety in the construction industry - Table 4.

Table 4. Obtained clusters

Cluster number	Voivodeship
1	Lubelskie, Podlaskie
2	Warmińsko-mazurskie, Lubuskie, Łódzkie, Podkarpackie
3	Dolnośląskie, Opolskie, Śląskie
4	Kujawsko-pomorskie, Świętokrzyskie, Zachodniopomorskie
5	Mazowieckie, Pomorskie, Wielkopolskie, Malopolskie

4 Summary

The aim of the research was to classify Polish voivodeships into appropriate groups, which are characterized by a similar level of occupational safety in the construction industry. Based on a literature review of the subject, a number of factors that characterize the construction industry in Poland were identified. Four factors were then selected for further analysis. Due to the purpose and nature of the research, indicators were created on the basis of the selected factors and describe the voivodeships regarding:

- the frequency of accidents with regards to the number of people employed in the construction industry in a specific voivodeship,
- the value of construction and assembly production per one person employed in the construction industry,
- employment density in the construction industry.

These parameters were used to describe voivodeships in the applied cluster analysis. As a result of the calculations, a division of voivodeships into 5 groups was obtained. In the future, this classification will be used to create mathematical models for conducting comparative analyses between particular groups and also for predicting parameters characterizing various aspects that describe the accident rate in the construction industry.

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.

References

1. B. Hoła, T. Nowobilski, J. Rudy, E. Borowa-Błazik, *Technical Transactions*, **7**, 31-39 (2017)
2. T. Stępień, *Technical Transactions*, **1-B**, 265-281 (2014)
3. B. Hoła, T. Nowobilski, I. Szer, J. Szer, *Procedia Engineering*, **11** (2017)
4. A. Hoła, B. Hoła, M. Szóstak, *IOP Conference Series - Materials Science and Engineering*, **251**, 1-8 (2017)
5. B. Hoła, *Archives of Civil and Mechanical Engineering*, **15**, 419-426 (2015)
6. B. Hoła, M. Sawicki, M. Skibniewski, *Journal of Civil Engineering and Management*, **21**, 1014-1026 (2015)
7. B. Hoła, M. Szóstak, *Archives of Civil Engineering*, **61**, 19-34 (2015)
8. Central Statistical Office. *Statistical yearbook 2008-2016* (Warsaw 2009-2017)
9. Central Statistical Office. *Occupational accidents 2008-2016* (Warsaw 2009-2017)
10. <http://stat.gov.pl/metainformacje/slownik-pojec/pojecia-stosowane-w-statystyce-publicznej/184,pojecie.html?pdf=1> (access: 01.2018)
11. A. Stanisław, *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, Tom 3: Modele liniowe i nieliniowe* (Kraków, 2007)
12. S. Wierzchoń, M. Kłopotek, *Algorytmy analizy skupień* (WNT, Warszawa, 2015)
13. T. Grabiński, *Metody taksonometrii* (Akademia Ekonomiczna w Krakowie, Kraków, 1992)
14. R. Mojena, *Computer Journal*, **20**, 359-363 (1977)