

PAPER • OPEN ACCESS

The application of cluster analysis to identify the occupational profile of people injured in accidents in the Polish construction industry

To cite this article: M Szóstak 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **456** 012027

View the [article online](#) for updates and enhancements.

You may also like

- [Data Transmission Security in Computer Network Communication](#)
Fei Chen
- [University Ideological and Political Education Management Based on K-means Mean Value Algorithm](#)
Jiafeng Ma
- [Clustering analysis on inbound tourism market in Shandong province](#)
Tong Yu



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

The application of cluster analysis to identify the occupational profile of people injured in accidents in the Polish construction industry

M Szóstak

Wrocław University of Science and Technology 27 Wybrzeże Wyspiańskiego st.
50-370 Wrocław, Poland

mariusz.szostak@pwr.edu.pl

Abstract. Cluster analysis is a tool used to analyze data and to group similar data into homogeneous groups called clusters. The aim of cluster analysis is to detect "natural" clusters - groups that can be interpreted in a sensible way - in a set of analyzed data. The objective of the research is to present the occupational characteristics of people injured in occupational accidents in the Polish construction industry using cluster analysis. The article presents the results obtained from the analysis of occupational accidents, which occurred in the years 2008-2014 in five Polish voivodeships: Dolnoslaskie, Kujawsko-Pomorskie, Lubelskie, Lubuskie and Slaskie. The impact of the following features was identified and examined in the paper: the status of the victim's employment, his occupation, age, work experience, and also the preparation of the employee to perform duties at a workplace. 361 occupational accidents were analyzed. On the basis of the conducted cluster analysis, a professional profile of people who are most often injured in occupational accidents in the Polish construction industry was developed.

1. Introduction

The construction industry in Poland, as well as in the world, is characterized by a high level of threat to the life and health of employees. Dangerous events lead to occupational accidents, which may result in material damage, injuries of severity varying degrees, and even the death of an employee. The phenomenon of employees being injured in accidents during work is defined as the accident rate. The accident rate can be considered and analyzed in both a detailed and general sense. In the general sense, the accident rate is defined as the sum of accidents that occurred at some time, usually during a year, and is presented by means of various indicators, e.g. the accident frequency ratio, the accident severity index. On the other hand, investigation of the accident rate within a detailed approach means that individual accidents are assessed.

Based on a thorough analysis of the subject literature, six main research areas related to the issue of the accident rate in the construction industry were identified, namely: analysis of the mechanisms of the occupational accidents occurrence and the formation of accident models [1-3], analysis of the causes of occupational accidents [4-8], identification of direct factors that affect the accident rate and occur on a construction site or in its close vicinity [9, 10], identification of indirect factors that affect the accident rate and which are related to construction companies and their surroundings [11, 12], analysis of legal regulations [13, 14], analysis of the accident rate with regards to factors that generate costs and material losses [15-18].



However, it should be remained, that the most important, according to occupational safety, is the man who in the accident process has a triple role, namely: the decision maker, the perpetrator of the accident and, above all, the victim. The man is a participant in the investment process and performs one of the following roles: investor, contractor, subcontractor, supplier of equipment and materials, construction supervision inspector, designer, etc. In addition, the construction industry is characterized by a very high variability of working conditions. As a result, employees in this section of the economy are exposed to many hazardous factors that cause accidents. Very often the source of these factors is a human being, and it is therefore important to conduct research and analyzes that will allow the occupational profile of the employee who is the most often injured in occupational accidents in the construction industry to be determined. Based on the analysis of the subject literature, it was found that modern analyzes related to a construction employee have involved, among others: the determination of personal factors and causes of occupational accidents regarding electric works [19], the analysis of occupational and individual characteristics of people injured in fatal accidents [20], the determination of the people occupational profile who are the most often injured in occupational accidents in the construction industry [21], the establishment of relationships between the selected factors that characterize people injured on scaffolding and the physiological parameters that change with age [22], the external factors impact on the productivity of employees [23], the determination of the relationships between the age of an injured person and which height from a fall occurred and resulted in the death of the injured person [24].

As it can be concluded from the literature review, up until now no professional profile of people injured in occupational accidents, which would take into account the existing relationships between individual characteristics of a construction worker, such as: employment status, occupation, work experience, age, and also preparation of an employee to work, has been developed.

The scientific research selected in the literature review was carried out using various research tools and techniques. The author's observation shows that multidimensional statistical analyzes, including cluster analysis, are now of great interest. Cluster analysis, as a tool for exploratory data analysis, involves segmenting a set of data into subsets in order to extract homogeneous objects in the analyzed set [25]. The method of grouping data is used in many scientific disciplines, including: mining and geology (e.g. in order to search for geothermal resources with high and low potential [26]), medicine, (e.g. for grouping and comparing medicines [27]), meteorology (e.g. in order to assess the local impact of predicted climate change [28]), civil engineering (e.g. for forecasting the work productivity of brickwork contractors [29], the analysis of risk factors for the occurrence of occupational accidents and potential accidental events at workplaces that use construction cranes [30], as well as for the analysis of Polish regions with regards to occupational safety [31]).

The aim of the article is to develop people professional profiles, who are the most often injured in occupational accidents in the Polish construction industry using cluster analysis.

2. Research methodology

The National Labor Inspectorate is the basic institution for the supervision and control of complying with the labor law in Poland. Institutions of the National Labor Inspectorate are obliged to investigate the circumstances and causes of fatal, heavy and collective accidents. After a post-accident investigation, a labor inspector draws up the post-accident protocol, which includes any conclusions from the investigation. The post-accident protocol contains information on the circumstances and causes of the examined event, in particular: information on the time and place of the incident, performed activities, used materials and equipment, data on the injured people, and also a description of the event course that takes into account its circumstances and causes. The protocols contain the following data that characterize an injured person: sex, year of birth, citizenship, employment status in a legal sense, the performed occupation, experience at the occupied workplace in the enterprise in which the accident took place, and also preparation to work. In addition, based on the analysis of data contained in documentation, it is possible to determine the number of hours worked by an injured person, which is counted from the time of starting work to the moment of the accident.

On the basis of collected documents regarding occupational accidents in the construction industry that occurred in five voivodeships in the years 2008-2014, which were made available by District Labor Inspectorates in Poland, a database of occupational accidents was created. The study covered 361 occupational accidents that took place in the Dolnoslaskie, Kujawsko-Pomorskie, Lubelskie, Lubuskie and Slaskievoivodeships.

In order to determine the professional profile of a person who is most likely to be injured in an accident in the construction industry, the following individual characteristics were assigned to each person $o_l (l = 1, \dots, L)$ who is injured in an occupational accident: employment status in a legal sense (v_l), the profession performed by the injured person (z_l), work experience (s_l), the age of the injured person (y_l), the preparation of an employee to perform duties at a workplace, which includes: initial training, i.e. general ($p_{1,l}$) and on-the-job training ($p_{2,l}$), periodic training ($p_{3,l}$), and medical examinations ($p_{4,l}$).

The vector of general characteristics of each employee can be represented by the following formula:

$$o_l = [v_l, z_l, s_l, y_l, p_{1,l}, p_{2,l}, p_{3,l}, p_{4,l}], l = 1, \dots, L, \quad (1)$$

These features, in the process of a large number of injured people analysed, will allow a professional profile of an employee who is the most often injured in occupational accidents while working in the construction industry to be created. Collected information on all people injured in occupational accidents can be presented in the form of a data matrix O .

$$O = \begin{bmatrix} v_1 & z_1 & s_1 & y_1 & p_{1,1} & p_{2,1} & p_{3,1} & p_{4,1} \\ v_2 & z_2 & s_2 & y_2 & p_{1,2} & p_{2,2} & p_{3,2} & p_{4,2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ v_l & z_l & s_l & y_l & p_{1,l} & p_{2,l} & p_{3,l} & p_{4,l} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ v_L & z_L & s_L & y_L & p_{1,L} & p_{2,L} & p_{3,L} & p_{4,L} \end{bmatrix} \quad (2)$$

Each row of matrix O includes data concerning one injured person, whereas each column of the matrix contains a set of data concerning one particular feature, which is common to all employees injured in accidents. Each of the mentioned features ($v_l, z_l, s_l, y_l, p_{1,l}, p_{2,l}, p_{3,l}, p_{4,l}$) which describe a construction employee, takes a certain numerical value from an appropriate interval, or it is described in words. Therefore, the qualitative variables concern: the employment form of an injured person (v_l), occupation (z_l), and employee preparation to work ($p_{1,l} - p_{4,l}$), while the quantitative variables are the work experience (s_l) and the age of the injured person (y_l).

Cluster analysis was used to define the occupational profile of an employee who is the most frequently injured in occupational accidents in the construction industry. The use of this method allows a group of objects to be found in a large set of objects, each of which is described by a set of attributes that are similar to each other [32]. The idea behind the method is to divide the set of objects into several groups, which objects have similar features.

There are many algorithms for dividing objects into clusters. Hierarchical and non-hierarchical methods are usually used in analyses [33]. The article presents the objects grouping with the use of the k-means method, which is one of the non-hierarchical methods. In this method, clusters are obtained as a result of division, and none of them is a sub-cluster of another cluster. In the k-averages method, the number of clusters is assumed, or calculated using the v-cross-check algorithm. This algorithm transfers objects to different clusters while both minimize the variability aiming inside the clusters and also maximize the cluster variation between clusters.

In the conducted analysis, every person who is injured in an occupational accident is an object $o_l; (l = 1, \dots, L)$, which is described by 8 variables ($v_l, z_l, s_l, y_l, p_{1,l}, p_{2,l}, p_{3,l}, p_{4,l}$). In the analysed phenomenon, there is an 8-dimensional space (R^8), which objects o_l are located. It should be noted that the quantitative variables in the conducted research are determined by means of different quantities. For example, variable s_l is work experience of the l -th person injured in an occupational accident is determined by the number of days, while y_l is the age of the l -th person injured in an occupational accident is determined by the number of years. In order to be able to compare the

quantitative variables, the data should be transformed into comparable values. For this purpose, the variables must be standardized.

The aim of the analysis is to divide the set of all objects (people injured in accidents) into clusters (groups). Individual clusters will include objects that are close together in 8-dimensional space (R^8). Individual clusters ($S_1, S_2, \dots, S_k$) are disjointed sets, which common part is an empty set of ($S_1 \cap S_2 \cap \dots \cap S_k = \emptyset$).

In order to determine the similarity between pairs of objects, the measure of dissimilarity (or similarity) was used. The analyses assume that the smaller the value of dissimilarity, the more similar objects is compared to each other. In the analysed case, each object has a feature vector assigned to it, which components for quantitative variables make numbers, whereas for qualitative variables, they are interpreted as labels. For example, if an interesting variable is information about an injured person being subjected to general training (qualitative variable), then it may take the following values: 0 - lack of information, 1 - if the victim has not been subjected to compulsory training, 2 - when the victim was trained.

In the case of mixed data, i.e. when part of the variables is qualitative, and part is quantitative, and such a case occurs in the conducted analysis, the most popular and used measure of difference and similarity are Gower coefficients [34]. The similarity measure proposed by Gower for objects with quantitative and qualitative variables is a weighted sum of difference partial coefficients $\delta(l, l+1)$, which are determined for each variable in the o_l and o_{l+1} , objects.

3. Analysis of results

Cluster analysis was carried out on a set of 361 people injured in occupational accidents in the Polish construction industry. As a result of the conducted calculations, 3 clusters were obtained. Figures 1-3 present the obtained diagrams of the qualitative cardinality variables in the individual clusters. The density of the probability for the variable - the age of an injured person - in individual clusters is shown in Figure 4, while Figure 5 shows the density of the probability for the quantitative variable - work experience.

Figure 1 presents a diagram of the qualitative variable cardinality is employment status, taking into account the division of the analysed data set into 3 clusters.

Analysis of employment status showed that:

- in cluster 1, the largest group of employees injured in occupational accidents are full-time employees hired for a definite period. This number amounts to 27 people,
- in cluster 2, the largest group of employees injured in occupational accidents in the construction industry are full-time employees hired for a definite period. This number amounts to 109 people,
- in cluster 3, the largest group of people injured in occupational accidents are full-time employees hired for an indefinite period. This number amounts to 89 people.

Moreover figure 1 presents a diagram of the qualitative variable cardinality is the profession performed by the injured person, taking into account the division of the analysed data set into 3 clusters. A profession analysis performed by injured people showed the most common occupation in all the clusters were construction workers performing shell and core works.

Figure 2 presents diagrams of the qualitative variables cardinality with regards to initial training (general and on-the-job training), which takes into account the division of the analysed data set into 3 clusters.

Completion analysis of the compulsory initial training by injured people showed that:

- in cluster 1, the largest group are people who have not been subjected to compulsory initial training in the form of general and on-the-job training. This number amounts to 70 and 69 people, respectively.
- in clusters 2 and 3, the most numerous group are people who have undergone compulsory initial training, both in the field of general and on-the-job training. This number in cluster 2 is equal to 167 and 160 people, respectively, while in cluster 3, to 106 and 99 people, respectively.

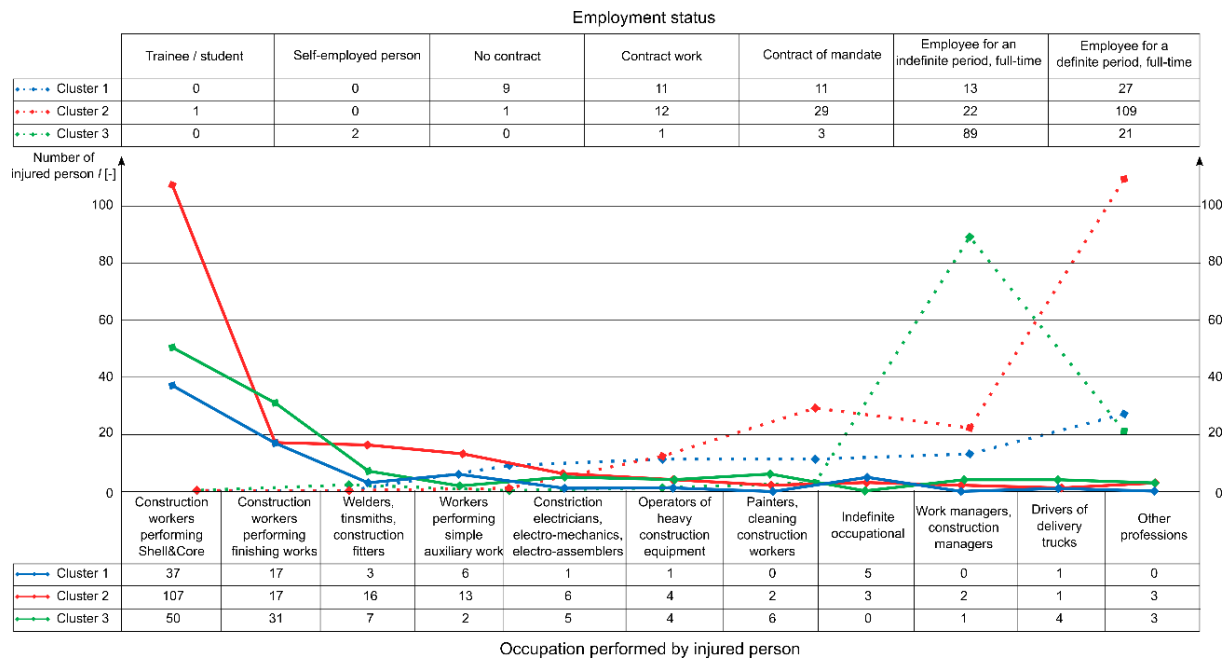


Figure 1. Diagram of the cardinality of the qualitative variable: employment status v_l and the occupation performed by an injured person z_l (own work).

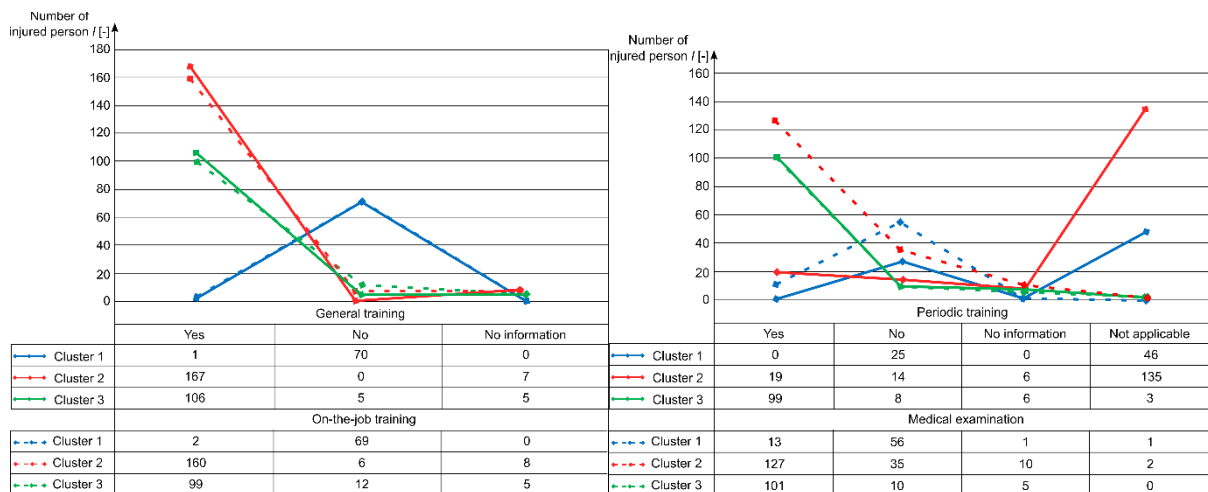


Figure 2. Diagram of the cardinality of the qualitative variable: general training $p_{1,l}$ and on-the-job training $p_{2,l}$ (own work).

Figure 3. Diagram of the cardinality of the qualitative variable: periodic training $p_{3,l}$ and medical examination $p_{4,l}$ (own work).

Figure 3 presents the diagrams of the qualitative variables cardinality with regards to periodic training and medical examinations, which takes into account the division of the analyzed data set into 3 clusters.

The analysis of the periodic training completion by injured people showed that:

- in clusters 1 and 2, the most numerous group are injured people who did not have to undergo periodic training, i.e. their work experience was shorter than one year. This number in cluster 1 amounts to 46, and in cluster 2 to 135 people.

- in cluster 3, the largest group are people who have undergone periodic training, which means that the length of service of people injured in this group was longer than 1 year. This number amounts to 99 people.

The analysis of numerical data regarding medical examinations that employees should be subjected to at each workplace showed that:

- in cluster 1, the largest group are people who did not undergo mandatory medical examinations before they were allowed to work. This number amounts to 56 people.
- in clusters 2 and 3, the most numerous group are people who were correctly admitted to work, i.e. they obtained a medical certificate regarding the absence of contraindications to perform work. This number in cluster 2 amounts to 127 people, while in cluster 3, to 101 people.

Figure 4 presents the probability density function of the quantitative variable, which is the age of an injured person. Figure 5 presents the probability density function of the quantitative variable, which is work experience of an injured person.

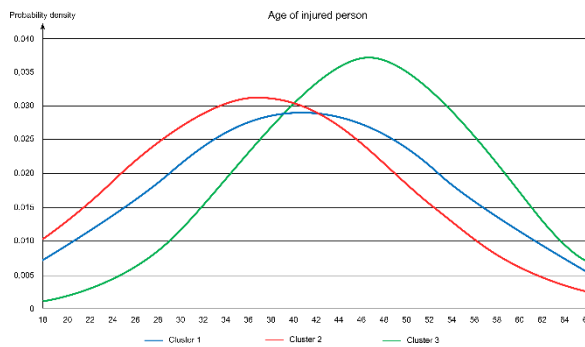


Figure 4. Probability density function of the quantitative variable: age of an injured person y_l (own work).

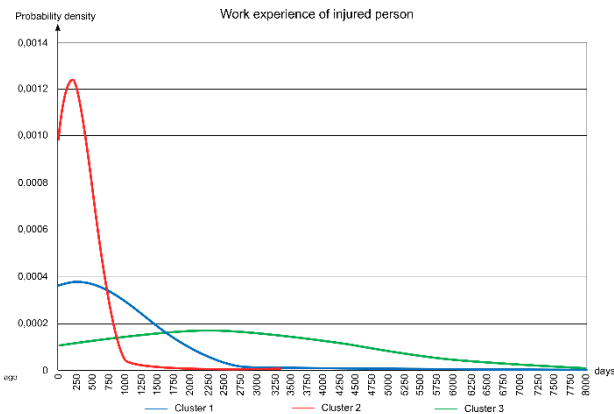


Figure 5. Probability density function of the quantitative variable: work experience of an injured person s_l (own work).

Analysis of the age of an injured person showed that average age and median value is equal to:

- in cluster 1: $\bar{y}_l^1 = 41$ years (40-49); median $M_{y_l}^1 = 42$ years,
- in cluster 2: $\bar{y}_l^2 = 37$ years (30-39); median $M_{y_l}^2 = 34$ years,
- in cluster 3: $\bar{y}_l^3 = 46$ years (40-49); median $M_{y_l}^3 = 50$ years.

Analysis of an employee's work experience showed that the average length of experience and the median is equal to:

- in cluster 1: $\bar{s}_l^1 = 283$ days (1 year and less); median $M_{s_l}^1 = 8$ days,
- in cluster 2: $\bar{s}_l^2 = 217$ days (1 year and less); median $M_{s_l}^2 = 90$ days,
- in cluster 3: $\bar{s}_l^3 = 2318$ days (over 6 years); median $M_{s_l}^3 = 1460$ days (approximately 4 years).

4. Conclusion

The analysis of people injured in occupational accidents in the Polish construction industry, which was carried out with the use of cluster analysis, allowed the following conclusions to be formulated:

- as the analysis result of a set of 361 people injured in occupational accidents in the construction industry, three clusters were obtained,
- cluster 1 includes employees hired as a construction worker of shell and core works for a definite period of time and who have not been properly prepared to perform work, i.e. they have not been properly trained in the area of general and on-the-job training, and have also not

obtained a certificate of no contraindications to work. The average age of injured people varies between 40 and 49 years. Employees included in this cluster had work experience of less than one year. 71 injured people were found in this cluster, which is 20% of all the analyzed injured people,

- cluster 2 includes employees hired as a construction worker of unfinished works for a definite period of time and who have completed initial training in the field of occupational safety and received a certificate of no contraindications to work. The average age of injured people varies between 30 and 39 years. Employees included in this cluster had work experience of less than one year. 174 injured people were found in this cluster, which is 48% of all injured people,
- cluster 3 consists of employees hired for an indefinite period as a construction worker of unfinished works, who have undergone initial training in the field of occupational safety and periodic training and also received a certificate of no contraindications to perform work. The average age of injured people varies between 40 and 49 years. Employees included in this cluster had work experience of more than 4 years. 114 victims were found in this cluster, which constitutes 32% of the total number of injured people.
- in all three clusters, the profession performed by the injured person is a construction worker performing shell and core works,
- individual clusters differ with regards to the preparation of an employee to perform work. The opposite to cluster 2 (the largest of the distinguished) is cluster 1. In this cluster, injured people were not prepared to work in any way. A lack of proper training, a lack of sufficient knowledge about safe work at the workplace, as well as short work experience were the main causes of an occupational accident,
- injured people assigned to cluster 1 and 2 have much less work experience than employees from cluster 3.

The occupational profile of people who are most frequently injured in occupational accidents in the Polish construction industry indicates the main personal characteristics, which significantly contribute to an employee being injured in an occupational accident. It should be noted, however, that the occupational profile obtained in the conducted analyzes is characteristic for people employed as a construction worker of shell and core works. In order to learn about the occupational profile of people who are injured in accidents at other workplaces, e.g. those related to finishing works (construction workers) or electrical works (electricians, electro-mechanics), a similar analysis should be carried out using the proposed research methodology in the data set of interest.

The obtained professional profile will allow appropriate preventive actions, which aim to improve work safety, to be formulated. It should be emphasized that all construction workers should be the recipients of these activities, and not only those who have been characterized by particular clusters. Other participants of the investment process who are also exposed to hazards and may suffer from accidents while their activities performing at a construction site should not be forgotten.

Where the author wishes to divide the paper into sections the formatting shown in table 2 should be used.

References

- [1] Hoła B and Szóstak M 2017 *Technical Transactions, B, Civil Engineering* **114** 4 81–90
- [2] Bellamy L J, Ale B J M, Whiston J Y, Mud M L, Baksteen H, Hale A R, Papazoglou I A, Bloemhoff A, Damen M and Oh J I H I 2008 *Safety Science* **46** 2 186–97
- [3] Li R Y M, Ho D C W and Tang B 2018 *Advances in Intelligent Systems and Computing* 67–76
- [4] Hoła A, Sawicki M and Szóstak M 2018 *Applied Sciences* **8** 1 48
- [5] Vasconcelos B and Barkokébas J B 2015 *Procedia Manufacturing* **3** 4392–9
- [6] Salguero-Caparros F, Suarez-Cebador M and Rubio-Romero J C 2015 *Safety Science* **72** 329–36
- [7] Raviv G, Shapira A and Fishbain B 2016 *Safety Science* **91** 298–309
- [8] Carrillo-Castrillo J A 2017 *Int. J. Occup. Safety Ergonom* **23** 2 240–50
- [9] Drozd W 2016 *J. Civil Eng. Env. Arch.* **33** 63 165–72

- [10] Opoyemi S W, Razali A H and Mohd S M 2018 *Int. J. Built Env. Sustain* **5** 1 78–92
- [11] Hoła B, Nowobilski T, Szer I and Szer J 2017 *Procedia Engineering* **208** 35–42
- [12] Stępień T 2014 *Technical Transactions, B, Civil Engineering* **111** 1-B 265–81
- [13] Irumba R 2014 *Safety Science* **64** 109–120
- [14] Guha H and Biswas P 2013 *Int. J. Sci. & Eng. Res.* **5** 2138–43
- [15] Warhrer G M, Dong X S, Miller T, Haile E, Men Y 2007 *Accident Analysis & Prevention* **39** 6
- [16] Hoła A, Hoła B, Sawicki M and Szóstak M 2016 *MATEC Web Conferences* **86** 07005
- [17] Forteza F J, Carretero-Gómez J M and Sesé A 2017 *Safety Science* **94** 61–76
- [18] Asanka W A and Ranasinghe M 2015 Study on the impact of accidents on construction projects *Proc. 6th Int. Conf. Structural Eng. Construction Management* Kandy, Sri Lanka Vol **4** 58–66
- [19] Castillo-Ros J, Suárez-Cebador M, Rubio-Romero J C and Aguado J A 2017 *Safety Science* **91** 286–97
- [20] Villanueva V and Garcia A M 2011 *Accident Analysis & Prevention* **43** 1 123–7
- [21] Hoła B and Szóstak M 2017 *Procedia Engineering* **208** 43–51
- [22] Sawicki M and Szóstak M 2018 *MATEC Web Conferences* (in print)
- [23] Plebankiewicz E, Juszczak M and Malara J 2014 *Scientific Review – Engineering and Environment Sciences* **65** 271–8
- [24] Saiful M I, Razwanuk I and Mahmud T 2017 *Jordan J. Civil Engineering* **11** 2 267–78
- [25] Nerurkar P, Shirke A, Chandane M and Bhirud S 2018 *Procedia Computer Science* **125** 770–9
- [26] Lindsey C R, Neupane G, Spycher N, Fairley J P, Dobson P, Wood T, McLing T and Conrad M 2018 *Geothermics* **72** 358–70
- [27] Wallner F L, Hopkins M H, Woodworth N, Bark T L, Olofsson P and Tilevik A 2018 *Pharmacological Research* **128** 244–51
- [28] Falquina R and Gallardo C 2017 *Atmospheric Research* **197** 224–31
- [29] Florez L and Cortissoz J C 2017 *Procedia Engineering* **196** 359–65
- [30] Raviv G, Fishbain B and Shapira A 2016 *Safety Science* **91** 192–205
- [31] Hoła B and Nowobilski T 2018 *Sustainability* **10** 5 1637
- [32] Stanisław A 2007 An accessible course of statistics with the use of STATISTICA PL using examples from medicine *Multivariate Analysis* StatSoft **3**
- [33] Krzyśko M, Wołyński W, Górecki T and Skorzybut M 2008 *Learning Systems* (Warsaw: Wydawnictwo Naukowo-Techniczne)
- [34] Wierzchoń S and Kłopotek M 2015 *Cluster Analysis Algorithms* (Warsaw: Wydawnictwo Naukowo-Techniczne)