

# CLUSTER ANALYSIS IN RESEARCH OF ACCIDENT RATE IN CONSTRUCTION SECTOR

**W. DROZD<sup>1</sup>**

The construction site and its elements create circumstances that are conducive to the formation of risks to work safety during the execution of works. Analysis indicates the critical importance of these factors in the set of characteristics that describe the causes of accidents in the construction industry. This paper attempts to analyse the characteristics of the construction site to indicate their importance in defining the circumstances of an accident at work. The research was carried out on the basis of data from the register kept by the District Labour Inspectorate in Krakow, Poland. Main substantive tasks include isolating patterns of accidents on site and identifying those of the analysed characteristics that are important in defining these patterns. In terms of methodology, the paper presents a method of analysing data resources by using means of conceptual grouping in the form of cluster analysis.

*Keywords:* construction site, work safety, accident at work, cluster analysis

## 1. INTRODUCTION

The Accidents at construction site are random phenomena that are difficult or impossible to foresee. Therefore, their investigation and determination of links between their characteristic features is not easy [1, 2, 3]. This is particularly true for analyses of factors which generate undesirable effects for human life and health.

Up to present, few studies are available concerning the safety of building work. Among them: in paper [4] described the preconditions and components of high safety standards in the construction industry from the perspective of both experienced construction workers and first-line managers.

---

<sup>1</sup> PhD., Eng., Institute of Construction Management, Tadeusz Kościuszko Cracow University of Technology,  
Warszawska 24 St., 31-155 Kraków, Poland., e-mail: wdrozd@ztob.pk.edu.pl

The article [5] presents the results of a study that identified the major factors affecting safety in tower-crane environments and evaluated the degree to which each factor influences ongoing safety on site. In [6] a behavioural observation measure of safety performance was developed. No relationship was found between safety climate and the safety performance measure. While identical safety climate factors cannot apply to all organizations, some general safety climate factors may emerge.

Work [7] discusses empirical research aimed at why construction workers engage in unsafe behavior. Interviews were conducted in Hong Kong with workers who had been accident victims. The results substantiate the significant role of management; safety procedure; psychological and economic factors; self-esteem; experience; performance pressure; job security; and education as well as safety orientation and training. The influences of these factors on the safety behavior of workers are discussed along with implications of the research for management of the construction industry.

The objective of the study in [8] was to determine the factors comprising the safety climate of the Finnish National Road Administration. The safety climate of the workers consisted of four dimensions: attitude towards safety in the organization, changes in work demands, appreciation of the work, and safety as a part of productive work. The safety climate of the supervisors consisted of four dimensions: changes in job demands, attitude towards safety within the organization, value of the work, and safety as a part of productive work. Both the workers and supervisors felt that safe work habits improve production, that the prevention of accidents is the responsibility of everyone, and that safety is a part of skilful job performance.

Paper [9] investigates local construction workers' behaviour, perception and attitude toward safety, and attempts to link the research findings to the influence of national culture. The paper also reveals that workers operating in a more collective and higher uncertainty avoidance environment, are more likely to have safety awareness and beliefs, which can exhibit safer on-site behaviour.

This paper attempts to identify and profile the accidents patterns based on the data kept by the District Labour Inspectorate in Krakow, Poland. The data concerned the period from 2014 to 2016, and the number of observations was 65. The main objective was to isolate the patterns of accidents at construction sites and to indicate those of the analysed characteristics that are important in defining these patterns. In terms of methodology, the paper presents a method of analysing data resources by using means of conceptual grouping in the form of cluster analysis. There are various known data exploration techniques, that is techniques which indicate new relationships in datasets. One of them is cluster analysis which involves searching for groups of elements forming relatively homogenous classes. The paper presents a successful attempt to use this method in investigation of accident rate in construction sector.

## 2. RESEARCH METHODOLOGY

The cluster analysis is a tool for exploratory data analysis which aims at arranging the group elements so that the degree of connection between the elements belonging to the same group is as high as possible and as low as possible with elements of other groups [10]. The cluster analysis can be used to detect structures in data without an interpretation or explanation. This means that the cluster analysis only detects structures in data, but does not explain why they occur. The clusters are determined based on certain statistical criteria, and the interpretation of result and evaluation from the practical point of view is the researcher's task.

The paper attempts to determine clusters of variables which characterise accidents at construction sites that occurred in the investigated time and place.

An important problem in such analyses is to determine the number of sought after clusters [11]. In Statistica, it is possible to define this parameter in advance, before starting calculations, and then find out the division of an "interesting", expert-based interpretation. An automatic strategy of division determination can also be applied, using the so-called  $v$ -fold cross validation. This method involves dividing the whole sample to  $v$  subsets (in practice usually 5 or 10) or random disjoint sub-samples. The same analysis is then applied to observations from  $v-1$  sets (learning samples), and the resultant model is used for subset  $v$  (a sample or a subset that was not used to determine the cluster, i.e. the test sample), and the "prognostic accuracy" is calculated (i.e. how well the observations from sample  $v$  are assigned to homogenous clusters using the current solution calculated from  $v-1$  learning samples). The results for successive  $v$  repetitions are aggregated (averaged) and give a single evaluation of the model stability, i.e. accuracy of assigning observations to clusters.

The automatic selection of the optimum number of clusters involves successive dividing into 2, 3, ... clusters. A division quality criterion is then calculated for each cluster (by means of cross validation), and the division which gives the best value is chosen.

## 3. SELECTION OF VARIABLES

The cluster analysis was performed using the "Generalized Cluster Analysis" module of the Statistica software. The data on the accidents came from the District Labour Inspectorate in Krakow and covered the period from 2014 to 2016. The "Accidents 2014 – 16" set with independent variables

was created. It is presented in Tables from 1a to 1c (shown are only 2 typical observations out of 65 in total).

Table 1a. Accidents 2014 – 16. Source: own

Lp	Sw	Wi (lata)	Stp (lata)	Ldn (dni)	Sz	Zwd	Mw
1	L	39	8	16	UPO	EXCAV.OPER	SITE
2	C	40	13	16	DG	LABOURER	SITE

Table 1b. Accidents 2014 – 16 (continued). Source: own

Pp	Cw	Cmc	Wdo	Cmo	Wyu	Cmu
RZ	PNZ	RM	POR	RM	POR	EI
TYNK	PNN	RU	UW	RU	UNO	PP

Table 1c. Accidents 2014 – 16 (continued). Source: own

PRZYCZYNA														
BK	BŚI	BŚZ	NCM	NPD	POŚ	PNO	PUZ	SPF	SZ	TN	UKM	WZ	ZN	ZO
						1	1			1		1		1
	1	1	1					1	1					1

Where:

- Mw – accident location,  
 Stp – victim seniority,  
 Wi – victim age,  
 Ldn – days on sick leave,  
 Sz – victim employment status (UPO – employment contract for a specified period, UPN – employment contract for an unspecified period, UZ – contract mandate, UD – contract for specific work, DG – self-employed),  
 Zwd – victim trade (CARPENTER-JOINER, ROOFER, ELECTRICIAN, STONE MASON, GEOLOGIST, FORKLIFT DRIVER, PAINTER, GAS PIPING FITTER, PLUMBER, BRICKLAYER-PLASTERER, EXCAVATOR OPERATOR, TOWER CRANE OPERATOR, LABOURER, WELDER, STEEL FIXER-CONCRETE PLACER),  
 Pp – work process (RZ - earthworks, TYNK - plastering, MUR-masonry works, PRZEB - reconstruction, MON -erection, DESK - formwork, DEM - disassembly, DACH – roof works, INST – rough-ins, TRANSP - transport, BET - concreting),  
 Cw – **activity performed by the victim when accident occurred**,  
 DEM – disassembly of building parts,

- KN – contact with a tool,
- PNN – work with a non-mechanized tool,
- PNZ – work with a mechanized tool,
- PODN – manual transport (lifting, lowering),
- SCH – movement (coming down, going up, going from to),
- SM – storage method,
- UK – losing control of a tool,
- Cmc** – **material factor related to the activity performed by the victim when accident occurred,**
- Wdo** – **incident which is a deviation from normal state,**
- BK – absence of required qualifications (no medical tests, no training),
- BŚI – absence of personal protective equipment (incorrect use),
- BŚZ – absence of collective protective equipment,
- NCM – incorrect condition of material factor,
- NPD – incorrect comings and goings,
- PNO – performing works ordered by a supervisor, outside the job description,
- POŚ – tripping, slipping,
- PUZ – performing works without eliminating hazards,
- SPF – incorrect psycho-physical state (under influence of alcohol, fatigue),
- SZ – wilful misconduct (absence of supervision, absence of concentration),
- TN – violations of the OHS rules being tolerated by the supervisors,
- UKM – losing control of a machine,
- WZ – entry to a hazard area,
- ZN – sudden incident,
- ZO – incorrect organisation of works,
- Cmo** – **material factor related to deviation,**
- Wyu** – **incident causing injury,**
- OE – cut,
- OMS – overstrain of the musculoskeletal system,
- POR – electric shock,
- PRZ – crushing,
- UNO – hitting immobile factors,
- UPZ – fall into a depression,
- URU – being hit or caught by moving parts,
- USP – being hit by falling factors,
- UW – fall from height,
- WY – explosion,
- ZAS – being buried in an excavation.
- Cmu** – **material factor being a source of injury,**
- DP – portable ladder,
- EB – building element,
- EI – rough-in element,
- EN – power tool,
- ŁO – shovel,
- MB – construction materials,
- OD – chips,

---

PP	– horizontal surface (ground),
RM	– mobile machinery and equipment,
RU	– scaffolding,
SP	– falling object,
SZW	– lift shaft,
ŚC	– wall,
UR	– Spoil.

The EM algorithm (Expectation Maximisation; Witten, Frank 2000) was used to determine clusters (groups of elements forming homogenous classes). This algorithm assumes that data come from the “mixing” of samples from probability distributions with different parameters and attempts to recreate such distributions so as to obtain the best fit. It can be used to numerical data, and also to variables with categories.

#### 4. RESULTS

The collected data on accidenta are both quantitative and qualitative (categorised). To obtain a higher cohesion of all results, the EM algorithm was used, indicating all instances of irregularities as categorised variables. The EM algorithm (Expectation Maximisation; Witten, Frank 2000 ) assumes that data come from the “mixing” of samples from probability distributions with different parameters and attempts to recreate such distributions so as to obtain the best fit. It can be used to numerical data, and also to variables with categories. The features which differentiate the clusters the most were examined using the chi-square statistics. The chi-square test results for the used segmentation are presented in the sheet Independence tests for random variables (Table 2). For each categorised variable, the chi-square test evaluated whether its distribution is significantly different for various clusters.

$$(1) \quad \chi^2 = \sum \frac{(O-E)^2}{E}$$

where:

E	–	expected cell density,
O	–	observed cell density.

The statistically significant results are marked in red. In addition, the chi-square value and the probability value can be treated as measures of the degree to which a variable differentiates the cluster; the greater the chi-square value/ the lesser the p, the greater the differentiation.

Using the variables Mw (accident location), Pp (work process), Cw (activity performed by the victim when accident occurred) , Cmc (material factor related to Cw), Wdo (incident which is

a deviation from normal state), Cmo (material factor related to deviation), Wyu (incident causing injury), Cmu (material factor related to Wdo) from the “Accidents 2014-16” sheet, and the personal variables from the same sheet: Sz (employment status), Wi (victim age), Stp (victim seniority), Zwd (victim trade), and the EM algorithm, the cross-validation indicated that the optimum number of clusters is 2.

Table 2. Independence tests for random variables „Accidents 2014 – 16”.

Number of clusters: 2, total number of learning cases: 65. Source: own

Variable	Chi-square <sup>2</sup>	p-value <sup>3</sup>
Sz (employment status)	3.64	0.6020
Zwd (victim trade)	31.86	0.0008
Mw (accident location)	1.76	0.1842
Pp (work process)	26.66	0.0052
Cw (activity performed by the victim when accident occurred)	15.84	0.0032
Cmc (material factor related to Cw)	59.04	0.0000
Wdo (incident which is a deviation from normal state)	41.58	0.0003
Cmo (material factor related to Cw)	54.75	0.0000
Wyu (incident causing injury)	38.50	0.0000
Cmu (material factor being a source of injury)	41.42	0.0000
BK (absence of required qualifications)	0.14	0.7105
BŚI (absence of personal protective equipment)	1.79	0.1804
BŚZ (absence of collective protective equipment)	3.97	0.0465
NCM (incorrect condition of material factor)	2.54	0.1113
NPD (incorrect comings and goings)	8.02	0.0046
POŚ (tripping, slipping)	6.11	0.0134
PNO (performing works ordered by a supervisor, outside the job description)	0.56	0.4555
PUZ (performing works without eliminating hazards)	5.02	0.0251
SPF (incorrect psycho-physical state)	0.24	0.6219
SZ (wilful misconduct)	2.14	0.1435

<sup>2</sup> Pearson’s Chi-square is a basis of significance test for qualitative (categorised) variables. This method is based on calculating the number of “*expected*” instances in a contingency table (i.e. the instances which would be expected if there was no relationship between variables).

<sup>3</sup> p-value – probability value, significance limit level (0.05) – probability of making the type I error, that is numerical expression of statistical significance.

<b>TN</b> (violations of the OHS rules being tolerated by the supervisors)	1.05	0.3044
<b>UKM</b> (losing control of a machine)	3.87	0.0492
<b>WZ</b> (entry to a hazard area)	5.17	0.0230
<b>ZN</b> (sudden incident)	9.23	0.0024
<b>ZO</b> (incorrect organisation of works)	1.15	0.2843

The greater the Chi-square is and the less the p-value is, the larger are differences between clusters 1 and 2 in relation a given variable.

For quantitative variables, the mean values and the range were compared in both clusters (Table 3). The significance of variability was evaluated using the F test of analysis of variance (ANOVA) – ratio of inter-group variability to intra-group variability:

$$(2) \quad F = \text{ratio: } \sum_{i=1}^K n_i (\bar{Y}_i - \bar{Y})^2 / (K-1) \quad \text{to: } \sum_{i=1}^K \sum_{j=1}^{n_i} n_i (Y_{ij} - \bar{Y}_i)^2 / (N - K)$$

where:

- $n_i$  – number of observations in the  $i$ th group,
- $\bar{Y}_i$  – sample mean in the  $i$ th group,
- $\bar{Y}$  – general mean of data,
- $Y_{ij}$  – observation  $j$  in  $i$  from group  $K$ ,
- $K$  – number of groups,
- $N$  – total sample size.

Table 3. ANOVA for quantitative variables in “Accidents 2014 – 16”.

Number of clusters: 2, total number of learning cases: 65. Source: own

Variable	Inter-group SS	df	Intra-group SS	df	F	p-value
Wi (years)	669.747	1	8478.71	63	4.98	0.0293
Stp (years)	97.210	1	1709.93	63	3.58	0.0630

The determined clusters differ significantly in relation to the “victim age” variable (Wi) ( $p=0.0293$ ). On the other hand, no statistical differences were found for the “seniority” variable (Stp).

Table 4. Quantitative variable statistics: Wi (victim age, in years) – “Accidents 2014 – 16”.

Number of clusters: 2, total number of learning cases: 65. Source: own.

Variable values	Cluster 1	Cluster 2	Total
Minimum	22.0	20.0	20.0



Maximum	63.0	62.0	63.0
Mean	44.6	38.0	42.2
Standard deviation	11.7	11.5	12.0

The most important for interpretation is the p-value column (Table 3) which says that the age of people involved in accidents from the first cluster is significantly different than in cluster 2 (Table 4 – descriptive statistics indicates that the average age in both clusters was 44.6 and 38 years, respectively), and that the differences in seniority are less significant. The remaining results in the table are rather auxiliary.

The results were illustrated on two box plots and raw data. The first (Fig. 1.) presents graphically the average value, range of variability, and untypical values of the variable, if any, in both clusters. The other plot (Fig. 2.) shows distribution of raw values of the variable.

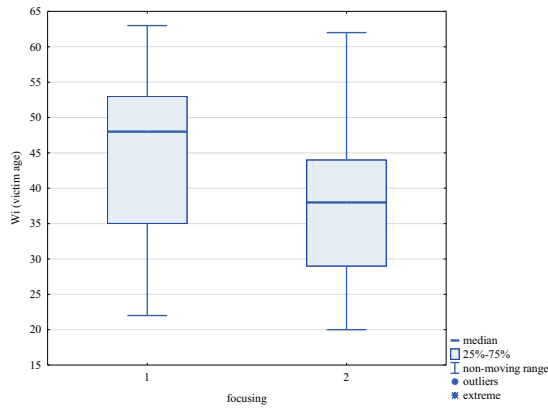


Fig. 1. Box plot of quantitative variable:  $W_i$  (victim age) in relation to clusters 1 and 2. Classification by results.

Source: own.

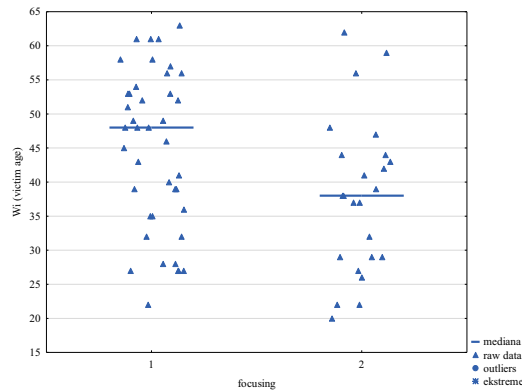


Fig. 2. Raw data lot of quantitative variable:  
 $W_i$  (victim age) in relation to clusters 1 and 2.

Source: own.

Below, figure 3 and table 5 present distributions of individual qualitative variables within clusters 1 and 2. The condition for a variable to be included in the plot was its occurrence, totally in both clusters, more than 5 times.

Distribution of variable: Cmc - material factor related to the activity performed, for “Accidents 2014 – 16”, included the following variables (Table 5):

Table 5. Sample size table for qualitative variable: Cmc – “Accidents 2014 – 16”.

Number of clusters: 2, total number of learning cases: 65. Source: own

Variable: Cmc (material factor related to the activity performed)			
Coding variables	Cluster 1	Cluster 2	Total
RM (mobile machinery and equipment)	16	0	16
RU (scaffolding)	9	0	9
DP (portable ladder)	4	0	4
EB (building element)	8	1	9
MB (construction materials)	0	14	14
SZW (lift shaft)	0	2	2
EN (power tool)	3	0	3
NONE	0	2	2
EI <sub>g</sub> (gas installation element)	0	4	4
LO (shovel)	1	1	2

The distribution of variable Cmc - material factor related to the activity performed, within the determined clusters, for “Accidents 2014 – 16”, is dominated by mobile machinery and equipment, scaffolding, building element, and construction materials (Fig. 3.).

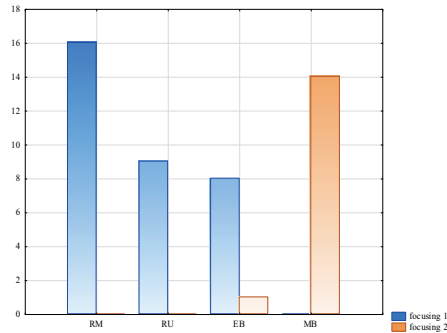


Fig. 3. Variable distribution plot: Cmc (material factor related to the activity performed).

Source: own

## 5. SUMMARY

The paper analyses the construction site accidents in the Małopolska region in 2014 – 16 which were recorded in the register kept by the District Labour Inspectorate in Krakow, Poland.

The basic objective was to determine the impact of selected site feature on safety hazards during construction works, expressed by worker behaviour, type and status of conditions for occurrence of an accident, as well as examination of suitability of cluster analysis as one of the data extraction techniques to perform such research. This objective was achieved by:

- analysing the records of the District Labour Inspectorate in Krakow which are the accidents database;
- selecting data on site features or workplaces where undesirable incidents took place, that is on material factors, incident types, and incident causes;
- identification and profiling of accident patterns.

Cluster analysis, as a method of conceptual grouping, is a good tool to indicate patterns in multidimensional sets which characterise construction site accidents. Application of this method in analyses of data described by independent variables allows profiling the construction site accidents. The variables which play an important role in investigating the accident patterns are visible in the method's algorithm.

The performed cluster analysis allowed determining groups of variables (clusters) characterising construction site accidents that occurred in the investigated time and place.

Using the variables Mw (accident location), Pp (work process), Cw (activity performed by the victim when accident occurred), Cmc (material factor related to Cw), Wdo (incident which is a deviation from normal state), Cmo (material factor related to deviation), Wyu (incident causing injury), Cmu (material factor being a cause of injury) from the "Accidents 2014-16" sheet, and the personal variables from the same sheet: Sz (employment status), Wi (victim age), Stp (victim seniority), Zwd (victim trade), and the EM algorithm, the cross-validation indicated 2 clusters. The age of people involved in accidents from the first cluster is significantly different than in cluster 2. The average age in both clusters was 44.6 and 38 years, respectively. The differences in seniority were less significant.

## REFERENCES

1. W. Drozd, Characteristics of construction site in terms of occupational safety, *Journal of Civil Engineering, Environment and Architecture*, Rzeszów University of Technology, January - March 2016, p. 165 – 172, 2016.
2. W. Drozd, M. Kowalik, Decision-makers in the analysis of accidents in construction, *Scientific Review Engineering and Environmental Sciences*, SGGW, Warsaw 2017.
3. B. Hoła, Work safety in construction processes. Publisher: Wrocław University of Technology. Wrocław 2016.
4. M. Törner, P. Pousette, Safety in construction – a comprehensive description of the characteristics of high safety standards in construction work, from the combined perspective of supervisors and experienced workers, *Journal of Safety Research* 40, p. 399-409, 2009.
5. A. Shapira, F. ASCE, B. Lyachin, Identification and Analysis of Factors Affecting Safety on Construction Sites with Tower Cranes, *Journal of Construction Engineering and Management* 135, 2009.
6. A. Iglendod, D. Klitherland, Safety climate factors, group differences and safety behaviour in road construction, *Safety Science* 39, p. 157-188, 2001.
7. R. M. Choudhry, D. Fang, Why operatives engage in unsafe work behavior: Investigating factors on construction sites, *Safety Science* 46, p. 566-584, 2008.
8. T. Niskanen, Safety climate in the road administration, *Safety Science* 17, 237-255, 1994.
9. S. Mohamed, T.H.Ali, W.Y.V.Tam, National culture and safe work behaviour of construction workers in Pakistan, *Safety Science* 47, p. 29-35, 2009.
10. R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large database, *Proceedings of the ACM SIGMOD Conference on Management of Data*, p. 207-216, 1993.
11. R. Agrawal, R. Srikant, Fast Algorithms For Mining Association Rules in Large Databases, *Proceedings of the 20th International Conference on Very Large Databases*, p. 487-499, 1994.

## LIST OF FIGURES AND TABLES:

Fig. 1. Box plot of quantitative variable: Wi (victim age) in relation to clusters 1 and 2. Classification by results

Rys. 1. Wykres ramkowy zmiennej ilościowej: Wi (wiek poszkodowanych) względem skupień 1 i 2.  
Klasyfikacja wyników

Fig. 2. Raw data lot of quantitative variable: Wi (victim age) in relation to clusters 1 and 2

Rys. 2. Wykres danych surowych zmiennej ilościowej: Wi (wiek poszkodowanych) względem skupień 1 i 2

Fig. 3. Variable distribution plot: Cmc (material factor related to the activity performed)

Rys. 3. Tytuł rysunku Wykres rozkładu zmiennej: Cmc (czynnik materialny związany z czynnością wykonywaną przez poszkodowanego)

Tab. 1. Accidents 2014 – 16

Tab. 1. Wypadki 2014 – 16

Tab. 2. Independence tests for random variables „Accidents 2014 – 16”

Tab. 2. Testy niezależności dla zmiennych jakościowych „wypadki 2014 – 16”

Tab. 3. ANOVA for quantitative variables in “Accidents 2014 – 16”

Tab. 3. Wyniki ANOVA dla zmiennych ilościowych występujących w analizie „wypadki 2014 – 16”

Tab. 4. Quantitative variable statistics: Wi (victim age, in years) – “Accidents 2014 – 16”

Tab. 4. Statystyki zmiennej ilościowej: Wi (wiek poszkodowanych, w latach) - „wypadki 2014 – 16”

Tab. 5. Sample size table for qualitative variable: Cmc – “Accidents 2014 – 16”

Tab. 5. Tabela licznosci dla zmiennej jakościowej: Cmc - „wypadki 2014 – 16”

## ANALIZA SKUPIEŃ W BADANIU WYPADKOWOŚCI W SEKTORZE BUDOWLANYM

*Słowa kluczowe:* plac budowy, bezpieczeństwo pracy, wypadek przy pracy, analiza skupień

### STRESZCZENIE:

Wypadki przy pracy, jako zdarzenia na terenie budowy, są zjawiskami losowymi, trudnymi lub niemożliwymi do przewidzenia. Dlatego ich badanie i wskazanie związków, między charakteryzującymi je cechami, nie jest łatwe. Dotyczy to w szczególności analiz, w których przedmiotem badań są czynniki generujące niepożądane następstwa dla życia i zdrowia ludzi.

W niniejszym artykule podjęto próbę identyfikacji i profilowania wzorców wypadków, na podstawie danych z prowadzonego w Okręgowym Inspektoracie Pracy, w Krakowie - rejestrze. Dyskutowanymi były zdarzenia wypadkowe w latach 2014 – 16. Liczba obserwacji wynosiła 65. Zasadniczy cel skoncentrowano na wyodrębnieniu wzorców wypadków budowlanych i wskazaniu na te spośród analizowanych cech, które mają istotne znaczenie w definiowaniu tych wzorców.

Cel ten został osiągnięty poprzez:

- przeanalizowanie zasobów Okręgowego Inspektoratu Pracy w Krakowie, stanowiących bazę danych o zdarzeniach wypadkowych;
- wybór danych o cechach terenu budowy lub stanowiskach pracy, na których doszło do niepożądanego zdarzenia w związku z podejmowaną pracą, a więc czynników materialnych, rodzajów zdarzeń i przyczyn wypadków;
- identyfikację i profilowanie wzorców wypadków.

W zakresie metodologicznym wykorzystano analizę zasobów danych za pomocą grupowania pojęciowego, w postaci analizy skupień. W artykule przedstawiono udaną próbę wykorzystania tej metody w obszarze badań wypadkowości w budownictwie. Przeprowadzona analiza skupień pozwoliła określić grupy zmiennych (klasy) charakteryzujące wypadki na placu budowy, które wystąpiły w badanym czasie i miejscu.

Wykorzystując zmienne z arkusza „wypadki 2014-16”: Mw (miejsce wypadku), Pp (proces pracy), Cw (czynność wykonywana przez poszkodowanego), Cmc (czynnik materialny związany z czynnością wykonywaną przez poszkodowanego w chwili wypadku), Wdo (wydarzenie będące odchyleniem od stanu normalnego), Cmo (czynnik materialny związany z odchyleniem), Wyu (wydarzenie powodujące uraz), Cmu (czynnik materialny będący źródłem urazu) oraz dodatkowo zmienne metryczkowe, z tego samego arkusza: Sz (status zatrudnienia), Wi (wiek poszkodowanego), Stp (staż pracy poszkodowanego), Zwd (zawód wykonywany przez poszkodowanego), walidacja krzyżowa wskazała 2, jako optymalną liczbę skupień. Osoby ulegające wypadkom z pierwszego skupienia istotnie różnią się wiekiem od osób ze skupienia 2. Średni wiek w obu skupieniach, to odpowiednio 44,6 i 38 lat, natomiast różnice w stażu pracy są mniej znaczące