

CLUSTER ANALYSIS OF FATAL ACCIDENTS SERIES IN THE INFOR.MO DATABASE: ANALYSIS, EVIDENCE AND RESEARCH PERSPECTIVES

M. LOMBARDI & G. ROSSI

Safety Engineer, University of Roma La Sapienza, Rome, Italy.

ABSTRACT

The state of the application of the techniques of cluster analysis does not include the work accidents. The applications more established for statistical data analysis include pattern recognition, image analysis and information retrieval.

The aim of this study is to provide a quantitative assessment, based on techniques of statistical processing of historical data in order to highlight the causality between the accident and predictive recurring events. On the basis of information provided by the analysis, it is possible to propose preventive strategies targeted to reducing the number of accidents (mainly the fatal accidents).

Based on the collection of fatal accidents in the Infor.MO database (INAIL), we proceeded to aggregate accident cases registered in order to provide cluster analysis, which with reference to generators of the danger flow mortal areas, could show typical accidents, namely preferential genesis that, proposing the causes of the same energy mortal flow, could explain a large number of events.

In order to run the analysis, a methodological assumption that describes the phenomenon of accidents, like any algebraic entity, as the case represented in algebraic space, is requested.

The n dimensions useful to describe the phenomenon are the n generators of the danger areas. Based on this premise, each accident can be represented by the Boolean n -tuple of coordinates in space R_n . This purpose allows to transform the descriptions of accidents in algebraic and statistic case study on which to apply the statistical cluster analysis protocol.

Applying this method to the analysis of fatal accidents in the Infor.MO database, related to the ATECO Construction Sector (F), with particular reference to the ‘falls from heights’, has showed successful clustering.

The results of the analysis aim to check the effective purposes of prevention/protection, in a perspective of maximum efficiency.

Keywords: Accidents at work, Boolean analysis, cluster analysis, falls from heights, fatal accidents database, slipping.

1 INTRODUCTION

For the purpose of the analysis, the historical data of the accidents are available on the Infor.MO database.

The Infor.MO project is together a protocol to collect accident data and a rational approach to reconstruct the accident events, represented by synthetic descriptors of harmful event, characterised in terms of energy.

The sample, extracted from databases and used in this analysis, consists of the events for the Construction Sector (‘F’) in the years 2002–2004.

The variables for the code, according to the ESAW model, are as follows:

Report of the accident occurred.

Year.

Number of accidents.

Fatalities.

Part of the lesion.

Nature of the lesion.

Sex.
Nationality.
Knowledge of the Italian language.
Number of years spent in Italy.
Educational qualification.
Employment contract.
Position.
Job seniority.
Economic sector activity.
Accident location.
Material agent accident.
Environment.
Type of accident (referring to model 'you learn from your mistakes').
Accident or deviation.

The accidents were distinct from fatalities and disability.

The extract accidents do not exhibit a high degree of homogeneity, due to the fact that the accident reconstructive main descriptors are missing in many files. However, sometimes by interpreting the event description, we filled the gaps where possible.

Figure 1 is a file type example from Infor.Mo database (INAIL website).

2 PRELIMINARY STATISTICAL ANALYSIS

Preliminary analysis of selected data has been done normalising events with respect to the number of workers in the economic sector of 'building construction' in the considered time interval.

The next step of the analysis is classifying the cases of disabilities and fatalities for 'Deviation' (accident) and 'Deviation Type' (energy change).

We identified seven different types of accidents leading to deviation (the last category is actually a residual category – cf. Table 1) and for each type of deviation a subset containing the number of cases and energy change by which these occur.

For each mode data were stratified with reference to three types of dangerous energy change:



1. Energy exchange.
2. Energy release.
3. Inappropriate application of energy.

It shows the relative frequency for deviation of fatal accident (Table 1) and histogram (Fig. 2).

For each accident, determinants or modulators are defined, i.e. the areas' generators of dangerous mortal flow for the injured.

According to the model 'you learn from your mistakes', the following determinants have been selected:

- Activity of the injured.
- Activities of third parties.
- Tools and equipments.
- Materials.
- Environment.
- Personal protective equipment.

Ti trovi in ISPESL / Statistiche / Infurtuni Mortali / I dati / INFOR.MO. / Risultati / Caso 2 / Dettagli infortunio

Id infortunio 2

Tipologia: Mortale

DATI RELATIVI ALL'INFORTUNIO

Data infortunio: 2002	Ora solare: 10.45	Ora ordinale:	Giorno della settimana: Lunedì	Infortunio collettivo:
Luogo infortunio:				
<input type="checkbox"/> in strada <input type="checkbox"/> in itinere <input checked="" type="checkbox"/> sul lavoro				
Descrizione del ciclo di lavoro svolto nell'unità organizzativa/reparto: DEMOLIZIONE DI UN VECCHIO FABBRICATO DI 3 PIANI. RECUPERO MATERIALE EDILE.				
Descrizione dell'evento: L'INFORTUNATO, TITOLARE DI PROPRIA DITTA ARTIGIANA, SI TROVAVA IN UN CANTIERE EDILE DURANTE LA DEMOLIZIONE DI UN VECCHIO FABBRICATO, AVEVA AVUTO IL PERMESSO DI ACCESSO AL CANTIERE DA PARTE DEL TITOLARE DELL'IMPRESA CHE ESEGUIVA I LAVORI DI DEMOLIZIONE, ALLO SCOPO DI RECUPERARE I VECCHI MATTONI DEL FABBRICATO. IL TITOLARE DELL'IMPRESA AVEVA RACCOMANDATO ALL'INFORTUNATO DI NON RECARSI NELLE ZONE DI PERICOLO, DELIMITATE DA TRANSENNE. CIO' NONOSTANTE L'INFORTUNATO, DOPO ESSERSI RECATO AL SECONDO PIANO DELL'EDIFICIO, OLTREPASSAVA UNA TRANSENNA E POSIZIONAVA UN BANCALE DI LEGNO ACCANTO AD UN MUCCOIO DI MACERIE. MENTRE SISTEMAVA I MATTONI SUL BANCALE SI E' VERIFICATO IL CROLLO DELLA VOLTA CON SPROFONDAMENTO DELLE MACERIE, DEL BANCALE E DELL'INFORTUNATO AL PIANO SOTTOSTANTE DA UN'ALTEZZA DI CIRCA METRI 2,10. IL CEDIMENTO DELLA VOLTA E' STATO VEROSIMILMENTE CAUSATO DAL PESO ECCESSIVO DEL BANCALE (CIRCA 300 KG) E DALLA SUA POSIZIONE RISPETTO ALLA TRAVE METALLICA DI SOSTEGNO. L'INFORTUNATO NON INDOSSAVA L'ELMETTO DI PROTEZIONE, NE DISPOSITIVI ANTICADUTA. INOLTRE NON AVEVA PROVVEDUTO A RAFFORZARE IL PIANO DI LAVORO NE A VISIONARE L'ANDAMENTO DELLE TRAVI DI SOSTEGNO DELLA STRUTTURA. L'INFORTUNATO E' DECEDUTO IL GIORNO SUCCESSIVO A CAUSA DI EMORRAGIA SUBARACNOIDEA MASSIVA E POLITRAUMA.				
Scambio di energia: Scambio				
Cittadinanza: ITALIA	In Italia dall'anno:	Conoscenza della lingua Italiana:	Titolo di studio:	
Sesso: M	Rapporto lavoro: Autonomo con dipendenti - Titolare con dipendenti	Mansione: ARTIGIANO EDILE Codice ISCO 88 (ex ISTAT 2001):	injured data	
Aziendatà nella mansione: In azienda: anni - mesi - giorni - In complesso: anni - mesi - giorni -	Aziendatà complessiva di lavoro: In azienda: anni - mesi - giorni - In complesso: anni - mesi - giorni -	Sede anatomica della lesione: CRANIO		
Natura della lesione: CONTUSIONE	Codice attività INAIL: 3110 - LAVORI GENERALI DI COSTRUZIONE E CANTIERISTICA	Codice attività ATECO: 45.45.1 - Attività non specializzate di lavori edili	conditions accident	
Modalità di accadimento (ESAW)	Modalità di accadimento (ESAW)			
Modalità di accadimento (ESAW) Tipo di luogo: 022 - Cantiere, fabbricato in demolizione, restauro, manutenzione Tipo di lavoro: 24 - Restauro, riparazione, ampliamento - di ogni tipo Attività fisica specifica: 41 - Prendere in mano, afferrare, strappare, tenere in mano, deporre - su un piano orizzontale Agente materiale dell'attività fisica specifica: 14.01.02.01 - Mattoni, tegole Deviazione: 34 - Scivolamento, caduta, crollo di agente materiale posto al di sotto (che trascina la vittima) Agente materiale della deviazione: 02.01.99.00 - Altre parti in altezza di un edificio Contatto: 31 - Movimento verticale, schiacciamento su/contro (risultato di una caduta) Agente materiale del contatto: 01.02.01.00 - Superfici in generale				

Figure 1: Infor.MO database: (a) accident report case study and (b). injured data and conditions accident.

3 CLUSTER ANALYSIS MODEL

The clustering techniques group the selected records (statistical unit) into clusters (groups). The goal is to define both most homogeneous groups within and most diverse groups between themselves [1, 2].

The application of cluster analysis to the sample described is designed to test the hypothesis of familiarity of the elements contained in the individual group. For the purpose of cluster

Table 1: Relative and absolute frequency distribution of fatal accidents.

Accident type	Relative frequency distribution	Absolute frequency distribution
To be run down	0.038	21
Electric failure	0.055	30
Breaking material	0.060	33
Loss of control of the vehicle	0.088	48
Slipping	0.192	105
Falls from heights	0.266	146
Other deviation	0.301	164
Sum	1	547

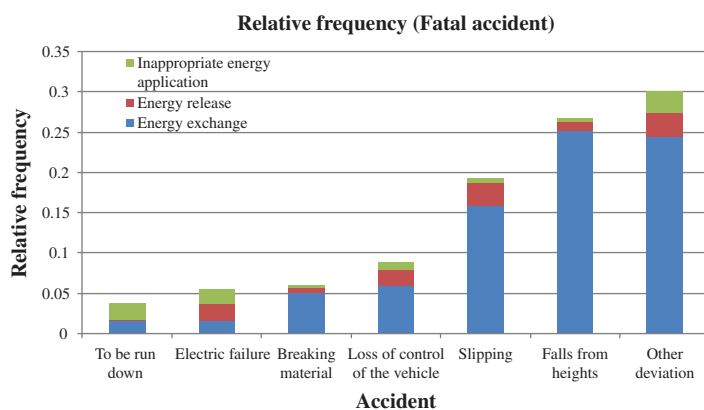


Figure 2: Relative frequency histogram and accident mode.

analysis application, the data have been manipulated for the transformation of each accident in the statistical event normalised with the Boolean logic (0 if absent, 1 if present) of the modulators (or determinants) of the accident.

The starting point consists of a data matrix cases \times variables; the arrival point is represented by information on the individual reference group for each case [3].

The analysis, based on the non-hierarchical method, produces a partition from a fixed number k of groups when defined the variables on which the Euclidean distance is calculated and a number ' k ' of clusters. In the classification stage, the cases are aggregated into clusters based on the criterion of 'proximity' to the initial cluster centroid, then the centroids of the groups are redefined and the cases are reclassified into the clusters in order to minimise the variance within each cluster and to maximise the variance between the clusters.

Cluster analysis has been set on the samples obtained by the stratification based on the modality of accident divided as follows:

- To be run down.
- Electric failure.
- Breaking material.

- Loss of control of the vehicle.
- Slipping.
- Falls from heights.

For the sample size, the analysis has been developed with regard to the conditions of accident:

- Loss of control of the vehicle.
- Slipping.
- Falls from heights.

3.1 Cluster analysis: sample 4 – loss of control of the vehicle

Table 2 shows the selected records for the modality of accident ‘Loss of control of the vehicle’.

Table 2: Sample 4 – loss of control of the vehicle: Boolean logic application of the determinants.

Accident report number	Activities of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment
1	0	1	0	1	1	1
2	1	1	1	1	0	0
3	1	1	0	1	0	0
4	1	0	1	0	1	0
5	1	0	1	0	0	0
6	0	1	1	0	0	0
7	1	0	0	1	0	1
8	1	0	1	0	0	0
9	1	0	0	0	1	0
10	1	1	0	0	0	0
11	1	1	1	0	0	0
12	0	1	1	0	0	1
13	1	0	1	0	0	0
14	0	0	0	1	1	0
15	0	0	0	0	1	0
16	1	0	0	0	0	1
17	0	0	1	0	0	0
18	1	1	0	0	0	0
19	1	1	1	0	0	0
20	1	0	1	0	0	0
21	1	1	0	1	0	0
22	0	1	0	1	0	1

Table 3: Sample 4: number and composition of selected clusters' centroids of six determinants.

Centroids	Activity of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment	Number of cases
Cluster A	0.38	0.63	0.13	0.75	0.38	0.50	8
Cluster B	0.86	0.43	0.71	0.07	0.14	0.07	14
Mean	0.62	0.53	0.42	0.41	0.26	0.29	Valid 22

The accident reports have been normalised with the Boolean logic of the modulators shown on the front row of the table. The value '0' means that the modulator is not involved in the causes of the accident, whereas the value '1' means that the modulator is the cause of the accident.

The application of cluster analysis to sample 4 shows the existence of two main clusters, cluster A and cluster B, with some degree of stability classification, as a function of the size of the proposed solution and the objectivity of the evidence [4, 5].

Table 3 shows the clusters with regard to the values of centroids in the variables 'Determinants' and the relative consistency.

It seems clear that the distinctive feature between the two clusters (A and B) is constituted by determining 'Tools and equipments' and 'Materials', the values of which is consistently polarised on the two clusters (Table 3).

The cluster A collects values with 'Tools and equipments' tending to 0, i.e. the contribution of tools and equipments makes little impact on the activity of the injured.

With regard to other modulators:

- 'Activity of injured' and 'Activities of third parties' have a main weight, then the causes of the accident could be found right in them.
- 'Environment' and 'Personal protective equipments' have a secondary influence with minimal accident contribution.

The cluster B collects values with 'Tools and equipments' tending to 1, i.e. the contribution of tools and equipments makes main impact on the activity of the injured.

With regard to other modulators:

- 'Activity of the injured' and 'Activities of third parties' have a main weight, then the cause of the accident is negligence of the worker.
- 'Materials', 'Environment' and 'Personal protective equipment' have a secondary influence with minimal accident contribution.

The first step of the *k*-means non-hierarchical analysis is to define the clusters A and B by setting the initial coordinates of the centroids (Table 4) and using the criterion of the Euclidean distance to form partitions in order to minimise the variance within each cluster and to maximise the variance between the clusters.

Table 5 shows clusters grouping of accident cases: it is useful to identify cases 'typical' (most representative) of the groups according to the minimal distance from the centroid of the group.

Table 4: Sample 4: initial cluster centroids.

Cluster	A	B
Activity of the injured	0.00	1.00
Activities of third parties	1.00	0.00
Tools and equipments	0.00	1.00
Materials	1.00	0.00
Environment	1.00	0.00
Personal protective equipment	1.00	0.00

Table 5: Cluster reference and distance.

Accident report number	Cluster reference	Cluster distance
1537	A	1.00
836	B	1.15
941	A	1.00
1199	B	1.02
1294	B	0.56
1394	B	1.08
1416	A	1.12
1447	B	0.56
1738	B	1.21
1851	B	0.94
1961	B	0.68
2012	A	1.41
2123	B	0.56
2249	A	1.12
2278	A	1.32
2370	B	1.26
2665	B	1.02
2669	B	0.94
2679	B	0.68
2683	B	0.56
2719	A	1.00
2725	A	0.87

The comparison between the means of the clusters in the k -means solution to two clusters (Fig. 3) shows that cases of injury belonging to the cluster B are caused by human error, direct or of third parties, in order to use work equipment and tools while they are irrelevant the variables 'Materials', 'Environment' and 'Personal protective equipment'. The cluster

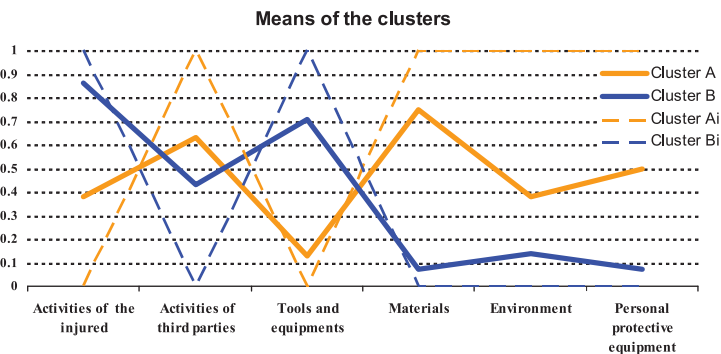


Figure 3: Graphical ‘Means of clusters’ non-hierarchical solution of *k*-means to two clusters.

Table 6: Analysis of variance (ANOVA) – loss of control of the vehicle.

	Cluster		Error		F	Sig.
	Mean squares	Degrees of freedom	Mean squares	Degrees of freedom	Mean squares	<i>p</i>
Activity of the injured	1.183	1	.179	20	6.594	.018
Activities of third parties	.196	1	.265	20	.741	.400
Tools and equipments	1.768	1	.187	20	9.474	.006
Materials	2.344	1	.121	20	19.305	.000
Environment	.274	1	.179	20	1.529	.231
Personal protective equipment	.935	1	.146	20	6.386	.020

A is only slightly polarised: the effect of the variable ‘Materials’ is combined with other variables (‘Activities of third parties’ and ‘Environment’).

Table 6 shows the solutions of analysis of variance (ANOVA), which confirm significant differences between the clusters. This result confirms the hypothesis that the clusters are constructed so as to be as homogeneous as possible within them and maximally different from the other clusters. The consequence is that if the variance within the group is much lower than the variance between groups, the *F*-test is high and significant. Therefore, there is a considerable difference between the means of the clusters. The variables that differentiate the clusters A and B are indicated in the table by the *F*-test values.

3.2 Cluster analysis: sample 5 – slipping

Table 7 shows the selected records for the modality of accident ‘Slipping’.

The accident reports have been normalised with the Boolean logic of the modulators shown on the front row of the table. The value ‘0’ means that the modulator is not involved in the cause of the accident, whereas the value ‘1’ means that the modulator is the cause of the accident.

Table 7: Sample 5 – slipping: Boolean logic application of the determinants.

Accident report number	Activities of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment
1	1	0	0	0	1	0
2	1	1	0	0	0	0
3	1	1	1	0	0	0
...
41	1	1	1	1	0	1

Table 8: Sample 5: number and composition of selected (five) clusters' centroids of six determinants.

Centroids	Activities of injured	Activities of the third parties	Tools and equipments	Materials	Environment	Personal protective equipment	Elements
Cluster A	0.55	0.27	0.82	0.09	1.00	0.09	11
Cluster B	0.67	0.83	0.33	0.17	0.00	0.00	6
Cluster C	1.00	0.00	0.00	0.22	1.00	0.22	9
Cluster D	0.80	0.10	0.80	0.80	0.00	0.10	10
Cluster E	0.20	0.00	1.00	0.00	0.20	0.60	5
Mean	0.64	0.24	0.59	0.26	0.44	0.20	Valid 41

The application of cluster analysis to sample 5 shows the existence of three main clusters, cluster A, cluster C and cluster D (see Table 8), which are selected for imposed solutions up to six clusters (objectiveness of the solution) [4, 5]. Two clusters include 30 cases of 41.

Interesting conclusions can be forwarded on the contribution of all the determinants that are selective determinants of cluster size.

The cluster A collects values with:

- 'Activity of the third parties' tending to 0.
- 'Materials' tending to 0.
- 'Personal protective equipment' tending to 0.

The cluster includes mainly those accidents that have not been caused by the external dangerous flow that has involved the injured during the work activity. The negative interaction between the worker, the environment and the equipments has produced a flow of danger to the worker.

The cluster C collects values with:

- 'Activity of the injured' tending to 1.
- 'Environment' tending to 1.

The cluster collects, although not exclusively, the cases of accident due to negligence of the workers; the collected cases of accident do not have responsibility for generating a third party, while there is always an ‘environmental triggering cause’.

The cluster D collects values with:

- ‘Activity of the injured’ tending to 1.
- ‘Tools and equipments’ tending to 1.
- ‘Materials’ tending to 1.

The other determinants do not affect the patterns of injury.

In this case, the cluster includes again the cases of accident due to negligence of the workers using the equipment and the materials.

To confirm this trend, a second test for imposed solution up to two clusters (Table 9) was performed.

As noted above, the negligence of the workers for the collected cases of accident is confirmed.

In the cluster A, the negative interaction of the worker with the equipment and the work environment is more evident; in the cluster B, the determinants ‘Tools and equipments’, ‘Materials’ and ‘Environment’ combined with the accident.

The comparison between the means of the clusters in the *k*-means solution to two clusters (Fig. 4, Clusters A and B) shows that the cases of injury belonging to the cluster B are caused mostly by human direct error in order to use work equipment and tools and materials, while they are irrelevant the variables ‘Activities of third parties’, ‘Environment’ and ‘Personal

Table 9: Sample 5: number and composition of selected (two) clusters’ centroids of six determinants.

Centroids	Activity of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment	Number of cases
Cluster A	0.71	0.14	0.48	0.14	1.00	0.19	21
Cluster B	0.65	0.30	0.70	0.45	0.00	0.15	20
Mean	0.68	0.22	0.59	0.30	0.50	0.17	Valid 41

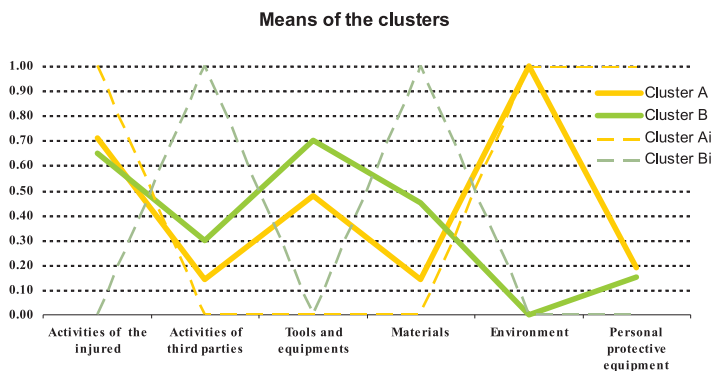


Figure 4: Graphical ‘Means of clusters’ non-hierarchical solution of k means to two clusters.

protective equipment'. The cluster A is only slightly polarised: the effect of the variable 'Environment' is combined with the variable 'Tools and equipments'.

The variables 'Activities of third parties' and 'Personal protective equipment' do not affect the mode of occurrence of accidents.

For the accident case 'Slipping', Table 10 presents the solutions of ANOVA, which confirm significant differences between the clusters A and B. This result confirms the hypothesis that the clusters are constructed so as to be as homogeneous as possible within them and maximally different from the other clusters. The consequence is that if the variance within the group is much lower than the variance between groups, the *F*-test is high and significant. Therefore, there is a considerable difference between the means of the clusters. For the six analysed variables, the *F*-test values are indicated in the table. The clusters A and B differ in the variables 'Materials' and 'environment': the related values of *F*-test are shown in the table.

3.3 Cluster analysis: sample 6 – falls from heights

The application of cluster analysis to sample 6 (Table 11 collection data) shows the existence of four main clusters (Table 12) that are selected for imposed solutions up to 12 clusters (objectiveness of the result) [4, 5].

Table 10: Analysis of variance (ANOVA) – slipping.

	Cluster		Error		<i>F</i>	Significance
	Mean squares	Degrees of freedom	Mean squares	Degrees of freedom	Mean squares	<i>p</i>
Activity of the injured	.042	1	.227	39	.187	.668
Activities of third parties	.253	1	.174	39	1.457	.235
Tools and equipments	.513	1	.242	39	2.120	.153
Materials	.966	1	.193	39	<u>5.011</u>	<u>.031</u>
Environment	10.244	1	.000	39	<u>.000</u>	<u>.000</u>
Personal protective equipment	.017	1	.148	39	.113	.738

Table 11: Sample 6: falls from heights: Boolean logic application of the determinants.

Accident report number	Activities of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment
1	0	0	1	0	0	0
2	0	1	0	0	0	1
3	1	0	0	0	1	0
...
72	1	0	1	0	1	0

Table 12: Sample 6: number and composition of selected (four) clusters, centroids of six determinants.

Centroids	Activities of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment	Elements
Cluster A	0.77	0.12	1.00	0.00	0.42	0.12	26
Cluster B	0.25	0.94	0.38	0.19	0.25	0.38	16
Cluster C	0.78	0.06	0.00	0.00	0.72	0.28	18
Cluster D	0.67	0.00	0.33	1.00	0.58	0.08	12
Mean	0.62	0.28	0.43	0.30	0.49	0.22	Valid 72

The contribution of all the determinants that are selective determinants of cluster size can be summarised as shown below:

The clusters A and C collect values with:

- ‘Activity of the third parties’ tending to 0.
- ‘Materials’ tending to 0.
- ‘Personal protective equipment’ tending to 0.

The clusters collect mainly those accidents that have caused by the dangerous flow due to the injured worker during the work activity. The negative interaction between the worker, the environment (Cluster C) and the equipments (Cluster A) has produced a flow of danger to the worker.

The cluster B collects values with:

- ‘Activity of the third parties’ tending to 1.

The other determinants have a minimal effect on the damage models.

The cluster D collects values with:

- ‘Activity of the third parties’ tending to 0;
- ‘Personal protective equipment’ tending to 0.

In this case also, the negative interaction between the worker (‘activity of the injured’), the ‘Environment’ and the ‘Materials’ has produced a flow of danger to the worker.

Interesting considerations can be forwarded on the contribution of the determinants ‘Activity of the injured’ (clusters A, C and D), ‘Tools and equipments’ (cluster A) and ‘Materials’ (cluster D).

- Cluster A seems to select the sample mainly with respect to the value of determinant ‘Activities of injured’, specifically the cluster collects mainly accidents happened for negligence of the worker. It is possible to note that on 26 cases included in cluster A, modulators regarding ‘Environment’, ‘Personal protective equipment’, ‘Activities of third parties’ have values tending to 0.
- Cluster B includes mainly accidents happened for the main role of determinants ‘Activities of third parties’. On the contrary, we found the values tending to 0 for determinants ‘Activity of the injured’, ‘Environment’ and ‘Materials’; concurrent factors are the modulators ‘Tools and equipments’ and ‘Personal protective equipment’.

In order to verify the possibility of clustering the cases of accident with respect to two main variables ('Activity of the injured' and 'Activities of third parties'), the analysis was repeated by requiring solutions to three and two clusters (Tables 13 and 14).

In Tables 13 and 15, the initial and final centroids of the determining variables, clusters and number of elements are shown.

The contribution of all the determinants can be summarised as shown below:

- The clusters A, B and C are polarised on the variables 'Activity of the injured' and 'Activities of third parties'.
- Cluster A selects the sample mainly with respect to the value of determinant 'Activities of the injured'.
- Cluster B selects the sample mainly with respect to the value of determinant 'Activities of third parties'.
- Cluster C collects values with 'Activities of injured' and 'Activities of third parties' tending to 0. The other modulators have an average trend, possible causes of accident have to be found in almost all the modulators summarising the work environment and working methods.

In every case collected in the clusters A and B, the negligence of the workers is evident: the determinants 'Activities of injured' and 'Activity of the third parties' are the main classification variables.

In the cluster A, the negative interaction of the worker with the equipment and the work environment is more evident; in the cluster B, the determinants 'Tools and equipments' and 'Environment' combine with the accident. In this case, the determinant 'Activity of the injured' cannot be excluded. The flow of the danger is generated from the activities of the workers.

Table 13: Sample 6: initial centroids of six determinants of selected (three) clusters.

Centroids	Activities of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment
Cluster A	1.00	0.00	0.50	0.50	0.50	0.50
Cluster B	0.00	1.00	0.50	0.50	0.50	0.50
Cluster C	0.50	0.50	0.50	0.50	0.50	0.50
Mean	0.50	0.50	0.50	0.50	0.50	0.50

Table 14: Sample 6: number and composition of selected (three) clusters' centroids of six determinants.

Centroids	Activities of the injured	Activities of third parties	Tools and equipments	Materials	Environment	Personal protective equipment	Elements
Cluster A	1.00	0.00	0.53	0.21	0.53	0.18	38
Cluster B	0.42	1.00	0.47	0.16	0.32	0.26	19
Cluster C	0.00	0.00	0.47	0.27	0.60	0.20	15
Mean	0.47	0.33	0.49	0.21	0.48	0.21	Valid 72

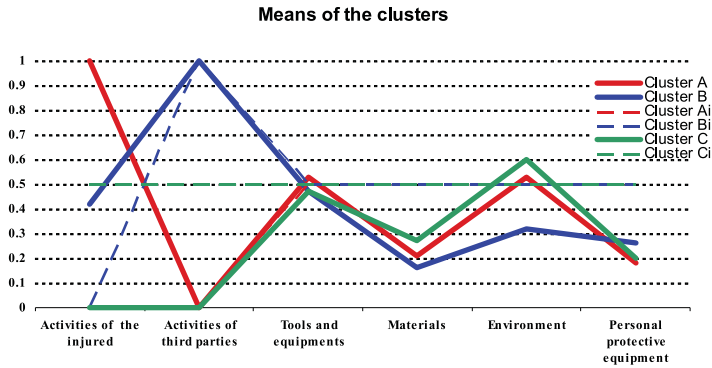


Figure 5: Graphical ‘Means of clusters’ non-hierarchical solution of *k*-means to three clusters.

Table 15: Analysis of variance (ANOVA) – falls from heights.

	Cluster		Error		<i>F</i>	Significance
	Mean squares	Degrees of freedom	Mean squares	Degrees of freedom	Mean squares	<i>p</i>
Activity of the injured	5.990	2	.067	69	89.234	.000
Activities of third parties	6.993	2	.000	69	.000	.000
Tools and equipments	.028	2	.260	69	.108	.898
Materials	.050	2	.171	69	.292	.748
Environment	.404	2	.249	69	1.621	.205
Personal protective equipment	.040	2	.171	69	.235	.791

The comparison between the means of the clusters in the *k*-means solution to three clusters (Fig. 5, Clusters A, B and C) shows that the cases of injury belonging to the cluster A are caused mostly by human direct error in order to use work equipment and tools and materials while they are irrelevant the variables ‘Activities of third parties’, ‘Materials’ and ‘Personal protective equipment’. The cluster B is only slightly polarised: the effect of the variable ‘Activities of third parties’ is combined with the variable ‘Activities of the injured’.

The variables ‘Materials’ and ‘Personal protective equipment’ do not affect the mode of occurrence of accidents.

Table 14 presents the solutions of ANOVA, which confirm significant differences between the six determinants. The determinants that classify the analyzed cases are the variables ‘Activities of the injured’ and ‘Activities of third parties’.

4 CONCLUSION

In the course of time, work accidents have been considered and interpreted generally by in-formal characterisations.

The present study aims to preliminary formalise the representation of accident cases-study represented in the space R^6 of algebraic determinants.

This representation tool has allowed the cluster analysis implementation for sample stratified by mode of accidents, in order to identify recurrent injured.

Particularly significant results have come out from the cluster analysis applied to the sample 6: Falls from heights.

In this case, more than 90% of accidents were included in three well-defined clusters.

In addition, each cluster is characterised by values well polarised and generally exclusive, according to the type of modulator.

The results of this analysis, which will be verified on larger samples size and which will be compared with evaluations made in areas selected by analogy, address the possibility, with reference to this category of the accidents, of a specific monitoring of single determinants, which same to be able to cause the accident, each one individually, in case of loss of the protective function [6].

REFERENCES

- [1] Abdi, H. & Williams, L.J., Principal component analysis. *Wiley Interdisciplinary Reviews. Computational Statistics*, **2**, pp. 433–459, 2010.
- [2] Dillon, W.R. & Goldstein M., *Multivariate Analysis Methods and Applications*. Wiley: New York, 1984.
- [3] Greenacre, M.J., *Correspondence Analysis*, John Wiley & Sons, Inc: New York, 2010.
- [4] Greenacre, M.J., Interpreting multiple correspondence analysis. *Applied Stochastic Models and Data Analysis*, **7(2)**, pp. 195–210, 1991. doi: <http://dx.doi.org/10.1002/asm.3150070208>
- [5] Lebart, L., Morineau, A. & Warwick, K.M., *Multivariate Descriptive Statistics*. Wiley: New York, 1984.
- [6] Rand, W.M., Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, **66(336)**, pp. 846–850, 1971. doi: <http://dx.doi.org/10.1080/01621459.1971.10482356>