

Multivariate Statistical Analysis Approach to Cluster Construction Workers based on Labor Productivity Performance

Diego Calvetti N.




Department of Civil Engineering, Faculty of Engineering, University of Porto, Porto, Portugal
(diego.calvetti@fe.up.pt) ORCID [0000-0001-9893-0377](https://orcid.org/0000-0001-9893-0377)

Abstract

In the construction industry, the direct workforce is one of the most important drivers of the work process. Identifying and quantifying labor productivity impact factors allows the diagnosis of recurring problems during the construction phase. Understanding how these factors influence the productive and the nonproductive states according to the characteristics of workers or group of workers is an essential tool to boost productivity. This paper introduces a multivariate statistical analysis approach to cluster workers based on the characteristics of the actions that are performed during the daily construction tasks. This study analyzed the data from a field experiment based on human observation of actions of 10 welders during a week in a pipe-shop. The case study conducted step by step presented in this work indicates retention of 50% and 40% of the total sample in segmented workers clusters.

Author Keywords. Multivariate Statistical Analysis, Workers, Labor Productivity, Performance, Continuous Improvement

Type: Research Article

 Open Access  Peer Reviewed  CC BY

1. Introduction

Construction is a labor-intensive industry. Consequently, the performance on construction is mainly dependent on labor productivity of the workers ([Jarkas and Bitar 2012](#)). Organizations are naturally interested in monitoring the performance of workers. Monitoring labor productivity performance allows managers to obtain important information for application in a planning of actions to continuous improvement ([Alder 2001](#)).

A continuous process improvement eliminates activities that do not add value and this allows processes to operate at increased levels of efficiency ([Project Management Institute 2013](#)). There are many factors that impact productivity in the construction industry, among which, those that lead to variations in production rates are difficult to control. However, there are other factors that can be easily identified and can be used to boost production rates. Attention to labor productivity management should be focused on the sources and causes of delays and, the perception of what are the actions those workers spending excessive efforts ([Christian and Hachey 1995](#)). A productivity continuous improvement process may be based on the identification of the impacts, quantification of the effects and, ranking of the most damages. Based on that process the implementation of actions to correct the most relevant problems should be made. As a result, that plan of actions may be generic for all workers. Now that, if a more detailed plan is desired that could be the origin of too many plans that could be impossible to control.

A multivariate statistical analysis approach to cluster construction workers based on labor productivity performance aims to be a tool to boost the effectiveness of productivity continuous improvement process. This paper first introduces and characterizes the data collected in field experiment. Then, this work describes the methodology applied to cluster the workers in case productive and nonproductive states. In the sequence, in order to provide a guide for further applications, the analysis is conducted step by step and the results are discussed and, finally, a conclusion is made.

2. Data Collected in Field Experiment

2.1. Sample characteristics

The experiment was carried out at Jambeiro Caldeiraria e Usinagem, located in the city of Jambeiro in the State of São Paulo, Brazil. The manufacturing plant of the Company has a total area of 73,117 m² with a constructed area of 16,492 m². Jambeiro Caldeiraria stands out in the manufacture of pipes for numerous projects for the Brazilian heavy construction industry.

The pipe manufacturing process is generally determined on the basis of the pipe design drawings, where they are divided into isometrics, and further subdivided into smaller parts called spools. In this way, a spool contains the union of elements, such as straight stretches of pipe and connections, among others, mostly consolidated through welding. The experiment was carried out at the service fronts of the welders under analysis, inside a metallic building, with full closure (cover and sides), characteristic of indoor activities. The experiment was carried out with ten welders, during their usual activities within the regular workday. The selection of welders occurred in a random manner, according to the indication of Jambeiro Company, among qualified welders in accordance with the standards used in the Brazilian industry.

The activities performed during the evaluation period consisted of the welding of carbon steel and low alloy steel pipes by the Tungsten Inert Gas (TIG) and Flux-cored arc welding (FCAW) methods. Because it is a pipe-shop, the processes take place in a production line, the pipe manufacturing pavilion arrangement is segmented into a given physical area by each type of activity. That beginning with the cutting, through the adjustment and coupling, in sequence arriving at the specific area of welding and, continues after until the last area where the heat treatments and tests for quality certification take place.

The purpose of the data collection procedure was to detect worker activity or inactivity. Observations are taken on a timely basis, whether or not timing is taken, but rather a marking of the action the welder is performing at the very moment of observation. During the data collection of this experiment the observation of the activity of each welder occurred in a time not less than 5 minutes for the same welder, where at random during the working day were collected data of the 10 welders who participated in the experiment.

In the model of [Adrian \(2004\)](#), the data collected through the human observation are stratified according to the work state, be it performing activity (productive) or not performing activity (nonproductive). The number of points of each individual in each variable is transformed into a percentage value based on the ratio of the total observations. The observations occurred between 07/30/2015 and 06/08/2015 and during this period 3,577 evaluations were collected regarding the activities of the ten welders under analysis.

2.2. Productive state

The productive state is that when the worker is performing some work activity. That activity may be direct connected to work or could be a contributory, support or auxiliary task ([Adrian 2004](#)). [Table 1](#) contains 9 variables identified as belonging to the productive state. Percentage values are the ratio of the totality of observations of each individual in each identified variable.

welders	welding	materials organizing	cleaning	making purge	changing diffuser	sanding	assessing task	adjusting spool	stamping
Welder 1	32,53	5,07	0,27	0,00	4,00	10,93	0,00	2,67	0,53
Welder 2	34,03	4,42	0,52	0,26	0,78	12,47	2,34	0,52	1,82
Welder 3	32,11	4,74	0,00	2,63	2,89	10,26	5,00	1,05	2,11
Welder 4	29,74	4,36	0,00	0,00	3,08	13,59	3,08	3,85	0,26
Welder 5	31,38	3,38	0,00	2,46	3,08	10,77	3,08	5,23	1,54
Welder 6	26,75	4,68	0,26	1,04	2,08	10,91	2,08	1,04	0,00
Welder 7	33,98	5,18	0,00	0,00	3,56	6,47	4,85	0,32	0,00
Welder 8	28,35	4,64	0,26	0,00	0,00	12,89	5,15	0,00	0,77
Welder 9	22,51	4,45	0,52	0,26	1,31	17,28	2,36	4,45	1,05
Welder 10	28,63	5,10	0,00	0,00	0,00	14,12	4,71	0,00	0,78

Table 1: Productive state

2.3. Nonproductive state

The nonproductive state is that when the worker is not performing any work activity. For example, a worker may be chatting with other workers, drinking water, or walking unnecessarily (Adrian 2004). Table 2 contains 10 variables identified as belonging to the nonproductive state. Percentage values are the ratio of the totality of observations of each individual in each identified variable.

welders	walking	mobilization	without joints	cargo handling	waiting crane	waiting inspector	human needs	analyzing specifications	security meeting	without machine
Welder 1	3,73	11,73	0,00	0,27	12,53	1,33	6,40	0,80	3,47	3,73
Welder 2	4,16	12,99	0,00	0,52	15,58	2,34	3,12	0,78	3,38	0,00
Welder 3	6,84	14,47	0,00	0,26	1,58	0,53	11,58	0,53	3,42	0,00
Welder 4	3,08	11,54	0,00	0,00	0,51	1,79	13,08	0,00	3,33	8,72
Welder 5	2,46	14,15	0,00	0,00	5,23	0,62	13,23	0,62	2,77	0,00
Welder 6	4,42	23,64	0,00	0,52	8,83	1,56	8,31	0,52	3,38	0,00
Welder 7	3,88	25,89	0,00	0,32	0,00	1,62	7,77	1,94	4,21	0,00
Welder 8	2,58	5,67	13,66	0,77	7,22	3,87	9,54	0,52	3,35	0,77
Welder 9	5,24	15,45	0,00	0,52	9,95	5,76	4,45	1,05	3,40	0,00
Welder 10	4,71	10,59	0,00	0,00	2,75	3,14	20,78	1,18	3,53	0,00

Table 2: Nonproductive state

3. Clustering Methodology

The purpose of the multivariate statistical analysis is to verify the formation of clusters with regard to welders and the actions collected that they performed during the welding process in the period of the field experiment. For the development of the statistical calculations, IBM® SPSS® Statistics version 25 software was used mainly, Microsoft Excel software was also used for graphical formatting of data from SPSS and also in the development of charts and figures. The methodology of this work begins by applying Principal Component Analysis (PCA) on the individuals data, in order to identify the trend of cluster formation and the existence of outliers, as well as which elements have the best projection and thus have a lower deformation. After that, the clustering method is developed by means of the hierarchical classification, obtaining the indication of the behavior of the clustering of individuals and identifying the number of clusters and which elements have greater capacity of connection between them. Finally, the analysis by non-hierarchical classification allows the evaluation of the dimensionality of the clusters formed by the individuals, where a synthesis analysis of the mathematical results is carried out in relation to the aspects of the labor productivity of the welders under study. Next, the methodology is shown step by step:

- a. Application of PCA to identify outliers, possible links tendencies and elements with better projection in the dimensions;
- b. Application of hierarchical classification with several connection methods to verify the natural formation of clusters;
- c. Analysis of the hierarchical classification regarding the trends of repetitive ramifications in dendrograms (indication of the formation of large clusters);
- d. Analysis of the hierarchical and non-hierarchical classifications in terms of the formation of strong connections between elements (always connected and, which do not divide even in a large number of clusters);
- e. Graphic development of the formation of clusters in the form of synthesis (indicate strong links and, to visualize clusters according to and close to the natural formation of clusters);
- f. Application of the non-hierarchical classification according to the number of natural clusters to perform a synthesis of practical analysis of how individuals clustering and what the average values of the variables in these clusters.

4. Results and Discussion

4.1. Welders clustering in the productive state

The Principal Component Analysis developed does not seek to retain individuals, firstly because of the small sample, and mainly because of the reason that all the elements collected are of extreme relevance for the analysis of welders' productivity under evaluation.

In [Figure 1](#), it can be seen that the dimension 1 is determined by individuals (1; 5; 3; 7; 10; 6; 8), and individuals (1; 5; 3; 7) at least over 0.5 in absolute value can be accepted as determinants, it can still be evaluated that individual (1) is strongly associated with dimension 1. The dimension 2 is determined by individuals (8; 10; 1; 7; 9; 4), and individuals (8; 10; 1) at least over 0.5 in absolute value can be accepted as determinants. It can still be verified that individuals (8; 10) are strongly associated with dimension 2. These individuals (8; 10) must present a great tendency of connection and formation of a common cluster, in counterpoint, the individual (9) is clearly an outlier and by tendency should not make connections and therefore must present itself in a single cluster.

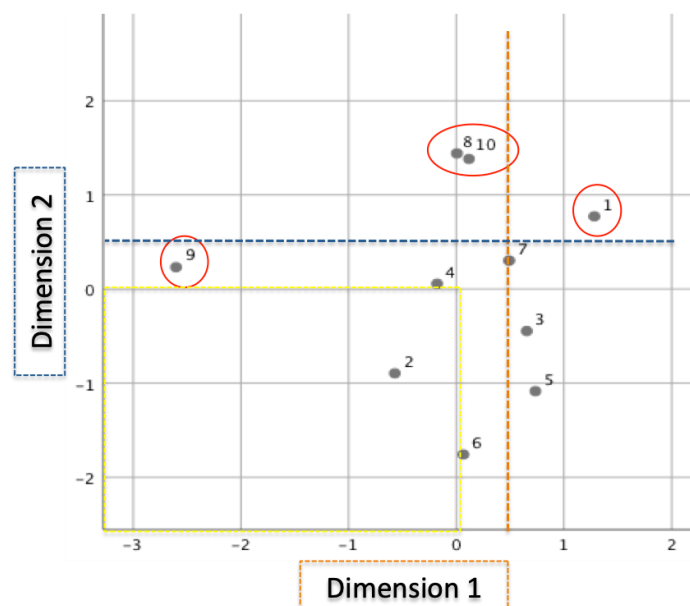


Figure 1: PCA productive state

Clustering methods differ in how the distances between groups are determined. According to [Marôco \(2011\)](#), the best recommendation is to use several agglomerative hierarchical methods simultaneously to verify if the cluster formations are "natural" (similar between the methods) or "artificial" (dissimilar between the methods). Thus, simulations were developed by: 1. Nearest neighbor; 2. Farthest neighbor; 3. Between-groups linkage; 4. Median clustering; 5. Centroid clustering; 6. Ward's method.

In the classification of individuals, based on the distance matrix, the Euclidean distance, as well as the squared Euclidean distance, was tested as a measure of an interval. However, the detailed studies were developed in the squared Euclidean distance because it performs a greater separation of different groups, which facilitates the graphical analyses of the dendrogram, a fact mainly identified in Ward's, centroid and median methods.

To evaluate how many clusters to retain in each simulation in addition to the visual analysis of the dendrogram, a graphic evaluation based on the distances between clusters was carried out ([Marôco 2011](#)). However, in all cases, the graphic evaluation based on the distances between clusters was observed to decide how many clusters to retain.

In [Figure 2](#), the clustering of welders according to their productive occupation is presented by the nearest neighbor method, which was retained 5 clusters, 3 clusters with one individual each, another two clusters with 4 and 3 individuals.

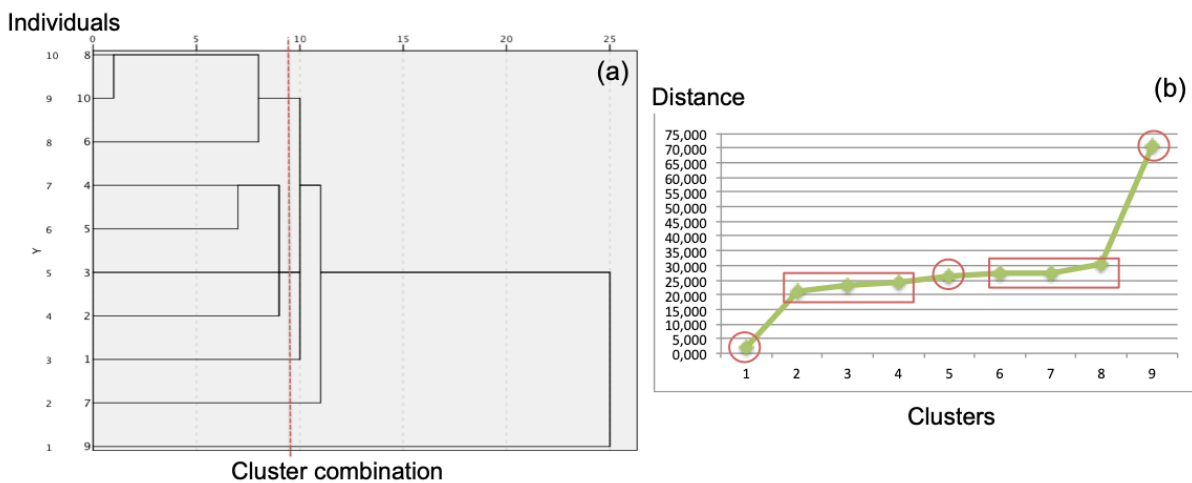


Figure 2: Clustering method by nearest neighbor, (a) dendrogram and (b) distances between clusters

In [Figure 3](#), the clustering of the welders according to their productive occupation is presented by the furthest neighbor method, which was also retained 5 clusters, 2 clusters with one individual each (9) and (7), one cluster with 2 individuals and another two clusters with 3 individuals. As it can be seen, the cluster with 3 individuals (6; 10; 8) is the same as that identified in the previous method. However, there is a division where the nearest neighbor method had a cluster with 4 individuals (2; 3; 5; 4), the furthest neighbor method has one cluster with individuals (3; 2) and another with (1; 5; 4).

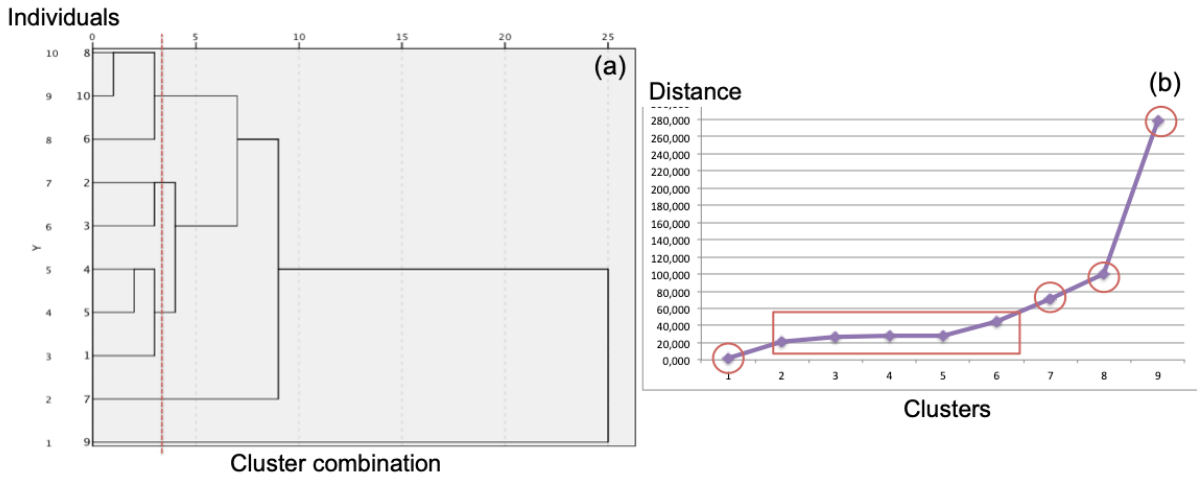


Figure 3: Clustering method by furthest neighbor, (a) dendrogram and (b) distances between clusters

Figure 4 shows the formation by the between-groups linkage method, which was also retained 5 clusters with exactly the same configuration as the furthest neighbor method.

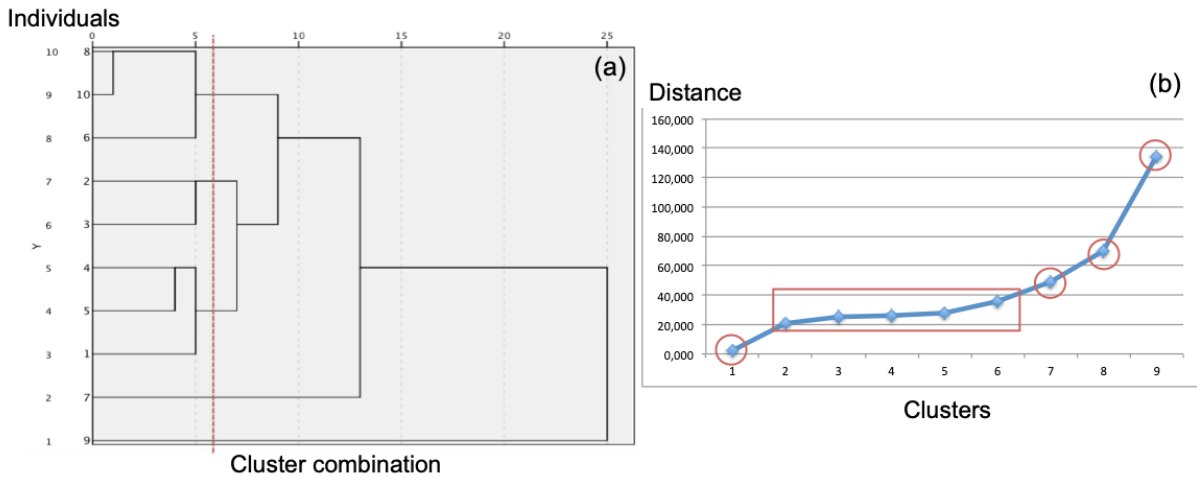


Figure 4: Clustering method by between-groups linkage, (a) dendrogram and (b) distances between clusters

In Figure 5, the clustering of welders according to their productive occupation is presented by the median clustering method, as in all previous cases, individuals (9) and (7) are in single clusters. It is also common among the above-mentioned methods the clustering with the individuals (6; 10; 8), and now in this method, the other individuals (3; 2; 1; 5; 4) being grouped into a single cluster.

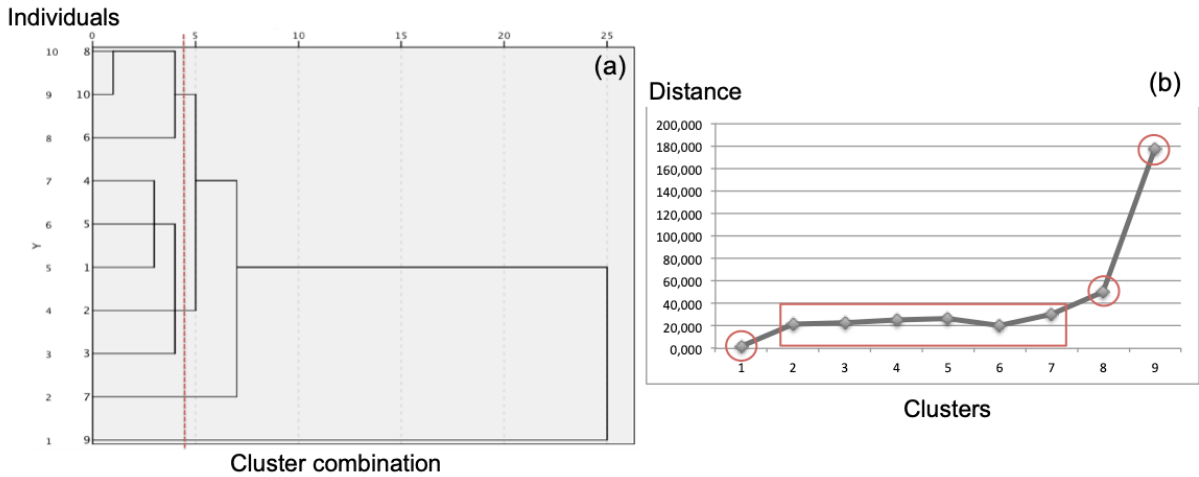


Figure 5: Clustering method by median clustering, (a) dendrogram and (b) distances between clusters

In [Figure 6](#), the centroid clustering method is shown that has exactly the same configuration as the median clustering method.

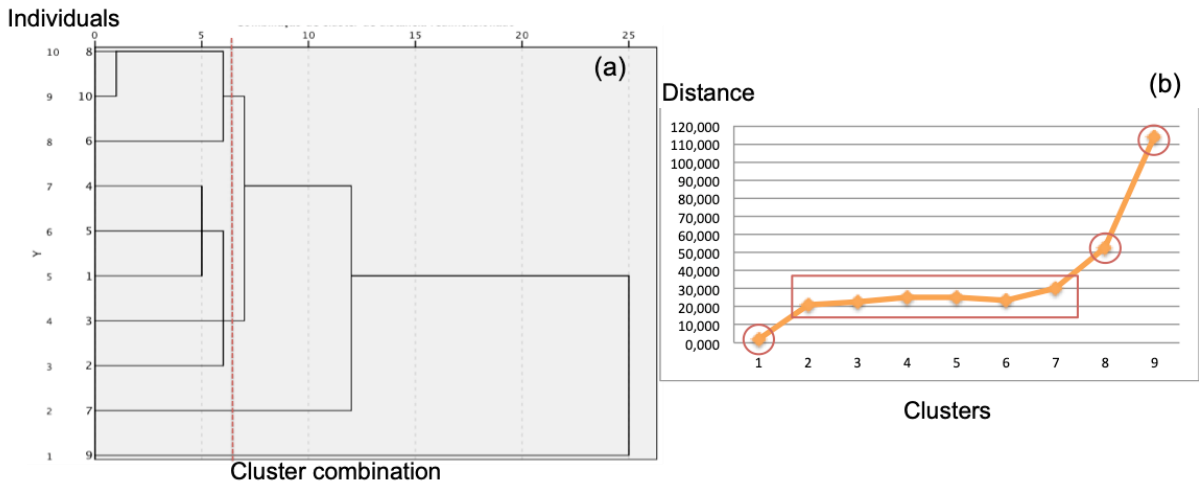


Figure 6: Clustering method by centroid clustering, (a) dendrogram and (b) distances between clusters

Finally, Ward's method has the same result determined by the furthest neighbor method and the between-groups linkage method, [Figure 7](#).

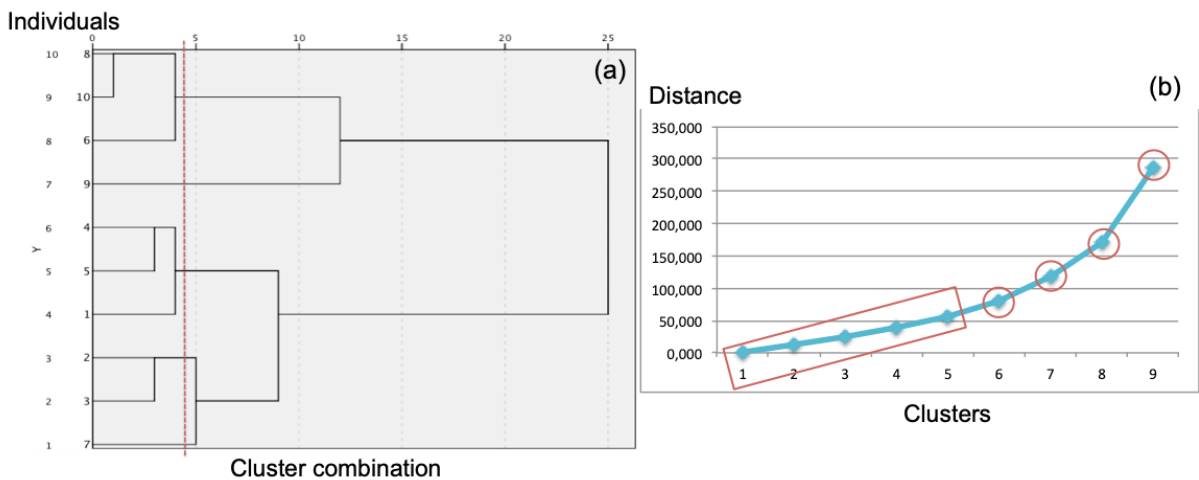


Figure 7: Clustering method by Ward's method, (a) dendrogram and (b) distances between clusters

As can be seen, there are small differences between the methods in the clusters formations, which indicate a "natural" formation of the clusters determined, with a tendency to form the number of 5 clusters.

A strong connection between the pairs of individuals (8; 10), (4; 5) and (3; 2) is highlighted. The individuals (9) and (7) do not have a tendency to bond with others but to form single clusters. The individual (6) has a strong connection with the individuals (8; 10), generating in the clusters always retained a single group with these (6; 8; 10). The individual (1) connects with (4; 5) and very often (1) form a cluster of these three individuals (1; 5; 4).

According to the Principal Component Analysis the individuals (6; 8; 10), (9) and (1) have a better projection, these main components tend to form a single cluster in the case of individuals (6; 8; 10), and not to connect in any group in the case of the individual (9), where only the individual (1) has facility of other connections. This analysis confirms the indications of PCA that indicate the individual (9) as an outlier, and a linking tendency among individuals (8; 10).

In summary, it can be observed a tendency of formation of 4 to 5 clusters with the individuals (9) and (7) not making connections and other two clusters with the strong grouping of the individuals (6; 8; 10) and different connections of the others (8; 10), (4; 5) and (2; 3) individuals, there are also strong pairs of links between individuals (1; 3; 2; 5; 4), [Figure 8](#).

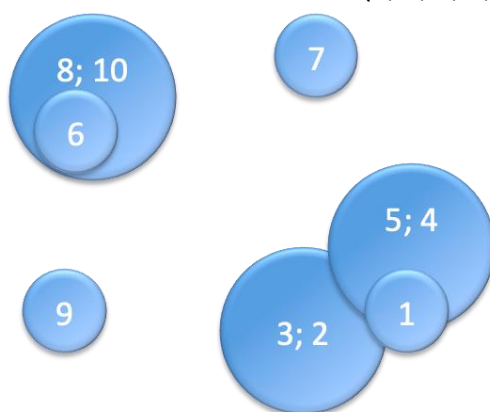


Figure 8: Welders clustering overview - hierarchical classification

The non-hierarchical classification was performed with a different "K" number of clusters aiming to ratify the formation of clusters according to the hierarchical classification previously performed, [Table 3](#). The synthesis of clusters formation is shown in [Figure 9](#), which is verified by the previous classification. The tendency of individuals (9) and (7) to not form cluster is clear, and once again there is a strong connection between pairs of individuals (4, 5) and (8; 10), the latter also tend to cluster in the cluster of individuals (6; 8; 10).

Number of Clusters	Welders to 4 clusters	Welders to 5 clusters	Welders to 6 clusters	Welders to 7 clusters	Welders to 8 clusters
1	7	9	1	1	1
2	1; 2; 3; 5	1; 2	2; 3	2	2
3	4; 6; 8; 10	7	9	3	3
4	9	6; 8; 10	4; 5	6; 8; 10	9
5		3; 4; 5	7	4; 5	4; 5
6			6; 8; 10	9	6
7				7	7
8					8; 10

Table 3: K clustering – non-hierarchical classification

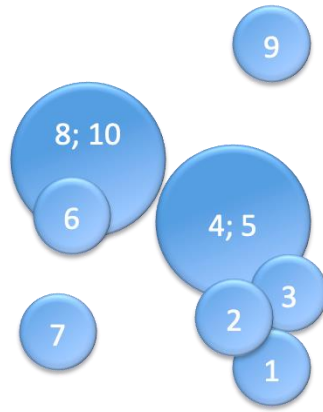


Figure 9: Welders clustering overview – non-hierarchical classification

Considering the possible graphical evaluation of the formation of clusters, and according to the trend of the formation of the number of 5 clusters in the hierarchical classification, this clustering configuration is adequate for a synthesis of the analysis and discussion of the results in order to interpret the characteristics related to the individuals and variables as presented in Figure 10 and Tables 4, 5, and 6.

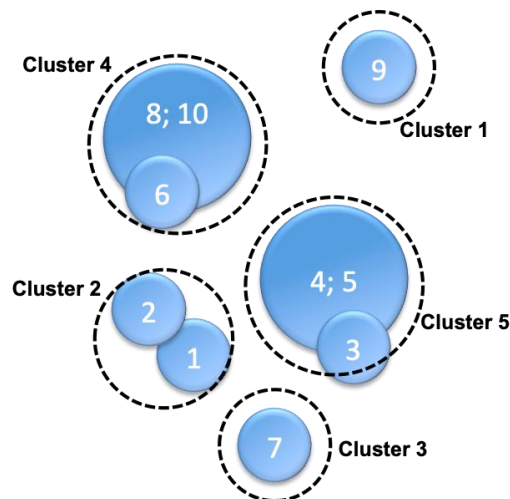


Figure 10: Welders clustering by non-hierarchical classification

Cluster association		
Welder	Cluster	Distance
1	2	2,609
2	2	2,609
3	5	3,406
4	5	3,260
5	5	2,405
6	4	3,334
7	3	0,000
8	4	1,567
9	1	0,000
10	4	2,034

Table 4: Cluster association

Distances between end cluster centers					
Cluster	1	2	3	4	5
1		12,566	16,690	8,429	10,709
2	12,566		6,792	6,525	4,232
3	16,690	6,792		9,187	7,115
4	8,429	6,525	9,187		5,363
5	10,709	4,232	7,115	5,363	

Table 5: Distances between end clusters centers

End Cluster Centers					
Activities	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
welding	22,51	33,28	33,98	27,91	31,08
materials organizing	4,45	4,74	5,18	4,80	4,16
cleaning	0,52	0,39	0,00	0,17	0,00
making purge	0,26	0,13	0,00	0,35	1,70
changing diffuser	1,31	2,39	3,56	0,69	3,02
sanding	17,28	11,70	6,47	12,64	11,54
assessing task	2,36	1,17	4,85	3,98	3,72
adjusting spool	4,45	1,59	0,32	0,35	3,38
stamping	1,05	1,18	0,00	0,52	1,30

Table 6: End clusters centers

Cluster 1 is the farthest from the others, the individual (9) previously identified, as a Principal Component clearly shows no tendency to link with any other individual. Cluster 1 (9) has an index of 22.51% on activity in the welding variable, a lower degree within the 5 identified clusters, another singular characteristic is to have the highest index among all of the sanding action of 17.28% and adjusting spool of 4.45%. It can be inferred that this fact is directly related to the type of welding performed by this one, where the process indicates a greater need for sanding execution between welding passes.

The cluster 5 (4; 5; 3) is close to the cluster 2 (2; 1) and cluster 4 (8; 10; 6), these three clusters have similar occupancy (% of action) in the welding and sanding variables. Cluster 3 (7) clearly sees itself further away from these clusters due to its low action occupation in the sanding activity (6.47%) approximately 45% smaller than clusters 5, 2 and 4. It can be inferred that the cluster 3 (7) also performs as cluster 1 (9) a type of welding process different from the other welders.

Beyond the type of welding, another extremely relevant factor in the welding process is relative to the diameter of the pipe to be welded. These two factors combined with the particular characteristics of each welder, such as concentration, dedication, and technical ability, together with the inherent imprecision aspects of data collection through human observation, impact the identified indices of the welding productive occupation, and thus, therefore, impact on the formation and interrelation of clusters. As a result, it is these factors that influence the formation and the linking tendency between cluster 5 (4; 5; 3), cluster 2 (2; 1) and cluster 4 (8; 10; 6), as well as the others clusters.

Most important, the identification of clusters of welders through the multivariate analysis makes it possible to segment these clusters based on their characteristics of productive actions. This is of great value in the evaluation of productivity, since it allows an analysis of

each specific group and allows the implementation of actions in process improvement appropriate to each group of individuals, increasing the effectiveness of these actions.

4.2. Welders clustering in the nonproductive state

In the same way that the Principal Components Analysis was performed for the productive state analysis, the analysis of the nonproductive state developed also does not seek to retain individuals, firstly because of the small sample, and mainly because of the reason that all the elements collected are extreme relevance to the analysis of the unproductivity of welders under evaluation.

In Figure 11, it can be verified that the dimension 1 is determined by the individuals (7; 10; 1; 2; 4; 9; 6); the individuals (7; 10; 1; 2) at least over 0.5 in absolute value can be accepted as determinants. The dimension 2 is determined by the individuals (10, 9, 5, 3, 8), and the individuals (10; 9) for having a component loading at least 0.5 in absolute value can be accepted as determinants, (10) is strongly associated with dimension 2. The individual (8) is clearly an outlier and by trend should not effect connections and therefore should present in a single cluster, also the individual (10) is distant graphically of the other individuals, what can be inferred as a non-tendency of connections.

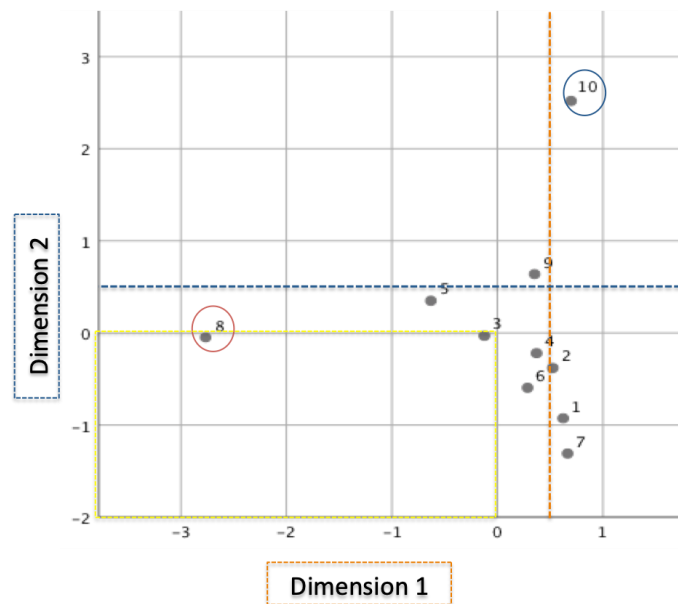


Figure 11: PCA nonproductive state

In Figure 12, the clustering of welders according to their non-productive state is presented by the nearest neighbor method, which was retained 7 clusters, 5 groups with one individual each, two other groups with 3 and 2 individuals.

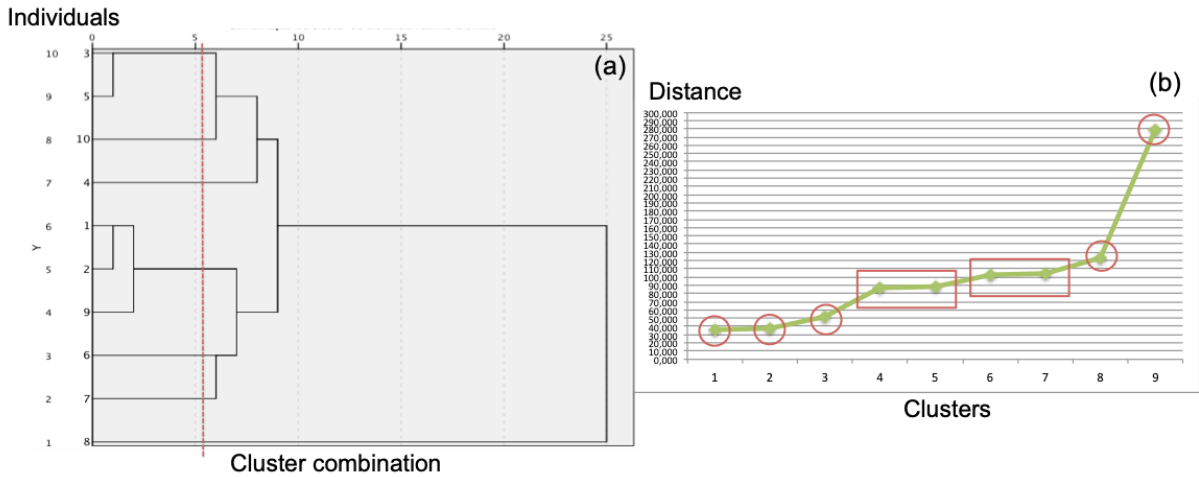


Figure 12: Clustering method by nearest neighbor, (a) dendrogram and (b) distances between clusters

In [Figure 13](#), the clustering of the welders according to their non-productive state is presented by the furthest neighbor method, where 5 clusters were retained, 2 clusters with one individual each (8) and (10), one cluster with 3 individuals and two clusters with 2 individuals. As can be seen, the cluster with 3 individuals (9; 2; 1) is the same as that identified in the previous method; however, clusters of individuals (7; 6) and (4; 5; 3) occur.

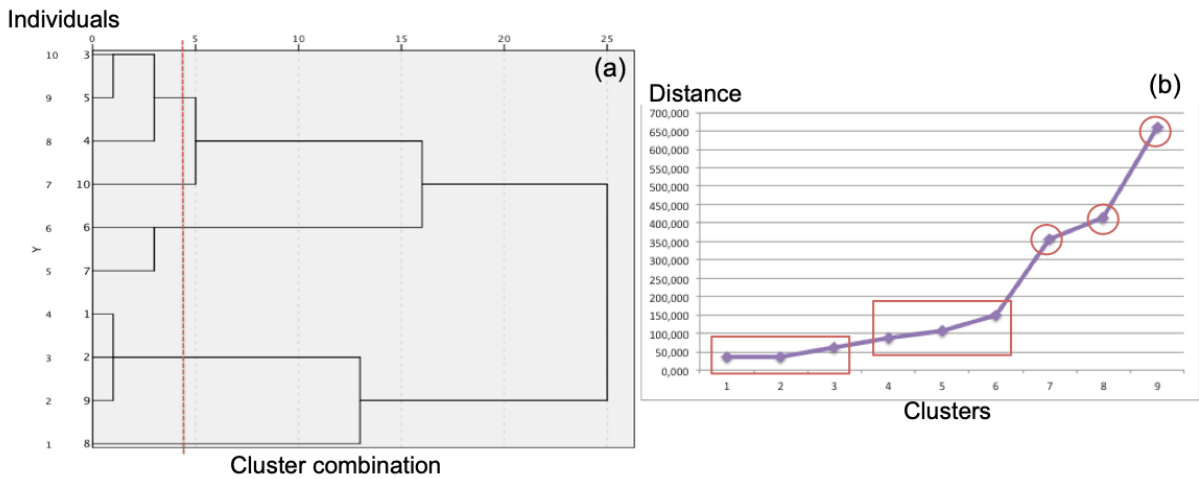


Figure 13: Clustering method by furthest neighbor, (a) dendrogram and (b) distances between clusters

[Figure 14](#) shows the formation by the between-groups linkage method of 6 clusters with almost the same configuration as the furthest neighbor method, only slightly altering the configuration of the dendrogram and the individual (4) forming a single cluster.

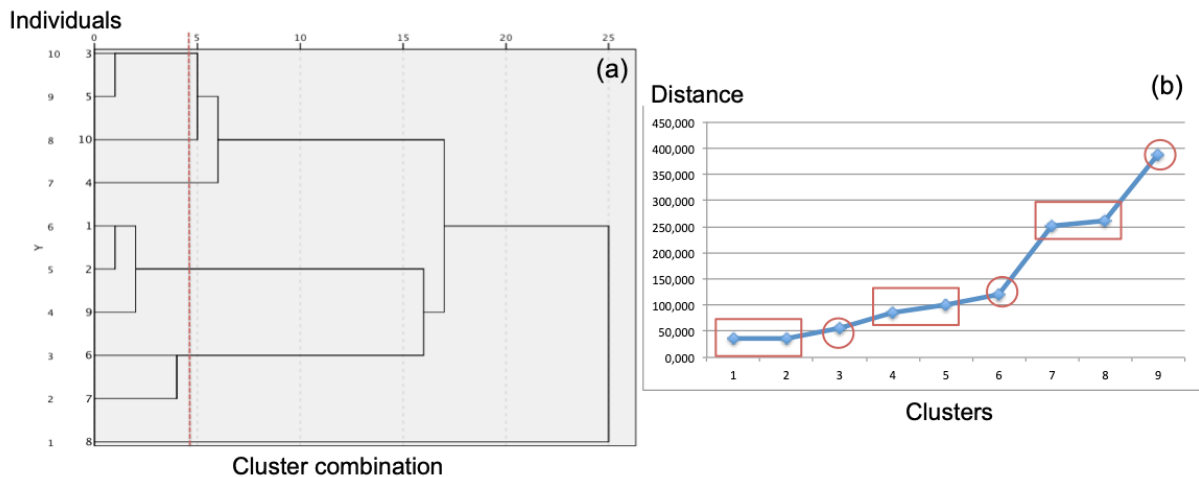


Figure 14: Clustering method by between-groups linkage, (a) dendrogram and (b) distances between clusters

In the median clustering method there is also the formation of 6 clusters, [Figure 15](#), and also exactly the same formation of clusters of the between-groups linkage method.

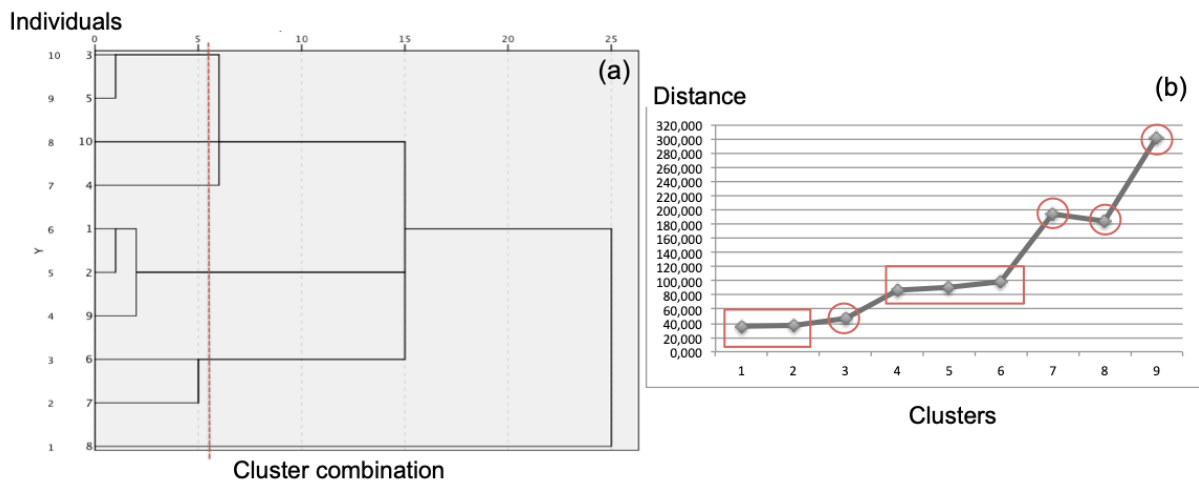


Figure 15: Clustering method by median clustering, (a) dendrogram and (b) distances between clusters

[Figure 16](#) shows the centroid clustering method that has exactly the same configuration as the between-groups linkage method and the median clustering method.

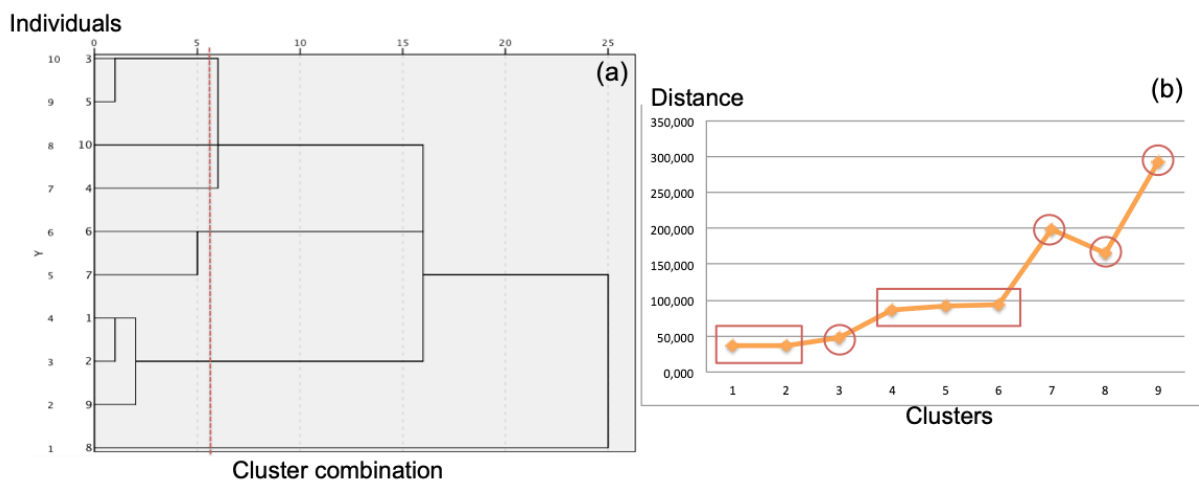


Figure 16: Clustering method by centroid clustering, (a) dendrogram and (b) distances between clusters

Finally, in Ward’s method has the same result determined by the methods of the between-groups linkage method, the median clustering method and, the centroid clustering method, [Figure 17](#).

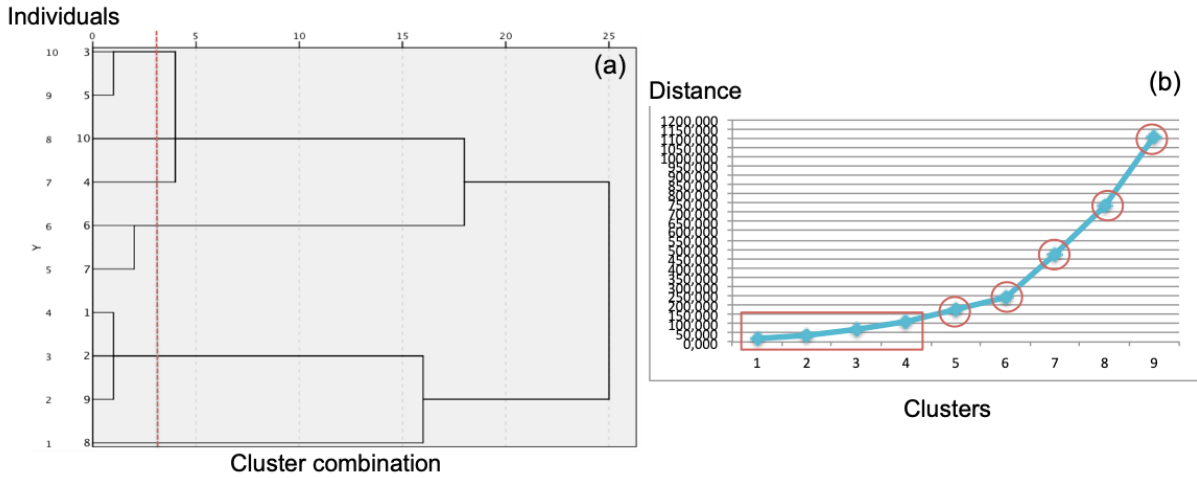


Figure 17: Clustering method by Ward’s method, (a) dendrogram and (b) distances between clusters

As can be seen, there are small differences between the methods in the clusters formations, which indicates a "natural" formation of the determined clusters, with a tendency to form the number of 6 clusters. There is a strong connection between the pairs of individuals (3; 5), (1; 2) and (6; 7). The individual (8) does not have a tendency to bind but to form single cluster. The individual (9) has a strong connection with the individuals (1; 2), generating the tendency of forms a single cluster with these (1; 2; 9). Individuals (10) and (4) connect with (3; 5) and very often one cluster is form with these four individuals (3; 5; 10; 4).

In synthesis, it can be observed a tendency of formation of 6 clusters, and in a broader view a formation of 4 biggest clusters. The individual (8) clearly does not have a tendency to connect with other individuals. The individuals (6; 7) always present themselves together in the formation of a single cluster. As the individuals (9; 2; 1) always present themselves together in the formation of a single cluster. The individuals (3; 5; 10; 4) can form a single cluster or 3 clusters by disconnection of 2 individuals (10) and (4), [Figure 18](#). This analysis confirms the indications of the PCA that indicates the individual (8) as an outlier.

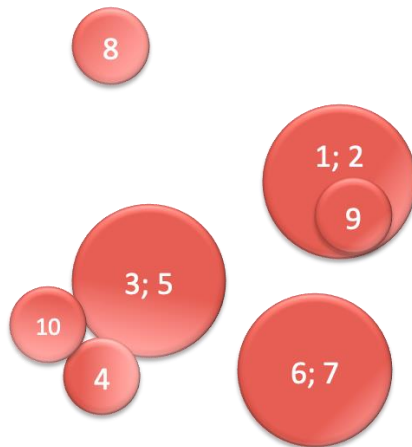


Figure 18: Welders clustering overview - hierarchical classification

The non-hierarchical classification was performed with different "K" number of clusters, [Table 7](#), aiming to ratify the formation of clusters according to the hierarchical classification previously performed. The synthesis of clusters formation is shown in [Figure 19](#), which is

verified by the previous classification. It is clear the tendency of non-connection of the individual (8) to form clusters. The pair of individuals (1; 2) have a strong link between them and a tendency to form a cluster of 3 individuals (1; 2; 9). The individuals (3; 5) have a strong connection and could connect to 2 other individuals (4) and (10) with this forming a cluster of 4 individuals (3; 5; 4; 10). Finally, the pair of individuals (6; 7) has a tendency to connect with each other and as the increase in the number of clusters can separate into single clusters.

Number of clusters	Welders to 3 clusters	Welders to 4 clusters	Welders to 6 clusters	Welders to 7 clusters	Welders to 8 clusters
1	3; 4; 5; 10	6; 7	8	10	10
2	1; 2; 8; 9	1; 2; 9	1; 2; 9	1; 2; 9	1; 2
3	6; 7	8	6; 7	3; 5	3; 5
4		3; 4; 5; 10	4	4	4
5			3; 5	8	9
6			10	6	6
7				7	7
8					8

Table 7: K clustering – non-hierarchical classification

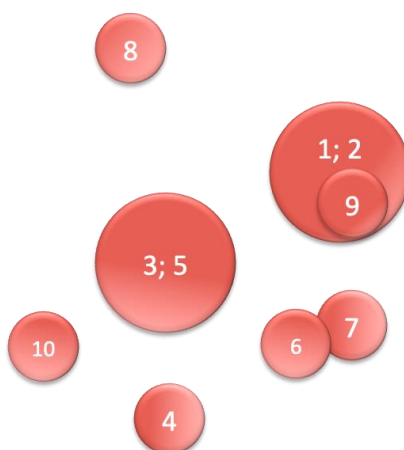


Figure 19: Welders clustering overview – non-hierarchical classification

As indicated in the hierarchical classification the tendency is the formation of 6 clusters, then this retention of clusters in the non-hierarchical classification is adequate for a synthesis of the analysis and discussion of the results in order to interpret the characteristics related to the individuals and variables in the nonproductive state, as presented in [Figure 20](#) and [Tables 8, 9, and 10](#).

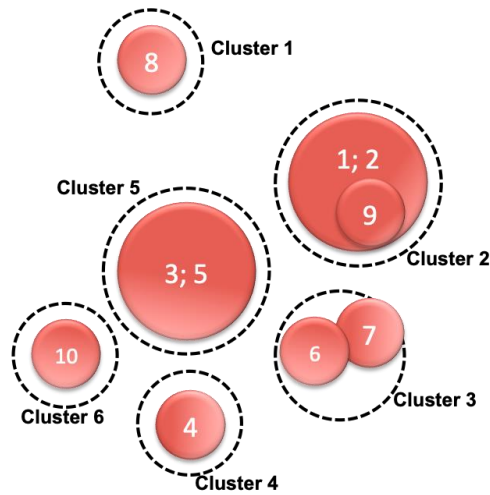


Figure 20: Welders clustering by non-hierarchical classification

Cluster association		
Welder	Cluster	Distance
1	2	3,966
2	2	3,630
3	5	2,994
4	4	0,000
5	5	2,994
6	3	4,648
7	3	4,648
8	1	0,000
9	2	4,577
10	6	0,000

Table 8: Cluster association

Distances between end cluster centers						
Cluster	1	2	3	4	5	6
1		17,440	23,881	18,629	17,329	19,079
2	17,440		14,614	16,820	12,475	19,204
3	23,881	14,614		17,164	11,470	19,212
4	18,629	16,820	17,164		9,844	12,131
5	17,329	12,475	11,470	9,844		9,576
6	19,079	19,204	19,212	12,131	9,576	

Table 9: Distances between end clusters centers

Activities	End Cluster Centers					
	Cluster					
	1	2	3	4	5	6
mobilization	5,67	13,39	24,76	11,54	14,31	10,59
human needs	9,54	4,66	8,04	13,08	12,40	20,78
waiting crane	7,22	12,69	4,42	0,51	3,40	2,75
walking	2,58	4,37	4,15	3,08	4,65	4,71
security meeting	3,35	3,42	3,79	3,33	3,10	3,53
waiting inspector	3,87	3,14	1,59	1,79	0,57	3,14
analyzing specifications	0,52	0,88	1,23	0,00	0,57	1,18
cargo handling	0,77	0,44	0,42	0,00	0,13	0,00
without joints	13,66	0,00	0,00	0,00	0,00	0,00
without machine	0,77	1,24	0,00	8,72	0,00	0,00

Table 10: End clusters centers

It can be verified that the formation of clusters is directly related to an "anomaly", whether it is a time spent excessively in an action differently from the other individuals. In cluster 1, the individual (8) was the only one who did not have joints available to perform work. The individual (4) in the cluster 4 was clearly impacted by a problem in his soldering machine, being also impossible to perform work. In cluster 6 we have the individual (10) who spent twice as much time as the others with human needs, this variable consolidates WC actions, drinking water, cooling the body and interacting with colleagues. Cluster 3 that contains the individuals (6; 7) presents an excessive time spent in initial mobilization. In turn, cluster 2 with individuals (1; 2; 9) was more impacted by the crane's delay in moving the large diameter pipes to the workstations of these welders. Finally, cluster 5 composed of individuals (3; 5) did not present any time spent in an extreme way in a single variable.

When the nonproductive state is evaluated through others analysis as for example Pareto analysis and sensitivity analysis that generalizes the big "problems" of all sample, and distortions could happen. On the other hand, the multivariate analysis approach generates the opportunity for a better understanding of the main impacts, and consequently how individuals and groups of individuals are affected by a certain variable. In this way, the multivariate analysis allows the evaluation and implementation of actions in process improvement appropriate to each cluster of individuals, increasing the effectiveness of these actions.

5. Conclusions

Through the multivariate analysis in the evaluation of the productive state, it is possible to identify the natural formation of 5 clusters of welders and this clustering is shown pertinent based on the evaluation of the characteristics of the variables of each center of these final clusters. It is clear the clustering of the welders was drive by the actions connected to the welding process; in that way, actions to boost de occupancy of the productive state must focus in improve the technics of welding.

As far as the evaluation of the nonproductive state is possible to identify the natural formation of 6 clusters of welders and also this grouping is pertinent based on the evaluation of the characteristics of the variables of each center of these final clusters. The clustering of welders in the nonproductive state was drive by anomaly actions; the precise diagnosis by the

multivariate analysis facilitates the adequate improvement actions for the appropriate patient group.

As presented, 50% and 40% of the total sample was retained in clusters, respectively relative to the productive and nonproductive states. Therefore, according to convenience, or for example, in the case of a greater sampling, the developed method may be repeated in this sense by forming sub-clusters. In this way, according to the analysis of the results and convenience to the management, it is possible to seek an adequate number of clusters to act on improvement actions.

The welders clustering in the productive and nonproductive states were different. The clustering of welders regarding their labor productivity performance is directly related to the identified direct work of each individual in each activity. The interpretation of the formation of clusters of welders and the identification of which variables drives that is extremely important for the analysis and understanding of the productive and nonproductive states of these. Also, for example, in the pursuit for continuous improvement of the production process, the creation of these clusters will first prevent each worker from demanding a specific action plan, as well as avoiding that the actions are applied in a generic sense, in that way not interpreting the singular characteristics of each worker or group of workers. Therefore, the multivariate analysis regarding cluster formation is a relevant tool to increase the effectiveness of productivity improvement actions.

References

- Adrian, James J. 2004. *Construction productivity: Measurement and improvement*. Champaign, IL: Stipes Publishing.
- Alder, G. Stoney. 2001. "Employee reactions to electronic performance monitoring: A consequence of organizational culture". *The Journal of High Technology Management Research* 12, no. 2 (september): 323-42. [https://doi.org/10.1016/S1047-8310\(01\)00042-6](https://doi.org/10.1016/S1047-8310(01)00042-6).
- Christian, John, and Daniel Hachey. 1995. "Effects of delay times on production rates in construction". *Journal of Construction Engineering and Management* 121, no. 1 (march): 20-26. [https://doi.org/10.1061/\(ASCE\)0733-9364\(1995\)121:1\(20\)](https://doi.org/10.1061/(ASCE)0733-9364(1995)121:1(20)).
- Jarkas, Abdulaziz M., and Camille G. Bitar. 2012. "Factors affecting construction labor productivity in Kuwait". *Journal of Construction Engineering and Management* 138, no. 7 (october): 811-20. [https://doi.org/10.1061/\(asce\)co.1943-7862.0000501](https://doi.org/10.1061/(asce)co.1943-7862.0000501).
- Marôco, João. 2011. *Análise estatística com o SPSS Statistics*. Pêro Pinheiro: Report Number - Análise e Gestão da Informação.
- Project Management Institute. 2013. *A guide to the project management body of knowledge: PMBOK® guide*. [Philadelphia], Pennsylvania: Project management Institute, Inc.