

Cluster Analysis for Data Processing in Educational Research

Daniel Morin Ocampo^a
 Luiz Caldeira Brant de Tolentino Neto^b

^a Universidade Federal de Santa Maria (UFSM), Coordenadoria de Educação Básica, Técnica e Tecnológica (CEBTT), Colégio Técnico Industrial de Santa Maria (CTISM), Departamento de Ensino, Santa Maria, RS, Brasil.

^b Universidade Federal de Santa Maria (UFSM), Centro de Educação, Departamento de Metodologia do Ensino, Santa Maria, RS, Brasil.

Received for publication on 22 Mar. 2019. Accepted, after revision, on 18 Apr. 2018.

Assigned editor: Renato P. dos Santos.

ABSTRACT

Quantitative approaches to educational research have been undervalued and consequently less widely used. In this sense, this paper aims to present and analyze the techniques of Cluster Analysis as a possibility for research in sciences area. Therefore, the main hierarchical and non-hierarchical techniques of Cluster Analysis are presented, as well as some of their applications in educational research found in the literature. Cluster Analysis is adequate to simplify or elaborate hypotheses on massive data, such as large-scale educational research. The studies in the area of education that used Cluster Analysis methods proved to be fruitful to elicit results that collaborate with the area.

Keywords: Cluster Analysis; Educational Research; Quantitative Method.

A Análise de Cluster para o Tratamento de Dados em Pesquisas da Área do Ensino

RESUMO

As abordagens quantitativas para pesquisas no ensino têm sido pouco valorizadas e, consequentemente, menos utilizadas. Nesse sentido, este artigo objetiva apresentar e analisar as técnicas de análise de cluster como possibilidade para pesquisas na área de ciências. Desta maneira, são apresentadas as principais técnicas hierárquicas e não hierárquicas de análise de cluster, bem como algumas de suas aplicações em pesquisas da área de ensino encontradas na literatura. A análise de cluster é adequada para simplificar ou elaborar hipóteses em dados massivos, como os de pesquisas educacionais de larga escala. Os estudos da área de ensino que utilizaram os métodos da análise de cluster mostram-se proficuas para suscitar resultados que colaboram com a área.

Palavras-chave: Análise de cluster; Pesquisa em Ensino; Métodos Quantitativos.

Corresponding author: Daniel Morin Ocampo. E-mail: daniel.ocampo@ufsm.br

INTRODUCTION

Researches based on quantitative analysis of data has lost space in research in educational and teaching areas, especially since the 1980s, because, justifiably, the number of papers with a qualitative approach has received more emphasis in these areas. This occurs because researchers in these areas have been interested in other aspects than student performance, such as social interactions at school, curricular issues and pedagogical work (Dal-Farra & Fetters, 2017). It should be emphasized that education and teaching are Human and Social sciences (as such, they make use of essentially qualitative objects), and work with the “universe of meanings, motives, aspirations, beliefs, values, and attitudes” (Minayo, 2000, p.22).

However, the characteristic of the qualitative approach is also its limitation, because investigating the population or a representative sample, as in the case of large-scale studies, makes impracticable to use qualitative methods for data analysis, since qualitative methods seek to comprehend the whole through the part. Large-scale studies that focus on (for example) students’ opinions, interests, and difficulties are fundamental, both to generate a corpus of knowledge that subsidizes other research and to the elaboration of public education policies.

In quantitative research, the researcher seeks to analyze the phenomena in the visible region (Minayo, 2000), that is, the quantitative approach is based on objectivity and therefore is commonly associated with positivism. In this sense, quantitative research plays a fundamental role in generating reality overviews, often focusing on representative samples of a given population, through numbers that may serve as a scope for later qualitative research.

Although there are differences in objectives and epistemological perspective, as both approaches describe the situations of teaching and education, the discussion is currently less polarized in “quantitative versus qualitative”, and more focused on how one approach limitations can be solved by the other (Dal-Farra & Fetters, 2017).

This paper aims to present the Cluster Analysis as a possibility for data analysis in the educational area researches, considering the advantages and limitations of quantitative research. Therefore, the most commonly used methods of Cluster Analysis will be presented, showing examples found in the literature of studies that have made use of these techniques. Initially, we will introduce the concept of Cluster Analysis and its most common applications.

WHAT IS AND WHEN CLUSTER ANALYSIS CAN BE USED

Cluster Analysis or grouping analysis (clustering) is the name given to a series of methods of multivariate data analysis, which seeks to approximate objects (which can be respondents of instruments, companies, products, etc.) into homogeneous groups, according to their characteristics in multiple dimensions (Hair et al., 2009; Henry,

Tolan & Gorman-Smith, 2005). These clusterings can serve as a basis for understanding characteristics of a certain population, which has shown high value in large-scale researches in several areas, such as epidemiological studies, market profile analysis or even biological taxonomy, to classify groups of animals and vegetables. Moreover, even if little explored in education and teaching, it is necessary to highlight this method potential for these areas, especially in large-scale research.

Hair (2009) points out that Cluster Analysis does not have an end in itself, but plays a role in conceptual development. Usually, Cluster Analysis is used to reduce data (when there is a high number of observations so that without partitions the data loses meaning) or the generation of hypotheses (if the researcher wishes to formulate or examine a hypothesis).

Data reduction has been an essential factor for the use of Cluster Analysis in studies that focus on public education policies and educational management. As in the study of Sant'Anna (2013), which sought to classify the municipalities of Rio de Janeiro as to the supply of day care, educational expenditures, and the proportion of children enrolled in daycare centers. Such researches help public administration and decision-making in public policies and serve as a reference for other researchers as, instead of analyzing all 90 municipalities studied as a single conglomerate, it looks at each of the six resulting groups on the research.

Likewise, researches that use Cluster Analysis for profiling students, teachers or institutions have been widespread. This fact corroborates the premise of Cluster Analysis for generating hypotheses. An example that illustrates this situation is the research of Freire and Motokane (2016), who elaborated four profiles of teachers according to their conceptions on science and ecology. From the results of this study, it was possible to infer a series of hypotheses, such as the generalization of those profiles to different teachers publics.

In the Norwegian context, Schreiner (2006) used Cluster Analysis to elaborate typologies of 15-year-old students in relation to their interest in science and technology. This approach generated five groups: Non-Selective Enthusiasts, with a high interest in all the subjects studied; Non-Selective Reluctant, who have a low interest in all science and technology topics investigated; Non-Selective Undecided, who have no high or low interest in any of the themes; Selective Girls, a group consisting predominantly of girls who claim to have a high interest in health issues; and finally, Selective Boys, predominated by boys who stands out for the high interest in engineering and technology. Schreiner's study brings relevant data to public education policies and to the recurrent discussion of the Scandinavian countries about the low interest of young people in pursuing scientific careers.

In Ocampo's research (2018), five typologies of Brazilian students were elaborated (with the support of Cluster Analysis) in relation to their positions regarding environmental challenges. When the typologies found were compared with the characteristics of the current typologies for environmental education

(Sauvé, 2005), it was possible to note the influence of different theoretical currents in each typology, demonstrating the importance of environmental education in the construction of the environmental challenges.

Either to reduce the data or to create hypotheses, Cluster Analysis is employed when the researcher aims to approach the objects by their similarities and to distance them by their differences in relation to the chosen characteristic. Thus, it is important that the objects of the same clustering resemble the chosen characteristic, that is, each clustering must have high internal homogeneity. In the same way, each grouping must present high external heterogeneity, that is, objects of distinct groupings must differ by the chosen characteristic. For this, Hair (2009) points out three important care for Cluster Analysis: firstly, how similarity will be measured; secondly, how the groupings will be formed (which method best fits the research) and finally, how many groups will be formed. To solve the question of the measure of similarity we need to find a way to measure the proximity between the objects.

MEASURES OF SIMILARITY MEASUREMENT

The literature points to many forms of measurement (Henry et al., 2005; Hair et al., 2009; Corrar et al., 2012), to cite the most commonly employed:

Euclidean distance: The measure of a straight line connecting two points P_1 and P_2 in an n-dimensional space. This measure can be calculated by the formula:

$$Distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \dots}$$

where x, y, z... are the coordinates of the points P_1 and P_2 in each of the dimensions in the n-dimensional space.

Squared Euclidean distance: This measure is calculated by the square of the coordinate differences of points P_1 and P_2 . Strictly speaking, it is the Euclidean distance itself elevated to the square that eliminates the root extraction operation. This simplification facilitates data computational modeling since the root extraction demands higher processing.

Manhattan distance: This distance, also called “city-block distance”, in an analogy, is like the route taken by a pedestrian in a city, skirting blocks. The Manhattan distance value is calculated by the module of the coordinate differences of points P_1 and P_2 , which makes the calculation very simplified. However, it can generate complications in successive applications. As shown in Figure 1, the Euclidean distance presents a geometric representation more intuitive when compared to Manhattan distance.

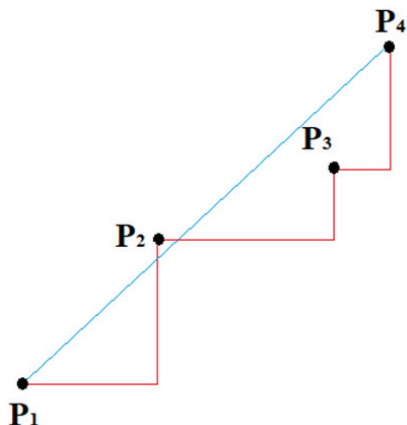


Figure 1. Example of the distinction between Euclidean distance and Manhattan distance. In blue the Euclidean distance and in red the Manhattan distance.

Chebyshev distance: This metric is calculated by the maximum absolute difference between the points of the group, that is, $D(P1, P2) = \max_i(|P1 - P2|)$. For example, a circle of radius “r” would be represented in Chebyshev distance as a “2r” side rectangle. Hair (2009) points out that this distance measure is sensitive to scale changes, which leads to the importance of data standardization.

Correlation coefficient: The use of the correlation coefficient as a measure comes from the researcher’s choice for a search for patterns and not for magnitude. Since (usually) Cluster Analysis offers higher emphasis on measurements, the use of this type of measurement is rare. A factorial analysis usually follows the correlation coefficient, which hinders its computational processing.

There are a variety of other ways of measuring the distance between the objects one wishes to group, each appropriate for different situations. In most cases, especially in teaching and educational research, Euclidean distance and squared Euclidean distance are the best options, the first for its intuitive geometric representation and the second for being merely a simplification of the former. Henry (2005) suggests the use of Euclidean distance considering that most of its inconveniences can be solved when standardizing the data.

To illustrate, let us observe the case in Figure 2. In this situation, through Euclidean distance is simple to intuitively elaborate 3 clusterings (circled in green, blue, and red) for the objects, because their distance in relation to the two variables (x, y) presented suggests the agglomeration in 3 clusters. Seemingly, these groups present high internal homogeneity and external heterogeneity.

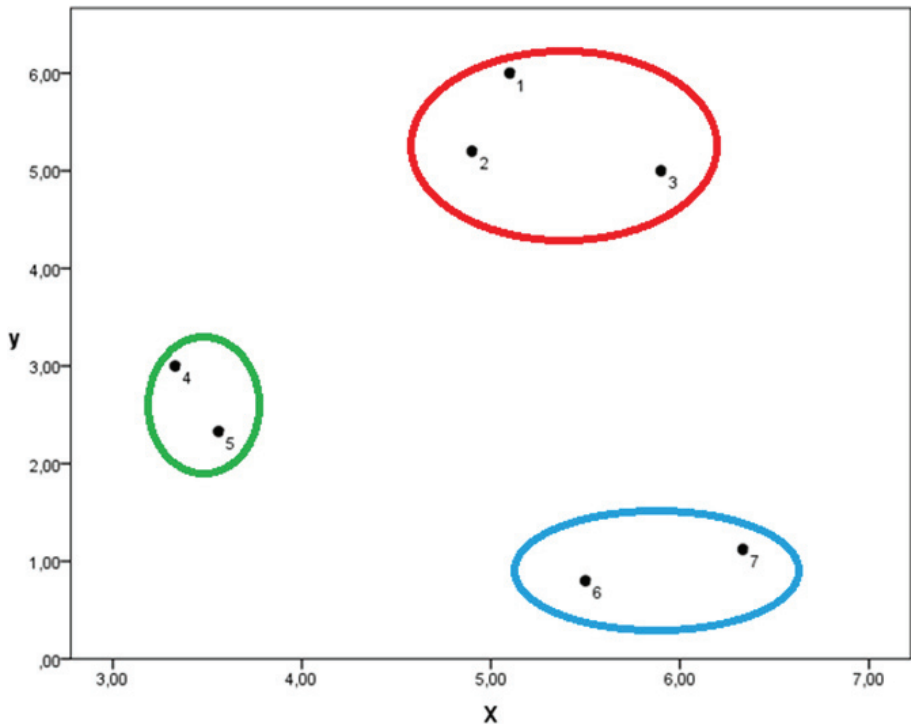


Figure 2. Illustration to exemplify Cluster Analysis. We illustrate the groupings of objects 1,2,3 – 4,5 – 6,7 marked by the blue, yellow and red circumscriptions, respectively.

A representation as shown in Figure 2 is not always possible, since in a few cases the objects are measured in only two variables (both chart axes), so it is called multivariate analysis. For this reason, the most common graphic representation is the dendrogram, a diagram in which connects the objects that group together first, thus forming a kind of “tree” (in Greek, “dendron”). For example, we elaborated Figure 3, a dendrogram that represents the same situation presented in Figure 2.

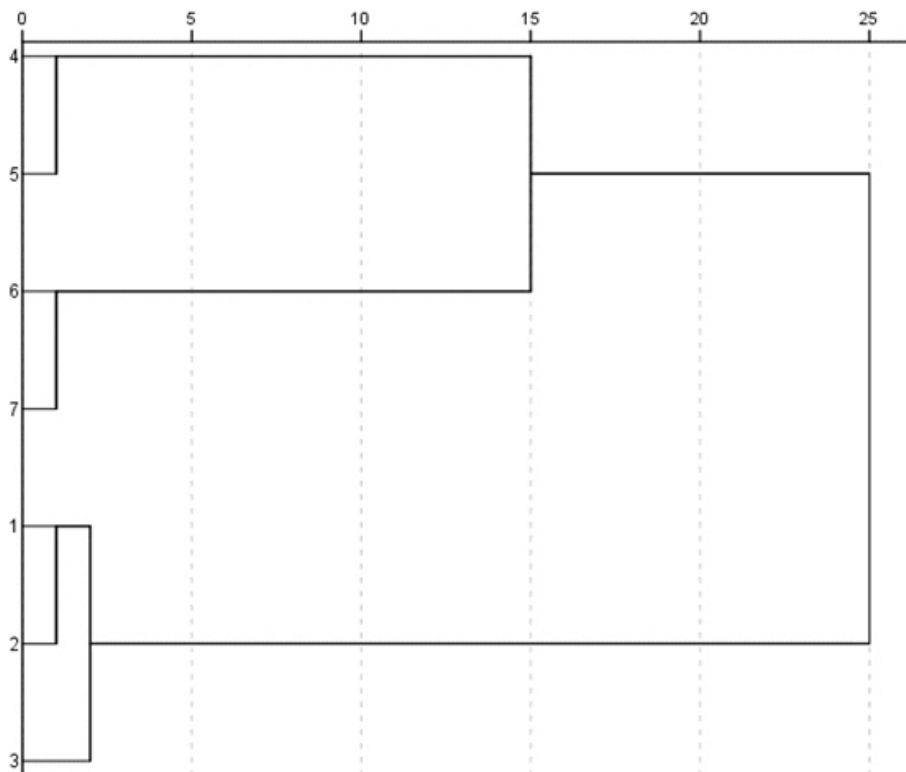


Figure 3. Dendrogram representative of the situation illustrated in Figure 2.

It is possible to observe that in this dendrogram (Figure 3) three groups are formed quickly, precisely because of their proximity. However, after some interactions (horizontal axis), we can observe the agglomeration of two groups, which results in a group with objects 4, 5, 6, and 7 and another with items 1, 2, and 3.

Thus, note that just finding a way to measure proximity is not enough. Hair (2009) points out that a consistent and coherent grouping algorithm is also necessary, as well as a choice based on the number of clusters, as shown in stage 4 of Figure 4.

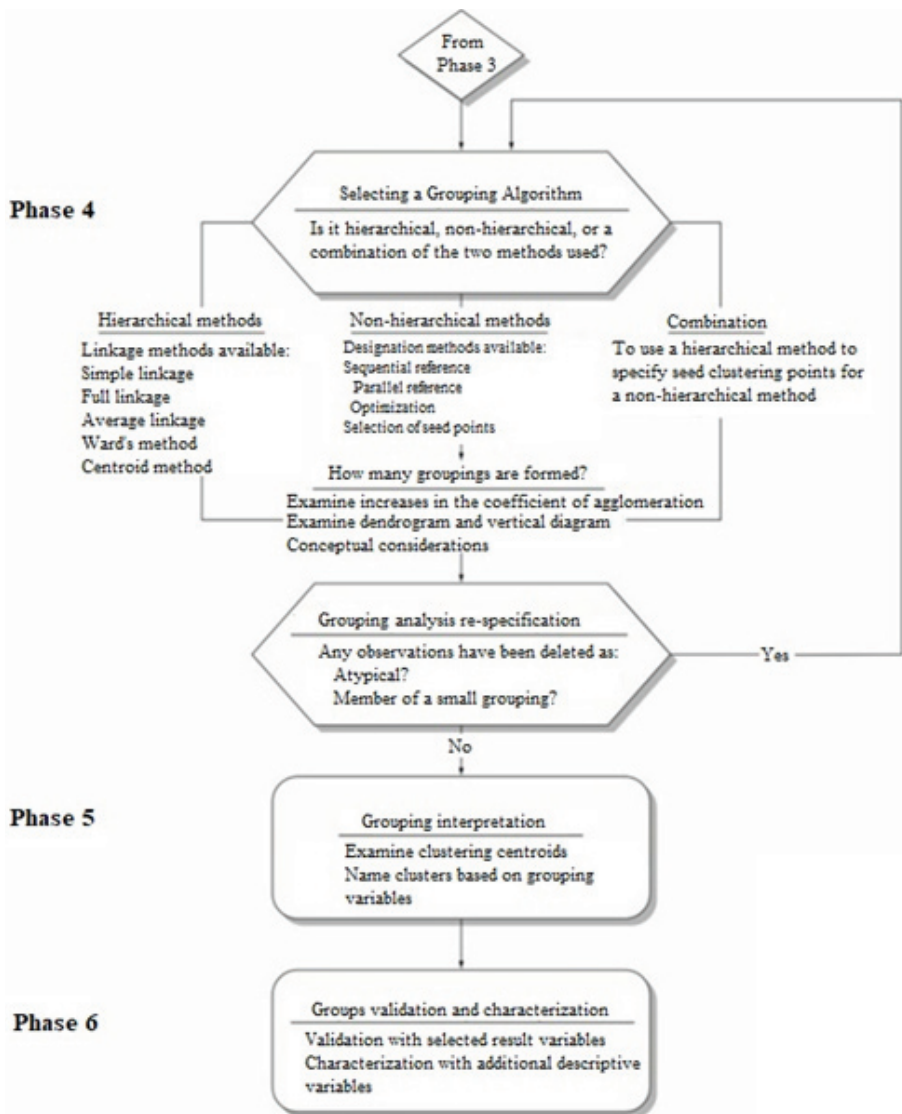


Figure 4. Stage of choice of the clustering method. (Hair et al., 2009, p.448)

On choosing the right amount of groupings, there are some techniques that help, such as the Elbow method.¹ However, this stage of Cluster Analysis is endowed with

¹ This method consists in the elaboration of a graph of dispersion of the variance percentage in relation to the number of clusters, since, by definition, it generates a decreasing monotonic curve, the intention is to try to form an “elbow” in the distribution of points. See Kodinariya & Makwana (2013) for more on this theme.

subjectivity, since the researcher, exploring the data, knowing the characteristics of its production and running simulations, chooses which is the number of groups more adequate for the research (Henry et al., 2005; Hair et al., 2009; Ocampo et al., 2018).

Moreover, it is necessary to define how to elaborate the groupings. Therefore, it is essential to choose the appropriate clustering method to group the data. Henry (2005) points out that is possible to classify the most used clustering methods into two categories: hierarchical and non-hierarchical. Note that the choice of the method used is fundamental to avoid errors arising from the method inadequacy for the sample.

HIERARCHICAL METHODS

Hierarchical methods consist of creating a hierarchy that, according to Corrar (2012), approaches – again – the image of a tree. It is possible to construct this tree in two ways: by grouping the objects or dividing successively a group of objects until eventually all will be separated. The most common is the use of agglomerative methods, where we begin with each object in isolation, that is, n groups of size 1, and group them until we obtain a single group of size n , as shown in Figure 5.

It is important to note that in the hierarchical methods, the result of the previous step aligns to the result of the later step, which generates the tree configuration. As seen in Figure 5, the group formed in Step 2 is aligned to form the group of Step 3. Due to this dependence of each cluster and the need for $n-1$ interactions, the hierarchical algorithms are not suitable to be applied in samples with a high number of objects.

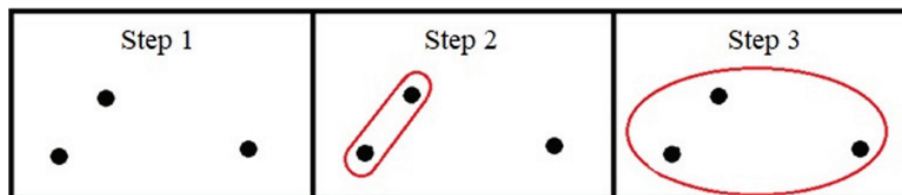


Figure 5. Example of agglomerative procedure 2.

It must be clear that the objects to be grouped are not necessarily the subjects of the research. To evaluate the satisfaction of 461 students in the Statistics disciplines, Zanella, Lopes, and Seidel (2009) used Ward's hierarchical method (which will be discussed later) to create groupings, considering that the grouping objects were the 18 courses in which those students were originally from. Souza and Silva (2009) has another research that illustrates this situation, in which aimed to find a profile of 184 students according to their academic performance in disciplines. Therefore, the researchers used Ward's method to group the ten school disciplines, making them objects of grouping.

There are several algorithms for hierarchical clustering, and because of the high applicability of Cluster Analysis methods, more and more algorithms are in development. The five most commonly used algorithms, according to the literature (Henry et al., 2005; Hair et al., 2009; Lattin et al., 2011; Corrar et al., 2012), will be presented and discussed (for convenience) from the Euclidean distance perspective.

Nearest neighbor: Also known as simple binding, this algorithm finds the two objects separated by the shortest distance and places them in the same group. After this, it groups the next two closest objects, and if one of them is part of a grouping, this grouping incorporates both – this can also occur if the two objects are part of distinct clusterings – and so on until obtaining a single clustering with all n objects. Despite its versatility, one of its problems is that an object will be added to a grouping since it is close to any object in the grouping, even if it is distant from the others – and because of it, this method “tends to be extremely short-sighted” (Lattin, 2011, p.229, our translation).

Furthest neighbor: Analogous to the previous method, the furthest neighbor (or full binding) method connects nearby objects. However, the measure between an object and a grouping is determined by the higher distance between this object and each of the objects belonging to the grouping, that is, the distance between an object p_1 and a grouping C , formed by elements $c_1, c_2, c_3, \dots, c_n$, is calculated by the formula:

$$d_{p_1 C} = \max \{d_{p_1 c_1}, d_{p_1 c_2}, \dots, d_{p_1 c_n}\}$$

This technique eliminates the problem mentioned in the nearest neighbor method since this is a method that produces groupings with comparable diameters. This algorithm is sensitive to data discrepancy since each connection affects the subsequent formation of the solution.

Average linkage: This method can be classified as a balance between the two previous algorithms, as to measure the similarity between two groupings, the algorithm calculates the arithmetic average between the distance of all the objects that constitute the two groupings. Although this technique solves the limitations of the other two methods, its complexity of computational processing makes it less convenient.

Centroid method: In this method, the centroid distance of each cluster calculates the distance between two clusters. To better understand what centroid is, consider a group consisting of three objects (points 1, 2, 3) grouped into two variables, as shown in Figure 6. In this case, the centroid would be the barycenter of the triangle, that is, the point of intersection of the three medians of this triangle (the red straight segments).

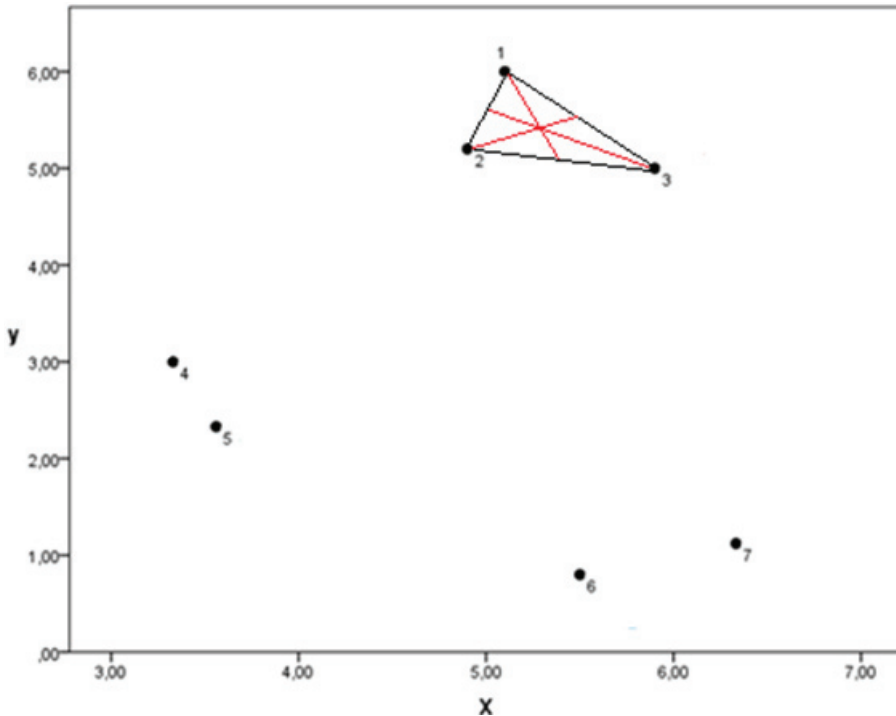


Figure 6. Example of the centroid for a grouping of three objects into two variables. Each vertex of the triangle is an object belonging to the cluster, and the centroid is the point of intersection of the three medians of the triangle that are represented by the red segments.

A characteristic of this method is that the centroid migrates according to the addition of objects in the cluster, which gives robustness to the method since similarity is stipulated taking into account all the cluster points. In addition, this method is less affected by outliers than other methods.

The centroid method is particularly widespread in biology, in taxonomic research. However, this method may lead to confusing results, since, as Hair (2009) points out, there are cases where “the distance between the centroids of one pair may be smaller than the distance between the centroids of another fused pair in an earlier combination” (p.452, our translation).

Ward’s Method: The methods described so far use one same principle, to cluster nearby objects. Ward’s method uses a different strategy, rather than joining close observations, or groupings, it seeks to form clusters with minimal internal variance. Therefore, this algorithm tends to generate clusterings of equal sizes, convex and compact (Lattin, 2011).

Practically, Ward's method tends to group objects quickly, leading to creating clusters of the same size. However, atypical observations can easily distort this method because of the sum of squares dependence.

Ward's method is quite helpful in choosing the number of clusters needed for a sample, especially if one wants to combine a hierarchical and a non-hierarchical technique. The association of this technique with a non-hierarchical method will be better understood below, with the description of non-hierarchical methods.

NON-HIERARCHICAL METHODS

Non-hierarchical methods, often called k-means, have been growing in acceptability, especially in large-scale researches, since hierarchical techniques are inappropriate for analyzing quite large samples. In addition to the ability to handle large samples, k-means methods have resiliency for samples with atypical observations, which is common when dealing with massive data.

For this reason, non-hierarchical methods can potentially analyze researches aimed to create agglomerations according to student performance in large-scale assessments. Leoni & Sampaio (2017) developed research with this approach, using the k-means method to group the 103 schools in the Vale do Paraiba (Brazil) according to their performance in the National High School Examination (in Portuguese, ENEM).

Unlike the hierarchical methods, the k-means techniques do not involve the construction of a process close to the image of a "tree". The procedure is to determine in advance the quantity k of clusters to be obtained and, from n interactions, to distribute the objects of the sample in these clusters.

The literature points to two essential steps for the k-means clustering process: I – The choice of Seeds (pre-defined starting points); II – The designation of each observation to a Seed, this group of observations around each Seed will constitute the grouping (Hair et al., 2009; Lattin et al., 2011; Corrar et al., 2012).

The choice of Seeds can be random – this is how some software determines the Seeds, as is the case with Microsoft SPSS. This way usually generates a large number of interactions, which can hinder computational processing. Another way of choosing Seeds is by the researcher, either by previous research or some other multivariate analysis applied to the sample, as a hierarchical algorithm.

The interactive method for organizing the Clusters starts after choosing the Seeds. To each interaction are elaborated groupings that contemplate the set of objects next to each Seed. The Seeds reallocation occurs after determining the member objects of each cluster, and the centroid of the objects belonging to the group associated with that Seed (as exemplified in Figure 7) determined the new position. Then, new groupings organize the objects, according to the proximity of the objects to the new position of the

Seeds. Each interaction will reallocate the Seeds and reformulate the clusters until there is convergence.

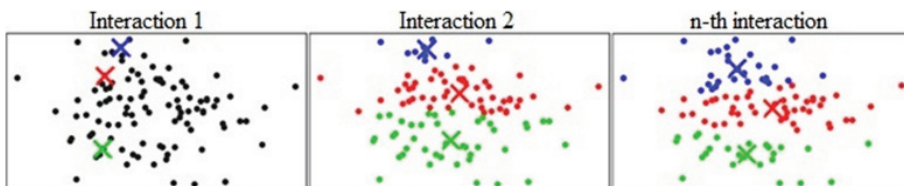


Figure 7. Example of the k-means method for 3 clusters: Objects that are intended to be grouped into 3 clusters in relation to two variables. The first interaction arranged the Seeds, the second altered the Seeds from the centroid, and the nth interaction obtained the final clusters.

The choice of the number of clusters to be reached is fundamental in this method since the algorithm will force the gathering of this number of groups, even if it is not suitable. For this reason, methods that help to determine the number of clusters (such as Ward's method or the elbow technique), repeated simulations, and the researcher in-depth study of the sample become fundamental so that there is no error in the sample by an improper choice of clusters. In the study by Camara et al. (2012), the researchers prepared three groups of gaucho adolescents in relation to their lifestyles. To determine how many clusters would be created, the researchers applied Ward's method, which "revealed the best configuration of the mode" (Camara et al., 2012, p.136, our translation).

As illustrated throughout this section, there are many possibilities for using Cluster Analysis for studies in the educational and teaching areas. Large-scale researches are the main options to use Cluster Analysis, but they are not the only ones. The methods, techniques, and algorithms presented in this section can be efficient as part of the solution process in the research problem of these studies' area, mainly for the data simplification and for the generation of hypotheses, as mentioned above.

FINAL CONSIDERATIONS

The complexity of teaching requires that the research on the subject draw on different approaches. Although currently most of its studies use qualitative approaches, quantitative methods such as Cluster Analysis can contribute to the advancement of knowledge in the area.

It is possible to find research on several topics in the area that have made use of Cluster Analysis techniques, which demonstrates its versatility and potential. The variety of algorithms for Cluster Analysis allows the researcher to be able to find the right method to solve the research problem.

In addition, Cluster Analysis can serve as an organizing technique, which structures and simplifies the massive samples, facilitating for subsequent research to explore further the problems studied. Thus, with respect to the educational area, it is possible to

observe that Cluster Analysis gains robustness when succeeded by qualitative research, which reinforces the importance of allying results from qualitative and quantitative approaches.

Over the years, research in the educational field has increased and been valued. However, the impact of the discoveries made has been mild in practices and, especially, on public education policies. In this sense, it is necessary that large-scale researches, with representative samples, be performed, so that research in the educational field receives its due recognition.

AUTHORS' CONTRIBUTION STATEMENTS

L.C.B.T.N. acted as a supervisor and, in this way, supervised and revised the methodological outline. In addition, L.C.B.T.N. contributed to the possibilities of applying cluster analysis in science teaching, considering his experience with the area in question. D.M.O was responsible for the review of the literature that originated the methods chosen to be part of this study. In addition, D.M.O was responsible for writing the text, which had its final version prepared and revised in conjunction with L.C.B.T.N.

DATA AVAILABILITY STATEMENT

The present research is a literature revision about cluster analyses and applications to the education research area. Data sharing is not applicable to this work, as no new data was created or analyzed in this study.

REFERENCES

- Camara, S. C., et al. (2012). Estilos de Vida de Adolescentes Escolares no Sul do Brasil. *Revista Aletheia*, 37, 133-148.
- Corrar, L. J., et al. (2012). *Análise Multivariada*. 1 ed. São Paulo: Atlas.
- Freire, C. C. & Motokane, M. T. (2016). Análise Fatorial e Análise de Agrupamento no Mapeamento de Concepções Epistemológicas de Professores Sobre a Ciência e a Ecologia. *Investigações em ensino de ciências*, 21(3), 152-175.
- Dal-Farra & R. A., Fetters, M. D. (2017). Recentes avanços nas pesquisas com métodos mistos: aplicações nas áreas de Educação e Ensino. *Acta Scientiae*, 19(3), 466-492.
- Hair, J. F., et al. (2009). *Análise Multivariada de Dados*. 6. ed. Porto Alegre: Bookman.
- Henry, D. B., Tolan, P. H., & Gorman-Smith, D. (2005). Cluster Análisis in Family Psychology Research. *Journal of Family Psychology*. 19(1), 121-132.

- Kodinarya, T. M. & Makwana, P. R. (2013). Review on determinig number of *cluster* in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*. 1(6), 90-95.
- Lattin, J., et al. (2011). *Análise de Dados Multivariados*. 1 ed. São Paulo: Cengage Learning.
- Leoni, R., C. & Sampaio, N. A., S. (2017). Desempenho das escolas públicas e privadas da região do Vale do Paraíba: uma aplicação da técnica de agrupamentos k-means com base nas variáveis do ENEM 2015. *Cadernos do IME*, 42, 31-43.
- Minayo, M. C. (2000). *Pesquisa Social: Teoria, Método e Criatividade*. 17 ed. Petrópolis: Editora Vozes.
- Ocampo, D. M., et al. (2018). Diferentes Perfis de Estudantes Brasileiros Frente aos Desafios Ambientais: Resultados de uma Pesquisa de Larga Escala. *Educação Ambiental em Ação*, 65(7), 1-19.
- Sant'Anna, A. P., et al. (2013). Aplicação da Composição Probabilística e do Método das K-Médias à Classificação de Municípios Quanto à Oferta de Creches. *Caderno do IME*, 33(1), 17-31.
- Sauvé, L. (2005) *Uma cartografia das correntes em educação ambiental*. In: Sato, M. & Carvalho, C. M. Educação ambiental – pesquisa e desafios. Porto Alegre: Artmed.
- Schreiner, C. (2006). *Exploring a ROSE-Garden: Norwegian youth's orientations towards science – see an signs of late modern identities*. (315 pg.). PHD. Thesis – Faculty of Education, University of Oslo.
- Souza, A., M., Silva, F., M. (2009). Perfil dos alunos dos colégios militares: um enfoque multivariado. *Ciência e Natura*, 31(2), 7-24.
- Zanella, A., Lopes, L., F., D. Seidel, E., J. (2009). Diagnóstico do Ensino-Aprendizagem e satisfação dos alunos nas disciplinas de estatística da UFSM. *GEPROS*, 4(3), 123-140.