Employee's clustering based on the Big Five Model in a fitness franchising

Anderson Bertarello Fernandes and **Michel José Anzanello**
Department of Production Engineering and Transportation
Federal University of Rio Grande do Sul
Av. Osvaldo Aranha 99, Porto Alegre, RS 90035-190 Brazil
anderson2069@hotmail.com, anzanello@producao.ufrgs.br

Guilherme Luz Tortorella and **Carlos Ernani Fries,**
Department of Production and Systems Engineering
Federal University of Santa Catarina
Campus Trindade, C.P. 476, Florianópolis, SC 88040-900, Brazil
gtortorella@bol.com.br, carlos.fries@ufsc.br

Abstract

Increasingly competition in the service sector makes companies look for best management practices. In this context, clustering techniques are gaining space through the identification of actions that fit better to the employees' profiles. In addition, Personnel Psychology techniques, with special emphasis on the personality analysis area, present a wide application in studies and companies. This article approaches the clustering by k-means and Fuzzy C-Means, assessed by Silhouette Index (SI), and the principal component analysis of the samples in a gym franchise unit. The samples were characterized by both socio-demographic characteristics and their relationship with the company, as well as their scores on personality tests based on Five Factors Model (FFM). The results indicate the formation of two groups that differ especially in age, time in the unit, salary and job role, also tracking the highest scores on FFM. The formed clusters were analyzed managerially, making it possible to propose specific actions for each group.

Keywords
Five Factors Model; Clustering; Principal Component Analysis.

1. Introduction

The dynamism of the Brazilian economy in recent years resulted in increased competition, especially in the fitness industry, which entails in the emergence of new market entrants (Ueda, 2005). For instance, economic indicators from Porto Alegre, one of the major cities in Brazil, in October of 2013 compared to April of the same year, demonstrate the opening of 28 new establishments over the previous period in this sector, reaching a total of 518 (excluding informal ones) (Prefeitura Municipal de Porto Alegre, 2014a and 2014b). These new establishments are composed by both independent entrepreneurs and expansion or creation of new franchises (Pinheiro and Pinheiro, 2006; Carvalho et al., 2014).
In particular, the franchising sector is featured by highly standardized processes. In this context, the human resources management (HRM) presents a key role in the implementation of practices that lead to satisfactory results, creating a competitive advantage (Silva, 2001). Liu et al. (2007), based on an analysis with more than 19,000 organizations, state that HRM is especially critical for the services sector, emphasizing its importance in the planning of these resources. HRM is understood as a set of business practices involving a company's employees aimed at achieving individual and organizational goals. This area is usually subdivided into three main blocks: personnel, organizational and human factors psychology (Landy and Conte, 2013).

Personnel psychology (PP) is comprehended as the field of psychology responsible for the recruitment, selection, training, performance evaluation, promotion, transfer and dismissals (Landy and Conte, 2013). It assumes that people are different in terms of their working attributes and behaviors, and the information about these differences can be used to predict, maintain and increase employee satisfaction and performance (Ahmada and Schroeder, 2003). Although the PP uses various techniques within the different areas, it is clear that employee grouping techniques (clusters) present few literature evidences concerning its application (Azevedo, 2013); whenever applied, they are together with learning curves, as the study of Uzumeri and Nembhard (1998). Another technique usually combined with PP is the Five-Factor Model (FFM), which consists in a personality model based on five factors: (i) neuroticism, (ii) extroversion, (iii) kindness, (iv) scrupulosity, (v) openness to new experiences (Vecchione et al., 2012). This technique has been used since the 1960s in the USA and Europe in order to predict the performance of employees (Barrick and Mount, 1991).

This paper proposes an approach for grouping employees of a gym franchise using multivariate analysis techniques. The descriptive variables of the employees' characteristics include the five personality factors from the FFM, and the evaluations of performance, economic, social and demographic. The observations (employees) are grouped through techniques that include k-means and fuzzy c-means (FCM). Further, the quality of the formed groups is assessed using the Silhouette Index (SI). To enhance the formed groups, variables are by applying PCA (Principal Component Analysis). This article presents contributions to both practice and theory. In practical terms, the achievement of consistent groups of employees enables a better understanding of their characteristics in order to establish proper management techniques with these groups. Moreover, its application can be expanded to other similar business units, providing support for decision-making processes. In terms of theoretical perspective, this article breaks new ground by integrating results from the FFM technique as input variables for clustering structures.

This rest of this paper is divided into four sections, besides this introduction. Section two brings a literature review with regards to cluster formation techniques such as *k*-means, FCM, PCA, SI and FFM. Section 3 presents the proposed method for generation of groups of employees, while the fourth section presents the results from the application of the proposed method. Section 5 presents the conclusions of the study and recommendations for future studies about grouping employees and personality models.

2. Theoretical Framework

2.1 Multivariate analysis tools

The general aim of multivariate tools is the simplification of the data used in analysis. It aims to reduce the variety and complexity of the original data, making the analysis simpler (Rencher, 2002). In particular, this section discusses two multivariate techniques applied in the article: PCA and clustering tools.

The purpose of the Principal Component Analysis (PCA) is to find an orthogonal basis vectors that maximize the variance of a set of data (Hansen et al., 2014). Therefore, the number of set dimensions is reduced, preserving the maximum of its variability, searching for the dimension in which the observations are further apart (Rencher, 2002). The operation of the PCA is to calculate the eigenvector (vi) of the covariance matrix, in which the magnitude of an eigenvector λi is equal to the sum of the variance in the correspondent eigenvector vi direction, called the principal component (PC). The next major component is the linear combination in which the sum of the variances in the orthogonal direction to vi be the maximum as possible. The principal components (PCs) resulting , $Z_1, Z_2,..., Z_n$ are linear combinations of the original variables $X_1, X_2,..., X_p$, through the weights $w_1,w_2,...,w_p$. They can be written as $Z_n = w_1X_1 + w_2X_2 + ... + X_pw_p$ (Jolliffe, 2002). It is worthy noticing that PCA should be used in a data set in which the values of each variable of the samples are approximately in the same range (Härdle and Simar, 2003). Further, the vectors may be normalized or not, depending on the specific use (Hair et al., 2005). To define the number of PCs to be retained, some techniques are suggested by Rencher (2002): (i) the sum of the variance of PC approaches 80% of the original variance; (ii) select the PC whose eigenvalues λ are higher than the average of the eigenvalues; (iii) use the technique of scree graph, and realize the natural distinction between large and small eigenvalues, and (iv) test the significance of Z corresponding to the largest eigenvalues. The PCA has been widely applied in different scenarios. Xu et al. (2014) present a monitoring of multimodal processes through this technique, while Zhou et al. (2014) applied PCA techniques aimed at face recognition. In other fields of science, Põldaru et al. (2014) used PCA to assess the quality of life in cities of Estonia, while Tan et al. (2015) used the technique to study the relationship between economic, social and environmental indicators in China.

Regarding the multivariate technique known as Cluster Analysis, it aims to find patterns within a multivariate sample and group them (Härdle and Simar, 2003). The clusters should be formed in a manner that its intrinsic homogeneity is the maximum possible while the heterogeneity in relation to others is the highest possible (Hair et al., 2010). It differs from the classification technique, since in clustering the groups and the number in each group are unknown. Clustering techniques have been increasingly used in various fields of study, among them for decision-making and pattern recognition (Taboada and Coit, 2007) and business intelligence (Tseng et al., 2012). The application of systematic groupings in Human Resource Management (HRM) as Landy and Conte (2013) point, has also gained importance, although little utilization (Azevedo, 2013). Some authors have used techniques of clustering associated with learning curves in assembly lines of the footwear industry, through PCA techniques, FCM, k-means and Kernel methods (Anzanello and Fogliatto, 2007; Azevedo, 2013). In medicine fields, Dreachslin et al. (2000) used techniques for groups of nurses treating patients using hierarchical techniques. Yang et al. (2006) use mixed-variable fuzzy c-means (MVFCM) techniques for the formation of cells in an industry.

Groups analysis algorithms can be classified into two groups: hierarchical and non-hierarchical. According to Rencher (2002), hierarchical techniques are those that involve a process of successive steps, in which it is highlighted the agglomerative and divisive algorithms. In the first one, according to Hair et al. (2010), the samples are grouped according to their distances (usually Euclidean); the formed clusters continue to be grouped according to their proximity until all clusters rise into a single group. However, divisive algorithms make the opposite way, from a single cluster that should be successively divided according to their internal distance. Once assigned to a particular group, the object cannot be reallocated, which ensures process hierarchy. They are generally represented by the dendograms (Rencher, 2002).

The non-hierarchical methods are those in which objects or samples are assigned to a cluster, from the definition of a number of groups to be formed (Corrar et al., 2009). Two non-hierarchical methods will be addressed in this study. The first is the k-means method based on Euclidean distance. It allows an object, once allocated to a particular group, to be reallocated to another if their similarity is greater with objects of that other group (Rencher, 2002). The method starts with the division of the samples into k clusters defined by researcher (Hair et al., 2010). Then it is calculated the centroids for each cluster formed and the distances (Euclidean) of the centroids for each observation. Following, a reallocation of the samples to the nearest centroids of clusters is made. This method is iterative, repeating until no significant variation in distances from each sample into each of the centroids (Rencher, 2002). The other non-hierarchical method to be addressed is the fuzzy c-means (FCM). Velmurugan (2014) state that the FCM is similar to k-means both in structure and in behavior, but it does not consider that the barrier among clusters is so rigid. In this way each observation belongs to different clusters in certain degree. This degree can be calculated as a function of the distance of this vector to the centroid or other representative vector of the cluster, which aims to minimize the objective function (1):

$$J_m = \sum_{j=1}^{k}\sum_{i=1}^{N} u_{ij}^{m} d_{ij} \tag{1}$$

where:

c_j - dimension center d of the cluster C_j
u_{ij} - degree of adhesion of x_{ij} in C_j cluster
N - number of data points
m - parameter, a real number greater than 1
k - number of groups
d_{ij} - Euclidean distance between the point of data x_{ij} and the center of the cluster C_j. The following restriction is valid for u_{ij}:

$$\sum_{j=1}^{k} u_{ij} = 1, \quad for\ i = 1,...,N \tag{2}$$

The iterative process of optimizing the objective function (1) is given by upgrading u_{ij} and the centers of the clusters c_j given by (3) and (4).

$$u_{ij} = \frac{1}{\sum_{j=k-1}^{c}\left(\frac{x_i - c_j}{x_i - c_k}\right)^{\frac{2}{m-1}}} \tag{3}$$

$$C_j = \frac{\sum_{i=1}^{N} x_i u_{ij}^m}{\sum_{i=1}^{N} u_{ij}^m} \tag{4}$$

The iterative procedure above is finished when $max_{ij}\{|u_{ij}(k+1) - u_{ij}(k)|\} < \zeta$, where ζ is a stopping criterion between 0 and 1 and k is the iteration steps. This procedure converges to a local minimum. The samples are allocated to each group by a member function, which explains the fuzzy behavior of the algorithm, since it belongs to more than one group at a time. Then, it is built one matrix $\boldsymbol{U}$ (whose factors are numbers between 0 and 1), representing the degree of association between the samples and cluster centers (Bezdek et al., 1984). Thus, the algorithm is completed with the next iteration (*i*) initialize the matrix $\boldsymbol{U} = [u_{ij}]$ with $U_{(0)}$; (*ii*) in every k step calculate the center of the vector $C_{(k)} = [c_j]$ with $U_{(k)}$; (*iii*) update $U_{(k)}$ with $U_{(k+1)}$; and (*iv*) if $|| U_{(k+1)} - U_{(k)} || < \zeta$ for. Otherwise, return to (*ii*). The FCM generates clusters in which the weight of the member function has a natural interpretation, not probabilistic. Thereby, it is heavily used in the natural sciences, as that can capture this behavior.
For Rousseeuw (1987), the Silhouette Index (*SI*) shown in equation (5), is a way of quantitatively evaluating the quality of a clustering. This index evaluates how a sample is similar to the others in its group compared to the closest (Stroieke et al., 2013):

$$SI(i) = \frac{b(i) - a(i)}{max[b(i), a(i)]} \tag{5}$$

where:

$a(i)$ - average distance of the sample i in relation to the other group that has been allocated
$b(i)$ - average distance of i sample in relation to the nearest neighbor group.

Therefore, the value of $SI(i)$ ranging from -1 to $+1$, with values near $+1$ denoting observations properly inserted in the destination cluster, while the opposite is denoted by -1.

2.2 Five Factors Model

The personality can be defined as the set of emotional and behavioral characteristics of an individual that is generally maintained constant both over time and in a variety of situations (Landy and Conte, 2013). When considering the usual way of an individual responding to incentives, personality is related to happiness and behavior, not only in everyday life but also at work. These behaviors include performance, absenteeism, effective teams, turnover, unproductive behaviors and job satisfaction, among others (Steele et al., 2008). For Barrick et al. (2005), personality, viewed in isolation, has more predictive significance on performance at work than the intelligence or experience alones. In addition, there are fewer differences in measurements of age, gender, ethnic subgroups than intelligence.
The interest in personality measurement has increased with the emergence of a taxonomy for the dimensions of personality, called Big Five or Five-Factors Model (FFM) (Digman, 1990). Early researches on the subject refer to McDougall (1932) and Cattell (1947) who pointed out sixteen factors for personality as first order and eight as second order. Further studies reduced number of factors, which were consistently validated by several researchers (Barrick and Mount, 1991; Hogan, 1992; Stewart, 1999; Moon, 2001; Clarke and Robertson, 2005; Vecchione et al., 2012; Landy and Conte, 2013): (i) extroversion, (ii) neuroticism, (iii) kindness, (iv) scrupulosity, (v) openness to new experiences. The first one refers to the characteristics of being sociable, talkative, ambitious, energetic, and assertive, and may be broken into two: ambition and sociability. The second factor, also called as emotional stability, determines the tendency that an individual presents to be anxious, depressed, angry, embarrassed, emotional, worried and unsure. The third factor refers to individuals who tend to be courteous, flexible, reliable, cooperative, friendly, pleasant, humorous, tolerant and generous. Individuals with lower levels of kindness are less likely to easily manage relationships with other individuals, including following safety standards. The factor scrupulosity, which includes characteristics such as responsibility, prudence, persistence, preparation of plans and objectives driven, was the first factor to attract the attention of researchers. For some scholars, scrupulosity would be better divided in realization (related to hard work, persistence and desire to perform well) and reliability (related to being

disciplined, organized, respectful of norms, honest and trustworthy). It is possibly the most important personality variable on the desktop and can be considered equivalent to intelligence "g" in the field of non-cognitive intelligence (Schmidt and Hunter, 1992). Finally, there is openness to new experiences, which reflects individuals who are imaginative, curious, original, independent, creative, educated and artistically sensitive.
For Campbell (1990), personality is more related to motivational aspects of work than its technical facet, presenting a better predictive capacity of what the employee probably will do (linked to motivation) rather than what he can do (linked to their knowledge and skills). Hence, Barrick et al. (1993) state that the greater the degree of autonomy of a task the more the personality factors influence in a good performance. As pointed out by Cheung (2004), although the five factors can vary across countries, it is applicable and has substantial results in different cultures. Several systematic measurements have been suggested with regards to FFM applicability. Landy and Conte (2013) highlight the fact that the tests may be divided into two types, screen-out and screen-in. The first one was initially developed and aimed to eliminate candidates in job selections, who might represent potential problems. The second type, developed later, includes a systematic aimed to add information about the positive attributes of respondents.

3. Methodological procedures

The proposed method relies on four operational steps: (i) data collection, (ii) data processing, (iii) generation and evaluation of quality of groups, and (iv) elaboration of management strategies based on clusters.
Initially, employees are characterized by two blocks of information. The first relies on features of socio-demographic and economic backgrounds. The qualitative profile data have been processed in order to generate quantitative data suitable for subsequent clustering. These data were collected by consulting the records of the restaurant itself and the questionnaires. The second part of the data aims to characterize the personality traits of employees through the FFM methodology. These data were collected by means of printed questionnaires applied to a group of employees and coaches. The questions were answered through a Likert scale, ranging from 1 to 5. Among the interviewed respondents there were teachers, interns, receptionists, commercial area and managers. The answers were arranged in matrices, with each column representing a variable of interest and each line the data of an employee.
The second step of the method is analyzing the collected data in order to assess potential inconsistencies and ensure a consistent profile of groups of employees. First, samples with missing information were excluded. Then, standardization techniques were applied to transform each variable in a range between *0* and *1*. This procedure was used for both data blocks and intended to prevent calculation distortion originated by large scale variables, when forming the clusters.
Based on the standardized data, clusters of employees are formed. First, it is performed the clustering of standardized variables (CSV) through the agglomerative hierarchical techniques in order to obtain a rough estimative of the number of clusters to be generated. Then, we proceed to the k-means clustering and FCM, changing the number of clusters, k. The quality of each group was evaluated by an average Silhouette Index (*SI*), obtained from each clustered observation. The value of k responsible for higher average *SI* is selected for interpretation. Alternatively, PCA is applied on the standardized variables in order to obtain the Clustering of Latent Variables (CLV). The idea is to assess the formation of clusters based on a reduced number of uncorrelated variables. The latent variables are clustered using k-means and FCM techniques. Next, it is evaluated the quality of the obtained clusters obtained through the average *SI*, as done previously. Finally, it is generated a comparative table between clusters using original and latent variables.
The last step of the method consists of two stages. First, it is verified if the clusters number, obtained using the previous steps, is appropriate with the current operation. The idea is to obtain a number of clusters that satisfies both the mathematical and managerial aspects. Second, the human resources strategy is adapted in such a way that takes into consideration the different profiles of employees and their different characteristics and demands.

4. Results

The study was conducted in a unit of a fitness franchising in the city of Porto Alegre. It is part of a holding that possess other several franchising units. The franchising is highly recognized in Brazil due to its quality level, and has been inspired on international franchising with full service style, in which all classes and methods are included in the monthly fee. The facility operates 17 hours a day, 363 days a year, offering 32 different kinds of classes, as well as providing monitoring and physical evaluation. Inaugurated in 2010, the unit has 58 employees, including 6 trainees and 41 employees (including coaches, maintenance assistants, reception and commercial area). Moreover, it has eight managers: marketing, sales, maintenance, evaluation, gym, cardio respiratory, fitness, IT and Financial, and general manager.

The variables used in the study had two distinct sources. The first was a printed questionnaire for the unit's employees. The questionnaire was divided into two large blocks. The first consisted of general questions about the employee that included age, gender, time in the unit, job satisfaction, role, level of education, and how the transport to work was performed. The second consisted of a Screen-in type questionnaire called IPIP Big-Five Factor Markers translated to Portuguese, which comprises 50 questions that should be answered on a Likert scale (ranging from strongly agree to strongly disagree). This questionnaire BFF resulted in a score of 1 to 5 for each of the areas. The company's database was also used to search for information such as employees' salaries, and to dispel doubts of the answers of the questionnaire. Forty-five out of fifty-six total employees responded to the questionnaire. Two were on vacation, seven declined and two wanted to deliver only after the collection period.

Processing the data was in three distinct stages. The first was to determine whether any of the employees who answered the questionnaire had overlooked some crucial information that could not be obtained through the company's records, or answered wrongly the IPIP. From the total, only one invalidated the analysis through the Big Five Model, since he pointed out more than one option in several statements. The second stage consisted of the conversion of qualitative data into quantitative, as well as deciding which variables would be used. Variables whose physical quantity made real sense were selected: age (years), time in the unit (years), job satisfaction (score 0-10), salary and job function (many managers and employees work in two or more sectors in the gym, and assigned values to their likely position in the unit - 1 for trainees, 2 to base positions in sectors, 3 to leaders and 4 for managers). As for the sectors, each employee had a weight of 1 (i.e., if acted in more than one sector, weights were summed to the respective employee). The final score was obtained by multiplying the position by the sector. The last stage consisted in the standardization of observations, obtaining values ranging from *0* to *1*. In particular, neuroticism had its score (and percentile) standardized on a scale in which the lowest value was *1*.

After data standardization, we proceeded to the clustering analysis, which resulted in the dendogram of Figure 1; on the x-axis there are the samples, and on the y-axis there is the distance between related samples. Figure 1 suggests that the cluster number for the sample oscillates between 2 and 6.

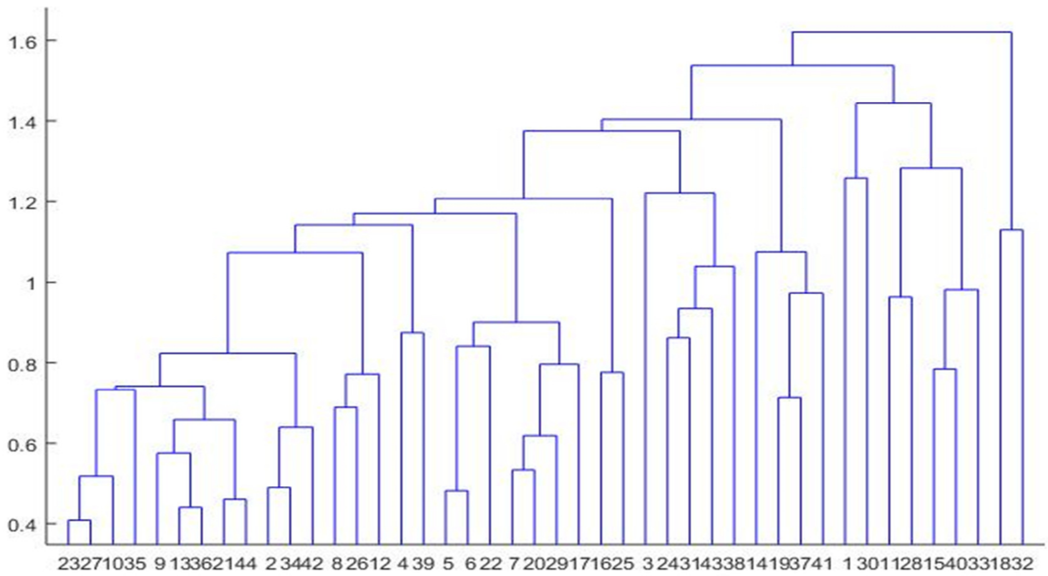

Figure 1. Dendogram representing the hierarchical clustering

Next, the clustering of standardized variables was proceeded, applying k-means and FCM techniques. For clustering with FCM technique, the difference of the objective function in the iteration was used as a stopping criterion; whenever less than 1.10-10. Then, *SI* was used to evaluate the formation of clusters, as displayed in Table 1. It is worth to notice that similar values were obtained to the function of probability of belonging to each group.

Table 1. *SI* values for the variation of cluster numbers and clustering techniques of standardized variables

	k-means	Fuzzy *c*-means
k = 2	0.329789	0.310947
k = 3	0.197415	-0.069335
k = 4	0.212450	0.006923
k = 5	0.142659	0.006923
k = 6	0.178136	0.006923

Table 2 shows the results of PCA. It is noticed that the first five eigenvectors account for 80.5% of the observed variance, and represent the decision criteria regarding the number of PCs to be retained. In addition, the average of the eigenvectors is equal to 1, showing that the sixth eigenvalue (1.062806), besides representing 7.0% of explained variance, is above average, which suggests another cut-off point to determine the number of PCs to be retained. Due to these reasons, the number of PCs retained for analysis was six.

Table 2. Eigenvalues and explained variance

	Eigenvalues	Explained variance	Cumulative explained variance
1	3.877373	25.8%	25.8%
2	2.898883	19.3%	45.1%
3	2.027182	13.5%	58.6%
4	1.805323	12.0%	70.7%
5	1.467383	9.7%	80.5%
6	1.062806	7.0%	87.5%
7	0.730504	4.8%	92.4%
8	0.556430	3.7%	96.1%
9	0.290844	1.9%	98.1%
10	0.200138	1.3%	99.4%
11	0.036887	0.2%	99.6%
12	0.022764	0.1%	99.8%
13	0.013220	0.08%	99.9%
14	0.007541	0.05%	99.9%
15	0.002722	0.018%	100.0%

In the sequence, a new clustering analysis was made by k-means and FCM on the six latent variables retained; again, the number of clusters ranged from 2 to 6. Table 3 compiled the SI values obtained for each type of clustering according to the variation of *k*.

Table 3. *SI* values for the variation of cluster numbers and clustering techniques of standardized variables

	k-means	Fuzzy *c*-means
k = 2	0.32804	0.348015
k = 3	0.25657	0.193935
k = 4	0.13597	0.105294
k = 5	0.21615	0.150594
k = 6	0.19960	0.153856

Through the analysis of the average *SI*, it is clear that the largest, 0.348015, was generated by the FCM with 2 clusters (the largest value generated by *k*-means is also coming from *k*=2). It is noteworthy, however, that the values of *SI* denote an average quality clustering, justified by the structure of the data. Figure 2 shows *SI* values for each observation grouped for 2 clusters. In management terms, the consideration of two clusters makes sense within the current gym context, without any management or operational impediment to this classification. To examine how the clusters were distributed along the characteristics of employees, Table 4 consolidated the average characteristics of each group.

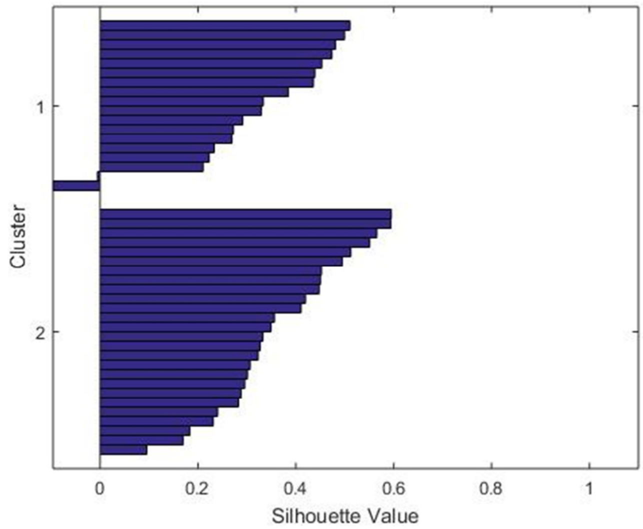

Figure 2. SI distribution by clusters

Table 4. Cluster's characteristics average

Characteristics	Group 1	Group 2
Age	31.50	26.29
Time in unit	3.63	1,13
Job role	4.39	1.83
Grade	8.42	9.16
Salary	US$ 560.85	US$ 279.15
Extroversion	3.42	3.20
Scrupulosity	3.77	3.60
Neuroticism	2,38	2.82
Kindness	4,04	3.76
Openness to new experiences	3,67	3.39
Percentile extroversion	60.94%	53.79%
Percentile scrupulosity	62.83%	57.08%
Percentile neuroticism	25.00%	41.58%
Percentile kindness	57.17%	40.92%
Percentile openness to new experiences	28.06%	17.04%

Table 4 shows the distinction between the two groups especially in the variables job role (encompassing all managers in Cluster 1), time in unit and salary. These three variables tend to have a higher correlation, since the company's policy stimulates developing managers within the unit, before hiring a third party. However, the variable salary also had significant influence, since all commercial area employees (employees with the highest salary for the position range) are in cluster 1, except the first employee who had entered for three months in the unit. The factors from BFF also showed alignment in the formation of groups, because all the factors on which a higher score represents a positive characteristic were higher in groups (and smaller neuroticism values). Also there is a difference in the average of the factor related to the good performance of the work (scrupulosity) in Group 1. In addition, the neuroticism factor had an average of almost 0.5, which indicates people with higher self-control trends, especially in relation to negative feelings. It should be noted that the degree of job satisfaction within the unit was higher in cluster 2, showing that employees at lower wages and lower positions are more satisfied with the unit. However, individuals with more experience in the unit and more professional ambitions, tend to have more requirements (justifying the reduction of the score).

In consequence of the formed groups, different management actions could be addressed. For Group 1, consisting of more experienced individuals, it is recommended to take actions to ensure greater financial stability, especially related to profit sharing. Furthermore, it is important to create the introduction of differentiated training routines, in particular those related to staff management and information management, seen as flawed even in the current structure. Moreover, it would be suitable the adoption of practices that reinforce the good atmosphere that exists in the workplace and to motivate the search for new practices, especially based on the fact that the average openness to new experiences is higher in this group. It is suggested, for example, the adoption of training in other Brazilian units, incentives for conferences and courses participation, among others.

In group 2, two distinct points need to be improved. The first one refers to training (both technical and procedures) necessary for beginners or for those who do not have the closest bond with the gym (because many coaches teach in other academies). In addition, training to develop personal skills is recommended to better train them and future leaders. On the other hand, it is important to ensure that employees of this group visualize the perspective of growth, development and better payment within the unit. If this group does not envision future opportunities, there will be inevitable shutdowns, turnovers and the need for basic training without greater chances of deepening. Finally, it is suggested the application of constant tests to more accurately assess employees identifying gaps and, hence, train them properly.

5. Conclusions

This study aimed to find the proper HRM practices for a gym franchise unit through the creation of groups of employees using multivariate analysis of clustering. Two main motivations guided the research. The first one was the insertion of personality techniques based on the Big Five Model in groups of studies, while the second was related to the formation of consistent groups to address HRM practices, which can be A competitive edge in the gym franchising.

The method was based on non-hierarchical clustering techniques, *k*-means and fuzzy c-means, and their quality assessed by Silhouette Index (*SI*). In addition, variables were initially clustered using agglomerative hierarchical techniques to know the approximated number of clusters (*k*) to be tested. Furthermore, it was also applied the principal component analysis (PCA) in the database and performed the clustering with the two techniques previously mentioned in the standardized variables. Once obtaining these clusters, the interpretation and proposition of appropriate management practices to the profile of each cluster could be settled.

The results were obtained by applying the method in a database of a gym's employees. This database was composed of both variables from the available information in the company's records and a survey with employees. The database consisted of two large blocks: socio-demographic and relationship characteristics derived from its personality traits based on the FFM. After clustering the standardized and latent variables, it was obtained the best SI value of 0.348015 when applied FCM. Employees with longer time in unit, higher salary and greater responsibility, as well as higher FFM scores were part of a single group. Finally, for each group, HRM practices were suggested.

For future studies, it is suggested the addition of other variables to complete the information about employees. Among them, the household income is suggested, which differs substantially because many coaches also provide the personal training service that carries a substantial increase in income. Furthermore, the application of the method in other franchising units is also indicated to verify common aspects and enable standardization of HRM policies.

References

Ahmada, S., and Schroeder, R. G., The impact of human resource management practices on operational performance: recognizing country and industry differences, *Journal of Operations Management*, vol. 50, no.1, pp.19–43, 2003.

Anzanello, M. J., and Fogliatto, F. S., Curvas de aprendizado: estado da arte e perspectivas de pesquisa, *Revista Gestão e Produção*, vol. 14, no. 1, pp. 109-123, 2007.

Azevedo, B. B., Agrupamento de funcionários com perfis semelhantes de aprendizado utilizando técnicas multivariadas. Dissertation (*Master in Production Engineering*, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, 2013.

Barrick, M. R., and Mount, M. K., The Big Five Personality Dimensions and Job Performance: A Meta-Analysis, *Personnel Psychology*, vol. 44, no. 1, pp. 1-26, 1991.

Barrick, M. R., and Mount, M. K., Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance, *Journal of Applied Psychology*, vol. 78, n. 1, pp. 111-118, 1993.

Barrick, M. R., and Mount, M. K., Yes, the personality matters: Moving on to more important matters, *Human Performance,* vol. 18, no. 4, pp. 359-372, 2005.

Bezdek, J. C., Ehrlich, R., and Full, W., FCM: the Fuzzy c-means clustering algorithm, *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984.

Campbell, J. P., *The role of theory in industrial and organizational psychology - Handbook of industrial and organizational psychology*, 2nd Edition, Consulting Psychologists Press, Palo Alto, CA, 1990.

Carvalho, M. J., Buchmann, C., and Gonçalves, C., Qualidade e satisfação no fitness: Percepção segundo o sexo dos sócios, *Revista Portuguesa de Ciências do Desporto*, Issue S1A/S1R, pp. 996-1008, 2014.

Cattell, R. B., Confirmation and clarification of primary personality factors, *Psychometrika*, vol. 12, no. 3, pp. 197-220, 1947.

Cheung, F. M., Use of Western and Indigenously Developed Personality Tests in Asia, *Applied Psychology*, vol. 53, n. 2, pp. 173–191, 2004.

Clarke, S., and Robertson, I. T., A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings, *Journal of Occupational and Organizational Psychology*, vol. 70, n. 3, pp. 355–376, 2005.

Corrar, L. J., Paulo, E., and Dias Filho, J. M., *Análise multivariada: para os cursos de administração, ciências contábeis e economia*. Ed. Atlas, São Paulo, Brazil, 2009.

Digman, J. M., Personality structure: Emergence of the five factor model, *Annual Review of Psychology*, vol. 41, pp. 417-440, 1990.

Dreachslin, J. L., Hunt, P. L., and Sprainer, E., Workforce diversity: implications for the effectiveness of health care delivery teams, *Social Science & Medicine*, vol. 50, n. 10, pp. 1403–1414, 2000.

Härdle, W. C., and Simar, L., *Applied Multivariate Statistical Analysis*, Method & Data Technologies, 2003.

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C., *Análise Multivariada de Dados*, 5th Edition, Bookman, Porto Alegre, Brazil, 2005.

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E., *Multivariate Data Analysis*, 7th Edition, Pearson, 2010.

Hansen, T. J., Abrahamsen, T. J., and Hansen, L. K., Denoising by semi-supervised kernel PCA preimaging, *Pattern Recognition Letters*, vol. 49, pp. 114–120, 2014.

Hogan, R., and Hogan, J., *Hogan Personality Inventory Manual*. Tulsa, OK, 1992.

Prefeitura Municipal de Porto Alegre, *Indicadores Econômicos de Porto Alegre de Abril de 2013*, Available at: http://lproweb.procempa.com.br/pmpa/prefpoa/smic/usu_doc/indicadores_abril__2013.pdf. Accessed in: August 20th 2016, 2014.

Prefeitura Municipal de Porto Alegre, *Indicadores Econômicos de Porto Alegre de Outubro de 2013*, Available at: http://lproweb.procempa.com.br/pmpa/prefpoa/smic/usu_doc/indicadores__outubro__2013.pdf >. Accessed in: August 20th 2016, 2014.

Jolliffe, I. T., *Principal Component Analysis*, 2nd Edition, Springer series in statistics,New York, NJ, 2002.

Landy, F. J., and Conte, J. M., *Work in the 21st Century: An Introduction to Industrial and Organizational Psychology*, 4th Edition, Wiley, 2013.

Liu, Y., Combs, J. G., Ketchen Jr, D. J., and Ireland, R. D., The value of human resource management for organizational performance, *Business Horizons*, vol. 50, pp. 503–511, 2007.

McDougall, W., Of the words character and personality, *Character Personality*, vol. 1, pp. 3-16, 1932.

Moon, H., The two faces of conscientiousness: Duty and achievement striving in escalation of commitment dilemas, *Journal of Applied Psychology*, vol. 86, n. 3, pp. 533-540, 2001.

Pinheiro, I. A., and Pinheiro, R. R., Organização científica do trabalho reinventa um mercado tradicional: o caso do fitness, *RAE-eletrônica*. Fundação Getulio Vargas – Escola de Administração de Empresas de São Paulo, São Paulo, Brazil, vol. 5, n. 2 (jul/dec), pp. 1-26, 2006.

Põldaru, R., and Roots, J. A., PCA-DEA approach to measure the quality of life in Estonian counties, *Socio-Economic Planning Sciences*, vol. 48, pp. 65-73, 2014.

Rencher, A. C., *Methods of Multivariate Analysis*. 2th Edition, John Wiley & Sons, New York, NY, 2002.

Rousseeuw, P. J., Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987.

Schmidt, F. L., and Hunter, J. E., Development of causal models of job performance, *Current Directions in Psychological Science*, vol. 1, pp. 89–92, 1992.

Silva, R. O., *Teorias da Administração*, Pioneira Thomson Learning, São Paulo, Brazil, 2001.

Steele, P., Schmidt, J., and Schultz, J., Refining the relationship between personality and subjective well-being, *Psychological Bulletin*, vol. 134, pp. 138-528, 2008.

Stewart, G., Trait Bandwidth and Stages of Job Performance: Assessing Differential Effects for Conscientiousness and its Subtraits, *Journal of Applied Psychology*, vol. 84, pp. 959-968, 1999.

Stroieke, R. E., Fogliatto, F. S., and Anzanello, M. J., Análise de conglomerados em curvas de aprendizado para formação de agrupamentos homogêneos de trabalhadores, *Produção*, vol. 23, pp. 537-547, 2013.

Taboada, H. A., and Coit, D. W., Data Clustering of Solutions for Multiple Objective System Reliability Optimization Problems, *Quality Technology & Quantitative Management*, vol. 4, n. 2, pp. 191-210, 2007.
Tan, F., and Lu, Z., Study on the interaction and relation of society, economy and environment based on PCA–VAR model: As a case study of the Bohai Rim region, China, *Ecological Indicators*, vol. 48, pp. 31–40, 2015.
Tseng, J., and Li, S., Quantifying volatility clustering in financial time series, *International Review of Financial Analysis*, vol. 23, pp.11–19, 2012.
Ueda, V. Os novos empreendimentos imobiliários e as transformações recentes no espaço urbano de Porto Alegre, *Proceedings of X Encontro de Geógrafos da América Latina*, São Paulo, USP, Brazil, 2005.
Uzumeri, M., and Nembhard, D., A Population of Learners: A New Way to Measure Organizational Learning, *Journal of Operations Management*, vol. 16, n. 5, pp.515-528, 1998.
Vecchione, M., Alessandri, G., and Barbaranelli, C., The Five Factor Model in personnel selection: Measurement equivalence between applicant and non-applicant groups, *Personality and Individual Differences*, vol. 52, pp.503–508, 2012.
Velmurugan, T. Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data, *Applied Soft Computing*, vol. 19, pp. 134–146, 2014.
Xu, X., Xie, L., and Wang, S., Multimode process monitoring with PCA mixture model, *Computers and Electrical Engineering*, published online, 2014.
Yang, M., Hungb, W., and Chenga, F., Mixed-variable fuzzy clustering approach to part family and machine cell formation for GT applications, *International Journal of Production Economics*, vol.. 117, n. 1, pp. 30–4.1, 2006.
Zhou, C., Wang, L., Zhang, Q., and Wei, X., Face recognition based on PCA and logistic regression analysis, *Optik*, vol. 125, no. 20, pp. 5916–5919, 2014.

Biography

Michel José Anzanello is Associate Professor in the Department of Production Engineering and Transportation of the Federal University Rio Grande do Sul (UFRGS), Porto Alegre, Brazil. He holds a Bachelor's degree in Chemical Engineering from UFRGS, and a PhD from Rutgers - The State University of New Jersey. His research include among others the use of learning curve modeling for assignment of products to teams of workers, studies of systems reliability, lean production in winery industry, and modeling of working systems using statistical tools.

Anderson Bertarello Fernandes holds a Bachelor's degree in Production Engineering from UFRGS. He has been active as controller, project manager, and financial analyst in Brazilian industries that are focused primarily in food and beverage.

Guilherme Luz Tortorella is Associate Professor in the Department of Production and Systems Engineering of the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil. His Master in Production Systems and PhD in Production Engineering were earned from UFRGS. He also has twelve years of experience in the automotive industry with international activities in Mexico, England, USA, and Uruguay.

Carlos Ernani Fries is currently Associate Professor in the Department of Production and Systems Engineering of the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil. Mr. Fries holds a Bachelor degree in Civil Engineering as well as a Master and PhD in Production Engineering from UFSC. He has taught courses in Operations Research applied to Manufacturing and Logistics, Decision Theory, Statistics and Forecasting Models among others. His research interests include manufacturing, simulation, optimization, management games, data analysis applied to logistics, and application of big data tools. He is member of IEOM, POMS, and INFORMS.