IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-02 E-ISSN: 2347-2693

Clustering Analysis of the Departments of Medical Faculty Hospitals Based on Some Variables: Adnan Menderes University

Özel Sebetci, Ph.D.[1*] and GökhanAksu, M.S.[2]

[1*] *Özel Sebetci is AssistantProfessor at Department of ComputerTechnology in University of Adnan Menderes.*
[2]*Department of Vocational High School, University of Adnan Menderes, Aydın, Turkey.*

www.ijcseonline.org

Received: 09/Jan/2016 Revised: 14/Jan/2016 Accepted: 07/Feb/2016 Published: 29/Feb/ 2016

Abstract-The objective of this study is to cluster 40 different departments at the Adnan Menderes University Hospital according to some variables. The data used in the study was obtained from the 2014 statistics of the Adnan Menderes University. Among the hierarchical clustering approaches, complete link method was used as the study attempted to determine the way of merging or partitioning of clusters. The study determined that all independent variables had a significant effect in clustering as the result of the ANOVA test which was done for clustering purpose. Chi-square means and discriminant analysis method was used in order to provide evidence on the validity of the cluster results obtained in the analysis. The results of the study were discussed in two stages. The first stage included the analysis of all 40 units while variables of relevant expense, package loss and SGK deduction were evaluated. The second stage included the analysis of 37 units while the variables of expense, lecturer, assistant, nurse, number of personnel, number of polyclinic rooms, polyclinic area, number of service beds and service area were evaluated. Upon the analysis in the first stage, it was determined the units were gathered under 5 clusters. The analysis showed that both the orthopaedics and oncology units were a cluster on their own while the units of hematology and brain surgery were included in the same cluster. The fourth clusters consist of the units of Cardiovascular Surgery, General Surgery, Emergency and Cardiology while the fifth cluster consists of the all other units. As a result of the analysis in the second stage, it was observed that the number of clusters and units within clusters didn't vary. In order to determine the validity of the results of the study, it was determined that the number of clusters obtained by calculating Wilk's lambda coefficient was the same with the number of clusters determined by the complete link method. According to the findings obtained in the study, it was determined that the units with highest expenses made a single unit and it is believed that the expenses of both units were significantly different from other units.

Keywords: Clustering, discriminant analysis, ANOVA, hospital, clinics

1. Introduction

It is becoming more important to obtain significant information among the data in order to support strategic decision making and provide reliable estimates together with the rapid increase in the number and volume of database. In order to provide future estimation, one needs to see the information available and applications conducted on these matters. Today, there are various algorithms and software for this purpose. The works of analysts became very easy thanks to these algorithms and software [1]. In order to reveal after a series of processes, the information that are hidden in huge amounts of data and very useful for the researches, the data should be analyzed and turned into intelligible data [2].

Human brain tends to find similarities on data. Aristoteles, the Greek philosopher, classified the living creatures on the basis of their living environments (air, water, land) [3]. Cluster analysis refers to the method which enables classification by gathering the units under the study within certain groups based on their similarities, reveal of common qualities of units and general definitions on these classes. The clustering method means the reorganization of data sets, that are the information on a sample with units, and these units to create groups based on some of their qualities [4]. The purpose of the cluster analysis is to divide n pieces of units or objects into homogenous clusters within themselves and heterogeneous clusters between themselves based on p pieces of variables [5]. Cluster analysis is a multivariate statistical method that intends to divide a set of observations into a limited number of groups and clusters. The division process at this stage is carried out in a manner to obtain similar observations in the same group and different observations in the different groups [6]. Distance measures, correlation measures or similarity measures of the data are used in order to determine the similarities between units or objects [7]. Although similarity measures are a convenient method for categorical data, distance and correlation measures are a convenient method for metric data [6]. Distance and similarity are reverse to each other and the fact that the coefficient obtained for similarity is small means that the units are distant from each other and big coefficient means

© 2016, IJCSE All Rights Reserved

that the units are close each other [8]. When the correlation coefficients obtained for dual units are compared, the pairs with big coefficient are in similar clusters and the pairs with small coefficient are in different clusters. On the other hand, having small distance coefficient means that the units are close to each other and vice versa. The most frequently used methods for distance calculations in the relevant literature are simple euclidian distance, quadratic Euclidean distance and city-block distance [8]. Duran and Odellstate that the most preferred method to determine the distance between two points in the two-dimensional space is the Euclidean distance method [9]. The used distance measures are useful but easily affected by the measurement units of the variables. For example, two individuals in a certain measurement unit are the most distant to each other while they become closer to each other when the measurement units vary which changes the order of distances between individuals. Therefore, it would be better to standardize the variables before calculating the distance [10].

Cluster analysis is a method that is commonly used in many studies and it provides groupings and summary information based on the basic qualities of units in social sciences like medicine, sociology, psychology, economics and marketing [11].The relevant literature includes several different approaches for the cluster analysis. However, the agreed point is that the clustering methods are divided into two like "hierarchical cluster" and "non-hierarchical cluster" methods. The hierarchical methods are also divided into two as merging methods (single link, complete link, average link, Ward's method and centroid method) and splitting methods (split average and automatic interaction detection). Non-hierarchical methods are divided into four: chi-average method, metoid partitioning method, accumulation clustering and fuzzy clustering method [12]. Hierarchical clustering methods are recommended for small sample group with number of samples under 250. Anderberg states that the non-hierarchical methods are used when researchers cannot predict the number of clusters in the concerned data set [13]. In addition, it is recommended that non-hierarchical methods should be used when the number of clusters is determined by the researcher [14]. The most distinct difference between the two methods is that the units are clustered until making a single cluster in hierarchical methods while the units in non-hierarchical methods are assigned to the pre-determined clusters [7].

A cluster of seven steps is used to organize the cluster process. These steps may vary depending on the practice and are listed as follows [15]. Firstly, the units to be clustered should be selected. Then, one should be careful to select the elements in the sample which represent the general nature of the cluster. Then, variables to be used in the cluster analysis should be selected. At this point, it is important to make sure that the selected variables contain sufficient information to allow cluster of individuals. Then, the researcher should decide whether the data will be standardized or not. After that, the distance or similarity criteria to be used in the analysis should be determined. Following the determination of the similarity criteria, one should select the cluster method in line with the purpose of the study. Hence, it is possible to reach different results by different methods. It is seen that different approaches provide different results in determining the number of clusters [8]. In determining the number of cluster, k refers to the number of cluster and n refers to the number of individuals participating in the research with $k=\sqrt{(n/2)}$ being one of most popular formulas in the relevant literature. However, this formula appears to be usable for small sample while it won't provide good results if the sample volume grows [16]. By a different approach, the value obtained at a point where the difference between the number of clusters and the distances of them are the highest can be determined as the total number of cluster [17].In a different approach, Wilk's lambda value is used in determining the number of clusters and it is aimed to approximate the λ value to zero. In practice, it is recommended to determine the number of clusters when this value goes under 0.01 for the first time as the number of clusters. The last and most important step in the cluster analysis is to test and interpret the results and apply them in other fields. Interpretation requires the researcher to be knowledgeable on the field of the application area. It includes the determination whether the tested clustering process is significant. At this stage, it is recommended that the consistency of the results obtained by different methods should be examined or the data group should be divided into two groups to review the clusters and cluster centres in order to determine the validity of the results [8]. A different approach recommends the comparison of clusters that are randomly obtained by the data group and the clusters obtained when all variables are included in the analysis. Çakmak recommends that the cluster results should be tested by the discriminant analysis in order to determine validity [18].

Health care organizations are among the organizations with the most complex structural and functional qualities. It is known that university hospitals in particular have a complex structure as they have a variety of employees from janitors to professors [19]. The health team providing service at all levels keeps the data on patient recorded on computer. Considering that the recorded information

© 2016, IJCSE All Rights Reserved

doubles every two years [20], this unprocessed data hardly provides any information [21]. In order to make future estimates from huge amounts of data, one needs to go back to past and see the types of information and applications on these matters available in the past. Today, there are several algorithms and software developed for this purpose. These algorithms and software are generally based on estimation or classification techniques and aim at creating the classification schemes of the empiric data which can be used to estimate the behaviour of objects currently unknown from the available data [22].

This study uses the cluster analysis method in order to provide significant results from the big data groups at hospitals. Cluster methods are used in this study to reveal the meaningful and useful data from the database of hospitals and to take more correct decisions for future. This study aims at determining the way of clustering of 40 different units at the Adnan Menderes University hospital according to some variables. It is believed with this study that decisions on units will be more correct and logical by determining similar and different aspects of the units included at hospitals. It is attempted to find response to the following sub problems within the scope of the basic problem of the study.

H_1: How are the 40 units operating at the University hospital clustered according to the variables of total expense, package loss and SGK deduction.

H_2: How are the 37 units operating at the University hospital clustered according to the variables of total expense, number of lecturers, number of assistants, number of nurses, number of personnel, number of polyclinic room number, polyclinic area, service bed rooms and service field?

2. Method

The data used in the study was obtained from the 2014 statistics of the Adnan Menderes University. Among the hierarchical clustering approaches, complete link method was used as the study attempted to determine the way of merging or partitioning of clusters. ANOVA test was applied to determine whether the independent variables to be used in the clustering analysis have a significant effect in clustering. In addition, chi-square means and discriminant analysis method was used in order to provide evidence on the validity of the cluster results obtained in the analysis. Discriminant analysis is a statistical technique used when the dependent variables are categorical and dependent variables are continuous [23]. It is defined as the assignment of a unit which is measured over the analysis to one of the different masses that are known to be in finite numbers [24]. The purpose of discriminant analysis isto assign the units to their groups and to determine their main original masses by minimizing the wrong classification probability [25].

The study covers the data analysis in two stages. At the first stage, all of the 40 units wereanalyzed and the variables of the unit like expense, package loss and SGK deduction were discussed. The second stage covered the analysis of 37 units and the variables of the unit like expense, number of lecturers, number of assistants, number of nurses, number of personnel, number of polyclinic room number, polyclinic area, service bed rooms and service field were discussed. The second sub problem of the study is that the units of family practice, radiology and nuclear medicine are not included in the analysis because they don't have data on the number of service beds and service fields as they provide ambulatory care. As lost data are not included in the analysis, the number of analyzed units in the second sub problem is reduced from 40 to 37.

3. Findings

This section provides information on the classification of countries, impact levels of the variables used in classification, determination of number of clusters and validity of the obtained results through 3 different affective qualities of the students by cluster analysis.

3.1 Findings on the First Sub Problem

The first sub problem of the study intends to determine the way of clustering the units operating at the university hospital. Table 1 includes the results on the variance analysis which was done in order to determine whether the variables included in the analysis for this purpose had a significant effect on clustering.

Table 1. Variance Analysis Table on the Variables Used in Clustering

	Cluster		Error		F	p
	Square mean	Degress of freedom	Square mean	Degress of freedom		
Expense	.317	7	.018	60	17.569	.000
Package loss	.221	7	.009	60	23.587	.000
SGK Deduction	.251	7	.010	60	24.116	.000

© 2016, IJCSE All Rights Reserved

As shown in Table 1, all of the independent variables used for the cluster analysis has a significant effect on clustering ($p<.05$). According to this result, it can be said that the clusters made by the variables used to cluster 40 different units have statistically significant difference. Figure 1 includes the dendogram obtained by the hierarchical cluster analysis carried out to determine the way of clustering of 40 units operating at the university hospital according to package loss, SGK deduction and other values among total expenses.

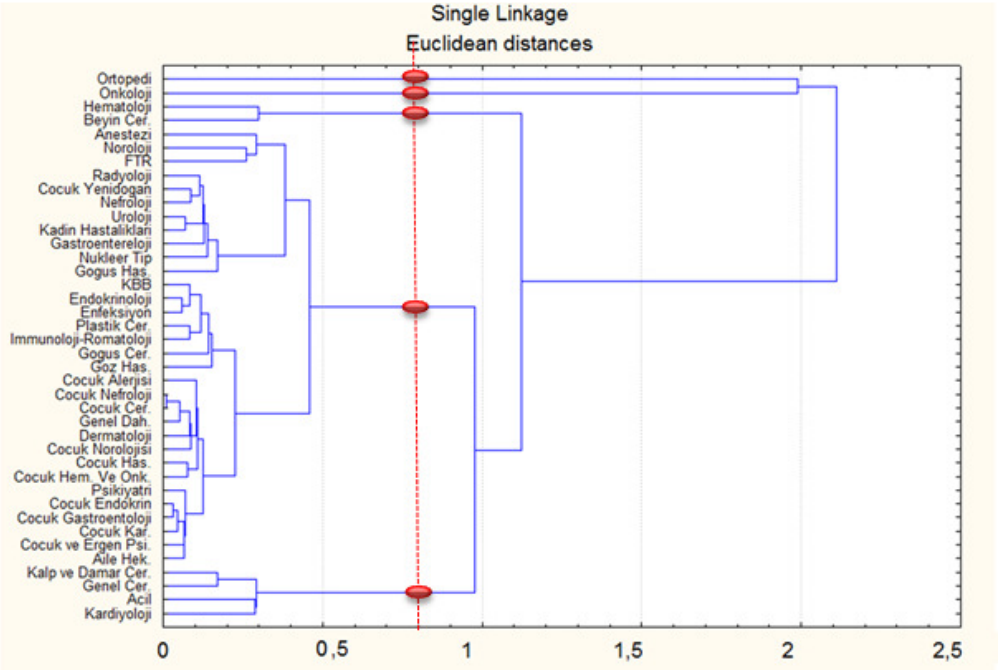

Figure 1. Dendogram as a result of Cluster Analysis

When we examine the Figure 1, we see clusters at each unit distance. The dendogram contains 5 clusters in a distance of 1 unit while the number of clusters is reduced to 3 in a distance of 1,5 units and to 2 in a distance of 2 units. One of the most important questions to be answered at this stage is how to determine the number of clusters. Djomou, Monkam and Woafosuggest the use of cutting index in determining the number of clusters in the dendogram[26]. As Figure 1 shows, the number of clusters over 1,5 distance value is very low and the number of clusters below the distance value of 0,3 is very high. According to this result, the cutting index to be used to determine the most convenient number of cluster is found to be between 0,30 and 1,50. In general, it is determined that the cutting index corresponds to the middle point of this distance, in another words there will be clusters in the number of intersectionbetween the rectangular line segment around the index value of 0,90 approximately and the dendogram.Accordingly, when the cutting index was accepted to be 0,90, 5 intersections are seen between the line segment and dendogram and it is believed that the appropriate number of clusters will be 5. Table 2 includes the findings as a result of the discriminant analysis on the validity of the number of clusters in the study.

Table 2. Wilk's Lambda Values

Cluster Number	Wilk's Lambda Value	Correct Classification Percentage
K=2	.254	% 99,90
K=3	.104	% 97,10
K=4	.060	% 92,60
K=5	**.009**	**% 94,10**
K=6	.006	% 91,40
K=7	.004	% 97,10
K=8	.009	% 94,10
K=9	.006	% 91,40

© 2016, IJCSE All Rights Reserved

K=10	.004	% 97,10

When we examine the Table 2, it is seen that the number of cluster is 5 when the Wilk's lambda value is below .01 which is regarded as the critical value. Accordingly, it was determined that the number of clusters determined by the dendogram is similar with the analysis results. Table 3 includes the clusters according to this result and the information on the countries in these clusters.

Table 3. Clustering Table on the Hospital Units

Number of Cluster	Number of Countries in the Cluster	Units
1	1	Orthopaedics
2	1	Oncology
3	2	Hematology, brain surgery
4	32	Anaesthesia, Neurology, FTR, Radiology, Neonatology, Nephrology, Urology, Gynaecological Diseases, Gastroenterology, Nuclear Medicine, Chest Diseases, Ear Nose and Throat, Endocrinology, Infection, Plastic Surgery, Immunology-Romatology, Thoracic Surgery, Eye Diseases, Children Allergy, Children Nephrology, Children Surgery, General Internal Medicine, Dermatology, Children Neurology, Pediatrics, Children Hematology and Oncology, Psychiatry, Endocrine, Gastroenterology, Children Cardiology, Children and Adolescent Psychology, Family Physician.
5	4	Cardiovascular surgery, General Surgery, Emergency, Cardiology.

Table 3 shows that the Oncology unit and the Orthopaedics unit is a cluster on its own. It is believed that the reason for these two units to make a single cluster is that both total expenses and service areas are far more than the other units. Especially total expenses of these units are far more than the average value obtained for 40 units. The study shows that a cluster is made by the units of cardiovascular, general surgery, emergency and cardiology. These four units are similar to each other with respect to total expense, number of nurses, number of personnel and service area. In addition, the other 32 units gather to make the last cluster in the study. Figure 2 shows the results of analysis with the multiple scaling techniques of the positions of the units at the hospital taking into consideration the distance measures on two-dimensional plane by the variables considered within the scope of the study.

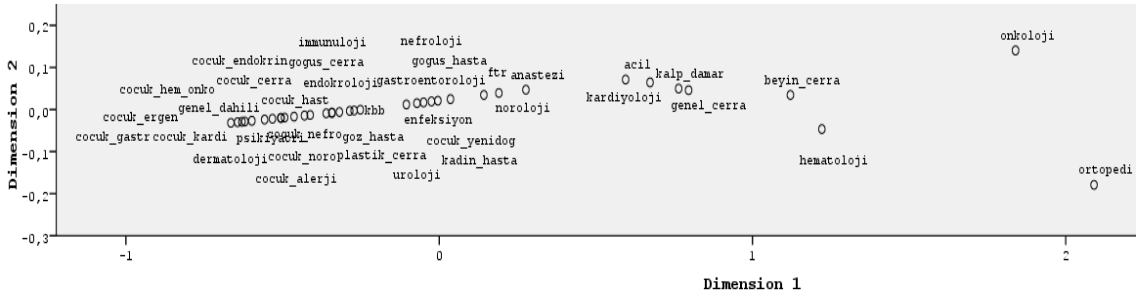

Figure 2. Multiple Scaling Results of the Hospital Units

When we examine the Figure 2, it was determined that the findings of the cluster analysis were similar with the multiple scaling techniques. The oncology unit and the orthopaedics unit which tend to create a cluster on their own in the cluster analysis have an apparent difference from the other units as a result of the multidimensional scale analysis. Again, the cluster analysis where the brain surgery and hematology units make a single cluster is similar with the multidimensional scaling. Likewise, the points of these two units on the two dimensional plane are close to each other. In addition, the units in this cluster are very close to each other as a result of the multidimensional scaling analysis of the 5th cluster consisting of the units of cardiovascular, general surgery, emergency and cardiology in the cluster analysis. According to these results, it is believed that the results obtained by the hierarchical cluster are valid.

3.2. Findings on the Second Sub Problem

The second sub problem of the study intends to determine the way of clustering of the units operating at the

© 2016, IJCSE All Rights Reserved

University hospital according to total expense, number of lecturers, number of assistants, number of nurses, number of personnel, number of polyclinic room number, polyclinic area, service bed rooms and service field. Table 4 includes the results of the variance analysis conducted to determine whether the variables included in the analysis have a significant effect in clustering.

Table 4. Variance Analysis Table on the Variables Used in Clustering

	Cluster		Error		F	p
	Square mean	Degress of freedom	Square mean	Degress of freedom		
Expense	.317	7	.018	60	17.569	.000
Package loss	.221	7	.009	60	23.587	.000
SGK deduction	.251	7	.010	60	24.116	.000
Number of lecturers	.240	7	.010	60	24.184	.000
Number of assistants	.191	7	.012	60	15.807	.000
Number of nurses	.263	7	.017	60	15.364	.000
Number of Personnel	.263	7	.016	60	16.487	.000
number of Polyclinic room number	.171	7	.011	60	15.807	.000
Polyclinic area	.246	7	.017	60	15.364	.000
Service bed rooms	.123	7	.014	60	16.487	.000
Service field	.238	7	.017	60	12.051	.000

As shown in Table 4, all of the independent variables used for cluster analysis has a significant effect ($p<.05$) on clustering. It can be said that clusters made by the variables used to cluster 37 different units have statistically significant difference from each other. Figure 3 includes the dendogram after the hierarchical clustering analysis conducted to determine the way of clustering of 37 units operating at the university hospital according to the 9 independent variables included in the study.

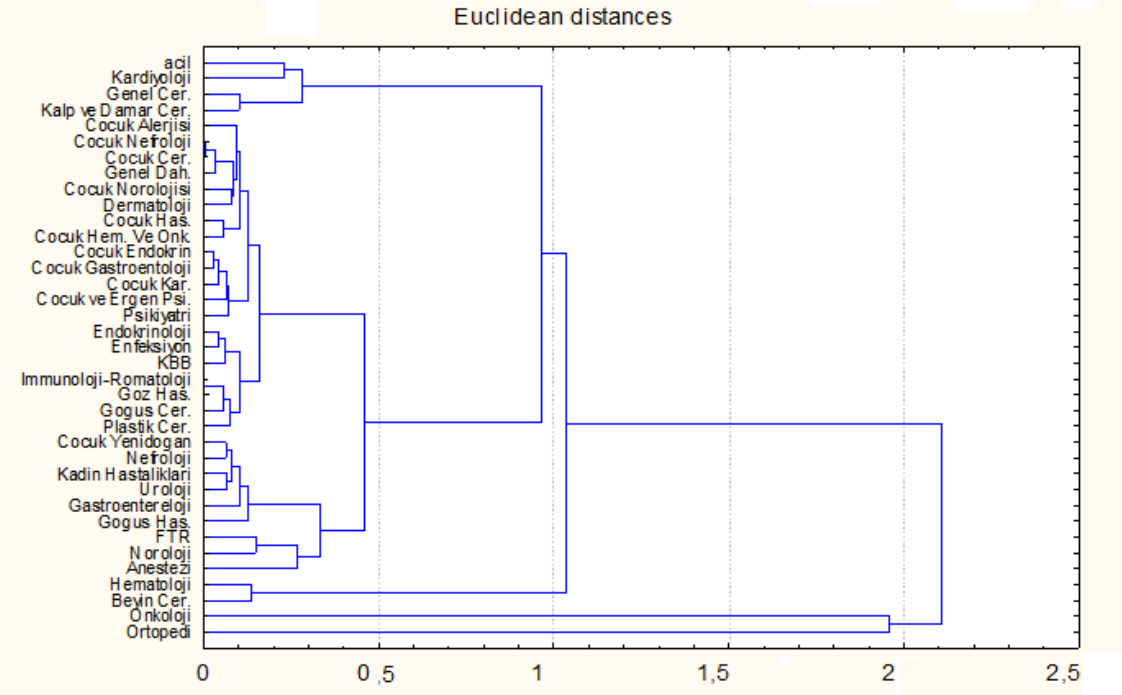

Figure 3. Dendogram Obtained as a result of the Cluster Analysis

When we examine the Figure 3, we see clusters at each unit distance. The dendogram contains 5 clusters in a distance of 1 unit while the number of clusters is reduced to 3 in a distance of 1,5 units and to 2 in a distance of 2 units. As Figure 3 shows, the number of clusters over 1,4 distance value is very low and the number of clusters below the distance value of 0,2 is very high. According to this result, the cutting index to be used to determine the most convenient number of cluster is found to be between 0,20 and 1,40. In general, it is determined that the cutting index

© 2016, IJCSE All Rights Reserved

corresponds to the middle point of this distance, in another words there will be clusters in the number of intersection between the rectangular line segment around the index value of 0,80 approximately and the dendogram. Accordingly, when the cutting index was accepted to be 0,80, 5 intersections are seen between the line segment and dendogram and it is believed that the appropriate number of clusters will be 5. Table 5 includes the findings as a result of the discriminant analysis on the validity of the number of clusters in the study.

Table 5. Wilk's Lambda Values

Number of Cluster	Wilk's Lambda Value	Correct Classification Percentage
K=2	.314	% 92,90
K=3	.216	% 93,10
K=4	.104	% 93,80
K=5	**.009**	**% 94,10**
K=6	.008	% 91,40
K=7	.006	% 92,10
K=8	.004	% 96,10
K=9	.002	% 98,40
K=10	.001	% 98,60

When we examine the Table 5, it is seen that the number of cluster is 5 when the Wilk's lambda value is below .01 which is regarded as the critical value. Accordingly, it was determined that the number of clusters determined by the dendogram is similar with the analysis results. Table 6 includes the clusters according to this result and the information on the countries in these clusters.

Table 6. Clustering Table on the Hospital Units

Cluster Number	Number of Countries in the Cluster	Units
1	4	Cardiovascular surgery, General Surgery, Emergency, Cardiology
2	29	Anaesthesia, Neurology, FTR, Radiology, Neonatology, Nephrology, Urology, Gynaecological Diseases, Gastroenterology, Chest Diseases, Ear Nose and Throat, Endocrinology, Infection, Plastic Surgery, Immunology-Romatology, Thoracic Surgery, Eye Diseases, Children Allergy, Children Nephrology, Children Surgery, General Internal Medicine, Dermatology, Children Neurology, Pediatrics, Children Hematology and Oncology, Psychiatry, Children Endocrine, Children Gastroenterology, Children Cardiology, Children and Adolescent Psychology
3	2	Hematology, brain surgery
4	1	Oncology
5	1	Orthopaedics

Table 6 shows that the Oncology unit and the Orthopaedics unit is a cluster on its own. It is believed that the reason for these two units to make a single cluster is that both total expenses and service areas are far more than the other units. Total expenses, number of lecturers, number of assistants, number of nurses, number of personnel, number of polyclinic room number, polyclinic area, service bed rooms and service field of these units are far more than the average value obtained for 40 units. The study shows that a cluster is made by the units of cardiovascular, general surgery, emergency and cardiology. Particularly the fact that the service area of the oncology unit (1006m2) is three times of the average value ($\bar{x}$=338m^2) explains why this unit is a single cluster on its own. Also the fact that the total expense of the unit of orthopaedics (9.844.830□) is around 4 times of the average total expense ($\bar{x}$=2.211.401□) explains why this unit is a single cluster of its own. In addition, it is observed that the units of hematology and brain surgery make one cluster. Looking at the independent factors affecting the clustering of these units, we see that the total expense, polyclinic area, polyclinic room number and service bed numbers for both units are very close to each other. Finally, it is observed that the units of cardiovascular, general surgery, emergency and cardiology make one cluster. These four units are similar to each other with respect to total expense, number of nurses, number of personnel and service area. In addition, the other 32 units gather to make the last cluster in the study. Figure 4 shows the results of analysis with the multiple scaling techniques of the positions of the units at the hospital taking into consideration the distance measures

© 2016, IJCSE All Rights Reserved

on two-dimensional plane by the variables considered within the scope of the study.

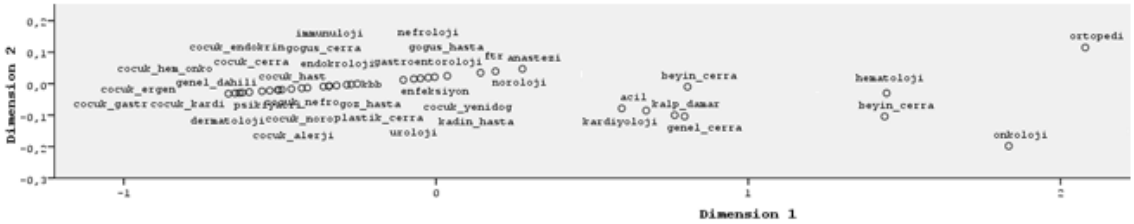

Figure 4. *Multiple Scaling Results of the Hospital Units*

When we examine the Figure 4, it was determined that the findings of the cluster analysis were similar with the multiple scaling techniques. The oncology unit and the orthopaedics unit which tend to create a cluster on their own in the cluster analysis have an apparent difference from the other units as a result of the multidimensional scale analysis. Again, the cluster analysis where the brain surgery and hematology units make a single cluster is similar with the multidimensional scaling. Likewise, the points of these two units on the two dimensional plane are close to each other. In addition, the units in this cluster are very close to each other as a result of the multidimensional scaling analysis of the 5th cluster consisting of the units of cardiovascular, general surgery, emergency and cardiology in the cluster analysis. According to these results, it is believed that the results obtained by the hierarchical cluster are valid.

4. Results and Recommendations

This study attempted to the determine the way the units operating at the university hospital in the Aydın province are clustered according to the variables of total expense, number of lecturers, number of assistants, number of nurses, number of personnel, number of polyclinic room number, polyclinic area, service bed rooms and service field. Hierarchical methods were used in the study as it intends to determine how to merge or split the clusters [27]. Complete link method among the hierarchical clustering methods was used for this purpose.

In order to the determine the validity of the results of the study, Wilk's lambda coefficient was calculated for different cluster numbers by the chi average method, which is a non-hierarchical method and it was determined that the number of clusters thus obtained is the same with the number of clusters determined by the complete link method. The study also determined that the units of orthopaedics and oncology, which tend to make one cluster on their own, are significantly different from other countries with respect to position in the 2-dimensional graphics obtained by the multidimensional scaling method. In addition, the units of hematology and brain surgery in the 3rd cluster and the cardiovascular, general surgery, emergency and cardiology units in the 5th cluster can be easily distinguished in visual from the other units. However, as there are 32 units in the 4th cluster, which is too many, it was determined that the multidimensional scaling method was not quite successful in distinguishing the elements in the cluster. As a result of the hypothesis tested in both the first and the second sub problem, the findings were seen to be similar. According to this result, it was determined that there is no change in the units in clusters as there are too many differences between units despite the significant effect of the variables like the expense, physical conditions and number of personnel of the unit in clustering the units.

With this study, the way of clustering the units in hospitals according to total expense, number of personnel and physical facilities and the similarities and difference between the units in the clusters were revealed in order to give an idea about the comparative status of the units with respect to specified variables. This requires that the managers at the universities should address to the units of orthopaedics, oncology, hematology and brain surgery departments differently from other departments in their decisions on units. The findings of the study are believed to be important particularly for the budgets and investments to be allocated to the unit. We believe that application of the study with national and international expansion will further clarify its effectiveness and provide more insight to the managers.

REFERENCES

[1] Mannila, H. Data mining: Machine learning, statistics, anddatabases. *Proc. Of the 8th International Conference on Scientificand Statistical Database Management*, **1996**, 1-6

[2] Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth. Fromdataminingtoknowledge discovery in databases. *AI Magazine,* **1996**, 37-54.

[3] Yünel Y. K-means Kümeleme Algoritmasının Genetik Algoritma Kullanılarak Geliştirilmesi, p. 1-2.Zeytinoğlu, F. Ç. (2014) Kümeleme Analizi: Kültür İstatistiklerine Göre İllerin Sınıflandırılmasına Yönelik

© 2016, IJCSE All Rights Reserved

Bir Çalışma, *İstanbul Ticaret Üniversitesi Sosyal Bilimleri Dergisi*,**2010**, 13 (25), 301-320.
[4] Çokluk, Ö, Şekercioğlu, G. ve Büyüköztürk, Ş. *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve Lisrel Uygulamaları,* **2014**,Ankara: Pegem Akademi.
[5] Kaya, V. ve Türkmen, A. Küresel Krizin Üst Orta Gelir Gurubu Ülkelere Makro Ekonomik Yansımaları, *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, **2013,** 27 (4), 317-338
[6] Neil, T.H. *AppliedMultivariate Analysis,*Springer-Verlag, New York, **2002**
[7] Kalaycı, Ş. *SPSS Uygulamalı Çok Değişkenli İstatistik*, Ankara: Asil Yayın Dağıtım, **2014**
[8] Alpar, R. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler,* Ankara: Detay Yayıncılık, **2011**
[9] Duran, B.S. ve Odell, P.L. Cluster analysis: lecturenotes in economicsandmathematicalsystems, *Econometrics.* New York: Springer-Verlag.**1974**
[10] Aldenderfer, M. S. ve Blashfield, R. K. *Cluster Analysis*, Beverly Hills: SagePublications, **1984**
[11] Zeytinoglu, I. U.,Denton, M., Davies, S., Baumann, A., Blyte, J. ve Boos, L.Retainingnurses in theirhospitalsand in theprofession: Effects of jobpreference, unpaidovertime, importance of earningsandstress. *HealthPolicy*, **2006**,79, 57-72.
[12] Gürsoy, U.T. ve Şimşek, G. *Veri Madenciliği ve Bilgi Keşfi*. Ankara: PegemAkademi, **2009**
[13] Anderberg, M. R. *Cluster Analysis forapplications*, New York: AcademicPress, **1973**
[14] Özdamar, K. *Paket Programlar ile İstatistiksel Veri Analizi*. Eskişehir: Kaan Kitabevi, **2004**
[15] Milligan, G. W. ve Cooper, M. C. A Study of theComparability of ExternalCriteriaforHierarchical Cluster Analysis,*MultivariateBehavioralResearch*, **1986**, 21,441-458.
[16] Tatlıdil, H. *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, H.Ü. Fen Fakültesi İstatistik Bölümü, Ankara, **1992**
[17] Mardia, K. V., Kent, J. T. ve Bibby, J. M. *Multivariate Analysis,* London: AcademicPress, **1989**.
[18] Çakmak, Z. Kümeleme Analizinde Geçerlik Problemi ve Kümeleme Sonuçlarının Değerlendirilmesi, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi,* **1999**,3, 187- 205.
[19] Aslan, Ş., Özata, M. ve Atayeter, C. Sağlık İşletmelerinde Ekip yönetimi: Fırsatlar ve Sınırlılıklar,*Standard Ekonomik ve Teknik Dergi,***2004**, 43 (516), 17-23.
[20] Frawley, W.J. Piatetsky-Shapiro, G. Matheus, C.J. *Knowledge discovery in databases*, AAAI/MIT Press, **1991**, pp. 1–27
[21] Larose, D. T. *k-NearestNeighborAlgorithmknowledge in Data: An Introduction to DataMining*, ABD: Wiley Online Library, **2005**
[22] Weiss, S. M.andKulikowski, C. A. *Computersystemsthatlearn*. San Mateo, CA: Morgan Kaufmann,**1991**
[23] Poulsen, J. ve French, A. *Discriminantfunctionanalysis* (DA).Retrieved August152015, fromhttp://userwww.sfsu.edu/efc/classes/biol710/discrim/discrim.pdf, **2003**
[24] Aaker,D. A, Kumar, V. andDay, S. G. *Marketing Research*, Fifth Edition, Canada, John WileySons, Inc. **1997**.
[25] Klecka, W. *Discriminant Analysis*. London: Sage Publications, **1980**.
[26] Djomou, Z. Y.,Monkam, D. ve Woafo, P. Variabilityandtrends of local/regional scale surface climate in northern Af ricaduring the twentieth century, *Theoretical and Applied Climatolog,* **2013**,117, 625-641.

Author Profile

*Özel Sebetci is AssistantProfessorat Departmentof ComputerTechnology İn Adnan MenderesUniversity, Turkey.

** Gökhan Aksu is an Instructor at Aydın Vocational School in Adnan Menderes University, Turkey.

© 2016, IJCSE All Rights Reserved