American Research Journal of Humanities & Social Science (ARJHSS)
E-ISSN: 2378-702X
Volume-03, Issue-04, pp 58-64
April-2020
www.arjhss.com

Research Paper **Open Access**

Classification by the Provincial Culture Structure in Turkey: Fuzzy Clustering Analysis Approach Based on Factor Analysis

Prof. Dr. Ünal Halit ÖZDEN[1], Associate. Prof. Dr. Seda BAĞDATLI KALKAN[2]

[1,2]*Istanbul Commerce University, Department of Statistics, İstanbul*
**Corresponding Author: Seda BAĞDATLI KALKAN*

ABSTRACT:- Turkey Statistical Institute of Statistics culture each year to the provinces in Turkey (TSI) are presented to the public with the work done by. The purpose of this study is to determine the culture profile of the provinces and to plan the studies to be done to ensure equality between provinces. Data on cinema, theater, public libraries and museums are collected within the context of cultural statistics of the provinces. In this study, cinema, theater and library statistics are included in the analysis process since there are no museums in each province. In the study, it is aimed to classify the provinces according to their cultural structure. Accordingly, four stages were carried out. In the first stage, classical cluster analysis was applied and a logical cluster was not achieved in the provinces. In the second stage, after applying factor analysis to the data, classical cluster analysis was applied to factor scores, but again, it was found that there was no logical clustering in terms of the unity of the provinces. In the third stage, Fuzzy cluster analysis was applied to the data. At this stage, it was determined that there was no logical cluster in terms of the unity of the provinces and the number of provinces in the clusters. At the last stage, Fuzzy cluster analysis was applied to factor analysis scores. As a result, it is determined that 3 more clusters are formed that are much more logical than other methods. According to this analysis, the results are interpreted in detail.

Keywords:- Cultural Structure of the Provinces, Fuzzy Cluster Analysis, Factor Analysis Jel Codes: C10, C38

I. INTRODUCTION

Culture is the whole material and spiritual values created in the process of social development and the tools used to create them and convey them to the next generations, which show the measure of the sovereignty of the human to its natural and social environment. As can be understood from the definition of culture, culture reveals the differences of societies compared to other societies. The cultural level of a society also significantly affects the level of development. Therefore, studies on culture are of great importance for social development. However, determining the cultural structure of the provinces is also an important factor in terms of planning tourism and cultural policies. TUIK helps to identify the cultural deficiencies of the provinces with the cultural statistics it shares with the public every year. In addition, the determination of the culturally rich provinces will enable the evaluation of these provinces from other perspectives.

Literature provinces in Turkey, to determine the structure of culture have been some studies. Çakmak, Uzgören, & Keçek classified the provinces using their hierarchical and non-hierarchical cluster analysis techniques according to their cultural statistics. By carrying out this study for 1990 and 2000, they provided comparison of provinces at the cultural level. As a result of the study, Istanbul was found as a single cluster in all cluster analyzes. Çakır Zeytinoğlu carried out the classification of the provinces in accordance with the cultural statistics determined by TUIK with the classical cluster analysis in his study in 2014. At the end of the study, three and six clusters are compared. In both cases, Istanbul was designated as a cluster alone. In the literature, it is seen that cluster analysis is frequently used in the classification of provinces by using other

[1]uozden@ticaret.edu.tr, Istanbul Commerce University, Department of Statistics, İstanbul
[2]sbagdatli@ticaret.edu.tr, Istanbul Commerce University, Department of Statistics, İstanbul

variables besides culture. Overhead, clustering analysis using twofold in 2018 in his study has identified potential second-tier cities in Turkey. Allahverdi & Alagöz classified the provinces with cluster analysis in terms of tax revenues in their study in 2019. Bulut, the classification of his work in 2019 by the provinces of living index in Turkey was carried out using cluster analysis. In this study, Fuzzy cluster analysis was preferred over classical cluster analysis. This is because the Fuzzy cluster has more information about the actual location of the units, since the cluster membership degrees, that is, the probability of entering the cluster, are calculated and the units are not forced to belong to only one cluster. In many studies, it is seen that the provinces are classified using Fuzzy cluster analysis. Bülbül & Camkıran classified the banks with classic and fuzzy approaches in their study in 2018. As a result of the study, it is concluded that classical and fuzzy methods do not give very different results but fuzzy cluster analysis is more suitable for some banks. Şimşek Kandemir classified the provinces according to the accommodation statistics in his study in 2018. In the study, the fact that the cluster possibilities of some provinces are very close to each other is concluded with fuzzy clustering because it is the subject.

This study consists of four stages. In the first stage, classical cluster analysis was applied and a logical cluster was not achieved in the provinces. In the second stage, after applying factor analysis to the data, classical cluster analysis was applied to factor scores, but again, it was found that there was no logical clustering in terms of the unity of the provinces. In the third stage, Fuzzy cluster analysis was applied to the data. At this stage, it was determined that there was no logical cluster in terms of the unity of the provinces and the number of provinces in the clusters. At the last stage, Fuzzy cluster analysis was applied to factor analysis scores. As a result, it is determined that 3 more clusters are formed that are much more logical than other methods.

II. MATERIALS AND METHODS

2.1. Data Source

In this study, the cultural statistics of the provinces that are shared with the public regularly every year were used by TUIK. Finally, since the culture statistics for 2018 were published, 2018 data were used in the study. TUIK collects data from 21 variables under the heading of cultural statistics of provinces. These variables are under 4 main headings: public libraries, museums, cinema and theater. In this study, the variables related to museums were excluded from the analysis since there are no museums in each province. The variables used in the study are shown in Table 1.

Table 1: Variables Used in the Study

Public libraries: Number of benefits per thousand
Public libraries: Number of libraries
Public libraries: Number of items on loan
Public libraries: Number of benefits
Cinema: Number of shows
Cinema: Number of spectators
Cinema: Number of movie theaters
Theater: Number of shows
Theater: Number of spectators
Theater: Number of theater halls

2.2. Fuzzy Clustering

All of the clustering algorithms used in classical cluster analysis assigns the observations precisely to a cluster. So the probability of an observation in any set is 1 or 0. In the Fuzzy cluster analysis, the possibilities of each observation to take place in the clusters, ie cluster membership probabilities, are determined. Certainty in classical clustering approaches causes erroneous results in some cases. In the data set, when there are observation units that are equidistant to each of the homogeneous sets, it is not known exactly to which set the units will be assigned. In this case, the possibilities of belonging to clusters gain importance. Fuzzy clustering analysis allows all data points to be elements of more than one cluster simultaneously with different membership values. The most widely used method in fuzzy cluster analysis is Fuzzy c-means clustering algorithm. In this algorithm, it is aimed to minimize the cost function calculated from distances and cluster memberships (Bağdatlı Kalkan, 2019) .The purpose function of the algorithm is shown in equation 1.

$$J = (X;\ U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^{m} \left\| x_j - v_i \right\|^2 \qquad (1)$$

In equation 1, c is the number of clusters and n is the number of units. *V is the cluster center, the prototype vector.* $\left\| x_j - v_i \right\|^2$ *if* D_{ij}^2 *expressed by Euclidean distance.*. *U,* μ_{ij} the membership matrix and *m* is the weighting

exponent coefficient (Yang, 1993). Fuzzy c-means algorithm decides which unit to assign to which cluster by using membership degrees. Each unit is assigned to that cluster, which membership is large. However, each unit can also become a member of other clusters with certain membership degrees (Atal, 2015). In fuzzy cluster analysis, specific indexes have been created to decide on the number of clusters or to check whether the analysis can provide an appropriate classification. These indices are called Cluster Validity Index. There are many cluster validity indexes in the literature. However, there is no exact information as to which index works best. Cluster validity indexes can be formulated in different ways for special cases in data sets. Therefore, new indices continue to be developed.

III. RESULTS

This study consists of four stages. In the first stage, k-means cluster analysis was applied to the data and no logical distinction was achieved. In the second stage, after applying factor analysis to the data, k-means cluster analysis was applied to factor scores. As a result of this analysis, 3 clusters were formed, one of which was Istanbul. There are 67 provinces in the second cluster and 13 provinces in the 3. cluster. However, it was determined that there is not a very logical cluster in terms of the unity of the provinces. In the third stage, fuzzy cluster analysis was applied to the data. As a result of this analysis, similar to the previous stage results, 3 clusters were formed, namely Istanbul one cluster. However, no logical distinction regarding provinces has been achieved. At the last stage of the study, factor analysis was applied to the data and fuzzy clustering analysis was performed with factor analysis scores. As a result of this analysis, 3 clusters were formed, but it was determined that the most logical results were achieved in terms of the unity of the provinces. Therefore, the results of the analysis of the last stage will be interpreted in this study. In the study, factor analysis was performed with SPPS 24, and cluster analysis with R Studio. "Ppclust", "factoextra", "cluster" and "fclust" packages were used in the R Studio program.

Factor analysis results are shown in Table 2 and Table 3.

Table 2: KMO and Bartlett Test Results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.871
Bartlett's Test of Sphericity	Approx. Chi-Square	2019.156
	df	45
	Sig.	.000

As seen in Table 2, since the KMO test is 87.1% (> 0.50), the data set is suitable for factor analysis. Bartlett's sphericity tests whether there is a high correlation between variables. Since p <0.05, it is concluded that there is a high correlation between the data. After this stage, Table 3 was created to decide the number of factors and to determine the percentage of variance explained.

Table 3: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.001	70.013	70.013	7.001	70.013	70.013	6.282	62.817	62.817
2	1.698	16.977	86.990	1.698	16.977	86.990	2.417	24.173	**86.990**
3	.753	7.527	94.517						
4	.284	2.835	97.352						
5	.206	2.059	99.411						
6	.035	.350	99.762						
7	.017	.166	99.928						
8	.004	.044	99.971						
9	.001	.015	99.986						
10	.001	.014	100.000						
Extraction Method: Principal Component Analysis.									

As can be seen in Table 3, there are two factors. Two factors explain about 87% of the total variance. Table 4 was created to determine the factors where the variables are located.

Table 4: Rotated Component Matrix

	Component	
	1	2
Cinema: Number of spectators	.974	
Theater: Number of shows	.972	
Cinema: Number of movie theaters	.970	
Cinema: Number of shows	.970	
Theater: Number of spectators	.968	
Theater: Number of shows	.963	
Public libraries: Number of benefits		.865
Public libraries: Number of items on loan	.517	.739
Public libraries: Number of libraries		.732
Public libraries: Number of benefits per thousand		.589
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.		
Rotation converged in 3 iterations.		

Table 4 shows the rotated component matrix created using the Varimax rotation method. When this table is examined, the first factor can be named as the cinema and theater factor, and the second factor as the library factor.
Table 5 shows the results of fuzzy cluster analysis applied to factor scores.

Table 5: Cluster Memberships

İller	Cluster 1 Membership Degree	Cluster 2 Membership Degree	Cluster 3 Membership Degree	Cluster Membership
İstanbul	0.000247502	0.999515266	0.000237233	2
Tekirdağ	0.975933643	0.001834991	0.022231366	1
Edirne	0.992479442	0.000365125	0.007155432	1
Kırklareli	0.996149539	0.000195996	0.003654465	1
Balıkesir	0.672526258	0.007364934	0.320108808	1
Çanakkale	0.999923575	0.000004331	0.00007209	1
İzmir	0.302346528	0.03641	0.6612	3
Aydın	0.14190369	0.004402	0.8537	3
Denizli	0.730857803	0.00698	0.2622	1
Muğla	0.889282571	0.003907	0.1068	1
Manisa	0.001304671	0.00007287	0.9986	3
Afyonkarahisar	0.974129892	0.001109	0.02476	1
Kütahya	0.670867682	0.006701	0.3224	1
Uşak	0.999385631	0.00003549	0.0005789	1
Bursa	0.857969304	0.008341	0.1337	1
Eskişehir	0.97138042	0.00203	0.02659	1
Bilecik	0.978080772	0.001563	0.02036	1
Kocaeli	0.888212682	0.01146	0.1003	1
Sakarya	0.968273395	0.002547	0.02918	1
Düzce	0.985732888	0.001001	0.01327	1
Bolu	0.986355758	0.0009653	0.01268	1
Yalova	0.933584773	0.006361	0.06005	1

Ankara	0.313700477	0.09813	0.5882	3
Konya	0.246990922	0.0673	0.6857	3
Karaman	0.993012523	0.0003825	0.006605	1
Antalya	0.677559719	0.01296	0.3095	1
Isparta	0.086602484	0.008251	0.9051	3
Burdur	0.998282914	0.00009658	0.001621	1
Adana	0.855232012	0.005747	0.139	1
Mersin	0.771441143	0.006392	0.2222	1
Hatay	0.745883221	0.006725	0.2474	1
Kahramanmaraş	0.994234702	0.000377	0.005388	1
Osmaniye	0.99050972	0.000455	0.009035	1
Kırıkkale	0.785363436	0.005552	0.2091	1
Aksaray	0.976819397	0.001045	0.02214	1
Niğde	0.99762618	0.0001311	0.002243	1
Nevşehir	0.039875603	0.001686	0.9584	3
Kırşehir	0.200509381	0.03082	0.7687	3
Kayseri	0.001785086	0.0001102	0.9981	3
Sivas	0.046672084	0.003708	0.9496	3
Yozgat	0.997439675	0.0001325	0.002428	1
Zonguldak	0.97503039	0.001947	0.02302	1
Karabük	0.979524382	0.0009153	0.01956	1
Bartın	0.92175148	0.007896	0.07035	1
Kastamonu	0.960197749	0.003393	0.03641	1
Çankıri	0.961758155	0.001601	0.03664	1
Sinop	0.982481301	0.0008335	0.01669	1
Samsun	0.907618889	0.003234	0.08915	1
Tokat	0.958932112	0.001629	0.03944	1
Çorum	0.124683099	0.003901	0.8714	3
Amasya	0.990615412	0.0006265	0.008758	1
Trabzon	0.974520515	0.001176	0.0243	1
Ordu	0.116767914	0.003662	0.8796	3
Giresun	0.749232222	0.005904	0.2449	1
Rize	0.982754056	0.001256	0.01599	1
Artvin	0.951615443	0.004165	0.04422	1
Gümüşhane	0.973462882	0.001982	0.02455	1
Erzurum	0.386442292	0.007031	0.6065	3
Erzincan	0.984976033	0.0007202	0.0143	1
Bayburt	0.93717306	0.005727	0.0571	1
Agrı	0.993935902	0.0003541	0.00571	1
Kars	0.987798283	0.0006776	0.01152	1
Iğdır	0.98710277	0.0007749	0.01212	1
Ardahan	0.88449987	0.01358	0.1019	1

Malatya	0.975522156	0.001116	0.02336	1
Elazığ	0.503583724	0.00742	0.489	1
Bingöl	0.988446257	0.000738	0.01082	1
Tunceli	0.936027405	0.005957	0.05802	1
Van	0.991193944	0.0005886	0.008217	1
Mus	0.971405316	0.002214	0.02638	1
Bitlis	0.833729716	0.004774	0.1615	1
Hakkari	0.977494515	0.001568	0.02094	1
Gaziantep	0.982013418	0.0009902	0.017	1
Adıyaman	0.932490348	0.002431	0.06508	1
Kilis	0.973075057	0.001387	0.02554	1
Şanliurfa	0.996107843	0.0002083	0.003684	1
Diyarbakır	0.0945068	0.003391307	0.902101894	3
Mardin	0.9987596	0.000066485	0.00117393	1
Batman	0.9953227	0.000284435	0.004392895	1
Şırnak	0.9642733	0.00287793	0.03284873	1
Siirt	0.9684297	0.002442314	0.029128012	1

When Table 5 is examined, 66 provinces are in the 1st cluster, only Istanbul is in the 2nd cluster and 14 provinces are in the 3rd cluster. Provinces are assigned to the cluster with a high membership level. To study the structure of the clusters, Table 6 was created.

Table 6: Final Cluster Prototype

	Cinema and Theater Factor	Library Factor
Cluster 1	-0.08736893	-0.4189177
Cluster 2	7.98176535	0.6896953
Cluster 3	-0.28506304	1.5689843

When Table 6 is examined, the cluster with the highest cinema and theater factor is the 2nd cluster and then the 1st. and finally the third cluster is located. The cluster with the highest library factor is 3. The cluster is then cluster 2 and the last is cluster 1.
After this stage, the cluster validity indices were calculated and shown in Table 7.

Table 7: Cluster Validity Indices

Dunn's Fuzziness Coefficient	**0.8643361**
Partition Entropy	0.245713
Partition Coefficient	0.8643361
Modified Partition Coefficient	0.7965041
Fuzzy Silhouette Index	0.833974

Validity indices shown in Table 7 are used to decide the number of clusters rather than the validity of cluster analysis. Of these indices, only the Partition Entropy value is mini-candle, and the other indices are the maximum desired. Cluster validity indices do not provide precise information, but give insight into the validity of cluster analysis.

IV. CONCLUSIONS

The concept of culture, at first glance, is considered only as a concept regarding human beings. However, culture is made up of societies' unique customs and traditions. It carries a culture of society to future generations. Therefore, the concept of culture has a very important role in our lives from past to present. Measuring the level of culture on the basis of both people and communities is very complex. Because the concept of culture differs from person to person and therefore from society to society. In this study, rather than measuring the level of culture provinces in Turkey, cultural structures which are intended to elicit similar cities.

For this purpose, cultural data of provinces collected by TUIK were used. These data are determined as cinema, theater, library and museum data. Since there are no museums in every province, museum data are excluded from the study. In the study, fuzzy cluster analysis was applied to factor analysis scores. As a result of the application, 3 clusters were obtained. The first cluster consists of Tekirdağ, Edirne, Kırklareli, Balıkesir, Çanakkale, Denizli, Muğla, Afyonkarahisar, Kütahya, Uşak, Bursa, Eskişehir, Bilecik, Kocaeli, Sakarya, Düzce, Bolu, Yalova, Karaman, Antalya, Burdur, Adana, Mersin, Hatay, Kahramanmaraş, Osmaniye, Kırıkkale, Aksaray, Niğde, Yozgat, Zonguldak, Karabük, Bartın, Kastamonu, Çankırı, Sinop, Samsun, Tokat, Amasya, Trabzon, Giresun, Rize, Artvin, Gümüşhane, Erzincan, Bayburt, Ağrı, Kars, Iğdır, Ardahan, Malatya, Elazığ, Bingöl, Tunceli, Van, Muş, Bitlis, Hakkari, Gaziantep, Adıyaman, Kilis, Şanlıurfa, Mardin, Batman, Şırnak and Siirt provinces. The second cluster consists of 1 city, only Istanbul. The third cluster consists of İzmir, Aydın, Manisa, Ankara, Konya, Isparta, Nevşehir, Kırşehir, Kayseri, Sivas, Çorum, Ordu, Erzurum and Diyarbakır provinces. As it is seen, Istanbul is a cluster on its own since it is very different from other provinces. In previous studies, it is seen that Istanbul is a cluster alone. Although Istanbul is the best cluster in cinema and theater factor, it is in the 2nd place in library factor. The worst cluster in cinema and theater factor is the third cluster. It is a surprising result that there are cities such as İzmir and Ankara in this cluster, but when the data are analyzed, it is seen that the number of spectators is low despite the number of shows. Although the living conditions in these provinces are much better than in other provinces, it is concluded that it cannot devote time to activities such as theater and cinema especially because of the high traffic and business life. In terms of library factor, it is seen that cluster 3 is the best. Consequently, 2018 provides relevant information since this study is carried out with 2018 data only. The realization of this study with the data announced in the following years and a comparative analysis of the results will enable the identification of similar cities in terms of culture. Therefore, it will help to form the necessary policies for the cultural level of the provinces to be at the best level. As a new method in the study, it was tried to contribute to the literature by applying fuzzy cluster analysis to factor analysis scores. In addition, it is thought that expanding the study with various culture-related variables and other analysis methods will contribute to the literature.

REFERENCES

[1]. Allahverdi, M., & Alagöz, A. (2019). İllerin Vergi Gelirleri Açısından Sınıflandırılmasında Kümeleme Analizi Kullanımı. *Maliye Degisi*, s. 441-473.

[2]. Atal, S. (2015). Yayınlanmamış Yüksek Lisans Tezi. *Bulanık Kümeleme Analizi ve OECD Ülkelerinin Gelişmişlik Bakımından Kümelendirilmesi*. Eskişehir: Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Ensitüsü.

[3]. Bağdatlı Kalkan, S. (2019). Classification of European Union Member and Candidate Countries Acorrding to Ease of Doing Business Index Using Fuzzy Clustering. *Eurasian Econometrics, Statistics and Emprical Economics Journal*, s. 25-35.

[4]. Bulut , H. (2019). Türkiye'deki İllerin Yaşam Endekslerine Göre Kümelenmesi. *Süleyman Demirel Enstitüsü Fen Bilimleri Dergisi*, s. 74-82.

[5]. Bulut, H. (2018). *R Uygulamaları İle Çok Değişkenli İstatistiksel Yöntemler.* İstanbul: Nobel Yayıncılık.

[6]. Bülbül , Ş., & Camkıran, C. (2018). Bankaların Klasik ve Bulanık Yaklaşımlarla Sınıflandırılması. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, s. 367-385.

[7]. Çakır Zeytinoğlu, F. (2014). Kümeleme Analizi: Kültür İstatistiklerine Göre İllerin Sınıflandırılmasına Yönelik Bir Çalışma. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, s. 301-320.

[8]. Çakmak, Z., Uzgören, N., & Keçek, G. (2005). Kümeleme Analizi Teknikleri İle İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerinin İncelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, s. 1-21.

[9]. Gezer, B. (2018, Kasım). Yayınlanmamış Yüksek Lisans Tezi. *Türkiye'nin Potansiyel İkinci Kademe Şehirlerinin Belirlenmesi ve Performans Analizi*. İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

[10]. Şimşek Kandemir, A. (2018). Bulanık Kümeleme Analizi ile Türkiye'deki İllerin Konaklama İstatistiklerine Göre Sınıflandırılması. *Seyahat ve Otel İşletmeciliği Dergisi*, s. 657-558.

[11]. Türk Dil Kurumu. (tarih yok). *Türk Dil Kurumu*. Mart 2020 tarihinde https://sozluk.gov.tr/?q=&aranan= adresinden alındı

[12]. Yalçın, M. O., & Güler Dinçer, N. (2015). Bulanık Kümeleme Analizi Kullanarak Türkiye'yi Ziyaret Eden Yabancı Turistlerin Profilinin Belirlenmesi. *Finans Politik & Ekonomik Yorumlar*, s. 25-37.

[13]. Yang, M. (1993). A Survey of Fuzzy Clustering. *Mathematical and Computer Modelling*, s. 1-16.

Corresponding Author: Dr. Seda BAĞDATLI KALKAN
[1,2]Istanbul Commerce University, Department of Statistics, İstanbul