

## TÜRKİYE' DE İLLERİN EĞİTİM DÜZEYLERİNE GÖRE KÜMELEME ANALİZİ İLE İNCELENMESİ

Dicle CENGİZ\*  
Faruk ÖZTÜRK\*\*

### ÖZET

Ülkelerin gelişmişlik düzeylerini belirleyen önemli göstergelerden biri de eğitimidir. Türkiye de bu alanda politikalar üretmekte ve gelişmiş ülkelerin standartlarına ulaşmak istemektedir. Bu bağlamda öncelikle okullaşma oranının artırılmasına yönelik politikalar ile her ilin eğitilmiş nüfusunun artırılması ve gelişme düzeyinin yükseltilmesi amaçlanmaktadır.

Türkiye'deki illerin eğitim düzeylerini belirleyen birden çok parametre vardır. Bu çalışmada illerin eğitim düzeyleri, okuma-yazma bilmeyen, okuma-yazma bilen fakat bir okul bitirmeyen, ilköğretim mezunu, ortaöğretim mezunu, yükseköğretim mezunu, yüksek lisans mezunu, doktora mezunu ve eğitim düzeyi bilinmeyen oranları yardımı ile tespit edilmeye çalışılacaktır. Bu nedenle illerin benzerliklerinin belirlenmesinde çok değişkenli istatistik analiz yöntemlerinden olan kümeleme analizi kullanılmıştır.

*Anahtar Kelimeler:* Eğitim düzeyi, Çok değişkenli analiz, Kümeleme analizi, K-Ortalama yöntemi.

### THE STUDY OF THE CITIES IN TURKEY WITH CLUSTER ANALYSIS ACCORDING TO THEIR EDUCATION LEVELS

### ABSTRACT

Education is one of the important indicators determining the development levels of countries. Turkey produces politics and wills to catch the developed countries standards in that area.

There are numbers of parameters to determine the education levels of the cities in Turkey. In that study, the education levels of the cities are defined with the

\* Doç. Dr., İstanbul Ticaret Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü.

\*\* İstanbul Ticaret Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, mail: mfarukoz@yahoo.com

rates of; illiterate, literate but haven't finished a school, graduated from a primary school, graduated from a secondary school, graduated from a university, being a grad student, having a doctorate and unknown. Therefore, among the multivariate analysis methods cluster analysis is used for determining the similarities of the cities.

**Key Words:** *Education level, Multivariate analysis, Cluster analysis, K-Means method.*

## 1. GİRİŞ

Son yıllarda her alanda ilerleme kaydeden Türkiye, eğitim alanında da Avrupa Birliği Standartlarını yakalamayı amaçlamaktadır. Bu nedenle eğitim alanında politikalar üretmekte ve yatırımlar yapmaktadır. Özel okulların da katkısıyla eğitimin her kademesine yönelik derslikler açılmakta, öğretmen kadrosu artırılmaktadır. Bu amaçlar doğrultusunda Türkiye'nin her ilinde eğitimli birey oranlarının artırılması hedeflenmektedir.

Bu çalışmanın amacı, Türkiye'deki iller arasındaki benzerlikleri eğitim düzeyini belirleyen faktörler yönüyle belirleyip, benzer özellikler taşıyan illeri çok değişkenli istatistiksel analiz tekniklerinden olan kümeleme analiziyle kümelere ayırmaktır.

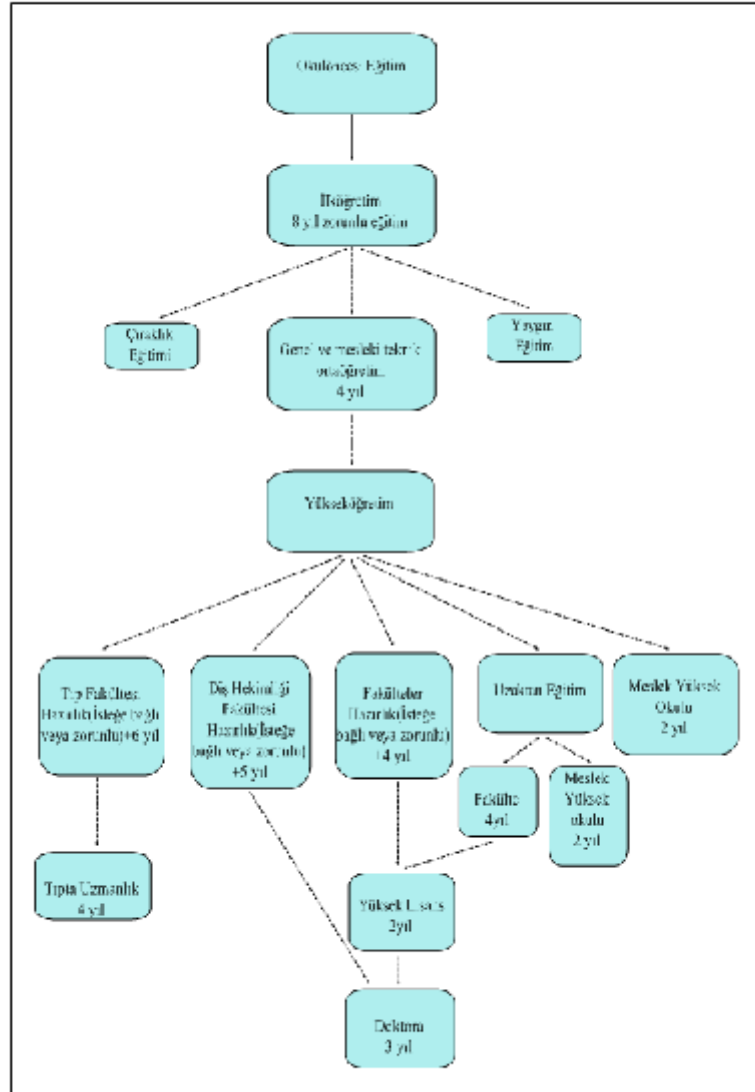
## 2. TÜRK EĞİTİM SİSTEMİNİN YAPISI

Günümüzde eğitim faaliyetleri; örgün ve yaygın olmak üzere iki temel başlık altında, Milli Eğitim Bakanlığı (MEB) bünyesinde yürütülmektedir. Örgün eğitim, belirli yaş grubundaki ve aynı seviyedeki bireyler için, belli bir amaç doğrultusunda ve program dahilinde, bir okul çatısı altında yapılan düzenli eğitimidir. Örgün eğitim kademeleri; okul öncesi eğitim, ilköğretim, ortaöğretim, yükseköğretim şeklindedir.

Yaygın eğitim ise, örgün eğitim sistemine hiç girmemiş, herhangi bir eğitim kademesinde bulunan veya bu kademelerin birinden ayrılmış olan bireylere verilen eğitim sistemidir. Yaygın eğitim; halk eğitimi, çıraklık eğitimi ve uzaktan eğitim şeklinde verilmektedir. Ayrıca, ortaöğretim seviyesinde örgün eğitim veren meslekî ve teknik öğretim kurumlarında açılan meslek kursları ile de yaygın eğitim verilmektedir. (<http://www.meb.gov.tr>, 2011).

Türkiye'deki eğitim sistemi şematik olarak Şekil 1.1'de verilmiştir.

**Şekil 1.1** Türkiye’de uygulanan eğitim sistemine dair sistematik şema



### 3. KÜMELEME ANALİZİ

Canlılar veya nesnelerin gruplandırılması çok eskilerden beri bilim adamlarının ilgilendikleri bir alan olmuştur. Sınıflandırma; insanların yaşadıkları çevreyi veya kainatı daha iyi anlayabilmesine olanak sağlayan önemli bir argümandır.

Verilerin sayısındaki artış, birimlerin sınıflandırılmasını daha da zorlaştırmış ve yeni teknikler bulmayı gerektirmiştir. Bu gereksinim sonucu kümeleme analizi kavramı ortaya çıkmıştır. Günümüzde halen en çok kullanılan yöntemlerden olan kümeleme analizi 1939 yılında ilk kez Tryon tarafından kullanılmıştır. Bu yöntem 1960 yıllarından sonra gelişmiş ve geniş uygulama alanlarına sahip olmuştur. Bu alanda önemli bir katkıyı, 1963 yılında yazdığı “Sayısal Sınıflandırma İlminin Temelleri” adlı kitapla Robert Sokal ve Peter sağlamıştır (Anderberg, 1973: 553) .

Çok değişkenli istatistik analiz yöntemlerinden olan kümeleme analizi, birimleri benzerliklerine göre kümelere ayırmak için kullanılan bir yöntemdir (Tatlıdil, 2002: 329). Kümeleme analizi neticesinde küme içi birimler, kümeler arasındaki birimlere göre karakteristik özellikleri bakımından daha çok benzeşirler. Bunun sonucu olarakta, küme içi homojenlik, kümeler arası heterojenlik sağlanmış olur.

#### 3.1. Benzerlik Ölçütleri

Kümeleme analizinde genel amaç küme içi homojenliği, kümeler arası heterojenliği sağlamaktır. Bu da, karakteristik özellikleri bakımında benzer bireylerin aynı kümede toplanmasıyla sağlanmış olur. Bireylerin benzerlikleri uzaydaki konumları ile ilgilidir. Uzaydaki konumları itibari ile birbirine uzaklıkları daha az olan bireyler aynı kümede toplanmış olacaktırlar. Bireyler gruplanırken birbirine yakınlıkları bazı uzaklık ölçütlerine göre hesaplanır. Veri matrisinde bulunan  $n$  tane birimin,  $p$  değişkene göre uzaklıkları  $D$  matrisi ile ifade edilmektedir.  $D$  matrisinin elemanları ise,  $i$ . birim ile  $j$ . birim arasındaki uzaklıklığı ifade eden  $d_{ij}$  lerdir.  $S$ , birimler arasındaki benzerlik düzeyini gösteren kovaryans matrisidir.

### a) Nicel Veriler İçin Uzaklık Ölçütleri

Nicel veriler için kullanılan başlıca uzaklık ölçütleri aşağıda verilmiştir;

Öklid uzaklığı: 
$$d_{ij} = \left[ \sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{1/2}$$

Ölçekli Öklit Uzaklığı: 
$$d_{ij} = \left[ \sum_{k=1}^p w_k^2 (x_{ik} - x_{jk})^2 \right]^{1/2}$$

City-Block(Manhattan) uzaklığı: 
$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

Minkowski uzaklığı: 
$$d_{ij} = \left[ \sum_{k=1}^p |x_{ik} - x_{jk}|^q \right]^{1/q}$$

Mahalanobis uzaklığı: 
$$Md_{ij} = (x_i - x_j)' S^{-1} (x_i - x_j)$$

Hotelling T<sup>2</sup> Uzaklığı: 
$$T^2 = \frac{n_1 n_2}{n} (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j)$$

### b) Nitel Veriler İçin Kullanılan Uzaklık Ölçütü

Değişkenlerin nicel olmadığı durumlarda, nitel değişkenler için uzaklık ölçütü kullanılır. Nitel değişkenler için uzaklık ölçütü;

$$w_k = \begin{cases} 1 & ; \text{nicel veriler için} \\ \frac{1}{k'} & ; \text{nitel veriler için} \\ da & \end{cases}$$

olmak üzere;

$$d_{ij} = \frac{1}{p} \sum_{k=1}^p w_k |x_{ik} - x_{jk}|$$

şeklindedir.

### 3.2. Kümeleme Yöntemleri

Hangi uzaklık ölçütünün kullanılacağına karar verildikten sonra, ikinci adım kümeleme yönteminin belirlenmesidir. Birimlerin benzerliklerine göre kümelenmesinde kullanılabilecek çeşitli yöntemler vardır. Bu yöntemler hiyerarşik ve hiyerarşik olmayan kümeleme yöntemleri diye iki başlık altında toplanırlar.

#### a) Hiyerarşik Kümeleme Yöntemi

Yöntem, aşama sıralı kümeleme yöntemi olarak da bilinir. Hiyerarşik kümeleme yönteminde, gruplayıcı(agglomerative) ve parçalayıcı(divisive) olmak üzere iki ana teknik vardır. Her birimin başlangıçta bir küme olarak kabul edildiği, daha sonra en yakın iki kümenin yeni bir kümede birleştirilip, küme sayısının bu şekilde her seferinde bir azaltıldığı yönetime gruplayıcı hiyerarşik yöntem denir. Oluşan küme yapıları “dendogram” adı verilen şekilde gösterilebilir(Johnson, Wichern, 2002: 681).

Toplamsal hiyerarşik yöntemdeki sürecin tam tersi olan yönetime de parçalayıcı hiyerarşik yöntem denir. Bu yöntemde başlangıçta tüm birimler bir kümede toplanır. Sonra benzer olmayan birimler başka bir kümeye atanır. Bu yolla küme sayısı her seferinde bir artırılmış olur. Herbir birim bir küme oluşturana kadar bu işlem devam ettirilir.

Gruplayıcı kümeleme yönteminde, başlangıçta N küme mevcut iken sonunda 1 küme elde edilir(parçalayıcı yöntemde tam tersi). Bu yöntem uygun küme sayısının ne olduğu konusunda bilgi vermez.

Uygulamalarda çoğunlukla gruplayıcı hiyerarşik kümeleme yöntemi kullanılmaktadır. Gruplayıcı hiyerarşik yöntemler arasında en çok bilinenleri; tek bağlantı tekniği, tam bağlantı tekniği, grup ortalama tekniği ve ward yöntemidir. Tek bağlantı tekniği, birimler arasındaki en yakın uzaklıklar baz alınarak uygulanan kümeleme tekniğidir. Tam bağlantı tekniğinde maksimum, grup ortalama yönteminde ise ortalama uzaklıklar baz alınmaktadır. Diğerlerinden farklı olarak ward yönteminde ise “Hata kareler toplamı(SS)” prensibinden faydalanılır. Bu yöntemde, hata kareler toplamında minimum artışı sağlayan kümelerin birleştirilmesi amaçlanmaktadır.

### **b) Hiyerarşik Olmayan Kümeleme Yöntemi**

Küme sayısı konusunda ön bilgi varsa veya uygun küme sayısına karar verilmiş ise bu durumda, çok uzun zaman alan hiyerarşik yöntemler yerine hiyerarşik olmayan yöntemler kullanılmaktadır (Anderberg, 1973: 554).

Hiyerarşik olmayan kümeleme yönteminde değişkenlerden ziyade nesnelere gruplandırılır ve bu nesnelere K adet küme ayrılır. Küme sayısı K başlangıçta belirlenebileceği gibi kümeleme sürecine bağlı olarak ortaya çıkabilir (Johnson, Wichern, 2002: 687).

Hiyerarşik olmayan kümeleme yöntemlerinin en sık kullanılanı k-ortalama yöntemidir. Yöntemin işleyiş aşamaları şu şekildedir:

- İlk k birim, k tane kümeye atanır. K tane kümenin herbiri bir elemanlı olup bu elemanlar küme merkezini oluşturmaktadır.
- Geriye kalan birimler en yakın oldukları kümelerle atanırlar. Bütün birimler k tane kümeye atanmış olurlar. Bu atama işlemi gerçekleştikten sonra yeni oluşan elemanlarla küme merkezleri tekrar hesaplanır.
- Hesaplanan küme merkezlerine göre, bir birim en yakın olduğu kümeyle tekrar atanır ve küme merkezleri yeniden hesaplanır.

Kümeler arası atama işlemi sonlanıncaya kadar bu işlem devam eder ve böylece kümeler içi homojenlik, kümeler arası heterojenlik sağlanmış olur.

### **3.3. Küme Sayısının Belirlenmesi**

Kümeleme analizinden sağlıklı bir sonuç alabilmek için küme sayısının ne olacağı önem arz etmektedir. Küme sayısının ne olacağını kesin olarak söyleyen bir yöntemden bahsetmek mümkün olmamakla beraber, bu yöntemlerden bazıları aşağıda verilecektir.

Bu yöntemlerden en çok bilineni;

$$k \cong \left( \frac{n}{2} \right)^{\frac{1}{2}}$$

eşitliğiyle verilmektedir. Küçük örneklemlerle araştırmalarda kullanılması tavsiye edilen bu yöntem, büyük örneklemlerle araştırmalarda sağlıklı sonuçlar vermeyebilmektedir (Everitt, 1974:122).

Rousseuw (1987) tarafından tanımlanan Silhouette indeksi;

$$sil(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

şeklindedir.  $a(i)$ ;  $i$ . birimin kendi kümesindeki tüm noktalara olan ortalama uzaklıklarını,  $b(i)$ ;  $i$ . birimin diğer kümelerdeki tüm noktalara olan ortalama uzaklıkların minimumunu göstermektedir. Uygun küme sayısının belirlenmesi için ortalama Silhouette değeri:

$$sil(C) = \frac{1}{n} \sum_{s_i \in S} sil(i)$$

ifadesiyle verilmekte ve bu indeks değerine göre, maksimum ortalama Silhouette değerine ulaşılan küme sayısı uygun küme sayısı olarak kabul edilmektedir.

Küme sayısına karar vermede kullanılan diğer bir ölçüt Mariott tarafından önerilen ve  $M$  harfiyle gösterilen ölçüttür. Bu ölçüt;

$$M = k^2 |W|$$

eşitliğiyle tanımlanmaktadır(Marriot 1971:501-514). Burada  $W$ , grup içi kareler toplamı matrisidir. Bu ölçüte göre  $M$  değerini minimum yapan  $k$  değeri uygun küme sayısı olarak kabul edilmektedir.

Birim sayısı 30'un üzerinde olduğunda, küme sayısının belirlenmesinde kullanılan ölçütlerden Wilk's Lamda ölçütünün duyarlılığı diğer ölçütlere göre en yüksek düzeye ulaşır(Dinçer ve Özdamar, 1992: 17-33). Wilks olabilirlik oran istatistiği;

$$\Delta = \frac{|W|}{|W + B|} = \frac{|W|}{|T|}$$

şeklinde hesaplanmaktadır.  $T$ : çarpımlar ve kareler toplam matrisi,  $B$ : gruplar arası çarpımlar ve kareler toplam matrisi ve  $W$ : grup içi çarpımlar ve kareler toplam matrisini göstermektedir. Lamda ölçütü sıfır ile bir arasında değer alan bir ölçüttür. Bu ölçütün değeri tam olarak sıfır değeri olmamakla beraber, küme sayısının artırılması Wilk's Lamda değerini sıfıra yaklaştırmaktadır. Uygulamada bu değer 0.01'in altına ilk düştüğü durumdaki küme sayısı analiz için uygun küme sayısı olarak belirlenmiştir.



#### 4. İLLERİN EĞİTİM DÜZEYLERİNİN KÜMELEME ANALİZİ İLE BELİRLENMESİ

Bu analiz için kullanılan veriler, Türkiye İstatistik Kurumu(TÜİK)'nun, adrese dayalı altı yaş üstü nüfus verileri olup, 2010 yılına aittir. Elde edilen verilere göre, analize dahil edilecek değişkenler; okuma yazma bilmeyen, okuma yazma bilen fakat herhangi bir okula gitmeyen, ilköğretim mezunu, ortaöğretim mezunu, yüksekokul mezunu, yüksek lisans mezunu, doktora mezunu ve eğitim düzeyi bilinmeyen oranları şekilde belirlenmiştir. Değişkenler arasındaki ölçek farkını ortadan kaldırmak amacıyla, herhangi bir değişkene ait nüfus sayısı toplam nüfusa oranlanarak herbir değişken standartlaştırılmıştır.

Eğitim düzeyi bilinmeyen oranı: Eğitim durumuna dair nüfusta kaydı bulunmayan kişilerin oranını belirtmektedir.

Aşağıdaki tablolarda değişkenlere ait istatistikler verilmiştir.

**Tablo 4.1** Değişkenlere ait tanımlayıcı istatistikler

	Tanımlayıcı İstatistikler				
	N	Ortalama	Standart Sapma	Minimum	Maksimum
Okuma yazma bilmeyen	81	.06910074	.037398708	.000563	.155187
Okuma yazma bilen fakat bir okul bitirmeyen	81	.20983924	.100421092	.001772	.497212
Sekizyıllık eğitim mezunu	81	.46197386	.096997585	.209425	.679415
Lise veya dengi okul mezunu	81	.16316297	.057213265	.000896	.287691
Yüksekokul veya fakülte mezunu	81	.04999174	.029375410	.000432	.154722
Yüksek lisans mezunu	81	.00331561	.001612424	.000315	.010083
Doktora mezunu	81	.00116533	.000873350	.000101	.005944
Bilinmeyen	81	.04145051	.017310206	.003604	.115131

Türkiye'deki illerin eğitim durumları itibariyle benzerliklerini saptamak için hiyerarşik olmayan kümeleme analizi yöntemlerinden olan k-ortalama yöntemi uygulanmıştır. SPSS 17 paket programı kullanılarak yapılan analiz neticesinde ikiden sekize kadar küme yapıları oluşturulmuştur.

**Tablo 4.2** Kümeler tablosu (K=2)

KÜME	İLLER					
K=1	Adana	Bursa	İzmir	Muğla	İstanbul	Karaman
	Afyon	Çanakkale	Kastamonu	Nevşehir	Tekirdağ	Kırıkkale
	Amasya	Çankırı	Kayseri	Niğde	Tokat	Batman
	Antalya	Çorum	Kırklareli	Ordu	Trabzon	Bartın
	Artvin	Denizli	Kırşehir	Rize	Tunceli	Ardahan
	Aydın	Edirne	Konya	Sakarya	Uşak	Yalova
	Balıkesir	Eskişehir	Kütahya	Samsun	Yozgat	Karabük
	Bilecik	Giresun	Malatya	Sinop	Zonguldak	Kilis
	Bolu	Gümüşhane	Manisa	Isparta	Aksaray	Osmaniye
	Burdur	Hatay	Kahramanmaraş	Bayburt	Düzce	
K=2	Adıyaman	Bitlis	Erzurum	Kars	Siirt	Şırnak
	Ağrı	Diyarbakır	Gaziantep	Kocaeli	Sivas	Iğdır
	Ankara	Elazığ	Hakkari	Mardin	Şanlıurfa	
	Bingöl	Erzincan	Mersin	Muş	Van	

**Tablo 4.3** Kümeler tablosu (K=3)

KÜME	İLLER				
K=1	Amasya	Çanakkale	Kırklareli	Rize	Aksaray
	Artvin	Çankırı	Konya	Sakarya	Karaman
	Aydın	Denizli	Kütahya	Sinop	Batman
	Balıkesir	Edirne	Manisa	Tekirdağ	Bartın
	Bilecik	İzmir	Nevşehir	Uşak	Karabük
	Burdur	Kastamonu	Niğde		
K=2	Adana	Elazığ	Mersin	Ordu	Kırıkkale
	Adıyaman	Erzincan	İstanbul	Samsun	Ardahan
	Afyon	Erzurum	Kars	Sivas	Yalova
	Ankara	Eskişehir	Kayseri	Tokat	Kilis
	Antalya	Gaziantep	Kırşehir	Trabzon	Osmaniye
	Bingöl	Giresun	Kocaeli	Tunceli	Düzce
	Bolu	Gümüşhane	Malatya	Yozgat	
	Bursa	Hatay	Kahramanmaraş	Zonguldak	
Çorum	Isparta	Muş	Bayburt		
K=3	Ağrı	Hakkari	Muş	Şanlıurfa	Şırnak
	Bitlis	Mardin	Siirt	Van	Iğdır
	Diyarbakır				

**Tablo 4.4** Kümeler tablosu (K=4)

KÜME	İLLER			
K=1	Çanakkale	İzmir	Kütahya	Aksaray
	Çankırı	Kastamonu	Rize	Batman
K=2	Amasya	Edirne	Nevşehir	Zonguldak
	Antalya	Gaziantep	Niğde	Bayburt
	Artvin	Gümüşhane	Ordu	Karaman
	Aydın	Hatay	Sakarya	Bartın
	Balıkesir	Isparta	Samsun	Ardahan
	Bilecik	Kayseri	Sinop	Yalova
	Bolu	Kırklareli	Tekirdağ	Karabük
	Burdur	Konya	Tokat	Kilis
	Bursa	Manisa	Uşak	Osmaniye
	Çorum	Kahramanmaraş	Yozgat	Düzce
K=3	Denizli	Muğla		
	Adana	Elazığ	Mersin	Malatya
	Adıyaman	Erzincan	İstanbul	Sivas
	Afyon	Erzurum	Kars	Trabzon
	Ankara	Eskişehir	Kırşehir	Tunceli
K=4	Bingöl	Giresun	Kocaeli	Kırıkkale
	Ağrı	Hakkari	Siirt	Şırnak
	Bitlis	Mardin	Şanlıurfa	İğdır
	Diyarbakır	Muş	Van	

**Tablo 4.5** Kümeler tablosu (K=5)

KÜME	İLLER			
K=1	Ankara	Erzincan	Kocaeli	Sivas
	Elazığ			
K=2	Adana	Gaziantep	Malatya	Yozgat
	Adıyaman	Giresun	Kahramanmaraş	Bayburt
	Afyon	Hatay	Muğla	Kırıkkale
	Bingöl	Mersin	Ordu	Ardahan
	Bolu	İstanbul	Samsun	Yalova
	Çorum	Kars	Tokat	Kilis
	Erzurum	Kayseri	Trabzon	Osmaniye
	Eskişehir	Kırşehir	Tunceli	
K=3	Çanakkale	İzmir	Kütahya	Aksaray
	Çankırı	Kastamonu	Rize	Batman
K=4	Ağrı	Hakkari	Siirt	Şırnak
	Bitlis	Mardin	Şanlıurfa	İğdır
	Diyarbakır	Muş	Van	
K=5	Amasya	Bursa	Manisa	Uşak
	Antalya	Denizli	Nevşehir	Zonguldak
	Artvin	Edirne	Niğde	Karaman
	Aydın	Gümüşhane	Sakarya	Bartın
	Balıkesir	Isparta	Sinop	Karabük
	Bilecik	Kırklareli	Tekirdağ	Düzce
	Burdur	Konya		

**Tablo 4.6** Kümeler tablosu (K=6)

KÜME	İLLER				
K=1	Adana	Çorum	Kars	Ordu	Bayburt
	Adıyaman	Edirne	Kayseri	Sakarya	Karaman
	Afyon	Erzurum	Kırklareli	Samsun	Kırıkkale
	Amasya	Eskişehir	Kırşehir	Tekirdağ	Bartın
	Antalya	Gaziantep	Konya	Tokat	Ardahan
	Aydın	Giresun	Malatya	Trabzon	Yalova
	Balıkesir	Gümüşhane	Manisa	Tunceli	Karabük
	Bilecik	Hatay	Kahramanmaraş	Uşak	Kilis
	Bolu	Isparta	Muğla	Yozgat	Osmaniye
	Burdur	İstanbul	Nevşehir	Zonguldak	Düzce
Bursa					
K=2	Çanakkale	İzmir	Kütahya	Aksaray	Batman
	Çankırı	Kastamonu	Rize		
K=3	Artvin	Denizli	Niğde	Sinop	
K=4	Ağrı	Hakkari	Muş		
K=5	Bingöl	Diyarbakır	Siirt	Van	Iğdır
	Bitlis	Mardin	Şanlıurfa	Şırnak	
K=6	Ankara	Erzincan	Mersin	Kocaeli	Sivas
	Elazığ				

**Tablo 4.7** Kümeler tablosu (K=7)

KÜME	İLLER				
K=1	Bitlis	Mardin	Şanlıurfa	Şırnak	Iğdır
	Diyarbakır	Siirt	Van		
K=2	Çanakkale	İzmir	Kütahya	Aksaray	Batman
	Çankırı	Kastamonu	Rize		
K=3	Artvin	Denizli	Niğde	Sinop	
K=4	Ağrı	Hakkari	Muş		
K=5	Adana	Erzurum	Kayseri	Ordu	Bayburt
	Adıyaman	Gaziantep	Kırşehir	Samsun	Ardahan
	Afyon	Giresun	Malatya	Tokat	Kilis
	Bingöl	Hatay	Kahramanmaraş	Trabzon	Osmaniye
	Çorum	Kars	Tunceli	Yozgat	
K=6	Ankara	Erzincan	Mersin	Kocaeli	Sivas
	Elazığ				
K=7	Amasya	Burdur	İstanbul	Sakarya	Bartın
	Antalya	Bursa	Kırklareli	Tekirdağ	Yalova
	Aydın	Edirne	Konya	Uşak	Karabük
	Balıkesir	Eskişehir	Manisa	Zonguldak	Düzce
	Bilecik	Gümüşhane	Muğla	Karaman	
	Bolu	Isparta	Nevşehir	Kırıkkale	

**Tablo 4.8** Kümeler tablosu (K=8)

KÜME	İLLER				
K=1	Adana	Erzurum	Mersin	Ordu	Bayburt
	Adıyaman	Gaziantep	Kars	Tokat	Ardahan
	Afyon	Giresun	Malatya	Tunceli	Kilis
	Bingöl	Hatay	Kahramanmaraş	Yozgat	Osmaniye
	Çorum				
K=2	Çanakkale	Çankırı	Kütahya	Aksaray	Rize
K=3	Amasya	Bursa	Kırklareli	Sakarya	Karaman
	Antalya	Edirne	Kırşehir	Samsun	Kırıkkale
	Aydın	Eskişehir	Konya	Tekirdağ	Bartın
	Balıkesir	Gümüşhane	Manisa	Trabzon	Yalova
	Bilecik	Isparta	Muğla	Uşak	Karabük
	Bolu	İstanbul	Nevşehir	Zonguldak	Düzce
Burdur	Kayseri				
K=4	Ağrı	Hakkari	Muş		
K=5	İzmir	Kastamonu	Batman		
K=6	Ankara	Elazığ	Erzincan	Kocaeli	Sivas
K=7	Artvin	Denizli	Niğde	Sinop	
K=8	Bitlis	Mardin	Şanlıurfa	Şırnak	Iğdır
	Diyarbakır	Siirt	Van		

Kümeleme analizinde en önemli husus uygun küme sayısının belirlenmesidir. Bölüm 3.3’de anlatılan küme sayısı belirleme yöntemlerinden olan wilk’s lamda ölçütü ve silhouette indeks değerleri, elde edilen farklı sayıdaki kümeler için hesaplanmıştır.

Wilk’s lamda ölçütünün hesaplanması için MANOVA analizi uygulanmış ve aşağıdaki değerler elde edilmiştir;

**Tablo 4.9** Wilk’s Lamda değerleri

KÜME SAYISI	WILK’S LAMDA DEĞERİ
K=2	0.3327
K=3	0.1499
K=4	0.0343
K=5	0.0183
<b>K=6</b>	<b>0.0064</b>
K=7	0.0031
K=8	0.0021

Silhouette indeks değerlerinin hesaplanması için ise MATLAB programı kullanılmış ve şu değerler elde edilmiştir;

**Tablo4.10** Silhouette indeks değerleri

KÜME SAYISI	SİLHOUETTE İNDEKS DEĞERİ
K=2	0.4362
K=3	0.3009
K=4	0.3232
K=5	0.2643
<b>K=6</b>	<b>0.4401</b>
K=7	0.3437
K=8	0.3334

Tablo 4.9’da verilen wilk’s lamda değerlerine göre ve Tablo 4.10’da verilen silhouette indeks değerlerine göre uygun küme sayısının altı olduğu görülmektedir. Bu sebeple, yapılan analiz için küme sayısı altı olarak belirlenmiştir.

Aşağıdaki tabloda, altı kümeli yapının oluşmasında değişkenlerin ayırma gücünü gösteren ANOVA analizi sonuçları verilmiştir.

**Tablo 4.11** Anova analizi sonuçları

	ANOVA					
	Kümeler		Hata		F	P
	Ortalama Kare	Sd	Ortalama Kare	Sd		
Okuma yazma bilmeyen	,009	5	,001	75	10,570	,000
Okuma yazma bilen fakat bir okul bitirmeyen	,144	5	,001	75	125,830	,000
Lise veya dengi okul mezunu	,036	5	,001	75	32,316	,000
Yüksekokul veya fakülte mezunu	,004	5	,001	75	6,313	,000
Yüksek lisans mezunu	,000	5	,000	75	3,441	,007
Doktora mezunu	,000	5	,000	75	3,976	,003
Bilinmeyen	,002	5	,000	75	7,447	,000
Sekizyillik eğitim mezunu	,122	5	,002	75	62,977	,000

Bütün değişkenler için ANOVA istatistik değerleri 0.05’den küçük olduğu görülmektedir. Bu nedenle, bütün değişkenlerin kümeler arasında anlamlı farklılık oluşmasında etkili olduğu söylenebilir.

## 5. SONUÇ

Türkiye coğrafi ve kültürel olarak kozmopolit bir yapıya sahiptir. Kültürel yönleri itibariyle, coğrafi olarak birbirine yakın iller benzer özellikler göstermektedir. Bu çalışmada da eğitim yönleri itibariyle benzer özellikler gösteren illerin hangileri olduğu analiz edilmiştir.

Türkiye'deki illerin eğitim durumlarını gösteren değişkenler belirlenmiş ve bu değişkenlerle kümeleme analizi yapılmıştır. Analiz safhasında hiyerarşik olmayan k-ortalama yöntemi kullanılmış ve uygun küme sayısının altı olduğu belirlenmiştir. Uygun küme sayısı 6 olarak belirlenmiş ve bu altı kümeye göre elde edilen küme yapıları şöyle gerçekleşmiştir:

Birinci küme, Türkiye'nin farklı bölgelerinden olan 51 ilden oluşmaktadır. Bu iller arasında İstanbul, Adana, Antalya, Balıkesir, Bursa, Edirne, Erzurum, Kayseri, Konya, Samsun, Trabzon gibi büyük iller bulunmaktadır. Birinci kümenin oluşumunda öne çıkan değişkenler yüksek lisans mezunu ve doktora mezunu oranları olmuştur. Üniversite bünyesinde verilen lisans ve doktora eğitimi, bu eğitim kurumlarının daha çok olduğu büyük illerde daha yaygın olarak yapılmaktadır. Bu nedenle bu illerin, yüksek lisans mezunu ve doktora mezunu oranları yönüyle benzerlik göstermeleri doğal bir durumdur.

İkinci kümeyi oluşturan iller Çanakkale, Çankırı, İzmir, Kastamonu, Kütahya, Rize, Aksaray, Batman olmuştur. Türkiye'nin coğrafi olarak birbirine uzak olan illerinin bir arada olması dikkat çekicidir. Coğrafi olarak en heterojen yapıdaki kümedir. Bu küme yardımıyla, illerin eğitim düzeyinin belirlenmesinde coğrafi konumun bir sınır olamayacağı söylenebilir. Bu kümeyi oluşturan en önemli etkenler okuma-yazma bilen fakat herhangi bir okula gitmeyen, yüksek lisans mezunu ve doktora mezunu oranları olmuştur. Bu illerin okuma-yazma bilen fakat bir okul bitirmeyen oranları, diğer illere ait oranlardan büyük bir şekilde farklılaşmaktadır.

Üçüncü kümede de, ikinci kümeye benzer bir durum dikkat çekmektedir. Artvin, Niğde, Denizli, Sinop gibi farklı bölgelere ait iller bir kümede toplanmaktadır. Yukarıda ifade ettiğimiz gibi; coğrafi farklılaşmalar, iller arasında, eğitim düzeyleri yönüyle benzerlik göstermesine engel teşkil etmemiştir. Bu kümeyi oluşturan önemli faktörler lise veya dengi okul mezunu, yüksek lisans mezunu ve doktora mezunu oranları olduğu söylenebilir.

Dördüncü ve beşinci kümeler aynı bölgedeki illerden oluşması yönüyle, coğrafi olarak en homojen kümeler oldukları görülmektedir. Güney Doğu Anadolu bölgesindeki iller bu iki kümede gruplanmaktadır. Burada ilginç olan Batman'ın, bu illerden farklılaşarak ikinci kümede yer almasıdır. Bunun nedeni olarak; okuma-yazma bilen fakat bir okul bitirmeyen oranıyla dördüncü kümeden, okuma-yazma bilen fakat bir okul bitirmeyen, sekiz yıllık eğitim mezunu ve lise ve dengi okul mezunu oranları itibarıyla de beşinci kümeden farklılaşmasıdır.

Altıncı kümeyi oluşturan iller ise Ankara, Elazığ, Erzincan, Mersin, Kocaeli, Sivas olmuştur. Ankara diğer büyük illerden farklılık göstermiştir. Farklılaşmanın nedeni olarak okuma- yazma bilen fakat bir okul bitirmeyen, sekiz yıllık eğitim mezunu ve lise ve dengi okul mezunu oranlarının etkili olduğu görülmektedir.

Güney Doğu Anadolu bölgesindeki iller genel olarak, eğitim düzeyleri yönüyle benzerlik göstermektedir. Bu bölgenin, Türkiye'nin en az gelişmiş bölgesi olmasının, eğitim yönüyle de paralellik arz ettiği söylenebilir. Her alanda Avrupa Birliği Standartlarını yakalamayı hedef olarak belirlemiş olan Türkiye bu durumu göz önünde bulundurmalıdır.

#### KAYNAKÇA

Anderberg, M. R., *Cluster Analysis For Applications*, Academic Press, NewYork, 1973.

Everitt, B., *Cluster Analysis*, Heinmann, London 1974.

Eğitim İstatistikleri, <http://www.tuik.gov.tr/VeriBilgi>, (13.9.2011).

Dinçer, K. S., Özdamar, K., "Kümeleme çözümlemesinde uygun kümeleme ölçütlerinin karşılaştırılması", *Hacettepe Fen ve Mühendislik Bilimleri Dergisi*, Sayı 14, Hacettepe Üniversitesi, 1992, sayfa:17-33.

Johnson, R., Wichern, D., *Applied Multivariate Statistics*, 5. Ed., Prentice-Hall, New Jersey 2002.

Marriot, F. H. C., "Practical Problems in a method of Cluster Analysis", *Biometrics*, 27, 1971, p:501-514.

Rousseuw, P. J., "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis", *Journal of Computational and Applied Mathematics*, 20, 1987, p:53-65.

Tatlıdil, H., *Uygulamalı Çok Değişkenli İstatistiksel Analizi*, Ziraat Matbaacılık, Ankara 2002.

Türk Eğitim Sistemi, <http://www.meb.gov.tr>, (2.10.2011).