

## Türkiye'deki İllerin Yaşam Endekslerine Göre Kümeleneşmesi

Hasan BULUT

Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55200, Samsun  
(ORCID: <https://orcid.org/0000-0002-6924-9651>)

(Alınış / Received: 16.07.2018, Kabul / Accepted: 29.01.2019, Online Yayınlanma / Published Online: 19.04.2019)

### Anahtar Kelimeler

Yaşam memnuniyeti,  
Kümeleme analizi,  
Küme geçerlilik indeksleri,  
EM kümeleme,  
K ortalamalar

**Özet:** Kümeleme analizi çok değişkenli veri yapısına sahip gözlemleri benzerlik ya da farklılık ölçütlerine bakarak sınıflamayı amaçlayan yöntemler topluluğu olarak tanımlanabilir. Kümeleme analizinde elde edilen kümelerin kendi içinde homojen, kümeler arasının ise heterojen olması istenir. Bu çalışma, ülkemizdeki illerin yaşam memnuniyet endeks değerlerine göre kümeleneşmesini amaçlamaktadır. Çalışmada kullanılan endeks değerleri Türkiye İstatistik Kurumu (TÜİK) tarafından 2015 yılında yürütölen ve sonuçları 2016 yılında açıklanan "İllerde Yaşam Endeksi" çalışmasından elde edilmiştir. Çalışmada 11 endeks değeri illere ait memnuniyet göstergeleri olarak alınmıştır ve elde edilen çok değişkenli veri yapısı kullanılarak iller kümeleneştir. Ayrıca çalışmada küme sayısının ve kümeleme yönteminin belirlenmesinde küme geçerlilik indekslerinden yararlanılmıştır.

## The Clustering of Cities in Turkey According to Indexes of Life Satisfaction

### Keywords

Life satisfaction,  
Cluster analysis,  
Cluster validation index,  
EM clustering,  
K means clustering

**Abstract:** Cluster analysis can be defined as the group of methods which aim to classify multivariate observations by being used similarity/dissimilarity measures between observations. In cluster analysis, it is wanted that within clusters are homogeneous and between clusters are heterogeneous. In this study, clustering of cities in Turkey is aimed according to life satisfaction index values. The index values used in study are obtained from results of "Life Satisfaction in Cities" study which are performed by TÜİK in 2015 and declared in 2016. In study, 11 index values are taken as satisfaction indicators of cities and the cities are clustered by using obtained multivariate data. Moreover, cluster validation indexes are used to determine cluster number and cluster method.

### 1. Giriş

Türkiye İstatistik Kurumu (TÜİK) tarafından 2015 yılında yürütölen "İllerde Yaşam Endeksi" çalışmasının sonuçları 2016 yılında kamuoyu ile paylaşılmıştır. İllerde yaşam endeksi çalışması; konut, çalışma hayatı, gelir ve servet, sağlık, eğitim, çevre, güvenlik, sivil katılım, altyapı hizmetlerine erişim, sosyal yaşam ve yaşam memnuniyeti olmak üzere illerdeki yaşamı 11 farklı endeks değeri ile ele almaktadır. Endeksler 0 ile 1 arasında değer almakta ve 1'e yaklaştıkça daha iyi bir yaşam düzeyini ifade etmektedir [1].

TÜİK yaptığı çalışmada 41 göstergelyi kullanarak 11 endeks değerini, bu 11 endeks değerini kullanarak da tek bir "İllerde Yaşam Endeksi" değerini bileşik endeks yaklaşımı ile hesaplamıştır. Bileşik endekslerin hesaplanmasında da Min-Max yönteminden yararlanmıştır. Min-Max yöntemi ile

endeks değerleri 0-1 aralığında normalize edilmiştir [1].

TÜİK'in yapmış olduđu bu çalışmadan sonra literatürde konu ile ilgili farklı yaklaşımlar kullanılarak çalışmalar yapılmıştır. Aydın ve Ertürk [2] yaptıkları çalışmalarında, TÜİK'in il bazında yapmış olduđu çalışmaya dayanarak, İstatistiki Bölge Birimleri Sınıflandırması (İBBS) Düzey-2'de yer alan 26 İstatistiki bölge birimlerini yaşam memnuniyet endekslerine göre karşılaştırmışlardır. Aydın ve Yalçın [3] ise Güneydođu Anadolu Bölgesi'nde yer alan 9 ili yaşam memnuniyetlerine göre çok boyutlu ölçekleme tekniğini kullanarak incelemişlerdir. Alpaykut [4] ülkemizdeki illerin yaşam memnuniyetini temel bileşenler analizi kullanarak elde ettiđi temel bileşenlere göre TOPSİS yöntemi ile sıralamış ve bu sonuçları TÜİK'in yapmış olduđu memnuniyet sıralaması ve Sosyo-Ekonomik

Gelişmişlik Endeksi (SEGE)-2011 sıralaması ile karşılaştırmıştır.

Ayrıca ülkemizde illerin kümelmesini konu alan çalışmalar sıklıkla gerçekleştirilmektedir. Buna göre Atalay ve Tortum [5] 1997-2006 yılları arasında şehir dışında yaşanan trafik kazalarındaki ölüm ve yaralanma oranlarına göre illeri k-ortalamlar ve bulanık c-ortalamlar algoritmaları ile 5 kümeye ayırmışlardır. Çelik [6] ülkemizdeki illeri 2010 yılına ait 10 sağlık göstergesine göre 7,10 ve 15 kümeye ayırmıştır. Benzer bir çalışmada Tekin [7] illeri 2013 yılına ait 16 sağlık göstergesi bakımından Ward kümeleme algoritması ile 5, 7, 11, 13 ve 21 kümeye ayırmıştır. Kandemir [8] ise ülkemizdeki illeri bulanık kümeleme yöntemiyle 2016 yılına ait konaklama istatistiklerini dikkate alarak kümelemiştir.

Bu çalışmanın temel amacı ise ülkemizde bulunan illeri, 11 farklı endeks bakımından almış oldukları değerlerden oluşan çok değişkenli veri yapısına dayanarak kümelemektir. Kümeleme analizinde en önemli sorunlardan birisi olan küme sayısı belirleme sırasında küme geçerlilik endekslerinden yararlanılmıştır. Ayrıca çalışmada çok değişkenli gözlemlerin kümelmesi amacıyla hiyerarşik olmayan kümeleme algoritmalarından birisi olan k-ortalamlar yöntemi ve karma çok değişkenli normal modele dayanan EM kümeleme algoritmasından yararlanılmıştır.

## 2. Materyal ve Metot

Literatürde çok sayıda kümeleme algoritması mevcuttur. Bu çalışmada modele dayalı algoritmalar EM algoritması ile en popüler kümeleme yöntemi olan k-ortalamlar algoritması kullanılmıştır. Bu bölümde söz konusu algoritmalar tanıtılmıştır. Ayrıca kullanılacak kümeleme yöntemi ve küme sayısına karar vermek için kullanılacak küme geçerlilik endeksleri de bu bölümde tanıtılmıştır.

### 2.1. K ortalamlar kümeleme algoritması

En sık kullanılan hiyerarşik olmayan kümeleme yöntemi k-ortalamlar algoritmasıdır. Bu algoritmada küme sayısının önceden bilinmesi gerekir. Algoritmanın amacı, başlangıç çekirdeklerine bağlı olarak elde edilen kümelerin içi homojen ve kümeler arası heterojen olacak şekilde bir kümeleme yapısını elde etmektir. Algoritma adımları genel olarak aşağıdaki gibi tanımlanabilir.

- i. k tane küme çekirdeği belirlenir.
- ii. Gözlemler kendisine en yakın olan çekirdeğin olduğu kümeye atanır.
- iii. Kümeye atanmış elemanların ortalama vektörü hesaplanarak küme çekirdekleri güncellenir. Eğer bir gözlem için kendi bulunduğu kümenin çekirdeğinden daha yakın bir çekirdek mevcut ise, gözlem yakın olan kümeye aktarılır.

iv. Tüm geçişler durana kadar Adım (iii) tekrarlanır [9].

Algoritmanın sonuçları küme sayısına ve başlangıç küme çekirdeklerine bağlıdır. Küme sayısının seçimi için Bölüm 2.3'de verilen küme geçerlilik endekslerinden yararlanılabilir.

Başlangıç küme çekirdeklerinin belirlenmesi için ise,

- k çekirdeğin rastgele belirlenmesi
- İlk k gözlemin çekirdek olarak alınması
- Karşılıklı olarak en uzak olan k gözlemin çekirdek olarak alınması
- Öncelikle hiyerarşik kümeleme algoritmalarından biri kullanılarak, elde edilen küme merkezlerinin k çekirdek olarak alınması. yaklaşımları önerilmiştir [9,10].

### 2.2. EM kümeleme algoritması

Daha önce tanıtılan k ortalamlar algoritması hiyerarşik olmayan bir algoritmadır. Bu bölümde modele dayalı bir kümeleme algoritması olan EM (Expectation-Maximization) algoritması tanıtılacaktır. Modele dayalı yaklaşım verinin karma bir dağılımdan (genellikle karma çok değişkenli normal dağılımdan) geldiğini varsayar. Bir başka deyişle, her küme farklı bir çok değişkenli dağılıma sahip olarak bir araya gelmiş ve böylece karma dağılımlı bir veri seti elde edilmiş olur. EM algoritması bu karma dağılım modelleri için sıklıkla kullanılmaktadır [10,11].

$X_1, X_2, \dots, X_n$  karma çok değişkenli normal modelden çekilmiş n birimlik bir örneklem olsun. Bu örneklem çekildiği karma çok değişkenli normal dağılım modeli:

$$f(\mathbf{x}; \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{g=1}^k \pi_g f(\mathbf{x}; \boldsymbol{\theta}_g) \quad (1)$$

şeklinde tanımlanır. Burada  $\pi_g > 0$  karışım parametresi olarak adlandırılır ve  $g.nci$  kümeye ait gözlemlerin tüm gözlemler içindeki oranını ifade eder. Ayrıca  $\boldsymbol{\theta}_g$   $g.nci$  kümeye ait çok değişkenli normal dağılımın parametre vektörüdür. EM kümeleme yaklaşımında  $(X_1, X_2, \dots, X_n)$  verisi tamamlanmamış veri olarak adlandırılır. Bu nedenle veriye etiket değişkeni adı verilen bir  $Z$  değişkeni ilave edilir.

Bu etiket değişkeni  $Z = \{Z_1, Z_2, \dots, Z_n\}$  ve  $Z'_i = [z_{1i}, z_{2i}, \dots, z_{ki}]$  şeklinde tanımlanır.  $z_{gi} = 1$  olması,  $i.nci$  gözlemin  $g.nci$  kümede olduğu anlamına gelir ve aksi halde  $z_{gi} = 0$  olacaktır.  $X_i$  gözlem vektörlerinden oluşan veriye etiket değişkeni de eklenerek tamamlanmış veri seti elde edilir  $\{X_1, X_2, \dots, X_n, Z_1, Z_2, \dots, Z_n\}$  ve bu tamamlanmış verinin olabirlik fonksiyonu:

$$L = f(x_1, x_2, \dots, x_n, z_1, z_2, \dots, z_n; \pi, \theta) \\ = \prod_{i=1}^n \prod_{g=1}^k [\pi_g f(x; \theta_g)]^{z_{gi}} \quad (2)$$

şeklinde tanımlanır. Böylece log olabilirlik fonksiyonu da Denklem 3.'de verildiği gibi yazılabilir:

$$\ln L(\pi, \theta; x_1, x_2, \dots, x_n, z_1, z_2, \dots, z_n) \\ = \sum_{i=1}^n \sum_{g=1}^k z_{gi} \ln[\pi_g f(x; \theta_g)] \quad (3)$$

EM algoritmasında amaç;  $\ln L$  log olabilirlik fonksiyonunun değerini maksimum yapan  $z_{gi}$  değerlerini bulmaktır. Bu  $z_{gi}$  değerleri gözlemlerin hangi kümede olduğunu ifade ettiği için, bu değerlerin bulunması ile gözlemler de kümeleneceklerdir. EM algoritması E (Expectation) ve M (Maximization) adımlarından oluşan iteratif bir algoritmadır.

**E Adımı:**  $z_{gi}$  değerleri bilinmediğinden dolayı bunun yerine  $z_{gi}$ 'nin Bayes teoremine göre Denklem 4'deki gibi hesaplanan şartlı beklenen değeri  $E(z_{gi} | x; \pi, \theta)$  kullanılır.

$$\hat{z}_{gi} = E(z_{gi} | x; \pi, \theta) = \frac{\pi_g f(x; \theta_g)}{\sum_{g=1}^k \pi_g f(x; \theta_g)} \quad (4)$$

Algoritmanın bu adımında Denklem 4.'de verilen şartlı beklenen değer bulunur.

**M Adımı:** Bu adımda Denklem 5. ile verilen  $\ln L$  log-olabilirlik fonksiyonu;

$$\widetilde{\ln L}(\pi, \theta; x_1, x_2, \dots, x_n, z_1, z_2, \dots, z_n) \\ = \sum_{i=1}^n \sum_{g=1}^k \hat{z}_{gi} \ln[\pi_g f(x; \theta_g)] \quad (5)$$

$\sum_{g=1}^k \pi_g = 1$  kısıtı altında maksimize edilerek parametreler güncellenir. EM yaklaşımı iteratif bir algoritmadır ve log-olabilirlik değeri durağanlaşana kadar devam eder [10,11].

### 2.3. Küme geçerlilik endeksleri

Küme sayısı ve kümeleme yönteminin belirlenmesi amacıyla literatürde önerilmiş olan çok sayıda küme geçerlilik endeksi vardır. Bu endekslerden bazıları bu alt bölümde tanıtılmıştır.

**i. Bağlanırlık endeksi (connectivity index):** Bu endeks değeri bir gözleme en yakın olan gözlemlerin başka bir kümede olduğu durumda büyüdüğü bir yaklaşım sunar.  $nn_{i(j)}$ ,  $i_{nci}$  gözleme  $j_{nci}$  en yakın komşu olarak tanımlansın. Bağlanırlık (Connectivity) endeksi Denklem 6.'da verildiği gibi tanımlanır.

$$Conn(C) = \sum_{i=1}^n \sum_{j=1}^L x_{i,nn_{i(j)}} \quad (6)$$

Burada L kaç komşuluğa bakılacağına karar veren bir parametre değeri ve  $x_{i,nn_{i(j)}}$

$$x_{i,nn_{i(j)}} = \begin{cases} 0 & , \quad i \text{ ve } nn_{i(j)} \text{ aynı kümede} \\ 1/j & , \quad i \text{ ve } nn_{i(j)} \text{ farklı kümede} \end{cases} \quad (7)$$

olup, her bir gözlem için hesaplanır. Buna göre bağlanırlık endeksi sıfırdan sonsuza kadar değer alır ve en küçük olması istenir [12,13].

**ii. Silüet genişlik endeksi (silhouette width):** Silüet genişliği her bir gözlemin silüet değerinin ortalamasıdır.  $i_{nci}$  gözlemin silüet değeri

$$S(i) = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (8)$$

şeklinde hesaplanır. Burada  $a_i$ ,  $i_{nci}$  gözlem ile aynı kümede olan gözlemler arasındaki uzaklıklar ortalamasıdır. Daha sonra eldeki tüm kümeler için  $d(x_i, C_j)$  ( $j = 1, 2, \dots, k$ ) uzaklıkları hesaplanır. Bu uzaklıklardan en küçüğü  $b_i$  olarak alınır. Silüet genişliği  $[-1, 1]$  aralığında değerler alır ve maksimum olması istenir [13,14].

**iii. Dunn endeksi:** Dunn endeksi gözlemlerin diğer kümelere bulunan gözlemler ile en küçük uzaklığının küme içinde bulunan gözlemlerin aralarındaki en büyük uzaklığa oranlanması ile bulunur ve Denklem 9.'daki gibi hesaplanır.

$$D(C) = \frac{\min_{x_i \in C_{m_1}, x_j \in C_{m_2}} d(x_i, x_j)}{\max_{x_i, x_j \in C_m} d(x_i, x_j)}, \quad m_1 \neq m_2 = 1, 2, \dots, k \quad (9)$$

Dunn endeksi de tıpkı Bağlanırlık endeksi gibi sıfırdan sonsuza kadar değerler alabilmektedir. Ancak Bağlanırlık endeksinin aksine, Dunn endeksinin büyük olması elde edilen kümeleme sonucunun daha iyi olduğunu göstermektedir [12,13].

### 3. Bulgular

Bu çalışmada kullanılan veri seti 81 il için TÜİK tarafından yayınlanan 11 endeks değerinden oluşmaktadır ve bu endeks değerlerine göre illerin kümeleneceği amaçlanmıştır. Bu endeks değerleri Tablo 1.'de verilmiş olup, ilgili veri seti Türkiye İstatistik Kurumu'nun web sayfasından alınmıştır [1].

Kümeleme analizinde modele dayalı EM algoritması ve hiyerarşik olmayan algoritmalar k-ortalama yöntemini kullandığından dolayı hangi algoritma sonucunun daha geçerli olacağı ve kullanılacak küme

sayısının değeri, Bölüm 2.3'te tanımlanan küme geçerlilik endekslerine dayanarak belirlenmiştir. Farklı kümeleme algoritması ve küme sayısı için küme geçerlilik endeks değerleri Tablo 2'de verilmiştir.

Tablo 2.'ye göre Silüet ve Bağlılık endeks değerleri 2 küme ile EM algoritmasının ve Dunn endeksi de 5 küme ile k-ortalamlar yönteminin kullanılmasını önermişlerdir. Dolayısıyla bu çalışmada iki farklı kümeleme yapısı elde edilmiş ve yorumlanmıştır. Bu değerlere dayanarak çizilen ve Şekil 1'de verilen grafikler yardımıyla da en uygun yöntem ve küme sayısı rahatlıkla belirlenebilecektir.

Küme sayısını 2 olarak EM algoritması ile illerin endekslerine göre kümelene sonuçları Tablo 3'de verilmiştir. Tablo 3'e göre ülkemizdeki iller yaşam kalitesini belirleyen 11 endeks yönünden iki gruba

ayrılmaktadır. Küme 2'de bulunan iller genellikle ülkemizin Doğu ve Güneydoğu bölgelerinde yer almaktadır. Kümelere ait memnuniyet endeks değerlerinin ortalamaları ise Tablo 5'te verilmiştir.

**Tablo 1.** Çalışmada kullanılan Endeksler.

Kod	Endeks Adı
$X_1$	Konut Endeksi
$X_2$	Çalışma Hayatı Endeksi
$X_3$	Gelir ve Servet Endeksi
$X_4$	Sağlık Endeksi
$X_5$	Eğitim Endeksi
$X_6$	Çevre Endeksi
$X_7$	Güvenlik Endeksi
$X_8$	Sivil Katılım Endeksi
$X_9$	Altyapı Hizmetlerine Erişim Endeksi
$X_{10}$	Sosyal Yaşam Endeksi
$X_{11}$	Yaşam Memnuniyeti Endeksi

**Tablo 2.** Küme Geçerlilik endeksleri ile kümeleme yöntemi ve küme sayısına karar verilmesi

Küme Sayısı	Silüet Endeksi		Bağlılık Endeksi		Dunn Endeksi	
	k Ort.	EM	k Ort.	EM	k Ort.	EM
k=2	0.436	0.443*	29.59	26.94*	0.222	0.168
k=3	0.257	0.252	79.79	86.07	0.194	0.135
k=4	0.235	0.237	95.47	94.05	0.230	0.250
k=5	0.228	0.230	115.43	112.78	0.260*	0.205
k=6	0.189	0.215	137.38	136.10	0.201	0.182
k=7	0.201	0.206	141.75	141.51	0.182	0.218
k=8	0.190	0.200	152.20	151.26	0.224	0.218
k=9	0.187	0.189	163.02	156.65	0.182	0.235
k=10	0.203	0.186	163.81	166.27	0.195	0.235

\*İlgili kritere göre Optimal küme sayısı ve kümeleme yöntemi

**Tablo 3.** EM algoritmasına göre illerin kümelenmesi

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

Ayrıca illerin kümelenme yapısı Türkiye Siyasi Haritası üzerinde Şekil 2'deki gibi verilmiştir. Bu harita ile ülkemizdeki illerin memnuniyet ya da memnuniyetsizlik durumunun bölgesel bir soruna da işaret ettiği rahatlıkla görülmektedir.

Küme 1 ve Küme 2'de bulunan illerin endeksler bakımından aldıkları değerlerin ortalamaları arasında fark olup-olmadığı da incelenmiştir.

Öncelikle kümeler için her bir endeks bakımından normallik testleri gerçekleştirilmiş ve sonuçlar Tablo 4'te verilmiştir. Normallik varsayımını sağlayan ( $X_2, X_4, X_5, X_6, X_{10}$  ve  $X_{11}$ ) endeksler için t testi, varsayımı sağlamayan endeksler için ise Mann-Whitney testi kullanılmış ve elde edilen p değerleri Tablo 5'te ilave bir sütun olarak EM algoritmasına ait sonuçlara

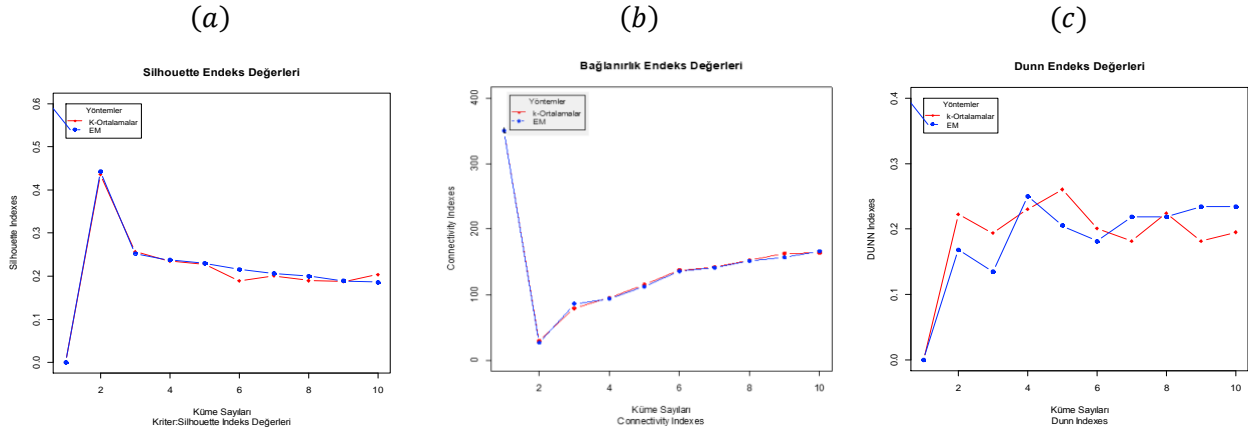
eklenmiştir. Bu sonuçlara bakıldığında Küme 1 ve Küme 2 arasında sadece Güvenlik Endeksi ( $X_7$ ) ve Yaşam Memnuniyeti Endeksi ( $X_{11}$ ) bakımından istatistiksel olarak anlamlı bir farklılık olmadığı görülmektedir. Küme 2'de yer alan illerin ülkemizin Doğu ve Güneydoğu bölgesinde yer alması ve bölgede yaşanan terör olayları düşünüldüğünde elde edilen bu sonuç oldukça çarpıcıdır. Diğer tüm endeksler yönünden Küme 1, Küme 2'ye göre daha yüksek değere sahiptir.

Küme sayısını 5 alarak k ortalamalar algoritması ile illerin endekslere göre kümelenme sonuçları Tablo 6'da verilmiştir. Tablo 6'ya göre ülkemizdeki iller yaşam memnuniyetini ölçen 11 endeks yönünden beş gruba ayrılmaktadır. Bu gruplarda, bir başka deyişle kümelerde, yer alan iller Türkiye siyasi haritası üzerinde Şekil 3'deki gibi gösterilebilir.

**Tablo 4.** Normallik testlerine ait sonuçlar

Endeks	EM Kümeleme Sonucu					k-ortalamlar Kümeleme Sonucu										
	Küme1		Küme2		Test İstatistiği	p- değeri	Küme1		Küme2		Küme3		Küme4		Küme5	
	Test İstatistiği	p- değeri	Test İstatistiği	p- değeri			Test İstatistiği	p- değeri	Test İstatistiği	p- değeri	Test İstatistiği	p- değeri	Test İstatistiği	p- değeri		
$X_1$	0.95	0.02	0.97	0.86	0.92	0.02	0.96	0.73	0.96	0.67	0.95	0.66	0.98	0.93		
$X_2$	0.98	0.29	0.96	0.67	0.97	0.45	0.97	0.90	0.95	0.49	0.92	0.34	0.94	0.62		
$X_3$	0.95	0.02	0.98	0.96	0.96	0.27	0.97	0.87	0.84	0.01	0.89	0.17	0.90	0.39		
$X_4$	0.98	0.32	0.96	0.51	0.97	0.54	0.85	0.03	0.95	0.36	0.88	0.12	0.96	0.85		
$X_5$	0.98	0.38	0.99	0.99	0.98	0.85	0.95	0.65	0.94	0.26	0.98	0.97	0.94	0.70		
$X_6$	0.98	0.41	0.95	0.37	0.96	0.28	0.88	0.06	0.98	0.92	0.88	0.13	0.92	0.49		
$X_7$	0.96	0.02	0.98	0.93	0.92	0.02	0.92	0.29	0.98	0.95	0.95	0.62	0.87	0.22		
$X_8$	0.92	0.00	0.95	0.48	0.89	0.00	0.90	0.13	0.92	0.15	0.91	0.28	0.85	0.16		
$X_9$	0.94	0.00	0.98	0.93	0.97	0.38	0.94	0.45	0.90	0.06	0.94	0.58	0.95	0.70		
$X_{10}$	0.98	0.37	0.92	0.14	0.95	0.13	0.93	0.32	0.98	0.95	0.80	0.01	0.88	0.27		
$X_{11}$	0.98	0.57	0.95	0.39	0.95	0.09	0.93	0.34	0.97	0.80	0.94	0.52	0.86	0.19		

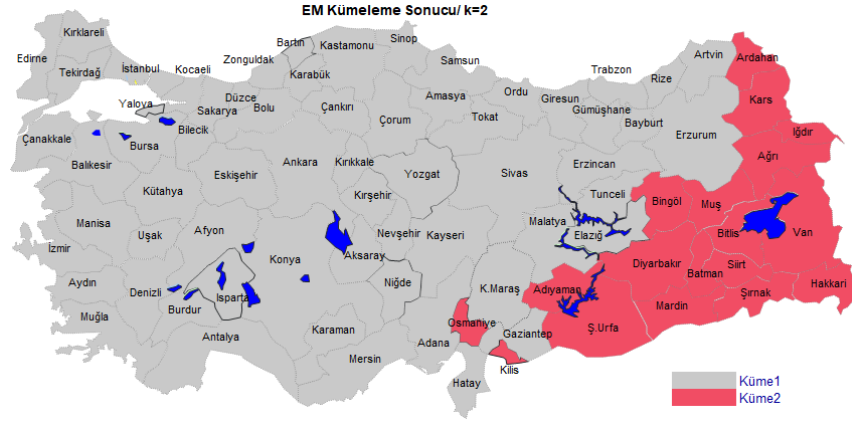
Normallik testi olarak Shapiro-Wilk Testi kullanılmıştır.



**Şekil 1.** Endeks değerlerini kullanarak uygun kümeleme yöntemi ve küme sayısının belirlenmesi (a) Silüet endeks değerleri (b) Bağlılık endeks değerleri (c) Dunn endeks değerleri

**Tablo 5.** İki farklı kümeleme algoritmasından elde edilen kümeler için ortalama endeks değerleri

Endeks	EM Sonucu			k Ortalamalar Sonucu					p değeri
	Küme1	Küme2	p değeri	Küme1	Küme2	Küme3	Küme4	Küme5	
x1	0.769	0.354	<0.001	0.373 <sup>a</sup>	0.767 <sup>b</sup>	0.643 <sup>c</sup>	0.812 <sup>d</sup>	0.319 <sup>a</sup>	<0.001
x2	0.592	0.362	<0.001	0.253 <sup>a</sup>	0.601 <sup>b</sup>	0.481 <sup>c</sup>	0.620 <sup>b</sup>	0.379 <sup>d</sup>	<0.001
x3	0.471	0.161	<0.001	0.196 <sup>a</sup>	0.449 <sup>b</sup>	0.316 <sup>c</sup>	0.545 <sup>d</sup>	0.141 <sup>a</sup>	<0.001
x4	0.631	0.466	<0.001	0.432 <sup>a</sup>	0.640 <sup>b</sup>	0.543 <sup>c</sup>	0.651 <sup>b</sup>	0.470 <sup>a</sup>	<0.001
x5	0.589	0.334	<0.001	0.224 <sup>a</sup>	0.587 <sup>b</sup>	0.536 <sup>b</sup>	0.606 <sup>b</sup>	0.339 <sup>c</sup>	<0.001
x6	0.635	0.436	<0.001	0.409 <sup>a</sup>	0.620 <sup>b</sup>	0.603 <sup>b</sup>	0.657 <sup>b</sup>	0.426 <sup>a</sup>	<0.001
x7	0.622	0.600	0.125>0.05	0.550 <sup>a</sup>	0.675 <sup>b</sup>	0.571 <sup>a</sup>	0.594 <sup>ab</sup>	0.610 <sup>ab</sup>	0.002<0.01
x8	0.451	0.309	<0.001	0.247 <sup>a</sup>	0.452 <sup>b</sup>	0.396 <sup>b</sup>	0.473 <sup>b</sup>	0.304 <sup>a</sup>	<0.001
x9	0.487	0.261	<0.001	0.154 <sup>a</sup>	0.439 <sup>b</sup>	0.416 <sup>b</sup>	0.551 <sup>c</sup>	0.280 <sup>d</sup>	<0.001
x10	0.458	0.303	<0.001	0.294 <sup>a</sup>	0.479 <sup>b</sup>	0.358 <sup>a</sup>	0.473 <sup>b</sup>	0.299 <sup>a</sup>	<0.001
x11	0.547	0.503	0.193>0.05	0.819 <sup>a</sup>	0.722 <sup>a</sup>	0.324 <sup>b</sup>	0.459 <sup>b</sup>	0.461 <sup>b</sup>	<0.001



Şekil 2. Ülkemizdeki illerin memnuniyet endekslerine göre EM algoritması ile iki kümeye ayrılması



Şekil 3. Ülkemizdeki illerin memnuniyet endekslerine göre k ortalamlar algoritması ile beş kümeye ayrılması

Tablo 6. k-ortalamlar algoritmasına göre illerin kümelenmesi

Küme1	Hakkâri	Siirt	Şırnak
Küme2	Afyonkarahisar	Aksaray	Amasya
	Artvin	Balıkesir	Bayburt
	Bolu	Çanakkale	Çankırı
	Düzce	Erzincan	Erzurum
	Giresun	Gümüşhane	Isparta
	Kahramanmaraş	Kırıkkale	Kırşehir
	Konya	Kütahya	Manisa
	Nevşehir	Niğde	Sakarya
	Sinop	Uşak	
	Küme3	Adana	Bingöl
Elazığ		Gaziantep	Hatay
Malatya		Mersin	Ordu
Osmaniye		Tunceli	Yozgat
Küme4	Ankara	Antalya	Aydın
	Bartın	Bilecik	Burdur
	Bursa	Denizli	Edirne
	Eskişehir	İstanbul	İzmir
	Karabük	Karaman	Kastamonu
	Kayseri	Kırklareli	Kocaeli
	Muğla	Rize	Samsun
	Sivas	Tekirdağ	Tokat
	Trabzon	Yalova	Zonguldak
	Küme5	Adıyaman	Ağrı
Batman		Bitlis	Diyarbakır
Iğdır		Kars	Kilis
Mardin		Muş	Şanlıurfa
Van			

Bu kümelerin endeks ortalamaları da yine Tablo 5'te verilmiştir. Bu grupların endeks değerlerinin ortalamaları yönünden aralarında fark olup-olmadığı incelenmiştir. Varsayımları sağlayan endeksler ( $X_1, X_2, X_5, X_9, X_{10}$  ve  $X_{11}$ ) için tek yönlü varyans analizi (ANOVA), diğer endeksler için ise Kruskal-Wallis H testi kullanılmış ve elde edilen p değerleri yine Tablo 5'te ilave bir sütun olarak k ortalamalar algoritmasına ait sonuçlara eklenmiştir.

Sonuçlar incelendiğinde Güvenlik endeksi dışındaki tüm endeksler için %99.9 güvenle, Güvenlik endeksi için ise %99 güven düzeyinde kümeler arasında istatistiksel olarak anlamlı bir farklılık vardır. Bu farklılıklara ilişkin çoklu karşılaştırma sonuçları da Tablo 5'te verilmiştir.

Tablo 5'te k ortalamalar algoritmasına ait sonuçlarda ilgili endeks ortalamasının üzerinde bulunan harf kodları incelendiğinde, harf kodunda aynı/ortak harfe sahip olan kümeler arasında istatistiksel olarak anlamlı bir farklılık yok iken, farklı harfe sahip kümeler arasında ilgili endeks ortalamaları bakımından istatistiksel olarak anlamlı bir farklılık olduğu söylenebilir.

Konut endeksi ( $X_1$ ) bakımından Küme 1 ve Küme 5 arasında anlamlı bir farklılık yoktur. Bu kümelere yer alan iller ülkemizin ağırlıklı olarak Doğu ve Güneydoğu bölgelerinde yer almaktadır. Konut endeksi bakımından İstanbul, Ankara, İzmir, Antalya, Bursa, Samsun gibi illerin yer aldığı Küme 4 en yüksek değere sahiptir. Küme 4'ü konut endeksi bakımından Küme 2 ve Küme 3 takip etmektedir.

Çalışma hayatı ( $X_2$ ) endeksi bakımından Küme 2 ve Küme 4 arasında istatistiksel olarak anlamlı bir farklılık yoktur. Diğer tüm kümeler çalışma hayatı endeksi bakımından anlamlı bir farklılığa sahiptir. Bu durum ülkemizde istihdam noktasında bölgeler/kümeler arası bir eşitsizliği gözler önüne sermektedir. Küme 4 ve Küme 2'yi sırasıyla Küme 3, Küme 5 ve Küme 1 izlemektedir.

Gelir ve servet endeksi ( $X_3$ ) bakımından Küme 1 ve Küme 5 arasında anlamlı bir fark olmayıp, geri kalan tüm kümeler arasında anlamlı bir farklılık vardır. Bu endeks bakımından en iyi durumda olan Küme 4'dür. Bu kümeyi sırasıyla Küme 2 ve Küme 3 takip etmektedir. En kötü durumda olan Küme 1 ve Küme 5 arasında ise anlamlı bir farklılık yoktur.

Sağlık endeksi ( $X_4$ ) bakımından en iyi durumdaki kümeler Küme 4 ve Küme 2 olup, aralarında istatistiksel olarak anlamlı bir farklılık yoktur. Bu kümelere sonra en yüksek endeks ortalaması Küme 3'e ait olup, en düşük endeks ortalamasına sahip olan Küme 1 ve Küme 5 arasında da istatistiksel olarak anlamlı bir farklılık yoktur.

Eğitim endeksi ( $X_5$ ) bakımından Küme 2, Küme 3 ve Küme 4 arasında anlamlı bir fark yoktur. Bu kümeleri

sırasıyla Küme 5 ve en düşük endeks değerine sahip olan Küme 1 takip etmektedir.

Çevre endeksi ( $X_6$ ) bakımından Küme 2, Küme 3 ve Küme 4 arasında anlamlı bir fark yokken, Küme 1 ve Küme 5 arasında da aynı şekilde istatistiksel olarak anlamlı bir farklılık yoktur. En yüksek Çevre endeks ortalaması Küme 4'e ait olup, bu kümeyi sırasıyla Küme 2, Küme 3, Küme 5 ve Küme 1 takip etmektedir.

EM algoritmasına ait sonuçlarda da beklentilerin aksi bir sonuçla karşılaşılan Güvenlik endeksi ( $X_7$ ) bakımından Küme 1, Küme 3, Küme 4 ve Küme 5 arasında istatistiksel olarak anlamlı bir farklılık yoktur. Güvenlik endeksi bakımından anlamlı farklılığa sahip olan kümeler Küme 2-Küme 1 ve Küme 2-Küme 3'tür. Fakat Küme 2 ile geriye kalan Küme 4 ve Küme 5 arasında istatistiksel olarak anlamlı bir farklılık yoktur. Bu duruma çalışmanın yapıldığı dönemde ülkemizin çeşitli illerinde yaşanan terör saldırılarının neden olduğu düşünülmektedir. Güvenlik endeksi bakımından en yüksek ortalamaya sahip olan Küme 2'dir. Endeks ortalamaları bakımından Küme 2'yi sırasıyla Küme 5, Küme 4, Küme 3 ve Küme 1 takip etmektedir.

Sivil katılım endeksi ( $X_8$ ) bakımından tıpkı Çevre endeksinde olduğu gibi Küme 2, Küme 3 ve Küme 4 arasında anlamlı bir fark yokken, Küme 1 ve Küme 5 arasında da aynı şekilde istatistiksel olarak anlamlı bir farklılık yoktur. Sivil katılım endeksi bakımından en yüksek ortalamaya sahip olan Küme 4'ü sırasıyla Küme 2, Küme 3, Küme 5 ve Küme 1 takip etmektedir.

Altyapı hizmetlerine erişim endeksi ( $X_9$ ) bakımından Küme 2 ve Küme 3 arasında istatistiksel olarak anlamlı bir farklılık yokken, geri kalan tüm kümeler arasında anlamlı farklılıklar vardır. Altyapı hizmetlerine erişim endeksi bakımından kümeler, pek çok endeks bakımından olduğu gibi Küme 4, Küme 2, Küme 3, Küme 5 ve Küme 1 şeklinde sıralanmaktadırlar.

Sosyal yaşam endeksi ( $X_{10}$ ) bakımından Küme 2 ve Küme 4 arasında anlamlı bir farklılık yokken, Küme 1, Küme 3 ve Küme 5 arasında da anlamlı bir farklılık yoktur. Sosyal yaşam endeksi bakımından kümeler Küme 2, Küme 4, Küme 3, Küme 5 ve Küme 1 şeklinde sıralanmaktadırlar.

Son olarak, yaşam memnuniyeti endeksi ( $X_{11}$ ) bakımından hemen hemen tüm endekslerde en sonlarda yer alan Küme 1 ile, hemen hemen tüm endekslerde en iyi durumda olan, ya da en iyi durumda olan küme ile arasında istatistiksel olarak anlamlı bir fark olmayan Küme 2 arasında, istatistiksel olarak anlamlı bir farklılık yoktur. Geriye kalan Küme 3, Küme 4 ve Küme 5 arasında da istatistiksel olarak anlamlı bir farklılık yoktur. Yaşam memnuniyeti endeksi bakımından en yüksek ortalamaya sahip Küme 1'i sırasıyla Küme 2, Küme 5, Küme 4 ve Küme 3 izlemektedir. İçerisinde ülkemizin en önemli şehirlerini barındıran Küme 4'ün yaşam

memnuniyeti endeksi bakımından sıralamada bu kadar gerilerde kalması ve hemen hemen tüm endekslerde en gerilerde yer alan Küme 1'in yaşam memnuniyeti endeksi bakımından tüm kümelerden daha iyi durumda olması başlı başına bir çalışmanın konusu olabilir.

#### 4. Tartışma ve Sonuç

1949 yılında Bakanlar Kurulu tarafından onaylanan ve resmi gazetede yayınlanarak yürürlüğe giren İnsan Hakları Evrensel Beyannamesinin 21/2 maddesinde "Herkesin ülkesinin kamu hizmetlerinden eşit olarak yararlanma hakkı vardır" denilmektedir [15,16]. TÜİK tarafından yürütülen ve paylaşılan çalışmanın sonuçları ise ülkemizdeki illerin çeşitli endeksler bakımından kamu hizmetlerinden eşit olarak yararlanıp-yararlanmadığını ortaya koymuştur. Buna göre Alpaykut [4] yaptığı çalışmada en iyi yaşamın olduğu illeri İstanbul, Ankara ve İzmir, en kötü yaşamın olduğu illeri ise Mardin, Şanlıurfa ve Siirt olarak belirlemiştir. Bu sonuç ülkemizde bölgeler arasında farklılık olduğunu ifade etmektedir. Benzer şekilde Aydın ve Ertürk [2] ise yaptıkları çalışmada Ankara'nın doğusunda yer alan illerin sosyo-ekonomik gelişme düzeyleri bakımından ülke ortalamasının altında olduğunu ve bu durumun bölgesel bir eşitsizliğe işaret ettiğini ifade etmişlerdir. Aydın ve Yalçın [3] ise Güneydoğu Anadolu Bölgesinde kamu hizmetlerinden memnuniyet düzeylerine göre iller arasında önemli bir farklılık olduğunu belirlemiştir. Bu sonuçlar ülkemizdeki illerin yaşam memnuniyetlerine göre kümelenmesi gerektiğini göstermektedir.

Bu çalışmada söz konusu "İllerde Yaşam Endeksi" çalışmasından elde edilen endeks değerleri kullanılarak 81 il, birbirleriyle benzer olanlar bir araya gelecek şekilde kümelenmişlerdir. Daha önce de açıklandığı üzere, iki farklı kümeleme sonucu elde edilmiştir.

Birinci kümeleme sonucunda ülkemizdeki iller iki kümeye ayrılmıştır. Bu iki kümeden bir tanesi Doğu ve Güneydoğu Anadolu bölgelerinde yer alan illerden, diğeri ise geriye kalan illerden oluşmuştur. Bu kümeler arasında Eğitim, Sağlık, Alt yapı hizmetlerine erişim gibi kamunun tüm vatandaşlara eşit olarak sağlaması gereken endeksler yönünden %99.9 güvenle istatistiksel olarak anlamlı bir farklılık bulunmuştur. Çalışmanın bu aşamasında Küme 2 ve Küme 1 arasında Güvenlik endeksi bakımından anlamlı bir farklılık olmadığı sonucu elde edilmiştir. TÜİK tarafından veri toplama çalışmasının yürütüldüğü 2015 yılının yaz aylarından itibaren, Küme 2'de bulunan bazı illerde yaşanan terör saldırıları ve güvenlik güçlerimizin bu terör örgütlerine karşı yürüttüğü operasyonların varlığı da göz önüne alındığında, elde edilen bu sonuç beklentilerin aksi bir durumu gözler önüne sermiştir.

İkinci kümeleme sonucu incelendiğinde ise ülkemizdeki iller beş farklı kümeye ayrılmıştır. Bu yeni kümeleme sonucu ile bir öncesine göre kümeler daha net bir şekilde ayrılmış/ayrışmış illerden meydana gelmiştir. Böylece tüm endeksler bakımından kümeler arasında istatistiksel olarak anlamlı farklılıklar bulunmuştur. Bu kümeleme sonuçlarına göre de kümeler sağlık, eğitim vb. gibi endeksler yönünden Doğu ve Güneydoğu Anadolu bölgelerinde yer alan illerin oluşturduğu Küme 1 ve Küme 5 gerilerde yer almaktadır. Önceki kümeleme sonucunda olduğu gibi güvenlik endeksi yönünden beş küme arasında beklentilerin aksi bir durum söz konusudur. Güvenlik endeksi bakımından Hakkâri, Şırnak ve Siirt illerinden oluşan Küme 1 ile sadece Küme 2 arasında istatistiksel olarak anlamlı bir farklılık vardır. Geriye kalan kümeler ile Küme 1 arasında anlamlı bir farklılık yoktur. Bu duruma TÜİK tarafından çalışmanın yapıldığı dönemde ülkemizin çeşitli illerinde görülen terör saldırılarının neden olduğu düşünülmektedir.

İkinci kümeleme yaklaşımında belki de en çarpıcı sonuç ise Yaşam memnuniyet endeksi bakımından en yüksek ortalamaya Hakkâri, Şırnak ve Siirt illerinin oluşturduğu Küme 1'in sahip olmasıdır. Ayrıca güvenlik endeksi bakımından Küme 1'den farklı olan tek küme olan Küme 2 ile Küme 1 arasında da yaşam memnuniyet endeksi bakımından istatistiksel olarak anlamlı bir farklılık yoktur. Ülkemizin en gelişmiş metropollerini içerisinde barındıran Küme 4'ün ise yaşam memnuniyet endeksi bakımından sondan ikinci sırada kalması oldukça düşündürücüdür. Bu sonucun sosyologlar tarafından çalışılması ve açıklanması gereken bir durumu ortaya koyduğu açıktır. "Büyük şehir yaşantısı insanların yaşam memnuniyetini düşürür mü?", "İnsanların sahip oldukları imkânların artması, tatmin duygularını köreltir mi?" gibi soruların cevaplarını aramak, elbette bu çalışmanın konusu değildir. Ancak söz konusu sorulara cevap arayan bir çalışmanın yapılması gerektiği de bu çalışmanın sonuçları ile ortaya konulmuştur.

Sonuç olarak ülkemizdeki illerin kümelenmesi, esasında illerde yaşayan bireylerin bir takım göstergelere göre sahip oldukları imkânların kümelenmesidir. Bu çalışma ile de gözler önüne serildiği gibi ülkemizde bölgeler ve kümeler arasında çeşitli göstergeler ve endeksler bakımından önemli farklılıklar mevcuttur. Bu çalışma ile karar alıcılara hangi kümeler arasında hangi endeksler bakımından farklılık olduğu gösterilmiştir ve karar alıcıların faaliyetlerinde bu farklılıkları azaltıcı yönde hareket etmeleri ve yatırımlarını bu sonuçlara göre yapmaları gerektiği düşünülmektedir. Bu yönüyle çalışmanın bilimsel olmasının yanı sıra günlük hayata ve topluma da katkı sağlayabileceğine inanılmaktadır.



**Kaynakça**

- [1] TÜİK, 2016. İllerde Yaşam Endeksi. <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=24561> (Erişim Tarihi: 08.05.2018).
- [2] Aydın, B., Ertürk, N. 2017. Türkiye'de Yaşam Memnuniyeti Endeksi Çerçevesinde Bölgesel Karşılaştırmalar. *Politik Ekonomik Kuram*, 1(2), 118-142.
- [3] Aydın, N., Yalçın, E. 2017. The Comparison of Satisfaction Levels of Southeastern Anatolian Provinces by Multidimensional Scaling Analysis. *Alphanumeric Journal*, 5(1), 15-36.
- [4] Alpaykut, S. 2017. Türkiye'de illerin yaşam memnuniyetinin temel bileşkenler analizi ve TOPSIS yöntemiyle ölçümü üzerine bir inceleme. *Journal of Suleyman Demirel University Institute of Social Sciences*, 29(4), 367-395.
- [5] Atalay, A., Tortum, A. 2010. Türkiye'deki illerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 16(3), 1997-2006.
- [6] Çelik, Ş. 2013. Kümeleme analizi ile sağlık göstergelerine göre Türkiye'deki illerin sınıflandırılması. *Doğuş Üniversitesi Dergisi*, 14 (2), 175-194.
- [7] Tekin, B. 2015. Temel Sağlık Göstergeleri Açısından Türkiye'deki İllerin Gruplandırılması: Bir Kümeleme Analizi Uygulaması. *Çankırı Karatekin Üniversitesi İİBF Dergisi*, 5(2), 389-416.
- [8] Kandemir, A. Ş. 2018. Bulanık Kümeleme Analizi ile Türkiye'deki İllerin Konaklama İstatistiklerine Göre Sınıflandırılması. *Journal of Travel and Hospitality Management*, 15(3), 657-668.
- [9] Aggarwal, C. C., Reddy, C. K. 2014. *Data Clustering Algorithms and Applications*. Taylor & Francis Group, New York, 616s.
- [10] Bulut, H. 2017. Çok değişkenli verilerde robust kümeleme analizi ile boyut indirgeme ve birimlerin sınıflandırılması. *Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi*, 91s., Samsun.
- [11] Yang, M. S., Lai, C. Y., Lin, C. Y. 2012. A robust EM clustering algorithm for gaussian mixture models. *Pattern Recognition*, 45(4), 3950-3961.
- [12] Brock, G., Pilhur, V., Datta, S., Datta, S. 2015. *clValid: An R Package for Cluster Validation*, *Journal of Statistical Software*, 25(4), 1-22.
- [13] Bulut, H., Öner, Y., Sözen, Ç. 2017. Clustering of Member and Candidate Countries of the European Union. *International Journal of Sciences: Basic and Applied Research*, 36(7), 18-25.
- [14] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. 2015. *cluster: Cluster Analysis Basics and Extensions*. <https://cran.r-project.org/web/packages/cluster/cluster.pdf> (Erişim Tarihi: 10.06.2018).
- [15] Bulut, H. 2014. Çok Değişkenli İstatistiksel Analizde Robust İstatistiklerin Kullanımı. *Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*, 77s., Samsun.
- [16] Bakanlar Kurulu Kararı. 1949. İnsan Hakları Evrensel Beyanamesi, <http://www.resmigazete.gov.tr/arsiv/7217.pdf> (Erişim Tarihi: 03.07.2018)