



SIVAS CUMHURİYET ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü
Ekonometri Ana Bilim Dalı

**FARKLI KÜMELEME ANALİZİ YÖNTEMLERİ İLE TÜRKİYE'DEKİ
İLLERİN SOSYO-EKONOMİK ÖZELLİKLERİNE GÖRE
SINIFLANDIRILMASI VE KARŞILAŞTIRILMASI**

Yüksek Lisans Tezi

Yasemin KARAOĞLAN

Sivas
Aralık 2018

**Sivas Cumhuriyet Üniversitesi
Sosyal Bilimler Enstitüsü
Ekonometri Ana Bilim Dalı**

**FARKLI KÜMELEME ANALİZİ YÖNTEMLERİ İLE TÜRKİYE'DEKİ
İLLERİN SOSYO-EKONOMİK ÖZELLİKLERİNE GÖRE
SINIFLANDIRILMASI VE KARŞILAŞTIRILMASI**

Yüksek Lisans Tezi

Yasemin KARAOĞLAN

**Tez Danışmanı
Doç. Dr. Necati Alp ERİLLİ**

**Sivas
Aralık 2018**

KABUL VE ONAY

Üniversite : Sivas Cumhuriyet Üniversitesi
Enstitü : Sosyal Bilimler Enstitüsü
Ana Bilim Dalı : Ekonometri
Bilim Dalı : Ekonometri
Tezin Başlığı : Farklı Kümeleme Analizi Yöntemleri ile Türkiye'deki İllerin Sosyo-Ekonomik Özelliklerine Göre Sınıflandırılması ve Karşılaştırılması
Savunma Tarihi : 21/12/2018
Danışmanı : Doç. Dr. Necati Alp Erilli

Unvanı-Adı Soyadı

Jüri Başkanı : Dr. Öğretim Üyesi Engin KARAKIŞ
Üye : Doç. Dr. Necati Alp ERİLLİ
Üye : Dr. Öğretim Üyesi Özge GÜNDOĞDU

İmza

Shoshur.
Alp Erilli
Özge Gündoğdu

Oy Birliği

Oy Çokluğu

Yasemin KARAOĞLAN Tarafından hazırlanan "Farklı Kümeleme Analizi Yöntemleri ile Türkiye'deki İllerin Sosyo-Ekonomik Özelliklerine Göre Sınıflandırılması ve Karşılaştırılması" başlıklı tez kabul edilmiştir./.../.....

Prof. Dr. Ahmet ŞENGÖNÜL
Enstitü Müdürü

ETİK İLKELERE UYUNLUK BEYANI

Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü bünyesinde hazırladığım bu yüksek lisans tezinin bizzat tarafımdan ve kendi sözcüklerimle yazılmış orijinal bir çalışma olduğunu ve bu tezde;

- 1- Çeşitli yazarların çalışmalarından faydalandığımda bu çalışmaların ilgili bölümlerini doğru ve net biçimde göstererek yazarlara açık biçimde atıfta bulunduğumu;
- 2- Yazdığım metinlerin tamamı ya da sadece bir kısmı, daha önce herhangi bir yerde yayınlamamışsa bunu da açıkça ifade ederek gösterdiğimi;
- 3- Başkalarına ait alıntılanan tüm verileri (tablo, grafik, şekil vb. de dahil olmak üzere) atıflarla belirttiğimi;
- 4- Başka yazarların kendi kelimeleriyle alıntıladığım metinlerini, tırnak içerisinde veya farklı dizerek verdiğim yine başka yazarlara ait olup fakat kendi sözcüklerime ifade ettiğim hususları da istisnasız olarak kaynak göstererek belirttiğimi;

Beyan ve bu etik ilkeleri ihlal etmiş olmam halinde bütün sonuçlarına katlanacağımı kabul ederim.

25.12/2018

Yasemin KARAOĞLAN



TEŐEKKÜR

Tez alıőmam sũresi boyunca alıőmalarımı yœnlendiren, yardımlarını, bilgi birikimini ve œnerilerini esirgemeyen yoęun alıőma temposunda deęerli vaktini sũrekli bana ayıran saygıdeęer hocam Do. Dr. Necati Alp ERİLLİ' ye ve bu dœnem boyunca emeklerini bilgi birikimlerini esirgemeyen bœlœm hocalarıma teőekkũr ederim.

Ayrıca bu sũrete sũrekli deneyimlerinden faydalandıęım Prof. Dr. Ahmet ŐENGœNœL, Yrd. Do. Dr. İlkay NOYAN YALMAN ve Yrd. Do. Dr. Őebnem ZORLUTUNA' ya vermiő oldukları destek ve emeklerinden dolayı teőekkũrlerimi sunuyorum.

Tezin hazırlanması aőamasında dua ve desteklerini esirgemeyen anneme, babama ve kardeőlerime, yũksek lisansa baőlamam iin bana destek olan bu sũrete sabır, anlayıő ve fedakârlık gœsteren hayat arkadaőım, eőim Ahmet KARAOęLAN' a ve son olarak varlıęıyla, sevgisiyle en stresli anımda yũzũmũ gũldũren canım kızım Zeynep İlay KARAOęLAN' a en iten duygularımla teőekkũr ediyorum.

İÇİNDEKİLER

İÇİNDEKİLER	i
TABLO LİSTESİ.....	v
ÖZET	vii
ABSTRACT.....	ix
GİRİŞ	1
BÖLÜM 1	3
1. LİTERATÜR TARAMASI.....	3
BÖLÜM 2	9
2. MATERYAL VE METOD	9
2.1. Sosyo Ekonomik Gelişme	9
2.2. Kümeleme Analizi	16
2.2.1. Kümeleme Analizinde Kullanılan Benzerlik ve Uzaklık Ölçüleri.....	19
2.2.1.1. Öklid (Euclidean) Uzaklığı.....	19
2.2.2.2. Karesi Alınmış Öklid Uzaklığı.....	20
2.2.2.3. Manhattan (City-Block) Uzaklığı.....	20
2.2.2.4. Chebychev Uzaklığı	20
2.2.2.5. Minkowski Uzaklığı	21
2.2.2.6. Mahalanobis Uzaklığı.....	21
2.2.2.7. Canberra Uzaklığı.....	21
2.2.2.8. Hotelling T^2 Uzaklığı.....	21
2.2.2.9. Biserial Korelasyon Ölçüsü	22
2.2.2.10. Pearson Korelasyon Ölçüsü.....	22
2.2.2.11. Spearman Sıra Korelasyon Katsayısı	22
2.3. Kümeleme Analizi Yöntemleri	22
2.3.1. Hiyerarşik (yığılmalı) Kümeleme Yöntemleri	23
2.3.1.1. Toplanmış Hiyerarşik Kümeleme.....	24
2.3.1.1.1. Tek Bağlantı Yöntemi.....	24
2.3.1.1.2. Tam Bağlantı Yöntemi.....	25
2.3.1.1.3. Ortalama Bağlantı Yöntemi	26
2.3.1.1.4. Ward's Bağlantı Yöntemi	27
2.3.1.2. Bölünmeli Kümeleme Yöntemi.....	27

2.3.2. Hiyerarşik Olmayan Kümeleme Yöntemi	27
2.3.2.1. k-Ortalamlar Tekniği.....	28
2.3.2.2. En Çok Olabilirlik Yöntemi.....	29
2.4. Küme Geçerlilik İndeksi.....	29
2.4.1. Küme Sayısının Belirlenmesi ve Küme Geçerlilik İndeksleri.....	30
2.4.2. Klasik Kümeleme Analizinde Küme geçerlilik indeksleri	31
2.4.2.1.Silhouette İndeksi.....	32
2.4.2.2. Calinski ve Harabazs İndeksi.....	33
2.4.2.3. Krzanowski ve Lai indeksi.....	33
2.4.2.4. Dunn indeksi	34
2.4.2.5.Wilk's lambda istatistiği	34
BÖLÜM 3.....	39
3. UYGULAMA.....	39
3.1. 2002 Yılı İçin Bulgular ve Sonuçlar.....	39
3.1.1. 2002 Yılı Sosyo-Ekonomik Verilerinin Gruplar Arası Yöntemi İle Kümeleme Sonuçları.....	39
3.1.2. 2002 Yılı Sosyo Ekonomik Verilerin Grup İçi Yöntemi İle Kümeleme Sonuçları	41
3.1.3. 2002 Yılı Sosyo Ekonomik Verilerin En Yakın Komşuluk Yöntem İle Kümeleme Sonuçları.....	41
3.1.4. 2002 Yılı Sosyo Ekonomik Verilerin En Uzak Komşuluk Yöntemi İle Kümeleme Sonuçları.....	42
3.1.5. 2002 Yılı Sosyo Ekonomik Verilerin Centroid Kümeleme Yöntemi İle Sonuçları	42
3.1.6. 2002 Yılı Sosyo Ekonomik Verilerin Median Kümeleme Yöntemi İle Sonuçları	42
3.1.7. 2002 Yılı Sosyo Ekonomik Verilerin Ward Yöntemi İle Sonuçları.....	43
3.2. 2008 Yılı İçin Bulgular ve Sonuçlar.....	43
3.2.1. 2008 Yılı Sosyo Ekonomik Verilerinin Gruplar Arası Yöntemi İle Kümelenme Sonuçları.....	43
3.2.2. 2008 Yılı Sosyo Ekonomik Verilerinin Grup İçi Yöntemi İle Kümelenmesi Sonuçları	44

3.2.3. 2008 Yılı Sosyo Ekonomik Verilerinin En Yakın Komşuluk Yöntemi İle Kümelenmesi Sonuçları.....	44
3.2.4. 2008 Yılı Sosyo Ekonomik Verilerinin En Uzak Komşuluk Yöntemi İle Kümelenmesi.....	44
3.2.5. 2008 Yılı Sosyo Ekonomik Verilerinin Centroid Kümeleme Yöntemi İle Kümelenmesi Sonuçları.....	45
3.2.6. 2008 Yılı Sosyo Ekonomik Verilerinin Median Kümeleme Yöntemi İle Kümelenmesi Sonuçları.....	45
3.2.7. 2008 Yılı Sosyo Ekonomik Verilerinin Ward Yöntemi İle Kümelenmesi Sonuçları.....	45
3.3. 2013 Yılı İçin Bulgular ve Sonuçlar	46
3.3.1 2013 Yılı Sosyo Ekonomik Verilerin Gruplar Arası Yöntemi İle Kümelenmesi Sonuçları.....	46
3.3.2. 2013 Yılı Sosyo Ekonomik Verilerin Grup İçi Yöntemi İle Kümelenmesi Sonuçları.....	46
3.3.3. 2013 Yılı Sosyo Ekonomik Verilerinin En Yakın Komşuluk Yöntemi İle Kümelenmesi Sonuçları.....	46
3.3.4. 2013 Yılı Sosyo Ekonomik Verilerinin En Uzak Komşuluk Yöntemi İle Kümelenmesi Sonuçları.....	47
3.3.5. 2013 Yılı Sosyo Ekonomik Verilerinin Centroid Kümeleme Yöntemi İle Kümelenmesi Sonuçları.....	47
3.3.6. 2013 Yılı Sosyo Ekonomik Verilerinin Median Kümeleme Yöntemi İle Kümelenmesi Sonuçları.....	48
3.3.7. 2013 Yılı Sosyo Ekonomik Verilerinin Ward Yöntemi İle Kümelenmesi Sonuçları.....	48
3.4. Kümeleme Yöntemleri Arasında Korelasyon Analizleri	48
SONUÇ ve TARTIŞMA	51
KAYNAKLAR	55
EKLER	63
ÖZ GEÇMİŞ	87

TABLO LİSTESİ

Tablo No	Tablo Adı	Sayfa
Tablo 3.1.	2002 yılı Sosyo ekonomik verilerinin Gruplar Arası yöntemi ile kümeleme sonuçları.....	40
Tablo 3.2.	2002 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri	49
Tablo 3.3.	2008 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri	49
Tablo 3.4.	2013 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri	50



ÖZET

Bu çalışmada; Türkiye'deki 3 farklı yıla 55 değişken ait 81 ilin sosyo-ekonomik verilerin klasik kümeleme analizi yöntemlerinden hiyerarşik kümeleme analizi olan 7 farklı (Gruplar Arası, Grup İçi, En Yakın Komşuluk, En Uzak Komşuluk, Centroid Metodu, Medyan, Ward's Yöntemi) teknik ile analiz yapılarak karşılaştırmalı olarak incelenmiştir. Verilere yapılan analizlerin uygunluğu klasik kümeleme analizi geçerlilik indekslerinden olan Silhoutte küme geçerlilik indeksi ile araştırılmıştır ve her veri seti için uygun küme sayısının 5 olmasına karar verilmiştir. Türkiye istatistik kurumunun internet sitesi üzerinden alınan veriler 2002, 2008 ve 2013 yıllarına ait sosyo-ekonomik dinamikleri kuvvetli veriler kullanılmıştır. Sosyo-ekonomik veriler için farklı yıllara ait değişkenlerin ölçüm düzeyleri farklılık göstermesinden dolayı çalışmanın uygulama kısmında değişkenler için standartlaştırılma yapılmıştır. Bu yıllara ait yapılan küme geçerlilik indeksi sonuçlarında da 2002 yılı için grup içi ve medyan kümeleme arasında yüksek ilişki gözlemlenmiştir. 2008 yılı geçerlilik indeksi sonuçlarında ise Yakın Komşuluk Kümeleme ile Medyan Kümeleme sonuçları aynı çıktığı görülmektedir. Yüksek ilişkili kümeleme yöntemleri olsa da, sonuçlar 2002 yılına ait sonuçlar kadar yüksek olmadığı incelenmiştir. 2013 yılında ise Gruplar Arası Kümeleme ile Medyan Kümelemenin aynı sonuçlar verdiği, En Uzak Kümeleme ile Ward Kümelemenin yüksek korelasyon katsayısına sahip olduğu görülmektedir. Klasik kümeleme analizi yöntemlerinden ward's yöntemi 2002, 2008 ve 2013 yılları için genel olarak gerçek sosyo-ekonomik yapılarla bağdaşan sonuçlar çıkarmıştır.

Anahtar Kelimeler: Kümeleme Analizi, Hiyerarşik kümeleme analizi, Sosyo - Ekonomik, Küme geçerlilik, Silhoutte, Korelasyon



ABSTRACT

In this study; 3 belonging to different years in Turkey 55 variations of the 81 provinces of the socio-economic data of classic cluster analysis method of hierarchical clustering analysis with 7 different (between, within groups, Nearest Neighborhood, furthest neighbor, centroid method, median, Ward's Method) techniques In this study, it was analyzed by comparative analysis. The suitability of the analyzes to the data was investigated with the Silhouette cluster validity index, which is one of the classical clustering analysis validity indices, and it was decided that the appropriate number of sets for each data set would be 5. Data received via the website of Turkey's statistical agency in 2002, 2008 and 2013, the dynamics of socio-economic data were used strong. As the measurement levels of variables for different years differ for socio-economic data, standardization was applied for variables in the application part of the study. The cluster validity results of these years showed a high correlation between intra-group and median clustering. In the results of 2008 validity index, it is seen that Close Cluster Clustering and Median Clustering results are the same. Although there are high within groups clustering methods, the results are not as high as the results for 2002. In 2013, it was observed that between groups and Median Clustering gave the same results, and Ward Clustering with the furthest neighbor Clustering had a high correlation coefficient. The wardtirs method, which is one of the classical clustering analysis methods, has shown concordant with the actual socio-economic structures for the years 2002, 2008 and 2013.

Key Words: Cluster Analysis, Hierarchical clustering analysis, Socio - Economic, Cluster validity, Silhouette, Correlation

GİRİŞ

Dünyanın varoluşundan beri insanoğlu hayatının her evresinde karar verme eylemindeydir. Bu karar verme durumunda, elimizde var olan bilgileri düşünce sistemimizden geçirerek sonuca varmaya çalışırız. Sonuca varma aşamasından önce elimizde yeterli bilgilerin olup olmadığını ve bu bilgiye nereden ve nasıl ulaşılması gerektiği gibi soruları sorgulama yoluna gideriz. Sorgulamanın temelinde aslında karmaşık bir yapı yatmaktadır. Zira insanoğlunun problemleri çeşitli türlerde ve birbirine benzemeyen yapılarda olabilmektedir. İhtimallerin bu kadar çok olduğu bir hayatta alınacak kararların ve içinde bulunan koşulların en iyi şekilde değerlendirilmesi karar alma sürecinde büyük önem arz etmektedir. Bu süreçte karşılaşılan problemlerin çeşitliliği ve sorulara aranan cevapların bilimselliği araştırmacıyı istatistik bilimine yönlendirecektir.

İstatistik, belirsizlik altında en doğru bilgiyi vermeye çalışan, gelecek hakkında geçmişle bağlantılı olarak öngörü yapabilen bilim dalıdır. İstatistik, bu öngörüyü üç aşamada gerçekleştirmektedir. Birinci aşama, çalışma yapılacak olay hakkında verilerin toplanmasıdır. İkinci aşama, toplanan verilerin sınıflandırılmasıyla (gruplandırılması) istatistiksel yöntemler uygulanarak olaylar arasında ilişkinin belirlenmesidir. Üçüncü aşama ise, olaylar arasındaki ilişkinin yorumlanması bu yoruma en uygun modellemenin yapılması ve öngöründe bulunmaya en yakın sistemleştirilmenin yapılmasıdır. İstatistik bilimi, kâinatın her evresindeki sayısal veya sayısal hale getirebilen her tür süreci sistemleştirebilen bilimdir.

Bilimsel olayları açıklamak ve öngöründe bulunmak için bazen tek değişkene bazen de çok değişkene ait verilerin toplanması gerekebilir. Araştırmacı tek değişkenle çalışacak ise tek değişkenli istatistiksel yöntemleri kullanması gerekirken, birden çok değişkenin yer aldığı bir analizde ise çok değişkenli istatistiksel yöntemleri kullanması daha gerçekçi bir yaklaşım olacaktır. Uygulamalarda birden çok değişkenin var olması, tek değişkenli istatistiksel yöntemleri kullanmayı sınırlandırarak çok değişkenli analiz tekniklerinin doğmasına neden olmuştur. Ayrıca tek değişkenli istatistiksel yöntemlerde çalışmanın zorluğu ve uyulması gereken varsayımlardır. Buna karşılık çok değişkenli istatistiksel yöntemlerin en önemli özelliği ise, normal dağılımın var olduğu kitle içerisinden çekilmiş olmasıdır.

Birden çok deęişkenin yer aldığı çalışmalarda arařtırmacının MANOVA, Faktör Analizi, Çok Deęişkenli Regresyon Analizi ve Kümeleme Analizi gibi farklı istatistiksel analizlerden yararlanması daha doğru olacaktır. Ancak mevcutta var olan çok deęişkenli istatistiksel yöntemlere ek olarak yeni geliştirilen yöntemlerle birleřtirmeler yapılarak yeni çözüm yolları aranmaktadır. Geliştirilen bu alternatif yöntemlerle bilimsel arařtırmalara farklı bakış açıları kazandırılmaya çalışılmaktadır. Bu bağlamda; yapay sinir aęları, bulanık mantık, bayesci istatistik, aktüerya, veri madencilięi gibi yeni alanlar istatistik ve ekonometri bilimlerinin ilerlemesine katkı sağlamaktadır (Erilli 2009).

Bu çalışmada çok deęişkenli istatistiksel analizlerden klasik kümeleme analizi üzerinde durulmuştur. Uygulamada TÜİK' ten alınmış 3 farklı döneme ait sosyo-ekonomik verilerin, farklı kümeleme analizi yöntemleri kullanılarak kümelendirilmiştir. Yöntemler arası benzerlik veya farklılıklar ortaya konmaya çalışılmıştır. Çalışmanın giriş kısmından sonra, ikinci bölümde literatür taramasına yer verilmiştir. Çalışmamızın üçüncü kısmında sosyo ekonomik veriler hakkında açıklamalar yapılmıştır. Uygulama kısmında çalışmada tanıtılan bazı Klasik Kümeleme yöntemleri kullanılarak, SPSS.25'de analizler yapılmıştır. Türkiye'deki illerin farklı dönemlere göre gelişmişlik düzeyindeki artış ya da azalışları kümeleme yöntemiyle gruplandırılmaya çalışılmıştır.

BÖLÜM 1

1. LİTERATÜR TARAMASI

Akat (2007); çalışmasında öncelikli olarak askeri yapı ve bunu etkileyen temel faktörler üzerinde durmuş ve belirlenen faktörlere göre kümeleme analizi yapılmıştır. Hiyerarşik ve hiyerarşik olmayan kümeleme tekniklerine değinmiştir. Sonuç olarak; 8 adet kümeden oluşan yapı uygun bulunmuştur. Farklı kümeleme metotlarından faydalanılmış sonuç olarak ward en uygun yöntem olarak bulunmuştur.

Akgöz (2010); Türkiye’de faaliyet gösteren kamu, özel, yabancı sermayeli bankaların, mali tablolarından üretilen oran verileri alınarak benzerlik ve farklılıkları yönüyle çok değişkenli istatistiksel yöntemlerden kümeleme analizi yöntemiyle gruplanmaya çalışılmıştır. Kamu, özel, yabancı sermayeli ayrımı yapılmaksızın bu gruplanmanın amacı hangi bankalar arasında homojen yapı olduğu ortaya çıkarmaktır. Hiyerarşik kümeleme analizlerinden “ward” yöntemi ile bankalar yeniden sınıflandırılmış ve ANOVA analizleri yapılmıştır.

Akın (2008); 2004’te yapılan hane halkı bütçe anketi sonuçlarına göre yapılan uygulamada merkeze dayalı bölümleyici kümeleme tekniklerinden biri olan CLARA yöntemi ve veri madenciliğinde kullanılan yoğunluk tabanlı kümeleme tekniğinden yararlanarak uygulama yapılmıştır.

Alkan (2012); çalışmasında kümeleme analizine teorik olarak değinilmiş ve diskriminant analizi de anlatılmıştır. Farklı kümeleme yöntemleriyle Bingöl, Elazığ, Malatya, Tunceli il ve ilçe merkezlerindeki hanelerin yıllık elektrik tüketim değerleri dikkate alınarak 5 kümeye ayrılmıştır. Elde edilen kümeler, diskriminant analizi ile doğru sınıflara atanma oranları elde edilmiştir.

Atbaş (2008); çalışmasında önceden bilinmeyen küme sayısını belirlemeye yönelik bir çalışmadır. Tek bağlantı, tam bağlantı, ward yöntemi ve k-ortalamlar yöntemi kullanılmıştır. Uzaklık ölçüsü olarak Öklid uzaklığı ve karesel Öklid uzaklığı kullanılmıştır. 81 ilin farklı suç türüne göre cezaevine giren hükümlü sayıları illerin nüfuslarına oranlanmıştır. Kümelerin kalitesi küme geçerlilik indeksleriyle test edilmiştir.

Çam (2014); çalışmasında ölçek tipine göre uzaklıkların belirlenmesiyle ilgili çalışılmış ve kümeleme analizi teknikleriyle ilgili teorik yapısı anlatılırken veri madenciliği konusu da teorik olarak anlatılmıştır. Uygulamada bir hastanedeki hastaların verilerine veri madenciliği kümeleme analizi yöntemlerinden olan k-ortalamlar ve yoğunluk tabanlı kümeleme algoritması uygulanmıştır. Hastaların demografik verileri ise ki-kare, kruskall-wallis H ve mann-whitney U testleri ile incelenmiştir.

Çakmak ve ark. (2005); çalışmasında küreselleşmeye bağlı olarak binlerce yıllık değişen kültürel değerlerin birikiminin çok değişkenli istatistiksel analiz yöntemlerinden kümeleme analizi ile açıklamaya çalışmışlardır.

Darakçı (2011); çalışmasında veri madenciliği kavramlarının nasıl oluştuğu, bir veri madenciliği uygulamasında hangi süreçler olduğu, veri madenciliğinde kullanılan metotlar ve teknikler, kümeleme analizi yöntemlerinden k-ortalamlar yöntemi ile ele alınarak anlatılmıştır.

Dinler (2014); Türkiye’deki 81 ilin hayvancılık potansiyeli yedi farklı kümeleme analizi yöntemi ve beş farklı uzaklık ölçüsü ile karşılaştırmalı olarak incelenmiştir. Çalışmada TÜİK(2012) yılına ait hayvancılık verilerinden yararlanılmıştır.

Doğan (2008); çalışmasında kümeleme analizi teorik olarak anlatılmış ve tekniğin pazarlama, finans ve ekonomi alanındaki uygulamaları irdelenmiştir.1998 - 2006 yıllarındaki Türk ticari bankaları verilerinden yararlanılarak uygulama yapılmıştır.

Ergün (2008); çalışmasında amacı ürün kategorileri ve ürün sınıfları arasındaki ilişkileri ortaya çıkarmaktır. Bu amaçla bir perakendeci işletmenin, alışveriş detaylarını içeren veri madenciliği yöntemleri (GRI ve apriori, hiyerarşik kümeleme analizi) ile incelenmiştir.

Erilli (2009); çalışmasında kümeleme ve bulanık kümeleme analizi arasındaki farklar belirtilmiştir. Gözlem sayıları ve değişken sayıları arttıkça veya küme yapıları birbirine çok yakın olduğu durumlarda bulanık kümeleme analizinin diğer kümeleme yöntemlerine göre daha başarılı olduğu görülmüştür. Uygulamada 81 ilin sosyo-ekonomik verileri incelenmiştir.

Erilli (2012); çalışmasında kümeleme analizi, bulanık kümeleme analizi ve bulanık sayılarda bulanık kümeleme analizi yöntemleri teorik olarak tanıtılmış ve aralarındaki fark dile getirilmiştir. Uygulama kısmında ise satranç oyunlarının maç skorlarına ve ECO kodlarına göre sınıflandırılmasına çalışılmıştır.

Erişoğlu (2011); çalışmasında kümeleme analizi tekniklerine uygun uzaklık ölçülerinin kullanılmasıyla, bu tekniklerin performanslarının artırması amaçlamıştır. Bu amaçla 7 uzaklık ölçüsü hem gerçek veri seti hem de simülasyon çalışmalarıyla incelenmiştir.

Gül (2014); çalışmasında çok değişkenli istatistik yöntemlerinden kümeleme analizi tekniklerine değinmiştir. AB'ye Üye Ülkelerin ve Türkiye'nin kriz öncesi, kriz yılı ve sonrası yıllar Maastricht Kriterleri doğrultusunda seçilen ekonomik değişkenlere göre hangi küme grubunda olduğu ortaya koymuştur.

Gün (2011); çalışmasında güvenlik açısından biyometrik sistemlerin önemi ve yüz tanımanın biyometrik tanıma sistemleri arasındaki yerini konu edinmiştir. Yapay sinir ağları, k-ortalamlar ve bulanık c-ortalamlar algoritmasına dayalı sınıflandırmaya yüz tanıma işlemi yapılmıştır.

Köse (2015); çalışmasında veri madenciliği, müşteri, perakendecilik gibi teorik kavramlara kapsamlı yer verilmiştir. Konya merkezli 9 ilde faaliyet gösteren bir işletmenin alışveriş fişleri toplanmıştır. Toplanan veriler üzerinden veri madenciliği birliktelik kuralları analizi ve kümeleme analizi yöntemlerinden hiyerarşik kümeleme metodu uygulanmıştır.

Kutsal (2008); ülkeler bölgelerinde lojistik bir üs haline gelerek, lojistiğin sunduğu ekonomik gelişim fırsatlarından en iyi şekilde faydalanmaya çalışmaktadır. Bu çalışmada lojistik bir üs olmak için gerekli kriterler incelenmiştir. Bu kriterler için kümeleme analizi metodu kullanılarak, Türkiye'nin lojistik bir üs olma potansiyeli incelenmiştir. Sonuç olarak ; lojistik üs olma potansiyeline sahip ülkelerin yani potansiyelin kuvvetli olduğu ülkelerin ilk üç kümede yer aldıkları gözlenmiştir.

Nacaroğlu (2010); çalışmasında deprem etkisiyle oluşan alt yapıdaki boru hasarlarıyla ilgili coğrafi bilgi sistemi kümeleme analizi yöntemlerini birlikte değerlendirilmiştir. Veri seti üzerinde bulanık c-ortalamlar kullanılarak küme geçerlilik indeksleri uygulaması yapılmıştır.

Onan (2013); çalışmasında kümeleme analizinde karınca kümeleme algoritmalarının kullanılmasına yönelik gerçekleştirilen çalışmalar incelenmiş, karınca kümeleme yaklaşımını, geleneksel bir kümeleme algoritması olan k-ortalama algoritmasıyla bütünleştirerek daha etkin çözümler elde etmeyi amaçlayan antClass algoritması temel alınmıştır.

Sayın (2010); çalışmasında sosyo ekonomik gelişme ile ilgili temel kavramlardan bahsedilmiştir. Bu çalışmanın asıl amacı olan karışım(hibrid) kümeleme analizi hakkında tamamlayıcı teknik bilgilere yer verilmiştir. Oluşturulan kümelerin geçerliliğini test etmekte olan yöntemler hakkında bilgi verilmiştir.

Selanik (2007); çalışmasında amacı Türk tarımının Avrupa Birliği içindeki yerini belirlemek ve hangi ülkelerle grup oluşturduğunu bulmaktır. Bu amaçla uygulama bölümünde 15 ülke ile Türkiye kümeleme analizi teknikleriyle gruplandırılmış ve Ward's yöntemine değinilmiştir.

Servi (2009); çalışmasında karma dağılım modeli, çok değişkenli kümeleme analizi ve karma kümeleme analizi incelenmiştir. Çok değişkenli kümeleme analizinde kavramlar, tanımlar, gösterimler ve yöntemler anlatılmıştır. Uygulama bölümünde ise; karma normal kümeleme analizi uzaktan algılanmış çok bantlı uydu görüntüsünün kümelmesi ve alan bazında eğitilmiş sınıflandırılması uygulamasına yer verilmiştir.

Silahtaroglu (2004); çalışmasında son zamanlarda gelişme gösteren veri madenciliği teknikleri üzerinde durulmuştur. Veri depolama kapasitelerindeki artışa bağlı olarak algoritmalara olan ihtiyaç dile getirilmiştir.

Şekerler (2008); çalışmasında klasik ve bulanık kümeleme yöntemleri için trafik kaza verileri, paket programları yardımı ile 2 parametrenin kullanıldığı bir analize tabi tutularak kümelmiş ve bunun sonucunda ve ortaya çıkan sonuçlar yorumlanmıştır. Ayrıca belirlenen kara noktalar için farklı çözüm önerileri getirilmiştir.

Şimşek (2006); çalışmasında çok değişkenli istatistiksel kümeleme tekniklerinden kümeleme analizi, çok boyutlu ölçekleme analizi, doğrulayıcı ve açıklayıcı faktör analizi ile elde edilen yapı geçerliliği kanıtları ve sonuçları karşılaştırmıştır.

Ulutugay (2004); bu çalışmada bulanık kümeleme kavramı üzerinde durulmuştur. Rastgelelik ile bulanık kümeleme kavramları karşılaştırılmıştır.

Yalçın (2013); çalışmasında teorik olarak hiyerarşik kümeleme analiz ve hiyerarşik olmayan kümeleme analizine değinilmiştir. Çalışmanın uygulamasında Elazığ iline ait organize sanayi bölgesinde bulunan iş yerlerinde, gürültü düzeyi 95 db'in üstünde olan ve en az 10 işçinin çalıştığı 6 fabrikada çalışan işçilerdeki işitme kaybına göre fabrikalar hiyerarşik kümeleme tekniklerinden olan ward kümeleme metoduyla sınıflandırılmıştır.

Yaşar (2009); teorik olarak doğrudan yabancı yatırımcıların yöntemlerine değinilmiştir. 1980-2005 yılındaki değişkenlerle kümeleme analizi uygulanmış ve benzer özellikler gösteren ülkeler aynı grupta yer almıştır. Kişi başına GSMH, gelişmişlik sıralaması, ülkenin yükselen piyasa olup olmaması, OECD üyeliği, 1000 kişiye düşen sabit hatlı ve mobil telefon sayısı küme ayrıştırmasındaki en etkili değişkenler olmuştur.

Yazdanoğlu (2011); bu çalışmada üç çeşit kümeleme yaklaşımı, k-means, two stage ve self organizing map yöntemleri, yaşam biçimleri veri seti üzerinde analiz edilmiştir. Yapılan çalışmanın ilk konusu tüketici segmentini yaşam biçimlerine uyarlamak, ikinci kısmı ise üç kümeleme yaklaşımını karşılaştırmaktır.

Yıldırım (2007); çalışmasında bütün kümeleme yaklaşımlarının yanı sıra, detaylı olarak hiyerarşik olmayan kümeleme yaklaşımına dayanan ve veri setindeki veri grupları arasında kesin ayrımının söz konusu olmadığı durumlarda başarıyla uygulanan Fuzzy (Bulanık) Kümeleme Algoritmaları incelenmiş ve ağırlıklı olarak Fuzzy (Bulanık) C-Ortalamlar tekniği ele alınmıştır.

Yıldız (1998); çalışmasında kümeleme analizi, diskriminant analizi ve tercih çözümünü konularına teorik olarak değinilmiştir. Tamsayı programlaması ve gruplandırılması teorisinden bilgi verilmiştir. Matematiksel formülasyonlar hazırlanıp, bu matematiksel modellere kümeleme analizinde yapılan işlemler yapılarak kümeleme oluşturulmaya çalışılmıştır.

Yılmaz (2011)' a göre, Türkiye'deki iller sosyo-ekonomik gelişmişlik düzeyleri açısından önemli farklılıklar göstermektedir. 81 ilin gelişmişlik düzeylerinin tahmin edilme amacıyla kalkınma büyüme sosyo- ekonomik gelişmişlik, dengesizlik

kavramlarına değinilmiştir. Bu amaçla diskriminant analizi, temel bileşenler, faktör ve kümeleme analizi kavramlarına kısaca değinilmiştir. Sonuç olarak; faktör analizine göre 5 faktör elde edilirken, kümeleme analizinde iller 3 gruba ayrılmıştır.

Yılmaz (2013); çalışmasında 31 ülkenin 2008 ve 2009 yılı sermaye yeterlilik oranlarını kullanarak kümeleme analizi yöntemiyle sınıflandırılmış, Basel III kriterlerinin getirdiği yaptırımları, Türk bankacılık sektörünü 2012 yıl öz kaynak yapısı ele alınarak etkileyip etkilemediği incelenmiştir.



BÖLÜM 2

2. MATERYAL VE METOD

2.1. Sosyo Ekonomik Gelişme

Sosyo-ekonomik gelişme bir toplumun demografik yapısının, kırsal nüfus, doğum ve ölüm oranlarının, yatay ve dikey hareketliliğin, şehirleşme hızının, aile büyüklüğünün, gelir dağılımının ekonomik gelişmeye paralel değişikliklere uğraması olarak tanımlanmaktadır (Sayın 2010).

Sosyo-ekonomik gelişme, kişi başına düşen milli gelirin artırılması şeklinde özetlenebilecek iktisadi büyüme kavramıyla beraber, yapısal ve insani gelişmeyi içine alan ve ölçülebilir bütün sosyal değişkenleri içermektedir (Dinçer, Özaslan, Kvasoğlu 2003: 14).

Bir ülkenin farklı bölgelerinin gelişme düzeylerini ekonomik, sosyal, tarihsel, coğrafi, iklimsel ve stratejik faktörler belirlemektedir. Bölgelerin sosyo ekonomik gelişme farklılıklarının ortaya koyulması yatırım projeleri ve politika yapıcılarının bölge halkına yapacağı yatırımlar bakımından büyük önem arz etmektedir. Bu bağlamda gelişmenin/büyümenin nasıl ölçüleceği ve daha fazla artış gösterebileceği ekonomideki tartışmalı konulardan biridir.

Gelir artışlarıyla ölçülen ekonomik *büyüme*, toplumsal gelişmişliği yeterince açıklayamayan, yalnızca talep genişlemesini açıklayabilen dar kapsamlı bir parametredir. Büyüme kavramı için en belirgin özelliğin “bir ekonominin üretim kapasitesinde, sayısal olarak ölçülebilen genişleme veya miktar artışı” olduğu söylenebilir. Ekonomik kalkınma kavramı, niteliksel değişmeye işaret etmektedir. Bu durumda ekonomik kalkınma hem daha fazla çıktı hem de teknik ve kurumsal yapıdaki değişimleri kapsamaktadır. Büyüme ile bir karşılaştırma yapılarak denebilir

ki, ekonomik büyüme daha çok aynı şeydeki basit artış sürecini, ekonomik kalkınma ise daha fazla ve farklı olanın yer aldığı yapısal değişme sürecini ifade eder (Flammang 1979). Ekonomik kalkınma sadece ekonomik boyutlarla sınırlanmayan, toplumu sosyolojik, psikolojik ve politik tüm boyutlarıyla kuşatan karmaşık bir süreçtir (Yavilioğlu 2002).

Gelişme kavramı ise zaman içinde farklı anlamlar ifade ede gelmiştir. Gelişme kavramı, 19. yüzyılda “ekonomik büyüme” anlamına gelirken, temel ölçütler milli ya da kişisel gelir, yaratılan katma değer, sanayi sektöründe üretim/çalışan hacmi, vb. olarak tanımlanmaktaydı. 20. yüzyılın ilk çeyreğinde “sosyal refah” içeriğine kavuşan gelişme kavramı, gelir yanında insanların/toplumların sahip oldukları kolaylıkları (fiziksel ve sosyal altyapı, vb.) da kapsar hale gelmiştir. 20. yüzyılın son çeyreğinde ise gelişme “yaşam kalitesi” ile ölçülmeye başlanmıştır. Yaşam kalitesi, özellikle nitelikli doğal, fiziksel, sosyal ve kültürel çevrenin varlığı ve tüketilmesi/tüketilme olanağına kavuşulması anlamını taşımaktadır (Üstünişik 2007).

1970'lere kadar kişi başı GSYİH (veya GSMH) gelişmenin temel ölçülerinden biri olarak kabul edilirken bu yıldan itibaren insani gelişimin ölçülmesinde tek başına bu kıstasın yeterli olup olmadığı tartışmalı hale gelmiştir. Ancak bu yaklaşım gelişmenin sadece ekonomik boyutunu öne çıkardığı, sosyal boyutuna vurgu yapmadığı için zamanla açıklayıcılık gücünü kaybetmiştir (Yıldız Sivri Berber 2012: 148). Özellikle gelir dağılımı hakkında çok az bilgi taşıması, bu değişkenin gelişimi ölçmedeki yetersizliğini ortaya koymaktadır. Buna bağlı olarak 1970'li yıllardan sonra gelişmişlik düzeyinin ölçümünde sosyo-ekonomik değişkenlere verilen önem giderek artmıştır. Gelişmenin en iyi ölçüsü olarak görülen kişi başına gelir sorgulanır hale gelmiş ve gelişmeyi ölçmek için birden çok refah ölçütünün bileşiminden oluşan bir kıstas arayışı başlamıştır (Kılıç 2011).

Gelişme için milli gelirin artması yeterli değildir. Bunun yanında örneğin okuryazarlık oranının, sanayinin istihdam ve milli gelirdeki paylarının vb. artması gerekmektedir (Manisalı 1975: 2). Örneğin az gelişmiş bölgelerde istihdam olanakları kısıtlı olduğundan işsizlik söz konusudur ve buna bağlı olarak bu bölgelerde gelirler düşüktür. Ücretlerin ve dolayısıyla satın alma gücünün düşüklüğü bu bölgelerde sosyal hizmetlerin gelişmemesine neden olur. Bu bölgelerde sağlık hizmetleri geridir, eğitim hizmetleri yeterince gelişmemiştir. Az gelişmiş bölgelerin gelir düzeyi ve gelir artış hızı, diğer bölgelerin gelir düzeyleri ve gelir artış hızlarından düşüktür. Tarım hakim iktisadi faaliyettir ve modern tarım girdilerinden yeterince faydalanılmamaktadır. Doğum oranı dolayısıyla nüfus artış hızı genellikle ülke nüfus artış hızından yüksektir. Gelişmiş bölgelere göre düşük sosyo-ekonomik göstergeler göze çarpmaktadır. Emeğin istihdam olanakları sınırlı olduğu gibi, sabit sosyal altyapı yatırımları da yetersizdir.

Üretken yatırımlar bölgeye çekilememektir. Tasarruf oranı gelir düzeyine bağlı olarak düşüktür (Kaya 2009: 7).

“Büyüme”, “kalkınma” ve “gelişme” kavramları, denk oluşumları içerdiği düşünce itibariyle, eşanlamlarda kullanılmış kavramlardır. Bununla birlikte, İkinci Dünya Savaşı sonrasında yaşanan gelişmeler, ulusal kalkınma sürecinde iktisadi büyüme olgusunun tek başına yeterli bir kavram olmadığını anlaşılmış, yaşamın diğer alanlarındaki sosyal, kültürel, çevresel ve mekânsal kavramların dengeli şekilde birlikte incelenmesi gerekmektedir. 1970’li yıllar öncesinde “kalkınma” ve “gelişme”, kavramlarındaki değişme büyük ölçüde, ulusal gelirdeki artışlarla eşit görülmekteydi. Kalkınmada temel amaç, üretim ve istihdam yapısını, tarımdan ziyade, sanayi ve hizmetler sektörleri doğrultusunda dönüştürmek olunca, bu yaklaşıma denk olabilecek şekilde uygun olarak ülke refahındaki değişimlerin temel göstergesi olarak “fert başına ulusal gelir” kullanılmaya başlandı. Ancak, 1960’lardan sonra meydana gelen gelişmeler fert başına ulusal gelir kavramının yetersizliğini vurgularken, 1970’lerde kalkınma kavramının yeniden tanımlanması gerektiğine ihtiyaç duyulmuştur. Gelişmeyi; insani, sosyal, kültürel, çevresel ve mekânsal boyutlarıyla da tanımlama amacı taşıyan yeni yaklaşım, ekonomik büyüme kavramı yanında; yoksulluk, işsizlik, gelir dağılımı ve bölgesel dengesizliklerin de gelişme kavramlarının birlikte değerlendirilmesini gerektiğini öne çıkarmıştır.

Sosyo ekonomik gelişme farklılıkları 18. Yüzyılda endüstri devriminden sonra hızlanmış ikinci dünya savaşı sonrasında ise giderek artış göstermiştir. Bölgelerdeki sosyo ekonomik farklılıkların gelişmiş ülkelerde daha az yaşandığı gözlenirken, gelişmekte olan ülkelerde daha fazla yaşanmaktadır.

Dünya ekonomisinde yaşanan küreselleşme süreci, ulusal ve uluslararası düzeyde faaliyet gösteren özel sektöre etkisinin yanı sıra kamu yönetimine yönelik iktisadi değişimlerde de önemli yapısal değişikliklerle birlikte artarak hız kazanmıştır. Bu süreç, bölgesel ve yerel ekonomilerin de küresel ekonomide birer aktör olarak yer almasını sağlamış ve rekabeti artırmıştır. Yaşanan gelişmeler kalkınma anlayışını makro ölçekten mikro ölçeğe indirirken, kent ve bölgeler ise yeni kalkınma anlayışının sürükleyici unsurları durumuna getirmiştir.

Bu bağlamda Sosyo-ekonomik gelişmişlik farklılıklarından kaynaklanan yoğun göçler büyük bir nüfus baskısı oluşturarak ülke genelinde tüm mekânsal birimleri etkileyen yerleşme/ kentleşme sorunlarına neden olmaktadır. Ülke ölçeğinde yaygınlaşan yerleşme sorunları ise zamanla nitelik değiştirerek iktisadi, toplumsal ve siyasal yapılara da nüfuz etmekte ve derinleşmektedir. Göç dalgalarıyla büyüyen kentlerde beliren; eğitim ve sağlık hizmetleri yetersizlikleri ile arsa ve konut ihtiyacı, su, enerji, altyapı ve genel olarak belediye hizmetlerinin yetersizliği, trafik sıkışıklığı, kalabalık, gürültü ve çevre kirliliği gibi büyük kent sorunları, kamu yatırım ihtiyacını da artırarak, kamu maliyesine ek bir yük getirmektedir. Böylece; bölgelerarası sosyo-ekonomik gelişmişlik farklılıkları, sadece geri kalmış yöreler açısından değil, gelişmiş bölgeler açısından da düzeltilmesi gereken bir sorun olmaktadır. Az gelişmiş bölgelerin gelişme sürecine katılabilmesi, bu bölgelerin özel sektör yatırımları yanı sıra, kamu yatırım harcamaları ve kamunun diğer harcama kalemlerinden daha fazla pay almasını gerektirmektedir. Bundan dolayı da kamu maliyesinin güçlü ve sağlam bir yapıda olması gerekir. Etkin ve başarılı bölgesel gelişme politikaları, güçlü bir kamu maliyesi yapısıyla birebir ilişki içindedir. Bu çerçevede, hem kamu gelirlerinin artırılması, hem de gelirlerin rasyonel alanlarda kullanılması tüm ülkeler için önem kazanmaktadır (Dinçer, Özaslan, Kavasoglu 2003).

Kısaca; gelişme için milli gelirin artması yeterli değildir. Bunun yanında örneğin okuryazarlık oranının, sanayinin istihdam ve milli gelirdeki paylarının vb. artması gerekmektedir (Manisalı 1975: 2).

Gelişmekte olan ülkeler arasındaki Türkiye’de bölgeler ve iller arasında sosyo ekonomik gelişmişlik düzeyleri açısından önemli farklılıklar göstermektedir. Gelişmişlik, kişiler üzerindeki etkileriyle beraber, ilin sosyo-ekonomik yapısını bir bütün olarak da ilgilendirmektedir. İl genelinde birikimi ve gelişme potansiyelini yansıtan toplam büyüklüğü gösteren değerlerin, ortalama ya da fert başına düşen değerlere göre daha iyi bir gösterge olduğu ileri sürülebilir (Hacıhasanoğlu 1980). Diğer taraftan, illerin sahip oldukları nüfusunun, sosyo-ekonomik gelişmelerden etkilenen bir değişken olduğu göz ardı edilmemelidir.

Türkiye’de ülke idari bölünüşünde son yıllarda çok sayıda mekânsal değişiklik yapılmıştır. 1988 yılında 67 olan il sayısı, 1996 yılına ulaştığında 79 ile, 2003 yılında

ise 81 ile ulaşmıştır. Dolayısıyla, illerde zaman içerisinde meydana gelen gelişmeler ile değişikliklerin izlenmesini sağlayan ve gelişme dinamiğini yansıtan değişkenlerin, seri olarak elde edilmesi olanağı belirli ölçüde ortadan kalkmış bulunmaktadır (Dinçer, Özaslan, Kavasoglu 2003).

Türkiye'nin bölgeleri ve illeri sosyo-ekonomik gelişme süreci değişkenler bakımından tekdüze olmayıp, farklı sosyo-ekonomik özellikler göstermektedir. Bu farklılıkların anlaşılması ve incelenmesi üzerinde çalışılması politika yapıcılar için büyük önem arz etmektedir. Çünkü bölgesel eşitsizliklerin minimum düzeye indirilebilmesinin başlangıç noktası, bölgesel gelişme farklılıklarının tespit edilip üzerinden çalışılmasıdır.

Bölgelerarası sosyo-ekonomik gelişme farklılıkları, diğer ülkelerde olduğu gibi ülkemizde de görülmektedir. Araştırma sonuçlarından da görülebileceği gibi, genel olarak ülkenin Batı'sında yer alan Marmara, Ege, İç Anadolu ve Akdeniz Bölgeleri'ni görece olarak gelişmiş bölgeler olarak tanımlayabiliriz. Diğer yandan, özellikle Doğu Anadolu, Karadeniz'in dağlık bölgeleri ve Güneydoğu Anadolu'nun bazı yöreleri; gelir istihdam ve genel olarak sosyo ekonomik değişkenler bakımından, Türkiye ortalamalarının oldukça altında kalmaktadır.

Sosyal gelişme ise, gelir dağılımındaki büyük farkları, mümkün olduğu kadar gidermektedir. Eğer bir ekonomide kişilerin gelirleri arasında büyük farklar varsa sosyal gelişmeden söz edilemeyecektir. Böyle durumlarda bölge farklarına göre vergi politikaları uygulanmaktadır. Örneğin, devlet fakir yörelerden daha düşük, zengin yörelerden ise daha yüksek vergi almaktadır. Ülkemizde de görülen bu tür vergi uygulamaları sosyal kalkınmanın bir göstergesi olabilir (Özgüven 1988: 96-100).

Türkiye'de, Cumhuriyet'in başlangıcından beri dengeli bölgesel kalkınmaya yönelik bir politika izlenmekle beraber, bu çabalar daha çok planlı dönemle birlikte yoğunluk kazanmıştır. Planlı dönemde, bölgesel gelişmişlik farklılıkları ülkenin en önemli sorunlarından biri olarak görülerek, bölgelerarası dengeli kalkınma ilkesi, planlı kalkınmanın toplumsal hedefleri arasında yer almıştır (Dinçer, Özaslan, Kavasoglu 2003). Türkiye'de Planlı dönemden önce bölge planlaması yapılmış, bir imar aracı olarak ele alınan bölge planlamasında deneyimler ortaya konulmuş çıkan sonuçlara göre bölgelerin sorunları genel çerçevede ele alınması bölgesel olarak örgütlenilememenin

sebepleri arasında yer almıştır. 1960 yılından günümüze kadar uzanan planlı dönemde ise planlama genel olarak ele alınan bir bütün gibi düşünülen kavramlardan çıkarılarak; fiziksel, sosyal ve ekonomik değişkenlerinde içerildiği günümüz Türkiye ihtiyaçlarına uygun bütüncül bir yaklaşım çerçevesinde değerlendirilmeye başlanmıştır. Ayrıca, planlı dönemde bölgesel gelişmeye ve bölgeler arası gelişmişlik farklılıklarının en aza indirilmesi başlıca hedefler arasında olmuştur. Hazırlanan tüm kalkınma planlarında öncelikli hedef olan bölgeler arası dengesizliklerin en aza indirilmesi amacıyla pilot bölgelere noktasal girişimlerle bölgesel kurumsallaşmada farklılıklar giderilmeye çalışılmıştır.

Planlı dönemde yaşanan gelişmelerden biri, bölgeler arası gelişmişlik farklılıklarının somut verilerle belirlenmesi ve bölgesel gelişmeyi hızlandırmak amacı doğrultusunda kaynak tahsislerinin yapılmasıdır. Geri kalmış yörelere, bu yörelerden elde edilen bütçe gelirin çok üzerinde harcama yapmak, planlı dönemin yatırım ve harcama politikalarının temel özelliklerinden biri olmuştur (Dinçer, Özaslan, Kavasoglu 2003). Bu amaçla, tüm plan ve programlarda kamu yatırımlarının dağılımında, sosyo ekonomik yatırımlar bakımından geri kalmış bölgelere öncelik verilmiştir. Kamu yatırım politikaları yanında, özel sektöre bu bölgelerde yatırıma teşvik etmek amacıyla yapılan devlet yardımları ve uygulanan personel politikaları da, bölgeler arası dengesizliği gidermede kullanılan başlıca kamu teşviklerinden olmuştur.

Nitekim, dengesizlikleri azaltma doğrultusunda kamu sektörü, yalnızca iktisadi rasyonellerle değil, aynı zamanda toplumsal rasyonellerle de kaynak dağılımına yeniden yön verme ihtiyacı duymaktadır. Bu amaçla özel sektör yatırımlarını teşvik etmektedir (Dinçer, Özaslan, Kavasoglu 2003).

Ancak bu konuda geçmiş dönemlerde Türkiye ekonomisi, yapısal dönüşüm ve uluslararası piyasalara entegre olma bakımından büyük ilerlemeler kaydetmiş olmasına rağmen, bölgeler arası gelişmişlik farklılıkları özellikle büyük kentleşme (metropol) bugün de önemli bir sorun olarak varlığını devam ettirmektedir. Türkiye'nin uzun dönemde kaydettiği ekonomik büyüme performansı, bölgeler arası gelişmişlik farklılıklarının giderilmesinde beklenen olumlu etkiyi yaratamamış çünkü hala mevcutta var olan göç sorunu verimliliğin az olduğu bölgelerden büyümenin yaşam standartlarının fazla olduğu bölgeye doğru gerçekleşmektedir.

AB'ye katılım sürecinde önemli bir aşamaya geldiğimiz son zamanlarda, gerek AB'ye üye ülkeler ve Türkiye arasında, gerekse Türkiye'nin kendi içerisinde bölgeler arasında önemli ölçüde ekonomik ve sosyal gelişmişlik düzeyi bakımından bölgeler ve ülkeler arasında önemli farklılıkları bulunmaktadır. Bu nedenle iller ve bölgeler arası gelişmişlik farklılıklarının azaltılması yönünde oluşturulacak plan ve stratejilerin hazırlanması aşamasında, bölgelerin gelişmişlik yönlerinin sosyo- kültürel ve ekonomik değişkenler yardımıyla karşılaştırılması ve değerlendirilmesi, bu konuda çalışmalar yapılması, politika üretilmesi çok önemlidir.

Bu nedenle ölçülmek istenen sosyo ekonomik düzeyin ölçülmesinde kullanılan gelişme kavramının öncelikle tanımı ve tanım içerisinde hangi değişkenlerin kullanılması gerektiği ortaya konmalıdır. Gelişme kavramına insan ihtiyaçlarına ve değişen ihtiyaçlara göre zaman içinde farklı anlamlar yüklenmiştir. On dokuzuncu yüzyılda gelişme sadece ekonomik büyüme anlamına gelirken, milli ya da kişisel gelir, yaratılan katma değer, sanayi sektöründe üretim düzeyi ve çalışan hacmi gibi bazı göstergeler gelişme kavramını açıklamada kullanılmış ölçütlerden olmuştur. Yirminci yüzyılın başlarına gelindiğinde ise değişen dünya ile birlikte gelişme temel kavramı insanların sosyal refah düzeylerinde iyileşme, gelir yanında toplumların sahip oldukları fiziksel ve sosyal alt yapıyı da içine alır hale gelmiştir. Yirminci yüzyılın son çeyreğinde ise gelişme artık yaşam kalitesi ile ölçülmeye başlanmış, yaşam kalitesi özellikle nitelikli doğal, fiziksel, sosyal ve kültürel çevrenin varlığı ve tüketilme olanağına kavuşulması anlamında incelemeye değer kavramlar haline gelmiştir.

Gelir artışlarıyla ölçülen ekonomik büyüme, toplumsal gelişmişliği yeterince açıklayamayan, yalnızca talep genişlemesini açıklayabilen dar kapsamlı bir parametredir. Diğer yandan gelişme kavramı, fiziki kapasite büyüklüğü ve gelir artışı gibi iktisadi gelişmeler yanında, bunların gelir grupları ve bölgelerarası dağılımı ile sosyal ve kültürel birikimlerin göstergeleştirildiği bir bütünü, diğer bir ifadeyle “toplumsal gelişme” düzeyini ifade etmektedir (Dinçer, Özaslan, Kavasoglu 2003). Gelişme kavramı, ülke genelinde toplumsal, siyasal, kültürel ve benzeri değişkenlerdeki yapısal değişimleri kapsayan çok boyutlu içeriğe sahiptir. Bu anlayışla gelişme; ülkenin ekonomik, sosyal, siyasal ve kültürel yapılarındaki ilerlemeyi kapsamakta ve sonuçta ülkenin ulusal büyümesi ilerlemektedir. Belirtilen kapsamda bu çalışmaya konu olan sosyo-ekonomik gelişme olgusu da; kişi başına düşen milli gelirin artırılması şeklinde

özetlenebilecek iktisadi büyüme kavramıyla beraber, yapısal ve insani gelişmeyi içine alan ve kayıt edebildiğimiz bütün nicel ve nitel sosyal değişkenleri kapsamaktadır. Bir ülkede illerin ve bölgelerin gelişmişlik düzeylerinin ölçülmesinde sosyo ekonomik alanlarda seçilen gelişmişlik düzeylerini en iyi biçimde yansıtabilecek çok sayıda değişken kullanılmaktadır. Sosyo-ekonomik değişkenler üç temel başlık altında toplayabiliriz. Bunlar sosyal, ekonomik, coğrafi unsurların bileşenlerinden oluşan değişkenlerdir.

Sosyal göstergeler altında demografik, eğitim, sağlık, istihdam ve sosyal güvenlik göstergeleri; ekonomik göstergeler altında mali ve finansal, imalat sanayi, tarım, dış ticaret, enerji, konut, altyapı son zamanlarda geri dönüşüm göstergeleri gibi değişkenler; coğrafi göstergeler altında rakım, denize kıyısı olup olmama, iklim türü ve arazi yapısı bitki örtüsü, sıcaklık değerleri gibi göstergeler kullanılmaktadır.

Özetlemek gerekirse, gelişme için milli gelirin artması yeterli değildir. Bunun yanında örneğin okuryazarlık oranının, sanayinin istihdam ve milli gelirdeki paylarının vb. artması gerekmektedir (Manisalı 1975). Genel olarak ülkelerin veya coğrafi alanların milli gelirlerindeki artış oranları, büyüme hızı olarak kabul edilmektedir. Fakat bu oranın düşüklüğü ve yüksekliği ülkelerin gelişme veya kalkınma sürecinde hangi aşamada bulunduğu hakkında bilgi vermez. Ülkelerin veya coğrafi alanların kalkınmalarını tamamlayıp tamamlayamadıklarına karar verebilmek için birtakım sosyoekonomik faktörlere ve bu faktörlerdeki değişmelere bakmak ve bunları tümünü bir arada düşünerek bir değerlendirme yapmak gerekmektedir (Dinçer, Özaslan, Kavasoglu 2003).

2.2. Kümeleme Analizi

Kümeleme analizi son zamanlarda iş ve bilim dünyası için araştırma problemlerin de araştırmacıya uygun, işe yarayan bilgiler sunmaya çalışan çok değişkenli istatistiksel analizlerden biridir. Kümeleme analizi küme içerisinde bireyleri ya da nesnelere ayrıntılı olacak şekilde birbirlerine çok benzer şekilde benzerliklerine göre sınıflandırırken, küme arasında farklı olacak şekilde gruplamak amacıyla geliştirilmiş bir yöntemdir. Ayrıca kümeleme analizi çalışma yapılacak veri içerisindeki benzer örneklerin yakınlıklarına göre sınıflandırma yapabilen kullanışlı ve tercih edilen bir metottur.

Kümeleme analizinin temel amacı değişkenlerin sahip oldukları özellikleri baz alarak sınırlandırmaktır. Kümeleme analizi araştırmacı için politika üretimi sağlayabilecek ve üzerinde çalışılabilen bir teknik olması nedeniyle son yıllarda gündemde olan analiz yöntemlerinden biridir.

Bu yöntem özellikle bilim ve iş alanında birçok durumda uygulanabilen, en kolay yorumlanabilen ve en etkili olan yöntem olma özelliğini taşır. Bu nedenle hemen hemen tüm bilim alanlarında bu yöntemden yararlanılmaktadır (Özdamar 2004).

Ziraat başta olmak üzere sosyal bilimlerde, tıpta, tüm mühendislik bilimlerinde yaygın olarak uygulama imkânı bulabilen, kümeleme analizi çok değişkenli varyans analizi, lojistik regresyon analizi, çok boyutlu ölçekleme gibi çok değişkenli analizlerle birlikte çalışılan bir tekniktir.

Aslında Kümeleme analizinin ana teması, gözlenen birey veya nesnelere arasındaki benzerlikleri veya uzaklıkları/yakınlıkları tespit etmektir. Benzerlik, uzaklık kavramının tersi olup, büyük bir sayı çıktığında iki nesnenin birbirine yakın olduğunu, küçük bir sayı çıktığında ise iki nesnenin birbirinden uzak olduğunu göstermektedir (Nakip 2003).

Kümeleme analizi birbirine benzer olan bireylerin aynı gruplarda toplanmasını amaçlaması bakımından diskriminant analiziyle, birbirine benzer değişkenlerin aynı gruplarda toplanmasını amaçlaması nedeniyle de faktör analizi ile benzerlik göstermekte olup veri indirgeme özelliği vardır (Çakmak 1999: 188). Diğer çok değişkenli analiz yöntemlerinde normallik varsayımı ön planda tutulurken, kümeleme analizinde uzaklık değerlerinin normalliği yeterli görülmektedir (Tatlıdil 1992: 252).

Kümeleme Analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine göre belirli gruplar içinde toplayarak sınıflandırma yapmayı, birimlerin ortak özelliklerini ortaya koymayı ve bu sınıflar ile ilgili genel tanımlamalar yapmayı sağlayan bir yöntemdir. Burada amaç; gruplanmamış verileri benzerliklerine göre sınıflandırmak ve araştırmacıya uygun, işe yarar özetleyici bilgiler elde etmede yardımcı olmaktır (Tatlıdil 2002). Kümeleme analizi sonucunda küme sayısı bilinmemekte (bilinmesi halinde zaten kümeleme analizi tekniklerinin kullanılması anlamsızdır) sadece verilerin var olan durumuna ilişkin sonuçlar vermesi nedeniyle gelecekte kullanılabilmesi söz konusu olmaktadır.

Kümeleme Analizi, önceden belirlenen seçme kriterine göre birbirine çok benzeyen birey ya da nesnelere aynı küme içinde sınıflandırır. Analizin sonucunda oluşan kümelerin kendi içindeki homojenliği yüksek ve kümeler arası heterojenliği düşük olacaktır (Kalaycı 2005). Bir başka deyişle; kümeleme analizi herhangi bir veri setine ait nesnelere, benzer özelliklerine göre homojen gruplara ayrılmasıdır.

Bütün bu tanımları hepsini toplarsak; kümeleme analizi, X veri matrisindeki gruplanmamış nesnelere, bireylerin ya da değişkenlerin benzerliklerine göre alt gruplara sınıflandırılması olarak tanımlanabilen yöntemler topluluğudur.

Bu genel amaçların yanı sıra özel amaçlardan da söz edebiliriz:

- i. Gerçek tiplerin belirlenmesi
- ii. Model uydurmanın kolaylaştırılması
- iii. Gruplar için ön tahmin
- iv. Hipotezlerin testi
- v. Veri yapısının netleştirilmesi
- vi. Veri indirgenmesi
- vii. Aykırı değerlerin bulunması (Tatlıdil 2002).

Kümeleme analizi teknikleri uygulamasında veri tipleri kategorik (kalitatif), sayısal (kantitatif) ya da hem kategorik hem de sayısal verilere uygulanabilmektedir. Veriler fiziksel bazı süreçlerin göstergesi iken değişkenler nicel ya da nitel olarak farklı değerler olarak kayıt edilmiş olabilir.

Kümeleme analizi diğer kümelerdeki nesnelere göre benzer özellikler gösteren bireylerin oluşturduğu çalışmadır. Kümeleme analizinin gruplarında aranan benzerlik matematiksel benzerliktir. Kümeleme analizi, aynı gruptaki nesnelere birbirine benzer ya da ilişkili olması, farklı gruptakilerin birbirlerinden değişik ya da ilişkilerinin düzeylerinin olmamasıdır. Aynı gruptakilerin biriyle olan ilişkilerinin korelasyonları ne kadar yüksek ya da farklı gruptakilerin birbirleriyle ne kadar az ilişkili olması o kadar iyi kümenin yapıldığının sonuç göstergesi olmaktadır.

Kümeleme analizinin en kritik konusu küme sayısına karar vermektir. Araştırmacının küme sayısına karar vermede özneliği minimize etmesi gerekmektedir. Küme sayısının doğru olarak seçilmiş olması oluşturulacak olan kümelerin kalitesini belirlemektedir. Uygun küme sayısının belirlenmesi amacıyla birçok yöntem

geliştirilmiştir. Ancak günümüzde yayınlanan birçok bilimsel makalede küme sayısının belirlenmesinde net bir yöntem yoktur (Atbaş 2008).

2.2.1. Kümeleme Analizinde Kullanılan Benzerlik ve Uzaklık Ölçüleri

Kümeleme analizinde “Benzerlik” en temel kavramdır. Analizde ele alınan yöntemlerin tümü benzerlik ölçütlerine dayanır. Bireyler arasındaki benzerliğin ölçülmesinin farklı yolları bulunmaktadır. Benzerliği ölçmenin en önemli iki yolu, korelasyona dayalı ölçüler ve uzaklık ölçüleridir (Junhagen 2000).

Kümeleme analizinde değişkenleri ya da bireyleri benzerliklerine göre sınıflandırma yapabilmek için korelasyon ölçüleri ya da katsayıları karşılaştırma ölçütleri kullanılmaktadır. Veriler metrik ise uzaklık veya korelasyon, değilse karşılaştırma (ortaklık) ölçütleri kullanılmaktadır. Bir başka deyişle bireyleri sınıflandırmak için uzaklık ölçüleri, değişkenleri sınıflandırmak için korelasyon ölçüleri kullanılmaktadır.

Bir veri setindeki birimlerin (değişkenlerin) düzenli şekilde gruplandırılması ve sınıflandırılması birimlerin birbiriyle olan uzaklıkları kullanılarak gerçekleştirilmektedir. Birimler arasında gruplandırma yapmak benzerlik gösteren birimleri bir arada toplamak ve incelemek değişkenler arasındaki neden-sonuç ilişkilerini açıklamayı daha da kolaylaştıracaktır. [0, 1] aralığı dışında değerler alabilen ve benzerlik ya da benzemezliğin ölçüsü olan bu değerlere “Uzaklık” değeri adı verilir (Tatlidil 1996).

Kümeleme analizinde kullanılan uzaklık ölçüleri şöyle tanımlanmaktadır.

2.2.1.1. Öklid (Euclidean) Uzaklığı

Uygulamada en çok kullanılan uzaklık ölçüsüdür. Çok boyutlu uzayda geometrik mesafedir. İki nokta arasındaki Pisagor teoreminin uygulamasıdır. Hesaplaması aşağıdaki gibi yapılmaktadır.

$$d(x_i - y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

Öklid uzaklık bağlantısı standartlaştırılmış değil, ham veriden hesaplanır (Erilli 2009). Bu metot kümeleme analizinde sınıflandırmadaki vektörlerin birbirine ne kadar benzediğinin ya da benzemediğinin avantajını sağlamaktadır.

2.2.2.2. Karesi Alınmış Öklid Uzaklığı

Uzak nesnelere daha büyük ağırlık verebilmek için Öklid uzaklığının karesi alınarak hesaplama yapılmıştır (Erilli 2009). Aşağıdaki gibi hesaplanmıştır.

$$d(x_i - y_i) = \sum_{i=1}^n (x_i - y_i)^2 \quad (2.2)$$

2.2.2.3. Manhattan (City-Block) Uzaklığı

Hesaplama yöntemi olarak; karşılıklı olarak koordinatların ortalama farklılığıdır. Birçok sonuçta, Öklid uzaklığıyla benzer sonuçlar karşımıza çıkmaktadır. Manhattan uzaklığı aşağıdaki gibi hesaplanmaktadır. Matematiksel hesaplama yöntemi olarak iki koordinat noktası arasındaki ortalama farklılıktır.

$$d(x_i - y_i) = \sum_{i=1}^n |x_i - y_i| \quad (2.3)$$

2.2.2.4. Chebychev Uzaklığı

Chebychev uzaklığı, iki gözlem vektörü arasındaki maksimum mesafeyi ifade etmektedir. Terim büyüklüğünün farklı olduğu veri setlerinde, uzaklığın hesaplanmasında terim büyüklüğünün fazla olması dışındaki diğer boyutlarda değişik özellikler dikkate alınmayacağından bu uzaklığın kullanılması tavsiye edilmemektedir. Chebychev uzaklığı şöyle hesaplanmaktadır.

$$d(x_i - y_i) = \text{Max}_i |x_i - y_i| = \lim_{k \rightarrow \infty} \left(\sum_{i=1}^n |x_i - y_i|^k \right)^{1/k} \quad (2.4)$$

2.2.2.5. Minkowski Uzaklığı

Minkowski uzaklığı, kümeleme analizinde nicel verilerin benzerliğinin hesaplanması için kullanılan metrik bir uzaklık ölçüsüdür. Veri seti içerisindeki kompakt ve birbirlerinden iyi ayrılmış kümeleri belirlemede Minkowski uzaklığı iyi sonuçlar vermektedir (Mao, Jain 1996). Aşağıdaki gibi hesaplanmaktadır.

$$d(x_i - y_i) = \left(\sum_{i=1}^n |x_i - y_i|^k \right)^{1/k}, k \geq 1 \quad (2.5)$$

2.2.2.6. Mahalanobis Uzaklığı

Mahalanobis uzaklığı iki nokta arasındaki uzaklığı hesaplarırken değişkenler arasındaki korelasyon ve kovaryans ilişkisini hesaba katan bir uzaklık ölçüsü olarak kullanılmıştır. Hesaplaması aşağıdaki gibi yapılmaktadır.

$$d(x_i - y_i) = (x_i - x_j)' S^{-1} (x_i - x_j) \quad (2.6)$$

2.2.2.7. Canberra Uzaklığı

Ölçeğe duyarlı olmayan bu uzaklık ölçüsü veri matrisindeki tüm değişkenlerin pozitif olduğunda kullanılması tavsiye edilmektedir. Aşağıdaki şekilde hesaplama yapılmaktadır.

$$d(x_i - y_i) = \sum_{i=1}^m \left| \frac{x_i - y_i}{x_i + y_i} \right| \quad (2.7)$$

2.2.2.8. Hotelling T^2 Uzaklığı

İki birimin (kümenin) ortalama vektörlerinin karşılaştırıldığında Hotelling T^2 kullanılır. Aşağıdaki gibi hesaplanır.

$$d(x_i, x_j) = T^2 = \frac{n_1 n_2}{n} (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j) \quad (2.8)$$

Buradaki S $p \times p$ kovaryans matrisini göstermektedir.

Kümeleme analizinde kullanılan bazı korelasyon ölçüleri aşağıda açıklanacaktır:

2.2.2.9. Biserial Korelasyon Ölçüsü

$$I_b = \frac{\overline{x_p - x_q} \cdot p \cdot q}{S_i \cdot y} \quad (2.9)$$

Şeklinde hesaplanmıştır.

2.2.2.10. Pearson Korelasyon Ölçüsü

Dağılım normal olduğunda iki sürekli değişken arasındaki matematiksel ilişkinin belirlenmesinde kullanılır. “r” ile gösterilmektedir.

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)} \sqrt{\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}} \text{ şeklinde gösterilmektedir.} \quad (2.10)$$

2.2.2.11. Spearman Sıra Korelasyon Katsayısı

Pearson korelasyon ölçüsünü özel bir halidir ve p ile gösterilmektedir.

$$p = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2.11)$$

2.3. Kümeleme Analizi Yöntemleri

Kümeleme analizi, birçok kaynakta alt bölümlere ayrılmış olsa da temelde hangi alt bölümlere ayrılacağı kesin çizgileri yoktur (Grabmeir, Rudolph 2002; Hartigan 1975). İlk olarak 1939 yılında literatüre giren Kümeleme Analizinde; 1950’li yıllarda verilerin hiyerarşik olarak gruplanmasına çalışılmış, 1960’lı yıllarda ise veriyi hiyerarşik olmayan alt gruplara bölerek kümeleme teknikleri aşamalı olarak geliştirilmiştir. En çok bilinen ayırım hiyerarşik (aşamalı) ve hiyerarşik olmayan (aşamalı olmayan) kümelemedir. Bu ayırım kümeleme yönteminin seçimine göre yapılmaktadır. Bunun dışında son yıllarda iki aşamalı (Two step cluster) gibi yöntemler de kullanılmaya başlanmıştır.

Arařtırmacı hangi benzerlik/uzaklık ölçüsünü kullanacağına karar verdikten sonra, kümeleme işleminin nasıl olacağına karar vermek zorundadır. Birimlerin benzerliklerine göre kümelere dâhil edilmesinde kullanılabilir çeşitli yaklaşımlar vardır. Bu yaklaşımlardan biri, en çok benzer iki birimi aynı gruba atamakla başlayıp tüm birimlerin aynı gruba atanması ile biten hiyerarşik bir yaklaşımdır. Bir başka yaklaşım ise tüm verilerin ortalama değerlerine en yakın değerlere sahip birimlerin aynı kümeye atanmasını esas alan yaklaşımdır. En çok kullanılan bu iki yaklaşım dışında diğer yaklaşımlar da mevcuttur. Tüm yaklaşımlarda en önemli ölçüt, kümeler arası farklar ile kümeler içi benzerliklerin maksimum olmasını sağlamaktır. En çok kullanılan kümeleme algoritmaları hiyerarşik ve hiyerarşik olmayan kümeleme adı altında iki kategoride toplanmaktadır (Blashfield, Aldenferder 1978).

Literatürde klasik kümeleme analizinin yanı sıra son yıllarda alternatif yöntemler de uygulanmaktadır. Bulanık kümeleme, soft set kümeleme, rough set kümeleme gibi yeni yöntemler, kümeleme literatüründe çalışma alanları bulmaktadır.

2.3.1. Hiyerarşik (yığılmalı) Kümeleme Yöntemleri

Hiyerarşik kümeleme yöntemleri özellikle küçük örneklemeler (tipik olarak $n < 250$) için uygundur (Yalçın 2013). Hiyerarşik kümeleme yöntemleri birimleri birbirleri ile değişik aşamalarda bir araya getirerek ardışık biçimde kümeler belirlemeyi ve bu kümelere girecek elemanların hangi uzaklık (ya da benzerlik) seviyesinde o kümenin elemanı olduğunu belirleyen tekniksel analizdir. Hiyerarşik tekniklerin ağaç diyagramları ile gösterilen sonuçlarına dendogram denir (Lorr 1983).

Kümeleme sürecinin başlangıcında her birey bir kümedir. Süreç sonunda ise tüm bireyler bir kümede toplanır. Bu sürecin işleyişi aşağıda maddeler halinde sıralanmıştır (Tatlıdil 1996: 334).

1. Öncelikle n adet birey, n adet küme olmak üzere işleme başlanır.
2. En yakın iki küme d_{ij} (değeri en küçük olan) birleştirilir.
3. Küme sayısı bir indirgenerek yinelenmiş uzaklıklar matrisi bulunur.
4. 2 ve 3 numaralı adımlar $n-1$ kez tekrarlanır.

Hiyerarşik kümelemenin avantajlarını şu şekilde sıralayabiliriz.

- i. Düşümlerin seviyesine ilişkin esneklik özelliğinin olması
- ii. Benzerlik ve uzaklık biçimlerini ele almanın kolaylığı
- iii. Çeşitli niteliklere uygulanabilme özelliğinin olması

Hiyerarşik kümelemenin dezavantajlarını ise şu şekilde sıralayabiliriz.

- i. Bitirme kriterinin belirsizliği
- ii. Birçok hiyerarşik kümeleme algoritmasının, geliştirilme amaçlarına göre arada bulunan kümelere tekrar ulaşılmasına izin vermez (Güler 2006).

Hiyerarşik kümeleme yöntemleri toplanmış hiyerarşik kümeleme ve bölen hiyerarşik kümeleme olarak ikiye ayrılır.

2.3.1.1. Toplanmış Hiyerarşik Kümeleme

Toplama teknikleri $\left\{\frac{1}{2}[n(n-1)]\right\}$ olası gözlem çifti arasındaki bir benzerlik veya uzaklık matrisinin hesaplanması ile başlar. Başlangıçta her gözlem bir kümedir. Benzerlik veya uzaklık matrisine göre en yakın iki küme birleştirilir. Daha sonra küme sayısı bir indirgeyerek benzerlik matrisi tekrar oluşturulur ve n birim aşamalı olarak sırasıyla $n, (n-1), (n-2), \dots, (n-r), \dots, 3, 2, 1$ kümeye yerleştirilir (Everitt 1974). Burada hiyerarşik kümeleme yöntemleri içerisinde en çok bilinen tek bağlantı, tam bağlantı, ortalama bağlantı ve ward yöntemleri kısaca açıklanacaktır.

2.3.1.1.1. Tek Bağlantı Yöntemi

En yakın komşuluk olarakta bilinen bu yöntem en kısa mesafe esasına dayanmaktadır. İlk olarak Florek ve diğerleri (1951)ve daha sonra sırasıyla Sneath (1957) ve Johnson(1967) tarafından uygulanmış olup uzaklık veya benzerlik matrisinden yararlanılarak birbirine en yakın iki gözlem veya küme birleştirilmekte ve birleştirme işlemi yinelenmektedir (Şentürk, 1995; Fırat, 1995).

Birleştirmede benzerlik türü ölçümlerde;

$$s_k(i, j) = \max(s_{ki}, s_{kj}) \quad (2.12)$$

Uzaklık türü ölçümlerde ise;

$$d_k(i, j) = \min(d_{ki}, d_{kj}) \text{ kriterleri göz önüne alınır.} \quad (2.13)$$

En yakın komşuluk yöntemi hiyerarşik kümeleme yönteminin en basit yöntemidir. Tek bağlantı yönteminin yapılış aşaması aşağıdaki gibidir.

- i. X veri matrisinin d Öklid uzaklık matrisi hesaplanır.
- ii. Eğer istenirse d matrisinden benzerlik matrisi hesaplanır.
- iii. Benzerlik veya d matrisinde en küçük değerli birimler hiyerarşik olarak birbirleri ile birleştirilir ve (ij) kümesi oluşturulur. i. ve j. kümeleri birbirleri ile birleştirildikten sonra D veya benzerlik matrisinde j. kümeye ilişkin satır silinir ve (ij) kümesinin uzaklığı i. kümenin uzaklığı olarak kalır.

Tüm kümeler birbirleri ile birleştirilinceye kadar 3 numaralı satırdaki işlemler tekrarlanır (Özdamar 2002).

2.3.1.1.2. Tam Bağlantı Yöntemi

Literatürde en uzak komşuluk yöntemi olarak ta bilinmektedir. Johnson tarafından önerilen bu yöntem, tek bağlantı yöntemine çok benzemektedir ancak formül olarak iki küme arasındaki uzaklık değeri olarak kümedeki eleman çiftleri arasındaki uzaklığın maksimum değeri alınır. Tek bağlantı yöntemine çok benzemektedir ancak burada iki küme arasındaki uzaklık olarak her kümedeki eleman çiftleri arasındaki uzaklığın en büyüğü alınır.

Benzerlikte;

$$s_k(i, j) = \min(s_{ki}, s_{kj}) \quad (2.14)$$

Uzaklık türü ölçümlerde ise;

$$d_k(i, j) = \max(d_{ki}, d_{kj}) \text{ kriteri esas alınmaktadır.} \quad (2.15)$$

Tam Bağlantı Kümeleme Yöntemi için aşağıdaki işlem sırası izlenir.

- i. Her birey, kendi kümesine yerleştirilir. Tüm sırasız birey çiftleri için bireyler arasındaki uzaklığın listesi oluşturulur ve bu liste küçükten büyüğe doğru sıralanır.

- ii. Birbirlerine en uzak (uzaklık değeri en büyük) olan birey ya da kümeler birleştirilir.
- iii. Bütün bireyler birleştirilen bir kümenin üyesi ise işlem bitirilir. Aksi takdirde 2. maddeye geri dönülür.

2.3.1.1.3. Ortalama Bağlantı Yöntemi

Sokal ve Michener tarafından önerilmiş olan bir tekniktir. Bu yöntem iki küme arasındaki fark, bir küme arasındaki eleman çiftleri ile diğer bir kümedeki eleman çiftleri arasındaki ortalama fark olarak alınır (Everitt 1981). Ortalama bağlantı yöntemi tam bağlantı yöntemine benzer dendogramlar çizmektedir. Ancak her bir yöntemde uzaklık farklı tanımlandığı için birleştirmeler farklı seviyelerde ortaya çıkabilmektedir (Fırat 1995).

Bir birimin m. küme olarak hangi birim ya da kümelerle birleştirileceği, birimlerin yeni oluşturulan uzaklıkları dikkate alınarak belirlenir. m. kümenin daha önce oluşan k. ve l. kümelerden hangisi ile birleşerek oluşacağını belirlemek için j. küme ile k. ve l. kümelerin uzaklıklarına bakılır. Bu uzaklıklar k ve l kümelerinin eleman sayısı ile çarpılarak ağırlıklandırılır. Elde edilen toplam yeni oluşacak m. küme eleman sayısına bölünür.

m. kümenin j. küme ile olan uzaklığı şu şekildedir.

$$d_{mj} = \frac{(N_k d_{kj} + N_l d_{lj})}{N_m} \quad (2.16)$$

Birbirlerine en çok benzeyen çiftin bulunmasında grup içi benzerlik ortalaması k ve l kümelerine ait çiftlerin benzerlik ölçülerinde ve birim sayılarından yararlanılarak hesaplanır (Özdamar 2002).

Kullanılmakta olan paket programlarının çoğu tek ve tam bağlantı tekniklerini kullanmaktadır. Tek bağlantı yöntemi sağlıklı sonuçlar vermesi dolayısıyla tercih edilmekte ancak işlemlerin uzun sürmesi açısından da sakıncalıdır. Tam bağlantı tekniği ise aynı küme içerisindeki bireylerin uzaklıklarının belli bir değerden küçük olması durumunda tüm kümelerin sağlıklı oluşturulmasını garanti edememektedir. Ortalama bağlantı tekniği, bu iki uç teknik arasında sonuçlar vermesi nedeniyle alternatif bir yöntem olarak önerilmektedir (Tatlıdil 2002).

2.3.1.1.4. Ward's Bağlantı Yöntemi

En küçük varyans yöntemi olarak da bilinen Ward's bağlantı tekniği, diğer kümeleme yöntemleri gibi kümeler arasındaki uzaklıkları hesaplamak yerine, küme içindeki varyansı minimize, homojenliği maksimum kılacak kümeler oluşturur. Bu amaçla hata kareler toplamı formülünden yararlanır.

$$ESS = \sum_{i=1}^n X_i^2 - 1/n \left(\sum_{i=1}^n X_i \right)^2 \quad (2.17)$$

Burada X_i , i nci gözlemin skorudur (Aldenderfer, Blashfield 1984: 43).

2.3.1.2. Bölünmeli Kümeleme Yöntemi

Bölünmeli Kümeleme Yöntemi, başlangıçta tüm birimlerin bir küme oluşturduğunu kabul ederek birimleri hiyerarşik n birimi sırasıyla 1, 2, 3,..., n-1, n kümeyle ayırmayı amaçlayan bir yöntemdir. Toplanmış hiyerarşik kümeleme yönteminin tersidir. Toplanmış yöntemle ilişkin sonuçlardan bölen yöntemle ilişkin sonuçlar elde edilebilir (Yıldırım 2007).

2.3.2. Hiyerarşik Olmayan Kümeleme Yöntemi

Araştırmacının küme sayısı hakkında ön bilgisinin olduğu durumlarda kullanılır. Araştırmacı anlamlı olabilecek küme sayısına karar vermişse uzun zaman alan hiyerarşik teknikler yerine hiyerarşik olmayan yöntemleri tercih edebilmektedir. Ayrıca, bu yöntemlerin kurumsal dayanaklarının daha güçlü olması diğer bir tercih sebebidir (Selanik 2007: 13).

Hiyerarşik Olmayan Kümeleme yöntemleri kümeleri oluştururken çoğunlukla yerel ya da tümel olarak tanımlanmış olan bir benzerlik kriterini optimize etmeye çalışır. Bu fonksiyonun muhtemel en iyi değerini bulmak için tüm kayıtlar üzerinde çeşitli kombinasyonlar denemek mümkün değildir. Bunun yerine pratik alanda birkaç başlangıç durumu için algoritma işletilir ve sonuç kümeleri de bu önceki çalışmaların sonuçlarından elde edilerek oluşturulur (Ünler 2006). Bu yöntemde değişkenlerin kümelere ayrılması rastgele yapılmaktadır.

Hiyerarşik olmayan yöntemlerde birimlerin uygun oldukları kümelere toplanmaları ve n birimin k sayıda kümeyle parçalanması hedeflenmektedir. Birimlerin

ayrılacakları küme sayısı belirlendikten sonra, kümeler için belirlenen küme belirleme kriterlerine göre birimlerin hangi kümelere girebileceklerine karar verilir ve atama işlemi gerçekleştirilir (Özdamar 1999: 303). Hiyerarşik olmayan kümeleme analizi yöntemlerinden en çok kullanılan iki tanesi k-ortalamlar ve en çok olabilirlik yöntemidir.

2.3.2.1. k-Ortalamlar Tekniği

Bu teknik 1965 yılında Forgy ve 1967 yılında MacQueen tarafından ortaya atılan hata kareler azaltma ilkesinden sonra, 1975’de Hartigan tarafından geliştirilmiştir. Bu teknik uzun yıllar boyunca kullanılan en önemli hiyerarşik olmayan kümeleme analizi tekniklerinden olmuştur.

Kümeleme analizlerinde uzaklık olarak Öklid uzaklığı kullanıldığından; $\|x_{ij} - v_j\|^2, 1 \leq j \leq k$ hesaplaması gereğince bireyler en yakın olan kümeye sınıflandırılmaktadır. Burada x_{ij} , j . kümedeki i . nesneyi; v_j ise j . küme merkezini ifade etmektedir (Cebeci Kayaalp Yıldız, 2015). Kullanılan uygun uzaklık ölçüsü şu şekilde hesaplanmaktadır.

$$j(x; v) = \sum_{j=1}^k \sum_{x \in c_j} \|x_j^{(i)} - v_j\|^2 \quad (2.18)$$

K-Ortalamlar yöntemi, ilk önce n adet nesneden rasgele k adet nesne seçer ve bu nesnelerin her biri, bir kümenin merkezini veya orta noktasını temsil eder. Geriye kalan nesnelere her biri kendisine en yakın olan küme merkezine göre kümelere dağılırlar. Yani bir nesne hangi kümenin merkezine yakınsa o kümeye yerleşir. Ardından her küme için ortalamlar hesaplanır ve hesaplanan bu değer o kümenin yeni merkezi olur. Bu işlem tüm nesnelere kümelere yerleşinceye kadar devam eder. K-Ortalamlar Yöntemi, sadece kümenin ortalaması tanımlanabildiği durumlarda kullanılabilir. Kullanıcıların k değerini, yani oluşacak küme sayısını belirtme gerekliliği bir dezavantaj olarak görülebilir. Burada esas dezavantajlı durum ise dışarıda kalan aykırı gözlemler olarak adlandırılan nesnelere karşı olan duyarlılıktır (Han 2000).

Zayıf bir optimuma yakınsama eğilimi; ölçeklemeye ve öteki dönüşümlere duyarlılık; ölçüm hatalarına ve aykırı değerlere duyarlılık; yanlılık (yani yanlı parametre değerleri yakınsama) gibi K-ortalama yönteminin ana problemlerini

listelemişlerdir. Yine de, en önemli problem K' nın uygun bir değerini hesaplamaktır (Estivill – Castro, Yang 2004).

2.3.2.2. En Çok Olabilirlik Yöntemi

Diskriminant analizinde de kullanılan en çok olabilirlik yönteminde her bir gözlem (her bir x_i gözlem vektörünün bağımsız ve aynı olasılık yoğunluk fonksiyonuna sahip olduğu varsayımı altında) büyük olabilirlik değerini verecek biçimde daha önceden belirlenen kümelere atanır.

Kuramsal dayanağı güçlü olmakla birlikte en çok olabilirlik tekniği uzun makine zamanı gerektirmesi nedeniyle yaygın olarak kullanılmamaktadır (Tatlıdil 1996).

2.4. Küme Geçerlilik İndeksi

Kümeleme analizi, veri setinin altında yatan doğal grupları ortaya koyan ve birçok alanda yaygın olarak kullanım alanı bulabilen çok değişkenli istatistiksel bir yöntemdir. Kümeleme analizinde birbirine benzeyen değişkenleri belirleyerek kümeleme analizi yapılmaktadır. Üzerinde çalışmaya başladığımız veri seti hakkında sonuç elde edebilmek için homojen ve uygun kümeleme ihtiyacı vardır. Kümeleme analizinde, anlamlı sonuçlara ulaşmada optimal küme sayısının belirlenmesi birçok araştırmacının sıklıkla karşılaştığı önemli sorunlardan biridir. Çünkü kümeleme analizinde küme sayısı önceden bilinmez. Ancak kümeleme analizi tekniklerinden bazılarında (k-ortalamlar) küme sayısı hakkında önceden bilginin olması gerekmektedir.

Bu yüzden farklı kümeleme analizlerinde çıkan farklı sınıflandırma sonuçlarından ve çıkan sınıflandırmanın kalitesinin ve doğruluğunun sorgulanması önemli aşamalardan biridir. Bu bağlamda yapılan çalışma sonucunda oluşan küme sayısının değerlendirilmesinde ve küme sayısının en optimal küme sayısı olduğuna karar verilmesinde küme geçerlilik indeksleri kullanılmaktadır. Ayrıca bazı karmaşık yapılar içeren verilerde, küme üyeliklerindeki kararsızlıklar nedeniyle küme geçerlilik indekslerindeki sonuçlardan birbirleriyle çelişki oluşturabilecek sonuçlar verebilmektedir. Bu amaçla 21 farklı küme geçerlilik indeksi kullanılmaktadır. 21 küme geçerlilik indeksi (Krzanowski-Lai (KL), Calinski-Harabazs (CH), Dunn, Silhouette, Gamma, Tau, PtBiserial, CCC, Davies-Bouldin (DB), Gplus, Cindex, McClain,

Sdindex, Sdbw, Duda, PseudoT2, Ratkowsky, Gap, Ball, Dindex ve Hubert) veri gruplarında farklı uzaklık ölçüleri kullanılarak kümelerin geçerliliklerinin karşılaştırması yapılabilir.

Küme geçerlilik indekslerinin farklı kümeleme algoritmalarına göre farklı performans göstermeleri muhtemel bir durumdur. Ancak küme geçerlilik indeksinin herhangi bir kümeleme yöntemine göre veri setinin en uygun küme sayısının kararının doğru verilmesi o indeksin optimal küme sayısını belirlemedeki başarısını göstermektedir.

2.4.1. Küme Sayısının Belirlenmesi ve Küme Geçerlilik İndeksleri

Küme sayısına karar vermede ve oluşan küme sayısının kalitesine karar verme, gerek klasik kümeleme analizinde gerekse bulanık kümeleme analizinde en kritik konulardan biridir. Bu amaçla küme geçerliliği (cluster validation) teknikleri büyük önem taşımaktadır. Çalışmacının kümeleme analizindeki, küme sayısına karar vermede öznelliği minimuma indirmesi gerekmektedir. Bu konuda 1970'li yıllarda geliştirilmiş ve çokta güvenilir sonuçlar vermeyen bazı testlerden yararlanılmaktadır. Küme sayısının belirlenmesindeki önerilen yaklaşımlardan en çok bilinen ve en pratik olanı;

$$k = (n/2)^{1/2} \text{ şeklinde hesaplanmaktadır.} \quad (2.19)$$

Buradaki k küme sayısı n ise birim sayısını göstermektedir. Bu eşitliğin büyük örneklemelerden daha çok küçük örneklemelerde kullanılması tavsiye edilmektedir. Büyük örneklem hacminde sağlıklı sonuçlara ulaşılması zorlaşmaktadır (Everitt 1974).

Diğer bir yöntem ise 1971 yılında Mariot tarafından önerilmiştir. Bu yüzden M harfi ile gösterilir.

$M = k^2 |W|$ biçimindedir (Marriot 1971). Burada W grup içi kareler toplamı matrisidir ve

$$W = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)(X_{ij} - \bar{X}_j) \quad (2.20)$$

n_j ; j. kümedeki birim sayısı

k ; küme sayısı

X_{ij} ; j. kümedeki i. birim değerleri

\bar{X}_j ; j. kümenin örneklem ortalama vektörü

biçiminde hesaplanır. En küçük M değerini veren küme sayısı gerçek küme sayısı olarak değerlendirilmektedir. Everitt (1974) farklı ölçümlerle ilgili yaptığı çalışmada, k^2W ifadesinin diğer yöntemlere göre daha iyi sonuç verdiğini ortaya koymuştur.

Küme sayısına karar vermede kullanılan bir diğer yaklaşım ise Lewis ve Thomas tarafından literatüre katılmıştır. Lewis ve Thomas' a göre küme sayısına karar vermek için iki kriter vardır. Kümeler, toplam varyansın %80'ini açıklamalı ve varyansta %5'e kadarlık bir artış durumunda yeni bir küme ilave edilebilmektedir (Ruiz 1998).

Küme sayısının belirlenmesinde kullanılan ölçütlerden Wilk's Lamda ölçütü, birim sayısı 30'un üzerinde olduğunda duyarlılığı diğer ölçütlere göre en yüksek düzeye ulaşmaktadır (Dinçer, Özdamar 1992). Everitt (1979) ve Anderberg (1973) gruplar arası varyansın grup içine göre maksimum olduğu durumları belirleme açısından Wilk's Lamda ölçütünün çok duyarlı olduğunu belirtmişlerdir.

Kümeleme analizi yöntemleri, oluşan kümelenmeleri doğrulamak için çok değişkenli bir yöntem olan Temel Bileşenler Analizinden faydalanabilir. Temel Bileşenler Analizi yönteminin kullanılmasıyla oluşturulan Temel bileşen skoru grafiğinden yorumlar yapılarak, bu grafik yardımıyla kümeleme analizi sonuçlarının doğruluğu denetlenebilir (Atbaş 2008) ve yorumu konusunda fikir verebilmektedir.

Küme sayısına karar verme teknikleri literatüre dahil edilmiş olsa bile araştırmacı çalışanın sonucunda küme sayısını bilgi düzeyi, mesleki tecrübesi sonuçların anlamlı olup olmamasıyla ve farklı hipotez testleriyle kontrol sağlayabilir. Küme sayısına karar verdikten sonra araştırmacı çalışmanın kalitesini test etmek amacıyla küme geçerlilik indekslerini kullanarak değerlendirilmelidir.

2.4.2. Klasik Kümeleme Analizinde Küme geçerlilik indeksleri

Kümeleme analizinde genellikle doğru ve en uygun küme sayısı çoğunlukla bilinmez. Kümeleme analizinde küme sayısını bulduktan sonra kaliteli ve doğru analiz için küme geçerlilik tekniklerine ihtiyaç vardır.

Kümeleme analizinde aynı küme içindeki nesnelere birbirine benzerlikleri fazla olması istenirken, diğer kümeler arasındaki benzerliğin ise minimum düzeyde olması gerekmektedir. Diğer bir ifadeyle küme içindeki varyansın düşük, kümeler arasındaki varyansın ise yüksek olması gerekmektedir.

Küme geçerliliği teknikleri küme içi ve kümeler arası değişimlerin farklı açılardan değerlendirilmesi ile birbirinden ayrılmaktadır (Bolshakova, Azuaje 2003). Ayrıca çok değişkenli hipotezlerin testlerinde kullanılan wilk's lamda ölçütünde küme geçerlilik indeksi küme sayısını belirlemekteki önemli tekniklerde biri olarak karşımıza çıkmaktadır.

2.4.2.1.Silhouette İndeksi

Rousseeuw (1987), yılında her birimin atandığı kümeye uyguladığını test amacıyla Silhouette indeksi önermiştir.

$a(i)$; i . birimin kendi kümesindeki tüm noktalara olan ortalama uzaklıklarını (benzerliğini)

$b(i)$; i . birimin diğer kümelerdeki tüm noktalara olan ortalama uzaklıkların minimumunu göstermek üzere,

Buradan i . birim Silhouette indeksi için;

$$sil(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.21)$$

olarak tanımlanır. Eğer $sil(i)$ değeri 1'e yaklaşırsa i . birimin atandığı kümeye daha uyduğu, $sil(i)$ değeri 0'e yaklaşırsa veya negatif olursa i . birimin atandığı kümeye uygun olmadığı sonucuna varılır. Negatif değerler yalnızca bir birim en uygun kümesine atanmadığında ortaya çıkar. Tüm kümelemenin kalitesi (geçerliliği) için bir doğal bir ölçü olan ortalama Silhouette değeri;

$$sil(c) = \frac{1}{n} \sum_{s_i \in s} sil(c) \quad (2.22)$$

Şeklinde verilmektedir. Bu indeks değerine göre, maksimum ortalama Silhouette değerine ulaşılan küme sayısı uygun küme sayısı olarak kabul edilir (Rousseeuw 1987).

2.4.2.2. Calinski ve Harabazs İndeksi

Calinski ve Harabazs (1974) yılında, kümeleme analizinde küme sayısı ve oluşturulan kümelerin küme geçerliliği kalitesini test etmek için önerilmiştir.

$$CH(k) = \frac{BSS(k)}{k-1} / \frac{WSS(k)}{n-k} \quad (2.24)$$

Şeklinde hesaplanır(Calinski, Harabasz 1974). Burada;

$W(k)$ küme içi uzaklıkların kareleri toplamını ifade eder ve

$$WSS(k) = \frac{1}{2} \sum_{j=1}^k \sum_{j \in c_i} d(i, j) \quad (2.25)$$

$B(k)$ ise kümeler arası uzaklıkların kareleri toplamıdır ve

$$BSS(k) = \frac{1}{2} \sum_{j=1}^k \sum_{\substack{j \in c_i \\ j \notin c_i}} d(i, j) \quad (2.26)$$

$WSS(k)$ ve $BSS(k)$ hesaplanırken karesel uzaklık ölçüleri kullanılmaktadır. Bu kritere göre, CH indeks değerini maksimum yapan küme sayısı, uygun küme sayısı olarak kabul edilir (Calinski, Harabasz 1974).

2.4.2.3. Krzanowski ve Lai indeksi

Krzanowski and Lai (1985) yılında, küme içi kareler toplamı (WSS) değerinin azalışı tabanında bir indeks önermişlerdir. Öncelikle Krzanowski ve Lai

$$DIFF(k) = k^{-2/p} WSS(K-1) - k^{-2/p} WSS(k) \quad (2.27)$$

Şeklinde tanımlanmıştır ve bu istatistiğe bağlı olarak,

$$KL(k) = \left| \frac{DIFF(k)}{DIFF(k+1)} \right| \quad (2.28)$$

İndeksini hesaplamışlardır. $DIFF(k) = k^{-2/p} WSS(k)$ alınarak da KL indeksi hesaplanmıştır. KL indeks değerinin hesaplanmasında amaç KL indeks değerinin maksimum olması amaçlanmaktadır. k , KL indeksi için uygun küme sayısı olsun. KL indeksinde uygun küme sayısının yani k 'nin belirlenmesindeki temel düşünce, $WSS(k)$ değerinin uygun küme sayısına kadar hızlı bir şekilde azaldığı ve uygun küme

sayısından sonra yavaş bir şekilde azaldığıdır. Yani $WSS(k)$ değerinin en hızlı azalış değerine ulaştığı küme sayısı uygun küme sayısı olarak alınmaktadır.

2.4.2.4. Dunn indeksi

(Dunn 1973) tarafından geliştirilmiş olan bu indeks kesin ve iyi ayrılmış kümelerin belirlenmesinde kullanılmak amacıyla geliştirilmiştir. Bu yüzden kümelemenin sonuçları zor bölümlleme algoritması sebebiyle tekrar hesaplanmalıdır.

$$DI(c) = \min_{j \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in c_i, y \in c_j} d(x, y)}{\max_{k \in c} \left\{ \max_{x, y \in c} d(x, y) \right\}} \right\} \right\} \quad (2.29)$$

Dunn indeksinin en büyük dezavantajı zor hesaplanmasıdır. Çünkü buradaki c küme sayısı ve n veri sayısı arttığında hesaplama bunlara bağlı olarak zorlaşır. Dunn indeksi optimum küme sayısını en büyük değerinde vermektedir.

2.4.2.5. Wilk's lambda istatistiği

Çok Değişkenli Varyans Analizi (MANOVA) iki ve daha fazla gruba ait çok değişkenli verilerin aynı ortalama vektörlü çok değişkenli normal dağılımdan gelip gelmediğinin testinde kullanılmaktadır (Johnson, Wichern 1988). Birden fazla bağımsız değişkeni, iki veya daha fazla bağımlı değişken arasındaki ilişkiyi eş zamanlı olarak incelemek istediğimizde MONOVA dan yararlanabiliriz. Monova'da bir veya birden fazla değişken bağımlı değişken test edilebilir, bağımsız değişkenler arasındaki değişkenler analiz edilebilir, hangi grubun diğerlerinden farklı olduğu karşılaştırmalar yapılabilir.

Veriler kümelere ayrıldıktan sonra, elde edilen kümelere MANOVA analizi uygulanabilir ve eğer iyi bir kümeleme yapılmış ise MANOVA analizinde de küme ortalamalarının birbirinden istatistiksel olarak farklı olması beklenir. Oluşturulacak kümeler için hesaplanan Wilks Lambda ölçütünün değerleri küme sayısının belirlenmesinde kullanılabilir (Tatlıdil 1996).

Wilk's lambda değeri 0 ile 1 arasında bir değer almaktadır. Wilk's lambda değeri küçüldükçe bağımsız değişkenin bağımlı değişken üzerindeki etkisi artmaktadır.

Küme sayısı arttıkça Wilks istatistiğinin değeri azalmakta ve sifira yaklaşmaktadır. Wilk's lambda değeri 0 çok yaklaşmamakta birlikte sifira en çok yaklaştığı değerde grup sayısı uygun küme sayısı kabul edilmektedir.

MANOVA uygulanabilmesi için; verilerin çok değişkenli normal dağılımdan seçilmesi, gruptaki kovaryansların homojen olması gibi varsayımlara ihtiyaç vardır (Krzanowski 1993).

Varyansların homejenliği testi ise Box'ın M testi yapılmaktadır. Varyansların homejenliği gruptaki birimlerin varyans-kovaryans matrislerinin istatistiksel olarak aynı olması anlamına gelmektedir. Varyansların homojenliği için hipotezler $\sum_i (i = 1, \dots, k)$ i . gruptaki birimlerin seçildikleri yığının varyans-kovaryans matrisi olmak üzere aşağıdaki şekilde oluşturulmaktadır.

$$H_0 : \sum_1 = \sum_2 = \dots = \sum_k \quad (2.30)$$

$$H_1 : \sum_i \neq \sum_j ; i, j = 1, \dots, k, i \neq j \quad (2.31)$$

Bu hipotezin test edilmesinde box'ın M istatistiği ise;

$$M = \sum_{i=1}^k (n_i - 1) \ln |S| - \sum_{i=1}^k (n_i - 1) \ln |S_i| \quad (2.32)$$

Şeklin de hesaplanmaktadır (Hawkins 1982). Burada n_i , i . gruptaki birim sayısı,

S_i , i . gruptaki birimlerden hesaplanan örneklem varyans-kovaryans matrisi ve

$$S = \frac{\sum_{i=1}^k (n_i - 1) S_i}{\sum_{i=1}^k (n_i - 1)} \quad (2.33)$$

ise tüm gruplar için örneklem ortak varyans-kovaryans matrisini göstermektedir. Box'ın M istatistiğinin değerlendirilebilmesi için şu formüller kullanılmaktadır.

$$C = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left[\sum_{n=1}^k \frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^k (n_i - 1)} \right] \quad (2.34)$$

gibi bir C çarpanı hesaplanır. Buradan MC istatistiğinin $p(p+1)(k-1)/2$ serbestlik dereceli χ^2 dağılımına sahip olduğu gösterilmiştir (Harris 1975). Hesaplanan MC χ^2 değeri tablo değerinden küçük ise $H_0: \sum_1 = \sum_2 = \dots = \sum_k$ şeklindeki hipotez red edilemez. Çıkan sonuca göre gruptaki varyans-kovaryans matrislerinin homojen olması MANOVA yönteminin uygulanabilmesi anlamına gelmektedir.

MANOVA iki ve daha fazla gruba ait çok değişkenli verilerin aynı ortalama vektörlü çok değişkenli normal dağılımdan gelip gelmediğinin testinde kullanılmaktadır. Varyans-kovaryans matrisleri homojen olan k tane grup için $\underline{\mu}_i$ ($i=1, \dots, k$) i . gruptaki birimlerin seçildikleri yığının ortalama vektörü olmak üzere, hipotezler aşağıdaki şekilde oluşturulmaktadır.

$$H_0 : \underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_k \quad (2.35)$$

$$H_1 : \underline{\mu}_i \neq \underline{\mu}_j \quad j \neq 1, \dots, k, i \neq j \quad (2.36)$$

kurulan bu hipotezlerin test edilmesinde çok değişkenli varyans analizinde üç model kullanımı ön plana çıkmaktadır. Bunlar Roy'un en büyük öz değerler yöntemi, Hotelling- Lawley iz yöntemi, wilk's lambda en çok olabilirlik yöntemidir. Bu bağlamda wilk's lambda yöntemi ön plana çıkmaktadır.

Bu metot wilk's tarafından 1932 yılında önerilmiştir. Wilk's yönteminde amaç değişkenler arasındaki bağlantı gösterilmeye çalışılmaktadır.

Wilks'in olabilirlik oran istatistiği, T ; genel çarpımlar ve kareler toplamı matrisi, B ; gruplar arası çarpımlar ve kareler toplamı matrisi ve W grup içi çarpımlar ve kareler toplam matrisi olmak üzere,

$$\Delta = \frac{|W|}{|W + B|} = \frac{|W|}{|T|} \quad (2.37)$$

Şeklin de hesaplanmaktadır (Tatlıdil 1996). Burada

$$B = \sum_{j=1}^k n_j (\overline{X}_j - \overline{X})(\overline{X}_j - \overline{X})' \quad \text{ve} \quad (2.38)$$

$$W = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)(X_{ij} - \overline{X}_j)' \quad \text{biçimindedir.} \quad (2.39)$$

Burada ; X_{ij} ; j. kümedeki i. birim değerleri

\overline{X}_j ; j. kümenin örneklem ortalama vektörü

\overline{X} ; Tüm birimlerin örneklem ortalama vektörü

n_j ; j. kümedeki birim sayısı

wilk's in değeri [0 1] arasında değişmektedir. Bu değer 0 yaklaşması H_0 hipotezinin kabul edilmeyeceğinin, 1 değerine yaklaşması ise kabul edileceğinin, yani yöntemin uygulandığı gruplar arasında fark olmadığını göstergesidir. Burada n' nin yeterince büyük olması durumunda,

$$L = -(n-1-(p+k)/2) \ln \Delta \quad (2.40)$$

istatistiğinin $(p)(k-1)$ serbestlik dereceli yaklaşık χ^2 dağılımına sahip olduğu gösterilmiştir. Hesaplanan L istatistiği χ^2 tablo değerinden büyük ise birimlerin aynı ortalama vektörlü yığınlardan geldiğini iddia eden H_0 hipotezi reddedilir (Johnson, Wichern 1988).

H_0 hipotezi reddedilirse birimlerin ortalama vektörlerinin en az ikisinin birbirinden farklı olduğu sonucuna varılır. H_0 hipotezi reddedilirse birbirinden farklı olan grupların incelemesi Hotelling T^2 yöntemi ile yapılabilmektedir. Hotelling T^2 istatistiği çok değişkenli iki kitle ortalama vektörünün eşitliğinin test edilmesinde kullanılan bir istatistiktir (Hawkins 1982). n_1 birinci gruptaki birim sayısı, n_2 ikinci gruptaki birim sayısı, S_1 birinci gruptaki birimlerden hesaplanan örneklem varyans-kovaryans matrisi, S_2 ikinci gruptaki birimlerden hesaplanan örneklem varyans-kovaryans matrisi ve

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n - 2} \quad (2.41)$$

ise ortak örneklem varyans-kovaryans matrisi olmak üzere,

T^2 istatistiği;

$$T^2 = \frac{n_1 n_2}{n} (\overline{X}_1 - \overline{X}_2)' S^{-1} (\overline{X}_1 - \overline{X}_2) \quad (2.42)$$

olarak hesaplanmaktadır (Krzanowski 1993).

Burada; \overline{X}_1 ; 1. gruptaki örneklem ortalama vektörü ;

\overline{X}_2 ; 2. gruptaki örneklem ortalama vektörü

Burada varyans- kovaryans bilinmiyor fakat 2 tane ortak grup için $\underline{\mu}_i$ ($i = 1, 2$)

i . gruptaki birimlerin seçildikleri yığının ortalama vektörü olmak üzere, hipotezler aşağıdaki şekilde oluşturulmaktadır.

$$H_0 : \underline{\mu}_1 = \underline{\mu}_2 \quad (2.43)$$

$$H_1 : \underline{\mu}_1 \neq \underline{\mu}_2 \quad (2.44)$$

Bu hipotezin test edilmesinde T^2 istatistiğine dayalı

$$F_{p, n-p-1} = \frac{n-p-1}{p} \frac{T^2}{n-2} \quad (2.45)$$

istatistiği kullanılır ve eğer $F_{p, n-p-1}$ hesaplanan değeri tablo değerinden büyük ise

$H_0 = \underline{\mu}_1 = \underline{\mu}_2$ hipotezi reddedilir ve iki grubun ortalama vektörlerinin birbirinden farklı olduğu sonucuna varılır (Seber 1984).

BÖLÜM 3

3. UYGULAMA

Bu çalışmada Türkiye İstatistik Kurumundan alınmış gelişmişlik araştırmalarında kullanılan üç farklı yıla ait 2002, 2008 ve 2013 yıllarının (55 değişkenli) sosyo-ekonomik verileri kullanılmıştır. 2002 yılının verileri DPT, 2008 ve 2013 yılları TÜİK kurumundan alınmıştır. Sosyo ekonomik özellikler bakımından Türkiye'deki 81 ilin benzerlikleri klasik kümeleme analizi teknikleriyle incelenmesi amaçlanmıştır.

Kümeleme analizi tekniklerinden 7 farklı yöntem (Gruplar Arası, Grup İçi, en yakın komşuluk, en uzak komşuluk, centroid metodu, medyan, ward's yöntemi) ile 81 ile ait farklı değişkenler sınıflandırılmaya çalışılmıştır. Bu değişkenler Sağlık, Eğitim, Çevre, Tarım, Nüfus, Mali, Sanayi, İnşaat ana başlıkları yer almaktadırlar ve TÜİK web sayfasındaki raporlardan tek tek elde edilmiştir. Analizlerde SPSS.25 paket programı kullanılmıştır. Veri setleri için uygun küme sayısı, Silhoutte küme geçerlilik indeksi yardımıyla araştırılmıştır. Küme sayısı olarak 2'den 7'ye kadar küme sayıları Silhoutte küme geçerlilik indeksi ile incelenmiş ve her veri seti için uygun küme sayısı 5 olarak belirlenmiştir. Farklı yıllara ait değişkenlerin ölçüm düzeyleri farklılık göstermesinden dolayı çalışmanın uygulama kısmında değişkenler için standartlaştırılma yapılmıştır.

3.1. 2002 Yılı İçin Bulgular ve Sonuçlar

3.1.1. 2002 Yılı Sosyo-Ekonomik Verilerinin Gruplar Arası Yöntemi İle Kümeleme Sonuçları

Bu yöntemde elde edilen sonuçlara göre 2. Kümede Ankara; 3. Kümede Bursa, İzmir; 4. Kümede İstanbul ve 5. Kümeyi ise Kocaeli tek başına oluşturmuştur. Gruplar Arası en iyi iller grubunu ayrı ayrı 4 farklı küme şeklinde sonuçlandırmıştır. Bu ayrı ayrı farklı 4 kümede yer alan toplamdaki 5 il Türkiye'nin sanayinin ve kültürel gelişmişliğin en fazla olduğu büyükşehirlerdir. 1. kümeyi ise geriye kalan 76 il oluşturmaktadır. 76 il içerisinde 2002 yılında büyükşehir statüsünde olan iller diğer gelişmişlik düzeyi daha az olan doğu ve güneydoğu illeri ile birlikte kümelemiştir. Bu iller sosyo ekonomik verilerin birçok kategorisinde (tarım, inşaat, mali, eğitim,

sağlık..vs) Grup İçi yöntemine göre aynı özelliklere sahip olduğunu göstermektedir. (Adana, Adıyaman, Afyon, Ağrı, Amasya, Antalya, Artvin, Aydın, Balıkesir, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Çanakkale, Çankırı, Çorum, Denizli, Diyarbakır, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Hakkari, Hatay, Isparta, Mersin, Kars, Kastamonu, Kayseri, Kırklareli, Kırşehir, Konya, Kütahya, Malatya, Manisa, Kahramanmaraş, Mardin, Muğla, Muş, Nevşehir, Niğde, Ordu, Rize, Sakarya, Samsun, Siirt, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Tunceli, Şanlıurfa, Uşak, Van, Yozgat, Zonguldak, Aksaray, Bayburt, Karaman, Kırıkkale, Batman, Şırnak, Bartın, Ardahan, Iğdır, Yalova, Karabük, Kilis, Osmaniye, Düzce)

Yukarıda verilen kümeleme analizi sonuçları tablo olarak, Tablo 3.1’de derlenmiştir. Özellikle bazı kümelerde birer veya ikişer illerin yer alması nedeniyle, diğer kümeleme analizleri tablo yerine metin olarak vurgulanmıştır.

Tablo 3.1. 2002 yılı Sosyo ekonomik verilerinin Gruplar Arası yöntemi ile kümeleme sonuçları

1. Küme			2. Küme	3. Küme	4. Küme	5. Küme
ADANA	EDİRNE	MANİSA	ANKARA	BURSA	İSTANBUL	KOCAELİ
ADİYAMAN	ELAZIĞ	MARDİN		İZMİR		
AFYON	ERZİNCAN	MERSİN				
AĞRI	ERZURUM	MUĞLA				
AKSARAY	ESKİŞEHİR	MUŞ				
AMASYA	GAZİANTEP	NEVŞEHİR				
ANTALYA	GİRESUN	NİĞDE				
ARDAHAN	GÜMÜŞHANE	ORDU				
ARTVİN	HAKKARİ	OSMANİYE				
AYDIN	HATAY	RİZE				
BALIKESİR	IĞDIR	SAKARYA				
BARTIN	ISPARTA	SAMSUN				
BATMAN	K.MARAŞ	SİİRT				
BAYBURT	KARABÜK	SİNOP				
BİLECİK	KARAMAN	SİVAS				
BİNGÖL	KARS	ŞANLIURFA				
BİTLİS	KASTAMONU	ŞIRNAK				
BOLU	KAYSERİ	TEKİRDAĞ				
BURDUR	KIRIKKALE	TOKAT				
ÇANAKKALE	KIRKLARELİ	TRABZON				
ÇANKIRI	KIRŞEHİR	TUNCELİ				
ÇORUM	KİLİS	UŞAK				
DENİZLİ	KONYA	VAN				
DİYARBAKIR	KÜTAHYA	YALOVA				
DÜZCE	MALATYA	YOZGAT				
		ZONGULDAK				

3.1.2. 2002 Yılı Sosyo Ekonomik Verilerin Grup İçi Yöntemi İle Kümeleme Sonuçları

Elde edilen sonuçlara göre 5. Kümede Kocaeli, 4.kümede İstanbul, 3. Kümede Ankara tek başına yer alırken 2. Kümede 14 il (Ağrı, Batman, Bingöl, Bitlis, Diyarbakır, Giresun, Hakkari, Iğdır, Mardin, muş, Siirt, Şanlıurfa, Şırnak, Van) ve geriye kalan 64 ilin hepsi 1. Kümeyi oluşturmuştur. Burada İstanbul, Ankara, Kocaeli'nin tek başına beşinci dördüncü ve üçüncüyü kümeyi oluşturması bu illerin sosyo ekonomik gelişmişlik düzeylerinin diğer illere göre daha fazla olmasından kaynaklanmaktadır. İstanbul ve Kocaeli'nin sanayi sektöründe öncü iller olması bu illeri tek başına kümelemiştir. 14 ilin yer aldığı 2. Küme ise sosyo ekonomik gelişmişliğin az olduğu illerden oluştuğu gözlemlenirken burada yer alan Karadeniz'in toprak yüzölçümü olarak küçük illerinden birisi olan Giresun'un yer alması dikkat çekmiştir. Yığılmanın 1. Kümede olduğu (64 il) gözlenebilmektedir.

3.1.3. 2002 Yılı Sosyo Ekonomik Verilerin En Yakın Komşuluk Yöntem İle Kümeleme Sonuçları

Ankara 2. Kümeyi, Giresun 3. Kümeyi, İstanbul 4.kümeyi ve Kocaeli tek başına 5. Kümeyi oluştururken bu yöntem sonuçlarına göre 5. Kümeyi bu iller dışında kalan 77 il (Adana, Adıyaman, Afyon, Ağrı, Amasya, Antalya, Artvin, Aydın, Balıkesir, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Diyarbakır, Edirne,Elazığ,Erzincan,Erzurum,Eskişehir,Gaziantep,Gümüşhane,Hakkari,Hatay,Isparta ,Mersin,İzmir,Kars,Kastamonu,Kayseri,Kırklareli,Kırşehir,Konya,Kütahya,Malatya,Manisa,Kahramanmaraş,Mardin,Muğla,Muş,Nevşehir,Niğde,Ordu,Rize,Sakarya,Samsun,Siirt,Sinop,Sivas,Tekirdağ,Tokat,Trabzon,Tunceli,Şanlıurfa,Uşak, Van,Yozgat,Zonguldak, Aksaray,Bayburt,Karaman,Kırıkkale,Batman,Şırnak,Bartın,Ardahan,Iğdır,Yalova,Karabük,Kilis, Osmaniye, Düzce) oluşturmaktadır. En yakın komşuluk yöntemi sonuçlarına beklenmeyen bir durum olarak dikkat çeken 3. Kümenin tek başına Giresun'un oluşturmuş olmasıdır. Karadeniz'in toprak yüz ölçümü bakımından küçük olmasına rağmen Türkiye tarım sektörü açısından önemli olan bir ürünü üreten bir il olması Giresun'u 3. Kümede tek başına yer almasına neden olmuştur. Ankara, İstanbul ve Kocaeli gibi sosyo ekonomik gelişmişliğin fazla olduğu iller yine bu yöntemde de bu ayrı ayrı kümelemiştir.

3.1.4. 2002 Yılı Sosyo Ekonomik Verilerin En Uzak Komşuluk Yöntemi İle Kümeleme Sonuçları

Bu yöntemin sonuçlarına göre İstanbul 5. Kümede; Bursa, İzmir, Kocaeli 4. kümede; Ankara ise 3. Kümeyi oluşturmaktadır. 1. Kümede ise 21 il (Adana, Antalya, Aydın, Balıkesir, Bilecik, Bolu, Denizli, Eskişehir, Gaziantep, Hatay, Kayseri, Kırklareli, Konya, Mersin, Manisa, Muğla, Sakarya, Samsun, Tekirdağ, Yalova, Zonguldak) yer almıştır. Bu illerin dışında kalan 55 il ise 2. Kümeyi oluşturmaktadır. İstanbul tek başına sosyo ekonomik gelişmişlik olarak her yöntemde tek başına bir küme oluşturabilirken bu yöntemde de 5. Kümede İstanbul yer almaktadır. Sanayi gelişmişliği benzer olan 3 il Bursa, İzmir ve Kocaeli aynı kümede yer almıştır. 1. Küme ise aralarında Adana, Samsun, Konya gibi tarım alanların çok işlendiği tarım sektörünün geliştiği beraberinde Muğla gibi turizmin aktif olduğu, Zonguldak gibi madencilikte öncülük yapan iller oluşturduğu bir küme olmuştur. Doğu ve güneydoğu gibi az gelişmiş illerinde beraberinde olduğu 2. Kümede yığılma olmuştur.

3.1.5. 2002 Yılı Sosyo Ekonomik Verilerin Centroid Kümeleme Yöntemi İle Sonuçları

Bu yöntem sonucuna göre İstanbul tek başına 4. Kümede yer alırken; Kocaeli 5. Kümeyi; Bursa ve İzmir ise 3. Kümeyi, Ankara 2. Kümeyi oluşturmaktadır. Bunların dışında kalan 77 il ise centroid kümeleme yöntemine göre 1. Kümeyi oluşturmaktadır. Doğal kaynaklarının, sosyo ekonomik birçok dinamiğin benzer olduğu Bursa ve İzmir bu yöntemde de 3. Kümede birlikte yer almıştır.

3.1.6. 2002 Yılı Sosyo Ekonomik Verilerin Median Kümeleme Yöntemi İle Sonuçları

Median kümeleme analizi sonuçlarına göre Kocaeli 5. Kümede; İstanbul 4. Kümede; Ankara 3. Kümeyi oluştururken; 2 kümeyi 15 il (Ağrı, Artvin, Bingöl, Bitlis, Batman, Giresun, Gümüşhane, Hakkari, Mardin, Muş, Siirt, Şanlıurfa, Şırnak, Tunceli, Van) oluşturmaktadır. Bu iller dışında geriye kalan 63 il ise 1. Kümeyi oluşturmaktadır. En yakın komşuluk yönteminde 3. Kümeyi tek başına oluşturan Giresun median yönteminde 2. Kümede gelişmişlik düzeyi az olan doğu ve güneydoğu illeriyle birlikte 2 kümede yer alması beklenmeyen bir durum olmuştur. 1. Kümeyi oluşturan 63 il

içerisinde gelişmişlik düzeyi birbirinden farklı olan İzmir ve Iğdır gibi illerin yer alması dikkat çeken bir sonuç olmuştur.

3.1.7. 2002 Yılı Sosyo Ekonomik Verilerin Ward Yöntemi İle Sonuçları

2002 yılı verilerinin ward's method sonuçlarına göre 4. Kümeyi Ankara ve İstanbul; İzmir, Bursa, Bilecek, Tekirdağ, Yalova, Kırklareli, Kocaeli ise 5. Kümeyi oluşturmaktadır. 2002 yılında büyükşehir (Adana, Kayseri, Gaziantep, Eskişehir vs) olan birçok ili de içinde barındıran 1. Kümede toplamda 25 il yer almaktadır. 2.kümede (Ağrı, Adıyaman, Ardahan, Batman, Bingöl, Bitlis, Diyarbakır, Hakkari, Iğdır, Kars, Mardin, Muş, Siirt, Şanlıurfa, Şırnak, Van) 16 il kümelenirken, geriye kalan 31 il 3. Kümede gruplaşmışlardır. 2002 yılı sosyo ekonomik verilerine göre 7 yöntem içerisinde en çok gerçek hayatla örtüşecek sonuçları Ward's yönteminde izlemekteyiz. İstanbul ve Ankara'nın Türkiye'nin iki büyük şehiri olarak aynı küme yer alması beklenen bir durumdur. 5. Kümede ise İzmir, Bursa, Kocaeli gibi sanayinin öncü illeri yer almaktadır. Burada Kırklareli'nin 2. kümede yer alması ilginç sonuçlardan birisidir. Tarımda, sanayide, turizmde gelişmiş olan birçok il 5. Kümede birlikte yer almıştır. 2. kümede ise sosyo ekonomik gelişmişlik düzeyi düşük olan doğu ve güneydoğu illeri birlikte kümelenmiştir.

3.2. 2008 Yılı İçin Bulgular ve Sonuçlar

3.2.1. 2008 Yılı Sosyo Ekonomik Verilerinin Gruplar Arası Yöntemi İle Kümelenme Sonuçları

2008 yılı sosyo ekonomik verilerinin Gruplar arası yöntemi ile kümelenmesi sonucu elde edilen sonuçlara göre İstanbul 5. Kümede; Antalya 4. Kümede; Ankara 3. Kümeyi tek olarak oluşturmuşlardır. İstanbul'un birçok kümeleme yönteminde tek başına bir kümeyi oluşturuyor olması birçok sosyo ekonomik etkenlerde ön sıralarda yer almasından kaynaklanmaktadır. Birçok sosyo ekonomik etkenin bir arada barındıran İstanbul (turizm, sağlık, nüfus, tarım.. v.s) bu bağlam da ayrı ele alınması gereken bir şehirdir. Dördüncü ve üçüncü kümeyi oluşturan Antalya ve Ankara ise yine birçok sosyo ekonomi değişkende (turizm, tarım gibi) öncü olan illerimizdendir. 1. Kümeyi ise 5 farklı büyük şehrimiz oluşturmaktadır. Bu şehirler Adana, Mersin, İzmir, Bursa, Konya'dır. Bu beş şehrin bir arada kümelenmesinde birçok değişkene dikkat çekebiliriz

(tarım, sanayi) bunların arasındadır. Geriye kalan 73 il ise 2. Kümeyi oluşturmaktadır. 2. Kümede dikkat çeken ise sosyo ekonomik gelişmişlik düzeyi çok düşük olan birçok ilin büyükşehirler ile birlikte kümelenmesidir. Kocaeli, Gaziantep gibi..

3.2.2. 2008 Yılı Sosyo Ekonomik Verilerinin Grup İçi Yöntemi İle Kümelenmesi Sonuçları

Bu yöntemde Türkiye'nin iki büyük ili 4.(Ankara) ve 5.(İstanbul) Kümeleri oluşturmuştur. 1. Kümeyi ise tarım sektöründeki büyük ilerimizden olan Akdeniz bölgesindeki iki ilimiz olan Adana ve Mersin oluşturmaktadır. Doğu ve güneydoğu illerinin bulunduğu 3. Kümede ise en dikkat çeken il ise Konya'dır. 57 tane ilin yığılmanın en fazla olduğu küme ise 2. Küme olmuştur.

3.2.3. 2008 Yılı Sosyo Ekonomik Verilerinin En Yakın Komşuluk Yöntemi İle Kümelenmesi Sonuçları

Beş büyük ilimizden üç tanesi kendi başlarına 5.(İstanbul), 2.(Ankara), 3.(Antalya) kümeler oluşturmaktadır. Bu yöntemde en dikkat çeken kümeleme ise 4. Kümede Eskişehir'in tek başına yer almasıdır. Eskişehir coğrafi konumunun yarattığı avantajlarla tarıma dayalı sanayinin son dönemlerde gelişme gösterdiği bir il olmuştur. Geriye kalan 77 il ise tek kümede toplanmıştır. Bu yığılmada ise gerçek ile bağdaşmayan bir sonuçla karşılaşmıştır. İzmir, Kocaeli, Bursa gibi sanayisi gelişmiş illerin sosyo ekonomik düzeyi düşük iller ile aynı kümede yer alması ilginç sonuç olarak karşımıza çıkmaktadır.

3.2.4. 2008 Yılı Sosyo Ekonomik Verilerinin En Uzak Komşuluk Yöntemi İle Kümelenmesi

2008 yılına ait 55 değişkenle yapılan kümeleme analizi sonucunda en uzak komşuluk yönteminde İstanbul yine 5. Kümede tek başına yer almıştır. 4.kümede ise Konya, Bursa, İzmir, Antalya yer almıştır. 4. Küme birçok değişken konusunda dengeli olan şehirlerden oluşması beklenen sonuç olarak karşımıza çıkmıştır. 3. Kümeyi ise nüfus, sanayi, eğitim gibi birçok değişkende İstanbul'dan sonra gelen Ankara oluşturmuştur. Yine bu kümeleme yönteminde 1. Kümeyi Adana ve Mersin oluşturmuştur. Adana ve Mersin birlikte kümeleyen en önemli değişkenin tarım(Çukurova) olduğu gözlenmiştir. Son olarak diğer bütün illerin kümelendiği 2.

Kümede 73 il birlikte yer almıştır. 2. Kümede ise Kocaeli, Gaziantep, Kayseri gibi sosyo ekonomik gelişmiş gösteren illerin gelişmişlik düzeyi düşük olan illerle birlikte olması ilginç sonuç olarak karşımıza çıkmaktadır.

3.2.5. 2008 Yılı Sosyo Ekonomik Verilerinin Centroid Kümeleme Yöntemi İle Kümelenmesi Sonuçları

Bu çalışmada Ankara 2. Kümede, Antalya 3., İstanbul 4.kümede ve Konya 5. Kümede tek başına yer almıştır. Bunun dışında kalan bütün iller 1. Kümeyi oluşturmuştur. 2008 yılı için 55 değişkenle yapılan centroid kümeleme yöntemi karşımıza gerçekte bağdaşmayan sonuçlar çıkarmıştır. Çünkü sosyo ekonomik gelişmişlik düzeyi ileri olan iller ile diğer illeri birlikte kümelemiş olması yorum oluşturabilme güçlüğü çıkarmıştır.

3.2.6. 2008 Yılı Sosyo Ekonomik Verilerinin Median Kümeleme Yöntemi İle Kümelenmesi Sonuçları

Median kümelemede yine gerçekte çok bağdaşmayan sonuçlar çıkarmıştır. Burada ilginç çıkan sonuç Eskişehir'i tek başına 4. Kümede kümelemesidir. Ankara 2. Kümede, Antalya 3. Kümede. İstanbul ise 5. kümede yer almıştır. Bunun dışındaki bütün illeri 1. Kümede birleştirilerek yığılmayı burada oluşturmuştur.

3.2.7. 2008 Yılı Sosyo Ekonomik Verilerinin Ward Yöntemi İle Kümelenmesi Sonuçları

Gerçeğe en yakın sonuçları 2008 yılı çalışmasında da yine ward kümeleme yöntemi vermiştir. Ankara (4.kümede) ve İstanbul (5.kümede) tek başına yer almıştır. Beklenen bir sonuç olarak karşımıza çıkmaktadır. 1. Kümede Adana, Antalya, Bursa, Konya, İzmir, Mersin gibi sosyo ekonomik olarak ileri gelişmişlik gösteren iller yer almıştır. Kocaeli ve Gaziantep' in de bu kümede olması beklenirdi. 3. Kümede ise hemen hemen hepsi doğu ve güneydoğu illerinden oluşmaktadır. 2008 yılına göre buradaki ilginç sonuç olarak Diyarbakır verebiliriz. Doğuda yer almasına rağmen Diyarbakır büyükşehir belediyesidir. Bu illerin dışında kalan bütün iller 2. Kümeyi oluşturmaktadır. 2. Kümede yer alan Kocaeli, Gaziantep gibi gelişmiş illerin Adıyaman, Çorum, Yozgat gibi illerle aynı kümede yer alması ilginç sonuç olarak karşımıza çıkmaktadır.

3.3. 2013 Yılı İçin Bulgular ve Sonuçlar

3.3.1 2013 Yılı Sosyo Ekonomik Verilerin Gruplar Arası Yöntemi İle Kümelenmesi Sonuçları

İstanbul 5. Kümede, Eskişehir ise 4. Kümede, Antalya 3. Kümede tek başına yer almıştır. 2. Kümeyi ise gelişmişlik düzeyi ileri olan üç il (Ankara, İzmir, Konya) birlikte oluşturmuştur. 2013 yılı verilerine ilginç sonuç olarak Eskişehir'in tek başına 4. Kümeyi oluşturmasıdır. Eskişehir son zamanlarda bulunduğu coğrafi konumu iyi değerlendirilerek gelişme tarımda, sanayide, eğitimde gelişme göstermiş iller arasındadır. Ankara, İstanbul, izmir gibi illeri yakın olması Eskişehir'i avantajlı il haline getirmiştir. Bu iller dışında kalan bütün illeri(75 il) Gruplar arası yöntemine göre 1. Kümede toplamıştır. Adana, bursa, Kocaeli gibi gelişmiş illerin bu kümede bulunması beklenmeyen sonuçtur.

3.3.2. 2013 Yılı Sosyo Ekonomik Verilerin Grup İçi Yöntemi İle Kümelenmesi Sonuçları

Bu kümelemede İstanbul 5. Kümede, Antalya 4. Kümeyi tek başına oluşturmuştur. 3. Kümede ise genelde doğu ve güneydoğu illerinin oluşturduğu gözlenmiştir. Buradaki ilginç sonuç ise 2013 yılına büyükşehir olan Diyarbakır ve Şanlıurfa'nın 3. Kümede yer almasıdır. 1. Kümeyi ise sosyo ekonomik gelişmişlik düzeyi yüksek olan iller oluşturmuştur. 1. Kümede yer bulan büyükşehir tarımda, turizmde, sanayide gelişme göstermiş illerdir. Yığılmanın olduğu 2. Küme ise gelişmişlik düzeyi benzer illerden oluşmaktadır. Burada ise sanayide gelişme göstermiş iki il olan Kocaeli ve Gaziantep'in yer alması beklenmeyen sonuçtur.

3.3.3. 2013 Yılı Sosyo Ekonomik Verilerinin En Yakın Komşuluk Yöntemi İle Kümelenmesi Sonuçları

2013 yılının sosyo ekonomik verilerinden en yakın komşuluk yöntemi ile yapılan bu çalışmada sonuçlar çok gerçekle bağdaşmayacak şekilde çıkmıştır. Yığılmanın olduğu 1. Kümede sosyo ekonomik yönden güçlü olan iller ile diğer iller (77 il) birlikte kümelenmiştir. Bu kümelenme sonucu ise çok gerçekçi olarak karşımıza çıkmamıştır. Antalya ise 2. Kümede, İstanbul 4. Kümede, Ankara 5. Kümede tek başına

küme oluştururken 3. Küme son zamanlarda sosyo ekonomik olarak gelişme gösteren Eskişehir yine tek başına küme oluşturmuştur.

3.3.4. 2013 Yılı Sosyo Ekonomik Verilerinin En Uzak Komşuluk Yöntemi İle Kümelenmesi Sonuçları

2013 yılı verilerine göre en uzak komşuluk yöntemi ile kümelenme gerçeğe yakın sonuçlar çıkarmıştır. İstanbul 5.kümede, Antalya ise 4. Kümede yine tek başına oluşturmuştur. Antalya tek başına küme oluşturabilecek birçok sosyo ekonomik değişken bahsedilebilir. Bunların başında turizm, tarım, eğitim gibi değişkenlerde Antalya ili son zamanlarda çok gelişme göstermiştir. Antalya'nın bu gelişmesi onu tek başına küme olarak çıkarmıştır. Bu çalışma sonucunda 3. Küme ise Konya, Ankara, İzmir yer almıştır. Bu üç büyükşehir kümelenmesinde öne çıkan değişken ise sanayi olarak düşünülebilir. Ankara ve İzmir ekonomik gelişmişlik göstermiş iller olurken, Konya son zamanlarda tarıma dayalı sanayileşmede gelişme göstermiş büyükşehirlerden olmuştur. Bu kümede Kocaeli, Gaziantep, Kayseri, Bursa gibi sanayileşmesi kuvvetli olan illerde olması beklenebilirdi. 2. Kümede ise genelde doğu ve güney doğu illeri yer almıştır. Bu illerin birçok sosyo ekonomik değişken yönünden zayıf olduklarını söyleyebiliriz. Bu kümede yer alan iki büyükşehir(Diyarbakır, Şanlıurfa) olması beklenmeyen sonuçtur. Yığılmanın olduğu 1. Kümede olmuştur. Birinci kümede yer alan iller aslında çok çelişkili olarak kümelenmiştir. Bursa, Kocaeli gibi sanayisi kuvvetli olan bir illerin Çorum, Sivas, Elazığ, Bilecik gibi hem nüfus olarak hem de sanayileşme olarak daha zayıf illerle birlikte kümelenmesi beklenmeyen sonuçtur.

3.3.5. 2013 Yılı Sosyo Ekonomik Verilerinin Centroid Kümeleme Yöntemi İle Kümelenmesi Sonuçları

Centroid kümeleme sonuçlarına göre Antalya, İstanbul, Konya tek başlarına birer küme oluşturmuşlardır. Ankara ve İzmir 2. Kümeyi birlikte oluşturmuşlardır. Yine beklenmeyen sonuç olarak 1. Kümede diğer bütün iller birlikte kümelenmiştir.

3.3.6. 2013 Yılı Sosyo Ekonomik Verilerinin Median Kümeleme Yöntemi İle Kümelenmesi Sonuçları

İstanbul, Antalya tek başına küme oluşturmuşlardır. Ankara, İzmir ve Konya üç büyükşehir birlikte 2. Kümeyi oluşturmuşlardır. Bu yöntemde beklenmeyen sonuç olarak 3. Kümeyi tek başına Eskişehir'in oluşturmasıdır. Bu illerin dışındaki bütün iller 1. Kümede sonuçları gerçekte bağdaşmayacak şekilde kümelenmişlerdir.

3.3.7. 2013 Yılı Sosyo Ekonomik Verilerinin Ward Yöntemi İle Kümelenmesi Sonuçları

2013 yılı sosyo ekonomik verilerine göre en uzak komşuluk yönteminden sonra gerçeğe en yakın sonuçları ward's kümeleme vermiştir. 5. Kümede İstanbul tek başına yer almıştır. Antalya, İzmir, Konya, Ankara gibi sosyo ekonomik gelişmişliği fazla olan 4 il üçüncü kümeyi birlikte oluşturmuştur. Kocaeli, Bursa gibi büyükşehirlerde bu kümede beklenebilirdi. Ward's yönteminde beklenmeyen sonuç olarak doğu ve güney doğu illerinin iki farklı kümede kümelenmiş olmasıdır. 2. Kümede gelişmişliğin zayıf olduğu hem de nüfus olarak az olan illerin yer aldığı küme olurken; 4. Küme genellikle doğu ve güneydoğuda daha fazla gelişmişliğin olduğu illerden oluşmuştur. Bu yöntemin sonuçlarına göre yaşılma (60 il) 1. Kümede olmuştur.

3.4. Kümeleme Yöntemleri Arasında Korelasyon Analizleri

Bu altbölümde, her veri seti için farklı yöntemlerle bulunan kümeleme sonuçları arasındaki korelasyon ölçülerine bakılmıştır.

Tablo 3.2'deki 2002 yılına ait sonuçlar arasındaki korelasyon değerlerine baktığımızda Gruplar Arası Kümeleme ile Centroid Kümeleme sonuçları aynı çıktığı görülmektedir. Benzer şekilde Medyan Kümeleme ve Gruplar İçi kümeleme sonuçları arasında da yüksek ilişki bulunmuştur. Tüm korelasyon değerleri $\alpha = 0,01$ için anlamlı çıkmıştır. Sadece (*) ile belirtilen değerler $\alpha = 0,05$ için istatistiksel anlamlı çıkmamışlardır.

Genel olarak, Ward Kümeleme yöntemi sonuçlarının diğer yöntemler ile düşük benzerlik gösterdiği görülmektedir. Ward yönteminin varyansları esas olması bu sonuçta etkili olduğu düşünülmektedir.

Tablo 3.2. 2002 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri

2002 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri							
	Gruplar Arası	Gruplar İçi	En Yakın Kom.	En Uzak Kom.	Centroid	Medyan	Ward
Gruplar Arası	1	,688	,805	,756	1,000	,680	,450
Gruplar İçi	,688	1	,808	,569	,688	,932	,219*
En Yakın Kom.	,805	,808	1	,560	,805	,799	,316
En Uzak Kom.	,756	,569	,560	1	,756	,569	,480
Centroid	1,000	,688	,805	,756	1	,680	,450
Medyan	,680	,932	,799	,569	,680	1	,258*
Ward	,450	,219*	,316	,480	,450	,258*	1

Tablo 3.3'deki 2008 yılına ait sonuçlar arasındaki korelasyon değerlerine baktığımızda En Yakın Komşuluk Kümeleme ile Medyan Kümeleme sonuçları aynı çıktığı görülmektedir. Yüksek ilişkili kümeleme yöntemleri olsa da, sonuçlar 2002 yılına ait sonuçlar kadar yüksek değildir. Burada da (**) ile belirtilen değerler hariç Tüm korelasyon değerleri $\alpha = 0,01$ için istatistiksel anlamlı çıkmıştır.

Tablo 3.3. 2008 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri

2008 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri							
	Gruplar Arası	Gruplar İçi	En Yakın Kom.	En Uzak Kom.	Centroid	Medyan	Ward
Gruplar Arası	1	0,504	0,722	0,438	0,423	0,722	0,553
Gruplar İçi	0,504	1	0,39	0,441	0,425	0,39	0,865
En Yakın Kom.	0,722	0,39	1	0,568	0,548	1	0,326
En Uzak Kom.	0,438	0,441	0,568	1	0,748	0,568	0,102**
Centroid	0,423	0,425	0,548	0,748	1	0,548	0,098**
Medyan	0,722	0,39	1	0,568	0,548	1	0,326
Ward	0,553	0,865	0,326	0,102**	0,098**	0,326	1

Tablo 3.4'de verilen 2013 yılına ait kümeleme yöntemleri sonuçları arasındaki korelasyon değerlerine baktığımızda ise Gruplar Arası Kümeleme ile Medyan Kümelemenin aynı sonuçlar verdiği, En Uzak Kümeleme ile Ward Kümelemenin yüksek korelasyon katsayısına sahip olduğu görülmektedir. Tüm korelasyon değerleri $\alpha = 0,01$ için istatistiksel anlamlı çıkmıştır.

Tablo 3.4. 2013 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri

2013 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri							
	Gruplar Arası	Gruplar İçi	En Yakın Kom.	En Uzak Kom.	Centroid	Medyan	Ward
Gruplar Arası	1	,371	,764	,648	,682	1,000	,418
Gruplar İçi	,371	1	,180	,631	,204	,371	,640
En Yakın Kom.	,764	,180	1	,544	,880	,764	,364
En Uzak Kom.	,648	,631	,544	1	,723	,648	,814
Centroid	,682	,204	,880	,723	1	,682	,476
Medyan	1,000	,371	,764	,648	,682	1	,418
Ward	,418	,640	,364	,814	,476	,418	1

3 dönemin korelasyon değerlerine bakıldığında 2013 yılına ait verilerinin diğerlerine göre daha yüksek korelasyon değerlerine sahip olduğu ve anlamsız bir korelasyon değerine de sahip olmadığı görülmektedir.

SONUÇ ve TARTIŞMA

İstatistik bilimi, belirli bir amaç için veri toplamak, toplanan verilerin uygun yöntemlerle analiz edilmesini yorumlanması bakımından her geçen gün önemi artan bir analiz aracı olmuştur.

İstatiksel analizler her alanda çalışma alanı bulurken, ülkelerde gelişmeyi artırmak için bu istatiksel analizlerde yararlanmaktadır. Ülkelerde toplanan nitelikli ve doğru veriler analiz edilmesiyle ulusal kalkınmayı artırmayı hedeflemişlerdir. Bu bağlamda ulusal kalkınmayı refah seviyeye ulaştırabilmek için bölgesel istatistikleri toplamak büyük önem arz etmiştir.

Bir ülkenin farklı bölgelerindeki sosyal, ekonomik, tarihsel, coğrafi, iklimsel gibi sosyo-ekonomik verilerin analiz edilmesi yatırımlar bakımından önemlidir. Ancak yapılacak bütün çalışmalara rağmen yapılan bölgeler arası gelişmişlik farkının yok edilmesi mümkün olmamış, ancak minimum düzeylere indirilmeye çalışılmıştır.

Bu çalışma da 81 ilin 2002, 2008 ve 2013 yılları sosyo ekonomik rakamları Türkiye İstatistik Kurumu yıllık bültenlerinden alınmış ve sosyo ekonomik benzerlik gösteren iller tespit edilmek amacıyla klasik kümeleme analizi teknikleri uygulanmıştır. Çalışmadaki amaç Türkiye'deki 81 ilin sosyo ekonomik benzerliklerine göre sınıflandırılmasının yapılmasıdır.

Üç farklı yılı ait veri setleri için küme sayısı, silhoutte küme geçerlilik indeksi yardımıyla belirlenmiştir. İncelen veri setleri için uygun küme sayısının 5 olarak belirlenmiştir. Aynı veri setlerine 7 farklı kümeleme tekniği için standardizasyon yapıldıktan sonra sonuçların değişimi incelenmiştir.

Belirlenen küme sayıları arasındaki korelasyon ölçüleride incelenmiştir. 2002 yılı için gruplar arası kümeleme ile Centroid Yöntemi için çıkan sonuç aynı olurken, medyan ve grup içi kümeleme sonucunda kümeler arası ilişkinin yüksek olduğu incelenmiştir. 2008 yılının korelasyon ilişkisi incelendiği çıkan korelasyon sonucunun 2002 yılı kadar yüksek olmadığı görülmüştür. 2013 yılı için ise uzak kümeleme ile ward kümelemenin ilişkisinin yüksek, gruplar arası kümeleme ile medyan aynı sonucunun aynı olduğu çalışmada gösterilmiştir.

7 farklı kümeleme analizi tekniğiyle yapılan çalışmada ortaya çıkan en önemli bulgulardan birisi iktisadi açıdan gelişmişliği benzerlik gösteren illerin aynı kümede birleşmesidir. İktisadi benzerlik gösteren illerin birleştiği kümeler incelendiğinde bu illerin kültürel açıdan da bariz benzerlikler gösteren iller olduğu somut bir gerçek olarak karşımıza çıkmaktadır.

2002, 2008 ve 2013 yıllarına ait veri setlerine yapılan kümeleme analizi bulgularından çıkarılacak sonuçlara göre genellikle İstanbul diğer tüm şehirlerden farklı olarak tek kümelenmiştir. Yani İstanbul ili için sosyo-ekonomik olarak diğer birçok büyükşehir statüsündeki ilden farklı olduğu yorumunu burada yapabiliriz.

2002 yılı sonuçlarını incelediğimizde sonuçları biraz daha karmaşık yapıda olduğu ancak 2008 ve 2013 yıllarına ait kümeleme sonuçlarının ise genelde sosyo-ekonomik benzerliğin bir arada olduğu iller şeklinde olduğu gözlenmiştir. 2002 yılı için Giresun ilinin tek başına bir kümede var olması bu yıl için belirleyici sosyo ekonomik kategorinin tarım olduğu yorumuna ulaşabiliriz. Ancak 2013 yılı aynı şekilde kümeleri incelendiğinde bu kategorinin sanayi ağırlıklı gelişme olduğu izlenmiştir.

Tablo 3.2’ deki 2002 yılına ait kümeleme yöntemleri arası yapılan korelasyon analizinde Gruplar Arası Kümeleme ile Centroid Kümeleme sonuçları aynı çıktığı görülmektedir. Benzer şekilde Medyan Kümeleme ve Gruplar İçi kümeleme sonuçları arasında da yüksek ilişki bulunmuştur. Tüm korelasyon değerleri $\alpha = 0,01$ için anlamlı çıkmıştır. Sadece (*) ile belirtilen değerler $\alpha = 0,05$ için istatistiksel anlamlı çıkmamışlardır. Tablo 3.3’ de ki En Yakın Komşuluk Kümeleme ile Medyan Kümeleme sonuçları aynı çıktığı görülmektedir. Yüksek ilişkili kümeleme yöntemleri olsa da, sonuçlar 2002 yılına ait sonuçlar kadar yüksek değildir. Tablo 3.4’ de ise yapılan kümeleme yöntemleri arasındaki korelasyon değerlerine baktığımızda ise Gruplar Arası Kümeleme ile Medyan Kümelemenin aynı sonuçlar verdiği, En Uzak Kümeleme ile Ward Kümelemenin yüksek korelasyon katsayısına sahip olduğu görülmektedir. Tüm korelasyon değerleri $\alpha = 0,01$ için istatistiksel anlamlı çıkmıştır.

Kümeleme analizi; gözlemleri benzerliklerine göre gruplamayı amaçlayan, çok değişkenli istatistiksel yöntemlerden en sık kullanılan yöntemdir. Amaç, gruplar arası heterojen, gruplar içi homojen yapıya en uygun sonuçları bulmaktır. Paket programların

artması ve hesaplama kolaylıkları sayesinde, kümeleme analizi hemen her alanda uygulanmaktadır.

Son yıllarda farklı yapıdaki veriler için farklı kümeleme analizi yöntemleri çalışılmaktadır. Bulanık Kümeleme, Soft Kümeleme, Gri Kümeleme, Rough Kümeleme gibi nispeten yeni ve alternatif yöntemler sayesinde uygun veri yapılarına daha uygun analizler yapılarak daha başarılı sonuçlar elde edilmesi amaçlanmaktadır.





KAYNAKLAR

- Akat Yiğit (2007). *Ülkelerin Askeri Benzerliklerine Göre Kümeleme Analizi Yardımıyla Sınıflandırılması*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Akgöz Elif (2010). *Türkiye’de Ticari Bankaların Finansal Oranlar Yardımıyla Sınıflandırılması: Kümeleme Analizi Yaklaşımı*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Kahramanmaraş.
- Aldenderfer M.S. ve Blashfield R.K., (1984). *Cluster Analysis*. Sage Univ. Papers Series, USA.
- Alkan Huri (2012). *Kümeleme Analizi İle Elektrik Tüketiminin Sınıflandırılması*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ocak 2012
- Anderberg M.R., (1973). *Cluster Analysis for applications*, *Academic Press*, New York. Page, s.(553–555)
- Atbaş Günay Azize Celile (2008). *Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ankara.
- Baday Yıldız Ezgi, Servi Uğur ve Berber Metin (2012). “Türkiye’deki illerin sosyo-ekonomik gelişmişlik sıralaması“ ,Erciyes Üniversitesi İİBF fakültesi dergisi, s (147-167) Kayseri.
- Blashfield, R.K. ve Aldenderfer, M.S., (1978). “The Literature on Cluster Analysis”, *Multivariate Behavioral Research*,13,p.271-295
- Bolshakova N. ve Azuaje F., (2003). *Cluster Validation Techniques for Genome Expression Data*, *Signal Processing*, 83(4), 825-833.
- Calinski R.B. ve Harabasz J., (1974). *A Dendrite Method For Cluster Analysis*, *Communications in Statistics 3, Theory and Methods*, 3, 1-27.
- Cebeci Zeynel, Yıldız Figen ve Kayaalp Tamer (2015) “K-Ortalamalar Kümelemesinde Optimum K Değeri Seçilmesi”, *2. Ulusal Yönetim Bilişim Sistemleri Kongresi, Bildiriler Kitabı* (Ed: Ü. Özen ve ark.), s. (231-242), Orka Ofset Matbaacılık, 8-10 Ekim 2015 Erzurum.

- Çakmak Zeki (1999). “Kümeleme Analizinde Geçerlilik Problemi ve Kümeleme Sonuçlarının Değerlendirilmesi”, *Dumlupınar Üniversitesi Sosyal Bilimler Akademi Matbaası Dergisi*, (3-Kasım), Isparta.
- Çakmak Zeki, Uzgören Nevin, Keçek Gülnur (2005). *Kümeleme Analizi Teknikleriyle İllerin Kültürel Yapıların Göre Sınıflandırılması ve Değişimlerinin İncelenmesi*, Dumlupınar üniversitesi, Isparta.
- Çam Selim (2014). *Veri Madenciliğinde Kümeleme Analizi Ve Sağlık Sektöründe Bir Uygulama*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Sivas.
- Darakçı Halil Çağdaş (2011). *Kümeleme Analizi Kullanılarak Benzin İstasyonlarının Operasyonel Değerlendirilmesi*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Dinçer Bülent ve Özaslan Metin (2004). İlçelerin sosyo ekonomik ve gelişmişlik sıralaması araştırması, T.C Başbakanlık Devlet Planlama Teşkilatı yayını, Ankara.
- Dinçer Bülent, Özaslan Metin ve Kavasoglu Taner (2003). İllerin sosyo ekonomik ve gelişmişlik sıralaması araştırması, T.C Başbakanlık Devlet Planlama Teşkilatı yayını, Ankara.
- Dinçer Kevser Setenay ve Özdamar Kazım (1992). Kümeleme çözümlemesinde uygun kümeleme ölçütlerinin karşılaştırılması, *Hacettepe Fen ve Mühendislik Bilimleri Dergisi* 14 s. (17–33), Ankara.
- Dinler Mehmet (2014). *Kümeleme Analizi Yöntemlerinin Hayvancılık Verilerinde Karşılaştırılmalı Olarak İncelenmesi*, Bingöl Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Bingöl.
- Doğan Barış (2008). *Bankaların Gözetiminde Bir Araç Olarak Kümeleme Analizi: Türk Bankacılık Sektörü İçin Uygulama*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, İstanbul.
- Ergün Ertuğrul (2008). *Ürün Kategorileri Arasındaki İlişkinin Birlikte Kuralları Ve Kümeleme Analizi İle Belirlenmesi ve Perakende Sektöründe Bir Uygulama*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, Afyon Karahisar.

- Erilli Alp Necati (2012). *Bulanık Sayıların Bulanık Kümeleme Analizinde Kullanımı ve Satranç Oyuncularının Sınıflandırılması*, Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, Samsun.
- Erilli Necati Alp, Yolcu Ufuk ve Eğrioğlu Erol ve Aladağ Ç. Hakan ve Öner Yüksel, (2011). Determining the Most Proper Number of Cluster in Fuzzy Clustering by Artificial Neural Networks. “*Expert Systems with Applications*”, 38, s.(2248-2252).
- Erilli Necati Alp (2009). *Kümeleme Analizine Bulanık Yaklaşım Algoritmaları Ve Uygulamaları*, On Dokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Samsun.
- Erişoğlu Murat (2011). *Uzaklık Ölçülerinin Kümeleme Analizine Olan Etkilerinin İncelenmesi ve Geliştirilmesi*, Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, Adana.
- Everitt B., (1978). *Cluster Analysis.*, London:Heinemann Educational Books Ltd. Blashfield, R.K., Aldenferder, M.S. “ The Literature on Cluster Analysis”, *Multivariate Behavioral Research*, 13, 271-295.
- Everitt B.S., (1979). Unresolved problems in cluster analysis, *Biometrics*, 35, s:(169-181)
- Fırat Oktay Seniye Ümit (1997). “Kümeleme Analizi: İstihdamın Sektörel Yapısı Açısından Avrupa Ülkelerinin Karşılaştırılması”, *İ.Ü. Sosyal Bilimler Dergisi*, *Cilt: III, Sayı:2*, s.(50-59), İstanbul.
- Flammang R.A., (1979). Economic Growth and Economic Development; Counterparts or Competitors?, *Economic Development and Cultural Change*, 28, (1): 50p.
- Forgy E. W., (1965). "Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classifications", *Biometrics*, *Cilt 21, No 3*, 768-769.
- Grabmeier J. ve Rudolph, A., (2002). Techniques of Cluster Algorithms In Data Mining, *data Mining and Knowledge Discovery* 6,303-360.

- Gül Yunus (2014). *2008 Yılı Küresel Kriz Sürecinde Türkiye'nin Maastricht Yakınlaşmasının Kümeleme Analizi Yöntemiyle İncelenmesi*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Malatya.
- Gün Ayşe Merve (2011). *Bulanık Kümeleme Analiziyle Görüntüdeki Yüzün Tanınması, Fen Bilimler Enstitüsü*, Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Hacıhasanoğlu Bilge (1980). *İller İçin Bir Gelişmişlik Göstergesi ve Sulama*, Başbakanlık DPT Yayını, Ankara.
- Han J. ve Kamber M., (2000). *Data Mining Concepts and Techniques*, M. K: Publishes, San Francisco, ABD.
- Hartigan A. ve Wong M. A., (1979). A K-means clustering algorithm. *Applied Statistics* 28, 100–108.
- Hartigan, J., (1975). *Clustering Algorithms*. John Wiley and Sons Ltd., USA.
- Hawkins D.M., (1982). *Topics in Applied Multivariate* . Cambridge University Pres. p.150-165
- Johnson A.R. ve Wichern D.W., (1988). *Applied Multivariate Statistical Analysis*. Prentice-Hall International Editions. New Jersey, p.554.
- Junhagen Sven (2000). “ From Complexity to Simplicity: on the Application of Three Tecniques for Multivariate Data Analysis”, MPP Working Paper No. 15/2000, Dec.
- Kalaycı Şeref (2005). *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*, Asil Yayıncılık, Ankara.
- Kaya Ayşe (2009). “*Türkiye’ de Bölgesel Net Mali Yansıma*”, T.C. Maliye Bakanlığı Strateji Geliştirme Başkanlığı, Ankara.
- Kılıç İbrahim, Saraçlı Sinan ve Kolukısaoglu Sıdika (2011). “Sosyo Ekonomik Göstergeler Bakımından İllerin Bölgesel Bazda Benzerliklerinin Çok Değişkenli Analizler İle İncelenmesi“, *İstatistikçiler Dergisi*, s.(57-68), Afyon.

- Köse Selçuk (2015). *Değerli Müşterilerde Ürün Kategorileri Arasındaki Satış İlişkilerinin Veri Madenciliği Yöntemlerinden Birliktelik Kuralları ve Kümeleme Analizi İle Belirlenmesi Ve Ulusal Bir Parekencide Örnek Uygulama*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Konya.
- Krzanowski W.J. ve Lai, Y.T., (1985). A criterion for determining the number of groups in a data set using sum of squares clustering. *Biometrics* 44, p.23-44.
- Krzanowski W.J., (1993). *Principles of Multivariate Analysis. A users perspective* clarendon pres. Oxford. p.150-178
- Kutsal Sakin (2008). *Türkiye'nin Lojistik Bir Üs Olabilme Potansiyelinin Kümeleme Analizi İle İncelenmesi*, Savunma Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ankara.
- LORR M., (1983). *Cluster Analysis for Social Sciences*. San Francisco: Jossey-Bass.
- Macqueen J., (1967). Some Methods for Classification and Analysis of Multivariate Observations. Fifth Berkeley Symposium on Mathematical Statistics and Probability, California Press Univ., California, USA.
- Manisalı Erol (1975). *Gelişme Ekonomisi*, İstanbul Üniversitesi İktisat Fakültesi Yayınları İstanbul.
- Mao J. ve Jain A. K., (1996). A Self-Organizing Network For Hyperellipsoidal Clustering (HEC). *IEEE Transactions on Neural Networks*, 7: 16-29.
- Marriot F.H.C., (1971). *Practical Problems in a method of Cluster Analysis*, *Biometrics*, 27.p.501-514.
- Nacaroğlu Engin (2010). *Deprem Etkisiyle Oluşan Boru Hasarlarının Coğrafi Bilgi Sistemleri (CBS) ve Kümeleme Analizi İle Değerlendirilmesi*, Fen Bilimleri Fakültesi, Yayınlanmamış Yüksek Lisans Tezi, Denizli.
- Nakip Mahir (2003). *Pazarlama Araştırmaları*, Seçkin Yayınevi, Ankara.
- Onan Aytuğ (2013). *Kümeleme Analizinde Melez Evrimsel Algoritmalar Üzerine Bir Çalışma*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, İzmir.

- Özdamar Kazım (2002). Paket Programlar ile İstatistiksel Veri Analizi 2, Kaan Kitabevi, Eskişehir.
- Özdamar Kazım (2004). Paket Programlar İle İstatistiksel Veri Analizi 5. Baskı, Kaan Kitabevi, Eskişehir.
- Özdamar Kazım (1999). Paket Programları İle İstatistiksel Veri Analizi, Kaan Kitabevi, Eskişehir.
- Özgüven Ali (1988). İktisadi Büyüme İktisadi Kalkınma Sosyal Kalkınma Planlama Ve Japon Kalkınması, Filiz Kitabevi, İstanbul.
- Öztürk Faruk (2012). *Kümeleme Analizi ve Uygulaması*, İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Rousseuw P.J., (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20, p. s:(53-65)
- Ruiz F.J.M., (1998). Strategic Group Analysis in Strategic Marketing: An application to Spanish Saving Banks, *Marketing intelligence & Planning*, 16/4. S:(1974), s.(1-27)
- Sayın Gökhan (2010). *Hibrid Ve İndirgenmiş Kümeleme Analizi İle Türkiye'deki İlçelerin Kümeleme Analizi İle Sınıflandırılması*, Yayınlanmamış Yüksek Lisans Tezi, Muğla.
- Seber G.A.F., (1984). *Multivariate Observation*, John Wiley and Sons, Inc. New York.
- Selanik Murat (2007). *Türk Tarımının Avrupa Birliği İçindeki Yerinin Kümeleme Analizi İle Belirlenmesi*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ankara.
- Servi Tayfun (2009). *Çok Değişkenli Karma Dağılım Modeline Dayalı Kümeleme Analizi*, Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, Adana.
- Silahtaroglu Gökhan (2004). *Veri Madenciliğinde Kümeleme Analizi Ve Öğretim Başarısının Değerlendirilmesine İlişkin Bir Uygulama*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, İstanbul.

- Şekerler Alper (2008). *Trafik Kaza Verilerinin Kümeleme Analizi Yöntemi İle İncelenmesi*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Denizli.
- Şentürk Aysan (1995). *Kümeleme Analizi Ve Bir Uygulama*, Yayınlanmamış Yüksek Lisans Tezi, Uludağ Üniversitesi S.(16-35), Bursa.
- Şimşek Dilek (2006). *Kümeleme Analizi, Çok Boyutlu Ölçekleme, Doğrulayıcı Ve Açıklayıcı Faktör Analizi İle Elde Edilen Yapı Gerçekliği Kanıtlarının Karşılaştırılması*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ankara.
- Tatlıdil Hüseyin (1992). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Engin Yayınları, Ankara.
- Tatlıdil Hüseyin (2002). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Ankara.
- Tatlıdil Hüseyin (1996). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Akademi Matbaası, Ankara.
- Ulutugay Gözde (2004). *Bulanık C- Ortalamalar Kümeleme Analizi Ve Uygulamaları*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, İzmir.
- Ünler Alper (2006). *Sezgisel Yöntemlerle K-Harmonik Ortalama Veri Kümeleme En İyilemesi*, fen bilimleri enstitüsü, Yayınlanmamış Doktora Tezi, Ankara.
- Üstünişik Naime Zerrin (2007). *Türkiye' deki İller ve Bölgeler Bazında Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması: Gri İlişkisel Analiz Yöntemi ve Uygulaması*, Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara.
- Yalçın Nazmiye (2013). *Kümeleme Analizi ve Uygulaması*, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Elazığ.
- Yaşar Ercan (2009). *Doğrudan Yabancı Yatırımcıların Belirleyicilerine Yönelik Kümeleme Analizi ve 1980—005 Dönemi Homojen Ülke Gruplarının Belirlenmesi*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, Kütahya.
- Yavilioğlu Cengiz (2002). *“Kalkınmanın Anlam bilimsel Tarihi ve Kavramsal Kökenleri”*, C.Ü. İktisadi ve İdari Bilimler Dergi, s (59-77) Sivas.

- Yazdaođlu Datev (2011). *Comparing Clustering Methods Based On Lifestyle Trends*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Yıldırım Serhat (2007). *Fuzzy Kümeleme Tekniđi Kullanarak Türkiye'deki Bir GSM Operatörü Kullanıcılarının Profili*, Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Yıldız İsmail (1998). *Kümeleme Analizi, Kümeleme Analizine Matematiksel Programlama Yaklaşımı Ve Bir Uygulama*, Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, Şanlıurfa.
- Yılmaz Ali (2013). *Sermaye Yeterlilik Oranlarının Kümeleme Analizi ve Basel III Kriterlerinin Türk Bankacılık Sektörü Sermaye Yeterliliđi Üzerine Etkileri*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Yılmaz Ünzile (2011). *Türkiye'de İllerin Sosyo-Ekonomik Düzeylerinin Faktör Analizi Ve Kümeleme Analizi İle İncelenmesi*, Sosyal Bilimler Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Trabzon.

EKLER

Ek 1. 2002 yılı Sosyo ekonomik verilerinin gruplar arası yöntemi ile kümeleme sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	ELAZIĞ	MERSİN	ANKARA	BURSA	KOCAELİ	İSTANBUL
ADYAMAN	ERZİNCAN	MUĞLA		İZMİR		
AFYON	ERZURUM	MUŞ				
AĞRI	ESKİŞEHİR	NEVŞEHİR				
AKSARAY	GAZİANTEP	NİĞDE				
AMASYA	GİRESUN	ORDU				
ANTALYA	GÜMÜŞHANE	OSMANIYE				
ARDAHAN	HAKKARİ	RİZE				
ARTVİN	HATAY	SAKARYA				
AYDIN	IĞDIR	SAMSUN				
BALIKESİR	ISPARTA	SİİRT				
BARTIN	K.MARAŞ	SİNOP				
BATMAN	KARABÜK	SİVAS				
BAYBURT	KARAMAN	ŞANLIURFA				
BİLECİK	KARS	ŞIRNAK				
BİNGÖL	KASTAMONU	TEKİRDAĞ				
BİTLİS	KAYSERİ	TOKAT				
BOLU	KIRIKKALE	TRABZON				
BURDUR	KIRKLARELİ	TUNCELİ				
ÇANAKKALE	KIRŞEHİR	UŞAK				
ÇANKIRI	KİLİS	VAN				
ÇORUM	KONYA	YALOVA				
DENİZLİ	KÜTAHYA	YOZGAT				
DİYARBAKIR	MALATYA	ZONGULDAK				
DÜZCE	MANİSA					
EDİRNE	MARDİN					

Ek 2. 2002 yılı Sosyo ekonomik verilerin Grup İçi yöntemi ile kümeleme sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	ESKİŞEHİR	OSMANİYE	AĞRI	ANKARA	İSTANBUL	KOCAELİ
ADIYAMAN	GAZİANTEP	RİZE	BATMAN			
AFYON	GÜMÜŞHANE	SAKARYA	BİNGÖL			
AKSARAY	HATAY	SAMSUN	BİTLİS			
AMASYA	İSPARTA	SINOP	DİYARBAKIR			
ANTALYA	İZMİR	SİVAS	GİRESUN			
ARDAHAN	K.MARAŞ	TEKİRDAĞ	HAKKARİ			
ARTVİN	KARABÜK	TOKAT	İĞDIR			
AYDIN	KARAMAN	TRABZON	MARDİN			
BALIKESİR	KARS	TUNCELİ	MUŞ			
BARTIN	KASTAMONU	UŞAK	SİİRT			
BAYBURT	KAYSERİ	YALOVA	ŞANLIURFA			
BİLECİK	KIRIKKALE	YOZGAT	ŞIRNAK			
BOLU	KIRKLARELİ	ZONGULDAK	VAN			
BURDUR	KIRŞEHİR					
BURSA	KİLİS					
ÇANAKKALE	KONYA					
ÇANKIRI	KÜTAHYA					
ÇORUM	MALATYA					
DENİZLİ	MANİSA					
DÜZCE	MERSİN					
EDİRNE	MUĞLA					
ELAZIĞ	NEVŞEHİR					
ERZİNCAN	NİĞDE					
ERZURUM	ORDU					

Ek 3. 2002 yılı Sosyo ekonomik verilerin nearest neighbor (en yakın komşuluk) yöntem ile kümeleme sonuçları

1.KÜME				2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	DÜZCE	MALATYA	TRABZON	ANKARA	GİRESUN	İSTANBUL	KOCAELİ
ADIYAMAN	EDİRNE	MANİSA	TUNCELİ				
AFYON	ELAZIĞ	MARDİN	UŞAK				
AĞRI	ERZİNCAN	MERSİN	VAN				
AKSARAY	ERZURUM	MUĞLA	YALOVA				
AMASYA	ESKİŞEHİR	MUŞ	YOZGAT				
ANTALYA	GAZİANTEP	NEVŞEHİR	ZONGULDAK				
ARDAHAN	GÜMÜŞHANE	NİĞDE					
ARTVİN	HAKKARİ	ORDU					
AYDIN	HATAY	OSMANİYE					
BALIKESİR	İĞDIR	RİZE					
BARTIN	İSPARTA	SAKARYA					
BATMAN	İZMİR	SAMSUN					
BAYBURT	K.MARAŞ	SİİRT					
BİLECİK	KARABÜK	SİNOP					
BİNGÖL	KARAMAN	SİVAS					
BİTLİS	KARS	ŞANLIURFA					
BOLU	KASTAMONU	ŞIRNAK					
BURDUR	KAYSERİ	TEKİRDAĞ					
BURSA	KIRIKKALE	TOKAT					
ÇANAKKALE	KIRKLARELİ						
ÇANKIRI	KIRŞEHİR						
ÇORUM	KİLİS						
DENİZLİ	KONYA						
DIYARBAKIR	KÜTAHYA						

Ek 4.2002 yılı Sosyo ekonomik verilerin furhest neighbor (en uzak komşuluk) yöntemi ile kümeleme sonuçları

1.KÜME	2.KÜME		3.KÜME	4.KÜME	5.KÜME
ADANA	ADİYAMAN	K.MARAŞ	ANKARA	BURSA	İSTANBUL
ANTALYA	AFYON	KARABÜK		İZMİR	
AYDIN	AĞRI	KARAMAN		KOCAELİ	
BALIKESİR	AKSARAY	KARS			
BİLECİK	AMASYA	KASTAMONU			
BOLU	ARDAHAN	KIRIKKALE			
DENİZLİ	ARTVİN	KIRŞEHİR			
ESKİŞEHİR	BARTIN	KİLİS			
GAZİANTEP	BATMAN	KÜTAHYA			
HATAY	BAYBURT	MALATYA			
KAYSERİ	BİNGÖL	MARDİN			
KIRKLARELİ	BİTLİS	MUŞ			
KONYA	BURDUR	NEVŞEHİR			
MANİSA	ÇANAKKALE	NİĞDE			
MERSİN	ÇANKIRI	ORDU			
MUĞLA	ÇORUM	OSMANİYE			
SAKARYA	DİYARBAKIR	RİZE			
SAMSUN	DÜZCE	SİİRT			
TEKİRDAĞ	EDİRNE	SİNOP			
YALOVA	ELAZIĞ	SİVAS			
ZONGULDAK	ERZİNCAN	ŞANLIURFA			
	ERZURUM	ŞIRNAK			
	GİRESUN	TOKAT			
	GÜMÜŞHANE	TRABZON			
	HAKKARİ	TUNCELİ			
	İĞDIR	UŞAK			
	ISPARTA	VAN			
		YOZGAT			

Ek 5. 2002 yılı Sosyo ekonomik verilerin centroid (clustering) kümeleme yöntemi ile sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	ELAZIĞ	MERSİN	ANKARA	BURSA	İSTANBUL	KOCAELİ
ADYAMAN	ERZİNCAN	MUĞLA		İZMİR		
AFYON	ERZURUM	MUŞ				
AĞRI	ESKİŞEHİR	NEVŞEHİR				
AKSARAY	GAZİANTEP	NİĞDE				
AMASYA	GİRESUN	ORDU				
ANTALYA	GÜMÜŞHANE	OSMANİYE				
ARDAHAN	HAKKARİ	RİZE				
ARTVİN	HATAY	SAKARYA				
AYDIN	İĞDIR	SAMSUN				
BALIKESİR	İSPARTA	SİİRT				
BARTIN	K.MARAŞ	SİNOP				
BATMAN	KARABÜK	SİVAS				
BAYBURT	KARAMAN	ŞANLIURFA				
BİLECİK	KARS	ŞIRNAK				
BİNGÖL	KASTAMONU	TEKİRDAĞ				
BİTLİS	KAYSERİ	TOKAT				
BOLU	KIRIKKALE	TRABZON				
BURDUR	KIRKLARELİ	TUNCELİ				
ÇANAKKALE	KIRŞEHİR	UŞAK				
ÇANKIRI	KİLİS	VAN				
ÇORUM	KONYA	YALOVA				
DENİZLİ	KÜTAHYA	YOZGAT				
DİYARBAKIR	MALATYA	ZONGULDAK				
DÜZCE	MANİSA					
EDİRNE	MARDİN					

Ek 6. 2002 yılı Sosyo ekonomik verilerin median(clustering) kümeleme yöntemi ile sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	ESKİŞEHİR	OSMANIYE	AĞRI	ANKARA	İSTANBUL	KOCAELİ
ADIYAMAN	GAZİANTEP	RİZE	ARTVİN			
AFYON	HATAY	SAKARYA	BATMAN			
AKSARAY	IĞDIR	SAMSUN	BİNGÖL			
AMASYA	ISPARTA	SİNOP	BİTLİS			
ANTALYA	İZMİR	SİVAS	GİRESUN			
ARDAHAN	K.MARAŞ	TEKİRDAĞ	GÜMÜŞHANE			
AYDIN	KARABÜK	TOKAT	HAKKARİ			
BALIKESİR	KARAMAN	TRABZON	MARDİN			
BARTIN	KARS	UŞAK	MUŞ			
BAYBURT	KASTAMONU	YALOVA	SİİRT			
BİLECİK	KAYSERİ	YOZGAT	ŞANLIURFA			
BOLU	KIRIKKALE	ZONGULDAK	ŞIRNAK			
BURDUR	KIRKLARELİ		TUNCELİ			
BURSA	KIRŞEHİR		VAN			
ÇANAKKALE	KİLİS					
ÇANKIRI	KONYA					
ÇORUM	KÜTAHYA					
DENİZLİ	MALATYA					
DİYARBAKIR	MANİSA					
DÜZCE	MERSİN					
EDİRNE	MUĞLA					
ELAZIĞ	NEVŞEHİR					
ERZİNCAN	NİĞDE					
ERZURUM	ORDU					

Ek 7. 2002 yılı Sosyo ekonomik verilerin ward's method yöntem ile kümeleme analizi sonuçları

1.KÜME	2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	BATMAN	ARTVİN	TRABZON	ANKARA
ANTALYA	BİNGÖL	BARTIN	TUNCELİ	İSTANBUL
AYDIN	BİTLİS	BAYBURT	YOZGAT	İZMİR
BALIKESİR	DIYARBAKIR	ÇANKIRI		KIRKLARELİ
BOLU	HAKKARİ	ÇORUM		KOCAELİ
BURDUR	İĞDIR	DÜZCE		TEKİRDAĞ
ÇANAKKALE	KARS	ERZİNCAN		YALOVA
DENİZLİ	MARDİN	ERZURUM		
EDİRNE	MUŞ	GİRESUN		
ELAZIĞ	SİİRT	GÜMÜŞHANE		
ESKİŞEHİR	ŞANLIURFA	K.MARAŞ		
GAZİANTEP	ŞIRNAK	KASTAMONU		
HATAY	VAN	KIRŞEHİR		
ISPARTA		KİLİS		
KARABÜK		KÜTAHYA		
KARAMAN		MALATYA		
KAYSERİ		NEVŞEHİR		
KIRIKKALE		NİĞDE		
KONYA		ORDU		
MANİSA		OSMANİYE		
MERSİN		RİZE		
MUĞLA		SAMSUN		
SAKARYA		SİNOP		
UŞAK		SİVAS		
ZONGULDAK		TOKAT		

Ek 8. 2008 yılı sosyo ekonomik verilerinin Gruplar Arası yöntemi ile kümelenme sonuçları

1.KÜME	2.KÜME			3.KÜME	4.KÜME	5.KÜME
ADANA	ADİYAMAN	ERZURUM	NEVŞEHİR	ANKRA	ANTALYA	İSTANBUL
BURSA	AFYON	ESKİŞEHİR	NİĞDE			
İZMİR	AĞRI	GAZİANTEP	ORDU			
KONYA	AKSARAY	GİRESUN	OSMANİYE			
MERSİN	AMASYA	GÜMÜŞHANE	RİZE			
	ARDAHAN	HAKKARİ	SAKARYA			
	ARTVİN	HATAY	SAMSUN			
	AYDIN	İĞDIR	SİİRT			
	BALIKESİR	ISPARTA	SİNOP			
	BARTIN	K.MARAŞ	SİVAS			
	BATMAN	KARABÜK	ŞANLIURFA			
	BAYBURT	KARAMAN	ŞIRNAK			
	BİLECİK	KARS	TEKİRDAĞ			
	BİNGÖL	KASTAMONU	TOKAT			
	BİTLİS	KAYSERİ	TRABZON			
	BOLU	KIRIKKALE	TUNCELİ			
	BURDUR	KIRKLARELİ	UŞAK			
	ÇANAKKALE	KIRŞEHİR	VAN			
	ÇANKIRI	KİLİS	YALOVA			
	ÇORUM	KOCAELİ	YOZGAT			
	DENİZLİ	KÜTAHYA	ZONGULDAK			
	DİYARBAKIR	MALATYA				
	DÜZCE	MANİSA				
	EDİRNE	MARDİN				
	ELAZIĞ	MUĞLA				
	ERZİNCAN	MUŞ				

Ek 9. 2008 yılı sosyo ekonomik verilerinin Grup İçi yöntemi ile kümelenmesi sonuçları

1.KÜME	2.KÜME			3.KÜME	4.KÜME	5.KÜME
ADANA	ADYAMAN	HATAY	TOKAT	AĞRI	ANKARA	
MERSİN	AFYON	ISPARTA	TRABZON	AKSARAY		
	AMASYA	İZMİR	UŞAK	BATMAN		
	ANTALYA	K.MARAŞ	YOZGAT	BİNGÖL		
	ARDAHAN	KARABÜK	NEVŞEHİR	BİTLİS		
	ARTVİN	KARAMAN		DİYARBAKIR		
	AYDIN	KARS		GAZİANTEP		
	BALIKESİR	KASTAMONU		HAKKARİ		
	BARTIN	KAYSERİ		İĞDIR		
	BAYBURT	KIRIKKALE		KİLİS		
	BİLECİK	KIRKLARELİ		KONYA		
	BOLU	KİRŞEHİR		MARDİN		
	BURDUR	KOCAELİ		MUŞ		
	BURSA	KÜTAHYA		OSMANİYE		
	ÇANAKKALE	MALATYA		SİİRT		
	ÇANKIRI	MANİSA		ŞANLIURFA		
	ÇORUM	MUĞLA		ŞIRNAK		
	DENİZLİ	NEVŞEHİR		TUNCELİ		
	DÜZCE	NİĞDE		VAN		
	EDİRNE	ORDU		YALOVA		
	ELAZIĞ	RİZE				
	ERZİNCAN	SAKARYA				
	ERZURUM	SAMSUN				
	ESKİŞEHİR	SİNOP				
	GİRESUN	SİVAS				
	GÜMÜŞHANE	TEKİRDAĞ				

Ek 10. 2008 yılı sosyo ekonomik verilerinin nearest neighbor(en yakın komşuluk) yöntemi ile kümelenmesi sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	ELAZIĞ	MARDİN	ANKARA	ANTALYA	ESKİŞEHİR	İSTANBUL
ADYAMAN	ERZİNCAN	MERSİN				
AFYON	ERZURUM	MUĞLA				
AĞRI	GAZİANTEP	MUŞ				
AKSARAY	GİRESUN	NEVŞEHİR				
AMASYA	GÜMÜŞHANE	NİĞDE				
ARDAHAN	HAKKARİ	ORDU				
ARTVİN	HATAY	OSMANİYE				
AYDIN	İĞDIR	RİZE				
BALIKESİR	ISPARTA	SAKARYA				
BARTIN	İZMİR	SAMSUN				
BATMAN	K.MARAŞ	SİİRT				
BAYBURT	KARABÜK	SİNOP				
BİLECİK	KARAMAN	SİVAS				
BİNGÖL	KARS	ŞANLIURFA				
BİTLİS	KASTAMONU	ŞIRNAK				
BOLU	KAYSERİ	TEKİRDAĞ				
BURDUR	KIRIKKALE	TOKAT				
BURSA	KIRKLARELİ	TRABZON				
ÇANAKKALE	KIRŞEHİR	TUNCELİ				
ÇANKIRI	KİLİS	UŞAK				
ÇORUM	KOCAELİ	VAN				
DENİZLİ	KONYA	YALOVA				
DİYARBAKIR	KÜTAHYA	YOZGAT				
DÜZCE	MALATYA	ZONGULDAK				
EDİRNE	MANİSA					

Ek 11. 2008 yılı sosyo ekonomik verilerinin furhest neighbor (en uzak komşuluk) yöntemi ile kümelmesi

1.KÜME	2.KÜME			3.KÜME	4.KÜME	5.KÜME
ADANA	ADİYAMAN	ERZURUM	NEVŞEHİR	ANKARA	ANTALYA	İSTANBUL
MERSİN	AFYON	ESKİŞEHİR	NİĞDE		BURSA	
	AĞRI	GAZİANTEP	ORDU		İZMİR	
	AKSARAY	GİRESUN	OSMANİYE		KONYA	
	AMASYA	GÜMÜŞHANE	RİZE			
	ARDAHAN	HAKKARİ	SAKARYA			
	ARTVİN	HATAY	SAMSUN			
	AYDIN	İĞDIR	SİİRT			
	BALIKESİR	ISPARTA	SİNOP			
	BARTIN	K.MARAŞ	SİVAS			
	BATMAN	KARABÜK	ŞANLIURFA			
	BAYBURT	KARAMAN	ŞIRNAK			
	BİLECİK	KARS	TEKİRDAĞ			
	BİNGÖL	KASTAMONU	TOKAT			
	BİTLİS	KAYSERİ	TRABZON			
	BOLU	KIRIKKALE	TUNCELİ			
	BURDUR	KIRKLARELİ	UŞAK			
	ÇANAKKALE	KIRŞEHİR	VAN			
	ÇANKIRI	KİLİS	YALOVA			
	ÇORUM	KOCAELİ	YOZGAT			
	DENİZLİ	KÜTAHYA	ZONGULDAK			
	DIYARBAKIR	MALATYA				
	DÜZCE	MANİSA				
	EDİRNE	MARDİN				
	ELAZIĞ	MUĞLA				
	ERZİNCAN	MUŞ				

Ek 12. 2008 yılı sosyo ekonomik verilerinin centroid yöntemi ile kümelenmesi sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	GAZİANTEP	RİZE	ANKARA	ANTALYA	İSTANBUL	KONYA
ADİYAMAN	GİRESUN	SAKARYA				
AFYON	GÜMÜŞHANE	SAMSUN				
AĞRI	HAKKARİ	SİİRT				
AKSARAY	HATAY	SİNOP				
AMASYA	İĞDIR	SİVAS				
ARDAHAN	İSPARTA	ŞANLIURFA				
ARTVİN	İZMİR	ŞIRNAK				
AYDIN	K.MARAŞ	TEKİRDAĞ				
BALIKESİR	KARABÜK	TOKAT				
BARTIN	KARAMAN	TRABZON				
BATMAN	KARS	TUNCELİ				
BAYBURT	KASTAMONU	UŞAK				
BİLECİK	KAYSERİ	VAN				
BİNGÖL	KIRIKKALE	YALOVA				
BİTLİS	KIRKLARELİ	YOZGAT				
BOLU	KİRŞEHİR	ZONGULDAK				
BURDUR	KİLİS					
BURSA	KOCAELİ					
ÇANAKKALE	KÜTAHYA					
ÇANKIRI	MALATYA					
ÇORUM	MANİSA					
DENİZLİ	MARDİN					
DİYARBAKIR	MERSİN					
DÜZCE	MUĞLA					
EDİRNE	MUŞ					
ELAZIĞ	NEVŞEHİR					
ERZİNCAN	NİĞDE					
ERZURUM	ORDU					
ESKİŞEHİR	OSMANİYE					

Ek 13. 2008 yılı sosyo ekonomik verilerinin median yöntemi ile kümelenmesi sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	GAZİANTEP	RİZE	ANKARA	ANTALYA	ESKİŞEHİR	İSTANBUL
ADİYAMAN	GİRESUN	SAKARYA				
AFYON	GÜMÜŞHANE	SAMSUN				
AĞRI	HAKKARİ	SİİRT				
AKSARAY	HATAY	SİNOP				
AMASYA	İĞDIR	SİVAS				
ARDAHAN	ISPARTA	ŞANLIURFA				
ARTVİN	İZMİR	ŞIRNAK				
AYDIN	K.MARAŞ	TEKİRDAĞ				
BALIKESİR	KARABÜK	TOKAT				
BARTIN	KARAMAN	TRABZON				
BATMAN	KARS	TUNCELİ				
BAYBURT	KASTAMONU	UŞAK				
BİLECİK	KAYSERİ	VAN				
BİNGÖL	KIRIKKALE	YALOVA				
BİTLİS	KIRKLARELİ	YOZGAT				
BOLU	KIRŞEHİR	ZONGULDAK				
BURDUR	KİLİS					
BURSA	KOCAELİ					
ÇANAKKALE	KÜTAHYA					
ÇANKIRI	MALATYA					
ÇORUM	MANİSA					
DENİZLİ	MARDİN					
DİYARBAKIR	MERSİN					
DÜZCE	MUĞLA					
EDİRNE	MUŞ					
ELAZIĞ	NEVŞEHİR					
ERZİNCAN	NİĞDE					
ERZURUM	ORDU					
KONYA	OSMANİYE					

Ek 14. 2008 yılı sosyo ekonomik verilerinin ward's tekniđi ile kmelenmesi sonuları

1.KME	2.KME	3.KME	4.KME	5.KME
ADANA	ADİYAMAN	KARABK	AĐRI	ANKARA
ANTALYA	AFYON	KARAMAN	AKSARAY	İSTANBUL
BURSA	AMASYA	KARS	BATMAN	
İZMİR	ARDAHAN	KASTAMONU	BİNGL	
KONYA	ARTVİN	KAYSERİ	BİTLİS	
MERSİN	AYDIN	KIRIKKALE	DİYARBAKIR	
	BALIKESİR	KIRKLARELİ	HAKKARİ	
	BARTIN	KİRŞEHİR	İĐDIR	
	BAYBURT	KOCAELİ	KİLİS	
	BİLECİK	KTAHYA	MARDİN	
	BOLU	MALATYA	MUŞ	
	BURDUR	MANİSA	OSMANİYE	
	ANAKKALE	MUĐLA	SİİRT	
	ANKIRI	NEVŞEHİR	ŞANLIURFA	
	ORUM	NİĐDE	ŞIRNAK	
	DENİZLİ	ORDU	TUNCELİ	
	DZCE	RİZE	VAN	
	EDİRNE	SAKARYA		
	ELAZIĐ	SAMSUN		
	ERZİNCAN	ŞİNOP		
	ERZURUM	ŞİVAS		
	ESKİŞEHİR	TEKİRDAĐ		
	GAZİANTEP	TOKAT		
	GİRESUN	TRABZON		
	GMŞHANE	UŞAK		
	HATAY	YALOVA		
	İSPARTA	YOZGAT		
	K.MARAŞ	ZONGULDAK		

Ek 15. 2013 yılı sosyo ekonomik verilerin Gruplar Arası yöntemi ile kümelenmesi sonuçları

1.KÜME	2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	EDİRNE	MANİSA	ANKARA	ANTALYA
ADİYAMAN	ELAZIĞ	MARDİN	İZMİR	ESKİŞEHİR
AFYON	ERZİNCAN	MERSİN	KONYA	İSTANBUL
AĞRI	ERZURUM	MUĞLA		
AKSARAY	GAZİANTEP	MUŞ		
AMASYA	GİRESUN	NEVŞEHİR		
ARDAHAN	GÜMÜŞHANE	NİĞDE		
ARTVİN	HAKKARİ	ORDU		
AYDIN	HATAY	OSMANİYE		
BALIKESİR	IĞDIR	RİZE		
BARTIN	ISPARTA	SAKARYA		
BATMAN	K.MARAŞ	SAMSUN		
BAYBURT	KARABÜK	SİİRT		
BİLECİK	KARAMAN	SİNOP		
BİNGÖL	KARS	SİVAS		
BİTLİS	KASTAMONU	ŞANLIURFA		
BOLU	KAYSERİ	ŞIRNAK		
BURDUR	KIRIKKALE	TEKİRDAĞ		
BURSA	KIRKLARELİ	TOKAT		
ÇANAĞKALE	KIRŞEHİR	TRABZON		
ÇANKIRI	KİLİS	TUNCELİ		
ÇORUM	KOCAELİ	UŞAK		
DENİZLİ	KÜTAHYA	VAN		
DİYARBAKIR	MALATYA	YALOVA		
DÜZCE	MANİSA	YOZGAT		
		ZONGULDAK		

Ek 16. 2013 yılı sosyo ekonomik verilerin Grup İçi yöntemi ile kümelmesi sonuçları

1.KÜME	2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	ADİYAMAN	K.MARAŞ	AĞRI	ANTALYA
AYDIN	AFYON	KARABÜK	BATMAN	İSTANBUL
BALIKESİR	AKSARAY	KARAMAN	BİNGÖL	
BURSA	AMASYA	KASTAMONU	BİTLİS	
DENİZLİ	ANKARA	KIRIKKALE	DİYARBAKIR	
İZMİR	ARDAHAN	KIRKLARELİ	HAKKARİ	
KAYSERİ	ARTVİN	KIRŞEHİR	İĞDIR	
KONYA	BARTIN	KİLİS	KARS	
MANİSA	BAYBURT	KOCAELİ	MARDİN	
MERSİN	BİLECİK	KÜTAHYA	MUŞ	
MUĞLA	BOLU	MALATYA	SİİRT	
SAKARYA	BURDUR	NEVŞEHİR	ŞANLIURFA	
SAMSUN	ÇANAKKALE	NİĞDE	ŞIRNAK	
	ÇANKIRI	ORDU	TUNCELİ	
	ÇORUM	OSMANİYE	VAN	
	DÜZCE	RİZE		
	EDİRNE	SİNOP		
	ELAZIĞ	SİVAS		
	ERZİNCAN	TEKİRDAĞ		
	ERZURUM	TOKAT		
	ESKİŞEHİR	TRABZON		
	GAZİANTEP	UŞAK		
	GİRESUN	YALOVA		
	GÜMÜŞHANE	YOZGAT		
	HATAY	ZONGULDAK		
	ISPARTA			

Ek 17. 2013 yılı sosyo ekonomik verilerinin nearest neighbor (en yakın komşuluk) yöntemi ile kümelmesi sonuçları

1.KÜME	2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	EDİRNE	MARDİN	ANTALYA	ESKİŞEHİR
ADYAMAN	ELAZIĞ	MERSİN		İSTANBUL
AFYON	ERZİNCAN	MUĞLA		KONYA
AĞRI	ERZURUM	MUŞ		
AKSARAY	GAZİANTEP	NEVŞEHİR		
AMASYA	GİRESUN	NİĞDE		
ANKARA	GÜMÜŞHANE	ORDU		
ARDAHAN	HAKKARİ	OSMANİYE		
ARTVİN	HATAY	RİZE		
AYDIN	İĞDIR	SAKARYA		
BALIKESİR	ISPARTA	SAMSUN		
BARTIN	İZMİR	SİİRT		
BATMAN	K.MARAŞ	SİNOP		
BAYBURT	KARABÜK	SİVAS		
BİLECİK	KARAMAN	ŞANLIURFA		
BİNGÖL	KARS	ŞIRNAK		
BİTLİS	KASTAMONU	TEKİRDAĞ		
BOLU	KAYSERİ	TOKAT		
BURDUR	KIRIKKALE	TRABZON		
BURSA	KIRKLARELİ	TUNCELİ		
ÇANAKKALE	KIRŞEHİR	UŞAK		
ÇANKIRI	KİLİS	VAN		
ÇORUM	KOCAELİ	YALOVA		
DENİZLİ	KÜTAHYA	YOZGAT		
DİYARBAKIR	MALATYA	ZONGULDAK		
DÜZCE	MANİSA			

Ek 18. 2013 yılı sosyo ekonomik verilerinin furhest neighbor (en uzak komşuluk) yöntemi ile kümelenmesi sonuçları

1.KÜME		2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	KARAMAN	AĞRI	ANKARA	ANTALYA	İSTANBUL
ADYAMAN	KASTAMONU	ARDAHAN	İZMİR		
AFYON	KAYSERİ	BATMAN	KONYA		
AKSARAY	KIRIKKALE	BİNGÖL			
AMASYA	KIRKLARELİ	BİTLİS			
ARTVİN	KİRŞEHİR	ÇANKIRI			
AYDIN	KİLİS	DİYARBAKIR			
BALIKESİR	KOCAELİ	GÜMÜŞHANE			
BARTIN	KÜTAHYA	HAKKARİ			
BAYBURT	MALATYA	İĞDIR			
BİLECİK	MANİSA	KARS			
BOLU	MERSİN	MARDİN			
BURDUR	MUĞLA	MUŞ			
BURSA	NEVŞEHİR	SİİRT			
ÇANAKKALE	NİĞDE	ŞANLIURFA			
ÇORUM	ORDU	ŞIRNAK			
DENİZLİ	OSMANİYE	TUNCELİ			
DÜZCE	RİZE	VAN			
EDİRNE	SAKARYA				
ELAZIĞ	SAMSUN				
ERZİNCAN	SİNOP				
ERZURUM	SİVAS				
ESKİŞEHİR	TEKİRDAĞ				
GAZİANTEP	TOKAT				
GİRESUN	TRABZON				
HATAY	UŞAK				
ISPARTA	YALOVA				
K.MARAŞ	YOZGAT				
KARABÜK	ZONGULDAK				

Ek 19. 2013 yılı sosyo ekonomik verilerinin centroid clustering yöntemi ile kümelenmesi sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	EDİRNE	MANİSA	ANKARA	ANTALYA	İSTANBUL	KONYA
ADİYAMAN	ELAZIĞ	MARDİN	İZMİR			
AFYON	ERZİNCAN	MERSİN				
AĞRI	ERZURUM	MUĞLA				
AKSARAY	ESKİŞEHİR	MUŞ				
AMASYA	GAZİANTEP	NEVŞEHİR				
ARDAHAN	GİRESUN	NİĞDE				
ARTVİN	GÜMÜŞHANE	ORDU				
AYDIN	HAKKARİ	OSMANİYE				
BALIKESİR	HATAY	RİZE				
BARTIN	İĞDIR	SAKARYA				
BATMAN	ISPARTA	SAMSUN				
BAYBURT	K.MARAŞ	SİİRT				
BİLECİK	KARABÜK	SİNOP				
BİNGÖL	KARAMAN	SİVAS				
BİTLİS	KARS	ŞANLIURFA				
BOLU	KASTAMONU	ŞIRNAK				
BURDUR	KAYSERİ	TEKİRDAĞ				
BURSA	KIRIKKALE	TOKAT				
ÇANAKKALE	KIRKLARELİ	TRABZON				
ÇANKIRI	KIRŞEHİR	TUNCELİ				
ÇORUM	KİLİS	UŞAK				
DENİZLİ	KOCAELİ	VAN				
DIYARBAKIR	KÜTAHYA	YALOVA				
DÜZCE	MALATYA	YOZGAT				
		ZONGULDAK				

Ek 20. 2013 yılı sosyo ekonomik verilerinin median clustering yöntemi ile kümelenmesi sonuçları

1.KÜME	2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	ANKARA	ANTALYA	ESKİŞEHİR	İSTANBUL
ADİYAMAN	İZMİR			
AFYON	KONYA			
AĞRI				
AKSARAY				
AMASYA				
ARDAHAN				
ARTVİN				
AYDIN				
BALIKESİR				
BARTIN				
BATMAN				
BAYBURT				
BİLECİK				
BİNGÖL				
BİTLİS				
BOLU				
BURDUR				
BURSA				
ÇANAKKALE				
ÇANKIRI				
ÇORUM				
DENİZLİ				
DİYARBAKIR				
DÜZCE				
EDİRNE				
ELAZIĞ	MUĞLA			
ERZİNCAN	MUŞ			
ERZURUM	NEVŞEHİR			
GAZİANTEP	NİĞDE			
GİRESUN	ORDU			
GÜMÜŞHANE	OSMANIYE			
HAKKARİ	RİZE			
HATAY	SAKARYA			
İĞDIR	SAMSUN			
ISPARTA	SİİRT			
K.MARAŞ	SİNOP			
KARABÜK	SİVAS			
KARAMAN	ŞANLIURFA			
KARS	ŞIRNAK			
KASTAMONU	TEKİRDAĞ			
KAYSERİ	TOKAT			
KIRIKKALE	TRABZON			
KIRKLARELİ	TUNCELİ			
KIRŞEHİR	UŞAK			
KİLİS	VAN			
KOCAELİ	YALOVA			
KÜTAHYA	YOZGAT			
MALATYA	ZONGULDAK			
MANİSA				
MARDİN				
MERSİN				

Ek 21. 2013 yılı sosyo ekonomik verilerinin ward's yöntemi ile kümelenmesi sonuçları

1.KÜME			2.KÜME	3.KÜME	4.KÜME	5.KÜME
ADANA	GİRESUN	SAMSUN	AĞRI	ANKARA	BATMAN	İSTANBUL
ADIYAMAN	GÜMÜŞHANE	SİNOP	ARDAHAN	ANTALYA	DIYARBAKIR	
AFYON	HATAY	SİVAS	BİNGÖL	İZMİR	MARDİN	
AKSARAY	ISPARTA	TEKİRDAĞ	BİTLİS	KONYA	SİİRT	
AMASYA	K.MARAŞ	TOKAT	HAKKARİ		ŞANLIURFA	
ARTVİN	KARABÜK	TRABZON	İĞDIR		ŞIRNAK	
AYDIN	KARAMAN	UŞAK	KARS			
BALIKESİR	KASTAMONU	YALOVA	MUŞ			
BARTIN	KAYSERİ	YOZGAT	TUNCELİ			
BAYBURT	KIRIKKALE	ZONGULDAK	VAN			
BİLECİK	KIRKLARELİ	ELAZIĞ				
BOLU	KIRŞEHİR	ERZİNCAN				
BURDUR	KİLİS	ERZURUM				
BURSA	KOCAELİ	ESKİŞEHİR				
ÇANAKKALE	KÜTAHYA	GAZİANTEP				
ÇANKIRI	MALATYA	NİĞDE				
ÇORUM	MANİSA	ORDU				
DENİZLİ	MERSİN	OSMANİYE				
DÜZCE	MUĞLA	RİZE				
EDİRNE	NEVŞEHİR	SAKARYA				

Ek 22. 2002 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri

2002 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri							
	Gruplar Arası	Gruplar İçi	En Yakın Kom.	En Uzak Kom.	Centroid	Medyan	Ward
Gruplar Arası	1	,688	,805	,756	1,000	,680	,450
Gruplar İçi	,688	1	,808	,569	,688	,932	,219*
En Yakın Kom.	,805	,808	1	,560	,805	,799	,316
En Uzak Kom.	,756	,569	,560	1	,756	,569	,480
Centroid	1,000	,688	,805	,756	1	,680	,450
Medyan	,680	,932	,799	,569	,680	1	,258*
Ward	,450	,219*	,316	,480	,450	,258*	1

Ek 23. 2008 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri

2008 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri							
	Gruplar Arası	Gruplar İçi	En Yakın Kom.	En Uzak Kom.	Centroid	Medyan	Ward
Gruplar Arası	1	0,504	0,722	0,438	0,423	0,722	0,553
Gruplar İçi	0,504	1	0,39	0,441	0,425	0,39	0,865
En Yakın Kom.	0,722	0,39	1	0,568	0,548	1	0,326
En Uzak Kom.	0,438	0,441	0,568	1	0,748	0,568	0,102**
Centroid	0,423	0,425	0,548	0,748	1	0,548	0,098**
Medyan	0,722	0,39	1	0,568	0,548	1	0,326
Ward	0,553	0,865	0,326	0,102**	0,098**	0,326	1

Ek 24. 2013 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri

2013 Yılına Ait Kümeleme Yöntemleri Sonuçları Arasındaki Korelasyon Değerleri							
	Gruplar Arası	Gruplar İçi	En Yakın Kom.	En Uzak Kom.	Centroid	Medyan	Ward
Gruplar Arası	1	,371	,764	,648	,682	1,000	,418
Gruplar İçi	,371	1	,180	,631	,204	,371	,640
En Yakın Kom.	,764	,180	1	,544	,880	,764	,364
En Uzak Kom.	,648	,631	,544	1	,723	,648	,814
Centroid	,682	,204	,880	,723	1	,682	,476
Medyan	1,000	,371	,764	,648	,682	1	,418
Ward	,418	,640	,364	,814	,476	,418	1

ÖZ GEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Yasemin KARAOĞLAN
Uyruğu : T.C
Doğum Tarihi Ve Yeri : 08.02.1987/Sivas
E-Posta : ysmn.ozdemir@windowslive.com

EĞİTİM

Derece	Kurum	Mezuniyet Yılı
Lisans	Cumhuriyet Üniversitesi	2007-2011
Yüksek Lisans	Cumhuriyet Üniversitesi	2014-2018

İŞ TECRÜBESİ

Tarih	Kurum	Görev
-------	-------	-------

YABANCI DİL BİLGİSİ

KPDS (.....) ÜDS (.....) TOEFL (.....) EILTS (.....)