

T.C.

İSTANBUL TİCARET ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK ANABİLİM DALI

İSTATİSTİK YÜKSEK LİSANS PROGRAMI

KÜMELEME ANALİZİ VE UYGULAMASI

Yüksek Lisans Tezi

Faruk ÖZTÜRK

0960Y11202

İstanbul, Ocak 2012

T.C.
İSTANBUL TİCARET ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI
İSTATİSTİK YÜKSEK LİSANS PROGRAMI

KÜMELEME ANALİZİ VE UYGULAMASI

Yüksek Lisans Tezi

Faruk ÖZTÜRK

0960Y11202

Danışman: Doç. Dr. Dicle TAŞPINAR CENGİZ

İstanbul, Ocak 2012

T.C.
İSTANBUL TİCARET ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ONAY SAYFASI

Yüksek lisans öğrencisi Faruk ÖZTÜRK'ün “Kümeleme Analizi ve Uygulaması” konulu tez çalışması jürimiz tarafından İstatistik Yüksek Lisans tezi olarak (oybirliği / oyçokluğu) ile başarılı bulunmuştur.

Adı- Soyadı	İmza
Tez Danışmanı :
Jüri Üyesi :
Jüri Üyesi :

Hazırlamış olduğum tez özgün bir çalışma olup YÖK ve İTİCÜ Lisansüstü Yönetmeliklerine uygun olarak hazırlanmıştır. Ayrıca, bu çalışmayı yaparken bilimsel etik kurallarına tamamıyla uyduğumu; yararlandığım tüm kaynakları gösterdiğimi ve hiçbir kaynaktan yaptığım ayrıntılı alıntı olmadığını beyan ederim. Bu tezin ihtiva ettiği tüm hususlar şahsi görüşüm olup İstanbul Ticaret Üniversitesinin resmi görüşünü yansıtmamaktadır.

ÖZET

Ülkelerin gelişmişlik düzeylerini belirleyen önemli göstergelerden biri de eğitimidir. Türkiye de bu alanda politikalar üretmekte ve gelişmiş ülkelerin standartlarına ulaşmak istemektedir. Bu bağlamda öncelikle okullaşma oranının artırılmasına yönelik politikalar ile her ilin eğitilmiş nüfusunun artırılması ve gelişme düzeyinin yükseltilmesi amaçlanmaktadır. Türkiye'deki illerin eğitim düzeylerini belirleyen birden çok parametre vardır. Bu çalışmada illerin eğitim düzeyleri yönüyle benzerlikleri tespit edilmiştir. Bu amaçla, birinci bölümde Türkiye'deki eğitim sistemi ve tarihi gelişimi incelenmiştir. İkinci bölümde kümeleme analizi anlatılmış, üçüncü bölümde ise bu anlizin uygulanmasına yer verilmiştir. Uygulama aşamasında, ilk önce illerin eğitim düzeylerini gösteren faktörler belirlenmiştir. TÜİK'den elde edilen verilere göre, analize dahil edilecek değişkenler okuma-yazma bilmeyenler, okuma-yazma bilen fakat herhangi bir okula gitmeyen, ilköğretim mezunu, lise veya dengi okul mezunu, yüksek okul mezunu, yüksek lisans mezunu, doktora mezunu ve eğitim durumu bilinmeyenler şeklinde belirlenmiştir. İller, hiyerarşik olmayan kümeleme analizi yöntemlerinden k-ortalama yöntemi ile gruplanmıştır. İki kümeden, sekiz kümeye kadar farklı sayıda küme yapıları oluşturulmuştur. Wilk's Lamda ve Silhouette indeks değerleri yardımıyla uygun küme sayısının altı olduğuna karar verilmiştir. Oluşan bu altı kümeden birincisi 51 ilden oluşmaktadır. Bu küme İstanbul, Adana, Antalya, Balıkesir, Bursa, Edirne, Erzurum, Kayseri, Konya, Samsun, Trabzon gibi büyük illeri içermektedir. Ankara'nın ise diğer büyük illerden farklı bir kümede olduğu görülmüştür. İkinci ve üçüncü kümenin farklı bölgelerdeki illerden, dördüncü ve beşinci kümenin ise aynı bölgedeki illerden oluştuğu gözlemlenmiştir.

ABSTRACT

Education is one of the important indicators determining the development levels of countries. Turkey produces politics and wills to catch the developed countries' standards in that area. Accordingly, with the politics for increasing the number of schools, increasing the educated population and the development level are intended. There are numbers of parameters to determine the education levels of the cities in Turkey. In that study, 81 cities in Turkey are clustered according to their educational level using the clustering analysis method. Firstly the factors that show the population's educational condition are determined. For this purpose the education system in Turkey and its historical process is examined. According to the data from TÜİK the variables in the analysis are defined as; illiterate, literate but not attended any school, graduated from a primary school, a secondary school, a college or a grad school and the ones having a doctorate. Secondly, the clustering analysis is examined in terms of methodology. Finally, with the defined variables, 81 cities are classified using the clustering analysis. Through the application period, among the non-hierarchical clustering analysis methods k-means is used. Clusters are composed ranging from two cluster structures to eight cluster structures and MANOVA analysis is practised on the clusters to determine their number. With the Willk's Lamda and Silhouette index value, it is found out that 6 is the most acceptable cluster number. The first of those clusters is composed of 51 cities. It's observed that, this cluster includes the big cities as; İstanbul, Adana, Antalya, Balıkesir, Bursa, Edirne, Erzurum, Kayseri, Konya, Samsun, Trabzon. It's viewed that Ankara is in a different cluster. Also It 's observed that while second and third clusters are composed of cities in different regions, fourth and fifth clusters are composed of cities in the same region.

TEŐEKKÜR

Çalıřmalarımı yönlendiren, yoğun çalıřma temposunda bana vakit ayıran danıřman hocam sayın Doç. Dr. Dicle TAŐPINAR CENGİZ'e, tezimin tashihinde yardımını esirgemeyen ve maddi-manevi destek olan eřim Gülseren ÖZTÜRK'e, yakın dostum Fatih İHTİYAROĞLU'na teőekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa No.
Özet (Abstract).....	ii
Teşekkür.....	iv
İçindekiler.....	v
Şekil Listesi.....	vii
Tablo Listesi.....	viii
Kısaltmalar.....	v
GİRİŞ.....	1
1. TÜRKİYE’DE UYGULANAN EĞİTİM SİSTEMİ	3
1.1.Örgün Eğitim.....	3
1.1.2.Okul Öncesi Eğitim.....	3
1.1.3.İlköğretim.....	4
1.1.4.Ortaöğretim.....	4
1.1.5.Meslekî ve Teknik Ortaöğretim.....	5
1.1.6.Özel Eğitim.....	6
1.1.7.Özel Öğretim.....	6
1.1.8.Yükseköğretim.....	7
1.2.Yaygın Eğitim.....	7
2. KÜMELEME ANALİZİ.....	9
2.1 Kümeleme Analizinin Tanımı.....	9
2.2 Değişken Setinin Belirlenmesi ve Veri Matrisi.....	10
2.3 Benzerlik Ölçütleri.....	11
2.3.1 Nicel Veriler İçin Benzerlik Ölçütleri.....	11
2.3.1.1 Öklid uzaklığı.....	11
2.3.1.2 Ölçekli Öklit Uzaklığı.....	11
2.3.1.3 City-Block(Manhattan) uzaklığı.....	12
2.3.1.4 Minkowski Uzaklığı.....	12
2.3.1.5 Mahalanobis uzaklığı.....	12
2.3.1.6.Hotelling T ² Uzaklığı.....	13

2.3.2 Nitel Veriler için Kullanılan Uzaklık Ölçütü.....	13
2.4 Kümeleme Analizi Yöntemi.....	13
2.4.1 Hiyerarşik Kümeleme Yöntemleri.....	14
2.4.1.1 Tek Bağlantı Yöntemi.....	14
2.4.1.2 Tam Bağlantı Yöntemi.....	15
2.4.1.3 Ortalama Bağlantı Yöntemi.....	15
2.4.1.4 Ward Yöntemi.....	16
2.4.2 Hiyerarşik Olmayan Kümeleme Yöntemleri.....	17
2.4.2.1 k-Ortalama Yöntemi.....	18
2.4.2.2 En Çok Olabilirlik Yöntemi.....	19
2.5. Küme Sayısı Belirleme Yöntemleri.....	19
2.5.1 Silhouette İndeksi.....	20
2.5.2 Mariott Yöntemi.....	20
2.5.3 Calinski ve Harabasz İndeksi.....	21
2.5.4. Lewis ve Thomas Kriteri.....	22
2.5.5 Wilk's Lambda İstatistiği.....	22
3.TÜRKİYE'DE İLLERİN EĞİTİM DÜZEYLERİ YÖNÜYLE KÜMELEME	
ANALİZİYLE KÜMELENMESİ.....	24
3.1.Kümeleme Analizi ile Yapılmış Bazı Çalışmalar.....	24
3.2.Analizde Kullanılan Veriler ve Değişkenler.....	26
3.3.Analizin Bulguları.....	27
SONUÇ.....	49
Ek 1.....	52
KAYNAKLAR.....	55

ŞEKİL LİSTESİ

	Sayfa No.
Şekil 1.1 Türkiye’de uygulanan tüm eğitim sistemine dair sistematikşema.....	12
Şekil 3.1 Okuma-yazma bilmeyen.....	27
Şekil 3.2 Okuma yazma bilen fakat bir okul bitirmeyen.....	28
Şekil 3.3 Sekizyıllık eğitim mezunu.....	28
Şekil 3.4 Lise veya dengi okul mezunu.....	29
Şekil 3.5 Yüksekokul veya fakülte mezunu.....	30
Şekil 3.6 Yüksek lisans mezunu.....	30
Şekil 3.7 Doktora mezunu.....	31
Şekil 3.8 Bilinmeyen.....	31
Şekil 3.9 Kümeleme yöntemlerinin wilk’s lamda değerine göre gösterimi.....	42
Şekil 3.10 Kümeleme yöntemlerinin Silhouette indeks değerine göre gösterimi.....	43
Şekil 3.11 Birinci küme.....	44
Şekil 3.12 İkinci küme.....	45
Şekil 3.13 Üçüncü küme.....	46
Şekil 3.14 Dördüncü küme.....	46
Şekil 3.15 Beşinci küme.....	47
Şekil 3.16 Altıncı küme.....	48

TABLO LİSTESİ

Tablo 3.1 Değişkenlerin karakteristik değerleri.....	32
Tablo 3.2 Değişkenlerin Normal Dağılım Uygunluk testleri.....	33
Tablo 3.3 İki küme için değişkenlerin anlamlılık testi.....	34
Tablo 3.4 İki küme için kümeler.....	35
Tablo 3.5 Üç küme için değişkenlerin anlamlılık testi.....	36
Tablo 3.6 K=3 için kümeler.....	36
Tablo 3.7 K=4 için değişkenlerin anlamlılık testi.....	37
Tablo 3.8 K=4 için kümeler.....	37
Tablo 3.9 K=5 için değişkenlerin anlamlılık testi.....	38
Tablo 3.10 K=5 için kümeler.....	38
Tablo 3.11 K=6 için değişkenlerin anlamlılık testi.....	39
Tablo 3.12 K=6 için kümeler.....	39
Tablo 3.13 K=7 için değişkenlerin anlamlılık testi.....	40
Tablo 3.14 K=7 için kümeler.....	40
Tablo 3.15 K=8 için değişkenlerin anlamlılık testi.....	41
Tablo 3.16 K=8 için kümeler.....	41
Tablo 3.17 Wilk's lamda değerleri.....	42
Tablo 3.18 Silhouette indeks değerleri.....	43

KISALTMALAR

CH(k): K ıncı küme için Calinski ve Harabasz indeks değeri

ÇPL: Çok Programlı Lise

K: Küme sayısı

LYS : Lisans Yerleştirme Sınavı

M: Mariott ölçütü

MEB: Milli Eğitim Bakanlığı

N: Gözlem sayısı

ÖSYM: Öğrenci Seçme ve Yerleştirme Merkezi

TÜİK: Türkiye İstatistik Kurumu

W : Grup içi kareler toplamı matrisi

YGS: Yüksek geçiş sınavı

$a(i)$: i . birimin kendi kümesindeki tüm noktalara olan ortalama uzaklığı

$b(i)$: i . birimin diğer kümelerdeki tüm noktalara olan ortalama uzaklıkların minimumu

$D_{k(i,j)}$; k.kümenin , i . ve j . kümelerle olan uzaklığı

d_{ij} : i . birim ile j . birim arasındaki uzaklık

x_{ik} : i . birimin k . değişken değeri

x_i : i . küme vektörü

\bar{x}_i : i . değişken vektörü

\bar{X} : küme merkezi

S: Kovaryans matris

SS: Hata kareler toplamı

GİRİŞ

Yaşama dair değerler ortaya koyabilmenin ve üretkenliğin vazgeçilmez unsuru eğitimidir. Toplumlar ancak eğitim sayesinde varlıklarını muhafaza edip, kalkınabilmişlerdir. Ülkelerin gelişmişlik düzeylerini göstermesi açısından eğitim büyük bir öneme sahiptir. Günümüzde gelişen ülkelerde eğitim ön saftadır.

Son yıllarda her alanda gelişme kaydeden Türkiye, eğitim alanında da Avrupa Birliği standartlarını yakalamayı amaçlamaktadır. Bu sebeple eğitim alanında politikalar üretmekte ve yatırımlar yapmaktadır. Özel okulların da yaptığı katkıyla eğitimin her kademesine yönelik derslikler açılmakta, öğretmen kadrosu artırılmaktadır. Bu bağlamda her ilin eğitilmiş nüfusunun artırılması ve gelişme düzeyinin yükseltilmesi amaçlanmaktadır.

Eğitim alanında yatırımlar yapılmasına karşın, Türkiye'nin her yerinde eğitim düzeyi aynı olmayıp, farklılık göstermektedir. Bu nedenle, bu çalışmada Türkiye'de, eğitim durumları yönüyle birbirine benzerlik veya farklılık gösteren illerin hangileri olduğu tespit edilmeye çalışılacaktır. Bu çalışmanın diğer bölümlerinde aşağıdaki sıralama takip edilecektir.

Birinci bölümde; Türkiye'de uygulanan eğitim sistemi incelenecektir. Eğitim sistemi "Örgün eğitim" ve "Yaygın eğitim" olmak üzere iki ana başlık altında verilir ve bu başlıklar altında eğitim kademeleri açıklanacaktır.

İkinci bölümde kümeleme analizi incelenecektir. Bu bölümde ilk olarak benzerlik ölçütlerinden bahsedilecektir. İkinci olarak hiyerarşik ve hiyerarşik olmayan kümeleme analizi yöntemleri açıklanacaktır. Son olarak da küme sayısı belirleme yöntemleri anlatılacaktır.

Üçüncü bölümde ise Türkiye'deki 81 il, illerin eğitim düzeyleri yönüyle, kümeleme analizi yöntemiyle kümelenecektir. Uygulama safahsında ilk olarak illere ait nüfusun eğitim durumlarını açıklayan değişkenler, Türkiye İstatistik Kurumun (TÜİK)'dan alınan veriler yardımıyla belirlenecektir. Bu değişkenlere göre, iller hiyerarşik olmayan kümeleme analizi yöntemlerinden olan k-ortalama yöntemiyle kümelenecektir. Küme sayısı belirleme yöntemleri ile de uygun küme sayısına karar verilecektir.

1. TÜRKİYE’DE UYGULANAN EĞİTİM SİSTEMİ

Türkiye, köklü bir eğitim geleneğine sahip olmasıyla birlikte Türk eğitim sistemi şartların ve dönemin gereksinimleri ölçüsünde farklı evrelerden geçmiştir. Günümüzde ise Türkiye’de uygulanan eğitim sistemi “Örgün Eğitim” ve “Yaygın Eğitim” olmak üzere iki ana bölümden oluşmaktadır.

1.1. Örgün Eğitim

Örgün eğitim, belirli yaş grubundaki ve aynı seviyedeki bireyler için, belli bir amaç doğrultusunda ve program dahilinde, bir okul çatısı altında yapılan düzenli eğitimidir. Örgün eğitim şu kademelerden oluşmaktadır;

- Okul öncesi
- İlköğretim
- Ortaöğretim
- Yükseköğretim

1.1.2. Okul Öncesi Eğitim

Okul öncesi eğitim; 36-72 ay arası yaşlardaki çocukların isteğe bağlı olarak eğitimini kapsamaktadır. Okul öncesi eğitim kurumları, bağımsız anaokulları, kız meslek liseleri bünyesinde eğitim veren anaokulları ve diğer öğretim kurumlarına bağlı ana sınıfları olarak eğitim faaliyetini sürdürmektedir. Okul öncesi eğitim isteğe bağlı olduğundan katılım oranı oldukça düşüktür. Ancak bu dönemdeki eğitimin, daha sonraki eğitim düzeylerindeki öğrenci başarısına olumlu katkıları bulunduğu için okul öncesi eğitimi yaygınlaştırma çalışmaları sürdürülmektedir (<http://www.meb.gov.tr>).

1.1.3. İlköğretim

İlköğretim, 6–14 yaş grubundaki çocukların eğitim-öğretimini kapsayan, kız ve erkek bütün vatandaşlar için devlet tarafından zorunlu olan bir eğitimidir. İlköğretim kurumları sekiz yıllık eğitim vermektedir. İlk beş yılı ilkokul, son üç yılı ortaokul düzeyinde eğitim vermektedir. Devlet okullarında eğitim ücretsiz olarak verilmektedir. Bu okulları bitirenler ilköğretim diplomasını almaya hak kazanırlar.

1.1.4. Ortaöğretim

Ortaöğretim; 14-17 yaş grubu vatandaşların, dört yıl süreli (yabancı dil hazırlığı olan okullarda beş yıl) eğitimini kapsayan genel liseler ile meslekî ve teknik liselerden oluşur. İlköğretimi tamamlayan her öğrenci, ortaöğretime devam hakkına sahiptir. Genel ortaöğretim türleri şunlardır;

- Genel Liseler
- Anadolu Liseleri
- Fen Liseleri
- Anadolu Öğretmen Liseleri
- Anadolu Güzel Sanatlar Liseleri
- Sosyal Bilimler Liseleri
- Spor Liseleri

Genel liseler dışındaki ortaöğretim kurumlarına giriş merkezi bir sınavla olmaktadır. Anadolu Güzel Sanatlar liseleri ile Spor liselerine ise yetenek sınavıyla öğrenci alınmaktadır. (<http://www.meb.gov.tr>)

1.1.5. Meslekî ve Teknik Ortaöğretim

Meslekî ve teknik ortaöğretim okulları; ülke endüstrisinin ve hizmet sektörünün gereksinim duyduğu iş alanlarına orta kademe teknik insan gücü yetiştiren ve öğrencileri yükseköğretime hazırlayan öğretim kurumlarıdır (<http://www.meb.gov.tr>). Meslekî ve teknik ortaöğretim türleri şunlardır;

- Erkek Teknik Öğretim Okulları
- Kız Teknik Öğretim Okulları
- Ticaret ve Turizm Öğretimi Okulları
- Din Öğretimi Okulları
- Özel Eğitim Okulları
- Sağlık Meslek Liseleri
- Tarım Meslek Liseleri
- Adalet Meslek Liseleri
- Tapu ve Kadastro Meslek Liseleri
- Anadolu Meteoroloji Meslek Liseleri

Erkek ve kız teknik öğretim okulları, başlangıçta geleneksel olarak sadece kız veya erkek öğrenciler için uygun olduğu düşünülen programların izlendiği okullar olarak eğitim vermesi planlanmıştır. Ancak bu okulların bir kısmı, meslek alanlarındaki çeşitli gelişmeler doğrultusunda benzer programların izlendiği karma eğitim kurumları hâline gelmiştir(<http://www.meb.gov.tr>).

Nüfusu az ve dağınık olduğu yerleşim yerlerinde eğitim binaları, öğretmen, yönetici ve diğer personelden azamî derecede yararlanarak kaynak israfını önlemek, ilköğretimi tamamlayan öğrencilere ilgi, istek ve yetenekleri doğrultusunda ortaöğretimden yararlanma olanağını sağlamak amacıyla çok programlı liseler (ÇPL) faaliyete geçirilmiştir(<http://www.meb.gov.tr>).

1.1.6. Özel Eğitim

Özel eğitim toplumdaki engelli vatandaşlar için, ihtiyaçlarını en iyi şekilde karşılayarak onları toplumla bütünleştirmek ve meslek sahibi yapmak amacıyla yapılan eğitimidir. Ülkemizde sekiz engel grubundaki çocuklara ve gençlere eğitim hizmetleri sunulmaktadır:

- Görme engelliler
- İşitme engelliler
- Ortopedik engelliler
- Zihinsel engelliler
- Uyum güçlüğü olan çocuklar
- Dil ve konuşma güçlüğü olanlar
- Üstün yetenekliler
- Uzun süreli hasta ve hastanede yatan çocuklar

Özel eğitim okulları eğitim sistemindeki kademelendirmeye göre yapılandırılmıştır. Diğer okullardan farklı olarak ilköğretim öncesinde hazırlık sınıfı bulunmaktadır. Ancak ilköğretime devam edebilecek durumda olan engelli öğrenciler hazırlık sınıfına alınmadan ilköğretime başlamaktadır(<http://www.meb.gov.tr>).

1.1.7. Özel Öğretim

Özel öğretim kurumları maliyeti, gelir ve gideri gerçek veya tüzel kişilerce karşılanan, denetim ve gözetimi ise Millî Eğitim Bakanlığınca yapılan kurumlardır. Bu kurumlar 625 Sayılı Özel Öğretim Kurumları Kanunu'na göre açılmakta ve faaliyet göstermektedir. Bu statüde eğitim-öğretim faaliyeti veren okullar;

- Her kademe ve türdeki özel okullar
 - Özel dershaneler
 - Özel meslekî ve teknik kurslar
 - Özel motorlu taşıt sürücü kursları
 - Özel öğrenci etüt eğitim merkezlerini
- şeklinde sınıflandırılabilir.

Örgün özel öğretim kurumları, özel Türk okulları, özel yabancı okullar, özel azınlık okulları, özel uluslararası okullar ve özel eğitim okulları olarak gruplandırılmaktadır(<http://www.meb.gov.tr>).

1.1.8.Yükseköğretim

Yükseköğretim kurumları sınavla öğrenci kabul eden, en az iki yıllık yükseköğrenim veren eğitim kurumlarını kapsar. Yükseköğretim kurumları: meslek yüksek okulları ve fakülteler olarak iki gruba ayrılırlar. Genel olarak ara eleman yetiştirmeyi amaçlayan meslek yüksek okullarının eğitim süresi iki yıldır. Fakültelerin eğitim süresi ise, eğitim verilen alana göre dört ile altı yıl arasında değişmektedir.

Yükseköğretime giriş merkezi bir sistemle ve ülke genelinde “Öğrenci Seçme ve Yerleştirme Merkezi” (ÖSYM) tarafından düzenlenen iki aşmalı (YGS ve LYS) sınavla yapılmaktadır. Bu sınav, lise mezunu veya mezun olabilecek durumdaki öğrencilerin kabul edildiği bir sınavdır.

1.2. Yaygın Eğitim

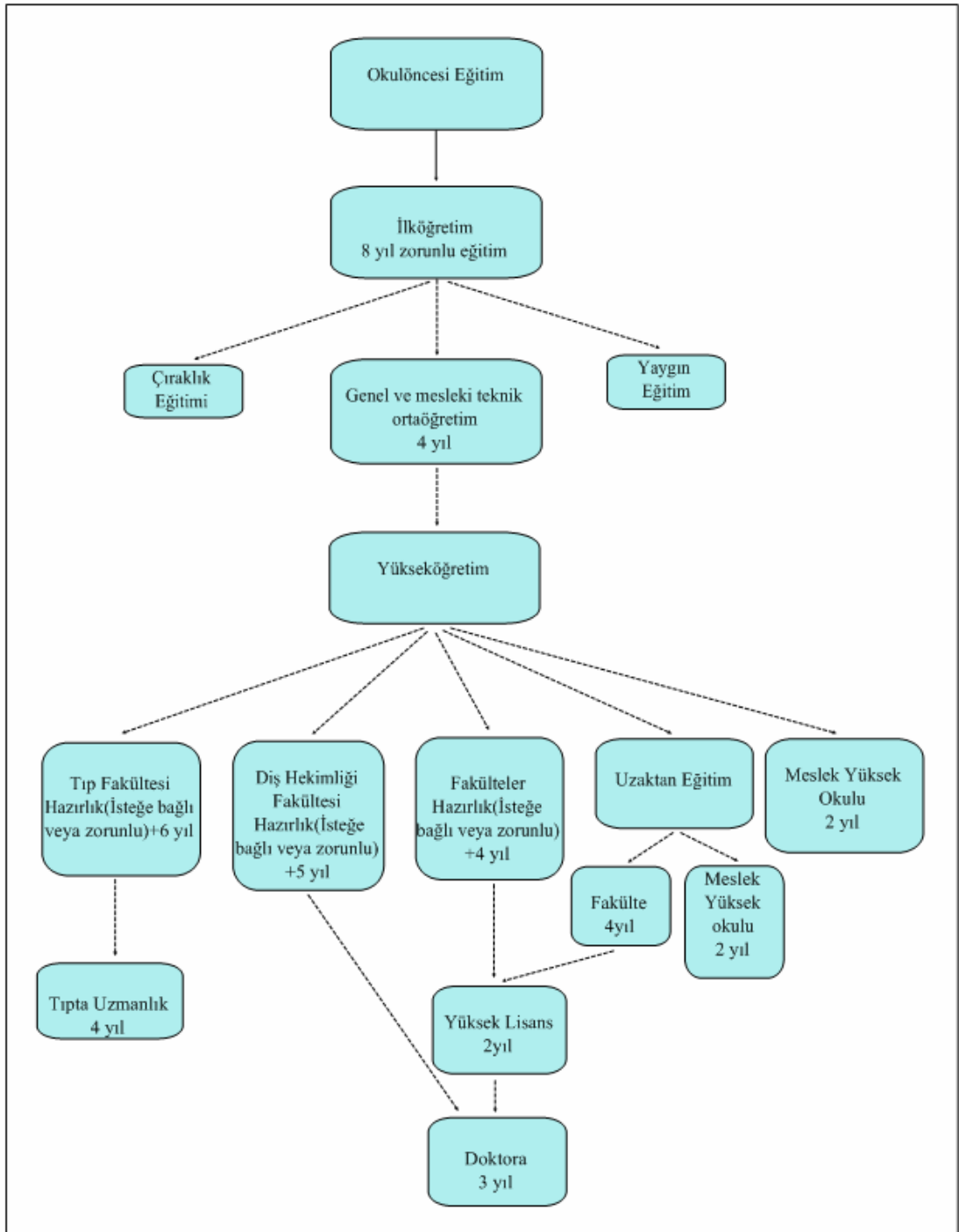
Yaygın eğitim, örgün eğitim sistemine hiç girmemiş, herhangi bir eğitim kademesinde bulunan veya bu kademelerden birinden ayrılmış olan bireylere eğitim verilen bir sistemdir. Yaygın eğitimin amaçları şöyle sıralanabilir:

- Yetişkinlere okuma-yazma öğretmek
- Temel bilgiler vermek
- Öğrencilerin en son devam ettikleri öğrenim kademesinde edindikleri bilgi ve kabiliyetlerini geliştirmek
- Hayatlarını kazanmalarını sağlayacak yeni imkânlar sunmak

Yaygın eğitim; halk eğitimi, çıraklık eğitimi ve uzaktan eğitim yoluyla uygulanmaktadır. Ayrıca, ortaöğretim seviyesinde örgün eğitim veren meslekî ve teknik öğretim kurumlarında açılan meslek kursları ile de yaygın eğitim verilmektedir(<http://www.meb.gov.tr>).

Aşağıda, Türkiye’de uygulanan tüm eğitim sistemine dair sistematik şema verilmiştir.

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



2. KÜMELEME ANALİZİ

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

Birimlerin sayısı arttıkça birimleri sınıflandırmak daha da zorlaşmış ve yeni teknikler bulmayı gerektirmiştir. Bu gereksinim sonucu kümeleme analizi kavramı ortaya çıkmıştır. Günümüzde gruplandırmada en sık kullanılan yöntemlerden olan kümeleme analizi 1939 yılında ilk kez Tryon tarafından kullanılmıştır. Bu kavram 1960 yıllarından sonra gelişmiş ve geniş uygulama alanlarına sahip olmuştur. Bu alanda önemli bir katkısı, 1963 yılında yazdığı “Sayısal Sınıflandırma İlminin Temelleri” adlı kitapla Robert Sokal ve Peter sağlamıştır (Anderberg, 1973) .

Günümüzde bu analiz tıptan ekonomiye varan bir yelpazede uygulanmaktadır (Hofman and Jarvis, 1998). Kümeleme analizinin tıp, biyoloji, kimya, ekonomi, veri madenciliği, sosyoloji, eğitim v.b. gibi çok uygulama alanları mevcuttur.

Kinnear ve Taylor (1979), tesadüfi olarak seçilen kişilere anket düzenlemiş ve katılımcıların eğitim, meslek, dini inanç, kültür, siyaset v.b. gibi sosyal ve kültürel eğilimleri tesbit edilmiştir. Elde edilen verilere Öklid uzaklık fonksiyonu kullanılarak benzerlik ölçümleri yapılmış ve kümeleme analizi ile bireyler kümelenmiştir. Yapılan analiz neticesinde bireylerin %47’ nin yeniliklere açık, geri kalanın ise eğitim ve gelir düzeyi düşük muhafazakarlardan oluşan iki kümeli bir yapı oluşturduğu sonucuna ulaşılmıştır.

2.1. Kümeleme Analizinin Tanımı

Kümeleme analizi son yıllarda çok sık kullanılan istatistiksel analiz yöntemlerinden biridir. Çok değişkenli istatistik analiz yöntemlerinden olan kümeleme analizi için çeşitli tanımlar mevcuttur.

Kümeleme analizi, birey veya nesnelere benzerliklerine göre kümeler veya gruplara ayırmak için kullanılan çok değişkenli istatistik analiz tekniğidir (Tatlıdil, 1996)

Çok boyutlu uzayda büyük ve karmaşık verilerin özetlenmesi ve tanımlanmasında yol gösterici bir araştırma yöntemidir. Kümeleme analizinde öncelikli amaç birey veya nesnelere temel özelliklerini dikkate alarak birbirleriyle benzerlikleri doğrultusunda onları gruplamaktır (Kalaycı, 2005).

Kümeleme analizi birbirine benzer olan bireylerin aynı gruplarda toplanmasını amaçlaması bakımından diskriminant analizi ile, birbirine benzer değişkenlerin aynı gruplarda toplanmasını amaçlaması nedeniyle de faktör analizi ile benzerlik göstermekte olup veri indirgeme özelliği vardır(Çakmak, 1999). Diskriminant analizinden farklı olarak, kümeleme analizinde gruplar önceden bilinmemekte ve grup sayısı analiz sürecinde değişebilmektedir.

Diğer çok değişkenli istatistik analizlerde önemli olan verilerin normalliği varsayımı, kümeleme analizinde çok önemli olmayıp uzaklık değerlerinin normalliği yeterli görülmektedir(Tatlıdil, 1996).

Kümeleme analizi neticesinde küme içi birimler, kümeler arasındaki birimlere göre karakteristik özellikleri bakımından daha çok benzeşirler. Bunun sonucu olarakta, küme içi homojenlik, kümeler arası heterojenlik sağlanmış olur.

2.2. Değişken Setinin Belirlenmesi ve Veri Matrisi

Analize dahil edilecek birimlere ait verilerden hangilerinin kümelerin oluşmasında etkili olup, hangilerinin etkili olmadığı belirlenir. Kümelerin oluşmasında etkin rol alan verilerin oluşturduğu matrise veri matrisi denir.

2.3. Benzerlik Ölçütleri

Kümeleme analizinde genel amaç küme içi homojenliği, kümeler arası heterojenliği sağlamaktır. Bu da, karakteristik özellikleri bakımında benzer bireylerin aynı kümede toplanmasıyla temin edilmiş olur. Bireylerin benzerlikleri uzaydaki konumları ile ilgilidir. Uzaydaki konumları itibari ile birbirine uzaklıkları daha az olan bireyler aynı kümede toplanmış olacaklardır. Bireyler gruplanırken birbirine yakınlıkları bazı uzaklık ölçütlerine göre hesaplanır. Değişkenlerin kesikli ya da sürekli olmalarına ya da değişkenlerin nominal, ordinal, aralık ya da oransal ölçekte olmalarına göre hangi hangi benzerlik ölçütünün kullanılacağına karar verilir. Benzerlik ölçütlerinden en çok kullanılanları şunlardır;

2.3.1. Nicel Veriler İçin Benzerlik Ölçütleri

2.3.1.1. Öklid uzaklığı

Öklid uzaklığı en sık kullanılan uzaklık ölçüsüdür. Geometrik olarak iki nokta arasındaki en yakın uzaklık şeklinde tanımlanabilir ve aşağıdaki formülle ifade edilir.

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

2.3.1.2. Ölçekli Öklit Uzaklığı

Değişkenlerin aynı ağırlıkta ölçeklendiği durumlarda kullanılan bir uzaklık ölçütüdür.

$$d_{ij} = \sqrt{\sum_{k=1}^p w_k^2 (x_{ik} - x_{jk})^2} \quad (2.2)$$

2.3.1.3. City-Block(Manhattan) uzaklığı

City-Block uzaklık ölçüsü, değişkenler arasında korelasyon olmadığı durumlarda kullanılan bir uzaklık ölçütüdür. Eğer değişkenler arasında korelasyon varsa City-Block uzaklık ölçüsüyle hesaplanan uzaklık ölçüleri baz alınarak yapılan kümeleme anlamlı olmayacaktır

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (2.3)$$

2.3.1.4. Minkowski uzaklığı

Minkowski uzaklık ölçüsü genel bir uzaklık ölçüsü, Öklid ve City-Block uzaklık ölçüleri ise Minkowski uzaklık ölçüsünün özel bir durumudur.

$$d_{ij} = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^q \right]^{1/q} \quad (2.4)$$

2.3.1.5. Mahalanobis uzaklığı

Değişkenler arası korelasyon olduğu durumlarda kullanılan bir uzaklık ölçütüdür.

$$Md_{ij} = (x_i - x_j)' S^{-1} (x_i - x_j) \quad (2.5)$$

Değişken sayısı p olmak üzere, S , $p \times p$ tipindeki kovaryans matrisini göstermektedir.

2.3.1.6. Hotelling T^2 Uzaklığı

İki kümenin ortalama vektörlerinin karşılaştırıldığı durumlarda Hotelling uzaklık ölçütü kullanılır.

$$T^2 = \frac{n_1 n_2}{n} (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j) \quad (2.6)$$

2.3.2. Nitel Veriler için Kullanılan Uzaklık Ölçütü

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

$$w_k = \begin{cases} 1 & ; \text{nicel veriler için} \\ \frac{1}{k'inci\ de\ g\ i\ s\ k\ e\ n\ i\ n} & ; \text{nitel veriler için} \\ \text{da\ g\ i\ l\ i\ m\ a\ r\ a\ l\ ı\ g\ ı} & \end{cases} \quad (2.7)$$

olmak üzere;

$$d_{ij} = \frac{1}{p} \sum_{k=1}^p w_k |x_{ik} - x_{jk}| \quad (2.8)$$

formülü ile hesaplanır.

2.4. Kümeleme Analizi Yöntemleri

Hangi uzaklık ölçütünün kullanılacağına karar verildikten sonra, ikinci adım kümeleme yönteminin belirlenmesidir. Birimlerin benzerliklerine göre kümelere

atanmasında kullanılabilir çeşitli yöntemler vardır. Bu yöntemlerden en çok kullanılanları hiyerarşik ve hiyerarşik olmayan kümeleme yöntemleridir.

2.4.1. Hiyerarşik Kümeleme Yöntemleri

Yöntem, aşama sıralı kümeleme yöntemi olarak da bilinir. Hiyerarşik kümelemede iki adet ana teknik vardır. Birincisi gruplayıcı (agglomerative) diğeri ise parçalayıcı (divisive) tekniktir. Her birimin başlangıçta bir küme olarak kabul edildiği, daha sonra en yakın iki kümenin yeni bir kümede toplanarak birleştirilip, küme sayısının bu şekilde her seferinde bir azaltıldığı yöntem gruplayıcı hiyerarşik yöntem denir. Bu süreç dendogram veya ağaç grafiği adı verilen şekilde gösterilebilir.

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

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

Uygulamalarda çoğunlukla gruplayıcı hiyerarşik kümeleme yöntemi kullanılmaktadır. Gruplayıcı hiyerarşik yöntemler arasında en çok tek bağlantı, tam bağlantı, grup ortalama yöntemi ve Ward yöntemi kullanılmaktadır.

2.4.1.1. Tek Bağlantı Tekniği

En yakın komşuluk olarak da bilinen tek bağlantı tekniği, uzaklıklar matrisini kullanarak birbirine en yakın (uzaklık değerleri en küçük) birey ya da nesnelere birleştirmeye dayanmaktadır (Johnson and Wichern, 1988).

Başlangıçta her bir birim bir küme olarak kabul edilir. Sonra birimler arası minimum uzaklıklar hesaplanır. Aralarındaki uzaklıkları en az olan birimleri içeren iki kümenin birleştirilmesiyle iki birimli yeni bir küme elde edilir. Bu işleme tüm birimleri kapsayan bir kümenin elde edilmesine kadar devam edilir.

Birleştirilen i ve j nci birimlerin oluşturduğu kümenin k ıncı küme ile ilişkisini gösteren uzaklık ölçütü;

$$D_{k(i,j)} = \min(d_{ki}, d_{kj}) \quad (2.9)$$

biçiminde ifade edilmektedir.

2.4.1.2. Tam Bağntı Tekniği

Tek bağlantı tekniğinden farklı olarak, birimler arasındaki maksimum uzaklıklar hesaplanır. Aralarındaki uzaklıkları en az olan birimlerin bir kümede toplanmasıyla, tek bağlantı tekniğindeki gibi kümeleme işlemi gerçekleştirilir. Tam bağlantı tekniğinde kümeler arasındaki uzaklıklar;

$$D_{k(i,j)} = \max(d_{ki}, d_{kj}) \quad (2.10)$$

şeklinde hesaplanır.

2.4.1.3.Ortalama Bağntı Tekniği

Tek bağlantı ve tam bağlantı yöntemlerinden tek farkı; uzaklıkların hesaplanmasındaki farklılıktır. Bir küme içindeki birimler ile diğer bir küme içindeki birimler arasındaki ortalama (yaygın olarak aritmetik ortalama) uzaklıklar kullanılır. Kümeler arası elemanların ortalama uzaklığı;

$$D_{k(i,j)} = \frac{\sum_{i,j} \sum_k d_{k(i,j)}}{N_{(i,j)} N_k} \quad (2.11)$$

formülüyle hesaplanır. Minimum uzaklığa sahip küme çifti gruplandırılır ve işlem tekrar edilir.

Yukarda bahsedilen üç yöntem algoritmik olarak aynıdır. Kümeler arası, birimlerin birbiriyle olan uzaklıkları hesaplanır ve benzerlikleri en fazla olan kümeler gruplandırılır. Bu yöntemleri birbirinden ayrılan tek yanları ise uzaklıkların hesaplanmasındaki farklılıktır.

2.4.1.4. Ward Metodu

Ward hiyerarşik kümeleme yöntemi; iki küme birleştirilirken “en az bilgi kaybını” sağlayacak bir yapıya oturtmayı önerir (Johnson, 2002). Bilgi kaybının en az seviyede kalmasını sağlamak için “Hata Kareler Toplamı(SS)” prensibinden faydalanılır. Bu teknikte amacın hata kareler toplamını en küçük seviyede tutmak olarak söylenebilir.

N birimli veri setinde, başlangıçta her bir birimi bir küme kabul edeceğimizden N tane kümemiz olacaktır. Bu durumda hata kareler toplamı;

$$SS = SS_1 + SS_2 + \dots + SS_N \quad (2.12)$$

şeklindedir ve başlangıçta SS=0 olmaktadır. Analizin her adımında mümkün olan tüm küme çiftleri dikkate alınır ve SS’de minimum artışı sağlayan küme çifti gruplandırılır (Johnson, 2002). Nesnelerin tek bir kümeye ait olduğu son adımda ise SS aşağıdaki şekilde hesaplanır;

$$SS = \sum_{i=1}^N (x_i - \bar{x})'(x_i - \bar{x}) \quad (2.13)$$

Ward Yöntemi kullanılarak yapılan analiz sonucunda, birimler değişik seviyelerde başarılı bir şekilde bir araya geldikleri "dendogram" adlı şemada gösterilir (Dibb 1998). Ward yöntemi aykırı noktalara duyarlı bir yöntemdir (Everitt *et al.*, 2001).

Hiyerarşik kümeleme analizi aykırı noktalara karşı fazla duyarlı bir yöntemdir. Bu sebeple analiz yapılırken farklı hiyerarşik teknikler uygulanıp bunlar karşılaştırılmalıdır. Uygulanan hiyerarşik kümeleme analizi sonuçları tutarlılık gösteriyorsa kümeleme işlemi gerçekleştirilir. Diğer bir yöntem ise veri setine bazı hatalı veriler ekleyerek sonucun eklenmeden önce ve sonra tutarlılık gösterip göstermediğine bakılmasıdır(Johnson, 2002). Hiyerarşik kümeleme analizinde karşılaşılabilecek diğer bir problem ise bir gruba atanan bir elemanın geri alınma ihtimalinin olmayışıdır.

2.4.2. Hiyerarşik olmayan kümeleme yöntemleri

Küme sayısına karar verilmiş ise bu durumda, uzun zaman alan hiyerarşik yöntemler yerine hiyerarşik olmayan yöntemler kullanılmaktadır.

Hiyerarşik olmayan kümeleme yöntemleri değişkenlerden ziyade nesnelere gruplandırmaya yarar ve bu nesnelere K adet kümeye ayırır. Küme sayısı K başlangıçta belirlenebileceği gibi kümeleme sürecine bağlı olarak ortaya çıkabilir(Johnson, 2002).

Hesaplanacak bir benzerlik matrisi olmadığından, ve işlem sırasında başlangıç verisinin hafızada tutulması gerekmediğinden hiyerarşik olmayan yöntemler çok daha

büyük ve kompleks veri setlerine hiyerarşik olanlara göre daha rahat uygulanır(Johnson, 2002).

Hiyerarşik olmayan kümeleme yöntemlerinin temel dezavantajı küme sayılarının daha önceden belirlenmesi ve küme seçimlerinin keyfi olmasıdır (Blashfield and Aldenferder, 1978).

Hiyerarşik olmayan kümeleme yöntemleri; k-ortalama yöntemi ve En çok olabilirlik yöntemidir. Hiyerarşik olmayan kümeleme yöntemlerinden en sık kullanılanı k-ortalama yöntemidir.

2.4.2.1. k – Ortalama yöntemi

Birimlerin, sayısı önceden belirlenen k tane kümeye rastgele atanmasıyla yöntem uygulanmaya başlanır. K tane kümenin ortalamaları hesaplanır ve bu ortalamalar kümenin merkezini oluşturur. Tüm birimlerin küme merkezlerine olan uzaklıkları hesaplanır. Hesaplanan bu uzaklık değerlerine göre, herhangi bir küme merkezine uzaklığı en yakın olan bir birimin bu merkezin ait olduğu kümeye atanması ile işlem devam eder. Oluşan yeni kümelerin merkezleri tekrar hesaplanır ve birimlerin kümelere atanma işlemi, herbir birimin küme merkezlerine olan uzaklık değerlerine göre, ait olduğu kümeye olan uzaklığının en az olması sağlanıncaya kadar devam eder.

Amaç, diğer kümeleme yöntemlerinde olduğu gibi küme içi benzerliklerinin maksimum, kümeler arası benzerliklerinin ise minimum olmasını sağlamaktır. Küme benzerliği, kümenin ağırlık merkezi kabul edilen bir birim ile kümedeki diğer birimler arasındaki uzaklıkların ortalama değeri ile ölçülmektedir (Han and Kamber 2001).

Yöntemin işleyiş aşamaları şu şekildedir;

— İlk k birim, k tane kümeye atanır. K tane kümenin herbiri bir elemanlıdır ve bu elemanlar küme merkezini oluşturmaktadır.

- Geriye kalan birimler yakın oldukları kümelere atanırlar. Bütün birimler k kümeye atanmış olurlar. Bu atama işlemi gerçekleşikten sonra yeni oluşan elemanlarla küme merkezleri tekrar hesaplanır.
- Hesaplanan küme merkezlerine göre, bir birim en yakın olduğu kümeye tekrar atanır ve küme merkezleri yeniden hesaplanır. Kümeler arası atama işlemi sonlanıncaya kadar bu işlem devam eder ve kümeler içi homojenlik, kümeler arası heterojenlik sağlanmış olur.

2.4.2.2. En çok olabilirlik yöntemi

Diskriminant analizinde de kullanılan en çok olabilirlik yönteminde her bir gözlem en büyük olabilirlik değerini verecek biçimde daha önceden belirlenen kümelere atanır. Kuramsal dayanağı güçlü olmakla birlikte en çok olabilirlik yöntemi yaygın olarak kullanılmamaktadır (Tatlıdil 1996).

2.5. Küme Sayısının Belirlenmesi ve Küme Geçerliliği Yöntemleri

Kümeleme analizinden sağlıklı bir sonuç alabilmek için küme sayısının ne olacağı önem arz etmektedir. Küme sayısının ne olacağının kesin olarak söyleyen bir yöntemden bahsetmek mümkün olmamakla beraber aşağıda bazı yöntemlerden bahsedilecektir.

Bu yöntemlerden en çok bilineni;

$$k = \left(\frac{n}{2} \right)^{\frac{1}{2}} \quad (2.14)$$

eşitliğiyle verilmektedir. Küçük örneklemlerle araştırmalarda kullanılması tavsiye edilir. Büyük örneklemlerle araştırmalarda kullanılması durumunda sağlıklı sonuçlara ulaşılması zorlaşır (Everitt 1974).

2.5.1.Silhouette İndeksi

Kümelenecek her bir birimin atandığı kümeye uygunluğunu tanımlamak amacıyla Rousseeuw (1987) Silhouette indeksi önermiştir.

$a(i)$; i . birimin kendi kümesindeki tüm noktalara olan ortalama uzaklıklarını
 $b(i)$; i . birimin diğer kümelerdeki tüm noktalara olan ortalama uzaklıkların minimumunu göstermek üzere, Silhouette indeksi;

$$sil(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.15)$$

şeklinde tanımlanır. Hesaplanan i . birime ait bu değer 1'e yaklaşırsa i . birimin atandığı kümeye daha uyduğu, 0'e yaklaşırsa veya negatif olursa i . birimin atandığı kümeye uygun olmadığı sonucuna varılır.

Yapılan tüm kümelemenin kalitesi için ise Ortalama Silhouette değeri;

$$sil(C) = \frac{1}{n} \sum_{s_i \in S} sil(i) \quad (2.16)$$

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

2.5.2 Mariott Yöntemi

Küme sayısına belirlemede kullanılan diğer bir ölçüt Mariott tarafından önerilmiştir ve M harfiyle gösterilmektedir. Bu ölçüt;

$$M = k^2 |W| \quad (2.17)$$

eşitliğiyle tanımlanmaktadır(Marriot 1971). Burada W , grup içi kareler toplamı matrisidir ve

$$W = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)(X_{ij} - \bar{X}_j) \quad (2.18)$$

şeklinde hesaplanmaktadır.

Mariott tarafından önerilen bu ölçüte göre M değerini minimum yapan k değeri uygun küme sayısı olarak kabul edilmektedir.

2.5.3. Calinski ve Harabasz İndeksi

Kümeleme analizinde, küme sayısının belirlenmesi ve oluşturulan kümelerin geçerliliğinin test edilmesinde kullanılan diğer bir yöntemde Calinski ve Harabasz indeks değeridir. Bu indeks değeri;

$$CH(k) = \frac{B(k)}{k-1} / \frac{W(k)}{n-k} \quad (2.19)$$

şeklinde hesaplanır (Calinski ve Harabasz 1974).

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

$$W(k) = \frac{1}{2} \sum_{j=1}^k \sum_{i, j \in C_i} d(i, j) \quad (2.20)$$

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

$$B(k) = \frac{1}{2} \sum_{j=1}^k \sum_{\substack{i, j \in C_i \\ j \notin C_i}} d(i, j) \quad (2.21)$$

şeklinde hesaplanırlar. Bu kritere göre, CH indeks değerini maksimum yapan küme sayısı, uygun küme sayısı olarak kabul edilir (Calinski Harabasz 1974).

2.5.4. Lewis ve Thomas Kriteri

Küme sayısının belirlenmesine kullanılacak bir diğer yaklaşım ise Lewis ve Thomas kriteridir. Lewis ve Thomas yaklaşımına göre küme sayısının belirlenmesinde uygulanması gereken iki kriter vardır. Kümeler, toplam varyansın %80'ini açıklamalı ve varyansta oluşacak %5'e kadarlık bir artış durumunda ise küme sayısı bir artırılmalıdır (Ruiz, 1998).

2.5.5. Wilk's Lambda İstatistiği

Kümelenecek birim sayısı 30'un üzerinde olduğunda, küme sayısının belirlenmesinde kullanılan ölçütlerden Wilk's Lamda ölçütü duyarlılığı diğer ölçütlere göre en yüksek düzeye ulaşmaktadır (Dinçer ve Özdamar, 1992). Gruplar arası varyansın grup içine göre maksimum olduğunun analiz edilmesinde Wilk's Lamda ölçütü çok duyarlı olmaktadır (Everitt 1979).

İki veya daha çok değişkenli verilerin aynı ortalama vektörlü çok değişkenli normal dağılımdan gelip gelmediği, Çok Değişkenli Varyans Analizi (MANOVA) ile test edilebilmektedir (Johnson and Wichern, 1988).

Kümeleme analizi neticesinde oluşturulan kümelerin ortalamalarının birbirinden istatistiksel olarak farklılığı MANOVA analiziyle test edilebilir ve

hesaplanan Wilk's Lamda ölçütü küme sayısının belirlenmesinde kullanılabilir(Tatlıdil, 1996).

Wilks olabilirlik oran istatistiği;

$$\Delta = \frac{|W|}{|W + B|} = \frac{|W|}{|T|} \quad (2.22)$$

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

3. TÜRKİYE’DE İLLERİN EĞİTİM DÜZEYLERİ YÖNÜYLE KÜMELEME ANALİZİYLE KÜMELENMESİ

Bu bölüm’de kümeleme analizinin uygulamalarına yer verilecektir. Türkiye’de illerin eğitim düzeyleri yönüyle kümelenmesi amaçlanmaktadır. Oluşturulacak kümelerin sayısının önceden bilinmemesi ve daha özet bilgi vermesi nedeniyle illerin kümelenmesi için, yöntem olarak kümeleme analizi tercih edilmiştir. İlk önce kümeleme analizi ile yapılmış bazı çalışmalardan bahsedilecek, daha sonra Türkiye’de illerin eğitim düzeyleri yönüyle, kümeleme analiziyle benzerlikleri tespit edilecektir.

3.1. Kümeleme Analizi ile Yapılmış Bazı Çalışmalar

Karabulut, Gürbüz ve Sandal(2004), yaptıkları araştırmada Türkiye’nin 81 iline ait 54 sosyo-ekonomik değişken (DPT) yardımıyla, aynı yapıyı gösteren homojen il grupları belirlenmiştir. Bunun için hiyerarşik kümeleme analizi kullanılmıştır. Küme sayılarının belirlenmesinde 81 il, istatistiksel olarak 7, 10 ve 15 gruba ayrılması test edilmiştir. Bulunan sonuçlara göre en anlamlı kümelemenin 15’li sınıflandırma sonucunda elde edildiğine karar verilmiştir. Analizde kullanılan Kareli Öklid ve Pearson Yakınlık Matrisi ölçütlerine göre en çok benzeyen illerin Bitlis ve Mardin, en az benzeyen illerin ise İstanbul ve Kars olduğu sonucuna ulaşılmıştır. Bu yöntemle illerin ayrılımlarında ve homojen yapı göstermelerine neden olan değişkenler yardımıyla illerin oluşturduğu farklı sosyo-ekonomik bölgeler belirlenmiştir. Bu sınıflandırmaya göre sosyo-ekonomik veriler bakımından gelişmiş olan İstanbul, Ankara, Kocaeli, İzmir ve aynı kriterler bakımından geri olan Bingöl tek başına bir grup oluşturduğu görülmüştür. Analiz sonucuna göre Türkiye’nin doğu yarısındaki illerden Artvin-Tunceli ve Kars-Ardahan kendi aralarında 2’li grup oluştururken, Marmara Bölgesine ait illerden Bilecik-Kırklareli-Tekirdağ 3’lü grup, Gazi Antep-Hatay-Mersin-Kayseri-Konya 5’li grup, Adana-Antalya-Bursa-Edirne-Eskişehir-Muğla ve Yalova; Elazığ-Malatya-Rize-Kırıkkale-Karabük-Kilis ve

Osmaniye kendi aralarında 7'li grup oluşturduğu görülmüştür. Anadolu'nun kuzeydoğusu ve İç Batı Anadolu illerinin geneli 11'li gruplar şeklinde iken, Orta Karadeniz, İç Anadolu ve Ege bölgesindeki Kütahya ve Afyon 15'li grup oluşturduğu tesbit edilmiştir. Türkiye'nin en gelişmiş illeri genelde Marmara, Ege ve Akdeniz bölgelerinde yer aldığı görülmüştür. Gelişmişlikte İstanbul ilk sıradadır. Ankara, Eskişehir, Konya ve Kayseri illeri dışındaki tüm İç Anadolu illeri ile Karadeniz Bölgesi illerinin büyük bir bölümü ve Doğu Anadolu'nun kuzeybatısındaki iller Türkiye içinde orta derecede bir gelişmişlik düzeyine sahipken, Doğu Anadolu'nun diğer illeri ve özellikle Güneydoğu Anadolu illeri Türkiye'nin en geri illeri olduğu sonucuna ulaşılmıştır. Türkiye genel olarak gelişmişlik düzeyi bakımından Zonguldak-Gazi Antep hattının doğusu ve batısı olmak üzere ikiye ayrılmıştır. Batıdaki iller doğudaki illere göre sosyo-ekonomik açıdan çok daha ileride olduğu sonucuna ulaşılmıştır.

Çakmak, Uzgören ve Kekeç(2005), yaptıkları çalışmada 73 il kültürel yapılarına göre iki ayrı dönem (1990 ve 2000 yılları) itibari ile sınıflandırılmış ve böylece 10 yıllık süreç içerisinde kültürel açıdan meydana gelen değişiklikleri incelenmiştir. Bu doğrultuda illerin kültürel yapıları bakımdan kümelenmesine yönelik analizde değişkenler; matbaa sayısı, kütüphanelerdeki toplam kitap sayısı, kütüphanelerden toplam yararlanan kişi sayısı, müzelerdeki mevcut kitap sayısı, müzelere bağlı taşınmaz kültür varlıkları sayısı, tiyatrolardaki toplam koltuk sayısı, turneye çıkan tiyatrolara giden toplam seyirci sayısı, sinema koltuk sayısı, sinemaya giden seyirci sayısı ve müzelerde kayıtlı mevcut kitap sayısı olarak belirlenmiştir. Araştırmada uzaklık matrisinin belirlenmesinde sadece kareli öklit uzaklığı (squared eucliden distance), illerin kümelendirilmesinde ise, hiyerarşik kümeleme tekniklerinden tek bağlantı tekniği ile Ward's tekniği ve hiyerarşik olmayan kümeleme yöntemlerinden K-ortalama tekniği kullanılmıştır. K-ortalama tekniğinde, çeşitli küme sayılarında değişkenlerin önem düzeyleri varyans analizi ile incelenmiştir. Elde edilen bulgular neticesinde kullanılan kümeleme tekniği ne olursa olsun bu on yıllık süreç içerisinde ele alınan değişkenler doğrultusunda İstanbul ilinin diğer tüm illerden kültürel açıdan farklı olduğu olgusunu değiştirmediği görülmüştür. Ayrıca birçok teknikte İzmir ve Ankara illerinin aynı kümeler içinde

yer aldıkları gözlenmiştir. Ancak iki dönem arasında en önemli farklılık 1990 yılına ilişkin analizlerde yer almayan Karaman ilinin 2000 yılında ayrı bir küme olarak karşımıza çıkmasıdır. Bunun nedeni, söz konusu ilimizin geçmişte önemli bir kültür merkezi olmasından kaynaklanmaktadır. Kümeleme analizinin ortaya koyduğu en önemli bulgulardan birisi de, bölgesel ve iktisadi gelişmişlik yönünden farklılıkların kültürel açıdan da bariz bir farklılığa yol açtığı belirlenmiştir. İktisadi açıdan gelişmiş illerin genelde tek bir küme oluşturması veya aynı kümeler içinde toplanmış olması bu sonucun somut bir göstergesi olduğu görülmüştür.

Turanlı, Özden ve Türedi (2006), Avrupa Birliği üye ülkeleri ile aday üye ülkeler arasında var olan ekonomik benzerlikleri ortaya koymak, aday ülkelerin üye olmak için yeterli olup olmadıklarını anlamaya yönelik bir araştırma yapmışlardır. Çalışmada AB'ye aday ve üye ülkelerin GSMH, enflasyon Oranı, işsizlik oranı, internet kullanım oranı, ömür boyu eğitim indeksi ve ithalat ihracat oranı bakımından ekonomik benzerliklerini incelemişlerdir. Ülkeleri gruplanması için kümeleme analizi yöntemlerinden olan hiyerarşik olmayan k-ortalama tekniği kullanılmıştır. Yapılan analizin neticesinde iki kümeli bir yapı elde etmişlerdir. Avrupa'nın temel iskeletini oluşturan ve onbeş ülkenin yer aldığı AB-15'in onüç ülkesi birinci kümededir. Bu kümede yer almayan iki ülkeden biri Yunanistan, bir diğeri ise Portekiz'dir. İkinci kümede ise AB-15'in en güçsüz iki ülkesinin yanı sıra yeni üye olan sekiz Merkez ve Doğu Avrupa ülkesi, birliğin iki adası Malta ve Kıbrıs Rum Kesimi ve dört aday üye ülkesi yer aldığı sonucuna ulaşmışlardır.

3.2 Analizde Kullanılan Veriler ve Değişkenler

Bu analiz için kullanılan veriler, TÜİK'in adrese dayalı altı yaş üstü nüfus verileri olup, 2010 yılına aittir. Elde edilen verilere göre, analize dahil edilecek değişkenler şu şekilde belirlenmiştir;

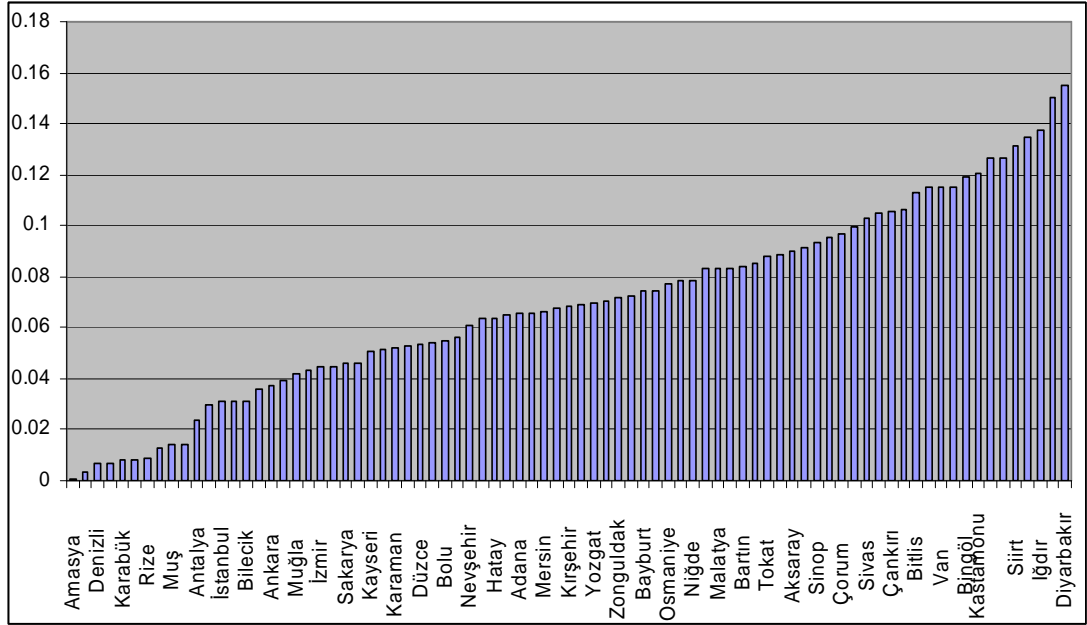
- Okuma yazma bilmeyen
- Okuma yazma bilen fakat herhangi bir okula gitmeyen
- İlköğretim mezunu

- Ortaöğretim mezunu
- Yüksek okul mezunu
- Yüksek lisans mezunu
- Doktora mezunu
- Eğitim durumu bilinmeyen oranları

Değişkenlerin arasındaki ölçek farkını ortadan kaldırmak amacıyla herbir değişken standartlaştırılmıştır. Değişkenlerin standartlaştırılmasında şu yol izlenmiştir; herhangi bir değişkene ait nüfus sayısı toplam nüfusa oranlanmıştır.

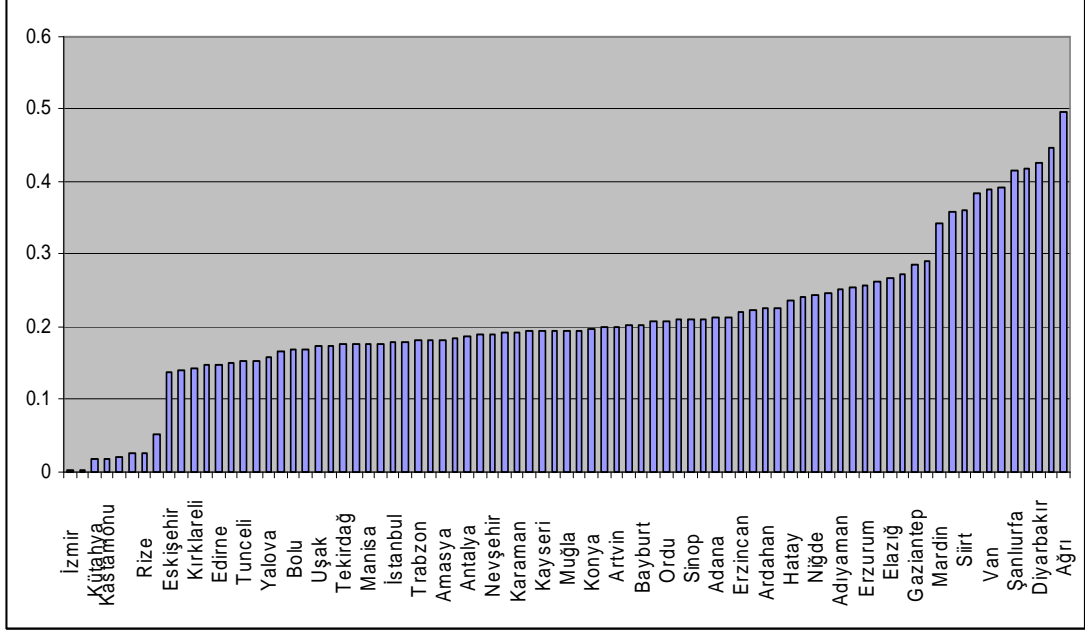
3.3. Analizin Bulguları

Standartlaştırılmış değişkenlerle yapılan analize ait bulgular, iller bazında aşağıdaki şekiller ve tablolarda verilmiştir.



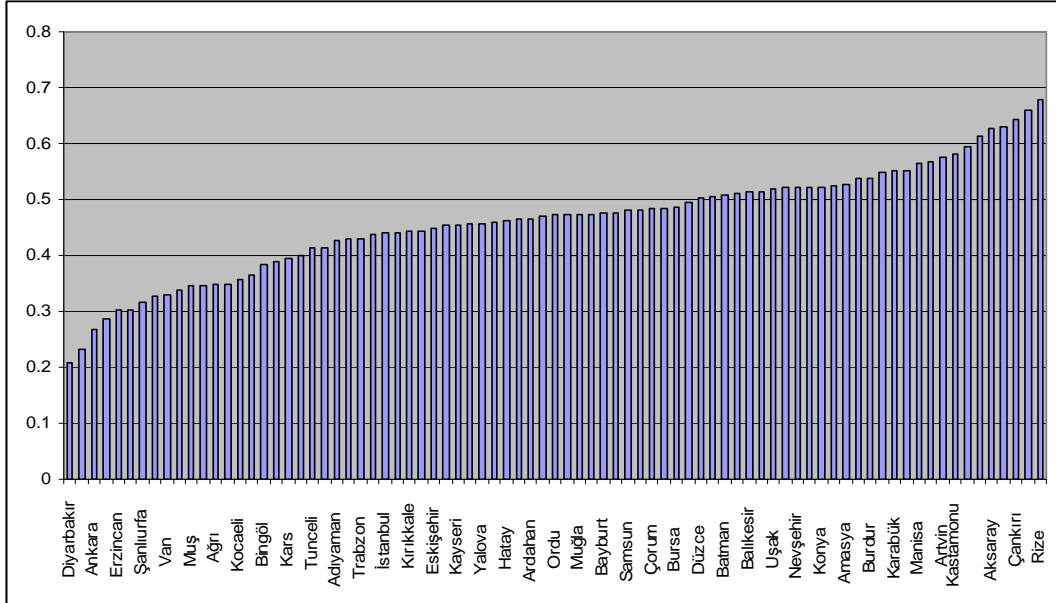
Şekil 3.1. Okuma-yazma bilmeyen oranları

Yukardaki grafikten de anlaşılacağı üzere, genellikle Türkiye'nin doğusunda kalan illerde okuma-yazma bilmeyen oranının yüksek olduğu görülmektedir.



Şekil 3.2. Okuma yazma bilen fakat bir okul bitirmeyen oranları

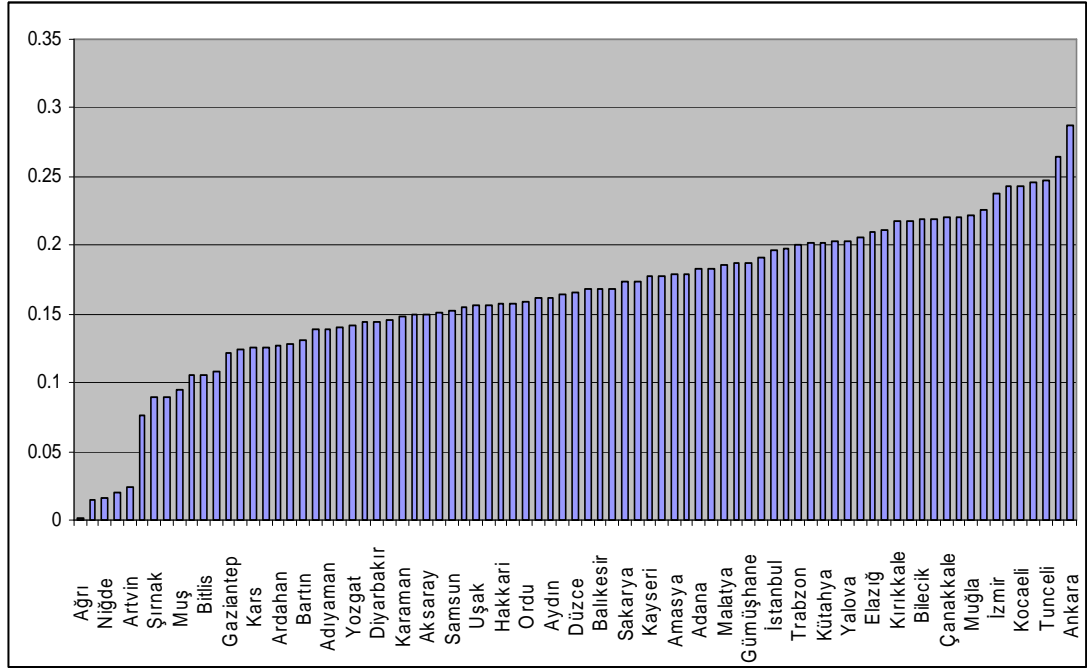
Bu grafikte de bir öncekine benzer bir durum görülmektedir. Kastamonu'nun konumu dikkat çekicidir; bir önceki değişkene göre oranı en yüksek olan iller arasında iken, bu değişkene göre oranı en düşük iller arasında bulunmaktadır.



Şekil 3.3 Sekiz yıllık eğitim mezunu oranları

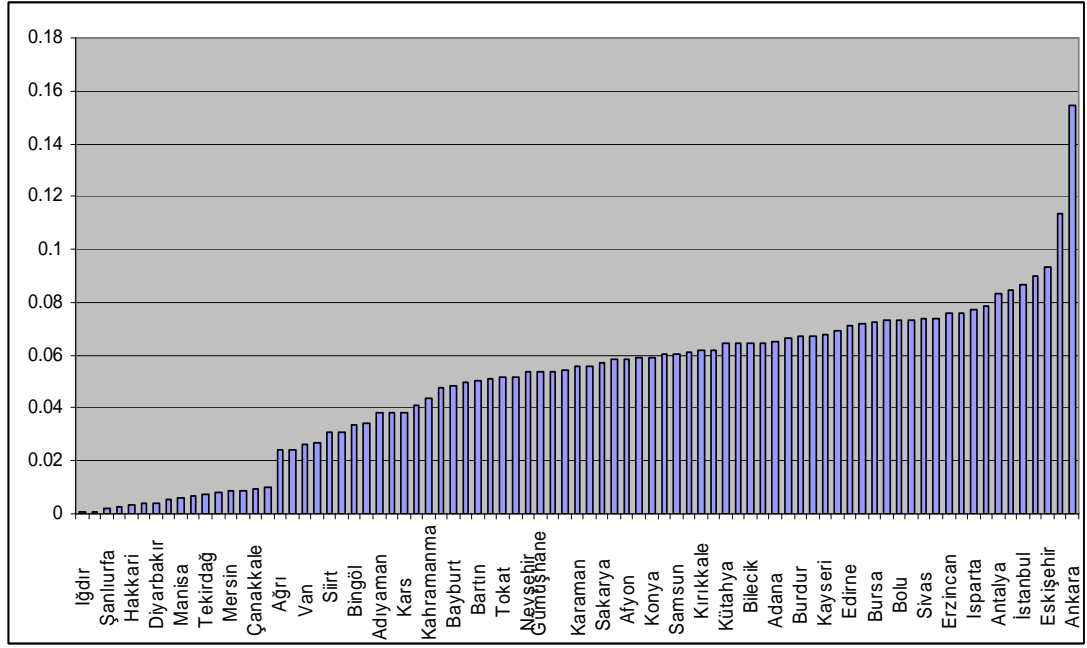
Sekiz yıllık eğitim mezunu oranlarına göre Rize en büyük orana sahiptir. Bu grafikte de Diyarbakır'dan sonra en düşük orana sahip olan Ankara'nın konumu

dikkat çekicidir. Bu değışkene göre iller arasında büyük bir kırılma görülmemektedir.



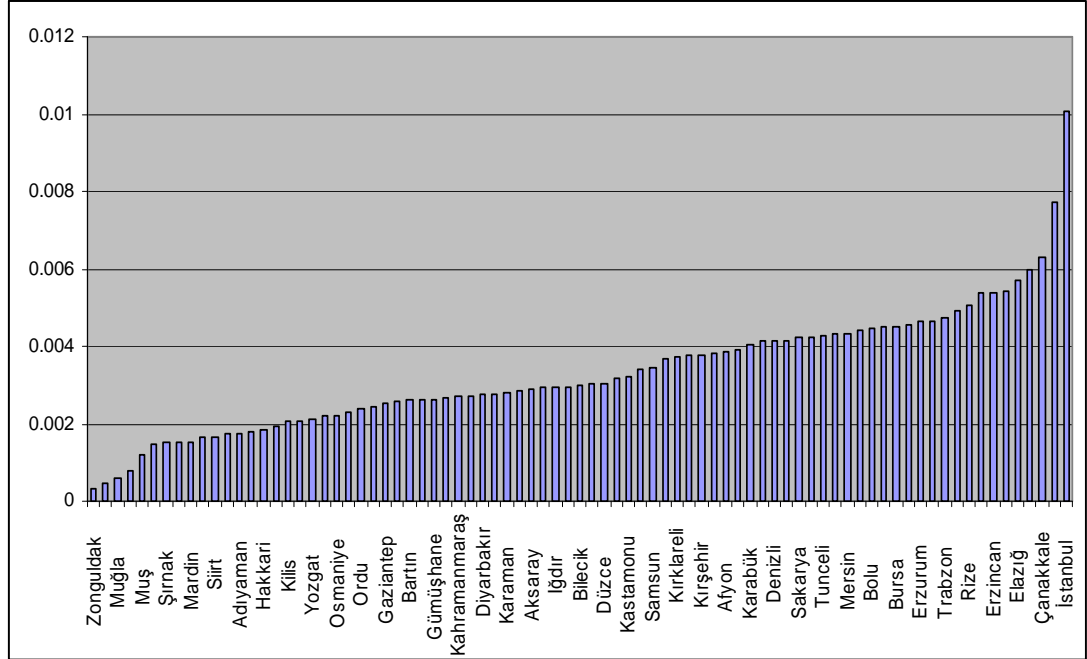
Şekil 3.4 Lise veya dengi okul mezunu oranları

Lise mezunu oranlarına göre İç Anadolu Bölgesi ve doğusundaki iller en yüksek değeri almaktadır. Doğu illeri, bu orana göre değeri en küçük olan illerdir. Burada ise Artvin ile Tunceli'nin konumları dikkat çekicidir. Diğer dikkat çeken bir husus da Şırnak'tan sonra gelen illerin değerlerinde büyük bir düşüş olmasıdır.



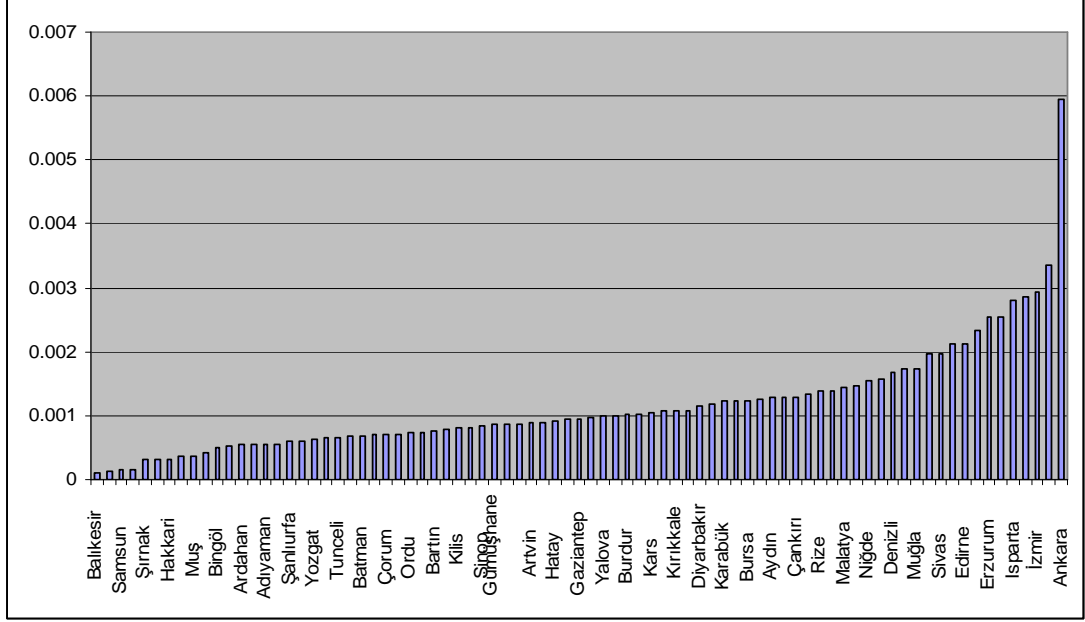
Şekil 3.5 Yüksekokul veya fakülte mezunu oranları

Yukarıdaki grafikte görüldüğü üzere Ankara diğer illerden açık bir şekilde ayrılmaktadır. Yine bu orana göre, Ağrı'dan sonra da bir kırılma görülmektedir.



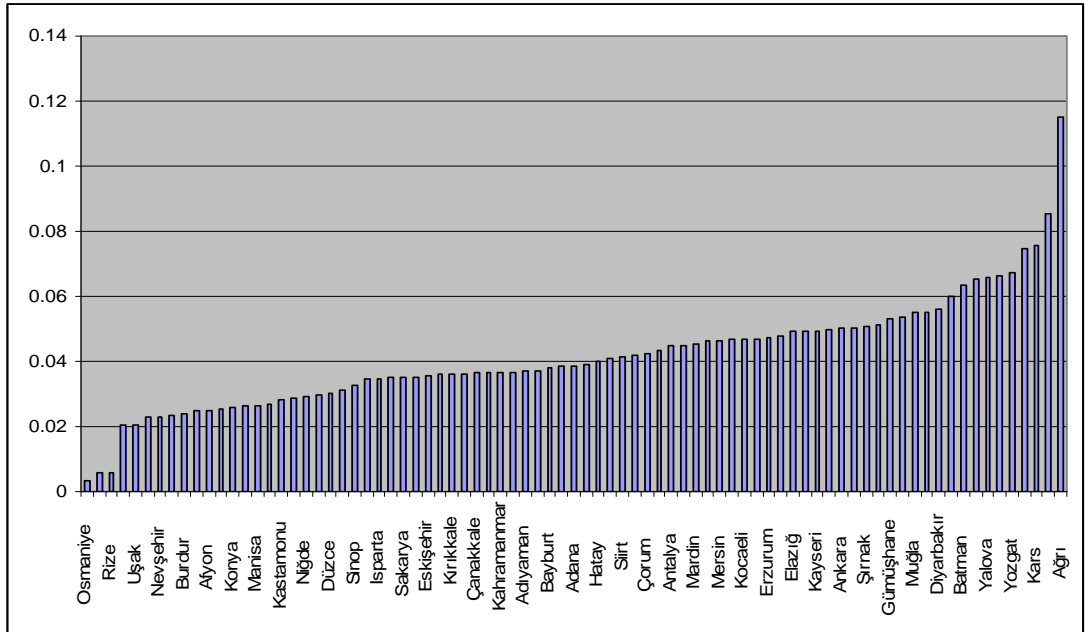
Şekil 3.6 Yüksek lisans mezunu oranları

İstanbul, yüksek lisans mezunu oranlarına göre en büyük değeri alırken, Zonguldak en düşük değeri almaktadır. Sıralamanın en altında yine doğu illeri yoğunlaşmaktadır.



Şekil 3.7 Doktora mezunu

Yüksek lisans mezunu oranlarında olduğu gibi Ankara diğer illerden ayrılmaktadır. Fakat diğer grafiklerin aksine doktora mezunu oranlarında önemli ölçüde bölgesel bir ayrışma görülmemektedir.



Şekil 3.8 Eğitim Düzeyi Bilinmeyen Oranları

Yukardaki grafiklerde herbir deęişken ayrıntılı olarak incelenmiştir. bu deęerlere göre illerin sıralamadaki konumları deęişebilmektedir. Kesin bir ayrışma olmamakla birlikte, eğitime dair oranlarda doęu illeri listenin sonlarında, batıda bulunan iller ise listenin başında yoğunlaşmaktadır.

Aşağıdaki bölümde ise bu deęişkenler istatistik analizleri yapılacaktır. Tablo 3.1 de deęişkenlerin karakteristik özellikleri, Tablo 3.2 normal dağılım testi sonuçları verilmektedir.

Tablo 3.1 Deęişkenlerin karakteristik deęerleri

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

Herbir deęişkene Kolmogorov-Smirnov testi uyguladığımızda Tablo 3.2 de verilen Z deęerleri ve anlamlılık düzeyleri elde edilir. 0,05'ten büyük olan anlamlılık deęerleri, bu deęişkenlerin %95 güvenilirlik düzeyinde tabloda belirtilen ortalama ve standart sapmaya sahip normal dağılıma uyduęu ifade etmektedir.

Tablo 3.2 Değişkenlerin Normal Dağılım Uygunluk testleri

One-Sample Kolmogorov-Smirnov Test					
	N	Mean	Std. Deviation	Kolmogorov-Smirnov Z	Asymp. Sig.
Okuma yazma bilmeyen	81	.06910074	.037398708	.474	.978
Okuma yazma bilen fakat bir okul bitirmeyen	81	.20983924	.100421092	1.417	.086
Sekizyıllık eğitim mezunu	81	.46197386	.096997585	.872	.432
Lise veya dengi okul mezunu	81	.16316297	.057213265	.892	.404
Yüksekokul veya fakülte mezunu	81	.04999174	.029375410	1.052	.218
Yüksek lisans mezunu	81	.00331561	.001612424	.917	.370
Doktora mezunu	81	.00116533	.000873350	1.540	.067
Eğitim düzeyi bilinmeyen	81	.04145051	.017310206	.910	.379

Türkiye’de 81 İl, hiyerarşik olmayan kümeleme yöntemlerinden olan k-ortalama yöntemiyle kümelere ayrılacaktır. K-ortalama yöntemiyle kümeleme, iki kümeli yapıdan sekiz kümeli yapıya kadar uygulanacaktır. Kümeleme analizinde en önemli durum küme sayısının belirlenmesidir. İkinci bölümde anlatılan küme sayısına karar verme ve küme geçerliliği tekniklerinden olan wilk’s lamda istatistiği ve Silhouette indeks değerine göre küme sayısı belirlenecektir.

Veriler kümelere ayrıldıktan sonra, elde edilen kümelere MANOVA analizi uygulanacak ve MANOVA analizi neticesinde de küme ortalamalarının birbirinden istatistiksel olarak farklı olması şartının sağlanması beklenmektedir. Bu şartı sağlayan küme sayısı uygun küme sayısı olarak kabul edilecektir. Yöntem olarak da wilk’s lamda değerinin 0.01 in altına ilk düştüğü durumdaki küme sayısı ile analizin yapılmasına devam edilecek küme sayısı olarak kabul edilecektir.

Küme sayısı belirlemede diğer bir kriter olan Silhouette indeks değerleri MATLAB yardımıyla hesaplanacaktır. Hesaplanan bu değerlere göre en büyük

Silhouette indeks deęerini veren küme sayısı uygun küme sayısı olarak belirlenecektir.

İki küme için k-ortalama yöntemi uygulanmıştır. Deęişkenlerin istatistiki anlamlılık deęerleri Tablo 3.3 de verilmiştir.

Tablo 3.3 İki küme için deęişkenlerin anlamlılık testi

ANOVA						
	Küme		Hata		F	P
	Kareler Ortalaması	Sd	Kareler Ortalaması	Sd		
Okuma yazma bilmeyen	.014	1	.001	79	11.611	.001
Okuma yazma bilen fakat bir okul bitirmeyen	.417	1	.005	79	84.594	.000
Sekizyıllık eğitim mezunu	.449	1	.004	79	117.041	.000
Lise veya dengi okul mezunu	.008	1	.003	79	2.611	.110
Yüksekokul veya fakülte mezunu	.005	1	.001	79	6.146	.015
Yüksek lisans mezunu	.000	1	.000	79	4.201	.044
Doktora mezunu	.000	1	.000	79	.024	.877
Bilinmeyen	.005	1	.000	79	20.454	.000

İki küme için uygulanan k-ortalama yönteminde, doktora mezunu ve lise ve dengi okul mezunu deęişkenlerinin anova istatistiki deęerleri 0.05 ten büyük çıkmaktadır. Bu deęişkenlerin, illerin kümelenmesinde etkili olmadığı söylenebilir.

Tablo 3.4 İki küme için kümeler

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

İki küme için yapılan analizde kümeler genel olarak Doğu Anadolu illeri ve diğerleri olarak ayrılmaktadır. Burada dikkat çeken durum, Ankara ve Kocaeli'nin Doğu Anadolu illeri ile aynı kümede, Batman ve Tunceli'nin de diğer kümede yer almasıdır.

Üç küme için k-ortalama yöntemi uygulanmış ve değişkenlerin anlamlılık istatistik değerleri Tablo 3.5 de verilmiştir. Görüldüğü gibi tüm değişkenler 81 ilin kümelenmesinde önemli bulunmuşlardır.

Tablo 3.5 Üç küme için değişkenlerin anlamlılık testi

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
Okuma yazma bilen fakat bir okul bitirmeyen	,284	2	,003	78	93,321	,000
Yüksekokul veya fakülte mezunu	,007	2	,001	78	10,824	,000
Bilinmeyen	,004	2	,000	78	16,848	,000
Okuma yazma bilmeyen	,007	2	,001	78	5,505	,006
Yüksek lisans mezunu	,000	2	,000	78	6,316	,003
Sekizyıllık eğitim mezunu	,266	2	,003	78	93,827	,000
Doktora mezunu	,000	2	,000	78	3,711	,029
Lise veya dengi okul mezunu	,028	2	,003	78	10,422	,000

Tablo 3.6 K=3 için kümeler

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

Dört küme için k-ortalama yöntemi uygulanmış ve değişkenlerin anlamlılık istatistik değerleri Tablo 3.7 de verilmiştir.

Tablo 3.7 K=4 için değişkenlerin anlamlılık testi

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	df		
Okuma yazma bilen fakat bir okul bitirmeyen	,237	3	,001	77	189,953	,000
Yüksekokul veya fakülte mezunu	,006	3	,001	77	8,548	,000
Bilinmeyen	,002	3	,000	77	9,535	,000
Okuma yazma bilmeyen	,007	3	,001	77	5,434	,002
Yüksek lisans mezunu	,000	3	,000	77	7,889	,000
Sekizyıllık eğitim mezunu	,193	3	,002	77	84,687	,000
Doktora mezunu	,000	3	,000	77	5,880	,001
Lise veya dengi okul mezunu	,027	3	,002	77	11,710	,000

Tablo 3.8 K=4 için kümeler

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

Beş küme için k-ortalama yöntemi uygulanmış ve değişkenlerin anlamlılık istatistik değerleri Tablo 3.9 da verilmiştir.

Tablo 3.9 K=5 için değişkenlerin anlamlılık testi

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
Okuma yazma bilmeyen	,009	5	,001	75	10,570	,000
Okuma yazma bilen fakat bir okul bitirmeyen	,144	5	,001	75	125,830	,000
Lise veya dengi okul mezunu	,036	5	,001	75	32,316	,000
Yüksek okul veya fakülte mezunu	,004	5	,001	75	6,313	,000
Yüksek lisans mezunu	,000	5	,000	75	3,441	,007
Doktora mezunu	,000	5	,000	75	3,976	,001
Bilinmeyen	,002	5	,000	75	7,447	,000
Sekizyıllık eğitim mezunu	,122	5	,002	75	62,977	,000

Tablo 3.10 K=5 için kümeler

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

Altı küme için k-ortalama yöntemi uygulanmış ve değişkenlerin anlamlılık istatistik değerleri Tablo 3.11 de verilmiştir.

Tablo 3.11 K=6 için değişkenlerin anlamlılık testi

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	Df		
Okuma yazma bilmeyen	,009	5	,001	75	10,570	,000
Okuma yazma bilen fakat bir okul bitirmeyen	,144	5	,001	75	125,830	,000
Lise veya dengi okul mezunu	,036	5	,001	75	32,316	,000
Yüksek okul veya fakülte mezunu	,004	5	,001	75	6,313	,000
Yüksek lisans mezunu	,000	5	,000	75	3,441	,007
Doktora mezunu	,000	5	,000	75	3,976	,003
Bilinmeyen	,002	5	,000	75	7,447	,000
Sekiz yıllık eğitim mezunu	,122	5	,002	75	62,977	,000

Tablo 3.12 K=6 için kümeler

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

Yedi küme için k-ortalama yöntemi uygulanmış ve değişkenlerin anlamlılık istatistik değerleri Tablo 3.13 de verilmiştir.

Tablo 3.13 K=7 için değişkenlerin anlamlılık testi

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	df		
Okuma yazma bilmeyen	,009	5	,001	75	10,570	,000
Okuma yazma bilen fakat bir okul bitirmeyen	,144	5	,001	75	125,830	,000
Lise veya dengi okul mezunu	,036	5	,001	75	32,316	,000
Yüksek okul veya fakülte mezunu	,004	5	,001	75	6,313	,000
Yüksek lisans mezunu	,000	5	,000	75	3,441	,007
Doktora mezunu	,000	5	,000	75	3,976	,005
Bilinmeyen	,002	5	,000	75	7,447	,000
Sekiz yıllık eğitim mezunu	,122	5	,002	75	62,977	,000

Tablo 3.14 K=7 için kümeler

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

Sekiz küme için k-ortalama yöntemi uygulanmış ve değişkenlerin anlamlılık istatistik değerleri Tablo 3.14 de verilmiştir.

Tablo 3.15 K=8 için değişkenlerin anlamlılık testi

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	df		
Okuma yazma bilmeyen	,009	5	,001	75	10,570	,000
Okuma yazma bilen fakat bir okul bitirmeyen	,144	5	,001	75	125,830	,000
Lise veya dengi okul mezunu	,036	5	,001	75	32,316	,000
Yüksek okul veya fakülte mezunu	,004	5	,001	75	6,313	,000
Yüksek lisans mezunu	,000	5	,000	75	3,441	,001
Doktora mezunu	,000	5	,000	75	3,976	,005
Bilinmeyen	,002	5	,000	75	7,447	,000
Sekiz yıllık eğitim mezunu	,122	5	,002	75	62,977	,000

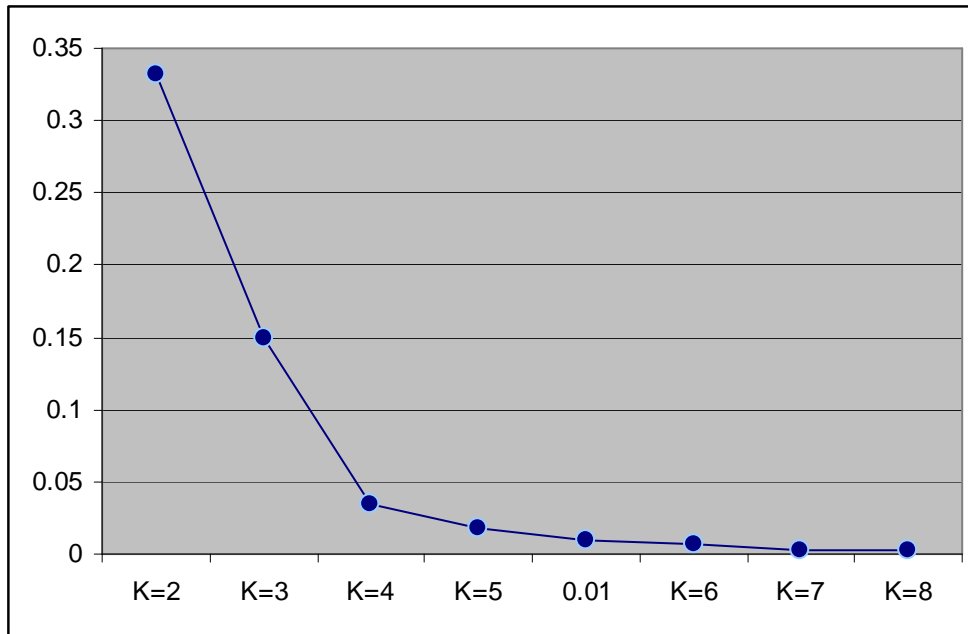
Tablo 3.16 K=8 için kümeler

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

Küme sayısının iki olduğu durum haricinde, bütün değişkenlerin ANOVA analizi neticesinde anlamlı sonuçlar verdiği ve bu değişkenlerin 81 ilin kümelenmesinde önemli olduğu görülmüştür. Yapılan kümelemenin neticesinde, ikiden sekize kadar elde edilen farklı sayıdaki kümelere MANOVA analizi uygulanmış ve Tablo 3.16 da verilen wilk's lamda değerleri elde edilmiştir.

Tablo 3.17 Wilk's lamda değerleri

KÜME SAYISI	WILK'S LAMDA DEĞERİ
K=2	0.3327
K=3	0.1499
K=4	0.0343
K=5	0.0183
K=6	0.0064
K=7	0.0031
K=8	0.0021



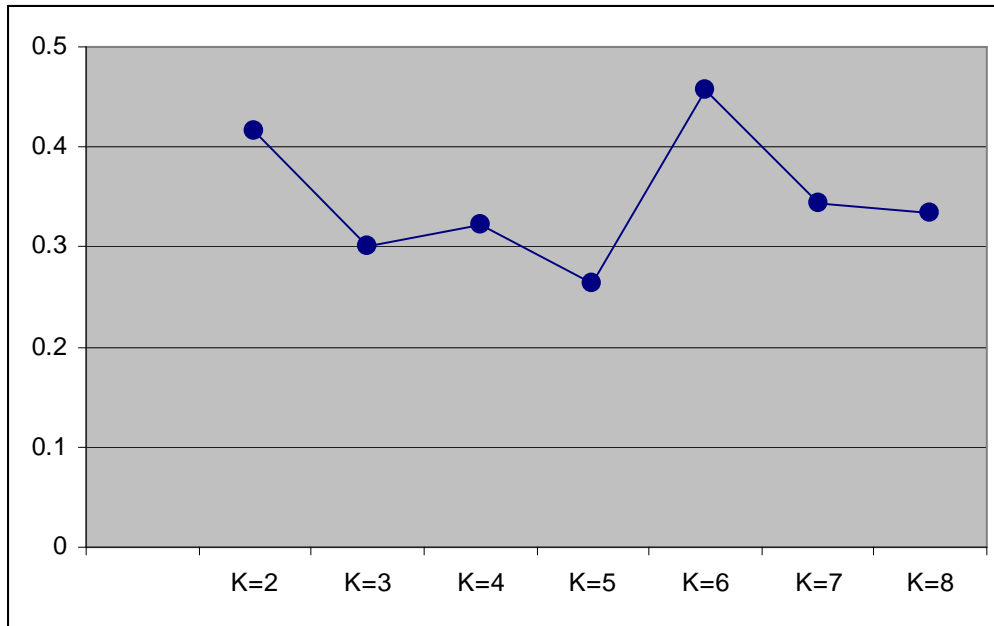
Şekil 3.9 Kümeleme yöntemlerinin wilk's lamda değerine göre gösterimi

Küme sayısı altı olduğunda, Wilk's lamda değeri ilk kez 0.01 altına düşmektedir. Wilk's lamda istatistiğine göre uygun küme sayısı altıdır.

Uygun küme sayısını belirleme kriterlerinden bir diğeri de Silhouette indeks değeridir. Silhouette indeks değerlerinin hesaplanması için MATLAB programı kullanılmış ve Tablo 3.17 de verilen değerler elde edilmiştir.

Tablo 3.18 Silhouette indeks değerleri

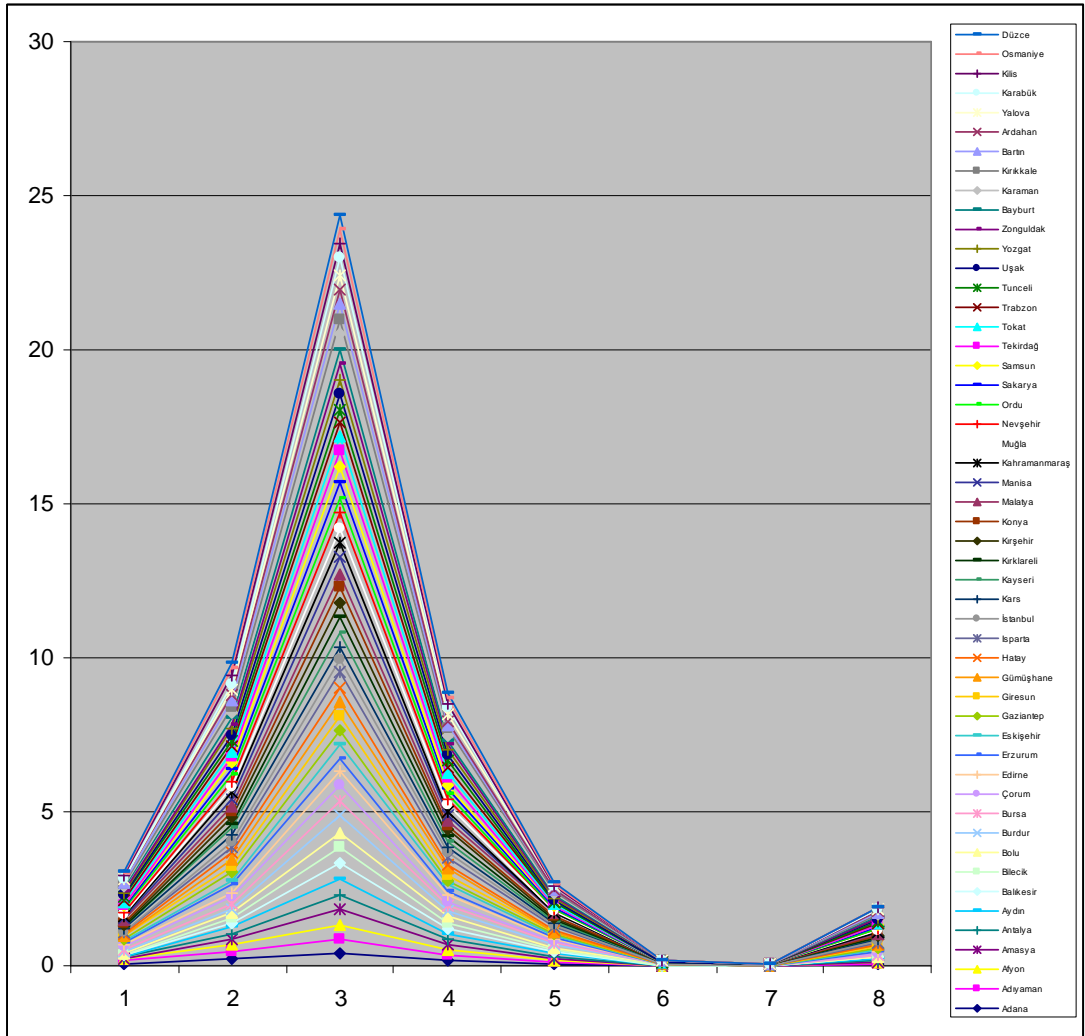
KÜME SAYISI	SİLHOUETTE İNDEKS DEĞERİ
K=2	0.4362
K=3	0.3009
K=4	0.3232
K=5	0.2643
K=6	0.4401
K=7	0.3437
K=8	0.3334



Şekil 3.10 Kümeleme yöntemlerinin Silhouette indeks değerine göre gösterimi

Silhouette indeks deęerini en byk yapan kme sayısı uygun kme sayısı olacaęından, bu deęere gre de uygun kme sayısı altıdır. İki farklı kritere gre uygun kme sayısı altı olarak bulunmuř oldu. Bu durumda, 81 il altı kmeye ayrılacaktır.

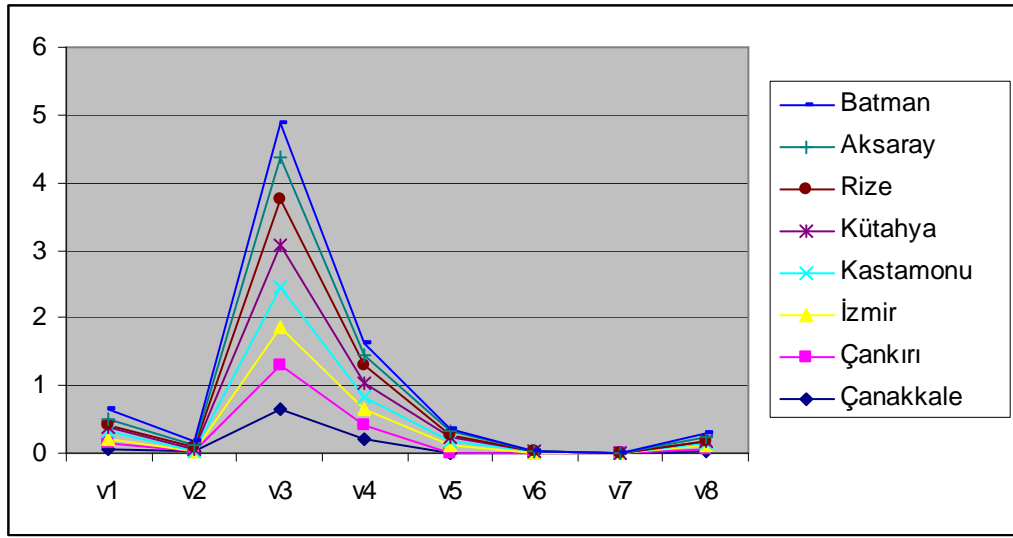
Ařaęıda, oluřturulan altı kme ve bu kmelerin oluřumunda hangi deęiřkenlerin daha etkin olduęunu gsteren řekiller yer almaktadır. Birinci kmeyi oluřturan iller ve bu kmenin oluřumunda etkili olan deęiřkenler řekil 3.11 de verilmiřtir.



řekil 3.11 Birinci kme

Birinci küme 51 ilden oluşmaktadır. Ankara hariç hemen hemen bütün büyük iller bu kümede gruplanmışlardır. Küme oluşumunda etkili değişkenler yüksek lisans mezunu ve doktora mezunu oranları olmuştur.

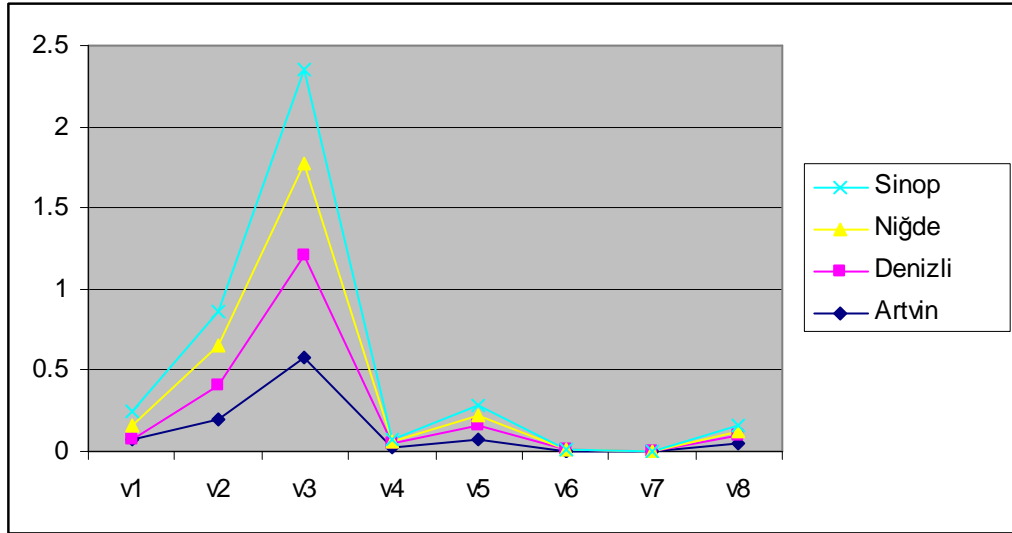
İkinci kümeyi oluşturan iller ve bu kümenin oluşumunda etkili olan değişkenler Şekil 3.12 de verilmiştir.



Şekil 3.12 İkinci küme

Kümeyi oluşturan iller Türkiye'deki çok farklı bölgelerdeki illerden oluşmaktadır. Şekil 3.12 den bu kümeyi belirleyen en önemli değişkenler okuma-yazma bilen fakat herhangi bir okula gitmeyen, yüksek lisans mezunu ve doktora mezunu oranları olmuştur.

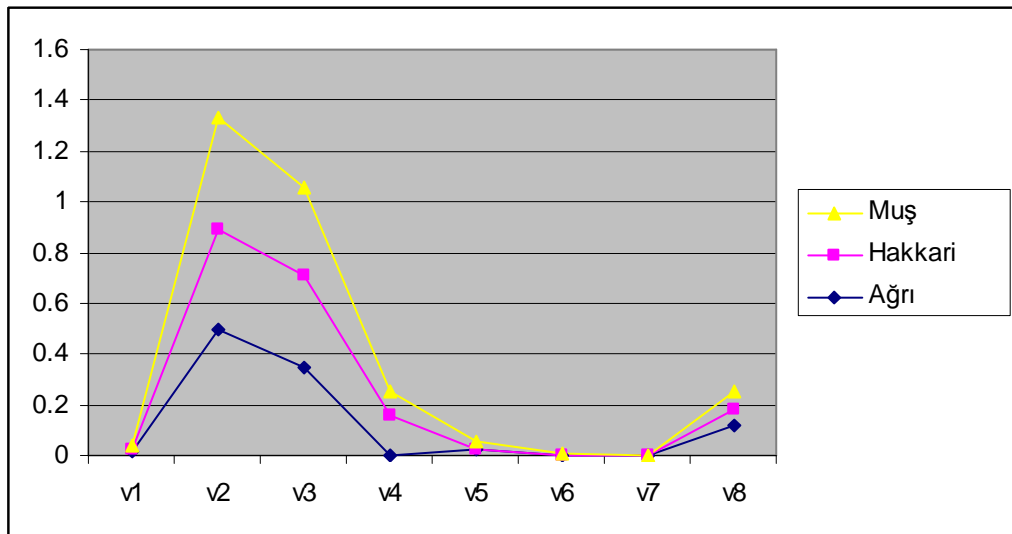
Üçüncü kümeyi oluşturan iller ve bu kümenin oluşumunda etkili olan değişkenler Şekil 3.13 de verilmiştir.



Şekil 3.13 Üçüncü küme

Dört ilden oluşan küme olmasına karşın farklı coğrafyalardan iller aynı kümede toplanmıştır. Özellikle Denizli ile Artvin'in aynı kümede olması dikkat çekicidir. Kümeyi belirleyen değişkenler lise veya dengi okul mezunu, yüksek lisans mezunu ve doktora mezunu oranları olduğu söylenebilir.

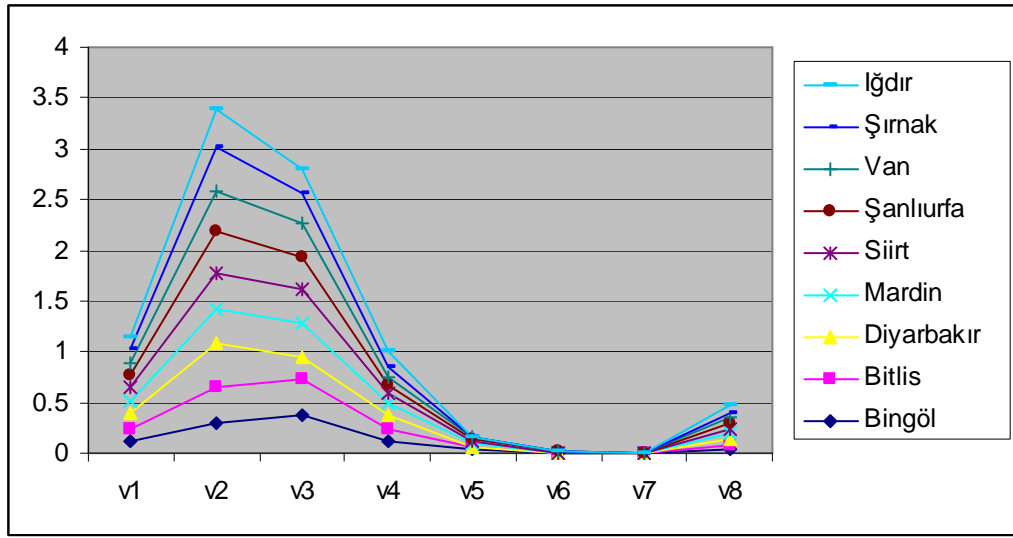
Dördüncü kümeyi oluşturan iller ve bu kümenin oluşumunda etkili olan değişkenler Şekil 3.14 de verilmiştir.



Şekil 3.14 Dördüncü küme

Kümelenen iller coğrafi konumları itibariyle birbirine yakın illerden oluşmaktadır. Bu yönüyle en homojen grup olduğu söylenebilir. Şekil 3.14 den dördüncü kümenin oluşumunda etkin değişkenler okuma-yazma bilmeyen, yüksek okul veya fakülte mezunu, yüksek lisans mezunu, doktora mezunu oranları olmuştur

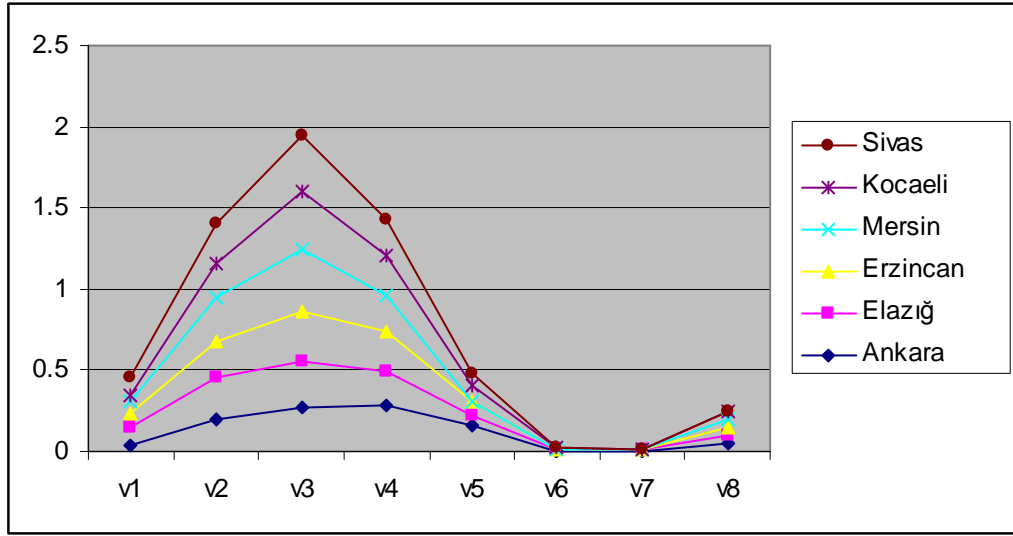
Beşinci kümeyi oluşturan iller ve bu kümenin oluşumunda etkili olan değişkenler Şekil 3.15 de verilmiştir.



Şekil 3.15 Beşinci küme

Bu kümeyi oluşturan illerin coğrafi olarak birbirine yakınlığı dikkat çekmektedir. Coğrafi konum itibari ile homojen olan diğer küme olduğu söylenebilir. Şekil 3.15 den beşinci kümenin oluşumunda etkili olan değişkenler yüksek okul veya fakülte mezunu, yüksek lisans mezunu, doktora mezunu oranları olduğu görülmektedir.

Altıncı kümeyi oluşturan iller ve bu kümenin oluşumunda etkili olan değişkenler Şekil 3.16 da verilmiştir.



Şekil 3.16 Altıncı küme

Bu kümeyi oluşturan iller farklı bölgelerdeki illerden oluşmaktadır. Şekil 3.16 dan bu kümenin oluşumunda en önemli değişkenler yüksek lisans mezunu, doktora mezunu oranları olduğu söylenebilir.

SONUÇ

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

İllerin gruplanması aşamasında yöntem olarak kümeleme analizi kullanılmıştır. Türkiye’de illerin eğitim durumlarını gösteren değişkenler belirlenmiş ve ANOVA analiziyle bu değişkenlerden hangilerinin illerin kümelenmesinde önemli oldukları belirlenmiştir. Analizin uygulanmasında fazla sayıda birim olması ve sonuçların daha kolay yorumlanabilmesi nedeniyle hiyerarşik olmayan kümeleme analiz yöntemlerinden k-ortalama yöntemi kullanılmıştır. İller iki kümeden sekize kadar farklı sayıda kümelere ayrılmıştır. Wilk’s lamda istatistiği, Silhouette indeks değerleri yardımıyla uygun küme sayısı belirlenmiştir. Wilk’s lamda değerlerinin hesaplanması için MANOVA analizi uygulanmıştır. Silhouette indeks değerlerinin hesaplanması için ise MATLAB programı kullanılmıştır. Bu iki kritere göre de uygun küme sayısı altı olarak belirlenmiştir. Analiz, SPSS 17 istatistik programıyla yapılmıştır.

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

İkinci kümeyi oluşturan iller Çanakkale, Çankırı, İzmir, Kastamonu, Kütahya, Rize Aksaray, Batman olmuştur. Türkiye’nin coğrafi olarak birbirine en uzak illerinin bir arada

olması sebebiyle dikkat çekicidir. Coğrafi olarak en heterojen yapıdaki kümedir. Bu küme yardımıyla, illerin eğitim düzeyinin belirlenmesinde coğrafi konumun bir sınır olamayacağı söylenebilir. Bu kümeyi oluşturan en önemli etkenler okuma-yazma bilen fakat herhangi bir okula gitmeyen, yüksek lisans mezunu ve doktora mezunu oranları olmuştur. Şekil 2.3 den de görülebileceği üzere bu iller okuma-yazma bilen fakat bir okul bitirmeyen oranları diğer illere ait oranlardan büyük bir şekilde farklılaşmaktadır.

Üçüncü kümede de, ikinci kümeye benzer bir durum dikkat çekmektedir. Artvin, Niğde Denizli, Sinop gibi farklı bölgelere ait iller bir kümede toplanmaktadır. Yukarıda ifade ettiğimiz gibi; coğrafi farklılaşmalar, iller arasında, eğitim düzeyleri yönüyle benzerlik göstermesine engel teşkil etmemiştir. Bu kümeyi oluşturan önemli faktörler lise veya dengi okul mezunu, yüksek lisans mezunu ve doktora mezunu oranları olduğu söylenebilir. Özellikle, şekil 3.4 ten lise ve dengi okul mezunu oranlarında, bu illerin değerlerinin diğerlerinden farklılaştığı görülmektedir.

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

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

Yapılan analiz neticesinde; Güney Doğu Anadolu bölgesindeki iller genel olarak benzerlik gösterdiği ve eğitim düzeyleri itibariyle diğer illerden geri kaldıkları gözlemlenmiştir. Türkiye'de politika üretenlerin, eğitim alanında yapacakları politikalarda bu durumu göz ardı etmemeleri gerekmektedir.

Sonu olarak; daha nce benzer analizlerde kullanılmıř olan bir yntem daha nce denenmemiř deęiřkenlerle uygulanmıřtır. Bu alıřma aıklayıcı ve tutarlı bilgiler iermekle beraber, bařaka deęiřkenlerin de analize dahil edilmesiyle kapsamı geniřletilmeye aıktır.

Ek 1. İllerin eğitim düzeyi oranları

	Okuma yazma bilmeyen	Okuma yazma bilen fakat bir okul bitirmeyen	İlkokul mezunu	İlköğretim mezunu	Ortaokul veya dengi okul mezunu	Lise veya dengi okul mezunu	Yüksekokul veya fakülte mezunu	Yüksek lisans mezunu	Doktora mezunu	Eğitim Durumun bilinmeyen
Adana	0.06535	0.21183	0.21500	0.16321	0.05156	0.18333	0.06548	0.00433	0.00139	0.03852
Adıyaman	0.10507	0.25310	0.19166	0.19755	0.03676	0.13858	0.03804	0.00173	0.00055	0.03695
Afyonkarahisar	0.07819	0.22393	0.36572	0.02030	0.05539	0.16794	0.05855	0.00387	0.00129	0.02482
Ağrı	0.01315	0.49721	0.17602	0.15218	0.01926	0.00090	0.02412	0.00152	0.00052	0.11513
Amasya	0.00056	0.18240	0.29176	0.18857	0.04643	0.17847	0.06716	0.00296	0.00071	0.04096
Ankara	0.03720	0.19377	0.02403	0.16865	0.07597	0.28769	0.15472	0.00178	0.00594	0.05024
Antalya	0.02381	0.18690	0.26499	0.16006	0.04836	0.18345	0.08299	0.00456	0.00013	0.04474
Artvin	0.07055	0.20052	0.31618	0.19059	0.06960	0.02393	0.07414	0.00375	0.00090	0.04983
Aydın	0.00664	0.17756	0.32198	0.17147	0.04335	0.16229	0.07344	0.00368	0.00129	0.03829
Balıkesir	0.05609	0.15289	0.32038	0.14273	0.04913	0.16834	0.07162	0.00382	0.00010	0.03488
Bilecik	0.03122	0.13958	0.27573	0.19294	0.04466	0.21864	0.06464	0.00298	0.00067	0.02894
Bingöl	0.11938	0.29003	0.16667	0.19107	0.02528	0.12579	0.03344	0.00173	0.00049	0.04613
Bitlis	0.11309	0.35964	0.14693	0.17427	0.02504	0.10605	0.03115	0.00165	0.00036	0.04182
Bolu	0.05464	0.16787	0.30771	0.15930	0.00439	0.20136	0.07299	0.00448	0.00212	0.02515
Burdur	0.05437	0.15163	0.32866	0.16478	0.04386	0.16187	0.06698	0.00302	0.00103	0.02379
Bursa	0.03937	0.16924	0.25884	0.16493	0.06285	0.19128	0.07257	0.00452	0.00124	0.03517
Çanakkale	0.04619	0.01920	0.42879	0.17940	0.05205	0.21996	0.00937	0.00629	0.00234	0.03641
Çankırı	0.10533	0.00219	0.36192	0.21423	0.06645	0.20599	0.00065	0.00466	0.00129	0.03729
Çorum	0.09669	0.19571	0.29609	0.18321	0.00377	0.12830	0.05083	0.00241	0.00071	0.04227
Denizli	0.00643	0.21102	0.36849	0.20528	0.05561	0.01961	0.08452	0.00412	0.00169	0.04325
Diyarbakır	0.15519	0.42644	0.15814	0.02191	0.02938	0.14460	0.00428	0.00276	0.00115	0.05617
Edirne	0.05137	0.14933	0.33081	0.12684	0.05184	0.18661	0.07098	0.00536	0.00213	0.02474
Elazığ	0.10603	0.26681	0.02303	0.20657	0.05754	0.20944	0.07297	0.00571	0.00285	0.04904
Erzincan	0.09554	0.22201	0.02635	0.21880	0.05725	0.24618	0.07604	0.00536	0.00126	0.05122
Erzurum	0.08559	0.25792	0.20419	0.15719	0.03752	0.15104	0.05186	0.00464	0.00254	0.04751
Eskişehir	0.02951	0.13871	0.25726	0.12913	0.06243	0.24279	0.09359	0.00775	0.00335	0.03548
Gaziantep	0.07252	0.28648	0.21862	0.20846	0.03814	0.12151	0.00419	0.00255	0.00094	0.04659
Giresun	0.09943	0.18388	0.23606	0.15431	0.04843	0.16853	0.05845	0.00318	0.00075	0.04699
Gümüşhane	0.00812	0.21023	0.25501	0.18110	0.04777	0.18750	0.05371	0.00264	0.00087	0.05305

Hakkari	0.01412	0.39247	0.12326	0.22080	0.02131	0.15719	0.00322	0.00185	0.00031	0.06549
Hatay	0.06332	0.23519	0.24684	0.17495	0.04010	0.14610	0.04996	0.00256	0.00091	0.04008
Isparta	0.04469	0.14785	0.25953	0.17413	0.04251	0.21165	0.07729	0.00493	0.00280	0.03461
Mersin	0.06654	0.26306	0.32832	0.00020	0.06040	0.22073	0.00872	0.00433	0.00133	0.04636
İstanbul	0.03110	0.17909	0.22516	0.15892	0.05638	0.19661	0.08666	0.01008	0.00254	0.05346
İzmir	0.04435	0.00177	0.30350	0.17948	0.06621	0.23842	0.11329	0.00080	0.00293	0.04925
Kars	0.09108	0.27287	0.21625	0.17338	0.00362	0.12575	0.03831	0.00230	0.00106	0.07537
Kastamonu	0.12639	0.02061	0.36028	0.18671	0.04849	0.16397	0.06123	0.00320	0.00087	0.02825
Kayseri	0.05078	0.19477	0.23927	0.16670	0.04771	0.17723	0.06804	0.00450	0.00156	0.04945
Kırklareli	0.03591	0.14289	0.31303	0.15087	0.05779	0.20305	0.06939	0.00374	0.00060	0.02273
Kırşehir	0.06826	0.18936	0.26599	0.17108	0.00561	0.19805	0.06681	0.00377	0.00107	0.02999
Kocaeli	0.04343	0.21215	0.26787	0.02138	0.06861	0.24371	0.08974	0.00597	0.00016	0.04698
Konya	0.05286	0.19657	0.31010	0.17353	0.03931	0.14069	0.05901	0.00044	0.00172	0.02575
Kütahya	0.07463	0.01847	0.35543	0.20745	0.05173	0.20228	0.06418	0.00421	0.00124	0.02038
Malatya	0.08316	0.20302	0.19064	0.16925	0.05361	0.18571	0.06470	0.00342	0.00143	0.04506
Manisa	0.06475	0.17757	0.34415	0.18233	0.03959	0.15610	0.00573	0.00272	0.00079	0.02627
Kahramanmaraş	0.08356	0.24061	0.22971	0.18222	0.04166	0.13857	0.04345	0.00271	0.00090	0.03660
Mardin	0.12650	0.34339	0.14360	0.18183	0.02267	0.10816	0.02672	0.00152	0.00031	0.04530
Muğla	0.04164	0.19548	0.39767	0.01820	0.05755	0.22182	0.01039	0.00058	0.00173	0.05493
Muş	0.01391	0.44578	0.14627	0.17758	0.02114	0.09477	0.02434	0.00118	0.00037	0.07466
Nevşehir	0.06068	0.18946	0.29227	0.17512	0.05299	0.14946	0.05348	0.00264	0.00086	0.02303
Niğde	0.07870	0.24421	0.31571	0.20352	0.04708	0.01572	0.06023	0.00414	0.00154	0.02915
Ordu	0.11507	0.20797	0.26207	0.16101	0.05003	0.15888	0.00514	0.00237	0.00074	0.03672
Rize	0.00848	0.02638	0.35487	0.25529	0.06925	0.26438	0.00884	0.00507	0.00139	0.00605
Sakarya	0.04572	0.17790	0.27982	0.17287	0.05194	0.17397	0.05713	0.00421	0.00146	0.03498
Samsun	0.06791	0.19874	0.26861	0.17255	0.03956	0.15214	0.06031	0.00346	0.00015	0.03657
Siirt	0.13124	0.36201	0.11964	0.18671	0.02050	0.10553	0.03067	0.00166	0.00043	0.04161
Sinop	0.09335	0.21053	0.34819	0.18932	0.04316	0.01480	0.06464	0.00268	0.00084	0.03249
Sivas	0.10283	0.24768	0.27711	0.00214	0.05820	0.22639	0.07376	0.00389	0.00197	0.00603
Tekirdağ	0.00364	0.17683	0.29989	0.18625	0.06411	0.21772	0.00726	0.00441	0.00100	0.03890
Tokat	0.08786	0.19503	0.26478	0.17637	0.03966	0.14459	0.05166	0.00287	0.00103	0.03615
Trabzon	0.06896	0.18184	0.22626	0.15573	0.04832	0.20020	0.07608	0.00472	0.00196	0.03592
Tunceli	0.11483	0.15248	0.20830	0.15930	0.04543	0.24762	0.00700	0.00429	0.00067	0.06008
Şanlıurfa	0.12063	0.41495	0.13847	0.16152	0.01747	0.07634	0.00231	0.00146	0.00060	0.06625

Uşak	0.06570	0.17290	0.29026	0.18099	0.04811	0.15589	0.06198	0.00294	0.00081	0.02042
Van	0.11484	0.39026	0.14391	0.16519	0.01960	0.08954	0.02587	0.00191	0.00097	0.04794
Yozgat	0.06958	0.20660	0.27564	0.15866	0.03955	0.14152	0.03818	0.00213	0.00062	0.06753
Zonguldak	0.07185	0.16742	0.27516	0.17187	0.04785	0.17956	0.05878	0.00032	0.00094	0.02626
Aksaray	0.09014	0.02630	0.34290	0.23378	0.04916	0.14966	0.05368	0.00289	0.00117	0.05033
Bayburt	0.07450	0.20390	0.23661	0.20466	0.03351	0.15778	0.04811	0.00220	0.00055	0.03819
Karaman	0.05217	0.19282	0.31829	0.20245	0.00378	0.14815	0.05561	0.00282	0.00071	0.02318
Kırıkkale	0.06329	0.17454	0.20985	0.16197	0.07013	0.21750	0.06152	0.00412	0.00108	0.03600
Batman	0.15015	0.05280	0.17931	0.29185	0.03735	0.17432	0.04742	0.00278	0.00069	0.06333
Şırnak	0.13435	0.41844	0.11532	0.17475	0.01339	0.08897	0.00236	0.00152	0.00031	0.05060
Bartın	0.08382	0.18006	0.28527	0.20171	0.03330	0.13060	0.05054	0.00263	0.00077	0.03128
Ardahan	0.08886	0.22664	0.25456	0.17422	0.03653	0.12710	0.03416	0.00206	0.00055	0.05532
İğdır	0.13732	0.38553	0.01831	0.21144	0.00343	0.15434	0.00043	0.00294	0.00067	0.08558
Yalova	0.03115	0.15799	0.23854	0.15339	0.06448	0.20343	0.07874	0.00545	0.00099	0.06585
Karabük	0.00809	0.18225	0.29021	0.20072	0.05916	0.21934	0.00798	0.00404	0.00124	0.02698
Kilis	0.08314	0.25383	0.19804	0.22437	0.03787	0.12468	0.04077	0.00205	0.00080	0.03443
Osmaniye	0.07716	0.22674	0.22449	0.17683	0.05445	0.17833	0.05564	0.00220	0.00055	0.00360
Düzce	0.05356	0.19095	0.27294	0.18846	0.04002	0.16519	0.05438	0.00303	0.00109	0.03039

KAYNAKLAR

- Anderberg, M.R. (1973). Cluster Analysis For Applications. New York: Academic Press
- Atbaş A.C.G. (2008) *Kümeleme analizinde küme sayısının belirlenmesi üzerine bir çalışma*
Yayımlanmış yüksek lisans tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü
- Blashfield, R.K. and Aldenderfer, M.S. (1978), The Literature on Cluster Analysis, *Multivariate Behavioral Research*, 13, 271-295
- Çakmak Z. (1999), Kümeleme Analizinde Geçerlilik Problemi ve Kümeleme Sonuçlarının Değerlendirilmesi, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 3, 187-205
- Çakmak Z., Uzgören N. ve Kekeç G. (2005) Kümeleme Analizi Teknikleri İle İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerinin İncelenmesi, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 12, 125-147
- Calinski, R.B. and Harabasz, J. (1974), A dendrite Method For Cluster Analysis, *Communications in Statistics*, 3, 1-27.
- Dibb, S. (1998), Market Segmentation: Strategies for success, *Marketing Intelligence & Planning*, 16(7), 394-406.
- Dinçer, K.S. ve Özdamar, K. (1992), Kümeleme çözümlemesinde uygun kümeleme ölçütlerinin karşılaştırılması. *Hacettepe Fen ve Mühendislik Bilimleri Dergisi*. 14: 17-33.
- Everitt, B. (1974). Cluster Analysis. Heinmann. London
- Everitt, B.S. 1979. Unresolved problems in cluster analysis, *Biometrics*, 35, 169-181
- Eğitim İstatistikleri. (b.t.). 13.9.2011, <http://www.tuik.gov.tr/VeriBilgi>
- Han, J. and Kamber, M. (2001), Data Mining Concepts and Techniques, Morgan Kaufmann Publishers Inc n Publishers Inc.
- Hofman I. and Jarvis, R. (1998), Robust and Efficient Cluster Analysis Using a Shared Near Neighbours Approach. 14. Pattern Recognition Conference, 1, 243
- Johnson, R.A (2002), Applied Multivariate Statistical Analysis, N.J: Prentice Hall Press,
- Johnson, A.R. and Wichern, D.W. (1988), Applied Multivariate Statistical Analysis, New Jersey: Prentice-Hall International Editions

- Kalaycı, Ş (2005) SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, Ankara: Asil Yayın Dağıtım
- Karabulut M., Gürbüz M. ve Sandal E.K.(2004), Hiyerarşik Kluster (Küme)Tekniği Kullanarak Türkiye’de İllerin Sosyo-Ekonomik Benzerliklerinin Analizi, Ankara Üniversitesi Coğrafi Bilimler Dergisi, 2(2), 72-85
- Kinnear, C.T. and Taylor, R.J. (1979),Marketing Research, New York: Mc Grow Hill Book
- Marriot, F.H.C.(1971), Practical Problems in a method of Cluster Analysis, Biometrics, 27,501-514
- Rousseuw, P.J. (1987), Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis.Journal of Computational and Applied Mathematics, 20,53-65.
- Ruiz, F.J.M. 1998. Strategic Group Analysis in Strategic Marketing: An application to Spanish Saving Banks, Marketing intelligence & Planning, 16(4)
- Tatlıdil, H. (1996). Uygulamalı Çok Değişkenli Analiz, Ankara: Cem Web Ofset Ltd. Şti.
- Turanlı M., Özden Ü.H., Türedi S. Avrupa Birliği’ne Aday ve Üye Ülkelerin Ekonomik Benzerliklerinin Kümeleme Analiziyle İncelenmesi, İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, 5 (9), 95-108
- Türk Eğitim Sistemi. (b.t). 2.10.2011,<http://www.munster.meb.gov.tr/mesistemi.htm>