

**TÜRKİYE'DE SAĞLIK AÇISINDAN
HOMOJEN İL GRUPLARININ BELİRLENMESİNE
İLİŞKİN İSTATİSTİKSEL BİR YAKLAŞIM**

BAHATTİN HAMARAT
Yüksek Lisans Tezi

İstatistik Anabilim Dalı
HAZİRAN - 1998

ÖZET
Yüksek lisans Tezi

**TÜRKİYE'DE SAĞLIK AÇISINDAN HOMOJEN İL GRUPLARININ
BELİRLENMESİNE İLİŞKİN İSTATİSTİKSEL BİR YAKLAŞIM**

BAHATTİN HAMARAT

Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. A. Fuat YÜZER
1998, Sayfa 75

Türkiye'de sağlık açısından gelişmenin dengeli olabilmesi için alınacak önlemler ve uygulanacak sağlık politikalarının belirlenmesi gerekmektedir. Bu nedenle, beşeri ve fiziki kaynakların dağılımı ile yatırımlarının dağılımı açısından birbirine benzer olan iller arasında ilişki sağlanmalıdır. Bu ilişkiye göre illere yapılacak yatırımlar arasında denge oluşturulmalıdır. Böylece sağlık açısından gelişmemiş ve gelişmekte olan iller, gelişmiş iller düzeyine yükseltilebilir. Bu çalışmada sağlık alanında bölgesel gelişme politikaları oluşturma süreçlerine analitik ve görsel bilgi sağlayacak, aynı özellikleri taşıyan homojen il gruplarının belirlenmesi amaçlanmıştır.

Bu çalışmada Türkiye'de bulunan illerin sağlık, nüfus, ve bazı ekonomik ve kültürel göstergelerine göre gruplandırılması yapılmıştır. Göstergelere ait 30 değişkenle çalışılmış ve il gruplarının elde edilmesinde Kümeleme Analizinden yararlanılmıştır. Farklı kümeleme teknikleri farklı uzaklık ölçülerine göre tek tek uygulanmış ve en uygun sonucu Karesel Öklid Uzaklığına göre Ward Kümeleme Yönteminin verdiği belirlenmiştir. Sonuç olarak iller 8 alt kümeye ayrılmıştır. İllerin gruplara doğru atanıp atanmadıklarının saptanması aşamasında ise Kümeleme Analizi sonuçlarına Ayırma (Diskriminant) Analizi uygulanmış ve illerin başarılı bir şekilde kümelendiği sonucu elde edilmiştir. Değişkenlerin ise diskriminasyonda önemli rollerinin olduğu saptanmıştır.

Anahtar Kelimeler: Kümeleme Analizi, Ayırma Analizi, İllerin Sağlık Düzeyi, Benzerlik, Farklılık.

ABSTRACT**Master of Science Thesis****A STATISTICAL APPROACH ON DETERMINING
THE HOMOGENOUS PROVINCE GROUPS ON HEALTH IN TÜRKİYE****BAHATTİN HAMARAT****Anadolu University****Statistical Basic Science Branch of Science Institute****Supervisor: Prof. A. Fuat YÜZER****1998, Page 75**

It is necessary to determine health policies to be applied and precautions to be taken for the growth on health to be in balance in Türkiye. For this reason the relation between the provinces which are similar themselves on the distribution of investments and on human and physical sources should be established. Then a balance should be formed between the investments to be made for the provinces according to this relation. Then the provinces which have not been developed on health and the provinces which are developing may be risen to the level of developed provinces. In this study determining the homogenous province groups bearing the same features which will provide analytical and visual information for the processes for forming regional development policies in the field of health has been aimed.

In this study grouping the provinces in Türkiye according to their health, population and some economical and cultural indicators has been made. The 30 variables in indicators have been studied and the clustering analysis has been used for obtaining the province groups. The different clustering techniques have been applied according to different distance measures one by one and it has been determined that Ward Clustering Method gave the most appropriate result according to Squared Euclidean Distance. As a result the provinces have been divided into 8 subsets. In the stage whether the provinces have been appointed into the groups incorrectly or not, discriminant analysis for the results of clustering analysis has been applied and the result of clustering the provinces successfully has been achieved. And then it has been determined that the variables have important parts in discrimination.

Keywords: Clustering Analysis, Discriminat Analysis, The Health Level of Provinces, Similarity, Dissimilarity.

İÇİNDEKİLER (devam)

	<u>Sayfa</u>
ÖZET	i
ABSTRACT	ii
İÇİNDEKİLER	iii
ŞEKİLLER DİZİNİ	vi
ÇİZELGELER DİZİNİ	vii
1- GİRİŞ	1
2- AMAÇ ve KAPSAM	4
3- ARAŞTIRMADA KULLANILAN GÖSTERGELER	5
3.1. Sağlık Göstergeleri	5
3.2. Nüfus Göstergeleri	11
3.3. Diğer Göstergeler	14
4. ANA ÇİZGİLERİYLE ÇOKDEĞİŞKENLİ VERİLERİN ANALİZİNDE KULLANILAN YÖNTEMLER	17
5. ÇOKDEĞİŞKENLİ VERİLERİN GRUPLANDIRILMASINDA KÜMELEME ANALİZİNİN KULLANILMASI	20
5.1. Ana Çizgileriyle Kümeleme Analizi	20
5.2. Kümeleme Analizi Uygulamasında Dikkat Edilmesi Gereken Bazı Önemli Noktalar	21
5.3. Kümeleme Analizinin Kullanım Amaçları	21
5.4. Kümeleme Analizinde İzlenen Adımlar	22
5.5. Kümeleme Analizi Uygulanacak Veri Matrisinin Oluşturulması	24
5.5.1. Değişkenlerin Seçimi	25
5.5.2. Verilerin Dönüşümü	26
5.5.2.1. Verilerin Standartlaştırılması	26

İÇİNDEKİLER (devam)

	<u>Sayfa</u>
5.5.2.2. Verileri $-1 \leq X \leq +1$ Aralığına Dönüşümü	27
5.5.2.3. Verileri $0 \leq X \leq +1$ Aralığına Dönüşümü	27
5.5.2.4. Maksimum Değer Bir Olacak Şekilde Dönüşüm	27
5.5.2.5. Dönüşüm Serisinin Standart Sapması Bir Olacak Şekilde Dönüşümü	28
5.5.2.6. Logaritmik Dönüşümler	28
6. BENZERLİK VE BENZERLİK ÖLÇÜLERİ	29
6.1. Uzaklık Türü Benzerlik Ölçüleri	32
6.1.1. Öklid Uzaklığı	32
6.1.2. Minkowski Uzaklığı	33
6.1.3. Mahalanobis D^2 Uzaklığı	34
6.1.4. Vektör Çarpım Uzaklığı	34
6.2. İlişki Türü Benzerlik Ölçüleri	35
6.2.1. Pearson İlişki Katsayısı	35
6.2.2. Sıra (Rank) Korelasyon Katsayısı	36
6.3. Birliktelik Türü Benzerlik Ölçüleri	36
6.3.1. Basit Eşleşme Katsayısı (The Simple Matching Coefficient)	37
6.3.2. Jaccard Katsayısı	38
6.3.3. Gower Katsayısı	38
7. KÜMELEME YÖNTEMLERİ	40
7.1. Aşamalı Kümeleme Yöntemleri	41
7.1.1. Tek Bağlantı Kümeleme veya En Yakın Komşu Kümeleme Yöntemi	44
7.1.2. Tam Bağlantı Kümeleme veya En Uzak Komşu Kümeleme Yöntemi	47
7.1.3. Ortalama Bağlantı Kümeleme Yöntemi	48

İÇİNDEKİLER (devam)

	<u>Sayfa</u>
7.1.4. Ward Yöntemi veya En Küçük Varyans Kümeleme Yöntemi	49
7.1.5. Merkezi Bağlantı Kümeleme Yöntemi	50
7.1.6. Lance ve Williams Doğrusal Modeli Ve Esnek Kümeleme Yöntemi	50
7.1.7. Ortanca (Medyan) Bağlantı Kümeleme Yöntemleri	51
7.2. Aşamalı Olmayan Kümeleme Yöntemleri	52
7.2.1. Bölünme Yöntemleri	54
7.2.1.1. k-Ortalamalar Yöntemi	54
8. TÜRKİYE'DE SAĞLIK AÇISINDAN HOMOJEN İL GRUPLARININ BELİRLENMESİ	56
9. SONUÇ VE ÖNERİLER	61
10. KAYNAKLAR	63
11. EKLER	67
EKLER	
EK-1 Farklı Kümeleme ve Uzaklık Ölçülerine Göre İllerin Kümelenmesi	67
EK-2 Ayırma Analizi Sonuçları	71
EK-3 Wilks-Lambda Değerleri ve Anlamlılık Seviyeleri	72
EK-4 Ayırma Analizi Sonucunda Değişkenlerin Kümelere Ait Ortalamaları ve Genel Ortalamaları	73
EK-5 SPSS İstatistik Paket Programında Değişkenlerin Gruplandırılması ..	74
EK-6 MINITAB İstatistik Paket Programında Değişkenlerin Gruplandırılması.....	75

ŞEKİLLER DİZİNİ

5.1. Kümeleme analizi karar süreci	23
6.1. S Benzerlik matrisi	30
6.2. İki boyutlu uzayda iki nokta arasındaki uzaklığın gösterimi	30
7.1. Aşamalı kümeleme yöntemlerinde ağaç grafiği	42
7.2. Aşamalı kümeleme yöntemleri	44
7.3. Beş birim arasındaki uzaklıklar için tek bağlantı dendogramı	47
7.4. Ortanca bağlantının analitik gösterimi	52
8.1. Ward Yöntemi Karesel Öklid uzaklığı 0.60 benzerlik düzeyine göre kümeleme analizi sonuçları	58

ÇİZELGELER DİZİNİ

3.1. Araştırmada kullanılan değişkenler	16
6.1. 2x2'lik olasılık tablosu	37
6.2. İkili veri için s_{ijk} ve w_{ijk} 'ya değer atama	39
7.1. Esnek kümeleme yönteminde bazı aşamalı kümeleme yöntemlerine uyan parametrelerin değerleri	51
8.1. Ward Yöntemine göre illerin kümelmesi	57

1. GİRİŞ

İktisadi gelişme gerek zaman gerek mekan açısından bir dizi farklılıklar gösteren dinamik bir olgudur. Bu olgu hem ülkeler arasında hem de ülke ölçeğinde bölgesel, sektörel ve toplumsal gruplar arasında farklılıklar göstermektedir.

Gelişmiş ya da gelişmekte olan bir çok ülkede olduğu gibi ülkemizde de bölgeler ve yörelerin farklı gelişme düzeylerine sahip olduğu gözlenmektedir. Bu farklı gelişme düzeyleri doğal kaynakların dengesizliği ya da yetersizliği, tarihsel, coğrafik ve üretim faktörlerinin bölgesel olarak farklı dağılımı gibi bir çok etkenin sonucunda oluşmaktadır.

Kalkınma sürecinde, ekonomik ve sosyal kalkınmanın ülkelerin her yerinde aynı zamanda gerçekleşmediği, kıt olan kaynakların etkin kullanılması amaçlanarak insan gücü kaynaklarının yeterliliği, pazar büyüklüğü, coğrafi konum, teşebbüs gücü, hammadde yeterliliği gibi sosyal ve ekonomik bir takım faktörlerin etkisiyle ülke içinde bazı yörelerin daha hızlı geliştiği ve sonuçta bölgesel kutuplaşmaların olduğu görülmektedir. Ulusal ekonomiler içinde “gelişmiş”, “gelişmekte olan” ve “geri kalmış” gibi farklı gelişmişlik düzeyinde yörelerin ve bölgelerin ortaya çıkması toplumu oluşturan bireyler arasında istihdam imkanları, gelir ve refah dağılımı açısından çeşitli eşitsizliklere ve dengesizliklere yol açmaktadır [1].

Sanayileşme ile ortaya çıkan bölgeler arası gelişmişlik farkı önce çok hızlı olarak artmakta, ülke belli bir kalkınma düzeyine erişince duraklamakta ve ekonomik kalkınmanın tüm ülke yüzeyine yayılmaya başladığı andan itibaren de, giderek azalma yönünde değişmektedir [2].

Az gelişmiş bölgelerde kişi başına düşen gelirin gelişmiş bölgelere göre düşük olması sonucu, yaşam ve gelişme düzeyi farklılık göstermektedir. Bu bölgelerde aktif nüfusa oranla sermaye ve doğal kaynakların arzı kıt olduğundan işgücünün büyük bir kısmı açık veya gizli işsiz durumundadır [2].

Genel olarak ülkelerin Gayri Safi Milli Hasıla (GSMH)'larının miktarı ve yıllara göre artış hızı, daha ileri gelişmişlik aşamalarına geçişte başarının bir ölçüsü olarak kabul edilmektedir. Öte yandan nüfus az gelişmişliğin önemli bir göstergesi

durumundadır. Az gelişmiş ülkelerdeki hızlı nüfus artışı kişi başına düşen geliri olumsuz yönde etkileme, tasarrufları kısıtlama, üretim yapısını değiştirme, işsizliği artırma gibi sosyo-ekonomik sorunlar yaratmaktadır [3].

Günümüzde ülkelerde olduğu gibi, bölgeler ve yörelerde sosyo-ekonomik bileşenlere bağlı olarak gelişmişlik düzeyleri belirlenerek, gelişmiş, az gelişmiş ve gelişmekte olan bölgeler ve yöreler adı altında sınıflandırılmaktadır. Bu sınıflandırma, kaynakların etkin kullanımının sağlanması ve buna bağlı olarak hızlı, dengeli ve istikrarlı bir kalkınma sürecine kavuşmak için oldukça önemlidir. Ekonomik kalkınmanın sağlanabilmesi ve dengeli bir biçimde yaygınlaştırılabilmesi için uyulması gereken önlem ve politikaların toplumun yapısına uygun olması, o toplumun gelişmişlik düzeyinin bilinmesiyle olanaklıdır.

Gelişmişliğin belirlenmesi ile ülkelerin, bölgelerin ve illerin zaman içindeki ekonomik ve sosyal yönden gösterdikleri değişiklikler izlenebilmekte ve kalkınma açısından önceliğe sahip yöreler belirlenebilmektedir. Tüm bunlar, yapılacak yatırımların yönlendirilmesinde temel oluşturmaktadır [4].

İllerin veya bölgelerin sosyo-ekonomik gelişmişliklerinin belirlenmesi söz konusu olduğunda öncelikle ekonomik, sosyal, demografik, kültürel, eğitim, tarımsal gelişme, sanayileşme ve sağlık gibi çok sayıda değişkene ilişkin veriye ihtiyaç vardır. Doğal kaynakların dengesizliği ya da yetersizliği, coğrafi etkenler gibi bir çok değişik etkinin sonucunda iller veya bölgeler arasında farklılıklar ortaya çıkmaktadır. Bölge veya illerin kendi aralarında gelişmişlik bakımından sınıflandırılması söz konusu olduğunda, bu değişkenlerin ortak bir bileşkesi ele alınarak sınıflandırmaya tabi tutulan bölge veya illerin ortaya koydukları gelişmişlik ile değişkenler tek tek ele alınarak sınıflandırma yapıldığında ortaya çıkan gelişmişlik düzeyi farklı olabilmektedir. Oysaki değişkenlerin tamamı ele alınıp sınıflandırma yapıldığında doğal sınıflar ortaya çıkacaktır.

Ekonomik ve sosyal hedeflerle sağlık arasında sıkı bağlar olduğundan sağlık, kalkınmanın gereçlerinden biri olarak görülmektedir. Modern bir sağlık politikası uygulaması ile ülke düzeyinde sağlığın gelişmesi arasında bir paralellik olacağından,

illerin veya bölgelerin sađlık düzeylerinin sınıflandırılması, yöresel gelişmişlik farklarının belirlenmesi açısından önem taşıyacaktır.

Bir ülkenin sađlık göstergeleri kullanılarak illerin veya bölgelerin gelişmişlik düzeylerine göre, varsa iller ya da bölgeler arası farklılıklar ortaya konabilir. Bölgeler ya da iller arasında dengeli kalkınmanın sađlanabilmesi için, bu göstergelere dayanılarak oluşturulacak bölgesel farklılıkları da içeren etkin planlama kararları alınabilir.

Türkiye’de demografik, eğitim ve sađlık göstergeleri baz alınarak illerin gelişmişlik düzeyleri öncelikle çeşitli kurum (Devlet Planlama Teşkilatı) veya kişiler tarafından gerek planlama, gerekse özel amaçlarla araştırılmış ve incelenmiştir. Bu çalışmada daha çok sađlık alanına ait göstergeler esas alınarak homojen il gruplarının belirlenmesine çalışılacaktır.

2. AMAÇ ve KAPSAM

Yukarıda da vurgulandığı gibi, sağlık hizmetleri arzı ve bu hizmetlerden yararlanma potansiyeli, sosyal gelişme düzeyini yansıtan ögeler arasında bulunmaktadır. Ayrıca, eğitilmiş işgücü olarak da nitelendirilebilecek sağlık personelinin gelişmiş yörelerde çalışma tercihi, illerin sosyo-ekonomik gelişme düzeyiyle doğrudan ilişkilidir [5]. Sağlık düzeyi ile sosyo-ekonomik yapı arasındaki ilişkiden dolayı sağlıkla ilgili göstergeler ülkenin sosyo-ekonomik gelişmişliğinin de bir göstergesi olmaktadır [6].

Sağlık düzeyini gösteren ölçütler; kaynakların etkili kullanılması, hizmette yaygınlık, süreklilik ve kalite sağlanarak halkın sağlık düzeyini yükseltmek amacıyla sağlık sisteminde finansman, yönetim ve organizasyon, insangücü, hizmet sunumu ve iletişim alanlarında yeniden yapılandırmalara gidilmesini kolaylaştıracaktır. Ayrıca, bir bölgeye götürülen sağlık hizmetinin yıllara göre etkinliğinin ölçülmesinde ve bölgeler arası karşılaştırmalar yapılmasında da belirleyici olacaktır.

Bu çalışmada illerin sağlık açısından grup oluşturmalarında önemli olduğu düşünülen, nüfus başına düşen hekim sayısı, bebek ölüm hızı, ana ölüm hızı, düşük hızı vb. gibi değişkenler aracılığı ile gelişmişlik düzeylerinin belirlenmesi amaçlanmış ve birbirine benzer özellik gösterenler bir araya getirilmeye çalışılmıştır. Bağlı olarak çalışmanın amacı, illerin sağlık ve demografik göstergeleri yönünden kümelenecek illerin sağlık alanında gelişmişlik düzeylerinde görülen farklılaşmalara göre, aynı özellikleri taşıyan homojen il gruplarının belirlenmesidir.

Bu çalışmada diğer bir amaç araştırma yapacak kişi ve kuruluşlara Türkiye'yi temsil edebilecek illerin seçiminde kolaylık sağlamaktır. Birbirine benzer iller bir araya getirileceğinden, araştırmacıya kısa ve öz bilgi sunmanın yanında bu kümelerden seçim yapmak zamandan ve emekten tasarruf sağlayacaktır.

Bu çalışmada Türkiye'nin 1996 yılındaki idari yapısı esas alınarak 79 il kapsama alınmıştır. Son idari yapıya göre il olan Osmaniye, Adana ili içinde gösterilmiştir. Çalışmada kullanılan tüm göstergeler 1990-1997 yıllarına ait olup kesitsel (cross sectional) verilerdir.

3. ARAŞTIRMADA KULLANILAN GÖSTERGELER

Kümeleme Analizi hem birimleri hem de değişkenlerin kümelere ayırma ve sınıflandırma olanağı sağlaması, analizin birbiriyle bağımlı olan çok sayıda göstergenin ve değişkenin kullanılmasına olanak verir.

Bu çalışmada, il ve bölgelerin sağlık açısından sınıflandırılmasını ayrıntılı bir biçimde ortaya koyabilmek için değişken sayısının, güvenilir ve tutarlı veri bulunabildiği ölçüde geniş tutulmasına çalışılmıştır.

Bu çalışmada değişkenler, belli bir dönem içindeki gelişmeleri değil, belirli bir yıl içindeki mevcut durumu yansıtmaktadır. Dolayısıyla yapılan analiz, bir kesit çalışması niteliğindedir.

Bilindiği gibi ülkemizde etkin bir veri tabanı sistemi bulunmadığından, sağlık alanında gelişmişlik düzeyi ile ilgili değişkenlere ait verilerin bulunmasındaki güçlükler halen devam etmektedir. Ayrıca, elde edilen verilerin doğruluğu ve güvenilirliği de tartışmalıdır. Türkiye’de bu alanda rasyonel bir veri kayıt sistemi ile sağlık bilişim sisteminin henüz kurulamamış olması, sağlıklı veri toplanamamasının en önemli nedeni olarak gösterilebilir.

Çalışmada, Ülke genelinde 79 ile ait sağlık açısından gelişmişlik seviyelerini yansıttığı varsayılan değişik göstergeler kullanılmıştır. Bunlar nüfus, sağlık ve diğer göstergeler adı altında eğitim, kültür, mali v.b göstergelerdir.

Daha sonra, bu göstergelere ait değişkenler yardımcı veriler aracılığıyla (topluma ya da risk altı topluma) oranlanarak, hız ya da oran haline getirilmiştir. İllerin nüfus büyüklüğünün etkisi altında bulunan değişkenler o ilin nüfusuna veya bu etkiden arındırarak çeşitli büyüklüklere bölünerek oransal hale getirilmiştir. Böylece göstergeler, illerin sağlık açısından gelişmişliğini etkileyecek nüfus büyüklüğü etkisinden arındırılmıştır.

3.1. Sağlık Göstergeleri

Bir ilin sağlık düzeyi, toplumun sosyo-ekonomik koşulları, kültürel düzeyi ve sağlık hizmetlerinin sayısı ve niteliği ile yakın ilişki göstermektedir. Sağlık açısından iller

arasında gruplandırma yapılırken illerin, sağlık hizmetlerine olan talebinin bilinmesi gerekmektedir [2].

Bu çalışmada illerin sağlık açısından düzeylerinin belirlenmesinde, doktor sayısı, eczane sayısı, aşılama oranı gibi değişkenler kullanılmıştır. Bu değişkenlere ait göstergeler aşağıda verildiği gibidir.

10000 Kişiye düşen uzman hekim sayısı (KBUHO): Doktor başına nüfusun azalması, diğer bir değişle nüfusa düşen doktor sayısının artması tedavi edilme olanağını arttırmakta sağlık hizmetlerinin yürütülmesinde aksamaları önlemektedir. 10000 kişiye düşen doktor sayısının yüksek olması o ildeki sağlık hizmetlerinin yaygınlığını göstermektedir. Dolayısıyla kişi başına düşen uzman hekim sayısının yüksek olması gelişmişliğe gösterge olacaktır. Bu oran;

$$\text{KBUHO} = (\text{Uzman Hekim Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmaktadır.

Bu göstergelere ait veriler 1995 yılına ait olup Sağlık Bakanlığı "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

10000Kişiye düşen pratisyen hekim sayısı (KBPHO): Kişi başına düşen pratisyen hekim sayısının yüksek olması gelişmişliğe gösterge olacaktır. Bu oran;

$$\text{KBPHO} = (\text{Pratisyen Hekim Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1995 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

10000 Kişiye düşen diş hekim sayısı (KBDHO): Kişi başına düşen diş hekim sayısının yüksek olması o ildeki sağlık hizmetlerinin yaygınlığını göstereceğinden gelişmişliğe olumlu katkı sağlayacaktır. Bu oran;

$$\text{KBDHO} = (\text{Diş Hekim Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1995 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

10000 Kişiyeye düşen toplam hekim sayısı (KBTHO): Kişiyeye başına düşen toplam hekim sayısının yüksek olması o ildeki sağlık hizmetlerinin yaygınlığını göstereceğinden gelişmişliğe olumlu katkı sağlayacaktır. Bu oran;

$$KBTHO = (\text{Toplam Hekim Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1995 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

10000 Kişiyeye düşen ebe sayısı (KDEO): Kentsel ya da kırsal yerleşim yerleri için sağlık göstergelerinin doğum öncesi (prenatal) ve doğum sonrası (neonatal) ayrı ayrı hesaplanması gerekir. Özellikle anne ve fetusun prenatal ve neonatal bakım dönemi izlenmesi gerekmektedir. Bu da bölgeye götürülen sağlık hizmeti ile mümkündür.

Bir ilde kişiyeye başına düşen ebe sayısının yüksek olması o ilin temel sağlık hizmetlerinin yaygınlığının bir göstergesi olabilir. Bu göstergeye ait veriler;

$$KDEO = (\text{Ebe Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1995 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

10000 Kişiyeye düşen eczacı sayısı (KBEO): Temel sağlık hizmetlerinde ya da yataklı tedavi kurumlarında sağlık hizmetlerinin örgütlenme ve yürütümünde eczacı, sağlık memuru ve hemşireye de ihtiyaç vardır. İllerde bu ihtiyacın sağlandığı oranda sağlık hizmetlerinin yaygınlığından ve gelişmişliğinden söz edilebilir. 10000 kişiyeye düşen eczacı sayısına ait oran;

$$NBEO = (\text{Eczacı Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1995 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

10000 Kişiyeye düşen sağlık memuru sayısı (KBSMO): Kişiyeye başına düşen sağlık memuru sayısının yüksek olması o ildeki temel sağlık hizmetlerinin yaygınlığını göstermektedir. Bu gösterge;

$$KBSMO = (\text{Sağlık Memuru Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1995 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

10000 Kişiye düşen hemşire sayısı (KBHO): Kişi başına düşen hemşire sayısının yüksek olması gelişmişliğe bir başka gösterge olacaktır. Bu gösterge;

$$KBHO = (\text{Hemşire Sayısı} / \text{İl Nüfusu}) * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1995 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

Yatak işgal oranı (YİO): Ülkemizde sağlık hizmetlerini sunan çeşitli kuruluşlar olup bunlardan hastaneler yataklı tedavi kurumlarının temelini oluşturmaktadır.

Bu oran, illerdeki yerleşik sağlık kuruluşlarının kapasite kullanımları hakkında açıklayıcı olabileceği düşünülerek çalışmaya alınmıştır. Tedavi için gerekli olan süre itibarıyla hastalıkların illere göre dağılımı istatistiki olarak "normal" varsayıldığında oranın değişmesi kapasitenin büyüklüğüne (yatak sayısına) bağlı olacaktır. Oran bütünüyle küçüldüğünde gelişmeye etkisi olumsuz, büyüdüğünde ise olumlu olacaktır. Yatak İşgal Oranı;

$$YİO = \frac{\text{Bir yılda hastanede yatan hastaların yatmış oldukları toplam gün sayısı}}{\text{Fiili yatak sayısı} * 365} * 100$$

şeklinde hesaplanmıştır. Bu oran için veriler Sağlık Bakanlığı "Yataklı Tedavi Kurumları İstatistik Yıllığı 1996" yayınından elde edilmiştir.

10000 Nüfusa düşen yatak sayısı (YBNO): Sağlık açısından gelişme düzeyini yansıtan önemli göstergelerden biri de yatak sayısıdır. İllerdeki sağlık merkezlerinde (Sağlık Ocağı veya Hastane gibi) mevcut olan yatak sayısı ne kadar fazla ise, o ildeki yerel halka o kadar çok sağlık hizmetinin götürüldüğünün bir göstergesidir. Bu şekilde il, sağlık düzeyi açısından daha iyi konuma gelebilir. Bu gösterge;

$$YBNO = \frac{\text{Kadro Yatak Sayısı}}{\text{Nüfus}} * 10000$$

şeklinde hesaplanmıştır. Bu göstergenin hesaplanmasında kullanılan veriler 1996 yılına ait olup Sağlık Bakanlığı'nın "Sağlık İstatistikleri 1996" yayınından elde edilmiştir.

Aşılama oranı: Koruyucu tedavi ve özellikle bulaşıcı hastalıkların kontrolü açısından aşı etkin bir araç durumundadır. İllerdeki bulaşıcı hastalıklara karşı aşılanan kişi sayısının hedef nüfusa oranının yüksekliği bu hastalıkların azaltılmasında anlamlı ilerleme kaydedildiğinin bir göstergesi olacaktır. Böylece aşı ile önlenbilir hastalıkların sıklığı azalacak ve illerde görülen olgu sayısı düşecektir.

Aşı doz ve oranları, Devlet İstatistik Enstitüsü tarafından projekte edilmiş 0 yaş grubu ve 1-4 yaş grubu hedef nüfuslar dikkate alınarak belirlenmektedir [7].

0 Yaş aşılama oranı(0YAO): 0 yaş aşılama da kullanılan aşılardan; BCG (verem aşısı), DBT1 (difteri, boğmaca, tetanoz 1.aşı), DBT2 (difteri, boğmaca, tetanoz 2.aşı), DBT3 (difteri, boğmaca, tetanoz 3.aşı), POL1 (çocuk felci 1.aşı), POL2 (çocuk felci 2.aşı), POL3 (çocuk felci 3.aşı) ve kızamık aşılardan oluşmaktadır. Bu oran;

$$0YAO = \frac{\left(\frac{\text{BCG Doz}}{\text{Hedef Nüfus}} + \frac{\text{DBT1 Doz}}{\text{Hedef Nüfus}} + \dots + \frac{\text{Kızamık Doz}}{\text{Hedef Nüfus}} \right)}{\text{Aşı Sayısı}} * 100$$

şeklinde hesaplanmıştır. Bu orana ait veriler 1996 yılına ait olup, Sağlık Bakanlığı "Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1996" yayınlanmamış verilerden elde edilmiştir.

1-4 Yaş aşılama oranı (1-4YAO): 1-4 yaş grubu aşılama da kullanılan aşılardan; DBTR (DPT rapel) ve POLR (polio rapel) den oluşmaktadır. Aşılama oranı;

$$1-4 YAO = \frac{\left(\frac{\text{DBTR Doz}}{\text{Hedef Nüfus}} + \frac{\text{POLR Doz}}{\text{Hedef Nüfus}} \right)}{\text{Aşı Sayısı}} * 100$$

şeklinde hesaplanmıştır. Bu orana ait veriler 1996 yılına ait olup, Sağlık Bakanlığı "Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1996" yayınlanmamış verilerden elde edilmiştir.

Ölü doğum hızı (ÖDH): Ölü doğumlar ve kendiliğinden düşüklükler, doğum öncesi bakım ve ana sağlığının, isteyerek ise aile planlaması hizmetlerinin değerlendirilmesinde önemli göstergelerdir. Ölü Doğum Hızı, bir yıl içinde ölü doğan bebek sayısının, o yıldaki toplam doğum sayısına oranıdır. Bu oranın düşüklüğü sağlık açısından iyi düzeyde olduğunu, tersi durumda ise sağlık açısından iyi düzeyde olmadığını gösterir. Bu hız;

$$\text{ÖDH} = \frac{\text{1 yılda ölü doğum sayısı}}{\text{1 yılda toplam doğum sayısı}} * 1000$$

şeklinde olup ölü doğum hızları, Sağlık Bakanlığı “Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1994-1995” yayınından elde edilmiştir.

Düşük hızı (DH): Ana-çocuk sağlığı ve aile planlaması hizmetlerinin tam ve uygun biçimde verilip verilmediğini veya aile planlamasının uygulanıp uygulanmadığını ortaya koyan bir göstergedir. İllerin sağlık düzeyini en iyi yansıtan hız 15-44 yaş grubundaki evli kadın nüfusuna göre hesaplanan hızdır. Bu hız istenmeyen gebeliklerin sıklığını objektif olarak gösteren bir ölçüttür [8]. Bu çalışmada kullanılan düşük hızı;

$$\text{DH} = \frac{\text{Düşük Sayısı}}{\text{Canlı Doğum Sayısı}} * 1000$$

eşitliği ile hesaplanmıştır. Bu hız için veriler 1995 yılına ait olup, Sağlık Bakanlığı “Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1994-1995” yayınından alınmıştır.

Bebek ölüm hızı (BÖH): Bir yıl içinde bir yaşına girmeden ölen bebeklerin o yılda canlı doğan bebek sayısına oranı ile bulunur. Bebek ölüm hızı bir ülkenin ya da bölgenin sağlık düzeyini gösteren önemli ölçütlerden biridir. Çünkü Bebek Ölüm Hızı bir toplumdaki ana-çocuk sağlığının düzeyini olduğu kadar, ülkenin çevre sağlığı koşullarını, toplumun sosyo-ekonomik düzeyini ve gelişme düzeyini çok iyi biçimde yansıtır [9]. Bu hız;

$$BÖH = \frac{\text{Bir yılda 1 yaşından küçük ölen sayısı}}{\text{Bir yıldaki canlı doğum sayısı}} * 1000$$

olarak hesaplanmıştır. Bu hız için veriler 1995 yılına ait olup, Sağlık Bakanlığı “Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1994-1995” yayınından alınmıştır.

Ana ölüm hızı (AÖH); Ana ölümü, gebeliğe ya da doğuma bağlı bir nedenden kadının ölmesidir. Gebelikte veya doğumda yeterli ölçüde sağlık hizmetlerinden yararlanan kadın, gebelik veya doğuma bağlı bir nedenden dolayı ölüm olasılığı azalacak ve buna bağlı olarak ana ölüm hızı düşecektir.

AÖH ile ilgili veriler Sağlık Bakanlığı “Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1994-1995” yayınından,

$$AÖH = \frac{\text{Bir yıl içinde ana ölüm sayısı}}{\text{Bir yılda canlı doğum sayısı}} * 100000$$

şeklinde hesaplanmıştır.

Yeterli içme suyu olan kırsal nüfus oranı (YİSNO): İçme suyu olan kırsal nüfus içinde ne kadarının yeterli içme suyuna sahip olduğunu gösteren bu gösterge, içme suyunun sağlıklı yakın bağlantılı olmasından dolayı önem kazanmaktadır. Bu oran;

$$YİSNO = \frac{\text{Yeterli içme suyu olan kırsal nüfus oranı}}{\text{Toplam kırsal nüfus}} * 100$$

şeklinde hesaplanmaktadır. YİSNO için veriler 1996 yılına ait olup “İller ve Bölgeler İtibariyle Çeşitli Göstergeler” DPT yayınından alınmıştır.

3.2. Nüfus Göstergeleri

Bölgesel dengesizliğin en belirgin göstergesi, bir ülkenin nüfus ve başlıca ekonomik faaliyetlerinin ilgili ülkenin belirli birkaç bölgesinde yoğunlaşması ve bunun sonucu olarak da öteki bölgelerin nüfus bakımından daha az kalabalık ve ekonomik açıdan daha az gelişmiş olmasıdır [2].

Sağlık hizmetlerinin örgütlenme, yürütme ve yeniden düzenleme aşamalarının uygun bir biçimde planlanmasında, nüfus yapısı ve özelliklerinin bilinmesi gerekir. Ayrıca il ve bölgelerin sağlık açısından gelişmişliklerinde etkisi olacak nüfus hareketleri ve nüfusun gelişiminin de belirlenmesi gerekir.

Bu çalışmada nüfus hareketleri, bir bölgede olagelen doğum, ölüm, göç gibi olaylarla, toplam nüfusun azalması ya da çoğalması ile ilgili nüfusun gelişmesine ait olan veriler birer gösterge olarak alınmıştır. Bu göstergeler aşağıda verildiği gibidir.

Yıllık nüfus artış hızı (1990-1995): Bir bölgenin nüfusunun bir yıl içinde doğumlar ve ölümlere bağlı olarak artış ya da azalmasını ifade eden bir kavramdır. Bu hız;

$$\text{Nüfus Artış Hızı} = \left(\frac{\text{Bir yıl içinde canlı doğum sayısı}}{\text{Yıl ortası nüfus}} - \frac{\text{Bir yıl içindeki ölüm sayısı}}{\text{Yıl ortası nüfus}} \right) * 1000$$

biçiminde hesaplanır [6].

Ölümler ve doğumlara ek olarak, içe ve dışa göçlerin de dikkate alındığı nüfusun azalması ya da çoğalmasını gösteren bir ölçüde Nüfusun Yıllık Toplam Artış Hızıdır. Bu hız;

$$\text{Toplam Nüfus Artış Hızı} = \frac{(\text{Doğum sayısı} + \text{içe göçler}) - (\text{Ölüm sayısı} + \text{dışa göçler})}{\text{Yıl ortası nüfus}}$$

ile bulunur [8].

Bir ilin nüfus artış hızının yüksek oluşu, o ilin çekim merkezi olma özelliğini, şehirleşmeyi ve diğer illerden göç almayı da yansıtmaktadır. Ayrıca ildeki sağlık düzeyinin önemini de ortaya çıkarabilmektedir. Böylece sağlık düzeyi yönünden o ilin hangi seviyede olduğuna dair bilgi verebilir. Bu çalışmada nüfus artış hızı verileri “İller itibariyle çeşitli göstergeler” DPT 1997 yayınından alınmıştır.

Nüfus yoğunluğu (NY): Birim alana düşen nüfusu gösteren bir göstergedir. Nüfus yoğunluğu göstergesi;

$$NY = \frac{\text{Nüfus}}{\text{Yüzölçümü}}$$

biçiminde hesaplanır.

İllerin kurulduğu coğrafi alanların özelliği, sanayi ve ticaretin gelişmiş olması gibi nedenler illerde nüfus yoğunluğunu artırmaktadır. Bu gibi illerde nüfus yoğunluğu artacağından, verilen sağlık hizmeti artacak ve sağlık açısından görülen gelişmenin önemi büyük olacaktır.

Nüfus yoğunluğunun bulunmasında 1997 yılı genel nüfus sayımı sonuçlarından yararlanılmıştır.

x Yaşa Göre Orantılı Ölüm Hızı (xYOÖH): Herhangi bir yaş grubunda görülen ölümlerin toplam ölümler içindeki yüzdesini, başka bir ifadeyle, toplam ölümler içindeki payını gösteren bir ölçüttür. Yaş grubuna göre Orantılı Ölüm Hızı da sağlık düzeyini gösteren ölçütlerden biridir. Geri kalmış veya az gelişmiş il ya da bölgelerde ölümlerin büyük bir oranı 0-4 yaşta yoğunlaşırken, gelişmiş il veya bölgelerde 65+ yaşta daha fazla yoğunlaşma vardır [10].

Orantılı Ölüm Hızı; 0-4, 5-44, 45-64 ve 65+ yaşlar için hesaplanmıştır.

$$X \text{ YOÖH} = \frac{X \text{ yaş grubunda bir yıl içinde ölenlerin sayısı}}{\text{Bir yıl içinde ölümlerin sayısı}} * 1000$$

şeklinde hesaplanmıştır.

Bu hızlara ait veriler “Sağlık Bakanlığı Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1994-1995” yayınından alınmıştır.

Kaba ölüm hızı (KÖH): Kaba ölüm hızındaki düşme o ilin belli sağlık koşullarına, eğitim seviyesine ve bunları sağlayacak ekonomik olanaklara ulaştığını belirlemiş olacaktır. Kaba Ölüm Hızı, bir bölgede bir yıl içinde görülen toplam ölüm sayısının bölge yıl ortası nüfusuna bölünmesiyle bulunur. Bu hız için veriler;

$$KÖH = \frac{\text{Bir yılda olan ölemlerin sayısı}}{\text{Yıl ortası nüfus}} * 1000$$

şeklinde hesaplanmaktadır. Bu hız için veriler 1995 yılına ait olup, Sağlık Bakanlığı “Temel Sağlık Hizmetleri Genel Müdürlüğü Çalışma Yıllığı 1994-1995” yayınından alınmıştır.

3.3. Diğer göstergeler

İllerin sağlık açısından sınıflandırılmasına etki edebilecek altyapı, eğitim, ekonomik v.b göstergeler diğer göstergeler adı altında toplanmıştır. Bu göstergelere ait veriler aşağıda verildiği gibidir.

Kişi başına gayri safi yurt içi hasıla (KBYİH): Bu değer hem ülkeler hemde ülke içerisindeki iller arasındaki gelişmişlik düzeylerinin karşılaştırma işleminde baş vurulan önemli değişkenler arasında yer almaktadır. Kişi başına gayri safi yurt içi hasıla değerinin yüksekliği ülke ya da ilin görece olarak gelişmiş olduğunu işaret etmektedir. Bir ilde bu değer yüksek olması o ilin daha gelişmiş olabileceğinin bir göstergesi olarak kabul edilebilir. Kişi Başına Gayri Safi Yurt İçi Hasıla değerleri için veriler 1996 yılına ait olup “İller ve Bölgeler İtibariyle Çeşitli Göstergeler” DPT yayınından alınmıştır.

Öğretmen başına düşen öğrenci sayısı (ÖBÖS): Sağlık düzeyi ile toplumu oluşturan fertlerin genel eğitim seviyesi arasında güçlü bir ilişki bulunmaktadır. Her türlü eğitim faaliyetinin gereği olarak okuma-yazma bilmenin temeli gerekli sayıda ve nitelikte öğretim elemanı istihdamından geçmektedir.

Bu oranların bir ilde diğer ile göre yüksek olması birinci ilin daha gelişmiş olabileceğinin bir göstergesi olarak kabul edilebilir. Bu göstergeler Milli Eğitim Bakanlığı 1997-1998 öğretim yılına ait verilerden hem ilk öğretim için hem de orta öğretim için,

$$ÖBÖS = \frac{\text{Öğrenci sayısı}}{\text{Öğretmen sayısı}}$$

şeklinde hesaplanmıştır.

Asfalt yol oranı (AYO): Bir ilin alt yapı göstergeleri, nüfusun çağdaş sağlık hizmetlerine ulaşabilme ve sağlık hizmetlerinden yararlanabilmelerini göstermesi nedeniyle çalışmaya dahil edilmiştir. Bu oran için veriler 1996 yılına ait olup “İller ve Bölgeler İtibariyle Çeşitli Göstergeler” DPT yayınından alınmıştır. Asfalt yol oranı;

$$AYO = \frac{\text{Asfalt yol uzunluğu}}{\text{Toplam yol uzunluğu}} * 100$$

şeklinde hesaplanmıştır.

Asfalt ve stabilize köy yolu oranı (ASKYO): Kırsal nüfusun sağlık hizmetlerinden yararlanabilme olanağını göstermesi nedeniyle çalışmaya dahil edilmiştir. Bu oran için veriler Köy Hizmetleri Genel Müdürlüğü’nden elde edilmiş ve

$$ASKYO = \frac{\text{Asfalt ve stabilize köy yolu uzunluğu}}{\text{Toplam yol uzunluğu}} * 100$$

şeklinde hesaplanmıştır. Veriler 1996 yılının göstergeleridir.

Kişi başına toplam tarımsal üretim değeri (KTTÜD): Kırsal kesimde kişi başına tarımsal üretim değerlerini belirleyen bir göstergedir. Bu oranın büyümesi halinde tarımsal gelişmenin olumlu yönde olduğu varsayılmaktadır. Kişi başına toplam tarımsal üretim değeri;

$$KTTÜD = \frac{\text{Toplam tarımsal üretim değeri}}{\text{Nüfus}}$$

şeklinde elde edilmiştir. Bu oran için veriler 1996 yılına ait olup “İller ve Bölgeler İtibariyle Çeşitli Göstergeler” DPT yayınından alınmıştır.

Araştırmamızın daha sonraki aşamalarında, yukarıda açıklanan hız ve oransal göstergeler değişken olarak anılacaktır. İl gruplarının oluşturulmasında üç ana göstergeden oluşan 30 değişken gösterimleri ve açıklamalarıyla Çizelge 3.1’de verilmiştir.

Çizelge 3.1. Araştırmada kullanılan değişkenler

Değişken Adı	Açıklama
X ₁	10000 kişiye düşen uzman hekim sayısı
X ₂	10000 kişiye düşen pratisyen hekim sayısı
X ₃	10000 kişiye düşen diş hekimi sayısı
X ₄	10000 kişiye düşen toplam hekim sayısı
X ₅	10000 kişiye düşen ebe sayısı
X ₆	10000 kişiye düşen eczacı sayısı
X ₇	10000 kişiye düşen sağlık memuru sayısı
X ₈	10000 kişiye düşen hemşire sayısı
X ₉	Yatak işgal oranı
X ₁₀	10000 kişiye düşen yatak sayısı
X ₁₁	0 (Sıfır) yaş aşılama oranı
X ₁₂	1-4 yaş aşılama oranı
X ₁₃	Ölü doğum hızı
X ₁₄	Düşük hızı
X ₁₅	Bebek ölüm hızı
X ₁₆	Ana ölüm hızı
X ₁₇	Yeterli içme suyu olan kırsal nüfus oranı
X ₁₈	Yıllık nüfus artış hızı (1990-1995)
X ₁₉	Nüfus yoğunluğu
X ₂₀	0-4 yaş orantılı ölüm hızı
X ₂₁	5-44 yaş orantılı ölüm hızı
X ₂₂	45-64 yaş orantılı ölüm hızı
X ₂₃	65+ yaş orantılı ölüm hızı
X ₂₄	Kaba ölüm hızı
X ₂₅	Kişi başına gayri safi yurtiçi hasıla değeri
X ₂₆	Öğretmen başına düşen öğrenci sayısı (1997-1998 İlköğretimde)
X ₂₇	Öğretmen başına düşen öğrenci sayısı (1997-1998 ortaöğretimde)
X ₂₈	Asfalt yol oranı
X ₂₉	Asfalt ve stabilize köy yolu oranı
X ₃₀	Kişi başına düşen toplam tarımsal üretim değeri

4. ANA ÇİZGİLERİYLE ÇOKDEĞİŞKENLİ VERİLERİN ANALİZİNDE KULLANILAN YÖNTEMLER

Bilimsel çalışmalarda ele alınan olaylar genellikle birden çok etkenin etkisi altındadır. Bununla birlikte, gözleme konu olan nesnelerin özellikleri de birbiriyle ilişki içinde bulunmaktadır. Dolayısıyla olayların çözümlenmesinde, inceleme konusu bir olayda tek bir faktörün etkisinin incelenmesi yerine birden çok faktörü eş zamanlı olarak göz önünde bulundurup sonuçlar çıkartma daha güvenilir olacaktır. Çünkü çok değişkenli istatistikler, genellikle kendi doğal çevrelerinden bir bütün olarak gözlenirler ve bütünlüğü sağlayan faktörlerin bağımlı yapısını açıklamaya çalışırlar [11]. Böylelikle incelenen olay ve çevresindeki çok sayıda içsel ve dışsal faktörleri dikkate alarak, problemi doğasındaki yapısında incelemek ve çözümlere ulaşmak mümkün olacaktır [12].

Çokdeğişkenli istatistiksel yöntemler değişik amaçlar için kullanılmaktadır. Bu amaçlar aşağıda kısaca açıklanmaktadır [12].

1-Veri indirgeme: p sayıda değişken içeren veri setinin değişimini açıklayan ve aralarında ilişki bulunmayan daha az sayıda değişkenle ($k < p$) açıklama yapmayı sağlamak.

2-Kümeleme ve sınıflama: Anakütle özellikleri bilinmeyen yapılar hakkında örnek kümeler (grup, sınıf) belirleme çalışmalarına yardımcı olmak. Daha önceden belirlenmiş gruplara yeni birimlerin atanmasını sağlamak.

3-Ölçekleme: p sayıda değişken içeren p boyutlu ölçümlerden daha az sayıda değişken kullanarak birimlerin gösterilmesini ve tanımlanmasını sağlamak. Birimlerin birbiri ile $k < p$ boyutlu ölçekte benzerlik ve farklılıklarını incelemek.

4-Çokdeğişkenli hipotezlerin test edilmesi: k toplumun ortalamalar vektörünün farklılığı üzerine kurulacak hipotezleri test etmek.

Bu yöntemlerden en yaygın olarak kullanılanları, Çok Değişkenli Varyans Analizi (Multivariate Analysis of Variance, MANOVA), Ayırma Analizi (Discriminant Analysis), Temel Bileşenler Analizi (Principal Component Analysis, PCA), Faktör Analizi (Factor Analysis), Çok Boyutlu Ölçekleme Analizi (Multidimensional Scaling),

Uyum Analizi (Corresponding Analysis) ve Kümeleme Analizi (Cluster Analysis) olarak ifade edilebilir [13]. Adı geçen çokdeğişkenli analizlerden bazıları aşağıda ana çizgileri ile ele alınmıştır.

Çokdeğişkenli Varyans Analizi (MANOVA) ikiden daha çok bağımsız ve bağımlı çokdeğişkenli normal dağılımlara dayalı hipotezleri test etmek üzere geliştirilmiş bir yöntemdir [12].

Ayrırma Analizi X veri matrisindeki değişkenlerin iki veya daha fazla gerçek gruplara ayrılmasını belirlemek amacıyla uygulanan bir yöntemdir. Bu yöntem, birimlerin p tane özelliğini ele alarak bu birimlerin doğal ortamdaki gerçek gruplarına, sınıflarına en iyi düzeyde atanmalarını sağlayacak fonksiyonlar bulmayı amaçlar. Ayrırma Analizi, bu fonksiyonlara göre yeni bir gözlemin eldeki gruplardan herhangi birine doğru olarak ve hata payı minimum olacak şekilde atamayı, sınıflamayı sağlayan bir yöntemdir. Bu açıklamalara göre Ayrırma Analizi;

- Grupları birbirlerinden ayırmayı sağlayan fonksiyonları bulmak,
- Hesaplanan fonksiyonlar aracılığı ile yeni gözlenen bir birimi sınıflama hatası minimum olacak şekilde elde edilen gruplardan herhangi birine atamaktır [12].

Çok Boyutlu Ölçekleme; n birim arasındaki uzaklık değerlerini kullanarak bu birimlerin çok boyutlu uzaydaki konumlarını, ilişki yapısını, mümkün olduğunca az boyutla gerçeğe en yakın bir biçimde ortaya koymayı amaçlayan çokdeğişkenli istatistiksel bir yöntemdir [11, 14].

Temel Bileşenler Analizi, p boyutlu (değişkenli) uzayı çok iyi tanımlayan p tane yeni dik değişken (temel bileşen ya da özvektör) elde etmek istendiğinde kullanılan bir yöntemdir [15]. Temel Bileşenler Analizi, veriler arasındaki ilişkilere dayanarak, verilerin daha anlamlı ve özet bir biçimde sunulmasını, ayrıca boyut (temel faktöre) indirgemeye yarayan bir analizdir [16]. Analizin bu özelliğinden dolayı; başka analizler için veri hazırlama tekniği olarak da kullanılmaktadır.

Temel Bileşenler Analizinin (TBA) veri indirgemesi, tahmin ve veri kümesini bazı yöntemlerin analiz edebileceği forma sokmak gibi üç temel amacı bulunmaktadır.

TBA daha önceden ortaya çıkarılmamış ilişkileri ortaya çıkarma ve sıradan sonuçlar diye nitelenemeyecek tahminler yapmaya izin veren bir yöntemdir [12].

Temel bileşenler bizzat kendileri bir sonuç olmaktan ziyade sonuç almayı sağlayıcı özelliğe sahiptirler. Çünkü temel bileşenler daha geniş incelemeler için bir ara adım özelliği taşırlar. Örneğin Çoklu Regresyon Analizi uygulanması istenilen fakat regresyon varsayımlarından birisi olan değişkenler arasında çoklu bağımlılık olmaması koşulunun yerine gelmemesi nedeniyle uygulanamadığı durumlarda çoklu regresyona bir giriş değeri olabilir ya da Kümeleme Analizine veri olabilirler [12, 17, 18].

Faktör Analizi p değişkenli bir olayda birbirini ile ilişkili değişkenleri bir araya getirerek, daha az sayıda ($m < p$) yeni (ortak) ilişkisiz değişken bulmayı amaçlar. Faktör Analizinin değişken sayısını azaltma ve değişkenler arasındaki ilişkilerden yararlanarak bazı yeni yapılar (common factor) ortaya çıkarmaktır. Bu son amaç değişkenleri sınıflayarak tek bir faktör adı altında birleştirmek ve ortak yeni açıklayıcı faktör yapıları oluşturmayı sağlamaya yöneliktir [12].

Kümeleme Analizinin temel amacı, karmaşık yapıdaki veri kümesinde yer alan birimlerin doğal sınıflandırılmasını bulma ve benzer birimlerin aynı kümede toplanmasını sağlamaktır. Kümeleme Analizi, birimlerin, değişkenlerin ya da birimlerin ve değişkenlerin bir arada kümelenmesini, sınıflandırılmasını içine alan yöntemlerin genel adıdır [19]. Böylece kümeleme analizi, değişkenler arasındaki benzerlikleri ve farklılıkları temel alarak en uygun kümelerin oluşmasını sağlayan bir gruplama tekniğidir [20]. İzleyen bölümde kümeleme analizi ayrıntıları ile ele alınacaktır.

5. ÇOKDEĞİŞKENLİ VERİLERİN GRUPLANDIRILMASINDA KÜMELEME ANALİZİNİN KULLANILMASI

5.1. Ana Çizgileriyle Kümeleme Analizi

Kümeleme Analizini açıklamadan önce analizin üzerine şekillendiği, matematiğin en temel kavramlarından birisi olan iyi tanımlanmış nesnelere oluşan küme kavramının açıklanması gerekir. Küme kavramı ortak özelliklere sahip birlikte ele alınan nesnelere topluluğu olarak açıklanabilir [21]. Bir kümenin oluşabilmesi için onu oluşturacak nesnelere ortak bir özellik taşımaları gerekir [22]. Kümeleme Analizinde, kümelerin oluşturulmasında göz önünde bulundurulmuş ortak özellik; benzerlik, yakınlık, uzaklık gibi kavramların ayrımlı biçimde tanımlanmasıyla küme; “istatistiksel anakütlenin yakın öğelerinden oluşan grup” ya da “test uzayında birbirine yakın noktaların bir araya getirilmesi” olarak açıklanabilir [23].

Bilimin temel işlevlerinden birisi de karmaşık durumları anlaşılır duruma getirmek ve gözlemlere ya da birimlerin doğal sınıflandırmasını bulmaktır. Kümeleme analizinin temel amacı, karmaşık yapılara sahip ya da hakkında kesin bilgi sahibi olunamayan veri kümesinden birimlerin doğal sınıflandırmasını belirlemek ve benzer birimleri aynı kümeye toplamaktır [20].

Kümeleme analizi, değişkenleri ya da birimleri gruplara yerleştiren çokdeğişkenli bir tekniktir. Böylece her gruptaki değişkenler ve birimler birbirlerine benzer ve diğer grupların üyelerine benzemezdir. Ortaya çıkan kümeler yüksek düzeyde küme içi homojenlik ve yüksek düzeyde kümeler arası heterojenlik göstermelidir [24, 25]. Böylelikle sınıflandırma başarılı ise geometrik olarak gösterildiğinde aynı kümelerde birimler birbirine yakın olurlar, farklı kümelerdeki birimlere ise uzak olurlar.

Pek çok farklı durumda kümeleme analizi veri analizi için yararlı bir araç olması yanında verilerin doğasıyla ilgili hipotezler geliştirmek ya da daha önce kurulmuş olan hipotezlerin incelenmesinde de yararlıdır [24]. Kümeleme analizinin en önemli özelliklerinden birisi, verilerin kategorik yapısı ile ilgili hipotezleri meydana getirmektir. Böylece kümeleme analizi verilerdeki yapıyı ve ilişkileri ortaya çıkarmak için kullanılabilir. Kümeleme analizinin sonuçları, diğer bilim dallarında sınıflandırma çalışmalarında doğrudan katkıda bulunabilmektedir. Kümeleme analizi yoluyla çok büyük bir veri kümesini nispeten özlü kısa tanımlara indirgemek mümkün olmaktadır. Verilerin kategorik yapıları ile ilgili çok az bilgi varsa ya da bu konuda hiç bir şey bilinmiyorsa, bu

durumda elde edilen verilerin sınıflandırılmasında yapılacak ilk iş gözlemlere uyan bir kategorik yapı oluşturmaktır. Sorun, verilerin temelinde yatan doğal sınıflamayı belirlemektir [26].

Gerçekte kümeleme analizi ile doğadaki bilinmeyen yapıların incelenmesinin yanısıra, karşılaşılan özel problem doğrultusunda doğal sınıfları belirlenmiş durumların da tekrardan ele alınması gerekebilir. Ayrıca geleneksel sınıflaması az sayıda değişkene göre belirlenmiş durumların, değişken sayısı artırılarak ele alınmaları ve önceki sınıflamanın değişip değişmediğini incelemek için de Kümeleme Analizinden yararlanılmaktadır [19].

Daha önce değinildiği gibi “Kümeleme Analizi değişkenlerin, birimlerin ya da birimlerin ve değişkenlerin bir arada kümelenmesini, sınıflandırılmasını içine alan yöntemlerin genel adıdır” [19].

5.2. Kümeleme Analizi Uygulamasında Dikkat Edilmesi Gereken Bazı Önemli Noktalar.

Kümeleme Analizi, farklı uygulamalar içeren yöntemlerin bir araya toplandığı bir istatistiksel yöntemdir. Kümeleme Analizinde birimlerin kümelenmesinde farklı hesaplama biçimleri olduğundan, çeşitli kümeleme yöntemleri ileri sürülmüştür [26]. Kümeleme analizinde kullanılan pek çok uzaklık ölçüsü ve bu ölçüler üzerinde kurulmuş pek çok yöntem olması, kuramda konuyla ilgili aşağıdaki genellemeleri zorunlu kılmıştır [27].

- Çoğu kümeleme analizi yöntemleri, kolay süreçler içerirler ve çoğu durumda istatistiki bir temele dayandırılmayabilirler.

- Kümeleme analizi yöntemleri birçok farklı disiplinden meydana gelmiştir. Her disiplin kendi eğilimine sahiptir ve verileri farklıdır. Analiz aşamasında bu eğilim dikkate alınmalıdır. Örneğin psikolojide önemli olan bir ayrıntı, biyolojide önemli olmayabilir. Kullanıcılar kendileri için en akla yatkın verileri en iyi temsil eden küme yöntemini seçmelidir.

- Değişik kümeleme yöntemleri, aynı veri kümesine uygulandığında farklı sonuçlar üretebilirler.

- Kümeleme analizinin temel amacı karşılaşılan özel soruna uygun bir çözüm aramaktır.

5.3. Kümeleme Analizinin Kullanım Amaçları

Kümeleme Analizi; birimleri, p değişkene göre hesaplanan ve benzerlik ölçüsü

olarak kullanılan bazı ölçüler kullanarak (similarity or dissimilarity measures) homojen gruplara bölmek amacıyla kullanılır. Kümeleme analizinin yararları; gruplara dayalı tahmin, hipotez testi, veri araştırması, hipotez genelleme olarak belirlense de [28], analizin genel hedefi bir nesnel kümesinin homojen alt gruplar veya homojen alt grupların bir hiyerarşik düzenlemesi içinde kısımlara veya alt bölümlere bölmektir. Daha özel olarak ifade edilir ise bu yöntem aşağıdaki amaçlar için düzenlenir [12].

- n sayıda birimi, nesneyi, oluşumu, p değişkene göre saptanan özelliklerine göre olabildiğince kendi içinde türdeş (homojen) ve kendi aralarında farklı (heterojen) alt gruplara ayırmak.

- p sayıda değişkeni, n sayıda birimde saptanan değerlere göre ortak özellikleri açıkladığı varsayılan alt kümelerle ayırmak ve ortak faktör yapıları ortaya koymak.

- Hem birimleri hem de değişkenleri birlikte ele alarak ortak n birimi p değişkene göre ortak özellikli alt kümelerle ayırmak.

- Birimleri, p değişken için saptanan değerlere göre, izledikleri biyolojik ve tipolojik sınıflamayı ortaya koymak (taksonomik sınıflandırma yapmak).

5.4. Kümeleme Analizinde İzlenen Adımlar

Uygulamada kümeleme analizi ile grupların oluşturulmasında veya yeniden elde edilmesinde izlenen adımlara aşağıda kısaca değinilmiştir [12].

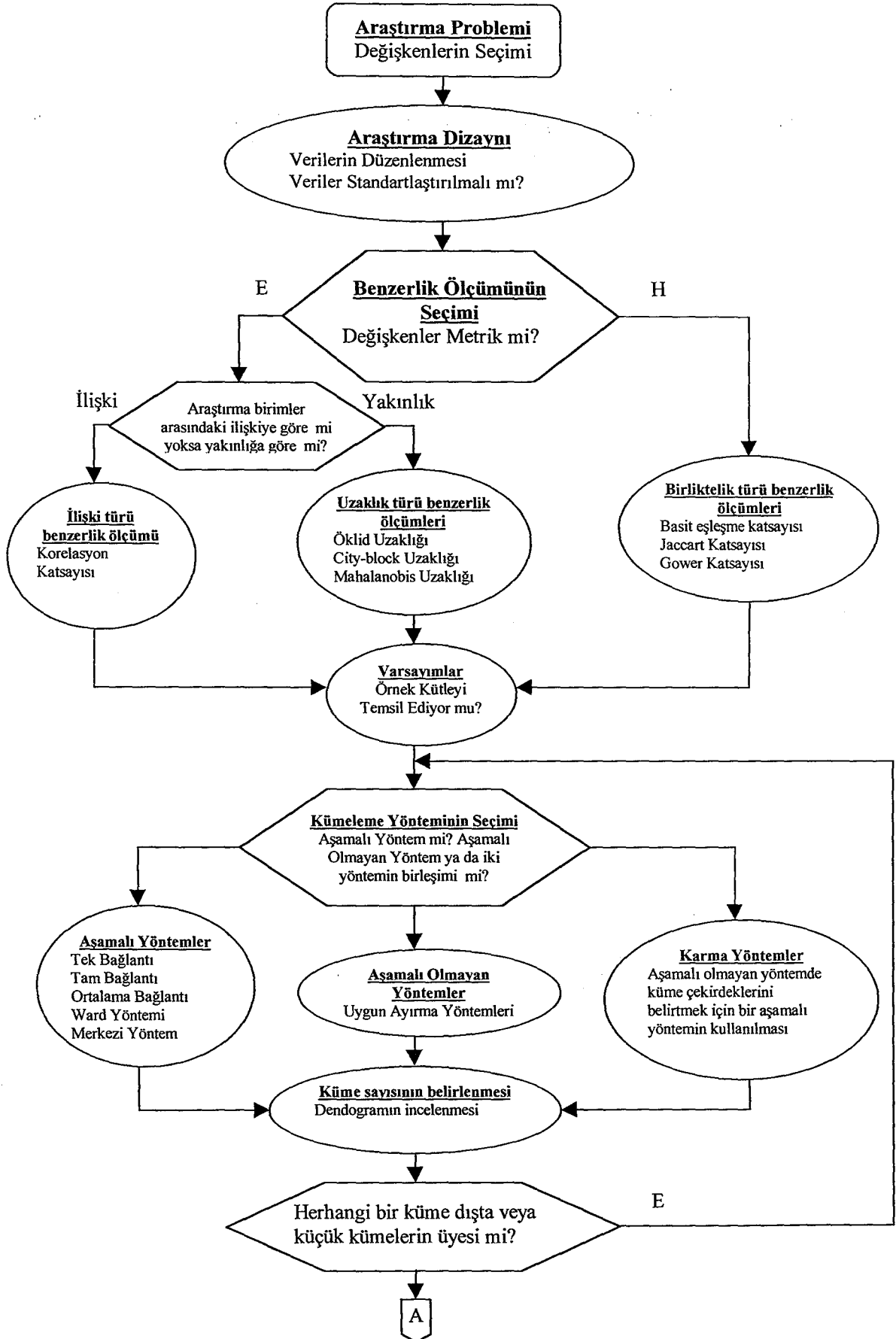
- Veri matrisinin belirlenmesi. Birim ya da değişkenlerin doğal gruplamaları hakkında kesin bilgilerin bulunmadığı anakütlelerden alınan n sayıda birimin p sayıda değişkenine ilişkin gözlemleri elde edilerek bunların ilgili tekniklerin uygulanabileceği bir matris formuna getirilmesi.

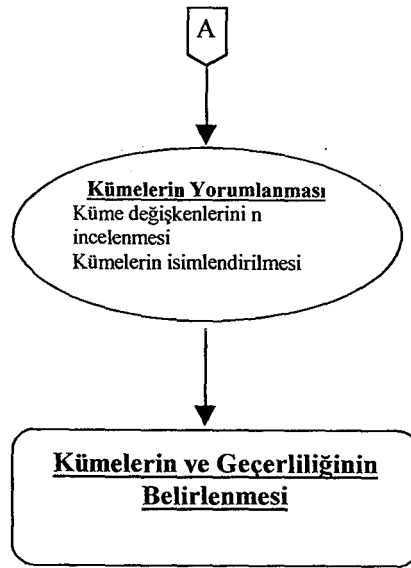
- Benzerlik ya da farklılık matrisinin belirlenmesi. Birimlerin ya da değişkenlerin birbirleri ile olan benzerliklerini ya da farklılıklarını gösteren uygun bir benzerlik ölçüsü ile birimlerin ya da değişkenlerin birbirine uzaklıkları hesaplanarak, benzerlik ya da farklılık matrisinin oluşturulması.

- Uygun kümeleme yöntemi yardımı ile benzerlik ya da farklılık matrislerine göre birimlerin ya da değişkenlerin uygun sayıda kümelerle ayrılması.

- Elde edilen kümelerin yorumlanması ve bu kümeleme yapısına dayalı olarak kurulan hipotezlerin doğrulanması için gerekli analitik yöntemlerin uygulanması.

Yukarıdaki genel açıklamalar şekil 5.1'de ayrıntılı olarak gösterilmiştir.





Şekil 5.1. Kümeleme Analizi Karar Süreci

5.5. Kümeleme Analizi Uygulanacak Veri Matrisinin Oluşturulması

Kümeleme Analizi, n birimin p değişkeninden elde edilen çok değişkenli veri matrisine uygulanır. Açık ki, bu matris n satır ve p sütunlu bir matristir [29]. Böyle bir X veri matrisi aşağıdaki gibi gösterilebilir.

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdot & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & \cdot & X_{2p} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ X_{n1} & X_{n2} & \cdot & \cdot & \cdot & X_{np} \end{bmatrix}$$

X matrisinin x_{ij} elemanı i . birey j . değişkenin değerini gösterir [30]. x_{ij} değerleri sayısal ya da isimsel (nominal) olarak ölçümlenebilecek ve kodlanmış değerler içeren türden olabilir. Eğer p tane değişken sayısal olarak ölçümlenebilecek türden ise veri matrisinde ham veri şekliyle yer alabilir. İsimsel olarak ölçümlenebilecek değişkenler ise 0, 1, 2, 3, gibi skorlanarak (puanlanarak) veri matrisinde yer alır [29].

$n \times p$ boyutlu X veri matrisinde yer alan p tane değişkenin ölçümünde kullanılan

ölçü birimleri ve bunların dağılma aralıkları birbirinden farklı olabilir. Bu durumda analize geçmeden önce veri matrisinde değişkenlerle ilgili bu farklılığın giderilmesi gerekir. Bunun için bilinen ve en çok başvurulan yöntem, verileri standart normal dağılıma sahip bir değişkene dönüştürmektir. Verilerin dönüştürülmesine izleyen bölümlerde değinileceğinden burada üzerinde durulmayacaktır. Dönüştürülmüş verilerden oluşan yeni veri matrisi elde edildikten sonra, p değişken yönünden benzeşimleri, benzerlik ya da farklılık (similarity or dissimilarity) ölçüsü adı verilen nesnel ölçüleri elde edilebilir. Böylelikle birim ya da değişkenler benzer özellik taşıyanları benzerlik düzeylerine göre bir araya getirilebilir.

5.5.1. Değişkenlerin Seçimi

Kümeleme Analizi uygulanacak X veri matrisinde yer alacak değişkenlerin, birimlerin nasıl sınıflandırılacağı üzerine büyük bir etkisi vardır. Kümeleme Analizi ile birlikte, kullanılan değişkenlerin seçimi, araştırma sürecindeki en kritik adımlardan birisidir. Bu noktada temel problem, analizin üzerinde şekillendiği benzerliğin genel kavramını en iyi vurgulayan değişkenleri elde etmektir. İdeal değişkenler, sınıflandırmayı destekleyecek rasyonel değişkenler olmalıdır [27]. Başka bir anlatımla seçilmiş değişkenler araştırılmakta olan sınıflandırmanın tipi ile ilgili olmalıdır. Dikkate alınabilecek diğer bir sorun da her bireyde kaç tane değişkenin ölçülmesi gereğidir? Uygulamada hangi değişkenlerin ölçüleceği sorununda olduğu gibi, kullanılacak değişkenlerin sayısını belirlemek için de sağlam kuramsal temeller yoktur ve bu yüzden de karşılaşılan probleme gözlemsel yaklaşılmalıdır [31].

Birimler arasındaki benzerliğin artması veya azalması sadece incelenen değişkenlere değil, aynı zamanda değişkenlerin sayısına da bağlıdır. Öte yandan kümeleme analizinin sonuçlarını etkileyebilecek ve kuvvetli ayrımlara neden olabilecek bir değişken analiz dışında bırakıldığında, bazı kümeler şekilsiz ve yorumu zor bir karışıklığın içine girebilirler. Bu nedenle sınıflandırma araştırmalarında, değişkenlerin çoğunluğu araştırma kapsamına alınmalıdır [26]. Ayrıca değişken sayısının birim (nesne) sayısından fazla olması da istenmeyen bir durumdur [17]. Çok fazla sayıdaki değişken, kümelerin kararlılığına zarar verebilir. Bu durumlarda kümeleme analizindeki

değişkenlerin indirgenmesi için uygun bir yöntem olan Temel Bileşenler Analizinin (TBA) değişkenlere uygulanmalıdır [32].

Değişkenlerin seçiminde dikkat edilmesi gereken diğer bir durum ise değişkenler arasındaki ilişkidir. Değişkenler arasındaki ilişkinin yüksek olması, değişkenlerin bağımsızlığını zedeleyebileceğinden istenmeyen bir durumdur. Bu durumda değişkenler arasında korelasyonu içeren Mahalonobis uzaklık ölçüsü kullanılarak analiz yapılmalıdır. Bir diğer yaklaşım ise değişkenlere Temel Bileşenler Analizi veya Faktör Analizi uygulayarak uygun değişkenleri belirlemek ve onlara Kümeleme Analizi uygulamaktır [12].

5.5.2. Verilerin Dönüşümü

Çoğu uygulamalarda kümelenecek olan birimleri tanımlayan değişkenler aynı ölçü birimlerinde ölçülmezler. Veri matrisinde yer alacak verilerin bazıları kategorik, sıra belirten veya bir aralık ölçümüne sahip değişken değerleri olabilir. Ayrıca değişkenlerin ortalama ve varyansları birbirlerinden çok farklı olduklarında büyük ortalama ve varyansa sahip değişkenler diğer değişkenlerin rollerini önemli oranda etkilerler. Bu gibi durumlarda verilerin standardize edilmesi ya da belirli aralıklarda gözlenen değerlere dönüştürülmesi uygun olmaktadır [12].

5.5.2.1. Verilerin Standartlaştırılması

Çoğu uygulamalarda kümelenecek olan birimleri tanımlayan değişkenler kategorik, sıralı ya da aralık ölçekli olabilirler. Özelliklerin ölçümü için ölçekler farklı olduğundan veri matrisi derlenirken verileri standartlaştırmak gerekir. Bu tip verilere uygulanan yöntem z skorlarına dönüştürme olup, dönüşüm,

$$Z_{ij} = \left(X_{ij} - \bar{X}_j \right) / S_j \quad (5-1)$$

eşitliği ile gerçekleştirilir [33].

Bu bir standardize edilmiş [z] veri matrisini ortaya çıkarır. Standardize edilmiş [z]'nin her sütununun ortalaması 0 ve standart sapması 1 olur [33].

5.5.2.2. Verilerin $-1 \leq X \leq +1$ Aralığına Dönüşümü

Bu dönüşüm veri dizisinde heterojen yapıda değerlerin ve aşırı uçlarda değerlerin bulunduğu durumlarda tercih edilmektedir. Değerler arasında artı ve eksi değerlerin bulunması halinde uygulanan yöntem olup, X_{\max} , dizideki en büyük değer olmak üzere dönüşüm,

$$x_i = \frac{X_i}{X_{\max}} \quad (5-2)$$

ile yapılır [34].

5.5.2.3. Verilerin $0 \leq X \leq +1$ Aralığına Dönüşümü

Eğer seride aşırı değerler varsa enbüyük değer X_{\max} , en küçük değer X_{\min} ve değişim aralığı $R = X_{\max} - X_{\min}$ olmak üzere, dönüşüm,

$$x_i = \frac{X_i - X_{\min}}{R} \quad (5-3)$$

ile yapılır. Bu dönüşüm ile veri değerleri 0-1 aralığında yer alır [34].

5.5.2.4. Maksimum Değer 1 Olacak Şekilde Dönüşüm

Bu dönüşüm dizideki değerlerin maksimum değerinin 1 olması istendiğinde uygulanan bir yöntemdir. Dönüşüm;

$$x_i = \frac{X_i}{X_{\max}} \quad (5-4)$$

ile yapılır. Eğer dizide maksimum değer sıfır ise dönüşüm,

$$x_i = \frac{X_i}{|X_{\max}|} \quad (5-5)$$

biçiminde yapılır [34].

5.5.2.5. Dönüşüm Serisinin Standart Sapması Bir Olacak Şekilde Dönüşümü

Bu dönüşüm yöntemi, yeni serinin standart sapmasının 1 olması istendiğinde uygulanır. Dönüşüm,

$$x_i = \frac{X_i}{S} \quad (5-6)$$

biçiminde yapılır. Eğer serinin standart sapması sıfır ise verilere dönüşüm uygulanmaz. Eğer mutlaka dönüşüm yapılması gerekiyor ise yukarıda açıklanan yöntemlerden uygun birisi ile dönüşüm yapılabilir [12].

5.5.2.6. Logaritmik Dönüşümler

Bazı veri tiplerinde aşırı uç değerler varsa, bu değerlerin analize katkıları diğer değerlerle aynı oranda değildir. Özellikle klinik kimya verilerinde iki tane çok yüksek ve patolojik değerler arasındaki büyük bir fark, bir yüksek değer ve bir ortalama değer arasındaki farklılık aynı öneme sahip olmayabilir. Bu tip verilere Lineer olmayan bir dönüşüm uygulanabilir. Burada Logaritmik bir dönüşüm, uç değerlerin sınıflandırma üzerinde çok büyük etkiye sahip olmasını önler. Böyle bir dönüşüm tercihen uç değerler üzerinde ve standardizasyondan önceki herhangi bir durumda yapılmalıdır [32].

6. BENZERLİK VE BENZERLİK ÖLÇÜLERİ

Kümeleme analizinde amaç, yakınlık veya benzerlik ölçümünü gerektiren karmaşık bir veri kümesinden hareketle daha basit bir grup yapısı üretmektir [35]. Sınıflandırmaların bütün yöntemlerinin bir ön gerekliliği iki farklı nesne arasındaki benzerlik veya uzaklık ölçümüdür [36].

Birimleri belirli özelliklerine göre sınıflandırma ya da gruplandırma, birimlere ilişkin daha düzenli ve güvenilir bilgiler verir. Sınıflandırma bir çok özellik yönünden benzerlik gösteren birimleri bir isim altında toplamak olarak açıklanabilir.

Ana çizgileri ile kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine göre belirli gruplar içinde toplayarak sınıflandırma yapmayı, birimlerin ortak özelliklerini ortaya koyma ve bu sınıflar ile ilgili genel tanımlamalar yapmayı sağlayan bir yöntemdir. Gerçekte iki değişkene göre incelenen ve farklı özellik taşıdıkları açıkça belli olan birimler, değişken sayısı artırıldığında ortak yönlerinin de arttığı ortaya çıkabilir ve çoğu benzer olan özelliklerinden dolayı aynı sınıfta yer alabilirler. Bu nedenden dolayı birimleri sınıflandırırken bu birimlerin p adet değişkenden hareketle (ölçülerek, tartılarak ya da nitel özellikler sonuç değerlerine göre sayısallaştırılarak) veri matrisleri oluşturmak ve n bireyi sınıflara ayırmak, model sınıflar oluşturmak bakımından en uygun yaklaşımdır. Böylelikle birimlerin, benzerlik gösterenleri bir sınıfta toplanacaktır [13].

p sayıda değişken itibari ile n birimin gözlenmesi ile elde edilen çokdeğişkenli verilerin analiz öncesinde benzerlik verisi durumuna getirilmesi gerekir [29]. Birimler ölçülen p değişken yönünden benzer, hemen hemen benzer ve farklı olmak üzere sınıflandırılabilir. Bu sınıflandırmayı bilimsel olarak yapmak için birimler arasındaki benzeşimleri benzerlik ya da farklılık (similarity or dissimilarity) ölçüsü adı verilen nesnel ölçülerle değerlendirmek gerekir [19].

Kümeleme ya da sınıflandırma birimler arasındaki benzerlikler dikkate alınarak yapılacak ise benzerlik verileri gözlemlenen n birimden oluşturulabilecek tüm olası $n(n-1)/2$ tane çift için benzerlik hesaplanarak elde edilir. Eğer değişkenler arasındaki benzerliklerin belirlenmesi ile ilgileniliyorsa p değişkenden oluşturulabilecek olası

$p(p-1)/2$ tane çift için benzerlik hesaplanacaktır [29].

Birimler veya değişkenler kümelenirken, yakınlık bazı uzaklık ölçütleri ile belirlenir [35]. Birimlerin ya da değişkenlerin kümelenmesi için, herhangi iki birim ya da değişken arasındaki uzaklığa dayanan benzerlik ölçülerinden yararlanır. Bunun için benzerlik ölçülerinin yer aldığı benzerlik matrisi kullanılır. S benzerlik matrisi, S_{jk} benzerlik ölçülerini içeren üçgen matris biçiminde, $n(n-1)/2$ elemana sahip bir matristir. S benzerlik matrisi şekil 6.1' deki gibidir.

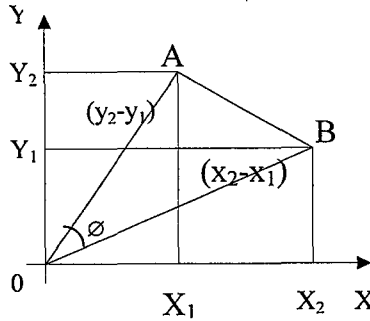
$$S = \begin{bmatrix} S_{21} & & & & & \\ S_{31} & S_{32} & & & & \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & & \cdot & & \\ \cdot & \cdot & & & \cdot & \\ S_{n1} & S_{n2} & \cdot & \cdot & \cdot & S_{n(n-1)} \end{bmatrix}$$

Şekil 6.1. S benzerlik matrisi

Benzerlik, iki boyutlu bir uzayda iki birimin birbirine olan uzaklıkları hesaplanarak belirlenebilir [27]. Koordinat sisteminde yer alan A ve B noktaları arasındaki doğrusal uzaklık, A' nın koordinat değerleri $A(x_1, y_1)$ ve B' nin koordinat değerleri $B(x_2, y_2)$ olmak üzere şekil 6.2'deki gibi gösterilebilir ve uzaklık Pisagor bağıntısına göre;

$$d(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

şeklinde hesaplanır.



Şekil 6.2. İki boyutlu uzayda iki nokta arasındaki uzaklığın gösterimi.

Noktaların geometrik olarak gösterimlerinde ikiden daha fazla boyut olduğunda noktalar arasındaki uzaklıkları çok boyutlu olarak hesaplamak gerekir [12]. Bu uzaklıklar birimler arası benzerlik ya da farklılık olarak isimlendirilir [27].

Kümeleme Analizinde, birimlerin p değişkene göre birbirleri arasındaki uzaklıkları hesaplamak için farklı uzaklık ölçü birimleri ileri sürülmüştür. Uzaklık ölçülerinin kullanılması orijinal veri matrisinin içerdiği verilerin ölçeğine göre farklılaşmaktadır. Oransal ölçekle elde edilmiş verilerde Minkowski ölçümünün özel bir yaklaşımı olan Öklid uzaklığı kullanılmaktadır. Birimlerin (cases) sınıflandırılması yapılmak istendiğinde uzaklık türü ölçülerle hesaplanmış benzerlik matrisinin tercih edilmesi, değişkenlerin sınıflandırılması yapılmak istendiğinde ise ilişki türü ölçülere dayalı farklılık matrislerinin tercih edilmesi uygun düşmektedir [37].

Genelde uzaklık ölçüleri doğrudan birim ya da değişkenlerin kümelenmesinde kullanılabileceği gibi birim ya da değişkenler arasındaki benzerlik ya da farklılıkların hesaplamasında da kullanılabilir. Veri matrisinde yer alan n birimin p değişkene göre uzaklıkları, uzaklık matrisi (Distance Matrix) adı verilen D matrisi ile gösterilir. D matrisinin elemanları d_{ij} ya da $d(i,j)$ biçiminde, birimlerin birbiri ile olan benzerlik düzeyleri ise benzerlik matrisi ile gösterilir. Benzerlik matrisinin elemanları D matrisinin elemanlarına göre belirlenir. Benzerlik matrisi elemanları $sim(i,j)$ ya da sim_{ij} biçiminde gösterilir ve $sim_{ij}=100(1-d_{ij}/\max(d_{ij}))$ biçiminde hesaplanır. Birimlerin birbirlerinden farklılıkları dis_{ij} şeklinde gösterilir ve benzerlik matrisinin elemanlarına göre hesaplanır. Farklılık matrisi elemanları $diss(i,j)$ ya da $diss_{ij}$ biçiminde gösterilebilir ve $diss(i,j)=100-sim(i,j)$ biçiminde hesaplanır. Değişkenler arasındaki benzerlikler ilişki matrisi R yardımı ile belirlenir. R matrisinin elemanları r_{ij} biçiminde gösterilir [12].

x ve y gibi verilen iki nokta arasındaki uzaklık, d ve uzaklık fonksiyonu $d(x,y)$ olarak yazılabiliyor ise bu durumda, benzerliğin doğru ölçüm olduğunu anlamak için kuramda dört ölçüt vardır [27, 33, 38].

1. Simetri: x ve y verilen iki nokta, aralarındaki uzaklık d ise

$$d(x, y) = d(y, x) \geq 0$$

2. Üçgen eşitsizliği: x , y ve z verilen üç nokta olsun ve aralarındaki uzaklıklar

$$d(x, y) \leq d(x, z) + d(y, z)$$

eşitsizliğini verir.

3. Aynı (benzer) olmayanların ayırt edilebilirliği:

$$d(x, y) \neq 0 \text{ ise } x \neq y$$

4. Benzerlerin (aynı olanların) ayırt edilemezliği:

$$d(x, x') = 0$$

Bir noktanın izdüşümü kendisine eşittir.

Herhangi iki birim ya da değişken arasındaki uzaklığa dayanan benzerlik ölçüleri; uzaklık (distance) türü benzerlik ölçüleri, ilişki (correlation) türü benzerlik ölçüleri ve birliktelik (association or similarity) türü benzerlik ölçüleri olarak üç grupta incelenebilir [31].

6.1. Uzaklık Türü Benzerlik Ölçüleri.

6.1.1. Öklid Uzaklığı.

Öklid uzaklığı birimlerin kümelenmesinde ve özellikle sayısal değişkenler için en sık kullanılan bir benzerlik ölçüsüdür [39]. Birimlerin kümelenmesinde en sık kullanılan Öklid uzaklığına dayalı n birimin oluşturduğu mümkün $n(n-1)/2$ çift arasındaki uzaklıklardan oluşan S benzerlik matrisinin elemanları S_{jk} ;

$$S_{jk} = d_{jk} = \left[\sum_{i=1}^p (x_{ij} - x_{ik})^2 \right]^{1/2} \quad (6-1)$$

olarak hesaplanır [26]. Burada S_{jk} ; i . değişken için j ve k birimleri arasındaki toplam uzaklığı belirtmektedir [19].

Kümeleme Analizinde; Öklid uzaklık ölçüsü, birimler ya da değişkenler arasındaki uzaklığı, değişken ölçü birimlerinden etkilenmeksizin belirten bir ölçüdür.

Bu üstünlüğü nedeniyle Öklid uzaklık ölçüsü, analizde oldukça sık kullanılmakta ve favori benzerlik ölçüsü olarak ileri sürülmektedir [40].

Kümeleme Analizlerinde Öklid uzaklığı Karesel Öklid uzaklığı olarak da kullanılmaktadır. (6.1) nolu eşitliğin karesi alındığında Kare Öklid uzaklığı ölçüsü elde edilmektedir ve

$$S_{jk}^2 = d_{jk}^2 = \sum_{i=1}^p (x_{ij} - x_{ik})^2 \quad (6-2)$$

şeklinde hesaplanmaktadır [38].

6.1.2. Minkowski Uzaklığı

X_{ij} , i. deęişkendeki j. birimin deęeri iken j ve k birimleri arasındaki Minkowski ölçümü $L \geq 1$ için

$$D_p = (X_j, X_k) = \left[\sum_{i=1}^n |x_{ij} - x_{ik}|^L \right]^{1/L} \quad (6-3)$$

dir. L'nin deęişik deęerlerini seçerek pek çok farklı uzaklık ölçüm fonksiyonları elde edilebilir. Böyle tanımlanan "City-Block", "taxi cap" veya L_1 ölçümü $L = 1$ alınarak sağlanır ve

$$S_{jk} = D_1 = (X_j, X_k) = \sum_{i=1}^n |x_{ij} - x_{ik}| \quad (6-4)$$

olarak hesaplanır.

Minkowski ölçümünde $L = 2$ alındığında L_2 ölçümü Öklid uzaklığı elde edilir. L deęerinin ∞ alındığı durumda, sınırsız L artışlarında Chebychev uzaklığı;

$$D_\infty (X_j, X_k) = \max_{i=1, \dots, n} |x_{ij} - x_{ik}| \quad (6-5)$$

elde edilir [26, 35].

6.1.3. Mahalanobis D^2 Uzaklığı

Mahalanobis D^2 uzaklığı Öklid uzaklığının karesi olarak hesaplanmaktadır. p değişkenli iki nokta arasındaki en büyük uzaklığı veren bir ifadedir. Mahalanobis uzaklığı Öklid ve diğer Minkowski uzaklık ölçülerinden farklı olarak değişkenler arasındaki ilişkiyi koruması nedeniyle tercih edilmektedir [31, 41]. Bu uzaklık;

$$d_{ij} = \sqrt{(x_i - x_j)^T C^{-1} (x_i - x_j)} \quad (6-6)$$

şeklinde hesaplanır. Burada C kovaryans matrisini x_i ve x_j , i . ve j . birimlerin değişkenlerinin aldıkları değerlerin vektörlerini göstermektedir [32].

6.1.4. Vektör Çarpım Uzaklığı

Vektör çarpım uzaklığı ölçüsü, p boyutlu bir uzayda noktalar arasındaki veri vektörleri ve görsel uzunlukları arasındaki açısal farkın benzerlik ölçüsü olarak alındığı bir ölçüdür [13].

Her bir veri kümesinde X ve Y değişkenlerine ait satır vektörleri;

$$X^T = (x_1, x_2, \dots, x_n)$$

$$Y^T = (y_1, y_2, \dots, y_n)$$

olarak gösterilir. Burada her vektörün i . bileşeni değişken üzerinde ölçülen i . veri kümesinin sonucudur ve üstyazım T , transpoze anlamına gelmektedir.

İki vektörün iç çarpımı ya da skaler çarpımı

$$\langle X, Y \rangle = X^T Y = \sum_{i=1}^n x_i y_j \quad (6-7)$$

olarak ifade edilir. Bu nicelik; X ve Y arasındaki çapraz çarpımların toplamıdır. $X^T X$ vektörünün kendisiyle iç çarpımı X ' in kareler toplamıdır. Kareler toplamının kare kökü Öklid normu veya vektörün uzunluğudur ve yaygın olarak $|X|$ veya $\|X\|$ şeklinde yazılır.

X ve Y arasındaki iç çarpım için alternatif bir ifade α , X ve Y arasındaki bir açı iken

$$X^T Y = |X| |Y| \cos \alpha \quad (6-8)$$

eşitliği elde edilir. $\cos \alpha$ iki nokta arasındaki uzaklığın ölçüsüdür ve

$$A(X, Y) = \cos \alpha = \frac{X^T Y}{|X| |Y|} = \frac{\sum_{i=1}^n x_i y_i}{\left(\left[\sum_{i=1}^n x_i^2 \right] \left[\sum_{i=1}^n y_i^2 \right] \right)^{1/2}} \quad (6-9)$$

şeklinde hesaplanır [26].

$\cos \alpha$ değeri X ve Y arasındaki benzerliğin bir ölçüsü olarak alınabilir ve birimlerin kümelenmesi yerine değişkenlerin kümelenmesi tercih edildiğinde kullanılan bir ölçüdür [19].

6.2. İlişki Türü Benzerlik Ölçüleri

6.2.1. Pearson İlişki Katsayısı

Açısal bir katsayı olan Pearson ilişki katsayısı, iki birim (nesne) ya da değişken arasındaki ilişkiyi belirlemek için kullanılan bir benzerlik ölçüsüdür. Bu ölçü kümeleme analizinde benzerliğin bir ölçüsü olarak kullanılır ve sınıflandırmalar buna göre yapılır. Daha çok değişkenler arası kümeleme söz konusu olduğunda kullanılmakta olan bir benzerlik ölçüsüdür [27]. Pearson ilişki katsayısı;

$$r_{jk} = \frac{\sum_{i=1}^p (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{s_j s_k} \quad (6-10)$$

şeklinde hesaplanmaktadır. p tane değişken için ölçümleri alınan herhangi iki birim (i. ve k. birim) arasındaki ilişkiyi veren bu katsayı da S_j j. birimin, S_k ise k. birimin standart sapmalarını belirlemektedir. \bar{x}_k ise k. birim için değişkenlerin değerlerinin ortalamasını ifade eder [26].

6.2.2. Sıra (Rank) Korelasyon Katsayısı

Benzerlik katsayısına, en yüksekten en düşüğe doğru sıralanan sonuçlar için ya da en yüksekten, en düşüğe doğru sıralanan nesnelere için, gerek duyulduğunda baş vurulan bir benzerlik katsayısıdır.

X_{ij} sıra korelasyon katsayısı, j değişkenine karşı i birimi tarafından verilen bir sıra olduğu zaman ve $d_{ij}^2 = \sum (x_{ij} - x_{hj})^2$ k elemanlarına karşı i ve h varlığı tarafından yapılan sıralar arasındaki farklılıkları gösterdiği zaman, Spearman sıra korelasyon katsayısı olan ρ_s ;

$$\tau_s = 1 - \frac{6 \sum d^2}{k^3 - k} \quad (6-11)$$

olarak hesaplanır [42].

6.3. Birliktelik Türü Benzerlik Ölçüleri

Birimler, anlamlı p boyutlu ölçümlerle gösterilemediği zaman birimlerin çiftleri kesin özelliklerin varlığı veya yokluğu temeline dayanarak karşılaştırılırlar. Benzer birimler benzemeyen birimlerden daha fazla özelliğe sahiptirler. Bu tip ikili değerler alan değişkenlerin tanımlanması durumunda birimler arasındaki benzerlikleri ölçmek için birliktelik türü benzerlik ölçülerinden faydalanılır [35]. Bu tip ölçümler ayrıca ikili değişkenlerle tanımlanan durumlar arasındaki benzerliği ölçmede kullanılır [27]. Bazı durumlarda bunlar herhangi bir niteliğin varlığı ya da yokluğu ile ilgilidir. Var-yok şeklinde iki değerli değişkenlerle ilgileniliyorsa ve i . birimde k . tür varsa $x_{ik} = \text{evet}$, aksi halde hayır olacaktır. Bu durumlar skorlanarak şöyle ifade edilebilir:

$$X_{ik} = \begin{cases} 1 & \text{evet ise} \\ 0 & \text{hayır ise} \end{cases}$$

olacaktır.

Birliktelik katsayısı hesaplamasında 2×2 olasılık tablosundaki i ve k birimleri için eşleşmelerin ve uyumsuz eşleşmelerin sıklıklarından elde edilir. Çizelge 6.1.

Çizelge 6.1. 2x2 lik olasılık tablosu

		k. birim		
		1	0	Toplam
i. birim	1	a	b	a+b
	0	c	d	c+d
Toplam		a+c	b+d	p = a+b+c+d

Bu olasılık tablosunda a kategorisi; i. birimin 1 değerine sahip değişkenlerinin k. birimindeki aynı özelliğinin de 1 değerine sahip olanlarının, birebir eşleşmelerin sıklığını gösterir. b kategorisi ise i. birimin 1 değerine sahip değişkenlerinin k. birimdeki aynı özelliğin 0 değerine sahip olanlarının, 1-0 eşleştirmelerin sıklığıdır. c ve d kategorileri değerleri de benzer şekilde olmama veya uyumlaşmama durumlarına göre eşleştirilerek sıklıkları belirlenir. Böylelikle i. ve k. birimler arasındaki benzerlikler, birliktelik türü benzerlik ölçüleriyle hesaplanabilir.

Bazı uygulamalarda gruplandırılacak bu birimlerden çok, değişkenlerin gruplandırılması düşünülebilir. Değişkenler ikili oldukları zaman veriler bir olasılık tablosu şeklinde tekrardan düzenlenebilir. Bu durumda birimlerin yerine değişkenler, kategorileri gösterirler. Değişkenlerin her bir çifti için tabloda kategorize edilen n birim vardır. Bilinen 0 ve 1 kodlamaları tekrardan değişkenler için düzenlenir. a, b, c ve d değerleri birimler için olduğu gibi değişkenler için de sıklık değerleri olarak elde edilir. Burada p değişken sayısı yerine n birim sayısı ($n = a+b+c+d$) gelir [35].

6.3.1. Basit Eşleşme Katsayısı (The Simple Matching Coefficient)

Bu katsayı; uyumlu olan değişkenlerin toplamının, değişkenlerin toplamına oranıdır. Birlikteliğin hesaplanmasında katsayı 1-1 ve 0-0 durumlarını dikkate alır [42]. Bu katsayı;

$$s = \frac{a + d}{a + b + c + d} \quad (6-12)$$

şeklinde hesaplanmaktadır. Burada s, 0 ve 1 arasında değişen iki durum arasındaki benzerliktir [27].

6.3.2. Jaccard Katsayısı

Mikrobiyolojik ve Taksonomik bulgularda ikili değerlere göre değerleri saptanan birimlerin belirli bir özelliğe sahip olanların pozitif ve negatif özellikler gösterenlere oranını belirten Jaccard Katsayısı bir benzerlik ölçüsü olarak ele alınmıştır [34]. İki tür arasındaki benzerliğin özelliklerinin karşılıklı varlığını kabul eden katsayı negatif eşleşmeleri hariç tutmak gerektiğinde hesaplanmaktadır [43]. Dolayısıyla d ye karşılık gelen olumsuzun (0-0 eşleşmelerin) dışlandığı, her iki birliktelik değerlerinin 1-1 durumunun dikkate alındığı bir katsayıdır [31, 42]. Bu katsayı;

$$s = \frac{a}{a + b + c} \quad (6-13)$$

şeklinde hesaplanmaktadır.

Jaccard katsayısı 0 ile 1 arasında değişmektedir [27].

6.3.3. Gower Katsayısı

İki durumlu, çok durumlu (sıralı ve niteliksel) ve niceliksel üç tip özelliğe uygulanabilen genel bir benzerlik katsayısıdır. Gower katsayısı iki j ve k bireyleri için $0 \leq s_{ijk} \leq 1$ sonucu ve i özelliği için w_{ijk} ağırlığını (tartısı) belirleyerek bir skor atamak üzere elde edilir. Bu katsayı;

$$s_G = \left(\sum_{i=1}^n w_{ijk} s_{ijk} \right) / \sum_{i=1}^n w_{ijk} \quad (6-14)$$

olarak tanımlanır. Burada w_{ijk} değişken ağırlık değeri, k değişkeni için 0 veya 1 değeri atanır. İkili değişkenlerde, kıyaslanan her iki durumda k bilinmiyorsa w_{ijk} sıfırdır. Negatif olarak tanımlanan eşleşmede w_{ijk} yine sıfır alınır. İkili değerler alan değişkenler hariç, sadece birimlerden birinin veya ikisinin de k . değişkenin değerleri bilinmediğinde 0 değerini alır.

İki durumlu özelliklerde eşlemeler için $s_{ijk} = 1$ ve uygunsuz eşlemeler için $s_{ijk} = 0$ dır. İki durumlu bir karakterin negatif durumu için iki tane birim eşlendiği zaman $w_{ijk} = 0$ dır.

Çok durumlu özellikler (sıralı ve niteliksel) için, durumlar arası eşlenmelerde $s_{ijk} = 1$ dir ve uygunsuz eşlenmede 0 dır.

Niceliksel özellikler için x_{ij} ve x_{ik} veri matrisinde tanımlandığı gibi ve R_i örnekteki veya bilinen ana kütle üzerindeki i özeliğinin değişim aralığı olduğu zaman;

$$s_{ijk} = 1 - \frac{|x_{ij} - x_{ik}|}{R_i} \quad (6-15)$$

olarak hesaplanır. Bu anakütlerdeki özelliklerin toplam değişim aralığı bilgisinden veya örnekteki gözlenen değişim aralığından belirlenir. Özellik durumları özdeş olduğu zaman $s_{ijk} = 1$ ve iki karakter j ve k birimlerinin değişim aralığının uçlarını birleştirdiğinde $s_{ijk} = 0$ olur. Genel bir yaklaşımla Gower benzerlik katsayısında ikili veriler için s_{ijk} ve w_{ijk} çizelge 6.2'deki gibi hesaplanabilir.

Çizelge 6.2. İkili veri için s_{ijk} ve w_{ijk} ' ya değer atama.

i durumu	1	1	0	0
j durumu	1	0	1	0
s_{ijk} sonucu	1	0	0	0
w_{ijk} ağırlığı	1	1	1	0

Sıra verilerde, karşılaştırma değerleri aynı ise $s_{ijk}=1$, eğer aynı değil ise $s_{ijk}= 0$ ' dır [27, 31, 43].

7. KÜMELEME YÖNTEMLERİ

Kümeleme Analizi, doğal gruplandırmalar veya birimlerin kümelendirmeleri için, verileri araştırmakta kullanılan bir araştırmacı istatistiksel süreçtir. Karmaşık bir veri kümesinden birim gruplarını belirleme işi mutlaka yakınlık ölçümünü gerektirir [44]. Bu yakınlık değerlerinden yararlanılarak birimlerin kümelere (gruplara) atanması işlemi analizin bir sonraki adımını oluşturur. Bu işlem, tek bir yaklaşımdan oluşan bir istatistiksel yöntem değil, birimlerin ya da değişkenlerin kümeleneğinde bir çok yöntemi içeren uygulamalardan yararlanılan istatistiksel süreçlerden oluşmaktadır.

Kümeleme analizinde pek çok yöntem bulunmakta, bu yöntemler veri matrisinden elde edilen benzerlik matrisinin analiz biçimine göre beş ana grupta incelenebilmektedir [13, 19]. Bunlar;

- Aşamalı Kümeleme Yöntemleri (Hierarchical Clustering Methods).
- Aşamalı Olmayan Kümeleme Yöntemleri (Nonhierarchical Clustering Methods).
- Ardışık Parçalama Yöntemleri (İterativ Partitioning Methods).
- Optimizasyon Yöntemleri (Optimizing Procedures).
- Diğer Kümeleme Yöntemleri.

Olarak isimlendirilmektedir.

Atama, Sıralama, Grafikselleştirme gibi parametrik olmayan yöntemler diğer kümeleme yöntemleri içinde yer almaktadır [19].

Burada dikkat edilmesi gereken ilk durum benzer nesnelere ya da birimleri gruplara veya kümelere yerleştirmede hangi yöntemin kullanılacağıdır. Hangi kümeleme algoritmasının veya kurallar dizisinin hangisinin en uygun olacağına karar verilmesidir. Örneğin biyologlar evrimsel gelişmeye ilgilerinden dolayı aşamalı yöntemleri güvenilirler ve kullanırlar [42].

Bütün kümeleme algoritmaları kümeler arası farklılıkları maksimize etmeyi amaçlarlar. Seçilen bir kümeleme algoritması, küme içi homojenliği ve kümeler arası heterojenliği sağlamalıdır [45].

Yukarıda sözü edilen kümeleme analizi yöntemleri arasında Aşamalı ve Aşamalı Olmayan Kümeleme Yöntemleri uygulamada yaygın kullanım alanı bulan yöntemlerdir. [11, 13, 19, 26, 42, 45].

Bu çalışmada gözlemler hem aşamalı hem de aşamalı olmayan kümeleme yöntemlerine göre analiz edileceğinden yalnız bu iki yöntem ele alınacaktır.

7.1. Aşamalı Kümeleme Yöntemleri

Aşamalı Kümeleme Yöntemleri, değişkenlerin kümelenmesinde p değişkenin $p(p-1)/2$ ya da birimlerin kümelenmesinde $n(n-1)/2$ tüm olası çiftlerinin aralarındaki ilişki veya uzaklık türü ya da birliktelik türü benzerlik ölçülerini dikkate alarak, değişkenleri ya da birimleri birbirlerine aşamalı bir biçimde bağlamayı amaçlayan yöntemlerdir [19, 26].

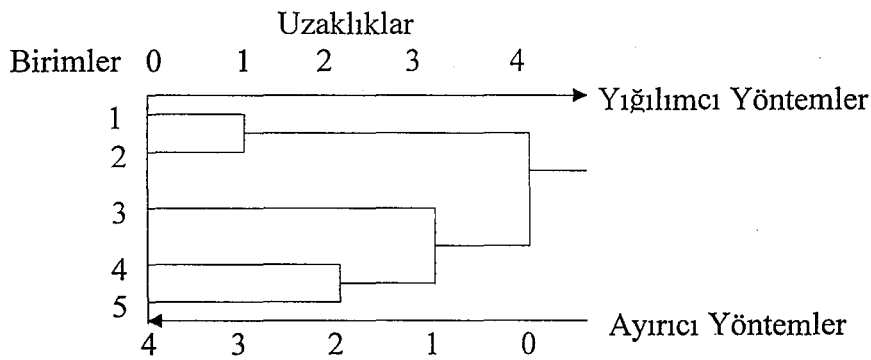
Aşamalı Kümeleme Yöntemleri, birimlerin benzerliklerini dikkate alarak belirli düzeylerde birbiri ile birleştirmeyi amaçlayan bir süreçtir [46]. Bu süreçle birimler için benzerlik düzeyine göre ağaç benzeri bir aşamanın kurulması amaçlanır. Aşamalı sınıflandırmanın sonucu dendogram adı verilen bir tür ağaç grafiği ile gösterilebilir. Ağaç grafiğinin aşamalı olarak bağlantıları, birimlerin oluşturdukları kümelerin birbirine olan uzaklık ve yakınlıklarını belirlemeye yardımcı olur. Bir kümenin heterojenliği ağaçta daha üst dallara tesadüf eden uygun dalların yüksekliği ile gösterilir. Bu nokta ne kadar yüksek olursa, grubun heterojenliği de o kadar yüksek olur. Böylelikle kümelerin birbirine bağlantı uzunlukları iki kümenin hangi aşamada birbiriyle benzer olduğunu belirtir. Benzerliklerine göre önce iki küme birleştirilir ve bu işlem tüm kümelerin tek bir kümede birleşmesine kadar devam eder [19, 32, 45, 47].

Aşamalı yöntemler ağaç benzeri bir yapının veya aşamanın kurulmasını kapsadığından ya ard arda gelen birleşmelerin serisiyle ya da ard arda gelen bölünmelerin serisiyle süreceğinden temel olarak yığılmcı ve ayırıcı yöntemler olmak üzere iki tip aşamalı kümeleme yöntemi vardır [38].

Yığılmcı yöntemlerde her bir nesne veya gözlem kendi kümesiyle başlar. Böylelikle başlangıçta nesnelere kadar çok küme vardır. Her adımda kümelerin sayısı bir indirgenerek en çok benzeyen nesnelere ilk olarak birleştirilirler. Bazı durumlarda üçüncü

birey ilk iki bireyle yeni bir küme oluşturmak için birleşebilir. Diğer benzerlik durumunda iki bireyin diğer bir gurubu yeni bir küme oluşturmak için birleşebilir. En sonunda benzerlik azaldığı için bütün alt gruplar tek bir kümede birleşirler. Bu nedenden dolayı yığılmcı yöntemlerden bazen çoğalma (buildup) yöntemleri olarak da söz edilir [30, 31, 35, 47].

Kümeleme süreci yığılmcı yöntemlerin tersinde işlediği zaman ayırıcı yöntem olarak adlandırılır. Ayırıcı yöntemlerde bütün gözlemleri (nesneleri) kapsayan bir büyük küme ile başlanır. Sonraki adımlarda en çok benzemeyen gözlemler ayrılırlar ve küçük kümelere dönüşürler. Bu süreç, her gözlem kendi başına bir küme olana kadar sürer [31, 35, 45, 47]. Şekil 7.1’de ayırıcı yöntemler sağdan sola doğru giderken yığılmcı yöntemler soldan sağa doğru giderler. En yaygın kullanılan paket programlar yığılmcı yöntemleri kullandığı için ayırıcı yöntemler daha fazla tartışılmayacaktır.



Şekil 7.1. Aşamalı Kümeleme Yöntemlerinde Ağaç Grafiği

Şekil 7.1’deki ağaç grafiğine göre yığılmcı ve ayırıcı kümeleme yöntemlerinin kuramsal uzaklıklara göre küme oluşturmaları kısaca aşağıda gösterilmiştir.

Uzaklık	Yığılmcı Yöntem		Ayırıcı Yöntem	
	Küme Sayısı	Kümelere	Küme sayısı	Kümelere
0	5	1,2,3,4,5	1	(1,2,3,4,5)
1	4	(1,2),3,4,5	2	(1,2),(3,4,5)
2	3	(1,2),(4,5),3	3	(1,2),(4,5),3
3	2	(1,2),(4,5,3)	4	(1,2),3,4,5
4	1	(1,2,3,4,5)	5	1,2,3,4,5

n birimden oluşan bir örnek için yığılmcı aşamalı kümeleme algoritması şöyledir:

1. Her biri bir birimi kapsayan n kümeleri ve $D = \{d_{ij}\}$ uzaklıklarına veya benzerliklerine sahip $n \times n$ simetrik matris hesaplanır.

2. Hesaplanan benzerlik matrisinde en benzer küme çiftleri araştırılır ve en benzer U ve V kümeleri arasındaki uzaklık d_{uv} olarak belirlenir.

3. Benzer olan U ve V kümeleri birleştirilerek (UV) yeni oluşturulmuş küme kabul edilir ve

a. Benzerlik matrisinde U ve V kümelerini temsil eden satır ve sütun iptal edilir ve matristen çıkartılır.

b. (UV) kümesi ve kalan kümeler arasındaki uzaklığı veren satır ve sütunu ekleyerek benzerlik matrisi güncelleştirilir.

4. Tüm birimler tek bir kümede toplanana kadar 2. ve 3. Adımlar $n-1$ kez tekrar edilir ve birleştirilen kümelerin, birleşmenin yapıldığı benzerlik düzeyleri ile kümelerin özellikleri, aşamaları kaydedilir [35].

Küme oluşturan birimlerin birbirlerine birleştirilmesinde benzerlik matrisinin farklı şekilde elde edilmesi, bulunan kümelerin de farklı olmasını doğuracaktır. Benzerlik matrisine bağlı olarak küme geliştirmekte kullanılan en yaygın yığılmcı yöntemler şunlardır [31].

1. Tek Bağlantı Kümeleme veya en yakın komşu kümeleme yöntemi.
(Single-Linkage or the Nearest-Neighbor Method)
2. Tam Bağlantı Kümeleme veya En uzak Komşuluk Kümeleme Yöntemi.
(Complete-Linkage or Furthest-Neighbor Method)
3. Ortalama Bağlantı Kümeleme Yöntemi.
(Average-Linkage Method)
4. Ward Yöntemi veya En Küçük Varyans Kümeleme Yöntemi.
(Ward's Method)
5. Küresel Ortalama Bağlantı Kümeleme Yöntemi.
(Centroit Method)

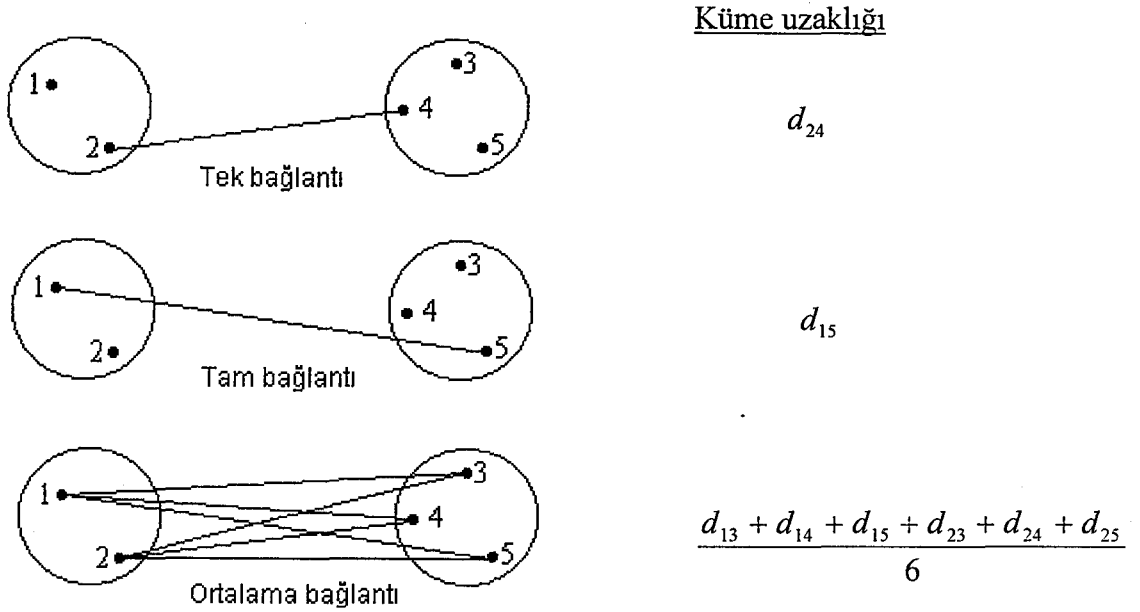
6. Ortanca (Medyan) Bağlantı Kümeleme Yöntemi.

(Median Linkage Cluster Analysis)

7. Lance ve Williams Doğrusal Modeli ve Esnek Kümeleme Yöntemi.

(Lance & Williams Flexible Clustering Method)

Şekil 7.2'de bazı aşamalı kümeleme yöntemleri gösterilmiştir [35, 48].



Şekil 7.2. Aşamalı kümeleme yöntemleri

7.1.1. Tek Bağlantı Kümeleme veya En Yakın Komşu Kümeleme Yöntemi

Tek Bağlantı Kümeleme Yöntemi birimler arasındaki en küçük uzaklığa ya da en çok benzerliğe dayandırıldığından en yakın komşu kümeleme yöntemi olarak da bilinmektedir [47, 48]. Tek Bağlantı Kümeleme Yönteminde i . ve j . birimlerin birleştirilmesiyle oluşturulan yeni kümenin başka herhangi bir k kümesi veya birimi ile ilişkisi uzaklık türü benzerlik ölçüleri kullanılıyor ise;

$$d_k(i,j) = \min(d_{ki}, d_{kj})$$

şeklinde belirlenir. Eğer ilişki türü ya da birliktelik türü benzerlik ölçüleri kullanılıyor ise;

$$d_k = \max(s_{ki}, s_{kj})$$

biçiminde bulunur [19, 26].

Tek Bağlantı Kümeleme Yöntemi'nde; ilk olarak en küçük uzaklık ya da en çok benzer birimler belirlenir ve bunlar ilk kümeyi oluştururlar. Sonraki adımda tekrar en küçük uzaklık hesaplanır, bu uzaklığa göre iki birimli yeni bir küme oluşturulur ya da üçüncü birim bir küme oluşturmak için ilk iki kümeyle birleşir. Bu süreç bütün birimler bir kümede toplanana dek devam eder [45, 47, 49].

Birleştirme yapılırken kümelerin eleman sayısının birden fazla olması koşulu yoktur. Bir birim yalnız başına bir küme oluşturabilir [12].

Tek Bağlantı Kümeleme Yöntemine açıklık kazandırmak için aşağıdaki beş birim çifti arasındaki kuramsal uzaklık matrisini esas alan algoritma ele alınmıştır.

$$d_{ij} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & & & & \\ 9 & 0 & & & \\ 3 & 7 & 0 & & \\ 6 & 5 & 9 & 0 & \\ 11 & 10 & 2 & 8 & 0 \end{bmatrix} \end{matrix}$$

Adım 1: Her bir birim bir küme olarak iki tane en yakın birim birleştirilir.

$$d_{ij} = \min (d_{ij}) = d_{53} = 2$$

Olduğundan 5 ve 3 birimlerini 2 uzaklığında (35) kümesinde birleştirilir.

Adım 2: Bulunan bu (35) kümesi ile diğer kümeler arasındaki en yakın komşu uzaklıklar bulunur ve uzaklıklar matrisi güncelleştirilir.

$$d_{(35)1} = \min \{ d_{31}, d_{51} \} = \min \{ 3, 11 \} = 3$$

$$d_{(35)2} = \min \{ d_{32}, d_{52} \} = \min \{ 7, 10 \} = 7$$

$$d_{(35)4} = \min \{ d_{34}, d_{54} \} = \min \{ 9, 8 \} = 8$$

ve diğer uzaklıklar da,

$$d_{12} = 9$$

$$d_{14} = 6$$

$$d_{24} = 5$$

olarak bulunur ve uzaklık matrisi;

$$d_{ij} = \begin{matrix} & (35) & 1 & 2 & 4 \\ (35) & \begin{bmatrix} 0 & & & \\ 1 & 3 & 0 & \\ 2 & 7 & 9 & 0 \\ 4 & 8 & 6 & 5 & 0 \end{bmatrix} \end{matrix}$$

olarak elde edilir.

Adım 3: Tüm birimler bir büyük kümede toplanana dek adım 1 ve 2 tekrar edilir.

Yenilenmiş uzaklık matrisinde küme çiftleri arasındaki en küçük uzaklık

$$d_{ij} = \min (d_{ij}) = d_{(35)1} = 3$$

olduğundan 1 kümesini (35) kümesi ile 3 uzaklığında (135) kümesinde ve uzaklık matrisini tekrar güncelleştirildiğinde yeni uzaklıklar şu şekilde olur.

$$d_{(135)2} = \min \{ d_{12}, d_{(35),2} \} = \min \{ 9, 7 \} = 7$$

$$d_{(135)4} = \min \{ d_{(35)4}, d_{14} \} = \min \{ 8, 6 \} = 6$$

$$d_{24} = 5$$

bulunur. Yenilenmiş uzaklık matrisi,

$$d_{ij} = \begin{matrix} & (135) & 2 & 4 \\ (135) & \begin{bmatrix} 0 & & \\ 2 & 7 & 0 \\ 4 & 6 & 5 & 0 \end{bmatrix} \end{matrix}$$

olarak hesaplanır. Yenilenmiş uzaklık matrisinde küme çiftleri arasındaki minimum uzaklık

$$d_{ij} = \min d_{ij} = d_{24} = 5$$

dir. 2 ve 4 birimleri 5 uzaklığında birleştirilerek yeni bir (24) kümesi oluşmaktadır.

Burada (135) ve (24) şeklinde iki farklı küme oluşmuş olur. Bu kümelerin en yakın komşu uzaklığı

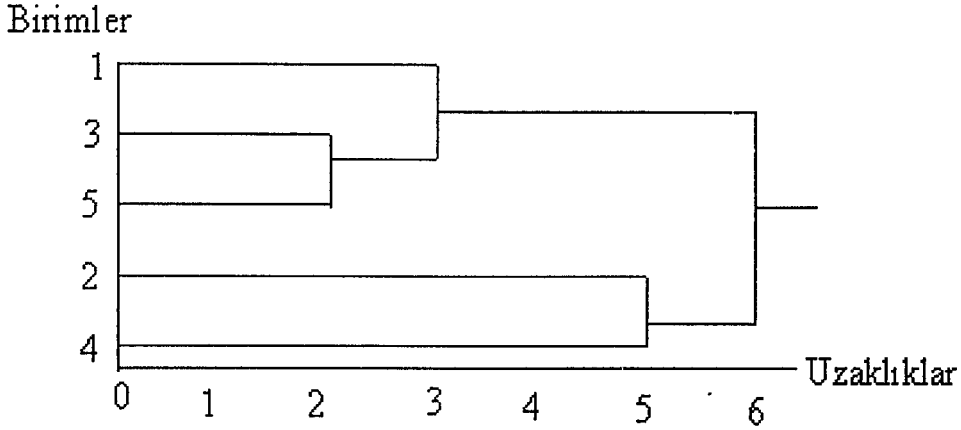
$$d_{(135)(24)} = \min \{ d_{(135)2}, d_{(135)4} \} = \min \{ 7, 6 \} = 6$$

bulunur. Yenilenmiş uzaklık matrisi de,

$$d_{ij} = \begin{matrix} & (135) & (24) \\ (135) & \begin{bmatrix} 0 & \\ (24) & 6 & 0 \end{bmatrix} \end{matrix}$$

olarak hesaplanır. Sonuç olarak birbirinden farklı bu iki küme 6 uzaklığında beş birim (1,2,3,4,5) tek bir kümede birleştirilirler [35].

Tek Bağlantı Kümeleme'nin sonuçları $n(n-1)/2$ uzaklıkları küçükten büyüğe doğru sıralanarak bir dendogram veya ağaç diyagram şeklinde şekil 7-3' deki gibi grafiksel olarak gösterilebilirler. Ağaçtaki her dal birleştirilen kümeleri gösterir.



Şekil 7-3. Beş birim arasındaki uzaklıklar için Tek Bağlantı dendogramı

Tek Bağlantı yönteminin en önemli avantajı matematiksel özellikleridir. Benzerlik matrisinin transformasyonu için tek bağlantı yöntemi değişmez ve veriler içindeki bağlardan etkilenmez. Bu özellik diğer aşamalı yöntemlerde yoktur. Bu yöntemin önemli dezavantajlarından birisi pratik kullanımda uzun zincirler oluşturan kümelerin ortaya çıkmasıdır [27]. İlgili yayınlarda Tek bağlantı yönteminin verdiği sınıflamaların diğer yöntemlere göre daha etkin olduğu belirtilmektedir [37].

7.1.2. Tam Bağlantı Kümeleme veya En Uzak Komşu Kümeleme Yöntemi

Tam Bağlantı Kümeleme Yöntemi, Tek Bağlantı Kümeleme Yönteminin tersine birimler arasındaki maksimum uzaklığa dayandırılır. Kümedeki bütün birimler birbirine maksimum uzaklık ya da minimum benzerlikte bağlandığından dolayı bu yöntem tam bağlantı ya da en uzak komşuluk olarak adlandırılır [11, 27, 31, 45, 47, 48, 49]. Tek bağlantı Kümeleme Yöntemine çok benzeyen bu yöntemde tek farklılık iki küme arasındaki uzaklık olarak her kümedeki eleman çiftleri arasındaki uzaklığın en büyüğü veya benzerliğin en az olanı ele alınmasıdır.

Kümelenecek birimler arasındaki en küçük uzaklığa ya da maksimum benzerliğe sahip olan (UV) kümesi belirlendikten sonra bu küme ile diğer bir W kümesi arasındaki uzaklıklar,

$$d_{(uv)w} = \max\{d_{uw}, d_{vw}\}$$

ile belirlenir. Birimler ilişki türü veya birliktelik türü benzerlik ölçülerine göre kümelenecekler ise (UV) ve W kümeleri arasındaki benzerlik

$$s_{(uv)w} = \min\{s_{uw}, s_{vw}\}$$

şeklinde belirlenir [13,19,26,42].

7.1.3. Ortalama Bağlantı Kümeleme Yöntemi

Ortalama Bağlantı Kümeleme Yöntemi, tek bağlantı ve tam bağlantı gibi başlar fakat kümeleme ölçütü bir kümedeki birimlerden diğer kümedeki bireylere olan ortalama uzaklıktır. Bu yöntem tek bağlantı veya tam bağlantıda olduğu gibi uç değerleri kullanmaz ve ayırma işlemi, uç değerlerin birtek çiftinden çok kümenin bütün elemanlarına dayandırılır. Ortalama bağlantı yöntemi kümeleri küçük varyanslarla birleştirmeye yöneliktir [42,45,47,48,49]. Bu yöntem tam bağlantı ve tek bağlantı yöntemleri arasında sonuçlar vermesi nedeniyle bir seçenek olarak önerilmektedir [11,27].

Ortalama bağlantı yöntemi uzaklıklar veya benzerlikler matrisinden yararlanılarak değişkenlerin veya birimlerin gruplandırılması için kullanılabilir. Örneğin, uzaklık veya benzerlik matrisinden belirlenen en çok benzer U ve V birimlerinin (UV) kümesini oluşturduğu varsayalım. Bu kümeye her hangi bir W kümesi arasındaki uzaklık d_{ik} ; W kümesindeki k birimi ile (UV) kümesindeki i birim arasındaki uzaklık olsun.

$N_{(uv)w}$ ve N_w ; (U) ve W kümelerindeki birimlerin sayısı olduğunda

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)} N_w} \quad (7-1)$$

ile belirlenir [35].

7.1.4. Ward Yöntemi veya En Küçük Varyans Kümeleme Yöntemi

Ward Yöntemi ile kümeler arası uzaklıklar hesaplanmaz daha çok küme içi homojenliği maksimize eden kümeler oluşturur. Genel olarak grup içi kareler toplamı, homojenliğin ölçümü olarak kullanılır. Bu yöntemle toplam grup içi veya küme içi karelerin toplamı minimize edilmeye çalışılır. Açıktır ki minimize edilen küme içi kareler toplamı en az olur [49].

Ward Yöntemi olguların fonksiyonel ilişki içinde olduğu durumlarda ve birimlerin belirlendiği kümenin ortalamasından her birimin kareli sapmalarının toplamıyla ölçüldüğü gibi, birimlerin küme içinde gruplandırmasından kaynaklanan bilgi kaybının ortaya çıktığı veri yapılarına uygulanmasında çokça kullanılan bir yöntemdir [31, 32, 48].

Bu yöntemde, hata kareler toplamının sıfır olduğu kabul edildiğinden başlangıçtaki her birimin kendisinin bir küme olduğu varsayılır. Dolayısıyla birleştirilecek ilk kümede iki birimin en yakın olanı araştırılır [31, 49].

n birimden elde edilen p değişkene ilişkin verilere göre birimler k kümeye ayrılmak istendiğinde öncelikle k. kümenin j. biriminin i. skorunu x_{ijk} ve k. kümedeki birimlerin i. değişken için ortalaması \bar{x}_{ik} olarak gösterilsin. Bu durumda k kümesinin ortalamalar vektörüne göre Öklid uzaklıkları toplamı olan k kümesinin hata kareler toplamı W_k ,

$$W_k = \sum_{i=1}^p \sum_{j=1}^{n_k} (x_{ijk} - \bar{x}_{ik})^2 = \sum_{i=1}^p \sum_{j=1}^{n_k} x_{ijk}^2 - n_k \sum_{i=1}^p \bar{x}_{ik}^2 \quad (7-2)$$

şeklinde hesaplanır. W_k değeri tüm kümelerde hesaplanarak toplam küme içi hata kareler toplamı W,

$$W = \sum_{k=1}^{k=n} W_k \quad (7-3)$$

şeklinde hesaplanır [19, 26].

Bu değerler araştırıldıktan sonra W'de en küçük artışa sahip olan p ve q kümeleri birleştirilerek t kümesi elde edilir. W'deki bu artış,

$$DW_{pq} = W_t - W_p - W_q \quad (7-4)$$

eşitliği ile hesaplanır. Böylece n birim $(n-1)$ kümeye ayrılmış olur. Böylelikle küme sayısı $k=1$ oluncaya kadar W artış değerleri bulunarak birimlerin aşamalı biçimde birbirine bağlanmaları sağlanır [19].

7.1.5. Merkezi Bağlantı Kümeleme Yöntemi

Merkezi Bağlantı Yönteminde iki küme arasındaki uzaklık, bunların merkezleri (ortası) arasındaki uzaklık olarak belirlenir. Bu yöntemde her adımda yeni bir birey veya bireyler grubu var olan kümeye ekleneceğinden, küme merkezleri yer değiştirecektir [47]. Eğer bir kümenin bir gözlemi varsa o zaman merkezin kendisi gözlemdir. Süreç kendi merkezleri arasındaki uzaklığa göre grupları birleştirerek sürer. En kısa uzaklıklı gruplar öncelikle birleştirilirler [48]. Her aşamada en yakın ortalama vektörle iki küme veya merkez birleştirilir. Yeni küme grup merkezinin koordinatlarıyla yer değiştirir. Böylece kümeler arası uzaklık küme merkezleri arası uzaklık olarak tanımlanır [42].

Her g_i grubu \bar{x}_j merkeziyle tanımlanır ve iki tane g_p ve g_q grupları arasındaki uzaklık uygun grupların kareli Öklit uzaklığı ise en küçük uzaklıklı bu gruplar birleştirildiğinde merkezi yöntem elde edilir. Uzaklıklar arasındaki eşitlik ,

$$d_{ki} = \frac{n_p}{n} d_{pi} + \frac{n_q}{n} d_{qi} - \frac{n_p n_q}{n^2} d_{pq}, \quad (n = n_p + n_q) \quad (7-5)$$

şeklindedir [32].

Bu yöntemin dezavantajı birleştirilmiş olan iki grubun hacimleri çok farklı olduğunda yeni grubun merkezi büyük grubun merkezine çok yakın olacaktır. Böylelikle daha küçük grubun karakteristik özellikleri kaybolacaktır. Birleştirilecek olan grupların hacimlerinin eşit olması durumunda, yeni oluşturulan grup iki grubun arasında kalmış olacaktır. Bu nedenden dolayı yöntemin kullanılması, birleştirilecek olan grupların hacimlerinin eşit olduğu varsayıldığı durumda uygun olacaktır [31, 42].

7.1.6. Lance ve Williams Doğrusal Modeli ve Esnek Kümeleme Yöntemi

Lance ve Williams tüm aşamalı kümeleme yöntemlerinde hesaplanabilecek doğrusal bir model geliştirmişler ve bu modele dayanarak uzaklık türü benzerlik matrisini yenilemişlerdir.

Lance ve Williams, g_p ve g_q gruplarını birleştirerek oluşturulan g_r grubunu elde ettikten sonra g_r grubu ile diğer g_i grubu arasındaki uzaklıkları ele alır. d_{ij}^2 ; g_i ve g_j grupları arasındaki uzaklığın ölçümü iken ,

$$d_{ir}^2 = \alpha_p d_{ip}^2 + \alpha_q d_{iq}^2 + \beta d_{pq}^2 + \gamma |d_{ip}^2 - d_{iq}^2| \quad (7-6)$$

doğrusal modeli ile benzerlik matrisi güncelleştirilmiş olur. Modelde α_p , α_q , β ve γ değerleri çizelge 7.1'deki gibi değiştirilerek bütün aşamalı yöntemler elde edilebilir. Çizelge 7.1'deki n_p ; g_p grubundaki birimlerin sayısıdır. g_q deki birimlerin sayısı n_q ve $n_r = n_p + n_q$ dir.

$$\gamma = 0 \text{ ve } \alpha_p, \alpha_q \text{ ve } \beta \text{ arasında}$$

$$\alpha_p + \alpha_q + \beta = 1$$

$$\alpha_p = \alpha_q$$

$$\alpha < 1$$

ilişkisi olduğunda Lance ve Williams'ın esnek staretejisi en iyi aşamalı yapıyı belirlemeye çalışır. $\beta = 1$ olduğunda tam zincir ortaya çıkar. β sıfıra düştüğünde ve gittikçe negatif olduğunda güçlü gruplandırmalar elde edilir [28, 50].

Çizelge 7.1. Esnek kümeleme yönteminde bazı aşamalı kümeleme yöntemlerine uyan parametrelerin değerleri.

Yöntem	α_p	α_q	β	γ
Tek Bağlantı	$\frac{1}{2}$	$\frac{1}{2}$	0	-1/2
Tam Bağlantı	$\frac{1}{2}$	$\frac{1}{2}$	0	1/2
Ortalama Bağlantı	n_p / n_r	n_q / n_r	0	0
Ward Yöntemi	$(n_p + n_i) / (n_r + n_i)$	$(n_q + n_i) / (n_r + n_i)$	$-n_i / (n_r + n_i)$	0
Ortanca (Medyan)	$\frac{1}{2}$	$\frac{1}{2}$	-1/2	0

7.1.7. Ortanca (Medyan) Bağlantı Kümeleme Yöntemi

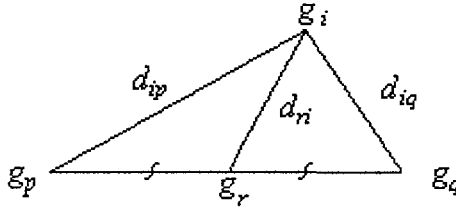
Ortanca Bağlantı yönteminde g_p ve g_q gruplarını birleştirerek oluşturulan herhangi g_i ve g_r grubu arasındaki yeni d_{ri}^2 uzaklıklarını hesaplamanın yöntemi g_p ve g_q 'nin göreceli büyüklüklerinden bağımsızdır. Böylece, örnekteki birimlere özdeş ek

birimler eklenirse kümeleme yöntemi etkilenmez. Bu yöntemde r ve i grupları arasındaki uzaklık,

$$d_{ri}^2 = \frac{1}{2}(d_{rp}^2 + d_{rq}^2) - \frac{1}{4}d_{pq}^2 \quad (7-7)$$

şeklinde hesaplanır.

Eğer uzaklık, kareli Öklid (d_{ri}^2) uzaklığı ise g_r geometrik olarak g_p ve g_q gruplarını gösteren noktaları birleştiren doğrunun orta yerine yerleştirir. d_{ri}^2 değeri g_i den çizilen $g_i g_p g_q$ üçgeninin ortanca uzunluğunun karesidir. Bu durum şekil 7.4'de gösterilmektedir.



Şekil 7.4. Ortanca Bağlantının analitik gösterimi.

Bu yöntem aynı zamanda metrik olmayan $\{d_{ij}^2\}$ uzaklık ölçümlerinde de kullanıma uygundur [28].

7.2. Aşamalı Olmayan Kümeleme Yöntemleri

Aşamalı olmayan kümeleme yöntemlerinde aşamalı kümeleme yöntemlerinin aksine temel düşünce; n biriminin k kümeye parçalanması ve birimlerin k kümeye ayrılmasında parçalama işlemi rasgele ya da verilerin incelenmesi ile sistematik olarak yapılabilir olmasıdır. Böylelikle n birimin ilk k bölünmesi belirlendikten sonra küme elemanlarını kümeler arası yerlerini değiştirecek en iyi bölünme bulunabilir.

Aşamalı olmayan kümelemede; veriler her bölünmenin bir kümeyi temsil etmesiyle, k bölünmelerine veya gruplarına bölünürler. Böylece aşamalı kümelemeye karşı, kümelerin sayısının öncelikli olarak belirlenmesi gerekir [35, 49].

Birimlerin ayrılacakları küme sayısı belirlendikten sonra, kümeler için belirlenen küme belirleme kriterlerine göre birimlerin hangi kümelere girebileceklerine karar verilir ve atama işlemleri yapılır [19].

Aşamalı olmayan kümeleme yöntemlerinden bazıları, kümelerin çekirdeğini oluşturacak olan çekirdek noktalar seçilerek, bu noktaları oluşacak kümenin merkezleri kabul ederek bu çekirdekler etrafında oluşacak kümelere birimlerin dağıtılmasını önerir. Bazı yöntemler ise küme sayısını deneme yoluyla belirlemeyi ve $k = 2, 3, \dots, m$ ise $m < n$ olacak şekilde bir ardışık parçalamayı önerir [19].

Aşamalı olmayan kümeleme yöntemlerinde uzaklıkların (benzerlikler) bir matrisi belirlenmek zorunda olmadığından, ayrıca bu yöntemlerin kuramsal dayanaklarının daha güçlü olması ve benzerlik matrisi yerine X veri matrisini ya da dönüştürülmüş veri matrisinin ele alınabilmesi yöntemlerin tercih nedenidir.

Aşamalı olmayan kümeleme yöntemlerinde çekirdek noktaların seçiminde yöntemden yararlanacak araştırmacının elde ettiği veriler hakkında ayrıntılı bilgi sahibi olması önem taşımaktadır. Çekirdek noktaların belirlenmesinde farklı yöntemler önerilmiştir. Aşağıdaki adımlar çekirdek noktaların nasıl genelleştirildiğini kısaca göstermektedir [19, 26, 51].

- 1- Veri kümesindeki ilk k veri birimleri çekirdek nokta olarak seçilir.
- 2- Veri birimleri birden m'ye kadar numaralanır ve sistematik olarak $m/k, 2m/k, \dots, (k-1)m/k$ ve m inci gözlemler çekirdek nokta olarak alınır.
- 3- Veri kümesinden herhangi k tane veri birimi öznel olarak seçilir. Bunlar çekirdek nokta olarak belirlenir.
- 4- Veri birimleri birden m'ye kadar numaralanır ve 1-m aralığında rasgele sayılara karşı gelen k tane veri birimi seçilir.
- 5- Her bir değişken için dağılım aralığına göre koordinatlar vektörü olarak k nokta türetilir. Türetilen bu noktalar çekirdek nokta olarak kabul edilir. Bu noktalar veri kümesindeki birimlere karşılık gelen noktalardan farklı olabilirler.
- 6- Veri birimlerini arzu edilen k kümeye parçalayacak bir parçalama kalıbı seçilir ve her kümenin merkezleri çekirdek noktaları olarak alınır.

Aşamalı olmayan kümeleme yöntemleri; Bölünme Yöntemleri, Yoğunluk ve Dağılım Arama Yöntemleri ve En İyi Bölünme Ölçütleri olarak üç temel bölümde incelenebilir. Bu çalışmada Bölünme Yöntemleri ve aşamalı olmayan yöntemler içinde yer alan k-Ortalamalar Yöntemine değinilecektir.

7.2.1. Bölünme Yöntemleri

Bölünme yöntemlerinde k bölünmelerinin ortalamaları ya da k çekirdek nokta başlangıç noktası olarak alınıp uzaklık ölçütüne göre en yakın birimler kümelere atanarak kümeler oluşturulur. Bunun için Forgy (1965) aşağıdaki adımlar dizisini içeren basit bir algoritma önermiştir [26].

1- Kümelemeye istediğiniz herhangi bir biçimle başlayın. Eğer çekirdek noktaların bir kümesiyle başlandıysa adım 2 ye, veri birimlerinin ayrılmasıyla başlandı ise adım 3 e gidin.

2- Her veri birimini en yakın çekirdek nokta ile bir kümeye ayırın.

3- Çekirdek noktaları, oluşturulan çekirdek kümelerin ortalaması olarak alın ve uzaklık ölçütüne göre kümeleri atayın.

4- Süreç her bir veri biriminin adım 2 deki küme üyeliğini değiştirmeyene kadar kısaca küme elemanlarının yerlerinin değişmez olmasına kadar adım 2 ve 3 değişimli olarak uygulanır.

Aşamalı olmayan yöntemler içinde sayılan yaklaşımlardan en yaygın kullanılanları bölünme yöntemleri içerisinde ifade edilebilen Mac Queen'in "k-ortalama" kümeleme yöntemidir.

7.2.1.1. k-Ortalamalar Yöntemi

Mac Queen'in k-ortalamalar adını verdiği yöntem gözlemleri kümelerin önceden tanımlanmış sayısına gruplandırmakla işleme başlar [52]. Böylece her biri tek gözlemden oluşan k tane küme ile işleme başlanır ve her yeni bir gözlem en yakın ortalama gruba eklenir. Gruba yeni bir gözlem eklendikten sonra küme ortalaması yeniden hesaplanır. Bu süreç tüm gözlemler gruplara atanıncaya kadar devam eder. Tüm

gözlemler gruplara atandıktan sonra atandıkları küme ortalamasından daha yakın küme ortalaması varsa, gözlemlerin yerleri değiştirilmekte o kümeye atanmaktadır [51, 53].

k-ortamalar yönteminde, her seferinde bir birim küme merkezlerini değiştirir ve kümelerin eski ve yeni merkezleri yeniden hesaplanır. Eğer i birimi v kümesinden w kümesine taşınıyorsa yeni merkezlerin koordinatları,

$$\bar{x}_f(v') = \frac{1}{n_v - 1} (n_v \bar{x}_f(v) - x_{if})$$

ve

(7-8)

$$\bar{x}_f(w') = \frac{1}{n_w + 1} (n_w \bar{x}_f(w) + x_{if})$$

ile verilirler. v' ve w' yeni kümeleri temsil etmektedir. Bu yöntemde herhangi bir birimin taşınabilmesi yeni merkeze daha yakın ise gerçekleşebilir. Böylece merkezlere olan uzaklıkların kareleri toplamı azalmaktadır [54].

p değişken için elde edilen n tane gözlem ve k küme için k-ortalama yöntemi aşağıdaki adımları izleyerek birimlerin kümelenmesini gerçekleştirir. Mac Queen, n gözlemin k kümede toplanması için aşağıdaki yolu göstermiştir [26].

1- İlk k gözlemi her biri bir elemanlı küme olarak alınır ve bunların her biri birer küme ortalaması olarak kabul edilir. Tüm birimlerin küme ortalamalarına olan uzaklıkları hesaplanır.

2- Geriye kalan (n-k) birimin her biri en yakın küme ortalaması olan kümeye atanır. Her atamadan sonra küme ortalaması yeniden hesaplanır.

Bütün veri birimleri 2. adımda k kümeye atandıktan sonra küme ortalamaları yeni çekirdek nokta olarak alınır ve en yakın ortalama göre atama işlemi küme elemanlarının yerlerinin değişmez olmasına kadar tekrarlanır.

8. TÜRKİYE' DE SAĞLIK AÇISINDAN HOMOJEN İL GRUPLARININ BELİRLENMESİ

Bu çalışmada Türkiye'de bulunan illerin sağlık düzeylerine göre gruplandırılmasında Kümeleme Analizi kullanılmıştır. Böylelikle illerin, sağlık açısından kendi içlerinde benzer, birbirleri arasında ise farklılık gösteren kümelere ayrılması amaçlanmıştır.

Çalışmaya sağlık açısından illerin yapısal niteliklerini belirleyici varsaydığımız 30 değişken kullanılarak başlanmıştır. Dolayısıyla veri matrisi 79 il'e ait 30 değişken kümesini içermektedir ve 79x30 boyutundadır. Değişken kümesindeki veriler çizelge 3.1'de verilmiştir. Değişkenlerin sayısal değerleri TL, km, kişi gibi çok farklı birimlerde ve farklı dağılım aralığına sahip olduklarından, her değişkene ait veriler;

$$z = \frac{X_i - \bar{X}}{s}$$

yaklaşımıyla standartlaştırılarak dönüştürülmüştür. Standartlaştırma işlemi ile farklı birimlerde ve büyüklüklerde olan değişkenlerin, bu yüzden daha fazla ya da daha az ağırlıklandırılmaları önlenmiştir.

İstatistiksel analizler, amaca uygun olarak SPSS ve MINITAB istatistiksel paket programları kullanılarak bilgisayar ortamında yapılmıştır. Bu paket programlarda veri kümesi olarak, dönüştürülmüş veri matrisi kullanılmıştır. İllerin gruplandırılmasında; n sayıda birimi, p sayıda değişkene göre saptanan özellikler itibarı ile gruplar, olabildiğince kendi içinde türdeş (homojen) ve kendi aralarında farklı (heterojen) kümelere ayıran Kümeleme Analizi kullanılmıştır. Elde edilen verilerin analizinde; farklı küme sayısına göre k-ortalamar yöntemi ve aşamalı kümeleme yöntemleri farklı uzaklık ölçülerine göre tek tek uygulanmış ve öncelikle görsel olarak uygunsuz gruplamaları veren yöntemler dışlanmıştır. Daha sonra gruplamaları görsel olarak uygun olan kümeleme analizi sonuçlarına, birimleri en az hata ile ait oldukları gruplara ayırmak için geliştirilmiş Diskriminant (ayırma) Analizi uygulanmıştır [55]. Kümeleme ve ayırma analizi sonuçları birbiri ile karşılaştırılarak Aşamalı Kümeleme

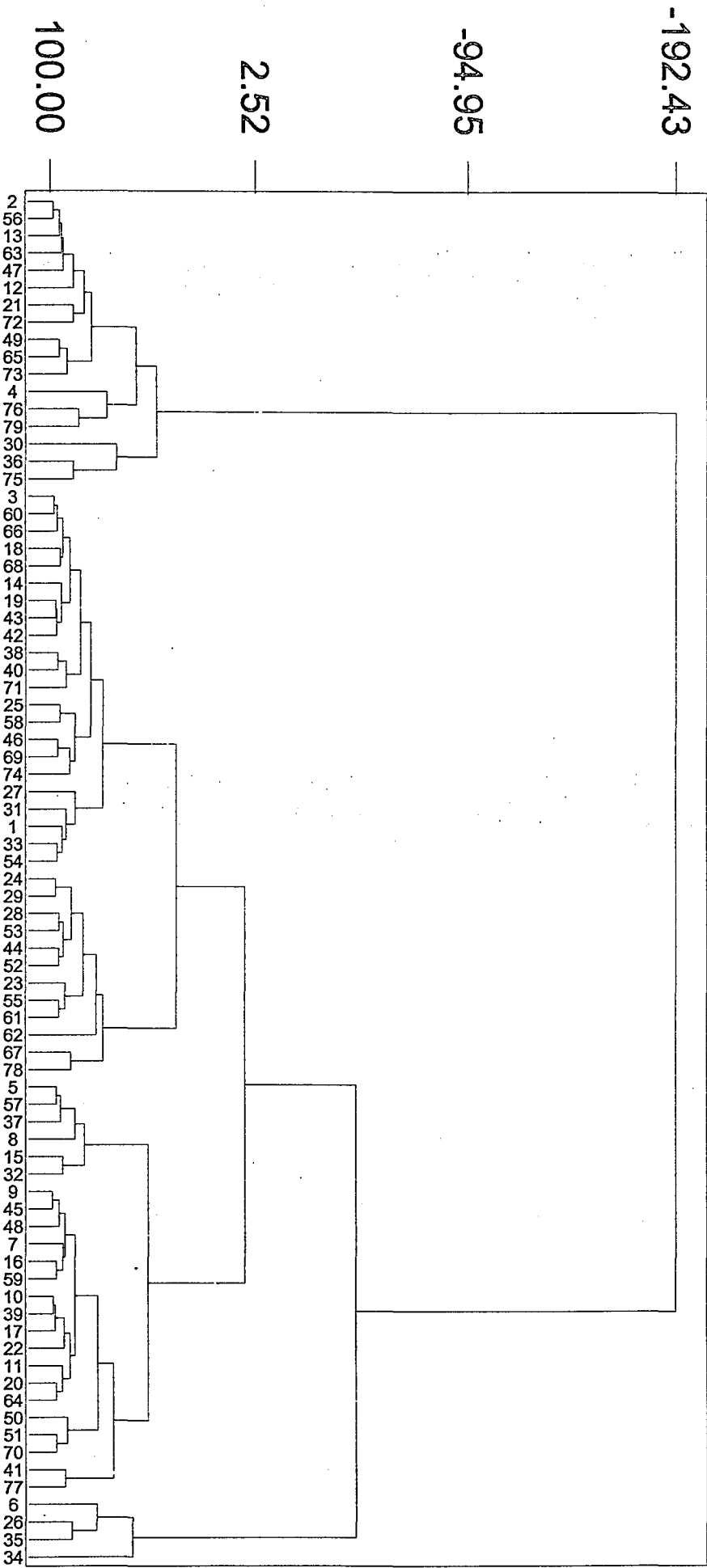
Yöntemlerinden Ward Yönteminin en uygun sonuçları oluşturduğuna karar verilmiştir. Farklı uzaklık ya da benzerliklere göre kümeleme yöntemlerine ait sonuçlar Ek 1’de birer dendogram halinde verilmiştir. Ayrıca Ward Yöntemi kümeler arası uzaklıkları hesaplama dan küme içi homojenliği maksimize edip grup içi kareler toplamını minimize etmeye çalıştığından, çalışmanın amacına uygun küme gruplarının elde edilmesi sağlanmıştır.

Bu şekilde 79 ilin küme içi homojenliğin maksimum ve küme içi varyansın minimum olacak şekilde 8 alt küme ye ayrıldığı belirlenmiştir. Sonuçlara ait dendogram şekil 8.1’de verilmiştir. Dendogramda küme oluşturan birim sayıları illerin plaka numarasını göstermektedir. Bu kümelerin elde edilmesinde Ward yöntemi, Karesel Öklid uzaklığı ve 0.60 benzerlik düzeyi kullanılmıştır. Elde edilen sonuçlar çizelge 8.1’de özetlenmiştir.

Çizelge 8.1. Ward Yöntemine göre illerin kümelene mesi

Küme No	Kümeye Yer Alan İller
1	Adana, Afyon, Aksaray, Bartın, Bayburt, Bolu, Çankırı, Çorum, Erzurum, Gaziantep, Hatay, İçel, Kayseri, Kırıkkale, Kırşehir, Konya, Kütahya, K. Maraş, Sakarya, Sivas, Tokat, Yozgat
2	Adıyaman, Batman, Bingöl, Bitlis, Diyarbakır, Mardin, Muş, Siirt, Şırnak, Şanlıurfa, Van
3	Ağrı, Iğdır, Kilis
4	Amasya, Artvin, Burdur, Isparta, Kastamonu, Sinop
5	Ankara, Eskişehir, İstanbul, İzmir
6	Antalya, Aydın, Balıkesir, Bilecik, Bursa, Çanakkale, Denizli, Edirne, Karaman, Kırklareli, Kocaeli, Manisa, Muğla, Nevşehir, Niğde, Tekirdağ, Uşak, Yalova
7	Elazığ, Erzincan, Giresun, Gümüşhane, Karabük, Malatya, Ordu, Rize, Samsun, Trabzon, Tunceli, Zonguldak
8	Ardahan, Hakkari, Kars

Similarity



Şekil 8.1. Ward Yöntemi Karasal Öklid uzaklığı 0.60 Benzerlik düzeyine göre kümeleme analizi sonuçları

Kümeleme analizi sonucunda elde edilmiş olan gruplarda yer alan illerin doğru kümelenip kümelenmediklerinin araştırılması aşamasında Diskriminant (Ayırma) analizinden yararlanılmıştır. Bu analiz SPSS paket programında yapılmış ve özet tabloları ve sonuçları Ek-2’de verilmiştir. Elde edilen sonuçlara göre kümeleme analizi sonucunda elde edilen gruplara illerin atanmasında yüksek düzeyde başarı sağlanmıştır. Ayırma analizinde kullanılan değişkenlerden X_{12} (1-4 yaş aşılama oranı) ve X_{22} (45-64 yaş orantılı ölüm hızı) değişkenlerinin illeri ayırmada (kümelemede veya diskriminasyonda) etkili olmadıkları diğer değişkenlerin ise önemli düzeyde diskriminantı sağladıkları saptanmıştır. Diskriminant Analizi sonucunda değişkenlerin modele katkılarına ilişkin Wilk’s Lambda değerleri ve anlamlılık seviyeleri Ek-3’de verilmiştir. Bu sonuçlara göre X_{12} ve X_{22} değişkenleri her ne kadar ayırma analizinde önemli bir değişken olarak görünmese de yapılan faktör analizi sonucunda bu değişkenlerin de göz ardı edilemeyecek yüklerinin olduğu sonucu elde edilmiştir. Bu değişkenlerin kümeleme analizinde il gruplarının oluşmasında önemli rollerinin olduğu saptanmıştır. Ayırma analizi sonucunda değişkenlerin kümelere ait ortalamaları ve genel ortalamaları hesaplanmış ve Ek-4’de verilmiştir. Bu ortalamalara göre kümelerin oluşmasına ait ifadeler kullanılabilir. Bu ifadeler okuyucuya bırakılmıştır. Örneğin, 10000 kişiye düşen uzman hekim sayısı değişkeni X_1 ’e bakılacak olursa 5. kümede bu değişkene ait ortalamanın Genel (Türkiye) ortalamasına göre oldukça yüksek olduğu görülecektir ki bu da İstanbul, Ankara, İzmir ve Eskişehir illerinde uzman hekim sayısının (9.8) oldukça yüksek olduğu buna karşın 8. Kümede yer alan Hakkari, Kars ve Ardahan illerinde ise uzman hekim sayısının oldukça düşük (1.03667) olduğu net bir şekilde görülmektedir. İllere ait kümelerin tanımlanmasında, birlikte seçiciliğe sahip değişkenlerin yapısal nitelikleri göz önünde tutulmalıdır. SPSS istatistik paket programında; Ward Kümeleme Yönteminde Karesel Öklid Uzaklığı kullanılarak değişkenlerin kümelenmesi de yapılmıştır. Ayrıca MINITAB istatistiksel paket programında korelasyon matrisi ölçümü kullanılarak değişkenlere göre Ward Kümeleme Yöntemi yapılmış ve elde edilen sonuçların birbirinin aynısı olduğu görülmüştür. Sonuçlar Ek-5 ve Ek-6 da verilmiştir. Bu sonuçlar incelenildiğinde değişkenlerin üç ana

grupta ve yedi alt grupta kümelandikleri görülebilir. Bu kümelere giren deęişkenlerin illerin kümelenmesinde birlikte seçicilik özellięi gösterdikleri söylenebilir.

9. SONUÇ VE ÖNERİLER

Kümeleme Analizi, n birimden elde edilen x_p değişkenlerini dikkate alarak birimlerin alt kümelere ve gruplara ayrılmasını sağlayan çokdeğişkenli bir analiz yöntemidir. Kümeleme Analizinde, değişkenler arası uzaklıklardan giderek birimlerin tüm değişkenler yönünden ait oldukları kümelere atanmaları amaçlanmaktadır.

Çok sayıda göstereyi birlikte analiz eden bir yaklaşım ve yöntemin benimsendiği bu çalışmada, Türkiye genelinde 79 ilin sağlık açısından homojen il grupları nüfus, sağlık, bazı ekonomik ve kültürel alanlardan seçilmiş 30 adet değişken kullanılarak belirlenmeye çalışılmıştır. Çalışmanın bu temel amacı yanında diğer araştırmacılara kısa, öz açıklayıcı bilgiler sunmak bir başka amacı oluşturmaktadır.

Çalışmada farklı kümeleme yöntemleri uygulanmış ve en uygun sonucu Ward Kümeleme yönteminin verdiği tespit edilmiştir. İllerin kümelenmesi sonucu küme içi homojen, kümeler arası heterojen 8 küme elde edilmiştir. Bu kümelerin oluşmasında değişkenlerin rolleri araştırılmış ve 30 değişkenden 28 tanesi illerin gruplaşmasında önemli düzeyde ayırmayı sağlayıcı rolleri olduğu tespit edilmiştir. Kümeleme Analizi sonucu elde edilen kümelere illerin atanmasında yüksek düzeyde (%100) bir başarı sağlanmıştır.

İllerin gruplaştırılmalarında farklı kümeleme algoritmaları farklı sonuçlar vermektedir. Sonuçları itibari ile yaklaşık aynı sonuçları veren kümeleme algoritmalarında bazı birimlerin farklı kümelerde yer aldığı gözlenmiştir. Aynı kümeleme yönteminde farklı benzerlik ölçülerine göre de bazı birimlerin küme içi homojenliği sağlayamadığı ortaya çıkmaktadır. Bu tip durumlarda birimlerin küme üyeliklerinde bir bulanıklık söz konusu olmaktadır. Bulanık kümeleme (Fuzzy Clustering) yöntemi bu tip durumları tanımlamak için daha iyi donatılmıştır [54]. Bulanık kümeler, kümedeki birimin üyeliği olarak tanımlanan 0 ile 1 arasındaki her bir birimi belirleyen fonksiyonlardır. Birbirine çok benzeyen birimler aynı kümede yüksek üyelik ilişkisine göre yer alırlar [56]. Bundan dolayı Bulanık Kümeleme Yöntemi, birimlerin kümeye ya da kümelere ait olabilme katsayılarını hesaplar. Böylelikle birimin en yüksek üyelik katsayısına sahip olduğu kümeye atanacağından [54, 56, 57] kümeleme analizi çalışmalarında Aşamalı ya da Aşamalı olmayan kümeleme yöntemleri sonuçları

ile Bulanık Kümeleme yöntemleri sonuçları karşılaştırılabilir ve en uygun sonuçları veren kümeleme algoritmaları seçilebilir.

Kümeleme Analizi sonucu elde edilen m sayıda kümenin elemanlarının p değişken yönünden türdeşliğini ve farklı gruplarda türdeş olmayışlarını çokdeğişkenli varyans analizi kullanılarak Hotteling-Lawley iz ölçütleri ile oluşturulan küme sayısına bağlı olarak belirlenen küme sayısına göre değişimi incelenebilir [58].

KAYNAKLAR

1. DİNÇER, B., *İlçelerin Sosyo-Ekonomik Gelişmişlik Sıralaması*. DPT, Ankara, 1996.
2. DİNLER, Z., *Bölgesel İktisat*, Uludağ Üniversitesi, Bursa, 1991.
3. İllerin Sosyo-Ekonomik Gelişmişlik Düzeyinin Temel Bileşenler Analizi ile saptanması., KULAKOĞLU, Ö. Uzmanlık tezi., DİE,1995, 64.
4. Türkiye İllerinin Sağlık Düzeylerine İlişkin Bir Çalışma., KARAHAN, E. Biyoistatistik Bilim Uzmanlığı Tezi., Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara, 1995.
5. DİNÇER, B., ÖZASLAN, M., ve SATILMIŞ, E., *İllerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması*. DPT, Ankara,1996.
6. ÖZDAMAR, K., *Biyoistatistik ve Bilgisayar*. Anadolu Üniversitesi AÖF Yayınları, Eskişehir, 1997.
7. U NICEF., *Türkiye 'de Anne ve Çocukların Durum Analizi*. T.C. Hükümeti-UNICEF İş Birliği Programı, Ankara, 1996.
8. SÜMBÜLOĞLU, K., ve ÇELEBİOĞLU, A.T., *Sağlık İstatistiği*. Copy Center Yayınları, Ankara, 1992.
9. TEZCAN, S., *Türkiye 'de Bebek ve Çocuk Ölümleri*. H.Ü Tıp Fak. Halk Sağlığı Ana Bilim Dalı, Ankara, 1985.
- 10.SÜMBÜLOĞLU, K., *Sağlık Alanına Özel İstatistiksel Yöntemler*. Özdemir Yayıncılık, Ankara,1994.
- 11.TATLIDİL, H., *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Engin Yayınları, Ankara,1996.
- 12.ÖZDAMAR, K., *Anadolu Üniversitesi Fen Fakültesi Çokdeğişkenli İstatistiksel Analiz Ders Notları*, Eskişehir, 1998.
- 13.Kümeleme Çözümlemesinde Uygun Kümeleme Ölçütlerinin Karşılaştırılması., DİNÇER, K.S. Tıbbi Biyoloji Ana Bilim Dalı Biyoistatistik Bilim Dalında Doktora Tezi., Anadolu Üniversitesi, Sağlık Bilimleri Enstitüsü, 1992,128.
14. TATLIDİL, H., *Türkiye 'nin AT Üyeliği ve Beşeri Kalkınma Durumu*, Hazine Dergisi, Sayı 6,1997/2,53-74,1997.
- 15.ALPAR, R., *Uygulamalı Çokdeğişkenli İstatistiksel Yöntemlere Giriş-1*. Bağırgeçen Yayınevi, Ankara, 1997.
16. Çok Boyutlu Veriler'in Bazı İstatistiksel Analiz Yöntemleri ve Uygulamaları. EMİN, S.E., Doktora tezi., Gazi Üniversitesi, Fen Bilimleri Enstitüsü, 1984, 50.
17. KOUTSOYIANNIS, A., *Ekonometri Kuramı*. Çev: ŞENESEN, Ü., ŞENESEN, G., Teknik Üniversite Matbaası, İstanbul, 1992.

KAYNAKLAR (devam)

18. AKKAYA, Ş., *Ekonometri II*. Anadolu Matbaacılık, İzmir, 1991.
19. ÖZDAMAR, K., Hastalık Olgularının İncelenmesinde Kümeleme Çözümlemesinin Kullanılması. Anadolu Üniversitesi Yayınları, Eskişehir, 1988.
20. COX, K. K., ENIS, B. M., *The Marketing Research Process*. Pacific Palisades, GoodYear Publishing, Co. Inc., 1972.
21. İNAL, C., GÜNAY, S., *Olasılık ve Matematiksel İstatistik*. H.Ü. Fen Fakültesi Basımevi, Ankara, 1978.
22. KARA, İ., *Olasılığa Giriş*. E.İ.T.İ.A. Yayınları, Eskişehir, 1982.
23. Çok Değişkenli Verilerin Kümelenendirilmesi için İstatistiksel bir Yöntem., ÇETİNEL, B. Doktora Tezi., Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 1982, 72.
24. CHATFIELD, C., COLLINS, A. J., *Multivariate Analysis*. Chapman and Hall, Londra, 1980.
25. WEIERS, R. W., *Marketing Research*, Prentice-Hall International C., Londra, 1988.
26. ANDERBERG, M. R., *Cluster Analysis for Applications*. Academic Press, Inc., New York, 1973.
27. ALDENDERFER, M. S., BLASHFIELD, R. K., *Cluster Analysis*. Sage Publications, London, 1984.
28. HOWKINS, D. M., *Topics in Applied Multivariate Analysis*. Cambridge University Press., London, 1982.
29. Banka Müşterilerinin Demografik ve Sosyo-ekonomik Özellikler Bakımından Gruplandırılmasında Kümeleme Çözümlemesi ve bir Uygulama., YILDIZ, Z. Yüksek Lisans Tezi., Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir, 1989, 42.
30. EVERITT, B. S., DUNN, G., *Applied Multivariate Analysis*. John Wiley & Sons, Inc., New York, 1991.
31. EVERIT, B. S., *Cluster Analysis*. John Wiley & Sons Inc., New York, 1993.
32. VOGT, W., NAGEL, D., SATOR, H., *Cluster Analysis In Clinical Chemistry: A Model*. John Wiley & Sons, Inc., New York, 1987.
33. FERREGUT, C., Cluster Analysis as a Tool in the Classification of Structures for Safety Differentiation, Canadian Journal of Civil Engineering, 15, 4, 601-8, 1988.
34. SPSS, *Statistical Algorithms*, SPSS Inc., USA, 1991.
35. JOHNSON, R.A, WICHERN, D. W., *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., New Jarsey, 1988.

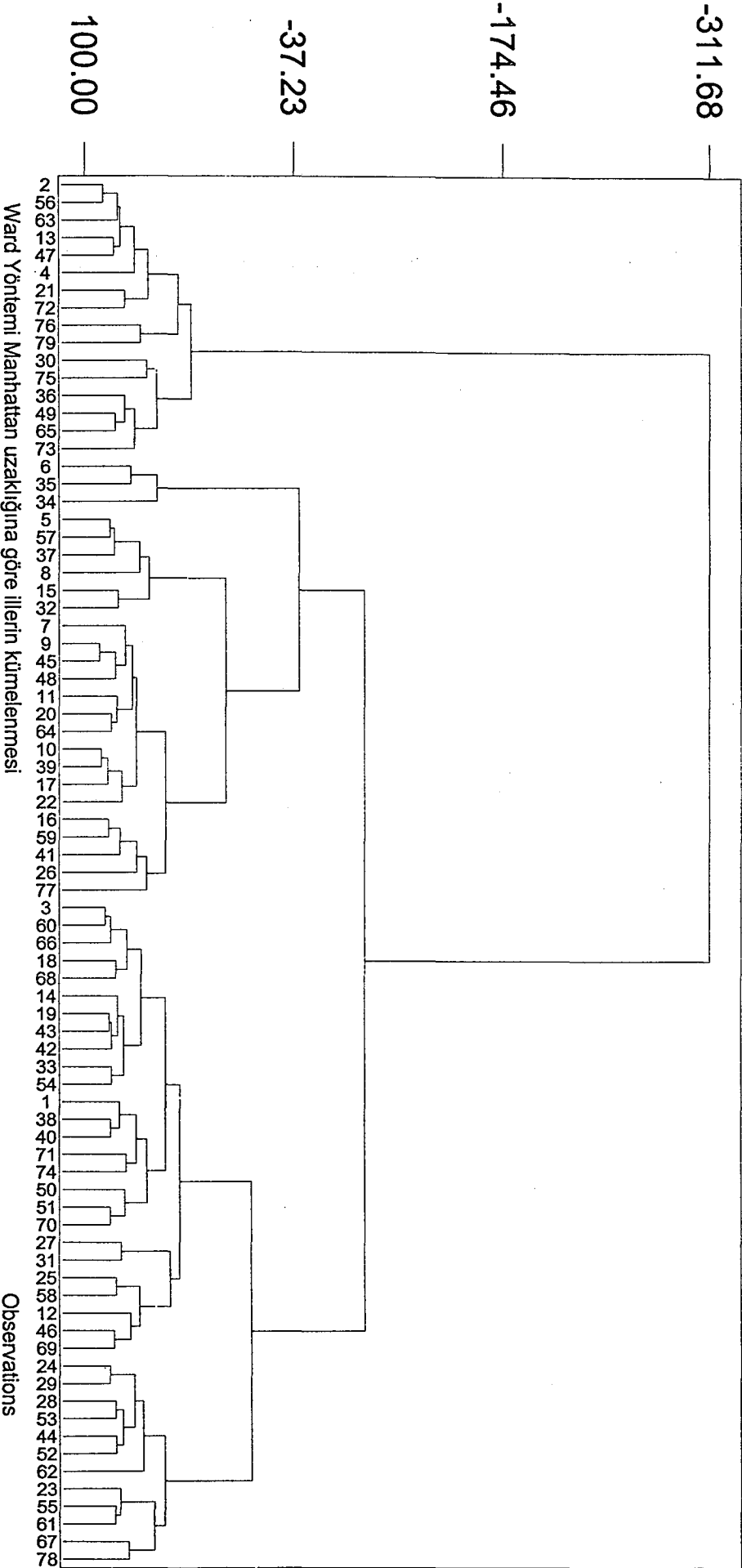
KAYNAKLAR (devam)

36. KURTULUŞ, K., *Pazarlama Araştırmaları*. İ. Ü. İşletme Fakültesi Yayınları, İstanbul, 1981.
37. BAĞCI, H., SHAREEF, S. R., ÖZDAMAR, K., *Bacllus Thuringiensis Varyetelerinin Sınıflandırılmasında Sayısal Taksonominin Uygulanması*, Doğa, 15, 70-81, 1991.
38. MARDIA, K. V, KENT, J. T., BIBBY, J. M., *Multivariate Analysis*. Academic Press, London, 1979.
39. MANLY, B. F. J., *Multivariate Statistical Methods*. Chapman & Hall, New York, 1994.
40. KENDALL, M., *Multivariate Analysis*. Charles Griffin & Company Ltd, London, 1980.
41. HARTUNG, J., ELPELT, B., *Multivariate Statistik*. Oldenbourg Verlag GmbH, München, 1995.
42. LORR, M., *Cluster Analysis for Social Scientists*. Jossey-Bass Publishers, London, 1983.
43. SNEATH, P. H. A., SOKAL, R. R., *Numerical Taxonomy*. W. H. Freeman and Company, San Fransisco, 1973.
44. MATHIEU, R. G., GIBSON, J. E., *Methodology for Large-Scale R&D Planning Based on Cluster Analysis*, IEEE Transaction on Engineering Management, 40, 3, 283-91, 1993.
45. HAIR, J. F., ANDERSON, R. E, TATHAM, R. L., BLACK, W. C., *Multivariate Data Analysis with Readings*. Prentice-Hall, Int. Inc., New Jersey, 1995.
46. JAIN, A. K., DUBES, R. C., *Algorithms For Clustering Data*. Prentice-Hall, New Jersey, 1988.
47. HAIR, J. F., ANDERSON, R. E, TATHAM, R. L., *Multivariate Data Analysis with Readings (second edition)*. Prentice-Hall, Int. Inc., New Jersey, 1990.
48. AAKER, D. A, KUMAR, V., DAY, G. S., *Marketing Research*. John Wiley & Sons, Inc., New York, 1995.
49. SHARMA, S., *Applied Multivariate Techniques*. John Wiley & Sons, Inc., New York, 1996.
50. MURTAGH, F., HECK, A, *Multivariate Data Analysis*. D. Reidel Publishing Company, New York, 1987.
51. SÖZER, E., ÇELİK, Y., KUTSAL, A., *Kümeleme Çözümlemesinde Çekirdek Noktaların Seçimi*, İstatistik Sempozyumu'86., ODTÜ ve AÜ, 15-17 Eylül 1986.
52. MINITAB Release 11.12, Help, Minitab Inc., 1996.

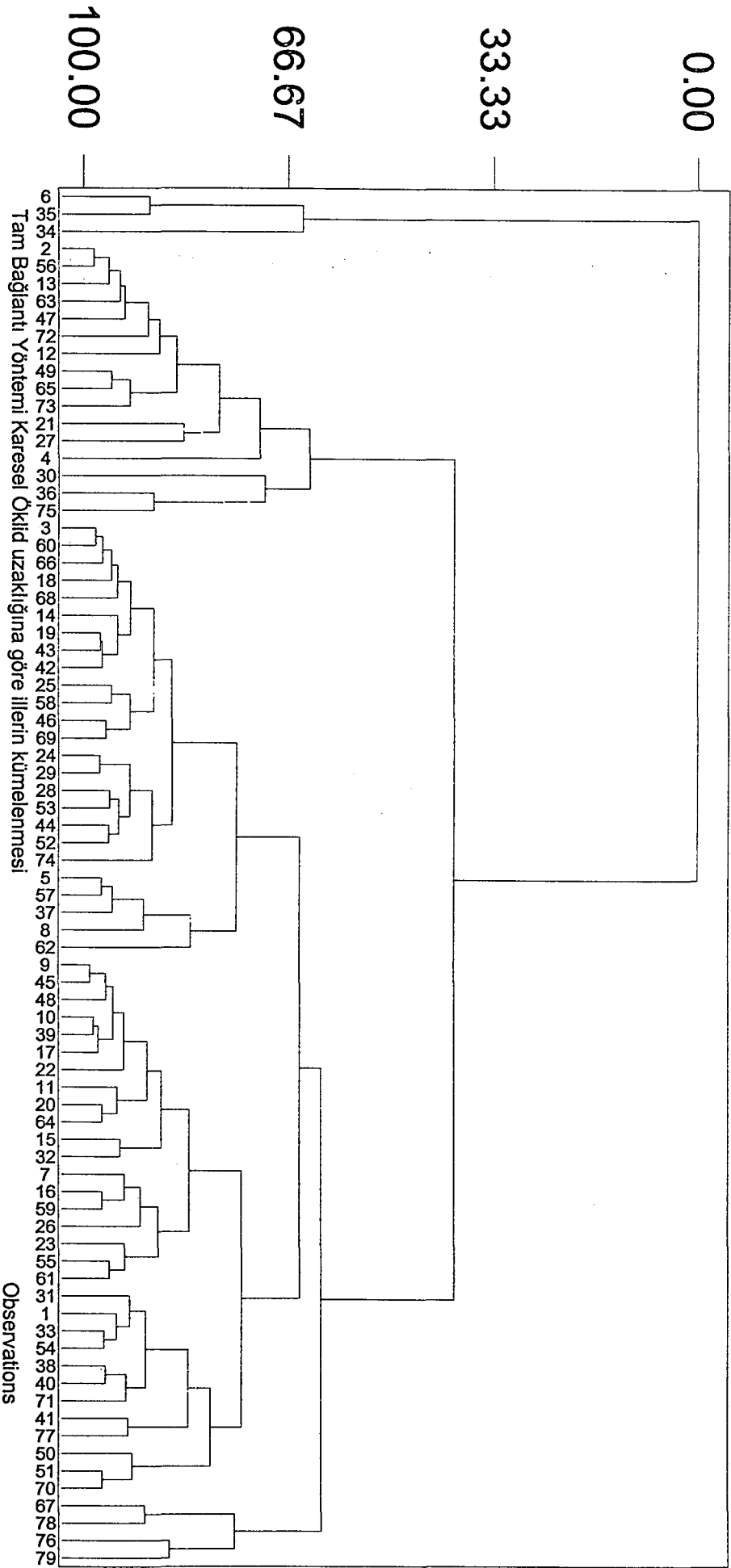
KAYNAKLAR (devam)

53. UYSAL, M., SÖZER, E., Ülkelerin Sosyo-Ekonomik Gelişmişlik Göstergelerine Göre Gelişme Düzeylerinin Saptanması, I. Ulusal Ekonometri ve İstatistik Sempozyumu Bildirileri, Dokuz Eylül Üniversitesi, İzmir, 1993.
54. KAUFMAN, L., ROUSSEEUW, P. J., *Finding Groups In Data; An Introduction to Cluster Analysis*, John Wiley & Sons, Inc., New York, 1990.
55. TATLIDİL, H., BİLEN, L., *İller İtibariyle Türkiye 'de Gelişmişlik Düzeyinin Çeşitli İstatistiksel Yöntemlerle Belirlenmesi*. TC. Başbakanlık Hazine Müsteşarlığı Araştırma ve İnceleme Dizisi, Ankara, 1996.
56. BEZDEK, J. C., PAL, S. K., *Fuzzy Models For Pattern Recognition; Methods That Search For Structures In Data*. IEEE Press, New York, 1992.
57. TERANO, T., ASAI, K., SUGENO, M., *Fuzzy Systems Theory and Its Applications*. Academic Press. Inc, New York, 1991.
58. DİNCER, K. S., ÖZDAMAR, K., Kümeleme Çözümlemesinde Uygun Kümeleme Ölçütlerinin Karşılaştırılması, Hacettepe Fen ve Mühendislik Bilimleri Dergisi, 14, 17-33,1993.

Similarity



Similarity



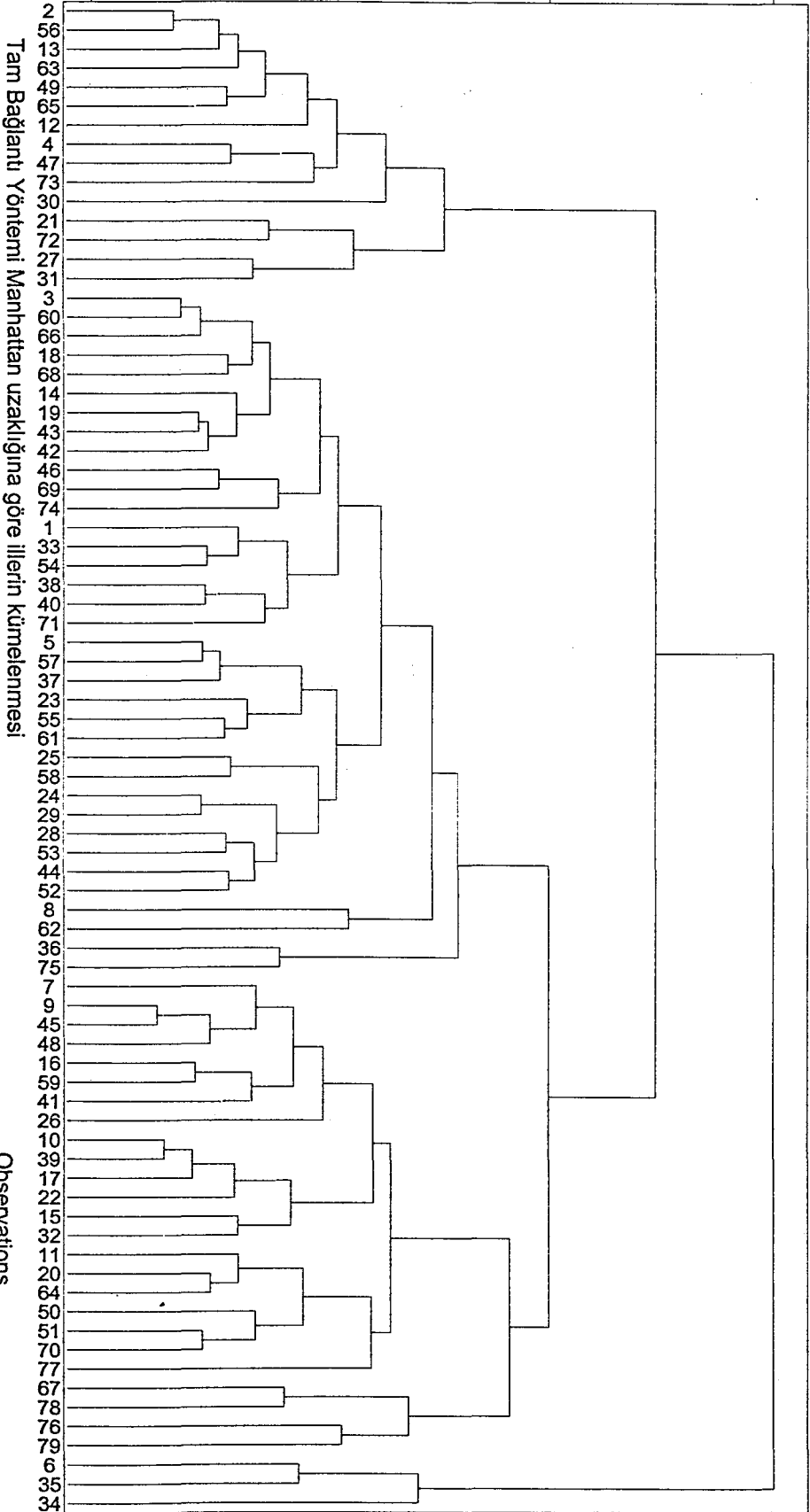
Similarity

0.00

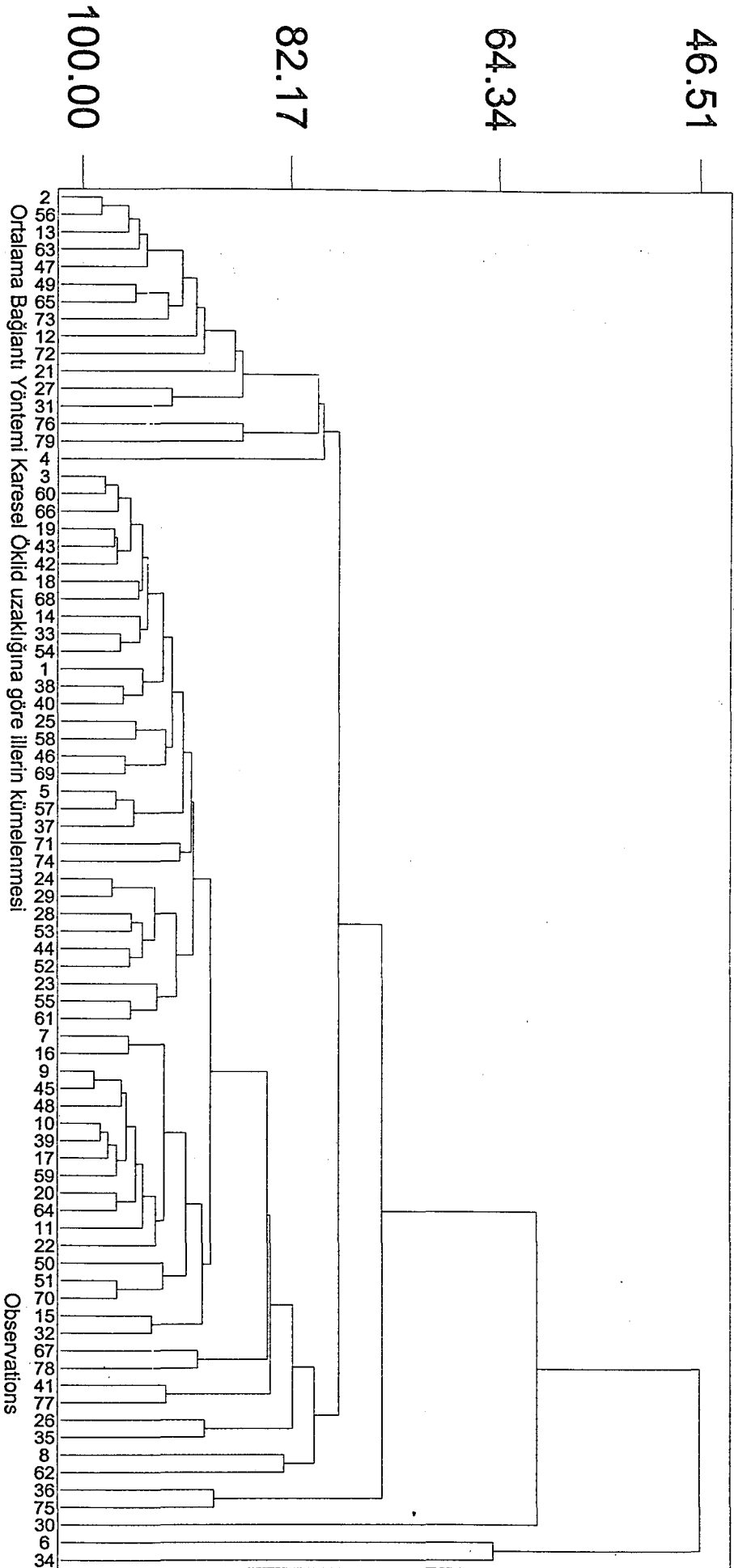
33.33

66.67

100.00



Similarity



Classification results -

Actual Group	Cases	No. of Predicted Group Membership							
		1	2	3	4	5	6	7	8
Group 1	22	22 100,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%
Group 2	11	0 ,0%	11 100,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%
Group 3	3	0 ,0%	0 ,0%	3 100,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%
Group 4	6	0 ,0%	0 ,0%	0 ,0%	6 100,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%
Group 5	4	0 ,0%	0 ,0%	0 ,0%	0 ,0%	4 100,0%	0 ,0%	0 ,0%	0 ,0%
Group 6	18	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	18 100,0%	0 ,0%	0 ,0%
Group 7	12	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	12 100,0%	0 ,0%
Group 8	3	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	0 ,0%	3 100,0%

Percent of "grouped" cases correctly classified: 100,00%

Wilks' Lambda değerleri ve anlamlılık seviyeleri

Wilks' Lambda (U-statistic) and univariate F-ratio
with 7 and 71 degrees of freedom

Variable	Wilks' Lambda	F	Significance
X1	,25710	29,3088	,0000***
X2	,38238	16,3830	,0000***
X3	,26462	28,1867	,0000***
X4	,24093	31,9563	,0000***
X5	,47627	11,1538	,0000***
X6	,17697	47,1720	,0000***
X7	,53619	8,7735	,0000***
X8	,36585	17,5813	,0000***
X9	,59137	7,0085	,0000***
X10	,41061	14,5590	,0000***
X11	,58001	7,3445	,0000***
X12	,93914	,6574	,7071 ^{ns}
X13	,47772	11,0888	,0000***
X14	,50750	9,8429	,0000***
X15	,58725	7,1291	,0000***
X16	,20061	40,4175	,0000***
X17	,59305	6,9599	,0000***
X18	,63158	5,9167	,0000***
X19	,74386	3,4926	,0028**
X20	,45129	12,3322	,0000***
X21	,52406	9,2114	,0000***
X22	,87039	1,5104	,1778 ^{ns}
X23	,40242	15,0621	,0000***
X24	,40503	14,8995	,0000***
X25	,40860	14,6807	,0000***
X26	,48647	10,7070	,0000***
X27	,44566	12,6164	,0000***
X28	,44495	12,6527	,0000***
X29	,58446	7,2114	,0000***
X30	,61158	6,4419	,0000***

ns : P>0.05

* : P<0.05

** : P<0.01

*** : P<0.001

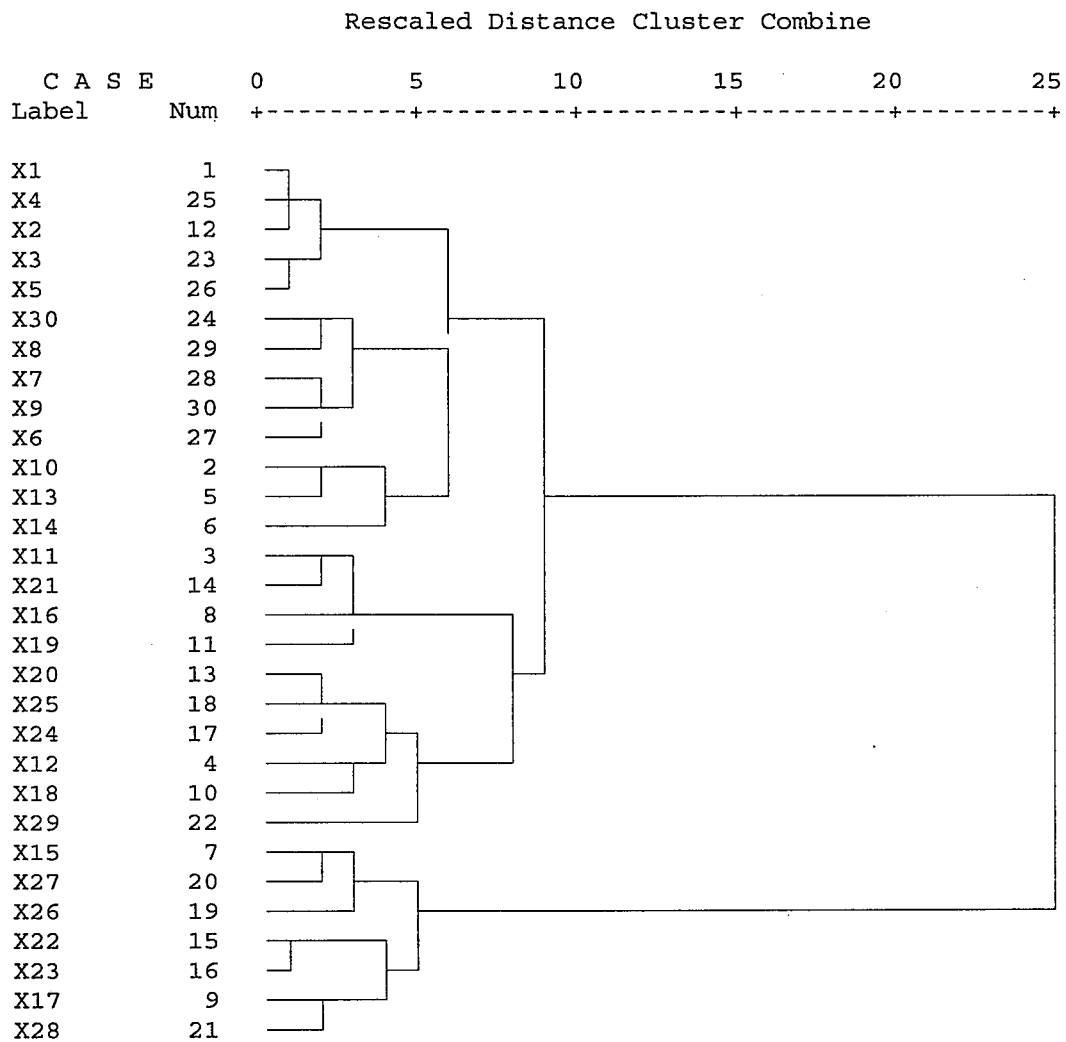
KÜME	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
1	2,2759	4,6550	,9486	7,8764	6,1205	2,064	5,6291	7,3927	52,691	17,259
2	,8881	3,0700	,3072	4,2646	3,8236	1,201	3,7391	4,8936	36,782	9,8727
3	,7500	3,4700	,3966	4,6200	6,2733	1,3633	2,7100	5,1667	36,267	8,3667
4	2,9050	6,3416	1,0616	10,308	12,220	2,3317	9,2283	13,115	43,550	32,067
5	9,8000	10,9850	3,2525	24,035	5,6450	5,2950	6,6100	13,880	64,475	35,125
6	3,5700	5,4661	1,7711	10,804	9,1656	3,1556	6,2550	10,684	52,650	15,855
7	2,4066	5,3833	,9250	8,7142	8,5083	2,0342	5,6950	10,023	52,400	23,242
8	1,0366	4,3966	,4033	5,8433	7,6133	,8633	4,6100	6,3200	28,700	11,033
Genel	2,7211	5,1235	1,1267	8,9699	7,3589	2,2998	5,6920	8,8319	48,752	18,275

KÜME	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀
1	86,500	68,955	18,455	39,500	34,818	59,223	84,968	7,7373	75,455	21,036
2	65,364	70,091	14,000	45,000	37,818	140,345	70,018	22,354	60,764	26,345
3	67,000	81,667	6,0000	105,667	16,000	123,200	53,267	9,226	49,667	11,633
4	87,333	76,167	23,667	45,000	35,000	58,950	83,800	-7,377	40,283	11,267
5	88,250	69,250	11,750	38,750	20,250	37,775	91,375	22,350	523,275	9,875
6	89,555	76,167	16,833	40,500	25,444	43,139	86,850	16,978	121,105	11,022
7	86,833	70,167	11,000	32,833	17,500	88,700	68,025	-5,374	88,058	11,258
8	69,333	82,667	19,333	53,667	59,333	965,467	78,400	-16,713	29,033	30,500
Genel	83,063	72,506	15,949	42,911	29,962	107,068	79,524	8,606	102,986	16,704

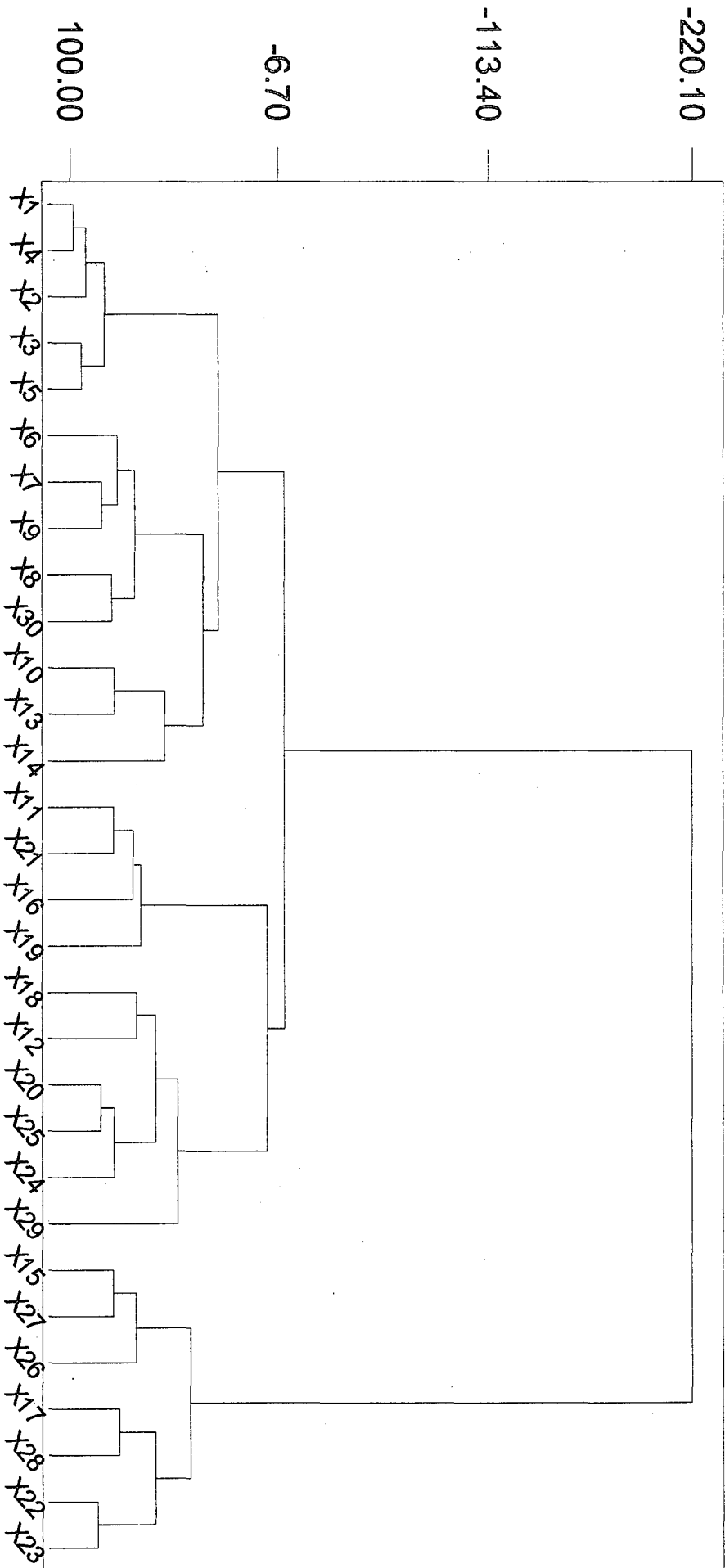
KÜME	X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₂₅	X ₂₆	X ₂₇	X ₂₈	X ₂₉	X ₃₀
1	9,609	20,063	49,091	2,892	1160079	28,841	14,505	30,805	73,318	40,650
2	13,436	20,345	38,882	1,246	602163	39,218	20,964	18,826	49,645	29,190
3	16,633	17,200	54,567	0,503	722013	27,533	15,867	17,670	58,133	35,233
4	7,900	19,850	61,000	4,195	1301900	18,883	10,150	28,478	59,583	49,233
5	11,400	24,675	54,075	2,288	2296297	30,950	14,875	52,862	81,325	20,025
6	7,894	20,388	60,672	3,802	2161091	25,500	12,394	42,854	70,711	56,956
7	8,300	21,125	59,300	1,963	1202161	25,408	14,242	15,739	51,175	31,308
8	13,167	20,833	35,433	1,113	430292	32,400	24,200	13,747	46,267	44,933
Genel	9,915	20,476	52,844	2,639	1340818	28,439	14,991	29,388	64,962	40,915

EK-5

Dendrogram using Ward Method Squared Distance Z Scores



Similarity



Variables