

**T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

KÜMELEME ANALİZİ VE UYGULAMASI

YÜKSEK LİSANS TEZİ

Nazmiye YALÇIN

(101133103)

Anabilim Dalı: İstatistik

Programı: Olasılık Teorisi ve Olasılık Süreçleri

Danışman: Doç. Dr. Mahmut IŞIK

Tezin Enstitüye Verildiği Tarih: 8 Ocak 2013

OCAK-2013

KÜMELEME ANALİZİ VE UYGULAMASI

Nazmiye YALÇIN

**Yüksek Lisans Tezi
İstatistik Anabilim Dalı
Danışman: Doç. Dr. Mahmut IŞIK
OCAK-2013**

**T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

KÜMELEME ANALİZİ VE UYGULAMASI

YÜKSEK LİSANS TEZİ

Nazmiye YALÇIN

(101133103)

**Tezin Enstitüye Verildiği Tarih : 8 Ocak 2013
Tezin Savunulduğu Tarih : 24 Ocak 2013**

**Tez Danışmanı : Doç. Dr. Mahmut IŞIK (F.Ü)
Diğer Jüri Üyeleri : Doç. Dr. Sinan ÇALIK (F.Ü)
Yrd. Doç. Dr. Reşat YILMAZER (F.Ü)**

OCAK-2013

ÖNSÖZ

Tez konusunun belirlenmesi ve yürütülmesi aşamasında, her türlü yardımı ve desteğini esirgemeyen kıymetli danışman hocam Doç. Dr. Mahmut IŞIK'a, tezimdaki verileri elde etmemde yardımcı olan ve kullanmama izin veren Fırat Üniversitesi Tıp Fakültesi öğretim üyesi Prof.Dr. Yasemin AÇIK'a, çalışmamda yardımlarını esirgemeyen değerli arkadaşım Arş.Gör. Adem DOĞANER'e teşekkür eder, saygılarımı sunarım.

Nazmiye YALÇIN

ELAZIĞ- 2013

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	II
İÇİNDEKİLER.....	III
ÖZET.....	IV
SUMMARY.....	V
ŞEKİLLER LİSTESİ.....	VI
TABLolar LİSTESİ.....	VI
SEMBOLLER LİSTESİ.....	VI
1.GİRİŞ	
1.1. Temel Tanımlar	1
1.2. Amaç.....	5
2. MATERYAL VE METOD.....	8
2.1. Kümeleme Analizinin İşlevleri.....	8
2.2. Kümeleme Yöntemleri.....	8
2.2.1. Hiyerarşik Yöntemler.....	9
2.2.2. Hiyerarşik Olmayan Yöntemler.....	15
2.3. Kümeleme Analizinin Aşamaları.....	21
2.3.1. Değişkenlerin Standardizasyonu ve Dönüştürülmesi.....	22
2.3.2. Benzerlik ve Farklılık Ölçümlerinin Seçimi.....	23
2.4. Sayıtlar ve Analizin Diğer Gereklilikleri.....	28
2.5. Kümelerin Değerlendirilmesi.....	30
2.6. Araştırma Deseni.....	31
3. UYGULAMA.....	33
3.1. Uygulamanın Amacı.....	33
3.2. Uygulama Verileri ve Düzeni.....	33
3.3. Örneklem Seçimi.....	37
3.4. Gürültülü İş Yerlerinde İşitme Kaybı.....	37
3.5. Odyometrik Ölçüm.....	39
3.6. Sesin Şiddeti Nasıl Ölçülür.....	40
3.7. Kümeleme Yöntemlerinin Belirlenmesi.....	41
3.8. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların Beşli Sınıflandırılması.....	41

3.9. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların İkili Sınıflandırılması.....	41
4. BULGULAR.....	48
4.1. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların Beşli Kümelenmesi.....	48
4.2. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların İkili Kümelenmesi.....	48
5. SONUÇLAR VE TARTIŞMA.....	50
6. ÖNERİLER.....	52
KAYNAKLAR.....	53

ÖZET

Kümeleme analizi (cluster analysis) grupları kesin olarak bilinmeyen, birimleri, değişkenleri birbiriyle benzer alt kümelere (grup, sınıf) ayırmaya yardımcı olan çok değişkenli istatistiksel analiz yöntemlerinden biridir.

Bu çalışma genel olarak iki bölümde oluşmaktadır. Birinci bölümde kümeleme analizi ve kümeleme tekniklerine ilişkin bilgiler sunulmuştur. İkinci bölüm olan uygulamada ise Ward's Kümeleme yöntemi ile, Elazığ ili organize sanayi bölgesinde (OSB) bulunan iş yerlerinde, gürültü düzeyi 95 dB'in üstünde olan ve en az on işçinin çalıştığı 6 fabrikada, çalışan işçilerden oluşan işitme kaybına göre fabrikalar sınıflandırılmıştır.

Anahtar Kelimeler: Kümeleme Analizi, Odyometrik Ölçüm, Gürültü

SUMMARY

Groups of cluster analysis help to discriminate absolutely unknown units, variables to similar sub cluster and it is one of the multivariable statistical analysis method.

This study is generally consist of two section. In the first section respected informations about cluster analysis and cluster techniques. In the second section application is classified workers hearing loss of 6 factories has minimum ten workers and noise level over 95 dB of Elazığ Organized Industrial (OSB) using Ward's cluster method.

Key Words : Cluster Analysis, Audiometric Measurement, Noise

ŞEKİLLER LİSTESİ

	<u>Sayfa No</u>
Şekil 1.1. Küme İçerisi ve Kümeler Arası Uzaklık.....	1
Şekil 1.2. Aynı Noktalardan Oluşan Bir Setin Değişik Yollarla Kümelenmesi.....	6
Şekil 2.1. Tek Bağlantı Yöntemi.....	11
Şekil 2.2. Tam Bağlantı Yöntemi.....	12
Şekil 2.3. Ortalama Bağlantı Yöntemi.....	13
Şekil 2.4. Ward's Bağlantı Kümeleme Yöntemi.....	14
Şekil 2.5. İki Boyutlu Noktalar Kullanılarak Gösterilen Değişik Türdeki Kümeler	20
Şekil 2.6. Kümeleme Analizi Akış Şeması.....	31
Şekil 3.1 Ward's Yöntemiyle Yapılan Beşli Kümeleme Analizinin Dendogram Grafiği..	43
Şekil 3.2 Ward's Yöntemiyle Yapılan Beşli Kümeleme Analizi.....	44
Şekil 3.3 Ward's Yöntemiyle Yapılan İkili Kümeleme Analizi.....	46
Şekil 3.4 Ward's Yöntemiyle Yapılan İkili Kümeleme Analizinin Dendogram Grafiği...	47

TABLÖLAR LİSTESİ

	<u>Sayfa No</u>
Tablo 3.1 Odyometrik Verilerin Kategorize Edilmesi.....	34
Tablo 3.2. Fabrikalara Göre Odyometri Ölçümlerinin Ortalamaları.....	36
Tablo 3.3 Ward's Yöntemi İle Oluşturulan Beşli Birleştirme Tablosu.....	42
Tablo 3.4 Ward's Yöntemi İle Oluşturulan Beşli Kümeleme.....	42
Tablo 3.5 Ward's Yöntemi İle Oluşturulan İkili Birleştirme Tablosu.....	45
Tablo 3.6 Ward's Yöntemi İle Oluşturulan İkili Kümeleme.....	45
Tablo 4.1 Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların Beşli Kümelenmesi.....	48
Tablo 4.2. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların ikili Kümelenmesi.....	49

SEMBOLLER LİSTESİ

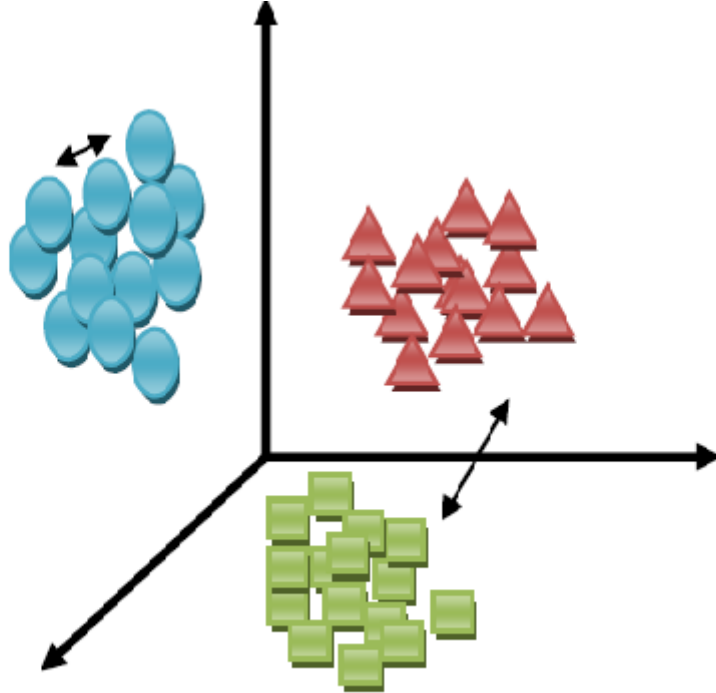
OSB	: Organize Sanayi Bölgesi
Hz	: Hertz
dB	: Desibel

1.GİRİŞ

1.1. Temel Tanımlar

Kümeleme analizi farklı yapıdaki verilerin küme yapısını ve küme sayısını araştırır. Kümeleme analizi grupları kesin olarak bilinmeyen, birimleri, değişkenleri birbiriyle benzer alt kümelere (grup, sınıf) ayırmaya yardımcı olan çok değişkenli istatistiksel analiz yöntemlerinden biridir. Kümeleme analizinin temel amacı birimleri sahip oldukları karakteristik özellikleri temel alarak gruplandırmaktır.

Kümeleme analizini, Hair ve arkadaşları; temel amacı objeleri sahip oldukları özelliklere göre gruplara ayırmak olan bir grup çok değişkenli teknik olarak tanımlamaktadırlar [1]. Kümeleme analizinin esas amacı; birimlerde bulunan karakteristik özellik dikkate alınarak gruplandırıldığı için; küme içerisinde çok benzer biçimde, kümeler arasında farklı olacak şekilde kümelenir. Kümeleme işlemi başarılı olursa geometrik çizim yapıldığında birimler küme içerisinde çok yakın, kümeler ise birbirinden uzak olacaktır. “Şekil 1.1”



Şekil 1.1. Küme içerisi ve kümeler arası uzaklık.

Kümeleme analizi için değişik tanımlar yapılmıştır. Aldenderfer ve Blashfield'e göre kümeleme analizi, sınıflandırma oluşturmak için kullanılabilen çok sayıda işlemi ifade eden kapsamlı (genel) bir kavramdır. Bu işlemler ampirik olarak yöntemin oluşturduğu gruplar ya da kümelerden gelir. Daha net bir ifadeyle kümeleme yöntemi, birimlerden oluşan bir örneklem hakkında bilgi içeren veri setleri ile başlayan ve bu birimleri benzer (homojen) gruplar şeklinde tekrar düzenlemeyi sağlayan çok değişkenli bir istatistiksel yöntemdir [2].

Kümeleme analizi farklı yapıda verilerin küme yapısını ve küme sayısını araştırır. Bu analiz; grupta yapısını bulurken küme içindeki gözlemlerin aynı yapıda, kümeler arasındaki gözlemlerin ise farklı yapılarda olmasını amaçlar. Gözlemler için bu ayrıştırmalar benzerlik ve farklılık ölçütleri kullanılarak yapılır. Bu ölçütler genellikle uzaklık ölçütlerine dayandırılarak bulunur. Bazı durumlarda korelasyon da kullanılır. Kümeler için sınıflandırma yapılırken gözlemlere ait grafiklerden de yararlanılabilir. İki değişkenli olması durumunda gözlemlere ait saçılım grafiği oluşturulabilir. İki değişkenden fazla veri olması durumunda ise temel bileşenler analizi kullanılarak iki değişkene indirgenir ve grafik çizilebilir.

Kümeleme analizi, hemen hemen tüm bilim dallarında kullanılan bir yöntemdir. Örneğin; Tıp, biyoloji, psikoloji, sosyoloji, arkeoloji gibi belirsizlik koşullarının ve karmaşık oluşumların bulunduğu bilim alanlarında ise daha yoğun olarak yararlanılan bir yöntemdir. Örneğin, tıp alanında; hastalıkların sınıflandırılması, psikiyatride; paranoya, şizofreni gibi semptomların doğru sınıflandırılması (teşhis profilleri oluşturması), laboratuvar bulguları ile klinik bulguların oluşturduğu veri matrislerinden hastalık alt gruplamalarının ya da yeni semptomların tanımlanması gibi amaçlarla kümeleme analizinden yararlanılmaktadır [3].

Kümeleme analizi 20. Yüzyılda fark edilmesine rağmen bununla ilgili literatür son yıllarda elde edilmiştir. Bu metodun gelişiminde önemli çalışmalardan Robert Sokal ve Peter Sneath adlı iki biyolog tarafından 1963 yılında yayınlanan 'Principles of Numerical Taxonomy' adlı kitap olmuştur,[4]. Sokal ve Sneath bu kitaplarında biyolojik sınıflandırmaların üretimi için etkili prosedürler araştırılmıştır. Böylece organizmalar arasındaki benzerlik derecesi ortaya konacak ve kümeleme analizinin kullanımı ile nispeten benzer organizmalar aynı gruplara yerleştirilecektir. Kümeleme analizi bu çalışmadan sonra rağbet görmüştür. 1963 ile 1975 yılları arasında bu metoda ilişkin çalışma sayısı katlanarak artmıştır. Bilgisayar teknolojisinin gelişmesi ile kümeleme analizi

yöntemlerinde ve uygulanabilir alanlarda gelişmeler yaşanmıştır. Farklı kümeleme teknikleri ortaya sunulmuş ve farklı disiplinlerde kümeleme analizi yöntemleri uygulama sahası bulmuştur. Kümeleme analizi sosyal bilimlerin birçok bilim dallarında da kendisine yer bulmuştur [2].

Kümeleme analizi kendi içerisinde popüler bir alan olmasına rağmen, hala; faktör analizi, diskriminant analizi çok boyutlu ölçekleme gibi çok boyutlu istatistiksel yöntemlere göre kıyaslandığında çok bilinen bir yöntem olmadığı anlaşılmaktadır.

Literatürde kümeleme analizi için birçok algoritma öne sürülmüştür. Ancak kümeleme yöntemlerini genel olarak iki temel algoritma altında toplamak mümkündür. Bunlardan biri "hiyerarşik kümeleme yöntemleri" diğeri ise "hiyerarşik olmayan yöntemleridir."

Hiyerarşik kümeleme yöntemleri özellikle küçük örneklemeler (tipik olarak $n < 250$) için uygundur. Hiyerarşik kümeleme analizi yapmak için araştırmacılar, benzerlik ya da uzaklığın nasıl tanımlanacağına ve kümelerin nasıl birleşeceğine ya da ayrılacağına karar vermek durumundadırlar [1]. Hiyerarşik kümeleme yöntemlerinin, araştırmacının incelediği veri setinde kaç grup bulunduğunu başlangıçta bilmediği durumlarda çok uygun bir yöntem olduğunu belirtilmektedir. Ayrıca bu yöntemin araştırmacılara, inceledikleri veri setinde daha önce gözlemlenmemiş ilişkileri gözleme ve ilkeleri keşfetme olanağı vermesi nedeniyle de faydalı olduğunu belirtmektedir. Hiyerarşik olmayan kümeleme yöntemleri, küme sayısının belirlenmiş olduğu durumlarda kullanılır. Bir başka deyişle araştırmacının küme sayısı konusunda ön bilgisinin olduğu durumlarda ya da küme sayısına karar verilmiş olması durumunda kullanılması önerilmektedir [3].

Bugün yeryüzünde milyonlarca farklı canlı türünün yaşadığı tahmin edilmekte ve yine benzer şekilde, bir tür içerisinde de milyonlarca farklı canlı yer almaktadır. İşte sayıları milyarları aşan canlıları tek tek gözlemek, incelemek ve tanımlamak mümkün olmadığından, bilim adamları çalışmalarını kolaylaştırmak ve birliktelik sağlanmak gibi nedenlerle, canlı türlerini benzerliklerine göre daha az sayıda gruba indirgeme yoluna gitmişlerdir. Burada temel amaç, canlıları bir sisteme oturtarak (sistemati olarak), doğayı daha anlaşılır hale getirmektir. Bu kaygı elbette sadece canlı türleri ya da canlılarla uğraşan disiplinler için söz konusu değildir. Bütün bilimler kendi çalışma alanlarını yapılandırmayı sağlayan sınıflandırmalar oluşturmakla ilgilenmişlerdir. Biyoloji alanında, yaşayan tüm canlıları sınıflamak için taksonomi oluşturmak, psikolojide kişilik ya da diğer psikolojik özelliklere göre sınıflandırmalar yapmak, pazar araştırmalarında pazarların

bölümlendirilmesine ilişkin analizler yapmak (segmentation analysis), ziraat alanında bitkileri, tıpta hastalıkları veya hastalıklara/semptomlara göre tedavileri sınıflamak gibi farklı disiplinlere ilişkin örnekler üretmek mümkündür [1].

Sınıflandırma en temel kavramsal yaklaşımlardan biridir. Çocuklar yaşamlarının erken dönemlerinde çevrelerindeki objeleri sınıflamayı ve kendi dillerindeki isimlerle ilişkilendirmeyi öğrenirler. Sınıflandırma bilim uygulamaları için de temel bir süreçtir; çünkü sınıflandırma sistemleri bilimde kuram geliştirme için gerekli kavramları içerir [2].

Kümeleme analizi, sınıflandırma oluşturmak için kullanılabilen çok sayıda işlemi ifade eden kapsamlı bir kavramdır. Bu işlemler ampirik olarak yöntemin oluşturduğu gruplar ya da kümelerden gelir. Daha net bir ifadeyle kümeleme yöntemi, birimlerden oluşan bir örneklem hakkında bilgi içeren veri setleri ile başlayan ve bu birimleri benzer (homojen) gruplar şeklinde tekrar düzenlemeyi sağlayan çok değişkenli bir istatistiksel yöntemdir [2], [1] Kümeleme analizin temel amacı objeleri sahip oldukları özelliklere göre gruplara ayırmak olan bir grup çok değişkenli teknik olarak tanımlamaktadır.

Kümeleme analizi, seçilen özellikler açısından birbirlerine diğerlerinden daha çok benzeyen bireyleri ya da objeleri bir kümede toplar. Böylece de küme içi homojenlikler, kümeler arası ise heterojenlikler maksimum kılınır. Eğer yapılan kümeleme işlemi başarılı, ise, geometrik gösterimde küme içinde yer alan nesneler birbirine yakın iken, farklı kümelerde yer alan nesneler birbirinden uzakta olacaklardır [1]. Kümeleme analizi araştırmacının belirlediği ölçütlere göre tüm gözlemler, “indirerek verileri azaltmakta ya da üst grup niteliğinde düşünülebilecek grup oluşturmaktadır”.

Kümeleme analizinin genel amacı, belirli özelliklerine göre birimlerin benzerliklerini ortaya koymak ve bu benzerlikleri esas alarak birimleri doğru kategorilere sınıflandırmaktır. Analizin bu yönü, çok değişkenli analiz tekniklerinden biri olan diskriminant analizine benzemektedir. Ancak kümeleme analizinde, diskriminant analizinin tersine birimlerin anlık durumu gözlemlendiği için, geleceğe yönelik tahminler yapmak mümkün olamamaktadır. Diskriminant analizi ise, daha çok yeni bir gözlemin önceden bilinen bir gruba uygun bir şekilde dahil edilmesi amacıyla kullanılmakta ve geleceğe göre öngörülerde bulunulmasına izin vermektedir [5]. Doğal sınıflamaları hakkında açıkça bilgi bulunmayan durumlarda, alt grupların yapılarını belirlemede kümeleme analizi kullanılırken doğa gruplamaları açıkça bilinen toplumlarda alt kümelerin irdelenmesi diskriminant analizi ile yapılır [6]. Faktör analizi ile karşılaştırıldığında da kümeleme analizinin objeleri gruplama, faktör analizinin ise değişkenle gruplama amacına

yönelik olduğu ifade edilebilir. Ayrıca faktör analizi gruplandırmayı verilerdeki değişmelere (varyans-kovaryansa) bağlı olarak yaparken, kümeleme analizi yakınlıklara (proximity) bağlı olarak yapar [1].

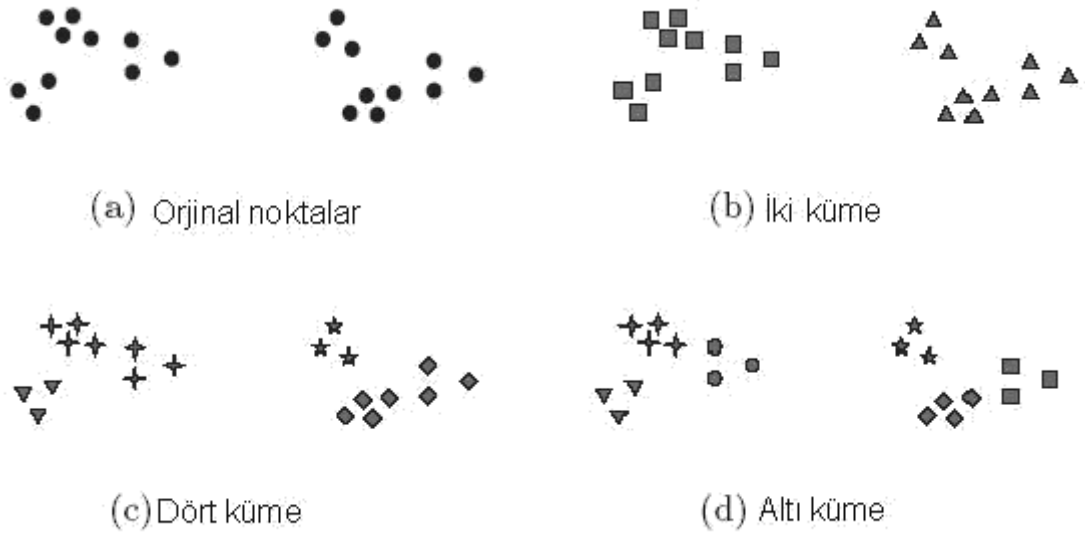
Popülaritesine karşılık kümeleme yöntemlerinin, faktör analizi, diskriminant analizi ve çok boyutlu ölçekleme gibi diğer bazı çok değişkenli istatistiksel yöntemlerle karşılaştırıldığında daha az anlaşıldığı ifade edilebilir. Sosyal bilimler alanında kümeleme ile ilgili çalışmalar incelendiğinde, çelişkili, tutarsız ve anlaşılması güç terminoloji, yöntem ve yaklaşımlar kullanıldığı görülebilir. Bu durum konuyu daha da karmaşık hale getirmektedir [2].

1.2. Amaç

Birimler hakkında düzenli bilgiler verilmesi için birimleri gruplamak ya da sınıflandırmak gerekmektedir. Sınıflandırma pek çok özellik açısından yakınlık ya da benzerlik gösteren birimleri gruplandırmaktır. Örneğin sadece iki değişkene göre incelendiğinde özellikleri birbirinden farklı olan birimlerin, değişken sayısı arttırıldığında birbirlerine benzedikleri ve bu benzer özelliklerinden dolayı bir sınıfta yer aldıkları görülebilir.

Kümeleme analizi veri nesnelerini yalnızca nesneleri tanımlayan ve ilişkilerini ortaya koyan verilerden çıkarılacak bilgiler ışığında gruplar. Amaç aynı grup içerisindeki nesnelerin birbirine benzer veya ilişkili olması; farklı gruptakilerin ise birbirinden farklı olması yada ilişkilerinin bulunmamasıdır. Aynı gruptakilerin birbirine benzeme oranı yada farklı gruptakilerin ise birbirinden farklı olma oranları kümelemenin ne kadar iyi olduğunun yada kümelerin birbirlerinden ne kadar kesinlikle ayrıldıklarının göstergesidir.

Bir çok uygulamada, küme kavramı net bir şekilde tanımlanmamıştır. Bir kümeyi ortaya koyan şeylerin neler olduğuna karar verme güçlüğüne daha iyi anlayabilmek için, Şekil 1.2’i dikkate alınırsa; Bu şekil 20 değişik nokta ve bu noktaları kümelere ayırmak için üç farklı yol göstermektedir. İşaretlerin şekilleri küme üyeliklerini göstermektedir. [17]



Şekil 1.2. Aynı noktalardan oluşan bir setin değişik yollarla kümelenmesi

Şekil 1.2 (b) ve Şekil 1.2 (d) sırasıyla veriyi iki ve altı parçaya ayırır. Bununla beraber, iki tane büyükçe kümenin her birinin daha küçük üç alt kümeye net olarak bölünmesi yalnızca insanın görme sistemine ilişkin bir yanıltma olabilir. Aynı zamanda, şunu söylemek de mantıksız olmaz “noktalar dört tane küme oluşturur”, bu da Şekil 1.2 (c)’ de görülmektedir. Bu şekil bize bir küme tanımının kesin ve net olmadığını, en iyi kümelemenin de verinin doğasına ve arzu edilen sonuçlara bağlı olduğunu gösterir.

Kümeleme analizi; birimleri ve nesneleri düzenleyerek sınıflara ayıran çok değişkenli istatistik analizlerinden birisidir. Bu nesnelerin kümelenmesindeki amaç, birimlerin kendi içlerinde birbirine benzeyecek şekilde gruplanmasını sağlamaktır. Benzerlik gösteren birimleri bir araya toplamak ve bu grubun ortak özelliklerini incelemek ve tanımlamak, değişkenler arasındaki neden- sonuç ilişkilerini açıklamayı daha da kolaylaştırır [7].

Aldenderfer ve Blasfield (1984) kümeleme analizinin dört temel hedefe ulaşmada yararlı olduğunu belirtmektedirler:

1. Bir sınıflama geliştirmek,
2. Birimleri gruplama amacıyla kavramsal sınıflama cetvelleri geliştirmek,
3. Veri keşfi yoluyla hipotez üretmek,
4. Hipotez test etmek veya bir veri setinde başka yöntemlerle yapılmış olan gruplamaların varlığını belirlemek.

Özdamar (1999) ise kümeleme analizinin kullanım amaçlarını şöyle özetlemektedir:

1. n sayıda bireyi p değişkene ilişkin özelliklerine göre, kendi içinde olabildiğince homojen ve kendi aralarında da heterojen alt kümelere ayırmak,

2. p sayıda değişkeni, n sayıda bireyde saptanan değerlere göre ortak özellikleri açıkladığı varsayılan alt kümelere ayırmak ve ortak faktör yapıları ortaya koymak,

3. Hem bireyleri, hem de değişkenleri birlikte ele alarak n bireyi p değişkene göre ortak özellikli alt kümelere ayırmak,

4. p değişkene göre saptanan değerler bakımından bireylerin biyolojik ve taksonomik sınıflamasını ortaya koymak.

Kümeleme analizi yukarıda söz edilen genel amaçların dışında, bir takım özel amaçlar için de kullanılabilir [8].

- Gerçek tiplerin (cinslerin-ırkların) belirlenmesi,
- Model uydurmanın kolaylaştırılması,
- Gruplar için ön tahmin,
- Hipotezlerin testi,
- Veri yapısının netleştirilmesi,
- Veri indirgemesi (veriler yerine kümelerin değerlendirilmesi),
- Uç değerlerin (outliers) belirlenmesi.

2.MATERYAL VE METOD

2.1. Kümeleme Analizinin İşlevleri

Kümeleme analizi, birimleri benzerliklerine göre tasnif eden çok değişkenli istatistiksel yöntemler topluluğudur. Bu yöntemler belirli işlemleri bazı parametrelere göre gerçekleştirmektedir. Kümeleme tekniklerinin başlıca işlevleri;

- a. n sayıda birim ve p sayıda değişkenin olduğu varsayılın. n sayıdaki birimve p adet değişkenin belirlenen özelliklerine göre gruplar içerisinde benzer ve gruplar arasında farklı alt kümeler oluşturmak,
- b. Uygun bir uzaklık ölçüsünün belirlenerek birimler arası uzaklıkları belirlemek,
- c. Kümeleme yöntemleri ile birimleri alt gruplara ayırmak,
- d. Grupları ve elde edilen grafikleri analiz edip ve yorumlayarak hipotezleri test etmek.

2.2. Kümeleme Yöntemleri

Literatürde kümeleme analizi için birçok algoritma öne sürülmüştür. Ancak kümeleme yöntemlerini genel olarak iki temel algoritma altında toplamak mümkündür. Bunlardan biri "hiyerarşik kümeleme yöntemleri; diğeri ise "hiyerarşik olmayan yöntemler"dir. Aşağıda bu iki temel algoritma içerisinde yer alan yöntemlere de yer verilerek bir sınıflandırma sunulmaya çalışılmıştır.

/. *Hiyerarşik Yöntemler*

A Birleştirici / Toplamalı (Agglomerative) Yöntemler / Algoritmalar

- a) Bağlantı Teknikleri
 - Tek Bağlantı (En yakın komşuluk)
 - Tam Bağlantı (En uzak komşuluk)
 - Ortalama Bağlantı
- b) Varyans Teknikleri
 - Ward's Yöntemi (Ward's Hata Kareler Toplamı)
- c) Merkezileştirme Teknikleri
 - Medyan
 - Centroid

B. Ayrıcı / Ayrımlı / Bölünmeli (Divisive) Yöntemler/ Algoritmalar

- a) Bölünmüş Ortalamalar (Splinter-Average Distance)
- b) Otomatik Etkileşim Belirleme (Automatic Interaction Detection-AID)

II. *Hiyerarşik Olmayan Yöntemler*

- A. K-Ortalama (K-Means) Yöntemi
- B. Metoid Parçalama Yöntemi
- C. Yığma / Yığılma Yöntemi
- D. Bulanık (Fuzzy) Kümeleme Yöntemi

2.2.1 Hiyerarşik Yöntemler: Hiyerarşik kümeleme yöntemleri özellikle küçük örneklemeler (tipik olarak $n < 250$) için uygundur. Hiyerarşik kümeleme analizi yapmak için araştırmacılar, benzerlik ya da uzaklığın nasıl tanımlanacağına ve kümelerin nasıl birleşeceğine ya da ayrılacağına karar vermek durumundadırlar [9,2]. Hiyerarşik kümeleme yöntemlerinin, araştırmacının incelediği veri setinde kaç grup bulunduğunu başlangıçta bilmediği durumlarda çok uygun bir yöntem olduğu belirtilmektedir. Ayrıca bu yöntemin araştırmacılara, inceledikleri veri setinde daha önce gözlemlenmemiş ilişkileri gözleme ve ilkeleri keşfetme olanağı vermesi nedeniyle de faydalı olduğu belirtilmektedir.

Bu çalışmada daha önce de değinildiği üzere, birleştirici/toplamalı (agglomerative) hiyerarşik kümeleme yöntemlerinin daha sık kullanılmasından yola çıkılarak, bu yöntemlerin tanıtılmasına yer verilmiştir. Bu nedenle öncelikle bölünmeli/ayırıcı (divisive) hiyerarşik kümeleme yönteminin kısa bir tanıtımına yer verilerek birleştirici hiyerarşik kümeleme yönteminin daha ayrıntılı tanıtılmasına çalışılacaktır.

Bölünmeli/ayırıcı (divisive) hiyerarşik kümeleme yönteminde bütün gözlemler başlangıçta tek bir küme olarak kabul edilir. Ardından küme sayısı bir indirgenerek benzerlik matrisi tekrar oluşturulur ve benzerlik veya uzaklık matrisine göre benzer birimler bir araya getirilerek n birim aşamalı olarak sırasıyla 1, 2, 3... ($n-r$) ($n-3$), ($n-2$), ($n-1$) ve n kümeye yerleştirilir [9].

Birleştirici/toplamalı (agglomerative) hiyerarşik kümeleme yönteminde başlangıçta her gözlem ya da birim kendi başlarına bir küme oluşturur. Bir başka deyişle n tane birey, n tane küme olmak üzere işlemlere başlanır. Daha sonra küme sayısı bir indirgenerek benzerlik matrisi, tekrar oluşturulur. Benzerlik veya uzaklık matrisine göre n birim aşamalı olarak sırasıyla n , ($n-1$), ($n-2$), ... ($n-r$), ... 3, 2, 1 kümeye yerleştirilir [9]. Bir diğer deyişle

işlem, her adımda küme sayısının azalması ile devam eder ve en sonunda bütün birimler büyük tek bir kümede toplanır.

Birleştirici hiyerarşik kümeleme yönteminde, birimlerin birbirleri ile birleştirilmesinde değişik yaklaşımlar (kümeleme algoritmaları) uygulanmaktadır. Bunlar içerisinde en yaygın kullanılanlar şunlardır .

- 1) Tek Bağlantı Kümeleme / En Yakın Komşuluk Yöntemi (Single Linkage / Nearest Neighbour Method)
- 2) Tam Bağlantılı Kümeleme / En Uzak Komşuluk Yöntemi (Complete Linkage Metod / Farthest Neighbour Method)
- 3) Ortalama Bağlantı Kümeleme Yöntemi (Average Linkage Metod)
- 4) Ward's Bağlantı Kümeleme Yöntemi (Ward Linkage Metod)
- 5) Merkezi Bağlantı Kümeleme Yöntemi / Kitle Merkezi Yöntemi (Centroid Linkage Method)

1. *Tek Bağlantı Kümeleme Yöntemi:* Tek bağlantı kümeleme yöntemi, hiyerarşik kümeleme analizleri içerisinde en basit olanıdır. En yakın komşu olarak da bilinen bu yöntemde uzaklıklar matrisi kullanılarak birbirine en yakın (uzaklık değerleri en küçük) birey ya da kümeler birleştirilmekte ve bu birleştirilme işlemi artarda tekrarlanarak yürütülmektedir. Yöntem, en yakın iki noktanın bulunması ile başlar. İki nokta bir küme oluşturur. Üçüncü noktanın eğer kümedeki herhangi iki noktadan birine olan uzaklığı, kümelenmemiş olan diğer noktalara uzaklıklardan daha küçükse, bu kümeye atanır. Tek bağlantı yöntemi elipsler şeklinde dağılmayan değerleri aynı kümede toplayabilen az sayıdaki kümeleme yönteminden biridir. Örneğin; koordinat sisteminde U şeklinde dağılan noktalar bu yönteme göre aynı kümenin içerisinde yer alabilir. Bu sebeple kümenin zıt taraflarındaki değerler birbirinden oldukça farklı olabilir. Zincirleme etkisiyle ortaya çıkan bu durum sebebiyle çoğu birbirinden oldukça farklı üyeler aynı kümede yer alabilirler. Bu yöntem birbirinden yeterince ayrık olan kümeleri tespit etmekte oldukça iyi, buna karşılık birbirinden çok az farklı olan kümeleri ayırmada ise yetersizdir “Şekil 2.1”. Tek bağlantı yönteminin sonuçlar, bir ağaç diyagramında veya dendogramda gösterilebilir. Ağaç yapıdaki dallar, kümeleri göstermektedir [1,2,9,10].



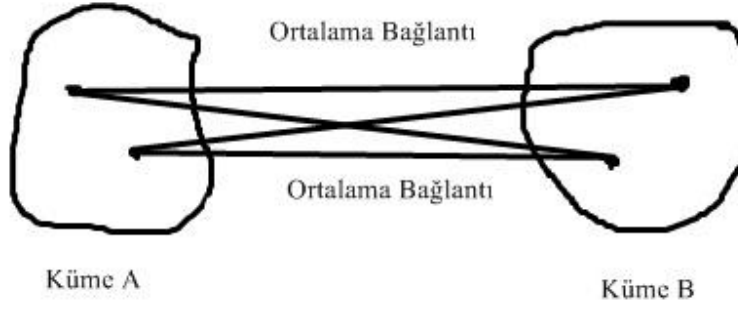
 ekil 2.1. Tek Baęlantı Y ntemi

2) *Tam Baęlantılı K meleme Y ntemi*: Tek baęlantı k meleme y ntemi ile benzerlik g steren bu y ntemde k meleme yaklařımı en b y k (maksimum) uzaklıęa dayanmaktadır “ ekil.2.2”. En uzak komřuluk olarak da bilinen bu y ntem, en uzak iki noktanın bir k meye atanması ile bařlar ve aynı mantıkla k melerin dięer k melerle birleřtirilmesiyle son bulur. Noktaların k melerle veya k melerin k melerle birleřtirilmesinde  l  t olarak en b y k uzaklık dikkate alınır. İki k me arasındaki uzaklık, birinci k medeki bir noktanın, ikinci k medeki bir noktaya olan en b y k uzaklıęıdır Tam baęlantı y ntemi, X-Y koordinat sisteminde birbirine yakın noktaların elips řeklinde daęılım g stermesi durumunda bu deęerleri k melemede, tek baęlantı teknięine g re daha iyi sonu lar vermektedir. Tek baęlantı y ntemindeki zincirleme etkisini ortadan kaldırır ve daha derli toplu /  zl  (compact) k meleme sonu ları oluřturur. Her ne kadar verilerin sadece bir y n n  ( yeler arasındaki en b y k uzaklık) temsil etse de, arařtırmacılar geniř ranjlı k meleme uygulamaları i in bu y ntemi uygun bulmaktadırlar [1,2] Tam baęlantı teknięi, aynı k me i erisinde bulunan g zlemlerin uzaklıklarının belirli bir deęerden k  k olması durumunda, t m k melerin saęlıklı bir bi imde oluřturulmasını garanti edememektedir [8].



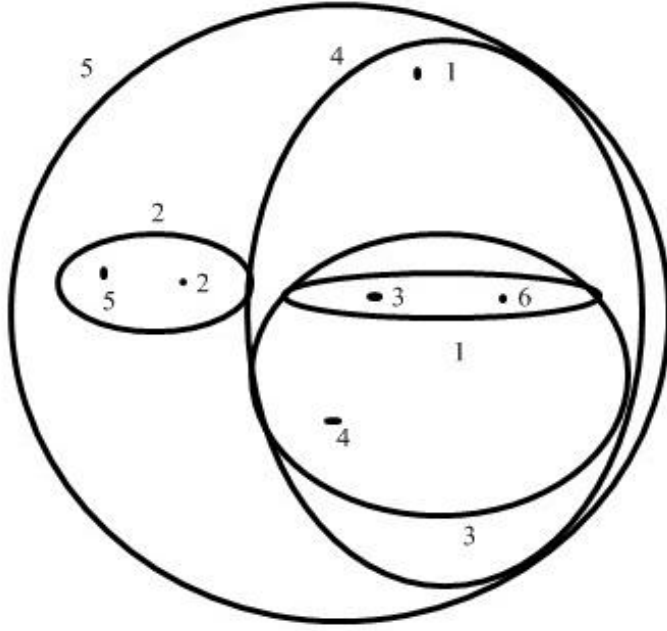
Şekil 2.2. Tam Bağlantı Yöntemi

3) *Ortalama Bağlantı Kümeleme Yöntemi:* Gözlemler arasındaki benzerliğin ortalamasına dayanmaktadır. İki küme arasındaki uzaklık, birinci kümedeki noktaların, ikinci kümedeki noktalara olan uzaklıklarının ortalaması alınarak bulunmaktadır “Şekil2.3”. Tek bağlantı yönteminde birbirine en yakın, tam bağlantı yönteminde ise birbirine en uzak komşulardan başlanarak kümeleme yapıldığından, bu iki teknik uç değerlerin etkisine oldukça açıktır. Ortalama bağlantı yöntemi bu dezavantajı ortadan kaldıran bir yöntem olarak kullanılmaktadır. Çünkü benzerlik, uçlarda yer alan bir üye çifti yerine, kümedeki tüm üyelerin benzerliğine dayalıdır. Bundan dolayı da uç değerlerden daha az etkilenir. Bu yöntem küçük küme içi değişimlere sahip kümeler üretme eğilimindedir ve aynı zamanda, aşağı yukarı eşit grup içi varyansa sahip kümeler üretir. Genellikle tam bağlantı ve ortalama bağlantı tekniklerinde benzer ağaç diyagramları (dendogramlar) oluşmaktadır. Ancak her bir yöntemde uzaklık farklı tanımlandığı için birleştirmeler farklı seviyelerde ortaya çıkabilmektedir [11,1].



 ekil 2.3. Ortalama Baęlantı Y ntemi

4) *Ward's Baęlantı K meleme Y ntemi*: En k   k varyans y ntemi olarak da bilinen Ward's baęlantı y ntemi, dięer k meleme y ntemleri gibi k meler arasındaki uzaklıkları hesaplamak yerine, k me i i hata kareler toplamını minimize ederek, homojenlięi maksimum kılacak k meler olu turur “ ekil 2.4”. Her a amada elde edilen k melerden hata kareler toplamı en k   k olanlar birle tirilir. Ama  k me i inde homojenlięi, k meler arasında heterojenlięi maksimum olan k meler olu turmaktır [1,8].



Şekil 2.4. Ward's Bağlantı Kümeleme Yöntemi

5) *Merkezi Bağlantı Kümeleme Yöntemi / Kitle Merkezi Yöntemi*: Bu yöntemde iki küme arasındaki benzerlik, iki küme merkezinin uzaklığı ile belirlenir. Küme merkezleri, küme değişkenlerine ilişkin gözlemlerin ortalama değeridir. Bu yöntemde bireyler her seferinde gruplanır ve yeni merkez hesaplanır. Küme merkezleri kümelerde birleşmeler meydana geldikçe ya da oluştuğu yer değiştirir / taşınır. Bir diğer deyişle, var olan bir kümeye yeni bireyler eklendikçe küme merkezi değişir. Bu yöntem biyoloji gibi bilim dallarında çok kullanılmakla birlikte, genellikle karmaşık sonuçlar üretir. Bu karmaşıklık yöntemin tersinmesinden / ters dönmesinden (reversal) kaynaklanır. Bir başka deyişle, bir merkez çifti arasındaki uzaklık, daha önceki adımda birleşmiş olan bir başka merkez çifti arasındaki uzaklıktan daha az olabilir. Bu yöntemin avantajı, ortalama bağlantı yönteminde olduğu gibi, diğer hiyerarşik yöntemlere oranla uç değerlerden daha az etkilenmesidir [1]

Yukarıda söz edilen yöntemler dışında, "McQuitty Bağlantı Yöntemi" ve "Ortanca Bağlantı Yöntemi (median linkage method)" gibi yöntemler de bulunmaktadır.

Kümeleme analizi; birkaç adımdan oluşan bir çözüm sürecidir. Veri girişi analizin ilk aşamasıdır. Yani ilk olarak doğal sınıflamaları hakkında kesin bilgilerin bulunmadığı

anak t lelerden alınan n sayıda birimin incelenen p sayıda deęiřkene iliřkin g zlem deęerleri elde edilir. B ylece veri matrisi oluřturulmuř olur. Daha sonra verinin  l  m tipine uygun bir benzerlik  l  s   le nesnelerin uzaklıklar matrisi elde edilir. Uygun k meleme teknięi se ilir ve uygulanır. Teknięin uygulanması sonucu nesneler k melere ayrılmıř olur. K meleme sonu larının anlamlılıęının yorumlandıęı ařama ise, analizin son ařamasıdır [6].

Sonu  olarak hiyerarřik tekniklerin  zelliklerini ř yle  zetlemek m mk nd r [11,12].

- Hiyerarřik k meleme teknikleri, k meleri ardarda birleřtirme s recidir.
- Hiyerarřik k meleme y ntemleri, veri matrisindeki birimlerin/deęiřkenlerin bařlangı ta ka  k me oluřturduęuna ve k me elemanlarını belirlemede bařlangı ta hangi  l  t n se ildięine g re iki temel gruba ayrılırlar: "Birleřtirici/Toplamalı Hiyerarřik K meleme Y ntemleri (Agglomerative)" ve "Ayrırcı/B l nmeli Hiyerarřik K meleme Y ntemleri (Divisive)"
- Birleřtirici/Toplamalı tekniklerde n adet nesne ardıl olarak gruplarda birleřtirilir.
- Ayrırcı/B l nmeli tekniklerde n adet nesne ardıl olarak daha k   k k melere ayrılır.
- Birleřtirici/Toplamalı teknikler, Ayrırcı/B l nmeli tekniklere g re daha sık kullanılmaktadır.
- Tek baęlantı, tam baęlantı, ortalama baęlantı, merkezileřtirme ve Ward's baęlantı y ntemi,  ok yaygın olarak kullanılan y ntemlerdir.
- Hiyerarřik k meleme analizi u  deęerlere olduk a duyarlıdır. Bu nedenle yapılması gereken, farklı hiyerarřik tekniklerin uygulanıp bunların karřılařtırılmasıdır. Eęer farklı tekniklerin sonu ları genel hatlarıyla tutarlılık g steriyorsa k meleme iřlemi yapılır.
- B y  k veri setlerinde hiyerarřik k meleme analizi iřlemleri  ok uzun s rmektedir.
- Hiyerarřik y ntemle yapılan toplamalı ve b l nmeli gruplandırmalar geri alınamazlar. Bir grup, dięeri ile bir kez birleřtirildikten sonra, daha sonraki adımlarda kesinlikle ayrılamaz.
- Hiyerarřik tekniklerin aęa  diyagramları ile g sterilen sonu larına dendogram denir.

2.2.2 Hiyerarřik Olmayan Y ntemler: K me sayısının belirlenmiř olduęu durumlarda kullanılır. Bir bařka deyiřle arařtırmacının k me sayısı konusunda  n

bilgisinin olduđu durumlarda ya da küme sayısına karar verilmiş olması durumunda kullanılması önerilmektedir [3].

Çeşitli kümeleme türleri birbirinden ayırt edilirken dikkat edilecek bazı hususlar vardır; hiyerarşik (iç içe) kümelemeye karşılık bölmesel (iç içe olmayan), seçkin (exclusive) kümelemeye karşın örtüşen (overlapping), bulanık (fuzzy) ve son olarak tama karşın kısmi kümeleme. Hiyerarşik(iç içe) kümelemeye karşın bölmesel(iç içe olmayan) kümeleme. Üzerinde en çok tartışmanın yapıldığı kümeleme türlerini birbirinden ayırma kriteri onların iç içe olup olmadıkları ile ilgilidir, yada daha geleneksel bir ifade ile hiyerarşik yada bölmesel olmaları ile ilgilidir. Bir bölmesel kümeleme basitçe veri nesnelerinin örtüşmeyen alt kümeler ayrılmasıdır öyle ki; her bir veri nesnesi yalnızca bir kümede bulunur. Örneğin Şekil 7(b) ve (d)'deki her bir küme koleksiyonu bir bölmesel kümelemedir. Kümelerin alt kümeler sahip olması durumunda ise hiyerarşik kümeleme yapılmış olunur. Hiyerarşik kümeler ağaçlar şeklinde organize edilmiş iç içe geçmiş alt kümelerden oluşur. Yaprak düğümler (leaf node) dışında ağaçtaki her bir düğüm(küme), kendi alt kümelerinin bir birliği ve ağacın kökü ise tüm nesneleri içeren bir kümedir. Her zaman değil fakat sıklıkla, ağacın yaprakları her bir veri nesnesi için tek kümeyi (singleton cluster) temsil eder. Kümelerin iç içe geçmelerine izin verilmesi durumunda, Şekil 7 (a)'nın bir değerlendirmesi onun iki kümeye sahip olduğunu ve sonrasında da bu iki kümeden her birinin de üç alt kümeye sahip olduğunu ifade eder. Şekil 2'de gösterilen kümeler, sırasıyla bakıldığında, hiyerarşik bir sınıflandırma teşkil ederler öyle ki; sırasıyla her bir seviyede 1, 2, 4 ve 6 küme yer alır. Son olarak, bir hiyerarşik kümelemenin aslında bir dizi bölmesel kümeleme olduğunu ve bu bölmesel kümelemenin de bu dizinin herhangi bir elemanı alınarak elde edilebileceği söylenebilir, yani hiyerarşik ağacın bazı seviyelerden kesilmesi gibi. Seçkin, örtüşen ve bulanık kümeleme. Şekil 2' de gösterilen tüm kümelemeler seçkin kümelemedir çünkü her biri bir nesneyi yalnızca bir kümeye koyar. Bazı durumlarda bir noktanın birden fazla kümede yer alması mantıklı olabilir ve bu durumlar seçkin olmayan kümeleme ile daha iyi açıklanabilir. En genel şekilde, bir örtüşen ve seçkin olmayan kümeleme bir nesnenin aynı anda birden fazla gruba(sınıfa) ait olmaları gerçeğini ortaya çıkarmada kullanılır. Örneğin; bir kişi üniversitede hem bir öğrenci hem de bir çalışan olabilir. Seçkin olmayan kümeleme aynı zamanda bir nesnenin birden fazla kümeye ait olabilmesi durumunda bunlardan herhangi birine konması için de kullanılabilir. Şekil 2' deki noktalardan iki küme arasına denk gelebilecekleri düşünülür. Onu rastgele bir kümeye koymak yerine, aynı uzaklıktaki tüm kümelere konur. [17]

Bulanık kümelemede, bir nesne belirli bir ağırlık değeriyle tüm kümelere ait olur. Bu ağırlık değeri 0 (hiç ait olmama) ile 1 (tamamıyla aitlik) arasında değerler alır. Diğer bir deyişle, kümeler mantık setleri olarak ele alınırlar. (Matematiksel olarak bir bulanık set içinde bir nesne herhangi bir sete 0 ile 1 arasında değerler alan bir ağırlık değeriyle aittir. Bulanık kümelemede, bir nesne için toplam ağırlık değerinin 1 olması gibi bir kısıt ortaya konulur.) Benzer şekilde, olasılıklı kümeleme teknikleri de her bir noktanın her bir kümeye aitliğine dair bir olasılık hesaplar ve bu olasılıklar toplamı da 1 olmak zorundadır. Üyelik ağırlıklarının yada olasılıkları toplamının 1 olması sebebiyle, bulanık yada olasılı kümeleme gerçek birden fazla sınıflandırma (ture multiclass) durumunu açıklamazlar, örneğin bir öğrenci çalışanı durumunda bir nesne birden çok sınıfa aittir. Bunun yerine, bu yaklaşımlar bir nesnenin rastgele yalnızca bir kümeye atanmasının önüne geçildiği ve aslında birden çok kümeye yakın olduğu durumlar için elverişlidir. Pratikte, bir bulanık yada olasılı kümeleme bir seçkin kümelemeye dönüştürülür; şöyle ki bir nesne ağırlığının yada olasılık değerinin en fazla olduğu kümeye atanır.

Tam kümelemeye karşı kısmi kümeleme. Tam kümeleme her nesneyi bir kümeye atarken; kısmi kümeleme bunu yapmaz. Kısmi kümeleme ardındaki neden bir nesnenin aslında iyi tanımlanmış bir gruba ait olamayışıyla ilgilidir. Çoğu kez bir veri seti içerisindeki nesneler, bir gürültüyü (noise), küme dışında kalmayı (outlier) yada ilgi çekmeyen bir arkaplanı (uninteresting background) temsil edebilirler. Örneğin bazı gazete makaleleri ortak bir temayı paylaşabilir, küresel ısınma gibi; fakat bazıları da çok daha genel yada tek bir çeşit tema olabilir. Bu yüzden son ay makalelerinin en önemli başlıkları bulmak için, yalnızca ortak bir temanın geçtiği kümeler içinde aramalar yapmak isteyebiliriz. Diğer durumlarda, nesnelerin tam bir kümelemesi istenir. Örneğin, dokümanları tarama ihtiyacıyla organize etmek için kümeleme kullanan bir uygulama öyle ki, bu uygulama tüm dokümanların taramamasını garanti eder.

Kümeleme nesneleri faydalı gruplara(kümelere) ayırmayı amaçlar, burada fayda veri analizinin hedefleri tarafından tanımlanır. Doğal olarak, pratikte fayda sağlayan değişik türde kümeler vardır. Bu kümeler arasındaki farkı görsel olarak ortaya koyabilmek için, Şekil 2.5’de ki gibi her nesnelerimizi iki boyutlu uzayda noktalar olarak gösteririz. Bununla beraber üstünde durmak istediğimiz nokta burada verilen kümeler diğer başka veri türleri için de aynı ölçüde geçerliliğe sahiptirler.

İyi ayrılmış böyle bir küme bir nesneler setidir öyle ki; küme içindeki her bir nesne aynı küme içindeki bir diğer nesneye benzer yada yakın iken küme dışındaki nesnelerden farklı veya bu nesnelere uzaktır. Kimi zaman küme içindeki nesnelerin birbirlerine yeterince benzer olduklarını belirtmek için belirli bir eşik kullanılır. Kümenin bu ideal tanımı yalnızca verinin doğal sınıfları yani birbirlerinden yeterince uzak olan sınıfları içermesi durumunda geçerli veya doyurucu olabilir. Şekil 2.5 (a) iyi ayrılmış bir sınıf için bir örnektir ve bu sınıf iki boyutlu uzayda iki grup noktadan oluşur. Farklı gruplar içinde bulunan herhangi iki nokta arasındaki uzaklık aynı grup içindeki herhangi iki nokta arası uzaklıktan daha fazladır. İyi ayrılmış sınıflar küre biçiminde olmak zorunda değildir, fakat bir şekle sahip olabilirler.

Prototip Tabanlı böyle bir küme bir nesneler setidir öyle ki; küme içindeki her bir nesne kümeyi tanımlayan prototipe benzer yada yakın iken diğer küme prototiplerinden farklı yada bu prototiplere uzaktır. Sürekli özelliklere sahip veriler için, prototip bir ağırlık merkezidir yani kümedeki tüm noktaların ortalaması. Ağırlık merkezinin anlamlı olmadığı durumlarda, örneğin veri kategorik özelliklere sahip ise; bu durumda prototip bir medoid'dir yani kümeyi en iyi temsil edecek noktadır. Birçok veri türü için, prototip en merkez nokta olarak düşünülebilir ve bu gibi durumlar için prototip tabanlı sınıfları merkez tabanlı sınıflar olarak değerlendirilir. Doğal olarak bu kümeler küresel şekle sahip olma eğilimindedirler. Şekil 2.5 (b) böylesi bir duruma örnektir.

Graf Tabanlı eğer veri düğümlerin nesneleri, dalların nesneler arası bağlantıları temsil ettiği bir graf olarak temsil ediliyorsa, bu durumda küme bir bağlı bileşen (connected component) olarak tanımlanabilir öyle ki, grup içindeki nesneler birbirlerine bağlı ve grup dışındaki nesnelere ise bağlantıları yoktur. Graf tabanlı kümeler önemli bir örnek, komşuluk tabanlı kümelerdir. Bu kümelerde, iki nesne ancak aralarında belirli bir mesafe var ise birbirlerine bağlıdır. Bu durumda, komşuluk tabanlı kümelerde, bir nesne küme içerisindeki başka bir nesneye başka kümedeki herhangi bir noktadan daha yakındır. Şekil 2.5 (c) bu türden kümeleri iki boyutlu uzayda noktalar halinde göstermektedir. Kümenin bu tanımı, kümeler düzensiz veya birbirlerine geçmiş durumda iken yararlıdır; fakat bununla beraber gürültünün olması durumunda ise sorunlarla karşılaşılabilir, çünkü Şekil 2.5 (c)' de iki küresel sınıf olarak gösterildiği gibi noktalar arasındaki küçük bir köprü iki birbirinden farklı kümeyi birleştirebilir.

Graf tabanlı kümelerin başka türleri de mümkündür. Böyle bir yaklaşım bir kümeyi bir komite (clique) olarak tanımlar, öyle ki grafdaki bir düğüm seti birbirleriyle tamamen

bağlıdırlar. Özellikle, nesneler arasındaki uzaklığın derecesine göre nesneleri birbirlerine bağladığımızda, bir nesne setinin bir komite oluşturmasıyla bir sınıf oluşur. Prototip tabanlı kümeler gibi, bu kümeler de küresel olma eğilimindedirler.

Yoğunluk Tabanlı bir küme, düşük yoğunluğa sahip bir bölge tarafından sarılmış ve yüksek nesne yoğunluğuna sahip bir bölgedir. Şekil 2.5 (d), Şekil 2.5 (c)' deki veriye gürültü eklenerek elde edilmiş veri için çeşitli yoğunluk tabanlı kümeleri göstermektedir. Şekil 2.5 (c)' deki gibi iki dairesel küme birleştirilmemiştir, çünkü aralarındaki köprü gürültü içerisinde kaybolmaktadır. Benzer şekilde, Şekil 2.5 (c)'deki eğri de yine gürültü içerisinde kaybolmakta ve Şekil 2.5 (d)' de bir küme teşkil etmemektedir. Bir kümenin yoğunluk tabanlı tanımlaması daha çok kümeler düzensiz yada birbirlerine geçmiş iken ve aynı zamanda gürültü ve dışsallar (outlier) var iken kullanılır. Buna karşın, bir kümenin komşuluk tabanlı tanımı, Şekil 2.5(d)' deki veri için geçerli olmayacaktır; çünkü gürültü kümeler arası köprüler oluşturma eğiliminde olacaktır. [17]

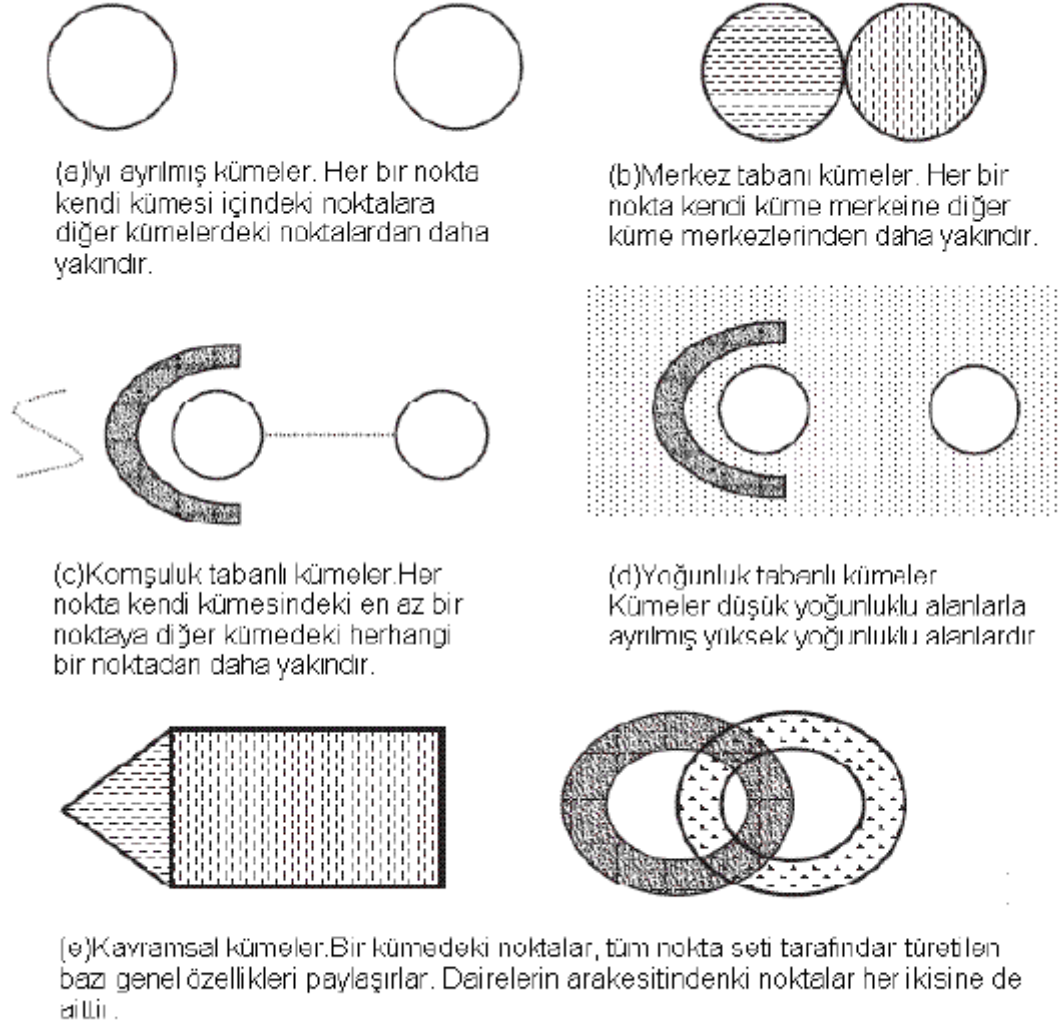
Paylaşılmış özellik(kavramsal kümeler) daha genel bir şekilde, bir kümeyi bazı özellikleri paylaşan nesne setleri olarak tanımlanabilir. Bu tanım daha önce yapılmış tüm küme tanımlamalarını kapsar, yani merkez tabanlı kümelerdeki nesneler ortak özellikleri paylaşırlar öyle ki, onlar ağırlık merkezine veya medoide en yakındırlar. Bununla beraber paylaşılmış özellik yaklaşımı yeni küme türlerini de kapsar. Şekil 2.5 (e)' deki kümeleri ele alınırsa;. Bir üçgensel alan (küme), bir dikdörtgensel alana yakındır ve iki tane birbirine geçmiş daire (küme) vardır. Her iki durumda da, bir kümeleme algoritması kümeleri tespit edebilmek için özel bir küme kavramına ihtiyaç duyacaktır. Böylesi kümeleri bulma işlemi, kavramsal kümeleme olarak adlandırılır.

Yol Haritası

- K-ortalamları. Bu bir prototip tabanlı ve kısmi kümeleme tekniğidir ve kullanıcının belirleyeceği sayıda ve merkezlerince temsil edilecek K tane küme bulmaya çalışır.

- Toplamsal hiyerarşik kümeleme. Bu kümeleme yaklaşımı hiyerarşik kümeleme üreten birbirleriyle sıkıca ilişkili kümeleme tekniklerinin bir koleksiyonuna başvurur şöyle ki; her bir noktayı tekil bir küme olarak ele alır, sonrasında ardışıl olarak en yakın iki kümeyi birleştirir, tek ve tam- kapsayan (all-encompassing) küme kalıncaya kadar bu işlem devam eder. Bu tekniklerden bazıları graf tabanlı bazıları da prototip tabanlı yaklaşımlar açısından bir doğal anlama sahiptirler.

- DBSCAN. Bu yoğunluk tabanlı bir kümeleme algoritması olup, küme sayısının algoritma yardımıyla otomatik olarak belirlendiği kısmi kümeler oluşturur. Düşük yoğunluk bölgelerindeki noktalar gürültü olarak değerlendirilip ihmal edildiğinden, DBSCAN tam bir kümeleme üretmez.



Şekil 2.5. İki boyutlu noktalar kullanılarak gösterilen değişik türdeki kümeler

Hiyerarşik olmayan yöntemlerin hiyerarşiklerden temel farkları şöyle özetlenebilir:

- Herhangi bir hiyerarşik işleyiş yoktur.

- Hiyerarşik yöntemlerde nesneler adım adım her nesne tek bir kümede oluncaya kadar ya da tam tersi şekilde kümelenir. Hiyerarşik olmayan yöntemlerde ise nesneler başlangıçta belirlenen kümelere atanarak işlem yapılır.
- Hiyerarşik yöntemlerde her adımda hangi nesnenin hangi kümede olduğu belirirken, hiyerarşik olmayan yöntemlerde sadece nesnelerin en sondaki üyelik durumları önemlidir.
- Hiyerarşik yöntemlerde dendogramlar kullanılırken, hiyerarşik olmayan yöntemlerde dendogramlar hiç bir şey ifade etmez.
- Hiyerarşik olmayan yöntemler, hiyerarşik yöntemlere göre daha büyük veri kümelerine uygulanabilir [14].

2.3 Kümeleme Analizinin Aşamaları

Amaçların belirlenmesi ve değişkenlerin seçiminden sonra araştırmacı aşağıdaki soruları cevaplandırmalıdır [1].

- Örneklem büyüklüğü yeterli midir?
- Verilerde uç değer (outliers) var mı ve kaldırılabilir mi?
- Gözlemlerin benzerlikleri nasıl belirlenmelidir?
- Veriler standartlaştırılmalı mıdır?

Bu sorulara yanıt vermek için pek çok farklı yaklaşım kullanılabilir. Ancak hiç biri, bu sorulardan birinin kesin yanıtı olmaya yeterli değildir ve farklı yaklaşımlar aynı veriler için farklı sonuçlar üretecektir. Bu yüzden kümeleme analizi de faktör analizi gibi, araştırma deseninden ve araştırmacının yöntem seçiminden, diğer çok değişkenli yöntemlere göre daha fazla etkilenecektir.

Kümeleme analizi birey ya da birimlerin benzerlik ya da uzaklıklarına göre yapılır. Bu nedenle de analizin ilk adımı, benzerlik ya da uzaklık matrisinin oluşturulmasıdır. Ancak, benzerlik ölçümleri hesaplanmadan önce, verilerin standartlaştırmasının gerekip gerekmediğine karar verilmelidir; çünkü bu ölçümler farklı ölçeklere ve değişkenler arasındaki büyüklük farklarına duyarlıdır.

2.3.1 Değişkenlerin Standardizasyonu ve Dönüştürülmesi

Tüm değişkenler aynı ölçek düzeyinde ölçüldüğünde (örneğin bir grup tutum sorusu), genellikle verilerin standardizasyonuna ihtiyaç duyulmazken, değişkenlerin farklı ölçek düzeyinde ölçülmesi durumunda, bu oldukça önemli bir konu haline gelir [1].

En yaygın kullanılan standardizasyon şekli, her bir değişkeni standart puanlara dönüştürmektir. Aynı zamanda Z puanlar, olarak da bilinen bu dönüştürmede, her bir ham puandan ortalama çıkartılır ve standart sapmaya ulaşılır. Bu dönüştürme şekli tüm bilgisayar programlarında bulunmakta ve kümeleme analizi sürecine de doğrudan dahil edilebilmektedir [1].

Z puanlarına dönüştürme: Normal dağılım gösterdiği varsayılan aralık ya da oran ölçeğiyle elde edilmiş verilere uygulanır. Z puanlarına dönüştürme:

$$Z = \frac{x - \bar{x}}{s} \text{ formülü ile yapılır.}$$

Bu dönüştürmeden başka aşağıda ifade edilen farklı dönüştürme yöntemleri de bulunmaktadır [3] .

-1 ≤ x ≤ 1 aralığına dönüştürme: Değerler arasında pozitif ve negatiflerin heterojen yapıdaki değerlerin ve uç değerlerin olduğu durumlarda tercih edilen bir tekniktir. Teknik, değişkenin bütün değerlerini maksimum değere oranlar. X max, dizideki en büyük değeri ifade eder.

0 ≤ x ≤ 1 aralığına dönüştürme: Heterojen yapıya sahip değerlerin bulunduğu ve/veya uç değerlerin yer aldığı durumlarda değerleri pozitif 0-1 aralığında değiştirmek üzere uygulanır.

Maksimum değer 1 olacak şekilde dönüştürme: Dizideki değerlerin maksimum 1 olacak şekilde dönüştürülmesi için uygulanan bir tekniktir.

Dönüşüm ortalaması 1 olacak şekilde dönüştürme: Dizinin ortalamasının pozitif ve 1 olması istendiği durumlarda uygulanan tekniktir.

Dönüşüm dizisinin standart sapması 1 olacak şekilde dönüştürme: Yeni dizinin standart sapmasının 1 olmasının istendiği durumda kullanılan bir tekniktir.

2.3.2. Benzerlik ve Farklılık Ölçümlerinin Seçimi

Kümeleme analizinin ilk adımı benzerlik ya da uzaklık matrisinin oluşturulmasıdır. Bu matris, satırlarında ve sütunlarında analiz birimleri olan ve hücrelerinde her bir gözlem çifti için benzerlik ya da uzaklık ölçümleri olan bir tablodur. Birbirine en yakın kümelerin geniş gruplar olarak birleştirilmesi ya da bir nokta ile bir küme arasındaki ilişkinin ölçüt olarak kullanıldığı gözlemler arası uzaklıklar ya da kümeler arası benzerlik ve uzaklıklar için çeşitli ölçümler mevcuttur. Uzaklık ölçümleri, iki gözlemin birbirinden ne kadar uzak olduğunu gösterir. Birbirine benzer olan gözlemler arası uzaklıklar azdır. Benzerlik ölçümleri ise iki gözlemin birbirine ne kadar benzer olduğunu gösterir. Ancak tüm ölçümleri aslında "uzaklık ölçümü" olarak adlandırmak mümkündür; çünkü aynı işlevi görürler. Burada dikkat edilmesi gereken önemli bir nokta, eğer uzaklıkları tanımlamak için iki ya da daha fazla değişken kullanılıyorsa, bunlardan birinin daha baskın olacağıdır. Bunu önlemek için değişkenlerin standardizasyonu yararlı olacaktır [1].

Kümeleme aynı küme içerisindeki gözlemlerin birbirine benzer, diğer kümelerdeki gözlemlerden farklı olacak şekilde yapılmasıdır. Bu amaç için benzerlik ve farklılık kavramları kullanılır. Benzerlik iki nesne veya iki özellik arasındaki ilişkinin kuvveti olarak açıklanır. Bu nicel değer alınan ölçeğe veya veri tipine göre değişik yollardan elde edilir. Farklılık ise, iki nesne arasındaki zıtlık ya da uyumsuzluğun bir ölçüsü olan farklılıkları ölçer. Benzerlik ve farklılık ölçümleri gözlemlerin birbirinden ayırt edilmesini sağlar ve bu sayede gözlemler gruplara ayrılır [15, 16].

Değişken tipleri kesikli ve sürekli olmak üzere iki kategoride sınıflandırılabilir. Değişkenin aldığı değerlerin sayısı sonlu veya sayılabilir sonsuzlukta ise bu değişkene kesiklidir denir, eğer değişken birden çok aralıkta her değeri alabiliyorsa bu değişken sürekli. Kümeleme analizinde değişkenlerin ölçek türleri büyük önem taşır. Stevens ölçüm düzeylerini isimsel, sıralı, aralık ve oransal olmak üzere dört sınıfa ayırmıştır [16].

Birimlerin deęişkenlere göre birbirleri arasındaki uzaklıkları hesaplamak amacıyla çeşitli uzaklık ölçü birimleri ileri sürülmüştür. Bu ölçü birimleri veri matrisinde yer alan deęişkenlerin ölçü birimlerine göre farklılık gösterir. Eđer deęişkenler oransal ya da aralıklı ölçekle elde edilmiş deęerler ise uzaklık ya da ilişki türü ölçülerden yararlanılır. Eđer ikili gözlemlere göre ölçümler yapılmış ise birimler arasındaki benzerlik ve farklılık ölçülerinden yararlanılır.

Birimlerin birbirleri ile olan benzerlik düzeyleri benzerlik (similarity, sim) matrisi ile gösterilir. sim matrisinin elemanları sim_{ij} ile gösterilir ve $sim_{ij}=100(1-d_{ij} / \max(d_{ij}))$ biçiminde hesaplanır. Birimlerin birbirinden farklılıkları (dissimilarity, diss) matrisinden yararlanılarak hesaplanır. Diss matrisinin elemanları $diss_{ij}$ ile gösterilir ve $diss_{ij}=100-sim_{ij}$ biçiminde hesaplanır [3]

Kümeleme analizinde birimler arasındaki uzaklıkların hesaplanmasında sıklıkla kullanılan ölçüler aşağıdaki gibi verilebilir:

Benzerliğin ölçülmesinin farklı yolları bulunmaktadır. Kullanılacak benzerlik ölçümlerinin belirlenmesinde deęişkenlerin metrik veya kategorik olması önemlidir. Metrik deęişkenler için benzerliği ölçmenin en önemli iki yolu, korelasyona dayalı ölçüler ve uzaklık ölçüleri, kategorik veriler içinse ortaklık ölçüleridir. Aşağıda söz konusu bu ölçümler (indekslerden) bazılarına ilişkin kısa açıklamalar sunulmaktadır [13].

a)Metrik (Interval) veriler için:

Öklid uzaklığı ve Kareli öklid uzaklığı (Euclidean Distance - Squared Euclidean Distance): En yaygın uzaklık ölçüsüdür. Nesne sayısının 100'den fazla olması durumunda öklit uzaklığının kullanılması tavsiye edilir. Öklid uzaklığı bir üçgenin hipotenüs uzunluğudur ve X gözlemlerine ilişkin farkların karesinin, Y gözlemlerine ilişkin farkların karesi ile toplanmasının kareköküne eşittir, iki boyutlu düzlemde iki nokta arasındaki mesafe, iki noktanın X ve Y koordinatlarının ayrı ayrı farklarının hipotenüsüne eşittir. Öklid uzaklığı ve kareli öklid uzaklığı, standartlaştırılmış verilerle deęil, ham verilerle hesaplanır. Deęişkenlerin ölçek farklılıkları öklid uzaklıklarını önemli ölçüde etkiler.

Kosinüs (Cosine): Deęerlerin iki vektörü arasındaki açının kosinüsüne dayalı metrik düzeyde benzerlik ölçümüdür.

Pearson Korelasyon: Metrik düzeyde benzerlik momentler çarpımına dayanır. Değişken kümelemeye karşı birimleri ya da bireyleri kümelemede araştırmacılar sütunlarında değişkenler, satırlarında bireyler / birimler olan normal veri tablosunda yer değişikliği yaparlar (transpose). Böylelikle, sütunları bireyler ve satırları değişkenler olarak kullanmak yoluyla, korelasyonun artık bireyler arasında olması sağlanır ve bu korelasyonlar da benzerlik matrisinin hücrelerini oluşturur.

Mutlak değer (Absolute values): Pearson korelasyon yüksek pozitif çıkabileceği gibi, yüksek negatif de çıkabilir. Araştırmacıların mutlak değer almaları gerekebilir.

Chebychev Uzaklığı: Uzaklıkları tanımlamak için kullanılan iki ya da daha fazla boyuttan (değişkenden) herhangi birindeki bir gözlem çifti arasındaki maksimum mutlak farktır. Çiftler ya da eşler (pairs), diğer boyutlardaki benzerlikler dikkate alınmaksızın, tek bir boyuttaki fark dikkate alınarak tanımlanır.

Blok Uzaklığı: "Manhattan uzaklığı" ya da "City-blok uzaklığı" olarak da bilinen bu uzaklık, uzaklıkları tanımlamak üzere kullanılan iki ya da daha fazla boyuttaki mutlak (absolute) farkların ortalamasıdır. Bir başka deyişle bu yöntem, değişkenler arasındaki farkın mutlak değerini esas almaktadır. Uzaklık, gözlemlerin koordinatları arasındaki farkın mutlak değerleri toplamına dayalı olarak hesaplanmaktadır.

Minkowski Uzaklığı: Genellenmiş uzaklık fonksiyonudur, n sayıda birim ve p sayıdaki değişkene göre birimler ya da değişkenler arasındaki uzaklıkları hesaplamak için kullanılan uzaklık ölçülerine genel olarak "Minkowski uzaklık ölçüsü" adı verilmektedir. Mutlak farkların toplamının p 'nci kökünün, madde değerleri arasındaki p 'nci kuvvete olan uzaklığıdır. $p=1$ olduğunda, Minkowski uzaklığı ile City-blok uzaklığı birbirine eşittir. İki kategorili (ikilem) verilerde $p=1$ olduğunda, Minkowski uzaklığı Hamming uzaklığına, $p=2$ olduğunda ise, Öklit uzaklığına eşittir. K sayıda değişken standardize edilmediğinde ve farklı ölçek düzeyinde ölçüldüğünde, geniş ölçekli (large scale) değişkenler daha baskın olur.

Uyarlanmış (Customized) Minkowski Uzaklığı: Minkowski uzaklığının genellenmiş halidir. Madde değerleri arasındaki p 'nci kuvvet üzerinde mutlak farkların toplamının r 'nci kökü alınarak hesaplanır.

Mahalanobis Uzaklığı: Doğrudan standardizasyon olanağı sağlayan Öklid uzaklık ölçüsüdür. Hem standart sapmaları dikkate alarak standardizasyon olanağı sağlar; hem de grup içi varyans kovaryansları toplayarak değişkenler arasındaki korelasyonu ayarlar.

Hotelling T^2 Uzaklığı: İki grup ya da kümenin ortalama vektörlerinin karşılaştırılması amacıyla kullanılan bir uzaklık ölçüsüdür.

b) Kategorik (Counts) veriler için:

Ki-kare uzaklığı: İki frekans setinin eşitliği için ki-kare testine dayanır.

Phi-kare uzaklığı: Ki-kare ölçüsünü, birleşik frekansın kare kökünü almak yoluyla normalleştirir.

c) İkilemler (Binary) için:

-Öklid uzaklığı ve Kareli öklid uzaklığı: Metrik ölçek için olan uzaklığın ikilem veriler için olan şeklidir.

-Büyüklik farkı (Size difference): 0 ile 1 arasında değişen asimetri indeksidir.

-Örüntü farkı (Pattern difference): 0 ile 1 arasında değişen uzaklık ölçüsüdür.

-Varyans: 2×2 'lik tabloda $(b+c)/4n$ formülüyle hesaplanır ve 0 ile 1 arasında değişir.

-Dağılım (Dispersion): -1 ile 1 arasında değişen benzerlik ölçüsüdür.

-Biçim (Shape): 0 ile 1 arasında değişen uzaklık ölçüsüdür.

-Basit eşleşme (Simple matching): Eşlerin (çiftlerin), toplam değer sayısına oranıdır.

-Phi dört noktalı korelasyon: Pearson korelasyon katsayısının ikili eşidir (analog) ve -1 ile 1 arasında değişir.

-Lambda: Goodman ve Kruskal'ın lambdası, bir maddeden diğerini kestirmede (çift yönlü kestirim) hatanın oransal olarak azaltılması olarak yorumlanır. 0 ile 1 arasında değişir ve 1, mükemmel yordamaya karşılık gelir.

-Abderberg'in D'si: Lambda'nın bir başka şeklidir. Çift yönlü kestirimde bir maddeden diğerini kestirmede hatanın gerçek (actual) azatılımıdır ve 0 ile 1 arasında değişir.

-Hamann: Eşleşmiş çiftlerin sayısından eşleşmeyenlerin sayısının çıkartılıp toplam madde sayısına bölünmesi ile elde edilir. Bu indeks -1 ile 1 arasında değişir.

-Jaccard: Bu indekste ortak yokluklar (joint absences) gözardı edilir. Eşleşmiş çiftlerle eşleşmeyenler eşit ağırlıklandırılır. Aynı zamanda benzerlik oranı olarak da bilinir. Bu indeks ikilem olan veriler için en fazla önerilen indekslerden biridir.

-Kulczynski-1: Ortak varlıkların (joint presence), tüm eşleşmeyenlere oranıdır. Alt sınırı 0'dır; üst sınırı yoktur. Teorik olarak, hiç eşleşmeyen kalmadığı durumlarda tanımsızdır

(undefined). Ancak paket programlar bu tanımsızlık durumunda 9999.999 gibi rastgele değerler ya da daha büyük derler atamaktadır.

-Kulczynski-2: Bir özelliğin bir maddede varlığının, diğer maddede olma olasılığına ilişkin koşullu olasılıktır.

-Lance ve I/Villiams: Bray-Curtis metrik olmayan katsayı olarak da bilinir, $(b+c) / (2a+b+c)$ formülünü kullanan 2x2'lik tablolara dayanır. Bu formülde a, her iki maddede de olan bireyleri gösterirken, b ve c, bir maddede olan, diğerinde olmayan çapraz köşeleri gösterir. 0 ile 1 arasında değişir.

-Ochiai: Benzerlik ölçümlerindeki kosinüsün ikilem veriler için olan karşılığıdır. 0 ile 1 arasında değişir.

-Rogers ve Tanimoto: Eşleşmeyen verileri iki kat (double) ağırlıklandırır.

-Russel ve Rao: Eşleşenleri ve eşleşmeyenleri eşit ağırlıklandırır. İkilemler için yaygın olarak kullanılan bir benzerlik ölçüsüdür.

-Sokal ve Sneath-1: Eşleşen verileri iki kat (double) ağırlıklandırır.

-Sokal ve Sneath-2: Eşleşmeyen verileri iki kat ağırlıklandırırken, yokluklar (absence) gözardı edilir.

-Sokal ve Sneath-3: Eşleşenlerin eşleşmeyenlere oranını gösteren bir indekstir.

-Sokal ve Sneath-4: Tahmin ortalaması tek bir yönde alınarak, bir maddedeki özelliğin, diğer maddedeki değerle eşleşmesi şeklinde açıklanabilecek olan koşullu olasılıktır.

-Sokal ve Sneath-5: Pozitif ve negatif çiftlerin koşullu olasılığının geometrik ortalamasının karesidir. 0 ile 1 arasında değişir.

-Yule Q: Goodman ve Kruskal'ın Gamma'sının 2x2'lik tablolara özel şeklidir ve -1 ile 1 arasında değişir.

Hangi uzaklık ölçümünün seçileceğine karar vermede araştırmacıların dikkat etmesi gereken noktalar şöyle özetlenebilir [1].

Farklı uzaklık ölçümleri ya da değişkenlerin ölçüldüğü ölçek düzeyindeki değişiklikler, farklı kümeleme çözümlerine neden olabilir. Bundan dolayı birkaç farklı ölçümün kullanılması ve elde edilen sonuçların kuramsal yapı ya da daha önceden bilinen örüntülerle karşılaştırılması önerilir.

Değişkenler ilişkili olduğunda (pozitif ya da negatif) Mahalanobis uzaklığı en uygun uzaklık ölçümü olacaktır; çünkü bu ölçüm korelasyonları düzelterek tüm değişkenlerin eşit ağırlıklandırılmasını sağlar.

Eğer araştırmacılar değişkenlerin eşit olmayan (unweighted) şekilde ağırlıklandırılmasını istiyorsa, bu durumda diğer yöntemleri tercih edebilirler.

2.4 Sayıtlar ve Analizin Diğer Gereklilikleri

Çoklu regresyon, varyans analizi gibi tekniklerde birkaç değişken yer alır. Bu değişkenler -bağımsız değişkenler ya da ortak değişkenler (covariates)- bir bağımlı değişkeni yordar ya da açıklarlar. Dolayısıyla da bu yöntemler "bağımlılık yöntemleri (dependence methods)" olarak adlandırılır. Faktör analizi, çok boyutlu ölçekleme ve kümeleme analizi gibi yöntemler ise, değişkenler arasındaki ilişkileri dikkate alır. Genellikle yordama amaçlı kullanılmaz ve bir "p" değeri yoktur. Araştırmacılar analiz sonuçlarını yorumlayarak en iyi modele karar verirler

Tüm modeller analize giren verilerin birbirleriyle ilişkili olmasını gerektirir. Faktör analizi için bunun anlamı korelasyondur. Çok boyutlu ölçekleme ve kümeleme analizi birbirinden farklı ve çeşitli verileri (örn. benzerlik ya da uzaklıklar) kullanabilir. Bu da çok boyutlu ölçekleme ve kümeleme analizinin, faktör analizinden daha esnek olduğu şeklinde yorumlanabilir. Bu yöntemlerin tümünün en önemli sayıltısı, analize giren verinin "geçerli (valid)" olmasına ilişkindir. Daha önce de değinildiği gibi, bu yöntemler bağımlılık yöntemlerinin sahip olduğu kestirisel istatistik mantığına sahip değildir. Dolayısıyla da, sadece girdinin kalitesi ölçüsünde iyi ya da kötü sonuçlar üretirler Kümeleme analizi, çok boyutlu ölçekleme gibi teknikler, yukarıda da değinildiği gibi, örneklemden evrene ilişkin parametrelerin kestirildiği, örneklemin evrenin temsilcisi olma olasılığının değerlendirildiği bir anlam çıkartıcı istatistik tekniği değildir. Kümeleme analizi, bir grup gözlemin yapısal özelliklerini nicelleştirme yöntemidir. Güçlü matematiksel temellere sahip olmakla birlikte, güçlü istatistiksel temellere sahip değildir. Normallik, doğrusallık ve varyansların homojenliği gibi diğer teknikler için önemli olan sayıtlar, kümeleme analizinde çok az etkilidir. Ancak araştırmacıların odaklanması gereken iki kritik nokta vardır. Bunlar, "örneklemin temsiliyeti" ve "değişkenler arasındaki çoklu bağlantı (multicollinearity)" problemidir [1].

2.4.1 Örneklemin Temsiliyeti: Çok ender olarak araştırmacılar kümeleme analizinde kullanmak için evrenin tam sayısına sahiptirler. Genellikle örneklem üzerinde çalışılır ve bu grupların evrenin yapısını temsil etmesi ve araştırmacıların da bu temsiliyeye

güvenmesi gerekir. Araştırmacılar kümeleme analizinin sadece örneklemin temsiliyeti oranında iyi sonuçlar vereceğinin farkında olmalıdırlar. Bu nedenle de tüm çabalar, örneklemin evreni temsil etmesi ve ancak bu sağlandığında elde edilen sonuçların evrene genellenebileceği üzerinde yoğunlaşmalıdır [1].

2.4.2 Değişkenler Arası Çoklu Bağlantı: Çoklu bağlantı problemi çok değişkenli analizlerde, aralarında çoklu bağlantı problemi olan değişkenlerin gerçek etkilerinin anlaşılmasını güçleştirdiğinden dolayı önemli bir sorundur. Ancak kümeleme analizindeki etkisi farklıdır; çünkü çoklu bağlantı aslında bir çeşit dolaylı ağırlıklandırma oluşturur. Örneğin, bir grup cevaplayıcının 10 değişken dikkate alınarak kümelendiği varsayalım. Bu 10 değişken bir hizmete yönelik tutum ifadeleri olsun. Çoklu bağlantı incelendiğinde, iki grup değişken belirlensin. İlk grup sekiz ifadeden; diğer grup ise, geriye kalan iki ifadeden oluşsun. Eğer bizim amacımız gerçekten cevaplayıcıları hizmet boyutlarına (bu örnekte iki grup değişken tarafından temsil edilmektedir) göre gruplamaksa, orijinal 10 değişkenin kullanılması yanıltıcı olacaktır. Çünkü her değişken küme içerisinde eşit ağırlıklandırılacaktır; oysaki ilk boyut, diğer boyuta oranla benzerlik ölçümünü etkileme konusunda dört kat daha fazla (bir boyutta iki madde varken, diğer boyutta sekiz madde olduğundan) şansa (olasılığa) sahiptir. Sonuç olarak benzerlik, iki maddeye sahip ikinci boyuta göre, sekiz maddeye sahip ilk boyut tarafından ağırlıklı olarak etkilenecektir [1].

Kümeleme analizinde çoklu bağlantı durumu, araştırmacı tarafından gözlenemeyen ancak analizi etkileyen bir ağırlıklandırma işlemi olarak rol oynar. Bu nedenle araştırmacılar kümeleme analizinde kullanılan değişkenler arasında önemli bir çoklu bağlantı probleminin olup olmadığını mutlaka incelenmeli ve eğer çoklu bağlantı problemi bulunursa, ya her bir grupta yer alan değişken sayısının eşitlenmesi ya da Mahalanobis uzaklığı gibi bu korelasyonu dengeleyecek bir uzaklık ölçüsünün kullanılması gerekir [1].

Tatlídil (1996) ise kümeleme analizinin sayıltılarına ilişkin olarak, kümeleme analizinde de diskriminant analizinde olduğu gibi normallik sayıltısı bulunmakla birlikte, bu sayıltının prensipte kaldığını ve uzaklık değerlerinin normalliğinin yeterli görüldüğünü ifade etmektedir. Ayrıca bu analizde kovaryans matrislerine ilişkin herhangi bir sayıltı bulunmadığını da belirtmektedir.

Başka bir tanımda kümeleme analizinin tahminden çok tanımlama ile, değişkenlerden çok bireylerle, kriter-tahmin değişkenleri arasındaki ilişkilerden çok, tüm veri setindeki ilişkilerle ilgilendiğini ifade etmektedir. Ayrıca bu analizde değişkenler

arasındaki ilişkinin doğrusal olması şartı bulunmadığını ve seçilen kümeleme yöntemine göre sınıflamalı, sıralamalı, aralıklı veya karma verilere uygulanabileceğini ifade etmektedir.

Sonuç olarak kümeleme analizinin sayıltılarına ilişkin tartışmalar şöyle özetlenebilir:

- Bir grup gözlemin yapısının nicelleştirilmesine yarayan objektif bir teknik olan kümeleme analizi, bir kestirisel istatistik tekniği değildir.
- Uç değerlerin incelenmesi önemlidir.
- Verilerin ve kümelerin çoklu bağlantı problemi açısından incelenmesi önemlidir.
- Seçilen örneklemin evreni temsil etmesi ve eksik örnekleme (undersampling) yapılmaması önemlidir. Evrene genelleme garanti altına alınmalıdır.

2.5. Kümelerin Değerlendirilmesi

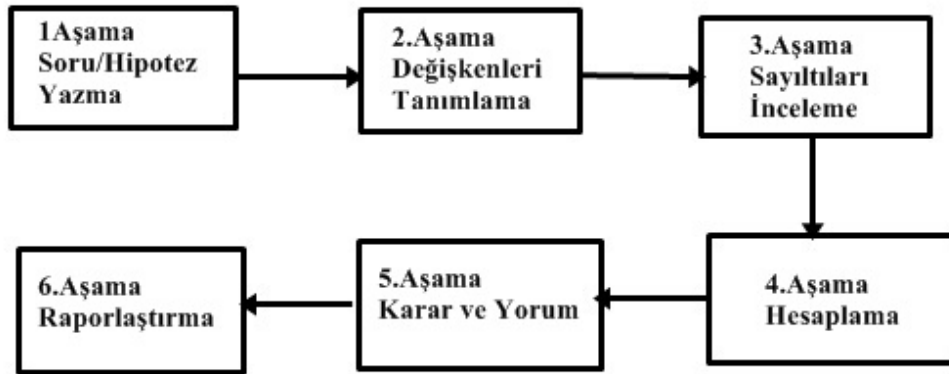
Kümeleme analizinde kullanılan farklı algoritmalarla veya aynı algorithmada alınan farklı parametre başlangıç değeri ile farklı birçok kümeleme yapısı elde edilebilir. Aynı veri kümesinden elde edilen farklı kümeleme yapılarının anlamlı olup olmadığı kümeleme geçerliliği olarak adlandırılır [18,19] Herhangi bir kümeleme yapısı içermeyen bir veri kümesine uygulanan bir kümeleme algoritması sonucu elde edilen kümeleme yapısı anlamsız olacaktır. Kümeleme sonucu ileri aşama analizlere başlanmadan önce veride bir kümeleme yapısının varlığından emin olmak için bazı testler uygulanabilir.

Kümeleme geçerlilik kriterleri kümeleme algoritmalarının tipine bağlı olarak üç kategoride toplanmaktadır [20]. Bunlar dışsal (external) kriterler, içsel (Internal) kriterler ve oransal (Relative) kriterlerdir. Bir veri kümesi X , bu veri kümesinden elde edilen kümeleme yapısı C ile gösterilsin. Dışsal (external) kriterlerde veri hakkında sahip olunan ön bilgi ile kümeleme algoritması sonunda elde edilen C kümeleme yapısı karşılaştırılır. Örneğin kümeleme sonunda elde edilen gözlemlerin küme etiketleri ile daha önceden bilgi sahibi olunan gözlemlerin kategori etiketleri karşılaştırılır. İçsel (Internal) kriterlerde ise veri hakkında bir ön bilgi kullanılmaz. Kümeleme yapısının geçerliliği elde edilen yapı içerisinde değerlendirilir. Örneğin X veri kümesinin C kümeleme yapısı benzerlik matrisi kullanılarak değerlendirilir. Oransal (Relative) kriterlerde ise farklı kümeleme algoritmaları ile elde edilen kümeleme yapıları birbiri arasında veya aynı kümeleme algoritması fakat farklı parametre değerleri ile elde edilen kümeleme yapıları birbirleriyle karşılaştırılır. Dışsal (external) ve içsel (Internal) kümeleme geçerlilik kriterleri istatistiksel

hipotez testleri ile ilişkilidir [21], kümeleme geçerlilik kriterlerinde ortak olarak kullanılan sıfır hipotezlerini üç grupta toplamıştır.

2.6 Araştırma Deseni

Bireyleri, objeleri benzer özelliğe sahip homojen gruplara ayırarak sıralamayı amaçlayan çok değişkenli keşfedici bir istatistik olan küme analizi, temelde korelasyonel araştırma desenlerinde kullanılan bir veri analizi tekniğidir. Korelasyonel araştırmalar, iki ya da daha çok değişken arasındaki ilişkinin, herhangi bir şekilde bu değişkenlere müdahale edilmeden incelendiği araştırmalardır “Şekil 2.6”. Korelasyonel araştırmalar, değişkenler arasındaki ilişkilerin açığa çıkarılmasında, bu ilişkilerin düzeylerinin belirlenmesinde etkili ve bu ilişkilerle ilgili daha üst düzey araştırmaların yapılması için gerekli ipuçlarını sağlayan araştırmalardır.



Şekil 2.6. Kümeleme Yöntentemi Akış Şeması

Yukarıdaki şekil incelendiğinde, birinci aşamanın, araştırma sorularının ya da hipotezlerinin yazılması olduğu görülmektedir. Kümeleme analizinin kullanımını gerektiren bir araştırma sorusu, "Üniversite öğrencileri istatistik başarı puanı, istatistik kaygı puanı ve istatistik tutum puanlarına göre nasıl gruplandırılabilir?" şeklinde yazılabilir.

İkinci aşamayı, bireylerin veya objelerin gruplandırmasında temel alınan değişkenlerin tanımlanması oluşturur.

Üçüncü aşama analize ilişkin sayıtların test edilmesidir. Sayıtlar ve Analizin Diğer Gereklilikleri" başlığı altında bu sayıtların neler olduğuna değinilmiştir.

Dördüncü aşama ilgili işlem adımlarının izlenmesi ile analizin gerçekleştirilmesidir.

Beşinci aşamada araştırma sorusunun cevaplanmasına ya da hipotezin desteklenip desteklenmeme durumuna ilişkin bir karara varılır.

Altıncı ve son aşamada ise, analiz sonuçlarının raporlaştırılarak sunulmasıdır.

3.UYGULAMA

3.1 Uygulamanın Amacı

Kümeleme Analizi yöntemleri değişkenleri tasnif etme ve sınıflandırma işlemleri için kullanılmaktadır. Bu yöntem kullanılırken benzerlik ve farklılık ölçülerinden yararlanılmaktadır. Bu çalışmanın amacı ise; kümeleme analizi ile Elazığ Organize Sanayi Bölgesi (OSB)nde bulunan F1, F2, F3, F4, F5 ve F6 fabrikalarında çalışan işçilerden oluşan gürültüye bağlı işitme kaybı baz alınarak, bu fabrikaları gürültü düzeylerine göre sınıflandırmaktır. Küme sayısı önce 5, sonrada 2 olarak belirlenip ikinci işlemde daha genel bir gruplandırmak amaçlanmıştır.

3.2 Uygulama Verileri ve Düzeni

Elazığ Organize Sanayi Bölgesinde bulunan F1, F2, F3, F4, F5 ve F6 fabrikalarında çalışan 197 işçinin, Fırat Tıp Merkezi Kulak Burun Boğaz Kliniği'nde yapılan odyometrik testler ile işitme düzeyleri ölçülmüştür. Odyometri sonucuna göre sağ ve sol kulaklar ayrı ayrı olarak 250, 500, 1000, 2000, 4000, 6000, 8000, 10000, 12000, 14000 ve 16000 Hz'lik sonuçları dikkate alınmıştır. Odyometri sonuçları sürekli değişkendir. Öncelikler verileri kesikli değişkene çevirmek için veriler kategorize edilmiştir. Her değişken 6 kategoriye ayrıldı. Her test değeri 20 eşit aralığa bölündü. (Tablo 3.1)

Tablo 3.1 Odyometrik Verilerin Kategorize Edilmesi

Kategori →	1	2	3	4	5	6
Odyometri Sonucu (Hz) ↓						
SA250	0-20	21-40	41-60	61-80	81-100	101-120
SA500	0-20	21-40	41-60	61-80	81-100	101-120
SA1000	0-20	21-40	41-60	61-80	81-100	101-120
SA2000	0-20	21-40	41-60	61-80	81-100	101-120
SA4000	0-20	21-40	41-60	61-80	81-100	101-120
SA6000	0-20	21-40	41-60	61-80	81-100	101-120
SA8000	0-20	21-40	41-60	61-80	81-100	101-120
SA10000	0-20	21-40	41-60	61-80	81-100	101-120
SA12000	0-20	21-40	41-60	61-80	81-100	101-120
SA14000	0-20	21-40	41-60	61-80	81-100	101-120
SA16000	0-20	21-40	41-60	61-80	81-100	101-120
OS250	0-20	21-40	41-60	61-80	81-100	101-120
OS500	0-20	21-40	41-60	61-80	81-100	101-120
OS1000	0-20	21-40	41-60	61-80	81-100	101-120
OS2000	0-20	21-40	41-60	61-80	81-100	101-120
OS 4000	0-20	21-40	41-60	61-80	81-100	101-120
OS 6000	0-20	21-40	41-60	61-80	81-100	101-120
OS 8000	0-20	21-40	41-60	61-80	81-100	101-120
OS 10000	0-20	21-40	41-60	61-80	81-100	101-120
OS 12000	0-20	21-40	41-60	61-80	81-100	101-120
OS 14000	0-20	21-40	41-60	61-80	81-100	101-120
OS 16000	0-20	21-40	41-60	61-80	81-100	101-120

Çalışmaya 197 işçi katılmıştır. 197 işçinin sağ ve sol kulaklarında 22 ayrı odyometri ölçümü yapılmıştır. Eğer analiz bu şekilde yapılacak olursa gruplandırma tek tek işçileri ve bunların her bir ayrı ölçümleri dikkate alınarak yapılacağından amaca çok cevap vermeyecektir. O yüzden herbir ölçümün fabrikalara göre ayrı ayrı ortalamaları alınarak yeni bir tablo oluşturulmuştur. (Tablo 3.2)

Tablo 3.2. Fabrikalara Göre Odyometri Ölçümlerinin Ortalamaları

Fabrika Adı →	F1	F2	F3	F4	F5	F6
Odyometri Sonucu(Hz) ↓						
SA250	25,00	29,21	23,86	22,40	25,00	20,66
SA500	19,87	23,02	20,45	18,20	17,75	15,33
SA1000	16,42	14,34	15,45	14,20	13,00	11,33
SA2000	14,80	17,10	18,86	13,60	12,50	11,33
SA4000	24,48	38,28	34,31	17,60	27,25	28,00
SA6000	27,59	39,86	33,40	32,40	26,00	22,66
SA8000	29,15	32,50	30,00	30,40	25,25	23,66
SA10000	25,90	31,44	30,45	24,90	20,75	23,66
SA12000	29,81	35,55	29,82	24,20	24,50	26,33
SA14000	31,38	41,11	33,55	28,60	25,85	33,80
SA16000	45,12	49,13	48,86	38,00	45,70	45,07
OS250	23,18	26,97	18,64	25,80	23,75	18,33
OS500	20,19	22,37	16,82	19,40	20,00	16,33
OS1000	14,81	16,18	13,41	11,80	17,00	12,67
OS2000	14,61	18,03	17,96	9,60	16,50	10,67
OS 4000	23,83	42,76	35,91	17,40	33,50	25,67
OS 6000	31,77	44,34	32,95	32,40	25,75	23,67
OS 8000	26,36	33,95	27,27	27,40	26,25	26,00
OS 10000	25,06	31,05	25,50	23,80	26,25	25,00
OS 12000	26,61	35,60	33,23	22,80	28,00	30,67
OS 14000	34,12	42,18	35,68	30,40	29,05	31,40
OS 16000	45,60	50,00	46,68	41,40	43,10	45,67

Küme sayısı önceden belli olduğu için hiyerarşik kümeleme yöntemi kullanılması gerekmektedir. Hiyerarşik kümeleme yöntemlerinden Ward's Yöntemi uygulanarak 5 kümeye ayrılması amaçlanmıştır. Daha sonra ise 2 kümeye ayrılarak daha genel bir gruplandırma yapılmıştır. Verilerin analizi ve incelenmesi SPSS paket programında gerçekleştirilmiştir. Sonuçlar tablo ve dendogram grafikleri ile sunularak işitme kaybına bağlı olarak fabrikaların gürültü düzeyleri sınıflandırılmıştır.

3.3 Örneklem Seçimi

Çalışma Elazığ ili organize sanayi bölgesinde (OSB) bulunan iş yerlerinde gürültü düzeyi 95 dB'in üstünde olan ve en az on işçinin çalıştığı işyerlerindeki işçiler üzerinde yapılmıştır. Bu altı iş yerinden ikisi mermer, üçü tekstil, biri ise un fabrikasıdır. Bu altı iş yerinde çalışan toplam 370 işçi vardır. Bu 370 işçinin 206'sı (%70,3) odyometri ölçümü yaptırmıştır. Bu 206 kişinin; 9'u daha önceden geçirilmiş kulak rahatsızlığı nedeniyle çalışmaya dahil edilmedi. Böylece çalışmaya dahil edilmek için gerekli tüm kriterleri karşılayan 197 işçi araştırmanın kapsamına alındı.

Çalışma Yapılan Fabrikalar:

1. F1
2. F2
3. F3
4. F4
5. F5
6. F6

3.4 Gürültülü İş Yerlerinde İşitme Kaybı

Gürültüye bağlı işitme kaybı; dünyada ve ülkemizde sık görülen ve geriye dönüşümsüz olan önemli bir meslek hastalığıdır. Kanada ve Amerika Birleşik Devletleri (ABD)'nde en sık görülen on meslek hastalığı arasında yer alır. Amerika'da 11 milyon çalışan, iş yeri ortamındaki gürültünün potansiyel zararlı etkisi ile karşı karşıya kalmaktadır ve ABD popülasyonunun yaklaşık %10'unda çeşitli derecelerde işitme kaybı vardır. İsveç'te toplam çalışanların %9'u zararlı düzeydeki gürültüye sürekli olarak maruz

kalmaktadır. Son yıllarda yapılan çalışmalara göre; ülkemizde mesleki gürültü nedenli işitme kaybı olanların sayısının 200 bini aştığı belirtilmektedir.

Gürültü; işitme kaybı, acı hissi, sinir ve dolaşım sistemi bozuklukları, hormonal dengenin bozulması gibi fiziksel etkilerinin yanı sıra rahatsızlık, uyumsuzluk, uykuya geç başlama, uyuyamama, ve yorgunluk, huzursuzluk, konsantrasyon bozukluğu, sinirlilik, libido azalması gibi psikolojik etkilere de yol açar. Ayrıca gürültü konuşmaları engelleyerek iş güvenliğini de azaltır .

Mesleki gürültüye bağlı işitme kaybı; gürültüye maruz kalma süresinin yanı sıra, gürültünün tipi (sürekli ya da vuruş şeklinde) frekansı ve şiddetine de bağlıdır. Gürültüye bağlı işitme kaybı; şiddeti 90 dB'in üzerindeki seslerde oluşur ve işitme kaybı bilateralidir. İşitme kaybı ilk olarak 4 000 Hz frekansında oluşur. Oluşan işitme kaybı genç yaşlarda görülür ve sensörinöral tipte olup geri dönüşümsüzdür. Genel olarak 90 dB'de izin verilen maksimum maruz kalma zamanı 8 saattir. Bu zaman her 5 dB'lik artışta yarıya iner ve her 5 dB'lik azalışta iki katına çıkar. Gürültünün etkisinde kalınan süre ifadesi; kişinin sürekli olarak gürültünün etkisi altında kaldığı süreyi ve aralıklı olarak gürültünün etkisinde kaldığı toplam yılları kapsar. Yani belirli yükseklikteki sesin etkisinde belirli bir süre kalmak işitme kaybına yol açacağı gibi belirli bir süre zararlı olmayacak düzeydeki sesin etkisinde çeşitli aralıklarla yıllarca kalmak da işitme kaybına yol açabilir.

Uzun süre şiddetli gürültüye örneğin 90 dB'nin üzerindeki seslere maruz kalan kişilerde geçici ve sürekli olmak üzere iki tür işitme kaybı olabilir. Geçici işitme kayıpları, uzun süre gürültüye maruziyet sonucunda ortaya çıkan ve belli bir süre dinlendikten sonra iyileşebilen işitme kayıplarıdır. 90 dB'lik bir gürültüye 100 dakika maruz kalma sonucunda ortaya çıkan yaklaşık 18-20 dB'lik bir işitme kaybının ortadan kalkabilmesi için gerekli olan iyileşme süresi yaklaşık olarak 1000 dakikadır. Yani ortaya çıkan işitme kaybının iyileşmesi için maruz kalma süresinin yaklaşık 10 katı kadar iyileşme süresine ihtiyaç vardır. Gürültü düzeyi arttıkça oluşan işitme kaybının arttığı ve iyileşme süresinin uzadığı bilinmektedir. Gürültülü ortamlarda çalışan kişilerin bahsedilen iyileşme sürelerine sahip olmaları mümkün değildir. Gürültüye 8 saatlik maruziyet sonucunda en fazla 16 saatlik bir dinlenme süresi vardır. Böylece bu işitme kayıpları yığılmalı bir biçimde oluşarak sürekli işitme kayıplarına yol açmaktadır.

Gürültüden korunmak için alınacak önlemler; gürültüyü kaynağından azaltmak, gürültüyü kaynak ile alıcı arasındaki yolda azaltmak ya da gürültüyü gürültüye maruz kalan kişide engellemek şeklinde olabilir. Gürültüyü, gürültüye maruz kalan kişide

engellemek için; gürültüye maruz kalan kişiyi tecrid etmek, kişisel koruyucu kullanmak, gürültüye maruziyet süresini azaltmak (rotasyonla çalışmak), iş programını değiştirmek gerekir. İşveren, çalıştırdığı işçiler için işyerinde güvenlik araçlarını bulundurmamakla yetinemez. Bu araçları çalışma sırasında işçilere vermek ve işçiler tarafından kullanıldığını denetlemekle de yükümlüdürler.

Gürültüye bağlı mesleki işitme kaybının dünyada sık görülen önemli bir meslek hastalığı olması nedeniyle birçok ülkede konu ile ilgili yasal düzenlemeler oluşturulmuş ve maruz kalınabilecek gürültü miktarı ile ilgili limitler getirilmiştir. Ancak; bu yasal düzenlemelere rağmen düzenli olarak; işe giriş ve periyodik odyometrik ölçümler yapılmamakta, mesleki gürültüye maruziyet düzeyleri ve yol açtığı işitme kaybı ya da sağlık sorunları bilinmemekte ve iş yerlerinin gerekli önlemleri alıp almadığı yeterince incelenmeden denetimsiz bir şekilde gürültüye maruziyet sürmektedir.

Elazığ ilinde de ülke geneline benzer şekilde, gürültülü iş yerlerindeki gürültü miktarı ve bunun sağlık üzerine etkileri ile ilgili; eğitim, korunma önlemleri, denetim, işe giriş ve periyodik muayenelerin yetersiz olduğu düşüncesi araştırmacıları bu konuya yönlendirmiştir. [22].

3.5 Odyometrik Ölçüm

Odyometri işitme testidir. İşitme kaybı farklı türlerdedir ve saptanması için farklı teknikler uygulanır. Kısmi ya da tam işitme kaybı durumu mevcuttur. İşitme kaybına neden olan rahatsızlıklar arasında kulak iltihabı, kızıl, grip, gen faktörü, yaralanma, yaşlılık ve nedensiz işitme kayıpları vardır. Her işitme kaybında farklı özellikler mevcuttur bu nedenle doktorlar farklı uygulama yapmaktadırlar.

İnsan kulağı 16-20000 Hz arasındaki seslere karşı duyarlı olmasına rağmen her frekansı aynı şiddette algılayamamaktadır. Saf ses odyometrisi yapılırken insan kulağının duyarlı olduğu, 250, 500, 1000, 2000, 4000, 6000 Hz'lerdeki işitme eksikliklerine bakılır.

İşitme incelemesinde hasarın derecesi ve işitme kaybının tipi ortaya konmalıdır. Ayrıca kaybın kalıcı, düzelebilir ya da ilerleyip ağırlaşan nitelikte mi olduğu ve tıbbi, fizik, cerrahi ya da protez tedavilerden hangisinin en iyi sonucu vereceği saptanmalıdır. Bunun için hastanın işitme yelpazesinin belirlenmesi, daha az işitilen ya da işitilmeyen titreşimlerin bulunması gerekir. Olgulara göre düşük, orta ya da yüksek titreşimler söz konusu olacaktır. Hasta genellikle işittiği ve işitemediği sesleri işaret ederek hekimi yönlendirir. Her titreşim için işitme kaybı miktarı saptanır. Fizyologlarca işitme eşiği

olarak tanımlanan, işitmeyi uyarabilecek güçteki en düşük ses şiddeti bulunur. Başka gürültülerin işitme eşiğinin saptanmasını etkilememesi İçin inceleme sessiz bir ortamda yapılır; incelemeden geçecek hasta sestten yalıtılmış bir kabine alınır. Kabin içindeki kulaklıklar dışındaki odyometreye bağlıdır. Hastanın her İki kulağına sırasıyla belirli titreşimlerde saf sesler verilir. En düşük titreşimlerden (125-cps [saniyedeki devir]) en yüksek titreşimlere (16.000 cps [saniyedeki devir]) doğru gidilir. İncelenen her ton için odyometrenin bir düğmesine basarak sesin desi-bel (dB) üzerinden ölçülen şiddeti kademeli biçimde artırılır. Hasta her ton için elinde bulunan düğmeye basarak ya da elini kaldırarak algılamının başladığı düzeyi belirtir. Bu düzey söz konusu ton elinde bulunan hastanın işitme eşiğini oluşturur. Her iki kulak için toplanan tek tek değerler odyogram denen bir grafik üzerinde gösterilir. Tek tek noktaların birleştirilmesiyle elde edilen eğriler işitme kaybının tipi ve şiddeti hakkında kesin tanı konmasını sağlar.

3.6 Sesin Şiddeti Nasıl Ölçülür

Sesin şiddet birimi desibel (dB)'dir. Bir sesin şiddedini belirtirken birim olarak dB kullanılır. İnsan kulağının duyabildiği en küçük ses 0 dB olarak kabul edilir. Bu oran logaritmik olarak artar. Yani 20 dB, 10 dB'den 10 kat daha şiddetli, 40 dB, 10 dB'den 1000 kat daha şiddetlidir. İnsanın 0 ila 180 dB arasındaki sesleri duyduğu kabul edilir. Bazı seslerin şiddeti şu şekilde belirtilebilir.

- 0 dB İnsanın duyabildiği en düşük ses şiddeti
- 30 dB Fısıltı ile konuşma
- 60 dB Normal konuşma veya daktilo sesi
- 90 dB Kamyon sesi veya çim biçme makinesi sesi
- 100 dBAsfalt delme makinesi
- 115 dB Konser veya barlarda yüksek sesli müzik
- 140 dB Jet uçağı sesi

Genel olarak 85 dB üzerindeki sesin kulağı zararlı olacağı kabul edilir.

3.7. Kümeleme Yöntemlerinin Belirlenmesi

Çalışmada Elazığ ili organize sanayi bölgesinde (OSB) bulunan iş yerlerinde, gürültü düzeyi 95 dB'in üstünde olan ve en az on işçinin çalıştığı 6 fabrikada, çalışan işçilerden oluşan işitme kaybına göre fabrikalar gruplandırılacaktır. Fabrikalar gruplandırılırken grup sayısı başta belirlenebilir. Birde gruplandırma yapılırken fabrikaların benzerlikleri dikkate alınacaktır. Bu özelliklerinden dolayı hiyerarşik kümeleme yöntemi kullanılmalıdır. Ayrıca örneklem sayımızın da 250'den az olması (197) hiyerarşik kümeleme yöntemini uygun kılmaktadır. Küme içindeki homojenlik, kümeler arasındaki heterojenlik maksimum olan kümeler oluşturacağımız için Hiyerarşik kümeleme yöntemlerinden Ward's Kümeleme yöntemini uygulanacaktır. Bu yöntemle önce 5 gruba ayrılacak, daha sonra ise 2 gruba ayrılarak genel bir kümeleme yapılacaktır.

3.8. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların Beşli Sınıflandırılmaları

Elazığ ili organize sanayi bölgesinde (OSB) bulunan iş yerlerinde, gürültü düzeyi 95 dB'in üstünde olan ve en az on işçinin çalıştığı 6 fabrikada, çalışan işçilerden oluşan işitme kaybına göre fabrikalar öncelikle 5 kümeye ayırma işlemi gerçekleştirilmiştir. Kümeleme işlemi yapılmadan önce, 197 işçinin (her biri için ayrı ayrı) sağ ve sol kulaklarında yapılan 22 ayrı odyometri ölçümün, fabrikalara göre ayrı ayrı ortalamaları alınarak yeni bir tablo oluşturulmuştur (Tablo 3.2). Bulunan bu ortalamalara göre kümeleme işlemi yapılmıştır. Kümeleme işlemi aşamalı olarak devam etmiştir (Tablo 3.3). Kümeleme sonucu tablo ve dendogramlarla sunulmuştur.

Tablo 3.3 Ward's Yöntemi İle Oluşturulan Beşli Birleştirme Tablosu

Birleştirme Tablosu

Aşama	Birleştirilmiş Kümeler		Katsayılar	İlk Kümeleşmenin Görüldüğü Aşamalar		Sonraki Aşamalar
	Küme 1	Küme 2		Küme 1	Küme 2	
1	1	6	2,000	0	0	2
2	1	5	4,667	1	0	4
3	2	3	9,167	0	0	5
4	1	4	15,500	2	0	5
5	1	2	30,167	4	3	0

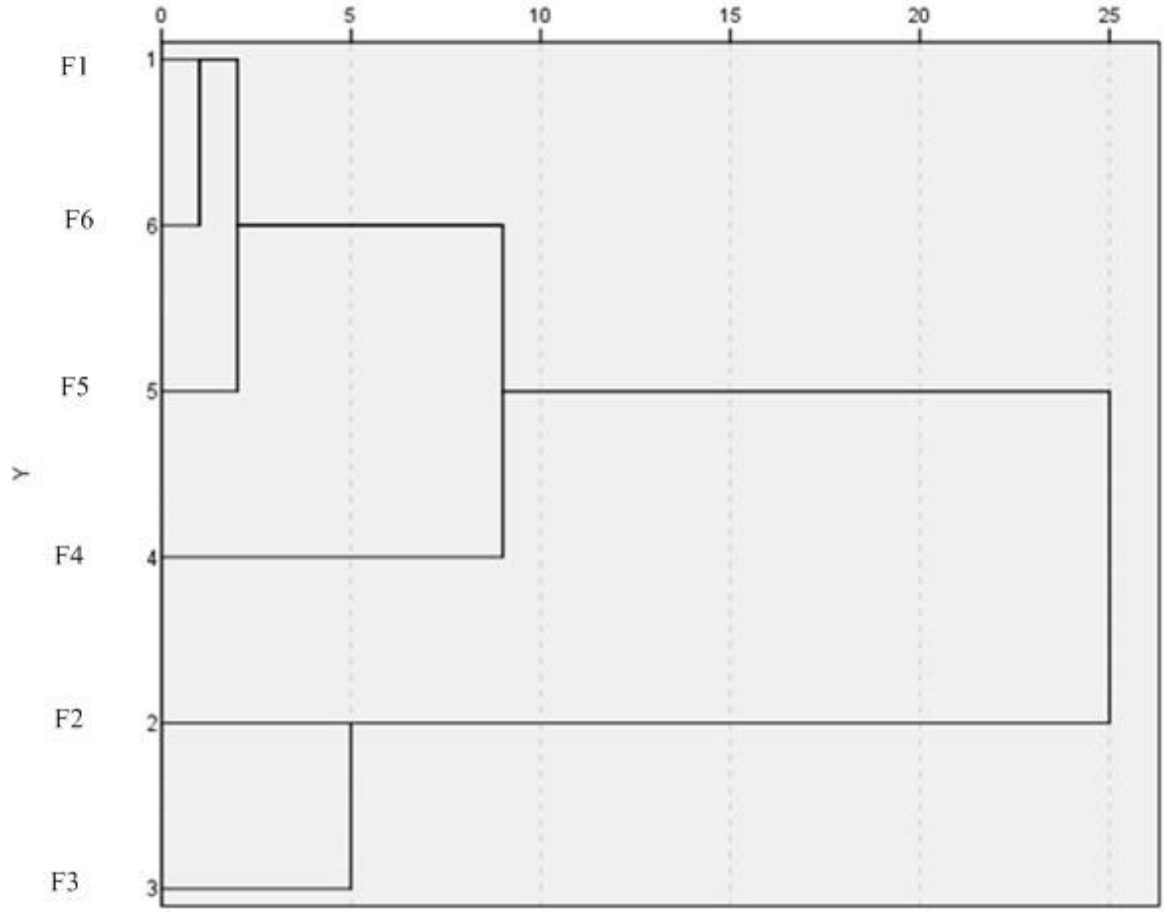
Kümeleme işlemi sonucu, gürültü bakımından birbirine en yakın fabrikalar F1 ile F6 çıkmıştır. F1 ile F6'nın gürültü düzeyinin aynı şiddette işitme kaybına neden olmalarından dolayı aynı kümeye yerleştirilmişlerdir. Daha sonra F1 ile F5 yakın bulunmuştur. En uzak ise F5 ile F2 görülmektedir (Şekil 3.1). Fabrikaların kümelere göre sınıflandırılması Tablo 3.4 ile verilmiştir.

Tablo 3.4 Ward's Yöntemi İle Oluşturulan Beşli Kümeleme

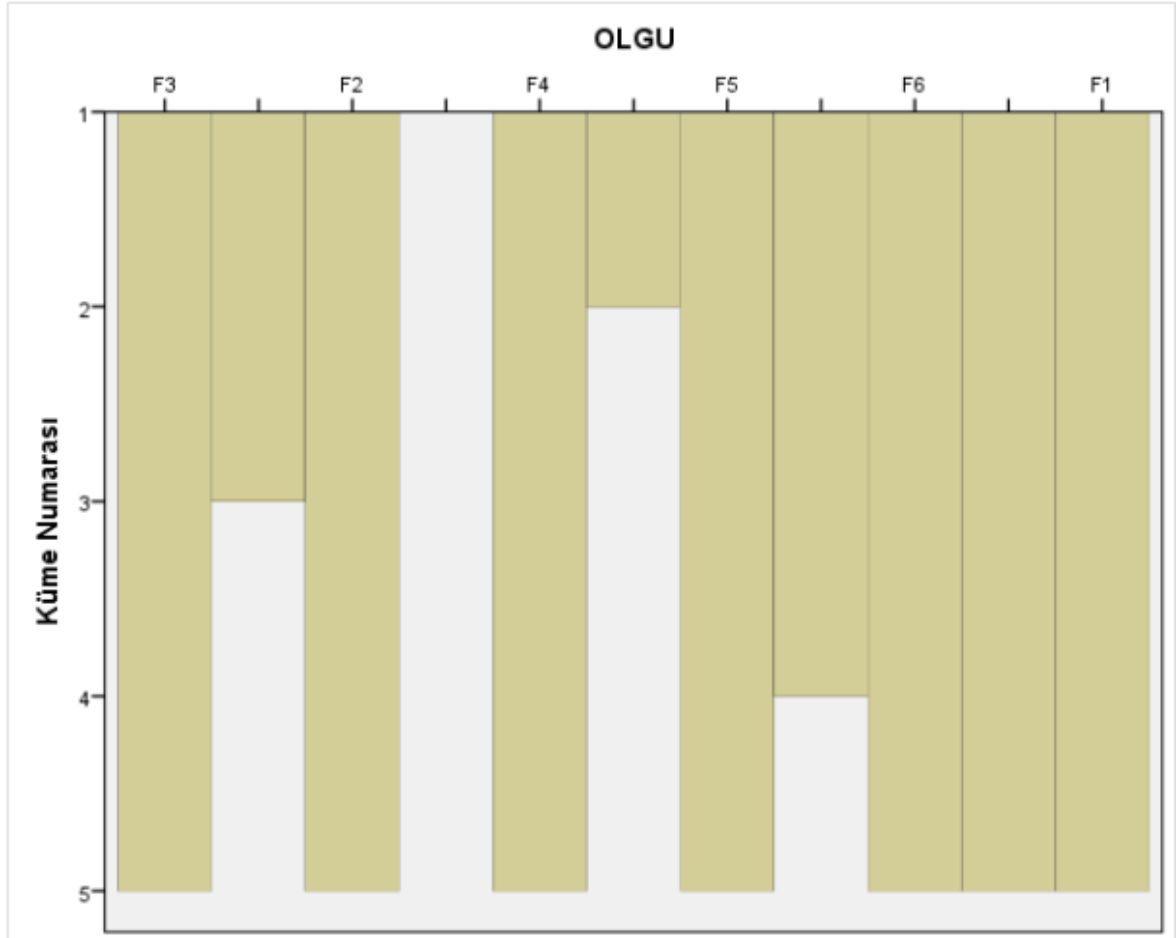
Küme Elemanları

Fabrikalar	5'li Kümeleme
1:F1	1
2:F2	2
3:F3	3
4:F4	4
5:F5	5
6:F6	1

Birleştirilmiş Küme Uzaklıkları Ölçeği



Şekil 3.1 Ward's Yöntemiyle Yapılan Beşli Kümeleme Analizinin Dendogram Grafiği



Şekil 3.2 Ward's Yöntemiyle Yapılan Beşli Kümeleme Analizi

3.9. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların İkili Sınıflandırılmaları

Daha sonra ise ikili bir kümeleme yapılmıştır. Bu kümeleme işleminde küme sayısı daha az olduğundan daha genel bir çalışma olmuştur. Dendogramda da görüldüğü gibi gürültü düzeylerinin neden olduğu işitme kaybı 5'li kümelemeyle aynı. Fabrikaların birbirine olan benzerlikleri veya uzaklıklarından bir değişiklik yoktur. Kümeleme yapılırken ise F3 ile F2 aynı kümeye koyulurken diğer dört fabrika ise aynı küme içinde kalmışlardır (Tablo 3.6)

Tablo 3.5 Ward's Yöntemi İle Oluşturulan İkili Birleştirme Tablosu

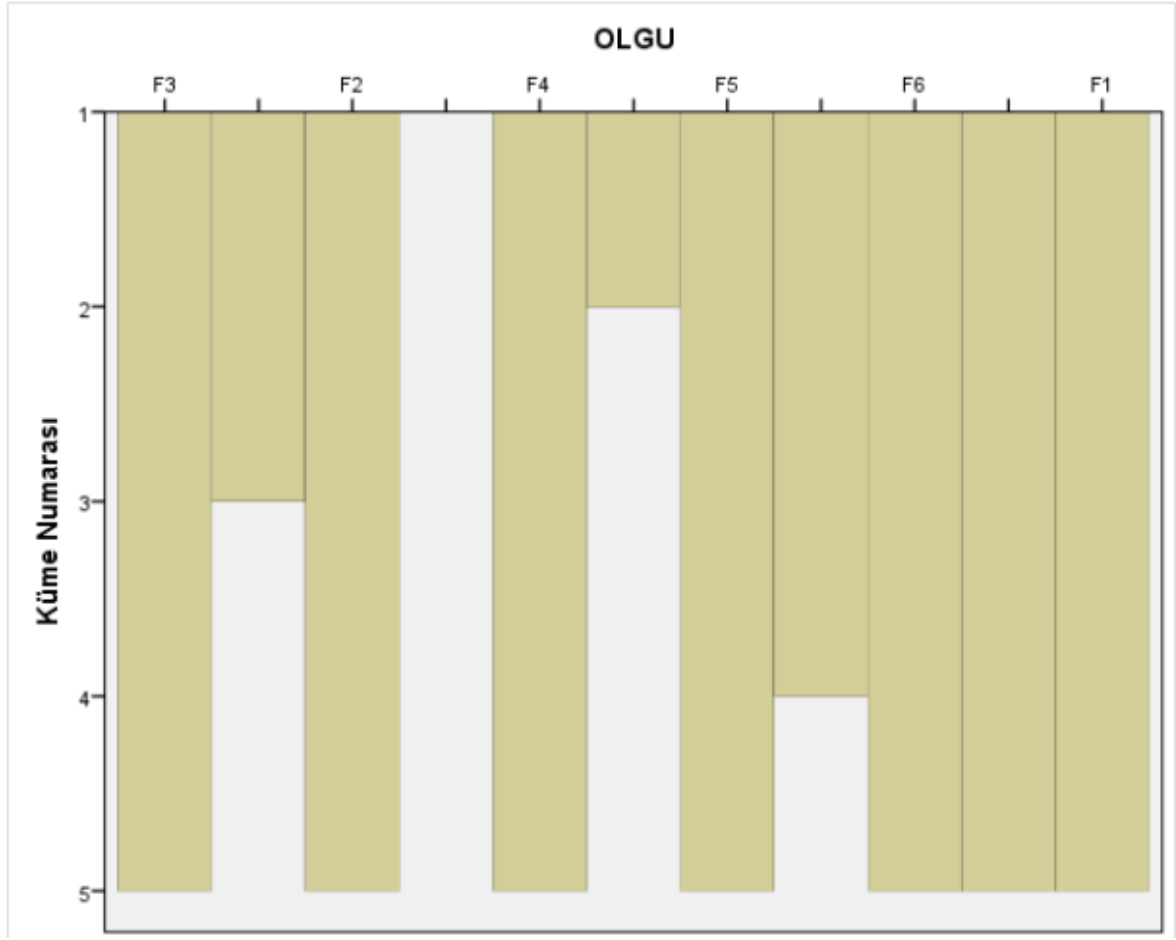
Birleştirme Tablosu

Aşama	Birleştirilmiş Kümeler		Katsayılar	İlk Kümeleşmenin Görüldüğü Aşamalar		Sonraki Aşama
	Küme 1	Küme 2		Küme 1	Küme 2	
1	1	6	2,000	0	0	2
2	1	5	4,667	1	0	4
3	2	3	9,167	0	0	5
4	1	4	15,500	2	0	5
5	1	2	30,167	4	3	0

Tablo 3.6 Ward's Yöntemi İle Oluşturulan İkili Kümeleme

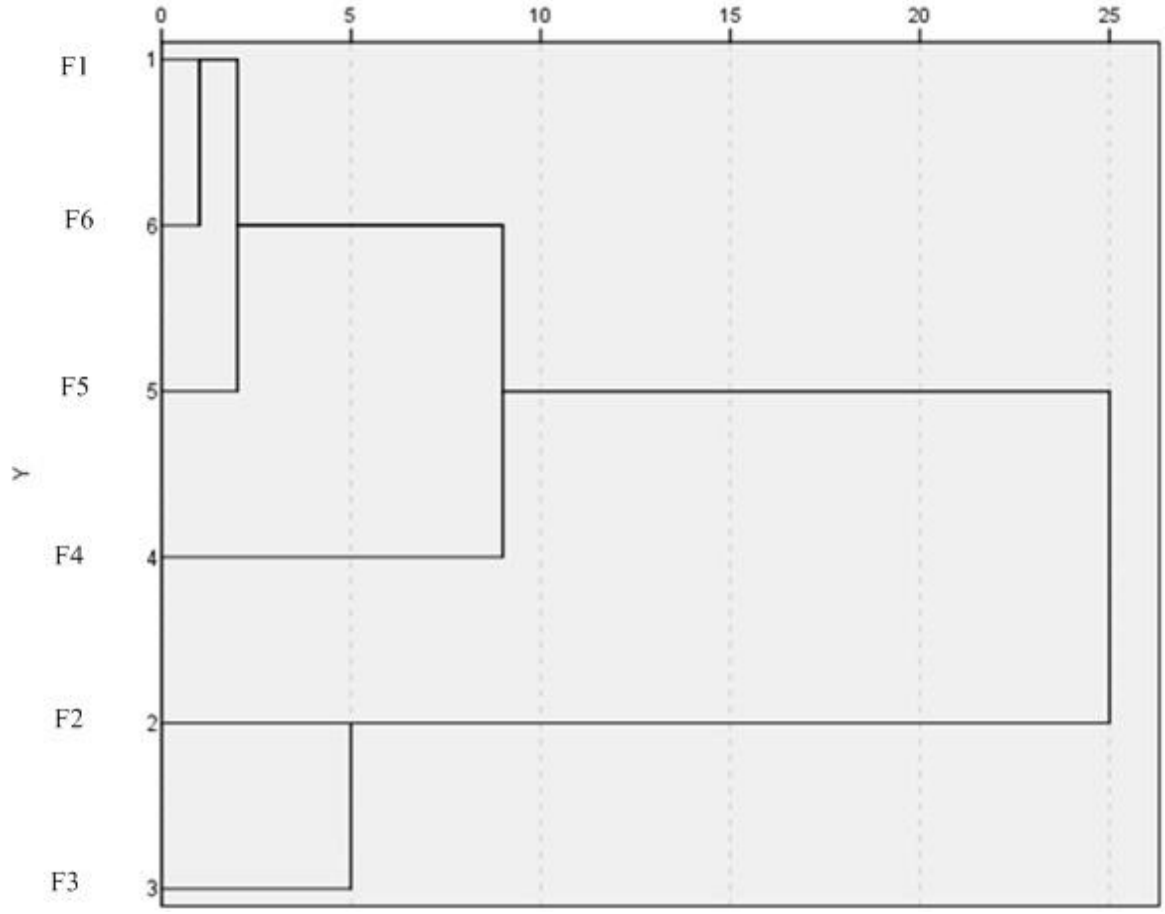
Küme Elemanları

Fabrikalar	2'li Kümeleme
1:F1	1
2:F2	2
3:F3	2
4:F4	1
5:F5	1
6:F6	1



Şekil 3.3 Ward's Yöntemiyle Yapılan İkili Kümeleme Analizi

Birleştirilmiş Küme Uzaklıkları Ölçeği



Şekil 3.4 Ward's Yöntemiyle Yapılan İkili Kümeleme Analizinin Dendogram Grafiği

4.BULGULAR

4.1 Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların Beşli Kümelenmesi

Ward's Kümeleme Analizi ile, Elazığ Organize Sanayi Bölgesi (OSB)nde bulunan F1, F2, F3, F4, F5 ve F6 fabrikalarında çalışan işçilerden oluşan, gürültüye bağlı işitme kaybı baz alınarak, bu fabrikaların gürültü düzeylerine göre beşli kümelenmesi yapılmıştır. Bu kümelenme aşamalar halinde yapılmıştır. Kümeleme işlemi sonucu, gürültü bakımından birbirine en yakın fabrikalar F1 ile F6 çıkmıştır. F1 ile F6'nın gürültü düzeyinin aynı şiddette işitme kaybına neden olmalarından dolayı aynı kümeye yerleştirilmişlerdir. Daha sonra F1 ile F5 yakın bulunmuştur. En uzak ise F5 ile F2 görülmektedir. F1 bir tekstil fabrikası, F6 ise bir un fabrikasıdır. Bu iki fabrikada aynı tip makineler kullanılmamaktadır. Dolayısıyla gürültü düzeyleri de aynı değildir. Bu fabrikalar gürültüyü azaltıcı olarak aldıkları tedbirler ile veya da işçilerin çalışma saatlerindeki değişiklikler ile, işçilerin işitme kayıpları birbirine yakın çıkmıştır. Yine dendograma bakıldığında F5 ile F2'nın işçilerinde oluşan işitme kaybı birbirine oldukça uzak çıkmıştır. F5 bir tekstil fabrikası, F2 ise mermer fabrikasıdır. Mermer kesimleri için kullanılan makinelerin oldukça gürültülü çalışmaları ve yeterli önlemlerin alınmaması işçilerin işitme düzeylerinde oldukça zararlı olmuştur.

Tablo 4.1 Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların Beşli Kümelenmesi

Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
F1 F6	F2	F3	F4	F5

4.2 Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların ikili Kümelenmesi

F1, F2, F3, F4, F5 ve F6 fabrikalarında çalışan işçilerin gürültüye bağlı işitme kayıplarına göre, bu fabrikaların gürültü düzeyleri ikili kümelendirilmiştir. Bu kümelenme sonucu iki mermer fabrikası olan F3 ile F2 aynı kümeye düşer iken, diğer dört fabrikada aynı grup içinde kalmışlardır.

Tablo 4.2. Ward's Kümeleme Yöntemi İle İşitme Kaybına Göre Fabrikaların ikili Kümelenmesi

Küme 1	Küme 2
F1	F2
F4	F3
F5	
F6	

5. SONUÇLAR VE TARTIŞMA

Yapılan çalışmada Ward's Kümeleme yöntemi kullanılarak işçilerin gürültüye bağlı işitme kayıpları ile ilgili verileri incelenerek sınıflama yapılmıştır.

Elazığ Organize Sanayi Bölgesi (OSB)nde bulunan F1, F2, F3, F4, F5 ve F6 fabrikalarında çalışan işçilerin gürültüye bağlı işitme kayıplarına göre, bu fabrikaların gürültü düzeylerinin beşli ve ikili kümelenmesi yapılmıştır. Bu fabrikaların gürültü düzeyi 95 dB'in üstünde ve en az on işçi çalışmaktadır. Çalışma 6 fabrikada çalışan 197 işçi ile yapılmıştır.

Gürültü, oluşturduğu işitme kaybının yanı sıra küçümsenmeyecek oranlarda tinnitus ve vertigo gibi otolojik semptomlara; sinir ve dolaşım sistemi bozuklukları, hormonal dengenin bozulması gibi fiziksel etkilerinin yanı sıra rahatsızlık, uyumsuzluk, yorgunluk, huzursuzluk, konsantrasyon bozukluğu, sinirlilik gibi psikolojik etkilere de yol açmaktadır. Sunulan çalışmada gürültüye maruz kalan işçilerin yarıya yakın kısmında kulak çınlaması, baş dönmesi kolay öfkelenme, uyku düzensizliği, konsantrasyon güçlüğü, dikkat dağınıklığı, performans düşüklüğü, kendini yorgun ve bitkin hissetme gibi şikayetlerin olduğu saptandı [22].

Mesleki gürültüye bağlı işitme kaybı; gürültüye maruz kalma süresinin yanı sıra, gürültünün tipi (sürekli ya da vuruş şeklinde) frekansı ve şiddetine de bağlıdır. Gürültüden korunmak için; gürültüyü kaynağından azaltmak, gürültüyü kaynak ile alıcı arasındaki yolda azaltmak ya da gürültüyü gürültüye maruz kalan kişide engellemek şeklinde olabilir. Gürültüyü, gürültüye maruz kalan kişide engellemek için; gürültüye maruz kalan kişiyi tecrid etmek, kişisel koruyucu kullanmak, gürültüye maruziyet süresini azaltmak (rotasyonla çalışmak), iş programını değiştirmek gerekir. Daha başkada önlemler alınabilir. Çalışmada üç tekstil fabrikası, iki mermer fabrikası ve bir un fabrikaları yer almıştır. Bakıldığında tekstil fabrikalarında birbirine benzer makineler kullanıldığı halde, işçilerin ölçümlerine bakıldığında işitme kayıpları çok da yakın çıkmamıştır. Keza mermer fabrikaları da aynı makineleri kullandığı halde aynı zararı vermemiştir.

Sonuç olarak; Elazığ Organize Sanayi Bölgesinde bulunan gürültülü iş yerlerinde çalışan işçilerde gürültüye bağlı işitme kaybı oldukça yüksektir ve ortamdaki gürültü miktarı ve çalışma süresi arttıkça gürültüye bağlı işitme kaybı oranları yükselmektedir.

Gürültünün insan sađlıđına olumsuz etkilerine ve işitme kaybı oluřturmasına karřın, toplumumuzda halen bir risk olarak algılanmamaktadır. Ayrıca, gürültü ile ilgili mevzuatta gürültü limit deđerleri konusunda farklı yasa ve tüzükler bulunmaktadır. Mevzuattaki bu farklılıđın giderilmesi, işyerlerine yönelik olarak etkin denetim ve kontrollerin yapılması yerel yönetimlerin ve işverenlerin bu konudaki sorumluluklarını yerine getirmeleri konusunda eğitilmeleri ve toplumsal duyarlılık oluřturulması gerekmektedir. Ayrıca; İşverenlerin işyeri gürültü düzeyi sınırına uymaları, işçilerin işe başlarken ve periyodik olarak en az yılda bir kez odyometrik tarama yapılmaları ve gürültüyü kaynađından azaltmak, gürültüyü kaynak ile alıcı arasındaki yolda azaltmak ya da gürültüyü gürültüye maruz kalan kişide engellemek amacıyla bir takım önlemlerin alınması gerekmektedir.

6. ÖNERİLER

Kümeleme analizi ile ses düzeyi yüksek fabrikalarda çalışan işçilerde oluşan, gürültüye bağlı işitme kaybı dikkate alınarak, bu fabrikaları gürültü düzeylerine göre sınıflandırılmaları başarılı bir şekilde yapılmıştır. Bu da gösteriyor ki; kümeleme analizinin sınıflandırma işlemleri, farklı çalışma sahalarında da uygulanabileceği düşünülmüştür.

Yapılan çalışmada da görüldüğü gibi, fabrika sahipleri çalışanlarından oluşan gürültüye bağlı işitme kayıplarını bu çalışma sonucunda da daha iyi analiz edip gerekli önlemleri almalılar. Mesleki gürültüye bağlı işitme kaybı; gürültüye maruz kalma süresinin yanı sıra, gürültünün tipi, frekansı ve şiddetine de bağlıdır. Gürültüden korunmak için; gürültüyü kaynağından azaltmak, gürültüyü kaynak ile alıcı arasındaki yolda azaltmak ya da gürültüyü gürültüye maruz kalan kişide engellemek şeklinde olabilir. Gürültüyü, gürültüye maruz kalan kişide engellemek için; gürültüye maruz kalan kişiyi tecrid etmek, kişisel koruyucu kullanmak, gürültüye maruziyet süresini azaltmak, iş programını değiştirmek gerekir.

Gürültüye bağlı işitme kayıpları farklı yöntemlerle de incelenebilir, yaptığımız çalışmayla karşılaştırılıp işçilere en zarar veren fabrikalar belirlenip gerekli önlemler alınabilir.

KAYNAKLAR

- [1] **Hair, J.F., Anderson, R. E., Tatham, R. L., & Black, W. C.**, 1995, Multivariate Data Analysis with Readings. Prentice-Hall.
- [2] **Aldenderfer, M.S., Blashfield R.K.**, 1984, Cluster Analysis, Sage Publications.
- [3] **Özdamar, K.:** 1999, Paket Programlarla İstatistiksel Veri Analizi 2, Kaan Kitabevi, Eskişehir.
- [4] **Sokal, R., Sneath, P.**, 1963 Principles of Numerical Taxonomy, San Francisco
- [5] **Selanik, M.**, 2007 Türk tarımının Avrupa Birliği içindeki yerinin kümeleme analizi ile belirlenmesi. Yayınlanmamış yüksek lisans tezi. Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- [6] **Akın, Y.**, 2008 Veri madenciliğinde kümeleme algoritmaları ve kümeleme analizi, Yayınlanmamış doktora tezi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- [7] **Çokluk, Ö., Şekercioğlu, G., Büyüköztürk, Ş.**, 2010 Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve Lisrel Uygulamaları Ankara.
- [8] **Tatlıdil, H.**, 1996 Uygulamalı çok değişkenli istatistiksel analiz Ankara.
- [9] **Everitt, B.**, 1974 Cluster analysis London.
- [10] **Şentürk, A.**, 1995 Kümeleme Analizi ve bir uygulama. Yayınlanmamış yüksek lisans tezi Bursa.
- [11] **Fırat, S.Ü.**, 1997 Kümeleme analizi İstihdamın sektörel yapısı açısından Avrupa ülkelerinin karşılaştırılması İstanbul.
- [12] **Lorr, M.:** 1983 Cluster Analysis for Social Scientists, Jossey-Bass Publishers, London,.
- [13] **Garson, G.D.**, 2009 Cluster Analysis
- [14] **Johnson, R.A., Wichern, D.W.**, 1988 Applied multivariate statistical analysis prentice Hall New Jers
- [15] **Doğan, İ.**, Kümeleme Analizi ile Seleksiyon. Turk J Vet Anim Sci. 2002;26: 47-53
- [16] **Servi, T.**, Çok Değişkenli Karma Dağılım Modeline Dayalı Kümeleme Analizi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Adana, 2009
- [17] **Yıldız, D.**, Kümeleme Analizi Ders notları, Yıldız Üniversitesi, İstanbul.
- [18] **Gordon, A.D.** (1998). Cluster validation. In: C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H.H. Bock and Y. Baba, Editors, Studies in Classification, Data Analysis, and

Knowledge Organization: Data Science, Classification, and Related Methods, Springer, Tokyo (1998), pp. 22–39.

[19] **Halkıdı, M., Batistakis, Y., And Vazirgiannis, M.** (2002A). Cluster Validity Methods: Part I. SIGMOD Record 31(2): 40-45.

[20] **Theodoridis, S. And Koutroumbas, K.** (2006). Pattern Recognition, thirded. Elsevier Academic Pres.

[21] **Jain, A.K. And Dubes, R.C.** (1988). Algorithms for Clustering Data. Prentice Hall, Englewood Cliffs, NJ.

[22] **Açık Y, Karlıdağ T, Akgün D, Keleş E, Karadağ M, Alpay HC, Deveci SE, Yalçın Ş.,** (2007) Elazığ İli Organize Sanayi Bölgesi’ndeki Gürültülü İşyerlerinde Çalışan İşçilerin İşitme Düzeyleri ve İşitme Düzeylerinin Lokal veya Sistemik Semptomlar ile İlişkisi. TTB Mesleki Sağlık ve Güvenlik Dergisi. 29:39-48.

ÖZGEÇMİŞ

1968 yılında Elazığ'da doğan Nazmiye YALÇIN, ilk, orta ve lise eğitimini Elazığ'da tamamlamış olup, üniversite eğitimini de Fırat Üniversitesi Mühendislik Fakültesi Jeoloji Mühendisliği'nde 1993 yılında tamamlayarak mezun olmuştur. Halen Fırat Üniversitesi Fen Fakültesi İstatistik Bölümü Uygulamalı İstatistik Anabilim Dalında uzman olarak çalışmaktadır.