TÜRKİYE'DEKİ İLLERİN HAYVANCILIK İSTATİSTİKLERİ BAKIMINDAN ÇOK DEĞİŞKENLİ ANALİZ TEKNİKLERİ İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

Süleyman ALDEMİR

Danışman
Doç. Dr. İbrahim KILIÇ

İSTATİSTİK ANABİLİM DALI

Haziran 2019

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TÜRKİYE'DEKİ İLLERİN HAYVANCILIK İSTATİSTİKLERİ BAKIMINDAN ÇOK DEĞİŞKENLİ ANALİZ TEKNİKLERİ İLE İNCELENMESİ

Süleyman ALDEMİR

Danışman
Doç. Dr. İbrahim KILIÇ

İSTATİSTİK ANABİLİM DALI

Haziran 2019

TEZ ONAY SAYFASI

Süleyman ALDEMİR tarafından hazırlanan "Türkiye'deki İllerin Hayvancılık İstatistikleri Bakımından Çok Değişkenli Analiz Teknikleri ile İncelenmesi" adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca 13/06/2019 tarihinde aşağıdaki jüri tarafından **oy birliği** ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **İstatistik Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Doç. Dr. İbrahim KILIÇ

İmza

Başkan : Dr. Öğr. Üyesi İlkay DOĞAN
Gaziantep Üniversitesi, Tıp Fakültesi

Üye : Doç. Dr. İbrahim KILIÇ
Afyon Kocatepe Üniversitesi, Veteriner Fakültesi

Üye : Doç. Dr. Sinan SARAÇLI
Afyon Kocatepe Üniversitesi, Fen-Edebiyat Fakültesi

Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
...../...../..... tarih ve
..... sayılı kararıyla onaylanmıştır.

.....
Prof. Dr. İbrahim EROL
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI
Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

13/06/2019

Süleyman ALDEMİR

ÖZET
Yüksek Lisans Tezi

TÜRKİYE'DEKİ İLLERİN HAYVANCILIK İSTATİSTİKLERİ BAKIMINDAN ÇOK DEĞİŞKENLİ ANALİZ TEKNİKLERİ İLE İNCELENMESİ

Süleyman ALDEMİR
Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı
Danışman: Doç. Dr. İbrahim KILIÇ

Bu araştırmada, kümeleme analizi, faktör analizi ve çok boyutlu ölçekleme analizi gibi çok değişkenli istatistiksel tekniklere teorik anlamda yer verilmiştir. Bununla birlikte, TÜİK'den elde edilen 2018 yılı hayvancılık istatistiklerine ait toplam 29 değişkeni içeren hayvan varlığı ve hayvansal üretim verileri çok değişkenli analiz teknikleri ile incelenmiştir. Çalışmada öncelikle kümeleme analizine yer verilmiş, en uygun küme yapısı hiyerarşik kümeleme yöntemlerinden Ward yöntemi ile elde edilmiş ve iller 2 kümeye ayrılmıştır. Hiyerarşik olmayan kümeleme analizlerinde uygun küme sayısı, bazı parametrelerin yanı sıra küme üyelik kodları kullanılarak diskriminant (ayırma) analizi ile belirlenmiş olup, fuzzy (bulanık) kümeleme yönteminin 5 küme yapısında en etkili yöntem olduğu sonucuna varılmıştır. Ayrıca, çalışmada faktör analizinden elde edilen faktör yükleri kullanılarak kişi başına düşen hayvansal üretim ve hayvan varlığına ait ham veriler ve standartlaştırılmış veriler ile illerin hayvancılık gelişmişlik endeksleri hesaplanmıştır. Her iki sıralamada da ilk sıradaki ilin Ardahan son sıradaki ilin ise İstanbul ili olduğu görülmüştür. Çalışmada son olarak metrik çok boyutlu ölçekleme tekniği kullanılmıştır.

2019, viii + 113 sayfa

Anahtar Kelimeler: Çok değişkenli analiz, Kümeleme analizi, Çok boyutlu ölçekleme, Hayvancılık istatistikleri

ABSTRACT
M.Sc. Thesis

EXAMINING OF PROVINCES IN TURKEY IN TERMS OF LIVESTOCK STATISTICS WITH MULTIVARIATE ANALYSIS TECHNIQUES

Süleyman ALDEMİR

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Statistics

Supervisor: Assoc. Prof. İbrahim KILIÇ

In this study, multivariate statistical techniques such as clustering analysis, factor analysis and multidimensional scaling analysis are given theoretically. In addition, animal existence containing 29 variables of the livestock statistics obtained from TÜİK in 2018 and animal production data are examined with multivariate analysis techniques. Firstly, in the study clustering analysis is included, the most appropriate cluster structure is obtained by the Ward method which is one of the hierarchical clustering methods and, the provinces examined in this study are divided into two groups. The number of the suitable clusters in non-hierarchical clustering analysis is determined by discriminant analysis by using cluster parameters as well as some parameters, moreover; Fuzzy clustering method is concluded to be the most effective method in the cluster structure of 5. Furthermore, the per capita animal production, the raw data of the animal existence, and livestock development indices of the provinces with standardized data are calculated by using factor loadings obtained from factor analysis in the study. It is seen that Ardahan is the first province whereas İstanbul is the last province in both rankings. Finally, metric multidimensional scaling technique was used.

2019, viii + 113 pages

Keywords: Multivariate analysis, Cluster analysis, Multidimensional scaling, Livestock statistics

TEŞEKKÜR

Bu araştırmanın konusu, deneysel çalışmaların yönlendirilmesi, sonuçların değerlendirilmesi ve yazımı aşamasında yapmış olduğu büyük katkılarından dolayı tez danışmanım Sayın Doç. Dr. İbrahim KILIÇ'a, araştırma ve yazım süresince her konuda öneri ve eleştirileriyle yardımlarını gördüğüm hocalarıma ve arkadaşlarıma teşekkür ederim. Bu araştırma boyunca maddi ve manevi desteklerinden dolayı aileme ayrıca teşekkür ederim.

Süleyman ALDEMİR

AFYONKARAHİSAR, 2019

İÇİNDEKİLER DİZİNİ

Sayfa

KISALTMALAR DİZİNİ

Kısaltmalar

r	Korelasyon Katsayısı
R^2	Belirleme (Determinasyon) Katsayısı
λ	Özdeğer (Özvektörlerce açıklanan varyans)
D	Uzaklık Matrisi
S	Benzerlik Matrisi
k	Küme Sayısı
n	Birim Sayısı
sd	Serbestlik Derecesi
MD	Mahalonobis Uzaklığı
KT	Kareler Toplamı
KO	Kareler Ortalaması
EP	Etki Payı
SC	Gölge İstatistiği veya Siluet Katsayısı
HKT	Hata Kareler Toplamı
Cov	Kovaryans Katsayısı
VIF	Varyans Şişkinlik Faktörü
FCM	Fuzzy C-means Algorithm (Bulanık C-Ortalamalar algoritması)
PAM	Partitioning Around Medoids (Medoidler etrafında bölünme)
ÇBÖ	Çok Boyutlu Ölçekleme
PAF	Pricipal-Axis Factoring (Temel Eksen Faktörü)
KMO	Kaiser-Mayer-Olkin
TÜİK	Türkiye İstatistik Kurumu
TekBKY	Tek Bağlantı Kümeleme Yöntemi
OrtBKY	Ortalama Bağlantı Kümeleme Yöntemi
TamBKY	Tam Bağlantı Kümeleme Yöntemi
k-OrtKY	k-Ortalamalar Kümeleme Yöntemi

ŞEKİLLER DİZİNİ

Sayfa

ÇİZELGELER DİZİNİ

Sayfa

1. GİRİŞ

Günümüzde farklı problemlerin gelişmesi, büyük veri topluluklarının oluşması ve karmaşık verilerin çözümü için gerekli olan bilgiye ihtiyaç duyulmaktadır. Karmaşık verileri bilgiye dönüştürmek oldukça zor olsa da modern istatistiksel yöntemler ile istenilen bilgiye ulaşmak kolaylaşmakta bununla beraber anlamlı sonuçlar elde edilmektedir. Çok değişkenli istatistiksel yöntemler istatistik biliminin en önemli konularından biri olmakla beraber diğer bilim dallarında da yüksek öneme sahiptir.

İstatistik bilimi genel olarak örneklerden elde edilen sonuçları kitleye genelleme, geçmiş bilgiler ile geleceğe dönük tahminler elde etme, bir amaç dâhilinde verilerin çözümlenmesine, sınıflandırılmasına ve sonuçların yorumlanmasına dayanmaktadır. Tek değişkenli analiz teknikler ya da yöntemler ile bilimsel araştırmalarda yeterli sonuç alınamadığında ihtiyaç duyulan yöntemler çok değişkenli analiz teknikleridir. Çok değişkenli analiz tekniklerinin, paket programlarının gelişmesi ile uygulanması kolaylaşmış, böylece veri setlerini analiz etmede kullanım yaygınlığı gittikçe artmıştır.

Hayvancılık bir ülkenin gelişmesi için gerekli olan en önemli alanlardan biridir. Hayvan varlığı ve hayvansal üretimin artmasıyla, ülkemizde ihtiyaç duyulan hayvansal üretimin karşılanması sağlanmakta ayrıca ihracat yapılarak ülke ekonomisine önemli katkısı bulunmaktadır. Ancak günümüzde yapılan araştırmalar, ülkemizin hayvan varlığı ve hayvansal üretimin gelişmesi gerektiğini göstermektedir. Basit bir yaklaşım ile hangi ilde hangi hayvan türünün yetiştirilmesi uygunsa söz konusu iller için o hayvan türüne ağırlık verilebilir. Bu sonuçlara göre gerekli üretim planlaması uygulanabilir. Bu çalışma, çok değişkenli istatistik tekniklerinin, hayvansal araştırmalarda yararlanma imkânını göstermek amacıyla yapılmıştır. Çok değişkenli analiz tekniklerinden bazıları (faktör analiz, kümeleme analizi vb.) etraflı olarak açıklanmaya çalışılmış, analizlerin yapılması ve sonuçların yorumlanması, 81 il için 2018 yılı değerleri alınarak 29 değişken üzerinden değerlendirilmiştir. Çalışmada, 81 ilde 2018 yılına ait hayvan varlığı ve hayvansal üretim miktarlarının adet ve kg cinsinden değerleri içeren veri seti kullanılmıştır. Veriler Türkiye İstatistik Kurumu (TÜİK)'ndan derlenmiştir.

2. LİTERATÜR BİLGİLERİ

2.1 Çok Değişkenli Analiz Teknikleri

2.1.1 Çok Değişkenli İstatistiğin Tanımı

Çok değişkenli istatistikler karmaşık veri setlerini analiz etmede kullanım yaygınlığı gittikçe artan tekniklerdir. Bu teknikler, çok sayıda bağımsız ve/veya bağımlı değişkenle ve tüm bu değişkenlerin değişik düzeylerde birbiri ile ilişkili olduğu durumlarda analize imkân sağlar. Karmaşık araştırma sorularına tek değişkenli analizler ile cevap verebilmenin zorluğundan ve çok değişkenli analiz yapabilme yeteneği olan paket programların ulaşılabilirliğinden dolayı çok değişkenli istatistikler oldukça yaygın bir şekilde kullanılmaya başlanmıştır (Tabachnick and Fidell 2015).

Değişkenlik gösteren olaylar, bir tek değişken etkisiyle değil, çok sayıda (genellikle sonsuz sayıda) bağımlı değişken veya bağımsız değişkenlerin ortak etkileriyle oldukça karmaşık bir yapı göstermektedir. Bu nedenle herhangi bir olgunun tanımlanması sadece bir değişkene göre değil, çok sayıdaki değişkene göre yapılmalıdır. Aksi halde diğer değişkenlerin ve bu değişkenler arasındaki olası ortak etkiler dikkate alınmamış olur. Bu nedenlerden dolayı olguları gerçek yapılarına yakın bir şekilde tanımlamak amacıyla, olguyu tanımlayan sonsuz sayıdaki özelliğinden, ölçülebilen azami sayıdaki özellik ölçülerek, çok boyutlu tanımlamak gerekmektedir (Albayrak 2006).

Çok değişkenli analiz teknikleri popülerdir çünkü organizasyonların bilgi yaratmalarını sağlamakta ve böylece karar vermelerini geliştirmektedir. Çok değişkenli analiz tüm istatistiksel teknikleri ifade etmektedir. Bu durum aynı zamanda, araştırılmakta olan bireyler veya nesneler üzerindeki çoklu ölçümleri analiz edecektir. Bu nedenle, ikiden fazla değişkenin eşzamanlı analizi, çok değişkenli olarak kabul edilebilir (Hair *et al.* 2014).

Albayrak (2006) literatürde tek ve çok değişkenli analiz kavramları ile ilgili çeşitli tanımlar yapılmıştır. Tek değişkenli analizde, verilerden bilgi üretilirken tek bir değişkenle; iki değişkenli analizde, eş zamanlı olarak iki değişkenle; çok değişkenli analizde ise, eş zamanlı olarak iki ve ikiden fazla değişkenle ilgilenilmektedir. Çok değişkenli istatistiksel analiz; aynı anda çok sayıda bağımlı değişken veya bağımsız değişkeni ya da bağımlı ve bağımsız değişken ayrımı yapılmadan çok sayıda değişken ile ilgilenildiğinde söz konusu olmaktadır (Shin 1996). Çok değişkenli istatistiksel analiz teknikleri ikiden fazla değişkeni bir olay üzerinde eş zamanlı çözümleyen tüm istatistik teknikleridir (Sheth 1971).

2.1.2 Çok Değişkenli İstatistik Tekniklerinin Kullanım Amacı

Çok değişkenli istatistik tekniklerinin, birden fazla amaç için kullanılabilmektedir. Bu amaçlardan önemli olanları aşağıdaki gibi sıralanabilir (Tatlıdil 1992):

a) Basitleştirme ve Boyut İndirgeme: Birden fazla değişken tarafından belirlenmiş olguları daha aza indirerek reel ya da suni değişkenlerle temsil edilmesini sağlamak, daha basitlik ve anlaşılabilirlik sağlamak.

b) Birimlerin Sınıflandırılması: Gözlemlenmiş birimler ya da değişkenlerin farklı sınıflardan oluşup oluşmadıkları belirlenmeye çalışılır. Diğer bir biçimde ise birimlerin önceden belirlenen sınıflardan hangilerine ait olduklarının belirlenmesidir. Bazı olgularda ise değişkenlerin de gruplandırılması söz konusu olabilmektedir.

c) Bağımlılık Yapısının İncelenmesi: Değişkenlerin kovaryans ve korelasyonlarından yararlanılarak bağımlılığın kaynakları ve sonuçları araştırılır. Örneğin, değişkenlerden bir ya da daha fazlasının bağımlı, ötekilerin bağımsız olduğu regresyon analizinin amacı değişkenler arasındaki bağımlılık yapısını ortaya çıkarmaktır.

d) Hipotez Testleri ve Hipotez Oluşturma: Tek değişkenli tekniklerden bazılarının hipotez testleri genelleştirilerek, çok değişkenli hipotez testlerini oluşturmaktadır. Bu genelleştirilmiş durumlar haricinde hipotez testlerinde başka çok değişkenli istatistik

yöntemleri de kullanılabilmektedir. Bu yöntemlere, açıklanmaya çalışılan olay hakkında yeni model ve hipotezler ortaya çıkarmak amacıyla da başvurulabilmektedir.

e) Sıralama ve Ölçekleme: Ölçekleme birçok değişkenden faydalanılarak incelenmek istenilen birimleri daha az boyuta indirgeyebilmektir. Genel olarak bazı uygulamaların amacı birimleri belirli ölçülere göre sıralamasını gerçekleştirmektir. Böylece grafik gösterimlerinden de yararlanılarak birimlerin, karşılaştırılması, yakınlık ya da uzaklıklarının belirlenmesi kolaylıkla yapılabilmektedir.

Özdamar (2013) kitabında çok değişkenli istatistiksel yöntemlerin değişik amaçlar ile uygulandığını belirtmiş ve bunları kısaca sıralamıştır.

Veri İndirgeme: Gözlemlenen veri setinde p sayıda değişkenin değişik biçimlerini açıklayan ve aralarında ilişki bulunmayan daha az sayıda değişkenle ($k < p$) veri yapısını açıklamayı sağlamak.

Kümeleme ve Sınıflama: Daha önceden sınıflandırılmış kümelere yeni gelen birimlerin atanmasını sağlamak. Özelliklerinden bazıları bilinmeyen yapıların, kümeleri (grup, sınıf) oluşturmalarına ya da kümelere ayrılmalarına yardımcı olmak.

Ölçekleme: p sayıda değişken içeren p boyutlu ölçümlemelerden daha az sayıda değişken kullanarak birimlerin gösterilmesini, tanımlanmasını sağlamak. Birimlerin birbirleri ile $k<p$ boyutlu ölçekte benzerlik ve farklılıklarını incelemek.

Çok Değişkenli Hipotezlerin Test Edilmesi: k kitlesi için oluşturulan çok değişkenli hipotezleri test etmek.

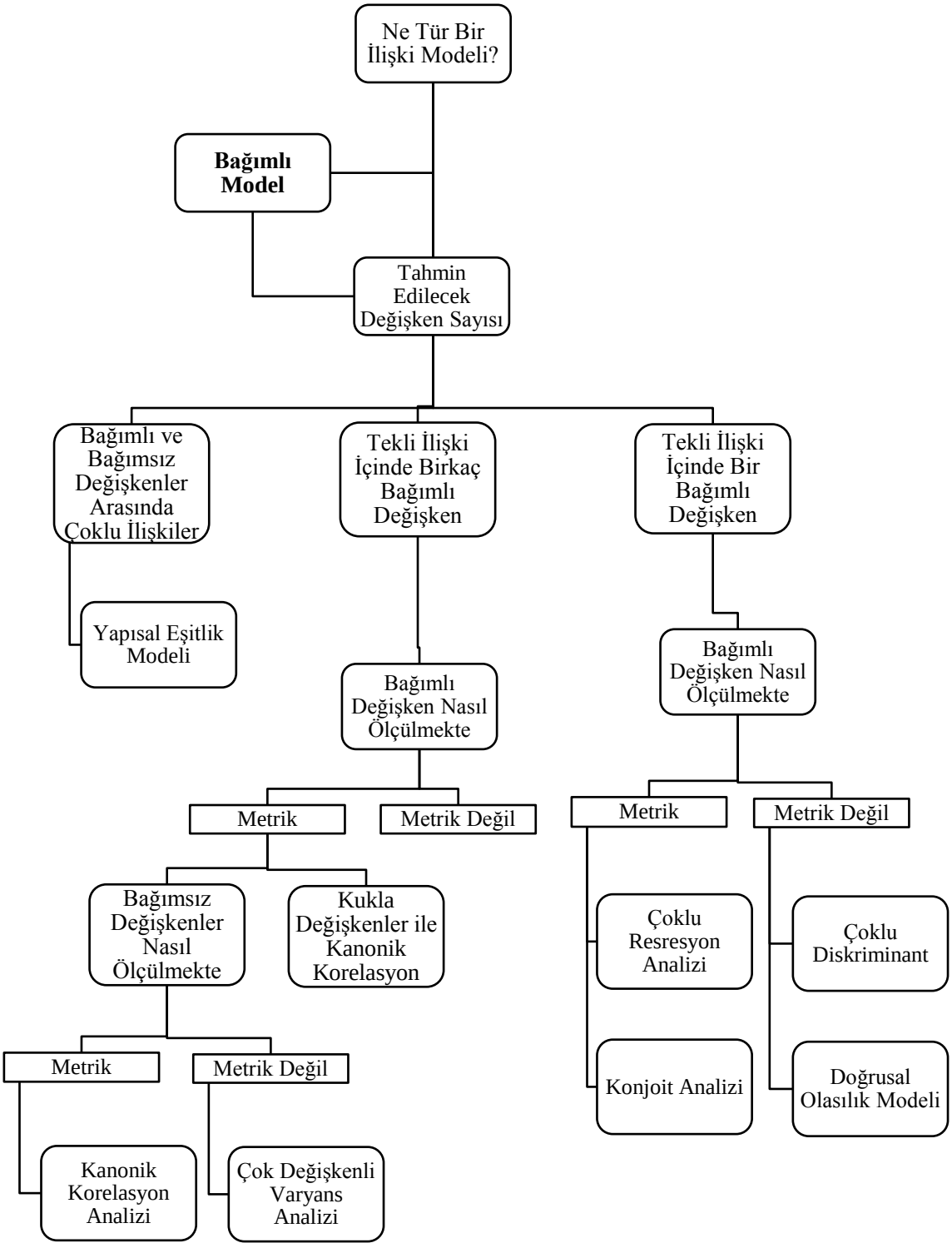
Çok değişkenli istatistiksel analizde en önemli varsayım "n sayıda birimden elde edilen p sayıda değişkenin çok değişkenli normal dağılım gösterdiği" varsayımıdır. Çok değişkenli istatistik tekniklerinde çözümlemeler için matris ve vektör işlemlerinden yararlanılır. Çok değişkenli istatistiksel yöntemlerle çalışmak isteyen uygulamacıların matris ve vektör cebiri ile -çok az da olsa- tanışık olması gerekir (Özdamar 2013).

2.1.3 Çok Değişkenli İstatistik Tekniklerinin Seçimi ve Sınıflandırılması

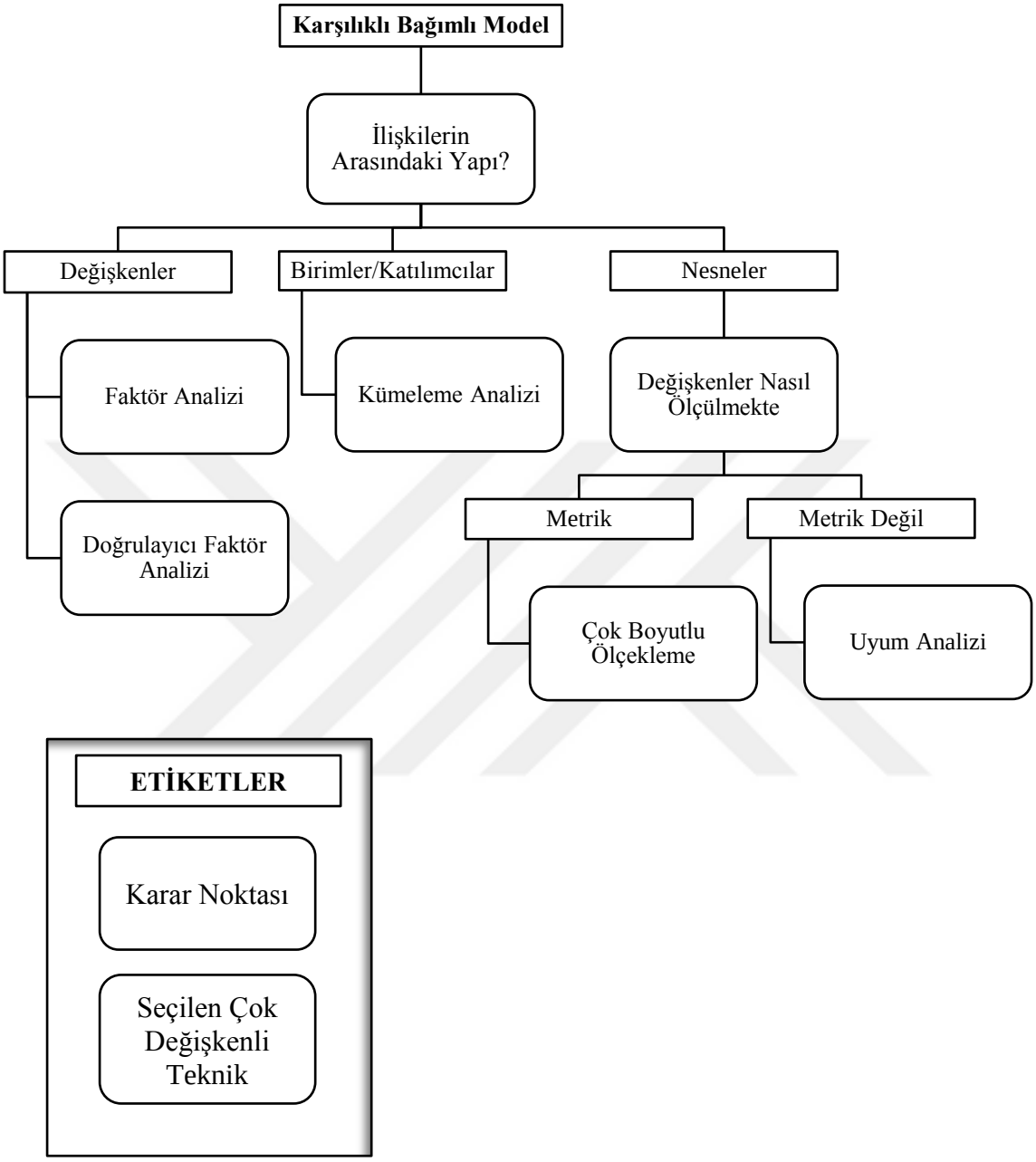
Çok değişkenli tekniklerin tanınmasına yardımcı olmak ve çok değişkenli yöntemlerin kullanımı hakkında Şekil 2.1'de bir sınıflandırma sunulmaktadır. Araştırmacı, araştırma amacı ve verinin doğası hakkında bilgi vermesinden sonra bu sınıflandırma üç yargılamaya dayanmaktadır. Uygun çok değişkenli tekniğin seçimi bu üçünün cevaplarına bağlıdır (Hair *et al.* 2014).

1. Analize tabi olan değişkenler, hipoteze göre bağımsız ve bağımlı sınıflandırmalara bölünebilir mi?
2. Bu değişkenler bağımlı ya da bağımsız sınıflandırılmalara ayrılabiliyorsa, tek bir analizde kaç değişken bağımlı olarak kabul edilebilir?
3. Hem bağımlı hem de bağımsız değişkenler nasıl ölçülür.

Çok değişkenli istatistiksel tekniklerin uygulanması düşünüldüğünde, cevaplanması gereken ilk soru, -İncelenen veri setinde değişkenler bağımsız ve bağımlı sınıflandırmalara ayrılabilir mi?- olmalıdır. Şekil 2.1'de bağımlı çok değişkenli teknikler ve karşılıklı bağımlı çok değişkenli teknikler verilmektedir. Bağımlı çok değişkenli teknikler, bağımlı bir değişken veya değişken kümesinin olduğu ve bu bağımlı değişkenlerin, bağımsız değişkenler olarak bilinen diğer değişkenler tarafından tanımlanması, tahmin edilmesi veya açıklanabilmesidir. Bağımlı çok değişkenli tekniklere örnek olarak çoklu regresyon analizi gösterilebilir. Buna karşılık olarak karşılıklı bağımlı çok değişkenli teknikler tek bir değişken veya grubun olmadığı değişkenlerin bağımlı veya bağımsız olarak tanımlanabildiği tekniklerdir. Karşılıklı bağımlı çok değişkenli istatistik tekniklerinde değişkenler eş zamanlı olarak analiz edilmektedir. Faktör analizi bu tekniğe örnek olarak gösterilebilir (Hair *et al*. 2014).



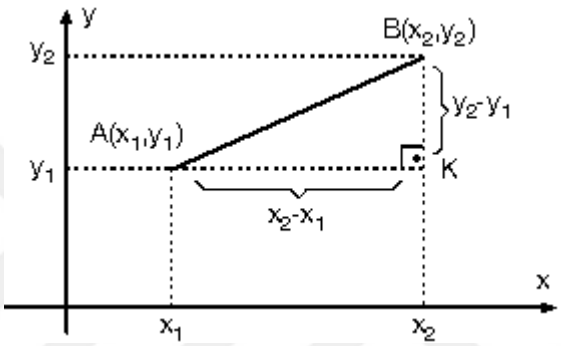
Şekil 2.1 Çok değişkenli bir tekniğin seçimi



Şekil 2.1 (Devam) Çok değişkenli bir tekniğin seçimi

2.1.4 Uzaklık ve Benzerlik Ölçüleri

Birim ya da değişkenler arasındaki uzaklık/benzerlik hesaplanırken geometrik yaklaşımlardan yararlanılır. Geometride koordinat sisteminde yer alan iki nokta arasındaki uzaklık Pisagor bağıntısına göre bulunabilir. Koordinat sistemin de yer alan A ve B noktaları arasındaki doğrusal uzaklık A'nın koordinat değerleri $A(x_1, y_1)$ ve B'nin koordinat değerleri $B(x_1, y_1)$ olmak üzere Pisagor bağıntısına göre; $d(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ şeklinde hesaplanır.



Şekil 2.2 Koordinat sisteminde iki nokta arsındaki uzaklığın gösterimi

Noktalar arasından ki uzaklıkların geometrik olarak gösterimlerinde iki veya daha fazla boyut söz konusu ise söz konusu noktalar arasındaki uzaklığı çok boyutlu olarak hesaplamak gerekmektedir. Değişkenler ya da birimler arasındaki uzaklıkları hesaplamada kullanılan uzaklık ölçülerinin genel anlamda Minkowski uzaklık ölçüsü olarak adlandırılmaktadır (Özdamar 2013).

Uzaklık (benzemezlik) ya da benzerlik ölçülerinin kullanılmasına karar verilirken verinin ölçüm birimi (veri tipi) dikkate alınır; çünkü sürekli-kesikli sayısal veriler, nitelik veriler, sıklık sayıları ve iki kategorili veriler için geliştirilmiş birçok ölçü söz konusudur. Bazı uzaklık ve benzerlik ölçülerinin veri tipine göre sınıflaması Çizelge 2.1'de verilmiştir (Alpar 2011).

Çizelge 2.1 Uzaklık ve benzerlik ölçülerinin sınıflandırılması

Uzaklık (Dissimilarity) Ölçüleri	**Benzerlik (Similarity) Ölçüleri**
Sayısal Veriler	
Öklit Uzaklık Ölçüsü	Pearson İlişki Katsayıları
Karesel Öklit Uzaklık Ölçüsü	Kosinüs Benzerlik Ölçüsü
Chebychev Uzaklık Ölçüsü	
Manhattan Uzaklık Ölçüsü	
Minkowski Uzaklık Ölçüsü	
Karl-Pearson Uzaklık Ölçüsü / Standartlaştırılmış Öklit Uzaklık Ölçüsü	
Korelasyon Uzaklığı Ölçüleri	
İkili Veriler	
Kare Öklit Uzaklık Ölçüsü	Russel ve Rao Benzerlik Ölçüsü
Öklit Uzaklık Ölçüsü	Basit Benzerlik Ölçüsü
Büyüklük Farkları Uzaklık Ölçüsü	Jaccard Benzerlik Ölçüsü
Biçim Farkları Uzaklık Ölçüsü	Parçalı Benzerlik Ölçüsü
Değişim Uzaklık Ölçüsü	Rogers ve Tanimato
Durum Uzaklık Ölçüsü	Sokal ve Sneath Benzerlik Ölçüsü 1
Lance ve Williams Uzaklık Ölçüsü	Sokal ve Sneath Benzerlik Ölçüsü 2
	Sokal ve Sneath Benzerlik Ölçüsü 3
	Sokal ve Sneath Benzerlik Ölçüsü 4
	Sokal ve Sneath Benzerlik Ölçüsü 5
	Kulczynski Benzerlik Ölçüsü 1
	Kulczynski Benzerlik Ölçüsü 2
	Hamann Benzerlik Ölçüsü
	Goodman ve Kruskal Lamda Benzerlik Ölçüsü
	Anderberg D Benzerlik Ölçüsü
	Yule Q Benzerlik Ölçüsü
	Yule Y Benzerlik Ölçüsü
	Ochiai Benzerlik Ölçüsü
	Fi 4 Nokta Benzerlik Ölçüsü
	Yayılım Benzerlik Ölçüsü
Sıklık Sayılar	
Ki-Kare Uzaklık Ölçüsü	
Phi-Kare Uzaklık Ölçüsü	

2.1.4.1 Öklid ve Karesel Öklid Uzaklık Ölçüsü

$n * p$ boyutlu ve i. ve j. gözlemler, nesneler ya da birimler arasındaki uzaklıkları doğrudan ölçü biçiminde (Öklid uzaklığı) ya da karesel uzaklıklar (Karesel Öklid uzaklığı) biçiminde hesaplayan bir ölçüdür. Öklid uzaklığı $d(i,j) = \sqrt{\sum_{k=1}^{p}(x_{ik} - x_{jk})^2}$ şeklinde hesaplanır. Burada $i = 1, 2, \dots\dots\dots, n$; $j = 1, 2, \dots\dots\dots, n$ ve $k = 1, 2, \dots\dots\dots, p$ dir. N birim sayısı ve p değişken sayısıdır.

Karesel Öklid uzaklığı ise; Öklid uzaklığının hesaplanmasında olduğu gibi hesaplanır (Özdamar 2013). Karesel Öklid uzaklığı formülü eşitlik (2.1)'de verilmiştir.

$$d(i,j)^2 = \sum_{k=1}^{p}(x_{ik} - x_{jk})^2 \qquad (2.1)$$

2.1.4.2 Chebychev Uzaklığı

Bu uzaklık iki birim arasındaki mutlak maksimum uzaklığa eşittir. Chebychev uzaklık formülü eşitlik (2.2)'de verilmiştir.

$$d(x,y) = \max|x_i - y_i| \qquad (2.2)$$

$d(x,y)$: x ve y noktaları arasındaki uzaklık, hata parametresidir.

x, y: Aralarındaki uzaklık hesaplanan nesneleri uzayda temsil eden noktalardır (Demiralay ve Çamurcu 2005).

2.1.4.3 Manhattan City-Block Uzaklık Ölçüsü

Farkların mutlak değerlerinin toplamı olarak tanımlanır. Kesikli sayısal veriler için daha çok önerilir. Çizelge 2.1'de iki nokta arasındaki en kısa uzaklık Öklit uzaklığı ile verilirken, Manhattan City-Block uzaklığı noktalı çizgi ile gösterilen üçgen kenarlarının toplamı olarak düşünülmelidir (Alpar 2011).

Manhattan City-Block uzaklık formülü eşitlik (2.3)'te verilmiştir.

$$d_{ij} = \sum_{k=1}^{p} \left| x_{ik} - x_{jk} \right| \tag{2.3}$$

2.1.4.4 Minkowski Uzaklık Ölçüsü

Minkowski uzaklığı, Manhattan ve Öklid mesafelerinin bir genellemesidir. Minkowski uzaklık formülü eşitlik (2.4)'te verilmiştir.

$$d_{ij} = \left[\sum_{i=1}^{p} \left| x_{ik} - x_{jk} \right|^{\lambda} \right]^{1/\lambda} \tag{2.4}$$

Burada $\lambda = 1$ olduğunda Manhattan uzaklık değerini, $\lambda = 2$ olduğunda ise Öklid mesafesini elde ederiz. Buradan anlaşıldığı üzere artan λ değerleri farklı birimleri, benzer birimlere kıyasla daha çok artırdığı söylenebilir (Hewson 2009).

2.1.4.5 Pearson ve Karesel Pearson Uzaklığı

Pearson uzaklığı, standardize Öklid uzaklığı olarak da isimlendirilmektedir. Öklid uzaklığının değişkenin varyansına oranlanmış uzaklığıdır. Pearson uzaklık formülü eşitlik (2.5)'te verilmiştir.

$$d_p(i,j) = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk} \right)^2 / s_k^2} \tag{2.5}$$

$i, j = 1, 2, \ldots\ldots, n;\ \ k = 1, 2, \ldots\ldots, p.$ Bu eşitliklerde s_k^2 k. değişkenin varyansını belirtmektedir. Karesel Pearson uzaklığı ise Pearson uzaklığının karesi olarak hesaplanır (Özdamar 2013).

2.1.4.6 Mahalanobis Uzaklığı

Mahalanobis uzaklığı, i ve k gözlemleri için, iki değişken arasındaki kovaryans ve korelasyon katsayılarını dikkate alarak aşağıdaki eşitlik ile hesaplanır. Mahalanobis uzaklık formülü eşitlik (2.6)'da verilmiştir.

$$MD_{ik}^2 = \frac{1}{1-r^2}\left[\frac{(x_{i1}-x_{k1})^2}{s_1^2} + \frac{(x_{i2}-x_{k2})^2}{s_2^2} - \frac{2r(x_{i1}x_{k1})(x_{i2}-x_{k2})}{s_1 s_2}\right] \qquad (2.6)$$

s_1^2 ve s_2^2 sırasıyla birinci ve ikinci değişkenlerin varyanslarıdır ve r iki değişken arasındaki korelasyon katsayısıdır. P değişkenli bir uzayda, Mahalanobis uzaklığı iki gözlem değeri için şu şekilde verilir;

$$MD_{ik}^2 = (x_i - x_k)' S^{-1} (x_i - x_k) \qquad (2.7)$$

Denklemde x, $p * 1$ boyutunda koordinasyonların vektörü ve S $p * p$ lik kovaryans matrisidir. Bağımsız değişkenler için S, üçgen matrisi, bağımsız standartlaştırılmış değişkenler için S, birim matrisi olacaktır (Sharma 1996).

2.1.4.7 Korelasyon Katsayısı ve Korelasyon Uzaklık Ölçüsü

Korelasyon katsayısı, iki değişken arasındaki ilişkinin düzeyini (derecesini, şiddetini, gücünü) ve yönünü belirlemek amacıyla hesaplanır. Korelasyon katsayısı "r" harfi ile ifade edilir ve -1 ile +1 arasında bir değer almaktadır. Burada, rakamların mutlak büyüklüğü, değişkenler arasındaki söz konusu ilişkinin düzeyini ve rakamların işareti ise yönünü belirler (Ural ve Kılıç 2011).

X ve Y gibi iki değişkenin aralarındaki lineer ilişkinin ölçüsü için Pearson Korelasyon katsayısı denilen ρ parametresi kullanılmaktadır. i ve j değişkenleri arasındaki örneklem için Pearson Korelasyon katsayısı eşitlik (2.8)'de verilmiştir (Akdeniz 2007).

$$r_{ij} = \frac{\sum_{k=1}^{n} x_i x_j - \sum_{k=1}^{n} x_i \sum_{k=1}^{n} x_j}{\sqrt{\left[\sum_{k=1}^{n} x_i{}^2 - \frac{\left(\sum_{k=1}^{n} x_i\right)^2}{n}\right]\left[\sum_{k=1}^{n} x_j{}^2 - \frac{\left(\sum_{k=1}^{n} x_j\right)^2}{n}\right]}} \quad (i, j = 1, 2, \ldots\ldots, p) \tag{2.8}$$

Bu kapsamda korelasyon uzaklık ölçüleri:

$$d_{ij} = \left(1 - r_{ij}\right)/2, \tag{2.9}$$

$$d_{ij} = 1 - \left|r_{ij}\right|, \tag{2.10}$$

$$d_{ij} = 1 - r_{ij}{}^2, \tag{2.11}$$

$$d_{ij} = 1 - r_{ij}. \tag{2.12}$$

Bu ölçüler arasından literatürde en çok eşitlik (2.12) ile karşılaşılır. Korelasyon katsayısı -1 ile +1 arasında değişirken bu eşitlikle elde edilen uzaklık ölçüsü 0 ile 2 arasında değerler alır (Alpar 2011).

2.1.4.8. Cosine (Açısal) Benzerlik Ölçüsü

+1 ile -1 aralığında değer alır ve iki veri noktasının özellik vektörleri arasındaki açının kosinüsü, açısal benzerlik ölçüsü olarak tanımlanmaktadır. Bu benzerlik ölçüsü aşağıda verilen eşitlik (2.13) ile hesaplanmaktadır (Vatansever 2008).

$$s_{ij} = \frac{X_i^T X_j}{\sqrt{X_i^T X_i}\sqrt{X_j^T X_j}} \tag{2.13}$$

2.1.4.9 Jaccard Benzerlik Ölçüsü

Mikrobiyolojik ve Taksonomik bulgularda sonuçları ikili değerlere göre saptanan birimlerin belirli bir özelliğe sahip olanlarının pozitif ve negatif özellikler gösterenlere oranını belirten Jaccard katsayısı bir benzerlik ölçüsü olarak ele alınmıştır (Özdamar 2013).

$$J = Sim_{kl} = (nS^{+})/(nS^{+} + nd) \qquad (2.14)$$

biçiminde hesaplanır. Burada nS^{+}, k ve l birimlerinde pozitif karakterlerin sayısını; nd ise birimlerden birinde pozitif diğerinde negatif olan karakterlerin sayısını göstermektedir.

2.1.5 Çok Değişkenli İstatistik Tekniklerinin Varsayımları

Bu bölümde çok değişkenli istatistik tekniklerinin varsayımları ve bu varsayımların olası etkileri ele alınacaktır. Ayrıca verilerin söz konusu varsayımlara olan uygunluğunun araştırılmasında tavsiye edilen yöntemlere değinilecektir. Bu varsayımlar (Albayrak 2006):

- **Çoklu normal dağılım:** Veriler çok değişkenli normal dağılıma uyar.
- **Sabit Varyans:** Tüm gruplar için kovaryans matrisleri eşittir.
- **Çoklu doğrusal bağıntı:** Bağımsız değişkenler arasında anlamlı doğrusal ilişki yoktur.
- **Doğrusallık:** Bağımsız ile bağımlı değişkenler arasındaki ilişki doğrusaldır.
- **Otokorelasyon olmaması:** Değişkenlerin birimlerinin söz konusu değerleri birbirinden bağımsızdır.

2.1.5.1 Çok Değişkenli Normal Dağılım ve Aşırı Gözlemler

Bazı değişkenli süreçlerde ve birçok istatistiksel testte, çok değişkenli normallik temelde kabul edilen bir varsayımdır. Çok değişkenli normallik varsayımı, her değişken ve değişkenlerin doğrusal kombinasyonlarının normal dağılıma sahip oluşudur. Bu varsayım yerine geldiğinde, artıklarda normal dağılır ve birbirinden bağımsızdır. Çok değişkenli normalliğin test edilmesi, uygulamada çok pratik olmayan bir yöntemdir çünkü normallik için değişkenlerin sonsuz sayıda doğrusal kombinasyonunu test etmek gerekecektir. Bu yönde hâlihazırda elimizde olan testler ise aşırı duyarlıdır (Tabachnick and Fidell 2015).

Doğal yaşamda gerçek veriler normal dağılım göstermeyebilir, fakat teorik olarak gerçek popülasyonun normal dağılım gösterdiği varsayımı, yararlı bir yaklaşımdır. Birçok sorunu standardize yaklaşım ile çözümleme kolaylığı sağlar. Çok değişkenli normal dağılımın birçok avantajı vardır. Bunlar kısaca (Özdamar 2013):

- Çok değişkenli normal dağılım, matematiksel olarak incelenmesi, çözümlemeler yapılması kolay olan bir dağılımdır. Ayrıca doğada sürekli değişken içeren birçok çok değişkenli olay normal dağılım göstermektedir.
- Çok değişkenli normal dağılım dışındaki diğer dağılımlarla ilgili matematiksel yaklaşımlar hem yaygın kullanıma sahip değildir hem de birçok uygulamacı normal dağılım dışındaki dağılımlarla ilgili ayrıntılı bilgiye sahip değildir.
- Normal dağılım bazı durumlarda popülasyondaki dağılımların asimptotik yapısına uygundur. Normal dağılım çoğu durumda gerçek popülasyon modeli olarak görev yapar.
- Çok değişkenli istatistiklerin örneklem dağılımları yaklaşık olarak çok değişkenli normal dağılım gösterirler. Örneğin alındığı dağılım formuna bakılmaksızın çok değişkenli istatistiklerin örneklem dağılımları merkezi limit teoremi etkisinden dolayı normal dağılım gösterir.

Çok değişkenli istatistik yöntemlerinin çoğu, X veri matrisinin çok değişkenli normal dağıldığı varsayımı temeline dayandırılmıştır. Çok değişkenli normal dağılım, tek değişkenli normal dağılımın, değişken sayısı $(p) \geq 2$ için genellenmiş durumudur.

Bilindiği gibi, tek değişkenli normal dağılım:
x: İncelenen değişken
μ: İncelenen değişkenin ortalaması
σ: İncelenen değişkenin standart sapması
olmak üzere

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2}\right)[(x-\mu/\sigma)]^2} \quad -\infty < x < +\infty \qquad (2.15)$$

ile verilen bir olasılık yoğunluk fonksiyonuna sahiptir ve bu fonksiyonda: ortalama (μ) ile ± 1 standart sapma $(\pm 1\sigma)$ aralığında gözlemlerin % 68.26'sı, ± 2 standart sapma $(\pm 2\sigma)$ aralığında gözlemlerin %95.44' ü ve ± 3 standart sapma $(\pm 3\sigma)$ aralığında gözlemlerin yaklaşık % 99'u bulunur ve aşağıdaki gibi yazılır:

$$p(\mu - 1\sigma \leq x \leq \mu + 1\sigma) = 0.6826$$
$$p(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0.9544$$
$$p(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0.9974$$

Eğer bir x dağılımı μ ve σ^2 ile normal dağılım gösteriyor ise bu dağılımın normal dağıldığına ilişkin gösterim $x \sim N(\mu, \sigma^2)$ şeklinde olur. Örneğin bir x dağılımı $\mu = 25$ ve $\sigma^2 = 4$ ile normal dağılım gösteriyor ise $x \sim N(25, 4)$ yazılır.

Eşitlik (2.15)'deki $[(x - \mu / \sigma)]^2$ terimi, x değişkenine ilişkin gözlemlerin ortalamaya olan uzaklığının standart sapma ile standartlaştırılmış halinin yani uzaklığının karesidir,

$$\frac{(x-\mu)^2}{\sigma^2} = (x-\mu)\left(\frac{1}{\sigma^2}\right)(x-\mu) = (x-\mu)(\sigma^2)^{-1}(x-\mu) \qquad (2.16)$$

şeklinde ifade edilir.

Bu yaklaşım p ∗ 1 boyutlu x gözlem vektörü için genelleştirilirse

Σ: p ∗ p boyutlu ve p ranklı kovaryans matrisi

x: p ∗ 1 boyutlu gözlem vektörü

μ: p ∗ 1 boyutlu ortalama vektörü olmak üzere

$$(x - \mu)'\Sigma^{-1}(x - \mu) \tag{2.17}$$

şeklinde yazılabilir ve bu ifade kısaca, "x gözlem vektöründen μ ortalama vektörüne olan genelleştirilmiş kare uzaklık" şeklinde ifade edilir. Bu uzaklığa Mahalanobis uzaklığı da denir. Çok değişkenli normal yoğunluk fonksiyonunu elde ederken, tek değişkenli uzaklık (2.16), çok değişkenli genelleştirilmiş uzaklık eşitlik (2.17) ile yer değiştirirken, $\sqrt{2\pi\sigma^2} = (2\pi)^{1/2}(\sigma^2)^{1/2}$, $(2\pi)^{p/2}|\Sigma|^{1/2}$ ile yer değiştirir ve bir x gözlem vektörü için p boyutlu normal dağılım fonksiyonu

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2} \tag{2.18}$$

ile verilir. Burada $-\infty < xi < +\infty$ ve $i = 1, \dots, p$ olup, p boyutlu normal yoğunluk $N_p(\mu, \Sigma)$ ile gösterilir. Örneklemler için μ yerine $\bar{x}$, Σ yerine S yazılır. Özet olarak çok değişkenli normal dağılım, ortalama vektörü ve varyans-kovaryans matrisi ile tanımlanır (Alpar 2011).

Çok Değişkenli Normallik Testleri

Çok değişkenli normal dağılımlar da, tek değişkenli normal dağılım varsayımında olduğu gibi betimsel, grafiksel ve istatistiksel yöntemler bulunmaktadır (Gnanadesikan 1997). Çok değişkenli basıklık ve çarpıklık katsayısı, çok değişkenli normallik varsayımının araştırılabilmesi için betimsel yöntemler olarak kullanılabilmektedir. Grafiksel yöntem olarak çoklu saçılma grafiğinin yanı sıra marjinal grafikler, Kowalski'nin çizgi grafikleri ve çok değişkenli Q-Q grafikleri de kullanılmaktadır (Thode 2002). Tek değişkenli normal dağılım için bu değişkenlerin saçılma grafiklerinin elips veya elipse benzer şekilde olduğunda çok değişkenli normallik

varsayımının sağlandığı ifade edilmektedir. Üç değişkenli normal dağılım için ise uygun programlar yardımıyla üç boyutlu saçılma grafikleri çizilebilmesine rağmen dört ve üzeri değişken için çok boyutlu uzayda grafik çizmenin son derece zor olduğu belirtilmektedir (Demir vd. 2016).

Aşırı Gözlemler (Sapan Değerler)

Çok değişkenli sapan birimler, analizde kullanılacak bağımsız değişkenler arasındaki kareli Mahalanobis uzaklıklar (MD^2) hesaplanarak saptanabilmektedir. Şöyle ki; analizdeki her birim için hesaplanan kareli Mahalanobis uzaklıkları analizdeki değişken sayısına (p, serbestlik derecesine) bölünerek elde edilen değer (MD^2/df) t dağılımına uymaktadır. Herhangi bir birimin sapan değer olarak değerlendirilmesi için ilgili birimin %1 anlamlılık düzeyinde anlamlı olması gerekmektedir. Yani, (MD^2/df) değerinin (t) 5.014'den büyük olması gerekmektedir. Genel olarak, sapan değerlerin analizden çıkarılması daha çok normal dağılıma uygun değişken ve böylece daha az dönüştürülmesi gereken değişken söz konusu olmaktadır. Ancak bu durumun her sapan değerin analizden çıkartılması gerektiği şeklinde yorumlanmamalıdır. Bu anlamda sapan değerleri iyi ve kötü huylu olmak üzere iki gruba ayırmak mümkündür. İyi huylu sapan değerler ana kütlenin özelliklerini yansıtan, kötü huylu sapan değerler ise ana kütlenin özelliğini yansıtmayan birimlerdir. Bu anlamda kötü huylu sapan birimlerin analizden mutlaka atılması gerekirken, iyi huylu sapan birimlerin analizden çıkartılmadan önce herhangi bir uygun dönüşümle düzeltilip düzeltilemeyeceği araştırılmalıdır (Albayrak 2003).

2.1.5.2 Çoklu Doğrusal Bağlantı

Çoklu doğrusal bağlantı değişkenlerin korelasyon matrisinde birbirleriyle aşırı derecede ilişki olduğu zaman ortaya çıkan problemlerdir. Çoklu doğrusallık olduğu zaman değişkenler oldukça yüksek derecede birbirleriyle ilişkilidir (ör., 0.90 ve yukarısı). Tekillikte ise değişkenlerden bazıları tekrardır çünkü değişkenlerden biri, iki veya daha fazla değişkenin kombinasyonudur (Tabachnick and Fidell 2015).

İki bağımsız değişken arasındaki doğrusal ilişki doğrusal bağlantı, ikiden çok bağımsız değişkenler arasındaki doğrusal ilişki ise çoklu doğrusal bağlantı olarak tanımlanmaktadır. İki değişken arasında ki ilişki +1 ise aynı, -1 ise zıt yönlü tam bir bağımlılık; sıfıra eşitse tam bir bağımsızlık söz konusudur. Çoklu bağlantı halinde bir bağımsız değişken diğer bir bağımsız değişken veya değişkenler tarafından tam olarak tahmin edilmesi durumunda gerçekleşen ekstrem duruma tekillik denilmektedir (Albayrak 2006).

Çoklu Doğrusal Bağlantının Nedenleri

Çoklu doğrusal bağlantı genellikle aşağıdaki gibi sebeplerden dolayı ortaya çıkmaktadır (Alpar 2011).

1. Örneklemin evreni temsil etmemesi nedeni ile ortaya çıkan çoklu bağlantı problemi.
2. Çoklu bağlantının ortaya çıkabileceği bir diğer durum, bağımsız değişkenler arasında gerçekten ilişki olması durumudur.
3. Gözlem sayısı değişken sayısından küçük olduğu durumlarda ($n<p+1$) tam çoklu bağlantı ortaya çıkmaktadır.

Çoklu Doğrusal Bağlantının Ortaya Çıkarılması

Çoklu bağıntının var olduğunu ortaya çıkartan birçok gösterge vardır. Genel anlamda iki değişken arasındaki korelasyon katsayısının 1'e yakın olması çoklu bağlantı olabileceğini düşündürür (Alpar 2011). Regresyon analizinin varsayımlarından birisi modeli oluşturan, bağımsız değişkenler arasında bir ilişki olmaması varsayımıdır. Bu varsayımın sağlanamaması durumunda, yani bağımsız değişkenler arasında doğrusal bir ilişki söz konusu ise, çoklu doğrusal bağıntı problemi oluşmaktadır. Bağımsız değişkenler arasında doğrusal bir ilişki söz konusu olduğunda regresyon katsayıları etkilediğinden, gerçekte olması gerekenden oldukça farklı kestirimler ortaya çıkabilmektedir. Ayrıca çoklu bağıntı, regresyon katsayılarının standart hata kestirimleri ve buna bağlı olarak da hesaplanan t istatistiğinin olması gerekenden farklı çıkmasına neden olacaktır (Aktaş ve Yılmaz 2003).

Genel olarak bağımsız değişkenlerin kendi aralarında var olan ikişerli kuvvetli ilişki çoklu doğrusal bağlantısının varlığının bir göstergesidir. Korelasyon katsayısının 0.80 den büyük bir değere sahip olması çok değişkenli bir modelde çoklu doğrusal bağlantı önemli bir problem arz etmektedir. Fakat bu kriterin yetersiz kaldığı söylenebilir. Zira bazen bağımsız değişkenler arasında karşılıklı olarak ikili zayıf bir ilişki olması durumunda da çok değişkenli doğrusal bağlantı problemi ortaya çıkabilmektedir. Bu sebepten bu ikinci kriterin sadece iki bağımsız değişkenli modellerde kullanılması sakıncalı olmaktadır. Bu son durumda kısmi korelasyon katsayıları kriteri kullanılmaktadır. Bu kriter şöyledir: Y ile X_2, X_3, X_4 arasında ki regresyonu ele alalım. Burada Y ile bağımsız değişkenler arasında çoklu belirlilik katsayısı R^2 çok yüksek iken $r_{12.34}, r_{13.24}$, ve $r_{14.23}$ kısmi korelasyon katsayılarının değerleri düşükse, X_2, X_3 ve X_4 değişkenlerinin aralarında kuvvetli ilişki olduğu sonucuna varılabilir; bu durumda en azından X'ler den birinin gereksiz olduğu modelden çıkarılması gerektiğine karar verilir. Fakat R^2 yüksek değerli ve kısmi korelasyon katsayıları da yüksek değerli ise, çoklu doğrusal bağlantının varlığına rahatça karar verilemez. Çoklu doğrusal bağlantının araştırılmasında üç boyut vardır (Atalay 1984). Bunlar çoklu doğrusal bağlantının varlığının belirlenmesi, şiddetinin saptanması ve bir veri kümesinde ki yerleşim formunun tespitidir (Baldemir 1997).

Açıklayıcı değişkenler arasında çoklu doğrusal bağlantının olup olmadığını test eden bir diğer istatistik ise Varyans Şişkinlik Faktörüdür (VIF, Variance Inflation Faktor). VIF yüksek değer alıyorsa regresyon katsayılarının varyansları büyür ve açıklayıcı değişkenlerin Y üzerine etkileri yanlış değerlendirilir. VIF değeri (Özdamar 2013):

- **VIF=1** ise çoklu doğrusal bağımlılık yoktur.
- **1 < VIF < 5** ise orta düzeyde çoklu doğrusal bağımlılık vardır. Modelde düzeltme yapmak gerekmez Ancak araştırmacı verilerine göre düzeltme planlayabilir.
- **5 < VIF < 10** ise yüksek düzeyde çoklu doğrusal bağımlılık vardır. Bağımsız değişkenler de düzeltme yapılmalıdır. Önemsiz ilişki düzeyine sahip değişkenler (P > 0.25) sırası ile modelden çıkarılır ve analiz yenilenir. Çoklu bağımlılığa neden olan bağımsız değişken bulunarak düzeltilir ya da modelden çıkarılır.

- **VIF > 10** ise çok yüksek düzeyde önemli çoklu doğrusal bağımlılık vardır. Model geçersizdir. Uygun yöntemler ile bağımsız değişkenler düzeltilir yeniden model oluşturulur. Gerekli düzeltme sağlanamıyorsa alternatif analizlere başvurulur.

2.1.6 Faktör Analizi

Faktör analizi; birbiriyle ilişkili çok sayıdaki değişkeni az sayıda, daha anlamlı, kolay anlaşılabilir ve birbirinden bağımsız faktörler haline getiren ve yaygın olarak kullanılan çok değişkenli istatistik tekniklerinden biridir (Cengiz ve Kılınç 2007).

Faktör analizinin işleyiş yapılarını ve faktör kavramını kısaca özetlemek amacıyla yapılan tanımlama ve açıklamadan bazıları aşağıda verilmiştir (Alpar 2011);

1. Faktör analizinde p sayıdaki değişkenin açıkladığı bir oluşumu ya da yapıyı genel anlamda, kendi içinde ilişkisi bulunan; fakat aralarında birbirleriyle ilişkisi bulunmayan daha az sayıdaki ($k<p$) yeni faktörler ya da değişkenler ile açıklamaya çalışan bir yöntemler bütünüdür. Faktör analizi sonucunda elde edilen yeni faktörler orijinal değişkenlerin doğrusal bileşenleri olup birbirine diktir (faktörler arasındaki ilişki katsayıları sıfırdır); ancak her faktörü oluşturan temel değişkenler arasındaki ilişkiler oldukça yüksektir.

2. Faktör analizi yorumlanması güç, birbiri ile ilişkili çok sayıda değişkenden, en az bilgi kaybı ile bağımsız, kavramsal açıdan anlamlı az sayıda yeni değişkenler (faktörler, boyutlar) bulmayı, ortaya çıkarmayı amaçlayan çok değişkenli yöntemler bütünüdür.

3. Faktör analizi anlaşılması güç çok değişkenli bir yapının, daha az temel değişkenle açıklanıp açılanamayacağının merak edildiği durumlarda kullanılan tekniklerdir.

4. Faktör analizi, aralarında ilişki bulunan veri setlerini birbirinden bağımsız daha az sayıda veri yapılarına dönüştürmek ya da bir başka ifade ile oluşumların nedenini açıklayabilen ya da açıkladığı düşünülen değişkenleri (faktörleri/boyutları/bileşenleri) belirlemek ve gerekli görüldüğü takdirde isimlendirmek amacıyla kullanılan teknikler

bütünüdür. Bu tanımlama ile beraber, faktör analizi birçok değişkenin birkaç başlık altında gruplanıp gruplanmadığına dair bilgi veren bir yöntemler bütünü olarak da tanımlanabilir.

5. Faktör analizi bağımlı, açıklanması kolay olmayan ve birçok değişkenin oluşturduğu topluluktan, bu oluşumu temsil edebilen birbirinden tamamen ya da göreceli olarak bağımsız daha az ve kavramsal olarak anlamlı faktörlerin türetilmesini amaçlayan yöntemler bütünüdür.

Faktör analizi; temel bileşenler analizine benzeyen bir yöntemdir. Her iki yöntemde de veri indirgeme söz konusudur. Ancak faktör analizi değişkenleri gruplayarak ortak faktörler tanımlama özelliğine sahiptir. Faktör analizinin temel iki amacı bulunmaktadır (Özdamar 2013).

- Değişkenlerin toplam sayılarını azaltabilmek (veri indirgeme).
- Değişkenlerin aralarında oluşan ilişkilerden faydalanarak yeni faktör yapıları elde etmektir.

Değişkenler arasındaki karşılıklı ilişkileri inceleyen fazlaca çok değişkenli istatistiksel analiz türü vardır. Bunlardan biriside faktör analizidir. Bu analizi uygulamada çeşitli yöntemler ve teknikler geliştirilmiştir. Bu nedenle faktör analizi belirli yöntemden çok bir teknikler seti olarak tanımlanabilir. Faktör analizinin en öncelikli amaçlarından biriside değişkenler arasında var olan söz konusu karşılıklı bağımlılığın kökenini araştırmaktır. Faktör analizinde veri matrisi, ayırma analizi ve diğer çok değişkenli analizlerden farklı olarak kriter ve tahmin değişkenleri alt setlerine bölüştürülmez, bunun yerine tüm değişkenler arasındaki (veya veri matrisi içerisindeki) ilişkiler analizde incelenir. Bu ilişkilere dayanarak verilerin daha anlamlı ve özet bir biçimde sunulması sağlanır (Kurtuluş 1992).

2.1.6.1 Faktör Analizinin Varsayımları

Faktör analizindeki varsayımlar istatistik olmaktan çok kavramsal varsayımlardır. İstatistiksel açıdan normal dağılımdan ve doğrusallıktan sapmalar sadece hesaplanan korelasyon değerlerini azaltmaktadır. Eğer yeni oluşturulacak faktörlerin anlamlı olup olmadığı araştırılacak ise sadece normallik varsayımı gereklidir. Fakat bu test çok sık kullanılmamaktadır. Gerçekte faktör analizinde değişkenlerin aralarındaki ilişkiler ortaya çıkarılarak oluşturulmasından dolayı, belirli düzeyde çoklu bağlantının olması arzu edilmektedir (Hair *et al.* 1998). Böylece faktör analizi kavramsal olarak seçilen anlamlı değişkenler için uygun olmaktadır. Bununla beraber faktör analizinde temel varsayım değişkenlerde gizli yapıların var olduğunu kabul etmektir. Değişkenler setinin faktör analizine kavramsal olarak uygunluğu ve geçerliliği tamamen araştırmacının sorumluluğuna aittir. Çünkü faktör analizinde değişkenler arkasında yatan gizli yapılar belirlenirken sadece değişkenler arasında var olan korelasyonlardan yararlanılarak belirlenmektedir (Albayrak 2006).

2.1.6.2 Faktör Analizinin Aşamaları

Faktör analizi genelde beş aşamada incelenir (Alpar 2011);

İlk aşama; verinin faktörlenebilir bir yapıda olup olmadığının ve gerekli varsayımların ve kısıtlayıcıların sağlanıp sağlanmamasının incelenmesidir.

İkinci aşama; faktörleşmeyi gösterecek olan faktör yükleri matrisinin faktör çıkarma yöntemlerinden biri ile elde edilmesi aşamasıdır. Bu yöntemler arasında en sık kullanılan iki tanesi; temel bileşenler yöntemi ve en çok olabilirlik yöntemidir.

Üçüncü aşama; özdeğerlerin incelenmesi, yamaç grafiğinin çizimi vb. yaklaşımlarla kaç faktörün dikkate alınacağına karar vermeye çalışılan bir aşamadır.

Dördüncü aşama; ikinci aşamanın bir alt bölümü olarak da düşünülebilecek olan ve faktörleri daha kolay yorumlayabilecek bir yapıya getirme amacını güden faktör döndürme aşamasıdır. Bu amaçla kullanılan birçok faktör döndürme yöntemi vardır.

Son aşama olarak düşünülecek aşama ise elde edilen bulguların tümel olarak yorumlanma aşamasıdır.

Bu adımların birçoğu istatistiksel bakımdan gerekli olmasına rağmen, analizin en önemli adımlarından biri, elde edilen sonuçların yorumlanabilir olmasıdır. Faktörler, o faktör üzerinde yoğunlaşan değişkenlerin ortak özelliklerine göre adlandırılırlar. Bu nedenle, herhangi bir faktörle yüksek ilişkide olan bir değişkenin, diğer faktör (ya da faktörlerle) zayıf bir ilişkide olması istenir. Bu durum, faktör eksenlerinin döndürülmesi ile sağlanabilir (Khalaf 2007).

2.1.6.3. Değişkenlerin ve Veri Setinin Uygunluğu

Veri setinin faktör analizi için uygun olup olmadığını değerlendirmek amacıyla 3 yöntem kullanılır. Bunlar korelasyon matrisinin oluşturulması, Barlett testi ve Kaiser-Meyer-Olkin (KMO) testleridir (Kalaycı 2018).

1- Veri setinin faktör analizi için uygun olup olmadığının tespit edilmesinde ilk adım, değişkenler arasındaki korelasyon katsayılarının incelenmesidir. İstenen, değişkenler arasındaki korelasyonların yüksek olmasıdır. Çünkü değişkenler arasındaki korelasyonlar ne kadar yüksek ise, değişkenlerin ortak faktörler oluşturması o kadar yüksektir. Başka bir ifade ile değişkenler arasında yüksek korelasyonların varlığı, değişkenlerin ortak faktörlerin değişik biçimlerdeki ölçümleri olduğunu gösterir. Değişkenler arasında düşük korelasyonların varlığı ise, değişkenlerin ortak faktörler oluşturamayacaklarının işaretidir.

2- Barlett testi, korelasyon matrisinde değişkenlerin en azından bir kısmı arasında yüksek oranlı korelasyonlar olduğu olasılığını test eder. Analize devam edilebilmesi için "Korelasyon matrisi birim matristir" sıfır hipotezinin reddedilmesi gerekir. Eğer sıfır hipotezi reddedilirse, veri setinin faktör analizi için uygun olduğunu gösterir (Kalaycı 2018).

3- KMO testi gözlenen korelasyon katsayıları büyüklüğü ile kısmi korelasyon katsayılarının büyüklüğünü karşılaştıran bir indekstir. KMO oranının 0,5'in üzerinde

olması gerekir. Oran ne kadar yüksek olursa veri seti faktör analizi yapmak için o kadar iyidir denilebilir. KMO değerleri için Çizelge 2.2'de bir kriterler verilmiştir (Kalaycı 2018).

Çizelge 2.2 KMO uygunluk testi için önerilen kriterler

KMO Ölçüsü	Önerilen Düzey
0,90+	Mükemmel
0,80+	Çok İyi
0,70+	İyi
0,60+	Orta
0,50+	Zayıf
0,50-	Kabul Edilemez

2.1.6.4. Faktör Analizi Modeli

X, p değişkenli N birimlik rasgele veri matrisi olsun. X' in kovaryans matrisi Σ ve ortalama vektörü μ olsun. X gözlem vektörü ile gözlenemeyen faktörler arasında iki tür faktör modeli kurulabilir (Özdamar 2013):

Ortogonal Faktör Modeli, X ile doğrusal olarak bağımlı, fakat aralarında bağımsız olan k tane gözlenemeyen ortak faktörler $F_1, F_2, \dots, F_k$ olduğunu ve p tane hata diye isimlendirilen özel faktörlerin bulunduğunu varsayarak faktörlerin belirlenmesini amaçlar.

Oblik Faktör Modeli ise X ile eğrisel olarak (nonlinear) bağımlı olan k tane gözlenemeyen ortak faktörler $F_1, F_2, \dots, F_k$ olduğunu ve p tane hata diye isimlendirilen özel faktörlerin bulunduğunu varsayarak faktörlerin belirlenmesini amaçlar. Genel olarak ortogonal faktör modeli kullanılır ve faktör modeli denildiğinde doğrusal model anlaşılacaktır.

Faktör analizi modeli;

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \cdots + l_{1k}F_k + \varepsilon_1$$
$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \cdots + l_{2k}F_k + \varepsilon_2$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pk}F_k + \varepsilon_p$$

şeklinde yazılır.

Burada l_{ij} katsayısı, faktör yükü olarak isimlendirilir ve i. değişkenin j. faktör üzerindeki yükünü belirtir.

Matris formunda faktör analizi modeli ise eşitlik (2.20)'de verilmiştir.

$$X - \mu = LF + \varepsilon \qquad (2.20)$$

Burada $X - \mu$, $(p * 1)$ boyutlu fark vektörü,
L, $(p * k)$ boyutlu faktör yükleri matrisi,
F, $(k * 1)$ boyutlu faktör vektörü,
ε, $(p * 1)$ boyutlu hata vektörüdür.

Modelde yer alan L matrisi p değişkeninin her birinin k sayıda $(k \leq p)$ faktör üzerindeki yüklerini belirten l_{ij} katsayılarını içerir. Bu katsayılar faktör yükleri olarak isimlendirilirler. ε_i hatası ise sadece X_i cevabı ile ilgilidir. Ortogonal faktör analizi modelinde varsayımlar aşağıdaki gibi özetlenebilir.

- F faktör matrisinin beklenen değeri sıfırdır $(E(F) = 0)$.
- F matrisinin kovaryansı $Cov(F)$ birim matristir $(Cov(F) = E[FF'] = I)$.
- ε'nin beklenen değeri sıfırdır. (E(ε)=0).
- ε'nin kovaryansı $Cov(e) = E[ee'] = \psi = \begin{bmatrix} \psi_1 & 0 & \dots & \vdots \\ 0 & \psi_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix}$

k ortak faktörlü bir faktör modeli;

$$X = \mu + LF + \varepsilon \tag{2.21}$$

şeklinde yazılır.

X'in kovaryansı, faktör modeline göre

$$S = \mathrm{Cov}(X) = LL' + \psi \tag{2.22}$$

X ile F arasındaki kovaryans ise

$$\begin{gathered}\mathrm{Cov}(X, F) = E(X - \mu)F' = L \\ E(FF') + E(\varepsilon F') = L\end{gathered} \tag{2.23}$$

olarak belirlenir (Özdamar 2013).

Varyansın Bileşenleri

Klasik faktör analizi modelinde, herhangi bir j değişkeninin varyans bileşenleri,

$$s_j^2 = \frac{\sum z_{ji}^2}{n} = \sum_{k=1}^{m} a_{jk}^2 \left(\frac{\sum f_{ki}^2}{n}\right) + d_j^2 \frac{\sum u_{ji}^2}{n} + 2 \sum_{1>k=1}^{m} a_{jk} a_{il} \left(\frac{\sum f_{ki} f_{li}}{n}\right) + 2d_j \sum_{k=1}^{m} a_{jk} \left(\frac{\sum f_{ki} u_{ji}}{n}\right) \tag{2.24}$$

biçiminde gösterilebilir. Standart formdaki her değişkenin varyansı bire eşit olduğundan yukarıdaki ilişki,

$$s_j^2 = 1 = \sum_{k=1}^{m} a_{jk}^2 + d_j^2 + 2 \sum_{1>k=1}^{m} a_{jk} a_{jl} r_{f_k f_l} \left(\frac{\sum f_{ki} f_{li}}{n}\right) + 2d_j \sum_{k=1}^{m} a_{jk}\, r_{f_k u_j} \tag{2.25}$$

biçimine dönüşür.

Artık faktörler her zaman ortak faktörlerden bağımsızdır. Eğer ortak faktörler de kendi aralarında ilişkisiz olurlarsa, formül (2.25)'deki ilişki şöyle olur:

$$s_j^2 = 1 = \sum_{k=1}^{m} a_{jk}^2 + d_j^2 = a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 + d_j^2 \qquad (2.26)$$

Burada a_{jk}^2, f_2 faktörünün z_j'nin varyansına katkısıdır. f_k ortak faktörünün bütün değişkenlerin varyansına toplam katkısı şöyle tanımlanır;

$$V_k = \sum_{j=1}^{1} a_{jk}^2 \qquad (k = 1, 2, \cdots \cdots, m\ ) \qquad (2.27)$$

ve ortak faktörlerin, değişkenlerin toplam varyansına toplam katkısı ise,

$$V = \sum_{k=1}^{m} V_k \qquad (2.28)$$

V/p oranı, faktör analizinin bütünlüğünün göstergesidir.

Toplam birim varyansın bileşenlerini ortaya çıkaran formül (2,26)'da ilişkisi, aynı zamanda faktör analizinin iki önemli kavramına açıklık getirmektedir. Bu kavramlar, ortak faktör varyansı ve artık faktör varyansı kavramlarıdır (Khalaf 2007).

2.1.6.5. Faktörlerin Türetilmesi

Faktör analizinde faktör türetmek için çeşitli yöntemler geliştirilmiştir. Temel bileşenler yöntemi, uygulamada en yaygın kullanılan yöntemlerinden biridir. Nitekim paket programların birçoğunda, faktör türetme yöntemi olarak temel bileşenler yönteminden yararlanılmaktadır (Khalaf 2007).

Faktör türetme modelleri öncelikli olarak iki temel amaçta oluşturulmaktadır: İlk olarak değişkenlerin aralarındaki korelasyonları daha iyi bir şekilde yeniden oluşturabilmek ve değişkenlerin varyansını en yüksek düzeyde açıklamaktır. Araştırmacı amacına göre faktör türetme yönteminin seçimini gerçekleştirmelidir (Cengiz ve Kılınç 2007).

Bütün çıkarma teknikleri, kombinasyonda R'yi yeniden üreten dik bileşenler ya da faktörlerin bir setini hesaplar. Varyansı en üst düzeye çıkartma ya da artık korelasyonları en aza indirgeme gibi çözümleri elde etmek için kullanılan kriterler teknikten tekniğe değişmektedir. Ancak, büyük bir örneklemli, çok sayıda değişken ve benzer ortak varyans kestirimleri içeren çözümlerde farklılıklar küçüktür. Aslında, bir faktör analizi çözümünün durağanlığının testi, hangi çıkartma tekniğinin kullanıldığına bakılmaksızın onun görünümüdür. Söz konusu faktör çıkartma teknikleri çizelge 2.3'te verilmiştir. Son olarak faktör analizi kullanan bir araştırmacı, veri hafiyeliğine karşı tehlikeleri göz önünde bulundurmalıdır. Ön inceleme amacıyla çıkartma teknikleri olarak temel bileşenler analizi ya da temel faktörleri kullanmak ve akabinde de her bir seferde faktör sayısı, ortak varyans kestirimleri ve döndürme yöntemlerini değiştirerek diğer yöntemlerden bir ya da daha fazlasını kullanmak oldukça yaygındır. (Tabachnick and Fidell 2015).

Çizelge 2.3 Faktör türetme süreçlerinin özeti

Türetme (Çıkarma) Tekniği	**Analizin Hedefi**	**Özel İşlevleri**
Temel Bileşenler	Dik bileşenler tarafında en yüksek varyans çıkartması	Ortak, özgün ve hata varyanslarının bileşenler içinde karıştığı, matematiksel olarak belirlenen, ampirik çözüm.
Temel Faktörler	Dik faktörler tarafından en yüksek varyans çıkartması	Değişkenlerden gelen özgün ve hata varyansını ortadan kaldırmak için ortak varyansların kestirimi.
En Yüksek Olabilirlik Faktörlendirilmesi	Gözlenen korelasyon matrisinin örnekleme olabilirliğini en yüksek yapacak evren için faktör yüklerini kestirmek	Özelikle doğrulayıcı faktör analizinde kullanışlı olan faktörler için manidarlık testlerine sahiptir.
İmaj Faktörlendirmesi	Amprik faktör analizi sağlamak	Hata varyansının ve özgün varyansın ortadan kaldırılmasıyla matematiksel olarak belirlenen bir çözüm üretmek için diğer tüm değişkenlerin ortak varyanslar olarak alındığı, bir değişkenin çoklu regresyonuna dayanan varyansları kullanır.
Alfa Faktörlendirilmesi	Dik faktörlerin genellenebilirliğini en yükseğe çıkarmak	
Ağırlıklandırılmamış En Küçük Kareler	Artık korelasyonların karesini en düşüğe indirmek	
Genelleştirilmiş En Küçük Kareler	Artık korelasyonların karesini en düşüğe indirmeden önce değişkenleri paylaşılan varyansla ağırlıklandırmak	

Temel Bileşenler Yöntemi İle Faktörlerin Türetilmesi

Temel bileşenler yöntemi, faktörleri veri matrisinin içerdiği toplam varyansın (ya da toplam bilginin) en büyük bölümünü açıklayacak biçimde türetmeye çalışır. Bu yöntemle faktörler, geometrik olarak faktör eksenlerinin dik döndürülmesi sonucu, birbirinden bağımsız olarak bulunur. Böylece faktörlerin birbirinden faklı bilgileri içermeleri sağlanmış olur. Temel bileşenler yöntemiyle faktörler, aşağıdaki aşamalar izlenerek türetilir: Birinci aşamada n bireyin p sayıda değişken üzerinden almış oldukları nicel değerleri içeren X_{p*n} boyutlu ham veri matrisi oluşturulur. Değişkenlerin ölçü birimleri ve varyansları birbirine yakın ise, analiz varyans - kovaryans matrisi üzerinden, değilse korelasyon matrisi üzerinden yürütülür. Uygulamada genellikle değişkenler farklı ölçeklerle ölçülmüş olduklarından, korelasyon matrisi kullanılır. Bu amaçla, X_{p*n} ham veri matrisi, standartlaştırılarak ölçü birimlerinin etkisinden arındırılır ve R_{p*p} boyutlu korelasyon matrisi, elde edilir (Ergeç 1984).

2.1.6.6 Uygun Faktör Sayısının Belirlenmesi

Faktör analizinde en fazla değişken adedi kadar faktör türetilmektedir. Değişken sayısınca faktör türetildiğinde değişkenler faktörlere eşit bir şekilde dağıtılacağından faydalı bir analiz elde edilememiş demektir. Burada amaç, değişkenler arasındaki ilişkilerden yüksek derecede temsil edecek az sayıda faktör elde etmektir. Gerekli faktör sayısını belirlemek için toplam varyansın her bir faktör tarafından yüzde kaçının açıklandığına bakılması gerekir. Faktör analizinde değişkenler standardize edildiğinden toplam varyans değişken sayısına eşittir (Albayrak 2006). Veri setinde yer alan p değişkeni açıklamak üzere belirlenecek faktör sayısı aşağıdaki kurallara göre belirlenir.

Kaiser Kriteri (Kaiser Criterion): Bu katsayı faktör sayısının karar verilmesine ve faktörler tarafından açıklanan varyansı hesaplamak için oluşturulan bir katsayıdır (Büyüköztürk 2011). S ya da R matrisinin birden büyük kök ($\lambda \geq 1$) sayısı kadar faktör belirlenmektedir (Özdamar 2013).

Cattell Scree Test (Yamaç Eğilim Testi, ScreePlot): Bu yöntem, faktörlerin öz değerleri ile çizilen yığın grafiğinin incelenmesine dayanır. Söz konusu grafiğin dikey ekseninde öz değerler, yatay ekseninde ise faktörler yer almaktadır. Grafik, faktörlerin öz değerleriyle eşleştirilmesi sonucunda bulunan noktaların birleştirilmesiyle elde edilir. Grafikte yüksek ivmeli, hızlı düşüşlerin yaşandığı faktör, önemli faktör sayısını verir (Büyüköztürk 2011)

Varyans Yüzdesi Kriteri: Her bir değişken için açıklanması gereken minimum varyans için gerekli faktör sayısının türetilmesidir. Eğer teorik veya pratik nedenlerle her değişken için minimum bir ortak varyans gerekliyse, bu anlamda değişkenleri uygun şekilde tanımlayabilecek faktör sayısı türetilmelidir. Yöntem değişkenlerin açıklanma düzeyini dikkate almayan ve sadece açıklanan toplam varyansı esas alan yaklaşımlardan farklılaşmaktadır (Albayrak 2006).

Joliffe Kriteri (0,7'den Büyük Özdeğer Sayısı Kadar Faktör Alınması): Bu yaklaşımda 0,7 değerinin aşağısında kalan tüm faktörler modelden çıkarılır (Kalaycı 2018).

Anlaşılabilirlik (Comprehensibility): Seçilecek faktör sayısının değişkenlerin doğasıyla açıklanabilir olacak kadar seçilmesi yaklaşımıdır. Her bir faktörü açıklamakta etkin olan değişkenlerin oluşturduğu yapıların doğal durumlarla uyuşan, mantıklı olarak açıklanabilir olması gerekir. Bu koşul, verilerin birden fazla kez değişik sayıda ($k>2$) faktör olarak faktör analizi yapılması ve uygun çözüme ulaşılması ile sağlanabilir. Pratik bir yaklaşım olarak faktör sayısına karar verirken, verilerin incelenmesi ve açıklayıcı en iyi şekilde verecek bir faktör yapısının deneme ile elde edilmesi tercih edilebilir. Ancak faktör sayısı değiştirilerek, anlamlı bir faktör yapısı ortaya konulmalı ve doğal yapıya uygun çözümlere ulaşılmalıdır. Böylece orijinal değişken yapısına uygun faktör yapısı belirlemek, oluşan faktör yapılarını pratik uygulama alanlarına göre yorumlamak mümkün olabilir (Özdamar 2013).

2.1.6.7 Faktör Yüklerinin Belirlenmesi

Faktör analizinde faktör yüklerini içeren B matrisinin belirlenmesi, faktör analizinin en önemli konusudur. Çünkü faktörler, bu katsayılara göre belirlenmektedir. Bunun için ortak faktör uzayının belirlenmesi gerekmektedir. Bunun için yukarıda kısaca açıklanan modeller kullanılmaktadır. Temel-eksen modelini, faktör analizinin özel bir durumu olan temel bileşen modeliyle karıştırmamak gerekmektedir. PAF (Principal-Axis Factoring – Temel Eksen Faktörü) modeli indirgenmiş korelasyon matrisine dayanmaktadır. Modelin varsayımları dikkate alındığında korelasyon matrisinden indirgenmiş korelasyon matrisi aşağıdaki gibi yazılmaktadır (Albayrak 2006).

$$R = \frac{X'X}{n-1} = \frac{(FB' + S)'(FB' + S)}{n-1} \tag{2.29}$$

ve burada $X = FB' + S$ dir.

$$R = \frac{(BF' + S')(FB' + S)}{n-1}, \tag{2.30}$$

$$R = \frac{BF'FB' + BF'S + S'FB' + S'S}{n-1}, \tag{2.31}$$

$$R = \frac{BF'FB' + \psi + \psi + S'S}{n-1}, \tag{2.32}$$

ve F_i ve S_i bağımsız ise,

$$R = B\frac{F'F}{n-1}B' + \frac{S'S}{n-1}, \tag{2.33}$$

faktörler bağımsız ise $F'F/n-1 = I$ dır. Bu nedenle yeni durum:

$$R = BB' + \frac{s's}{n-1}. \tag{2.34}$$

Burada $S'S/(n-1)$ ifadesi köşegen değerleri spesifik faktörlerin varyansını $\sigma_{s_i}^2$ gösteren köşegen bir matristir. Bu nedenle korelasyon matrisi köşegen değerlerinden bu varyansalar çıkartılırsa aşağıdaki BB'matrisi elde edilmektedir.

$$BB' = R - S'S/(n-1) \tag{2.35}$$

$$BB' = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{12} & 1 & \dots & r_{2p} \\ . & . & . & \dots \\ . & . & . & \dots \\ r_{1p} & r_{2p} & \dots & 1 \end{bmatrix} - \begin{bmatrix} \sigma_{s_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{s_2}^2 & \dots & 0 \\ . & . & . & \dots \\ . & . & . & \dots \\ 0 & 0 & \dots & \sigma_{s_p}^2 \end{bmatrix} = \begin{bmatrix} 1-\sigma_{s_1}^2 & r_{12} & \dots & r_{1p} \\ r_{12} & 1-\sigma_{s_1}^2 & \dots & r_{2p} \\ . & . & . & \dots \\ . & . & . & \dots \\ r_{1p} & r_{2p} & \dots & 1-\sigma_{s_1}^2 \end{bmatrix} \tag{2.36}$$

Görüldüğü gibi BB' matrisinin köşegen değerleri orijinal değişkenlerin toplam varyansından spesifik varyans çıkarıldıktan sonra elde edilen ortak varyansları ve köşegen dışındaki değerleri değişkenler arasındaki korelasyonları göstermektedir. Ortak varyanslar h ile gösterilirse indirgenmiş korelasyon matrisi aşağıdaki gibi yazılmaktadır.

$$\overline{R} = BB' = \begin{bmatrix} h_1^2 & r_{12} & \dots & r_{1p} \\ r_{12} & h_2^2 & \dots & r_{2p} \\ . & . & . & \dots \\ . & . & . & \dots \\ r_{1p} & r_{2p} & \dots & h_p^2 \end{bmatrix} \tag{2.37}$$

Temel eksen modelinde, indirgenmiş korelasyon matrisinden ($\overline{R}$) hareket edilmektedir. Bu yöntemin esası, $\overline{R}$'nin öz değer ve öz değer vektörlerinin belirlenmesine yöneliktir. Bunun için belli kısıtlıyıcı koşullar altında varyansın belli bir yönde maksimumlaştırılması gerekmektedir. $R = BB' + \psi$ eşitliğinden hareket edildiğinde $R = BB'$'nün kesin bir çözümü olabilmesi için, birinci faktörün ağırlıklarının kareleri toplamını ortak varyansta maksimuma ve diğer faktörler içinde, her defasında kalan ortak varyansta maksimuma ulaştıracak şekilde belirlenmesi gerekmektedir. Bunun için;

$$r_{ij} = b_{i1}b_{j1} \qquad (i,j = 1,2,\dots\dots,p \text{ ve } i < j \text{ maksimum} \rightarrow V_1 = \sum_{i=1}^{p} b_{il} \tag{2.38}$$

olması gerekmektedir. Burada V_1, ilk faktörün açıkladığı varyans değerini göstermektedir. Kısıtlayıcı koşullarda bir fonksiyonun maksimum yapılması için

Lagrange çarpanından yararlanılmaktadır. Bu durumda, p adet homojen denklem sistemi elde edilmektedir:

$$\begin{aligned} (1-\lambda)b_{11} + r_{12}b_{21} + \cdots + r_{1p}b_{pl} = 0 \\ r_{21}b_{11} + (1-\lambda)b_{21} + \cdots + r_{2p}b_{pl} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ r_{p1}b_{11} + r_{p2}b_{21} + \cdots + (1-\lambda)b_{pl} = 0 \end{aligned} \qquad (2.39)$$

Bu eşitlik sistemi, öz değerleri yansıtan reel ve simetrik bir matristir. Genel formülle yazılırsa: $Ru_1 = \lambda_1 u_1$ veya $(R - \lambda_1 I)u_1 = 0$ olmaktadır. Burada λ_1'ler R'deki öz değerler ve u_1'ler de bunlara ilişkin öz vektörlerdir. Yukarıdaki eşitlik sisteminin sıfırdan farklı çözümü için zorunlu ve yeterli koşul, sistemin determinantının sıfıra eşit olmasıdır:

$$\begin{bmatrix} (1-\lambda) & r_{12} & \dots & r_{1p} \\ r_{21} & (1-\lambda) & \dots & r_{2p} \\ . & . & . & \dots \\ . & . & . & \dots \\ r_{pl} & r_{p2} & \dots & (1-\lambda) \end{bmatrix} = 0 \qquad (2.40)$$

(2.39) eşitliği sistemin karakteristik denklemleri, (2.40) eşitliği ise karakteristik determinantıdır. Sistemin determinantı için (2.40) eşitliğinde p reel çözüm söz konusudur: $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$. Bu lamda değerleri korelasyon matrisinin öz değerleridir. (2.40)'den elde edilen λ_1 değerinin, (2.39)'da yerine konması durumunda yukarıdaki kısıtlayıcı koşulları sağlayan ve $\sum_{i=1}^{p} u_{il}^2$ için maksimum olan bir çözüm vektörü $(u_{11}, \dots, u_{p1})$ elde edilmektedir. Aynı şekilde λ_2 için, kalan ortak varyansta bir maksimum olan $\sum_{i=1}^{p} u_{i2}^2$ için $(u_{22}, \dots, u_{p2})$ öz vektörü elde edilmektedir. Benzer biçimde (2.40)'tan elde edilen λ_p değerinin (2.39)'da yerine konması durumunda yukarıdaki kısıtlayıcı koşulları sağlayan ve $\sum_{i=1}^{p} u_{i1}^2$ için maksimum olan bir çözüm vektörü $(u_{1p}, u_{2p}, \dots, u_{pp})$ elde edilmektedir (Albayrak 2006).

$$b_{il} = \frac{u_{il}\sqrt{\lambda_1}}{\sqrt{u_{11}^2 + u_{21}^2 + \cdots + u_{pl}^2}} \quad (2.41)$$

2.1.6.8 Faktörlerin Döndürülmesi

Faktör analizinde araştırmacı elde ettiği m kadar önemli sayılabilecek faktörü, daha basite indirgeyerek kolay açıklayabilmek ve bağımsızlık elde edebilmek amacıyla bir eksen döndürmesine tabii tutabilir. Faktör döndürmesi, çözümlerin temel matematiksel özelliklerini değiştirmez. Eksenlerin döndürülmesi sonrasında değişkenlerin bir faktördeki yükü artarken diğer faktörlerdeki yükleri azalır. Böylece faktörler, kendileriyle yüksek ilişki veren değişkenleri bulurlar ve faktörler daha kolay yorumlanabilir (Büyüköztürk 2011).

Bazı durumlarda faktör yüklerinden bilgi elde edilmek zorlaşabilmektedir. Bu sebeple elde edilen faktör yapılarını daha anlaşılır bir şekle getirmek için onları belirli bir açı ile döndürmek uygun olur. Bu işlemi bir mikroskop altında bir preparatı en iyi biçimde görebilmek için lameli mikrometrik olarak ileri geri döndürerek manipile etmeye benzetebiliriz. Faktör rotasyonu işlemi, **L** matrisini **TT'=I** koşulunu sağlayan bir ortogonal matris ile çarpılarak yeni faktör yükleri matrisi elde etmektedir (**L*=LT**). Faktör rotasyonu ile faktörlere atfedilen varyans, spesifik varyans, korelasyon (ya da kovaryans) matrisi değişmez. Faktör yükleri matrisinin bağımsız yapıyı elde etmek üzere döndürülmesi ile orijinal değerlendirmek mümkün olur. Döndürme işlemi bir matematiksel yaklaşımdır. Her bir faktörü ağırlıklı olarak etkili olan değişkenlerin belirgin olarak ortaya konmasını sağlar (Özdamar 2013).

Farklı döndürme yöntemleri, farklı faktörlerin teşhisiyle sonuçlanabilir. Faktör döndürmesinde iki farklı yöntem kullanılmaktadır. Eğer eksenler doğru açılarda, konumlarını değiştirmeden, 90 derecelik açı ile döndürülürse döndürme "dik döndürme" olarak adlandırılmaktadır. İkinci yöntem ise her faktörün birbirinden bağımsız olarak döndürüldüğü "eğik döndürme" dir. Bu döndürmede eksenlerin birbirine dik olması gerekli değildir. Dik döndürmede tek bir açıya gereksinim duyulurken, eğik döndürmede iki farklı açı bulunmaktadır (Pazarlıoğlu vd. 1999).

Dik faktör döndürme yöntemleri arasında en sık kullanılanları (Bulam 2005);

1. Quartimax yöntemi
2. Varimax yöntemi

Quartimax yöntemi iki faktör olması durumunda en iyi sonuç veren yöntemlerden biridir. Bu yöntemde basit yapıya ulaşmada faktör yükleme matrisinin satırları göz önünde bulundurulur. Quartimax yöntemi basit yapıların kurulmasında çok başarılı sonuçlar vermektedir. Bu yöntemin tam tersi olan yöntem varimax yöntemidir. Bu yöntemde basit yapıya ulaşmada faktör yükleme matrisinin sütunlarına öncelik verir ve her sütundaki yük değerleri 1'e yaklaştırılırken geriye kalan çok sayıda yük değeri 0'a yaklaştırılır. Varimax yöntemini maksimize eden sadece ve sadece tek bir faktör yükleme matrisi vardır. Bu yöntemde de diğer yöntemlerde olduğu gibi faktör varyanslarının maksimum olmasını sağlayacak biçimde döndürme yapılır. Bu yöntem faktörlerin dik döndürülmelerini elde eden bir analitik yaklaşım olarak çok iyi sonuçlar vermektedir (Pazarlıoğlu vd. 1999),

Eğik döndürme son yıllarda çok kullanılan ve daha iyi sonuç veren yöntemlerdir. Değişkenleri gösteren her bir noktanın döndürülmüş eksenler üzerindeki izdüşümlerinin yorumlanmasında verilen noktaların eksenler üzerindeki izdüşümleri eksenlere paralel doğrularla bulunur. Eğik döndürme yöntemleri arasında en çok kullanılanları:

1. Oblimax yöntemi
2. Quartimin yöntemi
3. Oblimin yöntemi
4. Direk oblimin yöntemi

olarak bilinirler (Bulam 2005).

2.1.7 Kümeleme Analizi

Kümeleme analizi, sınıflandırma yapabilmek için çok sayıda işlemi kapsayan birçok işlemi bir arada yapabilen kavramlar bütünüdür. 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 görece benzer (homojen) gruplar şeklinde tekrar düzenlemeyi sağlayan çok değişkenli bir istatistiksel yöntemdir (Aldenderfer and Blashfield 1984). Hair vd. (2006) kümeleme analizinin temel amacını nesnelerin ya da birimlerin özeliklerine göre sınıflandırarak gruplara ayırmak olan bir çeşit çok değişkenli teknik olarak tanımlamaktadır (Çokluk vd. 2010).

Çok boyutlu uzayda büyük ve karmaşık verilerin özetlenmesi ve tanımlanmasında yol gösterici bir yöntem olan kümeleme analizi (cluster analysis), son yıllarda en sık kullanılan çok değişkenli istatistiksel yöntemlerden biridir. Kümeleme analizi ile gözlenen birimlerin değişik sınıflar oluşturup oluşturmadıkları belirlenir. Everitt (1974) tarafından yapılan tanıma göre kümeleme, benzer olan bireylerin aynı gruba, benzer olmayan bireylerin ise farklı grup veya gruplara girmesi anlamına gelmektedir. Sharma (1996) ise kümeleme analizini, bir araştırmada incelenmek istenen değişkenlere ait birimlerin arasındaki birbirine benzeyenleri belli bir grup içerisinde toplayarak sınıflandırma yapmak, birimlerin ortak özelliklerini ortaya koymayı ve bu sınıflar ile ilgili genel tanımlamalar yapmayı sağlayan bir yöntem olarak tanımlamıştır (Kılıç 2008).

Kümeleme analizinde genel olarak amaç, kümelenmemiş verileri birbirleriyle benzerlik durumlarına göre sınıflamak (gruplamak) ve araştırmacıya; uygun, işe yarar özetleyici bilgiler elde etmek için yardımcı olmaktır. Birimlerin ya da nesnelerin gruplandırılmasında kümeleme ve diskriminant analizleri birbirleriyle benzerlik göstermektedir ancak iki yöntemin önemli farklılıkları da mevcuttur. Öncelikle, diskriminant analizinde grup sayısı (küme sayısı) bilinmektedir ve bu sayı analiz süresi içerisinde değişmemektedir. Araştırmacı, birimleri ya da nesneleri önceden bilinen bu gruplara ya da kümelere sınıflandırmaktadır. Ayrıca diskriminant analizi ile oluşan

sonuçlar, bilgiler (ayırma fonksiyonu) gelecekte kullanılabilmektedir. Bununla birlikte kümeleme analizinde küme sayısı önceden bilinmemekte (bilinmesi durumunda zaten kümeleme analizinin kullanılması anlamsızdır) ve sadece verilerin mevcut durumuna ilişkin sonuçlar vermesi nedeniyle gelecekte kullanılabilmesi söz konusu olamamaktadır (Tatlıdil 1992).

Birimleri ya da değişkenleri, değişkenler arası benzerlik (similarity) ya da farklılıkları (dissimilarity) dayalı olarak hesaplanan bazı ölçülerden yararlanarak homojen gruplara bölmek amacıyla kullanılan kümeleme analizi temel olarak dört değişik amaca yönelik işlev görür (Özdamar 2013).

1. n sayıda birimi, nesneyi, oluşumu, p değişkene göre saptanan özelliklerine göre olabildiğince kendi içinde türdeş (homojen) ve kendi aralarında farklı (heterojen) alt gruplara (küme) ayırmak,
2. p sayıda değişkeni, n sayıda birimde saptanan değerlere göre ortak özellikleri açıkladığı varsayılan alt kümelere ayırmak ve ortak faktör yapıları ortaya koymak,
3. Hem birimleri hem de değişkenleri birlikte ele alarak ortak n birimi p değişkene göre ortak özellikli alt kümelere ayırmak,
4. Birimlerin, p değişkene göre saptanan yapılar aracılığı ile toplumdaki doğal olarak oluşturdukları düşünülen biyolojik ve tipolojik sınıfları ortaya koymak.

Grupların (kümelerin) elde edilmesi ve dolayısıyla veride var olan durumun belirlenmesi kümeleme analizinin temel amacıdır. Diğer bir deyişle kümeleme analizinin ileriye yönelik kestirim, vb. amaçlı kullanımı da söz konusu değildir. Kümeleme analizi ve faktör analizi arasında da bir benzerlik vardır. Şöyle ki kümeleme analizi ile değişkenleri kümelemek de olanaklı olmakla birlikte, bu amaç faktör analizinin en temel amacıdır. Bu nedenle, değişkenlerin kümelenmesi gibi bir amaç için faktör analizinden yararlanmak daha çok tercih edilmektedir (Alpar 2011).

Kümeleme analizi, genel amaçlarının yanı sıra aşağıdaki gibi bir takım özel amaçlar içinde kullanılabilmektedir (Tatlıdil 1992).

- Gerçek tiplerin (cinslerin-ırkların) belirlenmesi
- Model uydurmanın kolaylaştırılması
- Gruplar için ön tahminde bulunulması
- Hipotezlerin testi
- Veri yapısının netleştirilmesi
- Veri indirgenmesi (veriler yerine kümelerin değerlendirilmesi)
- Aykırı değerlerin tespit edilmesi

2.1.7.1 Kümeleme Yöntemleri

Kümeleme yöntemleri hiyerarşik ve hiyerarşik olmayan (bölümlemeli, aşamalı) kümeleme yöntemleri olarak iki grupta incelenmekte ancak bu konuda yapılan araştırmalar bu algoritmaların daha alt bölümlere ayrılabileceğini göstermektedir (Berkhin 2004). Hiyerarşik kümeleme yöntemlerinde özellikle teorik yapının daha kolay anlaşılabilmesi için dendogramlardan (ağaç grafiği) yararlanılır. Hiyerarşik kümeleme yöntemlerinin en çok kullanılanları; tek bağlantılı, tam bağlantılı, ortalama bağlantı, merkezi ve Ward metotlarıdır. Hiyerarşik olmayan kümeleme yöntemlerinde küme sayısının önceden belirlenebilmesi ya da küme sayısının durumu hakkında bir ön bilginin olması, araştırmacı tarafından küme sayısına anlamlı olacak şekilde önceden belirlenmesi durumlarında tercih edilmektedir. Hiyerarşik olmayan kümeleme yöntemlerinde ise daha çok Mac Queen tarafından geliştirilen k-Ortalama tekniği ve en çok olabilirlik teknikleri tercih edilen yöntemler arasındadır (Sarıman 2011).

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 yöntemlerdir. Aşağıda bu iki temel algoritma içerisinde yer alan yöntemlere de yer verilerek bir sınıflandırma sunulmaya çalışılmıştır (Çokluk vd. 2010).

1. Hiyerarşik Kümeleme Yöntemleri

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)

2. Hiyerarşik Olmayan Kümeleme Yöntemleri

a. K-Ortalama (K-Means) Yöntemi

b. Medoid Parçalama Yöntemi

c. Yığma / Yığılma Yöntemi

d. Fuzzy (Bulanık) Kümeleme Yöntemi

Hiyerarşik (Aşamalı) Kümeleme Yöntemleri

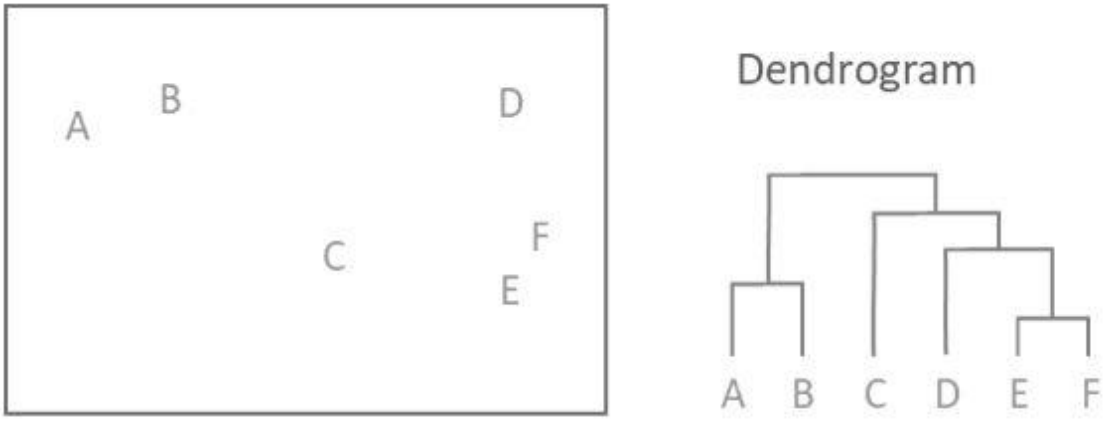
Hiyerarşik kümeleme yöntemleri, veri setlerinin aralarındaki uzaklık değerlerini ve benzerliklerini elde ederek, veri setinde bulunan birimlerin hiyerarşik bir şekilde ayrışmasını sağlar. Bu yöntemlerde araştırma sırasında ağaç veri yapısı olarak da bilinen dendogram kullanılmaktadır. Bu dendogramlar, hiyerarşik kümeleme yöntemleri ile elde edilen kümeleri görselleştirir (Pektaş 2013).

Veri matrisinde var olan değişkenlere ait birimlerin, başlangıç aşamasında kaç küme oluşturduğuna ve küme elemanlarını belirlemede başlangıçta hangi kriterlerin seçildiğine göre aşamalı yöntemler iki ana gruba ayrılır (Özdamar 2013).

a) Birleştirici aşamalı kümeleme yöntemleri
b) Ayırıcı aşamalı kümeleme yöntemleri

Birleştirici aşamalı kümeleme teknikleri ilk önce her birimin bir küme oluşturduğunu kabul eder ve sırasıyla $n, n-1, n-2, \ldots, n, 3, 2, 1$ bu birimleri aşamalı olarak kümeye yerleştirmek amacıyla kullanılan tekniklerdir. Bu yöntemler birimlerin birbirleri ile hangi aşamada ve hangi benzerlik düzeyinde ortak özelliklere sahip kümeler oluşturduklarını göstermeleri açısından yaygın olarak kullanılan aşamalı kümeleme yaklaşımıdır. Ayırıcı aşamalı kümeleme yöntemleri ise başlangıçta tüm birimlerin bir küme oluşturduğunu varsayarak, n birim sırasıyla aşamalı olarak $1, 2, 3, \ldots., n-r, n-3, n-2, n-1, n$ küme oluşturmayı amaç edinen bir tekniktir. Ayırıcı aşamalı kümeleme yöntemleri literatürde sıklıkla kullanılan yöntemler değildir. Birleştirici aşamalı kümeleme yönteminde, her birimin başlangıç için bir küme oluşturduğu kabul edilmektedir. Sonraki aşamada ise birbirleriyle yüksek derecede benzerlik göstermekte olan iki birim birer küme oluşturmaktadır. Son aşamada ise oluşan kümeye farklı farklı benzerlik düzeyleri olan diğer birimler eklenerek birimlerin tümü bir küme oluşturacak şekilde birbirleriyle bağlanırlar. Bağlantılar, uzaklıklar ve birimlerin bağlanma düzeyleri dendogram adı verilen ağaç grafikleri ile gösterilirler (Özdamar 2013).

Ağaç diyagramlarında, değişkenlerin hangi aşamalarda ve hangi uzaklık (ya da benzerlik) düzeylerinde bir araya geldiklerini görmek olanaklıdır. Bir ağaç grafiği örneği Şekil 2.3'de verilmiştir. Bu diyagramların bir özelliği, aynı yere bağlanan herhangi iki gözlemin yer değiştirilebilmesidir. Örneğin Şekil 2.3'de aynı yere bağlanan A ve B gözlemleri yer değiştirebilir. Şekil 2.3'deki ağaç diyagramında E ve F'nin en çok benzer olduğunu görebiliriz, çünkü onları birleştiren bağlantının yüksekliği en küçüktür. Sonraki en çok benzeyen iki gözlem ise A ve B'dir. Şekil 2.3'deki dendrogramda, dendrogramın yüksekliği kümelerin birleştirildiği sırayı gösterir. Bu durumda, dendrogram bize kümeler arasındaki büyük farkın A ve B kümeleri ile C, D, E ve F kümeleri arasındaki fark olduğunu göstermektedir (Alpar 2011).



Şekil 2.3 Hiyerarşik kümeleme yöntemleri için ağaç diyagramı

■ Tek Bağlantı Kümeleme Yöntemi (TekBKY)

En sade hiyerarşik kümeleme tekniğidir. En yakın komşuluk tekniği olarak ta bilinmektedir. Tek bağlantı kümeleme yönteminde uzaklıkların oluşturduğu matristen faydalanılmakta ve birbirleri ile en yakın birim veya kümeler birleştirilmektedir. Bu birleştirmeler tüm birimlerin kümelere atanması sonlanana dek devam edilir. Birleştirme yapılırken kümelerin eleman sayısının birden fazla olması koşulu yoktur. Bir birim yalnız başına bir küme oluşturabilir. İki küme arasındaki uzaklık, bir kümedeki bir gözlem ve diğer kümedeki bir gözlem arasındaki minimum mesafedir. Bu yöntemde eğer i. ve j. birimler birleştirilmiş ise birleştirilen kümenin k. küme ile ilişkisi uzaklık ölçütü olarak (Dinler 2014),

$$d_{k(i,j)} = \min(d_{ki}, d_{kj}) \qquad (2.42)$$

biçiminde ifade edilmektedir.

Eşitlikte;

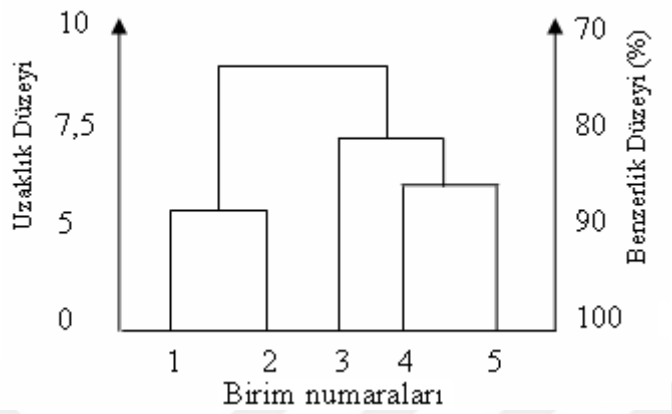
$d_{k(i,j)}$, k. kümenin i. ve j. kümelerle olan uzaklığını,

d_{kj}, k. kümenin j. kümeye olan uzaklığını,

d_{ki}, k. kümenin i. küme ile olan uzaklığını

göstermektedir.

TekBKY'nin işleyişini daha açık bir şekilde göstermek üzere beş birime (bireye) ait Şekil 2.4'te verilen dendrogram örneği aşağıda da gösterilmiş ve kısaca açıklanmıştır.



Şekil 2.4 TekBKY'ne ilişkin örnek dendrogram

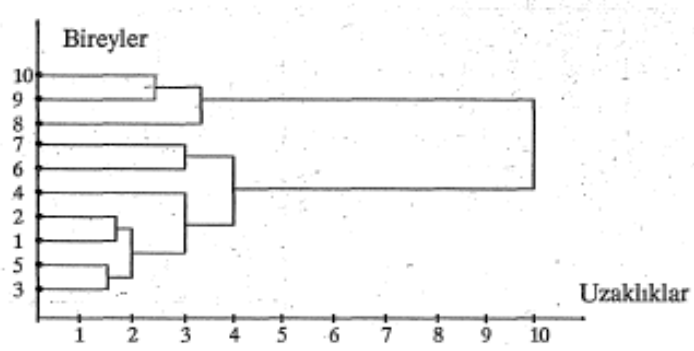
Şekil 2.4.'te görüldüğü gibi, öncelikle uzaklık düzeyi minimum olan veya benzerlik düzeyi maksimum olan 1 ve 2. birimler bir küme oluşturmuş ve daha sonra oluşan yeni uzaklık veya benzerlik düzeylerine göre birbirlerine en yakın olan 4 ve 5. birimler bir küme oluşturmuştur. Bu durumun tekrarlanmasıyla birlikte önce 3 ile 4, 5. birimler birleşmiş ve son olarak 1, 2 ile 3, 4, 5. birimler birleşmiştir. Böylece; 5 civarındaki uzaklıkta (yaklaşık %90 benzerlikte) [(1,2)-3-4-5] şeklinde dört küme, 6 civarındaki uzaklıkta (yaklaşık %85 benzerlikte) [(1,2)-3-(4,5)] şeklinde üç küme, 7 civarındaki uzaklıkta (yaklaşık %80 benzerlikte) [(1,2)-(3,4,5)] şeklinde iki küme ve daha yüksek uzaklıkta (veya daha düşük benzerlikte) [(1,2,3,4,5)] şeklinde bir küme oluşmaktadır (Kılıç 2008).

■ Tam Bağlantı Kümeleme Yöntemi (TamBKY)

En uzak komşuluk (furthest neighbour - complete Iinkage - CLINK) olarak da bilinen bu teknik Johnson tarafından önerilmiştir. Tek bağlantı tekniğine çok benzeyen bu teknikte tek farklılık iki küme arasındaki uzaklık olarak her kümedeki eleman çiftleri arasındaki uzaklığın en büyüğü ele alınmaktadır. Bu farklılığı yukarıdaki (2.42) eşitliğine paralel olarak,

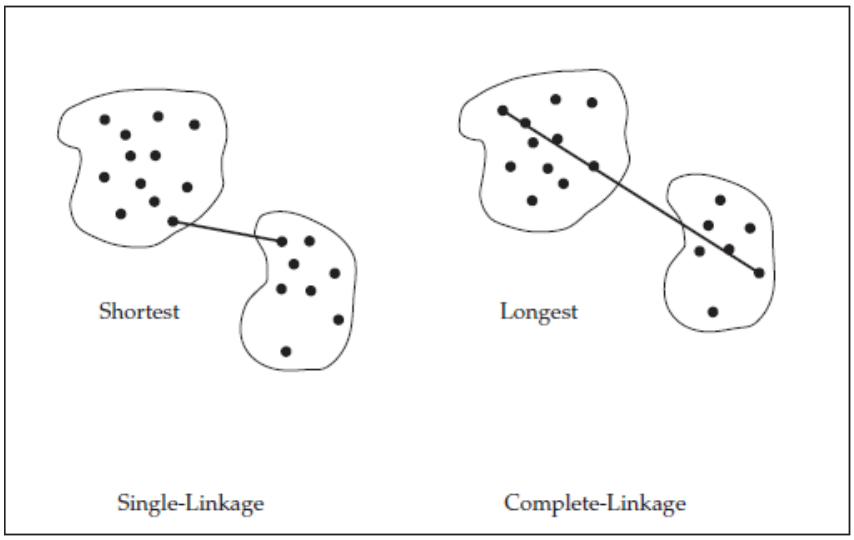
$$d_{k(i,j)} = \max(d_{ki}, d_{kj}) \quad (2.43)$$

biçiminde gösterebiliriz. Bu farklılığın şekilsel ifadesi de Şekil 2.5'de verilmiştir (Tatlıdil 1992).



Şekil 2.5 TamBKY için örnek dendrogram

Tam bağlantı yöntemi (ayrıca en uzak komşuluk olarak da bilinir) tek bağlantı algoritması ile benzerlik göstermektedir. Bu yöntem her kümedeki gözlemler arasındaki maksimum mesafeye dayanmaktadır. Eğer benzerlik ölçüleri kullanılmışsa kümeler arasında, tüm gözlemleri içine alabilen en küçük (minimum çap) bireylerin kümelendirilmesiyle işlem yapılır. Bu yönteme tam bağlantı adı verilir çünkü kümedeki tüm birimler birbirine maksimum mesafeden bağlanmıştır. Bu teknik, tek bağlantı ile tanımlanan zincirleme problemini ortadan kaldırmaktadır ve en kompakt kümeleme çözümlerini oluşturduğu düşünülmektedir. Şekil 2.6'da, en kısa (tek bağlantı) ve en uzun (tam bağlantı) mesafeleri karşılaştırılmaktadır (Hair *et al.* 2014).



Şekil 2.6 Tam bağlantı ve tek bağlantı kümeleme yöntemlerinin karşılaştırılması

■ Ortalama Bağlantı Kümeleme Yöntemi (OrtBKY)

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

m. kümenin j. küme ile olan uzaklığı (d_{mj}); (2.44)' deki eşitlik ile belirlenir.

$$d_{mj} = \left(N_k d_{kj} + N_l d_{lj}\right)/N_m \qquad (2.44)$$

Eşitlikte,

N_k = k'ıncı kümenin eleman sayısını

N_l = l'inci kümenin eleman sayısı

N_m = m'inci kümenin eleman sayısını

göstermektedir.

En çok benzer çiftin bulunmasında grup içi benzerlik ortalaması k ve l kümelerine ait çiftlerin benzerlik ölçülerinden ve birim sayılarından yararlanılarak hesaplanır (Özdamar 2013).

■ Ward Bağlantı Kümeleme Yöntemi

N bireyden elde edilen p karaktere ilişkin verilere göre bireyler (gözlem değerleri) k kümeye ayrılmak istensin. Bu durumda k kümesinde n_k birey vardır ve bireylerin i nci karakter ortalaması (Kayaalp vd. 2000); $\overline{X}_{ik}$;

$$\overline{X}_{ik} = \frac{\sum_{j=1}^{j=n_k} x_{ijk}}{n_k} \tag{2.45}$$

olarak gösterilir.

Eşitlikte,
x_{ijk} = k'ncı kümenin j nci bireyinin i nci gözlem değerini
n_k = k'ncı kümenin eleman sayısını
göstermektedir.

k. kümede yer alan n_i noktanın k kümesinin ortalamalar vektörüne olan Öklid uzaklıkları toplamı, hata kareler toplamı (HKT) ile gösterilecek olursa;

$$HKT_k = \sum_{i=1}^{p}\sum_{j=1}^{n_k}\left(X_{ijk} - \overline{X}_{ik}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{n_k} X_{ijk} - n_k \sum_{i=1}^{p} \overline{X}_{ik}^{\,2} \tag{2.46}$$

şeklinde hesaplanır. HKT_k değeri tüm $k = 1, 2, \ldots\ldots\ldots, n$ kümelerde hesaplanarak kümeler içi HKT'nın toplamı eşitlik (2.47) gibi olur.

$$HKT = \sum_{j=1}^{k=1} HKT_k \tag{2.47}$$

Bu değerler hesaplandıktan sonra bireylerin kümelerde birleştirilmesi aşamalı olarak aşağıdaki şekilde yapılır.

1- İlk başta $HKT_k = 0$ olacak şekilde her birim bir küme kabul edilir.
2- İkinci aşamada HKT' ında en küçük artışı sağlayan (u) ve (v) birleştirilerek (uv) kümesi oluşturulur. HKT' daki bu artış;

$$\Delta HKT_{uv} = HKT_{(uv)} - HKT_{(u)} - HKT_{(v)} \qquad (2.48)$$

olarak hesaplanır ve bu suretle n birim (n-1) kümeye ayrılmış olur.
3- Küme sayısı $k = 1$ oluncaya kadar 2 adım tekrarlanarak tüm bireylerin aşamalı olarak birbirine bağlanmaları sağlanır.

Böylece her aşamada HKT'da ki oluşan minimum artış, birleştirilen kümelerin (küme ortalamaları) merkez noktaları arasındaki Öklid uzaklığının karesi ile orantılı gerçekleşmiş olur. Ward bağlantı yöntemi centroid ve medyan bağlantı kümeleme yöntemlerinin karma ve ağırlıklı biçimidir (Kayaalp vd. 2000).

■ Medyan (Ortanca) Bağlantı Kümeleme Yöntemi

Bu yöntem, genellikle değişkenlerin değerlerinin sıralı ölçekle elde edildiği veya ölçüm değerleri yerine skor değerleri ele alındığında ilgili kümelerin ortaya çıkarılmasında kullanılır. Ortalama Bağlantı Kümeleme yönteminde olduğu gibi, m kümesinden bir birim ve q kümesinden bir birim alınarak oluşturulan tr çiftinin benzerlik ölçüsü S_{tr}, m ve q kümelerine ait bu birimlerin benzerlik ölçülerinin toplanması $(S_{tr} = S_{mr} + S_{qr})$ ile hesaplanır. Kümeler arasındaki mesafe;

Uzaklık türü bir ölçüt kullanıldığında,

$$d_{tr} = \frac{1}{2}(d_{mr} + d_{qr}) + \frac{1}{4}(d_{mq}). \tag{2.49}$$

Benzerlik (ilişki) türü bir ölçüt kullanıldığında ise,

$$s_{tr} = \frac{1}{2}(s_{mr} + s_{qr}) + \frac{1}{4}(1 - s_{mq}) \tag{2.50}$$

formülleri ile hesaplanır. Kümelerde yer alan birimler, yeni oluşan kümelerle olan uzaklık veya yakınlıkları dikkate alınarak birbirleri ile birleştirilir ve tüm birimler birbirleri ile uygun bir biçimde birleşinceye kadar (2.49) veya (2.50) eşitliğinde verilen işlem tekrarlanır (Kılıç 2008).

■ Küresel (Merkezi) Bağlantı Kümeleme Yöntemi (Centroid Linkage)

Centroid yönteminde, A ve B şeklindeki iki değişkenin ortalama vektörleri (genellikle centroid olarak adlandırılanlar) arasındaki Öklid uzaklığından yararlanılarak kümeler oluşturulur.

$$D(A, B) = d(\bar{y}_A, \bar{y}_B) \tag{2.51}$$

Burada $\bar{y}_A$ ve $\bar{y}_B$, A ve B gözlem vektörleri için ortalama vektörlerdir. Centroid yöntemi öklid uzaklığı kullanılarak hesaplanmaktadır. Daha önceki bölümlerde Öklid uzaklığı tanımlanmış ve $d(\bar{y}_A, \bar{y}_B)$ Öklid uzaklığı ile hesaplanır. Burada $\bar{y}_A$ ve $\bar{y}_B$ $\bar{y}_A = \sum_{i=1}^{n_a} y_i / n_A$ eşitliği ile bulunur. Böylece her adımda $\bar{y}_A$ ve $\bar{y}_B$'nin küme merkezleri arasındaki en küçük mesafe birleştirilir. İki küme (A ve B) birleştirildikten sonra, yeni AB kümesinin ağırlıklı ortalamaya göre centroidi verilir (Rencher 2003).

$$\bar{y}_{AB} = \frac{n_A \bar{y}_A + n_B \bar{y}_B}{n_A + n_B}. \tag{2.52}$$

■ **McQuitty Bağlantı Kümeleme Yöntemi (Mcquitty Linkage Method)**

m. kümenin oluşumunda k. ve l. kümelerin j. küme ile olan uzaklıkları toplamının yarısı (ortalaması) dikkate alınarak belirlenir. Ağırlıksız ortalama bağlantı yöntemi olarak da bilinmektedir. Yeni oluşan m ve j kümeleri arasındaki uzaklık;

$$d_{mj} = \left(d_{kj} + d_{lj}\right)/2 \qquad (2.53)$$

biçiminde belirlenir (Özdamar 2013).

Hiyerarşik (Aşamalı) Olmayan Kümeleme Yöntemleri

Gözlemleri, incelenen değişkenler yönünden kendi içlerinde homojen ve kendi aralarında heterojen yapıda kümelere ayırmayı sağlayan yöntemler bütünüdür. Bu yöntemlerde, küme sayısı önceden belirlenir. Diğer bir yaklaşımla, eğer oluşacak küme sayısı hakkında önsel bir bilgi var ise hiyerarşik olmayan yaklaşımları kullanmak daha çok tercih edilmektedir. Hiyerarşik olmayan kümeleme yöntemlerinin hiyerarşik kümeleme yöntemlerine göre bazı avantajları şöyle özetlenebilir (Alpar 2011):

1. Hiyerarşik kümeleme yöntemleri daha çok küçük veri setleri için uygun iken, hiyerarşik olmayan kümeleme yöntemleri çok daha büyük veri setlerine (n>1000) kolaylıkla uygulanabilmektedir; çünkü bu yöntemlerin başlangıcında hiyerarşik küme yöntemlerindeki gibi gözlem sayısı boyutlarında (n*n) benzerlik ve uzaklık matrisleri hesaplanmaz.

2. Hiyerarşik olmayan kümeleme yöntemlerinin hiyerarşik kümeleme yöntemlerine göre bir diğer avantajı verideki aykırı değerlere karşı daha az duyarlı olmasıdır.

Hiyerarşik olmayan kümeleme yöntemleri değişkenler ile ilgilenmekten önce n adet birimin k sayısı kadar kümeye atanması ile ilgilenmektedir. Gözlem sayısının çok fazla olduğu durumlarda ya da küme sayısı önceden biliniyorsa yani küme sayısı hakkında bir ön bilgi mevcut ise hiyerarşik olmayan kümeleme yöntemlerinin tercih edilmesi söz konusu olmaktadır. Çünkü bu yöntemlerde uzaklıklar matrisinin kullanılması zorunluluğu yoktur ve doğrudan ham verilerle (X matrisi ile) çalışılır (Çakmak 1999).

Hiyerarşik olmayan kümeleme metotları bölmeli metotlar olarak da bilinmektedir. Bölmeli metotları, hiyerarşik metotlara oranla daha büyük veri setlerine uygulanabilir. Bölmeli metotlarında oluşturulacak k adet kümede her bir küme en azından bir birim içerir ve her birim yalnızca bir grupta bulunur. Bölmeli metotlarda işlemler şu sıra ile yapılır: İlk olarak başlangıç küme merkezleri gelişi güzel olarak seçilir. Birimlerin, belirlenen kümelerin merkezlerine olan uzaklıklarına göre yeni küme merkezleri oluşturulur. Bu işlemler birbirlerinden farklı, içlerinde homojen, birbirleri arasında benzerlik bulunmayan k adet küme oluşturuluncaya kadar sürdürülür (Pektaş 2013).

Hiyerarşik olmayan kümeleme yöntemlerinin gerek teorik dayanaklarının hiyerarşik kümeleme yöntemlerine göre daha güçlü olması gerekse küme sayısı konusunda ön bilgi olması ya da araştırmacının anlamlı olacak küme sayısına karar verebilmesi aşamalı kümeleme yöntemlerine göre tercih edilmesini sağlamaktadır (Doğan 2002).

Hiyerarşik olmayan yöntemlerin hiyerarşiklerden temel farkları şöyle özetlenebilir (Akat 2007, Johnson and Wichern 1998, Çokluk vd. 2010):

- 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 hiçbir şey ifade etmez.
- Hiyerarşik olmayan yöntemler, hiyerarşik yöntemlere göre daha büyük veri kümelerine uygulanabilir.

■ K – Ortalamalar Kümeleme Yöntemi

En eski kümeleme algoritmalarından olan k-Means, 1967 yılında J.B. MacQueen tarafından geliştirilmiştir. En yaygın kullanılan gözetimsiz öğrenme yöntemlerinden birisi olan k-Means'in atama mekanizması, her verinin sadece bir kümeye ait olabilmesine izin verir. Bu nedenle, keskin bir kümeleme algoritmasıdır. Merkez noktanın kümeyi temsil etmesi ana fikrine dayalı bir metottur (Han and Kamber 2006). Eşit büyüklükte küresel kümeleri bulmaya eğilimlidir (Işık ve Çamurcu 2007).

Bu yöntemde küme sayısı; en az küme sayısı 2, en fazla küme sayısı ise gözlem sayısına eşit ya da daha az olacak şekilde belirlenir. K-Ortalama yönteminin amacı gözlemleri, sayısı araştırıcı belirlenen kümelere sınıflamaktır. Sonuçta, k-Ortalama yöntemi algoritmaları yardımıyla gözlemler; kümeler arasındaki değişkenlik en büyük, kümeler içi değişkenlik en küçük olacak şekilde farklı kümelere yerleştirilir (Alpar 2011).

K-Ortalamalar yönteminin sağlanabilmesi için iki koşul sağlanmalıdır (Pektaş 2013):

1. Veri setindeki değişkenler en azından aralıklı ölçekte bulunmalıdır.
2. Oluşturulacak olan küme sayısı başlangıçta biliniyor olmalıdır.

Bu yöntemde birimlerin az sayıdaki kümeye yerleştirilmesi iterative bir biçimde yapılır. Birimler her iterasyonda farklı kümelere atanarak en uygun çözüm permutasyonel bir yaklaşım ile belirlenir. Bu yöntemde başlangıçta seçilen küme merkezi noktaları değişmeksizin kümeleme yapılır ve diğer birimlerin kümelere yerleştirilmesi iterative yöntemle yapılır. Her atamada oluşan i. kümenin diğer kümelerin homojenitelerine göre birimlerin atandıkları kümelerden çıkarılarak başka bir kümeye atanması mümkündür. Atama işlemleri; her kümeleme aşamasında denetlenen "küme içi homojenite" ve "kümeler arası heterojenite"nin maksimizasyonunun sağlanması ile son bulur. K-Means yönteminde çekirdek noktaların gözlenen değerlerden seçilmesi zorunlu değildir. Olasılık kurallarına göre k çekirdek nokta seçilerek kümeleme yapılabilir (Özdamar 2013).

K-Ortalamalar yönteminin kullandığı algoritma aşağıdaki şekilde işlem yapar (Pektaş 2013):

a) K adet birim başlangıç küme merkezi olarak rasgele seçilir.
b) Küme merkezi olmayan birimler belirlenen uzaklık ölçütlerine göre başlangıç küme merkezlerinin ait oldukları kümelere atanır.
c) Yeni küme merkezleri oluşturulan k adet başlangıç kümesindeki değişkenlerin ortalamaları alınarak oluşturulur.
d) Birimler en yakın oldukları oluşturulan yeni küme merkezlerine birimlerin uzaklıkları hesaplanarak kümeye atanır.
e) Bir önceki küme merkezlerine olan uzaklıklar ile yeni oluşturulan küme merkezlerine olan uzaklıklar karşılaştırılır.
f) Uzaklıklar makul görülebilir oranda ise d adımına dönülür.
g) Eğer çok büyük bir değişiklik olmamış ise iterasyon sona erdirilir.

İterasyonun durdurulması için kullanılan ölçütlerden birisi, kareli hata ölçütleridir. Bu ölçüt p veri uzayında bir nokta, m_i ise C_i kümesine ait ortalama ya da küme merkezi olmak üzere şu biçimdedir:

$$E = \sum_{i=}^{k} \sum_{p \varepsilon C_i} |p - m_i|^2 \qquad (2.54)$$

■ Fuzzy (Bulanık) Kümeleme Yöntemi

Klasik küme kuramında bir eleman o kümenin ya elemanıdır ya da değildir, kısmi üyelik söz konusu olamaz. Nesnenin üyelik değeri 1 ise kümenin tam elemanı, 0 ise elemanı değildir yani klasik kümelerde elemanların üyelikleri {0,1} değerlerini alır. Bulanık mantık, insanların günlük yaşantısında nesnelere verdiği üyelik değerlerini, dolayısıyla insan davranışlarını taklit eder. Klasik kümelerdeki bu keskin durumun aksine, bulanık kümelerde elemanların üyelik dereceleri [0,1] aralığında sonsuz sayıda değerler alabilir. Keskin kümelerdeki soğuk-sıcak, hızlı-yavaş, aydınlık-karanlık gibi ikili değişkenler, bulanık mantıkta biraz soğuk, biraz sıcak, biraz aydınlık gibi esnek betimleyicilerle esnetilerek gerçek yaşam şartlarına benzetilirler (Erilli 2009).

Günümüzde birçok bulanık kümeleme yöntemi geliştirilmiş bulunmaktadır. Bunların çoğu bulanık C-Ortalamalar yönteminden köken almakta; bu yöntemin çeşitli durumlar için geliştirilmiş modifikasyonlarına dayanmaktadır. İlk olarak $m = 2$ için Dunn (1973) ve $m > 1$ için Bezdek (1973) ortaya konulan ve Bezdek (1981) tarafından iyileştirilen bulanık C-ortalamalar algoritması (FCM: Fuzzy c-means algorithm) en popüler bulanık kümeleme algoritmalarından biridir. Nitekim tarım, astronomi, kimya, jeoloji, görüntü işleme, medikal tanı, şekil analizi, hedef tanıma gibi çok sayıda alanda kullanılmakta olduğu bildirilmektedir (Cebeci ve Yıldız 2015).

Fuzzy (bulanık) kümeleme yönteminde kullanılan algoritmada bulanık kümeleme tekniği aşağıdaki amaç fonksiyonunun minimizasyonunu amaçlar. Bu amaç fonksiyonunda üyelik fonksiyonları şu kısıtlara sahiptir (Şahin ve Hamarat 2002):

1) $u_{iv} \geq 0$ ise $i = 1, \ldots\ldots\ldots\ldots, n$ ve $v = 1, \ldots\ldots\ldots\ldots\ldots, k$
2) $\sum_{v=1}^{k} u_{iv} = 1 = \%100$ ise $i = 1, \ldots\ldots\ldots\ldots, \ n$

Burada her bir i birimi ve her bir v kümesi u_{iv} nin bir üyesi olacaktır. u_{iv}, i. birimin v kümesine ne kadar ait olduğunu gösterir. Bu şartlar altında amaç fonksiyonu aşağıdaki gibidir.

$$C = \sum_{v=1}^{k} \frac{\sum_{i,j=1}^{n} u_{iv}^2 u_{jv}^2 d(ij)}{2\sum_{j=1}^{n} u_{jv}^2}. \qquad (2.55)$$

Burada, d(ij), i ve j. birimler arasındaki uzaklık (benzerlik); u_{iv}, i. birimin v. kümeye bilinmeyen üyeliğini tanımlar. Bulanık kümelemede her bir birimin tüm kümelere olan üyelik katsayıları toplamı daima bir olacak şekilde pozitiftir. Bulanık kümelemenin, kesin kümelemeden ne kadar uzakta olduğu Dunn ayrıştırma katsayısıyla değerlendirilir. Bu katsayı elde edilen kümenin ne kadar bulanık olduğuna ilişkin bir fikir verir. Dunn Ayrıştırma Katsayısı, tüm üyelik katsayılarının (u_{iv}) kareler toplamının birim sayısına bölünmesiyle (Şahin ve Hamarat 2002):

$$F_k = \sum_{i=1}^{n}\sum_{v=1}^{k} u_{iv}^2/n. \quad (2.56)$$

F(k) her zaman [1/k, 1] aralığında bulunur. Böylelikle birimlere ilişkin üyelik matrisi elde edilir. Küme sayısından bağımsız olarak bu katsayı aşağıdaki formül (Eşitlik 2.55) ile normalleştirilir (Kılıç vd. 2011):

$$F_k(u) = \frac{kF(u) - 1}{k - 1}. \quad (2.57)$$

Bulanıksızlık Endeksi (Nonfuzziness Index) olarak isimlendirilen Normalleştirilmiş Dunn Katsayısı ($F_k(u)$) 0 ile 1 aralığında bir değer alır. Ayrıca, uygun kümelemede çekirdek sayısı ve bu çekirdek noktalarına göre belirlenen kümelerin uygunluğu veya kümelerin kararlılık yapısı için gölge istatistiği (Silhoutte Coefficient ($\overline{SC}$)) kullanılır ve bu değerlerin ortalaması Ortalama Gölge İstatistiği ($\overline{SC}$) olarak bilinir. Sonuç olarak, bulanık kümeleme yönteminde Ortalama Gölge İstatistiği ($\overline{SC}$) ve Dunn ayrıştırma katsayısı (F(u)) uygun küme yapısını veren parametrelerdir ve bir veri setinde uygun kümeleme yapısı olması için $\overline{SC}$ değerinin en az 0.50 olması beklenir (Kılıç vd. 2011).

■ Medoid Kümeleme Yöntemi

K-Medoids algoritmasının temeli, verinin çeşitli yapısal özelliklerini temsil eden k tane temsilci nesneyi bulma esasına dayanır (Kaufman and Rousseeuw 1987). Temsilci nesne medoid olarak adlandırılır ve kümenin merkezine en yakın noktadır. Bir grup nesneyi k tane kümeye bölerken asıl amaç, birbirine çok benzeyen nesnelerin bir arada bulunduğu ve farklı kümelerdeki nesnelerin birbirinden benzersiz olduğu kümeleri bulmaktır. En yaygın kullanılan k-Medoids algoritması, Kaufman ve Rousseeuw (1987) tarafından geliştirilmiştir (Kaufman and Rousseeuw 1990). Temsilci nesne, diğer nesnelere olan ortalama uzaklığı minimum yapan kümenin en merkezi nesnesidir. Bu nedenle, bu bölünme metodu her bir nesne ve onun referans noktası arasındaki benzersizliklerin toplamını küçültme mantığı esas alınarak uygulanır. Kümeleme literatüründe temsilci nesnelere çoğunlukla merkez tipler (centrotypes) denilmektedir.

PAM (Partitioning Around Medoids) algoritmasında temsilci nesneler medoid olarak adlandırılmaktadır (Kaufman and Rousseeuw 1990). Amacın k tane nesneyi bulmak olmasından dolayı, k-Medoids metodu olarak adlandırılmaktadır. k adet temsilci nesne tespit edildikten sonra her bir nesne en yakın olduğu temsilciye atanarak k tane küme oluşturulur. Sonraki adımlarda her bir temsilci nesne temsilci olmayan nesne ile değiştirilerek kümelemenin kalitesi yükseltilinceye kadar ötelenir. Bu kalite nesne ile ait olduğu kümenin temsilci nesnesi arasındaki ortalama benzersizlik maliyet fonksiyonu kullanılarak değerlendirilir (Işık ve Çamurcu 2007).

Uygun kümelemede çekirdek sayısı ve bu çekirdek noktalarına göre belirlenen kümelerin uygunluğu için "gölge (siluet) istatistiği"nden yararlanılır. Gölge istatistiği (s) aşağıdaki biçimde hesaplanır (Özdamar 2013).

1. A kümesindeki n birimden i. birimin tüm diğer birimlere olan uzaklıkları ortalaması a, aşağıdaki formül ile belirlenir.

$$a = (1/n)\sum_{j=1}^{n} d_{ij};\ n\epsilon A. \qquad (2.58)$$

2. A kümesi dışında fakat i. birimin en yakın komşu olduğu ve elemanları arasındaki ortalama farklılığın en küçük olduğu B kümesindeki elemanlar ile i. birimi uzaklıklarının ortalaması b, belirlenir.

$$b = (1/n)\sum_{j=1}^{n} d_{ij};\ n\epsilon B,\ \ i\epsilon A. \qquad (2.59)$$

3. a ve b ortalama değerleri kullanılarak i. birimin gölge istatistiği s aşağıdaki kurallara göre hesaplanır.

Eğer A kümesi eleman sayısı $n = 1$ ise $s = 0$

Eğer $a < b$ ise $s = 1 - a/b$

Eğer $a > b$ ise $s = b/a - 1$

Eğer $a = b$ ise $s = 0$

Tüm birimler için gölge istatistiği hesaplanır. s istatistiği -1 ile +1 arasında değişim gösterir. Gölge istatistiği, i. birimin kendi kümesi içindeki diğer birimlerle farklılığını en yakın komşu diğer kümedeki birimlerin farklılığı ile karşılaştırmayı sağlar.

s, +1'e yakın ise, i. birim doğru sınıflandırılmıştır. Küme içi elemanları ile olan farklılığı en yakın komşu küme ile olan uzaklıklarından daha küçüktür.

s, sıfıra yakın ise, i. birim A ve B kümeleri arasındadır. A kümesine keyfi olarak atandığı varsayılır.

s, -1'e yakın ise, i. birim A kümesine yanlış olarak atanmıştır. En yakın komşu kümenin elemanı olması gerekir.

Kaufman ve Rousseeuw (1990)'a göre medoid kümelemede küme sayısını belirlemek için tüm birimlerin s değerleri ortalamasından (ortalama gölge istatistiği, $(\overline{SC})$ istatistiği) yararlanılır. $(\overline{SC})$'nin en büyük olduğu kümeleme çözümü uygun çözüm olarak alınır. k=2,3,…. için kümeleme çözümlerinden maksimum değerli $(\overline{SC})$'ye sahip çözüm uygun kümeleme çözümü kabul edilir. Ortalama gölge istatistiği $(\overline{SC})$'nin en büyük olduğu k küme sayısı çözümü uygun kümeleme çözümü olarak kabul edilir (Özdamar 2013).

2.1.7.2 Küme Sayısının Belirlenmesi

Kümeleme yöntemlerinden hiyerarşik olmayan kümeleme yöntemlerinde analize başlamadan önce küme sayısının belirlenmesi gerekir. Hiyerarşik yöntemlerde analiz öncesi böyle bir belirleme gerekmemektedir. Hiyerarşik kümeleme yöntemlerinde elde edilen dendogram yardımıyla küme sayısına görsel olarak karar verilebilmektedir. Bu nedenle verilere hiyerarşik olmayan kümeleme yöntemlerini uygulamadan önce hiyerarşik yöntemler uygulanarak muhtemel küme sayıları belirlenebilir (Çakmak 1999).

Friedman ve Rubin (1967) tarafından geliştirilen ve eşitlik (2.60) yardımıyla elde edilen k değerinin hesaplanması küme sayısının belirlenmesinde kullanılan en pratik yollardan birisidir. Bu yöntemde küme sayısı (k),

$$k = \sqrt{n/2} \tag{2.60}$$

biçiminde belirlenmektedir (Hartigan 1975). Fakat bu formülün, örneklem hacmi büyük ise genel olarak iyi sonuçlar vermediği düşünülmektedir.

Marriot (1971) küme sayısının belirlenmesinde başka bir yöntem önermiştir. Bu yöntemde W, grup içi kareler toplamı matrisi olmak üzere;

$$M = k^2|W| \tag{2.61}$$

eşitliğinden elde edilen bu kriterde en küçük M değerini veren küme sayısı (k), uygun küme sayısı olarak ele alınmaktadır (Kılıç 2008).

Küme sayısının belirlenmesinde temel bileşenler analizinden de sıklıkla yararlanılmaktadır. Özellikle toplam değişkenliğin ilk iki temel bileşenle açıklanabildiği durumlarda, ilk iki temel bileşen yardımıyla gözlemler için elde edilen temel bileşen skorlarının saçılım grafiği, kümeleme sonuçlarının doğru yorumu için bir rehber olabilmektedir (Alpar 2011).

2.1.8 Çok Boyutlu Ölçekleme

Çok Boyutlu Ölçekleme (ÇBÖ) analizi günümüzde oldukça yaygın bir kullanıma sahip olup, temeli Young ve Household tarafından ÇBÖ'ye öncülük eden çalışmalarına ve yine aynı yılda Richardson'un bu analizle ilgili uygulamalarına dayanmaktadır. Fakat ÇBÖ analizinde gerçek anlamda ilk çalışmalar 1950'li yıllarda başlamıştır. Öklit uzaklık modeline dayanan bu çalışmalar, Young ve Household tarafından ispatlanmış olan matematiksel temelli teoreme dayanmaktadır (Yiğit 2007).

Tarihsel olarak, ÇBÖ'nün gelişiminde iki önemli aşama olmuştur. Bunlardan birincisi, klasik veya metrik yaklaşımdır. Bu yaklaşım aynı zamanda 'Princeton' ya da daha çok 'Torgerson' yaklaşımı olarak bilinir. Klasik ÇBÖ'nün ilk temelleri Princeton üniversitesinde Messick ve Abelson (1956) ve Torgerson (1952)'nin içinde bulunduğu psychometrik grubu tarafından yapılmıştır. Ayrıca Torgerson (1952) ilk defa ÇBÖ analizinin uygulanılabilir olduğunu göstermiştir. İkinci önemli aşama ise on yıl sonra Beil Telephone laboratuvarında Shepard (1962) tarafından ÇBÖ analizinin metrik olmayan yaklaşımının bulunmasıyla olmuştur. Aynı zamanda bu 'Shepard-Kruskal' yaklaşımı olarak da bilinmektedir. Shepard (1962)'ın arkadaşı Kruskal (1964) tarafından bu yaklaşıma kavramsal ve hesaplanabilen yenilikler getirilmiştir. Verilerin farklı türleriyle ilgilenmek amacıyla yalnızca Shepard ve Kruskal değil aynı zamanda J.-J. Chang, S.C. Johnson, E.T. gibi kişiler bu yaklaşımlardan yola çıkarak yeni yaklaşımlar oluşturmuşlardır. C.H. Coombs ve Michigon Üniversitesindeki çalışma arkadaşları tarafından metrik olmayan yada ordinal verilerin çok boyutlu gösterimi için yeni düzenlemeler yaptılar. Ancak; bu özel metot gerçek verilerin çok boyutlu analizi için yaygın biçimde kullanılmamaktadır. Bunun ilk nedeni az sayıdaki nesneler hariç bilgisayar programları için dönüştürülebilecek yeterli formülün olmamasıdır. Diğer nedense, metrik olmayan verilerle çok önemli metrik bilgileri elde etmenin yetersiz olmasıdır (Yiğit 2007).

Çok Boyutlu Ölçekleme – ÇBÖ (Multidimensional Scaling- MDS) n tane nesne (birey - gözlem) arasındaki uzaklık değerlerini kullanarak bu nesnelerin çok boyutlu uzaydaki konumlarının, ilişki yapısını veren resmini ortaya koymayı amaçlamaktadır. Çok değişkenli analiz, n*p boyutlu veri matrisi ile ilgilenmektedir. Bazı analizlerde (kümeleme analizi başta olmak üzere) X veri matrisi yerine n tane bireyin uzaklıklarından elde edilen n*n boyutlu D uzaklıklar matrisi kullanılmakta idi. ÇBÖ'de de yine uzaklıklar matrisi göz önüne alınmakta ve bu matrisin simetrikliği nedeniyle işlemler, **1/2** n (n-l) tane uzaklık değerleri kullanılarak sürdürülmektedir. ÇBÖ'nün genelde amacı, mümkün olduğunca az boyutla, nesnelerin yapısı (uzaklık değerlerini kullanarak) orijinal şekle yakın bir biçimde ortaya koymaktır (Tatlıdil 1992).

Farklılıkların yanında benzerliklerin ortaya konulmasında da yararlanılan bir yöntem olan ÇBÖ analizi; tıp, psikoloji, sosyal bilimler, gibi birçok alanda kullanılan bir yöntemdir. Özetle ÇBÖ k boyutlu bir uzayda gösterilebilen nesneleri orijinal konumlarına çok yakın bir biçimde daha az boyutlu kavramsal bir uzayda göstererek, nesneler arası ilişkileri belirlemeye yardımcı olur. Analizin genel amacı, mümkün olduğunca az boyutla, nesnelerin yapısını (uzaklık değerlerini kullanarak) orijinal şekle yakın bir biçimde ortaya koymaktır. Bu teknik vasıtasıyla çok boyutlu veri matrisinde nesne veya bireyler arasındaki karmaşık ilişkilerin daha kolay anlaşılabilir ve açıklanabilir boyutlara indirgenebilmesi sağlanabilmektedir (Kalaycı 2018).

2.1.8.1 Çok Boyutlu Ölçeklemede Varsayımlar ve Yaklaşımlar

ÇBÖ yöntemi, birçok yöntemi içine alan bir yöntemler ailesidir. Ancak temel uygulama adımları klasik ÇBÖ yönteminde uygulanan adımlara benzerlik gösterir. Bu adımlar altı adımda özetlenebilir (Kalaycı 2018):

1. Veri tipine göre transformasyon yöntemlerinden uygun olan birinin seçilmesi ve bu seçime bağlı olarak verilerin dönüştürülerek elde edilmesidir.

2. Veri tipine bağlı olarak uygun uzaklık matrisinin hesaplanması

3. P değişkenli p boyutlu veri matrisine sahip olan n nesne ya da birimin kaç boyutlu bir uzayda gösterilebileceğine karar verilir. Uygulamada genellikle 2, 3, 4 gibi boyutlar seçilir ve bu boyutların her biri için ÇBÖ çözümleri elde edilir. Ayrıca her bir k için elde edilen çözümlerin orijinal uzaklık matrisine uygunluğu (stress ölçüsü) hesaplanır ve uygun çözümün hangi boyutta gerçekleştiğine ve hangi çözümün uygulanacağına karar verilir.

4. Veri uzaklıklarına göre konfigürasyon uzaklıkları d_{ij}'nin regresyonu verinin tipine göre hesaplanır. Regresyon yöntemlerinden veri tipine göre uygun olan biri seçilir. Belirlenen regresyon denklemi aracılığıyla tahmini konfigürasyon uzaklıkları belirlenir. Bu tahmini uzaklıklara fark adı verilir. Bu uzaklıklardan elde edilen matrise de farklılık matrisi denilmektedir.

5. Konfigürasyon uzaklıkları ile tahmini uzaklıklar arasındaki uygunluğu belirlemek için uygun bir istatistik olan stress istatistiği hesaplanır.

6. K boyutuna göre birim ya da nesnelerin koordinatları elde edilir. Bu koordinatlar k boyutlu uzayda gösterilerek her birim ya da nesnenin diğer birim ya da birimlere göre konumları görüntülenir. Bu görüntüler yorumlanarak birimler arası ilişkiler belirlenmeye çalışılır.

Çok Boyutlu Ölçekleme, verilerin tipine bağlı olarak n nesne/birim arasında hesaplanan yakınlık/benzerlik/uzaklık/farklılık matrisine göre iki şekilde uygulanır. İlki metrik çok boyutlu ölçekleme, ikincisi metrik olmayan çok boyutlu ölçekleme şeklindedir. ÇBÖ yaklaşımının belirlenmesinde veri tipinin önemi büyüktür. Veri matrislerinin içeriği veri tiplerine göre nesnelerin/birimlerin benzerlik/uzaklık matrisi ve uygulanacak ÇBÖ aşağıdaki gibi belirlenir (Özdamar 2013):

✓ Kategorik verilerden (isimsel, sıralı, skor) hesaplanan uzaklık/benzerlik matrisine metrik olmayan ÇBÖ uygulanır.

✓ Nicel verilerden hesaplanan uzaklık/benzerlik matrislerine metrik ÇBÖ uygulanır.

✓ Karma ölçekli veriler içeren veri matrislerinin değerleri uygun skor değerlere dönüştürülür. Uzaklık/benzerlik matrisi metrik değerler olarak hesaplanır ve metrik ÇBÖ uygulanır.

2.1.8.2 Metrik Çok Boyutlu Ölçekleme

Metrik ÇBÖ'de öncelikle $n * n$ boyutunda $D = (\delta_{ij})$ uzaklık matrisi dikkate alınır. Orijinal uzaklık matrisi D'deki (δ_{ij}) değerlerine yaklaşık olarak eşit olan, k $(k < p)$ boyutta noktalar arası uzaklık değerlerinin (koordinatların) bulunması ana amaçtır. Nesneleri temsil edecek koordinatların bulunması ise uzaklıklar matrisinin koordinatlarının elde edileceği B matrisinin öz vektörlerinin elde edilmesi sürecidir. Dolayısıyla, metrik çok boyutlu ölçeklemede ilk hedef, uzaklıklar matrisinin koordinatlarının elde edileceği B matrisinin bulunmasıdır. Bu amaçla (Alpar 2011):

1) Değişkenlerin birimleri farklı ise uzaklık ölçülerinde eşit etkiye sahip olmaları için standartlaştırma yoluna gidilir.

2) Veri tipine uygun olarak seçilen uzaklık ölçüleri kullanılarak $n * n$ boyutlu uzaklık matrisi $D = (\delta_{ij})$ elde edilir.

3) Asıl köşegen elemanları sıfır olduğu için D matrisi pozitif yarı tanımlı bir matris değildir. Bununla birlikte, D matrisinin elemanlarına dayalı olarak pozitif tanımlı B_{n*n} matrisi de elde edilebilmektedir. Bu çerçevede önce A matrisi elde edilir.

4) $n * n$ boyutlu A matrisi, δ_{ij}; D matrisinin elemanlarını göstermek üzere

$$A = (a_{ij}) = \left(-\frac{1}{2}\delta_{ij}^2\right) \tag{2.62}$$

yardımıyla bulunur.

B matrisi elde edilir. Bu amaçla,

$$\bar{a}_{i.} = \sum_{j=1}^{n} a_{ij}/n, \tag{2.63}$$

$$\bar{a}_{j.} = \sum_{i=1}^{n} a_{ij}/n, \tag{2.64}$$

$$\bar{a}_{..} = \sum_{ij} a_{ij}/n^2, \tag{2.65}$$

olmak üzere, A matrisinin her bir elemanından, ait olduğu satır ve sütun ortalamaları çıkartılıp A'nın genel ortalaması eklenerek eşitlik (2.66)'ya ulaşılır.

$$B = (b_{ij}) = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..} \tag{2.66}$$

B matrisi yukarıda ki süreç yerine eşitlik (2.67) yardımı ile elde edilebilir.

$$B = -\frac{1}{2}\left[I_n - \frac{1}{n} i_n i_n'\right] D^2 \left[I_n - \frac{1}{n} i_n i_n'\right] \quad (2.67)$$

Burada, I_n: $n * n$ boyutunda birim matris, i_n: $n * 1$ boyutunda birim vektör ve D^2; D matrisinin elemanlarının karelerinin alınması ile elde edilen matristir.

5) B matrisinin öz değerleri ve öz vektörleri bulunur. Eğer B matrisi pozitif yarı tanımlı q ranklı ise B'nin q tane öz değeri pozitif, $n - q$ tane öz değeri ise sıfır olacaktır. Buradan, B matrisi $B = V\Lambda V'$ ile ifade edilebilir. Burada V: B' nin öz vektörleri matrisi, Λ: ise köşegen elemanları öz değerler olan matristir.

6) B matrisin öz değerleri ve öz vektörleri elde edildikten sonra $\sqrt{\lambda_i} v_i$ yardımıyla grafiksel gösterimi için koordinatlar bulunur.

7) Gerektiğinde gerçek değerler ile kestirilen değerler arasındaki farklılığa dayalı bir ölçü olan stres değeri elde edilebilir. Her bir boyuta karşı gelen stres değerlerinin çizimi sonucunda elde edilecek yamaç grafiği yardımıyla çalışmadaki uygun boyut sayısı da belirlenebilmektedir.

$$\text{Stres} = \sqrt{\frac{\sum (d_{ij} - \delta_{ij})^2}{\sum d_{ij}^2}}, \quad (2.68)$$

$$\text{Stres} = \sqrt{\frac{\sum (d_{ij} - \delta_{ij})^2}{\sum \delta_{ij}^2}}. \quad (2.69)$$

d_{ij}: gösterimdeki uzaklıklar (analiz sonucu kestirilen uzaklıklar)

δ_{ij}: orijinal uzaklıklar

şeklinde ifade edilmektedir.

Uygun boyut sayısı belirlendikten sonra orijinal uzaklıklar ile gösterim için kestirilen uzaklıkların saçılım grafiği çizilerek uyumun doğrusallığı incelenebilir. Eğer tam uyum var ise 45 derecelik açıya sahip bir doğru görünümü elde edilecektir. Kestirilen uzaklıkların orijinal uzaklıklara ne kadar uyum sağladığı, bir diğer deyişle doğrusal ilişkinin kuvveti R^2 ile ölçülebilir. Metrik çok boyutlu konumlar orjin etrafında döndürülebilir, sabit eklenerek orjin değiştirilebilir, eksenler çevrilebilir veya tüm sonuç tekdüze olarak yayılabilir ya da sıkıştırılabilir ve bunların tümü nesnelerin göreli konumları değiştirilmeden yapılır (Alpar 2011).

2.1.8.3 Metrik Olmayan Çok Boyutlu Ölçekleme

Metrik olmayan ölçekleme yöntemlerinde δ_{ij} değerlerini belirlemede kullanılan tek bilgi, d_{ij} uzaklık değerlerinin sıra sayılarıdır. Zaten metrik olmayan yaklaşımda D uzaklıklar matrisi değil, farklılık ölçümleri (benzemezlik) matrisi olarak düşünülmektedir. Bu nedenle verilere ilişkin benzerlik matrisinin bilinmesi durumunda metrik olmayan yöntemler gündeme gelmektedir. Sıralanmış d_{ij} bilgilerinden yararlanılarak ölçeklemeye, şekil oluşturmaya ilişkin geliştirilen ilk algoritma Shepard–Kruskal algoritmasıdır. Metrik olmayan ölçeklemenin temeli sayılabilecek bu algoritmanın adımları aşağıdaki gibidir (Tatlıdil 1992).

Adım 1: D benzemezlik matrisinin (köşegen elemanları hariç) tüm elemanlarının sıralanması,

$$dr_1 s_1 < \cdots\cdots\cdots\cdots < dr_m s_m; m = \frac{1}{2}\ n(n-1) \tag{2.70}$$

ve d_{rs}'lerle monotonik olarak ilişkilendirilen d^*_{rs} değerlerinin tanımlanması. Bu ilişkilendirme aşağıdaki gibi bir koşulu sağlamalıdır.

$$d_{rs} < d_{uv} \rightarrow\ d^*_{rs} \leq d^*_{uv}; \text{tüm}\quad r < s, \qquad u < v \tag{2.71}$$

Adım 2: Çok boyutlu uzaydaki gerçek şekil ile indirgenmiş boyutlu (k-boyut) uzayda kestirilen şekil arasındaki farklılığın ifadesi olan stress değerinin hesaplanması. $\widehat{X}^{1}_{n*k}$, R^{k} uzayında $\hat{d}_{rs}$ uzaklık değerleri ile ifade edilen bir şekil ise, $\widehat{X}$' nın (karesel) stress'i aşağıdaki gibi tanımlanır.

$$S^2(\widehat{X}) = \min \sum_{r<s} \left(d^*_{rs} - \hat{d}_{rs}\right)^2 / \sum \hat{d}^2_{rs}. \qquad (2.72)$$

Burada d^*_{rs}'lar $S^2(\widehat{X})$'yi minimize eder ve $\hat{d}_{rs}$'ların sıralanmalarının d_{rs}'nin sıralanmaları ile iyi uyuşup uyuşmadığını gösterir. Eğer sıralamalar tam olarak birbirleriyle uyuşuyor ise $S^2(\widehat{X}) = 0$ değeri alır ki böylesi bir sonuç hedeflenen teorik bir sonuçtur. (2.72) nolu eşitlikte payda değeri stress katsayısını standartlaştırdığı için bu katsayı; $r = 1, \ldots\ldots\ldots, n$ ve $c \neq 0$ olmak üzere $y_r = cx_r$ türü ya da A dik bir matris, b katsayılar vektörü olmak üzere $y_r = Ax_r + b$ türü dönüşümlerde değişmezdir.

Adım 3: Her k boyut için en küçük stress değerli şekle, k boyuta uyan en iyi şekil adı verilir ve

$$S_k = \min S(\widehat{X}) \qquad (2.73)$$

biçiminde gösterilir. Burada S_k, k'nın azalan bir fonksiyonudur.

Adım 4: Gerçek (doğru) boyutu belirlemek amacıyla, $S1, S2, \ldots\ldots, Sk$ değerleri hesaplanmakta ve bu işlemlere küçük stress değeri elde edilince son verilmektedir.
Elde edilen şeklin gerçek şekille uygunluğunun bir ölçüsü olarak Kruskal tarafından geliştirilmiş tolerans oranlarından yararlanılmaktadır.

ÇBÖ analizinin çıktılarının yorumlanmasında stres değeri karar verme kriteri olarak kullanılmaktadır. Düşük bir stres değeri kararın uygunluğunu yansıttığı gibi, yüksek bir stres değeri kötü bir uyuma işaret etmektedir. Çizelge 2.4'te stress değerlerinin sınıflandırılmasına ilişkin Kruskal (1964) stres değerinin yorumlanması için çözümün uygunluğunu yansıtan bir rehber sunmuştur (Burmaoğlu 2011).

Çizelge 2.4 Stres değerlerinin sınıflandırılması

Stress(s)	**Uygunluk**
>0.20	Yetersiz
0.10<0.20	Vasat
0.05<0.10	İyi
0.025<0.05	Çok İyi
0.00<0.025	Mükemmel

Metrik olmayan ÇBÖ'de iki farklı stres ölçüsü:

$$\text{Stres1} = \sqrt{\frac{\sum\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum d_{ij}^2}}, \tag{2.74}$$

$$\text{Stres2} = \sqrt{\frac{\sum\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum\left(d_{ij} - \bar{d}\right)^2}}. \tag{2.75}$$

d_{ij}: orjinal uzaklıklar

$\hat{d}_{ij}$: benzerlik verisinden türetilen uzaklıklar

$\bar{d}$: haritadaki ortalama uzaklık $\left(\sum d_{ij}/n\right)$

Eğer uyum ölçüsü daha önceden belirlenerek seçilmiş bir durma değeri ile karşılaşmaz ise, uyun ölçüsünün daha fazla küçülmesi için yeni bir biçim bulunur. Kestirilen d_{ij}, orijinal d_{ij}'ye yaklaştıkça stres değeri küçülür. Nesneler biçim içine orijinal uzaklıklar ile en iyi örtüşecek şekilde konumlandırıldığında stres en aza indirgenir. Doyurucu bir stres değerine ulaşıldığında, boyutsallıklar azalır ve kabul edilebilir bir uyum ölçüsünde en düşük boyutsallığa ulaşana kadar işlem tekrarlanır (Alpar 2011).

3. MATERYAL ve METOT

Bu çalışmada Türkiye'deki illerin hayvancılık istatistikleri bakımından çok değişkenli istatistik analiz teknikleri ile sınıflandırılması amaçlanmıştır. Bu amaç dâhilinde Türkiye İstatistik Kurumu'ndan (TÜİK) 2018 yılına ait illerin hayvancılık istatistiklerinden elde edilen toplam 29 değişken kullanılmıştır. Söz konusu 29 değişken hayvan varlığı ve hayvansal üretim olarak iki ana faktör veya boyutta Çizelge 3.1'de görülmektedir.

Çalışmada, TÜİK tarafından yayınlanan güncel istatistikler kullanılmıştır. Araştırmada kullanılan değişkenler Türkiye'deki hayvan varlığının ve hayvansal üretimin kişi başına yıllık üretilen değerleri hesaplanarak kullanılmıştır. Kişi başına yıllık üretilen hayvan varlığı ve hayvansal üretimin hesaplanmasında, 2018 yılına ait TÜİK adrese dayalı nüfus sayım istatistikleri kullanılmıştır. Kişi başına yıllık üretilen hayvan varlığı ve hayvansal üretim, nüfus verilerine bölünerek elde edilmiştir.

Kişi başına yıllık üretilen hayvansal üretim ve hayvan varlığı istatistiklerinin, hem adet hem de kg gibi farklı ölçü birimlerinden oluşması ve rakamsal büyüklüklerinin de çok büyük farklılık göstermesinden dolayı standartlaştırılmış veriler kullanılmıştır. Veri dönüşümü veya standartlaştırma, istatistik analizlerde normalleştirme, durağanlaştırma ve doğrusallaştırma gibi üç temel amaçta yapılmaktadır (Kalaycı 2018). Bazen değişkenlerin aşırı uçlarda yer alan değerleri kümeleme üzerinde olumsuz etkilerde bulunmaktadır. Bu gibi durumlarda verilerin standardize ya da belirli aralıklarda gözlenen değerlere dönüştürülmesi uygun olmaktadır (Özdamar 2013). Verilerin dönüştürülmesinde birçok yöntem mevcuttur. Bunlardan bazıları, $-1 \leq x \leq +1$ aralığına dönüştürme, $0 \leq x \leq +1$ aralığına dönüştürme, z skorlarına dönüştürme gibi tekniklerdir. Bu çalışmada, kullanılan standartlaştırma yöntemi, $z_i = \frac{X_i - \bar{x}}{S}$ formülü ile hesaplanan z skorlarına dönüştürme tekniğidir.

Çizelge 3.1 Araştırmada kullanılan değişkenler

Değişkenler		Ölçü Birimi	Kodu
No	**Hayvan Varlığı**		
1	Kültür Sığır Sayısı	Adet	X1
2	Sağılan Kültür Sığır Sayısı	Adet	X2
3	Melez Sığır Sayısı	Adet	X3
4	Sağılan Melez Sığır Sayısı	Adet	X4
5	Yerli Sığır Sayısı	Adet	X5
6	Sağılan Yerli Sığır Sayısı	Adet	X6
7	Manda Sayısı	Adet	X7
8	Sağılan Manda Sayısı	Adet	X8
9	Merinos Koyunu Sayısı	Adet	X9
10	Sağılan Merinos Koyunu Sayısı	Adet	X10
11	Yerli ve Diğer Koyun Sayısı	Adet	X11
12	Sağılan Yerli ve Diğer Koyun Sayısı	Adet	X12
13	Keçi (Tiftik ve Kıl) Sayısı	Adet	X13
14	Sağılan (Tiftik ve Kıl) Keçisi Sayısı	Adet	X14
15	Et Tavuğu Sayısı	Adet	X15
16	Yumurta Tavuğu Sayısı	Adet	X16
17	Diğer Kümes Hayvanların Sayısı	Adet	X17
18	Arı Kovanı Sayısı	Adet	X18
No	**Hayvansal Üretim**		
19	Kültür Sığır Sütü	kg	X19
20	Melez Sığır Sütü	kg	X20
21	Yerli Sığır Sütü	kg	X21
22	Manda Sütü	kg	X22
23	Merinos Koyun Sütü	kg	X23
24	Yerli ve Diğer Koyun Sütü	kg	X24
25	Keçi (Tiftik ve Kıl) Sütü	kg	X25
26	Merinos Koyun Yapağı	kg	X26
27	Yerli ve Diğer Koyun Yapağısı	kg	X27
28	Tiftik ve Keçi Kılı	kg	X28
29	Bal ve Balmumu	kg	X29

Araştırmada kullanılan veriler, çok değişkenli istatistik yöntemlere ve kullanılan analiz tekniklerine (kümeleme analizi gibi) ilişkin çoklu normal dağılım, doğrusallık, çoklu doğrusal bağıntı, otokorelasyonun olmaması gibi varsayımların gerçekleştiği görülmüş ve bu yöntemler araştırmanın amacına uygun olarak kullanılmıştır. Verilerin çok değişkenli normal dağılıma sahip olup olmadığını test etmek için birçok yöntem mevcuttur. Henze-Zirkler testi, Royston testi, Mardia testi, Doornik-Hansen testi bunlardan bazılarıdır. Bu çalışmada çoklu normallik varsayımı Mardia testi ile test edilmiş olup sonuçlar Çizelge 3.2'de verilmiştir.

Çizelge 3.2 Mardia testi sonuçları

	Katsayı	**İstatistik**	**SD**	**P**
Çarpıklık	855,3	1,15	4495	0,34
Basıklık	1281	40,49	4495	0,12

p: Mardia testi olasılık değeri

Çizelge 3.2'deki Mardia basıklık ve çarpıklık değerlerinin her ikisinde de çoklu normalliğin karşılanması gerekmektedir. Bu durumda hem çarpıklık değeri hem de basıklık değerine göre çoklu normal dağılım gerçekleşmektedir ($p>0,05$).

Çok değişkenli varsayımların gerçekleşmesiyle beraber illerin nasıl bir dağılım gösterdiğini ortaya koyabilmek için kümeleme analizi uygulanmış ayrıca hiyerarşik ve hiyerarşik olmayan kümeleme yöntemlerini karşılaştırmak amacıyla hiyerarşik kümeleme yöntemlerinden TekBKY, TamBKY, McQuitty ve Ward kümeleme yöntemi, hiyerarşik olmayan kümeleme yöntemlerinden K-Ortalamalar kümeleme yöntemi ve fuzzy kümeleme yöntemi kullanılmıştır. Kullanılan bu kümeleme yöntemlerinde uzaklık ölçüleri olarak TekBKY yönteminde Öklid uzaklık ölçüsü, TamBKY ve Ward yöntemlerinde Karesel Öklid, McQuitty yönteminde ise Manhattan uzaklık ölçüleri kullanılmıştır. Ayrıca fuzzy kümeleme analizinde ise Öklid uzaklık ölçüsü uygulanmıştır.

Hiyerarşik olmayan kümeleme yöntemlerinde, hiyerarşik kümeleme yöntemlerine ilişkin kullanılan farklı teknikler ile elde edilen küme sayısı kullanılabilir. Ancak, hiyerarşik olmayan kümeleme yöntemlerinde küme sayısının belirlenmesi için en çok tercih edilen yöntem; $k = 2, 3, 4, \ldots$ biçiminde ardışık olarak küme sayısını bir artırarak oluşan kümelemede birimlerin hangi kümeye ait olduğuna ilişkin küme kodlarını (küme üyelik skorlarını) kullanarak yeni veri yapısına diskriminant (ayırma) analizi uygulamak ve Wilk's Lamda değerini bulmaktır. Burada, en yüksek önemliliğe sahip Wilk's Lamda değerine ait küme sayısı uygun parçalama olarak kabul edilmektedir (Özdamar 2013). Bununla birlikte diskriminant analizi sonucunda doğru sınıflandırma oranları elde edilerek de küme sayıları belirlenebilmektedir (Kılıç 2008). Bu çalışmada hiyerarşik ve hiyerarşik olmayan kümeleme yöntemlerinin uygulanması sonucunda küme sayısını belirlenmesi için küme üyelik kodları kullanılarak diskriminant analizi uygulanmıştır.

Diskriminant analizi, X veri setindeki değişkenlerin iki veya daha fazla gerçek gruplara ayrılmasını sağlayan, birimlerin p tane özelliğini ele alarak doğal ortamdaki gerçek gruplarına optimal düzeyde atanmalarını sağlayacak fonksiyonlar türeten bir yöntemdir. Diskriminant analizi grupların kovaryans matrislerinin eşit olup olmamasına göre farklı biçimlerde uygulanmakta olup, doğrusal (linear) ve karesel (quadratic) diskriminant analizi olarak iki grupta incelenmektedir. Doğrusal diskriminant analizi, tüm grupların kovaryans matrislerinin benzer olduğunu varsayar. Karesel ayırma analizi ise tüm grupların kovaryans matrislerinin benzer olduğu varsayımını kullanmaz. Doğrusal ve karesel diskriminant analizlerinin temel amacı, açıklayıcı değişkenlere ilişkin belirlenmiş ayırma fonksiyonlarına göre gözlemleri iki veya daha fazla gruba ayırmayı ve yeni gözlemleri bu gruplara atamayı sağlamaktır (Özdamar 2013).

Çalışmada, Türkiye'deki illerin hiyerarşik ve hiyerarşik olmayan kümeleme yöntemleri ile sınıflandırılması sonucunda uygun küme yapısı ve sayısı elde edildikten sonra, söz konusu kümelerin hayvansal üretim ve hayvan varlığına ait 29 değişken bakımından karşılaştırılması, parametrik test varsayımları gerçekleştiği için (normal dağılım, varyansların homojenliği, v.b.) parametrik testlerden bağımsız örneklemler için t testi (independent samples t test) ve tek faktörlü varyans analiziyle (one way ANOVA) yapılmıştır. Varyans analizi sonucunda kümeler arasındaki farklılıklar çoklu karşılaştırma testlerinden Tukey testi ile gerçekleştirilmiştir.

Araştırmada ayrıca Türkiye'de illerin hayvancılık göstergeleri bakımından sıralanması için bir hayvancılık gelişmişlik endeksi oluşturmak amacıyla temel bileşenler yöntemi kullanılarak faktör analizi uygulanmıştır. Değişkenlere ilişkin verilerin faktör yükleri ile çarpılarak toplanması ile elde edilen her il için endeks değerleri illere göre sıralamaya konulmuş ve hayvancılık göstergeleri bakımından en gelişmiş durumdaki iller tespit edilmiştir. Burada, verilere ilişkin bir karşılaştırma yapmak için standartlaştırılmış verilerin yanı sıra kişi başına yıllık üretilen ham veriler de kullanılmıştır.

Çalışmada, illerin hayvancılık istatistiklerine göre iki boyutlu bir uzayda konumlarının grafiksel olarak belirlenmesi ve daha geçerli bir değerlendirme elde edilmesi amacıyla Türkiye'deki illerin hayvancılık istatistiklerini oluşturan hayvan varlığı ve hayvansal üretimden oluşan toplam 29 değişkene çok boyutlu ölçekleme analizi uygulanmıştır. ÇBÖ (çok boyutlu ölçekleme) teknikleri metrik ve metrik olmayan ÇBÖ şeklinde iki yöntemden oluşmaktadır. Hangi yöntemin seçileceğini belirlemede veri tipine bakılmaktadır. Bu çalışmada veri yapısından dolayı metrik ÇBÖ teknikleri uygulanmıştır. Ayrıca ÇBÖ analizinde uzaklık ölçüsü olarak Karesel Öklid uzaklık ölçüsü kullanılmıştır. Bu araştırmada verilerin analizi için farklı istatistik paket programlarından yararlanılmıştır.

4. BULGULAR

Türkiye'deki illerin hayvancılık istatistikleri bakımından çok değişkenli istatistik analiz teknikleri ile sınıflandırılması amaçlanan bu çalışmadaki bulgular üç temel aşamada gerçekleştirilmiştir. Bunlar aşağıda kümeleme yöntemine ilişkin bulgular, faktör analizine ilişkin bulgular ve çok boyutlu ölçekleme analizine ilişkin bulgular seklinde aşağıda sunulmuştur.

4.1 Kümeleme Yöntemine İlişkin Bulgular

4.1.1 Hiyerarşik Kümeleme Analizine İlişkin Bulgular

■ Tek Bağlantı Kümeleme Yöntemi (TekBKY)

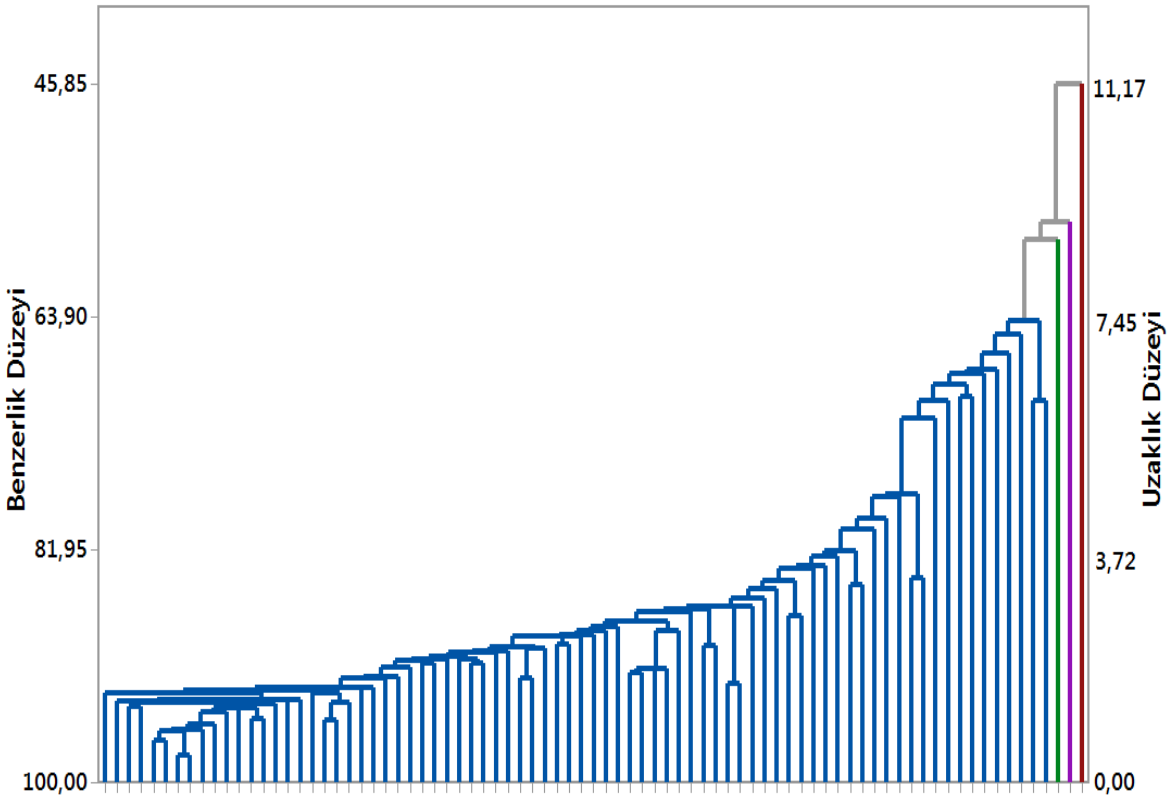
Türkiye'deki 81 il, hayvancılık istatistikleri bakımından TekBKY ile sınıflandırılmış ve farklı küme sayıları (k = 2, 3, 4, …) için uzaklık, benzerlik değerleri ve diskriminant analizi Doğru Sınıflandırma Oranları (DSO) Çizelge 4.1'de sunulmuştur. Ayrıca analize ilişkin dendrogram Şekil 4.1'de verilmiştir.

Çizelge 4.1 İllerin TekBKY ile kümelenmesi

Küme Sayısı	Uzaklık Düzeyi	Benzerlik Düzeyi	DSO(%)
2	8,9	56,6	100
3	8,7	57,9	100
4	7,4	64,2	100
5	7,2	65,3	92,6
6	6,9	66,7	54,3

Şekil 4.1'deki dendogram ve Çizelge 4.1'deki TekBKY sonuçları incelendiğinde; illerin hayvancılık istatistikleri bakımından 56,6'lık benzerlik düzeyi ve 8,9 uzaklık düzeyinde 2 kümede toplandığı görülmektedir. Benzerlik düzeyleri 56,6'dan 57,9'a çıktığında küme sayısı 2'den 3'e, 64,2'e çıktığında ise küme sayısı 4'e yükselmiştir. Bu bulgular,

beklendiği gibi birimler arasındaki uzaklık düzeyi azaldıkça ve/veya benzerlik düzeyi arttıkça hayvancılık istatistikleri bakımından illerin küme sayısının da arttığını göstermektedir.



Şekil 4.1 İllerin TekBKY ile kümelenmesine ilişkin dendrogram

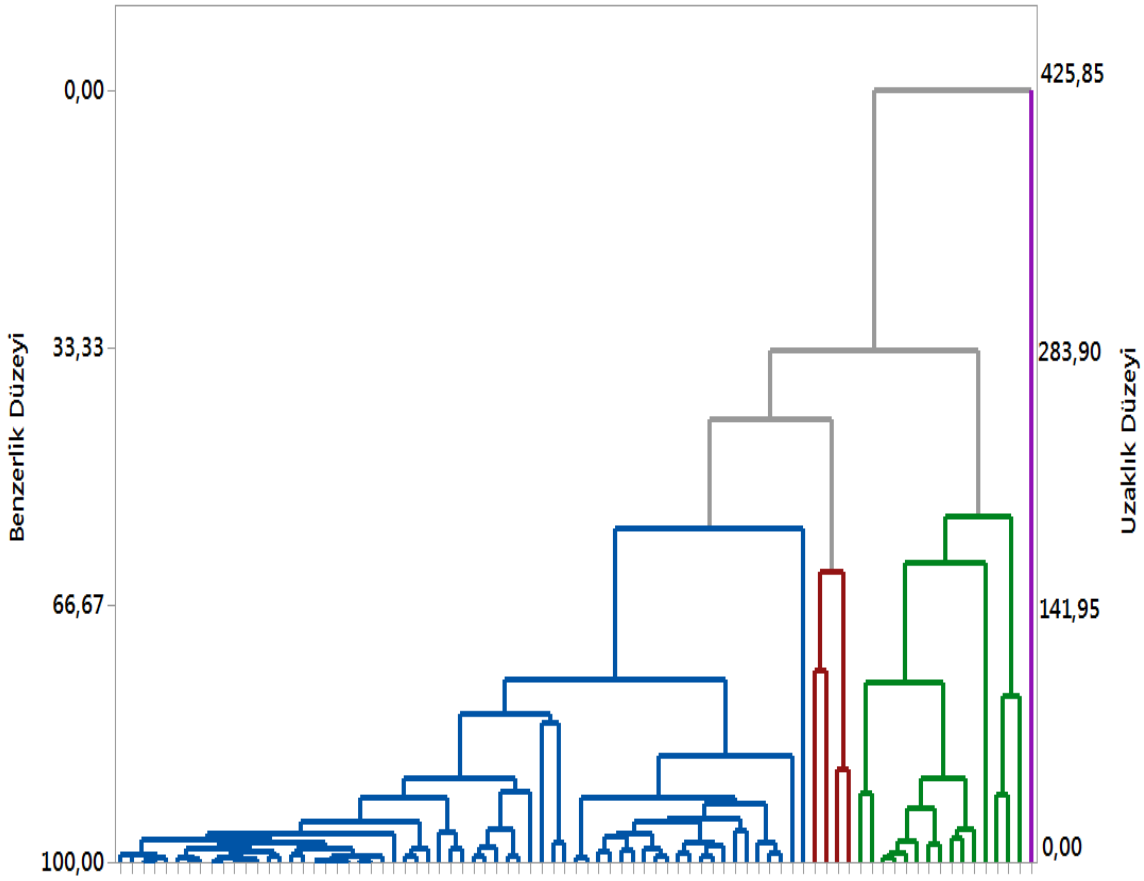
Şekil 4.1'deki dendrogram incelendiğinde, iki küme için Ardahan ili (en sağda) tek başına bir küme oluştururken diğer iller de ayrı bir kümede sınıflanarak diğer kümeyi oluşturmaktadır.

■ Tam Bağlantı Kümeleme Yöntemi (TamBKY)

Türkiye'deki illerin hayvancılık istatistikleri bakımından TamBKY ile sınıflandırılmasına ilişkin bulgular Çizelge 4.2'de ve elde edilen dendrogram Şekil 4.2'de sunulmuştur.

Çizelge 4.2 İllerin TamBKY ile kümelenmesi

Küme Sayısı	Uzaklık Düzeyi	Benzerlik Düzeyi	DSO(%)
2	283,90	33,33	100
3	244,14	42,67	95,1
4	190,82	55,19	95,1
5	184,08	56,77	95,1
6	165,13	61,23	93,8



Şekil 4.2 İllerin TamBKY ile kümelenmesine ilişkin dendrogram

Şekil 4.2'deki dendogram ve Çizelge 4.2'deki TamBKY sonuçları incelendiğinde; illerin hayvancılık istatistikleri bakımından 33,33'lük benzerlik düzeyi ve 283,9 uzaklık düzeyinde 2 kümede toplandığı görülmektedir. Benzerlik düzeyinin 42,67'ye yükselmesi ile 3 küme oluşmaktadır. Benzerlik düzeyleri 42,67'den 55,19'a çıktığında

küme sayısı 3'den 4'e, 56,77'ye çıktığında ise küme sayısı 5'e yükselmiştir. Ayrıca Şekil 4.2'de illerin TamBKY ile kümelenmesine ilişkin dendrogram detaylı bir şekilde incelendiğinde, iki küme için Ardahan ili (en sağda) tek başına bir küme oluştururken diğer iller de ayrı bir kümede sınıflanarak diğer kümeyi oluşturmaktadır.

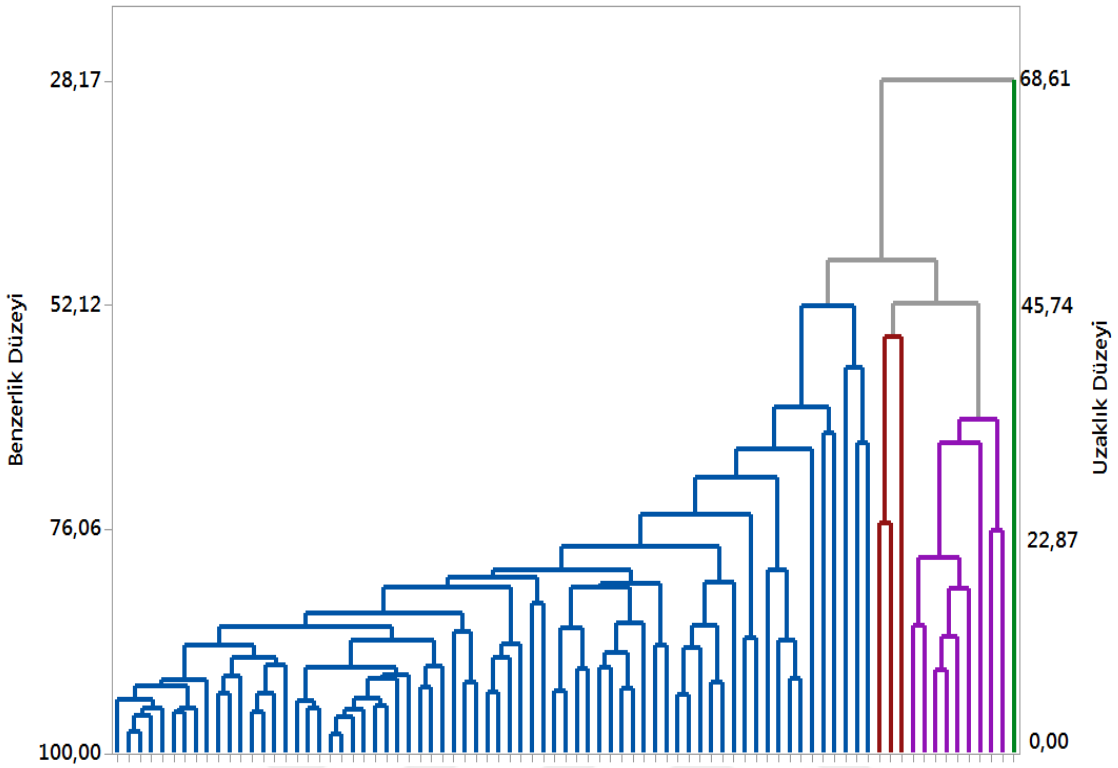
■ McQuitty Kümeleme Yöntemi

Türkiye'deki illerin hayvancılık istatistikleri bakımından McQuitty kümeleme yöntemi ile sınıflandırılmasına ilişkin bulgular Çizelge 4.3'de ve elde edilen dendrogram Şekil 4.3'de sunulmuştur.

Çizelge 4.3 İllerin McQuitty yöntemi ile kümelenmesi

Küme Sayısı	Uzaklık Düzeyi	Benzerlik Düzeyi	DSO(%)
2	50,22	47,42	100
3	45,74	52,12	96,3
4	45,66	52,19	95,1
5	42,46	55,55	95,1
6	39,35	58,80	95,1

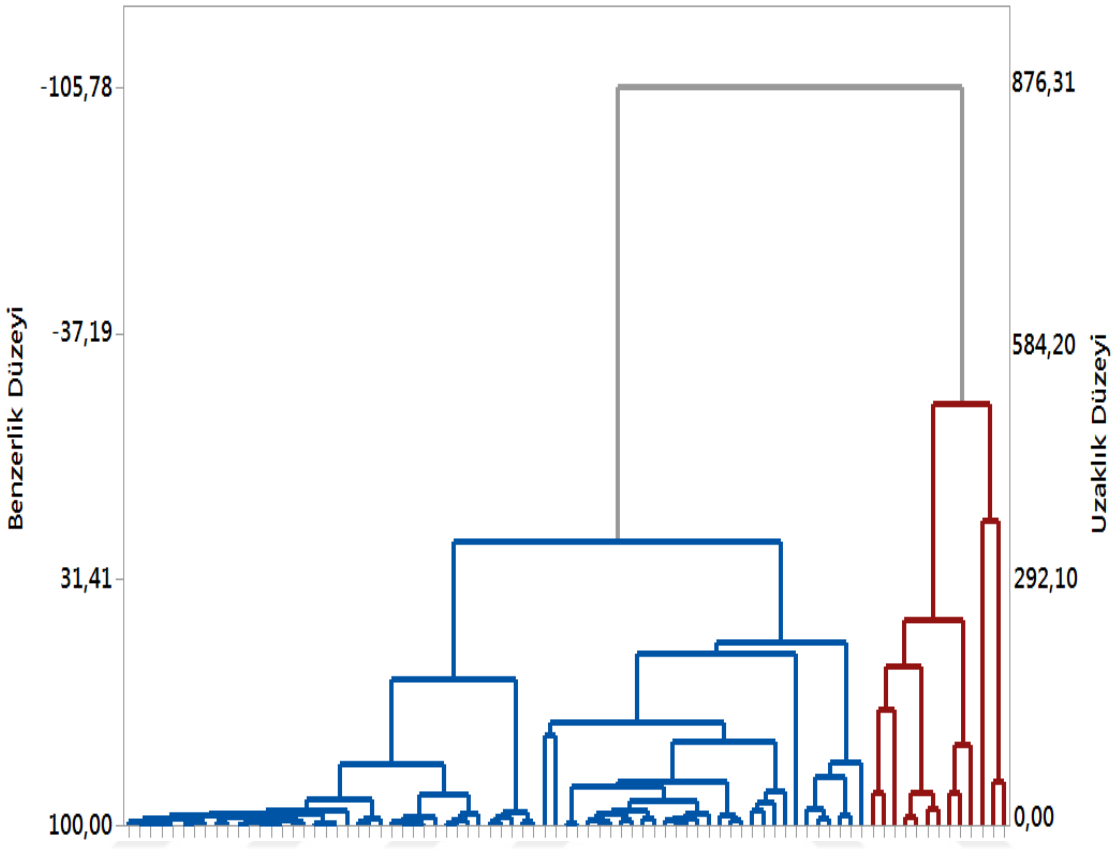
Şekil 4.3'deki dendogram ve Çizelge 4.3'deki McQuitty sonuçları incelendiğinde; illerin hayvancılık istatistikleri bakımından 47,42'lik benzerlik düzeyi ve 50,22 uzaklık düzeyinde 2 kümede toplandığı görülmektedir. Benzerlik düzeyleri 47,42'den 52,12'ye çıktığında küme sayısı 2'den 3'e, 52,19'a çıktığında ise küme sayısı 4'e, 55,55'e çıktığında ise küme sayısı 5'e yükselmiştir. Ayrıca dendrogram detaylı bir şekilde incelendiğinde, iki küme için Ardahan ili (en sağda) tek başına bir küme oluştururken diğer iller de ayrı bir kümede sınıflanarak diğer kümeyi oluşturmaktadır.



Şekil 4.3 İllerin McQuitty yöntemi ile kümelenmesine ilişkin dendrogram

■Ward Kümeleme Yöntemi (En Küçük Varyans Kümeleme Yöntemi)

Türkiye'deki illerin hayvancılık istatistikleri bakımından Ward kümeleme yöntemi ile sınıflandırılmasına ilişkin bulgular Çizelge 4.4'de ve elde edilen dendrogram Şekil 4.4'de sunulmuştur. Çizelge 4.4 ve Şekil 4.4'deki Ward kümeleme yöntemine ilişkin sonuçlar incelendiğinde, -17,30 benzerlik düzeyinde ve 499,54 uzaklık düzeyinde illerin 2 kümede toplandığı görülmektedir. Benzerlik düzeyi -17,30'dan 14,95'e çıktığında küme sayısı 2'den 3'e, 20,97'ye çıktığında küme sayısı 4'e, 42,86'ya çıktığında küme sayısı 5'e, 49,26'ya çıktığında ise küme sayısı 6'ya yükselmiştir. Bu bilgiler ile beraber Çizelge 4.4'deki DSO değerleri incelendiğinde en uygun küme sayısının 2 küme olduğu, diğer bir ifade ile illerin 2 temel sınıfta toplandığı görülmektedir.



Şekil 4.4 İllerin Ward yöntemi ile kümelenmesine ait dendrogram

İllerin detaylı olarak verildiği Çizelge 4.4'deki 1. kümede il sayısı 68, 2. kümedeki il sayısı ise 13'tür. 2. kümeyi Ağrı, Ardahan, Bingöl, Bitlis, Eskişehir, Iğdır, Hakkâri, Karaman, Kilis, Muş, Siirt, Şırnak, Tunceli illeri oluşturmaktadır. 2. kümeyi oluşturan bu illerin büyük bir çoğunluğunun doğu illeri olduğu dikkat çekmektedir. Küme sayısı 2'den 3'e çıktığında 3. kümenin Ardahan, Eskişehir ve Karaman illerinden ve küme sayısının 4'e yükselmesi ile Ardahan ili tek başına bir küme oluşturmaktadır.

Ward kümeleme yöntemi ile edilen 2 temel kümenin 29 değişken bakımından karşılaştırılmasına ilişkin t-test sonuçları Çizelge 4.5'te verilmiştir. Çizelge 4.5'te verilen hayvan varlığı ortalamaları her 1000 kişiye göre, hayvansal üretim ise kişi başına yıllık üretilen ortalamalardan oluşmaktadır.

Çizelge 4.4 İllerin Ward yöntemi ile kümelenmesi

Ölçüt	İller
UD=499,54 **BD**= -17,30 **KS**=2 **DSO**=100	**Küme1 (n=68):** Adana, Adıyaman Afyon, Aksaray, Amasya, Ankara, Antalya, Artvin, Aydın, Balıkesir, Bartın, Batman, Bayburt, Bilecik, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Diyarbakır, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, İstanbul, İzmir, Kahramanmaraş, Karabük, Kars, Kastamonu, Kayseri, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mardin, Mersin, Muğla, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Uşak, Van, Yalova, Yozgat, Zonguldak **Küme2 (n=13):** Ağrı, Ardahan, Bingöl, Bitlis, Eskişehir, Iğdır, Hakkâri, Karaman, Kilis, Muş, Siirt, Şırnak, Tunceli
UD=362,19 **BD**= 14,95 **KS**=3 **DSO**=100	**Küme1 (n=68):** Adana, Adıyaman Afyon, Aksaray, Amasya, Ankara, Antalya, Artvin, Aydın, Balıkesir, Bartın, Batman, Bayburt, Bilecik, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Diyarbakır, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, İstanbul, İzmir, Kahramanmaraş, Karabük, Kars, Kastamonu, Kayseri, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mardin, Mersin, Muğla, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Uşak, Van, Yalova, Yozgat, Zonguldak **Küme2 (n=10):** Ağrı, Bingöl, Bitlis, Iğdır, Hakkâri, Kilis, Muş, Siirt, Şırnak, Tunceli **Küme3 (n=3):** Ardahan, Eskişehir, Karaman
UD=243,35 **BD**= 45,86 **KS**=5 **DSO**=96,3	**Küme1 (n=38):** Adana, Adıyaman, Amasya, Ankara, Antalya, Bartın, Batman, Bilecik, Bursa, Diyarbakır, Düzce, Elazığ, Gaziantep, Giresun, Hatay, İstanbul, İzmir, Kahramanmaraş, Karabük, Kayseri, Kırıkkale, Kocaeli, Malatya, Manisa, Mardin, Mersin, Nevşehir, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Van, Yalova, Zonguldak **Küme2 (n=30):** Afyonkarahisar, Aksaray, Artvin, Aydın, Balıkesir, Bayburt, Bolu, Burdur, Çanakkale, Çankırı, Çorum, Denizli, Edirne, Erzincan, Erzurum, Gümüşhane, Isparta, Kars, Kastamonu, Kırklareli, Kırşehir, Konya, Kütahya, Muğla, Niğde, Ordu, Sinop, Sivas, Uşak, Yozgat **Küme3 (n=10):** Ağrı, Bingöl, Bitlis, Hakkâri, Iğdır, Kilis, Muş, Siirt, Şırnak, Tunceli **Küme4 (n=1):** Ardahan **Küme5 (n=2):** Eskişehir, Karaman
UD=216,07 **BD**= 49,26 **KS**=6 **DSO**=96,3	**Küme1 (n=38):** Adana, Adıyaman, Amasya, Ankara, Antalya, Bartın, Batman, Bilecik, Bursa, Diyarbakır, Düzce, Elazığ, Gaziantep, Giresun, Hatay, İstanbul, İzmir, Kahramanmaraş, Karabük, Kayseri, Kırıkkale, Kocaeli, Malatya, Manisa, Mardin, Mersin, Nevşehir, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Van, Yalova, Zonguldak **Küme2 (n=30):** Afyonkarahisar, Aksaray, Artvin, Aydın, Balıkesir, Bayburt, Bolu, Burdur, Çanakkale, Çankırı, Çorum, Denizli, Edirne, Erzincan, Erzurum, Gümüşhane, Isparta, Kars, Kastamonu, Kırklareli, Kırşehir, Konya, Kütahya, Muğla, Niğde, Ordu, Sinop, Sivas, Uşak, Yozgat **Küme3 (n=7):** Şırnak, Muş, Kilis, Iğdır, Hakkâri, Bingöl, Ağrı **Küme4 (n=1):** Ardahan **Küme5 (n=3):** Siirt, Tunceli, Bitlis **Küme6 (n=2):** Eskişehir, Karaman

UD: Uzaklık düzeyi BD: Benzerlik düzeyi KS: Küme sayısı DSO: Doğru sınıflandırma oranı

Çizelge 4.5'te, iki küme arasında t testi sonucunda 22 değişken açısından anlamlı farklılık tespit edilmiştir. T testi sonuçlarına göre hayvan varlığına ait değişkenlerden 13 (melez sığır sayısı, sağılan melez sığır sayısı, yerli sığır sayısı, sağılan yerli sığır sayısı, manda sayısı, sağılan manda sayısı, merinos koyunu sayısı, sağılan merinos koyunu sayısı, yerli ve diğer koyun sayısı, sağılan yerli ve diğer koyun sayısı, keçi (tiftik ve kıl) sayısı, sağılan tiftik keçisi sayısı, arı kovanı sayısı), hayvansal üretime ait 9 değişken (melez sığır sütü, yerli sığır sütü, manda sütü, merinos koyun sütü, yerli ve diğer koyun sütü, keçi (tiftik+kıl) sütü, merinos koyun yapağı, yerli ve diğer koyun yapağısı, tiftik ve keçi kılı) bakımından iki küme arasında anlamlı farklılıklar bulunmuştur ($p<0,05$). Ortalama değerleri incelendiğinde, farklılık gözlenen 22 değişkenin tamamında 2. küme değerlerinin daha yüksek değerlere sahip olduğu görülmektedir. Diğer bir ifade ile 2. kümede yer alan ve büyük bir çoğunluğu doğu illerinden oluşan 13 il (Ağrı, Ardahan, Bingöl, Bitlis, Eskişehir, Iğdır, Hakkâri, Karaman, Kilis, Muş, Siirt, Şırnak, Tunceli) hayvancılık göstergeleri bakımından daha gelişmiş illerdir. Söz konusu 13 ili kapsayan 2. küme hayvan varlığı açısından oldukça yüksek değerlere sahiptir. Aynı şekilde hayvansal üretimde kültür sığır sütü ve bal-balmumu dışında 2. küme çok yüksek değerleri içermektedir. Bu duruma kişi başına yıllık üretilen ortalama melez sığır sütünün 1. küme için 151,7 kg iken 2. küme için 432,8 kg, kişi başına yıllık üretilen yerli ve diğer koyunların sütünün 1. küme için 20,1 kg iken 2. küme için 79,6 kg ve kişi başına yıllık üretilen ortalama yerli ve diğer koyun yapağının 1. küme için 0,9 kg iken 2. küme için 3,3 kg olması örnek verilebilir.

Çizelge 4.5 Ward kümeleme yöntemi ile elde edilen kümelerin karşılaştırılmasına yönelik t-testi sonuçları

Hayvan Varlığı (Her Bin Kişide**)**	**Kod**	**1.Küme (n=68)** $\overline{X}$ ±SS	**2.Küme (n=13)** $\overline{X}$ ±SS	**P**
Kültür Sığır Sayısı	X1	174,0±0,15	101,0±0,08	0,09
Sağılan Kültür Sığır Sayısı	X2	66,2±0,06	39,0±0,03	0,12
Melez Sığır Sayısı	X3	148,2±0,19	389,8±0,72	**0,01***
Sağılan Melez Sığır Sayısı	X4	55,3±0,08	161,3±0,31	**0,01***
Yerli Sığır Sayısı	X5	26,1±0,03	84,4±0,08	**0,00***
Sağılan Yerli Sığır Sayısı	X6	10,1±0,01	35,6±0,03	**0,00***
Manda Sayısı	X7	3,7±0,00	8,4±0,01	**0,01***
Sağılan Manda Sayısı	X8	1,5±0,00	3,7±0,00	**0,01***
Merinos Koyunu Sayısı	X9	31,4±0,05	160,7±0,20	**0,00***
Sağılan Merinos Koyunu Sayısı	X10	14,2±0,02	95,8±0,11	**0,00***
Yerli ve Diğer Koyun Sayısı	X11	495,4±0,44	1798,6±1,23	**0,00***
Sağılan Yerli ve Dİğ.Koy.Say.	X12	254,6±0,24	1051,0±0,72	**0,00***
Keçisi Sayısı (Tiftik+Kıl)	X13	145,6±0,13	585,1±0,42	**0,00***
Sağılan Keçi Sayısı	X14	67,4±0,06	328,6±0,26	**0,00***
Et Tavuğu Sayısı	X15	8863,6±13,07	8935,9±7,49	0,98
Yumurta Tavuğu Sayısı	X16	1932,2±3,70	888,2±1,25	0,32
Diğer Kümes Hayvanların Sayısı	X17	124,4±0,44	159,3±0,26	0,78
Arı Kovanı Sayısı	X18	156,5±0,19	280,4±0,22	**0,04***
Hayvansal Üretim (Kişi Başı-kg)	**Kod**	**1.Küme (n=68)** $\overline{X}$ ±SS	**2.Küme (n=13)** $\overline{X}$ ±SS	**P**
Kültür Sığır Sütü	X19	254,6±237,28	138,2±116,28	0,09
Melez Sığır Sütü	X20	151,7±223,83	432,8±868,11	**0,02***
Yerli Sığır Sütü	X21	13,2±14,54	46,8±45,48	**0,00***
Manda Sütü	X22	1,5±1,97	3,5±4,49	**0,01***
Merinos Koyun Sütü	X23	0,7±1,23	4,5±5,01	**0,00***
Yerli ve Diğer Koyun Sütü	X24	20,1±19,33	79,6±58,98	**0,00***
Keçi (Tiftik+Kıl) Sütü	X25	6,9±6,84	34,9±27,79	**0,00***
Merinos Koyun Yapağı	X26	0,1±0,16	0,5±0,64	**0,00***
Yerli ve Diğer Koyun Yapağısı	X27	0,9±0,75	3,3±2,30	**0,00***
Tiftik ve Keçi Kılı	X28	0,1±0,07	0,3±0,24	**0,00***
Bal ve Balmumu	X29	2,1±3,48	2,3±2,01	0,79

P: t-test olasılık değeri *: $p<0,05$

■Hiyerarşik Kümeleme Yöntemlerinin Karşılaştırılması

Türkiye'deki illerin hayvancılık istatistikleri bakımından sınıflandırılmasına yönelik uygulanan hiyerarşik kümeleme yöntemleri sonucu, belirlenen küme sayıları için özet bulgular Çizelge 4.6'da verilmiştir.

Çizelge 4.6 Hiyerarşik kümeleme yöntemlerine ilişkin özet istatistikler

KS	TekBKY		TamBKY		McQuitty		Ward	
	UD(BD)	*DSO*	*UD(BD)*	*DSO*	*UD(BD)*	*DSO*	*UD(BD)*	*DSO*
2	8,9(56,6)	%100	283,90(33,33)	%100	50,22(47,42)	%100	499,54(-17,3)	%100
3	8,7(57,9)	%100	244,14(42,67)	%95,1	45,74(52,12)	%96,3	362,19(14,95)	%100
4	7,4(64,2)	%100	190,82(55,19)	%95,1	45,66(52,19)	%95,1	336,53(20,97)	%98,8
5	7,2(65,3)	%92,6	184,08(56,77)	%95,1	42,46(55,55)	%95,1	243,35(42,86)	%96,3
6	6,9(66,7)	%54,3	165,13(61,23)	%93,8	39,35(58,80)	%95,1	216,07(49,26)	%96,3

KS: Küme sayısı, UD: Uzaklık düzeyi, BD: Benzerlik düzeyi, DSO: Doğru sınıflandırma oranı

Çizelge 4.6'daki bulgular, TekBKY, TamBKY, McQuitty, Ward kümeleme yöntemlerine ait uzaklık ve benzerlik düzeyleri ile diskriminant analizi DSO değerlerinden oluşmaktadır. Söz konusu kümeleme yöntemlerinde 2 kümede uzaklık ve benzerlik düzeyleri birbirinden farklı olmasına rağmen DSO değerlerinin tamamı %100'dür. 3. kümede Ward ve TekBKY'de %100 iken TamBKY'de %95,1 ve McQuitty yönteminde %96,3'e düşmüştür. Diğer yöntemlerde de görüldüğü üzere küme sayısı arttıkça DSO değerleri düşmektedir. DSO değerlerinin %100 olması önemli bir gösterge olarak değerlendirilmiştir. Bu çerçevede 2 kümenin en uygun küme yapısı olduğu sonucuna varılmıştır. TekBKY de uzaklık değeri 8,9 ve benzerlik değeri 56,6, TamBKY'de uzaklık değeri 283,9 ve benzerlik değeri 33,33, McQuitty yönteminde uzaklık değeri 50,22 ve benzerlik değeri 47,42, Ward kümeleme yönteminde ise uzaklık değeri 499,57 ve benzerlik değeri -17,30'dur. Burada kümeleme yöntemlerinin karşılaştırılmasında DSO değerleri %100 olsa da uzaklık ve benzerlik düzeyleri incelendiğinde uzaklık değerinin en yüksek, benzerlik düzeyinin en düşük olduğu (ki

istenen durumdur) yöntem Ward kümeleme yöntemidir. Bu bilgiler ışığında hiyerarşik kümeleme yöntemleri arasında Ward kümeleme yöntemi en uygun yöntem olarak görülmektedir.

4.1.2 Hiyerarşik Olmayan Kümeleme Analizine İlişkin Bulgular

■ K - Ortalamalar (k-Means) Kümeleme Yöntemi

Türkiye'deki illerin hayvancılık istatistikleri bakımından k-Ortalamalar kümeleme yöntemi ile sınıflandırılmasına ilişkin bulgular Çizelge 4.7, Çizelge 4.8'de ve bu kümeleme yöntemi ile edilen 2 temel kümenin 29 değişken bakımından karşılaştırılmasına ilişkin t-test sonuçları Çizelge 4.9'da sunulmuştur.

Çizelge 4.7 İllerin k-Ortalamalar kümeleme analizi sonuçlarına göre DSO değerleri

Küme Sayısı	DSO(%)	Wilk's Lamda	SD	P
2	88,8	0,141	28	0,000*
3	87,6	0,015	50	0,000*
4	87,6	0,002	84	0,000*
5	86,3	0,000	116	0,000*
6	86,3	0,000	140	0,000*

*: $p<0,05$

Çizelge 4.7'deki diskriminant analizi ile elde edilen DSO değerleri incelendiğinde küme sayılarının tamamı istatistiksel açıdan anlamlı bulunmuştur ($p<0,05$). Bununla beraber 2 kümeli yapı %88,8 DSO değeri ile en yüksektir. 3 ve 4 küme sayılarında DSO değerleri %87,6 ile eşittir. 5 ve 6 küme sayılarında da %86,3 ile eşitlik söz konusudur. Bu çerçevede en yüksek DSO değerinin 2 kümeli yapıda olması nedeniyle 2 kümenin en uygun küme yapısı olduğu sonucuna varılmıştır.

Çizelge 4.8 İllerin k-ortalamalar yöntemi ile kümelenmesi

Küme Merkezinden Uzaklık Düzeyi	İller
Küme Sayısı: 2 **Küme1:** 5,59 **Küme2:** 4,35	**Küme1 (n=2):** Ardahan, Kars, **Küme2 (n=79):** Adana, Adıyaman, Ağrı, Afyon, Aksaray, Amasya, Ankara, Antalya, Artvin, Aydın, Balıkesir, Bartın, Batman, Bayburt, Bilecik, Bingöl, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Diyarbakır, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, İstanbul, İzmir, Kahramanmaraş, Karabük, Kastamonu, Kayseri, Kilis, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mardin, Mersin, Muğla, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Uşak, Van, Yalova, Yozgat, Zonguldak, Eskişehir, Karaman, Bitlis, Hakkâri, Iğdır, Muş, Tunceli, Şırnak, Siirt
Küme Sayısı: 3 **Küme1:** 0,00 **Küme2:** 3,58 **Küme3:** 7,09	**Küme1 (n=1):** Ardahan **Küme2 (n=71):** Adana, Adıyaman, Ağrı, Afyon, Aksaray, Amasya, Ankara, Antalya, Artvin, Aydın, Balıkesir, Bartın, Batman, Bayburt, Bilecik, Bingöl, Bolu, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Diyarbakır, Düzce, Edirne, Elazığ, Eskişehir, Erzincan, Erzurum, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, İstanbul, İzmir, Kahramanmaraş, Karabük, Kars, Kastamonu, Kayseri, Kilis, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mardin, Mersin, Muğla, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Uşak, Van, Yalova, Yozgat, Zonguldak **Küme3 (n=9):** Bitlis, Burdur, Hakkâri, Iğdır, Karaman, Muş, Siirt, Şırnak, Tunceli,
Küme Sayısı: 4 **Küme1:** 0,00 **Küme2:** 7,09 **Küme3:** 6,54 **Küme4:** 3,34	**Küme1 (n= 1):** Ardahan **Küme2 (n=9):** Bitlis, Burdur, Hakkâri, Iğdır, Karaman, Muş, Siirt, Şırnak, Tunceli **Küme3 (n=3):** Afyon, Bayburt, Bolu **Küme4 (n=68):** Adana, Adıyaman Aksaray, Amasya, Ankara, Antalya, Artvin, Aydın, Ağrı, Bingöl, Balıkesir, Bartın, Batman, Bilecik, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Diyarbakır, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, İstanbul, İzmir, Kahramanmaraş, Karabük, Kars, Kastamonu, Kayseri, Kilis, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mardin, Mersin, Muğla, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Uşak, Van, Yalova, Yozgat, Zonguldak
Küme Sayısı: 5 **Küme1:** 3,99 **Küme2:** 0,00 **Küme3:** 2,78 **Küme4:** 5,29 **Küme5:** 6,08	**Küme1 (n= 26):** Afyon, Aksaray, Amasya, Balıkesir, Bartın, Bolu, Burdur, Bayburt, Çanakkale, Çankırı, Çorum, Edirne, Erzincan, Erzurum, Gümüşhane, Kars, Kastamonu, Kırklareli, Kırşehir, Konya, Kütahya, Niğde, Sivas, Tokat, Uşak, Yozgat, **Küme2 (n=1):** Ardahan **Küme3 (n=45):** Adana, Adıyaman, Ankara, Antalya, Artvin, Aydın, Batman, Bilecik, Bingöl, Bursa, Denizli, Diyarbakır, Düzce, Elazığ, Gaziantep, Giresun, Hatay, Isparta, İstanbul, İzmir, Kahramanmaraş, Karabük, Kayseri, Kilis, Kırıkkale, Kocaeli, Malatya, Manisa, Mardin, Mersin, Muğla, Nevşehir, Ordu, Osmaniye, Rize, Sakarya, Samsun, Şanlıurfa, Sinop, Tekirdağ, Trabzon, Van, Yalova, Zonguldak, Eskişehir, **Küme4 (n=3):** Ağrı, Iğdır, Muş, **Küme5 (n=6):** Bitlis, Hakkâri, Siirt, Karaman, Şırnak, Tunceli

Çizelge 4.8'deki illerin küme merkezlerinden ortalama uzaklık düzeyleri ile küme sayısına göre illerin dağılımı görülmektedir. Birinci kümedeki il sayısı 2, ikinci küme ise 79 ilden oluşmaktadır. K-Ortalamalar tekniğine göre iller hayvancılık istatistikleri bakımından 2 kümeye ayrıldığında 1. küme Ardahan ve Kars illerinden oluşmaktadır.

Çizelge 4.9'daki t testi sonuçlarına göre hayvan varlığına ait değişkenlerden 7 (melez sığır sayısı, sağılan melez sığır sayısı, yerli sığır sayısı, sağılan yerli sığır sayısı, merinos koyunu sayısı, sağılan merinos koyunu sayısı ve diğer kümes hayvanların sayısı), hayvansal üretime ait 4 değişken (melez sığır sütü, yerli sığır sütü, merinos koyun sütü, merinos koyun yapağı) bakımından iki küme arasında anlamlı farklılıklar bulunmuştur ($p<0,05$). Ortalama değerleri incelendiğinde, farklılık gözlenen 11 değişkenin tamamında 1. küme değerlerinin daha yüksek değerlere sahip olduğu görülmektedir. Diğer bir ifade ile 1. kümede yer alan Ardahan ve Kars illeri hayvancılık göstergeleri bakımından daha gelişmiş illerdir. Söz konusu 2 ili kapsayan 1. küme hayvan varlığı açısından oldukça yüksek değerlere sahiptir. Aynı şekilde hayvansal üretimde melez sığır sütü, yerli sığır sütü, merinos koyun sütü, merinos koyun yapağı değişkenleri 1. kümede çok yüksek değerleri içermektedir. Bu duruma kişi başına yıllık üretilen ortalama melez sığır sütünün 1. küme için 2267,98 kg iken 2. küme için 114,38 kg, kişi başına yıllık üretilen yerli sığır sütünün 1. küme için 83,53 kg iken 2. küme için 16,96 kg ve kişi başına yıllık üretilen merinos koyun sütünün 1. küme için 6,32 kg iken 2. küme için 1,18 kg olması örnek verilebilir. Çizelge 4.9'da hayvan varlığı ortalamaları her 1000 kişi için yıllık üretilen ortalamalardan, hayvansal üretim ise kişi başına yıllık üretilen ortalamalardan oluşmaktadır.

K-Ortalamalar ve Ward kümeleme yöntemi ile oluşan küme sayılarının aynı olduğu görülmektedir. Çizelge 4.5 ile Çizelge 4.9 karşılaştırıldığında k-Ortalamalar kümeleme yönteminde anlamlı farklılıklar bulunan değişkenlerin tamamı Ward kümeleme yönteminde de bulunmuş fakat Ward kümeleme yönteminde daha fazla değişken anlamlı farklılık göstermiştir. Bununla beraber Ward kümeleme yönteminde bir kümede 13 il yer alırken k-Ortalamalar yönteminin bir kümesinde ise 2 il yer almaktadır.

Çizelge 4.9 k-Ortalamalar kümeleme yöntemi ile elde edilen kümelerin karşılaştırılmasına yönelik t-testi sonuçları

Hayvan Varlığı (Her Bin Kişide)	**Kod**	**1.Küme (n=2)** $\overline{X}$ ±SS	**2.Küme (n=79)** $\overline{X}$ ±SS	**P**
Kültür Sığır Sayısı	X1	276,7±0,15	159,4±0,14	0,26
Sağılan Kültür Sığır Sayısı	X2	92,3±0,04	61,0±0,06	0,45
Melez Sığır Sayısı	X3	1909,4±1,19	143,4±0,16	**0,00***
Sağılan Melez Sığır Sayısı	X4	800,9±0,53	53,9±0,06	**0,00***
Yerli Sığır Sayısı	X5	144,4±0,05	32,7±0,04	**0,00***
Sağılan Yerli Sığır Sayısı	X6	63,9±0,04	12,9±0,02	**0,00***
Manda Sayısı	X7	0,2±0,00	4,5±0,01	0,33
Sağılan Manda Sayısı	X8	0,001±0,00	1,9±0,00	0,34
Merinos Koyunu Sayısı	X9	256,3±0,18	47,0±0,10	**0,00***
Sağılan Merinos Koyunu Sayısı	X10	130,0±0,09	24,7±0,05	**0,00***
Yerli ve Diğer Koyun Sayısı	X11	1094,3±0,53	694,7±0,79	0,48
Sağılan Yerli ve Dİğ.Koy.Say.	X12	493,6±0,19	379,6±0,47	0,73
Keçisi Sayısı (Tiftik+Kıl)	X13	71,1±0,06	219,8±0,26	0,43
Sağılan Keçi Sayısı	X14	24,8±0,01	111,4±0,15	0,43
Et Tavuğu Sayısı	X15	7799,7±11,00	8902,4±12,40	0,90
Yumurta Tavuğu Sayısı	X16	1044,2±0,14	1782,9±3,48	0,77
Diğer Kümes Hayvanların Sayısı	X17	1012,3±0,14	107,7±0,40	**0,00***
Arı Kovanı Sayısı	X18	287,1±0,09	173,6±0,20	0,42
Hayvansal Üretim (Kişi Başı-kg)	**Kod**	**1.Küme (n=2)** $\overline{X}$ ±SS	**2.Küme (n=79)** $\overline{X}$ ±SS	**P**
Kültür Sığır Sütü	X19	347,78±151,32	233,12±227,41	0,48
Melez Sığır Sütü	X20	2267,98±1432,9	144,38±175,85	**0,00***
Yerli Sığır Sütü	X21	83,53±48,55	16,96±22,72	**0,00***
Manda Sütü	X22	0,04±0,02	1,90±2,63	0,33
Merinos Koyun Sütü	X23	6,32±4,38	1,18±2,50	**0,00***
Yerli ve Diğer Koyun Sütü	X24	35,99±11,22	29,44±36,73	0,80
Keçi (Tiftik+Kıl) Sütü	X25	2,46±1,46	11,65±16,33	0,43
Merinos Koyun Yapağı	X26	0,79±0,54	0,15±0,31	**0,00***
Yerli ve Diğer Koyun Yapağısı	X27	1,54±0,93	1,25±1,44	0,78
Tiftik ve Keçi Kılı	X28	0,05±0,04	0,13±0,15	0,48
Bal ve Balmumu	X29	3,27±1,84	2,06±3,31	0,61

P: t-test olasılık değeri *: p<0,05

■ **Fuzzy (Bulanık) Kümeleme Yöntemi**

Türkiye'deki illerin hayvancılık istatistikleri bakımından fuzzy kümeleme yöntemi ile sınıflandırılmasına ilişkin bulgular Çizelge 4.10, Çizelge 4.11'de, bu kümeleme yöntemi ile edilen 5 temel kümenin 29 değişken bakımından karşılaştırılmasına ilişkin tek faktörlü varyans analizi test sonuçları Çizelge 4.12'de, oluşan kümelerin illere göre dağılımı Çizelge 4.13'te sunulmuştur.

Çizelge 4.10 İllerin hayvancılık istatistiklerine ilişkin bulanık kümeleme sonuçları

Küme Sayısı (k)	DSO(%)	$\overline{SC}$	F(u)	$F_c(u)$	D(u)	$D_c(u)$
2	86,4	0,304289	0,6386	0,2772	0,2366	0,4731
3	87,7	0,348329	0,6791	0,5186	0,2024	0,3036
4	82,7	0,328273	0,6823	0,5764	0,2128	0,2838
5	90,1	0,436060	0,6871	0,6189	0,1919	0,2398
6	88,1	0,352075	0,6803	0,6063	0,2168	0,2602

Çizelge 4.10'daki fuzzy kümeleme yöntemi ile hesaplanan gölge istatistiği veya siluet katsayısı ($\overline{SC}$ =0,4360), Dunn parçalama katsayısı (F(u)= 0,6871) ve normalize Dunn katsayısının ($F_c(u)$= 0,6189) en yüksek değerleri 5 küme yapısında oluşmuş ve en iyi veya en kararlı sınıflandırılma elde edilmiştir. Aynı zamanda Kaufmann katsayısı (D(u)= 0,1919) ve normalize Kaufmann katsayısının ($D_c(u)$=0,2398) sıfıra yaklaşması ile beraber en uygun küme sayısının 5 olduğu sonucuna varılmaktadır. En uygun küme sayısının belirlenmesinde fuzzy kümeleme yöntemine ait istatistikler haricinde diskriminant analizi DSO değerleri incelendiğinde ise en yüksek değerin %90,1 ile yine 5 küme yapısında oluştuğu görülmektedir. Bu bilgiler ışığında elde edilen tüm sonuçlara göre illerin 5 kümeye ayrılması ile en uygun küme yapısına ulaşılmıştır.

Çizelge 4.11 İllerin küme üyeliği ve küme üyelik olasılıkları

İller	Küme Üyeliği	1.Küme Olasılığı	2. Küme Olasılığı	3. Küme Olasılığı	4.Küme Olasılığı	5. Küme Olasılığı
*Adana***	*5*	*0,1867*	*0,1866*	*0,1675*	*0,1774*	***0,2817***
Adıyaman	4	0,0054	0,0055	0,0050	**0,9789**	0,0053
Afyon	2	0,0000	**0,9224**	0,0776	0,0000	0,0000
Ağrı	2	0,0100	**0,7601**	0,0062	0,0076	0,2161
Aksaray	1	**0,9996**	0,0001	0,0001	0,0001	0,0001
Amasya	5	0,0227	0,0356	0,0000	0,0225	**0,9191**
Ankara	2	0,0062	**0,9762**	0,0060	0,0058	0,0059
Antalya	4	0,0054	0,0055	0,0050	**0,9789**	0,0053
Ardahan	2	0,0000	**0,6193**	0,0000	0,0000	0,3807
Artvin	1	**0,5902**	0,0000	0,0000	0,0000	0,4098
*Aydın***	*3*	*0,1491*	*0,1277*	***0,3849***	*0,1459*	*0,1924*
Balıkesir	1	**0,5513**	0,3686	0,0800	0,0000	0,0000
Bartın	3	0,1025	0,0896	**0,5184**	0,0828	0,2067
Batman	4	0,0075	0,0060	0,0037	**0,7970**	0,1859
Bayburt	1	**0,5591**	0,4409	0,0000	0,0000	0,0000
Bilecik	1	**0,9837**	0,0043	0,0041	0,0040	0,0040
Bingöl	4	0,0000	0,0000	0,0000	**0,5275**	0,4725
*Bitlis****	*1*	***0,3263***	*0,2592*	*0,0000*	*0,2607*	*0,1538*
Bolu	1	**0,6776**	0,3024	0,0201	0,0000	0,0000
Burdur	1	**0,8398**	0,0000	0,0000	0,1602	0,0000
Bursa	2	0,0041	**0,9678**	0,0207	0,0040	0,0034
Çanakkale	1	**0,5832**	0,0000	0,0000	0,4168	0,0000
Çankırı	1	**0,6073**	0,0000	0,0989	0,0000	0,2937
Çorum	3	0,0000	0,0000	**1,0000**	0,0000	0,0000
Denizli	4	0,0000	0,0000	0,0455	**0,9545**	0,0000
Diyarbakır	5	0,0368	0,0248	0,0250	0,0643	**0,8491**
Düzce	1	**0,8343**	0,0033	0,0045	0,0034	0,1544
Edirne	2	0,0000	**1,0000**	0,0000	0,0000	0,0000
Elazığ	2	0,0054	**0,9385**	0,0185	0,0281	0,0095
Erzincan	5	0,0000	0,0000	0,0000	0,0000	**1,0000**
Erzurum	5	0,0000	0,0000	0,0000	0,0000	**1,0000**
Eskişehir	2	0,0040	**0,9679**	0,0206	0,0040	0,0034
Gaziantep	3	0,0971	0,1415	**0,5704**	0,1037	0,0874
Giresun	5	0,0000	0,0000	0,0000	0,0000	**1,0000**
Gümüşhane	1	**0,6720**	0,0000	0,0000	0,0000	0,3280
*Hakkâri****	*1*	***0,3109***	*0,2613*	*0,0000*	*0,2277*	*0,2001*
*Hatay****	*3*	*0,1944*	*0,1965*	***0,2156***	*0,1962*	*0,1973*
Iğdır	2	0,0000	**0,9825**	0,0016	0,0000	0,0159
Isparta	1	**0,9927**	0,0001	0,0003	0,0000	0,0069
*İstanbul****	*3*	*0,1945*	*0,1963*	***0,2157***	*0,1963*	*0,1973*
İzmir	3	0,1389	0,1412	**0,4222**	0,1433	0,1545

***: Bulanık yapıya sahip iller

Çizelge 4.11 (Devam) İllerin küme üyeliği ve küme üyelik olasılıkları

İller	Küme Üyeliği	1.Küme Olasılığı	2. Küme Olasılığı	3. Küme Olasılığı	4.Küme Olasılığı	5. Küme Olasılığı
K.Maraş	4	0,0051	0,0052	0,0047	**0,9801**	0,0049
Karabük	2	0,0045	**0,9727**	0,0077	0,0039	0,0112
*Karaman****	*1*	***0,3258***	*0,2284*	*0,0531*	*0,2362*	*0,1565*
Kars	1	**0,4297**	0,3616	0,0000	0,0000	0,2087
Kastamonu	2	0,0000	**0,6273**	0,0000	0,0000	0,3727
Kayseri	3	0,0733	0,1214	**0,6124**	0,0607	0,1322
Kilis	4	0,0077	0,0061	0,0036	**0,8030**	0,1796
Kırıkkale	1	**0,9719**	0,0045	0,0035	0,0047	0,0154
Kırklareli	4	0,0000	0,0000	0,0000	**1,000**	0,0000
Kırşehir	1	**0,9605**	0,0000	0,0395	0,0000	0,0000
*Kocaeli****	*3*	*0,1944*	*0,1962*	***0,2159***	*0,1962*	*0,1973*
Konya	2	0,0000	**0,9218**	0,0782	0,0000	0,0000
Kütahya	4	0,0000	0,0000	0,0655	**0,9345**	0,0000
*Malatya****	*3*	*0,1944*	*0,1962*	***0,2159***	*0,1962*	*0,1973*
Manisa	1	**0,9688**	0,0063	0,0177	0,0039	0,0032
Mardin	4	0,0076	0,0060	0,0036	**0,8044**	0,1784
Mersin	1	**0,5422**	0,0025	0,0021	0,4508	0,0023
Muğla	4	0,0000	0,0000	0,0000	**1,0000**	0,0000
Muş	1	**0,4062**	0,2213	0,0032	0,3276	0,0417
Nevşehir	3	0,0000	0,0000	**1,0000**	0,0000	0,0000
Niğde	4	0,0000	0,0000	0,0712	**0,9288**	0,0000
*Ordu****	*5*	*0,1860*	*0,1855*	*0,1694*	*0,1772*	***0,2819***
*Osmaniye****	*3*	*0,1944*	*0,1962*	***0,2160***	*0,1961*	*0,1973*
Rize	1	**1,0000**	0,0000	0,0000	0,0000	0,0000
Sakarya	1	**0,9833**	0,0045	0,0040	0,0043	0,0040
Samsun	5	0,0254	0,0206	0,0261	0,019	**0,9089**
Şanlıurfa	5	0,0332	0,0301	0,0178	0,0314	**0,8875**
*Siirt****	*1*	***0,3337***	*0,2654*	*0,0000*	*0,2538*	*0,1470*
Sinop	1	**0,8368**	0,0034	0,0044	0,0034	0,1520
Şırnak	4	0,0076	0,0060	0,0035	**0,8069**	0,1760
Sivas	5	0,0000	0,0000	0,0000	0,0000	**1,0000**
Tekirdağ	2	0,0064	**0,9760**	0,0060	0,0058	0,0059
Tokat	5	0,0256	0,0207	0,0262	0,0192	**0,9084**
*Trabzon****	*5*	*0,1830*	*0,1702*	*0,1749*	*0,1798*	***0,2921***
Tunceli	4	0,0000	0,0000	0,0000	**0,6339**	0,3661
Uşak	1	**0,9969**	0,0005	0,0011	0,0008	0,0008
Van	5	0,0334	0,0302	0,0179	0,0317	**0,8869**
Yalova	1	**0,9841**	0,0043	0,0040	0,0039	0,0038
Yozgat	5	0,0203	0,0194	0,0877	0,0239	**0,8487**
Zonguldak	1	**0,9840**	0,0043	0,0040	0,0039	0,0038

***: Bulanık yapıya sahip iller

Çizelge 4.11'de küme üyelik olasılıkları verilmiş ve bu olasılık değerleri illerin her bir kümeye ne kadar bağlı olduklarını göstermektedir. 1. küme de 28 il, 2. küme de 13 il, 3. kümede 12 il, 4. ve 5. kümede ise 14'er il yer almaktadır. Bu illerin kümelere göre dağılımı Çizelge 4.13'te verilmiştir. Söz konusu illerden Adana, Hatay, İstanbul, Kocaeli, Malatya, Ordu, Osmaniye, Siirt ve Trabzon'un bulundukları kümedeki olasılık değerleri diğer 4 kümeye de oldukça yakındır. Olasılık değerlerindeki bu yakınlık illerin kararsız bir yapıya sahip oldukları yani bulanık bir yapıda olduklarını göstermektedir. Adıyaman, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Giresun, Iğdır, Isparta, Kahramanmaraş gibi illerin ise olasılık değerlerinin 1 veya 1'e çok yakın bir değer alarak bulundukları kümeye net olasılık değeri ile ayrılmışlardır.

Çizelge 4.12'deki tek faktörlü varyans analizi sonuçlarına göre kümeler arasında 21 değişkende anlamlı farklılık bulunmuştur. Bu çerçevede tek faktörlü varyans analizi sonuçlarına göre hayvan varlığına ait değişkenlerden 12, hayvansal üretime ait 9 değişken bakımından 5 küme arasında anlamlı farklılıklar bulunmuştur ($p<0,05$). Ortalama değerleri incelendiğinde; et tavuğu sayısı 1. kümede yüksek değerlere sahiptir, aynı zamanda bu kümede, çoğunlukla Manisa, Balıkesir, Uşak gibi tavukçuluk açısından gelişmiş iller yer almaktadır. Melez sığır sayısı, yerli sığır sayısı, merinos koyun sayısı, melez sığır sütü, yerli sığır sütü, merinos koyun sütü, yerli ve diğer koyun sütü gibi değişkenlerin 2. kümede yüksek değerlere sahip olmasıyla beraber bu kümede daha çok Ankara, Afyonkarahisar, Konya gibi hem küçükbaş hayvan hem de büyükbaş hayvan üreticiliği açısından gelişmiş iller yer almaktadır. 4. kümede sağılan yerli ve diğer koyunların sayısı, keçi sayısı, keçi sütü gibi değişkenler yüksek değerlere sahiptir ve bu kümede Kilis, Mardin, Muğla illeri gibi keçi yetiştiriciliği açısından gelişmiş iller daha çok yer almaktadır. 5. kümede ise bal ve balmumu değişkeninin ortalaması yüksek değere sahiptir. Ortalama değerlere, kişi başına yıllık üretilen ortalama merinos koyun sütünün 2. küme için 3,6 kg ile en yüksek olması, kişi başına yıllık üretilen ortalama tiftik ve keçi kılının 0,24 kg ile 3. kümede en yüksek olması, kişi başına yıllık üretilen ortalama keçi sütünün 23,31 kg ile 4. kümede en yüksek olması örnek verilebilir. Çizelge 4.12'deki hayvan varlığı ortalamaları her 1000 kişi için yıllık üretilen ortalamalardan, hayvansal üretim ise kişi başına yıllık üretilen ortalamalardan oluşmaktadır.

Çizelge 4.12 Fuzzy kümeleme yöntemi ile elde edilen kümelerin karşılaştırılmasına yönelik tek faktörlü varyans analizi sonuçları

Kod	1.Küme (n=28) $\overline{X}$ ±SS	2.Küme (n=13) $\overline{X}$ ±SS	3.Küme (n=12) $\overline{X}$ ±SS	4.Küme (n=14) $\overline{X}$ ±SS	5.Küme (n=14) $\overline{X}$ ±SS	P
X1	192,59±0,18	185,23±0,14	123,72±0,09	144,74±0,13	130,85±0,10	0,83
X2	72,21±0,07	71,09±0,06	45,26±0,03	60,35±0,06	48,02±0,04	0,80
X3	193,33±0,26^{b}	351,05±0,73^{a}	77,29±0,05^{b}	92,01±0,09^{b}	211,02±0,18^{b}	**0,02***
X4	74,69±0,11^{b}	142,70±0,32^{a}	26,35±0,02^{c}	36,14±0,03^{c}	77,92±0,07^{b}	**0,02***
X5	37,03±0,04^{b}	51,52±0,08^{a}	10,18±0,01b	33,08±0,04^{b}	41,48±0,03^{b}	**0,03***
X6	15,86±0,02^{c}	20,33±0,03^{a}	3,89±0,00^{b}	13,36±0,01^{c}	14,73±0,01^{c}	**0,04***
X7	4,67±0,01	2,93±0,00	3,13±0,01	3,59±0,01	7,28±0,01	0,17
X8	2,09±0,00	1,17±0,00	1,24±0,00	1,53±0,00	2,98±0,00	0,17
X9	64,79±0,11^{b}	160,82±0,15^{a}	4,57±0,00^{c}	10,83±0,01^{c}	7,89±0,01^{c}	**0,00***
X10	35,99±0,07^{b}	76,96±0,07^{a}	1,74±0,00^{c}	8,84±0,02^{c}	4,04±0,01^{c}	**0,00***
X11	723,93±0,62^{c}	931,10±1,37^{a}	208,27±0,15^{d}	886,57±0,78^{b}	699,04±0,60^{c}	**0,02***
X12	389,40±0,38^{a}	495,88±0,78^{a}	104,50±0,09^{b}	508,29±0,49^{a}	375,28±0,34^{a}	**0,02***
X13	277,58±0,33^{b}	99,15±0,07^{c}	71,67±0,05^{c}	422,86±0,27^{a}	118,97±0,07^{b}	**0,00***
X14	142,26±0,20^{a}	42,58±0,04^{b}	33,29±0,02^{b}	220,27±0,16^{a}	59,51±0,04^{b}	**0,00***
X15	20561±14,31^{a}	4070,3±6,18^{b}	1775,3±1,75^{c}	2231,9±2,89^{b}	2692,3±2,79^{b}	**0,00***
X16	1871,07±2,8	3392,3±7,05	1828,5±2,28	932,87±0,74	817,32±1,13	0,22
X17	254,48±0,68	123,46±0,24	31,53±0,03	59,06±0,05	42,42±0,03	0,23
X18	210,66±0,20^{a}	97,72±0,09^{b}	82,53±0,07^{b}	230,07±0,28^{a}	207,57±0,19^{a}	**0,02***
Kod	**1.Küme (n=28)** $\overline{X}$ ±SS	**2.Küme (n=13)** $\overline{X}$ ±SS	**3.Küme (n=12)** $\overline{X}$ ±SS	**4.Küme (n=14)** $\overline{X}$ ±SS	**5.Küme (n=14)** $\overline{X}$ ±SS	**P**
X19	275,36±285,9	275,8±224,1	171,9±129,1	230,6±226,8	180,39±142,1	0,82
X20	203,04±301,5^{b}	390,9±878,4^{a}	70,55±52,53^{c}	92,92±84,83^{c}	216,3±212,3^{b}	**0,02***
X21	20,82±28,23	27,11±40,61	5,31±5,40	17,12±18,83	19,14±11,77	**0,04***
X22	2,07±2,81	1,20±1,57	1,17±2,24	1,52±3,06	2,93±2,73	0,16
X23	1,74±3,36^{b}	3,60±3,22^{a}	0,13±0,16^{b}	0,45±0,85^{b}	0,20±0,37^{b}	**0,00***
X24	29,21±27,30^{a}	39,64±64,91^{a}	8,36±6,66^{b}	39,06±35,76^{a}	29,83±27,26^{a}	**0,02***
X25	14,98±21,50^{b}	4,16±3,66^{b}	3,35±2,41^{b}	23,31±17,24^{a}	6,08±3,84^{b}	**0,00***
X26	0,19±0,34^{b}	0,51±0,50^{a}	0,01±0,02^{b}	0,03±0,04^{b}	0,02±0,05^{b}	**0,00***
X27	1,31±1,17^{a}	1,63±2,39^{a}	0,37±0,27^{b}	1,58±1,54^{a}	1,27±1,02^{a}	**0,02***
X28	0,16±0,19^{b}	0,07±0,05^{b}	0,04±0,03^{b}	0,24±0,16^{a}	0,07±0,04^{b}	**0,00***
X29	2,09±1,91^{a}	0,81±0,65^{b}	0,84±1,05^{b}	2,49±4,10^{a}	3,96±5,79^{a}	**0,01***

P: Varyans analizi olasılık değeri *: $p<0,05$

a, b, c, d: Farklı harfleri içeren kümeler arasında anlamlı farklılık vardır.

Çizelge 4.13 Fuzzy kümeleme yöntemi ile oluşan kümelerde illerin dağılımı

Kümeler	**İller**
1.Küme (n=28)	Aksaray, Artvin, Balıkesir, Bayburt, Bilecik, *Bitlis*, Bolu, Burdur, Çanakkale, Çankırı, Düzce, Gümüşhane, *Hakkâri*, Isparta, *Karaman*, Kars, Kırıkkale, Kırşehir, Manisa, Mersin, Muş, Rize, Sakarya, *Siirt*, Sinop, Uşak, Yalova, Zonguldak
2. Küme (n=13)	Afyonkarahisar, Ağrı, Ankara, Ardahan, Bursa, Edirne, Elazığ, Eskişehir, Iğdır, Karabük, Kastamonu, Konya, Tekirdağ
3. Küme (n=12)	*Aydın*, Bartın, Çorum, Gaziantep, *Hatay*, *İstanbul*, İzmir, Kayseri, *Kocaeli*, *Malatya*, Nevşehir, *Osmaniye*
4. Küme (n=14)	Adıyaman, Antalya, Batman, Bingöl, Denizli, Kahramanmaraş, Kilis, Kırklareli, Kütahya, Mardin, Muğla, Niğde, Şırnak, Tunceli
5. Küme (n=14)	*Adana*, Amasya, Diyarbakır, Erzincan, Erzurum, Giresun, *Ordu*, Samsun, Şanlıurfa, Sivas, Tokat, *Trabzon*, Van, Yozgat

■ Hiyerarşik Olmayan Kümeleme Yöntemlerinin Karşılaştırılması

Türkiye'deki illerin hayvancılık istatistikleri bakımından sınıflandırılmasına yönelik uygulanan hiyerarşik olmayan kümeleme yöntemleri sonucu belirlenen ve küme üyelik kodları kullanılarak yapılan diskriminant analizi DSO değerleri özet olarak Çizelge 4.14'te verilmiştir.

Çizelge 4.14 Kümeleme yöntemlerine ilişkin doğru sınıflandırma oranları

Küme Sayısı	K-Ortalamalar DSO(%)	Fuzzy DSO(%)
2	88,8	86,4
3	87,6	87,7
4	87,6	82,7
5	86,3	90,1
6	86,3	88,1

Çizelge 4.14'deki bulgulara göre, illerin hayvancılık istatistiklerine göre k-Ortalamalar kümeleme yönteminde 2 kümeye ayrıldığında, diskriminant analizi DSO değerleri en yüksek düzeydedir. Bununla beraber fuzzy kümeleme yönteminde 5 kümeye ayrıldığında en yüksek DSO değeri oluşmuştur. Hiyerarşik olmayan kümeleme yöntemleri kendi aralarında doğru sınıflandırılma oranlarına göre karşılaştırıldığında ise fuzzy kümeleme yönteminin DSO değeri 5 küme yapısında en yüksek değere sahiptir. Dolayısıyla bu bilgiler ışığında hiyerarşik olmayan kümeleme yöntemleri arasında fuzzy kümeleme yöntemi 5 küme yapısı ile en uygun kümeleme yöntemidir.

4.2 Faktör Analizine İlişkin Bulgular

Bu bölümde Türkiye'deki illerin hayvancılık istatistikleri bakımından gelişmişliğini sıralamak için hayvan varlığı ve hayvansal üretimden oluşan toplam 29 değişkene faktör analizi uygulanmıştır. Burada amaç illerin hayvancılık istatistiklerine göre bir gelişmişlik endeks değeri elde etmektir.

Bu doğrultuda, il bazında 29 değişkene ait verilerin standardizasyonu yapılarak faktör analizi uygulanmıştır. Bununla beraber kişi başına yıllık üretilen hayvan varlığı ve hayvansal üretim ham verileri ile de hayvancılık gelişmişlik endeksi elde edilmiştir. Orijinal faktör yüklerinden bilgi elde edilmesi zor olabilmektedir. Faktörleri daha kolay bir şekilde görüntülemek adına faktör rotasyonu yapılmıştır. Rotasyon için döndürme teknikleri denenmiş ve en uygun tekniğin varimax rotasyonu olduğuna karar verilmiş ve verilere uygulanmıştır.

Uygulanan faktör analizine ilişkin faktör yükleri, KMO ve Bartlett's test sonuçları, özdeğeri (eigen value) Çizelge 4.15'te sunulmuştur. Faktör analizi sonucunda, faktör yükleri 0,067 ile 0,905 arasında değişmekte olup, KMO (KMO=0,662) ve Bartlett's test (ki-kare=6274,439; $p<0,05$) sonuçları faktör analizinin uygulanabilirliğini ve örneklem yeterliliğini ortaya koymuştur. Bununla birlikte ilk faktörün özdeğerinin 11,57 olduğu ve bu faktörün toplam varyansın %51,13'ünü açıkladığı görülmektedir.

Çizelge 4.15 Faktör analizi sonuçları

Değişken	Kod	Faktör Yükleri
Kültür Sığır Sayısı	X1	0,764
Sağılan Kültür Sığır Sayısı	X2	0,790
Melez Sığır Sayısı	X3	0,899
Sağılan Melez Sığır Sayısı	X4	0,905
Yerli Sığır Sayısı	X5	0,885
Sağılan Yerli Sığır Sayısı	X6	0,872
Manda Sayısı	X7	0,467
Sağılan Manda Sayısı	X8	0,475
Merinos Koyunu Sayısı	X9	0,625
Sağılan Merinos Koyunu Sayısı	X10	0,878
Yerli ve Diğer Koyun Sayısı	X11	0,885
Sağılan Yerli ve Diğer Koyunların Sayısı	X12	0,902
Keçisi Sayısı (Tiftik+Kıl)	X13	0,534
Sağılan Keçi Sayısı	X14	0,526
Et Tavuğu Sayısı	X15	0,436
Yumurta Tavuğu Sayısı	X16	0,432
Diğer Kümes Hayvanların Sayısı	X17	0,342
Arı Kovanı Sayısı	X18	0,788
Kültür Sığır Sütü	X19	0,900
Melez Sığır Sütü	X20	0,874
Yerli Sığır Sütü	X21	0,445
Manda Sütü	X22	0,358
Merinos Koyun Sütü	X23	0,506
Yerli ve Diğer Koyun Sütü	X24	0,310
Keçi (Tiftik+Kıl) Sütü	X25	0,105
Merinos Koyun Yapağı	X26	0,162
Yerli ve Diğer Koyun Yapağısı	X27	0,099
Tiftik ve Keçi Kılı	X28	0,097
Bal ve Balmumu	X29	0,067
Öz Değer (Eigen Value)	**11,57**	
Varyans Açıklama Oranı (%)	**51,13**	
KMO Değeri (Kaiser-Meyer-Olkin)	**0,662**	
Bartlett's Test	**Ki-Kare** **SD** **P**	**6274,439** **406** **0,00***

P: Bartlett's testi olasılık değeri *: p<0,05

Elde edilen faktör yükleri ağırlık olarak kabul edilmiş olup her bir değişkenle çarpılarak hayvancılık gelişmişlik endeksi elde edilmiştir. Bu çerçevede, Çizelge 4.16 ve Çizelge 4.17'de faktör analizi yardımıyla her bir il için hayvancılık gelişmişlik endeksi olarak adlandırabileceğimiz ve her ili temsil eden sayısal değerler elde edilmiş ve bu değerlere göre de bir sıralama yapılmıştır. Bu sıralama, gerek illerin kişi başına yıllık üretilen hayvancılık istatistikleri ve gerekse bu istatistiklerin standardizasyonu alınarak elde edilen veri setinden oluşmaktadır.

İllerin hayvancılık istatistikleri bakımından gelişmişliğini gösteren Çizelge 4.16'da standartlaştırılmış veri sıralamasına bakıldığında en gelişmiş ilin Ardahan ili olarak görüldüğü, ilk 10 ilin ise çoğunluğunu doğu illeri oluşturmaktadır. Söz konusu bu iller en baştan sırasıyla Ardahan, Iğdır, Muş, Karaman, Kars, Tunceli, Burdur, Ağrı, Siirt Bitlis illeridir. Oluşan hayvancılık gelişmişlik endeksinde ilk sıradaki ilin Ardahan ili olması ve hiyerarşik kümeleme analizlerinden TekBKY, TamBKY ve McQuitty kümeleme analizlerinde Ardahan ilinin tek başına bir küme oluşturması dikkat çekmektedir. Hayvancılık gelişmişlik endeksinde son sıradaki iller ise sırasıyla sondan başlayarak İstanbul, Kocaeli, Yalova, Hatay, Gaziantep, Bursa, Trabzon, Antalya, Zonguldak, Rize illerinden oluşmuştur. Bu sıralamada ise daha çok sanayi açısından gelişmiş iller yer almaktadır.

Faktör yükleri ile hesaplanan, endeks değerleri ile oluşturulan bu sıralama yıllık kişi başına üretilen ham veriler ile yapıldığında ilk sıradaki il değişmemektedir. Ardahan ili hem standartlaştırılmış veriler ile yapılan sıralamada hem de yıllık kişi başına üretilen ham verilerde ilk sırada yer almaktadır. Çizelge 4.17'de ham veriler ile yapılan sıralama verilmiştir. Söz konusu bu iller en baştan sırasıyla Ardahan, Kars, Burdur, Bayburt, Erzurum, Kastamonu, Aksaray, Niğde, Çankırı, Iğdır illeridir. Standartlaştırılmış veriler ile oluşan sıralamaya göre 6 ilin değiştiği gözlemlenmektedir. Son sıradaki iller ise sondan sırasıyla İstanbul, Yalova, Kocaeli, Ankara, Rize, Antalya, Mersin, Bursa, Hatay ve Şırnak illerinden oluşmaktadır.

Çizelge 4.16 İllerin standartlaştırılmış verilere göre hayvancılık gelişmişlik sıralaması

Sıra	İl Adı	Endeks	Sıra	İl Adı	Endeks
		Standartlaştırılmış Veri			
1	Ardahan	34,14	42	Elazığ	-1,35
2	Iğdır	19,94	43	Çorum	-1,66
3	Muş	19,29	44	Kütahya	-1,69
4	Karaman	16,50	45	Bartın	-1,93
5	Kars	15,42	46	Uşak	-2,55
6	Tunceli	15,19	47	Batman	-2,57
7	Burdur	14,26	48	Aydın	-2,98
8	Ağrı	11,82	49	Mardin	-3,00
9	Siirt	10,28	50	Samsun	-3,01
10	Bitlis	9,00	51	Giresun	-3,80
11	Bayburt	8,75	52	Denizli	-4,21
12	Afyonkarahisar	6,79	53	Ordu	-4,73
13	Balıkesir	5,16	54	Manisa	-4,87
14	Bolu	4,71	55	Adıyaman	-4,94
15	Kırşehir	4,70	56	Şanlıurfa	-5,00
16	Erzurum	4,41	57	Kırıkkale	-5,41
17	Aksaray	4,32	58	Düzce	-5,66
18	Bingöl	4,26	59	Kayseri	-5,78
19	Hakkâri	4,24	60	Bilecik	-5,81
20	Çankırı	3,79	61	Nevşehir	-6,13
21	Kastamonu	3,78	62	Kahramanmaraş	-6,39
22	Amasya	3,35	63	Karabük	-6,45
23	Eskişehir	3,02	64	Tekirdağ	-6,49
24	Niğde	2,81	65	Malatya	-6,81
25	Erzincan	2,78	66	Mersin	-7,17
26	Sinop	2,33	67	Ankara	-7,51
27	Şırnak	2,09	68	İzmir	-7,55
28	Yozgat	1,60	69	Osmaniye	-7,57
29	Çanakkale	1,49	70	Adana	-7,70
30	Tokat	1,48	71	Sakarya	-7,73
31	Sivas	1,05	72	Rize	-7,80
32	Konya	0,75	73	Zonguldak	-8,01
33	Kilis	0,66	74	Antalya	-8,03
34	Edirne	0,33	75	Trabzon	-8,33
35	Diyarbakır	-0,12	76	Bursa	-8,60
36	Kırklareli	-0,41	77	Gaziantep	-9,11
37	Gümüşhane	-0,48	78	Hatay	-9,45
38	Van	-0,79	79	Yalova	-9,69
39	Isparta	-0,99	80	Kocaeli	-10,49
40	Artvin	-1,06	81	İstanbul	-11,54
41	Muğla	-1,13			

Çizelge 4.17 İllerin ham verilere göre hayvancılık gelişmişlik sıralaması

Sıra	İl Adı	Endeks	Sıra	İl Adı	Endeks
		Ham Veri			
1	Ardahan	3157,45	42	Diyarbakır	280,27
2	Kars	1555,74	43	İzmir	246,97
3	Burdur	1285,18	44	Malatya	242,20
4	Bayburt	1204,06	45	Adıyaman	239,00
5	Erzurum	1046,93	46	Bitlis	238,52
6	Kastamonu	898,24	47	Kayseri	236,04
7	Aksaray	851,03	48	Samsun	231,64
8	Niğde	819,35	49	Giresun	217,72
9	Çankırı	787,46	50	Kahramanmaraş	211,66
10	Iğdır	786,62	51	Bilecik	210,38
11	Sivas	658,79	52	Eskişehir	207,48
12	Muş	638,52	53	Tekirdağ	196,20
13	Yozgat	632,36	54	Trabzon	194,31
14	Afyonkarahisar	602,58	55	Batman	193,39
15	Kırşehir	567,62	56	Kırıkkale	191,61
16	Gümüşhane	563,83	57	Ordu	187,57
17	Edirne	550,78	58	Manisa	182,42
18	Çanakkale	543,26	59	Sakarya	177,08
19	Konya	535,05	60	Osmaniye	173,20
20	Kırklareli	529,99	61	Van	170,73
21	Balıkesir	506,61	62	Hakkâri	168,58
22	Bolu	503,67	63	Mardin	156,16
23	Ağrı	501,56	64	Karabük	154,60
24	Çorum	496,55	65	Şanlıurfa	144,92
25	Uşak	491,18	66	Siirt	141,96
26	Erzincan	486,53	67	Zonguldak	141,87
27	Tokat	439,94	68	Adana	132,92
28	Aydın	433,48	69	Kilis	120,04
29	Amasya	433,47	70	Düzce	108,47
30	Tunceli	432,01	71	Gaziantep	106,73
31	Kütahya	412,02	72	Şırnak	98,97
32	Bingöl	411,40	73	Hatay	93,21
33	Karaman	410,80	74	Bursa	92,28
34	Artvin	407,09	75	Mersin	89,97
35	Isparta	380,83	76	Antalya	82,05
36	Denizli	375,72	77	Rize	73,93
37	Sinop	373,92	78	Ankara	55,51
38	Elazığ	358,20	79	Kocaeli	53,10
39	Muğla	318,76	80	Yalova	48,55
40	Bartın	317,36	81	İstanbul	7,03
41	Nevşehir	295,11			

Araştırmada uygulanan kümeleme yöntemlerinden Ward kümeleme yöntemi incelendiğinde, illerin 2 kümeye ayrıldığı görülmektedir. 2. kümeyi Ağrı, Ardahan, Bingöl, Bitlis, Eskişehir, Iğdır, Hakkâri, Karaman, Kilis, Muş, Siirt, Şırnak, Tunceli illeri oluşturmuş, 1. küme ise diğer illerden oluşmaktadır. Çizelge 4.16'da illerin faktör yükleri ile elde edilen endeks değerlerine göre yapılan sıralama incelendiğinde ilk 10 ilden 8 ilin Ward kümeleme yöntemiyle oluşan 2. kümedeki 8 il ile birebir aynı iller olduğu görülmektedir. Bu durum faktör analizi ile Ward kümeleme yönteminin bu çalışmada birbirini desteklediği göstermektedir.

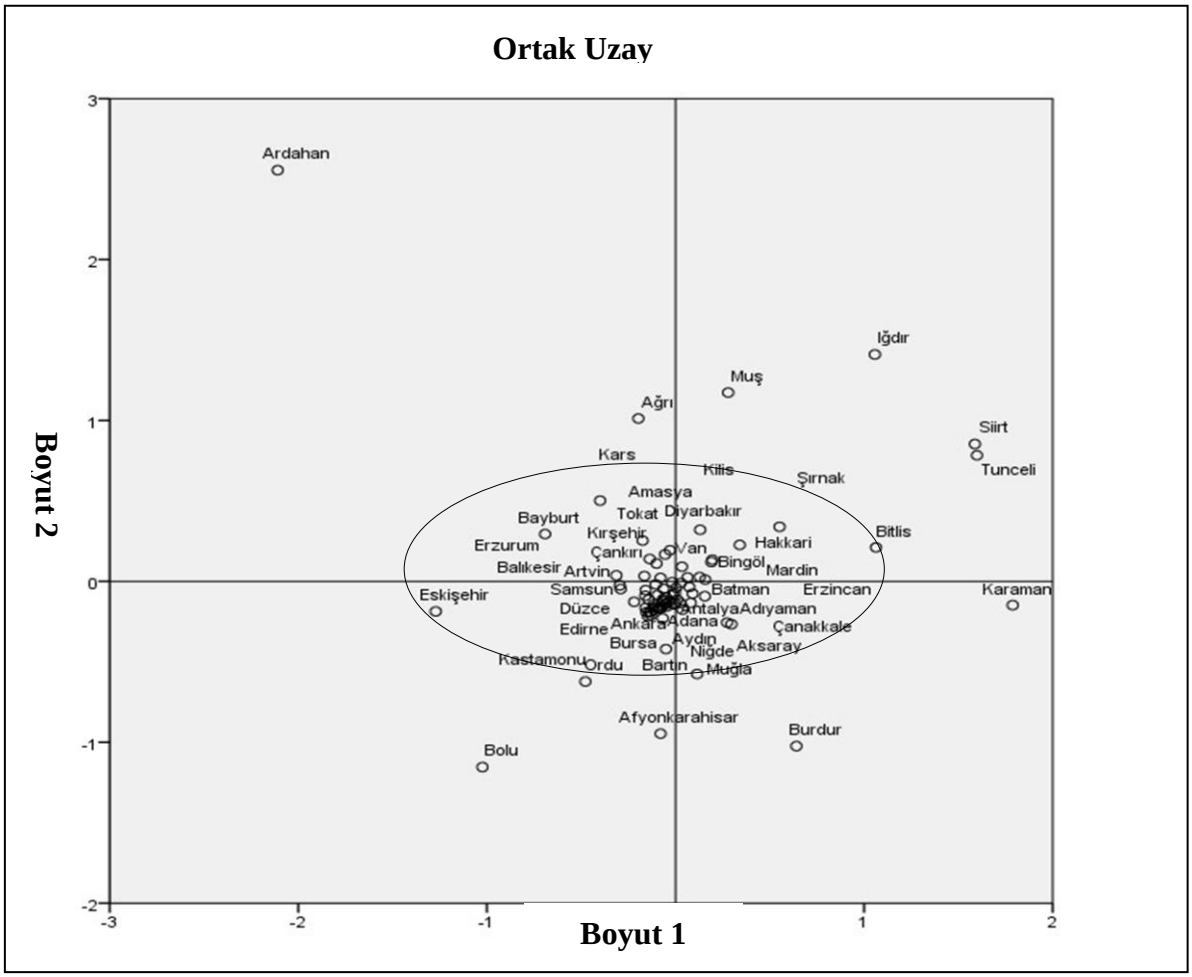
4.3 Çok Boyutlu Ölçekleme Analizine İlişkin Bulgular

Bu bölümde Türkiye'deki illerin hayvancılık istatistiklerini oluşturan hayvan varlığı ve hayvansal üretimden oluşan toplam 29 değişkene ÇBÖ (çok boyutlu ölçekleme) analizi uygulanmıştır. Çok boyutlu ölçekleme analizi ile illerin hayvancılık istatistiklerine göre iki boyutlu bir uzayda birimlerin konumları grafiksel olarak belirlenmiş ve daha geçerli bir değerlendirme elde edilmesi hedeflenmiştir.

Çok boyutlu ölçekleme teknikleri metrik ve metrik olmayan ÇBÖ şeklinde iki yöntemden oluşmaktadır. Hangi yöntemin seçileceğini belirlemede veri tipine bakılmaktadır. Eğer veriler isim, kategorik ya da skor tabanlı veriler ise metrik olmayan ÇBÖ teknikleri kullanılır. Metrik ÇBÖ tekniklerinde ise veriler oranlı ya da eşit aralıklı ölçek ile elde edilmiş olmalıdır. Bu bilgiler ışığında çalışmada verilerin standardizasyonu yapılarak metrik ÇBÖ tekniklerinden yararlanılmıştır.

ÇBÖ analizi ile illerin iki boyutlu uzay haritası için koordinat değerleri ve boyut sayısının uygunluğunu ölçen stress istatistik değeri Çizelge 4.18'de, illerin iki boyutlu uzay haritası (ÇBÖ grafiği) ise Şekil 4.7'de sunulmuştur. Çizelge 4.18'deki stress istatistiği 0.04 gibi sıfıra yakın bir değere sahiptir ve iller iki boyutlu uzayda iyi bir çözüm göstermektedir. Bir diğer ifadeyle stress istatistiğinin 0,025 ile 0,05 arasında bir değer alması orijinal uzaklıklar ile haritada görselleştirilen uzaklık değerleri arasında çok iyi bir uyumun olduğunu belirtmektedir. Bu durumda elde edilen sonuçlar elimizde bulunan veri kümesini iyi bir şekilde temsil etmektedir. Çizelge 4.18'deki R^2 değeri

%98 gibi yüksek bir değere sahiptir. Bu değer, ÇBÖ modelinin verileri ne kadar temsil ettiğini yansıtmaktadır. Bu bilgiler ışığında bu çalışmada yapılan ÇBÖ analizin güvenilir olduğu söylenebilir.



Şekil 4.7 İllerin iki boyutlu uzay haritası (ÇBÖ grafiği)

Şekil 4.7'deki ÇBÖ grafiğinde, illerin büyük bir çoğunluğunun orijin etrafında tek bir küme oluşturdukları görülmektedir. Bu kümedeki illerin, hayvan varlığı ve hayvansal üretim bakımından benzer iller olduğu yorumu yapılabilir. Ancak; Ardahan, Ağrı, Muş, Iğdır, Siirt, Tunceli, Kars, Kilis, Şırnak, Bitlis, Karaman, Bolu, Burdur ve Afyonkarahisar diğer illerden belirgin bir şekilde ayrılmaktadır. Bu durum söz konusu bu illerin, küme içerisinde kalan illere göre hayvancılık göstergeleri bakımından farklı iller olduğunu göstermektedir. Diğer bir ifade ile küme dışında kalan bu illerin hayvansal üretim ve hayvan varlığı bakımından etkili iller olduğu sonucu elde edilir.

Çizelge 4.18 İllerin iki boyutlu uzay haritası için koordinat değerleri

Sıra	İl Adı	Boyut1	Boyut2	Sıra	İl Adı	Boyut1	Boyut2
1	Adana	-0,047	-0,157	43	Karabük	-0,144	-0,113
2	Adıyaman	-0,013	-0,081	44	Karaman	1,788	-0,148
3	Afyon	-0,079	-0,946	45	Kars	-0,400	0,502
4	Ağrı	-0,197	1,012	46	Kastamonu	-0,295	-0,025
5	Aksaray	0,296	-0,266	47	Kayseri	-0,066	-0,111
6	Amasya	-0,029	0,193	48	Kilis	0,131	0,321
7	Ankara	-0,159	-0,165	49	Kırıkkale	-0,094	-0,084
8	Antalya	-0,011	-0,143	50	Kırklareli	0,081	-0,134
9	Ardahan	-2,110	2,556	51	Kırşehir	-0,137	0,139
10	Artvin	-0,165	0,033	52	Kocaeli	-0,131	-0,191
11	Aydın	-0,048	-0,139	53	Konya	-0,064	-0,047
12	Balıkesir	-0,314	0,037	54	Kütahya	-0,003	-0,073
13	Bartın	-0,051	-0,421	55	Malatya	-0,069	-0,123
14	Batman	0,064	0,024	56	Manisa	-0,068	-0,230
15	Bayburt	-0,692	0,294	57	Mardin	0,128	0,027
16	Bilecik	-0,035	-0,124	58	Mersin	0,026	-0,135
17	Bingöl	0,196	0,134	59	Muğla	0,115	-0,576
18	Bitlis	1,063	0,211	60	Muş	0,279	1,173
19	Bolu	-1,023	-1,154	61	Nevşehir	-0,071	-0,151
20	Burdur	0,641	-1,024	62	Niğde	0,273	-0,256
21	Bursa	-0,118	-0,177	63	Ordu	-0,478	-0,623
22	Çanakkale	0,156	-0,092	64	Osmaniye	-0,056	-0,142
23	Çankırı	-0,079	0,022	65	Rize	-0,129	-0,211
24	Çorum	-0,052	-0,098	66	Sakarya	-0,154	-0,191
25	Denizli	0,000	-0,137	67	Samsun	-0,291	-0,048
26	Diyarbakır	0,034	0,091	68	Şanlıurfa	-0,018	-0,006
27	Düzce	-0,160	-0,090	69	Siirt	1,587	0,853
28	Edirne	-0,219	-0,127	70	Sinop	-0,174	0,253
29	Elazığ	0,003	-0,040	71	Şırnak	0,552	0,339
30	Erzincan	0,158	0,011	72	Sivas	0,091	-0,076
31	Erzurum	-0,100	0,111	73	Tekirdağ	-0,100	-0,162
32	Eskişehir	-1,271	-0,187	74	Tokat	-0,055	0,167
33	Gaziantep	-0,077	-0,167	75	Trabzon	-0,117	-0,161
34	Giresun	-0,104	-0,019	76	Tunceli	1,599	0,783
35	Gümüşhane	-0,158	-0,054	77	Uşak	0,037	-0,176
36	Hakkâri	0,341	0,226	78	Van	0,191	0,121
37	Hatay	-0,089	-0,169	79	Yalova	-0,137	-0,191
38	Iğdır	1,057	1,410	80	Yozgat	0,028	-0,010
39	Isparta	0,073	-0,035	81	Zonguldak	-0,116	-0,144
40	İstanbul	-0,146	-0,216	**Stress**		0,04260	
41	İzmir	-0,083	-0,167				
42	K.Maraş	0,009	-0,113	**R^2**		0,98553	

Çizelge 4.18'deki koordinat değerleri ile Şekil 4.7'deki ÇBÖ grafiği, 1. boyuta göre en benzer illerin Elazığ ve Kütahya, en çok farklılık gösteren illerin ise Ardahan ve Bolu illerinin olduğunu göstermektedir. 2. boyuta göre ise birbirine en çok benzeyen iller Çankırı ve Mardin illeri iken en farklı olan iller ise Ardahan ve Karaman illeridir.

Şekil 4.7'deki ÇBÖ grafiğinde diğer illere göre en uzak mesafedeki ilin Ardahan ili olduğu görülmektedir. Bu araştırmada faktör analizi sonucu elde edilen, illerin hayvancılık istatistikleri bakımından gelişmişliğini gösteren Çizelge 4.16'daki sıralamada, en gelişmiş ilin Ardahan ili olması, bu iki analizin birbiri ile benzer sonuçlar verdiğini göstermektedir. Aynı zamanda hiyerarşik kümeleme yöntemlerinden TekBKY, TamBKY ve McQuitty kümeleme yöntemlerinde, Ardahan ili tek başına bir küme oluşturmaktadır. Araştırmada uygulanan kümeleme yöntemlerinden Ward kümeleme yönteminde 2. kümeyi Ağrı, Ardahan, Bingöl, Bitlis, Eskişehir, Iğdır, Hakkâri, Karaman, Kilis, Muş, Siirt, Şırnak, Tunceli illeri oluşturmuş, 1. küme ise diğer illerden oluşmaktadır. Çizelge 4.16'da faktör analizi ile elde edilen ve illerin endeks değerlerine göre yapılan sıralamada ilk 10 il en baştan sırasıyla Ardahan, Iğdır, Muş, Karaman, Kars, Tunceli, Burdur, Ağrı, Siirt Bitlis illeridir. ÇBÖ grafiğinde ise; Ardahan, Ağrı, Muş, Iğdır, Siirt, Tunceli, Kars, Kilis, Şırnak, Bitlis, Karaman, Bolu, Burdur ve Afyonkarahisar illeri diğer illere göre farklılık gösteren iller olarak ayrılmaktadır. Bu durum; faktör analizi, Ward kümeleme yöntemi ve ÇBÖ analizinden benzer sonuçların elde edildiğini göstermektedir. Diğer bir ifade ile bu üç çok değişkenli analiz yönteminin sonuçları bu çalışmada birbirini desteklemektedir.

5. TARTIŞMA ve SONUÇ

Çok değişkenli analiz teknikleri kullanılmasıyla beraber büyük ve karmaşık veri setlerini analiz etmek kolaylaşmıştır. Günümüzde sağlık, ekonomi, eğitim gibi birçok bilim dalı çok değişkenli istatistik tekniklerinden yararlanmaktadır. Bunun başlıca sebebi tek değişkenli istatistiksel yöntemlerin yetersiz kalması ve uymak zorunda olduğu varsayımlar söylenebilir.

Hayvancılık bir ülkenin en temel konularından olan tarım sektörünün bir kolu olarak görülmektedir. Hayvan varlığı ve hayvansal üretime yönelik temel bilimler içerisinde istatistik vazgeçilmez bir konu olarak yerini almıştır. Bu bilgiler çerçevesinde, Türkiye'deki illerin hayvan varlığı ve hayvansal üretim istatistikleri bakımından çok değişkenli istatistiksel teknikler ile incelenmesi amacı taşıyan bu çalışmada TÜİK'den elde edilen 2018 yılına ait hayvancılık istatistikleri verileri çerçevesinde 29 değişken kullanılmıştır. Bu değişkenlerin 18'i hayvan varlığı 11'i ise hayvansal üretim ile ilgilidir. Literatür incelendiğinde hayvancılık ve istatistik ile ilgili ulusal ölçekte birçok çalışma mevcuttur.

Bu çalışmada öncelikle Türkiye'deki illerin hayvancılık istatistikleri bakımından sınıflandırılmasını belirlemek için kümeleme analizi yapılmıştır. Bununla beraber hiyerarşik ve hiyerarşik olmayan kümeleme analizlerinden yararlanılmış ve hiyerarşik yöntemlerden; TekBKY, TamBKY, McQuitty ve Ward kümeleme yöntemleri, hiyerarşik olmayan yöntemlerden ise k-Ortalamalar ve fuzzy (bulanık) kümeleme yöntemleri uygulanmıştır. Uygulanan bu kümeleme yöntemlerinden hiyerarşik kümeleme yöntemlerini kendi aralarından karşılaştırabilmek için benzerlik, uzaklık ölçüleri ve diskriminant analizi DSO (doğru sınıflandırma oranları) değerleri kullanılmıştır. Hiyerarşik yöntemlerin tamamında 2 küme sayısında en yüksek uzaklık düzeyi ve en düşük benzerlik düzeyine ulaşılmıştır. Aynı zamanda küme sayılarının eşit olduğu durumda ve 2 küme yapısında en yüksek uzaklık düzeyi ve en düşük benzerlik düzeyi Ward kümeleme yöntemiyle elde edilmiştir. Hiyerarşik olmayan kümeleme yöntemlerini karşılaştırabilmek için diskriminant analizi DSO değerleri elde edilmiştir. Oluşan DSO değerlerinde k-Ortalamalar kümeleme yönteminde %88,8 ile 2 küme

yapısında, fuzzy kümeleme yönteminde ise %90,1 ile 5 küme yapısında en yüksek değere ulaşılmıştır. Buna göre fuzzy kümeleme yönteminde 5 küme yapısında oluşan DSO değerinin daha yüksek olduğu tespit edilmiştir. Bu çerçevede hiyerarşik kümeleme yöntemlerinden Ward kümeleme yönteminde, hiyerarşik olmayan yöntemlerden ise fuzzy kümeleme yönteminde en uygun kümeleme yapısına ulaşıldığı görülmüştür. Kümeleme analizi ile ilgili hayvancılık alanında literatürde birçok çalışmaya rastlamak mümkündür. Nitekim Dinler (2014) illerin hayvancılık bakımından incelenmesinde çok değişkenli istatistik tekniklerinden kümeleme analizi ile bir çalışma ortaya koymuştur. TÜİK tarafından elde edilen hayvancılık istatistikleri ile 44 değişken üzerinden hiyerarşik kümeleme analiz yöntemleri ve uzaklık ölçülerini karşılaştırmalı olarak incelemiştir. Çalışmasında hiyerarşik kümeleme yöntemlerinden ortalama (average), merkezi (centroid), TamBKY (complete), McQuitty, ortanca (median), TekBKY (single) ve Ward kümeleme yöntemlerini kullanmıştır. Bu kümeleme yöntemlerinden Ward kümeleme yöntemi dışındaki yöntemlerin benzer kümeler oluşturduğunu gözlemlemiştir. 2018 yılı itibariyle bu araştırmada Dinler (2014)'ün elde ettiği sonuca benzer olarak, diğer hiyerarşik kümeleme yöntemleri benzerlik gösterirken Ward kümeleme yöntemi diğer hiyerarşik yöntemlere göre farklılık göstermektedir. Kılıç vd. (2012) tarafından yapılan çalışmada, illerin hayvancılık istatistikleri bakımından sınıflandırılması amacıyla çok değişkenli istatistik tekniklerinden fuzzy kümeleme yöntemi kullanılmıştır. Araştırmadaki veriler TÜİK'den elde edilmiş ve hayvansal üretim ve hayvan varlığına ilişkin 26 değişkenden yararlanılmıştır. Fuzzy kümeleme analizi ile küme sayısının 2 olması durumunda en kararlı yapının oluştuğu sonucuna varılmış olup, birinci kümenin 67 ilden, ikinci kümenin ise 14 ilden oluştuğu saptanmıştır. Bu çalışmada 2018 yılı itibariyle fuzzy kümeleme yöntemiyle iller 5 küme yapısı oluşturmuştur. Bu durum Türkiye'deki illerin daha homojen bir yapıdan daha heterojen bir yapıya geldiğini göstermektedir. Diğer bir ifade ile bazı illerde hayvancılığın daha çok azaldığı, illerin hayvancılık benzerlikleri bakımından heterojenleştiği ifade edilebilir.

Kümeleme analizi ile hayvancılık alanında yapılan bir diğer çalışmada (Kılıç 2008) Karayaka ve Bafra koyunları vücut ölçülerine göre sınıflandırılmıştır. Karakaya ve Bafra koyunlarını vücut ölçülerine göre hiyerarşik ve hiyerarşik olmayan kümeleme

yöntemleri ile incelemiştir. Çalışmasında hiyerarşik kümeleme yöntemlerinden TekBKY, TamBKY ve OrtBKY'ni, hiyerarşik olmayan kümeleme yöntemlerinden ise k-Ortalamalar, medoid ve fuzzy kümeleme yöntemlerini uygulamıştır. Karakaya ve Bafra koyunlarının vücut ölçülerine göre yapılan kümeleme analizine yönelik olarak hiyerarşik kümeleme yöntemlerinden TekBKY ve OrtBKY'de Karayaka koyunlarının tamamına yakınının ve Bafra koyunlarının da çok büyük bir bölümünün tek kümede toplandığı sonucuna ulaşmıştır. TamBKY ve hiyerarşik olmayan kümeleme yöntemlerinden k-Ortalamalar kümeleme yöntemi ve medoid kümeleme yöntemine göre de Karayaka koyunları 3 ve Bafra koyunları ise 7 kümede toplandığı görülmektedir. Fuzzy kümeleme yönteminin sonucunda ise Karayaka koyunları 2 kümeye ve Bafra koyunları ise 4 kümeye ayrılmıştır. Çok değişkenli istatistik tekniklerinden kümeleme analizi ile Ankara keçisinin tiftik özellikleri yönünden sınıflandırılması amacıyla Çelik ve Akçapınar (2006) tarafından bir çalışma gerçekleştirilmiş ve çalışmada tek bağlantı kümeleme yöntemini kullanılmıştır. Ankara (tiftik) keçisinin 9 yıllık kayıtlarından elde edilen verilere kümeleme analizi uygulanmıştır ve Ankara keçisi sürüsünün kümeleme analizi yöntemiyle 4-7 kümeye ayrıldığı sonucuna ulaşılmıştır.

Bu çalışmada Türkiye'de illerin hayvancılık göstergeleri bakımından sıralanması için bir hayvancılık gelişmişlik endeksi oluşturmak amacıyla temel bileşenler yöntemi kullanılarak faktör analizi uygulanmıştır. Değişkenlerin faktör yükleri ile çarpılarak toplanması ile elde edilen her il için endeks değerleri illere göre sıralamaya konulmuş ve hayvancılık göstergeleri bakımından en gelişmiş durumdaki iller tespit edilmiştir. Burada, verilere ilişkin bir karşılaştırma yapmak için standartlaştırılmış verilerin yanı sıra kişi başına yıllık üretilen ham veriler de kullanılmıştır. Faktör analizi sonucunda özdeğeri 11,57 olan birinci faktör, toplam varyansın %51,13'ünü açıklamaktadır. Hayvancılık istatistikleri bakımından her iki veri yapısında da oluşan sıralamada Ardahan ili ilk sırayı almıştır. Standartlaştırılmış veriler ile yapılan sıralamada ilk 10 ili en baştan sırasıyla Ardahan, Iğdır, Muş, Karaman, Kars, Tunceli, Burdur, Ağrı, Siirt Bitlis illeri oluştururken son sıradaki iller ise sırasıyla sondan başlayarak İstanbul, Kocaeli, Yalova, Hatay, Gaziantep, Bursa, Trabzon, Antalya, Zonguldak, Rize illerinden oluşmuştur. Aynı zamanda ilk sıradaki bu 10 ilin 8'inin, Ward kümeleme yöntemiyle oluşan 2. kümedeki iller ile birebir aynı iller olduğu tespit edilmiştir. Bu

durum faktör analizi ile Ward kümeleme yönteminin bu çalışmada birbirini desteklediğini göstermektedir. Ham veriler ile yapılan sıralamada ilk 10 il en baştan sırasıyla Ardahan, Kars, Burdur, Bayburt, Erzurum, Kastamonu, Aksaray, Niğde, Çankırı, Iğdır illeri iken son sıradaki iller ise sondan sırasıyla İstanbul, Yalova, Kocaeli, Ankara, Rize, Antalya, Mersin, Bursa, Hatay ve Şırnak illerinden oluşmaktadır. Belirtilmelidir ki literatürde Türkiye'deki illerin hayvancılık göstergeleri bakımından gelişmişlik endekslerinin elde edildiği bir çalışma ile karşılaşılmamıştır. Ancak, farklı alanlarda (ekonomi, turizm, sağlık vb.) illerin ve ilçelerin gelişmişlik endekslerinin oluşturularak sıralandığı pek çok çalışma mevcuttur. Nitekim Albayrak (2012) tarafından yapılan çalışmada Türkiye'de illerin sosyoekonomik gelişmişlik düzeylerini belirlemek amacıyla çok değişkenli istatistik tekniklerinden ve faktör analizinden yararlanılmıştır. Faktör analiziyle, her bir il için sosyoekonomik gelişmişlik endeksi olarak tanımlanabilecek sayısal değerler elde edilmiş, sıralamada ilk üç il olarak İstanbul, Ankara ve İzmir illeri, sondan başlayarak son üç il ise Hakkâri, Ağrı, Şırnak illeri belirlenmiştir. Seçilmiş ve Sarı (2010) tarafından, Türkiye'de illerin turizm gelişmişlik endeksinin oluşturulmasına yönelik bir çalışma gerçekleştirilmiş ve çok değişkenli istatistik tekniklerinden temel bileşenler analizi kullanılmıştır. Elde edilen endeks değerlerinin sıralaması sonucu Türkiye'de turizm alanında popüler olan illerden Antalya, İstanbul, Muğla, İzmir, Aydın ve Balıkesir ilk altı sırada yer almıştır.

Bu çalışmada son olarak, illerin hayvancılık istatistiklerine göre grafiksel olarak bir değerlendirme elde edilmesi amacıyla çok değişkenli istatistik tekniklerinden ÇBÖ (çok boyutlu ölçekleme) analizi kullanılmıştır. ÇBÖ grafiğine göre 1. boyutta yüksek benzerlik gösteren iller Elazığ ve Kütahya illeri iken en fazla farklılığı Ardahan ve Bolu illeri göstermektedir. 2. boyuta göre ise birbirine en çok benzeyen iller Çankırı ve Mardin illeri iken en farklı olan iller ise Ardahan ve Karaman illeridir. Ayrıca ÇBÖ grafiğine göre Ardahan, Ağrı, Muş, Iğdır, Siirt, Tunceli, Kars, Kilis, Şırnak, Bitlis, Karaman, Bolu, Burdur ve Afyonkarahisar diğer illerden belirgin bir şekilde ayrılmaktadır. Ulusal ölçekte literatürde ÇBÖ analizi ile ilgili yapılan birçok çalışma mevcuttur. Hayvancılık alanında Çelik (2014) Türkiye'deki 81 ilin hayvancılık istatistikleri bakımından incelenmesinde ÇBÖ tekniğini kullanarak bir çalışma ortaya koymuştur. Çalışmada TÜİK'den elde edilen sığır sayısı, manda sayısı, koyun sayısı,

keçi sayısı, tavuk sayısı, hindi sayısı, ördek sayısı, at sayısı, eşek sayısı ve katır sayısı verileri kullanılmıştır. Şırnak, Antalya, Siirt ve Bitlis diğer illerden en farklı iller olarak sonuçlanmış, Tunceli, Hakkâri, Van, Şanlıurfa, Siirt, Bitlis ve Şırnak illeri en fazla pozitif etki yapan iller durumunda görülmüştür. Çelik (2014) tarafından yapılan çalışma ile bu çalışmada kullanılan değişkenler açısından farklılık olsa da Tunceli, Hakkâri, Şanlıurfa, Siirt, Bitlis ve Şırnak gibi illerin diğer illere göre farklılık oluşturması açısından iki çalışma da benzerlik göstermektedir. Küçükbaş hayvancılığı koyunculuk açısından incelemek amacıyla Gevrekçi ve vd. (2011) tarafından yapılan çalışmada çok değişkenli istatistiklerden yararlanılmıştır. Çalışmada Batı Anadolu'daki 11 il ÇBÖ ve kümeleme analizleri kullanılarak sınıflandırılmaya çalışılmıştır. Araştırmada, 2003-2008 yıllarına ait TÜİK'den elde edilen koyun sayısı, sağılan koyun sayısı, koyun süt verimi, kesilen koyun-kuzu sayısı, koyun-kuzu eti üretimi (ton), kırkılan koyun sayısı, yapağı üretimi verileri kullanılarak Batı Anadolu illeri dört ana grup oluşturmuştur. Bu gruplar Afyonkarahisar - Balıkesir; İzmir - Manisa; Bursa - Çanakkale – Denizli - Kütahya - Uşak ve Aydın - Muğla şeklinde oluşmuştur.

Hayvancılık istatistikleri ile ilgili 2010 yılından sonra yapılan çalışmalar incelendiğinde ve 2019 yılı itibariyle yapılan bu çalışma ile karşılaştırıldığında, Türkiye'deki illerin daha heterojen bir yapıya sahip olduklarını göstermiştir. Türkiye'deki 81 ilin hayvan varlığı ve hayvansal üretimine ait 2018 yılı hayvancılık istatistikleri verileri çerçevesinde 29 değişkenden oluşan bu çalışmada çok değişkenli istatistik tekniklerinden kümeleme analizi, faktör analizi ve çok boyutlu ölçekleme analizi yapılmış olup elde edilen bilgilerin literatüre katkı sağlayacağı değerlendirilmektedir. Bununla birlikte ilerleyen yıllarda yapılacak çalışmalar, bu çalışma ile karşılaştırıldığında hayvancılık bakımından illerdeki gelişmişlik düzeylerinin nasıl yer değiştiği anlaşılabilecek olup bu durum ilgili kişi, yönetici ya da kurumlara önemli bilgiler sunabilecektir.

6. KAYNAKLAR

Akat, Y. (2007). Ülkelerin Askeri Benzerliklerine Göre Kümeleme Analizi Yardımıyla Sınıflandırılması. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Akdeniz, F. (2007). Olasılık ve İstatistik. 13. Baskı, Nobel Kitabevi. Adana.

Aktaş, C. ve Yılmaz, V. (2003). Çoklu Bağıntılı Modellerde Liu ve Ridge Regresyon Kestiricilerinin Karşılaştırılması. *Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi*, **2:** 189-194.

Albayrak, A. S. (2003). Türkiye'de İllerin Sosyo-Ekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Çok Değişkenli İstatistik Yöntemlerle İncelenmesi. Yayınlanmamış Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü. İstanbul.

Albayrak, A. S. (2006). Uygulamalı Çok Değişkenli İstatistik Teknikleri. Asil Yayın Dağıtım. Ankara.

Albayrak, A. S. (2012). Türkiye'de İllerin Sosyoekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, **1(1)**: 153-177.

Aldenderfer, M. S. and Blashfield, R. K. (1984). Cluster Analysis. Beverly Hills: Sage Publications. London.

Alpar, R. (2011). Uygulamalı Çok Değişkenli İstatistiksel Yöntemler, Detay Yayıncılık. Ankara.

Atalay, M. (1984). Çoklu Doğrusal Bağıntının Ölçülmesi Sorunu. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, **(6):** 97-106.

Atan, M., Göksel, A., Karpat, G. (2002). Üniversite Öğrencilerinin Başarılarını Etkileyen Faktörlerin Çok Değişkenli İstatistiksel Analiz Yöntemleri ile Tespiti. XI. Eğitim Bilimleri Kongresi, Yakın Doğu Üniversitesi, Kıbrıs.

Baldemir, E. (1997). Çoklu Doğrusal Bağlantısının Tespit Metotları Ve Türkiye İçin Enflasyon Uygulaması. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, **1:** 173-191.

Berkhin, P. (2004). Survey Of Clustering Data Mining Techniques. Accrue Software: San Jose, CA.

Bezdek, J.C. (1973). Fuzzy Mathematics in Pattern Classification. PhD Thesis, Cornell University, Ithaca, New York.

Bezdek, J.C (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers. Norwell, MA, USA (ISBN:0306406713).

Bulam, N. (2005). Faktörleri Belirlemede Kullanılan Yöntemler ve Bir Uygulama. Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Samsun.

Burmaoğlu, S. (2011). Satınalma Alternatiflerinin Çok Değişkenli İstatistiksel Yöntemlerle Belirlenmesi: Keskin Nişancı Tüfekleri Üzerine Bir Uygulama. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, **15(1):** 369-382.

Büyüköztürk, Ş. (2011). Sosyal Bilimler İçin Veri Analizi El Kitabı. Pegem Akademi, Ankara.

Çokluk, Ö., Büyüköztürk, Ş. ve Şekercioğlu, G. (2010). Sosyal Bilimler İçin Çok Değişkenli İstatistik: SPSS ve Lisrel Uygulamaları. Pegem Akademi. Ankara.

Cebeci, Z. ve Yıldız, F. (2015). Bulanık C-Ortalamalar Algoritmasının Farklı Küme Büyüklükleri için Hesaplama Performansı ve Kümeleme Geçerliliğinin Karşılaştırılması. 9. Ulusal Zootekni Bilim Kongresi, Bildiriler Kitabı, s. 227-239. Konya.

Cengiz, D. ve Kılınç, B. (2007). Faktör Analiz İle 2006 Dünya Kupasına Katılan Takımların Sıralamasının Belirlenmesi. *Marmara Üniversitesi İ.İ.B.F. Dergisi*, **23(2).**

Çakmak, Z. (1999). Kümeleme Analizinde Geçerlilik Problemi Ve Kümeleme Sonuçlarının Değerlendirmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, **(3).**

Çelik, B. ve Akçapınar, H. (2006). Ankara Keçisinin Tiftik Özellikleri Yönünden Kümeleme Analizi ile İncelenmesi. *Lalahan Hayvancılık Araştırma Enstitüsü Dergisi*, **46(1).**

Çelik, Ş. (2014). Çok Boyutlu Ölçekleme Analizi ile Hayvancılık Açısından Türkiye'de İllerin Sınıflandırılması. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. **31(4):** 159-164.

Demir, E., Saatçioğlu, Ö., ve İmrol, F. (2016). Uluslararası Dergilerde Yayımlanan Eğitim Araştırmalarının Normallik Varsayımları Açısından İncelenmesi. *Current Research in Education, **2*****(3):** 130-148.

Demiralay, M. ve Çamurcu, A. Y. (2005). Cure, Agnes Ve K-Means Algoritmalarındaki Kümeleme Yeteneklerinin Karşılaştırılması. İ*stanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, **8:** 1-18.

Dinler, M. (2014). Kümeleme Analizi Yöntemlerinin Hayvancılık Verilerinde Karşılaştırmalı Olarak İncelenmesi. Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Bingöl.

Doğan, İ. (2002). Kümeleme Analizi İle Seleksiyon. *Turk J Vet Anim Sci*, **26:** 47-53.

Dunn, J.C. (1973). A Fuzzy Relative of the Isodata Process and Its Use in Detecting Compact Well-Separated Clusters. *J. Cybernetics*, **3 (3):** 32–57.

Ergeç, S. M., (1984). Çok Boyutlu Verilerin Bazı İstatistiksel Analiz Yöntemleri ve Uygulamaları. Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Erilli, N. A. (2009). Kümeleme Analizine Bulanık Yaklaşım Algoritmaları ve Uygulamaları. 19 Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Samsun.

Everitt, B. (1974). Cluster Analysis. London: Heinemann Educational Books Ltd.

Friedman, H. P. and Rubin, J. (1967). On Some İnvariant Criteria For Grouping Data. Journal Of The American Statistical Association, **62(320):** 1159-1178

Gevrekçi, Y., Ataç, F. E., Takma, Ç., Akbaş, Y. ve Taşkın, T. (2011). Koyunculuk Açısından Batı Anadolu İllerinin Sınıflandırılması. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, **17(5):** 755-60.

Gnanadesikan, R. (1997). Methods For Statistical Data Analysis Of Multivariate Observations (Second Edition). United States: John Wiley & Sons, Inc.

Gujurati, D.N. (1988). "Basic Econometrics", John Wiley and Sons, New York, 720 p.

Hair, J. F., R.E. Anderson, R. L. Tatham and W. C. Black (1998). Multivariate Data Analysis. Prentice Hall, New Jersey.

Hair J. F., Black Wc., Babin Bj., Anderson Re. and Tatham Rl. (2006). Multivariate Data Analysis. New Jersey.

Han, J. and Kamber, M. (2006). Data Mining Concepts and Techniques. Morgan Kauffmann Publishers Inc.

Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2014). Multivariate Data Analysis: Pearson new international edition. Essex: Pearson Education Limited. New Jersey.

Hartigan, J. A. (1975). Clustering Algorithms. New York: John Wiley and Sons, Inc.

Hewson, P. J. (2009). Multivariate Statistics with R. *URl:* http://local. disia. unifi. it/rampichini/analisi_multivariata/Hewson_ notes. pdf.

Işık, M. ve Çamurcu, A. Y. (2007). K-Means, K-Medoids Ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, **11:** 31-45.

Johnson, R. A. and Wichern, D. W. (1998). Multivariate Statistical Analysis.

Kalaycı, Ş. (2018). SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri. Dinamik Akademi Yayınları, Ankara.

Kaufman, L. and Rousseeuw, P. J., (1987). Clustering by Means of Medoids. Statistical Data Analysis Based on The L1–Norm and Related Methods, edited by Y. Dodge, North-Holland.

Kaufman, L. and Rousseeuw, P. J. (1990). Partitioning Around Medoids. Finding Groups İn Data: An İntroduction To Cluster Analysis, 68-125.

Kayaalp, T., Yazgan, E., Şahinler, S. (2000). Aşamalı Kümeleme Analizi (Hıerarchıcal Cluster Analysis) Yöntemlerinin Karşılaştırılmalı Olarak İncelenmesi. İstatistik Araştırma Sempozyumu.

Khalaf, K. (2007). Faktör Analizi ve Bir Uygulaması. Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Kılıç, İ. (2008). Canlı Ağırlık Ve Bazı Vücut Ölçüleri Kullanılarak Karayaka Ve Bafra (Sakız X Karayaka G1) Koyunlarının Çok Değişkenli İstatistiksel Yöntemler İle İncelenmesi. Doktora Tezi, Ankara üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.

Kılıç, İ., Emir, O. ve Kılıç, G. (2011). Bulanık Kümeleme Analizi İle Ülkelerin Turizm İstatistikleri Bakımından Sınıflandırılması. *İstatistikçiler Dergisi*: İstatistik ve Aktüerya, **4(1):** 31-38.

Kılıç, İ., Lenger, Ö. F. ve Bozkurt, Z. (2012). Bulanık Kümeleme Analizi ile Türkiye'deki İllerin Hayvancılık istatistikleri Bakımından Sınıflandırılması. *Kocatepe Veteriner Dergisi:* **5(1).**

Kruskal, J. B. (1964). Multidimensional Scaling By Optimizing Goodness Of Fit To A Nonmetric Hypothesis. *Psychometrika*. **29(1):** 1-27.

Kurtuluş, K. (1992). Pazarlama Araştırmaları. İstanbul Üniversitesi İşletme Fakültesi Yayınları. İstanbul.

Marriot, F. H. C. (1971). Practical Problems İn A Method Of Cluster Analysis. *Biometrics,* **27:** 501-514.

Messick, S. J. and Abelson, R. P. (1956). The Additive Constant Problem İn Multidimensional Scaling. *Psychometrika*, **21(1):** 1-15.

Özdamar, K. (2013). Paket Programlar ile İstatistiksel Veri Analizi. Nisan Kitabevi, Ankara.

Pazarlıoğlu, V. M., Emeç, H., Erdoğan, S. (1999). Dokuz Eylül Üniversitesi Öğrencilerinin Yüksek Öğretim Beklenti Değişkenlerinin Faktör Analizi ile İncelenmesi. IV. Ulusal Ekonometri ve İstatistik Sempozyumu Bildirileri, Antalya.

Pektaş, A. O. (2013). SPSS İle Veri Madenciliği. Dikeyeksen Yayın Dağıtım, İstanbul.

Rencher, A. C. (2003). Methods Of Multivariate Analysis. John Wiley and Sons. London.

Sarıman, G. (2011). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means Ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, **15(3):** 192-202.

Seçilmiş, C. ve Sarı, Y. (2010). Türkiye'de İllerin Turizm Gelişmişlik Endeksinin Oluşturulmasına Yönelik Bir Araştırma. *Elektronik Sosyal Bilimler Dergisi*. **9(32):** 117-132.

Sharma, S. (1996). Applied Multivariate Techniques. John Wiley and Sons, Inc. New York.

Shepard, R. N. (1962). The Analysis Of Proximities: Multidimensional Scaling With An Unknown Distance Function. I. *Psychometrika*, **27(2):** 125-140.

Sheth, J. N. (1971). The Multivariate Revolution İn Marketing Research. *Journal Of Marketing*, **35(1):** 13-19.

Shin, K. (1996). SPSS Guide For DOS Version 5 And Windows 6.1. 2. 2th Edit, Irwin, Chicago.

Şahin, M. ve Hamarat, B. (2002). G10, Avrupa Birliği ve OECD Ülkelerinin Sosyo-Ekonomik Benzerliklerinin Fuzzy Kümeleme Analizi İle Belirlenmesi. ODTÜ Uluslarası Ekonomi Kongresi VI, Ankara.

Şimşek Gürsoy, U. T. (2009). Veri Madenciliği ve Bilgi Keşfi. Pegem Akademi, Ankara.

Tabachnick, B. G. ve Fidell, L. S. (2015). Çok Değişkenli İstatistiklerin Kullanımı. (Çeviri Editörü: Baloğlu, M.), Nobel Yayın Dağıtım. Ankara.

Tatlıdil, H. (1992). Uygulamalı Çok Değişkenli İstatistiksel Analiz. Engin Yayınları. Ankara.

Thode, H. C. (2002). Testing For Normality. Marcel Dekker, Inc. United States.

Torgerson, W. S. (1952). Multidimensional Scaling: I. Theory And Method. *Psychometrika*, **17(4):** 401-419.

Ural, A. ve Kılıç, İ. (2011). Bilimsel Araştırma Süreci Ve SPSS İle Veri Analizi. Detay Yayıncılık. Ankara.

Vatansever, M. (2008). Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı Ve Uygulanması. Yüksek Lisans, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Yiğit, E. (2007). Çok Boyutlu Ölçekleme Yöntemlerinin İncelenmesi Ve Bir Uygulama. Çanakkale Onsekiz Mart Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Çanakkale.

ÖZGEÇMİŞ

Adı Soyadı : Süleyman ALDEMİR

Doğum Yeri ve Tarihi : Afyon / 05.08.1987

Yabancı Dili : İngilizce

İletişim (Telefon/e-posta) : 0505 247 4638 / (suleyman.aldemir03@gmail.com)

Eğitim Durumu (Kurum ve Yıl)

Lise : Afyon Fatih Lisesi, (2001-2005)

Lisans : Afyon Kocatepe Üniversitesi, İstatistik Bölümü, (2009-2013)

Çalıştığı Kurum/Kurumlar ve Yıl : Sosyal Yardımlaşma ve Dayanışma Vakfı (2016)