

HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

85341

**n<p BOYUTLU BİYOLOJİK VERİLERDE FARKLI KÜMELEME
YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ**

İrfan ÖZTÜRK

**DOKTORA TEZİ
ZOOTEKNI ANABİLİM DALI**

**T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ**

1999
ŞANLIURFA

85341

HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

n<p BOYUTLU BİYOLOJİK VERİLERDE FARKLI KÜMELEME
YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ

İrfan ÖZTÜRK


Prof. Dr. Abuzer YÜCEL
Fen Bil. Enst. Müdürü

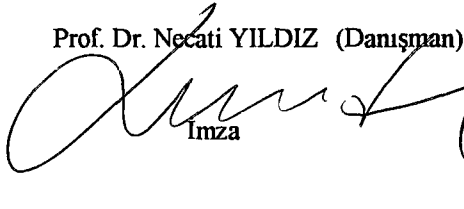
DOKTORA TEZİ
ZOOTEKNI ANABİLİM DALI

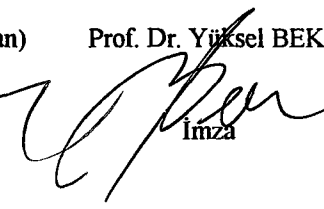
Bu tez 20.10.9/1999 tarihinde aşağıdaki jüri tarafından değerlendirilerek oy birliği /
oy çokluğu ile kabul edilmiştir.

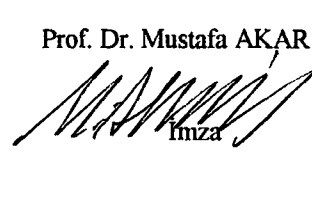
Prof. Dr. Necati YILDIZ (Danışman)

Prof. Dr. Yüksel BEK

Prof. Dr. Mustafa AKAR


İmza


İmza


İmza

TEŞEKKÜR

Bu çalışmanın her aşamasında yardımlarını esirgemeyen danışmanım Prof. Dr. Necati YILDIZ'a,

Doktora tez konumu belirlemede yardımcı olan ve doktora derslerinin bir çoğunu şahsından büyük bir özveri ile almış olduğum ayrıca çok değerli zamanını bana ayırarak tezimi baştan sonuna kadar büyük bir titizlikle inceleyen Prof. Dr. Yüksel BEK'e,

Hem literatür temininde, hem de takılmış olduğum noktalarda yardımlarını esirgemeyen Osman Gazi Üniversitesi Tıp Fakültesi öğretim elemanlarından Prof. Dr. Kazım ÖZDAMAR ve asistanı Arş. Gör. Cengiz BAL'a

Tezimde kullanmış olduğum fıstık (*Pistia vera L.*) tiplerine ilişkin verileri esirgmeden temin eden Arş. Gör. İzzet AÇAR'a,

Tezimi baştan sonuna kadar okuyup, ifade hatalarının düzeltilmesinde yardımcı olan Yrd. Doç. Dr. Turgay ŞENGÜL'e,

Her türlü yardımını esirgemeyen Arş. Gör. Paki TURGUT'a,

Başta, çok değerli bölüm başkanımız Doç. Dr. Ramazan ŞEVİK ve diğer bütün bölüm öğretim elemanları ile oda arkadaşım ve aynı uzmanlık dalında birlikte çalıştığım Arş. Gör. Zeki DOĞAN'a bana sabırla sürekli maddi ve manevi destek olduğundan, en derin şükranlarımı sunarım.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZ.....	V
ABSTRACT	VI
ŞEKİLLER	VII
ÇİZELGELER.....	X
İSTATİSTİK TERİMLERİ LİSTESİ.....	XIII
1. GİRİŞ.....	1
2. ÖNCEKİ ÇALIŞMALAR.....	4
2.1. Kümenin Tanımlanması ve Sınıflandırılması ile İlgili Çalışmalar.....	4
2.2. Kümeleme Analizi Teknikleri ve Kriterlerinin Araştırılması ile İlgili Çalışmalar	14
3. MATERYAL VE METOT	21
3.1. Materyal	21
3.2. Metot.....	21
3.2.1. Sayısal sınıflandırma analizinde kullanılacak veri setinin yapısı	21
3.2.2. Benzerlik (similarity) ve uzaklık (distance measure) ölçütleri.....	24
3.2.2.1. Minkowsky uzaklık ölçütü.....	28
3.2.2.2. Manhattan (City- Blok) uzaklık ölçütü.....	28
3.2.2.3. Euclid (Öklid) uzaklık ölçütü.....	28
3.2.2.4. Tartılı Euclid (Öklid) uzaklık ölçütü.....	30
3.2.2.5. Mahalanobis- D^2 uzaklık ölçütü.....	31
3.2.2.6. Chebychev uzaklık ölçütü.....	32
3.2.2.7. Canberra uzaklık ölçütü.....	33
3.2.2.8. Hotelling T^2 uzaklık ölçütü.....	33
3.2.2.9. Vektör çarpım uzaklık ölçütü.....	34
3.2.3. İlişki türü benzerlik ölçüleri	35
3.2.3.1. Korelasyon katsayıları (nicel veriler için iki değerli değişkenler).....	36
3.2.3.2. İkili ve nitel verilerle ilgili benzerlik ölçüleri	37
3.2.3.3. İki yönlü tablolarda birliktelik ölçüsü	39
3.2.4. Kümeleme yöntemleri.....	41

3.2.5. Aşamalı kümeleme (hierarchical clustering) yöntemleri	43
3.2.5.1. Tek Bağlantı (en yakın komşu) kümeleme yöntemi (Single Linkage Cluster Analysis).....	45
3.2.5.2. Tam Bağlantı (en uzak komşu) kümeleme yöntemi (Complete Linkage Cluster Analysis).....	46
3.2.5.3. Gruplar İçi Ortalama Bağlantı kümeleme yöntemi (Within-Groups Linkage Cluster Analysis).....	47
3.2.5.4. Gruplar Arası Ortalama Bağlantı kümeleme yöntemi (Between-Groups Linkage Cluster Analysis).....	48
3.2.5.5. Merkezi (Centroid) Bağlantı kümeleme yöntemi.....	49
3.2.5.6. Ortanca (Median) Bağlantı kümeleme yöntemi	51
3.2.5.7. Ward kümeleme tekniği.....	52
3.2.5.8. McQuitty Bağlantı kümeleme yöntemi	53
3.2.6. Aşamalı olmayan (Nonhierarchical) kümeleme yöntemleri.....	54
3.2.6.1. K-Ortalama yöntemi.....	55
3.2.7. En iyi bölümlenme (partitioning) ölçütleri.....	57
3.2.7.1. Wilks'in olabilirlik oran testi	59
3.2.7.2. Hotelling-Lawley iz yöntemi	60
3.2.8. Küme sayısının belirlenmesi	61
3.2.9. Kümeleme analizinden önce değişken sayısının azaltılması.....	63
3.2.9.1. Ana Bileşenler Analizi (Principal Component Analysis).....	63
3.2.9.2. Ayırma (Discriminant Analysis) analizi.....	66
4. BULGULAR VE TARTIŞMA.....	70
4.1. Kümeleme Analizinde Kullanılacak Verilerin Yapısı ve Özellikleri	70
4.2. Verilerin Kümeleme Analizi İçin Hazır Hale Getirilmesi	71
4.3. Uzaklık (Distance) Matrisinin Elde Edilmesi	75
4.4. Aşamalı Kümeleme Yöntemleri	75
4.4.1. Tek Bağlantı (En Yakın komşu) Kümeleme Yöntemi İle Fıstık Tiplerinin Kümelenmesi.....	75
4.4.2. Tam Bağlantı (En Uzak Komşu) Kümeleme Yöntemi İle Fıstık Tiplerinin Kümelenmesi.....	77
4.4.3. Gruplar İçi Ortalama Bağlantı Kümeleme Yöntemi	80
4.4.4. Gruplar Arası Ortalama (Grup Ortalama) Bağlantı kümeleme yöntemi	82
4.4.5. Merkezi (Centroid) Bağlantı kümeleme yöntemi	83
4.4.6. Ortanca (Median) Bağlantı kümeleme yöntemi	85
4.4.7. Ward kümeleme tekniği	87
4.4.8. McQuitty Bağlantı Kümeleme Yöntemi	92
4.5. Aşamalı Olmayan K-Ortalama Yöntemi İle Fıstık Tiplerinin Kümelenmesi.....	94
4.5.1. Korelasyon Analizi İle Elemine Edilmiş Değişkenlerin K-Ortalama Yöntemi İle Sınıflandırılması.....	94

4.5.2. Ana Bileşenler Analizi İle Boyutu İndirgenmiş Olan Değişkenlere İlişkin K-Ortalama Yöntemi Sonuçları.....	95
4.6. Stepwise Diskriminant Analizi Yöntemi İle Değişkenlerin Elemine Edilerek Kümelenmesi.....	96
4.5. Wilk's Lambda ve Hotelling Lawley İz Ölçütleri İle Farklı Küme seviyelerinin Test edilmesi.....	106
4.8. Küme Sayısının Belirlenmesi.....	112
5. SONUÇLAR.....	118
6. KAYNAKLAR.....	121
7. EKLER.....	125
8. ÖZET.....	141
9. SUMMARY.....	142



ÖZ

Doktora Tezi

n<p BOYUTLU BİYOLOJİK VERİLERDE FARKLI KÜMELEME YÖNTEMLERİNİN KARŞILAŞTIRMALI OLARAK İNCELENMESİ

İrfan ÖZTÜRK

Harran Üniversitesi
Fen Bilimleri Enstitüsü Zootečni Anabilim Dalı
Biyometri ve Genetik Bilim Dalı

1999, Sayfa: 142

Araştırma, n<p boyutlu olan, 24 farklı Antepfıstığı (*Pistacia vera L.*) tipi ve bu tiplerden alınan 38 adet değişken üzerinde yapılmıştır. Ancak bu tip bir veri matrisine kümeleme analizine ilişkin bazı çok değişkenli test istatistiklerinin uygulanabilmesi için değişken sayısının (n>p) azaltılması gerekmektedir. Değişken sayısının azaltılmasında Ana Bileşenler (Principal Component) analizi, Ayırma (Diskriminant) analizi ve Korelasyon analizinden yararlanılmıştır.

Sözü edilen metotlarla indirgenen değişkenler, farklı kümeleme metotlarıyla karşılaştırmalı olarak incelenmiştir. Sonuçta, kümelemede en uygun yöntemin Ana Bileşenler analizi ile birlikte kullanılan Ward metodunun olduğu belirlenmiştir.

Küme sayısının belirlenmesinde ise en uygun ölçütün C_{max} , Wilks Lambda ve Hotelling Lawley İz istatistiklerinin olduğu belirlenmiştir.

ANAHTAR KELİMELEER: Kümeleme Analizi, Sayısal Sınıflandırma, Ana Bileşenler Analizi, Ayırma Analizi, Korelasyon Analizi.

ABSTRACT

Ph. D. Thesis

A COMPARATIVE STUDY OF DIFFERENT CLUSTERING METHODS ON THE BIOLOGICAL DATA WITH $n < p$ DIMENSIONS

İrfan ÖZTÜRK

Harran Üniversitesi

Graduate School of Natural and Applied Science

Department of Animal Science

Division of Biometry and Genetic

1999, Page: 142

This research was carried out on 38 variables taken from 24 types of pistachio (*Pistacia vera* L.) with $n < p$ dimensions. However, in order to apply the some multivariate test statistics of clustering analysis to this type data matrix, the number of variables ($n \geq p$) must be decreased. For decreasing the number of variables, analysis of principle components, analysis of Discriminant and analysis of Principal Components and analysis of Correlation were used.

Variables decreased by the methods mentioned above, was comparatively examined by different clustering methods. As a result it was shown that the most suitable method for clustering is the Ward method used with Principle Components analysis.

Also it was concluded that the most suitable scales (measurement) to determine the number of clusters are Wilk's Lambda, C_{\max} and Hotelling Lawley Trace statistics.

KEYWORDS: Cluster Analysis, Numerical Taxonomy, Principal Component Analysis, Discriminant Analysis, Correlation Analysis.

ŞEKİLLER

	<u>Sayfa</u>
Şekil 2.1. En yakın gözlem değerlerinin şematik olarak basamaklar halinde birleştirilmesi.....	5
Şekil 2.2. Şematik olarak birleştirilmiş olan gözlem değerlerine ilişkin ağaç grafiği.....	6
Şekil 2.3. a. Kümeleme analizi akış diyagramı-1.....	12
Şekil 2.3. b. Kümeleme analizi akış diyagramı-2.....	13
Şekil 3.1. Çok değişkenli veri matrisi şeması.....	22
Şekil 3.2. $D_{n \times n}$ boyutlu uzaklık matrisi şeması.....	25
Şekil 3.3. Uzayda iki noktanın birbirine olan uzaklığının gösterimi.....	26
Şekil 3.4. Aşamalı (hierarchy) kümelemenin yapısı ve ağaç grafiği.....	44
Şekil 3.5. Merkezi Bağlantı yönteminin şematik gösterimi.....	50
Şekil 3.6. Ana Bileşenler dönüşümünün iki boyutlu uzayda gösterimi.....	64
Şekil 3.7. Diskriminant analizinde boyut indirgeme.....	67
Şekil 4.1. En etkili olan Ana Bileşen sayısının (Scree Graph) grafik metodu ile belirlenmesi	73
Şekil 4.2. İlk iki Temel Bileşen vasıtasıyla elde edilen grupların (birimlerin) dağılımı.....	74
Şekil 4.3. Korelasyon analizi ile boyutu indirgenmiş olan değişkenlere ilişkin Tek Bağlantı yöntemi ile elde edilen ağaç grafiği.....	76
Şekil 4.4. Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlere ilişkin Tek Bağlantı yöntemi ile elde edilen ağaç grafiği.....	77
Şekil 4.5. Korelasyon analizi ile elemine edilen değişkenlerle Tam Bağlantı yöntemine göre elde edilen ağaç grafiği.....	78
Şekil 4.6. Orijinal değişkenlerle Tam Bağlantı kümeleme yöntemine göre elde edilen ağaç grafiği.....	78
Şekil 4.7. Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlere ilişkin fıstık tipleri ile Tam Bağlantı yöntemine göre elde edilen	

ağaç grafiği.....	79
Şekil 4.8. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklığı kullanılarak Grup İçi Ortalama Bağlantı yöntemi ile elde edilen ağaç grafiği.....	80
Şekil 4.9. Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlerle Grup İçi Ortalama Bağlantı yöntemine göre elde edilen ağaç grafiği.....	81
Şekil 4.10. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklığı kullanılarak Gruplar Arası Ortalama Bağlantı yöntemiyle elde edilen ağaç grafiği.....	82
Şekil 4.11. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenler vasıtasıyla Gruplar Arası Ortalama Bağlantı yöntemine göre elde edilen ağaç grafiği.....	83
Şekil 4.12. Korelasyon Analiziyle elemine edilmiş değişkenlerle Öklid uzaklığı kullanılarak Merkezi Bağlantı yöntemine göre elde edilen ağaç grafiği.....	84
Şekil 4.13. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlerle Merkezi Bağlantı yöntemine göre elde edilen ağaç grafiği.....	85
Şekil 4.14. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklıkları kullanılarak Ortanca Bağlantı yöntemi ile elde edilen ağaç grafiği.....	86
Şekil 4.15. Ana Bileşenler analizi ile boyutu indirgenen değişkenlerin Öklid uzaklığı kullanılarak Ortanca Bağlantı yöntemiyle elde edilen ağaç grafiği	87
Şekil 4.16. Korelasyon analizi ile elemine edilen değişkenler vasıtasıyla Ward yöntemi kullanılarak fıstık tiplerinin kümelenmesine ilişkin ağaç grafiği	88
Şekil 4.17. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward yöntemi ile elde edilen ağaç grafiği.....	89
Şekil 4.18. Orijinal değişkenlerle elde edilen Ward yöntemine ilişkin ağaç grafiği.....	90
Şekil 4.19. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Öklid yöntemi ile elde edilen ağaç grafiği	90
Şekil 4.20. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Manhattan yöntemi ile elde edilen ağaç grafiği.....	91

Şekil 4.21. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Pearson Kare yöntemi ile elde edilen ağaç grafiği.....	91
Şekil 4.22. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Pearson yöntemi ile elde edilen ağaç grafiği.....	92
Şekil 4.23. Korelasyon analizi ile elemine edilen değişkenlere ilişkin McQuitty yöntemi ile elde edilen ağaç grafiği.....	93
Şekil 4.24. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin McQuitty yöntemi ile elde edilen ağaç grafiği	93
Şekil 4.25. Küme sayısı 2 olması durumunda Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlere ilişkin Ward yöntemi ile elde edilen ağaç grafiği.....	99
Şekil 4.26. Küme sayısı 3 olması durumunda Stepwise Diskriminant Analizi ile boyutu indirgenen değişkenlere ilişkin Ward yöntemi ile elde edilen ağaç grafiği	100
Şekil 4.27. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 4 kümeye ayrıldıklarında Ward yöntemi ile elde edilen ağaç grafiği	100
Şekil 4.28. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 5 kümeye ayrıldıklarında Ward yöntemi ile elde edilen ağaç grafiği	101
Şekil 4.29. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 2 kümeye ayrıldıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği	103
Şekil 4.30. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 3 kümeye ayrıldıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği.....	103
Şekil 4.31. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 4 kümeye ayrıldıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği	104
Şekil 4.32. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 5 kümeye ayrıldıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği.....	105

ÇİZELGELER

	<u>Sayfa</u>
Çizelge 3.1. İki birey için iki yönlü birleştirme (association) tablosu.....	37
Çizelge 3.2. Bazı benzerlik katsayıları.....	39
Çizelge 3.3. İki yönlü tablolar.....	40
Çizelge 3. 4. Çok değişkenli varyans analiz (ÇDVAT) tablosu.....	59
Çizelge 4.1. Kümeleme analizinde kullanılan fıstık tipleri ve kod Numaraları.....	70
Çizelge 4.2. Karakterlere ilişkin tanımlama ve değişken tipleri.....	128
Çizelge 4.3.-1. Ceylanpınar Ceylanpınar Tarım İşletmesi (TİGEM)'den 24 farklı tipte erkek fıstık ağacına ilişkin 38 adet karakter üzerinden alınmış ölçümler.....	129
Çizelge 4.3.-2. Ceylanpınar Ceylanpınar Tarım İşletmesi (TİGEM)'den 24 farklı tipte erkek fıstık ağacına ilişkin 38 adet karakter üzerinden alınmış ölçümler.....	130
Çizelge 4.4.-1. Değişkenlere (karakterlere) ilişkin korelasyon analizi sonuçları....	131
Çizelge 4.4.-2. Değişkenlere (karakterlere) ilişkin korelasyon analizi sonuçları....	132
Çizelge 4.5. Korelasyon analizi ile (yüksek derecede ilişkili olan karakterleri) elemine edilmiş olan değişkenler.....	133
Çizelge 4.6. Yüksek derecede ilişkili değişkenler elemine edildikten sonra elde edilen yeni Korelasyon analizi sonuçları.....	134
Çizelge 4.7. Korelasyon analizi ile elemine edilmiş değişkenlerin standardize edilmiş sonuçları.....	135
Çizelge 4.8. Ana Bileşenler (Principal Component Analysis) analizi sonuçları....	72
Çizelge 4.9. Varyansın yüzde 89 (%)'nu açıklayan ana bileşenler.....	136
Çizelge 4.10. Ana Bileşenler analizi ile elde edilen yeni değişkenlere ilişkin scorlar.....	137
Çizelge 4.11. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklık matrisi.....	138

Çizelge 4.12. Korelasyon analizi ile elemine edilen değişkenlerin Öklid uzaklığının karesine (Öklid Kare) ilişkin uzaklık matrisi.....	139
Çizelge 4.13. Ana Bileşenler analizi ile boyutu indirgenen değişkenler ilişkin fıstık tipleri arasındaki Öklid uzaklıkları.....	140
Çizelge 4.14.-1. Ana Bileşenler analizi ile boyutu indirgenen değişkenler ilişkin fıstık tipleri arasındaki Öklid Kare uzaklıkları.....	141
Çizelge 4.14.-2. Ana Bileşenler analizi ile boyutu indirgenen değişkenler ilişkin fıstık tipleri arasındaki Öklid Kare uzaklıkları.....	142
Çizelge 4.15. Korelasyon analizi ile elemine edilen değişkenler vasıtasıyla fıstık tiplerinin K-Ortalama yöntemine göre kümelere atanması.....	95
Çizelge 4.16. Ana Bileşenler analizinden yararlanarak K-Ortalama yöntemine göre fıstık tiplerinin kümelere atanması.....	96
Çizelge 4.17. Stepwise Diskriminant analizi ile seçilmiş olan etkili değişkenler ve önem seviyeleri.....	97
Çizelge 4.18. Stepwise Diskriminant analizinden yararlanarak elemine edilen değişkenlerle Ward yöntemine göre fıstık tiplerinin kümelere atanması.....	98
Çizelge 4.19. Stepwise Diskriminant analizinden yararlanarak elemine edilen değişkenlerle Tam Bağlantı yöntemine göre fıstık tiplerinin kümelere atanması.....	102
Çizelge 4.20. Stepwise Diskriminant analizinden yararlanarak elemine edilen değişkenlerle k-Ortalama yöntemine göre fıstık tiplerinin kümelere atanması.....	106
Çizelge 4.21. Korelasyon analizi ile elemine edilmiş değişkenlere ilişkin farklı küme sayılarının Wilk's Lambda test istatistiği ile önemlilik düzeylerinin belirlenmesi.....	107
Çizelge 4.22. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlere ilişkin farklı küme sayılarının Wilk's Lambda test istatistiği ile önemlilik düzeylerinin belirlenmesi.....	108
Çizelge 4.23. Stepwise Diskriminant analizi ile boyutu indirgenmiş değişkenlere ilişkin farklı küme sayılarının Wilk's Lambda ölçütü ile önemlilik düzeylerinin belirlenmesi.....	109
Çizelge 4.24. Korelasyon analizi ile elemine edilmiş değişkenlere ilişkin farklı küme sayılarının Hotelling Lawley Trace (İz) ölçütü ile önemlilik düzeylerinin belirlenmesi.....	110

Çizelge 4.25. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlere ilişkin farklı küme sayılarının Hotelling Lawley Trace (İz) ölçütü ile önemlilik düzeylerinin belirlenmesi.....	111
Çizelge 4.26. Stepwise Diskriminant analizi ile boyutu indirgenmiş değişkenlere ilişkin farklı küme sayılarının Hotelling Lawley Trace (İz) ölçütü ile önemlilik düzeylerinin belirlenmesi.....	112
Çizelge 4.27. Ward yöntemi ile gruplara ayrılan fıstık tiplerinin, $M = k^2 W $ ölçütü ile en uygun küme sayısının belirlenmesi.....	113
Çizelge 4.28. Tam Bağlantı yöntemi ile gruplara ayrılan fıstık tiplerinin, $M = k^2 W $ ölçütü ile en uygun küme sayısının belirlenmesi.....	114
Çizelge 4.29. k-Ortalama yöntemi ile gruplara ayrılan fıstık tiplerinin, $M = k^2 W $ ölçütü ile en uygun küme sayısının belirlenmesi.....	114
Çizelge 4.30. Ward yöntemine göre C_{max} eşitliği vasıtasıyla küme sayısının belirlenmesi.....	115
Çizelge 4.31. Tam Bağlantı kümeleme yöntemine göre C_{max} eşitliği vasıtasıyla küme sayısının belirlenmesi.....	116
Çizelge 4.32. K-Ortalama yöntemine göre C_{max} eşitliği vasıtasıyla küme sayısının belirlenmesi.....	117

İstatistik Terimleri Listesi

Between-Groups linkage	Gruplar arası bağlantı
Centroid linkage	Merkezi bağlantı
Cluster analysis	Kümeleme analizi
Complete linkage	Tam bağlantı
Correlation	Korelasyon
Dendogram	Ağaç grafiği, ağaç diyagramı
Discriminant Analysis	Diskriminant, Ayırma analizi
Distance measure	Uzaklık ölçütü
Eigen value	Özdeğer
Hierarchical clustering	Aşamalı kümeleme, aşamasıralı kümeleme
Likelihood ratio test (LRT)	Olabilirlik oran testi (OOT)
Median linkage	Ortanca bağlantı
Multivariate analysis of variance (MANOVA)	Çok değişkenli varyans analizi (ÇDVA)
Nominal	Adlandırma
Nonhierarchical clustering	Aşamalı olmayan kümeleme
Ordinal	Sıralama
Partitioning	Bölümleme
Principal component analysis (PCA)	Temel / Ana bileşenler analizi (ABA)
Similarity matrix	Benzerlik matrisi
Single linkage	Tek bağlantı
Trace	İz (matrisin köşegen elemanlarının toplamı)
Within-Groups Linkage	Gruplar içi Ortalama bağlantı

1. GİRİŞ

Canlı organizmalar karakter benzerlikleri veya farklılıkları bakımından, doğadaki varyasyonun çok boyutlu bir modelini oluşturmaktadır (1). Türlerin alt bölümleri arasındaki ekolojik ve genetik etkilerin tümünün oluşturduğu varyasyon, türlere ait populasyon yapılarını ortaya koyar (2). Bu varyasyonların saptanması değişik tür ve çeşitler arasındaki taksonomik ilişkileri aydınlatmak için önemli olduğu gibi, ıslah çalışmalarında germplasm havuzlarını (germplasm pools) genişletmek açısından da son derece önemlidir (3).

Varyasyonun saptanmasında sayısal sınıflandırma (Numerical Taxonomy) metotları önemli rol oynamaktadır. Bu metotlar ilk defa 1957'de Sneath tarafından organizmalardaki varyasyonu saptamak ve sınıflamaya somutluk getirmek maksadıyla kullanılmıştır. Sneath'ın uygulaması bakteriler üzerinde, Sokal ve Michener'in ise arılar üzerinde yapılmıştır (4).

Günümüzde bütün araştırma ve denemelerden elde edilen veriler çeşitli istatistik metotlarla değerlendirilmektedir. Benzer şekilde bitkiler ve hayvanların doğru olarak sınıflandırılmaları için de uygun istatistik analizlerin kullanılması gerekmektedir.

Özellikle gen bankalarında korumaya alınan materyalin birbirlerine benzeyen veya ayrılan karakterlerinin ilgili kayıtlarını analiz etmek kolay değildir. Çünkü, benzerliğin ve ayrılığın saptanmasında genetik, agronomik, morfolojik, kimyasal vb. pek çok özellik dikkate alınmaktadır. Bütün bunlar ele alındığı zaman, her bir özelliğin ölçülebilir verilerle ifadesi, bu ifadelerin benzerliğini veya farklılığını ortaya koyacak bir dizi hesap işlemlerinin yapılmasını gerektirmektedir (5).

Karakter sayısı arttıkça, benzerlik veya ayrılığı ortaya koyan hesaplama işlemleri önemli bir sorun olmaktadır. Çünkü klasik istatistiksel yöntemler, karakter adedi ve karşılaştırılan örnek sayısı arttıkça yetersiz kalmaktadır. Bu sorun, fazla sayıdaki karakteri formüle eden "sayısal sınıflandırma" analizinin ortaya konması ile büyük ölçüde çözüme kavuşmuştur.

Sayısal sınıflandırma, sadece hayvan ve genetik kaynaklı materyalin değerlendirilmesinde kullanılmakla kalmayıp, hayvan ve bitki ıslahı çalışmalarında da genetik varyasyonu ortaya koyma, hat seçimi, ırk saptama, hibrit indeksleme,

konukçu - patojen - vektör ilişkilerini belirleme, toprak örneklerinde saptanan özelliklerden yararlanarak toprak haritaları yapma, herhangi bir bitkinin ekimi için yöreler saptama, tıpta-karaciğer ve böbrek hastalıklarının belirtileri, beyincik işlevleri, ölüm halindeki hastaların psikolojik durumları, hastalık salgınlarının yayılmaları ve gelişme modelleri gruplandırılarak tedavi yöntemleri geliştirme gibi birçok bilim dalında uygulanabilmektedir (5).

Günümüzde Biyoloji, Ziraat ve Tıp alanında yapılan araştırmalardan elde edilen önemli derecede karmaşık veriler, matrisler haline getirilerek Çok Değişkenli analiz yöntemleri ile değerlendirilmektedir (6).

Çok değişkenli analiz yöntemlerinden en yaygın olarak kullanılanları; Çok Değişkenli Varyans analizi (Multivariate Analysis of Variance), Faktör analizi (Factor Analysis), Ana Bileşenler analizi (Principal Component Analysis), Ayırma analizi (Discriminant Analysis), Uyum analizi (Corresponding Analysis), Sayısal Sınıflandırma analizi (Numerical Taxonomy) veya diğer bir ifade ile Kümeleme analizi (Cluster Analysis) olarak sıralanabilir (6), (7).

Veri matrisindeki birim ve/veya değişkenler hakkında ayrıntılı şekilde bilgi edinilmek isteniyorsa, bunların en çok benzer olanları bir araya getirilerek sınıflandırılırlar. Sınıflandırma sonucunda aynı grup içindeki birimler (bireyler) benzer özelliklere sahipken diğer gruptaki birimler farklı özelliklere sahip olmaktadır (8).

1960'lı yıllardan sonra bilgisayarın matematik ve istatistik alanlarında yoğun bir şekilde kullanılmaya başlanması ile çözümü çok zor olan sayısal sınıflandırma yöntemlerinin kolaylıkla kullanılabilme olanağı doğmuştur.

Sayısal Sınıflandırma analizleri genel olarak iki grup altında toplanmaktadır. Bunlar;

- Aşamalı (hierarchical) yöntemler
- Aşamalı olmayan (nonhierarchical) yöntemler olarak sınıflandırılabilir.

Bu araştırma, 24 farklı fıstık ağacı tipinden alınan 38 adet değişken üzerinde uygulanmıştır. Ancak bilindiği üzere kümeleme analizi ile birlikte bazı Çok Değişkenli test istatistiklerinin uygulanabilmesi için veri matrisinin en azından ($n \geq p$) boyutunda olması gerekmektedir. Değişken sayısının birim (fıstık tipleri) sayısından fazla olduğu durumlarda, ($n < p$ boyutunda ki) bu tipteki verileri analiz edebilmek

için, deęişken sayısının bazı boyut indirgeme yöntemleri ile azaltılması gerekmektedir. Bu nedenle Korelasyon analizi, Ana Bileşenler analizi ve Stepwise Diskriminant analizi ile deęişkenlerin boyutu ($n \geq p$)'ye indirgenmiş ve gerçek deęişkenlerle elde edilen kümelemeye en yakın boyut indirgeme yöntemi tespit edilmeye çalışılmıştır. Bununla birlikte yukarıda sözü edilen kümeleme yöntemlerinden en yaygın olarak kullanılanları, boyutu indirgenmiş olan 24 Farklı Fıstık ağacı tipi üzerine uygulanarak en iyi küme yapısını ortaya koyan yöntem tespit edilmeye çalışılmıştır.

Farklı kümeleme yöntemlerine göre elde edilen kümeler, küme sayısını belirlemede kullanılan C_{max} , k , ve M ölçütleri ile Wilk's Lambda ve Hotelling Lawley İz istatistikleri kullanılarak test edilip en iyi küme yapısı belirlenecektir.

2. ÖNCEKİ ÇALIŞMALAR

Araştırmayla ilgili kaynaklar iki alt başlık altında incelenecektir. Birinci kısımda kümenin tanımlanması ve sınıflandırılması ile ilgili kaynaklar, ikinci kısımda ise kümeleme analizi ile ilgili teorik çalışmalar, karşılaştırmalar ve kümeleme kriterleri ile ilgili kaynaklar özetlenecektir. Tekrardan kaçınmak için bilhassa araştırmada uyguladığımız metotlar hakkında yapılan açıklamalar, ‘metotlar’ kısmında özetlenmiştir.

2.1. Kümenin Tanımlanması ve Sınıflandırılması ile İlgili Çalışmalar

Küme kavramının günümüzde çok değişik şekillerde tanımlanmasının asıl sebebi kümeyi oluşturan bireyleri bir araya getiren ortak özelliklerin farklı olmasıdır. Bu ortak özellikler; benzerlik, yakınlık, uzaklık gibi değişik kavramlarla tanımlandığında küme kavramı için tek bir açıklama yapmak oldukça zordur (9).

Kendall ve Buckland (10), kümelerin çeşitli özelliklerini dikkate alarak bir tanımlama yapmışlar ve çalışmalarında kümeyi “İstatistiksel evrenin yakın öğelerinden oluşan bir grup” olarak tanımlamışlardır. Bu tanımlamalarda, birim-birey, nesne-gözlem şeklinde geçen iki önemli kavram bulunmaktadır. Bireylere ilişkin vasıflar ise karakter, değişken veya özellik olarak birbirleriyle eş anlamlı olarak kullanılmaktadır.

Bu çalışmada, daha çok birey veya birim ile karakter veya değişken terimleri kullanılacaktır.

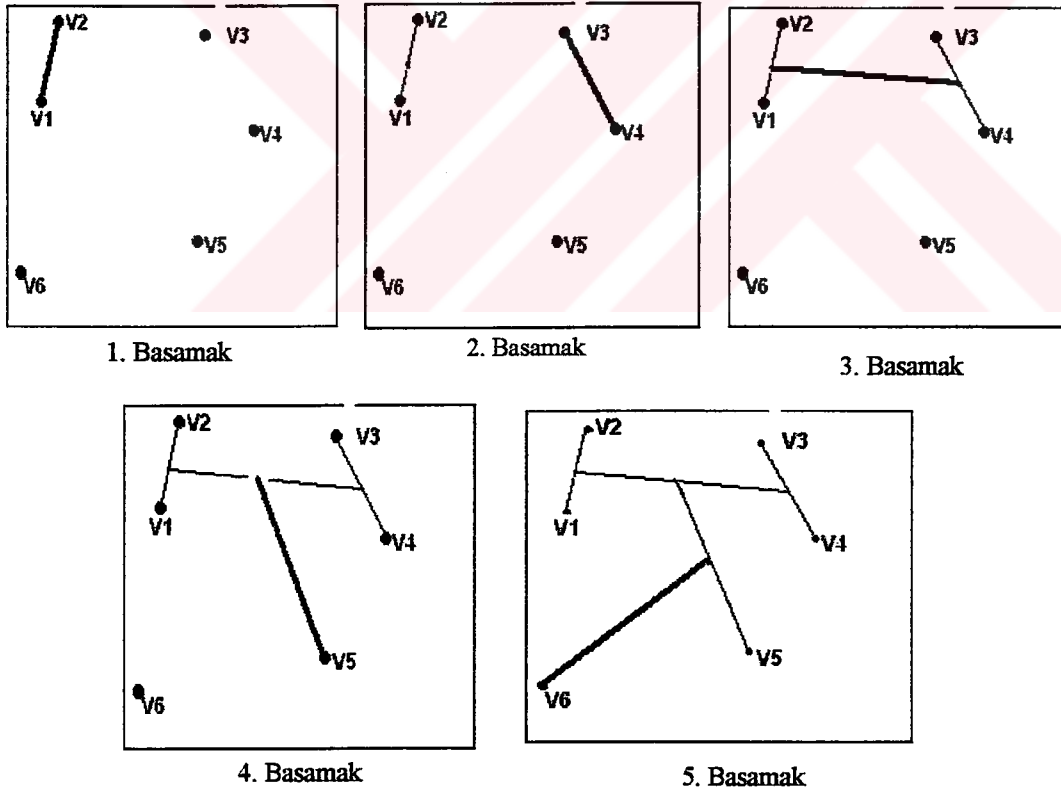
Everitt (11), farklı kaynaklarda, gruplandırma işleminin kullanım alanlarına göre değişik şekillerde isimlendirildiğini bildirmiş ve bunları; Tipleme (Typology), Yığıma (Clumping), Q-Analizi (Q- Mode Analysis), Sınıflandırma (Classification), Sayısal sınıflandırma (Numerical Taxonomy), Örüntü tanımlama (Pattern Recognition) ve Kümeleme analizi (Cluster Analysis) şeklinde sıralamıştır.

Özdamar (12), Populasyonu meydana getiren bireylere ait birden fazla karakterin gözlemlenmesiyle elde edilen verilerin yorumlanmasının veya bireylerin yapısı hakkında tam olarak bilgi sahibi olmadan sınıflandırma yapılmasının oldukça güç olduğunu belirtmiştir.

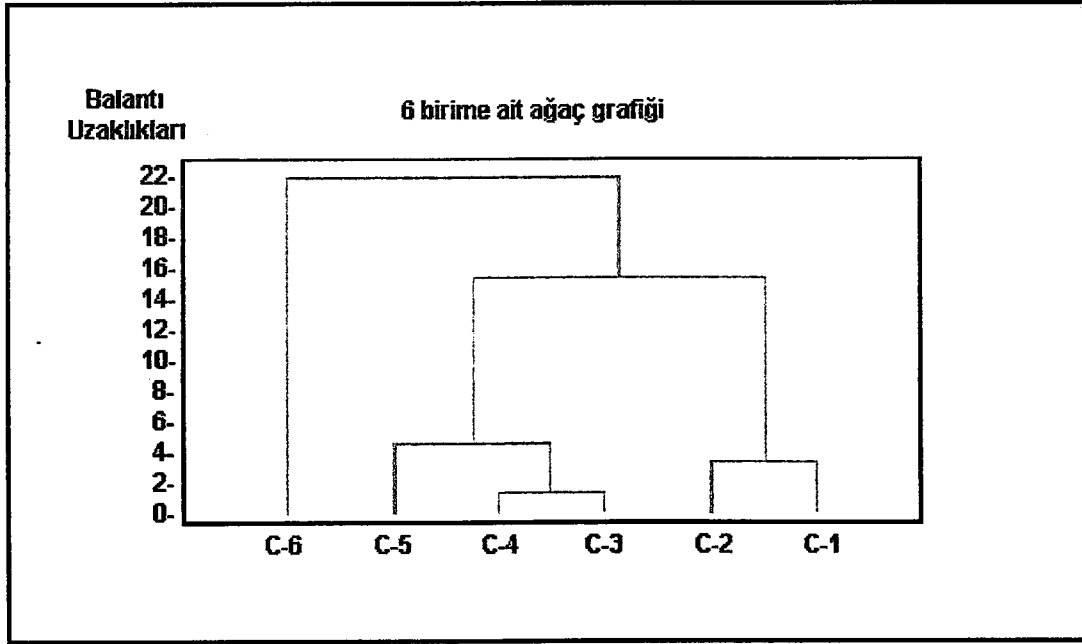
Sneath and Sokal (13), Sayısal sınıflandırma analizinin; karakterleri itibariyle birbirine benzer olan bireylerin aynı gruplarda toplanması amacıyla yapılan çalışmalar olarak tanımlanabileceğini bildirmişlerdir.

Bu tanımlamaya göre bireylerin aynı grupta toplanmasını sağlayan onların ortak karakterleridir. Bu demektir ki bir bireyden ne kadar fazla karakter (nitel veya nicel) ölçülürse, bireyler arasındaki ortak özellikler artacağından benzer karakterlere sahip olanlar aynı sınıfta yer alacaktır. Burada her bir bireyin bir küme olarak kabul edildiği, daha sonra birbirine en yakın olan bireylerin aynı grupta toplanarak küme sayısı (n)'den (n-1)'re indirildiği belirtilmiştir. Bu işlemin, bütün bireyler bir tek sınıf oluşturuncaya kadar devam edeceği ifade edilmiştir.

Sayısal sınıflandırma, "Test uzayında birbirine en yakın noktaların bir araya getirilmesi" tanımından yararlanılarak, şematik olarak ve ağaç grafiği (dendogram) ile Şekil 2.1.ve Şekil 2.2'de gösterilmiştir.



Şekil 2.1. En yakın gözlem değerlerinin şematik olarak basamaklar halinde birleştirilmesi



Şekil 2.2. Şematik olarak birleştirilmiş olan gözlem değerlerine ilişkin ağaç grafiği

Bireylerin sınıflandırılması ilk önce Biyoloji ve Zooloji alanlarında yapılmıştır. Eski Roma, Yunan ve Hindistan medeniyetlerinde, insanlar çeşitli özellikleri bakımından sınıflandırılmıştır. Bu konudaki ilk basılı eser, 18. yüzyılda ve İngiltere de yaşayan bitki ve hayvanların sınıflandırılması hakkında verilmiştir. Asıl gelişme 1930 yılından sonra doğa bilimlerinde olmuştur (11).

1960'tan sonra kümeleme analizi ile ilgili birçok yöntemler geliştirilmiştir. Özellikle bu çalışmalar, 1960-1970 yılları arasında yoğunlaşmıştır. Sayısal sınıflandırma ile ilgili ilk eser, 1963 yılında Sokal ve Sneath tarafından "Sayısal Sınıflandırma Yöntemleri" adıyla basılmıştır. Özellikle, 1965 yılından sonra bilgisayarların yaygın olarak kullanılmaya başlanmasıyla sayısal sınıflandırma (kümeleme) analizinin önemi daha da artmış ve bu sahadaki çalışmalar hızlanarak birçok metodlar pratikte kullanılabilir hale gelmiştir (14).

Kümeleme analizi konusundaki ilk eser 1970 yılında Tyron ve Bailey tarafından yayınlanmıştır. Daha sonra bunu 1973'de Anderberg, 1973'de Sneath ve Sokal, 1974'de Everitt ve 1974'de Duran ve Odell'in kitapları izlemiştir. 1975 yılında ise Hartiganın Bilgisayar algoritmaları ile ilgili kitabı yayınlanmıştır (15).

TAN (5), sayısal sınıflandırma analizleri ile doğal sınıflara hemen hemen yakın sonuçların elde edilmesinin, birçok bilim dallarında kullanılmasına sebep olduğunu bildirerek bu bilim dallarını aşağıdaki gibi açıklamıştır.

A) Sosyal bilimler

ARKEOLOJİ - Arkeolojik kazılarda bulunan yazıtların tarihsel zamanlarını saptama;

DEMOGRAFI - İdari bölgelerdeki kayıtlardan yararlanarak geçmiş zamanlardaki insanların göç kalıplarının sosyal ve genetik sonuçlarını anlama;

DİL BİLİM - Diller ve konuşma çeşitliliğinden yararlanarak, dillerin evrimsel ayrıntılarını tanımlama ve kelimelerin anlamlarını açıklama;

EĞİTİM PSİKOLOJİSİ - Öğrencilerle ilgili özelliklerden yararlanarak akademik erişim ve düzenleme sınavları için tipolojik kalıplar geliştirme;

EKONOMİ - Girdi-çıkıtı (Input-output) akış tablolarından yararlanılarak, bir ekonominin tekdüze sektörlerini çıkartmak; böylece ekonomik yapıların kalıplarını kolaylaştırma;

HUKUK - Avrupa Konseyi antlaşmalarındaki hükümlerden yararlanılarak, yürürlükte olan EEC kanunlarının işletimsel sınıflandırmalarını çıkarma;

KÜTÜPHANECİLİK - Yayınların konu ve aktarılan bölümlerini kümeleyerek, bilgi indeksleme, özetleme ve erişim için etkili konu gruplarını oluşturma;

POLİTİKA - Oylama sonuçlarından yararlanarak, oylama çelişkilerini açıklama ve gelecekteki oy verme eğilimlerini tasarılama;

SOSYOLOJİ - Toplumlara ait sosyal özellikler, davranış ve kurallardan yararlanarak, sosyal sınıfların tipolojik kalıplarını geliştirme ve tanımlama;

B) Fen ve mühendislik

ASTRONOMİ - Yıldızlar ve yıldız kümelerinin fotoğraflarından yararlanılarak, güneş sistemlerinin oluşumu konusundaki teorileri geliştirme;

JEOLOJİ - Borelog ölçümlerinden yararlanılarak mineral çıkarmada stratigrafik yorumlama yapma;

KİMYA - Kimyasal yapıların bilinmeyen biyolojik, fiziksel ve kimyasal özelliklerini daha iyi saptama;

METEOROLOJİ - Gelişen hava durumu kalıplarından, geniş alanlar için hava tahmin modellerindeki genel değişimleri bulup çıkarma;

MÜHENDİSLİK - Kümeleme tasarım yöntemlerine göre desenlenmiş makina modelleri oluşturarak, güneş sistemlerinin fotoğraflarından, el yazısı karakterlerinden, parmak izinden, radar sinyallerinden, röntgen ve elektrokardiyogramlardan özellik tanımlama;

TOPRAK İLMİ - Toprak örneklerinde saptanan özelliklerden yararlanılarak toprak haritaları yapma ve herhangi bir bitkinin ekimi için uygun yöreler saptama;

UZAY BİLİMİ - Çok yönlü ışık dağılım verileri (multispectral scanner data) kullanılarak uzay uçuşları için kayıtlanan değişik özellikleri sınıflandırma ve uçuşları kolaylaştırma;

VERİ TABANI (DATABASE) DESENLEME - Bilişim dizge elementlerinden yararlanılarak veri tabanı seçimleri ve erişim işlevlerinin etkinliğinden en iyi şekilde yararlanma;

C) Doğa bilimleri ve tıp

ANTROPOLOTİ - Antropometrik veriler kullanarak, geçmiş zamanlara ait insan ırkları arasındaki genetik farklılıkları açıklama ve bunların evrimsel tarihlerini yeniden düzenleme;

BİYOKİMYA - Protein ve genlerin aminoasit kompozisyonlarını kullanarak, yeni türlerin orijinine temel olan mutasyonların etkilerini açıklama;

BOTANİK - Vejetasyon örneklerinin çeşitli özelliklerini çözümleyerek, doğal bitki topluluklarının ekolojilerini tanımlama ve tarımsal gelişmeler için yöreler saptama veya toprak muhafazası için önlemler alma, öneriler yapma;

ENTOMOLOJİ - Arthropodalar, sivrisinekler, arılar, kabukluların çeşitli özelliklerini kümeleyerek bunların türleri arasındaki varyasyonu saptayarak, taksonomik tanımlamalarını yapmak ve işletimsel taksonomilerini geliştirme;

MİKROBİYOLOJİ - Bakteri ve virüslerde ırk saptama, yeni cinslerini tanımlama; var olan taksonomilerini geliştirme; hastalık etmeni olarak tanımlanan

yeni organizmaların (bakteri veya virüs) saptanan özelliklerinden yararlanarak aşı veya ilaçlar geliştirme;

PALEONTOLOJİ - Fosillerden yararlanarak, bunların fenetik ilişkilerini ölçümleme ve evrimsel teorilerini oluşturma;

PSİKİYATRİ - Psikopatolojik belirtileri kullanarak, psikiyatrik bozukluk belirtilerini sınıflandırma ve tedavi salık verme;

SİTOLOJİ - Kan örneklerini ölçümleyerek, kan gruplarını ve plazma tiplerini tanımlama ve anormal hücreleri tanıma yöntemlerini geliştirme;

TARIM - Çeşitli yörelerdeki çiftlik sayılarından yararlanarak bir bölgedeki değişik tarımsal işlevlerin dağılım haritalarını yapma üretimdeki değişiklikleri izleme;

TARIM EKONOMİSİ – Tarımsal ürünlerle yapılan çalışmalardan yararlanarak yeni çeşitlerin durağanlığı ve eskilerinden farklılığını ölçümleme ve bunların rekoltelerini tahmin;

TIP - Karaciğer ve böbrek hastalıklarının belirtileri, beyincik işlevleri, ölüm halindeki hastaların psikolojik durumları, hastalık salgınlarının yayılmaları ve gelişme modelleri kümelenerek tedavi yöntemlerini geliştirme;

D) Endüstri ve yönetim

BORSA - Fiyat serilerinin benzerliklerinden yararlanarak pazar eğilim kümelerini ölçümleme, sektör göstergelerini oluşturma ve tahvil yönetim performanslarını saptama;

ENDÜSTRİ EKONOMİSİ - Firmaların malları, işlevleri ürünleri ve rekabet şekillerinden gidilerek, endüstri sektörlerini, tanımlama, gelişme programlarını planlama ve türevsel performanslarını görüntüleme;

EV DEKORASYON - Ailelerin kullandıkları mobilyalarla ilgili verilerden yararlanarak., değişik tipteki ailelerin konut gereksinimleri ve dekorasyonlarının tanımlanması;

İŞ PSİKOLOJİSİ - İş tanımlarından yararlanarak, acemi işçiler, iş eğitimi, personel yönetimi ve işgücü planlaması için uygun iş sınıfları oluşturma;

PAZARLAMA - Ailelerin kullandıkları tüketim mallarının markaları ve özelliklerinden yararlanarak tüketim tiplerini tanımlama, ürünün çekiciliğini ölçümleme ve satış artışını sağlayan unsurları desenleme;

ŞEHİRCİLİK - Değişik yörelerdeki insan popülasyonlarının özelliklerinden gidilerek, yoksul yöreleri tanımlama ve yerleşme, refah iyileştirme programlarını gözleme;

TOPLU PLANLAMA - Belediyelerin parasal bütçelerinin durumundan yararlanarak, toplumsal planlamaları sınıflandırma ve merkezi vergilendirmeyi kolaylaştırma;

YÖNEYLEM - Bir organizasyonun dağıtım fonksiyonları, kontrol ve iletişim özelliklerinden yararlanılarak, büro yerleşimlerini optimize etme konularında uygulanmıştır.

Everitt (11), kümeleme yöntemlerinin araştırmacının amacına göre kullanımının değiştiğini, ancak genel olarak kullanım amaçlarının aşağıdaki şekilde sıralanabileceğini belirtmiştir.

- a) Gerçek tiplmeyi bulmak
- b) Model uyumu sağlamak
- c) Gruplar için ön tahmin hesapları
- d) Hipotez testi
- e) Veri yapısının bulunması
- f) Hipotez genelleştirmesi
- g) Boyut indirgeme

Bu sıralama Chatfield and Collins (16) tarafından daha genel olarak aşağıdaki şekilde yapılmıştır.

- a) Kümelerin belirlenmesi
- b) Boyut indirgeme
- c) Hipotez genelleştirmesi
- d) Kümeler üzerinde ön kestirim

Sayısal sınıflandırmaya ilişkin kullanım amaçları diğer bazı arařtırmacılar tarafından deęişik şekillerde belirtilmiştir. Ancak genel itibariyle sayısal sınıflandırmada temel amaç veri yapısının arařtırılmasıdır.

Öner (17), bir popülasyondaki çok deęişkenli gözlem deęerlerini tanımlamanın oldukça güç olduğunu ve istenilen bilginin çıkarılıp özetlenmesi gerektiğini belirtmiştir. Bu bakımdan arařtırmacı bireyleri kontrol edilebilir gruplara ayırmadıkça, istenilen bilginin kolaylıkla elde edilemeyeceğini ve bu nedenle sayısal sınıflandırma analizlerinin veri düzenlemede kullanılabileceğini belirtmiştir. Böylece elde edilen gruplar içinde homojenliği, gruplar arasında da heterojenliği sağlayarak, çok sayıda gözlem deęeri yerine birkaç grup ile çalışma kolaylığı temin edilmiş olmaktadır. Sayısal sınıflandırmanın amacının, bu aşamada meydana gelebilecek bilgi kaybını minimuma indirmek olduğu bildirilmektedir.

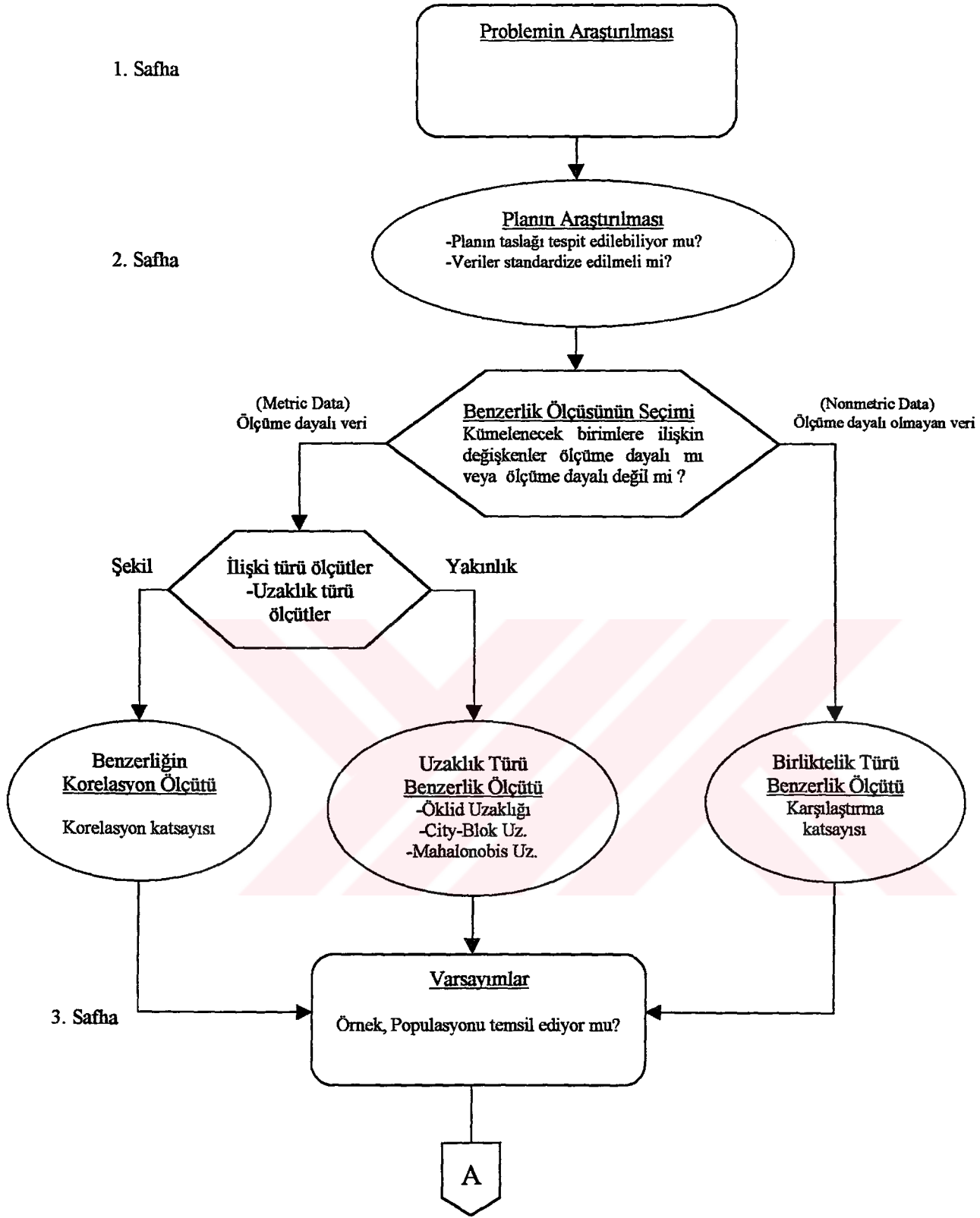
Anderberg (8), arařtırmacıların sayısal sınıflandırma analizinden en iyi yararı elde etmek için, aşağıdaki hususlara dikkat etmeleri gerektiğini bildirmiştir.

a) Deęişik sınıflandırma yöntemlerinin, aynı veri kümesinden farklı sonuçlar verebilmesidir. Bundan dolayı tek bir yöntemle bağımlı kalmayıp birden fazla yöntemin denenmesi daha sağlıklı sonuç verecektir.

- b) Ortaya çıkan gruplar kesin sonuçlar olmayıp, sadece bir ön tahmindir.
- c) Sayısal sınıflandırma yöntemleri hipotez tahminleri için bir amaçtır.
- d) Sınıflandırma sonucunda, veri kümesinin karışık yapısı ortaya çıkabilir.
- e) Yine sınıflandırmanın neticesinde veri yapısının gerçek görünümü ortaya çıkabilir.

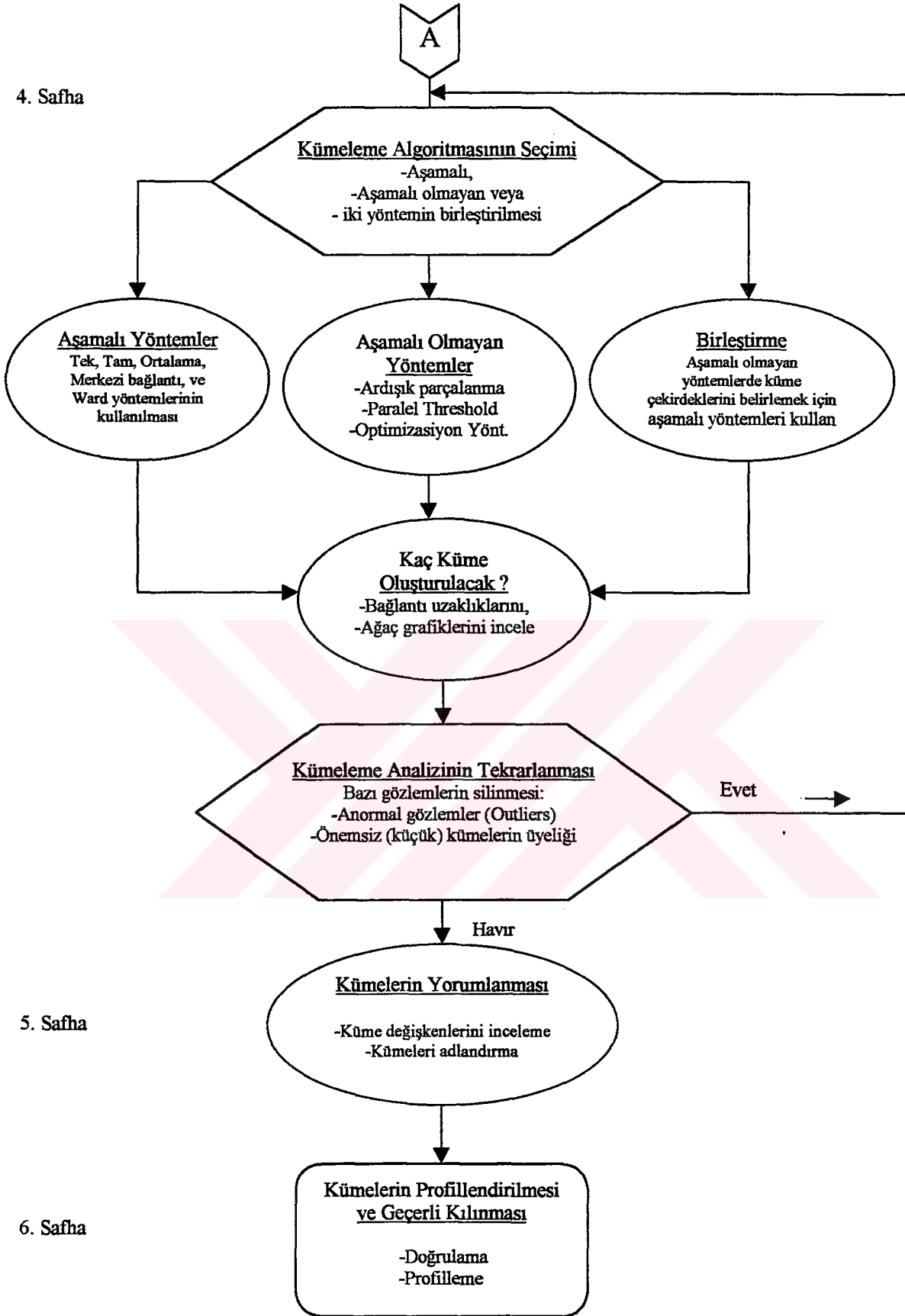
f) Verilerde küme yapısının olmaması veya bir tek küme olması gibi özel bir durumla da karşılaşılabilir.

Hair et all (40), Kümeleme analizi yapılırken takip edilecek yolu bir akış diyagramı ile Şekil 2.3.a. ve Şekil 2.3.b.'deki şemada aşama aşama özetlemiştir.



Şekil 2.3.a. Kümeleme analizi akış diyagramı-1

4. Safha



Şekil 2.3.b. Kümeleme analizi akış diyagramı-2

2.2. Kümeleme Analizi Teknikleri ve Kriterlerinin Araştırılması ile İlgili Çalışmalar

Edwards and Cavalli-Sforza (18), kümeleme analizinde farklı bir yaklaşımda bulunarak, çok boyutlu uzaydaki noktalar arası ilişkilerin araştırılmasında Çok Değişkenli Varyans analizi tekniğini kullanmışlardır. Sözü edilen teknik vasıtasıyla noktaları en yoğun iki kümeye ayırıp, bu işlemi ardışık olarak tekrar etmişlerdir. Sonuçta elde ettikleri bir ağaç grafiği (Dendogram) ile birimleri birbirinden ayırmışlardır.

Scott and Knott (19), varyans analizinde muameleleri mantıklı bir şekilde homojen gruplara ayırmanın bazen faydalı olacağını belirtmişlerdir. Bu amaçla sık sık çeşitli karşılaştırma prosedürlerinin kullanıldığını, ancak en doğrusunun kümeleme analizi tekniklerinin kullanılması olduğunu ifade etmişlerdir. Bu yaklaşımı birçok veri kümeleri ile göstermişler ve gruplar arasındaki farklılıkların önemini test etmek için 'olabilirlik oran testi' ni geliştirmişlerdir.

Jardine and Sibson (20), yaptıkları çalışmada sınıflandırma sistemlerinin oluşturulmasında kullanılan kümeleme metotlarının işlettirilmesinin, veri matrisinde elde edilen uzaklık ölçüleri ile sağlandığını bildirmişlerdir. Bu çalışmada, kümeleme metotlarının özelliklerini tartışmak üzere teorik bir çalışma çerçevesinde bir taslak çıkarmışlar ve kümeleme metodunu sağlayacak kesin şartları maddeler halinde vermişlerdir. Bunları sırasıyla;

1. $d(a, b) \neq 0$ $|P| > 1$
2. $d(a,b) \geq 0$ $a,b \in P$
3. $d(a,b) = 0$ $a = b$
4. $d(a,b) = d(b,a)$ $a,b \in P$
5. $d(a,c) \leq \max \{ d(a,b), d(b,c) \}$ $a,b,c \in P$
6. $d(a,c) \leq d(a,b) + d(b,c)$ $a,b,c \in P$

şeklinde sıralamışlardır. Bu şekildeki uygulamalarda tek bağlantı yönteminin matematiksel olarak en iyi yöntem olduğunu belirtmişlerdir.

Fisher and Van Ness (21), bir A şahsının genel veya özel bir uygulama için belirlediği kümeleme prosedürleri, kendi çalışması için mantıklı ve tatmin edici bazı sonuçlar verirken, bir B şahsı için aynı sonuçları vermeyebileceğini belirterek en iyi kümeleme prosedürünü belirlemenin mümkün olmadığını ifade etmişlerdir. Araştırmacılar kendi çalışmaları için dokuz kümeleme prosedürü (kabul edilebilirlik şartı) belirlemişler ve tanımladıkları bu kabul edilebilirlik şartlarına göre Tek ve Tam bağlantı ile merkezi bağlantı kümeleme yöntemlerini karşılaştırmışlardır. Sonuç olarak, Tek bağlantı yönteminin matematiksel olarak en iyi yöntem olduğunu söylemişlerdir.

Everitt (22), yaptığı çalışmada, Fisher ve Van Ness (21) ile Jardine ve Sibson (20)'nın çalışmalarını inceleyerek her iki çalışmada sunulmuş olan şartların pratikte kullanılmasının çok zor teorik şartlar olduğunu ifade ederek, söz konusu araştırmacıların bu önerilerini uygulamada gösteremediklerini açıklamıştır.

Kuiper and Fisher (23), yaptıkları çalışmada;

1. Tek bağlantı
2. Tam bağlantı
3. Merkezi bağlantı
4. Ortanca (median)
5. Grup ortalama
6. Ward'ın kareler toplamı metodu

olmak üzere altı Aşamalı kümeleme metotlarının performansını iki değişkenli ve çok değişkenli Monte Carlo simülasyon örnekleriyle karşılaştırılmışlardır. Kümeleme metotlarına ilişkin doğru küme sayılarını tespit ederek, birimlerin (nokta çiftlerinin) doğru kümelerine atanıp atanmadıklarını belirlemeye çalışmışlardır. Bununla birlikte hangi metotların, birimleri doğru kümelerine atamalarında etkin olduğunu karşılaştırmaya çalışmışlar ve sonuç olarak,

1. Tek bağlantı yönteminin iyi düzenlenmiş ortamda daha iyi sonuç verdiğini
2. Merkezi bağlantı ile grup ortalama yönteminin benzer tipteki örnek gruplarında aynı sonuçları verdiğini ve Ward yöntemine en çok benzeyen yöntemin Tam bağlantı yöntemi olduğunu

3. Küme sayısının artmasıyla sınıflandırmanın daha iyileştiğini
4. Çok değişkenli normal dağılımdan çekilmiş örnekler (birimler) için Ward yönteminin hemen hemen gerçek doğal sınıflamaya yakın sonuçlar verdiğini
5. Cauchy dağılımında anormal gözlemler nedeniyle tüm yöntemlerin zayıf sonuçlar verdiğini
6. Tam bağlantı ile Ward yönteminin yoğun ve eşit büyüklükteki örneklerde en iyi sonucu verdiğini belirtmişlerdir.

Cormack (24), kümeleme analizi ile büyük miktarda çok değişkenli verilerin özetlenmesinin çok zor olduğunu ve sık sık savunulamayacak şekilde kullanıldığını dile getirmiştir. Kümeleme ve benzerliğin gereğinden fazla tanımlamalarının sunulduğunu, bununla birlikte birçok yöntemin temellerinin tartışıldığını ve uygulamada bir kısım sınırlandırmalar getirildiğini belirtmiştir. Sayısal sınıflandırmanın başlı başına tatminkar bir yöntem olmadığını bildirmiştir.

Marriott (25), kümeleme analizi üzerinde yaptığı çalışmada grup içi dağılımı matrislerin determinantlarının minimize edilmesine dayanan sınıflandırma metodunu pratikte kullananların bakış açısına göre tartışmıştır. Sonuçların hem yorumlanmasında hem de hesaplanmasında bir kısım zorluklar çıktığını belirtmiş ve bu problemlerin çözümünün ise doğal kümelerin sayısının doğru tahmin edilmesine bağlı olduğunu söylemiştir. Uniform bir dağılıma uygulandığında metodun örnekleme davranışının incelendiğini, ancak kesin bir sonuç çıkarılmadığını bildirmiştir. Sonuçta yine en uygun olanının çok değişkenli varyans analizinden yararlanarak,

	S.D.	Matris Dağılımı
Grup arası	$g - 1$	B
Grup içi	$n - g$	W (Hata)
Toplam	$n - 1$	T

W'nin (hata) grup içi dağılımının minimize edilmesi olduğunu belirtmiştir.

Stoddard (26), kümeleme analizinde, bireyler arasındaki uzaklığın ölçekten bağımsız olmasının genel bir problem olduğunu belirtmiş ve makalesinde, standart bir karşılaştırmada ölçümlerin ölçeklendirilmesi için bir prosedür sunmuştur. Bu prosedürün rutin bir referans olarak, özellikle laboratuvar sonuçlarının sınıflandırılmasında kullanılabileceğini ifade etmiştir. Prosedürün, sınıflandırma için gerekli bilgiyi içermesine rağmen, değişkenlik taşıyacağını bildirmiştir.

Hartigan (27), çalışmasında gözlemlerin doğrudan veri matrisi üzerinden kümelenebileceğine dair bir metot sunmaktadır. Eğer veriler birkaç birim üzerinden elde edilen değişkenlerden ibaret ise, bu durumun, algoritmaların ya birimlerin kümelenebilmesinde veya değişkenlerin kümelenebilmesinde rol oynayabileceğini söylemiştir. Ayrıca bu yaklaşımın temel avantajı veriler üzerindeki kümelerin doğrudan yorumlanabilmesidir. Ancak bu uygulamanın pratikte çok fazla kullanılmadığı belirtilmiştir.

Murtagh (28), kümeleme analizinde en önemli problemin sonuçların geçerliliği olduğunu bildirmekte ve sonuçların geçerliliğini muhtemel bir yaklaşımla incelemiştir. Araştırmacı etiketlenilmemiş ağaç grafiklerin yapısını incelemiş ve rastgele ağaç grafiklerinin bir modelini önermiştir. Ağaç grafiği yapısını belirleyen katsayının olası dağılımının hesaplanması için algoritmalar sunmakla beraber, sonuçları farklı ağaç grafikleri ile ve Monte Carlo simülasyonu ile elde etmiş ve bunu veriler üzerinde tartışmıştır.

Blashfield (29), Monte Carlo teknikleri ile türetilen veri gruplarından karışık dağılıma sahip rastgele örnekler çekmiştir. Elde ettiği örnek veri gruplarını Tek bağlantı, Tam bağlantı, Merkezi bağlantı ve Ward kümeleme yöntemleri ile kümeleyerek kümelerin gerçek sınıflara olan yaklaşımını incelemiştir. Sonuç olarak, karışık dağılımlı modeli en iyi belirleyen metodun Ward yöntemi olduğunu tespit etmiştir.

Backer (30), yaptığı çalışmada gözlem hataları olan verileri kullanmak suretiyle Tek bağlantı ve Tam bağlantı yöntemleri ile elde edilmiş ağaç grafikleri

arasındaki ilişkiyi Goodman-Kruskal Gamma katsayısı ile ölçmüştür. Sonuçta, Tek bağlantı yönteminin Tam bağlantı yöntemine göre gözlem hatalarından daha fazla etkilendiğini belirtmiştir.

Gower (31), Grup ortalamaları, Ortanca (Median) ve Merkezi (Centroid) bağlantı yöntemlerini karşılaştırarak, gözlemler arası korelasyon katsayısının kullanılmasındaki mahsurlardan söz ederek Öklid uzaklık dönüşümlerinin yapılmasını önermiştir. Ayrıca P_i ve P_j kümelerinin birleştirilmesinde n ve m gözlem sayısı iken,

$$S_{k,ij} = \left(\frac{n}{n+m} \right) S_{ki} + \left(\frac{m}{n+m} \right) S_{kj} + \left(\frac{n \cdot m}{(n+m)^2} \right) \cdot (1 - S_{ij}) \quad (2.1)$$

ilişkisinin kullanılmasını önermektedir. Burada,

S_{ij} = i. birimin j. birime olan uzaklığıdır

S_{ki} = k. birimin i. birime olan uzaklığıdır

S_{kj} = k. birimin j. birime olan uzaklığıdır

$S_{k,ij}$ = k. birimin daha önceden birleştirilmiş ij. birimine olan uzaklığıdır

Yine aynı çalışmasında, kümelerdeki gözlem sayısına bakmadan elde edilen doğrusal ilişki ile Ortanca yöntemini önermiştir.

$$S_{k,ij} = \frac{1}{2} \cdot (S_{ki} + S_{kj}) + \frac{1}{4} \cdot (1 - S_{ij}) \quad (2.2)$$

Lance and Williams (32), Grup Ortalama yönteminde yenileme matrisi oluşturulurken d_{ij} gözlemleri ile d_k gözlemlerinin aritmetik ortalamaları alındığını, ancak benzerlik türü matrislerde ortalama almanın doğru olmadığını, bunun yerine

$$d_{ij} = \cos \left[\frac{1}{n_i n_j} \cdot \sum_{i,j} \cos^{-1} \cdot S_{ij} \right] \quad (2.3)$$

ilişkisini önermişlerdir.

Bağcı ve ark. (33), çalışmalarında *Bacillus thuringiensis*'in 27 varyetesi ile *Bacillus sphaericus* varyetesi olmak üzere 28 varyetenin 22 değişik antibiyotiğe karşı olan duyarlılıklarını ölçerek, milimetre cinsinden elde edilen değerleri standardize

ederek Öklid ve Öklid uzaklıklarının karesini (Öklid kare) hesaplamışlardır. Öklid uzaklığını kullanarak, Tek Bağlantı, Tam Bağlantı, Gruplar Arası Ortalama Bağlantı ve Gruplar İçi Ortalama Bağlantı kümeleme metotlarını bireyleri kümelere ayırmışlardır. Kümeleme çözümlerini ise kofenitik korelasyon katsayıları ölçütünü kullanarak değerlendirmişlerdir. Sonuçta Gruplar Arası Ortalama Bağlantı yöntemi ile Tek Bağlantı kümeleme yönteminin benzer sonuçlar verdiğini ve her iki metotta da dokuz ayrı küme oluştuğunu saptamışlardır. Bununla birlikte, bütün metotlarda *Bacillus Sphaericus* varyetesinin, *Bacillus Thuringiensis* varyetelerinden ayrı birer sınıfta yer aldığını bildirmişlerdir.

Arnold (34), toplam varyans kovaryans matrisinin (T) determinantı, küme içi varyans kovaryans matrisinin (W) determinantına oranının maksimum değerlerinin logaritması olarak alınan bir C katsayısı hesaplanarak kümenin bu kritere göre parçalanması gerektiğini ileri sürmüştür ve bu istatistiğin,

$$C = \log \left[\max \left(\frac{|T|}{|W|} \right) \right] \quad (2.4)$$

olarak hesaplanması gerektiğini bildirmiştir.

Punj and Stewart (35), yağma kümeleme yöntemi (Hill Climbing Method) üzerinde çalışarak, birimlerin en yakın ortalamalı kümeye atanması yerine önceden belirlenen bir istatistiksel kritere göre bir kümeden diğerine hareket ederek en uygun durum sağlanıncaya kadar yeniden atama işlemlerinin sürdürüldüğünü beyan etmişlerdir. Burada sözü edilen istatistiksel kriterin toplam küme içi varyans kovaryans matrisi (W) ve kümeler arası varyans kovaryans matrisi (B) ile ilgili olarak geliştirilebileceğini ileri sürmüşlerdir. Bu kriterin,

$$\text{tr } W, \text{tr} \left[\left(W^{-1} B \right) \right], |W|, \left[\left(W^{-1} B \right) \right] \text{'nin}$$

en büyük özdeğeri (eigen value) olarak alınabileceğini ifade etmişlerdir. Ayrıca n tane birimin parçalanıp k-kümeye ayrılarak C katsayısının optimum olduğu değere ulaştığında elde edilen k kümenin, n birimin en uygun kümeleneceği olarak ileri sürülmüştür.

Dinçer ve Özdamar (36), yaptıkları çalışmada g_1 , g_2 , g_3 , r_{cs} , Wilk's Lambda, ve Hotelling Lawley İz istatistikleri rasgele çekilmiş 10, 20, 30, 40 ve 50 birimlik gruplar ile şartlı çekilmiş 20, 30, 40, 50 ve 60 birimlik gruplar üzerinde test edilmiştir. Altı değişkene (karakter) göre değerlerin Öklid uzaklıkları kullanılarak benzerlik matrisleri elde edilmiş ve küme sayısı 2 ile 5 arasında olacak şekilde k-ortalama yöntemi ile kümeleyip, Tek Bağlantı yöntemi ile bağlantıları belirlemişlerdir. Rasgele ve şartlı gruplarda, kümeleme ölçütleri gruplarda benzerlik ve uyuma göstermemiştir. Rasgele grupların birim sayıları artarken Wilk's Lambda ve Hotelling Lawley İz istatistikleri dışındaki kümeleme ölçütleri küme sayılarında düzensiz artış ve azalma olduğunu bildirmişlerdir. Bu yüzden en uygun kümelemeyi belirleme ölçütü olarak Wilk's Lambda ve Hotelling Lawley İz istatistiklerinin olduğunu belirtmişlerdir.



3. MATERYAL ve METOT

Bu bölümde, değerlendirmeye alınan veriler kısaca tanımlandıktan sonra, kümeleme analizinde kullanılacak veri setinin yapısı, kümeleme analizinin yapılabilmesi için gerekli olan dönüşümler ve ilgili uzaklık ölçütleri tanımlanmış ve karşılaştırılacak olan yöntemler anlatılmıştır. Ayrıca kümeleme analizini çözümlenmede yardımcı olan diğer çok değişkenli analiz yöntemlerine de yer verilmiştir.

3.1. Materyal

Bu çalışmanın verilerini, Şanlıurfa TİGEM Ceylanpınar Tarım İşletmesinde yapılan bir araştırmanın sonuçları oluşturmuştur (37). Ölçümler, 24 farklı erkek Antepfıstığı (*Pistacia vera* L.) tipleri üzerinde gerçekleştirilmiş olup, fıstık ağaçlarını sınıflandırmak amacıyla 38 adet farklı karakterden ölçüm alınmıştır. Çalışmada kullanılan veriler EK Çizelge 4.3'te verilmiştir.

3.2. Metot

Kümeleme analizi için ilgili verilerin yapısı ve metotların istatistiksel formülleri aşağıda verilmiş olup, bununla birlikte bazı çok değişkenli kümeleme kriterleri de ele alınmıştır.

3.2.1. Sayısal sınıflandırma analizinde kullanılacak veri setinin yapısı

Sayısal sınıflandırma analizinde, bir bireyin birden fazla karakterine ilişkin ölçümler alınmaktadır. Bu veriler ölçülerek, sayılarak veya adlandırılarak elde edilebilir. Değerlendirilen karakterlerin özelliğine göre ölçüm ıskalası ve değişken tipi değişmektedir. Ölçümü yapılan karakterlerin nitel olması halinde ölçüm ıskalası” adlandırma (nominal) veya sıralama (ordinal) ölçekli, değişken tipi ise kesiklidir. Çünkü bu tip karakterlerin sonuçları daha çok sayım ile belirlenir.

Yapılan tasnifte bir dereceleme söz konusu değilse ve sınıfların bir üstünlüğü yok ise adlandırma ölçüm ıskalası kullanılır. Örneğin, Hastalar-normaller, erkekler-

kızlar, sigara içenler-içmeyenler gibi. Ancak bazen sınıflamada bir mertebelenme vardır. Örneğin, bir kovandaki arıların abdomenlerindeki kıl uzunluklarının uzun tüylü, kısa tüylü, orta tüylü şeklinde sınıflandırılması. Bu durumda ölçüm ıskalası sıralama (ordinal) ıskalasıdır (38).

Sürekli değişkenlerde kullanılan iki ayrı ölçek vardır. Bunlar; aralık ölçeği ve oran ölçeğidir.

Kısacası özetleyecek olursak;

- Adlandırma (Nominal) veya sınıflayıcı ölçek
- Sıralayıcı (Ordinal) ölçek
- Aralıklı (Interval) ölçek
- Oran (Ratio) ölçek

Burada adlandırma ve sıralayıcı ölçekler, sözel değişkenler için kullanılır ve kesikli özelliktedir. Aralık ve oran ölçekler ise sayısal değişkenleri ölçmekte kullanılır ve sürekli özelliktedir (8), (38).

Sayısal sınıflandırma analizinde kullanılan veriler çok değişkenli veriler olup, bunlar $n \times p$ boyutlu X veri matrisi ile gösterilir. Yani, n tane bireyin p tane karakterine ilişkin bir veri matrisidir. Şekil 3.1’de verilen matriste bireyleri satırlar karakterleri ise sütunlar göstermektedir.

Bireyler		KARAKTERLER						
		1	2	3	.	.	.	p
B	1	X_{11}	X_{12}	X_{13}	.	.	.	X_{1p}
I	2	X_{21}	X_{2p}
R	3	X_{31}	X_{3p}
E
Y
L
ER	n	X_{n1}	X_{n2}	X_{n3}	.	.	.	X_{np}

Şekil 3.1. Çok değişkenli veri matrisi şeması

$$X = [X_{ij}]_{np}, \quad i=1,2,\dots,n \quad j=1,2,\dots,p$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & x_{1P} \\ x_{21} & \cdot & \cdot & \cdot & x_{2P} \\ x_{31} & \cdot & \cdot & \cdot & x_{3P} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdot & \cdot & x_{nP} \end{bmatrix}_{n \times p}$$

Bireylere ilişkin karakterler oran ölçeğinde verilmesi halinde, veri matrisinde bu değerler olduğu gibi alınır. Ancak bütün karakterlerin nicel özellikte olması beklenemez. Bazı karakterler nitel özellikte olup ölçümler adlandırma veya sıralayıcı ölçekli olabilir. Bu durumda değişkenler kodlanarak (puan verilerek) sayısallaştırılabilir. Örneğin, arıların abdomenlerine ilişkin kıl uzunlukları kısa, uzun ve orta şeklinde ölçümü yapılmış ise, veri matrisinde bu isimler yerine {0,1,2...} gibi tam sayılı puanlar verilir.

Bu işlemlerden sonra elde edilen $n \times p$ boyutlu X veri matrisi ham verilerden ibarettir. Çünkü her bir karakterin ölçümünde kullanılan ölçü birimi farklıdır. Hatta aynı ölçüm birimine sahip olsa dahi değişkenlerin dağılım aralıkları farklı olabilir. Bu nedenle, analize geçmeden önce değişkenler arasındaki sözü edilen farklılığın giderilmesi gerekir.

Dönüştürme işlemleri aşağıdaki şekillerde yapılabilir,

- 1- Standart normal değişkene dönüştürme
- 2- Değişkenlerin belli aralığa indirgenmesi
- 3- Ağırlıklı değişkene dönüştürme
- 4- Değişkenlerin ortalamaya göre ölçeklendirilmesi

Bunlardan genellikle sayısal sınıflandırma analizinde standart normal değişkene dönüştürme işlemi daha yaygın olarak kullanılmaktadır (12), (39).

Yukarıdaki açıklamalara örnek verecek olursak; veriler binari (ikili) karakterlerden ibaret ise örneğin; tohumun yüzü tüylü ve tüysüz diye adlandırılmışsa sözü edilen karakterler veri matrisinde tüylü tohum (1), tüysüz tohum (0) olarak kodlanabilir. Bazı yazarlar ise aynı örneği sırasıyla (+), (-) olarak kodlamaktadırlar. Eğer karakterler çok durumlu (multistate) nitel karakterler ise, örneğin; kısa, uzun ve çok uzun olarak kayıt edilen meyve boyu düzenli multisate bir karakter olup

puanlama (skorlama); kısa (1), uzun (2), çok uzun (3) şeklinde yapılır. Çiçek renginin kırmızı, mavi ya da beyaz olabilirdiği ise düzensiz karakter tipini oluşturur. Bu durumda kırmızı (1), mavi (2), beyaz (3) olarak kodlanır.

Nicelik bildiren karakterler ise kesikli ve sürekli olabilir. Örneğin; bir çiçekteki petal sayısı kesikli bir rakam olup, bitkinin boyu, tohumun bindane ağırlığı ise sürekli bir rakam olup veri matrisine olduğu gibi kayıt edilir (5).

3.2.2. Benzerlik (similarity) ve uzaklık (distance measure) ölçütleri

Sınıflandırmada esas olan bireylerin benzerliklerine göre sınıflandırılmasıdır. Yani birbirine en çok benzeyenlerin değişik gruplar altında toplanması esasına dayanır. Bu benzerlik ölçüleri ise;

- 1- Uzaklık (distance like measures) türü ölçüler
- 2- İlişki katsayıları (Correlation like measures) türü ölçüler
- 3- Birliktelik (association) katsayıları
- 4- Olasılığa dayalı (based on probability) benzerlik ölçüleri ve diğerleri olarak sıralanabilir (13).

Kümeleme analizinde veri setinin oluşturulmasından sonra ilk adım, bireyler veya karakterler arasındaki benzerlik katsayısının saptanmasıdır. Burada bireyler veya karakterler denmesinin sebebi, araştırmacı bireyleri sınıflandırmak istediği gibi bazen de karakterleri sınıflandırmayı arzu edebilir.

Sözü edilen benzerlik katsayılarının saptanması, bütün gözlemler için ikili gözlemler arası benzerlik katsayıları hesaplanarak $n \times n$ (bireyler arası) veya $p \times p$ (karakterler arası) boyutlu bir benzerlik matrisinin dizayn edilmesi ile mümkündür (8).

Uzaklık (distance) matrisi D, elemanlarını ise;

$$d_{ij} = d(x_i, x_j) \quad (i, j = 1, 2, \dots, n)$$

ile Şekil 3.2'deki gibi gösterilmiştir. D uzaklık matrisi $n \times n$ boyutlu olup,

$$C(n;2) = \frac{n(n-1)}{2}$$

tane uzaklık değerinden oluşmaktadır.

$$D_{n \times n} = \begin{bmatrix} 0 & d_{12} & d_{13} & \dots & \dots & \dots & d_{1n} \\ & 0 & d_{23} & \dots & \dots & \dots & d_{2n} \\ & & 0 & \dots & \dots & \dots & \dots \\ & & & 0 & \dots & \dots & \dots \\ & & & & 0 & \dots & \dots \\ & & & & & 0 & d_{(n-1)n} \\ & & & & & & 0 \end{bmatrix}_{n \times n}$$

Şekil 3.2. $D_{n \times n}$ boyutlu uzaklık matrisi şeması

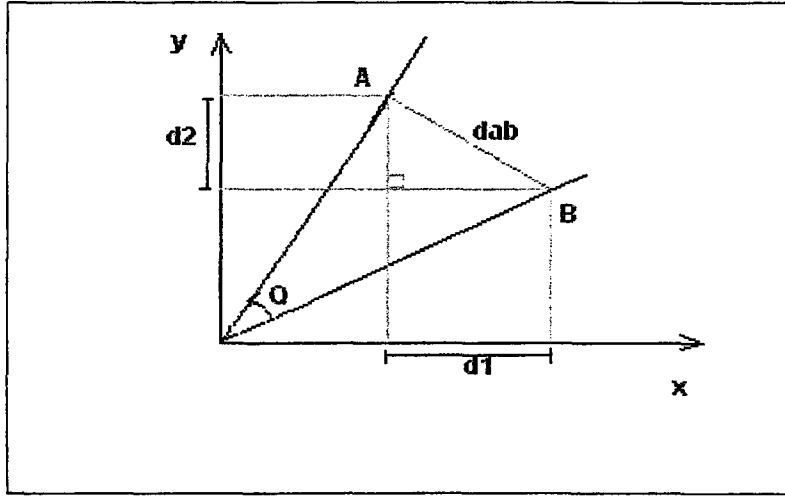
Birimler arasındaki mesafenin tespit edilmesinde en çok kullanılan ölçü, uzaklık (distance) ölçüleridir. Benzerlik matrisinin elemanları, D matrisinin elemanlarına göre belirlenir. Benzerlik matrisi (S) ile gösterilirken, elemanları ise $S_{(ij)}$ ile gösterilecektir. Benzerlik matrisi elemanları;

$$S_{(ij)} = 100 \left(\frac{1 - d_{(ij)}}{\max [d_{(ij)}]} \right)$$

biçiminde hesaplanır (11), (12).

Benzerliğin sayısal olarak ifade edilmesi metrik sistem kavramıyla açıklanabilir.

Bireyler koordinat sisteminde noktalarla gösterilir ve bu noktalar arasındaki benzerlikler, noktaların birbirine olan metrik uzaklığı kadardır. Bunu iki boyutlu bir uzayda A ve B gibi iki bireyin birbirine olan uzaklığını Şekil 3.3'deki gibi gösterebiliriz.



Şekil 3.3. Uzayda iki noktanın birbirine olan uzaklığının gösterimi

$d_{ab}^2 = d_1^2 + d_2^2$ olarak belirlenir. Çok değişkenli bir yapıda n bireyin p karakteri için çok boyutlu bir uzayda yerleri vektör olarak belirtilir. İki gözlem değerinin çok boyutlu uzayda birbirine uzaklıkları Pisagor teoreminin genellemesine göre belirlenir (6), (40).

Uzaklık ölçüleri gözlem çiftlerinden negatif olmayan reel sayılara tanımlı bir fonksiyondur. Aynı zamanda her uzaklık, metrik bir ölçüdür. Bu neden ile önce metrik tanımını vermekte yarar vardır. Burada vereceğimiz x_i , i birey. x_j , j birey; i ve j birey arasındaki uzaklık $d(x_i, x_j)$ olarak alınacaktır.

Tanım: X gözlemlerinin bir kümesi ve $d: X \times X \Rightarrow \mathbb{R}^+$ biçiminde tanımlı bir fonksiyon olsun. Eğer d fonksiyonu aşağıda verilecek olan şartları sağlıyorsa, o zaman d fonksiyonuna uzaklık ölçüsü denir.

1- Her $x_i, x_j \in X$ için $d(x_i, x_j) \geq 0$

Yani bütün x_i, x_j çiftleri arasındaki uzaklık her zaman pozitiftir.

2- Her $x_i, x_j \in X$ için $d(x_i, x_j) = d(x_j, x_i)$

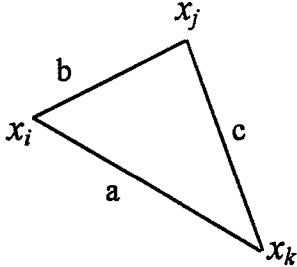
Yani uzaklık fonksiyonunun simetrik özelliği vardır. (Bu nedenle benzerlik matrisi üst veya alt üçgen matrisi olarak alınabilir.)

3- Her $x_i, x_j \in X$ için $x_i = x_j$ ise $d(x_i, x_j) = 0$

[$x_i \neq x_j$ ise, $d(x_i, x_j) \neq 0$]

Yani, aynı birimler arasındaki uzaklık 0'dır. (Farklı birimler arasındaki uzaklık 0'dan farklıdır.)

$$4- \text{ Her } x_i, x_j, x_k \in X \text{ için } d(x_i, x_k) \leq d(x_j, x_i) + d(x_j, x_k)$$
$$a \leq b + c \text{ olup,}$$

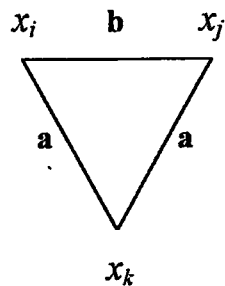


Üçgen eşitsizliği adı verilen bu özelliğe, x_i ve x_j birimleri arasındaki uzaklık, üçüncü bir nokta olan x_k noktası alındığında, x_i ile x_k ve x_k ile x_j arasındaki uzaklıkların toplamından küçüktür (8), (41), (42).

Johnson (43), 5. bir özellik tanımlayarak aşamalı yapının bu özelliğe bağlı olduğunu belirtmiştir. Buda ultrametrik özelliğidir.

$$5- \text{ Her } x_i, x_j, x_k \in X \text{ için } d(x_i, x_j) \leq \max \{d(x_i, x_k), d(x_k, x_j)\}$$

Bu özellik gereği X içindeki her (x_i, x_j, x_k) noktaları arasındaki uzaklık, bir ikiz kenar üçgen oluşturur ve eşit kenarların uzunluğu diğer kenardan daima büyük ya da eşittir.



Çalışmada kullanılacak uzaklık ölçütleri sırasıyla aşağıda verilmiştir.

3.2.2.1. Minkowsky uzaklık ölçütü

Sayısal sınıflandırma analizinde nicel veriler için kullanılan uzaklık ölçütü, Minkowsky ölçütü olup aynı zamanda genel uzaklık ölçüsüdür. x_i ve x_j (p) tane karaktere göre ölçülmüş i . ve j . bireyler olarak tanımlandığında, Minkowsky ölçütü;

$$d_{\lambda}(x_i, x_j) = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^{\lambda} \right]^{1/\lambda}, \quad \lambda \geq 1 \text{ için; } (i, j = 1, 2, \dots, n) \quad (3.1)$$

olarak verilebilir (6).

3.2.2.2. Manhattan (City- Blok) uzaklık ölçütü

Bu ölçüt Minkowsky uzaklık ölçütünün (3.1), $\lambda=1$ olduğu durumdaki özel hali Manhattan uzaklık ölçütünü verir.

$$d_1(x_i, x_j) = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^1 \right]^{1/1} \Rightarrow d_1(x_i, x_j) = \left[\sum_{k=1}^p |x_{ik} - x_{jk}| \right] \quad (3.2)$$

3.2.2.3. Euclid (Öklid) uzaklık ölçütü

Bu ölçüt Minkowsky uzaklık ölçütünün (3.1), $\lambda=2$ olduğu durumdaki özel hali Öklid uzaklık ölçütünü verir.

$$d_2(x_i, x_j) = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^2 \right]^{1/2} \quad (3.3)$$

Özdamar (12), Mariott (1974) ve Kendall (1975)'in Öklid uzaklığında gözlem değerleri standardize edildikten sonra kullanıldığından, karakterlerin farklı ölçü

birimleri ile ölçülmüş olmasından etkilenmediğini belirtmiştir. Bu üstünlüğünden dolayı, sayısal sınıflandırma analizinde en sık kullanılan uzaklık ölçüsü olduğunu, ayrıca gruplar arası hata kareler ortalamasını minimize ettiği için genellikle tercih edildiğini bildirmiştir.

Everitt (11), bir karakterin ölçüm birimi değiştirildiğinde Öklid uzaklığının ham verilerden ziyade,

$$Z_{ij} = \frac{x_{ij} - \mu}{\sigma_j} \quad i=1,2,\dots,n \quad j=1,2,\dots,p \quad (3.4)$$

biçiminde standartlaştırılmış veriler üzerinde kullanılmasının daha tatmin edici olduğunu belirtir. Burada;

x_{ij} ; i . bireyin j . karakterine ait gözlem değeri,

σ_j ; j . karaktere ait standart sapmadır.

Öklid uzaklığının bazı özelliklerini aşağıdaki maddeler halinde sıralayabiliriz.

a) Pozitif yarı tanımlılık özelliği vardır. Elemanları, Öklid uzaklığından bulunan herhangi bir simetrik matris,

$$A = -\frac{1}{2} \left[d_{ij}^2 \right] \quad 1 \leq i, j \leq n$$

olsun. I birim matris ve $\underline{\ell}^T = [1 \ 1 \ \dots \ 1]$ birlerden ibaret bir vektör olmak üzere,

bir merkezileştirme matrisi,

$$H = I - \left(\frac{\underline{\ell} \cdot \underline{\ell}^T}{n} \right)$$

biçiminde ölçeklendirildiğinde, $C = HAH$ şeklinde tanımlanan matris pozitif yarı tanımlıdır (44).

b) $i, j = 1, 2, \dots, n$ için d_{ij} Öklid uzaklıkları x_j ve x_i . gözlem vektörleri dik dönüşümlerden etkilenmezler.

Bazı kümeleme yöntemlerinde bireyler arasındaki benzerliği hesaplamak için Öklid uzaklığına ait eşitliğin karesi alınarak işlem yapılır. Bu durumda ölçüt Öklid kare uzaklığı olarak adlandırılır (8).

$$(d_2(x_i, x_j))^2 = \left(\left[\sum_{k=1}^p |x_{ik} - x_{jk}|^2 \right]^{1/2} \right)^2 \Rightarrow$$

gerekli sadeleştirmeler yapıldıktan sonra,

$$d_2^2(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}|^2 \quad (3.5)$$

elde edilir.

3.2.2.4. Tartılı Euclid (Öklid) uzaklık ölçütü

Bireylere ilişkin karakterlerin aynı ağırlıkta ölçeklenmemiş olması durumunda Öklid uzaklığının standartlaştırılması ile yeni bir uzaklık olan tartılı Öklid veya Karl Person uzaklığı elde edilir.

$$d_2^*(x_i, x_j) = \left\{ \sum_{k=1}^p W_k^2 (x_{ik} - x_{jk})^2 \right\}^{1/2} \quad \text{veya} \quad W_k = \frac{1}{S_k} \quad \text{ise}$$

$$d_2(x_i, x_j) = \left\{ \sum_{k=1}^p \left(\frac{x_{ik} - x_{jk}}{S_k} \right)^2 \right\}^{1/2} \quad i, j = 1, 2, \dots, n \quad (3.6)$$

Burada, $W_k = 1/S_k$ k. karakterin standart sapma değerinin (S_k), ya da değişim aralığı değerinin tersidir. W_k 'nin S_k değerinin tersi olması durumunda elde edilen uzaklığa literatürlerde genellikle Karl-Pearson uzaklığı da denmektedir. Karakterlerin bağımsız olması durumunda 3.6 No'lu eşitliğindeki gibi tanımlanan Karl-Pearson uzaklığı; bağımsızlık varsayımı gerçekleşmemesi durumunda;

$$d(x_i, x_j) = \left[\sum_{k=1}^p a_{kk} (x_{ik} - x_{jk})^2 + 2 \sum_{k < \ell = 1}^p a_{k\ell} (x_{ik} - x_{jk})(x_{\ell i} - x_{\ell j}) \right]^{1/2} \quad (3.7)$$

biçiminde tanımlanmaktadır.

Burada; $a_{k\ell}$, $1 \leq k, \ell \leq p$, katsayıları uzaklığı negatif yapmayacak biçimde sayılar olup, koordinat eksenlerine uygulanan θ dönme açısı ile karakterlerin varyans ve kovaryanslarına bağlı olarak değer alır (6), (17).

3.2.2.5. Mahalanobis- D^2 uzaklık ölçütü

Eşitlik 3.7 ile verilen uzaklığın en genel biçimi Mahalanobis- D^2 'si ya da genelleştirilmiş Öklid uzaklığı olarak bilinmektedir. $a_{k\ell}$, $1 \leq k, \ell \leq p$ katsayılarının oluşturduğu matris V varyans kovaryans matrisi olmak üzere herhangi iki x_j ve x_i gözlemleri arasındaki Mahalanobis- D^2 uzaklığı,

$$d(x_i, x_j) = D^2 = (x_i - x_j)^T \cdot V^{-1} \cdot (x_i - x_j) \quad (3.8)$$

biçiminde ifade edilir. Bu uzaklık iki Küme arasında;

$$d(\bar{x}_i, \bar{x}_j) = D^2 = (\bar{x}_i - \bar{x}_j)^T \cdot S_V^{-1} \cdot (\bar{x}_i - \bar{x}_j) \quad (3.9)$$

ile ifade edilir (11), (41).

Burada; S_V = Örneğe ait iki kümeden hesaplanan birleştirilmiş (pooled) kovaryans matrisidir. 3.9 No'lu eşitlikten elde edilen D^2 uzaklığı yardımıyla test istatistiğine ait F_h kritik değeri aşağıdaki gibi hesaplanabilir.

$$F_h = \frac{n_1 \cdot n_2 (n_1 + n_2 - p - 1)}{p \cdot (n_1 + n_2) \cdot (n_1 + n_2 - 2)} \cdot D^2 \quad (3.10)$$

Bu değer,

$$F_c = F_{\alpha; p, n_j + n_i - p - 1} \quad (3.11)$$

cevvel değeri ile karşılaştırılır. Burada;

n_i ve $n_j = i.$ ve $j.$ kümelerdeki birim sayısı

$p =$ her bir birime ilişkin ölçülen karakter sayısıdır (41).

Mahalanobis- D^2 uzaklığı deęişken uzayına uygulandıęında Öklid uzaklıęının tam tersine ölçme ölçeęinden etkilenmez. Ayrıca Öklid uzaklıęı gibi dik dönüşümlerden de etkilenmez (45). D^2 uzaklıęına çok benzer bir uzaklık ölçütü Rolf (46) tarafından tanımlanmıştır. i kümesinden j kümesine uzaklıęın bir ölçüsünü veren bu ölçüt $i, j = 1, 2, \dots, n$ için,

$$D = \left[\left(\bar{x}_i - \bar{x}_j \right)^T \cdot S_i^{-1} \cdot \left(\bar{x}_i - \bar{x}_j \right) \cdot |S_i| \right]^{1/2} \quad (3.12)$$

biçimindedir. Burada,

\bar{x}_i ve \bar{x}_j ; p adet karakter üzerinde i ve j kümeleri 3.9 No'lu eşitlikle verilen

Mahalanobis- D^2 için ortalamaları kapsayan sütun vektörleri,

S_i ; p deęişken üzerinde i kümesi için küme içi varyans- kovaryans matrisidir,

$|S_i|$; S_i matrisinin determinantıdır.

3.12 No'lu eşitlięi 3.8 veya 3.9 eşitlięine göre iki bakımdan farklılık göstermektedir. Bu farklılıklar sırasıyla;

1- 3.12 No'lu eşitlięinde birleştirilmiş varyans-kovaryans matrisi yerine sadece bir grubun varyans-kovaryans matrisinin kullanılmış olmasıdır. Rolf (46) bunun nedenini farklı varyans-kovaryans matrislerinin homojenlik durumu ile ilgili varsayımda bulunabilmek için hiç bir fikre sahip olmaması mantıęını kullanmaktadır.

2- Sınıflandırma analizinde hacim ölçütü hesabında kullanılan $|S_i|$ 'nin formüle dahil edilmesidir (17).

3.2.2.6. Chebychev uzaklık ölçütü

Minkowski metrięinin; $\lambda = \infty$ için,

$$d_\infty(x_i, x_j) = \sup \left\{ |x_{ik} - x_{jk}| \right\} \quad k=1, 2, \dots, p \quad (3.13)$$

Chebychev uzaklıęı elde edilir.

3.2.2.7. Canberra uzaklık ölçütü

Lance ve Williams (32) Manhattan uzaklığının bir türevi olan Canberra uzaklığını önermişlerdir.

$$d(x_i, x_j) = \frac{\sum_{k=1}^p |x_{ik} - x_{jk}|}{\sum_{k=1}^p (x_{ik} + x_{jk})} \quad (3.14)$$

Ancak bu uzaklık ölçütü uygulamada pek yaygın olarak kullanılmamaktadır.

Uygulamada bütün değişkenlerin sürekli yapıda olması mümkün değildir. Bazı karakterlerin kesikli (adlandırma, sıralayıcı) özellikte olduğu durumlarla da karşılaşmaktadır. Böyle durumlarda şimdiye kadar verilen formüller doğrudan kullanılamamaktadır. Onların yerine

$$W_k \begin{cases} 1; \text{ nicel veriler için} \\ \frac{1}{k. \text{ de\u0131işkenin de\u0131işim aralığı}}; \text{ nitel veriler için} \end{cases}$$

olmak üzere

$$d(x_i, x_j) = \frac{1}{p} \cdot \sum_{k=1}^p W_k |x_{ik} - x_{jk}| \quad (3.15)$$

formülünden bulunmaktadır. Nicel veriler (Yukarıda verilen tüm formüller) için geliştirilmiş formüllerin tümünde W_k tartı değeri kullanılarak uzaklıklar yaklaşık olarak hesaplanabilir (41), (8), (11).

3.2.2.8. Hotelling T^2 uzaklık ölçütü

Bu uzaklık ölçütü iki grup ortalama vektörün karşılaştırılmasında kullanılmaktadır.

$$T^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} (\bar{x}_i - \bar{x}_j)^T S^{-1} (\bar{x}_i - \bar{x}_j) \quad (3.16)$$

Burada sırasıyla \bar{x}_i, \bar{x}_j ; i . ve j . kümelere ilişkin $p \times 1$ boyutunda karakter ortalamaları vektörü,

n_1, n_2 ; 1.ve 2. Kümenin birey sayılarıdır.

S; bileşik kovaryans matrisi olup A_1 ve A_2 her iki gruba ilişkin kareler ve çarpımlar toplamı matrisleri olmak üzere;

$$S = \frac{1}{n-2}(A_1 + A_2) \quad (3.17)$$

eşitliği ile bulunur (41).

3.16 No'lu eşitliğinden elde edilen T^2 değeri ya doğrudan kullanılarak ya da F değerine dönüştürülerek hipotez testi sonuçlandırılır. T^2 tablo değerinin kolay bulunamaması nedeniyle uygulamalarda genellikle F dönüşümü yapılarak,

$$F = \frac{n_1 + n_2 - p - 1}{p \cdot (n_1 + n_2 - 2)} \cdot T^2 \Rightarrow \frac{n - p - 1}{p \cdot (n - 2)} \cdot T^2 \sim F_{\alpha; p, n-p-1} \quad (3.18)$$

F tablo değeri kritik değer olarak kullanılabilir. Aynı şekilde

$$T^2 = \frac{p(n-2)}{n-p-1} \cdot F_{\alpha; p, n-p-1} \quad (3.19)$$

eşitliğinden elde edilen T^2 değeri de 3.16 No'lu formülden elde edilen T^2 test istatistik değeri ile karşılaştırılarak karar verilebilmektedir. Bu karşılaştırmalar sonucunda H_0 hipotezi ret edilmeyene kadar işleme devam edilmektedir. Ret edilme durumunda ise farklılığa neden olan karakter ya da karakterlerin belirlenmesi gerekir (41).

3.2.2.9. Vektör çarpım uzaklık ölçütü

Bu ölçütü iki boyutlu uzayda noktalar arasındaki veri vektörleri ve görsel uzunlukları arasındaki açısal farkın benzerlik ölçüsü olarak alındığı bir ölçüdür.

Bir veri setinde X ve Y değişkenlerine ilişkin sıra vektörleri,

$$X^T = (X_1, X_2, \dots, X_n) \quad \text{ve} \quad Y^T = (Y_1, Y_2, \dots, Y_n)$$

olarak gösterilebilir.

Gözlem uzayında iki vektörün içsel çarpımı veya scalar çarpımı (inner product or scalar product),

$$\langle X, Y \rangle = X^T \cdot Y = \sum_{i=1}^n X_i Y_i \quad (3.20)$$

olarak ifade edilir ve $X^T \cdot Y$ 'nin karekökü, vektör uzunlukları olarak ele alınır.

Bir vektörün kendisinin transpozu ile çarpımı ($X^T \cdot X$), X'lerin kareler toplamı olarak adlandırılır ve genellikle $\|X\|$ şeklinde gösterilir. X ve Y'nin çarpımı kareler toplamı cinsinden tekrar ele alındığından

$$X^T Y = \|X\| \cdot \|Y\| \cdot \cos \alpha \quad (3.21)$$

eşitliği elde edilir. Burada $\cos(\alpha)$ iki nokta arasındaki uzaklığın ölçüsü olarak ele alınmaktadır ve

$$\cos \alpha = \frac{X^T Y}{\|X\| \cdot \|Y\|} = \frac{\sum_{i=1}^n X_i Y_i}{\left[\left(\sum_{i=1}^n X_i^2 \right) \cdot \left(\sum_{i=1}^n Y_i^2 \right) \right]^{1/2}} \quad (3.22)$$

şeklinde hesaplanır. $\cos \alpha$ değeri X ve Y arasındaki benzerliğin bir ölçütü olarak alınacak. Bireyler yerine, bireylere ilişkin karakterler kümelenecek istendiğinde yararlanılmaktadır (8), (12).

Diğer bir özelliği ise; iki vektör arasındaki açı ölçüm vektörlerinin uzunluklarından etkilenmez. Veri matrisi pozitif kabul edildiğinde açı, $(0.0 < \cos(\alpha) < 1.0)$ arasında değişmektedir.

Çetinel (47)'in bildirdiğine göre İmbrie ve Pudry, bu ölçüyü Q-Analizi olarak adlandırılan Sayısal Sınıflandırma veya Kümeleme analizinde benzerlik ölçüsü olarak kullanmışlardır.

3.2.3. İlişki türü benzerlik ölçüleri

Herhangi iki birey arasındaki ilişkinin bir ölçüsü olan benzerlik katsayıları $(0,1)$ aralığında değer alır. Bu katsayılar, verilerin ikili nitel ve nicel türden olmalarına göre farklı şekillerde tanımlanmaktadır.

3.2.3.1. Korelasyon katsayıları (nicel veriler için iki değerli değişkenler)

Korelasyon Katsayıları ile benzerliğin belirlenmesi, herhangi iki gözlem (birey ya da karakter) arasındaki ilişki göz önüne alınarak hesaplanan benzerlik ölçüleridir.

En yaygın olarak kullanılan Pearson ilişki katsayısıdır. Söz konusu bu ilişki katsayısı ölçü birimi içermeyen, ölçekten bağımsız birlikte değişimin bir ölçüsüdür.

X ve Y iki gözlem değeri olmak üzere bunlar arasındaki korelasyon katsayısı,

$$r_{(X,Y)} = \frac{\text{COV}(X, Y)}{\sqrt{[\text{Var}(X) \cdot \text{Var}(Y)]}} = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\left\{ \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] \cdot \left[\sum_{i=1}^n (Y_i - \bar{Y})^2 \right] \right\}^{1/2}} \quad (3.23)$$

olarak hesaplanmaktadır.

p tane karaktere göre i . birey ile j . birey arasındaki ilişki katsayısı,

$$r_{ij} = \frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i) \cdot (x_{jk} - \bar{x}_j)}{S_i \cdot S_j} \quad (3.24)$$

olarak bulunur (11), (42).

Burada;

$S_i = i$. bireyin (p) tane karakter için aldığı değerlerin standart sapmasıdır.

$S_j = j$. bireyin yine aynı (p) karakterler için aldığı değerlerin standart sapmasını vermektedir.

Pearson ilişki katsayısı $-1 < r < 1$ arasında değerler alır. r 'nin 0 (sıfır) olduğu durumda, gözlemler arasında herhangi bir ilişki olmadığını ifade eder.

Korelasyon katsayısı, kümeleme analizinde kullanılması bakımından, benzerlik ölçüleri içinde en çok tartışılan katsayılardan biridir. Korelasyon katsayısının hesaplanmasında bir gözlemin farklı değişkenler üzerindeki ortalaması kullanılmaktadır.

Çetinel'in (47) bildirdiğine göre, Eades, 1965'te yaptığı çalışmasında sınıflandırma problemlerinde korelasyon katsayısının kullanılmayacağını ve uzaklığa dayanan ölçütlerin daha iyi sonuç verdiğini belirtmiştir.

Ancak verilerin kodlanarak (yani 0-1-2... gibi puanlarla) elde edildiği durumlarda bireyler arasındaki benzerlik ölçüsü olarak Pearson ilişki katsayısını kullanılabileceği ifade edilmiştir (8).

3.2.3.2. İkili ve nitel verilerle ilgili benzerlik ölçüleri

Burada esas olan bu tür verilerde karakterlerin varlığına ve yokluğuna göre benzerlik katsayılarının hesaplanmasıdır. Bu aşağıda verilen bir örnekle açıklanabilir.

İki bireye ilişkin on tane karakterin varlığı {1} ile yokluğunda {0} olmak üzere kayıtları alınmış olsun. Örneğin karakterlerden biri tüylülük durumunu bildiriyorsa tüylü olan bireyin o karakteri {1} ile kodlanır, tüysüz olan bireyin konu olan o karakteri {0} olarak kodlanır. Diğer bütün karakterlerde benzer şekilde kodlanır. Sonra Çizelge 3.1'de verilen iki yönlü birleştirme (association) tablosu oluşturularak {1} ve {0} kodlu karakterlerin sayısı sayılarak tabloya kayıd edilir. Bu kodlar vasıtasıyla bireyler arasındaki ilişki hesaplanır. İki birey arasındaki benzerlik katsayısı U_{ij} ile gösterilecek olursa, U_{ij} 'yi hesaplamada Çizelge 3.1'de verilen iki yönlü birleştirme (association) tablosundan yararlanılacaktır (5)(11).

Çizelge 3.1. İki birey için iki yönlü birleştirme (association) tablosu

j .birey i .birey	1	0	Toplam
1	a	b	$a+b$
0	c	d	$c+d$
Toplam	$a+c$	$b+d$	$a+b+c+d = p$

Burada; $p = a+b+c+d$ olmak üzere karakter sayısını göstermektedir.

a : Her iki deneme biriminde ortak olan karakterlerin sayısı

b : i . deneme biriminde olup da, j . deneme biriminde olmayan karakterlerin sayısı

c : i . deneme biriminde olmayıp da, j . deneme biriminde olan karakterlerin sayısı

d : Her iki deneme biriminde olmayan karakterlerin sayısı

Bu terimleri aşağıdaki şekilde tanımladıktan sonra Çizelge 3.2'de ikili ve nitel verilere ilişkin bazı benzerlik katsayıları verilmiştir.

$$a = \sum_{k=1}^p x_{ik} \cdot x_{jk}$$

$$a + b = \sum_{k=1}^p x_{ik}$$

$$b = \sum_{k=1}^p x_{ik} (1 - x_{jk})$$

$$c + d = \sum_{k=1}^p (1 - x_{ik})$$

$$c = \sum_{k=1}^p (1 - x_{ik}) \cdot x_{jk}$$

$$a + c = \sum_{k=1}^p x_{jk}$$

$$d = \sum_{k=1}^p (1 - x_{ik}) \cdot (1 - x_{jk})$$

$$b + d = \sum_{k=1}^p (1 - x_{jk})$$

Örneğin iki birey arasındaki on tane karakter esas alınarak bunların varlığı ve yokluğu aşağıdaki şekilde kodlanmış olsun;

	Karakterler	1	2	3	4	5	6	7	8	9	10
Birey 1		1	0	0	0	1	1	0	0	1	0
Birey 2		0	0	0	0	1	0	0	1	1	0

Bu durumda iki yönlü birleştirme tablosu aşağıdaki şekilde oluşturulur ve ilgili U_{ij} değerleri hesaplanarak bireyler arasındaki benzerlik katsayıları hesaplanır.

i .birey \ j .birey	1	0	Toplam
1	2	1	3
0	2	5	7
Toplam	4	6	10

Anderberg (8), bazı benzerlik katsayılarını Çizelge 3.2’de ki gibi vermiştir.

Çizelge 3.2. Bazı benzerlik katsayıları

No	Benzerlik katsayıları	İsimler
1	$U_{ij} = \frac{a}{a+b+c}$	(Jakkard)* Pay ya da paydada 0-0 seçeneğine ait değer yok, yani (0-0) hücresi ihmal edilir.
2	$U_{ij} = \frac{a+d}{p}$	(Sokal ve Wichener- Roger ve Tanimoto) 1-1 ve 0-0 hücresi değerleri için eşit ağırlık
3	$U_{ij} = \frac{a+d}{b+c}$	(Sokal ve Sneath)
4	$U_{ij} = \frac{2a}{(a+b) \cdot (a+c)}$	(Krzanowski – Dice-Sorenson)
5	$U_{ij} = \frac{a}{p}$	(Russel ve Rao) Pay’da (0-0) hücre değeri yok
6	$U_{ij} = \frac{a}{2} \cdot \left(\frac{1}{a+b} + \frac{1}{a+c} \right)$	Kulczynski
7	$U_{ij} = \frac{a}{a+2 \cdot (b+c)}$	(Sokal ve Sneath) Pay ve paydada 0-0 hücre değeri yok, (1-0) ve (0-1) hücreleri için iki katı ağırlık verilir.
8	$U_{ij} = \frac{a}{\{(a+b) \cdot (a+c)\}^{1/2}}$	(Ochiai)
9	$U_{ij} = \frac{(ad-bc)}{(ad+bc)}$	Wilkinson,1984 Gama katsayısı

3.2.3.3. İki yönlü tablolarda birliktelik ölçütü

İki yönlü tablolarda iki adet değişken bulunup, bu değişkenlere ilişkin sınıflar genel olarak aşağıdaki Çizelge 3.3’de verilir.

Çizelge 3.3. İki yönlü tablolar

Y X	Y ₁ Y ₂ Y _q	Top
X ₁	f ₁₁ f ₁₂ f _{1q}	f _{1.}
X ₂	f ₂₁ f ₂₂ f _{2q}	f _{2.}
.	.	.
.	.	.
X _p	f _{p1} f _{p2} f _{pq}	f _{p.}
Top	f _{.1} f _{.2} f _{.q}	f _{..}

Burada X ve Y sınıflar olup ;

f_{ij} ise X_i ve Y_j sınıflarında ortaya çıkan frekanstır. İki yönlü tablolar genel olarak χ^2 tabloları olarak da adlandırılmakla birlikte, beklenen ve gözlenen frekanslar arasındaki çıkarım olup

$$f_{ij} = \frac{f_{i.} \cdot f_{.j}}{f_{..}} \quad (3.25)$$

ile tanımlanır. χ^2 değeri,

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q \left[\frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*} \right] \quad (3.26)$$

biçiminde verilir. f_{ij} = Deneysel, yani gözlenen frekans,

f_{ij}^* = Beklenen teorik frekans olup

$$f_{ij}^* = \frac{f_{i.} \cdot f_{.j}}{f_{..}} \quad (3.27)$$

biçiminde elde edilir (8).

χ^2 den yararlanarak, olabilirlik kareler ortalaması (olabilirlik katsayısı) olarak bilinen ϕ^2 aşağıdaki şekilde hesaplanır;

$$\phi^2 = \frac{\chi^2}{f \dots} \quad (3.28)$$

Bu ilişkiye dayanarak, Tschuprow ilişki katsayısı ve karesel olmayan tablolar için Cramer katsayısını örnek verilebilir (8).

2. Olabilirlik katsayısı ise

$$\phi^2 = \left[\frac{\phi}{1 + \phi} \right] = \left[\frac{\chi^2}{f \dots + \chi^2} \right] \quad (3.29)$$

olarak verilir (8), (17).

Özet olarak bireyler arasındaki benzerlikleri ölçmek için bir çok yol mevcuttur. Araştırmacıların çoğunun bireylere ait karakterleri sınıflandırmak için korelasyonları, bireyleri sınıflandırmak için ise Çizelge 3.2' deki katsayıları ya da uzaklık ölçütlerini kullandıkları görülmektedir (6).

3.2.4. Kümeleme yöntemleri

Bireylerin ya da bireylere ilişkin karakterlerin kümelenmesinde birçok yöntem geliştirilmiştir. Bu yöntemler, veri matrisinden elde edilen benzerlik matrisinin çözümlene şekline göre değişik şekillerde sınıflandırılmıştır.

Tarihsel gelişimi içinde kümeleme yöntemleri farklı şekillerde sınıflandırılmışlardır.

Çetinel (40), Ball (1966)'nın kümeleme yöntemlerini aşağıdaki şekilde sınıflandırdığını bildirmiştir.

- 1- Olasılıksal sınıflandırma
- 2- İşaret belirleyici
- 3- Kümeleme
- 4- Yığılma (clumping)
- 5- Özdeğer
- 6- En küçük tepe arayıcı (minimum mode seeking)
- 7- Diğer yöntemler şeklindedir.

Lance and Williams (32) ise,

- 1- Basitleştirici yöntemler
- 2- Öğelerin birden fazla grupta görüldüğü yığıma yöntemleri
- 3- Bölümleme (divisive) yöntemi
- 4- Bir araya getirme (agglomerative) yöntemleri olarak sınıflandırmıştır.

Duran and Odell (48); kümeleme yöntemlerini

- 1- Bölümleme (divisive)
- 2- Bir araya getirme (agglomerative) yöntemleri olmak üzere iki gruba ayırmıştır.

Diğer bir sınıflandırma şekli ise,

- 1- Aşamalı kümeleme yöntemleri (Hierarchical Clustering Methods).
- 2- Aşamalı olmayan kümeleme yöntemleri (Nonhierarchical Clustering Methods).
- 3- Ardışık parçalama Yöntemi (Iteratif Partitioning Methods).
- 4- Optimizasyon yöntemleri (Optimizing Procedures).
- 5- Diğer kümeleme yöntemleri olarak verilebilir (6), (28).

Anderberg (8), ise yukarıda verilen sınıflandırma şekillerini daha da genelleştirerek

- 1- Aşamalı kümeleme yöntemleri
- 2- Aşamalı olmayan kümeleme yöntemleri

olarak ikiye ayırmıştır. Çünkü kümeleme analizi denince akla ilk olarak bu iki grup gelmektedir. Anderberg (8)'in sınıflandırma biçimi genel kabul gören bir sınıflandırma biçimi olup, bu çalışmada da sözü edilen sınıflandırma biçimi benimsenmiş ve her iki başlık altında önemli görülen yöntemler ayrıntılarıyla incelenmiştir.

3.2.5. Aşamalı (hiyerarşik) kümeleme yöntemleri

Bireylerin kümelenmesinde n birimin

$$\frac{n(n-1)}{2} = \binom{n}{2} \quad (3.30)$$

ya da karakterlerin kümelenmesinde p karakterin

$$\frac{p(p-1)}{2} = \binom{p}{2} \quad (3.31)$$

adet mümkün olan tüm çiftlerin arasındaki benzerlik düzeylerine göre, birbirlerine aşamalı biçimde bağlamayı amaçlayan yöntemlere aşamalı kümeleme yöntemleri denir (8), (12).

Aşamalı kümeleme yöntemleri iki farklı şekilde incelenebilir:

1- Mevcut her bir birey veya karakteri bir küme farz ederek birbirlerine en çok benzeyenleri ard arda birleştirmek ki, bunlara “yığılmalı aşamalı yöntem” denir.

2- Yine mevcut bütün birey veya karakterler tek bir küme (grup) farz edilerek, birbirlerine en az benzeyenleri ard arda bölümleyerek kümelere ayırma ki, bunlara “ayrık aşamalı yöntem” denir.

Yığılmalı aşamalı yöntemde, başlangıçta birey veya karakter sayısı kadar küme vardır. Daha önce bölüm 3.2.2.’de bahsedilen uzaklık ve benzerlik matrislerinden yararlanarak birey veya karakterler adım adım birbirleriyle benzerlik düzeylerine göre birleştirilerek ağaç grafiği elde edilir. Bu yöntemleri kullanarak yapılan kümeleme işlemlerinde genel olarak aşağıdaki sıra takip edilir (8).

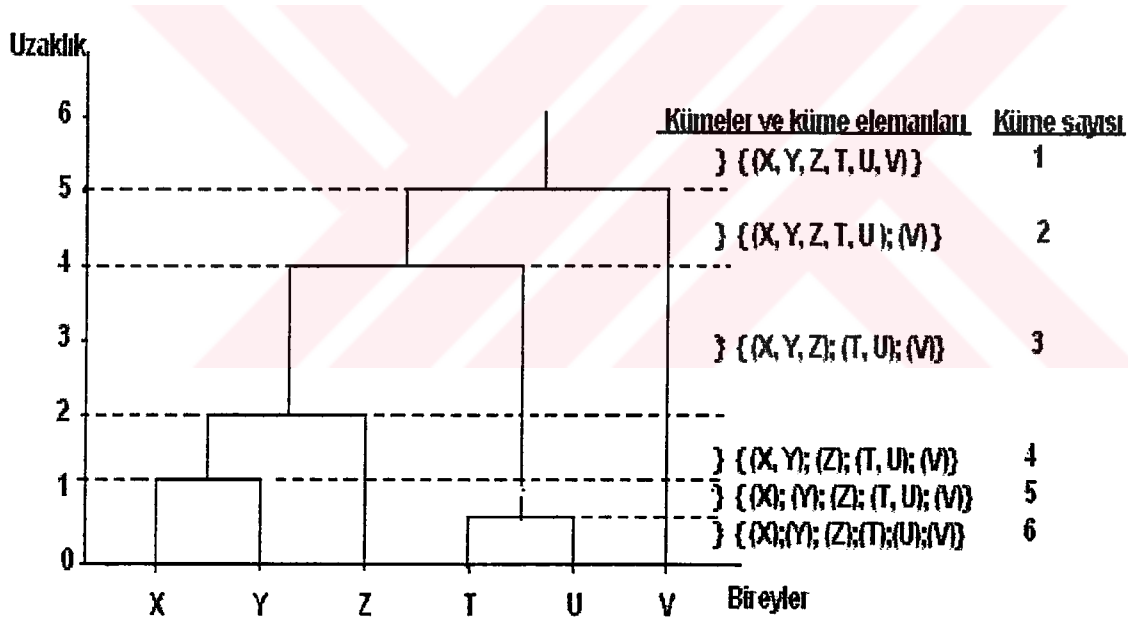
1. *Adım* - Gözlem değerleri 1’den n’e kadar sıraladığında, her biri bir gözlem değerine sahip n tane ayrı ayrı küme belirlenir.
2. *Adım* - Benzerlik veya uzaklık matrisine göre en benzer ya da en yakın (i) ve (j) kümeleri birbirleriyle birleştirilir.
3. *Adım* - (i) ve (j) kümeleri birleştirildikten sonra küme sayısı bir indirgenerek, yinelenmiş benzerlik matrisi oluşturulur.
4. *Adım* - İkinci ve üçüncü adımlar tüm bireyler tek bir küme oluşturuncaya kadar (n-1) kez tekrarlanır.

5. *Adım* - Benzerlik ölçütlerinin değerlerine göre birbirleriyle birleştirilerek küme oluşturan bireyler birleşme aşamalarına göre bir ağaç grafiği ile gösterilir.

i 'nci ve j 'nci kümenin birleştirilmesinden sonra oluşacak olan yeni küme ile diğer kümeler arasındaki benzerlikler belirlenerek üçüncü adımda yinelenmiş benzerlik matrisi oluşturulur. Bu matrisin oluşturulmasında çeşitli yaklaşımlara başvurulur. Bunun neticesinde başvuru farklı yaklaşımlar bir çok farklı aşamalı kümeleme yöntemlerinin ortaya çıkarılmasına sebep olmuştur.

Sözü edilen yöntemlere geçmeden önce ağaç grafiğinin nasıl elde edildiği bir örnekle açıklanırsa;

6 gözlem değerine ilişkin ağaç grafiği Şekil 3.4 deki gibi verilmiş olsun. Burada; Y ekseninde uzaklıklar, X ekseninde ise gözlem değerleri yani (X,Y,Z,T,U,V) harfleri ile isimlendirilmiş olan bireyler bulunsun.



Şekil 3.4. Aşamalı (hiyerarşik) kümelemenin yapısı ve ağaç grafiği.

Şekilde (3.4)'de görüldüğü gibi ilk başta, yani 0 (sıfır) uzaklıkta her bir birey bir kümeyi temsil ederek 6 adet küme oluşmaktadır. Daha sonra T ile U bireylerinin birbirlerine en yakın olması münasebetiyle bu iki birey birer dal ile birleştirilerek

(T, U) kümesi oluşturulur ve küme sayısı 5'e iner. Bir sonraki en yakın uzaklıkta ise X ve Y bireyleri birleşerek bir küme oluşturmakta ve bunun neticesinde küme sayısı 4'de inmektedir. Bundan sonraki uzaklıkta ise Z bireyi (X,Y) kümesine dahil olarak yeni bir (X, Y, Z) kümesi oluşturmakta ve küme sayısı 3'e inmektedir.

(T,U) bireylerinin oluşturduğu küme (X, Y, Z) bireylerinin oluşturduğu küme ile birleşerek küme sayısı (X, Y, Z, T, U) ve (V) olmak üzere ikiye indirgenmektedir.

En son uzaklıkta ise (V) kümesi (X, Y, Z, T, U) kümesine dahil olarak (X,Y,Z, T,U,V) bireylerinden müteşekkil tek bir küme oluşur.

3.2.5.1. Tek Bağlantı (en yakın komşu) kümeleme yöntemi (Single Linkage Cluster Analysis)

Tek bağlantı kümeleme yöntemi Aşamalı kümeleme yöntemleri arasında en basit olanıdır. Bundan dolayı en sık kullanılan yöntemlerden birisidir. Bu yöntemin diğer bir önemli özelliği ise, sınıflandırılacak birey sayısı az olduğu durumlarda küme yapısını belirlemede en iyi sonucu veren yöntemlerden biri olmasıdır.

Söz konusu yöntem literatürlerde değişik şekillerde adlandırılmaktadır. Bunlar sırasıyla;

- En yakın komşu (Nearest neighbor cluster analysis)
- Minimum yöntem (Minimum distance method)
- Basit bağlantı çözümlemesi

şeklinde karşımıza çıkmaktadır (12).

Tek bağlantı kümeleme yönteminde her bir birey veya bireylere ilişkin karakter bir küme olarak ele alınır. Yani başlangıçta n tane küme mevcuttur. Tek elemanlı olan bu kümelerden birbirine en benzer veya en yakın olanları birleştirilir. Bundan dolayı $D=\{d_{ij}\}$ uzaklık veya $S=\{s_{ij}\}$ ilişki türü benzerlik matrisi içindeki en küçük uzaklığı veya en benzer bireyleri tespit ederek birleştirilir. Başlangıçta (u) ve (v) kümeleri birleştirilmiş ise 3. adımda (u, v) kümesine en yakın olan (w) kümesi arasındaki uzaklık ve benzerlik ölçümleri aşağıdaki şekilde hesaplanabilir. Eğer

bireyler arasında, d_{ij} gibi uzaklık türü ölçüler (distance-like measure) kullanılmış ise;

$$d(u,v)_w = \min \{d_{uw}, d_{vw}\} \quad (3.32)$$

eşitliği ile hesaplanır. Eğer s_{ij} gibi ilişki türü ölçüler (correlation-like measure) kullanılmış ise, iki birey arasındaki benzerlik;

$$S(u,v)_w = \max \{S_{uw}, S_{vw}\} \quad (3.33)$$

eşitliği ile hesaplanır.

Buradaki d_{uw} ve d_{vw} uzaklık değerleri, S_{uw} ve S_{vw} benzerlik değerleri sırasıyla (u) ve (w) kümesi ile (v) ve (w) kümelerinin en yakın komşuları arasındaki uzaklık veya en çok benzeyen kümeler arasındaki benzerliktir. Bu işlemlerin sonunda ilgili matrisin boyutu ($n-1$) indirgenir ve matris yenilenir. Yöntem tekrarlanarak devam eder (8).

3.2.5.2. Tam Bağlantı (en uzak komşu) kümeleme yöntemi (Complete Linkage Cluster Analysis)

Literatürlerde Tam Bağlantı kümeleme yöntemi sırasıyla;

- En uzak komşu (Furthest neighbor analysis) analizi
- Maksimum yöntem (Maximum distance method)
- Çap yöntemi (Diameter method)

gibi isimlerle anılmaktadır (12).

Sözü edilen yöntem Tek bağlantı kümeleme yöntemine çok benzemekle beraber, Tek Bağlantı kümeleme yönteminden ayıran tek özelliği küme elemanları arasındaki uzaklık veya benzerliklerin en yüksek değerinin alınmasıdır. Yani kümeler teşkil edilirken öncelikle en uzak komşu niteliğine sahip bireyler birbirleriyle birleştirilerek bir araya getirilmektedir (8).

Tam bağlantı kümeleme tekniği basit olmakla beraber küme yapısını belirlemede iyi sonuç veren bir tekniktir.

Bu yöntemde uzaklık veya ilişki matrisi elde edildikten sonra n tane bireyden birbirine en yakın ya da birbirine en benzer olan (u) ve (v) gibi iki küme (uv) kümesi

olarak birleştirildikten sonra 3. bir (w) kümesi ile arasında uzaklığın (veya benzerliğin) belirlenmesi için uzaklık türü ölçüler (distance-like measure) kullanıldığında

$$d(u, v)_w = \max \{d_{uw}, d_{vw}\} \quad (3.34)$$

ilişki türü ölçüler (correlation-like measure) kullanıldığında ise,

$$S(u, v)_w = \min \{S_{uw}, S_{vw}\} \quad (3.35)$$

bağıntılarından yararlanır (8).

3.2.5.3. Gruplar İçi Ortalama Bağlantı kümeleme yöntemi (Within-Groups Linkage Cluster Analysis)

Bu yöntemde, D uzaklık matrisinde yer alan d_{ij} elemanları, (u) kümesinden bir gözlem değeri ve (v) kümesinden de bir gözlem değeri (birey) alınarak oluşturulan tüm olası birleşimlerle ilgili benzerlik ölçülerinin toplamı olarak alınır. (u) ve (v) kümeleri birleştirildiğinde oluşacak yeni (uv) kümesi ile herhangi bir (w) kümesi arasındaki çiftlerin benzerlik ölçüleri toplamını,

$$d(u, v)_w = d_{uw} + d_{vw} \quad (3.36)$$

olarak verebiliriz. Tekrarlanan benzerlik matrisinde $d(u, v)_w$ değeri yer alır.

T_u , (u) kümesi içindeki bireylerin tüm mümkün çiftleri arasındaki benzerliklerin toplamı, N_u ' da (u) kümesi içindeki bireylerin sayısı olarak kabul edilecek olursa; kümeler birleşmeden önce $T_u = 0$ ve $N_u = 1$ olacaktır. Yani ilk başta her birey bir kümeyi temsil etmektedir ve henüz bireyler birleştirilmediğinden bireylerin benzerlikleri toplamı 0 dır. (u) ve (v) birleştirildiğinde yeni oluşacak (uv) kümesi içindeki bireylerin tüm mümkün çiftleri arasındaki benzerliklerin toplamı,

$$T(uv) = T_u + T_v + d_{uv} \quad (3.37)$$

(uv) kümesindeki birimlerin sayısı ise

$$N(uv) = N_u + N_v \quad (3.38)$$

şeklinde hesaplanır.

(u) ve (v) kümelerine ait çiftlerin benzerlik ölçülerinden ve birey sayılarından yararlanarak en benzer çiftin bulunmasında grup içi benzerlik ortalaması;

$$\begin{aligned}\bar{d}_{uv} &= \frac{T(uv)}{N(uv) \cdot \frac{(N(uv) - 1)}{2}} \\ &= \frac{T_u + T_v + d_{uv}}{(N_u + N_v) \cdot \frac{(N_u + N_v - 1)}{2}}\end{aligned}\quad (3.39)$$

olarak hesaplanabilir.

Hesaplanan bu ortalamalar minimum olduğunda (u) ve (v) kümeleri birleştirilir. (u) ve (v)'nin satır ve sütunları D matrisinden çıkarılır. Bu işlemler tekrarlanmak suretiyle tüm bireylerin kümelerde birleştirilmesi sağlanıncaya kadar kümeleme işlemi devam eder (8), (12), (49).

3.2.5.4. Gruplar Arası Ortalama Bağlantı kümeleme yöntemi (Between-Groups Linkage Cluster Analysis)

Gruplar arası ortalama bağlantı yöntemi, gruplar arası uzaklığı iki gruptaki bireylerin tüm mümkün çiftleri arasındaki uzaklıkların aritmetik ortalaması olarak tanımlanmaktadır. Bu yöntem iki birey veya bir birey ile bir grup arasındaki uzaklığın bir ölçüsü olarak kullanılabilirdiği gibi iki grup arasındaki uzaklığın bir ölçüsü olarak da kullanılabilir (11), (50).

Sözü edilen teknikte Gruplar içi ortalama bağlantı kümeleme tekniğinde izlenen yol takip edilmek suretiyle yinelenen benzerlik matrisi oluşturulur. Daha önce söz edildiği gibi d_{uv} ; (u) ile (v) kümeleri arasındaki benzer çiftlerin benzerliklerinin toplamıdır. N_u , (u) kümesindeki birey sayısı N_v ise, (v) kümesindeki birey sayısı olmak üzere; $N_u \cdot N_v$ benzer grup çiftlerinin sayısıdır.

(u) ve (v) kümeleri için Gruplar arası benzerlik ortalaması,

$$\bar{d}_{uv} = \frac{d_{uv}}{N_u \cdot N_v}\quad (3.40)$$

olarak hesaplanır.

Bu metodun Gruplar içi ortalama bağlantı tekniğinden farkı, benzer grup çiftlerinin benzerlikleri toplamının (T_u ve T_v) hesaplamaya katılmamasıdır.

Bu teknik Lance and Williams (50) tarafından grup ortalaması olarak adlandırılmıştır.

3.2.5.5. Merkezi (Centroid) Bağlantı kümeleme yöntemi

Bu yöntem, kümeler arası uzaklıklar ve küme merkezleri arası uzaklıklar olarak tanımlanmıştır. Çünkü kümeler merkezleri ile ifade edilmektedir. Kümelerin birleştirilmesi küme merkezleri arası uzaklığa göre yapılmaktadır. İlk birleştirilecek kümeler, kümeler arası uzaklığı en az olanlar olacaktır (11).

(uv) kümesinin (w) kümesine olan uzaklığı;

$$d(uv)w = \frac{Nu \cdot duw + Nv \cdot dvw}{N(uv)} - \frac{Nu \cdot Nv \cdot duv}{N(uv)^2} \quad (3.41)$$

biçiminde hesaplanır.

Gower (44), Merkezi bağlantı kümeleme yönteminin uzaklık türü ölçütlerle kullanılmasının daha iyi sonuç vereceğini, ancak benzerlik türü ölçütler mevcut ise bunların doğrudan doğruya kullanılması halinde bir kısım sakıncaların ortaya çıkacağını bildirmiştir. Bu nedenle, benzerlik türü ölçütlerin öncelikle uzaklık türü ölçülere dönüştürülmesi gerektiğini önermiştir.

Everitt (11) herhangi bir uzaklık tipi ölçütün benzerlik tipi ölçüte dönüşümünün,

$$S_{uv} = 1/1 + d_{uv}$$

biçiminde mümkün olduğunu bildirmiştir. Burada S_{uv} , $0 < S_{uv} < 1$, (u) ve (v) bireyleri arasındaki benzerliktir.

Gower (44) ise, bu işlemin tersine dönüşümünün oldukça güç olduğunu, ancak; benzerlik katsayıları matrisinin pozitif yarı tanımlı olma şartı altında benzerlik tipi ölçütlerinde uzaklık tipi ölçütlere 3.42 No'lu eşitlik vasıtasıyla dönüştürülebileceğini bildirmiştir.

Burada; $S_{uu} = S_{vv}$ ise,

$$d_{uv} = \sqrt{S_{uu} + S_{vv} - 2S_{uv}} = \sqrt{2 \cdot (S_{vv} - S_{uv})}$$

$S_{vv}=1$ ise,

$$d_{uv} = \sqrt{2 \cdot (1 - S_{uv})} \quad (3.42)$$

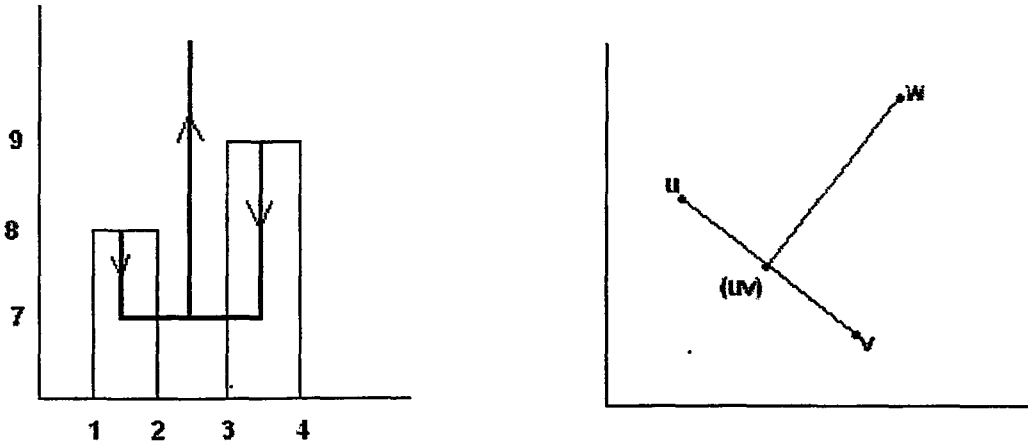
3.42 No'lu eşitlik vasıtasıyla hesaplanan uzaklık matrisi elde edilirken, S benzerlik katsayıları matrisinin özdeğer ve özvektörlerinden yararlanılmıştır (6),(17).

$$S(uv)w = \frac{N_u}{N_u + N_u} \cdot S_{uw} + \frac{N_v}{N_u + N_v} \cdot S_{vw} - \frac{N_u \cdot N_v}{(N_u + N_v)} \cdot S_{uv} \quad (3.43)$$

(u) ve (v) kümeleri ilk aşamada birleştirildikten sonra oluşacak olan (uv) kümesinin en yakın 3. bir küme olan (w) kümesi ile birleştirilmesinde kullanılacak eşitlik (3.43) da verilmiştir (8).

Anderberg (8), bu yöntemin en iyi sonucu D benzerlik matrisinin Öklid uzaklığının karesinin kullanılması ile vereceğini bildirmiştir.

Merkezi bağlantı yönteminin şematik olarak gösterimini Şekil 3.5'de ki gibidir.



Şekil 3.5. Merkezi Bağlantı yönteminin şematik gösterimi

3.2.5.6. Ortanca (Median) Bağlantı kümeleme yöntemi

Bu yöntem daha ziyade bireylerin karakterlerine ilişkin değerlerinin sıralı ölçek ile elde edildiği veya ölçüm değerleri yerine sıra puanları kullanıldığında ilgili kümelerin belirlenmesinde kullanılır. Merkezi bağlantı kümeleme tekniğinde birleştirilen iki kümeye ait gözlem sayıları $N(u)$ ve $N(v)$ birbirinden çok farklı ise yeni kümenin merkezi büyük kümeye daha yakın olacak ve belki de büyük kümenin içinde kalacak ve küçük kümenin özellikleri hemen hemen kaybolacaktır.

Gower (31), bu durumu Merkezi kümeleme yönteminin sakıncalı yönü olarak bildirmektedir. Bu sakıncadan kurtulmak ve yeni kümenin yerinin daima birleştirilmiş olan iki kümenin arasında olmasını sağlamak için birleştirilen kümeler eşit boyutlu kabul edilerek küme boyut dan bağımsız kılınabilir. Ya da Gower (31)'in önerisi olan birleştirilecek kümelerdeki gözlem sayısına bakılmaksızın her bir kümeye birim ağırlıkları verilebilir.

Eğer S_{jk} ilişki türü bir benzerlik matrisi ise (u) ve (v) kümelerinin birleştirilmesi ile oluşan (uv) kümesi başka bir (w) kümesi ile birleştirilecek ise bu birleştirmede kullanılacak olan ilişki;

$$S(uv)w = \frac{1}{2} \cdot S_{uw} + \frac{1}{2} \cdot S_{vw} + \frac{1}{4} \cdot (1 - S_{uv}) \quad (3.44)$$

biçimindedir. Ancak elimizdeki veriler uzaklık türü benzerlik matrisini meydana getiriyorsa Merkezi bağlantı kümeleme yönteminde uygulandığı gibi 3.44 No'lu eşitliğe, 3.42 eşitliği ile verilen dönüşüm uygulandığı zaman, bu eşitlik uzaklık tipi ölçütler için;

$$d(uv)w = \frac{1}{2} \cdot d_{uw} + \frac{1}{2} \cdot d_{vw} - \frac{1}{4} \cdot d_{uv} \quad (3.45)$$

biçimine dönüşür (44), (8).

3.2.5.7. Ward kümeleme tekniği

“En küçük varyans yöntemi” olarak da adlandırılan Ward metodu, çok aşamalı bir kümeleme yöntemidir. Bu yöntemin en önemli özelliği birey sayısı (n) çok büyük olduğunda ($n > 50$), diğer aşamalı kümeleme yöntemlerine göre daha iyi sonuçlar vermesidir (8).

İstatistiksel anlamda, kümeleme yayılımının küçültülmesine dayanan Ward metodu kümeler içi varyansı en küçük yapmayı amaçlar. Kümeleme analizinin herhangi bir aşamasında kümelerdeki bireylerin gruplanmasından dolayı bir miktar bilgi kaybı olacaktır. Bu bilgi kaybı, belli bir kümeye ait olan her noktanın küme ortalamasından sapmalarının kareler toplamı ile (HKT) ölçülebilir.

Bu metodun ikinci önemli bir özelliği ise benzerlik matrisi elemanlarının Öklid uzaklığının karesi kullanılarak oluşturulmasıdır. n bireyden elde edilen p karaktere ilişkin verilere göre bireyler (gözlem değerleri) k kümeye ayrılmak istensin. Bu durumda X_{ijk} ; k . kümenin, j . bireyinin, i . değerini gösterir. k kümesinde n_k birey vardır ve bireylerin i . karakter ortalaması \bar{X}_{ik} ;

$$\bar{X}_{ik} = \frac{\sum_{j=1}^{n_k} X_{ijk}}{n_k} \quad (3.46)$$

olarak gösterilir.

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

$$\begin{aligned} HKT_k &= \sum_{i=1}^p \sum_{j=1}^{n_k} (X_{ijk} - \bar{X}_{ik})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_k} X_{ijk}^2 - n_k \sum_{i=1}^p \bar{X}_{ik}^2 \end{aligned} \quad (3.47)$$

şeklinde hesaplanır. HKT_k değeri tüm $k = 1, 2, \dots, n$ kümelerde hesaplanarak kümeler içi HKT'nın toplamı

$$HKT = \sum_{i=1}^{k=n} HKT_k \quad (3.48)$$

olur.

Bu değerler hesaplandıktan sonra bireylerin kümelerde birleştirilmeleri 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) kümeleri birleştirilerek (uv) kümesi oluşturulur. HKT'daki bu artış;

$$\Delta HKT_{uv} = HKT_{(uv)} - HKT_{(u)} - HKT_{(v)} \quad (3.49)$$

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 (8).

Böylece her aşamada HKT'ında ki oluşan minimum artış, birleştirilen kümelerin (küre ortalamaları) merkez noktaları arasındaki Öklid uzaklığının karesi ile orantılı olarak gerçekleşmiş olur.

3.2.5.8. McQuitty Bağlantı kümeleme yöntemi

Herhangi bir (uv) kümesinin (w) kümesine olan uzaklığı, (u) ve (v) kümelerinin (w) kümesine olan uzaklıkları toplamının yarısı olarak belirlenir. Tartısız ortalama bağlantı yöntemi olarak da bilinen yöntemde kümeler arası uzaklık;

$$d(uv)w = \frac{d(uw) + d(vw)}{2} \quad (3.50)$$

şeklinde hesaplanır (51).

3.2.6. Aşamalı Olmayan Kümeleme Yöntemleri

Daha önce açıklandığı gibi aşamalı kümeleme yöntemlerinde, başlangıçta mevcut olan n birey (veya p karakter), n tane kümeyi temsil etmekteydi. 2. adımda ise, daha önce hesaplanmış olan benzerlik matrisleri vasıtasıyla birbirine en yakın bireyler belirlenerek, aşama aşama tüm bireyler birbirine bağlanarak tek bir küme oluşturuluyordu. Ancak, burada herhangi bir kümeye bir defa girmiş olan bir birey, işlemin ileri ki aşamalarında kendisine daha uygun bir küme bulunması halinde dahi ilk girdiği kümeden çıkamamaktaydı. Bu durum, aşamalı yöntemler için bir sakıncadır. Fakat buna karşılık bireyler arasındaki benzerlik düzeylerini ve bireyler arası ilişkileri inceleme olanağı sağlanmaktadır.

Aşamalı olmayan yöntemlerde ise sadece bireylerin kümelenme olanağı olup, karakterlerin kümelenme olanağı yoktur. Burada bireylerin ilk bölümlenmesi seçildikten sonra küme elemanlarının yerlerini değiştirerek en iyi bölümlenme yapılabilmektedir. Yani, n bireyin k kümeye ayrılması söz konusu olup, bireylerin k kümeye ayrılması ya önceden araştırmacının ön bilgisine dayanarak belirlenir veya parçalanma işlemi rastgele yapılarak kümeleme işlemi sonucu bulunabilir (6), (8).

Kümeleme işlemi esnasında benzerlik matrisi bulunmaz ve ham veri matrisinin de bilgisayarda depolanması gerekmez. Bu nedenle aşamalı yöntemlerin aksine geniş veri matrisinde kullanılabilir. Eğer veri matrisi benzerlik matrisine dönüştürülecek olursa birey sayısının çok olması durumunda aşamalı olmayan yöntemlerde sorunlar çıkabilmektedir (6).

K-ortalama yöntemi iki şekilde uygulanabilir:

- 1) Bireylerin gruplara başlangıç dağıtımı ile işe başlanır veya
- 2) Kümelerin çekirdeğini teşkil edecek olan çekirdek noktaların başlangıç seti oluşturularak işe başlanır.

Bu çekirdek noktalar ise ya tesadüfen seçilir veya bireyler gruplara başlangıçta tesadüfen dağıtılır.

Araştırmacının veriler hakkında ayrıntılı bilgiye sahip olması, çekirdek noktalarının seçiminde büyük önem arz etmektedir. Çekirdek noktalarının tespit edilmesinde birçok farklı yöntem önerilmiştir. Bu yaklaşımlardan bazıları aşağıda sıralanmıştır.

1. MacQueen(53), veri setindeki ilk k bireyin çekirdek nokta olarak seçilmesini önermiştir.
2. Bireyler 1 den n 'e kadar numaralanarak, 1 ile n arasında rastgele sayılar seçilmek suretiyle, bu sayılara karşılık gelecek sıra numarasına sahip bireyler çekirdek noktayı temsil eder (8).
3. Veri setindeki bireyler 1'den n 'e kadar sıralanır ve bu defa sistematik olarak $a/k, 2a/k, \dots, (k-2)a/k$ ve $(k-1)a/k$ numaralı bireyler çekirdek nokta olarak seçilir. Burada, $a \leq (n/k)$ olarak alınan rastgele sayıdır (8).
4. Her bir birey için dağılım aralığına göre koordinatlar vektörü olarak k nokta türetilir. Türetilen bu noktalar çekirdek nokta olarak kabul edilir. Bu noktalar veri setindeki birimlere karşılık gelen noktalardan farklı olabilir.
5. Veri setindeki birimler birbirleri ile kesişmeyen, k tane kümeye parçalanır ve her kümenin küresel ortalamaları çekirdek noktalar olarak alınır (8), (52).

Küme sayısı hakkında bir ön bilgi yok ise, aşamalı yöntemlerin ağaç grafiklerinden yararlanmak suretiyle de küme sayısına karar verilebilir. Ancak, küme sayısı hakkında ön bilgi var ise aşamalı olmayan kümeleme yöntemleri tercih edilir.

3.2.6.1. k-Ortalama yöntemi

MacQueen (53) tarafından K-Ortalama diye adlandırılan yöntemin işleyiş mantığı; n bireyin k kümeye ayrılmasında bireyin p boyutlu uzayda gösterimi olan en yakın ortalamaya sahip çekirdek noktalı bir kümeye atanmasını içerir. Diğer bir ifadeyle, bu yöntemde bireyler kümeler içi kareler toplamı en küçük olacak şekilde k kümeye bölünmektedir. Yani;

$$X_1, X_2, X_3, \dots, X_n$$

her bir p karakterli gözlem vektörleri, çok boyutlu X uzayında birer nokta olarak düşünülecek olursa ve aynı uzayda, her grup birey için küme merkezleri olarak

$$a_{1n}, a_{2n}, \dots, a_{kn} ; j = 1, 2, \dots, k$$

seçildiğinde;

$$W_n = \frac{1}{n} \sum_{i=1}^n \min_{1 \leq j \leq k} \|X_i - a_{jn}\|^2 \quad (3.51)$$

kuralına göre bireyler en yakın kümeye sınıflanmaktadır (41),(53).

Sözü edilen yöntemin n gözlem k küme için ($k < n$), bilgisayar algoritmasındaki işleyişi aşağıdaki gibi maddeler halinde verilebilir;

1. Araştırmacının elde etmiş olduğu verilerin ön bilgilerine dayanarak, ilk k nokta çekirdek nokta olarak alınır. Bu çekirdek noktaların her biri birer küme ortalaması olarak kabul edilir. Küme ortalamalarından her bireyin uzaklıkları hesaplanır.
2. Geriye kalan $(n-k)$ bireyin her biri, ortalaması en yakın olan kümeye atanır. Her atamadan sonra oluşan yeni kümenin ortalaması tekrar hesaplanır. Böylece çekirdek noktaların, verileri en uygun kümeye ayırmayı sağlayacak biçime dönüşmesi sağlanır. Uzaklıklar genellikle standartlaştırılmış ya da standartlaştırılmamış gözlemlerle Öklid uzaklığı kullanarak hesaplanmaktadır (6).
3. Bireylerin tamamının 2. adımdaki işlemin tekrarlanması ile kümelere (k -kümeye) atanması bittikten sonra oluşan kümelerin ortalamaları tekrar hesaplanarak yeni çekirdek noktalar belirlenir ve bir sonraki adım için bu noktalar kullanılır. Bu durumda her bireyin en yakın ortalamaya göre atama işlemi tekrarlanır. Bu atama işlemleri, kümeler arası birey geçişi duruncaya kadar tekrarlanır. Deneyimler çoğu ana değişimlerin ilk yerleştirme adımında oluştuğunu göstermektedir (53).

Anderberg (8)'in bildirdiğine göre, Forgy isimli araştırmacı, gözlemlerin kümelere atanmasında aşağıdaki gibi bir algoritma tanımlamıştır.

1. Algoritmaya başlarken ilk önce çekirdek noktalar bulunmuş ise 2. adıma geçilir, ya da ilk uygun bölümlenme belirlenmiş ise 3. adıma geçilir.
2. Her gözlemi en yakın çekirdek noktaya atayarak küme oluşturulur. (Çekirdek noktalar tüm algoritma boyunca değişmez, sabit kalır)
3. Uygun ilk bölümlenme belirlenmiş ise çekirdek noktalar; oluşturulan çekirdek kümelerin ortalamaları olarak alınır ve yakınlık ölçütüne göre atama yapılır.

4. İkinci ve üçüncü adım algoritma yakınsayınca kadar; yani küme elemanlarının geçişi duruncaya kadar sürdürülür.

Burada, Anderberg (8) ile Mac Queen (53)'in ileri sürdüğü algoritmalar arasında önemli bir fark vardır. Bunlar ;

- 1- Mac Queen (53)'in yönteminde her gözlemin kümeye atanmasından sonra küme çekirdeğinin yenilenmesi,
- 2- Anderberg (8)'in algoritmasında ise çekirdek noktaların sabit kalmasıdır.

3.2.7. En iyi bölümlene ölçütleri

En iyi bölümlene ölçütünden amaç, bireylerin ayrılacakları küme sayısı belirlendikten sonra, küme içi homojenliği ve kümeler arası heterojenliği en iyi sağlayacak ölçüt geliştirmektir. Bu amaçla g_1 , g_2 , g_3 , İz W ($\text{tr}W$), $|W|$, r_{cs} , LR(Likelihood Ratio), Wilks Lambda, Hotelling Lawley iz değeri gibi bir çok kümeleme ölçütü tanımlanmıştır (18), (35), (54).

Kümeleme analizinde belirli ölçütleri kullanarak, kümeleme analizi sonucu elde edilen farklı kümelerden, en iyi kümeleri saptama düşüncesini uygulayan ilk araştırmacı Edwards ve Cavalli-Sforza (18) olmuştur. Bunun için çok değişkenli varyans analizinin temel denklemi olarak bilinen

$$T = W + B$$

matris eşitliği kullanılmıştır. Burada;

$$T = \text{Genel kareler toplamı}$$

$$W = \text{Kümeler içi kareler toplamı}$$

$$B = \text{Kümeler arası kareler toplamı matrisleridir.}$$

Kümeler arası uzaklığı kullanan ölçüt değerlerini maksimize veya kümeler içi uzaklığı kullanan ölçüt değerlerini minimize etmek suretiyle veriler kümelere atanır.

Bu amaçla kullanılan kümeleme ölçütlerinden biri;

$$\text{Max} [iz(B)] = \text{Max} \left(\sum_{i=1}^p b_{ii} \right) \quad (3.52)$$

biçimindedir. Edwards ve Cavalli-Sforza (18) ilk önce n gözlemlili veri kümesinin ikiye bölünmesini 3.52 No'lu eşitlik yardımıyla maksimize ettikten sonra, ortaya çıkan bölünmeleri de incelemişlerdir.

Ancak, Gover (31) bir araştırmasında, çok büyük örnekler için bu ölçütün bilgisayar yardımıyla dahi kullanımının imkansızlığını belirtmiştir.

Bu konuda yapılan çalışmalardan elde edilen bilgilerin genelinde, kümelerin test edilmesinde en etkili yöntemlerin Olabilirlik Oran Testi (Likelihood Ratio Test) ve Küme Ara Kesit Testi (Union Intersection Test) yaklaşımları üzerine kurulu olan çok değişkenli varyans analizi testleri olduğu belirtilmiştir. Sözü edilen testlerden kümeleme analizinde en yaygın olarak kullanılanları Wilk's Lambda ve Hotelling Lawley İz yöntemleridir. Bu testlerin tümünde $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ olduğu varsayılmakta ve tek yönlü çok değişkenli varyans analizi olarak adlandırılmaktadır.

Bu testlerde,

$$H_0: \tau_1 = \tau_2 = \dots = \tau_k = 0$$

$$H_1: \tau_i \neq \tau_j \quad i, j = 1, \dots, k \text{ ve } i \neq j \text{ için}$$

hipotezleri kurularak analize başlanmaktadır. Her bir $p \times 1$ boyutlu örnek grubu için $x_j \sim N_p(\mu_j, \sigma^2)$ varsayımı geçerli olmak üzere; $i=1, \dots, n_j$, n_j her örnek grubundaki gözlem sayısını, $j=1, \dots, k$; k grup sayısını göstermek üzere aşağıdaki eşitliklerden yararlanılmaktadır.

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})(x_{ij} - \bar{x})' = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})' + \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)'$$

$$T_{pxp} = B_{pxp} + W_{pxp} \quad (3.53)$$

Burada,

T = Genel çarpımlar ve kareler toplamı matrisi (Total matrix),

B = Gruplar arası çarpımlar ve kareler toplamı matrisi (Between group matrix),

W = Gruplar içi çarpımlar ve kareler toplamı matrisidir (Within group matrix).

Bu etkileri gösteren değerler Çizelge 3.4'deki çok değişkenli varyans analizi tablosunda (ÇDVAT) verilmiştir.

Çizelge 3. 4. Çok değişkenli varyans analiz tablosu

Varyasyon Kaynakları	Serbestlik Derecesi	Matris Kareleri ve Çapraz Çarpımlar Toplamı
Gruplar	$k-1$	$B = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})'$
Hata	$\sum_{j=1}^k n_j - k$	$W = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)'$
Genel	$\sum_{j=1}^k n_j - 1$	$B + W = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})(x_{ij} - \bar{x})'$

Yukarda sözü edilen ve bu çalışmada kullanılmış olan test istatistikleri sırasıyla aşağıda açıklanmıştır;

3.2.7.1. Wilks'in olabilirlik oran testi

Wilks tarafından önerilen bu yöntem, geliştirilmiş olabilirlik oran testi (generalized likelihood ratio principle) temeline dayalıdır. Test istatistiği olarak önerilen Λ ,

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W + B|} \quad (3.54)$$

veya

$$\Lambda = \prod_{i=1}^p \frac{1}{1 + \lambda_i} \quad (3.55)$$

biçimindedir. 0 ile 1 arasında değerler alan bu test istatistiğinde, Λ 'nın 0'a yakın olması H_0 hipotezinin red edileceğini, 1'e yakın olması red edilemeyeceğini gösteren bir işarettir. Bu yöntemde kritik değer olarak,

$$L = -\left(n-1 - \frac{(p+k)}{2}\right) \ln \Lambda = -\left(n-1 - \frac{(p+k)}{2}\right) \ln \left(\frac{|W|}{|W+B|}\right) \sim \chi^2_{p(k-1), \alpha} \quad (3.56)$$

dağılımı nedeniyle $p(k-1)$ serbestlik dereceli Khi-kare tablo değerleri kullanılmaktadır. Eğer sonuç, hesap değeri $\geq \chi^2_{p(k-1), \alpha}$ ise H_0 hipotezi reddedilir ve elde edilen kümeler anlamlı derecede farklıdır denir (41).

Wilks Λ yerine F istatistiği kullanılması için,

$$F = \frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \cdot \frac{r \cdot t - 2 \cdot u}{p \cdot q} \sim F_{(p \cdot q, r \cdot t - 2 \cdot u) \alpha} \quad (3.57)$$

dönüşümü yapılır. Burada sırasıyla;

v = Hata serbestlik derecesi, k = grup (küme) sayısı, p = değişken sayısı ve

$$r = v - (p + q + 1)/2$$

$$u = (pq - 2)/4$$

$$t = \sqrt{\frac{(p^2 q^2 - 4)}{(p^2 q^2 - 5)}} \text{ , Eğer } (p^2 q^2 - 5) > 0 \text{ değilse } t=1 \text{ dir.}$$

$$q = k-1 \text{ dir (6), (55).}$$

3.2.7.2. Hotelling-Lawley iz yöntemi

Bu yöntemde, (BW^{-1}) matrisinin determinantından elde edilen köşegen elemanlarının toplamı; yani $[\text{İz}(BW^{-1})]$ Hotelling- T^2 değerini vermektedir. Diğer bir ifade ile λ_r , BW^{-1} matrisinin determinantından elde edilen r 'inci öz değer olmak üzere,

$$T_o^2 = \sum_{r=1}^p \lambda_r \quad (3.58)$$

test istatistiği kullanılmaktadır. T_o^2 değerinin büyüklüğü H_0 hipotezinin reddedileceğini ifade etmektedir. n değerinin yeterince büyük olması durumunda,

$$HL = nT_o^2 \sim X_{p(k-1),\alpha}^2 \quad (3.59)$$

benzerliğinden yararlanarak $p(k-1)$ serbestlik dereceli Khi-kare ($\chi^2_{p(k-1)}$) tablo değerleri kritik değer olarak kullanılmaktadır. Ayrıca,

$$F = \frac{p(n-p-1) + 2}{(k-1)p^2} \cdot T_o^2 \quad (3.60)$$

dönüşümü yapılarak, $F_{[(k-1)p^2, p(n-p-1)+2],\alpha}$ serbestlik dereceli F tablo değerleri de kritik değer olarak kullanılabilir (41).

3.2.8. Küme sayısının belirlenmesi

Kümeleme analizinde, en önemli sorunlarından biri, küme sayısına karar vermektir. İncelenen tüm yöntemlerde bu problem görülmektedir. Günümüzde halen bir çok araştırmacı en uygun küme sayısını belirleme konusunda çalışma yapmaktadırlar.

Her ne kadar aşamalı yöntemlerle elde edilen ağaç grafikleri kümelerin yapısını aşağı yukarı vermekte ise de, ağaç grafikleri, değişik aşamalı kümeleme yöntemleri ile farklı sonuçlar vermektedir. Bundan dolayı bu yolla küme sayısını tespit etmek oldukça zordur.

Bu konuya çözüm getirmek amacıyla, küme sayısına ani artış veya azalışlarla karar verilmek istenmiştir. Ancak burada da karar verme eyleminde olasılıksal ölçütleri kullanılmadığından, işlemin yetersiz kaldığı tespit edilmiştir.

Küme sayısının (k) belirlenmesinde en pratik yol,

$$k \cong \left(\frac{n}{2}\right)^{1/2} \quad (3.61)$$

değerinin hesaplanmasıdır. Ancak, bu eşitlik örnek hacminin büyük olması durumunda iyi sonuçlar vermemektedir (41).

Probleme ilk defa doğrudan yaklaşan Marriott (25) olmuştur. En iyi küme sayısını bulmak için, gruplar içi kareler ve çarpımlar toplamı (HKÇT) matrisini $|W|$ kullanan ölçütü incelemiştir.

Küme sayısı 3.62 No'lu eşitlikten bulunmakta ve en küçük M değerini veren küme sayısı gerçek küme sayısı olarak değerlendirilmektedir.

$$M = k^2 |W| \quad (3.62)$$

Burada; k = küme sayısını vermektedir.

Bu ölçüt minimize edilerek doğru küme sayısına ulaşılabilmektedir (25).

Ayrıca gözlemlerle ilgili bir kümeleme yapısının olup olmadığına veya tek bir grupta gösterilip gösterilmeyeceğine karar vermek için 3.63 No'lu istatistiğin kullanılabileceği belirtilmektedir.

$$k^2 \frac{|W|}{|T|} \quad (3.63)$$

Eğer bu ölçütün değeri, mümkün olan tüm alt bölümler için birden daha büyük ise, gözlemler tek grup olarak kabul edilmektedir.

Bu konuyla ilişkin diğer bir çalışmada, Calinski ve Harabasz (56), B ve W sırasıyla gruplar arası ve gruplar içi kareler ve çarpımlar toplamı matrisi olmak üzere;

$$C = \frac{\left[\frac{iz(B)}{k-1} \right]}{\left[\frac{iz(W)}{N-k} \right]} \quad (3.64)$$

eşitliğini maksimum yapan C oranının kullanılmasını önermiştir. Burada C oranını maksimum yapan k değeri aynı zamanda küme sayısını vermektedir. Eğer;

- a- C 'nin değeri k ile birlikte tek düze artarsa verilerde küme yapısı olmadığına,
- b- C 'nin değeri k ile birlikte tek düze azalırsa verilerde aşamalı yapı bulunduğuna,
- c- C 'nin değeri k 'da en büyük değere ulaşıyorsa, uygun küme sayısının k olduğuna karar verilir.

3.2.9. Kümeleme analizinden önce değişken sayısının azaltılması

3.2.9.1. Ana Bileşenler analizi (Principal Component Analysis)

Çok değişkenli istatistiksel analizlerde, n tane bireye ilişkin p tane değişken incelenmektedir. Ancak bir bireyden ölçülen p sayıdaki değişkenin bazıları arasında yüksek derecede korelasyon (bağımlılık) bulunabilmektedir, bu da değişkenlerin bağımsızlık varsayımını bozar. Ayrıca kümeleme analizi ile elde edilecek grupların Wilks Lambda veya Hotelling-Lawley gibi çok değişkenli test istatistikleri ile test edilebilmesi için veri matrisinin en azından kare matris ($n=p$) veya ($n>p$) boyutunda olması şarttır. Aksi halde ($n<p$) olması durumunda hata matrisi tekil (singular) olacağından çok değişkenli varyans analizi yapılamaz. Bu nedenle, değişken sayısının azaltılması bu problemi bir derece ortadan kaldıracaktır (51), (57), (58), (59)

Günümüz şartlarında çok sayıda değişkeni analiz etmek her ne kadar kolay ise de, bunlara ilişkin sonuçları yorumlamak oldukça zor olabilmektedir. Halbuki varyasyonun büyük bir kısmını açıklamaya yeterli olan daha az sayıdaki değişkenin analiz sonuçlarını yorumlamak daha kolaydır. Bununla birlikte daha öncede konu edildiği gibi, kümeleme analizinde kullanılan pek çok uzaklık ölçüsü ve bu ölçüler üzerinde kurulmuş çok sayıda yöntem mevcuttur. Gerek uzaklık ölçütlerine gerekse seçilen yöntemle ilgili olarak kümeleme analizi sonuçları çok farklı olabilmekte ve araştırmacıyı kararsızlığa düşürmektedir. Bu nedenle, son yıllarda kümeleme analizinde, daha çok Ana Bileşenlerden yararlanılmaktadır. Bu yolla hem değişken sayısı azaltılmakta, hem de Çok Değişkenli analiz yöntemlerini olumsuz yönde etkileyen değişkenler arasındaki bağımlılık yapısı yok edilmektedir. Ayrıca ilk iki temel bileşen üzerindeki gözlem değerlerinin çiziminden ayrıntılı bilgi çıkarmak mümkün olabilmektedir. Elde edilen grafik araştırmacıya kümeleme analizi sonuçlarının doğruluğu ve yorumu hakkında da fikir verebilmektedir (41).

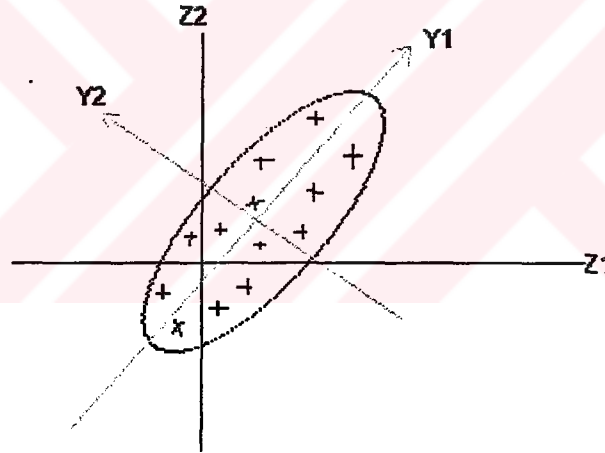
Ana bileşenler analizinin mantıksal açıklamasını yapacak olursak; n birey ve p değişkenden oluşan X veri matrisinin p boyutlu uzaydaki durumunu düşünelim. Veri matrisindeki her bir birey bir noktayı göstermek üzere, çok sayıda noktadan oluşan bir bulut olarak ifade edilebilir. Değişkenler arasında tam bir bağımsızlık söz konusu olamayacağından, bulut şeklinde ifade edilen geometrik şeklin eksenleri

birbirlerine dik olmayacak ve tanımı da yapılamayacaktır. Halbuki bu noktalar eksenleri birbirine dik bir elipsoid içine alınacak olursa daha ayrıntılı ve açıklayıcı bilgi edinilebilir. Bu amaçla uygulanan dönüştürmede, noktaların ilk eksenler boyunca sahip oldukları toplam varyans değiştirilmeden yeni eksenler birbirine dik hale getirilmektedir.

Uygulamada değişkenlerin ölçü birimlerinin birbirinden farklı olmasından $X_{p \times n}$ boyutlu ham veri matrisi değil, bunların standardize edilmiş değerlerinden oluşan $Z_{p \times n}$ standart matrisi kullanılacaktır. Bu durumda $T_{p \times p}$ bir dönüştürme matrisi olmak üzere,

$$Y_{p \times n} = T'_{p \times p} \cdot Z_{p \times n}$$

biçiminde olmaktadır. Yani birbirleri ile ilişkili z_{ij} değerlerinden dönüştürme sonunda, birbirleri ile ilişkisiz y_{ij} değerleri elde edilmektedir. Bu dönüştürme Şekil 3.6'da verilmiştir.



Şekil 3.6. Ana Bileşenler dönüşümünün iki boyutlu uzayda gösterimi

Ana Bileşenler analizinde (ABA), Varyans kovaryans yapısını izah etmek için p adet değişken yerine, bunların linear bileşkesi olan k adet ana bileşen kullanılır. Ana bileşenler analiz yöntemi daha sonraki çok değişkenli istatistik analizler için bir temel yapı oluşturur. Sözü edilen analizler; çok değişkenli regresyon, kümeleme (Cluster) analizi ve faktör analizidir (41), (55), (60).

Ana bileşenler analizinde Varyans-Kovaryans matrisi (Σ), ya da korelasyon matrisi kullanılır. Değişkenlere ilişkin vektörlerimiz,

$$X' : \{X_1, X_2, \dots, X_p\}$$

şeklinde olsun. Bunlara ilişkin özdeğerlerimiz,

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p \geq 0$$

ve özvektörlerimiz,

$$e_i' : \{e_1, e_2, e_3, \dots, e_p\}$$

şeklinde olsun. Sözü edilen Özdeğer ve Özvektörleri kısaca tanımlayacak olursak, A p'inci mertebeden bir kare matris olsun. λ bir sabit olmak üzere bu matrisin karakteristik denklemi aşağıdaki eşitlikten elde edilir.

$$|A - \lambda \cdot I_p| = 0 \quad \text{ya da} \quad \lambda^p + b_{p-1}\lambda^{p-1} + \dots + b_1\lambda + b_0 = 0 \quad (3.65)$$

Yukarıdaki p'inci dereceden polinom denklemin elde edilen p tane köküne A matrisinin özdeğerleri (ya da karakteristik kökleri) denir. A matrisinin özdeğerlerinin gerçek değerler olabilmesi için $A=A^I$ özelliği sağlanmalıdır. Aksi durumda özdeğerler karmaşık değerli olacaktır. Ayrıca $\lambda_1, \lambda_2, \dots, \lambda_p$, A matrisinin özdeğerleri ise,

$$|A| = \prod_{j=1}^p \lambda_j \quad (3.66)$$

dır. Bu durumda p'inci mertebeden bir A matrisinin rankı p'den küçük ise bu matrisin en az bir özdeğeri sıfırdır ve determinant değeri de sıfır olacaktır.

A: $p \times p$ boyutlu bir matris, λ bir sabit olmak üzere, $e: p \times 1$ boyutlu sıfır olmayan vektör aşağıdaki şartı sağlıyorsa, bu vektöre A matrisinin λ ile ilgili özvektörü adı verilir.

$$(A - \lambda I) e = 0$$

Simetrik matrisin özvektörleri diktir ve bu özellik çok değişkenli birçok analizin ortaya çıkmasını sağlamıştır (39), (41), (55).

Bunlardan yararlanarak temel bileşenler aşağıdaki eşitlik vasıtasıyla bulunabilir.

$$\underline{Y}_i = \underline{e}_i' x = e_{i1}x_1 + e_{i2}x_2 + \dots + e_{ip}x_p \quad i = 1, 2, \dots, p \quad (3.67)$$

3.2.9.2. Diskriminant (ayırma) analizi

Diskriminant analizi, kümeleme analizinden farklı olmakla birlikte, birbirlerini destekleyen ortak yönleri bulunmaktadır. Çalışmada, Diskriminant analizinin, kümeleme analizini destekler mahiyette olan bazı özelliklerinden yararlanılacağından bu konu detaylı olarak ele alınmayacaktır.

Bazı gruplar bir kısım özellikler bakımından benzerlik gösterirken, bazı özellikler yönünden de farklılık göstermektedir. Bu grupları anlamlı bir biçimde birbirinden ayırıp fonksiyonel olarak tanımlayabilmek için, k gruba ilişkin p tane özelliğinden yararlanılır. Sözü edilen özellikler (değişkenler) vasıtasıyla her bir grubu tanımlayacak bir fonksiyon geliştirilir. Hesaplanan bu fonksiyonlar vasıtasıyla, sonradan elde edilen p değişkenli bir birimin, hata payı minimum olacak şekilde gruplardan herhangi birisine atanır. Kısacası Diskriminant analizinin iki temel fonksiyonu vardır. Bunlar sırasıyla;

- 1- Grupları birbirinden ayırmaya yarayan fonksiyonları bulmak,
- 2- Elde edilen fonksiyonlar vasıtasıyla yeni gözlenen bir birimi sınıflama hatası minimum olacak şekilde gruplardan herhangi birine atamaktır.

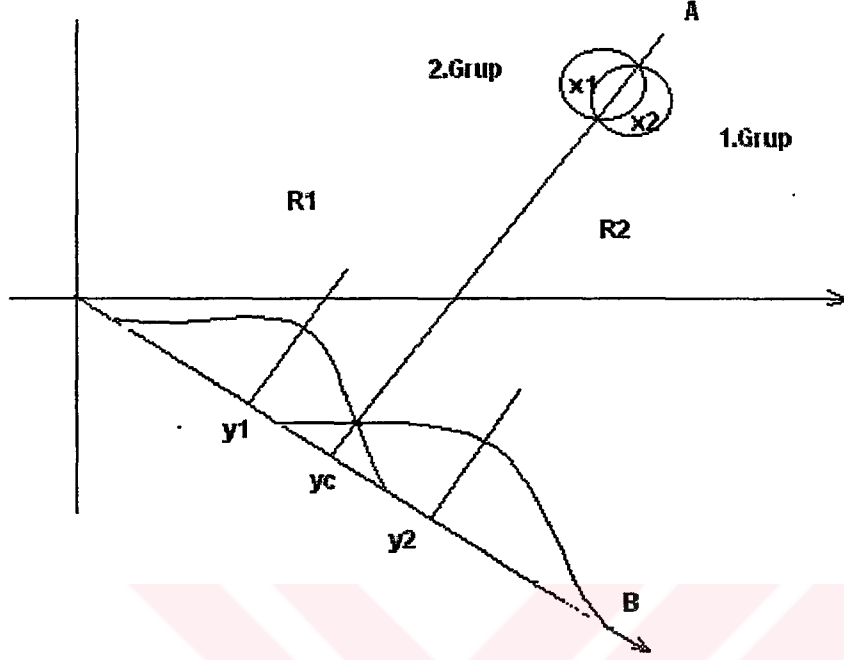
Bununla birlikte,

- a- Birimlerin sınıflandırılmasında en etkili olan değişkenleri, Diskriminant analizinin 'Stepwise' tekniğinden yararlanarak tespit etmek,
- b- Kanonik (Canonical) Diskriminant fonksiyonlarından yararlanılarak her bir genotip grubu için boyutsuz ortamda sınır harita grafiğini (territorial map) veya serpilme diyagramını elde etmek mümkündür (59).

Bir veri setine Diskriminant analizini uygulamak için bu veri setinin gereğinden fazla değişken içermemesi gerekir. Ayrıca, Diskriminant analizi, grupların merkezlerini oluşturmada ve ayırma fonksiyonlarının grupları birbirlerinden maksimum etkinlikte ayırmasını sağlamada gerekli olan değişkenleri kapsamaları gerekir. Bu nedenle veri setinde ayırma fonksiyonlarının belirlenmesinde gerekli olmayan değişkenleri ayıklamak amacıyla Stepwise Diskriminant analizi uygulanır.

Ayrıca, yeni gözlenen birimlerin uygun gruplara atanmasında etkin sınıflandırma modellerinin kurulması için de optimal sınıflamayı sağlayıcı bir

modelde gereksiz deęişkenlerin bulunmaması için Stepwise Diskriminant analizi teknięinden yararlanır.



Şekil 3.7. Diskriminant analizinde boyut indirgeme

Stepwise Diskriminant analizi, çoklu regresyon analizinde uygulanan aşamalı regresyon yaklaşımında yer alan kurallara uygun olarak yürütülür (59).

Diskriminant analizinin kümeleme analizi ile olan ilişkisine gelince; grupları belli olmayan birimler kümeleme analizi ile birbirinden ayrıldıktan sonra, bu birimlerin ilgili kümelere doğru atanıp atanmadıkları Diskriminant analizi sonuçları ile test edilmektedir (55), (59).

$k > 2$ ve $p \geq 2$ olan veri setlerine Diskriminant analizi uygulamak için, grup sayısı kadar Diskriminant fonksiyonu hesaplanır.

k grup için ortak kovaryans matrisi aşağıdaki gibi hesaplanır.

$$S_{ortak} = S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_k - 1)S_k}{n_1 + n_2 + \dots + n_k - k} \quad (3.68)$$

Bundan yararlanarak Diskriminant fonksiyonları;

$$y_i(x) = \bar{x}_i' S^{-1} x - \frac{1}{2} \bar{x}_i' S^{-1} \bar{x}_i + \ln p_i \quad (3.69)$$

biçiminde hesaplanır. Burada p_i , k grup içinde her bir grubun ilk olasılığını (prior probability) belirtmektedir. $\sum p_i = 1$ olmalıdır.

Eğer p_i değeri verilmemişse, $p_i = n_i / N$ oranı olarak alınır (59).

İkiden çok grup olması durumunda, elde edilen Diskriminant fonksiyonunun önem kontrolünde kullanılan kriterlerden ilki, 3.54 No'lu eşitlikte verilen ve Wilks tarafından geliştirilmiş olan ve genelleştirilmiş varyans olarak bilinen Λ dır.

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W + B|}$$

Burada W ve T , daha önce tanımlandığı gibi sırasıyla grup içi ve toplam varyans kovaryans matrisleridir. Önerilen Λ değerinin küçük bulunması gruplar arası farklılığın önemli olduğunun bir işaretidir ki, bu değer aynı zamanda çok değişkenli varyans analizinin de temelini teşkil etmektedir.

Gruplardaki birey sayısının yeterince büyük olması durumunda,

$$m = n - 1 - \frac{1}{2}(p + k) \quad (3.70)$$

olmak üzere, Λ değeri kullanılarak aşağıdaki test istatistik değeri bulunmaktadır.

$$\chi^2 = -m \log(\Lambda) \sim \chi^2_{p(k-1); \alpha} \quad (3.71)$$

Eşitlikteki p , n , k sırasıyla; değişken sayısı, birey sayısı ve grup sayısıdır.

Daha önce de, sözü edilen uzaklık-benzerlik ölçütlerinden Mahalanobis- D^2 uzaklık ölçütü grupları birbirinden ayırmak için kullanılmaktadır. Grupların birbirlerine olan uzaklığın karesini veren Mahalanobis uzaklık matrisinin elemanları (d_{ij}) nin

$$d(\bar{x}_i, \bar{x}_j) = D_{ij}^2 = (\bar{x}_i - \bar{x}_j)^T \cdot S_V^{-1} \cdot (\bar{x}_i - \bar{x}_j)$$

biçiminde hesaplandığı daha önceki bölümde (3.2.2.5) verilmişti.

Sınıflandırma kuralı ise, yeni gözlenen x gözlem vektörü ,

$$-\frac{1}{2}D_{ij}^2 + \ln p_i \quad (3.72)$$

değeri en az olan π_i popülasyonuna atanır (41).



4. BULGULAR ve TARTIŞMA

4.1. Kümeleme Analizinde Kullanılacak Verilerin Yapısı ve Özellikleri

Bu çalışmada, Açar (1997) tarafından Şanlıurfa Ceylanpınar Tarım İşletmesinde (TİGEM) yapılan bir araştırma sonucunda elde edilen veriler kullanılmıştır. Ölçümler 24 farklı fıstık tipine ait erkek fıstık ağaçları (*Pistacia vera* L.) üzerinde gerçekleştirilmiştir. Fıstık ağaçlarını sınıflandırmak amacıyla 38 adet farklı karakter için ölçüm alınmıştır. Bu karakterler (değişkenler) morfolojik ve fenolojik gözlemlere bağlı olarak AÇAR (1997) tarafından geleneksel metotlara dayanarak teşhis edilmiştir. Fıstık tipleri erkenci ve geççi olmak üzere temel iki gruba ayrılmaktadır. Bununla birlikte bu ikisi arasında yer alan ve bazen erkenci tiplerle, bazen de geççi tiplerle kümelenen bir kısım melez tipler orta geççi tipler olmak üzere üçüncü bir grup oluşturmuştur. Sözü edilen fıstık tipleri Çizelge 4.1.'de verilmiştir.

Çizelge 4.1. Kümeleme analizinde kullanılan fıstık tipleri ve kod numaraları

Erkenci Fıstık Tipleri	Orta Geççi Fıstık Tipleri	Geççi Fıstık Tipleri
3	1	7
4	2	13
5	8	14
6	9	17
10	12	18
11	16	19
15	22	20
24	23	21

Fıstık tiplerine ilişkin deęişkenlerin bir kısmı sürekli deęişken özellięi gösterirken, bir kısmı da kesikli deęişken özellięi göstermektedir. Bunun yanı sıra bazı karakterler Nominal özellikte olup, bu tip deęişkenlere puanlar verilmiştir. Kümeleme analizinde kullanılmak üzere her bir karakter yeniden tanımlanıp, bu karakterlere ilişkin deęişken tipleri Ek Çizelge 4.2’de verilmiştir.

Ek Çizelge 4.2 deęişkenlere ilişkin veri yapısını ve bunlara ait hesaplanacak en uygun uzaklık ölçütlerini göstermektedir. Daha öncede belirtildięi gibi veriler, farklı ölçüm ıskalalarına sahip ise en uygun uzaklık ölçüsü Öklid (Euclid) veya Öklid uzaklığının karesi (karesel Öklid) dir.

4.2. Verilerin Kümeleme Analizi İçin Hazır Hale Getirilmesi

Ek Çizelge 4.3’de verilen 24 farklı erkek fıstık ağacına ilişkin ölçümler hem kesikli hem de sürekli deęişken özellięine sahip oldukları için verilerin standardize edilmesi gerekmektedir. Ancak daha önce verilere ilişkin MANOVA’nın yapılabilmesi için $n < p$ boyutunda olan veri matrisinin en azından $n \geq p$ boyutuna indirgenmesi gerekmektedir. Bu nedenle, Ek Çizelge 4.4’de verilen korelasyon analizi sonuçlarından yararlanarak, aralarında yüksek derecede (** : $P < 0.01$) ilişki bulunan deęişkenler tespit edilip çıkarılmıştır. Bu suretle hem fıstık ağaçlarının sınıflandırılmasında birbirinden bağımsız olan karakterler (deęişkenler) tespit edilmiş, hem de veri matrisinin boyutu ($n \geq p$) indirgenmiştir. Korelasyon analizi ile seçilmiş ve Kümeleme analizinde kullanılacak olan karakterler Ek Çizelge 4.5’de verilmiştir.

Ek Çizelge 4.5’de verilen karakterlere ilişkin korelasyon analizi sonuçları Ek Çizelge 4.6’de görülmektedir. Çizelgede, korelasyon katsayısı 0.47’den büyük ($P < 0.01$) olan tüm deęişkenlerin elemine edildięi görülmekte ve yüksek korelasyonlu deęişkenler bulunmamaktadır.

Ek Çizelge 4.5’de elemine edilmiş olan deęişkenlerin, standardize edilmiş olan deęerleri Ek Çizelge 4.7’de verilmiştir.

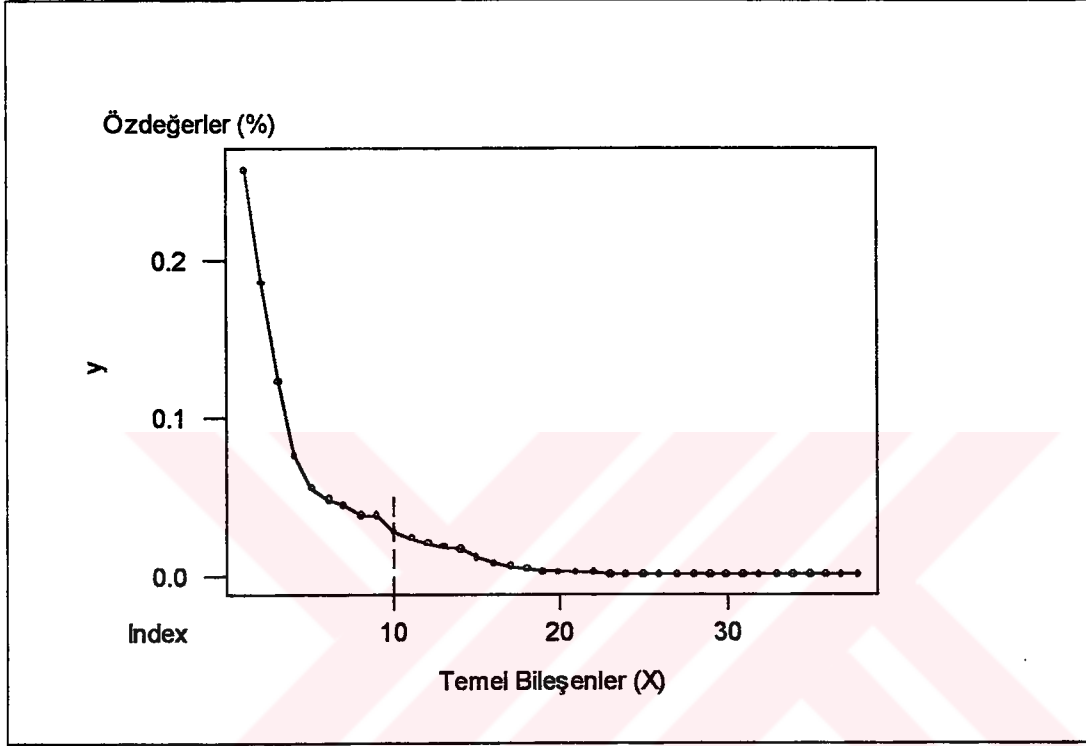
Bunun yanı sıra boyut indirgemede ikinci bir yöntem olarak, Ana Bileşenler analizinden de (Principal Component Analysis) yararlanarak ÇDVA ‘nın

yapılabilmesi için $n < p$ boyutunda olan veri matrisinin boyutu indirgenerek değişken sayısı azaltılmıştır. Ayrıca bireylerin kümelenmesinde en etkili ve en fazla varyasyonu açıklayan Ana Bileşenler Çizelge 4.8'deki gibi elde edilmiştir.

Çizelge 4.8. Ana Bileşenler (Principal Component Analysis) analizi sonuçları

Bileşenler	Özdeğer	Varyans	Eklemeli Varyans (Cumulative)
1	9.7703	0.257	0.257
2	7.0284	0.185	0.442
3	4.6561	0.123	0.565
4	2.9303	0.077	0.642
5	2.1037	0.055	0.697
6	1.8353	0.048	0.745
7	0.0440	0.044	0.789
8	0.0380	0.038	0.828
9	0.0370	0.037	0.864
10	0.0270	0.027	0.892
11	0.0230	0.023	0.914
12	0.0200	0.020	0.935
13	0.6365	0.017	0.952
14	0.5900	0.016	0.967
15	0.4076	0.011	0.978
16	0.2923	0.008	0.986
17	0.1746	0.005	0.990
18	0.1576	0.004	0.994
19	0.0890	0.002	0.997
20	0.0595	0.002	0.998
21	0.0428	0.001	0.999
22	0.0272	0.001	1.000
23	0.0000	0.000	1.000
24	0.0000	0.000	1.000
25	0.0000	0.000	1.000
26	0.0000	0.000	1.000
27	0.0000	0.000	1.000
28	0.0000	0.000	1.000
29	0.0000	0.000	1.000
30	0.0000	0.000	1.000
31	0.0000	0.000	1.000
32	0.0000	0.000	1.000
33	0.0000	0.000	1.000
34	0.0000	0.000	1.000
35	0.0000	0.000	1.000
36	0.0000	0.000	1.000
37	0.0000	0.000	1.000
38	0.0000	0.000	1.000

Analizde kullanılacak en etkili Ana Bileşenlerin sayısını belirlemek amacı ile özdeğerlerin, varyansı açıklama oranlarından yararlanarak Ana (temel) Bileşenlere ilişkin grafik (Scree Graph) çizdirilmiştir. Böylece, Ana Bileşenlerin sayısı Şekil 4.1’de verilen grafik vasıtasıyla belirlenmiş olmaktadır.



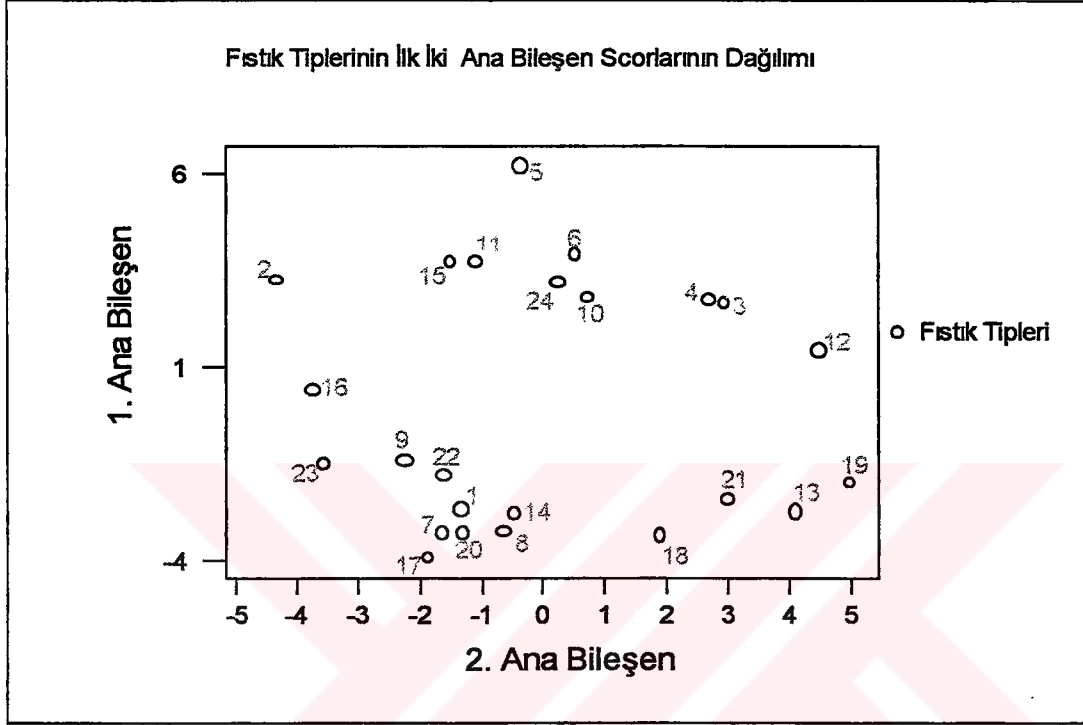
Şekil 4.1. En etkili olan Ana Bileşen sayısının (Scree Graph) grafik metodu ile belirlenmesi

Grafik (Scree graph) vasıtasıyla belirlenen ve varyansın % 89'unu açıklayan ilk on temel bileşen 3.67 No'lu eşitlik vasıtasıyla elde edilerek Ek Çizelge 4.9'de verilmiştir.

Kümeleri ve küme sayılarını belirlemeden önce bölüm 3.2.9.1.'de belirtildiği gibi ilk iki temel bileşen vasıtasıyla birimlerin dağılımını ve kaç kümeye ayrılacaklarını Şekil 4.2'de çizilmiş olan grafik vasıtasıyla tahmin etmek mümkündür.

Grafikten de anlaşılacağı gibi birimlerin oluşturabileceği olası küme sayısı iki veya üç olabilir. Bu durum, özellikle k-Ortalama yöntemini kullanmak isteyen

arařtırmacılar için bir ön bilgi mahiyetini taşıyabilmektedir. Ancak, bu grafięe bakılarak kesin bir hüküm vermek doğru olmayıp, bunu doğrulamak için kümeleme analizi metotlarına başvurulmuřtur. Bununla birlikte bir kısım test istatistięi kriterleri ile de sonuçların anlamlı olup olmadıęı test edilmiřtir.



řekil 4.2. İlk iki Temel Bileřen vasıtasıyla elde edilen grupların (birimlerin) daęılımı

Kümeleme analiziyle fıřık tiplerinin kümelenmesini saęlayacak ve Ana Bileřenler analizi ile boyutu indirgenmiř olan deęiřkenlere iliřkin Skorlar Ek Çizelge 4.10'de verilmiřtir.

Ayrıca, üçüncü bir metot olarak bölüm 4.6'da verilmiř olan Stepwise Diskriminant analizinden yararlanılarak birimlerin kümelere atanmasında en etkili rol oynayan deęiřkenler elemine edilerek deęiřkenlerin boyutu indirgenmiř ve elde edilen etkili deęiřkenlerle kümeleme analizi yapılmıřtır.

4.3. Uzaklık Matrisinin Elde Edilmesi

Daha önce de belirtildiği gibi birimler arasındaki uzaklıkları belirlemek için bölüm 3.2.2.'de verilen uzaklık ölçülerinden yararlanılarak uzaklık matrisi elde edilmektedir. Ek Çizelge 4.5'de verilen ve korelasyon analizi sonucu elemine edilmiş değişkenlerden yararlanılarak birimler arasındaki Öklid uzaklığını belirlemede 3.3 No'lu eşitlik kullanılarak Ek Çizelge 4.11'deki uzaklık matrisi elde edilmiştir.

Aynı şekilde Öklid uzaklığının karesini elde etmek için 3.5 No'lu eşitlikten yararlanılarak Ek Çizelge 4.12'deki uzaklık matrisi elde edilmiştir.

Benzer şekilde Ana Bileşenler Analizi ile boyutu indirgenen Ek Çizelge 4.10'daki değişkenlere ait skorlara ilişkin Öklid ve Öklid uzaklıklarının karesi sırasıyla Ek Çizelge 4.13 ve Ek Çizelge 4.14'de verilmiştir.

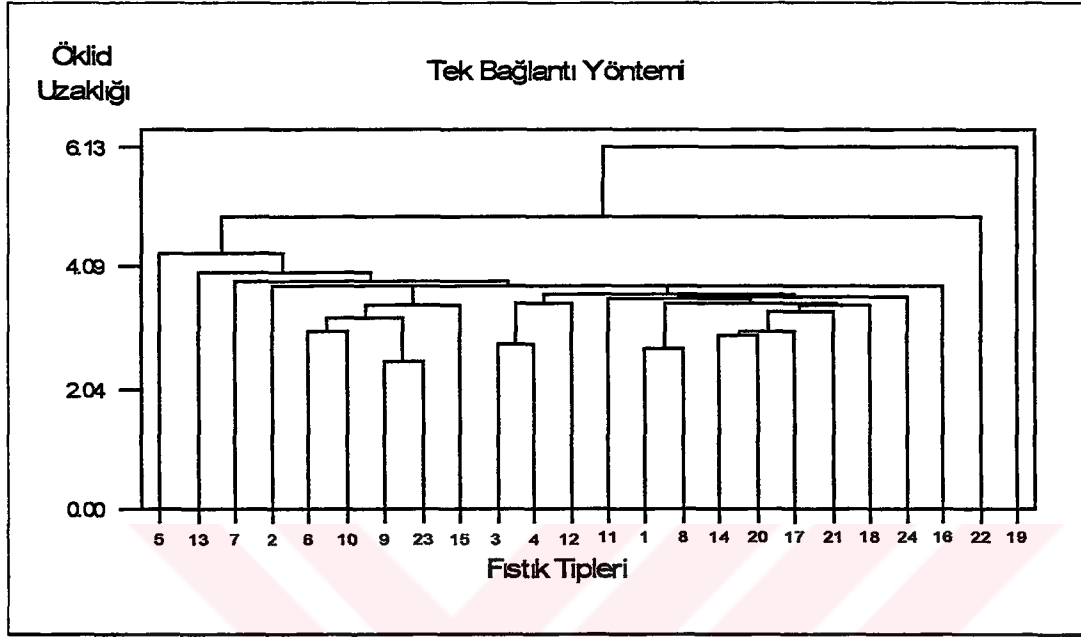
4.4. Aşamalı Kümeleme Yöntemleri

Kümeleme analizi yapmak için çok sayıda kümeleme yöntemi önerilmiş olmasına karşılık, bu çalışmada, daha çok literatürlerde yaygın olarak kullanılan yöntemler karşılaştırılmıştır.

4.4.1. Tek Bağlantı (en yakın komşu) kümeleme yöntemi ile fıstık tiplerinin kümelenmesi

Fıstık tipleri, korelasyon analizi ile elemine edilmiş olan değişkenlerle, 3.32 No'lu eşitlikten yararlanılarak kümelenecek olursa Şekil 4.3'deki ağaç grafiği elde edilmektedir. Ağaç grafiğinde görüldüğü gibi Tek Bağlantı yöntemi, Öklid uzaklığına göre fıstık tiplerini anlamlı bir şekilde sınıflandıramamıştır. Çünkü erkenci fıstık tipleri ile geççi fıstık tiplerini birbirinden ayıramamakla birlikte 19 No'lu fıstık tipi ile 22 No'lu fıstık tipini tek başına bir kümelemiş, diğer erkenci ve geççi fıstık tiplerini de karışık olarak üçüncü bir kümede sınıflandırmıştır. Diğer bütün uzaklık ölçüleri ile de Tek Bağlantı yöntemine ilişkin kümeleme analizi

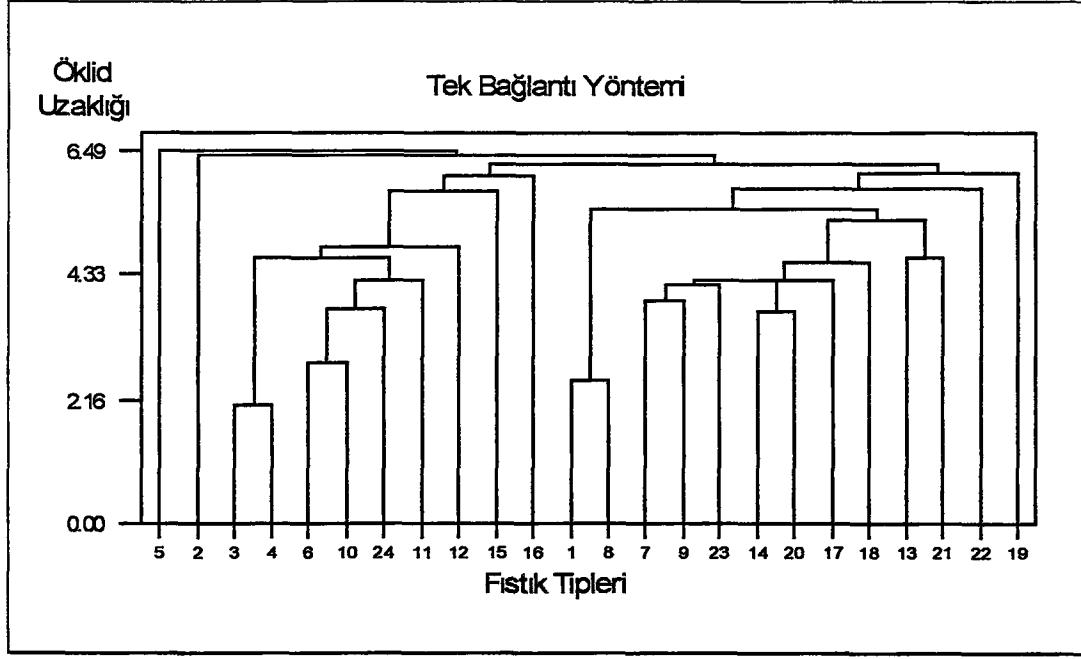
yapılmış, ancak fıstık tipleri bu uzaklık ölçüleriyle de anlamlı kümeler oluşturamamıştır. Bu nedenle, diğer uzaklık ölçülerine ilişkin ağaç grafiklerine tekrar yer verilmemiştir.



Şekil 4.3. Korelasyon analizi ile boyutu indirgenmiş olan değişkenlere ilişkin Tek Bağlantı yöntemi ile elde edilen ağaç grafiği

Ana Bileşenler Analizi ile boyutu indirgenmiş olan değişkenler vasıtasıyla fıstık tiplerini sınıflandıracak olursak Şekil 4.4'deki ağaç grafiği elde edilir.

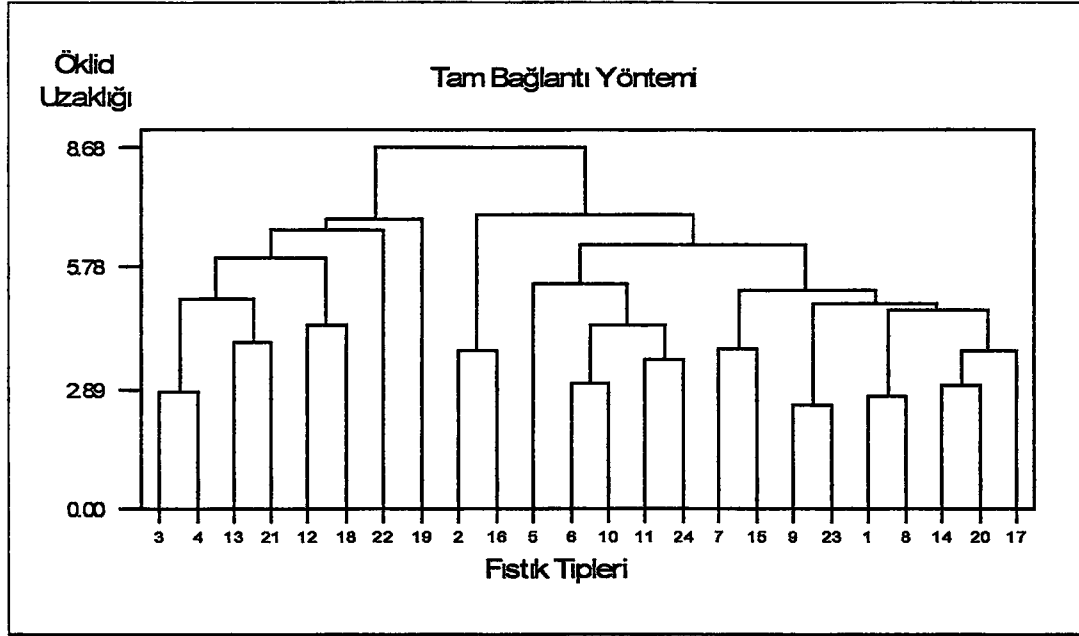
Grafikten' de görüldüğü gibi bu yöntem ile de anlamlı bir kümeleme elde edilememiştir. 2 No'lu ve 5 No'lu fıstık tipleri diğerlerinden ayrılarak tek başına kümelenebilirlerdir. Ayrıca fıstık tiplerine ilişkin bağlantı uzaklıkları birbirine çok yakın olup, açık bir şekilde birbirinden ayrılmamıştır. Bu yöntemle anlamlı sonuçlar elde edilemediğinden istatistik analizi sonuçları verilmemiştir.



Şekil 4.4. Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlere ilişkin Tek Bağlantı yöntemi ile elde edilen ağaç grafiği

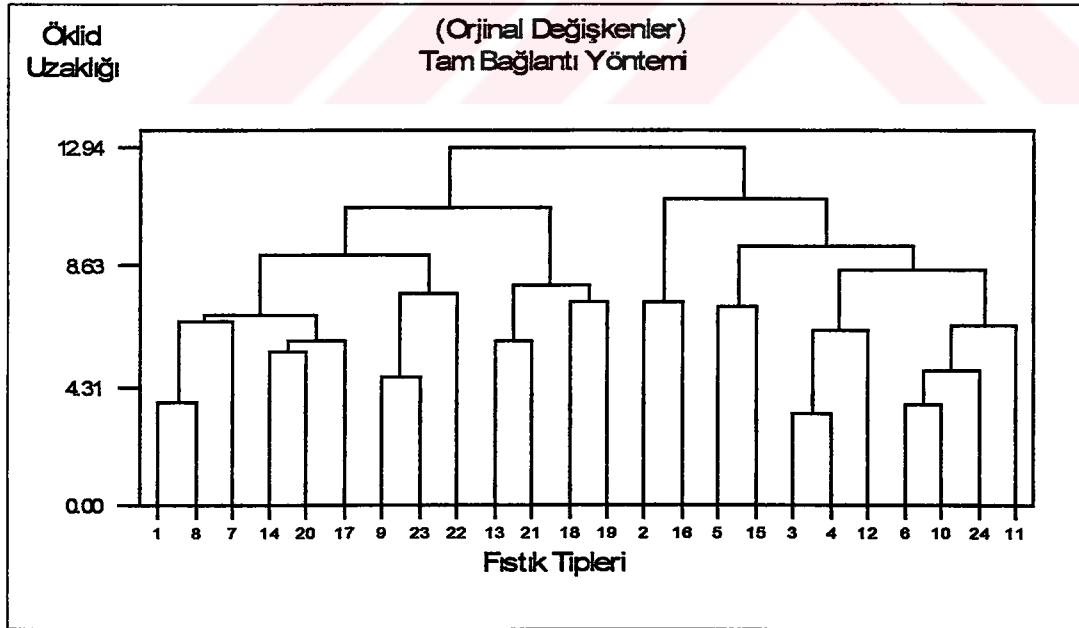
4.4.2. Tam Bağlantı (en uzak komşu) kümeleme yöntemi ile fıstık tiplerinin kümelenmesi

3.34 No'lu eşitlikten yararlanarak fıstık tiplerini korelasyon analizi ile elemine edilen değişkenlere göre kümelediğimizde, Şekil 4.5' deki ağaç grafiği elde edilmiştir. Grafikten de görüldüğü gibi Tam Bağlantı yöntemi Öklid uzaklığına göre fıstık tiplerini iki kümeye ayırmıştır. Ancak 15 No'lu erkenci fıstık tipini geççi tiplerle birlikte sınıflandırmıştır. Bu ise çok önemli bir hata olarak kabul edildiğinden kümeleme bu noktadan hatalı kabul edilmektedir. Bununla birlikte diğer bütün uzaklık ölçüleri ile de Tam Bağlantı yöntemine ilişkin kümeleme analizi yapılmıştır. Sonuçta Manhattan uzaklığının dışındaki diğer uzaklık ölçülerinin de aynı sonucu verdiği gözlenmiştir. Manhattan uzaklığından elde edilen kümeler ise tutarsızdır. Çünkü burada da erkenci tipler geççi tiplerle birlikte hatalı kümeler oluşturmuştur.



Şekil 4.5. Korelasyon analizi ile elemine edilen değişkenlerle Tam Bağlantı yöntemine göre elde edilen ağaç grafiği

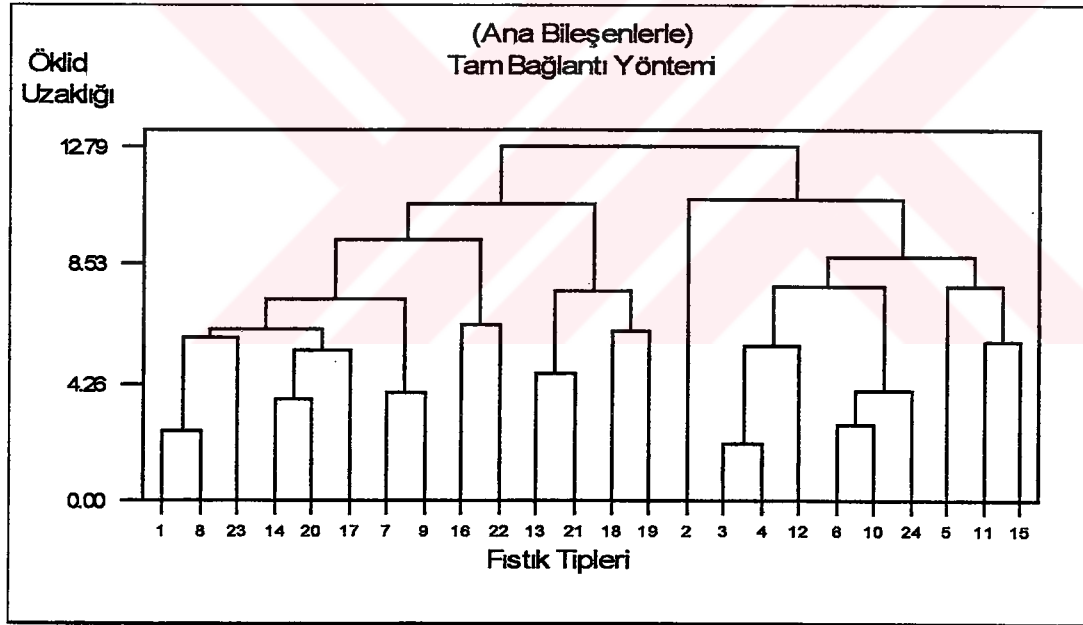
Benzer işlemler orijinal değişkenlerle yapıldığında, Şekil 4.6'da verilen ağaç grafiğindeki küme yapısı elde edilir.



Şekil 4.6. Orijinal değişkenlerle Tam Bağlantı kümeleme yöntemine göre elde edilen ağaç grafiği

Ağaç grafiğinde, orijinal değişkenlerle elde edilen kümelerde 2 ile 16 No'lu orta geççi tiplerin erkenci tiplerle birlikte kümelendiği görülmektedir. Bunun nedeninin orta geççi fıstık tiplerinin, erkenci fıstık tipleri ile geççi fıstık tiplerinin birer melezi olmasından kaynaklandığı söylenilebilir. Tam Bağlantı kümeleme yönteminin alt grupları ayırmada pek başarılı olmadığı görülmüştür.

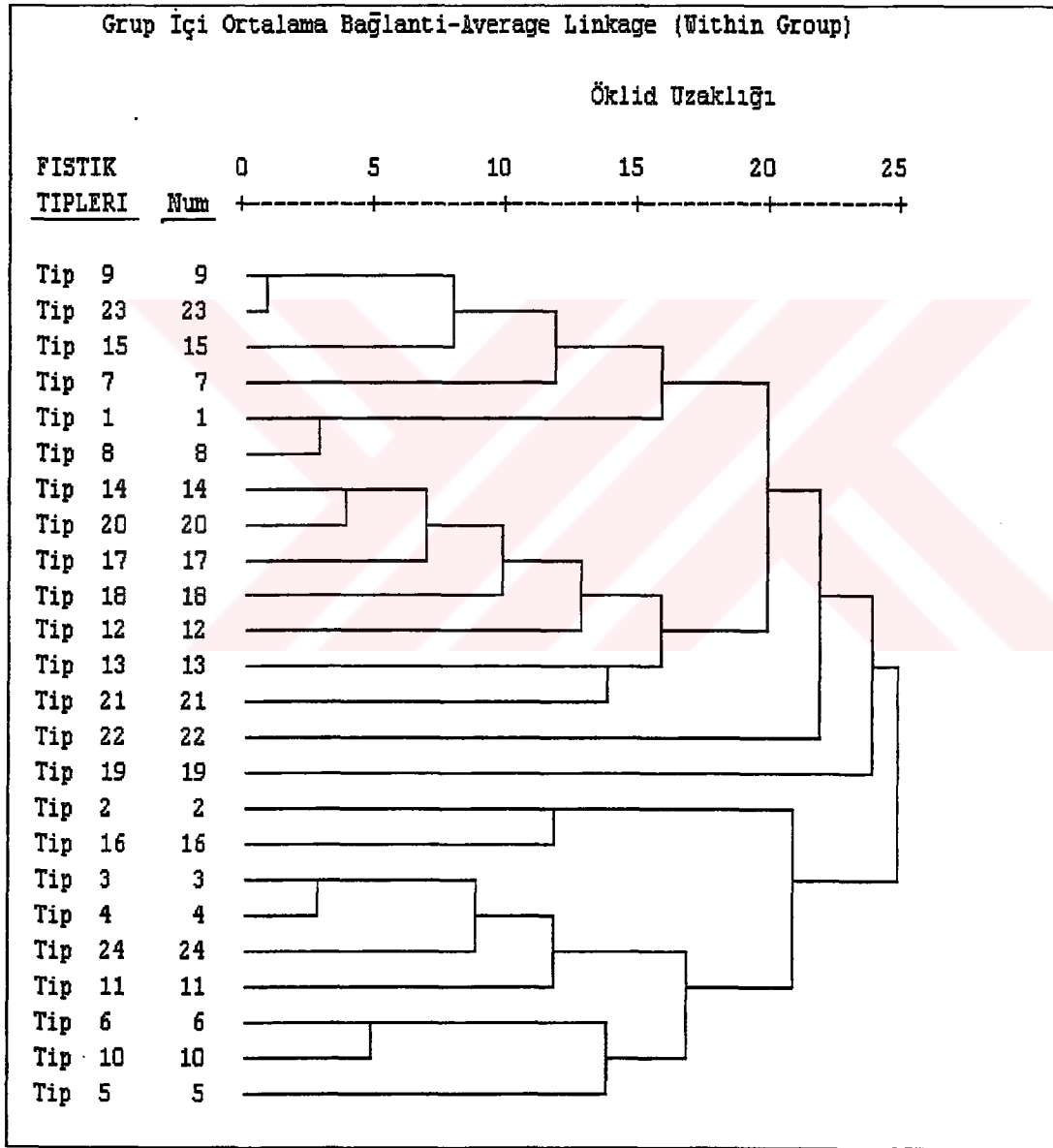
Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlerle fıstık tiplerini sınıflandırdığımızda Şekil 4.7'deki ağaç grafiği elde edilmiştir. Bu yöntemle, fıstık tipleri iki kümeye ayrılmış ve orta geççi bir tip olan 2 No'lu fıstık ağacı erkenci tiplerle birlikte kümelendi, diğer fıstık tipleri ise geççi tiplerle ikinci bir küme oluşturmuştur. Ana Bileşenler analiziyle boyutu indirgenmiş olan değişkenlerin kümelmesiyle elde edilen sonuçlar şekil olarak daha anlamlı görünmektedir.



Şekil 4.7. Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlere ilişkin fıstık tipleri ile Tam Bağlantı yöntemine göre elde edilen ağaç grafiği

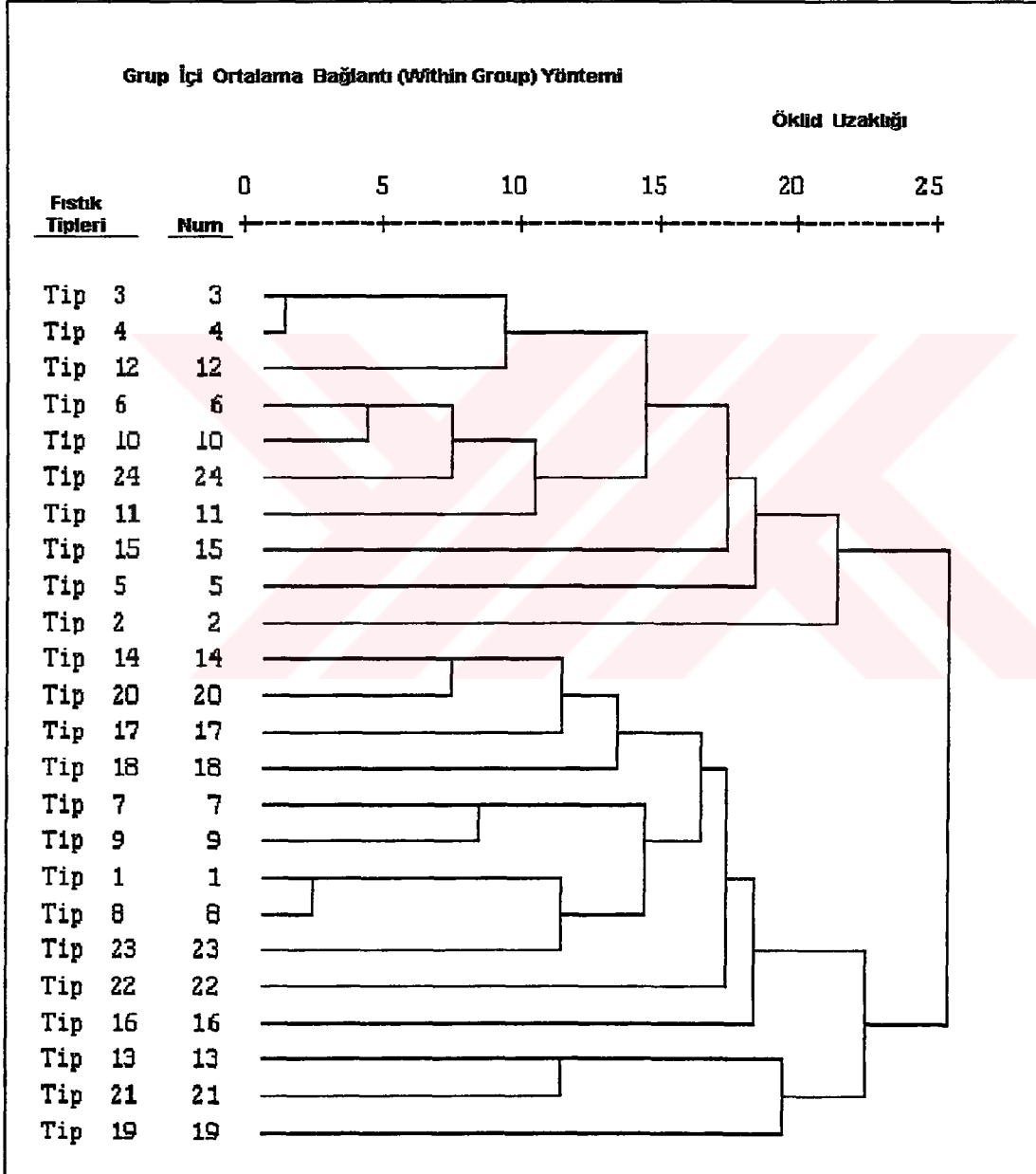
4.4.3. Gruplar İçi Ortalama Bağlantı kümeleme yöntemi

3.36 ve 3.39 No'lu eşitliklerden yararlanılarak Şekil 4.8 ve 4.9'de verilen ağaç grafiği elde edilmiştir. Grup İçi Ortalama Bağlantı kümeleme yöntemine göre korelasyon analizi ile elemine edilen değişkenlerle yapılan kümeleme analizi sonucunda, fıstık tipleri anlamlı gruplara ayrılmamıştır. Çünkü 2 ile 16 No'lu orta geççi olan iki fıstık tipi erkenci tiplerin arasında yer almıştır. Bundan daha önemlisi ise, 15 No'lu erkenci fıstık tipi geççi tiplerle birlikte hatalı olarak kümelenmiştir.



Şekil 4.8. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklığı kullanılarak Grup İçi Ortalama Bağlantı yöntemi ile elde edilen ağaç grafiği.

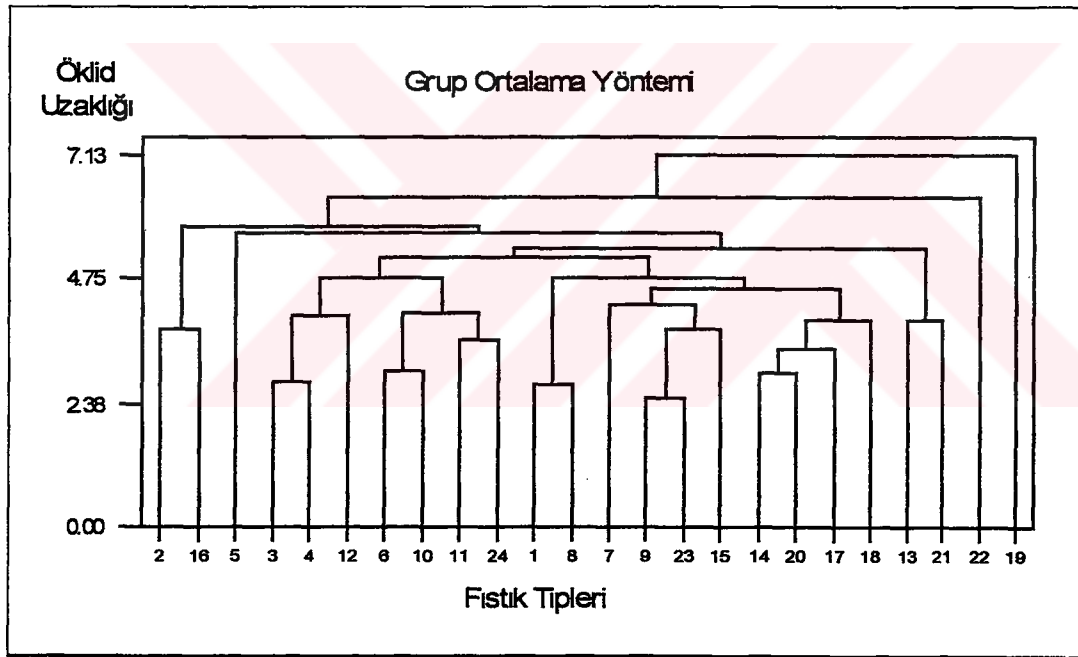
Aynı yöntem, Ana Bileşenler analizi ile elde edilen ve boyutu indirgenmiş skor değişkenleri üzerine uygulanacak olursa Şekil 4.9'deki ağaç grafiği elde edilmektedir. Burada, fıstık tiplerinin erkenci ve geççi olmak üzere iki gruba ayrıldığı gözlenmektedir. Ancak, orta geççi tiplerden olan 2 ile 12 No'lu fıstık tipleri erkenci tiplerle birlikte küme oluşturmuştur. Fıstık tipleri, Ana bileşenler analizi ile elde edilen ağaç grafiğinde, korelasyon analizi ile elde edilen değişkenlere ilişkin ağaç grafiğinden daha belirgin bir şekilde birbirinden ayrılmıştır.



Şekil 4.9. Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlerle Grup İçi Ortalama Bağlantı yöntemine göre elde edilen ağaç grafiği

4.4.4. Gruplar Arası Ortalama (Grup Ortalama) Bağlantı kümeleme yöntemi

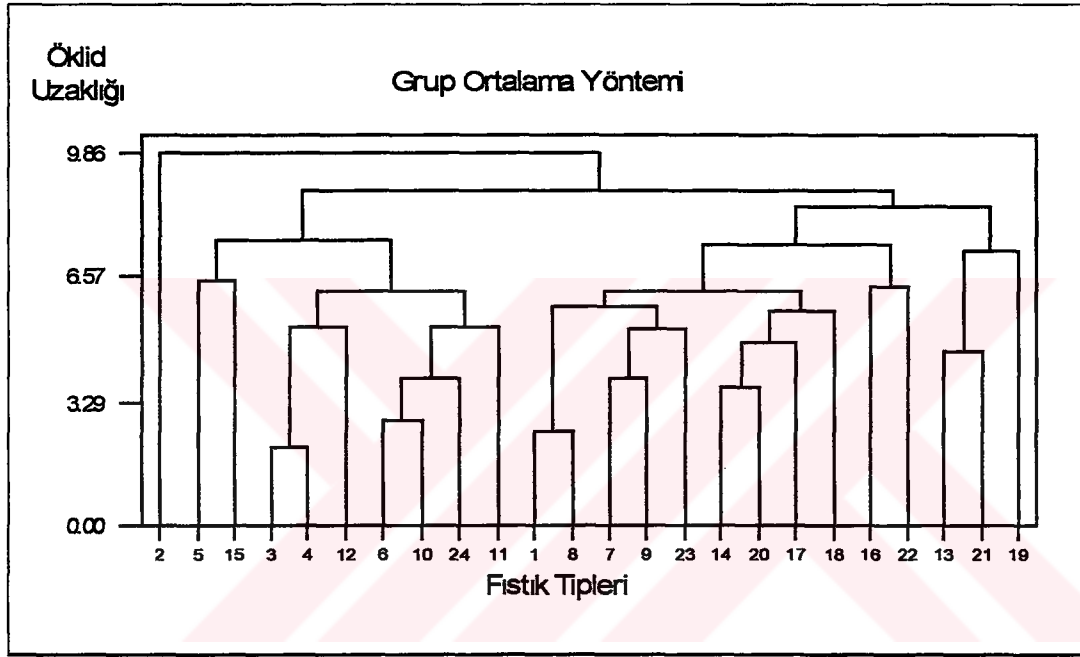
3.40 No'lu eşitlikten yararlanılarak korelasyon analizi ile elemine edilen değişkenlerle fıstık tiplerini kümelediğimizde, Şekil 4.10 deki ağaç grafiği elde edilmektedir. Grafikten' de görüldüğü gibi bu yöntem 19 ve 22 No'lu fıstık tiplerini tek başına birer küme, diğer fıstık tiplerini ise ayrı bir küme olarak algılamıştır. Ancak, bu şekilde elde edilen sonuçlar gerçek küme yapısını yansıtmamaktadır. Çünkü; burada da birbirinden ayrılması gereken erkenci ve geççi fıstık tipleri, birbirinden ayrılamamış ve karışı olarak kümeleneşlerdir. Yöntem, Öklid uzaklığı dışındaki uzaklık ölçülerinde de anlamsız sonuçlar verdiğinden ayrıca yer vermeye gerek görülmemiştir.



Şekil 4.10. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklığı kullanılarak Gruplar Arası Ortalama Bağlantı yöntemiyle elde edilen ağaç grafiği

Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlerle fıstık tiplerini sınıflandıracak olursak Şekil 4.11'deki ağaç grafiği elde edilmektedir. Ancak analiz sonuçları, sözü edilen yöntemin fıstık tiplerini anlamlı kümelere

ayırmadığını göstermiştir. Fıstık tiplerini iki kümeye ayırdığımızda 2 No'lu fıstık tipinin tek başına bir grupta, diğerlerinin ise ayrı bir grupta kümelendiği görülmektedir. Ancak fıstık tipleri üç kümeye ayrılarak incelendiğinde, 2 No'lu orta geççi fıstık tipinin tek başına bir küme oluşturduğunu, diğer erkenci ve geççi çeşitlerin ise birbirlerinden ayrılıp iki ayrı küme oluşturdukları gözlenmektedir. Geriye kalan orta geççi çeşitler ise erkenci ve geççi fıstık tiplerinin arasına dağılmıştır.

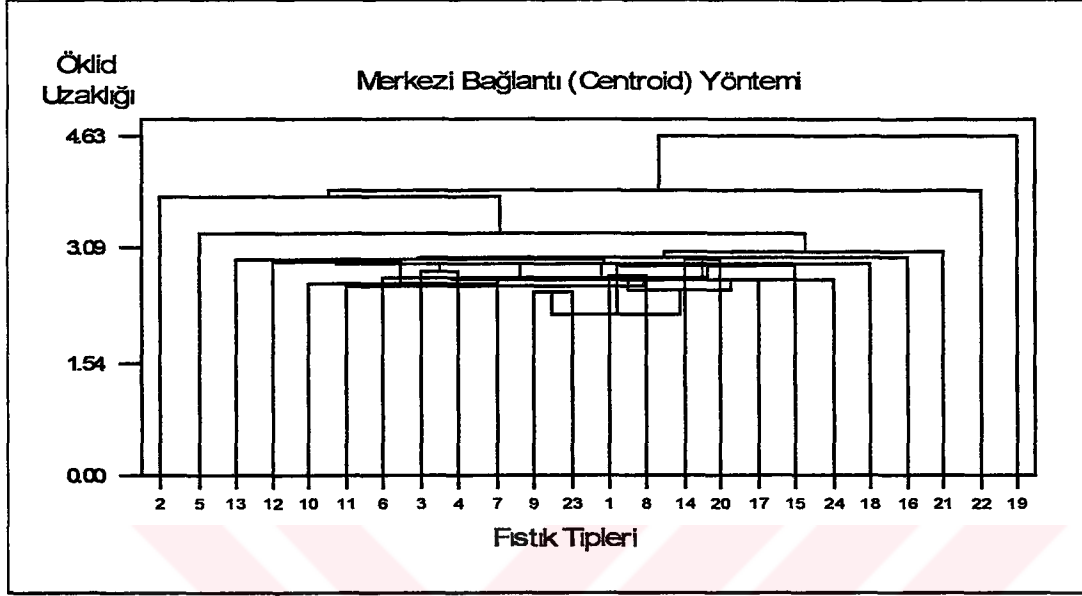


Şekil 4.11. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenler vasıtasıyla Gruplar Arası Ortalama Bağlantı yöntemine göre elde edilen ağaç grafiği

4.4.5. Merkezi (Centroid) Bağlantı kümeleme yöntemi

3.41 No'lu eşitlikten yararlanarak korelasyon analizi ile elemine edilmiş değişkenler vasıtasıyla fıstık tipleri kümelendiğinde, Şekil 4.12' deki ağaç grafiği elde edilmektedir. Ağaç grafiğinde Merkezi Bağlantı yöntemi Öklid uzaklığına göre fıstık tiplerini anlamlı kümelere ayırmamakla birlikte, oldukça karışık bir bağlantı gerçekleştirmiştir. Bununla birlikte kümeleme analizini yapmış olduğumuz Minitap

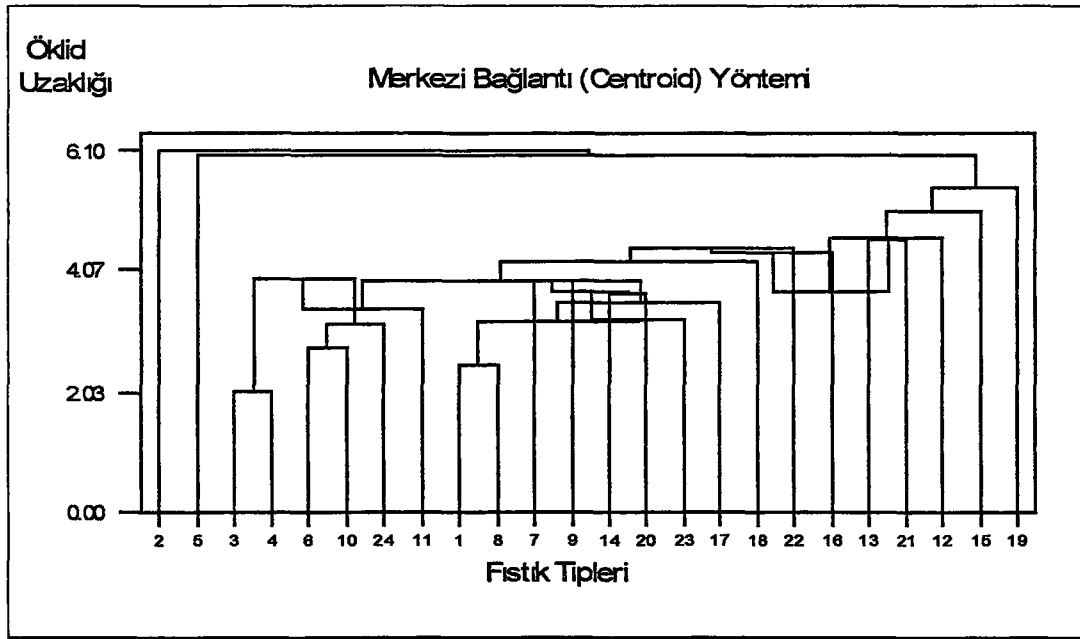
paket programından, fıstık tiplerinin ayrılabilirleri küme yapılarına ilişkin menü işaretlenmiş olmasına rağmen, buradan da anlamlı bir küme yapısının olmadığı cevabı alınmıştır.



Şekil 4.12. Korelasyon Analiziyle elemine edilmiş değişkenlerle Öklid uzaklığı kullanılarak Merkezi Bağlantı yöntemine göre elde edilen ağaç grafiği

Diğer uzaklık ölçüleri ile de Merkezi Bağlantı yöntemine ilişkin kümeleme analizi yapılmış, ancak bu ölçütlerle de fıstık tipleri anlamlı gruplara ayrılmamıştır.

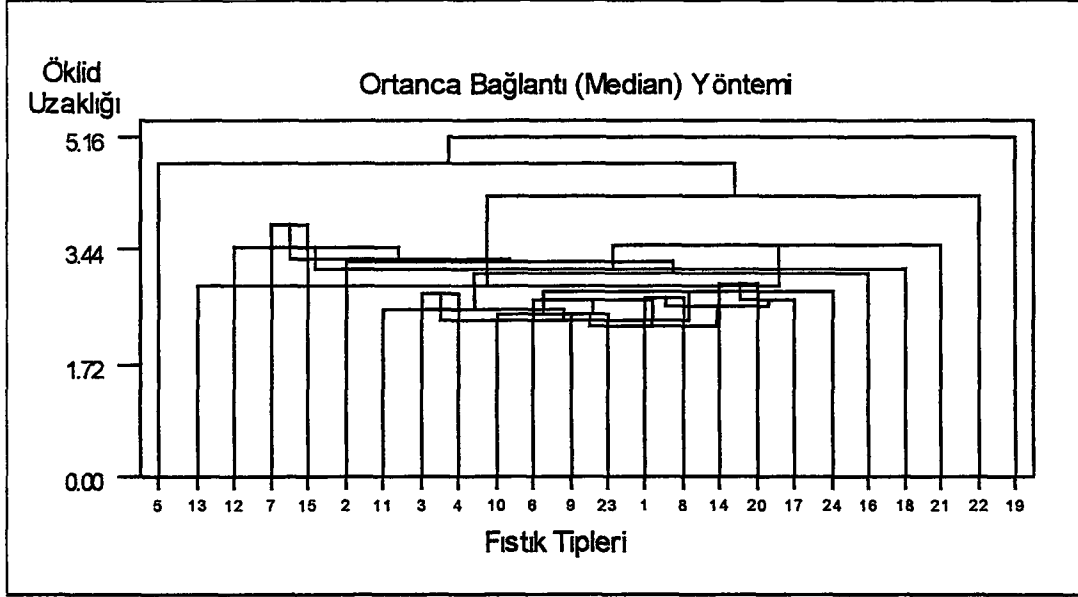
Ana Bileşenler Analizi ile boyutu indirgenmiş olan değişkenlere ilişkin Merkezi Bağlantı kümeleme yöntemi kullanılarak fıstık tipleri sınıflandırıldığında yine anlamlı bir küme yapısı elde edilememiştir. Bu sonuç, Şekil 4.13 deki ağaç grafiğinde belirgin bir şekilde görülebilmektedir. Burada 2 ve 5 No'lu fıstık tipleri tek başına ayrı bir küme oluşturmuş, diğer fıstık tipleri ise erkenci ve geççi olmak üzere karışık olarak bu kümelere sırayla belirli uzaklıklarda bağlanmışlardır. Bu nedenle, diğer test istatistiklerinin bu yöntem için uygulanmasına gerek duyulmadı.



Şekil 4.13. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlerle Merkezi Bağlantı yöntemine göre elde edilen ağaç grafiği

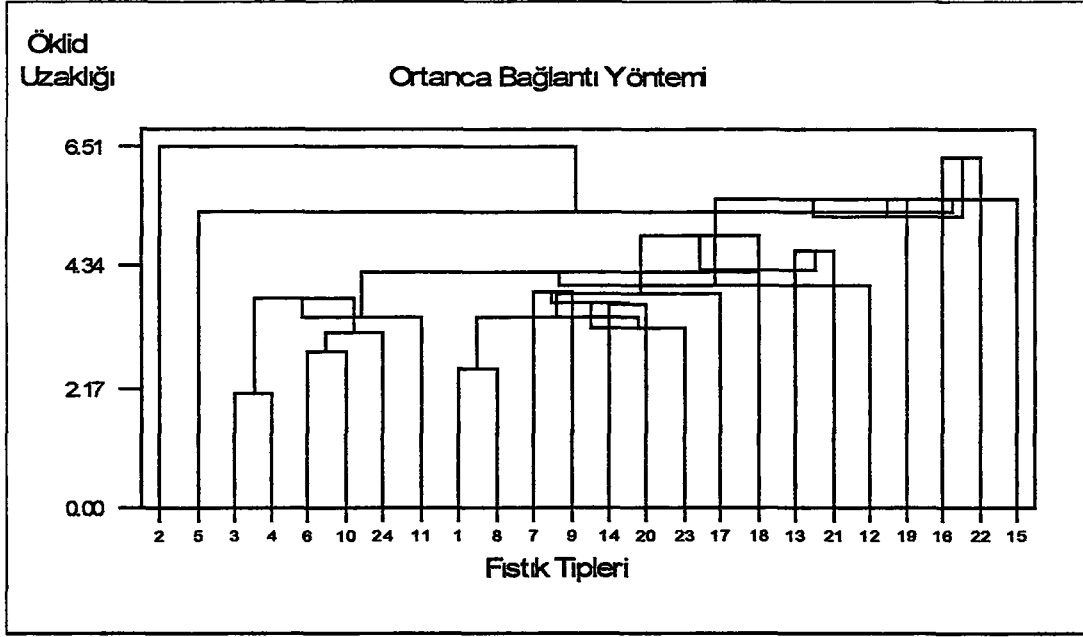
4.4.6. Ortanca (Median) Bağlantı kümeleme yöntemi

3.45 No'lu eşitlikten yararlanılarak fıstık tipleri kümelendiğinde, Şekil 4.14' deki ağaç grafiği elde edilmektedir. Ağaç grafiğinde de görüldüğü gibi Ortanca Bağlantı yöntemi Öklid uzaklığına göre fıstık tiplerini anlamlı kümelere ayırmamıştır. 5, 19 ve 22 No'lu fıstık tipleri tek başına birer küme oluşturarak erkenci ve geççi fıstık tiplerinin karışık olarak kümelendikleri dördüncü kümeye en uzak noktada bağlanmışlardır. Diğer uzaklık ölçülerinden yararlanılarak ta Ortanca Bağlantı yöntemine ilişkin kümeleme analizi yapılmıştır. Ancak, bu ölçütler ile de fıstık tipleri anlamlı gruplara ayrılmamıştır. Bu sebeple, sözü edilen diğer uzaklık ölçülerine ilişkin ağaç grafiklerine yer verilmemiştir.



Şekil 4.14. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklıkları kullanılarak Ortanca Bağlantı yöntemi ile elde edilen ağaç grafiği

Sözü edilen eşitlikten yararlanılarak Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlerin yardımıyla fıstık tipleri sınıflandırıldığında Şekil 4.15'deki ağaç grafiği elde edilmektedir. Ancak bu değişkenlerle Ortanca Bağlantı kümeleme yöntemi fıstık tiplerini birbirinden ayırıp, anlamlı bir şekilde sınıflandıramamıştır. Ağaç grafiğinde 2 No'lu fıstık tipi ile 5 No'lu fıstık tipi tek başına bir küme oluştururken, diğer fıstık tipleri birbirleriyle anlamsız bir şekilde bağlanmışlardır. Alt kümeler incelendiğinde, bu yöntemde de erkenci tiplerle geççi tiplerin birbirleri ile karışık kümelandikleri sözü edilen ağaç grafiğinden görülebilmektedir.



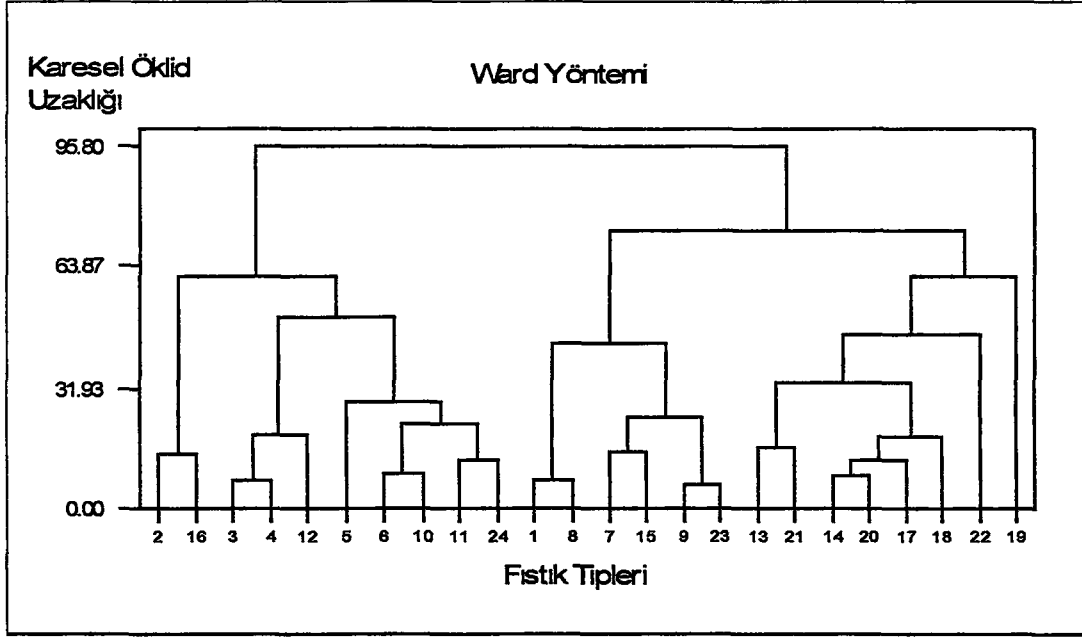
Şekil 4.15. Ana Bileşenler analizi ile boyutu indirgenen değişkenlerin Öklid uzaklığı kullanılarak Ortanca Bağlantı yöntemiyle elde edilen ağaç grafiği

4.4.7. Ward kümeleme tekniği

3.47, 3.48 ve 3.49 No'lu eşitliklerden yararlanarak Korelasyon analizi ile elemine edilen değişkenler vasıtasıyla Ward kümeleme yöntemi ile elde edilen ağaç grafiği aşağıda Şekil 4.16'de verilmiştir.

Ağaç grafiğini incelediğimizde erkenci bir tip olan 15 No'lu fıstık tipi, geççi fıstık tipleri ile birlikte; orta geççi tipler olan 2 ile 16 No'lu fıstık tipleri ise erkenci fıstık tipleri ile birlikte sınıflanmıştır.

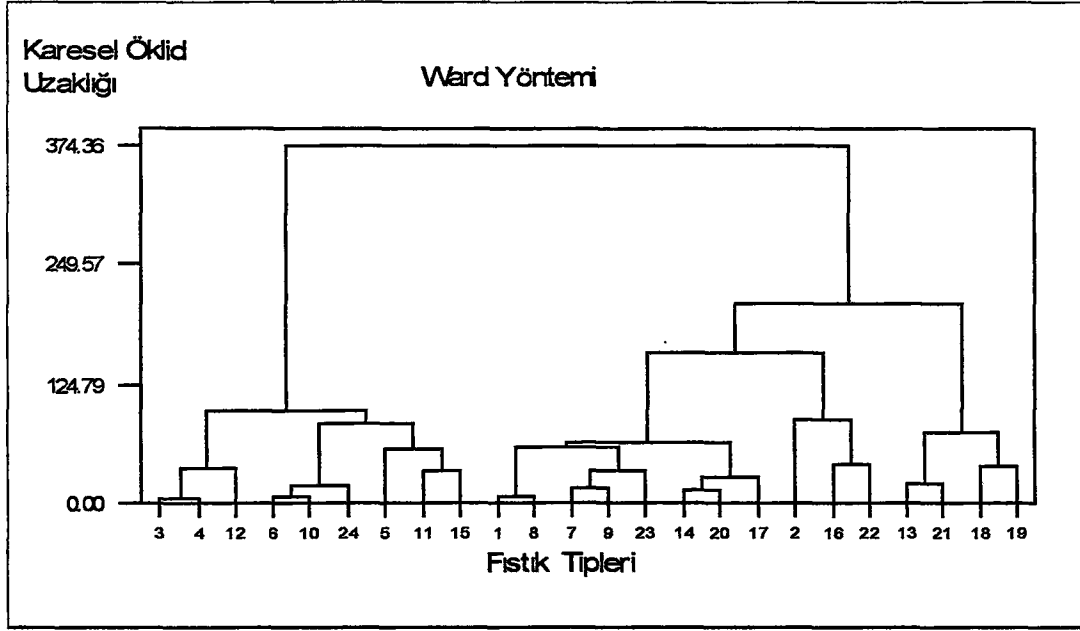
Diğer uzaklık ölçüleri kullanılarak çizilen ağaç grafiklerinden Manhattan ile Pearson Kare uzaklığı ile elde edilen sonuçlar Öklid uzaklığının karesi ile aynı sonucu vermiştir. Pearson uzaklığı ile Öklid uzaklığından yararlanarak elde edilen ağaç grafikleri ise birbirine benzerlik göstermiş olup yine hatalı bir kümeleme söz konusudur.



Şekil 4.16. Korelasyon analizi ile elemine edilen değişkenler vasıtasıyla Ward yöntemi kullanılarak fıstık tiplerinin kümelenmesine ilişkin ağaç grafiği

Benzer işlemler Ana Bileşenler analizi ile elemine edilen değişkenlere uygulandığında, fıstık tiplerinin gerçeğe yakın bir şekilde sınıflandıkları Şekil 4.17'deki ağaç grafiğinde görülmektedir. Özellikle Öklid uzaklığının karesinden (Öklid kare) elde edilen ağaç grafiği vasıtasıyla erkenci ve geççi fıstık tipleri aynı gruplar altında toplanmıştır, orta geççi fıstık tipleri ile birlikte kümelenmiştir. Sadece 12 No'lu orta geççi bir fıstık tipi erkenci tiplerle birlikte sınıflanmıştır. Bunun sebebini araştırdığımızda ise, erkenci fıstık tipleri ile geççi fıstık tiplerinin bir melezi olduğu tahmin edilen 12 No'lu fıstık tipinin bazı özellikler bakımından erkenci fıstık tiplerine daha çok benzerlik göstermesinden kaynaklanmaktadır.

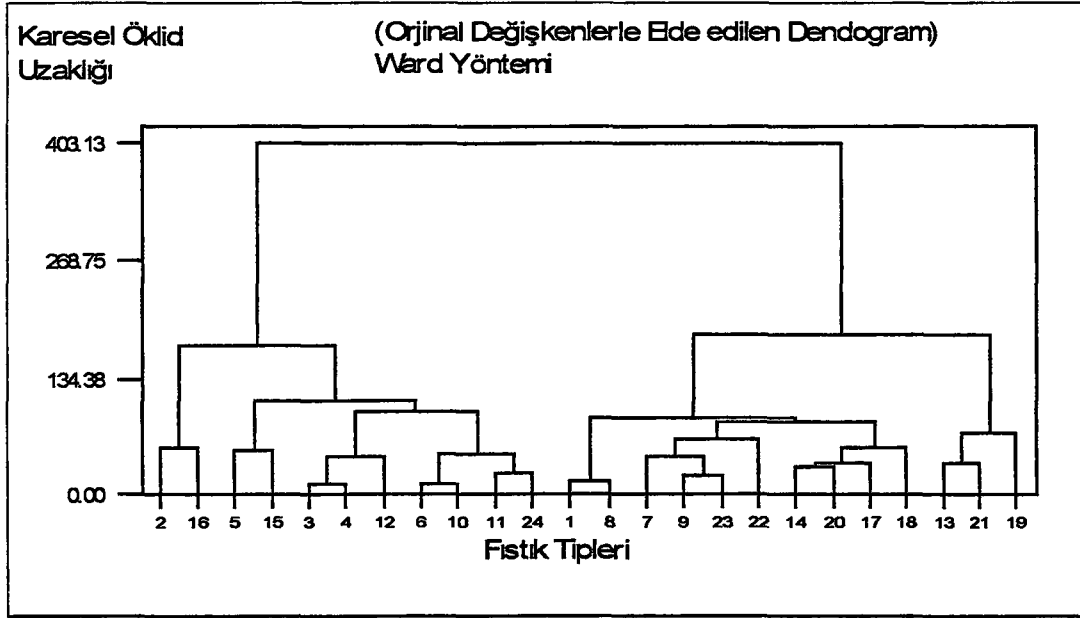
Geççi ve orta geççi fıstık tiplerinin birlikte oluşturmuş oldukları ikinci kümenin alt gruplarını incelediğimizde; 13, 18, 19, 21 No'lu geççi fıstık tiplerinin kendi aralarında ayrı bir küme oluşturdukları 4.18 No'lu ağaç grafiğinden görülmek mümkündür. Geri kalan 7, 14, 17, 20 No'lu geççi fıstık tipleri ise orta geççi fıstık tipleri ile ikinci bir alt küme oluşturmuştur.



Şekil 4.17. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward yöntemi ile elde edilen ağaç grafiği

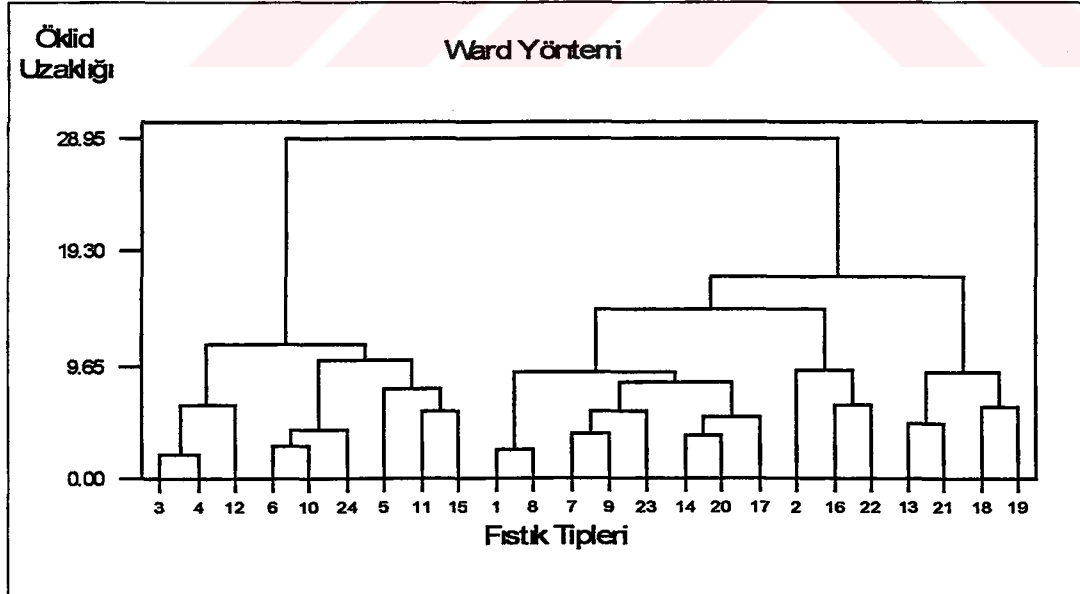
Ward yönteminin en iyi küme yapısını vermesinden dolayı orijinal değişkenlere ilişkin ağaç grafiği Şekil 4.18’de verilmiştir. Böylece bu yöntemi kullanmak suretiyle hangi değişken yapısının (boyut indirgeme yöntemi ile boyutu indirgenmiş değişkenlerin) daha iyi sonuç verdiği kıyaslanabilecektir.

Şekil 4.18’da ki ağaç grafiğini incelediğimizde; Orijinal değişkenlerle elde edilen kümelerde 2 ile 16 No’lu orta geççi fıstık tipleri erkenci tiplerle kümelenmiştir. Erkenci ve geççi tipler ise kendi aralarında doğru olarak kümelenmiştir. Orijinal değişkenlere ilişkin en yakın küme yapıları Ana bileşenler analizinden elde edilmiştir. En iyi kümeleme, Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlerden elde edildiğinden test istatistikleri ve ilgili yorumlamalar Ward yöntemi üzerinde yapılmıştır.



Şekil 4.18. Orjinal değişkenlerle elde edilen Ward yöntemine ilişkin ağaç grafiği

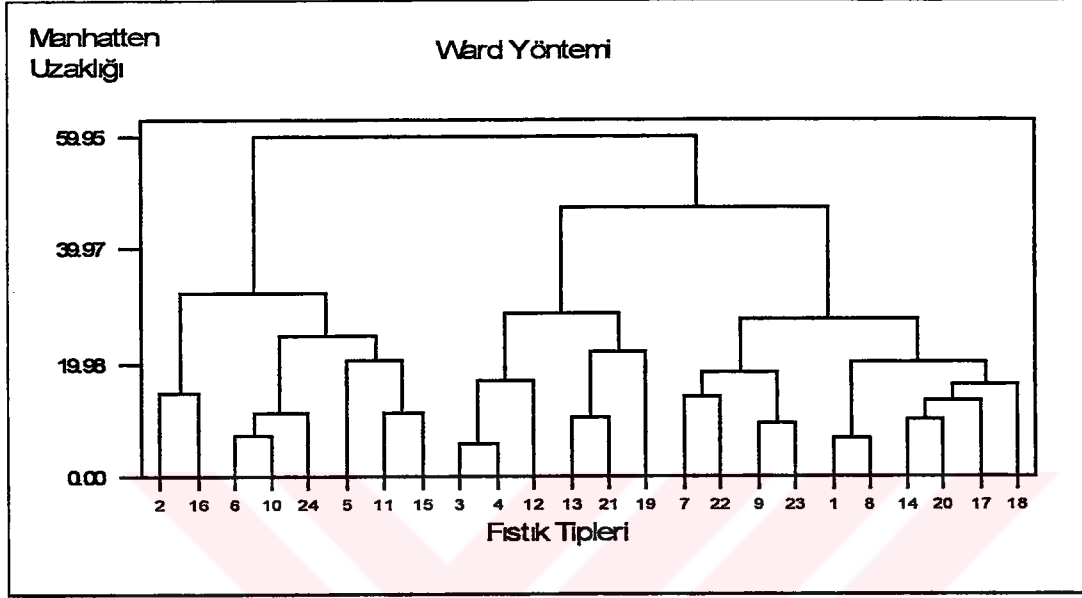
En iyi sonucu (Ana Bileşenler-Ward) yöntemi verdiğiinden, farklı uzaklık ölçüleri ile elde edilen ağaç grafikleri karşılaştırmak için; sadece bu yöneme ilişkin diğer uzaklıklarla elde edilen ağaç grafikleri aşağıdaki şekillerde sırasıyla verilmiştir.



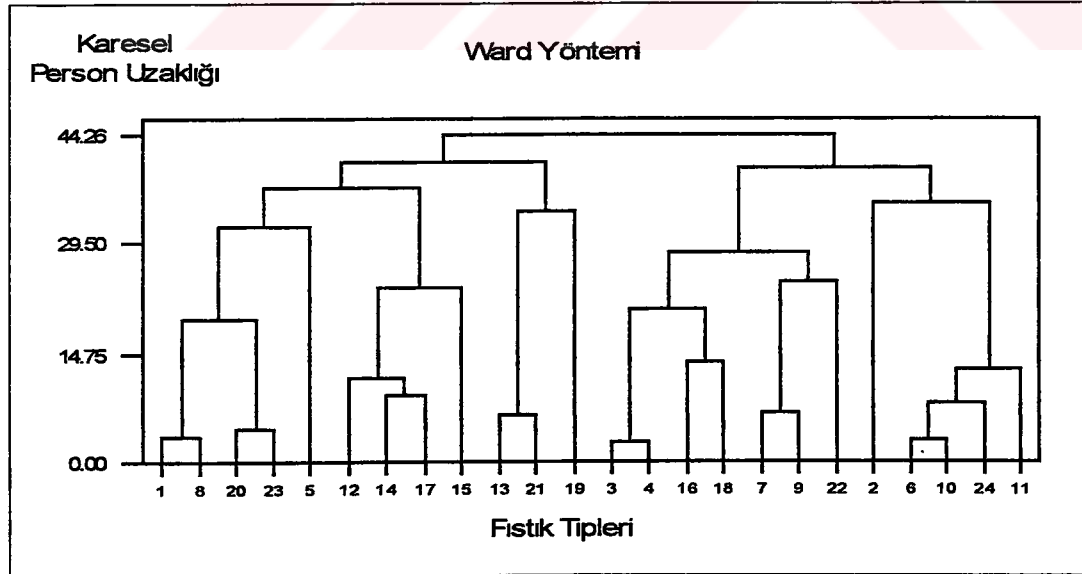
Şekil 4.19. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Öklid yöntemi ile elde edilen ağaç grafiği .

Şekil 4.19 da verilen Ward- Öklid uzaklığına ilişkin ağaç grafiği , Ward- Öklid uzaklığının karesi ile aynı sonuçları vermiştir.

Şekil 4.20’de verilen Ward-Manhattan yöntemi ise erkenci fıstık tipleri ile geçici fıstık tiplerini karışık olarak kümelemiştir.



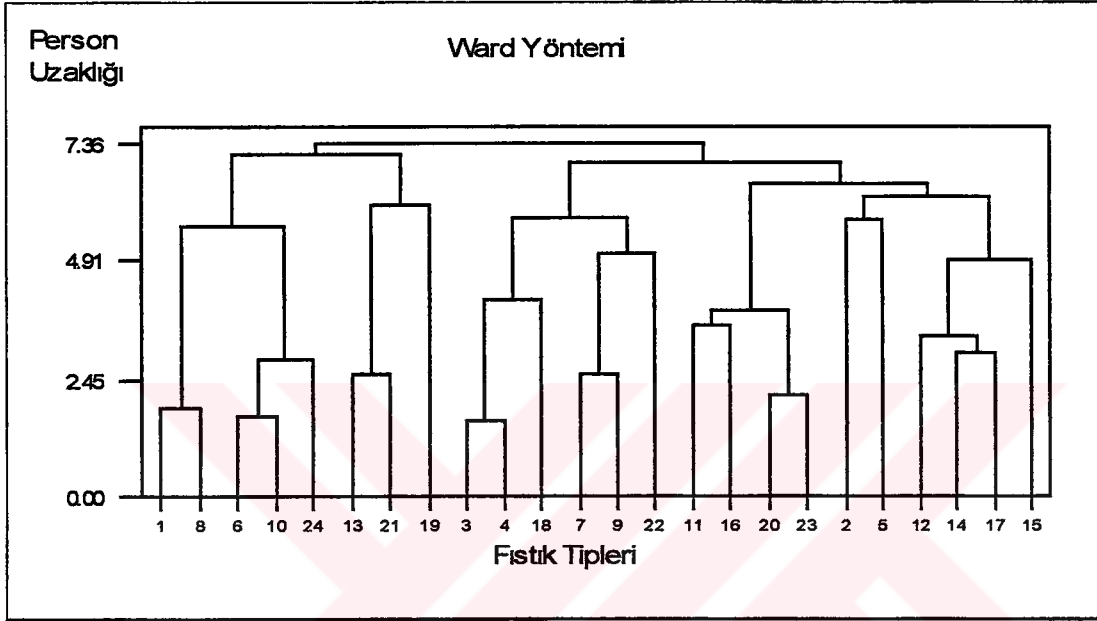
Şekil 4.20. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Manhattan yöntemi ile elde edilen ağaç grafiği .



Şekil 4.21. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Pearson Kare yöntemi ile elde edilen ağaç grafiği .

Şekil 4.21'deki Ward- Person uzaklığının karesine ilişkin ağaç grafiğinde erkenci ve geççi fıstık tipleri hatalı kümelenmiştir.

Temel iki küme olarak ele alınacak olursa, 7 ve 18 No'lu geççi fıstık tipleri, erkenci fıstık tiplerinin arasında yer almaması; bununla birlikte 5 ve 15 No'lu erkenci fıstık tiplerinin geççi fıstık tipleri arasında yer almaması gerekirdi.

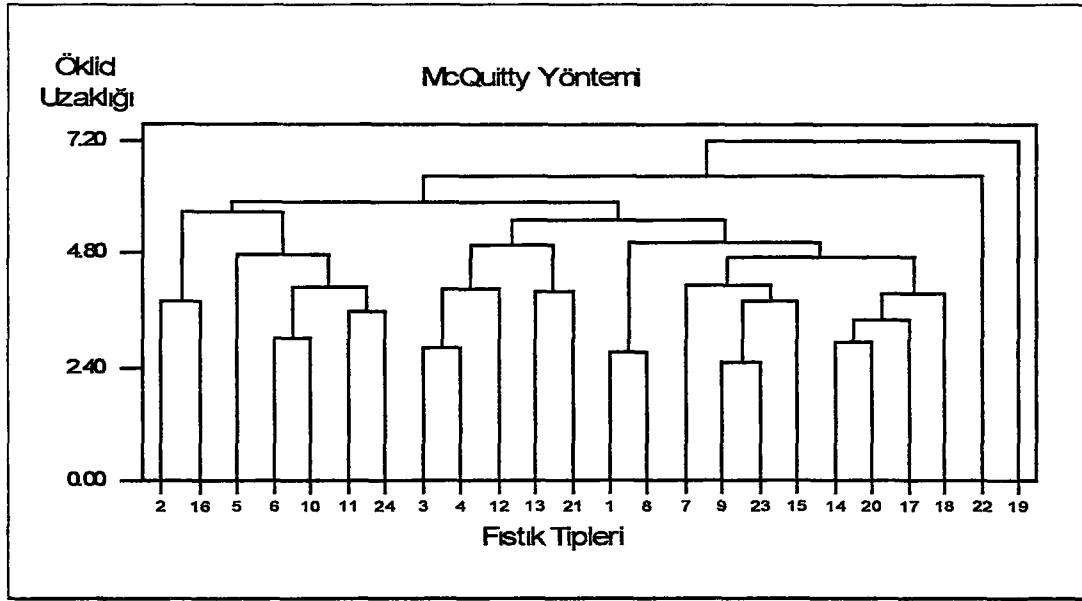


Şekil 4.22. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin Ward-Pearson yöntemi ile elde edilen ağaç grafiği .

Şekil 4.22'de verilen Ward-Person uzaklığına ilişkin ağaç grafiğinde fıstık tipleri hatalı bir şekilde kümelenmiştir. Erkenci fıstık tipleri ile geççi fıstık tipleri burada da karışık olarak kümelenmişlerdir.

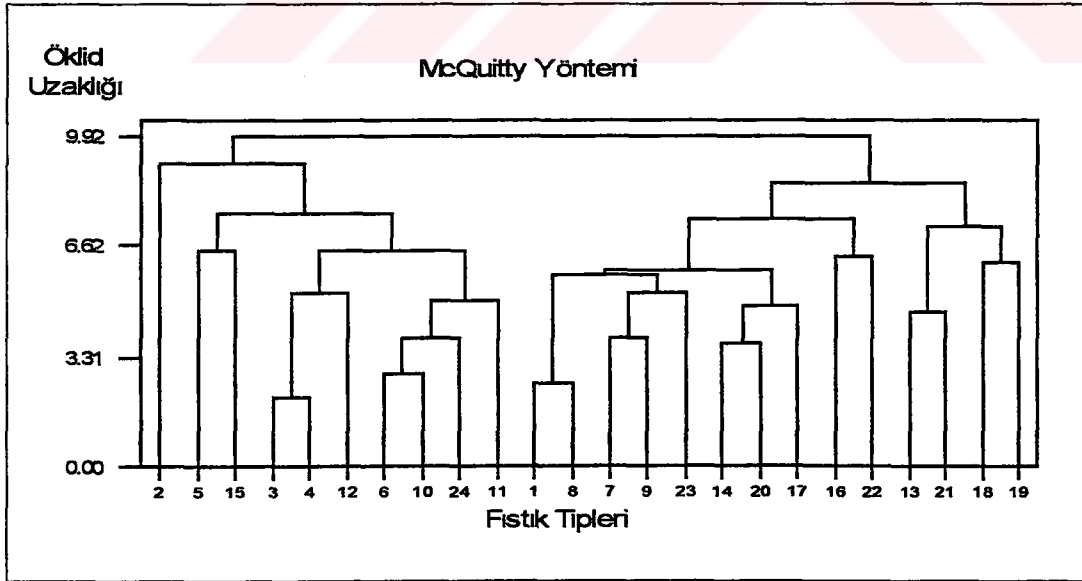
4.4.8. McQuitty Bağlantı Kümeleme Yöntemi

3.50 No'lu eşitliğinden yararlanarak Şekil 4.23'de elde edilen ağaç grafiği vasıtasıyla da fıstık tipleri anlamlı gruplara ayırlanamamıştır. 19 No'lu fıstık tipi ile 22 No'lu fıstık tipleri tek başına birer küme oluştururken, diğer fıstık tipleri de ayrı bir üçüncü küme oluşturmuşlardır.



Şekil 4.23. Korelasyon analizi ile elemine edilen değişkenlere ilişkin McQuitty yöntemi ile elde edilen ağaç grafiği

Benzer şekilde aynı yöntem kullanılarak Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlerle fıstık tipleri sınıflandırıldığında Şekil 4.24'deki ağaç grafiği elde edilmektedir. Bu yöntemde de fazla hatalı bir kümeleme olmamakla birlikte, 2 ve 12 No'lu orta geççi fıstık tipleri erkenci tiplerle kümelenmiştir.



Şekil 4.24. Ana Bileşenler analizi ile boyutu indirgenen değişkenlere ilişkin McQuitty yöntemi ile elde edilen ağaç grafiği

4.5. Aşamalı Olmayan K-Ortalama Yöntemi İle Fıstık Tiplerinin Kümelenmesi

Hem k-Ortalama yöntemi, korelasyon analizi ile elemine edilmiş, hem de Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlere ilişkin fıstık tipleri kümelenecek en anlamlı küme yapısı tespit edilmiştir.

K-Ortalama yönteminde en önemli sorun küme sayısının belirlenmesidir. Özdamar (59) küme sayısını belirlemek için;

- a- Aşamalı kümeleme yöntemlerinden elde edilen ağaç grafiklerini inceleyerek karar vermek,
- b- Olasılıklı olarak başlangıç noktalarını belirlemek,
- c- $k=2, 3, 4, \dots$ biçiminde ardışık olarak küme sayısını bir artırarak oluşan kümelemede birimlerin hangi kümeye ait olduğuna ilişkin küme kodlarını (grup kodları) kullanarak yeni veri yapısına Diskriminant analizi uygulamak ve Wilk's Lambda değerine sahip olan küme sayısını uygun parçalama olarak kabul etmek,
- d- İlk k birimin p değişkenler ortalamasını başlangıç (seed) ortalama vektörü olarak ele alıp birimleri bu kümelere atamak,

yaklaşımlarından birisinin seçilebileceğini bildirmiştir. Bu çalışmada, k-Ortalama yöntemi için, yukarıda verilen (c) maddesi esas alınarak çözüme gidilmeye çalışılmıştır.

4.5.1. Korelasyon analizi ile elemine edilmiş değişkenlerin K-Ortalama yöntemi ile sınıflandırılması

İlgili küme sayıları 3.51 No'lu eşitlikten yararlanılarak önceden araştırmacı tarafından belirlenmektedir. Bu amaçla, tüm ihtimaller denenerek fıstık tiplerinin 2 ile 5 grup arasındaki kümelere ayrıldığında hangi kümelere yer alacakları Çizelge 4.15'de verilmiştir. Ancak oluşturulan kümeler incelendiği zaman erkenci ve geççi fıstık tiplerinin birbiri ile karışık halde kümelendiğini görmekteyiz.

Çizelge 4.15. Korelasyon analizi ile elemine edilen değişkenler vasıtasıyla fıstık tiplerinin k-Ortalama yöntemine göre kümelere atanması

Küme Sayıları	Küme Numarası	Fıstık Tipleri (Küme Elemanları)
2	1	(1, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 17, 18, 19, 20, 23)
	2	(2, 10, 11, 16, 21, 22, 24)
3	1	(1, 7, 8, 9, 14, 15, 17, 20, 23)
	2	(2, 10, 16)
	3	(3, 4, 5, 6, 11, 12, 13, 18, 19, 21, 22, 24)
4	1	(1, 8, 14, 17, 23)
	2	(2, 16)
	3	(3, 5, 6, 10, 11, 24)
	4	(4, 7, 9, 12, 13, 15, 18, 19, 20, 21, 22)
5	1	(1, 8, 14, 17, 23)
	2	(2, 16)
	3	(3, 24)
	4	(4, 7, 9, 11, 12, 13, 15, 18, 19, 20, 21, 22)
	5	(5, 6, 10)

4.5.2. Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlere ilişkin K-Ortalama yöntemi sonuçları

Ana Bileşenler analizi ile boyutu indirgenmiş olan değişkenlere ilişkin k-Ortalama yöntemi 3.51 No'lu eşitlikten yararlanılarak uygulanıldığında Çizelge 4.16'deki sonuçlar elde edilmektedir. Burada oluşan kümeleri incelediğimizde, sonuçların korelasyon analizi ile elemine edilen değişkenlerle elde edilen kümelere daha isabetli olduğunu görebilmekteyiz. Ancak burada elde edilen kümelere de erkenci fıstık tiplerinin arasında orta geççi bazı fıstık tiplerini görmekteyiz. Örneğin, fıstık tiplerini iki küme ayırdığımızda orta geççi olan 2 ve 16 No'lu fıstık tiplerinin erkenci fıstık tipleri ile birlikte sınıflandırıldığını görmekteyiz.

Fıstık tiplerini üç kümeye ayırdığımızda 19 No'lu geççi fıstık tipinin erkenci fıstık tipleri ile birlikte kümelendiğini görülmektedir. Buna karşılık Çizelge 4.16'de, 15 No'lu erkenci fıstık tipinin erkenci fıstık tiplerinin arasından çıkıp, 2 ve 16 No'lu orta erkenci fıstık tipleri ile ayrı bir küme oluşturduğu görülmektedir.

Diğer kümeleme seviyelerinde daha anlamsız kümelerin oluştuğu yine aynı Çizelgeden izlenebilmektedir.

Çizelge 4.16. Ana Bileşenler Analizinden Yaralanarak k-Ortalama Yöntemine Göre Fıstık Tiplerinin Kümelere Atanması

Küme Sayıları	Küme Numarası	Fıstık Tipleri (Küme Elemanları)
2	1	(1, 7, 8, 9, 13, 14, 17, 18, 19, 20, 21, 22, 23)
	2	(2, 3, 4, 5, 6, 10, 11, 12, 15, 16, 24)
3	1	(1, 7, 8, 9, 13, 14, 17, 18, 20, 21, 22, 23)
	2	(2, 15, 16)
	3	(3, 4, 5, 6, 10, 11, 12, 19, 24)
4	1	(1, 7, 8, 9, 13, 14, 17, 18, 20, 23)
	2	(2, 16)
	3	(3, 5, 21, 22, 24)
	4	(4, 6, 10, 11, 12, 15, 19)
5	1	(1, 7, 8, 9, 14, 17, 20, 23)
	2	(2, 16)
	3	(13, 18, 19, 21, 22)
	4	(3, 4, 6, 7, 10, 11, 12, 24)
	5	(5, 15)

4.6. Stepwise Diskriminant Analizi Yöntemi İle Değişkenlerin Elemine Edilerek Kümeleneşmesi

Değişkenlerde boyut indirgemenin diğeri bir yolu ise Diskriminant analizinin Stepwise tekniğı ile en etkili değışkenlerin seçilmesidir. Bunun uygulaması řu şekilde yapılmıřtır. En anlamlı sonuçlar verdiğı tespit edilen Ward, Tam Bağılantı ve

k-Ortalama yöntemlerine göre fıstık tipleri, orijinal değişkenlerin yardımıyla 2, 3, 4, 5 küme oluşturacak şekilde sınıflandırılmıştır. Elde edilen sonuçlar, Stepwise Diskriminant analizine tabi tutularak sınıflandırmada en etkili olan değişkenler tespit edilmiştir. Sonuçta her küme için, birimlerin atanmasında rol oynayan etkin değişkenlerin farklı oldukları belirlenmiştir. Sözü edilen etkili değişkenler Çizelge 4.17’de verilmiştir.

Çizelge 4.17. Stepwise Diskriminant analizi ile seçilmiş olan etkili değişkenler ve önem seviyeleri

Küme Sayısı	Kümeleme Yöntemleri								
	Ward Yöntemi			Tam Bağlantı Yöntemi			k-ortalama Yöntemi		
	Değişken	Wilk's L.	P	Değişken	Wilk's L.	P	Değişken	Wilk's L.	P
2	C27	,35485	,0000	C27	,35485	,0000	C27	,35485	,0000
	C21	,18586	,0000	C21	,18586	,0000	C21	,18586	,0000
	C13	,11418	,0000	C13	,11418	,0000	C13	,11418	,0000
	C32	,08496	,0000	C32	,08496	,0000	C32	,08496	,0000
	C14	,06175	,0000	C14	,06175	,0000	C14	,06175	,0000
	C28	,03105	,0000	C28	,03105	,0000	C25	,03105	,0000
3	C27	,32926	,0000	C23	,17553	,0000	C26	,20724	,0000
	C13	,12775	,0000	C20	,10026	,0000	C06	,05486	,0000
	C17	,05761	,0000	C11	,04479	,0000	C11	,03684	,0000
	C18	,03336	,0000				C18	,02346	,0000
	C31	,01959	,0000				C31	,01354	,0000
	C15	,01092	,0000						
	C24	,00685	,0000						
4	C23	,11594	,0000	C23	,15567	,0000	C27	,19592	,0000
	C13	,04578	,0000	C13	,04028	,0000	C13	,04341	,0000
	C20	,01974	,0000	C20	,01776	,0000	C06	,01728	,0000
	C24	,01114	,0000	C11	,00667	,0000	C34	,00687	,0000
	C21	,00547	,0000	C24	,00306	,0000	C18	,00315	,0000
	C19	,00205	,0000	C19	,00165	,0000	C09	,00163	,0000
	C35	,00109	,0000	C17	,00061	,0000			
				C35	,00030	,0000			
5	C23	,11558	,0000	C11	,08036	,0000	C24	,15194	,0000
	C11	,03633	,0000	C24	,02228	,0000	C13	,02808	,0000
	C09	,00907	,0000	C06	,00557	,0000	C06	,00865	,0000
	C29	,00266	,0000	C18	,00174	,0000	C17	,00205	,0000
	C31	,00093	,0000	C22	,00056	,0000	C19	,00057	,0000
				C13	,00018	,0000	C38	,00026	,0000
				C20	7,40E-05	,0000	C34	,00012	,0000
				C19	3,20E-05	,0000	C22	4,69E-05	,0000
				C17	1,03E-05	,0000	C01	1,79E-05	,0000

Çizelge 4.17'yi incelediğimizde fıstık tiplerinin belirlenmesinde 38 adet değişken yerine genel olarak en etkili olan;

{C01, C06, C09, C11, C13, C14, C17, C18, C19, C20, C21,
C22, C23, C24, C25, C26, C27, C28, C31, C32, C34, C35, C38}

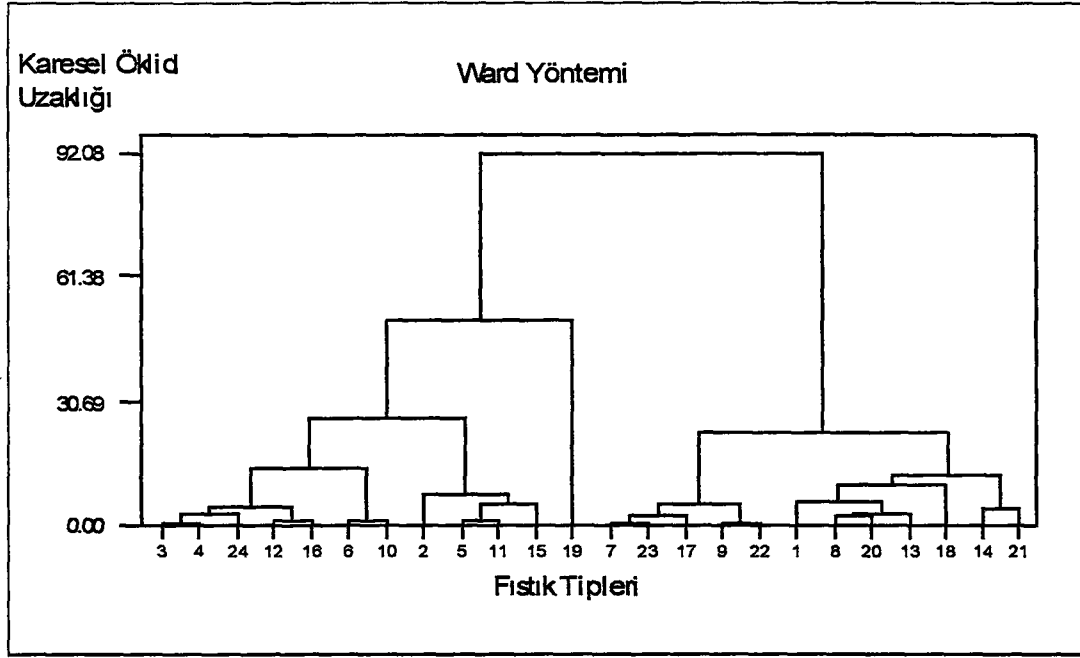
değişkenlere ilişkin ölçümlerin alınması yeterli olacaktır.

Diskriminant analizi ile boyutu indirgenmiş değişkenlerin yardımıyla kümeleme analizi yapılmış olup, Ward yöntemine göre elde edilen sonuçlar Çizelge 4.18.'de verilmiştir. Burada her bir kümeyi meydana getiren değişkenler farklı olduklarından, bunlara ilişkin ağaç grafikleri ayrı ayrı Şekil 4.25, Şekil 4.26, Şekil 4.27, Şekil 4.28'de verilmiştir.

Çizelge 4.18. Stepwise Diskriminant analizinden yaralanarak elemine edilen değişkenlerle Ward yöntemine göre fıstık tiplerinin kümelere atanması

Küme Sayıları	Küme Numarası	Fıstık Tipleri (Küme Elemanları)
2	1	(1, 7, 8, 9, 13, 14, 17, 18, 20, 21, 22, 23)
	2	(2, 3, 4, 5, 6, 10, 11, 12, 15, 16, 19, 24)
3	1	(1, 7, 8, 9, 14, 15, 17, 22, 23)
	2	(2, 3, 4, 5, 6, 10, 11, 12, 16, 24)
	3	(13, 18, 19, 20, 21)
4	1	(1, 2, 7, 8, 9, 16, 17, 22, 23)
	2	(3, 4, 11, 12, 15)
	3	(5, 6, 10, 24)
	4	(13, 14, 18, 19, 20, 21)
5	1	(1, 7, 8, 9, 14, 17, 18, 20, 22, 23)
	2	(2, 16)
	3	(3, 4, 6, 10, 11, 24)
	4	(5, 15)
	5	(12, 13, 19, 21)

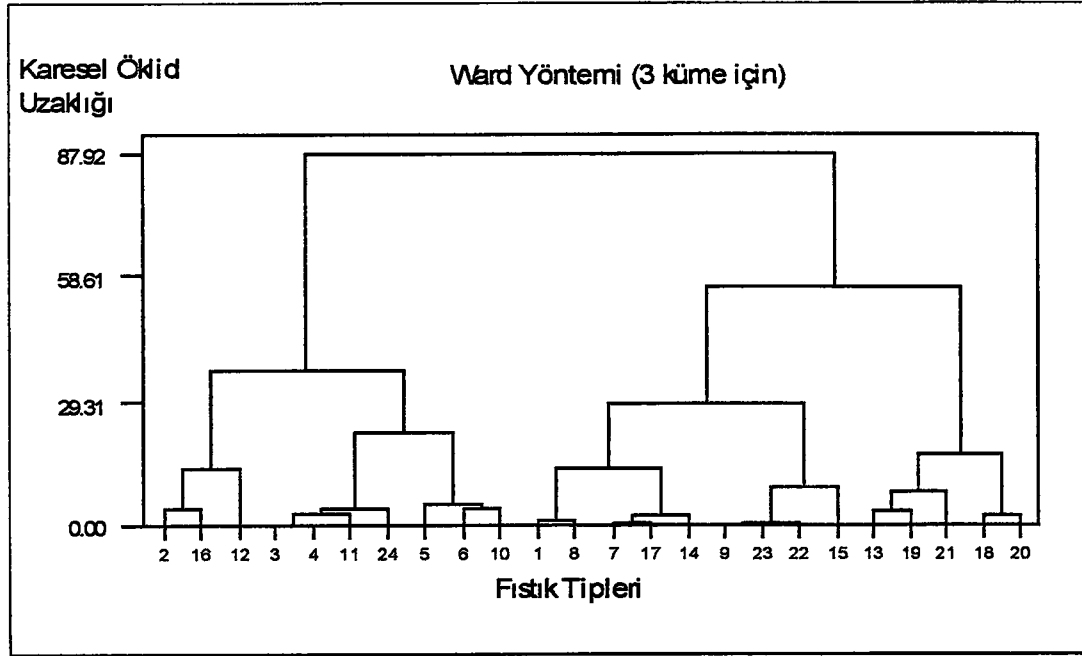
Çizelge 4.18'de özetlemiş olan sonuçlar aşağıdaki şekillerde ağaç grafikleri vasıtasıyla sırasıyla ayrı ayrı verilmiştir.



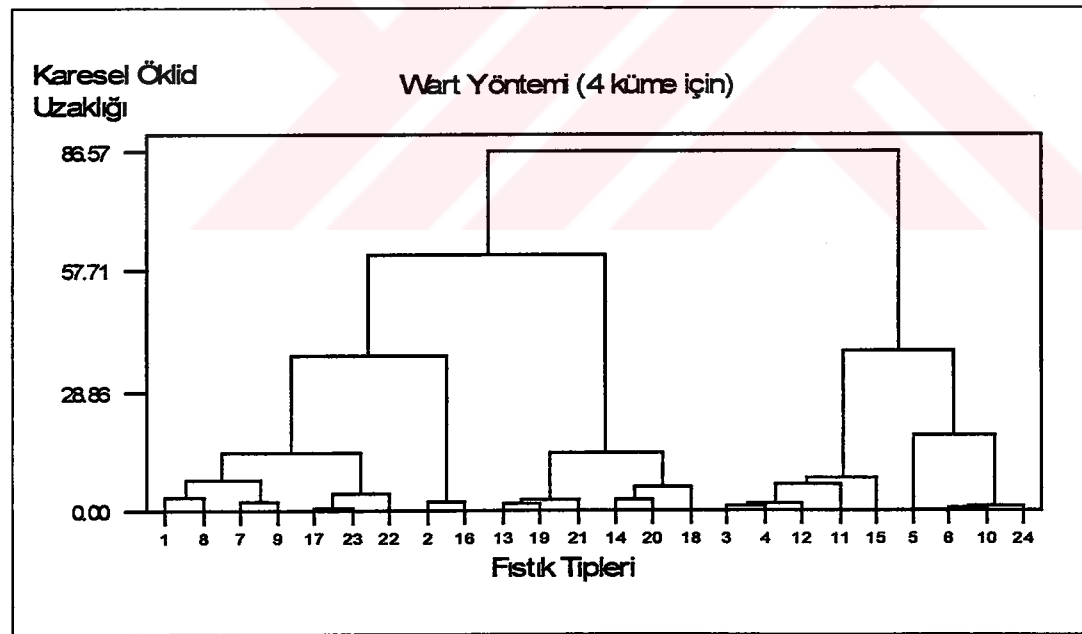
Şekil 4.25. Küme sayısı 2 olması durumunda Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlere ilişkin Ward yöntemi ile elde edilen ağaç grafiği .

Şekil 4.25'deki ağaç grafiğinde, birimler iki küme oluşturacak şekilde Stepwise tekniği ile elemine edilen değişkenler kullanılmıştır. Fıstık tiplerini geççi ve erkenci olmak üzere iki grup olarak kabul edecek olursak; bütün erkenci fıstık tipleri kendi aralarında problemsiz bir şekilde kümelendiğini, ancak 19 No'lu geççi fıstık tipi erkenci tiplerle hatalı olarak kümelendiği gözlenmektedir.

Şekil 4.26'de ise küme sayısı üç olarak ele alındığında, 15 No'lu erkenci fıstık tipi geççi tiplerle birlikte kümelendiği görülmektedir. Diğer fıstık tipleri ise hemen hemen doğru kümelere atanmıştır. Sadece alt kümeler incelendiğinde, orta geççi bazı fıstık tiplerinin kendi aralarında kümelenerken erkenci ve geççi fıstık tiplerinin arasında bir alt grup oluşturdukları görülmektedir. Örneğin 2, 16 ve 12 No'lu orta geççi tipler kendi aralarında kümelendikleri halde erkenci tiplere yakınlık göstermektedir. 13 ,19, 21, 18 ve 20 No'lu geççi tipler kendi aralarında doğru olarak kümelendirken; 7, 17 ve 14 No'lu geççi tipler kendi aralarında ayrı bir küme oluşturarak orta geççi tiplere yakınlık göstermiştir.

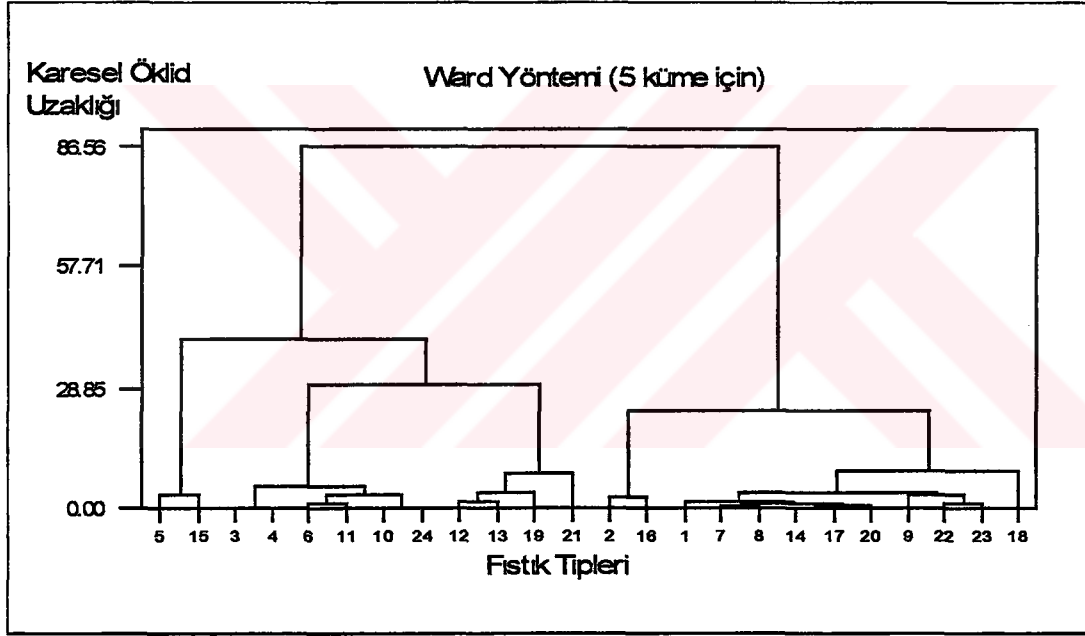


Şekil 4.26. Küme sayısı 3 olması durumunda Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlere ilişkin Ward yöntemi ile elde edilen ağaç grafiği



Şekil 4.27. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 4 kümeye ayırdıklarında Ward yöntemi ile elde edilen ağaç grafiği

Şekil 4.27'deki ağaç grafiğini incelediğimizde erkenci fıstık tiplerinin tamamının doğru olarak kümelendiğini görmekteyiz. Ancak 19 No'lu orta geççi fıstık tipinin bu kümenin arasında yer aldığı ve bunun sebebinin de sözü edilen fıstık tipinin Stepwise analizi ile seçilen değişkenler bakımından erkenci tiplere benzerlik göstermesinden kaynaklandığı söylenilebilir. Geççi fıstık tiplerinden 7 ve 17 hariç (13, 19, 21, 14, 20, 18) diğerlerinin kendi aralarında ayrı bir küme oluşturdukları tespit edilmiştir. 7 ve 17 No'lu fıstık tiplerinin orta geççi fıstık tipleri ile kümelенmesinin sebebi, bu fıstık tiplerinin fenotipik olarak bazı orta geççi fıstık tiplerine (C13) Uç Yaprak Tabanı, (C35) Salkımdaki Çiçek Tozu Sayısı ve (C21) Karagöz Genişliği bakımından benzerlik göstermesi olarak kabul edilebilir.



Şekil 4.28. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 5 kümeye ayrıldıklarında Ward yöntemi ile elde edilen ağaç grafiği

Fıstık tipleri Şekil 4.28.'deki ağaç grafiğinde görüldüğü gibi 5 kümeye ayrıldığında; erkenci tipler olan 5 ile 15 No'lu fıstık tiplerinin bir küme; 3, 4, 6, 11, 10, 24 No'lu fıstık tiplerinin ise kendi aralarında ayrı bir küme oluşturdukları görülmektedir. Ancak üçüncü bir küme oluşturan orta geççi 12 No'lu fıstık tipi ile

13, 19, 21, 2 ve 16 No'lu geççi fıstık tiplerinin kendi aralarında oluşturmuş oldukları küme, üçüncü aşamada diğer orta geççi ve geççi tiplerden bağımsız olarak erkenci fıstık tiplerinin oluşturdukları kümeye yakınlık göstermişlerdir.

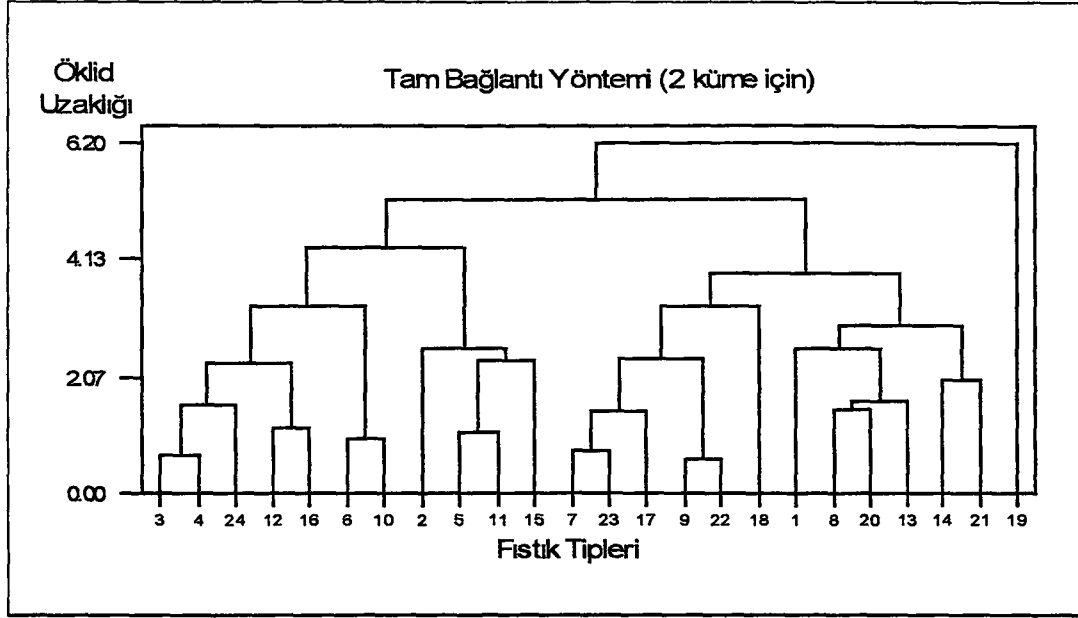
Diskriminant analizi ile boyutu indirgenmiş değişkenlerle kümeleme analizi yapılarak, Tam Bağlantı yöntemine göre elde edilen sonuçlar Çizelge 4.19'de verilmiştir.

Çizelge 4.19. Stepwise Diskriminant analizinden yaralanarak elemine edilen değişkenlerle Tam Bağlantı yöntemine göre fıstık tiplerinin kümelere atanması

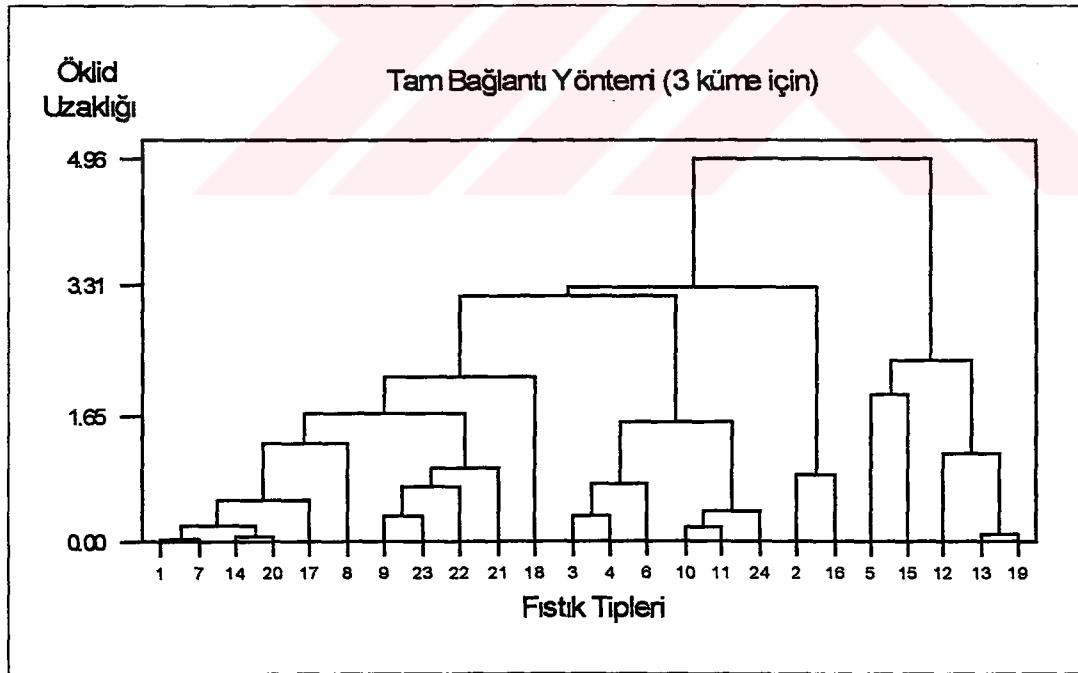
Küme Sayıları	Küme Numarası	Fıstık Tipleri (Küme Elemanları)
2	1	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24)
	2	(19)
3	1	(1, 3, 4, 6, 7, 8, 9, 10, 11, 14, 17, 18, 20, 21, 22, 23, 24)
	2	(2, 16)
	3	(5, 12, 13, 15, 19)
4	1	(1, 7, 8, 9, 17, 22, 23)
	2	(2, 3, 4, 6, 10, 11, 12, 15, 16, 24)
	3	(5)
	4	(13, 14, 18, 19, 20, 21)
5	1	(1, 8)
	2	(2, 16)
	3	(3, 4, 5, 6, 10, 11, 12, 15, 24)
	4	(7, 9, 14, 17, 20, 22, 23)
	5	(13, 18, 19, 21)

Tam Bağlantı yöntemine göre elde edilen ağaç grafikleri sırasıyla Şekil 4.29, Şekil 4.30, Şekil 4.31, Şekil 4.32'de verilmiştir.

Fıstık tipleri Tam Bağlantı yöntemine göre iki kümeye ayrıldığında Şekil 4.29'deki ağaç grafiği elde edilmektedir. Ward yönteminde erkenci fıstık tipleri ile birlikte kümelenen 19 No'lu geççi fıstık tipi burada tek başına bir küme oluşturmuştur. Diğer erkenci ve geççi fıstık tipleri ise kendi aralarında doğru olarak kümelenmişlerdir. Orta geççi fıstık tipleri ise, erkenci ve geççi fıstık tipleri arasında dağılım göstermiştir.

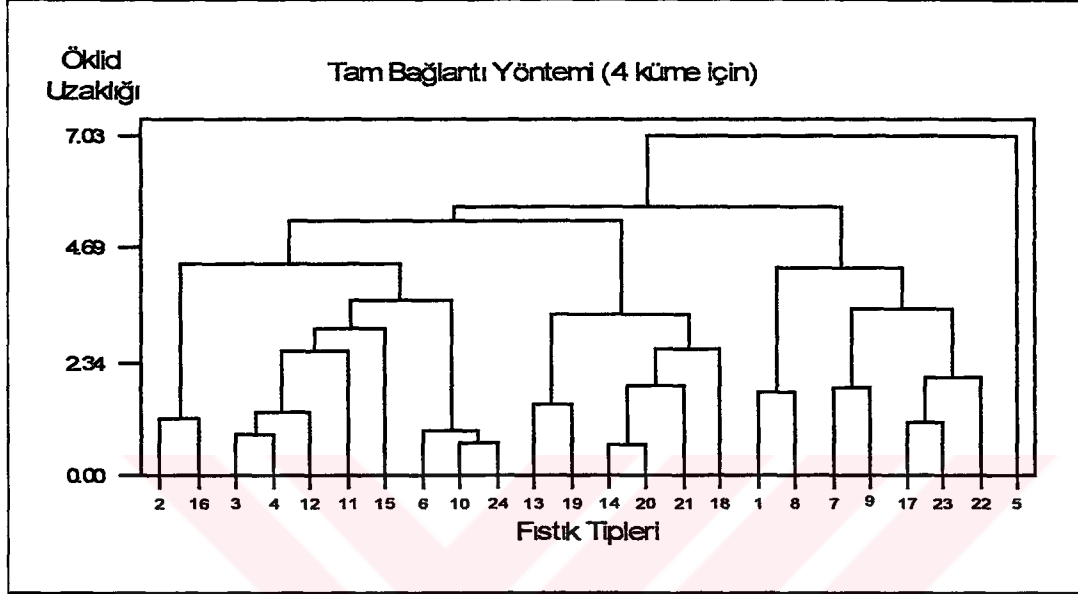


Şekil 4.29. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 2 kümeye ayrıldıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği



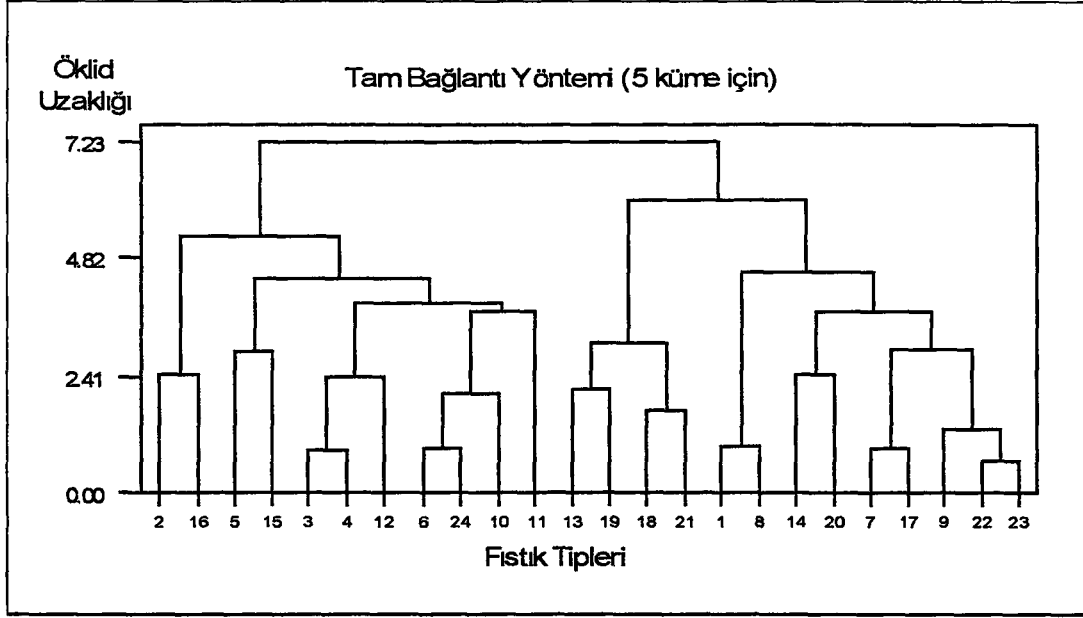
Şekil 4.30. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 3 kümeye ayrıldıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği

Şekil 4.30'deki ağaç grafiğinde birimler, üç küme esas alınarak, değişkenler elemine edilip kümeleme yapılmıştır. Ağaç grafiği üç kümeye ayrılarak incelendiğinde fıstık tiplerinin karışık olarak birbirleri ile kümelendikleri görülmektedir.



Şekil 4.31. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 4 kümeye ayrıldıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği

Şekil 4.31 incelendiğinde 5 No'lu fıstık tipinin tek başına kümelendiği, geççi fıstık tiplerinden 13, 19, 14, 20, 21, 18 No'lu tiplerin ayrı bir küme oluşturdukları ve erkenci tiplerin tamamının ise bazı orta geççi fıstık tipleri ile birlikte kümelendiği gözlenmektedir. Son olarak herhangi bir kümeye dahil olmayan bazı geççi fıstık tiplerle, orta geççi fıstık tipleri kendi aralarında ayrı bir küme oluşturmuşlardır.



Şekil 4.32. Stepwise Diskriminant analizi ile boyutu indirgenen değişkenlerle fıstık tipleri 5 kümeye ayırdıklarında Tam Bağlantı yöntemi ile elde edilen ağaç grafiği

Şekil 4.32 incelendiğinde erkenci ve geççi fıstık tiplerinin başarılı bir şekilde iki ayrı grupta kümelendikleri görülmektedir. Orta geççi fıstık tipleri ise bu iki grup arasında dağılım göstermiştir.

Çizelge 4.20 incelendiğinde, fıstık tipleri iki kümeye ayrıldığında Ward yöntemine benzer sonuçlar elde edilmiştir. Yani erkenci ve geççi fıstık tipleri ayrı ayrı kümelenebilir olup orta geççi fıstık tipleri ise bu iki küme arasında dağılım göstermiştir.

Fıstık tipleri üç kümeye ayrıldığında ise geççi tipler olan 13, 18, 19 No'lu fıstık tipleri erkenci fıstık tipleri ile birlikte hatalı olarak kümelenebilirlerdir.

Fıstık tipleri dört kümeye ayrıldığında ise (3, 4, 6, 10, 11, 24) No'lu erkenci fıstık tiplerinin kendi aralarında bir küme oluşturdukları ve 5 ile 15 No'lu erkenci fıstık tiplerinin ise orta geççi fıstık tipleri ile ayrı bir küme oluşturdukları gözlenmiştir. Geççi fıstık tipleri de aynı şekilde kendilerine en benzer olan bazı orta geççi fıstık tipleri ile iki ayrı küme oluşturmuştur.

Birimler beş kümeye ayrıldığında oluşan alt kümelerde geççi ve erkenci fıstık tiplerinin birbirleri ile karışık olarak kümelendikleri gözlenmektedir.

Çizelge 4.20. Stepwise Diskriminant analizinden yaralanarak elemine edilen değişkenlerle k-Ortalama yöntemine göre fıstık tiplerinin kümelere atanması

Küme Sayıları	Küme Numarası	Fıstık Tipleri (Küme Elemanları)
2	1	(1, 7, 8, 9, 13, 14, 17, 18, 19, 20, 21, 22, 23)
	2	(2, 3, 4, 5, 6, 10, 11, 12, 15, 16, 24)
3	1	(1, 7, 8, 9, 14, 17, 20, 21, 22, 23)
	2	(2, 5, 15, 16)
	3	(3, 4, 6, 10, 11, 12, 13, 18, 19, 24)
4	1	(1, 7, 8, 9, 14, 17, 20, 22, 23)
	2	(2, 5, 15, 16)
	3	(12, 13, 18, 19, 21)
	4	(3, 4, 6, 10, 11, 24)
5	1	(1, 7, 8, 17)
	2	(2, 9, 15, 16, 20, 22, 23)
	3	(3, 13)
	4	(4, 12, 14, 18, 19, 21)
	5	(5, 6, 10, 11, 24)

4.7. Wilk's Lambda ve Hotelling Lawley İz Ölçütleri İle Farklı Küme seviyelerinin Test edilmesi

Her bir yöntemle ilişkin elde edilen ağaç grafiklerinde, Ana Bileşenler analizi ile elde edilen kümelerin daha anlamlı sonuçlar verdiği tespit edilmişti. Hatta Ana Bileşenler analizi ile elde edilen sonuçlar orijinal verilerin kümeleneşine daha yakın sonuçlar verdiği gözlenmiştir. Bu fark, sırasıyla; Çizelge 4.21, Çizelge 4.22 ve Çizelge 4.23 ile Çizelge 4.24, Çizelge 4.25 ve Çizelge 4.26'de verilen Wilk's Lambda ve Hotelling Lawley İz ölçütlerine ilişkin test istatistikleri sonuçlarından da görülebilir. Ancak orijinal değişkenlere ilişkin veri matrisi ($n < p$) boyutunda

olduğundan Wilk's Lambda ve Hotelling Lawley İz ölçütlerine ilişkin sonuçlar verilememiştir.

Wilk's Lambda ve Hotelling Lawley İz ölçütlerine ilişkin sonuçlara bakarak değişken seçiminde en iyi rol oynayan yöntemi (Ana Bileşenler, Korelasyon, Stepwise) belirlememiz mümkün değildir. Çünkü değişken sayıları farklıdır. Ancak daha önceki bölümde incelediğimiz ağaç grafiklerine bağlı olarak, en anlamlı sonuçları Ana Bileşenler analizi ile elemine edilen değişkenlerle elde edildiğini söyleyebilmekteyiz. Bu nedenle değişken seçimi yöntemine bağlı olarak kümeleme yöntemleri kendi içinde karşılaştırılacaktır.

Korelasyon analizi ile elemine edilen değişkenlerle ilgili küme yapısı Wilk's Lambda test istatistiğine göre test edilmesi halinde, fıstık tipleri 2 ve 3 kümeye ayrıldığında en iyi küme yapısını Ward yönteminin verdiğini Çizelge 4.21'den görebilmekteyiz.

Çizelge 4.21. Korelasyon analizi ile elemine edilmiş değişkenlere ilişkin farklı küme sayılarının Wilk's Lambda test istatistiği ile önemlilik düzeylerinin belirlenmesi

Kümeleme Yöntemi	Küme Sayısı	Wilk's Lambda	Test İstatistiği F	Önemlilik Düzeyi P
K-Ortalama	2	0.11480	4.112	0.025
	3	0.01750	3.061	0.015
	4	0.00117	3.595	0.002
	5	0.00025	2.680	0.007
Ward Yöntemi	2	0.03661	14.035	0.000
	3	0.00470	6.338	0.000
	4	0.00014	7.845	0.000
	5	0.00001	6.918	0.000
Tam Bağlantı	2	0.06787	7.325	0.004
	3	0.00490	6.203	0.000
	4	0.00005	11.156	0.000
	5	0.00000	12.920	0.000

k-Ortalama yönteminin 2 ve 3 küme için oluşturmuş olduğu küme yapısı ise $P > 0.01$ düzeyinde önemsiz çıkmıştır. Aynı çizelgede fıstık tipleri 4 ve 5 kümeye ayrıldığında ise Tam Bağlantı yönteminin en iyi küme yapısını verdiği tespit edilmiştir.

Ana Bileşenleri esas alarak kümeleme yöntemleri kendi aralarında önemlilik seviyelerine göre değerlendirecek olursa; fıstık tipleri 2 ve 4 kümeye ayrıldığında en iyi küme yapısını Ward yönteminin verdiğini Çizelge 4.22’de görmemiz mümkündür. Fıstık tipleri 3 kümeye ayrıldığında Tam Bağlantı yönteminin en önemli küme yapısını verdiğini yine aynı cetvelden görebiliriz. Fıstık tipleri 5 kümeye ayrıldığında ise k-Ortalama yöntemi en iyi küme yapısını vermiştir.

Çizelge 4.22. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlere ilişkin farklı küme sayılarının Wilk’s Lambda test istatistiği ile önemlilik düzeylerinin belirlenmesi

Kümeleme Yöntemi	Küme Sayısı	Wilk’s Lambda Ölçütü	Test İstatistiği F	Önemlilik Düzeyi P
K-Ortalama	2	0.06038	20.231	0.000
	3	0.01975	7.339	0.000
	4	0.00272	7.121	0.000
	5	0.00009	10.695	0.000
Ward Yöntemi	2	0.02893	43.640	0.000
	3	0.00627	13.956	0.000
	4	0.00083	11.209	0.000
	5	0.00013	9.629	0.000
Tam Bağlantı	2	0.04243	29.339	0.000
	3	0.00576	14.612	0.000
	4	0.00120	9.777	0.000
	5	0.00022	8.220	0.000

Stepwise diskriminant analizi ile elemine edilmiş deęişkenlerin sayısı her yöntem için farklı olduğundan kümeleme yöntemlerini kendi aralarında karşılaştırmamız doğru deęildir. Ancak en anlamlı sonuçları veren Tam Bağlantı, k-Ortalama ve Ward yöntemlerine göre elde edilen kümeler Çizelge 4.23’de verilen sonuçlara göre $P < 0.01$ düzeyinde önemli çıktığı gözlenmektedir.

Yöntemlerin karşılaştırılmasında ise sadece ağaç grafikleri incelenerek, hangi yöntemin fıstık tiplerini en iyi şekilde kümelere atadığı belirlenmiştir.

Çizelge 4.23. Stepwise Diskriminant analizi ile boyutu indirgenmiş deęişkenlere ilişkin farklı küme sayılarının Wilk’s Lambda ölçütü ile önemlilik düzeylerinin belirlenmesi

Kümeleme Yöntemi	Küme Sayısı	Wilk’s Lambda Ölçütü	Test İstatistięi F	Önemlilik Düzeyi P
K-Ortalama	2	0.031	88.41	0.000
	3	0.017	22.33	0.000
	4	0.002	21.32	0.000
	5	0.003	4.65	0.000
Ward Yöntemi	2	0.05	57.40	0.000
	3	0.02	12.85	0.000
	4	0.00	13.00	0.000
	5	0.00	20.53	0.000
Tam Bağlantı	2	0.04	69.51	0.000
	3	0.10	13.47	0.000
	4	0.00	10.59	0.000
	5	0.00	14.79	0.000

Çizelge 4.24’u incelediğimizde Hotelling Lawley İz Ölçütlerine göre k-Ortalama yönteminden elde edilen kümeler $P > 0.01$ düzeyinde önemsiz çıkmıştır. Fıstık tiplerinin ikiye ayrılması durumunda en iyi küme yapısını Ward yöntemi; 3, 4 ve 5 kümeye ayrılması durumunda ise Tam Bağlantı yöntemi vermiştir.

Çizelge 4.24. Korelasyon analizi ile elemine edilmiş değişkenlere ilişkin farklı küme sayılarının Hotelling Lawley Trace (İz) ölçütü ile önemlilik düzeylerinin belirlenmesi

Kümeleme Yöntemi	Küme Sayısı	Hotelling Lawley İz Ölçütü	Test İstatistiği F	Önemlilik Düzeyi P
k-Ortalama	2	7.71064	4.112	0.025
	3	13.17461	2.635	0.039
	4	30.22194	3.134	0.012
	5	42.34516	2.470	0.032
Ward Yöntemi	2	26.31562	14.035	0.000
	3	34.43377	6.887	0.001
	4	70.26158	7.286	0.000
	5	97.66573	5.697	0.000
Tam Bağlantı	2	13.73398	7.325	0.004
	3	38.28128	7.656	0.000
	4	245.59517	25.469	0.000
	5	953.11792	55.599	0.000

Ana bileşenler analizi ile elemine edilen değişkenler, Hotelling Lawley İz ölçütüne göre değerlendirilecek olursa; fıstık tiplerinin 2 ve 4 kümeye ayrılması durumunda en iyi küme yapısını Ward yönteminin verdiğini Çizelge 4.25'den görmek mümkündür. Fıstık tiplerinin 3 kümeye ayrılması durumunda en iyi küme yapısını Tam Bağlantı yöntemi, 5 kümeye ayrılması durumunda ise k-Ortalama yöntemi vermiştir.

Çizelge 4.25. Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlere ilişkin farklı küme sayılarının Hotelling Lawley Trace (İz) ölçütü ile önemlilik düzeylerinin belirlenmesi

Kümeleme Yöntemi	Küme Sayısı	Hotelling Lawley İz Ölçütü	Test İstatistiği F	Önemlilik Düzeyi P
k-Ortalama	2	15.56241	20.231	0.000
	3	15.59810	8.579	0.000
	4	23.41867	7.546	0.000
	5	73.53192	15.626	0.000
Ward Yöntemi	2	33.56928	43.640	0.000
	3	37.27117	20.499	0.000
	4	55.35712	17.837	0.000
	5	68.05547	14.462	0.000
Tam Bağlantı	2	22.56820	29.339	0.000
	3	40.64567	22.355	0.000
	4	44.54142	14.352	0.000
	5	61.49244	13.067	0.000

Stepwise diskriminant analizi ile elde edilen sonuçlar Hotelling Lawley İz ölçütüne göre karşılaştırılmamaktadır. Çünkü, her yöntem için elemine edilen değişken sayıları farklıdır. Ancak Çizelge 4.26 incelendiğinde elde edilen kümelerin $P < 0.01$ düzeyinde önemli çıktığı belirlenmiştir.

Çizelge 4.26. Stepwise Diskriminant analizi ile boyutu indirgenmiş değişkenlere ilişkin farklı küme sayılarının Hotelling Lawley Trace (İz) ölçütü ile önemlilik düzeylerinin belirlenmesi

Kümeleme Yöntemi	Küme Sayısı	Hotelling –Lawley İz Ölçütü	Test İstatistiği F	Önemlilik Düzeyi P
k-Ortalama	2	31.20	88.41	0.000
	3	14.81	23.70	0.000
	4	32.19	24.44	0.000
	5	21.43	5.65	0.000
Ward Yöntemi	2	20.26	57.4	0.000
	3	12.54	12.54	0.000
	4	29.19	17.61	0.000
	5	29.99	20.24	0.000
Tam Bağlantı	2	24.53	69.51	0.000
	3	5.04	15.12	0.000
	4	29	14.1	0.000
	5	96.19	25.38	0.000

4.8. Küme Sayısının Belirlenmesi

Küme sayısına karar vermek için bölüm 3.2.8’de belirlenen ölçütleri irdeleyecek olursak,

$$k \cong \left(\frac{n}{2} \right)^{1/2} \text{ ölçütüne göre;}$$

$$k \cong \left(\frac{24}{2} \right)^{1/2} \cong 3.46 \cong 3$$

fıstık tiplerinin üç kümeye ayrılması gerekmektedir. Ancak bu yöntemin istatistiksel dayanağı olmadığından çok itibar edilen bir ölçüt olmadığı ve daha çok büyük örneklerde iyi sonuç verdiği daha önceden belirtilmişti.

$M = k^2 |W|$ ölçütüne göre; M'yi minimize edecek " k " (küme sayısı) değerini tespit edecek olursak; ölçüt Ward yöntemi ile belirlenen kümelere uygulandığında Çizelge 4.27'deki sonuçlar elde edilir. Çizelgeden de görüldüğü gibi ölçüt pek güvenilir sonuçlar vermemektedir. Çünkü küme sayısı arttıkça M ölçütünü minimize edecek sonuçlarda onun paralelinde küçülmektedir. Benzer sonuçlar Çizelge 4.28 ve Çizelge 4.29'de verilen Tam Bağlantı ve k-Ortalama yöntemi için de elde edilmiştir.

Çizelge 4.27. Ward yöntemi ile gruplara ayrılan fıstık tiplerinin, $M = k^2 |W|$ ölçütü ile en uygun küme sayısının belirlenmesi.

Değişken Tipleri	Küme Sayıları	$M = k^2 W $ Sonuçları	En Uygun Küme Sayısı
Korelasyon analizi ile elemine edilen değişkenler	2	20.76E+30	
	3	28.41E+29	
	4	21.37E+28	
	5	12.73E+27	Min M
Ana Bileşenler analizi ile elemine edilen değişkenler	2	6.00E+16	
	3	2.93E+16	
	4	6.91E+15	
	5	1.63E+15	Min M
Stepwise Diskriminant analizi ile elemine edilen değişkenler	2	34416674657	
	3	5,70502E+12	
	4	2715327701	Min M
	5	9846561289	

Çizelge 4.28. Tam Bağlantı yöntemi ile gruplara ayrılan fıstık tiplerinin, $M = k^2 |W|$ ölçütü ile en uygun küme sayısının belirlenmesi.

Değişken Tipleri	Küme Sayıları	$M = k^2 W $ Sonuçları	En Uygun Küme Sayısı
Korelasyon analizi ile elemine edilen değişkenler	2	26.88E+30	
	3	30.78E+29	
	4	57.07E+27	
	5	38.99E+26	Min M
Ana Bileşenler analizi ile elemine edilen değişkenler	2	8.81E+16	
	3	2.68E+16	
	4	9.94E+15	
	5	2.79E+15	min M
Stepwise Diskriminant analizi ile elemine edilen değişkenler	2	28651794146	
	3	54721,37	min M
	4	1866008910	
	5	200805861	

Çizelge 4.29. k-Ortalama yöntemi ile gruplara ayrılan fıstık tiplerinin, $M = k^2 |W|$ ölçütü ile en uygun küme sayısının belirlenmesi.

Değişken Tipleri	Küme Sayıları	$M = k^2 W $ Sonuçları	En Uygun Küme Sayısı
Korelasyon analizi ile elemine edilen değişkenler	2	48.40E+30	
	3	19.80E+30	
	4	18.0E+29	
	5	38.50E+28	min M
Ana Bileşenler analizi ile elemine edilen değişkenler	2	12.53E+16	
	3	9.22E+16	
	4	1.73E+16	
	5	1.13E+15	min M
Stepwise Diskriminant analizi ile elemine edilen değişkenler	2	2,27E+10	
	3	1,01369E+11	
	4	3105208495	min M
	5	1,0505E+14	

Min M ölçütü diğer ölçütlerle aynı sonuçları vermemiş olup, küme sayısı arttıkça M ölçütü küçülmekte ve tutarsız bir sonuç vermektedir.

$$C = \left[\frac{iz(B)}{k-1} \right] / \left[\frac{iz(W)}{N-k} \right] \text{ ölçütüne göre; Ward yöntemi ile kümelere}$$

ayrılan fıstık tiplerinin kaç kümeye ayrılacağı belirlenmesinde, eşitlikteki C katsayısını maksimize eden (k) değeri küme sayısı olarak kabul edilecektir. Her bir (k) değeri için C katsayısının aldığı değer Çizelge 4.30'de verilmiştir.

Çizelge 4.30. Ward yöntemine göre C_{max} eşitliği vasıtasıyla küme sayısının belirlenmesi

Değişken Tipleri	Küme Sayıları	$C = \left[\frac{iz(B)}{k-1} \right] / \left[\frac{iz(W)}{N-k} \right]$ sonuçları	En Uygun Küme Sayısı
Korelasyon analizi ile elemine edilen değişkenler	2	5.84	
	3	17.39	Max
	4	12.54	
	5	9.13	
Ana Bileşenler analizi ile elemine edilen değişkenler	2	6.98	Max
	3	6.31	
	4	6.08	
	5	5.56	
Stepwise Diskriminant analizi ile elemine edilen değişkenler	2	1,99	
	3	3,34	
	4	19,70	Max
	5	14,46	

Çizelge 4.30'ü incelediğimizde Korelasyon analizi ile elde edilen sonuçlara göre fıstık tiplerinin 3 kümeye, buna karşılık Ana Bileşenler analizi sonuçlarına göre Ward yöntemi ile kümelenen fıstık tiplerinin ise 2 kümeye ayrılması gerekmektedir. Bununla birlikte Wilk's Lambda ve Hotelling Lawley İz istatistiklerine göre; Ana

Bileşenler analizi istatistiksel olarak daha anlamlı çıktığından Ward yöntemine göre sınıflandırılan fıstık tiplerinin iki kümeye ayrılması gerektiğine karar verilmiştir.

Tam Bağlantı kümeleme yöntemine göre fıstık tipleri farklı kümelere ayrıldıklarında, her bir küme grubu için çok değişkenli varyans analizi ile elde edilen Kümeler arası kareler ve çarpımlar toplamı (B) ile Kümeler içi kareler ve çarpımlar toplamı (W) vasıtasıyla C katsayısını max. yapan küme sayısı Cetvel 4.31'de tespit edilmiştir.

Çizelge 4.31. Tam Bağlantı kümeleme yöntemine göre C_{max} eşitliği vasıtasıyla küme sayısının belirlenmesi

Değişken Tipleri	Küme Sayıları	$C = \left[\frac{iz(B)}{k-1} \right] / \left[\frac{iz(W)}{N-k} \right]$ sonuçları	En Uygun Küme Sayısı
Korelasyon analizi ile elemine edilen değişkenler	2	4.44	Max
	3	3.94	
	4	2.69	
	5	2.15	
Ana Bileşenler analizi ile elemine edilen değişkenler	2	7.76	Max
	3	5.41	
	4	5.83	
	5	5.44	
Stepwise Diskriminant analizi ile elemine edilen değişkenler	2	91,68	Max
	3	4,96	
	4	10,81	
	5	7,17	

Çizelge 4.31'de ki sonuçlara göre Tam Bağlantı kümeleme yönteminin ayırmış olduğu kümeler içinde max C katsayısını veren küme sayısı 2 olarak tespit edilmiştir. Ancak fıstık tiplerini gerçeğe en yakın sınıflandırması bakımından Ward kümeleme yönteminden elde edilen sonuçlar daha güvenilir olmaktadır.

K-Ortalama Yöntemine göre her bir kümeleme seviyesi için Çok Değişkenli Varyans Analizi ile hesaplanan (B) ve (W) vasıtasıyla max C katsısı Çizelge 4.32’de belirlenmiştir.

Çizelge 4.32. K-Ortalama yöntemine göre C_{\max} eşitliği vasıtasıyla küme sayısının belirlenmesi

Değişken Tipleri	Küme Sayıları	$C = \left[\frac{iz(B)}{k-1} \right] / \left[\frac{iz(W)}{N-k} \right]$ sonuçları	En Uygun Küme Sayısı
Korelasyon analizi ile elemine edilen değişkenler	2	10.52	
	3	11.95	Max
	4	2.57	
	5	2.13	
Ana Bileşenler analizi ile elemine edilen değişkenler	2	7.67	Max
	3	5.63	
	4	3.57	
	5	5.60	
Stepwise Diskriminant analizi ile elemine edilen değişkenler	2	4.43	
	3	2.77	
	4	8.27	
	5	9.08	Max

K-Ortalama yöntemine göre fıstık tipleri farklı kümelere ayrıldığında Cetvel 4.32’deki sonuçlara göre, en uygun küme sayısının Korelasyon analizi sonuçlarına göre 3 olduğu belirlenmiştir. Ancak daha önce de belirtildiği gibi Wilk’s Lambda ve Hotelling Lawley İz istatistikleri sonuçlarına göre Ana Bileşenler analizi ile elde edilen sonuçlar daha anlamlı olduğu için k-Ortalama yöntemine göre fıstık tiplerinin 2 küme ayrılması gerekmektedir.

5. SONUÇLAR

Araştırma sonuçlarına göre, kümeleme analizlerini sağlıklı bir şekilde uygulayabilmek için, eğer uygulamada kullandığımız veri seti gibi ($n < p$), değişken sayısı kümelenecek birim sayısından fazla ise bu durumda veri setinin boyutu en azından ($n \geq p$) şekline indirgenmesi gerekmektedir. Bunu sağlamak için Korelasyon analizi, Ana Bileşenler analizi ve Stepwise Diskriminant analizi gibi değişken sayısını azaltan yöntemler kullanılabilir. Bu çalışmada her üç yöntemde kullanılmış olup, elde edilen değişken gruplarına kümeleme analizi yöntemleri uygulanmış ve sonuçlar karşılaştırılmıştır.

Korelasyon analizi yöntemiyle, yüksek derecede bağımlı değişkenler tespit edilerek elemine edilmiştir. Böylece fıstık tipleri üzerinde ölçümü yapılan değişkenlerin boyutu ($n > p$) şekline indirgenmiş ve oluşturulan kümelerin istatistiksel olarak anlamlılığının test edilmesi için Wilk's Lambda ile Hotelling Lawley İz istatistiklerinin hesaplanabilmesi sağlanmıştır.

Ana Bileşenler analizi ile de varyasyonu en iyi açıklayan Ana Bileşen skorları tespit edilmiştir. Genel Varyasyonun % 89.2'sini açıklayan Çizelge 4.8'deki 10 adet ana bileşen seçilmiş olup, bu değişkenler gerçek değişkenlerin yerine kümeleme analizinde kullanılmıştır. Böylece hem değişkenlerin boyutu ($n > p$) şekline indirgenmiş, hem de her iki boyut indirgeme yöntemi için bağımsız değişken ve skorları içeren değişkenler elde edilmiştir. Geri kalan değişken skorlarından 11-22 arasındaki varyansı çok düşük, 23-38 arasındaki değişken skorlarının

olarak daha anlamlı çıktığı belirlenmiştir. Bu durumda kümeleme yapılırken özellikle çok sayıda ($n < p$) anlamsız değişkenin bulunduğu durumlarda, orijinal değişkenler yerine Ana Bileşenler analizi ile boyutu indirgenmiş değişkenlerin kullanılması önerilebilir. Böylece küme yapısını olumsuz yönde etkileyecek olan anlamsız ve yüksek korelasyonlu değişkenlerin etkisi elemine edilmiş olur.

Sözü edilen değişken tiplerine göre kümeleme yöntemleri karşılaştırılmış ve en iyi küme yapısını ortaya koyan yöntemin Ward yöntemi ile en iyi uzaklık ölçütünün, Öklid uzaklığının karesi olduğu tespit edilmiştir. Bununla birlikte Öklid uzaklığı, Öklid uzaklığının karesi ile benzer sonuçlar verdiğini; ancak diğer uzaklık ölçüleri ile elde edilen küme yapılarının anlamsız olduğu tespit edilmiştir. Bu sonuçlar Kuiper and Fisher (23)'in literatür bilgisi ile uyum sağlamaktadır.

Ayrıca Blashfield (29); Monte Carlo teknikleri ile karışık dağılıma sahip veriler türeterek, oluşturulan örneklerle farklı kümeleme yöntemlerini uygulamıştır. Sözü edilen karışık dağılımlı modeli en iyi belirleyen yöntemin Ward yöntemi olduğunu belirlemiş olup, bu durum çalışmamızda ki sonuçları desteklemektedir.

Ancak en iyi kümeleme sonuçlarını veren yöntemin, her zaman Ward yöntemi olduğunu söylemek pek doğru değildir. Çünkü birimlerin kümelenmesinde, kümeleme yöntemlerinin gösterdikleri başarı veri yapısına bağlı olarak değişmektedir. Yani Verilerin dağılım şekli değiştikçe, farklı bir kümeleme yöntemi daha iyi sonuçlar verebilmektedir. Bununla ilgili bir kısım çalışmalar yapılmış olup, Fisher and Van Ness (21)'in de açıklamaları bu doğrultudadır.

İkinci sırada en anlamlı küme yapısını veren kümeleme yöntemleri Tam Bağlantı, McQuitty ve Grup İçi Ortalama Bağlantı kümeleme Yöntemleridir. Bu yöntemlerin Ward yönteminden farkı, orta geççi bir fıstık tipi olan 2 No'lu fıstık tipinin erkenci fıstık tipleri ile birlikte kümelenmesidir. Aslında, bu fıstık tipinin erkenci fıstık tipleri ile geççi fıstık tiplerinin bir melezi olduğu bilinmektedir. Sözü edilen fıstık tipi her ne kadar orta geççi fıstık tipleri ile birlikte değerlendiriliyorsa da büyük bir olasılıkla melez olduğundan, diğer bazı özellikler bakımından erkenci fıstık tiplerine de yakınlık göstermektedir.

Fıstık tiplerinin kaç kümeye ayrılacağı konusunda ise test istatistikleri genel olarak 2 ve 3 küme olmak üzere farklı sonuçlar vermiştir. Ancak en güvenilebilir

sonuçları Ana Bileşenler analizi ortamında veren C_{\max} ölçütü ile Wilk's Lambda ve Hotelling Lawley İz ölçütleri küme sayısının 2 olması gerektiğini göstermektedirler.

Dinçer ve Özdamar (36)'ın çalışmasında Wilk's Lambda ve Hotelling Lawley İz istatistiklerinin en iyi sonuçları verdiği bildirilmekte idi. Bu durum, çalışmamızda bulduğumuz sonuçları destekler mahiyettedir.



6. KAYNAKLAR

- (1) Stebbins, G.L., Variation and evolution in plants. Columbia Univ. Press, New York, 1950.
- (2) Jain, S.K., Population structure and the effects of breeding system. In Crop Genetic Resources for Today and Tomorrow, eds. Frankel, O.H. and Hawkes, J.G., 1975.
- (3) Krull, C.F. and Borlaug, N.E., The utilisation of collections in plant breeding and production. In Genetic Resources in Plants Eds. Frankel, O.H. and Bennett, E., 1970.
- (4) Stace, C.A., Plant taxonomy and biosystematics. Arnold, London, 1980.
- (5) Tan, A., Sayısal taksonomik yöntemlerle varyasyonun saptanması. Ege-Bölge Ziraat Araştırma Enstitüsü Yay. No: 30, Menemen- İzmir, 1983.
- (6) Johnson, A.R. and Wichern, D.W., Applied multivariate statistical analysis. Prentice Hall International Inc. New Jersey, 1992.
- (7) Tyron, R.C. and Bailey, D.E., Cluster analysis. McGraw Hill Book Company, New York, 1970.
- (8) Anderberg, M.R., Cluster analysis for applications. Academic Press, New York, 1973.
- (9) Kara, İ., Olasılık. Bilim Teknik Yayınevi, İstanbul, 1982.
- (10) Kendall, S.M. and Buckland, W.R., A dictionary of statistical terms, Hafner, New York, 1960.
- (11) Everitt, B., Cluster Analysis. Heinemann, London, 1974.
- (12) Özdamar, K., Hastalık olgularının incelenmesinde kümeleme çözümlemesinin kullanılması. T.C. Anadolu Üniversitesi Yayınları No:295, Tıp Fakültesi Yayınları No:25, Eskişehir, 1988.
- (13) Sneath, P.A.H. and Sokal, R.R., Numerical taxonomy. W.H. Freeman and Company, San Francisco, 1973.
- (14) Aldenderfer, M.S. and Blasfield, R.K., Cluster analysis. Sage Publication, California, 1984.
- (15) Alçal, H., Veri yapısının araştırılması ve doğal kümelerin belirlenmesi amacıyla kümeleme çözümlemesi ve bir uygulama. Yüksek Lis. Tezi (Yayınlanmamış), Anadolu Üni., 1991.
- (16) Chatfield, C. and Collins, A.J., Introduction to multivariate analysis. Chapman and Hall, London, 1980.
- (17) Öner, Y., Kümeleme çözümlemesi ve bir geometrik yaklaşım. Doktora Tezi (Yayınlanmamış), Gazi Üniversitesi, Ankara, 1993.
- (18) Edwards, A.W.F. and Cavalli-Sforza, L.L., A method for cluster analysis. Biometrics, 21: 362-375, 1965.

- (19) Scott, A.J. and Knott, M., A cluster analysis method for variance. *Biometrics*, 30: 507-512, 1974.
- (20) Jardine, N. and Sibson, R., The construction of hierarchic and non-hierarchic classifications. *The Comp. J.*, 11: 177-184, 1968.
- (21) Fisher, L. and Van Ness, J.W., Admissible clustering procedures. *Biometrika*, 58: (1), 91-104, 1971.
- (22) Everitt, B.S., Unresolved problems in cluster analysis. *Biometrics*. 35: 169-181, 1979.
- (23) Kuiper, F.K. and Fisher L., A monte carlo comparison of six clustering procedures: *Biometrics*, 31: 777-783, 1975.
- (24) Cormack, R.M., A review of classification. *J. Royal statist. Soc., Seri A*, 134: 321-367, 1971.
- (25) Marriott, F.H.C., Practical problems in a method of cluster analysis. *Biometrics*, 27: 501-514, 1971.
- (26) Stoddard, A.M., Standardisation of measures prior to cluster analysis. *Biometrics*, 35: 765-773, 1979.
- (27) Hartigan, J.A., Direct clustering of a data matrix. *JASA*, 67: (337), 123-129, 1972.
- (28) Murtagh, F., A probability theory of hierarchic clustering using random dendograms. *J. Statist. Comput. Simul.*, 13: 145-157, 1983.
- (29) Blashfield, R.K., Mixture model test of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bull.*, 83: 377-388, 1976.
- (30) Backer, F.B., Stability of two hierarchical grouping techniques case I: Sensivity of data errors, *JASA*, 69: 440-445, 1975.
- (31) Gower, J.C., A comparison of some methods of cluster analysis. *Biometrics*, 23: 623-628, 1967.
- (32) Lance, G.N. and Williams, W.T., A General Theory of classificatory sorting strategies: 1-Hierarchical systems. *Computer Journal*, 9: 373-380, 1967.
- (33) Bağcı, H., Shareef, S.R. ve Özdamar, K., *Bacillus Thuringiensis* varyetelerinin sınıflandırılmasında sayısal taksonominin uygulanması. *Doğa*, 15: 70-81, 1991.
- (34) Arnold, S.J., A test for cluster. *Journal of Marketing Research*, XVI: 545-551, 1979.
- (35) Punj, G. and Stewart, D.W., Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20: 134-148, 1983.
- (36) Dinçer, K.S. ve Özdamar, K., Kümeleme çözümlenmesinde uygun kümeleme ölçütlerinin karşılaştırılması. *Hacettepe Fen ve Mühendislik Bilimleri Dergisi*, 14: 17-33, 1992.
- (37) AÇAR, İ., Ceylanpınar tarım işletmesinde seçilmiş bazı erkek antep fıstığı tiplerinin morfolojik ve biyolojik özellikleri üzerinde bir araştırma. *Harran Üniv. Fen Bil. Enst. Y. Lisans Tezi (Basılmamış)*. Ş.Urfa, 1997.

- (38) Yıldız, N. ve Bircan, H., Uygulamalı istatistik. Dicle Üni., Ziraat Fak. Yayınları Şanlıurfa, 1992.
- (39) Überla, K., Faktorenanalyse. Springer-Verlag Berlin Haidelberg, New York, 1968.
- (40) Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C., Multivariate data analysis. Prentice-Hall, Inc. A Simon & Schuster Company, New Jersey, 1995.
- (41) Tathdil, H., Uygulamalı çok değişkenli istatistik analiz. Hacatepe Üni., Ankara, 1992.
- (42) Späth, H., Cluster analysis algorithms for data reduction and classification of objects. John Wiley & Sons, New York, 1980.
- (43) Johnson, S.C., Hierarchical clustering schemes. Psychometrika, 32: (3), 241-254, 1967.
- (44) Gower, J. C., Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, 53: (3-4), 325-338, 1966.
- (45) Rao, C.R., Cluster analysis applied to a study of race mixture in human population, Van Ryzin, J., (Der.), Classification and clustering. Academic Press, New York, 1977.
- (46) Rohlf, F.J., Adaptive hierarchical clustering schemes. Syst. Zool., 19: 58-82, 1970.
- (47) Çetinel, B., Çok değişkenli verilerin kümelendirilmesi için istatistiksel bir yöntem. Doktora Tezi (yayınlanmamış), Hacettepe Üni., Ankara, 1982.
- (48) Duran, S.S. and Odell, P.L., Cluster Analysis, Heineman, London, 1974.
- (49) Hawkins, D. M., Müller, M.W. and Krooden, J.A., Topics in applied multivariate analysis. Cambridge University Press, USA, 1982.
- (50) Lance, G. N. and Williams, W. T., Computer Program for Hierarchical polythetic classification., Comput. J. 9, 1: 60-64, 1966.
- (51) Minitap Inc., Minitap release, V 10.51, 1995.
- (52) Sözer, E.E., Çelik, Y. ve Kutsal, A., Kümeleme çözümlemesinde çekirdek noktaların seçimi. İstatistik Sempozyumu 86, Orta Doğu Teknik Üni. ve Ankara Üni., 15-17 Eylül, 1986.
- (53) Mac Queen, J., Some methods for classifications and stopping rules: An evaluation. Comp. J., 20: 359-363, 1967.
- (54) Krzanowski, W.J. and Lai, Y.T., A Critetion for Determining the Number of Grups in a Data Set Using Sum-of-Squares Clustering. Biometrics 44: 23-34, 1988.
- (55) Bek, Y., Çok değişkenli varyans analizi ders notları. Çukurova Üni., Adana, 1998.
- (56) Calinski, T. and Harabasz, J., A dendrite method for cluster analysis. Unpublished manuscript, 1971.
- (57) Norusis, M.J., SPSS/PC, SPSS İnc., V.6.0 Chicago, 1993.
- (58) Statistica Inc., Statistica release, V 10.51., 1995.
- (59) Özdamar, K., Çok değişkenli varyans analizi ders notları. Osmangazi Üni. Eskişehir, 1998.

(60) Sharma, S., Applied Multivariate Techniques. John Wiley & Sons, Inc. New York. 493 p., 1996.



7. EKLER



Çizelge 4.2. Karakterlere ilişkin tanımlama ve değişken tipleri

Tanımlama	Karakterler	Değişken Tipi
C1	Büyüme Biçimi	Nominal
C2	Sürgün Uzunluğu	Sürekli
C3	Dal Sayısı	Kesikli
C4	Sürgün Rengi	Nominal (Sıralama)
C5	Dal Kalınlığı	Sürekli
C6	Yaprak Uzunluğu	Sürekli
C7	Yaprak Genişliği	Sürekli
C8	Yaprakçık Sayısı	Kesikli
C9	Uç Yaprak Uzunluğu	Sürekli
C10	Uç Yaprak Genişliği	Sürekli
C11	Uç Yaprak İriliği	Nominal (Sıralama)
C12	Uç Yaprak Şekli	Nominal
C13	Uç Yaprak Tepesi	Nominal
C14	Uç Yaprak Tabanı	Nominal
C15	Uç Yaprak Kenarı	Nominal
C16	Yaprak Rengi	Nominal (Sıralama)
C17	Karagöz Ağırlığı	Sürekli
C18	Karagöz Şekli	Nominal
C19	Karagöz Rengi	Nominal (Sıralama)
C20	Sürgündeki Karagöz Sayısı	Kesikli
C21	Karagöz Uzunluğu	Sürekli
C22	Karagöz Genişliği	Sürekli
C23	Gözlerin Kabarması	Nominal
C24	Gözlerin Patlaması	Nominal
C25	Çiçeklenme Başlangıcı	Nominal
C26	Tam Çiçeklenme	Nominal
C27	Çiçeklenme Sonu	Nominal
C28	Çiçeklenme Süresi	Kesikli
C29	Salkım Sayısı	Kesikli
C30	Salkım Uzunluğu	Sürekli
C31	Bir Salkımdaki Çiçek Sayısı	Kesikli
C32	Verim Potansiyeli	Sürekli
C33	Çiçek Tozu Üretimi	Sürekli
C34	Salkım Ağırlığı	Sürekli
C35	Bir Çiçekteki Anter Sayısı	Kesikli
C36	Bir Anterdeki Çiçek Tozu Sayısı	Kesikli
C37	Çiçek Tozu Canlılık Oranı	Oranlı (Ratio)
C38	% 20 Sakkarozda Çiçek Tozu Çimlenme Oranı	Ratio

Çizelge 4.3.-1. Ceylanpınar Tarım İşletmesi (TİGEM)'den 24 farklı tipte erkek fıstık ağacına ilişkin 38 adet karakter üzerinden alınmış ölçümler

Karakterler F. Tipleri	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
1	2	12.11	9.75	2	10.05	15.70	14.54	3	9.40	6.60	3	1	2	4	2	2	4.04	2	4
2	1	2.96	8.75	4	8.40	14.70	13.14	6	6.84	3.73	2	2	1	4	1	1	1.25	1	4
3	1	5.20	10.50	1	7.55	19.55	16.41	5	9.44	4.73	2	2	1	1	1	3	1.90	3	2
4	1	2.54	9.00	1	9.83	19.51	15.75	5	8.94	4.58	2	2	1	1	1	3	1.64	3	1
5	1	5.31	11.25	2	7.20	15.35	13.34	7	5.46	3.28	1	1	1	3	2	3	1.65	2	4
6	1	3.62	11.00	1	7.76	17.94	17.46	5	8.19	4.31	2	2	1	1	2	2	2.51	3	4
7	2	11.46	14.25	2	9.27	16.41	15.21	3	9.90	7.26	3	5	3	2	2	2	2.94	2	4
8	3	10.30	7.75	1	9.14	16.75	14.69	3	9.53	7.03	3	1	2	2	2	2	4.14	2	4
9	2	2.51	10.00	2	7.22	16.40	14.39	3	10.53	6.74	3	3	3	2	2	2	1.72	3	3
10	1	7.46	9.25	1	8.96	18.56	16.68	5	8.69	5.00	2	1	1	1	2	2	2.80	2	4
11	1	11.92	5.50	2	8.58	16.63	14.20	4	7.73	4.41	2	2	1	3	1	2	1.50	3	1
12	1	13.09	6.50	1	11.56	21.90	17.33	6	8.53	4.31	2	1	1	1	1	3	1.40	2	2
13	2	11.66	8.25	1	7.68	21.65	18.20	5	9.69	6.09	2	1	1	2	2	1	3.42	3	2
14	1	9.82	4.75	1	9.62	17.50	16.19	3	9.10	5.39	3	1	2	4	2	2	2.90	2	1
15	1	15.28	5.75	2	8.74	17.28	13.64	7	5.41	3.44	1	4	2	3	2	2	1.18	2	3
16	1	15.90	4.50	2	6.63	14.70	13.24	3	8.68	6.36	2	1	1	2	1	2	1.15	2	4
17	2	30.46	7.25	1	11.23	16.21	14.61	3	9.33	6.08	3	1	3	2	2	2	2.42	2	4
18	2	10.61	4.25	2	11.52	19.91	17.90	3	11.28	6.13	2	2	1	1	1	3	2.57	3	1
19	2	8.13	1.50	1	9.84	20.46	16.94	4	10.25	5.38	2	4	1	1	2	3	2.39	3	1
20	3	16.67	4.50	1	9.05	13.79	12.25	3	8.90	6.39	3	1	2	3	1	2	2.42	3	1
21	1	11.78	2.50	1	8.22	19.18	16.18	5	8.36	5.71	3	1	1	4	1	2	2.30	3	1
22	1	13.62	7.00	1	9.06	14.79	15.10	2	8.85	7.29	3	5	3	2	2	2	1.93	3	4
23	3	12.60	5.00	2	7.69	14.09	11.73	3	8.15	5.99	3	1	3	2	2	2	1.79	3	4
24	1	10.15	6.00	2	6.29	17.80	16.26	5	8.56	4.96	2	2	1	2	1	2	2.53	3	4

Çizelge 4.3.-2. Ceylanpınar Tarım İşletmesi (TİGEM)'den 24 farklı tipte erkek fıstık ağacına ilişkin 38 adet karakter üzerinden alınmış ölçümler

Karakterler F. Tipleri	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38
1	6.38	11.97	7.63	16	19	34	39	42	9	5.95	6.13	509.0	93.33	6.53	55.45	5	54375	89.62	93.72
2	2.92	7.62	5.59	14	21	37	40	42	6	2.82	7.12	249.7	96.43	3.34	23.70	6	29688	81.94	93.68
3	7.81	9.66	5.55	9	19	32	35	39	8	7.10	6.28	373.0	90.88	4.84	35.56	5	25625	90.54	65.71
4	7.25	8.87	5.48	9	19	32	35	39	8	6.75	7.75	321.3	93.10	3.56	38.87	5	46250	87.43	74.18
5	10.19	8.49	5.53	8	13	29	32	36	8	9.20	6.68	282.7	92.24	4.02	28.60	7	34375	95.23	87.23
6	6.69	10.71	5.93	10	14	30	33	36	7	6.00	8.80	407.3	92.72	5.06	33.16	6	42703	95.64	97.36
7	6.44	11.19	6.81	16	22	38	42	46	9	6.15	4.46	339.7	95.54	2.09	38.28	5	45000	93.10	84.16
8	7.63	11.34	7.39	17	20	35	39	43	9	6.65	6.20	354.3	87.21	6.14	56.54	6	56771	85.70	94.39
9	6.38	9.89	5.41	14	21	35	39	44	10	6.05	5.45	195.3	94.90	4.03	32.60	5	56250	92.30	95.63
10	5.50	11.20	6.57	10	15	31	34	36	6	5.15	7.99	561.3	93.64	4.01	35.47	6	30208	88.91	92.54
11	5.19	9.06	5.21	10	15	29	32	35	7	5.00	6.76	308.7	96.39	2.18	23.38	5	38125	92.59	92.63
12	9.00	8.49	5.81	10	19	36	40	43	8	8.30	7.75	759.0	92.22	4.32	60.18	5	29375	84.30	85.72
13	9.63	11.39	7.66	13	25	39	43	47	9	8.60	6.66	758.3	89.35	2.51	45.72	5	46250	79.72	75.16
14	6.06	9.02	5.87	16	24	39	43	47	9	5.60	4.63	262.0	92.38	2.66	48.12	5	40000	85.78	92.27
15	7.38	7.36	5.07	11	19	33	37	40	8	7.00	6.80	282.0	94.92	3.25	19.86	5	67500	87.82	92.18
16	3.88	8.75	4.45	16	24	36	39	43	8	3.90	4.29	219.7	100.00	1.41	12.67	6	32552	82.63	92.69
17	5.50	10.15	6.56	16	23	41	45	48	8	5.50	5.14	315.0	100.00	3.02	51.95	6	40625	87.69	95.03
18	6.88	10.03	6.68	17	25	39	43	48	10	6.05	5.28	250.0	88.00	1.44	44.45	6	43750	84.18	90.15
19	9.81	10.07	6.52	13	26	40	43	46	7	8.40	6.21	527.7	35.61	1.58	35.43	5	26875	84.65	84.12
20	6.19	11.00	6.20	16	23	39	42	47	9	5.90	5.67	387.3	95.36	2.68	33.04	5	43125	87.78	89.35
21	7.50	10.53	6.58	14	26	42	45	49	8	6.80	6.13	822.0	90.67	2.31	45.89	5	22500	79.18	76.63
22	5.94	10.05	6.01	13	21	33	38	42	10	5.38	5.02	307.7	90.53	1.53	32.50	7	25893	74.20	80.73
23	6.38	10.53	5.61	15	21	36	40	44	9	5.90	5.56	265.7	92.55	5.03	26.77	6	58333	92.02	95.18
24	5.88	10.15	5.91	10	17	29	33	36	8	5.63	7.58	552.3	95.74	4.00	28.70	6	28646	85.54	86.13

Çizelge 4.4 -1. Değişkenlere (karakterlere) ilişkin korelasyon analizi sonuçları

	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19
D01	1	0,28	-0,1	-0,09	0,18	-0,25	-0,3	-0,54**	,43*	,61**	,54**	-0,11	,50**	-0,04	0,29	-0,12	,49**	0,15	0,03
D02	0,28	1	-0,35*	-0,22	,36*	-0,21	-0,21	-0,34	0,04	0,34	0,29	-0,12	,41*	0,13	0,08	-0,16	0,09	-0,12	0,09
D03	-0,10	-0,35*	1	0,16	-0,16	-0,11	0,02	0,16	-0,07	-0,04	-0,06	0,16	0,18	-0,13	0,31	-0,08	0,16	-0,27	,52**
D04	-0,09	-0,22	0,16	1	-0,25	-0,45*	-0,47**	0,18	-0,35*	-0,24	-0,22	0,12	-0,01	,41*	-0,19	-0,34	-0,34	-0,46*	,34*
D05	0,18	0,36*	-0,16	-0,25	1	0,3	0,28	-0,2	0,33	0,13	0,2	0,03	0,13	-0,14	0,02	0,33	0,21	-0,19	-0,32
D06	-0,25	-0,21	-0,11	-0,45*	0,30	1	,88**	,36*	0,32	-0,22	-0,31	-0,04	-0,54**	-0,48**	-0,15	,36*	0,14	0,27	,51**
D07	-0,3	-0,21	0,02	-0,47**	0,28	0,88**	1	0,13	,47*	-0,03	-0,15	0,05	-0,44*	-0,48**	0,23	0,34	0,29	-0,34	
D08	-0,54**	-0,34	0,16	0,18	-0,2	0,36*	0,13	1	-0,69**	-0,89**	-0,80**	-0,11	-0,60**	0,03	-0,18	0,12	-0,39*	-0,23	-0,02
D09	0,43*	0,04	-0,07	-0,35*	0,33	0,32	0,47*	0,69**	1	,72**	,56**	0,07	0,18	-0,42*	0,01	0,13	,50**	,38*	-0,27
D10	0,61**	0,34	-0,04	-0,24	0,13	-0,22	-0,03	-0,89**	0,72**	1	,80**	0,15	,60**	-0,05	0,24	-0,21	,53**	0,21	0,11
D11	0,54**	0,29	-0,06	-0,22	0,2	-0,31	-0,15	-0,80**	0,56**	0,80**	1	-0,04	,66**	0,21	0,18	-0,28	,46*	0,12	0,03
D12	-0,11	-0,12	0,16	0,12	0,03	-0,04	0,05	-0,11	0,07	0,15	-0,04	1	,35*	-0,18	0,24	0,07	-0,19	0,14	0,07
D13	0,50**	0,41*	0,18	-0,01	0,13	-0,54**	-0,44*	-0,60**	0,18	0,60**	0,66**	0,35*	1	0,14	,55**	-0,23	0,14	-0,04	,36*
D14	-0,04	0,13	-0,13	0,41*	-0,14	-0,48**	-0,48**	0,03	-0,42*	-0,05	0,21	-0,18	0,14	1	0,01	-0,50**	0,06	-0,38*	0
D15	0,29	0,08	0,31	-0,19	0,02	-0,15	-0,02	-0,18	0,01	0,24	0,18	0,24	0,55**	0,01	1	-0,2	,42*	-0,15	,44*
D16	-0,12	-0,16	-0,08	-0,34	0,33	0,36*	0,23	0,12	0,13	-0,21	-0,28	0,07	-0,23	-0,50**	-0,20	1	-0,2	0,26	-0,35*
D17	0,49**	0,09	0,16	-0,34	0,21	0,14	0,34	-0,39*	0,50**	0,53**	0,46*	-0,19	0,14	0,06	0,42*	-0,20	1	0,01	0,09
D18	0,15	-0,12	-0,27	-0,46*	-0,19	0,27	0,29	-0,23	0,28*	0,21	0,12	0,14	-0,04	-0,38*	-0,15	0,26	0,01	1	-0,46*
D19	0,03	0,09	0,52**	0,34*	-0,32	-0,51**	-0,34	-0,02	-0,27	0,11	0,03	0,07	0,36*	0,00	0,44*	-0,35*	0,09	-0,46*	1
D20	0,12	-0,15	0,01	-0,50**	0,12	0,59**	0,41*	0,34*	0,04	-0,15	-0,29	0,02	-0,20	-0,29	0,26	0,50**	0,19	0,32	-0,32
D21	0,58**	0,13	0,15	-0,41*	0,04	0,03	0,23	-0,53**	0,60**	0,69**	0,59**	-0,12	0,25	-0,14	0,34	-0,19	0,82**	0,34	0,13
D22	0,49**	0,16	0,12	-0,30	0,38*	0,28	0,41*	-0,26	0,49**	0,48**	0,41*	-0,11	0,11	0,02	0,36*	-0,19	0,90**	0,01	0,01
D23	0,64**	0,44*	-0,26	0,11	0,29	-0,35*	-0,23	-0,73**	0,49**	0,75**	0,69**	-0,06	0,51**	0,30	0,16	-0,34*	0,43*	-0,19	0,05
D24	0,37*	0,35*	-0,53**	-0,11	0,27	0,14	0,10	-0,45*	0,53**	0,54**	0,44*	0,08	0,20	0,12	-0,08	-0,14	0,15	0,11	-0,40*
D25	0,44*	0,43*	-0,43*	-0,12	0,43*	0,12	0,07	-0,36*	0,44*	0,46*	0,49**	-0,07	0,24	0,19	0,01	-0,17	0,21	-0,07	-0,34
D26	0,47*	0,47*	-0,40*	-0,12	0,46*	0,10	0,08	-0,42*	0,47**	0,53**	0,54**	-0,03	0,34	0,20	0,08	-0,19	0,26	-0,08	-0,28
D27	0,51**	0,44*	-0,39*	-0,16	0,41*	0,08	0,05	-0,47**	0,52**	0,58**	0,57**	-0,02	0,37*	0,18	0,06	-0,13	0,26	0,02	-0,33
D28	0,47**	0,20	-0,01	-0,18	0,12	-0,13	-0,04	-0,59**	0,46*	0,68**	0,49**	0,16	0,58**	0,03	0,19	0,07	0,28	0,33	-0,08
D29	0,10	-0,10	0,04	-0,50**	0,11	0,57**	0,37*	0,37*	0,00	-0,16	-0,30	0,01	-0,17	-0,27	0,25	0,50**	0,16	0,30	-0,30
D30	-0,37*	-0,45*	0,20	-0,06	-0,09	0,41*	0,33	0,71**	-0,37*	-0,68**	-0,53**	-0,19	-0,59**	-0,26	-0,14	0,07	-0,08	0,05	0,02
D31	-0,10	0,05	-0,17	-0,42*	0,09	0,63**	0,56**	0,31	0,10	-0,04	0,00	-0,25	-0,40*	-0,06	-0,12	-0,04	0,31	0,17	-0,23
D32	-0,18	0,18	0,39*	0,25	-0,19	-0,39*	-0,32	0,02	-0,29	-0,02	0,10	-0,31	0,20	0,28	-0,2	-0,36*	-0,14	-0,26	0,34
D33	0,20	-0,25	0,44*	-0,01	-0,10	-0,07	-0,11	0,18	-0,14	-0,13	0,07	-0,39*	0,06	0,00	0,23	0,02	0,34	-0,17	0,42*
D34	0,28	0,19	0,06	-0,49**	0,66**	0,46*	0,49**	-0,19	0,49**	0,31	0,43*	-0,30	0,10	-0,06	0,16	0,16	0,65**	-0,07	-0,19
D35	-0,10	0,04	0,12	0,18	-0,19	-0,42*	-0,20	-0,03	-0,28	-0,02	-0,14	-0,02	0,07	-0,11	0,19	-0,02	-0,10	-0,17	,62**
D36	0,53**	0,09	0,18	0,13	0,06	-0,23	-0,33	-0,15	-0,02	0,17	0,11	0,03	0,48**	0,10	0,44*	-0,19	0,22	-0,05	0,15
D37	0,13	-0,23	0,51**	0,11	-0,16	-0,19	-0,23	0,14	-0,17	-0,28	-0,15	-0,08	0,10	-0,11	0,23	0,24	-0,02	-0,01	0,18
D38	0,26	0,20	-0,05	0,37*	0,07	-0,51**	-0,41*	-0,22	-0,15	0,04	0,13	-0,16	0,30	0,23	0,33	-0,31	0,05	-0,40*	0,42*

* : p < 0,05; ** : p < 0,01

Çizelge 4.4.-2. Değişkenlere (karakterlere) ilişkin korelasyon analizi sonuçları

	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31	D32	D33	D34	D35	D36	D37	D38
D01	0,12	,58**	,49**	,64**	,37*	,44*	,47*	,51**	,47**	0,1	-,37*	-,01	-,018	0,2	0,28	-,01	,53**	0,13	0,26
D02	-,015	0,13	0,16	,44*	,35*	,43*	,47*	,44*	0,2	-,01	-,45*	0,05	0,18	-,025	0,19	0,04	0,09	-,023	0,2
D03	0,01	0,15	0,12	-,026	-,53**	-,43*	-,40*	-,39*	-,016	0,04	0,2	-,017	,39*	,44*	0,06	0,12	0,18	,51**	-,005
D04	-,050**	-,41*	-,03	0,11	-,011	-,012	-,012	-,016	-,018	-,50**	-,006	-,42*	0,25	-,001	-,49**	0,18	0,13	0,11	,37*
D05	0,12	0,04	,38*	0,29	,43*	,46*	,46*	,41*	0,12	0,11	-,009	0,09	-,019	-,01	,66**	-,019	0,06	-,016	0,07
D06	0,59**	0,03	0,28	-,35*	0,14	0,12	0,1	0,08	-,013	,57**	,41*	,63**	-,39*	-,007	,46*	-,42*	-,023	-,019	-,51**
D07	0,41*	0,23	,41*	-,023	0,1	0,07	0,08	0,05	-,004	,37*	0,33	,56**	-,032	-,011	,49**	-,02	-,033	-,023	-,41*
D08	0,34*	-,53**	-,026	-,73**	-,45*	-,36*	-,42*	-,47**	-,59**	,37*	,71**	0,31	0,02	0,18	-,019	-,003	-,015	0,14	-,022
D09	0,04	,60**	,49**	,49**	,53**	,44*	,47**	,52**	,46*	0	-,37*	0,1	-,029	-,014	,48**	-,028	-,002	-,017	-,015
D10	-,015	,69**	,48**	,75**	,54**	,46*	,53**	,58**	,68**	-,016	-,68**	-,004	-,002	-,013	0,31	-,002	0,17	-,028	0,04
D11	-,029	,59**	,41*	,69**	,44*	,49**	,54**	,57**	,49**	-,03	-,53**	0	0,1	0,07	,43*	-,014	0,11	-,015	0,13
D12	0,02	-,012	-,011	-,006	0,08	-,007	-,003	-,002	0,16	0,01	-,019	-,025	-,031	-,039*	-,03	-,002	0,03	-,008	-,016
D13	-,02	0,25	0,11	,51**	0,2	0,24	0,34	,37*	,58**	-,017	-,59**	-,40*	0,2	0,06	0,1	0,07	,48**	0,1	0,3
D14	-,029	-,014	0,02	0,3	0,12	0,19	0,2	0,18	0,03	-,027	-,026	-,006	0,28	0	-,006	-,011	0,1	-,011	0,23
D15	0,26	0,34	,36*	0,16	-,008	0,01	0,08	0,06	0,19	0,25	-,014	-,012	-,02	0,23	0,16	0,19	,44*	0,23	0,33
D16	0,50**	-,019	-,019	-,34*	-,014	-,017	-,019	-,013	0,07	,50**	0,07	-,004	-,36*	0,02	0,16	-,002	0,19	0,24	-,031
D17	0,19	,82**	,90**	,43*	0,15	0,21	0,26	0,26	0,28	0,16	-,008	0,31	-,014	0,34	,65**	-,01	0,22	-,002	0,05
D18	0,32	0,34	0,01	-,019	0,11	-,007	-,008	0,02	0,33	0,3	0,05	0,17	-,026	-,017	-,007	-,017	-,005	-,001	-,40*
D19	-,032	0,13	0,01	0,05	-,40*	-,034	-,028	-,033	-,008	-,03	0,02	-,023	0,34	,42*	-,019	,62**	0,15	0,18	,42*
D20	1	0,12	0,32	-,033	0,01	0,04	0,04	0,08	0,18	,99**	0,21	,45*	-,51**	0,12	,39*	-,017	0,03	0,08	-,45*
D21	0,12	1	,78**	,37*	0,14	0,19	0,22	0,24	0,3	0,09	-,011	,36*	-,011	0,25	,45*	-,006	0,12	0,05	-,002
D22	0,32	,78**	1	,39*	0,27	,38*	,43*	,40*	0,24	0,28	-,002	,50**	-,024	0,2	,73**	-,011	0,11	-,018	-,009
D23	-,033	,37*	0,39*	1	,71**	,73**	,77**	,78**	,50**	-,36*	-,74**	-,022	0,01	-,017	0,27	-,006	0,31	-,027	,37*
D24	0,01	0,14	0,27	0,71**	1	,92**	,91**	,93**	,40*	-,002	-,64**	0,12	-,032	-,53**	0,22	-,032	-,004	-,60**	-,02
D25	0,04	0,19	0,38*	0,73**	0,92**	1	,99**	,97**	0,27	0,02	-,55**	0,19	-,026	-,41*	,39*	-,032	-,001	-,44*	-,005
D26	0,04	0,22	0,43*	0,77**	0,91**	0,99**	1	,98**	,37*	0,02	-,59**	0,18	-,023	-,37*	,44*	-,029	0,05	-,47*	-,002
D27	0,08	0,24	0,40*	0,78**	0,93**	0,97**	0,98**	1	,50**	0,06	-,65**	0,13	-,02	-,38*	,43*	-,029	0,09	-,44*	-,006
D28	0,18	0,3	0,24	0,50**	0,40*	0,27	0,37*	0,50**	1	0,19	-,62**	-,019	0,13	-,006	0,29	-,001	,41*	-,016	-,008
D29	0,99**	0,09	0,28	-,036*	-,002	0,02	0,02	0,06	0,19	1	-,62**	-,45*	-,44*	0,12	,38*	-,02	0,06	0,12	-,45*
D30	0,21	-,011	-,002	-,074**	-,064**	-,055**	-,059**	-,062**	0,2	0,2	1	,41*	-,003	,44*	0	-,011	0,21	-,006	-,006
D31	0,45*	,36*	0,50**	-,022	0,12	0,19	0,18	0,13	-,019	0,45*	0,41*	1	-,022	0,08	,48**	-,031	-,034	-,41*	-,006
D32	-,051**	-,011	-,024	0,01	-,032	-,026	-,023	-,02	0,13	-,044*	-,003	-,022	1	0,2	-,01	0,16	0,21	0,15	0,19
D33	0,12	0,25	0,2	0,17	-,053**	-,041*	-,037*	-,038*	-,006	0,12	0,44*	0,08	0,2	1	0,32	0,02	,37*	,46*	0,2
D34	0,39*	,45*	0,73**	0,27	0,22	0,39*	0,44*	0,43*	0,29	0,38*	0,01	0,48**	-,01	0,32	1	-,024	0,05	-,015	-,012
D35	-,017	-,006	-,011	-,006	-,032	-,032	-,029	-,029	-,001	-,02	0	-,031	0,16	0,02	-,024	1	-,021	-,01	0,26
D36	0,03	0,12	0,11	0,31	-,004	-,001	0,05	0,09	0,41*	0,06	-,011	-,036*	0,21	0,37*	0,05	-,021	1	,39*	,44*
D37	0,08	0,05	-,018	-,027	-,060**	-,044*	-,047*	-,044*	-,016	0,12	0,21	-,034	0,15	0,46*	-,015	-,010	0,39*	1	0,33
D38	-,045*	-,002	-,009	0,37*	-,02	-,005	-,002	-,006	-,008	-,045*	-,006	-,041*	0,19	0,20	-,012	0,26	0,44*	0,33	1

* : p < 0,05; ** : p < 0,01

Çizelge 4.5. Korelasyon analizi ile (yüksek derecede ilişkili olan karakterleri) elemine edilmiş olan değişkenler

Karakterler F.Tipleri	C5	C12	C14	C15	C17	C18	C26	C28	C29	C32	C33	C35	C36	C37	C38
1	10,05	1	4	2	4,04	2	39	9	5,95	93,33	6,53	5	54375	89,62	93,72
2	8,40	2	4	1	1,25	1	40	6	2,82	96,43	3,34	6	29688	81,94	93,68
3	7,55	2	1	1	1,90	3	35	8	7,10	90,88	4,84	5	25625	90,54	65,71
4	9,83	2	1	1	1,64	3	35	8	6,75	93,10	3,56	5	46250	87,43	74,18
5	7,20	1	3	2	1,65	2	32	8	9,20	92,24	4,02	7	34375	95,23	87,23
6	7,76	2	1	2	2,51	3	33	7	6,00	92,72	5,06	6	42703	95,64	97,36
7	9,27	5	2	2	2,94	2	42	9	6,15	95,54	2,09	5	45000	93,10	84,16
8	9,14	1	2	2	4,14	2	39	9	6,65	87,21	6,14	6	56771	85,70	94,39
9	7,22	3	2	2	1,72	3	39	10	6,05	94,90	4,03	5	56250	92,30	95,63
10	8,96	1	1	2	2,80	2	34	6	5,15	93,64	4,01	6	30208	88,91	92,54
11	8,58	2	3	1	1,50	3	32	7	5,00	96,39	2,18	5	38125	92,59	92,63
12	11,56	1	1	1	1,40	2	40	8	8,30	92,22	4,32	5	29375	84,30	85,72
13	7,68	1	2	2	3,42	3	43	9	8,60	89,35	2,51	5	46250	79,72	75,16
14	9,62	1	4	2	2,90	2	43	9	5,60	92,38	2,66	5	40000	85,78	92,27
15	8,74	4	3	2	1,18	2	37	8	7,00	94,92	3,25	5	67500	87,82	92,18
16	6,63	1	2	1	1,15	2	39	8	3,90	100,00	1,41	6	32552	82,63	92,69
17	11,23	1	2	2	2,42	2	45	8	5,50	100,00	3,02	6	40625	87,69	95,03
18	11,52	2	1	1	2,57	3	43	10	6,05	88,00	1,44	6	43750	84,18	90,15
19	9,84	4	1	2	2,39	3	43	7	8,40	35,61	1,58	5	26875	84,65	84,12
20	9,05	1	3	1	2,42	3	42	9	5,90	95,36	2,68	5	43125	87,78	89,35
21	8,22	1	4	1	2,30	3	45	8	6,80	90,67	2,31	5	22500	79,18	76,63
22	9,06	5	2	2	1,93	3	38	10	5,38	90,53	1,53	7	25893	74,20	80,73
23	7,69	1	2	2	1,79	3	40	9	5,90	92,55	5,03	6	58333	92,02	95,18
24	6,29	2	2	1	2,53	3	33	8	5,63	95,74	4,00	6	28646	85,54	86,13

Çizelge 4.6. Yüksek derecede ilişkili değişkenler elemine edildikten sonra elde edilen yeni Korelasyon analizi sonuçları

Karakterler	C5	C12	C14	C15	C17	C18	C26	C28	C29	C32	C33	C35	C36	C37	C38
C5	1	0,03	-0,14	0,02	0,21	-0,19	0,46	0,12	0,11	-0,19	-0,1	-0,19	0,06	-0,16	0,07
C12	0,03	1	-0,18	0,24	-0,19	0,14	-0,03	0,16	0,01	-0,31	-0,39	-0,02	0,03	-0,08	-0,16
C14	-0,14	-0,18	1	0,01	0,06	-0,38	0,2	0,03	-0,27	0,28	0	-0,11	0,1	-0,11	0,23
C15	0,02	0,24	0,01	1	0,42	-0,15	0,08	0,19	0,25	-0,2	0,23	0,19	0,44	0,23	0,33
C17	0,21	-0,19	0,06	0,42	1	0,01	0,26	0,28	0,16	-0,14	0,34	-0,1	0,22	-0,02	0,05
C18	-0,19	0,14	-0,38	-0,15	0,01	1	-0,08	0,33	0,3	-0,26	-0,17	-0,17	-0,05	-0,01	-0,4
C26	0,46	-0,03	0,2	0,08	0,26	-0,08	1	0,37	0,02	-0,23	-0,37	-0,29	0,05	-0,47	-0,02
C26	0,12	0,16	0,03	0,19	0,28	0,33	0,37	1	0,19	0,13	-0,06	-0,01	0,41	-0,16	-0,08
C29	0,11	0,01	-0,27	0,25	0,16	0,3	0,02	0,19	1	-0,44	0,12	-0,2	0,06	0,12	-0,45
C32	-0,19	-0,31	0,28	-0,2	-0,14	-0,26	-0,23	0,13	-0,44	1	0,2	0,16	0,21	0,15	0,19
C33	-0,1	-0,39	0	0,23	0,34	-0,17	-0,37	-0,06	0,12	0,2	1	0,02	0,37	0,46	0,2
C35	-0,19	-0,02	-0,11	0,19	-0,1	-0,17	-0,29	-0,01	-0,2	0,16	0,02	1	-0,21	-0,1	0,26
C36	0,06	0,03	0,1	0,44	0,22	-0,05	0,05	0,41	0,06	0,21	0,37	-0,21	1	0,39	0,44
C37	-0,16	-0,08	-0,11	0,23	-0,02	-0,01	-0,47	-0,16	0,12	0,15	0,46	-0,1	0,39	1	0,33
C38	0,07	-0,16	0,23	0,33	0,05	-0,4	-0,02	-0,08	-0,45	0,19	0,2	0,26	0,44	0,33	1

** : p < 0,01

Çizelge 4.7. Korelasyon analizi ile elemine edilmiş değişkenlerin standardize edilmiş sonuçları

Karakterler F.Tipleri	C5	C12	C14	C15	C17	C18	C26	C28	C29	C32	C33	C35	C36	C37	C38
1	0,87	-0,74	1,69	0,83	2,14	-0,85	0,05	0,67	-0,2	0,19	2,16	-0,82	1,17	0,49	0,73
2	-0,28	0,03	1,69	-1,16	-1,23	-2,54	0,29	-2,02	-2,37	0,45	-0,04	0,7	-0,86	-0,96	0,72
3	-0,87	0,03	-1,14	-1,16	-0,45	0,85	-0,92	-0,22	0,6	-0,01	1	-0,82	-1,2	0,67	-2,7
4	0,72	0,03	-1,14	-1,16	-0,76	0,85	-0,92	-0,22	0,35	0,17	0,11	-0,82	0,5	0,08	-1,67
5	-1,11	-0,74	0,75	0,83	-0,75	-0,85	-1,65	-0,22	2,05	0,1	0,43	2,22	-0,48	1,56	-0,07
6	-0,72	0,03	-1,14	0,83	0,29	0,85	-1,41	-1,12	-0,17	0,14	1,15	0,7	0,21	1,64	1,17
7	0,33	2,34	-0,2	0,83	0,81	-0,85	0,78	0,67	-0,06	0,37	-0,9	-0,82	0,39	1,15	-0,44
8	0,24	-0,74	-0,2	0,83	2,26	-0,85	0,05	0,67	0,28	-0,31	1,89	0,7	1,36	-0,25	0,81
9	-1,1	0,8	-0,2	0,83	-0,67	0,85	0,05	1,57	-0,13	0,32	0,44	-0,82	1,32	1	0,96
10	0,11	-0,74	-1,14	0,83	0,64	-0,85	-1,17	-2,02	-0,76	0,22	0,42	0,7	-0,82	0,36	0,58
11	-0,15	0,03	0,75	-1,16	-0,93	0,85	-1,65	-1,12	-0,86	0,44	-0,84	-0,82	-0,17	1,06	0,6
12	1,92	-0,74	-1,14	-1,16	-1,05	-0,85	0,29	-0,22	1,43	0,1	0,64	-0,82	-0,89	-0,52	-0,25
13	-0,78	-0,74	-0,2	0,83	1,39	0,85	1,02	0,67	1,64	-0,13	-0,61	-0,82	0,5	-1,39	-1,55
14	0,57	-0,74	1,69	0,83	0,76	-0,85	1,02	0,67	-0,44	0,11	-0,51	-0,82	-0,02	-0,24	0,55
15	-0,04	1,57	0,75	0,83	-1,32	-0,85	-0,44	-0,22	0,53	0,32	-0,1	-0,82	2,25	0,15	0,54
16	-1,51	-0,74	-0,2	-1,16	-1,35	-0,85	0,05	-0,22	-1,63	0,74	-1,37	0,7	-0,63	-0,83	0,6
17	1,69	-0,74	-0,2	0,83	0,18	-0,85	1,51	-0,22	-0,51	0,74	-0,26	0,7	0,03	0,13	0,89
18	1,9	0,03	-1,14	-1,16	0,36	0,85	1,02	1,57	-0,13	-0,24	-1,35	0,7	0,29	-0,54	0,29
19	0,73	1,57	-1,14	0,83	0,14	0,85	1,02	-1,12	1,5	-4,53	-1,26	-0,82	-1,1	-0,45	-0,45
20	0,18	-0,74	0,75	-1,16	0,18	0,85	0,78	0,67	-0,24	0,36	-0,5	-0,82	0,24	0,14	0,19
21	-0,4	-0,74	1,69	-1,16	0,04	0,85	1,51	-0,22	0,39	-0,03	-0,75	-0,82	-1,46	-1,49	-1,37
22	0,18	2,34	-0,2	0,83	-0,41	0,85	-0,19	1,57	-0,6	-0,04	-1,29	2,22	-1,18	-2,43	-0,86
23	-0,77	-0,74	-0,2	0,83	-0,58	0,85	0,29	0,67	-0,24	0,13	1,13	0,7	1,49	0,95	0,91
24	-1,74	0,03	-0,2	-1,16	0,31	0,85	-1,41	-0,22	-0,42	0,39	0,42	0,7	-0,95	-0,28	-0,2

Çizelge 4.9. Varyansın yüzde 89 (%)’nu açıklayan ana bileşenler

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
BUY,BICI	-0.211	-0.067	-0.164	0.054	0.209	0.157	0.032	-0.042	-0.263	0.138
SUR,UZUN	-0.154	-0.076	0.077	-0.095	0.130	-0.065	-0.267	0.386	-0.013	-0.361
DAL,SAYI	0.099	-0.067	-0.294	0.015	-0.147	-0.115	0.063	-0.146	0.419	0.129
SUR,REN	0.087	-0.228	0.116	-0.109	0.078	-0.041	0.061	-0.377	0.058	0.306
DAL,KALI	-0.119	0.117	0.011	-0.114	0.035	-0.005	-0.575	-0.114	0.113	-0.131
YAP,UZUN	0.004	0.347	0.009	-0.050	-0.024	-0.016	-0.056	-0.163	0.101	-0.074
YAP,GEN	-0.034	0.303	-0.045	-0.033	-0.227	-0.127	-0.076	-0.173	0.095	-0.078
YAP,SAY	0.243	0.126	0.034	-0.201	0.218	-0.094	0.085	-0.022	0.097	0.028
UCYAP,SA	-0.232	0.109	-0.054	0.157	-0.254	0.112	-0.088	-0.230	-0.005	0.101
UYAP,GEN	-0.274	-0.076	-0.083	0.143	-0.181	-0.024	0.076	0.055	0.024	0.037
UYAP,IRI	-0.251	-0.102	-0.075	0.000	-0.164	0.138	0.050	0.079	0.069	-0.085
UYAP,SEK	-0.005	-0.021	0.045	0.333	0.007	-0.379	0.048	-0.322	0.177	-0.269
UYAP,TEP	-0.163	-0.212	-0.124	0.173	0.134	-0.111	-0.053	0.067	0.202	-0.240
UYAP,TAB	-0.025	-0.168	0.101	-0.284	0.187	0.060	0.272	0.064	0.130	0.010
UYAP,KEN	-0.057	-0.064	-0.289	0.045	0.179	-0.394	0.060	-0.039	-0.107	-0.226
YAP,RENG	0.063	0.189	-0.003	0.273	0.101	0.102	-0.387	0.078	0.005	0.289
KARAG,AG	-0.188	0.063	-0.286	-0.166	-0.081	-0.023	0.073	-0.076	-0.074	0.062
KARAG,SE	-0.057	0.157	-0.010	0.365	-0.064	0.237	0.223	0.135	-0.152	-0.175
KARAG,RE	0.046	-0.232	-0.215	-0.074	-0.086	-0.346	-0.045	0.081	-0.010	0.097
KARAG,YO	0.006	0.280	-0.136	0.096	0.333	-0.123	0.087	0.130	-0.001	0.158
KARAG,UZ	-0.194	0.046	-0.278	-0.015	-0.204	0.048	0.176	0.054	-0.177	-0.044
KARAG,GE	-0.195	0.121	-0.228	-0.214	-0.050	-0.130	0.070	-0.071	-0.034	0.059
GOZ,KABA	-0.271	-0.154	0.047	-0.078	0.021	0.026	-0.045	-0.113	-0.074	0.143
GOZ,PAT	-0.252	0.049	0.243	-0.002	0.067	-0.053	0.104	-0.087	-0.009	0.085
CICEK,BA	-0.254	0.046	0.192	-0.143	0.120	-0.052	0.016	-0.067	-0.014	0.049
TAM,CICE	-0.272	0.032	0.160	-0.133	0.130	-0.082	-0.006	-0.059	0.038	0.031
CICEK,SO	-0.280	0.031	0.152	-0.061	0.142	-0.031	0.023	-0.023	0.069	0.093
CICEK,SU	-0.200	-0.039	-0.072	0.253	0.132	0.058	0.032	0.141	0.310	0.186
SALK,SAY	0.012	0.270	-0.137	0.094	0.359	-0.104	0.076	0.161	0.054	0.114
SALK,UZ	0.211	0.155	-0.135	-0.184	-0.059	0.070	-0.002	-0.089	-0.129	-0.171
SAL,CIC,	-0.047	0.266	-0.030	-0.272	-0.045	-0.048	0.194	0.163	-0.014	-0.160
VER,POT	0.053	-0.199	-0.039	-0.133	-0.129	0.223	-0.022	0.254	0.456	-0.067
CI,TOZ,U	0.065	-0.029	-0.349	-0.161	0.072	0.206	-0.026	0.016	0.019	0.156
SALK,AG	-0.165	0.178	-0.179	-0.219	-0.003	0.047	-0.234	0.057	0.211	0.074
ANTER,SA	0.063	-0.142	-0.054	0.049	-0.148	-0.362	-0.179	0.327	-0.216	0.348
ANT,CITO	-0.047	-0.135	-0.211	0.049	0.364	0.183	0.011	-0.203	0.070	-0.179
CANL,OR	0.125	-0.071	-0.250	0.112	0.156	0.259	-0.082	-0.167	-0.043	-0.034
CITO,CIM	-0.006	-0.240	-0.086	-0.128	0.101	0.042	-0.266	-0.159	-0.331	-0.148

Çizelge 4.10. Ana Bileşenler analizi ile elde edilen yeni değişkenlere ilişkin skorlar

Fıstık Tipleri	Değişken Skorları									
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
1	-2,64401	-1,32315	-3,85107	2,3129	0,18242	-0,80059	0,01899	0,62701	0,30834	0,40814
2	3,26686	-4,35467	3,28612	3,64722	-0,41623	1,09391	-0,02804	2,4176	-0,31359	1,26483
3	2,67888	2,9444	-0,21783	-1,81813	-1,07033	-1,11475	0,15801	0,2679	1,58385	0,96986
4	2,76402	2,68992	0,31139	-1,54522	-0,02009	-1,64619	-0,90776	0,72717	1,15563	0,01104
5	6,25388	-0,35645	-1,68793	-0,54264	2,60599	1,37652	0,73521	-1,92143	0,32532	2,30988
6	3,89785	0,54833	-3,33103	-0,09299	-1,25268	0,01288	-0,41957	0,51125	-1,16629	-1,05769
7	-3,28847	-1,64449	-1,48922	-1,32709	0,04215	1,62043	-0,0259	2,29467	1,72997	-0,29006
8	-3,22946	-0,64082	-3,92071	1,0291	0,25862	-0,10826	-0,30707	-0,45551	-0,91808	1,13534
9	-1,40968	-2,25912	-1,22621	-2,71698	0,37108	-1,01108	0,75992	1,81562	0,57157	-0,0107
10	2,85476	0,72947	-2,30912	1,66517	-2,18978	0,87487	-1,2121	-0,03421	-1,1517	-0,94701
11	3,75091	-1,10454	1,30131	-0,56003	-0,6966	-2,12634	-0,10506	0,18242	-0,60337	-1,63984
12	1,44957	4,51302	0,65218	1,61706	1,35822	-0,2941	-2,29375	-0,87666	1,33863	-0,11789
13	-2,71932	4,12227	-0,87182	1,63351	-0,27191	0,9533	2,98621	0,29448	0,56817	-0,31311
14	-2,76079	-0,47103	1,4846	0,95162	0,8535	-0,51897	-0,5051	0,47971	0,74451	-0,2378
15	3,73948	-1,4999	1,2747	-0,1871	3,957	0,9563	0,77733	0,35321	0,46227	-1,97725
16	0,40451	-3,71928	3,86204	-0,56052	-1,54587	-0,07245	-0,23366	-0,96004	-0,55876	0,76461
17	-3,96709	-1,85867	-0,03132	1,29732	0,83547	0,82984	-2,44551	-1,84195	0,38764	-1,25491
18	-3,34349	1,89672	1,75119	-1,26874	-0,1325	-0,5264	-2,33307	0,77225	-0,59224	1,63822
19	-1,95306	5,00069	1,87361	-1,77133	1,30392	1,82549	0,17446	1,00753	-3,51078	-0,08237
20	-3,26736	-1,29349	1,01427	-0,16053	0,37992	-2,55437	0,98825	-1,18399	-0,59825	-0,26086
21	-2,36511	3,01331	2,96774	2,34032	-0,7563	-0,53634	2,14114	-0,94856	0,73011	-0,05697
22	-1,80424	-1,61134	0,80754	-2,81175	-2,4381	3,12698	0,09569	-1,6669	1,24716	-0,48194
23	-1,50944	-3,5523	-1,20695	-1,32246	1,27263	-1,21822	0,9947	-1,19946	-1,36833	0,20137
24	3,20081	0,2371	-0,4435	0,19128	-2,63053	-0,14246	0,98667	-0,66231	-0,37177	0,02512

Çizelge 4.11. Korelasyon analizi ile elemine edilen değişkenlere ilişkin Öklid uzaklık matrisi

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	,00	6,75	6,80	6,04	6,22	5,34	5,24	2,73	4,97	5,40	6,04	5,94	5,57	3,48	5,20	6,95	4,58	6,18	8,28	4,57	6,32	7,85	4,51	5,96
2	6,75	,00	7,22	6,63	6,76	6,64	6,60	6,90	7,06	5,08	5,19	6,17	7,80	5,25	5,97	3,79	5,21	7,01	8,68	5,70	5,84	7,10	6,54	5,57
3	6,80	7,22	,00	2,79	5,62	5,14	5,64	6,46	5,53	5,33	4,84	4,63	5,05	6,30	6,20	5,78	6,45	6,03	6,82	4,81	4,95	6,63	5,58	3,68
4	6,04	6,63	2,79	,00	5,73	4,73	4,82	5,67	4,66	4,85	3,83	3,49	4,62	5,35	4,68	5,16	5,13	4,33	6,43	3,65	4,85	6,02	4,69	3,97
5	6,22	6,76	5,62	5,73	,00	4,35	6,18	5,54	5,56	4,76	5,42	5,90	6,23	5,85	5,47	5,83	5,71	6,87	7,89	5,84	6,65	6,89	4,66	4,70
6	5,34	6,64	5,14	4,73	4,35	,00	5,29	4,53	4,06	3,01	4,12	5,74	6,10	5,62	4,97	5,52	4,94	5,99	7,20	5,02	6,87	6,88	3,25	3,78
7	5,24	6,60	5,64	4,82	6,18	5,29	,00	5,22	3,93	5,40	5,24	5,52	4,99	4,07	3,85	5,83	4,37	4,78	6,30	4,38	5,83	5,61	5,03	5,51
8	2,73	6,90	6,46	5,67	5,54	4,53	5,22	,00	4,79	4,58	6,37	5,58	4,86	4,20	5,28	6,37	4,20	5,35	7,56	4,81	6,50	6,90	3,81	5,26
9	4,97	7,06	5,53	4,66	5,56	4,06	3,93	4,79	,00	5,62	4,59	5,82	5,06	4,46	3,44	5,32	4,96	5,11	7,26	3,77	6,11	6,17	2,49	4,67
10	5,40	5,08	5,33	4,85	4,76	3,01	5,40	4,58	5,62	,00	4,41	4,90	5,98	4,96	5,49	4,67	3,96	5,85	7,00	5,14	6,20	6,60	4,69	4,04
11	6,04	5,19	4,84	3,83	5,42	4,12	5,24	6,37	4,59	4,41	,00	5,21	6,26	4,92	4,56	4,14	5,25	5,50	7,32	3,57	5,20	6,71	4,74	3,60
12	5,94	6,17	4,63	3,49	5,90	5,74	5,52	5,58	5,82	4,90	5,21	,00	5,32	4,97	5,44	5,46	4,21	4,44	6,44	4,25	4,94	6,73	5,52	5,34
13	5,57	7,80	5,05	4,62	6,23	6,10	4,99	4,86	5,06	5,98	6,26	5,32	,00	4,36	5,60	6,09	5,27	4,96	6,13	4,06	3,99	5,91	5,08	5,14
14	3,48	5,25	6,30	5,35	5,85	5,62	4,07	4,20	4,46	4,96	4,92	4,97	4,36	,00	4,50	4,86	3,03	4,65	6,92	2,95	4,17	6,14	4,40	5,29
15	5,20	5,97	6,20	4,68	5,47	4,97	3,85	5,28	3,44	5,49	4,56	5,44	5,60	4,50	,00	5,52	4,88	5,80	7,10	4,65	6,27	6,40	4,16	5,57
16	6,95	3,79	5,78	5,16	5,83	5,52	5,83	6,37	5,32	4,67	4,14	5,46	6,09	4,86	5,52	,00	4,75	5,22	7,96	4,16	4,90	5,68	5,01	3,78
17	4,58	5,21	6,45	5,13	5,71	4,94	4,37	4,20	4,96	3,96	5,25	4,21	5,27	3,03	4,88	4,75	,00	3,87	7,04	3,80	5,35	6,01	4,12	5,54
18	6,18	7,01	6,03	4,33	6,87	5,99	4,78	5,35	5,11	5,85	5,50	4,44	4,96	4,65	5,80	5,22	3,87	,00	6,39	3,45	5,21	4,98	5,01	5,36
19	8,28	8,68	6,82	6,43	7,89	7,20	6,30	7,56	7,26	7,00	7,32	6,44	6,13	6,92	7,10	7,96	7,04	6,39	,00	6,81	6,59	6,98	7,48	7,29
20	4,57	5,70	4,81	3,65	5,84	5,02	4,38	4,81	3,77	5,14	3,57	4,25	4,06	2,95	4,65	4,16	3,80	3,45	6,81	,00	3,34	6,03	3,78	3,97
21	6,32	5,84	4,95	4,85	6,65	6,87	5,83	6,50	6,11	6,20	5,20	4,94	3,99	4,17	6,27	4,90	5,35	5,21	6,59	3,34	,00	5,94	6,03	4,70
22	7,85	7,10	6,63	6,02	6,89	6,88	5,61	6,90	6,17	6,60	6,71	6,73	5,91	6,14	6,40	5,68	6,01	4,98	6,98	6,03	5,94	,00	6,44	5,36
23	4,51	6,54	5,58	4,69	4,66	3,25	5,03	3,81	2,49	4,69	4,74	5,52	5,08	4,40	4,16	5,01	4,12	5,01	7,48	3,78	6,03	6,44	,00	4,39
24	5,96	5,57	3,68	3,97	4,70	3,78	5,51	5,26	4,67	4,04	3,60	5,34	5,14	5,29	5,57	3,78	5,54	5,36	7,29	3,97	4,70	5,36	4,39	,00

Çizelge 4.12. Korelasyon analizi ile elemine edilen değişkenlerin Öklid uzaklığının karesine (Öklid Kare) ilişkin uzaklık matrisi

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0	45,62	46,30	36,44	38,69	28,53	27,45	7,43	24,66	29,18	36,54	35,28	31,00	12,11	27,02	48,29	21,01	38,23	68,62	20,92	39,89	61,68	20,30	35,46
2	45,62	0	52,10	43,90	45,66	44,14	43,56	47,60	49,83	25,83	26,92	38,05	60,76	27,55	35,60	14,34	27,16	49,11	75,28	32,51	34,09	50,38	42,80	31,07
3	46,30	52,10	0	7,79	31,60	26,45	31,82	41,67	30,56	28,41	23,43	21,44	25,53	39,69	38,39	33,46	41,58	36,37	46,49	23,13	24,55	43,92	31,14	13,54
4	36,44	43,90	7,79	0	32,78	22,41	23,26	32,12	21,71	23,55	14,65	12,20	21,31	28,63	21,87	26,60	26,33	18,78	41,35	13,31	23,53	36,23	22,02	15,77
5	38,69	45,66	31,60	32,78	0	18,89	38,21	30,67	30,94	22,65	29,43	34,75	38,80	34,16	29,95	33,95	32,65	47,13	62,31	34,07	44,20	47,44	21,76	22,05
6	28,53	44,14	26,45	22,41	18,89	0	27,95	20,51	16,46	9,07	17,01	32,90	37,22	31,57	24,74	30,52	24,38	35,91	51,79	25,18	47,26	47,37	10,57	14,26
7	27,45	43,56	31,82	23,26	38,21	27,95	0	27,29	15,44	29,21	27,48	30,48	24,86	16,57	14,82	33,98	19,06	22,85	39,65	19,21	33,93	31,51	25,29	30,39
8	7,43	47,60	41,67	32,12	30,67	20,51	27,29	0	22,98	20,99	40,53	31,14	23,66	17,62	27,88	40,62	17,67	28,61	57,17	23,10	42,28	47,65	14,53	27,62
9	24,66	49,83	30,56	21,71	30,94	16,46	15,44	22,98	0	31,55	21,09	33,82	25,56	19,88	11,84	28,34	24,64	26,08	52,67	14,24	37,33	38,07	6,21	21,83
10	29,18	25,83	28,41	23,55	22,65	9,07	29,21	20,99	31,55	0	19,42	24,06	35,75	24,56	30,18	21,85	15,67	34,26	48,98	26,41	38,50	43,53	22,01	16,32
11	36,54	26,92	23,43	14,65	29,43	17,01	27,48	40,53	21,09	19,42	0	27,19	39,23	24,17	20,79	17,13	27,60	30,27	53,64	12,77	27,07	44,99	22,50	12,96
12	35,28	38,05	21,44	12,20	34,75	32,90	30,48	31,14	33,82	24,06	27,19	0	28,31	24,68	29,58	29,78	17,76	19,75	41,47	18,07	24,45	45,26	30,48	28,50
13	31,00	60,76	25,53	21,31	38,80	37,22	24,86	23,66	25,56	35,75	39,23	28,31	0	19,03	31,36	37,11	27,74	24,55	37,63	16,46	15,96	34,97	25,80	26,38
14	12,11	27,55	39,69	28,63	34,16	31,57	16,57	17,62	19,88	24,56	24,17	24,68	19,03	0	20,26	23,61	9,19	21,60	47,92	8,70	17,43	37,76	19,39	27,94
15	27,02	35,60	38,39	21,87	29,95	24,74	14,82	27,88	11,84	30,18	20,79	29,58	31,36	20,26	0	30,51	23,84	33,63	50,43	21,59	39,37	40,93	17,27	31,03
16	48,29	14,34	33,46	26,60	33,95	30,52	33,98	40,62	28,34	21,85	17,13	29,78	37,11	23,61	30,51	0	22,59	27,27	63,32	17,32	24,03	32,28	25,15	14,26
17	21,01	27,16	41,58	26,33	32,65	24,38	19,06	17,67	24,64	15,67	27,60	17,76	27,74	9,19	23,84	22,59	0	15,00	49,57	14,46	28,63	36,16	16,98	30,69
18	38,23	49,11	36,37	18,78	47,13	35,91	22,85	28,61	26,08	34,26	30,27	19,75	24,55	21,60	33,63	27,27	15,00	0	40,82	11,88	27,10	24,81	25,08	28,75
19	68,62	75,28	46,49	41,35	62,31	51,79	39,65	57,17	52,67	48,98	53,64	41,47	37,63	47,92	50,43	63,32	49,57	40,82	0	46,44	43,40	48,77	55,93	53,18
20	20,92	32,51	23,13	13,31	34,07	25,18	19,21	23,10	14,24	26,41	12,77	18,07	16,46	8,70	21,59	17,32	14,46	11,88	46,44	0	11,16	36,31	14,26	15,72
21	39,89	34,09	24,55	23,53	44,20	47,26	33,93	42,28	37,33	38,50	27,07	24,45	15,96	17,43	39,37	24,03	28,63	27,10	43,40	11,16	0	35,33	36,35	22,12
22	61,68	50,38	43,92	36,23	47,44	47,37	31,51	47,65	38,07	43,53	44,99	45,26	34,97	37,76	40,93	32,28	36,16	24,81	48,77	36,31	35,33	0	41,45	28,76
23	20,30	42,80	31,14	22,02	21,76	10,57	25,29	14,53	6,21	22,01	22,50	30,48	25,80	19,39	17,27	25,15	16,98	25,08	55,93	14,26	36,35	41,45	0	19,31
24	35,46	31,07	13,54	15,77	22,05	14,26	30,39	27,62	21,83	16,32	12,96	28,50	26,38	27,94	31,03	14,26	30,69	28,75	53,18	15,00	22,12	28,76	19,31	0

Çizelge 4.13. Ana Bileşenler analizi ile boyutu indirgenen değişkenler ilişkin fıstık tipleri arasındaki Öklid Uzaklıkları

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0	10,25	8,98	8,95	10,69	7,72	5,52	2,5	6,07	7,18	9,14	9,06	7,2	5,7	9,85	9,47	5,97	7,94	10,63	6,23	8,59	8,83	5,91	8,02
2	10,25	0	10,46	10,03	9,87	9,69	10,25	11,2	9,78	8,8	7,51	10,9	12,08	8,57	7,89	6,41	10,37	10,97	12,79	10,1	10,46	10,56	9,64	8,09
3	8,98	10,46	0	2,07	7,46	5,77	8,54	8,91	7,24	6,31	5,85	5,63	7,65	7,65	8,1	8,68	10,1	7,38	8,48	8,42	7,78	8,29	8,87	4,62
4	8,95	10,03	2,07	0	7,63	5,68	8,54	9,02	7,1	6,24	4,97	4,84	8,14	7,11	7,16	8,47	9,6	7,01	8,18	8,12	7,91	8,76	8,67	5,17
5	10,69	9,87	7,46	7,63	0	6,84	11,27	10,55	9,87	7,83	7,78	8,86	11,46	10,69	6,49	10,03	11,8	11,77	11,89	11,29	11,8	10,75	9,37	6,99
6	7,72	9,69	5,77	5,68	6,84	0	8,83	7,87	7,63	2,8	5,49	7,75	9,09	8,87	7,64	9,4	9,88	9,71	10,04	9,42	10,23	9,16	8,16	4,01
7	5,52	10,25	8,54	8,54	11,27	8,83	0	5,67	3,9	8,61	9,19	9,72	7,6	5,04	9,11	8,34	6,04	6,67	9,4	6,62	8,8	5,84	6,29	8,58
8	2,5	11,2	8,91	9,02	10,55	7,87	5,67	0	6,15	7,37	9,59	9,04	7,05	5,96	10,3	9,46	5,69	7,09	9,5	6,04	8,65	8,17	5,47	8,16
9	6,07	9,78	7,24	7,1	9,87	7,63	3,9	6,15	0	8,29	6,99	9,75	8,55	5,46	7,91	7,15	7,29	6,87	9,49	5,37	9,11	6,56	4,18	7,31
10	7,18	8,8	6,31	6,24	7,84	2,8	8,61	7,37	8,29	0	5,95	6,91	8,33	8	8,22	8,67	8,66	8,94	9,65	9,01	9,01	8,51	8,45	3,75
11	9,14	7,51	5,85	4,97	7,78	5,49	9,19	9,59	6,99	5,95	0	7,85	10,05	7,36	5,79	6,05	9,34	8,89	10,14	7,47	8,87	8,68	7,23	4,26
12	9,06	10,9	5,63	4,84	8,86	7,75	9,72	9,04	8,75	6,91	7,85	0	7,35	7,03	8,28	10,05	8,67	7,15	7,9	8,94	6,87	10,02	10,25	7,37
13	7,2	12,08	7,65	8,14	11,46	9,09	7,6	7,05	8,55	8,33	10,05	7,35	0	6,54	10,33	10,75	8,62	7,52	6,99	7,53	4,63	8,85	9,3	8,07
14	5,7	8,57	7,65	7,11	10,69	8,87	5,04	5,96	5,46	8	7,36	7,03	6,54	0	7,82	6,27	4,24	4,54	7,92	3,68	5,3	6,8	5,82	7,58
15	9,85	7,89	8,1	7,16	6,49	7,64	9,11	10,3	7,91	8,22	5,79	8,28	10,33	7,82	0	8,1	9,43	10,3	10,27	8,99	9,93	9,53	7,86	7,54
16	9,47	6,41	8,68	8,47	10,03	9,4	8,34	9,46	7,15	8,67	6,05	10,05	10,75	6,27	8,1	0	7,64	7,81	10,59	6,33	8,43	6,34	6,46	6,78
17	5,97	10,37	10,1	9,6	11,8	9,88	6,04	5,69	7,29	8,66	9,34	8,67	8,62	4,24	9,43	7,64	0	6,57	9,84	5,47	8,01	6,79	6,24	9,25
18	7,94	10,97	7,38	7,01	11,77	9,71	6,67	7,09	6,87	8,94	8,89	7,15	7,52	4,54	10,3	7,81	6,57	0	8,1	5,9	6,69	7,51	7,87	8,64
19	10,63	12,79	8,48	8,18	11,89	10,04	9,4	9,5	9,49	9,65	10,14	7,9	6,99	7,92	10,27	10,59	9,84	6,1	0	8,87	7,6	9,57	10,13	9,55
20	6,23	10,1	8,42	8,12	11,29	9,42	6,62	6,04	5,37	9,01	7,47	8,94	7,53	3,68	8,99	6,33	5,47	5,9	8,87	0	6,16	7,35	4,23	7,86
21	8,59	10,46	7,78	7,91	11,8	10,23	8,8	8,65	9,11	9,01	8,87	6,87	4,63	5,3	9,93	8,43	8,01	6,69	7,6	6,16	0	8,62	9,23	7,83
22	8,83	10,56	8,29	8,76	10,75	9,16	5,84	8,17	6,56	8,51	8,68	10,02	8,85	6,8	9,53	6,34	6,79	7,51	9,57	7,35	8,62	0	7,15	7,38
23	5,91	9,64	8,87	8,67	9,37	8,16	6,29	5,47	4,18	8,45	7,23	10,25	9,3	5,82	7,86	6,46	6,24	7,87	10,13	4,23	9,23	7,15	0	7,56
24	8,02	8,09	4,62	5,17	6,99	4,01	8,58	8,16	7,31	3,75	4,26	7,37	8,07	7,58	7,54	6,78	9,25	8,64	9,55	7,86	7,83	7,38	7,56	0

Çizelge 4.14.-1. Ana Bileşenler analizi ile boyutu indirgenen değişkenler ilişkin fıstık tipleri arasındaki Öklid Kare uzaklıkları

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	105,13	80,57	80,02	114,1	59,61	30,52	6,26	36,88	51,60	83,51	82,17
2	105,13	0	109,42	100,63	97,33	93,99	104,96	125,40	95,57	77,46	56,41	118,70
3	80,57	109,42	0	4,26	55,69	33,27	72,99	79,39	52,45	39,78	34,27	31,67
4	80,02	100,63	4,26	0	58,15	32,31	72,99	81,28	50,45	38,98	24,73	23,46
5	114,18	97,33	55,69	58,15	0	46,84	127,09	111,35	97,40	61,39	60,53	78,47
6	59,61	93,99	33,27	32,31	46,84	0	77,94	61,94	58,26	7,82	30,18	60,01
7	30,52	104,96	72,99	72,99	127,09	77,94	0	32,20	15,21	74,10	84,54	94,40
8	6,26	125,40	79,39	81,28	111,35	61,94	32,20	0	37,88	54,25	91,97	81,74
9	36,88	95,57	52,45	50,46	97,40	58,26	15,21	37,88	0	68,77	48,84	95,01
10	51,60	77,46	39,78	38,98	61,39	7,82	74,11	54,25	68,77	0	35,44	47,79
11	83,51	56,41	34,27	24,73	60,53	30,18	84,54	91,97	48,84	35,44	0	61,59
12	82,17	118,70	31,67	23,46	78,47	60,01	94,40	81,74	95,01	47,79	61,59	0
13	51,78	145,94	58,46	66,26	131,36	82,64	57,69	49,73	73,12	69,43	101,08	53,95
14	32,50	73,50	58,53	50,58	114,21	78,71	25,42	35,56	29,77	63,94	54,14	49,42
15	97,00	62,29	65,66	51,23	42,11	58,43	82,99	106,01	62,49	67,59	33,51	68,53
16	89,76	41,03	75,41	71,76	100,70	88,36	69,54	89,57	51,09	75,24	36,64	101,02
17	35,69	107,55	102,00	92,16	139,12	97,60	36,47	32,35	53,08	74,97	87,15	75,20
18	63,12	120,24	54,41	49,07	138,63	94,34	44,44	50,20	47,19	79,97	78,98	51,16
19	113,07	163,69	71,95	66,95	141,48	100,83	88,29	90,28	90,09	93,09	102,86	62,45
20	38,78	102,08	70,83	65,99	127,46	88,69	43,85	36,46	28,80	81,25	55,84	79,96
21	73,71	109,46	60,52	62,63	139,28	104,73	77,42	74,86	82,95	81,22	78,69	47,26
22	77,99	111,51	68,72	76,79	115,51	83,97	34,08	66,75	42,98	72,42	75,29	100,32
23	34,97	92,91	78,63	75,10	87,74	66,52	39,55	29,91	17,44	71,48	52,33	105,11
24	64,26	65,43	21,34	26,69	48,84	16,08	73,59	66,51	53,41	14,03	18,12	54,27

Çizelge 4.14.-2. Ana Bileşenler analizi ile boyutu indirgenen değişkenler ilişkin fıstık tipleri arasındaki Öklid Kare uzaklıkları

	13	14	15	16	17	18	19	20	21	22	23	24
1	51,78	32,50	97,00	89,76	35,69	63,12	113,07	38,78	73,71	77,99	34,97	64,26
2	145,94	73,50	62,29	41,03	107,55	120,24	163,69	102,08	109,46	111,51	92,91	65,43
3	58,46	58,53	65,66	75,41	102,00	54,41	71,95	70,83	60,52	68,72	78,63	21,34
4	66,26	50,58	51,23	71,76	92,16	49,07	66,95	65,99	62,63	76,79	75,10	26,69
5	131,36	114,21	42,11	100,70	139,12	138,63	141,48	127,46	139,28	115,51	87,74	48,84
6	82,64	78,71	58,43	88,36	97,60	94,34	100,83	88,69	104,73	83,97	66,52	16,08
7	57,69	25,42	82,99	69,54	36,47	44,44	88,29	43,85	77,42	34,08	39,55	73,59
8	49,73	35,56	106,01	89,57	32,35	50,20	90,28	36,46	74,86	66,75	29,91	66,51
9	73,12	29,77	62,49	51,09	53,08	47,19	90,09	28,80	82,95	42,98	17,44	53,41
10	69,43	63,94	67,59	75,24	74,97	79,97	93,09	81,26	81,22	72,42	71,48	14,03
11	101,08	54,14	33,51	36,64	87,15	78,98	102,86	55,84	78,69	75,29	52,33	18,12
12	53,95	49,42	68,53	101,02	75,20	51,16	62,45	79,96	47,26	100,32	105,11	54,27
13	0	42,81	106,79	115,52	74,38	56,53	48,84	56,68	21,40	78,40	86,53	65,08
14	42,81	0	61,23	39,32	17,93	20,65	62,73	13,56	28,09	46,25	33,87	57,46
15	106,79	61,23	0	65,53	88,92	106,06	105,49	80,81	98,65	90,86	61,70	56,79
16	115,52	39,32	65,53	0	58,31	60,92	112,21	40,10	71,03	40,18	41,76	45,92
17	74,38	17,93	88,92	58,31	0	43,21	96,92	29,87	64,17	46,15	38,99	85,51
18	56,53	20,65	106,06	60,92	43,21	0	37,25	34,79	44,70	56,34	61,89	74,65
19	48,84	62,73	105,49	112,21	96,92	37,25	0	78,69	57,73	91,61	102,52	91,14
20	56,68	13,56	80,81	40,10	29,87	34,79	78,69	0	37,99	54,02	17,87	61,71
21	21,40	28,09	98,65	71,03	64,17	44,70	57,73	37,99	0	74,31	85,11	61,25
22	78,40	46,25	90,86	40,18	46,15	56,34	91,61	54,02	74,31	0	51,12	54,46
23	86,53	33,87	61,70	41,76	38,99	61,89	102,52	17,87	85,11	51,12	0	57,12
24	65,08	57,46	56,79	45,92	85,51	74,65	91,14	61,71	61,25	54,46	57,13	0

8. ÖZET

Bu çalışma ile, sistematik konusunda görsel olarak veya geleneksel yöntemleri kullanarak sınıflandırma yapan araştırmacılar için, özellikle ($n < p$) boyutlu biyolojik verilerde istatistiksel temele dayanan en uygun boyut indirgeme yöntemi ile kümeleme yöntem ve kriterlerinin ortaya konulması amaçlanmıştır. Araştırma, çeşit ve türleri saptanma safasında olan 24 farklı Antepfıstığı (*Pistacia vera L.*) tipi ve bu tiplerden alınan 38 farklı değişken üzerinde yapılmıştır.

Çalışmada araştırmacıların, her bir bireye ilişkin çok sayıda değişken (karakter) belirlemeleri durumunda, en anlamlı karakterleri kullanmaları için farklı şekillerde boyut indirgeme yöntemleri karşılaştırılmış. Bu karşılaştırmadan en anlamlı boyut indirgeyen yöntemin Ana Bileşenler analizi olduğu belirlenmiştir. Sözü edilen yöntem; uzayda farklı açılarla tanımlı olan gerçek değişkenleri, dik açılarla tanımlı skorlara dönüştürmekte ve varyasyonun büyük bir kısmını açıklamaya yeterli olan daha az sayıdaki değişkenleri belirlemeye yardımcı olmaktadır. Bu yöntemle göre fıstık tiplerinden elde edilen 38 adet değişken yerine, (Çizelge 4.8'de verilen) varyasyonun %89.2'sini açıklayan ilk 10 adet Ana Bileşen skorları belirlenmiş ve bu Ana Bileşenler kümeleme analizinde gerçek değişkenlerin yerine kullanılmıştır.

Sözü edilen değişkenlerle, literatürlerde önemli görülen bazı kümeleme yöntemleri karşılaştırılmış ve en uygun küme yapısını ortaya koyan kümeleme yöntemin Ward yöntemi olduğu tespit edilmiştir. Aynı zamanda, bazı bölümlenme ölçütleri ve küme sayısını saptayan kriterler karşılaştırılarak en uygun ölçüt ve kriterler saptanmıştır. Bunlar sırasıyla, C_{max} ölçütü ile Wilks Lambda ve Hotelling Lawley İz istatistikleridir. Bu ölçütler, fıstık tiplerinin iki kümeye ayrılması gerektiğini göstermektedir. Bu duruma göre (3, 4, 5, 6, 10, 11, 12, 15, 24) No'lu fıstık tipleri erkenci fıstık tipleri olarak bir kümede kümelenecek olmakla birlikte, orta geççi bir fıstık tipi olan 12 No'lu fıstık tipi diğer bazı özellikler bakımından erkenci çeşitlere daha çok benzerlik gösterdiğinden onlarla birlikte kümelenecektir. (1, 2, 7, 8, 9, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23) No'lu orta geççi ve geççi fıstık tipleri ise ikinci bir küme oluşturmuştur.

9. SUMMARY

The objective of this study is to put forward the method of dimension increasing and criteria and method for clustering based on statistical foundation especially on the biological data with ($n < p$) dimension, for the researchers who work on taxonomy visual or traditional methods. The research has been carried out on 38 different variables taken from 24 different pistachio (*Pistacia vera* L.) types at the stage of determining the species and varieties.

In the study the dimension reduction methods were compared in different ways for the researchers who determine a large number of variables (Characteristics) related to each individual to use the most suitable characteristics. In this comparison, it has been determined that the most suitable method is the Principal Component analysis method. This method make it possible to determine the smaller number of variables which are sufficient to explain the most of the variation by transferring the real variables described by different angles in space into scores described by right angles. According to this method, instead of 38 variables (shown in table 4.8) which were obtained from types of pistachios, the first 10 principle component scores which can explain 89.2 % of the variation in the principle component clustering analysis were determined and used in place of real variables.

Above mentioned method was compared with some important clustering methods in literature and it was concluded that the Ward clustering method was the most suitable method to determine the structure of the population. At the same time, the most suitable criteria were determined by comparing some dividing criteria with those that determine the number of clusters. These are C_{max} criteria, Wilks Lambda and Hotelling Lawley Trace statistics respectively. These criteria show that the pistachio types need to divide into two clusters. Accordingly, number 3, 4, 5, 6, 10, 11, 12, 15 and 24 are clustered in the same group as early growing variety. Number 12 show some similarities to medium growing pistachio varieties. But it has many common characteristics with the early growing types. So it was clustered with early growing group. The numbers 1, 2, 7, 8, 9, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, constituted the second cluster as medium late and late growing pistachio types.