



**İÇ GÖÇ VERİLERİNİ KULLANARAK KÜMELEME ANALİZİNDE ELDE
EDİLEN KÜMELERİN GEÇERLİLİĞİNİN SINANMASI**

Hasan Gürcan ÖZGEN

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

HAZİRAN 2019

Hasan Gürçan ÖZGEN tarafından hazırlanan “İÇ GÖÇ VERİLERİNİ KULLANARAK KÜMELEME ANALİZİNDE ELDE EDİLEN KÜMELERİN GEÇERLİLİĞİNİN SINANMASI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İstatistik Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Semra ERBAŞ

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Başkan: Prof. Dr. Jülide YILDIRIM ÖCAL

Ekonomi Ana Bilim Dalı, TED Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Üye: Doç. Dr. Hülya OLMUŞ

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Tez Savunma Tarihi: 24/06/2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....
Prof. Dr. Sena YAŞYERLİ
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
 - Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
 - Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
 - Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
 - Bu tezde sunduğum çalışmanın özgün olduğunu,
- bildirim, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Hasan Gürcan ÖZGEN

24/06/2019

İÇ GÖÇ VERİLERİNİ KULLANARAK KÜMELEME ANALİZİNDE ELDE EDİLEN KÜMELERİN GEÇERLİLİĞİNİN SINANMASI

(Yüksek Lisans Tezi)

Hasan Gürcan ÖZGEN

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2019

ÖZET

Kümeleme analizinde, anlamlı ve geçerli sonuçlara ulaşabilmek birçok araştırmacının hem hedefi hem de karşılaştığı en önemli sorunlardan biridir. Bu bakımdan elde edilen kümeleme çözümlerinin kalitesini test etmek amacıyla geliştirilen bazı yöntemler bulunmaktadır. Bu yöntemlerden biri de, çalışmada incelenen çok değişkenli bir istatistik analiz tekniği olan diskriminant analizinin kümeleme sonuçlarına uygulanmasıdır. Bu çalışmada, SPSS Paket Programı ve Minitab Paket Programı kullanılarak Türkiye İstatistik Kurumu 2017 yılı 15 yaş ve üzeri iç göç verileri üzerinde hiyerarşik kümeleme analizi tekniklerinden tek bağlantı kümeleme tekniği, tam bağlantı kümeleme tekniği, merkezi bağlantı kümeleme tekniği, medyan bağlantı kümeleme tekniği ve Ward tekniklerinden her birine karesi alınmış Öklid, Pearson, City-Block(Manhattan) ve Minkowski uzaklıkları uygulanarak elde edilen kümeler karşılaştırılmış ve ortaya çıkan bu kümelerin her biri için varyans-kovaryans matrislerinin eşitliğine göre doğrusal veya karesel diskriminant analizi uygulanarak meydana gelen kümelerin geçerliliği sorgulanmıştır. Bu çerçevede söz konusu uygulama özelinde, hangi uzaklık ölçüsünün hangi kümeleme tekniğinde nasıl kümeler oluşturduğu incelenmiş ve doğrusal veya karesel diskriminant analizi gerçekleştirilerek hangi oranda doğru sınıflandırma ortaya koyduğu belirlenmeye çalışılmıştır. Gerçekleştirilen analiz sonuçlarına göre, tek bağlantı kümeleme tekniğinde tüm uzaklıkların aynı oranda doğru sınıflandırma yaptığı, tam bağlantı kümeleme tekniği ve medyan bağlantı kümeleme tekniğinde Pearson uzaklığına göre elde edilen kümelerin sırasıyla % 97,5 ve % 98,8 oranında en fazla doğru sınıflandırma oranına sahip olduğu, merkezi bağlantı kümeleme tekniğinde karesi alınmış Öklid uzaklığının % 98,8 ile en fazla oranda doğru sınıflandırma gösterdiği ve Ward tekniğinde ise Pearson ve Manhattan uzaklıklarının % 96,3 ile en fazla doğru sınıflandırma oranına sahip oldukları belirlenmiştir. Bununla birlikte, Türkiye’de 81 il bazında illere göç edenlerin eğitim durumu konusunda durum tespiti yapılmış olup elde edilen bu sınıflandırmaların ilgili politika, uygulama ve diğer akademik çalışmalara yol gösterici nitelikte olacağı düşünülmektedir.

Bilim Kodu : 20512

Anahtar Kelimeler : Benzerlik Katsayıları, Kümeleme Analizi, Diskriminant Analizi, Göç, İl, Eğitim Durumu

Sayfa Adedi : 93

Danışman : Prof. Dr. Semra ERBAŞ

TESTING THE VALIDITY OF CLUSTERS USED IN CLUSTER ANALYSIS USING INTERNAL MIGRATION

(M. Sc. Thesis)

Hasan Gürcan ÖZGEN

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

June 2019

ABSTRACT

In cluster analysis, finding significant and valid results is both the goal and one of the most important problems of many researchers. In respect to this, there are some methods developed to test the quality of resulting cluster analyses. One of these methods examined in our study is to apply a discriminant analysis, which is a multivariate statistical analysis technique, to the clustering results. In this study, SPSS Software Package and Minitab Software Package were used to apply Euclidean Squared, Pearson's, City-Block (Manhattan) and Minkowski distances on each of these particular hierarchical cluster analysis methods, consisting of single linkage clustering method, complete linkage clustering method, centroid linkage clustering method, median linkage clustering method and Ward's method, using the Turkish Statistical Institute data on internal migration of people aged 15 and older, to compare the resulting clusters, and to question the validity of the resulting clusters by applying linear or quadratic discriminant analysis depending on the equality of the variance-covariance matrices. Within the scope of the aforementioned operation, it was examined which distance measurement created what kind of cluster in which clustering method, and it was attempted to be determined that at what rate the accurate classifications were done. According to the results of the analysis, in the single linkage clustering method all distances had the same rate of accurate classification, in the complete linkage clustering method and the median linkage clustering method the clusters which are obtained according to the Pearson's distance had the most accurate classification rate of 97.5% and 98.8% respectively, in the centroid linkage clustering method the Euclidean Squared distance had the most accurate classification rate with 98.8%, and in Ward's method Pearson's and Manhattan distances had the most accurate classification rate with 96.3%. In addition to this, educational statuses of immigrants throughout 81 provinces of Turkey were determined on the basis of provinces, and it is believed that the resulting classifications will provide guidance to politics, practical applications, and other academic studies.

Science Code : 20512

Key Words : Similarity Coefficients, Cluster Analysis, Discriminant Analysis, Immigration, Province, Educational Status

Page Number : 93

Supervisor : Prof. Dr. Semra ERBAŞ

TEŐEKKÖR

Çalıőmalarımın her aőamasında bilgi, öneri ve tecrübeleri ile beni yönlendirerek çalıőmalarıma katkı ve yardımlarını sunan çok deęerli danıőman hocam Sayın Prof. Dr. Semra ERBAŐ'a ve tüm eęitim hayatım boyunca emek ve katkılarını hiçbir zaman esirgemeyen aileme teőekkürlerimi sunarım.



İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	xiii
SİMGELER VE KISALTMALAR.....	xiv
1. GİRİŞ.....	1
2. KÜMELEME ANALİZİ	5
2.1. Kümeleme Analizinin Tanımı ve Genel Amacı.....	5
2.2. Kümeleme Analizinde Kullanılan Benzerlik Ölçütleri	8
2.2.1. Öklid uzaklığı.....	8
2.2.2. Karesi alınmış Öklid uzaklığı	9
2.2.3. Pearson uzaklığı	9
2.2.4. Karesi alınmış Pearson uzaklığı.....	10
2.2.5. Manhattan (City-Blok) uzaklığı.....	10
2.2.6. Minkowski uzaklığı.....	10
2.2.7. Hotelling T^2 uzaklığı	11
2.2.8. Mahalanobis uzaklığı	11
2.3. Kümeleme Teknikleri.....	11
2.3.1. Hiyerarşik kümeleme teknikleri.....	12
2.3.2. Hiyerarşik olmayan kümeleme teknikleri	20
2.4. Küme Sayısına Karar Verme	22
3. DİSKRİMİNANT ANALİZİ.....	25

	Sayfa
3.1. Diskriminant Analizinin Amacı ve Varsayımları.....	26
3.2. Diskriminant Analizinin Adımları	28
3.3. Diskriminant Fonksiyonlarının Belirlenmesi.....	29
3.4. Diskriminant Fonksiyonlarının Hatalı Sınıflandırma Olasılıkları.....	32
4. İLLERİN, ALDIĞI GÖÇÜN EĞİTİM DURUMUNA GÖRE KÜMELENMESİ ÜZERİNE BİR UYGULAMA	35
4.1. Göç Kavramı ve Türkiye’de İç Göç.....	35
4.2. İllerin, Aldığı Göçün Eğitim Durumu Bakımından Kümelenmesi	36
4.2.1. Tek bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması..	38
4.2.2. Tam bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması.	45
4.2.3. Merkezi bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması.....	53
4.2.4. Medyan bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması.....	61
4.2.5. Ward kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması	69
5. SONUÇ VE ÖNERİLER	77
KAYNAKLAR	83
EKLER.....	87
EK-1. Uygulama veri seti	88
EK-2. Diskriminant analizi sınıflama sonuçları (Tek bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı).....	91
ÖZGEÇMİŞ	93

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Veri matrisi	6
Çizelge 2.2. Benzerlik/farklılık matrisi	7
Çizelge 4.1. Betimleyici istatistikler.....	37
Çizelge 4.2. İllerin, aldığı göçün eğitim durumuna göre kümeleme analizi ANOVA tablosu	38
Çizelge 4.3. Tek bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler.....	38
Çizelge 4.4. Box-M test sonucu.....	39
Çizelge 4.5. Tek bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	39
Çizelge 4.6. Tek bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler	40
Çizelge 4.7. Box-M test sonucu.....	40
Çizelge 4.8. Tek bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	41
Çizelge 4.9. Tek bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler	41
Çizelge 4.10. Box-M test sonucu.....	42
Çizelge 4.11. Tek bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	42
Çizelge 4.12. Tek bağlantı kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler	43
Çizelge 4.13. Box-M test sonucu.....	43
Çizelge 4.14. Tek bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	43
Çizelge 4.15. Tek bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler	44
Çizelge 4.16. Tek bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları	44
Çizelge 4.17. Tam bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler	45

Çizelge	Sayfa
Çizelge 4.18. Box-M test sonucu.....	45
Çizelge 4.19. Tam bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	46
Çizelge 4.20. Tam bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler.....	47
Çizelge 4.21. Box-M test sonucu.....	47
Çizelge 4.22. Tam bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	48
Çizelge 4.23. Tam bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler	48
Çizelge 4.24. Box-M test sonucu.....	49
Çizelge 4.25. Tam bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	49
Çizelge 4.26. Tam bağlantı kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler	50
Çizelge 4.27. Box-M test sonucu.....	50
Çizelge 4.28. Tam bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	51
Çizelge 4.29. Tam bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler	51
Çizelge 4.30. Tam bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları	52
Çizelge 4.31. Merkezi bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler	53
Çizelge 4.32. Box-M test sonucu.....	53
Çizelge 4.33. Merkezi bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	54
Çizelge 4.34. Merkezi bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler	54
Çizelge 4.35. Box-M test sonucu.....	55
Çizelge 4.36. Merkezi bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	55

Çizelge	Sayfa
Çizelge 4.37. Merkezi bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler.....	56
Çizelge 4.38. Box-M test sonucu.....	56
Çizelge 4.39. Merkezi bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	57
Çizelge 4.40. Merkezi bağlantı kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler.....	57
Çizelge 4.41. Box-M test sonucu.....	58
Çizelge 4.42. Merkezi bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	58
Çizelge 4.43. Merkezi bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler	59
Çizelge 4.44. Merkezi bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları	60
Çizelge 4.45. Medyan bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler	61
Çizelge 4.46. Box-M test sonucu.....	61
Çizelge 4.47. Medyan bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	62
Çizelge 4.48. Medyan bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler	62
Çizelge 4.49. Box-M test sonucu.....	63
Çizelge 4.50. Medyan bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	63
Çizelge 4.51. Medyan bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler.....	64
Çizelge 4.52. Box-M test sonucu.....	64
Çizelge 4.53. Medyan bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	65
Çizelge 4.54. Medyan bağlantı kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler.....	65
Çizelge 4.55. Box-M test sonucu.....	66

Çizelge	Sayfa
Çizelge 4.56. Medyan bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	66
Çizelge 4.57. Medyan bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler.....	67
Çizelge 4.58. Medyan bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları	68
Çizelge 4.59. Ward kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler	69
Çizelge 4.60. Box-M test sonucu.....	69
Çizelge 4.61. Ward kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	70
Çizelge 4.62. Ward kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler	70
Çizelge 4.63. Box-M test sonucu.....	71
Çizelge 4.64. Ward kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	71
Çizelge 4.65. Ward kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler.....	72
Çizelge 4.66. Box-M test sonucu.....	72
Çizelge 4.67. Ward kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları.....	73
Çizelge 4.68. Ward kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler...	73
Çizelge 4.69. Box-M test sonucu.....	74
Çizelge 4.70. Ward kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları	74
Çizelge 4.71. Ward kümeleme tekniği - Tüm uzaklıklara göre kümeler.....	75
Çizelge 4.72. Ward kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları	76
Çizelge 4.73. Tüm kümeleme teknikleri - Diskriminant analizi sınıflandırma sonuçları	76

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Öklid uzaklığı	9
Şekil 2.2. Kümeleme teknikleri	12
Şekil 2.3. Ağaç diyagramı (dendogram)	13
Şekil 2.4. k-ortalama kümeleme tekniğinin görsel ifadesi.....	22



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

%

Yüzde

Kısaltmalar

Açıklamalar

HKT

Hata Kareler Toplamı

sd

Serbestlik Derecesi

TÜİK

Türkiye İstatistik Kurumu

1. GİRİŞ

Canlı ve cansız varlıkların sınıflandırılması yüzyıllardır bilim insanlarının ilgilendiren bir konudur. Sınıflandırma; insanların içinde buldukları ortamı veya evreni daha kolay kavrayabilmesine imkân sağlayan önemli bir araçtır. Elemanların sayısı çoğaldıkça elemanları sınıflandırmak daha da güçleşmiştir ve yeni teknikler keşfetmeyi gerektirmiştir. Bu ihtiyaçlar “kümeleme analizi” kavramını doğurmuştur (Öztürk, 2012).

Kümeleme analizi, Linnaeus’un 1753 yılında hayvanların ve bitkilerin sınıflandırılması üzerine yaptığı çalışmaya dayanan bir araştırma metodu olmakla birlikte, ilk kez John Snow tarafından 1854 yılında Londra’daki bir kolera salgını sırasında kullanılmış olup, bu olay kümeleme analizinin en basit ve bilinen ilk uygulaması olarak dikkat çekmektedir (Hofman and Jarvis, 1998).

Özellikle 1960’lı yıllardan sonra gelişme gösteren kümeleme analizi, geniş bir uygulama sahasına sahip olmuştur. Günümüzde bakıldığında; veri madenciliği, bankacılık, pazarlama, tıp, sosyoloji, ekonomiye varan bir yelpazede uygulandığı görülmektedir (Anderberg, 1973).

Kümeyi, birbiriyle benzerlik gösteren veya aynı türden olan objelerin oluşturdukları gruplar olarak tanımlamak mümkündür (TDK, 1932). Kümeleme analizi ise; bu kümeleri bir araya getirme tekniğidir. Başka bir deyişle göre kümeleme analizi; homojen olan grupların iç içe durumları veya objelerin kategorize edilebilmesi için başvurulan teknikler sınıfıdır. Bu sayede kümeleme analizi, objelerin sınıfı oluşturan benzer ya da aynı özelliklerinin daha iyi yansıtılmasına olanak sağlar. Kümeleme, bir denetimsiz öğrenme biçimidir ve var olan sınıflar içinden benzer özellikleri olanların bir araya getirilmesidir.

Kümeleme analizinin amacı, birbirine çok benzeyen birimlerden oluşan fakat nitelikleri yönünden farklı olan kümelerin ortaya çıkarılması ve bilgi depolarının farklı gruplara ayrılmasıdır (Yılmaz, 2011). Ayrıca kümeleme analizi, araştırmacıya, incelenen kümeyi, kendi içinde aynı cins ancak diğerleriyle aynı olmayacak şekilde kümelere ayırma imkanı sunan bir çok değişkenli istatistiksel analiz tekniğidir (Kantardzic, 2003).

Kümeleme analizi, birimlerin benzerlik veya farklılıklarını göz önüne alarak onları homojen gruplara ayırmaktadır. Birimler arasındaki benzerlik veya farklılıklar genel olarak uzaklık ölçüleri olarak isimlendirilir. Uzaklık ölçüleri, farklılık ölçüleri ve benzerlik ölçüleri olmak üzere 2 temel başlık altında toplanabilir. Tercih edilecek kümeleme tekniğini uygulamadan önce verinin yapısına uygun olacak şekilde uzaklık ölçüsü seçilmesi önemli olup, gerçekleştirilen kümeleme analizinin performansı, seçilecek yakınlık ölçüsünün uygunluğuna bağlıdır (Carvalho, 2015).

Birimlerin benzerliklerine göre kümelere dağılmasında kullanılabilecek birçok yaklaşım bulunmakla birlikte; bu yaklaşımlardan biri, en çok benzer iki birimi aynı gruba atamakla başlayıp tüm birimlerin aynı gruba atanması ile biten hiyerarşik bir yaklaşımdır. Bir diğer yaklaşım ise tüm verilerin ortalama değerlerine en yakın değerlere sahip birimlerin aynı kümeye atanmasını esas alan yaklaşımdır. Bütün yaklaşımlarda en önemli husus, kümeler arası farklar ile kümeler içi benzerliklerin maksimum olmasını sağlamaktır. En çok kullanılan kümeleme teknikleri hiyerarşik ve hiyerarşik olmayan kümeleme adı altında iki kategoride toplanmaktadır (Tatlıldil, 2008).

Kümeleme analizi, diğer çok değişkenli istatistiksel tekniklerle de birleştirilebilmektedir. Gerçekleştirilen analiz sonucunda elde edilen sonuçları doğrulamak adına oluşan kümelerin istatistiksel güvenilirliğinin değerlendirilmesi amacıyla kümeleme analizi, diskriminant analizi ile birleştirilebilir. Böylece kümeleme analizi sonuçlarının, diskriminant analizi ve sonucunda elde edilen fonksiyonlar yardımıyla kontrol edilebilme imkanı sağlanmaktadır (Yıldız, 1989).

Diskriminant analizi, bireylerin sahip oldukları p sayıda değişken kullanılarak bu bireylerin ait oldukları grupların belirlenmesi ya da mevcut grupların birbirlerinden ayrılmasını sağlayacak optimum fonksiyonun bulunmasında kullanılan çok değişkenli istatistik teknikleri arasında yer almaktadır (Çamdeviren, 2000). Diskriminant analizi sayesinde araştırmacılar, aynı anda çok sayı ve çeşitte değişken ile birden fazla sayıdaki örnek grubu arasındaki benzerlik ve farklılıkları belirleyebilme imkanını elde etmektedir. Gruplar arasında varyans-kovaryans eşitliğinde doğrusal diskriminant analizi kullanırken, eşit olmaması durumunda ise karesel diskriminant analizinin kullanılması önerilmiştir.

Benzerlik katsayılarının farklı kümeleme teknikleri üzerinde gösterdikleri performansı inceleyen birçok çalışma bulunmaktadır. Bu çalışmalardan biri olan Finch (2015) küme uzaklık ölçülerini karşılaştırdığı çalışmada, simülasyon sonuçlarına göre doğru grupta potansiyellerinin % 60 ila % 90 arasında olduğunu ve kümeleme teknikleri arasında en iyi performansı Ward tekniğinin gösterdiğini ortaya koymuştur. Ayrıca, Shirkorshidi (2015) kümelemede benzerlik ve farklılık katsayılarının karşılaştırılması üzerine yaptığı bir çalışmada, söz konusu katsayıların farklı kümeleme teknikleri üzerinde uygulandığında gösterdiği performansı incelemiştir. Tek bağlantı ve ortalama bağlantı kümeleme tekniklerinde Öklid uzaklığının genellikle Manhattan ve Minkowski uzaklıklarından daha iyi performans gösterdiğini değerlendirmiştir.

Ele alınan bu çalışmanın amacı da, kümeleme analizinde kullanılan bazı uzaklık ölçülerini farklı kümeleme teknikleri üzerinde uygulayarak oluşan kümeleri karşılaştırmak ve ortaya çıkan bu kümelerin her biri için diskriminat analizi uygulayarak söz konusu meydana gelen kümelerin geçerliliğini araştırmaktır. Dolayısıyla hangi uzaklık ölçüsünün hangi kümeleme tekniğinde en iyi sonucu verdiği ortaya konulmaya çalışılacaktır.

Çalışmanın uygulama bölümünde ise, Türkiye’de iç göç hareketi içinde yer alan illere yönelen 15 yaş ve üzeri nüfusun eğitim durumu incelenmiştir. Türkiye İstatistik Kurumu verilerine göre; 2017 yılı içinde Türkiye’de 81 ilin, aldığı göçün eğitim durumuna göre kümeleme analizi ile gruplaması yapılmıştır. Elde edilen grupların geçerliliği diskriminant analizi uygulanarak sorgulanmıştır.

Literatürde iç göç konusunda yapılan pek çok araştırma bulunmaktadır. Munro (1974), makro düzeyde Türkiye özelinde iç göçü araştıran en önemli çalışmalardan birine sahip olup Türkiye’de 67 ilin göç verilerini regresyon analizi kullanarak iller arasında iç göç temelinde etkili olan değişkenleri araştırmıştır. Çelik (2002), 1975-1990 verilerini kullanarak Türkiye’de iç göçü araştırmış ve göçmenler nitelikleri itibarıyla eğitim, yaş ve cinsiyet bakımından incelenmiştir. Söz konusu araştırma sonucunda göçmenlerin daha çok eğitim seviyesi yüksek erkeklerden oluştuğu saptanmıştır. Sönmez ve Er (2007), 81 ile ilişkin 2004 yılı göç verilerini kullanarak iç göç hareketlerini iki farklı kümeleme tekniğini uygulayarak yaptıkları çalışmada, büyük şehirlerden dışarıya doğru göçlerin çoğunluğunun köylere değil, üç büyük il dışında kalan diğer illere yapıldığını ortaya koymuştur. Bülbül ve Köse (2010), 2008 yılı göç verilerini kullanarak 12 bölgeye yönelik olarak göç

hususunun çeşitli değişkenlere göre belirlenen uzaklıklara göre bölgeler arasındaki benzerlikleri ve farklılıkları belirlemeyi amaçlamışlardır. Garipağaoğlu (2016), 1990 yılı genel nüfus sayımı sonuçlarına göre göç alan 21 ile yönelen nüfusun eğitim durumunu araştırmıştır. Söz konusu araştırma ile yer değiştiren bu nüfusun eğitim durumunu iller özelinde incelemiş olup göçe katılanların çoğunluğunun eğitim seviyesinin düşük yani % 50,6 oranında ilkokul mezunu olduğu ancak düşük olsa bile Türkiye ortalamasının çok üstünde bir yükseköğretim mezunu kesimin de göçe iştirak ettiğini saptamıştır.

Göç analizinin farklı boyutlarının yer aldığı araştırmalarda anket tekniği, örnekleme tekniği, regresyon analizi, en küçük kareler tekniği, kümeleme analizi ve panel veri analizi gibi istatistiksel teknikler kullanılarak iç göç olgusu analiz edilmiş olup bazılarında göç nedenleri ortaya konulurken bazılarında ise göçün sonuçları üzerinde durulmuştur. Kimi araştırmalar ise iç göç ile ilgili mevcut durum tespitine yönelerek bu konuya odaklanmıştır. Bu kapsamda, çalışmadaki bu incelemenin özünü de iç göç hareketi içinde yer alan illere yönelen nüfusun eğitim durumu teşkil etmektedir.

Çalışmanın giriş kısmından sonra, ikinci ve üçüncü bölümünde sırasıyla kümeleme analizi ve diskriminant analizi incelenerek literatüre değinilmiştir. Uygulama bölümünde ise söz konusu alanda güncel son veri olan, TÜİK 2017 yılı 15 yaş ve üzeri iç göç verileri kullanılarak Türkiye’de 81 il bazında illere göç edenlerin eğitim durumu araştırılarak durum tespiti yapılacak olup çalışmada yapılan bu sınıflandırmaların ilgili politika, uygulama ve diğer akademik çalışmalara yol gösterici nitelikte olacağı düşünülmektedir.

2. KÜMELEME ANALİZİ

2.1. Kümeleme Analizinin Tanımı ve Genel Amacı

Kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklere göre belirli gruplar içinde toplayarak, birimlerin ortak özelliklerini ortaya koymayı ve bu gruplar ile ilgili genel tanımlar yapmayı sağlayan bir tekniktir. Kümeleme analizinde birincil amaç verileri niteliklerine göre, grup içinde olası en büyük homojenliğe ve gruplar arasında olası en büyük heterojenliğe sahip gruplar oluşturarak bir araya getirmektir (Kalaycı, 2008).

Kümeleme analizinin genel amacı ise gruplanmamış verileri benzerliklerine göre gruplandırarak ve araştırmacıya uygun, işe yarar özet bilgiler elde etmede yardımcı olmaktır. Kümeleme analizi sözü edilen amaçlar dışında; gruplar için ön tahmin, hipotez testi, veriler yerine kümelerin değerlendirilmesi ve aykırı değerlerin bulunması gibi amaçlarla da kullanılmaktadır (Romesburg, 1984).

Anderberg(1973)'e göre kümeleme analizi, aşağıda belirtilen amaçlara ulaşmada faydalı olacaktır;

- İnsan tiplerini belirleme ve ayırt etme tekniği(tipoloji) veya sınıflandırma geliştirmek,
- Kuruluşları gruplama hedefiyle kavramsal sınıflama ölçekleri geliştirmek,
- Verileri sınıflayarak belli alanda kanıt üretmek,
- Hipotez test etmek ya da veri setinde başka metotlarla yapılmış olan gruplamaların varlığını saptamak (Doğan, 2008).

Özdamar (2004)'a göre ise kümeleme analizinin kullanım amaçları:

- n sayıda bireyi p değişkenine bağlı niteliklerine göre, kendi içinde mümkün olduğunca homojen ve kendi aralarında da heterojen alt kümelere ayırmak,
- p sayıda değişkeni, n sayıda bireyde saptanan bulgulara göre ortak özellikleri açıkladığı varsayılan alt kümelere ayırmak ve ortak faktör yapıları ortaya koymak,
- Bireyleri ve değişkenleri birlikte ele alarak n bireyi p değişkene göre ortak nitelikli alt kümelere ayırmak,

- p deęişkene göre saptanan bulgular açısından bireylerin biyolojik ve taksonomik sınıflamasını ortaya koymaktır (Yalçın, 2013).

Belirtilen amaçlara ulaşmada kullanılan kümeleme analizinde başarı elde edebilmek için;

- Kayıt dosyasının belirlenmesi
- Verilerin ve verilere ait temel özelliklerin saptanması
- Kümeleme yapılacak temel özelliklerin belirlenmesi
- Başlangıç merkezlerin belirlenmesi
- Uzaklık ölçümünün seçimi
- Kümeleme ölçütlerinin belirlenmesi
- Kümeleme algoritmaları ve bilgisayar uygulaması
- Küme sayısının belirlenmesi
- Nihai sonuçların yorumlanması gibi yollar izlenmelidir (Üstünel, 2018).

Kümeleme analizinde geçerli ve anlamlı sonuçlara ulaşabilmek için iki hususun mutlaka sağlanması gerekmekte olup, ilki önemli deęişkenlerin seçilmesi ikincisi ise küme sayısının isabetli seçilmesidir (Carvalho, 2015).

Kümeleme analizinin temel uygulama adımları aşağıdaki gibidir:

Adım 1- Veri Matrisinin Belirlenmesi: Birimlerin olaęan sınıflandırılmaları hakkında net bilgilerin bulunmadığı yığınlardan seçilen n sayıda birimin p sayıda deęişkeniyle alakalı gözlemler yapılarak Çizelge 2.1.'deki gibi matris haline getirilir.

Çizelge 2.1. Veri matrisi

	1.Deęişken	2.Deęişken	3.Deęişken	p. Deęişken
1.Birim	X_{11}	X_{12}	X_{13}	X_{1p}
2.Birim	X_{21}	X_{22}	X_{23}	X_{2p}
3.Birim	X_{31}	X_{32}	X_{33}	X_{3p}
....
n. Birim	X_{n1}	X_{n2}	X_{n3}	X_{np}

Adım 2- Benzerlik ya da Farklılık Matrisinin Belirlenmesi: Birimlerin/değişkenlerin birbirleri ile olan benzerliklerini ya da farklılıklarını ortaya koyan uygun bir benzerlik ölçüsü ile birimlerin/değişkenlerin birbirlerine uzaklıkları hesaplanır. Bulunan uzaklıklara ilişkin Çizelge 2.2. şeklinde benzerlik/farklılık matrisi hazırlanır.

Çizelge 2.2. Benzerlik/farklılık matrisi

	1. Birim	2. Birim	3. Birim	n. Birim
1. Birim	0	d_{12}	d_{13}	d_{1n}
2. Birim	SİMETRİ	0	d_{23}	d_{2n}
3. Birim	SİMETRİ	SİMETRİ	0
....	SİMETRİ	SİMETRİ	SİMETRİ	0	$d_{n-1,n}$
n. Birim	SİMETRİ	SİMETRİ	SİMETRİ	SİMETRİ	0

Adım 3- Kümelerin Oluşturulması: Doğru küme metodu yardımı ile benzerlik/farklılık matrisine göre birimlerin/değişkenlerin uygun sayıda oluşturulan kümelere bölünmesi sağlanır. Gruplandırma yapılırken, sıklıkla kullanılan kümeleme tekniklerinden hiyerarşik ve hiyerarşik olmayan kümeleme tekniklerinden uygun olanlar seçilir. Birimler birbirlerine olan uzaklık veya yakınlık derecelerine göre uygun sayıda kümelere bölünür.

Adım 4- Analiz ve Sonuç: Ulaşılan kümelerin analiz edilmesi ve bu kümeleme şekline dayanarak kurulan varsayımların doğrulanması için gerekli olan analitik metotların uygulanarak, sonuçların hassaslığının ve anlamlılığının tartışması yapılır. Elde edilen sonuçların doğru olmaması durumunda (değişkenlerin uygun olmaması ve/veya küme sayısının uygun belirlenmemiş olması nedeniyle) değişkenler, küme sayısı, uzaklık ölçüleri ve kümeleme metotları gibi değişkenler gözden geçirilerek süreç başından tekrar edilir.

Kümeleme analizinin pek çok olumlu tarafları bulunmaktadır. Kümeleme analizinde ilişkilerin görüntülenmesi, sonuçların grafik haline getirilmesi açısından önemlidir. Bu sayede görsel sonuçlardan benzerlikler kolayca tespit edilir. Ayrıca aykırı değerler de kolaylıkla saptanır. Bazı tekniklerde daha küçük veri toplulukları üzerinde çalışılması gerekmekte olup, kümeleme analizi ile büyük veri toplulukları daha küçük kümelere bölünmesi sayesinde bu tür tekniklerin uygulanabilmesi için ortam oluşturulur (Green, 1989).

2.2. Kümeleme Analizinde Kullanılan Benzerlik Ölçütleri

Kümeleme analizinin genel hedefi küme içindeki homojenliği, kümeler arasında ise heterojenliği oluşturmaktır. Bu da belirgin özellikleri açısından benzer bireylerin aynı kümede bir araya gelmesiyle sağlanmış olur. Bireylerin benzerlikleri uzaydaki konumlarıyla alakalı olup uzayda birbirlerine daha yakın olan bireyler aynı kümede bir araya gelmiş olacaklardır. Bireyler sınıflandırılırken birbirlerine mesafeleri çeşitli uzaklık ölçülerine göre bulunur (Öztürk, 2012).

Uzaklık, farklılıkları ölçerken farklılıklar ise çeşitli niteliklere göre iki nesnenin zıtlık veya uyumsuzluklarının bir ölçümüdür. Uzaklık ölçümleri bir gözlemin ötekiler içinde fark edilmesini sağlar ve bunun sayesinde gözlemler benzerlik ya da farklılıklarına göre sınıflara ayrılabilir. Birimler alt gruplara ya da kümelere bir kere dahil olduktan sonra her grubun belirgin özellikleri anlaşılabilir ve kümelerin nitelikleri açıklanabilir. Gruplama sayesinde bilginin düzenlenmesi ve incelenmesi daha kolay sağlanır. Yeni veriler belirlenen alt grup ya da kümelere göre ayrılabilir ve nitelikleri öngörülebilir. Veriler sayesinde basit hale getirilerek daha etkin değişkenler üzerinde çalışmalar yapılabilir ve veri kümesi içindeki gizli yapı ortaya çıkarılabilir. Verinin içindeki kümelene yapısına ve bu kümelerin değişken tahminine göre plan yapıp karar alınabilir (Servi, 2009).

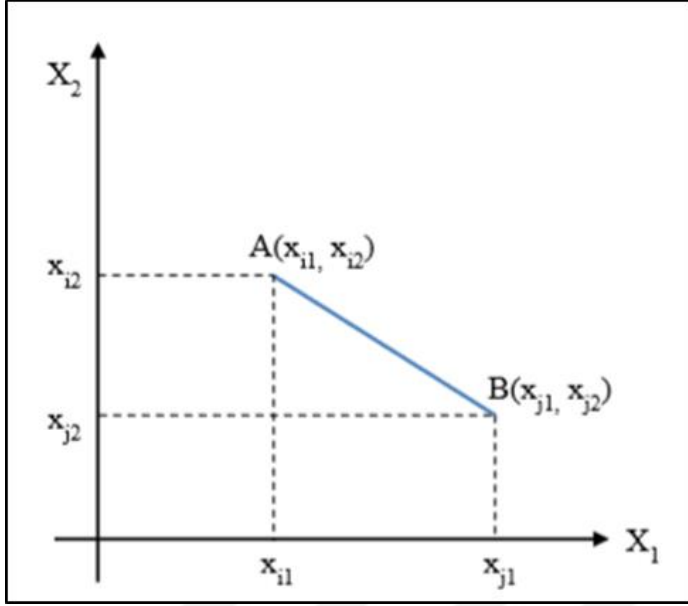
Verilerin gruplandırılmasında üzerinde etkili olan değişkenlerin kesikli veya sürekli olmalarına ya da değişkenlerin sınıflama, sıralama, aralık ya da oransal ölçekte olmalarına dayanarak kullanılacak uzaklık ölçüsü belirlenir (Öztürk, 2012). Uzaklık ölçülerinden en çok kullanılanları şunlardır:

2.2.1. Öklid uzaklığı

Öklid uzaklık ölçüsü, uzaklık ölçüleri içinde en sık kullanılan ölçülerdendir. En önemli niteliklerinden biri; iki birim arası mesafenin, analize yeni birimlerin dahil olmasıyla değişmemesidir. Boyutları birbirine benzeyen ölçeklere sahip olan kayıtlarda kullanılan yararlı bir uygulamadır. İki birim arasındaki Öklid uzaklığı, karşılık olan değerler arasındaki farklarının karelerinin toplamının karekökünü hesap etmeyi kapsar (Üstünel, 2018). Bu kapsamda, p değişken sayısı ve n birim sayısı olmak üzere $i, j = 1, 2, 3, \dots, n$, i . ve

j. birimleri arasındaki uzaklık şekilde de gösterilen Öklid uzaklık ölçüsü kullanılarak formüldeki gibi hesaplanır (Tatlıldil, 2008):

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (2.1)$$



Şekil 2.1. Öklid uzaklığı

2.2.2. Karesi alınmış Öklid uzaklığı

Karesi alınmış Öklid uzaklık ölçümü, iki birim arasında birimlere karşılık gelen değerlerin farklarının karelerinin toplamını hesaplamasıyla bulunur. Öklid uzaklık ölçüsüyle aynı denklemi kullanılmasıyla beraber, Öklid uzaklığının karesidir. Karesi alınmış Öklid uzaklık ölçüsü formülü aşağıda gösterilmiştir (Tatlıldil, 2008):

$$d(i, j) = (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2 \quad (2.2)$$

2.2.3. Pearson uzaklığı

İki birim arasındaki uzaklığın bulunduğu Pearson uzaklık ölçüsü formülü aşağıdaki gibidir:

$$d(i, j) = \sqrt{\frac{(x_{i1} - x_{j1})^2}{s_1^2} + \frac{(x_{i2} - x_{j2})^2}{s_2^2} + \dots + \frac{(x_{ip} - x_{jp})^2}{s_p^2}} \quad (2.3)$$

Formülde kullanılan S_p , uzaklığın hesaplandığı değişkene ait varyanstır ve farklı gruplarla ilgili daha önceden bilgi sahibi olunmadığından, uzaklık hesaplanmasında S değerinin kullanılması yanlış olur. Bu sebeple Pearson uzaklık ölçüsünün yerine sıklıkla Öklid uzaklık ölçüsü kullanılır (Dinler, 2014).

2.2.4. Karesi alınmış Pearson uzaklığı

Karesi alınmış Pearson uzaklık ölçüsü formülü aşağıdaki gibidir:

$$d(i, j) = \frac{(x_{i1} - x_{j1})^2}{S_1^2} + \frac{(x_{i2} - x_{j2})^2}{S_2^2} + \dots + \frac{(x_{ip} - x_{jp})^2}{S_p^2} \quad (2.4)$$

2.2.5. Manhattan (City-Blok) uzaklığı

Manhattan uzaklığı veya City-Blok uzaklığı isimleriyle kullanılan bu uzaklık, uzaklıkları tanımlamak için kullanılan iki ya da daha çok boyuttaki mutlak farkların toplanmasından elde edilir. Başka deyişe göre bu metot, değişkenler arasındaki farkın mutlak değerini baz almaktadır. Uzaklık, gözlemlerin koordinatları arasındaki farkın mutlak değerleri toplamıyla bulunur (Johnson and Wichern, 1988).

$$d(i, j) = (|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|) \quad (2.5)$$

2.2.6. Minkowski uzaklığı

Uzaklıkların belirlenmesinde bir diğer kullanılan ölçüde Minkowski olarak bilinen uzaklık ölçüsüdür. Minkowski uzaklık ölçüsü aşağıdaki gibi gösterilmektedir.

$$d(i, j) = \left[|x_{i1} - x_{j1}|^m + |x_{i2} - x_{j2}|^m + \dots + |x_{ip} - x_{jp}|^m \right]^{\frac{1}{m}} \quad (2.6)$$

Minkowski uzaklık ölçüsü formülünde m değeri büyük ve küçük farklara verilen ağırlığı etkiler. $m=1$ olarak kabul edilirse formül, Manhattan uzaklık ölçüsü formülüne dönüşür. Eğer $m=2$ olarak kabul edilirse formül Öklid uzaklık ölçüsü formülüne dönüşür (Tatlıdil, 2008).

2.2.7. Hotelling T² uzaklığı

İki grup veya kümenin ortalama vektörlerinin kıyaslanması için başvurulan Hotelling T² değeri, uzaklık ölçütü olarak da kullanılmaktadır.

$$d(i, j) = T^2 = \frac{n_1 \cdot n_2}{n} (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j) \quad (2.7)$$

S, p×p kovaryans matrisini ifade etmektedir.

i, j=1,2,...,n'dir ve n birim sayısıdır (Tatlıldil, 2008).

2.2.8. Mahalanobis uzaklığı

Mahalanobis uzaklığı aslında merkez ve her bir veri noktası arasındaki uzaklığı göstermektedir. Verideki değişkenler birbirleriyle ilişkili değil ise, S bir birim matris yapısına dönüşür ve Mahalanobis uzaklığı karesi alınmış Öklid uzaklığına eşit olur. Doğrudan birleştiren, standart bir metot olan Mahalanobis uzaklık ölçüsü, iki değişken arasında bir ilişki var ise bu iki değişken arasındaki kovaryans ya da korelasyonu dikkate alan Mahalanobis uzaklığının kullanılması uygundur. Bu çerçevede, p sayıda değişkenli bir değerlendirmede i ve j birimleri arasındaki Mahalanobis uzaklık ölçüsü aşağıdaki formülle bulunur (Sharma 1996).

$$d(i, j) = D^2 = (x_i - x_j)' S^{-1} (x_i - x_j) \quad (2.8)$$

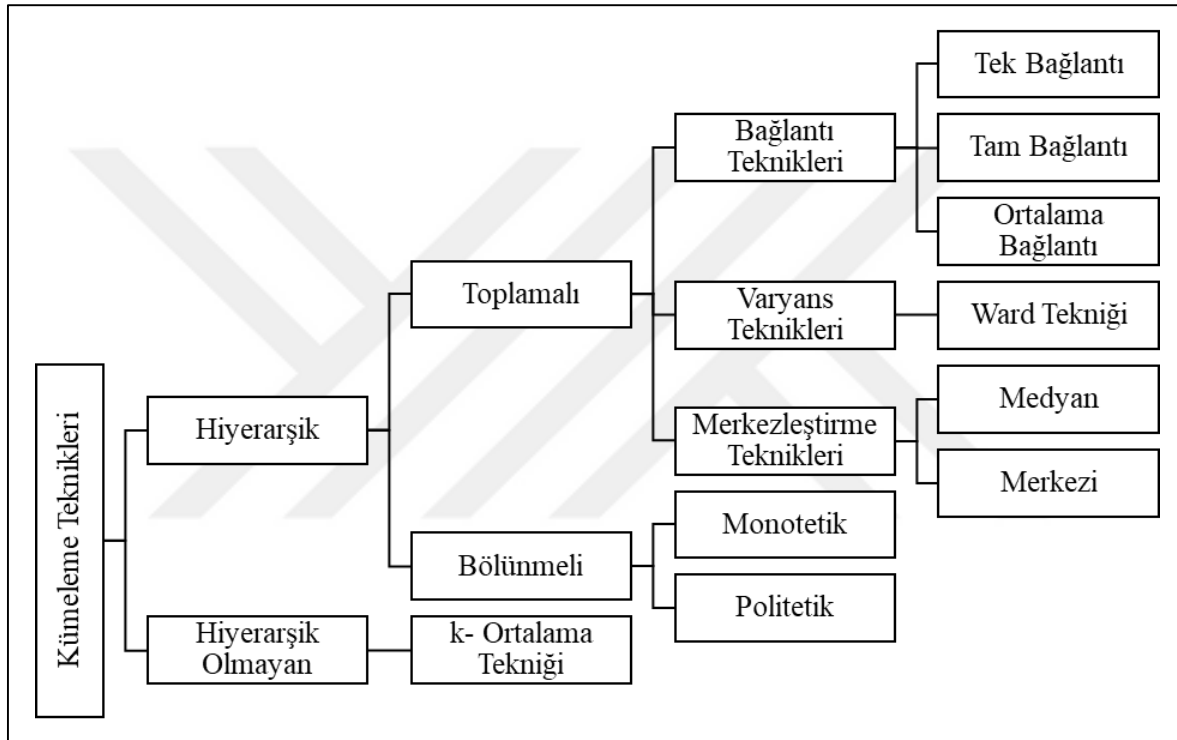
S, p×p kovaryans matrisini ifade etmektedir.

2.3. Kümeleme Teknikleri

Kümeleme teknikleri; uzaklık matrisini kullanarak objeleri ya da değişkenleri kendi aralarında heterojen ve kendi içinde homojen olacak şekilde gruplar kurmaya olanak sağlar. Kullanılacak benzerlik ölçüsü seçildikten sonra kümeleme işleminin ne şekilde yapılacağına karar verilir. Kullanılacak kümeleme metodunun doğru tercih edilmesi yapılan analizin hedefine ulaşmasında büyük önem taşır. Bu sebeple kümeleme

metotlarının temel nitelikleri doğru kavranmalı ve uygulanan metodun sonuçları doğru analiz edilmelidir (Doğan, 2008).

2000’li yıllardan beri kümeleme çalışmalarıyla ilgili birçok teknik geliştirilmiştir. Bu teknikleri standartlaştırmak ve onları kategorize etmek oldukça güçtür. Kaynaklarda kabul edilen kümeleme tekniklerinin hiyerarşik teknikler ve hiyerarşik olmayan teknikler şeklinde iki grupta toplandığı görülür. (Taştatan, 2018).



Şekil 2.2. Kümeleme teknikleri

2.3.1. Hiyerarşik kümeleme teknikleri

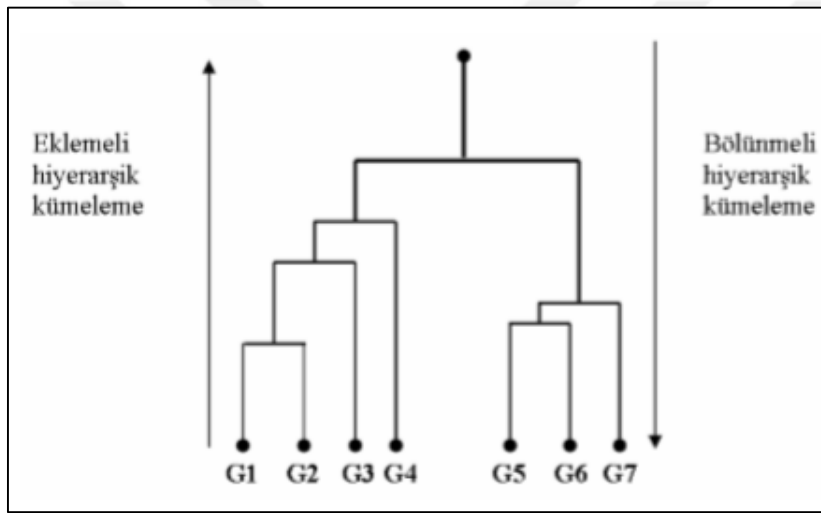
Hiyerarşik kümeleme tekniklerinin bir diğer adı bağlantı teknikleridir. Bu teknikler birimleri birbirleri ile farklı aşamalarda toplayarak; ardışık kümeler kurmaya ve bu kümelere dahil olacak birimlerin nasıl bir uzaklık veya benzerlik düzeyinde dahil olduğunu saptamaya yöneliktir (Doğan, 2008).

Hiyerarşik kümelemede, araştırma sonucunda elde edilen birimler iç içe bölünmelerin ya da birleşmelerin bir sırası şeklinde gruplanır. Hiyerarşik kümeleme teknikleri uygun bir hesaplama tekniği kullanarak verideki doğru ifade edilmiş kümeleri bulmaya çalışır. n

sayıda birimden oluşan bir araştırma sonucu g farklı küme yapısı içeriyorsa, n gözlemin g farklı kümeye bölünmelerinin sayısı,

$$N(n, g) = \frac{1}{g!} \sum_{k=1}^g \binom{g}{k} (-1)^{g-k} k^n \quad (2.9)$$

formülüyle hesaplanır. Bu değer $\frac{g^n}{g!}$ ile yaklaşık olarak da hesaplanabilir (Duran and Odell, 1974). Hiyerarşik kümeleme tekniğinde, dendogram olarak bilinen ağaç diyagramı kullanılır. Ağaç diyagramı, hiyerarşik kümeleme tekniğiyle elde edilen kümelerin görselleştirilmesini sağlar. Küme sayısına görsel olarak karar verilir (Servi, 2009).



Şekil 2.3. Ağaç diyagramı (dendogram)

Toplamalı ve bölünmeli şeklinde iki hiyerarşik teknik bulunmaktadır. Toplamalı hiyerarşik teknikte her birim en başta bir küme olarak varsayılır. Daha sonra en yakın iki küme yeni bir kümede toplanarak bütünleştirilir. Bu sayede her basamakta küme sayısı bir azalmış olur. Bölünmeli hiyerarşik teknikte ise aşamalar toplamalı hiyerarşik tekniğin tam tersidir. Bu teknikte araştırma sonuçlarından elde edilen büyük bir küme küçük parçalara bölünerek yeni kümeler oluşturulmaya başlanır. Farklı olan araştırma sonuçları seçilerek daha küçük kümeler elde edilir. Her birim kendi başına küme oluşturana dek uygulamaya devam edilir (Dinler, 2014).

n sayıda birimden küme oluşturmak için kademeli olan tekniklerde kullanılan algoritmaların temel basamakları şu şekildedir:

- Birimler arasındaki mesafeler $n \times n$ boyutlu simetrik bir matrisin ($D = \{d_{ij}\}$) gösterdiği n küme ile hesaplanmaya başlanır.
- Birbirine en yakın (D matrisindeki en küçük değer) iki küme bir araya getirilir.
- Küme sayısı bir azaltılarak tekrarlanmış uzaklar matrisi elde edilir.
- Tek küme kalıncaya dek 2. ve 3. basamaklar $n-1$ kere tekrarlanır (Çakmak, 1999).

Hiyerarşik kümeleme tekniklerinin; iki kümenin bir araya getirilmesine karar verildikten sonra işlem geri alınamaması yani analizin başında yapılan parçalamada meydana gelen sorunlar analizin sonuna kadar sürmesi, herhangi bir hedef fonksiyonunun direkt minimize edilememesi, çeşitli uzaklık ölçüleriyle farklı sonuçlar bulunması gibi dezavantajları bulunmaktadır (Taştan, 2018).

En çok kullanılan 7 hiyerarşik kümeleme teknikleri şunlardır:

- Tek Bağlantı Kümeleme Tekniği
- Tam Bağlantı Kümeleme Tekniği
- Ortalama Bağlantı Kümeleme Tekniği
- McQuitty Bağlantı Kümeleme Tekniği
- Merkezi Bağlantı Kümeleme Tekniği
- Medyan Bağlantı Kümeleme Tekniği
- Ward Kümeleme Tekniği

Tek bağlantı kümeleme tekniği

En yakın komşuluk olarak da bilinen tek bağlantı kümeleme tekniği, hiyerarşik kümeleme teknikleri arasında en kolayıdır. Bu teknik, uzaklıklar matrisiyle birbirine en yakın (uzaklık değerleri en küçük) kişi ya da objeleri tespit ederek birleştirmeyi amaçlamaktadır (Johnson and Wichern, 1988).

Tek bağlantılı kümeleme tekniği, birbirine en yakın iki birimin tespit edilmesiyle başlar. Bu iki birim ile bir küme oluşturulur. Daha sonrasında, üçüncü birimin küme içindeki herhangi iki birimden birine olan mesafesi, kümelenmemiş durumdaki öteki birimlere olan mesafelerinden daha az ise, bu kümeye dahil edilir. Bu işlem küme dışında bir birim kalmayıncaya kadar devam eder. Yeni veriler, mevcut bir küme ya da bu kümede var olan

bir birim ile ortak nitelikler taşıyorsa bir araya getirilmektedir. Bu durum zincirleme etkisi nedeniyle farklı sonuçların birleştirilmesine sebep olmaktadır. Yani tek bağlantılı kümeleme tekniği birbirinden yeteri kadar bağımsız olan kümeleri saptamakta oldukça iyi, ancak birbirinden farkı az olan kümeleri ayırt etmede ise yeterli değildir. Birleştirme yapılırken, kümelerin eleman sayısının birden fazla olma koşulu yoktur ve bir birim yalnız başına da bir küme oluşturabilir (Sangün, 2007). Bu teknikte i. ve j. birimler bir araya getirilmiş ise bir araya getirilen kümenin k. küme ile ilişkisi, uzaklık ölçütü olarak

$$d_{k(i,j)} = \min(d_{(k,i)}, d_{(k,j)}) \quad (2.10)$$

şeklinde gösterilmektedir. Eşitlikte;

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

$d_{(k,j)}$, k. kümenin j. kümeyle olan uzaklığına,

$d_{(k,i)}$, k. kümenin i. küme ile olan uzaklığına karşılık gelmektedir (Rencher, 2002).

Tam bağlantı kümeleme tekniği

Tam bağlantılı kümeleme tekniğinin tek bağlantılı tekniğinden ayrılan yönü, birimler arasındaki en uzak mesafelerin hesaplanmasıdır. Aralarındaki mesafeleri en çok olan birimlerin aynı kümede bir araya getirilmesiyle, tek bağlantı tekniğindeki gibi kümeleme işlemi yapılır (Öztürk, 2012).

Tam bağlantı tekniğinde aynı kümede bulunan birimlerin mesafelerinin belirli bir değerden az olması halinde bütün kümelerin doğru oluşturulması garantilenememektedir. Diğer bir adıyla “En Uzak Komşuluk Tekniği” olarak bilinen bu teknikte, uç noktalarda bulunan birimler için fazlasıyla hassas bir tekniktir. Bu teknik, maksimum uzaklıktaki iki birimin bir kümeyle dahil edilmesiyle başlar ve aynı şekilde kümelerin öteki kümelerle bir araya getirilmesiyle sonuçlanır. Birimlerin kümelerle ya da kümelerin diğer kümelerle bir araya getirilmesinde, birbirlerine olan uzaklıklarının maksimum seviyede olması temel ölçüttür (Green, 1989). Bu teknik;

$$d_{k(i,j)} = \max(d_{(k,i)}, d_{(k,j)}) \quad (2.11)$$

şeklinde gösterilmektedir. Eşitlikte;

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

$d_{(k,j)}$, k. kümenin j. kümeye olan uzaklığına,

$d_{(k,i)}$, k. kümenin i. küme ile olan uzaklığına karşılık gelmektedir (Rencher 2002).

Ortalama bağlantı kümeleme tekniği

Bu teknikte, bir kümedeki birim çiftleri ile öteki kümedeki birim çiftleri arasındaki ortalama fark, bu iki küme arasındaki fark olarak kabul edilir. Bu tekniğin farklılaştırılmış biçimleri de vardır. En sık kullanılan biçiminde birim çiftleri arasındaki mesafenin aritmetik ortalaması hesaplanmaktadır. Genel olarak tam bağlantı ve ortalama bağlantı tekniklerinde birbirine benzeyen dendogramlar meydana gelmektedir. Fakat her bir teknikte uzaklık değişik biçimlerde tanımlandığı için bütünleştirmeler farklı düzeylerde görülmektedir (Çakmak, 1999).

Tek bağlantılı kümeleme tekniğinde birbirine minimum uzaklıktaki, tam bağlantılı kümeleme tekniğinde ise birbirine maksimum uzaklıktaki birimlerden başlanarak kümeleme işlemi yapıldığından, bahsedilen iki teknik uç değerlerden etkilenmeye açıktır. Ortalama bağlantılı kümeleme tekniği bu olumsuz durumu yok eden bir teknik olarak kullanılmaktadır. Çünkü benzerlik, uçlarda bulunan bir birim çiftine değil, kümedeki bütün birimlerin benzerliğiyle ilişkilidir. Bu sebeple de ortalama bağlantılı kümeleme tekniği uç değerlerden minimum düzeyde etkilenir. Bu teknik hemen hemen eşit grup içi varyansa sahip kümeler ortaya çıkartmaktadır ve küçük küme içi değişimleri olan kümeler oluşturma eğilimindedir (Yalçın, 2013). Ortalama bağlantılı kümeleme tekniği;

$$d_{(i,j)k} = (N_i d_{(i,k)} + N_j d_{(j,k)}) / N_i + N_j \quad (2.12)$$

şeklinde gösterilir. Eşitlikte;

$d_{(i,j)k}$: i ve j'inci kümenin k'inci kümenin küme ile olan uzaklığı,

$d_{(i,k)}$: i'inci kümenin k'inci kümeyle olan uzaklığı,

$d_{(j,k)}$: j'inci kümenin k'inci kümeyle olan uzaklığıdır.

N_i : i'inci kümedeki toplam birim sayısı,

N_j : j'inci kümedeki toplam birim sayısıdır (Rencher, 2002).

McQuitty bağlantı kümeleme tekniği

Söz konusu teknik, ağırlıksız ortalama bağlantı kümeleme tekniği olarak da bilinmekte olup, k'nın i ve j kümeleri arasındaki uzaklığı aşağıda verilen eşitlik yardımı ile bulunabilir (Rencher, 2002).

$$d_{(i,j)k} = (d_{(i,k)} + d_{(j,k)})/2 \quad (2.13)$$

$d_{(i,k)}$: i ve j'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(i,k)}$: i'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(j,k)}$: j'inci kümenin k'inci küme ile olan uzaklığını göstermektedir.

Merkezi bağlantı kümeleme tekniği

Merkezi bağlantı kümeleme tekniği, ortalama bağlantı kümeleme tekniğinin özel bir biçimi olup kümeler arası uzaklıklar ve küme merkezleri arası uzaklıklar olarak tanımlanmaktadır. Bu kapsamda, k'nın i ve j kümeleri arasındaki uzaklığı aşağıda verilen eşitlik yardımı ile bulunabilir (Rencher, 2002).

$$d_{(i,j)k} = \frac{N_i d_{(i,k)} + N_j d_{(j,k)}}{N_i + N_j} - \frac{N_i N_j d_{(i,j)}}{N_i^2 + N_j^2} \quad (2.14)$$

$d_{(i,k)k}$: i ve j'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(i,k)}$: i'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(j,k)}$: j'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(i,j)}$: i'inci kümenin j'inci küme ile olan uzaklığını,

N_i : i'inci kümedeki toplam birey sayısını,

N_j : j'inci kümedeki toplam birey sayısını göstermektedir.

Medyan bağlantı kümeleme tekniği

Medyan bağlantı kümeleme tekniği, daha çok ölçüm değerlerinin yerine sıra puanları kullanılarak elde edilen kümelerdir. Bu çerçevede, k'nın i ve j kümeleri arasındaki uzaklığı aşağıda verilen eşitlik yardımı ile bulunabilir.

$$d_{(i,j)k} = \frac{d_{(i,k)} + d_{(j,k)}}{2} - \frac{d_{(i,j)}}{4} \quad (2.15)$$

$d_{(i,k)k}$: i ve j'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(i,k)}$: i'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(j,k)}$: j'inci kümenin k'inci küme ile olan uzaklığını,

$d_{(i,j)}$: i'inci kümenin j'inci küme ile olan uzaklığını göstermektedir (Rencher, 2002).

Ward kümeleme tekniği

Ward tekniği, minimum varyans tekniği olarak da bilinmekte olup küme içindeki homojenliğin en üst seviyede olmasına odaklanan bu teknikte kümeler arasındaki mesafeler dikkate alınmaz. Homojenlik küme içi kareler toplamı ile hesaplanırken, bu

durumun en alt düzeyde olması için çalışılmaktadır. Minimum değeri üzerinde çalışılan küme içi kareler toplamı, Hata Kareler Toplamı (HKT) ismiyle de bilinir (Taşatan, 2018).

Ward tekniğinde, başlangıçta her biri farklı bir alt grup olan pek çok birimin belirli değişkenleri dikkate alınarak, en az bilgi kaybıyla benzer yönlerine göre kümelemek olasıdır. Bu sayede pek çok bireysel birimlerin aralarındaki ilişkileri anlamak ve yorumlamak daha basit olacaktır. Bu teknik, bütünleştirici hiyerarşik kümeleme tekniğinin, n birimden başlayarak ve her basamakta birbirine benzeyen birimleri ya da kümeleri bir araya getirerek sonunda tek kümeye ulaşmayı esas almaktadır. Kümelemenin her basamağında, ilk basamakta birimler arasında sonraki basamaklarda ise birimler ve kümeler arasında olabilecek her türlü birleşme ihtimalinin üzerinde durularak, bilgi kaybını en aza indiren birimlerin ya da kümelerin bir araya getirilmesi amaçlanmaktadır (Sharma, 1996). Bu teknikte k'nın i ve j kümelerine uzaklığı;

$$d_{(i,j)k} = \frac{(N_k + N_i)d_{(i,k)} + (N_k + N_j)d_{(j,k)} + N_k d_{(i,j)}}{N_k + N_i + N_j} \quad (2.16)$$

şeklinde gösterilir. Eşitlikte;

$d_{(i,j)k}$: i ve j'inci kümenin k'inci kümenin küme ile olan uzaklığı,

$d_{(i,k)}$: i'inci kümenin k'inci küme ile olan uzaklığı,

$d_{(j,k)}$: j'inci kümenin k'inci küme ile olan uzaklığı,

$d_{(i,j)}$: i'inci kümenin j'inci küme ile olan uzaklığıdır.

N_i : i'inci kümedeki toplam birim sayısı,

N_j : j'inci kümedeki toplam birim sayısı,

N_k : k'inci kümedeki toplam birim sayısıdır (Rencher, 2002).

2.3.2. Hiyerarşik olmayan kümeleme teknikleri

Küme sayısının belirlendiği durumlarda, uzun bir süreç gerektiren hiyerarşik kümeleme teknikleri yerine hiyerarşik olmayan kümeleme teknikleri tercih edilir. Hiyerarşik olmayan kümeleme teknikleri değişkenlerin yerine gözlemleri gruplandırmaya yardımcı olur ve bu gözlemleri k sayıda kümeye böler. Küme sayısı k 'ya ilk başta karar verilebileceği gibi kümeleme işlemine bağlı olarak da karar verilebilir (Johnson, 2002). Çünkü uzaklık matrisine karar verme zorunluluğu ve temel verinin bilgisayarın işlem gördüğü süre boyunca depolanma zorunluluğu yoktur. Hiyerarşik olmayan kümeleme teknikleri, hiyerarşik kümeleme tekniklerine göre daha büyük kümelere kullanılabilir (Johnson and Wichern, 1988).

Hiyerarşik olmayan kümeleme teknikleriyle elde edilecek k sayıda kümede her bir küme minimum bir birim bulundurur ve her birim sadece bir kümede yer alır. Hiyerarşik olmayan kümeleme tekniğinde, ilk başta küme merkezleri rastgele seçilir. Daha sonra, birimlerin seçilen kümelerin merkezlerine olan mesafelerine göre yeni küme merkezleri belirlenir. Bu adımlar birbirilerinden farklı, kendi içlerinde birbiriyle benzer olan, birbirleri arasında ise benzer olmayan k sayıda küme elde edilinceye dek devam ettirilir (Tatlıdil, 2008).

Hiyerarşik olmayan kümeleme tekniklerini, hiyerarşik kümeleme tekniklerinden ayrılan özellikleri aşağıdaki gibi özetlenebilir:

- Herhangi bir hiyerarşik mekanizma bulunmaz.
- Hiyerarşik tekniklerde birimler aşama aşama her birim tek kümede toplanana kadar veya tam tersi biçiminde kümelenir. Hiyerarşik olmayan tekniklerde ise birimler ilk başta karar verilen kümelere dahil edilerek işlem gerçekleştirilir.
- Hiyerarşik kümeleme tekniklerinde her aşamada hangi birimin hangi kümede bulunduğu belirlenirken, hiyerarşik olmayan kümeleme tekniklerinde ise yalnızca birimlerin son durumdaki konumları önemlidir.
- Hiyerarşik kümeleme tekniklerinde dendogramlar kullanılırken, hiyerarşik olmayan kümeleme tekniklerinde ise dendogramların bir anlamı yoktur.
- Hiyerarşik olmayan kümeleme teknikleri, hiyerarşik kümeleme tekniklerine göre daha büyük kümeler için uygulanabilir (Yalçın, 2013).

k- ortalama tekniđi

Bir hiyerarşik olmayan kümeleme tekniđi olan k-ortalama tekniđi kümeleme analizinde en çok tercih edilen tekniklerinden biri olmakla birlikte, k-ortalama tekniđi, hata kareler toplamının en az seviyede olmasına dayanarak verideki en uygun ayrılmayı tespit etmeyi hedefleyen uyumlaştırma tekniđidir. (Han and Kamber, 2001).

Bununla birlikte, k-ortalama tekniđi büyük verileri işlemede başarılı olduđu için daha çok bu hususta tercih edilen bir tekniktir ve k-ortalama tekniđinin uygulama basamakları aşağıdaki gibi özetlenebilir:

Adım 1: Başlarken kullanılacak küme merkezlerini bulmak için k sayıda merkez belirlenir. Küme merkezleri gelişi güzel veya farklı teknikler aracılığıyla belirlenebilir.

Adım 2: Her verinin belirlenen merkezlere uzaklığı tespit edilir. Bulunan sonuçlara göre tüm veriler k sayıda kümeden kendilerine minimum yakınlıktaki kümeye dahil edilir.

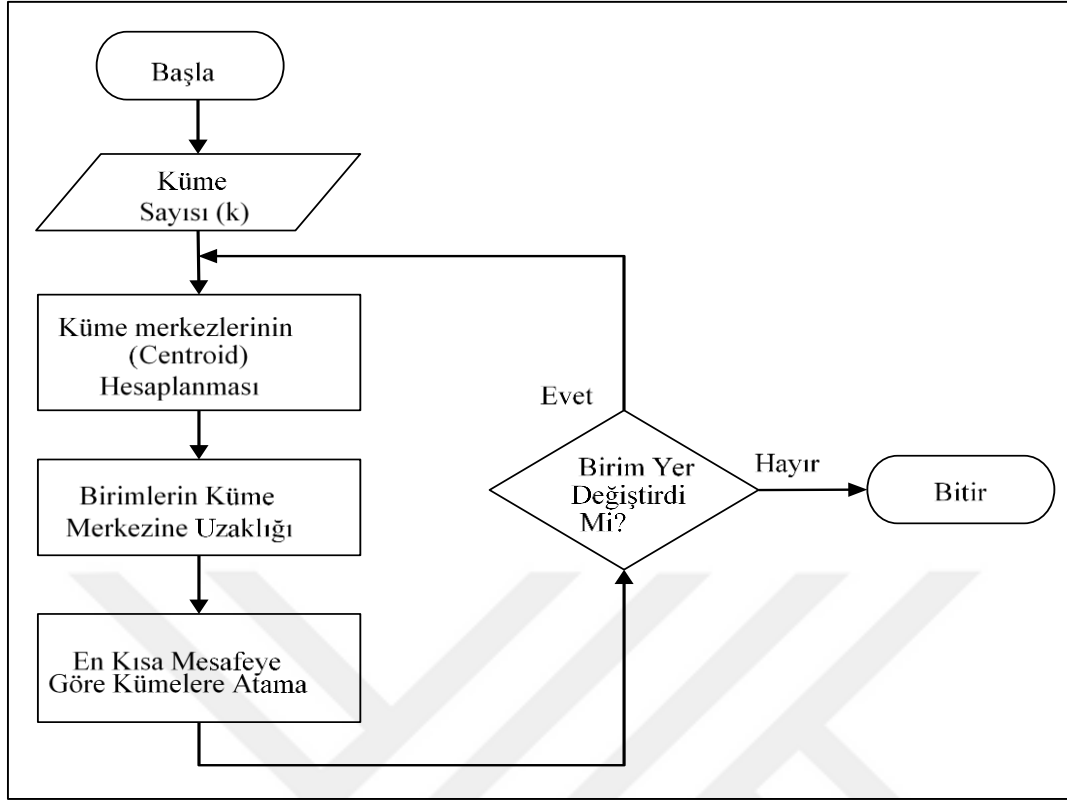
Adım 3: Elde edilen kümelerin yeni merkezleri, kümedeki bütün verilerin ortalaması alınarak tekrar hesaplanır.

Adım 4: Adım 2 ve Adım 3, küme merkezleri deđişmeyene kadar yinelenir.

Bu çerçevede, k-ortalama tekniđinde, karşı bir durum söz konusu olmadıkça, Öklid uzaklık ölçüsü tercih edilir (Üstünel, 2018).

Bu teknikte birimler, gruplar içindeki kareler toplamı minimum olacak biçimde k kümeye bölünür. x_1, x_2, \dots, x_n deđişkenlerinin her biri p deđişkenli gözlem vektörleri, çok boyutlu X uzayında farklı birer noktayı temsil ederken, aynı uzayda $a_{1n}, a_{2n}, \dots, a_{kn}$ her grup birim için küme merkezleri olarak tespit edildiğinde, aşağıdaki formüle göre birimler minimum uzaklığı veren kümeye atanmaktadır. Bu teknik uç verilerden etkilenmekte olup, k-ortalama tekniđinin şeması şekildeki gibi gösterilebilir (Taşatan, 2018).

$$W_N = \frac{1}{n} \sum_{i=1}^n \min \|x_i - a_{in}\|^2 \quad (2.17)$$



Şekil 2.4. k-ortalama kümeleme tekniğinin görsel ifadesi

En çok olabilirlik tekniği

En çok olabilirlik tekniğinde her bir birim, en büyük olabilirlik değerini sağlayacak şekilde daha önceden tespit edilen kümelere dahil edilir. Kuramsal anlamda güçlü olmasının yanında en çok olabilirlik tekniği çok sık tercih edilmez (Tatlıdil, 2008).

2.4. Küme Sayısına Karar Verme

Kümeleme analizinde küme sayısına karar verme önemli konulardan biridir. Bu kapsamda, günümüzde küme sayısının ne olacağını kesin olarak söyleyen bir teknikten bahsetmek mümkün olmamakla beraber ilk önerilen ve en çok bilinen teknik olarak;

$$k = (n/2)^{1/2} \quad (2.18)$$

biçiminde hesaplanmaktadır. Burada k küme sayısı, n birim sayısını göstermektedir. Küçük örneklemli araştırmalarda kullanılması tavsiye edilmekle birlikte, büyük örneklemli araştırmalarda kullanılması sağlıklı sonuçlara ulaşılmasını zorlaştırmaktadır (Everitt 1974).

Küme sayısı konusunda bir diğer teknik ise Marriott (1971) tarafından aşağıdaki eşitlikte olduğu şekilde önerilmiştir.

$$M = k^2|W| \quad (2.19)$$

biçimindedir ve M harfi ile gösterilir. Burada W , grup içi kareler toplamı matrisidir. M değerini minimum yapan k değeri uygun küme sayısı olarak alınmaktadır (Marriott, 1971).

Kümeleme analizi sonucunda elde edilen kümelerin geçerlilik problemi bahse konu analizin en zor kısımlarından biridir. Veri setinin analizi ve kümeleme çözümünün elde edilmesinden sonra ulaşılan sonuçların anlamlılığı ve güvenilirliği konusunda bir garanti yoktur. Dolayısıyla, elde edilen kümeleme sonuçlarının niteliğini test etmek hedefiyle geliştirilen ve ortaya konulan bazı teknikler bulunmaktadır. Söz konusu tekniklerden birisi de, çalışmamızda incelenecek olan ve çok değişkenli bir istatistik analiz tekniği durumunda bulunan diskriminant analizinin kümeleme analizi sonuçlarına uygulanmasıdır.



3. DİSKRİMİNANT ANALİZİ

Literatürde diskriminant analizi ile ilgili ilk çalışma, doğrusal diskriminant fonksiyonu kavramını literatüre kazandıran Fisher tarafından 1936 yılında yapılmıştır. Söz konusu çalışmada Fisher, dağılımdan bağımsız bir teknik geliştirmiştir. Bu teknik, gözlemlerin alındığı sınıfların eşit varyanslara sahip normal dağılımlı olmaları ile örtüşmektedir.

Diskriminant analizi, birimlerin dahil edileceği grupların kaç tane olacağı ve hangi birimin hangi grupta var olduğunun önceden belirli olduğu için sıklıkla tercih edilen istatistiksel bir analiz tekniğidir. Yani diskriminant analizinin ana hedefi birimleri gruplandırmaktır. Diskriminant analizinin en önemli niteliklerinden bir tanesi seçilen sınıflardan herhangi birine yeni birimleri dahil etmek veya var olan elamanların hangi sınıftan geldiğini tespit etmektir. Diskriminant analizi çok boyutlu uzayda sınıflanma gerçekleştiren birimlerin doğru bir biçimde parçalanıp parçalanamayacağı, oluşturulacak sınıf adeti, her sınıfa dahil edilecek birimlerin seçilmesi, gruplandırma oranı ve sınıflara bölmeye katkı sağlayan niteliklerin hangileri olduğu gibi sorulara yanıt vermektedir. Var olan ya da ileride elde edilecek birimleri sınıflardan herhangi birine dahil etmek için bir ölçüt üretme tekniği olan diskriminant analizi, hatalı gruplandırma ihtimalini minimum seviyeye indirgeyerek birimleri ait oldukları sınıflara dahil etmektir (Şahin, 2017).

Diskriminant fonksiyonları olan matematiksel eşitlikler, analize dahil edilen grupların benzerlik düzeylerine göre sınıflandırılmasında kullanılmaktadır. Diğer bir deyişle diskriminant analizinde fonksiyonlar yardımıyla, birden fazla grubun benzerlik ve farklılıklarının belirlenmesi sağlanmaktadır. Buna ek olarak diskriminant analizinde, farklılığı sağlayan temel değişkenin ve gruplarda farklılaşmayı sağlayan faktörlerin belirlenmesi de sağlanmaktadır. Yapılan analiz sonuçlarının test edilmesi kapsamında, analiz sonucu yapılan gruplandırma ile gerçek grup üyeliklerinin karşılaştırılması yapılabilmektedir (Erçetin, 1993).

Diskriminant analizi ile gruplandırmalar tespit edilirken, belirli değişkenler baz alınarak bir değerlendirme yapılması da mümkün olmaktadır. Gruplardaki değişkenlerin etkileşim seviyelerinin ve bunların diğer değişkenlerle arasındaki farklılıkların düzeyi ortaya konulabilmektedir (Tümer, 2001; Cangül, 2006). Grupların birbirlerinden ayrılmasını sağlamak amacıyla kullanılan değişkenlere literatürde “diskriminant değişkenleri” adı

verilmektedir. Dolayısıyla diskriminant analizi yapılırken grup farklılıkları, diskriminant değişkenleri vasıtasıyla belirlenmektedir (Klecka, 1980; Cangül, 2006).

Diskriminant analizi, yanlış sınıflandırma ihtimalini azaltmak için oldukça etkili bir teknik sunmaktadır. Eğer bir değişken tahmin edilen bir grubun üyesi değilse, yanlış sınıflandırma yapıldığı değerlendirilmesi yapılmaktadır (Demirhan, 1997; Cangül, 2006). Bu analiz, araştırmacıların p tane özelliği üzerinden değerlendirilen değişkenleri, gerçek gruplarına en yakın şekilde atamasını sağlayacak fonksiyonlar belirlemelerini sağlayan bir tekniktir. Hata olasılığının en aza indirgenerek yapıldığı analizler de istatistiksel değerlendirmelerin de isabet oranı artmaktadır. Analizlerde bazı durumlarda analize dahil edilen gruplara ait bir üyeliğin belirlenemediği de söz konusu olabilmektedir. Analizlerde grup üyeliklerinin belirlenemediği durumların analizden dışlanması, analize dahil edilmemesi söz konusu olabilmektedir. Böyle bir durumla karşılaşılması durumunda analiz kapsamında belirlenen matematiksel denklemler ışığında ayrı bir grup oluşturulması mümkündür.

3.1. Diskriminant Analizinin Amacı ve Varsayımları

Diskriminant analizinin amaçlarını aşağıdaki gibi belirtmek mümkündür:

- Sınıf üyeliği hakkında öngöründe bulunmak yani bir birimin hangi değişken sınıfına dahil olacağını tespit etmek için kullanılabilir. Analiz yapılmadan belirlenen sınıflandırmalar, birimlerin sınıflandırılmasında etkin olabileceği düşünülen değişkenlere göre yapılmalıdır. Diskriminant analizi ile bahsedilen şekilde sınıflama yapılıp yapılamayacağına ya da sınıfların birbirinden farkının olup olmadığına karar verilmektedir. Bu doğrultuda diskriminant analizinin oldukça geniş kullanım yelpazesi vardır. Örnek verilecek olursa işletmecilikte pazar veya fiyat bölümlendirmesi konularında kullanıldığı gibi, eğitim, antropoloji, psikoloji, tıp, tarih, planlama, ekonomi, finans ve bankacılık gibi alanlarda da tercih edilmektedir.
- İki veya daha fazla sayıdaki sınıfı birbirinden ayıran en önemli değişkenleri tespit etmektedir.
- Değişkenlerin ortalamalarının sınıflar arasında nasıl farklılık gösterdiğini bulmak için kullanılabilir. Yani ayırt edici değişkenlere göre birimleri (iller, ülkeler, firmalar, ürünler, bireyler vb.) diskriminant değerleriyle gruplandırmak amacıyla teknikler üretmektir. Diskriminant analizi, sınıflar arası varyans ve sınıflar içi varyans oranını en

üst seviyede tutarak bu hedefi gerçekleştirir. Varyans oranının en üst seviyede olması, sınıflar arası varyansın maksimum, sınıflar içi varyansın ise minimum olması halinde mümkündür. Böyle bir durumda birimler birbirlerinden farklı olacak biçimde sınıflandırılmaktadır. Diskriminant analizi farklılaşan birimlerin tespit edilmesine de, bu şartı sağlayan birimlere ait değişkenlerin saptanmasına da yardım etmektedir.

- Amaç, değişkenler aracılığıyla biçimlendirilen sınıflar arasında, farklılık boyutlarının bütünü ve miktarını tespit etmektir. Böylece diskriminant analizi ile aynı niteliklerin bulunduğu yeni birimin hangi sınıfa dahil edileceği saptanabilmektedir. Yeni birimin dahil edileceği sınıfı bulmada, bir ön inceleme ile tespit edilen ve sınıfları parçalamada kullanılan diskriminant fonksiyonlarından faydalanılır.

Yukarıda bahsedilen ilk üç hedefi uygulamak için yapılan analize, grupları tanımlama amaçlı diskriminant analizi; dördüncü hedefi uygulamak için yapılan analize ise karar verme amaçlı diskriminant analizi denir. Bahsedilen hedeflerden de anlaşılacağı üzere; diskriminant analizi, sınıf farklılıklarını incelemekte ve birimleri sınıflara dahil etmektedir. Bu sebeple diskriminant analizi bir çeşit profil analizi ya da analitik tahmin metodu olarak da bilinebilir (Girginer ve Kamışlı, 2010; Yapıcı, 2015).

Diskriminant analizi, birbirine girmiş ortak özelliklere sahip grupların birbirinden ayrılması için kullanılan grup ortalama vektörlerini birbirinden ayıracak fonksiyonlar için geliştirilen bir teknik olmakla birlikte, önemli bazı varsayımları bulunmaktadır. Analizin farklılık gücü dayandığı varsayımların sağlanmasına veya bu varsayımlara karşı sağlam durmasına bağlıdır. Özellikle modelin başarı oranının umulandan az olduğu durumlarda, doğru yorumlama yapabilmek için bu varsayımların test edilmesi gerekir. Diskriminant analizinin varsayımlarını aşağıdaki gibi açıklamak mümkündür:

g = sınıf sayısı

p = değişken sayısı

n_i = i . sınıftaki birim sayısı

n = tüm sınıflardaki birim sayısı olmak üzere varsayımlar;

- Ana kütle belirli niteliklere göre sınıflanabilir. Birbirleri arasında fark bulunan birden fazla sınıf söz konusu olmalıdır. $g \geq 2$.
- Her sınıfta minimum iki durum olmalıdır: $n_i \geq 2$.

- Birimler ana kütleden rastgele olarak seçilmiştir.
- Özel formüller kullanılmadığı sürece her bir sınıf için kovaryans matrisleri hemen hemen eşittir. Sınıfların ortalamaları ve kovaryans matrisi daha önce bilinir. Sınıfların kovaryans matrisleri eşittir. Bu varsayım gerçekleşmediği takdirde, diskriminant analizinin karesel şekli kullanılabilir.
- Her bir sınıf, değişkenler üzerindeki çok değişkenli dağılıma sahip bir örnek sınıftan seçilir ve değişkenlerin çok boyutlu normal dağılımı vardır.
- Sınıfların birim sayılarının eşit olmadığı durumlarda, üyelerin tahmin edilen olasılıklarının bilindiği kabul edilir.
- Herhangi bir durumun yanlış sınıflandırılmasının maliyeti önceden bellidir (Cangül, 2016).

Diskriminant analizi varsayımları arasında yer alan grup varyans-kovaryans matrislerinin homojen olması durumu gerçekleşmediği zaman, doğrusal diskriminant analizi yerine karesel diskriminant analizi kullanılarak diskriminant fonksiyonlarının hesaplanması yapılmalıdır. Karesel diskriminant analizinde iki ayrım bölgesi arasındaki sınır doğrusal değil, karesel bir yüzey alanına sahiptir. Kovaryans matrisleri arasındaki fark çok fazla olmadığında, buna ek olarak veri setinin sahip olduğu birim sayısı orta ya da az düzeylerde olduğunda doğrusal diskriminant analizinin daha gerçeğe yakın sonuçlar verdiğini söylemek mümkündür.

3.2. Diskriminant Analizinin Adımları

Diskriminant analizinin adımlarını aşağıdaki gibi belirtmek mümkündür:

- Önsel grup üyelikleri belirlenir.
- Değişkenler için sınıflar arasında fark olup olmadığı, Wilks'in istatistiği ile tespit edilir. Bu doğrultuda yapılacak test sonucunda sınıflar arasında kayda değer bir fark bulunuyorsa analiz sürdürülür. Şayet kayda değer bir fark tespit edilmezse tüm sınıfların ortalamalarının aynı olduğu, yani sınıf farkı bulunmadığı söylenebilir. Böyle bir durumda diskriminant analizi kullanılamaz.
- Kullanılacak değişkenler tespit edilir. Değişken tercihinde önsel bilgi veya istatistiki teknikler uygulanabilir.

- Değişkenler arasında çoklu bağlantının varlığı araştırılır. Bu amaçla birleştirilmiş sınıf içi korelasyon matrisi incelenir. Bu matristeki korelasyon değerleri mutlak değerce % 75'den çok ise değişkenlerden bir bölümünün çıkarılması gerekir. Bu adımın sonunda değişken kümesi tespit edilmiş olur.
- $W^{-1}B$ matrisinin temel değerleri ve bu temel değerlere ilişkin temel vektörler bulunur. Bu temel vektörler, diskriminant fonksiyonları için ihtiyaç duyulan ağırlıkları verir.
- Diskriminant fonksiyonlarının anlamlılık testi de bu temel değerlerin kullanılmasıyla uygulanır. Eğer fonksiyonlardan biri anlamlıysa yaptığı ayrımın başarıya ulaştığı söylenebilir.
- Standartlaştırılmamış diskriminant fonksiyonu ile her birim için diskriminant fonksiyonu değerleri bulunur. Bu değerler sınıflandırma aşamasında kullanılacaktır.
- Sınıf üyelikleri için önsel olasılıklar saptanır. Sonrasında bu ihtimaller ve diskriminant değerleri kullanılarak sonsal ihtimaller saptanır. Bireyde bulunan maksimum sonsal olasılık bulunur. Bu ihtimali veren sınıfın o birimin ait olduğu sınıf tahmin edilir ve birim sınıflandırılmış olur.
- Her bir birim sınıflandırıldıktan sonra, diskriminant fonksiyonunun başarılı olup olmadığı doğru sınıflandırma oranı incelenerek saptanır (Yaprak, 2007).

3.3. Diskriminant Fonksiyonlarının Belirlenmesi

Diskriminant analizinde birimlerin genel ortalamalardan sapma miktarlarının karelerinin toplamı olan genel kareler toplamı ve çapraz çarpımlar matrisi, gruplar arası kareler toplamı ve çapraz çarpımlar matrisi ile grup içi kareler toplamı çapraz çarpımlar matrislerinin toplamlarından meydana gelmektedir.

Genel kareler toplamı ve çapraz çarpımlar matrisi "T" ile gösterilmektedir.

$$T = [t_{ij}]_{p \times p} \quad (3.1)$$

olarak ifade edilmektedir. Söz konusu matrisin birimleri aşağıda sunulan formüle göre hesaplanmaktadır:

$$t_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_i)(x_{jkm} - \bar{x}_j) \quad i, j=1,2,\dots,p \quad (3.2)$$

Yukarıdaki formüldeki değişkenlerden, “g” grup sayısını; “ x_{ikm} ” i. değişkenin, k. grupta, m. birey için değerini; “ n_k ” k. gruptaki birey sayısını; \bar{x}_i i. değişkenin genel ortalamasını ve son olarak “p” değişken sayısını ifade etmektedir.

N, tüm gruplardaki toplam birey sayısını göstermektedir. T matrisinin her bir biriminin N-1 değerine bölünmesi sonucunda “genel kovaryans matrisi” adı verilen Σ_T değeri elde edilmektedir. Bu matrisin ana köşegenini oluşturan birimleri, varyansı; geriye kalan birimleri ise kovaryansı göstermektedir. Kovaryanslara bakarak iki değişkenin birlikte ne ölçüde değiştiğine dair yorumlar yapılabilmektedir.

T matrisinin her bir biriminin, aynı sıra ve sütunundaki köşegen birimlerinin çarpımlarının karekökünün alınmasıyla “genel korelasyon matrisi (R)”ne ulaşılmaktadır. Bu matrisin birimleri aşağıdaki formülle bulunmaktadır:

$$r_{ij} = \frac{t_{ij}}{\sqrt{t_{ii}t_{jj}}} \quad i, j=1, \dots, p \quad (3.3)$$

Grup içi kareler toplamı çapraz çarpımlar matrisi (W)’ne ait birimler ise aşağıdaki şekilde hesaplanmaktadır:

$$w_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_{ik})(x_{jkm} - \bar{x}_{jk}) \quad i, j=1, 2, \dots, p \quad (3.4)$$

Benzer şekilde burada da, “g” grup sayısını, “ x_{ikm} ” i. değişkenin, k. grupta, m. birey için değerini; “ n_k ” k. gruptaki birey sayısını; \bar{x}_{ik} , k. gruptaki i. değişkenin genel ortalamasını ve son olarak p değişken sayısını ifade etmektedir. Eğer W matrisinin her bir birimi (N-g) değerine bölünürse, grup kovaryans matrislerinin ağırlıklı ortalamalarına, başka bir deyişle “grup içi kovaryans matrisi (S)” değerine ulaşılmaktadır.

Yukarıda açıklanan T ve W matrisleri arasındaki fark ise “gruplar arası kareler toplamı ve çapraz çarpımlar matrisi” olarak ifade edilen ve “B” ile gösterilen matrisi vermektedir. W ve B matrislerinin, hem grup içi hem de gruplar arası ilişkilere ait bilgileri içermesi sebebiyle, gruplar arası farklılıkların belirlenmesinde bir ölçüt olarak kullanılabilir. Bu ölçüt, iki matrisin varyanslarının birbirlerine oranlanmasıyla elde edilen λ katsayısıdır.

$$\lambda = \frac{v'Bv}{v'Wv} \quad (3.5)$$

Yukarıda gösterilen λ değerini maksimum yapan değer, ölçüt olarak alınmaktadır. λ 'nın köklerini yani özdeğerleri veren, her bir λ değeri için v vektörüne ulaşılmaktadır. λ 'nın i 'inci kökü için bulunan vektör v_i ile gösterilirse, bunlara ait i ' inci fonksiyon (ya da diskriminant fonksiyon) aşağıdaki şekilde gösterilmektedir:

$$f_i = v_i'X \quad (3.6)$$

Dolayısıyla p tane değişken söz konusu olduğunda, k sayıda grup ve m sayıda birey için fonksiyon gösterimi aşağıdaki şekilde olmaktadır:

$$f_{km} = v_1X_{1km} + v_2X_{2km} + \dots + v_pX_{pkm} \quad (3.7)$$

Yukarıdaki fonksiyondan elde edilen v_i değerleri “diskriminant fonksiyonlarının ham katsayıları” olarak ifade edilmektedir. Fonksiyonların yorumlanmasını daha kolay hale getirmek için bazı dönüştürmeler ile yeni katsayılar elde edilmesi mümkündür. Bahse konu yeni katsayılar, aşağıda sunulan dönüştürme formülleri yardımıyla bulunabilmektedir:

$$u_i = v_i\sqrt{N-g} \quad (3.8)$$

$$u_0 = -\sum_{i=1}^p u_i\bar{x}_i \quad (3.9)$$

Katsayılardaki dönüştürme işlemi sonrası elde edilen yeni katsayılar sayesinde diskriminant fonksiyonlarının eksen merkezinde değişme meydana gelmektedir. Tüm diskriminant fonksiyonları eksenlerinin sıfır olduğu, başka bir deyişle sistemin merkezi kabul edilen nokta yeni eksek merkezi haline gelmektedir. Bu durum, bireylerin diskriminant fonksiyonundaki değerlerine bakılarak, sistemin merkezinden hangi yönde ve ne kadar uzaklıkta olduğunun da anlaşılmasını kolaylaştırmaktadır.

Daha önce ifade edilen standart olmayan katsayılar, her bir değişkenini sınıflandırma işlemine tek başına olan katkısını göstermektedir. Fakat, verilerin standardize edilerek

tekrar hesaplanması ile aşağıdaki dönüşüm formülü de kullanılarak standart nitelikteki katsayılarla ulaşılmaktadır (Cangül, 2006).

$$c_i = u_i \sqrt{\frac{w_{ii}}{N - g}} = v_i \sqrt{w_{ii}} \quad (3.10)$$

Yukarıdaki dönüşüm formülündeki değişkenlerden;

w_{ii} : i değişkeni için kareler toplamı

N : Tüm gruplardaki toplam birey sayısı

g : Grup sayısı

olarak ifade edilmektedir.

Sonuç olarak, diskriminant fonksiyonlarının değişkenlerle olan ilişkilerini gruplar bazında incelemek amacıyla; “toplam yapısal katsayılar” olarak ifade edilen aşağıda sunulan formül kullanılmaktadır:

$$c_{kj}^* = \frac{v_{kj} \sqrt{t_{kk}}}{\sqrt{\sum_{i=1}^g \sum_{m=1}^p v_{ij} v_{mj} t_{im}}} \quad k=1, \dots, p; j=1, \dots, r \quad (3.11)$$

Toplam yapısal katsayılar ise aşağıdaki formülle bulunmaktadır:

$$S_{ij} = \sum_{k=1}^p \frac{t_{ik} c_{kj}^*}{\sqrt{t_{ii} t_{kk}}} \quad i=1, \dots, p; j=1, \dots, r \quad (3.12)$$

3.4. Diskriminant Fonksiyonlarının Hatalı Sınıflandırma Olasılıkları

$X = (X_1, X_2, \dots, X_p)'$ elde edilmiş ölçüm değerlerine karşılık gelen p boyutlu, X vektörü, Π_i popülasyondan kaynaklanan rastgele bir vektörü temsil etmektedir. Bunlardan hareketle, olasılık yoğunluk fonksiyonu; $f_i(x)$ ile gösterildiğinde, ($i=1, 2, \dots, g$) yani vektörün alabileceği tüm değerler p boyutlu R^p örnek uzayında (tüm örneklemeleri içeren kümede), sınıflandırma problemi ile bu örnek uzay öyle (R_1, R_2, \dots, R_g) bölgelerine ayrılır ki, X vektörü herhangi bir i bölgesinde iken (R_i), X gözlem vektörünün elde edildiği birey G_i grubunda yer alacak şekilde sınıflandırılmaktadır (Cangül, 2006). Normalde G_i grubunda yer alan

bireyin doğru grup olan G_i de sınıflandırılmış olma olasılığı ise $P(i|i)$ ile gösterilmektedir. Bu olasılık fonksiyonu aşağıdaki şekilde ifade edilmektedir.

$$P(i|i) = P(G_i|G_i, R) = \int_{R_i} f_i(x) dx \quad (3.13)$$

$$dx = dx_1, dx_2, \dots, dx_p$$

Buna ek olarak normalde G_j grubuna ait olan bir bireyin yanlışlık yapılarak G_i grubuna dahil edilme olasılığı ise $P(j|i)$ ile gösterilmektedir.

$$P(j|i) = P(G_j|G_i, R) = \int_{R_j} f_i(x) dx = 1 - \int_{R_i} f_i(x) dx \quad (3.14)$$

$$dx = dx_1, dx_2, \dots, dx_p$$

Bu iki formülden hareketle bir G_i grubunun üyesinin yanlış bir grupta sınıflandırılma olasılığını $P(i)$ ile gösterirsek, aşağıdaki şekilde formüle edilmektedir.

$$P(i) = \sum_{j=1, j \neq i}^g P(j|i) = 1 - P(i|i) \quad (3.15)$$

Ancak burada q_i önsel olasılık olarak ifade edilen olasılıklar dahil edilmemiştir. Bu olasılıklar da göz önüne alındığında toplam hatalı sınıflandırma olasılığı aşağıdaki şekilde ifade edilmektedir:

$$P(R, f) = \sum_{i=1}^g q_i P(i) = 1 - \sum_{i=1}^g q_i P(i|i) \quad (3.16)$$



4. İLLERİN, ALDIĞI GÖÇÜN EĞİTİM DURUMUNA GÖRE KÜMELENMESİ ÜZERİNE BİR UYGULAMA

4.1. Göç Kavramı ve Türkiye’de İç Göç

Göç, yapılan bir tanıma göre canlıların hemen hepsi için geçerli ve hayatlarını devam ettirebilmeleri için önemli bir yer değiştirme eylemidir (Yalçın, 2004). Bir diğer tanımda ise göç, bireysel olarak veya grupların ekonomik, eğitim, kültürel vb. sebeplerden ötürü bir yerden başka bir yere gitmeleri şeklinde açıklanmıştır (Kızılcelik ve Erjem, 1992). Göçün birçok değişik tanımı yapılabilmekle birlikte, en genel biçimde insanların bir yerden bir başka yere farklı nedenlerden dolayı hareket etme işlevidir. Erder (1986), bu hareket etme işlevinin göç olarak anlamlandırılabilmesi için kayda değer uzaklık ile birlikte etki oluşturacak bir zaman dilimi içerisinde meydana gelmesinin önemini belirtmiştir.

Göç tanımını sadece belli bir çerçeveye göre açıklamanın çok doğru olmadığı değerlendirilmektedir. Bu sebeptir ki, göç kavramı genel olarak üç temel başlıkta açıklanmaktadır. Bunlar; göç hareketi içinde bulunan bireylerin iradelerine göre (zorunlu ve gönüllü göçler), göç edilen yerde kalma sürelerine göre (kalıcı ve geçici göçler) ve göç edilen yerlerin konumlarına göre (dış ve iç göçler) olan göç biçimleridir (Perruchoud, 2009).

Çalışmada incelenen konunun iç göç hareketi içinde yer alması dolayısıyla yukarıda bahsedilen göç biçimlerinden iç göç şekli üzerinde daha fazla durulması gerekmektedir. İç göçü en basit düzeyde açıklamak gerekirse, bir ülke sınırları içinde gerçekleşen göç hareketi şeklinde belirtmek mümkündür. Bu kapsamda Tekeli (1998), iç göç kavramını süre, mekan ve irade hususlarını belirterek açıkladığı tanım oldukça önemlidir. Söz konusu tanıma göre iç göç, belirli bir süre içerisinde belli bir yerde yaşayan bireylerin, iradelerine bağlı olarak yaşam yerlerini bahse konu yerleşme alanının dışı olarak değiştirenlerin miktarı olarak tanımlanmıştır.

Türkiye’de, bilhassa İkinci Dünya Savaşı ile birlikte 1950’lerde başlayan, 1960’larda hızlanma eğilimi gösteren ve 1980’li yıllarla beraber sanayileşme faaliyet ve etkinlikleri ile paralel olarak kayda değer bir iç göç aktivitesi yaşanmıştır. Bununla birlikte, başta İstanbul ve yakın çevresi olmak üzere sanayide pilot bölgeler tespit edilmiş ve bu bölgelere sanayi

faaliyetlerin başlaması sonrasında, kırsal kesimden göç hareketi başlamıştır. Sonrasında ise; tarımda modernleşme sonucu işsizlik, eğitim-sağlık, doğal afetler ve hızlı nüfus artışı gibi nedenler iç göçlere ivme kazandıran yeni etmenleri oluşturmuştur. Göçler ilk başlarda köyden kente doğru olurken, daha sonraları da küçük ile orta ölçekli kentlerden büyük kentlere doğru olarak devam etmiş ve bu durum İstanbul, Ankara ve İzmir gibi şehirlerin aşırı büyümesine neden olmuştur (Bostan, 2017).

Ülkemizde iç göç hareketi içinde yer alan nüfus incelendiğinde, Türkiye İstatistik Kurumu 2017 yılı iç göç verilerine göre 2.684.820 kişi yani toplam nüfusun % 3,3'ü iç göç hareketi içinde bulunmaktadır. Bu nüfusun yaklaşık olarak % 51'i kadınlardan meydana gelmektedir. Ayrıca, söz konusu nüfus içinde yaş gruplarına baktığımızda ise % 21,3 ile en çok orana sahip yaş grubunun 20-24 yaş arası olduğu dikkat çekmektedir.

Bununla birlikte, bahse konu iç göç hareketi içinde yer alan kitlelerin ise eğitim durumunun araştırılıp, ortaya konması demografik açıdan büyük önem taşımakta olup, göçün önemli bir açıklayıcı faktörünün okur-yazarlık ve eğitim durumu olduğunu ortaya koymak gerekmektedir. Bu bağlamda; göç alan illere yönelen nüfusun okur-yazarlık oranı, okur-yazar nüfusun eğitim durumu gibi göstergeler söz konusu alanda tespitlerin ortaya konulması bakımından değer taşımaktadır. Ayrıca, nüfusun öğrenim düzeyini belirleyen bu hususların iller arası dağılımlarının da dikkate alınması gerekir. Böylece incelenen dönem içerisinde göç hareketi içinde yer alan nüfusun daha çok hangi eğitim seviyesinden oluştuğu ve eğitim durumları genelinde hangi illerin benzer özellik gösterdiği yorumunu yapabilme imkanı sağlayacaktır.

4.2. İllerin, Aldığı Göçün Eğitim Durumu Bakımından Kümelmesi

Çalışmanın bu uygulama bölümünde, kümeleme analizinde kullanılan bazı uzaklık ölçüleri farklı kümeleme teknikleri üzerinde uygulanarak oluşan kümeler karşılaştırılmış ve ortaya çıkan bu kümelerin her biri için diskriminant analizi uygulanarak söz konusu meydana gelen kümelerin geçerliliği sorgulanmıştır. Bu hususta, hangi uzaklık ölçüsünün hangi kümeleme tekniğinde en iyi sonucu verdiği ortaya konulmaya çalışılmıştır.

Uygulama olarak gerçek bir veri seti olan ve söz konusu alanda güncel son veri olan, TÜİK 2017 yılı 15 yaş ve üzeri iç göç istatistiklerindeki veriler kullanılmıştır. Söz konusu

verilere ilişkin olarak 81 il, göç hareketi içinde yer alanların eğitim durumlarına göre okuma-yazma bilmeyen, ilkokul, ortaokul, lise ve dengi meslek okulu ve lisans ve üstü eğitim durumu kategorilerinde olmak üzere 5 değişken bakımından incelenmiştir. Çok değişkenli istatistiksel analiz tekniklerinden kümeleme analizi ve diskriminant analizi, bahse konu verilere uygulanarak SPSS Paket Programı ve Minitab Paket Programında gerçekleştirilmiştir.

Ayrıca, her bir il için aldığı göçe göre ilgili değişken özelinde yüzdesel ağırlıklar hesaplanarak elde edilen veriler kullanılmış ve analizler gerçekleştirilmiştir. Böylece, illere göç edenlerin yüzdesel ağırlıklı olarak hangi eğitim seviyesine sahip olduğu her bir il bazında belirlenmiştir.

Küme sayısına karar vermek için, $k = (n/2)^{1/2}$ formülünden yararlanılmış ve küme sayımız $k = (81/2)^{1/2}$ eşitliğinden $k \cong 7$ olarak belirlenmiş olup elde edilen bu küme sayısına göre iller göç hareketi içinde yer alanların eğitim durumuna göre aldığı göç bakımından 7 kümeye ayrılarak kümeleme analizi gerçekleştirilmiştir.

Çizelge 4.1. Betimleyici istatistikler

Eğitim Durumu	N	Aralık	En küçük	En büyük	Ortalama
Okuma_Yazma_Bilmeyen	81	3,8	,5	4,3	1,740
İlkokul	81	16,4	4,2	20,6	10,090
Ortaokul	81	20,5	9,3	29,8	16,522
Lise_Dengi_Meslek_Okulu	81	37,1	23,4	60,5	40,632
Lisans_ve_üstü	81	33,9	17,9	51,8	31,021
Geçerli N	81				

Betimleyici istatistiklere bakılarak illere göç edenlerin eğitim durumu genel olarak incelendiğinde, ortalama olarak göç edenlerin % 1,7'sinin okuma-yazma bilmeyen, % 10,1'inin ilkokul, % 16,5'inin ortaokul, % 40,6'sının lise ve dengi meslek okulu ve % 31'inin lisans ve üstü mezunu olduğu görülmektedir. % 60,5 ile en fazla lise ve dengi meslek okulu nüfusun göç ettiği il olarak Karabük ili dikkat çekerken, % 51,8 ile kendi içinde en fazla lisans ve üstü nüfus ağırlığının göç ettiği il olarak Şırnak dikkat çekmektedir. Nüfus ağırlığı olarak en fazla ortaokul mezununun göç ettiği il ise Tekirdağ olup, Gümüşhane ili % 17,9 ile ağırlık olarak en az lisans ve üstü nüfus göç ettiği il olarak dikkat çekmektedir.

Çizelge 4.2. İllerin, aldığı göçün eğitim durumuna göre kümeleme analizi ANOVA tablosu

Eğitim Durumu	Küme		Hata		F	p-değeri
	Ortalamanın Karesi	sd	Ortalamanın Karesi	sd	Ortalamanın Karesi	sd
Okuma_Yazma_Bilmeyen	4,876	6	,291	74	16,765	,000
İlkokul	110,916	6	2,884	74	38,462	,000
Ortaokul	118,134	6	4,009	74	29,464	,000
Lise_Dengi_Meslek_Okulu	821,793	6	8,567	74	95,925	,000
Lisans_ve_üstü	564,880	6	8,910	74	63,397	,000

ANOVA sonuçları, değişkenlerin kümelemede etkili olup olmadığını belirleme noktasında önemli durumdadır. Bunun için gerçekleştirilen analiz sonucunda Çizelge 4.2.'deki değerler karşımıza çıkmaktadır ve değişkenlerin kümelemede önemli olup olmadığı noktasında;

H_0 : Değişkenler kümelemede etkili değildir.

H_s : Değişkenler kümelemede etkilidir.

hipotezi kurularak değişkenler sorgulanmıştır.

Söz konusu hipoteze karşılık olarak alınan göçün eğitim durumu bakımından illerin kümelendirilmesinde bütün değişkenlerin önemli olduğu saptanmıştır ($p < 0,05$).

4.2.1. Tek bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması

Çizelge 4.3. Tek bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler

Küme No	İller
1	Diğer İller
2	Çankırı, Giresun, Kastamonu, Sinop
3	Gümüşhane
4	Hakkari
5	Şırnak
6	Sivas
7	Tekirdağ, Tokat

Tek bağlantı kümeleme tekniğinde karesi alınmış Öklid uzaklığı uygulanarak oluşturulan kümelerde Gümüşhane, Hakkari, Şırnak ve Sivas illeri tek başına birer küme oluştururken Tekirdağ ve Tokat illeri beraber bir küme oluşturmuşlardır. Bununla birlikte, coğrafi açıdan benzer nitelikte olarak tanımlayabileceğimiz Çankırı, Giresun, Kastamonu ve Sinop illeri de birlikte bir küme oluşturmuştur. Kalan diğer iller ise bir kümede toplanmıştır.

Çizelge 4.4. Box-M test sonucu

Box's M		112,312
F	Yaklaşık	1,679
	sd1	40
	sd2	1407,840
	p-değeri	,007

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s: En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den küçük olarak bulunmuş ve yokluk hipotezi reddedilmiştir. Bu durumda, grupların varyans-kovaryans matrisleri eşit olmadığından karesel diskriminant analizi kullanılmış ve elde edilen kümelerin doğru sınıflandırma oranları bu şekilde belirlenmiştir.

Çizelge 4.5. Tek bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Tek Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	70	0	1	0	0	0	0	71
2	0	4	0	0	0	0	0	4
3	0	0	1	0	0	0	0	1
4	0	0	0	1	0	0	0	1
5	0	0	0	0	1	0	0	1
6	0	0	0	0	0	1	0	1
7	0	0	0	0	0	0	2	2
1	98,6	,0	1,4	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 98,8'i doğru sınıflandırılmıştır.

Tek bağlantı kümeleme tekniğine karesi alınmış Öklid uzaklığı uygulanarak elde edilen bu kümeleri, diskriminant analizi ile sorguladığımızda % 98,8 oranında doğru sınıflandırma görülmüştür. Yalnızca 1. kümede yer alan Ordu ilinin 3. kümede yer alması gerektiği EK-2'deki tabloya göre belirlenmiştir. Bundan sonraki diğer tüm tahmini grup üyelikleri de bu şekilde belirlenmiş olup, örnek teşkil etmesi açısından EK-2 verilmiştir.

Çizelge 4.6. Tek bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler

Küme No	İller
1	Diğer İller
2	Çankırı, Giresun, Kastamonu, Sinop
3	Gümüşhane
4	Kocaeli
5	Yalova, Ordu
6	Tekirdağ
7	Tokat

Tek bağlantı kümeleme tekniğine Pearson uzaklığı uygulanarak meydana gelen kümelere karesi alınmış Öklid uzaklığına göre oluşturulan kümelere farklı olarak Tekirdağ ve Tokat illeri ayrı birer küme oluşturmuşlardır. Gümüşhane ili benzer şekilde ayrı bir küme olarak yer alırken Yalova ve Ordu illeri ise beraber bir küme oluşturmuşlardır.

Çizelge 4.7. Box-M test sonucu

Box's M	102,103
F	Yaklaşık 1,464
	sd1 30
	sd2 1415,310
	p-değeri ,010

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den küçük olarak bulunmuş ve yokluk hipotezi reddedilmiştir. Bu durumda, grupların varyans-kovaryans matrisleri eşit olmadığından karesel diskriminant analizi kullanılmış ve elde edilen kümelerin doğru sınıflandırma oranları bu şekilde belirlenmiştir.

Çizelge 4.8. Tek bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Tek Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	70	0	0	0	1	0	0	71
2	0	4	0	0	0	0	0	4
3	0	0	1	0	0	0	0	1
4	0	0	0	1	0	0	0	1
5	0	0	0	0	2	0	0	2
6	0	0	0	0	0	1	0	1
7	0	0	0	0	0	0	1	1
1	98,6	,0	,0	,0	1,4	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 98,8'i doğru sınıflandırılmıştır.

Tek bağlantı kümeleme tekniğine Pearson uzaklığı uygulanarak elde edilen bu kümeleri, diskriminant analizi ile sorguladığımızda % 98,8 oranında doğru sınıflandırma yaptığı görülmekle birlikte 1. kümede yer alan İstanbul ilinin 5. kümede yer alması gerektiği ortaya çıkmıştır.

Çizelge 4.9. Tek bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler

Küme No	İller
1	Diğer İller
2	Çankırı, Giresun, Kastamonu, Sinop
3	Gümüşhane
4	Hakkari
5	Şırnak
6	Sivas
7	Tekirdağ, Tokat

Çizelge 4.10. Box-M test sonucu

Box's M		112,312
F	Yaklaşık	1,679
	sd1	40
	sd2	1407,840
	p-değeri	,007

$H_0: \Sigma_1=\Sigma_2=\Sigma_3=\Sigma_4=\Sigma_5=\Sigma_6=\Sigma_7$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den küçük olarak bulunmuş ve yokluk hipotezi reddedilmiştir. Bu durumda, grupların varyans-kovaryans matrisleri eşit olmadığından karesel diskriminant analizi kullanılmış ve elde edilen kümelerin doğru sınıflandırma oranları bu şekilde belirlenmiştir.

Çizelge 4.11. Tek bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Tek Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	70	0	1	0	0	0	0	71
2	0	4	0	0	0	0	0	4
3	0	0	1	0	0	0	0	1
4	0	0	0	1	0	0	0	1
5	0	0	0	0	1	0	0	1
6	0	0	0	0	0	1	0	1
7	0	0	0	0	0	0	2	2
1	98,6	,0	1,4	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 98,8'i doğru sınıflandırılmıştır.

Çizelge 4.14. (devam) Tek bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

1	98,6	,0	1,4	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 98,8'i doğru sınıflandırılmıştır.

Tek bağlantı kümeleme tekniğine Manhattan ve Minkowski uzaklıkları uygulanarak meydana gelen kümeler karesi alınmış Öklid uzaklığı uygulanarak elde edilen kümeler ile aynı olup, söz konusu kümelere uygulanan diskriminant analizi sonuçları da % 98,8 ile aynı çıkmıştır.

Çizelge 4.15. Tek bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler

Küme	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
1	Diğer İller	Diğer İller	Diğer İller	Diğer İller
2	Çankırı, Giresun, Kastamonu, Sinop	Çankırı, Giresun, Kastamonu, Sinop	Çankırı, Giresun, Kastamonu, Sinop	Çankırı, Giresun, Kastamonu, Sinop
3	Gümüşhane	Gümüşhane	Gümüşhane	Gümüşhane
4	Hakkari	Kocaeli	Hakkari	Hakkari
5	Şırnak	Yalova, Ordu	Şırnak	Şırnak
6	Sivas	Tekirdağ	Sivas	Sivas
7	Tekirdağ, Tokat	Tokat	Tekirdağ, Tokat	Tekirdağ, Tokat

Çizelge 4.16. Tek bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları

Tek Bağlantı Kümeleme Tekniği	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
Diskriminant Analizi Doğru Sınıflandırma Oranı(%)	98,8	98,8	98,8	98,8

Tek bağlantı (en yakın komşuluk) kümeleme tekniğine uygulanan uzaklık ölçülerine genel olarak baktığımızda Pearson uzaklığından farklı olarak karesi alınmış Öklid, Manhattan ve Minkowski uzaklıklarının aynı sonucu verdiği görülmektedir. Gümüşhane ili tüm uzaklık kümelerinde ayrı bir küme olarak yer alırken Çankırı, Giresun, Kastamonu ve Sinop illeri

yine tüm uzaklıklarda beraber olarak bir küme oluşturmuşlardır. Bununla birlikte, tüm uzaklıklara göre elde edilen kümeleri karesel diskriminant analizi ile sorguladığımızda % 98,8 oranında doğru sınıflandırma yapıldığı belirlenmiştir.

4.2.2. Tam bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması

Çizelge 4.17. Tam bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler

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

Tam bağlantı (en uzak komşuluk) kümeleme tekniğine karesi alınmış Öklid uzaklığı uygulanarak meydana gelen kümeler incelendiğinde Hakkari ve Şırnak illeri ile Tekirdağ ve Tokat illerinin ayrı birer küme oluşturduğu belirlenmiştir. Ayrıca, Hatay hariç coğrafi olarak benzer konumda bulunan doğu illerinin birlikte bir kümede toplandığı dolayısıyla bu illere göç edenlerin eğitim durumlarının benzer nitelikte olduğu ortaya çıkmıştır.

Çizelge 4.18. Box-M test sonucu

Box's M	133,166
F	Yaklaşık 2,885
	sd1 40
	sd2 7388,851
	p-değeri ,073

$$H_0: \Sigma_1=\Sigma_2=\Sigma_3=\Sigma_4=\Sigma_5=\Sigma_6=\Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.19. Tam bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Tam Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	14	0	0	0	0	0	0	14
2	0	20	0	0	0	0	0	20
3	1	0	9	0	0	0	0	10
4	1	1	0	13	0	0	0	15
5	0	3	0	0	15	0	0	18
6	0	0	0	0	0	2	0	2
7	0	0	0	0	0	0	2	2
1	100,0	,0	,0	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	10,0	,0	90,0	,0	,0	,0	,0	100,0
4	6,7	6,7	,0	86,7	,0	,0	,0	100,0
5	,0	16,7	,0	,0	83,3	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 92,6'sı doğru sınıflandırılmıştır.

Söz konusu kümeleme tekniği ile elde edilen kümelerin doğruluğunu diskriminant analizi ile sorguladığımızda % 92,6 oranında doğru sınıflandırma yapıldığı görülmektedir. Bu kapsamda, 3. kümede yer alan Hatay ilinin 1. kümede yer alması gerektiği belirlenmiş olup, küme gruplarında yaptığımız yorum ile örtüşmektedir. Ayrıca, 4. kümede yer alan Yozgat ilinin 1. kümede, yine 4. kümede yer alan Manisa ilinin 2. kümede ve 5. kümede yer alan Bilecik, Nevşehir, Kırşehir illerinin 2. kümede yer alması gerektiği diskriminant analizi sonucu ortaya çıkmıştır.

Çizelge 4.20. Tam bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler

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

Tam bağlantı kümeleme tekniğine Pearson uzaklığı uygulanarak elde edilen kümeler, karesi alınmış Öklid uzaklığı uygulanarak elde edilen kümelere göre daha homojen bir şekilde dağılma göstermiştir. Coğrafi olarak benzer nitelikte olan illerden Ordu hariç Bursa, İstanbul, Kocaeli ve Yalova birlikte bir kümeyi oluştururken yine benzer coğrafi özellikte olan iller 6. kümede bir araya gelmiştir. Bu çerçevede, yine benzer coğrafi özelliklere sahip illere göç edenlerin eğitim durumunun benzer nitelikte olduğu yorumu getirilebilir. Ayrıca, Tekirdağ ve Tokat illeri de yine karesi alınmış Öklid uzaklığında olduğu gibi birlikte bir kümede yer almıştır. Bununla birlikte, yine benzer coğrafi özellikte olan birçok doğu bölgesi illeri bir kümede toplanmıştır.

Çizelge 4.21. Box-M test sonucu

Box's M	119,266
F	Yaklaşık 1,774
	sd1 50
	sd2 1293,821
	p-değeri ,065

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.22. Tam bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Tam Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	16	1	1	0	0	0	0	18
2	0	19	0	0	0	0	0	19
3	0	0	10	0	0	0	0	10
4	0	0	0	22	0	0	0	22
5	0	0	0	0	5	0	0	5
6	0	0	0	0	0	5	0	5
7	0	0	0	0	0	0	2	2
1	88,9	5,6	5,6	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 97,5'i doğru sınıflandırılmıştır.

Bahse konu kümeleme tekniğine göre elde edilen kümelerin diskriminant analizi ile geçerliliğine bakıldığında oluşturulan kümelerin % 97,5 oranında doğru sınıflandırıldığı belirlenmiştir. Yalnızca, 1. kümede yer alan Balıkesir ilinin 2. kümede ve yine 1. kümede yer alan Iğdır ilinin 3. kümede yer alması gerektiği tahmin edilmiştir.

Çizelge 4.23. Tam bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler

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

Çizelge 4.24. Box-M test sonucu

Box's M		97,707
F	Yaklaşık	2,055
	sd1	40
	sd2	4010,448
	p-değeri	,053

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.25. Tam bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Tam Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	16	0	0	0	0	0	0	16
2	0	28	0	0	0	0	0	28
3	0	0	8	0	0	0	0	8
4	2	0	0	13	0	0	0	15
5	0	1	0	0	9	0	0	10
6	0	0	0	0	0	2	0	2
7	0	0	0	0	0	0	2	2
1	100,0	,0	,0	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	13,3	,0	,0	86,7	,0	,0	,0	100,0
5	,0	10,0	,0	,0	90,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 96,3'ü doğru sınıflandırılmıştır.

Tam bağlantı kümeleme tekniğine Manhattan uzaklığı uygulanarak elde edilen kümeler diskriminant analizi ile test edildiğinde % 96,3 oranında doğru sınıflandırma yapıldığı belirlenmiştir. Bununla birlikte, 4. kümede yer alan Manisa ve Yozgat illerinin 1. kümede yer alması ve 5. kümede yer alan Çanakkale'nin 2. kümede yer alması gerektiği tahmin edilmiştir.

Çizelge 4.26. Tam bağlantı kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler

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

Çizelge 4.27. Box-M test sonucu

Box's M	133,166
F	Yaklaşık 2,885
	sd1 40
	sd2 7388,851
	p-değeri ,073

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.28. Tam bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Tam Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	14	0	0	0	0	0	0	14
2	0	20	0	0	0	0	0	20
3	1	0	9	0	0	0	0	10
4	1	1	0	13	0	0	0	15
5	0	3	0	0	15	0	0	18
6	0	0	0	0	0	2	0	2
7	0	0	0	0	0	0	2	2
1	100,0	,0	,0	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	10,0	,0	90,0	,0	,0	,0	,0	100,0
4	6,7	6,7	,0	86,7	,0	,0	,0	100,0
5	,0	16,7	,0	,0	83,3	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 92,6'sı doğru sınıflandırılmıştır.

Tam bağlantı kümeleme tekniğine uygulanan Minkowski uzaklığına göre elde edilen kümeler karesi alınmış Öklid uzaklığına göre elde edilen kümeler ile aynı olup, diskriminant analizi geçerlilik sonucu da yine % 92,6 ile aynıdır.

Çizelge 4.29. Tam bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler

Küme	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
1	Adana, Adıyaman, Ankara, Ardahan, Balıkesir, Gaziantep, İzmir, Kahramanmaraş, Malatya, Mersin, Muğla, Osmaniye, Siirt, Tunceli	Adana, Adıyaman, Ankara, Ardahan, Balıkesir, Gaziantep, Elazığ, Hatay, Iğdır, İzmir, Kahramanmaraş, Kilis, Malatya, Mersin, Muğla, Osmaniye, Siirt, Tunceli	Adana, Adıyaman, Ankara, Ardahan, Balıkesir, Gaziantep, Hatay, Iğdır, İzmir, Kahramanmaraş, Malatya, Mersin, Muğla, Osmaniye, Siirt, Tunceli	Adana, Adıyaman, Ankara, Ardahan, Balıkesir, Gaziantep, İzmir, Kahramanmaraş, Malatya, Mersin, Muğla, Osmaniye, Siirt, Tunceli
2	Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Bartın, Bingöl, Denizli, Elazığ, Erzurum, Eskişehir, Karaman, Kars, Kayseri, Kilis, Kırklareli, Konya, Rize, Sakarya, Sivas	Afyonkarahisar, Aksaray, Amasya, Antalya, Artvin, Aydın, Bingöl, Çorum, Denizli, Erzurum, Eskişehir, Karaman, Kars, Kayseri, Konya, Manisa, Rize, Samsun, Yozgat	Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Bartın, Bilecik, Bingöl, Bolu, Denizli, Düzce, Elazığ, Erzurum, Eskişehir, Karaman, Kars, Kayseri, Kırşehir, Kilis, Kırklareli, Konya, Nevşehir, Niğde, Rize, Sakarya, Sivas, Trabzon, Zonguldak	Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Bartın, Bingöl, Denizli, Elazığ, Erzurum, Eskişehir, Karaman, Kars, Kayseri, Kilis, Kırklareli, Konya, Rize, Sakarya, Sivas

Çizelge 4.29. (devam) Tam bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler

3	Ağrı, Batman, Bitlis, Diyarbakır, Hatay, Iğdır, Mardin, Muş, Şanlıurfa, Van	Ağrı, Batman, Bitlis, Diyarbakır, Hakkari, Mardin, Muş, Şanlıurfa, Şırnak, Van	Ağrı, Batman, Bitlis, Diyarbakır, Mardin, Muş, Şanlıurfa, Van	Ağrı, Batman, Bitlis, Diyarbakır, Hatay, Iğdır, Mardin, Muş, Şanlıurfa, Van
4	Antalya, Bursa, Çankırı, Çorum, Giresun, Gümüşhane, İstanbul, Kastamonu, Kocaeli, Manisa, Ordu, Samsun, Sinop, Yalova, Yozgat	Bartın, Bayburt, Bilecik, Bolu, Burdur, Çanakkale, Düzce, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kırklareli, Kırşehir, Kütahya, Nevşehir, Niğde, Sakarya, Sivas, Trabzon, Uşak, Zonguldak	Antalya, Bursa, Çankırı, Çorum, Giresun, Gümüşhane, İstanbul, Kastamonu, Kocaeli, Manisa, Ordu, Samsun, Sinop, Yalova, Yozgat	Antalya, Bursa, Çankırı, Çorum, Giresun, Gümüşhane, İstanbul, Kastamonu, Kocaeli, Manisa, Ordu, Samsun, Sinop, Yalova, Yozgat
5	Bayburt, Bilecik, Bolu, Burdur, Çanakkale, Düzce, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kırşehir, Kütahya, Nevşehir, Niğde, Trabzon, Uşak, Zonguldak	Bursa, İstanbul, Kocaeli, Ordu, Yalova	Bayburt, Burdur, Çanakkale, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kütahya, Uşak	Bayburt, Bilecik, Bolu, Burdur, Düzce, Çanakkale, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kırşehir, Kütahya, Nevşehir, Niğde, Trabzon, Uşak, Zonguldak
6	Hakkari, Şırnak	Çankırı, Giresun, Gümüşhane, Kastamonu, Sinop	Hakkari, Şırnak	Hakkari, Şırnak
7	Tekirdağ, Tokat	Tekirdağ, Tokat	Tekirdağ, Tokat	Tekirdağ, Tokat

Çizelge 4.30. Tam bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları

Tam Bağlantı Kümeleme Tekniği	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
Diskriminant Analizi Doğru Sınıflandırma Oranı(%)	92,6	97,5	96,3	92,6

Tam bağlantı kümeleme tekniğine uygulanan tüm uzaklıklara göre genel olarak bakıldığında oluşturulan kümelerin doğru sınıflandırma oranı diskriminant analizine göre % 97,5 ile en çok Pearson uzaklığına göre elde edilen kümelerde olmuştur. Daha sonra ise % 96,3 ile en yüksek ikinci sınıflandırma Manhattan uzaklığı ile elde edilmiştir. Karesi alınmış Öklid ve Minkowski uzaklıklarına göre edilen kümeler ise aynı olup % 92,6 ile tam bağlantı kümeleme tekniğinde diğer uzaklıklara göre daha düşük bir sınıflandırma oranı sağlamışlardır. Ayrıca, Tekirdağ ve Tokat il çifti tüm uzaklıklara göre elde edilen kümelerde ayrı bir küme oluşturmuşlardır.

4.2.3. Merkezi bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması

Çizelge 4.31. Merkezi bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler

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

Çizelge 4.32. Box-M test sonucu

Box's M	82,213
F	Yaklaşık 1,574
	sd1 40
	sd2 1305,420
	p-değeri ,013

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den küçük olarak bulunmuş ve yokluk hipotezi reddedilmiştir. Bu durumda, grupların varyans-kovaryans matrisleri eşit olmadığından karesel diskriminant analizi kullanılmış ve elde edilen kümelerin doğru sınıflandırma oranları bu şekilde belirlenmiştir.

Çizelge 4.33. Merkezi bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Merkezi Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	24	0	0	0	0	0	0	24
2	0	30	0	0	0	0	0	30
3	1	0	9	0	0	0	0	10
4	0	0	0	8	0	0	0	8
5	0	0	0	0	5	0	0	5
6	0	0	0	0	0	2	0	2
7	0	0	0	0	0	0	2	2
1	100,0	,0	,0	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	10,0	,0	90,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 98,8'i doğru sınıflandırılmıştır.

Merkezi bağlantı kümeleme tekniğine karesi alınmış Öklid uzaklığı uygulanarak elde edilen kümeler söz konusu kümeleme Tekniğindeki en yüksek sınıflandırma oranına % 98,8 ile sahip olduğu diskriminant analizi sonucu belirlenmiştir. Bu kapsamda, yalnızca 3. kümede yer alan Hatay ilinin 1. kümede yer alması tahmin edilmiştir.

Çizelge 4.34. Merkezi bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler

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

Çizelge 4.35. Box-M test sonucu

Box's M		160,666
F	Yaklaşık	2,182
	sd1	60
	sd2	4855,506
	p-değeri	,096

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.36. Merkezi bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Merkezi Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	6	2	0	0	0	0	0	8
2	0	16	0	0	0	0	0	16
3	0	0	10	0	2	0	0	12
4	1	0	0	7	0	0	0	8
5	0	1	0	0	15	0	0	16
6	0	1	1	0	0	11	0	13
7	0	0	0	0	0	0	8	8
1	75,0	25,0	,0	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	83,3	,0	16,7	,0	,0	100,0
4	12,5	,0	,0	87,5	,0	,0	,0	100,0
5	,0	6,3	,0	,0	93,8	,0	,0	100,0
6	,0	7,7	7,7	,0	,0	84,6	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 90,1'i doğru sınıflandırılmıştır.

Merkezi bağlantı kümeleme tekniğine uygulanan Pearson uzaklığına göre elde edilen kümeler, diskriminant analizi sonucu % 90,1 ile en düşük sınıflama oranına sahip kümeler olarak dikkat çekmektedir. Toplamda 8 il, farklı kümelere tahmin edilmiştir.

Çizelge 4.37. Merkezi bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler

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

Çizelge 4.38. Box-M test sonucu

Box's M	78,254
F	Yaklaşık 2,111
	sd1 30
	sd2 2337,336
	p-değeri ,056

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.39. Merkezi bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Merkezi Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	48	0	0	0	0	0	0	48
2	1	9	0	0	0	0	0	10
3	0	0	1	0	0	0	0	1
4	0	0	0	8	0	0	0	8
5	1	0	0	0	9	0	0	10
6	0	0	0	0	0	2	0	2
7	0	0	0	0	0	0	2	2
1	100,0	,0	,0	,0	,0	,0	,0	100,0
2	10,0	90,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	10,0	,0	,0	,0	90,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 97,5'i doğru sınıflandırılmıştır.

Merkezi kümeleme tekniğine uygulanan Manhattan uzaklığı ile elde edilen kümeler incelendiğinde önceki diğer uzaklıklara göre elde edilen kümelerden farklı olarak Bartın ili tek başına bir küme olmuştur. Hakkari, Şırnak ve Tekirdağ, Tokat il çiftleri ise diğer uzaklıklarda olduğu gibi yine birer küme meydana getirmiştir. Diskriminant analizi sonucu elde edilen kümelerin doğru sınıflandırma oranı ise % 97,5 ile oldukça yüksektir.

Çizelge 4.40. Merkezi bağlantı kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler

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

Çizelge 4.41. Box-M test sonucu

Box's M		64,066
F	Yaklaşık	1,510
	sd1	30
	sd2	869,151
	p-değeri	,040

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den küçük olarak bulunmuş ve yokluk hipotezi reddedilmiştir. Bu durumda, grupların varyans-kovaryans matrisleri eşit olmadığından karesel diskriminant analizi kullanılmış ve elde edilen kümelerin doğru sınıflandırma oranları bu şekilde belirlenmiştir.

Çizelge 4.42. Merkezi bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Merkezi Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	52	2	0	0	1	0	0	55
2	0	8	0	0	0	0	0	8
3	0	0	1	0	0	0	0	1
4	0	0	0	8	0	0	0	8
5	0	0	0	0	5	0	0	5
6	0	0	0	0	0	2	0	2
7	0	0	0	0	0	0	2	2
1	94,5	3,6	,0	,0	1,8	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 96,3'ü doğru sınıflandırılmıştır.

Merkezi kümeleme tekniğine Minkowski uzaklığı uygulanarak elde edilen kümeler Manhattan uzaklığına göre elde edilen kümeler ile benzer nitelikte olup, diskriminant

analizi ile yapılan geçerlilik sonucunda oluşan kümelerin % 96,3 oranında doğru sınıflandırıldığı hesaplanmıştır. Bu kapsamda 1.kümede yer alan Iğdır ve Siirt illerinin 2.kümede ve yine 1.kümede yer alan Ordu ilinin 5.kümede yer alması diskriminant analizi neticesinde tahmin edilmiştir.

Çizelge 4.43. Merkezi bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler

Küme	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
1	Adana, Adıyaman, Ankara, Antalya, Ardahan, Balıkesir, Bursa, Çorum, Gaziantep, İstanbul, İzmir, Kahramanmaraş, Kocaeli, Malatya, Manisa, Mersin, Muğla, Ordu, Osmaniye, Samsun, Siirt, Tunceli, Yalova, Yozgat	Adana, Ankara, Batman, Hatay, Iğdır, Kahramanmaraş, Mardin, Tunceli	Adana, Adıyaman, Afyonkarahisar, Aksaray, Amasya, Ankara, Antalya, Ardahan, Artvin, Aydın, Balıkesir, Bilecik, Bingöl, Bolu, Çanakkale, Çorum, Denizli, Düzce, Elazığ, Erzurum, Eskişehir, Gaziantep, İzmir, Kahramanmaraş, Karaman, Kars, Kayseri, Kilis, Kırklareli, Kırşehir, Konya, Kütahya, Malatya, Manisa, Mersin, Muğla, Nevşehir, Niğde, Osmaniye, Rize, Sakarya, Samsun, Siirt, Sivas, Trabzon, Tunceli, Yozgat, Zonguldak	Adana, Adıyaman, Afyonkarahisar, Aksaray, Amasya, Ankara, Antalya, Ardahan, Artvin, Aydın, Balıkesir, Bilecik, Bingöl, Bolu, Bursa, Çanakkale, Çorum, Denizli, Düzce, Elazığ, Erzurum, Eskişehir, Gaziantep, Hatay, Iğdır, İstanbul, İzmir, Kahramanmaraş, Karaman, Kars, Kayseri, Kilis, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mersin, Muğla, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Samsun, Siirt, Sivas, Trabzon, Tunceli, Yalova, Yozgat, Zonguldak
2	Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Bartın, Bilecik, Bingöl, Bolu, Çanakkale, Denizli, Düzce, Elazığ, Erzurum, Eskişehir, Karaman, Kars, Kayseri, Kırşehir, Kilis, Kırklareli, Konya, Kütahya, Nevşehir, Niğde, Rize, Sinop, Sakarya, Sivas, Trabzon, Zonguldak	Adıyaman, Antalya, Ardahan, Balıkesir, Elazığ, Gaziantep, İzmir, Kilis, Malatya, Manisa, Mersin, Muğla, Osmaniye, Samsun, Siirt, Yozgat	Ağrı, Batman, Bitlis, Diyarbakır, Hatay, Iğdır, Mardin, Muş, Şanlıurfa, Van	Ağrı, Batman, Bitlis, Diyarbakır, Mardin, Muş, Şanlıurfa, Van
3	Ağrı, Batman, Bitlis, Diyarbakır, Hatay, Iğdır, Mardin, Muş, Şanlıurfa, Van	Afyonkarahisar, Bilecik, Bolu, Çanakkale, Düzce, Eskişehir, Kırşehir, Kütahya, Nevşehir, Niğde, Trabzon, Zonguldak	Bartın	Bartın

Çizelge 4.43. (devam) Merkezi bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler

4	Bayburt, Burdur, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Uşak	Ağrı, Bitlis, Diyarbakır, Hakkari, Muş, Şanlıurfa, Şırnak, Van	Bayburt, Burdur, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Uşak	Bayburt, Burdur, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Uşak
5	Çankırı, Giresun, Gümüşhane, Kastamonu	Aksaray, Amasya, Artvin, Aydın, Bingöl, Çorum, Denizli, Erzurum, Karaman, Kars, Kayseri, Kırklareli, Konya, Rize, Sakarya, Sivas	Bursa, Çankırı, Giresun, Gümüşhane, İstanbul, Kastamonu, Kocaeli, Ordu, Sinop, Yalova	Çankırı, Giresun, Gümüşhane, Kastamonu, Sinop
6	Hakkari, Şırnak	Bartın, Bursa, Çankırı, Giresun, Gümüşhane, İstanbul, Kastamonu, Kocaeli, Ordu, Sinop, Tekirdağ, Tokat, Yalova	Hakkari, Şırnak	Hakkari, Şırnak
7	Tekirdağ, Tokat	Bayburt, Burdur, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Uşak	Tekirdağ, Tokat	Tekirdağ, Tokat

Çizelge 4.44. Merkezi bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları

Merkezi Bağlantı Kümeleme Tekniği	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
Diskriminant Analizi Doğru Sınıflandırma Oranı(%)	98,8	90,1	97,5	96,3

Merkezi bağlantı kümeleme tekniğine uygulanan tüm uzaklık ölçülerine genel olarak baktığımızda en yüksek doğru sınıflandırma oranının % 98,8 ile karesi alınmış Öklid uzaklığına göre elde edilen kümelerde olduğu görülmektedir. Daha sonra ise % 97,5 oranı ile Manhattan ve % 96,3 oranı ile Minkowski uzaklıklarına göre elde edilen kümelerin de oldukça yüksek oranda doğru sınıflandırıldığı belirlenmiştir. Söz konusu kümeleme tekniğinde Pearson uzaklığın ise % 90,1 oranında en düşük sınıflandırmaya sahip uzaklık ölçüsü olarak dikkat çekmektedir.

4.2.4. Medyan bağlantı kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması

Çizelge 4.45. Medyan bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler

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

Çizelge 4.46. Box-M test sonucu

Box's M	178,502
F	Yaklaşık 3,813
	sd1 40
	sd2 5524,454
	p-değeri ,122

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.47. Medyan bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Medyan Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	23	0	0	0	2	0	0	25
2	1	9	0	0	0	0	0	10
3	2	0	11	1	0	0	0	14
4	1	0	0	9	0	0	0	10
5	0	0	0	0	18	0	0	18
6	0	0	0	0	0	2	0	2
7	0	0	0	0	0	0	2	2
1	92,0	,0	,0	,0	8,0	,0	,0	100,0
2	10,0	90,0	,0	,0	,0	,0	,0	100,0
3	14,3	,0	78,6	7,1	,0	,0	,0	100,0
4	10,0	,0	,0	90,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 91,4'ü doğru sınıflandırılmıştır.

Medyan bağlantı kümeleme tekniğine uygulanan karesi alınmış Öklid uzaklığına göre elde edilen kümelerin, diskriminant analizi sonucunda % 91,4 oranında doğru sınıflandırıldığı tahmin edilmiştir. 3.kümede yer alan Aydın ve Denizli, 2.kümede yer alan Hatay ve 4.kümede yer alan Yozgat illerinin 1.kümede yer alması gerektiği tahmin edilmiştir.

Çizelge 4.48. Medyan bağlantı kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler

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

Çizelge 4.49. Box-M test sonucu

Box's M		106,782
F	Yaklaşık	1,905
	sd1	40
	sd2	900,025
	p-değeri	,101

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.50. Medyan bağlantı kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Medyan Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	3	0	1	0	0	0	0	4
2	0	33	0	0	0	0	0	33
3	0	0	10	0	0	0	0	10
4	0	0	0	22	0	0	0	22
5	0	0	0	0	5	0	0	5
6	0	0	0	0	0	5	0	5
7	0	0	0	0	0	0	2	2
1	75,0	,0	25,0	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 98,8'i doğru sınıflandırılmıştır.

Medyan bağlantı kümeleme tekniğine uygulanan Pearson uzaklığına göre elde edilen kümeler, diskriminant analizi sonuçlarına göre söz konusu kümeleme tekniğine uygulanan

uzaklıklar arasında en fazla oran olan % 98,8 oranında doğru sınıflandırılmıştır. Yalnızca Iğdır ilinin 1.küme yerine 3.kümede yer alması gerektiği tahmin edilmiştir.

Çizelge 4.51. Medyan bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler

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

Çizelge 4.52. Box-M test sonucu

Box's M	98,156
F	Yaklaşık 4,294
	sd1 20
	sd2 2801,149
	p-değeri ,069

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.53. Medyan bağlantı kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Medyan Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	23	3	0	0	0	0	0	26
2	0	35	0	0	0	1	1	37
3	0	0	3	0	0	0	0	3
4	0	1	0	9	0	0	0	10
5	0	0	0	0	2	0	0	2
6	1	0	0	0	0	0	0	1
7	0	0	0	0	0	0	2	2
1	88,5	11,5	,0	,0	,0	,0	,0	100,0
2	,0	94,6	,0	,0	,0	2,7	2,7	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	10,0	,0	90,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	100,0	,0	,0	,0	,0	,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 91,4'ü doğru sınıflandırılmıştır.

Medyan bağlantı kümeleme tekniğine Manhattan uzaklığı uygulanarak elde edilen kümeler incelendiğinde, önceki uzaklıklara göre elde edilen kümelerden farklı olarak Kars ilinin tek başına bir küme oluşturması dikkat çekmektedir. Ayrıca, aynı coğrafi bölgede yer alan Batman, Mardin ve Şanlıurfa üçlüsünün de birlikte bir küme oluşturması dikkat çeken bir diğer noktadır. Bununla birlikte, elde edilen kümelerin diskriminant analizi ile sorgulanması sonucu elde edilen kümelerin % 91,4 oranında doğru sınıflandırıldığı belirlenmiştir.

Çizelge 4.54. Medyan bağlantı kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler

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

Çizelge 4.55. Box-M test sonucu

Box's M		92,411
F	Yaklaşık	2,636
	sd1	30
	sd2	3552,046
	p-değeri	,064

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.56. Medyan bağlantı kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Medyan Bağlantı	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	23	0	0	0	0	0	1	24
2	1	31	0	0	0	0	0	32
3	1	0	9	0	0	0	0	10
4	0	2	0	8	0	0	0	10
5	0	0	0	0	2	0	0	2
6	0	1	0	0	0	0	0	1
7	0	0	0	0	0	0	2	2
1	95,8	,0	,0	,0	,0	,0	4,2	100,0
2	3,1	96,9	,0	,0	,0	,0	,0	100,0
3	10,0	,0	90,0	,0	,0	,0	,0	100,0
4	,0	20,0	,0	80,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	100,0	,0	,0	,0	,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 92,6'sı doğru sınıflandırılmıştır.

Medyan bağlantı kümeleme tekniğine Minkowski uzaklığı uygulanarak elde edilen kümeler incelendiğinde Kars ili, Manhattan uzaklığı ile benzer olarak ayrı bir küme oluşturmuştur. Ayrıca, 3.kümede yer alan Hatay ve 2.kümede yer alan Kilis illerinin

1.kümede, 4.kümede yer alan Çanakkale ve Kütahya illeri ile 6.kümede yer alan Kars ilinin 2.kümede yer alması gerektiği diskriminant analizi sonucu tahmin edilmiştir.

Çizelge 4.57. Medyan bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler

Küme	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
1	Adana, Adıyaman, Afyonkarahisar, Aksaray, Ankara, Ardahan, Artvin, Balıkesir, Bingöl, Elazığ, Erzurum, Eskişehir, Gaziantep, İzmir, Kahramanmaraş, Kars, Kayseri, Kilis, Konya, Malatya, Mersin, Muğla, Osmaniye, Siirt, Tunceli	Adana, Hatay, Iğdır, Kahramanmaraş	Adana, Adıyaman, Ağrı, Aksaray, Ankara, Ardahan, Balıkesir, Bingöl, Bitlis, Diyarbakır, Elazığ, Gaziantep, Hatay, Iğdır, İzmir, Kahramanmaraş, Kayseri, Kilis, Malatya, Mersin, Muğla, Muş, Osmaniye, Siirt, Tunceli, Van	Adana, Adıyaman, Ankara, Antalya, Ardahan, Balıkesir, Bursa, Çorum, Gaziantep, İstanbul, İzmir, Kahramanmaraş, Kocaeli, Malatya, Manisa, Mersin, Muğla, Ordu, Osmaniye, Samsun, Siirt, Tunceli, Yalova, Yozgat
2	Ağrı, Batman, Bitlis, Diyarbakır, Hatay, Iğdır, Mardin, Muş, Şanlıurfa, Van	Adıyaman, Afyonkarahisar, Aksaray, Amasya, Ankara, Antalya, Ardahan, Artvin, Aydın, Balıkesir, Bingöl, Çorum, Denizli, Elazığ, Erzurum, Eskişehir, Gaziantep, İzmir, Karaman, Kars, Kayseri, Kilis, Konya, Malatya, Manisa, Mersin, Muğla, Osmaniye, Rize, Samsun, Siirt, Tunceli, Yozgat	Afyonkarahisar, Amasya, Antalya, Artvin, Aydın, Bartın, Bilecik, Bolu, Bursa, Çankırı, Çorum, Denizli, Düzce, Erzurum, Eskişehir, Giresun, Gümüşhane, İstanbul, Karaman, Kastamonu, Kırklareli, Kırşehir, Kocaeli, Konya, Manisa, Nevşehir, Niğde, Ordu, Rize, Sakarya, Samsun, Sinop, Sivas, Trabzon, Yalova, Yozgat, Zonguldak	Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Bartın, Bilecik, Bingöl, Bolu, Çankırı, Denizli, Düzce, Elazığ, Erzurum, Eskişehir, Giresun, Gümüşhane, Karaman, Kastamonu, Kayseri, Kilis, Kırklareli, Kırşehir, Konya, Nevşehir, Niğde, Rize, Sakarya, Sinop, Sivas, Trabzon, Zonguldak
3	Amasya, Aydın, Bartın, Çankırı, Denizli, Giresun, Gümüşhane, Karaman, Kastamonu, Kırklareli, Rize, Sakarya, Sinop, Sivas	Ağrı, Batman, Bitlis, Diyarbakır, Hakkari, Mardin, Muş, Şanlıurfa, Şırnak, Van	Batman, Mardin, Şanlıurfa	Ağrı, Batman, Bitlis, Diyarbakır, Hatay, Iğdır, Mardin, Muş, Şanlıurfa, Van

Çizelge 4.57. (devam) Medyan bağlantı kümeleme tekniği - Tüm uzaklıklara göre kümeler

4	Antalya, Bursa, Çorum, İstanbul, Kocaeli, Manisa, Ordu, Samsun, Yalova, Yozgat	Bartın, Bayburt, Bilecik, Bolu, Burdur, Çanakkale, Düzce, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kırklareli, Kırşehir, Kütahya, Nevşehir, Niğde, Sakarya, Sivas, Trabzon, Uşak, Zonguldak	Bayburt, Burdur, Çanakkale, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kütahya, Uşak	Bayburt, Burdur, Çanakkale, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kütahya, Uşak
5	Bayburt, Bilecik, Bolu, Burdur, Çanakkale, Düzce, Edirne, Erzincan, Isparta, Karabük, Kırıkkale, Kırşehir, Kütahya, Nevşehir, Niğde, Trabzon, Uşak, Zonguldak	Bursa, İstanbul, Kocaeli, Ordu, Yalova	Hakkari, Şırnak	Hakkari, Şırnak
6	Hakkari, Şırnak	Çankırı, Giresun, Gümüşhane, Kastamonu, Sinop	Kars	Kars
7	Tekirdağ, Tokat	Tekirdağ, Tokat	Tekirdağ, Tokat	Tekirdağ, Tokat

Çizelge 4.58. Medyan bağlantı kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları

Medyan Bağlantı Kümeleme Tekniği	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
Diskriminant Analizi Doğru Sınıflandırma Oranı(%)	91,4	98,8	91,4	92,6

Medyan bağlantı kümeleme tekniğine uygulanan tüm uzaklıklara göre elde edilen kümeler incelendiğinde tüm küme oluşumlarında Tekirdağ ve Tokat il çiftinin birlikte ayrı bir küme oluşturduğu gözlemlenmiştir. Bununla birlikte, Hakkari ve Şırnak il çifti karesi alınmış Öklid, Manhattan ve Minkowski uzaklıklarında ayrı bir küme oluştururken, Kars ili Manhattan ve Minkowski uzaklıklarında tek başına bir küme meydana getirmiştir. Ayrıca, söz konusu kümeleme tekniğine Pearson uzaklığı uygulanarak meydana gelen kümelerin %

98,8 ile diğer uzaklıklardan farklı olarak en yüksek doğru sınıflandırma oranına sahip oldukları tespit edilmiştir.

4.2.5. Ward kümeleme tekniği ile uzaklık ölçülerinin karşılaştırılması

Çizelge 4.59. Ward kümeleme tekniği - karesi alınmış Öklid uzaklığına göre oluşan kümeler

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

Çizelge 4.60. Box-M test sonucu

Box's M	160,213
F	Yaklaşık 2,121
	sd1 60
	sd2 3321,928
	p-değeri ,113

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit

olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.61. Ward kümeleme tekniği - karesi alınmış Öklid uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Ward Metot	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	15	0	1	0	0	0	0	16
2	0	19	0	0	0	1	0	20
3	0	0	10	0	0	0	0	10
4	0	0	0	10	0	0	0	10
5	0	0	0	0	6	0	0	6
6	0	1	0	0	1	10	0	12
7	0	0	0	0	0	0	7	7
1	93,8	,0	6,3	,0	,0	,0	,0	100,0
2	,0	95,0	,0	,0	,0	5,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	8,3	,0	,0	8,3	83,3	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 95,1'i doğru sınıflandırılmıştır.

Ward kümeleme tekniğine karesi alınmış Öklid uzaklığı uygulanarak elde edilen kümelerin uygulanan diskriminant analizi sonucunda % 95,1 oranında doğru sınıflandırıldığı belirlenmiştir.

Çizelge 4.62. Ward kümeleme tekniği - Pearson uzaklığına göre oluşan kümeler

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

Çizelge 4.63. Box-M test sonucu

Box's M		142,297
F	Yaklaşık	2,255
	sd1	50
	sd2	2056,053
	p-değeri	,098

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.64. Ward kümeleme tekniği - Pearson uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Ward Metot	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	14	1	1	0	0	0	0	16
2	0	15	0	1	0	0	0	16
3	0	0	10	0	0	0	0	10
4	0	0	0	10	0	0	0	10
5	0	0	0	0	22	0	0	22
6	0	0	0	0	0	5	0	5
7	0	0	0	0	0	0	2	2
1	87,5	6,3	6,3	,0	,0	,0	,0	100,0
2	,0	93,8	,0	6,3	,0	,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	,0	,0	100,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 96,3'ü doğru sınıflandırılmıştır.

Ward kümeleme tekniğine uygulanan Pearson uzaklığına göre elde edilen kümeler incelendiğinde, karesi alınmış Öklid uzaklığına göre birlikte bir kümeyi oluşturan 6. ve 7.kümeler bu uzaklık sonucunda elde edilen kümelerde bölünmüş ve ayrı birer küme oluşturmuşlardır. Bununla birlikte, 1.kümede yer alan Balıkesir ilinin 2.kümede, yine

1.kümede yer alan Iğdır ilinin 3.kümede ve 2.kümede yer alan Rize ilinin 4.kümede yer alması gerektiği diskriminant analizi sonucu tahmin edilmiştir.

Çizelge 4.65. Ward kümeleme tekniği - City-Block(Manhattan) uzaklığına göre oluşan kümeler

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

Çizelge 4.66. Box-M test sonucu

Box's M	153,697
F	Yaklaşık 2,107
	sd1 60
	sd2 5427,713
	p-değeri ,102

$$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6 = \Sigma_7$$

H_s : En az iki Σ_i ($i=1, \dots, 7$) eşit değildir.

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.67. Ward kümeleme tekniği - City-Block(Manhattan) uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Ward Metot	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	14	0	0	0	0	0	0	14
2	0	14	0	0	0	0	0	14
3	1	0	11	0	0	0	0	12
4	0	0	0	13	1	0	0	14
5	0	0	0	0	10	0	0	10
6	0	0	0	1	0	9	0	10
7	0	0	0	0	0	0	7	7
1	100,0	,0	,0	,0	,0	,0	,0	100,0
2	,0	100,0	,0	,0	,0	,0	,0	100,0
3	8,3	,0	91,7	,0	,0	,0	,0	100,0
4	,0	,0	,0	92,9	7,1	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	,0	,0	10,0	,0	90,0	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 96,3'ü doğru sınıflandırılmıştır.

Ward kümeleme tekniğine Manhattan uzaklığı uygulanarak elde edilen kümelerin, diskriminant analizi sonucu % 96,3 oranında doğru sınıflandırıldığı tahmin edilmiştir. Bununla birlikte, 6.kümede yer alan Çanakkale ilinin 4.kümede, 3.kümede yer alan Hatay ilinin 1.kümede ve 4.kümede yer alan Rize ilinin 5.kümede yer alması gerektiği diskriminant analizi sonucunda tespit edilmiştir.

Çizelge 4.68. Ward kümeleme tekniği - Minkowski uzaklığına göre oluşan kümeler

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

Çizelge 4.69. Box-M test sonucu

Box's M		154,342
F	Yaklaşık	2,066
	sd1	60
	sd2	3984,415
	p-değeri	,105

Box-M testinin sonucuna göre p değeri $\alpha=0,05$ 'den büyük olarak bulunmuş ve yokluk hipotezi kabul edilmiştir. Bu kapsamda, grupların varyans-kovaryans matrisleri eşit olduğundan doğrusal diskriminant analizi kullanılmış ve elde edilen kümelerin geçerliliği bu şekilde belirlenmiştir.

Çizelge 4.70. Ward kümeleme tekniği - Minkowski uzaklığı kümelerinin diskriminant analizi sınıflandırma sonuçları

Ward Metot	Tahmini Grup Üyelikleri							Toplam
	1	2	3	4	5	6	7	
1	15	0	1	0	0	0	0	16
2	0	19	0	0	0	1	0	20
3	0	0	10	0	0	0	0	10
4	0	0	0	10	0	0	0	10
5	0	0	0	0	7	0	0	7
6	0	1	0	0	1	9	0	11
7	0	0	0	0	0	0	7	7
1	93,8	,0	6,3	,0	,0	,0	,0	100,0
2	,0	95,0	,0	,0	,0	5,0	,0	100,0
3	,0	,0	100,0	,0	,0	,0	,0	100,0
4	,0	,0	,0	100,0	,0	,0	,0	100,0
5	,0	,0	,0	,0	100,0	,0	,0	100,0
6	,0	9,1	,0	,0	9,1	81,8	,0	100,0
7	,0	,0	,0	,0	,0	,0	100,0	100,0

Orijinal grupların % 95,1'i doğru sınıflandırılmıştır.

Ward kümeleme tekniğine uygulanan Minkowski uzaklığı sonucunda elde edilen kümeler karesi alınmış Öklid uzaklığı sonucu elde edilen kümeler ile benzer nitelikte olup sadece

Erzincan ili yer değiştirerek karesi alınmış Öklid uzaklığında 6.kümede yer alırken, Minkowski uzaklığında 5.kümede yer almıştır. Ayrıca elde edilen kümeler, karesi alınmış Öklid uzaklığında olduğu gibi % 95,1 oranında doğru sınıflandırılmıştır.

Çizelge 4.71. Ward kümeleme tekniği - Tüm uzaklıklara göre kümeler

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

Çizelge 4.72. Ward kümeleme tekniği - Diskriminant analizi sınıflandırma sonuçları

Ward Kümeleme Tekniği	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
Diskriminant Analizi Doğru Sınıflandırma Oranı(%)	95,1	96,3	96,3	95,1

Ward kümeleme tekniğine uygulanan uzaklıklara göre elde edilen kümeler incelendiğinde, “Çankırı, Giresun, Gümüşhane, Kastamonu, Sinop, Tekirdağ ve Tokat” il grubunun karesi alınmış Öklid, Manhattan ve Minkowski uzaklıklarında ayrı bir küme oluşturduğu, Pearson uzaklığında ise söz konusu il grubunun iki kümeye bölünerek Tekirdağ ve Tokat il çiftinin ayrı bir küme oluşturduğu gözlemlenmiştir. Ayrıca, coğrafi olarak benzer olan “Ağrı, Batman, Bitlis, Diyarbakır, Hakkari, Mardin, Muş, Şanlıurfa, Şırnak, Van” il grubunun da karesi alınmış Öklid, Pearson ve Minkowski uzaklıklarında aynı şekilde bir küme oluşturduğu dikkat çekmektedir. Bununla birlikte, karesi alınmış Öklid ve Minkowski uzaklıklarına göre elde edilen kümeler % 95,1 oranında, Pearson ve Manhattan uzaklıklarında elde edilen kümelerin ise % 96,3 oranında doğru sınıflandırıldığı doğrusal diskriminant analizi sonucu tespit edilmiştir.

Çizelge 4.73. Tüm kümeleme teknikleri - Diskriminant analizi sınıflandırma sonuçları

Diskriminant Analizi Doğru Sınıflandırma Oranları(%)	Karesi Alınmış Öklid	Pearson	City-Block (Manhattan)	Minkowski
Tek Bağlantı Kümeleme Tekniği	98,8	98,8	98,8	98,8
Tam Bağlantı Kümeleme Tekniği	92,6	97,5	96,3	92,6
Merkezi Bağlantı Kümeleme Tekniği	98,8	90,1	97,5	96,3
Medyan Bağlantı Kümeleme Tekniği	91,4	98,8	91,4	92,6
Ward Kümeleme Tekniği	95,1	96,3	96,3	95,1

5. SONUÇ VE ÖNERİLER

Günümüz dünyasında artık hemen her şeyin verilerden ibaret olduğu göz önüne alındığında, verilerin analizi kadar bu analizleri gerçekleştirirken doğru tekniklerin kullanılması da bir o kadar önemlidir. Bu analizler arasında yaygın ve geçerli bir şekilde kullanılan kümeleme analizi, bizlere ilgilenilen veri setine ait birimlere yönelik çıkarsama ve yorum yapmayı kolaylaştırmaktadır. Bu sayede, araştırılan konuya ilişkin bir ön bilgi sahibi olunabilmekle birlikte söz konusu veri setine yönelik bir model geliştirme imkanı da sunmaktadır. Bununla birlikte, analiz sonucunda elde edilen gruplara ait birimlerin, doğru sınıflanıp sınıflanmadığı veya hangi oranda doğru sınıflandırıldığı da son derece önemli olup, araştırmacı tarafından sorgulanmalıdır.

Kümeleme analizinde, gerçekleştirilen analizler neticesinde anlamlı ve geçerli sonuçlara ulaşabilmek birçok araştırmacının hem hedefi hem de karşılaştığı en önemli sorunlardan biridir. Özellikle, analizin başında küme sayısının belirlenmesi gerekmektedir. Uygun küme sayısının belirlenmesinden sonra hangi kümeleme analiz tekniğinin kullanılacağına karar vermekte bir o kadar önem arz etmektedir. Verilerin analizinde mümkün olan tüm hiyerarşik kümeleme tekniklerine önemli uzaklık ölçülerinin uygulanarak denenmesi, araştırmacının verilerini daha doğru bir şekilde sınıflandırmasına ve meydana gelen kümeleri karşılaştırarak incelemesine imkan sağlayacaktır.

Verilere uygulanan kümeleme analizi sonucunda oluşturulan kümelerin geçerliliği ve birimlerin doğru kümede sınıflandırılması sorunu, kümeleme analizinin en zor kısımlarından biridir. Bu bakımdan elde edilen kümeleme çözümlerinin kalitesini test etmek amacıyla geliştirilen bazı yöntemler bulunmaktadır. Bu yöntemlerden biri de, çalışmada incelenen çok değişkenli bir istatistik analiz tekniği olan diskriminant analizinin kümeleme sonuçlarına uygulanmasıdır. Söz konusu diskriminant analizinde, elde edilen grupların varyans-kovaryans matrislerinin eşitliği durumunda doğrusal diskriminant analizi, farklılığı durumunda ise karesel diskriminant analizi kullanılmaktadır. Bununla birlikte, kovaryans matrisleri arasındaki fark çok fazla olmadığında, buna ek olarak veri setinin sahip olduğu birim sayısı orta ya da az düzeylerde olduğunda doğrusal diskriminant analizinin daha gerçeğe yakın sonuçlar verdiğini söylemek mümkündür (Gilbert, 1969).

Bu çalışmada, TÜİK 2017 yılı 15 yaş ve üzeri iç göç verileri kullanılarak kümeleme analizinde kullanılan bazı önemli uzaklık ölçüleri farklı hiyerarşik kümeleme teknikleri üzerinde uygulanarak elde edilen kümeler karşılaştırılmış ve ortaya çıkan bu kümelerin her biri için diskriminat analizi uygulanarak söz konusu meydana gelen kümelerin geçerliliği sorgulanmaya çalışılmıştır. Dolayısıyla söz konusu uygulama, hangi uzaklık ölçüsünün hangi kümeleme tekniğinde nasıl kümeler oluşturduğu incelenerek diskriminant analizi neticesinde hangi oranda doğru sınıflandırma ortaya koyduğu sonucu ile birlikte, Türkiye’de 81 il bazında illere göç edenlerin eğitim durumu konusunda durum tespiti yapılmış olup elde edilen bu sınıflandırmaların ilgili politika, uygulama ve diğer akademik çalışmalara yol gösterici nitelikte olacağı düşünülmektedir.

Araştırmada, hiyerarşik kümeleme analizi tekniklerinden tek bağlantı kümeleme tekniği, tam bağlantı kümeleme tekniği, merkezi bağlantı kümeleme tekniği, medyan bağlantı kümeleme tekniği ve Ward tekniklerinden her birine karesi alınmış Öklid, Pearson, City-Block(Manhattan) ve Minkowski uzaklıkları uygulanarak kümeleme analizi gerçekleştirilmiştir.

Tek bağlantı kümeleme tekniğine uygulanan uzaklık ölçülerine bakıldığında karesi alınmış Öklid, Manhattan ve Minkowski uzaklıklarının Pearson uzaklığından farklı olarak aynı kümeleri oluşturduğu görülmüştür. Tüm uzaklıklara göre elde edilen kümeler varyans-kovaryans matrisleri eşit olmadığından karesel diskriminant analizi ile geçerliliği sorgulanmış ve hepsinde % 98,8 ile aynı oranda doğru sınıflandırma yapıldığı belirlenmiştir. Bununla birlikte, Gümüşhane ili tüm uzaklık kümelerinde ayrı bir küme olarak yer alırken Çankırı, Giresun, Kastamonu ve Sinop illeri yine tüm uzaklıklarda beraber olarak bir küme oluşturmuşlardır.

Tam bağlantı kümeleme tekniğine uygulanan uzaklık ölçülerine göre oluşturulan kümelerin doğru sınıflandırma oranı varyans-kovaryans matrisleri eşit olduğu için doğrusal diskriminant analizine göre % 97,5 ile en çok Pearson uzaklığına göre elde edilen kümelerde olmuştur. Daha sonra ise % 96,3 ile en yüksek ikinci sınıflandırma Manhattan uzaklığı ile elde edilmiştir. Karesi alınmış Öklid ve Minkowski uzaklıklarına göre edilen kümeler ise aynı olup % 92,6 ile tam bağlantı kümeleme tekniğinde diğer uzaklıklara göre daha düşük bir sınıflandırma oranı sağlamışlardır. Ayrıca, tüm uzaklıklara göre elde edilen kümelerin hepsinde Tekirdağ ve Tokat il çifti ayrı bir küme oluşturmuşlardır.

Merkezi bağlantı kümeleme tekniğine uygulanan uzaklık ölçülerine göre elde edilen kümeler incelendiğinde, gerçekleştirilen karesel diskriminant analizi sonucu en yüksek doğru sınıflandırma oranının % 98,8 ile karesi alınmış Öklid uzaklığına göre elde edilen kümelerde olduğu görülmüştür. Daha sonra ise % 97,5 oranı ile Manhattan ve % 96,3 oranı ile Minkowski uzaklıklarına göre elde edilen kümelerin de oldukça yüksek oranda doğru sınıflandırıldığı belirlenmiştir. Söz konusu kümeleme tekniğinde Pearson uzaklığı ise % 90,1 oranında en düşük sınıflandırmaya sahip uzaklık ölçüsü olarak dikkat çekmiştir. Bununla birlikte, Hakkari ve Şırnak il çifti ile Tekirdağ ve Tokat il çiftleri karesi alınmış Öklid, Manhattan ve Minkowski uzaklıklarında ayrı birer küme oluştururken; Bartın ili Manhattan ve Minkowski uzaklıklarında tek başına bir kümeyi oluşturmuştur.

Medyan bağlantı kümeleme tekniğine uygulanan uzaklıklara göre elde edilen kümeler incelendiğinde tüm küme oluşumlarında Tekirdağ ve Tokat il çiftinin birlikte ayrı bir küme oluşturduğu gözlemlenmiştir. Bununla birlikte, Hakkari ve Şırnak il çifti karesi alınmış Öklid, Manhattan ve Minkowski uzaklıklarında ayrı bir küme oluştururken, Kars ili Manhattan ve Minkowski uzaklıklarında tek başına bir küme meydana getirmiştir. Ayrıca, söz konusu kümeleme tekniğine Pearson uzaklığı uygulanarak meydana gelen kümelerin doğrusal diskriminant analizi sonucu % 98,8 ile diğer uzaklıklardan farklı olarak en yüksek doğru sınıflandırma oranına sahip oldukları tespit edilmiştir.

Ward kümeleme tekniğine uygulanan uzaklıklara göre elde edilen kümeler incelendiğinde, “Çankırı, Giresun, Gümüşhane, Kastamonu, Sinop, Tekirdağ ve Tokat” il grubunun karesi alınmış Öklid, Manhattan ve Minkowski uzaklıklarında ayrı bir küme oluşturduğu, Pearson uzaklığında ise söz konusu il grubunun iki kümeye bölünerek Tekirdağ ve Tokat il çiftinin ayrı bir küme oluşturduğu gözlemlenmiştir. Ayrıca, coğrafi olarak benzer olan “Ağrı, Batman, Bitlis, Diyarbakır, Hakkari, Mardin, Muş, Şanlıurfa, Şırnak, Van” il grubunun da karesi alınmış Öklid, Pearson ve Minkowski uzaklıklarında aynı şekilde bir küme oluşturduğu dikkat çekmiştir. Bununla birlikte, karesi alınmış Öklid ve Minkowski uzaklıklarına göre elde edilen kümeler % 95,1 oranında, Pearson ve Manhattan uzaklıklarında elde edilen kümelerin ise % 96,3 oranında doğru sınıflandırıldığı doğrusal diskriminant analizi sonucunda tespit edilmiştir.

Genel olarak bakıldığında, hiyerarşik olmayan kümeleme tekniklerine uygulanan uzaklık ölçülerine göre elde edilen kümeler ve uygulanan diskriminant analizleri sonucunda tek

bağlantı kümeleme tekniğinde tüm uzaklıkların aynı oranda doğru sınıflandırma yaptığı ancak elde edilen kümelerin diğer oluşturulan kümelere göre homojen bir dağılım göstermediği görülmüştür. Tam bağlantı kümeleme tekniği ve medyan bağlantı kümeleme tekniğinde Pearson uzaklığına göre elde edilen kümelerin sırasıyla % 97,5 ve % 98,8 oranında en fazla doğru sınıflandırma oranına sahip olduğu, merkezi bağlantı kümeleme tekniğinde karesi alınmış Öklid uzaklığının % 98,8 ile en fazla oranda doğru sınıflandırma gösterdiği ve Ward tekniğinde ise Pearson ve Manhattan uzaklıklarının % 96,3 ile en fazla doğru sınıflandırma oranına sahip oldukları belirlenmiştir.

Benzerlik katsayılarının performansı üzerine gerçekleştirilen birçok çalışmada elde edilen ve ortaya konan farklı tespitler bulunmakla birlikte; bu çalışmada da, Ward tekniğinin tüm uzaklık katsayılarına göre oluşturulan kümelerinde diğer kümeleme tekniklerine nazaran daha homojen kümeler oluşturduğu ve her bir uzaklık katsayısının % 95'in üzerinde ve birbirine daha yakın performans gösterdiği ortaya konulmuştur.

Çalışmada ele alınan iç göç hareketi içindeki 15 yaş ve üzeri nüfus, 2017 yılı TÜİK verilerine göre 2.231.299 olmakla birlikte toplam nüfusun yaklaşık olarak % 3'ünü oluşturmaktadır. Ayrıca, söz konusu nüfusun % 51,5'inin kadınlar ve % 25,6 ile en fazla oranda 20-24 yaş arası aktif ve genç nüfustan meydana geldiği görülmektedir. Gerçekleştirilen analiz sonuçlarına göre, iç göç hareketi içinde yer alan nüfusun % 1,7'sinin okuma-yazma bilmeyenlerden oluştuğu ortaya çıkmış olup bu oran 15 yaş ve üzeri nüfus için Türkiye ortalamasının (% 3,8) altındadır. Ayrıca, iç göç hareketi içindeki nüfusun yaklaşık % 10'unun ilkokul mezunu, % 16,5'inin ortaokul mezunu olmakla birlikte ortalama % 40,6'sının lise ve dengi meslek okulu mezunu ve % 31'inin lisans ve üstü mezunu olduğu ortaya çıkmıştır. Söz konusu oranlara bakıldığında da, lise ve dengi meslek okul ile lisans ve üstü mezun oranlarının 15 yaş ve üzeri nüfus için Türkiye ortalamasından daha yüksek olması (sırasıyla % 23,1 ve % 17,1) göze çarpan bir diğer sonuç olarak dikkat çekmektedir. Bu durum, iç göç hareketi içinde yer alan bireylerin Türkiye ortalamasına göre daha eğitilmiş bireylerden oluştuğu şeklinde yorumlanabilir. Ayrıca, bahse konu 20-24 yaş arası nüfus oranının fazlalığı ve lise ve üstü eğitim seviyesine sahip bireylerin oranlarının çokluğu ise, iç göç hareketinin belli oranda eğitim amaçlı olarak gerçekleştiğini kanıtlar nitelikte olduğu söylenebilir. Bununla birlikte; söz konusu yaş grubunun genç olarak nitelendirilen 15-24 yaş grubu arasında yer alması, iç göç hareketinin yaklaşık % 72'sinin lise ve üstü eğitim seviyesinde olduğu ve günümüz

üniversite mezunu/genç işsiz rakamlarını göz önüne aldığımızda iç göç hareketi içinde yer alanların istihdam yani iş bulma odaklı olarak da hareket ettiği değerlendirilebilir.

İlleri, göç edenlerin eğitim durumlarına göre değerlendirdiğimizde ise elde edilen hemen hemen tüm kümelerde coğrafi olarak benzer nitelikte olan “Giresun, Gümüşhane, Kastamonu, Sinop” il grubunun, yine coğrafi olarak aynı bölgede olan “Hakkari, Şırnak” il çiftinin ve diğer Güneydoğu illerinin ayrı birer küme oluşturduğu görülmüştür. “Tekirdağ, Tokat” il çifti ise yine çoğu kümede ayrı bir küme meydana getirerek dikkat çekmiştir.

Elde edilen hemen hemen tüm kümelerde ayrı bir küme oluşturan “Hakkari, Şırnak” il çiftinin ortalama olarak aldığı göçlerde ağırlığın yaklaşık yarısı yani % 49,75 oranında lisans ve üstü mezunlarının oluşturması da Türkiye geneli alınan göçlerin ortalamasını (yaklaşık % 31) düşündüğümüzde dikkat çekmiştir. Bu durumun sebeplerinin neler olabileceğinin daha detaylı olarak araştırılmasının gerektiği düşünülmele birlikte yerel bazda göç hareketleri alanında inceleme yapmayı düşünen araştırmacılara söz konusu iller özelinde bir ön bilgi sağlayıp yol göstermektedir. Bununla birlikte, bahse konu illerin herhangi bir eğitim, tarım ve sanayi merkezi olmadığı göz önüne alındığında genel olarak üniversite mezunu olan kamu ve asker meslek grubunda kişilerin zorunlu hizmet veya tayin durumları nedeniyle belli bir zaman dilimi için göç hareketi içinde bulunarak söz konusu illerin aldığı göçün ağırlığının bu kişilerden oluşmasına neden olduğu şeklinde açıklanabilir.

Ayrıca, Karabük ili aldığı göçlerin % 60,5’ini lise ve dengi meslek okulu mezunu nüfustan alarak dikkat çekmektedir. İllere göç edenlerin ortalama olarak % 40,6’sının lise ve dengi meslek okulu mezunu olduğu göz önüne alındığında bu oranın Karabük özelinde oldukça yüksek olduğu görülmektedir. Bu durumun ortaya çıkmasında, Karabük ilinde faaliyet gösteren demir-çelik fabrikalarında “ara eleman” olarak istihdam edilmek üzere ihtiyaç duyulan meslek lisesi ve yüksekokul mezunlarının, diğer illere kıyasla göç edenler arasında daha ağırlıklı olarak bu şehre göç etmesinin etkili olduğu düşünülmektedir.

Sonuç olarak bu çalışma, hiyerarşik kümeleme teknikleri üzerinde önemli uzaklık ölçülerini uygulayarak meydana getirdikleri kümelerin diskriminant analizi ile geçerliliğini sınamakta ve benzerlik katsayılarının gösterdikleri performansı ortaya koymaktadır. Bununla birlikte, göç hareketi içinde yer alanların eğitim durumunun belirlenmesi ve illerin

bu durum bakımından benzerlik ya da uzaklıklarına göre kümelenmesi amacıyla bir durum tespiti niteliğinde gerçekleştirilmiştir. Ayrıca, gerçekleştirilen analiz ve ortaya çıkan sonuçlar çerçevesinde açıklayıcı ve tutarlı bilgiler içeren çalışmanın tespit edilenler ışığında, bu durumun daha derinlemesine ve nedenlerinin neler olabileceği şeklinde araştırmacılar tarafından detaylı ve yerel düzeyde incelenmesinin faydalı olabileceği sonucu ortaya konulmuştur.



KAYNAKLAR

- Anderberg, M.R. (1973). *Cluster analysis for applications*. New York: Academic Press, 553-555.
- Bostan, H. (2017). Türkiye’de iç göçlerin toplumsal yapıda neden olduğu değişimler, meydana getirdiği sorunlar ve çözüm önerileri. *Coğrafya Dergisi*, 35, 1-16.
- Bülbül, S., Köse, A. (2010). Türkiye’de bölgelerarası iç göç hareketlerinin çok boyutlu ölçekleme tekniği ile incelenmesi. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 39(1), 75-94.
- Cangül, O. (2016). *Diskriminant analizi ve bir uygulama denemesi*. Yüksek Lisans Tezi, Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Bursa, 26-74.
- Carvalho N. B. (2015). A discriminant function for validation of the cluster analysis and behavioral prediction of the coffee market. *Food Research International*, 77, 400-407.
- Çakmak, D. Z. (1999). Kümeleme analizinde geçerlilik problemi ve kümeleme sonuçlarının değerlendirilmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 9, 191-192.
- Çamdeviren, H. (2000). *Lojistik regresyon ve diskriminant analizi*. Doktora Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 89-91.
- Çelik, F. (2002). İç göçlerin seçkinlik yaklaşımı ile analizi. *Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 13, 276-296.
- Demirhan, N. (1997). *Kümeleme analizi ile konfeksiyon üretiminde Beden-Drop ölçülerinin belirlenmesi ve bir uygulama*. Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 18-19.
- Dinler, M. (2014). *Kümeleme analizi tekniklerinin hayvancılık verilerinde karşılaştırılmalı olarak incelenmesi*. Yüksek Lisans Tezi, Bingöl Üniversitesi Fen Bilimleri Enstitüsü, Bingöl, 14-26.
- Doğan, B. (2008). *Bankaların gözetiminde bir araç olarak kümeleme analizi: türk bankacılık sektörü için bir uygulama*. Doktora Tezi, Kadir Has Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 76-94.
- Duran, B., Odell, P. (1974). *Cluster analysis; A survey*. New York: Springer Verlag, 22.
- Erçetin, Y. (1993). *Diskriminant analizi ve bankalar üzerine bir uygulama*. Ankara: Türkiye Kalkınma Bankası A.Ş., 2-10.
- Erder, S. (1986). *Refah toplumunda “Getto” ve Türkler*. İstanbul: Teknografik Matbaacılık A.Ş., 12-15.
- Everitt, S.B. (1974). *Cluster analysis*. London: Heinmann Educational Books, 115.

- Finch H. (2015). Comparison of distance measures in cluster analysis with dichotomous data. *Journal of Data Science*, 3(2005), 85-100.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eungenicis*, 7, 179-188.
- Garipağaoğlu, N. (2016). Türkiye’de göç alan illere yönelen nüfusun eğitim durumu. *Marmara Üniversitesi Marmara Coğrafya Dergisi*, 3(2), 71-86.
- Gilbert E.S. (1969). The effect of unequal variance-covariance matrices on fisher’s linear discriminant function. *Biometrics*, 25(3), 505-515.
- Girginer, N., Kamışlı, M. (2010). İşlem bazlı manipülasyonun istatistiksel sınıflandırma analizleriyle belirlenmesi. *İstanbul Üniversitesi Ekonometri ve İstatistik Dergisi*, 11, 12-13.
- Green, E.P. (1989). *Analysing multivariate data*. Philadelphia: Duhok, 427.
- Han, J., Kamber, M. (2001). *Data mining concepts and techniques*. Massachusetts: Morgan Kaufmann Publishers, 34.
- Hofman, I., Jarvis, R. (1998). *Robust and efficient cluster analysis using a shared near neighbours approach*. 14. Pattern Recognition Conference-Volume 1, Brisbane, 243.
- İnternet: Türk Dil Kurumu. (1932). *Güncel Türkçe sözlük*. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.tdk.gov.tr%2F&date=2019-05-20>, Son Erişim Tarihi: 12.03.2019.
- İnternet: Türkiye İstatistik Kurumu. (2018). *Göç istatistikleri*. URL: <http://www.webcitation.org/query?url=https%3A%2F%2Fbiruni.tuik.gov.tr%2Fmedas%2F%3Fkn%3D95%26locale%3Dtr&date=2019-05-20>, Son Erişim Tarihi: 15.03.2019.
- Johnson, R. (2002). *Applied multivariate statistical analysis*. New Jersey: Prentice Hall Press, 12-22.
- Johnson, R., Wichern, D. (1988). *Applied multivariate statistical analysis*. New Jersey: Prentice- Hall International Editions, 554.
- Kalaycı, Ş. (2008). *SPSS uygulamalı çok değişkenli istatistik teknikleri* (3. Baskı). Ankara: Asil Yayın Dağıtım Ltd.Şti, 350–369.
- Kantardzic, M. (2003). *Data mining: Concepts, models and algorithms*. New York: IEEE Press and John Wiley, 18.
- Kızılcılık, S., Erjem Y. (1992). *Açıklamalı sosyoloji terimler sözlüğü*. Konya: Emre Yayıncılık, 185.
- Klecka, W. R. (1980). *Discriminant analysis*. Thousand Oaks: Sage Publishing and Beverly Hills, 14-20.
- Marriott, F.H.C. (1971). Practical problems in a method of cluster analysis. *Biometrics*, 27, 501-514.

- Munro, J. (1974). Migration in Turkey. *Economic Development and Cultural Change*, 22(4), 634–653.
- Özdamar, K. (2004). *Paket programlar ile istatistiksel veri analizi (çok değişkenli analizler)*. Eskişehir: Kaan Kitabevi, 502.
- Öztürk, F. (2012). *Kümeleme analizi ve uygulaması*. Yüksek Lisans Tezi, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 11-19.
- Perruchoud, R. (Ed.). (2009). *Göç terimleri sözlüğü* (Çev. B. Çiçekli). Cenevre: Uluslararası Göç Örgütü, 19.
- Rencher, A. L. (2002). *Methods of multivariate analysis*. USA: John Wiley and Sons, Inc., 708.
- Romesburg, H. (1984). *Cluster analysis for researchers*. Malabar: Krieger Publishing, 40.
- Sangün, L (2007). *Temel bileşenler analizi, ayırma analizi, kümeleme analizleri ve ekolojik verilere uygulanması üzerine bir araştırma*. Doktora Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana, 24-30.
- Servi, T. (2009). *Çok değişkenli karma dağılım modeline dayalı kümeleme analizi*. Doktora Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana, 30.
- Sharma, S. (1996). *Applied multivariate techniques*. New York: John Wiley and Sons Inc, 27.
- Shirkhorshidi, A.S. (2015). A comparison study on similarity and dissimilarity measures in clustering continuous data. *Article in Plos One*, 1-21.
- Sönmez, H., Er, F. (2007). Türkiye’de illere göre iç göç hareketlerinin modern kümeleme teknikleri ile incelenmesi. *Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 20(1), 17-32.
- Şahin, D. (2017). Kümeleme analizi ile doğu Avrupa ülkelerinin ekonomik özgürlükler açısından değerlendirilmesi. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 9.
- Taşatan, A. (2018). *Hiyerarşik kümeleme tekniklerinde küme eleman sayısının eşitlenmesine yönelik bir yaklaşım önerisi ve gerçek karayolu uzaklık verilerine dayalı kümeleme analizi*. Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli, 16-42.
- Tatlıdil, H. (2008). *Uygulamalı çok değişkenli analiz*. Ankara: Cem Web Ofset Ltd. Şti, 335-369.
- Tekeli, İ. (1998). Türkiye’de içgöç sorunsalı yeniden tanımlanma aşamasına geldi. A. İçduygu (Ed.), *Türkiye’de içgöç konferans Bolu-Gerede, 6-8 Haziran 1997 içinde* (s.7-21). İstanbul: Tarih Vakfı Yurt Yayınları.

- Tümer, M. (2001). *Kuzey Kıbrıs Türk Cumhuriyeti imalat sanayinde faaliyet gösteren kobileri ayırtıran faktörlerin tespiti*. Gazimağusa: Doğu Akdeniz Üniversitesi, 296-303.
- Üstünel, M. (2018). *k-ortalamlar algoritmasına dayalı kümeleme analizi sistemi ve perakendecilik sektöründe uygulaması*. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 15-26.
- Yalçın, C. (2004). *Göç sosyolojisi*. Ankara: Anı Yayıncılık, 21.
- Yalçın, N. (2013). *Kümeleme analizi ve uygulaması*. Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ, 7-25.
- Yapıcı, A. K. (2015). *Elit orta ve uzun mesafe koşucuların fizyolojik belirleyicilerinin araştırılması: Bir diskriminant analizi uygulaması*. Doktora Tezi, Gazi Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara, 29.
- Yaprak, B. (2007). *İlköğretim öğrencilerinin algıladıkları anne-baba tutumunun diskriminant analiziyle belirlenmesi ve benlik saygısı ile olan ilişkisinin değerlendirilmesi üzerine bir uygulama*. Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir, 15-34.
- Yıldız, Z. (1989). *Banka müşterilerinin demografik ve sosyo-ekonomik özellikler bakımından gruplandırılmasında kümeleme çözümlemesi ve bir uygulama*. Yüksek Lisans Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir, 39.
- Yılmaz, Ş. K. (2011). Kümeleme analizi ve pazarlamada kullanımı. *Akademik Yaklaşımlar Dergisi*, 2(1), 99.



EKLER

EK-1. Uygulama veri seti

Birim No	2017, 15+ yaş İllere Göç Eden Nüfusun Eğitim Durumu	Okuma Yazma Bilmeyen (%)	İlkokul (%)	Ortaokul (%)	Lise ve Dengi Meslek Okulu (%)	Lisans ve üstü (%)
1	Adana	2,3	9,1	16,9	34,0	37,8
2	Adıyaman	2,2	9,1	18,6	34,9	35,1
3	Afyonkarahisar	0,8	7,9	13,4	47,7	30,2
4	Ağrı	2,8	8,9	15,5	28,8	44,0
5	Aksaray	1,3	8,4	15,0	43,3	32,0
6	Amasya	1,4	11,2	17,4	41,6	28,4
7	Ankara	1,6	9,2	15,0	36,9	37,3
8	Antalya	1,5	11,2	20,4	37,3	29,6
9	Ardahan	2,2	11,2	17,4	34,8	34,4
10	Artvin	1,0	8,9	12,9	45,5	31,7
11	Aydın	1,2	9,8	16,1	43,2	29,7
12	Balıkesir	1,0	10,6	15,1	38,8	34,6
13	Bartın	2,0	11,6	19,1	45,3	22,0
14	Batman	3,4	9,4	22,9	26,0	38,2
15	Bayburt	1,4	6,9	9,3	59,4	22,9
16	Bilecik	0,8	8,9	16,1	47,0	27,3
17	Bingöl	2,4	8,4	15,3	42,3	31,7
18	Bitlis	3,0	8,8	18,4	29,7	40,2
19	Bolu	1,0	9,7	14,3	50,2	24,8
20	Burdur	0,5	7,9	12,4	55,8	23,3
21	Bursa	2,0	11,7	21,8	34,9	29,6
22	Çanakkale	0,6	8,1	12,4	50,1	28,9
23	Çankırı	1,8	18,8	19,5	38,4	21,6
24	Çorum	2,1	13,7	17,3	36,2	30,6
25	Denizli	0,8	8,3	15,7	45,2	30,0

EK-1. (devam) Uygulama veri seti

26	Diyarbakır	2,7	7,8	18,1	28,6	42,8
27	Düzce	0,8	9,1	17,1	47,8	25,2
28	Edirne	0,9	7,7	11,8	55,8	23,8
29	Elazığ	2,0	8,1	14,5	41,6	33,8
30	Erzincan	2,1	10,4	11,9	52,6	23,0
31	Erzurum	2,2	8,3	12,8	46,5	30,2
32	Eskişehir	0,9	8,5	13,8	46,8	30,0
33	Gaziantep	1,7	7,7	17,7	37,3	35,6
34	Giresun	2,5	18,1	20,1	38,3	21,0
35	Gümüşhane	1,8	16,7	20,9	42,6	17,9
36	Hakkari	1,5	4,2	14,7	32,0	47,7
37	Hatay	1,1	7,6	15,3	34,1	41,9
38	Iğdır	2,9	7,9	16,7	32,7	39,8
39	Isparta	0,6	6,3	10,2	55,0	27,8
40	İstanbul	2,9	14,7	21,6	32,7	28,1
41	İzmir	1,4	9,2	16,5	37,1	35,8
42	Kahramanmaraş	1,3	8,2	16,6	35,2	38,7
43	Karabük	0,9	7,3	10,2	60,5	21,0
44	Karaman	0,7	8,0	18,4	42,8	30,1
45	Kars	1,7	6,8	12,3	45,7	33,5
46	Kastamonu	2,6	19,0	16,5	40,5	21,4
47	Kayseri	1,3	8,7	15,6	42,4	32,1
48	Kilis	1,2	6,9	16,1	41,4	34,4
49	Kırıkkale	1,3	7,6	13,3	53,7	24,1
50	Kırklareli	0,9	11,2	16,7	45,0	26,2
51	Kırşehir	1,5	10,0	14,0	46,6	28,0
52	Kocaeli	2,1	13,0	24,3	34,5	26,0
53	Konya	0,9	8,3	13,6	46,2	31,0

EK-1. (devam) Uygulama veri seti

54	Kütahya	0,6	7,3	12,0	51,5	28,6
55	Malatya	2,1	9,9	15,8	38,4	33,9
56	Manisa	1,4	10,7	18,7	38,3	30,9
57	Mardin	4,3	9,8	20,4	26,6	39,0
58	Mersin	2,1	9,6	17,2	34,8	36,3
59	Muğla	0,8	9,0	16,6	37,6	36,0
60	Muş	3,0	8,4	16,4	27,0	45,1
61	Nevşehir	1,1	9,3	13,9	48,3	27,3
62	Niğde	1,1	9,8	15,3	48,7	25,1
63	Ordu	2,3	16,1	21,2	34,5	25,8
64	Osmaniye	1,7	8,2	17,4	35,8	36,8
65	Rize	1,6	11,3	17,6	40,8	28,7
66	Sakarya	1,4	10,8	18,5	43,8	25,6
67	Samsun	1,4	12,2	17,7	38,5	30,3
68	Şanlıurfa	3,4	9,4	18,5	23,4	45,4
69	Siirt	2,9	6,5	17,3	37,7	35,5
70	Sinop	2,7	20,6	19,6	36,1	21,0
71	Şırnak	2,1	5,2	13,8	27,2	51,8
72	Sivas	2,8	14,0	14,1	42,5	26,7
73	Tekirdağ	2,3	16,7	29,8	29,0	22,2
74	Tokat	2,3	18,7	27,3	31,2	20,4
75	Trabzon	1,3	9,4	13,3	48,8	27,3
76	Tunceli	2,5	8,2	14,1	37,3	37,9
77	Uşak	0,6	7,2	13,0	56,6	22,6
78	Van	2,6	8,1	17,9	28,8	42,6
79	Yalova	2,0	15,0	20,2	36,5	26,3
80	Yozgat	1,8	11,8	17,0	37,9	31,6
81	Zonguldak	1,2	9,1	14,2	50,3	25,2

EK-2. Diskriminant analizi sınıflama sonuçları (Tek bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı)

Birim No	Gerçek Grup No	En İyi Grup			
		Tahmin Edilen Grup No	P(D>d G=g)		P(G=g D=d)
			p	sd	
1	1	1	,882	4	1,000
2	1	1	,832	4	1,000
3	1	1	,810	4	1,000
4	1	1	,143	4	,985
5	1	1	,978	4	1,000
6	1	1	,873	4	1,000
7	1	1	,596	4	,999
8	1	1	,465	4	1,000
9	1	1	,768	4	1,000
10	1	1	,631	4	1,000
11	1	1	,968	4	1,000
12	1	1	,205	4	1,000
13	1	1	,153	4	,996
14	1	1	,031	4	1,000
15	1	1	,049	4	1,000
16	1	1	,720	4	1,000
17	1	1	,617	4	1,000
18	1	1	,482	4	1,000
19	1	1	,834	4	1,000
20	1	1	,516	4	1,000
21	1	1	,295	4	,998
22	1	1	,659	4	1,000
23	2	2	,829	4	,986
24	1	1	,100	4	,956
25	1	1	,726	4	1,000
26	1	1	,367	4	,995
27	1	1	,430	4	1,000
28	1	1	,480	4	1,000
29	1	1	,901	4	1,000
30	1	1	,036	4	,898
31	1	1	,389	4	1,000
32	1	1	,880	4	1,000
33	1	1	,590	4	,999
34	2	2	,779	4	,907
35	3	3	1,000	4	,941
36	4	4	1,000	4	,802
37	1	1	,092	4	,871
38	1	1	,443	4	1,000
39	1	1	,359	4	1,000
40	1	1	,061	4	,855

EK-2. (devam) Diskriminant analizi sınıflama sonuçları (Tek bağlantı kümeleme tekniği - karesi alınmış Öklid uzaklığı)

41	1	1	,786	4	,999
42	1	1	,423	4	,992
43	1	1	,136	4	1,000
44	1	1	,109	4	,999
45	1	1	,613	4	1,000
46	2	2	,770	4	,989
47	1	1	,979	4	1,000
48	1	1	,489	4	,997
49	1	1	,383	4	1,000
50	1	1	,662	4	1,000
51	1	1	,822	4	1,000
52	1	1	,020	4	,660
53	1	1	,835	4	1,000
54	1	1	,628	4	1,000
55	1	1	,921	4	1,000
56	1	1	,784	4	1,000
57	1	1	,011	4	1,000
58	1	1	,944	4	1,000
59	1	1	,264	4	,997
60	1	1	,163	4	,982
61	1	1	,918	4	1,000
62	1	1	,872	4	1,000
63	1	3(**)	,736	4	,611
64	1	1	,754	4	,999
65	1	1	,895	4	1,000
66	1	1	,619	4	1,000
67	1	1	,511	4	,999
68	1	1	,148	4	,996
69	1	1	,058	4	1,000
70	2	2	,932	4	,999
71	5	5	1,000	4	,826
72	6	6	1,000	4	,950
73	7	7	,834	4	1,000
74	7	7	,834	4	,932
75	1	1	,822	4	1,000
76	1	1	,470	4	1,000
77	1	1	,293	4	1,000
78	1	1	,440	4	,995
79	1	1	,074	4	,776
80	1	1	,660	4	1,000
81	1	1	,841	4	1,000

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : ÖZGEN, Hasan Gürcan
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 17.11.1987, İzmir
 Medeni hali : Bekar
 Telefon : 0 (544) 621 08 68
 e-mail : gurcan271@gmail.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / İstatistik Bölümü	Devam ediyor
Lisans	Hacettepe Üniversitesi / İstatistik Bölümü	2010
Lise	Batıkent Süper Lisesi	2005

İş Deneyimi

Yıl	Yer	Görev
2013-Halen	Aile, Çalışma ve Sosyal Hizmetler Bakanlığı	İstatistikçi

Yabancı Dil

İngilizce

Yayımlar

Özgen, H.G., Erbaş, S. (2018). Türkiye’de iç göç hareketlerinin eğitim seviyesine göre kümelenmesi üzerine bir çalışma. *Sosyal Politika Çalışmaları Dergisi*, 18(40/2), 141-160.

Hobiler

Bilgi Teknolojileri, Futbol, Müzik



GAZİ GELECEKTİR..