

T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ



OECD ÜLKELERİNİN KÜMELEME VE
DİSKRİMİNANT ANALİZİ İLE
SINIFLANDIRILMASI ÜZERİNE BİR
İNCELEME

YÜKSEK LİSANS TEZİ

DANIŞMAN

Prof. Dr. Mehmet GÜNGÖR

HAZIRLAYAN

Ayşegül HAMARAT

MALATYA-2022

**T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI**

YÜKSEK LİSANS TEZİ

**OECD ÜLKELERİNİN KÜMELEME VE DİSKRİMİNANT ANALİZİ
İLE SINIFLANDIRILMASI ÜZERİNE BİR İNCELEME**

Ayşegül HAMARAT

**Danışman
Prof. Dr. Mehmet GÜNGÖR**

MALATYA, 2022

ONUR SÖZÜ

Prof. Dr. Mehmet Güngör' ün danışmanlığında yüksek lisans tezi olarak hazırladığım **“OECD ÜLKELERİNİN KÜMELEME VE DİSKRİMİNANT ANALİZİ İLE SINIFLANDIRILMASI ÜZERİNE BİR İNCELEME”** başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığını ve yararlandığım bütün yapıtların hem metin içinde hem de kaynakçada yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

Tarih

Ayşegül HAMARAT

İmza



TEŐEKKÖR SAYFASI

Tez alıŐmalarımın yÖrÖtÖlmesi esnasında, bana danıŐmanlık ederek, izlemem gereken yolda beni yÖnlendiren ve her tÖrlÖ olanaĐı saĐlayan saygıdeĐer danıŐmanım Prof. Dr. Mehmet GÖNGÖR'e teŐekkÖrlerimi sunarım.

alıŐmalarım sırasında benden desteklerini ve anlayıŐlarını esirgemeyen, gÖc veren, maddi ve manevi olarak bana her an varlıklarını hissettiren anne- babama, ok deĐerli aileme her birine ayrı ayrı teŐekkÖrÖ bir bor bilirim. Saygılarımı sunarım.



ÖZET

Kümeleme analizi, birbirine benzeyen verileri aynı küme içinde sınıflandırır, yani gruplar içinde benzerlik yüksek olduğunda birbirlerine yakın birimleri, gruplar arasında ise uzak birimleri kümelere atama işlemi yapan çok değişkenli istatistiksel analiz yöntemidir. Diskriminant analizi, temel amacı ayırma olan, gözlemlerin özelliklerinden yararlanarak atanacağı grupları belirleyen veya daha önceden belirlenmiş grupları birbirinden ayıracak fonksiyonu bulurken kullanılan çok değişkenli istatistiksel analiz tekniklerinden biridir.

Yapılan bu çalışmada, diskriminant ve kümeleme analizi ile 2020 yılı baz alınarak 37 OECD ülkesi; ekonomik büyüme, askeri harcamalar, eğitim harcamaları, sağlık harcamaları, sübvansiyon ve transfer harcamaları, yatırım harcamaları değişkenleri açısından sınıflandırılmıştır. Kümeleme analizinde Belçika, Hollanda, Avusturya, Finlandiya, Fransa, İngiltere, Kanada, Almanya, İsviçre, Norveç, İsveç, Avustralya, Yeni Zelanda, İspanya, Japonya ve ABD'nin bir küme oluşturduğu; Macaristan, Slovenya, İtalya, Portekiz, Yunanistan, Kolombiya, İsrail, Güney Kore, Şili, Danimarka, İzlanda, Meksika, İrlanda ve Türkiye'nin ise ikinci bir küme oluşturduğu görülmüştür. Diskriminant analizi sonucunda, birinci kümede bulunan 16 ülkenin tamamını doğru sınıfladığı ve ikinci kümede de 21 ülkenin tamamını doğru sınıfladığı görülmüştür.

Kamu harcamaları bakımından transfer sübvansiyon harcamaları ve sağlık harcamaları diğer ülkelere kıyasen yüksek olanlar aynı grupta kümelenmiştir. Ekonomik büyümenin normal şartlar altında olması gerektiği oranı ile kamuya yapılan harcamaların ilişkisi bakımından pozitif yönde bir ilişki olduğu gözlemlenmiştir.

Anahtar Kelimeler: Diskriminant Analizi, Kümeleme Analizi, Kamu Harcaması.

ABSTRACT

Cluster analysis is a multivariate statistical analysis method that assigns similar data to clusters, that is, assigns units close to each other when similarity is high within groups, and assigns distant units to clusters between groups. Discriminant analysis is one of the multivariate statistical analysis techniques, the main purpose of which is separation, determining the groups to be assigned by benefiting from the characteristics of the observations, or finding the function that will separate the previously determined groups from each other.

In this study, 37 OECD countries, based on 2020 with discriminant and cluster analysis, economic growth, military expenditures, education expenditures, health expenditures, subsidy and transfer expenditures, and investment expenditures are classified in terms of variables. In the cluster analysis, Belgium, Netherlands, Austria, Finland, France, England, Canada, Germany, Switzerland, Norway, Sweden, Australia, New Zealand, Spain, Japan and the USA formed a cluster; Hungary, Slovenia, Italy, Portugal, Greece, Colombia, Israel, South Korea, Chile, Denmark, Iceland, Mexico, Ireland and Turkey formed a second cluster. As a result of the discriminant analysis, it was seen that all 16 countries in the first cluster were classified correctly and all 21 countries in the second cluster were correctly classified.

In terms of public expenditures, transfer subsidy expenditures and health expenditures that are higher than in other countries are clustered in the same group. It has been observed that there is a positive relationship between the rate of economic growth that should be under normal conditions and the rate of public expenditure.

Keywords: Discriminant Analysis, Cluster Analysis, Public Expenditure.

İÇİNDEKİLER

ONUR SÖZÜ.....	i
TEŞEKKÜR SAYFASI.....	ii
ÖZET	iii
ABSTRACT	iv
ŞEKİLLER LİSTESİ.....	vii
TABLolar LİSTESİ.....	viii
1.GİRİŞ.....	1
2. KÜMELEME ANALİZİ.....	2
2.1. Kümeleme Analizinin Tanımı	2
2.2. Kümeleme Analizinin Amacı.....	3
2.3. Kümeleme Analizinin Özellikleri.....	3
2.4. Kümeleme Analizinin Aşamaları	3
2.5. Kümeleme Analizinde Uzaklık ve Benzerlik Ölçüleri.....	4
2.5.1. Benzerlik Ölçüleri.....	4
2.5.2. Korelasyon Ölçüleri.....	5
2.5.3. Uzaklık Ölçüleri.....	5
2.6. Kümeleme Yöntemleri	6
2.6.1. Hiyerarşik (Aşamalı) Yöntemler	6
2.6.2. Hiyerarşik Olmayan Yöntemler.....	8
2.6.3. İki Aşamalı Kümeleme Yöntemi	10
3. DİSKRİMİNANT ANALİZİ (AYIRMA ANALİZİ)	11
3.1. Diskriminant Analizi Tanım.....	11
3.2. Diskriminant Analizinin Amacı	11
3.3. Diskriminant Analizinin İşleyişi.....	12
3.4. Diskriminant Analizinin Regresyon Analizi ve ANOVA ile İlişkisi	12

3.5. Diskriminant Analizi Varsayımları	13
3.6. Diskriminant Analizi Uygulama Aşamaları.....	14
3.7. İki Grup İçin Diskriminant Analizi	14
3.8. İki'den Fazla Grup İçin Diskriminant Analizi.....	15
3.9. Diskriminant Analizinin Geçerliliği.....	15
3.9.1. Sınıflandırma Tutarlılığı Tablosu	16
4. UYGULAMA	17
4.1. Kamu Harcamaları ve Ekonomik Büyüme.....	17
4.2. Uygulamanın Amacı.....	17
4.4. Veri Seti ve Yöntem.....	20
4.5. Hiyerarşik Kümeleme Yöntemi Uygulama	20
4.6. Diskriminant Analizi ile Uygulama	23
4.6.1. Bulgular	23
5. SONUÇ	27

ŞEKİLLER LİSTESİ

Şekil 2.1. Kümeler Arası Heterojen ve Kümeler İçinde Homojenlik	2
Şekil 2.2. Kümeleme Analizi Aşamaları.....	3
Şekil 2.3. Kümeleme Yöntemleri.....	6
Şekil 2.4. Hiyerarşik Kümeleme Yöntemi.....	6
Şekil 2.5. Hiyerarşik Olmayan Yöntemler.....	9
Şekil 4.1. Ağaç Grafiği (Dendrogram)	23



TABLÖLAR LİSTESİ

Tablo 3.1. Sınıflandırma Tutarlılığı Tablosu.....	16
Tablo 4.1. Değişkenler Tablosu.....	21
Tablo 4.2. Durum Özeti Tablosu.....	21
Tablo 4.3. Hiyerarşik Kümeleme Analizi Sonuçları Tablosu.....	22
Tablo 4.4. Box' M Testi Tablosu.....	24
Tablo 4.5. Özdeğer İstatistiği Tablosu.....	24
Tablo 4.6. Wilks' Lambda İstatistiği Tablosu.....	25
Tablo 4.7. Ayırma Fonksiyon Katsayısı Tablosu.....	25
Tablo 4.8. Yapı Matrisi Tablosu.....	26
Tablo 4.9. Kanonik Diskriminant Katsayıları Tablosu.....	26
Tablo 4.10. Grupların Ortalama Ayırma Fonksiyon Değerleri Tablosu	27
Tablo 4.11. Diskriminant Analizi Sınıflandırma Sonuçları Tablosu.....	27

1.GİRİŞ

Çok deęişkenli istatistiksel analizler, incelenen olay ve çevresindeki çok sayıda içsel ve dışsal faktörleri dikkate alarak, problemi doğasındaki yapısına ilişkin bilgilere göre incelemek ve çözümlere ulaşmak için geliştirilmiş yöntemler bütünüdür (Özdamar, 2002: 1). Çok deęişkenli istatistiksel analiz yöntemlerinin farklı amaçları vardır. Bu amaçlar şu şekildedir (Tatlıdil, 1996: 11-12):

1. Basitleştirme ve Boyut İndirgeme
2. Birimlerin Kümelenmesi ve Sınıflandırılması
3. Baęımlılık Yapısının İncelenmesi
4. Sıralama ve Ölçekleme
5. Çok Deęişkenli Hipotezlerin Oluşturulması ve Test Edilmesi

Diskriminant analizi ve kümeleme analizi çok deęişkenli istatistiksel yöntemlerden olan iki analiz türüdür. Her iki analiz de sınıflandırma amacı taşıması yönüyle birbirine çok benzemektedir fakat analizlerdeki farklılıklar iki analizi birbirinden ayırmıştır. Kümeleme analizi; büyük veri kümelerinden daha sade veri kümelerine ulaşmak ve araştırmacıya özetleyici bilgiler sunmak için veri gruplandırmasını yapar. Kümeleme analizinde daha önceden küme sayısı hakkında herhangi bir bilgi olmadığı zaman kullanılması uygundur. Eğer analiz yapılacak kümeler hakkında bilgiler mevcut ise kümeleme kullanılması anlamlı olmayacaktır. Kümeleme analizini hiyerarşik yöntemler ve hiyerarşik olmayan yöntemler olarak iki ana başlık olarak sınıflandırmak mümkündür. Sınıflandırma yapılırken benzerlikleri belirleyebilmek için uzaklık ve benzerlik ölçütleri kullanılır. Diskriminant analizinde ise X veri setindeki deęişkenlerin iki veya daha fazla gerçek gruplara ayrılmasını sağlayan, birimlerin gruplarına, optimal düzeyde ayrılmalarını sağlayacak fonksiyonlar türeten bir yöntemdir (Özdamar, 2002: 341).

Diskriminant analizinde; baęımlı deęişken kategorik, baęımsız deęişken ise sayısaldır. Bu analizde birimleri, gruplara atanmasını sağlayacak bir diskriminant fonksiyonu bulunur, bu fonksiyon grup ortalamaları arasındaki farkı en büyükleyecek şekilde belirlenir (Tatlıdil, 1996: 203).

Çalışmanın giriş bölümü birinci bölümdür. İkinci bölümünde, kümeleme analizi odak alınarak açıklanmaya çalışılmıştır. Üçüncü bölümde ise diskriminant analizi ve yöntemleri açıklanmaya çalışılmıştır. Dördüncü bölümde gerçekleştirilen uygulamanın bulgularına yer verilmiştir.

2. KÜMELEME ANALİZİ

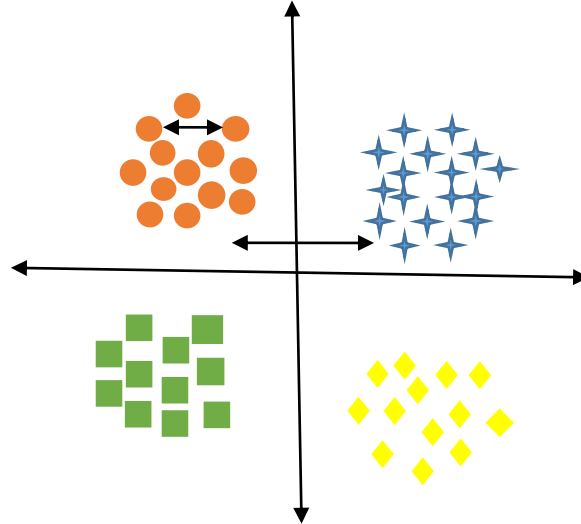
2.1. Kümeleme Analizinin Tanımı

Küme, birbirine benzeyen veya aynı cinsten olan nesnelerin oluşturduğu gruplara denir. Oluşturulan kümeleri gruplandırmaya da kümeleme denir. Çok değişkenli istatistiksel analizlerden biri olan kümeleme analizi ham gruplanmamış x veri matrisinde ki gözlemleri sahip oldukları özellikler çerçevesinde kümelemek amacıyla geliştirilmiş yöntemler topluluğu olarak tanımlanabilir (Alpar, 2011: 309). Bazen elde edilen çok sayıda veriyi gruplamak ve anlamlandırmak zorlaşmaktadır. Analiz ile analizcinin belirlediği kriterler baz alınarak kümeleme yapılmakta ve özet bilgiler oluşturulmaktadır.

Birbirine benzeyen veriler aynı küme içinde sınıflandırılır yani gruplar içinde benzerlik yüksek olduğundan birbirlerine yakın birimleri, gruplar arasında ise uzak birimlerin olduğunu görmek mümkündür. Tüm bunların sonucu olarak yapılan kümeleme işleminden sonra her bir öbek kendi içinde bağdaşık olmalı, farklı öbekler arasında bağdaşıklık olmamalıdır.

Analiz çok geniş kullanım alanına sahiptir. Bilim alanlarının büyük bölümünde aktif olarak kullanılmaktadır. Bunlardan bazıları pazarlama, bankacılık, eğitim, ziraat, tıp, biyoloji, zooloji ve sosyoloji gibi alanlarda kullanımına olanak sağlıyor. Tıpta hastalıkların, biyolojide canlı çeşitlerinin, ziraatta bitkilerin sınıflandırılmasında yaygın kullanım alanına sahiptir.

Şekil 2.1. Kümeler Arası Heterojen ve Kümeler İçinde Homojenlik



2.2. Kümeleme Analizinin Amacı

Kümeleme analizinin amacı, gruplanmamış gözlemleri benzerlik ve uzaklık ölçüsüne göre kümeleyerek ve araştırmaya uygun, özet bilgiler elde etmede yardımcı olmaktır. Analiz bu amaçlarla birlikte; gruplar için ön tahmin, hipotez testi ve uç değerlerin saptanması gibi amaçlarla da kullanılmaktadır.

Yukarıda bahsedilen amaçlara ulaşmak için aşağıda belirtilen aşamalar izlenmelidir:

- Verilere ait temel özelliklerin belirlenmesi
- Kümelere ait temel özelliklerin belirlenmesi
- Başlangıç merkezlerin belirlenmesi
- Uygun olan uzaklık ölçümünün belirlenmesi
- Kümeleme ölçütlerinin belirlenmesi
- Küme sayısının belirlenmesi
- Elde edilen sonuçların yorumlanması

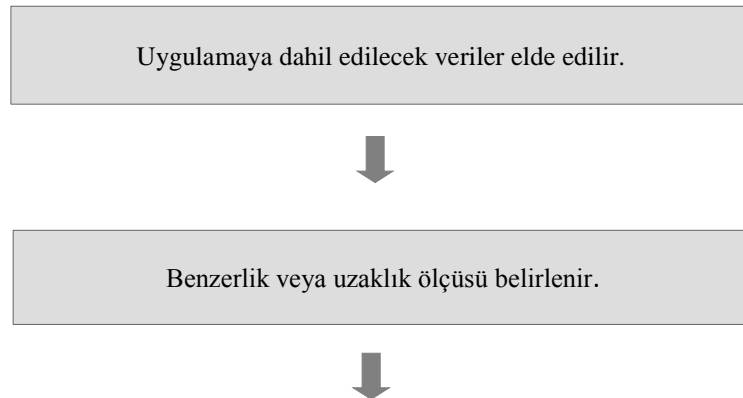
2.3. Kümeleme Analizinin Özellikleri

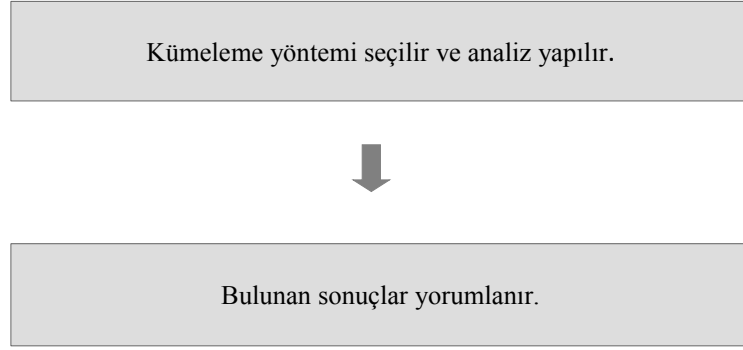
Kümeleme analizi değişkenleri, bağımlı ve bağımsız değişken olarak ayırım yapmaksızın sınıflandırma yapar. Verileri benzerliklerine göre sınıflandırarak gruplandırır. Analizde benzerlikler tayin edildikten sonra gruplar oluşturulmaktadır. Benzerliklerin oluşturulması analizcinin tercihine bırakıldığı için genellikle uzaklaşır. Sadece elde olan veri seti ile sonuçlar vereceğinden dolayı daha sonra kullanılamaz.

2.4. Kümeleme Analizinin Aşamaları

Analizi yapabilmek için açıklanacak olan adımları aşağıdaki şekilde belirtildiği gibi izlemek gerekiyor.

Şekil 2.2. Kümeleme Analizi Aşamaları





İlk olarak analizde kullanılacak olan verileri elde edip verileri bilgisayar ortamına aktararak veri girişi sağlanır. Yani ilk aşamada veri matrisini oluşturmak için ana kütlede seçilen n sayıdaki birimin incelenen p sayıda değişkenin gözlenen değerleri elde edilir. İkinci aşamada ise elde edilen verilere uygun benzerlik ölçüsü ile verilerin uzaklık matrisi oluşturulur. Bu ölçüme bakılarak üçüncü aşamada uygun olan kümeleme tekniğine karar verilerek uygulanır. Kümeleme işlemi tamamlanmış olup veriler kümelere ayrılmış olur. Son aşama ise analizden elde edilen sonuçların yorumlanması ile analiz tamamlanır. Yapılan kümeleme analizinin başarısı test etmek için geometrik çizimlerden yararlanılabilir. Eğer başarılı bir analiz gerçekleşmişse her bir kümedeki birimler kendi içinde benzer, diğer kümeler arasında farklılık olması gerekir.

2.5. Kümeleme Analizinde Uzaklık ve Benzerlik Ölçüleri

Analizde elde edilen verilerin aralarındaki benzerlikler itibariyle kümelemeyi yapar. Kümeleme analizi, birimleri değişkenler arası benzerlik ya da uzaklıklara dayalı olarak hesaplanan bazı ölçülerden yararlanarak homojen gruplar oluşturmaya çalışır (Özdamar, 2002: 279). Bu işleme geçerken hangi ölçütün kullanılması gerektiğine karar vermek gerekmektedir. Burada verinin tipinin kategorik- metrik, sıklık sayıları veya iki kategorili olması ölçütü belirlemede öncülük etmektedir.

Kümeleme analizinde elde edilen verilerin normal dağılıma uyması yönünde bir kural olmakla birlikte bu kural sadece prensipte kalır. Aslında burada önemli olan uzaklık değerlerinin normal dağılıma uyması gerektiğidir.

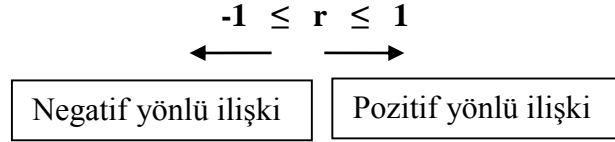
2.5.1. Benzerlik Ölçüleri

İki veri arasındaki benzerliklere göre yani verilerin birbirine olan benzerliklerini belirleyebilme ölçütü olarak benzerlik ölçüleri kullanılır. Bu değer sıfır ile bir arasında

değerler alabilmektedir. Benzerlik ölçülerinin; Pearson İlişki Katsayısı, Kosinüs Benzerlik Ölçüsü gibi çeşitleri vardır.

2.5.2. Korelasyon Ölçüleri

Benzerliklerin ölçümünde kullanılan diğer bir ölçü birimi ise korelasyon ölçümüdür. İki veri arasındaki ilişkiyi belirleyen ölçüm -1 ile 1 arasında değerlere sahip olur.



2.5.3. Uzaklık Ölçüleri

İki veri arasındaki uzaklığın ölçülebilmesi için uzaklık ölçüleri kullanılır. Uzaklık ölçüsü de benzerlik ölçüsü gibi sıfır ile bir arasında değere sahiptir. Veri tipinin farklılıklarına göre kullanılacak uzaklık ölçüleri değişkenlik gösterebilir. Mesela sayısal veriler için; Öklit uzaklığı, Minkowski Uzaklık Ölçüsü, Chebychev Uzaklık Ölçüsü, Manhattan City-Block Uzaklık Ölçüsü kullanılabilir. Sıklık sayıları için; Ki-kare Uzaklık Ölçüsü, Phi-Kare Uzaklık Ölçüsü kullanılabilir.

Uzaklık ölçüsünün hesaplanması için birçok yöntem vardır. Uzaklık ölçülerinden en çok kullanılan birkaç yöntem değinilecektir.

Öklid uzaklık ölçüsü, uzaklık ölçülerinin hesaplanması için diğer ölçülere göre daha fazla kullanılan ölçülerdendir. En önemli özelliği; iki birim arası mesafe analize daha sonra eklenen birimlerin uzaklık değerini değiştirmemesidir. İki nesne arasındaki mesafeyi ölçmek için öklid uzaklığı, değerler arasındaki farklarının karelerinin toplamının karekökünü alarak hesaplar. Öklid bağıntısı ile iki nokta arasındaki uzaklığın hesaplanması üçgenin hipotenüs hesaplaması olarak ifade edilir.

Öklid uzaklık ölçüsü ikiden fazla değişken için genel formülü aşağıdaki gibi hesaplanır (Hair vd., 2014: 431):

$$D(x_i, x_k) = \sqrt{\sum w_j (x_{ij} - x_{ik})^2} \quad (2.1)$$

Manhattan Uzaklığı, City-Blok uzaklığı olarak da bilinen uzaklık yönteminde iki nokta arasındaki farkın mutlak değerlerinin toplamı uzaklık değerini verecektir.

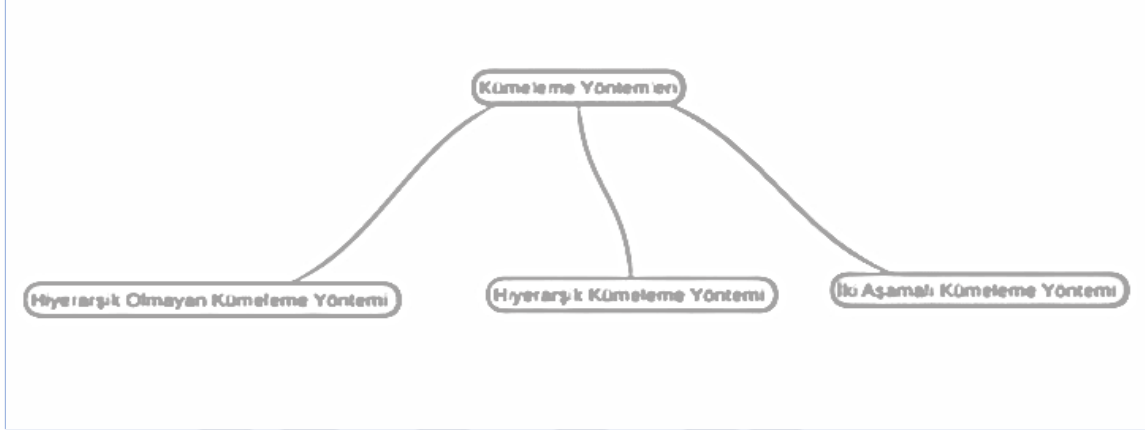
Kare öklid uzaklık ölçüsü, iki nokta arasındaki farkın karelerinin alındıktan sonra toplanması ile hesaplanması yapılır. Yani öklid uzaklığının karesi alınmış olduğu durumdur.

$$D(i, j) = (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2 \quad (2.2)$$

2.6. Kümeleme Yöntemleri

Kümeleme yöntemleri üç ana yöntem ayrılmaktadır. Bunlar; hiyerarşik ve hiyerarşik olmayan kümeleme yöntemi ve iki aşamalı kümeleme yöntemidir. Hiyerarşik ve hiyerarşik olmayan yöntemlerin kendi altında farklı yöntemleri mevcuttur. Bunu aşağıdaki gibi ifade etmek mümkündür;

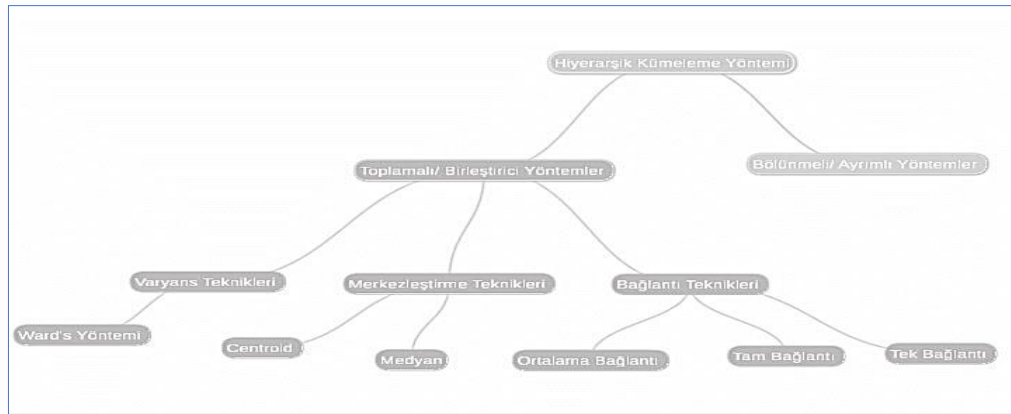
Şekil.2.3. Kümeleme Yöntemleri



2.6.1. Hiyerarşik (Aşamalı) Yöntemler

Hiyerarşik kümeleme yöntemleri bağlantı teknikleri olarak da adlandırılır. Yöntemin işleyişi gözlemleri birbirleri ile aşama aşama toplayarak; art arda kümeler kurar ve bu kümelere atanacak değerlerin uzaklık veya benzerlik ölçüsünün düzeyini belirler.

Şekil 2.4. Hiyerarşik Kümeleme Yöntemi



Hiyerarşik yöntemleri iki ana başlıkta incelenir. Bu ana başlıklar altında birçok yöntemi incelemek mümkündür. Bu yöntemler aşağıdaki şekilde ifade edildiği gibidir:

Hiyerarşik kümeleme yöntemleri daha çok küçük örneklemeler (genellikle $n < 250$) için kullanılması daha uygundur. Analizi yaparken benzerlik veya uzaklık ölçüsünün seçilmesi ve kümelerin birleşmesi veya ayrılması ile ilgili tekniklere karar vermek gerekmektedir.

Analizin başlangıcında kaç grup oluşturulması gerektiğinin bilinmediği durumlarda hiyerarşik kümeleme yöntemlerinin kullanılması uygundur. Analizcinin ön göremediği benzerlikleri analiz değerlendirmeye alacağından dolayı bu durum analizin avantajı olmuştur.

Hiyerarşik kümeleme yöntemlerinde işlemler büyük kümeden benzerliklere göre ayrılan kümelerin varlığıyla yani tümdengelim yöntemiyle veya tümevarımsal olarak bütün verileri tek küme olarak görüp kümeleri benzerliklerine göre birleştirerek ilerleyen yöntemdir. Kolay bir kullanıma sahiptir ama veri sayısının fazla olduğu durumlarda gruplandırma işlemi uzun zaman alacağından küçük veri setlerinde kullanılması uygun görülmüştür.

2.6.1.1. Toplamalı/Birleştirici Yöntemler

Birleştirici yöntemlerde her gözlem tek başına bir kümedir. Yani gözlem sayısı ile küme sayısı birbirine eşittir. Birçok sayıdaki küme benzerliklerine göre birleştirilmeye başlar. Tek küme kalana kadar bu aşamalar tekrarlanır.

Birleştirici yöntemleri kullanan algoritmaların işleniş yapısı genel olarak aşağıdaki gibi dört adımlıdır (Johnson, 1998: 323-324):

1. Tüm gözlemler birer küme olarak başlanır (yani, her gözlem tek üyeli bir küme oluşturur), böylece küme sayısı gözlem sayısına eşittir.
2. Bir benzerlik ölçüsünü kullanarak, en benzer iki kümeyi yeni bir kümede birleştirilir (iki gözlem içeren durum), böylece küme sayısını bir azaltır.
3. En benzer iki kümeyi yeni bir kümeleme işleminde birleştirmek için benzerlik ölçüsünü kullanarak kümeleme işlemi tekrarlanır.
4. Her bir adımda en benzer iki kümeyi yeni bir kümede birleştirerek bu işleme devam edilir. Tüm gözlemler bir kümede tutulana kadar işlem toplamda $(n-1)$ kez tekrarlanır.

2.6.1.1.1. Merkezileştirme Teknikleri

Küme içindeki gözlemlerin ortalaması bulunur ve kümeleme bu değer etrafında gerçekleştirilir.

2.6.1.1.2. Bağlantı Teknikleri

Kümeler arasında farklı şekillerde bağlantıların birleştirilmesi ile ortaya çıkan tekniklerdir.

Tam Bağlantı: En uzak komşuluk olarak da adlandırılan bu yöntemde isminden de anlaşılacağı gibi birbirine en uzak iki gözlem seçilir ve bir kümeye atanır. Bu şekilde kümeler birleştirilerek sonuca ulaşılabilir.

Tek Bağlantı: Tam bağlantı yönteminin tam tersi şeklinde ilerletilir. En yakın komşuluk olarak da adlandırılan bu yöntem isminden de anlaşılacağı gibi birbirine en yakın iki gözlem seçilir ve bu iki gözlem bir küme oluşturur. Seçilen gözleme en yakın gözlemler seçilerek kümeye atanır benzerlik olmayan gözlemler ise farklı kümelere atanır. Bu işlemler yapılırken uzaklık matrisinden faydalanılarak yapılır. Bu yöntemin dezavantajı olarak birbirine yakın değerler seçilirken bir gözleme yakın olan değer başka bir gözleme uzak kalabileceğinden dolayı bu yöntemi ayrıklığı net olan analizlerde kullanmak daha iyi sonuçlar vermektedir.

Ortalama Bağlantı: Kümelemeye başlarken ortanca değeri bulup kümelemeye ortadaki değeri esas alarak başlar. Tek bağlantı ve ortalama bağlantı teknikleri uç değerlere duyarlı kalmamasına rağmen ortalama bağlantı yöntemi uç değer problemini en aza indirger.

2.6.1.1.3. Varyans Teknikleri

Varyans hesaplamasına göre kümeleme yapan bir yöntemdir.

Ward's Yöntemi: Kümeler içinde en küçük hata kareler toplamının olduğu kümeleri birleştirir. Her aşamada yeniden hesaplanan hata kareler toplamına göre aşamalar ilerletilir.

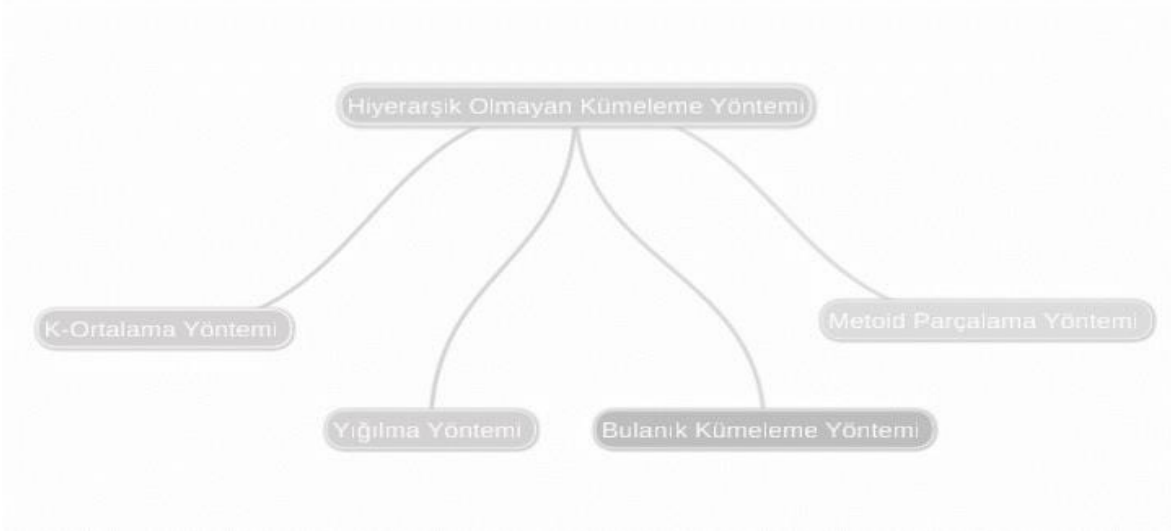
2.6.1.2. Bölünmeli/Ayrımlı Yöntemler

Bölünmeli yöntemlerde bütün gözlemleri ilk başta hepsini bir küme olarak alır. Var olan tek küme benzerliklere göre ikinci aşamada bir alt kümeye daha ayrılır. Bu aşamalar her bir gözlemin tek bir kümede olana kadar yapılır ve analiz sonlandırılır.

2.6.2. Hiyerarşik Olmayan Yöntemler

Analize başlamadan önce küme sayısının belli olduğu zamanlarda hiyerarşik olmayan kümeleme yöntemleri kullanılır. Veri sayısının çok büyük olduğu durumlarda kullanılması daha uygundur. Uç değerlere daha az duyarlı olduğundan uç değer problemi en aza iner. Yöntemler; bulanık kümeleme yöntemi, metoid parçalama yöntemi, yığılma yöntemi ve k-ortalama yöntemi gibi yöntemleri kapsar. Bu yöntemlerden en sık kullanılan k-ortalama yöntemidir. Bu yöntemlerin içinden bulanık kümeleme yöntemine ve k-ortalama yöntemine kısaca değinilecektir. Bu yöntemler Şekil 2.5'de ifade edildiği gibidir:

Şekil 2.5. Hiyerarşik Olmayan Yöntemler



Analizci hiyerarşik olmayan kümeleme yöntemi uygularken iki aşama gerçekleştirir. Birinci aşama: daha önceden belirlemiş olduğu kümelerin başlama noktasını belirler veya rastgele olarak belirlenmesini sağlar. İkinci aşama: Başlama noktası belirlenen kümelere verilerin atama işlemi bu aşamada sağlanır.

2.6.2.1. Bulanık Kümeleme Yöntemi

Bulanık kümeleme yöntemi karmaşık yapıdaki veri grupları için yani birbirinden belirgin şekilde ayrılamayan, hangi veri hangi kümeye daha uygun olduğunun tam olarak belirlenemediği durumlarda kullanılması uygundur. Klasik kümelemede verinin kümeye olan üyeliği bellidir. Üyelik varsa kümeye dahil edilir, üyelik yoksa kümeye dahil edilemez fakat bulanık kümelemede üyelik derecesi oranında üyeliği mümkün olur.

Diğer kümeleme yöntemlerinde olduğu gibi bulanık kümeleme yönteminde de uzaklık ölçüleri kullanılarak analiz yapılır. Klasik kümeleme analizi yöntemlerindeki her bir verinin tek bir kümeye atanma sınırını aşarak, bulanık kümeleme yönteminde bir verinin belli üyelik dereceleriyle bütün kümelere ait olmasını mümkün kılar. Daha önce herhangi bir kümeye ataması yapılamayan bir verinin üyelik derecesinin belirlenip bulanık kümelemede üyelik derecesi oranında bir kümeye aitliği sağlanır.

Bulanık kümeleme yönteminde, grup üyeliklerinin olasılıklarının toplamı en fazla 1 olabilir. Bir gözlemin atanabileceği kümeler 0 ile 1 arasında değer alır. Bulanık kümeleme analizi uygulanırken iki durumun sağlanması gerekir. Birincisi küme içinde hata minimizasyonu sağlamak ikincisi ise kümelerin sayısını doğru seçmektir. Bu iki durum birbirine ters düştüğü için birlikte sağlanması sıkıntılıdır çünkü küme sayısı küçültülmeye çalışılırken hatalar maksimum olur.

2.6.2.2. K- Ortalama Yöntemi

K- ortalama yönteminde her gözlemin tek bir kümeye aitliği sağlanır. Bu yöntemde, gözlemlere en yakın olan kümeye atama yapılır. En sık kullanılan kümeleme yöntemi olan k- ortalama, küme sayısı en az iki en çok ise veri içindeki gözlem sayısı ile aynı olacak şekilde verilerin kümelenebilir. Kümeler içinde homojenlik, kümeler arasında heterojenlik olacak şekilde kümelenebilir.

Hata kareler toplamının en az seviyede olmasından güç alarak verilerdeki en uygun ayrılmayı tespit etmeyi amaçlayan tekniktir. K- ortalama tekniğinde, genellikle öklid uzaklık ölçüsü kullanılır.

2.6.3. İki Aşamalı Kümeleme Yöntemi

Veri yapısının tek bir yapıda olmayıp iki farklı yapıdaki veri tipinin yani kategorik ve sürekli değişkenlerin bir arada bulunduğu veri kümesinde kümeleme analizi yapılması gerekiyorsa iki aşamalı kümeleme yöntemi kullanılır. Veri sayısının büyük olduğu durumlarda kullanılabilir. Yöntem uygulanırken küme sayısını analizcinin belirlemesine gerek kalmadan uygun sayıda kümeyi kendiliğinden oluşturur. Bu yöntemde iki uzaklık ölçüsü kullanılabilir. Bunlar log-olabilirlik uzaklığı ve öklid uzaklığıdır. Adından da anlaşılacağı gibi iki aşamadan oluşan bir süreçten geçer.

Birinci aşama: Veri setinde her bir gözlemi ön aşama olarak kümelenebilir.

İkinci aşama: İlk etapta kümelenebilir kümeleri tekrar ele alarak hiyerarşik kümeleme yaklaşımı uygulanır.

3. DİSKRİMİNANT ANALİZİ (AYIRMA ANALİZİ)

3.1. Diskriminant Analizi Tanım

Çok değişkenli istatistiksel analizlerden biri olan diskriminant analizini 1930'lu yıllarda Ronald A. Fisher ilk bilimsel araştırma olarak tanıtmıştır. Diskriminant analizi, X veri setindeki değişkenlerin iki veya daha fazla gerçek gruplara ayrılmasını sağlayan, birimlerin p tane özelliğini ele alarak bu birimlerin doğal ortamdaki gerçek gruplarına, sınıflarına optimal düzeyde ayrılmalarını sağlayacak fonksiyonlar türeten bir yöntemdir (Özdamar, 2002: 341). Kümelenen grupların birbirlerinden nasıl farklılaştığının, gruplarda gözlemlerin hangi kümede bulunacağını ve kümelemede öne çıkan özelliklerin hangileri olduğunun bilgisini veren analiz türüdür.

Diskriminant analizinde yapısındaki grup sayısı bilinmekte ve bu verilerden faydalanarak bir ayırma modeli elde edilmektedir. Kurulan bu model yardımı ile veri kümesine yeni alınan gözlemlerin gruplara atanması yapılmaktadır (Başarır, 1990: 1).

Genel anlamda ayırma olup, bireylere ait p tane özellikten yararlanarak ait oldukları grupları belirlemede veya mevcut grupları birbirinden ayıracak en iyi fonksiyonu bulmada kullanılan çok değişkenli istatistik tekniklerinden birisidir. Bazı araştırmacılar diskriminant analizinde ayırma fonksiyonu katsayılarının hesaplanmasında başvurulan yöntemlere göre diskriminant analizini, kanonik diskriminant analizi, en çok olabilirlik diskriminant analizi ve bayes diskriminant analizi şeklinde adlandırılırlar.

3.2. Diskriminant Analizinin Amacı

Diskriminant analizinin temel iki amacı bulunmaktadır. Bunlar; ayırma ve sınıflandırmadır. Analizin bu amaçlarından dolayı eğer diskriminant analizi bir ayırma fonksiyonu belirlemeye yönelik olarak uygulanmış ise tanımlayıcı diskriminant analizi ve eğer sınıflama amacıyla uygulanmış ise ayırıcı diskriminant analizi olarak isimlendirilir (Özdamar, 2002: 320). Analiz, farklılığa sebep olan değişkenleri belirler ve böylece grupları farklılaştıran faktörlerin belirlenmesini sağlar. Analiz sonuçlarına göre yapılan sınıflama ile önceden belli olan grupların karşılaştırılması fonksiyonun başarısını test eder.

Diskriminant analizinin amaçlarını aşağıdaki gibi açıklamak mümkündür (Kalaycı, 2010: 335):

1. Grup üyeliğini tahmin etmek, başka bir deyişle, bir verinin hangi değişken grubuna gireceğine karar vermek için kullanılabilir.

2. Ayırma fonksiyon eşitliğini kullanarak, verilerin gruplara ayrılmasına yardımcı olur.
3. Bağımsız değişkenlerin aritmetik ortalamalarının gruplar arasında nasıl değiştiğini tespit etmek için kullanılabilir.
4. Grupları ayırmada etkili olan ve olamayan değişkenleri belirlemek için kullanılabilir.
5. Bağımlı değişkenin varyansının ne kadarının bağımsız değişkenler tarafından açıklanabildiğini belirlemek için kullanılabilir.

3.3. Diskriminant Analizinin İşleyişi

Diskriminant analizinde; bağımlı değişken kategorik, bağımsız değişken ise sayısaldir. Bu analizde birimleri, gruplara atanmasını sağlayacak bir diskriminant fonksiyonu bulunur, bu fonksiyon grup ortalamaları arasındaki farkı en büyüklenecek şekilde belirlenir (Tatlıdil, 1996: 203).

Analiz iki grup veya daha fazla gruba ayırma yapılması durumunda diskriminant fonksiyonları üreterek ayırma işlemini bu formüller üzerinden gerçekleştirir. Analiz ayırma yapabilmek için her gruba ait formüller çıkarır. Bu formüller diskriminant fonksiyonu olarak adlandırılır. Diskriminant fonksiyonlarının varlığıyla sınıflandırma kolaylıkla yapılır. Fonksiyonunun avantajı ile veri grubuna daha sonradan eklenmek istenen gözlemleri analize dahil edilebilir. Formüller aracılığıyla hangi gruba dahil edileceği tespit edilir ve gruba ataması yapılır.

Oluşturulan fonksiyonlara önemlilik testi yapılır daha sonra grup içi varyans-kovaryans matrisi oluşturulur.

3.4. Diskriminant Analizinin Regresyon Analizi ve ANOVA ile İlişkisi

Analiz, regresyon analizi ve ANOVA (tek yönlü varyans analizi)'nin çok değişkenli uzantısı olan MANOVA (çok yönlü varyans analizi) ile yakınlık gösterir. Diskriminant analizi; bir bağımlı kategorik değişken ve birden fazla bağımsız metrik değişken ile işlev görür. Regresyon analizi ve ANOVA ile diskriminant analizinin farklılığı burada karşılaşılmaktadır. Analizlerin üçünde de bağımlı değişken sayısı bir tane bağımsız değişken sayısı ise birden fazladır fakat değişkenlerin yapısı farklılık göstermektedir. Regresyon analizinde bağımlı ve bağımsız değişken metrik yapıda iken ANOVA'da bağımlı değişken metrik, bağımsız değişken kategorik diskriminant analizinde ise bağımlı değişken kategorik, bağımsız değişken metrik yapıda olması gereklidir.

Regresyon analizinde bağımsız ve bağımlı değişkenlerin değerlerini baz alarak bağımlı değişkenin bağımsız değişkene göre sahip olacağı değerler kestirilir. Diskriminant analizinde ise bağımsız değişkenlerin sınıf üyelikleri tahmin edilir Üçten fazla grup var olduğunda regresyon analizi, gözlemleri gruplara ayırması diskriminant analizine kıyasen eksiklikleri vardır.

3.5. Diskriminant Analizi Varsayımları

Bir gruba ya da bir veri setine diskriminant analizi uygulayabilmek için grubun veya veri setinin ANOVA ve MANOVA uygulaması için gerekli olan aşağıdaki özellikleri taşımalıdır (Özdamar, 2002: 343-344):

- X veri matrisi çok değişkenli normal dağılım göstermelidir.
- X matrisinde yer alan değişkenlerin varyans ve kovaryansları homojen olmalı ve çok değişkenli toplumdan çekilmiş örnek olmalıdır.
- Değişkenler arasında çok yönden bağımlılık bulunmamalıdır.
- Değişkenlerin varyansı ve ortalamaları arasında bir korelasyon bulunmalıdır.
- X matrisinde grupların birbirinden ayrılmasında rol almayacak gereksiz değişkenler bulunmamalıdır.

Bunun yanında grup sayısı ikiden fazla olmalı, her grup içinde en az iki gözlem bulunmalı, gruplara atama yapılırken kullanılan diskriminant fonksiyonları (ayırıcı değişken) bir başka diskriminant fonksiyonu ile doğrusal bileşimi olmamalıdır. Analiz için gerekli olan iki temel varsayım, değişkenlerin normalliği ve varyans- kovaryans eşitliğinin sağlanmasıdır. Bunlar aşağıda açıklanmıştır:

-Çok Değişkenli Normallik Varsayımı

İstatistiksel analizlerin genelinde normal dağılıma uygunluk öncelikle bakılması gereken kriterdir. Veri setinin normal dağılıma uygunluğu ile analiz yapılır. Çok değişkenli normallik varsayımında birden fazla değişkenin her biri ayrı ayrı veya diğer kombinasyonlarının hepsinin normal dağılıma uygun olduğu varsayılmaktadır. Çok değişkenli normal dağılım için Ki-Kare testinden yararlanılabilir (Şahin, 2017: 11).

-Varyans- Kovaryans Matrislerinin Homojenliği Varsayımı

Varyans- kovaryans matrisleri homojenlik göstermediğinde ve örneklem büyüklükleri yeterli olmadığında, diskriminant fonksiyonları tahmin edilirken istatistiksel anlamlılığı olumsuz yönde etkilenebilir. Ayrıca, örneklem büyüklükleri yeterli olduğu halde varyans- kovaryans matrislerinin homojen olmaması gözlemleri kovaryansı daha yüksek olan

gruplara sınıflandırabilmektedir (Şahin, 2017: 11). Varyans- kovaryans matrisinin homojenliğini test etmek için Box M testi kullanılır.

Örneklem büyüklüğü: Grupların n sayılarının, yani örneklem büyüklüklerinin eşit olması gerekmemektedir. En küçük grupta yer alan birey ya da birim sayısının, bağımsız değişken sayısından fazla olması gerekliliğidir (Çokluk vd., 2010: 112).

Uç değerler: Diskriminant analizi uç değerlere çok duyarlıdır. Öncelikle her grupta tek yönlü (univariate) ve çok yönlü (multivariate) uç değerlerin taranması ve bunların ya dönüştürülmesi ya da çıkartılması gerekir (Çokluk vd., 2010: 112).

Çoklu doğrusal bağlantı: Eğer bağımsız değişkenlerden biri, diğer bir bağımsız değişkenle yüksek korelasyon veriyorsa, çoklu doğrusal bağlantı problemi ortaya çıkar, dolayısıyla bağımsız değişkenler arasında çoklu doğrusal bağlantı probleminin olmaması gerekir (Çokluk vd., 2010: 112).

3.6. Diskriminant Analizi Uygulama Aşamaları

Analiz uygulama aşamaları üç ana aşama olarak gerçekleştirilir.

1. Aşama: Çıkarsama: Gruplar arasında farklılığın varlığına bakılır. Eğer farklılık varsa analize devam edilir farklılık yoksa analiz sonlandırılır yani farklılığın olmaması gruplar arasında eşitlik vardır ve grupların ayrılması mümkün olamayacaktır bu durumda diskriminant analizi yapılamaz. Farklılığın varlığı halinde bağımlı ve bağımsız değişkenlere karar verilir. Değişkenler seçildikten sonra aralarında çoklu bağlantı probleminin varlığına bakılır. Diskriminant fonksiyonları elde edilir ve fonksiyonlara göre grup üyelikleri belirlenerek sınıflandırma yapılır.

2. Aşama: Onaylama: Her gözlem ayırma yapıp kümelendirildikten sonra diskriminant fonksiyonunun önemliliğini test etmek için doğru sınıflandırma yüzdesine bakılır.

3. Aşama: Yorumlama: Bütün aşamalar gerçekleştirildikten sonra sınıflama ile oluşturulan gruplar analizci tarafından yorumlanır.

3.7. İki Grup İçin Diskriminant Analizi

Bireylere ait p tane özelliğin olması durumunda bu özelliklerin her birinin ayrı ayrı ele alınarak bireylerin sınıflara ayrılması gerçekten güç ve hatta bazı durumlarda olanaksızdır. Bu nedenle ayırma analizinde amaç, çok değişkenli problemin tek değişkenli biçime dönüştürülmesidir. Yani tüm değişkenlerin uygun ağırlıklarla katılacağı tek bir fonksiyonun elde edilmesidir (Tatlıdil, 1996: 203).

İki gruplu diskriminant analizi, elde edilen fazla sayıda birimlerin iki gruba ayrılması istenildiği zaman kullanılan yöntemdir. Diskriminant analizinin uygulanabilmesi için her gruba ait özel fonksiyonlar üretildiğinden ve her fonksiyona ait vektör hesaplaması yapıldığından daha önce bahsetmiştik. Bu iki gruplu diskriminant analizi içinde geçerlidir.

Diğer analizlerden farklı olarak diskriminant analizinde öncelikle gözlemlerin ait olduğu grupları belirler ve her grup için özellikleri belirleyebilmek adına bir diskriminant fonksiyonu belirlenir. İki gruplu diskriminant analizi için iki ayırıcı fonksiyon belirlenir.

Bir bireye ait n değişkenden elde edilecek olan fonksiyon,

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in} \quad (3.1)$$

Bu diskriminant (ayırıcı) fonksiyonu bulunurken, gruplar arası varyansı grup içi varyansa göre en büyük değeri alması gerekmektedir. Formül olarak aşağıdaki gibi gösterilmektedir;

$$F = \text{Max} \left(\frac{\text{Gruplar Arası Varyans}}{\text{Grup İçi Varyans}} \right) \quad (3.2)$$

3.8. İki'den Fazla Grup İçin Diskriminant Analizi

İki'den fazla grup için uygulanan diskriminant analizi, iki grup için uygulanan diskriminant analizinin genel bir şekilde ifade edilmesi ile uygulanan bir halidir. Analiz, elde edilen fazla sayıda birimlerin iki'den fazla gruba ayrılması istenildiği zaman kullanılan yöntemdir. İki gruplu diskriminant analizinde tek bir ayırıcı fonksiyon yeterliyken iki'den fazla gruplu analizde yeterli olmayabilir. Bundan dolayı ayırıcı fonksiyon sayısı gruplamadaki ihtiyaca göre artırılmaktadır. Burada önemli olan nokta gruplamada kullanılacak olan diskriminant fonksiyonlarını minimum düzeyde tutabilmektir.

İki'den fazla grup olması durumunda gözlemleri gruplara atama işlemi, gruplar arasında ayırma gücü en fazla olacak şekilde belli sayıda lineer bağıntı yardımıyla yapılmaktadır. İki'den çok grup olması halinde elde edilen ayırma fonksiyonunun önem kontrolünde kullanılan kriterlerden ilki, Wilks tarafından geliştirilmiş olan ve genelleştirilmiş varyans olarak bilinen Λ 'dır (Tatlıdil, 1996: 210).

Ayırıcı değişkenlerin üzerinde çok değişkenli bir ölçüsü olan Wilks'in Λ istatistiği 0 ile 1 değerleri arasında değişkenlik gösterir. Λ değerleri sıfıra yaklaştıkça iyi bir ayrımı göstermektedir, 1 değerine yaklaştıkça daha az ayrımın gerçekleştiğini bildirmektedir.

3.9. Diskriminant Analizinin Geçerliliği

Analizin başarısını ölçmek için kullanılan yaklaşım sınıflandırma tutarlılığı tablosu mevcuttur. Aşağıda bu yaklaşım incelenecektir.

3.9.1. Sınıflandırma Tutarlılığı Tablosu

Analizin geçerliliğini ölçmek için kullanılan en kolay yaklaşım sınıflandırma tutarlılığı tablosu yaklaşımıdır. Diskriminant analizi sonunda gruplara atanan üyelerin sınıflandırma tutarlılığı tablosu ile incelenmesidir.

Tablo 3.1. Sınıflandırma Tutarlılığı Tablosu

Gerçek Grup	Tahmin Edilen Grup			Toplam
	1	2	3	
1	F ₁₁			
2		F ₂₂		
3			F ₃₃	
Toplam				N

Tabloda verilen bilgilere göre F₁₁, F₂₂, F₃₃ ile ifade edilen köşegen birimleri doğru sınıflama sıklıklarını bunun dışındaki birimler oluşacak hatalı sınıflama sıklıklarını göstermektedir. Köşegen birimlerinin toplamı genel toplamda oranlandığı zaman bulunan sınıflandırma oranı geçerlilik oranı olarak adlandırılır.

$$\left(\frac{F_{11} + F_{22} + F_{33}}{n} \right) \quad (3.3)$$

Bu oranın 1 olması mükemmel bir sınıflandırma gerçekleştiğini gösterir.

Denklemlere ait H₀ ve H₁ hipotezi oluşturulur.

H₀: Diskriminant analizinde; bağımsız değişkenlerin doğrusal fonksiyonları istatistiksel açıdan yazılamaz.

H₁: Diskriminant analizinde; bağımsız değişkenlerin doğrusal fonksiyonları istatistiksel açıdan yazılabilir.

Şeklinde H₀ hipotezini θ istatistiği yardımıyla test edebiliriz. Press Q istatistiği de denilen bu istatistik "1" serbestlik dereceli bir χ^2 dağılımıdır (Alpar, 2011: 708-711).

$$Q = \frac{(n-k)\sum f_{ii}^2}{n(k-1)} \quad (3.4)$$

4. UYGULAMA

4.1. Kamu Harcamaları ve Ekonomik Büyüme

Dar anlamda kamu harcamaları denildiği zaman devletin para halinde yaptığı harcamaların toplamı akla gelmektedir (Aksoy, 1994: 105). Geniş anlamda kamu harcamaları ise başta devlet olmak üzere kamu tüzel kişilerinin ve kamu tarafından devredilen yetkilere sahip kuruluşların emretme yetkilerini uygulanması dolayısıyla yaptıkları harcamaları ifade etmektedir (Edizdoğan, 2008: 40-41). Kamu harcamaları genel olarak askeri, yatırım, transfer-sübvansiyon harcaması, eğitim, sağlık, gibi başlıklar altında toplanabilir.

Ekonomik büyüme, kişi başına düşen milli gelirin artışı, ülkenin uzun dönemde üretim ölçeğinin veya potansiyelinin genişlemesi veya daha üretken kullanılması sayesinde ortaya çıkması sonucu gayrisafi milli hâsıla da meydana gelen artışlarla tanımlanabilmektedir (Alkin, 1975: 220).

Kamu harcamaları ve ekonomik büyümenin ilişkisi yok sayılamaz. Kamu harcamalarının ekonomik büyümeye etkisi ekonomik yaklaşımlara göre farklılık göstermektedir.

Günümüzde her ekolden iktisatçılar gelişmiş ya da gelişmekte olan ekonomilerin kamu müdahalelerini ve devletin ekonomideki etkin rolünü kanıksamış bunun büyüme oranlarına yaptığı etkileri hesaplama yollarına başvurmuşlardır (Kar ve Taban, 2003: 147).

Wagner, ekonomilerin büyüme seviyelerinin artmasıyla beraber sosyoekonomik başlıklarının ilerleyeceğini ortaya koymuştur. Bu durum sosyal ve kültürel aktivitelerin artmasına sebep olarak kamudan hizmet yükseltecektir. Böylece kamu harcamalarının artış göstereceğini söylemektedir (Güder vd., 2016: 49). Wagner'e göre "kamu harcamalarındaki artış ekonomik büyümenin kaynağı değil sonucudur" şeklinde ifade edilmektedir (Henrekson, 1993: 407).

Keynes'e göre kamu harcamalarında meydana gelen bir artışın ekonomik büyümeye neden olacağı ortaya konulmuş kamu harcamalarının bu sayede milli gelir artışına olumlu yönde bir katkı sağlayacağı fikri vurgulanmıştır (Arısoy, 2005: 64). Ekonomik büyüme ve kamu harcamalarına ilişkin farklı görüşler nedeniyle birçok araştırmaya konu olmuştur.

4.2. Uygulamanın Amacı

Bu araştırmanın amacı istatistikte çok değişkenli bir analiz tekniği olan kümeleme analizi ve diskriminant analizinin kamu harcamaları ve ekonomik büyüme verileri ile bir

uygulamadan oluşmaktadır. Kümeleme analizi ve diskriminant analizi sınıflandırma için kullanılmaktadır. Diskriminant analizi ve kümeleme analizi ile sınıflandırılan OECD ülkelerinin gruplanmasına ilişkin sonuçlar tablolar yardımı ile yorumlanmıştır.

4.3. Literatür Taraması

Atılğan Yaşa ve Yüzbaşı Künc (2020), tarafından yapılan Kamu Harcamaları ve Sosyoekonomik Gelişmişlik Düzeyleri Bakımından İllerin Gruplandırılması: Türkiye Üzerine Bir Uygulama adlı çalışma da 2018 yılına ait verilerle Türkiye'deki illerin kamu harcamaları ve sosyo ekonomik açıdan gruplanmaları arasındaki farklılık ya da benzerliklerin ortaya çıkarılması amaçlanmıştır. Analiz olarak kümeleme analizi ve çok boyutlu ölçekleme analizi kullanılmıştır. Kümeleme analizinde k- ortalamalar tekniği ile analiz yapılmıştır. Kamu harcamaları ve sosyoekonomik gelişmişlik göstergeleri ele alındığında nüfus yoğunluğu fazla olan illerin kamu harcamaları bakımından ayrı bir grup oluşturduğu ve sosyal ekonomik açıdan gelişmemiş illerin kamu harcamalarından yüksek düzeyde faydalanan iller olduğu sonucu ortaya çıkmıştır.

Çetintürk ve Gençtürk (2020), tarafından yapılan OECD Ülkelerinin Sağlık Harcama Göstergelerinin Kümeleme Analizi ile Sınıflandırılması adlı çalışmada 2003-2017 yılları arasında 36 OECD ülkesinin sağlık harcaması verileri ile kümeleme analizi uygulaması yapılmıştır. Kümeleme analizi, hiyerarşik kümeleme yöntemi kullanılarak yapılmıştır. Analiz sonucunda OECD ülkeleri, sağlık harcama türlerine göre kümelenmiştir. Bu kümelemeye göre Türkiye'nin çeşitli sağlık harcama türlerinde en fazla benzerlik gösterdiği ülkeler belirlenmiştir. Diğer ülkeler ile kıyaslama yapıldığında Türkiye'nin gayri safi yurt içi hasıla içerisindeki sağlık harcamalarına ayırmış olduğu payın düşük bir seviyede olduğu gözlemlenmiştir.

Uysal ve Ersöz (2017), tarafından yapılan Türkiye'deki İllerin Yaşam Endeksinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi adlı çalışmasında Türkiye'de yaşam endeksi değerlerine göre farklılık ya da benzerlik gösteren illerin ortaya konulması için 81 ile ait 2015 yılı yaşam endeksleri ile kümeleme analizi değerlendirmiştir. Yöntemde Ward's yöntemi ve Kareli Öklid Uzaklığı ölçüsü kullanılmıştır. Kümeleme analizi sonuçlarının doğruluğuna bakmak için diskriminant analizi kullanılmıştır. İller, kümeleme analizi sonucunda üç kümeye ayrılmıştır. Diskriminant analizi sonucunda; analize dâhil edilen illerin hepsi doğru olarak sınıflandırılmıştır.

Ötken ve Süslü (2020), tarafından yapılan Kümeleme ve Ayırma Analizleri ile PISA 2012'ye Katılan Öğrencilerin Matematik Başarısı: Türkiye-Finlandiya Karşılaştırması adlı çalışmasında 2012 yılı Türkiye'nin uluslararası öğrenci değerlendirme programında (PISA) OECD ülkeleri ve aday ülkeler arasındaki yerinin, matematik hakkındaki özbenlik, özyeterlilik, motivasyon ve kaygı değişkenlerinin birlikte matematik başarısını sınıflama derecesi açısından incelemek için kümeleme ve diskriminant analizi yapmıştır. Kümeleme analizi sonuçlarına göre Türk öğrencilerinin matematik başarı puanlarının sınıflandırılması ile OECD ülkelerinin ortalamasının altında kaldığı sonucuna ulaşılmıştır. Diskriminant analizi sonucuna göre Türk öğrencilerinin özyeterlilik algılarının Finli öğrencilere göre daha fazla olduğu söylenebilir fakat özbenlik değişkeninin Türk öğrencileri tahminlenmesinde çok fazla etkili olmazken Finli öğrencilerde grup üyeliklerinin tahminlenmesinde etkili olduğu görülmüştür.

Burmaoğlu vd. (2009), tarafından yapılan Birleşmiş Milletler Kalkınma Programı Beşeri Kalkınma Endeksi Verilerini Kullanarak Diskriminant Analizi ve Lojistik Regresyon Analizinin Sınıflandırma Performanslarının Karşılaştırılması adlı çalışmasında 2007-2008 yılına ait Birleşmiş Milletler kalkınma programının beşeri kalkınma endeksi verileri ile diskriminant analizi ve lojistik regresyon analizi karşılaştırılmıştır. Diskriminant analizi sonucunda bağımsız değişkenlerin bağımlı değişkenleri açıklayabilme gücü %70,39' dur. Analizde kullanılan bağımsız değişkenler değerlendirilirken doğumda yaşam beklentisinin, gayri safi yurt içi hasılanın yüzdesi olarak ihraç edilen malların, kamu sağlık harcamalarının ve iletişim verilerinin ülkelerin gelişmişlik seviyesini yükseltmede pozitif etkisi olduğu tespit edilmiştir. Sınıflandırma sonuçlarına göre lojistik regresyon analizinde %100, diskriminant analizi sonucunda %92,5'lik bir başarı sağlanmıştır. İki analiz karşılaştırılırken varsayımları çok olan diskriminant analizine kıyasen varsayımları az olan lojistik regresyon analizini avantajlı hale getirmektedir.

Yalçın (2013), tarafından yapılan Kümeleme Analizi ve Uygulaması adlı çalışmasında Elazığ ilinde fabrikalarda çalışan işçilerin işitme kaybına ilişkin verileri kümeleme analizi ile sınıflandırılmıştır. Analiz uygulanırken Ward's kümeleme yöntemi kullanılmıştır. Öncelikle fabrikalar 5 kümeye ayrılmış gürültü bakımından birbirine en yakın iki fabrika aynı kümeye atanmıştır. Daha sonra fabrikalar 2 kümeye ayrılmış ve birbirine en yakın 4 fabrika aynı kümeye atanmıştır. Sonuç olarak; gürültülü iş yerlerinde çalışan işçilerin

gürültüye bağlı işitme kaybı ortamlardaki gürültü miktarı ve çalışma süresi ile ilişkili olmaktadır.

4.4. Veri Seti ve Yöntem

37 OECD (Ekonomik Kalkınma İşbirliği Örgütü) ülkesi için 2020 yılına ait yıllık veriler Dünya Bankası ve OECD resmi veri tabanından elde edilmiştir. Elde edilen değişkenler Tablo 4.1 Değişkenler tablosundaki gibidir:

Değişkenler	Kaynak
Askeri Harcama	OECD, https://data.oecd.org/
Eğitim Harcaması	
Sağlık Harcaması	
Sübvansiyon ve Transfer Harcaması	
Yatırım Harcaması	
Büyüme Hızı	
Toplam Harcama	

Kümeleme analizi yapılırken; hiyerarşik kümeleme yöntemlerinden Wards Yöntemi seçilerek kare öklid uzaklık ölçüsü ile uzaklıklar hesaplanmıştır. Diskriminant analizi uygulanırken Box M testi ve Wilks Lambda istatistiği kullanılmıştır. Sonuçlar, tablolar halinde ve dendrogram ile sunularak kamu harcamaları ve ekonomik büyüme verilerine göre ülkeler sınıflandırılmıştır. Bu analizler yapılırken çok değişkenli parametrik test varsayımlarının sağlandığı bilinmektedir.

4.5. Hiyerarşik Kümeleme Yöntemi Uygulama

4.5.1. Bulgular

Verilere ait durum özetini gösteren tablo aşağıda ifade edildiği gibidir;

Durum					
Geçerli veri		Eksik veri		Toplam	
N	Yüzde	N	Yüzde	N	Yüzde
37	100,0	0	0	37	100,0

Analiz çıktısına göre; 37 gözlem ile analiz yapılmıştır ve eksik veri olmadığı belirlenmiştir. Analiz kare öklid uzaklık ölçüsü ile Ward's yöntemi kullanılarak yapılmıştır.

2020 yılına ait kamunun yaptığı toplam harcamalar ve bu harcamalara ait alt birimleri olan eğitim harcamaları, sağlık harcamaları, askeri harcamalar, yatırım harcamaları,

sübvansiyon ve transfer harcamaları bunlarla birlikte büyüme hızı değişkenleri kullanılarak yapılan hiyerarşik kümeleme analizinin sonuçları Tablo 4.3'te ifade edildiği gibidir;

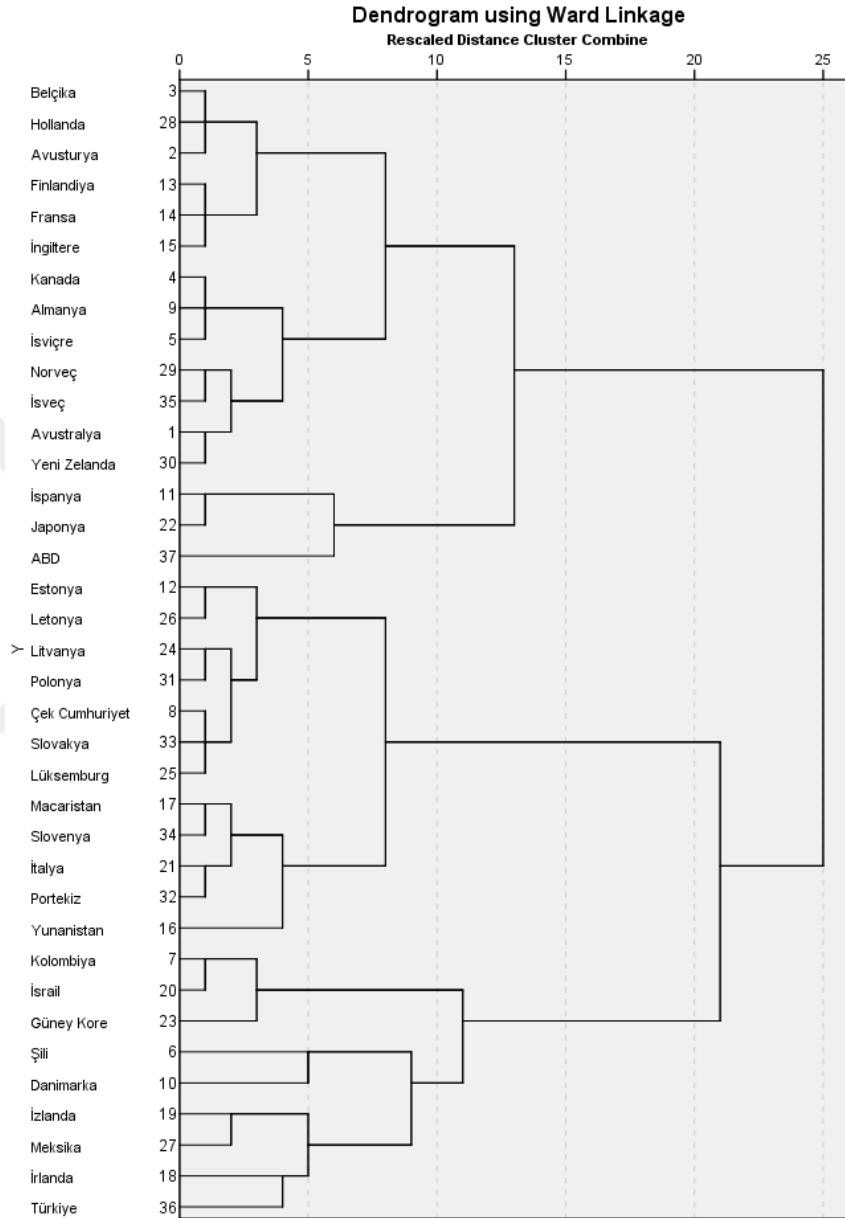
Tablo 4.3. Hiyerarşik Kümeleme Analizi Sonuçları

Aşama	Küme Birleştirme		Katsayı	İlk Görünen Küme Aşaması		Sonraki Aşama
	Küme 1	Küme 2		Küme 1	Küme 2	
1	3	28	0.242	0	0	4
2	8	33	0.686	0	0	12
3	17	34	1.298	0	0	17
4	2	3	2.013	0	1	23
5	4	9	2.854	0	0	13
6	29	35	3.700	0	0	20
7	11	22	4.560	0	0	29
8	21	32	5.510	0	0	17
9	13	14	6.527	0	0	15
10	24	31	7.811	0	0	19
11	1	30	9.236	0	0	20
12	8	25	10.775	2	0	19
13	4	5	12.359	5	0	25
14	12	26	13.958	0	0	21
15	13	15	15.566	9	0	23
16	7	20	17.392	0	0	22
17	17	21	19.623	3	8	26
18	19	27	22.183	0	0	28
19	8	24	24.778	12	10	21
20	1	29	27.614	11	6	25
21	8	12	32.043	19	14	30
22	7	23	36.619	16	0	33
23	2	13	41.255	4	15	31
24	18	36	46.915	0	0	28
25	1	4	52.797	20	13	31
26	16	17	60.113	0	17	30
27	6	10	68.039	0	0	32
28	18	19	76.923	24	18	32
29	11	37	87.437	7	0	34
30	8	16	100.098	21	26	35
31	1	2	113.306	25	23	34
32	6	18	129.353	27	28	33
33	6	7	148.992	32	22	35
34	1	11	171.516	31	29	36
35	6	8	207.498	33	30	36
36	1	6	252.000	34	35	0

Analiz sonucunda elde edilen bulgular değerlendirildiğinde, ilk olarak birinci aşamada birleşen ülkeler Belçika ve Hollanda'dır. Bu ülkeleri sırasıyla Avusturya-Finlandiya,

Avustralya-İspanya-Şili takip etmektedir. Benzerliği yüksek olan kümelerin atamaları bu şekilde devam etmektedir. Ülkelerin kümelenme aşamalarını aynı zamanda ağaç diyagramı ile incelenebilir.

Şekil 4.1. Ağaç Grafiği (Dendrogram)



Ağaç diyagramı (Dendrogram) sayesinde uzaklıklara göre oluşturulan kümelerin aşamalarını görebiliriz. Dendrogram incelendiğinde, Belçika, Hollanda, Avusturya, Finlandiya, Fransa, İngiltere, Kanada, Almanya, İsviçre, Norveç, İsveç, Avustralya, Yeni Zelanda, İspanya, Japonya ve ABD'nin bir küme oluşturduğu; Macaristan, Slovenya, İtalya, Portekiz, Yunanistan, Kolombiya, İsrail, Güney Kore, Şili, Danimarka, İzlanda, Meksika, İrlanda ve Türkiye'nin ise ikinci bir küme oluşturduğu görülmektedir.

4.6. Diskriminant Analizi ile Uygulama

4.6.1. Bulgular

Diskriminant analizi varsayımından varyans-kovaryans matrislerinin eşitliğine bakmak için Box M testi sonuçları aşağıdaki Tablo 4.4'te ifade edildiği gibidir;

Tablo 4.4. Box M Testi Tablosu

Box's M		91.446
F	Yaklaşık	1.982
	df1	28
	df2	970.879
	Anlamlılık	0.002

Güvenilir bir analiz gerçekleştirmek için değişkenlerin varyans ve kovaryanslarının eşitliği varsayımının sağlanması gerektiğine diskriminant analizi anlatılırken değinmiştik. Varyans-kovaryans matrisinin eşitliğinin test edilmesi için Box M testi kullanılmıştır. Box M değerine ilişkin olasılık değeri $p > 0.05$ 'den büyük olması gerekir. Analiz sonucuna göre 0.002 değeri 0.05'den küçük değer elde edildiği için varyans-kovaryans matrislerinin eşit olmadığı görülmektedir. Bu durumda log-determinant incelenmeli ya da kovaryans matris kullanımında Within-Groups yerine Separate-Groups seçilerek analize devam edilir. Separate-Groups seçildikten sonra Box's M için anlamlılık, 0.080 olarak elde edilmiştir.

Diskriminant fonksiyonlarının önemliliğini görebilmek için Wilks Lambda, öz değer, kanonik korelasyon değerine bakılır. Özdeğer istatistiği aşağıda Tablo 4.5'te ifade edildiği gibidir;

Tablo 4.5. Özdeğer İstatistiği Tablosu

Fonksiyon	Öz Değer	Açıklanan Varyans (%)	Birikimli Açıklanan Varyans (%)	Kanonik Korelasyon
1	3,422	100,0	100,0	0,880

Analizde Toplam varyansı açıklama oranı Tablo 4.5'te verilmiştir. Bu tabloya göre bağımlı değişken iki kategorili olduğu için 1 tane fonksiyon yer almaktadır. Toplam varyansın %100'ünü tek fonksiyonda açıklamıştır.

Kanonik korelasyon katsayıları sütunu incelendiğinde de fonksiyonun gruplarla olan ilişkisi diskriminant fonksiyonları gruplar arası farklılıkları açıklamada önem derecelerine göre sıralanmıştır. Kanonik korelasyon gruplar arasındaki ilişkiyi ölçer. Kanonik korelasyon değeri (0,880)'dir. Bu değerın karesi alınarak yorumlaması yapılır. $(0,880^2=0,774)$ Bu sonuca göre bağımlı değişkendeki varyansın %77'sini açıklayabilmektedir.

Öz değer istatistiğinin büyüklüğüne göre bağımlı değişkendeki varyansın o fonksiyon tarafından açıklanacağı büyüklüğü gösterir. Genelde 0,40'dan büyük değerler özdeğerler iyi olarak kabul edilir. Tabloda görüldüğü gibi öz değer istatistiği 3,422'dir. Bu değer fonksiyonumuzun iyi bir ayırmacılık sağladığını gösterir.

Wilks Lambda istatistiği aşağıda Tablo 4.6'da ifade edildiği gibidir;

Tablo 4.6. Wilks Lambda İstatistiği Tablosu

Test Edilen		Serbestlik		
Fonksiyon	Wilks' Lambda	Ki-Kare	Derecesi	Anlamlılık
1	0,226	46,826	7	0,000

Wilks Lambda istatistiği 0-1 arasındaki değerlere sahiptir. Tablo 4.6'da Wilks Lambda istatistiği, varyans toplamalarının gruplar arasındaki farklar açısından açıklanmayan oranını gösterir. Tabloya göre Wilks Lambda değeri 0.226 bulunmuş ve fonksiyonumuz toplam varyansın %22'sini açıklayamamaktadır.

Ayırma fonksiyon katsayısı aşağıda Tablo 4.7'de ifade edildiği gibidir;

Tablo 4.7. Ayırma Fonksiyon Katsayısı

	Fonksiyon 1
Askeri	-0.358
Yatırım	-0.297
Büyüme	-0.113
Transfer	0.725
Eğitim	0.312
Harcama	-0.054
Sağlık	0.893

Diskriminant fonksiyonu katsayılarına ve yapı matrisine bakarak bağımsız değişkenlerin ne kadar önemli olduğunu değerlendirmek mümkün olur. Tablo 4.7'de görüldüğü gibi ülkelerin gruplara ayırmada sağlık harcaması ve transfer sübvansiyon harcaması önemli ayırt edici bağımsız değişkenlerdir. Sağlık harcaması 0.893 katsayıya sahipken, transfer sübvansiyon harcaması 0.725 katsayıya ve eğitim harcaması ise 0.312 katsayıya sahiptir. Diğer değişkenlerin ayırmada etkili bir değişken olmadığını söylemek mümkündür.

Yapı matrisi aşağıda Tablo 4.8'de ifade edildiği gibidir;

Tablo 4.8. Yapı Matrisi Tablosu

	Fonksiyon 1
Sağlık Harcaması	0,588
Transfer Sübvansiyon Harcaması	0,497
Ekonomik Büyüme	-0,151
Yatırım Harcaması	-0,134
Toplam Harcamalar	-0,127
Askeri Harcama	-0,126
Eğitim Harcaması	0,017

Yapı matrisi, 6 tane değişkenin ayırma fonksiyonu ile arasındaki korelasyonu gösterir. Tablo 4.8'e göre en yüksek korelasyona sahip değişken sağlık harcaması ve bunu takip eden transfer sübvansiyon harcaması değişkenidir. Ekonomik büyüme, yatırım harcaması, toplam harcamalar ve askeri harcama değişkenleri önemli bir tahmin edici değildir.

Kanonik diskriminant katsayısına ait Tablo 4.9 aşağıda ifade edildiği gibidir;

Tablo 4.9. Kanonik Diskriminant Katsayıları Tablosu

	Fonksiyon 1
Askeri Harcama	-0.116
Yatırım Harcaması	-0.084
Ekonomik Büyüme	-0.075
Transfer Sübvansiyon Harcaması	0.065
Eğitim Harcaması	0.105
Toplam Harcamalar	-0.006
Sağlık Harcaması	0.502
Sabit	-8.503

Diskriminant fonksiyonuna etki eden katsayı değerleri, Tablo 4.9'daki gibidir. Bu fonksiyon $D = -8,503 - 0,116(\text{Askeri Harcama}) - 0,084(\text{Yatırım Harcaması}) - 0,075(\text{Ekonomik Büyüme}) + 0,065(\text{Transfer Sübvansiyon Harcaması}) + 0,1055(\text{Eğitim Harcaması}) - 0,006(\text{Toplam Harcamalar}) + 0,502(\text{Sağlık Harcaması})$ şeklinde ifade edilir.

Gruplara ait ortalama ayırma fonksiyon değerleri aşağıda Tablo 4.10'da ifade edildiği gibidir;

Tablo 4.10. Grupların Ortalama Ayırma Fonksiyon Değerleri Tablosu

Ward Method	Fonksiyon 1
1	2,061
2	-1,570

Tablo 4.10’da iki gruba ait ortalama ayırma fonksiyon skorları gösterilmiştir. Birinci gruba ait ortalama değeri 2,061 ve ikinci gruba ait ortalama değeri -1,570’dir.

Diskriminant analizi sınıflandırma sonuçları aşağıda Tablo 4.11’de ifade edildiği gibidir;

Tablo 4.11. Diskriminant Analizi Sınıflandırma Sonuçları Tablosu

	Ward Method	Tahmin Edilen Grup Değerleri		Toplam
		1	2	
Sayı	1	16	0	16
	2	0	21	21
%	1	100,0	0,0	100,0
	2	0,0	100,0	100,0

Ayırma analizinin başarısını test etmek için doğru sınıflandırma yüzdesine bakılır. Doğru sınıflandırma yüzdesinin büyüklüğü analizde sınıflandırmanın başarısının büyüklüğünü gösterir. Birinci grupta olan 16 ülke doğru sınıflandırılmıştır, bu sayılar %100 doğru sınıflama yapıldığını gösterir. İkinci grupta ise 21 ülke doğru sınıflandırılmıştır, bu sayılar bize %100 oranında doğru sınıflama yapıldığını gösterir. Sonuç olarak 37 ülkenin tamamı doğru sınıflandırılmıştır.

5. SONUÇ

Diskriminant analizi ve kümeleme analizi verileri, gruplara ayırmayı amaçlayan çok değişkenli analizlerdir. Kümeleme analizinde çok değişkenli parametrik varsayımlara fazla gerek duyulmazken diskriminant analizinde çok değişkenli parametrik varsayımların sağlanması gerekir.

Kümeleme analizi sınıflandırmayı yapabilmek için uzaklık ve benzerlik gibi ölçütlerden faydalanırken diskriminant analizi diskriminant fonksiyonları oluşturarak ilerleme sağlar. Diskriminant analizi, veri grubuna daha sonradan eklenen gözlemleri analize dahil edebilir fakat kümeleme analizinde daha sonra eklenen gözlemin atanacağı kümenin belirlenmesi mümkün değildir.

2020 yılı OECD ülkelerine ait kamu harcamaları ve ekonomik büyüme verileri incelenmiştir. Kümeleme analizi sonuçlarına bakıldığında 37 ülkenin kümelenebileceği ile oluşturulan iki grupta; birinci küme 16 ülkeden, ikinci küme ise 21 ülkeden oluşmaktadır. Diskriminant analizi sonuçlarına bakıldığında ise her iki kümede yer alan ülkelerin tamamının doğru sınıflandığı görülmüştür.

Kümeleme ve diskriminant analizi sonucunda; Belçika, Hollanda, Avusturya, Finlandiya, Fransa, İngiltere, Kanada, Almanya, İsviçre, Norveç, İsveç, Avustralya, Yeni Zelanda, İspanya, Japonya ve ABD'nin birinci kümeyi oluşturduğu; Macaristan, Slovenya, İtalya, Portekiz, Yunanistan, Kolombiya, İsrail, Güney Kore, Şili, Danimarka, İzlanda, Meksika, İrlanda ve Türkiye'nin ise ikinci kümeyi oluşturduğu görülmektedir.

Ülkeler açısından verimli bir büyüme için gerekli ekonomik büyüme hızı oranının %2-%3 arasında olması beklenir. Buna göre genel itibari ile ekonomik büyüme düzeyi, bu oranlar arasında olan ülkeler aynı grupta kümelenebilir. Kamu harcamaları bakımından transfer sübvansiyon harcamaları, toplam harcamalar ve sağlık harcamaları diğer ülkelere göre yüksek olanlar aynı grupta kümelenebilir. Ekonomik büyümenin normal oranı ile kamuya yapılan harcamalar arasında pozitif yönde bir ilişki olduğu gözlemlenmiştir.

KAYNAKÇA

- Aksoy, Ş., (1994), “*Kamu Maliyesi*”, (2. Baskı), Filiz Kitapevi, İstanbul.
- Alkin, E., (1975), “*Gelir ve Büyüme Teorisi*”, (1. Baskı), İstanbul Üniversitesi Yayınları, İstanbul.
- Alpar, R., (2011), “*Çok Değişkenli İstatistiksel Yöntemler*”, (3.Baskı), Detay Yayıncılık, Ankara.
- Arısoy, İ. (2005), “Wagner ve Keynes Hipotezleri Çerçevesinde Türkiye’de Kamu Harcamaları ve Ekonomik Büyüme İlişkisi”, *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2(14), 63-80.
- Atılğan Yaşa, A. ve Yüzbaşı Küncü, G. (2020), “Kamu Harcamaları ve Sosyoekonomik Gelişmişlik Düzeyleri Bakımından İllerin Gruplandırılması: Türkiye Üzerine Bir Uygulama”, *Akademik Araştırmalar ve Çalışmalar Dergisi*, 12(22), 276-297.
- Başarır, G., (1990), “Çok Değişkenli Verilerde Ayrımsama Sorunu ve Lojistik Regresyon Analizi”, (*Yayınlanmış Doktora Tezi*), Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara (Türkiye).
- Çetintürk, İ. ve Gençtürk, M. (2020), “OECD Ülkelerinin Sağlık Harcama Göstergelerinin Kümeleme Analizi ile Sınıflandırılması”, *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, 11(26), 228-244.
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2010), *Sosyal Bilimler için Çok Değişkenli İstatistik Teknikleri*, (1. Baskı), Pegem Akademi, Ankara.
- Edizdoğan, N. (2008), *Kamu Bütçesi*, (7. Baskı), Ekin Basın Yayın Dağıtım. Bursa.
- Güder, F., Yücekaya, P., Şenyurt, A., “Kamu Harcamaları ile Ekonomik Büyüme İlişkisi: Türkiye İçin Keynesyen Görüş Mü? Wagner Kanunu Mu Geçerli? (2006- 2015 Dönemi)”, *Usb Dergi*, 1, 47-60.
- Hair, J.F., Anderson, R. E., Tatham, R. L. ve Black, W. C. (1995), “*Multivariate Data Analysis with Readings*”. Prentice-Hall.
- Henrekson, M. (1993), “*The Peacock – Wiseman Hypothesis, (Ed. Norman Gemmill) The Growth Of The Public Sector:Theories And International Evidence*”, England: Edward Elgar Publishing.
- Johnson, R. A. ve Wichern, D. W. (1988), “*Applied Multivariate Statistical Analysis*”, Prentice Hall New Jers.

- Kalaycı, Ş. (2010), “*SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*” (5.baskı), Asil Yayınları, Ankara
- Kar, M. ve Taban, S. (2003). “Kamu Harcama Çeşitlerinin Ekonomik Büyüme Etkisi”, *Ankara Üniversitesi SBF Dergisi*.
- Ötken, Ş. ve Süslü, A. (2020) “Kümeleme ve Ayırma Analizleri ile Pısa 2012’ye Katılan Öğrencilerin Matematik Başarısı: Türkiye-Finlandiya Karşılaştırması”, *Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi*, 7(7), 106-117.
- Özdamar, K., (2002), “*Paket Programlarla İstatistiksel Veri Analizi 2*”, (10. baskı), Kaan Kitabevi, Eskişehir.
- Servi, T. ve Erişoğlu Ü. (2020), “Türkiye’deki Şehirlerin Sosyo- Ekonomik Gelişmişlik Düzeylerinin İstatistiksel Analizi”, *Al-Farabi International Journal on Social Sciences*, 5(2), 174- 186.
- Şahin, G. (2017), “Sağlamlık Özelliğine Dayalı Tahmin Edicilerle Diskriminant Analizi”, (Yayımlanmamış Yüksek Lisans Tezi), Gazi Üniversitesi, İstatistik Anabilim Dalı, Ankara (Türkiye).
- Tatlıdil, H. (1996), “*Uygulamalı Çok Değişkenli İstatistiksel Analiz*”, Engin Yayınları, Ankara.
- Uysal, F. N., Ersöz, T. ve Ersöz, F. (2017), Türkiye’deki İllerin Yaşam Endeksinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi, *Ekonomi Bilimleri Dergisi*, 9(1), 49-65.
- Yalçın, N., (2013), Kümeleme Analizi ve Uygulaması, (Yayımlanmamış Yüksek Lisans Tezi), Fırat üniversitesi, İstatistik Anabilim Dalı, Elazığ.