

AYIRMA ANALİZİ VE BİR UYGULAMA

Mert ÜNAL

**YÜKSEK LİSANS TEZİ
İSTATİSTİK**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EKİM 2006

ANKARA

AYIRMA ANALİZİ VE BİR UYGULAMA

Mert ÜNAL

**YÜKSEK LİSANS TEZİ
İSTATİSTİK**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EKİM 2006

ANKARA

Mert ÜNAL tarafından hazırlanan AYIRMA ANALİZİ VE BİR UYGULAMA adlı bu tezin Yüksek lisans tezi olarak uygun olduğunu onaylarım.

Yrd .Doç .Dr. Selahattin ERGEÇ
Tez Yöneticisi

Bu çalışma, jürimiz tarafından oy birliği / oy çokluğu ile İstatistik Anabilim Dalında yüksek Lisans tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Fevzi KUTAY

Üye :Prof. Dr. Müslim EKNİ

Üye : Prof. Dr. Semra ERBAŞ

Üye : Prof. Dr. Aydın ÜNSAL

Üye : Yrd .Doç .Dr. Selahattin ERGEÇ

Tarih: 30/10/2006

Bu tez, Gazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Mert ÜNAL

AYIRMA ANALİZİ VE BİR UYGULAMA**(Yüksek Lisans Tezi)****Mert ÜNAL****GAZİ ÜNİVERSİTESİ****FEN BİLİMLERİ ENSTİTÜSÜ****Ekim 2006****ÖZET**

Belediyelerde su üretimi, dağıtımı ve tüketimi, Türkiye Enerji Sektöründe çok önemli bir yer teşkil etmektedir.

Bu çalışmada Türkiye İstatistik Kurumunun 2001 yılı Büyükşehir Belediyeleri Su İstatistiklerine ayırma analizi uygulanarak, Büyükşehir belediyeleri yeniden sınıflandırılmıştır. Oluşan yeni gruplara diğer 65 belediyenin yerleştirilmesi amaçlanmıştır.

Bilim Kodu : 205.1.009

Anahtar Kelimeler : Ayırma Analizi, Ayırma

Fonksiyonları, Büyükşehir Belediyeleri, Su İstatistikleri

Sayfa Adedi : 51

Tez Yöneticisi : Öğr. Gör. Yrd. Doç. Dr. Selahattin ERGEÇ

DISCRIMINANT ANALYSIS AND AN APPLICATION**(M.Sc. Thesis)****Mert ÜNAL****GAZI UNIVERSITY****INSTITUTE OF SCIENCE AND TECHNOLOGY****October 2006****ABSTRACT**

Water production, distribution and consumption in the municipalities constitutes a most important place in the Turkey's energy sector.

In this study, wen municipalities have been reclassified after the discriminant analysis was applied to the water statistics of the year 2001 taken from TURKSTAT (TURKISH STATISTICAL INSTITUTE). The rest of the municipalities totally 65, are presumed to be placed into the new groups that appeared after the classification process.

Science Code : 205.1.009

Key Words : Discriminant Analysis, Discriminant Functions, Wen Municipalities, Water Statistics

Page Number : 51

Adviser : Instructor Yrd. Doc. Dr. Selahattin ERGEÇ

TEŐEKKÜR

Çalıőmalarım boyunca deęerli yardım ve katkılarıyla beni yönlendiren Hocam Öğr. Gör. Dr. Selahattin ERGEÇ'e, manevi desteęiyle beni hiçbir zaman yalnız bırakmayan annem Yıldız YÜKSEL'e ve kardeőim Murat ÜNAL'a teőekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER.....	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
1. GİRİŞ.....	1
2. AYIRMA ANALİZİ.....	3
2.1. Tanım ve Tarihçe.....	3
2.2. Ayırma Analizinin Amaçları.....	6
2.3. Ayırma Analizinin Varsayımları.....	6
2.4. Veri Yapısı.....	7
2.5. Ayırma Analizinin Geometrik Yorumu.....	8
2.6. Ayırma Analizinin Temel Problemi.....	9
2.6.1. Betimsel amaçlı ayırma analizi.....	10
2.6.2. Karar amaçlı ayırma analizi.....	19
3. AYIRMA ANALİZİNDE UYGULANAN TESTLER.....	23
3.1. Çok Değişkenli Normalliğin Testi.....	23
3.2. Grup Kovaryans Matrislerinin Eşitliği Testi.....	24
3.3. Grup Ortalama Vektörlerinin Farklılığı Testi.....	25
3.4. Ayırma Fonksiyonlarının Anlamlılık Testi.....	26

	Sayfa
4.UYGULAMA.....	29
4.1. Araştırmanın Kapsamı.....	30
4.2. Analize Alınan Değişkenler ve Bu Değişkenlerin Tanımları.....	30
4.3. Araştırmaya Alınan Belediyelerin Analiz Öncesi Gruplanması	32
4.4. Analiz.....	33
4.4.1. Grup istatistikleri.....	33
4.4.2. Ayırma fonksiyonları ve anlamlılıkları.....	33
4.4.3. Belediyelerin sınıflandırılması.....	37
4.4.4. Örnek seçilen belediyenin oluşturulan gruplara atanması.....	37
5.SONUÇ.....	40
KAYNAKLAR.....	41
EKLER.....	43
EK-1. DİE 2001 yılı Büyükşehir belediyelerinin toplanan ve dağıtılan su miktarı, abone sayıları, gelirleri, çalışan sayıları, çalışanlarına ödediği ücretler, giderleri, işçisat toplamları, satın aldıkları elektriğin ücretleri, verimleri....	44
EK-2. Grup İstatistikleri.....	47
EK-3. Belediyelerin analiz öncesi ve analiz sonrası yer aldığı gruplar ile ayrıcı faktör değerleri.....	48
EK-4. Balıkesir belediyesinin 2001 yılı su istatistikleri.....	49
EK-5 Mahalanobis Uzaklık Değerleri.....	50
ÖZGEÇMİŞ.....	51

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. G - Grup İçin Ayırma Analizi Verisi.....	7
Çizelge 4.1. VERİM değişkeninin tolerans değeri.....	33
Çizelge 4.2. Ayırma fonksiyonları, özdeğerler, açıklanan varyans payları, kanonik ilişkiler.....	33
Çizelge 4.3. Ayırma fonksiyonlarının test istatistikleri.....	33
Çizelge 4.4. Standartlaştırılmış ayırma fonksiyonlarının katsayıları.....	34
Çizelge 4.5. Faktör yapı matrisi.....	35
Çizelge 4.6. Grup ortalamalarının ayırıcı faktör değerleri.....	35
Çizelge 4.7. Sınıflandırma sonuçları.....	36
Çizelge 4.8. Fisher'in ayırma fonksiyonları.....	38

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Ayırma analizinin geometrik yorumu.....	9
Şekil 4.1. İki Ayırıcı Fonksiyon üzerinde 3 grubun görünümü.....	36

1. GİRİŞ

Çok deęişkenli analiz, istatistięin uygulamalarda kullanılan önemli bir koludur. Bu analizde birbirleriyle iliřkili çok sayıda deęişken söz konusudur. Kullanılan birçok teknikle, çok sayıda deęişkenin oluşturduęu sistemin yapısı belirlenir ve olabildięince basit bir forma dönüřtürülerek herhangi bir konuda doęru karar için gerekli bilgi çıkartılır.

Çok deęişkenli analizde deney birimlerinden gözlem ya da ölçüm yoluyla elde edilen özellikler göz önüne alınır. Deęişken adı verilen bu özelliklerin çok sayıda olması, sorunun klasik istatistik teknikleri ile çözümüne olanak tanımamaktadır. Bu nedenle, 1940'lı yıllardan başlayarak günümüze kadar birçok teknik geliştirilerek sorunun çözümüne çalışılmıştır.

“Çok Deęişkenli Analiz Teknikleri” adı verilen bu tekniklerin asıl amacı; istatistięin öteki kollarında olduęu gibi, bilimsel çalışmaların sayı ile ifade edilebilen sonuçlarının özetlenmesi, yorumlanması ve karar verilirken kullanılmasının sağlanmasıdır. Bilimsel çalışmalarda ele alınan olaylar genellikle pek çok etkenin etkisi altındadır. Ayrıca, gözleme konu olan nesnelere özelliklerini de birbiriyle ilişkilidir. Bu nedenle, uygulamada çok sayıda deęişkenle karşılaşılmaktadır. Yapılan çalışmaların geçerli ve güvenilir sonuçlar verebilmesi için, inceleme konusu olayları bütün yönleriyle deęerlendirmek bir zorunluluktur. Bu zorunluluk sonucu arařtırmacı çok deęişkenli veri ve bunların analizi ile karşı karşıya kalır.

Son yıllarda yapılan bilimsel arařtırmalarda incelenen olayların analizinde, kısıtlayıcı varsayımlar altında geçerli olan tek deęişkenli analizlerin yeterli olmadığı görülmektedir. Tek deęişkenli analizlerle ilgili en önemli kısıt, olaydaki birçok faktörün deneysel olarak kontrol altında tutulması ve her defasında tek bir faktörün etkisinin incelenmesidir.

Çok deęişkenli analizler, birden çok özellięin analizi ile ilgilendięinden uygulamalarda deęişik amaçlarla kullanılmaktadır. Bu amaçlardan önemli olan

birkaç tanesi; basitleştirme ve boyut indirme, birimlerin sınıflandırılması, bağımlılık yapısının incelenmesi, hipotez testleri ve hipotez oluşturma, sıralama ve ölçekleme şeklinde ifade edilebilir [1].

Tıp, biyoloji, tarım ve eğitim başta olmak üzere birçok uygulama alanı bulunan çok değişkenli analiz yöntemlerinden biri de ayırma analizidir. Ayırma analizi gruplar arası farklılığın kökenini araştırmaya çalışan ve hangi yığından geldiği bilinmeyen bir birimin eldeki gruplardan birine sınıflandırılması sağlayan birçok değişkenli analiz türüdür.

Çalışmanın birinci bölümünde, ayırma analizinin tanımına, tarihçesine, amaçlarına, varsayımlarına, veri yapısına, geometrik yorumuna, betimsel amaçlı ayırma analizinin temel problemine, ayırma fonksiyonlarına ve karar amaçlı ayırma analizine yer verilmiştir.

Çalışmanın ikinci bölümünde, ayırma analizinde uygulanan testlere yer verilmiştir.

Üçüncü bölümde, belediyelerde su istatistikleri göz önüne alınarak 16 Büyükşehir Belediyesi yeniden sınıflandırılmıştır.

2. AYIRMA ANALİZİ

2.1. Tanım ve Tarihçe

Çok değişkenli analizde sıklıkla karşılaşılan problemlerden birisi sınıflandırma problemi. Araştırmacının ilgilendiği bireyler farklı yığınlardan (gruplardan) geliyor olabilir. Araştırmacı, bir bireyin p sayıda özelliğini ölçtüğünde, elindeki bireyin hangi gruptan geldiğini merak edebilir. Bu durumda sınıflandırma problemi, bireyin p sayıda özelliğini inceleyerek hangi gruptan (grup sayısı g olmak üzere) geldiğine karar verme problemi olarak nitelendirilebilir.

Anderson'un belirttiği gibi bu grupların p değişkenli olasılık dağılımlarına sahip oldukları varsayılır. Bu durumda herhangi bir bireyin bu gruplardan gelen bir rassal örnek olduğu söylenebilir. İşte sınıflandırma problemindeki temel soru; “ p tane değişkene ilişkin gözlem değerleri bilinen bireyin hangi olasılık dağılımından geldiği”dir.

Bu açıdan değerlendirildiğinde sınıflandırma problemi, bir istatistiksel karar verme sürecidir. Bu süreçte araştırmacı; bireyin hangi gruptan geldiğine karar vermelidir. Bazı durumlarda grupların olasılık dağılımları ve bu dağılımların parametreleri bilinmektedir. Ancak uygulamada genellikle her bir grubun p değişkene ilişkin bir dağılıma sahip olduğu varsayılır ve bu dağılımın parametreleri seçilen örnek aracılığıyla tahmin edilir. Ardından karar verme problemi çözülmeye çalışılır [1].

Bireyin hangi gruptan geldiğini tespit etmeye çalışan araştırmacı, p değişkeni kullanarak, bireyi uygun bir gruba atar. Bu aşamada her bir değişkenin atama kararında etkisi olduğu söylenebilir. Araştırmacının bir diğer amacı da, bireyleri sınıflandırmada hangi değişken ya da değişkenlerin daha etkili olduğunu belirlemek olabilir. Böylece bireylerin farklı gruplarda yer almalarına neden olan değişkenler tespit edilebilir. Hem sınıflandırma hem de grup ayırımına etki eden değişkenleri belirlerken p tane değişkenin fonksiyonu olan ayırma fonksiyon(lar)ı tanımlanır. Bu

fonksiyon(lar) aracılığıyla bireylerin sınıflandırılması ya da ayırma etki eden değişkenlerin saptanması mümkündür.

Bu açıklamalar doğrultusunda Ayırma Analizi'nin kısaca tanımını şu şekilde yapabiliriz; Ayırma analizi, birimlerin çok sayıdaki özelliğini dikkate alarak, bu özelliklere göre birimlerin doğal ortamdaki gerçek sınıflarına optimal düzeyde atanmaları amacıyla uygulanır [2]. Dolayısıyla ayırma analizi, grupların birbirlerinden en iyi şekilde ayırımını sağlar. Ayırma analizini kullananlar, açıklama ve tahminle ilgilenirler. Bu gruplarla en fazla ilişkili olan değişkenlerin hangileri olduğunu ve bunların grup üyeliğini ne kadar iyi tahmin edebildiğini belirlemek isterler [3].

Ayırma analizinin uygulama adımlarını ise aşağıdaki gibi sıralayabiliriz:

- Önsel grup üyelikleri belirlenir.
- Analize alınan gruplar arasında fark olup olmadığı, Bartlett'in Ki-kare test istatistiği ile test edilir. Test sonucu gruplar arasında anlamlı bir fark varsa, analize devam edilir.
- Kullanılacak değişkenler seçilir. Değişken seçiminde önsel bilgi ya da istatistikî yöntemler uygulanabilir.
- Değişkenler arasında çoklu bağlantının olup olmadığı incelenir. Bu amaçla birleştirilmiş grup içi korelasyon matrisi incelenir. Bu matristeki korelasyon değerleri mutlak değerce %75'den büyük ise değişkenlerden bir kısmının atılması gerekir. Bu adımın sonunda değişken kümesi belirlenmiş olur.
- $W^{-1}B$ matrisinin özdeğerleri ve bu özdeğerlere ilişkin özvektörler bulunur. Bu özvektörler, ayırma fonksiyonları için gerekli ağırlıkları verir. Ayırma fonksiyonlarının anlamlılık testi de bu özdeğerler kullanılarak yapılır. Eğer herhangi bir fonksiyon anlamlı ise yaptığı ayırımın başarılı olduğu söylenebilir.

- Standartlaştırılmamış ayırma fonksiyonu kullanılarak her bir birey için ayırma fonksiyonu değerleri elde edilir. Bu değerler sınıflandırma aşamasında kullanılacaktır.
- Grup üyelikleri için önsel olasılıklar belirlenir. Daha sonra bu olasılıklar ve ayırma skorları kullanılarak sonsal olasılıklar elde edilir. Bireyin sahip olduğu en büyük sonsal olasılık tespit edilir. Bu olasılığı veren grubun o bireyin ait olduğu grup olduğu tahmin edilir ve birey sınıflandırılmış olur.
- Her bir birey sınıflandırıldıktan sonra, ayırma fonksiyonunun başarısı, doğru sınıflandırma yüzdesi incelenerek tespit edilebilir [2].

Ayrırma analizinin geçmişine bakıldığında ayırma analizi ile ilgili çalışmaların 1930'lu yıllarda aynı problemi farklı perspektiften farklı araştırmacılar incelemiştir. Bunlar:

Fisher 1936 yılında, ayırma analizinde en çok kullanılan ölçüt olan ayırma fonksiyonunu ortaya koymuştur. Bu ölçüt, uzaklık ölçüleri, yanlış sınıflandırma olasılığı ve yanlış sınıflandırma maliyetini kapsayan optimal bir ayırma fonksiyonudur.

Roy 1939 yılında, p değişkenli iki normal dağılım için kovaryans matrislerinin eşitliğini test etmek için çalışmalar yapmıştır.

Welch 1939 yılında, yanlış sınıflandırma olasılığının ve yanlış sınıflandırma maliyetlerinin minimum yapıldığında en iyi ayırmanın elde edilebileceğini belirtmiştir.

Rao 1948 yılında, ikiden çok grubun olduğu durumlarda bireyleri minimum hata ile sınıflandırma yöntemi üzerinde çalışmıştır.

Anderson ve Bahadur 1962 yılında farklı ortalamalı ve farklı varyans kovaryans matrisli çok deęişkenli normal yığınını gruplara ayrılması üzerinde çalışmışlardır. Ayrıca Hotelling ve Mahalanobis'de benzer çalışmalarda bulunmuşlardır [5].

2.2. Ayırma Analizinin Amaçları

Ayırma analizinin amaçları, aşağıdakilerden biri veya tamamı olabilir:

1. Analiz öncesi tanımlanmış iki ya da daha fazla grubun ortalama özellikleri arasında önemli farklar olup olmadığının, bağımsız deęişkenlere (açıklayıcı deęişken) baęlı olarak istatistiksel olarak test edilmesi
2. Gruplar arasındaki farka her bir deęişkenin katkısının saptanması.
3. Grup içi deęişimine oranla gruplar arasındaki ayrımı maksimize eden tahmin deęişkenleri (doęrusal) kombinasyonunun belirlenmesi ve bu sayede başlangıçtaki açıklayıcı deęişken sayısından daha az sayıda deęişken ile gruplar arasındaki önemli farklılıkların açıklanması.
4. Analiz öncesi tanımlanmış olan grupların birinden geldięi varsayılan yeni bireylerin gruplara atanması ile ilgili yöntemlerin geliştirilmesi, yani yeni bireylerin grup üyeliklerinin saptanması [6].

2.3. Ayırma Analizinin Varsayımları

Ayırma analizinin varsayımları araştırmacının amaçları ile yakından ilgilidir.

1. Şayet araştırmacı yalnızca, grup içi deęişimine oranla gruplar arasındaki deęişmeyi maksimize eden ayırma fonksiyonunu saptamak istiyorsa (ki bu 2 ve 3 numaralı amaçlar demektir.), bütün gruplar için grup varyans ve dağılımlarının birbirine eşit olduğunu varsaymak yetecektir.

2. Şayet araştırmacı istatistiksel önem ile ilgileniyor ise (ki bu 1 nolu amaç demektir), ele alınan nitelikler veya değişkenler açısından grupların bilinmeyen fakat birbirine eşit kovaryans ve dağılımlara sahip olduğu ve bu dağılımların multinormal dağılımlar olduğunu varsaymak gerekecektir.

3. Son olarak, şayet araştırmacı yeni bireylerin gruplara atanması ile ilgileniyor ise (ki bu 4 nolu amaç demektir), yanlış sınıflandırma maliyetlerinin ve yanlış sınıflandırma olasılıklarının eşit olduğunu varsaymak zorunluluğu vardır [7].

Ayrırma Analizini probleminin irdelendiği veri yapısını inceleyelim.

2.4. Veri Yapısı Matrisi

Ayrırma analizine ilişkin veri yapısı, bazı notasyon değişiklikleri dışında, MANOVA'ya ilişkin veri yapısına benzer.

g. grupta (yığın) (G_1, G_2, \dots, G_p), p.değişkenin (X_1, X_2, \dots, X_p) her birine ilişkin n_j ($j:1, 2, \dots, g$) gözlem yapıldığı varsayıldığında aşağıdaki veri tablosu elde edilir [7].

Çizelge 2.1. G - Grup İçin ayrırma analizi verisi

Gruplar	Değişkenler					
	Birimler	X_1	X_2	X_3	...	X_p
1	1	X_{111}	X_{121}	X_{131}	...	X_{1p1}
	2	X_{112}	X_{122}	X_{132}	...	X_{1p2}

	N_1	X_{11N1}	X_{12N1}	X_{13N1}	...	X_{1pN1}
2	1	X_{211}	X_{221}	X_{231}	...	X_{2p1}
	2	X_{212}	X_{222}	X_{232}	...	X_{2p2}

	N_2	X_{21N1}	X_{22N2}	X_{23N2}	...	X_{2pN2}
.	
.	
.	
G	1	X_{g11}	X_{g21}	X_{g31}	...	X_{gp1}
	2	X_{g12}	X_{g22}	X_{g32}	...	X_{gp2}

	N_g	X_{g1N1}	X_{g2N2}	X_{g3N2}	...	X_{gpN2}

2.5. Ayırma Analizinin Geometrik Yorumu

Ayırma analizini önce geometrik olarak irdelemek, konunun daha kolay bir biçimde anlaşılmasını sağlayacaktır.

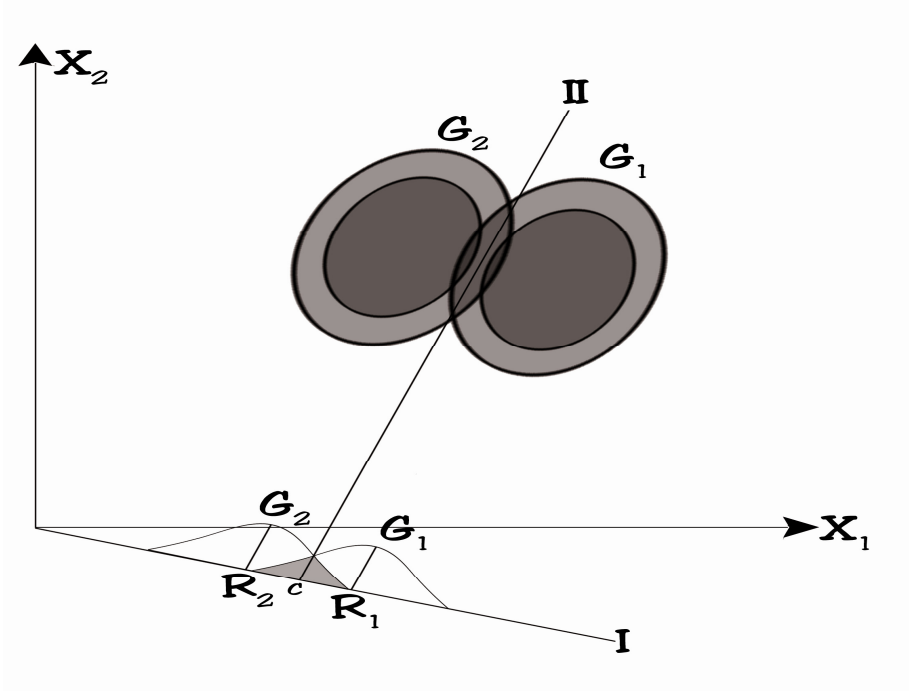
Bunun için, aralarında pozitif yönde az çok bir ilişki olan X_1 ve X_2 değişkeninin Π_1 yığınının N_1 ve Π_2 den N_2 tane birey için gözlemlendiğini varsayalım. Böylece N_1 ve N_2 tane bireyin bu değişkenler üzerinden aldığı her değere, oklidyen uzayda bir nokta karşılık gelecektir. Bu noktaların meydana getirdiği çok-değişkenli normal bulutlar ise, eşit yoğunlukta (ya da eşit frekanslı) birer elips olarak düşünülebilir. Ayrıca, bulutlar yığın merkezi civarında çok yoğun olup, merkezden uzaklaştıkça seyrekleşirler.

Örneğin, G_1 grubunun dış elipsi %90 nını ve dış elips ile ortak merkezli olan iç elipsi de noktaların %75 ini içerir. Elipslerin kesiştiği iki noktadan 2 doğrusu çizilir ve iki boyutlu uzayda bireyleri temsil eden noktaların izi, bu doğruya dikey olan I doğrusu üzerine düşürülürse G_1 ve G_2 grupları arasında kalan kesit elipslerinden çizilebilecek herhangi bir doğrunun vereceği ara kesitten daha küçük olur. Başka bir deyişle, I eksenini üzerinde grupların ortalama vektörlerinin izdüşümleri arasındaki varyans, herhangi bir doğrunun vereceği varyanstan daha büyük olur.

Bireylerin p tane değişken üzerinden aldığı değerlerin önemli eksenler üzerindeki izdüşümleri, bu değerlerin doğrusal bir dönüşümü sonucu bulunur. O halde, I eksenini herhangi bir bireyin p tane değişken üzerinden aldığı değerleri bir tek ayırıcı değere çeviren Ayırma eksenini (ya da ayırma fonksiyonunu) temsil edecektir.

Ayrıca, I ve II doğrularının kesiştiği C noktası, tek boyutlu ayırma uzayını R_1 ve R_2 bölgelerine ayırır. Böylece, R_1 bölgesine düşen noktanın temsil ettiği birey az bir hatayla G_1 grubuna ve R_2 bölgesine düşen noktanın temsil ettiği birey de G_2

grubuna sınıflandırılabilir. Böylece, Ayırma eksenini olan I eksenini, gruplar arası ayırmanın kökenini belirlediği gibi, bireylerin gruplara sınıflandırılması ile ilgili bilgileri de ayırmış olacaktır.



Şekil 2.1. Ayırma analizinin geometrik yorumu

Şekil 2.1.'de görüldüğü gibi, Ayırma eksenini (I), grupların ortak varyans-kovaryans matrisi ile çok değişkenli normal dağılımın gösterdikleri varsayımı altında oluşturulmuştur. Eğer grupların varyans-kovaryans matrisleri eşit değilse, grupların dağılımını birbirine benzemeyecek dolayısıyla, iki grubu ayıran sınır (II doğrusu) en iyi sınır olmayacaktır [4].

Şimdi de buraya kadar yapmış olduğumuz açıklamaların ışığı altında Ayırma Analizi'nin temel problemini istatistiksel olarak ifade etmeye çalışalım.

2.6. Ayırma Analizinin Temel Problemi

“Ayırma Analizi”ni;

- Betimsel amaçlı analiz
- Karar amaçlı analiz

olmak üzere iki ana başlık altında inceleyelim.

2.6.1. Betimsel amaçlı ayırma analizi

Betimsel amaçlı ayırma analizinin temel problemi, gruplar arası kareler toplamının gruplar içi kareler toplamına olan oranını (bir anlamda F oranını) maksimum kılacak ve p tane ayırıcı değişkenini birer doğrusal bileşeni olan ayırıcı faktörleri araştırma problemidir.

Bu problemi istatistiksel olarak aşağıdaki gibi ifade edebiliriz.

Her şeyden önce, p tane bireyin m tane gruba ayrıldığını varsayalım, Yani, k.ci grupta N_k ve toplam birey sayısı $N = \sum_{k=1}^m N_k$ olsun. Ayrıca, her gruptaki bireyler için p tane ayırıcı değişkenin (X_1, X_2, \dots, X_p) nicel değerleri gözlemlenmiş olsun. O halde k.ci grubun ortalama vektörü

$$\bar{X}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} X_{ik} \quad ; k=1, 2, \dots, m \quad (2.1)$$

Genel ortalama vektörü ise,

$$\bar{\bar{X}} = \frac{1}{N} \sum_{k=1}^m \sum_{i=1}^{N_k} X_{ik} \quad (2.2)$$

olup, çok-değişkenli varyans analizi (MANOVA) tablosundan toplam çarpımlar ve kareler toplamı,

$$T = \sum_{k=1}^m \sum_{i=1}^{N_k} (X_{ik} - \bar{\bar{X}})(X_{ik} - \bar{\bar{X}})' \quad (2.3)$$

Gruplar arası çarpımlar ve kareler toplamı,

$$B = \sum_{k=1}^m \sum_{i=1}^{N_k} (\bar{X}_k - \bar{X})(\bar{X}_k - \bar{X})' \quad (2.4)$$

Gruplar içi çarpımlar ve kareler toplamı da

$$W = \sum_{k=1}^m W_k = \sum_{k=1}^m \sum_{i=1}^{N_k} (X_{ik} - \bar{X}_k)(X_{ik} - \bar{X}_k)' \quad (2.5)$$

Ayrıca, her üç çarpımlar ve kareler toplamı arasında,

$$T = B + W \quad (2.6)$$

ilişkisi vardır. Bundan başka bu değerler serbestlik dereceleri ile bölüldüğünde, toplam varyans, gruplar arası varyans ve gruplar içi varyans sırasıyla,

$$S_X = \left(\frac{1}{N-1} \right) T \quad (2.7)$$

$$S_B = \left(\frac{1}{m-1} \right) B \quad (2.8)$$

$$S_W = \left(\frac{1}{N-m} \right) W \quad (2.9)$$

biçiminde elde edilir [8].

Y değişkenini X ayırıcı (ya da tahmin) değişken vektörünün doğrusal bileşeni, C'de X'in öz vektörü olarak,

$$Y = C' X \quad (2.10)$$

$$= C_1 X_1 + C_2 X_2 + \dots + C_p X_p$$

biçiminde yazılır.

Y, X'lerin bir doğrusal bileşeni olduğundan, k.ci grup için ortalama,

$$\bar{Y}_k = \bar{X}_k C$$

ve genel ortalaması da,

$$\bar{Y} = \bar{X} C$$

olarak bulunur.

Y'nin toplam varyansı,

$$\text{Var}(Y) = C' S_X C$$

Ayrırma analizinde m tane grup olduğundan, Y'nin gruplar içi kareler toplamı, her grup için ayrı ayrı hesaplanan grup içi kareler toplamının toplamına eşit olacaktır. O halde, Y'nin gruplar içi kareler toplamı,

$$\begin{aligned} SS_W(Y) &= SS_1(Y) + SS_2(Y) + \dots + SS_m(Y) \\ &= C'W_1C + C'W_1C + \dots + C'W_1C \\ &= C'(W_1 + W_2 + \dots + W_m)C \\ &= C'WC \end{aligned} \tag{2.11}$$

eşitliği ile bulunur.

Y'nin gruplar arası kareler toplamı $[SS_B(Y)]$, bulmak için ilk olarak, X değişken vektörünün gruplar arası çarpımlar ve kareler toplamını içeren B matrisini belirlemek gerekir.

\bar{X}_{ik} : i.ci değişkenin k.ci gruptaki ortalaması

\bar{X}_i : i.ci değişkenin genel ortalamasıdır.

$$\bar{X} = \begin{bmatrix} \bar{X}_{11} & \bar{X}_{21}, \dots, \bar{X}_{p1} \\ \bar{X}_{12} & \bar{X}_{22}, \dots, \bar{X}_{p2} \\ \cdot \\ \cdot \\ \cdot \\ \bar{X}_{1m} & \bar{X}_{2m}, \dots, \bar{X}_{pm} \end{bmatrix}$$

genel ortalamalar matrisi de, her satırı p tane değişkenin genel ortalama vektörünü içeren pxm dereceden,

$$\bar{\bar{X}} = \begin{bmatrix} \bar{X}_1 & \bar{X}_2, \dots, \bar{X}_p \\ \bar{X}_1 & \bar{X}_2, \dots, \bar{X}_p \\ \cdot \\ \cdot \\ \cdot \\ \bar{X}_1 & \bar{X}_2, \dots, \bar{X}_p \end{bmatrix}$$

matrisi ile ifade edilirse, B matrisi aşağıdaki gibi tanımlanabilir:

$$B = (\bar{X} - \bar{\bar{X}})' (\bar{X} - \bar{\bar{X}}) \quad (2.12)$$

Eş. 2.16 eşitliğinin her iki tarafı önden C' ve sondan C ile çarpılırsa,

$$\begin{aligned} C'BC &= C' (\bar{X} - \bar{\bar{X}})' (\bar{X} - \bar{\bar{X}}) C \\ &= (\bar{X}C - \bar{\bar{X}}C)' (\bar{X}C - \bar{\bar{X}}C) \end{aligned}$$

sonucuna varılır.

Böylece,

$$\begin{aligned} SS_b(Y) &= \sum_{k=1}^m N_k (\bar{Y}_k - \bar{Y})^2 \\ &= (\bar{X}C - \bar{\bar{X}}C)' (\bar{X}C - \bar{\bar{X}}C) \\ &= C'BC \end{aligned} \quad (2.13)$$

biçiminde ifade edilebilecektir [5].

Ayırma Ölçütü

Tek değişkenli varyans analizinde

$$SS_T = SS_B + SS_W \quad (2.14)$$

biçiminde olup, gruplar arası farklılık F oranı ile test edilir. Eş. 2.14 te yer alan,

SS_T : Genel kareler toplamı,

SS_B : Gruplar arası kareler toplamı,

SS_W : Gruplar içi kareler toplamıdır.

F oranı ise, gruplar arası kareler ortalamasının gruplar içi kareler ortalamasına oranı olup, aşağıdaki gibi ifade edilir [9]:

$$F = \frac{SS_B / (m-1)}{SS_W / (N-m)} = \frac{SS_B}{SS_W} \cdot \frac{(N-m)}{(m-1)} \quad (2.15)$$

Yukarıdaki eşitlikte yer alan $(N-m) / (m-1)$ terimi, verilen her problem için sabit olduğundan, gruplar arasındaki ayırım, gruplar içi değişime orantılı olarak sadece SS_B / SS_W terimi ile ölçülebilir. Fisher tarafından,

$$\frac{SS_b(Y)}{SS_w(Y)} = \frac{C'BC}{C'WC} = \lambda \quad (2.16)$$

biçiminde tanımlanmıştır.

Bu oran ise, grup farklılıklarını optimum bir düzeyde ölçen bir ölçüt olarak alınacaktır [5].

Ayırma ölçütünün enbüyüklenmesi

Fisher'in ayırma probleminin çözümü için öne sürdüğü ölçüt, gruplar arası kareler toplamının, gruplar içi kareler toplamına oranını maksimum yapacak şekilde C katsayılar vektörünü bulmaktır [10].

Bu oranın maksimum olmasını, sınıfları birbirinden maksimum ölçüde ayırmak ve her bir sınıf içindeki gözlemlerin birbirine mümkün olduğunca yakın olmalarını sağlamak amacıyla isteriz. Bunun için de λ 'nın C 'ye göre kısmi türevi alınıp sifıra eşitlenir.

$$\frac{\partial \lambda}{\partial C} = \frac{2[(BC)(C'WC) - (C'BC)(WC)]}{(C'WC)^2} = 0$$

pay ve payda $C'WC$ ile bölünüp, $C'BC/C'WC$ yerine λ konursa,

$$(B - \lambda W) C = 0 \quad (2.17)$$

Veya

$$(W^{-1}B - \lambda I) C = 0$$

Eşitliği elde edilir [1].

λ_i ($i = 1, 2, \dots, r$) değerleri $W^{-1}B$ matrisinin özdeğerleri olarak,

$$|W^{-1}B - \lambda I| = 0 \quad (2.18)$$

karakteristik denkleminde bulunup, bu özdeğerlere karşı gelen C_i özvektörleri de Eş. 2.17'den elde edilerek, grupları birbirinden en iyi biçimde ayıran Ayırma eksenleri (ya da Ayırma fonksiyonları) oluşturulur [1,5].

Ayırma fonksiyonları

$W^{-1}B$ matrisinin en büyük öz değeri λ_1 ' e karşı gelen C_1 öz vektörü yardımıyla, p tane ayırıcı değişkenin 1.ci doğrusal bileşeni aşağıdaki biçimde oluşturulur:

$$\begin{aligned} Y_1 &= C_1 x \\ &= C_{11}x_1 + C_{12}x_2 + \dots + C_{1p}x_p \end{aligned} \quad (2.19)$$

Burada $x = X - \bar{X}$ olduğundan, 1.ci doğrusal bileşenin ortalaması,

$$E(Y_1) = C_1' E(x) = 0 \quad (2.20)$$

ve varyansı da,

$$\text{Var}(Y_1) = \theta_1 = C_1' \left(\frac{1}{N-1} T \right) C_1 = C_1' S(X) C_1 \quad (2.21)$$

olup, Y_1 doğrusal bileşeni,

$$f_1 = \frac{1}{\sqrt{\theta_1}} Y_1 = \theta_1^{-1/2} C_1 x \quad (2.22)$$

şeklinde standardize edilerek, 1.ci ayırma fonksiyonu elde edilir.

$$U_1 = \theta_1^{-1/2} C_1 \quad (2.23)$$

biçiminde tanımlanırsa, 1.ci ayırma fonksiyonumuz,

$$f_1 = U_1' x \quad (2.24)$$

şeklinde ifade edilebilir.

Betimsel amaçlı ayırma analizi sonucu türetilen ayırma fonksiyonlarının tanımlanması istenilir. Bu betimlemeyi sağlıklı bir biçimde yapabilmek için, her şeyden önce, bu analiz yardımıyla oluşturulan ayırma fonksiyonlarına en çok katkıda bulunan (ya da ayırıcı gücü fazla olan) ayırıcı değişkenleri saptamak gerekir. Bu da ancak, katsayılar vektörünü ölçü birimlerinin etkisinden arındırmakla mümkün olur.

Bunun için, katsayılar vektörünün her elemanı, kendisine karşı gelen değişkenin standart sapmasına çarpılarak standardize edilir. p tane değişkenin standart sapması, $S(X)$ matrisinin diyagonal elemanlarının kareköküne eşit olup D_σ ile gösterilirse, standardize katsayılar vektörü,

$$V_1 = D_\sigma U_1 \quad (2.25)$$

ilişkisiyle bulunur. Böylece, ayırıcı değişkenlerin 1.ci ayırma fonksiyonu olan oransal katkıları U_1 vektörü yerine, V_1 vektörü ile ölçülecektir.

Aynı sonuca X_i değişken vektörünü standardize ederek de varılabilir. Nitekim

$$Z = D_{1/\sigma} X \quad \text{veya} \quad X = D_\sigma Z$$

olup Eş. 2.24 te X_i yerine yukarıdaki eşiti konursa,

$$\begin{aligned} f_1 &= U_1' D_\sigma Z \\ &= V_1' Z \end{aligned} \quad (2.26)$$

biçiminde ifade edilen ve grup ortalamalarını maksimum düzeyde ayıran 1.ci ayırma fonksiyonu elde edilir.

Eğer, 1.ci ayırma fonksiyonu, gruplar arası ayırımın kökenini tanımlamada yetersiz ise, o zaman $W^{-1}B$ matrisinin ikinci en büyük öz değeri olan λ_2 'ye karşı gelen 2.ci ayırma fonksiyonuna gereksinme duyulur. Bu faktör ise,

$$f_2 = V_2' Z \quad (2.27)$$

biçiminde ifade edilir. Bilindiği gibi bu faktör 1.ci faktörden bağımsız olup, gruplar arası ayırımı değişik bir yönde (ya da boyutta) irdelemeye çalışır. Bu ilk iki faktör de yetersiz bulunduğu, bunlardan bağımsız olan 3.ci faktör,

$$f_3 = V_3' Z \quad (2.28)$$

aranılır. Bu süreç, r tane faktör bulunana kadar devam eder. [5,11]

Ayrırma fonksiyonlarının (eksenlerinin) sayısı

Ayrırma fonksiyonlarının sayısı, W matrisinin rankı p ve B matrisinin rankı (m-1) olduğuna göre, $W^{-1}B$ matrisinin rankı (ya da maksimum ayırma sayısı),

$$r = \min(p, m-1)$$

olacaktır.

Buna göre, (m-1) değeri p'den küçük ise, ayırma uzayının olası maksimum rankı (m-1)'e eşit olur. Eğer $(m-1) \geq p$ ise, o zaman p tane ayırma türetilecek, gruplar arası ayırım p boyutlu uzayda betimlenmeye çalışılır. Ancak, pratikte genellikle ilk üç ayırma fonksiyonu ile yetinilmektedir.

Ayrıca, ayırma fonksiyonu sayısının belirlenmesi, bunların ayırıcı gücüne bağlıdır. O halde, bu ayırıcı gücü ölçecek bir ölçüye gereksinme vardır. Bunun için λ_i ($i = 1, 2, \dots, r$) değerleri kullanılabilir. Böylece, ayırmaların ayırıcı gücü (ya da açıklayıcı payı),

$$\lambda_i / \sum_{i=1}^r \lambda_i \quad ; \quad i=1, 2, \dots, S, \dots, r \quad (2.29)$$

olup, λ_1 'e karşı 1.ci ayırma fonksiyonu en çok ayırım gücüne, λ_2 'ye karşı gelen 2.ci ayırma fonksiyonu ise, birinciden az fakat üçüncüden fazla ayırım gücüne sahip olacaktır. Böylece, S tane λ değeri, istatistiksel olarak anlamlı çıkarsa, gruplar arası ayırımın kökenini araştırmak için S tane ayırma fonksiyonu geliştirilir [5].

Faktör yapısı

R tane ayırma fonksiyonu ile p tane ayırıcı değişken (ya da tahmin değişkeni) arasındaki korelasyon katsayılarını içeren faktör yapısı matrisi de,

$$A=RV \quad (2.30)$$

eşitliğinden elde edilir.

Bu matris, fonksiyonların ya da ayırıcı faktörlerin tanımlanmasında büyük kolaylıklar sağlar [11].

2.6.2. Karar amaçlı ayırma analizi

Karar amaçlı ayırma analizi tekniği (ya da sınıflandırma tekniği) ile mevcut, yığınlardan birine ilişkin olduğu varsayılan fakat hangi yığından geldiği bilinmeyen bir bireyin, en az hatayla, eldeki yığınlardan birine sınıflandırılmasına karar verilir. Bu amaçla, çeşitli karar kuralları geliştirilmiştir. Bunlar,

- Geleneksel sınıflandırma,
- Gruplara ait olma olasılıklarına göre sınıflandırma

olmak üzere iki ana başlık altında toplanabilir.

Geleneksel sınıflandırma kuralı

Geleneksel sınıflandırma yöntemi, her grup için p tane değişkenin doğrusal bileşeni olan sınıflandırma istatistiklerinin (ya da sınıflandırma fonksiyonlarının) türetilmesini gerektirir. Mahalanobis'in genelleştirilmiş uzaklık fonksiyonundan hareketle bulunan bu istatistiklerinin yardımıyla, hangi yığından geldiği bilinmeyen X_i bireyinin eldeki gruplardan birine atanmasına çalışılır [1].

$$D_{ik}^2 = (\mathbf{X}_i - \bar{\mathbf{X}}_k)' \mathbf{S}_w^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}_k)$$

biçiminde tanımlanan genelleştirilmiş uzaklık fonksiyonunun sağ tarafı açılırsa,

$$\begin{aligned} D_{ik}^2 &= (\mathbf{X}_i' \mathbf{S}_w^{-1} \bar{\mathbf{X}}_k' \mathbf{S}_w^{-1}) (\mathbf{X}_i - \bar{\mathbf{X}}_k) \\ &= \mathbf{X}_i' \mathbf{S}_w^{-1} \mathbf{X}_i - \mathbf{X}_i' \mathbf{S}_w^{-1} \bar{\mathbf{X}}_k - \bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \mathbf{X}_i + \bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \bar{\mathbf{X}}_k \\ &= \mathbf{X}_i' \mathbf{S}_w^{-1} \mathbf{X}_i - 2\bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \mathbf{X}_i + \bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \bar{\mathbf{X}}_k \end{aligned}$$

eşitliğine varılır. k indisine bağlı olmayan $\mathbf{X}_i' \mathbf{S}_w^{-1} \mathbf{X}_i$ terimi gruptan gruba değişmediği için, bireyi temsil eden \mathbf{X}_i vektörünün doğrusal bir bileşeni halinde yukarıdaki uzaklık fonksiyonu, k indisine bağlı olarak,

$$f_k(\mathbf{X}_i) = \bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \bar{\mathbf{X}}_k - 2\bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \mathbf{X}_i, \quad k=1, 2, \dots, m \quad (2.31)$$

biçiminde yazılabilir. Eş. 2.31 değerine k .ci grubun sınıflandırma değeri denir. Dolayısıyla, \mathbf{X}_i gözlemi sınıflandırma değeri en küçük olan gruba dâhil edilir. Başka bir deyişle,

$$f_k(\mathbf{X}_i) = \min \{f_1(\mathbf{X}_i), f_2(\mathbf{X}_i), \dots, f_m(\mathbf{X}_i)\} \quad (2.32)$$

ise, \mathbf{X}_i gözlemi k .ci gruba atanır.

Eş 2.31'deki fonksiyon aşağıdaki biçimde de ifade edilebilir:

$$G_k(\mathbf{X}_i) = \bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \mathbf{X}_i - (1/2) \bar{\mathbf{X}}_k' \mathbf{S}_w^{-1} \bar{\mathbf{X}}_k, \quad k=1, 2, \dots, m \quad (2.33)$$

Burada dikkat edilmesi gereken konu $G_k(\mathbf{X}_i)$ fonksiyonunun Fisher'in ayırma fonksiyonu olmasıdır.

O zaman karar kuralı,

$$G_k(\mathbf{X}_i) = \max \{G_1(\mathbf{X}_i), G_2(\mathbf{X}_i), \dots, G_m(\mathbf{X}_i)\} \quad (2.34)$$

biçimine dönüşür.

Ayrıca, E.ş 2.31 ve Eş. 2.33 i incelediğimizde $f_k(X_i)$ ile $G_k(X_i)$ arasında

$$f_k(X_i) = -2 G_k(X_i)$$

biçiminde bir ilişki olduğunu kolaylıkla görebiliriz [5].

Gruplara ait olma olasılıklarına göre sınıflandırma

Bireylerin her grupta çok-değişkenli normal dağılım gösterdikleri varsayımı altında sınıflandırma değerleri, gruplara ait olma olasılıklarına kolaylıkla dönüştürülebilir. Böylece, her hangi bir bireyin minimum ki-kare değerine göre, eldeki gruplardan birine atanması kuralı, maksimum olasılığa sahip olduğu gruba sınıflandırılması kuralına özdeş olur. Nitekim $X_i = X_1, X_2, \dots, X_p$ değişken vektörünün her grupta çok-değişkenli normal dağıldığı koşulu altında uzaklık ölçüsü olarak kullanılan ki-kare değeri belirli bir olasılığa orantılı olarak bulunabilir.

Grupların ortak dağılım matrisi ile çok-değişkenli normal dağıldıkları ve analiz öncesi olasılıkların eşit olduğu varsayıldığında,

$$P(G_k|X_i) = \frac{e^{-D_{ik}^2/2}}{\sum_{j=1}^m e^{-D_{ij}^2/2}}, \quad k=1, 2, \dots, m \quad (2.35)$$

biçimine dönüşür. Burada,

$$\sum_{k=1}^m P(G_k|X_i) = 1 \quad (2.36)$$

Ve

$$D_{ik}^2 = (X_i - \bar{X}_k)' S_w^{-1} (X_i - \bar{X}_k) \quad (2.37)$$

dir.

Denkleminden de anlaşılacağı gibi, $P(G_k | X_i)$ olasılığının maksimum olması, D_{ik}^2 'nin minimum olmasına bağlıdır. O halde, minimum ki-kare değerine göre yapılan sınıflandırma, maksimum $P(G_k | X_i)$ olasılığına göre yapılan sınıflandırmaya özdeş olacaktır.

Analiz sonrası (aposteriori) olasılıklara göre geliştirilen karar kuralı ise,

$$R_k = P(G_k | X_i) = \max_{k=1, 2, \dots, m} (P(G_1 | X_i), P(G_2 | X_i), \dots, P(G_m | X_i)) \quad (2.38)$$

biçiminde ifade edilir [16].

3. AYIRMA ANALİZİNDE UYGULANAN TESTLER

Ayırma analizinin birtakım varsayımları olduğundan daha önce bahsetmiştik. Analizin sağlıklı bir şekilde yürütülebilmesi için bu varsayımların sağlanıp sağlanmadığının kontrol edilmesi gerekir.

Kısaca varsayımlar;

- Verilerin çok değişkenli normal dağıldığı bir yığından geldiği,
- Grupların eşit varyans-kovaryansa sahip oldukları,
- Grup ortalamaları arasında farklılık olup olmadığı,
- Oluşturulan ayırma fonksiyonlarının anlamlı olup olmadıklarıdır.

Şimdi bu varsayımların kontrolü için gerekli testleri aşağıda inceleyelim.

3.1. Çok Değişkenli Normalliğin Testi

Çok değişkenli normallik varsayımının geçerli olup olmadığının kontrol edilmesindeki amaç, gözlemlerin çok değişkenli normal yığınlardan gelip gelmediklerinin test edilmesi ve bunun sonucunda gözlem verileri üzerinde uygun düzeltmelerin yapılabilmesidir.

Bir veri setinin normalliğinin test edilmesi için D^2 uzaklık ölçüsü kullanılır.

$$D_{ik}^2 = (X_i - \bar{X}_k)' S_W(x)^{-1} (X_i - \bar{X}_k) \quad (3.1)$$

Burada S ortak gruplar içi kovaryans matrisi olup,

$$S_W = \left(\frac{1}{N-m} \right) W \quad (3.2)$$

dir.

Bu yöntemin en büyük avantajı p tane değişkene doğrudan uygulanabilmesidir.

Uygulama aşamaları sırasıyla şöyledir;

1. D_{ij}^2 , uzaklık kareleri hesaplanır,
2. Bulunan değerler küçükten büyüğe doğru sıralanır,
3. D_{ij}^2 , $p(i-1/2)n$ serbestlik derecesiyle χ^2 dağılımı gösterir ve D_{ij}^2 çiftlerinin grafiği çizilir, bu doğru ki-kare grafiği olarak adlandırılır,
4. Elde edilen ki-kare grafiğinin şeklinin bir doğru oluşturması halinde, çok değişkenli normalliğin geçerliliğine karar verilir [1].

3.2. Grup Kovaryans Matrislerinin Eşitliği Testi

Ayırma analizinde grupların kovaryans matrislerinin eşit olması durumunda “doğrusal ayırma fonksiyonları”, grupların kovaryans matrislerinin farklı olması durumunda ise “karesel ayırma fonksiyonları” uygun olmaktadır. Bu nedenle ayırma analizine başlamadan önce grup kovaryans matrislerinin eşitliğinin test edilmesi gerekir.

m tane p değişkenli normal grubun kovaryans matrislerinin eşitliğini test etmek için

$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_m$ hipotezini ele alalım.

Bu hipotezi test etmek için k. gruptan N_k , $k=1, 2, \dots, m$ sayıda bireyden oluşan bir rasgele örnek için S_k örnek kovaryans matrisi, Σ_k 'nin sapmasız bir tahmini olsun.

H_0 hipotezi doğru ise, ortak kovaryans matrisi,

$$S=(N-m) \sum_{k=1}^m (N_k -1)S_k \quad ; \quad N=\sum_{k=1}^m N_k \quad (3.3)$$

şeklinde olacaktır. Bu durumda test istatistiği şu şekilde ifade edilir:

$$V=(N-m)\ln \left| S \right| - \sum_{k=1}^m (N_k -1)\ln \left| S_k \right| \quad (3.4)$$

ve

$$C^{-1}=1 - \frac{2p^2 + 3p - 1}{6(p+1)(m-1)} \left| \frac{1}{\sum (N_k -1)} - \frac{1}{(N-m)} \right| \quad (3.5)$$

ölçü faktörünü kullanılarak, VC^{-1} miktarının örnek hacmi büyüdükçe $1/2(m-1)p(p+1)$ serbestlik dereceli bir ki-kare dağılımına yaklaştığı gösterilmiştir.

$$VC^{-1} \leq \chi_{1/2(m-1)p(p+1)}^2 \quad (3.6)$$

Bu durumda H_0 hipotezinin kabul edileceği yani grupların kovaryans matrislerinin farklı olmadığı kararına varılır [12].

3.3. Grup Ortalama Vektörlerinin Farklılığı Testi

Ayrırma analizine başlamadan önce grup ortalama vektörlerinin farklılığının test edilmesi gereklidir. Eğer yapılan testler sonucunda grup ortalama vektörleri arasında önemli bir farklılık olmadığı kararına varılırsa, ayrırma analizi gereksiz olacaktır. Çünkü bu durumda grupları tam anlamıyla birbirinden ayırmak mümkün olmayacak ve bireylerin gruplara sınıflamasında yeterince iyi ölçüler geliştirilemeyecektir.

Bu amaçla,

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_m \text{ hipotezine karşılık}$$

H_1 : En az iki μ birbirinden farklıdır (μ 'lardan en az biri farklıdır) hipotezi test edilir.

H_0 hipotezini test etmek için, S. S. Wilks tarafından geliştirilen ve Wilks'in Lambdası olarak adlandırılan test istatistiği kullanılır. Bu istatistik,

$$\Lambda = \frac{|W|}{|T|} \text{ dir.} \quad (3.7)$$

M. S. Bartlett Λ istatistiğine dayanarak, $p(m-1)$ serbestlik derecesiyle yaklaşık ki-kare test istatistiğinin aşağıdaki biçiminde geliştirilmiştir.

$$\chi^2_{p(m-1)} = -[N-1-(p+m)/2] \ln \Lambda \quad (3.8)$$

Eş. 3.8 ilişkisinden hesaplanan χ^2 değeri, $\alpha = 0,05$ anlamlılık düzeyinde $p(m-1)$ serbestlik derecesine karşı gelen χ^2 tablo değerinden büyük ise, $\chi^2_H > \chi^2_{p(m-1)}$ en az iki grup ortalama vektörünün birbirinden farklı olduğuna karar verilecektir [5].

3.4. Ayırma Fonksiyonlarının Anlamlılık Testi

Bir önceki bölümde, ayırma sayısının $W^{-1}B$ matrisinin sıfırdan farklı öz değerleri sayısına ve bu sayının da $r = \min(p, m-1)$ ' e eşit olduğuna değinmiştik. Oysa pratikte gruplar arası ayırım daha az boyutlu bir sistemde betimlenmek istenir. Bu yüzden r tane ayırma arasından, istatistiksel bakımdan anlamlı olanlarını belirlemek gerekir. Bunun için, gruplar arasındaki farkın önem kontrolünde kullanılan Wilks test istatistiğinden yararlanır. (Λ) test ölçütü, λ ayırma ölçütüne bağlı olarak aşağıdaki gibi bulunur:

$$1/\Lambda = |T|/|W| = |W^{-1}T|$$

T yerine $W + B$ konursa,

$$\begin{aligned} 1/\Lambda &= |W^{-1}(W+B)| \\ &= |I+W^{-1}B| \\ &= \prod_{i=1}^r (1+\lambda_i) \end{aligned}$$

Buradan

$$\Lambda = 1 / \prod_{i=1}^r (1+\lambda_i) \quad (3.9)$$

ilişkinine varılır.

Bartlett, (Λ) ölçütünü kullanarak, $p(m-1)$ serbestlik derecesiyle yaklaşık ki_kare dağılımını gösteren aşağıdaki test istatistiğini geliştirmiştir.

$$\begin{aligned} X_{p(m-1)}^2 &= -[N-1-(p+m)/2] \ln \Lambda \\ &= [N-1-(p+m)/2] \ln [(1+\lambda_1)(1+\lambda_2) \dots (1+\lambda_r)] \\ &= [N-1-(p+m)/2] \sum_{i=1}^r \ln(1+\lambda_i) \end{aligned} \quad (3.10)$$

Bu istatistik, seçilen α anlamlılık düzeyinde $X_{p(m-1)}^2$ tablo değeri ile karşılaştırılarak, gruplar arasındaki farkın önemli olup olmadığı test edilir. Gruplar arasındaki farklılık istatistiksel bakımdan önemli ise, ayırmalardan en az birinin önemli olduğu anlamına gelir.

İlk s tane ayırma anlamlı çıktıktan sonra, geri kalan (r-s) tane ayırmanın anlamlılık testi ise, aşağıdaki biçimde yapılabilir:

$$\chi^2 = -[N-1-(p+m)/2] \ln \Lambda' \quad (3.11)$$

Burada

$$\Lambda' = \prod_{i=s+1}^r 1/(1+\lambda_i)$$

ve serbestlik derecesi ise

$$s.d. = (p-s)(m-s-1)$$

dir [5].

4. UYGULAMA

Türkiye'de de su, mülkiyeti ve işletmeciliği bakımından kamuya aittir. Ulusal çapta sorumlu kurum Devlet Su İşleri Genel Müdürlüğü'dür. Kırsal yerleşmelerin su yönetimi Köy Hizmetleri Genel Müdürlüğü'ne, kentsel yerleşmelerdeki su yönetimi İller Bankası Genel Müdürlüğü'ne verilmiştir. Her üç kurum da bölge müdürlükleri temelinde örgütlenmiş, genel olarak bölge sınırları birbirlerine koşut belirlenmiştir. Yetki alanlarında planlama ve yatırımları gerçekleştiren bu kurumlar, tesislerin işletmesini genel olarak yerel yönetimlere, bunlar tarafından kurulan birliklere, kooperatiflere devretmektedirler. Yerleşmeler bazında sorumlu kurumlar yerel yönetimlerdir. Su işletmeciliği, yerel yönetimlere tekel hakkı olarak verilmiş, buna bağlı olarak su hizmetinin özel sektöre gördürülmesi söz konusu olduğunda imtiyaz sözleşmesi yapmak yükümlülüğü getirilmiştir.

Belediye içme suyu ve kanalizasyon hizmeti, 1980'li yıllara kadar odağında İller Bankası'nın bulunduğu, kamu kaynaklarına ve kamu kredilerine dayanan bir yatırım ve finansman modeli eliyle gerçekleştirilmiştir. 1980'li yıllardan bu yana İller Bankası odaklı model değişmeye başlamıştır. Sektör doğrudan belediyelere bırakılarak finansmanda dış kredi kullanımı genişletilmiştir. Büyükşehir belediyelerine bağlı olarak kullanılan su ve kanalizasyon idareleri en büyük harcamacı kurumlar haline gelmiş bulunmaktadır.

Devlet İstatistik Enstitüsü her yıl belediyelerde su istatistiklerini yayınlamaktadır. Bu istatistikleri, genel sanayi sayımında tam sayım olarak gerçekleştirmektedir. Diğer yıllarda ise nüfusu 20 000 üzeri tam sayım, nüfusu 20 000 altına ise örnekleme metodu uygulanarak yaklaşık 400 belediye seçilmektedir. Seçilen belediyeler Türkiye su istatistiklerinin % 75'e yakınına kapsamaktadır.

Su istatistikleri toplanırken birçok sorunla karşılaşılmaktadır. Bu problemler kısaca iki ana başlık altında özetlenebilir:

- Büyükşehir belediyeleri dışında belediyelerin kendilerine ait su birimlerinin olmaması,
- Belediyelerin göndermiş olduđu verilerin bir önceki yıllla tutarlılık göstermemesidir.

Sektör uzmanları, Büyükşehir belediyelerini nüfus büyüklüklerine göre sınıflamış, böylece deđişkenler için ortalama deđerler elde ederek diđer belediyelerin verilerini kontrol etmeyi amaçlamışlardır. Fakat birçok belediyede sadece nüfus büyüklüğünün yeterli olmadığı görülmüştür.

Örneğin, Şanlı Urfa ilinin nüfusu 1 500 000 civarında olmasına rağmen, konutta yaşayan kişi sayısının yüksek olması sebebiyle Urfa belediyesinin su verileri, daha az nüfuslu belediyelerin verileriyle uyumaktadır.

Bu çalışmada, nüfus büyüklüklerine göre üç grup altında toplanan 16 Büyükşehir belediyesine ayırma analizi uygulanmıştır.

Çalışmanın amacı; analize alınan deđerşkenler bakımından grupları yeniden bir sınıflandırmaya tabi tutmak ve analiz öncesi gruplandırma dışında bırakılan illeri, eldeki gruplardan birine atamaktır.

4.1. Araştırmanın Kapsamı

Çalışmada DİE'nin 2001 Büyükşehir Belediyelerinde Su İstatistikleri verileri kullanılmıştır (16). Büyükşehir belediyeleri, toplam 3138 belediyenin, istatistikler bakımından %61'ini kapsamaktadır. Bu sebeple Büyükşehir belediyelerinin verilerinin alınması uygun görülmüştür.

4.2. Analize Alınan Deđerşkenler ve Bu Deđerşkenlerin Tanımları

Analize alınan deđerşkenler ve bu deđerşkenlerin tanımları şöyledir:

Belediyelerde Toplanan Su Miktarı (TSM) : Belediyelerin su kaynaklarından (kuyu, akarsu, baraj vb.) topladığı veya satın aldığı su miktarıdır. 2001 yılı için belediyelerde toplanan su miktarı EK-1’de verilmiştir.

Belediyelerin Dağıttı Su Miktarı (DSM) : Belediyelerin abonelerine dağıttı su miktarıdır. 2001 yılı için belediyelerin dağıttığı su miktarı EK-1’de verilmiştir.

Belediyelerde Kayıtlı Olan Abone Sayıları (Abone) : Belediyelerde kayıtlı olan aboneler, Resmi Kuruluşlar, Sanayi İşletmeleri ve Ticarethaneler, Meskenler ve İnşaatlar, Park ve Bahçeler, Halk Çeşmeleri ve diğerleri olmak üzere ayrılırlar. Analimize toplam abone sayısı alınmıştır. 2001 yılı için belediyelerde abone sayıları EK-1’de verilmiştir.

Belediyelerin Su Gelirleri (Gelirler) : Belediyelerin abonelerinden ve diğer hizmetlerden elde ettiği toplam gelir. 2001 yılı için belediyelerin su gelirleri EK-1’de verilmiştir.

Belediyelerde İşçi Saat Toplamı (Saat) : Belediyelerin su işinde çalıştırdığı işçilerin, çalıştıkları saatlerin toplamı. 2001 yılı için belediyelerin işçi saat toplamları EK-1’de verilmiştir.

Belediyelerde Çalışan Sayısı (Çalışan) : Belediyelerde yıl için de su hizmetinde çalışan işçi ve memurların ortalama sayısı. 2001 yılı için belediyelerde çalışan işçi ve memur sayıları EK-1’de verilmiştir.

Belediyelerin Çalışanları İçin Yaptığı Ödemeler (Ödemeler) : Belediyelerin çalışanları için yaptığı maaş, fazla mesai, ikramiye, prim, tazminatlar, sosyal yardımlar ödemeleri ile ssk ve emekli sandığı ödemelerini kapsar. 2001 yılı için belediyelerin çalışanlarına yaptığı ödemeler EK-1’de verilmiştir.

Belediyelerin Satın Aldığı Elektrik (Elektrik) : Belediyelerin yıl içinde aldığı elektriğin alış değeri. 2001 yılı için belediyelerin satın aldıkları elektrik değerleri EK-1’de verilmiştir.

Belediyelerin Su Motorlarından Elde Ettiği Verim (Verim) : Belediyelerin su çıkarmada kullandıkları motorların saniye/litre oranları. 2001 yılı için belediyelerin verimleri EK-1’de verilmiştir.

Belediyelerin Diğer Giderleri (Diğer) : Belediyelerin satın aldığı hammadde ve yardımcı madde giderleri ile hizmet giderlerini (kira gideri ve faiz ödemeleri) kapsar. 2001 yılı için belediyelerin diğer giderleri EK-1’de verilmiştir.

4.3. Araştırmaya Alınan Belediyelerin Analiz Öncesi Gruplanması

Çalışmada yer alan belediyeler 2000 yılı nüfus sayımı sonucundaki büyüklüklerine göre 3 gruba ayrılmıştır. Nüfus büyüklükleri; 1 500 000 ve aşağısı, 1 500 000 –3 000 000, 3 000 000 ve üzeri olarak sınıflanmıştır. Nüfus gruplamasına göre belediyeler ve grupları şöyledir:

1. Grup :

- ESKİŞEHİR
- ADAPAZARI
- ERZURUM
- KAYSERİ
- KOCAELİ
- SAMSUN
- GAZİANTEP
- DİYARBAKIR

2. Grup

- İÇEL
- ANTALYA
- ADANA
- BURSA
- KONYA

3. Grup

- İZMİR
- ANKARA
- İSTANBUL

4.4. Analiz

Türkiye Büyükşehir Belediyelerine EK-1’de verilen 10 değişken üzerinden ayırma analizi uygulanmış ve örnek seçilmiş bir belediyenin, oluşan yeni gruplardan hangisine dâhil olacağı araştırılmıştır.

Çalışmada SPSS 11 paket programı kullanılmıştır. Analiz sonuçları tablolar halinde düzenlenmiş ve yorumları ile birlikte aşağıda verilmiştir.

Ayrıca STATISTICA 7.0 programı yardımıyla uzaklık ölçüleri bulunmuştur.

4.4.1. Grup istatistikleri

Araştırmaya alınan 10 değişkenin, 3 gruba ve toplama göre ortalamaları ve standart sapmaları EK-2’de verilmiştir.

4.4.2. Ayırma fonksiyonları ve anlamlılıkları

Gruplar arası ayırım $m-1=3-1=2$ ayırma uzayında yapılabilmektedir. Bu 2 ayırıcı uzayın özdeğerleri, ayırma fonksiyonlarıyla açıklanan varyans yüzdeleri ve bunların birikimli yüzdeleri ile ayırma fonksiyonları ile ayırıcı değişkenler arasındaki ilişkiyi veren kanonik korelasyonlar Çizelge 4.2’de verilmiştir. Ayrıca analiz öncesi yapılan testlerde VERİM değişkeninin anlamsız olduğu saptanmış ve sonuçları Çizelge 4.1’de verilmiştir.

Çizelge 4.1. VERİM değişkeninin tolerans değeri

Değişken	Gruplar Arası Kovaryans	Tolerans	Minimum Tolerans
VERİM	8168028,5647	0,000	0,001

VERİM değişkeninin tolerans değeri minimum tolerans değerinden ufak olduğundan VERİM değişkeni analizde kullanılmamıştır.

Çizelge 4.2. Ayırma fonksiyonları, özdeğerler, açıklanan varyans payları, kanonik ilişkiler

Ayırma Fonksiyonları	Özdeğerler	Açıklanan Varyans %	Birikimli %	Kanonik Korelasyon
1	55,6739	98,6717	98,6717	0,9911
2	0,7494	1,3282	100	0,6545

Ayırma fonksiyonlarının hangilerinin anlamlı olduğuna karar vermek için uygulanan testin sonucu Çizelge 4.3'te verilmiştir.

Çizelge 4.3. Ayırma fonksiyonlarının test istatistikleri

Fonksiyonların Testi	Wilks' Lambda	Ki-Kare Değerleri	Serbestlik Derecesi	Anlamlılık Düzeyi
1	0,0100	41,3694	18	0,001
2	0,5716	5,0336	8	0,753

1. ayırma fonksiyonunun anlamlılık düzeyi 0.05'ten küçük olduğundan belediyeler 1. ayırma fonksiyonu ile ayrılırlar.

Analize alınan değişkenlerin görel katkıları Çizelge 4.4'de verilmiştir. Bu çizelge bize değişkenlerden hangilerinin ayırma ne kadar katkı yaptıklarını göstermektedir.

Çizelge 4.4. Standartlaştırılmış ayırma fonksiyonlarının katsayıları

Değişkenler	Fonksiyon
TSM	-4,1067
DSM	-1,4002
GİDER	5,2113
ABONE	1,5976
GELİR	-10,9925
ÇALIŞAN	-5,4034
ÜCRET	7,3084
SAAT	5,8266
ELEKTRİK	2,0860

Çizelge 4.4'den anlaşılacağı gibi gruplar arası ayırma en önemli katkısı yapan değişkenler sırasıyla GELİR, ÜCRET ve SAAT olduğunu görebiliriz. Ayrıca 1. ayırma fonksiyonu bakımından ayırıcı gücü düşük olan değişken ise DSM'dir.

Çizelge 4.4'den faydalanarak ayırıcı faktör eşitliği,

$$f_1 = -4,1067z_1 - 1,4002z_2 + 5,2113z_3 + 1,5976z_4 - 10,9925z_5 \\ - 5,4034z_6 + 7,3084z_7 + 5,8266z_8 + 2,0860z_9$$

olarak yazılabilir. Z değişkenlerinin, orijinal değişkenlerin standart normal dağılıma dönüşmüş halleri olduğunu unutmamak gerekir.

Ayırıcı faktör katsayıları, ayırıcı faktör ile ayırıcı değişkenler arasındaki ilişkileri tam olarak yansıtamadığından, bu faktörleri tanımlayabilmemiz için faktör yapısını bulmamız gerekmektedir.

Ayırıcı faktörler ile bu faktörleri oluşturan ayırıcı değişkenler arasındaki korelasyon katsayılarını içeren faktör yapı matrisi Çizelge 4.5'da verilmiştir.

Çizelge 4.5. Faktör yapı matrisi

Değişkenler	Fonksiyon 1
VERİM	-0,1994
TSM	-0,1994
GELİR	-0,1933
ÇALIŞAN	-0,1924
DSM	-0,1903
ÜCRET	-0,1729
ABONE	-0,1611
SAAT	-0,1332
GIDER	-0,1318
ELEKTRİK	-0,1276

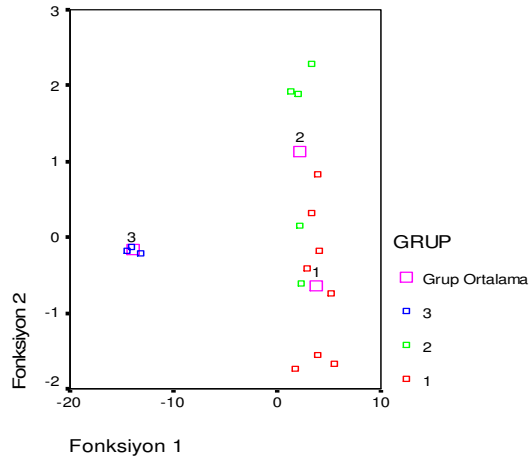
Tüm değişkenlerine ayırıcı faktör ile düşük bir korelasyon katsayısına sahip olduğu görülmektedir.

Grupların ayırıcı faktör değerleri Çizelge 4.6’da verilmiştir.

Çizelge 4.6. Grup ortalamalarının ayırıcı faktör değerleri

GRUP	FAKTÖR 1	FAKTÖR 2
1	3,8257	-0,6417
2	2,2337	1,1280
3	-13,9250	-0,1686

İki ayırma fonksiyonu üzerinde 3 grubun gösterimi Şekil 4.1’de verilmiştir.



Şekil 4.1. İki Ayırıcı Fonksiyon üzerinde 3 grubun görünümü

4.4.3. Belediyelerin sınıflandırılması

Belediyelerin analiz öncesi ve analiz sonrası yer aldığı gruplar ile ayrııcı faktör değerleri EK-3’da verilmiştir.

Analiz sonrası belediyelerin çoğunun kendi gruplarına atandığı görülmektedir.

Çizelge 4.7’deki sınıflandırma sonuçlarına baktığımızda analiz sonrası % 93.8’ lik doğru bir sınıflandırma yapıldığı görülmektedir. Nitekim, belediyeler içinden bir Bursa belediyesi, analiz öncesi ikinci grupta iken, analiz sonrası birinci gruba atanmıştır. Bunun başlıca nedeni, Bursa belediyesinin dağıttığı su miktarının düşük olmasıdır.

Çizelge 4.7. Sınıflandırma sonuçları

		GRUP	Analiz Sonrası Grup Üyeliği			Toplam
			1	2	3	
Orijinal	Sayı	1	8	0	0	8
		2	1	4	0	5
		3	0	0	3	3
	%	1	100	0	0	100
		2	20	80	0	100
		3	0	0	100	100

4.4.4. Örnek seçilen belediyenin oluşturulan gruplara atanması

Bu bölümde Balıkesir ili örnek olarak seçilmiş ve eldeki gruplardan birine atanması incelenmiştir. Balıkesir ilinin 2001 yılı su istatistikleri Ek-4’te verilmiştir.

Gruplara ait olma olasılıklarına göre sınıflandırma yöntemi kullanılmış ve D_{ik}^2 uzaklık değerleri EK-5’te verilmiştir.

$$P(G_k|X_i) = \frac{e^{-D_{ik}^2/2}}{\sum_{j=1}^m e^{-D_{ij}^2/2}}, \quad k=1, 2, \dots, m \quad \text{formülünden hesaplanan olasılıklar:}$$

$$P(G_1|X) = 0,52491 \quad \text{max}$$

$$P(G_2|X) = 0,29412$$

$$P(G_3|X)=0,17647$$

1. olasılık değeri maksimum olduğundan Eş. 2.38 'deki karar verme kuralı gereğince Balıkesir ili 1. gruba dâhil edilir.

Şimdide Fisher'in ayırma fonksiyonu yardımıyla Balıkesir ilini atayalım. Böylece ayırma modelimizi de ortaya koymuş olacağız.

Fisher'in ayırma fonksiyonları 3 grup için Çizelge 4.8'de verilmiştir. Bu katsayılar, bağımsız değişkenlerin grupların ayrılmasına ne kadar katkıda bulduklarını (ne düzeyde iyi tahmin edici olduklarını) tanımlamaktadır [13].

Çizelge 4.8. Fisher'in ayırma fonksiyonları

	GRUP		
	1	2	3
TSM	1,88909E-07	2,72609E-07	1,00074E-06
DSM	1,40158E-07	1,9412E-07	5,30493E-07
GIDER	-1,73329E-06	-2,10826E-06	-1,23193E-05
ABONE	-7,93021E-06	1,17091E-05	-5,81466E-05
GELİR	2,1172E-07	3,61061E-07	3,27157E-06
ÇALIŞAN	0,0175	0,0265	0,1138
ÜCRET	-2,11246E-06	-3,21774E-06	-1,38104E-05
SAAT	-3,93834E-06	-1,30212E-05	-5,06091E-05
ELEKTRİK	-6,80563E-07	-1,02366E-06	-4,86548E-06
(Sabit)	-5,7440	-12,7859	-206,8785

$$f_1 = -5,7440 + 1,88909E-07 X_1 + 1,40158E-07 X_2 - 1,73329E-06 X_3 - 7,93021E-06 X_4 + 2,1172E-07 X_5 + 0,0175 X_6 - 2,11246E-06 X_7 - 3,93834E-06 X_8 - 6,80563E-07 X_9$$

$$f_2 = -12,7859 + 2,72609E-07 X_1 + 1,9412E-07 X_2 - 2,10826E-06 X_3 + 1,17091E-05 X_4 + 3,61061E-07 X_5 + 0,0265 X_6 - 3,21774E-06 X_7 - 1,30212E-05 X_8 - 1,02366E-06 X_9$$

$$f_3 = -206,8785 + 1,00074E-06 X_1 + 5,30493E-07 X_2 - 1,23193E-05 X_3 - 5,81466E-05 X_4 + 3,27157E-06 X_5 + 0,1138 X_6 - 1,38104E-05 X_7 - 5,06091E-05 X_8 - 4,86548E-06 X_9$$

Balıkesir ilinin EK-4'teki verileri kullanılarak oluşturulan fonksiyonlar aşağıdadır:

$$f_1 = -3214,06 \quad \max$$

$$f_2 = -7155,4681$$

$$f_3 = -115834,192$$

Eş. 2.34 deki karar kuralı gereğince Balıkesir ili 1. gruba atanır.

5. SONUÇ

Uygulamada Türkiye Büyükşehir Belediyelerine, 10 deęişken üzerinden ayırma analizinin bütün varsayımlarının sağlanıp sağlanmadığı test edilerek analiz edilmeye çalışılmıştır.

Analiz sonrası belediyelerin çoęu kendi gruplarında kalmıştır. Analiz sonrası grubu deęişen belediye Bursa'dır.

Sanayisi son derece gelişmiş olan ve bu sebeple sanayi bölgesine dağıttığı su miktarı ve buradan elde ettiği gelirin yüksek olmasına rağmen Bursa'nın 3. grubu zorlayamadığını ve analiz sonrası 1. gruba atandığını görüyoruz. Bunun başlıca nedeni, Bursa belediyesinin dağıttığı su miktarı oldukça düşük olmasıdır. Ayrıca sektöründe çalıştırdığı işçi ve memur sayısının oldukça yüksek olmasının da, gelir-gider dengesini bozduğunu söyleyebiliriz.

Sonuç itibariyle Türkiye'de su istatistiklerinin kısmen de olsa illerin nüfusuna bağlı olduğunu söyleyebiliriz.

KAYNAKLAR

1. Tatlıdil, H., ‘Uygulamalı Çok Değişkenli İstatistiksel Analiz’, *Akademi Matbaası*, Ankara, 1, 265-278 (1992).
2. Özdamar, K., ‘Paket Programlar İle İstatistiksel Veri Analizi-2’, *Kaan Kitabevi*, Eskişehir, 341-394 (2002).
3. Akgül, A. ve Çevik, O., İstatistiksel Analiz Teknikleri: SPSS’te İşletme Yönetimi Uygulamaları, *Emek Ofset*, Ankara, 1-85 (2003).
4. Ünsal, A., Güler, H., ‘Türk Bankacılık Sektörünün Lojistik Regresyon Ve Diskriminant Analizi İle İncelenmesi’, *Gazi Üniversitesi, İ.İ.B.F. Dergisi*, 6:3-4 (2005).
5. Emin (Ergeç), S., ‘Çok Boyutlu Verilerin Bazı İstatistiksel Analiz Yöntemleri ve Uygulamaları’, Doktora Tezi, *Gazi Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, 51-130 (1984).
6. Şen, Şenkal, Güneş, “Bankalarda Mali Başarısızlık ve Türkiye’deki Mali Başarısızlığa Uğrayan Bankaların Kantitatif Yöntemler Yardımıyla Tahmini”, Doktora Tezi, *Marmara Üniversitesi*, İstanbul,15-20 (1998).
7. Kurtuluş, K., ‘Pazarlama Araştırmaları’, *İstanbul Üniversitesi Yayınları*, 54: 468 (1976).
8. Tatsuoka, M. M., ‘Multivariate Analysis: Techniques for Educational and Psychological Research’, *John Wiley & Sons, Inc.*, New York, 87-98 (1971).
9. Malhotra K.N., “Marketing Research An Applied Orientation”, Second Edition, *Prentice Hall International Edition*, 45-55 (1996).
10. Çakmak Z., “En İyi Ayırma Modelinin Belirlenmesinde Kullanılan Değişken Seçme Yöntemleri”, Anadolu Üniversitesi Kütahya İktisadi ve İdari Bilimler Fakültesi, *15. Yıl Ağarmanı*, Kütahya, 289-299 (1989).
11. Cooley, W. W., Lohnes P.R., ‘Multivariate Data Analysis’, *John Wiley & Sons, Inc.*, 246 (1978).
12. Çetin, E., İ., ‘Çok Değişkenli Analizlerin Pazarlama İle İlgili Araştırmalarda Kullanımı: 1995-2002 Arası Yazın Taraması’, *Akdeniz İ.İ.B.F. Dergisi*, 5:32-47, (2003).
13. Kiang, M.Y., “A comparative assessment of classification methods”, *Decision Support Systems*, 441– 454 (2003).

14. Ünsal, A., 'Diskriminant Analizi ve Uygulaması Üzerine Bir Örnek', *Gazi Üniversitesi, İ.İ.B.F. Dergisi*, 3: 19-36 (2000).
15. Tatlıdil, H., Özel, M., 'Firma Derecelendirme Çalışmaları Konusunda Çok Değişkenli İstatistiksel Analize Dayalı Karar Destek Sistemlerinin Kullanımı', *Bankacılar, Türkiye Banka Birliği*, ISSN 13-0217 54: 46-58 (2005).
16. Cankur, N., Ş., Coşkun, İ., Ediz, B., 'El Kullanımının Ayırma Fonksiyonu İle Saptanması', *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 28 (1): 5-7, (2002).
17. Tengilimoğlu, D., 'Kamu Ve Özel Sektör Örgütlerinde Liderlik Davranışı Özelliklerinin Belirlenmesine Yönelik Bir Alan Çalışması', *Elektronik Sosyal Bilimler Dergisi*, ISSN:1304-0278,4 (14): 10-16 (2005).
18. Güler, A., Kaftanoğlu, O., Bek, Y., Yeninar, H., 'Türkiye'deki Önemli Balarısı (Apis melliferaL.) Irk ve Ekotiplerinin Morfolojik Karakterler Açısından İlişkilerinin Diskriminant Analiz Yöntemiyle Saptanması', *Tr. J. of Veterinary and Animal Sciences*, 23: 337-343 (1999).
19. Özdamar, K., 'Paket Programlar İle İstatistiksel Veri Analizi-2', *Kaan Kitabevi*, Eskişehir, 341-394 (2002).
20. Anderson, T., W., 'An Introduction to Multivariate Statistical Analysis', *John Wiley & Sons, Inc.*, 126-153 (1974).
21. Lachenbruch, P., A., 'Discriminant Analysis', *A division of Macmillian Publishing Co.*, 2-41 (1975).

EKLER

EK-1. DİE 2001 yılı Büyükşehir belediyelerinin toplanan ve dağıtılan su miktarı, abone sayıları, gelirleri, çalışan sayıları, çalışanlarına ödediği ücretler, giderleri, işçisat toplamları, satın aldıkları elektriğin ücretleri, verimleri

Çizelge 1.1 Belediyelerinin toplanan ve dağıtılan su miktarı

Grup	Belediyeler	Toplanan Su Miktarı (m ³ /yıl)	Dağıtılan Su Miktarı (m ³ /yıl)
2	ADANA	95 699 733	41 194 661
1	ADAPAZARI	37 843 200	22 527 488
3	ANKARA	278 620 560	200 392 767
2	ANTALYA	73 360 985	26 945 488
2	BURSA	85 123 440	57 268 166
1	DİYARBAKIR	43 201 008	15 757 037
1	ERZURUM	38 454 000	17 962 438
1	ESKİŞEHİR	32 604 350	17 968 630
1	GAZİANTEP	92 587 120	34 687 255
2	İÇEL	98 485 760	20 404 870
3	İSTANBUL	631 281 142	416 603 502
3	İZMİR	216 256 726	97 423 293
1	KAYSERİ	38 898 670	22 139 647
1	KOCAELİ	101 337 820	21 829 933
2	KONYA	90 998 875	30 589 782
1	SAMSUN	54 750 000	20 695 840

Çizelge 1.2 Belediyelerin abone sayıları, gelirleri, çalışan sayıları

Belediyeler	Abone (Kişi)	Gelirler (Milyon TL)	Çalışan (Kişi)
ADANA	331 969	32 522 091	243
ADAPAZARI	61 889	4 827 688	60
ANKARA	1 039 446	179 090 831	3 433
ANTALYA	274 157	22 158 876	273
BURSA	363 031	38 073 693	751
DİYARBAKIR	102 598	10 553 528	174
ERZURUM	78 769	8 062 402	540
ESKİŞEHİR	180 565	24 413 011	487
GAZİANTEP	217 136	33 614 151	670
İÇEL	186 758	13 115 355	308
İSTANBUL	3 012 588	409 284 844	6 279
İZMİR	660 626	104 110 284	1 309
KAYSERİ	189 128	18 820 092	145
KOCAELİ	188 059	18 688 997	284
KONYA	200 370	38 151 289	365
SAMSUN	133 193	10 590 793	308

EK-1. DİE 2001 yılı Büyükşehir belediyelerinin toplanan ve dağıtılan su miktarı, abone sayıları, gelirleri, çalışan sayıları, çalışanlarına ödediği ücretler, giderleri, işçisaat toplamları, satın aldıkları elektriğin ücretleri, verimleri

Çizelge 1.3 Belediyelerin çalışanlarına ödediği ücretler, giderleri, işçisaat toplamları

Belediyeler	Ücret (Milyon TL)	Diğer Giderler (Milyon TL)	İşçisaat (Saat)
ADANA	3 218 845	765 001	384 144
ADAPAZARI	496 000	155 379	144 960
ANKARA	32 379 602	16 834 576	3 611 208
ANTALYA	2 382 537	1 107 313	224 688
BURSA	9 100 501	1 809 092	1 316 720
DIYARBAKIR	3 462 283	183 690	372 064
ERZURUM	3 690 115	545 741	719 968
ESKİŞEHİR	4 222 798	1 230 537	804 528
GAZİANTEP	5 624 920	1 519 705	1 087 200
İÇEL	1 871 228	4 525 914	333 408
İSTANBUL	66 250 146	44 708 747	12 382 000
İZMİR	11 251 025	2 214 088	1 541 408
KAYSERİ	876 168	2 353 207	330 992
KOCAELİ	5 523 608	1 975 982	546 016
KONYA	3 040 780	6 802 216	797 280
SAMSUN	1 312 091	534 744	514 608

EK-1. DİE 2001 yılı Büyükşehir belediyelerinin toplanan ve dağıtılan su miktarı, abone sayıları, gelirleri, çalışan sayıları, çalışanlarına ödediği ücretler, giderleri, işçisat toplamları, satın aldıkları elektriğin ücretleri, verimleri

Çizelge 1.4 Belediyelerin satın aldıkları elektriğin ücretleri, verimleri

Belediyeler	Satın Alınan Elektrik (Milyon TL)	Verim (saniye/litre)
ADANA	14 419 200	3 035
ADAPAZARI	642 695	1 200
ANKARA	8 802 134	8 835
ANTALYA	2 792 360	2 326
BURSA	851 276	2 699
DIYARBAKIR	1 686 345	1 370
ERZURUM	3 665 255	1 219
ESKİŞEHİR	1 354 998	1 034
GAZİANTEP	8 000 000	2 936
İÇEL	2 133 771	3 123
İSTANBUL	46 038 050	20 018
İZMİR	12 384 488	6 857
KAYSERİ	1 898 890	1 233
KOCAELİ	443 296	3 213
KONYA	2 394 658	2 886
SAMSUN	603 342	1 736

EK-2. Grup İstatistikleri

Çizelge 2.1 Gruplara ait ortalama ve standard sapmalar

GRUP	Değişkenler	Ortalama	Standard Sapma	Birim Sayısı	
1	TSM	54959521	26801372	8	
	DSM	21696034	5779066	8	
	GIDER	1062373	834812,9	8	
	ABONE	143917,1	57930,71	8	
	GELİR	16196333	9581693	8	
	ÇALIŞAN	333,5	213,2443	8	
	ÜCRET	3150998	2031583	8	
	SAAT	565042	298902,5	8	
	ELEKTRİK	2286853	2531394	8	
	VERİM	1742,625	849,8346	8	
	2	TSM	88733759	9976393	5
		DSM	35280593	14415221	5
		GIDER	3001907	2587518	5
ABONE		271257	77913,21	5	
GELİR		28804261	10924908	5	
ÇALIŞAN		388	207,9351	5	
ÜCRET		3922778	2943784	5	
SAAT		611248	450138,5	5	
ELEKTRİK		4518253	5582329	5	
VERİM		2813,8	316,5734	5	
3	TSM	375386143	223794548,5	3	
	DSM	238139854	162903755	3	
	GIDER	21252470,3	21589056,34	3	
	ABONE	1570886,67	1262835,381	3	
	GELİR	230828653	159029807,5	3	
	ÇALIŞAN	3673,6666	2493,7251	3	
	ÜCRET	36626924,3	27744470,4	3	
	SAAT	5844872	5755132,481	3	
	ELEKTRİK	22408224	20542269,17	3	
	VERİM	11903,3333	7096,7585	3	

EK-3. Belediyelerin analiz öncesi ve analiz sonrası yer aldığı gruplar ile ayrıca faktör değerleri

Çizelge 3.1 Belediyelere ilişkin faktör değerleri

Belediyeler	Analiz Öncesi	Analiz Sonrası	Faktör 1	Faktör 2
ESKİŞEHİR	1	1	3,9772	-1,5521
ADAPAZARI	1	1	4,0311	-0,1715
ERZURUM	1	1	5,5126	-1,6714
KAYSERİ	1	1	3,8515	0,8269
KOCAELİ	1	1	3,3149	0,3252
SAMSUN	1	1	2,8818	-0,4071
GAZİANTEP	1	1	1,7925	-1,7372
DİYARBAKIR	1	1	5,2443	-0,7470
İÇEL	2	2	3,3091	2,2826
ANTALYA	2	2	1,3505	1,9229
ADANA	2	2	2,0621	1,8958
BURSA	2	1	2,2533	-0,6055
KONYA	2	2	2,1936	0,1442
İZMİR	3	3	-13,2156	-0,2105
ANKARA	3	3	-14,4635	-0,1730
İSTANBUL	3	3	-14,096	-0,1222

EK-4. Balıkesir belediyesinin 2001 yılı su istatistikleri

Çizelge 4.1 Balıkesir belediyesinin deęişken deęerleri

Deęişkenler	Deęerler
TSM	15 768 000
DSM	6 710 102
GIDER	44 945
ABONE	68 597
GELİR	5 890 307
ÇALIŞAN	103
ÜCRET	1 132 921
SAAT	108 160
ELEKTRİK	1 408 043
VERİM	560

EK-5 Mahalanobis Uzaklık Değerleri

Çizelge 5.1 Grupların mahalanobis uzaklık değerleri

Belediyeler	Gruplar	G_1:1	G_2:2	G_3:3
ADANA	G_2:2	19,1139	9,2448	248,0931
ADAPAZARI	G_1:1	7,0014	11,7875	314,6049
ANKARA	G_3:3	332,0627	274,8375	9,2886
ANTALYA	G_2:2	17,0785	4,7331	222,4790
BURSA	G_2:2	8,9581	9,4148	255,1878
DİYARBAKIR	G_1:1	3,6725	14,5727	345,1079
ERZURUM	G_1:1	7,5409	22,9717	364,9525
ESKİŞEHİR	G_1:1	5,9664	15,5938	299,9625
GAZİANTEP	G_1:1	8,6129	11,8613	241,5404
İÇEL	G_2:2	14,8971	7,9103	295,5712
İSTANBUL	G_3:3	316,7777	260,6428	9,2974
İZMİR	G_3:3	284,3599	231,5556	9,2812
KAYSERİ	G_1:1	5,3766	5,5533	294,8352
KOCAELİ	G_1:1	10,0791	10,5358	290,8999
KONYA	G_2:2	12,9597	10,2819	252,2017
SAMSUN	G_1:1	4,9230	6,7835	275,2576

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : ÜNAL, Mert
Uyruğu : T.C.
Doğum tarihi ve yeri : 17.01.1978 Ankara
Medeni hali : Bekar
Telefon : 0 (312) 212 11 79
e-mail : munal@mail.com

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	Gazi Üniversitesi /İstatistik Bölümü	2006
Lisans	Gazi Üniversitesi/ İstatistik Bölümü	2001
Lise	Cumhuriyet Lisesi	1996

İş Deneyimi

Yıl	Yer	Görev
2004-2006	Ankara Üniversitesi Geliştirme Vakfı Okulları Özel İlköğretim Okulu	Bilgisayar Usta Öğreticisi
2002-2004	TÜİK	Sözleşmeli Personel

Yabancı Dil

İngilizce

Hobiler

Sinema, Kitap, Bilgisayar teknolojileri, Basketbol