T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ

AYIRMA ANALİZİ ÜZERİNE BİR İNCELEME

YÜKSEK LİSANS TEZİ

DANIŞMAN
Prof. Dr. Mehmet GÜNGÖR

HAZIRLAYAN
Koray ÇİFTÇİ

MALATYA-2019

T.C.

İNÖNÜ ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

AYIRMA ANALİZİ ÜZERİNE BİR İNCELEME

Hazırlayan

Koray ÇİFTÇİ

Danışman

Prof. Dr. Mehmet GÜNGÖR

MALATYA-2019

T.C.

İNÖNÜ ÜNİVERSİTESİ

SOSYAL BİLİMLERİ ENSTİTÜSÜ

AYIRMA ANALİZİ ÜZERİNE BİR İNCELEME

YÜKSEK LİSANS

DANIŞMAN	HAZIRLAYAN
Prof. Dr. Mehmet GÜNGÖR	**Koray ÇİFTÇİ**

Jürimiz 25.07.2019 tarihinde yapılan savunma sınavı sonucunda bu yüksek lisans tezini (oybirliği /oyçokluğu) ile başarılı bulunarak Ekonometri Anabilim dalında yüksek lisans tezi olarak kabul edilmiştir.

Jüri Üyelerinin Unvan Ad Soyadı imzası

1. Prof. Dr. Mehmet Güngör .....
2. Doç. Dr. Yunus BULUT .....
3. Dr. Öğr. Üyesi Fahrettin ÖZBEY.....
4. .....
5. .....

İnönü Üniversitesi Sosyal Bilimler Enstitüsü Yönetim Kurulunun ..... tarih ve .....sayılı kararıyla bu tezin kabulü onaylanmıştır.

Prof. Dr. Mehmet KUBAT

Sosyal Bilimler Enstitüsü Müdürü

ONUR SÖZÜ

"Prof. Dr. Mehmet GÜNGÖR'ün danışmanlığında yüksek lisans tezi olarak hazırladığım **AYIRMA ANALİZİ 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."

Koray ÇİFTÇİ

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kağıt ve elektronik kopyalarının İnönü Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- ✓ Tezimin tamamı her yerden erişime açılabilir.
- ✓ Tezim sadece İnönü Üniversitesi yerleşkelerinden erişime açılabilir.
- ✓ Tezimin … yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

Koray ÇİFTÇİ

TEŞEKKÜR

Yüksek lisans eğitimim süresince ilgisi ve desteği ile yanımda olan, ayırma analizi konusunda detaylı bir araştırmaya yönlendirerek, konuya hakim olmamı sağlayan saygıdeğer hocam Tez Danışmanım Prof. Dr. Mehmet Güngör'e,

Tezimin her aşamasında engin bilgi ve tecrübesiyle desteklerini esirgemeyen, her zaman yanımda hissettiğim değerli abim, hocam Doç. Dr. Necati ÇİFTÇİ' ye,

Bana her şeyden önemli olan manevi eğitim ve terbiyeyi veren, hayatımı kazanmam için gerekli özeni gösteren anneme ve babama,

Tezin hazırlanması döneminde ilgi ve desteğimi vermede eksiklikler yaşadığım, bu konuda sabır ve sevgi ile bekleyen biricik oğlum Ömür Tuna ÇİFTÇi' ye,

Teşekkürlerimi sunarım.

ÖZET

Koray ÇİFTÇİ, Ayırma Analizi Üzerine Bir İnceleme,

Yüksek Lisans Tezi, Malatya, 2019

İstatistiğin temelinde belirlenen bir araştırma konusu hakkında örnek veriler elde ederek bu veriler üzerine geliştirilmiş birçok istatistiksel analiz uygulayarak analiz sonuçlarına göre konu hakkında genelleme yapmaktır. Bu sayede seçilmiş örneklem üzerindeki çalışma ile ele alınan konunun bütün evreni hakkında tahmin yürütülebilir. Ancak bilimsel araştırmalarda üzerinde çalışılan konunun analizinde, tek değişkenli analizlerin yeterli olmadığı görülmektedir. Bu kısıtlılık nedeniyle çok değişkenli istatistiksel analiz yöntemleri geliştirilmiştir. Bu yöntemlerden biri de ayırma analizidir.

Bu çalışmanın birinci bölümünde çok değişkenli analiz yöntemleri tanıtılmış. Bu yöntemlerin temelini teşkil eden çeşitli varsayımalar açıklanmıştır.

İkinci bölümde; ayırma analizinin, tanımı, amacı, dayandığı varsayımlar, iki ve daha fazla grup olması halinde ayırma analizi, ayırma fonksiyonlarının elde edilişi ve anlamı ve ayırma analizinin teorik alt yapısı incelenmiştir.

Üçüncü bölümde, Birleşmiş Milletler Kalkınma Programı tarafından yayınlanan 2018 İnsani Gelişme Endeksi verileri kullanılarak ülkeleri çok yüksek, yüksek, orta ve düşük insani gelişmiş ülkeler gruplarına ayırmak için ayırma analizi SPSS 22 paket programı kullanılarak uygulanmış. Seçilmiş bir ülke, analiz sonrası elde edilen fonksiyonlar sayesinde doğru sınıflandırılıp sınıflandırılmadığı test edilmiştir.

Anahtar Kelimeler: Ayırma Analizi, SPSS, İnsani Gelişme Endeksi

ABSTRACT

Koray ÇİFTÇİ, An Investigation on Discriminant Analysis

Master's Thesis, Malatya, 2019

By obtaining sample data about a research topic on the basis of statistics, generalization is made according to the results of the analysis by applying a number of statistical analyzes developed on these data. In this way, it can predict the whole universe of the subject covered by the study on the selected sample. However, it is seen that unvariate analyzes are not sufficient in the analysis of the subject studied in scientific research. Because of this limitation, multivariate statistical analysis methods have been developed. One of these methods is discriminant analysis.

In the first part of this study, multivariate analysis methods are introduced. The various assumptions underlying these methods are explained.

In the second part; discriminant analysis, definition, aim, basis of assumptions, discriminant analysis in case of two or more groups, the acquisition and meaning of separation functions and theoretical background of discriminant analysis are examined.

In the third part, discriminant analysis was applied by using SPSS 22 package program to divide countries into very high, middle and low human developed countries by using the 2018 Human Development Index data published by United Nations Development Program. A selected country was tested for correct classification by means of the functions obtained after analysis.

Key words: Discriminant Analysis, SPSS, Human Development Index

İÇİNDEKİLER

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

KISALTMALAR LİSTESİ

BM = Birleşmiş Milletler

CI = Condition İndex

Cov = Kovaryans

DW = Durbin Watson

GAV = Gruplar Arası Varyans

GİV = Grup İçi Varyans

GSMG = Gayri Safi Milli Gelir

GSMH = Gayri Safi Milli Hasıla

GSYİH = Gayri Safi Yurt İçi Hasıla

HDI = Human Development Index

HDR = Human Development Report

İGE = İnsani Gelişme Endeksi

İGR = İnsani Gelişme Raporu

Max = Maksimum

Min = Minimum

SGP = Satın Alma Gücü Paritesi

UNDP = United Nations Development Programme

Var = Varyans

VIF=Variance Increase Factors

BÖLÜM 1

1. ÇOK DEĞİŞKENLİ İSTATİSTİK TEKNİKLERİ

1.1. Çok değişkenli istatistik nedir?

Bilimsel çalışmalarda karşılaşılan problemlerin en önemlilerinden biri, araştırmalar sonucu elde edilen birden çok özellik arasında herhangi bir ilişkinin varlığını incelemede tek değişkenli istatistik tekniklerinin yetersiz kalmasıdır. Nitekim gerçek hayatta karşılaşılan olaylar birçok etkenin etkisi altındadır. Bir sonuca ait özelliği, birçok özellik etkilemekle birlikte etkileyen bu özellikler arasında da bir ilişki vardır. Bilimsel çalışmaların sonucunun güvenilir ve sağlıklı olması için sonucu etkileyen özelliklerin olabildiğince fazlası dikkate alınmalıdır. Tek değişkenli istatistik tekniklerinin bu konudaki sınırlılıkları nedeni ile özellikle 20. Yüzyılın ilk çeyreğinden itibaren çok değişkenli istatistik teknikleri geliştirilmeye başlanmıştır. Geliştirilen bu teknikler sayesinde tek değişkenli istatistiksel tekniklerde varsayılan sınırlılıklar ortadan kalkacağından araştırma sonuçlarının geçerliliği ve güvenilirliği artacaktır. Nitekim geliştirilen çok değişkenli istatistik teknikleri araştırma konusunu olabildiğince tüm yönleriyle ele aldığından araştırma sonuçları daha sağlıklı ve doğru olmaktadır. Bu nedenlerle araştırma sonucunun daha güvenilir ve geçerli olmasını isteyen araştırmacı çok değişkenli veri setleri ve bunların analizi ile karşı karşıya kalmaktadır.

Çok değişkenli istatistik tekniklerinde n tane nesneye ilişkin p tane değişken incelenmektedir. Nesne ve değişken sayılarının çok olması durumunda, gösterim kolaylığı sağlaması için vektör ve matrislerden yararlanılmaktadır. Bu değişkenlerin birbiriyle ilişkili ve çok sayıda olması istatistiksel analiz tekniklerinde sorun yaratmaktadır. Ayrıca, çok sayıda değişkenle çalışmak, işlem yükünü arttıracağı ve elde edilecek sonuçların yorumunda bazı güçlüklere neden olacağı için istenen bir durum değildir. Bilgisayar programlarının ve de istatistik paket programlarının çok geliştiği günümüzde işlem yükü çok sorun görülmese de, çok sayıda değişkene ilişkin analiz sonuçlarının yorumlanması ve özetlenmesi gerçekten zor olabilmektedir. Böyle durumlarda, çok değişkenli istatistik yöntemlerine başvurulmaktadır (Tatlıdil,1992: 7).

Tek değişkenli ve çok değişkenli istatistikleri değişken sayısı ve varsayımları açısından karşılaştıran bir tablo aşağıda verilmiştir.

Tablo 1. Tek Değişkenli ve Çok Değişkenli İstatistiklerin Karşılaştırılması

	Tek değişkenli istatistik	**Çok değişkenli İstatistik**
Değişken sayısı (P)	P=1	P≥2
Varsayımı	Değişkeni etkileyen diğer faktörler türdeş ya da sabittir.	N birimli p tane değişken normal dağılım gösterir.

Bir bilimsel araştırmada incelenen olayın analizinde tek değişkenli istatistiğin yeterli olmayacağı acıktır. Çünkü tek değişkenli yöntemler kısıtlayıcı varsayımlar altında geçerlidir. En önemli kısıtlayıcı ise olaydaki birçok faktörün deneysel olarak kontrol altında tutulması ve her defasında tek bir faktörün etkisinin incelenmesidir. Oysaki, çok değişkenli istatistikte bazı kontrollü denemeler dışında böyle bir kısıtlayıcıdan ya da özellikten söz etmek olanaklı değildir. Çünkü çok değişkenli istatistikler genellikle kendi doğal çevrelerinde birçok yönleriyle gözlenirler. Bu durum, gözlenen özellikler arasındaki bağımlılık yapısının incelenmesini ön plana çıkarmaktadır. Bir başka anlatım ile çok değişkenli istatistik; inceleme konusu olayı bir bütün olarak ele almakta ve bütünlüğü sağlayan değişkenlerin bağımlılık yapısını açıklamaya çalışmaktadır. Bu durumda çok değişkenli istatistiğin en önemli amacının değişkenler arasındaki bağımlılık yapısının analizi olduğu iddia edilebilir (Tatlıdil, 1992: 11).

Çok değişkenli istatistik teknikleri, birden çok özelliğin analizi ile ilgilendiğinden uygulamalarda değişik amaçlarla kullanılabilmektedir. Bu amaçları şöyle sıralayabiliriz:

a) Basitleştirme ve boyut indirgeme: Çok sayıda değişken tarafından belirlenen bir sistem daha az sayıda gerçek ya da yapay değişkenlerle temsil edilmeye, böylece basitlik sağlanmaya çalışılır. Yani veri setinin varyasyonunu açıklayan ve aralarında ilişki bulunmayan daha az sayıda değişkenle veri yapısını açıklamaya çalışmaktır.

b) Birimlerin sınıflandırılması: Gözlenen birimlerin değişik sınıflar oluşturup oluşturmadıkları belirlenmeye çalışılır. Sınıflandırmanın diğer bir biçimi de birimlerin

önceden tanımlanmış sınıflardan hangilerine ait olduklarının belirlenmesidir. Bazı durumlarda birimler yerine değişkenlerin gruplandırılması söz konusudur.

c) Bağımlılık yapısının incelenmesi: Değişkenlerin kovaryans ve korelasyonlarından yararlanılarak bağımlılığın kaynakları ve sonuçları incelenir. Bir ya da daha fazla bağımlı değişkenin olduğu, diğerlerinin bağımsız olduğu regresyon analizinin amacı değişkenler arasındaki bağımlılık yapısını ortaya çıkarmaktır.

d) Hipotez testleri ve hipotez oluşturma: Çok değişkenli verilerde kullanılan hipotez testlerinden bazıları tek değişkenli teorinin çok değişkenli duruma genelleştirilmiş uzantılarıdır. Bunların dışında bazı çok değişkenli analiz yöntemleri de hipotez testlerinde kullanılmaktadır.

e) Sıralama ve ölçekleme: Birimlerin belli ölçülere göre sıralanması bazı uygulamaların temel amacıdır. Ölçekleme ise çok sayıda değişkenden yararlanarak birimlerin daha az boyutla gösterilmesidir.

Çok değişkenli istatistiksel teknikler; Ziraat, Tıp, Ekonomi, Biyoloji, Kimya, Jeoloji, Sosyoloji, Arkeoloji, Eğitim gibi birçok bilim dalında kullanılmaktadır. Bu denli geniş bir uygulama alanı olan çok değişkenli istatistiksel analiz tekniklerinin bazı varsayımlarla kısıtlı olduğunu söylemekte yarar vardır. Bu varsayımlardan en önemlisi, kullanılacak verilerin çok değişkenli normal dağılımlı kitleden çekilmiş olduğu biçimindeki varsayımdır (Tatlıdil, 1992: 12).

Aşağıda çok değişkenli istatistik tekniklerinin varsayımları detaylı bir şekilde sunulmuştur.

1.2. Çok değişkenli istatistik tekniklerin varsayımları

1.2.1. Normallik varsayımı: Çok değişkenli istatistiksel analiz tekniklerinin büyük çoğunluğunda örneklemlerin çok değişkenli normal dağılımlı kitlelerden geldiği kabul edilmektedir. Bu varsayım, bazı işlemleri ve sonuçların yorumlanmasını kolaylaştırmanın ötesinde, dağılım teorisi açısından da gereklidir. Çünkü; kanonik korelasyon, ayırma (diskriminant) analizi, çok değişkenli varyans analizi gibi yöntemler tamamiyle normallik varsayımı üzerine kurulmuştur (Tatlıdil, 1992: 53). Çok değişkenli tekniklerin en temel varsayımlarından biri olan normallik varsayımından sapmalar

anlamlı ise, t ve F istatistiklerin hesaplanmasında bu varsayım gerekli olduğundan, elde edilen testler geçerliliğini yitirmektedir. Tek değişkenli ve çok değişkenli teknikler tek değişkenli normallik varsayımına; çok değişkenli teknikler ayrıca çoklu normallik varsayımına dayanmaktadır (Kalaycı vd., 2010: 209). İçerisinde ayırma analizinin de bulunduğu birçok istatistiksel teknikte olduğu gibi normallik varsayımından sapma sınıflandırma oranını ve istatistik testlerin gücünü etkilemektedir.

Tek değişkenli normal dağılım gösteren veri setleri kolaylıkla test edilebilmekte ve normallik varsayımına uymayan değişkenlerin dağılımını normal dağılıma dönüştürmek için birkaç dönüşüm uygulanabilmektedir. Çoklu normal dağılım ise her bir değişkenin tek değişkenli normal dağılıma uyduğunu ve ilgili değişkenlerin kombinasyonlarının da normal olduğunu varsaymaktadır. Yani, bir değişken çoklu normal dağılıma uyuyorsa aynı zamanda tek değişkenli normal dağılıma da uyuyor demektir. Ancak, bunun tersi her zaman söylenemez. Bu nedenle, tüm değişkenlerin tek değişkenli normal dağılım göstermesi, garanti olmasa da, çok değişkenli normal dağılımı göstermeye yardım edecektir (Kalaycı vd., 2010: 209).

Sonuç olarak; gerçek yaşam durumlarının çoğu normal dağılım göstermektedir. Matematiksel olarak incelenmesi, çözümlenmesi ve elde edilen sonuçların yorumlanması bakımından kolay olan çok değişkenli normal dağılım varsayımı çoğu problemin incelenmesinde oldukça tutarlı bir yaklaşımdır.

1.2.1.1. Tek değişkenli normal dağılım: Tek değişkenli normal dağılımı tartışmanın iki amacı vardır. Birincisi, çok değişkenli normal dağılım testinin daha zor ve karmaşık olması ve bu testi sağlamanın tek değişkenli testlere bağlı olmasıdır. İkincisi, çok nadir de olsa, tüm marjinal dağılımlar normal dağılıma uysa da çoklu normal dağılım sağlanamamasıdır. Tek değişkenli normallik varsayımı sağlandığı halde çoklu normal dağılımın sağlanamadığı durumlara nadir de olsa rastlanmaktadır (Kalaycı vd., 2010: 209).

Sürekli bir X değişkeninin ortalaması μ, varyansı σ^2 ve normal dağılıyorsa dağılımın olasılık yoğunluk fonksiyonu şöyledir

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2} \qquad -\infty < x < \infty \text{ ve } \sigma^2 > 0 \qquad (1.1)$$

Normal dağılmış bir değişken şöyle gösterilir:

$$X \sim N(\mu,\sigma^2)$$

Olasılık yoğunluk fonksiyonundaki ortalama ve standart sapma değerlerini μ=0 ve σ=1 olarak seçersek dağılımımız *Standart Normal Olasılık Dağılımı* olarak adlandırılır. Ortalaması μ ve standart sapması σ olan bir değişken normal dağılmış ise bunu standart normal dağılıma dönüştürmek için *X* değişkenini, standartlaştırılmış normal değişken *Z'* ye çevirmek gerekir. Bunun için aşağıdaki eşitlik kullanılır:

$$Z = \frac{x-\mu}{\sigma} \qquad (1.2)$$

(1.2) de verilen eşitliği (1.1) eşitliğinde yerine yazarsak şunu elde ederiz:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}Z^2\right)} \qquad (1.3)$$

Bu yol izlenirse,

$$X \sim N(0,1)$$

X' in ortalaması 0 ve varyansı 1 olan normal dağılıma uyduğu anlamına gelir. Bir başka ifade ile standartlaştırılmış bir *Z* değişkenidir.

Normal dağılım fonksiyonunun grafiği çan şeklindedir. Şekil 1' de ortalaması μ, varyansı σ^2 olan ve normal dağılan bir *X* değişkeninin grafiği verilmiştir. Bu eğrinin altında kalan ve ortalamadan ±1, ±2 ve ±3 standart sapmaya sahip yaklaşık alanlarda gösterilmiştir. Bu alanlar olasılıkları gösterir ve *X* değişkeni için

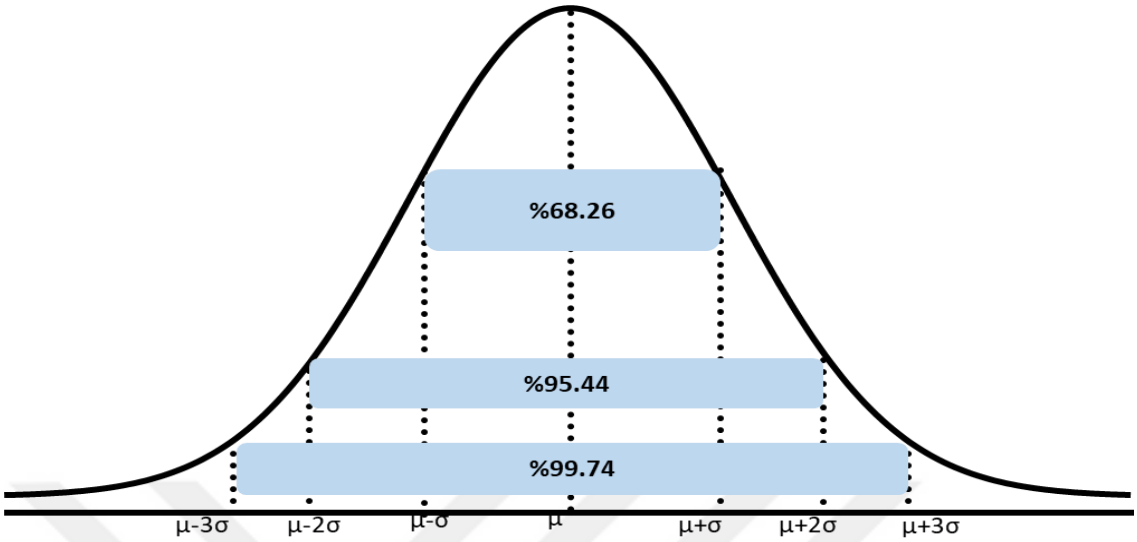
$$P(\mu-\sigma<X<\mu+\sigma)=0,6826$$

$$P(\mu-2\sigma<X<\mu+2\sigma)=0,9544$$

$$P(\mu-3\sigma<X<\mu+3\sigma)=0,9974$$

yazılabilir.

Şekil 1. Ortalaması μ ve varyansı σ^2 olan bir normal yoğunluk ve eğri altından kalan seçilmiş alanlar



Bu şu anlama gelir; *X* sürekli rassal değişkenindeki veriler %68,26 olasılıkla ortalamadan bir standart sapma aralığında, %95,44 olasılıkla ortalamadan iki standart sapma aralığında ve %99,74 olasılıkla ortalamadan üç standart sapma aralığındadır.

1.2.1.2. Çok değişkenli normal dağılım: Çok değişkenli normal yoğunluk, tek değişkenli normal yoğunluğun *p≥2* boyuta genelleştirilmesidir.

X=$[X_1, X_2, \dots, X_P]$ rassal vektörü için p-boyutlu bir normal yoğunluk fonksiyonu, i=1,2,…,p ve $-\infty < x_i < \infty$ olmak üzere

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\sum|^{\frac{1}{2}}} e^{-(x-\mu)'\sum^{-1}(x-\mu)/2} \qquad (1.4)$$

şeklindedir.

Burada ∑ matrisi X' in varyans-kovaryans matrisidir.

1.2.2. Kovaryans matrislerinin eşitliği varsayım: Farklı varyans problemi genelde normallik problemlerinden kaynaklandığı gibi değişkenlerin dağılımından da kaynaklanabilmektedir. Farklı varyans durumunda regresyon analizinin hataları incelendiğinde hataların dağılımı bazen bir koniye benzemektedir. Bu durumda koni eğer sağa açılıyorsa değişkenin tersi sola açılıyorsa değişkenin karekökü alınmalıdır. Bazı dönüşümler ise verilerin türüne bağlıdır (Hair vd., 2014: 77).

Tek değişkenli analiz durumunda kovaryans matrisi sabit bir sayı ve bağımlı değişkenin varyansı bütün hücreler için eşit olduğu zaman varsayım sağlanmaktadır. Ancak ayırma (diskriminant) analizinde kovaryans matrislerinin tüm hücreleri eşit olduğu zaman varsayım sağlanabilmektedir. Örneğin, kovaryans matrisinde üç varyans $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ ve kovaryans $(\sigma_{12}, \sigma_{13}, \sigma_{23})$ bulunan üç bağımlı değişkeni ele alalım. Kovaryans matrisleri eşitliği varsayımının sağlanabilmesi için matrislerdeki ilgili altı öğenin eşit olması gerekir. Bu nedenle ayıra analizinde kovaryans matrislerinin eşitliği varsayımından sapma olasılığı ANOVA analizinden daha yüksektir (Sharma vd., 1996: 383).

Ayırma analizi gibi birçok istatistik teknikler kovaryans matrislerinin eşitliği varsayımında bulunmaktadır. Bu amaçla ilk kez 1949 yılında Box tarafından M istatistiği tanıtılmıştır. Box-M istatistiği tek değişkenli (p=1) eş varyans testi olan Barlett-Box F testinin genelleştirilmesi ile elde edilmiştir. Diğer bir anlatımla p=1 için hesaplanacak Box-M istatistiği Barlett-Box F istatistiğine eşittir. Bunun için p sayıdaki değişken için ölçülen g sayıda grubun olduğunu ve her gruptaki birim sayısının n_j ile gösterildiğini varsayalım. Ayrıca grup içi kovaryanslar S_j ile gösterilsin. Böylece Box-M, Barlett-Box F istatistiğinin genelleştirilmesiyle aşağıdaki gibi hesaplanmaktadır (Kalaycı vd., 2010: 217).

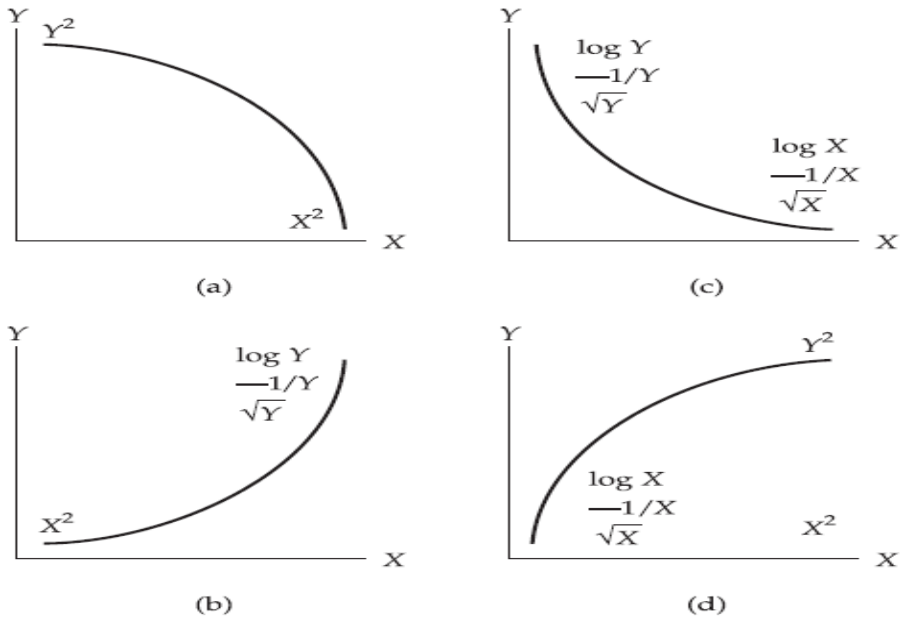
$$Box - M = (N - g)Ln|S_w| - \sum_{j=1}^{g}(n_j - 1)Ln(S_j) \qquad (1.5)$$

Kovaryans matrislerinin eşitliğinin test edilmesinde kullanılan tüm istatistikler normallik varsayımına duyarlıdır. Anlamlı Box-M istatistiği eşit olmayan kovaryans matrislerini veya normallikten sapmayı veya her ikisini gösterir. Bu nedenle Box-M istatistiğini kullanmadan önce çoklu normal dağılımın sağlanması gerekir (Kalaycı vd., 2010: 218).

1.2.3. Doğrusallık varsayımı: Korelasyon katsayılarına dayanan aralarında ayırma analizinin de bulunduğu birçok istatistik tekniklerin örtülü varsayımlarından birisi de doğrusallık varsayımıdır. Doğrusal olmayan etkileşimler için hesaplanacak doğrusal korelasyonlar gerçek ilişkiyi her zaman daha düşük gösterecektir (Kalaycı vd., 2010: 220).

İki değişken arasındaki doğrusallığı sağlamada çok sayıda dönüşüm uygulanabilmekte, ancak doğrusal olmayan ilişkiler dört grup altında toplanmaktadır. Her grafik kuadratik fonksiyonun hem bağımlı hem de bağımsız değişkenlere uygulanan dönüşümleri göstermektedir.

Şekil 2. Doğrusallık varsayımı için dönüşüm seçimi (Hair vd., 2014, 76).



Örneğin; değişkenler arası ilişki Şekil 2-a da olduğu gibi ise, her iki değişkenin karesi alınarak ilişki doğrusallaştırılabilir. Değişkenler arası ilişki Şekil 2-b de olduğu gibi ise, iki değişken arasındaki ilişkiyi doğrusallaştırmak için bağımsız değişkenin karesi alınıp bağımlı değişkene sırasıyla LogY, -1/Y ve $\sqrt{Y}$ dönüşümleri uygulanır ve bunlardan doğrusallığı sağlayan dönüşüm seçilir. Değişkenler arası ilişki Şekil 2-c de olduğu gibi ise, doğrusallığı sağlayabilmek için 9 dönüşüm seçeneği söz konusu olmaktadır. Bu dönüşümler doğrusallık sağlanana kadar sırasıyla denenerek gerçekleştirilir. Çoklu dönüşüm seçenekleri söz konusu olduğunda doğrusallık sağlanana kadar en üstteki dönüşümden aşağıya doğru hareket edilerek dönüşümler gerçekleştirilmelidir (Kalaycı vd., 2010: 221).

1.2.3.1. Logaritmik dönüşüm: Bu dönüşüm negatif sayıların logaritması alınamayacağı için değerleri pozitif olan değişkenlere uygulanabilmektedir. Ancak tüm

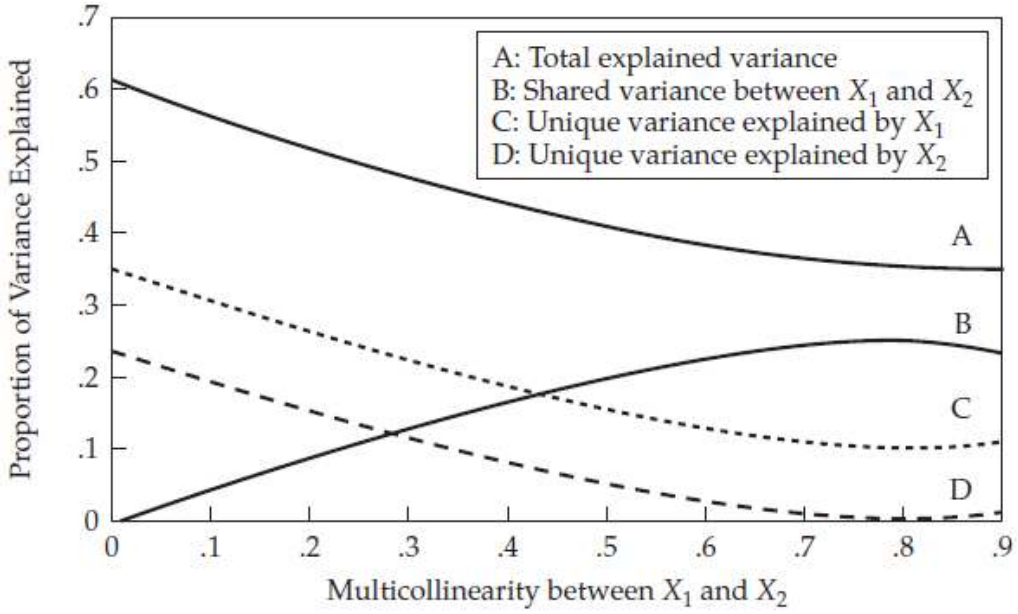
negatif değerleri pozitif yapacak bir sabit sayı değişken değerlerine ilave edilerek bu sorunun üstesinden gelinebilmektedir. Bağımlı değişken bağımsız değişkenler ile sürekli artan bir eğimi gösteriyorsa modeli doğrusallaştırmak için kullanılmaktadır (Kalaycı vd., 2010: 221).

1.2.3.2. Kare dönüşümü: Değişkenler arasında Şekil 2-a'daki gibi aşağıya doğru dışbükey bir ilişki olması durumunda ilişkiyi doğrusallaştırmak için kullanılır.

1.2.4. Çoklu doğrusal bağlantı varsayımı: Bağımsız değişkenler arasındaki ilişki çoklu doğrusal bağlantı olarak tanımlanmaktadır. İki değişken arasındaki ilişki +1 ise aynı, -1 ise zıt yönlü tam bir bağımlılık; sıfıra eşitse tam bir bağımsızlık söz konusudur (Kalaycı vd., 2010: 222). Eğer bir modelde herhangi bir bağımsız değişken diğer bir bağımsız değişken veya değişkenler tarafından tam olarak açıklanıyorsa bu durumda diğer değişkenler tarafından açıklanan değişken gereksiz olacağından modelden çıkartılmalıdır.

Çoklu doğrusal bağlantı, bir bağımsız değişkenin diğer bağımsız değişkenlerle olan ilişkisinin derecesine göre bağımsız değişkenin tahmin gücünü azaltır. Çoklu doğrusal bağlantı arttıkça bağımsız değişken tarafından açıklanan spesifik varyans azalmakta, ortak varyans yüzdesi ise artmaktadır (Kalaycı vd., 2010: 223). Dolayısıyla modele yüksek çoklu doğrusal bağlantılı değişkenler alındıkça modelin genel tahmin gücü daha artmaktadır. Hair vd. çoklu doğrusal bağlantıyı da açıkladıkları eserlerinde iki bağımsız değişken arasındaki çoklu doğrusal bağlantının derecesine göre spesifik ve ortak varyansı aşağıdaki grafikteki gibi göstermişlerdir.

Şekil 3. Çoklu Doğrusal bağlantının Derecesine Göre Spesifik ve Ortak Varyans (Hair vd., 2014: 198).



Şekil 3'te çeşitli çoklu doğrusal bağlantı düzeylerine göre X_1 ve X_2 bağımsız değişkenleri arasındaki korelasyon sıfır ise yani çoklu doğrusal bağlantı yoksa değişkenlerin her biri bağımlı değişkeni sırasıyla %35 ve %25 oranındaki değişimleri açıkladığı söylenebilir. Bağımsız değişkenler arsındaki çoklu doğrusal bağlantı düzeyi arttıkça bağımlı değişkenin açıklanan toplam varyansı azalmaktadır.

1.2.4.1. Çoklu doğrusal bağlantı probleminin doğurduğu sonuçlar: Tam çoklu doğrusal bağlantının varlığı halinde, katsayılar tanımsız ve bu katsayıların standart hataları sonsuz olmaktadır.

Çoklu bağlantı halinde katsayıların varyans ve kovaryansları artmaktadır.

Bunun sonucunda veya çok sayıda bağımsız değişkenden dolayı modelin çoklu korelasyon katsayısı (R^2) yüksek, ancak bağımsız değişkenlerden hiçbiri çok azı anlamlı çıkmaktadır.

Çoklu bağlantının varlığı halinde bağımsız değişkenlerden bazılarının modele alınmaması gerekebilir. Fakat hangi değişkenlerin modelden çıkartılacağına karar

vermek için kullanabileceğimiz basit kurallar bulunmamaktadır (Kalaycı vd., 2010: 224).

1.2.4.2. Çoklu doğrusal bağlantının saptanması: Çoklu doğrusal bağlantıyı açıklayıp sonuçlarını tartıştıktan sonra bu durumu nasıl tespit edeceğimizi inceleyelim.

İlk olarak bağımsız değişkenler arasındaki korelasyon matrisini inceleyebiliriz. İki bağımsız değişken arasındaki basit korelasyon katsayısı anlamlı ise çoklu doğrusal bağlantıya yol açabilir. Bu durum her zaman geçerli olmayabilir. Basit korelasyon katsayısı (r) çoklu korelasyon katsayısından (R) küçük ise (r<R) çoklu bağlantı problemine neden olmaya bilir.

Bir diğer yaklaşım, modele bağımsız değişkenler ilave edildikçe, R^2 katsayısındaki değişimleri incelemektir. R^2'de önemli bir gelişme sağlanamazsa bu durum çoklu doğrusal bağlantıya işaret edebilir.

Çoklu doğrusal bağlantıyı tespit etmekte kullanılan diğer bir yaklaşım kısmi korelasyon katsayılarının incelenmesidir. İki değişken arasındaki basit korelasyon katsayıları anlamlı olduğu halde kısmi korelasyon katsayıları anlamsız çıkıyorsa, bu durum çoklu bağlantı için bir işaret olabilir.

Diğer önemli bir yaklaşım varyans artış faktörleridir(VIF). VIF değerlerinin hesaplanmasını göstermek için aşağıdaki gibi üç bağımsız değişkeni içeren modeli ele alalım:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e \tag{1.6}$$

Üç bağımsız değişkenli bir modelin VIF değerleri şu şekilde hesaplanmaktadır.

İlk olarak X_1 bağımsız değişkeni bağımlı değişken olarak alınıp diğer bağımsız değişkenler arasındaki çoklu korelasyon katsayısı (R^2) hesaplanır. Böylece X_1 bağımsız değişkeni için varyans artış faktörü;

$$\text{VIF}(X_1) = \frac{1}{1-R_1^2} \tag{1.7}$$

olarak hesaplanmaktadır.

İkinci olarak, X_2 değişkenini bağımlı değişken olarak alıp X_1 ve X_3 bağımsız değişkenleri arasındaki R^2 hesaplanır. Böylece X_2 için varyans artış faktörü;

$$\text{VIF}(X_2) = \frac{1}{1-R_2^2} \tag{1.8}$$

olarak hesaplanır.

Son olarak, X_3 değişkenini bağımlı değişken olarak alıp X_1 ve X_2 bağımsız değişkenleri arasındaki R^2 hesaplanır. Böylece X_3 için varyans artış faktörü;

$$\text{VIF}(X_3) = \frac{1}{1-R_3^2} \tag{1.9}$$

olarak hesaplanır.

Bağımlı değişken (Y) ile bağımsız değişkenler (X_i) arasında ilişki yoksa R^2=0 ve VIF=1 olacaktır. Bağımlı ve bağımsız değişkenler arasında tam bir ilişki varsa R^2=1 ve VIF=∞ olacaktır. Örneğin; R^2 = 0,8 ise, VIF = $\frac{1}{1-0,8}$ = 5 olacaktır.

Çoklu bağlantı probleminin saptanmasında kullanılan diğer bir yöntem, değişkenlerin toleranslarını hesaplamaktır. Tolerans değeri,

$$T = 1 - R_i^2 \tag{1.10}$$

şeklinde hesaplanmaktadır. Yani, daha küçük tolerans daha büyük VIF demektir (Gujarati, 2012: 338).

Çoklu bağlantı problemini saptamadı kullanılan diğer bir yöntem yardımcı regresyon eşitliklerinden yararlanarak *F* değerleri hesaplamaktır. Eşitlik (1.6) daki üç bağımsız değişkenli modeli ele alacak olursak; bağımlı kabul edilen her bir bağımsız değişkenle diğer bağımsız değişkenler arasında $R_{1.23}$, $R_{2.13}$ ve $R_{3.12}$ çoklu korelasyon katsayıları hesaplanır. Daha sonra çoklu korelasyon kat sayılarından yararlanarak her bağımsız değişken için bir *F* değeri hesaplanır. Örneğin; X_1 bağımsız değişkeni için *F* değeri şu şekilde hesaplanmaktadır (Gujarati, 2012: 336):

$$F_{1.23} = \frac{\frac{R_{1.23}^2}{(k-1)}}{\frac{(\ 1-R_{1.23}^2)}{(n-k)}} \tag{1.11}$$

Formülde n, toplam örnek birim sayısını; k ise, sabit terim dahil tahmin edilecek parametre (modeldeki bağımlı ve bağımsız değişken) sayısını göstermektedir. Hesaplanan F değeri, belirli bir alfa anlamlılık düzeyi ve $df_1 = k - 2$ ve $df_2 = n - k + 1$ serbestlik dereceli kritik F değeri ile karşılaştırılmaktadır. Hesaplanan F değeri ($F_{1.23}$) kritik F değerlerinden büyük ise, X_1 değişkeniyle diğer bağımsız X_2 ve X_3 değişkenleri arasındaki ilişkinin anlamlı olduğuna karar verilmektedir (Kalaycı vd., 2010: 225).

Çoklu bağlantı probleminin saptanmasında kullanılan diğer yöntem koşullu endeks sayılarının (CI = Condition Index) hesaplanmasıdır. Koşullu endeks (CI) sayıları aşağıdaki gibi hesaplanmaktadır:

$$CI = \sqrt{\frac{V_{max}}{V_{aj}}} \tag{1.12}$$

Formülde V_{max} maksimum açıklanan varyansı; V_{aj}, i. Değişken tarafından açıklanan toplam varyansı göstermektedir. CI, 10-30 arasında ise orta düzeyde, 30'u aşarsa çok güçlü çoklu bağlantı problemi var demektir (Kalaycı vd., 2010: 226).

1.2.4.3. Çoklu doğrusal bağlantı probleminin çözümü: Çoklu bağlantı probleminin bazı çözüm önerileri şunlardır:

1) Bir veya daha çok bağımsız değişken modelden çıkarılabilir.

2) Örnek büyütülerek çoklu bağlantı problemi çözülebilir.

3) Birbiriyle ilişkili olan iki değişken, bu iki değişkenin toplamını ifade eden tek bir değişken olarak modele dahil edilebilir.

4) Değişkenler, farkları alınarak dönüştürülebilir.

5) Birbirinden bağımsız bileşenler oluşturan temel bileşenler regresyon yöntemi kullanılabilir.

1.2.5. Hataların bağımsızlığı ve otokorelasyon varsayımı: Hataların bağımsızlığı herhangi bir zaman serisinin veya eşleştirilmiş zaman serilerinin değerleri arasındaki korelasyondur. Çapraz korelasyon, iki farklı zaman serisi arasındaki ilişkidir. Sıra korelasyonu, hem otokorelasyon hem de çapraz korelasyon anlamı taşır. Pratikte,

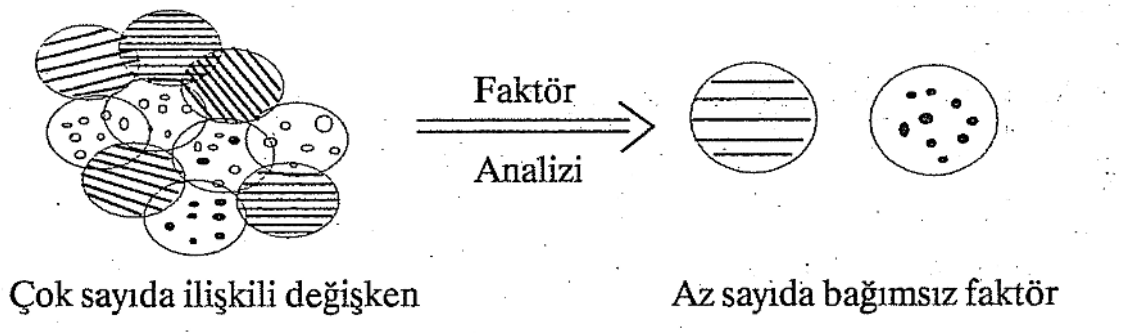
sıra korelasyonu ve otokorelasyon eş anlamda kullanılmaktadır. Anlamlı otokorelasyon modelin yanlış tanımlandığını gösterebilir. Yani, modelde önemli bir değişken unutulmuş veya fonksiyonel ilişki yanlış tanımlanmış olabilir (Kalaycı vd., 2010: 226). Otokorelasyonun saptanması için Durbin Watson (DW) istatistiği kullanılmaktadır.

1.3. Çok değişkenli istatistiksel analiz yöntemleri

Bu bölümde çok değişkenli analiz yöntemlerinin varsayımlarının gevşetilmesi ile oluşturulmuş çok değişkenli istatistiksel yöntemlerin tanımları ve kullanım amaçları kısaca ele alınacaktır.

1.3.1. Faktör analizi: p değişkenli bir olayda birbiriyle ilişkili değişkenleri bir araya getirerek az sayıda yeni ilişkisiz değişken bulmayı amaçlar. Faktör analizinde kovaryans matrisi ya da korelasyon matrisi ile işe başlanır. Genelde korelasyon matrisi kullanılmaktadır. Korelasyon matrisinin faktörleştirilmesi esasına dayalı faktör analizinde faktörleştirmede kullanılan pek çok yöntem kullanılmaktadır (Tatlıdil, 1992: 141). Bu yöntem bir boyut indirgeme ve bağımlılık yapısını yok etme yöntemidir. Amaç değişkenler arasındaki karşılıklı bağımlılığın kökenini araştırmaktır. Çok sayıdaki değişken arasındaki ilişkileri analiz etmeyi ve bu değişkenlerin temelinde yatan ortak faktörleri açıklamayı amaçlamaktadır (Cangül, 2006: 4).

Şekil 4. Faktör Analizinin Şekilsel İfadesi



Çok sayıda ilişkili değişkenden az sayıda bağımsız faktör oluşturmayı amaçlayan faktör analizi, bu faktör gruplarını kullanarak bağımlı değişken daha kolay açıklanmakta ve yorumlanmaktadır.

1.3.2. Kümeleme analizi: Kümeleme analizinin genel amacı, gruplanmamış verileri benzerliklerine göre sınıflandırmak ve araştırmacıya; işe yarar özetleyici bilgiler elde etmede yardımcı olmaktır. Kümeleme analizinde küme sayısı bilinmemekte sadece verilerin mevcut durumuna ilişkin sonuçlar vermesi nedeniyle gelecekte kullanılabilmesi söz konusu olamamaktadır. Çok değişkenli analiz yöntemlerinin varsayımlarından normallik varsayımı prensipte kalmakta, uzaklık değerlerinin normalliği yeterli görülmektedir. Ayrıca kümeleme analizinde kovaryans matrisine ilişkin herhangi bir varsayım bulunmamaktadır. Bireylerin sınıflanması, ait oldukları kitlelerin belirlenmesi ile uğraşan kümeleme analizinin amacı, sınıflama, nümerik taksonomidir (Tatlıdil, 1992: 252).

1.3.3. Kanonik korelasyon analizi: En gelişmiş ve en karmaşık ilişki analizi olan kanonik korelasyon analizi çok boyutlu kitleden çekilmiş iki ya da daha çok değişken kümesi arasındaki ilişki ile ilgilenir. Raslantı değişkenler kümesinin doğrusal fonksiyonları arasındaki maksimum korelasyonları bulmaya çalışan kanonik korelasyon analizinde tüm formülasyonlar iki raslantı değişken kümesi için geliştirilmiş olup küme sayısının ikiden çok olması durumlarında bu formüller geliştirilerek kullanılmaktadır. Daha önce de değinildiği gibi kanonik korelasyonda her bir kümenin raslantı değişkenlerinin, maksimum korelasyonlu ve birim varyanslı birer doğrusal bileşimleri elde edilmektedir. Daha sonra, bulunan bu çiftten bağımsız, maksimum korelasyonlu ve birim varyanslı ikinci bir doğrusal bileşim çifti bulunmaktadır. Bu işlemlere küçük değişken kümesindeki değişken sayısı kadar yeni doğrusal bileşim çifti elde edilinceye değin devam edilmektedir (Tatlıdil, 1992: 173).

1.3.4. Temel bileşenler analizi: İncelenen birçok özellik bakımından X değişken kümesinin varyans yapısını, p adet orijinal değişken yerine, k adet değişken ($k<p$) ve bu değişkenlerin doğrusal bileşenleri olan yeni değişkenler ile ifade etmek amacıyla kullanılır (Özdamar, 2004). Kısaca değişkenler arasındaki bağımlılığı yok etmek ve boyut indirgemek için kullanılır. Kendisi bir analiz olmakla birlikte çoklu regresyon analizinde çoklu doğrusal bağlantıyı gidermek ve değişken kümelerinin boyutunu indirgemek için de kullanılır. Bu yöntem, bütün değişkenlerdeki varyansı maksimum açıklayacak faktörün bulunmasını amaçlar.

1.3.5. Çok değişkenli varyans analizi (MANOVA): Tek yönlü ve iki yönlü MANOVA olmak üzere ikiye ayrılır. Tek yönlü MANOVA, birden fazla bağımlı

değişkene tek bir bağımsız değişkenin etki ettiği durumlarda kullanılır. Bu analizde H_0 hipotezi, bağımlı değişkenlerin hiçbirinde faktördeki gruplara göre bir ortalama farklılığının olmadığıdır. H_1 hipotezi ise, en az bir bağımlı değişkende faktörün en az iki grubuna göre bir ortalama farklılığı olduğudur. Dolayısıyla, bağımsız değişkenin gruplarından sadece ikisi arasında bile bağımlı değişkenin ortalamaları arasında bir fark gözlemlenirse H_0 hipotezi reddedilir (Kalaycı vd., 2010: 155). Kısaca, bağımsız değişkenler ve iki ve ya daha fazla bağımlı değişken arasındaki ilişkiyi araştırır.

1.3.6. Çoklu doğrusal regresyon analizi: Birden çok bağımsız değişkenle kurulan regresyon modeline çoklu doğrusal regresyon modeli denir. Çoklu doğrusal regresyon modeli;

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon \tag{1.13}$$

şeklinde ifade edilir. Burada;

Y, Bağımlı değişken

X_i , Bağımsız değişkenler

β_i , Tahmin edilecek parametreler

ε , Hata terimi

dir.

Çok değişkenli analiz yöntemlerinin tüm varsayımları geçerlidir. Bu analiz sadece bir bağımlı değişkenin birden çok bağımsız değişken ile ilişkisinin incelenmesinde kullanılır. Bağımsız değişkendeki değişimlerin bağımlı değişkende nasıl bir değişim oluşturacağını tahmin eder.

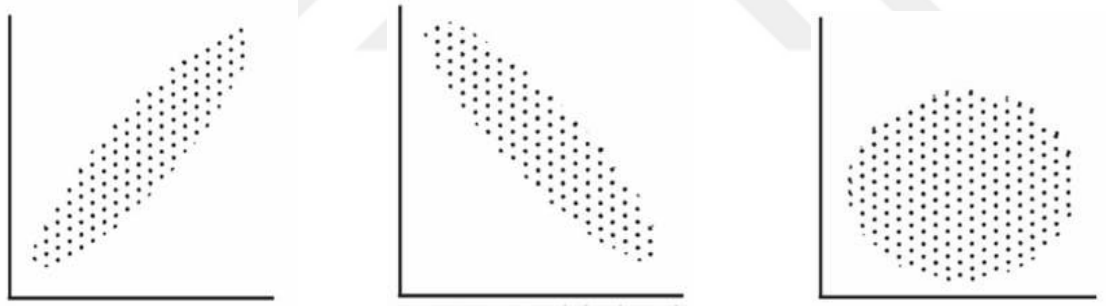
1.3.7. Lojistik regresyon analizi: Bağımlı değişkenin kategorik olduğu regresyon modelidir. Bağımsız değişkenler kategorik, sürekli ve ya hem kategorik hem sürekli olabilirler. Amaç, lojistik ayırt etme fonksiyonunu kullanarak verileri gruplara atamaktır. Ayırma analizinden farkı bağımsız değişkenlerin kategorik olabilmesidir.

1.3.8. Kovaryans analizi: Kovaryans analizi, varyans analizinin bir uzantısıdır. Varyans analizi, grup ortalamaları arasındaki farkları bulmak için kullanılır. Analizde, birkaç bağımsız değişkenin bir bağımlı değişken üzerindeki etkisi ortaya koyulmaya çalışılır. Kovaryans analizi ise, iki veya daha fazla sayıdaki grupta, bir bağımlı değişkenin ortalamalarının karşılaştırılmasında, bağımlı değişkeni etkileyen başka bir bağımlı değişkenin etkisinin yok edilmesi veya arttırılması söz konusudur. Kovaryans

analizi, varsayımlarının karşılanması durumunda yararlı ve güçlü bir istatistiksel yöntemdir. Kovaryans analizi sırasında, grup ortalamaları arasındaki fark ölçülürken regresyon analizi ve varyans analizi birlikte kullanılır. Yani Kovaryans analizi, varyans analizi ile regresyon analizinin bir kombinasyonuduır (Kalaycı vd., 2010: 185).

1.3.9. Korelasyon analizi: Korelasyon analizi, iki değişken arasındaki doğrusal ilişkiyi veya bir değişkenin iki veya daha fazla değişken ile olan ilişkisini test etmek, varsa bu ilişkinin derecesini ölçmek için kullanılan istatistiksel bir yöntemdir (Kalaycı vd., 2010: 115). Korelasyon analizinin amacı, bağımsız değişkendeki değişimin bağımlı değişkeni hangi yönde ne kadar etkileyeceğini göstermektir. Bu analizin uygulanabilmesi için her iki değişkeninde sürekli ve normal dağılmış olmaları gerekmektedir. Korelasyon analizi sonucunda, ilişki olup olmadığı ve eğer ilişki varsa bu ilişkinin derecesi korelasyon katsayısı ile hesaplanır. Bu katsayı *r* ile gösterilir ve -1 ile +1 arasında bir değer alır. +1 aynı yönlü (pozitif) ilişkiyi, -1 zıt yönlü (negatif) ilişkiyi, 0 ise değişkenler arasında ilişki olmadığını gösterir.

Şekil 5. Farklı Korelasyon Durumları



Pozitif yönlü ilişki Negatif yönlü ilişki İlişki yok

1.3.10. Çok boyutlu ölçekleme analizi: Çok boyutlu ölçekleme analizi, nesneler arasındaki ilişkilerin bilinmediği, fakat aralarındaki uzaklıkların hesaplanabildiği durumlarda uzaklıklardan yararlanılarak nesneler arasındaki ilişkileri ortaya koymaya yarayan istatistiksel bir yöntemdir. Çok boyutlu ölçekleme analizinin uygulama alanı son derece geniştir. Hem metrik hem de metrik olamayan değişkenlere uygulanabilir. Ayrıca benzerlik ve farklılıklara göre değişik nesnelerin mümkün olan en az boyuttaki en iyi düzenlemesine ulaşma imkanını da vermektedir (Kalaycı vd., 2010: 379).

Çok boyutlu ölçekleme analizinin amacı; n tane nesne arasındaki uzaklık değerlerini kullanarak bu nesnelerin çok boyutlu uzaydaki konumlarını, ilişki yapısını veren resmini ortaya koymaktır. Uzaklıklara dayalı bu yöntemde genel olarak Öklit uzaklıklarının kullanılmasına karşın, asimetrik uzaklıkların bulunması durumunda Öklit yerine diğer uzaklık ölçütlerinden de yararlanılmaktadır. Bunların yanı sıra benzerlik ölçütleri de uzaklık ya da farklılık ölçütleri yerine kullanılabilmektedir (Tatlıdil, 1992: 269).

BÖLÜM 2

2. AYIRMA ANALİZİ

2.1. Ayırma analizinin tanımı ve amaçları

Ayırma analizi, kategorik bağımlı değişkenler ile metrik bağımsız değişkenler arasındaki ilişkileri tahmin etmeyi amaçlayan çok değişkenli istatistik tekniklerinden biridir (Kalaycı vd., 2010: 335).

Ayırma analizi, hatalı sınıflandırma olasılığını en aza indirgeyerek birimleri (bireyleri) ait oldukları gruplara ayırmak, çekilmiş oldukları kitleleri belirlemek amacına yönelik istatistiksel bir karar verme yöntemidir (Tatlıdil, 1992: 202).

Ayırma analizi, gruplar arasında çeşitli değişkenlere bağlı olarak farklılıklarını belirlemesine olanak sağlamaktadır. Bireyleri en az hata ile ait oldukları gruplara atamak amaçlanmaktadır. Bunun için bireylerin ait olduğu kitlenin belirlenmesi için bunu sağlayacak ayırma fonksiyonunu bulmak esastır. Bazı araştırmalar bu ayırma fonksiyonunun katsayılarının hesaplanmasında başvurulan yöntemlere göre ayırma analizini, kanonik ayırma analizi, en çok olabilirlik ayırma analizi, Bayes ayırma analizi gibi çeşitli şekillerde adlandırmaktadırlar.

Ayırma analizinin temeli, incelenen bireyin kitlesinin belirlenmesini sağlayacak bir fonksiyonun bulunmasıdır. Bu fonksiyonun bulunmasında, belirlenecek grupların ortalamaları arasındaki farklılığın maksimum olması amaçlanmaktadır (Tatlıdil, 1992: 203).

Ayırma analizi, X veri setindeki değişkenlerin iki ve daha fazla gerçek gruplara ayrılmasını sağlayan, birimlerin p tane özelliğini ele alarak bu birimlerin doğal ortamdaki gerçek gruplarına, sınıflarına en uygun düzeyde atanmalarını sağlayacak fonksiyonlar türeten bir yöntemdir (Özdamar, 2004).

Ayırma analizi, birbiriyle yakından ilişkili birkaç istatistiksel yaklaşımı kapsayan geniş bir kavramdır. Bu yaklaşımlar iki kategoride ele alınabilir. İlk kategori,

gruplar arası farklılıkları yorumlama da kullanılırken ikincisi, birimleri gruplara ayırmak için kullanılır. Eğer ayırma analizi bir ayırma fonksiyonu belirlemek için kullanılmış ise buna Tanımlayıcı Ayırma Analizi, birimleri sınıflamak için kullanılmışsa Tahmin Edici Ayırma Analizi olarak adlandırılır.

Tahmin edici ayırma analizi, geleceği yordamaya çalışan birçok bilimsel alanda kullanılmaktadır. Örneğin; yaşam beklentisi, ekonomik büyüme, akademik başarı, bir eğitim programının başarısı, seçmenin oy vermesi, satış gelirleri tahminleri, aile planlaması çalışmaları vb. birçok çalışmada kullanılabilir. Bu örneklerin her biri bir bağımlı değişken ile birlikte bir veya daha fazla bağımsız değişken içerir. Tanımlayıcı ayırma analizi, bağımsız değişkenlerin grupları ayırmada etkinliğini araştırmak ve bunlardan yararlanarak ayırma fonksiyonu oluşturarak grupları tanımlamada hangi değişkenlerin etkili olduğunu göstermektir.

Bu tanımlamalar doğrultusunda ayırma analizinin tanımını şu şekilde yapabiliriz: Ayırma analizi, bireylerin çok sayıdaki özelliğini (bağımsız değişkenler) dikkate alarak bireyleri en az hata ile optimal düzeyde mensubu oldukları gruplara ayırmak, gruplara ayırmada hangi özelliklerin (bağımsız değişkenler) etkili olduğuna karar vermek ve bireyin hangi gruptan çekildiğini belirlemede kullanılan çok değişkenli bir istatistik yöntemdir.

Ayırma analizi bir veya birden çok amaç için kullanılabilir. Ayırma analizinin amaçlarını maddeler halinde ifade edecek olursak:

1) Bireyin grup üyeliğini tahmin etmek, başka bir deyişle, bir verinin (gözlem, denek, vaka) hangi değişken grubuna gireceğine karar vermek için kullanılabilir.

2) Ayırma fonksiyon eşitliğini kullanarak, verilerin gruplara ayrılması için kullanılabilir.

3) Bağımsız değişkenlerin aritmetik ortalamalarının gruplar arasında nasıl değiştiğini tespit etmek için kullanılabilir.

4) Bağımlı değişkenin varyansının ne kadarının bağımsız değişkenler tarafından açıklanabildiğini belirlemek için kullanılabilir.

5) Gruplara ayırmada etkili olan ve olmayan bağımsız değişkenleri belirlemek için kullanılabilir.

6) Verilerin tahmin edildiği gibi sınıflandırılıp sınıflandırılmadığını test etmek için kullanılabilir (Kalaycı vd., 2010: 335).

2.2. Ayırma analizinin varsayımları

Çok değişkenli istatistik yöntemleri ve varsayımlarını incelediğimiz birinci bölümdeki varsayımlardan ayırma analizinde yanlış sınıflandırma olasılığını ortadan kaldırmak için kullanılan bazıları şunlardır:

1) Değişkenlerin çoklu normal dağılıma sahip olmaları,

2) Bütün gruplar için kovaryans matrislerinin eşit olması ve

3) Bağımsız değişkenler arasında çoklu doğrusal bağlantı probleminin olmaması gerekir.

Bu varsayımlardan bazılarının hafif derecede ihmal edilmesinin ayırma analizi sonuçlarını önemli ölçüde etkilemediği çeşitli araştırmacılar tarafından gösterilmiştir. Bunlardan Lachenbruch 1975 yılındaki çalışmasında çoklu normal dağılım ve eşit kovaryans varsayımlarının hafif derecede ihmalinin analizi sonuçlarını önemli ölçüde etkilemediğini göstermişken, Klecka 1980 yılındaki çalışmasında çoğu zaman normal dağılım kuralını ihlal eden dikotom (iki sonuçlu) değişkenlerin analiz sonuçlarını etkilemeyebileceğini göstermiştir. Yine de araştırmalarda kullanılan veri grubunun bu varsayımlardan sapmalarının ayırma analizi sonuçları tam olarak gerçeği yansıtmayacağı bilinmelidir.

2.3. Ayırma analizinde kullanılan veri seti büyüklüğü ve veri yapısı

Ayırma analizi için gerekli olan veri seti büyüklüğü, her bir değişken için minimum 20 olmak üzere en az 100 veridir (Kalaycı vd., 2010: 335).

g. grupta $(G_1, G_2, \dots, G_p)$, p. değişkenin $(X_1, X_2, \dots, X_p)$ her birine ilişkin n_j

(j:1,2, …, g) gözlem yapıldığı varsayıldığında aşağıdaki veri tablosu elde edilir (Kurtuluş, 1976: 468).

Şekil 6. G – Grup İçin Ayırma 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}

2.4. İki gruplu ayırma 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, 1992: 203).

İki gruplu ayırma analizi için açıklanan varyansın yüzde yüzüne karşılık gelen bir ayırma fonksiyonu mevcuttur. Eğer birden fazla ayırma fonksiyonu varsa bunlardan ilki, bunlardan en büyüğü ve en önemlisidir. Ayırma fonksiyonu belli şartları gerçeklediği bilinen bağımsız değişkenlerin bir doğrusal birleşimidir. Bir bireye ait p tane değişkenden bulunacak olan fonksiyon,

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \cdots + a_p x_{ip} \qquad (2.1)$$

şeklinde gösterilebilir. Burada $x_1, x_2, \dots, x_p$ bağımsız değişkenleri, $a_1, a_2, \dots, a_p$ bağımsız değişkenlere ait katsayıları ve a_0 ise sabit katsayıyı göstermektedir. Bu fonksiyona ayırma fonksiyonu adı verilmektedir. Bu fonksiyon bulunurken, gruplar arası varyansın grup içi varyansa oranının en büyük olması yani gruplar arası varyans en büyük iken grup içi varyansın en küçük olması gerekir. Kısaca,

$$F = Max\left(\frac{Gruplar\ Arası\ Varyans}{Grup\ \text{İçi}\ Varyans}\right) \tag{2.2}$$

oranının en büyük olması gerekir. Buradan hareketle a_j katsayılarının bulunmasında kullanılan ilk eşitlik Fisher tarafından aşağıdaki şekliyle ifade edilmiştir.

$$f(a_1, a_2, \dots, a_p) = \frac{a^{\iota}\boldsymbol{B}a}{a^{\iota}\boldsymbol{W}\boldsymbol{a}} \tag{2.3}$$

Burada;

a: px1 boyutlu katsayılar vektörünü,

B: pxp boyutlu (her grup ortalama vektörünün genel ortalama vektöründen farklarından elde edilen) gruplar arası varyans matrisini,

W: pxp boyutlu (her gruptaki bireylerin kendi ortalama vektörlerinin farklarından elde edilen) grup içi varyans matrisini göstermektedir. Şayet,

$$C = \frac{1}{(n_1+n_2-2)}\ W \tag{2.4}$$

$\bar{x}_j^{(1)}$ birinci grubun j' inci değişkenin ortalaması

$\bar{x}_j^{(2)}$ ikinci grubun j' inci değişkenin ortalaması olsun.

$$d_j = (\bar{x}_j^{(1)} - \bar{x}_j^{(2)}) \tag{2.5}$$

$$d^{\iota} = (d_1, d_2, \dots, d_p) \tag{2.6}$$

şeklinde alınırsa (2.3)' deki eşitlik şu şekilde yazılabilir.

$$f(a_1, a_2, \dots, a_p) = \frac{n_1 n_2}{n_1+n_2} \frac{a^{\iota} d d^{\iota} a}{a^{\iota} C a} \quad (2.7)$$

Bu eşitlik, p bilinmeyenli denklem sistemidir. İki grubun grup içi varyanslarının (kovaryans) eşitliği varsayımından,

$$a = C^{-1} d \quad (2.8)$$

olduğu gösterilebilir. Bu durumda bağımsız değişkenlerimize ait katsayılar $(a_1, a_2, \dots, a_p)$, grup içi kovaryans matrisinden elde edilen bir *C* matrisinin tersi (C^{-1}) ile grup ortalama vektörlerinden elde edilen fark değerleri vektörlerinin (*d*) çarpımından elde edilir. Yani birinci grup için ayırma fonksiyonu ortalaması, (2.1) nolu eşitlikte x'ler yerine bu grubun ortalama değerleri, ikinci grup için ayırma fonksiyonu ortalaması yine aynı eşitlikte x'ler yerine ikinci grubun ortalama değerleri yazılarak bulunur.

$$\bar{y}^{(1)} = a_0 + a_1 \bar{x}_1^{(1)} + a_2 \bar{x}_2^{(1)} + \cdots + a_p \bar{x}_p^{(1)} \quad (2.9)$$

$$\bar{y}^{(2)} = a_0 + a_1 \bar{x}_1^{(2)} + a_2 \bar{x}_2^{(2)} + \cdots + a_p \bar{x}_p^{(2)} \quad (2.10)$$

İki grubun varyans-kovaryans matrislerinin ortak olması durumunda, bulunan ayırma fonksiyonunun varyansı da ortak olacaktır. Ayırma fonksiyonu varyansını matris gösterimiyle şu şekilde ifade edebiliriz:

$$Var(y) = \sum_{i=1}^{p} \sum_{j=1}^{p} c_{ij} a_i a_j = a^{\iota} C a \quad (2.11)$$

(2.8) eşitliğinden yararlanarak,

$$Var(y) = d^{\iota} C^{-1} C C^{-1} d = d^{\iota} C^{-1} d \quad (2.12)$$

elde edilebilir.

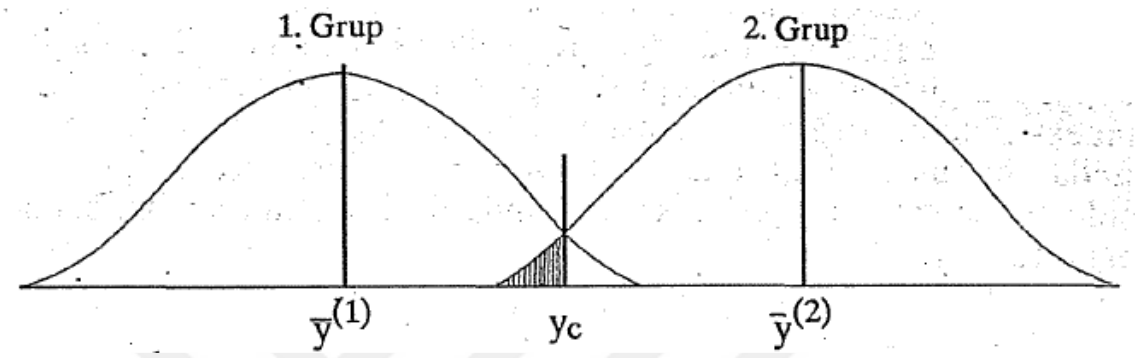
Kovaryans matrislerinin ortak olduğu varsayılan iki grup olması durumunda y ile ifade edilen ayırma fonksiyon değişkeni de $\bar{y}_1$ ve $\bar{y}_2$ ortalamaları ve $\sigma^2 = d^{\iota} a$ varyansı ile normal dağılacaktır. Bu durumda ayırma fonksiyonuna ait her grup için z değeri şu şekilde hesaplanır:

$$z_{y_i} = \frac{y - \bar{y}^{(i)}}{\sqrt{Var(y)}} \quad (2.13)$$

Bu eşitlikten faydalanarak i=1 için birinci gruba ait bir z değeri, i=2 için ikinci gruba ait bir z değeri bulunabilir.

y_c, iki grup arasındaki sınır noktasını göstermek üzere, iki grup olması durumunda hatalı sınıflandırma durumu Şekil.7' de gösterilmiştir.

Şekil 7. Hatalı Sınıflandırma Durumu (Tatlıdil, 1992: 205)



Şekilde taralı kısma düşen bireyler gerçekte ikinci grubun elemanları iken birinci grubun elemanları olarak sınıflandırılmışlardır. Bu nedenle ikinci gruptaki bireylerin yanlış sınıflandırma olasılığı, taralı alanın ikinci grubun toplam alanına oranı şeklinde bulunmaktadır. Kısaca ikinci grupta yer alan bir bireyin yanlışlıkla birinci gruba sınıflandırma olasılığı, y_c sınır noktası iken (2.13) nolu eşitlikte verilen z değerine karşılık gelen olasılık olacaktır. Dolayısıyla birinci ve ikinci gruptaki bir bireyin doğru sınıflandırılma olasılıkları şöyledir;

$$1 - P(z > \frac{y_c - \bar{y}^{(1)}}{\sqrt{Var(y)}}) \tag{2.14}$$

$$1 - P(z > \frac{y_c - \bar{y}^{(2)}}{\sqrt{Var(y)}}) \tag{2.15}$$

şeklinde bulunmaktadır.

Ayırma analizinin amacının, bireyleri minimum hata ile doğru gruplara atamak için bir ayırma fonksiyonu bulmak olduğu hatırlandığında bu amaca ulaşmak için en önemli aşamanın (2.1) nolu ayırma fonksiyonundaki bağımsız değişken katsayılarının $(a_1, a_2, \dots, a_p)$ doğru tahmin edilmesi ve y_c ile ifade edilen sınır noktasının doğru belirlenmesi olduğu görülecektir.

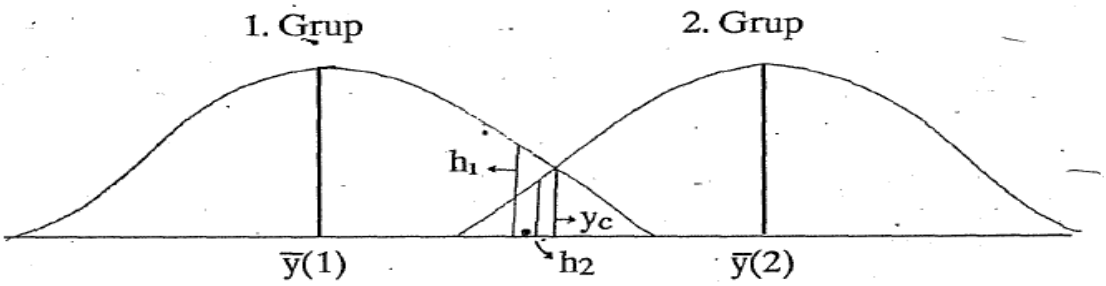
(2.11) nolu eşitlikte verilmiş olan değere Mahalanobis D^2 uzaklığı denmektedir. Çok değişkenli normal dağılımlı ve ortak varyans-kovaryans matrisine sahip iki kitleden gelen örneklemler için D^2 uzaklığı değeri F dağılımına uymaktadır. Bunu matematiksel olarak şu şekilde gösterebiliriz;

$$F = \frac{n_1 n_2 (n_1+n_2-p-1)}{p(n_1+n_2)(n_1+n_2-2)} \; D^2 \qquad (2.16)$$

(2.16) nolu eşitlikten yararlanılarak elde edilen ayırma fonksiyonunun ayırıcı özelliğinin önemli olup olmadığının testinde F tablosunun uygun serbestlik derecelerindeki tablo değerleri kritik değer olarak kullanılmaktadır. Daha önce de belirtildiği gibi $Var(y) = \bar{y}^{(1)} - \bar{y}^{(2)} = D^2$ dir ve (2.16) nolu eşitlikten elde edilen F değeri (ki D^2 ye bağlı olduğu için fonksiyonun ayırma gücünü göstermektedir) p ve $n_1 + n_2 - p - 1$ serbestlik dereceli, α yanılma düzeyindeki F tablo değeri ile karşılaştırılır. Hesap edilen değerin tablo değerini aşması durumunda, bulunan ayırma fonksiyonunun bireyleri gruplara ayırma özelliğinin iyi olduğu sonucuna ulaşılır (Tatlıdil, 1992: 206).

2.4.1. Yanlış sınıflamaya ilişkin olasılıkların tahmini: Ayırma analizinin en önemli sorununun bireylerin yanlış gruplara atanması olduğu daha önce ifade edilmişti. Yanlış sınıflandırmanın genellikle her iki grup ortalamalarına aynı uzaklıkta bulunan bireyler için söz konusu olduğu söylenebilir. Şekil 8 incelendiğinde yanlış sınıflandırma olasılıklarının bulunmasının daha anlaşılır olmasını sağladığı söylenebilir.

Şekil 8. Bireylerin Gruplara Ait Olma Olasılıklarının Şekilsel İfadesi (Tatlıdil, 1992: 206).



Herhangi bir birey y_c noktasında yer almış ise hatalı sınıflandırılma olasılığının büyük olacağı daha önce ifade edilmişti. Grupların birey sayıları eşit ise her iki grup

içinde hatalı sınıflandırma olasılığı eşit olacaktır. Şekil 8' de nokta şeklinde verilmiş bireyin birinci ve ikinci gruba ait olma olasılıklarını bulmak için;

h_1: bireyin birinci gruba ait normal dağılım eğrisine uzaklığı,

h_2: bireyin ikinci gruba ait normal dağılım eğrisine uzaklığı olsun.

Normal dağılımın olasılık yoğunluk fonksiyonundan bu yükseklikler;

$$h_1 = \frac{1}{\sqrt{2\pi Var(y)}} e^{\frac{-\left(y-\bar{y}^{(1)}\right)^2}{2} Var(y)} \tag{2.17}$$

$$h_2 = \frac{1}{\sqrt{2\pi Var(y)}} e^{\frac{-\left(y-\bar{y}^{(2)}\right)^2}{2} Var(y)} \tag{2.18}$$

şeklinde yazılabilir. Grupların eşit sayıda birey içermesi durumunda, bireyin gruplara ait olma olasılıkları;

P_1: bireyin birinci gruba ait olma olasılığı,

P_2: bireyin ikinci gruba ait olma olasılığı olmak üzere;

$$P_1 = \frac{h_1}{h_1+h_2} \tag{2.19}$$

$$P_2 = \frac{h_2}{h_1+h_2} \tag{2.20}$$

$$P_1 + P_2 = 1 \tag{2.21}$$

şeklinde gösterilir.

Eğer grup birey sayıları birbirinden farklı ise,

n_1: birinci grubun birey sayısı,

n_2: ikinci grubun birey sayısı olmak üzere;

İlk olarak q_1 ve q_2 oranları şu şekilde bulunur:

$$q_1 = \frac{n_1}{n_1+n_2} \tag{2.22}$$

$$q_2 = \frac{n_2}{n_1+n_2} \tag{2.23}$$

$$q_1 + q_2 = 1 \tag{2.24}$$

Ön olasılıklar olarak adlandırılan bu değerler kullanılarak bireylerin gruplara ait olma olasılıkları (2.19) ve (2.20) deki eşitliklere benzer şekilde hesaplanır.

$$P_1^l = \frac{q_1 P_1}{q_1 P_1 + q_2 P_2} \tag{2.25}$$

$$P_2^l = \frac{q_2 P_2}{q_1 P_1 + q_2 P_2} \tag{2.26}$$

$$P_1^l + P_2^l = 1 \tag{2.27}$$

Bulduğumuz bu sonuçlardan büyük olana birey sınıflandırılır.

Sonuç olarak gruplara atanmak istenen bireyin (2.1) nolu ayırma fonksiyonu ile verilen değeri hesaplandığında,

$$y - y_c < 0 \tag{2.28}$$

ise birey birinci gruba eğer,

$$y - y_c > 0 \tag{2.29}$$

ise birey ikinci gruba sınıflanır.

Bağımlı değişkenin ikili katagorik ve bağımsız değişkenlerin metrik olduğu model aynı zamanda lojistik regresyon analizi ile de analiz edilebilir. Ayırma analizinin Lojistik regresyon analizinden ayrıldığı nokta bağımlı değişkenin ikiden çok katagorili olduğu modeldir. Bundan sonraki bölümde dolayısı ile ikiden fazla katagori olması durumunda ayırma analizi incelenecektir.

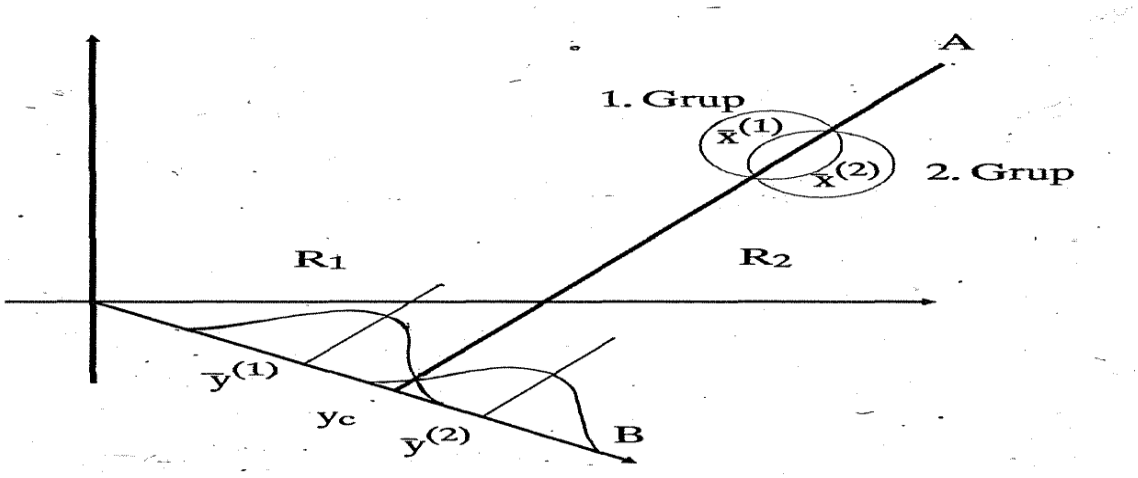
2.5. İkiden çok grup olması halinde ayırma analizi

İkiden çok gruplu ayırma analizi, temelde iki gruplu ayırma analizine dayanmaktadır. İki gruplu ayırma analizinde olduğu gibi ikiden çok gruplu ayırma analizinde de p değişkenli olmak üzere bu sefer ikiden çok sonlu sayıda kitle bulunmaktadır.

Bireyler, kitleler arasında ayırma gücü en büyük olacak biçimde belli sayıda doğrusal bağlantılar yardımıyla sınıflandırılmaktadır. Bireylere ait p değişkeni, p boyutlu uzayda tanımlayacak öyle eksenler bulunmalı ki veri toplulukları bu eksenler boyunca birbirinden olabildiğince ayrılabilsin. Ayırma analizi, bu eksenleri belirtecek doğrusal bağıntılara ait katsayıların bulunması ile ilgilidir. İkiden çok grup olması halinde bulunacak ayırma fonksiyonu sayısı; k, grup sayısını, p, değişken sayısını göstermek üzere min(k-1,p) tane olacaktır (Tatlıdil, 1992: 208).

Yukardaki tanımlamalar ışığında aslında ayırma analizini bir boyut indirgeme tekniği olarak da düşünebiliriz. Nitekim bireyler tam ölçüm uzayından bir alt uzaya indirgenir. İki boyutlu bir uzaydan tek boyutlu bir uzaya indirgeme aşağıdaki şekilde gösterilmektedir.

Şekil 9. Ayırma Analizinde Boyut İndirgeme (Tatlıdil, 1992: 209)



Şekilden de görüleceği gibi iki boyutlu uzaydaki noktaların B doğrusu üzerine izdüşümleri arasındaki çakışma en az olacaktır. Ayrıca fonksiyon, bireye ait iki ayrı değişken değerini tek bir ayırıcı fonksiyon değerine dönüştürmektedir.

Yine şekil 9'dan da görüleceği üzere, ayırma ekseni B, ayırma analizinin varsayımları altında oluşturulmuştur. Nitekim grupların varyans-kovaryans matrisleri eşit olmazsa, grupların dağılımları birbirine benzemez ve bu nedenle iki grubu ayıran sınır olan A doğrusu en iyi sınır olmayacaktır.

Bireylerin k tane normal dağılımlı ve ortak varyans-kovaryans matrise sahip kitleden gelmesi durumu için geliştirilen yöntem, Fisher'in iki grubunun ayrımında kullanılan yönteminin devamı biçimindedir. Anderson tarafından geliştirilen yöntem k grubun birbiriyle kıyaslanması esasına dayanmaktadır. Bu düşünce ışığında $\binom{k}{2} = \frac{k(k-1)}{2}$ tane ayırıcı fonksiyonun bulunması gerekmektedir. Bu fonksiyonlar, karşılaştırılan iki grubun olasılık yoğunluk fonksiyonlarının birbirine oranlanmasından elde edilmektedir. Aşağıda $P_i(x)$,

$$P_i(x) = \frac{1}{(2\pi)^{p/2}|\sum|^{1/2}}\ exp\left[-\frac{1}{2}\left(x-\mu^{(i)}\right)^{\iota}\sum^{-1}\left(x-\mu^{(i)}\right)\right] \qquad (2.30)$$

i-inci grupa ait olasılık yoğunluk fonksiyonu olmak üzere, i ve j gibi iki grubun karşılaştırılması amacıyla,

$$f_{ij} = \frac{P_i(x)}{P_j(x)} = \frac{exp\left[-\frac{1}{2}\left(x-\mu^{(i)}\right)^{\iota}\sum^{-1}\left(x-\mu^{(i)}\right)\right]}{exp\left[-\frac{1}{2}\left(x-\mu^{(j)}\right)^{\iota}\sum^{-1}\left(x-\mu^{(j)}\right)\right]}$$

$$= x^{\iota}\sum^{-1}\left(\mu^{(i)}-\mu^{(j)}\right) - \frac{1}{2}\left(\mu^{(i)}-\mu^{(j)}\right)^{\iota}\sum^{-1}\left(\mu^{(i)}-\mu^{(j)}\right) \qquad (2.31)$$

fonksiyonu kullanılmaktadır. Bu fonksiyonun yardımıyla i-inci ve j-inci grupları birbirinden ayıran bölgeler tanımlanırken aşağıdaki s sabiti kullanılmaktadır.

$$S = \frac{q_j M(\frac{i}{j})}{q_i M(\frac{j}{i})} \qquad (2.32)$$

Burada M(i/j), j-inci grupta olması gereken bireyi i-inci gruba sınıflamanın maliyeti, M(j/i) i-inci grupta olması gereken bireyi j-inci gruba sınıflamanın maliyeti, q_i ve q_j değerleri ise (2.22) ve (2.23) nolu eşitliklerde verildiği gibi ön olasılıklardır (eğer ön olasılıklar ve maliyetler eşit ise s sabiti bire eşit olmaktadır). Bu durumda bölgelere sınıflama,

$$\begin{matrix} f_{ij}(x) \geq InS \ \ ise\ R_i \\ f_{ij}(x) < InS \ \ ise\ \ R_j \end{matrix} \quad \text{i,j=1,2,…,k ve i≠j için} \qquad (2.33)$$

biçiminde olmaktadır.

Yukarıda verilen (2.31) nolu eşitlikle ayırma fonksiyonu grup parametrelerine bağlıdır. Eğer grup hakkında bilgi mevcut değilse, her gruptan çekilen örneklemler yardımı ile grup parametreleri tahmin edilerek ayırma fonksiyonu,

$$f_{ij}(x) = x^{\iota}C^{-1}\left(\bar{x}^{(i)} - \bar{x}^{(j)}\right) - \frac{1}{2}\left(\bar{x}^{(i)} - \bar{x}^{(j)}\right)^{\iota}C^{-1}\left(\bar{x}^{(i)} - \bar{x}^{(j)}\right) \qquad (2.34)$$

biçiminde yazılmaktadır. Kullanılan C matrisi,

$$C = \frac{1}{(\sum_{i=1}^{k} n_i) - k}\ W \qquad (2.35)$$

biçimindedir.

Yukarıda tanımlanan $f_{ij}(x) = -f_{ji}(x)$ özelliği taşıyan fonksiyonun kullanımı ile bireylerin hangi gruba sınıflandırılacağına karar verilmektedir. k = 3 ve $p \geq 2$ olması durumunda;

-eğer $f_{12}(x) \geq 0$ ve $f_{13}(x) \geq 0$ ise birey birinci gruba

-eğer $f_{21}(x) \geq 0$ ve $f_{23}(x) \geq 0$ ise birey ikinci gruba

-eğer $f_{31}(x) \geq 0$ ve $f_{32}(x) \geq 0$ ise birey üçüncü gruba

sınıflandırılır.

Bireylerin sınıflandırılmasında kullanılan fonksiyon dikkatle incelenecek olursa; bireylerin p tane değişkeninin göz önüne alındığı ve bu değişkenlere göre bireylerin birbirine olan uzaklıklarına dayalı ayırma fonksiyonlarının elde edildiği görülür.

Tüm bu anlatımlardan anlaşılacağı üzere, iki gruplu ayırma analizinde tek bir ayırma fonksiyonu grupları ayırmada yeterli iken, ikiden daha fazla grup olması halinde tek bir ayırma fonksiyonu grupları ayırmada yeterli değildir. Bu nedenle daha önce de değinildiği gibi grup sayısının bir eksiği ve bağımsız değişken sayısından en az olan sayı kadar ayırma fonksiyonuna ihtiyaç vardır. Bulunan ilk fonksiyon ise ayırma gücü en büyük olan fonksiyon olacaktır. Bulunan ikinci ayırma fonksiyonu ilkine göre daha az güçlü eğer varsa bulunacak üçüncüsüne göre daha kuvvetli ayırma gücüne sahip olacaktır. Bu fonksiyonların kuvveti bu şekilde ifade edilebilir.

2.5.1. Çok gruplu ayırma analizinde ayırma fonksiyonlarının üretilmesi: İkiden çok grup olması durumunda bulunacak ayırma fonksiyonlarının elde edilmesinde de iki gruplu ayırma analizinde olduğu gibi varyans oranlarının en büyüklenmesi gerekir.

Fisher tarafından tanımlanan iki varyans oranı,

$$\lambda = \frac{a^{\iota}Ba}{a^{\iota}Wa} \qquad \vdots \quad max \tag{2.36}$$

şeklindedir. λ'nın maksimum olmasını, grupları birbirinden maksimum ölçüde ayırmak ve her bir grup içindeki gözlemlerin birbirine mümkün olduğunca yakın olmalarını sağlamak amacıyla isteriz. Buradan elde edilecek λ_i özdeğerlerine karşılık gelen özvektörler yukarıda belirtilen koşulları sağlayan ayırıcı fonksiyonlar olacaktır. Dolayısıyla (2.30) daki eşitliğin a'ya göre türevi alınıp 0'a eşitlenirse;

$$\frac{\partial\lambda}{\partial a} = \frac{2[(Ba)(a^{\iota}Wa) - (a^{\iota}Ba)(Wa)]}{(a^{\iota}Wa)(a^{\iota}Wa)} = 0 \tag{2.37}$$

payı ve paydayı $a^{\iota}Wa$ ile bölüp, $\frac{a^{\iota}Ba}{a^{\iota}Wa}$ yerine λ konursa,

$$\frac{2(Ba)(a^{\iota}Wa)}{(a^{\iota}Wa)} - \frac{2(a^{\iota}Ba)(Wa)}{(a^{\iota}Wa)} = 0$$

$$2(Ba) - 2\lambda(Wa) = 0$$

$$(Ba) - \lambda(Wa) = 0$$

$$(B - \lambda W)a = 0 \qquad \text{veya} \tag{2.38}$$

$$(W^{-1}B - \lambda I)a = 0 \tag{2.39}$$

Eşitlikleri elde edilir.

λ_i (i=1,2,…,r) değerleri $W^{-1}B$ matrisinin öz değerleri olarak,

$|W^{-1}B - \lambda\mathrm{I}|a = 0$ determinantının çözümünden bulunur. Ayırma fonksiyonu sayısı; r = min(k-1,p) olmak üzere elde edilen $\lambda_1, \lambda_2, \ldots, \lambda_r$ özdeğerlerine karşılık gelen r tane

özvektör, istenen ayırma fonksiyonları olacaktır. k-1 < p olması durumunda $W^{-1}B$ matrisi simetrik olmayacaktır. Bu durumda özdeğerler karekök yöntemiyle veya özel logaritmalar kullanılarak elde edilirler.

λ_1, $W^{-1}B$ matrisinin öz değeri ve $a^{(1)} = (a_0^{(1)}, a_1^{(1)}, \dots, a_p^{(1)})$ birinci özdeğere karşılık gelen öz vektör (katsayılar) ise, birinci ayırma fonksiyonu,

$$y^{(1)} = a_0{}^{(1)} + a_1{}^{(1)}x_{i1} + a_2{}^{(1)}x_{i2} + \cdots + a_p{}^{(1)}x_{ip} \qquad (2.40)$$

biçiminde ifade edilir. $y^{(1)}$ en büyük ayırma kriteri λ_1'e sahiptir.

λ_2, $W^{-1}B$ matrisinin öz değeri ve $a^{(2)} = (a_0^{(2)}, a_1^{(2)}, \dots, a_p^{(2)})$ ikinci özdeğere karşılık gelen öz vektör (katsayılar) ise, ikinci ayırma fonksiyonu,

$$y^{(2)} = a_0{}^{(2)} + a_1{}^{(2)}x_{i1} + a_2{}^{(2)}x_{i2} + \cdots + a_p{}^{(2)}x_{ip} \qquad (2.41)$$

olarak elde edilir. $y^{(2)}$ de ikinci en büyük ayırma kriteri λ_2'e sahiptir. $y^{(2)}$ ayırma fonksiyonu ile $y^{(1)}$ ayırma fonksiyonu arasındaki korelasyon sıfırdır.

Benzer şekilde,

$$y^{(3)} = a_0{}^{(3)} + a_1{}^{(3)}x_{i1} + a_2{}^{(3)}x_{i2} + \cdots + a_p{}^{(3)}x_{ip} \qquad (2.42)$$

$y^{(1)}$ ve $y^{(2)}$ ile korelasyonu sıfır olan üçüncü en büyük ayırma kriterine sahip $y^{(3)}$ ayırma fonksiyonu elde edilir. $y^{(3)}$ ikinci en büyük ayırma kriteri λ_3'e sahiptir.

Benzer şekilde r. ayırma fonksiyonu $y^{(r)}$, λ_r'ye karşılık gelen $a^{(r)} = (a_0^{(r)}, a_1^{(r)}, \dots, a_p^{(r)})$ ağırlıkları kullanılarak elde edilir. $y^{(r)}$, $y^{(1)}$, $y^{(2)}$,…, $y^{(r-1)}$ ayırma fonksiyonları ile korelasyonu sıfır olan en büyük ayırma kriteri $y^{(r)}$'ye sahiptir.

Bu şekilde elde edilen tüm ayırma (ayırma) fonksiyonları arasındaki korelasyon 0'dır.

Bu aşamadan sonra istatistiksel olarak anlamlı olan ayırma fonksiyonları belirlenmelidir.

2.5.2. Ayırma fonksiyonlarının önem kontrolü: İkiden çok grup olması durumunda elde edilen ayırma fonksiyonunun önem kontrolünde kullanıla kriterlerden ilki, Wilks tarafından geliştirilmiş olan ve genelleştirilmiş varyans olarak bilinen Λ'dır.

Çok değişkenli normal dağılımlı ve ortak varyans-kovaryans matrisli iki ana kütleden gelen örneklemler için Mahalonobis Uzaklık Değeri-D^2 istatistiği, F dağılımına uyar ve F istatistiği yardımıyla bulunan ayırma fonksiyonlarının grupları ayırma özelliklerinin önemliliği kontrol edilebilir.

Grup sayısının ikiden fazla olduğu durumlarda bulunacak olan ayırma fonksiyonlarının önem kontrolünde Wilks'in Λ kriteri kullanılır. Λ kriteri ile ayırma fonksiyonlarının ayırma değerleri arasında cebirsel bir ilişki vardır. Bu ilişki,

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W+B|} \tag{2.43}$$

burada W grup içi, T ise toplam varyans matrisleridir. Önerilen Λ değerinin küçük bulunması gruplar arası farklılığın önemli olduğunun bir işaretidir ki bilindiği gibi bu değer aynı zamanda çok değişkenli varyans analizinin de temelini teşkil etmektedir.

Gruplardaki birey sayısının yeterince büyük olması durumunda m=n-1-1/2(p+k) olmak üzere, Λ değeri kullanılarak aşağıdaki test istatistik değeri bulunmaktadır.

$$X^2 = -mlog(\Lambda) \sim x^2_{p(k-1):\alpha} \tag{2.44}$$

Tatsuoka(1971), Cooley ve Lohnes (1972), Λ oranının ayırma fonksiyonlarının sayısını belirlemede kullanılabileceğini göstererek bir yöntem geliştirmişlerdir. Bu yöntem,

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

$$= |W^{-1}(W+B)| = |W^{-1}W + W^{-1}B|$$

$$= |1 + W^{-1}B| \tag{2.45}$$

$W^{-1}B$ matrisinin özvektörler matrisi P = $\left[e_1, e_2, \ldots, e_p\right]_{pxp}$ olsun. P ortogonalbir matris olduğundan,

$$\frac{1}{\Lambda} = |PP^{\iota}||1 + W^{-1}B| = |P^{\iota}(1 + W^{-1}B)P|$$

$$= |P^{\iota}P + P^{\iota}W^{-1}BP| = \text{Köş}[\lambda_i\,]$$

$$= \prod_{i=1}^{r}(1 + \lambda_i\,) \tag{2.46}$$

ve böylece

$$\Lambda = \prod_{i=1}^{r}\frac{1}{1 + \lambda_i} \tag{2.47}$$

elde edilir. Bu ifadenin doğal logaritması alınırsa,

$$In(\Lambda) = -In\left(\frac{1}{\Lambda}\right) = -In\left(\prod_{i=1}^{r}(1 + \lambda_i)\right) = -\sum_{i=1}^{r} In\,(1 + \lambda_i)$$

$$= -[(1 + \lambda_1) + (1 + \lambda_2) + \cdots + (1 + \lambda_r)] \tag{2.48}$$

olmak üzere (2.44) eşitliği,

$$X^2 = m\sum_{i=1}^{r} In(1 + \lambda_i)$$

$$= \left[n - 1 - \frac{1}{2}(p + k)\right]\sum_{i=1}^{r} In(1 + \lambda_i) \tag{2.49}$$

biçiminde gösterilebilir. Burada her bir λ_i'ye karşılık olarak X_i^2 değerinin bulunabileceği ve bunların da (p+k-2i) serbestlik derecesi ile yaklaşık khi-kare dağılacağı düşüncesinden hareketle,

$$X^2 = X_1^2 + X_2^2 + \cdots + X_r^2 \tag{2.50}$$

biçimindeki r tane ayırıcı fonksiyondan kaç tanesinin ele alınacağı ya da ihmal edilenlerin ayırmada önemli etkisinin bulunup bulunmadığının belirlenmesinde bu bilgilerden yararlanılmaktadır.

Örneğin i'inci ayırıcı fonksiyon için,

$$X_i^2 = \left[n - 1 - \frac{1}{2}(p+k)\right] In(1+\lambda_i) \tag{2.51}$$

olmak üzere toplam r tane olan ayırıcı fonksiyondan ilk m tanesinin önemli bulunmasından sonra geriye kalan (r-m) tane ayırıcı fonksiyonun önemliliğinin testinde,

$$X_{r-m}^2 = X^2 - \sum_{i=1}^{m} X_i^2 \tag{2.52}$$

olarak bulunan değer, p(k-1)-$\sum_{i=1}^{m}(p+k-2i)$ serbestlik dereceli khi-kare tablo değeri ile karşılaştırılmaktadır. Test işlemleri için önce hipotezler kurulur;

$H_0 =$ bulunan ayırma fonksiyonları önemsizdir.

$H_1 =$ bulunan ayırma fonksiyonlarından en az biri önemlidir.

Bulunan ayırma fonksiyonlarının test edilmesi için bulunan ilk fonksiyondan (en güçlü olan) başlamak üzere sıra ile devam edilir. Bu işlemler, ayırıcı fonksiyonların önemsizliği kabul edilinceye kadar devam ettirilir.

$X_1^2 \rightarrow$ birinci ayırma fonksiyonunu

$X_2^2 \rightarrow$ ikinci ayırma fonksiyonunu

$\vdots$

$X_r^2 \rightarrow$ r-inci ayırma fonksiyonunu

test etmektedir.

$$X_i^2 > X_{1-\alpha,p+k-2i}^2$$

olduğunda H_0 hipotezi reddedilir. Bu i-inci ayırma fonksiyonunun bireyleri sınıflara ayırmada önemli olduğu anlamına gelir.

r = min (k-1,p) tane ayırma fonksiyonundan m < r olmak üzere; m tanesinin önemli olduğunu varsayalım. Geriye kalan r – m tanesinin önemli olup olmadığını anlamak için şu hipotezler oluşturularak bu hipotezler de test edilirken, X^2 test istatistiği kullanılır.

H_0 = geriye kalan r-m tane ayırma fonksiyonu önemsizdir.

H_1 = geriye kalan r-m tane ayırma fonksiyonu önemlidir.

$$X_{r-m}^2 = X_{p-(k-1)}^2 - \sum_{i=1}^{m} X_i^2 - X_{p(k-1)-\sum_{i=1}^{m}(p+k-2i)}^2 \quad (2.53)$$

eğer,

$$X_{r-m}^2 \geq X_{1-\propto;p(k-1)-\sum_{i=1}^{m}(p+k-2i)}^2 \quad (2.54)$$

ise H_0 hipotezi reddedilir ve geriye kalan ayırma fonksiyonlarından en az bir tanesi ayırma için önemlidir kararına varılır. m'in sayısı bir arttırılarak test işlemi tekrarlanır.

$$X_{r-m}^2 < X_{1-\propto;p(k-1)-\sum_{i=1}^{m}(p+k-2i)}^2 \quad (2.55)$$

olduğunda ise H_0 hipotezi kabul edilir ve geriye kalan r-m tane ayırma fonksiyonunun ayırma için önemsiz olduğuna karar verilir.

Bu arada, r tüm ayırma fonksiyon sayısı ve m önemli bulunan ayırma fonksiyon sayısı olmak üzere, test işlemlerine m=r/2 alınarak başlanmasının zaman kaybını azaltması açısından önemli olduğu söylenmektedir (Tatlıdil, 1992: 212).

2.5.3. Özel durumlarda kullanılan ayırma analizi yöntemleri:

Şimdiye kadar incelediğimiz gerek Fisher yöntemi ve gerekse bunun genelleştirilmesi olan Anderson yöntemi kullanılarak doğrusal ayırma fonksiyonu bulunduğunda tanımlanan bu fonksiyondaki bireyler normal dağılımlı ve ana kütleler ortak varyans-kovaryans matrisine sahiptir. Doğrusal ayırma fonksiyonu olarak adlandırılan bu fonksiyonun,

$$f = f_{ij} = L(x) = \left\{x^{\iota} - \frac{1}{2}\left(\bar{x}^{(i)} + \bar{x}^{(j)}\right)^{\iota}\right\} S^{-1}\left(\bar{x}^{(i)} + \bar{x}^{(j)}\right) \quad (2.56)$$

kullanılabilmesi için verilerin normal dağılımlı ve kitlelerin ortak varyans-kovaryans matrisine sahip olması gereklidir. Bu varsayımlardan sapmalar durumunda alternatif fonksiyonlar kullanılmaktadır.

Karesel ayırma fonksiyonu: Verilerin normal dağıldığı ancak grupların varyans-kovaryans matrislerinin farklı olmaları durumunda kullanılan fonksiyondur. $Q(x) = \frac{1}{2} log \frac{|S_j|}{|S_i|} - \frac{1}{2}\left(\bar{x}^{(i)\iota} S_i^{-1} \bar{x}^{(i)} - \bar{x}^{(j)\iota} S_j^{-1} \bar{x}^{(j)} + x^{\iota}\left(S_i^{-1}\bar{x}^{(i)} - S_j^{-1}\bar{x}^{(j)}\right)\right) - \frac{1}{2} x^{\iota}\left(S_i^{-1} - S_j^{-1}\right)x$ (2.57)

Başlangıçta iki grup olması halinde geliştirilen bu fonksiyon, ikişerli alınarak çok grup olması durumunda da kullanılmaktadır. Fonksiyonda S_i veS_j sırasıyla i-inci ve j-inci gruba ilişkin varyans-kovaryans matrisleridir. $S_i = S_j$ = S alındığında karesel fonksiyon doğrusal fonksiyona eşit olur.

Fonksiyon değeri Q(x) < 0 ise bireyin R_j bölgesine değilseR_i bölgesine sınıflandığı karesel ayırma fonksiyonunda hatalı sınıflandırma olasılığı;

$$R_{Q(x)} = [1 + exp(Q(x) - log(\hat{q}_j/\hat{q}_i))]^{-1} \quad (2.58)$$

eşitliği ile ifade edilir.

Verilerin karışık olması durumunda yani bazılarının sürekli ve normal dağılımlı, bazılarının ise kesikli olması durumunda doğrusal ayırma fonksiyonuna bazı karesel terimler eklenerek yeni bir fonksiyon oluşturulmaktadır. Bu fonksiyona da **Yüksek derece terimli doğrusal ayırma fonksiyonu** (doğrusal ayırma fonksiyonu + karesel ayırma fonksiyonu) adı verilmektedir.

$$LQ(x) = b_0 + b_1 w_1 + \cdots + b_r w_r \quad (2.59)$$

biçimindeki fonksiyonda katsayılar birinci ve ikinci dereceden değişkenlere aittir. Bu yöntemde hatalı sınıflandırma olasılığı,

$$R_{LQ}(x) = [1 + exp(LQ(x) - log(\hat{q}_j/\hat{q}_i))]^{-1} \quad (2.60)$$

Bu fonksiyonun doğrusal ayırma fonksiyonunda olduğu gibi tahmin edilen katsayılar, doğrusal ayırma fonksiyonunun sonuçlarına çok yakın değerler almaktadırlar (Tatlıdil, 1992: 215).

Yüksek dereceli terimleri kapsayan fonksiyonda olduğu gibi değişkenlerden bazılarının sürekli, bazılarının kesikli olması durumlarında ayırma analizine alternatif olarak lojistik regresyon analizinden faydalanılabilir.

Pratikte hemen hemen tüm bilim dallarında yaygın kullanılmakta olan ayırma analizi aynı zamanda uygulamasında ve yorumunda çok da hata yapılan bir konudur. İkinci bölümde anlatılanlar özetlenecek olursa Ayırma Analizi;

- Hangi ana kitleden geldiği belli olmayan herhangi bir bireyin uygun kitleye atanmasıdır. Eski verilerden grup ya da grupların konum ve yayılma parametrelerinin bilinmesi durumunda yeni çekilen herhangi bir gözlemin bu gruplara atanıp atanmayacağına karar verme konusu ayırma fonksiyonları yardımıyla yapılmakta ve bu işlemlere karar amaçlı ayırma analizi adı verilmektedir.
- Ayırma analizi gelecekte kullanabilir fonksiyonlar da ortaya koyabilir. Bu özelliğiyle ayırma analizi kümeleme analizinden ayrılmaktadır. Gelecekte kullanılmasından ötürü bu ayırma analizine tahmin edici ayırma analizi denilmektedir.

BÖLÜM 3

3. İNSANİ GELİŞMİŞLİK ENDEKSİ VERİLERİ ÜZERİNE AYIRMA ANALİZİ UYGULAMASI

3.1. İnsani gelişme

Ülkelerin gelişmiş ve gelişmekte olan ülkeler olarak ikiye ayrıldığı II. Dünya Savaşı'ndan sonra ekonomi politikaları yeni bir boyut kazanmıştır. Bunun sonucunda gelişmiş ülkeler büyümeyi amaç edinirken gelişmekte olan ülkeler ise kalkınmayı esas amaç edinmiştir. II. Dünya Savaşı'ndan sonra birkaç on yıl büyüme ve kalkınma ölçümleri için sadece parasal parametreler kullanılmıştır. Bu dönemde ülkelerin milli geliri ve bu gelirden ülke vatandaşlarının aldığı pay, büyüme ve kalkınma politikaları hazırlanmasında birincil parametre olarak kullanılmıştır. Yalnız, milli gelir ve kişi başına düşen milli gelir rakamları yüksek olan ülkelerde yaşayan insanların özellikle gelir adaletsizliği nedeniyle ülkelerin gelişmişlik seviyeleriyle paralel refah içinde yaşayamadıkları, gelişmeye ve kalkınmaya ulaşamadıkları fark edilmiştir. Milli gelirdeki artışın toplumun ekonomik göstergeleri dışındaki eğitim, sağlık, sosyal, demografik vb. göstergeleri üzerinde etkisinin yetersizliğinden dolayı özellikle 1970'lerden sonra kullanılan parasal parametrelere ilave olarak ülkelerin sosyal, sağlık, eğitim, demografik vb. özelliklerini gösteren parametrelerin kullanılması gereği tartışmaları yapılmıştır. 1980'lere gelindiğinde "sürdürülebilir kalkınma" yeni bir kavram olarak önem kazanmış ve 1990'larda imzalanan uluslararası antlaşmalarla küresel bir uygulama planı haline gelmiştir. 1987'de Birleşmiş Milletler Dünya Komisyonu tarafından hazırlanan Brundtland Raporuna göre iktisadi kalkınma, sosyal eşitlik ve çevre üzerine dikkat çeken, odağına insanı yerleştiren çok boyutlu bir yaklaşıma geçilmiş ve bütüncül gelişme parametreleri üzerinde durulmuştur.

Sürdürülebilir kalkınma görüşü, ülkelerin sosyal ve ekonomik gelişme politikalarında ortak yanı "sürdürülebilirlik" olarak belirlemektedir. Sürdürülebilirliğin çok boyutlu yapısındaki anlayışa bağlı olarak oluşturulan politikaların geri bildirimlerini gerçeğe en yakın şekilde hesaplayabilmek için bütün toplumsal yapıları birer faktör olarak hesaplama parametrelerine dahil eden analizler gerekmektedir. Ekonomik

kalkınmanın, sosyal parametrelere tam olarak daima pozitif yönde etkilemediği ya da istenen düzeyde pozitif bir etkisinin olmadığını savunan görüşler teorik olarak halen tartışılmakta ve uygulamada çeşitli istatistiksel analizler yapılmaktadır. Bu analizlere çok büyük bir katkı da 1990 yılında başlanarak Birleşmiş Milletler Kalkınma Programı (United Nations Development Programme – UNDP) tarafından dünya ülkeleri dahil edilerek hesaplanan ve her yıl çeşitli parametreler ve faktörler de eklenerek raporlanan ve yıllık bazda yayımlanan gelişme raporlarıyla sağlanmaktadır. İlki 1990 yılında yayımlanan ve her yıl düzenli olarak yayımlanan bu raporlara İnsani Gelişme Endeksi (Human Development Index –HDI) adı verilmektedir.

3.2. İnsani gelişme endeksi

Milli gelirde ciddi bir artışın olması bir ülke için gelişmiş olarak ifade edilmesinde yeterli değildir. Ekonomik yönden gelişmiş birçok ülkenin çeşitli sosyal sorunlarını çözemediklerinin görülmesiyle ekonomik büyüme ve insani gelişme arasındaki ilişkinin daha gerçekçi olarak doğru bir şekilde kurulması gereği ortaya çıkmıştır. Sonuç olarak kalkınma fiziki ve ekonomik kalkınmanın yanı sıra, insanların seçeneklerini artırmasını da içerecek bir kavrama evirilmiştir. Daha önce de ifade edildiği gibi ilki 1990 yılında yayımlanan İnsani Gelişme Raporlarında (İGR), gelir dışında insani gelişmişliği ölçmek için gelir dışı parametrelerin de kullanıldığı bir takım endeksler yayınlanmaya başlanmıştır. Kişilerin seçeneklerini artırma süreci olan bu endekslerdeki temel anlayışın kısıtlarından en önemlisi bu seçeneklerin sonsuz ve değişken olmasıdır. Eğitim, sağlık, sosyal güvenlik hizmetlerine ve temiz suya erişim, elektrik ve internet hizmetlerine erişim, istihdam, toplumsal faaliyetlerde yer alabilme ile karar alma süreçlerine katılabilme bu parametrelerden sadece bazılarıdır. Yalnız bu yaşam standartlarının belirlenmesinde yukarda belirtiğimiz alanlardaki parametrelerle ilgili ve mevcut durumda olan verileri bulmak oldukça zordur. Sonuçta, insani gelişmişlik için yaşam standartlarının belirlenmesinde olabildiğince kapsamlı parametrelerin içerilmiş olması gerekmektedir. Ancak, veriler söz konusu olduğunda bazı parametrelerden vazgeçmek durumunda kalınmaktadır. Tüm bunlar düşünüldüğünde gelirin bu hesaplamalarda sıklıkla kullanılan bir parametre olmasının temel nedeninin ölçülebilir ve ulaşılabilir bir veri olmasından kaynaklandığı görülecektir.

3.2.1. İnsani gelişme endeksi bileşenleri: Birleşmiş Milletler Kalkınma Programı (UNDP), ölçülebilir ve ulaşılabilir parametreler olması dolayısıyla gelir, sağlık, cinsiyet farklılıkları, eğitim, çevre, elektrik ve internet hizmetlerine erişim, istihdam vb. insani gelişmeyi ölçmeyi amaçlayan parametreler kullanmaktadır. İnsani Gelişim Endeksi hesaplamalarında kullanılan sağlık, gelir ve eğitim faktörleri bütün gelişim aşamalarında öne çıkan üç temel bileşendir. Bu üç temel faktörü ölçmek için kullanılan değişkenler ise doğumda beklenen yaşam süresi, beklenen okullaşma yılı, ortalama okullaşma yılı ve kişi başına Gayri Safi Milli Gelir (GSMG)'dir. İGE, toplumun yaşam kalitesini belirlemek için sosyo-ekonomik bir endeks olarak dünyaca kabul görmektedir. İGE birbiriyle ilişkili boyutları kapsayacak üç faktöre bölümlenir. Bu faktörler, topluma tam katılım için gerekli olan eğitim, sağlık ve asgari yaşam standardını yakalamayı sağlayacak düzeyde bir gelirdir.

3.2.2. İnsani gelişme endeksi metodolojisi: İnsani Gelişme Endeksi'nin hesaplanmasında ana unsur olan üç endeksin hesaplanma yöntemleri aşağıdaki gibidir.

1. Gelir Endeksi: kişinin seçeneklerini artırmasına temel teşkil eden ve yaşam standardının kalitesine ifade eden endeks, kişi başı GSMH esas alınarak hesaplanır.

$$I_I = \frac{In(GNI) - In(GNI)_{min}}{In(GNI)_{max} - In(GNI)_{min}} \tag{3.1}$$

Formüldeki GNI, logaritması alınmış satın alma gücü paritesine (SGP) göre uyarlanan kişi başına GSMH' yı, $In(GNI)_{min}$100 değerine eşitliği temsil eden notasyondur.

2. Sağlık Endeksi: Sağlıklı ve uzun bir yaşamı ifade eden endeksi doğumda yaşam beklentisini kullanarak formüle ederiz.

$$I_H = \frac{L - L_{min}}{L_{max} - L_{min}} \tag{3.2}$$

Formüldeki L doğumda yaşam beklentisini, L_{min} 25yaş sınırını, L_{max} 85 yaş sınırını gösteren notasyondur.

3. Eğitim Endeksi: bilgiye ulaşmayı ifade eden endeksi ortalama okullaşma yıllarını ve okullaşma yıllarını kullanarak hesaplarız.

$$I_E = \frac{\left(\frac{MYS - MYS_{min}}{MYS_{max} - MYS_{min}} + \frac{EYS - EYS_{min}}{EYS_{max} - EYS_{min}}\right)}{2} \quad (3.3)$$

Formüldeki MYS ortalama okullaşma yılını, EYS beklenen okullaşma yılını gösteren notasyonlardır. Hesaplama yapılırken EYS ve MYS'nin eşit ağırlıklandırılmış ortalamaları alınır.

İGE, ölçüm parametreleri faktörlerin her biri için hesaplanmış $I_I, I_H \; ve \; I_E$ endeks değerlerinin geometrik ortalaması ile hesaplanır.

$$HDI = \sqrt[3]{I_I \times I_H \times I_E} \quad (3.4)$$

İfade edilen şekilde hesaplanan endeks değeri 0 ile 1 arasında bir değer alır. Bu değerin 0!a yakın olması durumunda düşük insani gelişmişliği, 1'e yakın olması yüksek insani gelişmişliği ifade eder.

Bu şekilde hesaplanan ülkelerin İGE skorları, ülkeleri gelişmişlik düzeyine göre dört gruba ayırır. Bu dört grup aşağıdaki tablodaki gibidir.

Şekil 10. İGE skorlarına göre ülkelerin gelişmişlik seviyeleri

İnsani Gelişmişlik Seviyesi	Skor Aralığı
Düşük İnsani Gelişmişlik	0-0.479
Orta İnsani Gelişmişlik	0.480-0.670
Yüksek İnsani Gelişmişlik	0.671-0.780
Çok Yüksek İnsani Gelişmişlik	0.781-1

3.3. Ayırma analizinin İGE verilerine uygulanması

3.3.1. Araştırmanın kapsamı: Çalışmada Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından 2018 yılında yayımlanan İnsani Gelişme Raporu (HDR)'nun hesaplanmasında ve hazırlanmasında kullanılan çeşitli veriler kullanılmıştır.

Farklı dört grupta bulunan ülkelerden rastgele seçilen 55 ülke verileri üzerine uygulama yapılmıştır.

3.3.2. Analize alınan değişkenler ve tanımları: Ayırma analizi uygulaması için analize seçilen değişkenler ve seçilmiş bu değişkenlerin tanımları şöyledir:

Toplam Nüfus: Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından yayınlanan 2018 İnsani Gelişme Endeksi (HDI) raporunda yer alan seçilmiş ülkelere ait 2017 nüfuslarıdır.

Tüberküloz Oranı: Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından yayınlanan 2018 İnsani Gelişme Endeksi (HDI) raporunda yer alan seçilmiş ülkelere ait 2016 yılı verilerine göre her 100000 kişide görülen tüberküloz hastalığı sayısını göstermektedir.

Toplam Doğurganlık Oranı: World Bank Data'da yer alan seçilmiş ülkelere ait 2017 yılı için ülkelerdeki kadın sayısı başına doğurganlık oranını göstermektedir.

Doğumda Yaşam Beklentisi: Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından yayınlanan 2018 İnsani Gelişme Endeksi (HDI) raporunda yer alan seçilmiş ülkelere ait 2017 yılı verilerine göre doğumdan sonra insanlarda beklenen yaşam süresini yıl olarak göstermektedir.

Gayri Safi Yurtiçi Hasıla: World Bank Data'da yer alan seçilmiş ülkelere ait 2017 yılı verilerine göre ülkelerin ülke sınırları içerisinde üretmiş oldukları tüm mal ve hizmetlerin değerini USD cinsinden göstermektedir.

Elektrik Tüketimi: World Bank Data'da yer alan seçilmiş ülkelere ait 2017 yılı verilerine göre ülkelerin kişi başına tükettikleri elektrik miktarını kilowatt-saat (kW-h) cinsinden göstermektedir.

İlköğretim Brüt Kayıt Oranı: Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından yayınlanan 2018 İnsani Gelişme Endeksi (HDI) raporunda yer alan seçilmiş ülkelere ait ilköğretim çağındaki nüfusun yüzdesi olarak 2012-2017 yılları arasındaki ilköğretim brüt kayıt oranını göstermektedir.

Toplam İşsizlik: Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından yayınlanan 2018 İnsani Gelişme Endeksi (HDI) raporunda yer alan seçilmiş ülkelere ait 2017 yılı toplam işgücünün yüzdesi olarak verilmiş değerdir.

3.3.3. Araştırmaya alınan ülkelerin analiz öncesi grup üyeliklerinin gösterilmesi: Çalışmada yer alan ülkeler 2018 yılında UNDP tarafından yayımlanan 2018 HDR' daki dört gruptan, seçilmiş değişkenler doğrultusunda eksik veri olmayacak şekilde seçilmiş 55 ülkeden oluşmaktadır. Bu gruplar ve grup üyelikleri aşağıdaki gibidir.

Tablo 2. Ayırma analizi için seçilmiş ülkeler

ÇOK YÜKSEK İNSANİ GELİŞMİŞ ÜLKELER (1. GRUP)
Norveç
Avustralya
İrlanda
İsviçre
Danimarka
Amerika Birleşik Devletleri
Birleşik Krallık
Finlandiya
İsrail
Fransa
Estanya
Kıbrıs
Litvanya
Şili
Macaristan
Hırvatistan
Arjantin
Bulgaristan

YÜKSEK İNSANİ GELİŞMİŞ ÜLKELER (2. GRUP)
İran İslam Cumhuriyeti
Kosta Rika
Türkiye
Arnavutluk
Meksika
Sri Lanka
Brezilya
Azerbaycan
Lübnan
Ermenistan
Tayland
Ukrayna
Peru
Kolombiya
Jamaica

ORTA İNSANİ GELİŞMİŞ ÜLKELER (3. GRUP)
Güney Afrika
Endonezya
Vietnam
Bolivya
El Salvador
Kırgızistan
Guatemala
Hindistan
Honduras
Bangladeş
Gana
Kamboçya
Nepal
Pakistan

DÜŞÜK İNSANİ GELİŞMİŞ ÜLKELER (4. GRUP)
Tanzanya
Zimbabve
Benin
Senegal
Togo
Fil Dişi Sahili
Etiyopya
Kongo Demokratik Cumhuriyeti

3.3.4. Analiz: Yukarıda verilen seçilmiş BM üyesi 55 ülkenin İGE verilerine 8 değişken üzerinden ayırma analizi uygulanmış ve seçilmiş bir başka ülkenin bu dört gruptan hangisine dâhil olacağı araştırılmıştır.

Çalışmada SPSS 22 paket programı kullanılarak analiz sonucu elde edilen sonuçlar istatistiksel tablolar halinde düzenlenmiş ve bu tablolar aşağıda yorumlanmıştır.

Gruplara ait her bir değişkene göre ortalama ve standart sapmalar aşağıda Tablo 3'deki gibidir.

Tablo 3. Toplam ve Gruplara Ait Her Bir Değişkene Göre Ortalama ve Standart Sapmalar

GRUP İSTATİSTİKLERİ

GRUPLAR		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
1.0	TOPLAM NÜFUS	33980545.2222222240000	75597810.35294619000000	18	18.000
	TÜBERKÜLOZ ORANI	12.4944444444444	12.12515480663927	18	18.000
	TOPLAM DOĞURGANLIK ORANI	1.8388888888889	.35169988691712	18	18.000
	DOĞUMDA YAŞAM BEKLENTİSİ	79.7045257452574	2.78696459187578	18	18.000
	İLKÖĞRETİM BRÜT KAYIT ORANI	102.0981050000000	6.374664950151111	18	18.000
	GSYİH	1712171539828.6736000000000	4523663764241.81400000000000	18	18.000

	ELEKTRİK TÜKETİMİ	7695.997518377233	5321.21682904336800	18	18.000
	TOPLAM İŞSİZLİK	7.1666666666667	2.63996434913540	18	18.000
2.0	TOPLAM NÜFUS	49762631.6666666640000	58130621.17611709000000	15	15.000
	TÜBERKÜLOZ ORANI	48.066666666667	47.06032096707172	15	15.000
	TOPLAM DOĞURGANLIK ORANI	1.8333333333333	.24397501823713	15	15.000
	DOĞUMDA YAŞAM BEKLENTİSİ	75.9347317073171	2.32756058007720	15	15.000
	İLKÖĞRETİM BRÜT KAYIT ORANI	105.0000000000000	7.48331477354788	15	15.000
	GSYİH	872218421403.9238000000000	1050818249098.23170000000000	15	15.000
	ELEKTRİK TÜKETİMİ	2133.4620245733336	805.27946144309710	15	15.000
	TOPLAM İŞSİZLİK	8.5333333333333	4.62086365870937	15	15.000
3.0	TOPLAM NÜFUS	160076099.3571428700000	349191331.93155270000000	14	14.000
	TÜBERKÜLOZ ORANI	217.3571428571429	194.37735193446520	14	14.000
	TOPLAM DOĞURGANLIK ORANI	3.9357142857143	5.02724992001012	14	14.000
	DOĞUMDA YAŞAM BEKLENTİSİ	70.1622857142857	3.90873228373759	14	14.000
	İLKÖĞRETİM BRÜT KAYIT ORANI	107.0000000000000	10.22816623913518	14	14.000
	GSYİH	1186277663191.9995000000000	2566415791999.84860000000000	14	14.000
	ELEKTRİK TÜKETİMİ	972.4730650785714	1044.05147205233860	14	14.000
	TOPLAM İŞSİZLİK	5.2428571428571	6.66075928747724	14	14.000
4.0	TOPLAM NÜFUS	39907006.5000000000000000	36793935.01311579000000	8	8.000
	TÜBERKÜLOZ ORANI	174.1250000000000	98.08223590436751	8	8.000

	TOPLAM DOĞURGANLIK ORANI	4.6625000000000	.71301672991794	8	8.000
	DOĞUMDA YAŞAM BEKLENTİSİ	62.144250000000	4.3549662865680	8	8.000
	İLKÖĞRETİM BRÜT KAYIT ORANI	103.250000000000	17.902513789968	8	8.000
	GSYİH	83169155878.306	66882243260.005	8	8.000
	ELEKTRİK TÜKETİMİ	195.88152546500	154.67453362271	8	8.000
	TOPLAM İŞSİZLİK	3.5125000000000	1.4227111944653	8	8.000
Total	TOPLAM NÜFUS	71243831.490909	187050101.69346	55	55.000
	TÜBERKÜLOZ ORANI	97.852727272727	136.54637233548	55	55.000
	TOPLAM DOĞURGANLIK ORANI	2.7818181818182	2.7604408470448	55	55.000
	DOĞUMDA YAŞAM BEKLENTİSİ	73.693244345898	6.8131797682593	55	55.000
	İLKÖĞRETİM BRÜT KAYIT ORANI	104.30483436363	9.8918380759117	55	55.000
	GSYİH	1112283719448.8	2934110179125.6	55	55.000
	ELEKTRİK TÜKETİMİ	3376.5738329856	4358.1369778599	55	55.000
	TOPLAM İŞSİZLİK	6.5181818181818	4.6584487526584	55	55.000

Tablo 4. Çok Değişkenli Test Sonuçları

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	,999	4600,316[b]	8,000	44,000	,000
	Wilks' Lambda	,001	4600,316[b]	8,000	44,000	,000
	Hotelling's Trace	836,421	4600,316[b]	8,000	44,000	,000
	Roy's Largest Root	836,421	4600,316[b]	8,000	44,000	,000
GRUPLAR	Pillai's Trace	1,438	5,297	24,000	138,000	,000
	Wilks' Lambda	,082	7,301	24,000	128,215	,000
	Hotelling's Trace	5,717	10,164	24,000	128,000	,000
	Roy's Largest Root	4,807	27,642[c]	8,000	46,000	,000

Tablo 4' deki test sonuçlarına göre değişkenlerimizin ülkeleri gruplara ayırdığı söylenebilir.

Tablo 5. Box's M istatistiği Sonuçları

Test Results[a]

Box's M		662,091
F	Approx.	6,691
	df1	72
	df2	4951,423
	Sig.	,000

Tablo 5'deki sonuçlara göre gruplarımızın kovaryans matrislerinin eşit olduğu söylenebilir.

Gruplar arası ayırımı m-1=4-1=3 ayırma fonksiyonu gerçekleştirmektedir. Bu 3 ayırma fonksiyonunun özdeğerleri, ayırma fonksiyonlarıyla açıklanan varyans yüzdeleri ve bunların birikimli yüzdeleri ile ayırma fonksiyonları ile ayırıcı değişkenler arasındaki ilişkiyi veren kanonik korelâsyonlar Tablo 7'de verilmiştir.

Ayırma fonksiyonlarımızın anlamlı olarak gruplarımızı birbirinden ayırıp ayıramadığını anlamamız için Wilks' Lambda ve Özdeğerler (Eigenvalue) tablolarımızı birlikte değerlendirmeliyiz.

Tablo 6. Ayırma Fonksiyonlarına Ait Wilks' Lambda ve Ki-Kare Değerleri

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	.082	119.926	24	.000
2 through 3	.477	35.488	14	.001
3	.766	12.770	6	.047

Tablo 7. Ayırma Fonksiyonlarına Ait Özdeğerler, Varyans, Birikimli Varyans ve Kanonik Korelasyon Değerleri

ÖZDEĞERLER

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	4.807	84.1	84.1	0.910
2	0.605	10.6	94.7	0.614
3	0.305	5.3	100.0	0.483

Tablo 6'da significiant (sig.) değerlerinin hepsi 0,05 den küçük olduğundan her üç fonksiyonumuzun da gruplarımızı birbirinden anlamlı olarak ayırdığını söyleyebiliriz.

Tablo 7'deki Canonical Correlation değerinin karesi açıklanan varyansımızı gösterir. 1. fonksiyonumuz için bu değer $0.910^2 = 0{,}83$'dir. Bu değer bağımlı değişkenimizde 1. fonksiyonun %83 varyans açıklayabildiğini gösterir. Aynı şekilde 2. fonksiyonun %38 ve 3. fonksiyonun %23 bağımlı değişkenimiz üzerinde varyans açıkladığı söylenebilir.

Eigenvalue değerleri de fonksiyonlarımızın tüm değişkenlerdeki varyansı açıklayabilme gücünü göstermektedir. Tablomuzdaki değerlere bakacak olursak 1. fonksiyon 2. fonksiyona göre 2. fonksiyon da 3. fonksiyona göre daha fazla tüm değişkenlerdeki varyansı açıklayabilme gücüne sahiptir diyebiliriz.

% of Variance değerimiz, toplam açıklanan varyansın ne kadarının yüzde olarak ayırma fonksiyonlarımız tarafından açıklandığını göstermektedir. Bu doğrultuda 1.

fonksiyonumuz toplam varyansın % 84.1'ini açıklarken 2. fonksiyonumuz % 10.6'sını ve 3. fonksiyonumuz % 5.3'ünü açıklamaktadır.

Standartlaştırılmış kanonik ayırma fonksiyonlarımıza ait değişken katsayıları Tablo 8'de gösterilmektedir.

Tablo 8. Ayırma Fonksiyonlarına Ait Değişken Katsayıları

Standartlaştırılmış Kanonik Ayırma Fonksiyon Katsayıları

	Function		
	1	2	3
TOPLAM NÜFUS	-.068	.265	.202
TÜBERKÜLOZ ORANI	-.247	.855	.724
TOPLAM DOĞURGANLIK ORANI	-.072	-.151	.054
DOĞUMDA YAŞAM BEKLENTİSİ	.842	.676	.159
GSYİH	.167	.018	-.144
ELEKTRİK TÜKETİMİ	.208	-.727	.618
İLKÖĞRETİM BRÜT KAYIT ORANI	.004	.356	-.001
TOPLAM İŞSİZLİK	.479	-.099	-.663

Bu tablomuz her bir bağımsız değişkenimizin 3 ayırma fonksiyonunda ayrı ayrı yüklerini görüyoruz. Bu değerlerin mutlak değerce en büyük değere sahip olanları fonksiyondaki en büyük yüke sahip olanı olduğunu söyleyebiliriz. Örneğin; 1. fonksiyonda Doğumda Yaşam Beklentisi değişkenimizin en büyük yüke sahip olan değişken olduğu söyleyebilirken İlköğretim Brüt Kayıt Oranı değişkenimizin en az yüke sahip olan değişkenimiz olduğu söylenebilir. Benzer şekilde 2. fonksiyonumuz için Tüberküloz Oranı değişkeni en büyük yüke sahipken GSYİH değişkenimiz en az yüke sahip olan değişkenimizdir. 3. fonksiyonumuz için ise Tüberküloz Oranı değişkeni en büyük yüke sahipken İlköğretim Brüt Kayıt Oranı değişkenimizin en az yüke sahip olan değişkenimiz olduğu söylenebilir. Bu şekilde büyükten küçüğe sıraladığımız değişkenlerimizi aynı zamanda fonksiyon üzerlerindeki yüklerine göre de sıralamış oluruz.

Her bir değişkenimizin her bir ayırma fonksiyonumuzla ne kadar korelasyona sahip olduğu Tablo 9'da gösterilmiştir.

Tablo 9. Ayırma Fonksiyonlarındaki Her Bir Değişkene Ait Yapı Matrisi

Yapı Matrisi

	Function		
	1	2	3
DOĞUMDA YAŞAM BEKLENTİSİ	.868*	.193	.148
TÜBERKÜLOZ ORANI	-.336	.418*	.408
TOPLAM NÜFUS	-.060	.301*	.218
İLKÖĞRETİM BRÜT KAYIT ORANI	-.042	.228*	-.029
ELEKTRİK TÜKETİMİ	.392	-.500	.684*
TOPLAM İŞSİZLİK	.156	.120	-.346*
TOPLAM DOĞURGANLIK ORANI	-.209	.014	.246*
GSYİH	.069	.043	.194*

1. fonksiyonumuzla en yüksek korelasyon değerine sahip değişkenimiz DOĞUMDA YAŞAM BEKLENTİSİ değişkenimizdir. 2. ve 3. fonksiyonlarımızla en yüksek korelasyon değerine sahip değişkenimiz ELEKTRİK TÜKETİMİ değişkenimizdir.

Tablo 10. Analiz sonrası sınıflandırma sonuçları

Sınıflandırma Sonuçları

			Predicted Group Membership				
		GRUPLAR	1.0	2.0	3.0	4.0	Total
Original	Count	1.0	13	5	0	0	18
		2.0	1	13	1	0	15
		3.0	0	3	10	1	14
		4.0	0	0	0	8	8
	%	1.0	72.2	27.8	.0	.0	100.0
		2.0	6.7	86.7	6.7	.0	100.0
		3.0	.0	21.4	71.4	7.1	100.0
		4.0	.0	.0	.0	100.0	100.0

Sınıflandırma sonuçları tablomuz (Tablo 10) analiz sonrası gruplandırmaları göstermektedir. Buna göre;

- 1. gruptaki 18 ülkenin 13'ü doğru gruplandırılırken 5'i 2. gruba sınıflandırılarak yanlış sınıflandırılmıştır. Yani 1. gruptaki ülkelerin %72,2'si doğru gruplandırılırken %27,8'i yanlış gruplandırılmıştır.
- 2. gruptaki 15 ülkenin 13'ü doğru gruplandırılırken 1 ülke 1. gruba, 1 ülke de 3. Gruba sınıflandırılarak yanlış sınıflandırılmıştır. Yani 2. gruptaki ülkelerin %86,7'si doğru gruplandırılırken %13,4'ü yanlış gruplandırılmıştır.
- 3. gruptaki 14 ülkenin 10'u doğru gruplandırılırken 3 ülke 2. gruba, 1 ülke de 4. gruba sınıflandırılarak yanlış sınıflandırılmıştır. Yani 3. gruptaki ülkelerin %71,4'ü doğru gruplandırılırken %28,6'sı yanlış gruplandırılmıştır.
- 4. gruptaki 8 ülkenin 8'i de doğru gruplandırılmıştır.

Sonuçta ülkelerin %80'ninin doğru gruplandırıldığı söylenebilir.

Tablo 8'deki veriler kullanılarak ayırıcı faktör eşitliklerimiz şu şekilde oluşturulur:

$$f_1 = -0{,}068z_1 - 0{,}247z_2 - 0{,}072z_3 + 0{,}842z_4 + 0{,}167z_5 + 0{,}208z_6 + 0{,}004z_7 + 0{,}479z_8$$

$$f_2 = 0{,}265z_1 + 0{,}855z_2 - 0{,}151z_3 + 0{,}676z_4 + 0{,}018z_5 - 0{,}727z_6 + 0{,}356z_7 - 0{,}099z_8$$

$$f_3 = 0{,}202z_1 + 0{,}724z_2 + 0{,}054z_3 + 0{,}159z_4 - 0{,}144z_5 + 0{,}618z_6 - 0{,}001z_7 - 0{,}663z_8$$

Buradaki z değişkenleri orijinal değişkenlerimizin standart normal dağılıma dönüştürülmüş formlarıdır.

Dört grup için Fisher'in doğrusal ayırma fonksiyon katsayıları aşağıda Tablo 11'de gösterilmiştir. Bu katsayılar ülkeleri, bağımsız değişkenlerimizin gruplara ayırmada hangisinin ne kadar katkıda bulunduğunu göstermektedir.

Tablo 11. Fisher'in Ayırma Fonksiyonlarının Katsayıları

Sınıflandırma Fonksiyon Katsayıları

	GRUPLAR			
	1.0	2.0	3.0	4.0
TOPLAM NÜFUS	-1.487E-8	-1.442E-8	-1.129E-8	-1.424E-8
TÜBERKÜLOZ ORANI	.122	.124	.143	.127
TOPLAM DOĞURGANLIK ORANI	.312	.274	.333	.510
DOĞUMDA YAŞAM BEKLENTİSİ	10.097	9.877	9.443	8.382
İLKÖĞRETİM BRÜT KAYIT ORANI	1.473	1.504	1.524	1.446
GSYİH	1.554E-12	1.542E-12	1.349E-12	1.240E-12
ELEKTRİK TÜKETİMİ	-.002	-.003	-.003	-.002
TOPLAM İŞSİZLİK	1.007	1.021	.560	.461
(SABİT)	-476.658	-460.448	-430.444	-349.603

Tablo 11'deki katsayılar kullanılarak dört grubumuz için Fisher'in ayırma fonksiyonlarını şu şekilde ifade edebiliriz.

$$f_1 = -476{,}658 - (1{,}487\text{x}10^{-8})X_1 + 0{,}122X_2 + 0{,}312X_3 + 10{,}097X_4 + 1{,}473X_5 + (1{,}554\text{x}10^{-12})X_6 - 0{,}002X_7 + 1{,}007X_8$$

$$f_2 = -460{,}448 - (1{,}442\text{x}10^{-8})X_1 + 0{,}124X_2 + 0{,}274X_3 + 9{,}877X_4 + 1{,}504X_5 + (1{,}542\text{x}10^{-12})X_6 - 0{,}003X_7 + 1{,}021X_8$$

$$f_3 = -430{,}444 - (1{,}129\text{x}10^{-8})X_1 + 0{,}143X_2 + 0{,}333X_3 + 9{,}443X_4 + 1{,}524X_5 + (1{,}349\text{x}10^{-12})X_6 - 0{,}003X_7 + 0{,}560X_8$$

$$f_4 = -349{,}603 - (1{,}424\text{x}10^{-8})X_1 + 0{,}127X_2 + 0{,}510X_3 + 8{,}382X_4 + 1{,}446X_5 + (1{,}24x10^{-12})X_6 - 0{,}002X_7 + 0{,}461X_8$$

Herhangi bir ülkeye ait sekiz değişkenin verileri bu fonksiyonlarda yerlerine ayrı ayrı yazılarak her bir fonksiyonun değeri bu ülke verileri için bulunur. Bunlardan maksimum olan değere Fisher'in karar kuralı gereği ülke sınıflandırılır.

3.3.5. Örnek seçilen ülkenin oluşturulan gruplara atanması: Birleşmiş Milletler Kalkınma Programı-UNDP tarafından 2018 yılında yayınlanan İnsani Gelişme Endeksi sıralamasında 2. Grupta (Yüksek İnsani Gelişmiş Ülkeler) yer alan ve analize almadığımız bir ülke olan Paraguay'a ait veriler Tablo 12' de gösterilmiştir.

Tablo 12. Paraguay'a Ait İGE Verileri

X_1 Toplam Nüfus	6811297
X_2 Tüberküloz Oranı	42.0
X_3 Toplam Doğurganlık Oranı	2.5
X_4 Doğumda Yaşam Beklentisi	73.21
X_5 İlköğretim Brüt Kayıt Oranı	106
X_6 GSYİH	36054281572
X_7 Elektrik Tüketimi	1563.505329
X_8 Toplam işsizlik	5.8

Bu değerler Fisher'in ayırma fonksiyonlarında yerine yazılarak fonksiyon değerleri şu şekilde bulunur:

$$f_1 = 5{,}8406$$

$$f_2 = 5{,}9218$$

$$f_3 = 3{,}2480$$

$$f_4 = 2{,}6738$$

Fisher'in karar kuralı gereği fonksiyon değerlerinden hangisi en büyükse ülkemiz o gruba atanmalıdır. Dört fonksiyon değerine baktığımızda en büyük değerin f_2 olduğu görülmektedir. Dolayısıyla Paraguay ayırma fonksiyonlarımız yardımıyla 2. gruba atanmalıdır. Paraguay 2. gruptan analize almadığımız test için seçtiğimiz bir ülke idi. Ayırma analizimiz sonucunda elde ettiğimiz fonksiyonlar sayesinde Paraguay ülkesi doğru gruplandırılmıştır.

3.3.6. Sonuç: Uygulama bölümünde Birleşmiş Milletler Kalkınma Programı (UNDP) tarafından yıllık olarak yayınlanan 2018 İnsani Gelişme Endeksi raporundan seçilmiş 55 ülkeye, 8 değişken üzerinden ayırma analizinin bütün varsayımlarının sağlanıp sağlanamadığı test edilerek analiz edilmeye çalışılmıştır.

Veri setimizde çok yüksek insani gelişmiş ülkeler 1, yüksek insani gelişmiş ülkeler 2, orta insani gelişmiş ülkeler 3 ve düşük insani gelişmiş ülkeler 4 olarak kodlanarak SPSS 22 paket programıyla ayırma analizi uygulanmıştır.

Ayırma analizi, temelinde üç önemli varsayıma dayanmaktadır. Bunlardan biri de verilerin çoklu normal dağılıma uygun olması gerektiği varsayımıdır. Test çoklu normallikten sapma durumuna karşı oldukça hassastır ve normallikten sapma durumunda grup varyans ve kovaryans matrislerinin de eşit olmama eğilimine sahip olduğu söylenebilir. Ayırma fonksiyonlarımızın optimal olabilmesi ve bireyleri sınıflandırmada yanlış sınıflandırma olasılığını minimize etmek için her grubun çoklu normal dağılıma sahip bir kitleden çekilmesi ve kovaryans matrislerinin eşit olması istenmiştir. Bunun için yapılan Box M testi sonucunda $\alpha < 0,05$ anlamlılık düzeyinde H_0 hipotezinin reddedilemeyerek grupların kovaryans matrislerinin eşit olduğu görülmüştür. Çoklu normallik ve kovaryans matrislerinin eşitliği test edildikten sonra analizimize alınan sekiz değişkenin çoklu doğrusal bağlantılarının olup olmadığına bakıldı. Bunun için bağımsız değişkenlerimizin korelasyon matrisleri incelenmiştir. Korelasyon matrisine bakıldığında bağımsız değişkenlerimiz arasında anlamlı bir ilişkinin olmadığı görülmüştür.

Sınıflandırma sonuçları tablomuza bakıldığında (Tablo 10) 1. gruptaki 18 ülkenin 13'ü, 2. gruptaki 15 ülkenin 13'ü, 3. gruptaki 14 ülkenin 10'u ve 4. gruptaki 8 ülkenin tamamının doğru sınıflandırıldığı görülmüştür. Doğru sınıflandırma yüzdesi böylece toplamda % 80 bulunmuş olup bu da iyi bir sınıflandırma oranıdır.

Analize alınmayan ve 2018 insani gelişme endeksi raporunda 2. grupta yer alan Paraguay Fisher'in ayırma fonksiyonları yardımıyla 2. gruba atanarak doğru sınıflandırılmıştır.

Uygulama sonucuna göre Ayırma Analizinin ülkeler için yapılacak olan insani gelişme endeksi sınıflandırılmasında kullanılabileceği söylenebilir.

KAYNAKLAR

Akgül, A. Ç., (2003) *İstatistiksel Analiz Teknikleri, SPSS'te İşletme Yönetimi Uygulamaları,* Emek Ofset, Ankara.

Akın, F.,. (2002) *Ekonometri,* Beta Basım, Bursa.

Akkaya, Ş., (1998) *Ekonometri II,* Erkam Matbaa, İstanbul.

Alpar, R., (2003) *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1*. Nobel Yayın Dağıtım, Ankara.

Cangül, O., (2006) *Diskriminant Analizi ve Bir Uygulama Denemesi*, (Yayınlanmamış Yüksek Lisans Tezi), Uludağ Üniversitesi, Ekonometri Anabilim Dalı, İstatistik Bölümü, Bursa.

Cooley, W. W., (1978) *Multivariate Data Analysis,* John Wiley & Sons, Inc., London.

Çakmak, Z., (1992) *Çoklu Ayırma ve Sınıflandırma Analizinin Eğitimde Öğrencilerin Meslek Seçimine* Uygulanması, Anadolu Üniversitesi Yayınları,Eskişehir.

Çokluk, Ö., vd., (2014) *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları,* (3. Baskı), Pegem Akademi, Ankara.

Doğan, H. G., Z., Gürler, (2013). "Türkiye'nin İnsani Gelişme Endeksinin Analitik Olarak Değerlendirilmesi" *, Iğdır Üniversitesi Fen Bilimleri Dergisi,* 2013/3(2), ss. 69-76.

Duyar, İ. T., "Antropometrik Boyutlarda Gözlenen Cinsiyet Farklılıklarının Diskriminant Analizi Kullanılarak İncelenmesi", *Morfoloji Dergisi*, 1998/6(1), ss.10-14.

Everitt, B. D., (1992) *Applied Multivariate Data Analysis,* Oxford Universty Press, New York.

Fisher, R. A., (1976) Discriminant Analysis, *Data Analysis Strategies and Designs for Substance Abuse Research*, 1976/1, ss. 201-220.

Geoffry, J. M., (2004) *Discriminant Analysis And Statistical Pattern Recognition,* John Wiley & Sons, New Jersey.

Gujarati, D. N. (2012) Temel Ekonometri, (1. baskı), Literatür Yayınları, İstanbul.

Gümüş, D., (2014) *Diskriminant Analizi ve Bireysel Emeklilik Üzerine Bir Uygulama,* (Yayınlanmamış Yüksek Lisans Tezi), İstanbul Üniversitesi, İşletme Anabilim Dalı, Sayısal Yöntemler Bilim Dalı, İstanbul,

Günsoy, G., "İnsani Gelişme Kavramı ve Sağlıklı Yaşam Hakkı", *ZKÜ Sosyal Bilimler Dergisi*, 2005/1,(2), ss. 35-52.

Gürses, D., ""İnsani Gelişme" ve Türkiye", *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,* Haziran 2019/12(21), ss. 339-350.

Hair, J. F. (2014). *Multivariate Data Analysis.* Harlow: Pearson.

Huberty, C. O., (2006). *Applied MANOVA and Discriminant Analysis,* Wiley Interscience, Harlow.

İnsani Gelişme Vakfı, <*http://ingev.org/hakkimizda/insani-gelisme-nedir*>, (24.03.2019).

Kalaycı, Ş., vd., (2010) *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri,* (5. Baskı), Öz Baran Ofset, Ankara.

Klecka, W. R., (1980), *Discriminant Analysis,* Sage Publication, London.

Kurtuluş, K., (1976), *Pazarlama Araştırmaları,* İstanbul Üniversitesi İşlatme Fakültesi Yayını No: 54, Sermet Matbaası, İstanbul.

Oruç, M., (1982), *İstatistik Yöntemler,* (1. Baskı), Ankara Üniversitesi Basımevi, Ankara.

Özdamar, K., (2002), *Paket Programlar İle İstatistiksel Veri Analizi-2,* Kaan Kitabevi, Eskişehir.

Öztürk, A, (2006) *Esnek Ayırma Analizi ve Bir Uygulama,* (Yayınlanmamış Doktora Tezi), Eskişehir Osmangazi Üniversitesi, Biyoistatistik Anabilim Dalı, Eskişehir.

Sangün, L, (2007) *Temel Bileşenler Analizi, Ayırma Analizi, Kümeleme Analizi ve Ekolojik Verilere Uygulanması Üzerine Bir Araştırma,* (Yayınlanmamış Doktora Tezi), Çukurova Üniversitesi, Su Ürünleri Anabilim Dalı, Adana.

Sharma, S., (1996) *Applied Multivariate Techniques,* John Willey & Sons, USA.

Tacq, J., (1997) *Multivariate Analysis Techniques,* Sage Pub., London.

Tatlıdil, H., (1992) *Uygulamalı Çok Değişkenli İstatistiksel Analiz,* Akademi Matbaası, Ankara.

Tatsuoka, M., (1971), *Multivariate Analysis: Techniques for Educational and Psychological Research,* John Wiley & Sons, Inc., New York.

UNDP, Human Development Indicators, <http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf>, (25.03.2019).

Ünalan, E., (2016) *Kamu Harcamaları ve İnsani Gelişme Arasındaki İlişk*i, (Yayınlanmamış Yüksek Lisans Tezi), Ankara Üniversitesi, Maliye Anabilim Dalı, Ankara.

Ünsal, A., "Diskriminant Analizi ve Uygulaması Üzerine Bir Örnek", *Gazi Üniversitesi, İ.İ.B.F. Dergisi*, 2000/1, ss. 19-36.

Ünver, Ö. G., (1999), Uygulamalı İstatistik Yöntemler, (3. Baskı), Siyasal Kitabevi, Ankara.

Yakut, E., B., Elmas, "İşletmelerin Finansal Başarısızlığının Veri Madenciliği Ve Diskriminant Analizi Modelleri İle Tahmin Edilmesi", *Afyon Kocatepe Üni. İİBF Dergisi*, 2013/15(1), ss. 261-280.