ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERDE BİPLOT TEKNİĞİ

Bilal Barış ALKAN

İSTATİSTİK ANABİLİM DALI

ANKARA
2011

Her hakkı saklıdır

ÖZET

Doktora Tezi

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERDE BİPLOT TEKNİĞİ

Bilal Barış ALKAN

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Doç. Dr. Cemal ATAKAN

Bu çalışmada, literatürde oldukça karmaşık bir teoriye sahip olan biplot tekniğinin gerçek verilere uygulanması ile birlikte daha anlaşılır bir şekilde sunulması amaçlanmıştır. Bunun yanısıra çok değişkenli istatistiksel yöntemlerde biplot tekniğinin kullanılmasının veri kümesinde saklı kalmış yapıların ortaya çıkarılmasında son derece başarılı olduğu vurgulanmıştır.

Çalışmanın birinci bölümünde, konuya giriş ve önceki çalışmalar ele alınmıştır. İkinci bölümünde, biplot tekniği hakkında temel teorik kavramlar, Gabriel (1971) ile Gower ve Hand (1996)'in biplot yaklaşımlarından bahsedilmiştir. Daha sonraki bölümlerde sırasıyla Temel Bileşenler Analizi Biplot, Kanonik Değişken Analizi Biplot, Uzaklık Analizi Biplot, Bağlantı Biplot, Alan Biplot için temel kavramlar ve gerekli matematiksel teori verilmekte ve istatistiksel araştırmalarda bu yöntemlerin kullanımının önemi vurgulanmaktadır. Çalışmanın sekizinci bölümünde ise ekonomi, gıda ve enerji alanından elde edilen veri kümelerine önceki bölümlerde ayrıntılı olarak incelenen yöntemlerin uygulamalarına yer verilmektedir. Son bölümde sonuç ve öneriler tartışılacaktır.

Ekim 2011, 111 sayfa

Anahtar Kelimeler: Tekil Değer Ayrıştırması, Temel Bileşenler Analizi Biplot, Kanonik Değişken Analizi Biplot, Uzaklık Analizi Biplot, Bağlantı Biplot, Alan Biplot.

ABSTRACT

Ph.D. Thesis

BIPLOT TECHNIQUE IN MULTIVARIATE STATISTICAL METHODS

Bilal Barış ALKAN

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Assoc. Prof. Dr. Cemal ATAKAN

In this study, it was attempted to evidently present the biplot technique which has a quite complex theory in literature with real-life applications and to demonstrate that the use of biplot and the multivariate statistical methods have proven extremely effective in discovering the covered structures in data sets.

In the first section of the study, the introduction to the topic and the preceding studies have been dealt with. In the second section, the fundamental theoretical concepts of biplot, also the Gabriel's (1971) and Gower and Hand (1996)'s biplot approaches were argued. In the following sections, the basic concepts of Principal Component Analysis Biplot, Canonical Variate Analysis Biplot, Analysis of Distance Biplot, Collinearity Biplot, Area Biplot and necessary mathematical theory were respectively given and it was focused on the significance of the use of these theories in statistical researches. In the eighth section, the applications of the methods that have been examined in details in the previous sections were explored by obtaining the data sets from the economy, food and energy fields. The last section is to discuss the conclusions and recommendations.

October 2011, 111 pages

Key Words: Singular Value Decomposition, Principal Component Analysis Biplot, Canonical Variate Analysis Biplot, Analysis of Distance Biplot, Collinearity Biplot, Area Biplot.

TEŞEKKÜR

Tez çalışmamın her aşamasında ilgi ve önerileri ile beni yönlendiren değerli danışman hocam, Doç. Dr. Cemal Atakan (Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü)'a sonsuz teşekkürlerimi sunarım.

Tez İzleme Komitesi üyelerim Doç. Dr. Mehmet YILMAZ (Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü) ve Yrd. Doç. Dr. Mehmet UYSAL (Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü)'a tez çalışmam süresince verdikleri katkılardan dolayı çok teşekkür ederim.

Doktora süresince gerek aldığım derslerde, gerekse ders dışı sohbetlerde engin bilgilerinden yararlandığım her biri alanlarında birer otorite olan Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü öğretim üyelerine teşekkürlerimi sunarım.

Ekonomi alanındaki uygulama çalışmalarında verdiği katkılardan dolayı Yrd. Doç. Dr. Afşin ŞAHİN (Gazi Üniversitesi Bankacılık ve Sigortacılık Yüksek Okulu, Bankacılık Bölümü)'e çok teşekkür ederim.

Çalışmalarım süresince birçok fedakarlık göstererek her zaman yanımda olan ve beni her koşulda destekleyen sevgili eşim, oğlum ve aileme sonsuz teşekkürlerimi sunarım.

Bilal Barış ALKAN
Ankara, Ekim 2011

İÇİNDEKİLER

ÖZET …….. i

ABSTRACT ……..... ii

TEŞEKKÜR ……..... iii

SİMGELER DİZİNİ ……. vii

ŞEKİLLER DİZİNİ ……..…… viii

ÇİZELGELER DİZİNİ …… ix

1. GİRİŞ VE ÖNCEKİ ÇALIŞMALAR ……..... 1

1.1 Giriş …….. 1

1.2 Önceki Çalışmalar ……... 5

2. BİPLOT TEKNİĞİNİN TEMELLERİ…….. 10

2.1 Gabriel'in Biplot Yaklaşımı…… 10

2.1.1 Veri matrisinin ayrıştırması……...……..… 14

2.1.2 Grafiksel gösterimin özellikleri……...……. 14

2.1.3 Veri matrisinin varyans ayrıştırması……..……... 15

2.2 Ölçeklendirilmiş Biplot Eksenleri…….. 16

2.2.1 Alfa (α) ölçekleme……... 17

3. TEMEL BİLEŞENLER ANALİZİ BİPLOT…….. 19

3.1 Temel Bileşenler Analizi ve Biplot Yapılandırılması...……. 19

3.1.1 Gözlem noktalarının gösterimi…….…….…….. 20

3.1.2 İnterpolasyon ve kestirim kavramı…….…… 21

3.2 Temel Bileşenler Analizi Biplot İçin Uyum Ölçüleri…..…….. 22

4. KANONİK DEĞİŞKEN ANALİZİ BİPLOT…….……. 28

5. UZAKLIK ANALİZİ BİPLOT……..……... 33

5.1 Temel Koordinat Analizi……..……... 34

5.2 Uzaklık Analizi...…….. 40

5.3 Uzaklık Analizi ile Biplot Kullanımı…….. 41

6. BAĞLANTI BİPLOT…… 44

6.1 Çoklu Doğrusal Regresyon Modeli……. 45

6.2 Çoklubağlantı ve Negatif Etkileri……... 45

6.3 Geleneksel Çoklubağlantı Teşhis Yöntemleri…… 46
6.3.1 Varyans büyüme faktörleri……..…….……... 47
6.3.2 Özdeğerlerin kullanımı……...…… 48
6.3.2.1 Koşul indeksleri……....…… 48
6.3.2.2 Koşul sayısı…….....…… 49
6.3.2.3 Varyans ayrışım oranları……..……..... 49
6.4 Görsel Çoklubağlantı Teşhis Yöntemleri…..……. 50
6.4.1 Tableplot……..... 50
6.4.2 Klasik ve bağlantı biplot…….. 52
6.4.2.1 Bağlantı biplot ile varyans ayrışım oranlarının görselleştirilmesi…… 52
6.4.2.2 Bağlantı biplot ile koşul indekslerin görselleştirilmesi……...…… 53
7. ALAN BİPLOT…….. 54
7.1 Alan Biplot Oluşturmada Temel Bilgiler……... 54
8. GERÇEK VERİ UYGULAMALARI…….. 57
8.1 Uygulama 1- Gabriel'in Biplot Yaklaşımı……. 57
8.2 Uygulama 2- Gower ve Hand'in Biplot Yaklaşımı……... 61
8.3 Uygulama 3- Kanonik Değişken Analizi Biplot……. 66
8.4 Uygulama 4- Uzaklık Analizi Biplot……... 72
8.5 Uygulama 5- Bağlantı Biplot……... 75
8.6 Uygulama 6- Alan Biplot……. 81
9. TARTIŞMA VE SONUÇ…….. 89
KAYNAKLAR……... 93
EKLER……... 99
EK 1 Türkiye Sanayi Üretim Endeksi (2005=100), Haziran, 2005 – Haziran, 2009 genel ve ana kategorileri veri kümesi…… 100
EK 2 Çeşitli meyve türlerinin 100 gramında bulunan besin bileşenleri veri kümesi ..…… 102
EK 3 Beş farklı besin grubunda yer alan toplam 105 ürünün 100 gramında bulunan beş kolin değişkeninin yer aldığı veri kümesi……. 103
EK 4 Üç bölgeden alınan ham petrol özellikleri veri kümesi……. 106
EK 5 Avrupa ve Merkez Asya Ülkelerinin Temel Göstergelerine İlişkin Veri Kümesi……. 108

EK 6 Yüksek Gelirli OECD Ülkelerinin Temel Ekonomik Göstergeleri Veri Kümesi…….... 109
ÖZGEÇMİŞ…… 110

SİMGELER VE KISALTMALAR DİZİNİ

n	Gözlem Sayısı
p	Değişken Sayısı
R	Tüm Uzay
S	Biplot Uzayı
TDA	Tekil Değer Ayrıştırması
TBA	Temel Bileşenler Analizi
KDA	Kanonik Değişken Analizi
UA	Uzaklık Analizi
TKA	Temel Koordinat Analizi
ÇBÖ	Çok Boyutlu Ölçekleme
LDA	Lineer Diskriminant Analizi
MANOVA	Çok Değişkenli Tek Yönlü Varyans Analizi
GKKA	Genelleştirilmiş Kanonik Korelasyon Analizi
KT	Kareler Toplamları
Kİ	Koşul İndeksi
VAO	Varyans Ayrışım Oranları

ŞEKİLLER DİZİNİ

Şekil 1.1	Nokta saçılım grafiği……..	3
Şekil 1.2	Biplot grafiği……..…….	4
Şekil 2.1	Biplot eksenlerinin ölçeklendirilmesinin geometrik gösterimi……..	17
Şekil 3.1	Tüm uzay ve Biplot uzayı arasındaki ilişki……	21
Şekil 3.2	Karesel toplamların dik veya dik olmayan kırılması…….	23
Şekil 6.1	Tableplot'un temsili gösterimi……...	51
Şekil 7.1	Alan Biplot'un geometrik yorumu…….	55
Şekil 8.1	Gabriel'in Biplot grafiği……	61
Şekil 8.2	TBA Biplot grafiği…….	63
Şekil 8.3	TBA Biplot grafiği, 21 gözlem için……...	65
Şekil 8.4	Beş grubun birinci ve ikinci kanonik değişkenlerinin nokta saçılım grafiği……....	69
Şekil 8.5	KDA Biplot grafiği……	70
Şekil 8.6	Grup ortalamaları için KDA biplot grafiği……	71
Şekil 8.7	Ham petrol veri kümesi için UA Biplot grafiği…….	74
Şekil 8.8	Bölge ortalaması için UA Biplot grafiği……	75
Şekil 8.9	Veri kümesinin Tableplot grafiği……...	79
Şekil 8.10	Klasik Biplot grafiği……..	80
Şekil 8.11	Bağlantı Biplot grafiği……...	80
Şekil 8.12	Veri kümesi için Alan Biplot grafiği…….	86
Şekil 8.13	Enflasyon hedefleyen ülkeler için Alan Biplot grafiği……..	87
Şekil 8.14	PIGS ülkeleri için Alan Biplot grafiği……...	88

ÇİZELGELER DİZİNİ

Çizelge 8.1	Tekil değerler ve toplam varyans açıklama oranları…….	60
Çizelge 8.2	Korelasyon değerleri…….	60
Çizelge 8.3	İki boyutlu TBA Biplot için eksen yeterlilikleri ve kestirimleri...	63
Çizelge 8.4	İki boyutlu TBA Biplot grafiksel yaklaşımındaki besin bileşenlerinin Koordinatları…….	64
Çizelge 8.5	KDA Biplot grafiğinde eksenlerin kestirimleri…….	69
Çizelge 8.6	Gözlemlerin gruplara göre frekansları……..	69
Çizelge 8.7	UA Biplot grafiğinde eksenlerin kestirimleri…….....	74
Çizelge 8.8	Veri kümesinin varyans büyüme faktörleri ve parametre tahminleri……..	76
Çizelge 8.9	Koşul indeksleri ve varyans ayrışım oranları...…….	77
Çizelge 8.10	Modifiye edilmiş koşul indeksleri ve varyans ayrışım oranları…	77
Çizelge 8.11	1990-2007 yılları temel makroekonomik göstergeleri ortalamaları veri kümesinin z-skorları……..	84
Çizelge 8.12	İki boyutlu Alan Biplot'daki değişkenler arasındaki korelasyonlar…….	84

1. GİRİŞ VE ÖNCEKİ ÇALIŞMALAR

1.1 Giriş

Günlük yaşamda karşılaşılan olgular tek bir değişkenin değil, çok sayıda değişkenin etkisi altındadır. Bu nedenle eğer tanımlanacak yapı birden fazla değişkenin etkisi altında ise, bu yapının gerçeğe en yakın şekilde yorumlanabilmesi için çok değişkenli istatistiksel metotların kullanımı önemlidir.

İstatistiksel çalışmalarda verinin görsel yapısını ortaya çıkarmak için genellikle grafiklere başvurulur. Chambers vd. (1983) çalışmalarında, herhangi bir istatistiksel analiz için grafiksel gösterimlerin önemini vurgulamış ve iyi seçilmiş bir grafik kadar güçlü herhangi bir istatistiksel araç olmadığının altını çizmişlerdir. Bu tanım çok boyutlu veri yapısının yorumlanmasında kullanılan çok değişkenli istatistiksel analizler için özellikle doğrudur. Nokta saçılım grafikleri, değişkenlerin iki çifti arasındaki ilişkilerin yorumlanmasında kullanılan güçlü bir grafiksel araçtır (Tukey 1977). Örneğin, değişkenlerin iki çifti için tanımlanan bir korelasyon, veri içeriğinin çok kaba bir özetidir. Halbuki nokta saçılım grafiği, veri hakkındaki tüm bilgiyi verebilir. Ancak değişken sayısı ikiden fazla ise grafiksel olarak değişkenler arası ilişkilerin yorumlanması güçleşir. Çok değişkenli bir veri kümesinde, değişkenler arası ilişkinin yorumlanması için daha az boyutlu (genellikle 2 veya 3) uzaya yaklaşımlar elde edilmeye çalışılır. Bu durum göz ardı edilebilecek bir bilgi kaybına neden olsa da, ilişkilerin daha kolay yorumlanabilmesini sağlar. Biplot adı verilen grafiksel bir teknik ile çok değişkenli veri yapısının geometrik olarak yorumlanması sağlanır.

Biplot, iki değişkenli verilerin analizinde kullanılan nokta saçılım grafiklerinin çok değişkenli eşleri olarak göz önüne alınabilir. Biplot tekniği, çok değişkenli bir veri kümesinin ayrıntılı bir şekilde özetlenmesi yanında değişkenler arasındaki ilişkileri belirleme ve birimlerin sınıflandırılmasında oldukça başarılı bir tekniktir. Bu teknik görsel bir grafik yardımıyla veriyi tüm yönleri ile incelemeye imkan sağlamaktadır. Biplot sadece grafiksel bir teknik değil, aynı zamanda ayrıntılı istatistiksel analizler için gerekli bir metodoloji ile desteklenmektedir. Biplot da ki “Bi” ifadesi grafik boyutunu

değil birimlerin ve değişkenlerin aynı grafikte gösterileceğini ifade etmektedir. Bu teknik, tekil değer ayrıştırması prensibine dayanmaktadır.

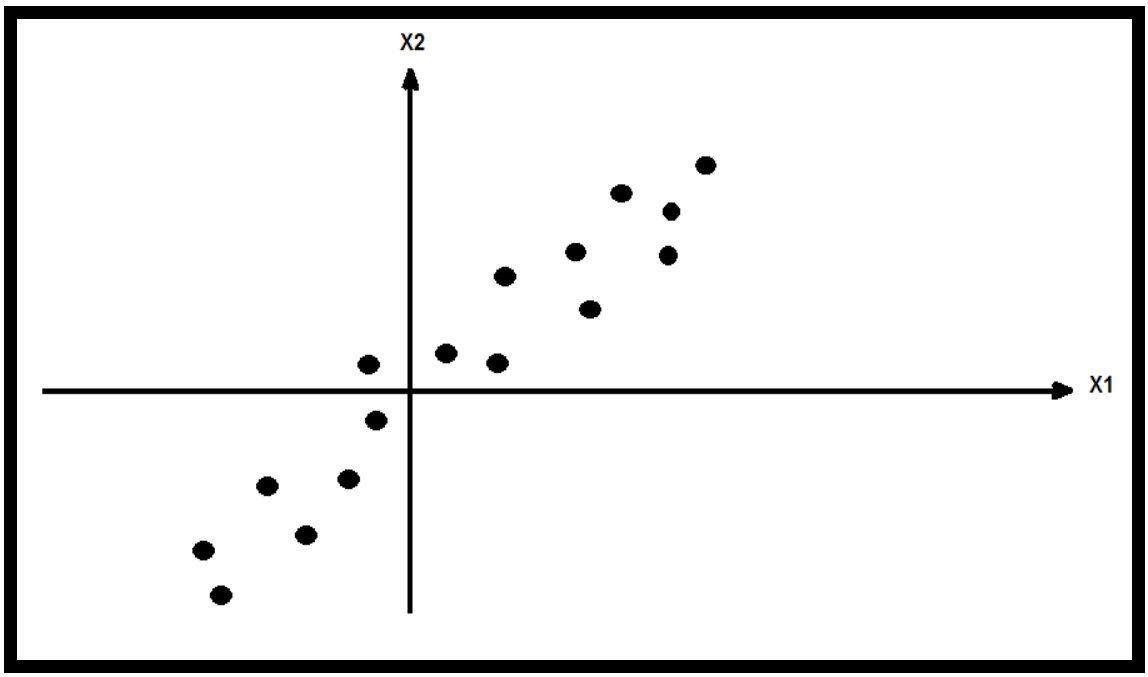
Biplotlar simetrik ve asimetrik olmak üzere iki ana türde incelenebilir. Simetrik biplotlar, iki yönlü bir tablonun satırları ve sütunlarıyla ilgili bilgi veren yaklaşımlar olarak tanımlanırken, asimetrik biplotlar ise bir veri matrisinin değişkenleri ve gözlem birimleri üzerine bilgi veren yaklaşımlar olarak tanımlanır. Simetrik biplotlarda, satırlar ve sütunlar bilgi kaybı olmaksızın yer değiştirebilirken, asimetrik biplotlarda obje türü farklı olduğundan böyle bir yer değişikliği mümkün değildir. Simetrik biplotlarda, hem gözlemler hem de değişkenler noktalar ile temsil edilirken asimetrik biplotlarda, gözlemler noktalar ile değişkenler ise vektörler ile temsil edilmektedir. Hangi biplot'un seçileceğine veri kümesinde yer alan değişken türlerine (nicel, nitel, sıralı vb.) bakılarak karar verilir (Gower vd. 2011).

İki değişkenli nokta saçılım grafiği genel literatürdeki ifadesiyle x-yatay eksen ve y-dikey eksen olmak üzere iki eksene sahiptir. Biplotlar ise değişken sayısı kadar eksene sahiptir. Şekil 1.1-1.2'de sırasıyla nokta saçılım ve biplot grafikleri örneklendirilmiştir.

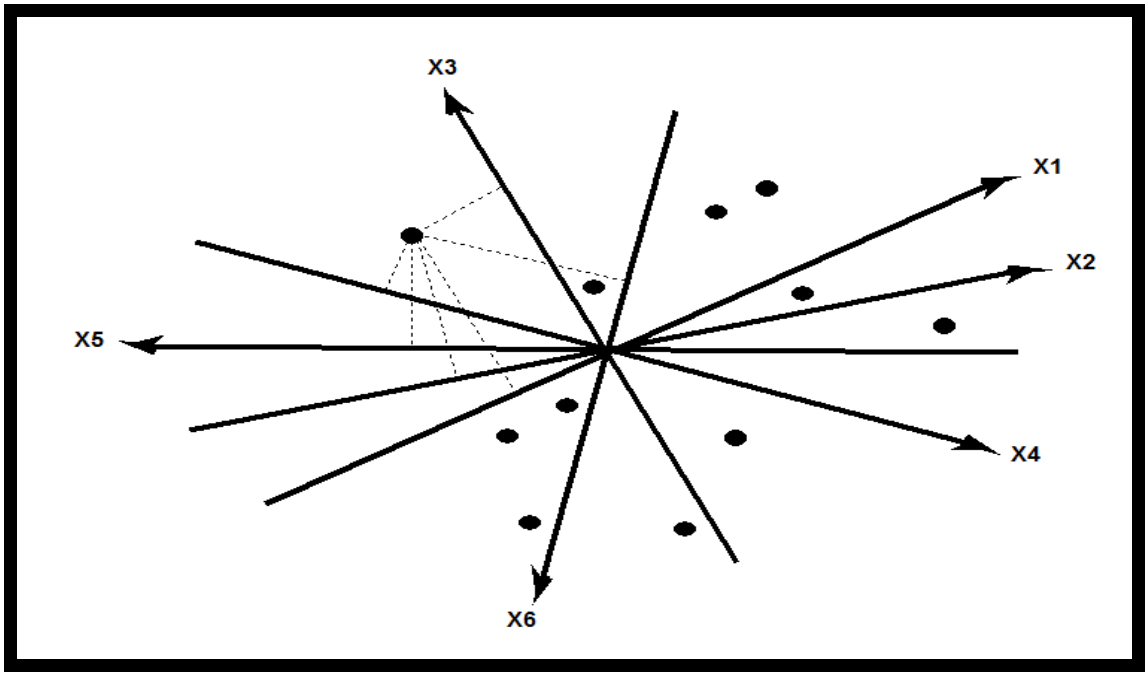
Bir nokta saçılım grafiğinde gözlemler noktalar ile gösterilir ve iki değişken üzerinde gözlemlerin dik izdüşümleri alındığında, gözlemlerin ilgili değişkenler üzerindeki değerleri elde edilir. Benzer olarak bir biplot'da, tüm değişkenler üzerinde herhangi bir gözlemin dik izdüşümü, ilgili gözlemin tüm değişkenler üzerinde aldığı değerleri verir. Nokta saçılım grafiklerinde ise gözlemlerin değişkenler üzerinde aldığı değerlere kesin ulaşılırken biplot'da bu mümkün değildir. Biplotlar indirgenmiş boyutlu bir uzayda (genellikle 2 veya 3 boyutlu) gösterilir. Biplot tekniği, indirgenmiş boyutlu bir uzayda değişkenler arasındaki korelasyonlardan yararlanır. Şekil 1.1'de verilen nokta saçılım grafiğinde X_1 ve X_2 değişkenleri yüksek pozitif korelasyona sahiptir. Böyle bir durum Şekil 1.2'deki biplot grafiğinde X_1 ve X_2 için söylenebilir. Ayrıca X_4 ve X_5 değişkenlerinin zıt yönlerde olup, bu değişkenlerin yüksek negatif korelasyona sahip olduğu düşünülür.

Bu tez çalışması, veri görselleştirilmesinde ve aşağıdaki soruların cevapları hakkında teorik yapının daha net anlaşılması ve uygulamadaki boşluğu ayrıntılı olarak doldurmayı amaçlamaktadır.

- Şekil 1.2'deki gözlem noktaları grafiğe nasıl yerleştirildi?
- Vektörlerle ifade edilen değişkenlerin yönü neden farklıdır?
- Biplot grafiğinin yorumu nasıl yapılacaktır?
- Farklı çok değişkenli istatistiksel yöntemlerde biplot tekniği nasıl kullanılır?



Şekil 1.1 Nokta saçılım grafiği



Şekil 1.2 Biplot grafiği

Çalışmanın ikinci bölümünde, biplot tekniği hakkında temel teorik kavramlar, Gabriel (1971) ile Gower ve Hand (1996)'in biplot yaklaşımları ele alınmıştır.

Üçüncü bölümde, Temel Bileşenler Analizi Biplot yapılandırması için temel kavramlar ve gerekli matematiksel bilgiler sunulmuştur.

Çalışmanın dördüncü bölümünde, gruplandırma ve diskriminant analizinde biplot tekniğinin kullanılmasına olanak sağlayan Kanonik Değişken Analizi Biplot için gerekli teorik yapı verilmiştir.

Beşinci bölümünde, veri kümesinde heterojenlik ve çarpıklık olduğunda çok değişkenli tek yönlü varyans analizine alternatif olarak Gower ve Krzonowski (1999) tarafından önerilen Uzaklık Analizi ile Biplot tekniğinin birlikte ele alınmasını sağlayan Uzaklık Analizi Biplot yöntemi için gerekli kavramlar ve teorik yapı verilmiştir.

Çalışmanın altıncı bölümünde, veri kümesindeki açıklayıcı değişkenler arasında güçlü yakın bağımlılığın var olmasının bir sonucu olarak ortaya çıkan çoklubağlantı probleminin teşhisinde kullanılan klasik yaklaşımlarla birlikte, tableplot ve bağlantı biplot olarak adlandırılan iki görsel çoklubağlantı teşhis yöntemi ele alınmıştır. Bu bağlamda gerekli teorik temeller bu bölümde verilmiştir.

Çalışmanın yedinci bölümünde, Gower vd. (2010) tarafından önerilen Alan Biplot tekniğinin teorik yapısı geometrik olarak ele alınmakta ve istatistiksel araştırmalarda Alan Biplot'un kullanımının önemi vurgulanmaktadır.

Çalışmanın sekizinci bölümünde ise ekonomi, gıda ve enerji alanından elde edilen veri kümelerine önceki bölümlerde ayrıntılı olarak incelenen yöntemlerin uygulamalarına yer verilmektedir.

Çalışmanın dokuzuncu bölümde, sonuçlar sunulmakta ve ileride yapılabilecek çalışmalar konusunda önerilerde bulunulmaktadır.

1.2 Önceki Çalışmalar

Gabriel (1971) bir veri matrisinin önemli özelliklerini göstermek için biplot tekniğini önermiştir. Bu teknik kısa zamanda birçok farklı alanda uygulanmış ve son derece kullanışlı olduğu ispatlanmıştır. Biyomedikal araştırmalarda (Gabriel ve Odoroff 1990), çok değişkenli süreç verilerinde (Sparks vd. 1997), iki yönlü çapraz tablolarda (Bradu ve Gabriel 1978, Gabriel 1995), dayanıklı (robust) yöntemlerde (Daigle ve Rivest 1992), büyüme eğrisi analizinde (Ojeda ve Juarez-Cerrillo 1996), uygunluk analizinde (Greenacre 1984, 1993) kullanılmıştır. Gabriel'in biplot tekniği, Gower ve Hand'in (1996) fikirleriyle yeni bir bakış açısı kazanmıştır.

Gardner (2001), sınıflandırma problemleri ve diskriminant analizinde Biplot tekniğinin kullanılmasına yönelik var olan yöntemin çeşitli genişlemelerini vermiştir.

Diskriminasyon ve sınıflandırma ile birlikte biplot tekniğinin kullanılmasının diskriminant fonksiyonları ve hata oranlarının sunulması yanında ekstra bilgiler verdiğini göstermiştir.

Martín-Rodríguez vd. (2002), bireylerin iki farklı grubundaki aynı değişkenlere uygulanmış olan iki biplot analizinden ortaya çıkarılan sonuçların etkileşimi ve karşılaştırılması için bir yöntem önermişlerdir. Ayrıca biplot geometrisi ve biplot analizinde iki alt uzayın sonuçlarının karşılaştırılması ile ilgili teorik yapıyı gerçek bir veri kümesi üzerinde bir uygulamayla vermişlerdir.

Aitchison ve Greenacre (2002), Biplot tekniğinin bileşik veri[1] analizinde kullanımı için gerekli teorik yapıyı vererek, bir uygulama üzerinde göstermişlerdir.

Krzanowski (2004), çok değişkenli varyans analizinin varsayımlarının sağlanmadığı durumlarda izlenecek yolun, bireylerin her bir çifti arasındaki uzaklıklar veya benzemezlik matrisi hesaplanarak elde edilecek veriye bir uzaklık analizi uygulanması ile bulunan çeşitli metrik ölçekleme grafiklerinin, analizin sonuçlarının yorumlanması için kullanılabileceğini ifade etmiştir. Çalışmasında, uzaklık analizindeki metrik ölçekleme gösterimlerinin tüm formları üzerine biplot'un dahil edilmesini ele almıştır.

Gower (2004), yapmış olduğu çalışmada biplot gösterimlerinin elde edilmesinde kullanılan matris yaklaşımlarının ana özelliklerini sunmuş ve basit bir geometri ile biplot teorisini açıklamıştır.

Gardner ve Roux (2005), yapmış oldukları çalışmada diskriminant analizinin çeşitli formları ile biplot tekniğinin nasıl kombine edildiğini göstermişlerdir. Çalışmalarında, karesel diskriminant analizi biplot, alt uzay diskriminant analizi biplot ve esnek diskriminant analizi biplot tanımlarını vermişlerdir.

[1] **Bileşik veri:** Bir bütünü oluşturan farklı parçaların bütün içindeki yüzde paylarını gösteren değerlerden oluşan verilerdir.

Roux ve Gardner (2005), farklı alanlardan elde edilen çok değişkenli verileri, temel bileşenler analizi biplot ve kanonik değişken analizi biplot ile analiz etmişler ve sınıflandırma bölgeleri ile kabul bölgelerinin biplotlarda nasıl gösterileceğini ele almışlardır.

Gardner vd. (2005), çalışmalarında kanonik değişken analizi ve çok değişkenli varyans analizine kovaryans matrislerinin heterojenliği ve küçük örnek hacmi durumunda bir alternatif önermişlerdir. Çalışmalarında, uzaklık analizinin grup kovaryans matrislerinin heterojenliği ve küçük örneklem genişliğinden etkilenmediğini ve uzaklık analizinin biplot tekniği ile birlikte ele alınmasının gereğini vurgulamışlardır.

Roux ve Gardner (2006), kategorik değişkenlerin diskriminant analizinde biplot tekniğini temel alan bir yaklaşım için gerekli teorik yapıyı ele almışlardır. Genelleştirilmiş biplot yönteminden bahsetmişlerdir.

Gardner vd. (2006), aynı değişkenlerle tanımlanmış gözlemlerin J sınıfının analizi için Kanonik Değişken Analizi (KDA)'nin aynı gözlemlerle tanımlanan değişkenlerin k kümesinin analizi için Genelleştirilmiş Kanonik Korelasyon Analizi (GKKA)'nin ve KDA ile GKKA arasındaki bağlantıları ortaya çıkarmak için değişkenlerin p kümesi ve gözlemlerin g sınıfı içine ayrılmış veri için, genelleştirilmiş procrustes analizinin[2] biplot tekniği ile birlikte kullanımını ele almışlardır.

Park vd. (2008) çalışmalarında, Temel Bileşenler Analizi Biplot, Faktör Analizi Biplot, Çok Boyutlu Ölçekleme Analizi Biplot ve Uygunluk Analizi Biplot olarak bilinen biplot tekniklerinin çeşitli tiplerini ele almışlar ve bu tekniklerin özelliklerini incelemişlerdir. Ayrıca, iyi bilinen gen ifade verisinin çeşitli türlerinin analizinde bu tekniklerin performanslarını karşılaştırmışlardır.

[2] İstatistikte şekillerin bir kümesinin dağılımının analizi için kullanılan şekil analizinin bir formudur.

Gardner-Lubbe vd. (2008), Temel Bileşenler Analizi (TBA) Biplot ve Kanonik Değişken Analizi (KDA) Biplot tekniklerinde uyum ölçülerini incelemişler ve iki farklı alandan elde ettikleri veri setlerinde bu iki tekniğin uygulamasını göstermişlerdir.

Blasius vd. (2009), Kategorik Temel Bileşenler Analizinde (KTBA) biplot tekniğinin kullanımı üzerine bir takım tespitlerde bulunmuşlardır.

Atakan vd. (2009), meyveler ve besin bileşenleri arasındaki ilişkilerin incelenmesinde biplot tekniğini kullanmışlardır. Gıda alanında yapılacak araştırmalarda, biplot tekniğinin kullanımının önemini vurgulamışlardır.

Şahin vd. (2009), kriz dönemlerinde finansal yatırım araçlarından (altın, euro, dolar, borsa, mevduat faizi) hangilerine doğru bir yönelim olduğunu ve yatırım araçları arasındaki ilişkilerin yönü ile derecesinin belirlenmesine yönelik çalışmalarında biplot tekniğini kullanmışlardır.

Friendly ve Kwan (2009), görsel bağlantı teşhis yöntemleri isimli çalışmalarında, tableplot ve bağlantı biplot yöntemlerini ele almışlardır. Çalışmalarında özellikle tableplot kullanımının önemi üzerinde durmuşlardır.

Gardner-Lubbe vd. (2009), gen açıklama verisinin analizinde Temel Bileşenler Analizi (TBA) biplot ve Uzaklık Analizi (UA) biplot kullanmışlar ve bu iki biplot tekniğini karşılaştırmışlardır. Gen açıklama verisi analizlerinde sınıfların ayrılması ile genlerin farklılıklarının açıklanmasında UA biplot'un daha optimal sonuç verdiğini tespit etmişler ve yüksek boyutlu gen açıklama verilerinin analizinde UA biplot'un TBA biplot'un bir alternatifi olabileceğini belirtmişlerdir.

Şahin vd. (2010), endüstriyel üretim içinde bir desen oluşturan ve karşılıklı etkileşimleri tartışmalara konu olmuş Ara İmalat Sanayi, Dayanıklı Tüketim Malları, Dayanıksız Tüketim Malları, İmalat Sanayi Endeksi, Sermaye Malları Üretimi ve Toplam Sanayi

Üretim Endeksi arasındaki ilişkiyi biplot tekniği ve yakınsama yöntemleriyle incelemişlerdir.

Alkan vd. (2011), Türkiye'deki 81 il'de çalışan sağlık personeli dağılımı, hangi illerin hangi sağlık personeli bakımından benzerlik gösterdiği, sağlık personelinin yıllara göre illerdeki değişimi ve sağlık personeli sayısının fazla olduğu illerde fazlalığın sebeplerinin ne olabileceğini biplot tekniği ile incelemişlerdir.

Alkan ve Atakan (2011) çalışmalarında, beş farklı besin grubunda yer alan toplam 105 ürünün 100 gramında bulunan beş kolin değişkeni için, besinlerin ortak kolin içeriklerinin sunulduğu USDA (US Department of Agriculture) veri tabanından derlenen veri kümesini Kanonik Değişken Analizi (KDA) biplot yöntemiyle analiz etmişler ve gıda araştırmalarında veride saklı kalmış bilgilerin ortaya çıkarılmasında KDA Biplot tekniğinin kullanımının son derece önemli olduğunu vurgulamışlardır.

2. BİPLOT TEKNİĞİNİN TEMELLERİ

İki değişken üzerinde n gözlemden oluşan bir veri kümesi, iki boyutlu bir nokta saçılım grafiğinde kolayca yorumlanabilir. İkiden fazla değişken olduğu durumda, çok değişkenli verinin görsel olarak ifadesi oldukça karmaşıktır. Biplot, çok değişkenli verinin daha az boyutta (iki veya üç) yorumlanması için kullanılan bir tekniktir. Boyut indirgemesi genellikle bilgi kaybına sebep olur. Biplotların amacı, bilgideki bu kaybın minimizasyonu için bazı kriterleri optimize etmektir. Optimize edilmiş kriterlere bağlı olarak, biplotların çeşitli tipleri bulunabilir. Farklı kriterler genellikle farklı uzaklık ölçülerine bağlıdır. Bu bölümde biplot tekniği ile ilgili temel yaklaşım ve kavramlar üzerinde durulacaktır.

2.1 Gabriel'in Biplot Yaklaşımı

Rankı k olan herhangi bir $n \times p$ boyutlu $\mathbf{X}$ [3] matrisi, her bir gözlem için bir nokta (satır etkileri) ve her bir değişken için bir vektör (sütun etkileri) olmak üzere grafiksel bir yaklaşım ile gösterilebilir. Gabriel'in yaklaşımı tekil değer ayrıştırmasına dayanmaktadır. $\mathbf{X}: n \times p$ matrisinin tekil değer ayrıştırması (TDA),

$$\mathbf{X}_{n\times p} = \mathbf{U}_{n\times k}\mathbf{\Gamma}_{k\times k}\left(\mathbf{V}_{p\times k}\right)^{T} \tag{2.1}$$

formunda verilir. Burada k, $\mathbf{X}$ matrisinin rankını göstermektedir. $\mathbf{\Gamma}$, köşegenleri $\mathbf{X}$'in sıfır olmayan $0 < \gamma_1 \le \gamma_2 \le ... \le \gamma_k$ tekil değerlerinden oluşan $k \times k$ tipinde köşegen bir matristir ve $\mathbf{\Gamma} = \text{diag}\{\gamma_1, \gamma_2, ..., \gamma_k\}$ biçiminde ifade edilmektedir. $\mathbf{\Gamma}$, $\mathbf{X}^T\mathbf{X}$ veya $\mathbf{X}\mathbf{X}^T$ matrislerinin özdeğerlerinin karekökleri alınarak oluşturulan köşegen matristir. $\mathbf{X}$ matrisinin karesel ve simetrik bir matris olmaması durumunda özdeğer ayrıştırması yerine tekil değer ayrıştırması tercih edilir. Özdeğerler matrisi ise köşegenleri azalan

[3] Tez boyunca, aksi belirtilmedikçe $\mathbf{X}$ matrisinin merkezileştirilmiş olduğu varsayılmaktadır.

sırada özdeğerlerden oluşan, $\mathbf{\Lambda} = \text{diag}\left\{\lambda_1, \lambda_2 ... \lambda_k\right\}$ köşegen matristir. $\mathbf{X}\mathbf{X}^T$ simetrik matrisinin birimleştirilmiş özvektörlerine $\mathbf{U}$ tekil vektörler matrisi karşılık gelirken, $\mathbf{X}^T\mathbf{X}$ simetrik matrisinin birimleştirilmiş özvektörlerine $\mathbf{V}$ tekil vektörler matrisi karşılık gelmektedir. Eşitlik (2.1)'de $\mathbf{U}$ ve $\mathbf{V}$ tekil vektörler matrislerinin ilk k sütunu alınmıştır.

$\mathbf{X}$ matrisine r boyutta yaklaşım,

$$\tilde{\mathbf{X}}^{(r)}_{n\times p} = \mathbf{U}_{n\times r}\mathbf{\Gamma}_{r\times r}\left(\mathbf{V}_{p\times r}\right)^T \tag{2.2}$$

biçiminde ifade edilir.

$\mathbf{X}$ matrisinin bu faktörleştirilmesi, değişkenler arasındaki korelasyon, değişkenlerin varyansları ve gözlemler arasındaki farklılıkların görsel bir incelemesini verir. Eğer $\mathbf{X}$'e r rank ile yaklaşım yeterli oluyorsa, çok değişkenli verinin yorumlanması için kullanışlı bir grafiksel araç elde edilmiş olur.

Eckart-Young teoremi, r ranklı $\tilde{\mathbf{X}}^{(r)}$ matrisi ile k ranklı $\mathbf{X}$ matrisine $\left(r \le k\right)$ optimal yaklaşımı matrislerin en küçük kareler yaklaşım teorisini kullanarak bulur. Yani, burada amaç hata kareler toplamını minimize etmektir (Eckart ve Young 1936).

Buna göre,

$$\min_{\text{rank}(B_{n\times p})=r} \left\|\mathbf{X}-\mathbf{B}\right\| = \left\|\mathbf{X}-\tilde{\mathbf{X}}^{(r)}\right\| = \text{tr}\left\{\left(\mathbf{X}-\tilde{\mathbf{X}}^{(r)}\right)\left(\mathbf{X}-\tilde{\mathbf{X}}^{(r)}\right)^T\right\} = \sqrt{\sum_{j=r+1}^{k}\gamma_j^2} = \sqrt{\sum_{j=r+1}^{k}\lambda_j} \tag{2.3}$$

eşitliği sağlanır.

Rankı r (r≤k) olan daha düşük ranklı $\tilde{\mathbf{X}}^{(r)}$ matrisi ile $\mathbf{X}$ matrisine yaklaşmak istediğimizi varsayalım. Bunu yapmak için ilk olarak yaklaşım hatası veya uyum iyiliği ölçüsü kavramları tanımlanmalıdır. Yaklaşım hatasının ölçüsü genellikle, $\mathbf{E} = \mathbf{X} - \tilde{\mathbf{X}}^{(r)}$ hata matrisinin Öklid normu olarak verilir. Bir matrisin karesel Öklid normunun, matrisin iç çarpımının iz'i olarak yazılabileceği bilgisi ışığında,

$$\|\mathbf{E}\| = \left\|\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right\| = [\mathrm{iz}(\mathbf{E}^T\mathbf{E})]^{1/2} = \left(\sum_{i=1}^{n}\sum_{j=1}^{p} e_{ij}^2\right)^{1/2} \tag{2.4}$$

eşitliği yazılabilir (Bartkowiak ve Szustalewicz 1995).

Burada problem, Öklid normu kullanıldığında minimum hata ile daha düşük ranklı matrislerle $\mathbf{X}$ matrisine nasıl yaklaşılacağıdır. Bu problem ilk olarak Householder ve Young (1938) tarafından ele alınmıştır.

Buna göre, rankı r (r≤k) olan bir $\tilde{\mathbf{X}}^{(r)}$ matrisi ile k ranklı bir $\mathbf{X}$ matrisine en iyi yaklaşım, $\mathbf{E} = \mathbf{X} - \tilde{\mathbf{X}}^{(r)}$ hata matrisinin Öklid normunun minimizasyonu ile, $\mathbf{X}$ matrisinin tekil değer ayrıştırmasının ilk r bileşeninin kullanılması ile elde edilebilir.

Yaklaşım hatası, $\mathbf{X}$ ve $\tilde{\mathbf{X}}^{(r)}$ matrislerinin tekil değer ayrıştırması ve Eckart-Young teoremi kullanılarak,

$$\|E\| = \min_{\text{rank}(\mathbf{B})=r} \|\mathbf{X}-\mathbf{B}\| = \left\|\mathbf{X}-\tilde{\mathbf{X}}^{(r)}\right\| = \left\|\mathbf{U}_{nxk}\boldsymbol{\Gamma}_{kxk}\left(\mathbf{V}_{pxk}\right)^T - \mathbf{U}_{nxr}\boldsymbol{\Gamma}_{rxr}\left(\mathbf{V}_{pxr}\right)^T\right\| = \left\|\sum_{j=1}^{k}\gamma_j u_j v_j^T - \sum_{j=1}^{r}\gamma_j u_j v_j^T\right\|$$

$$= \left\|\gamma_1 u_1 v_1^T + \gamma_2 u_2 v_2^T + \ldots + \gamma_r u_r v_r^T + \gamma_{r+1} u_{r+1} v_{r+1}^T + \ldots + \gamma_k u_k v_k^T - \gamma_1 u_1 v_1^T - \gamma_2 u_2 v_2^T - \ldots - \gamma_r u_r v_r^T\right\|$$

$$= \left\|\gamma_{r+1} u_{r+1} v_{r+1}^T + \ldots + \gamma_k u_k v_k^T\right\| \ , \ \left(\|X\| = \sqrt{\text{iz}(\mathbf{X}^T\mathbf{X})}\right)$$

$$= \sqrt{\sum_{j=r+1}^{k}\gamma_j^2}$$

$$= \sqrt{\sum_{j=r+1}^{k}\lambda_j}$$

şeklinde yazılabilir.

Eşitlik (2.2), $p \times p$ tipindeki bir $\mathbf{J}$ matrisi,

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_r & 0 : r \times (p-r) \\ 0 : (p-r) \times r & 0 : (p-r) \times (p-r) \end{bmatrix} \tag{2.5}$$

kullanılarak,

$$\tilde{\mathbf{X}}^{(r)} = \mathbf{U}\boldsymbol{\Gamma}\mathbf{J}\mathbf{V}^T = \mathbf{U}\mathbf{J}\boldsymbol{\Gamma}\mathbf{V}^T = \mathbf{U}\mathbf{J}\boldsymbol{\Gamma}\mathbf{J}\mathbf{V}^T \tag{2.6}$$

şeklinde yazılabilir. $\mathbf{J}$ için $\mathbf{J}^2 = \mathbf{J}$ ve $(\mathbf{I}-\mathbf{J})^2 = \mathbf{I}-\mathbf{J}$ sağlanır.

$\mathbf{UJ}$ ve $\mathbf{VJ}$ matrisleri, sırasıyla $\mathbf{U}$ ve $\mathbf{V}$ matrislerinin ilk r sütunlarının alındığını ifade eder. Bazı durumlarda $\mathbf{U}$ ve $\mathbf{V}$ matrislerinin ilk r sütunlarının gösterimi için $\mathbf{U}_r$ ve $\mathbf{V}_r$ gösteriminin kullanılması daha uygundur.

2.1.1 Veri matrisinin ayrıştırılması

$\mathbf{X}$ veri matrisinin elemanları, satır ve sütunlara karşılık gelen vektörlerin iç çarpımına eşittir. $\mathbf{U}$, $\mathbf{\Gamma}$ ve $\mathbf{V}$ matrisleri, (2.2) eşitliğinde verilen tekil değer ayrıştırması sonucunda elde edilmektedir. $\mathbf{A}: n \times k$ matrisinin satırları ve $\mathbf{B}: k \times p$ matrisinin sütunları sırasıyla gözlemler ve değişkenler için koordinatları sağlar. Buna göre,

$$\mathbf{X}_{n\times p} = \mathbf{A}_{n\times k}\mathbf{B}_{k\times p} = \left(\mathbf{U}_{n\times k}\mathbf{\Gamma}_{k\times k}\right)\left(\mathbf{V}_{p\times k}\right)^{T} \tag{2.7}$$

eşitliği yazılabilir. (2.7) eşitliğinde, $\mathbf{A}_{n\times k} = \mathbf{U}_{n\times k}\mathbf{\Gamma}_{k\times k}$ ve $\mathbf{B}_{k\times p} = \left(\mathbf{V}_{p\times k}\right)^{T}$ olduğu açıkça görülmektedir.

(2.6) eşitliği r boyutta yaklaşımda,

$$\tilde{\mathbf{X}}^{(r)} = \left(\mathbf{UJ\Gamma}\right)\left(\mathbf{VJ}\right)^{T} = \left(\mathbf{UJ\Gamma Q}\right)\left(\mathbf{VJQ}\right)^{T} = \tilde{\mathbf{A}}^{(r)}\tilde{\mathbf{B}}^{(r)} \tag{2.8}$$

şeklinde ifade edilebilir. Eşitlik (2.8), herhangi bir $r \times r$ tipinde ortogonal $\mathbf{Q}$ matrisi için geçerlidir (Gower vd. 2011).

2.1.2 Grafiksel gösterimin özellikleri

Biplot grafiksel yaklaşımı aşağıda verilen özellikleri sağlamalıdır.

(1) Grafiksel yaklaşımda herhangi iki gözlem çifti arasındaki uzaklıklar Öklid uzaklığıdır: $\left\|\mathbf{a}_i - \mathbf{a}_{i'}\right\|^2 = \left(\mathbf{x}_i - \mathbf{x}_{i'}\right)^T\left(\mathbf{x}_i - \mathbf{x}_{i'}\right)$, burada $\mathbf{a}_i$, $\mathbf{A}$'nın i. satırı ve $\mathbf{x}_i$, $\mathbf{X}$'in i. satırıdır.

(2) Orijinden i. gözleme öklid uzaklığı, $\|\mathbf{a}_i\|^2 = \left(\mathbf{x}_i^T \mathbf{x}_i\right)$, i. satırın toplam varyansa katkısını verir.

(3) Grafiksel yaklaşımda herhangi iki değişken vektörü arasındaki uzaklıklar Öklid uzaklığıdır: $\|\mathbf{b}_j - \mathbf{b}_{j'}\|^2 = \left(\mathbf{x}_j - \mathbf{x}_{j'}\right)^T \left(\mathbf{x}_j - \mathbf{x}_{j'}\right)$, burada $\mathbf{b}_j$, $\mathbf{B}$'nin j. sütunu ve $\mathbf{x}_j$, $\mathbf{X}$'in j. sütunudur.

(4) Orijinden j. değişken vektörüne Öklid uzaklığı, j. değişkenin standart sapması ile doğru orantılıdır. Yani, $\|\mathbf{b}_j\|^2 = \left(\mathbf{x}_j^T \mathbf{x}_j\right) = (n-1)s_j^2$ eşitliği sağlanır.

(5) Grafiksel yaklaşımda $\mathbf{b}_j$ ve $\mathbf{b}_{j'}$ vektörleri arasındaki açının kosinüsü yaklaşık olarak j ve j′ değişkenleri arasındaki korelasyonu verir.

$$r_{jj'} = \frac{\mathbf{b}_j^T \mathbf{b}_{j'}}{\|\mathbf{b}_j\| \|\mathbf{b}_{j'}\|} = \frac{s_{jj'}}{\sqrt{s_{jj} s_{j'j'}}} \approx \cos(\theta_{jj'}) \tag{2.9}$$

Eşitlik (2.9)'a göre, yüksek pozitif korelasyonlu değişken vektörleri arasındaki açı küçüktür. Eğer herhangi iki değişken vektörü arasındaki açı çok küçükse bu iki değişken arasındaki korelasyonunda yüksek olması beklenir. Vektörler arasındaki açı 90° ise bu iki vektör birbirine diktir ve ilgili değişkenler arasında herhangi bir ilişki yoktur. Eğer vektörler arasında geniş açı söz konusu ise negatif eğilimli bir ilişkiden söz edilebilir.

2.1.3 Veri Matrisinin Varyans Ayrıştırması

$\mathbf{X}$ veri matrisinin varyans ayrıştırmasını göstermek için ilk olarak bu veri matrisinin normu kavramı üzerine yoğunlaşmak gerekir. Buna göre norm yapısı kullanılarak,

$$\|\mathbf{X}\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2 = iz\left(\mathbf{X}^T\mathbf{X}\right) = \sum_{k=1}^{p}\gamma_k^2 = (n-1)\sum_{j=1}^{p} s_j^2 \qquad (2.10)$$

eşitliği yazılabilir. Burada γ_k, k. tekil değerdir. $\sum_{j=1}^{p} s_j^2$ ise **X** veri kümesinin p değişkenine ilişkin toplam varyansıdır.

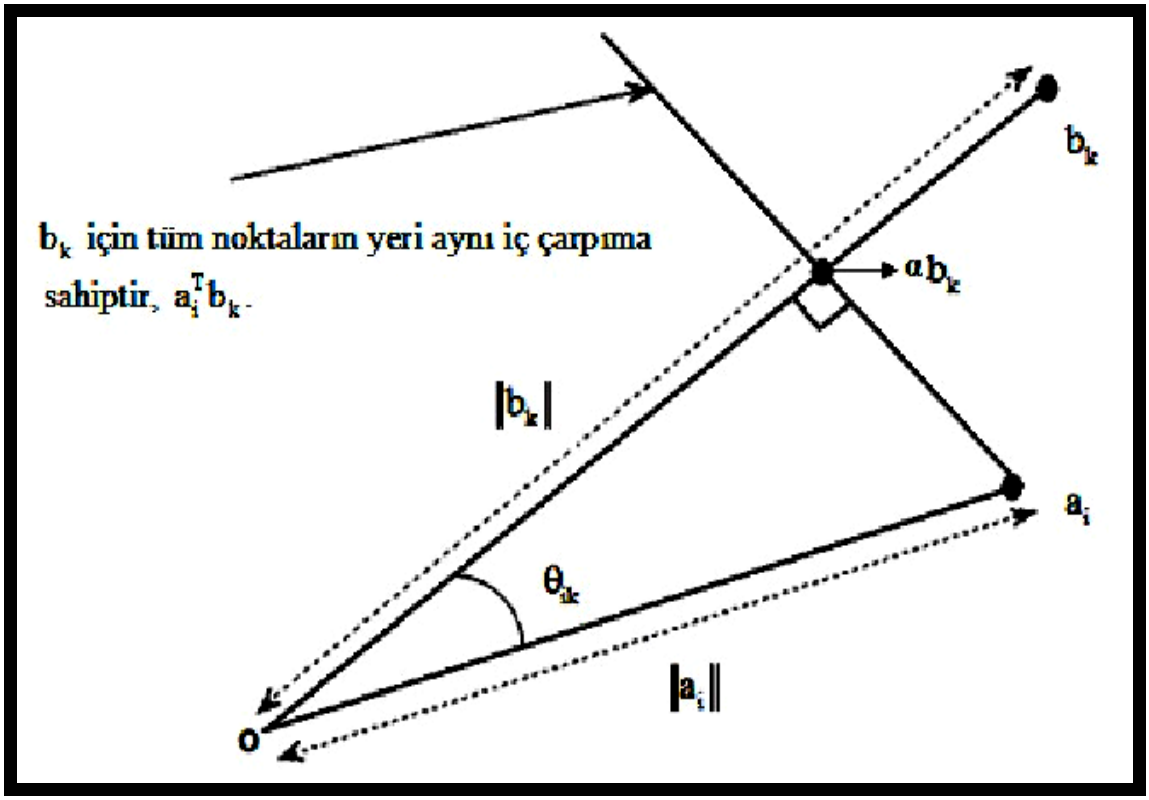
2.2 Ölçeklendirilmiş Biplot Eksenleri

Bölüm 2.1'de tanımlanan biplot yaklaşımının bir iç çarpım gösterimi kullandığı ifade edildi. Ölçeklendirilmiş eksenler, **AB** iç çarpımına dayanan tüm uygulamalar için ortak basit bir yöntem ile açıklanabilir.

Burada,

$$\mathbf{A}: n \times 2 = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ . \\ . \\ . \\ \mathbf{a}_n^T \end{bmatrix} \quad \text{ve} \quad \mathbf{B}^T: p \times 2 = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ . \\ . \\ . \\ \mathbf{b}_n^T \end{bmatrix} \qquad (2.11)$$

olarak tanımlansın. Böylece noktaların bir kümesinin koordinatları gibi **A** 'nın satırlarını ve **B** matrisinin sütunları yardımıyla ölçeklendirilecek eksenlerin koordinatlarını çizebiliriz.



Şekil 2.1 Biplot eksenlerinin ölçeklendirilmesinin geometrik gösterimi

Şekil 2.1'de i. $\mathbf{a}_i$ gözlem noktasını ve k. $\mathbf{b}_k$ eksenini gösterir. $\mathbf{a}_i^T\mathbf{b}_k$ iç çarpımı, $\mathbf{b}_k$ değişken vektörü üzerine indirilen $\mathbf{a}_i$ izdüşüm doğrusu üzerindeki tüm gözlem noktaları için sabittir. İç çarpım $\mu = \|\mathbf{a}_i\|.\|\mathbf{b}_k\|.\cos(\theta_{ik})$ değeridir. $\mathbf{b}_k$ ekseni üzerine gözlem noktasının izdüşümü $\alpha\mathbf{b}_k$'dır. Bu nedenle izdüşüm noktası μ değeri ile bu noktanın etiketlenmesi sonucu ölçeklendirme yapılır. μ ölçeklendirilmiş olan $\alpha\mathbf{b}_k$ noktası için, $\alpha\mathbf{b}_k^T\mathbf{b}_k = \mu$ iç çarpımının sağlanması gerekir. Öyleki, $\alpha = \mu/\mathbf{b}_k^T\mathbf{b}_k$ ve $\mu\mathbf{b}_k/\mathbf{b}_k^T\mathbf{b}_k$ bir μ değeri ile ölçeklendirilmiş $\mathbf{b}_k$ ekseni üzerindeki nokta koordinatlarını verir (Gower vd. 2011).

2.2.1 Alfa (α) Ölçekleme

Bir değişken vektörünün ölçeklendirilmesi yapıldıktan sonra grafiksel gösterimin daha net ve anlaşılır olabilmesi için α ölçeği kullanılır. **A** ve **B** matrisleri yardımıyla verilen noktalar çizileceği zaman, gözlem noktalarının dağılımının değişkenlerin temsil edildiği

vektörleri gölgelediği görülebilir. Yani grafiksel gösterim anlaşılır olmayabilir. Bu problem şu şekilde giderilebilir. $\mathbf{A}$, α ile ve $\mathbf{B}$ matrisi de α^{-1} ile ölçeklendirildiğinde iç çarpım değişmez ve,

$$\mathbf{AB} = (\alpha\mathbf{A})(\mathbf{B}\alpha^{-1}) \tag{2.12}$$

şeklinde yazılabilir.

Grafiksel gösterimin iyileştirilmesi için temel yollardan birisi olarak α seçimi kullanılabilir. α seçiminin bir yolu $\alpha\mathbf{A}$ ve $\mathbf{B}\alpha^{-1}$ deki noktaların ortalama karesel uzaklıklarının düzenlenmesidir. Eğer $\mathbf{A}$, n satıra ve $\mathbf{B}$ matrisi de p sütuna sahip ve her ikiside merkezileştirilirse,

$$\frac{\alpha^2 \|\mathbf{A}\|^2}{n} = \frac{\alpha^{-2} \|\mathbf{B}\|^2}{p} \tag{2.13}$$

elde edilir. Buradan,

$$\alpha^4 = \frac{n}{p} \frac{\|\mathbf{B}\|^2}{\|\mathbf{A}\|^2} \tag{2.14}$$

eşitliği yazılır. Bu α ölçekleme olarak tanımlanır (Gower vd. 2011).

3. TEMEL BİLEŞENLER ANALİZİ BİPLOT

Bu bölümde asimetrik biplot'un en basit ve en popüler şekli olan Temel Bileşenler Analizi (TBA) biplot ele alınacaktır. TBA biplot $\mathbf{X}$ veri kümesinin değişkenlerini ve gözlemlerini aynı uzayda incelediğinden asimetriktir. Ayrıca asimetrik form, veri matrisinin kovaryans veya korelasyon yapısını gösterir. Bu bölümde ilk olarak TBA'nın matematiksel ve geometrik özellikleri gözden geçirilecek ve TBA biplot tartışılacaktır.

3.1 Temel Bileşenler Analizi ve Biplot Yapılandırması

Temel Bileşenler Analizi bir boyut indirgeme tekniğidir. Bu indirgeme, temel bileşenler adı verilen değişkenlerin yeni bir kümesine dönüşüm ile yapılabilir. Temel bileşenler ilişkisizdir. TBA'da orijinal değişkenlerin varyanslarının toplamı, birbiriyle ilişkisiz daha az değişkenle ve mümkün olduğunca az hatayla açıklanmaya çalışılır. Bu bölümde TBA dönüşümünün farklı yönleri vurgulanmış ve daha farklı bir yaklaşım ele alınmıştır (Jolliffe 2002).

TBA'nın esas problemi $\tilde{\mathbf{X}}^{(r)}$ ile $\mathbf{X}$'e r boyutta veya eşdeğer bir ifadeyle r rank ile yaklaşmaktır. TBA'da $\mathbf{X}$'in sütunları farklı değişkenleri ima eder. TBA, yaklaşım temeli olarak en küçük kareler kriterini kullanır. $\mathbf{X}$ ve $\tilde{\mathbf{X}}^{(r)}$'nin elemanları arasındaki farkların kareler toplamı minimize edilir.

Cebirsel olarak,

$$\text{Minimum iz}\left\{\left(\mathbf{X}-\tilde{\mathbf{X}}^{(r)}\right)^{T}\left(\mathbf{X}-\tilde{\mathbf{X}}^{(r)}\right)\right\} \text{ veya Minimum } \left\|\left(\mathbf{X}-\tilde{\mathbf{X}}^{(r)}\right)\right\| \qquad (3.1)$$

şeklinde ifade edilir.

Geometrik olarak, p boyutta n noktanın koordinatlarını veren $\mathbf{X}$'in satırları düşünülür ve (3.1) eşitliğini minimum yapan ilk r boyut için gözlemlerin koordinatlarını veren r boyutlu bir uzay aranır. En iyi uyumu veren $\tilde{\mathbf{X}}^{(r)}$, $\mathbf{X}$'in dik izdüşümü olduğunda bulunur.

Eşitlik (2.1)'de verildiği gibi, $\mathbf{X}: n \times p = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$ ve r boyutta Eckart-Young yaklaşımı, $\tilde{\mathbf{X}}^{(r)} = \mathbf{U}\mathbf{\Gamma}\mathbf{J}\mathbf{V}^T = \mathbf{U}\mathbf{J}\mathbf{\Gamma}\mathbf{V}^T = \mathbf{U}\mathbf{J}\mathbf{\Gamma}\mathbf{J}\mathbf{V}^T = \mathbf{X}\mathbf{V}\mathbf{J}\mathbf{V}^T$, (2.3) ile verilen karesel hatayı minimize eder. Daha sonra merkezileştirilmiş $\mathbf{X}$'e r boyutta yaklaşımın koordinatları, $\mathbf{U}\mathbf{\Gamma}\mathbf{J}$, yani $(\mathbf{U}\mathbf{\Gamma})_r = \mathbf{X}\mathbf{V}_r$'nin ilk r sütunuyla ve $\mathbf{V}\mathbf{J}$ ile de eksenlerin yönleri verilir. Böylece eksenlerin yönünü veren $\mathbf{V}\mathbf{J}$'nin ilk p satırı ve r boyutta nokta saçılım grafiği ile $\mathbf{U}\mathbf{\Gamma}\mathbf{J}$'nin n satırını veren koordinatlar çizilir. Burada, $\mathbf{V}\mathbf{J}\mathbf{V}^T$'nin izdüşüm matrisi olduğunun unutulmaması gerekir. $\mathbf{V}$'nin sütunları temel bileşenler veya temel bileşen yükleri olarak adlandırılır. $\mathbf{X}\mathbf{V}$, temel bileşenler olarak yorumlanır. Ayrıca, $\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{\Gamma}^2 = \mathbf{\Lambda}$, köşegendir ve temel bileşenler ilişkisizdir. Buna göre toplam kareler toplamı,

$$\|\mathbf{X}\|^2 = \|\tilde{\mathbf{X}}^{(r)}\|^2 + \|\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\|^2 \tag{3.2}$$

şeklinde yazılabilir.

3.1.1 Gözlem Noktalarının Gösterimi

Geometrik olarak, gösterim için en iyi r boyut (genellikle, 2 veya 3) üzerinde $\mathbf{X}$'in gözlem noktalarının dik izdüşümlerinin Eckart-Young yaklaşımı ele alınır. Biplot uzayını S ile gösterelim. İki boyutlu S biplot uzayında p boyutlu gözlem noktalarının izdüşümünü düşünelim. $S = 2$ ve $p = 3$ olduğu durumda Şekil 3.1'de verilen içi dolu noktalar tüm uzayda (R) gözlem noktalarını, içi boş olan noktalar ise biplot uzayında ($S = 2$) gözlem noktalarını göstersin. Şekil 3.1'de oklar ile gösterilen uzaklıkların kareler toplamı minimize edilir. Bu şekil, tüm uzayda gözlem noktalarının indirgenmiş boyutlu uzaydaki gözlem noktalarına uzaklığının minimum olması gereğini gösterir. Bu uzaklıklar ne kadar küçük olursa gerçek yapıya yaklaşım o kadar daha iyi sağlanmış olur.

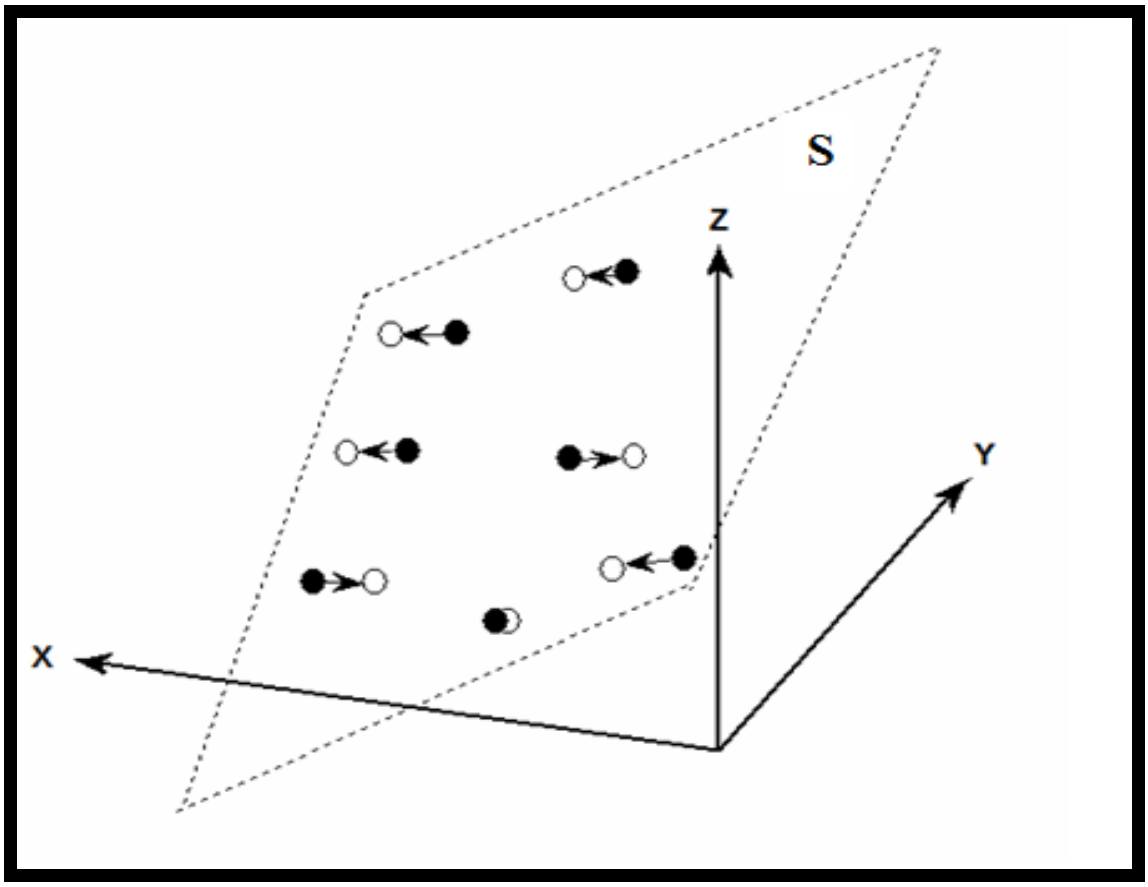
TBA biplot'da interpolasyon, biplot uzayı üzerinde her bir gözlem noktasının dik izdüşümü ile elde edilir.

$$\mathbf{x}_{\text{izdusum}}^{T} = \mathbf{x}^{T}\mathbf{V}_{r}\mathbf{V}_{r}^{T} \tag{3.3}$$

p kartezyen eksen üzerine izdüşürülmüş $\mathbf{x}$ gözleminin temsili (3.3) eşitliği ile bulunur. S biplot uzayında, dik eksenlere göre izdüşürülen gözlem noktaları gösterileceği zaman, izdüşürülen gözlem noktaları,

$$\mathbf{z}^{T} = \mathbf{x}^{T}\mathbf{V}_{r} \tag{3.4}$$

eşitliği ile verilir.



Şekil 3.1 Tüm uzay ve biplot uzayı arasındaki ilişki

3.1.2 İnterpolasyon ve Kestirim Kavramı

Şekil 3.1 sadece gözlem noktaları için bilgi verdiğinden, biplot için bir örnek değildir. X, Y ve Z değişkenleri üzerine bilgi vermemektedir. Bu bilgiyi eklemek için, biplot

eksenleri grafiğe eklenir. Bu eksenler, yeni gözlem noktaları eklemek için kullanılır. Tüm Biplot eksenleri orijinden geçer. Gower ve Hand (1996), orijinal değişkenler ve temel eksenler arasındaki ilişkiler için interpolasyon ve kestirim kavramlarını kullanmışlardır.

Herhangi bir yeni gözlem için verilen değişken değerleri $\mathbf{x}^* : \mathrm{p} \times 1$ olsun. S uzayında $\mathbf{z}^* : \mathrm{r} \times 1$ olarak adlandırılan gözlemin pozisyonunun bulunması süreci interpolasyon olarak tanımlanır. Kestirim ise orijinal değişkenlerin değerlerinin S biplot uzayında çıkarım (inferring), $\tilde{\mathbf{x}}^* : \mathrm{p} \times 1$, sürecidir. $\mathbf{x}^* : \mathrm{p} \times 1$'in S üzerinde interpolasyonu $\mathbf{z}^{*\mathrm{T}}_{1\times r} = \mathbf{x}^{*\mathrm{T}}_{1\times p}\mathbf{V}_{p\times r}$ ile verilirken, kestirim ise, $\tilde{\mathbf{x}}^{*\mathrm{T}}_{1\times p} = \mathbf{z}^{*\mathrm{T}}_{1\times r}\mathbf{V}^{\mathrm{T}}_{r\times p}$ ile verilir.

3.2 TBA Biplot İçin Uyum Ölçüleri

Gözlenen herhangi bir $\mathbf{x}$ vektörü,

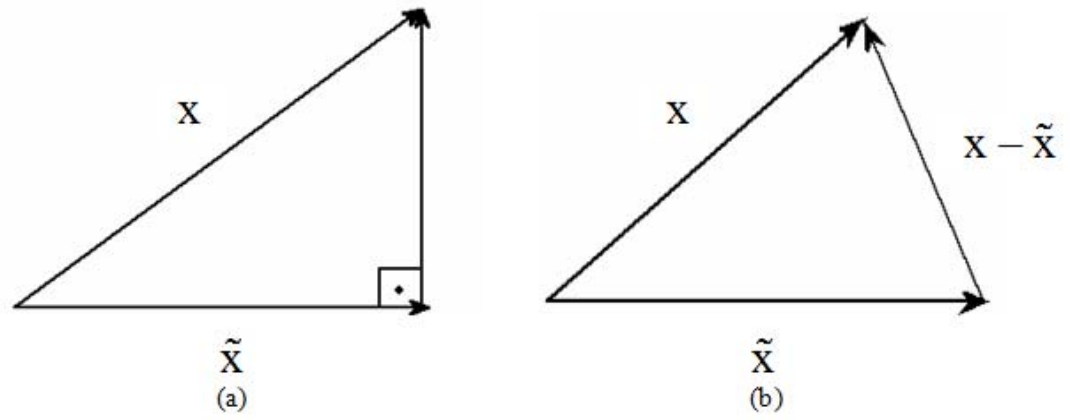
$$\mathbf{x} = \tilde{\mathbf{x}} + \left(\mathbf{x} - \tilde{\mathbf{x}}\right) \tag{3.5}$$

şeklinde parçalanabilir. Burada $\mathbf{x}$ vektörünün kestirimi $\tilde{\mathbf{x}}$ ile ifade edilirken hata ise $\left(\mathbf{x} - \tilde{\mathbf{x}}\right)$ ile tanımlanır. Uyumun kalitesi;

$$\frac{\mathrm{KT}\left(\tilde{\mathbf{x}}\right)}{\mathrm{KT}\left(\mathbf{x}\right)} = \frac{\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}} \tag{3.6}$$

varyans açıklanma oranı ile belirlenebilir.

(3.6) eşitliğinde verilen oran sadece dik ayrıştırma olduğunda çok daha anlamlıdır. Ayrıca (3.5) eşitliği, $\mathrm{KT}\left(\mathbf{x}\right) = \mathrm{KT}\left(\tilde{\mathbf{x}}\right) + \mathrm{KT}\left(\mathbf{x} - \tilde{\mathbf{x}}\right)$ şeklinde de yazılabilir.



Şekil 3.2 Karesel toplamların dik (a) ve dik olmayan (b) kırılması

Şekil 3.2-(a), dik kırılmayı gösterir. Herhangi bir dik olmayan analiz, dikliğe göre daha fazla hata kareler toplamına sahip olacaktır. Yani, Şekil 3.2-(b) ile verilen dik olmayan bir kırılma, Şekil 3.2-(a)'ya göre daha büyük bir hata kareler toplamına sahip olacaktır. Bunun sonucu olarak daha büyük bir varyans açıklama oranı verecektir. Hatta bu oranın 1'den büyük olduğu durumlarla da karşılaşılabilir. Böylece (3.6) eşitliği ile verilen oranın sadece diklik koşulu sağlandığında kullanışlıdır. Yani (3.6) eşitliğinde verilen oran, diklik koşulları sağlandığında güvenli bir ölçü olabilir (Gardner vd. 2008).

TBA'nın amacı, en küçük kareler düşüncesinde, en iyi uyumu veren r boyutlu S alt uzayını seçmektir. Uyum ölçüleri kareler toplamlarının oranları olmasından dolayı gerekli diklik koşullarının sağlanması önemlidir.

Bir $\mathbf{X}$ veri matrisi için, $\mathbf{X} = \mathbf{XVJ} + \left(\mathbf{X} - \mathbf{XVJ}\right) = \tilde{\mathbf{X}}^{(r)} + \left(\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right)$ eşitliğinden dikliğe ulaşmanın olası iki yolu vardır. Bunlar Tip A ve Tip B olarak tanımlanır.

$$\mathbf{XX}^{T} = \mathbf{XVJV}^{T}\mathbf{X}^{T} + \mathbf{XV}\left(\mathrm{I} - \mathrm{J}\right)\mathbf{V}^{T}\mathbf{X}^{T} = \tilde{\mathbf{X}}^{(r)}\tilde{\mathbf{X}}^{(r)T} + \left(\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right)\left(\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right)^{T} \tag{3.7}$$

ve

$$\begin{aligned}\mathbf{X}^{\mathrm{T}}\mathbf{X} = \|\mathbf{X}\|^2 &= \left\|\tilde{\mathbf{X}}^{(r)} + \mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right\|^2 \\ &= \left\|\tilde{\mathbf{X}}^{(r)}\right\|^2 + \left\|\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right\|^2 \\ &= \tilde{\mathbf{X}}^{(r)\mathrm{T}}\tilde{\mathbf{X}}^{(r)} + \left(\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right)^{\mathrm{T}}\left(\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right)\end{aligned} \tag{3.8}$$

dir.

(3.7) eşitliği ile verilen $\mathbf{X}$ veri kümesi için Tip A diklik olarak, (3.8) eşitliği ile verilen ise Tip B diklik olarak tanımlanır. $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ matrisinin köşegenleri $\mathbf{X}$'in satırlarının kareler toplamını verir ve gözlemleri ifade eder. $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ matrisinin köşegenleri $\mathbf{X}$'in sütunlarının kareler toplamlarını verir ve değişkenleri ifade eder.

Hatalar, $\mathbf{X}(\mathbf{I} - \mathbf{V}\mathbf{J})$'nin kareler toplamı olan,

$$\left\|\mathbf{X}(\mathbf{I} - \mathbf{V}\mathbf{J})\right\|^2 = \mathrm{iz}\left\{\left(\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right)\left(\mathbf{X} - \tilde{\mathbf{X}}^{(r)}\right)^{\mathrm{T}}\right\} = \mathrm{iz}\left\{\mathbf{X}\mathbf{V}(\mathbf{I} - \mathbf{J})\mathbf{V}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\right\} \tag{3.9}$$

eşitliği ile verilir.

Hatalar için kareler toplamı,

$$\mathrm{iz}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) - \mathrm{iz}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{V}\mathbf{J}\mathbf{V}^{\mathrm{T}}\right) = \mathrm{iz}(\mathbf{\Lambda} - \mathbf{\Lambda}\mathbf{J}) = \sum_{j=r+1}^{p} \lambda_j \tag{3.10}$$

olarak yazılabilir.

Gösterimin kalitesi(GK)'nin ölçüsü ise,

$$GK = \frac{iz(\mathbf{\Lambda J})}{iz(\mathbf{\Lambda})} = \frac{\sum_{j=1}^{r} \lambda_j}{\sum_{j=1}^{p} \lambda_j} = \frac{\sum_{j=1}^{r} \gamma_j^2}{\sum_{j=1}^{p} \gamma_j^2} \tag{3.11}$$

oranı ile bulunur.

Genel gösterim kalitesi sürecin sadece bir bölümü olup, r boyutta temsil edilen değişkenlerin gösterim yeterliliği hakkında da bilgi edinilebilmesi için incelenecek ölçü yeterlilik olarak adlandırılır. **VJ** 'nin satırlarının kareler toplamları, her bir değişkenin temsilinin yeterliliği ölçüsünü verir. Cebirsel olarak bu ölçü, $p \times p$ boyutlu $\mathbf{VJV}^T$ köşegen matrisi ile verilir.

Yeterlilik ölçüsü (YÖ),

$$(\text{YÖ})_j = \sum_{s=1}^{r} v_{js}\ ,\ j = 1,2,3,...,p. \tag{3.12}$$

ile verilir. Burada $\text{diag}(\mathbf{VJV}^T)$'nin köşegen elemanları değişkenler için yeterlilik ölçülerini verir. Bu ölçü j. değişkenin temsilinin yeterliliğinin bir ölçüsüdür (Gardner vd. 2008).

Yeterlilik verinin uyum ölçüsü değildir. Popüler bir ölçü olmasına rağmen, **X** veri kümesine yaklaşımın sağlanmasında bir takım sınırlamalara sahiptir. Yeterlilik, koordinat eksenlerinin hazırlanmasıyla daha ilgili iken, kestirimler ise **X** veri kümesi ile ilişkilendirilmiş değişkenlerin aldığı değerlerin kestirilmesinde daha başarılıdır. Burada yaklaşımdan kasıt, **X** veri kümesine daha düşük r boyutta yaklaşımı sağlayan $\tilde{\mathbf{X}}^{(r)}$'yi elde etmektir.

Eksen kestirimleri $\mathbf{\Pi} : p \times p$ matrisinin köşegen elemanları olarak,

$$\mathbf{\Pi} = \text{diag}\left(\tilde{\mathbf{X}}^{(r)T}\tilde{\mathbf{X}}^{(r)}\right)\left[\text{diag}\left(\mathbf{X}^{T}\mathbf{X}\right)\right]^{-1} = \text{diag}\left(\mathbf{V\Lambda JV}^{T}\right)\left[\text{diag}\left(\mathbf{V\Lambda V}^{T}\right)\right]^{-1} \tag{3.13}$$

eşitliği ile tanımlanır.

Benzer şekilde gözlem noktalarının kestirimleri ise,

$$\mathbf{\Psi} = \text{diag}\left(\tilde{\mathbf{X}}^{(r)}\tilde{\mathbf{X}}^{(r)T}\right)\left[\text{diag}\left(\mathbf{XX}^{T}\right)\right]^{-1} = \text{diag}\left(\mathbf{U\Lambda JU}^{T}\right)\left[\text{diag}\left(\mathbf{U\Lambda U}^{T}\right)\right]^{-1} \tag{3.14}$$

eşitliği ile bulunur.

$\mathbf{\Pi}$'nin köşegen elemanları, her bir değişkenin kestirim güçlerinin ölçüleridir ve bu ölçüler 1'i geçemez. Π_j, j. değişken için eksen kestirimini gösterir.

Eksen kestirimi değişken başına varyans açıklanma oranıdır ve

$$\Pi_j = \frac{\left(\mathbf{V\Lambda JV}^{T}\right)_{jj}}{\left(\mathbf{V\Lambda V}^{T}\right)_{jj}} \tag{3.15}$$

eşitliği yardımıyla bulunur. Burada, jj- indisi köşegenlerin alınacağını gösterir.

Eşitlik (3.11),

$$\text{GK} = \frac{\text{iz}\left(\Lambda\text{J}\right)}{\text{iz}\left(\Lambda\right)} = \sum_{j=1}^{p}\frac{\left(\text{V}\Lambda\text{V}^{T}\right)_{jj}}{\text{iz}\left(\Lambda\right)}\Pi_j \tag{3.16}$$

şeklinde yazılabilir.

Eksen kestirimleri her bir değişkenin gösterimlerinin kalitesi ile ilgilenirken, gözlem kestirimleri ise gözlemlerin kalitesini incelemeyi olanaklı kılar.

$\mathbf{X}$'e en yakın en küçük kareler yaklaşımı $\tilde{\mathbf{X}}^{(r)}$ için, eksen kestirimlerinin eksen yeterliliklerinden büyük veya eşit olması sağlanmalıdır (Gardner vd. 2008). $\mathbf{\Lambda = I}$ olduğunda eksen yeterliliği, Π_j dir.

4. KANONİK DEĞİŞKEN ANALİZİ BİPLOT

Çok değişkenli istatistiksel yöntemlerin çoğu, verideki grupların bulunması probleminin çözümü için geliştirilmiştir. İstatistiksel olarak grupların optimal ayrışımı için gerekli cebirsel türetimin sağlanması yanı sıra, bu yöntemlerde analiz sonuçlarının gösterimi grafiklerle birlikte olur (Gardner vd. 2005).

Gardner ve Roux (2005), gruplandırma problemleri ve diskriminant analizinde biplot tekniğinin kullanılmasına yönelik var olan metodolojinin çeşitli genişlemelerini vermişlerdir. Diskriminant analizi ile birlikte biplot tekniğinin kullanılmasının diskriminant fonksiyonları ve hata oranlarının elde edilmesi yanında, ekstra bilgiler verdiğini göstermişlerdir. Biplot tekniğinin gruplandırma problemlerine uygulanması sadece, grupların ayrılmasında farklı değişkenlerin oynadığı rol ve grup içi yapının ayrıntılı bir şekilde tanımlanması değil, aynı zamanda grupları ayırmanın derecesinin ölçülmesi için de kullanılır (Gardner ve Roux 2005).

Kanonik Değişken Analizi (KDA), gruplar arası değişimin grup içi değişime oranını maksimize eden değişkenlerin doğrusal birleşimlerinin bulunması ile grupların ayrılması için popüler bir tekniktir. KDA, lineer diskriminant analizi (LDA) ve çok değişkenli varyans analizi (MANOVA) ile yakın ilişkilidir. KDA, kanonik diskriminant analizi olarak da adlandırılmaktadır.

Kanonik Değişken Analizi Biplot (KDA Biplot) kullanımı ise, son yıllarda önem kazanan bir biplot yaklaşımı olarak karşımıza çıkmaktadır. KDA Biplot, diskriminant analizindeki biplot tekniğinin en temel uygulaması olup, doğrusal diskriminant analizinin grafiksel bir gösterimi olarak görülebilir (Gardner vd. 2005).

KDA Biplot, birimlerin optimal bir şekilde gruplara ayrılması ve gruplara ayrılmada farklı değişkenlerin rolünün görsel olarak yorumlanmasını sağlamaktadır.

KDA, p tane değişken ile tanımlanan bütün gözlemlerin g tane grubunun analizi ile ilgilenir. KDA, gruplar arası varyansın, gruplar içi varyansa oranını $\mathbf{v}_j^T\mathbf{B}\mathbf{v}_j \big/ \mathbf{v}_j^T\mathbf{W}\mathbf{v}_j$, $j = 1,2,3,...,p$, $\mathbf{V}^T\mathbf{W}\mathbf{V} = \mathbf{I}$ kısıtı altında maksimize eden değişkenlerin doğrusal birleşimlerinin incelenmesi ile gruplara ayırmayı amaçlamaktadır. Burada, $\mathbf{v}_j$, j. kanonik değişken olup, $\mathbf{V}$ ise sütunları $\mathbf{v}_j$'lerden oluşan $p \times p$ tipinde bir matristir. $\mathbf{B}$ ve $\mathbf{W}$ matrisleri sırasıyla gruplar arası ve grup içi varyans-kovaryans matrisleridir. $\mathbf{X}: n \times p$ merkezileştirilmiş veri matrisi olmak üzere $\mathbf{B}$ ve $\mathbf{W}$ matrisleri,

$$\mathbf{B}_{p\times p} = \sum_{\ell=1}^{g} n_\ell \left(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}}\right)\left(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}}\right)^T = \bar{\mathbf{X}}_{p\times g}^T \mathbf{N}_{g\times g}\, \bar{\mathbf{X}}_{g\times p} \tag{4.1}$$

$$\mathbf{W}_{p\times p} = \sum_{\ell=1}^{g}\sum_{c=1}^{n_\ell}\left(\mathbf{x}_{\ell c} - \bar{\mathbf{x}}_\ell\right)\left(\mathbf{x}_{\ell c} - \bar{\mathbf{x}}_\ell\right)^T = \mathbf{X}_{p\times n}^T \mathbf{X}_{n\times p} - \mathbf{B}_{p\times p} \tag{4.2}$$

biçiminde ifade edilir.

Burada,

$$\bar{\mathbf{x}}_\ell = \frac{1}{n_\ell}\sum_{c=1}^{n_\ell} x_{ic}\,, \quad \bar{x} = \frac{1}{n}\sum_{\ell=1}^{g} n_\ell \bar{x}_\ell \text{ ve } \mathbf{N}_{g\times g} = \text{köş}\left(n_1,...,n_g\right)$$

ile elde edilir. $\bar{\mathbf{X}}: g \times p$ matrisinin satırları grup ortalama vektörlerini, $\bar{\mathbf{x}}_\ell^T$, $\ell = 1,2,...,g$ gösterir.

Kanonik değişkenler matrisinin bulunması için ilk olarak, $\mathbf{W}_{p\times p}^{-1/2}\mathbf{B}_{p\times p}\,\mathbf{W}_{p\times p}^{-1/2}$ matrisinin özvektörleri bulunur ve bu özvektörler $\mathbf{M}_{p\times p}$ matrisi ile gösterilir. Aynı matrisin azalan sırada özdeğerleri ise $\mathbf{\Lambda} = \text{diag}\left(\lambda_1,...,\lambda_p\right)$ köşegen matrisi ile verilir. Kanonik değişkenler matrisi ise,

$$\mathbf{V}_{p\times p} = \mathbf{W}_{p\times p}^{-1/2}\mathbf{M}_{p\times p} \tag{4.3}$$

eşitliği ile bulunur.

Buradan tüm uzayda kanonik skorlar yani kanonik değişkenler olarak adlandırılan değişkenlerin yeni bir kümesi, $\mathbf{U}_{n\times p} = \mathbf{X}_{n\times p}\mathbf{V}_{p\times p}$ dönüşümü ile elde edilir. Daha sonra $\bar{\mathbf{U}}_{g\times p} = \bar{\mathbf{X}}_{g\times p}\mathbf{V}_{p\times p}$ ile kanonik ortalamalar olarak adlandırılan ortalamaların yeni bir kümesi elde edilir. Buradan,

$$\bar{\mathbf{U}}\bar{\mathbf{U}}^{T} = \bar{\mathbf{X}}_{g\times p}\mathbf{W}_{p\times p}^{-1}\bar{\mathbf{X}}_{p\times g}^{T} \tag{4.4}$$

eşitliğine göre, kanonik uzayda ortalamalar arasındaki Öklid uzaklıkları, orijinal uzayda Mahalanobis uzaklıklarıdır. Yani, kanonik uzay Öklid'dir. Bu uzay $r \le \min(p, g-1)$ boyutludur.

Eğer Öklid kanonik uzayı R ile gösterilirse, kanonik ortalamalar r boyutlu bir S alt uzayında gösterilebilir. Bu kanonik ortalamalar KDA Biplot'daki orijinal değişkenlerle ilişkili olabilir.

S alt uzayında gözlem noktalarının izdüşümlerinin koordinatları,

$$\mathbf{Z}_{n\times r} = \mathbf{X}_{n\times p}\,\mathbf{V}_{p\times r} \tag{4.5}$$

ve kanonik ortalamalar,

$$\bar{\mathbf{Z}}_{g\times r} = \bar{\mathbf{X}}_{g\times p}\,\mathbf{V}_{p\times r} \tag{4.6}$$

eşitlikleri ile bulunur.

Eşitlik (4.6), S alt uzayında kanonik ortalamaların r boyutlu koordinatlarının bir kümesini verir.

KDA Biplot elde etmek için kullanılan interpolasyon ve kestirim kavramlarının tanımları Bölüm 3.1.2'de verilmiştir.

r-boyutta gözlem kestirimleri,

$$\text{diag}\left(\tilde{\mathbf{Y}}_{\mathbf{g}}^{\mathbf{T}}\tilde{\mathbf{Y}}_{\mathbf{g}}\right)\left\{\text{diag}\left(\mathbf{Y}_{\text{g}}^{\text{T}}\mathbf{Y}_{\text{g}}\right)\right\}^{-1} \tag{4.7}$$

ile verilir.

Eşitlik (4.7)'de, $\mathbf{Y}_{\text{g}} = \left(\mathbf{I} - \mathbf{G}\left(\mathbf{G}^{\text{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\text{T}}\right)\mathbf{X}$ olup, $\mathbf{G}:\text{n}\times\text{g}$, gözlem-grup gösterge matrisidir. Eğer i gözlemi ℓ grubuna ait ise e_{ℓ}'nın $i.$ satırıdır. $i = 1,2,...,n$; $\ell = 1,2,...,\text{g}$. e_{ℓ} ise, $\mathbf{I}_{\text{n}}$ birim matrisinin ℓ. sütunudur.

$\tilde{\mathbf{Y}}_{\mathbf{g}}$ matrisi,

$$\tilde{\mathbf{Y}}_{\mathbf{g}} = \mathbf{Y}_{\text{g}}\mathbf{V}_{\text{r,int}}\left(\mathbf{V}_{\text{r,kes}}\right)^{\text{T}} \tag{4.8}$$

ile verilir. Eşitlik (4.8)'de, interpolasyon için temel matris, $\mathbf{V}_{\text{r,int}}$, $\mathbf{V}$ 'nin ilk r sütununu içerir. Kestirim için temel matris, $\mathbf{V}_{\text{r,kes}}$, $\left(\mathbf{V}^{-1}\right)^{\text{T}}$'nün ilk r sütununu içerir.

r-boyutta grup kestirimleri,

$$\text{diag}\left(\tilde{\bar{\mathbf{X}}}\mathbf{W}^{-1}\tilde{\bar{\mathbf{X}}}^{\text{T}}\right)\left\{\text{diag}\left(\bar{\mathbf{X}}\mathbf{W}^{-1}\bar{\mathbf{X}}^{\text{T}}\right)\right\}^{-1} \tag{4.9}$$

ile verilir. Eşitlik (4.9)'da, $\tilde{\bar{\mathbf{X}}} = \bar{\mathbf{X}}\mathbf{V}_{\text{r,int}}\left(\mathbf{V}_{\text{r,kes}}\right)^{\text{T}}$ dir.

r-boyutta eksen kestirimleri,

$$\operatorname{diag}\left(\tilde{\bar{\mathbf{X}}}^{\mathbf{T}}\tilde{\bar{\mathbf{X}}}\right)\left\{\operatorname{diag}\left(\bar{\mathbf{X}}^{\mathrm{T}}\bar{\mathbf{X}}\right)\right\}^{-1} \tag{4.10}$$

ile verilir.

5. UZAKLIK ANALİZİ BİPLOT

Veri kümesinde heterojenlik ve çarpıklık aşırı ise, çok değişkenli tek yönlü varyans analizi (MANOVA) gibi geleneksel çok değişkenli teknikler, gruplar arasındaki farklılıklar belirleneceği zaman yüksek derecede kuşkulu sonuçlar verebilmektedir.

Uzaklık analizi (UA), kitlelere (gruplara) ilişkin varyans-kovaryans matrislerinin heterojen olması durumunda tek yönlü MANOVA'ya alternatif bir çıkarımsal metot olarak Gower ve Krzonowski (1999) tarafından önerilmiştir. Çalışmalarında Temel Koordinat Analizi (TKA) ve Uzaklık Analizi (UA) olarak adlandırılan iki yaklaşımı ele almışlardır. TKA, Temel Bileşenler Analizine (TBA) paralel bir yaklaşım olup, analiz için gözlem çiftleri arasındaki uzaklıklardan oluşan matrisi kullanırken, TBA, değişkenler arasındaki korelasyonları veya varyans-kovaryansları içeren matrisi kullanır. UA ise MANOVA' ya paralel bir yaklaşım olup, kareler toplamlarından daha çok gözlemler arasındaki uzaklıkların parçalanması temeline dayanır (Fenty 2004).

TKA, klasik metrik ölçekleme olarak da bilinmekte ve çok boyutlu ölçekleme (ÇBÖ) teknikleri ailesine aittir. Bu teknikler indirgenmiş bir uzayda çok değişkenli veri kümesinin görsel temsilini bulmak için tasarlanmıştır. Veri kümesinde yer alan değişkenler arasındaki benzerlikleri yansıtan bir koordinat sisteminde, dik eksenlere ait gözlemlerin bir grafiğini verir. Dik eksenler, indirgenmiş boyutlu uzayda (genellikle iki boyutlu) indirgemeden kaynaklanacak bilgi kaybını minimize edecek şekilde seçilir.

Temel varyans formülü, n tane gözlem noktası arasındaki karesel Öklid uzaklıkları toplamı olarak,

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{t=1}^{n}\left(x_i-x_t\right)^2 \qquad (5.1)$$

biçiminde yazılır.

Eşitlik (5.1), uzaklık analizi ile varyans analizi arasındaki bağlantıyı vermektedir. MANOVA, toplam çapraz çarpımlar ve kareler toplamları matrisinin parçalanmasını temel alır. UA ise, ikili gözlem çiftleri arasındaki uzaklıkları içeren bir matrisin parçalanmasını temel alır. UA'da tek varsayım herhangi iki i ve t gözlemleri arasında bir uzaklığın (d_{it}) tanımlanabilmesidir. Ayrıca, tüm i ve t gözlemleri için $d_{it} = d_{ti}$ olmalıdır. d_{it}'yi hesaplamak için farklı uzaklık ölçüleri vardır (Fenty 2004).

İki gözlem arasındaki benzerlik, Öklid uzayında gözlemlerin grafiği ile gösterilebilir. Bu grafiksel yaklaşımda benzer gözlemler arasındaki uzaklık, benzer olmayan gözlemler arasındaki uzaklıktan daha az olacaktır. Uzaklığın en bilinen ölçüsü, Öklid uzaklığıdır. j boyut üzerinden ölçülen i gözleminin değeri x_{ij} ile gösterilsin. Bu durumda i ve t gözlemleri arasındaki uzaklık, d_{it},

$$d_{it} = \sqrt{\sum_{j=1}^{p} \left(x_{ij} - x_{tj} \right)^2} \tag{5.2}$$

ile verilir.

Öklid uzaklığı, Minkowski uzaklıkları olarak bilinen uzaklık ailesinin bir üyesidir (Coxon 1982). Bu form,

$$d_{it}^{(h)} = \sqrt[h]{\sum_{j=1}^{p} \left| x_{ij} - x_{tj} \right|^h} \tag{5.3}$$

ile ifade edilir. (5.3) eşitliğinde $h = 2$ olduğunda Öklid uzaklığını verir. Sadece Öklid uzaklığı, reel çözümlere ulaşacağından TKA ve UA'de önemli bir rol oynar.

5.1 Temel Koordinat Analizi

TKA çok fazla bilinen ve kullanılan bir yöntem değildir. Ancak genetik (Kosaki vd. 1996), antropoloji (Fox vd.1996), moleküler biyoloji (He ve Haymer 1999) gibi alanlarda uygulamaları yapılmıştır. TKA'nin gelişiminde Householder ve Young (1938), Gower (1966), Krzanowski ve Marriot (1994)'nin çalışmalarının katkıları büyüktür. Young ve Householder (1938), gözlemler arasındaki uzaklıkların bir kümesinin koordinatlarını yeniden yapılandırma fikrini ilk olarak ortaya atmışlardır.

$\mathbf{X}$, $n \times p$ boyutlu elemanları x_{ij}'ler olan bir veri matrisi olsun. Burada n gözlem sayısını, p değişken sayısını ifade eder. x_{ij} elemanı, j. boyutta ölçülen i gözleminin değeridir. p boyutlu bir Öklid uzayında n gözlemin koordinatları $\mathbf{x}_i \ (i = 1, 2, ..., n)$ ile verilir.

Burada $\mathbf{x}_i = \left(x_{i1}, ..., x_{ip}\right)^T$ olarak tanımlanmıştır. *i* ve *t* gözlemleri arasındaki Öklid uzaklığı,

$$\begin{aligned} d_{it}^2 &= \left(\mathbf{x}_i - \mathbf{x}_t\right)^T \left(\mathbf{x}_i - \mathbf{x}_t\right) \\ &= \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_t^T \mathbf{x}_t - 2\mathbf{x}_i^T \mathbf{x}_t \end{aligned} \tag{5.4}$$

ile verilir.

Gözlemler arasındaki uzaklıklar, $\mathbf{K}_{n \times n} = \left(\mathbf{X}_{n \times p}\right)\left(\mathbf{X}_{n \times p}\right)^T$ matrisi yardımıyla açıklanabilir. Burada $\mathbf{X} = \left[\mathbf{x}_1, ..., \mathbf{x}_n\right]^T$, $n \times p$ boyutlu koordinatlar matrisidir. $\mathbf{K}$ matrisinin rankı,

$$\text{Rank}\left(\mathbf{K}\right) = \text{Rank}\left(\mathbf{X}\mathbf{X}^T\right) = \text{Rank}\left(\mathbf{X}\right) = p$$

biçiminde elde edilir. **K** matrisi simetrik, pozitif yarı tanımlı, rankı p olan bir matris olup, p tane negatif olmayan özdeğere ve $n-p$ tane sıfır özdeğere sahiptir.

K bir iç çarpım matrisi olup,

$$\left[\mathbf{K}\right]_{it} = k_{it} = \mathbf{x}_i^T \mathbf{x}_t \tag{5.5}$$

şeklinde tanımlanır.

Eşitlik (5.4),

$$\frac{1}{n}\sum_{i=1}^{n} d_{it}^2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i + \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_t^T \mathbf{x}_t - \frac{2}{n}\sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_t \tag{5.6}$$

biçiminde yazılabilir.

Eğer veri kümesi sıfır ortalama ve birim varyansa göre standartlaştırılmış ise,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_t = 0 \tag{5.7}$$

olacağından (5.6) eşitliği,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} d_{it}^2 &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i + \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_t^T \mathbf{x}_t \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_t^T \mathbf{x}_t \end{aligned} \tag{5.8}$$

olarak yazılabilir.

Benzer şekilde,

$$\frac{1}{n}\sum_{t=1}^{n} d_{it}^2 = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_i^T\mathbf{x}_i + \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t^T\mathbf{x}_t$$
$$= \mathbf{x}_i^T\mathbf{x}_i + \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t^T\mathbf{x}_t \tag{5.9}$$

elde edilir.

Eşitlik (5.4) kullanılarak,

$$-2\mathbf{x}_i^T\mathbf{x}_t = d_{it}^2 - \mathbf{x}_i^T\mathbf{x}_i - \mathbf{x}_t^T\mathbf{x}_t$$
$$\mathbf{x}_i^T\mathbf{x}_t = -\frac{1}{2}\left(d_{it}^2 - \mathbf{x}_i^T\mathbf{x}_i - \mathbf{x}_t^T\mathbf{x}_t\right) \tag{5.10}$$

elde edilir. Buradan da,

$$\left[\mathbf{K}\right]_{it} = k_{it} = \mathbf{x}_i^T\mathbf{x}_t = -\frac{1}{2}\left(d_{it}^2 - \mathbf{x}_i^T\mathbf{x}_i - \mathbf{x}_t^T\mathbf{x}_t\right) \tag{5.11}$$

biçiminde yazılabilir.

Eşitlik (5.8) ve (5.9) kullanılarak,

$$\mathbf{x}_t^T\mathbf{x}_t = \frac{1}{n}\sum_{i=1}^{n} d_{it}^2 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^T\mathbf{x}_i \tag{5.12}$$

$$\mathbf{x}_i^T\mathbf{x}_i = \frac{1}{n}\sum_{t=1}^{n} d_{it}^2 - \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t^T\mathbf{x}_t \tag{5.13}$$

eşitlikleri elde edilir.

Eşitlik (5.11)'de (5.12) ve (5.13) eşitlikleri yerine yazılıp düzenlendiğinde,

$$\left[\mathbf{K}\right]_{it} = k_{it} = \mathbf{x}_i^T\mathbf{x}_t = -\frac{1}{2}\left(d_{it}^2 - \frac{1}{n}\sum_{t=1}^{n} d_{it}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{it}^2 + \frac{2}{n}\sum_{i=1}^{n} \mathbf{x}_i^T\mathbf{x}_i \right) \tag{5.14}$$

elde edilir.

Eşitlik (5.9) kullanılarak,

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} d_{it}^2 &= \mathbf{x}_i^T\mathbf{x}_i + \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t^T\mathbf{x}_t \\ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n} d_{it}^2 &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^T\mathbf{x}_i + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n} \mathbf{x}_t^T\mathbf{x}_t \\ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n} d_{it}^2 &= \frac{2}{n}\sum_{i=1}^{n} \mathbf{x}_i^T\mathbf{x}_i \end{aligned} \tag{5.15}$$

eşitliği elde edilir.

Eşitlik (5.14)'deki $\frac{2}{n}\sum_{i=1}^{n} \mathbf{x}_i^T\mathbf{x}_i$ ifadesi yerine (5.15)'de verilen karşılığı kullanılırsa,

$$\left[\mathrm{K}\right]_{it} = k_{it} = \mathbf{x}_i^T\mathbf{x}_t = -\frac{1}{2}\left(d_{it}^2 - \frac{1}{n}\sum_{t=1}^{n} d_{it}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{it}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n} d_{it}^2 \right) \tag{5.16}$$

elde edilir. Burada, $\left[\mathbf{D}\right]_{it} = \frac{1}{2}d_{it}^2$ alınırsa, (5.16) eşitliği

$$\left[\mathbf{K}\right]_{it} = k_{it} = \mathbf{x}_i^T\mathbf{x}_t = -\left(D_{it} - D_{i\bullet} - D_{\bullet t} + D_{\bullet\bullet}\right) \tag{5.17}$$

biçiminde yazılabilir. Burada, $D_{i\bullet} = \frac{1}{n}\sum_{t=1}^{n} D_{it}$, $D_{\bullet t} = \frac{1}{n}\sum_{i=1}^{n} D_{it}$, $D_{\bullet\bullet} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n} D_{it}$ olduğundan, (5.17) eşitliği,

$$\left[\mathbf{K}\right]_{it} = k_{it} = \mathbf{x}_i^T\mathbf{x}_t = -\left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{n}\right)\mathbf{D}\left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{n}\right) \tag{5.18}$$

şeklinde yazılabilir ve $\mathbf{1} = (1,1,...,1)^T$, n tane 1'den oluşan vektör, $\mathbf{I}$ ise $n \times n$ boyutlu birim matristir.

$\mathbf{K}$ matrisi simetrik ve pozitif yarı tanımlı bir matristir. Bu nedenle $\mathbf{K}$ matrisine spektral ayrışım uygulandığında,

$$\mathbf{K}_{n\times n} = \mathbf{V}_{n\times n}\mathbf{\Lambda}_{n\times n}\left(\mathbf{V}_{n\times n}\right)^T \tag{5.19}$$

elde edilir. Burada $\mathbf{\Lambda} = \text{diag}\left(\lambda_1, \lambda_2 ..., \lambda_p, ..., \lambda_n\right)$, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq ... \geq \lambda_n$, $\mathbf{K}$ matrisinin özdeğerlerinin oluşturduğu köşegen matris olup, $\mathbf{V} = \left[\mathbf{v}_1, ..., \mathbf{v}_n\right]$ ise $n \times n$ boyutlu $\mathbf{v}_i^T \mathbf{v}_i = 1$ olan özvektörler matrisidir.

$\mathbf{K}$ matrisi $n - p$ tane sıfır özdeğere sahip olduğundan (5.19) eşitliği,

$$\mathbf{K}_{n\times n} = \mathbf{V}_{n\times p}\mathbf{\Lambda}_{p\times p}\left(\mathbf{V}_{n\times p}\right)^T \tag{5.20}$$

biçiminde yeniden yazılabilir.

Eşitlik (5.20) yeniden düzenlenirse,

$$\mathbf{K}_{n\times n} = \left(\mathbf{V}_{n\times p}\mathbf{\Lambda}_{p\times p}^{1/2}\right)\left(\mathbf{V}_{n\times p}\mathbf{\Lambda}_{p\times p}^{1/2}\right)^T \tag{5.21}$$

olarak yazılabilir ve burada $\mathbf{X} = \mathbf{V}_{n\times p}\,\mathbf{\Lambda}_{p\times p}^{1/2}$ dir. Böylece gözlemlerin koordinatlarını veren p-boyutlu koordinat matrisi, $\mathbf{X}_{n\times p} = \mathbf{V}_{n\times p}\mathbf{\Lambda}_{p\times p}^{1/2}$ eşitliği ile bulunabilir. İki boyutlu bir grafiksel yaklaşım için koordinatlar ise, $\mathbf{X}_{n\times 2} = \mathbf{V}_{n\times 2}\mathbf{\Lambda}_{2\times 2}^{1/2}$ eşitliği ile elde edilir.

Başka bir deyişle, gözlemlerin koordinatları, $\mathbf{X}\mathbf{X}^T$ matrisinin özdeğerleri ve özvektörlerinin belirlenmesi ve daha sonra özdeğerlerin karekökünün oluşturduğu matris

ile özvektörler matrisinin çarpılması ile bulunur. $\mathbf{X}$ veri kümesinin değişkenleri arasındaki uzaklık matrisinin tanımlanması ile $\mathbf{X}$'in indirgenmiş boyutta sütun değişkenlerinin grafiği bulunabilir (Park vd. 2008).

Gower (1966), TBA yöntemi ile karışmaması için bu yönteme TKA adını vermiştir. Gower, TKA'nin TBA'nın bir eşi olduğunu, benzer çözümler içerdikleri, ancak TBA'da değişkenler arasındaki ilişkilerin $p \times p$ boyutlu varyans-kovaryans veya korelasyon matrisi temel alınırken, TKA'nın gözlemler arasındaki birlikteliklerin $n \times n$ tipinde bir uzaklık matrisini temel aldığını vurgulamıştır.

5.2 Uzaklık Analizi

n tane birimin g tane grup içine $n_1, n_2, ..., n_g$ büyüklüklerinde bölündüğünü varsayalım. Gruplandırma $n \times g$ boyutlu $\mathbf{G}$ matrisi ile gösterilebilir. Eğer i birimi r grubunda yer alıyorsa $g_{ir} = 1$, diğer durumda $g_{ir} = 0$'dır. $\mathbf{N} = \text{diag}\left(n_1, ..., n_g\right)$, $n_1, n_2, ..., n_g$ grup birim sayılarından oluşan köşegen bir matristir. $\mathbf{m} = \left(n_1, ..., n_g\right)$ ise, grup birim sayılarını içeren bir vektördür.

Grup ortalamalarının koordinatları ise,

$$\bar{\mathbf{X}}_{g \times p} = \mathbf{N}_{g \times g}^{-1} \left(\mathbf{G}_{n \times g}\right)^T \mathbf{X}_{n \times p} \tag{5.22}$$

eşitliği ile elde edilir.

Gower ve Krzanowski (1999), toplam karesel uzaklıklar toplamını $\left(\mathbf{T}\right)$, grup-içi $\left(\mathbf{W}\right)$ ve gruplar-arası $\left(\mathbf{B}\right)$ karesel uzaklıklar toplamlarının toplamı olarak $\mathbf{T} = \mathbf{W} + \mathbf{B}$ biçiminde ayrıştırmışlardır. Burada,

$$\mathbf{T} = \frac{1}{n} \mathbf{1}^T \mathbf{D} \mathbf{1} \tag{5.23}$$

$$\mathbf{W} = \sum_{r=1}^{g} \frac{1}{n_r} \mathbf{1}_r^T \mathbf{D}_{rr} \mathbf{1}_r \tag{5.24}$$

$$\mathbf{B} = \frac{1}{n} \mathbf{m}^T \bar{\mathbf{\Delta}}\, \mathbf{m} \tag{5.25}$$

dir.

$\mathbf{D}$ matrisi g^2 tane alt matrise bölünür, $\mathbf{D}_{rs}$ $(r,s = 1,...,g)$. Burada $n_r \times n_s$ boyutlu $\mathbf{D}_{rs}$, s. gruptaki her bir gözlem ve r. gruptaki her bir gözlem arasındaki 2 ile bölünmüş karesel uzaklıkları içerir. Eşitlik (5.24)'deki $\mathbf{D}_{rr}$, r gruptaki gözlemlerin tüm çiftleri arasındaki uzaklıkları içeren $\mathbf{D}$ matrisinin alt matrisidir.
Eşitlik (5.25)'deki $\bar{\mathbf{\Delta}}$, $g \times g$ boyutlu simetrik bir matris olup, r ve s grup ortalamaları arasındaki 2 ile bölünmüş karesel uzaklığın (r,s)'inci elamanını verir (Gardner vd. 2005).

5.3 Uzaklık Analizi ile Biplot Kullanımı

Bölüm 5.2'de verilen uzaklık analizi ile Biplot birlikte oluşturulabilir. Bu oluşum, MANOVA'ya eşlik eden KDA Biplot'a benzer ve gruplar arasındaki ayrışma ve örtüşmenin grafiksel temsilini sağlar.

UA Biplot, değişken sayısı, gözlem sayısından fazla ise etkileyici boyut indirgeme özelliklerine sahiptir (Gardner vd. 2009). Büyük veri kümelerin analizlerinde MANOVA varsayımlarının sağlanması zor olduğundan, UA kullanımı bu tür veri kümeleri için önemlidir.

Grup içi varyans-kovaryans matrisleri farklı veya tekil olduğunda, grupların ayrılması için biplot kullanımının gereği Gardner vd. (2005) tarafından ele alınmıştır. Bu biplot,

Gower ve Krzonowski (1999) tarafından ele alınan UA'nın temeline dayanır. UA, grup ortalamalarındaki değişimi optimal bir şekilde göstermeyi amaçlar.

KDA Biplot'un tüm genişlemeleri ve özellikleri UA Biplot için de mevcuttur. Gardner vd. (2009) çalışmalarında, TBA Biplot ve UA Biplot'un gen açıklama verisi üzerinde uygulamasını ele almışlar ve sonuç olarak bu iki Biplot yaklaşımının birkaç özellik hariç yakın sonuçlar verdiklerini tespit etmişlerdir. Eğer amaç bir veri kümesindeki değişimi yani varyansı araştırmak ise, TBA Biplot'un iki boyutlu bir yaklaşım ile bunu sağlayacağını, ancak amaç grup farklılıklarını araştırmak ise KDA Biplot ve UA Biplot'un daha uygun olduğunu göstermişlerdir. UA, grup ortalamaları arasındaki uzaklıkları maksimize eder. Grup yapısının grafiksel olarak incelenmesi için UA Biplot kullanılır.

Grup ortalamaları için $\bar{\mathbf{X}}$ koordinatları,

$$-\left(\mathbf{I}_g - \frac{\mathbf{1}\mathbf{1}^T}{g}\right)\mathbf{D}^*\left(\mathbf{I}_g - \frac{\mathbf{1}\mathbf{1}^T}{g}\right) = \bar{\mathbf{X}}\bar{\mathbf{X}}^T \tag{5.26}$$

eşitliğinin spektral ayrışımından bulunabilir. Burada, $\mathbf{D}^* = \mathbf{N}^{-1}\mathbf{G}^T\mathbf{D}\mathbf{G}\mathbf{N}^{-1}$'dir. Bu değer TKA kullanılarak elde edilebilir (Gardner vd. 2005). Eksenlerin orijini grup genişlikleri ile ağırlıklandırılmamış grup ortalamaları noktalarının merkezindedir.

Alternatif olarak ağırlıklandırılmış TKA,

$$-\left(\mathbf{I}_g - \frac{\mathbf{1}\mathbf{m}^T}{n}\right)\mathbf{D}^*\left(\mathbf{I}_g - \frac{\mathbf{m}\mathbf{1}^T}{n}\right) = \bar{\mathbf{X}}\bar{\mathbf{X}}^T \tag{5.27}$$

eşitliği ile de ifade edilebilir.

Bu durumda, grafiksel gösterimde, eksenlerin orijini n gözlem noktalarının merkezindedir.

İki boyutlu grafiksel yaklaşımda, gözlem noktalarını ve onlara karşılık gelen grup ortalamalarını göstermek için,

$$-\frac{1}{2}\left(\bar{\mathbf{X}}^{*\mathrm{T}}\bar{\mathbf{X}}^{*}\right)^{-1}\bar{\mathbf{X}}^{*\mathrm{T}}\left(\mathbf{I}_{\mathrm{g}}-\frac{1}{\mathrm{n}}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\left(\mathbf{d}_{\mathrm{i}}-\mathbf{d}_{0}\right) \tag{5.28}$$

eşitliği kullanılır (Gardner vd. 2005).

Eşitlik (5.28)'deki $\bar{\mathbf{X}}^{*}$, (5.27) eşitliğinin spektral ayrışımından bulunan $\bar{\mathbf{X}}$'nın iki boyutta yaklaşımını gösterir. Burada, $\mathbf{d}_{\mathrm{i}}$ vektörü, grafik üzerinde g grup ortalamasının her birine i. gözlem noktasının karesel uzaklıklarını içerir. $\mathbf{d}_{0}$ ise, her bir grup ortalamasının sırasıyla eksenlerin orijinine olan karesel uzaklığıdır.

Eğer ağırlıklandırılmış TKA kullanılırsa (5.28) yerine,

$$-\frac{1}{2}\left(\bar{\mathbf{X}}^{*\mathrm{T}}\bar{\mathbf{X}}^{*}\right)^{-1}\bar{\mathbf{X}}^{*\mathrm{T}}\left(\mathbf{I}_{\mathrm{g}}-\frac{1}{\mathrm{n}}\mathbf{1}\mathbf{m}^{\mathrm{T}}\right)\left(\mathbf{d}_{\mathrm{i}}-\mathbf{d}_{0}\right) \tag{5.29}$$

eşitliği kullanılır.

UA biplot elde etmek için, orijinal değişkenler ile ilgili bilgi, ölçeklendirilmiş eksenler şeklinde eklenmektedir (Gardner vd. 2005).

6. BAĞLANTI BİPLOT

Çoklu regresyon modelinde, açıklayıcı değişkenler arasında güçlü bir ilişki olmaması varsayımına bağlıdır. Açıklayıcı değişkenler arasında bir ya da daha fazla doğrusal bağıntının olması, varsayımın bozulmasına neden olur ve bunun sonucu olarak çoklubağlantı problemi ortaya çıkar.

Korelasyon, sadece iki değişken arasındaki doğrusal ilişkiyi ifade ederken çoklubağlantı, iki değişken arasında var olabildiği gibi, bir değişken ve diğer değişkenlerin doğrusal birleşimleri arasında da var olabilir. Yani korelasyon çoklubağlantının özel bir durumudur. Açıklayıcı değişkenler arasındaki yüksek korelasyon çoklubağlantının göstergesidir, ancak tersi doğru değildir. Yani açıklayıcı değişkenler arasında çoklubağlantı varsa, ilgili bu değişken çiftleri arasındaki korelasyon yüksek olmayabilir.

Açıklayıcı değişkenler arasındaki yüksek korelasyon ve etkili gözlemlerin bir sonucu olarak çoklu regresyon modellerinde ortaya çıkan tahmin problemleri ilk olarak Belsley vd. (1980) tarafından ele alınmıştır. Çalışmalarında, çoklubağlantıya neden olan kaynakların belirlenmesine yönelik teşhis yöntemlerini incelemişlerdir. Daha sonra Belsley (1991a) bağlantı teşhis yöntemlerinin görselleştirilmesi yönünde önemli çalışmalar yapmıştır. Ayrıca çoklubağlantı teşhisine ilişkin yöntemlerden açıklayıcı değişkenlerin varyans ayrışım oranlarını, koşul indekslerini ve özdeğerlerini bir tabloda vermiştir. Belsley'in önerdiği tablo gösterimlerinin bağlantı teşhis yöntemlerinin yaygın kullanımına katkısı oldukça fazladır. Ancak çoğu zaman problemli açıklayıcı değişkenleri ve onların etkilerini tanımlamak oldukça karmaşık bir iştir.

Chatterjee ve Price (1977), Weisberg (1980), Hocking (1983) tarafından çoklubağlantının belirlenmesi için birtakım yöntemler önerilmiştir. Bu yöntemlerin çoğu, korelasyon matrisi veya korelasyon matrisinin tersini temel almaktadır.

Friendly ve Kwan (2009), çalışmalarında çoklubağlantı teşhisini karmaşık bir bulmacaya benzetmişlerdir. Ayrıca bağlantı teşhisi için Tableplot olarak adlandırılan

görselleştirilmiş bir yapı önermişlerdir. Çalışmalarında, Gabriel (1971) tarafından önerilen biplot grafiksel yaklaşımının, bağlantı teşhisinde kullanımını tanımlamışlar ve bu biplot yaklaşımını bağlantı biplot olarak adlandırmışlardır.

Alt bölümlerde tableplot ve bağlantı biplot oluşturmak için gerekli teorik yapı kısaca özetlenecektir.

6.1 Çoklu Doğrusal Regresyon Modeli

Çoklu doğrusal regresyon modeli,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{6.1}$$

ile verilsin. Bu modelde, $\mathbf{y}$, $n \times 1$ boyutlu açıklanan değişken vektörü, $\mathbf{X}$, $n \times (p+1)$ boyutlu tam ranklı, ilk sütunu 1'lerden oluşan açıklayıcı değişkenler matrisidir, $\boldsymbol{\beta}$, $(p+1) \times 1$ boyutlu regresyon katsayıları vektörü olup, $\boldsymbol{\varepsilon}$ ise $n \times 1$ boyutlu rasgele hatalar vektörü olarak tanımlanmıştır, $E(\boldsymbol{\varepsilon}) = 0, \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$. $\boldsymbol{\beta}$'nın en küçük kareler tahmin edicisi $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ ile verilir.

$\hat{\boldsymbol{\beta}}$'nın varyansı ise,

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T\mathbf{X})^{-1} \tag{6.2}$$

ile elde edilir.

6.2 Çoklubağlantı ve Negatif Etkileri

Çoklu doğrusal regresyon modelinde, genellikle açıklayıcı değişkenlerin birbirlerinden bağımsız olduğu varsayılır. Ancak çoklu regresyon uygulamalarının çoğunda açıklayıcı değişkenler arasında doğrusal ilişki olabilir. Bu durumda açıklayıcı değişkenler arasında doğrusal bağımsızlık varsayımı geçerli olmaz ve çoklubağlantı problemi ortaya çıkar.

$\mathbf{X}_j$, $\mathbf{X}$ matrisinin j. sütun vektörü olmak üzere, eğer $\sum_{j=1}^{p} c_j \mathbf{X}_j = 0$ olacak şekilde tümü sıfır olmayan $c_1, c_2, ..., c_p$ sabitleri mevcutsa $\mathbf{X}_j$, $j = 1, 2, ..., p$, vektörleri doğrusal bağımlıdır ve tam çoklubağlantı söz konusudur. Bu durumda $\mathbf{X}$ matrisinin rankı p'den küçük olacaktır ve $\mathbf{X}^T\mathbf{X}$ matrisi tekil olacağından tersi bulunamayacaktır. Dolayısıyla regresyon katsayıları hesaplanamayacaktır. Eğer, $\sum_{j=1}^{p} c_j \mathbf{X}_j \cong 0$ eşitliği sağlanırsa, güçlü çoklubağlantı durumu söz konusudur. Bu durumda, $\mathbf{X}^T\mathbf{X}$ matrisinin tersi alınabilecek, fakat elde edilecek regresyon katsayıları verideki küçük değişikliklere veya modelden bir değişkenin çıkartılmasına (ya da eklenmesine) karşı oldukça duyarlı hale gelecektir. Açıklayıcı değişkenler arasında doğrusal bir ilişki söz konusu değilse, açıklayıcı değişkenler birbirlerine diktir.

Açıklayıcı değişkenler arasında çoklubağlantı olduğunda, $\hat{\boldsymbol{\beta}}$'nın varyansında aşırı büyüme, güvenilmez t ve F istatistikleri, regresyon katsayılarının işaretlerinin beklenenden farklılık göstermesi ve güven bölgelerinin yorumlanmasında karşılaşılan problemler gibi son derece önemli sıkıntılar ortaya çıkar (Farrar ve Glauber 1967, Silvey 1969, Marquardt 1970, Marquardt ve Snee 1975, Gunst ve Mason 1977, Willan ve Watts 1978). Sonuç olarak çoklubağlantı varlığında, çoklu doğrusal regresyon modeli güvensiz ve geniş güven aralığına sahip olacaktır.

6.3 Geleneksel Çoklubağlantı Teşhis Yöntemleri

Tüm korelasyonların düşük olması durumunda dahi çoklubağlantının varlığından söz edilebileceğinden, çoklubağlantının teşhisi için sadece korelasyon matrisinin incelenmesi yanıltıcı sonuçlara götürebilir. Bu yüzden, doğru ve tutarlı bir sonuca ulaşabilmek için, karar aşamasında tüm çoklubağlantı teşhis yöntemlerinden faydalanılması önemlidir.

Çoklubağlantının etkisini ölçmek için varyans büyüme faktörleri, koşul indeksleri, koşul sayısı ve varyans ayrışım oranları hesaplanmaktadır. Bu kavramlar alt bölümlerde açıklanmıştır.

6.3.1 Varyans büyüme faktörleri

Varyans büyüme faktörleri, her bir regresyon katsayısı için çoklubağlantının ölçüsünü verir. Bu ölçü, VIF_j ile gösterilir ve

$$VIF_j = \frac{1}{1 - R^2_{j/\bullet}} \tag{6.3}$$

olarak ifade edilir.

Varyans büyüme faktörleri (6.3) eşitliği yardımıyla açıklayıcı değişkenlerin korelasyon matrisinden hesaplanabilir.

Eğer $VIF_j = 1$ ise, $R^2_{j/\bullet} = 0$ olur. Bu durumda $\mathbf{X}_j$, diğer açıklayıcı değişkenler ile doğrusal ilişkili değildir. Eğer $VIF_j \to \infty$ ise, $R^2_{j/\bullet} = 1$ olur. Bu durumda $\mathbf{X}_j$, diğer açıklayıcı değişkenler ile tam bir doğrusal ilişki içindedir. Burada $R^2_{j/\bullet}$, diğer açıklayıcı değişkenler üzerinden $\mathbf{X}_j$'nin çoklu belirtme katsayısıdır (Rawlings vd. 1998, Friendly ve Kwan 2009)

$\widehat{\beta}_j$ tahmin edicisinin varyansı,

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_j\right) = \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \frac{\sigma^2}{(n-1)s_j^2}\left(\frac{1}{1 - R^2_{j/\bullet}}\right) = \frac{\sigma^2}{\mathbf{x}_j^T\mathbf{x}_j} VIF_j \tag{6.4}$$

ifadesi ile verilmektedir.

Eşitlik (6.4)'den, j. regresyon katsayısının VIF_j'ye doğrudan orantılı olduğu görülebilir. Burada $\mathbf{x}_j$, merkezileştirilmiş $\mathbf{X}$ matrisinin j. sütunudur. s_j^2 ise, $\mathbf{X}$ matrisinin j. sütununun örneklem varyansıdır (Theil 1971, Berk 1977, Belsley vd. 1980, Rawlings vd. 1998, Friendly ve Kwan 2009).

Eşitlik (6.4)'deki $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ matrisinin köşegen elemanları, $1/\left(1-R_{j/\bullet}^2\right)$ ile verilir (Rawlings vd. 1998, Kidwell ve Brown 1982).

VIF değerlerinin 10'dan büyük veya ortalama VIF değerinin 6'dan büyük olması çoklubağlantı probleminin varlığının göstergesidir (Hocking 1983).

6.3.2 Özdeğerlerin Kullanımı

Koşul sayısı, koşul indeksleri ve varyans ayrışım oranları gibi çoklubağlantı teşhis yöntemlerinde özdeğerler önemli bir yer tutar. Çoklubağlantı teşhis yöntemleri $\mathbf{R}_{XX}$ veya $\mathbf{X}^T\mathbf{X}$ matrisinin özdeğerlerini kullanır. Burada $\mathbf{R}_{XX}$, açıklayıcı değişkenler arasındaki korelasyonları veren korelasyon matrisidir (Rawlings vd. 1998).

6.3.2.1 Koşul İndeksleri (Kİ)

Koşul indeksleri (Kİ), her bir özdeğer için hesaplanabilir. Özdeğerler en büyük (max) özdeğerden başlayarak azalan sırada sıralanır. Koşul indeksleri, $K\dot{I}_j = \lambda_{max}/\lambda_j$ ile bulunur. Burada λ_p, $j=1,2,...,p$, en küçük özdeğere karşılık gelir. Literatürde bazı yazarlar Kİ değerlerini bulurken $\sqrt{\lambda_{max}/\lambda_j}$ formülünü kullanır (Paulson 2007). SAS and R programlarında $\sqrt{K\dot{I}_j}$ kullanılmaktadır.

Belsley (1991b)'e göre, 30'dan büyük koşul indeksleri sıkıntılıdır, ancak 100'den büyük koşul indeksleri tahminde potansiyel problemin göstergesidir. En büyük koşul indeksi, en büyük boyutta yer almaktadır. 30'un üzerinde koşul indeksi kaç tane varsa o kadar

doğrusal bağımlılık vardır. Doğrusal bağımlılığın hangi değişkenler arasında olduğunu anlamak için varyans ayrışım oranları incelenir.

6.3.2.2 Koşul Sayısı (KS)

Koşul sayısı, $\sqrt{\lambda_{max}/\lambda_{min}}$ oranı ile bulunur. Eğer KS, 100'den büyük ise güçlü çoklubağlantı söz konusudur (Belsley vd. 1980, Rawlings vd. 1998).

6.3.2.3 Varyans Ayrışım Oranları (VAO)

Büyük VIF_j değerleri bazı yakın bağlantılar içeren değişkenleri gösterir. Ancak her bir değişkenin diğer değişkenlerle içerdiği bağlantıları göstermezler. Bu düşünceyle Belsley vd. (1980), Belsley (1991b) her bir boyut için varyans katsayılarının bir ayrıştırması gibi her bir temel bileşen ile her bir değişkenin varyans oranlarının hesaplanmasını önermişlerdir (Friendly ve Kwan 2009).

Açıklayıcı değişkenlerin korelasyon matrisinin spektral ayrıştırması, $\mathbf{R}_{XX} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ ile verilir. Burada $\mathbf{\Lambda}$, $\mathbf{R}_{XX}$ matrisinin özdeğerlerinin azalan sırada yerleştirildiği köşegen matrisi ve $\mathbf{V}$ ise ilgili birim özvektörlerden oluşan ortogonal bir matristir. Ayrıca VIF_j, $\mathbf{R}_{XX}^{-1}$'in köşegen elemanları ile bulunabilir. Bu durumda VIF_j, $VIF_j = \sum_{k=1}^{p} v_{jk}^2 / \lambda_k$, $j = 1,...,p$, olarak yazılabilir.

Böylece VAO,

$$VAO_{jk} = \frac{v_{jk}^2 / \lambda_k}{\sum_{k=1}^{p} v_{jk}^2 / \lambda_k} = \frac{v_{jk}^2}{VIF_j \, \lambda_k} \tag{6.5}$$

biçiminde bulunabilir (Belsley vd. 1980, Fox 1984, Rawlings vd. 1998, Friendly ve Kwan 2009). Burada $v_{jk}\ (j,k=1,...,p)$, **V** matrisinin elemanlarıdır.

Belsley (1991a, 1991b)'ye göre eğer VAO, 0.50'den büyük (bazı kaynaklarda 0.50 yerine 0.90 alınmaktadır) ise iki veya daha fazla açıklayıcı değişken arasında çoklubağlantı meydana gelebilir.

6.4 Görsel Çoklubağlantı Teşhis Yöntemleri

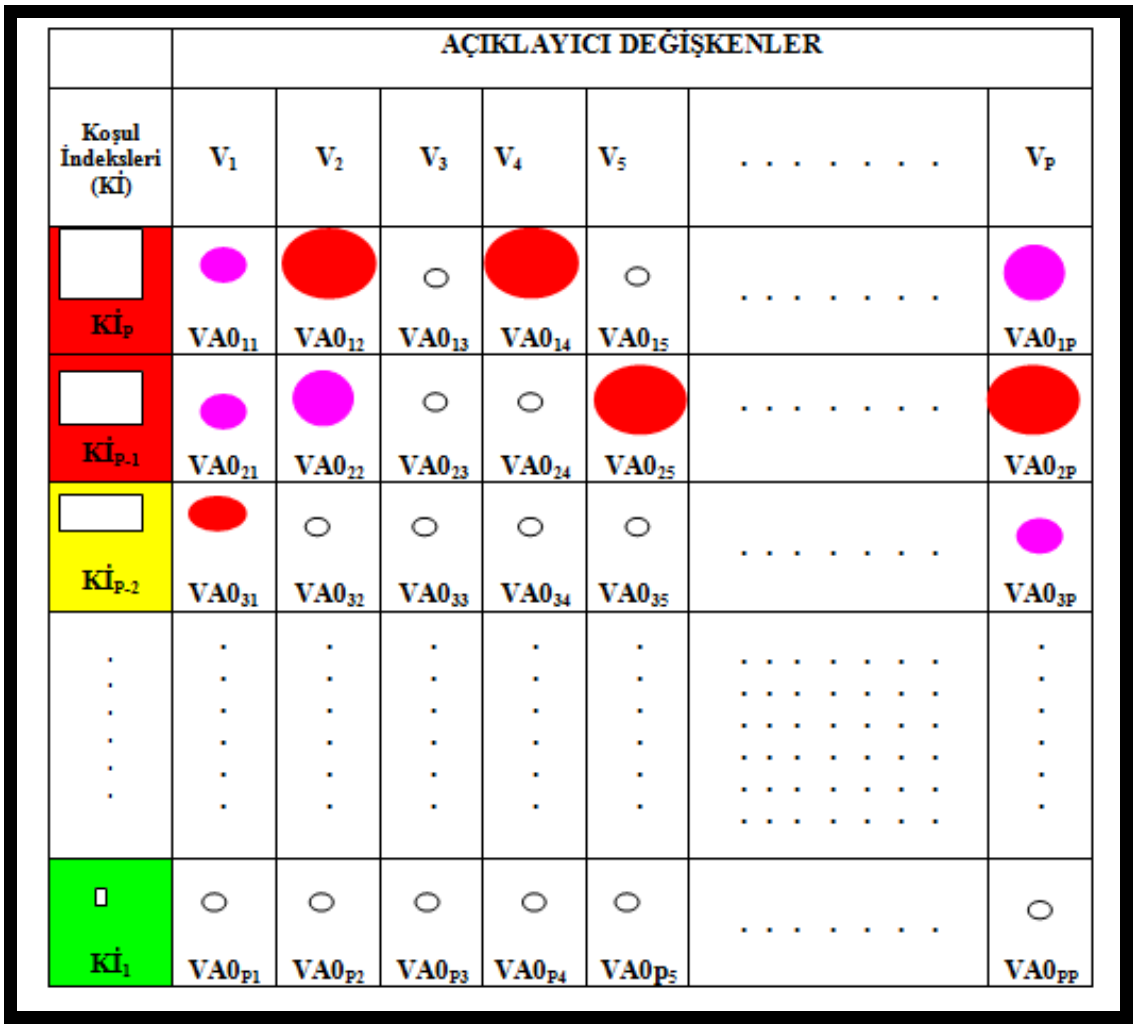
İstatistiksel analizlerde tek bir tabloda verilen bulgulardan genel yapıyı anlamak her zaman mümkün olmayabilir. Bu gibi durumlarda grafiksel gösterimin kullanımı oldukça önemlidir. Analiz sonucunun daha anlaşılır ve eksiksiz yorumlanabilmesi için veri yapısına ve analiz türüne bağlı olarak literatürde birçok grafiksel yöntem önerilmiştir.
Bu alt bölümde, çoklubağlantının görsel teşhisi için son dönemlerde geliştirilen Tableplot ve Biplot yöntemlerinin kullanımıyla ilgili temel bilgiler verilecektir. Ayrıca bu iki yöntemin gerçek bir veri kümesi üzerine uygulaması Bölüm 8.5'de verilecektir.

6.4.1 Tableplot

Tableplot, grafik ve çizelgenin bir kombinasyonu olarak düşünülebilir. Tableplot, mevcut bağlantı teşhis yöntemlerinin grafiksel anlamda sunumunu ve karmaşık sayısal veriler içinde saklı kalmış çoklubağlantıya katkı sağlayan önemli bilgilerin daha göze çarpar şekilde, önemsiz olduğu düşünülen yani çoklubağlantıya katkı sağlamayan bilgilerin daha arka planda kalmasını amaçlar.

Tableplot'un ilk sütunu büyükten küçüğe satırlar halinde sıralanmış koşul indekslerinden oluşur. Burada koşul indeksleri ilk sütunda beyaz karelerden oluşan alanlar ile gösterilir. Bu ilk sütun, koşul indekslerinin büyüklüklerine göre derecelendirilmiş ve arka planı ilgili derecelere göre renklendirilmiştir. Ayrıca diğer sütunların ise varyans ayrışım oranlarına (x100) göre içi renklendirilmiş daire gösterimi kullanılmıştır. Böylece büyük varyans ayrışım oranına sahip olan hücrede büyük daire yer alacak ve bu dairenin rengi çoklubağlantıya katkısına göre belirlenecektir.

Tableplot'un temsili gösterimi şekil 6.1'de verilmiştir. Şekil 6.1'de Sütun 1, koşul indekslerinin ($\sqrt{\text{Kİ}_j}$) değerlerini gösterir. Friendly ve Kwan (2009)'a göre, bir Tableplot gösterimindeki birinci sütundaki hücrelerin arka plan renkleri koşul indeksinin ciddiyetini ifade eder. Eğer $\sqrt{\text{Kİ}_j} < 5$ ise yeşil ("TAMAM"), $5 \le \sqrt{\text{Kİ}_j} < 10$ ise sarı ("UYARI") ve $\sqrt{\text{Kİ}_j} \ge 10$ ise kırmızı ("TEHLİKE") renkleri kullanılarak çoklubağlantının derecesinin görsel olarak belirlenmesi sağlanmaktadır. Diğer sütunların renklendirilmesi varyans ayrışım oranlarına (VAO, x100) göre yapılır. Burada, {0–20, 20–50, 50–100} oranlarına sırasıyla {Beyaz, Pembe, Kırmızı} renkleri karşılık gelir.



Şekil 6.1 Tableplot'un temsili gösterimi

6.4.2 Klasik ve bağlantı biplot

Klasik biplot yaklaşımı ve özellikleri Bölüm 2.1'de ele alınmıştır. Çoklu regresyonda klasik biplot, merkezileştirilmiş açıklayıcı değişkenlerin $(\mathbf{X})$ korelasyon matrisini $(\mathbf{R}_{XX})$ kullanır. Bu biplot, $\mathbf{R}_{XX}$'in temel bileşenler analizi ile indirgenmiş boyutlu uzayda (2 veya 3), gözlemler için temel bileşen skorları ve değişkenler için temel bileşen katsayılarının birlikte grafiğidir. Bu grafikte gözlemler noktalar ile, değişkenler ise vektörler ile temsil edilir. Klasik biplot grafiği, $\mathbf{R}_{XX}$'in en büyük özdeğerlerine karşılık gelen ilk iki veya üç boyut ile oluşturulabilir (Jolliffe 2002, Friendly ve Kwan 2009).

Klasik biplot'da değişken vektörleri üzerine gözlem noktalarının dik izdüşümleri merkezileştirilmiş veri matrisinin elemanlarını yaklaşık olarak verir. Herhangi iki vektör arasındaki açının kosinüsü bu değişkenler arasındaki korelasyonu indirgenmiş boyutlu uzayda yaklaşık olarak verir. Ayrıca koordinat eksenleri ve her bir değişken vektörü arasındaki açılar, koordinat eksenleri ve değişken vektörleri arasındaki korelasyonu yaklaşık olarak verir (Gabriel 1971, Gower ve Hand 1996, Aitchison ve Greenacre 2002).

Klasik biplot kullanışlı bir bilgi vermesine rağmen, biplot oluşturulurken en büyük varyans açıklama oranına sahip boyutları kullandığından, bağlantı teşhisi amacıyla kullanıldığında yanıltıcıdır. Ancak bağlantı problemi, en büyük boyutlarda saklıdır. Bu nedenle Friendly ve Kwan (2009), çoklubağlantı probleminin görsel teşhisinde en küçük özdeğerlere karşılık gelen boyutlar için biplot grafiksel yaklaşımının kullanımının gereğini vurgulamışlar ve bu biplot tipini de "Bağlantı Biplot" olarak adlandırmışlardır. Küçük varyans açıklama oranına sahip boyutlar oldukça büyük koşul indekslerini beraberinde getirir. Bu büyük koşul indeksleri ise çoklubağlantının göstergesidir.

6.4.2.1 Bağlantı biplot ile varyans ayrışım oranlarının görselleştirilmesi

p en büyük boyutu göstermek üzere, p. ve (p-1). boyutlar arasında oluşturulacak biplot "Bağlantı Biplot" olarak adlandırılır.

Boyut p ve Boyut (p-1) eksenleri üzerine değişken vektörlerinin izdüşümleri varyans ayrışım oranları ile orantılıdır. Bu değişken vektörlerinin göreceli uzunlukları bu iki yakın tekil boyutlar için her bir değişkenin bağlantıya katkısını belirlemede göz önüne alınabilir (Friendly ve Kwan 2009) .

Böylece Boyut p'nin büyük bölümünün hangi değişken ile açıklandığı ve daha sonraki en büyük katkının hangi açıklayıcı değişken ile yapıldığı görülmektedir. Aynı şekilde Boyut (p-1)'de hangi değişkenler arasında güçlü ilişkiler olduğu da görülmektedir. Ayrıca değişken uzayında merkezden uzak yerlerde yer alan ve aykırı değer gibi duran gözlemlerin de teşhisi yöntemin gücünü artırmaktadır. Aykırı değerlerin teşhisine yönelik farklı grafiksel yaklaşımlar da mevcuttur. Fakat bu bilginin "Bağlantı Biplot" ile de görülmesi önemlidir.

6.4.2.2 Bağlantı biplot ile koşul indekslerinin görselleştirilmesi

Koşul indeksleri, $\sqrt{\lambda_1/\lambda_p}$, oranları olarak tanımlanır. Oluşturulacak bir grafik sayısal koşul indekslerinden daha fazla bilgi verir. Boyut 1'e karşın Boyut p ile çizilecek bir Biplot'dan bu görülebilir. Burada Boyut p, en küçük özdeğere karşılık gelen boyuttur. λ_1'in λ_p'ye göreceli boyutunu göstermek için, bileşen skorları için bir veri elipsini bu grafikte göstermek kullanışlıdır. Genellikle açıklayıcı değişken vektörlerinin göreceli uzunlukları yaklaşık olarak bu değişkenlerin çoklubağlantıya katkısını gösterir. Gözlemler, açıklayıcı değişken vektörleri üzerine dik izdüşümleri alınarak ayrıca yorumlanabilir (Friendly ve Kwan 2009) .

Büyük koşul indekslerine katkısı olan gözlemler, en büyük boyutta yani Boyut p'de büyük izdüşüm ile yer alacaktır.

7. ALAN BİPLOT

Çok değişkenli analiz metotlarının çoğu, bir $\mathbf{c}^T\mathbf{d}$ iç çarpım ölçüsüyle ifade edilir. Geleneksel gösterim, **d** vektörü üzerinde **c** nin izdüşümü ve sonucun **d** vektörünün uzunluğu ile çarpılmasını içerir.

Bu bölümde, yeni bir alternatif görselleştirme olan Alan Biplot, iç çarpıma yaklaşmak için iki nokta ve orijin ile gerilen bir üçgenin alanını kullanır. Tanımlanan alanlar iki boyutludur. Ancak burada, yüksek boyutlu bir çözümü göstermek için 1'den daha fazla iki boyutlu grafiği birleştirmenin kolay olduğu da görülecektir. Alanların yorumlanması biraz zor olmasına rağmen, geleneksel biplot yöntemlerindeki izdüşüm yorumlama işinden daha az zor olduğu ifade edilebilir.

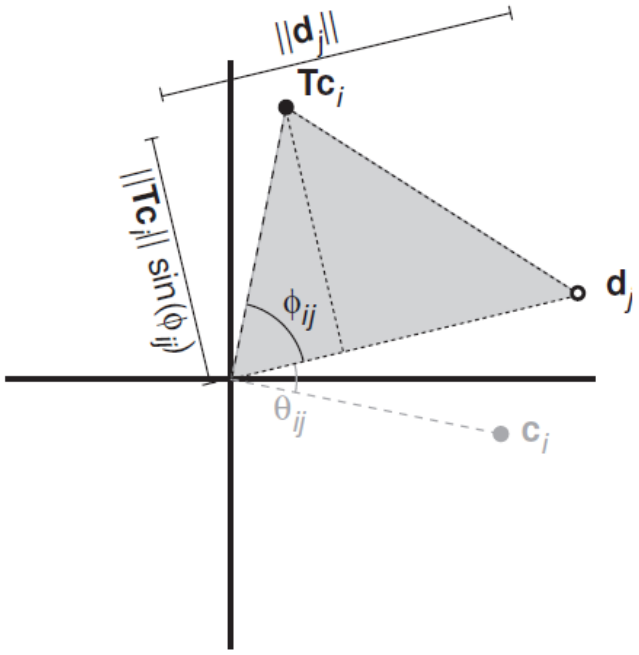
Alan Biplot, karşılaştırılabilir ve aynı ölçek üzerinden herhangi iki veri değeri olan biplot durumlarında özellikle kullanışlıdır. Bu bölümde temel bileşenler analizi için Alan Biplot'ların elde edilmesi, yorumlanması konuları ele alınacaktır. Ancak Alan Biplot'un genele uygulanabilirliği vardır.

7.1 Alan Biplot Oluşturmada Temel Bilgiler

İki boyutlu Alan Biplot fikri çok basittir. Noktaların bir kümesi (örneğin satır noktaları) 90° döndürülebilir. x_{ij} veri değerine yaklaşım, j sütun vektörünün son noktası ve döndürülmüş i gözlem noktası ve orijin ile belirlenen alandan bulunabilir. Aslında vektörlerin gösterimi gereksizdir. Her bir Alan Biplot'da ne ölçeklendirilmiş eksenlere ne de sütun noktasının vektör uzunluğuna gerek vardır. Üçgenlerle parçalanmış alanları karşılaştırmak göz becerisine dayanır. Veri kümesinde değerlerin herhangi bir çifti anlamlı bir şekilde karşılaştırılacağı zaman, TBA Alan Biplot özellikle kullanışlıdır. TBA'da genellikle veri z-skorlarına göre sıfır ortalamalı ve değişken başına eşit varyansla birimleştirilir. Bu dönüşümden sonra değerlerin herhangi bir çifti doğrudan karşılaştırılabilir.

Ölçeklendirilmiş biplot'da yaklaşık değerlerin görsel karşılaştırmasının cazibesi yanıltıcı olabilir. Buna karşılık Alan Biplot, satır ve sütunların herhangi bir çiftinin birleşim alanlarını direkt karşılaştırmaya izin verir. Alan Biplot geliştirmek için, $\mathbf{T}$; 90° lik bir dönüşüm matrisi olsun. Döndürülen satır noktaları, orijinden bir vektörün uzunluğunu değiştirmez. $90^\circ\left(\pi/2\right)$ döndürüldüğünde $\cos\left(\theta_{ij}\right)=\sin\left(\theta_{ij}+\frac{\pi}{2}\right)=\sin(\phi_{ij})$ dir.

Burada, $\phi_{ij}=\frac{\pi}{2}-\theta_{ij}$; $\mathbf{d_j}$ sütun noktası ve $\mathbf{Tc_i}$ döndürülmüş satır noktası arasındaki açıdır. Bu nedenle, $\mathbf{Tc_i^T d_j}=\left\|\mathbf{Tc_i}\right\|.\left\|\mathbf{d_j}\right\|.\sin\left(\phi_{ij}\right)$ olarak ifade edilebilir (Gower vd. 2010).



Şekil 7.1 Alan Biplot'un geometrik yorumu

Şekil 7.1, $\mathbf{c_i}$, $\mathbf{d_j}$ ve döndürülmüş versiyon $\mathbf{Tc_i}$'yi gösterir. ϕ_{ij} ve θ_{ij} açıları $\mathbf{d_j}$'den başlar. Üçgenin yüksekliği, $\left\|\mathbf{Tc_i}\right\|\sin\left(\phi_{ij}\right)$ olup, $\mathbf{d_j}$'ye diktir. Bu nedenle, $\left\|\mathbf{Tc_i}\right\|\sin\left(\phi_{ij}\right)$ yüksekliği ve $\mathbf{d_j}$ tabanı ile üçgenin alanı tam olarak $\frac{\mathbf{Tc_i^T d_j}}{2}$ dir. Bu alan $\frac{x_{ij}}{2}$'ye yaklaşıktır.

Alan Biplot'da veri değerlerini yorumlama kuralı yukarıda belirtildiği gibi bir üçgen oluşturmaktır.

x_{ij}'ye pozitif yaklaşımlar, $\mathbf{d_j}$'den saat yönünün tersi yönünde döndürme ile, negatif yaklaşımlar ise saat yönünde döndürme ile bulunur. Başka bir deyişle, döndürülmemiş nokta ve orijinden zihnen bir çizgi çizilmesi ile yaklaşımların işaretine ulaşılabilir. Eğer $\mathbf{d_j}$ ve $\mathbf{Tc_i}$, orijinden geçen aynı çizgi üzerinde ise sıfır alanlar meydana gelir. $\mathbf{d_j}$'ye paralel $\mathbf{Tc_i}$'den geçen tüm noktalar $\mathbf{d_j}$ ile aynı alanı üretir.

8. GERÇEK VERİ UYGULAMALARI

Çalışmanın bu aşamasında gıda, ekonomi ve enerji alanından elde edilen toplam altı farklı veri kümesine önceki bölümlerde incelenen yöntemlerin uygulamaları verilmiştir.

Bu çalışmada TBA biplot, KDA biplot, UA biplot yöntemleri için R sür.2.11.1 (R Development Core Team 2010) yazılımında Anthony la Grange tarafından geliştirilen BiplotGUI kütüphanesi; Bağlantı Biplot ve Gabriel (1971)'in biplot yaklaşımı için MATLAB'da yazılan kodlar; Tableplot için, R yazılımında Ernest Kwan ve Michael Friendly tarafından geliştirilen tableplot kütüphanesi; Alan Biplot için ise Gower vd. (2010) tarafından verilen MATLAB makrosu bir takım eklentilerle modifiye edilerek kullanılmıştır. Ayrıca veri kümelerinin düzenlenmesinde ve yardımcı analizlerin yapılmasında R yazılımında mevcut olan rgl, MASS, numDeriv gibi bir çok kütüphane ile SPSS (SPSS Inc. 2008), MINITAB (Minitab Inc. 2007) ve SAS (SAS Institute Inc. 2009) programları kullanılmıştır.

8.1 Uygulama 1- Gabriel'in Biplot Yaklaşımı

Ekonomide sermaye stoğu arttıkça üretim seviyesi yükselmekte, ancak bu artış hızı sınırlı kalmaktadır. Sermaye kullanımı arttıkça, sermayenin marjinal verimliliği de düşüş göstermektedir. Ancak sermaye stoğu Türkiye'de olduğu gibi yeterli seviyede değilse, işgücü ve verimlilik sabit olmak kaydıyla, kişi başına düşen sermaye miktarı arzu edilen seviyenin altında kalmaktadır. Bu da sürdürülebilir ekonomik gelişmenin önünde önemli bir engel teşkil etmektedir. Özel sektör yatırımlarındaki yavaş artış ve kamunun özellikle 2002 sonrası yatırımlarını ciddi oranda düşürmesi; işgücü talebinin ve sermayenin düşük seyretmesine yol açmaktadır. 2008 ekonomik ve finansal krizi sonrası düşen taleple beraber, sanayi sektörü kapasite kullanım oranı da azalmıştır. Bu durum istihdam piyasasında üretimin istihdam yaratamaması gibi bir sonucu beraberinde getirmiştir. Artış hızı düşük bir istihdam piyasasına rağmen, teknoloji seviyesinde ise göreli daha hızlı bir ivme kaydedilmiştir.

Teknolojik gelişme, kişisel geliri artırabilmekte, çeşitli iktisadi değişkenler arasında coğrafi ve kategorik yakınsama sağlamakta ve entegrasyon ekonomilerinin gelişimine katkıda bulunmaktadır. Özellikle 1990 sonrası dünyada hızla gelişen teknolojik faaliyetler, ara malının dayanıklı imalat sektörü ürünlerine dönüşümünde inovasyon yaratılmasına yol açmıştır. Teknoloji, sanayi sektöründe doğrudan içselleştirilmiş, verimlilik, rekabet ve istihdam yaratımına olumlu katkılarda bulunmuştur. Böylece kalkınma aşamasında sırasıyla tüketim, ara malı üretiminden sermaye malı üretimine geçiş süreci hızlanmıştır. Ülkeler arası teknoloji yakınsaması Barro (1991) ve Mankiw vd. (1992)'de olduğu gibi homojen teknoloji varsayımını ekonomik çerçevede güçlendirmiştir. Bu anlamda Türkiye'nin Avrupa Birliği'ne entegre olması Phillips ve Sul (2006)'da varsayılan heterojen teknoloji kavramını zayıflatmaya başlamıştır.

Ekonomide üretim sürecinde ara ve sermaye malları faktörlerinden yararlanılmakta ve bu mallar sanayi tesislerinde işlenerek nihai ürüne dönüşmektedir. Bir ülkenin her iki çeşit malı da gerekli oranda üretmesi, hızlı ve dengeli kalkınma bakımından gün geçtikçe ön plana çıkmaktadır. Ara malı imalatı üretimindeki artış hızının yükselmesi, ülkenin üretim sürecinde gücünü artırmakta ve dışa bağımlılığını azaltmaktadır. Ara malı dayanıksız tüketim malı üretiminde önemli oranda kullanılmaktadır. Gıda, içecek ve giyim gibi dayanıksız tüketim mallarını arz eden sektörler istihdama önemli katkıda bulunmaktadır. Ev eşyası, otomotiv gibi dayanıklı tüketim malları ise ara ve sermaye mallarına önemli oranda ihtiyaç duymaktadır. Kriz zamanlarında ve işçevrimlerinde tüketim ömürleri farklı dayanıklı ve dayanıksız malların etkilenme dereceleri değişebilmektedir.

İktisadi analizde görselliği artırmak, düşünce açılımı sağlamak ve kelimelerin yetersiz ya da kapsamının dar olduğu noktalarda çok boyutu daha az boyuta indirgeyen modern grafiksel yöntemlerden yararlanılmaktadır. Bu uygulamada, endüstriyel üretim içinde bir desen oluşturan ve karşılıklı etkileşimleri tartışmalara konu olmuş Ara İmalat Sanayi, Dayanıklı Tüketim Malları, Dayanıksız Tüketim Malları, İmalat Sanayi Endeksi, Sermaye Malları Üretimi ve Toplam Sanayi Üretim Endeksi arasındaki ilişki Gabriel'in Biplot tekniğiyle incelenmektedir.

Uygulama için Türkiye Sanayi Üretim Endeksi (2005=100), Haziran, 2005 – Haziran, 2009 genel ve ana kategorileri veri seti olarak kullanılmaktadır. Ele alınan ana sanayi grupları (MIG), Ara İmalat Sanayi (ARA), Dayanıklı Tüketim Malları (DAYANIKLI), Dayanıksız Tüketim Malları (DAYANIKSIZ), İmalat Sanayi Endeksi (İMALAT), Sermaye Malları Üretimi (SERMAYE) ve Toplam Sanayi Üretim Endeksi (TOPLAM)'dir. Serilerin bir önceki aya göre yüzde değişimleri alınmıştır.

Uygulamada, Bölüm 2.1'de verilen Gabriel'in Biplot yaklaşımı kullanılarak Türkiye sanayi üretiminin altı ana kategorisinde zaman içindeki meydana gelen değişikliklerin ve bu kategoriler arasındaki ilişkilerin görüldüğü bir portre ortaya koymak amaçlanmıştır. Bu çerçevede, çok değişkenli veri yapısının geometrik olarak yorumlanmasında Gabriel'in Biplot yaklaşımı kullanılmış ve değişkenler arasındaki ilişkilerin görsel analizi yapılmıştır.

Gabriel'in Biplot yaklaşımı ile analiz edilen verilere ilişkin sonuçlar çizelge 8.1, çizelge 8.2 ve şekil 8.1'de verilmiştir. Tekil değerler ve her bir tekil değerin toplam varyans açıklama oranına katkısı çizelge 8.1 de yer almaktadır. Bu çizelge incelendiğinde toplam varyansın %90 lık bir bölümünün ilk iki boyutta açıklandığı görülmüştür. Bu oran analiz sonucu elde edilen biplot grafiğinin açıklama gücünün yüksek olduğunu göstermektedir (Johnson ve Wichern 2002).

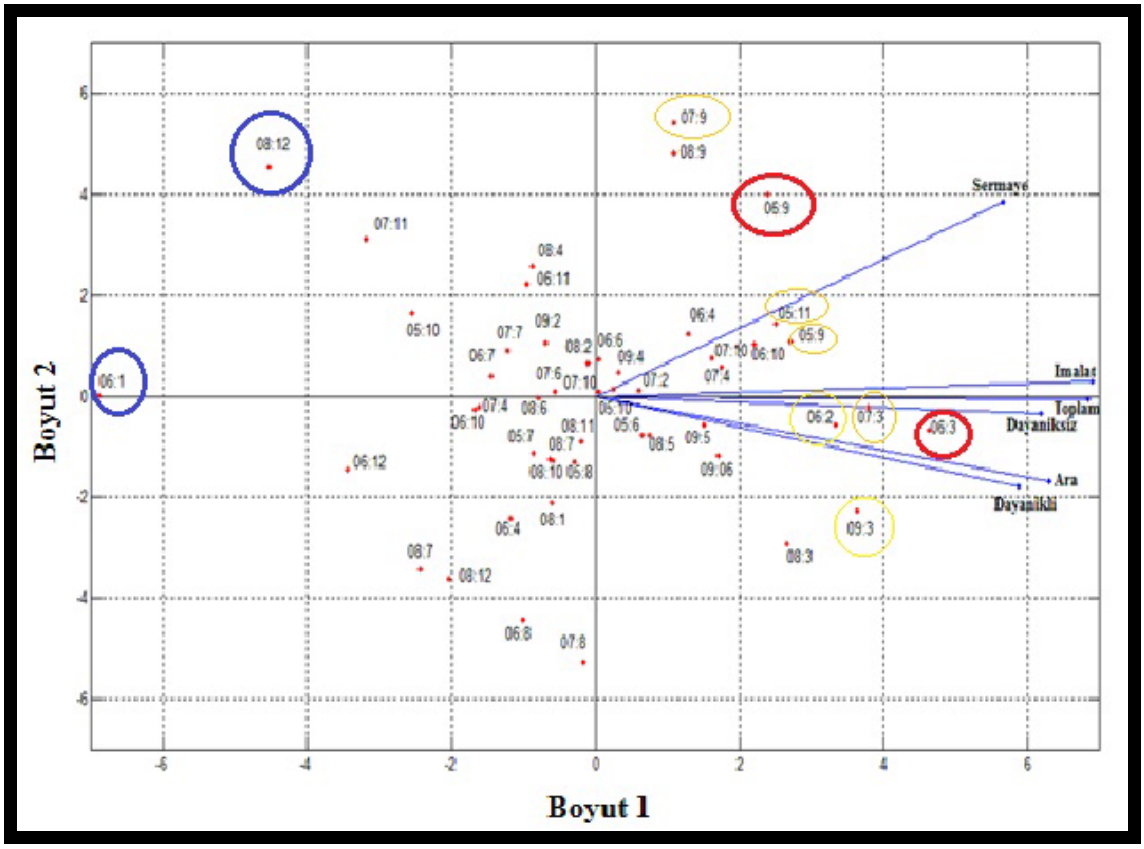
Şekil 8.1'de verilen biplot grafiği incelendiğinde imalat, toplam, dayanıksız, ara, dayanıklı ve sermaye üretimi endekslerindeki büyümeler arasında pozitif yönde bir ilişki olduğu görülmektedir. Hem imalat, toplam ve dayanıksız malların üretim değişkenleri arasında, hem de ara ve dayanıklı malların üretimi arasında yüksek pozitif bir ilişki görülmektedir. Bu ilişkilerin dereceleri ve değişken vektörleri arasındaki açılar çizelge 8.2'de verilmiştir. Değişken vektörleri arasındaki açının verilmesinin nedeni, biplot'da vektörler arasındaki açının kosinüs değerinin yaklaşık olarak korelasyonu vermesidir. Bir başka deyişle, daha küçük açıya sahip iki vektör arasındaki korelasyon daha büyüktür.

Çizelge 8.1 Tekil değerler ve toplam varyans açıklama oranları

Boyut	Tekil Değerler	Tekil Değerlerin Kareleri (Özdeğerler)	Açıklama Oranı	Kümülatif Oran
1	15,4766	239,5244	0,8317	0,8317
2	4,5944	21,1084	0,0733	0,9050
3	3,9076	15,2696	0,0530	0,9580
4	3,4083	11,6165	0,0403	0,9983
5	0,5962	0,3555	0,0012	0,9995
6	0,3545	0,1257	0,0005	1,0000
Toplam		288,0001		

Çizelge 8.2 Korelasyon değerleri

Değişkenler	Korelasyon	Değişken vektörleri arasındaki açı (θ°)
ARA-DAYANIKLI	0,9994	2
ARA-DAYANIKSIZ	0,9782	12
ARA-IMALAT	0,9539	17
ARA-SERMAYE	0,6530	49
ARA-TOPLAM	0,9682	14
DAYANIKLI-DAYANIKSIZ	0,9706	14
DAYANIKLI-IMALAT	0,9432	19
DAYANIKLI-SERMAYE	0,6271	51
DAYANIKLI-TOPLAM	0,9592	16
DAYANIKSIZ -IMALAT	0,9955	5
DAYANIKSIZ - SERMAYE	0,7962	37
DAYANIKSIZ - TOPLAM	0,9991	2
IMALAT-SERMAYE	0,8502	32
IMALAT- TOPLAM	0,9987	3
SERMAYE- TOPLAM	0,8218	35



Şekil 8.1 Gabriel'in Biplot grafiği

Şekil 8.1'de verilen Gabriel'in Biplot grafiğinde dikkat çeken aylar gözlemlendiğinde ise 2006'nın 3 ve 9. aylarında sermaye üretiminin seviyesinin yüksek olduğu görülmektedir. Bu ayları sırasıyla, 2007:03, 2007:09, 2005:11 ve 2005:09 ayları izlemiştir. 2006'nın 3. ayı ve 2006'nın 9. ayında sermaye üretimi yüksek iken, 2006'nın 1. ayı, 2008'in 12. ayı, 2006'nın 12. ayı, 2008'in 7. ayı ve 2008'in 12. ayında en düşük seviyededir. İmalat, toplam ve dayanıksız malların üretimi ise en fazla 2006'nın 3. ayında olmuş ve bu ayı sırasıyla 2007:03, 2009:03 ve 2006:02 ayları izlemiştir. Bu malların üretimi 2006'nın 1. ayında minimum seviyeye düşmüştür.

8.2 Uygulama 2- Gower ve Hand'in Biplot Yaklaşımı

Günlük beslenmenin bir parçası olarak yeterli miktarda meyve ve sebze tüketmek, bulaşıcı olmayan başlıca hastalıkları önlemeye yardımcı olabilir (Ebrahimof vd. 2006). Düşük düzeyde meyve alımı kanser, koroner kalp hastalığı ve inme için önemli bir risk

faktörüdür (Bashirian vd. 2008). Anemi, raşitizm, şişmanlık, basit guatr ve diş çürümesi gibi hastalıklar yanlış beslenme alışkanlıklarının bir sonucu olarak ortaya çıkmaktadır.

Kronik hastalıklara karşı meyvelerin koruyucu etkilerinin önemini ortaya çıkaran kanıtların sayısı her geçen gün artmaktadır. Dünya sağlık örgütü günde yaklaşık olarak 400 gr. meyve çeşidi tüketilmesini önermektedir (WHO 1990). Sağlıklı bir bedenin temel besin olarak yağlar, proteinler ve vitaminlere uygun miktarda ihtiyacı vardır. Enerji verimli organik besinler (karbonhidratlar, yağlar ve proteinler) ve karbon, hidrojen, oksijen elemanlarını katalize eden vitaminler suda yer alan biyoenerjetik reaksiyonlar için gerekli kaynağı sağlamaktadırlar (Driskell 2000).

Bu uygulamada, insan doğasına etki eden başlıca meyve tiplerinin organik ve inorganik kimyasal özelliklerinin Temel Bileşenler Analizi (TBA) Biplot tekniği ile incelenmesi amaçlanmaktadır. Bu çerçevede, çalışmada uygulama materyali olarak, Ek 2'de verilen 25 meyvenin içerdiği 5 besin bileşeni (su, protein, karbonhidrat, yağ, enerji) ile ilgili veri kümesi kullanılmıştır. Ek 2'deki satırlar meyveleri, sütunlar ise her bir meyvenin içinde bulunan 5 besin bileşeninin miktarlarını göstermektedir.

Ek 2'deki veri kümesine TBA Biplot tekniği uygulanarak meyve türleri ve besin bileşenleri arasındaki ilişkileri veren tüm uzaydaki veri yapısının indirgenmiş boyutlu (iki boyutlu) uzaydaki yaklaşımını temsil eden şekil 8.2'deki grafiksel yaklaşım ve Çizelge 8.3 elde edilmiştir.

Şekil 8.2'deki grafiksel yaklaşımın gösterim kalitesi 0.9944 olarak bulunmuştur. Bu oran tüm değişkenlerin iyi kestirildiğinin göstergesidir. Yani elde edilen TBA Biplot grafiğinin açıklama gücü oldukça iyidir.

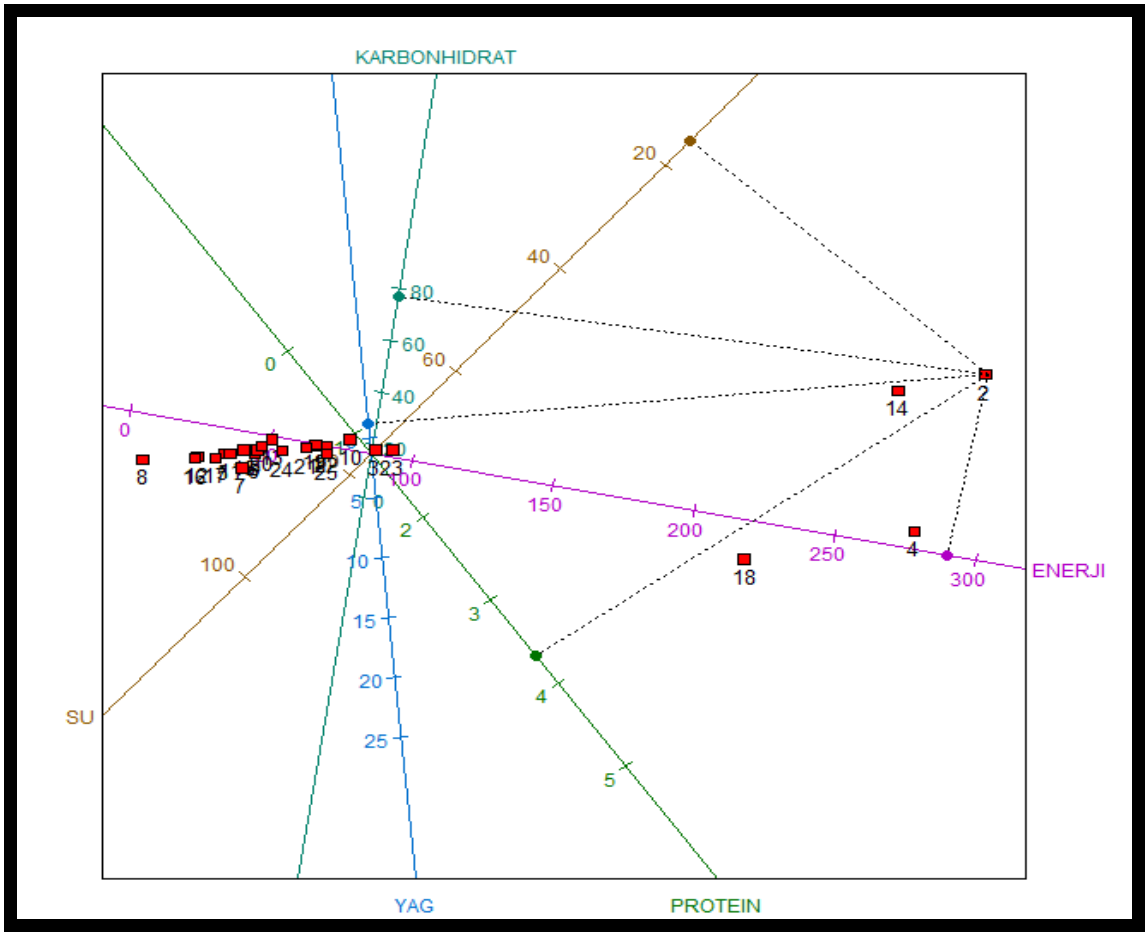
Çizelge 8.3, şekil 8.2'deki grafiksel yaklaşımdaki eksenlerin yeterlilikleri ve kestirimlerini sunar. Çizelge 8.3 incelendiğinde SU, PROTEİN, KARBONHİDRAT ve ENERJİ değişkenlerinin kestirimlerinin oldukça iyi olduğu görülmektedir. Eksen kestirimleri yüksek olması, herhangi bir gözlem noktasının o eksen üzerindeki dik

izdüşümü ile elde edilecek kestirim değerinin de gözlemin gerçek değerine yakın olması anlamında da yorumlanabilir.

Çizelge 8.3 İki boyutlu TBA Biplot için eksen yeterlilikleri ve kestirimleri

	SU	PROTEİN	KARBONHİDRAT	YAĞ	ENERJİ
Yeterlilik	0.1276	0.0006	0.8682	0.0420	0.9615
Kestirim	0.9397	0.8548	0.9840	0.5561	0.9998

Eksen kestirimlerinin eksen yeterliliklerinden büyük olduğu çizelge 8.3'den açık bir şekilde görülmektedir. Yani $\mathbf{X}$'e en yakın en küçük kareler yaklaşımı $\tilde{\mathbf{X}}^{(2)}$ elde edilir. Ayrıca SU, PROTEİN ve YAĞ değişkenlerinin temsil yeterliliklerinin düşük olduğu da çizelge 8.3'den görülmektedir.



Şekil 8.2 TBA Biplot grafiği

Çizelge 8.4, besin bileşenleri vektörlerinin indirgenmiş boyutlu (iki boyut) uzayda koordinatlarını sunar. Bu koordinatlar yardımıyla değişkenler arasındaki korelasyonlar (2.9) eşitliği yardımıyla hesaplanarak şekil 8.2'in yorumlanmasında kullanılır.

Şekil 8.2'de verilen TBA Biplot grafiği incelendiğinde, PROTEİN&YAĞ vektörleri $r = 0.8636$ ile; PROTEİN&ENERJİ vektörleri $r = 0.73$ oranlarında yüksek pozitif korelasyona sahip olduğu ve bunların orijinden benzer yöne uzandığı görülmektedir. Aynı zamanda, KARBONHİDRAT&YAĞ, $r = -0.9712$; KARBONHİDRAT&SU, $r = -0.8670$; KARBONHİDRAT&PROTEİN $r = -0.7188$ ile yüksek negatif korelasyona sahiptir. Bunlar grafikte zıt yönde yer alırlar. Ayrıca YAĞ&ENERJİ arasındaki korelasyondan bu iki değişken arasında pozitif yönde zayıf bir ilişkinin olduğu görülmüştür.

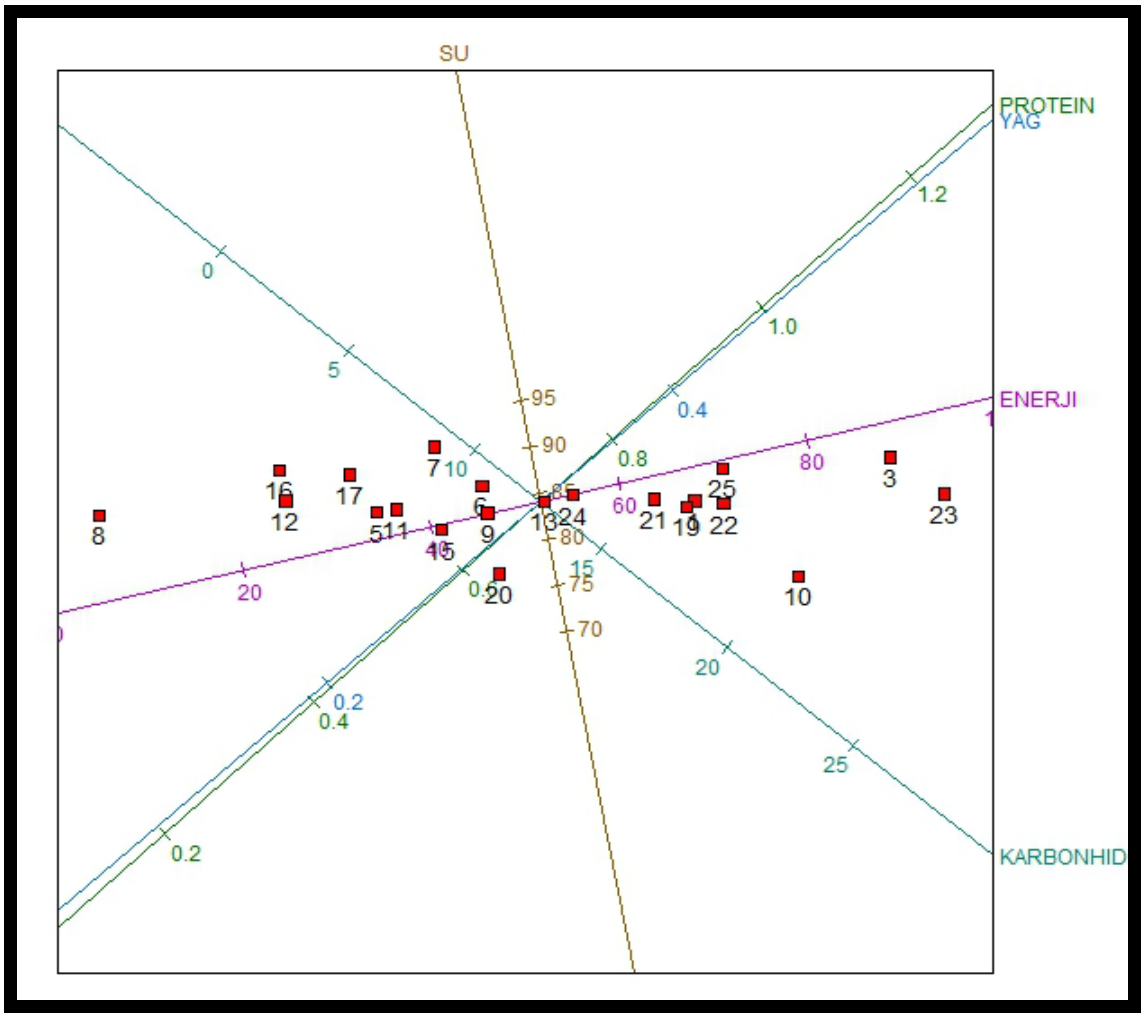
Gözlem noktalarının değişken vektörleri üzerine dik izdüşümleri yardımıyla gözlem noktaları açısından şekil 8.2 yorumlanabilir. Buna göre, YAĞ değişkenini en çok içeren meyvelerin sırasıyla AVOKADO ve KURU İNCİR olduğu görülmektedir. Diğer meyveler yağ içerikleri bakımından hemen hemen benzer bir yapı göstermektedir. PROTEİN ve ENERJİ bakımından en zengin meyvelerin KURU KAYISI, KURU ÜZÜM, AVOKADO VE KURU İNCİR olduğu grafiksel yaklaşımdan net bir şekilde anlaşılmaktadır. Karbonhidrat bakımından en zengin besinlerin ise sırasıyla KURU ÜZÜM, KURU KAYISI ve KURU İNCİR olduğu grafiksel yaklaşımdan çıkarılan diğer bir yorumdur. Ayrıca SU içerikleri bakımından en zengin meyveler ise 8-23 aralığında yer alan gözlem noktalarında verilen meyvelerdir.

Çizelge 8.4 İki boyutlu TBA Biplot grafiksel yaklaşımında besin bileşenlerinin koordinatları

Besin Bileşenleri	Boyut 1	Boyut 2
SU	-0.23796950	-0.26645439
PROTEIN	0.01402029	-0.01966618
KARBONHIDRAT	0.13871620	0.92142358
YAG	0.01862540	-0.20413789
ENERJI	0.96103322	-0.19473470

Şekil 8.2'de diğer gözlemlerden oldukça farklı bir yapı sergileyen 2 (kuru üzüm), 4 (kuru incir), 14 (kuru kayısı) ve 18 (avokado) gözlemleri veri kümesinden çıkarılıp analiz tekrarlandığında şekil 8.3'de verilen biplot grafiksel yaklaşımı elde edilir.

Şekil 8.3 incelendiğinde, 3 (yaş incir) ve 23 (karadut) gözlemlerinin PROTEİN, YAĞ ve ENERJİ içeriği bakımından oldukça benzer olduğu görülmetedir. SU içeriği açısından 10 (muz) gözlemi hariç geriye kalan 20 meyve türünde benzerlik görülmektedir. En az su içeriğinin 10 (muz) gözleminde olduğu şekil 8.3'den tespit edilmektedir. Ayrıca KARBONHİDRAT bakımından sırasıyla 8 (limon suyu), 16 (karpuz), 17 (kavun), 12 (çilek), 7 (greyfurt) meyveleri oldukça zayıf iken 3 (yaş incir), 10 (muz) ve 23 (karadut) meyvelerinde daha yüksektir.



Şekil 8.3 TBA Biplot grafiği, 21 gözlem için

8.3 Uygulama 3- Kanonik Değişken Analizi (KDA) Biplot

Günlük beslenmenin bir parçası olarak yeterli miktarda kolin alımı önemli hastalıkların önlenmesine yardımcı olabilir. Kolin hücre canlılığı için çok önemlidir. Düşük kolin alımı karaciğer için önemli bir risk faktörüdür. Ayrıca kolin eksikliği nörolojik bozukluklarda önemli bir rol oynayabilir. Aşırı kolin alınması ile hipotansiyon, terleme, ishal ve balık vücut kokusu gibi hastalıklar ortaya çıkabilir.

Kolin, 1864 yılında Adolph Strecker tarafından keşfedilmiş ve 1866 yılında kimyasal olarak sentezlenmiştir (Wikipedia, 2010). Kolin, 1998 yılında Amerika Birleşik Devletleri Tıp Enstitüsü Gıda ve Beslenme Kurulu tarafından vazgeçilmez bir besin olarak sınıflandırılmıştır (Institute of Medicine, 1998). Kolin suda eriyen temel bir besin olup, genellikle B kompleks vitaminler içinde gruplandırılır (Zeisel ve daCosta 2009).

Yetişkin kadınlar için günde 425 mg (miligram) kolin alımı yeterli olarak kabul edilirken, hamile ve emziren kadınlar için daha yüksek kolin alımı, yetişkin erkekler için ise 550 mg / gün kolin alımı makul kabul edilir.

Gebelik ve laktasyon (emzirme) dönemlerinde kolin gereksinimi özellikle yüksek olur. Hamile kadınlar için önerilen yeterli kolin alımı 450 mg / gün, laktasyondaki kadınlarda ise bu oran 550 mg / gün olarak elde edilmiştir (Zeisel ve daCosta 2009).

Kolin, besinlerin bir çoğunda bulunur. Kolinin en yoğun bulunduğu besin kaynakları karaciğer, yumurta ve buğday tohumudur. Besinlerde kolin serbest ve esterlenmiş biçimde (phosphocholine, lycerophosphocholine, phosphatidylcholine ve sphingomyelin) bulunur (Zeisel 2006).

Besinlerin toplam kolin içeriği, serbest kolin (FREECHO), glycerophosphocholine (GPC), phosphocholine (PCHO), phosphatidylcholine (PTDCHO) ve sphingomyelin (SM) bileşenlerinin toplamı olarak hesaplanır. Bunlar kolin değişkenleri olarak adlandırılır. Kolin değişkenlerinin gözlenen değerleri besin bileşenlerinde farklı

miktarlarda olabilir. Bu nedenle farklı besin gruplarının kolin değişkenleri açısından karşılaştırılması faydalı olacağı düşünülmüştür.

Uygulamada, beş farklı besin grubunda (G1: süt ve yumurta ürünleri, G2: tavuk ve hindi ürünleri, G3: meyve ve meyve ürünleri, G4: sebzeler ve sebze ürünleri, G5: baklagil ve baklagil ürünleri) yer alan toplam 105 ürünün 100 gramında bulunan beş kolin (FREECHO, GPC, PCHO, PTDCHO, SM) değerini içeren veri kümesi kullanılmıştır. Veriler besinlerin ortak kolin içeriklerinin sunulduğu USDA (US Department of Agriculture) veri tabanından derlenmiştir (Patterson vd. 2008).

Veri kümesine ilk olarak KDA uygulanmıştır. Elde edilen sonuçlar yorumlandığında, beş kolin değişkeninin ilk iki doğrusal birleşiminde, optimal bir şekilde beş grubun ayrıldığı görülmüştür. Bu ilk iki kanonik değişken ile gruplar arası değişimin sırasıyla %68.9 ve %20.2'sinin açıklandığı görülmüştür. Beş besin grubunun birinci ve ikinci kanonik değişken skorlarına göre oluşturulan nokta saçılım grafiği şekil 8.4'de verilmiştir. Şekil 8.4 incelendiğinde, bu beş besin grubunun oluşturduğu yapının, üç farklı grupta, yani Grup 1 ve Grup 2 kendileri birer grup iken Grup 3, Grup 4 ve Grup 5'in birlikte olduğu bir grupta açıklanabileceği görülmektedir.

Şekil 8.4'den beş grubun grup yapıları hakkında açık bir bilgi elde edilir. Fakat bu grafikte orijinal değişkenler hakkında bir bilgi olmamasının ciddi bir eksiklik olduğu görülmektedir. Bu eksiklik veriye KDA Biplot tekniğinin uygulanması ile giderilebilir

Veri kümesine iki değişkenli nokta saçılım grafiklerin çok değişkenli genişlemesi olarak düşünülen Gower ve Hand (1996)'in önerdiği KDA Biplot uygulanması ile şekil 8.5'de verilen grafik elde edilmiştir. Nokta saçılım grafiklerinde sadece iki eksen söz konusu iken biplot metodolojisi kullanılarak elde edilecek grafiksel yaklaşımda her biri bir eksen ile gösterilen p sayıda değişken söz konusudur.

Şekil 8.5'deki Biplot, grupları optimal bir şekilde ayırmayı ve veri noktaları arası uzaklıkları korumayı amaçlamaktadır. Şekil 8.5'deki her bir eksen ölçümlerin orijinal birimlerinde derecelendirilmiştir. Şekil 8.5'de verilen değişken eksenlerinin iki boyutta

kestirimleri çizelge 8.5'de verilmiştir. Çizelge 8.5 incelendiğinde 3 değişkenin de şekil 8.5 de verilen grafiksel gösterimde iyi temsil edildiği görülmektedir. FREECHO ve GPC değişkenlerinin iki boyutta açıklaması çok iyi değildir. Bu durum bu değişkenlerinin kestirimlerinin tutarlı olmayacağı şeklinde yorumlanabilir. Bu iki değişkenin kestirimlerinin daha yorumlanabilir olması için üç boyutlu bir grafiksel gösterim düşünülebilir.

Şekil 8.5'de gösterilen gözlem noktaları, tüm grup içi varyasyonun iki boyutta bir yaklaşımıdır.

Şekil 8.6'de sadece grup ortalamalarının yer aldığı KDA Biplot grafiği verilmiştir. Bu grafiğin beş değişken için beş grup ortalamasını ayırdığı görülmektedir. Grafiksel gösterimde, grup ortalama noktalarından eksenlere dik izdüşümler çizilerek elde edilecek uzaklıklara göre eksenler üzerinden grup ortalamaları yorumlanır. Bu bağlamda şekil 8.6 incelendiğinde, Grup 3 ve Grup 4 ortalamalarında PCHO ve SM değişkenleri açısından fazla bir farklılık olmadığı görülmektedir. Grup 2 ve Grup 3 ortalamaları arasında, FREECHO değişkenine göre grup ortalamaları bakımından bir değişiklik yoktur. PCHO ve SM değişkenleri bakımından grup ortalamaları arasındaki farklılıklar incelendiğinde Grup 2 de bu değişkenlerin etkisinin fazla olduğu şekil 8.6 den yorumlanabilmektedir. PCHO ve SM değişkenlerinin Grup 3, Grup 4 ve Grup 5 ortalamaları üzerindeki etkisi benzer bir yapı göstermektedir. GPC değişkenin etkisi en fazla Grup 1 de görülmektedir. PTDCHO değişkeni açısından en zengin Grup 2 iken, bu gruptaki besin ürünlerini aynı değişken açısından sırasıyla Grup 1, Grup 3, Grup 4, Grup 5 izler. FREECHO değişkenin en fazla olduğu besin grubu Grup 5 iken, en az olduğu besin grubu ise Grup 1'dir.

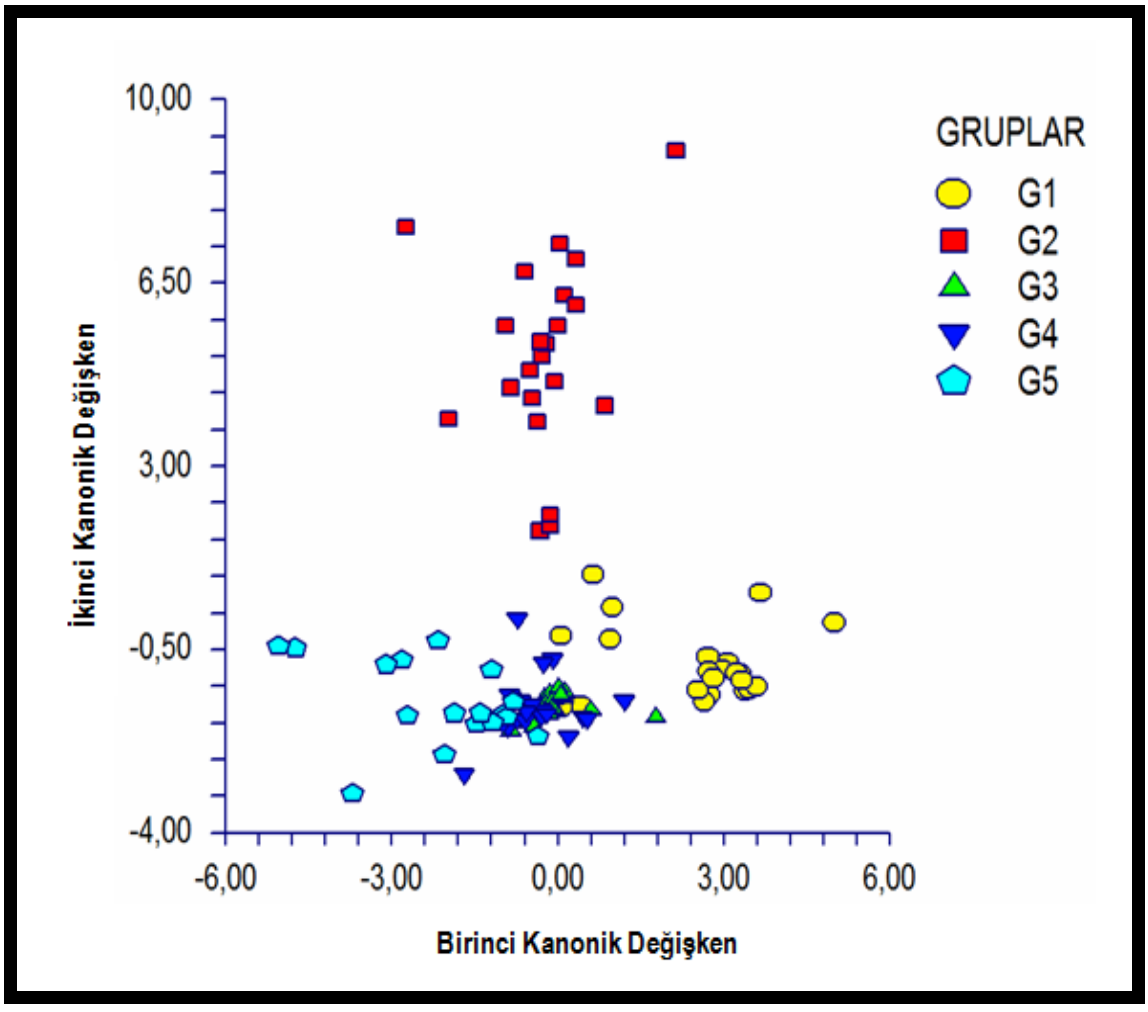
Analiz sonucunda grupların frekanslarında değişmeler olmuştur. Bazı gözlemler gerçek grupları dışında bir grupta yer almıştır. Gözlemlerin gruplandırılması çizelge 8.6'de verilmiştir. Çizelge 8.6 gözlemlerin sınıflandırılmasında mevcut diskriminant fonksiyonlarının ne kadar iyi olduğunu gösteren verileri sunar. Çizelge 8.6 da sunulan veriler ışığında hatalı sınıflandırma olasılığı 0.17 olarak bulunmuştur.

Çizelge 8.5 KDA Biplot grafiğinde eksenlerin kestirimleri

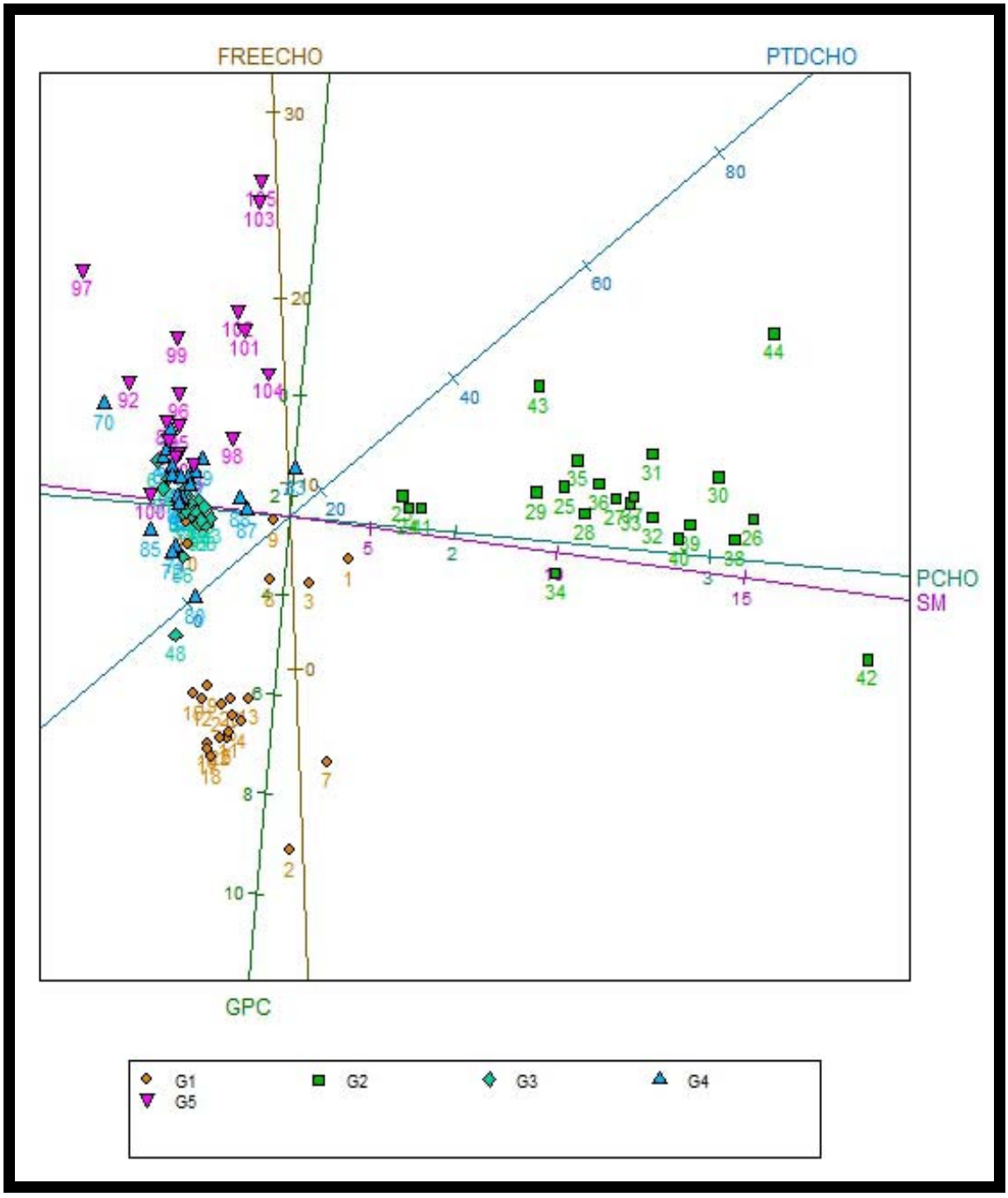
FREECHO	GPC	PCHO	PTDCHO	SM
0.60	0.70	0.83	0.93	0.99

Çizelge 8.6 Gözlemlerin gruplara göre frekansları

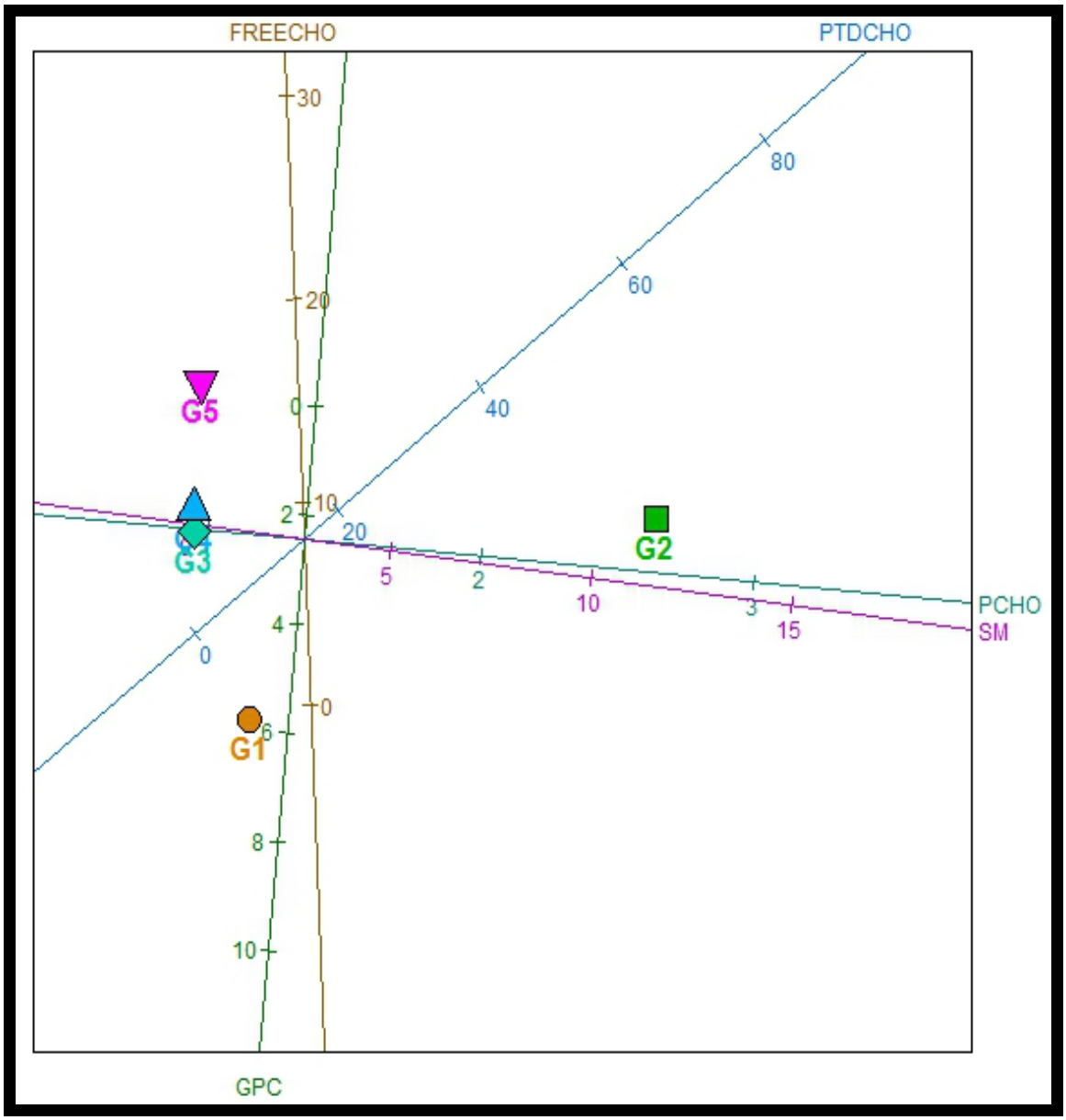
	Kestirilen (Predicted) Grup					
Gerçek(Actual)	Grup 1	Grup 2	Grup 3	Grup 4	Grup 5	Toplam
Grup 1	17	0	5	0	0	22
Grup 2	0	22	0	0	0	22
Grup 3	2	0	18	2	0	22
Grup 4	1	0	6	14	1	22
Grup 5	0	0	1	0	16	17



Şekil 8.4 Beş grubun birinci ve ikinci kanonik değişken skorlarının nokta saçılım grafiği



Şekil 8.5 KDA Biplot Grafiği



Şekil 8.6 Grup ortalamaları için KDA Biplot grafiği

8.4 Uygulama 4- Uzaklık Analizi Biplot

Ham petrol ve ürünleri günümüzde bir çok alanda kullanılan çok önemli bir enerji kaynağı olarak kabul edilir. Petrol karbon ve hidrojen elementlerinden meydana gelen hidrokarbonlardan oluşur. Karbon ve hidrojen elementleri çok çeşitli ve karmaşık molekül yapıları ortaya çıkarır (Wikipedia 2011).

Petrolde, birçok metal element yer alır. En bol bulunan elementler vanadyum ve nikeldir. Petroldeki vanadyum ve nikel miktarlarının (5-40g/ton) petrolün asfalten içeriğinin artması ile arttığı bilinmektedir (Uysal 2006).

Bu uygulamada, Gerrild ve Lantz (1969) tarafından elde edilen ham petrol örneklerinin analizini içeren veri kümesinin Johnson ve Wichern (2002)'de verilen düzenlenmiş şekli kullanılmıştır. Veri kümesi, üç kumtaşı bölgesinden (Bölge I: Wilhelm, Bölge II: Aşagi-Mulinia, Bölge III: Yukarı-Mulinia) alınan 56 ham petrol örneklerinin vanadyum, demir, berilyum, doymuş hidrokarbonlar, aromatik hidrokarbonlar kimyasal özelliklerini içermektedir. Veri kümesi EK 4'de verilmiştir.

Söz konusu değişkenler ile direkt petrol kalitesini belirlemeden ziyade petrolün kökeni (denizel vs.) ve oluşum koşulları gibi yorumlar yapılabilmektedir. Bunun yanında direkt olarak petrol kalitesi, API (American Petroleum Institute) denilen bir indeksle belirlenmekte bunun hesabı ise farklı yöntemlerle elde edilen parametrelere göre yapılmaktadır (Uysal 2006).

Veri kümesine R programında yer alan kütüphane ve fonksiyonların kullanımı ile UA Biplot uygulanmış ve şekil 8.7'de verilen grafiksel yaklaşım elde edilmiştir.

Şekil 8.7, bölgeleri optimal bir şekilde ayırmayı ve gözlem noktaları arasındaki uzaklıkları korumayı amaçlamaktadır. Şekil 8.7'de her bir eksen değişkenlerin orijinal birimlerinde derecelendirilmiştir.

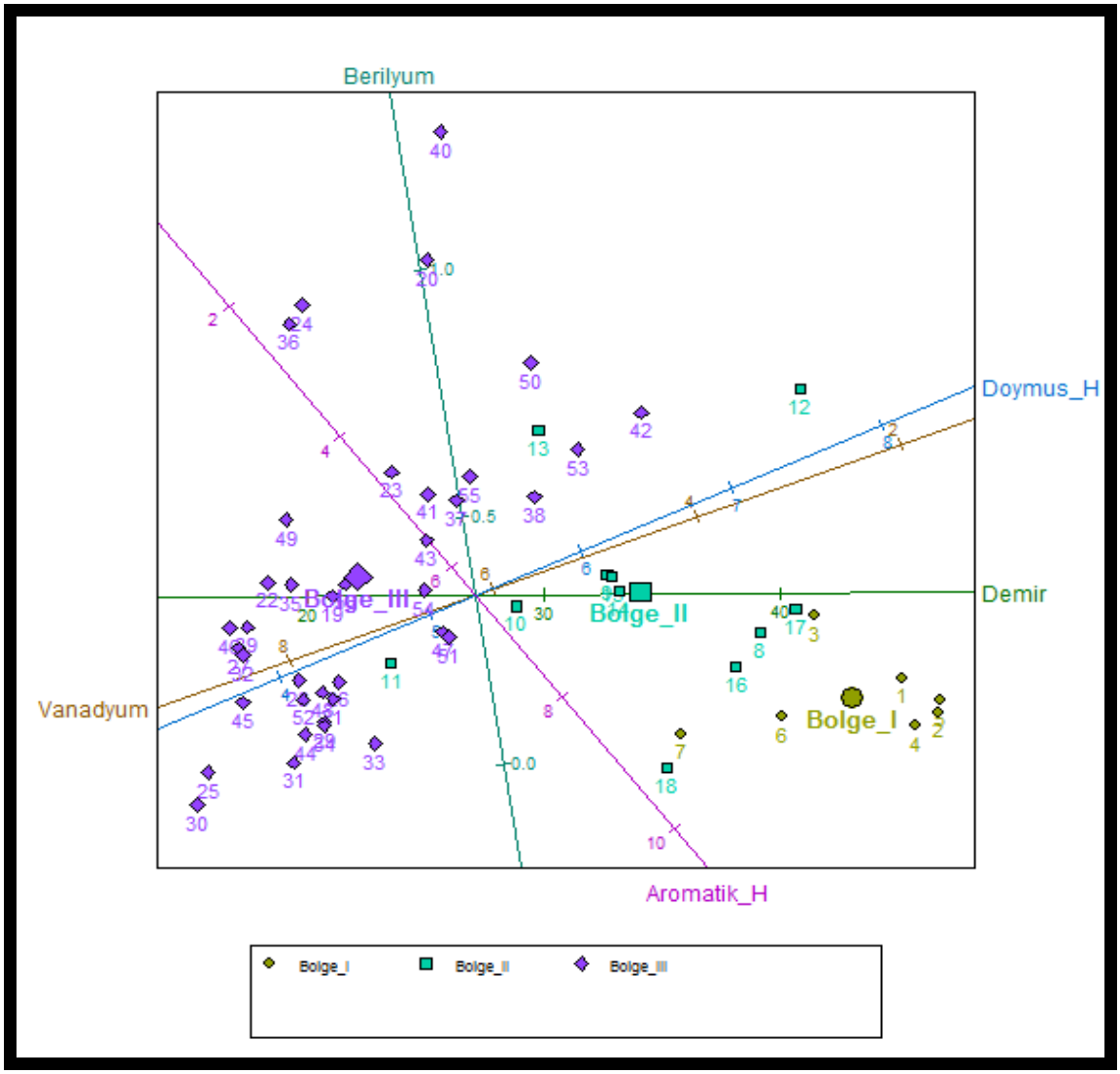
Çizelge 8.7 incelendiğinde, Berilyum ve Doymuş Hidrokarbonlar değişkenlerinin UA Biplot grafiğinde iyi temsil edildiği, Aromatik Hidrokarbonlar değişkenin ise çok iyi temsil edilmediği görülmektedir. Bu durum, Aromatik Hidrokarbonlar değişkeni üzerinden yapılacak kestirimlerin çok fazla tutarlı olmayacağı yorumunu beraberinde getirmektedir. Düşük eksen kestirimlerine sahip değişkenlerden elde edilecek kestirimlerin daha yorumlanabilir olması için üç boyutlu bir grafiksel gösterim düşünülebilir.

Şekil 8.7'de gösterilen gözlem noktaları tüm bölge içi varyasyonun iki boyutta bir yaklaşımıdır. İki boyutlu bu grafiksel yaklaşım ile toplam varyansın %65'i açıklanmaktadır.

Şekil 8.7, sadece bölge ortalamalarının yer aldığı UA biplot grafiğini verir. Bu grafiğin beş değişken için üç bölge ortalamasını ayırdığı görülmektedir. Grafiksel yaklaşımda, bölge ortalama noktalarından eksenlere dik izdüşümler çizilerek elde edilecek uzaklıklara göre eksenler üzerinden bölge ortalamaları yorumlanır.

Buna göre şekil 8.8'de verilen grafiksel yaklaşım incelendiğinde Bölge II ve Bölge III ortalamalarının Berilyum değişkeni açısından fazla bir farklılık göstermediği görülmektedir.

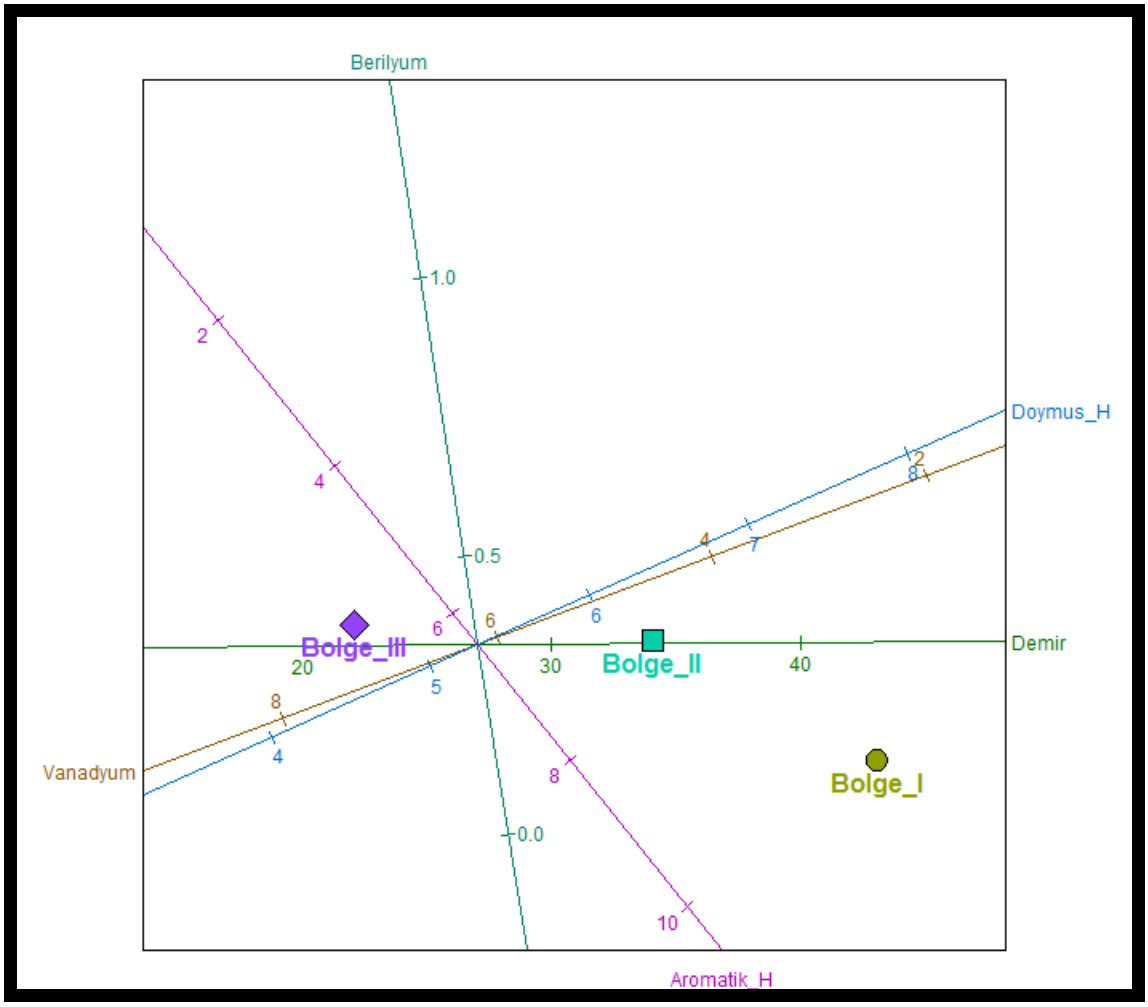
Aromatik_H, Vanadyum, Doymuş_H, Demir değişkenleri Bölge I ve Bölge II'ye göre Bölge III'de farklı bir yapı sergilemektedir. Demir, Aromatik_H ve Doymuş_H değişkenleri bakımından bölge ortalamaları arasındaki farklılıklar incelendiğinde Bölge I'de bu değişkenlerin etkilerinin fazla olduğu şekil 8.8'den yorumlanabilmektedir. Vanadyum değişkeni Bölge III ortalaması üzerinde etkiliyken, Doymuş_H değişkeni Bölge I ve Bölge III ortalamaları üzerindeki etkilidir. Bu üç bölge ortalaması düşünüldüğünde Demir oranı bakımından en zengin Bölge I iken, en fakir Bölge III olarak görülmektedir.



Şekil 8.7 Ham petrol veri kümesi için UA Biplot grafiği

Çizelge 8.7 UA Biplot grafiğinde eksenlerin kestirimleri

Vanadyum	Demir	Berilyum	Doymuş Hidrokarbonlar	Aromatik Hidrokarbonlar
0.60	0.59	0.83	0.79	0.40



Şekil 8.8 Bölge ortalamaları için UA Biplot grafiği

8.5 Uygulama 5- Bağlantı Biplot

Bu uygulamada, veri kümesi elde edilebilir olan ülkelerden derlenen 29 Avrupa ve Merkez Asya (tüm gelir grupları için) ülkeleri için kişi başına düşen gayri safi yurt içi hasıla (GSYİH), doğurganlık hızı (DH), tüketici fiyatları enflasyonu (TÜFE), kentsel nüfus (KN), ölüm oranı (ÖLO), toplam işsizlik oranı (TİO), hanehalkı nihai tüketim harcamaları (HHNTH) temel göstergeleri 2008 yılı verileri alınmıştır. Açıklanan değişken olarak GSYİH, açıklayıcı değişkenler olarak da yukarıda tanımlanan temel göstergeler alınmıştır. Çalışmada kullanılan veri kümesi Dünya Bankası (WorldBank) kalkınma göstergeleri veri tabanından derlenmiştir. Veri kümesi EK 5'de verilmiştir.

Çizelge 8.8 Veri kümesinin varyans büyüme faktörleri ve parametre tahminleri

Değişken	SD **	Parametre Tahmini	Standart Hata	t-değeri	p-değeri	Varyans Büyüme Faktörü
LOG(DH)	1	-1.89016	2.96402	-0.64	0.5302	223.70943
TÜFE	1	0.04473	0.00471	9.50	0.0001*	2.44856
KN	1	0.00992	0.01770	0.56	0.5809	193.28279
LOG(ÖLO)	1	0.16242	0.09557	1.70	0.1033	2.58381
TİO	1	-0.00143	0.00612	-0.23	0.8181	4.97038
HHNTH	1	-0.00453	0.00443	-1.02	0.3181	1.75220

** $p<0.05$, **SD, serbestlik derecesi*

Veri kümesi ilk olarak R programının yardımıyla analiz edilmiş ve GSYİH, DH, ÖLO değişkenlerinin logaritmik dönüşümleri ile çoklu regresyon modeli için varsayımlar sağlanmıştır. Analiz sonucunda modelin $R^2 = 0.927$ 'lik bir oranla oldukça iyi olduğu görülmüştür. Çoklu regresyon modeli çizelge 8.8'den,

$$\begin{aligned}\mathrm{LOG}(\mathrm{GSYİH}) = &-1.89016(\mathrm{LOG}(\mathrm{DH})) + 0.04473(\mathrm{TÜFE}) + 0.00992(\mathrm{KN}) \\ &+ 0.16242(\mathrm{LOG}(\mathrm{ÖLO})) - 0.00143(\mathrm{TİO}) - 00453(\mathrm{HHNTH})\end{aligned}$$

şeklinde elde edilir.

Çizelge 8.8'deki t-istatistiğine göre, TÜFE değişkeninin önemli olduğu ve LOG(DH), KN açıklayıcı değişkenleri için varyans büyüme faktörlerinin 10 değerinden yüksek olduğu tespit edilmiştir. Bu yüzden bu iki değişkenin çoklubağlantı problemine sebep olabileceği düşünülebilir. Ayrıca çizelge 8.8'deki katsayıların işaretleri incelendiğinde, LOG(GSYİH) ile LOG(DH), TİO, HHNTH arasında negatif bir ilişkinin, LOG(GSYİH) ile TÜFE, KN ve LOG(ÖLO) arasında ise pozitif bir ilişkinin varlığından söz edilebilir.

Çizelge 8.9, model için varyans ayrışım oranlarını ve koşul indekslerini vermektedir. Ayrıca, çizelge 8.9'un modifiye edilmiş versiyonu çizelge 8.10'da verilmiştir. Çizelge 8.10, en küçük iki boyuttaki (Boyut 5 ve Boyut 6) çoklubağlantıya açıklayıcı

değişkenlerin katkılarını göstermektedir. Çalışmada koşul indeksleri 10'dan büyük olan son iki satır ele alınmıştır. Çizelge 8.10 incelendiğinde, Boyut 6'da LOG(DH) ve KN açıklayıcı değişkenleri arasında, Boyut 5'de ise LOG(ÖLO) ve TÜFE açıklayıcı değişkenleri arasında yüksek yakın bağımlılık görülmektedir.

Şekil 8.9, tüm koşul indeksleri ve varyans ayrışım oranlarının bir tableplot grafiğini sunmaktadır. Şekil 8.9 incelendiğinde, en üstteki iki satırın kırmızı renk ile boyandığı görülmektedir. Bu iki satır en küçük iki boyutu ifade eder ve bu boyutların varyans ayrışım oranları da açıklayıcı değişkenlerin bu boyutlara katkılarını göstermektedir. 1 ile 4 arasındaki diğer boyutlar yeşil olarak boyanmıştır. Bu durum ilgili boyutlarda çoklubağlantı probleminin yokluğunun göstergesidir.

Çizelge 8.9 Koşul indeksleri ve varyans ayrışım oranları

		Varyans Ayrışım Oranları					
Boyutlar	Koşul İndeksi	LOG(DH)	TÜFE	KN	LOG(ÖLO)	TİO	HHNTH
Boyut 1	1.000	0.000	0.007	0.000	0.002	0.010	0.007
Boyut 2	2.228	0.001	0.005	0.001	0.000	0.000	0.483
Boyut 3	3.937	0.001	0.010	0.004	0.001	0.719	0.062
Boyut 4	4.785	0.003	0.566	0.005	0.002	0.021	0.375
Boyut 5	10.307	0.003	0.400	0.024	0.849	0.121	0.070
Boyut 6	32.118	0.992	0.012	0.965	0.146	0.130	0.003

Çizelge 8.10 Modifiye edilmiş koşul indeksleri ve varyans ayrışım oranları

		Varyans Ayrışım Oranları (x100)					
Boyutlar	Koşul İndeksi	LOG(DH)	TÜFE	KN	LOG(ÖLO)	TİO	HHNTH
Boyut 6	32.12	**99**	1	**97**	15	13	0
Boyut 5	10.31	0	**40**	2	**85**	12	7

Not: Modifiye edilmiş VAO değerlerinden 40'dan büyük olanlar koyu olarak gösterilmiştir.

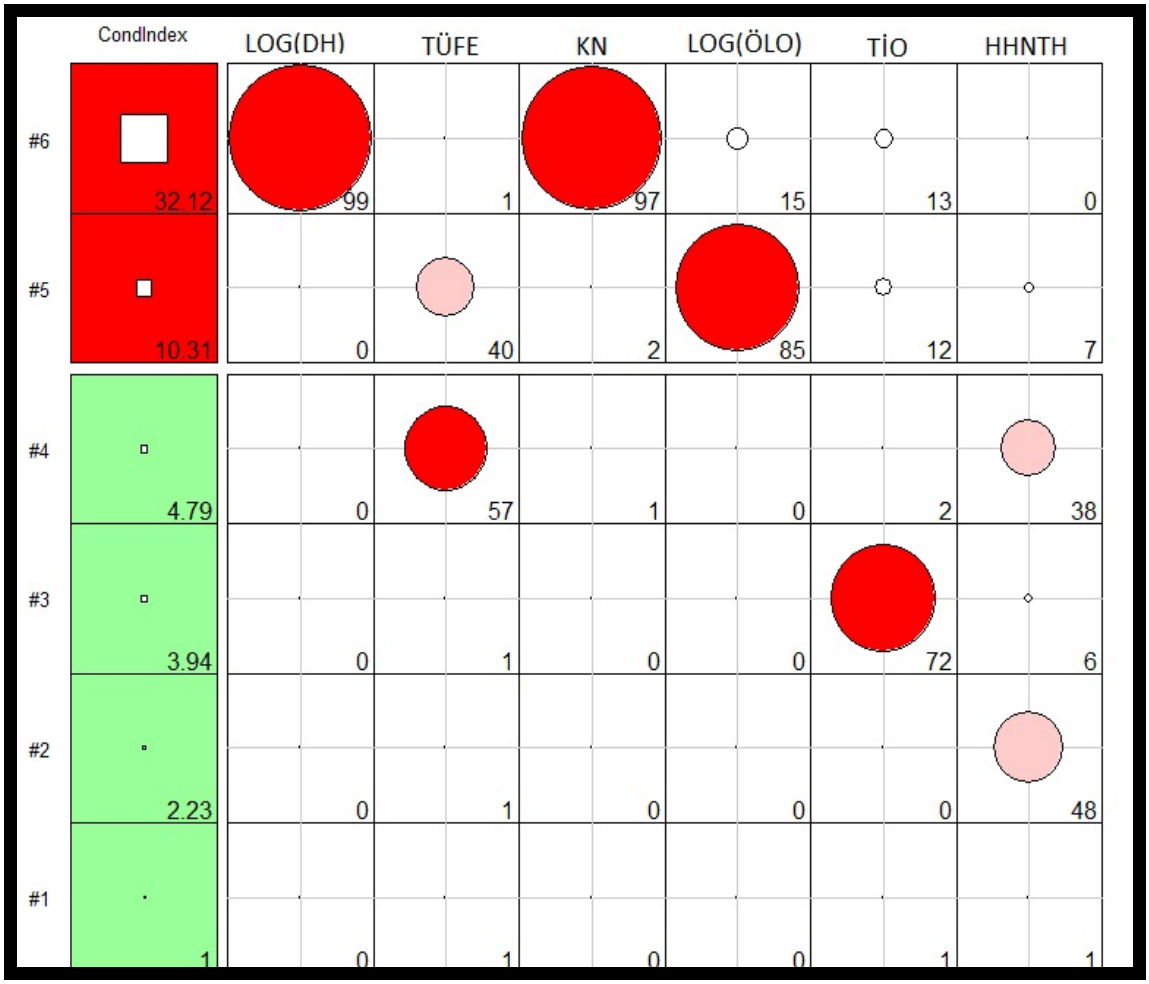
Şekil 8.10'da verilen klasik biplot, açıklayıcı değişkenler ile gözlemler arasındaki ilişkilerin incelenmesi için kullanılır. Ayrıca, indirgenmiş boyutlu (en büyük iki boyut) bir uzayda açıklayıcı değişkenlerin varyasyonunu verir. Bu iki boyut ile 6 boyutlu

uzaydaki toplam varyansın %78'lik bölümü açıklanabilmektedir. Toplam varyansın tümü 6 boyutta açıklanabilir. Toplam varyansın birinci boyutta %57.43'si, ikinci boyutta %20.48'i, üçüncü boyutta %9.89, dördüncü boyutta %7.83'ü, beşinci boyutta %4.33, altıncı boyutta %0.04'ü açıklanmaktadır.

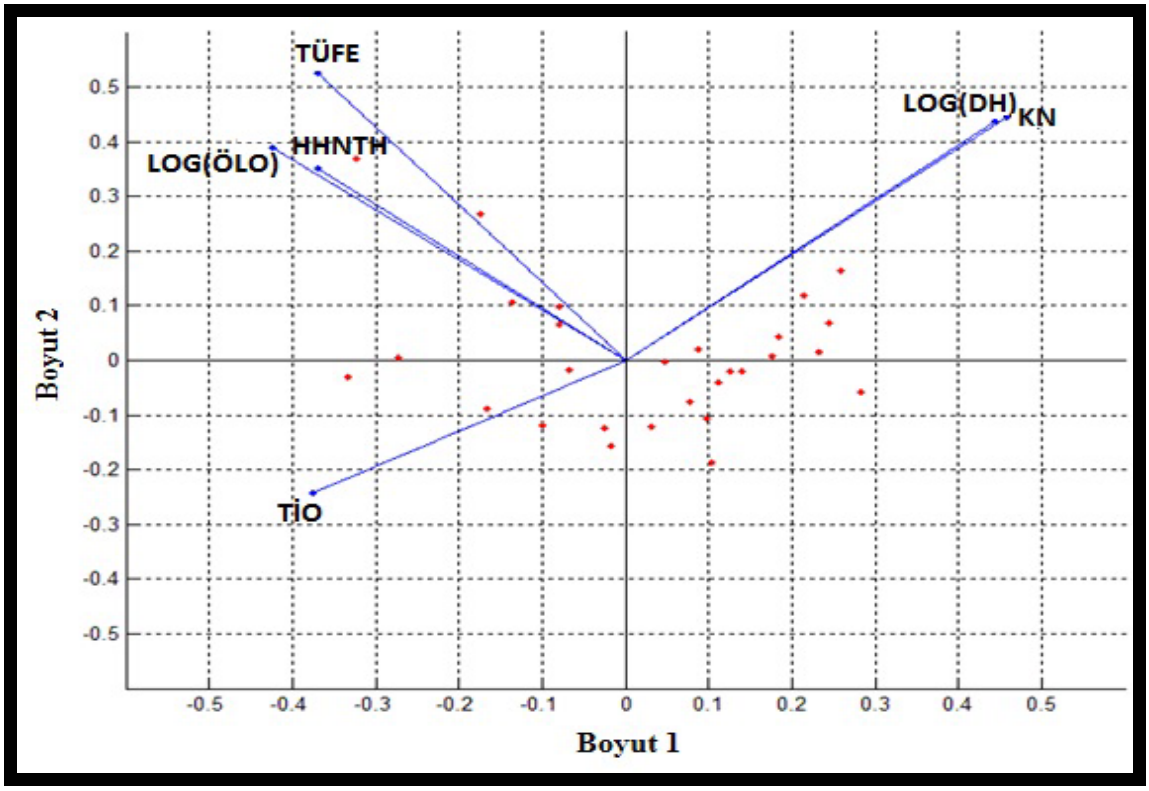
Şekil 8.10 incelendiğinde, yüksek ilişkili açıklayıcı değişkenlerin oluşturduğu iki küme görülmektedir. Birinci küme LOG(DH) ve KN açıklayıcı değişkenlerinden oluşurken, ikinci küme ise LOG(ÖLO), HHNTH ve TÜFE açıklayıcı değişkenlerinden oluşmaktadır. Ayrıca, LOG(DH) ve TİO, KN ve TİO açıklayıcı değişkenleri arasında negatif yönde yüksek bir korelasyonun varlığı söz konusudur.

Yukarıda yapılan yorumlar çoklubağlantının teşhisi söz konusu olduğunda kullanışlı değildir. Çünkü bağlantı problemi son boyutlarda saklıdır.

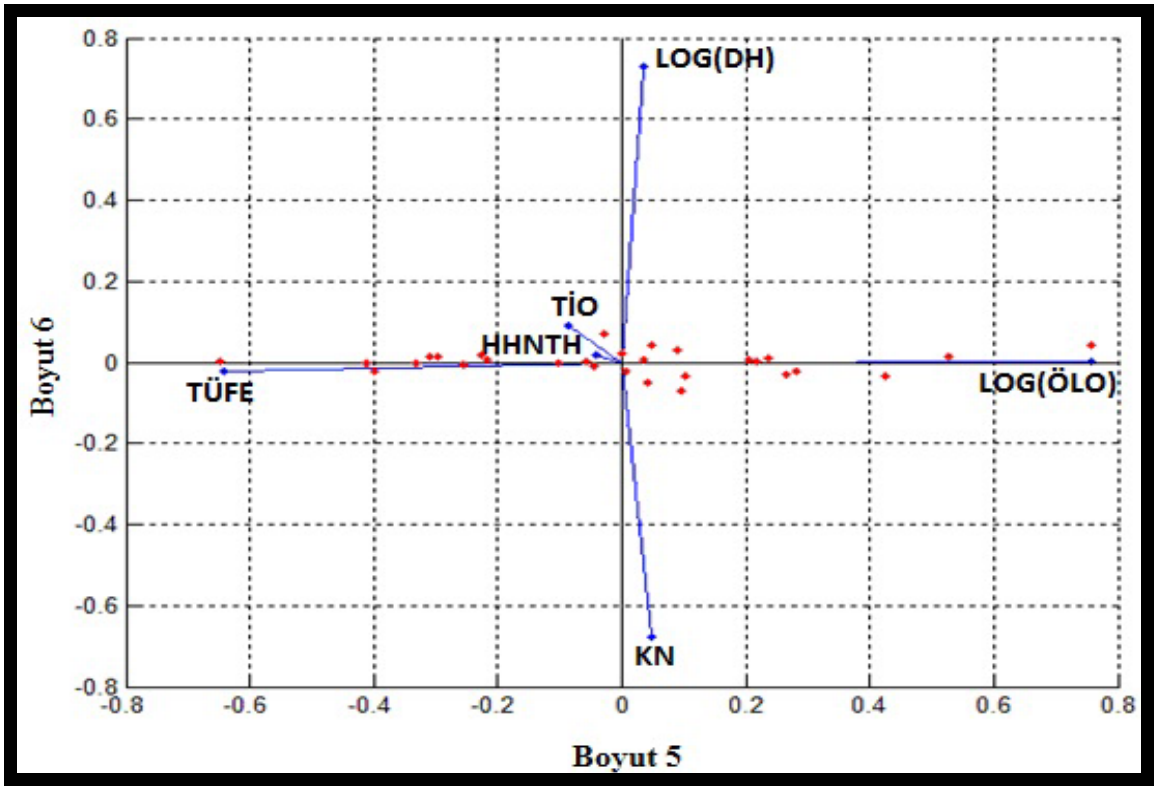
Şekil 8.11'da en büyük iki boyut için verilen Bağlantı Biplot grafiği toplam varyansın %4.37'lik bölümünü açıklamaktadır. Bu biplot yaklaşımı Tableplot ve Çizelge 8.10'dan elde edilen yorumları desteklemektedir. Şekil 8.11 incelendiğinde, Boyut 6 üzerinde LOG(DH)&KN ve Boyut 5 üzerinde TÜFE&LOG(ÖLO) açıklayıcı değişkenlerinin etkili olduğu görülmektedir. Ayrıca, TİO&HHNTH açıklayıcı değişkenlerinin Boyut 5 üzerinde etkisinin oldukça küçük olduğu elde edilen diğer bir yorumdur.



Şekil 8.9 Veri kümesinin Tableplot grafiği



Şekil 8.10 Klasik Biplot



Şekil 8.11 Bağlantı Biplot

8.6 Uygulama 6- Alan Biplot

Küresel ekonomik ve finansal kriz OECD (Organisation for Economic Co-operation and Development) ülkeleri arasında temel makroekonomik değişkenlerde bir farklılığa neden olmuştur. Yüksek gelirli OECD ülkeleri, yüksek gelirli OECD'de yer almayan ülkelere göre krizden daha çok etkilenmiştir. Bu temel görüşten, yüksek gelirli OECD ülkelerinin temel makroekonomik göstergelerinin kriz öncesi dönemde incelenmesinin önemli olduğu söylenebilir.

Bu uygulama, yüksek gelirli OECD üyesi 26 ülke için kişi başına gayri safi yurt içi hasıla (GSYİH), kişi başına hane halkı nihai tüketim harcamaları (HHNTH), enflasyon oranı (EO), dış ticaret (DT) ve işsizlik oranı (İO) göstergelerinin 1990-2007 arasındaki ortalamaları incelenmektedir. Çalışmada kullanılacak veri kümesi Dünya Bankası (WorldBank) kalkınma göstergeleri veri tabanından derlenmiştir. Ülke seçimi, yüksek gelirli OECD ülkelerinden veri seti elde edilebilir olanlara göre yapılmıştır. Ekonomik kriz temel değişkenleri bozduğu için 2008 yılı verileri veri kümesine dahil edilmemiştir. Sadece 2008 yılı için bulgular ayrıca incelenmiş fakat anlamlı bulunmamıştır. 2008 yılı verileri yapısal kırılma olarak değerlendirilmiştir.

Veri kümesinin analizi için, Gower vd. (2010) tarafından önerilen Alan Biplot yöntemi kullanılmıştır. Alan Biplot, Satır (ülke) noktaları 90° döndürüldükten sonra, döndürülmüş bir satır (ülke) noktası, bir sütun (değişken) noktası ve orijinin bir üçgen ile gerildiği alan, veri değerlerine yaklaşacaktır temeline dayanmaktadır.

Çalışmada, beş farklı ekonomik gösterge değişkenlerinin 1990-2007 yılları arasındaki ortalamaları verisine Alan Biplot yöntemi uygulanarak, ülkeler ve ekonomik göstergeler açısından meydana gelen değişikliklerin görüldüğü bir portre ortaya koymak amaçlanmaktadır.

Verilerin analizinde ve grafiklerin elde edilmesinde MATLAB ve R programlarından yararlanılmıştır.

Analiz için ilk olarak *z*-skorları elde edilmiştir. Daha sonra elde edilen z-skorlarının oluşturduğu veri kümesi Alan Biplot yöntemi ile analiz edilmiş ve şekil 8.12'de sunulan grafiksel yaklaşım elde edilmiştir. Ayrıca, toplam varyansın %67'lik bir bölümünün ilk iki boyutta açıklandığı görülmektedir. Bu oran analiz sonucu elde edilen Alan Biplot grafiğinin açıklama gücünün yeterli olduğunu göstermektedir.

Çizelge 8.11'de, 1990-2007 yılları arasında her bir ülkenin beş ekonomik göstergesinin ortalamalarının oluşturduğu orijinal veri kümesinin standartlaştırılmış biçimi olan z-skorları verilmiştir**.** Şekil 8.12 ise, çizelge 8.12'de verilen veri kümesine Alan Biplot uygulanarak elde edilmiştir. Bu grafiksel yaklaşımda, ülkeler küçük noktalar ile gösterilirken, kalkınma göstergeleri ise büyük noktalar ile temsil edilmektedir. Şekil 8.12'de, iki boyutlu Alan Biplot grafiğinde yer alan değişkenler arasındaki yaklaşık korelasyonlar ise çizelge 8.12'de verilmiştir. Çizelge 8.12 incelendiğinde, İO&EO(0.73) ve GSYİH&HHNTH (0.95) kalkınma gösterge değişken çiftlerinin kendi aralarında iyi derecede bir ilişkiye sahip oldukları; İO&DT (-0.87) değişken çiftinin ise negatif yönde güçlü bir ilişkiye sahip olduğu; GSYİH&İO (-0.25) değişkenleri arasında ise negatif yönde zayıf bir ilişki olduğu görülmektedir.

Şekil 8.12'in sol kısmında yer alan Kanada ve ABD ülkelerinin ayrı ayrı orijin ve GSYİH değişkeni ile oluşturdukları alanlar incelendiğinde bu alanların sıfıra yakın oldukları görülmektedir. Alan Biplot'da orijin, döndürülmüş satır noktası ve sütun noktası ile oluşturulacak bir üçgenin alanının iki katı ilgili gözlemin çizelge 8.11'de verilen z-skorunu yaklaşık olarak verir. Bu durumda, GSYİH değişkeni bakımından Kanada ve ABD'nin benzerlik gösterdikleri söylenebilir. GSYİH'nın Lüksemburg ile oluşturduğu alan Norveç ile oluşturduğu alandan daha büyüktür. Bu durum GSYİH ortalamasının Lüksemburg'da Norveç'e göre daha büyük olduğunu göstermektedir. Alan Biplot aynı zamanda ilgili değişkenler bakımından benzer olan ülkelerin bir kümesini ve diğer ülkelerden aykırı bir konumda olan ülkeleri görmemizi sağlar.

Grafiğin sağ tarafında İrlanda ve Kore ülkeleri yer almaktadır. Bu ülkelerin 1990–2007 yılları arası beş kalkınma göstergesi ortalamaları bakımından yakın bir duruş sergiledikleri görülmüştür. Ayrıca G7 ülkelerinden İtalya, Fransa, Almanya ve Japonya

ise GSYİH değişkeni bakımından benzer bir konumda yer almaktadır. EO bakımından G7 ülkelerinden İtalya, Fransa ve Kanada'nın yine yakın bir ilişki içinde olduğu şekil 8.12'den elde edilen diğer bir yorumdur. OECD ülkeleri arasında DT açısından birinci ülke Lüksemburg görünürken, bu ülkeyi sırasıyla İrlanda ve Kore takip etmektedir. DT bakımından en zayıf ülkelerin sırasıyla İspanya, Macaristan ve Yunanistan olduğu grafiksel yaklaşımdan açıkça anlaşılmaktadır.

İO ve EO bakımından en problemli ülkelerin ise İspanya, Macaristan, Yunanistan ve İsrail olduğu görülmektedir. Lüksemburg, İO ve EO'nun en düşük olduğu, DT ve GSYİH'nın en yüksek olduğu bir ülke olarak tanımlanabileceği yine şekil 8.12'den açıkça yorumlanmaktadır.

Şekil 8.13-8.14 de sırasıyla enflasyon hedeflemesi yapan ülkeler ile PIGS ülkeleri için Alan Biplot Grafikleri verilmiştir. Literatürde PIGS ülkeleri olarak Portekiz, İtalya, Yunanistan ve İspanya ele alınmaktadır. PIGS, uluslararası ekonomi basını, uluslararası tahvil analistleri ve akademisyenler tarafından sık kullanılan bir kısaltmadır. PIGS ülkeleri krize maruz kalmış borçlu ülkeler statüsünde değerlendirilmektedir.

Şekil 8.13'de verilen Alan Biplot grafikleri, enflasyon hedefleyen yüksek gelirli OECD ülkelerinden Avusturalya, Kanada, İsrail ve Norveç için oluşturulmuştur. Bu enflasyon hedefleyen dört ülkeyi incelediğimizde, İsrail'in diğer üç ülkeye göre İO, GSYİH ve HHNTH açısından oldukça farklı yapı sergilediği görülmektedir. Göreli olarak bu dört ülke DT açısından değerlendirildiğinde Norveç ön plana çıkmaktadır. Bu ülke'de DT ortalamanın üstünde yer almaktadır.

Çizelge 8.11 1990-2007 yılları temel makroekonomik göstergeleri ortalamaları veri kümesinin z-skorları

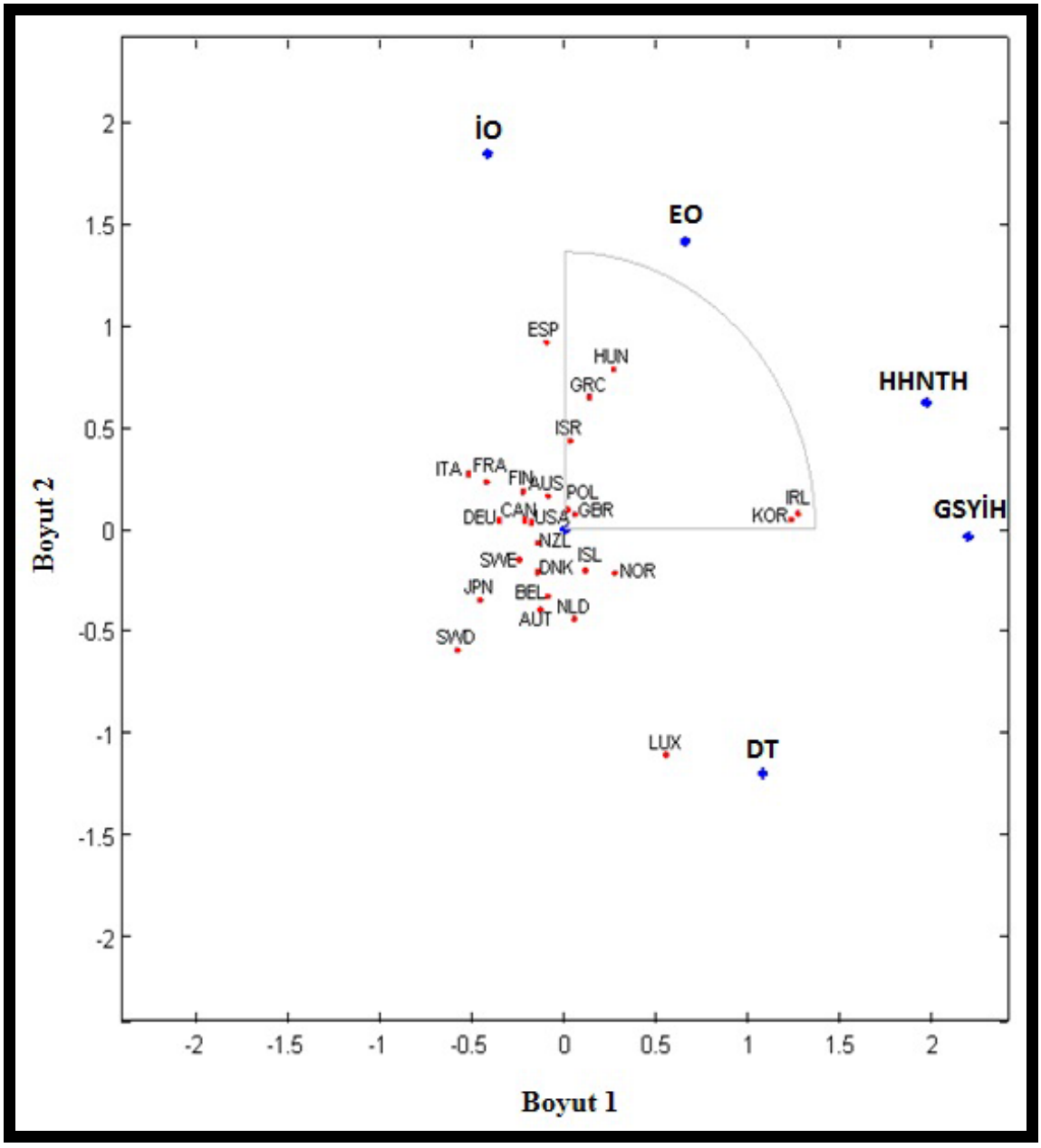
Ülke Kodu	Ülke Adı	GSYİH	HHNTH	EO	DT	İO
AUS	Avustralya	-0.1460	0.2752	-0.2448	-0.888	0.1435
AUT	Avusturya	-0.1757	-0.6378	-0.4435	0.1614	-0.8630
BEL	Belçika	-0.4074	-0.649	-0.4856	1.5593	0.3329
CAN	Kanada	-0.4646	-0.2793	-0.4410	-0.1675	0.4822
DNK	Danimarka	-0.3594	-0.3046	-0.4997	0.0488	-0.2947
FIN	Finlandiya	-0.1283	-0.4946	-0.5506	-0.2288	1.1499
FRA	Fransa	-0.7557	-0.6012	-0.5819	-0.6414	1.1155
DEU	Almanya	-0.4746	-0.8259	-0.5329	-0.3818	0.5935
GRC	Yunanistan	0.1883	0.5015	1.4514	-0.5952	0.8266
HUN	Macaristan	-0.1571	-0.1335	4.0984	0.7255	0.3764
ISL	İzlanda	-0.0741	0.3538	0.3345	-0.1469	-1.2562
IRL	İrlanda	3.1787	2.3953	-0.0989	1.4972	0.5950
ISR	İsrail	-0.0592	0.0017	1.1997	-0.0709	0.6907
ITA	İtalya	-0.9931	-1.0316	-0.063	-0.6823	0.9758
JPN	Japonya	-0.7823	-0.7433	-1.0454	-1.2472	-1.1060
KOR	Kore	2.8241	3.0205	0.3513	-0.2535	-1.2208
LUX	Lüksemburg	1.2321	-0.6214	-0.4107	3.5588	-1.3853
NLD	Hollanda	0.0241	-0.2824	-0.4107	0.9325	-0.6755
NZL	Yeni Zelanda	-0.3792	-0.0415	-0.4003	-0.4119	-0.2029
NOR	Norveç	0.2806	1.1153	-0.4631	-0.1322	-0.9204
POL	Portekiz	-0.2799	0.5339	0.4021	-0.2819	-0.3846
ESP	İspanya	0.0980	0.3032	0.0976	-0.6165	2.9505
SWE	İsveç	-0.3185	-0.822	-0.3586	-0.0360	-0.0536
SWD	İsviçre	-1.3152	-1.66	-0.6262	0.0155	-1.2533
GBR	İngiltere	-0.1017	0.4699	-0.067	-0.5246	-0.0785
USA	Amerika	-0.4539	0.1577	-0.2112	-1.1925	-0.5377

Çizelge 8.12 İki boyutlu Alan Biplot'daki değişkenler arasındaki korelasyonlar

	GSYİH	HHNTH	EO	DT	İO
GSYİH	1.0000	0.9530	0.3998	0.6798	-0.2549
HHNTH	0.9530	1.0000	0.6417	0.4300	0.0495
EO	0.3998	0.6417	1.0000	-0.3510	0.7333
DT	0.6798	0.4300	-0.3510	1.0000	-0.8691
İO	-0.2549	0.0495	0.7333	-0.8691	1.0000

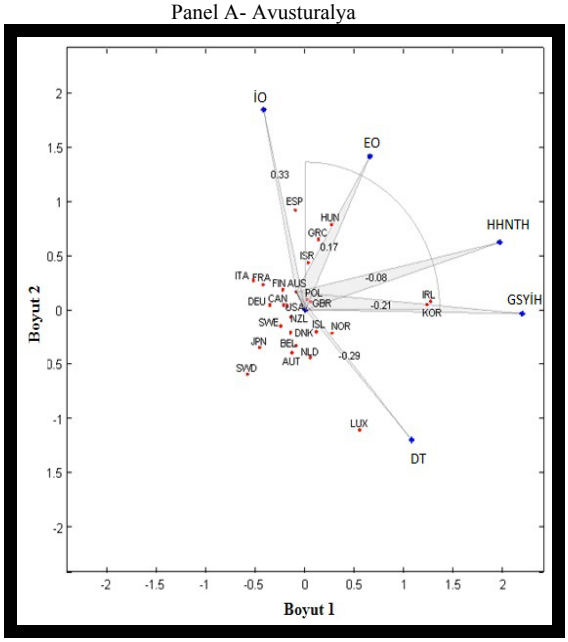
Şekil 8.14'de verilen Alan Biplot grafikleri, PIGS ülkelerin incelenmesi için oluşturulmuştur. PIGS ülkelerinden işsizlik ortalaması en yüksek olan ülke İspanya olarak görünmektedir. Enflasyon açısında PIGS ülkeleri değerlendirildiğinde İspanya ile Yunanistan, İtalya ile Portekiz yakın bir ilişki içerisinde seyretmişti. Ayrıca PIGS ülkelerinin dış ticaret açısından ortalamanın altında olduğu, GSYİH açısından ise Yunanistan hariç diğer üç ülkenin yine ortalamanın altında olduğu görülmektedir.

Diğer ülkeler için de benzer Alan Biplot grafikleri elde edilebilir. Yorumlar karşılaştırılabilir. Bu alan yaklaşımlarının şekil 8.12'de verilen Alan Biplot grafiği üzerinden zihnen yapılmasının bir kaç pratikten sonra oldukça kolay olduğu görülecektir. Alan Biplot yaklaşımının, değişkenler ve gözlemler arasındaki ilişkilerin görüldüğü garafiksel bir resim olduğu ve bu resmin netliğinin ise toplam varyans açıklanma oranıyla doğrudan ilişkili olduğu unutulmamalıdır.

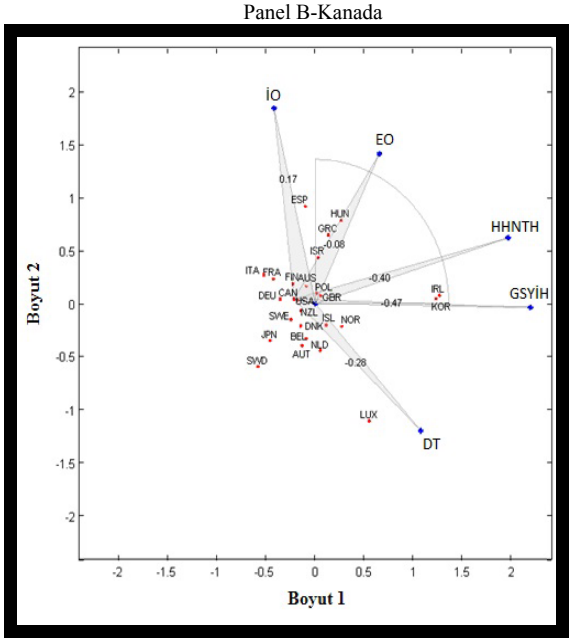


Şekil 8.12 Veri kümesi için Alan Biplot grafiği

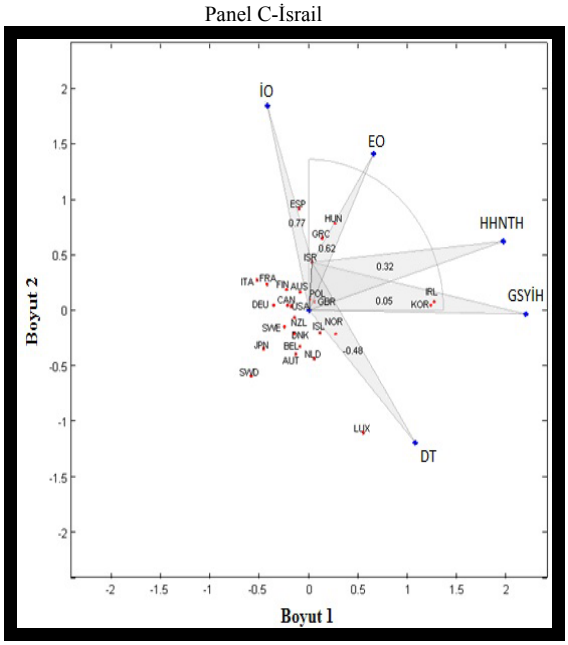
Panel A- Avusturalya



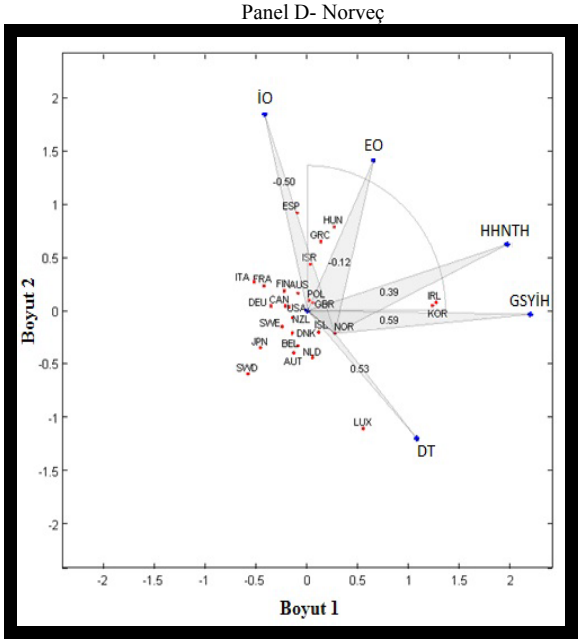
Panel B-Kanada



Panel C-İsrail

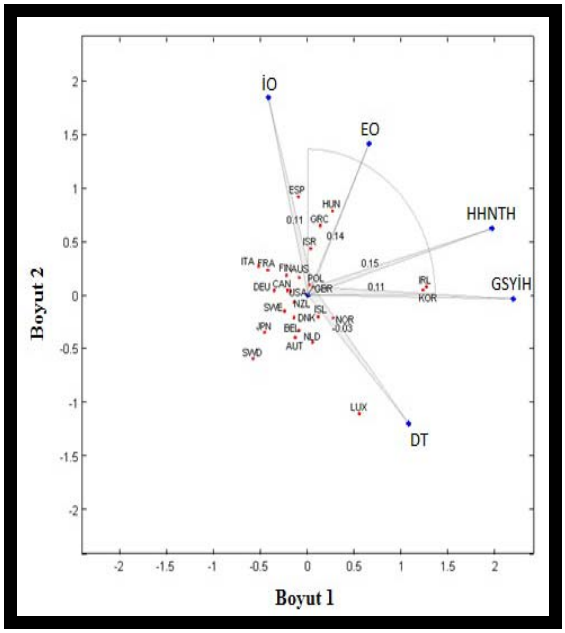


Panel D- Norveç

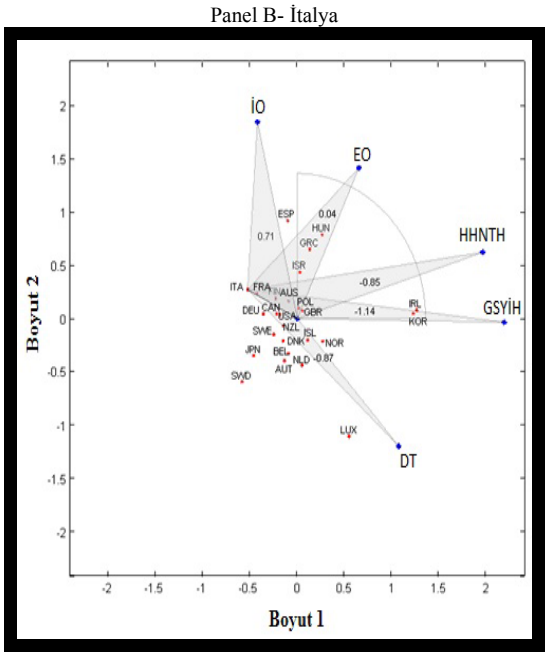


Şekil 8.13 Enflasyon hedefleyen ülkeler için Alan Biplot grafikleri

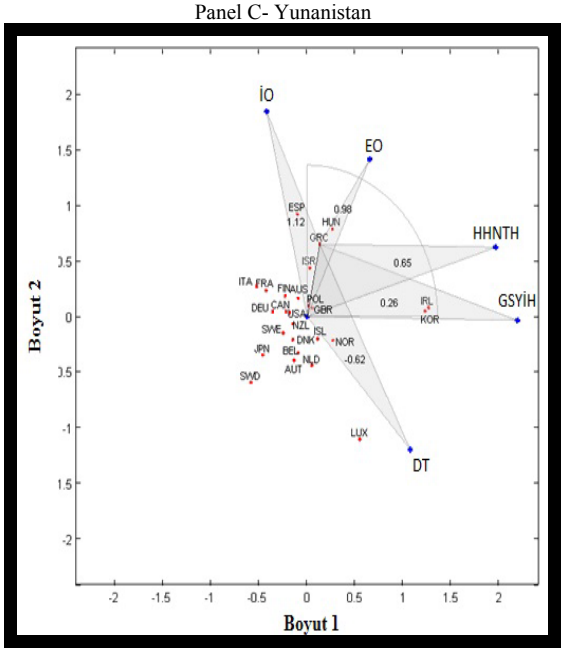
Panel A-Portekiz



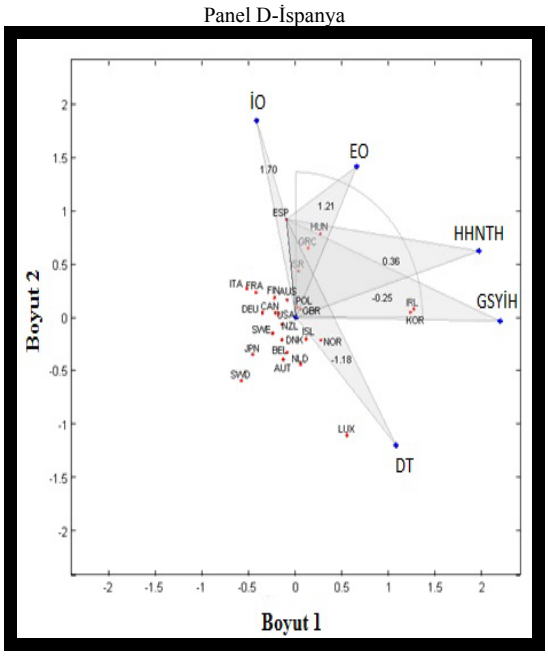
Panel B- İtalya



Panel C- Yunanistan



Panel D-İspanya



Şekil 8.14 PIGS ülkeleri için Alan Biplot grafikleri

9. TARTIŞMA VE SONUÇ

Bu çalışmada, çok değişkenli istatistiksel yöntemlerde biplot tekniğinin kullanımına yönelik konular ele alınmış ve kuramsal olarak açıklanmıştır. Biplot, veri kümesinde değişkenler arasındaki ilişkilerin, değişkenlere göre gözlemler arasındaki benzerliklerin veya farklılıkların görsel olarak yorumlanabilmesinde ve gözlemlerin sınıflandırılmasında oldukça başarılı bir tekniktir. Bu teknik görsel bir grafik yardımıyla veriyi tüm yönleri ile incelemeye imkan sağlamaktadır. Veri kümesine ardışık olarak uygulanacak bir çok klasik istatistiksel analiz ile elde edilecek yorumların tek bir biplot üzerinden elde edilebilmesi tekniğin gücünü ve önemini artırmaktadır.

Çalışmanın uygulama bölümünde ekonomi, gıda ve enerji alanlarından toplam altı gerçek veri kümesine farklı biplot teknikleri uygulanmış ve ilgili alanlarda veri açıklayıcı bir yöntem olarak biplot tekniğinin kullanım gereğinin önemi vurgulanmıştır.

Uygulama 1, endüstriyel üretim içinde bir desen oluşturan ve karşılıklı etkileşimleri tartışmalara konu olmuş değişkenler arasındaki ilişkileri Gabriel'in Biplot tekniğiyle incelenmektedir. Sermaye malları üretimindeki değişimin ortalaması ve oynaklığı ele aldığımız yıllar arasında en yüksek olan kategoridir. Ortalama üretimin en düşük olduğu alan ara malı üretimidir. Ara malı üretimindeki değişim aynı zamanda en az oynak ana kategoridir. Dayanıklı mal üretimi, dayanıksız mal üretimine göre daha oynak ve ortalaması daha yüksektir. Bu değerlerle beraber düşünüldüğünde, Türkiye'de 2005 sonrasında ara malı üretimi ile dayanıklı üretimindeki artışın en yüksek korelasyon göstermesi beklenen ve tutarlı bir sonuçtur. Kredi koşullarının iyileşmesi yatırımı artırmış, toplam büyümede bir artışla beraber otomotiv, beyaz eşya gibi ürünlere de talep artmıştır. İmalat sanayi içinde dayanıklı üretim malları önemli bir paya sahiptir ve özellikle Avrupa'ya olan ihracatımızda ciddi bir ivme kat etmiştir. Dolayısıyla dayanıklı imalat sanayi üretimi ile toplam imalat sanayi arasındaki yüksek korelasyon beklenen bir sonuçtur. Ara malı üretimi ile sermaye malı üretimi arasındaki korelasyonun düşük olması ilginç bir sonuçtur. Türkiye'de makine teçhizat ithalatı yıllar itibariyle artış göstermiş, sanayi üretiminde ara malı ithalatı artmıştır. Sermaye malı üretimi montaja ve ithal girdiye daha bağımlı olduğu için korelasyonun düşük çıkması makuldür. Ayrıca

Türkiye'de üretilen ara malı, sermaye üretimi için gerekli ve yeterli koşulu tam olarak sağlamamaktadır. Türkiye'de sermaye malı üretimindeki artış, dayanıklı ve dayanıksız mallar üretimi dışındaki sektörlerle daha fazla ilişkilidir.

Uygulama 2, insan doğasına etki eden başlıca meyve tiplerinin bazı organik ve inorganik kimyasal özelliklerinin Gower ve Hand (1996)'in biplot tekniği ile incelemektedir. Sonuçlar, protein ve enerji bileşenlerinin yüksek derecede ilişkili, protein-su ve karbonhidrat-su bileşenlerinin ise negatif yönde ilişkili, kuru üzüm ve kuru kayısının oldukça benzer bir yapı sergilediği, protein ve enerji bakımından en zengin meyvelerin kuru kayısı, kuru üzüm, avokado ve kuru incir, en çok yağ içeren meyvenin ise avokado olduğu şeklinde özetlenebilir.

Uygulama 3, Gower ve Hand (1996) ile Gardner ve Roux (2005) tarafından önerilen KDA Biplot tekniği kullanılarak, beş farklı besin grubu arasındaki ayrım ve örtüşme, gruplarda yer alan beş kolin değişkeni açısından incelemektedir. KDA Biplot ile tek bir grafik üzerinden gözlemler, grup ortalamaları ve değişkenlerin etkileri hakkında yorum yapılabilmektedir. KDA Biplot, farklı grupların kovaryans matrisleri arasındaki homojenliğin olmaması bazı şüpheler oluştursa bile, veri kümesindeki grupların ayrıntılı bir şekilde tanımlanması, grup yapısı ve çok değişkenli varyasyonun görsel gösteriminin kullanıcılar açısından faydası tartışılamaz (Gardner vd. 2005). Bu uygulamada kullanılan veri kümesinin analizi sonucunda, süt ve yumurta ürünleri ile sebze ve sebze ürünlerinde PCHO ve SM kolin değişkenleri bakımından önemli düzeyde bir farklılık görülmemiştir. Ayrıca tavuk ve hindi ürünleri ile meyveler ve meyve ürünleri arasında FREECHO içerme oranı açısından benzerlik gözlenmiştir. PCHO ve SM kolin değişkenleri bakımından diğer besin grupları arasındaki farklılıklar incelenmiş, tavuk ve hindi ürünlerinin oluşturduğu besin grubunda bu değişkenlerin oranının fazla olduğu sonucuna ulaşılmıştır. Meyve ve meyve ürünleri ile sebze ve sebze ürünlerinde PCHO-SM kolin değişkenleri açısından fark olmadığı görülmüştür. GPC kolin değişkeninin etkisi en fazla süt ve yumurta ürünlerinde görüldüğü analizden elde edilen diğer bir önemli sonuçtur. FREECHO değişkeninin ise, en fazla olduğu besin grubu baklagil ve baklagil ürünleri iken, en az olduğu besin grubunun süt ve yumurta ürünleri olduğu görülmektedir. Bir çok ardışık analiz yapılarak elde edilebilecek sonuçlar, sadece KDA

Biplot kullanılarak da elde edilebilir. Bu durum araştırmacıları işlem karmaşasından kurtarır ve veri kümesinin görsel, daha açıklayıcı olarak değerlendirilmesi sağlanabilir. Bu da KDA Biplot'un öneminin diğer bir göstergesi olarak kabul edilebilir.

Uygulama 4, üç kumtaşı bölgesinden alınan 56 ham petrol kimyasal özellikleri arasındaki ilişkileri incelemektedir. Bu uygulamadaki veri kümesine MANOVA varsayımlarını sağlamadığından UA Biplot uygulanmıştır. UA Biplot'da uygulama verisi için bölge ortalamalarının biplot eksenleri üzerlerine dik izdüşümleri her bir bölgenin özellikleri hakkında önemli bilgiler vermektedir. Uzaklık analizi, dağılımsal varsayımlara göre değişmemektedir. Üstelik küçük örneklemlerde ve grup varyans-kovaryans matrislerinin heterojenliğinden etkilenmez. Ayrıca bir hipotez testi gerektiği zaman permütasyon test prosedürleri kullanılabilir. KDA biplotların tüm özellikleri ve genişlemeleri, UA biplotlar için de mevcuttur. UA biplot, mevcut istatistiksel çıkarım prosedürleri ile birlikte ele alınması durumunda, bu yöntem veri analizi için daha da önemli bir araca dönüşür. Uzaklık analizi sonucu elde ettiğimiz grafiksel yaklaşımda, grup farklılıkları daha az belirgin olduğunda, gruplar arasındaki örtüşmenin derecesinin belirlenmesi ve ilgili gruplar arasındaki farklılıkların istatistiksel önemini saptamak için bootstrap veya permütasyon hipotez testi gibi parametrik olmayan çıkarım teknikleri uygulanır (Gower ve Krzanowski 1999).

Uygulama 5, 29 Avrupa ve Merkez Asya (tüm gelir grupları için) ülkelerinin 2008 yılı için bazı temel göstergelerinin oluşturduğu veri kümesinde çoklubağlantının görsel teşhisi için bağlantı Biplot ve Tableplot yöntemlerinin kullanışlılığını incelemektedir. Bağlantı biplot, en küçük özdeğerlere karşılık gelen son iki boyut olan Boyut (p-1) ve Boyut p'nin oluşturduğu bir uzayda, saklı olan ve çoklubağlantıya katkısı olan değişkenlerin belirlenmesi yanında, açıklayıcı değişkenler ve gözlemler arasındaki ilişkilerinde yorumlanabiliyor olması bakımından önemlidir. Elde edilecek bu Bağlantı Biplot grafiğinde, Boyut (p-1) ve Boyut p eksenleri üzerine değişken vektörlerinin izdüşümleri, ilgili değişkenlerin varyans oranları ile orantılıdır. Bu değişken vektörlerinin göreceli uzunlukları, her bir değişkenin çoklubağlantıya katkısını ifade etmektedir. Altıncı bölümde ayrıca çoklubağlantının görsel teşhisinde kullanılan Tableplot yönteminin de tanımı yapılmış ve bağlantı biplot ile gerçek bir veri kümesi

üzerinden karşılaştırılması ele alınmıştır. Özellikle ekonomi alanında yapılan araştırmalarda çoklubağlantı problemi ile sıklıkla karşılaşılmaktadır. Çoklubağlantı probleminin görsel teşhisinde kullanılan bu iki yöntem çok net yorumlar vermesine rağmen literatürde çok fazla bilinen yöntemler değildir. Bu bağlamda çalışmanın, bu yöntemlerin çoklubağlantının görsel teşhisinde kullanımının önemini göstermesi açısından literatüre ampirik olarak katkı sağlaması kaçınılmazdır.

Uygulama 6, 1990-2007 dönemi boyunca yüksek gelirli OECD ülkelerindeki temel makroekonomik göstergeleri Alan Biplot tekniği ile incelemektedir. Alan Biplot, bir veri kümesinin ij. elemanının yaklaşık değerini bulmak için döndürülmüş bir i satır noktası, bir j sütun noktası ve orijin arasında gerilen üçgenin alanını kullanır. Alan biplot'da sütun ve satır noktalarının farklı çiftleri ve orijin ile gerilen alan doğrudan karşılaştırılabilir. Bu Alan biplot'un ana avantajlarından biridir. Bu uygulama, OECD ülkeleri arasındaki ilişkilerin belirlenmesinde beş makroekonomik göstergenin rolü üzerine ampirik kanıt sağlamaktadır. Elde edilecek Alan Biplot grafiğinin görsel olarak incelenmesi ile değişkenlere göre ülkeler arasında net bir ayrım tespit edilmektedir. Alan Biplot grafiğinden elde edilen yorumların iktisadi olarak değerlendirilmesiyle, negatif şoklara karşı güvenlik açıklarını dikkate alan istikrarlı bir yapı sergileyen ekonomilere sahip ülkeler diğerleriyle kıyaslandığında krizden daha az etkilendikleri tespit edilmiştir. Ayrıca çalışmada G7, PIGS ve enflasyon hedefleyen ülkeler de Alan Biplot yaklaşımıyla incelenmiştir.

Bu tez çalışması, çok değişkenli bir veri kümesinin en iyi şekilde yorumlanmasında biplot tekniğinin oldukça başarılı olduğunu yapılan gerçek veri uygulamaları üzerinden net bir şekilde göstermektedir.

Daha sonraki çalışmalarda, dayanıklı (robust) ve bulanık (fuzzy) tekil değer ayrıştırması kullanılarak çalışmada ele alınan biplotların dayanıklı ve bulanık versiyonları teorik olarak elde edilip, klasik yöntemlerle karşılaştırılabilir. Ayrıca veri kümesinde kayıp gözlemler olması durumunda biplot tekniğinin kullanımı bir başka çalışma kapsamında değerlendirilebilir.

KAYNAKLAR

Aitchison, J. and Greenacre, M. 2002. Biplots of Compositional Data, Journal of the Royal Statistical Society: Series C, Vol. 51(4), pp. 375-392.

Alkan, B.B. and Atakan, C. 2011. Use of Canonical Variate Analysis Biplot in Examination of Choline Content Data of Some Foods, International Journal of Food Sciences and Nutrition, Vol. 62(2), pp. 171–174**.**

Alkan, B.B., Atakan, C. and Şahin, A. 2011. Measuring inequalities in the distribution of health workers by biplot approach: the case of Turkey. Journal of Economics and Behavioral Studies,Vol. 2(2), pp. 57-66.

Anonymous 1998. Institute of Medicine, Food and Nutrition Board USA. Dietary Reference Intakes: Thiamin, Riboflavin, Niacin, Vitamin B-6, Vitamin B012, Pantothenic Acid, Biotin, and Choline. Washington, D.C.: National Academy of Sciences.

Anonymous. 2007 Minitab Inc. Minitab, Version 15. PA, USA.

Anonymous. 1990 WHO.. Diet, nutrition and prevention of chronic diseases. Geneva: World Health Organization.

Anonymous. 2010. Wikipedia. Choline. The free encyclopedia. http://en.wikipedia.org/wiki/Choline Erişim Tarihi: 21.07.2010.

Anonymous. 2011. Wikipedia. Petroleum. The free encyclopedia. http://en.wikipedia.org/wiki/ Petroleum. Erişim Tarihi: 20.07.2011.

Anonymous. 2009. SAS Institute Inc. Cary, NC, USA.

Anonymous. 2008 SPSS . Inc. SPSS, Version 17. Chicago, USA.

Anonymous. 2010. R Development Core Team. R: Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Atakan, C., Alkan, B.B. and Şahin, A. 2009. Examination of relation between nutrient components and fruits: Biplot approach, International Journal of Food Sciences and Nutrition, Vol. 60(S1), pp. 181-189.

Barro, R.J. 1991. Economic Growth in a Cross Section of Countries, Quarterly Journal of Economics, Vol. 106(2), pp. 407-443.

Bartkowiak, A. and Szustalewicz, A. 1995. The Augmented Biplot and Some Examples of its Use, Machine Graphics & Vision, Vol. 4(3), pp. 161-185.

Bashirian, S., Allahverdpour, H. and Moeini, B. 2008. Fruit and vegetable intakes among elementary schools'pupils: Using five-a-day educational program, J Res Health Sci, Vol. 8, pp. 56-63.

Belsley, D. A., Kuh, E. and Welsch, R. E. 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

Belsley, D.A. 1991a. A Guide to Using the Collinearity Diagnostics, Computer Science in Economics and Management, Vol. 4, pp. 33–50.

Belsley, D. A. 1991b. Conditioning Diagnostics: Collinearity and Weak Data in Regression, New York, Wiley.

Berk, K. N. 1977. Tolerance and condition in regression computations, Journal of the American Statistical Association, Vol. 72, pp. 863-866.

Blasius, J., Eilers, P. and Gower, J.C. 2009. Better Biplots, Computational Statistics & Data Analysis, Vol. 53, pp. 3145-3158.

Bradu, D. and Gabriel, K.R. 1978. The Biplot as a Diagnostic Tool for Models of Two-way Tables, Technometrics, Vol. 20, pp. 47-68.

Chambers, J.M., Cleveland, S. B. and Kleiner, P.A. 1983. Tukey. Graphical Methods for Data Analysis, Wadsworth, Belmont, CA.

Chatterjee, S. and Price, B. 1977. Regression Analysis by Example. John Wiley, & Sons, New York.

Coxon, A.P.M. 1982. The User's Guide to Multidimensional Scaling, London: Heinnemann.

Daigle, G. and Rivest, L.P. 1992. A robust biplot, Canadian Journal of Statistics, Vol. 20(3), pp. 241-255.

Driskell, J.A. 2000. Sports nutrition. CRC Pres, New York.

Ebrahimof, S., Hoshyarrad, A., Hossein, A., Zandi, N., Larijani, B. and Kimiagar, M. 2006. Fruit and vegetable intake in postmenopausal women with osteoperia, ARYA J, Vol. 1, pp. 183-187.

Eckart, C. and Young, G. 1936. The Approximation of One Matrix by Another of Lower Rank, Psychometrika, Vol. 1(3), pp. 211-218.

Farrar, D.E. and Glauber, R.R. 1967. Multicollinearity in regression analysis: the problem revisited, Review of Economics and Statistics, Vol. 49, pp. 92-107.

Fenty, J. 2004. Analyzing distances, The Stata Journal, Vol. 4(1), pp. 1-26.

Fox, C.L., Martin, A.G. and Civit, S.V. 1996. Cranial variation in the Iberian Peninsula and the Balearic Islands: Inferences about the history of the population, American Journal of Physical Anthropology, Vol. 99(3), pp. 413-428.

Friendly, M. and Kwan, E. 2009. Where's Waldo: Visualizing Collinearity Diagnostics, The American Statistician, Vol. 63(1), pp. 56-65.

Gabriel, K.R. and Odoroff, C.L. 1990. Biplots in Biomedical Research. Statistic in Medicine, Vol. 9, pp. 469 - 485.

Gabriel, K.R. 1995. MANOVA Biplots for Two-way Contingency Tables, Recent Advances in Descriptive Multivariate Analysis, Oxford: Clarendon Pres.

Gabriel, K.R. 1971. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis, Biometrika, Vol. 58(3), pp. 453-467.

Gabriel K.R. 1978. Least squares approximation of matrices by additive and multiplicative models. J R Statist Soc, Ser. B., Vol. 40, pp. 186-196.

Gardner, S. 2001. Extensions of Biplot Methodology to Discriminant Analysis with Applications of Non-parametric Principal Components, Unpublished PhD thesis, University of Stellenbosch, Stellenbosch, South Africa.

Gardner, S. and Roux, N.J.L. 2005. Extensions of Biplot Methodology to Discriminant Analysis, Journal of Classification, Vol. 22, pp. 59-86.

Gardner, S., Roux N.J.L., Rypstra, T. and Swart, J.P.J. 2005. Extending a Scatterplot for Displaying Group Structure in Multivariate Data: A Case Study, ORiON, Vol. 21(2), pp. 111-124.

Gardner, S., Gower, J.C. and Roux, N.J.L. 2006. A Synthesis of Canonical Variate Analysis, Generalised Canonical Correlation and Procrustes Analysis, Computational Statistics and Data Analysis, Vol. 50, pp. 107-134.

Gardner-Lubbe, S., Roux N.J.L. and Gower, J.C. 2008. Measures of Fit in Principal Component and Canonical Variate Analyses, Journal of Applied Statistics, Vol. 35(9), pp. 947-965.

Gardner-Lubbe, S., Roux N.J.L., Maunders, H., Shah, V. and Patwardhan, S. 2009. Biplot methodology in exploratory analysis of microarray data, Statistical Analysis and Data Mining, Vol. 2(2), pp. 135-145.

Gerrild, P.M. and Lantz, R.J. 1969. Chemical Analysis of 75 Crude Oil Samples from Pliocene Sand Units, ELK Hills Oil Field,. U.S. Geological Survey Open-File Report, California.

Gower, J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis, Biometrika, Vol. 53, pp. 325–338.

Gower, J.C. and Hand, D.J. 1996. Biplots, Chapman&Hall, London.

Gower, J.C. 2004. The geometry of biplot scaling, Biometrika,Vol. 91,pp. 705-714.

Gower, J.C. and Krzanowski W.J. 1999. Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance, Applied Statistics, Vol. 48, pp. 505-519.

Gower, J.C., Groenen, P.J.F. and Velden, V.D. 2010. Area Biplots, Journal of Computational and Graphical Statistics, Vol.19(1), pp. 46-61.

Gower, J.C., Lubbe, S. and Roux, N.J.L. 2011. Understanding Biplots, John Wiley & Sons, Chichester, UK.

Greenacre, M.J. 1984. Theory and applications of correspondence analysis. Academic Press, London.

Greenacre, M.J. 1993. Biplots in correspondence analysis, Journal of Applied Statistics, Vol. 20, pp. 251-269.

Gunst, R.F. and Mason, R.L. 1977. Advantages of examining multicollinearities in regression analysis, Biometrics, Vol. 33(1), pp. 249-260.

He, M. and Haymer, D. S. 1999. Genetic relationships of populations and the origins of new infestations of the Mediterranean fruit fly, Molecular Ecology, Vol. 8(8), pp. 1247-1257.

Hocking, R.R. 1983. Developments in linear regression methodology: 1959-1982, Technometria, Vol. 25, pp. 219-249.

Householder, A.S. and Young, G. 1938. Matrix approximation and latent roots, American Mathematical Monthly, Vol. 45, pp. 165-171.

Johnson, R.A. and Wichern, D.W. 2002. Applied Multivariate Statistical Analysis. Prentice-Hall, New Jersey.

Jolliffe, I.T. 2002. Principal Component Analysis. Springer, New York.

Kidwell, J.S. and Brown, L.H. 1982. Ridge regression as a technique for analyzing models with muiticollinearity, Journal of Marriage and the Family, Vol. 44, pp. 287-299.

Kosaki K., Jones M. C. and Stayboldt, C. 1996. Zimmer phocomelia: Delineation by principal coordinate analysis, American Journal of Medical Genetics, Vol. 66(1), pp. 55-59.

Krzanowski, W.J. 2004. Biplots for Multifactorial Analysis of Distance, Biometrics, Vol. 60, pp. 517-524.

Krzanowski, W.J. and Marriott, F.H.C. 1994. Multivariate Analysis, Part I: Distributions, Ordination and Inference. London: Arnold.

Mankiw, N.G., Romer, D. and Weil, D.N. 1992. A Contrubution to the Empirics of Economic Growth, Quarterly Journal of Economics, Vol. 107, pp. 407-438.

Marquardt, D.W. 1970. Generalized inverses; ridge regression; biased linear estimation and non-linear estimation, Technometrics, Vol. 12, pp. 591-612.

Marquardt, D.W. and Snee, R.D. 1975. Ridge regression in practice. The American Statistician, Vol. 29, pp.3–19.

Martín-Rodríguez, J., Galindo-Villardón, M.P. and Vicente-Villardón, J.L. 2002. Comparison and integration of subspaces from a biplot perspective, Journal of Statistical Planning and Inference, Vol. 102(2), pp. 411-423.

Ojeda, M.M. and Juarez-Cerrillo, S.F. 1996. Biplot display for diagnostic in a two-level regression model for growth curve analysis, Computational Statistics and Data Analysis, Vol. 22, pp. 583-597.

Park, M., Lee J.W., Lee J.B. and Song S.H. 2008. Several biplot methods applied to gene expression data, Journal of Statistical Planning and Inference, Vol. 138(2), pp. 500-515.

Patterson, K.Y., Bhagwat, S.A., Williams, J.R., Howe, J.C. and Holden, J.M. 2008. USDA Database for the Choline Content of Common Foods Release Two, Agricultural Research Service Beltsville Human Nutrition Research Center, Beltsville, Maryland.

Paulson, D.S. 2007. Handbook of Regression and Modeling: Applications for the Clinical and Pharmaceutical Industries, Chapman and Hall/CRC, Boca Raton.

Phillips, P.C.B. and Sul, D. 2006. Economic Transition and Growth, University of Auckland, Mimeo.

Rawlings, J.O., Pantula, S.G. and Dickey, D.A. 1998. Applied Regression Analysis: A Research Tool, Springer-Verlag, New York.

Roux, N.J.L. and Gardner, S. 2005. Analysing Your Multivariate Data as a Pictorial: A Case for Applying Biplot Methodology?, International Statistical Review, Vol. 73(3), pp. 365-387.

Roux, N.J.L and Gardner, S. 2006. A biplot-based approach to discriminant analysis with categorical variables in the presence of reversals, The South African Statistical Journal, Vol. 40(1),pp. 1–31.

Silvey, S.D. 1969. Multicollinearity and imprecise estimation. J Roy Statist Soc, Vol. 31, pp. 539-552.

Sparks, R., Adolphson, A. and Phatak, A. 1997. Multivariate Process Monitoring Using the Dynamic Biplot, International Statistical Review, Vol. 65(3), pp. 325-349.

Şahin, A., Alkan, B.B., Akdi, Y. and Atakan, C. 2009. Biplot Representation of the Financal Invesment Returns, Arab Bank Review, Vol. 10(2), pp. 76-86.

Şahin, A., Atakan, C. and Alkan, B.B. 2010. Türkiye'de Yeni Tanımlı Sanayi Üretim Endeksi ve Alt Sanayi Gruplarındaki Büyümenin Biplot Yöntemi ile İncelenmesi, Journal of Yasar University, Cilt 18(5), s. 2958-2969.

Theil, H. 1971. Principles of Econometrics. Wiley, New York.

Tukey, J.W. 1977. Exploratory Data Analysis, Addison-Wesley, Reading, Massachusetts.

Uysal, A. 2006. Ham petrol fraksiyonlarının biyolojik bozunma sonrası fizikokimyasal özelliklerinin değişimi, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.

Weisberg, S. 1980. Applied Linear Regression. John Wiley & Sons, New York.

Willan, R.A. and Watts, G.D. 1978. Meaningful Multicollinearity Measures, Technometrics, Vol. 20(4), pp. 407-412.

Young, G. and Householder, A.S. 1938. Discussion of a set of points in terms of their mutual distances, Psychometrika, Vo. 3, pp.19-22.

Zeisel, S. 2006. The fetal origins of memory: the role of dietary choline in optimal brain development, Journal of Pediatrics, Vol. 149, pp.131-136.

Zeisel, S.H. and daCosta K.A. 2009. Choline: an essential nutrient for public health, Nutrition Reviews, Vol. 67 (11), pp. 615-23.

EKLER

EK 1 Türkiye Sanayi Üretim Endeksi (2005=100), Haziran, 2005 – Haziran, 2009 genel ve ana kategorileri veri kümesi

EK 2 Çeşitli meyve türlerinin 100 gramında bulunan besin bileşenleri veri kümesi

EK 3 Beş farklı besin grubunda yer alan toplam 105 ürünün 100 gramında bulunan beş kolin değişkeninin yer aldığı veri kümesi

EK 4 Üç bölgeden alınan ham petrol özellikleri veri kümesi

EK 5 Avrupa ve Merkez Asya Ülkelerinin temel göstergelerine ilişkin veri kümesi

EK 6 Yüksek gelirli OECD ülkelerinin temel ekonomik göstergeleri veri kümesi

EK 1 Türkiye Sanayi Üretim Endeksi (2005=100), Haziran, 2005 – Haziran, 2009 genel ve ana kategorileri veri kümesi

	Toplam	ARA	DAYANIKLI	DAYANIKSIZ	SERMAYE	IMALAT
Haz.05	3	2	1	10	-1	3
Tem.05	-2	-2	6	-9	-5	-4
Ağu.05	0	1	-1	2	-7	-1
Eyl.05	9	5	12	17	20	12
Eki.05	2	-1	-3	10	-1	2
Kas.05	-9	-11	-11	-8	-8	-9
Ara.05	11	12	5	7	22	11
Oca.06	-25	-22	-24	-30	-37	-28
Şub.06	11	13	20	11	22	13
Mar.06	18	20	18	16	26	20
Nis.06	-5	-1	7	-10	-11	-5
May.06	6	5	2	6	13	6
Haz.06	2	1	-7	6	1	1
Tem.06	-3	-1	-8	-10	-5	-5
Ağu.06	-3	-1	1	3	-25	-4
Eyl.06	7	2	11	12	33	10
Eki.06	-6	-8	-4	-2	-11	-6
Kas.06	9	5	12	9	19	9
Ara.06	-1	-1	-14	-3	2	-2
Oca.07	-13	-9	-5	-19	-21	-14
Şub.07	2	2	7	1	7	3
Mar.07	15	17	12	14	22	17
Nis.07	-5	-3	-5	-9	-8	-6
May.07	7	6	6	9	13	8
Haz.07	-1	-2	-3	1	-3	-2
Tem.07	-2	-1	-15	-2	-6	-3
Ağu.07	-1	1	13	3	-20	-2
Eyl.07	3	0	0	3	31	6
Eki.07	0	-4	3	7	0	1
Kas.07	7	5	6	7	13	8
Ara.07	-10	-13	-14	-16	-4	-12
Oca.08	-1	4	-1	-5	-10	-2
Şub.08	-1	-3	6	-1	5	0
Mar.08	9	13	25	3	11	10
Nis.08	-1	0	-15	-4	4	-1
May.08	4	4	2	7	1	3

EK 1 Türkiye Sanayi Üretim Endeksi (2005=100), Haziran, 2005 – Haziran, 2009 genel ve ana kategorileri veri kümesi (devam)

Aylar	Toplam	ARA	DAYANIKLI	DAYANIKSIZ	SERMAYE	IMALAT
Haz.08	-2	-3	-2	-1	-4	-3
Tem.08	0	2	-3	-3	-8	-2
Ağu.08	-8	-7	-9	1	-31	-10
Eyl.08	2	-3	7	4	30	5
Eki.08	-3	-5	1	6	-10	-2
Kas.08	0	-5	2	8	-7	0
Ara.08	-15	-20	-22	-18	-7	-18
Oca.09	-5	4	-13	-9	-27	-7
Şub.09	-4	-4	6	-7	6	-3
Mar.09	13	17	21	11	16	15
Nis.09	1	3	0	1	5	3
May.09	5	4	16	3	11	6
Haz.09	7	7	9	9	6	7

EK 2 Çeşitli meyve türlerinin 100 gramında bulunan besin bileşenleri

Etiketler	Meyveler	Su(gr)	Protein(gr)	Karbonhidrat(gr)	Yağ(gr)	Enerji(Kcal)
1	Uzum(Taze)	81.4	0.6	17.3	0.3	67
2	Uzum(Kuru)	18	2.5	77.4	0.2	289
3	İncir(Yas)	82	1.2	20.4	0.4	88
4	İncir(Kuru)	23	4.3	6.9	1.3	274
5	Portakal	86	0.8	8.5	0.1	35
6	Mandalina	87	0.8	11.6	0.2	46
7	Greyfurt	91	0.6	5.3	0.2	43
8	Limon(Suyu)	91	0.3	1.6	0.5	7
9	Elma	84	0.3	11.9	0.3	46
10	Muz	71	1.1	19.2	0.2	76
11	Şeftali	86	0.6	9.1	0.2	37
12	Çilek	89	0.6	6.2	0.3	26
13	Kayısı(Yas)	85.3	1	18.2	0.6	51
14	Kayısı(Kuru)	25	5	66.5	1	260
15	Armut	83	0.3	10.6	0.2	41
16	Karpuz	92.6	0.5	6.4	0.2	26
17	Kavun	90.6	0.8	7.7	0.3	33
18	Avokado	69	4.2	1.8	22.2	223
19	Erik(Kırmızı)	81.1	0.5	17.8	0.2	66
20	Ananas	77	0.5	11.6	0.3	46
21	Nar	82.3	0.5	16	0.3	63
22	Kiraz	80.4	1.3	17.4	0.3	70
23	Karadut	76.5	0.9	19.8	1.1	93
24	Ahududu	84.2	1.2	13.6	0.5	55
25	Ayva	83.8	0.4	15.3	0.1	71

EK 3 Beş farklı besin grubunda yer alan toplam 105 ürünün 100 gramında bulunan beş kolin değişkeninin yer aldığı veri kümesi

Etiketler	Besinler	Gruplar	FREECHO	GPC	PCHO	PTDCHO	SM
1	01001	G1	0.6	1.2	0.7	11	5.4
2	01046	G1	7.9	14	2.1	7.1	4.6
3	01009	G1	1.6	2.3	0.6	7.4	4.6
4	01012	G1	3.6	8.4	1.3	2.5	2.5
5	01015	G1	2.9	8.1	1.3	2	2
6	01014	G1	3.7	9.2	1.6	1.6	1.8
7	01017	G1	3.6	9.3	1.5	7.3	5.6
8	01029	G1	2.3	2.7	1	5.2	3
9	01040	G1	4.5	0.6	0	6.3	4.2
10	01069	G1	0.5	1.4	0	0.1	0.3
11	01049	G1	3.9	9	1.1	2.6	2.1
12	01199	G1	4	8	1.2	0.9	1
13	01056	G1	3.9	7.6	1.2	3.7	2.7
14	01124	G1	0.2	0.6	0	0.3	0
15	01103	G1	5.4	8.2	1.1	1.7	0.8
16	01082	G1	4	9.8	1.9	1.2	0.7
17	01085	G1	2.8	9.7	1.7	0.8	0.7
18	01079	G1	2.8	10	1.6	1.2	0.9
19	01077	G1	3.7	7.5	1.9	0.6	0.7
20	43261	G1	3.3	7.8	2	1.9	1.4
21	01121	G1	2.1	7.8	1.6	1.5	1.1
22	01117	G1	2.3	9.1	1.7	1	1.1
23	98042	G2	5.8	1.3	1.1	28	5.2
24	98043	G2	3.8	1.4	1.1	28	5
25	05006	G2	6	1	3.6	41	8.5
26	05013	G2	5.7	1.1	3.7	54	15
27	98084	G2	5.8	0.8	2.8	44	11
28	98085	G2	3.2	1.6	2.1	46	8.9
29	98110	G2	3.1	1.8	2.3	47	6
30	98086	G2	6	1.1	3	62	12
31	98111	G2	6.3	1.1	3.2	60	9
32	05009	G2	5.3	1.2	3.5	45	12
33	05011	G2	5.8	1.1	3.5	45	11
34	98087	G2	5.1	1.4	1.8	25	13
35	98112	G2	5.3	1	2.4	52	7.4
36	98088	G2	5.8	0.8	2.8	46	9.7
37	98089	G2	4.9	0.9	2.2	49	11
38	05320	G2	4	1.5	1.7	55	15

EK 3 Beş farklı besin grubunda yer alan toplam 105 ürünün 100 gramında bulunan beş kolin değişkeninin yer aldığı veri kümesi (devam)

Etiketler	Besinler	Gruplar	FREECHO	GPC	PCHO	PTDCHO	SM
39	05313	G2	5.7	1.4	1.5	52	14
40	05312	G2	4.6	1.4	1.6	48	14
41	98041	G2	5.5	2.1	1.5	32	4.9
42	05174	G2	9.5	3.1	3.6	41	25
43	05173	G2	41	0.3	5	24	19
44	05175	G2	25	0.9	3.7	84	14
45	98003	G3	0.3	0	0	3.1	0
46	09032	G3	7.1	0.8	0	6	0
47	09038	G3	8.6	0.7	2.5	2.2	0.1
48	09040	G3	3.2	5.6	0.5	0.4	0
49	09042	G3	1.8	0.8	0	6	0
50	09050	G3	3	0.6	0.7	1.8	0
51	09078	G3	1.3	1	0.5	2.6	0
52	09112	G3	3.6	1.2	0.3	2.6	0
53	09148	G3	2.6	1	0.5	3.6	0
54	09436	G3	1	0.3	0.3	0	0
55	09191	G3	1.7	1.2	0.6	2.7	0
56	09214	G3	8.1	4.2	0.6	7	0
57	09202	G3	4.7	1.1	0.5	2.1	0
58	09236	G3	0.8	1.1	0.5	3.7	0
59	09252	G3	2.2	0.2	0	2.7	0
60	09429	G3	5.1	0	0.5	0.1	0
61	09291	G3	6.7	0.9	0	2.5	0
62	09298	G3	9.4	0.3	1.1	0.3	0
63	09302	G3	3.1	1.6	0.6	7	0
64	09316	G3	0.6	0.9	0	4.2	0
65	09218	G3	5.2	1	0.6	3.5	0
66	09326	G3	3.1	0	0.7	0.2	0.1
67	11001	G4	11	0.6	1.8	1	0
68	11011	G4	12	0.6	2.4	0.9	0
69	11080	G4	4.1	0.6	0.9	0.4	0
70	11090	G4	18	0	0.4	0.1	0.1
71	11109	G4	6.1	2.9	1.3	0.3	0
72	11112	G4	9.7	4	2	1.5	0
73	11124	G4	6.8	0	1.1	0.8	0
74	11143	G4	5.3	0	0.7	0.2	0
75	11206	G4	3.5	0.5	0.7	0.9	0
76	35196	G4	13	0.3	1.9	4.5	0
77	11250	G4	5.9	0	2.4	0.2	0
78	11252	G4	4.8	0	1.5	0.4	0
79	11257	G4	7.9	0	3.4	0.5	0
80	11260	G4	5.9	5.1	1.4	4.5	0
81	11282	G4	4.4	0.6	0.7	0.5	0

EK 3 Beş farklı besin grubunda yer alan toplam 105 ürünün 100 gramında bulunan beş kolin değişkeninin yer aldığı veri kümesi (devam)

Etiketler	Besinler	Gruplar	FREECHO	GPC	PCHO	PTDCHO	SM
82	11294	G4	3.7	0.7	0.6	0.5	0
83	11313	G4	2.2	0.8	0.7	24	0
84	11333	G4	3.6	0	1.2	0.7	0
85	11354	G4	7.9	2.6	0.3	0.3	0
86	11429	G4	4.8	0	1	0.6	0
87	11478	G4	0.5	0.7	2.4	5.7	0
88	11644	G4	2.1	0.6	2.6	5.3	0
89	16006	G5	17	1.3	0.8	12	0
90	98019	G5	14	1.8	0.9	11	0
91	16010	G5	13	1.6	0.9	9.2	0
92	16011	G5	25	1.6	1	12	0
93	98020	G5	11	1.6	0.9	11	0
94	16034	G5	17	2.1	0.6	13	0
95	16039	G5	14	0.8	0.6	12	0
96	16038	G5	21	1.9	0	22	0
97	16037	G5	50	4.5	0	33	0
98	16043	G5	11	1.4	0.2	22	0
99	16042	G5	32	2.9	0.8	31	0
100	43133	G5	9.9	1.8	0.4	2.2	0
101	16097	G5	25	1.3	2.2	33	0
102	16098	G5	26	1.1	0.7	38	0
103	16399	G5	39	1	1.8	53	0
104	16087	G5	18	1.3	1.8	32	0
105	16108	G5	47	2.9	1.1	65	0

EK 4 Üç farklı bölgeden alınan ham petrol özellikleri veri kümesi

Gruplar	Vanadyum	Demir	Berilyum	Doymus_H	Aromatik_H
Bölge 1	3.9	51	0.2	7.06	12.19
Bölge 1	2.7	49	0.07	7.14	12.23
Bölge 1	2.8	36	0.3	7	11.3
Bölge 1	3.1	45	0.08	7.2	13.01
Bölge 1	3.5	46	0.1	7.81	12.63
Bölge 1	3.9	43	0.07	6.25	10.42
Bölge 1	2,7	35	0	5.11	9
Bölge 1	5	47	0.07	7.06	6.1
Bölge 2	3,4	32	0.2	5.82	4.69
Bölge 2	1,2	12	0	5.54	3.15
Bölge 2	8,4	17	0.07	6.31	4.55
Bölge 2	4,2	36	0.5	9.25	4.95
Bölge 2	4,2	35	0.5	5.69	2.22
Bölge 2	3,9	41	0.1	5.63	2.94
Bölge 2	3,9	36	0.07	6.19	2.27
Bölge 2	7,3	32	0.3	8.02	12.92
Bölge 2	4,4	46	0.07	7.54	5.76
Bölge 2	3	30	0	5.12	10.77
Bölge 2	6.3	13	0.5	4.24	8.27
Bölge 3	1.7	5.6	1	5.69	4.64
Bölge 3	7.3	24	0	4.34	2.99
Bölge 3	7.8	18	0.5	3.92	6.09
Bölge 3	7.8	25	0.7	5.39	6.2
Bölge 3	7.8	26	1	5.02	2.5
Bölge 3	9.5	17	0.05	3.52	5.71
Bölge 3	7.7	14	0.3	4.65	8.63
Bölge 3	11	20	0.5	4.27	8.4
Bölge 3	8	14	0.3	4.32	7.87
Bölge 3	8.4	18	0.2	4.38	7.98
Bölge 3	10	18	0.1	3.06	7.67
Bölge 3	7.3	15	0.05	3.76	6.84
Bölge 3	9.5	22	0.3	3.98	5.02
Bölge 3	8.4	15	0.2	5.02	10.12
Bölge 3	8.4	17	0.2	4.42	8.25
Bölge 3	9.5	25	0.5	4.44	5.95
Bölge 3	7.2	22	1	4.7	3.49
Bölge 3	4	12	0.5	5.71	6.32
Bölge 3	6.7	52	0.5	4.8	3.2
Bölge 3	9	27	0.3	3.69	3.3
Bölge 3	7.8	29	1.5	6.72	5.75
Bölge 3	4.5	41	0.5	3.33	2.27
Bölge 3	6.2	34	0.7	7.56	6.93

EK 4 Üç farklı bölgeden alınan ham petrol özellikleri veri kümesi (devam)

Gruplar	Vanadyum	Demir	Berilyum	Doymus_H	Aromatik_H
Bölge 3	5.6	20	0.5	5.07	6.7
Bölge 3	9	17	0.2	4.39	8.33
Bölge 3	8.4	20	0.1	3.74	3.77
Bölge 3	9.5	19	0.5	3.72	7.37
Bölge 3	9	20	0.5	5.97	11.17
Bölge 3	6.2	16	0.05	4.23	4.18
Bölge 3	7.3	20	0.5	4.39	3.5
Bölge 3	3.6	15	0.7	7	4.82
Bölge 3	6.2	34	0.07	4.84	2.37
Bölge 3	7.3	22	0	4.13	2.7
Bölge 3	4.1	29	0.7	5.78	7.76
Bölge 3	5.4	29	0.2	4.64	2.65
Bölge 3	5	34	0.7	4.21	6.5
Bölge 3	6.2	27	0.3	3.97	2.97

EK 5 Avrupa ve merkez Asya ülkelerinin temel makroekonomik göstergeleri 2008 veri kümesi

ÜLKE ADI	LOG(GSYİH)[a]	LOG(DH)[b]	TÜFE[c]	KN[d]	LOG(ÖLO)[e]	TİO[f]	HHNTH[g]
Avusturya	0.5027	1.8271	3.1823	67.1600	0.5441	3.8000	0.0380
Belçika	0.6522	1.9884	4.4892	97.3600	0.5911	7.0000	0.7152
Hırvatistan	0.7830	1.7580	6.0670	57.2800	0.6721	8.4000	1.0380
Çek C.	0.8031	1.8663	6.3551	73.5000	0.4771	4.4000	2.6965
Danimarka	0.5314	1.9379	3.3995	86.6800	0.5441	3.3000	-1.1530
Estonya	1.0156	1.8417	10.3656	69.4600	0.6721	5.5000	-4.8953
Finlandiya	0.6092	1.8014	4.0660	63.3000	0.4150	6.4000	1.2462
Fransa	0.4493	1.8885	2.8139	77.3600	0.5315	7.4000	-0.0289
Yunanistan	0.6183	1.7853	4.1524	61.0000	0.4771	7.7000	2.7707
İrlanda	0.6078	1.7877	4.0535	61.3400	0.5682	3.0000	-9.1408
İtalya	0.5248	1.8330	3.3478	68.0800	0.5315	6.0000	-3.1607
Letonya	1.1876	1.8333	15.4032	68.1200	0.8633	6.7000	-1.5223
Lüksemburg	0.5315	1.9161	3.4003	82.4400	0.2553	5.8000	-5.0780
Makedonya	0.8585	1.8254	7.2185	66.9000	1.0212	5.1000	2.8448
Moldova	1.1063	1.6208	12.7744	41.7600	1.1818	33.8000	6.4999
Hollanda	0.3953	1.9129	2.4848	81.8200	0.5911	4.0000	0.7519
Norveç	0.5759	1.8892	3.7662	77.4800	0.4624	2.8000	0.4434
Polonya	0.6384	1.7876	4.3494	61.3200	0.7634	2.6000	5.9041
Portekiz	0.4134	1.7742	2.5904	59.4600	0.5052	7.1000	1.6420
Romanya	0.8948	1.7343	7.8483	54.2400	1.0492	7.6000	8.5675
Rusya	1.1495	1.8624	14.1078	72.8400	1.0756	5.8000	10.8113
Sırbistan	1.1096	1.7163	12.8707	52.0400	0.8195	13.6000	8.0322
Slovak C.	0.6626	1.7525	4.5982	56.5600	0.7782	9.5000	5.8844
İspanya	0.6095	1.8872	4.0688	77.1200	0.5563	11.3000	-2.0779
İsveç	0.5362	1.9271	3.4370	84.5400	0.3802	6.2000	-0.7847
İsviçre	0.3843	1.8662	2.4225	73.4800	0.6128	3.4000	0.0490
Türkiye	1.0189	1.8368	10.4441	68.6800	1.2989	9.4000	-1.5445
Ukrayna	1.4020	1.8324	25.2319	67.9800	1.1367	6.4000	11.1165
Büyük B.	0.6014	1.9540	3.9937	89.9400	0.6721	5.6000	-0.2238

a-) GSYİH, Logaritma olarak, kişi başına düşen GSYİH büyümesi (%yıllık) olarak

b-) LOG(DH), Logaritma olarak, Doğurganlık oranı, toplam (kadın başına doğum)

c-) TÜFE, Enflasyon, tüketici fiyatları (yıllık%)

d-) KN, Kentsel nüfus (toplam%)

e-) LOG(ÖLO), Logaritma olarak, Ölüm oranı, bebek (1000 canlı doğumda)

f-) TİO, İşsizlik, toplam (toplam işgücünün%)

g-) HHNTH, Kişi başına büyüme, Hanehalkı nihai tüketim harcamaları (%yıllık)

EK 6 1990-2007 yılları arasında Yüksek gelirli OECD ülkelerinin temel ekonomik göstergeleri ortalamaları veri kümesi

Ülke Kodu	Ülke Adı	GSYİH	HHNTH	EO	DT	İO
AUS	Avustralya	2.0233	2.1947	2.8068	37.7293	7.4111
AUT	Avusturya	1.9937	1.4556	2.2415	85.4129	4.4889
BEL	Belçika	1.7631	1.4465	2.1218	148.9334	7.9611
CAN	Kanada	1.7062	1.7458	2.2487	70.4702	8.3944
DNK	Danimarka	1.8109	1.7253	2.0817	80.2971	6.1389
FIN	Finlandiya	2.0409	1.5715	1.9368	67.6810	10.3333
FRA	Fransa	1.4164	1.4852	1.8478	48.9346	10.2333
DEU	Almanya	1.6962	1.3033	1.9873	60.7298	8.7176
GRC	Yunanistan	2.3560	2.3778	7.6328	51.0353	9.3944
HUN	Macaristan	2.0122	1.8638	15.1640	111.0477	8.0875
ISL	İzlanda	2.0948	2.2583	4.4550	71.4063	3.3471
IRL	İrlanda	5.3325	3.9109	3.2220	146.1120	8.7222
ISR	İsrail	2.1097	1.9732	6.9166	74.8592	9.0000
ITA	İtalya	1.1801	1.1368	3.3241	47.0772	9.8278
JPN	Japonya	1.3900	1.3702	0.5291	21.4076	3.7833
KOR	Kore	4.9796	4.4170	4.5029	66.5587	3.4500
LUX	Lüksemburg	3.3950	1.4688	2.3350	239.7903	2.9722
NLD	Hollanda	2.1925	1.7433	2.3349	120.4521	5.0333
NZL	Yeni Zelanda	1.7911	1.9383	2.3645	59.3638	6.4056
NOR	Norveç	2.4479	2.8748	2.1859	72.0702	4.3222
POL	Portekiz	1.8899	2.4041	4.6473	65.2708	5.8778
ESP	İspanya	2.2662	2.2174	3.7812	50.0638	15.5611
SWE	İsveç	1.8516	1.3064	2.4832	76.4440	6.8389
SWD	İsviçre	0.8595	0.6281	1.7217	78.7824	3.3556
GBR	İngiltere	2.0674	2.3523	3.3127	54.2428	6.7667
USA	Amerika	1.7168	2.0996	2.9024	23.8908	5.4333

ÖZGEÇMİŞ

Adı Soyadı **:** Bilal Barış ALKAN
Doğum Yeri **:** Sinop
Doğum Tarihi **:** 03.12.1978
Medeni Hali **:** Evli
Yabancı Dili **:** İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : Atatürk Lisesi (1992–1995)
Lisans : Ondokuz Mayıs Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü (1996–2000)
Yüksek Lisans : Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı (2002–2004)

Çalıştığı Kurum/Kurumlar ve Yıl:

Ondokuz Mayıs Üniversitesi Sinop Fen-Edebiyat Fakültesi (2000–2007)

Sinop Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü (2007–)

Yayınları (SCI ve diğer)

1- Alkan, B.B. and Atakan, C. (2011). Use of Canonical Variate Analysis Biplot in Examination of Choline Content Data of Some Foods. ***International Journal of Food Sciences and Nutrition,*** 62(2), 171–174. **(SCI)**

2- Alkan, B.B., Atakan, C., Şahin, A. (2011). Measuring inequalities in the distribution of health workers by biplot approach: the case of Turkey. ***Journal of Economics and Behavioral Studies***, 2(2), 57-66.

3- Şahin, A., Atakan, C., **Alkan, B.B**. (2010). Türkiye'de Yeni Tanımlı Sanayi Üretim Endeksi ve Alt Sanayi Gruplarındaki Büyümenin Biplot Yöntemi ile İncelenmesi. ***Journal of Yasar University***, 18(5), 2958–2969.

4- Atakan, C., **Alkan, B.B.,** Şahin, A. (2009). Examination of Relation Between Nutrient Components and Fruits: Biplot Approach. ***International Journal of Food Sciences and Nutrition***, Vol. 60, No. S1,181-189. **(SCI)**

5- Şahin, A., **Alkan, B.B.**, Akdi, Y., Atakan, C. (2009). Biplot Representation of the Financial Investment Returns. ***Arab Bank Review***, Vol. 10, No. 2, 76-86.

6- Alkan, B.B., Atakan, C. (2010). Çok Değişkenli Veride Grup Yapısının Gösterimi İçin Kanonik Değişken Analizi Biplot Kullanımı. ***VII. İstatistik Günleri Sempozyumu***, ODTÜ, 28-30 Haziran, Ankara.

7- Şahin, A., **Alkan, B.B.**, Akdi, Y., Atakan, C.(2009). Biplot Representation of the Financial Investment Returns. ***1st International Conference on Critical Issues in Business and Economics***, 5-6 November, 2009, Gümüşhane.

8- Alkan, B.B., Atakan, C. (2008). Çok Değişkenli Bir Veri Kümesinde Değişkenler Arası İlişkinin İncelenmesi: Biplot. ***VI. İstatistik Günleri Sempozyumu***, Ondokuz Mayıs Üniversitesi 27-30 Ağustos, Samsun.