

T.C.
BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TEMEL BİLEŞENLER ANALİZİ VE BİR UYGULAMASI

Fatma ŞAHİN ELİK

Matematik Anabilim Dalı

Matematik Programı

OCAK 2025

T.C.
BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TEMEL BİLEŞENLER ANALİZİ VE BİR UYGULAMASI

Fatma ŞAHİN ELİK

DANIŞMAN
Dr. Öğr. Üyesi Özlem ORHAN

Matematik Anabilim Dalı

Matematik Programı

OCAK 2025

ONAY

Fatma ŞAHİN ELİK tarafından hazırlanan “**Temel Bileşenler Analizi ve Bir Uygulaması**” adlı tez çalışması 20/01/2025 tarihinde yapılan sınavla aşağıdaki danışman tarafından Bandırma Onyedli Eylül Üniversitesi Eğitim Enstitüsü Matematik Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Dr. Öğr. Üyesi Özlem ORHAN

(Danışman)

Prof. Dr. Hüseyin IŞIK

(Üye)

Doç. Dr. Zehra PINAR

(Üye)

Bu tezin kabulü, Lisansüstü Eğitim Enstitüsü Yönetim Kurulu’nun / / gün ve / sayılı kararı ile onaylanmıştır.

Prof. Dr. Ahmet AYDIN

Enstitü Müdürü

BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
YÜKSEK LİSANS TEZİ
ETİK BEYANI

Bandırma Onyedi Eylül Üniversitesi Lisansüstü Eğitim Enstitüsü tez yazım kılavuzuna göre hazırlamış olduğum “**Temel Bileşenler Analizi ve Bir Uygulaması**” adlı tezimin özgün bir çalışma olduğunu, tez hazırlanırken tüm aşamalarda bilimsel etik ilkelerine uygun davrandığımı, tez kapsamında sunulan tüm verileri bilimsel etik ilkelerine uygun elde ettiğimi, tezde yararlandığım tüm eserlere atıf yaptığımı ve kaynaklar kısmında bu eserleri gösterdiğimi beyan ederim. 20/01/2025

Fatma ŞAHİN ELİK

ÖNSÖZ

Bu tez çalışması, Temel Bileşenler Analizi (TBA) yönteminin televizyon reyting verileri üzerinden uygulanmasını ve bu sayede büyük veri kümelerinin daha etkili bir şekilde analiz edilmesini hedeflemektedir. Medya sektöründe veri yoğun analizlerin giderek daha fazla önem kazandığı günümüzde, anlamlı içgörüler elde etmek ve veriyi daha verimli bir şekilde işlemek büyük bir gereklilik haline gelmiştir. Bu çalışmada, temel bileşenler analizinin medya verileri üzerindeki pratik uygulamaları ayrıntılı bir şekilde ele alınarak, veri analizinin etkili yöntemleri konusunda önemli bir katkı sağlanması amaçlanmıştır.

Bu çalışmanın, veri analizi ve istatistiksel modelleme yöntemlerinin medya ve benzeri alanlarda nasıl daha etkin kullanılabileceğine dair önemli bir katkı sağlayacağı düşünülmektedir. Teorik ve pratik açıdan literatüre katkı sağlayarak, ilgili alandaki diğer çalışmalara yol göstermesi amaçlanmaktadır. Bu tez, veri analizi alanında daha ileri düzeyde araştırmalar yapılabilmesine ve medya sektöründe daha etkili kararlar alınabilmesine katkıda bulunmayı amaçlamaktadır.

Bu tez çalışmasının ortaya çıkmasında emeği geçen, bana rehberlik eden ve desteklerini hiçbir zaman esirgemeyen değerli danışmanım Dr. Öğr. Üyesi Özlem ORHAN'a içten teşekkürlerimi sunuyorum.

Eşim Eray ELİK'e ise ayrı bir teşekkür etmek istiyorum. Bana her zaman moral kaynağı olmuş, her anımda desteğini esirgememiştir. Sevgisi, anlayışı, bu tez çalışmasını başarıyla tamamlama yolunda benim en büyük motivasyon kaynağım olmuştur. Bu süreçte bana verdiği güç ve inanç, sadece akademik değil, aynı zamanda kişisel anlamda da büyük bir anlam taşımaktadır.

Aileme de sonsuz teşekkürlerimi sunmak istiyorum. Özellikle anneme ve babama, bana her zaman güven verdikleri, sabırla yanımda oldukları ve sevgiyle beni destekledikleri için minnettarım. Her birinin verdiği destek ve cesaret, bu tez çalışmasının sonuca ulaşmasında önemli bir rol oynamıştır.

ÖZET

YÜKSEK LİSANS TEZİ

TEMEL BİLEŞENLER ANALİZİ VE BİR UYGULAMASI

Fatma ŞAHİN ELİK

Bandırma Onyedİ Eylöl Üniversitesi

Lisansüstü Eğitim Enstitüsü

Matematik Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Özlem ORHAN

Ocak 2025, 88 sayfa

Bu tez çalışmasında, Temel Bileşenler Analizi (TBA) yöntemi incelenmiş ve yüksek boyutlu veri kümelerinde boyut indirgeme ve veri analiz süreçleri araştırılmıştır. TBA, verideki en fazla varyansı açıklayan doğrusal bileşenler oluşturmakta ve veri setinin boyutunu indirgerken önemli bilgileri korumaktadır. Bu yöntem, çok sayıda değişken içeren veri kümeleriyle çalışırken kullanılmaktadır. Veriyi daha anlaşılır hale getirmekte ve görsel olarak daha kolay yorumlanabilmesini sağlamaktadır.

Çalışmada, TBA'nın matematiksel temelleri (kovaryans matrisi, özdeğerler ve özvektörler) ayrıntılı olarak açıklanmış ve bu yöntemin uygulama adımları kapsamlı şekilde ele alınmıştır. Uygulama aşamasında, TBA yöntemi televizyon reyting verilerine uygulanarak farklı izleyici gruplarının kanal tercihleri ve izleme alışkanlıkları analiz edilmiştir. TİAK tarafından sağlanan AB ve 20+ABC1 gruplarına ait reyting verileri karşılaştırılarak bu grupların davranışsal farklılıkları ve benzerlikleri değerlendirilmiştir.

Anahtar Kelimeler: Temel Bileşenler Analizi, Boyut İndirgeme, Reyting, İzleyici Davranışları, Doğrusal Bileşen.

ABSTRACT

M.Sc. THESIS

PRINCIPAL COMPONENT ANALYSIS AND AN APPLICATION

Fatma ŞAHİN ELİK

Bandırma Onyedli Eylül University

Institute of Graduate Studies

Department of Mathematics

Supervisor: Asst. Prof. Özlem ORHAN

January 2025, 88 pages

This thesis examines the **Principal Component Analysis (PCA)** method and investigates dimensionality reduction and data analysis processes in high-dimensional datasets. PCA constructs linear components that explain the maximum variance in the data, reducing the dataset's dimensionality while preserving essential information. This method is widely used when working with datasets containing numerous variables, making the data more comprehensible and visually interpretable.

In this study, the mathematical foundations of PCA (covariance matrix, eigenvalues, and eigenvectors) are explained in detail, and the implementation steps of this method are thoroughly addressed. During the application phase, the PCA method is applied to television rating data to analyze different audience groups' channel preferences and viewing habits. Rating data from the AB and 20+ABC1 groups, provided by TİAK, are compared to assess behavioral differences and similarities between these groups.

Keywords: Principal Component Analysis, Dimensionality Reduction, Rating, Audience Behavior, Linear Component.

İÇİNDEKİLER

ÖZET	I
ABSTRACT	II
İÇİNDEKİLER.....	III
SİMGE VE KISALTMA LİSTESİ	VI
ŞEKİLLER LİSTESİ	VII
TABLO LİSTESİ.....	VIII
1.GİRİŞ.....	1
2. MATEMATİKSEL TEMEL	5
2.1. İSTATİSTİK.....	5
2.2. STANDART SAPMA.....	5
2.3. VARYANS	6
2.3.1. Popülasyon Varyansı Formülü.....	6
2.3.2. Örneklem Varyansı Formülü	7
2.4. KOVARYANS	8
2.4.1. Varyans ve Kovaryans Formülleri.....	8
2.5. KOVARYANS MATRİS	8
2.6. ÖZVEKTÖRLER VE ÖZDEĞERLER.....	10
2.7. TEKİL DEĞER AYRIŞIMI.....	10
2.7.1. Temel Kavramlar	10
2.7.2. Normlar ve Öklit Formu	11
2.7.3. Ortonormal ve Üniter Matrisler	11
2.8. DİKEY MATRİSLER	12
2.9. DİK İZDÜŞÜM	13
3. ÇOK DEĞİŞKENLİ ANALİZ VE TEMEL BİLEŞENLER ANALİZİ	15
3.1. ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ	15
3.1.1. Çok Değişkenli İstatistiksel Analizin Tanımı.....	15
3.1.2. Çok Değişkenli İstatistiksel Analizin Kullanımı Ve Önemi.....	17
3.1.3. Çok Değişkenli İstatistiksel Analiz Tekniklerinde Temel Bileşenler Analizinin Yeri.....	19
3.2. TEMEL BİLEŞENLER ANALİZİNE GİRİŞ.....	20
3.2.1. Temel Bileşenler Analizinin Tarihsel Gelişimi	20

3.2.2. Temel Bileşenler Analizinin Tanımı Ve Mahiyeti	22
3.2.3. Temel Bileşenler Analizinin Özellikleri.....	25
3.2.4. Temel Bileşenler Analizinin Gerekliliği Ve Sağladığı Yararları.....	27
3.2.4.1. TBA'nın Gerekliliği	28
3.2.4.2. TBA'nın Yararları.....	28
4. TEMEL BİLEŞENLER ANALİZİNİN TEORİK HESAPLARI	30
4.1. TEMEL BİLEŞENLER ANALİZİNE GEOMETRİK YAKLAŞIM.....	30
4.2. TEMEL BİLEŞENLER ANALİZİNE ANALİTİK YAKLAŞIM	32
4.2.1. Temel Bileşenlerin Hesaplanması	33
4.2.2. Temel bileşenlerin varyans-kovaryans matrisinden elde edilmesi	38
4.2.3. Temel bileşenlerin korelasyon (standartlaştırılmış değişkenlerden) matrisinden elde edilmesi	43
4.2.4. Analizde Temel Bileşenlerin Seçimi ve Sayısının Belirlenmesi Yöntemleri.....	44
4.2.4.1. Açıklanan Varyans Kriteri (Kümülatif Varyans Oranı)	44
4.2.4.2. Kaiser Kuralı (Özdeğerlerin Büyüklüğü)	45
4.2.4.3. Scree Plot (Cattell Scree Test)	45
4.2.4.4. Joliffe Kriteri.....	47
4.2.5. Analizde Örneklem Büyüklüğünün Yeterliliği	47
4.3. TEMEL BİLEŞENLER ANALİZİ UYGULAMA ADIMLARI.....	49
4.3.1. Veri Setinin Hazırlanması.....	49
4.3.2. Ortalamadan Çıkarma (Veri Standardizasyonu).....	49
4.3.3. Kovaryans Matrisinin Hesaplanması	50
4.3.4. Özdeğerler Ve Özvektörlerin Hesaplanması	50
4.3.5. Bileşenlerin Seçilmesi Ve Özellik Vektörünün Oluşturulması.....	51
4.3.6. Yeni Veri Setinin Türetilmesi.....	51
4.3.7. Sonuçların Değerlendirilmesi ve Yorumlanması.....	51
5. REYTING SİSTEMİNİN TEKNİK VE SOSYOLOJİK ANALİZİ.....	52
5.1. REYTING	52
5.2. REYTING ÖLÇÜM SİSTEMİ VE TÜRKİYEDEKİ REYTING SİSTEMİ	52
5.3. TİAK VERİLERİNİN YAPISI VE ANLAMI.....	53
5.4. REYTINGİN MEDYA İÇERİKLERİ VE TOPLUMSAL NORMLAR ÜZERİNDEKİ ETKİSİ	54
6. UYGULAMA: TV KANALI REYTING VERİLERİ ÜZERİNDE TEMEL BİLEŞENLER ANALİZİ (TBA) UYGULAMASI.....	56

6.1. ARAŞTIRMANIN METODOLOJİSİ	56
6.1.1. Araştırmanın Konusu	56
6.1.2. Araştırmanın Amacı.....	56
6.1.3. Araştırmanın Problemi	57
6.1.4. Araştırmanın Yöntemi	57
6.1.5. Veri Toplama Yöntemi.....	58
6.1.6. Örneklem Seçimi	58
6.1.7. Veri Toplama Aracı.....	58
6.1.8. Örneklem Büyüklüğü	59
6.2. VERİ SETİ VE ÖZELLİKLER.....	59
6.2.1. Zaman Dilimleri (Timebands):	59
6.2.2. Demografik Gruplar:	60
6.2.3. Kanal Adları ve Programlar:	60
6.2.4. Temel Bileşenler Analizi İçin Seçilen Değişkenler:	61
6.3. TEMEL BİLEŞENLER ANALİZİ İÇİN VERİ HAZIRLAMA SÜRECİ	61
6.3.1. Verinin Hazırlanması.....	61
6.3.2. Verinin Yüklenmesi ve Temizlenmesi.....	62
6.3.3. Veri Standardizasyonu.....	62
6.3.4. TBA Uygulama	63
6.3.5. Kovaryans Matrisi Hesaplanması.....	63
6.3.6. Özdeğerler ve Özvektörlerin Hesaplanması	65
6.3.7. Özdeğerleri Sıralama	66
6.3.8. Açıklanan Varyans ve Kümülatif Varyans	67
6.3.9.Yeni Veri Seti Oluşturma (Veriyi Dönüştürme) (Manuel)	69
6.3.10.TBA Sonucunun Analizi ve Yorum	75
6.3.10.1. Bileşenlerin Yorumlanması	75
6.3.10.2. Kanalların Konumlarına Göre Yorumlar.....	76
6.3.10.3. Stratejik Çıkarımlar.....	76
6.3.10.4. AB ve 20+ABC1 Grubunun Karşılaştırılması	77
7. SONUÇ VE ÖNERİLER	79
8. KAYNAKLAR.....	82
9. EKLER	86

SİMGE VE KISALTMA LİSTESİ

Simgeler	Açıklama
$\bar{A}$	A'nın ortalaması
$\ \cdot\ $	Öklid Normu
$var()$	Varyans
$cov()$	Kovaryans
Σ	Kovaryans matrisi
∂	Türev operatörü
Σ	Toplam sembolü
$\cong$	Yaklaşık eşitlik
μ	Popülasyon ortalaması
λ	Lagrange çarpanı
Φ	Lagrange fonksiyonu
$\oplus$	Direkt toplam
$tr(A)$	A matrisinin izi
A^T	A matrisinin tranzpozesi
ρ	Korelasyon katsayısı

Kısaltmalar	Açıklama
diag	Diagonal (köşegen matris)
TBA, PCA	Temel Bileşenler Analizi
TDA, SVD	Tekil Değer Ayrışımı
RTG	Reyting (İzlenme Oranı)

ŞEKİLLER LİSTESİ

Şekil 4-1 :Üç değişkenli bir veri matrisinin geometrik gösterimi	30
Şekil 4-2 :Koordinat sistemi üzerinde gözlem verilerinin dağılımı.....	30
Şekil 4-3: Ortalanmış verilerin koordinat sistemi üzerindeki dağılımları	31
Şekil 4-4 :Örnek bir gözlem verisinin Birinci Temel Bileşen (PC1) üzerindeki yansıması ..	31
Şekil 4-5: Temel Bileşenlerin geometrik gösterimi.....	32
Şekil 6-1:Veri yükleme Kodları.....	62
Şekil 6-2:Veri Standartlaştırma Kodları.....	63
Şekil 6-3:PCA Uygulama Kodları.....	63
Şekil 6-4:Kovaryans Matrisi Oluşturma Kodları.....	63
Şekil 6-5: Python kodları ile Kovaryans Matrisi Isı Haritası.....	64
Şekil 6-6:Özdeğerler ve Özvektörler Kodları	65
Şekil 6-7:Özdeğerleri Sıralama Kodları.....	67
Şekil 6-8:Python Kodları ile Özdeğerlerin Sıralanmış Gösterimi	67
Şekil 6 9:Açıklanan Varyans ve Kümülatif Varyans Kodları	69
Şekil 6-10:Python Kodları ile Açıklanan Varyans ve Kümülatif Varyans Grafiği	69
Şekil 6-11:Python Kodları ile Dönüştürülmüş Veri	71
Şekil 6-12: Dönüştürülmüş Veriyi 2D İle Görselleştirme Kodları.....	72
Şekil 6-13:Python Kodları ile PCA Bileşenlerini Görselleştirme.....	72
Şekil 6-14: PCA Bileşenlerinin 2D Renkli Gösterim Kodları.....	73
Şekil 6-15: PCA Bileşenlerinin 2D Renkli Gösterim Grafiği.....	74
Şekil 6-16: PCA Bileşenlerinin 2D Renkli Gösterim Grafiği (Özvektörler Dahil)	75

TABLO LİSTESİ

<u>Tablo</u>	<u>Sayfa</u>
Tablo 1:İki Boyutlu Veri Kovaryans Hesaplaması	9
Tablo 2: Kovaryans Matrisi	64
Tablo 3: Özdeğerler.....	65
Tablo 4: Özvektörler Matrisi.....	66
Tablo 5:Açıklanan Varyanslar ve Kümülatif Varyanslar	68



1.GİRİŞ

Veri analizi, işlenmemiş verilerin işlenmesi, düzenlenmesi ve yorumlanması sürecidir. Veri analizindeki amaç, verilerden değerli bilgiler almak, karar alma süreçlerini desteklemek ve belirlenen bir problemi veya hipotezi test etmektir. Veri analizi, bir dizi aşamadan oluşan sistematik bir süreçtir. İlk olarak analiz edilecek bilgiler toplanır. Bilgileri anket, deney, gözlem, mülakat, IoT cihazları veya veritabanları gibi çeşitli yöntem ve teknolojiler aracılığıyla toplanmaktadır. Toplanan verilerde eksiklik, hata ve tutarsızlık var ise temizleme aşamasında bu problemler düzeltilerek hatalı veriler kaldırılır. Verinin özelliklerini anlamak için istatistiksel özet, görselleştirme ve temel analizler uygulanır. Bu evrede verinin dağılımı, ortalaması, medyan, mod, varyans gibi istatistiksel işlemler hesaplanır. Veri analiz aşamasında ise istatistiksel yöntemler uygulanmaktadır. Örneğin; tanımlayıcı istatistikler, çıkarımsal istatistikler, regresyon analizi ve temel bileşenler analizi.

Temel bileşenler analizi kullanılarak yapılan analiz sonuçları hipoteze göre yorumlanmaktadır. Sonuçların anlamlılığı ve güvenilirliği üzerinde değerlendirme sağlanmaktadır. Verilerimizden elde ettiğimiz bulguları sunup raporlama işlemi yapılır. Bu sayede karar vericiler temel bileşenler analizi uygulamasının çıktılarından yararlanırlar. Araştırmacıların ve işletmelerin bilinçli karar vermelerini sağlar. Ürün geliştirme, pazarlama stratejileri ve operasyonel iyileştirmeler gibi alanlarda daha etkili stratejilerin belirlenmesine yardımcı olur. İş dünyasında; müşteri davranışlarını analiz etme, satış tahminleri yapma, sağlıkta; hastalık yayılımını inceleme, tedavi etkilerini değerlendirme, sosyal bilimlerde; toplumsal eğilimleri belirleme, finans alanında ise yatırım kararlarını destekleme ve piyasa verilerini incelemek için olanak sağlar.

Temel bileşenler analizi (TBA), çok boyutlu veri setlerinin boyutunu azaltarak verilerin daha anlaşılır hale getirilmesine yardımcı olan istatistiksel bir tekniktir. Yüksek boyutlu verilerde kalıpları bulmak zor olabilir; bu durum, grafiksel temsil imkanlarının sınırlı olduğu durumlarda daha da belirginleşir. Bu kalıplar benzer özellik gösteren eğilimler veya ilişkiler olabilir. Bir veri kümesinde belirli bir davranışın, değişkenin veya olgunun nasıl ortaya çıktığını, birbirleriyle nasıl etkileşimde bulunduğunu gösteren belirgin yapılarıdır. Veriyi anlamak için bize ipuçları sunmaktadır. TBA, veriyi analiz etmek için güçlü bir anahtardır.

TBA, yüksek boyutlu veri setlerinde gizli kalıpları ve ilişkileri ortaya çıkarmak için etkili bir boyut azaltma yöntemidir. Bu yöntem, verinin daha anlaşılır hale gelmesine yardımcı olur ve karmaşık verileri daha basit bir şekilde temsil etme imkânı sunar. Ayrıca, önemli bilgileri

koruyarak gereksiz gürültüyü azaltır; bu da daha doğru analizler ve sınıflandırmalar yapmaya olanak tanır. Bu nedenle TBA, veri analizi ve makine öğrenimi süreçlerinde sıklıkla tercih edilen bir tekniktir. Bu teknik, orijinal değişkenlerin lineer kombinasyonları olarak yeni değişkenler oluşturur. Oluşan bu yeni değişkenler "birincil bileşenler" (principal components) olarak adlandırılır ve seçilirken birbirleriyle korelasyonsuz olmalarına dikkat edilir. Bu sayede, her bir birincil bileşen, verinin farklı bir yönünü temsil eder ve veri setindeki değişkenlerin toplam varyansını en iyi şekilde açıklamaya çalışır. Bu özellik, TBA'nın veri analizi ve boyut azaltma süreçlerindeki etkisini artırır.

Birincil bileşenler arasındaki ilk bileşen, verideki maksimum varyansı açıklar ve verideki en büyük değişkenliği anlatır. İkinci bileşen ise ilk bileşen tarafından henüz açıklanmamış maksimum varyansı açıklar. Her yeni bileşen, önceki bileşenlerden bağımsız olacak şekilde seçilir ve bu şekilde devam eder. Oluşturulabilecek yeni değişkenlerin maksimum sayısı, orijinal değişkenlerin sayısına eşittir. Ancak genellikle hepsi gerekli değildir ve seçilmiş bir sayıda birincil bileşen kullanılır. Sadece ilk birkaç bileşen, verinin büyük kısmındaki varyansı yeterince iyi açıklayabilir. Bu nedenle, yalnızca en fazla varyansı açıklayan bileşenler seçilir. Bu seçim, modelin basitleştirilmesi ve analiz sürecinin hızlandırılması açısından önemlidir. Çünkü fazla değişken, modelin karmaşıklığını artırabilir ve gereksiz gürültü yaratabilir. TBA sayesinde, temel bilgiyi koruyarak daha yönetilebilir bir veri seti elde edilir. Temel Bileşenler Analizinin üç önemli noktası; varyansın maksimizasyonu, boyut azaltma ve korelasyon olarak ifade edilebilir (Bishop, 2006).

Büyük veri setleri, birçok disiplinin araştırma ve uygulama alanlarında giderek artan bir öneme sahiptir. Bu veri setlerinin etkin bir şekilde analiz edilebilmesi için, boyutlarının anlamlı ve önemli ölçüde azaltılmasını sağlayan yöntemlere ihtiyaç duyulmaktadır. Bu sayede, verilerdeki bilginin büyük bir kısmı korunarak, anlamlı yorumlar yapılabilmektedir. Bu amaç doğrultusunda çeşitli yöntemler geliştirilmiş olsa da, temel bileşenler analizi (TBA), köklü bir geçmişe sahip ve en yaygın kullanılan tekniklerden biri olarak dikkat çekmektedir. TBA'nın temel prensibi, bir veri setinin boyutunu azaltırken, mümkün olan en fazla 'değişkenliği' (istatistiksel bilgiyi) koruyarak veri setinin yapısını yansıtmayı sağlamaktır. Temel bileşenler analizi (TBA) ile ilgili en eski literatürler incelendiğinde bu süreç 1900'lü yıllara dayanır. TBA ile ilgili en eski literatür Pearson ve Hotelling tarafından yazılmış ancak elektronik bilgisayarların yaygın olarak kullanılabilir hale gelmesiyle birlikte, bu yöntemin trivial derecede küçük olmayan veri setlerinde uygulanabilirliği artmıştır (Jolliffe ve Cadima, 2016).

Ian T. Jolliffe ve Jorge Cadima'ya göre (Jolliffe ve Cadima, 2016), temel bileşenler analizi ile ilgili en eski makale, Karl Pearson tarafından yazılan "On Lines and Planes of Closest Fit to

Systems of Points in Space"dir. Bu makalede Pearson, çok deęişkenli verilerde en yakın uyum çizgileri ve düzlemleri hakkında, temel kavramları ortaya koymuş ve TBA'nın temelini oluşturmuştur. Pearson'ın çalışması, istatistik ve veri analizi alanında önemli bir dönüm noktası haline gelmiştir (Person, 1901).

Charles Spearman, 1904 yılında yayınladığı "General Intelligence" başlıklı makalesinde, ilk defa faktör analizi yöntemini geliştirmiştir. Bir dizi ölçümün hepsinin ortak olarak sahip olduğu unsurların, diğer tüm ölçümlerde de ortak olarak bulunacağını belirtmiştir. Devam eden yıllarda Spearman, 1910 yılında "The Correlation between Mental and Physical Tests" adlı makalesinde, zihinsel ve fiziksel testler arasındaki ilişkiyi incelemiştir. Bu çalışmada, farklı testlerden elde edilen sonuçların birbirleriyle ne kadar ilişkili olduğunu araştırarak, genel zekâ kavramını destekleyen bir temel oluşturmuştur. Ayrıca, bu makale, faktör analizi ve korelasyon gibi istatistiksel yöntemlerin uygulanmasının erken örneklerini sunmaktadır. Hotelling ise 1933 yılında "Analysis of a Complex of Statistical Variables into Principal Components" adlı çalışmasında çok deęişkenli istatistikte temel bileşenler analizi yöntemini sistematik bir şekilde tanımlamıştır. Verilerdeki varyansı maksimize eden temel bileşenleri belirlemeye yönelik çalışmalar yapmıştır (Spearman et al., 1904).

1970li yıllara gelindiğinde bilgisayar teknolojisinin gelişimi ile büyük ve geniş verilerde TBA uygulanmaya başlanmıştır. Sosyal bilimler alanında Raymond Cattell, Hans Eysenck, McCauley, David Funder araştırmacıları birey davranışlarını kişilik özelliklerini anlamak için TBA yöntemini kullanmışlardır (Cattel et al., 1949).

David A. Aaker, Philip Kotler, D. M. McDonald ve I. Dunbar, temel bileşenler analizi (TBA) kullanarak pazarlama stratejileri ve müşteri segmentasyonu üzerine kapsamlı analizler gerçekleştirmişlerdir. Bu araştırmalarda, TBA'nın müşteri verilerinin daha iyi anlaşılmasına ve pazarlama kararlarının daha etkili bir şekilde alınmasına olanak tanıdığı ortaya konmuştur (Aaker et al., 1996).

Genetik veriler alanında ise I. T. Jolliffe, Joe C. Tsai ve arkadaşları, TBA'yı kullanarak gen ekspresyonu ve diğer biyolojik veri setlerini inceleyen araştırmalar yapmışlardır. Bu çalışmalar, biyolojik verilerdeki gizli yapıların keşfine önemli katkılarda bulunmuştur (Jolliffe et al., 2002).

Finansal alanda, Eugene F. Fama, João Nobre, Rui Ferreira Neves ve Andrew Lo, TBA'nın risk ve getiri analizini geliştirmeye yönelik önemli katkılarda bulunmuşlardır. Bu araştırmalar, yatırım kararlarının daha bilinçli bir şekilde alınmasını sağlamıştır (Fama et al., 1970).

Makine öğrenimi alanında, Ian Goodfellow, TBA'yı kullanarak veri setlerinin boyutunu azaltma ve model performansını artırma üzerine çalışmalar gerçekleştirmiştir. Ayrıca, Yoshua

Bengio, derin öğrenme ve sinir ağıları konularında TBA'nın rolünü incelemiştir. Geoffrey Hinton ise TBA'nın veri setlerinin boyutunu azaltarak model performansını artırma potansiyelini vurgulamış ve bu bağlamda TBA'nın daha etkili ve verimli makine öğrenimi modellerinin geliştirilmesine katkıda bulunduğunu belirtmiştir (Goodfellow et al., 2016).

Temel Bileşenler Analizi (TBA), verilerin daha iyi anlaşılmasına ve analiz edilmesine olanak tanıyan önemli bir yöntemdir. Farklı disiplinlerdeki uygulamaları sayesinde, TBA'nın etkisi oldukça geniştir. Geçmişten günümüze, bu yöntemin gelişimi, istatistik ve veri bilimi alanlarındaki ilerlemelerle paralel bir seyir izlemiştir. Bu durum, TBA'nın çok yönlülüğünü ve geçerliliğini açıkça ortaya koymaktadır. Veri setlerinin boyutunu azaltırken, temel yapısal özellikleri ortaya çıkarması, bu yöntemi araştırmalarda vazgeçilmez bir araç haline getirmektedir. Bu nedenle, TBA'nın teorik temelleri ve pratik uygulamaları, günümüz veri odaklı dünyasında giderek daha fazla önem kazanmaktadır.

2. MATEMATİKSEL TEMEL

2.1. İSTATİSTİK

İstatistik, belirli bir amacı gerçekleştirmek üzere verilerin toplanması, düzenlenmesi, analizi ve yorumlanması sürecini kapsayan bir bilim dalıdır. Bu süreç, verilerin toplanmasından başlayarak, uygun bir şekilde sınıflandırılması ve özetlenmesiyle devam eder; ardından elde edilen bulgular tablo ve grafiklerle görselleştirilir. Verilerin güvenilirliği ve geçerliliği, sonuçların güven derecelerini belirleyen istatistiksel ölçütlerle açıklanır. Ayrıca, örneklem verilerinden elde edilen bulgular, belirli bir kitleye genellenebilir. İstatistiksel analizler, farklı değişkenler arasındaki ilişkileri inceleyerek, bu ilişkilerden geleceğe yönelik tahminlerde bulunmayı mümkün kılar. Deneysel düzenekler ve gözlemler aracılığıyla istatistiksel sonuçlar elde edilerek, bu bulguların genelleme ve doğrulama aşamaları gerçekleştirilir. İstatistiğin temel ilkesi, büyük veri kümelerinin daha küçük, anlamlı alt kümelere ayrılarak analiz edilmesi ve bu sayede daha büyük bir veri seti üzerinden anlamlı sonuçlar çıkarılmasıdır. Bu yönüyle, istatistik veri odaklı karar alma süreçlerini yönlendiren ve geniş çaplı analizler yapmayı mümkün kılan bir disiplindir (Cömert, 2015).

2.2. STANDART SAPMA

Standart sapma, bir veri kümesindeki bireysel gözlemlerin ortalamadan ne kadar sapma gösterdiğini ve dolayısıyla verilerin dağılımının genişliğini ölçen bir istatistiksel parametredir. Verilerin yayılımını özetlemek amacıyla kullanılan bu ölçüt, varyansın karekökü olarak tanımlanır. Varyans, her bir gözlemin ortalamadan olan farkının karesinin ortalaması iken, standart sapma, bu değerın karekökü alınarak elde edilir ve bu şekilde aynı birimlerde, daha anlaşılır bir ölçüm sağlar. Bir veri kümesinin ortalamasını hesaplamak için kullanılan formül şu şekildedir:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

Burada $\bar{X}$ veri kümesinin ortalamasını ifade eder, x_i her bir veri gözlemini ve n ise veri kümesindeki gözlem sayısını temsil eder. Standart sapma ise şu formülle hesaplanır:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - X)^2}$$

Bu formülde x_i , her bir gözlem değerini; X , veri kümesinin ortalamasını; n ise toplam gözlem sayısını belirtir. Her bir gözlemin ortalamadan farkının karesi alınarak toplanır, elde edilen toplam $n - 1$ 'e bölünür ve sonuç olarak karekökü alınır.

Standart sapma, verilerin değişkenliğini ve dağılımını anlamada yaygın olarak kullanılan bir ölçüttür. Küçük bir standart sapma, gözlemlerin ortalamaya yakın olduğunu, dolayısıyla verinin dar bir aralıkta toplandığını gösterirken, büyük bir standart sapma, verinin geniş bir aralığa yayıldığını ve daha fazla değişkenlik içerdiğini ifade eder. Bu nedenle, standart sapma, veri setlerinin içsel özelliklerini değerlendirmek ve farklı veri kümeleri arasındaki değişkenlikleri karşılaştırmak için önemli bir araçtır.

2.3. VARYANS

Varyans, bir veri kümesinin dağılımını ölçmek amacıyla kullanılan istatistiksel bir kavramdır ve bir veri kümesindeki değerlerin, ortalamadan ne kadar uzaklaştığını nicel olarak ifade eder. Varyans, genellikle bir veri setindeki değişkenliği ya da yayılmayı anlamak için başvuru olan önemli bir ölçümdür. Verilerin yayılımı, varyansın büyüklüğüyle doğru orantılıdır; yani varyansın yüksek olduğu bir veri kümesi, daha geniş bir dağılıma ve daha fazla çeşitliliğe sahipken, düşük varyans, daha dar bir dağılımı ve az değişkenliği işaret eder.

Varyans, her bir gözlemin, veri kümesinin ortalamasına olan farkının karesinin ortalamasını alarak hesaplanır. Bu hesaplama, uç değerlerin (outliers) etkisini de göz önünde bulundurur çünkü farkların kareleri alındığında bu uç değerlerin varyans üzerindeki etkisi artar. Varyansın matematiksel tanımı, popülasyon verisi ve örneklem verisi için farklı formüllerle yapılır.

2.3.1. Popülasyon Varyansı Formülü

Bir popülasyonun varyansı hesaplanırken, formül şu şekilde tanımlanır:

$$\sigma = \frac{1}{N} \sum_{j=1}^N (x_i - \mu)^2$$

Burada:

- σ^2 : Popülasyon varyansı,
- N : Veri kümesindeki toplam gözlem sayısı,
- x_i : Her bir veri noktasının değeri,
- μ : Popülasyonun ortalaması (ya da beklenen değeri).

Bu formülde, x_i her bir gözlemin değerini temsil ederken, μ popülasyonun ortalamasını ifade etmektedir. Varyans hesaplanırken, her bir gözlemin ortalamadan ne kadar saptığı, yani $(x_i - \mu)$ farkı alınır ve bu farkların kareleri hesaplanır. Tüm bu kare farkların toplamı, veri kümesindeki toplam gözlem sayısı N 'ye bölünerek varyans elde edilir. Bu işlem, veri setindeki yayılmayı nicel olarak ölçmek amacıyla yapılır.

2.3.2. Örneklem Varyansı Formülü

Eğer veriler bir örneklemde alınmışsa ve bu örneklem üzerinden popülasyon varyansı hakkında bir tahminde bulunulmak isteniyorsa, varyansın hesaplanması için kullanılan formül şu şekilde değişir:

$$s^2 = \frac{1}{n-1} \sum_{j=1}^N (x_i - \bar{x})^2$$

Burada:

- s^2 : Örneklem varyansı,
- n : Örneklemdeki gözlem sayısı,
- $\bar{x}$: Örneklem ortalaması.

Bu formülde, örneklemdeki her bir veri noktası ile örneklem ortalaması arasındaki farkların kareleri alınır ve bu karelerin toplamı hesaplanır. Toplam kare farklar, örneklemdeki gözlem sayısının bir eksiğine (yani $n - 1$) bölünerek örneklem varyansı elde edilir. Bu düzeltme, örneklem varyansının popülasyon varyansına daha doğru bir tahmin sunmasını sağlar. Örneklem varyansı, genellikle $n - 1$ ile bölünerek hesaplanır çünkü bu, örneklemde popülasyona yapılan genellemelerin doğruluğunu artıran bir düzeltme faktörüdür.

Varyans, veri setinin değişkenliğini ölçerken önemli bir araçtır. Varyans değeri ne kadar büyükse, veriler arasındaki farklar o kadar belirgindir ve dağılım geniştir. Öte yandan, düşük varyans, verilerin ortalama etrafında daha sıkı bir şekilde toplandığını gösterir. Varyansın hesaplanmasında, verilerin uç noktalarındaki sapmaların karelerinin alınması, uç değerlerin

etkisinin daha fazla vurgulanmasını sağlar, ancak bu aynı zamanda varyansın yüksek olmasına neden olabilir.

2.4. KOVARYANS

Kovaryans, olasılık teorisi ve istatistikte, iki değişkenin birlikte nasıl değiştiğini ölçen bir kavramdır. İki rastgele değişken arasındaki doğrusal ilişkiyi inceler. Standart sapma ve varyans, tek boyutlu veriler için geçerli iken, kovaryans çok boyutlu veri setlerinde iki değişken arasındaki ilişkiyi ölçmek için kullanılır.

2.4.1. Varyans ve Kovaryans Formülleri

Varyans, bir değişkenin ortalamadan sapmalarının karelerinin ortalamasıdır ve şu şekilde hesaplanır:

$$var(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

İki değişken arasındaki kovaryans ise şu formülle hesaplanır:

$$cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

Burada X_i ve Y_i , sırasıyla X ve Y değişkenlerinin gözlem değerleridir; $\bar{X}$ ve $\bar{Y}$ ise bu değişkenlerin ortalamalarıdır.

Kovaryansın değeri, değişkenler arasındaki ilişkinin yönünü belirtir. Pozitif kovaryans, her iki değişkenin aynı yönde değiştiğini (biri arttıkça diğeri de artar), negatif kovaryans ise zıt yönde değiştiklerini (biri arttıkça diğeri azalır) gösterir. Kovaryans sıfır ise, değişkenler arasında doğrusal bir ilişki olmadığını, yani bağımsız olduklarını ifade eder.

2.5. KOVARYANS MATRİS

İkiden fazla değişkenin ilişkisini analiz etmek için kovaryans matrisi kullanılır. Bu matrisin diyagonal elemanları, her bir değişkenin varyans değerlerini temsil eder. Ayrıca, kovaryans

matrisinin simetrik olmasının nedeni, $cov(a, b) = cov(b, a)$ özelliğidir; yani matrisin simetrik yapısı, her iki değişken arasındaki kovaryansın aynı değeri taşımasını sağlar.

$$C = \begin{pmatrix} cov(x, x) & cov(x, y) & cov(x, z) \\ cov(y, x) & cov(y, y) & cov(y, z) \\ cov(z, x) & cov(z, y) & cov(z, z) \end{pmatrix}$$

Gözlem No	X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
1	2	3	$2 - 6.5$ $= -4.5$	$3 - 7.5$ $= -4.5$	$(-4.5) \times (-4.5)$ $= 20.25$
2	4	6	$4 - 6.5$ $= -2.5$	$6 - 7.5$ $= -1.5$	$(-2.5) \times (-1.5)$ $= 3.75$
3	6	9	$6 - 6.5$ $= -0.5$	$9 - 7.5 = 1.5$	$(-0.5) \times 1.5$ $= -0.75$
4	8	10	$8 - 6.5$ $= 1.5$	$10 - 7.5$ $= 2.5$	$1.5 \times 2.5 = 3.75$
5	7	5	$7 - 6.5$ $= 0.5$	$5 - 7.5$ $= -2.5$	$0.5 \times (-2.5)$ $= -1.25$
6	5	7	$5 - 6.5$ $= -1.5$	$7 - 7.5$ $= -0.5$	$(-1.5) \times (-0.5)$ $= 0.75$
7	9	8	$9 - 6.5$ $= 2.5$	$8 - 7.5 = 0.5$	$2.5 \times 0.5 = 1.25$
8	10	11	$10 - 6.5$ $= 3.5$	$11 - 7.5$ $= 3.5$	$3.5 \times 3.5 = 12.25$
9	3	4	$3 - 6.5$ $= -3.5$	$4 - 7.5$ $= -3.5$	$(-3.5) \times (-3.5)$ $= 12.25$
10	11	12	$11 - 6.5$ $= 4.5$	$2 - 7.5 = 4.5$	$4.5 \times 4.5 = 20.25$
Toplam					72.5

Tablo 1:İki Boyutlu Veri Kovaryans Hesaplaması

Kovaryansı hesaplayalım:

$$\text{cov}(X, Y) = \frac{72.5}{10 - 1} = \frac{72.5}{9} \approx 8.06$$

2.6. ÖZVEKTÖRLER VE ÖZDEĞERLER

İki matrisin çarpılabilmesi, yalnızca boyutlarının uyumlu olduğu durumlarda mümkündür ve özvektörler, bu çarpma işlemine ilişkin özel bir durumu ifade eder. Bir matrisin bir vektör üzerindeki etkisi, o vektörün hem büyüklüğünü hem de yönünü değiştirebilir. Ancak, bazı durumlarda, bir matris, belirli vektörler üzerinde yalnızca büyüklük değişikliği yapar ve bu vektörlerin doğrultuları değişmeden kalır. Bu tür vektörler, matrisin özvektörleri olarak tanımlanır. Özvektörler, matrisin doğrultuyu değiştirmeyen, yalnızca büyüklüğü üzerinde etkili olan vektörleridir. Özdeğerler, özvektörler ve özuzaylar, bir matrisin karakteristik özelliklerini ortaya koyarak matris hakkında önemli bilgiler sağlar.

Özvektörler yalnızca kare matrisler için tanımlanabilir, bu nedenle özdeğer ve özvektör hesaplamaları genellikle kovaryans matrisleri gibi kare matrisler üzerinde yapılır. Ancak, her kare matrisin mutlaka özvektörleri yoktur. Bu durum, matrisin belirli özelliklerine ve yapısına bağlı olarak değişebilir.

2.7. TEKİL DEĞER AYRIŞIMI

Tekil Değer Ayrışımı (TDA), matrislerin çarpanlara ayrılmasını sağlayan bir yöntemdir ve sayısız uygulama alanı vardır. Bu alanlar arasında Google'ın PageRank algoritması, insan yüzü modelleme, gen analizi, veri sıkıştırma ve boyut azaltma gibi çok çeşitli konular bulunmaktadır.

2.7.1. Temel Kavramlar

- Skalerler küçük harflerle yazılır (örneğin: a).
- Vektörler kalın küçük harflerle yazılır (örneğin: $\mathbf{a}$).
- Matrisler kalın büyük harflerle yazılır (örneğin: $\mathbf{A}$).

Bir matrisin herhangi bir sütunu, bir vektör olarak ifade edilir. Ayrıca bir matrisin herhangi bir elemanı a_{ij} ile gösterilir.

2.7.2. Normlar ve Öklit Formu

L_2 Normu (Öklit Normu), bir vektörün büyüklüğünü hesaplamak için kullanılan bir matematiksel araçtır. Öklit mesafesi olarak da bilinir ve bir vektörün uzunluğunu ölçen yaygın bir normdur. Örneğin, $a \in C^I$ vektörünün L_2 normu şu şekilde tanımlanır:

$$\|a\|_2 = \sqrt{|a_1|^2 + |a_2|^2 + \dots + |a_I|^2} = \sqrt{(a^H a)}$$

Burada H , matrisin Hermit (karmaşık transpoz) işlemini ifade eder. Spektral Norm ise bir matrisin özdeğerleri ile ilgilidir. $A \in C^{I \times J}$ matrisinin spektral normu, $A^H A$ matrisinin en büyük özdeğerinin kareköküne eşittir:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^H A)}$$

Burada, $\lambda_{\max}(A^H A)$, $(A^H A)$ matrisinin en büyük özdeğerini ifade eder.

2.7.3. Ortonormal ve Üniter Matrisler

Bir matrisin sütunlarının ortogonal olması, farklı sütunlar arasındaki iç çarpımların sıfır olması anlamına gelir. Eğer matrisin sütunlarının normları bir ise, bu matris birim ortonormal matris (orthonormal matrix) olarak adlandırılır. Bir kare matrisin hem satır hem de sütunları birbirine dik ve normlanmışsa, bu matris üniter (unitary) matris olarak tanımlanır. Üniter matrisler, $A^H A = I$ denkleğini sağlar.

Bir $A \in C^{I \times J}$ matrisinin tekil değer ayrışımı (SVD), aşağıdaki şekilde ifade edilir:

$$A = U \Sigma V^H$$

Burada $U \in C^{I \times I}$ ve $V \in C^{J \times J}$, üniter matrislerdir. $\Sigma \in C^{I \times J}$ pseudodiagonal (sözde köşegen) bir matristir ve köşegen elemanları $\sigma_1, \sigma_2, \dots, \sigma_{\min(I,J)}$ olarak sıralanır, burada $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$. Bu ayrışım, A matrisinin tekil değerleri olarak adlandırılan $\sigma_1, \sigma_2, \dots, \sigma_r$ değerlerini ve özvektörleri $u_1, u_2, \dots, u_r$ ve $v_1, v_2, \dots, v_r$ içerir. U matrisi, AA^H matrisinin özvektörlerinden oluşur. V matrisi ise $A^H A$ matrisinin özvektörlerinden oluşur. Eğer A matrisinin sıfır olmayan tekil değerlerinin sayısı r küçükse, A , daha düşük ranklı bir matrisle yaklaşık edilebilir:

$$A = U_1 D V_1^H$$

Burada U_1 ve V_1 , sırasıyla AA^H ve $A^H A$ matrislerinin ilk r özvektörlerinden oluşur. D ise AA^H veya $A^H A$ matrislerinin sıfırdan farklı özdeğerlerinin karekökleriyle oluşturulmuş köşegen bir matristir:

$$D = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_r}) = \begin{bmatrix} \sqrt{d_1} & \dots & \\ \vdots & \ddots & \vdots \\ & \dots & \sqrt{d_r} \end{bmatrix}$$

Burada $d_1, d_2, \dots, d_r$ sıfırdan farklı özdeğerlerdir. Tersine işlem olarak, yukarıda belirtilen tekil değer ayrışımı (SVD) gösteriminden başlanarak, matrisin her bir bileşeni yeniden yapılandırılabilir. Bu süreçte, D matrisi sıfır matrisleriyle genişletilecek ve U_1 ile V_1 matrisleri, ortogonal matrisler olacak şekilde genişletilecektir. Özellikle, D matrisinin sıfır olmayan elemanları, orijinal matrisin tekil değerlerini temsil ederken, sıfır olan elemanlar, tekil değerlerin eksik veya gereksiz olduğu durumları gösterir. Bu genişletilmiş gösterim, A matrisinin tekil değer ayrışımını elde etmenin ters yönüdür ve D matrisinin köşegen elemanları, A 'nın tekil değerleri olarak adlandırılır.

2.8. DİKEY MATRİSLER

Dikey matrisi ifade etmek için Denklem kullanılabilir:

$$A = U \Sigma V^H = \sum_{i=1}^n \sigma_i u_i v_i^H$$

Bu denklemde, A matrisinin n dereceden bir matris olduğu, $\sigma_i u_i v_i^H$ terimlerinin ise birbirine dik olan n adet matrisin doğrusal birleşimi (veya lineer kombinasyonu) şeklinde yazılabileceği ifade edilmektedir. Burada, her bir terim $\sigma_i u_i v_i^H$ bir tekil değer σ_i ile ilişkili olarak, u_i ve v_i^H matrislerinin iç çarpımını ifade eder. Özellikle, A matrisine en büyük katkıyı sağlayan terim, ilk tekil değeri σ_1 ile ilişkili olan $u_1 v_1^H$ matrisidir. Bunun nedeni, σ_1 'in A 'nın en büyük tekil değeri olması ve $u_1 v_1^H$ matrislerinin Frobenius normunun, toplam norm üzerinde belirgin bir etkiye sahip olmasıdır.

Bir diğer deyişle, A matrisini bir oyuncak yapı olarak düşündüğümüzde, bu oyuncak çeşitli legolardan (yani, $u_i v_i^H$ matrisleri) oluşur. Buradaki her bir lego parçası, $u_i v_i^H$, belirli bir oranda

(yani, σ_i) kullanılarak matrisin toplam yapısı oluşturulur. Bu oranlar, tekil değerler tarafından belirlenir ve her bir parça matrisin katkısı, norm açısından tekil değerlerin büyüklüğü ile doğru orantılıdır.

2.9. DİK İZDÜŞÜM

Verilen bir matris A , genellikle $R^{n \times m}$ boyutlarında olup, sütunları bir alt uzayı $[A]$ gerdirir. Bu durumda, bir vektör $v \in R^n$, A 'nın sütun vektörlerinin oluşturduğu uzaya projekte edilebilir. Bu projeksiyon işlemi, vektörün, alt uzaya dik izdüşümünü almak olarak yorumlanabilir. Dik izdüşüm (ortogonal projeksiyon), belirli bir alt uzaya ait olmayan bir vektörün, o uzaya olan en kısa mesafeye sahip projeksiyonunu ifade eder. Daha matematiksel bir ifadeyle, dik izdüşüm, bir vektörün, o alt uzaya paralel olmayan bileşeninin sıfır olduğu projeksiyonudur.

Bir matris A , $R^{n \times m}$ boyutlarında bir matris olduğunu varsayalım. A 'nın sütunlarının gerdiği alt uzaya $[A]$ üzerine dik izdüşüm, projeksiyon matrisi $P[A]$ ile ifade edilir. Bu projeksiyon matrisi, aşağıdaki şekilde tanımlanır:

$$P[A] = A(A^T A)^+ A^T$$

Burada, $(A^T A)^+$ ifadesi, $A^T A$ 'nın Moore-Penrose tersini temsil eder. Moore-Penrose tersinin tanımı, genellikle bir matrisin tersinin bulunamadığı durumlar için geçerlidir ve geliştirilmiş ters olarak da adlandırılabilir. Bu projeksiyon matrisi, A 'nın sütunlarının oluşturduğu alt uzaya dik izdüşüm matrisidir.

Bir matrisin sütun vektörlerinin oluşturduğu alt uzay $[A]$, belirli bir vektör uzayının bir alt kümesi olarak tanımlanabilir. Bu alt uzay üzerine yapılan dik izdüşüm, vektörlerin uzaya paralel olan bileşenlerini korurken, uzaya dik olan bileşenleri sıfıra indirger.

Verilen matris A 'nın sütunlarının oluşturduğu alt uzaya dik izdüşüm alındığında, bu işlemde tersine dik izdüşüm de önemlidir. Dik izdüşüm, bir vektörün bu alt uzaya olan projeksiyonunu bulur, aynı zamanda bu vektörün alt uzaya dik olan bileşenini ortadan kaldırır.

Bir matrisin sütunlarının oluşturduğu alt uzayın dik olan bileşeni $[A]^\perp$ olarak gösterilir. Bu alt uzay, R^n uzayını $[A]$ alt uzayı ve $[A]^\perp$ alt uzayı şeklinde iki alt uzaya ayırır. Yani:

$$R^n = [A] \oplus [A]^\perp$$

Burada $\oplus$ işareti, iki alt uzayın doğal direkt toplamını ifade eder. Yani R^n uzayındaki her bir vektör, $[A]$ ve $[A]^\perp$ alt uzaylarının vektörlerinin toplamı olarak yazılabilir.

Dik izdüşüm matrisinin $[A]^\perp$ alt uzayına yapılacak projeksiyonu, şu şekilde tanımlanır:

$$P = [A]^\perp = I - P[A]$$

Burada, I birim matrisini temsil eder. Yani, $P = [A]^\perp$, A 'nın sütunlarının oluşturduğu alt uzaya dik izdüşümü ifade eder.



3. ÇOK DEĞİŞKENLİ ANALİZ VE TEMEL BİLEŞENLER ANALİZİ

Bu bölümde, çok değişkenli analiz yöntemleri ve bunun alt dalı olan Temel Bileşenler Analizi konusu ele alınmaktadır. İlk bölümde, verilerin birden fazla değişkeni içeren durumlarla nasıl başa çıkılacağı, istatistiksel yöntemlerin çok boyutlu veri setlerinde nasıl kullanıldığına dair temel kavramlar açıklanmaktadır. İkinci bölümde ise, temel bileşenler analizi (PCA) üzerinde durulmakta, bu yöntemle yüksek boyutlu verilerin daha düşük boyutlu bir forma indirgenmesi ve ana bileşenlerin nasıl elde edileceği anlatılmaktadır.

3.1. ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

Bu başlık altında, çok değişkenli istatistiksel analiz yöntemlerinin temel kavramları ve uygulamaları ele alınmaktadır. Birden fazla değişkenin aynı anda incelendiği durumlarda, veri setlerinin nasıl analiz edileceği, ilişkilerin nasıl ortaya konacağı ve değişkenler arasındaki etkileşimlerin nasıl değerlendirileceği üzerinde durulmaktadır.

3.1.1. Çok Değişkenli İstatistiksel Analizin Tanımı

İstatistiksel analiz yöntemleri değişken boyutu ile tek değişkenli ve çok değişkenli olmak üzere iki grupta sınıflandırılırlar. Tek değişkenli istatistik analiz yöntemleri incelenen değişken dışındaki etkili/ ilişkili tüm faktörleri katılımcılar/denekler için türdeş kabul eder (Özdamar, 1999). Gerçekte ise incelenen değişken dışındaki faktörler bakımından deneklerin bir örnek veya türdeş olması çoğu zaman mümkün değildir. Bu durum ancak deneysel, kısmen de yarı deneysel çalışmalarda sağlanabilir. Ayrıca tek değişkenli istatistiksel analiz yöntemlerinin sahip olması gereken normallik, toplanabilirlik, varyansların homojenliği, kovaryansın sıfır olması gibi birçok varsayımları vardır (Yıldız, vd. 2012). Bu nedenlerle tek değişkenli analizler ile gerçek durumun aydınlatılması mümkün olmayabilir. Çok değişkenli istatistiksel analizler, araştırılan olay ve ilgili olduğu değişkenleri dikkate alarak veya çok sayıdaki değişkeni daha az sayıda doğrusal faktörlerine indirgeyerek, değişkenler arasındaki karmaşık ilişkileri incelemek ve çözümlere ulaşmak için geliştirilmiş yöntemler bütünüdür.¹ Ayrıca tek değişkenli analiz yöntemleri için varsayımların gerçekleşmemesi durumunda çok değişkenli analizler gerçekçi bir yaklaşımdır. Bu nedenle çok değişkenli istatistiksel analizler veri çözümlemede önemli bir yere sahiptir (Baydemir, 2021).

Çok değişkenli istatistiğin üzerinde fikir birliğine varılmış tek bir tanımından bahsetmek mümkün değildir. Tanımlamalar araştırma konusuna ve kullanılan değişkenlerin ölçüm düzeyine bağlı olarak değişiklik göstermektedir. Çok değişkenli istatistik yöntemle çok sayıda bağımlı ve bağımsız değişkenlerin eş zamanlı olarak analiz edilmesini sağlamaktadır. Daha genel ifade etmek gerekirse, iki veya ikiden daha fazla bağımlı değişken ihtiva edecek şekilde oluşturulan istatistik analiz tasarımı (şekli) "çok değişkenli istatistik analiz olarak kabul edilmektedir (Tabachnik & Fidell, 1989).

Çok değişkenli istatistiksel analiz, incelenen olayı ve buna etki eden içsel (endojen) ile dışsal (eksojen) faktörleri dikkate alarak, problemin doğasına ilişkin yapısal bilgilere dayalı bir şekilde (gerçek oluşum sürecine uygun olarak) inceleme yapmayı ve çözüm yolları geliştirmeyi amaçlayan bir yöntemler bütünüdür (Özdamar, 2004).

Çok değişkenli istatistiksel teknikler, sosyal bilimler (örneğin eğitim, psikoloji, sosyoloji), doğa bilimleri ve tıp alanlarındaki araştırmalar da dahil olmak üzere çeşitli alanlarda kullanılmaktadır. Bu tekniklerin kullanımı, araştırma tasarımlarının ve ilişkili araştırma sorularının giderek daha karmaşık hale gelmesi nedeniyle daha yaygın hale gelmiştir. Tek bir sonuç ölçütü üzerinde izole bir tedavi koşulunun etkisini incelemek genellikle gerçekçi değildir, özellikle sosyal bilimlerde, çünkü araştırma çalışmalarındaki katılımcılar neredeyse her zaman insanlardır.

Hepimizin bildiği gibi, insanlar, bilgi, inançlar, duygular, düşünceler, tutumlar vb. ile dolu karmaşık varlıklardır. İnsan katılımcıları, tek bir bağımsız değişken ve tek bir bağımlı değişkenle incelemek, gerçekten pratik değildir çünkü bu değişkenler insan zihninde ya da davranışlar bütününde yalnızca bir arada var olmazlar. Bu iki değişken, birçok başka değişken tarafından etkilenebilir veya onları etkileyebilir. Araştırmacı, ilgi duyulan fenomeni açıklığa kavuşturmak ve doğru açıklamalar sunmak için birden fazla değişkeni aynı anda incelemeye istekli olmalıdır.

Pituch ve Stevens (2016), araştırma çalışmalarında özellikle değişen tedavi yöntemlerinin (örneğin öğretim yöntemleri, danışmanlık teknikleri vb.) etkilerinin incelenmesiyle ilgili olarak, birden fazla sonuç ölçütü (yani bağımlı değişkenler) kullanmanın üç nedenini sunmaktadırlar:

1. Herhangi bir tedavi genellikle katılımcıları birden fazla şekilde etkiler. Sadece bir kriter ölçütünü incelemek çok sınırlıdır. Bir tedavi koşulunun etkilerini tam olarak anlamak için, araştırmacı, katılımcıların koşullara nasıl tepki verebileceğini farklı açılardan incelemelidir.
2. Birden fazla sonuç ölçütü kullanarak, araştırmacı araştırılan fenomenin daha tamamlayıcı ve ayrıntılı bir tanımını elde edebilir.
3. Tedavi uygulamak pahalı olabilir, ancak birkaç bağımlı değişkenin ölçülmesi (aynı araştırma içinde) genellikle çok düşük maliyetlidir ve araştırmacının bilgi kazancını maksimize etmesine olanak tanır.

Şunu belirtmek gerekir ki, çok değişkenli istatistiksel analiz için uygun bir çalışma genellikle birkaç bağımlı değişkeni olan (sadece bir bağımlı değişkeni olan tek değişkenli çalışmalara karşı) bir çalışma olarak tanımlanır (Mertler, vd. 2021).

3.1.2. Çok Değişkenli İstatistiksel Analizin Kullanımı Ve Önemi

Çok değişkenli analizler için veriler sayısal olmalıdır. Nicel değişkenler, belirli değişkenlerdeki puanlardan oluşur. Kategorik değişkenler için ise, araştırmacı tarafından atanan kodlar kullanılır. Örneğin, okul konumu değişkeni için, kentsel okullara 1, banliyö okullarına 2, kırsal okullara ise 3 atanabilir.

Veriler, çok değişkenli teknikler kullanılarak analiz için birçok farklı formatta sunulabilir. Çoğu zaman, bir veri matrisi analiz edilir. Veri matrisi, ham puanların veya verilerin düzenlendiği bir yapıdır; burada satırlar katılımcıları veya vakaları, sütunlar ise değişkenleri temsil eder. Verilerin analiz için bir başka olası formatı korelasyon matrisidir.

Korelasyon matrisi, her satır ve her sütunun farklı bir değişkeni temsil ettiği ve kesişen hücrelerin iki değişken arasındaki korelasyon katsayısını içerdiği kare, simetrik bir matristir. Üçüncü bir seçenek ise, varyans-kovaryans matrisidir; bu da kare, simetrik bir matristir ve ana diyagonaldeki (yani bir değişkenin kendisiyle kesişen) elemanlar her değişkenin varyansını, diyagonal dışındaki elemanlar ise değişkenler arasındaki kovaryansları temsil eder. Son olarak, kareler toplamı ve çapraz çarpanlar matrisidir ki bu, varyans-kovaryans matrisinin öncüsüdür. Özellikle, henüz ortalama alınmamış sapma değerlerinden oluşan bir matristir.

Çok değişkenli istatistiksel analizlerde yer alan matematiksel hesaplamalar, daha önce bahsedilen herhangi bir matris üzerinde yapılabilir. Ancak, bu hesaplamalar matris cebiri adı verilen işlem ile yapılır. Matris cebiri, matematiksel işlemlerin sıralı bir sayısal değer dizisi üzerinde yapıldığı skaler cebirin bir uzantısıdır (Mertler, vd. 2021).

Çok değişkenli istatistiksel analizler, karmaşık işlemleri içermekle birlikte, günümüzde bilgisayar yazılımlarının sağladığı olanaklarla yaygın ve etkili bir şekilde uygulanmaktadır. Bu analiz yöntemleri, hemen hemen tüm disiplinlerde kullanılmakla birlikte, özellikle sağlık bilimleri, psikoloji, biyoloji, sosyoloji, ekonomi, eğitim, mühendislik, tarım ve turizm gibi alanlarda geniş bir kullanım alanına sahiptir (Akbulut ve Çapık, 2022).

Çok değişkenli bir analiz yapmanın amacı, başlangıçta her birinin eşit derecede önemli kabul edildiği, birkaç ilişkili değişkeni aynı anda dikkate almaktır (Manly, 2018).

Çok değişkenli istatistik, birden fazla değişkenin eş zamanlı olarak analiz edilmesine olanak tanıyan bir yöntem olduğundan, çeşitli alanlarda farklı amaçlarla kullanılmaktadır. Bu amaçlar şu şekilde sıralanabilir:

Veri Boyutunun İndirgenmesi: Çok sayıda değişkenin bulunduğu veri setlerinde, veri boyutunu küçültmek ve önemli bilgileri daha kolay analiz edilebilir bir şekilde sunmak için kullanılır. Örneğin, Principal Component Analysis (PCA) gibi tekniklerle boyut indirgeme yapılabilir.

İlişki ve Etkileşimlerin İncelenmesi: Değişkenler arasındaki ilişkilerin ve etkileşimlerin belirlenmesi amacıyla kullanılır. Bu, özellikle çoklu regresyon, MANOVA (Multivariate Analysis of Variance) veya faktör analizi gibi tekniklerle yapılır.

Veri Kümesinin Gruplandırılması: Değişkenler arasındaki benzerlikleri analiz ederek, veriyi gruplamak amacıyla kullanılır. Örneğin, kümeleme analizi (cluster analysis) ile benzer özelliklere sahip örnekler bir araya getirilir.

Modelleme ve Tahmin Yapma: Çoklu değişkenlerin bir arada incelenmesiyle daha doğru tahminler yapma amacı güder. Bu amaçla, çoklu regresyon, lojistik regresyon ve yapısal eşitlik modelleme (SEM) gibi yöntemler kullanılır.

Anlamlı Farkların Belirlenmesi: Çoklu değişkenler arasında anlamlı farkların olup olmadığını incelemek için kullanılır. MANOVA ve MANCOVA (Multivariate Analysis of Covariance) gibi analizler, farklı gruplar arasındaki çoklu sonuçları test etmek için yaygın olarak kullanılır.

Veri Setlerinde Anomalilerin Tespiti: Çok değişkenli teknikler, veri setlerindeki sıra dışı gözlemleri (outlier) tespit etmek için de kullanılabilir. Bu, özellikle veri temizleme aşamalarında önemli bir adımdır.

Daha Karmaşık Yapıların Modellenmesi: İnsan davranışları veya biyolojik süreçler gibi daha karmaşık yapıları anlamak için çok değişkenli istatistiksel yöntemler kullanılır. Örneğin, faktör analizi, gizli değişkenlerin veya faktörlerin ortaya konmasına yardımcı olabilir.

Çok değişkenli istatistiksel yöntemlerin uygulama alanı oldukça geniştir. Sağlık bilimlerinden (tıp, diş hekimliği, eczacılık) mühendislik, eğitim, psikoloji ve ekonomi gibi pek çok alanda sıklıkla uygulanmaktadır. (Johnson ve Dean, 2007).

3.1.3. Çok Değişkenli İstatistiksel Analiz Tekniklerinde Temel Bileşenler Analizinin Yeri

Çok değişkenli istatistiksel analiz yöntemleri, verilerdeki karmaşık ilişkileri anlamak için çeşitli teknikler sunar. Bu yöntemler arasında Çok Değişkenli Varyans Analizi (MANOVA), Çok Değişkenli Regresyon Analizi, Çok Değişkenli Lojistik Regresyon Analizi, Setler Arası Korelasyon Analizi, Temel Bileşenler Analizi (TBA), Faktör Analizi, Kümeleme Analizi, Ayırma Analizi, Çok Boyutlu Ölçekleme ve Yapısal Eşitlik Modelleri (SEM) yer alır. Bu teknikler, verilerin boyutlarını indirgemek, ilişkileri modellemek ve gruplar arasındaki farkları analiz etmek için kullanılır. Özellikle, büyük veri setlerinde boyut indirgeme ve ilişki modelleme için TBA ve SEM gibi yöntemler sıklıkla tercih edilmektedir (Akbulut ve Çapık, 2022).

Çok değişkenli bir veri setinde, p adet değişken arasındaki ilişkiler, değişkenlerin bağımlı ve bağımsız olarak ayrılabilirdiği durumlarda regresyon analizleri ile incelenebilir. Ancak, bu tür bir ayrım yapılamadığında ve değişkenler arasındaki ortak değişimlerin yorumlanması gerektiğinde boyut indirgeme yöntemlerine başvurulmaktadır. Bu yöntemlerden biri olan Temel Bileşenler Analizi (TBA), değişkenler arasındaki ilişkilerin incelenmesinde etkili bir araçtır.

Çok değişkenli istatistiksel analizlerde, n sayıda birey (veya nesne) üzerinden p sayıda değişken (özellik) incelenmektedir. Bu değişkenlerin çoğu birbirini etkileyen (bağımlı) olup, p sayısının büyük olması, analiz sürecinde bazı güçlükler yaratmaktadır. Örneğin, bir öğrencinin akademik başarı düzeyine ilişkin olarak sınav puanları, derse katılım oranı, ödev notları, ders dışı etkinliklere katılım, uyku süresi, çalışma alışkanlıkları, beslenme düzeni gibi çok sayıda değişken bulunmaktadır. Bu değişkenlerin bir kısmı birbirleriyle yüksek derecede ilişkilidir, bu da değişkenlerin bağımsızlık varsayımını ihlal etmektedir. Ayrıca, çok sayıda değişkenle yapılan analizlerde işlem yükü artmakta ve bu durum, sonuçların yorumlanmasında önemli zorluklar ortaya koymaktadır. Günümüzde bilgisayar teknolojilerinin gelişmesiyle işlem yükü bir ölçüde hafiflese de çok sayıda değişkeni içeren analizlerin sonuçlarının anlamlı bir şekilde yorumlanması ve özetlenmesi hala oldukça zorlayıcı olabilir. Bu bağlamda, başvurulacak önemli yöntemlerden biri, Temel Bileşenler Analizi (Principal Component Analysis, PCA) olup, bu analiz, değişkenler arasındaki bağımlılığı ortadan kaldırmayı veya boyut indirgeme sağlamayı amaçlamaktadır. Temel Bileşenler Analizi, bağımsız bir analiz yöntemi olarak uygulanabileceği gibi, diğer analiz türleri için de veri hazırlama tekniği olarak kullanılabilir (Tatlıdil, 1996).

3.2.TEMEL BİLEŞENLER ANALİZİNE GİRİŞ

Temel bileşenler analizinin tarihteki yeri, tanımı, özellikleri ve gerekliliği hakkında yapılan çalışmalar incelenecektir.

3.2.1. Temel Bileşenler Analizinin Tarihsel Gelişimi

İstatistik, "devlet" veya "hükümet" anlamına gelen Latince "status" veya İtalyanca "statista" kelimelerinden gelmektedir. Bu bağlamda istatistik ilk dönemlerde, "devlet işleri", "devletlerle ilgili veriler", "siyasal aritmetik" anlamlarına gelen isimlerle anılmıştır. Günümüzde ise istatistik, veri bilimi olarak tanımlanmaktadır. Uygulamalı matematiğin bir dalı olan istatistik, ham veriden bilgi çıkarma sanatıdır. Geniş anlamda ise, istatistik:

- Verilerin toplanması, sınıflandırılması, özetlenmesi, düzenlenmesi, analiz edilmesi ve yorumlanmasını;
- Verilerin belli bir kısmından yola çıkarak, bütünü hakkında çıkarımlarda bulunulmasını;
- Elde edilen sonuçların, karar vermede kullanılmasını içerir.

O halde, veriler sayıları, sayılar bilgiyi kapsar, istatistik ise bu bilginin özündeki doğayı araştırır. Sonuç olarak, istatistik, veri toplama ve veriyi anlama sanatı ve bilimidir.

İstatistik fikri ve yöntemleri, çeşitli uygulamalarda kullanmak üzere toplumun veri toplama isteğinin artmasıyla ortaya çıkmıştır. İlk başlarda hükümetler nüfusu öğrenmek ve vergilendirmek için istatistiğe ihtiyaç duymuştur. 17. ve 18. yüzyıllarda fiziksel bilimlerin gelişmeye başlamasıyla birlikte ağırlık, uzaklık ve fiziksel miktarların titiz bir şekilde ölçülmesi önem kazanmıştır. Bilim adamları hassasiyetle yaptıkları ölçümlerde değişkenlikler olduğunu fark etmiştir. Tek bir ölçüm yapmaktansa, değişkenlik göstermesine rağmen çok sayıda ölçümün daha değerli olduğu düşünülmüştür. İstatistiksel yöntemler öncelikle bilimsel ölçümleri analiz etmek amacıyla geliştirilmiştir.

19. yüzyıldan itibaren tarım, biyoloji ve davranış bilimleri, temel sorularına cevap bulmada istatistiğe güvenmeye başlamıştır. İstatistik yöntemlerin asıl gelişimi ise 20. yüzyılda olmuş ve hemen hemen tüm bilim alanlarında kullanımı yaygınlaşmıştır (Miran, 2021).

Temel Bileşenler Analizi (TBA), çok değişkenli istatistiksel analizlerde boyut indirgeme ve veri özetleme amacıyla yaygın olarak kullanılan bir tekniktir. Bu yöntem, veri setindeki temel yapıları ortaya çıkarmak için doğrusal dönüşümler kullanır ve boyutlar arası ilişkilerin daha iyi anlaşılmasını sağlar. Ancak TBA'nın kökenleri, zaman içinde farklı araştırmacılar tarafından geliştirilmiş bir dizi teorik çalışmanın sonucudur. Bu yazı, TBA'nın tarihsel gelişimini, ilk

alışmalarından gnmze kadar olan evrimini ve bu srecin nemli kilometre taşlarını ele almaktadır.

Temel Bileşenler Analizi'nin kkenleri, 19. yzyılın sonlarına kadar uzanır. Presendorfer ve Mobley (1988), TBA'nın kkeninin, Beltrami (1873) ve Jordan (1874) tarafından bağımsız bir şekilde geliştirilen "Tekil Değer Ayrışımı" (Singular Value Decomposition, SVD) analizine dayandığını belirtmişlerdir. Bu erken alışmalar, TBA'nın matematiksel temelini oluşturmuş, ancak metodun pratikte yaygın olarak kullanılabilir hale gelmesi için daha fazla gelişim gerekmektedir.

Temel Bileşenler Analizi'nin ilk tanımları, Karl Pearson (1901) tarafından yapılmıştır. Pearson, ok boyutlu verilerin daha dşk boyutlarda temsil edilmesi için bir yntem nerdi. Bu yaklaşım, verilerin "temel bileşenler" adı verilen doėrusal birleşimlerini kullanarak boyut indirgeme sağlamaktadır. Pearson'ın bu alışmaları, istatistiksel hesaplamalar açısından nemli bir temele dayanmakla birlikte, zellikle sayısal analizlerin zorlukları ve sınırlamaları konusunda dikkatli olunması gerektiğini vurgulamıştır. Pearson, verilerdeki temel bileşenlerin hesaplanmasını teorik olarak mmkn grse de ok sayıda deėişkenin yer aldığı veri setlerinde bu işlemin pratikte zorlayıcı olabileceğini belirtmiştir.

1933 yılında, Amerikan istatistikçi Harold Hotelling, Pearson'ın temel bileşenler analizini daha ileriye taşıyarak, bu tekniėi matematiksel olarak yapılandırmıştır. Hotelling, verilerin temel bileşenlerini, veri setindeki varyansı en iyi şekilde temsil eden doėrusal bileşimler olarak tanımlamış ve bu bileşenleri elde etmek için zdeėer/zvektr yntemini kullanmıştır. Hotelling'in alışması, "temel bileşenler metodu" olarak bilinen yaklaşımın gelişmesine yol açmıştır.

Hotelling'in yaklaşımı, temelde Pearson'ın alışmalarını içermektedir, ancak daha sistematik bir şekilde kovaryans matrisinden yararlanarak ve Lagrange arpanları gibi optimizasyon tekniklerini kullanarak temel bileşenleri elde etmiştir. Hotelling, ayrıca bu bileşenlerin, istatistiksel analize daha uygun hale gelmesi için "bileşenler" olarak adlandırılmasına karar vermiştir; bylece matematiksel literatrdeki "faktr" terimi ile karışıklıkların nne geilmiştir.

Hotelling'in temel bileşenler analizi, faktr analizi (factor analysis) ile bazı benzerlikler taşısada bu iki teknik arasında nemli farklar vardır. Her iki yntemde, ok deėişkenli verileri analiz etmek için doėrusal dnşmler kullanırken, temel bileşenler analizi, veri setindeki en fazla varyansı açıklayan doėrusal kombinasyonları bulmayı hedeflerken; faktr analizi, genellikle daha az sayıda "gizli" deėişken (faktr) tanımlamayı amaçlar. Hotelling'in alışmasındaki en nemli katkılardan biri, temel bileşenlerin, faktr analizi için uygun olmayan daha yksek boyutlu bir zm elde edilmesini sağlamasıdır.

Hotelling'in 1933'teki çalışmasının ardından, TBA, özellikle veri madenciliği, görüntü işleme, biyoinformatik, ekonomi gibi alanlarda daha yaygın kullanılmaya başlanmıştır. Ancak, TBA'nın yaygınlaşması, bilgisayar teknolojilerinin gelişmesiyle mümkün olmuştur. Pearson'ın verilerin dört ya da daha fazla değişkeni içerdiğinde hesaplamaların oldukça karmaşıklaşacağını belirttiği gibi, ancak bilgisayarların kullanımının artmasıyla, TBA, yüksek boyutlu veri setlerinin analizinde güçlü bir araç haline gelmiştir.

TBA'nın bilgisayar ortamında uygulanmasının yaygınlaşmasında, Anderson (1963) ve Rao (1964) gibi araştırmacıların katkıları büyük olmuştur. Anderson, özellikle örneklem dağılımlarına ve temel bileşenlerin istatistiksel özelliklerine odaklanırken, Rao'nun çalışmaları, TBA'nın daha geniş veri setlerine nasıl uygulanabileceği ve bu verilerin nasıl yorumlanacağı konusunda yeni perspektifler sunmuştur.

Sonraki yıllarda, TBA üzerine yapılan çalışmalar daha geniş bir yelpazeye yayılmış ve bu teknik, daha karmaşık analizler için uyarlanmıştır. Gower (1966), TBA ve diğer istatistiksel teknikler arasındaki ilişkileri incelemiş, Jeffers (1967) ise TBA'nın uygulamalı yönlerine dikkat çekerek, tekniğin teoriden öte, veri analizi sürecinde nasıl kullanılması gerektiğini açıklamıştır. Cooley ve Cohens (1971) tarafından yapılan çalışmalar, temel bileşenler analizinin bilgisayarlarla uygulanabilirliğini göstermiştir. Bu dönemde, bilgisayar teknolojisinin gelişmesiyle birlikte, TBA'nın kullanımı pratikte büyük bir hız kazanmıştır. Pearson'ın "p" sayısının dört ya da daha az olmadığı durumlarda hesaplamaların zor olduğu yönündeki yorumları, bilgisayarların yaygınlaşmasıyla aşılabilmektedir. Bu noktada, TBA'nın daha büyük "p" değerleri için verimli bir şekilde uygulanabilirliği, teknik anlamda önemli bir dönüm noktası oluşturmuştur. Preisendorfer ve Mobley (1988) tarafından yapılan çalışmalar, TBA'nın meteoroloji, okyanus bilimleri gibi alanlarda nasıl uygulanabileceğini ele almış ve bu kitap, TBA üzerine yazılmış önemli bir kaynağa dönüşmüştür.

Günümüzde, TBA hala istatistiksel literatürde önemli bir yere sahiptir. TBA, yalnızca boyut indirgeme değil, aynı zamanda verinin yapısal özelliklerini anlamak ve daha verimli modeller oluşturmak için geniş bir uygulama alanına sahiptir. Bu alanlar arasında ekonomi, biyoloji, jeoloji, iklimbilimi, psikoloji ve genetik gibi pek çok disiplinde yapılan uygulamalar yer almaktadır.

3.2.2. Temel Bileşenler Analizinin Tanımı Ve Mahiyeti

Temel bileşenler analizi TBA, aralarında korelasyon bulunan k adet değişkeni, aralarında korelasyon bulunmayan ve orijinal değişkenlerin doğrusal bileşenleri olan daha az sayıda (m adet)

değişkenle ifade etme yöntemidir. TBA orijinal k adet değişkenin açıkladığı durumu daha az değişkenle ($m < k$) açıklamakta olup, yeni değişkenler orijinal değişkenlerin doğrusal bileşenidir. TBA;

- Veri indirgemesi
- Tahmin işlemi
- Verileri bazı istatistiksel yöntemlerle analiz edilebilecek forma sokma
- Veri sıkıştırma amaçlarıyla kullanılmaktadır (Özdamar, 1999'dan aktaran Akbulut ve Çapık, 2022).

Temel Bileşenler Analizi (TBA), verilerdeki desenleri tanımlamanın ve verilerin benzerliklerini ve farklılıklarını vurgulamak için ifade etmenin bir yoludur. Yüksek boyutlu verilerde desenleri bulmak zor olabileceğinden, grafiksel temsilin lüksü mevcut değildir. TBA, verileri analiz etmek için güçlü bir araçtır. TBA'nın diğer ana avantajı, verilerdeki bu desenler bir kez bulunduğunda, boyut sayısını azaltarak verilerin sıkıştırılabilmesi ve çok fazla bilgi kaybı olmadan yapılabilmesidir. Temel Bileşenler Analizi, bir boyut azaltma tekniğidir. Orijinal değişkenlerin doğrusal kombinasyonları olan yeni değişkenler oluşturma tekniğidir. Yeni değişkenlere temel bileşenler denir ve birbirleriyle ilişkisiz olacak şekilde seçilirler (Gezmez, 2007).

Büyük veri setleri, çeşitli disiplinlerde giderek daha yaygın hale gelmektedir. Bu veri setlerini anlamlı bir şekilde analiz edebilmek için, verinin yoğunlukla içerdiği bilgiyi koruyarak boyutunun önemli ölçüde azaltılmasını sağlayan yöntemlere ihtiyaç duyulmaktadır. Bu amaçla birçok teknik geliştirilmiş olmakla birlikte, temel bileşen analizi (TBA) en eski ve en yaygın kullanılan yöntemlerden biridir. TBA'nın temel prensibi, bir veri setinin boyutunu azaltırken, verinin içinde barındırdığı istatistiksel değişkenliği mümkün olan en fazla şekilde muhafaza etmektir.

Çeşitli disiplinlerde kullanılsada, bazen yeniden keşfedilsede, temel bileşen analizi özünde istatistiksel bir tekniktir ve bu nedenle büyük ölçüde istatistikçiler tarafından geliştirilmiştir. Bu durum, "mümkün olduğunca fazla değişkenliği korumak" ifadesinin, orijinal veri setindeki değişkenlerin doğrusal fonksiyonları olan, ardışık olarak varyansı maksimize eden ve birbirleriyle korelasyonsuz yeni değişkenlerin (temel bileşenlerin) bulunmasına indirgenmesi anlamına gelir. Bu tür yeni değişkenleri, yani temel bileşenleri (PC'ler), bulmak, özdeğer/özvektör problemi çözmeye indirgenir (Jolliffe ve Cadima, 2016).

Matematiksel açıdan, Temel Bileşenler Analizi (TBA), değişkenler arasındaki simetrik kovaryans matrisinin kullanılmasıyla gerçekleştirilir. $N \times N$ boyutunda bir kare matris için, bu

matrisin N özdeğeri ve karşılık gelen N özvektörü hesaplanabilir. Temel bileşenler, bu özvektörler olarak tanımlanır ve en yüksek özdeğere sahip olan özvektör, ilk temel bileşen olarak kabul edilir. Veri setindeki en büyük varyansa sahip bileşenlerin, genellikle en önemli bileşenler olarak kabul edildiği varsayımı altında, temel bileşenler arasındaki ilişkiler görselleştirilerek, veri kümesindeki örnekler arasındaki mesafeler ve farklılıklar daha açık bir şekilde analiz edilebilir. Ayrıca, temel bileşenlere en fazla katkıyı sağlayan değişkenlerin belirlenmesi, biyolojik anahtar süreçler gibi karmaşık sistemlerde anlamlı içgörüler elde edilmesine olanak sağlar. Bu analizde, binlerce orijinal değişken yerine, genellikle veri kümesindeki toplam varyansın %90'ından fazlasını açıklayan yalnızca on veya daha az temel bileşenin bulunduğu gözlemlenmektedir (Saeed, 2022).

Boyut indirgeme, çok sayıda değişkenin daha az sayıda asal bileşene dönüştürülmesi işlemidir ve bu süreçte bilgi kaybı minimumda tutulur. Ancak, bu yöntem orijinal değişkenler arasındaki ilişkilerin yapısını inceleyemez. Bazı araştırmacılar, Asal Bileşenler Analizi (TBA) ve Açıklayıcı Faktör Analizi (EFA) yöntemlerini yanlışlıkla aynı şey olarak kabul etmektedir (Sharma, 1996:108).

Asal Bileşen Analizi, değişkenlerin açıklanan varyansını maksimize etmeyi amaçlayan bir tekniktir. Bu model, p sayıda bağımsız asal bileşenin her birinin, değişkenlere ait toplam varyansa en yüksek katkıyı sağladığı bir yapıyı ortaya koyar. Sonuç olarak, p adet değişken, yine p adet bağımsız asal bileşenle temsil edilir ve bu bileşenler sırasıyla toplam varyansa en fazla katkıyı yapacak şekilde sıralanır. İlk asal bileşen, toplam varyansa en yüksek katkıyı sağlarken, diğer bileşenler giderek azalan katkılar sunar. Bu özellik, sınırlı sayıda asal bileşenle, verinin toplam varyansının büyük bir kısmının açıklanabilmesine olanak tanır. Ayrıca, asal bileşen modelinde, değişkenin varyansı, ortak ve spesifik varyans olarak ayrılmadan, tümüyle ortak varyans olarak kabul edilmektedir.

Bu durumda, toplam varyansın ortak ve spesifik varyanslar olarak ayrılmadan bir bütün olarak ele alındığı ifade edilmektedir. Başka bir deyişle, Temel Bileşenler Analizi (TBA), değişken sayısının ve ortak faktör sayısının eşit olduğunu varsayar; bu ortak faktörlerden bazıları toplam varyansın önemli bir kısmını açıklar, geri kalan faktörler ise spesifik varyansı gösterir. Bu bağlamda, korelasyon matrisinin köşegen elemanları, Temel Bileşenler Analizinde 1 olarak kabul edilirken, Açıklayıcı Faktör Analizinde (EFA) bu elemanlar ortak varyans değerlerini ifade eder ve genellikle 1'den küçük olurlar. Yani, EFA, her bir değişkenin varyansını ortak ve spesifik varyans olarak iki bileşene ayırır ve ortak varyansın yalnızca ortak faktörlerden kaynaklandığını kabul eder. EFA'nın amacı, önce ortak varyansları tahmin etmek ve ardından değişkenler arasındaki korelasyonları ve ortak varyansları dikkate alarak ortak faktörleri belirlemektir. EFA, gözlemlenen ve ölçülen birçok özelliğin arkasında, gözlemlenemeyen ve ölçülemeyen gizli

boyutların var olduğunu kabul etmektedir (Sharma, 1996:108-109 ve Fabrigar, 1999:273-299'dan aktaran Albayrak, 2005).

$X_1, X_2, \dots, X_p$ ile gösterilen değişkenlerden bazılarının birbirleriyle ilişkili (bağımlı) olduğu ve değişken sayısının çok fazla, yani p 'nin yüksek olduğu durumlarda, belirtilen analizlerin uygulanması araştırmacılar için bazı zorluklar yaratabilir. Değişken sayısının aşırı fazla olması, hem işlem yükünü ve maliyeti artırmakta hem de elde edilen sonuçların yorumlanmasında zorluklara yol açabilmektedir.

p (bağımlı değişken sayısı) $> n$ (gözlem sayısı) ise regresyon analizinde çözümleme etkisiz ve hatta imkânsız olabilir. Öte yandan $p \cong s.d_{hata}$ iken varyans analizinde uygulanan testin gücü düşer, eğer $p > s.d_{hata}$ ise test uygulanamaz. Bağımsız (açıklayıcı) değişkenlerin yüksek derecede ilişkili olması durumunda regresyon analizinde regresyon parametrelerinin tahminleri tutarsız olabilir. Çünkü bu durumda çoklu bağlantı sorununun ortaya çıkması muhtemeldir. Bu tür sorunların giderilmesinde başvurulacak istatistiksel tekniklerin en önemlisi temel bileşenler analizidir.

TBA, başlangıç sistemindeki p tane X_j değişkenlerinin varyans-kovaryans veya korelasyon yapısını daha az sayıda ve bu değişkenlerin doğrusal bileşimleri olan yeni değişkenlerle ifade etme yöntemidir. Diğer bir ifade ile aralarında orta ya da yüksek derecede korelasyonlar olan p tane X_j değişkenlerinin açıkladığı yapıyı (varyans-kovaryans veya korelasyon) daha az sayıda ($k \leq p$) ve başlangıç değişkenlerin doğrusal fonksiyonları olan yeni değişkenlerle ifade etme yöntemine TBA denir.

TBA'nın üç temel amacı vardır. Bunlar:

- Değişkenler arasındaki bağımlılık yapısını ortadan kaldırmak (çoklu bağlantıyı gidermek)
- Boyut (veri) indirgemesi yapmak
- Başka istatistiksel analizler için veri hazırlamak şeklinde ifade edilir.

3.2.3.Temel Bileşenler Analizinin Özellikleri

Temel Bileşenler Analizi (TBA), çoklu iç ilişkilerden kaynaklanan veri karmaşıklığını ortadan kaldırmak amacıyla kullanılan bir faktör türetme tekniğidir. Bu yöntem, çok sayıda değişkenin doğrusal kombinasyonlarını kullanarak daha az sayıda bileşenle veri setinin varyans-kovaryans yapısını açıklamayı amaçlar. TBA, verilerin sadeleştirilmesini ve daha anlamlı bir şekilde analiz edilmesini sağlar. Bu analizde, değişkenler bağımlı ya da bağımsız olarak tanımlanamaz. Bunun yerine, her bir değişkenin normal dağılıma sahip olduğu, gözlemler arasında

doğrusal bir ilişki olduğu ve verilerin birbirinden bağımsız olduğu varsayımlarına dayanılır (Bayram, 2004; 131).

TBA'nın önemli bir avantajı, çoklu bağımsız değişkenlerin bir arada kullanılmasıdır. Bu, özellikle çoklu iç ilişkilerden kaynaklanan karmaşık veri setlerini sadeleştirmek için faydalıdır. Birden fazla değişkenin aynı anda kullanılması, verilerdeki farklı boyutları açıklamak için oldukça etkilidir. Temel bileşenler analizi, içsel ilişkileri minimize etmek için başvurulan bir yöntemdir. Veri setindeki bağımsız boyutların minimum sayısını belirlemek için tercih edilir (Billor ve Kırıl, 2005).

Temel bileşenler analizi, değişkenler arasındaki doğrusal ilişkileri inceleyerek her bir temel bileşeni türetir. Birinci temel bileşen, verilerdeki maksimum varyansı açıklayacak şekilde türetilir. İkinci temel bileşen ise geriye kalan varyansa maksimum katkıyı sağlar. Bu, temel bileşenler analizinin en önemli özelliklerinden biridir, çünkü az sayıda bileşenle toplam varyansın büyük bir kısmı açıklanabilir. Bu nedenle, TBA ile sadeleştirilmiş bir model oluşturulabilir ve yorumlanması daha kolay hale gelir (Hacıhasanoğlu, 1980;13 ve Albayrak, 1980; 89).

Geometrik olarak, temel bileşenler analizi, orijinal eksenlerin döndürülmesiyle yeni bir koordinat sistemi oluşturur. Bu döndürme işlemi, verinin en yüksek varyansını açıklayacak doğrusal kombinasyonları temsil eden yeni eksenler yaratır. İlk temel bileşen, verilerdeki maksimum varyansı açıklayacak şekilde türetilirken, diğer bileşenler ise giderek daha az varyans açıklayan yeni eksenler oluşturur. Bu özellik, verinin daha sadeleştirilmesine ve anlamlı bir şekilde analiz edilmesine olanak tanır (Tezbaşaran, 2016; 17).

Temel bileşenler analizi, yalnızca veri setini sadeleştirmekle kalmaz, aynı zamanda daha geniş analizlere olanak tanır. Örneğin, çoklu regresyon analizi uygulanması istenilen ancak verilerin çoklu bağımlılık koşulunu sağlamadığı durumlarda, veriler temel bileşen skorlarına dönüştürülerek regresyon analizi yapılabilir. Benzer şekilde, kümeleme analizine uygun olmayan veri setleri, temel bileşenler skorlarına dönüştürülerek kümeleme analizi için uygun hale getirilebilir (Özdamar, 2002; 214).

Temel Bileşenler Analizi, doğrusal bir analiz yöntemidir. Matematiksel olarak temel bileşenler, ppp adet rastgele değişkenin doğrusal kombinasyonları şeklinde tanımlanır. Geometrik açıdan ise bu doğrusal kombinasyonlar, orijinal eksenlerin döndürülmesiyle elde edilen yeni bir koordinat sistemini ifade eder. Bu yeni koordinat sistemi, verinin en yüksek varyansını temsil eder ve verinin daha basit bir şekilde yorumlanmasına olanak tanır (Demir, Esenbuğa, & Bilgin, 2016; 538). Ancak, TBA'da ilk temel bileşenin bağımsız değişkenle ilişkisiz olması, sonuçların geçerliliğini olumsuz etkileyebilir ve analiz sonuçlarının yetersiz olmasına neden olabilir (Aktürk, 2012; 2).

Sonuç olarak, temel bileşenler analizi, çoklu bağımsız değişkenlerin doğrusal kombinasyonları ile daha az sayıda bileşene indirgenmesini sağlayarak, verinin özetlenmesine olanak tanır. Bunun yanı sıra, veri setinin daha sadeleştirilmesine ve analiz sonuçlarının daha anlamlı hale getirilmesine yardımcı olur. Ancak, doğru sonuçlar elde edebilmek için, değişkenler arasındaki ilişkilerin dikkate alınması önemlidir.

Temel Bileşenler Analizi yöntemi, verinin çeşitliliğini daha etkili bir şekilde temsil edebilecek yeni bir boyut seti oluşturmayı hedefler. İlk boyut, verinin maksimum çeşitliliğini yansıtacak şekilde belirlenirken, ikinci boyut, birincisine dik olacak şekilde seçilir ve yine veri içindeki çeşitliliği en iyi şekilde açıklayacak biçimde yapılandırılır. TBA'nın birkaç karakteristik özelliği bulunur. Bunlar arasında, verinin içindeki en güçlü örüntülerin tespit edilmesi ve verinin çeşitli yönlerinin ortaya konulması yer alır. Ayrıca, TBA, verideki gürültülerin, örüntülerin aksine zayıf olduğundan, bu tür gürültülerin boyut küçültme işlemi sırasında temizlenmesini sağlar. Bu özellik, veri madenciliği ve diğer veri analiz algoritmalarında oldukça faydalıdır. Boyut küçültme işlemleri, daha küçük boyutlu veri setlerinin elde edilmesini sağlayarak, yüksek boyutlu verilere uygun olmayan analiz tekniklerinin, bu yeni veri seti üzerinde daha verimli bir şekilde uygulanmasına olanak tanır (Yıldız, Çamurcu, & Doğan, 2010).

3.2.4. Temel Bileşenler Analizinin Gerekliliği Ve Sağladığı Yararları

Temel Bileşenler Analizi (TBA), özellikle yüksek boyutlu veri setlerinde değişkenler arasındaki doğrusal ilişkileri anlamak ve karmaşıklığı azaltmak amacıyla kullanılır. Alptekin Günel'in (1977) belirttiği gibi, faktör analizinin temel amacı, çok sayıda değişkenin daha az sayıda temel bileşene indirgenmesidir. TBA da benzer şekilde, yüksek korelasyonlar ve çoklu bağımsızlık problemleri gibi zorluklarla karşılaşılan durumlarda etkili bir çözüm sunar.

TBA'nın gerekliği şu noktalarda özetlenebilir:

- a. Boyut İndirgeme: Çok sayıda değişkenin bulunduğu durumlarda TBA, veriyi daha yönetilebilir bir hale getirir ve anlamlı bir şekilde analiz edilebilmesini sağlar
- b. Çoklu Bağımlılık Sorunlarının Çözülmesi: TBA, değişkenler arasındaki yüksek korelasyonu ortadan kaldırarak daha güvenilir analizlere olanak tanır ve böylece multicollinearity problemini çözer.
- c. Veri Görselleştirme ve Yorumlama: TBA, verinin daha düşük boyutlarda temsil edilmesini sağlayarak, veri setindeki ana yapıları görselleştirir ve yorumlamayı kolaylaştırır (Günel, 1977).

Sonuç olarak, TBA, büyük veri setlerinde analiz ve yorumlamayı kolaylaştırmak, veri boyutunu küçültmek ve gürültüyü azaltmak için temel bir araçtır.

Temel Bileşenler Analizi, çok boyutlu veri setlerinde değişkenler arasındaki ilişkileri daha anlaşılır hale getirmek ve veriyi daha yönetilebilir bir formata indirmek amacıyla yaygın olarak kullanılan istatistiksel bir tekniktir. TBA, özellikle büyük veri setlerinde ve çoklu değişkenlerin yer aldığı durumlarda boyut indirgeme sağlar, veri analizi sürecini basitleştirir ve anlamlı sonuçlar elde edilmesini kolaylaştırır. Ancak, TBA'nın gerekliliği ve etkinliği, belirli koşullara bağlıdır. Bu yazıda, TBA'nın gerekliliği ve sağladığı yararlar, literatürden yapılan alıntılarla ele alınacaktır.

3.2.4.1. TBA'nın Gerekliliği

TBA'nın gerekliliği, genellikle değişkenler arasındaki korelasyonun yüksek olduğu ve veri setinin boyutunun büyük olduğu durumlarda ortaya çıkar. Güneş (2007)'e göre, değişkenler arasında yüksek derecede bir korelasyon bulunması, çoklu regresyon analizlerinde çoklu bağlantı (multicollinearity) sorunlarına yol açabilir. Bu durum, modelin tahmin gücünü azaltabilir ve güvenilir analizler yapılmasını engelleyebilir. Ayrıca, çok sayıda değişkenin bulunduğu veri setlerinde her bir değişkenin tek başına analiz edilmesi, işlem zorluğu, zaman kaybı ve maliyet gibi sorunları beraberinde getirir. Sütçü Asar (2014), yüksek boyutlu veri setlerinde bu tür problemlerin ortadan kaldırılabilmesi için TBA'nın önemli bir araç olduğunu belirtmektedir.

TBA'nın gerekliliğini belirlemek için Bartlett Küresellik Testi sıklıkla kullanılmaktadır. Bu test, veri setindeki değişkenler arasında bir ilişki olup olmadığını belirlemeye yardımcı olur. Eğer korelasyon matrisi birim matrisle (I) eşit olursa, değişkenler arasında ilişki bulunmadığı kabul edilir ve TBA'nın uygulanması anlamlı olmayacaktır. Güneş (2007), bu durumda, veri setinin bir küre oluşturduğunu ve dolayısıyla boyut indirgemesinin fayda sağlamayacağını ifade etmektedir. Ancak, test sonucunda korelasyon matrisi birim matrisle eşit değilse ve veriler elipsoid bir yapı oluşturuyorsa, TBA uygulamak anlamlıdır. Bu durum, değişkenler arasında önemli ilişkilerin bulunduğunu ve boyut indirgeme işleminin veriyi daha anlaşılır hale getireceğini gösterir

3.2.4.2. TBA'nın Yararları

TBA, özellikle çoklu değişkenlerin yer aldığı veri setlerinde çeşitli yararlar sağlar. Sütçü Asar (2014)'a göre, TBA'nın en önemli yararlarından biri, değişkenler arasındaki korelasyonu ortadan kaldırarak bağımsız yeni bileşenler elde etmesidir. Bu, özellikle regresyon analizleri gibi modelleme süreçlerinde büyük önem taşır. Yeni elde edilen bileşenler ortogonaldır (dikeydir), yani birbirleriyle ilişkisizdir. Bu sayede, temel bileşen skorları kullanılarak yapılan karşılaştırmalar daha güvenilir ve anlamlı olur.

TBA'nın diğ er bir  nemli yararı, veri setinin boyutunu indirgerken, veri setinin b y k bir kısmını temsil eden temel bile enlerin  ıkarılmasını saėlamasıdır. S t   Asar (2014), bu sayede veri setindeki  nemli bilgilerin kaybolmadan daha k   k bir boyutta analiz yapılabilceėini belirtmektedir. Boyut indirgeme i lemi, gereksiz deėi kenleri ortadan kaldırarak analiz s recini basitle tirir ve daha hızlı sonu lar elde edilmesini saėlar. Ayrıca, TBA, veri setindeki temel yapıları daha net bir  ekilde ortaya  ıkarır, b ylece daha anlamlı ve doėru sonu lar elde edilebilir.

Bir ba ka  nemli yarar, TBA'nın aşırı  ğrenme (overfitting) riskini azaltmasıdır. Y ksek boyutlu veri setlerinde,  ok sayıda deėi ken kullanılarak olu turulan modeller aşırı  ğrenme yapabilir ve bu da modelin genellenebilirliėini azaltabilir. TBA, gereksiz deėi kenleri ortadan kaldırarak modelin daha genellenebilir olmasını saėlar ve bu sayede daha saėlam sonu lar elde edilmesine yardımcı olur.

TBA'nın saėladığı bir diğ er  nemli avantaj ise veri g rselle tirmesidir.  ok boyutlu veri setlerinin g rselle tirilmesi, genellikle zordur. Ancak, TBA, verileri iki veya    temel bile ene indirger ve bu sayede verinin g rsel olarak daha anlaşılır bir  ekilde temsil edilmesini saėlar. Bu,  zellikle b y k veri setlerinde veri ke fi ve modelleme s re lerinde faydalıdır.

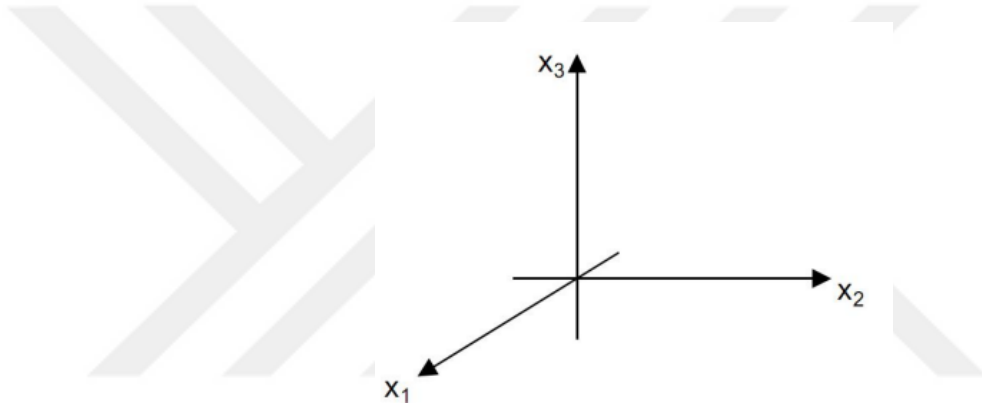
Sonu  olarak, TBA,  oklu deėi kenlerin yer aldığı ve y ksek boyutlu veri setlerinde verinin analiz edilmesi ve yorumlanması i in vazge ilmez bir tekniktir. Cooley ve Lohnes (1971)'a g re, TBA'nın bir diğ er yararı, temel bile enlerin ortogonal olması nedeniyle, her bir bile enin baėımsız olarak incelenebilmesi ve veri setinin daha anlamlı hale gelmesidir. Harris (2001) ise, TBA'nın, verinin boyutunu indirmekle birlikte,  ok sayıda deėi ken arasındaki karma ık ili kiyi basitle tirerek, daha a ık ve net sonu lar elde edilmesine olanak saėladığını vurgulamaktadır. Bu sebeplerle, TBA,  zellikle y ksek boyutlu veri setlerinde veriyi anlamlı bir  ekilde analiz etmek, anlamlı bile enler  ıkarmak ve modelin doėruluėunu artırmak i in son derece  nemli bir ara tır.

4. TEMEL BİLEŞENLER ANALİZİNİN TEORİK HESAPLARI

Bu bölümde temel bileşenler analizinin geometrik ve analitik yaklaşımları ele alınacaktır.

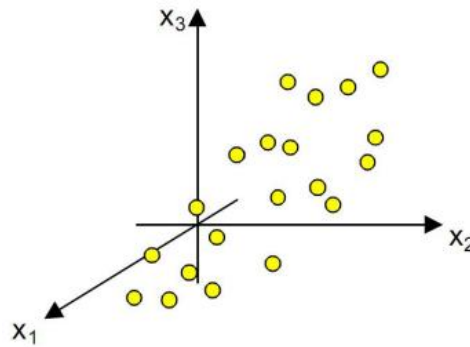
4.1. TEMEL BİLEŞENLER ANALİZİNE GEOMETRİK YAKLAŞIM

Değişkenleri sütunlar, bu değişkenlere ait gözlem değerlerini ise satırlar olarak içeren bir veri matrisi, değişken sayısı kadar boyuta sahip bir koordinat sistemi şeklinde kabul edilebilir. Örnek olarak, üç sütundan oluşan bir X matrisi için değişkenler X_1, X_2, X_3 olarak adlandırılabilir. Bu durumda, sistem geometrik olarak üç boyutlu bir koordinat düzleminde, Şekil 4-1'te gösterildiği gibi temsil edilebilir.



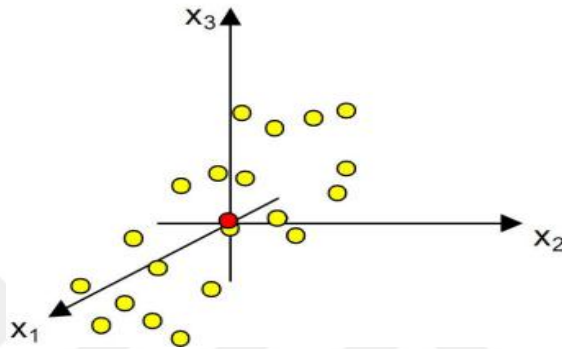
Şekil 4-1: Üç değişkenli bir veri matrisinin geometrik gösterimi

Veri matrisinin (X) her satırında yer alan gözlem değerleri, bu koordinat sistemi üzerinde bir noktayı temsil eder. Eğer değişkenlerin değişkenlikleri birbirine yakın veya eşitse, gözlem değerleri koordinat düzleminde Şekil 4-2'te gösterildiği gibi dağınık noktalar olarak görülebilir.



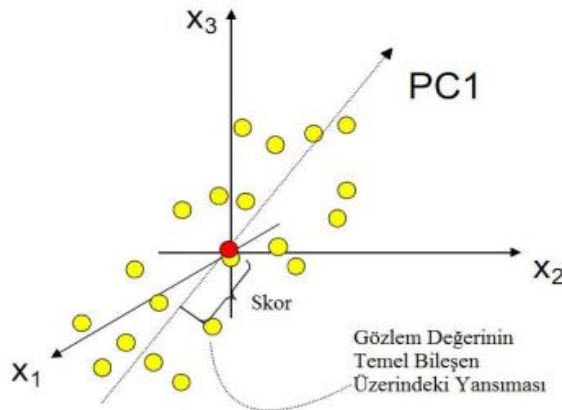
Şekil 4-2: Koordinat sistemi üzerinde gözlem verilerinin dağılımı

Koordinat sisteminde her bir değişkenin eşit düzeyde temsil edilebilmesi için, değişkenlerin varyanslarının eşit olması gerekmektedir. Örnek olarak, bir gerilim değişkeni -200 ile 200 arasında değişirken, bir diğer devir sayısı değişkeni yalnızca 100 ile 150 arasında bir aralıkta değişiyorsa, koordinat sistemindeki gösterimde daha yüksek varyansa sahip olan gerilim değişkeni baskın hale gelecektir. Bu durumu önlemek amacıyla, veriler ölçeklendirilir ve sıfır ortalamaya göre standartlaştırılır. Böylece, tüm değişkenlerin sistem üzerindeki etkileri dengelenmiş olur.



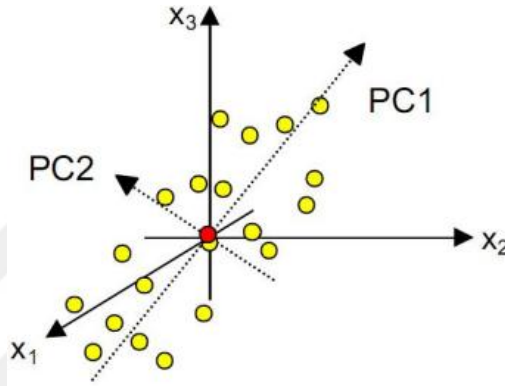
Şekil 4-3: Ortalanmış verilerin koordinat sistemi üzerindeki dağılımları

Temel Bileşenler Analizi (TBA), mevcut verilerdeki değişimleri göz önünde bulundurarak, sistemi daha az boyuta sahip ve bağımsız değişkenlerden oluşan yeni bir koordinat sistemine dönüştürmeyi amaçlayan bir yöntemdir. Bu doğrusal dönüşüm sonucunda oluşturulan yeni koordinat düzleminin eksenleri "Temel Bileşenler" olarak adlandırılır. Birinci Temel Bileşen(PC1), en küçük kareler yöntemiyle belirlenir ve çok boyutlu uzayda verilerin dağılımını en iyi şekilde temsil eden, verilerin ağırlık merkezinden geçen doğrultuda konumlanır.



Şekil 4-4: Örnek bir gözlem verisinin Birinci Temel Bileşen (PC1) üzerindeki yansıması

Üç boyutlu koordinat sisteminde yer alan her noktanın, PC1 hattına yansımaları (iz düşümü) belirlenebilir. Temel Bileşen hattı boyunca hesaplanan bu yansımalar, literatürde "skor" denir. Çoğunlukla, bir temel bileşen veri setini tam olarak tanımlamak için yeterli olmayabilir. Bu nedenle, veriyi daha doğru bir şekilde tanımlamak için birden fazla temel bileşen kullanılabilir ve bu bileşenlerin birbirlerine dik olması gerekir. Geçerli veriler elde edebilmek için kaç adet temel bileşenin kullanılacağına dair hesaplamalar bir sonraki bölümde açıklanacaktır. Örneğin, daha önce verilen üç değişkenli sistemde, ikinci Temel Bileşen (PC2), PC1'e dik olacak şekilde ve uygun en fazla veriyi simgeleyecek şekilde ortalama noktadan geçen bir doğru olarak belirlenir (Bayar, 2013).



Şekil 4-5: Temel Bileşenlerin geometrik gösterimi

4.2. TEMEL BİLEŞENLER ANALİZİNE ANALİTİK YAKLAŞIM

Temel Bileşenler Analizi, çok boyutlu verilerin boyutunu indirmek ve bu verileri daha anlamlı bir şekilde ifade etmek için sıkça kullanılan istatistiksel bir tekniktir. TBA, verinin temel bileşenlere ayrılması yoluyla, daha düşük boyutlu bir uzaya projeksiyon yaparak, verinin çoğunlukla varyans taşıyan özelliklerini öne çıkarır. Bu işlem, verinin yapısal özelliklerinin daha anlaşılır hale gelmesine olanak tanır. Temel bileşenlerin hesaplanması, kovaryans matrisinin özdeğerleri ve özvektörleri üzerinden gerçekleştirilir. Bu yaklaşımda her bir temel bileşen, orijinal değişkenlerin doğrusal bir kombinasyonu olarak türetilir ve her bir bileşen, sırasıyla verinin varyansına en fazla katkıyı sağlayacak doğrultuda yer alır.

Her bir temel bileşen (örneğin $Y_1, Y_2, \dots, Y_p$), başlangıçtaki değişkenlerin doğrusal kombinasyonlarından oluşur. Temel bileşenlerin katsayıları (a_{ij}), veri matrisinin içerdiği varyansı anlamak için önemli rol oynar. Bu katsayılar şu koşullara uyar;

Varyansın Maksimizasyonu: birinci temel bileşen, tüm veri setinin varyansını en fazla açıklayan bileşendir. İkinci temel bileşen, birinci bileşenden bağımsız olup, ikinci en büyük

varyansı açıklayan bileşendir. Bu şekilde, temel bileşenler sırasıyla azalan bir şekilde toplam varyansa katkı sağlar.

Temel Bileşen Ağırlıkları; her bir temel bileşenin katsayıları (a_{ij}) karelerinin toplamı 1 olmak zorundadır. Bu, bileşenlerin normlanmış (standartlaştırılmış) olduğunu gösterir. Örneğin, birinci temel bileşen için $a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1$ koşulu sağlanır. Bu normlanma, bileşenlerin ölçeklerinin birbirleriyle uyumlu olmasını sağlar.

Bağımsızlık (Ortalama Sıfır Korelasyon); temel bileşenler birbirinden bağımsız olmalıdır. Yani, her iki temel bileşen arasındaki korelasyon sıfırdır. Matematiksel olarak, bir temel bileşenin katsayılarının karelerinin toplamı ile diğer bileşenin katsayılarının karelerinin toplamı sıfır olmalıdır.

Veri Standartlaştırma, ikinci koşul (katsayıların karelerinin toplamının 1 olması) sağlanabilmesi için, verilerin standartlaştırılması gerekir. Yani, her bir değişkenin ortalaması 0, varyansı 1 olacak şekilde veriler düzenlenir. Bu adım, verilerin ölçü birimlerinin farklı olmasının temel bileşenler üzerindeki etkilerini ortadan kaldırır. Bu durumda;

- i. Temel bileşenler, başlangıçtaki değişkenlerin doğrusal kombinasyonlarıdır.
- ii. Her bir bileşenin varyansı sırasıyla en yüksekten en düşüğe doğru düzenlenir.
- iii. Temel bileşenler arasında ortogonal (bağımsız) ilişkiler vardır.
- iv. Temel bileşenlerin katsayılarının karelerinin toplamı 1 olmalıdır.
- v. Verilerin standardizasyonu gereklidir.

4.2.1. Temel Bileşenlerin Hesaplanması

Temel bileşenler analizi, bir veri setindeki p sayıda değişkenin doğrusal bileşenler (veya yeni değişkenler) aracılığıyla daha düşük boyutlu bir temsilini elde etmeyi amaçlar. Bu bileşenler, orijinal değişkenlerin doğrusal kombinasyonlarıdır ve her bir bileşen, verinin varyansını azalan sırayla açıklar.

Bir rassal vektör $X = [X_1 + X_2 + \dots + X_p]^T$ olsun. vektörün varyans-kovaryans matrisi Σ' 'nin özdeğerleri $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ sıralamasına sahiptir. Bu durumda, doğrusal bileşenler şu şekilde tanımlanır:

$$Y_1 = a_1^T X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a_2^T X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

$$\vdots$$

$$Y_p = a_p^T X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p$$

Bu doğrusal bileşenler, matris formunda şu şekilde ifade edilebilir:

$$Y = a^T X$$

Burada a , her bir temel bileşenin katsayılarını içeren bir matristir ve Y , doğrusal bileşenlerin oluşturduğu yeni bir vektördür. Bu doğrusal bileşenler, veri setinin boyutunu azaltarak, verinin varyansını daha basit ve özlü bir şekilde temsil etmek amacıyla kullanılır.

Amaç, X vektörünün doğrusal dönüşümü olan $Y = a^T X$ rassal değişkenlerinin varyansını, $a^T a = 1$ kısıtlaması altında maksimum yapmak için doğru bir şekilde belirlemektir. Burada a vektörü, temel bileşenlerin katsayılarını içeren vektördür. Bu doğrusal bileşenin varyansı, aşağıdaki şekilde hesaplanmaktadır:

$$E((a^T X)^2) = E(a^T X X^T a) = a^T \Sigma a$$

Bu ifade, $a^T \Sigma a$ şeklinde yazılmakta olup, Σ kovaryans matrisini temsil eder. Amaç, bu varyansı maksimum yapacak olan a vektörünü bulmaktır.

Bu doğrusal bileşenin varyansını maksimize etmek için, a vektörünü uygun şekilde optimize etmek gerekmektedir. Bunun için Lagrange çarpanı yöntemi kullanılarak aşağıdaki Lagrange fonksiyonu tanımlanır:

$$\Phi(a, \lambda) = a^T \Sigma a - \lambda(a^T a - 1)$$

Burada λ Lagrange çarpanıdır ve $a^T a = 1$ kısıtlamasını dikkate alır. Lagrange fonksiyonunun optimizasyonu için, a vektörüne göre kısmi türev alınır ve bu türev sıfıra eşitlenir. Bu türevi alarak şu denkleme ulaşılır:

$$\frac{\partial \Phi(a, \lambda)}{\partial a} = 2\Sigma a - 2\lambda a = 0$$

Bu denklem, şu şekilde sadeleştirilebilir:

$$\Sigma a = \lambda a$$

Bu denklem, kovaryans matrisi Σ ile a vektörünün özvektör olarak ilişkili olduğunu ve λ 'nın özdeğerini temsil ettiğini gösterir. Yani, a vektörü, Σ matrisinin özvektörü ve λ ise karşılık gelen özdeğeridir.

Sonuç olarak, çözüm, Σ matrisinin özvektörlerini ve karşılık gelen özdeğerlerini bulmakla elde edilir. Her bir özdeğer, doğrusal bileşenin varyansını ve her bir özvektör de bu doğrusal bileşenin katsayılarını belirler. Bu şekilde, temel bileşenler analizi ile maksimum varyansa sahip doğrusal bileşenler elde edilir.

Verilen denklemde, $a^T a = 1$ kısıtlaması altında, Σ matrisinin özdeğerleriyle ilişkili bir çözüm elde edebilmek için, aşağıdaki koşulun sağlanması gerekmektedir:

$$(\Sigma - \lambda I)a = 0$$

Bu denklemde, yalnızca λ özdeğeri için $\Sigma - \lambda I = 0$ koşulu sağlandığında, sıfırdan farklı bir çözüm bulunabilir. Bu, Σ matrisinin özdeğerlerine karşılık gelen bir karakteristik denkleme yol açar. Dolayısıyla, $\Sigma - \lambda I = 0$ ifadesi, λ 'nın bir p -dereceden polinomunun kökleriyle ilgilidir. Bu polinomun kökleri sırasıyla:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

olarak sıralanır. Burada, Σ matrisinin pozitif tanımlı ve simetrik olduğu varsayıldığında, elde edilen özdeğerlerin tamamının gerçekteki sayılar olduğu ve her bir özdeğerin sıfır ya da pozitif olduğu sonucuna varılır.

Her bir özdeğer, ona karşılık gelen bir özvektörle ilişkilidir. Özdeğerler, matrisin doğrusal bileşenlerinin varyanslarını temsil ederken, her bir özvektör, o özdeğere karşılık gelen doğrusal bileşeni tanımlar. Özdeğerlerden en büyük olanı, en yüksek varyansı temsil eder. Bu bağlamda, maksimum varyansı elde edebilmek için λ_1 özdeğeri seçilecektir.

Bununla birlikte, a_1^T özvektörü, Σ matrisinin en büyük özdeğeri olan λ_1 'e karşılık gelen normalleştirilmiş özvektördür. Bu özvektör, Σ matrisinin doğrusal dönüşümünü en iyi şekilde temsil eden doğrusal bileşeni verir. Dolayısıyla, $Y_1 = a_1^T X$ ifadesi, maksimum varyansı taşıyan doğrusal bileşeni ifade eder.

Sonuç olarak, her bir özdeğer ve özvektör çifti, veri setinin varyansını temsil eden doğrusal bileşenler oluşturur. Ancak, maksimum varyansı elde etmek ve veri setindeki en önemli doğrusal bileşeni seçmek için en büyük özdeğer olan λ_1 ve ona karşılık gelen özvektör olan a_1 seçilir. Bu doğrusal bileşen, veri setinin yapısını en iyi şekilde yansıtan ve en fazla bilgiyi taşıyan bileşendir.

Şimdi, Y_1 ile ilişkisiz en yüksek varyansı içeren doğrusal bileşenler arasında olan $a^T X$ değişkenine bakıldığında, eğer bu doğrusal bileşenler arasında korelasyon yoksa, aşağıdaki koşul sağlanır:

$$0 = E[(a^T X) Y_1]$$

Burada, beklenen değer (yani ortalama) ifadesini açacak olursak:

$$0 = E[(a^T X X^T a_1)] = a_1^T \Sigma a_1 = \lambda_1 a_1^T a_1$$

Bu denklemde, $a_1^T a_1 = 1$ olduğu için, elde edilen sonuç şu şekilde sadeleşir:

$$\lambda_1 = 0$$

Bu durum, hem istatistiksel olarak $a^T X$ ve Y_1 arasında bir ilişki bulunmadığını, hem de geometrik olarak a ve a_1 vektörlerinin iç çarpımının sıfır olduğunu gösterir. Yani, $a^T X$ ve Y_1 birbirine diktir (ortogonaldir).

Sonuç olarak, eğer $\Sigma \neq 0$ ise, $\lambda_1 \neq 0$ olmalıdır. Ancak, $\lambda_1 \neq 0$ olduğunda $a^T a_1 = 0$ olduğu için, $\lambda_1 a^T a_1 = 0$ ifadesi geçerli olur.

Lagrange çarpanları λ ve t_1 kullanılarak, aşağıdaki fonksiyon tanımlanır:

$$\Phi_2 = (a, \lambda, t_1) = a^T \Sigma a - \lambda(a^T a - 1) - 2t_1 a^T \Sigma a_1$$

Bu fonksiyonu maksimum yapmak için, a 'ya göre kısmi türev alınır ve sıfıra eşitlenir. Bu türevden şu eşitlik elde edilir:

$$\frac{\partial \Phi_2}{\partial a} = 2\Sigma a - 2\lambda a - 2t_1 \Sigma a_1 = 0$$

Bu denklemi, a_1^T soldan ile çarptığımızda, şu ifadeye ulaşılır:

$$a_1^T (2\Sigma a - 2\lambda a - 2t_1 \Sigma a_1) = 0$$

Eşitlik açıldığında:

$$2a_1^T \Sigma a - 2\lambda a_1^T a - 2t_1 a_1^T \Sigma a_1 = -2t_1 \lambda_1 = 0$$

Bu denklemin sonucunda, $t_1 = 0$ olduğu bulunur. Bu durumda, a vektörünün $(\Sigma - \lambda I)a = 0$ ifadesini sağlaması ve λ 'nın $\Sigma - \lambda I = 0$ eşitliğini karşılaması gerektiği sonucuna varılır.

Sonrasında, λ_2 için, $(\Sigma - \lambda_2 I)a = 0$ koşulunun sağlandığı gözlemlenir. Burada, $a_1^T a = 1$ olduğu ve $\lambda_1 a_1^T a_1 = 0$ eşitliğini sağlayan bir a vektörünün bulunduğu görülür. Bu durumda, $\lambda_1, \lambda_2, \dots, \lambda_p$ özdeğerler şeklinde sıralanır ve en büyük özdeğere karşılık gelen özvektöre a_2 denir. Bu özvektöre karşılık gelen doğrusal bileşen şu şekilde ifade edilir:

$$Y_2 = a_2^T X$$

Bu süreç devam ettiğinde, $(r + 1)$. adımda, ile ilişkisiz olan tüm aX doğrusal bileşenler arasında en yüksek varyansı barındıracak şekilde a vektörünü bulmak istiyoruz. Bu hedefi gerçekleştirmek için, doğrusal bileşenlerin varyansını maksimize eden bir a vektörü aramamız gerekir. Bunun matematiksel ifadesi şu şekilde verilebilir:

$$E(a^T X Y_i) = E(a^T X X^T a_i) = 0 \quad a^T \Sigma a_i = \lambda_i a^T a_i, \quad i = 1, 2, \dots, r.$$

Burada, λ ve $t_1, t_2, \dots, t_r$ Lagrange çarpanlarıdır. Bu durumu şu şekilde ifade edebiliriz:

$$\Phi_{r+1} = (a, \lambda, t_1, t_2, \dots, t_p) = a^T \Sigma a - \lambda(a^T a - 1) - 2 \sum_{i=1}^r t_i a^T \Sigma a_i$$

Maksimize etmek istediğimiz Y_i bileşenine göre, yukarıdaki ifadenin a 'ya göre türevini alıp sıfıra eşitlediğimizde şu denklemi elde ederiz:

$$\frac{\partial \Phi_{p+1}}{\partial a} = 2\Sigma a - 2\lambda a - 2 \sum_{i=1}^r t_i \Sigma a_i = 0$$

Denklemleri soldan a_i^T ile çarparsak,

$$2a_i^T a - 2\lambda a_i^T a - 2t_i a_i^T \Sigma a_i = 0$$

elde ederiz. $\lambda_i \neq 0$ olduğu durumda $-2t_i \lambda_i = 0$ ve $t_i = 0$ olur. Eğer $\lambda_i = 0$ olursa $\Sigma a_i = \lambda_i a_i = 0$ olur. Verilen eşitlikteki toplamda i . terim biter. Bu durumda a ve λ 'nın $(\Sigma - \lambda I)a = 0$, $\Sigma - \lambda I$ eşitliklerini sağlaması gerekir. İşlemler devam ettirilerek, $(r + 1)$. adımda λ_{r+1} ,

$$(\Sigma - \lambda_{r+1} I)a = 0, a^T a = 1 \text{ ve } E(a^T X Y_i) = E(a^T X X^T a_i) = 0$$

denklemlerini sağlayan a vektörü olsun. $\lambda_1, \lambda_2, \dots, \lambda_p$ en büyüğü olsun. Bu vektörü a_{r+1} ile ve buna karşı gelen doğrusal bileşeni $Y_{r+1} = a_{r+1}^T X$ ile gösterelim. $\lambda_{r+1} = 0$ ve $\lambda_i = 0$ ise $a_i^T \Sigma a_{r+1} = 0$ olmasını gerektirmemektedir. a_{r+1} sıfır olan λ_i li a_i ve a_{r+1} 'in doğrusal bir bileşeni olarak kabul edilebilir. Bu şekilde alınan a_{r+1} tüm a_i lerle diktir. Bu yöntem $a^T a = 0$ ve $(\Sigma - \lambda I)a = 0$ eşitliği karşılayan bir a vektörü bulunamayınca kadar sürdürülür.

4.2.2. Temel bileşenlerin varyans-kovaryans matrisinden elde edilmesi

$$E(a^T X)^2 = E(a^T X X^T a) = a^T \Sigma a$$

İfadesinden hareketle i .inci doğrusal bileşenin varyans-kovaryans matrisi şu şekilde ifade edilebilir:

$$\text{Var}(Y_i) = a_i^T \Sigma a_i$$

i . ve k . değişkenler arasındaki kovaryans matrisi

$$\text{Cov}(Y_i, Y_k) = a_i^T \Sigma a_k \quad i, k = 1, 2, \dots, p$$

eşitlikleri ifade edilir. Temel bileşenlerin varyansları mümkün olan en yüksek seviyede olan ve aralarında korelasyon bulunmayan $Y_1, Y_2, \dots, Y_p$ doğrusal bileşenleridir.

Birinci temel bileşen maksimum varyansa sahip doğrusal bileşendir. Yani, birinci temel bileşen $\text{Var}(Y_1) = a_1^T \Sigma a_1$ değerini maksimum yapar. $a_1^T \Sigma a_1$ değerinin a_1 gibi herhangi bir

sabitlerle çarpılması sonucu değerin yükseltilmesi sağlanabilir. Söz konusu belirsizliği gidermek amacıyla, katsayı vektörünün birim uzunluğa sahip olması koşulu getirilir. Bu koşul, aşağıdaki şekilde tanımlanır.

Birinci temel bileşen: $a_1^T a_1 = 1$ (a_1 vektörünün vektörün uzunluğu 1 olmalı) kısıtına bağlı olarak $Var(a_1^T X)$ ' i maksimize edilmelidir. $a_1^T X$ bir doğrusal kombinasyon olup, X verinin özellik matrisi ve a_1 bir vektördür. Amaç, bu doğrusal kombinasyonun varyansını maksimize etmektir, yani veri setindeki en fazla bilgiyi (değişkenliği) açıklayacak doğrusal bileşeni bulmaktır. Bu bileşen, verinin en fazla varyansını temsil eden doğrusal kombinasyonu ifade eder.

İkinci temel bileşen: $a_2^T a_2 = 1$ (yani ikinci bileşenin uzunluğu 1 olmalı) ve $Cov(a_1^T X, a_2^T X) = 0$ yani birinci bileşen ile ikinci bileşen arasındaki kovaryans sıfır olmalıdır. $a_2^T X$ doğrusal bileşendir. Bu kısıt, bileşenlerin birbirinden bağımsız olmasını sağlar (ortogonal olma durumu).

İkinci temel bileşen, verinin kalan varyansını temsil eder ve birinci temel bileşenden bağımsızdır. İkinci bileşen, birinci bileşenin açıklamadığı varyansı maksimize etmek üzere seçilir. Bu, yine bir özdeğer problemi olarak çözülür, ancak burada birinci bileşenin zaten en büyük varyansı kapsadığı için ikinci en büyük özdeğerin karşılık geldiği özvektör a_2 seçilir.

Özetle, birinci bileşen, verinin en fazla varyansını açıklarken, ikinci bileşen bu varyansın dışında kalan varyansı açıklayacak şekilde ortogonal (bağımsız) olarak seçilir.

i. inci temel bileşen: $a_i^T a_i = 1$ ve $k < i$ için $Cov(a_i^T X, a_k^T X) = 0$ kısıtlarına bağlı olarak $Var(a_i^T X)$ ' i maksimum yapan $a_i^T X$ doğrusal bileşenidir.

Sonuç 1: $[X_1, X_2, \dots, X_p]$ gibi öğelerle tanımlanan bir rassal vektörle bağlantılı olan varyans-kovaryans matrisi Σ olsun. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ olacak şekilde, Σ 'nin özdeğer-özvektör çiftleri $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ olmak üzere i . temel bileşen

$$Y_i = e_i^T X = e_{i1}^T X_1 + e_{i2}^T X_2 + \dots + e_{ip}^T X_p \quad i = 1, 2, 3, \dots, p$$

ile verilir.

Burada $Var(Y_i) = e_i^T \Sigma e_i = \lambda_i \quad i; 1, 2, 3, \dots, p$ ve $Cov(Y_i, Y_k) = e_i^T \Sigma e_k = 0 \quad i \neq k$ dır. Eğer bazı λ_i ler eşit ise ilişkili katsayı vektörleri e_i 'ler ve Y_i 'lerin seçimi tek değildir.

Σ yarı pozitif tanımlı bir matris ve $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ler bu matrisin öz değerleri ve $e_1, e_2, \dots, e_p$ lerde öz vektörleri olmak üzere, özdeğer problemine dayanarak optimizasyonlar yapılır.

$$\max_{a \neq 0} \frac{a^T \Sigma a}{a^T a} = \lambda_1 \quad a = e_1$$

Özvektörler normalleştirildiğinden, $e_1^T e_1 = 1$ olduğu belirtilir.

$$\max_{a \neq 0} \frac{a^T \Sigma a}{a^T a} = \lambda_1 = \frac{e_1^T \Sigma e_1}{e_1^T e_1} = e_1^T \Sigma e_1 = \text{Var}(Y_1)$$

$$\max_{a \perp e_1, e_2, \dots, e_k} \frac{a^T \Sigma a}{a^T a} = \lambda_{k+1} \quad k = 1, 2, 3, \dots, p-1$$

$a = e_{k+1}$ ve $e_{k+1}^T e_k = 0$ olmak üzere $\frac{e_{k+1}^T \Sigma e_{k+1}}{e_{k+1}^T e_{k+1}} = e_{k+1}^T \Sigma e_{k+1} = \text{Var}(Y_{k+1})$ sonucuna ulaşılır.

$e_{k+1}^T (\Sigma e_{k+1}) = \lambda_{k+1} e_{k+1}^T e_{k+1} = \lambda_{k+1}$ dolayısıyla $\text{Var}(Y_{k+1}) = \lambda_{k+1}$ olur. Burada e_1 nin e_k ya diktir diyebiliriz. Yani $i \neq k$ olmak üzere $e_1 e_k = 0$ dır. e_1 nin e_k ya dik olması $\text{cov}(Y_1, Y_k) = 0$ sonucunu verir.

Eğer bir matrisin özdeğerleri farklıysa, bu matrisin özvektörleri birbirine diktir (ortogondur). Özdeğerlerin eşit olduğu durumlarda özvektörlerin doğrusal bağımsızlıklarını kaybetmeyecekleri, fakat doğrusal kombinasyonlarla birbirlerine dik olacak şekilde seçilebileceği anlamına gelir. Bu yüzden herhangi iki e_1 ve e_k özvektörleri için, $e_i e_k = 0$ $i \neq k$ eşitliğini verir.

$\Sigma e_k = \lambda_k e_k$ olduğundan eşitliğin her iki tarafını e_i^T ile soldan çarparsak,

$$\text{cov}(Y_1, Y_k) = \text{Cov}(Y_i, Y_k) = e_i^T \Sigma e_k = e_i^T \lambda_k e_k = \lambda_k e_i^T e_k = 0 \quad i \neq k$$

eşitliğini elde ederiz.

Eğer özvektörler ortogonalse (yani birbirine dikse), o zaman iki bileşenin kovaryansı sıfır olur. Her bir temel bileşenin varyansı, o bileşenin özdeğeriyle eşittir. Özdeğerler, verinin doğrusal dönüşümündeki varyansı temsil eder.

Sonuç 2: $[X_1, X_2, \dots, X_p]$ olarak verilen bir rassal vektörle ilişkili olan varyans-kovaryans matrisi Σ olsun. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ olacak şekilde, Σ 'nin özdeğer-özvektör çiftleri $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ olsun ve temel bileşenler $Y_1 = e_1^T X, Y_2 = e_2^T X, \dots, Y_p = e_p^T X$ olmak üzere verilir. Böylece

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^P \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^P \text{Var}(Y_i)$$

eşitliği sağlanır.

İspat: $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \text{tr}(\Sigma)$ dir. (A'nın izi $\text{tr}(A)$), Σ' nin özdeğerlerinden oluşan diagonal bir matris ve P , Σ' nin özdeğerlerine karşılık gelen birim özvektörlerden oluşan ortogonal bir matris olsun. Böylece $P = (e_1, e_2, \dots, e_p)$ ve $PP^T = P^T P = I$ 'dır. Spektral ayrışımından, $\Sigma = PAP^T$ 'dır. Buradan,

$$\begin{aligned} \text{tr}(\Sigma) &= \text{tr}(PAP^T) \\ &= \text{tr}(P^T P) \\ &= \text{tr}(A) \\ &= \lambda_1 + \lambda_2 + \dots + \lambda_p \end{aligned}$$

dir. Böylece

$$\sum_{i=1}^P \text{Var}(X_i) = \text{tr}(\Sigma) = \text{tr}(A) = \sum_{i=1}^P \text{Var}(Y_i)$$

dir. Bu sonuç toplam kitle varyansının $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp}$ olduğunu ve buradan k inci temel bileşenin varyansının toplam varyansa oranının

$$\frac{\lambda_k}{\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp}} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, k = 1, 2, \dots, p$$

olduğunu gösterir.

Temel Bileşenler Analizi kullanarak, verideki büyük varyansı açıklayan birkaç temel bileşen seçilebilir. Eğer bu bileşenler %80 veya %90 gibi büyük bir varyansı açıklıyorsa, orijinal değişkenler yerine bu temel bileşenler kullanılabilir ve çok fazla bilgi kaybı yaşanmaz.

Her temel bileşenin katsayıları (e_{ik}), o bileşenin orijinal değişkenlerle ilişkisini gösterir. Bir temel bileşenin katsayıları ne kadar büyükse, o bileşen o değişkenin varyansını daha fazla açıklar. Bu katsayılar, değişkenler arasındaki korelasyonla da ilişkilidir.

Sonuç 3: Eğer bileşenler $Y_1 = e_1^T X, Y_2 = e_2^T X, \dots, Y_p = e_p^T X$, Σ varyans-kovaryans matrisinden elde edilen temel bileşenler ise

$$\rho_{X_k, Y_i} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} ; \quad i, k = 1, 2, \dots, p$$

X_k rasgele değişkeni ile Y_i temel bileşeni arasındaki korelasyon katsayısıdır. Burada $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ 'ler Σ 'nın özdeğer ve birim özvektör çiftleridir.

İspat: $Corr(X_k, Y_i) = \rho_{X_k, Y_i} = \frac{Cov(X_k, Y_i)}{\sqrt{Var(X_k)}\sqrt{Var(Y_i)}} ; i, k = 1, 2, \dots, p$ dir. $a_k^T = [0, \dots, 0, 1, 0, \dots, 0]$ olsun. Böylece, $a_k^T X = X_k$ 'dır.

$$\begin{aligned} Cov(X_k, Y_i) &= Cov(a_k^T X, e_i^T X) \\ &= a_k^T \Sigma e_i \\ &= a_k^T \lambda_i e_i \\ &= \lambda_i a_k^T e_i \\ &= \lambda_i e_{ik} \end{aligned}$$

olarak elde edilir. Ayrıca $Var(Y_i) = \lambda_i$ ve $Var(X_k) = \sigma_{kk}$ olduğundan

$$\rho_{X_k, Y_i} = \frac{Cov(X_k, Y_i)}{\sqrt{Var(X_k)}\sqrt{Var(Y_i)}} = \frac{\lambda_i e_{ik}}{\sqrt{\lambda_i} \sqrt{\sigma_{kk}}} = \frac{\sqrt{\lambda_i} e_{ik}}{\sqrt{\sigma_{kk}}} ; \quad i, k = 1, 2, \dots, p$$

Temel bileşenler ile değişkenler arasındaki korelasyonlar, genellikle temel bileşenlerin yorumlanmasına yardımcı olmakla birlikte, yalnızca belirli bir temel bileşene (örneğin, Y temel bileşeni) bir değişkenin (X) katkısını ölçer. Bu yaklaşım, diğer değişkenlerin etkilerini göz ardı eder ve yalnızca tek bir değişkenin bileşen içindeki önemini gösterir. Bu nedenle, bazı istatistikçiler, temel bileşenlerin yorumlanmasında korelasyonlar yerine katsayıların kullanılmasını tavsiye etmektedir. Katsayılar ve korelasyonlar, aynı bileşen için değişkenlerin önemini ölçerken farklı sonuçlar verebilse de genellikle bu farklar önemli ölçüde belirgin olmaz. Pratikte, büyük katsayılara sahip değişkenlerin yüksek korelasyonlara sahip olma eğilimi vardır. Bu yüzden, her iki ölçüm türü de temel bileşenlerin yorumlanmasında benzer sonuçlar ortaya koyabilir. Sonuç olarak, her iki ölçüm türü de temel bileşenlerin yorumlanmasına katkı sağlamak amacıyla birlikte incelenebilir.

4.2.3. Temel bileşenlerin korelasyon (standartlaştırılmış değişkenlerden) matrisinden elde edilmesi

Standartlaştırma, veri setindeki değişkenlerin farklı birimlerde ölçülmesi durumunda, bu değişkenleri karşılaştırılabilir hale getirmek için yaygın olarak kullanılan bir yöntemdir. Bu işlemde, her bir değişkenin değeri, o değişkenin standart sapmasına bölünerek normalize edilir. Bu sayede değişkenlerin varyansları küçültülür ve veri setindeki tüm değişkenler aynı ölçü biriminde ifade edilir. Sonuç olarak, dağılım matrisi yerine korelasyon matrisi üzerinde işlem yapılması gereklidir. Standartlaştırma, temel bileşenler analizi uygulanmadan önce, veri setindeki değişkenlerin aynı ölçü biriminde olmasını sağlamak için yapılır ve böylece analizin geçerliliği artırılır. $X^T = (X_1, X_2, \dots, X_p)$ rasgele vektörünün bileşenlerinin ölçü birimleri farklı ise, temel bileşenlerin ölçü biriminden etkisini arındırmak için değişkenlerin standartlaştırılması gerekir. $X^T = (X_1, X_2, \dots, X_p)$ rasgele vektörünün bileşenlerinin standartlaştırılmış vektörü $Z^T = (Z_1, Z_2, \dots, Z_p)$ 'dir. Burada $Z_i = \frac{(X_i - \mu_i)}{\sqrt{\sigma_{ii}}}$; $i = 1, \dots, p$ dir. Matris gösterimiyle $Z = \left(V_z^{-1}\right)^{-1} (X - \mu)$ 'dir. V_z^{-1} köşegen standart sapma matrisidir. Burada V_z^{-1} diagonal elemanları X_i rasgele değişkenlerinin standart sapmalarından oluşan $p \times p$ tipinde bir matrisdir. Ayrıca $E(Z) = 0$ ve $Cov(Z) = \left(V_z^{-1}\right)^{-1} \Sigma \left(V_z^{-1}\right)^{-1} = \rho$ 'dir.

Temel bileşenler analizi uygulandığında, korelasyon matrisinden elde edilen temel bileşenler, standartlaştırılmış değişkenlerden türetilmiş olur. Ancak, temel bileşenlerin elde edilmesinde kullanılan korelasyon matrisinin (ρ) ve varyans-kovaryans matrisinin (Σ) sonuçları birbirinden farklıdır. Özellikle, Σ matrisinden elde edilen özdeğerler ve birim özvektör çiftleri ile ρ matrisinden elde edilen özdeğerler ve birim özvektör çiftleri aynı değildir. Bunun nedeni, korelasyon matrisinin standartlaştırılmış verilerle ilişkili olması ve varyans-kovaryans matrisinin orijinal verilerle ilişkili olmasıdır. Standartlaştırma, her bir değişkenin varyansını birleştirip küçültür ve dolayısıyla her iki matrisin yapısal farklılıklar göstermesine yol açar. Bu da, iki matrisin temel bileşenlerinin farklılık göstermesine neden olur.

Sonuç: $Cov(Z) = \rho$ olan $Z^T = (Z_1, Z_2, \dots, Z_p)$ standartlaştırılmış rasgele değişkenler için i . temel bileşen $Y_i = e_i^T Z = e_i^T \left(V_z^{-1}\right)^{-1} (X - \mu)$; $i = 1, 2, \dots, p$ 'dir. Bununla birlikte

$$\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(Z_i) = \rho$$

$$\rho_{Y_i, Z_k} = e_{ik} \sqrt{\lambda_i}; i, k = 1, 2, \dots, p$$

eşitlikleri bulunur. Bu durumda; $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ 'ler ρ nun özdeğer ve özvektör çiftleridir ve $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ dır. Buradan k. temel bileşenin varyansının toplam varyansa oranı $\frac{\lambda_k}{\sum \lambda_k}$; $k=1, 2, \dots, p$ 'dir. Burada λ_k değerleri ρ matrisinin özdeğerleridir.

4.2.4. Analizde Temel Bileşenlerin Seçimi ve Sayısının Belirlenmesi Yöntemleri

Temel Bileşenler Analizi, yüksek boyutlu verilerin daha düşük boyutlarda, ancak mümkün olduğunca fazla bilgiyi koruyarak temsil edilmesini sağlayan bir tekniktir. Bu süreçte, verinin varyansını en iyi şekilde açıklayan yeni bileşenler (temel bileşenler) elde edilir. Ancak, bu bileşenlerin sayısının belirlenmesi önemlidir, çünkü fazla sayıda bileşen kullanmak gereksiz hesaplama yüküne yol açabilirken, çok az bileşen kullanmak ise önemli bilgilerin kaybına neden olabilir. Bu yüzden TBA'nın temel bileşen sayısını belirlemek için çeşitli yöntemler kullanılır. Bu yöntemler arasında Scree Plot, Kümülatif Varyans Oranı, Kaiser Kuralı, Eigenvalue Yöntemi ve Çapraz Doğrulama gibi teknikler yer alır. Bu yöntemler, verinin iç yapısını en iyi şekilde temsil edecek bileşen sayısını seçmeye yardımcı olur.

4.2.4.1. Açıklanan Varyans Kriteri (Kümülatif Varyans Oranı)

Bu yaklaşım, temel bileşenlerin toplam varyansa katkılarını dikkate alarak, her bir bileşenin açıkladığı varyans oranlarının hesaplanmasını ve bu oranların kümülatif olarak toplanmasını içerir. Bu yöntemle, veri setinin büyük bir kısmını açıklayacak olan bileşenlerin seçilmesi amaçlanır. Genellikle, %80 veya %90 gibi yüksek bir varyans açıklama oranı hedeflenir.

Matematiksel olarak ifade edildiğinde, bu kriter şu şekilde tanımlanabilir:

$$\sum_{j=1}^m \lambda_j \geq p$$

Burada:

- λ_j, j . temel bileşenin özdeğerini,
- p , açıklanan toplam varyans oranını (örneğin %80 veya %90) temsil eder,
- m , kümülatif varyans oranı p 'yi aşan en küçük bileşen sayısını ifade eder.

Bu eşik, kümülatif varyans oranının belirli bir değere (genellikle %80 veya %90) ulaşana kadar bileşenlerin sırasıyla dahil edilmesini sağlar. Bu kriter, temel bileşenlerin sayısını seçerken,

veri setinin açıklayıcı gücünü koruyarak, boyut indirgeme işleminde bilgi kaybını asgariye indirmeyi amaçlar.

4.2.4.2. Kaiser Kuralı (Özdeğerlerin Büyüklüğü)

Temel Bileşenler Analizi yönteminde, veri setindeki önemli yapıları temsil eden temel bileşenlerin sayısını belirlemek için kullanılan yaygın yöntemlerden biri Kaiser Kuralı'dır. Bu kural, temel bileşenlerin özdeğerlerine dayanarak, yalnızca özdeğeri 1'den büyük olan bileşenlerin önemli olduğunu kabul eder. Kaiser Kuralı'nın mantığı, bir bileşenin özdeğerinin, o bileşenin veri setine katkı sağladığı varyans miktarını temsil etmesidir. Bu bağlamda, özdeğeri 1'den küçük olan bileşenler, veri setinin açıklanan varyansına çok küçük bir katkı sağladığı için genellikle analiz dışı bırakılır.

Matematiksel olarak, temel bileşenler analizi sonucunda elde edilen özdeğerler $\lambda_1, \lambda_2, \dots, \lambda_m$ sırasıyla değerlendirilir ve Kaiser Kuralı şu şekilde ifade edilir:

$$\lambda_j > 1 \text{ ise bileşen } j \text{ seçilir.}$$

Burada:

- λ_j j. temel bileşenin özdeğerini ifade eder,
- m, toplam temel bileşen sayısını,
- j, her bir bileşenin sırasını belirtir.

Bu kritere göre, yalnızca özdeğeri 1'den büyük olan bileşenler veri setinin önemli yapıları temsil eden bileşenler olarak kabul edilerek seçilir. Özdeğeri 1'in altındaki bileşenler ise veri setinin varyansını açıklama bakımından yeterli katkıyı sağlamaz ve bu nedenle dışlanır.

Kaiser Kuralı, veriyi daha düşük boyutlara indirgerken, önemli bilgi kaybı yaşamadan, yalnızca anlamlı bileşenlerin seçilmesini sağlar. Bu yaklaşım, temel bileşen sayısının hızlı ve objektif bir şekilde belirlenmesine olanak tanır. Ancak, her ne kadar yaygın bir uygulama olsa da bu kuralın bazı durumlarda veri setinin daha fazla yapısını gözden kaçırabileceği ve alternatif yöntemlerle tamamlanması gerekebilir.

4.2.4.3. Scree Plot (Cattell Scree Test)

Scree Plot, temel bileşen analizi uygulamalarında kullanılan, temel bileşen sayısını belirlemeye yönelik yaygın bir görsel yöntemdir. Bu yöntem, özdeğerlerin büyüklük sırasına göre düzenlendiği ve bu değerlerin bileşen sayısına karşı grafiksel olarak gösterildiği bir eğim grafiği

şeklinde sunulur. Cattell'in önerdiği Scree Testi, bileşen sayısının belirlenmesinde, özdeğerlerin azalma eğilimindeki değişimleri inceleyerek önemli bileşenlerin seçilmesine olanak tanır.

Özdeğerlerin Hesaplanması ve Sıralanması: Temel bileşenler analizi uygulandığında, her bir bileşenin $\lambda_1, \lambda_2, \dots, \lambda_p$ özdeğeri hesaplanır. Özdeğerler, her bir bileşenin veri setindeki varyansın ne kadarını açıkladığını gösterir; bu nedenle, en büyük özdeğerlere sahip bileşenler, veri setinin büyük bir kısmını açıklayan önemli bileşenleri temsil eder.

Grafiksel Temsil (Scree Plot): Hesaplanan özdeğerler, bileşen sayısı $(1, 2, \dots, p)$ ile özdeğerlerin değerleri arasındaki ilişkiyi görsel olarak temsil eden bir çizgi grafik şeklinde sunulur. Burada, x – *ekseni* bileşen numarasını, y – *ekseni* ise her bir bileşenin özdeğerini gösterir. Bu grafik, genellikle özdeğerlerin bileşen sayısına göre nasıl azaldığını ve azalmanın eğilimsel olarak nasıl bir yamaç oluşturduğunu gözler önüne serer.

Eğimdeki Değişim ve "Dirsek" Noktasının Belirlenmesi: Scree plot üzerindeki grafik, başlangıçta özdeğerlerin hızlı bir şekilde azaldığını gösteren dik bir eğimle başlar. Ancak, belirli bir bileşen sayısına ulaşıldığında, eğim belirgin bir şekilde azalır ve grafik daha yatay bir hale gelir. Bu noktada, grafik üzerinde eğim kaymasının başladığı yer olarak tanımlanan dirsek (elbow) noktası gözlemlenir. Dirsek noktası, temel bileşenlerin sayısını belirlemek için kritik bir referans noktasıdır. Bu noktadaki bileşen sayısı, veri setindeki varyansı en yüksek ölçüde açıklayan bileşenlerin sayısını işaret eder.

Temel Bileşen Sayısının Seçimi: Eğimdeki kaymanın başladığı dirsek noktası, anlamlı bileşenlerin sayısını belirlemek için kullanılır. Dirsek noktasından önceki bileşenler, veri setinin büyük kısmını açıklayan bileşenlerdir ve bu bileşenler genellikle önemli olarak kabul edilir. Oysa dirsek noktasından sonraki bileşenler, açıklanan varyansın yalnızca çok küçük bir kısmını sağladığı için önemsiz olarak değerlendirilir.

Scree plot'ta, her bir temel bileşenin özdeğeri λ_j ile ifade edilir. Grafikte gözlemlenen eğim değişikliğinin kaydığı nokta, özdeğerlerin azalma hızının belirgin bir şekilde yavaşladığı ve bu noktada seçilecek bileşenlerin toplam varyansın çoğunu açıklayacak şekilde önemli olduğuna işaret eder. Bu durum matematiksel olarak şu şekilde ifade edilebilir:

$$\lambda_1 > \lambda_2 > \dots > \lambda_p$$

Grafikteki bu sıralama üzerinden yapılan görsel değerlendirme, genellikle dirsek noktasının öncesindeki bileşenleri temel bileşenler olarak kabul eder. Bu noktadaki bileşenler, veri

setindeki varyansı önemli ölçüde açıklayan bileşenlerdir ve temel bileşen sayısının belirlenmesinde referans alınan ana bileşenlerdir (Bulut,2019).

4.2.4.4. Joliffe Kriteri

Korelasyon matrisindeki özdeğerlerin analizi, temel bileşenler analizinde önemli bir adım olup, doğru temel bileşenlerin seçilmesi, modelin açıklayıcı gücünü ve doğruluğunu doğrudan etkileyen bir faktördür. Bu bağlamda, özdeğerlerin belirli bir eşik değeriyle değerlendirilmesi, hangi bileşenlerin anlamlı kabul edileceğini belirlemek için yaygın olarak kullanılan yöntemlerden biridir. Özellikle, Joliffe (1986) tarafından önerilen, özdeğeri 0.7 ve daha büyük olan bileşenlerin dikkate alınması yaklaşımı, temel bileşen seçiminde daha sıkı bir kriter olarak öne çıkmaktadır.

Kaiser ölçütü, temel bileşenlerin seçiminde daha geniş bir yaklaşım benimseyerek, yalnızca özdeğeri 1.0 veya daha büyük olan bileşenlerin önemli kabul edilmesini önerir. Bu ölçüt, modelin fazla bileşen içermesine ve dolayısıyla daha fazla veri içeriği (veya gürültü) barındırmasına neden olabilir. Diğer yandan, Joliffe ölçütü, özdeğeri 0.7 ve üzeri olan temel bileşenleri dikkate almayı önererek, Kaiser ölçütüne kıyasla daha katı bir seçim yapmayı amaçlar. Bu yaklaşım, daha anlamlı ve güvenilir bileşenlerin seçilmesine olanak tanırken, genellikle daha az sayıda bileşenle sonuçlanır.

Joliffe'nin önerdiği eşik değeri, Kaiser ölçütünün genellikle kabul edilen eşik değerinden daha düşük olduğundan, Joliffe kriteriyle seçilen temel bileşen sayısı, Kaiser kriterine göre yaklaşık olarak iki kat daha az olabilmektedir. Bu durum, veri setinin yapısına ve özdeğerlerin dağılımına bağlı olarak, gereksiz bileşenlerin modele dahil edilmesini engelleyerek, modelin daha fazla açıklayıcı güce sahip olmasına katkı sağlayabilir.

Sonuç olarak, Joliffe ölçütü, Kaiser ölçütünden daha az sayıda ancak daha güvenilir ve anlamlı bileşen seçilmesine olanak tanır. Bu da daha sağlam ve anlamlı bir veri modellemesi süreci sağlar, çünkü fazla sayıda bileşenin dahil edilmesi, aşırı uyum (overfitting) ve düşük genellenebilirlik gibi sorunlara yol açabilir (Ünal, 2020).

4.2.5. Analizde Örneklem Büyüklüğünün Yeterliliği

Temel Bileşenler Analizi (TBA) uygulamalarında literatürde örneklem büyüklüğü için belirli bir kurala rastlanmamaktadır. Bununla birlikte, TBA genellikle Çıkarımsal Doğrusal Regresyon Analizi (ÇDRA), Küme Analizi (KA) veya Faktör Analizi (FA) gibi yöntemlerin bir ara adımı olarak kullanılmaktadır. Genel bir bakış açısıyla, TBA için gereken örneklem

büyüklüğü, bu analizlerin ana adımlarına yönelik gerekli örneklem büyüklüğü ile uyumlu olmalıdır (Akbulut ve Çapık, 2022).

Çalışma grubundan elde edilen verilerin açımlayıcı faktör analizi (AFA) için uygun olup olmadığını değerlendirmek, genellikle Kaiser-Meyer-Olkin (KMO) testi ve Bartlett'in küresellik testi ile yapılır. Bu iki test, verilerin faktör analizi için uygun olup olmadığını belirlemek için kullanılır. Her iki testin amacı, verilerdeki korelasyon yapısının analiz için yeterli olup olmadığını ölçmektir (Kaya, 2013).

Kaiser-Meyer-Olkin (KMO) örneklem yeterliliği ölçütü, bir veri setinin faktör analizine uygunluğunu değerlendirmek amacıyla, değişkenler arasındaki izlenen korelasyon katsayılarının büyüklüklerini kısmi korelasyon katsayılarının büyüklükleri ile karşılaştırır. KMO değeri, eşleştirilmiş tüm değişkenler için kısmi korelasyon katsayılarının karelerinin toplamının, gözlenen korelasyon katsayılarının karelerinin toplamına göre daha küçük olduğu durumlarda 1'e yakın değer alır. Bu durum, değişkenler arasındaki ilişkinin büyük ölçüde doğrusal olduğunu ve faktör analizinin uygulanmasının uygun olacağını gösterir. Öte yandan, KMO değeri küçükse, değişkenlerin birbirleriyle yeterli düzeyde ilişkili olmadığı ve bu nedenle temel bileşenler analizinin yapılmasının önerilmediği anlamına gelir. KMO ölçütü şu şekilde yorumlanır: 0.90-1.00 arası mükemmel, 0.80-0.89 arası çok iyi, 0.70-0.79 arası iyi, 0.60-0.69 arası orta, 0.50-0.59 arası zayıf ve 0.50'nin altındaki değerler kabul edilemez olarak değerlendirilir.

Değişkenler arasındaki ilişkinin gücünü ölçen bir diğer gösterge ise kısmi korelasyon katsayılarıdır. Bir değişkenin doğrusal etkinin kaldırılması durumunda iki değişken arasındaki kısmi korelasyon katsayısının küçük olması, bu değişkenlerin ortak bir yapıyı paylaştığını ve dolayısıyla faktör analizine uygun olduğunu gösterir. Kısmi korelasyon katsayılarının negatif değeri ise “anti-imaj korelasyon” (anti-image correlation) olarak adlandırılmaktadır.

KMO ölçütü, bir bütün olarak soru grubunun genel analizle uyumluluğunu incelerken, Measures of Sampling Adequacy (MSA) değeri her bir değişkenin faktör analizine uygunluğunu ayrı ayrı belirler. MSA değerlerinin yorumlanması da KMO değeri ile paralellik göstermektedir (Alkan, 2008).

4.3. TEMEL BİLEŞENLER ANALİZİ UYGULAMA ADIMLARI

Temel Bileşenler Analiz çok boyutlu bir veri setini daha düşük boyutlu bir hale indirmek amacıyla kullanılan bir tekniktir. TBA, verideki en fazla varyansı açıklayan bileşenleri bulmaya çalışır ve bu sayede verinin daha anlamlı bir şekilde temsil edilmesini sağlar. Aşağıda TBA'nın adım adım nasıl uygulandığını açıklayan bir süreç bulunmaktadır:

4.3.1. Veri Setinin Hazırlanması

TBA uygulamadan önce verilerin uygun bir şekilde hazırlanması gerekir. Bu aşama genellikle şu adımları içerir:

Eksik Verilerin Düzenlenmesi: Eğer veri setinde eksik veriler varsa, bu eksiklikler bir ortalama, medyan veya diğer uygun yöntemlerle tamamlanmalıdır.

Veri Temizliği: Veri setindeki yanlış ya da tutarsız değerler düzeltilmelidir.

Veri Formatı: Verilerin sayısal formatta olması gerekir. Kategorik veriler, uygun sayısal değerlere dönüştürülmelidir.

Bu adımlar tamamlandığında, veriler TBA için hazır hale gelir.

4.3.2. Ortalamadan Çıkarma (Veri Standardizasyonu)

TBA uygulamasının bir diğer önemli adımı, verilerin standartlaştırılmasıdır. Çünkü TBA, her bir değişkenin varyansını dikkate alır ve farklı ölçeklerdeki değişkenlerin karşılaştırılabilir olabilmesi için tüm değişkenlerin aynı ölçeğe getirilmesi gerekir.

Her bir değişkenin ortalaması çıkarılır ve veriler z-skoru formülüyle dönüştürülür. Bu işlem, her bir değişkenin ortalamasının sıfır ve standart sapmasının bir olmasını sağlar.

$$Z = \frac{X - \mu}{\sigma}$$

Burada:

- X gözlemlenen veri değeri,
- μ o değişkenin ortalaması,
- σ o değişkenin standart sapmasıdır.

4.3.3. Kovaryans Matrisinin Hesaplanması

Veri setindeki her bir değişkenin varyansını ve değişkenler arasındaki kovaryansı görmek için kovaryans matrisi oluşturulmalıdır. Kovaryans, iki değişkenin nasıl birlikte değiştiğini gösterir.

Kovaryans Matrisinin Hesaplanması: Veri setindeki her iki değişkenin kovaryansı hesaplanarak bir kovaryans matrisi oluşturulur. Örneğin, veri setinde iki değişken X ve Y varsa, kovaryans şu şekilde hesaplanır:

$$cov(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

Bu matris, değişkenlerin birlikte nasıl hareket ettiklerini gösterir ve TBA için önemli bir adımdır.

4.3.4. Özdeğerler Ve Özvektörlerin Hesaplanması

Bir sonraki adım, özdeğerler ve özvektörlerin hesaplanmasıdır. Özdeğerler, her bir bileşenin (principal component) verinin toplam varyansını ne kadar açıkladığını gösterirken, özvektörler ise bu bileşenlerin yönlerini belirtir.

Özdeğerlerin ve Özvektörlerin Hesaplanması: Kovaryans matrisinden özdeğerler ve özvektörler hesaplanır. Özdeğerler, verinin ne kadar varyansı açıkladığını gösterirken, özvektörler ise bu bileşenlerin doğrultusunu gösterir. Bu işlem, genellikle lineer cebir kullanılarak yapılır. Özdeğerler ve özvektörler aşağıdaki denklemle elde edilebilir:

$$C \cdot v = \lambda \cdot v$$

Burada:

- C kovaryans matrisidir,
- v özvektör,
- λ ise özdeğerdir.

4.3.5. Bileşenlerin Seçilmesi Ve Özellik Vektörünün Oluşturulması

Bu adımda, hangi temel bileşenlerin (principal components) seçileceği belirlenir. Genellikle en yüksek özdeğerler ile ilişkilendirilen bileşenler seçilir, çünkü bu bileşenler verinin en fazla varyansını açıklamaktadır.

Bileşen Seçimi: En büyük özdeğerlere sahip olan özvektörler seçilir. Bu seçilen özvektörler, temel bileşenlerin doğrultusunu temsil eder.

Özellik Vektörünün Oluşturulması: Seçilen özvektörler birleştirilerek bir özellik vektörü oluşturulur. Bu vektör, verinin yeni bir alt uzaya projeksiyonunu sağlar.

4.3.6. Yeni Veri Setinin Türetilmesi

Son olarak, elde edilen özellik vektörü kullanılarak veriler yeni bir boyutta temsil edilir. Bu işlem, verinin projeksiyonu olarak bilinir.

Yeni Veri Setinin Oluşturulması: Orijinal veriler, seçilen özvektörlerle çarpılarak yeni bir veri seti elde edilir. Bu yeni veri seti, boyutları indirgenmiş ancak en fazla varyansı açıklayan bileşenlerden oluşur.

$$Y = X \cdot V$$

Burada:

- Y yeni veri setidir,
- X orijinal veri setidir,
- V seçilen özvektörlerin matrisidir.

Bu yeni veri seti, daha az sayıda boyutla aynı bilgiyi taşır ve analiz için daha uygun hale gelir.

4.3.7. Sonuçların Değerlendirilmesi ve Yorumlanması

TBA uygulandıktan sonra elde edilen yeni veriler ve bileşenler değerlendirilmelidir:

Bileşenlerin Anlamı: Seçilen bileşenlerin veriye kattığı anlamı yorumlamak önemlidir. Bileşenlerin her biri, orijinal değişkenlerin bir kombinasyonudur.

Varyansın Açıklanması: Hangi bileşenin ne kadar varyans açıkladığı, modelin ne kadar etkili olduğuna dair bilgi verir. Genellikle ilk birkaç bileşen çoğu varyansı açıklar.

5. REYTING SİSTEMİNİN TEKNİK VE SOSYOLOJİK ANALİZİ

Bu bölümde, reyting sistemlerinin hem teknik hem de sosyolojik açıdan analizi yapılmaktadır. Teknik açıdan, reyting sistemlerinin nasıl yapılandırıldığı, algoritmaların nasıl işlediği ve veri toplama süreçlerinin nasıl tasarlandığı incelenmektedir.

5.1. REYTING

Kitle iletişim araçları, gazete ve dergilerin yayına başladığı 19. yüzyıldan itibaren toplumun bilgi edinme gereksinimini karşılamıştır. 20. yüzyıla gelindiğinde ise bu rol, gazete ve dergilerden radyoya geçmiştir. 1950’li yıllardan itibaren ise radyo, yerini televizyona bırakmıştır. Televizyon, 20. yüzyılın ikinci yarısından itibaren en yaygın kullanılan kitle iletişim aracı olarak toplum üzerinde büyük bir etki yaratmış ve kitleleri yönlendirme kapasitesine sahip bir güç haline gelmiştir. Özellikle reklamcılık sektöründe televizyonun etkili bir mecra olduğunun fark edilmesi, izleyici kitlesinin ölçülmesi gerekliliğini gündeme getirmiş ve bu bağlamda reyting kavramı ortaya çıkmıştır.

Genel anlamda ölçümleme, bir ölçme işlemi sonucunda elde edilen verilerin sayısal ya da sembolik ifadelerle aktarılmasıdır. Çeşitli alanlarda uygulanabilen ve belirleyici bir kriter olarak değerlendirilen ölçümleme, televizyon yayınlarının hitap ettiği izleyici kitlesi tarafından ilgi çekici olup olmadığını değerlendiren nicel veriler sunmaktadır (Kuyucu,2020).

İzleyici ölçümü; izlenme oranı, izlenme payı ve benzeri verilerin elde edilmesi amacıyla gerçekleştirilen tüm ölçüm süreçlerini ifade eder. Reyting (izleyici oranı), belirli bir program ya da zaman diliminde her dakikaya düşen ortalama izleyici yüzdesini temsil eden bir göstergedir. İzlenme payı ise bir televizyon kanalının belirli bir zaman aralığında toplam izleyici kitlesinden aldığı payı ifade eder (Özbek vd., 2011: 89).

5.2. REYTING ÖLÇÜM SİSTEMİ VE TÜRKİYEDEKİ REYTING SİSTEMİ

Reyting ölçüm sistemi, televizyon kanallarının ve programlarının izlenme oranlarını belirlemek amacıyla geliştirilmiş, tüketici davranışlarını analiz ederek üreticilere pazarlama stratejileri sunan bir araçtır. İlk olarak 1963’te ABD’de ve 1966’da İngiltere’de uygulanmaya başlanmıştır.

Reyting ölçümleri, belirli bir örneklem grubunun televizyon cihazlarına takılan “peoplemeter” (insan ölçer) adı verilen cihazlar aracılığıyla gerçekleştirilir. Bu sistemde, örneklem

grupları, sosyo-ekonomik göstergelere göre seçilir ve ölçüm sonuçları izleyicilerin saatlik izleme alışkanlıklarına dayalı olarak A-B, C, diğer ve total gibi kategorilerde raporlanır. Örneklem grubu tüm izleyicileri temsil etmediği için sistemin genelleyici yapısı tartışma konusu olsada, bu bireylerin toplumsal kodlarının reyting sonuçlarına yansıdığı kabul edilmektedir.

Peoplemeter teknolojisinin yanı sıra, izleme ölçümlerinin tarihi radyo dönemine kadar uzanır. İlk uygulamalar telefon ve anket yöntemleri gibi araçlara dayanırken, günümüzde taşınabilir peoplemeter cihazlarıyla çoklu ölçüm yöntemlerine evrilmiştir. Dünyada izleyici ölçümlerini gerçekleştiren organizasyonel yapılar üç ana başlık altında toplanmaktadır:

- Araştırma Şirketine Ait Sistemler (Own Service-OS)
- Medya Sahipleri Komitesi (Media Owner Contract-MOC)
- Müşterek Endüstri Komitesi (Joint Industry Committee-JIC) (Aydın, 2018).

Türkiye’de izleyici ölçümü araştırmaları 1989 yılından bu yana gerçekleştirilmektedir. 1992 yılından itibaren ise bu süreç, Uluslararası Reklamcılık Derneği (IAA) bünyesinde kurulan TİAK tarafından yürütülmektedir. TİAK, yapısal olarak bir Ortak Endüstri Komitesi (OEK) niteliği taşımaktadır ve Reklam Verenler Derneği (RVD), Reklamcılar Derneği (RD) ile televizyon kanallarını kapsamaktadır. İzleyici örnekleminin seçimi, kontrolü, analiz edilmesi ve raporlanması gibi süreçler TİAK tarafından organize edilmektedir (Emanet,2017).

TİAK, televizyon izleme verilerinin sağlanması ve analiz edilmesi süreçlerinde stratejik kararlar alabilen, yazılı kurallar çerçevesinde sektör katılımcılarını bir araya getiren bir organizasyondur. Reyting ölçümleri, yalnızca televizyon programlarının izlenme oranlarını değil, aynı zamanda toplumsal normlar ve medya içeriklerini de şekillendiren bir araç olarak, medya ve toplum arasındaki karşılıklı etkileşimi anlamada önemli bir yere sahiptir (Aydın, 2018).

5.3. TİAK VERİLERİNİN YAPISI VE ANLAMI

Televizyon İzleme Araştırmaları Anonim Şirketi (TİAK A.Ş.), televizyon izleme ölçümlerini organize etmek ve denetlemek amacıyla Aralık 2010’da kurulmuş bir anonim şirkettir. TİAK A.Ş., televizyon izleme araştırmaları alanında, sektörün farklı paydaşlarını (reklam verenler, medya planlama ajansları ve televizyon kanalları) bir araya getiren, tüm tarafların dünya standartlarına uygun bir platformda fikir birliğiyle hareket etmelerini sağlayan kurumsal bir yapı sunmaktadır.

TİAK A.Ş.’nin temelleri, 1992 yılında gönüllülük esasına dayalı olarak faaliyet göstermeye başlayan bir komiteye dayanmaktadır. Bu komite, televizyon izleme araştırmalarını organize etmiş

ve denetlemiştir. Ancak, Rekabet Kurulu'nun 2 Kasım 2009 tarihli raporu doğrultusunda, 2010 yılında anonim şirket statüsüne geçerek yeniden yapılandırılmıştır.

TİAK'ın 4 ana görev ve sorumlulukları vardır.

- Televizyon İzleme Ölçümlerini Yürütme: Ölçüm faaliyetlerini, yüklenici firmalar aracılığıyla organize eder ve denetler.
- Kurallara Uygunluk: Televizyon izlenme ölçümlerinin, objektif olarak belirlenmiş kurallara ve dünya ölçüm standartlarına uygun şekilde gerçekleştirilmesini sağlar.
- Paydaş Temsiliyeti: İlgili tarafların (reklam verenler, medya planlama ajansları, televizyon kanalları) ortak platformda temsilini ve iş birliğini mümkün kılar.
- Şeffaflık ve Güvenilirlik: Ölçüm süreçlerini denetleyerek tarafsız ve güvenilir veri üretimini güvence altına alır.

TİAK A.Ş., televizyon izleme ölçüm süreçlerini yalnızca teknik bir faaliyet olarak değil, aynı zamanda sektörel iş birliği ve standardizasyon açısından stratejik bir alan olarak ele almaktadır. Bu yönüyle medya sektörü ve reklam dünyasında önemli bir role sahiptir.

5.4. REYTINGİN MEDYA İÇERİKLERİ VE TOPLUMSAL NORMLAR ÜZERİNDEKİ ETKİSİ

Reytinglerin medya içerikleri üzerindeki etkisi, toplumsal normların şekillenmesinde önemli bir rol oynamaktadır. Medya, ticari bir sektör olarak yüksek reyting elde etmek amacıyla içeriklerini sürekli olarak izleyici taleplerine göre şekillendirir. Bu durum, özellikle dramatik, duygusal ya da tartışmalı içeriklerin ön plana çıkmasına neden olabilir. Medyanın bu ticari odaklı yapısı, içeriklerin toplumda norm haline gelmiş değerleri yansıttığı gibi, aynı zamanda bu değerleri dönüştürebilir veya pekiştirebilir. Reytinglerin artırılması amacıyla seçilen içerikler, toplumsal davranışları da etkileyebilir; örneğin, şiddet, cinsiyet rolleri ve aile yapısı gibi konular medya aracılığıyla topluma sunulur ve bu normlar zaman içinde daha geniş bir kitleye yayılabilir.

Medyanın toplumsal normlar üzerindeki etkisi, içeriklerin toplumun değerleriyle doğrudan ilişkilidir. Özellikle reyting odaklı içeriklerin yaygınlaşması, toplumda norm haline gelmiş sosyal değerlerin medya aracılığıyla pekişmesini sağlayabilir. Bunun yanı sıra, medyanın toplumsal algıya etkisi de önemlidir, çünkü medya içerikleri, bireylerin toplum hakkındaki anlayışlarını şekillendirebilir. Bu etkileşim, toplumsal yapının zamanla nasıl değiştiğini gösteren önemli bir göstergedir. Bununla birlikte, medyanın etik sorumluluğu da göz ardı edilmemelidir. Medya

şirketlerinin içerik üretirken toplumsal normlara karşı duyarlı olması gerektiği tartışmaları, özellikle ticari kaygıların toplumsal etkilerle çeliştiği durumlarda daha belirgin hale gelir.

Reytinglerin medya içerikleri üzerinde oluşturduğu kültürel çeşitlilik algısı, farklı toplumsal grupların, etnik kimliklerin ve cinsiyet rollerinin temsil edilme biçimlerini doğrudan etkiler. Medya, kültürel çeşitliliği yansıtarak toplumsal normların şekillenmesinde önemli bir araçtır. Ancak popüler kültürün medya aracılığıyla yayılması, bazen toplumsal normları sadece yansıtmaktan öte, yeniden üretmelerine de neden olabilir. Bu bağlamda, reytinglerin arttırılması amacıyla yapılan içerik tercihlerinin toplumsal değerlerin ve normların evrimindeki etkileri, uzun vadede daha belirgin hale gelir.

Son olarak, medya içeriklerinin sosyal cinsiyet ve kimlik inşası üzerindeki etkisi de göz ardı edilemez. Medyanın sunduğu içerikler, cinsiyet rollerinin yeniden üretimine katkı sağlayabilir ve bu da toplumsal normları şekillendiren güçlü bir araç olabilir. Reytinglerin toplumsal kimlikler üzerinde yarattığı etki, bireylerin kendilerini nasıl tanımladıkları ve bu kimliklerin toplumsal yapı içinde nasıl yer aldığı konusunda önemli bir belirleyicidir. Medyanın, toplumsal eşitsizlikleri ve ayrımcılığı pekiştiren içeriklere yer vermesi, bu tür sosyal yapıları daha görünür kılabilir ve bu içeriklerin yaygınlaşması, toplumda daha geniş kesimler tarafından içselleştirilebilir.

6. UYGULAMA: TV KANALI REYTING VERİLERİ ÜZERİNDE TEMEL BİLEŞENLER ANALİZİ (TBA) UYGULAMASI

Bu bölümde temel bileşenler analizi ve uygulama alanındaki TV kanalı reyting verileri konusunun metodolojisi, veri setinin özellikleri, temel bileşenler analizi için veri hazırlama süreci ve reyting veri setindeki temel bileşenler ele alınacaktır.

6.1. ARAŞTIRMANIN METODOLOJİSİ

Çalışmanın temel araştırma yöntemi, kullanılan veri setleri, analiz teknikleri ve uygulama süreçleri detaylı olarak açıklanacaktır. Araştırmanın metodolojisi, veri analizi ve yorumlama süreçlerini sistematik bir çerçeveye oturtmak amacıyla yapılandırılmıştır.

6.1.1. Araştırmanın Konusu

Bu tez çalışması, çok değişkenli istatistiksel analiz yöntemlerinden biri olan Temel Bileşenler Analizi'nin (TBA) teorik altyapısını detaylı bir şekilde incelemek ve bu yöntemin pratik bir uygulama alanı olarak televizyon reyting verilerine uygulanmasını ele almaktadır. Çalışmada, TBA'nın boyut indirgeme ve veri analizindeki güçlü yönleri, Türkiye'deki televizyon izleyici davranışlarını temsil eden demografik reyting verileri üzerinde incelemeye alınmıştır.

6.1.2. Araştırmanın Amacı

Bu çalışmanın amacı, Temel Bileşenler Analizi (TBA) yönteminin teorik ve pratik yönlerini ortaya koyarak, televizyon reyting verileri gibi çok boyutlu veri setlerinde nasıl etkili bir şekilde uygulanabileceğini göstermektir. Özellikle TİAK AB ve ABC1 izleyici gruplarına ait reyting verileri üzerinde yapılan uygulamalar ile, TBA'nın boyut indirgeme sürecinde veri yapısını sadeleştirme, ilişkileri daha net bir şekilde ortaya koyma ve yorumlanabilirlik sağlamadaki katkıları değerlendirilmiştir. Bu bağlamda, çalışmanın temel amacı, TBA'nın hem istatistiksel bir araç hem de medya araştırmaları için stratejik bir analiz yöntemi olarak kullanılabileceğini kanıtlamaktır.

6.1.3. Araştırmanın Problemi

Reyting verileri gibi çok boyutlu ve karmaşık veri setlerinin doğru bir şekilde analiz edilmesi, medya içerik stratejilerinin geliştirilmesi ve hedef kitlelerin daha iyi anlaşılması açısından kritik bir öneme sahiptir. Ancak bu tür çok boyutlu verilerin geleneksel yöntemlerle analiz edilmesi, verideki ilişkilerin ve anlamlı desenlerin ortaya çıkarılmasını zorlaştırmaktadır. Bu çalışmanın problemi, televizyon reyting verilerinin analizinde çok boyutluluğun yarattığı karmaşıklığın, Temel Bileşenler Analizi gibi modern veri analizi yöntemleriyle nasıl çözülebileceğini araştırmaktır.

6.1.4. Araştırmanın Yöntemi

Bu çalışma, nicel araştırma yöntemlerinden biri olan Temel Bileşenler Analizi (Principal Component Analysis - PCA) yöntemini temel almaktadır. TBA, çok boyutlu veri setlerini analiz etmek ve varyansın büyük bir kısmını açıklayan bileşenleri belirlemek için kullanılan istatistiksel bir tekniktir. Araştırma, Türkiye televizyon kanallarının reyting verilerini ele alarak bu yöntemin uygulanmasını ve sonuçlarının yorumlanmasını kapsamaktadır.

TV reyting verileri üzerinde Temel Bileşenler Analizi (TBA) uygulamasını detaylı bir şekilde ele alınacaktır. Elimizdeki veri seti, farklı zaman dilimlerinde (prime time, total day, vb.) ve demografik gruplara göre (AB, 20+ABC1) kanal reytinglerini içermektedir. Bu veriler, kanal adı, reyting yüzdesi (RTG %), pazar payı (share) gibi özellikleri içermektedir. TBA, bu verilerin boyutunu indirgemek ve temel bileşenlerle veriyi daha anlaşılır hale getirmek amacıyla kullanılacaktır.

İlk adımda veriler normalize edilerek, TBA'nın doğru bir şekilde çalışabilmesi için gerekli olan ön işleme adımları yapılmıştır. Daha sonra, kovaryans matrisi hesaplanmış ve özdeğerler ile özvektörler çıkarılarak temel bileşenler belirlenmiştir. Bu bileşenler, verinin varyansını en iyi şekilde açıklamak için seçilmiştir.

TBA ile elde edilen bileşenler üzerinde yapılan analizler, farklı zaman dilimlerinin, demografik grupların ve kanal türlerinin reytingler üzerindeki etkisini ortaya koymaktadır.

Türkiye'de televizyon izleyicilerinin davranışlarını analiz etmek ve televizyon reyting verilerini anlamak amacıyla TV İzleyici Ölçümü kapsamında gerçekleştirilen bir çalışmadır. Araştırma, Türkiye İstatistiksel Araştırmalar Kurumu (TİAK) tarafından toplanan televizyon izleme verilerinin analizine dayanmakta olup, Temel Bileşenler Analizi (TBA) kullanılarak

verilerin temel bileşenleri belirlenmiştir. Aşağıda, kullanılan veri toplama yöntemleri, örneklem seçimi ve veri analiz yöntemleri ayrıntılı olarak açıklanmıştır.

6.1.5. Veri Toplama Yöntemi

Veri toplama, TİAK A.Ş. tarafından sağlanan televizyon izleme verileriyle yapılmıştır. TİAK, 40 ildeki hanelerden gelen televizyon izleme verilerini 365 gün 7/24 aralıksız bir şekilde toplayarak günlük, aylık ve yıllık raporlar oluşturur. Ölçüm, PeopleMeter cihazları kullanılarak gerçekleştirilmiştir. Bu cihazlar, her bir hane bireyinin hangi kanalı izlediğini ve izleme süresini kaydederek, televizyon izleme davranışlarını pasif olarak izler. Panelistler, izledikleri programları belirtmek için uzaktan kumanda aracılığıyla kendi tanımlanmış tuşlara basar, böylece her bir izleyiciye ait izleme verisi tanımlanabilir.

6.1.6. Örneklem Seçimi

Örneklem, TÜİK Nüfus İstatistikleri ve TİAK tarafından yapılan veri tabanı araştırmasına dayalı olarak seçilen çok aşamalı, tesadüfi örnekleme yöntemiyle belirlenmiştir. Araştırmanın örnekleme, 40 ilde yer alan, en az bir çalışan televizyonu olan haneler ve bu hanelerde yaşayan 5 yaş ve üzeri kişiler ile haneleri ziyaret eden ve izleme yapan misafirleri içermektedir. Evren, her ilde yerleşim yeri alt sınırının 10.000 nüfus olduğu yerleşim yerlerinden seçilmiştir. Haneler, yalnızca TÜİK tarafından tesadüfi olarak seçilmiş ve veri tabanı anketi yapılmış olan hanelerden oluşmaktadır. Bu seçim, TİAK'ın ölçüm sistemine dahil olan haneleri oluşturmuş ve yaklaşık olarak 14.000 kişi ile 750 günlük ziyaretçi üzerinden veri toplanmıştır.

6.1.7. Veri Toplama Aracı

Televizyon izleme verileri, PeopleMeter cihazları aracılığıyla toplanmaktadır. Bu cihaz, evde izlenen her programı ve kanalın izlenme oranlarını kaydederek, günlük izleme süreleri, izlenen kanal, izleyici grubu (cinsiyet, yaş, sosyo-ekonomik durum) gibi çeşitli bilgileri kaydeder. Panelistlerin izleme davranışları, cihazın uzaktan kumandasındaki tuşlar aracılığıyla kaydedilir, böylece hangi hane bireyinin hangi kanalı izlediği tespit edilebilir.

6.1.8. Örneklem Büyüklüğü

Araştırma kapsamında, günlük olarak elde edilen TV izleme verileri Brüt 4.400 / Net 4.000 hane ve bu hanelere ait yaklaşık 14.000 kişi üzerinden toplanmaktadır. Ayrıca, bu haneleri ziyaret eden ve televizyon izleyen yaklaşık 750 misafir de veri toplama sürecine dahil edilmiştir. Bu büyük örneklem, Türkiye'deki televizyon izleme davranışlarını temsil eden kapsamlı bir veri seti sunmaktadır.

6.2. VERİ SETİ VE ÖZELLİKLER

Bu çalışmada kullanılan veri seti, televizyon kanal reytinglerine ilişkin çeşitli parametreleri içermekte olup, bu veriler zaman dilimleri (timebands) ve demografik gruplara göre düzenlenmiştir. Veri seti, kanal isimleri, reyting yüzdesi (RTG), pazar payı (SHARE), yayınlanan programların izlenme oranları ve demografik grupların televizyon izleme alışkanlıklarına dair detaylı bilgileri barındırmaktadır. Bu veriler, kanal performansını ve izleyici davranışlarını incelemek amacıyla kapsamlı bir şekilde analiz edilecektir.

6.2.1. Zaman Dilimleri (Timebands):

Veri setinde yer alan zaman dilimleri, TV yayınlarının farklı saat dilimlerinde izlenme oranlarının nasıl değiştiğini ve bu zaman dilimlerinin reytingler üzerindeki etkisini anlamaya olanak tanımaktadır. Zaman dilimlerinin her biri, belirli izleyici davranışlarını yansıtmaktadır. Bu çalışmada, aşağıdaki ana zaman dilimlerine odaklanılacaktır:

Prime Time (20:00-23:00): Bu dilim, televizyon izleyicisinin en yoğun olduğu ve en yüksek reytinglerin elde edildiği saat dilimidir. Bu dilimde, genellikle daha büyük bütçeli ve dikkatlice seçilmiş programlar yayınlanmaktadır.

Total Day: Günün tamamında kanalın aldığı reytingler ve pazar payı oranları toplamını ifade eder. Bu, kanalın günlük izlenme alışkanlıkları hakkında genel bir bakış sağlar.

OPT (07:00-20:00): Bu dilim, sabah saatlerinden akşam saatlerine kadar olan zamanı kapsar ve genellikle sabah programları, haber bültenleri ve gündüz kuşağı programlarının izlendiği dönemi yansıtır.

OPT 07:00-18:00 ve Diğer Alt Dilimler: Bu alt dilimler, belirli saat dilimlerinin daha ayrıntılı analizini yapmamıza olanak tanır. Örneğin, 18:00-25:00 veya 18:00-24:00 gibi daha dar dilimler, izleyici davranışlarının zamanla nasıl evrildiğini görmek için kullanılmaktadır.

Zaman dilimlerinin belirlenmesi, farklı saatlerdeki izleyici kitlesinin programlara gösterdiği ilgiyi daha iyi anlamamıza yardımcı olur. Ayrıca, Prime Time gibi stratejik zaman dilimlerinde daha yüksek reytingler ve pazar payları beklenirken, daha az popüler saat dilimlerinde düşük performanslar gözlemlenebilecektir.

6.2.2. Demografik Gruplar:

Veri setinde bulunan demografik gruplar, izleyicilerin yaş, gelir düzeyi, eğitim seviyesi gibi özelliklerine göre ayrılmaktadır. Bu tür demografik segmentasyonlar, televizyon izleme alışkanlıklarını daha derinlemesine incelemek ve farklı izleyici gruplarına yönelik içerik stratejileri geliştirmek için önemlidir. Çalışmada öne çıkan iki ana demografik grup şunlardır:

AB Grubu (A, B): Genellikle yüksek gelir seviyesine sahip, eğitilmiş ve daha çok şehirli izleyicilerden oluşan bu grup, televizyon yayıncıları için önemli bir hedef kitledir. Bu izleyici grubunun televizyon izleme alışkanlıkları, diğer gruplara kıyasla farklılıklar gösterebilir.

20+ABC1 Grubu: Genç yetişkinlerden oluşan bu grup, özellikle popüler kültür ve sosyal medya etkileşimlerinin etkisiyle televizyon izleme alışkanlıklarında belirli eğilimler sergileyebilir. ABC1 sınıfı, genellikle daha yüksek gelirli ve eğitilmiş bireylerden oluştuğundan, reklam stratejileri açısından önemli bir hedef kitleyi oluşturur.

Bu demografik grupların analizi, televizyon programlarının izleyici kitlesine göre özelleştirilmesini sağlar. Örneğin, Prime Time diliminde AB grubu izleyicilerinin daha fazla ilgisini çeken programlar, içerik stratejilerinin bu kitleye yönelik özelleştirilmesini gerektirebilir.

6.2.3. Kanal Adları ve Programlar:

Veri seti, TV kanal adları ve her kanalın reyting yüzdesi (RTG) ile birlikte pazar payı (share) verilerini de içermektedir. Bu unsurlar, kanal performansının ve izleyici kitlesinin büyüklüğünün değerlendirilmesi açısından önemli göstergelerdir. Kanal adları, her kanalın yayınladığı programların izlenme oranları ile birlikte, kanalın pazar içerisindeki yerini ve izleyici kitlesinin büyüklüğünü ortaya koymaktadır.

Reyting Yüzdesi (RTG), bir kanalın belirli bir zaman dilimindeki izlenme oranını yansıtır. Bu metrik, bir kanalın izleyici kitlesi ile ne kadar etkileşimde olduğunu gösteren bir göstergedir.

Pazar Payı (Share), kanalın genel izleyici kitlesi içindeki payını ifade eder ve kanalın pazardaki rekabet gücünü değerlendirmek için kullanılır.

Bu değişkenler, kanal performansını analiz etmek ve programların hangi zaman dilimlerinde daha fazla ilgi gördüğünü belirlemek için kullanılacaktır. Özellikle, kanal isimleri ve

izlenme oranları arasındaki ilişkiyi çözümlemek, hangi kanalların belirli zaman dilimlerinde öne çıktığını anlamamıza olanak sağlayacaktır.

6.2.4. Temel Bileşenler Analizi İçin Seçilen Değişkenler:

Bu çalışmanın amacı, temel bileşenler analizi kullanarak televizyon kanal reyting verilerinin temel yapılarını ortaya çıkarmaktır. TBA, yüksek boyutlu veriyi daha düşük boyutlara indirmek ve verinin temel bileşenlerini keşfetmek için etkin bir yöntemdir. Bu bağlamda, aşağıdaki değişkenler analiz için seçilmiştir:

Zaman Dilimleri: Farklı zaman dilimlerinde elde edilen reyting ve pazar payı verileri, izleyici davranışlarının zamanla nasıl değiştiğini ve hangi dilimlerin daha fazla ilgi gördüğünü anlamaya yardımcı olacaktır.

Kanal Adları: Her kanalın izlenme oranları ve pazar payı, kanalın izleyici kitlesine yönelik içerik stratejilerini anlamamıza olanak tanıyacaktır.

Demografik Gruplar: Farklı demografik gruplara ait izleme alışkanlıkları, televizyon yayıncılarının hedef kitlesine yönelik içerik planlamaları yapabilmeleri için kritik öneme sahiptir.

Reyting Yüzdesi ve Pazar Payı (RTG ve Share): Bu metrikler, kanal performansını değerlendirmek ve hangi faktörlerin reytingleri daha fazla etkilediğini belirlemek için temel ölçütlerdir.

6.3. TEMEL BİLEŞENLER ANALİZİ İÇİN VERİ HAZIRLAMA SÜRECİ

6.3.1. Verinin Hazırlanması

Temel Bileşenler Analizi uygulamadan önce veri setinin doğru şekilde hazırlanması kritik bir adımdır. Bu süreçte kullanılan veri seti, TİAK AB grubunun televizyon izleme oranlarını içermektedir. Veriler, farklı zaman dilimlerine ait izlenme yüzdeleri (RTG % değerleri) ve kanal adlarını içeren sütunlardan oluşmaktadır. Çalışmanın amacı, zaman dilimlerine göre kanallar arasında varyansı açıklayan bileşenleri belirlemektir. Bu çalışmada, “tiakab.xlsx” isimli veri seti kullanılmıştır. Bu veri seti, çeşitli zaman dilimlerinde televizyon izleme oranlarını (RTG) içermektedir. Veri setinde yedi farklı zaman dilimi bulunmaktadır:

1. Toplam Günlük RTG (Total Day RTG %)
2. (20:00-23:00) RTG (%)
3. (07:00-20:00) RTG (%)
4. (07:00-18:00) RTG (%)
5. (18:00-25:00) RTG (%)

6. (18:00-24:00) RTG (%)
7. (19:00-24:00) RTG (%)

Bu zaman dilimlerini içeren veriler üzerinden **PCA** (Principal Component Analysis) uygulanacaktır.

6.3.2. Verinin Yüklenmesi ve Temizlenmesi

Veri seti Python ortamında pandas kütüphanesi kullanılarak yüklenmiştir. Yükleme işlemi sırasında eksik veya hatalı veri olup olmadığı kontrol edilmiş, gerekli durumlarda eksik veriler temizlenmiştir. TBA için hangi değişkenlerin analiz kapsamına alınacağı seçilmiştir. Analiz için, sadece RTG sütunları ve kanal adları çalışmaya dahil edilmiştir.

```
import pandas as pd

# Veriyi yükle
df = pd.read_excel("tiakab.xlsx")

# RTG sütunlarını seç
rtg_columns = [col for col in df.columns if 'RTG' in col]
rtg_data = df[rtg_columns]

# KANAL sütununu da almak
channel_names = df['KANAL'].values
```

Şekil 6-1:Veri yükleme Kodları

6.3.3. Veri Standardizasyonu

Verilerin standartlaştırılması, TBA'nın doğru ve etkili bir şekilde uygulanabilmesi için temel bir gerekliliktir. Analiz öncesinde her bir değişkenin, ortalamadan çıkarılarak birim varyansa ölçeklendirilmesi gerekir. Bu işlem, farklı birimlerde ifade edilen değişkenler arasında karşılaştırma yapılabilirliği sağlamak için kritik öneme sahiptir.

TBA'nın temel varsayımlarından biri, değişkenlerin aynı ölçekte olmasıdır. Bu bağlamda, her bir zaman dilimine ait sütun değerleri, sklearn.preprocessing. StandardScaler sınıfı kullanılarak standardize edilmiştir. Bu yöntem, her sütunun ortalamasını sıfıra, standart sapmasını ise bire eşitleyerek tüm değişkenleri aynı ölçeğe taşır. Standardizasyon, değişkenler arasındaki ölçek farklılıklarını ortadan kaldırarak, analiz sonuçlarının daha tutarlı ve güvenilir olmasını sağlar.

```
from sklearn.preprocessing import StandardScaler

# Veriyi ölçeklendir
scaler = StandardScaler()
scaled_data = scaler.fit_transform(rtg_data)
```

Şekil 6-2:Veri Standartlaştırma Kodları

6.3.4. TBA Uygulama

TBA işlemi, sklearn.decomposition. TBA sınıfı kullanılarak gerçekleştirilmiştir. Veri, iki boyuta indirgenmiş ve her bir kanalın bu bileşenler üzerindeki koordinatları hesaplanmıştır. TBA işlemi sırasında, özdeğerler varyansı açıklayan oranlara dönüştürülmüş ve her bir bileşenin veri setindeki toplam varyansa katkısı belirlenmiştir. Elde edilen bu iki boyut, veri setinin boyutunu azaltırken en yüksek varyansı koruyan eksenler olarak tanımlanmıştır.

```
from sklearn.decomposition import PCA

# PCA uygulama (iki bileşene indirgeme)
pca = PCA(n_components=2)
transformed_data = pca.fit_transform(scaled_data)
```

Şekil 6-3:PCA Uygulama Kodları

6.3.5. Kovaryans Matrisi Hesaplanması

Standartlaştırılmış veri kullanılarak kovaryans matrisi oluşturulmuştur. Kovaryans matrisi, değişkenlerin birbiriyle olan ilişkisini ve değişimin boyutlarını anlamak için temel bir adımdır. TBA sınıfı, bu matrisi otomatik olarak hesaplar ve ardından özdeğer ayrışımı yapar. Kovaryans matrisi, veri setindeki her iki değişken arasındaki doğrusal ilişkinin gücünü gösterir.

```
import numpy as np

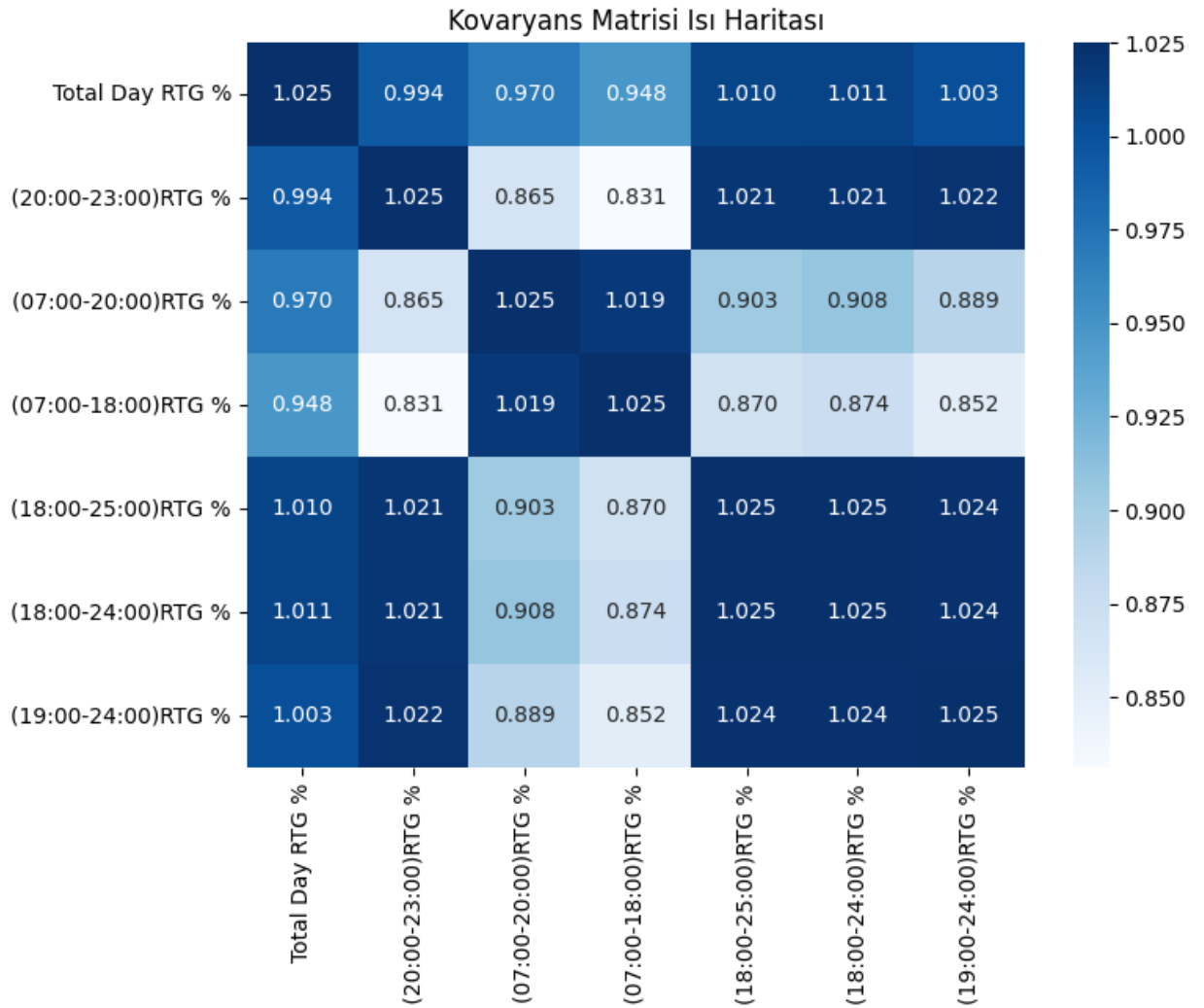
# Kovaryans matrisi
cov_matrix = np.cov(scaled_data.T)

print("Kovaryans Matrisi:")
print(cov_matrix)
```

Şekil 6-4:Kovaryans Matrisi Oluşturma Kodları

Tablo 2: Kovaryans Matrisi

1.025	0.085421	-0.01883	0.048681	0.213595	-0.01892	-0.1308
0.085421	1.025	0.131624	-0.0068	0.091375	-0.07631	-0.14488
-0.01883	0.131624	1.025	0.00995	0.121516	0.088774	-0.13899
0.048681	-0.0068	0.00995	1.025	0.028119	0.025901	-0.08173
0.213595	0.091375	0.121516	0.028119	1.025	0.144198	0.052166
-0.01892	-0.07631	0.088774	0.025901	0.144198	1.025	0.345594
-0.1308	-0.14488	-0.13899	-0.08173	0.052166	0.345594	1.025



Şekil 6-5: Python kodları ile Kovaryans Matrisi Isı Haritası

6.3.6. Özdeğerler ve Özvektörlerin Hesaplanması

TBA sınıfı, kovaryans matrisi üzerinden özdeğerler ve özvektörleri hesaplamıştır. Özdeğerler, her bir bileşenin veri setindeki varyansı açıklama gücünü ifade eder. Özvektörler ise bu bileşenlerin yönlerini temsil eder. İlk iki özvektör, en yüksek özdeğerlere karşılık geldiğinden, TBA bileşenleri olarak seçilmiş ve analiz bu bileşenler üzerinden yürütülmüştür.

```
# Özdeğerler ve özvektörler
eigenvalues, eigenvectors = np.linalg.eig(cov_matrix)

print("Özdeğerler:")
print(eigenvalues)

print("Özvektörler:")
print(eigenvectors)
```

Şekil 6-6:Özdeğerler ve Özvektörler Kodları

Özdeğerler genellikle verinin boyutlarını indirgeme aşamasında önemli rol oynar. Bu büyük olan özdeğer, verinin büyük kısmını açıklayan bileşeni temsil ederken, küçük olanlar düşük varyansa sahip bileşenlerdir. Bu tür bir durumda, büyük özdeğerleri seçip küçük olanları göz ardı etmek (boyut indirgeme) yaygın bir yaklaşımdır. Elde edilen özdeğerler ve özvektörler şu şekildeki gibi bir çıktı verebilir:

Tablo 3: Özdeğerler

	Özdeğerler
0	6.76775821e+00
1	3.98167016e-01
2	6.08991868e-03
3	2.28999928e-03
4	4.59636794e-04
5	4.26474530e-05
6	1.92569699e-04

Tablo 4: Özvektörler Matrisi

	Bileşen 1	Bileşen 2	Bileşen 3	Bileşen 4	Bileşen 5	Bileşen 6	Bileşen 7
Total Day RTG %	-0.38888	-0.02636	-0.12718	0.674837	-0.47249	0.205172	-0.3334
(20:00-23:00)RTG %	-0.37921	0.354589	-0.44461	-0.43402	-0.46537	-0.31351	0.171953
(07:00-20:00)RTG %	-0.36685	-0.52995	0.610514	-0.17504	-0.28574	-0.29781	0.104283
(07:00-18:00)RTG %	-0.35775	-0.62775	-0.54673	-0.13433	0.339415	0.207924	0.050513
(18:00-25:00)RTG %	-0.38452	0.244203	0.078349	0.445527	0.382392	-0.20802	0.631171
(18:00-24:00)RTG %	-0.38508	0.23079	0.104221	-0.09263	0.463554	-0.34638	-0.66644
(19:00-24:00)RTG %	-0.38245	0.29285	0.312257	-0.31696	0.047234	0.751295	0.057665

6.3.7. Özdeğerleri Sıralama

Hesaplanan özdeğerler, büyüklüklerine göre azalan sırada sıralanmış ve bu sıralama, veri setindeki en fazla varyansı açıklayan bileşenlerin belirlenmesini sağlamıştır. Özdeğerlerin büyüklüğü, her bir temel bileşenin veri setindeki toplam varyansa katkısını ifade eder. Bu doğrultuda, en büyük özdeğerlere sahip olan bileşenler, veri setindeki bilginin büyük bir kısmını temsil eder.

Analizde, ilk iki özdeğer, toplam varyansın önemli bir bölümünü açıklamaları nedeniyle tercih edilmiştir. Özdeğerlerin sıralanması, veri boyutunun küçültülmesi sürecinde en bilgilendirici bileşenlerin seçilmesine olanak tanır. Daha büyük özdeğerlere sahip bileşenlerin seçilmesi, veri boyutunu küçültürken bilgi kaybını en aza indirmeyi amaçlar. Bu yöntem, analizin hem etkinliğini hem de yorumlanabilirliğini artırır.

```
# Özdeğerleri sıralama
sorted_indices = np.argsort(eigenvalues)[::-1]
sorted_eigenvalues = eigenvalues[sorted_indices]
sorted_eigenvectors = eigenvectors[:, sorted_indices]

print("Sıralanmış Özdeğerler:")
print(sorted_eigenvalues)
print("Sıralanmış Özvektörler:")
print(sorted_eigenvectors)
```

Şekil 6-7:Özdeğerleri Sıralama Kodları



Şekil 6-8:Python Kodları ile Özdeğerlerin Sıralanmış Gösterimi

6.3.8. Açıklanan Varyans ve Kümülatif Varyans

TBA sonucunda elde edilen her bir bileşenin açıklanan varyansı, o bileşenin veri setindeki toplam varyansa katkısını yüzdelik oranla ifade eder. Bu oranlar genellikle azalan sırayla düzenlenir; çünkü ilk bileşenler, veri setindeki varyansın en büyük kısmını açıklayacak şekilde seçilir. Bu nedenle, açıklanan varyans, bileşenlerin veri setinin yapısal özelliklerini ne ölçüde yansıttığını değerlendirirken kritik bir ölçüttür.

Kümülatif varyans, bileşenlerin açıklanan varyans oranlarının toplamını ifade eder. Bu metrik, belirli bir bileşen sayısı ile toplam varyansın ne kadarının açıklandığını gösterir. Kümülatif

varyans genellikle grafiksel olarak sunulur ve grafik üzerindeki eğri, kaç bileşenin yeterli bilgi taşıdığına karar vermede yol gösterici olur.

Aşağıda, TBA analizi sırasında elde edilen açıklanan varyans ve kümülatif varyans sonuçları yer almaktadır:

Tablo 5:Açıklanan Varyanslar ve Kümülatif Varyanslar

Bileşen	Açıklanan Varyans	Kümülatif Varyans
Bileşen 1	0.943241563	0.943241563
Bileşen 2	0.055493661	0.998735223
Bileşen 3	0.000848769	0.999583993
Bileşen 4	0.000319164	0.999903156
Bileşen 5	6.40609E-05	0.999967217
Bileşen 6	2.6839E-05	0.999994056
Bileşen 7	5.9439E-06	1

Açıklanan Varyans: Her bir bileşenin, toplam varyansa yaptığı katkı oranını temsil eder. Örneğin, Bileşen 1 %94.32'lik bir varyans açıklarken, Bileşen 2 %5.55'lik bir katkı sağlamaktadır. Bu sonuç, ilk iki bileşenin toplamda %99.87'lik bir varyans açıklamakta olduğunu gösterir. Bu bileşenler, veri setinin büyük çoğunluğunu açıklamakta ve veri boyutunu azaltırken en fazla bilgi içeren bileşenlerdir.

Kümülatif Varyans: Kümülatif varyans, belirli bir sayıda bileşenin toplamda ne kadar varyansı açıkladığını gösterir. Bileşen 1 ve Bileşen 2, toplam varyansın %99.87'sini açıklarken, Bileşen 3'ten sonra açıklanan varyansın çok düşük olduğu görülmektedir. Bu durum, veri setinin büyük kısmının ilk iki bileşenle temsil edilebileceğini ve sonraki bileşenlerin çok fazla bilgi katkısı sağlamadığını gösterir.

Elde edilen sonuçlar, TBA'nın etkili bir şekilde çalıştığını ve verinin büyük çoğunluğunun yalnızca ilk iki bileşenle temsil edilebileceğini göstermektedir. İlk bileşen %94.32'lik bir varyansı açıklarken, ikinci bileşen %5.55'lik bir katkı sağlar. Bu iki bileşen, toplamda %99.87'lik bir varyans açıklayarak veri setinin temel yapısını temsil etmektedir. Bu da gösteriyor ki, veri seti sadece iki bileşenle yeterince açıklanabilir, bu nedenle diğer bileşenler analizde dikkate alınmamıştır. TBA, verinin boyutunu önemli ölçüde azaltırken, veri setinin özünü koruyan bir yöntem olarak etkili bir şekilde kullanılmıştır.

```

# Açıklanan varyans oranları ve kümülatif varyans
explained_variance_ratio = pca.explained_variance_ratio_
cumulative_variance_ratio = np.cumsum(explained_variance_ratio)

# DataFrame olarak düzenleme
variance_df = pd.DataFrame({
    "Bileşen": [f"Bileşen {i+1}" for i in range(len(explained_variance_ratio))],
    "Açıklanan Varyans": explained_variance_ratio,
    "Kümülatif Varyans": cumulative_variance_ratio
})

# Açıklanan varyans ve kümülatif varyans grafikleri
plt.figure(figsize=(14, 6))

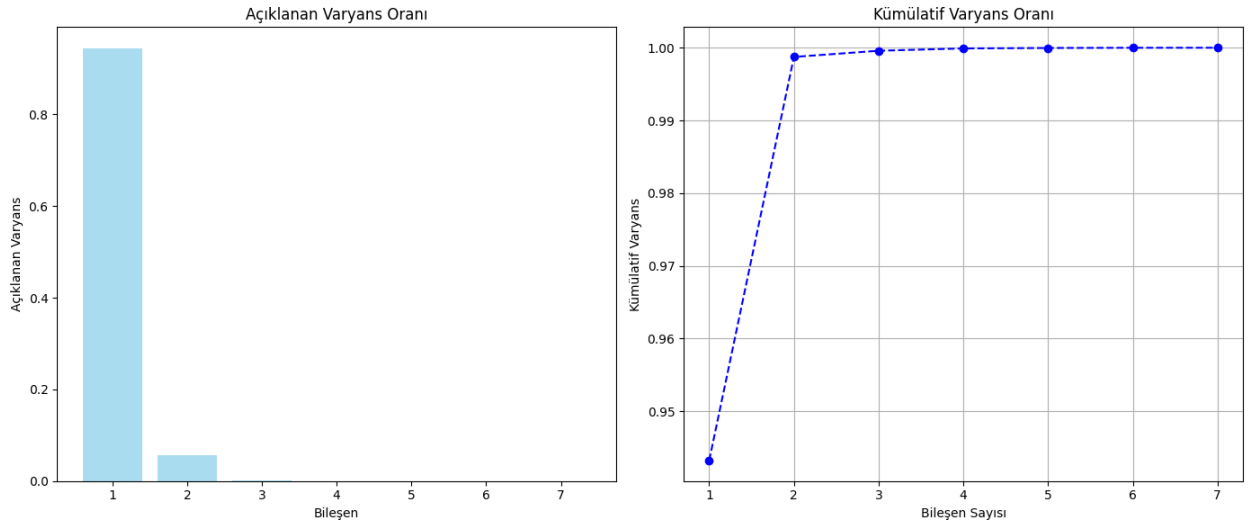
# Açıklanan varyans grafiği
plt.subplot(1, 2, 1)
plt.bar(range(1, len(explained_variance_ratio) + 1), explained_variance_ratio, color='skyblue', alpha=0.7)
plt.title('Açıklanan Varyans Oranı')
plt.xlabel('Bileşen')
plt.ylabel('Açıklanan Varyans')
plt.xticks(range(1, len(explained_variance_ratio) + 1))

# Kümülatif varyans grafiği
plt.subplot(1, 2, 2)
plt.plot(range(1, len(cumulative_variance_ratio) + 1), cumulative_variance_ratio, marker='o', linestyle='--', color='b')
plt.title('Kümülatif Varyans Oranı')
plt.xlabel('Bileşen Sayısı')
plt.ylabel('Kümülatif Varyans')
plt.xticks(range(1, len(cumulative_variance_ratio) + 1))
plt.grid()

plt.tight_layout()
plt.show()

```

Şekil 6 9:Açıklanan Varyans ve Kümülatif Varyans Kodları



Şekil 6-10:Python Kodları ile Açıklanan Varyans ve Kümülatif Varyans Grafiği

6.3.9.Yeni Veri Seti Oluşturma (Veriyi Dönüştürme) (Manuel)

Özvektörlerle veriyi dönüştürmek için, veriyi bu özvektörlerle çarpılır. Bu, veriyi daha düşük boyutlu bir uzaya projekte eder. TBA özvektör matrisini transpoze eder, böylece

(n_features, n_components) boyutunda bir matris elde edilir. Bu işlem, veriyle matris çarpımına uygun hale getirilmesi için yapılır.

[:, :2]: İlk 2 bileşenin özvektörlerini seçer. Bu, boyut indirgeme işleminde sadece en fazla varyansı açıklayan ilk iki bileşeni kullanılacağı anlamına gelir.

```
selected_eigenvectors = pca.components_.T[:, :2] # (7, 2) şeklinde
```

Seçilen Özvektörlerin boyutu: [7, 2]

```
transformed_data = np.dot(scaled_data, selected_eigenvectors)
```

StandardScaler ile ölçeklenmiş orijinal veri [41,7] Matris çarpımı yapar ve orijinal veriyi 2 bileşene indirir. Bu, orijinal 7 boyutlu veri kümesini 2 boyuta indirgeyerek her örneği 2D uzaya taşır.

```
print("Dönüştürülmüş Veri:")  
print(transformed_data)
```

Bu, her bir örneğin 2D uzayda temsilini gösterir. Matrisin boyutu [41, 2] olur. Dönüştürülmüş veri aşağıdaki gibi olur.

Dönüştürülmüş Veri:

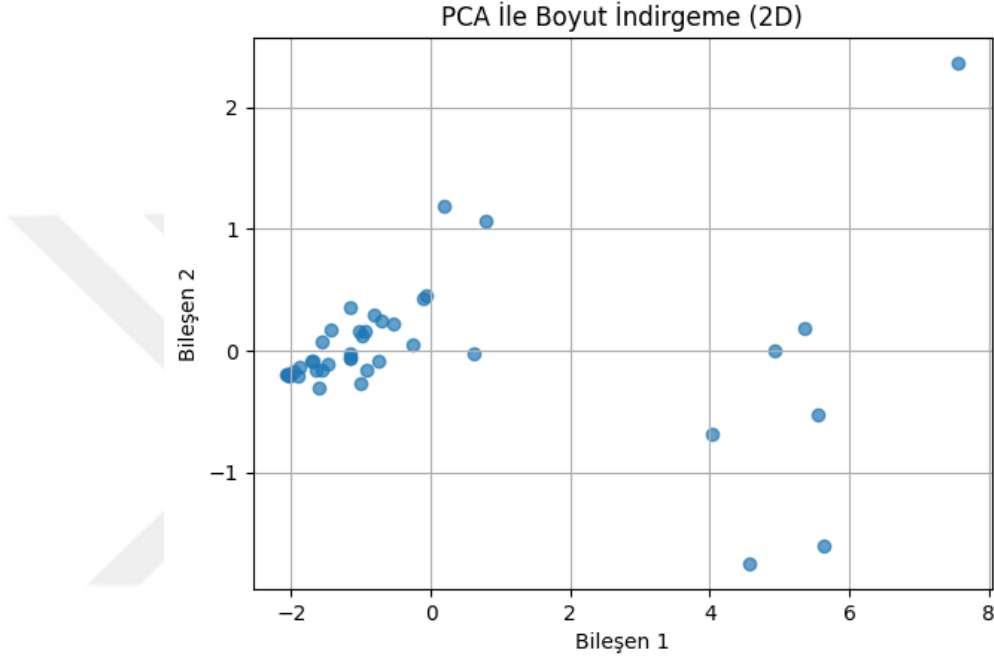
```
[[-1.14445037e+00 -6.48308065e-02]
 [-1.14445037e+00 -6.48308065e-02]
 [-9.73519169e-01  1.25601834e-01]
 [-1.00038987e+00 -2.75233922e-01]
 [ 7.55848899e+00  2.36313490e+00]
 [-2.06332757e+00 -1.92672573e-01]
 [-9.10764502e-01 -1.66112144e-01]
 [-1.15219231e+00  3.49970435e-01]
 [ 6.22921421e-01 -2.24805523e-02]
 [-2.04004667e+00 -1.99635365e-01]
 [-9.36405460e-01  1.55225772e-01]
 [-1.87000603e+00 -1.38466531e-01]
 [ 5.35255940e+00  1.86298788e-01]
 [-1.55148962e+00 -1.66189456e-01]
 [-2.49240334e-01  5.22692162e-02]
 [ 7.83992921e-01  1.06770864e+00]
 [-6.87064618e-02  4.51376653e-01]
 [ 4.92674161e+00 -4.86625298e-03]
 [-1.43543362e+00  1.65340823e-01]
 [-1.56359028e+00  7.48420901e-02]
 [-8.19149766e-01  2.88859932e-01]
 [ 5.54822175e+00 -5.30693411e-01]
 [ 4.03786139e+00 -6.92207805e-01]
 [-1.15147309e+00 -2.17623822e-02]
 [-1.02778215e+00  1.54466299e-01]
 [-1.47626868e+00 -1.07099221e-01]
 [-7.11315318e-01  2.42225107e-01]
 [ 4.56740459e+00 -1.75103370e+00]
 [-5.23949550e-01  2.19438719e-01]
 [ 1.92887798e-01  1.18390389e+00]
 [-1.04348450e-01  4.31458035e-01]
 [-2.02425904e+00 -2.10804686e-01]
 [-1.71143889e+00 -8.33142830e-02]
 [-7.57336832e-01 -8.16711758e-02]
 [-1.89912993e+00 -2.10242430e-01]
 [-1.95524128e+00 -1.69263715e-01]
 [ 5.64333308e+00 -1.61064079e+00]
 [-1.59831535e+00 -3.04717966e-01]
 [-1.63483581e+00 -1.62251470e-01]
 [-1.68616788e+00 -8.93720470e-02]
 [-2.04938830e+00 -1.91727643e-01]]
```

Şekil 6-11:Python Kodları ile Dönüştürülmüş Veri

Verimizi görselleştirdiğimiz zaman bileşenleri daha kolaylıkla analiz edebiliriz.

```
# Dönüştürülmüş veriyi görselleştirme (2D)
plt.scatter(transformed_data[:, 0], transformed_data[:, 1], alpha=0.7)
plt.title('PCA İle Boyut İndirgeme (2D)')
plt.xlabel('Bileşen 1')
plt.ylabel('Bileşen 2')
plt.grid(True)
plt.show()
```

Şekil 6-12: Dönüştürülmüş Veriyi 2D İle Görselleştirme Kodları



Şekil 6-13: Python Kodları ile PCA Bileşenlerini Görselleştirme

PCA (Principal Component Analysis), çok boyutlu verilerin daha düşük boyutlara indirgenmesi için yaygın olarak kullanılan bir yöntemdir. Bu işlemde, verinin en fazla varyansını açıklayan bileşenler seçilerek, verinin temel yapısı korunur. Yukarıdaki grafik, TBA yöntemiyle boyut indirgeme işlemi sonrasında elde edilen verilerin 2D düzlemde görselleştirilmesini sunmaktadır. Grafik, ilk iki ana bileşeni (principal components) kullanarak orijinal verinin 2D düzlemdeki temsilini göstermektedir. X eksenini, verinin en fazla varyansını açıklayan birinci bileşeni, Y eksenini ise ikinci bileşeni temsil etmektedir. Her bir nokta, orijinal veri setindeki bir örneği göstermekte olup, bu noktaların konumu, örneklerin bu iki bileşene göre nasıl sıralandığını belirtmektedir.

Grafikteki noktaların düzeni, verinin yapısına dair önemli bilgiler sunar. Özellikle, noktaların gruplaşma veya dağılımı biçimi, verinin içindeki benzerlikler veya farklılıklar hakkında bilgi verir. Örneğin, verinin belirli kümeler oluşturduğunu gösteren yoğun gruplar, veri setinde anlamlı sınıfların veya yapısal özelliklerin varlığını işaret edebilir. Ayrıca, veri noktalarının geniş

bir alana yayılması, verinin daha fazla varyansa sahip olduğunu ve bu varyansın birçok farklı bileşende dağılmış olduğunu gösterir. Sonuç olarak, TBA ile yapılan boyut indirgeme işlemi, veri setinin temel yapısını daha anlaşılır bir biçimde sunarak, veri analizi ve yorumlama süreçlerine önemli katkılar sağlar.

Verimizi daha iyi yorumlamak için kanal isimlerini ve her kanal için farklı renk ataması yapıyoruz. Özvektörleri de ekleyerek grafiğimizin anlaşılabilirliğini arttırabiliriz.

```
import matplotlib.pyplot as plt

# Renk paletini oluşturma
colors = plt.cm.get_cmap('tab20', len(channel_names))

# Grafik çizimi
plt.figure(figsize=(12, 8))

# Her kanalı farklı renklerle göstermek
for i, channel in enumerate(channel_names):
    plt.scatter(transformed_data[i, 0], transformed_data[i, 1], color=colors(i), label=channel)

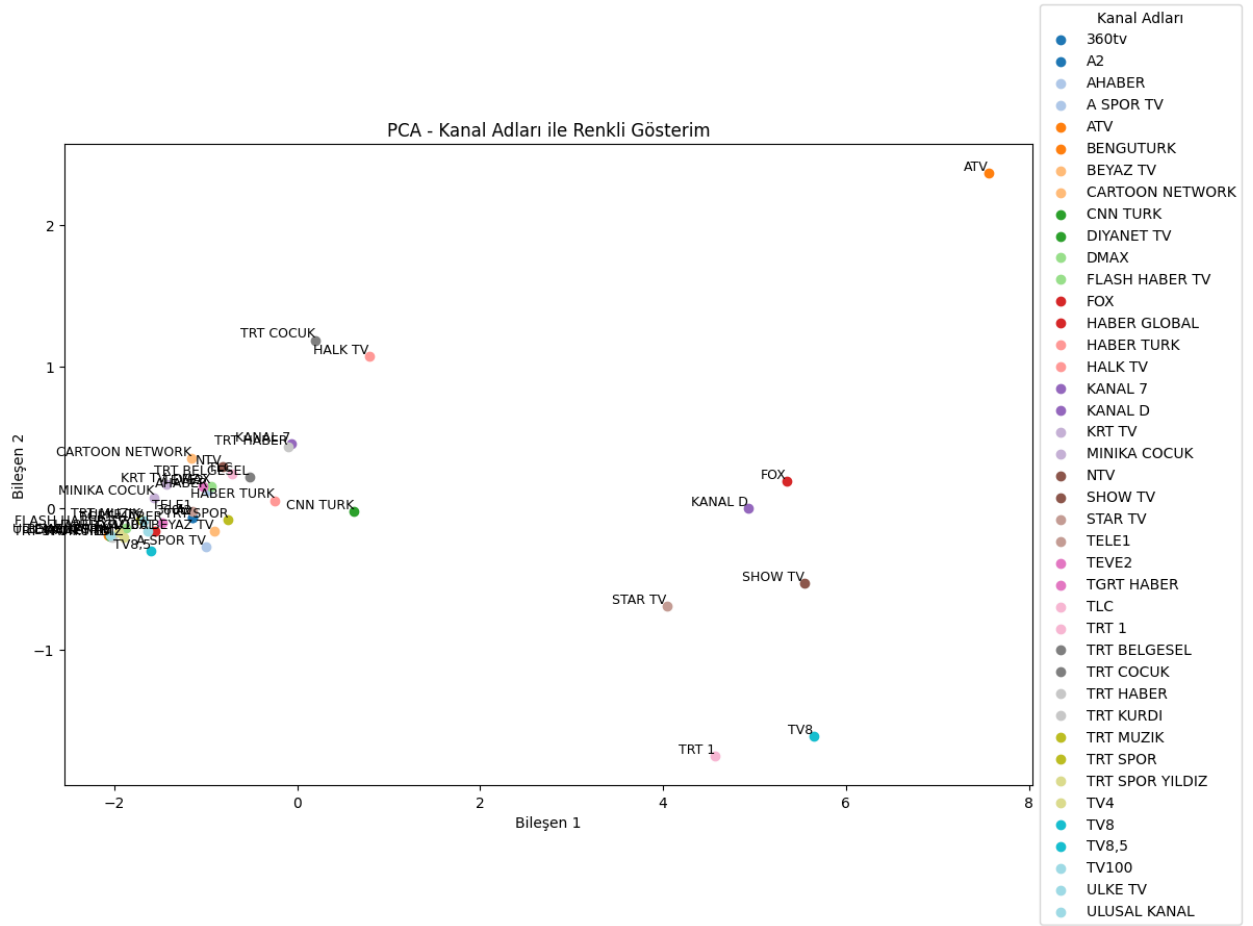
# Kanal adlarını her noktanın yanında yazma
for i, channel in enumerate(channel_names):
    plt.text(transformed_data[i, 0], transformed_data[i, 1], channel, fontsize=9, ha='right', va='bottom')

# Başlık ve etiketler
plt.title('PCA - Kanal Adları ile Renkli Gösterim')
plt.xlabel('Bileşen 1')
plt.ylabel('Bileşen 2')

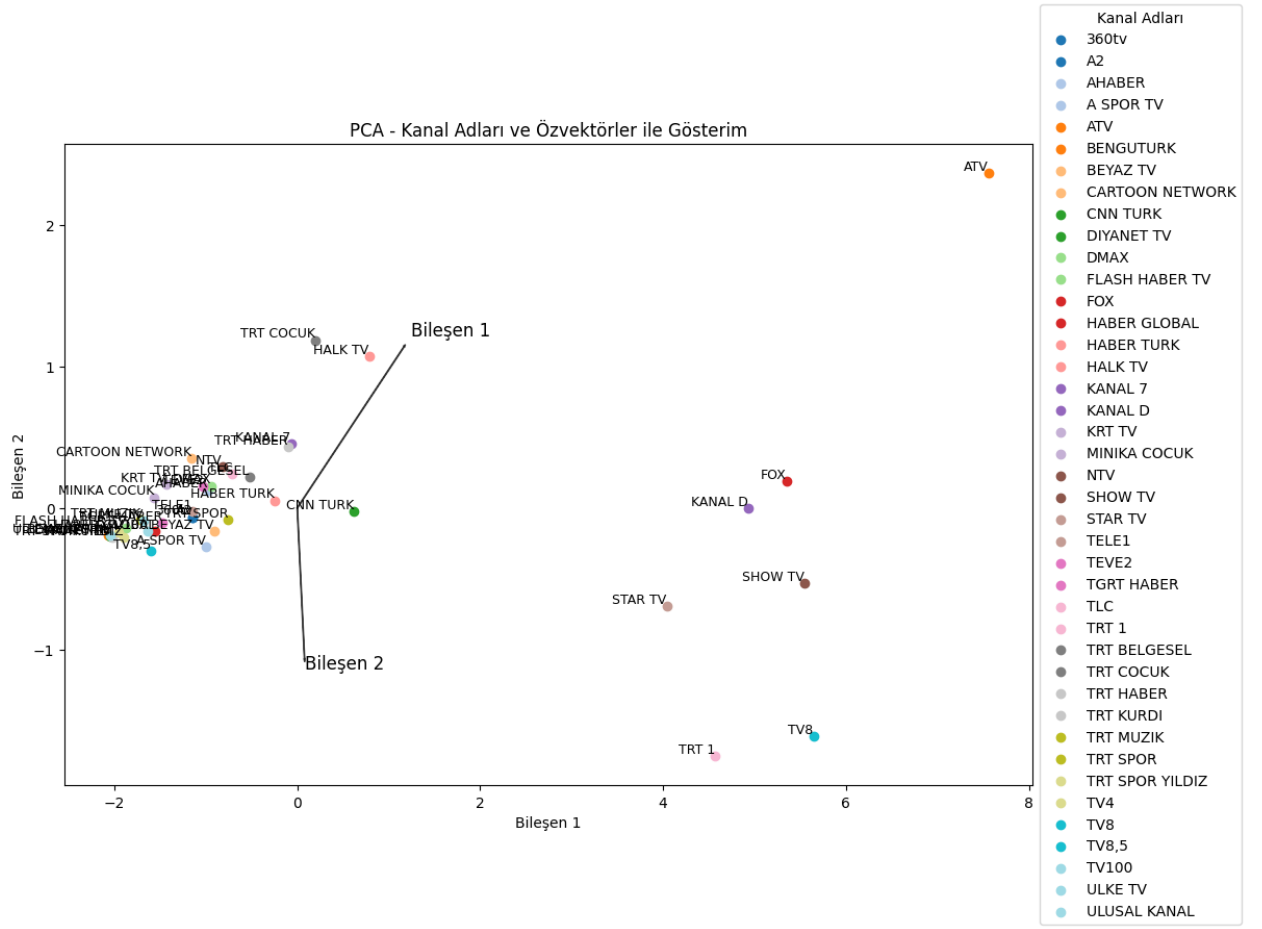
# Renkli gösterge (legend) ekleme
plt.legend(loc='center left', bbox_to_anchor=(1, 0.5), title='Kanal Adları')

# Grafik gösterimi
plt.show()
```

Şekil 6-14: PCA Bileşenlerinin 2D Renkli Gösterim Kodları



Şekil 6-15: PCA Bileşenlerinin 2D Renkli Gösterim Grafiği



Şekil 6-16: PCA Bileşenlerinin 2D Renkli Gösterim Grafiği (Özvektörler Dahil)

6.3.10.TBA Sonucunun Analizi ve Yorum

6.3.10.1. Bileşenlerin Yorumlanması

Bileşen 1 (X eksen), veri setindeki kanallar arasındaki genel izlenme trendlerini temsil etmektedir. Bu eksen, kanalların izlenme oranlarındaki çeşitliliği ve toplam varyans üzerindeki etkisini ölçmektedir. Şu bulgular dikkat çeker;

- ATV, X ekseninde en yüksek değeri alarak, genel izlenme oranları üzerinde baskın bir etkiye sahip olmuştur.
- FOX, KANAL D, SHOW TV, ve STAR TV, X ekseninde yüksek değerler almış olup, ulusal çapta önemli izleyici kitlesine sahip kanallar olarak dikkat çekmektedir.
- Sol tarafa yakın konumlanan ULUSAL KANAL, BEYAZ TV, ve benzeri kanallar, diğer ana akım kanallara kıyasla daha düşük izlenme oranlarına sahiptir.

Bileşen 2 (Y Eksen), izlenme oranlarındaki belirli bir zaman dilimi veya kategoriye özgü farklılıkları yansıtmaktadır. Bu eksen de şu bulgular dikkat çekmektedir:

- TRT ÇOCUK, Y ekseninde üst düzeyde bir konumda yer almakta olup, çocuk programlarıyla belirli zaman dilimlerinde yoğun bir izleyici kitlesine ulaştığını göstermektedir.
- HALK TV, yine Y ekseninde yüksek bir konumda bulunarak, haber ve tartışma programlarıyla belirli izleyici segmentlerinde önemli bir rol oynadığını işaret etmektedir.
- Daha merkezi bir konumda bulunan kanallar, izlenme oranlarının dış faktörlerden daha az etkilendiğini ve düşük varyansa sahip olduklarını göstermektedir.

6.3.10.2. Kanalların Konumlarına Göre Yorumlar

Yoğun Gruplar (Merkez): Merkezde yoğunlaşan kanallar arasında NTV, CARTOON NETWORK, MINIKA ÇOCUK, ve TRT BELGESEL yer almaktadır. Bu kanallar:

- Genellikle niş bir izleyici kitlesine hitap etmekte,
 - Daha dengeli izlenme oranları sergilemekte,
 - Ana akım kanallara kıyasla daha düşük varyans oluşturmaktadır.
- a) Sol Tarafta Yer Alan Kanallar: Sol tarafa yerleşen ULUSAL KANAL, BEYAZ TV, TV4, ve benzeri kanallar, izlenme oranlarının genelde daha düşük olduğunu ve izleyici kitlesinin sınırlı bir çeşitlilik gösterdiğini ifade etmektedir.
- b) Sağ Tarafta Ayırışan Kanallar: X ekseninde önemli şekilde ayırışan kanallar, izlenme oranlarındaki genel farklılıkları temsil etmektedir. Şu kanallar dikkat çekicidir:
- ATV, X ekseninde en yüksek konuma sahiptir ve bu durum, kanalın izlenme oranlarının diğer kanallara göre çok daha belirgin olduğunu göstermektedir.
 - FOX, KANAL D, ve SHOW TV, ana akım izleyici kitlesine hitap ederek hem Bileşen 1 hem de Bileşen 2'de çok boyutlu bir etkisini ortaya koymaktadır.
- c) Yukarda Konumlanan Kanallar: TRT ÇOCUK ve HALK TV, Y ekseninde yukarıda yer almakta olup, özel izleyici segmentlerinde etkili olduklarını göstermektedir. Bu konumlar, bu kanalların çocuk programları veya haber odaklı yayınlarıyla özel bir niş pazarı hedeflediğini işaret etmektedir.
- d) Alt Tarafta Yer Alan Kanallar: Alt tarafta yer alan kanallar arasında belirgin bir yoğunluk olmamakla birlikte, bu kanalların zaman dilimlerine göre izlenme oranlarındaki farklılıkların düzeyinin düşük olduğu ifade edilebilir.

6.3.10.3. Stratejik Çıkarımlar

Bu analiz, televizyon kanallarının izlenme oranlarına ilişkin önemli stratejik bilgiler sunmaktadır. Şu noktalar öne çıkmaktadır:

Ana Akım Kanallar: Özellikle ATV, FOX, ve KANAL D gibi kanalların, genel izleyici kitlesine hitap eden stratejiler geliştirerek prime-time reklam gelirlerini artırması olasıdır.

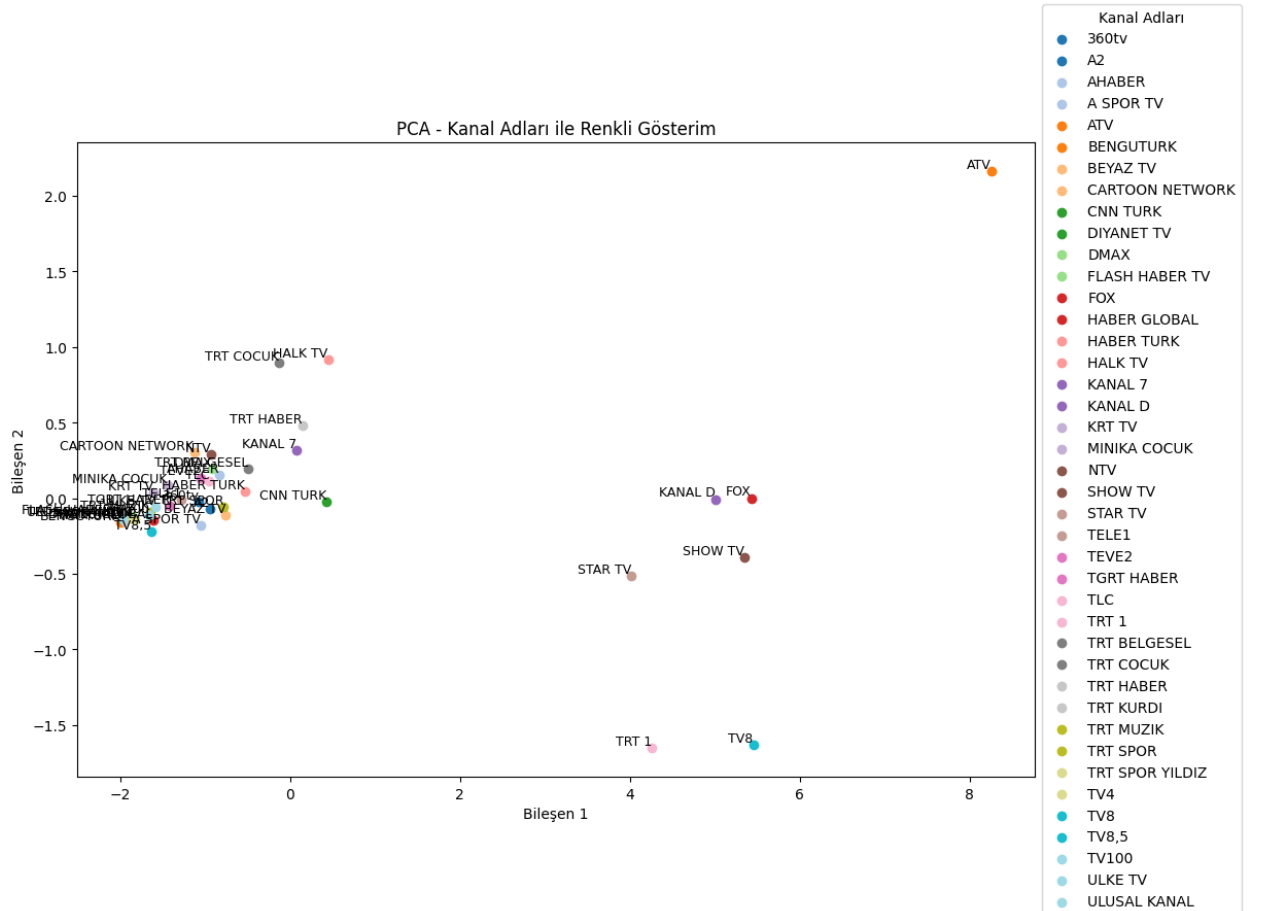
Niş Kanallar: TRT ÇOCUK ve MINIKA ÇOCUK gibi niş kanallar, özellikle sabah ve gündüz zaman dilimlerinde yoğunlaşan izleyici kitlesine uygun çocuk programlarıyla pazar payını genişletebilir.

Benzer Kanallarla İş Birlikleri: Merkezde yer alan ve benzer izleyici segmentlerine hitap eden kanallar için ortak içerik stratejileri geliştirilebilir.

Sonuç olarak, PCA ile elde edilen bu analiz, televizyon kanallarının izlenme oranlarına dair çok boyutlu bir bakış sağlamakta ve stratejik karar alma süreçlerinde önemli bir yol gösterici işlev görmektedir.

6.3.10.4. AB ve 20+ABC1 Grubunun Karşılaştırılması

AB grubuna uyguladığımız temel bileşenler analizi yöntemini 20+ABC1 grubuna da uygularsak aşağıdaki grafiği elde ederiz.



Şekil 6 17: PCA Bileşenlerinin 2D Renkli Gösterim Grafiği

AB Grubu, daha eğitimli ve yüksek gelir düzeyine sahip bireylerin tercihlerini yansıtmaktadır. Bu grup, genellikle haber kanalları ve kültürel içerik üreten kanalların daha merkezi ya da daha yüksek varyans açıklayan bileşenlerde yoğunlaştığı bir grafikte temsil edilir. Diğer yandan, ABC1 Grubu, daha geniş bir demografik kitlenin tercihlerini içerir ve orta ile üst sosyoekonomik grupları kapsar. Bu grafikte ise eğlence kanalları, örneğin ATV, Kanal D ve FOX gibi geniş izleyici kitlesine hitap eden kanallar, daha belirgin bir şekilde ayrılarak kendilerini gösterir.

ATV ve FOX gibi kanalların ayrışması; ABC1 grafiğinde ATV gibi popüler eğlence kanallarının birinci bileşen (Bileşen 1) boyunca daha uç noktalarda yer aldığı görülüyor. Bu, eğlence içeriklerinin 20+ ABC1 grubunda önemli bir faktör olduğunu gösteriyor. Kanal D ve FOX, hem eğlence hem de dizi içeriklerinde güçlü bir konuma sahip olduğundan benzer şekilde ayrışıyor.

Çocuk ve haber kanallarının konumları; TRT Çocuk veya Cartoon Network, her iki grafikte de genellikle üst bileşenlerde (Bileşen 2) yer alır, ancak ABC1 grafiğinde diğer kanallardan daha belirgin şekilde izole olmuş. Haber kanalları (örneğin, TRT Haber, CNN Türk) her iki grupta da merkezi pozisyonlarda bulunur; ancak ABC1'de nispeten daha düşük varyans açıklıyor.

Varyans ve izleme davranışlarındaki farklar; AB grubu grafiği daha dar bir demografik yapıyı temsil ettiğinden kanalların konumları arasında daha belirgin gruplamalar görülmektedir. ABC1 grafiği ise daha geniş bir kitleyi temsil ettiğinden kanallar arasındaki konum farkları daha homojen veya daha az belirgin olması beklenmektedir fakat AB grubunda olduğu gibi belirgin gruplaşmalar fark edilmiştir.

Stratejik çıkarımlar; ABC1 grubu, reklamcılık ve içerik stratejisi açısından daha geniş bir kitleyi temsil ettiğinden, bu gruba hitap eden kanalların ayrışma derecesi, kanal stratejilerinin etki alanını gösterebilir. AB grubu sonuçları ise daha niş ve odaklı içerik stratejileri geliştirmek için kullanılabilir.

7. SONUÇ VE ÖNERİLER

Bu çalışmada, çok değişkenli veri analizinin önemli bir aracı olan Temel Bileşenler Analizi (TBA), televizyon reyting verileri üzerinde uygulanmış ve yöntem, teorik ve uygulamalı yönleriyle detaylı bir şekilde ele alınmıştır. Araştırma kapsamında, TİAK tarafından sağlanan AB ve ABC1 izleyici gruplarına ait reyting verileri, TBA'nın veri boyutlarını indirgeme ve değişkenler arasındaki ilişkileri görselleştirme kabiliyeti çerçevesinde değerlendirilmiştir. Elde edilen bulgular, TBA'nın özellikle karmaşık ve çok boyutlu veri setlerinde etkili bir analiz aracı olduğunu ve veri içerisindeki temel desenleri ortaya koymada büyük bir avantaj sağladığını göstermektedir.

Araştırmanın bulguları, televizyon kanallarının izleyici kitlesinin günün farklı zaman dilimlerine göre değişiklik gösterdiğini, bazı kanalların özellikle "Prime Time" saat dilimlerinde yüksek izlenme oranlarına sahip olduğunu, diğerlerinin ise daha çok gündüz saatlerinde tercih edildiğini göstermektedir. Özellikle, ATV gibi kanalların hemen her zaman diliminde yüksek reyting ve pay oranlarına sahip olduğu gözlemlenirken, TRT 1 ve Kanal 7 gibi kanallar, izleyicilerini sabah ve öğle saatlerinde daha yoğun bir şekilde çekmektedir.

TBA analizi, bu kanalların reytinglerinin zamansal dağılımını anlamada önemli bir araç olmuş, izlenme oranlarındaki değişkenliklerin temel sebepleri ortaya konmuştur. Elde edilen ana bileşenler, televizyon pazarındaki ana eğilimleri ve kanal stratejilerini yansıtmaktadır. Ayrıca, araştırma, medya planlamacıları ve reklam verenler için değerli bilgiler sunarak, televizyon kanallarının daha hedeflenmiş içerik stratejileri geliştirmelerine ve reklam bütçelerini etkin bir şekilde dağıtmalarına yardımcı olacaktır.

TBA, reyting verileri gibi çok sayıda değişkenden oluşan karmaşık veri setlerinde boyut indirgeme açısından başarılı sonuçlar sunmuştur. Özellikle AB ve ABC1 izleyici gruplarına ait değişkenlerin, ilk birkaç temel bileşen aracılığıyla anlamlı bir şekilde özetlenebildiği gözlemlenmiştir. İlk iki bileşenin toplam varyansın büyük bir kısmını açıklaması, bu yöntemin veri boyutunu azaltırken bilgi kaybını minimumda tuttuğunu ortaya koymaktadır. TBA ile elde edilen bileşenlerin incelenmesi, AB ve ABC1 gruplarının televizyon izleme alışkanlıklarında belirgin farklılıklar olduğunu göstermiştir. Bu farklılıklar, özellikle prime-time kuşağı ve belirli kanalların izlenme oranlarında kendini göstermektedir. Çalışmada, her iki grup için de yüksek varyansa sahip kanallar ve zaman dilimlerinin belirlenmesi, medya stratejileri açısından değerli bilgiler sunmuştur. Reyting verileri üzerinde gerçekleştirilen TBA uygulaması, medya kuruluşlarının içerik stratejilerini planlarken izleyici gruplarının tercihlerini daha iyi anlamalarına olanak sağlamaktadır. Elde edilen bileşenlerin analiz edilmesi, kanalların rekabet stratejilerini şekillendirebilecek önemli çıkarımlar sunmuştur. TBA, yüksek boyutlu verilerin düşük boyutlu bir

uzayda görselleştirilmesine imkan tanımış ve televizyon kanallarının veya zaman dilimlerinin birbirine yakınlıkları ile farklılıklarını kolayca gözlemleme olanağı sağlamıştır. Bu durum, özellikle stratejik planlama ve karar destek sistemlerinde kullanılabilecek bir araç olarak TBA'nın önemini artırmaktadır.

Bu araştırma ile televizyon kanallarının zaman dilimlerine göre izlenme oranlarını incelemiş olsa da, gelecekteki çalışmalar bu verileri daha derinlemesine analiz edebilir. Örneğin, içerik türlerinin ve kanal programlarının izlenme oranları üzerindeki etkisini araştıran çalışmalar, kanal stratejileri hakkında daha fazla bilgi sağlayabilir. Ayrıca, dijital platformların televizyon izlenme davranışları üzerindeki etkisi de gelecekteki araştırmalar için önemli bir konu olabilir. Yapılacak daha kapsamlı analizler, özellikle sosyal medya ve dijital medya platformlarıyla televizyon izlenme alışkanlıklarının karşılaştırılması üzerine odaklanabilir.

TBA'nın daha kapsamlı bir analiz sağlamak amacıyla, sadece AB ve ABC1 gruplarını değil, farklı demografik grupları da kapsayan daha geniş veri setleri üzerinde uygulanması önerilmektedir. İzleyici profillerine dair daha detaylı ve genel geçer çıkarımlar yapılabilir.

Bu çalışmada kullanılan reyting verileri belirli bir dönemi kapsamakla birlikte, daha dinamik analizler için uzun dönemli veri setlerinin incelenmesi önerilmektedir. Zaman serisi analizleriyle entegre edilen TBA uygulamaları, izleyici davranışlarındaki değişimleri daha iyi anlamaya olanak sağlayabilir. TBA sonuçlarının daha geniş bir bağlama oturtulabilmesi için, faktör analizi, kümeleme analizi ve diskriminant analizi gibi diğer çok değişkenli istatistiksel yöntemlerle karşılaştırmalı analizlerin yapılması önerilmektedir. Bu tür karşılaştırmalar, TBA'nın avantajlarını ve sınırlamalarını daha net bir şekilde ortaya koyabilir. Bu tezde kullanılan TBA yaklaşımı, yalnızca televizyon reytingleri ile sınırlı kalmayıp, diğer sektörlerdeki çok boyutlu veri analizlerinde de uygulanabilir. Örneğin, müşteri segmentasyonu, ürün performansı analizi veya finansal risk değerlendirmesi gibi farklı alanlarda TBA'nın potansiyeli araştırılabilir. Araştırmacılar ve sektör profesyonelleri için TBA'nın daha kolay yorumlanabilir hale getirilmesi adına, görselleştirme tekniklerinin ve açıklayıcı araçların geliştirilmesi önerilmektedir. TBA sonuçlarının daha geniş bir kitle tarafından anlaşılmasını sağlayabilir. TBA, veri ön işleme veya boyut indirgeme adımları için makine öğrenmesi modelleriyle birlikte kullanılabilir. Özellikle sınıflandırma ve tahminleme modellerinde, TBA ile elde edilen bileşenlerin girdi olarak kullanılması, analiz gücünü artırabilir.

Bu çalışmanın sonuçları, TBA'nın çok değişkenli veri analizi için etkili bir araç olduğunu ve televizyon reytingleri gibi karmaşık veri setleri üzerinde indirgeme yaparak önemli katkılar sunduğunu göstermektedir. Bununla birlikte, önerilen yeni uygulama alanları ve metodolojik yaklaşımlar, gelecekteki çalışmalar için geniş bir araştırma alanı sunmaktadır.

Ayrıca, TBA'nın televizyon reyting verileri üzerindeki uygulamalarının, sadece içerik stratejileri ve reklam planlamasıyla sınırlı kalmayıp, televizyon yayın politikalarının şekillendirilmesine de katkı sağlayabileceği düşünülmektedir. Özellikle izleyici tercihleri doğrultusunda program içeriklerinin yeniden yapılandırılması, yayın akışlarının optimize edilmesi ve belirli saat dilimlerinde hedef kitleye yönelik içeriklerin planlanması açısından TBA'nın sunduğu özet bilgilerin karar alma süreçlerinde önemli bir rol oynayabileceği görülmüştür. Bu bağlamda, medya yöneticileri ve yayıncılar için daha fazla içgörü sunabilmek adına, TBA sonuçlarının zaman içinde nasıl değiştiğini inceleyen uzun vadeli analizlerin gerçekleştirilmesi önerilmektedir. Böylece, izleyici kitlesinin zaman içerisindeki eğilimleri ve yeni medya platformlarının televizyon izleme alışkanlıkları üzerindeki etkileri daha net bir şekilde ortaya konabilir.

Öte yandan, televizyon reyting verilerinin daha detaylı analiz edilmesi için farklı istatistiksel ve yapay zeka tabanlı yöntemlerle birlikte değerlendirilmesi de önemli bir araştırma alanı olarak öne çıkmaktadır. Örneğin, makine öğrenmesi algoritmaları ile TBA'nın entegre edilmesi, izleyici davranışlarını tahmin etmek ve gelecekteki reyting eğilimlerini öngörmek açısından güçlü bir yaklaşım sunabilir. Derin öğrenme modelleri ve zaman serisi analiziyle birleştirilen TBA, medya sektöründeki karar destek sistemlerini daha etkin hale getirebilir. Böylece, televizyon kanalları yalnızca geçmiş verileri analiz etmekle kalmayıp, aynı zamanda izleyici alışkanlıklarının gelecekte nasıl değişebileceğine dair proaktif stratejiler geliştirebilir. Bu tür çalışmalar, medya sektörünün dijitalleşme sürecinde rekabet avantajı elde etmesine ve izleyici kitlesine daha iyi hizmet sunmasına olanak sağlayacaktır.

8. KAYNAKLAR

- [1] Aaker, D. A. (1996). Building strong brands. *Managing Brand Equity*, 196.
- [2] Akbulut, Ö., & Çapık, C. (2022). Çok değişkenli istatistiksel analizler için örneklem büyüklüğü. *J Nursology*, 25(2), 111–116. <https://doi.org/10.5152/JANHS.2022.970637>
- [3] Aktürk, H. (2012). Temel Bileşenler Analizi ve Uygulama Alanları. *Veri Bilimi Araştırmaları*, 2(1), 2–3.
- [4] Albayrak, A. S. (2005). Türkiye’de illerin sosyoekonomik gelişmişlik düzeylerinin çok değişkenli istatistik yöntemlerle incelenmesi. *ZKÜ Sosyal Bilimler Dergisi*, 1(1), 159–177.
- [5] Albayrak, A. (1980). Veri Kümesi ve Faktör Analizi Üzerine Çalışmalar.
- [6] Alkan, Ö. (2008). Temel bileşenler analizi ve bir uygulama örneği (Yüksek lisans tezi). Atatürk Üniversitesi, Sosyal Bilimler Enstitüsü, Erzurum.
- [7] Anderson, T. W. (1963). *An introduction to multivariate statistical analysis*.
- [8] Aydın, G. (2018). *Son dönem değişen reyting ölçüm sisteminin medya yayın program içeriğine etkileri (Acun Ilıcalı örneği üzerinden)* (Yüksek lisans tezi). Avrasya Üniversitesi, Sosyal Bilimler Enstitüsü, Halkla İlişkiler Bilim Dalı, Trabzon.
- [9] Bayar, H. (2013). Asenkron jeneratör birimkoruma sisteminin gerçekleştirilmesi (Doktora tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, Elektrik Eğitimi Anabilim Dalı, Elektrik Eğitimi Programı). İstanbul.
- [10] Baydemir, M. (2021). Bilimsel araştırmalarda istatistiğin ve doğru yöntem seçimlerinin önemi. *Bitlis Eren Üniversitesi Sosyal Bilimler Dergisi*, 10(1), 29–34.
- [11] Bayram, N. 2004. Sosyal Bilimlerde SPSS İle Veri Analizi. Bursa: Ezgi Kitabevi.
- [12] Billor, N., & Kırıl, G. (2005). Yüksek boyutlu veri kümeleri için robust bacon temel bileşenler analizi. 7. *Ulusal Ekonometri ve İstatistik Sempozyumu*, 26–27 Mayıs 2005.
- [13] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- [14] Bulut, H. (2023). *R uygulamaları ile çok değişkenli istatistiksel yöntemler* (2. baskı). Nobel Akademik Yayıncılık.
- [15] Cattell, R. B. (1949). The description of personality: Basic traits resolved into the five source traits. *The Journal of Abnormal and Social Psychology*, 44(2), 166–178.
- [16] Cooley, W.W., & Lohnes, P.R. (1971). *Multivariate Data Analysis*. John Wiley & Sons, Inc., Canada, 364s.
- [17] Cömert Z. (2015). Temel bileşenler analizine genel bir bakış. Tarihinde erişildi. www.zafercomert.com [21.05.15].

- [18] Demir, M., Esenbuğa, H., & Bilgin, M. (2016). İleri düzey veri analizi yöntemleri ve uygulamalar. İstanbul: Bilimsel Yayınlar.
- [19] Emanet, H. (2017). *Sosyal medya kullanımının yaygınlaşmasının TV reytingine etkisi* (Yüksek lisans tezi). Bahçeşehir Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.
- [20] Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 3, 272–299.
- [21] Funder, D. C. (2016). *The Personality Puzzle* (7th ed.). W.W. Norton & Company.
- [22] Gezmez, E. (2007). Clustering MRI images with principal component analysis methods (Yüksek lisans tezi). Çukurova Üniversitesi.
- [23] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- [24] Günel, H. Alptekin. (1977). Faktör Analizi, İ.Ü. Orman Fakültesi Dergisi, Seri B Cilt: 27, Sayı: 1, Matbaa Teknisyenleri Basımevi, İstanbul.
- [25] Hacıhasanoğlu, B. (1980). İller İçin Bir Gelişmişlik Göstergesi ve Sıralama. DPT Yayın No:1695-KÖYD:19, Ankara: Devlet Planlama Teşkilatı.
- [26] Harris, R.J. (2001). *A Primer of Multivariate Statistics* (3rd ed.). Psychology Press. <https://doi.org/10.4324/9781410600455>
- [27] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441.
- [28] Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.
- [29] Jolliffe, I. T. (2004). *Principal component analysis* (2nd ed.). Springer Science+Business Media.
- [30] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- [31] Johnson, R., & Wichern, D. (2019). Multivariate statistical analysis in various fields. *Journal of Data Science*, 12(4), 123–145.
- [32] Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson.
- [33] Kaya, M. F. (2013). Sürdürülebilir kalkınmaya yönelik tutum ölçeği geliştirme çalışması. *Armara Coğrafya Dergisi*, 28, 175–193. <http://www.marmaracografya.com>
- [34] Koçak, İ. (1998). Temel Bileşenler Analizi ve Uygulaması. Yüksek Lisans Tezi, İnönü Üniversitesi Sosyal Bilimler Enstitüsü. Malatya.
- [35] Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson.

- [36] Kuyucu, M. (2020). Televizyon mecrasında reyting sorunsalı: Türkiye’de yayın yapan TV kanallarının izlenme oranlarının ölçümlenmesinde iletişim eğitimi alan gençlerin gözünde yaşanan aksaklıklar ve bu aksaklıklara yönelik çözüm önerileri. *USBAD Uluslararası Sosyal Bilimler Akademi Dergisi*, 2(4), 754-780. <https://doi.org/10.47994/usbad.788702>
- [37] Manly, B. F. J. (2018). *Multivariate statistical methods: A primer* (3rd ed.). CRC Press.
- [38] McCauley, L. G. (1995). The Five Factor Model: A psychological perspective. In *Personality and Individual Differences*, 18(1), 1–12.
- [39] McDonald, M., & Dunder, I. (1998). *Market Segmentation: How to Do It and How to Profit from It*. Macmillan Business, 376.
- [40] Mertler, C. A., Vannatta, R. A., & LaVenja, K. N. (2021). *Advanced and multivariate statistical methods: Practical application and interpretation* (7th ed.). Routledge. <https://doi.org/10.4324/9781003047223>
- [41] Miran, B. (2021). *Temel istatistik*. İzmir: İzmir Yayınevi.
- [42] Naik, P. A., & Tsai, C. L. (2004). Constrained inverse regression for incorporating prior information. *Journal of the American Statistical Association*, 99(465), 319–327.
- [43] Özbek B., Ayav T., Yatır M. N., Kirişken B., Televizyon İzleme Ölçüm Sistemi Tasarımı, *EMO Bilimsel Dergi*, 1(2) 89-94
- [44] Özdamar, K. (1999). *Paket Programlar ile İstatistiksel Veri Analizi Çok Değişkenli Analizler* (2nd ed.). Kaan Kitabevi, 137–496.
- [45] Özdamar, K. (2002). *Modern İstatistiksel Analiz Teknikleri*. Eskişehir: Anadolu Üniversitesi Yayınları.
- [46] Özdamar, K. (2004). *Paket Programlar ile İstatistiksel Veri Analizi-2 (Çok Değişkenli Analizler)*, 5. Baskı. Eskişehir: Kaan Kitabevi.
- [47] Rein, M., & Eysenck, H. J. (2013). *Dimensions of Personality*. New York: Psychology Press. <https://doi.org/10.4324/9780203793268>
- [48] Sharma, S. (1996). *Applied Multivariate Techniques*. John Wiley.
- [49] Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15(2), 201–293.
- [50] Spearman, C. (1910). The correlation between mental and physical tests. *British Journal of Psychology*, 3(1), 55–83.
- [51] Sütçü Asar, S. (2014). Veri Zarflama Analizinde Temel Bileşenler Analizinin Kullanımı (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

- [52] Tabachnick, B. G., & Fidell, L. S. (2020). *Çok Değişkenli İstatistiklerin Kullanımı* (6. baskı). Nobel Akademik Yayıncılık.
- [53] Tatlıdil, H. (1996). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Cem Web Ofset Ltd. Şti., Ankara.
- [54] Tezbaşaran, A. (2016). *Faktör Analizi ve Temel Bileşenler Yöntemi*. İstanbul: Akademik Yayınlar.
- [55] TİAK: <https://tiak.com.tr/> (Erişim Tarihi: 10.12.2024)
- [56] Ünal, F. (2020). Faktör analizinde belirli koşullar altında faktör sayısı belirleme yöntemlerinin karşılaştırılması (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü.
- [57] Yıldız, K., Çamurcu, Y., & Doğan, B. (2010). Veri madenciliğinde temel bileşenler analizi kullanımı: Bir vaka çalışması. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 24(1), 115–126.
- [58] Ziegler, D., & Matzke, S. (2021). A statistical method for principal component analysis. *Data Science Review*, 2(4), 142–153.
- [59] Zhang, S., & Zhang, Y. (2019). Applications of PCA in high-dimensional data. *Applied Computational Statistics*, 34(2), 93–104.
- [60] Zhao, Y., & Chen, S. (2017). A new approach to principal component analysis in bioinformatics. *Journal of Computational Biology*, 11(3), 258–275.
- [61] Zhao, Z., & Lu, J. (2018). Optimal principal component analysis for large datasets. *International Journal of Data Analysis*, 10(5), 255–268
- [62] Zhou, X., & Li, Z. (2020). Advanced principal component analysis for multivariate data. *Statistical Methods in Medical Research*, 29(6), 1567–1580.
- [63] Zhu, T., & Liu, J. (2016). Application of principal component analysis in environmental data analysis. *Environmental Research Letters*, 11(2), 245–260.

9. EKLER

EK1: TİAK AB Grubu RTG Değerleri

KANAL	Total Day RTG %	(20:00- 23:00) RTG %	(07:00- 20:00) RTG %	(07:00- 18:00) RTG %	(18:00- 25:00) RTG %	(18:00- 24:00) RTG %	(19:00- 24:00) RTG %
360 tv	0.11	0.29	0.10	0.09	0.21	0.23	0.24
A2	0.11	0.29	0.10	0.09	0.21	0.23	0.24
AHABER	0.14	0.25	0.14	0.12	0.24	0.24	0.25
A SPOR TV	0.13	0.38	0.09	0.07	0.29	0.30	0.33
ATV	1.07	2.39	1.13	0.99	2.00	2.22	2.27
BENGUTURK	0.00	0.01	0.01	0.01	0.01	0.01	0.01
BEYAZ TV	0.14	0.37	0.11	0.09	0.30	0.31	0.33
CARTOON NETWORK	0.11	0.15	0.15	0.14	0.15	0.16	0.15
CNN TURK	0.34	0.77	0.25	0.21	0.70	0.68	0.73
DIYANET TV	0.01	0.02	0.01	0.01	0.01	0.01	0.02
DMAX	0.15	0.26	0.14	0.13	0.24	0.24	0.25
FLASH HABER TV	0.03	0.05	0.03	0.03	0.05	0.05	0.05
FOX	0.80	2.15	0.75	0.52	1.81	1.99	2.25
HABER GLOBAL	0.07	0.16	0.05	0.05	0.13	0.14	0.15
HABER TURK	0.23	0.50	0.19	0.16	0.44	0.44	0.47
HALK TV	0.34	0.52	0.41	0.35	0.55	0.58	0.53
KANAL 7	0.24	0.47	0.26	0.22	0.41	0.44	0.42
KANAL D	0.77	2.19	0.63	0.51	1.73	1.88	2.07
KRT TV	0.08	0.11	0.10	0.10	0.11	0.11	0.11
MINIKA COCUK	0.06	0.10	0.08	0.08	0.09	0.10	0.09
NTV	0.16	0.27	0.16	0.16	0.25	0.25	0.25
SHOW TV	0.83	2.60	0.62	0.49	1.96	2.13	2.35
STAR TV	0.68	2.21	0.45	0.38	1.60	1.72	1.90
TELE1	0.11	0.25	0.11	0.09	0.21	0.23	0.23
TEVE2	0.13	0.23	0.14	0.12	0.23	0.22	0.22
TGRT HABER	0.07	0.18	0.07	0.06	0.14	0.15	0.16
TLC	0.17	0.32	0.18	0.15	0.28	0.29	0.29
TRT 1	0.71	2.71	0.37	0.29	1.87	2.06	2.34
TRT BELGESEL	0.19	0.39	0.18	0.17	0.33	0.34	0.36
TRT COCUK	0.26	0.37	0.36	0.34	0.34	0.38	0.36
TRT HABER	0.23	0.46	0.25	0.22	0.40	0.43	0.44
TRT KURDI	0.01	0.03	0.01	0.01	0.02	0.02	0.02

TRT MUZİK	0.04	0.10	0.05	0.05	0.08	0.08	0.09
TRT SPOR	0.17	0.38	0.13	0.11	0.33	0.33	0.35
TRT SPOR YILDIZ	0.02	0.07	0.02	0.02	0.05	0.06	0.06
TV4	0.02	0.04	0.02	0.02	0.03	0.03	0.03
TV8	0.85	2.86	0.48	0.37	2.22	2.36	2.59
TV8,5	0.06	0.18	0.03	0.03	0.15	0.15	0.17
TV100	0.06	0.13	0.05	0.04	0.11	0.12	0.12
ULKE TV	0.05	0.10	0.05	0.05	0.09	0.09	0.09
ULUSAL KANAL	0.01	0.01	0.01	0.01	0.01	0.01	0.01

EK2: TİAK AB Grubu RTG Değerleri

KANAL	Total Day RTG %	(20:00- 23:00) RTG %	(07:00- 20:00) RTG %	(07:00- 18:00) RTG %	(18:00- 25:00) RTG %	(18:00- 24:00) RTG %	(19:00- 24:00) RTG %
360tv	0,13	0,35	0,13	0,11	0,25	0,27	0,29
A2	0,15	0,42	0,13	0,12	0,3	0,32	0,35
AHABER	0,18	0,34	0,18	0,16	0,31	0,32	0,32
A SPOR TV	0,14	0,39	0,1	0,09	0,3	0,31	0,34
ATV	1,43	3,14	1,52	1,31	2,67	2,95	2,97
BENGUTURK	0	0,01	0,01	0	0,01	0,01	0,01
BEYAZ TV	0,19	0,47	0,15	0,12	0,38	0,39	0,42
CARTOON NETWORK	0,13	0,18	0,17	0,16	0,18	0,19	0,19
CNN TURK	0,38	0,84	0,29	0,24	0,76	0,75	0,8
DIYANET TV	0,01	0,02	0,01	0,01	0,01	0,02	0,02
DMAX	0,18	0,29	0,17	0,16	0,27	0,27	0,28
FLASH HABER TV	0,02	0,04	0,03	0,02	0,04	0,05	0,05
FOX	1	2,67	0,93	0,66	2,23	2,45	2,76
HABER GLOBAL	0,06	0,15	0,05	0,04	0,12	0,13	0,14
HABER TURK	0,23	0,48	0,19	0,17	0,43	0,43	0,45
HALK TV	0,37	0,53	0,45	0,39	0,57	0,6	0,54
KANAL 7	0,3	0,65	0,32	0,26	0,55	0,59	0,58
KANAL D	0,97	2,65	0,81	0,67	2,09	2,27	2,5
KRT TV	0,06	0,1	0,08	0,07	0,09	0,09	0,09
MINIKA COCUK	0,08	0,14	0,11	0,09	0,13	0,14	0,13
NTV	0,17	0,27	0,18	0,18	0,25	0,25	0,25
SHOW TV	1,01	2,92	0,8	0,63	2,28	2,48	2,7
STAR TV	0,84	2,57	0,59	0,52	1,88	2,01	2,21
TELE1	0,11	0,23	0,11	0,09	0,2	0,21	0,21
TEVE2	0,15	0,24	0,16	0,13	0,26	0,25	0,25
TGRT HABER	0,09	0,2	0,09	0,07	0,16	0,17	0,18

TLC	0,17	0,31	0,16	0,14	0,28	0,28	0,29
TRT 1	0,83	3,08	0,45	0,36	2,14	2,35	2,66
TRT BELGESEL	0,23	0,47	0,22	0,2	0,4	0,41	0,43
TRT COCUK	0,27	0,38	0,37	0,35	0,36	0,39	0,37
TRT HABER	0,32	0,6	0,34	0,3	0,54	0,57	0,58
TRT KURDI	0,01	0,03	0,01	0,01	0,02	0,02	0,02
TRT MUZIK	0,05	0,13	0,05	0,05	0,1	0,1	0,11
TRT SPOR	0,19	0,43	0,15	0,13	0,37	0,37	0,39
TRT SPOR YILDIZ	0,03	0,07	0,03	0,02	0,05	0,06	0,06
TV4	0,03	0,05	0,02	0,02	0,05	0,05	0,05
TV8	1,01	3,39	0,59	0,46	2,61	2,77	3,04
TV8,5	0,06	0,16	0,03	0,03	0,14	0,14	0,15
TV100	0,06	0,12	0,05	0,04	0,1	0,11	0,11
ULKE TV	0,07	0,13	0,06	0,06	0,12	0,12	0,12
ULUSAL KANAL	0,01	0,01	0,01	0,01	0,01	0,01	0,01