

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Selim GÜNDÜZ

YÜKSEK LİSANS TEZİ

**UZAKLIK FONKSİYONLARININ ÇOK BOYUTLU ÖLÇEKLEME
ALGORİTMALARINDAKİ ETKİNLİĞİNİN İNCELENMESİ VE
UYGULAMALAR**

İSTATİSTİK ANABİLİM DALI

ADANA, 2011

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

Selim GÜNDÜZ

**UZAKLIK FONKSİYONLARININ ÇOK BOYUTLU ÖLÇEKLEME
ALGORİTMALARINDAKİ ETKİNLİĞİNİN İNCELENMESİ VE
UYGULAMALAR**

İSTATİSTİK ANABİLİM DALI

ADANA, 2011

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**UZAKLIK FONKSİYONLARININ ÇOK BOYUTLU ÖLÇEKLEME
ALGORİTMALARINDAKİ ETKİNLİĞİNİN İNCELENMESİ VE
UYGULAMALAR**

Selim GÜNDÜZ

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANABİLİM DALI

Bu Tez 08./07./2011 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından
Oybirliği/Oyçokluğu ile Kabul Edilmiştir.

.....
Prof.Dr.Sadullah SAKALLIOĞLU
DANIŞMAN

.....
Prof.Dr.Selahattin KAÇIRANLAR
ÜYE

.....
Doç.Dr.K.Can KILIÇ
ÜYE

Bu Tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.
Kod No:

Prof. Dr. İlhami YEĞİNGİL
Enstitü Müdürü

Bu Çalışma Ç. Ü. Araştırma Projeleri Birimi Tarafından Desteklenmiştir.
Proje No: FEF2010YL29

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZ

YÜKSEK LİSANS TEZİ

**UZAKLIK FONKSİYONLARININ ÇOK BOYUTLU ÖLÇEKLEME
ALGORİTMALARINDAKİ ETKİNLİĞİNİN İNCELENMESİ VE
UYGULAMALAR**

Selim GÜNDÜZ

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI**

Danışman : Prof. Dr. Sadullah SAKALLIOĞLU
Yıl: 2011, Sayfa: 111
Jüri : Prof. Dr. Sadullah SAKALLIOĞLU
: Prof. Dr. Selahattin KAÇIRANLAR
: Doç. Dr. Kemal Can KILIÇ

Çok boyutlu ölçekleme (ÇBÖ) analizi, birimler arasındaki ilişkileri göstermek için veriyi daha düşük boyutta konfigürasyon ile görselleştiren istatistikî bir araçtır. Bu tez çalışmasında çok boyutlu ölçekleme analizinde, farklı uzaklık fonksiyonlarının stress değeri üzerindeki etkisi incelenmiştir. Çalışmada hem gerçek hem de simülasyon verileri kullanılmıştır. Öklid uzaklığı incelenen tüm çok boyutlu ölçekleme algoritmalarında stress değerini en iyileyen uzaklık ölçüsü olmuştur. Çalışmada ayrıca çok boyutlu ölçekleme algoritmalarının pratikte uygulamasını göstermek için iki uygulamaya yer verilmiştir. Birinci uygulamada Türkiye'deki birinci düzey istatistikî bölgeler karşılaştırılmıştır. İkinci uygulamada ise Türkiye ve 36 Avrupa ülkesi, seçilen 12 sosyo-ekonomik göstergeye göre karşılaştırılmıştır.

Anahtar Kelimeler: Klasik Çok Boyutlu Ölçekleme, SMACOF, Kruskal-Shepard Algoritması, Stress Değeri, Uzaklık

ABSTRACT

MSc THESIS

INVESTIGATION OF EFFECTS ON MULTIDIMENSIONAL SCALING ALGORITHMS OF DISTANCE FUNCTIONS AND APPLICATIONS

Selim GÜNDÜZ

**ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS**

Supervisor : Prof. Dr. Sadullah SAKALLIOĞLU
Year: 2011, Page: 111

Jury : Prof. Dr. Sadullah SAKALLIOĞLU
: Prof. Dr. Selahattin KAÇIRANLAR
: Assoc.Prof. Dr. Kemal Can KILIÇ

Multidimensional scaling (MDS) is a statistical tool for constructing a low-dimensional configuration to represent the relationships between the objects. In this study, the effects of distance functions on the stress will be examined in metric multidimensional scaling algorithms. To study the effects on stress value of distance measures used in both simulation and real data sets. The distance measurement having the best performance is Euclidean distance according to stress value. Also in this study, two applications are made to demonstrate the practical applicability of metric MDS algorithms. In first application, the first-level statistical region units will be compared in terms of socio-economic indicators by metric multidimensional scaling algorithms. In second application, it is aimed to examine the position of European states and Turkey and display possible similarities or dissimilarities between them, from the point of their socio-economic indicators.

Key Words: Classic Multidimensional Scaling, SMACOF, Kruskal-Shepard Algorithm, Stress Value, Distance

TEŐEKKÜR

Her zaman öđrencisi olmaktan büyük gurur duyacađım sayın Prof. Dr. Sadullah SAKALLIOđLU'na yoğun programına rađmen tez alıřmama verdiđi katkılardan dolayı sonsuz teőekkür ederim. Tez alıřmam sırasında karřılařtıđım tüm problem ve sıkıntılarında yardımlarını esirgemeyen bölüm başkanımız sayın Prof. Dr. Selahattin KAIRANLAR'a ve ukurova Üniversitesi İstatistik Bölümü'nün diđer öğretim elemanlarına teőekkür ederim.

Tanıdıđım günden bugüne yaşam sevincim olan, yanımda olmadıđında bile manevi desteđi ile bana güç veren sevgili eřim F. Feyza GÜNDÜZ' e ve hayatım boyunca moral ve motivasyon aısından hep yanımda olduklarını hissettiren aileme teőekkür ederim.

İÇİNDEKİLER

SAYFA

ÖZ	I
ABSTRACT	II
TEŞEKKÜR	III
İÇİNDEKİLER.....	IV
TABLO LİSTESİ	VII
ŞEKİLLER DİZİNİ	IX
1.GİRİŞ	1
2.ÖNCEKİ ÇALIŞMALAR.....	7
3. ÇOK BOYUTLU VERİ VE UZAKLIK FONKSİYONLARI.....	21
3.1. Çok Değişkenli Verinin Matris Gösterimi.....	21
3.2. Ölçek Türleri.....	22
3.2.1. Nominal (İsimsel) Ölçek	23
3.2.2. Ordinal (Sıralı) Ölçek.....	23
3.2.3. Aralık Ölçeği	24
3.2.4. Oran Ölçeği	24
3.3. Uzaklık Fonksiyonları	25
3.3.1. Nicel Değişkenler İçin Uzaklık Ölçüleri.....	26
3.3.1.1. City-Block (Manhattan) Uzaklığı.....	26
3.3.1.2. Chebyshev Uzaklığı.....	28
3.3.1.3. Öklid (Euclidean) Uzaklığı.....	30
3.3.1.4. Standartlaştırılmış Öklid Uzaklığı.....	31
3.3.1.5. Minkowski Uzaklığı	32
3.3.1.6. Mahalanobis Uzaklığı.....	33
3.3.2. Nitel Değişkenler İçin Benzerlik Ölçüleri.....	35
3.3.2.1 İki Sonuçlu (Binary) Değişkenler İçin Kullanılan Başlıca Benzerlik Ölçüleri	36
3.3.2.2. İki Sonuçtan Fazla Değer Alan Nitel Değişkenler İçin Benzerlik Ölçüleri	38

3.3.3. Nitel ve Nicel Değişkenler İçeren Gözlem Çiftleri İçin Benzerlik Ölçüleri.....	38
4. ÇOK BOYUTLU ÖLÇEKLEME (ÇBÖ) ANALİZİ	43
4.1. Çok Boyutlu Ölçeklemede Varsayımlar.....	46
4.2. Çok Boyutlu Ölçekleme Modelleri	47
4.3. Çok Boyutlu Ölçekleme Teknikleri	48
4.3.1. Metrik Çok Boyutlu Ölçekleme Tekniği.....	50
4.3.1.1. Klasik Metrik Çok Boyutlu Ölçekleme	52
4.3.1.2 Karmaşık Fonksiyonların Optimizasyonu İle Ölçekleme.....	53
4.3.2 Metrik Olmayan (Non-Metrik) Çok Boyutlu Ölçekleme Tekniği.....	55
4.3.2.1. Kruskal-Shepard Yaklaşımı	56
4.4. Stress Fonksiyonları	58
4.5. Boyut Sayısının Belirlenmesi	60
4.6. Boyutların Yorumlanması	61
4.7. Çok Boyutlu Ölçekleme Algoritmaları İçin Örnek Uygulamalar	63
4.7.1. Birinci Düzey İstatistikî Bölgelerin Gelişmişlik Karşılaştırılması	67
4.7.1.1. Klasik Çok Boyutlu Ölçekleme Analizi İle Birinci Düzey İstatistikî Bölgelerin Karşılaştırılması	67
4.7.1.2. SMACOF Algoritması ile Birinci Düzey İstatistikî Bölgelerin Karşılaştırılması.....	70
4.7.2. Avrupa Ülkelerinin Sosyo-Ekonomik Göstergelere Göre Gelişmişlik Düzeylerinin Karşılaştırılması	72
4.7.2.1. Klasik Çok Boyutlu Ölçekleme Analizi İle Avrupa Ülkelerinin Karşılaştırılması.....	73
4.7.2.2. SMACOF Algoritması ile Avrupa Ülkelerinin Karşılaştırılması.75	
5. UZAKLIK FONKSİYONLARININ STRESS ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ	79
5.1. Gerçek Veri Setleri İle Uzaklık Fonksiyonlarının Stress Üzerindeki Etkisinin İncelenmesi.....	79
5.2. Simülasyon Çalışması ile Uzaklık Fonksiyonlarının Stress Üzerindeki Etkilerinin İncelenmesi.....	86

5.2.1. Çok Değişkenli Normal Dağılımdan Rassal Örneklem Üretme.....	86
5.2.2.Simülasyon	88
6. SONUÇLAR VE ÖNERİLER	99
KAYNAKLAR.....	103
ÖZGEÇMİŞ	111

TABLULAR DİZİNİ**SAYFA**

Tablo 3.1. İki sonuçlu p sayıda değişken içeren bir gözlem çiftinin çapraz tablosu..	36
Tablo 3.2. İki sonuçlu p değişken içeren verinin gözlem çiftleri arasındaki yakınlık için kullanılan bazı eşleştirme tipli benzerlik ölçüleri	37
Tablo 3.3. Bahçe çiçeği verisi (Kaufman ve Rousseeuw 1990).....	40
Tablo 3.4. Bahçe çiçeği verisinin Gower (1971) tarafından önerilen benzerlik ölçüsüne göre ilk 10 bitki türü için elde edilen benzerlik matrisi.	41
Tablo 4.1. Gözlemler Arası Uzaklık Değerleri	49
Tablo 4.2. Birinci Düzey İstatistikî Bölgelerin Karşılaştırılmasında Kullanılan Sosyo-Ekonomik Değişkenler.....	63
Tablo 4.3. İstatistikî Bölge Birimlerinin Düzey 1 İtibariyle Sosyo-Ekonomik göstergeleri.....	64
Tablo 4.4. Türkiye ve 36 Avrupa ülkesinin karşılaştırılmasında kullanılan 12 sosyo-ekonomik değişken	65
Tablo 4.5. Türkiye ve 36 Avrupa ülkesine ait seçilmiş 12 sosyo-ekonomik gösterge	66
Tablo 4.6. Birinci Düzey İstatistikî Bölgelerin Karşılaştırılmasında Klasik ÇBÖ Algoritması ve Aralıklı ÇBÖ Modelinde Parametre Tahminleri ve Anlamlılık Testleri.....	67
Tablo 4.7. İstanbul Bölgesinin Diğer İstatistikî Bölgelere Olan Orijinal ve Konfigürasyon Uzaklıkları.....	70
Tablo 4.8. Birinci Düzey İstatistikî Bölgelerin Karşılaştırılmasında Rassal Başlangıçlı SMACOF Algoritması ve Aralıklı ÇBÖ Modelinde Parametre Tahminleri ve Anlamlılık Testleri.....	70
Tablo 4.9. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde Klasik ÇBÖ Algoritması ve Aralıklı ÇBÖ Modelinde Parametre Tahminleri ve Anlamlılık Testleri	73
Tablo 4.10. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde SMACOF Algoritması ve Aralıklı ÇBÖ odelinde Parametre Tahminleri ve Anlamlılık Testleri	76

Tablo 5.1. Çalışmada kullanılacak veri setlerinin örneklem hacimleri ve değişken sayıları.....	80
Tablo 5.2. Klasik metrik ÇBÖ algoritmasında farklı uzaklık fonksiyonlarına göre elde edilen stres değerleri.....	80
Tablo 5.3. Karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında farklı uzaklık fonksiyonlarına göre 10 tekrar sonucunda elde edilen ortalama stres değerleri(Başlangıç Koordinatları Rassal).....	82
Tablo 5.4. Karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında farklı uzaklık fonksiyonlarına göre elde edilen stres değerleri (Başlangıç Koordinatları Temel Bileşen Skorları).....	84
Tablo 5.5. Simülasyon çalışmasında çok değişkenli normal dağılımdan rassal örneklemeler üretilirken göz önünde bulundurulacak ortalama vektörleri ve kovaryans matrisleri.....	89
Tablo 5.6. Değişkenler arasındaki ilişkilerin ve terim büyüklüklerinin aynı olması durumunda farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri	91
Tablo 5.7. Değişkenler arasındaki ilişkilerin aynı ve terim büyüklüklerinin farklı olması durumunda farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri	92
Tablo 5.8. Değişkenler arasındaki ilişkilerin farklı ve terim büyüklüklerinin aynı olması durumunda farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri	93
Tablo 5.9. Değişkenler arasındaki ilişkilerin ve terim büyüklüklerinin farklı olması durumunda farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri	95

ŞEKİLLER DİZİNİ

SAYFA

Şekil 1.1. Yıllara göre içerisinde “çok boyutlu ölçekleme” kavramı geçen çalışma sayıları.....	4
Şekil 3.1. Çok değişkenli veride gözlem ve değişken vektörleri	21
Şekil 3.2. City-Block uzaklığının iki boyutlu düzlemde hesaplanması.....	27
Şekil 3.3. City-Block uzaklık fonksiyonunun üç boyutlu uzayda grafiği.....	28
Şekil 3.4. Chebsyhev uzaklığının iki boyutlu düzlemde hesaplanması.....	29
Şekil 3.5. Chebsyhev uzaklık fonksiyonunun üç boyutlu uzayda grafiği.....	29
Şekil 3.6. Öklid uzaklığının iki boyutlu düzlemde hesaplanması	30
Şekil 3.7. Öklid uzaklık fonksiyonunun üç boyutlu uzayda grafiği.....	31
Şekil 3.8. İki boyutlu düzlemde orijinden bir birim uzaklıktaki noktalardan oluşan yüzeyde L_1, L_2 ve L_∞ metrikleri.....	33
Şekil 3.9. Farklı I değerleri Minkowski uzaklık fonksiyonunun üç boyutlu uzayda grafiği	33
Şekil 4.1. ÇBÖ Yöntemine Göre İngiltere’deki Şehirlerin Haritalandırılması (Cox and Cox, 2001).....	44
Şekil 4.2. ÇBÖ Yöntemi ile Türkiye Haritası	45
Şekil 4.3. Gözlemler Arası İlişkinin Şekilsel İfadesi (Tatlıdil, 2002)	50
Şekil 4.4. Orijinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafiği	68
Şekil 4.5. İstatistiki Bölge Birimlerinin Düzey 1 İtibariyle ÇBÖ Analiziyle İki Boyutlu Gösterimi	69
Şekil 4.6. Orijinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafiği	71
Şekil 4.7. İstatistiki Bölge Birimlerinin Düzey 1 İtibariyle SMACOF Algoritması ile Elde Edilen İki Boyutlu Gösterimi	72
Şekil 4.8. Orijinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafiği	74
Şekil 4.9. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde Klasik ÇBÖ Algoritması ile Elde Edilen İki Boyutlu Gösterimi	75
Şekil 4.10. Orijinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafiği	76

Şekil 4.11. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde Rassal Başlangıçlı SMACOF Algoritması ile Elde Edilen İki Boyutlu Gösterimi.....	77
Şekil 5.1. Farklı uzaklık fonksiyonlarına göre metrik ÇBÖ algoritmalarının Karşılaştırılması.....	81
Şekil 5.2. Farklı veri setleri için rassal başlangıçlı karmaşık fonksiyonların optimizasyonu algoritmasında farklı uzaklık fonksiyonlarının kullanılması ile elde edilen stress değerleri	83
Şekil 5.3. Farklı veri setleri için temel bileşen başlangıçlı karmaşık fonksiyonların optimizasyonu algoritmasında farklı uzaklık fonksiyonlarının kullanılması ile elde edilen stress değerleri	85
Şekil 5.4. Farklı uzaklık fonksiyonlarına göre metrik ÇBÖ algoritmalarının Karşılaştırılması.....	85
Şekil 5.5. Durum I’de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimi.....	91
Şekil 5.6. Durum II’de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimi.....	93
Şekil 5.7. Durum III’de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimi.....	94
Şekil 5.8. Durum IV’de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimi.....	96

1.GİRİŞ

İnsanlar yaşamları boyunca birçok problemle karşılaşmakta ve bu problemleri çözmeleri için çok yönlü düşünüp, problemleri çok yönlü olarak değerlendirmeleri gerekmektedir. Çünkü bir olayı etkileyen pek çok faktör olabilir. Bundan dolayı faktörlerin tamamını göz önünde bulundurduktan sonra çözüm önerilerini getirmek gerekir.

Çok değişkenli istatistiksel analiz, istatistik bilimi başta olmak üzere birçok bilim dalında büyük bir öneme sahiptir. Tek değişkenli istatistiksel analizlerin problemlere cevap veremediği durumlarda çok değişkenli analizlere başvurulur. Çok değişkenli analizlerin yapılmalarının temel amaçları nesnelere sınıflandırmak, boyutlarını indirgemek ve problemleri daha basit hale getirmektir. Kısacası çok değişkenli analiz, incelenen olayın daha az sayıda parametre ile ifade edilebilmesidir. Diğer istatistik tekniklerinde olduğu gibi, çok değişkenli analiz tekniklerinin hedefleri de bilimsel çalışmaların ve araştırmaların sayı ile ifade edilebilen sonuçlarının özetlenmesi, yorumlanması ve karar verilirken bu sonuçların kullanılmasının sağlanmasıdır.

Çok değişkenli analiz yöntemlerinden biri de çok boyutlu ölçekleme (ÇBÖ) analizidir. ÇBÖ analizi, yüksek boyutlu veri kümesinde boyut indirgemeyi amaçlayan ve grafiksel olarak daha basit şekilde yorumlanmasına imkân sağlayan bir yöntemdir. Çok boyutlu ölçekleme analizi, yüksek boyutlu veri setlerinde birimlerin (veya değişkenlerin) birbirlerine göre konumlarını daha az boyutlu düzlemde grafiksel olarak göstermeyi amaçlar (Borg ve Groenen, 2005). ÇBÖ analizinde, $n \times n$ boyutlu uzaklık matrisindeki δ_{ij} değerlerini temsil edecek d_{ij} konfigürasyon uzaklıklarının elde edileceği indirgenmiş yeni koordinat düzlemi elde edilmektedir (Cox ve Cox, 2001). ÇBÖ analizinde kullanılan çok sayıda algoritma mevcuttur. Bu algoritmalar genel olarak metrik ve metrik olmayan algoritmalar olmak üzere iki ayrı sınıfta değerlendirilmektedir. İster metrik ister metrik olmayan algoritma olsun çok boyutlu ölçekleme algoritmalarındaki temel amaç orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uyumsuzluğun bir göstergesi olan stress değerini en küçükmeye çalışmaktır.

Çok boyutlu ölçekleme ile ilgili ilk çalışmalar Young ve Household (1938)'un birimler arasındaki Öklid uzaklık matrisindeki uzaklıkları koruyarak nasıl iki boyutlu koordinat sisteminde gösterileceğine yönelik çalışması ile Richerdson (1938)'in boyut indirgemek için önerdiği modelin uygulanabilirliğini gösterdiği çalışmadır.

ÇBÖ'nin gelişmesinde iki önemli aşama gerçekleşmiştir. İlk olarak bahsedilecek aşama, metrik yaklaşımlardır. 'Princeton' veya 'Torgerson' yaklaşımı olarakta bilinen bu yaklaşımla, Klasik ÇBÖ'nin temelleri Princeton Üniversitesinde Messick ve Albelson (1956) ve Torgerson'nun (1952) içinde bulunduğu Psychometrik grubu tarafından atılmıştır. Bunun yanında Torgerson (1952) ilk defa ÇBÖ analizinin uygulanabilirliğini göstermiştir (Davison, 1983). Bahsedilecek olan ikinci aşama ise, on yıl aradan sonra Beil Telephone Laboratuvarında Shepard (1962a, 1962b) tarafından metrik olmayan ÇBÖ analizi yaklaşımının bulunmasıdır. Bu yaklaşım 'Kruskal-Shepard' yaklaşımı olarak da bilinmektedir. Kruskal (1964), Shepard'ın (1962) bulduğu bu yaklaşıma kavramsal yenilikler getirmiştir. Metrik olmayan ÇBÖ analizi için Kruskal (1964b) tarafından kavramsal ve hesaplanabilen yenilikler getirilmesi ile metrik olmayan ÇBÖ analizinin kullanımı da yaygınlaşmıştır. Kruskal ve Shepard'ın yanı sıra J-J Chang, S.C. Johnson, E.T. Klemmer, L. Nakatani, M.Wish, J.D. Carroll, J.De Leeuw, T. Gleason, L. Guttman, J.C. Lingoes, W.S. Torgerson, F.Young gibi önemli bilim insanları da verilerin farklı türleriyle ilgilenmek amacıyla, bu yaklaşımlardan yola çıkarak yeni yaklaşımlar oluşturmuşlardır. C.H. Coombs ve Michigon Üniversitesindeki çalışma arkadaşları tarafından metrik olmayan ya da ordinal verilerin çok boyutlu gösterimi için yeni düzenlemeler yapılmıştır. Fakat bu özel metotlar gerçek verilerin çok boyutlu analizi için yaygın biçimde kullanılmamaktadır. Bunun sebepleri ise az sayıdaki nesnelere hariç bilgisayar programları için dönüştürülecek yeterli formülün olmaması ve metrik olmayan verilerle çok önemli metrik bilgileri elde etmenin yetersiz olmasıdır (Davison, 1983).

Alternatif en küçük kareler ölçeklemesi (ALSCAL) algoritması, metrik olmayan ÇBÖ algoritması içerisinde yaygın bir kullanıma sahip iken Takane ve arkadaşları (1977) tarafından önerilen karmaşık fonksiyonların optimizasyonu ile

ölçekleme (SMACOF) algoritmasının Groenen ve Borg (1997) tarafından metrik durum için tanımlanması ile SMACOF algoritması hem metrik hem de metrik olmayan durumlar içinde kullanılabilir hale gelmiştir.

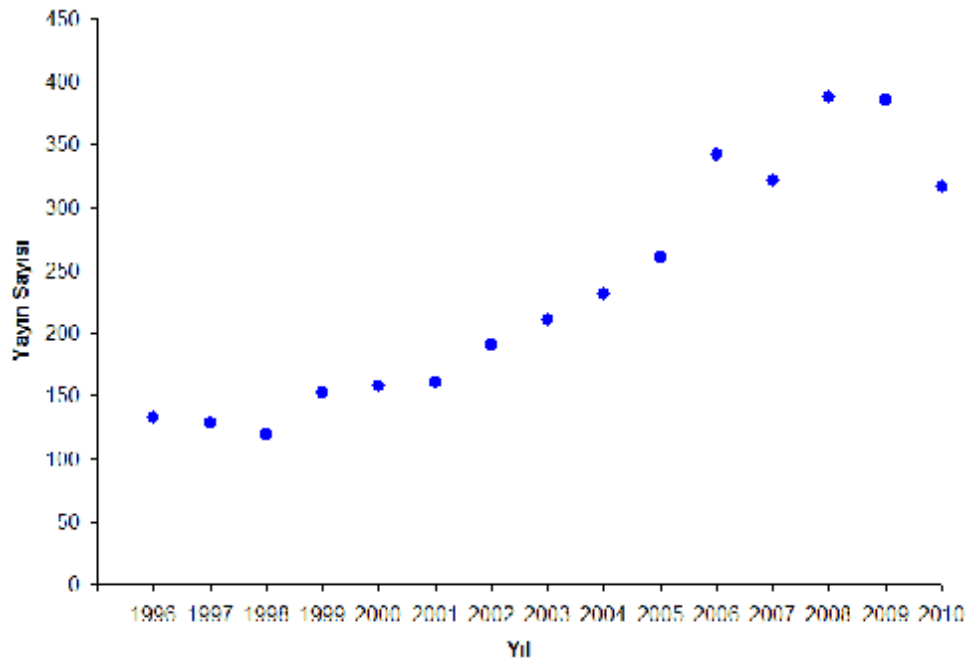
1960-1980 yılları arasında da metrik ve metrik olmayan ÇBÖ analizi ile ilgili pek çok uygulama yapılmış ve oldukça yol kat edilmiştir. Regresyon, kümeleme analizi, faktör analizi gibi çok değişkenli veri analizlerinin, boyutların yorumlanmasına ilişkin yaklaşımları, 1980 ve sonrasında ise çok boyutlu ölçekleme ile ilgili çalışmalar devam etmiştir. Çalışmaların sonunda da, üçlü çözümleme modelleri (Three way unfolding models), dağılımsal çok boyutlu ölçekleme modelleri (Stochastic MDS models), simetrik olmayan matris modelleri (nonsymmetric models) ve hibrid modeller denilen melez modeller ortaya çıkmıştır. Ancak en dikkat çeken çalışmalar barkod verilerinin çok boyutlu ölçeklemede kullanılması ve sayımsal ürün tasarımı rekabetçi eylem modelleri (prescriptive or normative product design and competitive action models) şeklinde sınıflandırılmasıdır (Carroll ve Green, 1997).

ÇBÖ analizi ilk olarak psikoloji alanında ortaya çıkmasına rağmen günümüzde psikolojinin yanında pazarlama, sosyal bilimler, tıp, eğitim bilimleri, sınıflama problemleri, arkeoloji ve kimya gibi birçok farklı alanda uygulamaları bulunmaktadır.

Bilgisayar programlarının gelişmesiyle, ÇBÖ tekniğinin uygulama aşamaları kolaylaşmış ve buna bağlı olarak kullanım alanları da yaygınlaşmıştır. Özellikle pazarlama araştırmalarında çok boyutlu ölçekleme analizi oldukça sık başvurulan bir çok değişkenli analiz türü olmuştur. ÇBÖ tekniğini iktisatçılar ve pazarlama araştırmacıları yaygın bir ürün alanında tüketici tepkilerini araştırmak için kullanmaktadır.

Thomson Reuters tarafından hazırlanan SCI, SCI-Expanded, SSCI ve AHCI indekslerince taranan çalışmalara ulaşılabilen, ISI Web of KnowledgeSM (<http://apps.isiknowledge.com>) arama sayfasında arama kutucuğuna “multidimensional scaling” terimi yazılıp yıllara göre arama yapılarak 1996 ile 2010 yılları arasında, yıllara göre içerisinde çok boyutlu ölçekleme kavramı geçen çalışma sayıları Şekil 1.1’de verilmiştir. Yıllara göre çalışma sayıları incelendiğinde

içerisinde çok boyutlu ölçekleme analizi kavramı geçen çalışma sayısının giderek arttığı görülmektedir. 1996 yılında çok boyutlu ölçekleme kavramı geçen çalışma sayısı 133 iken bu sayı 2010 yılında 316 olmuştur. 1996-2010 yılları arasında 120 farklı bilim alanında 3497 makale yayınlanmıştır. Bu makalelerin 412 si ekoloji alanında, 349 tanesi deniz ve tatlı su biyolojisinde, 230 tanesi bilgisayar ve yapay zekâ alanında, 215 tanesi çevre bilimlerinde, 175 tanesi ise bitki bilimlerinde yayınlanmıştır. İstatistik ve olasılık alanında yayınlanan makale sayısı ise 136 dır.



Şekil 1.1. Yıllara göre içerisinde “çok boyutlu ölçekleme” kavramı geçen çalışma sayıları

Çok boyutlu ölçekleme analizinde seçilen uzaklık fonksiyonunun stress değeri üzerindeki etkisinin inceleneceği çalışmamızdaki genel amaç farklı uzaklık fonksiyonlarının hangi algoritmalarda daha başarılı olduğunun tespit edilmesidir. Bu amaçla gerçek ve simülasyon verileri üzerinde farklı çok boyutlu ölçekleme algoritmalarında farklı uzaklık fonksiyonları ile elde edilen stress değerleri karşılaştırılacaktır. Çalışma kapsamında ayrıca çok boyutlu ölçekleme algoritmalarının kullanımına örnek olması amacıyla iki uygulamaya yer verilecektir. Uygulamalardan ilki Türkiye’deki 12 istatistikî bölgenin seçilen 10 sosyo-ekonomik

göstergeler bakımından karşılaştırılmasıdır. Diğer uygulama ise Türkiye ve 36 Avrupa ülkesinin 2005 yılına ait 12 sosyo-ekonomik göstergeye göre gelişmişlik değerlendirilmesinin yapıldığı çalışmadır.

Çalışma kapsamında yapılan uygulamaların özetlendiği ve çalışmanın amacı ile öneminin belirtildiği bu bölümün ardından, ikinci bölümde çok boyutlu ölçekleme analizi ile ilgili önceki çalışmalar verilecektir.

Çalışmanın üçüncü bölümünde öncelikle çok değişkenli veri yapısı tanıtarak çok boyutlu ölçeklemede temelde ilgilenilen uzaklık matrisinin özellikleri incelenecektir. Daha sonra uzaklık matrisinin elde edilmesinde kullanılan uzaklık ölçüleri değişken türlerine bağlı olarak incelenecektir. Bu bölümün ardından dördüncü bölümde de çok boyutlu ölçekleme analizi ile ilgili temel kavram ve algoritmalar bahsedilecektir Çok boyutlu ölçekleme algoritmaları metrik ve metrik olmayan algoritmalar olarak iki alt başlıkta ele alınacaktır. Burada özellikle klasik çok boyutlu ölçekleme ve karmaşık fonksiyonların optimizasyonu ile çok boyutlu ölçekleme algoritması (SMACOF) detaylı olarak incelenecektir. Stress değerleri, boyut sayısının belirlenmesi ve boyutların yorumlanması ile ilgili açıklamalar yapıldıktan sonra her algoritma için örnek iki uygulama yapılacaktır.

Çalışmanın beşinci bölümünde ölçme düzeyi en az eşit aralıklı ölçek olan birimler arasında kullanılan City-Block, Öklid ve Standartlaştırılmış Öklid uzaklıklarının klasik metrik çok boyutlu ölçekleme algoritması ve metrik karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında stress değeri üzerindeki etkileri incelenecektir. Uzaklık fonksiyonlarının stress değeri üzerindeki etkisi, öncelikle sınıflama literatüründe iyi bilinen farklı özellikler taşıyan gerçek veri setleri üzerinde gösterilecektir. Daha sonra gerçekleştirilecek simülasyon çalışmasında farklı örneklem hacmi, farklı ortalama vektörleri ve farklı kovaryans matris yapılarında çok değişkenli normal dağılımdan üretilen veri setlerinde uzaklık fonksiyonlarının etkisi incelenecektir.

Çalışmanın son bölümünde ise çok boyutlu ölçekleme analizi ile ilgili yapılan uygulamalardan ulaşılan sonuçlara ve önerilere yer verilip çalışmada hangi sonuçlara ulaşıldığı anlatılacaktır.

2.ÖNCEKİ ÇALIŞMALAR

Young ve Household (1938) ile Richerdson (1938)'in çalışmaları ile başlayan çok boyutlu ölçekleme analizi ile ilgili çalışmalar başlangıçta uygulamada karşılaşılan zorluklardan dolayı genelde teorik yönde olmuştur. Bilgisayar teknolojisinin gelişmesiyle sağlanan kazanımlar nedeniyle uygulamadaki zorluklar aşılmış ve buna bağlı olarak çok boyutlu ölçekleme analizinin kullanımı yaygınlaşmıştır.

Boyut indirgeme tekniklerinden biri olarak kabul edilen çok boyutlu ölçekleme analizi, çok boyutlu verinin daha düşük boyutlarda grafiksel olarak gösterimini sağlayan bir veri görselleştirme tekniğidir. Çok boyutlu verideki birim veya değişkenler arasındaki ilişkilerin grafiksel olarak görülmesini sağlamasından dolayı farklı birçok disiplinde yaygın olarak kullanılmıştır. Son 15 yıl içerisinde SCI, SCI-Expanded, SSCI ve AHCI indekslerince taranan dergilerde yaklaşık 3500 makale yayınlanmıştır. Bu yayınların disiplinlere yayılımı incelendiğinde 120 farklı disiplinde çok boyutlu ölçekleme ile ilgili çalışmalar yapıldığı görülmektedir. Bu bölümde uluslararası düzeyde yapılan çalışmalar ile yurtiçinde yapılan tez ve makale çalışmalarından bazılarına yer verilmiştir.

Kurt (1992), “Çok Değişkenli İstatistiksel Analiz Tekniklerinden Çok Boyutlu Ölçekleme ve Bir Uygulama” isimli çalışmasında deterjan markalarının özelliklerinin tüketiciler üzerindeki etkilerini irdelemiştir.

Çalış (1995), “Tüketici Marka Algılamaları ve Satın Alma Davranışı: Binek Otomobil Örneği” isimli çalışmasında, tüketicilerin marka algısı ve satın alma tutumlarını incelemiştir. Çalışmada 8 otomobil markasının (Skoda Favorit, Lada Samara, Opel Vectra, Doğan SLX, Renault 9, Broadway, Şahin, Toyota Corolla ve Ford Taunus) tüketici algılama haritası çok boyutlu ölçekleme analizi ile elde edilmiştir. Analiz sonuçlarına göre, ‘Lada Samara’ ve ‘Skoda Favorit’ birbirine en çok benzeyen, ‘Opel Vectra’ ve ‘Skoda Favorit’ birbirine en az benzeyen otomobil modelleri olarak ortaya çıkmıştır.

Fındıkkaya (1995), “Çok Boyutlu Ölçekleme Analizi ve Bir Uygulama Denemesi” isimli çalışmada ise 5 Ulusal gazetenin (Hürriyet, Sabah, Milliyet,

Zaman, Yeni Yüzyıl) okurlar tarafından nasıl algılandığını ve benzerliklerini incelemiştir.

Hall (2001), bir erkek ve bir kadın konuşmacının konuşma örnekleri arasındaki algısal uzaklıklardan oluşan uzaklık matrisine dayalı çok boyutlu ölçekleme analizi ile konuşma örneklerini kadın ve erkek konuşmacıların sözcükleri olarak sınıflandırmıştır.

Çılan ve Demirhan (2002), Türkiye’deki illeri sosyal ve ekonomik göstergeler bakımından çok boyutlu ölçekleme ve kümeleme analizi ile incelemiştir. Sosyal göstergeler ile gerçekleştirilen çok boyutlu ölçekleme analizi sonucu Şırnak, İstanbul ve Eskişehir illerinin kendilerine özgü diğer illerden farklı sosyal bir yapıya sahip olduğu görülmüştür. Ekonomik göstergeler bakımından gerçekleştirilen çok boyutlu ölçekleme analizi sonucuna göre ise İstanbul, Ankara ve İzmir illerinin diğer illerden ayrıldığı görülmüştür. Tüm sosyo ekonomik göstergeler bir arada incelendiğinde ise İstanbul ve Ankara ilinin diğer illerden belirli bir şekilde ayrıldığı tespit edilmiştir.

Candan (2003) “Büyük Ölçekli Perakendecilik Konumlama ve Algılama Haritalarının Kullanılması” isimli çalışmasında İstanbul’daki 7 büyük alışveriş merkezinin çok boyutlu ölçekleme tekniği kullanılarak tüketici üzerindeki algısı incelenmiştir. Belirli markaların birbirine yakın, bir kısmının ise birbirine hiç benzemediği sonucu ortaya çıkmıştır.

Filiz ve Çemrek (2003), Türkiye ile Avrupa Birliğine üye ülkelerin karşılaştırılmasında 15 ülkeyle Türkiye’nin ve 16 ülkenin kendi aralarında birbirleriyle olan farklılıklarını incelemiştir. Bu karşılaştırmada ÇBÖ analizinin yanında Faktör analizi de kullanılmıştır. Boyut sayısı 2 ve 3 alınarak her bir boyut için birbirine en uzak ve en yakın olan ülkeler belirlenmiştir.

Dura ve ark. (2004), Türkiye ile Avrupa Birliğinin karşılaştırıldığı çalışmalarında, beşeri sermaye açısından Türkiye’nin Avrupa Birliği karşısındaki kalkınma seviyesini incelemiştir. Beşeri sermaye göstergesi olarak gördükleri değişkenleri ÇBÖ analizi ile birlikte kümeleme analiziyle kullanmışlardır. Bunun sonucunda incelenen 26 ülkenin 5 ayrı grup oluşturduğu, Türkiye’nin tek başına bir grupta yer aldığı ve hiçbir AB ülkesi ile aynı gelişmişlik düzeyini paylaşmadığı sonucuna varılmıştır.

Ji ve Zha (2004), kablosuz eklemeli algılayıcı ağlarında, algılayıcı yerleştirilmesinde çok boyutlu ölçekleme analizinin kullanımını göstermişlerdir. Çalışmalarında ağdaki farklı algılayıcı sayıları ile düzenli ve düzensiz radyo frekansı durumlarında iteratif çok boyutlu ölçekleme algoritması ile klasik çok boyutlu ölçekleme algoritmalarını ortalama hata kriterine göre karşılaştırmışlardır. Karşılaştırma sonucunda iteratif algoritmanın klasik çok boyutlu ölçeklemeye göre daha etkili olduğunu belirtmişlerdir.

Ong ve ark (2004), karmaşık ağ sistemlerinin alt sistemlere bölünmesi ile gerçekleştirilen ağırlıklandırılmış çok boyutlu ölçekleme analizinin değerlendirilmesini kolaylaştırmak için yorumlayıcı yapısal modelleme yöntemi ile analitik ağ süreci prosedürlerini bir arada kullanılmasıyla elde ettikleri yeni bir yöntem önermişlerdir. Çalışmalarında gerçek bir veri seti üzerinde önerdikleri yöntemin sağladığı avantajları ortaya koymuşlardır.

Leung ve Lau (2004), city-blok uzaklığının kullanıldığı iki boyutlu ölçekleme analizinde stress değerini en küçükmek için yeni bir yöntem önermişlerdir. Önerilen yöntem iki adımdan oluşmaktadır. Birinci adımda stress fonksiyonunun yerel minimumunu bulmak için en küçük kareler regresyonu kullanılmaktadır. İkinci adımda ise küresel minimum değerini elde etmek için tavlama benzetim yöntemi uygulanmaktadır. Leung ve Leu (2004), önerdikleri yöntemin etkinliğini gerçek ve simülasyon verileri üzerinde göstermişlerdir. Simülasyon çalışmasında 30, 50 ve 70 birimlik örneklemeler kullanmışlardır. Çalışma sonucunda önerilen yöntemle elde edilen konfigürasyon uzaklıklarının orijinal uzaklıklara oldukça yakın olduğunu göstermişlerdir.

Yeloğlu ve ark (2005), ülkelerin karşılaştırılmasıyla ilgili yaptıkları çalışmada Türkiye'nin Avrupa Birliği ile olan uyum süreci içerisinde olduğu değişkenleri ele almışlardır. Türkiye bu değişkenler açısından diğer ülkelerle karşılaştırılmıştır. Bu çalışmada da ÇBÖ analizi ile birlikte kümeleme analizi kullanılmıştır. Grafikselleştirilerek yapılan değerlendirmede Türkiye'nin diğer ülkelerden anlamlı farklılıklar göstermediği sonucuna ulaşılmıştır.

Kaçar ve Azkan (2005), 2001 ve 2002 yıllarında Türkiye'nin çeşitli yerlerinden toplanan farklı hypericum türlerinin morfolojik özelliklerine göre

gerçekleştirilen çok boyutlu ölçekleme analizi ile türlerin gruplandırılmasını gerçekleştirmişlerdir. Öklid uzaklık fonksiyonunun kullanıldığı analiz sonucunda elde edilen iki boyutlu grafiksel gösterimde her bir türün farklı grup oluşturduğu belirgin bir şekilde gözlemlenmiştir.

Costa ve ark (2005), algılayıcı ağlarında düğüm yerleştirmeleri için dağıtılmış ağırlıklandırılmış çok boyutlu ölçekleme analizini kullanmışlardır. Ağlarda algılayıcıların yerleştirilmesinde iki önemli sorun vardır. Bunlardan birincisi ölçümlerin pahalı olması ve diğeri de ağdaki algılayıcı sayısının artması durumunda merkezin tahmininin oldukça güç olmasıdır. Costa ve ark., ağırlıklandırılmış çok boyutlu ölçekleme analizini temel alarak yeni bir dağıtılmış yerleştirme algoritmasını önermişlerdir. Gerçek bir ağ verisinde ulaşım süresi ve alınan sinyalin gücü değişkenleri bakımından önerdikleri yeni yöntemi, klasik çok boyutlu ölçekleme algoritması ve Patwari ve ark (2003), tarafından önerilen en çok olabilirlik yöntemi ile hata kareler ortalamasına göre karşılaştırılmışlardır. Karşılaştırma sonucuna göre alınan sinyalin gücü değişkeninde Patwari ve ark (2003) tarafından önerilen en çok olabilirlik yöntemi daha başarılı bulunurken, ulaşım süresi değişkeni bakımından yeni önerilen yöntem daha başarılı bulunmuştur.

Rouzier ve ark (2005), meme kanseri moleküler sınıflandırılmasının insan tümörlerinin gen yapılarına dayalı olarak yapılmasını önerdikleri çalışmalarında çok boyutlu ölçekleme analizini kullanmışlardır. Çalışmalarında kemoterapi öncesi dört farklı sınıfa ayırdıkları meme kanseri moleküllerinin kemoterapiye farklı tepkiler verdiklerini ortaya koymuşlardır. Meme kanseri moleküllerinin sınıflandırılmasında tam bağlantı aşamalı kümeleme tekniği ile çok boyutlu ölçekleme algoritmasını kullanmışlardır. Öklid uzaklığına dayalı gerçekleştirilen çok boyutlu ölçekleme ile elde edilen üç boyutlu grafiksel gösterimle meme kanseri moleküllerinin dört farklı sınıfa ayrılması gerektiğini ortaya koymuşlardır.

Cheung ve So (2005), sinyal varış sürelerinin ölçümünden elde edilen veri setinin değerlendirilmesinde basit bir mobil yerleştirme algoritmasıyla düzenledikleri yeni bir çok boyutlu ölçekleme algoritması kullanmışlardır. Algoritmanın üç veya daha fazla mobil baz istasyonunun yerleşiminin tahmininde oldukça etkin olduğunu ifade etmişlerdir. Önerdikleri yöntemin, klasik çok boyutlu ölçekleme algoritmasıyla

karşılaştırılmasında Cramer-Rao alt sınırı kullanmışlardır. Karşılaştırma sonucunda yeni yöntemin daha başarılı olduğunu belirtmişlerdir.

Ustaahmetoğlu (2005), “Marka Konumlandırma Çok Boyutlu Ölçekleme Tekniğinin Uygulanması” isimli çalışmasında Türkiye de faaliyet gösteren gazlı içecek firmaları üzerinde tüketicinin tepkisini ölçme amaçlı bir çalışma yapmıştır. Tüketicilere benzer özellik taşıyan birçok ürün sunulmuş ve ürünlerin birbirine yakın özellikler göstermesine rağmen tüketicinin belli ürünlere daha fazla rağbet gösterdiği gözlemlenmiştir. Tüketicinin bu davranışlarının altında yatan nedeni anlama ve ürün konumlandırması, stratejik yöntemin konusudur. Bu çalışmada da bu konular üstünde durulmuş ve Çok Boyutlu Ölçekleme Yöntemi ile elde edilen algılama haritaları ile konumlanan gazlı içecek markalarının; ürüne, pazara, rakiplerine ve tüketicilerine karşı geliştirip uygulaması gereken stratejiler ortaya çıkarılmıştır. Analiz sonucunda Pepsi, Coca Cola, ve Coca Cola Klasiğın tüketici algılamasında oldukça benzer olduğu tespit edilmiştir.

Brandes ve Pich (2006), klasik çok boyutlu ölçekleme algoritması için yeni bir örnekleme yaklaşımına dayalı bir teknik önermişlerdir. Yeni önerdikleri yöntemin çok yüksek boyutlu verilerde bile kolaylıkla uygulanabildiğini ve son derece hızlı çözüme ulaştığını belirtmişlerdir. Çalışmalarında önerdikleri yöntemi, Klasik, Landmark, Pivot, Süreç çok boyutlu ölçekleme algoritmaları ile yakınsama hızı ve orijinal uzaklıklara uyum bakımından karşılaştırmışlardır. Karşılaştırma sonucunda önerdikleri yöntemin diğer yöntemlere göre daha basit ve daha etkili olduğunu ifade etmişlerdir.

Şimşek (2006), yapı geçerliliğinin test edilmesinde kümeleme, çok boyutlu ölçekleme, doğrulayıcı ve açıklayıcı faktör analizlerini kullanmıştır. Çalışmada 47 maddeden oluşan çok boyutlu öfke ölçeği kullanılmıştır. Ölçek Hacettepe üniversitesi öğrenci yurtlarında toplam 542 kişiye uygulanmıştır. Çok boyutlu ölçekleme analizi sonucunda stress değerine göre uygun boyut sayısı 4 olmasına rağmen boyutlarda yer alan değişkenler bakımından diğer analizlerle farklılıklar göstermiştir.

Sığırlı ve ark (2006), Türkiye ve Avrupa birliğine üye ülkelerin sağlık göstergelerini çok boyutlu ölçekleme analizi ile incelemişlerdir. 1998-2004

periyodunda toplam 25 ülke için dünya sağlık raporundan elde edilen veri setine uygulanan çok boyutlu ölçekleme analizi sonucunda iki boyutlu konfigürasyon için stress değeri 0.02 olarak bulunmuştur. Çalışmaya dahil edilen 25 ülke iki boyutlu uzayda üç farklı sınıf oluşturmuştur. Türkiye, Slovakya, Macaristan ve çek cumhuriyeti dışındaki diğer ülkelerden birinci boyutta temel sağlık göstergeleri bakımından ikinci boyutta ise sağlık harcamaları ve milli gelirden sağlığa ayrılan pay bakımından farklılık göstermektedir.

Şahin ve Miran (2007), Bayındır ilçesi örneğinde çiftçi algılarına göre bitkisel ürünlerin risk haritasını çok boyutlu ölçekleme analizi ile elde etmişlerdir. Analiz sonucunda meyvelerin ve yem bitkilerinin risk düzeyinin düşük olduğu, diğer ürünlerin ise işletmelerde belirlenen ortalama risk oranına sahip olduğu görülmüştür. Risk değerlendirilmesinde birbirine en az benzeyen iki ürün patates ve zeytindir. Ürünlerin risk algılarının iki boyutlu konumlandırılmasında üç farklı küme elde edilmiştir.

Tetik (2007), Avrupa Birliğine üye 27 ülke ile katılımcı üye statüsündeki Türkiye'nin ekonomik, sektörel, demografik, kültürel, sosyal, psikoloji, ekonomik ve teknolojik dinamiklerden oluşan 100 değişkeni göz önünde bulundurarak Avrupa pazarının sınıflandırılmasını gerçekleştirmiştir. Sınıflandırma işlemi için kümeleme analizi ve çok boyutlu ölçekleme analizini kullanmıştır. Çok boyutlu ölçekleme analizi sonuçlarına göre Türkiye'nin tüm bu özellikler bakımından hiçbir Avrupa ülkesiyle benzer özellikler göstermediği ve kendi başına bir küme oluşturduğu görülmüştür.

Yiğit (2007), çok boyutlu ölçekleme yöntemlerini incelediği çalışmasında Türkiye'deki illerin sosyo ekonomik değişkenler bakımından konumlandırılmasını çok boyutlu ölçekleme analizi ile gerçekleştirmiştir. Sosyal değişkenler bakımından 9 değişkenin göz önünde bulundurulduğu karşılaştırma sonucunda İstanbul ve Kırşehir illeri diğer illerden belirli bir şekilde ayrılmıştır. Ekonomik özellikler bakımından 12 değişkenin kullanıldığı karşılaştırma sonucunda ise İstanbul ili yine tek başına bir küme oluşturmuştur. Ekonomik özellikler bakımında İstanbul iline en yakın il olarak İzmir bulunmuştur. Ankara ili ekonomik özellikler bakımından en iyi il iken Mardin ekonomik yönden en zayıf il olarak tespit edilmiştir.

Agarwal ve ark (2007) aynı düzlemde bulunan noktalar arasındaki Öklid uzaklıklarının metrik olmayan çok boyutlu ölçekleme algoritmasında oluşturduğu problemle ilgilenmişlerdir. Çalışmalarında problemin çözümü için aynı düzlemde bulunan noktalar arasındaki Öklid uzaklıklarını küçükten büyüğe doğru sıralayarak uzaklıklar arasındaki sıra ilişkilerini göz önünde bulundurmışlardır. Kruskal-Shepard algoritmasında dış bükey optimizasyon tekniklerini kullanarak problemin çözümüne ulaşmışlardır. Malzeme yansımalarının insan algısındaki ölçümlerini elde etmek için tasarlanan psikofizik denemelerden elde edilen veri setinin analizinde önerdikleri yöntemi kullanmışlardır. Uygulamalarında stress değerini en küçükleme bakımından genelleştirilmiş metrik olmayan çok boyutlu ölçekleme algoritması ismini verdikleri yeni yöntemin klasik çok boyutlu ölçekleme algoritmasından daha iyi olduğunu göstermişlerdir.

Pak (2007), Akdeniz çanağı ve Avrupa birliği ülkeleri içersinde Türkiye'nin turizmdeki yerini çok boyutlu ölçekleme analizi ile incelemiştir. Çalışma sonucunda Türkiye'nin Akdeniz çanağı ülkelerden turizm pazarındaki payının %3.06 olduğu ve bu değerle 8. sırada olduğu tespit edilmiştir.

Alan (2008), hazırlamış olduğu “Çok Boyutlu Ölçekleme Yöntemi ve Bir Uygulama” isimli çalışmasında Çok Boyutlu Ölçekleme Yönteminin, Pazarlama Araştırmalarında kullanımıyla ilgili bir prosedür geliştirerek sonuçların değerlendirilmesini amaçlamış ve bu doğrultuda Metrik ve Metrik olmayan yöntemleri ayrıntılı bir şekilde incelemiştir. Bununla ilgili bir bilgisayar programı geliştirilmiş ve yöntemin amaç fonksiyonlarıyla programın etkinliği belirlenmiştir. Alan (2008), çok boyutlu ölçekleme algoritmalarının incelendiği çalışmasının uygulamasında Türkiye'deki 8 sigara markasının algılanma haritasını çok boyutlu ölçekleme analizi ile elde etmiştir. Sigara markaları fiyat, tütün kalitesi, tadı ve kokusu, içiminin sert oluşu, paketinin şık oluşu, kolay ulaşılabilir, bulunabilir olması ve birbirine olan benzerliği bakımından değerlendirilmiştir. Analiz sonucunda Maltepe sigara markası diğer markalardan belirli bir şekilde ayrılmıştır.

Rosman ve ark (2008), çok boyutlu ölçekleme analizinin nümerik çözümünü hızlandırmak için vektör ekstrapolasyon yönteminin kullanılmasını önermişlerdir. Vektör ekstrapolasyonunu sabit noktalı iteratif algoritmalarının yakınsamasını

hızlandırmak için kullanmışlardır. SMACOF algoritmasında vektör ekstrapolasyonu kullanılarak yakınsamanın hızlandırıldığını gerçek ve simülasyon verileri üzerinde göstermişlerdir.

Güzel (2008), Türkiye'nin de yer aldığı seçilmiş 10 ülkenin hazır giyim sektörünü karşılaştırdığı çalışmada çok boyutlu ölçekleme analizini kullanmıştır. Çalışmada yer alan 10 ülke Öklid uzaklığının kullanıldığı çok boyutlu ölçekleme analizi ile 2 boyutlu düzlemde konumlandırılmıştır. Grafik incelendiğinde seçilen 10 ülke 5 farklı sınıfa ayrılmıştır. Çin, Türkiye ve Amerika Birleşik Devletleri tek başlarına birer sınıf oluştururken Hindistan, Brezilya, Yunanistan bir sınıf ve İtalya, İngiltere, Fransa ve Almanya diğer sınıfı oluşturmuştur.

Yenidoğan (2008), pazarlama araştırmalarında çok boyutlu ölçekleme analizinin incelendiği çalışmada spor ayakkabı markalarının nasıl ve hangi yönleri ile algılandığı ile ilgili bir çalışma gerçekleştirmiştir. Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi İşletme Bölümündeki 163 öğrenci ile gerçekleştirilen anket sonucunda 10 spor ayakkabı markası iki boyutlu düzlemde konumlandırılmıştır. Puma, Nike ve Adidas markalarının algılanma bakımından birbirine en yakın markalar olduğu Letton markasının ise algılanma bakımından diğer markalardan belirgin farklılık gösterdiği tespit edilmiştir.

Tokuçoğlu (2008), Ankara ilinde bulunan 4 büyük devlet üniversitesinde yapılan anket çalışması ile 6 kot markasının algılama haritasını çok boyutlu ölçekleme analizi ile oluşturmuştur. Çalışma sonucunda Mavi Jeans ve Levis markalarının tüketici algılamasında en benzer markalar olduğu Diesel, Wrangler, LitleBig ve Zara Jeans markalarının ise en az benzer algılanan markalar olduğu görülmüştür.

Rehimli ve ark (2008), Türkiye ve Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) ne üye ülkelerin kadın sağlığına ilişkin sağlık göstergeleri bakımından değerlendirilmesinde çok boyutlu ölçekleme analizini kullanmıştır. Stress değerine göre uygun boyut sayısının üç olarak tespit edildiği çalışmada 30 ülke üç grupta kümelendirilmiştir. Grafik incelendiğinde Türkiye'nin 29 ülke içerisinde en çok Meksika ile benzer özellik gösterdiği tespit edilmiştir.

Şahin ve ark (2008), tarım ve çevre özelliklerini göz önünde bulundurarak gerçekleştirdikleri çok boyutlu ölçekleme analizi ile Ege Bölgesinde birbirine benzer illeri belirlemişlerdir. İncelenen bölgede yer alan 12 ilin iki ve üç boyutlu gösterimlerinin yer aldığı çalışmada dört farklı küme yapısı elde edilmiştir. Muğla ve Afyon illeri Ege Bölgesindeki diğer illere göre farklı bir konumdadır. Oluşan diğer iki kümeden biri İzmir Aydın ve Manisa illerinden oluşmakta iken diğer kümede Denizli, Isparta, Burdur, Balıkesir, Çanakkale, Kütahya ve Uşak illeri yer almıştır. Çalışmada göz önünde bulundurulmuş 35 çevresel değişken içerisinde çayır-mera oranını ile alana düşen koyun sayısı bakımından iller benzer özellik göstermiştir.

Sarıyer (2009), yaşları 12 ila 19 arasında değişen sigara içen gençlerin sigara marka tercihini etkileyen faktörlerin belirlenmesinde çok boyutlu ölçekleme analizini kullanmıştır. Bu amaçla Yozgat il merkezindeki dört okulda okuyan 1204 öğrenciye anket yapılmıştır. Yapılan çok boyutlu ölçekleme analizi sonucunda, sigara içen genç pazarında sigara tercihini etkileyen faktörler marka ismi ve paket fiyatı olarak belirlenmiştir.

Ersöz (2009), Avrupa Birliğine üye ülkeler ile Amerika, Japonya, İsrail ve Türkiye'nin bulunduğu 30 ülkenin 2007 Avrupa İnovasyon Raporundan alınan 12 değişkene göre ülkelerin savunma güçlerinin inovasyonunu değerlendirmiştir. Çalışmada çok boyutlu ölçekleme analizi, kümeleme analizi ve ayırma analizi kullanılmıştır. Çok boyutlu ölçekleme analizi ile iki boyutlu düzlemde konumlandırılan ülkeler incelendiğinde Lüksemburg'un diğer ülkelere göre oldukça farklı olduğu görülmüştür. İsviçre, Japonya, ABD ve İsrail'in savunma inovasyonunda lider ülkeler olarak tespit edildiği çalışmada Türkiye'nin bu ülkelere göre daha düşük bir seviyede olduğu görülmüştür.

Çelik (2009), müşteri temsilcilerinin performanslarına göre çağrı merkezlerinin değerlendirilmesinde kümeleme ve çok boyutlu ölçekleme yöntemlerini kullanmıştır. Çalışmada 190 müşteri temsilcisi 2 boyutlu düzlemde konumlandırılmıştır. Grafik incelendiğinde, müşteri temsilcilerinin çağrı performanslarından farklı olarak, iptalden vazgeçirme, öneri ve satış performansları bakımından farklılaştıkları görülmüştür.

Tüzüntürk (2009), suç istatistikleri üzerine yaptığı çalışmada çok boyutlu ölçekleme analizini kullanarak Türkiye'deki 81 ili iki boyutlu düzlemde konumlandırmıştır. İl bazında İstanbul ve Ankara illerinin diğer illerden belirgin bir şekilde farklılık gösterdiği belirlenmiştir.

Cha ve ark (2009), pazarlama karmaasının düzenlenmesinde çok boyutlu ölçekleme analizinin kullanımını göstermişlerdir. Uygulamada çok boyutlu ölçekleme ile etki ve memnuniyet haritası oluşturmuşlardır. Oluşturulan haritanın kullanımı ile piyasada tüketici memnuniyetinin mevcut düzeyinin analiz edilebileceğini ve tüketici tercih ve tatminine dayalı olarak pazarlama karmaasının değiştirilmesinde pazarlamacılara yol gösterebileceğini ifade etmişlerdir.

Lala (2009), ülkelerin ekonomik koşullarını, politik yapılarını, mesleki eğitimlerini, iş kültürlerini, iş güçlerini ve fikir ayrılıklarını göz önünde bulundurarak gerçekleştirdiği çok boyutlu ölçekleme ile ülke resmi için robust bir ölçek geliştirmiştir. Meksika örneğinde robust bir ölçek geliştirmek için çok boyutlu ölçekleme ile birlikte açıklayıcı ve doğrulayıcı faktör analizini kullanmıştır.

Vera ve ark (2009), çok boyutlu verilerin sonlu karma dağılım modelleri ile kümelenmesinde uygun küme sayısının belirlenmesi için çok boyutlu ölçekleme analizini kullanmışlardır. Çalışmalarında çok boyutlu veriyi öncelikle çok boyutlu ölçekleme analizi ile iki boyuta indirgemişlerdir. İndirgenmiş boyutlardan elde edilen konfigürasyon uzaklıklarının normal dağılıma sahip olduğu varsayılarak veri seti içerisindeki gizli kümeler normal dağılımların karmaası ile ortaya çıkarılmıştır. Önerdikleri yöntemin uygulanabilirliğini hem gerçek hem de simülasyon verileri üzerinde göstermişlerdir.

Akküçük (2009), çok boyutlu ölçekleme tekniği olarak Torgersen ölçekleme yöntemi ile temel bileşenler analizini karşılaştırmıştır. Her iki yöntemin benzer ve farklı özelliklerin ortaya konduğu çalışmada R dilinde yazılan program kodları ile örnek uygulamalara yer verilmiştir.

Machado ve ark (2010), 2000-2009 dönemi boyunca dünya borsa indekslerinden elde edilen finansal veri serisinin dinamik özelliklerinin analizinde çok boyutlu ölçekleme ve kesirli fourier dönüşümünü kullanmışlardır. Bölgesel bir kriter altında seçilen 10 adet borsanın günlük değişimlerinin analizinde öncelikle

kesirli fourier dönüşümü ve kesirli hesap yöntemleri ile sinyal analizi gerçekleştirilmiştir. Çalışmanın ikinci aşamasında ise elde edilen çok boyutlu veri çok boyutlu ölçekleme analizi ile üç boyutlu uzayda temsil edilmiştir. Elde edilen grafiksel gösterim ile borsalar arasındaki ilişkiler tanımlanmıştır.

Masnack ve ark (2010), Fen, Teknoloji, Mühendislik ve Matematik alanlarında öğrenim gören yüksek okul ve kolej öğrencilerinin bilimsel mesleklere karşı örtük ve açık tutumlarının ölçülmesi ile elde edilen veri setinin analizinde çok boyutlu ölçekleme kullanmışlardır. Çalışmalarında yaş ve cinsiyetin diğer kariyer seçim kriterlerine göre daha az önemli olduğunu tespit etmişlerdir. Çalışmada ayrıca öğrencilerin bilimsel çalışmaların zorluğu, sosyal hayattan uzaklaşma ve yaratıcı fikir üretmedeki sıkıntılardan dolayı akademik kariyerdan uzaklaştıkları ortaya konmuştur.

Ding ve Davison (2010), en küçük kareler metodunun kullanıldığı metrik çok boyutlu ölçekleme analizinde boyut ve uyumun değerlendirilmesi için Akaike Bilgi Kriterinin (AIC) kullanılmasını önermişlerdir. Gerçek ve simülasyon verileri üzerinde çok boyutlu ölçekleme analizinde en iyi modelin belirlenmesinde Akaike Bilgi Kriterinin faydalı olabileceğini göstermişlerdir. En küçük kareler metodunda boyut ve model uyumunun AIC ile belirlendiği çalışma sonucunda Ding ve Davison (2010), AIC'in ağırlıklandırılmış çok boyutlu ölçeklendirme gibi diğer metrik çok boyutlu ölçeklendirme yöntemlerinde de kullanılabileceğini ifade etmişlerdir

Bülbül ve Köse (2010), Türkiye de bölgeler arası iç göç hareketlerini çok boyutlu ölçekleme yöntemi ile incelemişlerdir. Türkiye'deki 12 istatistikî bölgenin çok boyutlu ölçekleme analizi ile üç boyutlu uzayda konumlandırılması sonucunda üç farklı sınıfın olduğu tespit edilmiştir. Analiz sonuçlarına göre İstanbul ve Kuzeydoğu Anadolu bölgelerinin diğer bölgelerden anlamlı olarak farklı bir konumda olduğu ve Batı Marmara, Doğu Marmara ile Batı Anadolu bölgelerinin birbirlerine en benzer bölgeler olduğu belirlenmiştir.

Machado ve ark (2011), 15 borsa değeri arasındaki süre-değişim ilişkisini grafiksel olarak göstermek için çok boyutlu ölçekleme analizini kullanmışlardır. Çalışma sonucunda benzer davranış gösteren hisse senedi piyasalarını istikrarlı yada gelişmekte olan kümeler olarak gözlemlemişlerdir. Elde edilen grafiklerin çok

değişkenli ekonometrik modellerin oluşturulmasında kullanılabileceğini ifade etmişlerdir.

Bechavin ve ark (2011), yüksek boyutlu verilerin analizinde çok boyutlu ölçekleme analizinin etkinliğini tekil değer dönüşümünü kullanarak geliştirmişlerdir. İteratif çok boyutlu ölçekleme algoritmalarında başlangıç koordinatlarının, tekil değer dönüştürmesi ile elde edilen bileşenlerden seçilmesi durumunda rassal başlangıçlı algoritmalara göre daha avantajlı olduğunu ortaya koymuşlardır. Çalışmalarında farklı örneklem hacmi ve değişken sayısına sahip 13 adet veri seti kullanmışlardır. Önerdikleri yöntemin rassal başlangıç durumuna göre etkinliğini stress değerine göre değerlendirmişlerdir. Çalışma sonucunda önerdikleri yöntemin stress değerini en küçüklemede daha başarılı bulunduğunu ve ayrıca yakınsama hızının da rassal başlangıca göre daha iyi olduğunu ifade etmişlerdir.

Connor ve Siegrist (2011), İsviçre halkını temsil eden bir örneklem ile gerçekleştirdikleri deneysel çalışmada insanların biyoteknolojiye karşı davranışlarının analizinde çok boyutlu ölçekleme analizini kullanmışlardır. İsviçre de 2008 yılında 632 kişiyle yüz yüze görüşerek 29 farklı biyoteknolojik uygulamanın değerlendirilmesinden elde edilen veri seti çok boyutlu ölçekleme ile analiz edilmiş ve bilişsel bir harita oluşturulmuştur. Deneysel çalışmanın sonucunda biyoteknolojik uygulamalar için insan tercihlerinin istikrarlı olduğu ortaya konmuştur.

Dong ve Ding (2011), Amerika Birleşik Devletlerinde gençlerin tipik ve tipik olmayan davranış profillerinin analizinde çok boyutlu ölçekleme analizini kullanmışlardır. Çalışmada kullanılan veri seti dünya sağlık örgütü tarafından gerçekleştirilen okul çağı çocuklarının sağlık davranışlarının araştırılmasından elde edilmiştir. Araştırmada 10-15 yaş grubundaki çocuklara alkol kullanma, silah taşıma, fiziksel kavga sonucu yaralanma, esrar kullanma, kokain ve LSD gibi bağımlılık yapan maddeleri kullanma, sigara kullanma sıklıkları yanında arkadaş edinme kolaylığı, kendisini ne kadar mutlu hissettiği, kendisini ne kadar sempatik bulduğu ve kötü alışkanlıklardan kurtulma isteği sorulmuştur. Çalışmada demografik özellikler olarak cinsiyet, etnik kimlik, ikamet yeri, yaş ve eğitim düzeyi kullanılmıştır. Çalışma sonucunda gençlerin olumsuz davranışlarda daha fazla bulunduğu ve daha az olumlu davranışlar sergiledikleri gözlemlenmiştir.

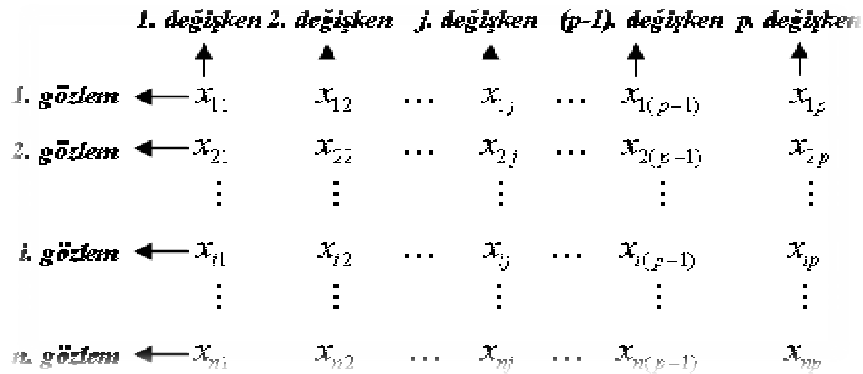
Dickes ve ark (2011), 47 Avrupa ülkesinin sosyal uyumunun değerlendirilmesinde çok boyutlu ölçekleme analizini kullanmışlardır. 2008 yılında Avrupa ülkelerinin değerlendirilmesi çalışmasında elde edilen veri seti kullanılmıştır. Çalışma sonucunda 47 Avrupa ülkesi kuzey, güney, batı, doğu, eski Sovyetler birliğinden ayrılan ülkeler ve Türkiye olarak 6 ayrı sınıfa ayrılmıştır. Türkiye kurumsal güven bakımından Kuzey ve Batı Avrupa ülkelerine daha yakın iken politik katılım, sosyo kültürel katılım ve önemli ilişkiler bakımından diğer ülke gruplarına daha yakındır.

3. ÇOK BOYUTLU VERİ VE UZAKLIK FONKSİYONLARI

Bu bölümde çok boyutlu bir verinin bir matris ile nasıl ifade edildiği ve değişkenlerin ölçülmesinde kullanılan ölçek türleri ile çok boyutlu ölçekleme analizinde temelde ilgilenilen uzaklık matrisi ve uzaklık matrisinin elde edilmesinde kullanılan uzaklık fonksiyonları incelenecektir.

3.1. Çok Değişkenli Verinin Matris Gösterimi

Bir araştırmada veya deneyde sosyal ya da fiziksel bir olayın anlaşılması için n sayıda birey, nesne ya da deneysel gözlemin her biri üzerinden gözlenen $p > 1$ sayıda değişken ya da karakteristiğin kayıt edilmiş ölçüm değerlerine çok değişkenli veri denir.



Şekil 3.1. Çok değişkenli veride gözlem ve değişken vektörleri

Böyle bir veri kümesi genellikle,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \mathbf{K} & x_{1j} & \mathbf{K} & x_{1(p-1)} & x_{1p} \\ x_{21} & x_{22} & \mathbf{K} & x_{2j} & \mathbf{K} & x_{2(p-1)} & x_{2p} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} & & \mathbf{M} & \mathbf{M} \\ x_{i1} & x_{i2} & \mathbf{K} & x_{ij} & \mathbf{K} & x_{i(p-1)} & x_{ip} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} & & \mathbf{M} & \mathbf{M} \\ x_{n1} & x_{n2} & \mathbf{K} & x_{nj} & \mathbf{K} & x_{n(p-1)} & x_{np} \end{bmatrix} \quad (3.1.)$$

biçimindeki $n \times p$ tipinde bir \mathbf{X} matrisi ile ifade edilir. Burada x_{ij} , i . gözlem ya da nesne üzerinden gözlenen j . değişken ya da karakteristik değerini ifade eder. Bu durumda (3.1)'deki eşitlikte gösterilen \mathbf{X} veri matrisi $i = 1, 2, \dots, n$ ve $j = 1, 2, \dots, p$ olmak üzere her biri $1 \times p$ boyutlu \mathbf{x}_i gözlem vektörleri,

$$\mathbf{X} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T)^T \quad (3.2.)$$

şeklinde de ifade edilebilir.

3.2. Ölçek Türleri

Ölçme en genel anlamda nesne ve olaylara, bazı kurallara uygun olarak sayılar vermektir (Balcı, 2005).

Ölçme ile ilgili temel teoriler Klasik Ölçme (Classical Measurement), Temsili Ölçme (Representational Measurement) ve Operasyonel Ölçme (Operational Measurement) olarak özetlenebilir. Klasik ölçme teorisinde gerçekte varolan ve sadece sayısal olan özelliklerin ölçülmesi anlayışı varken Temsili Ölçüm Teorisi ölçülen özelliklerin yine gerçeklik ile bağlantılı fakat sayısal olmayan bazı özelliklerin de ölçülebileceğini kabul eder. Operasyonel Ölçme Teorisinde ise ölçümün yapılan ölçme şeklinden ve adımlarından doğduğu anlayışı kabul edilir (Alan, 2008).

Özde kavramlar ölçülemez, onların somut göstergeleri ölçülür ve bunlara değişken adı verilir (Judd ve ark., 1991). Bu ifadeyi örneklendirmek gerekirse, bir kişinin futbol yeteneğinin olduğunun söylenebilmesi için, yeteneğin göstergeleri olarak kabul edilebilecek, kişinin topla oynama süresi, hızı, çabuk düşünüp çabuk uygulaması gibi değişkenlerle ölçülüp değerlendirilmesi gereklidir.

Çok boyutlu ölçekleme analizinde, veri elde edilirken değişkenler için hangi ölçeğin ve hangi veri türünün kullanıldığının bilinmesi gerekir. Çünkü ÇBÖ analizi için geliştirilen tekniklerden hangisinin kullanılacağına buna göre karar verilir.

Harvard'lı psikolog S.S.Stevens'in 1940'lı yılların başlarında psikofizikte kullanılan ölçeklerin hiyerarşisini tanımlamak için sınıflama türünden (nominal), sıralı (ordinal), aralıklı, oranlı terimlerini ilk defa kullandı ve geçerli ölçeklere göre istatistiksel prosedürü sınıfladı (Velleman ve Wilkinson, 1993).

3.2.1. Nominal (İsimsel) Ölçek

Veri ölçümü sınıflandırmadır, sadece veriler farklı gruplara ayrılır. Kısıtlaması en az olan ve bunun yanında en zayıf ölçektir. Bu ölçek ile ölçülmüş değişkenler durum, etiket, kategori veya isimlerle temsil edilir. Bu simgelere üstünlük veya sıralama gibi anlamlar yüklenemez. Örneğin plaka numaraları, göz rengi, cinsiyet, şehir isimleri, saç rengi, doğum yeri, öğrenci numaraları, kan grubu gibi. Her ne kadar isimsel değerler alan değişkenler sayısal değerlerle ilişkilendirilseler de bu sayısal değerlerle herhangi bir matematiksel işlem yapılamaz. Kan gurupları için A grubu olanların 1, B grubu olanların 2, 0 grubu olanların 3 ile kodlanması gibi. Bu değerler arasında bir sıralanma söz konusu değildir. Örneğin 0 grubunun 3 yerine 2 ile kodlanması, A grubunun 1 yerine 3 ile kodlanması veya B grubunun 2 yerine 1 ile kodlanması arasında bir fark yoktur.

3.2.2. Ordinal (Sıralı) Ölçek

Nominal ölçekten daha güçlü bir ölçektir. Nominal ölçekte değerlendirilen nesnelerin birbirlerine göre sırası veya yönü belirlenemezken Ordinal ölçekle belirlenebilir. Nominal ölçekte olduğu gibi değişkenler durum, etiket, kategori veya isimlerle temsil edilir. Bu ölçekle ölçülen değişken değerleri arasında sıralanma sırası önemlidir. Ama bu sıralama sayısal bakımdan bir sıralama değildir. Örneğin bir kişinin kullandığı bilgisayarından memnuniyetinin; çok memnun olmasının 1, memnun olmasının 2 ve memnun olmamasının 3 ile kodlanması gibi. Durumlarda değişken değerleri arasında bir önem sırası vardır. Diğer bir ifadeyle sıralanma en az önemli durumdan en çok önemli duruma doğru ya da tam tersi biçimde yapılır. Burada bilgisayarından çok memnun olan birinin sadece memnun olan birine göre iki

kat fazla memnun olduğu söylenemez. Sadece bu kişinin diğerine göre daha fazla memnun olduğu iddia edilebilir.

Sıralı ölçek ile ölçülmüş olgular üzerinde medyan, kartil, persentil, sıra korelasyonu gibi analizler yapılabilir (Yükselen, 2003).

3.2.3. Aralık Ölçeği

Nesneler arasındaki uzaklıklar ölçülebilir. Yani iki değer arasındaki farkın anlamlı olduğu sayısal veriler aralıklı ölçekle ölçülebilir. Aralıklı ölçekte ölçülmüş değişken değerleri pozitif veya negatif gerçel sayı değerleri alabilir. Ölçümler arası farklar eşit öneme sahiptir. Örneğin 10°C ile 20°C arasındaki fark ile -40°C ile -30°C arasındaki fark eşit öneme sahiptir. Aralıklı ölçekte mutlak bir sıfır değeri yoktur. Bu nedenden dolayı değerler arasındaki oran anlamsızdır. Örneğin 20°C ile 10°C arasındaki oran ile 40° ile 20°C arasındaki oran sayısal olarak aynı olmasına karşın sıcaklığın 20°C olması, 10°C 'den iki kat daha sıcak olduğu anlamına gelmez (Servi, 2009).

Aralıklı ölçeklerde önceki iki ölçekte yapılacak işlemlere ek olarak toplama, çıkarma işlemleri yapılabilir. Böylece bu ölçekteki verilerle aritmetik ortalama, standart sapma, pearson korelasyon katsayısı gibi istatistiki tekniklerin kullanılması mümkündür (Baykul, 1999).

3.2.4. Oran Ölçeği

Güçlü bir ölçek tipidir. Oran ölçeği, aralık ölçüm düzeyini kapsamakla birlikte değerler arasında oransal önem bulunmaktadır. Bu ölçekte bir değer diğerinden ne kadar oranda az veya çok olduğu anlamlıdır. Bunun nedeni bu ölçekte mutlak bir sıfır değeri vardır. Bu düzeyde sıfır değerini alan bir değişken için bu değişkenin ölçüldüğü gözlemde bulunmadığı anlamına gelir. Yıllık otomobil satışı 500 olan bir galeri sahibinin, yıllık otomobil satışı 250 olan başka galeri sahibinden yılda iki kat fazla otomobil sattığı söylenebilir.

3.3. Uzaklık Fonksiyonları

\mathbf{X} , $n \times p$ boyutlu bir veri matrisi olmak üzere $d: \mathbf{X} \times \mathbf{X} \rightarrow R$ şeklinde tanımlanan bir fonksiyon eğer tüm $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$ için

$$\text{i. } d(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \quad \text{(Pozitiflik)} \quad (3.3.)$$

$$\text{ii. } d(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_j, \mathbf{x}_i) \quad \text{(Simetri)} \quad (3.4.)$$

$$\text{iii. } d(\mathbf{x}_i, \mathbf{x}_i) = 0 \quad \text{(Özdeşlik)} \quad (3.5.)$$

koşullarını sağlıyorsa d , \mathbf{X} üzerinde bir **uzaklık fonksiyonu** olarak adlandırılır (Deza ve Deza 2006). Yukarıda belirtilen pozitiflik, simetri ve özdeşlik koşulları yanında eğer tüm $\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k \in \mathbf{X}$ için

$$\text{iv. } d(\mathbf{x}_i, \mathbf{x}_j) = 0 \text{ sadece ve sadece } \mathbf{x}_i = \mathbf{x}_j \quad \text{(Belirlilik)} \quad (3.6.)$$

$$\text{iv. } d(\mathbf{x}_i, \mathbf{x}_j) \leq d(\mathbf{x}_i, \mathbf{x}_l) + d(\mathbf{x}_j, \mathbf{x}_l) \quad \text{(Üçgen Eşitsizliği)} \quad (3.7.)$$

koşulları da sağlanıyorsa d uzaklık fonksiyonu **metrik** olarak isimlendirilir (Webb 2002). Çok boyutlu ölçekleme analizinde $i, j = 1, 2, \dots, n$ olmak üzere $(\mathbf{x}_i, \mathbf{x}_j)$ gözlem çiftleri arasındaki uzaklıklar d_{ij} olmak üzere n birimden oluşan veri setinden elde edilecek uzaklık matrisi \mathbf{D} ,

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \mathbf{L} & d_{1n} \\ d_{21} & d_{22} & d_{23} & \mathbf{L} & d_{2n} \\ d_{31} & d_{32} & d_{33} & \mathbf{L} & d_{3n} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ d_{n1} & d_{n2} & d_{n3} & \mathbf{L} & d_{nn} \end{bmatrix} \quad (3.8.)$$

şeklinindedir. Uzaklık tanımındaki özdeşlik koşulu gereğince $i = j$ durumunda $d_{ii} = 0$ ve simetri koşulu gereğince $d_{ij} = d_{ji}$ olduğundan uzaklık matrisi \mathbf{D}

$$\mathbf{D} = \begin{bmatrix} 0 & d_{12} & d_{13} & \mathbf{L} & d_{1n} \\ d_{12} & 0 & d_{23} & \mathbf{L} & d_{2n} \\ d_{13} & d_{23} & 0 & \mathbf{L} & d_{3n} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ d_{1n} & d_{2n} & d_{3n} & \mathbf{L} & 0 \end{bmatrix} \quad (3.9.)$$

şeklinde gösterilir.

3.3.1. Nicel Değişkenler İçin Uzaklık Ölçüleri

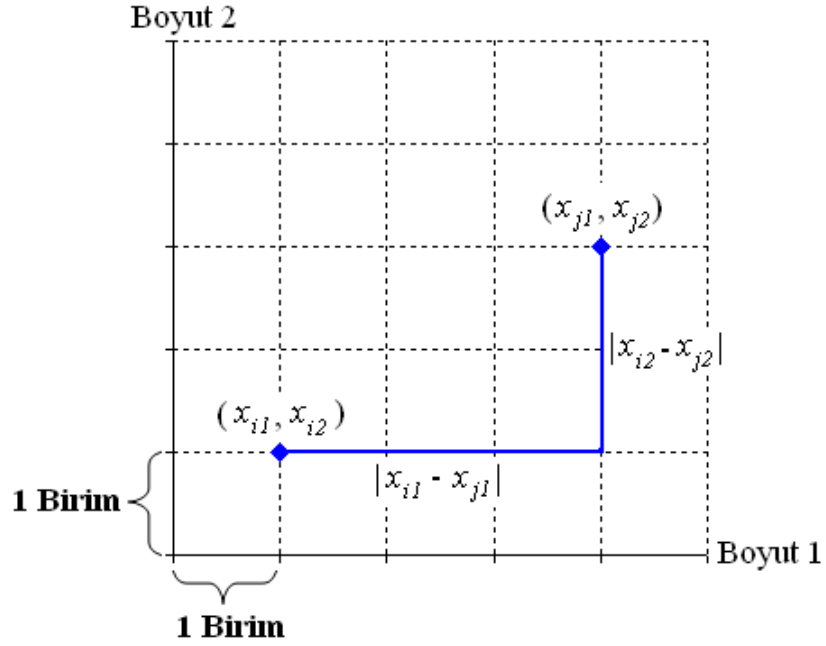
Nicel değişkenlerden oluşan veri setlerinde gözlem vektörleri arasındaki benzerlik, uzaklık ölçüleri ile belirlendiğinde iki gözlem vektörü arasındaki minimum uzaklığa sahip olan gözlem çiftleri daha benzer olarak tanımlanmaktadır. Bu kısımda, nicel değişkenlerden oluşan veri setlerinde benzerliğin belirlenmesinde yaygın kullanılan uzaklık ölçüleri ve bu uzaklık ölçülerinin özellikleri incelenecektir.

3.3.1.1. City-Block (Manhattan) Uzaklığı

City-Block uzaklığı birimlerin, değişkenleri arasındaki mutlak farkların toplamına eşittir. City-Block uzaklığı Manhattan uzaklığı olarak da bilinir. Bu ölçüt kullanıldığında farkın karesi alınmadığı için farklılıkların etkisi azalır. İki gözlem vektörü arasındaki City-Block uzaklığı $i, j = 1, 2, \dots, n$ ve $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$ olmak üzere

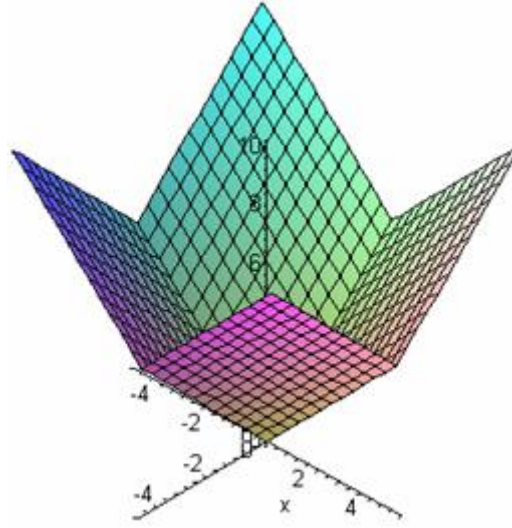
$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (3.10.)$$

eşitliği ile verilir. Değişken sayısının iki olması durumunda $p = 2$ için \mathbf{x}_i ve \mathbf{x}_j gözlem vektörleri arasındaki City-Block uzaklığının hesaplanması Şekil 3.2'de grafiksel olarak verilmiştir (Erişoğlu, 2011).



Şekil 3.2. City-Block uzaklığının iki boyutlu düzlemde hesaplanması

Şekil 3.2.'de iki boyutlu düzlemde gösterilen \mathbf{x}_i ve \mathbf{x}_j gözlemleri arasındaki Manhattan uzaklığı $3+2=5$ birimdir. City-Block uzaklığı aynı zamanda metriktir ve L_1 metriği olarak da isimlendirilir. Değişken sayısının iki ve değişkenlerin alacağı değerler $(-\infty, \infty)$ aralığında olmak üzere (x_{i1}, x_{i2}) noktasının orijin noktasına olan City-Block uzaklık fonksiyonunun grafiği Şekil 3.3.'de verilmiştir.



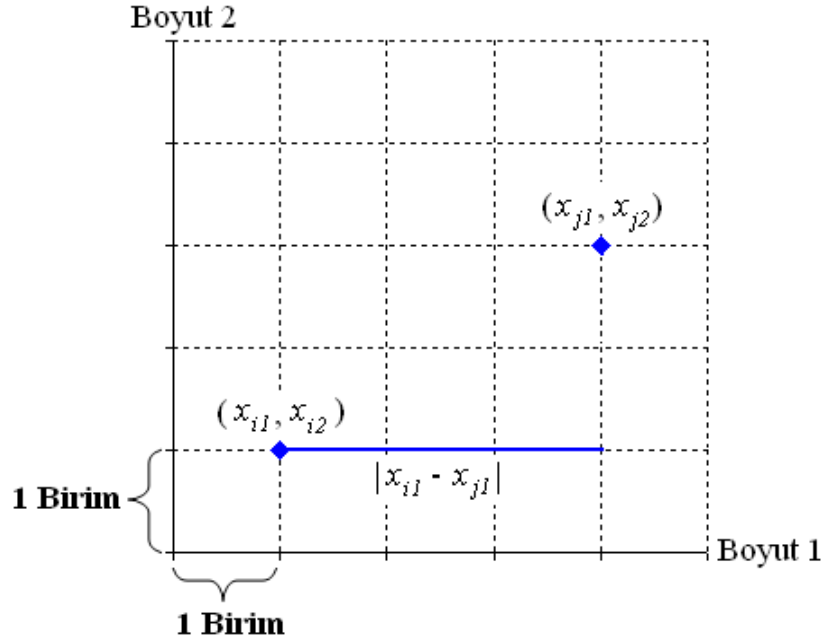
Şekil 3.3. City-Block uzaklık fonksiyonunun üç boyutlu uzayda grafiği

3.3.1.2. Chebyshev Uzaklığı

Chebyshev uzaklığı iki gözlem vektörü arasındaki maksimum farklılığa sahip boyuttaki farkı, iki gözlem vektörü arasındaki uzaklık olarak alan metrik bir uzaklık ölçüsüdür. Chebyshev uzaklığı, maksimum metrik veya L_∞ metriği olarak da isimlendirilir. İki gözlem vektörü arasındaki Chebyshev uzaklığı $i, j = 1, 2, \dots, n$ ve $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$ olmak üzere

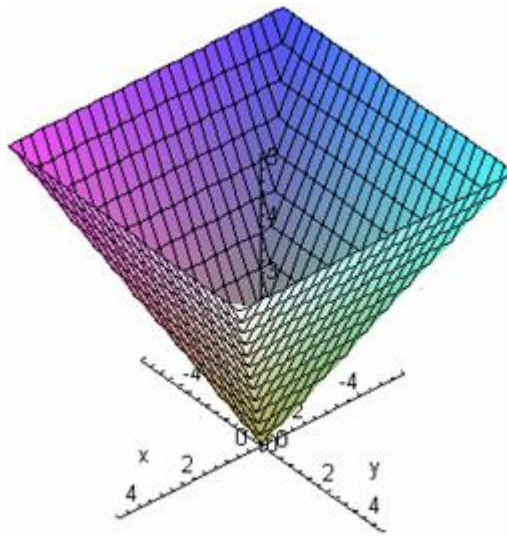
$$d(\mathbf{x}_i, \mathbf{x}_j) = \max_{k=1, \dots, p} \{ |x_{ik} - x_{jk}| \} \quad (3.11.)$$

eşitliği ile verilir. Veri setinde yer alan değişkenlerin ölçü birimleri farklı ise başka bir ifadeyle terim büyüklükleri farklı ise Chebyshev uzaklığında terim büyüklüğü fazla olan değişken diğer değişkenlere baskın gelecektir. Terim büyüklüklerinin farklı olduğu veri setlerinde, benzerliğin hesaplanmasında terim büyüklüğü fazla olan boyut dışındaki diğer boyutlardaki farklı özellikler göz önüne alınamayacağından Chebyshev uzaklığının kullanılması önerilmez. Değişken sayısının iki olması durumunda ($p = 2$ için) \mathbf{x}_i ve \mathbf{x}_j gözlem vektörleri arasındaki Chebyshev uzaklığının hesaplanması Şekil 3.4.'de grafiksel olarak verilmiştir.



Şekil 3.4. Chebyshev uzaklığının iki boyutlu düzlemde hesaplanması

Şekil 3.4.'de iki boyutlu düzlemde gösterilen \mathbf{x}_i ve \mathbf{x}_j gözlemleri arasındaki Chebyshev uzaklığı 3 birim olarak elde edilmiştir. Değişken sayısının iki olması durumunda (x_{i1}, x_{i2}) noktasının orijin noktasına olan Chebyshev uzaklık fonksiyonunun grafiği Şekil 3.5.'de verilmiştir.



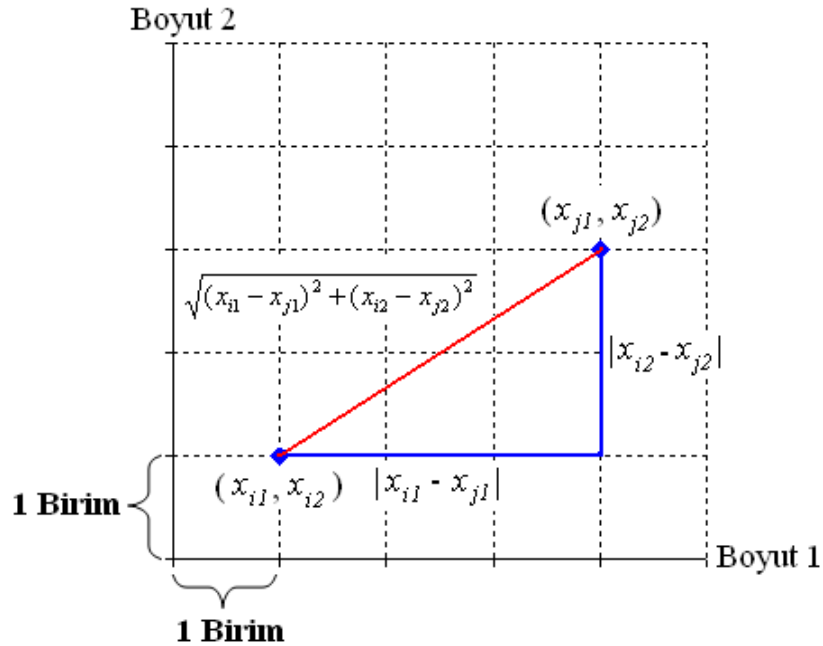
Şekil 3.5. Chebyshev uzaklık fonksiyonunun üç boyutlu uzayda grafiği

3.3.1.3. Öklid (Euclidean) Uzaklığı

ÇBÖ analizinde muhtemelen en çok bilinen ve en yaygın kullanılan uzaklık ölçüsü olan Öklid uzaklığı, gözlem vektörleri arasındaki farkların kareleri toplamının karekökünün alınması ile hesaplanır. Metrik bir uzaklık ölçüsü olan Öklid uzaklığı L_2 metriği olarak da isimlendirilir. İki gözlem vektörü arasındaki Öklid uzaklığı $i, j = 1, 2, \dots, n$ ve $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$ olmak üzere

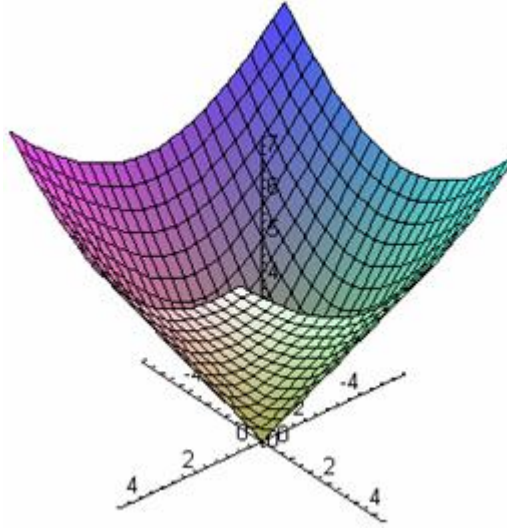
$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right)^{\frac{1}{2}} = \left[(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' \right]^{\frac{1}{2}} \quad (3.12.)$$

eşitliği ile verilir. Değişken sayısının iki olması durumunda ($p = 2$ için) Öklid uzaklığı Pisagor teoremine göre dik üçgende hipotenüs olarak isimlendirilen kenarın uzunluğuna karşılık gelir. Gözlem vektörleri \mathbf{x}_i ve \mathbf{x}_j arasındaki Öklid uzaklığının hesaplanması Şekil 3.6.'da grafiksel olarak verilmiştir.



Şekil 3.6. Öklid uzaklığının iki boyutlu düzlemde hesaplanması

Şekil 3.6.'da iki boyutlu düzlemde gösterilen \mathbf{x}_i ve \mathbf{x}_j gözlemleri arasındaki Öklid uzaklığı $\sqrt{4^2 + 2^2} = \sqrt{20}$ birim olarak elde edilmiştir. Veri setini oluşturan değişkenlerin ölçü birimlerinin farklı olması durumunda Öklid uzaklığı hesaplanırken terim büyüklüğü ve varyansı büyük olan değişkenin katkısı diğer değişkenlerden fazladır. Değişken sayısının iki olması durumunda (x_{i1}, x_{i2}) noktasının orijin noktasına olan Öklid uzaklık fonksiyonunun grafiği Şekil 3.7.'de verilmiştir.



Şekil 3.7. Öklid uzaklık fonksiyonunun üç boyutlu uzayda grafiği

3.3.1.4. Standartlaştırılmış Öklid Uzaklığı

Standartlaştırılmış Öklid uzaklığı Karl Pearson uzaklık ölçütü olarak da adlandırılır. Öklid uzaklığında her değişkenin varyansının ağırlık olarak kullanılmasıyla elde edilir. İki gözlem vektörü arasındaki standartlaştırılmış Öklid uzaklığı $i, j = 1, 2, \dots, n$ ve $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$ olmak üzere

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^p \frac{1}{s_k^2} (x_{ik} - x_{jk})^2 \right)^{\frac{1}{2}} \quad (3.13.)$$

eşitliği ile tanımlanır. Eşitlikte yer alan s_k^2 , k . değişkene ait örneklem varyansını göstermektedir ve \bar{x}_k , k . değişkene ait aritmetik ortalama olmak üzere

$$\bar{x}_k = \frac{\sum_{i=1}^n x_{ik}}{n}, \quad k = 1, \dots, p \quad (3.14.)$$

$$s_k^2 = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{n-1}, \quad k = 1, \dots, p \quad (3.15.)$$

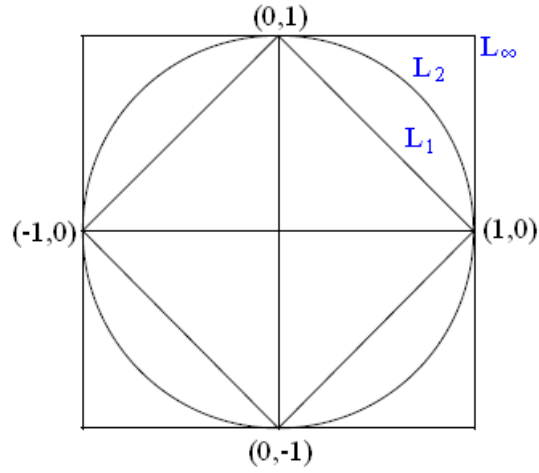
şeklinde hesaplanır. Standartlaştırılmış Öklid uzaklığı değişkenlerin ölçülmesinde kullanılan ölçü biriminden ve terim büyüklüklerinden etkilenmez.

3.3.1.5. Minkowski Uzaklığı

Nicel veriler için kullanılan uzaklık fonksiyonlarından biri de Minkowski uzaklığıdır. İki gözlem vektörü arasındaki Minkowski uzaklığı $I \geq 1$ ve $i, j = 1, 2, \dots, n$ için $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$ olmak üzere

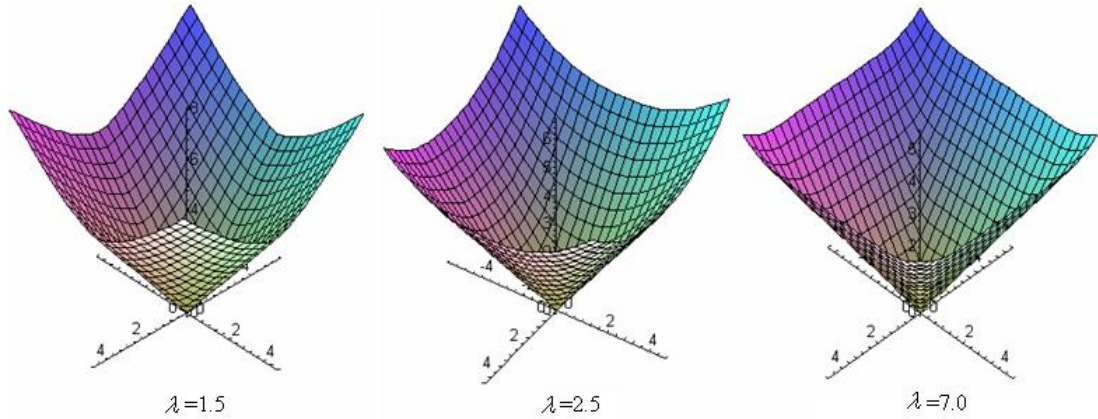
$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^I \right)^{1/I} \quad (3.16.)$$

eşitliği ile hesaplanır. Aynı zamanda Minkowski uzaklığı bir metrik uzaklık ölçüsüdür ve L_I metriği olarak da isimlendirilir. Minkowski uzaklığı genel bir uzaklık ölçüsüdür. Minkowski uzaklığında $I = 1$ alındığında Manhattan uzaklığı, $I = 2$ alındığında Öklid uzaklığı ve $I \rightarrow \infty$ için Chebyshev uzaklığı elde edilir. L_1, L_2 ve L_∞ metrikleri arasındaki geçişler birim çember üzerinde Şekil 3.8.'de verilmiştir.



Şekil 3.8. İki boyutlu düzlemde orijinden bir birim uzaklıktaki noktalardan oluşan yüzeyde L_1, L_2 ve L_∞ metrikleri

Farklı I değerleri, değişken sayısının iki ve değişkenlerin alacağı değerler $(-\infty, \infty)$ aralığında olmak üzere (x_{i1}, x_{i2}) noktasının orijin noktasına olan Minkowski uzaklık fonksiyonunun grafikleri Şekil 3.9.'da verilmiştir.



Şekil 3.9. Farklı I değerleri Minkowski uzaklık fonksiyonunun üç boyutlu uzayda grafiği

3.3.1.6. Mahalanobis Uzaklığı

Mahalanobis uzaklığı değişkenlerin varyanslarına ve karşılıklı (ikili) korelasyonlarına göre farklı özniteliklere farklı ağırlıklar verir. Diğer bir deyişle bu ölçü küme elemanlarının çok değişkenli Gauss dağılımdan geldiğini varsayar. **S**

örneklem kovaryans matrisi ve $i, j = 1, 2, \dots, n$ için $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$ olmak üzere Mahalanobis uzaklığı

$$d(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i - \mathbf{x}_j) \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j)']^{\frac{1}{2}} \quad (3.17.)$$

eşitliği ile tanımlanır. Eşitlik de yer alan \mathbf{S} örneklem kovaryans matrisi \mathbf{X} veri matrisi için

$$\mathbf{S} = \text{cov}(\mathbf{X}) = \frac{1}{n-1} \mathbf{X}' (\mathbf{I} - \frac{1}{n} \mathbf{J}) \mathbf{X} \quad (3.18.)$$

eşitliği ile hesaplanır. Kovaryans matrisi

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \mathbf{L} & s_{1p} \\ s_{21} & s_{22} & \mathbf{L} & s_{2p} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ s_{p1} & s_{p2} & \mathbf{L} & s_{pp} \end{bmatrix} \quad (3.19.)$$

şeklinindedir. \mathbf{S} matrisinin esas köşegen üzerindeki terimleri örneklem varyansını gösterirken ($s_1^2 = s_{11}$), diğer terimler değişkenler arasındaki kovaryansı gösterdiğinden \mathbf{S} matrisi varyans-kovaryans matrisi olarak da isimlendirilmektedir. \mathbf{S} örneklem kovaryans matrisinde $i, j = 1, 2, \dots, p$ olmak üzere $s_{ij} = s_{ji}$ olduğundan örneklem kovaryans matrisi simetrik bir matristir. \mathbf{X} veri matrisindeki değişkenler ilişkisiz ise yani $i, j = 1, 2, \dots, p$ ve $i \neq j$ olmak üzere $s_{ij} = 0$ ise Mahalanobis uzaklığı standartlaştırılmış Öklid uzaklığına dönüşür. $\mathbf{S} = \mathbf{I}$ olduğunda yani değişkenler arasında ilişkinin olmadığı ve değişkenlere ait varyansların birim varyansa eşit olması durumunda Mahalanobis uzaklığı karesel Öklid uzaklığına dönüşür. İki gözlem vektörü arasındaki karesel Öklid uzaklığı

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^p (x_{ik} - x_{jk})^2 = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' \quad (3.20.)$$

eşitliği ile elde edilir. Mahalanobis uzaklığı ölçü birimi ve terim büyüklüklerinden etkilenmez ve lineer dönüşümler altında değişmezdir(Duda ve ark. 2001).

Morrison (1967), Mahalanobis uzaklığında değişkenlerin ağırlıklandırarak genelleştirilmiş Mahalanobis uzaklığını önermiştir. Genelleştirilmiş Mahalanobis uzaklığı w_k , k .değişkene ait ağırlık ve $\mathbf{W} = \text{diagonal}(w_k)$, $k = 1, \dots, p$ olmak üzere

$$d(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i - \mathbf{x}_j)\mathbf{W}\mathbf{S}^{-1}\mathbf{W}(\mathbf{x}_i - \mathbf{x}_j)']^{\frac{1}{2}} \quad (3.21.)$$

şeklinde tanımlanır.

3.3.2. Nitel Değişkenler İçin Benzerlik Ölçüleri

Değişkenlerinin tümü isimsel ya da kategorik olan verilerde gözlem çiftleri arasındaki yakınlık, eşleştirme tipli benzerlik ölçüleri ile belirlenir. Ölçümler genellikle [0,1] aralığında ölçeklenmiş olsa da bazı durumlarda yüzdeler olarak da ifade edilebilir. i . ve j . gözlem arasındaki benzerlik katsayısı ya da değeri s_{ij} ile ifade edilebilir. Bu değer in sıfır olması iki gözlemin tüm değişkenleri cinsinden maksimum farklı olduğu anlamına gelir. İki gözlem arasındaki benzerlik yerine farklılık da ölçülebilir. İki gözlem arasındaki d_{ij} uzaklığı s_{ij} benzerlik katsayısından,

$$d = 1 - s, \quad (3.22.)$$

$$d = \frac{1-s}{s}, \quad (3.23.)$$

$$d = \sqrt{1 - s}, \quad (3.24.)$$

$$d = \sqrt{2(1 - s^2)}, \quad (3.25.)$$

$$d = -\ln s \text{ ve } d = \text{arccoss} \quad (3.26.)$$

dönüşümleri kullanılarak elde edilebilir.

3.3.2.1 İki Sonuçlu (Binary) Değişkenler İçin Kullanılan Başlıca Benzerlik Ölçüleri

Verideki tüm değişkenler sadece iki sonuçlu değer alıyorsa bu verideki gözlemler arasındaki yakınlık için birçok benzerlik ölçüsü tanımlanmıştır. İki sonuçlu değişkenlerin aldığı iki değerlerin öneminin eşit olup olmamasına göre bir gözlem çifti arasındaki benzerlik ölçüleri simetrik ve simetrik olmayan ölçüler şeklinde iki kategoriye ayrılır (Kaufman ve Rousseeuw 1990).

İkili ve nitel verilerle ilgili benzerlik ölçülerinde veriler, benzerlik katsayıları ve değişkenlerin varlığı ve yokluğu ilkesine göre hesaplanır. Herhangi bir değişkenin varlığı 1 ya da + ile, yokluğu 0 ya da – ile gösterilmektedir. Eğer herhangi i ve k bireyleri arasındaki benzerlik katsayısı d_{ik} ile gösterilirse, d_{ik} 'yi hesaplamada 2×2 çapraz tablosundan yararlanılmaktadır (Yiğit,2007). İki sonuçlu değişkenler içeren gözlem çiftleri arasındaki benzerlik ölçüsü hesaplanırken eşleştirme tipli benzerlik ölçütleri kullanılır. Çapraz tablo, iki sonuçlu değişkenler içeren gözlem çiftinin karşılıklı eşleşen değişken değerlerinin sayısından oluşur.

Çapraz tablosu olan Tablo 3.1 de verilen i . ve j . gözlemlerle ilgili karşılıklı değişkenlerinin ikisi de 1 değerini almışsa a , ikisi de 0 değerini almış ise d , j -nci değişken 1, i -nci değişken 0 değerini almışsa b , i -nci değişken 1, j -nci değişken 0 değerini almışsa c ile gösterilmiştir. Burada p değişken sayısını göstermek üzere $a + b + c + d = p$ olarak verilmiştir.

Tablo 3.1. İki sonuçlu p sayıda değişken içeren bir gözlem çiftinin çapraz tablosu.

		i. gözlem		Toplam
		Değişken değeri	1	
j. gözlem	1	a	b	a+b
	0	c	d	c+d
Toplam		a+c	b+d	$p = a+b+c+d$

Tablo 3.2. İki sonuçlu p değişken içeren verinin gözlem çiftleri arasındaki yakınlık için kullanılan bazı eşleştirme tipli benzerlik ölçüleri (Cox and Cox, 2001)

Ölçüm	Formül
Simpson	$s_{ij} = \frac{a}{\min[(a + b), (a + c)]}$
Hamman	$s_{ij} = \frac{a - (b + c) + d}{a + b + c + d}$
Yule	$s_{ij} = \frac{ad - bc}{ad + bc}$
Kulczynski	$s_{ij} = \frac{a}{b + c}$
Kulczynski	$s_{ij} = \frac{1}{2} \left(\frac{a}{a + b} + \frac{a}{a + c} \right)$
Eşleştirme Katsayısı	$s_{ij} = \frac{a + d}{a + b + c + d}$
Jaccard Katsayısı	$s_{ij} = \frac{a}{a + b + c}$
Rogers ve Tanimoto	$s_{ij} = \frac{a + d}{a + 2(b + c) + d}$
Sokal ve Sneath	$s_{ij} = \frac{a}{a + 2(b + c)}$
Sokal ve Sneath	$s_{ij} = \frac{2(a + d)}{2(a + d) + (b + c)}$
Gower ve Legendre	$s_{ij} = \frac{a + d}{a + \frac{1}{2}(b + c) + d}$
Gower ve Legendre	$s_{ij} = \frac{a}{a + \frac{1}{2}(b + c)}$
Dice, Czekanowski, Sorensen	$s_{ij} = \frac{2a}{2a + b + c}$

İki sonuçlu p değişkenden oluşan veriler için bir çok eşleştirme tipli benzerlik ölçüsü kullanılmakla birlikte en yaygın olarak kullanılanlar Tablo 3.2. de verilmiştir. İki sonuçlu p değişkenden oluşan gözlemler için bu kadar çok eşleştirme tipli benzerlik ölçüsünün kullanılmasının nedeni 0-0 eşleştirmesinin nasıl yorumlandığı

ile ilgilidir. Bazı durumlarda 0-0 eşleştirmesi, 1-1 eşleştirmesi ile eşdeğer olabilir. Örneğin cinsiyetin erkek için 1 değeri ve bayan için 0 değeri ile kodlanması veya tam tersi şeklinde kodlanması arasında bir fark yoktur. Bazı durumlarda ise 0-0 eşleştirmesi bir gözlem üzerinde herhangi bir özelliğin bulunup bulunmadığı ile ilgili olabilir. Örneğin, bir böceğin bir özelliği, kanadının olup olmaması ise kanatlı olmasının 1, olmamasının 0 ile kodlanması gibi .

3.3.2.2. İki Sonuçtan Fazla Değer Alan Nitel Değişkenler İçin Benzerlik Ölçüleri

Tümü iki sonuçtan daha fazla değerler alan nitel değişkenlerden oluşan veride bir gözlem çifti arasındaki yakınlık benzerlik katsayıları (skor değerleri) ile belirlenir (Everitt ve ark. 2001). Her biri p sayıda iki sonuçtan fazla değerler alan kesikli x_i ve x_j gözlem çifti arasındaki benzerlik için

$$S(x_i, x_j) = \frac{1}{p} \sum_{k=1}^p s_{ijk} \quad (3.27.)$$

şeklinde tanımlanan bir skor değeri kullanılır. Burada, s_{ijk} ;

$$s_{ijk} = \begin{cases} 0 & , \text{eğer } x_i \text{ ve } x_j \text{ değişken } k. \text{değişkende eşleşmiyorsa} \\ w & , \text{eğer } x_i \text{ ve } x_j \text{ değişken } k. \text{değişkende eşleşiyorsa} \end{cases} \quad (3.28.)$$

şeklinde tanımlanır.

3.3.3. Nitel ve Nicel Değişkenler İçeren Gözlem Çiftleri İçin Benzerlik Ölçüleri

Bazı değişkenleri sürekli, bazı değişkenleri kategorik olan verideki gözlem çiftleri arasındaki benzerliğin bulunması için çeşitli yaklaşımlar önerilmiştir. Bunlardan biri tüm değişkenlerin ikili değerlere dönüştürülmesi ve iki sonuçlu değişkenler içeren gözlem çiftleri için önerilen bir benzerlik ölçüsünü kullanmaktır. Başka bir yaklaşım ise her değişken tipine uygun uzaklıkların hesaplanması ve

bunların birleştirilmesidir. Daha karmaşık bir yaklaşım, Gower (1971) tarafından önerilmiştir. Gower kategorik değişkenler için genel benzerlik ölçüsünü,

$$s_{ij} = \frac{\sum_{k=1}^p w_{ijk} s_{ijk}}{\sum_{k=1}^p w_{ijk}} \quad (3.29.)$$

biçiminde ifade etmiştir. Burada s_{ijk} k. değişken değerine göre i. ve j. gözlemler arasındaki benzerlik ölçüsüdür. w_{ijk} ise i. ve j. gözlem k. değişkene göre karşılaştırmasında değişken değeri bulunmuyorsa 0 diğer durumlarda 1 değerini almaktadır. Gower (1971), verideki sürekli değişkenler için benzerlik ölçüsünü,

$$s_{ij} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k} \quad (3.30.)$$

biçiminde tanımlamıştır. Burada R_k , i. ve j. gözlemin k. değişken değerlerinin değişim aralığı (range) olarak tanımlanır. Gower (1971) katsayısının kullanımına ilişkin bir örnek aşağıda verilmiştir.

Bahçe çiçeği verisi (Kaufman ve Rousseeuw 1990) 18 bahçe çiçeğinin yedi özelliğine ilişkin değerler içerir. Çiçekler için ilgilenilen bu özellikler;

- 1: Bitki donduğunda bahçede bırakıldı mı? (1=evet, 0=hayır)
2. Bitki gölge altında mı? (1=evet, 0=hayır)
3. Bitki yumrulu mu? (1=evet, 0=hayır)
4. Bitkinin çiçek rengi nedir? (beyaz=1, sarı=2, pembe=3, kırmızı=4, mavi=5)
5. Bitkinin ekildiği toprak türü nedir? (kuru=1, normal=2, nemli=3)
6. Bitkinin boyu kaç cm dir?
7. Bitkiler arası ideal uzaklık kaç cm dir?

Bu özelliklerde sadece bitki türü yumrulu mu sorusunda incelenen iki bahçe çiçeği de yumrulu değil ise 0 diğer tüm durumlarda 1 ile ağırlıklandırılmıştır.

Tablo 3.3. Bahçe çiçeği verisi (Kaufman ve Rousseeuw 1990).

Bitki Türleri	Değişkenler						
	Kategorik					Sürekli	
	1	2	3	4	5	6	7
1. Begonya	0	1	1	4	3	25	15
2. Katır tırnağı	1	0	0	2	1	150	50
3. Kamelya	0	1	0	3	3	150	50
4. Yıldız çiçeği	0	0	1	4	2	125	50
5. Myosotis sylvatica	0	1	0	5	2	20	15
6. Küpe çiçeği	0	1	0	4	3	50	40
7. Sardunya	0	0	0	4	3	40	20
8. Kılıç çiçeği	0	0	1	2	2	100	15
9. Süpürge otu	1	1	0	3	1	25	15
10. Hortensis	1	1	0	5	2	100	60
11. Süsen çiçeği	1	1	1	5	3	45	10
12. Zambak	1	1	1	1	2	90	25
13. Vadi zambağı	1	1	0	1	2	20	10
14. Şakayık	1	1	1	4	2	80	30
15. Karanfil	1	0	0	3	2	40	20
16. Gül	1	0	0	4	2	200	60
17. İskoç gülü	1	0	0	2	2	150	60
18. Lale	0	0	1	2	1	25	10

Tablo 3.4. Bahçe çiçeği verisinin Gower (1971) tarafından önerilen benzerlik ölçüsüne göre ilk 10 bitki türü için elde edilen benzerlik matrisi.

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.09	1.00								
3	0.52	0.33	1.00							
4	0.53	0.41	0.41	1.00						
5	0.57	0.10	0.43	0.39	1.00					
6	0.77	0.21	0.71	0.48	0.56	1.00				
7	0.69	0.30	0.46	0.56	0.46	0.76	1.00			
8	0.51	0.43	0.29	0.74	0.51	0.32	0.51	1.00		
9	0.43	0.43	0.43	0.11	0.50	0.39	0.30	0.23	1.00	
10	0.24	0.42	0.42	0.38	0.61	0.39	0.14	0.30	0.45	1.00

4. ÇOK BOYUTLU ÖLÇEKLEME ANALİZİ

Tarih boyunca hep nesnelere kendi aralarında birbirleriyle kıyaslanmış ve benzerlik değerlendirilmesi yapılmıştır. n birimlik bir veri kümesi düşünüldüğünde a ve b nesnelere arasında bir benzemezlik ölçüsü vardır. Örneğin her biri ayrı bir imalathanede üretilmiş 10 tepsili baklava alınır. Gurmeler tarafından a. ve b. tepsilerin kıyaslanması için 0 ile 10 arasında tamsayı olarak puanlar verilerek benzemezlik ölçütü oluşturulsun. Gurme, öncelikle a. tepside sonra b. tepside bir dilim baklava alsın ve ardından puanlarla baklavaları kıyaslasın. 0 puan verilen baklavalar birbirlerinden ayırt edilemeyecek oranda benzerken, 10 puana doğru baklavaların tatlarının tamamen farklılaştığı görülür. Gurmeye 10 tepsili baklavayı ikişerli olarak karşılaştırması için 45 tane baklava çifti sunulur ve toplam benzemezlik kümesiyle veri analizi sağlanır.

Buradaki baklavaların her biri uzaydaki bir nokta ile temsil edilir ve noktalar arasındaki uzaklıklar, mümkün olduğunca uygun bir şekilde orijinal benzemezlik ölçüleri ile eşlenir. Ama boyut sayısı fazla olacağından daha düşük bir boyutta çalışma hedeflenir ki bu da çok boyutlu ölçekleme ile sağlanır.

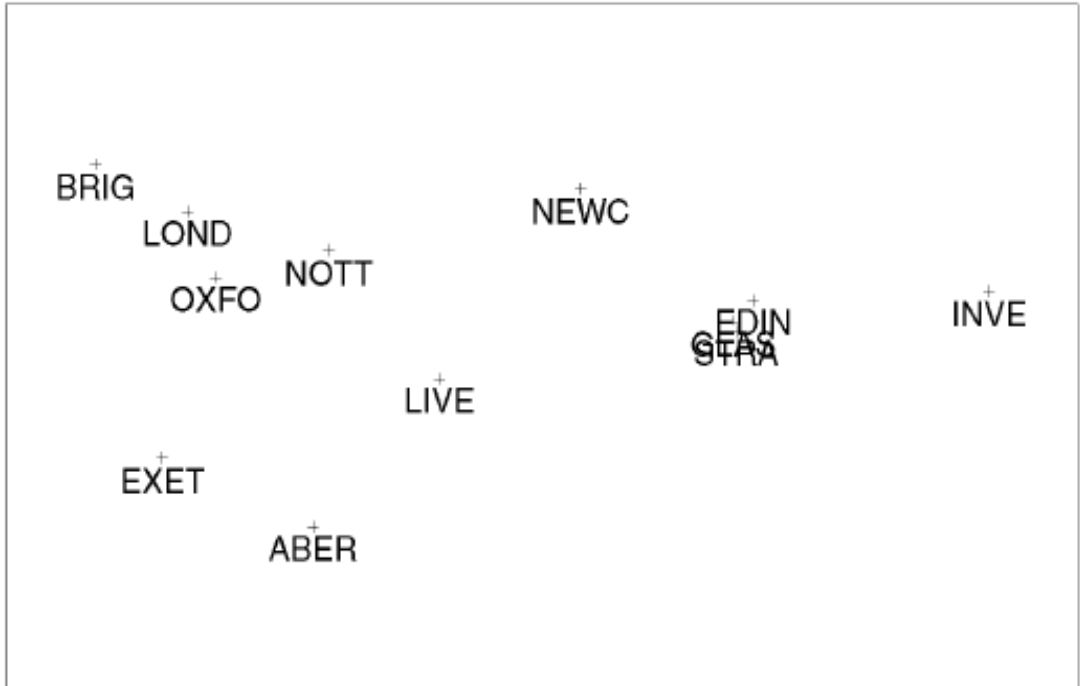
Çok boyutlu ölçekleme analizinde herhangi bir dağılım varsayımına gerek duyulmadan, hesaplanmış uzaklık ölçülerini bir fonksiyon yardımıyla kullanarak gösterim uzaklıkları elde edilmeye çalışılır. Amaç, nesnelere yapısının görsel olarak daha kolay bir biçimde ortaya konulmasıdır. Yani kısaca çok boyutlu ölçekleme analizi, bir boyut indirgeme yöntemidir.

Çok değişkenli istatistiksel yöntemlerden biri olan çok boyutlu ölçekleme analizi bir boyut indirgeme tekniği olarak kullanılmasının yanı sıra, nesnelere ya da bireylerin aralarındaki farklılıkları veya benzerlikleri sıralayan ve hipotez testlerinin kurulmasına ve verinin bağıllık yapısının incelenmesine yardımcı olan bir tekniktir.

Bir coğrafik bölgede ulaşım olanaklarına göre yerleşim yerlerinin (il, ilçe, kasaba) birbirlerine yakınlıklarını, idari yapı bakımından birbirleri ile olan ilişkilerini ve ulaşım, erişim, denetleme açısından uygun bir yapılanma ortaya koymak için yerleşim birimlerinin ele alınan özellikler bakımından benzerlik ya da farklılıklarını

rasyonel biçimde ifade edebilmek için çok boyutlu ölçekleme analizinden yararlanır.

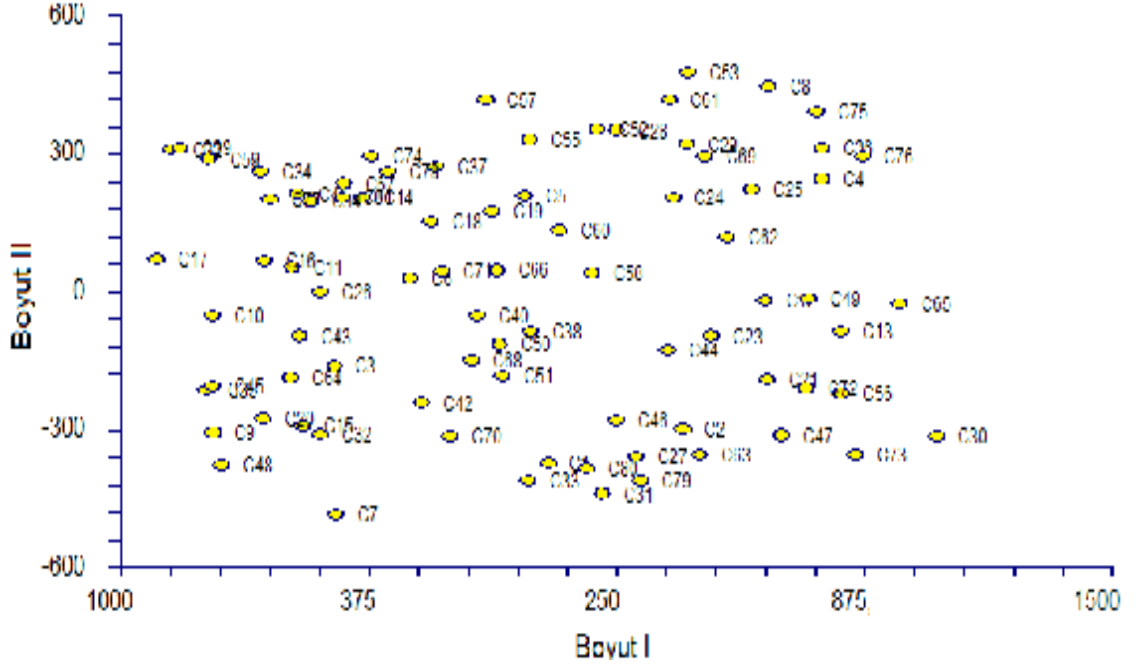
Çok boyutlu ölçeklemenin örneklenmesindeki klasik yollardan bir tanesi de yukarıda bahsedildiği gibi yerleşim yerlerinin birbirlerine yakınlıkları baz alınarak şehir haritasını yapılandırmak için şehirler arasındaki yolculuk sürelerini kullanmaktır. Greenacre ve Underhill (1982), Güney Afrika havalimanları arasındaki uçuş zamanlarını, Mardia (1979) ise İngiltere'deki bazı şehirler arasındaki yol uzaklıklarını kullanmıştır. İngiltere'deki 12 şehir arasındaki yolculuk süreleri, çok boyutlu ölçeklemede kullanılmak için ele alınmıştır. ÇBÖ tekniğiyle elde edilmiş noktaların konfigürasyonu Şekil 4.1'de gösterilmiştir. Atladaki İngiltere haritası ile Şekil 4.1'de elde edilen pozisyonlar çarpıcı bir şekilde benzerdir.



Şekil 4.1. ÇBÖ Yöntemine Göre İngiltere'deki Şehirlerin Haritalandırılması (Cox and Cox, 2001)

Benzer düşünce ile Türkiye'deki 81 ilin uzaklıklarını kullanarak ÇBÖ analizinde Türkiye haritasının yeniden çizim çalışması Şekil 4.2.'de gösterilmiştir. Yine şekilden de anlaşılacağı gibi atlastaki Türkiye haritasıyla neredeyse birebir

uyuşma gözlenmektedir. Hafif farklılığın nedeni olarak, il sınırlarının belli olmaması ve şehir merkezlerinin baz alınması gösterilebilir.



Şekil 4.2. ÇBÖ Yöntemi ile Türkiye Haritası

Şehirlerin ve ülkelerin haritalanmasının yanı sıra birçok alanda kullanılan ÇBÖ analizi çeşitli çok değişkenli veri analiz tekniklerini içerir. ÇBÖ, çok değişkenli verilerdeki nesnelere grafiksel gösterimini sağlayan teknikleri de kapsar. Örneğin baklava kıyaslamalarından elde edilen benzerlik ölçüleri, benzer baklava grupları bulmak için kümeleme analizinde kullanılabilir.

Çok boyutlu ölçekleme yönteminin amaçlarını kısaca şu dört maddeyle belirtebiliriz (Borg, I. ve Gronen P.J, 2005):

- 1- Araştırma amaçlı uygulanan ÇBÖ analizi, teorik olarak biçimsiz olan veriye uygulanır. Veri belirli bir teoriyle uyumlu olarak araştırılmazsa araştırmacı verinin yapısını anlamak için bu yöntemi kullanabilir.
- 2- Yapısal hipotezlerin test edilmesi için kullanılan ÇBÖ ile alan hakkında araştırma yapmaktan çok daha fazla şey bilinmesi durumunda konuların hangi faktörler ile ne kadar bağlantılarının olduğu test edilmeye çalışılır.

- 3- Psikologlar tarafından psikolojik yapının araştırılması amacıyla verinin altında yatan psikolojik boyutların belirlenmesinde kullanılır.
- 4- Benzerlik hükümlerinin bir modeli olarak ÇBÖ analizinden faydalanılabilir.

4.1. Çok Boyutlu Ölçeklemede Varsayımlar

Çok boyutlu ölçekleme yöntemi metodoloji, veri türleri, değişkenler arası ilişkiler vs. açısından herhangi bir varsayım içermemesine rağmen araştırmacılar algılamayla ilgili şu üç prensibi göz önünde bulundurmalıdır (Joseph ve ark, 1998);

- 1- **Boyut İlkesi:** Her cevaplayıcı, her uyarıcıyı aynı boyutta algılayamaz. Örneğin bazıları otomobili beygir gücü ve dış görünüşü açısından değerlendirirken bazıları ise bu faktörler yerine otomobili fiyatı ve iç konforu itibarıyla değerlendirebilirler.
- 2- **Önem İlkesi:** Bütün cevaplayıcılar bir uyarıcıyı aynı boyutta algılasalar bile aynı önem seviyesinde algılamazlar. Örneğin iki cevaplayıcı da bir kolalı içerde asit (karbonatlılık) olmasını boyut olarak kabul edebilirken; cevaplayıcılardan biri için bu özellik çok önemli, diğeri için önemsiz olabilir.
- 3- **Zaman İlkesi:** Bir uyarıcıya ilişkin boyut ve önem verme her zaman sabit kalmaz. Yani farklı zaman aralığında bir uyarıcıya ilişkin farklı olabilir. Örneğin bireylerin dün bir kolalı içeceği asitli olarak algıladıkları aynı zamanda asitlilik özelliği önem arz ederken bugün tam tersi olabilir.

Çok boyutlu ölçekleme yöntemi, veriler ile ilgili dağılım varsayımı gerektirmeyen bir yöntemdir. ÇBÖ, değişkenlerin tipine bağlı olarak hesaplanan nesnelere arasındaki uzaklıkları en az hata ile temsil edecek ÇBÖ gösterim uzaklıklarını herhangi bir regresyon yöntemi (doğrusal, polinomial, monotonic vb.) aracılığı ile belirlemeyi sağlar. ÇBÖ yönteminde veri uzaklık matrisinden elde edilen birim ya da nesnelere arası uzaklıkların daha az boyutlu bir uzayda grafiksel olarak gösterilmesine grafiksel gösterim adı verilir. Grafiksel gösterimi elde etmek için veri koordinatlarının en az hata ile grafiksel gösterim koordinatlarına dönüştürülmesi gerekmektedir. n nesne ya da birim arasındaki uzaklıklar simetrik ve yansımalıdır. n nesne arasında $n(n - 1)/2$ çift uzaklık hesaplanır. Bu orijinal

uzaklıklar mutlak uzaklıklar olarak alınarak işlenirler. Bu uzaklıklara göre, uygun ve daha az boyutlu bir geometrik gösterim elde etmek için orijinal uzaklıklara oldukça yakın bir gösterim koordinat sistemi elde edilmeye çalışılır. Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uygunluğu ölçen ölçüye stress ölçüsü adı verilir. Konfigürasyon uzaklıkları, orijinal değerlerden çeşitli yöntemlere göre hesaplanarak elde edilir.

ÇBÖ, verilerin tipine bağlı olarak metrik ÇBÖ ve metrik olmayan ÇBÖ olarak iki biçimde verilere uygulanır. Yaklaşımların belirlenmesinde veri tipinin önemi büyüktür.

ÇBÖ yöntemi, uzaklık matrislerinden yararlanarak çözüm yapar. Bu nedenle veri tipine uygun uzaklık matrislerini hesaplamak gerekir. ÇBÖ uzaklıklar matrisini farklılık matrisi olarak ele alır. Eğer veriler aralıklı ya da orantılı ölçekli olarak elde edilmiş ise uzaklık değerleri Öklid uzaklığı, Karesel Öklid uzaklığı, Chebychef, City-Block, Minkowski uzaklıkları biçiminde hesaplanabilir. (Özdamar, 2004).

ÇBÖ analizi nesne ya da birimlerin birbirine benzerliği temeline dayandığından, birimlerin birbirlerine benzerliğinin ölçülmesi önemlidir. Bu amaçla veri türüne bağlı olarak değişik benzerlik ölçüleri kullanılabilir. Benzerlik ya da benzemezlik ölçütlerinden hangisinin kullanılacağı konusunda teknik nedenlerden dolayı birçok bilim adamı, ÇBÖ analizinde benzemezliği (farklılığı) kullanmayı tercih etmektedir. Çünkü birimler arasındaki bu uzaklık ilişkisi dolaysız ve pozitifdir. Yani yüksek değerdeki benzemezlik (farklılık) daha kolay algılanmayı sağlamaktadır (Gigure, 2006).

4.2. Çok Boyutlu Ölçekleme Modelleri

Çok boyutlu ölçekleme algoritmalarını detaylı olarak ele almadan önce, ÇBÖ analizi için kullanılan bazı modelleri taslak olarak ele alalım. Tüm nesne çiftleri arasındaki benzemezlik ölçüsü δ_{ab} belirlenmiş olan n nesneli bir küme olduğunu varsayalım. Nesnelerin temsil edildiği n noktanın bir konfigürasyonu, k boyutlu bir uzayda aranır. Her nokta bir nesneyi temsil eder. Noktalar arasındaki uzaklıklarda

d_{ab} ile gösterilir. ÇBÖ analizinin amacı; uzaklıkları mümkün olduğunca uygun bir şekilde benzemezliklerle eşleyen bir konfigürasyon bulmaktır.

- **Klasik Ölçeklendirme:** $d_{ab} = \delta_{ab} \quad a, b = 1, 2, \dots, n$
Bu eşliğin sağlanması ile noktaların konfigürasyonunu bulmak mümkündür.
- **Metrik En Küçük Kareler Ölçeklendirmesi:** Uzaklıkları, dönüştürülmüş benzemezliklerle eşleyen bu yöntemde, f monoton bir fonksiyondur.
- **Tek Boyutlu Ölçeklendirme:** Nesnelere temsil eden noktaların konfigürasyonu sadece tek boyutta aranır. Tek boyut, sadece analizde kullanılacak nesnelere sıralar. Çok boyutlu ölçeklendirmenin özel bir durumudur.
- **Metrik Olmayan Ölçeklendirme:** Benzemezlik dönüşümünün metrik yapısı kaybolmuşsa metrik olmayan çok boyutlu ölçeklemeye geçilir.
$$\delta_{ab} < \delta_{a'b'} \Rightarrow f(\delta_{ab}) \leq f(\delta_{a'b'}) \quad 1 \leq a, b, a', b' \leq n.$$
Böylece benzemezliğin rank sırası dönüşümle korunmuş olur.
- **Procrustes Analizi:** İsmi, misafirlerinin boylarını yatağa sığdırmak için kol ve bacaklarını kırıp kısaltan veya çekip uzatan efsanevi bir devden alan bu analiz, yapılan iki konfigürasyonu kıyaslamak ve birini, genişleterek, döndürerek, büyütürük diğerine eşlemeye çalışır.
- **Biplots:** Hem nesnelere temsil eden noktaların bir konfigürasyonunu hem de grafikteki eksenleri temsil eder.

4.3. Çok Boyutlu Ölçekleme Teknikleri

Verilerin tiplerine bağlı olarak çok boyutlu ölçekleme analizi, genel olarak metrik ÇBÖ ve metrik olmayan ÇBÖ olarak iki gruba ayrılır. Bazı araştırmacı ve kaynaklara göre bu iki gruba yarı metrik ÇBÖ adında üçüncü bir grupta eklenmektedir.

Nicel ve metrik uzaklıklara dayalı verilere metrik ÇBÖ tekniği, skor, sıralı ve kategorik verilere ise metrik olmayan ÇBÖ tekniği uygulanır. Metrik olmayan ÇBÖ, metrik olan ÇBÖ'ye göre daha az varsayım gerektirmekte ve çözümlenmelerde en çok

tercih edilen yöntem olmaktadır. Metrik ÇBÖ'de gösterim ve veri uzaklıkları doğrusal ya da polinomial fonksiyon yöntemi ile hesaplanmaktadır. Metrik olmayan ÇBÖ'de varsayımların daha esnek uygulanması nedeniyle metrik ÇBÖ'ye göre daha küçük boyutta çözümler elde edilmesi mümkün olabilmektedir (Özdamar, 2004).

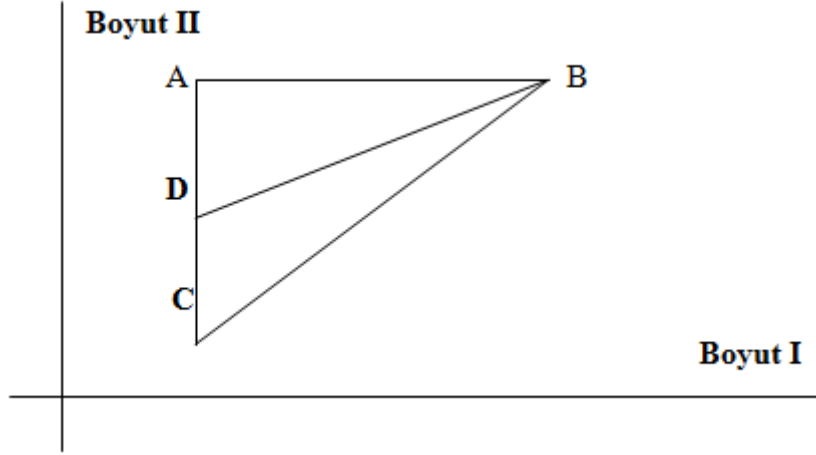
Metrik ve metrik olmayan yöntemlerin kullanımı ile elde edilen şekillerin; yorum, yansıma ve dönüşüme bağlı olarak bazı belirsizlikleri bulunmaktadır. Şekillerde bulunan tüm noktalar bir yerden başka bir yere kaydırılabildiği gibi şeklin tümü döndürülebilmekte veya yansıtılabilmektedir. Bu bilgilere bağlı olarak, sadece uzaklıkların büyüklük sırasına göre sıralanmalarını kullanarak çözüme giden metrik olmayan ölçekleme yöntemi ile temel bileşenlere benzer bir yaklaşım kullanan metrik ölçekleme yöntemine ayrıntılı bir şekilde değinmeden önce, uzaklık değerlerinin kullanımını basit bir örnekle vermek ÇBÖ'nin daha iyi anlaşılmasını sağlayacaktır (Tatlıdil, 2002).

A, B, C ve D gibi gözlemlerin aşağıda verilen uzaklık değerlerini kullanarak bu gözlemlerin çok boyutlu uzayda oluşturacağı şekli bulalım.

Tablo 4.1. Gözlemler Arası Uzaklık Değerleri

	A	B	C	D
A	0	6	6	2.5
B	6	0	9.5	7.8
C	6	9.5	0	3.5
D	2.5	7.8	3.5	0

Gözlemlerin temsil edilmesinde kaç boyutun olması gerektiği genelde bilinmez, ancak bir ya da iki boyutlu olması çizim yapılabilmesi açısından kolaylık sağlar. Bu nedenle yukarıdaki uzaklık değerleri kullanılarak A, B, C ve D gözlemleri arasındaki ilişki aşağıdaki şekille ifade edilebilir.



Şekil 4.3. Gözlemler Arası İlişkinin Şekilsel İfadesi (Tatlıldil, 2002)

Verilere ilişkin ilk değerlerin kullanımı ile gözlemler arasındaki ilişki hakkında bir şeyler söylemek çoğu kez mümkün olmamakta, ancak uzaklık değerleri kullanılarak aralarındaki ilişki yapısı şekille ortaya konabilmektedir. Nitekim yukarıdaki şekilden de görülebileceği gibi çizilen şekilde gözlemlerin yer aldığı doğrular, birbirlerine göre konumlarını, aralarındaki ilişkiyi ortaya koyma açısından önemlidir (Tatlıldil, 2002).

4.3.1. Metrik Çok Boyutlu Ölçekleme Tekniği

Metrik ÇBÖ analizinde girdi ve çıktı verileri metriktir. Verilerin metrik olması, aralıklı veya oranlı ölçekle ölçülebilmesi anlamına gelmektedir (Malhotra, 2004). Metrik ÇBÖ tekniği, uzaklık matrisindeki uzaklıkların oranlı ölçeklenmiş (veya en azından eşit aralıklı ölçeklenmiş) olduğunu varsayar (Fındıkkaya, 1995).

Metrik ölçekleme aslında Temel Koordinat Analizi ve Temel Bileşenler Analizi (Principal Component Analysis, PCA) denilen çok değişkenli istatistik yöntemleri ile aynıdır. Bu yöntemlerden tek farkı metrik ölçeklemede negatif özdeğerlerin var olmasına karşın Temel Bileşenlerde bütün özdeğerler pozitif olmaktadır (Tatlıldil, 2002).

Metrik ÇBÖ tekniği, oranlı ölçeklenmiş uzaklıkları, az sayıda boyut kullanarak grafiksel bir düzenleme elde etmek için kullanılır. Böyle bir grafiksel

düzenlemeyle elde edilen algılama haritalarındaki noktalar arasındaki uzaklık da orantılı olacaktır. Dolayısıyla, metrik ÇBÖ kullanılarak uyarıcılar arasındaki benzerlik ve tercih verilerine göre, uyarıcıların az boyutlu alanda grafiksel düzenlenmesi gerçekleştirilebilir. Böylece, uyarıcılar hakkında araştırma konusuna yönelik bazı bilgiler elde edilebilir (Akacan, 2001).

Metrik ÇBÖ tekniği $n \times (n - 1)/2$ tane uzaklık değerlerinin bulunduğu matrisi girdi olarak kabul edip, uygunluğuna bağlı olarak genelde iki ya da üç boyutlu bir harita çıktısı verir. Bu yöntemle, çıktı ve girdi matrislerinin arasındaki farklılığın en az olması hedeflenir.

Metrik ÇBÖ analizinde temel yaklaşım, $n \times p$ boyutlu veri matrisinden hesaplanan birimler arasındaki δ_{ij} uzaklıklarını daha düşük bir boyutta temsil etmek üzere

$$d_{ij} = f(\delta_{ij}) \quad (4.6.)$$

dönüşümünü gerçekleştirmektir. Eşitlikte yer alan f sürekli monoton bir fonksiyondur. Metrik ÇBÖ analizinde $f(.)$ fonksiyonunun farklı şekillerde tanımlanması ile farklı ÇBÖ modelleri oluşturulmuştur. Örneğin, Borg ve Groenen (1997) $f(.)$ fonksiyonunu

$$f(\delta_{ij}) = b\delta_{ij} \quad (4.7.)$$

tanımlayarak oransal ÇBÖ adını verdikleri modeli önermişlerdir. Ayrıca a ve b serbest parametreler olmak üzere aralıklı ÇBÖ modelinde $f(.)$ fonksiyonu

$$f(\delta_{ij}) = a + b\delta_{ij} \quad (4.8.)$$

eşitliğiyle tanımlanır. Bu modeller dışında $f(.)$ fonksiyonu, yüksek dereceli polinom fonksiyonları, logaritmik fonksiyonlar veya üstel fonksiyonlar olarak tanımlanabilir. Metrik ÇBÖ analizindeki metotların çoğu (4.6) nolu eşitlikteki dönüşümü sağlamak

için orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uyumun bir göstergesi olan

$$stress = (\text{ölçek faktörü})^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (d_{ij} - f(\delta_{ij}))^2 \quad (4.9.)$$

stress değerini en küçükmeye çalışan optimizasyon teknikleridir. Eşitlikte yer alan ölçek faktörü farklı biçimlerde tanımlanabilmekle birlikte çoğunlukla Kruskal(1964a) tarafından önerilen

$$\text{ölçek faktörü} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^2 \quad (4.10.)$$

ölçek faktörü kullanılır (Martinez ve Martinez, 2005).

4.3.1.1. Klasik Metrik Çok Boyutlu Ölçekleme

Klasik metrik ÇBÖ analizi aynı zamanda temel koordinat analizi olarak da bilinmektedir. Algoritmadaki temel amaç birimler arasında p boyutta elde edilen δ_{ij} uzaklıklarına yaklaşık eşit olacak şekilde k boyutta d_{ij} konfigürasyon uzaklıklarını elde etmektir. Uygulamalarda genellikle $k = 2$ alınır ve birimlerin iki boyutlu düzlemde saçılımı grafiksel olarak ortaya konur. Klasik metrik ÇBÖ algoritması (Rencher, 2002);

- i. $n \times n$ boyutlu **A** matrisi oluşturulur.

$$A = (a_{ij}) = \left(-\frac{1}{2} \delta_{ij}^2\right) \quad (4.11.)$$

- ii. $n \times n$ boyutlu **B** matrisi

$$\mathbf{B} = (b_{ij}) = (a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}) \quad (4.12.)$$

oluşturulur.

Eşitlikte yer alan $\bar{a}_{.i} = \sum_{j=1}^n a_{ij} / n$, $\bar{a}_{.j} = \sum_{i=1}^n a_{ij} / n$ ve $\bar{a}_{..} = \sum_i \sum_j a_{ij} / n^2$ şeklinde tanımlanır. Ayrıca **B** matrisi

$$\mathbf{B} = \left(\mathbf{I} - \frac{1}{n} \mathbf{J} \right) \mathbf{A} \left(\mathbf{I} - \frac{1}{n} \mathbf{J} \right) \quad (4.13.)$$

şeklinde tanımlanabilir. Eşitlikte yer alan **J** $n \times n$ boyutlu birlerden oluşan matrisi ve **I** $n \times n$ boyutlu birim matrisi göstermektedir.

iii. **B** matrisi simetrik olduğu için spektral dönüşüm ile

$$\mathbf{B} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}' \quad (4.14.)$$

şeklinde yazılabilir. Eşitlikte yer alan **V**, **B** matrisinin özvektörlerinin matrisi, **Λ**, **B** matrisinin özdeğerlerden oluşan köşegen matrisidir.

iv. k boyutlu uzaydaki koordinat değerleri

$$\mathbf{Y} = \mathbf{V}_k \mathbf{\Lambda}_k^{1/2} \quad (4.15.)$$

eşitliği ile elde edilir. \mathbf{V}_k , **B** matrisinin en büyük k özdeğerine karşılık gelen özvektörlerden oluşan matrisi, $\mathbf{\Lambda}_k^{1/2}$ ise **B** matrisinin en büyük k özdeğerinin karekökünden oluşan köşegen matrisi göstermektedir.

4.3.1.2 Karmaşık Fonksiyonların Optimizasyonu İle Ölçekleme

De Leeuw (1977) tarafından önerilen karmaşık fonksiyonların optimizasyonu ile ölçekleme (SMACOF) algoritması hem basit hem de metrik ve metrik olmayan uygulamalarda kullanılabilir. Bu çalışmada, Borg ve Groenen (1997) tarafından metrik durum için oluşturulan algoritma incelenecektir. SMACOF algoritmasında karmaşık fonksiyonlar kullanılarak stress değeri en küçüklenmeye

çalışılmaktadır. Konfigürasyon uzaklıklarının hesaplanıldığı k boyutlu düzlemdeki başlangıç koordinat değerleri \mathbf{X}^0 olmak üzere sıra stress değeri,

$$\sigma(\mathbf{X}^0) = \sum_{i < j} w_{ij} (d_{ij}(\mathbf{X}^0) - \delta_{ij})^2 \quad (4.16.)$$

şeklinde tanımlanır. Eşitlikte yer alan w_{ij} ağırlığı i . ve j . birimler arasındaki uzaklığın bilindiği veya hesaplanabildiği durumlarda 1, diğer durumlarda 0 değerini alır. SMACOF algoritmasının başlangıç adımında, (4.16.) nolu eşitlikte yer alan \mathbf{X}^0 rasgele alınabileceği gibi belirli bir sistematığe göre de seçilebilir (Martinez ve Martinez, 2005). Tüm birimler arasındaki orijinal uzaklıkların bilindiği veya hesaplanabildiği varsayımı altında, algoritmanın r . tekrarında $\mathbf{X}^{(r)}$,

$$\mathbf{X}^{(r)} = n^{-1} \mathbf{B}(\mathbf{X}^{(r-1)}) \mathbf{X}^{(r-1)} \quad (4.17.)$$

eşitliği ile elde edilir. Eşitlikte yer alan $\mathbf{B}(\mathbf{X}^{(r-1)})$ matrisinin değerleri $i \neq j$ durumunda

$$b_{ij} = \begin{cases} -\frac{\delta_{ij}}{d_{ij}(\mathbf{X}^{(r-1)})}, & d_{ij}(\mathbf{X}^{(r-1)}) \neq 0 \\ \mathbf{0}, & d_{ij}(\mathbf{X}^{(r-1)}) = 0 \end{cases} \quad (4.18.)$$

ve $i = j$ durumunda

$$b_{ii} = - \sum_{j=1, j \neq i}^n b_{ij} \quad (4.19.)$$

şeklinde hesaplanır. Algoritmanın ardışık iki adımındaki sıra stress değerleri arasındaki mutlak fark istenilen en küçük fark kriterine (EFK) eşit veya daha küçük oluncaya kadar algoritma işletilir. Yani;

$$|s(\mathbf{X}^r) - s(\mathbf{X}^{r-1})| \leq \text{EFK} \quad (4.20.)$$

oluncaya kadar işlem yapılır.

4.3.2 Metrik Olmayan (Non-Metrik) Çok Boyutlu Ölçekleme Tekniği

Metrik olmayan ifadesi analize giren verilerin ordinal ilişkileri içermesine karşın analizden çıkan verilerin metrik olması veya en azından aralıklı ölçekle ölçülmüş veri olması anlamına gelmektedir (Kurtuluş, 2007).

Metrik olmayan ölçektelemedeki problemlerin çözümü için ilk algoritma Shepard (1962a, 1962b) tarafından tanımlanmıştır. Fakat Shepard (1962a,1962b), tanımladığı algoritmada kayıp fonksiyonunu kullanmamıştır. Kruskal (1964a,1964b), Shepard (1962a,1962b) tarafından tanımlanan algoritma için stress adı verdiği kayıp fonksiyonunu kullanarak, bu fonksiyonu minimize etmeye çalışmıştır.

Metrik olmayan ölçekleme, uzaklıkların nümerik değerleri yerine büyüklük sıralarının kullanıldığı durumlarda da uygulanabilmekte ve konfigürasyon uzaklıkları d_{ij} değerlerini belirlemede kullanılan tek bilgi δ_{ij} uzaklık değerlerinin sıra sayıları olmaktadır. Bu yaklaşımda kullanılan genel algoritmada analitik bir çözüm mümkün olmadığından dolayı, stress değeri iteratif bir yaklaşımla minimize edilmeye çalışılır.

Metrik olmayan ÇBÖ'nin temeli sayılabilecek bu algoritmanın adımları aşağıdaki gibidir:

İlk adımda; D uzaklık matrisinin (köşegendeki sıfır elemanları hariç) tüm elemanları,

$$d_{i_1j_1} < d_{i_2j_2} < \dots < d_{i_mj_m}; \quad m = \frac{n(n-1)}{2} \quad (4.21.)$$

şeklinde sıralanır. Daha sonra d_{ij} 'lerle monotonik olarak ilişkilendirilen d_{ij}^* değerleri tanımlanır. Bu ilişkilendirme

$$d_{ij} < d_{uv} \rightarrow d_{ij}^* \leq d_{uv}^* \quad (\forall i < j, u < v \text{ için}) \quad (4.22.)$$

gibi bir koşulu sağlamalıdır.

İkinci adımda; çok boyutlu uzaydaki (p -boyutlu) gerçek şekille, indirgenmiş boyutlu (k boyutlu) uzayda kestirilen şekil arasındaki farklılığın ifadesi olan stress değeri hesaplanır. $\hat{X}_{n \times k}$, R^k uzayında \hat{d}_{ij} uzaklık değeri ile ifade edilen bir şekilse, \hat{X} 'nin stress değeri,

$$S(\hat{X}) = \text{en küçük} \left(\frac{\sum_{i < k} (d_{ik}^* - \hat{d}_{ik})^2}{\sum_{i < k} \hat{d}_{ik}^2} \right)^{\frac{1}{2}} \quad (4.23.)$$

eşitliği ile ölçülür.

Üçüncü adımda; her k boyut için en küçük stress değerine sahip olan şekle, k boyutta en iyi şekil adı verilir. En küçük stress değeri k -nci boyut

$$S_k = \min S(\hat{X}) \quad (4.24.)$$

eşitliği ile ifade edilir. Burada S_k , k 'nin azalan bir fonksiyonudur.

Son adımda ise; uygun boyut sayısını belirlemek amacıyla S_1, S_2, \dots, S_k değerleri hesaplanmakta ve bu işlemlere en küçük stress değeri elde edilince son verilmektedir (Sığırlı ve ark., 2006).

4.3.2.1. Kruskal-Shepard Yaklaşımı

Kruskal-Shepard algoritması, metrik olmayan ÇBÖ ile ölçeklenmek istenen veri nesnelere $d_{ij} \approx \hat{d}_{ij}$ gibi monoton bir ilişkiyi korumasını amaçlar. Benzemezlik değerleri artan bir sırada dizildiğinde, buna karşılık gelen uzaklık değerlerinin de aynı monoton artan ilişkiyi koruyarak ölçeklemenin yapılması amaçlanır (Alan, 2008). Bu amaçla Kruskal-Shepard algoritmasının işlem sırası aşağıdaki şekilde gösterilebilir:

- 1- İlk adım ölçeklenmek istenen nesne sayısı n ve elde edilecek görünümün boyutu k olmak üzere $n \times k$ boyutlu bir başlangıç konfigürasyonu belirlenir. Başlangıç konfigürasyonu rasgele olabileceği gibi metrik yöntem ile de elde edilebilir.
- 2- İkinci adım bu başlangıç konfigürasyonlarından d_{ij}^0 uzaklık değerlerinin hesaplanarak monoton regresyon ile \hat{d}_{ij}^0 (disparity) denilen değerlerin bulunmasıdır.
- 3- Üçüncü adımda \hat{d}_{ij}^0 değerleriyle yeni konfigürasyonların bulunması sağlanır. Yeni konfigürasyonlar Newton-Raphson veya Dik iniş (Steepest Descent) denilen optimizasyon yöntemleriyle elde edilir.
- 4- Eğer belirlenen stress ölçütü veya iterasyon sayısı sağlanırsa işlem biter.
- 5- Yoksa işlem yeni bulunan konfigürasyonlar ile tekrar d_{ij}^1 hesaplanır ve adım 2 ye geri dönlür.

3 nolu adımda yeni konfigürasyonlar aşağıdaki formül ile bulunabilir (Hardle, 2007):

$$x_{il}^{yeni} = x_{il} + \frac{\alpha}{n-1} \sum_{j=1, j \neq i}^n \left(1 - \frac{\hat{d}_{ij}}{d_{ij}} \right) (x_{jl} - x_{il}) \quad (4.25.)$$

Burada x_{il} , koordinat vektörlerinin bir elemanıdır. Yeni koordinat değerlerinin belirlenebilmesi için öncelikle α katsayısının belirlenmesi gerekir. Bu katsayı yeni noktanın ne kadar değişeceğini belirler. Bu yüzden çok önemlidir. Kruskal bu katsayının başlangıç değerini 0,2 olarak belirtmiştir (Hardle, 2007).

Kruskal'ın belirttiği α katsayısı,

$$\alpha^{şimdiki} = \alpha^{önceki} \times (AçıF) \times (GvF) \times (ISF) \quad (4.26.)$$

eşitliği ile bulunur (Cox and Cox, 2001). Eşitlikte yer alan AçıF, açılı faktörü olarak isimlendirilir ve θ şimdiki ve bir önceki gradyanlar arasındaki açı olmak üzere,

$$AçıF = (4.0)^{\cos^3\theta} \quad (4.27.)$$

şeklinde bulunur. Eşitlik (4.26)'da yer alan GvF , gevşeklik faktörü olarak isimlendirilir ve $k = \min\left[1, \left(\frac{\text{şimdiki stress}}{5 \text{ adım önceki stress}}\right)\right]$ olmak üzere,

$$GvF = \frac{1.3}{1+k^5} \quad (4.28.)$$

biçiminde elde edilir. Eşitlik (4.26)'da yer alan ISF ise iyi şans faktörü olarak isimlendirilir ve

$$ISF = \min\left[1, \left(\frac{\text{şimdiki stress}}{\text{önceki stress}}\right)\right] \quad (4.29.)$$

eşitliği ile bulunur.

4.4. Stress Fonksiyonları

ÇBÖ yöntemi kendi içinde stress fonksiyonları denilen hesap kriterleriyle sonuçlarını değerlendirmektedir. Burada esas olan giriş matrisindeki benzemezlik değerleri ile bulunan konfigürasyon uzaklığı arasındaki pozitif farkların minimum olmasıdır. Bunun için türetilen fonksiyonların değerleri minimize edilerek daha iyi sonuçlar elde edilir. Kruskall'ın bu amaçla tanımladığı Stress-1 ve Stress-2 ölçüm formülleri

$$stress - 1 = \left(\frac{\sum_{i<j} (d_{ij} - \bar{d}_{ij})^2}{\sum_{i<j} d_{ij}^2}\right)^{\frac{1}{2}} \quad (4.30.)$$

$$stress - 2 = \left(\frac{\sum_{i<j} (d_{ij} - \bar{d})^2}{\sum_{i<j} (d_{ij} - \bar{d})^2}\right)^{\frac{1}{2}} \quad (4.31.)$$

şeklindedir (Hardle and Simar, 2007).

Formüllerde d_{ij} bulunan konfigürasyonların birbirlerine uzaklıkları, \bar{d} ise bu uzaklık değerlerinin ortalamasını göstermektedir. Metrik olmayan yöntemde kullanılan stress formüllerinde \hat{d}_{ij} değerleri yer almaktadır. Bunun nedeni disparite değerlerinin benzemezlik değerlerinin monoton sıra ilişkisini yansıtmasıdır (Alan, 2008).

Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uyumsuzluğun ölçümünde kullanılan diğer bir ölçütte Guttman'ın uzaklaşma katsayısıdır. Guttman'ın uzaklaşma katsayısı K ,

$$K = (1 - \mu^2)^{\frac{1}{2}} \quad (4.32.)$$

eşitliği ile elde edilir. Eşitlikte yer alan μ ,

$$\mu = \frac{\sum_{i<j} d_{ij}^* d_{ij}(X)}{[\sum_{i<j} (d_{ij}^*)^2 \sum_{i<j} d_{ij}^2(X)]^{\frac{1}{2}}} \quad (4.33.)$$

formülü ile hesaplanır (Cox and Cox, 2001).

Başka bir stress ölçüsü ise S-Stress değeridir (Borg and Groenen, 1997). Bu stress değeri ALSCAL programı ile,

$$\sigma_{AL}(X) = \sum_{i<j} (d_{ij}^2(X) - \delta_{ij}^2)^2 \quad (4.34.)$$

minimize edilmeye çalışılır.

Uygulamada birçok faktör stress değerini etkilemektedir. Genellikle örneklem hacmi n değerinin büyük olması, stress değerini arttıran bir faktördür. Buna karşın ÇBÖ analizinde tercih edilen boyut sayısı k değerinin artması, stress değerini azaltan bir faktördür (Borg and Groenen, 1997).

Stress ölçüsünün yorumlanmasında Kruskal-Shepard tarafından geliştirilen tolerans oranlarından yararlanılmaktadır. Bu oranlar;

$stress \geq 0.20$	Kötü uyum
$0.10 \leq stress < 0.20$	Orta uyum
$0.05 \leq stress < 0.10$	İyi uyum
$0.025 \leq stress < 0.05$	Mükemmel uyum
$0 < stress < 0.025$	Tam uyum

şeklinde. Stress değerinin minimizasyonunda kullanılacak çok sayıda yöntem bulunmaktadır. Bu yöntemlerden bazıları,

- İteratif Majorizasyon (IM) Algoritması
- Hat Arama Algoritması (Line Search Method)
- Newton-Raphson Yöntemi
- En Dik İniş Yöntemi (Steepest Descent)
- Tavlama Benzetimi (Simulated Annealing)
- Genetik Algoritmalar
- Tabu Araması (Tabu Search)

olarak verilebilir (Wojciech, 2001).

Stress değerlerinin minimumlarını bulurken karşılaşılan sorun fonksiyonların yerel minimumlarının bulunmasıdır. Ancak, mutlak minimumlar karmaşık algoritmalar vasıtasıyla bulunabilmektedir. Yapay zekâ alanında, matematiksel yollarla fonksiyonların amaçlarına ulaşamadığı durumlar için, Tabu Araması, Genetik Algoritma ve Tavlama Benzetimi modelleri geliştirilmiştir.

4.5. Boyut Sayısının Belirlenmesi

Boyut sayısını belirlerken, göz önünde bulundurulacak en önemli faktör grafiğin kolay yorumlanabilir olmasını sağlamaktır. Zaten, analiz edilecek veri setinin boyutunu mümkün olduğunca indirgemek çok boyutlu ölçeklemenin temel amaçlarından biridir. Bu sayıyı belirlerken de stress değerinden yararlanılır. Stress değeri ne kadar küçükse, uygunluk da o kadar iyidir.

ÇBÖ'deki çalışmaların çok büyük bir bölümü 2 boyutlu ölçeklemeyi kullanmaktayken, teknolojik gelişmelere bağlı olarak grafiklerin daha rahat elde

edilmesi nedeniyle az da olsa 3 boyutlu sonuçlar da yayınlanmaktadır. Yöntemden alınacak sonuçların görsel olması gerektiğinden 1-2-3 boyutların kullanılması gerekir (Alan, 2008).

Boyut sayısının kaç olacağına karar verebilmek amacıyla bazı durumlarda $n - 1 \geq 4r$, $n(n - 1) \geq 2nr$ biçiminde hazırlanmış iki formülden biri, n uyarıcı veya nesne sayısı, r ise boyut sayısı olmak üzere belirlenebilmektedir (Kurt, 1992).

Başka bir yaklaşımda özdeğerlerin büyükten küçüğe doğru sıralanarak en uygun r. değer boyut olarak belirlenmelidir. Bu amaçla temel bileşenler analizinde kullanılan,

$$G = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^n |\lambda_j|} \geq \frac{2}{3} \quad (4.35.)$$

ölçütünden yararlanılabilmektedir (Tatlidil, 2002).

Boyut sayısının belirlenmesinde aşağıdaki öneriler göz önünde bulundurulabilir (Ustaahmetoğlu, 2005);

- 1) **Önceki bilgiler:** Teorik ya da önceki araştırmaların sonuçları boyutların adlarını ve sayısını belirlemede ışık tutucu olabilir.
- 2) **Uzaysal haritaların yorumlanabilirliği:** Üçten fazla boyutu olan görünümün ya da uzaysal bir haritanın yorumu zordur. En ideal boyut ikili olanıdır.
- 3) **Dirsek kriteri:** Stress değerleri karşısında boyut sayısını bir grafik üzerinde görerek boyut sayısına karar vermek mümkündür.

4.6. Boyutların Yorumlanması

Çok boyutlu ölçeklemede grafiğin yorumlanabilmesi için eksenlerin anlamlarının bilinmesi gerekmektedir.

Araştırmacı elde ettiği algılama haritası ışığında iki farklı yöntem kullanarak oluşan boyutları adlandırır. Bunlar;

a-Sübjektif Yöntem: Bu yönteme göre araştırmacı kendi yorumlarından veya cevaplayıcıların yorumlarından yararlanır. Bazı durumlarda eldeki veriler boyutları adlandırmayı kolaylaştırır. Mümkün olursa her cevaplayıcının yaptığı tercihlerin ya da tespit ettiği benzerliklerin uzaysal haritası ona gösterilir ve cevaplayıcıdan boyutlara isim vermesi istenebilir. Cevaplayıcıların nesnelere arası benzerlikleri ya da tercihleri hangi kritere göre yaptıklarını sorarak, oradan elde edilen bilgiler, uzaysal haritanın elde edilmesinde kullanılabilir (Nakip, 2003). Ya da konunun uzmanlarından boyutların değerlendirilmesi ve yorumlanması istenebilir. Eğer boyutların özelliklerine ilişkin nicel bir bağlantı olmazsa; yani cevaplayıcılara göre boyutlar öyle gibi görünüyorsa, sübjektif yöntem en iyi sonucu verir.

Benzer anlamda; araştırmacı boyutları bilinen tipik özelliklerine göre tanımlayabilir. Bu şekilde güvenilirlik doğrudan bilgisel ve duygusal boyutlar arasından yapılır. Böylece ortaya çıkan ilişki cevaplayıcıların geri dönüşlerinden olmayıp tamamen araştırmacının kararına bağlıdır

b-Objektif Yöntem: Sübjektif yöntemi tamamlayıcı ve daha biçimlendirilmiş bir metot olarak objektif yöntem geliştirilmiştir. En geniş kullanılan yöntem PROFİT (PROperty FİTting) her nesnenin özelliklerinin oranını toplar böylece her özelliğin en güvenilir algısal haritası elde edilir. Benzerlik cevapları içinde belirleyici özelliği tanımlamaya çalışmak ayrı yapılır. Tek boyuttaki her özelliğin uygunluğu kadar her özelliğe bir uygunluk ölçüsü verilir. Araştırmacı hangi özelliğin algısal konumu en iyi şekilde tanımladığını ve boyutları şekillendirdiğini belirleyebilir. Özellikler ve tanımlanmış boyutlar arasındaki uygunluk gereksinimi metrik çıktı kullanılarak küçültülebilir. Ayrıca boyutlar özgürce döndürüldüğünde yorumlarda da herhangi bir değişiklik olmaz (Joseph, 1998).

Boyutların yorumlanmasını kolaylaştırmak için koordinatların nesnelere atanmasında bir takım işlemler yöntemin uyum iyiliğini değiştirmeyeceği için kullanılmaktadır. Bu yöntemlerden bazıları,

- Belli bir eksene bir sabit atanması veya çıkarılması
- Herhangi bir boyuttaki eksenin ters çevrilmesi
- Eksenlerin bulunduğu düzlemin belli bir açıyla döndürülmesi
- Bütün koordinat değerlerini tutarlı bir faktör ile ölçekleme

şeklindedir (Everett, 2001)

4.7. Çok Boyutlu Ölçekleme Algoritmaları İçin Örnek Uygulamalar

Bu kısımda ele alınan ÇBÖ algoritmalarından Klasik ve SMACOF algoritmalarının pratikte nasıl uygulandıklarını göstermek için iki uygulamaya yer verilecektir. Ele alınacak veri setlerinden ilki, Türkiye'deki Birinci Düzey İstatistiki Bölgelerini oluşturan 12 bölgenin karşılaştırılması için seçilen 10 sosyo-ekonomik göstereye göre oluşturulan veri setidir. Birinci Düzey İstatistiki Bölgelerinin karşılaştırılması için seçilen 10 sosyo-ekonomik gösterge ve göstergelerin gösteriminde kullanılan simgeler Tablo 4.2.'de verilmiştir.

Tablo 4.2. Birinci Düzey İstatistiki Bölgelerin Karşılaştırılmasında Kullanılan Sosyo-Ekonomik Değişkenler

Simge	Yıl	Değişken	Birim
X_1	2000	Şehirleşme Oranı	Yüzde
X_2	2000	Tarım İşkolunda Çalışanların Toplam İstihdama Oranı	Yüzde
X_3	2000	Sanayi İşkolunda Çalışanların Toplam İstihdama Oranı	Yüzde
X_4	2000	Okur-Yazar Nüfus Oranı	Yüzde
X_5	2000	Üniversite Bitirenlerin 22+ Yaş Nüfusa Oranı	Yüzde
X_6	2000	Bebek Ölüm Oranı	Binde
X_7	2000	On bin Kişiye Düşen Hekim Sayısı	Kişi
X_8	2000	Fert Başına Gayri Safi Yurt İçi Hasıla	Milyon TL
X_9	2000	Fert Başına Banka Mevduatı	Milyon TL
X_{10}	2000	Yeşil Karta Sahip Nüfus Oranı	Yüzde

İstatistiki Bölge Birimlerinin Düzey 1 itibariyle araştırmada kullanılan sosyo-ekonomik göstergeleri Tablo 4.3.'de verilmiştir. Tablo 4.3.'de verilen veri seti, Dinçer, Özaslan ve Kvasoğlu (2003)'den derlenmiştir.

Tablo 4.3. İstatistiki Bölge Birimlerinin Düzey 1 İtibariyle Sosyo-Ekonomik göstergeleri

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
IST	90.69	8.13	32.15	93.39	12	39	20.58	2749.7	2846.3	4.87
BA	77.22	34.79	11.76	92.07	13	39.67	22.77	2156.9	2147.1	9.17
DM	67.36	40.95	20.64	91.63	7	42	11.22	2668.4	690.26	8.39
Ege	61.48	50.48	13.84	89.78	8	40.13	14.81	2129.7	708.28	10.84
BM	55.55	50.63	13.74	90.12	7	37.2	10.62	2090.7	512.58	10.09
Akd	59.78	54.97	8.78	88.16	8	37.13	10.06	1725.8	475.1	15.86
BKrd	49.39	65.67	7.92	85.51	6	45.4	9.21	1439.8	320.84	25.4
OA	56.47	64.14	7.97	87.26	6	43.38	9.58	1306.2	313.32	22.42
DKrd	49.37	68.51	5.26	85.79	6	35.5	7.97	1160.5	288.21	23.1
GDA	62.69	61.35	7.06	73.22	5	48.33	5.49	953.95	115.42	22.54
ODA	53.86	66	3.6	76.05	6	49.13	7.5	881.22	137.89	23.41
KDA	51.44	67.35	2.76	80.5	6	56.86	7.58	781	127.99	26.17

Seçilen sosyo-ekonomik göstergelerden tarım işkolunda çalışanların toplam istihdama oranı, bebek ölüm oranı ve yeşil karta sahip nüfus oranı değişkenleri ile gelişmişlik ters orantılı iken ele alınan diğer değişkenler ile doğru orantılı olarak tanımlanmaktadır. Örneğin gelişmişlik açısından tarım işkolunda çalışanların toplam istihdama oranının düşük, sanayi işkolunda çalışanların toplam istihdama oranının yüksek olması istenir.

Örnek diğer veri seti ise, Türkiye ve 36 Avrupa ülkesinin karşılaştırılması için seçilen 12 sosyo-ekonomik göstergeye göre oluşturulan veri setidir. Türkiye ve 36 Avrupa ülkesinin karşılaştırılmasında kullanılan 12 sosyo-ekonomik değişken ve değişkenlerinin gösteriminde kullanılan simgeler Tablo 4.4.'de verilmiştir.

Tablo 4.4. Türkiye ve 36 Avrupa ülkesinin karşılaştırılmasında kullanılan 12 sosyo-ekonomik değişken

Değişken	Simge
Tarımın GSYİH'ya katkısı	X_1
Kişi başına düşen CO2 emisyonu	X_2
Kişi başına düşen elektrik tüketimi	X_3
Kişi başına düşen enerji kullanımı	X_4
Sabit fiyatlarla kişi başına düşen milli gelir	X_5
Yüksek teknolojik ürünlerin ihracattaki payı	X_6
GSYİH'dan ithalata ayrılan pay	X_7
Sanayinin GSYİH'ya katkısı	X_8
Yıllık enflasyon	X_9
100 kişi içersinde internet kullanıcılarının sayısı	X_{10}
Doğuştta yaşam beklentisi	X_{11}
Bebek ölümlülüğü	X_{12}

Türkiye ve 36 Avrupa ülkesinin karşılaştırılmasında kullanılacak 12 sosyo-ekonomik göstergeler 2005 yılına aittir. Ükelere ait veriler Dünya Bankasının yayınladığı World Development Indicators (<http://dataworldbank.org/indicators>)'dan derlenmiştir. Bu veri setinde de bazı değişkenler (sabit fiyatlarla kişi başına düşen milli gelir, doğuştta yaşam beklentisi, yüksek teknolojik ürünlerin ihracattaki payı vb.) gelişmişlikle doğru orantılı iken bazı değişkenler (bebek ölümlülüğü, tarımın GSYİH ya katkısı vb.) ise gelişmişlik ile ters orantılıdır. Türkiye ve 36 Avrupa ülkesinin karşılaştırılmasında kullanılacak 12 sosyo-ekonomik değişkenlere ait değerler Tablo 4.5.'de verilmiştir.

Tablo 4.5. Türkiye ve 36 Avrupa ülkesine ait seçilmiş 12 sosyo-ekonomik gösterge

Ülke	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
Avs	2	9	7886	4135	33270	14	49	30	2	55	79	5
Blc	1	10	8510	5889	32210	9	83	24	3	58	79	5
Arn	23	1	1221	766	6280	5	46	22	3	6	76	17
Erm	21	1	1480	832	4210	1	43	45	3	5	73	27
Azr	10	4	2407	1653	3940	1	53	64	16	8	67	46
Blg	9	6	4122	2593	9260	5	76	29	4	20	73	13
Hrv	7	5	3476	2005	14790	11	49	28	3	33	75	7
Cek	3	12	6342	4418	19380	13	69	38	0	32	76	4
Dnm	1	9	6665	3636	33740	22	44	25	3	77	78	5
Est	5	14	5567	3744	15730	27	86	29	5	58	73	7
Fnl	3	10	16120	6554	30860	26	37	31	0	73	79	4
Frs	2	6	7945	4537	30910	20	27	21	2	43	80	5
Grc	17	1	1672	714	3550	25	52	27	8	6	71	32
Alm	1	10	7113	4187	31680	17	36	29	1	65	79	5
Ynn	4	9	5242	2794	24550	10	33	22	3	22	79	5
Mcr	4	6	3771	2757	16060	26	68	30	2	37	73	8
Izl	6	7	27987	12179	33660	27	44	24	3	85	81	3
Irl	2	10	6234	3647	32730	35	70	36	3	37	79	5
Itl	2	8	5669	3169	28030	8	26	27	2	48	80	4
Ltn	4	3	2702	1946	12880	5	62	22	10	42	71	10
Ltv	5	4	3104	2543	13860	6	65	33	7	34	71	9
Mkd	13	5	3417	1347	7510	1	62	30	4	23	74	16
Mld	20	2	1472	944	2530	4	92	16	9	15	68	20
Hll	2	8	6988	5048	35020	30	62	24	2	79	79	6
Nrv	2	11	25083	7151	47630	17	28	43	9	80	80	4
Ply	5	8	3437	2429	13480	4	37	31	3	35	75	7
Prt	3	6	4663	2574	20250	9	37	25	3	32	78	5
Rmn	10	4	2331	1775	9280	3	43	35	12	17	72	17
Rsy	6	11	5785	4586	11570	8	21	39	19	15	65	17
Slvk	4	7	4920	3503	15510	7	81	36	2	50	74	8
Slvy	3	7	6918	3629	23260	5	63	34	2	46	78	4
Isp	3	8	6147	3340	26980	7	31	30	4	44	81	5
Isv	1	5	15440	5784	32270	17	41	28	1	81	81	4
Isvcr	1	6	8305	3630	39140	22	42	27	0	68	81	5
Trk	11	3	1921	1201	10850	2	25	29	7	14	71	29
Ukr	10	7	3246	3041	5520	4	51	32	25	17	68	24
Ing	1	9	6233	3894	32850	28	30	24	2	65	79	6

4.7.1. Birinci Düzey İstatistikî Bölgelerin Gelişmişlik Karşılaştırılması

Bu bölümde oluşturulan Birinci Düzey İstatistikî Bölge birimlerinin seçilen 10 sosyo-ekonomik göstergeye göre karşılaştırılması yapılacaktır. Tablo 4.3.'de verilen veri setinden öncelikle birimler arasındaki uzaklık matrisi Öklid uzaklığına göre oluşturulacaktır. Elde edilen uzaklık matrisi temel alınarak metrik ÇBÖ algoritmalarından klasik ÇBÖ ve SMACOF algoritmaları uygulanacaktır. Her iki algortmada da ÇBÖ modellerinden aralıklı ÇBÖ modeli kullanılmış ve tercih edilen indirgenmiş boyut sayısı 2 olarak kabul edilmiştir.

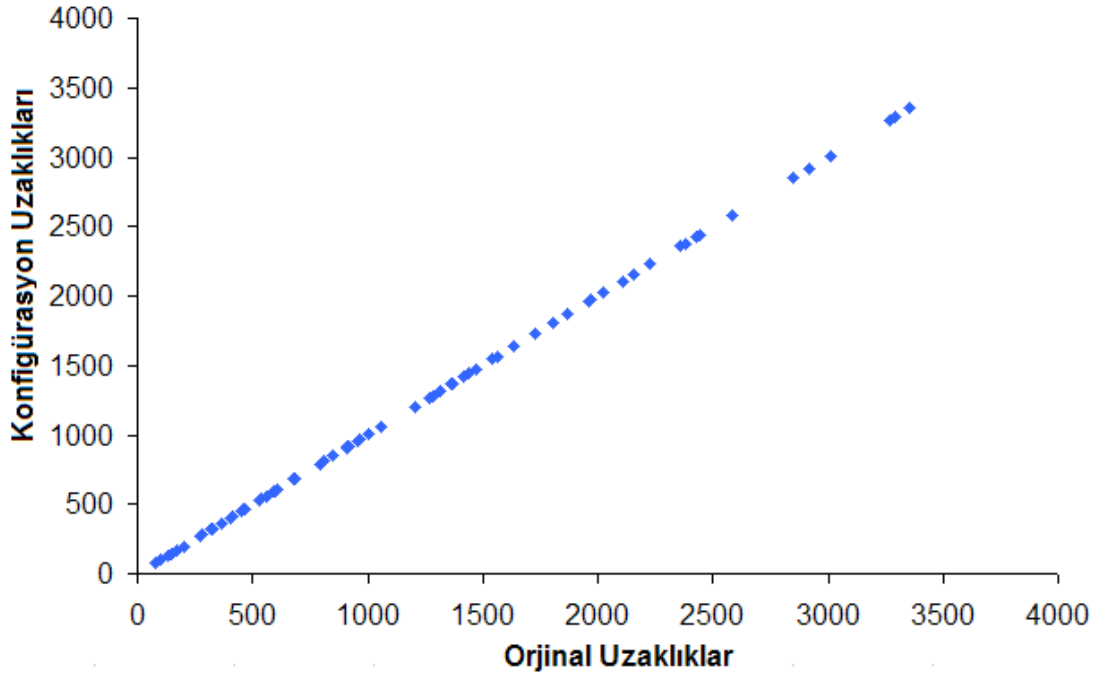
4.7.1.1. Klasik Çok Boyutlu Ölçekleme Analizi İle Birinci Düzey İstatistikî Bölgelerin Karşılaştırılması

Klasik ÇBÖ modeline göre Öklid uzaklığına göre hesaplanan uzaklık matrisi temel alınarak öncelikle eşitlik (4.11.) kullanılarak A matrisi oluşturulur. Daha sonra eşitlik (4.12.) veya (4.13.) kullanılarak elde edilen B matrisine eşitlik (4.14.) ile spektral dönüşüm uygulanır. Konfigürasyon koordinatları tercih edilen boyut sayısına göre eşitlik (4.15.) ile elde edilir. Konfigürasyon koordinatlarına göre oluşturulan konfigürasyon uzaklıkları ile orijinal uzaklıklar arasındaki uyumun (veya uyumsuzluğun) göstergesi olan stress değeri aralıklı ÇBÖ modeline göre hesaplanır. Aralıklı ÇBÖ modeline ilişkin parametre tahminleri ve stres değeri Tablo 4.6.'de verilmiştir.

Tablo 4.6. Birinci Düzey İstatistikî Bölgelerin Karşılaştırılmasında Klasik ÇBÖ Algoritması ve Aralıklı ÇBÖ Modelinde Parametre Tahminleri ve Anlamlılık Testleri

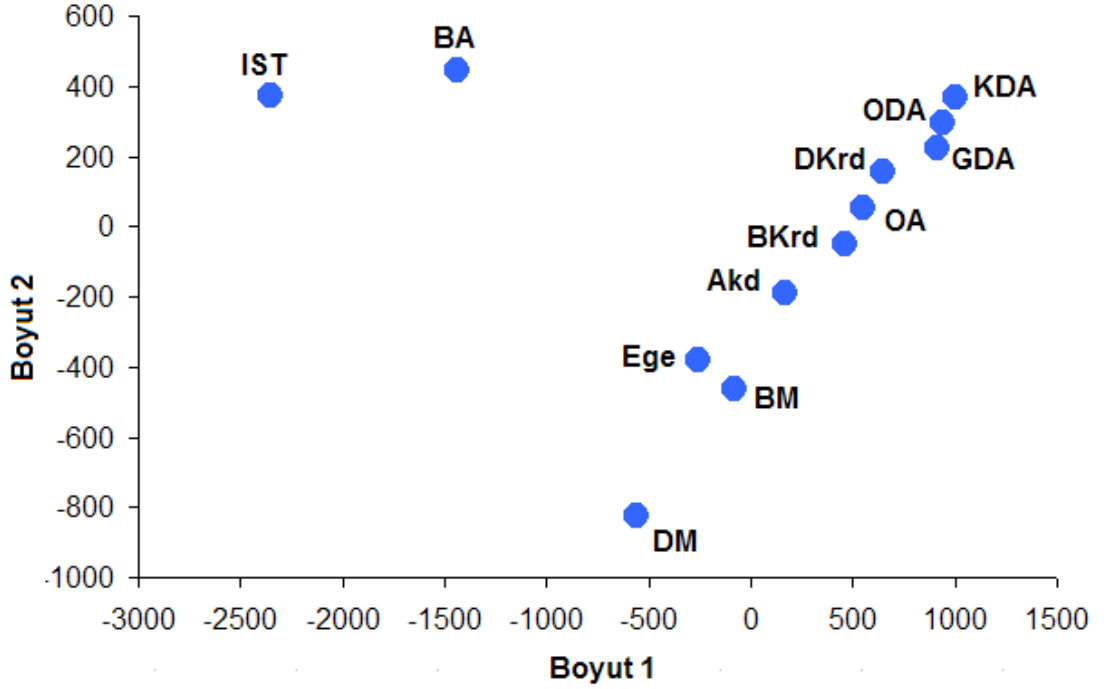
	Tahminler	S.Hata	t_h	P
<i>a</i>	-0,32792	0,040620	-8,072873	< 0.001
<i>b</i>	1,000134	0,000026	38426,622	< 0.001
stress	0.00012			

Düzyey 1'de yer alan 12 birimin seçilen 10 sosyo-ekonomik deęişkene göre konumlarının, iki boyutlu uzayda temsil edilmesinde stress deęeri 0.00012 olarak elde edilmiştir. Stress deęerinin 0.025 deęerinden daha düşük olması orijinal uzaklıklar ile konfigürasyon uzaklıkları arasında mükemmel uyum olarak belirtilmektedir. Buna göre, Düzyey 1'de yer alan 12 birimin metrik karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritması ile iki boyutlu düzlemde mükemmel bir şekilde temsil edilmiştir. Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki saçılım grafięi Şekil 4.4.'de verilmiştir.



Şekil 4.4. Orijinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafięi

Düzyey 1'de yer alan birimlerin seçilen sosyo-ekonomik deęişkenlere göre birbirlerine olan konumlarının iki boyutlu uzayda temsili gösterimi Şekil 4.5.'de verilmiştir.



Şekil 4.5. İstatistiki Bölge Birimlerinin Düzey 1 İtibariyle ÇBÖ Analiziyle İki Boyutlu Gösterimi

Şekil 4.5. incelendiğinde İstanbul, Batı Anadolu ve Doğu Marmara istatistiki bölgelerinin diğer bölgelerden belirgin bir şekilde ayrıldığı görülmektedir. Seçilen 10 sosyo-ekonomik değişken bakımından birbirine en benzer iki istatistiki bölge Ortadoğu Anadolu ile Güneydoğu Anadolu bölgeleridir. Seçilen sosyo-ekonomik değişkenler bakımından birbirine en benzemeyen iki bölge ise İstanbul ile Kuzeydoğu Anadolu bölgesidir. Seçilen sosyo-ekonomik göstergelere göre gelişmişlik sıralamasında birinci sırada bulunan İstanbul bölgesine diğer bölgelerin orijinal ve konfigürasyon uzaklıkları Tablo 4.7.'de verilmiştir.

Tablo 4.7. İstanbul Bölgesinin Diğer İstatistikî Bölgelere Olan Orijinal ve Konfigürasyon Uzaklıkları

		Orijinal Uzaklıklar	Konfigürasyon Uzaklıkları
İST	BA	917,35	917,13
İST	DM	2157,96	2157,95
İST	Ege	2226,76	2226,72
İST	BM	2425,68	2425,65
İST	Akd	2583,56	2583,53
İST	BKrd	2846,02	2845,98
İST	OA	2916,29	2916,27
İST	DKrd	3012,62	3012,56
İST	GDA	3269,18	3269,16
İST	ODA	3291,34	3291,33
İST	KDA	3357,35	3357,33

Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uyumun mükemmel olduğu Tablo 4.7.'den de anlaşılmaktadır. İstanbul bölgesine gelişmişlik bakımından en benzer bölgeler sırasıyla Batı Anadolu, Doğu Marmara, Ege, Akdeniz, Batı Karadeniz, Orta Anadolu, Doğu Karadeniz, Güneydoğu Anadolu, Ortadoğu Anadolu ve Kuzeydoğu Anadolu istatistikî bölgeleridir.

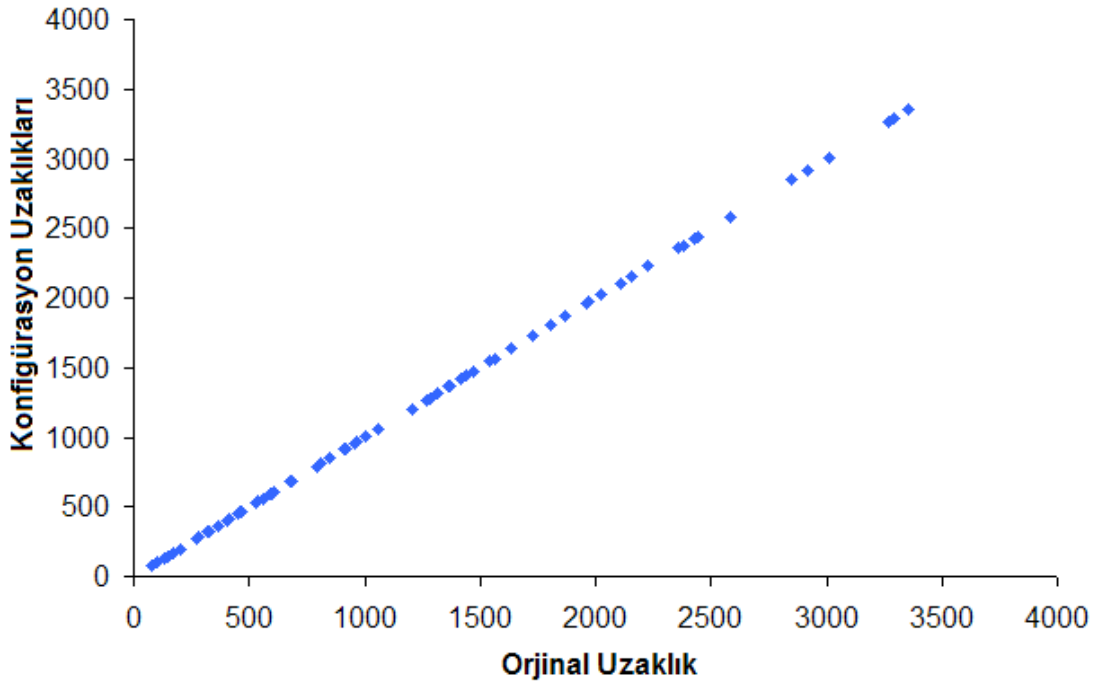
4.7.1.2. SMACOF Algoritması ile Birinci Düzey İstatistikî Bölgelerin Karşılaştırılması

Başlangıç koordinatlarının düzgün dağılımdan rassal olarak oluşturulduğu SMACOF algoritması ve aralıklı ÇBÖ modeline göre elde edilen parametre değerleri, anlamlılık testleri ve stress değeri Tablo 4.8.'de verilmiştir.

Tablo 4.8. Birinci Düzey İstatistikî Bölgelerin Karşılaştırılmasında Rassal Başlangıçlı SMACOF Algoritması ve Aralıklı ÇBÖ Modelinde Parametre Tahminleri ve Anlamlılık Testleri

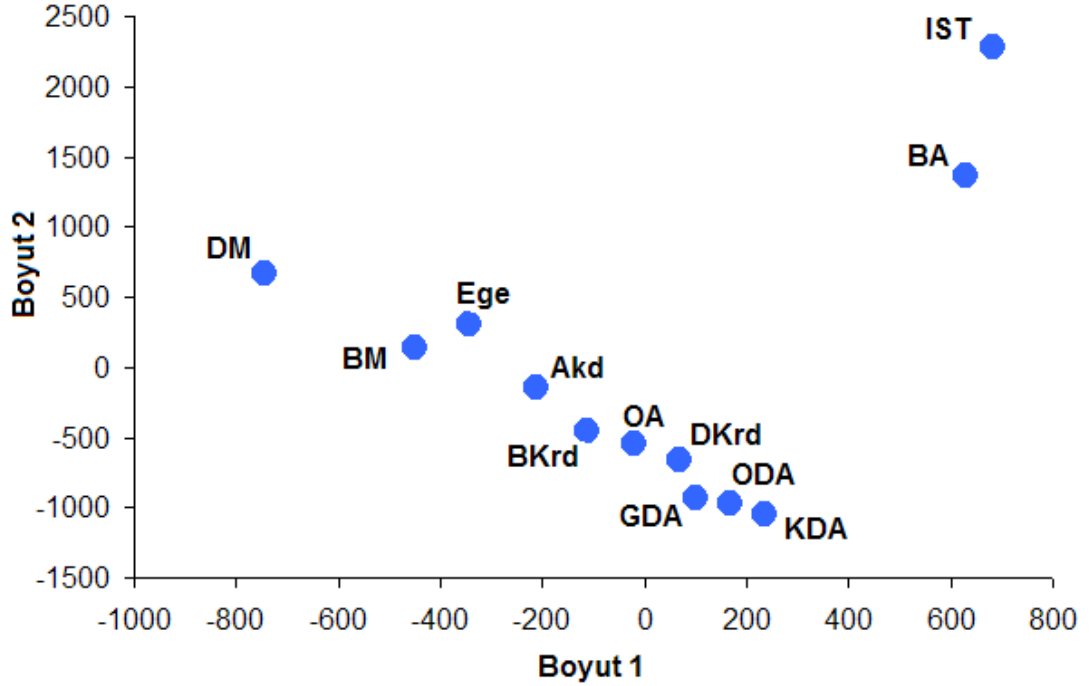
	Tahminler	S.Hata	t_h	P
a	-0.17301	0,036097	-4,792957	< 0.001
b	1.00009	0,000023	43239,78	< 0.001
stress	0.000107			

Düzey 1’de yer alan 12 birimin seçilen 10 sosyo-ekonomik değişkene göre konumlarının, iki boyutlu uzayda SMACOF algoritması ile temsil edilmesinde stress değeri 0.000107 olarak elde edilmiştir. Buna göre, Düzey 1’de yer alan 12 birimin metrik karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritması ile iki boyutlu düzlemde mükemmel bir şekilde temsil edilmiştir. Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki saçılım grafiği Şekil 4.6.’da verilmiştir.



Şekil 4.6. Orijinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafiği

Düzey 1’de yer alan birimlerin seçilen sosyo-ekonomik değişkenlere göre birbirlerine olan konumlarının iki boyutlu uzayda SMACOF algoritması ile temsili gösterimi Şekil 4.7’de verilmiştir.



Şekil 4.7. İstatistiki Bölge Birimlerinin Düzey 1 İtibariyle SMACOF Algoritması ile Elde Edilen İki Boyutlu Gösterimi

Şekil 4.7’de SMACOF algoritması ile elde edilen iki boyutlu gösterim, klasik ÇBÖ algoritması ile elde edilen iki boyutlu gösterimin saat yönünde 90° döndürülmesi ile elde edilecek görünümüne çok benzemektedir. Her iki algoritma ile elde edilen iki boyutlu gösterimle elde edilecek sonuçlar aynıdır. Rassal başlangıçlı SMACOF algoritması ile stress değerinde yaklaşık 0.00001 birimlik bir iyileşme sağlanmıştır.

4.7.2. Avrupa Ülkelerinin Sosyo-Ekonomik Göstergelere Göre Gelişmişlik Düzeylerinin Karşılaştırılması

Bu kısımda da, Türkiye ve seçilmiş 36 Avrupa ülkesinin belirlenen 12 sosyo-ekonomik değişken bakımından gelişmişlik değerlendirilmesinde ÇBÖ analizi ile gerçekleştirilecektir. Tablo 4.5.’de verilen veri setinden Öklid uzaklığı ile elde edilen uzaklık matrisinin temel alınacağı uygulamada, 37 ülkenin birbirlerine göre konumları iki boyutlu düzlemde klasik ÇBÖ ve rassal başlangıçlı SMACOF

algoritması ile belirlenecektir. Her iki uygulamada da ÇBÖ modeli olarak aralıklı ÇBÖ modeli kullanılacaktır.

4.7.2.1. Klasik Çok Boyutlu Ölçekleme Analizi İle Avrupa Ülkelerinin Karşılaştırılması

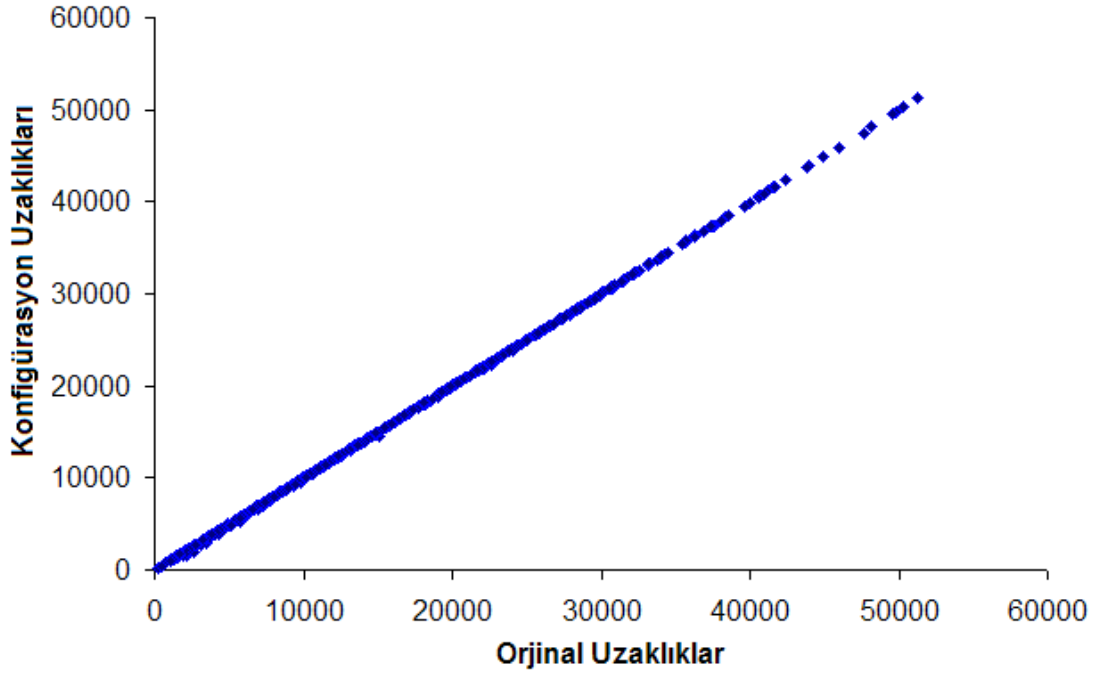
Türkiye ve 36 Avrupa ülkesinin seçilmiş 12 sosyo-ekonomik değişkene göre gelişmişlik değerlendirilmesinde klasik ÇBÖ algoritması ve aralıklı ÇBÖ modeline göre elde edilen parametre tahminleri ve stres değeri Tablo 4.9.'da verilmiştir.

Tablo 4.9. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde Klasik ÇBÖ Algoritması ve Aralıklı ÇBÖ Modelinde Parametre Tahminleri ve Anlamlılık Testleri

	Tahminler	S.Hata	t_h	p
a	-77,854190	5,779106	-13,471667	< 0.001
b	1,001676	0,000300	3337,817599	< 0.001
stress	0.004311			

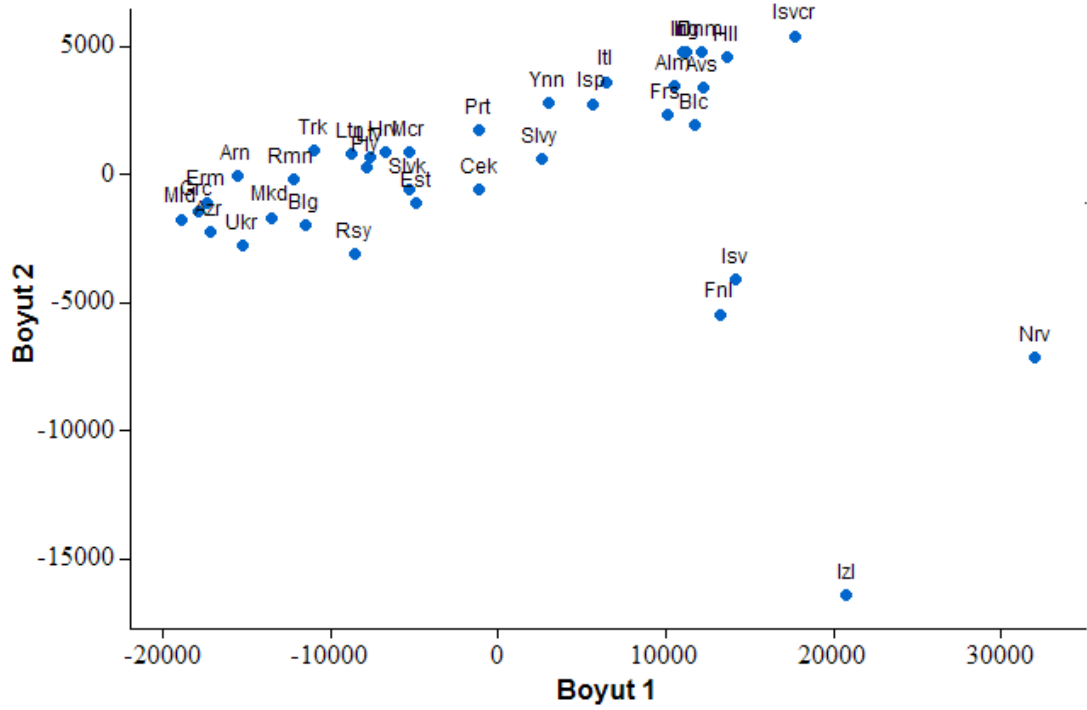
Türkiye ve 36 Avrupa ülkesinin klasik ÇBÖ algoritması ile iki boyutlu düzlemde konumlandırılmasında orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uyumun(veya uyumsuzluğun) bir göstergesi olan stress değeri 0.004311 olarak elde edilmiştir. Elde edilen bu stress değeri orijinal uzaklıklar ile konfigürasyon uzaklıkları arasında mükemmel uyumu göstermektedir. Tablo 4.9.'da yer alan parametrelerin anlamlılık testlerinde tüm parametreler anlamlı bulunmuştur.

Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki saçılım grafiği Şekil 4.8.'de verilmiştir.



Şekil 4.8. Orjinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafiği

Şekil 4.8. incelendiğinde orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki mükemmel uyum açık bir şekilde görülmektedir. Orijinal uzaklıklar ile konfigürasyon uzaklıkların saçılımı birinci açığortay üzerindedir. Türkiye ve 36 Avrupa ülkesinin klasik ÇBÖ algoritması ile elde edilen iki boyutlu düzlemdeki grafiği Şekil 4.9.'da verilmiştir.



Şekil 4.9. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde Klasik ÇBÖ Algoritması ile Elde Edilen İki Boyutlu Gösterimi

Şekil 4.8. incelendiğinde Norveç, İzlanda, Finlandiya ve İsveç'in diğer ülkelerden belirgin bir şekilde ayrıldığı görülmüştür. Göz önünde tutulan 12 sosyo-ekonomik göstereye göre en gelişmiş ülke Norveç'tir.

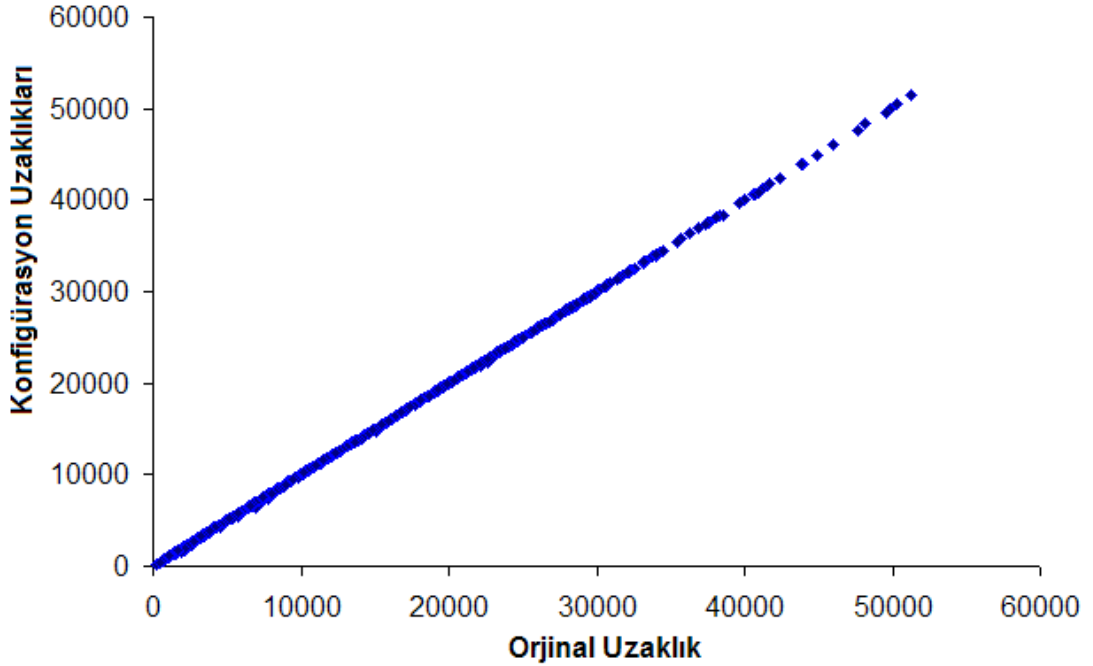
4.7.2.2. SMACOF Algoritması ile Avrupa Ülkelerinin Karşılaştırılması

Ülkeler arasındaki uzaklıkların indirgenmiş iki boyutlu koordinat düzlemde temsilinin rassal başlangıçlı SMACOF algoritması ve aralıklı ÇBÖ modeline göre elde edilmesinde hesaplanan parametre tahminleri, parametrelerin anlamlılık testleri ve stress değeri Tablo 4.10.'da verilmiştir.

Tablo 4.10. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde SMACOF Algoritması ve Aralıklı ÇBÖ Modelinde Parametre Tahminleri ve Anlamlılık Testleri

	Tahminler	S.Hata	t_h	p
a	-41,1838	4,242362	-9,70775	< 0.001
b	1,001765	0,00022	4547,306	< 0.001
Stress	0.003159			

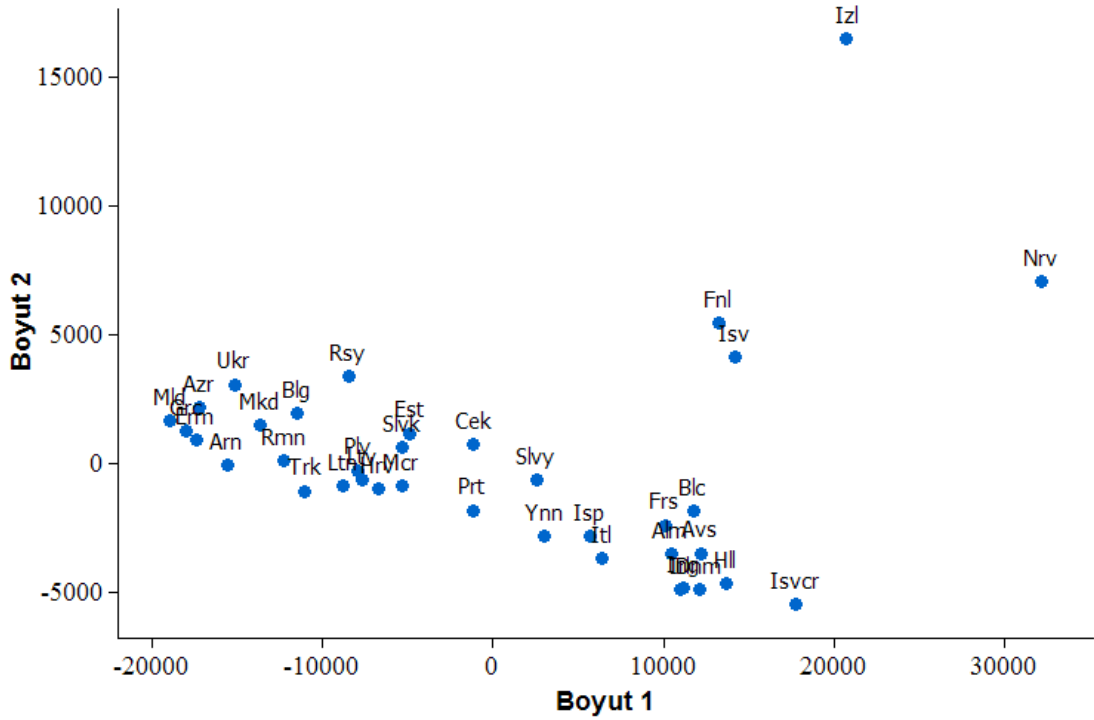
Tablo 4.10. incelendiğinde aralıklı ÇBÖ modelinin istatistiksel olarak anlamlı olduğu görülmektedir. Rassal başlangıçlı SMACOF algoritması ile elde edilen stress değeri 0.003159'dur. Rassal başlangıçlı SMACOF algoritması ile klasik çok boyutlu ölçekleme algoritmasına göre stress değerinde 0.001 birimlik bir iyileşme sağlanmıştır. Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki saçılım grafiği Şekil 4.10.'da verilmiştir.



Şekil 4.10. Orijinal uzaklıklar ile Konfigürasyon uzaklıklarının Saçılım Grafiği

Şekil 4.10. incelendiğinde orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki mükemmel uyum açık bir şekilde görülmektedir. Orijinal uzaklıklar ile

konfigürasyon uzaklıklarının saçılımı birinci açırtay üzerindedir. Türkiye ve 36 Avrupa ülkesinin rassal başlangıçlı SMACOF algoritması ile elde edilen iki boyutlu düzlemdeki grafiği Şekil 4.11.'de verilmiştir.



Şekil 4.11. Türkiye ve 36 Avrupa Ülkesinin Gelişmişlik Değerlendirilmesinde Rassal Başlangıçlı SMACOF Algoritması ile Elde Edilen İki Boyutlu Gösterimi

Şekil 4.11. incelendiğinde ülkeler arasındaki uzaklıkların gelişmişlik düzeyleri ile ilişkili olduğu anlaşılmaktadır. Gelişmişlik düzeyi en yüksek olarak tespit edilen Norveç'e diğer ülkelerin konfigürasyon uzaklıkları hesaplanmıştır. Bir ülkenin Norveç' olan uzaklığı ne kadar küçükse gelişmişlik düzeyinin o kadar yüksek olarak tanımlanacaktır. Türkiye, Norveç'e olan konfigürasyon uzaklığına göre gelişmişlik sıralamasında 28. sıradadır.

5. UZAKLIK FONKSİYONLARININ STRESS ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Bu bölümde uzaklık fonksiyonlarının metrik ÇBÖ algoritmalarında stress değeri üzerindeki etkisi incelenecektir. Uzaklık fonksiyonlarının stress değeri üzerindeki etkisi incelenirken hem gerçek hem de simülasyon verileri kullanılacaktır. Kullanılacak gerçek veri setleri özellikle sınıflama çalışmalarında yaygın olarak kullanılan veri setleridir. Simülasyon çalışmasında ise örneklem hacmi, terim büyüklüklerinin farklı olup olmaması ve kovaryans matrisinin küresel olup olmaması durumu göz önünde bulundurulacaktır.

5.1. Gerçek Veri Setleri İle Uzaklık Fonksiyonlarının Stress Üzerindeki Etkisinin İncelenmesi

Uzaklık fonksiyonlarının, metrik çok boyutlu ölçekleme algoritmalarında stress değeri üzerindeki etkisinin inceleneceği bu çalışmada konfigürasyon uzaklıkları ile orijinal uzaklıklar arasında (4.8.) nolu eşitlikte tanımlanan aralıklı ÇBÖ modeli kullanılmıştır. ÇBÖ algoritmalarında tercih edilen boyut sayısı 2 olarak sabitlenmiştir. Stress değeri hesaplanırken (4.30.) nolu eşitlik ile verilen formül kullanılmıştır. Çalışma kapsamında, özellikle sınıflama analizlerinde yaygın olarak kullanılan farklı örneklem hacmi ve farklı değişken sayılarına sahip 13 farklı veri seti kullanılmıştır. Veri setleri UCI Machine Learning Repository(Asuncion ve Newman, 2007) internet sitesinden alınmıştır. İncelenen veri setlerine ait örneklem hacimleri ve değişken sayıları Tablo 5.1.'de verilmiştir. Tablo 5.1'de verilen örneklem hacimleri bazılarında örneklem hacimleri algoritmalarından daha hızlı sonuç alabilmek amacı ile düşürülmüştür. Bu amaçla Tablo 5.1. de verilen Shuttle, Letter, Landsat, Atmosfer, Vehicle, Vowel ve Yeast veri setleri içersinden rassal olarak seçilen birimler ile örneklem hacmi indirgenmiş örneklem oluşturulmuştur.

Tablo 5.1. Çalışmada kullanılacak veri setlerinin örneklem hacimleri ve değişken sayıları

	<i>n</i>	<i>p</i>
Diabet	145	4
E.coli	336	7
Glass	214	9
Iris	150	4
Wine	178	13
Shuttle[*]	263	36
Letter[*]	158	16
Landsat[*]	67	36
Liver	345	6
Atmosfer[*]	351	32
Vehicle[*]	198	18
Vowel[*]	159	14
Yeast[*]	317	8

* : Örneklem hacmi rassal olarak indirgenmiş veriler

Çalışmadaki hesaplamalar Matlab programında gerçekleştirilmiştir. Klasik metrik ÇBÖ algoritmasında City-Block, Öklid ve Standartlaştırılmış Öklid uzaklık fonksiyonlarının kullanılması ile elde edilen stress değerleri Tablo5.2.'de verilmiştir.

Tablo 5.2. Klasik metrik ÇBÖ algoritmasında farklı uzaklık fonksiyonlarına göre elde edilen stress değerleri

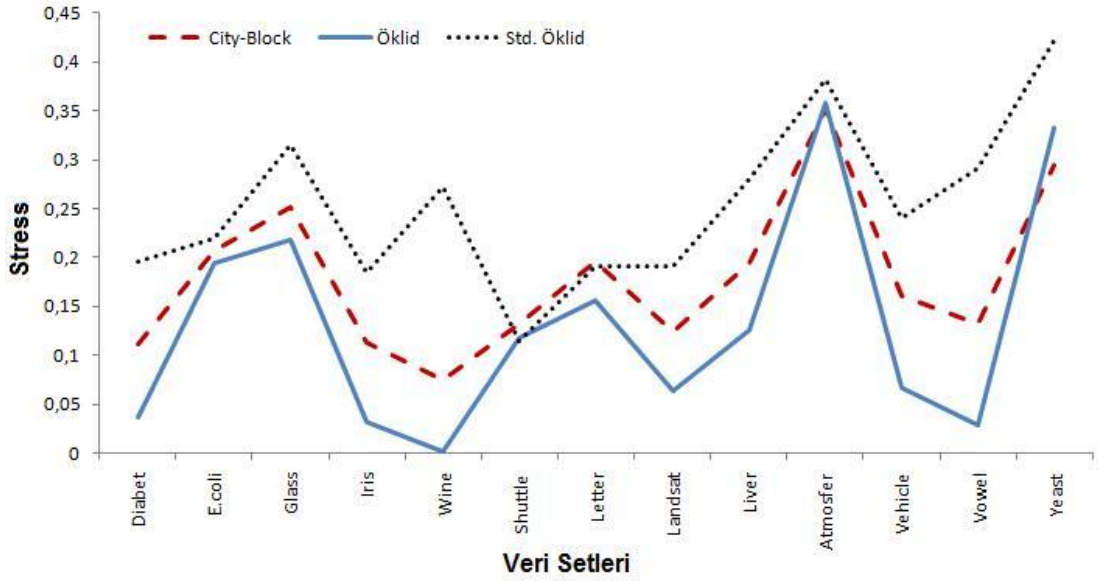
	City-Block	Öklid	Std. Öklid
Diabet	0.1119	0.0369	0.1957
E.coli	0.2072	0.1936	0.2203
Glass	0.2509	0.2187	0.3151
Iris	0.1133	0.0319	0.1852
Wine	0.0746	0.0008	0.2729
Shuttle	0.1318	0.1177	0.1141
Letter	0.1964	0.1555	0.1910
Landsat	0.1248	0.0636	0.1910
Liver	0.1944	0.1250	0.2802
Atmosfer	0.3524	0.3588	0.3816
Vehicle	0.1603	0.0670	0.2402
Vowel	0.1321	0.0279	0.2913
Yeast	0.2949	0.3320	0.4215
Ortalama	0.1804	0.1330	0.2539
Std. Hata	0.0223	0.0320	0.0237

5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ

Tablo 5.2. incelendiğinde Shuttle, Atmosfer ve Yeast veri setleri haricindeki diğer 10 veri setinde en düşük stress değeri Öklid uzaklığının kullanıldığı durumda elde edilmiştir. Shuttle veri setinde en iyi stress değeri standartlaştırılmış Öklid uzaklığı ile elde edilirken Atmosfer ve Yeast veri setlerinde en iyi stress değeri City-Block uzaklığı ile elde edilmiştir. Shuttle, Atmosfer ve Yeast veri setleri için hesaplanan stress değerleri incelendiğinde orijinal uzaklıklar ile konfigürasyon uzaklıkları arasında Shuttle verisinde orta uyumun, Atmosfer ve Yeast verisinde kötü uyumun olduğu görülmektedir. İncelenen veri setleri için hesaplanan stress değerlerinin ortalaması incelendiğinde en kötü stress değeri Standartlaştırılmış Öklid uzaklığının kullanılması durumunda elde edilmiştir.

Farklı veri setlerinde, farklı uzaklık fonksiyonları ile Klasik ÇBÖ algoritması için hesaplanan ve Tablo 5.2’de verilen stress değerlerinin değişimi Şekil 5.1’de verilmiştir.



Şekil 5.1. Farklı veri setleri için klasik metrik ÇBÖ algoritmasında farklı uzaklık fonksiyonlarının kullanılması ile elde edilen stress değerleri

Şekil 5.1 incelendiğinde Klasik ÇBÖ algoritmasında stress değerini en iyileyen uzaklık fonksiyonunun Öklid uzaklığı olduğu açıkça görülmektedir. Ayrıca City-Block uzaklığı ile elde edilen stress değerlerinin, standartlaştırılmış Öklid

uzaklığı ile elde edilen stress değerlerinden daha iyi olduğu rahat bir şekilde görülmektedir.

Karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında uzaklık fonksiyonlarının stress değerleri üzerindeki etkisi belirlenirken algoritmada durdurma kuralı için $EKF = 10^{-6}$ alınmıştır. Başlangıç koordinat değerlerinin rassal olarak belirlendiği karşılaştırmada, her bir veri seti için algoritma 10 kez tekrar edilmiş ve elde edilen stress değerlerinin ortalaması Tablo 5.3.'de verilmiştir.

Tablo 5.3. Karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında farklı uzaklık fonksiyonlarına göre 10 tekrar sonucunda elde edilen ortalama stress değerleri(Başlangıç Koordinatları Rassal)

	City-Block	Öklid	Std. Öklid
Diabet	0.0713	0.0532	0.0812
E.coli	0.1916	0.1715	0.4336
Glass	0.1848	0.1397	0.4472
Iris	0.0710	0.0429	0.0721
Wine	0.0244	0.0033	0.4726
Shuttle	0.1301	0.1082	0.1102
Letter	0.1960	0.1537	0.4643
Landsat	0.1217	0.0672	0.1201
Liver	0.1643	0.0954	0.2185
Atmosfer	0.3254	0.3123	0.3451
Vehicle	0.1087	0.0652	0.4399
Vowel	0.1540	0.0337	0.4823
Yeast	0.2654	0.2802	0.3124
Ortalama	0.1545	0.1174	0.3077
Std. Hata	0.0227	0.0259	0.0457

Tablo 5.3. incelendiğinde, karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında da her veri setinde en düşük stres değerleri Öklid uzaklığının kullanılması durumunda elde edilmiştir.

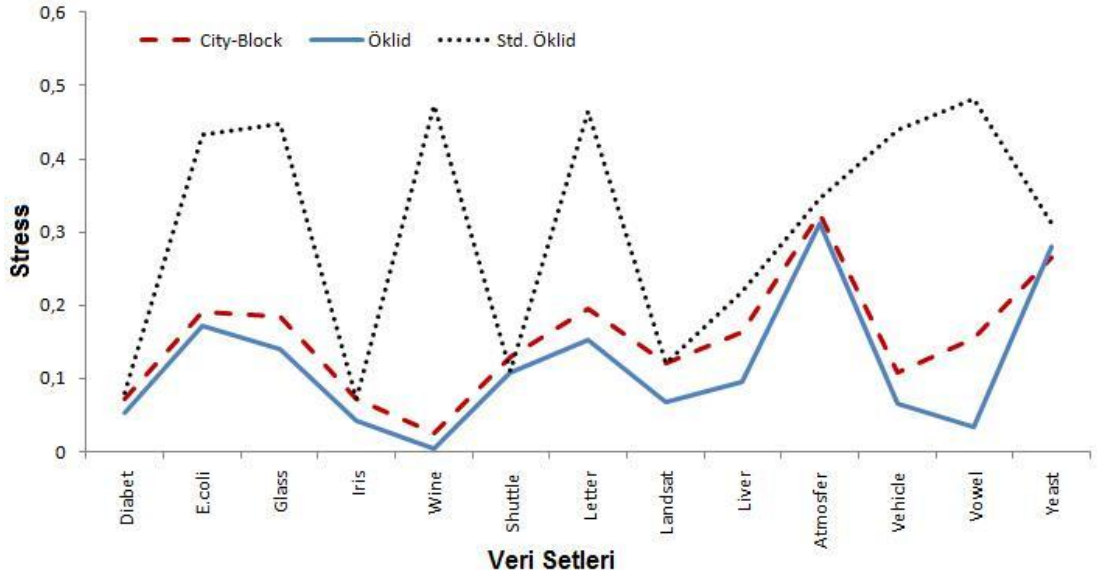
Tablo 5.2. ve Tablo 5.3.'deki ortalama stres değerleri incelendiğinde City-Block ve Öklid uzaklığının kullanılması durumunda Karmaşık Fonksiyonların Optimizasyonu ile Ölçekleme algoritmasından elde edilen stress değerleri Klasik Metrik ÇBÖ algoritmasından daha düşüktür. Buna karşın uzaklık fonksiyonunun Standartlaştırılmış Öklid uzaklığı olması durumunda Klasik Metrik ÇBÖ algoritması,

5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ

karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasından daha iyi ortalama stress değerine sahiptir.

Farklı veri setlerinde, farklı uzaklık fonksiyonları ile karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritması için hesaplanan ve Tablo 5.3’de verilen stress değerlerinin değişimi Şekil 5.2’de verilmiştir.



Şekil 5.2. Farklı veri setleri için rassal başlangıçlı karmaşık fonksiyonların optimizasyonu algoritmasında farklı uzaklık fonksiyonlarının kullanılması ile elde edilen stress değerleri

Şekil 5.2 incelendiğinde E.coli, Glass, Wine, Letter, Vehicle ve Vowel veri setlerinde Standartlaştırılmış Öklid uzaklığı oldukça kötü sonuçlar vermiştir. Bu veri setleri dışında kalan diğer veri setlerinde özellikle City-Block uzaklığı ile Standartlaştırılmış Öklid uzaklığı benzer sonuçlar vermiştir.

Karmaşık fonksiyonların optimizasyonu ile ölçeklemede başlangıç koordinatları rassal olarak belirlenebileceği gibi belirli bir yöntemle de belirlenebilir. Burada karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında başlangıç koordinat değerlerinin, Temel Bileşenler (TB) analizinden elde edilen bileşenlere göre belirlenmesi durumunda farklı uzaklık fonksiyonlarının stress değeri üzerindeki etkisi ortaya konacaktır. Başlangıç koordinatlarının kovaryans matrisi kullanılarak gerçekleştirilen TB analizinden elde edilen en yüksek

5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ

varyans açıklama oranına sahip iki bileşen skorları olarak seçilmesi durumunda farklı uzaklık fonksiyonlarına göre hesaplanan stress değerleri Tablo 5.4.'de verilmiştir.

Tablo 5.4. Karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında farklı uzaklık fonksiyonlarına göre elde edilen stress değerleri (Başlangıç Koordinatları Temel Bileşen Skorları)

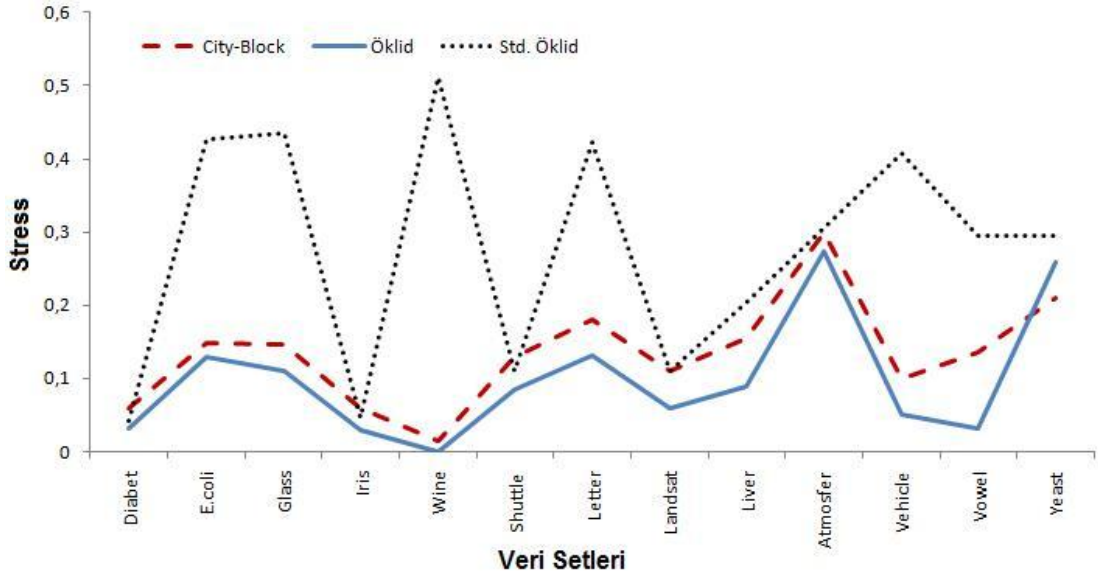
	City-Block	Öklid	Std. Öklid
Diabet	0.0588	0.0314	0.0415
E.coli	0.1475	0.1285	0.4265
Glass	0.1467	0.1109	0.4349
Iris	0.0591	0.0298	0.0460
Wine	0.0148	0.0007	0.5114
Shuttle	0.1284	0.0852	0.1094
Letter	0.1807	0.1324	0.4231
Landsat	0.1107	0.0592	0.1108
Liver	0.1541	0.0894	0.2045
Atmosfer	0.2998	0.2734	0.3054
Vehicle	0.0987	0.0498	0.4065
Vowel	0.1354	0.0319	0.2957
Yeast	0.2104	0.2587	0.2954
Ortalama	0.1342	0.0986	0.2778
Std. Hata	0.0202	0.0235	0.0447

Başlangıç koordinatlarının TB analizi ile belirlendiği karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasında en iyi stress değerleri Öklid uzaklığı kullanılarak elde edilmiştir. Ortalama stress değerlerine göre incelenen uzaklıklar içerisinde Öklid uzaklığından sonra stress değerini en küçükleyen uzaklık City-Block uzaklığıdır.

Farklı veri setlerinde, farklı uzaklık fonksiyonları ile başlangıç koordinatları TB analizinden elde edilen karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritması için hesaplanan ve Tablo 5.4'de verilen stress değerlerinin değişimi Şekil 5.3'de verilmiştir.

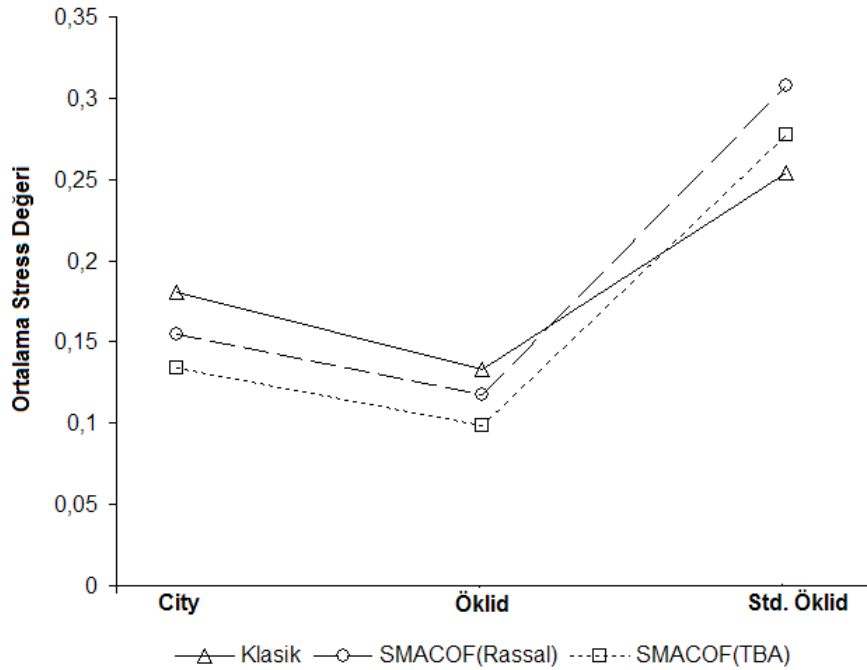
5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ



Şekil 5.3. Farklı veri setleri için temel bileşen başlangıçlı karmaşık fonksiyonların optimizasyonu algoritmasında farklı uzaklık fonksiyonlarının kullanılması ile elde edilen stress değerleri

Farklı metrik ÇBÖ algoritmalarında farklı uzaklık fonksiyonlarının kullanılması durumunda elde edilen ortalama stress değerleri Şekil 5.4.'de verilmiştir.



Şekil 5.4. Farklı uzaklık fonksiyonlarına göre metrik ÇBÖ algoritmalarının karşılaştırılması

Şekil 5.4.'e göre incelenen algoritmalar içerisinde en iyi stress değeri Öklid uzaklığının kullanıldığı, başlangıç koordinat değerlerinin temel bileşen skorları ile belirlendiği karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritmasından elde edilmiştir. Standartlaştırılmış Öklid uzaklığının kullanılması durumunda en iyi stress değeri klasik çok boyutlu ölçekleme algoritması ile elde edilmiştir.

5.2. Simülasyon Çalışması ile Uzaklık Fonksiyonlarının Stress Değeri Üzerindeki Etkilerinin İncelenmesi

Uzaklık fonksiyonlarının, Metrik ÇBÖ algoritmalarında stress değerine olan etkisi bu bölümde simülasyon çalışmasıyla ortaya konulacaktır. Farklı özellikler göz önünde bulundurularak çok değişkenli normal dağılımdan üretilen örneklemeler Klasik ve başlangıç koordinatının Temel Bileşenler Analizinden elde edildiği SMACOF algoritmaları ile iki boyutlu uzayda temsil edilecektir. Elde edilen stress değerlerine, farklı uzaklık fonksiyonlarının etkisi belirlenirken aynı zamanda örneklem hacmi, terim büyüklüklerinin aynı olup olmaması ve kovaryans matrisinin küresel olup olmamasının da stress değeri üzerinde etkilerinin olup olmadığı belirlenecektir. Kovaryans matrisinin küresel olması ifadesi ile değişkenler arasındaki ilişkilerin aynı olması ifade edilmektedir. Buna karşın kovaryans matrisinin küresel olmaması da değişkenler arasındaki ilişkilerin farklı olması anlamına gelmektedir.

Gerçek veri setlerinde temel bileşen başlangıçlı SMACOF algoritmasının rassal başlangıçlı SMACOF algoritmasına göre daha iyi sonuç vermesi ve rassallığın etkisini ortadan kaldırmak için simülasyon çalışmasında Klasik ÇBÖ algoritması sadece temel bileşen başlangıçlı SMACOF algoritması ile karşılaştırılacaktır.

5.2.1. Çok Değişkenli Normal Dağılımdan Rassal Örneklemeler Üretme

Ortalama vektörü μ ve kovaryans matrisi Σ olan çok değişkenli normal dağılıma sahip p değişkenli X_i rassal değişkeninin olasılık yoğunluk fonksiyonu,

$$f(\underline{x}) = \frac{1}{(2p)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\underline{x}-\mu)' \Sigma^{-1}(\underline{x}-\mu)\right) \quad (5.1.)$$

şeklinde dir. Çok deęişkenli normal dağılımdan pozitif tanımlı Σ matrisi kullanılarak Cholesky ayrışımı veya spektral ayrışım ile iki farklı yoldan rassal sayılar üretilmektedir.

Σ kovaryans matrisini, Cholesky ayrıştırması ile

$$\Sigma = \mathbf{A}'\mathbf{A} \quad (5.2.)$$

şeklinde yazmak mümkündür. Eşitlikte yer alan \mathbf{A} matrisi üst üçgen bir matristir. $\mathbf{Z}' = \begin{bmatrix} \mathbf{Z}'_1 & \mathbf{Z}'_2 & \mathbf{L} & \mathbf{Z}'_p \end{bmatrix}$ matrisi her bir elemanı standart normal dağılıma sahip ve birbirinden bağımsız rassal deęişkenler olmak üzere, çok deęişkenli normal dağılıma sahip X rassal deęişkeni,

$$\mathbf{X} = \mathbf{A}'\mathbf{Z} + \mu \quad (5.3.)$$

dönüşümüyle elde edilir. Σ kovaryans matrisini, spektral ayrışım ile

$$\Sigma^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{V}' \quad (5.4.)$$

şeklinde yazmak mümkündür. Eşitlikte yer alan \mathbf{V} , kovaryans matrisinin özvektörlerinden oluşan bir matris ve \mathbf{D} esas köşegen elemanları kovaryans matrisinin özdeğerlerinden oluşan diyagonal bir matristir. $\mathbf{Z}' = \begin{bmatrix} \mathbf{Z}'_1 & \mathbf{Z}'_2 & \mathbf{L} & \mathbf{Z}'_p \end{bmatrix}$ matrisi her bir elemanı standart normal dağılıma sahip ve birbirinden bağımsız rassal deęişkenler olmak üzere, çok deęişkenli normal dağılıma sahip X rassal deęişkeni,

$$\mathbf{X} = \Sigma^{\frac{1}{2}} \mathbf{Z} + \boldsymbol{\mu} \quad (5.5.)$$

dönüşümüyle elde edilir. Görüldüğü gibi çok değişkenli normal dağılımdan rassal sayı üretmek için öncelikle standart normal dağılımdan rassal sayılar üretmek gerekmektedir. Standart normal dağılımdan rassal sayı üretmek için, R_1 ve R_2 bağımsız Uniform değişkenler olmak üzere,

$$z = \sqrt{-2 \ln(R_1)} \sin(2pR_2) \quad (5.6.)$$

formülü ile standart normal dağılımdan rassal sayılar üretilir. \mathbf{X} $n \times p$ boyutlu veri matrisi, $\boldsymbol{\mu}$ ortalama vektörü ve Σ kovaryans matrisi olmak üzere, \mathbf{X} veri matrisi,

$$\mathbf{Z}_{n \times p} = (\mathbf{X}_{n \times p} - \mathbf{1}_{n \times 1} \boldsymbol{\mu}_{1 \times p}) \left(\sqrt{\text{diagonal}(\Sigma)} \right)_{p \times p}^{-1} \quad (5.7.)$$

formülü ile standartlaştırılır.

5.2.2. Simülasyon

Simülasyon çalışmasında örneklem hacminin stress üzerindeki etkisini ortaya koymak için 10, 20, 30 ve 50 birimlik rassal örneklem oluşturulacaktır. Büyük örneklem hacimlerinde SMACOF algoritmasının yakınsama hızının kötü olmasından dolayı simülasyon çalışmasında küçük örneklem hacimlerinde rassal örneklem oluşturulmuştur. City-Block, Öklid ve Standartlaştırılmış Öklid uzaklıklarının stress değeri üzerindeki etkileri incelenirken oluşturulan örneklem 6 değişken içerecektir. Çok değişkenli normal dağılımdan rassal örneklem oluşturulurken göz önünde bulundurulacak ortalama vektörleri ve kovaryans matrisleri Tablo 5.5.'de verilmiştir.

5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ
ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ

Tablo 5.5. Simülasyon çalışmasında çok değişkenli normal dağılımdan rassal örneklemeler üretilirken göz önünde bulundurulacak ortalama vektörleri ve kovaryans matrisleri

DURUM I						
$\mu =$	1.00	1.50	1.20	1.40	1.10	1.70
$\Sigma =$	1.00	1.40	0.99	1.57	0.99	2.80
	1.40	4.00	1.98	3.13	1.98	5.60
	0.99	1.98	2.00	2.21	1.40	3.96
	1.57	3.13	2.21	5.00	2.21	6.26
	0.99	1.98	1.40	2.21	2.00	3.96
	2.80	5.60	3.96	6.26	3.96	16.00
DURUM II						
$\mu =$	0.10	1.50	120.00	1500.00	1.10	20000.00
$\Sigma =$	0.40	0.89	4.43	14.00	0.63	44.27
	0.89	4.00	14.00	44.27	1.98	140.00
	4.43	14.00	100.00	221.36	9.90	700.00
	14.00	44.27	221.36	1000.00	31.30	2213.59
	0.63	1.98	9.90	31.30	2.00	98.99
	44.27	140.00	700.00	2213.59	98.99	10000.00
DURUM III						
$\mu =$	12.00	20.00	15.00	18.00	22.00	10.00
$\Sigma =$	19.79	3.60	12.82	8.41	38.82	4.64
	3.60	336.64	27.28	26.97	95.89	6.17
	12.82	27.28	380.73	145.26	385.61	13.47
	8.41	26.97	145.26	101.29	208.45	9.39
	38.82	95.89	385.61	208.45	1540.92	44.71
	4.64	6.17	13.47	9.39	44.71	11.14
DURUM IV						
$\mu =$	1.00	1.50	120.00	1500.00	1.10	20000.00
$\Sigma =$	0.40	1.80	1.20	3.00	0.90	0.30
	1.80	36.00	7.20	18.00	5.40	1.80
	1.20	7.20	64.00	12.00	3.60	1.20
	3.00	18.00	12.00	100.00	9.00	3.00
	0.90	5.40	3.60	9.00	9.00	2.40
	0.30	1.80	1.20	3.00	2.40	10000.00

Durum I'de, terim büyüklükleri ve değişkenler arasındaki ilişkiler aynı iken Durum II'de, terim büyüklükleri farklı, değişkenler arasındaki ilişkiler aynıdır. Durum III'de, terim büyüklükleri aynı ve değişkenler arasındaki ilişkiler farklı iken Durum IV'de, terim büyüklükleri ve değişkenler arasındaki ilişkiler farklıdır.

SMACOF (TBA) algoritmasında ardışık iki stress değeri arasındaki fark 10^{-4} oluncaya kadar algoritma çalıştırılmıştır. Orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uyumun ölçüsü olan stress değeri gerçek veri setlerinde olduğu gibi eşitlik (4.30.) kullanılarak hesaplanacaktır. Burada da konfigürasyon uzaklıkları ile orijinal uzaklıklar arasında (4.8.) nolu eşitlikte tanımlanan aralıklı ÇBÖ modeli kullanılmıştır.

Göz önünde bulundurulan her durum için 100'er örneklem üretilmiştir. Simülasyon çalışmasında 4 farklı örneklem hacmi ve 4 farklı durum için toplam 1600 rassal örneklem oluşturulmuştur. Üretilen her örnekleme 3 farklı uzaklık fonksiyonu ile 2 farklı ÇBÖ algoritması uygulanmıştır. Böylelikle gerçekleştirilen simülasyon çalışmasında 9600 tane ÇBÖ analizi uygulanmıştır.

Değişkenler arasındaki ilişkilerin ve terim büyüklüklerinin aynı olması durumunda farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında Klasik ve SMACOF (TBA) algoritmalarından elde edilen stress değerlerine ait ortalama ve standart hata değerleri Tablo 5.6. da verilmiştir.

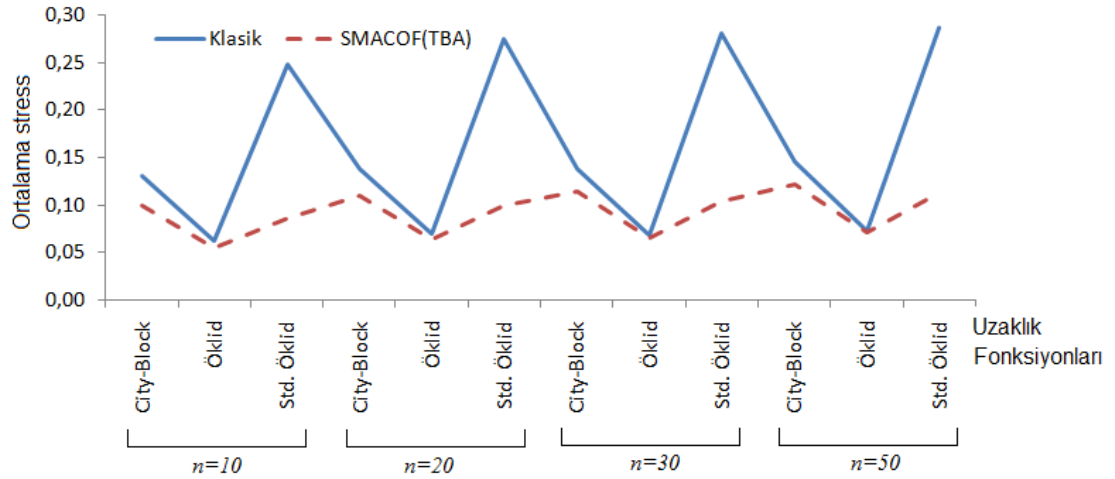
5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ

Tablo 5.6. Değişkenler arasındaki ilişkilerin ve terim büyüklüklerinin aynı olması durumunda (Durum I) farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında Klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri

Örneklem Hacmi	Uzaklık	Klasik ÇBÖ		SMACOF (TBA)	
		Ortalama	Std. Hata	Ortalama	Std. Hata
10	City-Block	0.1301	0.0026	0.0996	0.0026
	Öklid	0.0624	0.0028	0.0555	0.0021
	Std. Öklid	0.2480	0.0033	0.0859	0.0029
20	City-Block	0.1380	0.0022	0.1103	0.0020
	Öklid	0.0692	0.0022	0.0641	0.0017
	Std. Öklid	0.2746	0.0024	0.0998	0.0022
30	City-Block	0.1378	0.0017	0.1137	0.0016
	Öklid	0.0677	0.0013	0.0655	0.0012
	Std. Öklid	0.2807	0.0015	0.1037	0.0019
50	City-Block	0.1455	0.0015	0.1224	0.0013
	Öklid	0.0735	0.0013	0.0710	0.0011
	Std. Öklid	0.2869	0.0012	0.1125	0.0015

Durum I için elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimini gösteren grafik Şekil 5.5’de verilmiştir.



Şekil 5.5. Durum I’de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimi

Şekil 5.5. incelendiğinde başlangıç koordinatlarının Temel Bileşenler Analizinden elde edildiği SMACOF algoritması stress değerini en küçüklemeye Klasik ÇBÖ algoritmasından daha iyidir. Uzaklık fonksiyonları içerisinde her iki algoritmada da Öklid uzaklığı en iyi performans gösteren uzaklık ölçüsü olmuştur. Öklid uzaklığı kullanılması durumunda her iki algoritmadan elde edilen stress değerleri birbirlerine yakındır. Standartlaştırılmış Öklid uzaklığı Klasik ÇBÖ algoritmasında en kötü performans gösteren uzaklık ölçüsüdür. Tablo 5.6.'daki ortalama stress değerleri incelendiğinde örneklem hacmi artığında, beklendiği gibi stress değerleri de artmıştır.

Değişkenler arasındaki ilişkilerin aynı ve terim büyüklüklerinin farklı olması durumunda yani Durum II'de farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında Klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri Tablo 5.7.'de verilmiştir.

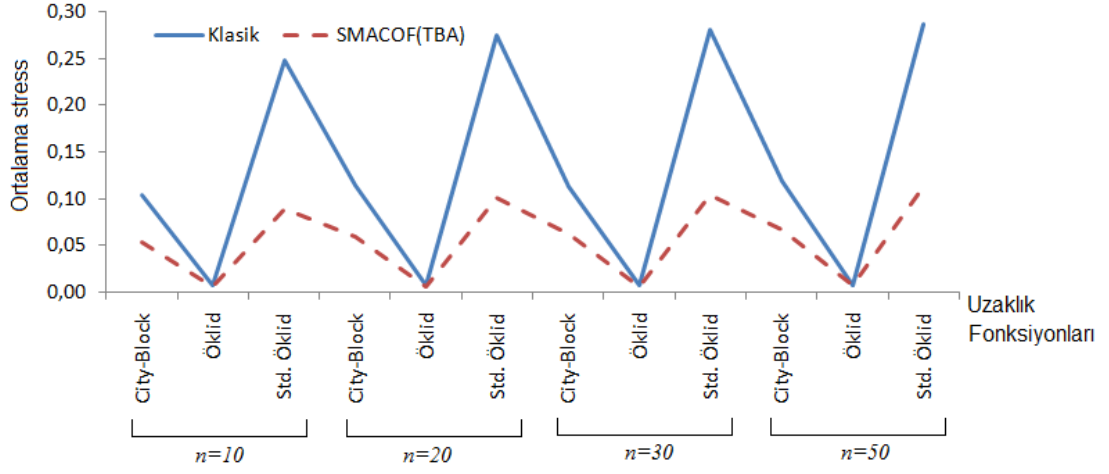
Tablo 5.7. Değişkenler arasındaki ilişkilerin aynı ve terim büyüklüklerinin farklı olması durumunda (Durum II) farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri

Örneklem Hacmi	Uzaklık	Klasik ÇBÖ		SMACOF (TBA)	
		Ortalama	Std. Hata	Ortalama	Std. Hata
10	City-Block	0.1044	0.0022	0.0537	0.0017
	Öklid	0.0073	0.0006	0.0061	0.0005
	Std. Öklid	0.2482	0.0033	0.0889	0.0029
20	City-Block	0.1143	0.0018	0.0599	0.0012
	Öklid	0.0077	0.0003	0.0066	0.0003
	Std. Öklid	0.2748	0.0024	0.1007	0.0022
30	City-Block	0.1132	0.0013	0.0628	0.0008
	Öklid	0.0074	0.0003	0.0066	0.0002
	Std. Öklid	0.2811	0.0015	0.1043	0.0019
50	City-Block	0.1182	0.0011	0.0671	0.0007
	Öklid	0.0077	0.0002	0.0072	0.0002
	Std. Öklid	0.2869	0.0012	0.1136	0.0014

Durum II için elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimini gösteren grafik Şekil 5.6' de verilmiştir.

5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ



Şekil 5.6. Durum II’de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimi

Değişkenler arasındaki ilişkilerin aynı ve terim büyüklüklerinin farklı olması durumunda da en başarılı algoritma başlangıç koordinatlarının Temel Bileşenler Analizinden belirlendiği SMACOF algoritmasıdır. Uzaklık fonksiyonları içerisinde Öklid uzaklığı her iki algoritmada da en başarılı uzaklık fonksiyonudur. Terim büyüklüklerinin farklı olması durumunda elde edilen ortalama stress değerleri terim büyüklüklerinin aynı olduğu duruma göre daha düşüktür. Beklendiği gibi standartlaştırılmış Öklid uzaklığı terim büyüklüklerinin aynı olup olmamasından etkilenmemiştir.

Değişkenler arasındaki ilişkilerin farklı ve terim büyüklüklerinin aynı olması durumunda yani Durum III’de farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında Klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri Tablo 5.8.’de verilmiştir.

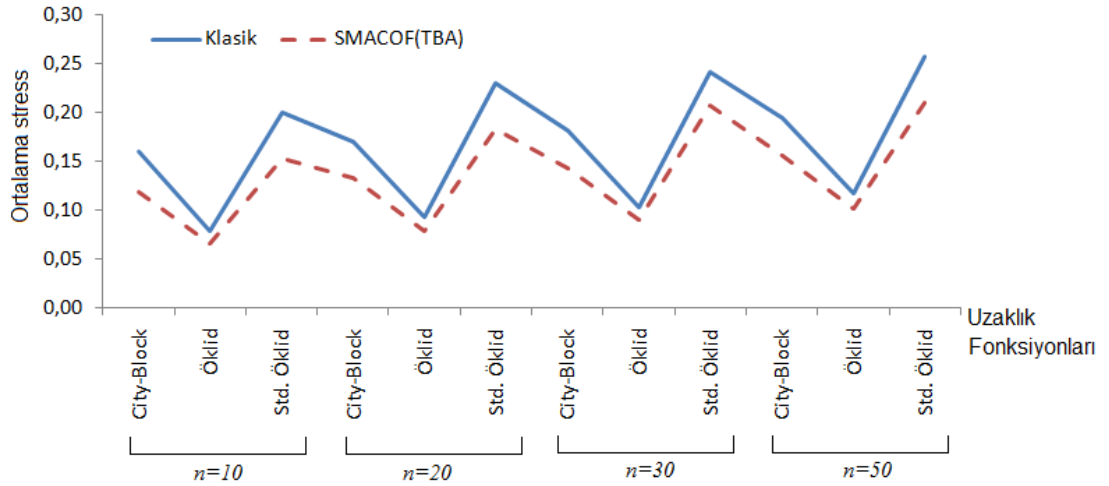
5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ
ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ

Tablo 5.8. Değişkenler arasındaki ilişkilerin farklı ve terim büyüklüklerinin aynı olması durumunda (Durum III) farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri

Örneklem Hacmi	Uzaklık	Klasik ÇBÖ		SMACOF (TBA)	
		Ortalama	Std. Hata	Ortalama	Std. Hata
10	City-Block	0.1596	0.0030	0.1184	0.0029
	Öklid	0.0783	0.0035	0.0661	0.0027
	Std. Öklid	0.1995	0.0033	0.1531	0.0056
20	City-Block	0.1701	0.0028	0.1322	0.0023
	Öklid	0.0925	0.0027	0.0783	0.0019
	Std. Öklid	0.2306	0.0024	0.1834	0.0050
30	City-Block	0.1815	0.0023	0.1424	0.0019
	Öklid	0.1027	0.0022	0.0896	0.0018
	Std. Öklid	0.2416	0.0024	0.2065	0.0072
50	City-Block	0.1942	0.0022	0.1564	0.0018
	Öklid	0.1172	0.0021	0.1008	0.0015
	Std. Öklid	0.2565	0.0019	0.2097	0.0041

Durum III için elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritma göre değişimini gösteren grafik Şekil 5.7' de verilmiştir.



Şekil 5.7. Durum III'de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritma göre değişimi

5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Selim GÜNDÜZ

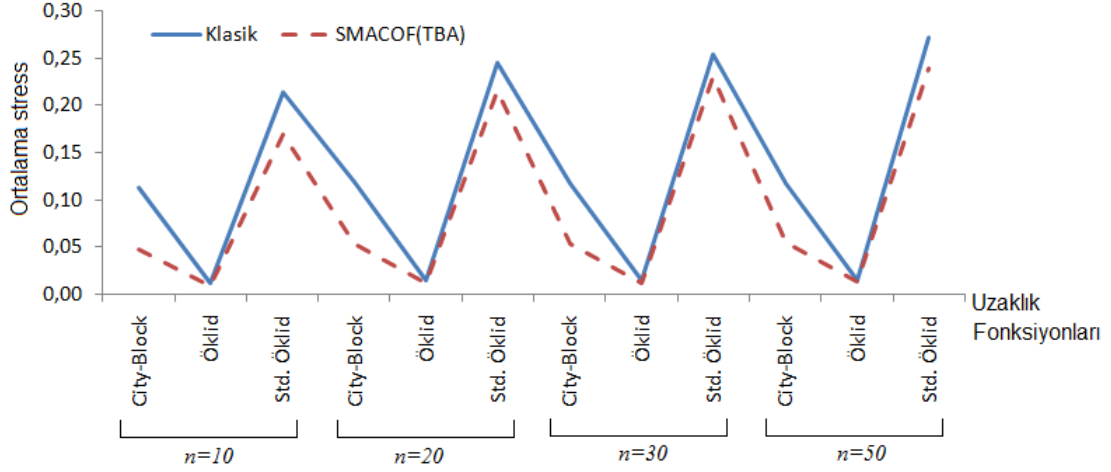
Değişkenler arasındaki ilişkinin farklı, terim büyüklüklerinin aynı olduğu durumda yine en başarılı algoritma SMACOF (TBA) ve en başarılı uzaklık fonksiyonu Öklid uzaklığı olmuştur. Durum III'de önceki durumlardan farklı olarak Standartlaştırılmış Öklid uzaklığı kullanıldığında her iki algoritmadan elde edilen stress değerleri arasındaki fark azalmıştır. Tablo 5.8.'deki stress değerleri incelendiğinde kovaryans matrisinin küresel olmaması yani değişkenler arasındaki ilişkilerin farklı olması durumunda elde edilen stress değerleri, terim büyüklüklerinin aynı, kovaryans matrisinin küresel olduğu Durum I'de elde edilen stress değerlerinden daha büyüktür.

Değişkenler arasındaki ilişkilerin ve terim büyüklüklerinin farklı olması durumunda yani Durum IV'de farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında Klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri Tablo 5.9.'da verilmiştir.

Tablo 5.9. Değişkenler arasındaki ilişkilerin ve terim büyüklüklerinin farklı olması durumunda (Durum IV) farklı örneklem hacmi ve farklı uzaklık fonksiyonlarında klasik ve SMACOF (TBA) algoritmalarında elde edilen stress değerlerine ait ortalama ve standart hata değerleri

Örneklem Hacmi	Uzaklık	Klasik ÇBÖ		SMACOF (TBA)	
		Ortalama	Std. Hata	Ortalama	Std. Hata
10	City-Block	0.1126	0.0013	0.0473	0.0013
	Öklid	0.0122	0.0007	0.0093	0.0005
	Std. Öklid	0.2137	0.0032	0.1692	0.0048
20	City-Block	0.1185	0.0018	0.0542	0.0011
	Öklid	0.0148	0.0005	0.0122	0.0005
	Std. Öklid	0.2447	0.0022	0.215	0.0083
30	City-Block	0.1178	0.0089	0.0534	0.0008
	Öklid	0.0147	0.0004	0.0125	0.0004
	Std. Öklid	0.2544	0.0023	0.2307	0.0091
50	City-Block	0.1168	0.0008	0.0544	0.0007
	Öklid	0.0147	0.0003	0.0132	0.0003
	Std. Öklid	0.2722	0.0021	0.2388	0.0080

Durum IV için elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimini gösteren grafik Şekil 5.8’ de verilmiştir.



Şekil 5.8. Durum IV’de elde edilen ortalama stress değerlerinin örneklem hacmi, uzaklık fonksiyonu ve ÇBÖ algoritmaya göre değişimi

Değişkenler arasındaki ilişkilerin ve terim büyüklüklerinin farklı olması durumunda da en başarılı algoritma SMACOF (TBA), en başarılı uzaklık fonksiyonu Öklid uzaklığı olmuştur. Kovaryans matrisinin küresel olmadığı durumlarda iki metrik ÇBÖ algoritma ile elde edilen stress değerleri arasındaki fark az iken, kovaryans matrisinin küresel olması durumunda her iki algoritma ile elde edilen stress değerleri arasındaki fark fazladır. Bu farklılık özellikle standartlaştırılmış Öklid uzaklığı kullanıldığında belirgin bir şekilde artmaktadır.

Farklı örneklem hacmi, farklı ortalama vektörleri ve farklı kovaryans matris yapılarında çok değişkenli normal dağılımdan üretilen rassal örneklemeler ile gerçekleştirilen simülasyon çalışması sonucunda en iyi yani en küçük stress değerleri Öklid uzaklığının kullanıldığı başlangıç koordinatlarının temel bileşenler analizinden elde edildiği karmaşık fonksiyonların optimizasyonu ile ölçekleme (SMACOF(TBA)) algoritması ile elde edilmiştir. İncelenen uzaklık fonksiyonlarından City-Block uzaklık fonksiyonu da Öklid uzaklığı gibi algoritmadan algoritmaya büyük farklılıklar göstermemiştir. Simülasyon sonunda,

5. UZAKLIK FONKSİYONLARININ STRESS DEĞERİ ÜZERİNDEKİ
ETKİSİNİN İNCELENMESİ Selim GÜNDÜZ

standartlaştırılmış Öklid uzaklığı en kötü performans gösteren uzaklık fonksiyonu olmuştur.

6. SONUÇLAR VE ÖNERİLER

Veri görselleştirme ve boyut indirgeme tekniklerinden biri olan çok boyutlu ölçekleme (ÇBÖ) analizi, birimler veya değişkenler arasındaki orijinal uzaklıkları daha düşük boyutta konfigürasyon koordinatlarında konumlandırarak çok değişkenli istatistiksel bir analizdir. ÇBÖ analizinde, özellikle çok boyutlu verilerde birimlerin birbirlerine göre konumları iki ya da üç boyutlu uzayda grafiksel olarak ortaya konmaya çalışılır. Çok boyutlu verilerde birimler arasındaki orijinal uzaklıklara karşı daha düşük boyutta konfigürasyon koordinatlarına göre elde edilen uzaklıklar konfigürasyon uzaklıkları olarak adlandırılır. Birimler arasındaki ilişkilerin daha düşük boyutlu uzayda gösteriminin değerlendirilmesinde genel olarak stress değeri kullanılır. Stress değeri, orijinal uzaklıklar ile konfigürasyon uzaklıkları arasındaki uyumun başka bir anlatımla uyumsuzluğun bir göstergesidir. Stress değerinin yüksek olması orijinal uzaklıklarla konfigürasyon uzaklıkları arasındaki kötü uyumu gösterirken, düşük stress değerleri orijinal uzaklıklarla konfigürasyon uzaklıkları arasındaki iyi uyumu gösterir.

ÇBÖ analizinde verilerin düşük boyutta temsil edilmesi için önerilen çok sayıda algoritma mevcuttur. Araştırmacıların çoğu ÇBÖ analizinde yer alan bu algoritmaları metrik ve metrik olmayan algoritmalar olarak iki alt sınıfa ayırırken bazı araştırmacılarda bu sınıfa yarı metrik algoritmalar diye üçüncü bir sınıfı eklemektedir. ÇBÖ analizi ile ilgili ilk çalışmalar metrik yaklaşımlardır. 'Princeton' veya 'Torgerson' yaklaşımı olarak bilinen bu yaklaşımla, klasik ÇBÖ'nin temelleri Princeton Üniversitesinde Messick ve Albelson (1956) ve Torgerson'nun (1952) içinde bulunduğu Psychometrik grubu tarafından atılmıştır. Bunun yanında Torgerson (1952) ilk defa ÇBÖ analizinin uygulanabilirliğini göstermiştir (Davison, 1983). Metrik olmayan yaklaşımdaki çalışmalar ise, on yıl aradan sonra Beil Telephone Laboratuvarında Shepard (1962a, 1962b) tarafından gerçekleştirilen çalışmadır. Bu yaklaşım 'Kruskal-Shepard' yaklaşımı olarak da bilinmektedir. Kruskal (1964), Shepard'ın (1962) bulduğu bu yaklaşıma kavramsal yenilikler getirmiştir. Metrik olmayan ÇBÖ analizi için Kruskal (1964b) tarafından kavramsal

ve hesaplanabilen yenilikler getirilmesi ile metrik olmayan ÇBÖ analizinin kullanımı da yaygınlaşmıştır.

Çalışmanın üçüncü bölümünde uzaklık fonksiyonları ele alınmıştır. Uzaklık ile ilgili temel tanım ve özellikler verildikten sonra nicel ve nitel veri durumunda kullanılan uzaklık fonksiyonları verilmiştir. Ayrıca hem nicel hem de nitel verilerin bulunduğu karma veri setlerinde kullanılan uzaklık fonksiyonları incelenmiş ve örnek bir uygulama ile ilgili uzaklıkların nasıl hesaplandığı gösterilmiştir.

Çalışmanın dördüncü bölümünde ÇBÖ algoritmaları metrik ve metrik olmayan algoritmalar başlığı altında ele alınmıştır. Metrik ÇBÖ algoritmalarından klasik ÇBÖ ve karmaşık fonksiyonların optimizasyonu ile ölçekleme (SMACOF) algoritmaları incelenirken, metrik olmayan ÇBÖ algoritmalarından Kruskal-Shepard algoritması incelenmiştir. İncelenen algoritmaların pratikte uygulamasını göstermek amacı ile iki uygulamaya yer verilmiştir.

Uygulamalardan ilki Türkiye'deki birinci düzey istatistiki bölgelerini oluşturan 12 bölgenin seçilen 10 sosyo-ekonomik değişkene göre ÇBÖ analizi ile karşılaştırılmasıdır. Elde edilen iki boyutlu düzlemdeki konumlandırma sonrası İstanbul, Batı Anadolu ve Doğu Marmara istatistiki bölgelerinin diğer bölgelerden belirgin bir şekilde ayrıldığı görülmektedir. Seçilen 10 sosyo-ekonomik değişken bakımından birbirine en benzer iki istatistiki bölge Ortadoğu Anadolu ile Güneydoğu Anadolu bölgeleridir. Seçilen sosyo-ekonomik değişkenler bakımından birbirine en benzemeyen iki bölge ise İstanbul ile Kuzeydoğu Anadolu bölgesidir. İstanbul bölgesinin en gelişmiş bölge olarak seçildiği karşılaştırmada bölgelerin gelişmişlik sıralaması bölgelerin İstanbul bölgesine olan konfigürasyon uzaklığı ile belirlenmiştir. İstanbul bölgesine gelişmişlik bakımından en benzer bölgeler sırasıyla Batı Anadolu, Doğu Marmara, Ege, Akdeniz, Batı Karadeniz, Orta Anadolu, Doğu Karadeniz, Güneydoğu Anadolu, Ortadoğu Anadolu ve Kuzeydoğu Anadolu istatistiki bölgeleridir.

Örnek ikinci uygulamada, Türkiye ve seçilen 36 Avrupa ülkesi göz önünde bulundurulmuş 12 sosyo-ekonomik gösterge bakımından ÇBÖ algoritması ile karşılaştırılmıştır. Türkiye ve 36 Avrupa ülkesinin gelişmişlik değerlendirilmesinde ÇBÖ algoritması ile elde edilen iki boyutlu gösterimi incelendiğinde Norveç,

İzlanda, Finlandiya ve İsveç'in diğer ülkelerden belirgin bir şekilde ayrıldığı görülmüştür. Gelişmişlik düzeyi en yüksek ülke olarak Norveç'e tespit edilmiştir. Gelişmişlik sıralaması belirlenirken diğer ülkelerin Norveç'e olan konfigürasyon uzaklıkları hesaplanmıştır. Bir ülkenin Norveç' olan uzaklığı ne kadar küçükse gelişmişlik düzeyinin o kadar yüksek olarak tanımlandığı uygulamada, Türkiye Norveç'e olan konfigürasyon uzaklığına göre gelişmişlik sıralamasında 28. sırada bulunmuştur.

Çalışmanın beşinci bölümünde, ÇBÖ analizinde farklı uzaklık fonksiyonlarının metrik çok boyutlu ölçekleme algoritmalarında stress değeri üzerindeki etkileri incelenmiştir. Uzaklık fonksiyonlarının stress değeri üzerindeki etkisi incelenirken özellikle sınıflama çalışmalarında yaygın bir kullanıma sahip olan farklı boyut ve yapılarıdaki Glass, Wine, İris, Letter, Liver, Vehicle, Yeast, Landsat, Shuttle, Diabet, Vowel, E.coli ve Atmosfer veri setleri kullanılmıştır. Tercih edilen boyut sayısının 2 olarak kabul edildiği çalışmada, Öklid, Standartlaştırılmış Öklid ve City-Block uzaklık fonksiyonlarının stress değeri üzerindeki etkileri Klasik, rassal başlangıçlı SMACOF ve Temel Bileşen başlangıçlı SMACOF algoritmalarında incelenmiştir. Karşılaştırma sonucunda tüm algoritmalarda stress değerini en iyileyen uzaklık fonksiyonu Öklid uzaklığı olmuştur. İncelenen algoritmalar içerisinde stress değerini en küçükleyen algoritma olarak başlangıç koordinatlarının temel bileşenler analizinden elde edilen bileşenlerin olduğu SMACOF algoritması olmuştur.

Uzaklık fonksiyonlarının metrik ÇBÖ algoritmalarında stress değeri üzerindeki etkisi ayrıca farklı koşulların göz önünde bulundurulduğu bir simülasyon çalışmasıyla da ortaya konmuştur. Simülasyon çalışmasında farklı örneklem hacimlerinde, farklı ortalama vektörleri ve farklı kovaryans matris yapılarında çok değişkenli normal dağılımından üretilen rassal örneklemeler kullanılmıştır. Simülasyon çalışmasında terim büyüklüklerinin farklı veya aynı olması durumunda, küresel veya küresel olmayan kovaryans matris yapılarında farklı uzaklık fonksiyonlarında stress değerindeki etkileri incelenmiştir.

Farklı örneklem hacmi, farklı ortalama vektörleri ve farklı kovaryans matris yapılarında çok değişkenli normal dağılımdan üretilen rassal örneklemeler ile gerçekleştirilen simülasyon çalışması sonucunda en iyi yani en küçük stress değerleri

Öklid uzaklığının kullanıldığı başlangıç koordinatlarının Temel Bileşenler Analizinden elde edildiği karmaşık fonksiyonların optimizasyonu ile ölçekleme algoritması ile elde edilmiştir. İncelenen uzaklık fonksiyonlarından City-Block uzaklık fonksiyonu da Öklid uzaklığı gibi algoritmadan algoritmaya büyük farklılıklar göstermemiştir. Simülasyon sonunda, Standartlaştırılmış Öklid uzaklığı en kötü performans gösteren uzaklık fonksiyonu olmuştur.

Kovaryans matrisinin küresel olmadığı durumlarda iki metrik ÇBÖ algoritma ile elde edilen stress değerleri arasındaki fark az iken, kovaryans matrisinin küresel olması durumunda her iki algoritma ile elde edilen stress değerleri arasındaki fark fazladır. Bu farklılık özellikle Standartlaştırılmış Öklid uzaklığı kullanıldığında belirgin bir şekilde artmaktadır.

Gerek gerçek, gerekse simülasyon verileri kullanılarak uzaklık fonksiyonlarının stress değeri üzerindeki etkisinin incelenmesi sonucunda elde edilen sonuçların nedeni yapılacak olan teorik çalışmalarla araştırılabilir.

KAYNAKLAR

- AGARWAL, S., WILLS, J., CAYTON, L., LANCKRIET, G., KRIEGMAN, D., and BELONGIE, S., 2007. Generalized non metric multi-dimensional scaling. Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics.
- AKACAN, P., 2001, "Pazarlama Arařtırmalarında Conjoint Analiz ve Çok Boyutlu Ölçekleme Analizi: Kozmetik Sektöründe Bir Uygulama", Gebze İleri Teknoloji Enstitüsü; Sosyal Bilimler Enstitüsü; İşletme Anabilim Dalı; Yüksek Lisans Tezi
- ALAN, S., 2008, Çok Boyutlu Ölçekleme Yöntemi Ve Bir Uygulama, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul
- BALCI, A., 2005, Sosyal Bilimlerde Arařtırma Yöntem, Teknik ve İlkeler, Kitap, 5.Baskı, Pagema Yay.
- BAYKUL, Y., 1999, İstatistik Metodlar ve Uygulamalar, Kitap, 3.Basım, Ankara
- BORG, I. and GROENEN, P.(2005), Modern Multidimensional Scaling: Theory and Application, Springer, USA.
- BRANDES, U. ve PICH, C., 2006. Eigensolver Methods for Progressive Multidimensional Scaling of Large Data. Proc. 14th Int'l Symp. Graph Drawing, 42-53.
- CANDAN F. B. 2003 "Büyük Ölçekli Perakendecilikte Konumlama ve Algılama Haritalarının Kullanılması", (Basılmamış Doktora Tezi, Atatürk Üniversitesi, Sosyal Bilimler Enstitüsü), Erzurum
- CARROLL J. Douglas and GREEN E. Paul. 1997, "Psychometric Methods in Marketing Research: Part II, Multidimensional Scaling", Journal of Marketing Research, Volume: 34
- CHA, J., KİM, S. and LEE, Y. 2009, Application Of Multidimensional Scaling For Marketing-Mix Modification: A Case Study On Mobile Phone Category, Expert Systems With Applications 36 (2009) 4884–4890
- CHEUNG, K.W. AND SO,H.C 2005,. Ieee Transactions On Signal Processing, Vol. 53, No. 2, February 2005

- COSTA, J. A., PATWARI, N. ve HERO, III A. O., 2005. Distributed Weighted-Multidimensional Scaling for Node Localization in Sensor Networks. *ACM Trans.on Sensor Networks*, 2 (1), 39-64.
- COX, T.F. and COX, M.A.A.(2001), *Multidimensional Scaling*, Second Edition, Chapman&Hall/CRC, USA.
- ÇALIŞ A. Figen 1995, "Tüketici Marka Algılamaları ve Satın Alma Davranışı: Binek Otomobil Örneği", (Basılmamış Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü), İzmir, 1995
- ÇİLAN, Ç.A. ve DEMİRHAN, A. 2002, Türkiye'nin İllere Göre Sosyo-Ekonomik Yapısının Çok Boyutlu Ölçekleme Tekniği Ve Kümeleme Analizi İle İncelenmesi, *Yönetim*, Sayı 42, Sayfa 39-50.
- DAVİSON, M. L. (1983). *Multidimensional Scaling*. New York, New York: John Wiley & Sons. 1-5p.
- DE LEEUW, J., 1977, Applications of convex analysis multidimensional scaling, In Barra, J.R., Brodeau, F., Romier, G. and van Cutsen, B. (eds.), *Recent Developments in Statistics*, Amsterdam: North Holland, 133-145.
- DEZA, M. M. and DEZA, E., 2006. *Dictionary of Distances*. Elsevier. 391p.
- DICKES, P., VALENTOVA, M. And BORSENBERGER, M. 2011, A Multidimensional Assessment Of Social Cohesion In 47 European Countries, *Fonds National Recherche Luksembourg*, Working Paper No: 2011-07
- DİNÇER , B., ÖZASLAN, M., KAVASOĞLU, T., 2003, İllerin ve Bölgelerin Sosyo-ekonomik Gelişmişlik Sıralaması Araştırması, <http://ekutup.9>
- DİNG, C. S. and DAVİSON, M. L. 2010, Assessing Fit and Dimensionality in Least Squares Metric Multidimensional Scaling Using Akaike's Information Criterion, *Educational and Psychological Measurement* 70: 199
- DONG, Y. and DİNG, C. 2011, Adolescent Risk Behaviors: Studying Typical And Atypical Individuals Via Multidimensional Scaling Profile Analysis, *Journal Of Adolescence*, 10.1016/j.adolescence.2011.02.006
- DUDA, R. O., HART, P.E. and STORK, D. G., 2001. *Pattern Classification*. John Wiley & Sons, INC. New York. 654p.

- DURA, C., ATİK, H., TÜRKER, O. 2004, Beşeri Sermaye Açısından Türkiye'nin Avrupa Birliği Karşısındaki Kalkınma Seviyesi, 3. Bilgi Ekonomi ve Yönetim Kongresi, Osmangazi Üniversitesi, İktisadi ve İdari Bilimler Fakültesi
- EVERETT, B.S., LANDAU, S. and LEESE, M., 2001. Cluster Analysis. Oxford University Press Inc., New York, NY. 237p.
- ERİŞOĞLU, M. 2011, Uzaklık Ölçülerinin Kümeleme Analizine Olan Etkilerinin İncelenmesi Ve Geliştirilmesi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Ana Bilim Dalı, Basılmamış Doktora Tezi, Adana, 199 s.
- FINDIKKAYA, A. 1995 “Çok Boyutlu Ölçekleme Analizi ve Bir Uygulama Denemesi” (Basılmamış Yüksek Lisans Tezi, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü), Bursa
- FİLİZ ,Z. ve ÇEMREK, F., 2005, “Avrupa Birliğine Üye Ülkeler ile Türkiye'nin Karşılaştırılması”. İstanbul Üniversitesi VII. Ulusal Ekonometri ve İstatistik Sempozyumu.
- GOWER, J. C., 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 27: 857-872.
- GROENEN, P.J., BORG ,I., 1997, Title Modern Multidimensional Scaling : Theory and Applications , Kitap , Springer , New York
- HALL, J. L. 2001, Application of multidimensional scaling to subjective evaluation of coded speech Multimedia Communications Research Laboratory, Bell Labs, Lucent Technologies, Murray Hill, New Jersey 07974-0636
- HARDLE , W.,SİMAR L., 2007, Applied Multivariate Statistical Analysis , E-Kitap, Springer <http://www.xplore-stat.de/ebooks/ebooks.html>
<http://www.worldbank.org/>
- JI, X. ve ZHA, H., 2004. Sensor Positioning in Wireless Ad-hoc Sensor Networks Using Multidimensional Scaling. In Proc. IEEE Infocom 2004, 2652-2661.
- JOSEPH F., RALPH A., RONALD T. And WILLİAM B., 1998, Multivariate Data Analysis, Fifth Edition, Prentice Hall, New Jersey

- KAÇAR, O., AZKAN, N. 2005. Türkiye’de Bulunan Bazı Hypericum Türleri Üzerinde Sayısal Taksonomik Bir Çalışma, F.Ü. Fen Ve Mühendislik Bilimleri Dergisi, 17(2), 424-434.
- KAUFMAN, L., and ROUSSEEUW, P. J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons, Inc. 342p.
- KRUSKAL, J.B. (1964a), Multidimensional scaling by optimizing goodnessof-fit to a nonmetric hypothesis, Psychometrika, 29, 1-27.
- KRUSKAL, J.B. (1964b), Nonmetric multidimensional scaling: A numerical method, Psychometrika, 29, 115- 129.
- KRUSKAL, J.B.,Wish ,M. , Multidimensional Scaling (12. Ed.),Kitap , Sage Publications , Bell
- KURT ,G., 1992, “Çok Değişkenli İstatistiksel Analiz Tekniklerinden Çok Boyutlu Ölçekleme ve Bir Uygulama”, (Basılmamış Doktora Tezi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü), İstanbul
- KURTULUŞ, K., 2007, Pazarlama Araştırmaları, Kitap, 7.Basım, Literatür Yay.
- LALA, V. 2009, A Multidimensional Scale For Measuring Country Image, Journal Of International Consumer Marketing, 21:51–66, 2009 Copyright _C Taylor & Francis Group, LLC ISSN: 0896-1530 Print / 1528-7068
- LEUNG, P. L. and LAU K. N., 2004, Estimating The City-Block Two-Dimensional Scaling Model With Simulated Annealing, European Journal Of Operational Research 158 518–524
- MACHADO, J.T., DUARTE, F.B. and DUARTE, G.M. 2010, Analysis Of Financial Data Series Using Fractional Fourier Transform And Multidimensional Scaling, Springer Science-Business Media B.V.
- MACHADO, J. T., DUARTE, F. B. and DUARTE G. M. 2011, Analysis Of Stock Market Indices Through Multidimensional Scaling, Commun Nonlinear Sci Numer Simulat <http://dx.doi.org/10.1016/j.cnsns.2011.04.027>
- MALHOTRA, K. Naresh, 2004, Marketing Research, An Applied Orientation, 4th edition, Prentice Hall, 2004.

- MARDÍA, K. V., KENT, J. T. and BİBBY, J. M., 1979, *Multivariate Analysis*. London, Academic Press.
- MARTÍNEZ, W.L. and MARTÍNEZ, A.R., 2005, *Exploratory Data Analysis with Matlab*, Chapman&Hall/CRC, USA.
- MORRISON, D., 1967. Measurement problems in cluster analysis. *Management Science (Series B, Managerial)*, 13(12): B775–B780.
- ONG, C. S., HUANG, J. J., TZENG, G.H. 2004, *Multidimensional Data In Multidimensional Scaling Using The Analytic Network Process*, *Pattern Recognition Letters*
- ÖZDAMAR, K., 2004, *Paket Programlar ile İstatistiksel Veri Analizi 2 (Çok Değişkenli Analizler)*, Kaan Kitabevi, Eskişehir
- PAK, V. 2007, *Akdeniz Çanağı Ve Avrupa Birliği Ülkeleri İçinde Çok Boyutlu Ölçekleme Analizi İle Türkiye Turizminin Yeri (Beykent Üniversitesi Sosyal Bilimler Enstitüsü ,Yüksek Lisans Tezi)*, İstanbul
- RENCHER, A.C., 2002, *Methods of Multivariate Analysis, Second Edition*, John Wiley&Sons,USA.
- ROSMAN, G., BRONSTEİN, A.M., BRONSTEİN, M. M., SİDİ, A. and KİMMEL, R. 2008, *Fast Multidimensional Scaling using Vector Extrapolation*, Technion-Computer Science Department Technical Report CIS
- ROUZIER, R., PEROU, C. M., SYMMANS, W. F., et al. 2005, *Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy*, <http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-04-2421>
- SHABBİR, H.A., REAST, J.And PALİHAWADANA, D. 2009, *25 Years Of Psychology & Marketing: A Multidimensional Review*, *Psychology & Marketing*, Vol. 26(12): 1031–1065
- SHEPARD, R.N. (1962a), *The analysis of proximities: multidimensional scaling with an unknown distance function I*, *Psychometrika*, 27, 125-140.
- SHEPARD, R.N. (1962b), *The analysis of proximities: multidimensional scaling with an unknown distance function II*, *Psychometrika*, 27, 219-246.

- SIĞIRLI, D., EDİZ, B., CANGÜR, Ş., ERCAN, İ., KAN, İ., 2006. Türkiye ve Avrupa Birliği'ne üye ülkelerin sağlık düzeyi ölçütlerinin çok boyutlu ölçekleme analizi ile incelenmesi, İnönü Üniversitesi Tıp Dergisi,81-85s.
- SERVİ, T. 2009, Çok Değişkenli Karma Dağılım Modeline Dayalı Kümeleme Analizi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Basılmış Doktora Tezi, Adana
- ŞAHİN, A., ATIŞ, E. VE MİRAN, B. 2008, Daha Etkin Tarım-Çevre Politikaları için Homojen Alanların Belirlenmesi: Ege Bölgesi Örneği, Ekoloji, 17, 67, 15-23 2008
- ŞİMŞEK, D. 2006, Kümeleme Analizi, Çok Boyutlu Ölçekleme Analizi, Doğrulayıcı ve Açıklayıcı Faktör Analizi ile Elde Edilen Yapı Geçerliği Kanıtların Karşılaştırılması, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Basılmış Yüksek Lisans Tezi, Ankara
- TETİK, N. 2007, Çok Değişkenli İstatistik Tekniklerle Ab Pazarının Bölümlendirilmesi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı İstatistik Bilim Dalı, Basılmamış Yüksek Lisans Tezi, İstanbul
- TAKANE, Y., YOUNG, F.W. and DE LEEUW, J.(1977), Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features, Psychometrika, 42, 7-67.
- TATLIDİL, H., 2002. Uygulamalı Çok Değişkenli İstatistiksel Analiz, Ziraat Tekniğinin Uygulanması, Kocaeli Üniversitesi, Sosyal Bilimler Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, Kocaeli
- TORGERSON, W.S. (1952), Multidimensional scaling: 1. Theory and method, Psychometrika, 17, 401-419.
- TÜZÜNTÜRK, S.2009, Çok Boyutlu Ölçekleme Analizi: Suç İstatistikleri Üzerine Bir Uygulama, Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi Cilt XXVIII, Sayı 2, 2009, s. 71-91
- USTAAHMETOĞLU, E. 2004, Marka Konumlandırma Çok Boyutlu Ölçekleme Tekniğinin Uygulanması, Kocaeli Üniversitesi, Sosyal Bilimler Enstitüsü, Yayınlanmış Yüksek Lisans Tezi, Kocaeli

- VELLEMAN, P.F. ,WILKINSON, L., 1993, Nominal, Ordinal, Interval, and Ratio Typologies are Misleading , Makale (Revised), The American Statistician
- VERA, J. F., MACÍAS, R. and HEISER, W. J. 2009, A Latent Class Multidimensional Scaling Model For Two-Way One-Mode Continuous Rating Dissimilarity Data, PSYCHOMETRIKA—VOL. 74, NO. 2, 297–315
- WEBB, A. R., 2002. Statistical Pattern Recognition. Second Edition. John &Wiley Sons. Chichester, England, 496p.
- WOJCIECH, B., 2001, Proximity Visualisation of Abstract Data , E-Kitap, www.pavis.org
- YELOĞLU ve ark., 2005, Avrupa Birliği ve Resmi İstatistikler: Ülkelerin Karşılaştırılması, 14üncü İstatistik Araştırma Sempozyumu, Ankara 5-6 Mayıs 2005, Bildiriler Kitabı, 83-97s., Ankara.
- YENİDOĞAN , T.G., 2008, Pazarlama Araştırmalarında Çok Boyutlu Ölçekleme Analizi: Üniversite Öğrencilerinin Marka Algısı Üzerine Bir Araştırma. Akdeniz İ.İ.B.F. Dergisi. 15, 138-169.
- YİĞİT, E. (2007), Çok Boyutlu Ölçekleme Yöntemlerinin İncelenmesi ve Bir Uygulama, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmamış Yüksek Lisans Tezi, Samsun.
- YOUNG, G. and HOUSEHOLDER, A.S. (1938), Discussion of a set of points in terms of their mutual distances, Psychometrika, 3, 19-22.
- YÜKSELEN, C., 2003, Pazarlama Araştırmaları, Kitap,2 .Baskı, Detay Yay. Ankara

ÖZGEÇMİŞ

17.07.1986 yılında Elazığ'da doğdu. İlk, orta ve lise öğrenimini Elazığ'da tamamladı. 2004 yılında başladığı Çukurova Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü'nden 2008 yılında mezun olup aynı yıl içerisinde Çukurova Üniversitesi İstatistik Bölümü'nde yüksek lisansa başladı. 2009 yılında İstatistik Bölümü'nde Araştırma Görevlisi oldu ve halen görevine devam ediyor.