

**T.C.
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI**

DOKTORA TEZİ

OPTİMAL ÖLÇEKLEME TEKNİKLERİ VE BİR UYGULAMA

**Engin KARAMAN
2502150230**

**TEZ DANIŞMANI
PROF. DR. ÇİĞDEM ARICIGİL ÇİLAN**

İSTANBUL-2019



T.C.
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

DOKTORA
TEZ ONAYI



ÖĞRENCİNİN;

Adı ve Soyadı : ENGİN KARAMAN Numarası : 2502150230
Anabilim Dalı / Anasanat Dalı / Programı : SAYISAL YÖNTEMLER Danışmanı : PROF.DR.ÇİĞDEM ARICIGİL ÇILAN
Tez Savunma Tarihi : 13.12.2019 Saati : 11:30
Tez Başlığı : OPTİMAL ÖLÇEKLEME TEKNİKLERİ VE BİR UYGULAMA.

TEZ SAVUNMA SINAVI, İÜ Lisansüstü Eğitim-Öğretim Yönetmeliği'nin 50. Maddesi uyarınca yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin **KABULÜNE** OYBİRLİĞİ / OYÇOKLUĞUYLA karar verilmiştir.

JÜRİ ÜYESİ	İMZA	KANAATİ (KABUL / RED / DÜZELTME)
PROF.DR.ÇİĞDEM ARICIGİL ÇILAN		KABUL
PROF.DR.SENİYE ÜMIT FIRAT		Kabul
PROF.DR.HALİM KAZAN		KABUL
PROF.DR.MELİH İNAL		KABUL
DOÇ.DR.BİLGE ACAR BOLAT		KABUL

YEDEK JÜRİ ÜYESİ	İMZA	KANAATİ (KABUL / RED / DÜZELTME)
PROF.DR.UMMAN TUĞBA GÜRSOY		—
DOÇ.DR.AYSEN Ş. KANDEMİR		—

ÖZ

OPTİMAL ÖLÇEKLEME TEKNİKLERİ VE BİR UYGULAMA

Engin Karaman

Sosyal Bilimler alanında yapılan araştırmalarda veriler daha çok kategorik değişkenlerden oluşur. Ancak bu tür verilerle çalışmak bir çok varsayımın sağlanmasında soruna neden olmaktadır. Kategorik verilere dönüşüm uygulanarak bu soruna çözüm geliştirmek bir çok araştırmacı için önem arz eder.

Bu çözümlerden biri de Optimal Ölçekleme ya da diğer adıyla Gifi Yöntemi'dir. Optimal Ölçekleme ile kategorik verilerin ölçek özellikleri bozulmadan hem sürekli veriye dönüşümü sağlanır hem de bilgi kaybı engellenir.

Bu tezde; cep telefonu kullanımının, uyku düzeni üzerine etkisinin olup olmadığını, var ise hangi değişkenlerce açıklandığı Regresyon Analizi ile belirlendi. Araştırma da toplam 491 kişiye uygulanan anket verileri analiz edildi. Veri seti tamamı kategorik 1 bağımlı, 25 bağımsız değişkenden oluşmaktadır. Bu kadar çok bağımsız değişkene aynı anda Regresyon Analizi uygulamak gerek verimlilik gerekse de öngörü geliştirme açısından zorluk yaratacaktır. Araştırmamızın temel savı, en az değişkenle en uygun modeli saptamaktır. Bunun için kullanılan yöntem, bir çok opsiyonel seçeneği içinde barındırarak araştırmacıya çözüm arayan bir sezgisel yöntem olan Genetik Algoritmalar. Böylece belirli bir amaç fonksiyonu doğrultusunda en iyi çözümü bularak optimal sonuca ulaşılabilecektir. Bunun için kullanılacak olan amaç fonksiyonu olarak, Hamparsum Bozdoğan (1988) tarafından geliştirilen model yalınlığı ve uyum iyiliğinin yanısıra modelin karmaşıklığını da dikkate alan ICOMP (Information Complexity) ve onun özel hali ICOMP (IFIM) belirlenmiştir.

Sonuç olarak Algoritma için gerekli kısıt ve seçimler yapılarak, 25 bağımsız değişken içerisinde bağımlı değişkeni açıklayabilecek ve amaç fonksiyonuna en uygun alt seti verecek regresyon modeli tespit edilmiş ve model oluşturulmuştur.

Anahtar Kelimeler: Optimal Ölçekleme, Gifi Yöntemi, Uyku Düzeni, ICOMP (IFIM), Genetik Algoritma.

ABSTRACT

OPTIMAL SCALING TECHNIQUES AND AN APPLICATION

Engin Karaman

In the studies conducted in the field of Social Sciences, the data consists of categorical variables. However, working with such data causes problems in providing many assumptions. Developing solutions to this problem by applying transformation to categorical data is important for many researchers. One of these solutions is Optimal Scaling or Gif Method. With Optimal Scaling, the categorical data is transformed into continuous data without disturbing the scale properties and information loss is prevented.

In this thesis; Regression analysis was used to determine whether mobile phone use has an effect on sleep patterns, and if so, which variables are explained. In the research, the survey data applied to a total of 491 people were analyzed. The data set consists of 25 independent variables, all of which are categorically dependent. Applying Regression Analysis to so many independent variables at the same time will create difficulties in terms of both productivity and forecasting. The main argument of our research is to determine the most appropriate model with the least variable. Genetic Algorithms, which is an intuitive method that seeks solutions to the researcher by incorporating many optional options, are used for this purpose. Thus, an optimal result will be achieved by finding the best solution for a specific purpose function. For this purpose, ICOMP (Information Complexity) and ICOMP (IFIM), which takes into account the complexity of the model as well as the model simplicity and goodness of fit developed by Hamparsum Bozdoğan (1988), have been determined as the objective function to be used for this purpose.

As a result, the necessary constraints and choices were made for the algorithm, and the regression model that could explain the dependent variable among the 25 independent variables and gave the most appropriate subset to the objective function was determined and the model was created

Key Words: Optimal Scaling, Gifi Method, Sleep Order, ICOMP (IFIM), Genetic Algorithm.

ÖNSÖZ

Bu tezin konusunun belirlenmesi, tasarlanması ve yazılması süresince çalışmamı sahiplenerek titizlikle takip eden danışmanım Prof. Dr. Çiğdem ARICIGİL ÇİLAN hocama;

Tez izleme sürecinde yapıcı eleştirilerini ve yönlendirmelerini esirgemeyen değerli hocalarım Prof. Dr. Seniye Ümit FIRAT ve Doç. Dr. Bilge ACAR BOLAT'a

Doktora ders döneminde bana çok değer katan İstanbul Üniversitesi Sayısal Yöntemler programının değerli öğretim üyelerine;

Tezin yazılım-kodlama bölümlerinde ve teorik alt yapısının geliştirilmesinde rehberliğini ve desteğini esirgemeyen Prof. Dr. Hamparsum BOZDOĞAN hocama;

Akademik dünyada varoluşumu borçlu olduğum ve emeğini üzerimden eksik etmeyen hocam Prof. Dr. Halim KAZAN'a;

Dostluğunun yanı sıra ihtiyaç duyduğum her anda yardımına koşan ve desteğini hiçbir zaman esirgemeyen Prof. Dr. Hasan LATİF'e;

Zorlu hayat koşulları altında bu konuma gelmemde en büyük payı olan Emekçi babam Eyüphan KARAMAN'a, cefakâr ve fedakâr annem Gülbin KARAMAN'a ve sanatçı kardeşim Aydın KARAMAN'a;

Bu zorlu süreçte her eksikimi telafi eden ve tüm boşlukları dolduran can yoldaşım Canan KARAMAN'a ve iki oğluma;

Son olarak, bugünlere gelmemde emeği geçen ve ismini yazamadığım tüm değerli insanlara;

Saygı, sevgi ve sonsuz teşekkürlerimi sunarım.

Engin KARAMAN

2019

İÇİNDEKİLER

ÖZ.....	iii
ABSTRACT.....	iv
ÖNSÖZ.....	v
TABLolar LİSTESİ.....	ix
ŞEKİLLER LİSTESİ.....	x
KISALTMALAR LİSTESİ.....	xi
GİRİŞ.....	1

BİRİNCİ BÖLÜM

İSTATİSTİK TAHMİN TEKNİKLERİ

1.1. EN ÇOK OLABİLİRLİK TAHMİNİ (MAXIMUM LIKELIHOOD ESTIMATION)	4
1.1.1. En Çok Olabilirlik Tahmini için Basit Bir Uygulama.....	6
1.2. EN KÜÇÜK KARELER YÖNTEMİ.....	9
1.3. TOPLAM EN KÜÇÜK KARELER TEKNİĞİ.....	11
1.4. YANLI (RİDGE) EN KÜÇÜK KARELER TEKNİĞİ.....	12
1.5. KİSMİ EN KÜÇÜK KARELER TEKNİĞİ.....	13
1.6. DALGALI (AĞIRLIKLİ) EN KÜÇÜK KARELER TEKNİĞİ.....	13
1.6.1. Kayıp Fonksiyon	14
1.6.2. ALS Algoritması ve Aşamaları.....	14
1.6.2.1. Normalleştirilmiş Skor Algoritması.....	15
1.6.2.2. Normalleştirilmiş Ağırlıklar Algoritması.....	16
1.6.3. Dalgali En Küçük Kareler’de Sayısallaştırma işlemi Üzerine Bir Uygulama.....	16

İKİNCİ BÖLÜM

OPTİMAL ÖLÇEKLEME TEKNİKLERİ VE TÜRLERİ

2.1. ÇOKLU UYGUNLUK YA DA HOMOJENLİK ANALİZİ (HOMALS).....	22
2.2. DOĞRUSAL OLMAYAN TEMEL BİLEŞENLER ANALİZİ.....	38
2.3. KATEGORİK REGRESYON ANALİZİ.....	40
2.4. DOĞRUSAL OLMAYAN KANONİK KORELASYON ANALİZİ (OVERALS)	46

ÜÇÜNCÜ BÖLÜM

MODEL SEÇİMİNDE BİLGİ KRİTERLERİ

3.1. ÇOKLU BELİRLİLİK KATSAYISI (R^2) KRİTERİ.....	51
3.2. KULLBACK- LEİBLER BİLGİ KRİTERİ (K-L UZAKLIĞI).....	52
3.3. AKAIKE BİLGİ KRİTERİ (AIC)	53
3.4. BAYES BİLGİ KRİTERİ (BIC)	53
3.5. BİLGİ KARMAŞIKLIĞI KRİTERİ (ICOMP).....	54
3.5.1. Fisher Bilgi Matrisi.....	56

DÖRDÜNCÜ BÖLÜM

GENETİK ALGORİTMA

4.1.GENETİK ALGORİTMA NEDİR?	58
4.1.1. Genetik Algoritmanın Adımları.....	59
4.1.2. Genetik Algoritmada Kullanılan Kavramlar.....	61
4.2.REGRESYON MODEL SEÇİMİNDE GENETİK ALGORİTMA KULLANIMI.....	64

BEŞİNCİ BÖLÜM

UYGULAMA

5.1.ARAŞTIRMANIN YÖNTEMİ.....	68
5.1.1. Uyku Nedir ?.....	68
5.1.2. Ölçek Oluşturma Süreci	71
5.1.3. Metodoloji	73
5.1.3.1. Demografik Özellikler Bilgileri.....	73
5.1.3.2. Veri Yapısı.....	75
5.1.3.3. Verilerin Sayısallaştırılması.....	77
5.1.3.4. Sayısallaştırılmış Veriler.	79
5.2.ANALİZ	82
SONUÇ VE TARTIŞMA	85
KAYNAKLAR.....	87
EKLER.....	104
ÖZGEÇMİŞ.....	106

TABLÖLAR LİSTESİ

Tablo 1: Dağılım Türlerine Göre En Çok Olabilirlik Tahmin Edicileri.....	5
Tablo 2: $L(\theta)$, En Çok Olabilirlik Tahmincisi.....	8
Tablo 3: Heterojen H Matrisinin Normalleştirilmiş Skor Algoritması İterasyon Özeti.....	18
Tablo 4: Heterojen H Matrisinin Normalleştirilmiş Skor Algoritması Sonuç İterasyon Özeti.....	20
Tablo 5: Kontenjans Tablosu (I x J Boyutlu)	23
Tablo 6: Satır Profilleri (r) Tablosu.....	24
Tablo 7: : Sütun Profilleri (c) Tablosu.....	25
Tablo 8: H Veri Matrisi (Kategorik hali)	28
Tablo 9: H Veri Matrisi (Kodlanmış hali)	28
Tablo 10: Gösterge Matrisi'nin (G_J)	29
Tablo 11: C İkili Marjinaler Matrisi.....	30
Tablo 12: D Tekli Marjinaler Matrisi.....	30
Tablo 13: Nesne Skorları (x)	31
Tablo 14: Kategori Sayısallaştırmaları (y)	31
Tablo 15: Q Veri Matrisi (Sayısallaştırılmış hali)	32
Tablo 16 : Değişkenler	76
Tablo 17: Ham Veri Matrisi (H)	78
Tablo 18: Sayısallaştırılmış Veri Matrisi (Q)	79
Tablo 19: GA sonucunda elde edilen en iyi Regresyon modelleri ve ICOMP(IFIM) Skorları.....	81
Tablo 20: Regresyon Modeli ve Değişkenler.....	85

ŞEKİLLER LİSTESİ

Şekil 1: Farklı Ana Kütle Dağılımları.....	7
Şekil 2: Serpilme Diyagramı.....	10
Şekil 3: Kategoriler Arasındaki İlişkiler Grafiği.....	35
Şekil 4: Nesneler Arasındaki İlişkiler Grafiği.....	36
Şekil 5: Ayrım Ölçüsü Grafiği	36
Şekil 6: Genetik Algoritma Akış Diyagramı.....	60
Şekil 7: Gen Yapısı Görseli.....	61
Şekil 8: Çaprazlama (Crossing-over) Süreci.....	63
Şekil 9: Cinsiyet Bilgileri.....	74
Şekil 10: Katılımcıların Mezuniyet Durumu.....	74
Şekil 11: Cep Telefonu Kullanımı, Uyku Düzeninizi Olumsuz Etkiler Mi?.....	75
Şekil 12: Tez Uygulama Akış Şeması.....	84

KISALTMALAR LİSTESİ

AIC	: Akaike Bilgi Kriteri
ALS	: Alternating Least Squares of Optimal
BIC	: Bayes Bilgi Kriteri
CATREG	: Kategorik Regresyon
ICOMP	: Information Complexity
DEKK	: Dalgalı En Küçük Kareler
DOKKA	: Doğrusal Olmayan Kanonik Korelasyon Analizi
DOTBA	: Doğrusal Olmayan Temel Bileşenler Analizi
EÇOB	: En Çok Olabilirlik Tahmin Yöntemi
EKK	: En Küçük Kareler
GA	: Genetik Algoritmanın
HOMALS	: Homojenlik Analizi
ICOMP	: Information Complexity
IFIM	: Ters Fisher Matrisi
KEKK	: Kısmi En Küçük Kareler
K-L BK	: Kullback- Leibler Bilgi Kriteri
MLE:	: Maximum Likelihood Estimation
NIPALS	: Lineer Olmayan Yinelemeli Kısmi En Küçük Kareler Regresyonu
NREM	: Non- Rapid Eye Movement Period
REM	: Rapid Eye Movement Period
WMON	: Monotonik Regresyon Yöntemi

GİRİŞ

Politika belirleme, öngörü oluşturma ve analiz kabiliyeti geliştirmede bilimsel teknik ve uygulamaların kullanılmadığı alanlar artık modern toplumlarda pek bir anlam ifade etmemektedir. Hurafe ve dogmadan sıyrılmanın çabası, kendini bilimsel ölçütlerle tanımlayan teknoloji ve yöntemleri insanların önüne sunmuştur. Bu büyük evrende İstatistik dünyası kendi sınırlarını da aşarak her bilim dalıyla bütünleşik hareket eder. Neredeyse analiz ve araştırma konularının işlendiği bütün disiplinlerde İstatistik olmazsa olmaz bir parçadır ve bu dinamik dünya kendini günden güne güncelleştirmekte ve sınırlarını daha da geliştirmektedir.

Bununla birlikte, İstatistiki yöntemlerin güvenilirliğini ve geçerliliğini sağlamak en önemli kıstasların başında gelmektedir. Uygulanan yöntem ne olursa olsun, o yönteme haiz belirli varsayımlar sağlanmadan harekete geçmek sonuçların doğruluğu konusunda soru işaretlerine yol açmaktadır. Özellikle Sosyal Bilimler alanında yapılan araştırmalarda veriler daha çok kategorik (nitel ve sınırlı sayıda şıkkı olan kesikli) değişkenlerden oluşur. Maalesef bu tür verilerle çalışmak birçok varsayımın sağlanmasında soruna neden olmaktadır. Örneğin kategorik verilere Regresyon Analizi uygulayabilmek için gerekli olan varsayımların ihlal edileceği aşikardır.

Akademik çalışmalar külliyatı bu tür sorunların çözümü adına sayısız yöntem ve denemelerle doludur. En yaygın şekliyle verilere çarpımsal dönüşüm uygulayarak düzgünleştirme buna örnektir. Bunlar içerisinde en sık kullanılanları; arcsin veya logaritmik dönüşüm, ters çevirme, karekök ya da kare alma, logit, probit vb.. işlemleridir (Tabachnick ve Fidell, 2013). Bununla birlikte çözüm olarak literatürde giderek yaygınlaşan yöntemlerden biri olarak Optimal Ölçekleme ya da diğer adıyla Gifi Yöntemi bu soruna çözüm olarak önerilmektedir.

Optimal Ölçekleme Teknikleri, kategorik verilerin ölçek özelliklerini bozmadan onları hem sürekli veriye dönüştürür hem de bilgi kaybını engelleyerek araştırmalarda ön açıcı bir yöntem olarak fırsat sunmaktadır. Deniz vd.(2010:87) belirttiği gibi sayısallaştırmaların ardından, kategorik veriler için ihlal edilen çok değişkenli normal dağılım varsayımı artık sağlanmaktadır. Böylece en önemli sorunlardan biri çözüme bağlanır.

Tezin uygulama bölümünde; cep telefonu kullanımının, uyku düzeni üzerine etkisinin olup olmadığını, var ise hangi değişkenlerce açıklandığı Regresyon Analizi ile belirlenecektir. Hayatın her alanında kendini zorunlu kılan ve kitle iletişiminin tüm yeni unsurlarını (telefon, internet, mesajlaşma, müzik, tv, sosyal medya, video, oyun v.d.) bir arada barındıran akıllı telefon, özellikle genç yetişkinler tarafından her yerde sürekli ve sınırsız kullanılmaktadır. Cep telefonun kullanıldığı yerlerden biri de yatak odasıdır, dolayısıyla uykuya ayrılan zaman, akıllı telefon tarafından istila halindedir. Akıllı telefon kullanımı; uyku süresinin azalması, uykunun bölünmesi gibi davranışsal uyku problemlerine (Rosen vd, 2016) yol açtığı kadar, günlük faaliyetlere uyum sağlamada güçlüklerle ve diğer psiko sosyal problemlere yol açmaktadır.

Bu araştırma da Türkiye'nin farklı bir çok yerinden göç alması sebebiyle İstanbul ve Kocaeli illerinden toplam 491 kişiye uygulanan anket verileri analiz edilecektir. Veri seti tamamı kategorik 1 bağımlı, 25 bağımsız değişkenden oluşmaktadır. Bu kadar çok bağımsız değişkene aynı anda Regresyon Analizi uygulamak gerek verimlilik açısından gerekse de öngörü geliştirme açısından kargaşa yaratabileceğinden araştırmamızın temel savı, en az değişkenle en uygun modeli saptamaktır. Peki bu kadar değişkenden oluşan bir veri seti içerisinde mümkün alt setler (yaklaşık 70 milyon) içinden en ideal regresyon modeli nasıl bulunabilir? Bu soruya verilecek olan cevap kendini Evrimsel Algoritmada göstermiştir.

Birçok opsiyonel seçeneği içinde barındırarak araştırmacıya çözüm arayan bir sezgisel yöntem olan Genetik Algoritmalar, belirli bir amaç fonksiyonu doğrultusunda en iyi çözümü bulma konusunda literatürde sık olarak kullanılır. Bu şekilde mümkün her modelin uygululuğunu denemekten kurtularak, algoritmanın bulduğu optimal sonuca ulaşılır. Buraya kadar her şey sorunsuz tasavvur edilmiştir. Asıl belirleyici soru modelin seçim kriteri olarak kullanılacak amaç fonksiyonun ne olması gerektiği? Geleneksel regresyon modellerinde en çok kullanılan seçim kriterleri; Çoklu Belirlilik Katsayısı (R^2), Akaike Bilgi Kriteri (AIC) ve Bayes Bilgi Kriteri (BIC) vb... olarak sıralanabilir. Bu çalışmada yaygın kullanılmamakla birlikte giderek artan ilgide atıf alan ve Hamparsum Bozdoğan (1988) tarafından geliştirilen model yalınlığı ve uyum iyiliğinin yanısıra modelin karmaşıklığını da dikkate alan ICOMP (Information

Complexity) ve onun özel hali ICOMP (IFIM) kriteri amaç fonksiyonu olarak belirlenmiştir.

Son olarak Algoritma için gerekli kısıt ve seçimler yapılarak, 25 bağımsız değişken içerisinde bağımlı değişkeni açıklayabilecek ve amaç fonksiyonuna en uygun alt seti verecek regresyon modeli tespiti gerçekleştirilecektir.

Yalın, tutarlı ve tahmin gücü yüksek modellerin oluşturulmasında kullanacağımız metodolojinin etkin bir çözüm olanağı sunacağı düşünülmektedir. Böylece kategorik verilerle çalışan araştırmacılara bir çözüm tekniği sunma açısından, bütünlüklü bir teknik içermesi ve ilk kez uygulanması açısından analiz yöntemimiz birçok alanda kullanılabilir. Bu da çalışmanın özgünlüğüne işaret etmesi bakımından önemlidir.

BİRİNCİ BÖLÜM

İSTATİSTİK TAHMİN TEKNİKLERİ

Çalışmanın bu bölümünde, Optimal Ölçeklemenin temel öğelerinden olan Dalgalı En Küçük Kareler (Alternating Least Squares of Optimal-ALS) Algoritması ayrıntılı bir şekilde incelenecektir. ALS tekniği analizine geçmeden önce, istatistik tahmin tekniklerine temel düzeyde değinilecektir.

1.1. EN ÇOK OLABİLİRLİK TAHMİNİ (MAXIMUM LIKELIHOOD ESTIMATION)

İstatistiki modellerde parametrelerin tahmin edilebilmesinde en çok kullanılan yöntemlerden biri En Çok Olabilirlik Tahmin Yöntemi (EÇOB)'dur. Yöntem, gözlenen veri kümesini elde etme olasılığını maksimum yapan bilinmeyen parametrelerin değerlerini verir (Bircan, 2004:190) Başka bir tanıma göre, gözlenen değişken değerleriyle elde edilecek tahmini değerler arasındaki benzerliği maksimum yapacak olan logaritmik benzerliğin belirlenmesidir (Orhunbilge, 2010:190) İstatistikçi ve de genetikçi olan İngiliz bilim insanı R. A. Fisher tarafından 1912 ile 1922 yılları arasında geliştirilen yöntem, ilk zamanlarda Bayes istatistiğinde sık sık kullanılmakla birlikte ilerleyen dönemlerde basitleştirilerek farklı alanlarda uygulanmaya başlanmıştır (Pfanzagl, 1994:207). En Çok Olabilirlik Tahmin Yönteminin ardında yatan temel ilke şu beklentidir:

“Rastsal bir olayın gerçekleşmesi, o olayın gerçekleşme olasılığı en yüksek olay olmasındandır.” (Yalta, 2004:84).

En Çok Olabilirlik Tahmin Yönteminin çalışma prensibi şu şekilde işler; ana kütlenin olasılık dağılımı biliniyor ise eldeki örneklemin gelme ihtimalinin en çok olduğu ana kütlenin hangi katsayılarla sahip olduğu saptanır, eğer ana kütlenin olasılık dağılımı bilinmiyorsa varsayımda bulunarak yola devam edilir.

Ana kütlenin olasılık yoğunluk fonksiyonu belirlenir. Bu fonksiyon aslında verilen bir θ değeri için koşullu olasılık formülüyle hesaplanır. Burda amaç, θ değerinin gözlemlenen değerlerde maksimum olasılığa sahip olmasıdır.

$X_1, X_2, X_3 \dots, X_n$ gözlenen ana kütle değerleri olmak üzere

$$f(x_1, x_2 \dots x_n | \theta) = f(x_1 | \theta, x_2 | \theta \dots x_n | \theta) \quad (1)$$

tüm ana kütlelerden alınan örneklemelerin olasılık yoğunluk fonksiyonu olmak üzere θ nin olabilirlik fonksiyonu;

$$L(\theta | x_1, x_2 \dots x_n) = f(x_1 | \theta, x_2 | \theta \dots x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \quad (2)$$

olarak bulunur.

$L(\theta)$ olabilirlik fonksiyonunu maksimum yapan değer θ nin en çok olabilirlik tahmincisi olarak belirlenir. Genelde olabilirlik fonksiyonunun en çoklanması yerine fonksiyonun logaritması alınarak maksimizasyon işlemi sağlanır (Akdi, 2011).

$$l(\theta) = \log L(\theta | x_1, x_2 \dots x_n) \quad (3)$$

EÇOB yöntemini kullanarak farklı dağılımlar için anakütle parametrelerinin tahmini yapılmaktadır. Dağılım türüne göre tahmin ediciler Tablo 1’de sunulmuştur.

Tablo 1: Dağılım Türlerine Göre En Çok Olabilirlik Tahmin Edicileri

Dağılım Türü	Parametreler	EÇOB Tahmin Edicisi
Binom Dağılım	n: Deney sayısı p: Başarı gerçekleşme olasılığı k: İlgilenilen özelliğin sayısı	$\tilde{p} = \frac{\sum_{i=1}^N k_i}{\sum_{i=1}^N n_i}$
Poisson	λ : Gerçekleşen ortalama olay sayısı n: Deney sayısı	$\tilde{\lambda} = \frac{\sum x_i}{n}$
Normal Dağılım	μ : Ortalama σ : Standart sapma	$\tilde{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n}$
Üstel Dağılım	x: Ortalama süre θ : Süre oranı	$\tilde{\theta} = \frac{\sum x_i}{n}$

EÇOB yöntemi, birçok istatistiksel modelleme alanında kullanılmaktadır. Bunlar aşağıdaki gibi sınıflanabilir:

- Logistik Regrasyon Analizi
- Açıklayıcı ve Doğrulayıcı Faktör Analizi
- Yapısal Eşitlik Modellemesi
- Bilgi Kriterleri Oluşumunda

Bunun dışında pek çok test tekniğinde de bütünleşik olarak kullanılmaktadır.

1.1.1. En Çok Olabilirlik Tahmini için Basit Bir Uygulama

$L(\theta)$ olabilirlik fonksiyonu tezin ilerleyen bölümlerinde sıklıkla teleffuz edileceğinden, basit bir uygulama üzerinden nasıl tahmin edildiği aşağıdaki örnekte gösterilmiştir.

Örnekte anakütlesinin normal dağıldığı varsayılan 6 yıllık tütün mahsülünün yıllık ürün miktarı kilogram cinsinden verilmiştir. Buna göre yıllara ait ürün hasılatı 24, 26, 32, 34, 36 ve 38 kg olarak tespit edilmiş olsun. Acaba örneklem üzerinden giderek anakütle parametreleri olan standart sapma ve ortalama hangi olasılıkla bulunur.

$$f(x_1, x_2, \dots, x_n | \theta) = f(x_1 | \theta, x_2 | \theta, \dots, x_n | \theta) \quad (4)$$

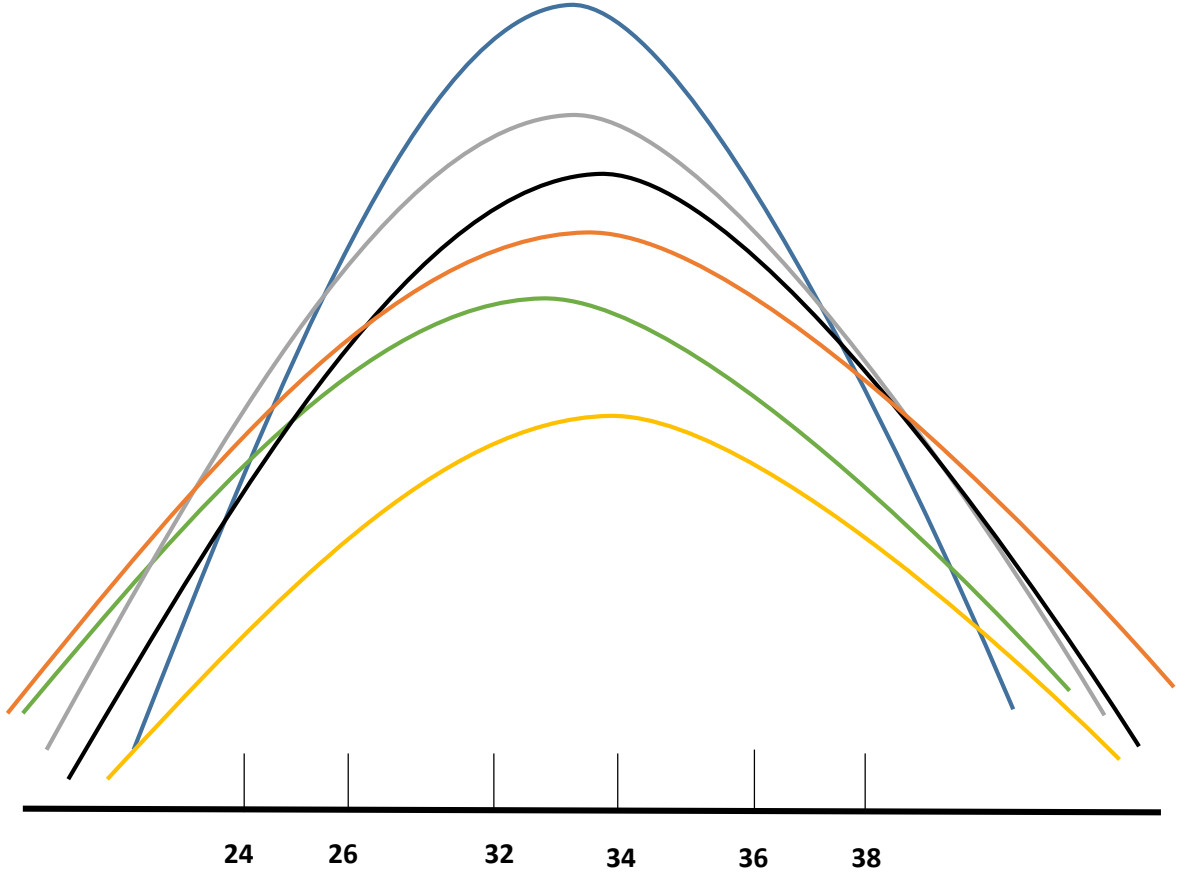
tüm ana kütlelerden alınan örneklemelerin olasılık yoğunluk fonksiyonu olmak üzere θ nın olabilirlik fonksiyonu;

$$L(\theta | x_1, x_2, \dots, x_n) = f(x_1 | \theta, x_2 | \theta, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \quad (5)$$

şeklinde hesaplanır. Buna göre anakütlesinin normal dağıldığını varsayıldığına göre formül

$$pr(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad (6)$$

parametrelerini kullanarak olasılık yoğunluk fonksiyonu hesaplar. Acaba bu örneklem hangi ana kütlede hangi olasılıkla gelmiş olabilir.



Şekil 1 : Farklı Ana Kütle Dağılımları

Örneğin; ortalaması (\hat{X}) 28, standart sapması (S) 6 olan anakütleden gelme olasılığı şu şekilde hesaplanır.

$$L(\mu, \sigma \setminus x_1, x_2, x_3 \dots x_n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad (7)$$

$$L(\mu = 28 \text{ ve } \sigma = 6 \setminus 24, 26, 32, 34, 36 \text{ ve } 38)$$

$$L(\mu = 28 \text{ ve } \sigma = 6 \setminus 24) \times L(\mu = 28 \text{ ve } \sigma = 6 \setminus 26) \times L(\mu = 28 \text{ ve } \sigma = 6 \setminus 32) \times \\ \times L(\mu = 28 \text{ ve } \sigma = 6 \setminus 34) \times L(\mu = 28 \text{ ve } \sigma = 6 \setminus 36) \times L(\mu = 28 \text{ ve } \sigma = 6 \setminus 38)$$

$$= \frac{1}{\sqrt{2\pi 6^2}} e^{-\frac{(24-28)^2}{2 \cdot 6^2}} x \frac{1}{\sqrt{2\pi 6^2}} e^{-\frac{(26-28)^2}{2 \cdot 6^2}} x \frac{1}{\sqrt{2\pi 6^2}} e^{-\frac{(32-28)^2}{2 \cdot 6^2}} x \dots$$

$$x \frac{1}{\sqrt{2\pi 6^2}} e^{-(36-28)^2/2 \cdot 6^2} x \frac{1}{\sqrt{2\pi 6^2}} e^{-(38-28)^2/2 \cdot 6^2} x$$

$$L(\theta) = 0,00000000330$$

bulunur. Yani örneklemin belirlenen ana kütleden gelme olasılığını bulmuş olduk. Bu işlemi tahmini olarak verilen başka ana kütle parametreleri için de kullanabiliriz. Yapılan bir çok denemeden sonra aşağıdaki tabloya ulaşabiliriz.

Tablo 2: $L(\theta)$, En Çok Olabilirlik Tahmincisi

Parametreler	$L(\theta)$
$L(\mu = 28 \text{ ve } \sigma = 6 \setminus 24, 26, 32, 34, 36 \text{ ve } 38)$	0,00000000330
$L(\mu = 30 \text{ ve } \sigma = 6 \setminus 24, 26, 32, 34, 36 \text{ ve } 38)$	0,00000000790
$L(\mu = 32 \text{ ve } \sigma = 5 \setminus 24, 26, 32, 34, 36 \text{ ve } 38)$	0,00000001140
$L(\mu = 28 \text{ ve } \sigma = 5 \setminus 24, 26, 32, 34, 36 \text{ ve } 38)$	0,00000000230
$L(\mu = 31 \text{ ve } \sigma = 5 \setminus 24, 26, 32, 34, 36 \text{ ve } 38)$	0,00000001090
$L(\mu = 30 \text{ ve } \sigma = 5 \setminus 24, 26, 32, 34, 36 \text{ ve } 38)$	0,00000000830

Tablo 2’den de anlaşılacağı gibi, denenen parametreler bakımından örneklemin gelme ihtimali en yüksek ana kütlenin ortalaması 32, standart sapması 5 olarak tahmin edilmiştir. En Çok Olabilirlik Tahmihcisi ($L(\theta)$) diğer dağılımlar üzerinden de bulunabilir. Son bir not olarak, görüldüğü gibi tahmin edicinin değerleri çok küçük sayılardır. Bu sayıları karşılaştırmak adına değerlerin logaritması alınabilir. Ancak 0-1 arasındaki sayıların logaritması negatif çıktığından literatürde genellikle ‘-2log’ çarpanı kullanılır. Bu şekilde mukayese de kolaylık sağlanır.

1.2. EN KÜÇÜK KARELER YÖNTEMİ

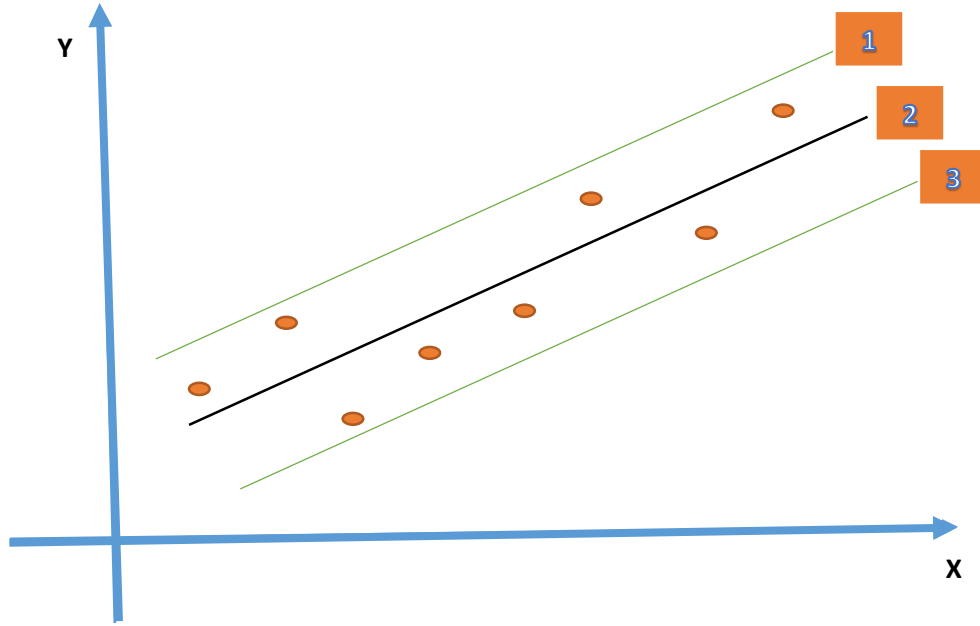
Doğrusal regresyon modellerinin parametrelerinin tahminde kullanılan çeşitli yöntemler arasında en yaygın kullanılan yöntem olan EKK (En Küçük Kareler) tahmincileri, klasik doğrusal regresyon modelinin varsayımları sağlandığında sapmasız doğrusal tahminciler içerisinde en küçük varyanslı olanlardır; yani en iyileridir. Bu yöntem 1806 yılında Legendre tarafından geliştirilse de yöntemi Gauss'un daha önce bulduğu bilinmektedir. Bu nedenle yöntem Legendre ile değil, Gauss ile anılır. Gauss, hata teriminin normal dağıldığı durumda EKK'nın en iyi olduğunu gösterdikten sonra, yöntem uygulamada standart bir mekanizma haline gelmiştir (Türkay, 2004:106).

EKK yöntemi, belli varsayımların sağlanması halinde bağımlı ve bağımsız değişkenler arasındaki ortalama ilişkiyi gerçeğe en yakın şekilde tahmin eden yöntemdir. EKK yönteminin varsayımları aşağıda verilmiştir:

- Hata terimi stokastik bir değişken* olmalıdır.
- Hata terimlerinin ortalaması sıfır olmalıdır.
- Hata terimleri normal dağılmalıdır.
- Hata terimi sabit varyanslıdır.
- Hata terimleri arasında ilişki olmamalıdır.
- Bağımsız değişkenler arasında güçlü bir korelasyon ilişkisi olmamalıdır (Tarı, 2016)

Çeşitli bağımsız değişkenler (X) karşısındaki bağımlı değişken (Y) değerlerinin dağılımını gösteren serpilme diyagramları incelendiğinde doğrusal bir eğilim gözüküyorsa X'in Y'ye göre matematik fonksiyonunun doğrusal olduğuna karar verilebilir (Orhunbilge, 2010:66).

* Stokastik Değişken: Hangi değer veya sonucu alacağı önceden bilinmeyen, değeri şansa bağlı olarak tesadüfi bir şekilde ortaya çıkan değişkene denir.



Şekil 2: Serpilme Diyagramı

Bu doğrusallığın tahmini için birçok lineer fonksiyon saptanır. Bu fonksiyonlar (1, 2 ve 3) içinden en optimalı, Y gözlem değerine en yakın tahmini veren \hat{Y} değerini saptayan doğrusal fonksiyon olacaktır. Burada temel amaç, gözlenen değerler ile tahmin değerleri arasındaki farkın mümkün olduğunca en küçüklenmesidir. Başka bir ifadeyle tahmin hatalarının (e) minimal seviyede olması hedeflenmektedir.

Serpilme diyagramında görüldüğü gibi, Y ile X arasındaki gerçek ilişkiyi en iyi temsil eden fonksiyon (2) numaralı doğrudur. Bu fonksiyon X ve Y gözlem değerlerine en yakın bulunan doğrudur. EKK yöntemi ile tahmin değişkenini oluşturacak olan doğrusal fonksiyon denklemi bulunur.

Y ve X arasındaki ilişki;

$$Y = \beta_0 + \beta_1 X + \epsilon \quad \text{şeklinde olacaktır.} \quad (8)$$

β_0 = Doğrunun y eksenini kestiği nokta

β_1 = Doğrunun eğimi

ϵ = Şansa bağlı hata terimi

Pratikte β_0 ve β_1 deęerleri bilinmiyorsa, ana kütleden bir örnek alınarak ana kütle parametreleri hakkında istenilen bilgiler üretilir. Bu noktada tahmini deęerler olarak b_0 ve b_1 kullanılır.

$$\hat{y} = b_0 + b_1x + e \quad (9)$$

\hat{y} = y'nin tahmini deęeri

Gözlenen deęer ile tahmin deęerleri arasındaki farkın (hatanın) en az olması gerektięi için istenilen özellik ařaęıdaki gibi olacaktır:

$$e = y - \hat{y} = y - b_0 - b_1x_i \quad (10)$$

Hata terimlerinin bir kısmı negatif bir kısmı pozitif olacaęından toplam sonuç sıfır çıkacaktır. Bu nedenle hata terimlerinin işaret farkını giderebilmek için karelerinin alınması çözüm olarak sunulmuřtur. Yönteme ismini veren olgu da budur, hataların karelerini en küçüklemek.

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1x_i)^2 \quad (11)$$

Bu fonksiyonun en küçük olabilmesi için b_0 ve b_1 parametrelerine göre birinci mertebeden türev 0'a eřitlenir. Süreç sonunda parametreleri veren ařaęıdaki denklemlere ulařılır.

$$\sum y = nb_0 + b_1 \sum x \quad (12)$$

$$\sum xy = b_0 \sum x + b_1 \sum x^2 \quad (13)$$

Böylece gözlemlere en uygun fonksiyon denklemi (doęru) bulunur. Bu denklemlere EKK Normal Denklemleri de denir. EKK yöntemi, varsayımları saęlandığı kořullarda tahmin geliştirme süreçlerinde etkin kullanılan bir yöntemdir.

1.3. TOPLAM EN KÜÇÜK KARELER TEKNİęİ

Doęrusal regresyon modellerinde deęişken tahminlerindeki *hatalar* için en yaygın olarak kullanılan tekniklerden bir tanesi de Toplam En Küçük Kareler Yöntemidir. Yöntem ayrıca Ortogonal Regresyon olarak da adlandırılır. Ortogonal Regresyon analizinde amaç, veri noktalarından uyumlu eğriye olan dik uzaklıkları en az yapmaktır. Eğer, varsayımlar saęlanırsa, Toplam En Küçük Kareler Yöntemi mükemmele yakın bir tahmin deęerlendirme yöntemidir (Öztürk, 2012:60). En Küçük

Kareler Yönteminden farklı olarak, bir değişkenin diğerine nasıl bağımlı olduğunu bulmaktan ziyade, iyi bir grafik temsilinin sağlanması amaç edilir (Costantini ve Bernardi, 2008:243).

1.4. YANLI (RIDGE) EN KÜÇÜK KARELER TEKNİĞİ

Çoklu doğrusal bağlantı sorununun olduğu analizlerden elde edilen EKK tahmin edicisinin varyans değerleri normalden büyük olmakta ve regresyon katsayıları olması gereken değerlerden uzaklaşmaktadır. Bu sorunu aşmak için 1970 yılında Hoerl ve Kennard tarafından önerilen Ridge tahmin edicisi kavramı ortaya atılmıştır. Yöntem parametreleri yanlı olarak tahmin ederek daha küçük varyansa sahip regresyon katsayıları elde eder. Sistemin temel mantığı $X'X$ matrisinin köşegen elemanlarına küçük bir k sabite ekleyerek tahmin geliştirmektir. (Bulut, 2011:20). Ridge tahmin edicisi aşağıdaki eşitlikle elde edilir.

$$\hat{\beta}_R = (X'X + kI)^{-1}X'y \quad (0 < k < 1) \quad (14)$$

k sabiti Ridge tahmin edicisinin verimliliğini etkilemektedir. $k=0$ olması durumunda tahmin edici EKK yöntemine dönüşmektedir. Varyanstaki azalma, k daki artışın karesindeki artıştan fazla olduğu sürece yöntemin performansı başarılı olarak görülmektedir.

1.5. KİSMİ EN KÜÇÜK KARELER TEKNİĞİ

Yanlı bir regresyon türü olan Kısmi En Küçük Kareler (KEKK) Tekniği ile Y değişkenlerini X değişkenleriyle birlikte döndürebilmek için X ile türdeş olan gizli değişkenler elde edilir. Bu gizli değişkenler kullanılarak boyut azaltma işlemi gerçekleştirilmiş olur (Bağcı 2017:15). Temel Bileşenler analizine benzer şekilde gizli değişkenler X 'in doğrusal ayırımlarını içermektedir. Farklı olan yanı ise gizli değişkenler belirlenirken bağımlı değişkenin ağırlığı da göz önüne alınır. KEKK ilk adımında X ve Y arasındaki kovaryansın büyük olması için gizli değişkenler hesaplar. Yöntem birçok algoritmayla çalışabilmektedir. Bunlar; Tek Değişkenli Kısmi En Küçük Kareler Regresyonu (PLS), Lineer Olmayan Yinelemeli Kısmi En Küçük Kareler Regresyonu (NIPALS) bu bütünleşik algoritmalarındandır. Sonuç olarak X bağımsız değişkenlerine uygulanan dönüşümler neticesinde elde edilen daha küçük

boyuttaki gizli deęişkenler vasıtasıyla, bağımlı deęişkene (Y) En Küçük Kareler Teknięi uygulanır. Süreç biraz yanlı olsa da hata varyanslarının düşük olması hedefi saęlanır.

1.6. DALGALI (AĞIRLIKLİ) EN KÜÇÜK KARELER TEKNİĞİ

Bu tez kapsamında işlenecek olan *Optimal Ölçekleme* konusunun temelini oluşturan algoritma, Dalgali En Küçük Kareler (Alternating Least Squares of Optimal-ALS) teknięidir. Lineer Çok Deęişkenli Analiz teknikleri Öklityen uzay özelliklerinden yararlanmakta ve deęişkenler arasındaki doğrusal ilişkiler üzerinde çalışmaktadır. NonLineer Çok Deęişkenli Analiz teknikleri ise kategorik deęişkenlere çeşitli dönüşümler uygulayarak sayısallaştırma işlemleri gerçekleştirir. Böylece minimum bir kayıp ile gerçek uzaydaki ilişkiler iki ya da üç boyutlu uzayda temsili olarak ifade edilebilecektir.

Özellikle Sosyal Bilimler alanında yapılan çalışmalar daha çok anket ve tutum ölçekleri gibi sınıflayıcı ya da sıralayıcı kategorik testler ile gerçekleşir. Ancak verilerin kategorik olması Çok Deęişkenli Analiz tekniklerinin uygulanabilmesi için gerekli koşul olan varsayımların (çoklu normallik, otokorelasyon, çoklu doğrusallık vb...) saęlanmasına engel teşkil eder. Nitel deęişkenlere kodlama gerçekleştirerek nicel bir deęer atanabilir fakat bu kategoriler arasında doğru bir ilişkiyi tanımlayamaz. Gerçek anlamda bir sayısallaştırma işlemini Dalgali En Küçük Kareler algoritmasını kullanan *Optimal Ölçekleme* yöntemi gerçekleştirir. *Optimal Ölçekleme* yöntemi, kurucusu ve geliştiricisi olan GİFİ ile özdeşleşerek Gifi Teknikleri olarak da adlandırılmaktadır. Tam olarak Optimal Ölçekleme teorisi 1980’li yılların başında Gifi, De Leeuw, Young ve Takane tarafından geliştirilmiş ve ilerleyen yıllarda birçok teknięin alt yapısını oluşturmuştur (Michailidis ve de Leeuw 1998:73).

Optimal Ölçekleme kavramına, test türleri ve uygulama yöntemleri daha sonraki bölümlerde ayrıntılı bir şekilde ele alınacaktır. Şimdi sayısallaştırma işlemlerinin nasıl gerçekleştięi ve bunun için gerekli olan Dalgali En Küçük Kareler algoritmasının nasıl yapıldıęını küçük bir örnekle açıklanacaktır. Örneęe geçmeden önce bilinmesi gereken bazı temel kavramları açıklamak algoritmanın anlaşılması açısından önemlidir.

1.6.1. Kayıp Fonksiyon

İlk olarak Gifi tarafından kullanılan Homojenlik Analizi Teorisi'nde kendine yer bulan Kayıp Fonksiyon; ham kategorik verilere sayısal dönüşümü uygulayacak fonksiyonun minimum kayıpla tespit edilmesidir. Homojenlik analizi, ayrıklığın homojenlikten en aza indirilmesi kriterine dayanmaktadır. Bu farklılık, kayıp fonksiyonu ile ölçülür (De Leeuw ve Mair 2007:2). Optimal Ölçekleme Tekniğinin temel amacı, sayısal dönüşümler yapacak bir fonksiyon üretmek ve bu fonksiyon sonucu elde edilen kayıpları belirli kısıtlar altında minimize etmektir. Kayıp fonksiyon notasyonu aşağıda verilmiştir:

$$\sigma(\mathbf{x}, \mathbf{a}) = \mathbf{m}^{-1} \sum_j \mathbf{SSQ}(\mathbf{x} - \mathbf{G}_j \mathbf{a}_j) \quad (15)$$

Notasyonların açılımı aşağıdaki gibidir:

\mathbf{x} : nesne skorları

\mathbf{a}_j : kategori sayısallaştırmaları vektörü

\mathbf{m} : H veri matrisinin sütun değişken sayısı

\mathbf{G}_j : j. değişken için Gösterge matrisi

$\mathbf{SSQ}()$: matris elemanlarının karelerinin toplamı

Kayıp fonksiyonun minimize edilmesi için bir takım kısıtlar oluşturulur.

$$\text{Kısıtlar : } \mathbf{x}'\mathbf{x} = nI \quad \text{ve} \quad \mathbf{u}'\mathbf{x} = 0 \quad (16)$$

Kısıtlar ile nesne skorları, ortalaması 0 ve varyansı 1 olan standart normal değişkenler haline dönüştürülecektir. Bu nedenle bu kısıtlar normalizasyon kısıtları olarak bilinmektedir (Gifi 1991:88).

1.6.2. ALS Algoritması ve Aşamaları

Dalgalı En Küçük Kareler algoritmasının başlangıç noktası $H_{n \times m}$ ile gösterilen data matrisini (sıra sayısı n; nesne sayısı ve sütun sayısı m; değişken sayısı) kayıp fonksiyon iterasyonlarla (Bauer-Rutishauser simultane iterasyonları olarak da bilinir) optimum seviyeye getirinceye kadar sayısal dönüşüme tabi tutmaktır. Algoritma iki amaç üzerinden çalışır: Birincisi nesne skorlarını (\mathbf{x}) oluşturmak ikincisi ise kategori puanlarını (\mathbf{a}) hesaplamaktır. Her ikisini aynı anda yapamaz. Ayrı ayrı yapılan her

adımında kayıp fonksiyon, a sabit tutulup x skoruna göre en küçüklenirken, daha sonraki adımda x sabit tutulup a puanına göre küçüklenir (Gifi 1991:87-88).

1.6.2.1. Normalleştirilmiş Skor Algoritması

Normalleştirilmiş Skor Algoritmasında nesne skorları $x'x = 1$ koşulu ile sınırlandırılmıştır. Algoritmanın başlaması için a puanları için önsel değerler ataması yapılır ($a \neq 0$). Ardından aşağıdaki adımlar izlenir (Gifi, 1991:88):

$$(1) \quad \text{Nesne skorlarını güncelle :} \quad \tilde{x} \leftarrow H\tilde{a}/m \quad (17)$$

$$(2) \quad \text{Normalizasyon :} \quad x^+ \leftarrow \tilde{x}(\tilde{x}'\tilde{x})^{-1/2} \quad (18)$$

$$(3) \quad \text{Ağırlık puanlarını güncelle:} \quad a^+ \leftarrow H'x^+ \quad (19)$$

$$(4) \quad \text{Yakınsaklık testi :} \quad x^+ \text{ ve } a^+ \text{ değerleri yeterince dengelenmedikçe} \\ \text{birinci adıma geri dön.} \quad (20)$$

Algoritmanın içeriği Tekil Değer Ayırıştırma* Yöntemiyle ilişkilidir. Algoritma K'nin ilk köşegen elemanı k_1 , Λ 'nın ilk köşegen elemanı λ_1 , L'nin ilk elemanı l_1 ile aşağıda verilen formüle göre yakınsama gerçekleştirir.

$$H'x^* = L\Lambda K'x^* = L\Lambda K'k_1 = \lambda_1 l_1 = a^* \quad (21)$$

Ayrıca göreceli kayıp, $R=H'H$ matrisinin baskın özdeğeriyle aşağıdaki şekilde hesaplanır(Gifi, 1991:89):

$$\frac{W}{T} = 1 - \frac{B}{T} = 1 - \frac{\lambda_1^2}{m} \quad (22)$$

Buna ek olarak $L\Lambda$ matrisi, Temel Bileşenler Analizinin yük matrisi olarak da bilinir.

* **Tekil Değer Ayırıştırması** : Her boyuttaki H matrisini, iki ortonormal matrise K (sol tekil vektör), L (sağ tekil vektör) ve bir köşegen matrise Λ (tekil değerlerin köşegen matrisi)'e ayıran yonteme denir. Sonuçta H matrisi;

$$H = K \Lambda L'$$

olarak ayrışırma uğrar.

1.6.2.2. Normalleştirilmiş Ağırlıklar Algoritması

Normalleştirilmiş Ağırlıklar Algoritmasında ağırlıklar $a'a = 1$ koşulu ile sınırlandırılmıştır. Normalleştirilmiş Skor Algoritmasında skor puanları (x) için geçerli olan adımlar bu sefer ağırlık puanı olan a için uygulanır. Algoritmanın başlaması için x puanları için önsel değerler ataması yapılır ($x \neq 0$). Ardından aşağıdaki adımlar izlenir:

$$(1) \quad \text{Ağırlıkları güncelle} \quad : \quad \tilde{a} \leftarrow H'x \quad (22)$$

$$(2) \quad \text{Normalizasyon} \quad : \quad a^+ \leftarrow \tilde{a}(\tilde{a}'\tilde{a})^{-1/2} \quad (23)$$

$$(3) \quad \text{Nesne skorlarını güncelle:} \quad x^+ \leftarrow Ha^+/m \quad (24)$$

$$(4) \quad \text{Yakınsaklık testi} \quad : \quad x^+ \text{ ve } a^+ \text{ değerleri yeterince dengelenmedikçe} \\ \text{birinci adıma geri dön.} \quad (25)$$

Süreç baştan sona kadar Normalleştirilmiş Skor Algoritması ile benzer şekilde işler. Aynı anda hem nesne skorlarını hem de kategori puanlarını bulan bir algoritma henüz geliştirilememiştir. İlerleyen bölümlerde ayrıntısıyla anlatılacak olan farklı Optimal Ölçekleme Test Tekniklerinde ihtiyaca bağlı olarak bu algoritmalarından biri ya da ikisi birden kullanılabilir.

1.6.3. Dalgalı En Küçük Kareler'de Sayısallaştırma işlemi Üzerine Bir Uygulama

Dalgalı En Küçük Kareler (DEKK-ALS) yönteminin sayısallaştırma işlemini nasıl yaptığını küçük bir örnekle açıklayalım. Burada Normalleştirilmiş Skor Algoritması kullanılarak x nesne skorları tespit edilecektir. Bunun için dönüşüme uğratılacak bir H matrisi oluşturulalım.

$$H = \begin{bmatrix} 3 & -4 & 0 \\ -1 & 2 & 1 \\ -5 & -1 & -2 \\ 3 & 3 & 1 \end{bmatrix} \text{ olsun;}$$

Şimdi her j sütununun (h_j) ağırlığı olan a vektörünü önsel olarak atayalım.

$$\tilde{a} = [1 \ 1 \ 1]$$

Bu ağırlıkları ilgili sütun değerleriyle çapıp yeni bir H matrisi elde edilir.

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} 3 & -4 & 0 \\ -1 & 2 & 1 \\ -5 & -1 & -2 \\ 3 & 3 & 1 \end{bmatrix} \rightarrow \text{yeni matris } H = \begin{bmatrix} 3 & -4 & 0 \\ -1 & 2 & 1 \\ -5 & -1 & -2 \\ 3 & 3 & 1 \end{bmatrix} \text{ olur.}$$

Şimdi algoritmanın birinci adımı olan güncellenmiş nesne skorları vektörü bulunur.

$$\tilde{x} = \sum H \tilde{a} = \begin{bmatrix} 3 + (-4) + 0 \\ -1 + 2 + 1 \\ -5 + (-1) + (-2) \\ 3 + 3 + 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -8 \\ 7 \end{bmatrix}$$

İkinci adım olarak skorların normalizasyonu* yapılır. Bunun için genellikle Öklit uzaklık fonksiyonu kullanılır. Bunun için x vektörünün uzunluğu bulunur.

$$\|x\| = \sqrt{(-1)^2 + (2)^2 + (-8)^2 + (7)^2} = 10,863$$

Normalizasyonunu son adımı olarak her bileşen, o vektörün büyüklüğüne bölünür.

$$\hat{x} = \frac{x}{\|x\|}$$

$$x^+ = \begin{bmatrix} -\frac{1}{10,863} \\ \frac{2}{10,863} \\ \frac{-8}{10,863} \\ \frac{7}{10,863} \end{bmatrix} = \begin{bmatrix} -0,092 \\ 0,184 \\ -0,736 \\ 0,644 \end{bmatrix}$$

ya da;

$$x^+ \leftarrow \tilde{x}(\tilde{x}'\tilde{x})^{-1/2} = \begin{bmatrix} -1 \\ 2 \\ -8 \\ 7 \end{bmatrix} \left(\begin{bmatrix} -1 & 2 & -8 & 7 \end{bmatrix} * \begin{bmatrix} -1 \\ 2 \\ -8 \\ 7 \end{bmatrix} \right)^{-1/2} = \begin{bmatrix} -0,092 \\ 0,184 \\ -0,736 \\ 0,644 \end{bmatrix}$$

* **Normalizasyon:** Vektörün uzunluğunun bir birim olacak şekilde ölçeklenmesidir. Bu amaçla vektörün tüm bileşenleri vektörün uzunluğuna bölünür.

olarak bulunur ve 3. adıma geçilir. Bu adımda güncellenmiş ağırlık puanları hesaplanır. Ancak burada algoritma H data matrisinin homojen olması (0 -1 arası değerler) varsayımını zorunlu kılar. Örneğimizdeki veriler heterojen olduğundan, H matrisinin sütunlarını D diagonal tekli marjinal matrisiyle güncellenir.

D matrisi, H data matrisinin transpozesiyle çarpımının diagonal elemanlarından oluşur.

$$D = diag(H'H) = diag\left(\begin{bmatrix} 3 & -1 & -5 & 3 \\ -4 & 2 & -1 & 3 \\ 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 3 & -4 & 0 \\ -1 & 2 & 1 \\ -5 & -1 & -2 \\ 3 & 3 & 1 \end{bmatrix}\right) =$$

$$diag\left(\begin{bmatrix} 44 & 0 & 12 \\ 0 & 30 & 7 \\ 12 & 7 & 6 \end{bmatrix}\right) = \begin{bmatrix} 44 & 0 & 0 \\ 0 & 30 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

Buna göre algoritma güncellenmiş ağırlık puanlarını aşağıdaki gibi hesaplar:

$$a^+ \leftarrow D^{-1}H'x^+ = \begin{bmatrix} 44 & 0 & 0 \\ 0 & 30 & 0 \\ 0 & 0 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 3 & -1 & -5 & 3 \\ -4 & 2 & -1 & 3 \\ 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} -0,092 \\ 0,184 \\ -0,736 \\ 0,644 \end{bmatrix}$$

$$a^+ = [0,117 \quad 0,113 \quad 0,383]$$

olarak bulunur. Bu şekilde 1. iterasyon tamamlanmıştır. Burada \tilde{a} ve $D^{-1}H'x^+$ değerlerinin birbirine kayıp minimum olacak kadar yakınsayıp yakınsamadığına bakılır. Eğer yakınsama sağlanmışsa iterasyon bitirilir, sağlanmamışsa 1. adıma tekrar dönülür. İterasyon işlemleri Tablo 3' te özetlenmiştir.

Tablo 3: Heterojen H Matrisinin Normalleştirilmiş Skor Algoritması İterasyon Özeti

İterasyon	\tilde{a}	$H\tilde{a}$	x^+	$D^{-1}H'x^+$
1.	1	-1	-0,092	0,117
	1	2	0,184	0,113
	1	-8	-0,736	0,383
		7	0,644	

Görüldüğü gibi yakınsama istenilen seviyede gerçekleşmemiştir. Bu yüzden iterasyon ikinci adıma geçecektir. Aynı şekilde H data matrisinin sütunları yeni bulunan $D^{-1}H'x^+$ ağırlık puanlarıyla çarpılıp güncellenir.

$$\begin{bmatrix} 0,117 & 0,113 & 0,383 \end{bmatrix}$$

$$H = \begin{bmatrix} 3 & -4 & 0 \\ -1 & 2 & 1 \\ -5 & -1 & -2 \\ 3 & 3 & 1 \end{bmatrix} \rightarrow H = \begin{bmatrix} 0,351 & -0,452 & 0 \\ -0,117 & 0,226 & 0,383 \\ -0,585 & -0,113 & -0,766 \\ 0,351 & 0,339 & 0,383 \end{bmatrix} \text{ olur.}$$

Ardından sırayla aşağıdaki değerler bulunur:

$$\tilde{x} = \sum H\tilde{a} = \begin{bmatrix} -1,101 \\ 0,492 \\ -1,464 \\ 1,073 \end{bmatrix}$$

$$x^+ \leftarrow \tilde{x}(\tilde{x}'\tilde{x})^{-1/2} = \begin{bmatrix} -0,053 \\ 0,261 \\ -0,777 \\ 0,569 \end{bmatrix}$$

$$a^+ \leftarrow D^{-1}H'x^+ = \begin{bmatrix} 44 & 0 & 0 \\ 0 & 30 & 0 \\ 0 & 0 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 3 & -1 & -5 & 3 \\ -4 & 2 & -1 & 3 \\ 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} -0,053 \\ 0,261 \\ -0,777 \\ 0,569 \end{bmatrix}$$

$$a^+ = [0,117 \quad 0,107 \quad 0,397]$$

Şeklinde bulunur. Dikkat edilirse a^+ vektörünün ilk elemanı ($a_{1 \times 1}^+$) durağanlaşmıştır. Yani kayıp minimize olmuş fakat diğer iki eleman için aynı şeyler istenilen hale gelmemiştir. Bu yüzden algoritma çalışmaya devam eder. Sonuç itibariyle iterasyon 6. adımda yakınsamayı sağlayarak kaybı minimize etmiştir. Tablo 4'te İterasyon özetini göstermektedir.

Tablo 4: Heterojen H Matrisinin Normalleştirilmiş Skor Algoritması Sonuç İterasyon Özeti

İterasyon	\tilde{a}	$H\tilde{a}$	x^+	$D^{-1}H'x^+$
1.	1	-1	-0,092	0,117
	1	2	0,184	0,113
	1	-8	-0,736	0,383
		7	0,644	
2.	0,117	-0,101	-0,053	0,117
	0,113	0,492	0,261	0,107
	0,383	-1,464	-0,777	0,397
		1,073	0,569	
3.	0,117	-0,077	-0,040	0,118
	0,107	0,494	0,260	0,105
	0,397	-1,490	-0,783	0,398
		1,071	0,563	
4.	0,118	-0,064	-0,033	0,119
	0,105	0,490	0,257	0,104
	0,398	-1,495	-0,785	0,398
		1,070	0,561	
5.	0,119	-0,058	-0,030	0,119
	0,104	0,487	0,255	0,103
	0,398	-1,497	-0,786	0,398
		1,068	0,561	
6.	0,119	-0,054	-0,028	0,119
	0,103	0,485	0,254	0,103
	0,398	-1,498	-0,787	0,398
		1,067	0,560	

Tablo 4'te görüldüğü üzere \tilde{a} ve $D^{-1}H'x^+$ arasındaki durağanlık $\varsigma = 0,001$ farkla sağlanmıştır. İterasyon sonucunda H matrisini oluşturan nesnelerin (sütunların) skorları belirlenmiştir. Buna göre H matrisinin nesne skor matrisi;

$$\tilde{x} = \begin{bmatrix} -0,028 \\ 0,254 \\ -0,787 \\ 0,560 \end{bmatrix}$$

olarak saptanmıştır. Aynı şekilde Normalleştirilmiş Ağırlıklar Algoritmasını kullanarak da kategorilerin puanları bulunabilir. Bu prosedürde hem x hem de a için tek bir çözüm bulunur. Genelde p boyutta farklı çözümle ilgilenebilir. Yani test tekniğine göre tekli ya da çoklu sayısallaştırma yapılabilir.

Çoklu sayısallaştırma yapılması durumunda, kategori puanları $(k \times p)$ boyutlu Y_j matrisiyle $(n \times p)$ boyutlu indirgenmiş nesne skorları matrisi elde edilir. Tekli sayısallaştırmanın yapıldığı testlerde ölçek düzeyi kısıtı devreye girebilmekte, Çoklu sayısallaştırmanın yapıldığı analizlerde ise bu kısıta gerek duyulmamaktadır (Michailidis ve Leeuw, 1998:319)

İKİNCİ BÖLÜM

OPTİMAL ÖLÇEKLEME TEKNİKLERİ VE TÜRLERİ

Gifi sisteminde Çok Boyutlu Ölçekleme Tekniğinden esinlenilmiştir; ki bu yöntem özellikle karmaşık çok değişkenli verinin sunumunda Öklidyen uzaklıklardan yararlanır (Leeuw, 2011:1). Burada temel amaç, değişkenler arasındaki ilişki çoklu bir şekilde ele alındığından oluşabilecek bilgi kaybını en aza indirmektir (Güç ve Başar, 2016:15). Optimal Ölçeklemenin amacı, sayısallaştırılmış değişkenlerin korelasyon matrisi özelliklerinin maksimize etmektir. Yöntem özellikle analizde seçilen bileşen sayısını belirten P adet sayısallaştırılmış değişkenin korelasyon matrisinin ilk P sayıdaki özdeğerini maksimize etmektedir (Linting vd., 2007:338). Gifi teknikleri gün geçtikçe çeşitlenmekte ve evrilmektedir. Bununla birlikte temel teknikler aşağıda açıklanmıştır.

2.1. ÇOKLU UYGUNLUK YA DA HOMOJENLİK ANALİZİ (HOMALS)

Tarihsel olarak Homojenlik kavramı, farklı değişkenleri aynı şeyler ölçekbilir fikriyle yakından ilişkilidir (Gifi, 1991:81). Homojenlik analizi ise değişkenler arasındaki ilişkileri iki ya da daha çok boyutlu çapraz tablolarla inceleyen istatistiksel yöntemdir. Yöntemin amacı karmaşık ve çok boyutlu yapıdaki bir veri matrisini minimum kayıpla ve olabilecek en fazla varyansla temsil edecek basit yapıdaki yeni bir veri matrisi ve grafik ile açıklamaktır (Alpar, 2013). Birden çok kategoriyi içinde barındıran değişkenlerle ilgilenen Homojenlik analizi, değişkenlerin homojenliğini maksimize edecek şekilde kategorileri puanlar (Meulman, 1982; Nishisato, 2014).

Uygunluk analizi çapraz tablolardaki boyut ve değişken sayısına göre iki farklı şekilde uygulanmaktadır. Bunların en basit hali olanı Basit Uygunluk Analizi, iki yönlü çapraz tabloların incelenmesinde kullanılırken, ikiden fazla değişkenin olduğu ve verilerin bir matris olarak kodlanıp çok yönlü çapraz tablolarda uygulanması ise Homojenlik Analizi ya da Çoklu Uygunluk Analizi olarak adlandırılır (Suner et Çelikoğlu, 2010a:45). Analiz çeşitli ülkelerde farklı isimlerde de adlandırılmaktadır. Amerika'da Optimal Ölçekleme, Almanya'da Homojenlik Analizi, Fransa'da Çoklu Uygunluk

Analizi, İsrail’de Skalogram Analizi ve Japonya’da Sayısallaştırma Metodu olarak isimlendirilir (Tenenhaus ve Young, 1985:91).

Basit Uygunluk Analizi; Correspondence Analizi, Uyum Analizi ve Karşılık Getirme Analizi gibi isimlerle de kullanılır. Yöntem ilk olarak 1935 yılında Hirschfeld tarafından üretilmiş daha sonrasında ise Fisher (1940) ve Guttman (1941) tarafından geliştirilmiştir. Bununla birlikte Uygunluk Analizinin tanınırlığını sağlayan kişi 1984 yılında yayınlanan “*Theory and Applications of Correspondence Analysis*” kitabının yazarı Greenacre’dir. Analizde kontenjans tablosu halinde sunulan verilerin satır ve sütunlarının birlikte değişimleri incelenir. Kontenjans tabloları, kategorik değişkenler arasındaki ilişkilerin ortaya çıkarılmasında kullanılan tablolardır. Bu tablolarda yer alan değerler değişken kategorilerinin veri kümesinde kaç kez tekrarlandığını gösterir (Çılan 2013:11). Analiz satır ve sütunların iki ya da üç boyutlu uzayda grafiksel gösterimini sağlaması bakımından oldukça kullanışlıdır.

Kontenjans ya da çapraz tablonun gösterimi aşağıdaki tabloda görüldüğü gibidir.

Tablo 5: Kontenjans Tablosu (I x J Boyutlu)

I / J	1	2	j-1	j	$\sum_{j=1}^J n_{ij}$
1	n_{11}	n_{12}	n_{1j-1}	n_{1j}	$n_{1.}$
2	n_{21}	n_{22}	n_{2j-1}	n_{2j}	$n_{2.}$
....
i-1	n_{i-11}	n_{i-12}	n_{i-1j-1}	n_{i-1j}	$n_{i-1.}$
i	n_{i1}	n_{i2}	n_{ij-1}	n_{ij}	$n_{i.}$
$\sum_{i=1}^I n_{ij}$	$n_{.1}$	$n_{.2}$	$n_{.j-1}$	$n_{.j}$	$n_{..}$

İlişkilerin grafiksel gösterimini yapabilmek için kategori profillerinin, marjinal oranların ve ki-kare uzaklıklarının hesaplanması gerekir. İlk ikisi için tablonun satır ve sütunlarında ayrı ayrı işlemler yapılır. Bütün hesapların temelinde kontenjans tablosundaki veriler baz alınır (Çılan 2013).

Profil; tabloda yer alan frekans değerlerinin satır ya da sütun toplamlarına bölünmesiyle elde edilen orandır. Eğer frekans değeri ilgili satırın toplam değerine bölünürse *satır profili*, sütun toplamına bölünürse *sütun profili* denir. Her biri için ayrı ayrı hesaplama yapılır.

Tablo 6: Satır Profilleri (r) Tablosu

I / J	1	2	j-1	j
1	$n_{11}/n_{1.}$	$n_{12}/n_{1.}$	$n_{1j-1}/n_{1.}$	$n_{1j}/n_{1.}$
2	$n_{21}/n_{2.}$	$n_{22}/n_{2.}$	$n_{2j-1}/n_{2.}$	$n_{2j}/n_{2.}$
....
i-1	$n_{i-11}/n_{i-1.}$	$n_{i-12}/n_{i-1.}$	$n_{i-1j-1}/n_{i-1.}$	$n_{i-1j}/n_{i-1.}$
i	$n_{i1}/n_{i.}$	$n_{i2}/n_{i.}$	$n_{ij-1}/n_{i.}$	$n_{ij}/n_{i.}$
$\sum_{i=1}^I n_{ij}/n_{..}$	$n_{.1}/n_{..}$	$n_{.2}/n_{..}$	$n_{.j-1}/n_{..}$	$n_{.j}/n_{..}$

Tablo 7: : Sütun Profilleri (c) Tablosu

I / J	1	2	j-1	j	$\sum_{j=1}^J n_{ij}/n_{..}$
1	$n_{11}/n_{.1}$	$n_{12}/n_{.2}$	$n_{1j-1}/n_{.j-1}$	$n_{1j}/n_{.j}$	$n_{1.}/n_{..}$
2	$n_{21}/n_{.1}$	$n_{22}/n_{.2}$	$n_{2j-1}/n_{.j-1}$	$n_{2j}/n_{.j}$	$n_{2.}/n_{..}$
....
i-1	$n_{i-11}/n_{.1}$	$n_{i-12}/n_{.2}$	$n_{i-1j-1}/n_{.j-1}$	$n_{i-1j}/n_{.j}$	$n_{i-1.}/n_{..}$
i	$n_{i1}/n_{.1}$	$n_{i2}/n_{.2}$	$n_{ij-1}/n_{.j-1}$	$n_{ij}/n_{.j}$	$n_{i.}/n_{..}$

Satır ve sütun profillerinin toplamaları birbirine eşit olmak zorundadır. Her iki profilde uzayda bir noktayı temsil eden birer vektördür. Bağımlılık ilişkisi satır ya da sütun bazında araştırılır. (Greenacre 2017:16). Bir profil değeri ortalamadan ne kadar farklıysa merkezden de o kadar uzaktadır (Clausen 1998). Her bir sütun ya da satır toplamalarının genel toplama bölünmesi sonucunda elde edilen orana kütle (mass) ya da marjinal oranlar denir. Profille aynı şekilde satır ve sütun kütleleri vardır. Örnek olarak ikinci satır için $n_{2j}/n_{2.}$, birinci sütun için $n_{1.}/n_{..}$ ağırlıklarını gösterebiliriz (Çılan 2013). Bir çok veri çok büyük hacimde vektör uzayı içerir. Bu tür değişken yapısını incelemek olanaksız olacağından tam bu noktada boyut indirgeme yöntemi devreye girer.

Profiller arasındaki uzaklıklar hesaplanırken en yaygın şekilde kullanılan yöntem, Öklid uzaklığının özel bir biçimi olan Ki-kare uzaklığıdır. Burada hem satır hem de sütun profilleri ayrı ayrı hesaplanır. Buna göre ağırlıklar hesaplanacaksa, Tablo 5 ve 6 daki oranlar kullanılacaktır. Satır profilleri için hesaplanan ağırlık formülü aşağıdaki gibidir:

$$d_{1I} = \sqrt{\frac{(r_{11}-r_{I1})^2}{n_{.1}/n_{..}} + \frac{(r_{12}-r_{I2})^2}{n_{.2}/n_{..}} + \dots + \frac{(r_{1J}-r_{IJ})^2}{\frac{n_{.J}}{n_{..}}}} \quad (26)$$

Sütun profilleri için hesaplanan ağırlık formülü ise aşağıdaki gibidir:

$$d_{1j} = \sqrt{\frac{(c_{11}-c_{1j})^2}{n_{1.}/n_{..}} + \frac{(c_{21}-c_{2j})^2}{n_{2.}/n_{..}} + \dots + \frac{(c_{I1}-c_{IJ})^2}{n_{I.}/n_{..}}} \quad (27)$$

Buradaki ağırlıklar ortalama profil elemanlarının tersine eşittir. Bundan sonra ise profillerin dağılımının yorumunda kullanılacak olan varyans ya da hareketsizlik (inertia) ölçümü yapılır. Hareketsizlik, profil noktalarının orjin etrafında ne şekilde dağıldığı hakkında bilgi verir. Hareketsizlik satır ya da sütun profilleri arasındaki uzaklıklarının karelerinin ortalaması olarak betimlenir. Hareketsizlik büyük ise bağımlılık yüksek, düşük ise ilişkinin bulunmadığını belirtir (Greenacre 1999). Hareketsizlik değeri aşağıdaki formül ile hesaplanır.

$$\Lambda^2 = \frac{x^2}{n} = \sum_i w_{ri} d_i^2 \quad (28)$$

Eşitlikte w_{ri} satır ağırlıklarını, d_i^2 uzaklıkları göstermektedir. Everitt'inde (2006) belirttiği gibi, Uygunluk analizi beklenen değerlerden saçılımın iki boyutlu grafikte gösterildiği ve hareketsizliğin (inertia) ayırtırmaya tabii tutulduğu bir yöntemdir. Hareketsizliği ayırtırmada kullanılan en önemli yöntem Bölüm 1.6. de anlatılan Tekil Değer Ayırtırmasıdır. Bu sayede satır ve sütun profil değerleri grafiksel yöntemde gösterilir. Sadece kare matrislere uygulanan özdeğer ayırtırmasının genel biçimi olan yöntem bütün matrislere uygulanabilir. Ayrıca tekil değerler özdeğerlerin kareköküne eşittir (Greenacre 2017). Ancak ikiden fazla değişkenin olduğu ilişkilerde Basit Uygunluk analizi yetersiz kalmaktadır. Bu konuda en çok kullanılan yöntemlerden biri de Homojenlik Analizidir.

Gifi, doğrusal olmayan çok değişkenli veri analizinin temel taşı olarak homojenlik analizini belirlemiş ve daha sonra parametrelere çeşitli kısıtlamalar uygulayarak diğer çok değişkenli teknikleri geliştirmiştir (Van der Burg, de Leeuw, ve Verdegaal 1988:178).

Homojenlik Analizinde bulunan değişkenlerin hepsi çoklu nominal ölçek seviyesindedir. Bu nedenle sayısallaştırma işlemi her boyut için yapılacağından Çoklu sayısallaştırma yöntemi kullanılır. HOMALS için kayıp fonksiyon notasyonu aşağıda verilmiştir.

$$\sigma(X; Y) = m^{-1} \sum_j SSQ(X - G_j Y_j) \quad (29)$$

Notasyonların açılımı aşağıdaki gibidir;

X : nesne skorları

Y_j : kategori sayısallaştırmaları vektörü

m : H veri matrisinin sütün değişken sayısı

G_j : j. değişken için Gösterge matrisi

Homojenlik analizinde kayıp fonksiyonun en küçükleme için gerekli kısıtlar aşağıda verilmiştir.

$$\textbf{Kısıtlar} : X'X = nI \text{ ve } u'X = 0 \quad (30)$$

Aşağıda optimal ölçeklemede kullanılan diğer teknikleri anlamak ve kodlamaları yapabilmek adına hepsinin temeli olan Homojenlik Analizi'ni tekniğinin aşamaları küçük ama gerçek bir örnekle açıklanmaktadır. Kocaeli'de eğitim veren bir kurumda 3. sınıf öğrencilerinin kan grubu, cinsiyet ve başarı seviyelerin ilişkin sınıf öğretmeni değerlendirmesi ve öğrenci özellikleri üzerine oluşan bilgiler aşağıdaki tabloda sunulmuştur. Sınıfın mevcudu 15 kişidir. Değişkenlerin kategorileri ve frekansları aşağıda verilmiştir.

cinsiyet	frekans	Kan grubu	frekans	başarı	frekans
Erkek (1)	7	A (1)	5	Kötü (1)	4
Kadın (2)	8	B(2)	4	Orta(2)	6
		AB(3)	2	İyi(3)	5
		0 (4)	4		

Buna göre öğrencilerin bilgilerinin sunulduğu orijinal veri matrisi (H) Tablo 8'de kodlanmış hali Tablo 9'da sunulmuştur.

Tablo 8: H Veri Matrisi (Kategorik hali)

Öğrenci	cinsiyet	Kan Grubu	başarı
1	erkek	A	orta
2	kadın	B	kötü
3	kadın	A	iyi
4	kadın	AB	kötü
5	erkek	0	orta
6	erkek	AB	iyi
7	kadın	B	orta
8	kadın	0	orta
9	erkek	A	iyi
10	kadın	0	orta
11	kadın	B	kötü
12	erkek	A	iyi
13	erkek	B	orta
14	kadın	0	kötü
15	erkek	A	iyi

Tablo 9: H Veri Matrisi (Kodlanmış hali)

Öğrenci	cinsiyet	Kan Grubu	başarı
1	1	1	2
2	2	2	1
3	2	1	3
4	2	3	1
5	1	4	2
6	1	3	3
7	2	2	2
8	2	4	2
9	1	1	3
10	2	4	2
11	2	2	1
12	1	1	3
13	1	2	2
14	2	4	1
15	1	1	3

İlk aşama verilerin kodlanması sürecidir. Bunun ardından sayısallaştırmanın gerçekleştirilmesi için Gösterge Matrisi'nin (G_f) oluşturulması gerekecektir. Bu matris

her j. nesnenin ait olduğu kategoriye dair bilgileri içeren ikili (binary) atamaları gösterir. Eğer nesne (öğrenci) ilgili değişkenin k. kategorisinde tanımlı ise o kategori hücresi 1, diğerleri ise 0 olarak atanır. Her değişkende bir adet 1 ataması yapılır. Buna göre yeni matris $n \times k_j$ boyutta olacaktır. Örneğimizin Gösterge Matrisi Tablo 10'daki gibi oluşmuştur.

Tablo 10: Gösterge Matrisi (G_j)

erkek	kadın	A	B	AB	0	kötü	orta	iyi
1	0	1	0	0	0	0	1	0
0	1	0	1	0	0	1	0	0
0	1	1	0	0	0	0	0	1
0	1	0	0	1	0	1	0	0
1	0	0	0	0	1	0	1	0
1	0	0	0	1	0	0	0	1
0	1	0	1	0	0	0	1	0
0	1	0	0	0	1	0	1	0
1	0	1	0	0	0	0	0	1
0	1	0	0	0	1	0	1	0
0	1	0	1	0	0	1	0	0
1	0	1	0	0	0	0	0	1
1	0	0	1	0	0	0	1	0
0	1	0	0	0	1	1	0	0
1	0	1	0	0	0	0	0	1

Matrisi yorumlamak gerekirse, son öğrencinin cinsiyetinin erkek, kan grubunun A ve de başarı durumunun iyi olduğu görünecektir. Klasik Gösterge Matrisi analiziyle kategoriler ve nesneler sayısallaştırılırken, ters Gösterge Matrisi'ni kullanarak her bireyin kategorilerini ve değişkenlerinin sayısallaştırılması gerçekleştirilebilir.

Gösterge Matrisi'ni üzerinden, C ($C_j = G_j G_l$) İkili Marjinaler Matris ile D Tekli Marjinaler Matrisleri hesaplanır. Burt Tablosu olarak da bilinen C İkili Marjinaler Matrisi analizde yer alan bütün değişkenlere ait kategoriler kadar satır ve sütuna sahip karesel bir matristir. Verilerdeki kategorilerin ikili kombinasyon halinde kaç kez tekrarlandığını sunar. D Tekli Marjinaler Matrisi ise C matrisinin bölünmüş diagonal (köşegen) matrisi olarak tanımlanır; $D = \text{diag}(C)$ (Gifi 1991:68). Tablo 11 ve Tablo 12 sırasıyla C ve D matrislerini göstermektedir.

Tablo 11: C İkili Marjinaller Matrisi

	Erkek	kadın	A	B	AB	0	kötü	orta	iyi
erkek	7	0	4	1	1	1	0	3	4
kadın	0	8	1	3	1	3	4	3	1
A	4	1	5	0	0	0	0	1	4
B	1	3	0	4	0	0	2	2	0
AB	1	1	0	0	2	0	1	0	1
0	1	3	0	0	0	4	1	3	0
kötü	0	4	0	2	1	1	4	0	0
orta	3	3	1	2	0	3	0	6	0
iyi	4	1	4	0	1	0	0	0	5

Tablo 12: D Tekli Marjinaller Matrisi

	Erkek	kadın	A	B	AB	0	kötü	orta	iyi
erkek	7	0	0	0	0	0	0	0	0
kadın	0	8	0	0	0	0	0	0	0
A	0	0	5	0	0	0	0	0	0
B	0	0	0	4	0	0	0	0	0
AB	0	0	0	0	2	0	0	0	0
0	0	0	0	0	0	4	0	0	0
kötü	0	0	0	0	0	0	4	0	0
orta	0	0	0	0	0	0	0	6	0
iyi	0	0	0	0	0	0	0	0	5

Tüm bu matrisler sayısallaştırma işlemlerinde yani Dalgacı En Küçük Kareler algoritmasında kullanılacağından önemlidir. Bu konu önceki bölümde anlatıldığı şekilde iki farklı algoritma (Normalleştirilmiş Ağırlıklar Algoritması ve Normalleştirilmiş Skor Algoritması) ile sayısallaştırma sağlanır. Buna ek olarak değişkenler arasındaki homojenliği en fazla sağlayan nesne skorları (x) ve kategori sayısallaştırmalarını (y) elde etmek için otonormalleştirme sürecine geçilir. Bunun için Tekil Değer Ayrıştırması ya da Gram-Schmidt prosedürü uygulanır. Bu süreçte x ve y için tek bir çözüm oluşur. Ancak sayısallaştırma işlemi tek de olabilir çoklu da olabilir. Eğer her değişken için p kez sayısallaştırma yapılırsa çoklu sayısallaştırma söz konusu olur. Yani bağımsız sayısallaştırmalar isteniyorsa çoklu sayısallaştırma, gerek yok ise

tekli sayısalılaştırma yapılır. Tekli sayısalılaştırmalarda ölçek düzeyi kısıtları mevcut iken, Çoklu sayısalılaştırmalarda kısıt önem arz etmez.

Örneğimizin iki boyuttaki nesne skorları (x) Tablo 13'te, kategori sayısalılaştırmalarını (y) Tablo 14'te gösterilmiştir.

Tablo 13: Nesne Skorları (x)

	x_1	x_2
1	0,77	0,88
2	-1,19	-0,98
3	0,74	-0,41
4	-0,76	-1,99
5	-0,06	1,56
6	0,97	-1,28
7	-0,84	0,45
8	-0,78	1,28
9	1,45	-0,13
10	-0,78	1,28
11	-1,19	-0,98
12	1,45	-0,13
13	-0,13	0,73
14	-1,12	-0,15
15	1,45	-0,13

Tablo 14: Kategori Sayısalılaştırmaları (y)

	y_1	y_2
erkek	0,84	0,21
kadın	-0,74	-0,19
A	1,17	0,01
B	-0,84	-0,19
AB	0,11	-1,63
0	-0,68	0,99
kötü	-1,06	-1,03
orta	-0,30	1,03
iyi	1,22	-0,42

Boyut sayısı istenirse 3 ve daha üzeri olabilirdi ama iki boyutlu olması gerek grafik gösteriminde kolaylık sağlaması açısından gerekse de hesap kolaylığı bakımından literatürde sık olarak tercih edilmektedir. Son olarak son Optimal Ölçekli veri matrisini (Q) bulacağız. Bunun için kategori sayısallaştırmaları tablosundaki ilk boyut kullanılacaktır. H veri matrisindeki her kategori kodunun karşılığına tablodaki kategori puanı atanarak Sayısallaştırma işlemi gerçekleşir. Oluşan Sayısallaştırılmış Matris (Q) Tablo 15’te sunulmuştur.

Tablo 15:Q Veri Matrisi (Sayısallaştırılmış hali)

Öğrenci	cinsiyet	Kan Grubu	başarı
1	0,84	1,17	-0,30
2	-0,74	-0,84	-1,06
3	-0,74	1,17	1,21
4	-0,74	0,11	-1,06
5	0,84	-0,68	-0,30
6	0,84	0,11	1,21
7	-0,74	-0,84	-0,30
8	-0,74	-0,68	-0,30
9	0,84	1,17	1,21
10	-0,74	-0,68	-0,30
11	-0,74	-0,84	-1,06
12	0,84	1,17	1,21
13	0,84	-0,84	-0,30
14	-0,74	-0,68	-1,06
15	0,84	1,17	1,21

Sayısallaştırma sürecini sağlayan kodlar MATLAB programında yazılmıştır. Algoritmanın aşamaları ve kodları aşağıda sunulmuştur.

```

function [Q,G,a]=sayisallastirma(H,itrsyn,xb,fark)
%
[n,m]=size(H);
% -----
%
%   Varsayılan parametereler
%
fark=0.000001;
itrsyn=50;
c=rand(n,1);
xb=c-(sum(c)/n)*ones(n,1);

% -----
%   G; Gösterge matrisi hesaplanır.
%   D1 tekli marjinler matrisinin tersidir.

[n,m]=size(H);
M=max(H);
tplmm=sum(M);
G=zeros(n,tplmm);
for i=1:n,
    jmax=0;
    for s=1:m,
        jmin=jmax+1;
        jmax=jmin+M(s)-1;
        b=zeros(1,M(s));
        b(H(i,s))=1;
        G(i,jmin:jmax)=b;
    end
end
d=sum(G);
D1=diag(d.*(-1));
% -----

if nargin==1,
    itrsyn=default_itrsyn;
end

if nargin>=3
    [n1,n2]=size(xb);
    if n1==n && n2==1,
        x0=xb;
    else
        if n1==1 && n2==n,
            x0=xb';
        else
            disp('*** oncel skor matrisi boyutu hatalidir');
            xb=default_xb;
        end
    end

    xb=default_xb;
end
% -----
%   durağanlık sağlanması

```

```

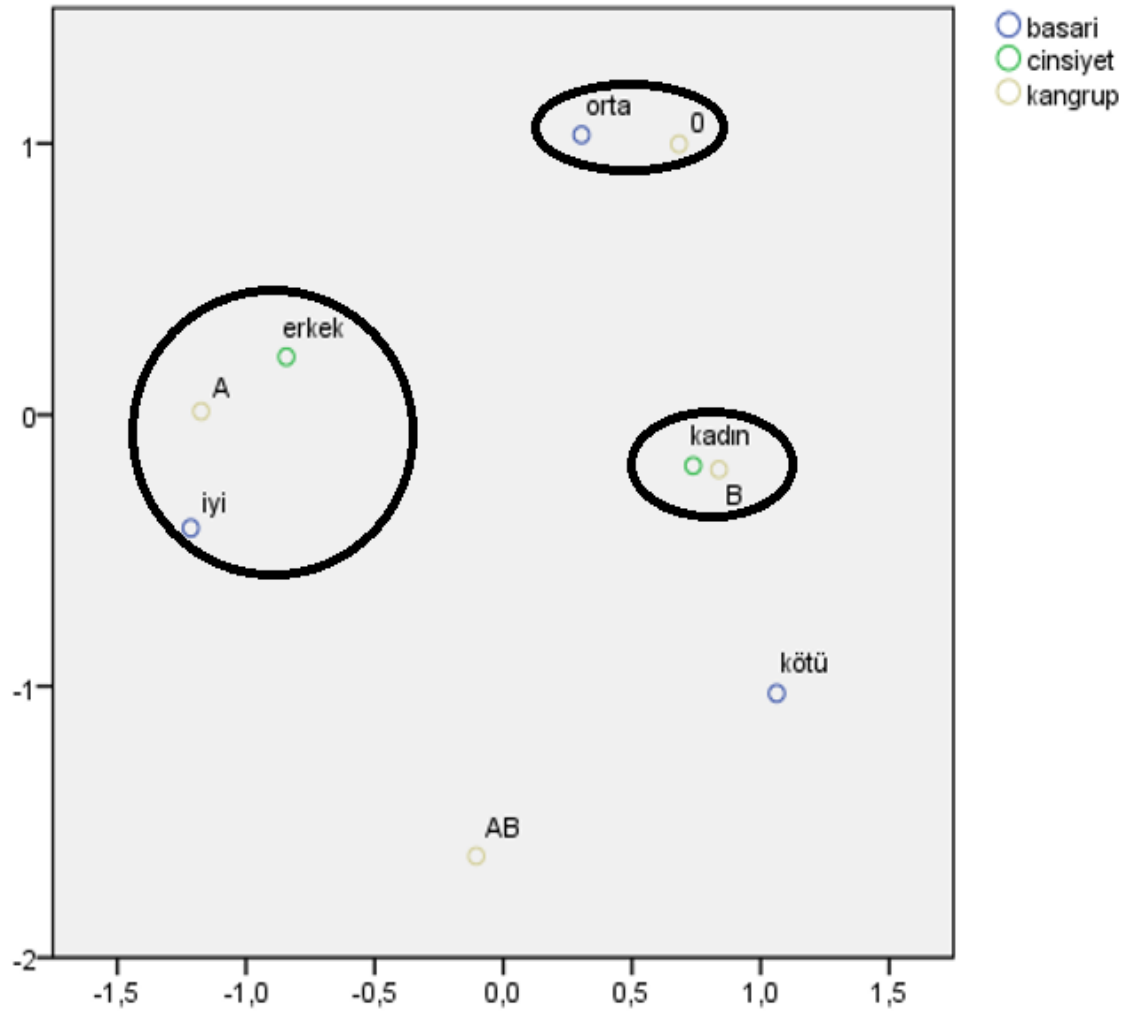
if nargin<4
    fark=default_fark;
end
% -----
%   Tüm yinelemeleri için depo matrislerini başlat.
%
X=zeros(n,itrsyn);
Y=zeros(tplmm,itrsyn);
Alfa=zeros(1,itrsyn);
% -----
%
%   x ve y oluştur
%
x=x0;
y=D1*G'*x;
alfa=norm(x*y'-G,'fro')/m;
X(:,1)=x;
Y(:,1)=y;
% -----
%
%   Iterasyon
%
for i=2:itrsyn
    x=G*y/m;
    lx=norm(x);
    x=sqrt(n)*x/lx;
    y=D1*G'*x;
    alfa=norm(x*y'-G,'fro')/m;
    X(:,i)=x;
    Y(:,i)=y;
    Alfa(i)=alfa;
    if norm(Y(:,i-1)-y)<fark
        disp('*** HOMALS1:Ulaşılan y için yakinsama');
        break
    end
end
num_iter=i;
X=X(:,1:num_iter);
Y=Y(:,1:num_iter);
Alfa=Alfa(:,1:num_iter);
a=Y(:,end);

%   M vektörü, değişkenler ve bu değişkenlerin kategorilerinin
%   sayısını belirtir.

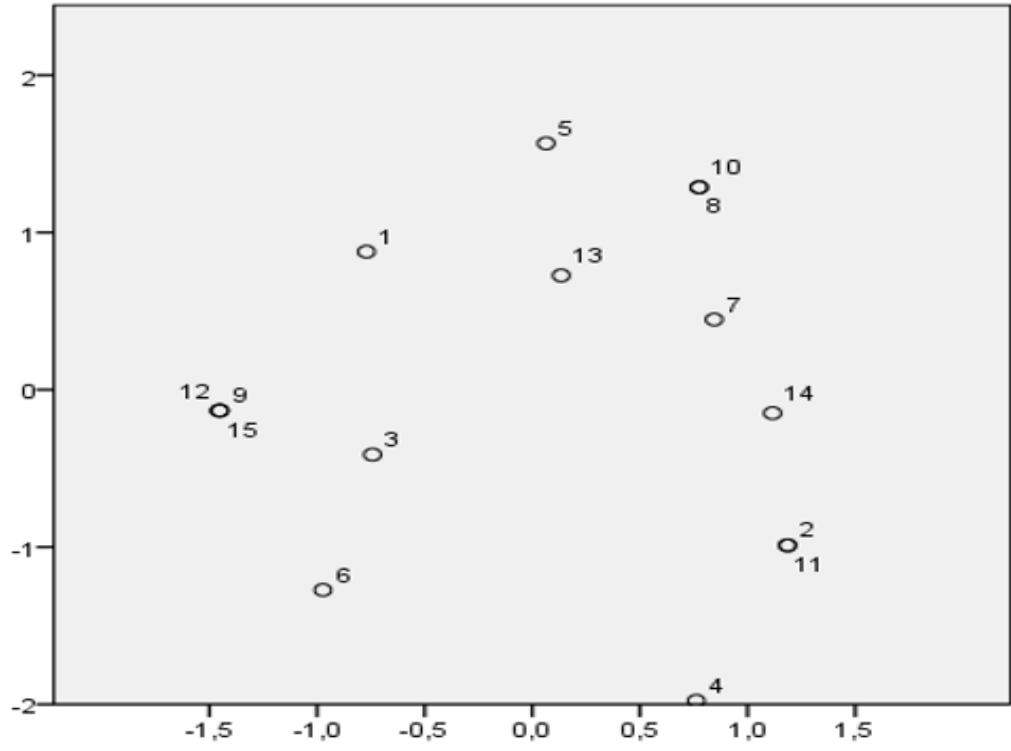
M = [ k1 k2 kn...]; %   ilk değişken n kategorili, ikinci m.
m=length(M);
[n,tplmm]=size(G);
Q=zeros(n,m);
jmax=0;
for alpha=1:m
    jmin=jmax+1;
    jmax=jmax+M(alpha);
    Q(:,alpha)=G(:,jmin:jmax)*a(jmin:jmax);
end
end

```

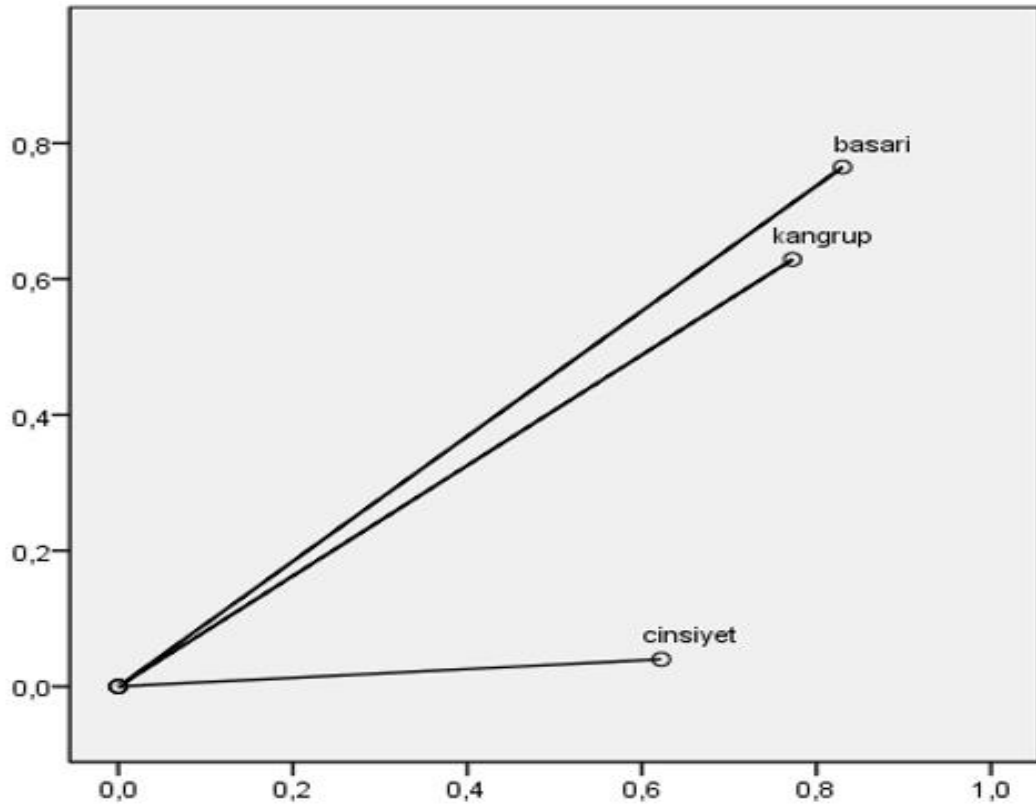
Yapılan dönüşüm sonucunda, aşağıda örnekteki nesnelerin ve kategorilerin Homejenlik Analizi sonucunda elde edilen şekiller (Şekil 3 ve 4) ile değişkenlerin aralarındaki ilişkinin yakınlığını gösteren Ayrım Ölçüsü grafiği Şekil 5'te gösterilmektedir.



Şekil 3: Kategoriler Arasındaki İlişkiler Grafiği



Şekil 4: Nesneler Arasındaki İlişkiler Grafiği



Şekil 5: Ayrım Ölçüsü Grafiği

Şekil 3, 4 ve 5 üzerinden aşağıdaki tespitler yapılabilir.

- Kan grubu A olan erkeklerin başarılıdır.
- 0 kan gruplu öğrenciler orta başarı düzeyindedir.
- B kan grupluların çoğunluğu kadındır.
- 3, 6, 9, 12, 15 numaralı öğrenciler ile diğerleri arasında farklı kümelenmeler oluşmuştur.
- Kan grubu ve başarı arasında yakın bir ilişki mevcuttur. Cinsiyet değişkeni ise diğerlerinden bağımsız görünmektedir.

Örneklem hacmi ve değişken boyutları geliştirilerek daha karmaşık araştırmalar gerçekleştirilebilir. Burada önemli olan nokta değişkenlerin ölçek düzeyinin nominal olmasıdır. Eğer sıralı veya numerik ölçek düzeyleri mevcut ise Doğrusal Olmayan Temel Bileşenler Analizi kullanılmak zorundadır.

Konuya yönelik Literatür taraması yapıldığında, yöntemin tek başına değil daha çok diğer tekniklerle bütünleşik bir tarzda işlendiği görülmektedir.

Bedward vd. (1992), toprak tiplerinin sınıflandırılması ve haritalandırılması üzerine Homojenlik Analizi kullanılarak optimum bir harita ölçeğinin ya da sınıflandırma gruplarının tanımlanabileceğini göstermiştir.

Ata (2007) tarafından gerçekleştirilen araştırmada, Akciğer kanseri hastalarının tedavi sonrası hastalığın nüks edip etmediğini 4 farklı faktör üzerinden değerlendirmek için Homojenlik Analizi Yöntemi kullanılmıştır. Sonuç olarak faktörlerin içindeki kategoriler arasında anlamlı ilişkiler saptanmıştır.

Sahin ve Cigizoglu (2010), 4 değişken üzerinden Türkiye'nin meteorolojik verilerine eksik değer enterpolasyonu ve Homojenlik Analizi uygulamıştır. Sonuç olarak, göreceli ve mutlak testlerin farklı sonuçlar verdiği grafiklerle saptamıştır.

Suner ve Çelikoğlu (2010b), Toplum Tabanlı Bir Çalışmada Çoklu Uygunluk Analizi ve Kümeleme Analizi ile Sağlık Kurumu Seçimi adlı makalesinde, bir Eğitim ve Araştırma Hastanesinde toplanan veriler kullanılarak, hastalık teşhisi konulan 348 kişinin sağlık kurumu seçimini 4 faktör üzerinden belirlemek için Homojenlik

analizini kullanmış ve de bu sonuçları Kümeleme Analiziyle karşılaştırmıştır. Sonuç olarak iki testin sonuçları arasında tutarlılık tespit etmiştir.

Abar ve Karaslan (2013), konut talebi olan Atatürk Üniversitesi personelinin özellikleri ile talep edilen konutun özellikleri arasındaki ilişkileri Homojenlik Analizi ile incelemiştir. Anket uygulaması üzerinden konut talep eden 1959 kişinin demografik özellikleri ve talep tercihleri grafikler kullanılarak yorumlanmıştır.

2.2. DOĞRUSAL OLMAYAN TEMEL BİLEŞENLER ANALİZİ

Davranış ve sosyal bilimler alanında çalışan araştırmacılar, analizler esnasında sık sık büyük sayıda değişkenle karşılaşır. Ancak bu değişkenleri mümkün olduğunca daha az kayıpla daha küçük bileşenlere azaltmayı arzu ederler. Bunun için en çok tercih edilen yöntemlerden biri Temel Bileşenler Analizi (TBA)'dir. Fakat Temel Bileşenler Analizi'nin iki önemli varsayımı vardır. Bunlardan ilki, değişkenler arasındaki ilişkinin doğrusal olması, diğeri ise değişkenlerin aralıksal ya da oransal olarak ölçeklendirilmesi zorunluluğudur. Bilindiği üzere bu varsayımlar maalesef belirtilen alanlarda sağlanamadığından araştırmalar ya sonlandırılmakta ya da varsayımlar ihlal edilmektedir. Bu sınırlamadan kaçınmak üzere doğrusal olmayan bir çok alternatif geliştirilmiştir. Bunlardan biri de Doğrusal Olmayan Temel Bileşenler Analizi (DOTBA)'dir. Yöntem ilk olarak Guttman, (1941) tarafından tanımlanmış, ardından Kruskal (1965), Shepard (1966), Young, Takane (1978) Van Rijckevorsel ve de Leeuw (1979) ile birlikte geliştirilmiş en son hali Gıfı (1990) tarafından formüle edilmiştir.

Kategorik verilere Temel Bileşenler Analizi uygulanmak istendiğinde Dalgacı En Küçük Kareler yöntemi devreye girer ve yöntem "Doğrusal Olmayan Temel Bileşenler Analizi"(DOTBA) olarak tanımlanır. Bununla birlikte metoda Kategorik Temel Bileşenler Analizi (CATCPA) ismi de konulmuştur (Linting vd., 2007:337). Eğer bütün değişkenler çoklu nominal Ölçekli ise Doğrusal Olmayan Temel Bileşenler Analizi, Homojenlik Analizi ile aynı sonuçları verir. Bununla birlikte değişkenler nümerik ise Doğrusal Olmayan Temel Bileşenler Analizi, Doğrusal Temel Bileşenler ile aynı sonuca ulaşacaktır (Filiz ve Çemrek 2007:211). Bununla birlikte nümerik değişkenler içeren verilere sadece Doğrusal Temel Bileşenler Analizi uygulamak tam

olarak doğru değildir. Örneğin yaş ve gelir arasındaki ilişki incelendiğinde (her ikisi de numeriktir) aralarında ki ilişki doğrusal olmayabilir (Linting vd., 2007:338). Genç ve yaşlı insanlar, 30-60 yaş arası yetişkin insanlardan daha az bir gelire sahip olduğu kabul edildiğinde oluşturulacak grafiğin yatay eksenini yaş, dikey eksenini gelir değişkeni olarak düşünülürse grafik doğrusal olmayan bir fonsiyon eğrisini gösterir. Hatta eğrinin “n” şeklinde çıkacağını tahmin edilir. Bu durumda tutarlı bir sonuç için bu değişkenler (yaş için 3'lü) kategorik hale getirilmelidir. Böylece Yöntem tutarlı sonuçlar elde edebilmek amacıyla Doğrusal Olmayan Temel Bileşenler Analizi'ne dönüşür. Sonuç olarak analiz tüm değişken tipleri için kullanılabilir.

Daha öncede belirtildiği gibi, bu tekniğin temelini de Homojenlik Analizi notasyonu oluşturur.

$$\sigma(X; Y) = m^{-1} \sum_j SSQ(X - G_j Y_j) \quad (31)$$

Farklı ya da ek olarak Doğrusal Olmayan Temel Bileşenler Analizi'nde çoklu değil, tekli sayısallaştırmalar yapılır. Bu kısıtı sağlamak üzere kayıp fonksiyonda yer alan Y_j değişkeni üzerine $Y_j = a_j y_i$ eklentisi yapılır. Böylece çoklu sayısallaştırmalar matrisi 1 rank sınırlaması eklenerek tekil hale dönüştürülür. Çoklu sayısallaştırmanın amacı, bir bütün olarak bir değişkeni temsil etmekten ziyade, bu değişken ve diğerleri arasındaki kategoriler arasındaki ilişkilerin doğasını optimal bir şekilde açığa çıkarmaktır. Bu amaç her değişken için ayrı ayrı bir sayısallaştırma yapılarak başılır. Bu nedenle çoklu sayısallaştırmalar sadece nominal sınıflı ölçeklere uygulanırken, diğer ölçek türleri için tekli sayısallaştırmalar yapılır.

Bu analiz tekniği de literatürde araştırmacıların ilgisini çekmektedir. Birçok ileri araştırma tekniğinin önsel boyut indirgeme aşamasında temel olarak kullanılmasının yanı sıra tek başına da kullanıldığı görülür.

Young ve diğerleri, (1978) çeşitli ölçek düzeyine sahip veriye (sanal), Dalgalı En Küçük Kareler Algoritmasını uygulayarak Optimal Ölçekleme gerçekleştirmiş ve geçerliliğini sınamıştır.

Bekker ve Burg 1988 yılında gerçekleştirdikleri çalışmalarda Doğrusal Olmayan Temel Bileşenler Analizi ile Varyans Analizi arasındaki ilişkileri incelemişlerdir.

Domínguez Lozano ve arkadaşları (2003), nadirlik ve tehdit seviyelerine ilişkin faktörleri tanımlamak ve açıklamak için Iberia'nın tehdit altındaki florasını analiz etmiştir. Flora havuzundan (588 tür) dağılmış 59 bitki türünü rastgele seçerek 12 değişken üzerinden Doğrusal Olmayan Temel Bileşenler Analizini kullanmıştır. Sonuçta, yazarlar dört nadir bitki sınıfı üretmiş ve bu grupların her biri için bazı özel stratejiler önermişlerdir.

Tomlinson (2003), 1984 ve 1985 yılları arasında İngiltere’de 9003 yetiştikine uygulanan Sağlık ve Yaşam Tarzları Araştırması'ndaki kategorik verilere Doğrusal Olmayan Temel Bileşenler Analizini uygulamış ve ilk olarak davranış kalıplarının kümelenmesini göstermek ve ikinci olarak da bu kümelenmelerin sınıf ve cinsiyet gibi geleneksel sosyal kategorilerle güçlü bir şekilde ilişkili olup olmadığını araştırmıştır.

Linting ve diğerleri (2007), SPSS 2004 programını kullanarak ampirik örnekler üzerinden CATPCA modülünü gerek teorik gerekse de pratik yönünü ele almıştır.

Altaş ve Kandur (2018), Türkiye’deki amatör düzeydeki koşucuların çeşitli yarış mesafelerine ve türlerine ilişkin tercihlerini belirlemeye amacıyla İdeal Nokta Modeli ve Vektör Modeli tekniklerine kullanarak amatör koşucuların yarış türlerine ilişkin tercih haritalarını belirlemiştir. Vektör Modeli tekniğini kullanan araştırmacılar Doğrusal Olmayan Temel Bileşenler Analizi algoritmasından yararlanmışlardır.

Karaman ve Arıcıgil Çılan (2018), Uluslararası Fen ve Matematik Eğilimleri Araştırması (TIMSS) sınavı 2015 uygulamasına Türkiye’den katılan 4. sınıf öğrencilerine uygulanan tutum ve bilgi anketi verilerini Doğrusal Olmayan Temel Bileşenler Analizi (DOTBA) ile analiz etmiştir. Buna göre öğrencilerin sahip oldukları teknolojik destek miktarı ve materyal çeşitliliği arttırıldıkça öğrencilerin tutum dereceleri ve başarı seviyelerinin de doğru orantılı olarak arttığı gözlemlenmiştir.

2.3 .KATEGORİK REGRESYON ANALİZİ

Kategorik değişkenleri sayısallaştırarak regresyon analizi yapmak son yıllarda önemli bir araç haline gelmiştir. Bu konuda çeşitli program ve modeller mevcuttur. Örneğin; Box-Cox Model (Box ve Cox 1964), Monoton dönüşümler (Kruskal 1965), MORALS (Young 1976), Spline dönüşümler (Ramsay 1988), İzdüşüm Takip Regresyonu (Friedman ve Stuetzle), ACE (Breiman ve Friedman 1985), GAM (Hastie ve

Tibshirani 1990) ve Yapay Sinir Ağları (Gaudart ve diğerleri 2004) bunların başlıcalarıdır (van der Kooij, Meulman, ve Heiser 2006:16). Şu an yaygın bir şekilde kullanılan Kategorik Regresyon (CATREG) analizinin temelini Gifi (1990) oluşturmuş ve öğrencileri Kooij, Meulman ve Heiser (2006) daha ayrıntılı çalışmalar geliştirerek, yöntemi günümüz uygulamaları için bir çok alanda tercih edilebilir bir hale getirmiştir.

Kategorik Regresyon (CATREG) analizi, hem bağımlı hem de bağımsız değişkenin kategorik olduğu ve de bağımsız değişkenlerin aralarında doğrusal olmayan ilişkilerin bulunduğu durumlarda uygulanabilen bir analiz tekniğidir. CATREG klasik regresyon analizine göre daha az varsayım gerektirir. Ayrıca kategorik verilerin çok sık kullanıldığı sosyal bilimler alanındaki çalışmalar içinde oldukça uygundur (Güç ve Başar, 2016:17). Analizde kategorik veriler yukarıda da anlatıldığı gibi, Dalgali En Küçük Kareler (Alternating Least Squares of Optimal-ALS) tekniği vasıtasıyla sayısal dönüşümlere uğrar. Böylece değişkenler kendi karakteristik özelliklerini yansıtacak şekilde evrilir (Shrestha, 2009:207). Sayısallaştırma yapılırken, en uygun doğrusal regresyon denkleminde ulaşılması amacı aranır. Yani optimal regresyon modelini elde etmek için bir çok doğrusal olmayan dönüşümlere başvurulur. Bu dönüşüm sayesinde bağımlı değişken ile bağımsız değişkenlerin her biri arasındaki ilişki en iyilenecektir (Altaş ve Giray, 2013:138).

Homojenlik Analizi ve Kategorik Regresyon Analiz'leri karşılaştırıldığında, ilki sadece çoklu nominal ölçme düzeyinde değişkenleri baz alan ve hipotez testlerini kullanmayan bir teknik iken, Kategorik Regresyon Analizi değişik ölçme düzeylerine sahip değişkenleri kullanan ve hipotez testleri uygulanan bir tekniktir. Bununla birlikte grafiksel analiz bakımından Homojenlik Analizi daha kullanışlıdır (Altaş ve Giray, 2013:140). CATREG modeli, sayısallaştırılan değişkenlere uygulanan klasik Lineer Regresyon modelidir. Tüm değişkenlere sayısallaştırma uygulandığında CATREG sonucu Lineer Regresyon ile aynı sonuca ulaşır (Van der Kooij vd., 2006:18). Ayrıca CATREG, Lineer Regresyona oranla, göreceli olarak küçük örneklerle uygulanabilen ve tutumsal veriler için çok uygun bir yöntemdir (Anastasios, Koutsouris, ve Konstadinos, 2010).

Kategorik Regresyon'da kayıp fonksiyonu aşağıdaki gibidir.

$$\sigma(y_r; y_j; b_j) = \|G_r y_r - \sum_{j=1}^m b_j G_j y_j\| \cdot \|G_r y_r - \sum_{j=1}^m b_j G_j y_j\| \quad (32)$$

Notasyonların açılımı aşağıdaki gibidir;

y_r : Bağımlı değişkenin kategori sayısallaştırmaları

y_j : Bağımsız değişkenin kategori sayısallaştırmaları

b_j : j. değişken için regresyon katsayısı

G_j : Bağımsız değişkenin Gösterge matrisi

G_r : Bağımlı değişkenin Gösterge matrisi

Dikkat edilirse notasyonda $\| \|$ Öklit normu kullanılmıştır. Bunun nedeni sayısallaştırılan değişkenlerin merkezileştirilerek kareler toplamın N olacak biçimde normlaştırmaktır (Van der Kooij vd., 2006:18).

Kategorik Regresyon Analizinde, Dalgalı En Küçük Kareler (Alternating Least Squares of Optimal-ALS) tekniği sayısallaştırmaları iki şekilde gerçekleştirir. İlk adımda, regresyon katsayıları ve bağımsız değişkenin sayısallaştırma vektörü sabit tutularak bağımlı değişken sayısallaştırma vektörü bulunur. İkinci adımda ise, bunun tam tersi yapılır. Yani bağımlı değişkenin sayısallaştırma vektörü sabit tutularak regresyon katsayıları ve bağımsız değişkenin sayısallaştırma vektörü hesaplanır (Güç, 2015:30).

İlk adımda bağımlı değişkenin sayısallaştırma aşamasında ölçüm düzeyinin dikkate alınması gerekir. Bilindiği gibi regresyon analizlerinde bağımlı değişkenlerin ölçüm düzeyi genellikle nominal ölçeklidir. Bu nedenle sayısallaştırma işlemi çoklu olarak yapılacağından vektör y_r biçiminde gösterilir. Eğer numerik ya da ordinal ölçekte bağımlı değişken söz konusu ise, y_r 'ye bazı kısıtlar eklenmesi gerekecektir. Mesela ordinal ölçekte değişkenler için ağırlıklandırılmış Monotonik Regresyon Yöntemi (WMON), numerik ölçekte değişkenler için ağırlıklandırılmış doğrusal regresyon yöntemi (WMON) kullanılarak sayısallaştırmalar gerçekleştirilir. Buna göre bağımlı değişken sayısallaştırma işlemi;

$$y_r = D_r^{-1} G_r' \sum_{j=1}^m b_j G_j y_j \quad (33)$$

formülasyonu ile hesaplanır. Burada $D_r = G_r' G_r$ bağımsız değişkenlerin kategorilerine ait tekli marjinal matrisidir.

İkinci adımda ise ilgili bağımsız değişkenlerin bağımlı değişkeni açıklama oranını veren v vektörü hesaplanır.

$$v = \sum_{j=1}^m b_j G_j y_j \quad (34)$$

ardından bu vektörden j 'inci değişkenin tahmin değer katkısı çıkartılarak sayısallaştırmaya (y_j) geçilir.

$$v_j = v - b_j G_j y_j \quad (35)$$

$$y_j = (D_j^{-1} G_j') (G_r y_r - v_j) \quad (36)$$

Regresyon katsayıları (b_j) aşağıdaki formülle elde edilir.

$$b_j = N^{-1} \hat{v}_j D_j y_j \quad (37)$$

Son olarak her değişkenin güncel katkısı tahmin değeri eklenir.

$$v = v_j b_j G_j y_j \quad (38)$$

değeri hesaplanır. bütün bağımsız değişkenlere güncelleme yapılarak regresyon katsayıları ve kategori sayısallaştırmalarına devam edilir. Bu hesaplamalar belirli bir durağanlığa ulaşılan kadar devam eder, eğer fark anlamsızlaşırsa döngü durdurulur (Van der Kooij vd. 2006).

Literatürde CATREG Analizi'ni kullanan çalışmalar aşağıda sunulmuştur.

Kraaij ve Wilde (2001), Geriatrik Depresyon Ölçeği ve Olumsuz Yaşam Olayları Anketi ile 194 yaşlı toplum örneklemiyle röportaj yaparak, olumsuz yaşam olaylarının insanların refahı için uzun vadeli sonuçları olup olamayacağını CATREG Analiziyle incelemiştir. Yaşlılıkta moral bozukluğunun nedenleri arasında, çocukluk dönemindeki olumsuz sosyo-ekonomik durumların, duygusal kötüye kullanım ve ihmalin, olumsuz sosyo-ekonomik durumların, cinsel istismarın, duygusal istismar ve ihmalin, ilişkisel stres ve problem davranışların olduğunu belirlemiştir. Öneri olarak da yaşam öykülerinin tanısal görüşmeye dahil edilmesini sunmuşlardır.

Dusseldorp ve Meulman (2001), doğrusal regresyonda etkileşim terimlerini seçme ve de regresyon ağaçlarının önemini istatistiksel olarak test etme problemlerini çözmek için yeni bir veri analizi stratejisi önermiştir. Önerilen strateji iki veri madenciliği tekniğini birleştirmektedir: regresyon ağaçları ve optimal ölçeklendirmeye regresyon analizi (CATREG). Kardiyak hastalardan elde edilen verilere yapılan bir uygulama üzerinden önermenin verimliliği test edilmiştir.

Shrestha (2009), Nepal Sağlık Araştırma Konseyi'nin yayınladığı hane halkı verileri üzerinden ev içinde ya da mutfakta oluşan hava kirliliğini tahmin etmek amacıyla Kategorik Regresyon Analizini kullanarak 4 bağımsız değişken üzerinden havadaki partekül (PM_{10}) miktarını ve karbon monoksit (CO) seviyesini tahmin etmeyi başarmıştır.

Capretz ve Ho (2010), Açık Kaynak Yazılım (Open Source Software) tahmin modelini Kategorik Regresyon (CATREG) ile geliştirmiştir. Burada değişkenler tek tek sayısallaştırılarak step-wise lineer regresyonu uygulanmış ve CATREG sonuçlarıyla karşılaştırmıştır. Sonuç olarak her iki regresyon modelinde de aynı bağımsız değişkenler anlamlı çıkmış ve modelin açıklama oranı birbirine yakın çıkmıştır.

Lima ve diğerleri (2010), uçucu parteküllerin içerisinde bulunan ağır metallerin (Cl, Ca, Cr, Cu, Zn, ve Pb) ayrıştırılmasını sağlayan Elektroliz Sürecine (EDR) etki eden değişkenlerin incelemesinde CATREG metodunu kullanmış ve en önemli değişkenlerin kül tipi ile anot-katot arası göç hızı olduğunu saptamıştır.

Graham ve arkadaşları (2011), Batı Pasifik kıyısındaki mercan resiflerin kurtarılması için gerekli olacak sürücülerin göreceli önemini değerlendirmek için CATREG analizini kullanmıştır.

Altaş ve Giray 2013 yılında yaptıkları çalışmada, World Values Survey (WVS) 2008 verilerini kullanarak, dünyadaki en önemli problem hakkındaki düşünceleri etkileyen sosyo demografik değişkenleri Kategorik Regresyon ile belirlemiş ve bu kategoriler arasındaki ilişkileri Homojenlik Analizi grafiği ile yorumlamışlardır. Çalışma sonucunda; gelir düzeyinin artması durumunda çevreye verilen önemin arttığı; genç-

orta yaş grubunun eğitim sorununu diğerlerine oranla daha çok önemseydiği bulgularına ulaşmışlardır.

Almanya menşeli çalışmada (Hartmann, Zeeck, ve Barrett 2010), yemek bozukluklarının kişiler arasında problemlere yol açıp açmadığını belirlemek üzere CATREG analizi uygulanmış, bir başka çalışmada da (Gundacker vd., 2017) hamilelik esnasında toprak yeme alışkanlığının doğacak bebekler için bir risk oluşturma durumu CATREG analiziyle incelenmiştir.

Çılan ve Can (2014), İstanbul Üniversitesi İşletme Ekonomisi Enstitüsü'nde öğrenim gören MBA öğrencilerinin performansını etkileyen bazı faktörleri belirlemek istemiştir. Bu değişkenler; demografik, sosyal, bireysel ve örgütsel faktörler olarak belirlenmiştir. Verilerin çoğu kategorik olduğu için Kategorik Regresyon Analizi (CATREG) kullanılmış ve sonuç olarak, yaş başarıyı etkileyen en önemli değişken olarak saptanırken medeni halinde MBA öğrencisinin başarısı üzerinde önemli bir etkiye sahip olduğu belirlenmiştir.

2016 yılında yapılan çalışmada, büyük bir ekonomik kriz geçiren Yunanistan'da, buhran sürecinde nikotun bağımlılığına etki eden faktörleri belirlemede CATREG modeli tahmin edilmiştir. Burada en gözen çarpan ayrıntı ise kadınların bu dönemde sigara kullanma oranının ve bağımlılık seviyesinin erkeklere oranla daha yüksek çıkmasıdır (Dimitrios, 2016).

Almeida ve Garrod (2018), Atlantik'in Avrupa kıyılarında bulunan yarı Portekiz sömürgesi Madeira adasına gelen turistlerle ilgili yapılan çalışmada turistlerin adayı tercih etmelerine etkise olan değişkenleri CATREG modeli ile test etmiştir. Adanın cazibesinin bağımlı değişken olarak belirlendiği modelde 28 adet bağımsız değişken mevcuttur. Bunlardan 8 tanesi anlamlı çıkmış ve en önemlisi ise ada halkının turistlere olan tutumu olarak belirlenmiştir.

2.4. DOĞRUSAL OLMAYAN KANONİK KORELASYON ANALİZİ (OVERALS)

Doğrusal Olmayan Kanonik Korelasyon Analizi çözüm tekniği ilk olarak 1981 yılında Gifi ve 1984 yılında Van der Burg de Leew ile Veerdegaal tarafından ortaya konmuştur. OVERALS ya da CANALS adıyla bilinen Doğrusal Olmayan Kanonik Korelasyon Analizi içinde birden çok değişkenin bulunduğu iki veya daha fazla küme arasındaki ilişkileri inceler. Analiz, değişik ölçüm düzeylerdeki değişkenlerin doğrusallığı ya da dağılımı ilgili herhangi bir varsayımda bulunmaz. Numerik değişkenlerle birlikte kategorik değişkenleri de analize katması ve de bu değişkenleri iki boyutlu haritalarda konumlandırması, analizinin en önemli avantajlarından biridir. Bununla birlikte Doğrusal Olmayan Kanonik Korelasyon Analizi'nde de, Doğrusal Olmayan Temel Bileşenler Analizinde olduğu gibi, değişkenlerin ölçüm düzeylerinin farklı olması sorun teşkil etmez. Yani analizde sıralayıcı, sınıflayıcı ve numerik ölçüm düzeylerinde ait değişkenler kullanılabilir (Bayram ve Ertaş, 2001:40). Değişkenlerin hepsi sayısal ise Doğrusal Olmayan Kanonik Korelasyon Analizi, Kanonik Korelasyon Analizi ile aynı sonucu verir.

Bülbul ve Giray'ın (2012:108) belirttiği gibi Doğrusal Olmayan Kanonik Korelasyon Analizi, kayıp fonksiyonunu belli kısıtlar altında minimize eden bir optimizasyon problemi olarak betimlenebilir. Gifi tekniklerinin temelini oluşturan, Homojenlik Analizi; Doğrusal Olmayan Kanonik Korelasyon Analizi'nin, her sette sadece bir çoklu nominal değişkenin bulunduğu özel bir durumudur (Van de Geer, 1993:95). Ayrıca, birden çok değişken setinin her birinde ölçek düzeyleri farklı kategorik değişken bulunduğu koşulda ise Doğrusal Olmayan Kanonik Korelasyon Analizi, Doğrusal Olmayan Temel Bileşenler analizine dönüşür (Giray 2001:116).

Doğrusal Olmayan Kanonik Korelasyon Analizi için kayıp fonksiyon aşağıdaki gibi olacaktır.

$$\textbf{Kayıp Fonksiyon} : \sigma(x, y) = K^{-1} \sum_j SSQ(x - \sum_{j,k} G_j y_j) \quad (39)$$

K küme sayısını; j_k , k. setteki değişken sayısını ifade etmektedir.

Literatürde Doğrusal Olmayan Kanonik Korelasyon analiziyle ilgili çalışmalar incelendiğinde konuya olan ilginin paket programların gelişmesiyle birlikte giderek

arttığı gözlenir. Daha çok Hollanda merkezli olan çalışmalar zaman ilerledikçe ilgi odağı olmaya başlamış, ülkemizde de 2000’li yılların başından itibaren araştırmalarda kullanılmıya başlanmıştır.

Golob (1986), Hollanda’da yaşayan 3800 kişiden oluşan 1700 aileye seyahat yapmalarında etkili olan haneye (20 adet değişken) ve kişiye ait demografik özellikleri (7 adet değişken) dikkate alarak bunlar arasındaki ilişkileri 3 meslek grubu segmentine ayırarak Doğrusal ve Doğrusal Olmayan Kanonik Korelasyon Analizi ile incelemiştir. Sonuç olarak en önemli etkinin yaşam biçimi olduğu ve ardından yaş ve gelir değişkenlerinin geldiğini belirlemiştir.

Van der Burg ve De Leeuw (1988) Delta Tekniği ile çözümün uygun olmadığı bazı durumlarda Doğrusal Olmayan Kanonik Korelasyon Analizi kategori ölçeklerinin, özdeğerleri ile standart hatalarını, Bootstrap ve Jackknife yöntemlerini kullanarak tahmin etmiş ve iki yöntemi karşılaştırarak değerlendirmede bulunmuşlardır. Sonuç olarak Bootstrap metodunun Jackknife metodundan daha iyi olduğunu vurgulamışlardır.

Bohacek ve Jonckheere (1998) yüzey elektromyografik (sEMG) sinyallerinin doğrusal olmayan matematiksel analizini Doğrusal ve Doğrusal Olmayan Kanonik Korelasyon Analizi ile modellemiş, doğrusal olmayan öngörücülerin, doğrusal öngörücülerden % 30 daha iyi performans gösterdiğini analiz etmişlerdir.

Bayram ve Ertaş (2001), Bursa’da yaşayan 953 hane halkı üzerinden tüketim harcamaları verilerine Doğrusal Olmayan Temel Bileşenler Analizi (PRINCALS) ve Doğrusal Olmayan Kanonik Korelasyon Analizi (OVERALS) yöntemlerini uygulayarak, tüketici harcama düzeni ve biçimini belirlemeye çalışmışlardır. Sonuç olarak PRINCALS ve OVERALS’ın aralarında çoklu doğrusal bağıntılı bulunan değişkenler arasındaki ilişkileri incelemede başarılı olduğu sonucuna varılmıştır.

Hsieh 2000 yılında gerçekleştirdiği çalışmada Doğrusal olmayan bir kanonik korelasyon analizi yöntemini üç ileri beslemeli sinir ağı metodunda kullanarak sentetik veriler yerine gerçek verilerle DOKKA’nın kullanılabileceğini göstermiştir.

Süt (2001), 3 farklı set üzerinden yaptığı çalışmada setler arasındaki ilişkileri Doğrusal Olmayan Kanonik Korelasyon Analizi ile incelemiştir. Birinci sette cinsiyet, medeni

durum, unvan, idari görev ve görev yeri değişkenleri, ikinci sette sigara ve alkol kullanımı değişkenleri, üçüncü sette ise güvensizlik, endişe, mutluluk ve hassasiyet değişkenleri ele alınmıştır. Sonuç olarak; cinsiyet ve alkol kullanımı değişkenleri ile hassasiyet değişkeni arasında ve unvan, görev yeri ve mutluluk değişkenleri ile sigara kullanımı değişkeni arasında ilişki bulunmuştur.

Kumar vd. (2002), çalışmalarında Doğrusal Olmayan Kanonik Korelasyon Analizi için yeni bir yöntem olan radyal tabanlı bir fonksiyon önermiş, bu metodun hataların belirlenmesinde ve teşhisinde kullanmanın faydalı olacağı görüşünü öne sürmüşlerdir.

Wu William ve Hsieh (2001), Tropikal Pasifik'teki deniz yüzey sıcaklığı ve rüzgar etkisini tespit etmek için sinir ağları yaklaşımı ile Doğrusal Olmayan Kanonik Korelasyon Analizini modellemişlerdir. Sonuçta kendi geliştirdikleri Doğrusal Olmayan Kanonik Korelasyon Analizini sonuçlarının korelasyon ilişkilerini Doğrusal Kanonik Korelasyon Analizinden daha iyi derecede bulduğu tespitinde bulunmuşlardır.

Wang vd. (2005), yapmış oldukları çalışmada fMRI verilerinde nöral aktivasyonu saptamak için doğrusal olmayan kanonik korelasyon analizinin, Doğrusal Kanonik Korelasyon Analizine oranla algılama performansını daha iyi gösterdiğini tespit etmişlerdir.

Cannon (2006) yılında yaptığı çalışmada, Doğrusal Olmayan Ana Öngörücü Analizi (NLPPA) gerçekleştirmek için yeni bir sinir ağı yaklaşımını ortaya koymuştur. Lorenz sistem denklemlerine NLPPA uygulanmış ve Doğrusal Olmayan Kanonik Korelasyon Analizi ve doğrusal çok değişkenli modellerle karşılaştırılmıştır. Sonuç olarak veri kümeleri gürültüyle bozulduğunda NLPPA'nın Doğrusal Olmayan Kanonik Korelasyon Analizi'den daha iyi performans gösterdiği bulunmuştur.

Girginer vd., 2007 yılında Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi'nin İşletme, İktisat ve Maliye bölümlerine kayıtlı 257 öğrenciye, Aiken (1979) tarafından geliştirilen 24 maddelik İstatistik Tutum Ölçeği anketini uygulayarak, istatistiğe yönelik tutumlarda üniversite öğrencileri arasındaki bireysel farklılıkları Doğrusal Olmayan Kanonik Korelasyon Analizi ile incelemişlerdir. Sonuç olarak öğrencilerin istatistik dersine karşı tutumlarını etkileyen en önemli nedenlerin İstatistik 1 ve İstatistik 2 dersindeki başarı notları, bu derslerin tekrar sayısı ve tekrar

ediliş nedenleri olarak belirlenmiş ayrıca demografik değişkenlerin derslere ilişkin tutumla anlamlı bir ilişkisi olmadığını tespit edilmiştir.

Theodosiu vd. (2008), farklı belge temsillerinin değişkenleri arasındaki korelasyonları saptamak amacıyla dokümanları tek bir yeni veri setiyle açıklamak adına Doğrusal Olmayan Kanonik Korelasyon Analizini kullanmıştır. Ayrıca yazarlar analiz, kümeleme ve sınıflandırma yöntemlerinde ilk aşama olarak kullanılabileceğini önermişlerdir.

Frie ve Janssen (2009), sosyal eşitsizlik, sağlık ve yaşam biçimi üzerine 695 kişiyle yaptıkları çalışmada Doğrusal Olmayan Kanonik Korelasyon Analizini kullanarak, sağlık özellikleri sosyo ekonomik koşullar ve yaşam tarzlarıyla birlikte analiz etmişlerdir. Sonuçta, setler arasındaki uyumun 1.444 olduğu ve özdeğerlerin varyansın % 50 sini açıkladığı görülmüş, değişkenler arasındaki ilişkiler üç boyutlu uzayda görüntülenmiştir. Analiz sayesinde, Bourdieus teorik yaklaşımına dayanarak bu değişkenler arasındaki karmaşık korelasyonların daha ayrıntılı bir şekilde yorumlanabilir ve sunulabilir olduğu görüşünü ileri sürmüşlerdir.

Yazıcı vd. (2010) yılında gerçekleştirdiği ampirik çalışmada K adet kategorik değişken kümesi arasındaki ilişkileri ve veri kümelerinin yapısal benzerliklerini DOKKA ile inceleyerek yöntemin tıp alanında uygulanabilirliğini araştırmışlardır.

Bülbül ve Giray (2011), sosyodemografik özellikler ile mutluluk algısı arasındaki ilişki derecesini, hangi kategorilerin birbirleri ile ilişkili oldukları, bunun yanı sıra hangi sosyodemografik özelliklerdeki bireylerin ne derece mutlu oldukları ve bu mutlulukların kişi ve kavram bazında kaynakları açısından homojen kümeler oluşturup oluşturmadıklarını araştırmak için Doğrusal Olmayan Kanonik Korelasyon Analizi kullanmışlardır.

Filiz ve Kolukısaoglu (2012), çalışmalarında müşterilerinin bir hizmet işletmesinden aldıkları hizmetten memnuniyet düzeylerini belirleme amacı ile toplanmış olan veriler üzerinden Doğrusal Olmayan Kanonik Korelasyon Analizi kullanmışlardır. Yapılan analiz sonucunda serbest meslek sahibi kişilerin servis personelinin görünüşünden ve odaların ortamı, donanımı ve fiziksel görüntüsünden memnun oldukları görülmüştür.

Sunulan yemeğin lezzet ve temizliğinden oldukça memnun olanların aynı zamanda sunulan yemeğin çeşitliliğinden de oldukça memnun oldukları sonucuna ulaşılmıştır.

Aktürk (2015), çalışmasında organik ve geleneksel ürünlere karşı tüketici talepleri arasındaki farkları, Doğrusal Olmayan Kanonik Korelasyon Analizi yöntemi ile incelemiştir. Sonuç olarak eğitim düzeyi yükseldikçe organik ürün tüketiminin arttığı saptanmıştır.

Akça ve Işıkhani (2017), Hacettepe Üniversitesi Sosyal Bilimler Meslek Yüksekokulu'nda gerçekleştirdikleri çalışmada 617 öğrenciye uyguladıkları ankette, kişilerin demografik özellikleri ile iş değerleri ve kariyer seçimleri arasındaki ilişkiyi ortaya koyabilmek için Doğrusal Olmayan Kanonik Korelasyon Analizi'nden (OVERALS) yararlanmışlardır.

Kazan ve Karaman (2018), farklı departmanlarda görevli çalışanların demografik özelliklerine göre iş ve sosyal hayat tutum seviyelerinin ölçmek için Doğrusal Olmayan Kanonik Korelasyon Analizini (DOKKA) kullanarak aynı ve farklı grupların analizi yaparak benzerliğe neden olan tutumlar karşılaştırılmıştır. Sonuç olarak bazı grupların birbiriyle benzerlik gösterdiği, bazılarının ise diğer gruplardan ayrı bir yönelim gösterdiğini tespit etmişlerdir.

ÜÇÜNCÜ BÖLÜM

MODEL SEÇİMİNDE BİLGİ KRİTERLERİ

İstatistik alanında en çok uğraşı gerektiren konulardan biri de, iyi ve etkili bir regresyon modeli tahmin etmektir. Model de hangi değişkenlerin olması gerektiği, bu değişkenlerin modeli iyi derecede açıklama kabiliyeti göstermesi ve en uygun model olduğuna kanaat getirilmesi gibi hususlar araştırmacılar tarafından daima tartışma konusu olmuştur. Ucal'ın da (2006:42) belirttiği gibi bu süreçte en önemli prensip, modeli en iyi şekilde yansıtabilecek en az sayıda değişkenin belirlenmesidir. 1960'lı yılların sonlarından itibaren kullanılmaya başlanan model seçim kriterleri (Akaike FPE, C_p testleri gibi) güncellenerek ve evrilerek günümüzde kullanılmaktadır. Bu çalışma kapsamında modellere uygulanmak üzere ICOMP olarak bilinen entropik istatistiksel karmaşıklık kriteri kullanılacaktır. Ancak kriterin teorik yapısını anlamak adına öncelikle literatürde sık kullanılan bilgi kriterleri incelenecektir.

3.1. ÇOKLU BELİRLİLİK KATSAYISI (R^2) KRİTERİ

Regresyon modellerinin uygunluğunu belirlemede en çok kullanılan kriterlerin başında R^2 katsayısı gelir (Alpar 2003). 0-1 arasında değer alabilen bu sayı ne kadar 1'e yaklaşırsa; bağımsız değişkenlerin, bağımlı değişkeni açıklamada o kadar başarılı olduğunu gösterir (Bayata ve Hattatoğlu 2014:215). Regresyon denkleminde elde edilen veriler üzerinden R^2 şu şekilde hesaplanır (Günel 2004).

$$R^2 = \frac{\text{Açıklanmış ya da regresyon kareler toplamı}}{\text{Gözlenen değerlerin kareler toplamı}} \quad (40)$$

Bununla birlikte her bağımsız değişken, bağımlı değişkenle anlamlı ilişkili olmasa bile R^2 değerinin artmasına neden olur. Bu sorun farklı değişken sayılarına sahip modellerin karşılaştırılmasında çelişki yaratır. Bağımsız değişkenlerin modele etkisinin olması gerekenden fazla olmasını önlemek için Düzeltilmiş Belirlilik Katsayısı (Adjusted R^2) geliştirilmiştir (Özdemir, 2017, p. 215). Düzeltilmiş R^2 için önerilen denklem aşağıdaki gibi hesaplanır (Greene 2012).

$$R_A^2 = 1 - \frac{n-1}{n-k-1} (1 - R^2) \quad (41)$$

Formülde n örneklem sayısını, k bağımsız değişken sayısını belirtir. Bu kısıtlamalar sonucunda $R_A^2 < R^2$ çıkacaktır. Her ne kadar Düzeltilmiş R^2 En Küçük Kareler Yöntemiyle tahmin edilmiş regresyon modellerinde yaygın kullanılsa da, Bayes sınıflamalarında yetersiz olduğu görülmüştür (Burnham ve Anderson, 2013:37).

3.2. KULLBACK- LEİBLER BİLGİ KRİTERİ (K-L UZAKLIĞI)

Kullback ve Leibler (1951), K-L bilgisi veya mesafesi olarak adlandırılan Boltzmann'ın entropisinin* negatif olduğu bir bilgi ölçütü türetmiştir. İstatistikte Kulback-Leibler uzaklığı bir X rasgele değişkeni için söz konusu olabilecek f ve g gibi iki dağılım arasındaki uzaklığın simetrik olmayan bir ölçüsüdür (Biçer ve Biçer 2017:501). K-L uzaklığı; ayırım için bilgi kriteri, ıraksama ve görel entropi gibi farklı isimlerle de adlandırılır (Arslan, 2008:18).

f ve g dağılımları arasındaki K-L bilgisi sürekli değişken fonksiyonları için, logaritmanın doğal logaritmayı ifade ettiği integrali alınarak aşağıdaki notasyonlarla ifade edilir (Burnham ve Anderson, 2013:51).

$$I(f, g) = \int f(x) \log\left(\frac{f(x)}{g(x/\theta)}\right) dx \quad (42)$$

Burada $I(f, g)$ notasyonu, g modeli f modeline yaklaştıkça kaybedilen bilgiyi ifade eder. Bir başka ifadeyle g nin f ye olan uzaklığıdır. Yani K-L uzaklığı ne kadar küçükse bu iki modelde birbirine o kadar yakındır. Bu formülasyon kesikli değişkenler için ise aşağıdaki gibidir.

$$I(f, g) = \sum_{x \in X} f(x) \log\left(\frac{f(x)}{g(x/\theta)}\right) dx \quad (43)$$

Sonuç olarak, bu bir “tutarsızlık” ölçüsüdür; ancak basit bir mesafe değildir, çünkü f den g ye kadar olan ölçü, g den f ye kadar olan ölçü ile aynı değildir. K-L mesafesi, belki de asgari varsayımlardan ve bunun katkı özelliğinin türetilmesi anlamında tüm

* Bilgi teorisinin önemli temel kavramlarından olan entropi, kısaca bir değişkenin belirsizlik ölçüsü olarak tanımlanır. Shannon ve Weaver (1963), tarafından belirsizliğin bilgi ile ilişkisi kuramsal olarak formüle edilmiştir. Olasılık fonksiyonu $p(x)$ olan bir değişkenin entropisi şu formülle bulunur.

$$E(x) = \sum_{x \in X} p(x) \log p(x)$$

bilgi ölçümlerinin en temel olanıdır. Modeller arasındaki K-L mesafesi, bilim ve bilgi teorisinde temel bir niceliktir ve olasılık çıkarımı ile birlikte model seçiminin mantıksal temelidir. Başka bir ifadeyle, x ve y iki değişken olsun, buna göre bilgi içeriği $f(x,y)$; x deki belirsizliğin ne kadarının y tarafından açıklandığı bilgisini verir.

3.3. AKAIKE BİLGİ KRİTERİ (AIC)

Akaike 1973 yılında Budapeşte'deki Kiado Akademisinin gerçekleştirdiği Bilgi Teorileri Sempozyumunda Kullback-Leibler bilgilerinin veya mesafesinin model seçiminde temel olarak kullanmayı önerir. Akaike bununla birlikte, K-L mesafesini aday modellerin her birinde ($g_i(x/\theta)$), hem f hem de parametrelerin (θ) tam bilgisi olmadan hesaplanamayacağını, buna çözüm olarak ampirik log-olabilirlik fonksiyonuna dayanarak, maksimum noktada K-L bilgisini tahmin etmenin kesin bir yolunu bulmuştur. En Çok Olabilirlik (ML) tahmin edicileri, uygun koşullar altında asimptotik olarak etkindir. Bu yüzden gerçek değer etrafındaki parametrelerin ani değişimlerine karşı hassastır. Akaike'nin, göreceli beklenen K-L mesafesi ile maksimize edilmiş log olasılığı arasındaki bu ilişkiyi bulması, model seçiminde ve karmaşık veri setlerinin analizinde önemli pratik ve teorik gelişmelere olanak sağlamıştır (Bozdoğan 1987). Bu anlamda Akaike Bilgi Kriteri (AIC) ilk model seçim kriteri olarak kabul edilir. Formülasyon aşağıda verilmiştir .

$$AIC(p) = -2 \log(ML \text{ fonksiyonu}) + 2d \quad (44)$$

d , modeldeki bağımsız parametre sayısını gösterir (Bozdoğan ve Haughton 1998:53). Bu kriter gere göre en küçük değere sahip model en iyi olarak belirlenir. Akaike'nin yukarıdaki formülasyonu baz alınarak, cezalandırılmış olabilirliğe dayalı birçok model seçim kriteri (AIC_C , $QAIC$, $CAICF$, TIC vb..) türetilmiştir (Pamukçu 2015).

3.4. BAYES BİLGİ KRİTERİ (BIC)

Schwarz tarafından 1978 yılında türetilen BIC (ya da SIC) bilgi kriteri, Akaike Bilgi Kriteri'nin asimptotik olarak tam optimallliği sağlayamadığı eleştirisi üzerine modifiye edilmiştir (Schwarz 1978). BIC, eşitliğin son kısmına örneklem hacmini (n) ekleyerek AIC'den farklılığını ortaya koyar. Buna göre formülasyon şu şekilde oluşur.

$$BIC = -2 \log(ML \text{ fonksiyonu}) + d \log(n) \quad (45)$$

Bayesian Bilgi Kriteri (BIC), doğru bir model seçmekte daha faydalı iken, AIC gelecekteki gözlemleri tahmin etmek için en iyi modeli bulmakta daha uygundur (Chakrabarti ve Ghosh, 2011). BIC, modeldeki değişken sayısını daha çok cezalandırmakta ve AIC'e göre daha az sayıda değişken içeren modelleri seçer.

Aho'ya (2014:3) göre; BIC doğru modeli seçerken, AIC gerçek modelden daha karmaşık bir model seçer. Bu, hipotezlerin karşılaştırıldığı ve hipotezlerden birinin doğru olduğu bir dünya görüşünü yansıtır. Eğer gerçek model mevcut modellerden biriye, daha yüksek boyutlu bir model seçme olasılığı (fazladan takma) AIC için pozitif bir sabite ve BIC için sıfıra dönüşür. Gerçek boyuttan daha küçük bir boyut modeli seçme olasılığı (donatı), hem AIC hem de BIC için sıfıra yaklaşır (Bozdoğan et Haughton, 1998:3)

3.5. BİLGİ KARMAŞIKLIĞI KRİTERİ (ICOMP)

Bu bölüme kadar anlatılan bilgi kriterlerinin en büyük eksikliği değişkenler arasındaki ilişkileri dikkate almamalarıdır. Pamukçu (2017)'nin da belirttiği gibi en iyi modelin seçimi, değişkenler arasındaki ilişkiler hakkında kesin bilginin olmadığı durumlarda tutarlılık sorunu yaşanmaktadır. Buradan hareketle Bozdoğan (1988) sadece model yalınlığı ve uyum iyiliğini değil, ek olarak modelin karmaşıklığını da devreye sokarak ICOMP (Information **Complexity**) ve onun özel hali ICOMP (IFIM) kriterlerini üretmiştir. ICOMP (IFIM) olarak adlandırılan en genel ICOMP formu, MLE'lerin iyi bilinen asimptotik optimallik özelliklerinden yararlanır ve bir modelin karmaşıklığını ölçmek için Ters Fisher Matrisi'ni (IFIM) kullanır. Her iki tip için kriterler arasında en küçük puana sahip model en iyi model olarak seçilir (Liu ve Bozdoğan, 2008a).

Bozdoğan (2003) karmaşıklığı: *“Tüm sistem ve alt sistemlerinin veya parçalarının basit bir numaralandırma kompozisyonu arasındaki bağımlılık derecesinin bir ölçüsü”* olarak tanımlar. Bu tanım, bir modelin karmaşıklığının yalnızca bir modele dahil edilen parametrelerin (veya değişkenlerin) sayısına değil, bunların birbirleriyle nasıl ilişkili olduklarına dayandığını gösterir. Başka tanımlara göre karmaşıklık; modellerin spesifik içeriğinden, yapısından veya olasılık özelliklerinden büyük ölçüde bağımsız olan istatistiksel modellerin genel bir özelliğidir. Genel olarak, istatistikte benzersiz

bir karmaşıklık tanımı yoktur, çünkü kavram van Emden'e göre “zor” dur (Emdem, 1971:8). Karmaşıklığın birçok yüzü vardır ve “Kolmogorov Karmaşıklığı”, “Shannon Karmaşıklığı ve “Stokastik Karmaşıklık” bunlara örnektir. Bilgi teorik kodlama teorisinde karmaşıklığı, model sınıfı tarafından elde edilebilecek veriler için en kısa kod uzunluğu olarak tanımlar ve buna Stokastik Karmaşıklık (SC) adını verir. Monash Okulu (örneğin, Wallace ve Freeman, 1987, Wallace ve Dowe, 1993, Baxter, 1996), veriyi içeren bir mesajı sıkıştırma yeteneklerine göre modellerin değerlendirilmesine dayanan Minimum Mesaj Uzunluğu (MML) olarak karmaşıklığı tanımlar (Bozdoğan, 1993:19).

Bilgi Karmaşıklığı (ICOMP) Kriteri yukarıdaki tanımları kullanır ve sonuç olarak model seçimi için bir ICOMP kriterler sınıfı elde edilir. ICOMP'un geliştirilmesi, kısmen AIC tarafından ve kısmen de bilgi karmaşıklığı kavramları ve endeksleri tarafından motive edilmiştir. ICOMP'un geliştirilmesi ve inşası, ilk olarak van Emden (1971) tarafından tanıtılan kovaryans karmaşıklığı endeksinin genelleştirilmesine dayanmaktadır. ICOMP, AIC'nin aksine, serbest parametre sayısını doğrudan cezalandırmak yerine, modelin kovaryans karmaşıklığını cezalandırır (Koc ve Bozdoğan, 2014). ICOMP, aşağıda verilen kayıp fonksiyonu ile tanımlanmaktadır.

$$Kayıp = Uyum eksikliği + Ceza puanı + Karmaşıklık ölçüsü$$

Formüldeki ilk iki kısım AIC kriterini tanımlarken, son kısım bilgi teorisinin ekleme özelliklerini kullanır. Bu terim, modelin rasgele hata terimleri ve parametre tahminleri arasındaki kovaryans ve bağımlılık yapısını betimler (Howe ve Bozdoğan, 2010). ICOMP kriterinin genel formülü şu şekildedir.

$$ICOMP = -2\log L(\hat{\theta}_k) + 2C(\Sigma_{Model}) \quad (46)$$

Notasyonların açılımı aşağıdaki gibidir;

$L(.)$: maksimize edilmiş olabilirlik fonksiyonu

$\hat{\theta}_k$: model içindeki θ parametre vektörünün maksimum olabilirlik tahmini

C : gerçek değerli karmaşıklık ölçüsü

Σ : modelin parametre vektörünün tahmini kovaryans matrisi

$C(\Sigma)$, hem varyanslar arasındaki eşitsizliği hem de kovaryansların Σ 'ya katkısını ölçer. Eğer bütün değişkenler birbirinden bağımsız ise $C(\Sigma) = 0$ 'a eşittir. Değişkenler arasında yüksek etkileşim var ise değer büyük çıkacaktır. Yalın ve tutarlı modeller oluşması açısından karmaşıklığın mümkün olduğunca küçük ve hatta sıfır çıkması istenir.

ICOMP'un birkaç çeşidi bulunmaktadır (ICOMP (IFIM), ICOMP (IFIM)_PEU, ICOMP_k, ICOMP_MISS, CCOMP ve türevleri). Bu çalışmada ICOMP (IFIM) olarak adlandırılan form açıklanacak ve kullanılacaktır. Bu kriter, hem MLE'lerin iyi bilinen asimptotik optimallik özelliklerini kullanır hem de bir modelin ters Fisher bilgi matrisinin (IFIM) bilgiye dayalı karmaşıklığını kullanır (Bozdoğan ve Haughton, 1998:55). Burada Σ kovaryans matrisi yerine Fisher Bilgi Matrisi kullanılmaktadır. Böylece formül aşağıdaki gibi özelleşmiş olur.

$$ICOMP(IFIM) = -2\log L(\hat{\theta}_k) + 2C(\hat{\mathcal{F}}^{-1}) \quad (47)$$

$$= n\log(2\pi) + n\log(\sigma^2) + n + C(\hat{\mathcal{F}}^{-1}) \quad (48)$$

Bilgi teorik model değerlendirme kriterlerinde $C(\hat{\mathcal{F}}^{-1})$ 'in kullanımı, bir modeldeki serbest parametre sayısını artırdıkça, parametre tahminlerinin doğruluğunun azaldığı gerçeğini dikkate alır. Cimrilik ilkesinde de belirtildiği gibi, ICOMP (IFIM), daha karmaşık ve belirsiz modelden ziyade daha doğru ve verimli parametre tahminleri sağlayan daha basit modeller seçer (Bozdoğan, 2003a:29). ICOMP (IFIM), AIC ve SBC'den daha makul bir ceza süresi sağlar ve daha doğru ve verimli parametre tahminleri sağlayan basit modelleri seçer.

3.5.1. Fisher Bilgi Matrisi

Herhangi bir kestiricinin minimum varyans sınırını bulmak için Fisher Bilgi Matrisi kullanılır. Bunun için öncelikle olabilirlik fonksiyonunun belirlenmesi gerekir (Şenoğlu, Özarıcı, ve Fidan 2001:370). Matris, θ 'nın yansız kestiricisinin doğruluğu ile orantılıdır ve θ parametresine ilişkin bilginin miktarını ölçer (Deniz Akıncı 2007:32). Fisher Bilgi Matrisi'nin log olabilirlik fonksiyonu altındaki kovaryanslar biçiminde tanımlanan $k \times k$ karakteristiğinde bir simetrik matristir.

$$\mathcal{F}(\theta)_{i,j} = E \left[\frac{\partial^2 \log f(X/\theta)}{\partial \theta_i \partial \theta_j} \right] \quad (49)$$

Fisher Bilgi Matrisi'nin tersi (Inverse Fisher Information Matrix-IFIM)), model kovaryans matrisi ile yakından ilişkili olup bazı bilgi kriterlerinde önemli rol oynar. Tahmihi $\hat{\theta}_k$ 'nın yardımı ile hesaplanan kovaryans matrisini kullanılmasıyla normal dağılım için IFIM şu şekilde hesaplanır (Bozdogan,1987).

$$\hat{\mathcal{F}}^{-1} = \begin{bmatrix} \hat{\Sigma}(\hat{\theta}) & 0 \\ 0 & (\frac{2}{n} D_{p+q}^+ [\hat{\Sigma}(\hat{\theta}) \otimes \hat{\Sigma}(\hat{\theta})] D_{p+q}^+) \end{bmatrix} \quad (50)$$

Bir modelin yanlış tanımlanmasını kontrol etmek için basit bir bilgi matrisi denkliği testi yapılır. IFIM'in köşegen elemanları, tahmini parametrelerin tahmini varyansını içerirken, karşılık gelen köşegenler kovaryansları içerir. Dolayısıyla, ICOMP, modelin tüm parametre alanını hesaba katan IFIM ile evrensel bir kriter sağlar (Koc ve Bozdogan, 2014:9).

DÖRDÜNCÜ BÖLÜM

GENETİK ALGORİTMA

4.1 GENETİK ALGORİTMA NEDİR?

Tarih 19. yüzyılın ortalarını gösterdiğinde kaşif rahip Darwin, canlı türlerinin ortak atalardan farklılaşarak evrimleştiği tezini ortaya atarak bilim dünyasında bitmek bilmeyen tartışmaların fitilini ateşlemiştir. 1859'da "Türlerin Kökeni" kitabı yayınlayan Darwin, doğada güçlü olan canlıların hayatta kaldığını, koşullara adapte olamayan türlerin elendiğini teorize etmiştir. Yavruların çeşitli nedenlerle ebeveynlerinden birazcık farklılık göstermeleri, sabit koşullar altında çoğu bireye dezavantaj sağlarken, değişen çevresel koşullarla birlikte bazılarının hayatta kalma konusunda avantaj sağlamaktadır. Avantajlı olanlar daha kolay ürer ve bu özelliklerini gelecek nesillere daha çok aktarırlar. Bu şekilde nesil ve popülasyon bazında avantajlı özellikler sayıca artar, dezavantajlı özellikler giderek azalır. Bu eleme-seçilim mekanizmasına Doğal Seçilim denir (Bakırcı, 2014) .

En iyi olan yaşar kanununa bağlı olarak işleyen ve bir popülasyon üyelerinin rekabet sonucu elenmelerini modelleyen Genetik Algoritma Yöntemi, her ne kadar 1967 yılında Bagley tarafından keşfedilmişse de (ardından Rosenberg, De Jong) kuramsal olarak John Holland, meslektaşları ve Michigan üniversitesindeki öğrencileri tarafından geliştirilmiştir (Altay, 2015:8). Holland'ın 1975 yılında yayınlanan "Doğal ve Yapay Sistemlerde Uyarılma" adlı kitabı Genetik Algoritmaları, doğada gözlemlenen evrimsel süreci taklit ederek çalışan arama ve en iyileme yöntemi olarak tarif eder. Algoritmanın amacı; karmaşık çok boyutlu arama uzayında en iyi kromozonların hayatta kalarak belirli bir durağanlık sonunda bütünsel en iyi çözümü aramaktır. Genetik Algoritmalar, evrimsel süreci simüle ederek, modeli belirli kısıtlar ve amaç fonksiyonu altında çözerek en iyilemeye çalışır. Yöntem özellikle polinom zamanda çözülemeyen zor problemleri (NP Hard) çözmek için evrim sürecinin bilgisayar aracılığıyla kodlandığı, en iyi ya da en iyiye yakın sonuçların hesaplandığı bir sezgisel arama yöntemidir.

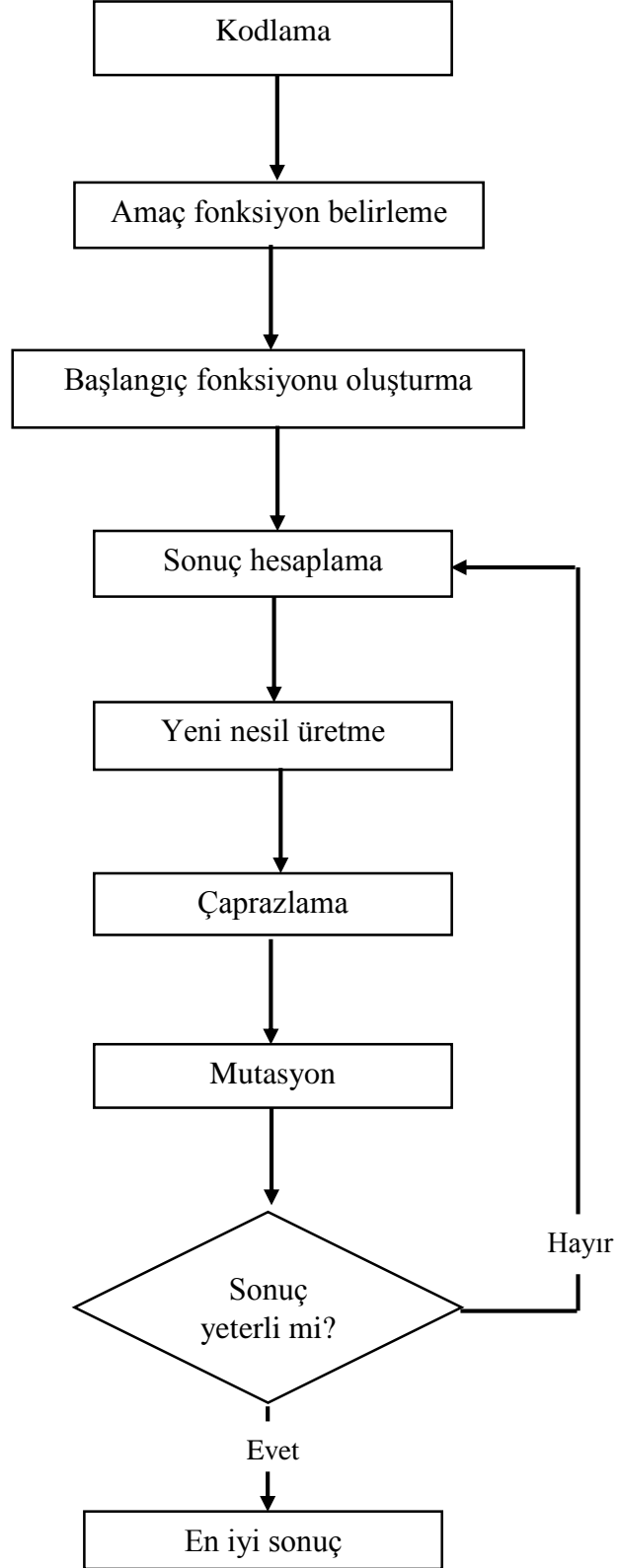
Bununla birlikte ilerleyen zamanlarda yöntem birçok alanda kullanılmıştır. Holland'ın ardından teorik yöntemi pratiğe döken Goldberg (1989) "Genetic Algorithms in

Search, Optimization and Machine Learning’’ kitabını yayımlamıştır. Ardından 1992’de John Koza, belirli görevleri yerine getirmek ve program geliřtirmek için Genetik Algoritmayı kullanmış ve metodunu “genetik programlama” olarak adlandırdı (Bozdogan 2003a:22). Günümüzde nerdeyse her alanda kullanılan yöntemin performansı gerek hibrit yöntemlerle gerekse de opsiyonel olarak gelişmektedir. Yöntemin en önemli ayrıcalığı, birçok arama algoritması lokal minima sorunu yaşarken, Genetik Algoritma bununla baş etmesidir (Minerva ve Paterlini, 2002:19). Genetik Algoritmada kromozomlar kendini devamlı kombine ettiğı için (çaprazlama ve mutasyonlar sayesinde) araştırma uzayı devamlı genişler.

4.1.1. Genetik Algoritmanın Adımları

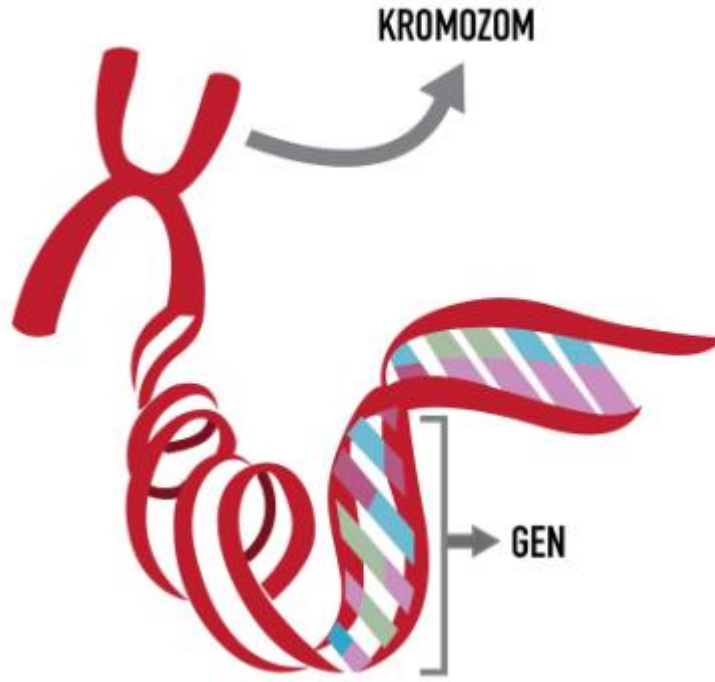
Genetik Algoritmanın çalışma akış diyagramı aşağıdaki şemada gösterildiğı gibi işlenmektedir.

Şekil 6: Genetik Algoritma Akış Diyagramı



4.1.2. Genetik Algoritmada Kullanılan Kavramlar

Algoritma işleyişinde kullanılan terimler ve analogiler konuyu anlama bakımından önemlidir. Algoritma elemanları ve tanımları aşağıda verilmiştir.



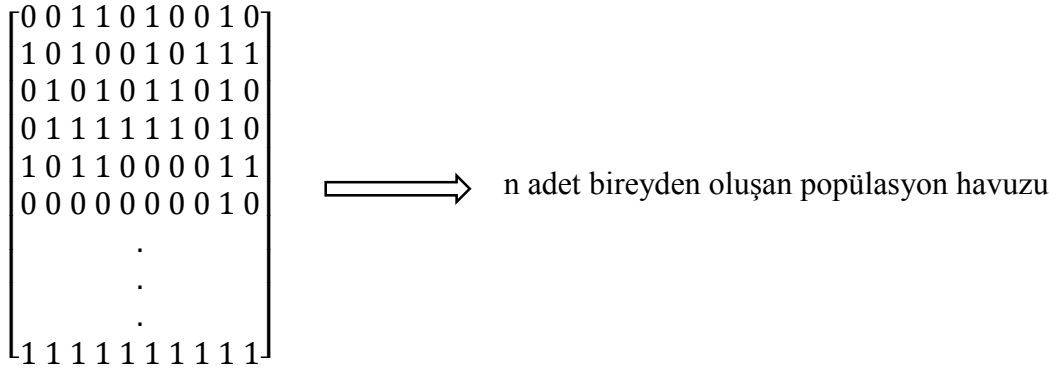
Şekil 7: Gen Yapısı Görseli

Gen: Bireylerin kalıtsal karakterlerini belirleyip, nesilden nesile aktaran kalıtım materyalidir. Her bir gen bir özelliği taşır. Genetik Algoritmada bir sayı ile ifade edilir. 0 veya 1 gibi.

Kromozom: Genlerin bir dizi halinde birleşerek oluşturdukları yapılara denir. Algoritmada bir bireyi tarif eder.

0 1 0 1 1 0 1 1 0 1 1 0

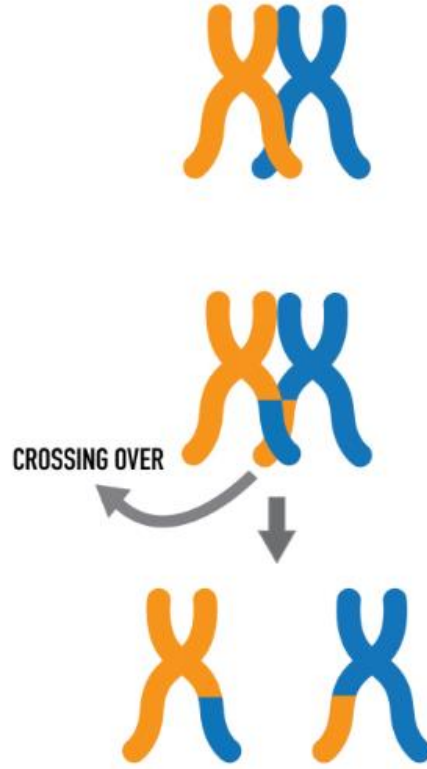
Popülasyon: Birden çok bireyin (kromozom) bir araya gelerek oluşturduğu topluluğa denir. Algoritmada aday çözümler topluluğunu temsil eder. İsteğe göre popülasyon havuzundaki birey sayısı belirlenir. Genellikle 10 ya da katları olarak atanabilir.



Amaç Fonksiyonu: Hangi bireylerin hayatta kalacağını hangisinin eleneceğini belirleyen ve iterasyonlar sonucunda en optimal çözümü tespit eden sonuç kriteridir. Fonksiyonun tipine göre en büyük ya da küçük değerlere sahip olan bireyler seçilir.

Seleksiyon : Popülasyon içerisinde amaç fonksiyonu tarafından hayatta kalabilen bireylerin seçilmesidir. Diğer kromozomlar elenecek, kalanlar tekrar algoritmaya tabii olacaktır. Bu sayede güçlü bireylerin hayatta kalması hedeflenir. Güçlü bireyler tekrar üreme işlemine girer. Bunun için literatürde bir çok yöntem bulunur. Bunlara; Rassal Üreme, Rulet Çarkı Yöntemi, Beklenen Değer Yöntemi, Boltzmann Yöntemi, Sıralı Seçim Yöntemi, Turnuva Yöntemi, Elitizm Yöntemi gibi örnekler verilebilir. Bu opsiyonların her biri denenerek uygun çözüm aranabilir.

Çaprazlama: Çaprazlama (Crossing-over) bu süreçte kullanılan en önemli operatördür. Amacı popülasyonda olmayan bireyleri yaratarak bireyler arasındaki çeşitliliğin arttırılmasıdır (Altay, 2015:15).



Şekil 8: Çaprazlama (Crossing-over) Süreci (Bakırcı 2011)

Algoritma içerisinde her çift için çaprazlama olabilmesi için bir olasılık katsayısı p_c belirlenir. Bu olasılığa bağlı olarak çaprazlama yapılıp yapılmayacağı belirlenir (Vural, 2015:16).

Mutasyon: Mutasyon, GA'da global bir arama gerçekleştirmek için kullanılan başka bir parametre veya operatördür. Bu esnada kromozomdaki bir gen değişikliğe uğrar. Mutasyon içinde çok küçük bir olasılıkla bir katsayı p_m belirlenir. Böylece arama süreci bir bölgeden başka bir bölgeye atlayarak lokal minimadan kurtulur (Liu ve Bozdoğan, 2008b:17).

Sonuç olarak döngü optimal çözümü buluna kadar aramaya devam eder. Eğer amaç fonksiyonunda belirli bir durağanlık sağlanırsa sona gelinir ve en iyi çözüme ulaşılır.

4.2. REGRESYON MODEL SEÇİMİNDE GENETİK ALGORİTMA KULLANIMI

Çok Değişkenli Lineer Regresyon modellerinin seçiminde Genetik Algoritma kullanımı ilk olarak Goldberg (1988) tarafından sunulmuştur. Bağımsız değişken sayısının çok fazla olduğu regresyon modellerinin her birinin test edilmesi gerek zaman açısından gerekse de maliyet açısından zordur. (Paterlini ve Minerva 2010), iki ayrı veri setine uyguladıkları analizlerde, Stepwise regresyon ile Genetik Algoritma regresyon yöntemlerini karşılaştırmıştır. Sonuç olarak ikinci yöntemin daha anlamlı sonuçlar ürettiğini ve de Stepwise yönteminin lokal minimada takıldığını belirlemiştir. Bununla birlikte çalışmamızda 1 adet bağımlı değişken, 25 adet bağımsız değişken mevcuttur. Basit bir hesaplamayla toplamda $2^{26} - 1$ adet (yaklaşık olarak 70 milyon) altset arasından en uygununu bulmak, neredeyse imkansızdır.

Genetik Algoritma, popülasyon adı verilen bir dizi çözüm (kromozomlarla temsil edilen) ile başlatılır. Bir nüfustan çözümler alınır ve yeni bir nüfus oluşturmak için kullanılır. Bu, yeni nüfusun eskisinden daha iyi olacağı ön kabulü vardır. Yeni çözümler (offsprings) oluşturmak üzere seçilen çözümler, uygunluk değerlerine göre seçilir (Bozdoğan, 2003a:17). Son olarak amaç fonksiyon değerinde durgunluk ve tekrarlar saptandığında algoritma nihai çözüme ulaşır.

Genetik Algoritma aracılığıyla en uygun regresyon modelinin seçilmesi süreci aşağıdaki adımlarla tanımlanabilir:

i) Mümkün regresyon modelleri için genetik kodlama şeması oluşturulması

Her regresyon modeli, dizedeki her geni-bağımsız değişkeni temsil edecek şekilde ikili (binary) kodla tarifiendirilerek bir dizi-kromozom oluşturulur. Burada değişkeninin (ya da genin) varlığını (1) veya yokluğunu (0) temsil eder. Her dize aynı uzunluğa sahiptir, ancak her biri farklı ikili kodlama içerir. Her kromozom bir regresyon modelini temsil eder.

Örneğin, $k = 5$ değişkenli bir regresyon modeli olsun. Kromozom ise $[0 \ 1 \ 0 \ 1 \ 1]$ şeklinde kodlansın. Dikkat edilirse 5 gen mevcuttur. Sonuç olarak bu modelin en uygun olduğunu varsayarlırsa, modeli açıklayan değişkenler; ikinci, dördüncü ve

beşinci bağımsız değişkenler olarak tespit edilirken, bir ve üçüncü değişkenler bağımlı değişkeni açıklama da anlamsız olarak belirmiştir.

ii) Model için başlangıç popülasyonun üretilmesi

Popülasyon büyüklüğü (yani, monte edilen model sayısı(n)) Genetik Algoritma'nın önemli bir parametresidir. Popülasyon büyüklüğü, bir popülasyonda kaç kromozom olduğunu belirtir. Çok az kromozom varsa, Genetik Algoritma çaprazlama(crossover) gerçekleştirmek için az imkan bulur ve arama alanının sadece küçük bir kısmını araştırır. Öte yandan, çok fazla kromozom varsa, Genetik Algoritma yavaşlar. Araştırmalar, bazı sınırlardan sonra (esas olarak kodlamaya ve soruna bağlı olan) popülasyon boyutunu arttırmanın yararlı olmadığını göstermektedir. İlk önce n rastgele üreme modellerinin ilk popülasyonunu başlatılır.

iii) Modellerin performanslarını değerlendirmek için bir uygunluk (amaç) fonksiyonun belirlenmesi

Literatürde regresyon model seçimi için Genetik Algoritma kullanımında bir çok amaç fonksiyonu AIC (Paterlini ve Minerva, 2010), kNN (Oluleye vd. 2014), ICOMP (Liu ve Bozdogan 2008b) kullanılmıştır. Araştırmacının inisiyatifine bağlı olan süreç, tutarlı ve sıralanabilir olmalıdır. Çalışmamızda amaç fonksiyonu olarak ICOMP (IFIM) bilgi kriteri kullanılacaktır. Buna göre en son nesil içerisinde en düşük puana sahip olan kromozom-birey hedef modelimiz olacaktır.

iv) Yeni nesiller üretmek için, ebeveynlerin çiftleşmesi ve üreme yöntemlerinin seçilmesi

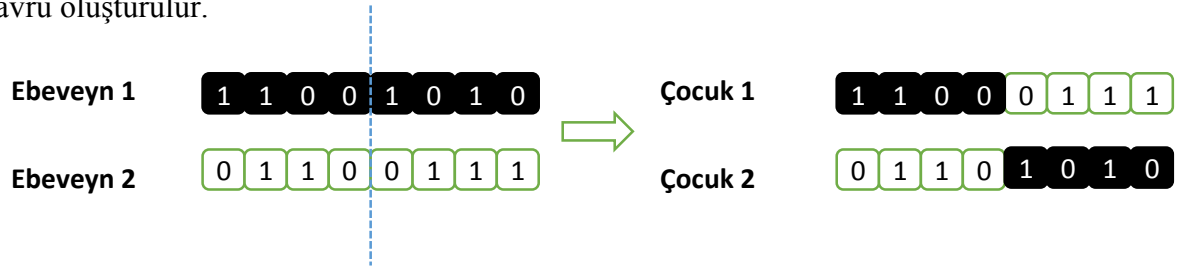
Çiftleşme bir çaprazlama işlemi olarak gerçekleştirilir. Bu işlem Genetik Algoritmada kullanılan en önemli operatördür ve amacı popülasyona yeni bireyler yaratarak bireyler havuzundaki çeşitliliği arttırmaktır (Altay, 2015:15) Geçiş için seçilen bir model, çaprazlama olasılığı (p_c) ile kontrol edilir. Çaprazlama olasılığı (p_c) genellikle araştırmacı tarafından belirlenir. $p_c = 0$ çarpma olasılığı, basitçe çiftleşme havuzunun üyelerinin bir sonraki nesle taşınması ve hiçbir yavru üretilmemesi anlamına gelir (Tek hücreli canlıların üreme şekli olan mitoz bölünme gibi). Çaprazlama olasılığı $p_c = 1$, çiftleşme havuzundan seçilen iki ana model arasında her zaman çiftleşmenin

(çaprazlama) meydana geldiğini gösterir; bu nedenle gelecek nesil sadece yavru modellerden oluşacaktır (önceki nesilden hiçbir yeni nesilde olmayacaktır).

Çaprazlama işlemi sırasında, geçiş noktası olarak her ana model çifti (dizeleri) boyunca rastgele bir nokta seçilir. Herhangi bir ebeveyn çifti için, kromozom geçiş noktasında iki parçaya bölünür ve bu noktanın sağındaki iki kromozomun kısımları iki yavru kromozomu oluşturmak için ebeveynler arasında değiştirilir. Her bir kromozomun uzunluğu boyunca rastgele seçilen bir nokta, modellerin kırıldığı ve daha sonra iki yeni model oluşturmak üzere başka bir ana parçaya yeniden bağlandığı çaprazlama noktası olur (Bozdoğan, 2003a:23). Birçok çaprazlama operasyon seçeneği bulunmaktadır. Bunlardan en çok kullanılanları şu şekilde özetlenebilir.

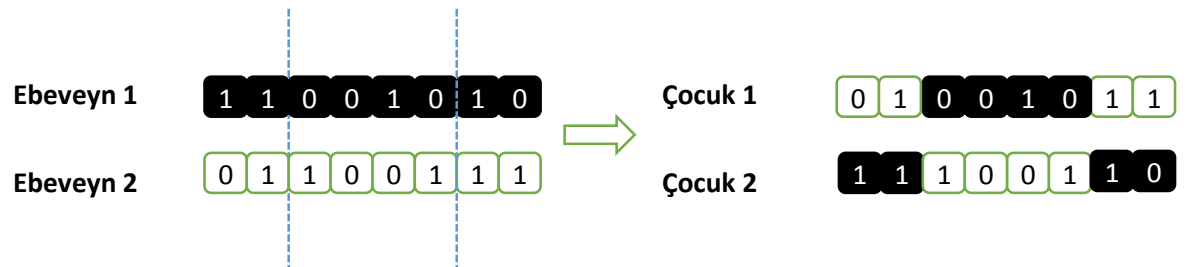
- *Tek noktalı çaprazlama*

En basit çaprazlama yöntemidir. Keyfi olarak iki gen arasında bir nokta seçilir. Ardından kromozomun sağ tarafındaki genler karşılıklı olarak değiştirilerek iki adet yavru oluşturulur.



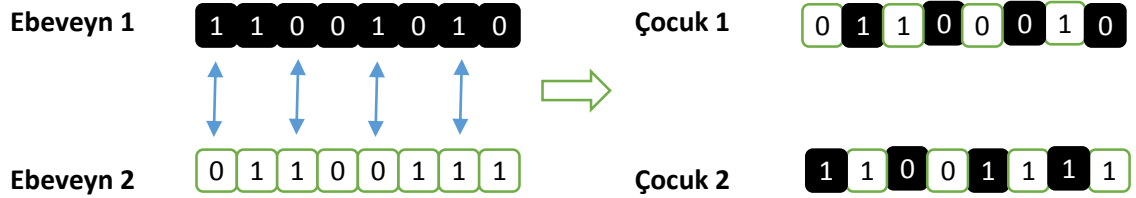
- *İki noktalı çaprazlama*

Aynı bir noktalı çaprazlama yönteminde olduğu gibi, burada da kromozom iki nokta da parçalanır ve karşılıklı genler yer değiştirir.



- *Tekdüze (Uniform) çaprazlama*

Daha fazla noktalı çaprazlama yöntemidir. Her kesim noktası arasındaki parçaların karşılıklı olarak değişimidir. Tesadüfi belirlenen genler arasındaki değişimi içerir.



Ayrıca kullanıcı ‘elitizm kuralı’ olarak adlandırılan seçim opsiyonuna da sahiptir. Yani iyi çözümlerden en az bir tanesi yeni popülasyona değişmeksizin geçer. Bulunan en iyi çözüm çalışmanın sonuna kadar canlı kalacaktır.

v) Yeni yavru modellerinin bileşimini değiştirmek için mutasyon uygulanması

Modellerin mutasyonu, arama işleminin sınırlı bir alanda arama yapmak yerine daha verimli başka bir alana atlayabilmesi için yeni değişken kombinasyonları oluşturmanın başka bir yolu olarak kullanılır. Rastgele seçilen bir genin 0 ile 1 veya 1 ile 0 arasında değişebileceği bir mutasyon oranı veya olasılığı (p_m) belirtilerek mutasyona izin verilir. Bu nedenle, rasgele seçilen bir bağımsız değişken modele eklenir veya modelden kaldırılır.

Belirli çaprazlama ve mutasyon oranlarına bağlı olarak, ikinci nesil tamamen yavru modellerden ya da yavru ve ebeveyn modellerinin bir karışımından oluşacaktır. İkinci jenerasyondaki modeller daha sonra üçüncü jenerasyonu üretmeye gider; süreç, araştırmacı tarafından kontrol edilen belirli sayıda nesil için birbiri ardına devam eder. Algoritma artık birbirine yakın amaç değerleri üretmeyi sürdürürse süreç sonlandırılır.

BEŞİNCİ BÖLÜM

UYGULAMA

Çalışmanın bu kısmında model değişkenleri hakkında bilgi verilecektir. Ardından veri analiz sonuçları raporlanıp yorumlanacaktır. Mevcut araştırmanın temel amacı, cep telefonu kullanımının, uyku düzeni üzerine etkisinin olup olmadığını araştırmak, var ise hangi değişkenlerce açıklandığını belirlemektir. Bu amaç doğrultusunda ilk olarak, cep telefonu ve uyku arasındaki ilişkiyi belirlemeye yönelik bir ölçek geliştirilmesini ihtiyaç duyulmuştur. Var olan araştırmalarda bu ihtiyacı karşılayacak bir ölçeğe rastlanmadığından, yeni bir ölçek geliştirilmiştir. Kategorik değişkenlerden oluşan ölçeğe sayısallaştırmalar uygulanarak, uygun modelin seçimi için tüm mümkün alt setler Genetik Algoritma ile taranmıştır. Sonuç olarak, amaç fonksiyonu değeri en az olan regresyon modeli belirlenmiştir. Araştırma için gerekli sayısallaştırma ve algoritma kodlamaları MATLAB 2015a programında gerçekleştirilmiştir.

5.1. ARAŞTIRMANIN YÖNTEMİ

5.1.1. Uyku Nedir?

Uyku, insanın neredeyse zamanın üçte birini geçirdiği ve yaşamında soluk alıp verme, boşaltım ve beslenme kadar önemli olan bir aktivitedir. Literatürde uyku ile ilgili birçok tanım mevcuttur. Bunlardan birkaçı şunlardır:

Uykunun Felsefesi adlı kitabın yazarı Macnish (1834:1) uykuyu; ölüm ve uyanıklık arasının tam ortası olarak tarif eder. Uyanıklığı, bütün zihinsel fonksiyonların aktif halde olması, ölümü ise bütün bunların pasife alınması olarak betimler. (Selvi vd. (2015) göre ise, Elektrofizyolojik yöntemlerle gözlenebilen beynin aktif bir durumudur ve belirli bir ritmi vardır. Başından sonuna kadar her şey sistemli şekilde belirli bir ritme uygun olarak gelişir. Uyku, bir süreç içerisinde değişik uyaranlarla ya da kendiliğinden sonlandırılan bir işlemdir. Uyku, algılama seviyesinin yükseldiği, dış dünyayı algılamanın zayıfladığı, çevreye duyarlılığın geçici olarak kaldırıldığı bir durumdur. Bir başka tanıma göre ise uyku, bilincin çevresel uyaranların bir kısmını veya bütünü algılamadığı, tepki şiddetinin zayıfladığı ve vücuttaki birçok organın performansının büyük oranda azaldığı dinlenme durumudur (Gohil ve Hannon, 2018). Amerikan Ulusal Akıl Sağlığı Enstitüsü ise uykuyu, tekrarlayan, beynin dinamik

fonksiyonel örgütlenmesindeki değişiklikleri yansıtan, davranışsal, fizyolojik ve vücut sağlığı açısından vücudu en uygun ve zinde hale getirme süreci olarak tanımlar (Vijayalaxmi vd., 2002:2578).

Tanımlardan da anlaşılacağı üzere, uyku salt bir bilinçsizlik hali olmanın dışında bütün vücudun iç dengesini düzenleyerek onu yaşama hazırlayan bir yenilenme sürecidir.

Uyku, sağlıklı yaşamın çok önemli bir bileşeni, fiziksel ve zihinsel faaliyetlerin yerine getirilmesini sağlayan çok önemli bir ihtiyaçtır. Uyku, fizyolojik ve biyolojik ihtiyacı karşılayan bir konu olarak öne çıksa da, sosyolojik, antropolojik, psikolojik, davranışsal ve tarihsel boyutları göz ardı edilemez. Uykuyu sadece biyolojik tabanda ele almak, konuyu basite indirgemek olur; uykuyu kolaylaştıran veya kısıtlayan çevresel koşullar gibi sosyo-kültürel faktörler önemlidir. Uyku, çok yönlü (multifaceted), multidisipliner, çok boyutlu, karmaşık bir fenomendir (Williams, 2002).

Doğanın gece gündüz ritmi, insanın biyolojik ritmini, diğer deyişle uyku ve uyanık olma zamanını işaret eder . Işığın uykuyu engelleyici etkisinden yola çıkan otoriteler, yatmadan önce cep telefonu, bilgisayar ve tablet ekranının uykuyu engellediği görüşündedirler (Adrien, Royant-Parole & Gronfier, 2016). Az uyku veya kalitesiz uyku, genç insanların öğrenme ve gelişmelerini olumsuz etkileyebiliyor. Matricciani, Old ve Petkov (2012), 1905 ve 2008 yılları arasında 20 ülkenin 103 yıllık verilerini incelemiş; çocuk ve ergenlerin uyku sürelerinde sürekli düşüşler tespit etmiştir. Yeterli uyku alamayan insanların hem fiziksel hem de zihinsel olarak olumsuz etkilendikleri yapılan çalışmalarda görülmüştür. Axelsson, Sundelin, Ingre, Van Someren, Olsson ve Lekander'in bulguları (2010) uykudan mahrum (deprived) insanların daha az sağlıklı, daha az çekici (less attractive) ve iyi dinlenmiş (well rested) hallerine göre daha yorgun görüldüğünü göstermektedir. Bu, uykusuzluğun insanların yüzüne yansıdığını ve davranışlar üzerinde potansiyel etkileri olduğunu göstermektedir. Şenol vd.,gençlerin %54.7'sinin kötü uyku kalitesine sahip olduğunu, gençlerin ders çalışırken (%60.3) ve ders dinlerken uykusuzluk problemi yaşayanların (%49.3), daha kötü uyku kalitesine sahip olduklarını tespit etmiştir.

Tüm canlı organizmalar için uyku çok önemlidir ve uyku olmadan canlılığın yaşam sürdürmesi imkansızdır. İnsanlarda uyku ihtiyacı; cinsiyet, yaş, sağlık durumu, beslenme alışkanlığı, fiziksel aktivite, bireysel özellikler ve çevresel koşullara göre farklılık gösterir. Bu nedenle uyku; insanın yaşam kalitesini ve iyilik halini önemli ölçüde etkiler (Åkerstedt vd. 1994:290). Düzenli ve kaliteli bir uyku süreci organizma için en önemli gereksinimdir. Bireyin kendisini yeni güne dingin ve zinde bir şekilde hazırlama süreci olarak tanımlanan uyku kalitesi; çok fazla neden ve faktörden etkilenir (Buysse 2014; Şenol vd. 2012).

Kabul edilen normlara göre, bebeklerde 20-22 saat, çocuklarda 10-12 saat, yetişkinlerde 6-8 saat, yaşlılarda ise 5-6 saatlik uykular idealdir (Şenel 2015:2). Sağlıklı, zinde, dinç ve verimli bir hayat için düzenli, yeterli ve kaliteli bir uyku en önemli koşuldur. Ancak gerek psikolojik gerekse de dışsal etmenler bireylerin çoğunda uyku düzeni ve kalitesinde olumsuz etkiye neden olur. Yaş ve coğrafyaya bağlı olarak değişen faktörlerin başında; cinsiyet, uyku alışkanlıkları, beslenme tarzı, hareketsiz yaşam, ideal olmayan koşullar, uyku öncesi cihaz kullanımı, üst solunum yolu bozuklukları, stres, çalışma hayatı düzensizlikleri ve ilaç kullanımı gibi nedenler sıralanabilir (Kiper ve Sunal 2009; Ruggiero 2003; Taylor ve Harper 2003).

Bu gibi nedenlerden dolayı uyku kalitesinde yaşanan yetersizlikler, bireylerin sağlığında ve sosyal yaşam ilişkilerinde problemleri de beraberinde getirir. Bunların en önde gelenleri; gündüz uykululuk hali, dürtüsellik, dikkat dağınıklığı, huysuzluk, asabiyet, akademik performansta düşüklük, öğrenme bozuklukları, sosyal ilişkilerde anlaşmazlık, kalp yetmezliği, patolojik derecede utangaçlık, yüksek tansiyon gibi birçok olumsuz durumlardır (Kaczor ve Skalski 2016; Nunes vd. 2015).

Ancoli-Israel vd. (1997) de belirttiği gibi, düzensiz uyku ve kötü uyku kalitesi; ağır kazalar, çalışma performansında kayıp, psikolojik işlevlerde ciddi bozukluk gibi ciddi sağlık problemlerini neden olmaktadır. Buradan hareketle uyku düzensizliğine yol açan nedenlerin incelenmesi ve bu etkenlerin düzenlenmesi ya da ortadan kaldırılması araştırmacılar tarafından devamlı işlenmektedir.

5.1.2. Ölçek Oluşturma Süreci

Bir önceki bölümde de vurgulandığı gibi uyku düzensizliği neden olan faktörlerin sayısının çok olduğunu ve literatürde buna dair pek çok çalışmanın olduğu belirtilmiştir. Bu çalışmada bu etkenlerden biri olan uyku öncesi cihaz, alet ya da etkinliklerin taraması yapılmıştır. Teknolojik gelişmelerin hızlı dünyasında bu tür cihaz ve ürünün özellikle gençlerin yaşamında vazgeçilmez bir unsur olduğu aşikardır. Aşağıda belirtilen çalışmalarda da belirtildiği gibi ciddi sağlık ve sosyal problemlere yol açan bu tür ürünlerin ilerleyen dönemlerde kontrol edilemez ya da düzenlenemezse kişi ve toplum sağlığı açısından nasıl bir olumsuz tabloyla karşılaşılacağına ipuçlarını göstermesi bakımından önem arz etmektedir.

Kore’de yapılan bir çalışma, cep telefonu bağımlılığının düşük uyku kalitesi riskini artırdığı, fakat uyku süresini düşürmediğini saptamıştır (Lee vd., 2017).

Garrison, Liekweg, ve Christakis (2011) çalışmalarında, İç içe (Nested) doğrusal regresyon modeli üzerinden çocukların uyku kalitelerini araştırmış sonuç olarak uyku öncesi şiddet içerikli film ya da video izleyen çocukların uyku problemi yaşadığını saptamıştır. Ayrıca medya kullanımının, özellikle akşamları veya yatak odasında televizyon ve bilgisayar gibi elektronik eşyaların bulunmasının çocukları uyku kalitesini olumsuz etkilediği belirlenmiştir.

Arora vd. (2014), tüm teknoloji türlerinin sık kullanımı ile uyku süresi (saat) arasında ters orantı olduğunu ve uyku öncesi sık müzik dinleyenler ile video oyunu oynayanların uyku başlangıç sürelerinin önemli ölçüde uzamış olduğunu tespit etmişlerdir. Ayrıca erken uyanma sıklığının, tüm teknoloji türlerinin sık kullanımıyla anlamlı şekilde ilişkili olduğu bu çalışmada saptanmıştır. En büyük etki, sık televizyon izleyicilerinde gözlenmiştir. Sonuç olarak uykuya dalmakta güçlük çekmek ile, cep telefonu kullanımı, video oyunu oynama ve internet kullanımı arasında büyük ölçüde ilişki olduğu ve en belirleyici olanın ise müzik dinlemek olduğunu vurgulamıştır.

Dworak ve Wiater (2013) çalışmalarında, çocuklarının medya kullanımlarını irdelemiş ebeveyn ve eğitimcilerin farkındalıklarını sorgulamışlardır. Bu çalışmada ayrıca çocuk ve ergen yatak odasının TV için uygun olmadığı ve bilgisayarla harcanan

zamanın yanı sıra TV ve video oyunlarının içerikleri de dahil olmak üzere toplam medya kullanımı hakkında talimatlar sağlanması gerekliliği sonucun varılmıştır.

Do vd. (2013) çalışmalarında internet kullanımında harcanan zamanı dikkate alarak, bildirilen uyku süresinin ergen sağlığı sonuçları ile olan ilişkilerini incelemiştir. Sonuç olarak Güney Kore'deki okul içi ergenler arasında aşırı İnternet kullanımının sadece olumsuz sağlık sonuçlarını doğurmadığı, aynı zamanda uyku yoksunluğu nedeniyle birçok olumsuz etkileri olabileceğini de saptamışlardır.

Aşırı İnternet kullanımının gündüz aşırı uyku hali ile ilişkisini inceleyen Choi vd. (2009), ergenlerde İnternet bağımlılığı ile gündüz aşırı uyku hali arasında güçlü bir ilişki olduğunu göstermişlerdir.

İnsanların günlük yaşamının olmazsa olmaz bir unsuru haline gelen cep telefonlarının sundukları hizmet ve iletişim olanakları bakımından faydaları saymakla bitmez. Bununla birlikte diyalektik olarak zıttının da ortaya çıkması şüphe götürmez. Aygıtla olan bağımlılık günlük faaliyetlere uyum sağlama güçlükleri ve diğer psiko-sosyal bir çok problemi de beraberinde getirir. Bir çok çalışmada cep telefonu bağımlılığının olumsuz etkileri üzerine araştırma yapılmıştır.

Yapılan literatür araştırmalar neticesinde cep telefonu kullanımının uyku düzeni üzerine etkisinin olup olmadığını dair bir çalışma saptanamamıştır. Bu amaçla yukarıda aktarılan bilgiler ve literatür taraması üzerinden 16 tanesi demografik olan 42 maddelik bir soru havuzu oluşturulmuştur. Carpenter'in (2018) önerilerine uygun olarak, soru ifadelerinin muğlaklığını önlemek ve anlaşılabilirliğini geliştirmek, konuya ve amaca uygunluğunu sağlamak için ilgili alan ve ölçek geliştirme uzmanlarına başvurulmuştur. İçerik açısından geçerliği incelenen soru havuzundan uzman görüşleri dikkate alınarak 1 soru elenmiş ve soru sayısı demografik sorular dışında 25'e düşürülerek taslak ölçek hazır hale getirilmiştir. Sonraki adımda taslak ölçek, 70 kişilik deney grubuna uygulanarak, soru ifadeleri anlaşılabilirlik ve geçerlilik testine tabi tutulmuştur. Ölçek, Hiçbir zaman'dan (1), Her zaman'a (3) cevap seçeneklerinin bulunduğu üçlü Likert olarak tasarlanmıştır.

5.1.3. Metodoloji

Bu araştırmanın ana kütlesi Türkiye'nin farklı bir çok yerinden göç alması sebebiyle İstanbul ve Kocaeli illeri seçilmiştir. Araştırma için rassal örnekleme yoluyla seçilen 654 kişiye yüz yüze anket uygulanmıştır. Toplanan anketler içerisinde eksik gözlem içerenler arındırıldıktan sonra, 491 kişilik anket verileri bir istatistik yazılımı aracılığıyla analiz edilmiştir.

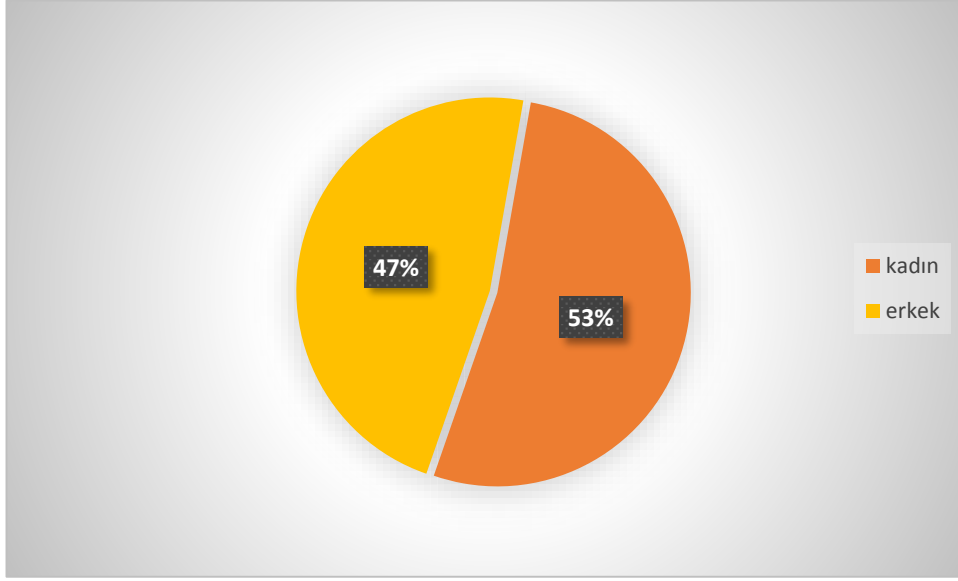
Ölçeğin faktör yapısının oluşturulmasında Açıklayıcı Faktör Analizi kullanılmıştır. Analiz sonucunda, Kaiser Meyer Olkin (KMO) değeri 0.865, Barlett Testi sonucu $p < 0.00$ çıkmıştır. Bunun anlamı, değişkenler arasında ilişki olmadığını test eden sıfır hipotez reddedilerek, verilerin faktör analizine uygunluğunun onaylanmasıdır.

Ölçeğin güvenirliği için Cronbach α katsayısına (Cronbach, 1951) bakılmıştır. Cronbach's Alpha katsayısı 0.851 çıktığından ölçeğin güvenirliğinin çok iyi olduğu söylenebilir. Çalışmada faktör sayısına herhangi bir sınırlama getirilmemiş ve öz değeri 1.00'den büyük olan faktörler ölçeğe alınmıştır. Ayrıca analize Varimax rotasyon işlemi uygulandı. Bu işlem sayesinde, daha az değişkenle faktör varyanslarının maksimum olmasını sağlayacak rotasyon yapılmış olur (Kline, 1994:64). Analiz sonucunda 6 faktör ve 25 değişkenden oluşan bir ölçek geliştirildi. Bu altı faktörün, ölçeği açıklama oranı % 58,2 olarak saptanmıştır. Bu veriler ışığında ölçeğin araştırmada kullanılması uygun bulunmuştur.

5.1.3.1. Demografik Özellik Bilgileri

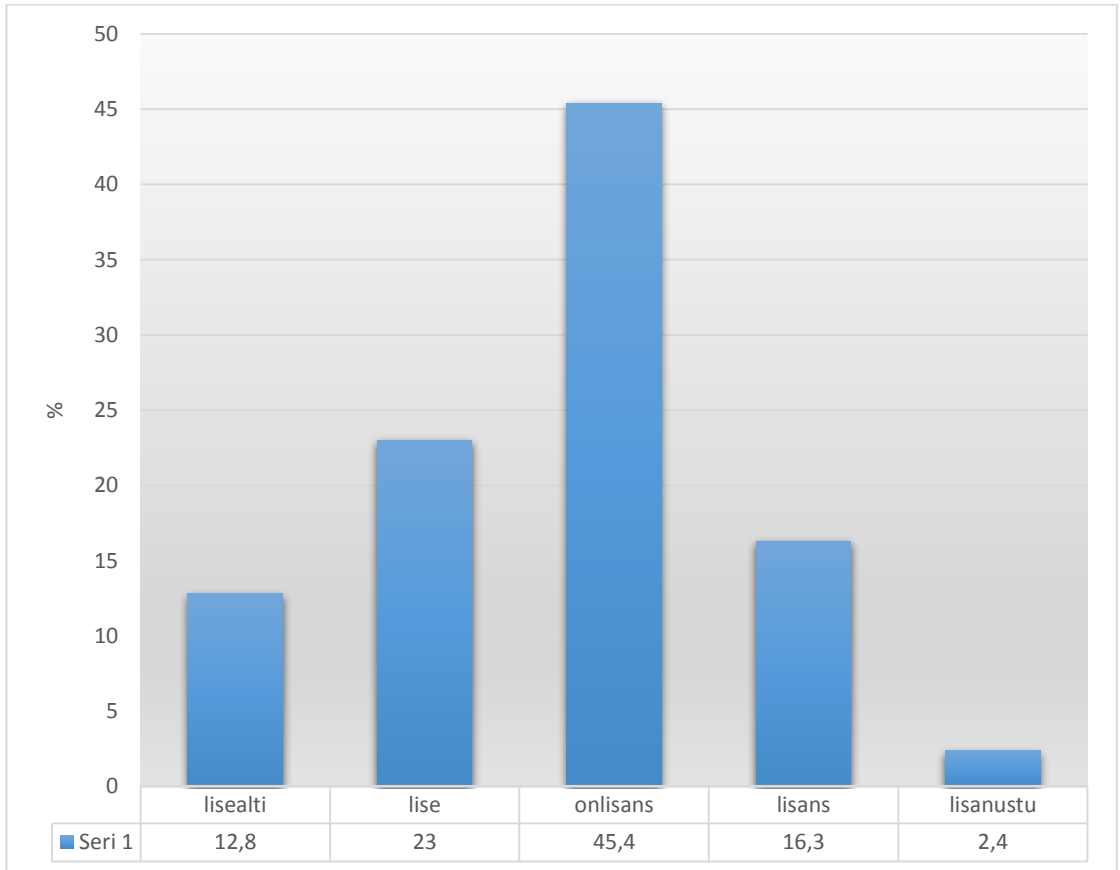
Ankete katılım gösteren kişilerin demografik yapısını gösteren bilgileri Şekil 9, 10 ve 11' de gösterilmiştir.

Katılım oranına bakıldığında en çok kadın katılımcı (258 kişi) olduğu görülmektedir. Ayrıca ankete katılanların yaş ortalaması 27,5 yıl olarak saptanmıştır.



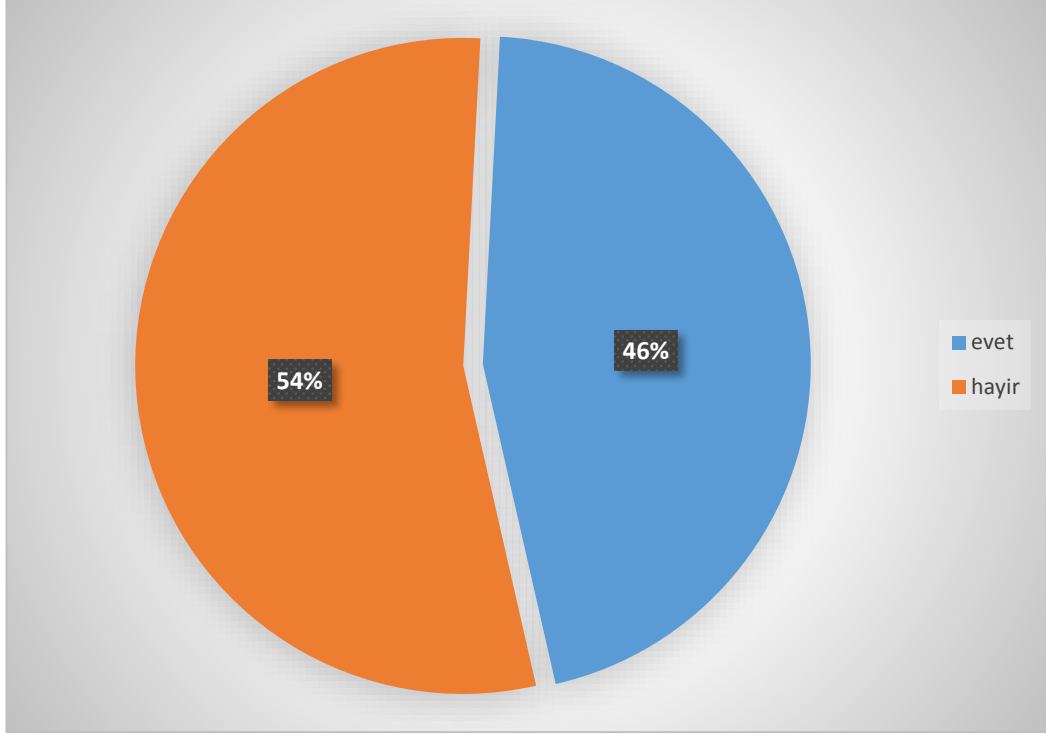
Şekil 9: Cinsiyet Bilgileri

Katılımcıların mezuniyet durumlarını gösteren Şekil 10 incelendiğinde, daha çok ön lisans ve üzeri tahsil derecesine sahip kişiler yoğunluk oluşturmuştur.



Şekil 10: Katılımcıların Mezuniyet Durumu

Bağımlı değişken sorusu olan “Cep telefonu kullanımı, uyku düzeninizi olumsuz etkiler mi?” Sorusuna verilen cevapların dağılımı Şekil 11’da gösterilmiştir.



Şekil 11: Cep Telefonu Kullanımı, Uyku Düzeninizi Olumsuz Etkiler Mi?

Ankete cevap verenlerin yarıya yakını (221 kişi), uyku öncesi cep telefonu kullanımının uykuları üzerine olumsuz etkide bulunduğunu beyan etmiştir.

5.1.3.2. Veri Yapısı

Ölçek; 1 bağımlı, 25 bağımsız değişkenden oluşmaktadır. Bu değişkenler aşağıda sembolleriyle tanımlanmıştır.

Tablo 16 : Değişkenler

Sembol	Değişken adı
Y	Cep telefonu kullanımı, uyku düzeninizi olumsuz etkiler mi?
X1	Gece yatağında uyumadan önce cep telefonumdan arayanlar olur.
X2	Gece yatağında uyumadan önce cep telefonumla birilerini ararım.
X3	Gece yatağında uyumadan önce cep telefonumdan sosyal medya hesaplarıma göz atarım.
X4	Gece yatağında uyumadan önce cep telefonumdan kısa mesajlar yazarım.
X5	Gece yatağında uyumadan önce cep telefonuma kısa mesajlar gelir.
X6	Gece yatağında uyumadan önce cep telefonumda oyun oynarım.
X7	Gece yatağında uyumadan önce cep telefonumdan video izlerim.
X8	Gece yatağında uyumadan önce cep telefonumu kulaklık ile kullanırım.
X9	Gece yatağında uyumadan önce cep telefonu uygulamalarını (aplikasyonlarını) incelerim.
X10	Gece yatağında uyumadan önce cep telefonumdan müzik dinlerim.
X11	Gece yatağında uyumadan önce cep telefonumdan film izlerim.
X12	Gece yatağında uyumadan önce cep telefonumu tamamen kapatırım.
X13	Gece yatağında uyumadan önce cep telefonumu sessize alırım.
X14	Gece yatağında uyumadan önce cep telefonumu titreşime alırım.
X15	Gece yatağında uyumadan önce cep telefonumu uçak moduna alırım.
X16	Gece yatağında cep telefonu elimde uykuya dalarım.
X17	Gece yatağında cep telefonu kulaklığı kulağında uykuya dalarım.

- X18 Gece uyuduktan sonra cep telefonuma çağrılar gelir.
- X19 Gece uyuduktan sonra cep telefonuma kısa mesajlar gelir.
- X20 Gece uyuduktan sonra uykumun belli bir anında kendiliğimden uyanırım, cep telefonuma göz atarım ve uyurum.
- X21 Uykumun ortasında cep telefonuma göz atmak isterim ve bunun için alarmını kurarım.
- X22 Gece uyumadan önce cep telefonumu yastığımın altına koyarım.
- X23 Gece uyumadan önce cep telefonumu kolay erişebileceğim bir yere koyarım.
- X24 Sabahları cep telefonumun alarmıyla uyanırım.
- X25 Sabah uyandığımda ilk önce cep telefonuma göz atarım.
-

Sonuç olarak bu 25 bağımsız değişken içerisinde bağımlı değişkeni ile anlamlı ilişkiye sahip bir model aranacaktır. Bunun için mümkün alt setlerin hepsinin hesaplanması zor olacağından yöntem olarak *Genetik Algoritma* yöntemine başvurulacaktır. Ama öncelikle tamamı kategorik olan değişkenleri sayısallaştırmak için *Optimal Ölçekleme* tekniği kullanılacaktır. Böylece *Regresyon Modeli* varsayımlardan aranır.

5.1.3.3. Verilerin Sayısallaştırılması

Çalışmada kullanılan verilerin ham halinin özet gösterimi Tablo 17’de verilmiştir.

Tablo 17: Ham Veri Matrisi (H)

	Y1	X1	X2	X3	X4	X20	X21	X22	X23	X24	X25
1	1	2	2	3	2	1	1	1	3	3	2
2	1	3	3	3	3	3	3	3	3	3	3
3	2	2	2	2	2	3	3	3	3	3	3
4	2	3	3	3	3	3	3	3	3	3	3
5	1	3	3	3	2	3	3	3	2	3	1
6	2	3	3	3	3	3	3	3	3	3	3
7	2	1	3	1	3	3	3	2	2	1	1
8	1	3	2	3	2	3	3	2	3	2	2
9	1	1	2	3	2	3	3	3	2	2	3
10	2	1	3	2	2	3	3	2	2	2	1
11	1	1	2	2	3	2	3	2	1	3	3
12	2	3	2	2	1	2	3	2	1	2	3
.
.
.
.
471	2	1	1	2	2	1	1	1	3	3	2
472	1	2	2	3	3	1	1	1	3	3	3
473	2	2	1	1	2	1	1	1	1	3	2
474	2	1	1	1	1	1	1	1	2	3	2
475	1	2	1	2	2	1	1	2	3	2	2
476	2	2	2	1	1	1	1	1	3	3	1
477	1	1	1	3	1	1	1	3	3	3	2
478	1	1	1	3	1	1	1	1	3	1	2
479	2	1	1	1	1	1	1	1	2	3	1
480	2	1	1	1	1	1	1	1	1	1	1
481	2	1	1	1	1	1	1	1	1	3	2
482	2	1	1	1	1	1	1	1	1	3	1
483	2	1	1	2	1	1	1	1	3	1	2
484	2	1	1	1	1	1	1	1	3	2	2
485	2	2	2	2	2	1	1	1	2	2	3
486	2	1	1	2	1	1	1	1	3	3	2
487	1	2	2	3	1	1	1	1	3	2	2
488	1	1	1	3	1	1	1	3	3	3	2
489	1	1	1	3	1	1	1	2	2	3	3
490	1	2	2	3	3	1	1	1	1	1	1
491	2	1	1	1	1	1	1	2	2	1	1

5.1.3.4. Sayısallaştırılmış Veriler

Tablo 17’de verilen ham veriler, önceki bölümlerde anlatıldığı gibi dönüşüme uğratarak sayısallaştırıldı. Sayısallaştırılan verilerin gösterimi Tablo 18’te verilmiştir.

Tablo 18: Sayısallaştırılmış Veri Matrisi (Q)

	Y1	X1	X2	X3	X4	X21	X22	X23	X24	X25
1	0,24	0,11	-0,12	0,18	0,01	0,22	-0,36	-0,11	-0,06	-0,52
2	-0,28	-1,11	-1,02	-0,44	-0,9	-1,13	-0,61	-0,11	-0,03	-0,52
3	0,24	0,11	-0,12	0,18	0,01	-1,13	-0,61	-0,11	-0,03	-0,52
4	0,24	0,11	-0,12	0,18	0,01	-1,13	-0,61	-0,11	-0,03	-0,52
5	0,24	0,11	-0,12	-0,44	0,01	-1,13	-0,61	-0,11	-0,03	-0,52
6	0,24	-1,11	-1,02	-0,44	-0,9	-1,13	-0,61	-0,11	-0,03	-0,52
7	-0,28	-1,11	-1,02	-0,44	0,01	-1,13	-0,61	0,04	-0,03	0,57
8	-0,28	-1,11	-0,12	-0,44	0,01	-1,13	-0,36	-0,11	-0,06	0,35
9	0,24	0,11	0,54	0,18	0,01	-1,13	-0,61	-0,11	-0,06	0,35
10	0,24	-1,11	-0,12	0,18	0,01	-1,13	-0,36	0,04	-0,06	-0,52
11	-0,28	0,51	-0,12	0,18	-0,9	-1,13	-0,36	0,51	-0,03	-0,52
12	0,24	0,51	0,54	0,18	0,01	-1,13	-0,61	0,51	0,19	-0,52
.
.
.
482	0,54	0,94	0,88	0,89	0,54	0,22	0,58	0,51	-0,06	0,57
483	0,54	0,94	0,88	0,89	0,54	0,22	0,58	0,04	-0,06	-0,52
484	0,54	0,94	0,88	0,14	0,54	0,22	0,58	0,04	-0,03	0,57
485	0,54	0,94	0,88	0,89	0,54	0,22	0,58	0,51	-0,03	0,35
486	0,54	0,94	0,88	0,89	0,54	0,22	0,58	0,51	-0,03	0,57
487	0,54	0,18	0,88	0,89	0,54	0,22	0,58	-0,11	-0,03	0,35
488	0,54	-0,44	0,88	0,89	-0,88	0,22	-0,36	0,04	-0,03	-0,52
489	-0,12	-0,44	-0,9	-0,92	0,54	0,22	0,58	0,51	0,19	0,57
490	0,54	0,94	0,88	0,89	0,54	0,22	-0,36	0,04	0,19	0,57
491	-0,12	-0,44	0,01	0,14	0,54	0,22	-0,36	-0,11	-0,03	-0,52

5.2. ANALİZ

Sayısal dönüşüme uğrayan değişkenleri kullanarak en uygun regresyon modeli oluşturmak çalışmanın asıl hedefidir. Bunun için mümkün alt setler içerisinde amaç fonksiyonuna en uygununu Genetik Algoritma yöntemi ile araştırılmıştır.

İlk olarak gen popülasyonun kaç bireyden oluşacağına karar verilmiştir. Yukarıda da belirtildiği gibi sayı ne çaprazlama (cross-over) imkanını sekteye uğratacak kadar az, ne de algoritmanın çalışmasına yavaşlatacak kadar çok olmalıdır. Yapılan simülasyon neticesinde popülasyon büyüklüğünün (N) 30 olması uygun görülmüştür.

Çaprazlama türü olarak literatürde en çok rağbet gören ‘İki noktalı çaprazlama (two point cross-over)’ yöntemi kullanılmış ardından algoritmanın belirli bir bölgede tekrara düşmesini engelleyecek ve evrimleştirecek mutasyon oranı veya olasılığı (p_m) katsayısı belirlenmiştir. Bu sayı yaygın kullanıldığı haliyle 0,01 yani %1 olarak atanmıştır.

Ayrıca algoritmaya opsiyonel olarak ‘Elitizm Kuralı’ şartı konulmuştur. Böylece kaliteli ebeveynlerdin en az bir tanesi yeniden oğul döllere değişmeksizin geçecektir. Algoritmanın bu şekilde en fazla 50 kuşak çalışmasına izin verildi. Bunun nedeniyle algoritmanın belirli bir durgunluk halinden sonra gereksiz performans sergilemesini önlemektir. Seçim işlemleri belirlendikten sonra, gerekli kodlamalar MATLAB 2015a programında işlenmiştir. Mümkün alt setler içinde en iyi 30 regresyon modelini gösteren tablo aşağıda verilmiştir.

Tablo 19 : GA sonucunda elde edilen en iyi Regresyon modelleri ve ICOMP(IFIM) Skorları

Sıra	Kromozomlar(İkili String)	Değişkenler	ICOMP(IFIM)
1	0000000001010000000000111	10-12-23-24-25	672,22
2	0000100001010000000000111	5-10-12-23-24-25	672,86
3	0000000001010010000000111	10-12-15-23-24-25	672,96
4	1000000001010000000000111	1-10-12-23-24-25	673,02
5	0000000001110000000000111	10-11-12-23-24-25	673,20
6	0000000001010000100000111	10-12-17-23-24-25	673,28
7	0000000001011000000000111	10-12-13-23-24-25	673,35
8	0000000001010000000001111	10-12-22-23-24-25	673,42
9	0100000001010000000000111	2-10-12-23-24-25	673,55
10	0010000001010000000000111	3-10-12-23-24-25	673,62
11	0000000001010000000100111	10-12-21-23-24-25	673,70
12	0000010001010000000000111	6-10-12-23-24-25	673,77
13	0000000001010000010000111	10-12-18-23-24-25	673,88
14	0000100000010000000000111	5-12-23-24-25	674,03
15	0000100100010000000000111	5-8-12-23-24-25	674,11
16	0000100100010000000000110	5-8-12-23-24	674,13
17	0000100100010000000000101	5-8-12-23-25	674,19
18	0000100100010000000000011	5-8-12-24-25	674,25
19	0001100100010000000000111	4-5-8-12-23-24-25	675,33
20	1000100000011000001000001	1-5-12-13-19-25	677,45
21	0101000100010001000000100	2-4-8-12-16-23	678,03
22	1100001001010001001000000	1-2-7-10-12-16-19	679,22
23	0011010100001000000011000	3-4-6-8-13-21-22	680,23
24	0110010001000000011001010	2-3-6-10-18-19-22-24	682,75
25	0010100011010100001001000	3-5-9-10-12-14-19-22	684,36
26	1001001100011000000000011	1-4-7-8-12-13-24-25	690,10
27	0011000110000010000110000	3-4-8-9-15-20-21	696,54
28	1011010101000000000000100	1-3-4-6-8-10-23	702,22
29	1000000000000000000000001	1-25	714,86
30	0100110001010000000000110	2-5-6-10-12-24-25	745,64

Algoritmanın sonuç ekranı aşağıda gösterilmiştir.

GA gave:

```
0000000001010000000000111    {10 12 23 24 25} as the best subset.

%%-- GA parameters
N = 30;    population size
ncrosspoints = 2; %Two point cross over
uniform = 0;    %CrossOver type 0=No uniform
mutrate = 0.1;
elitism = 1;
Ngen = 50;

%-----

%Analysis of the best subset {10 12 23 24 25}

logLike: -332.58
G02: 10.159
dev: 665.17
pearson: [490x1 double]
standardized: [490x1 double]
leverage: [490x1 double]
dbeta: [490x1 double]
sens: [21x1 double]
spec: [21x1 double]
AIC: 677.17
ICOMP: 672.22
VIF: [5x1 double]
probs: [490x1 double]
auc: 0.57946
```

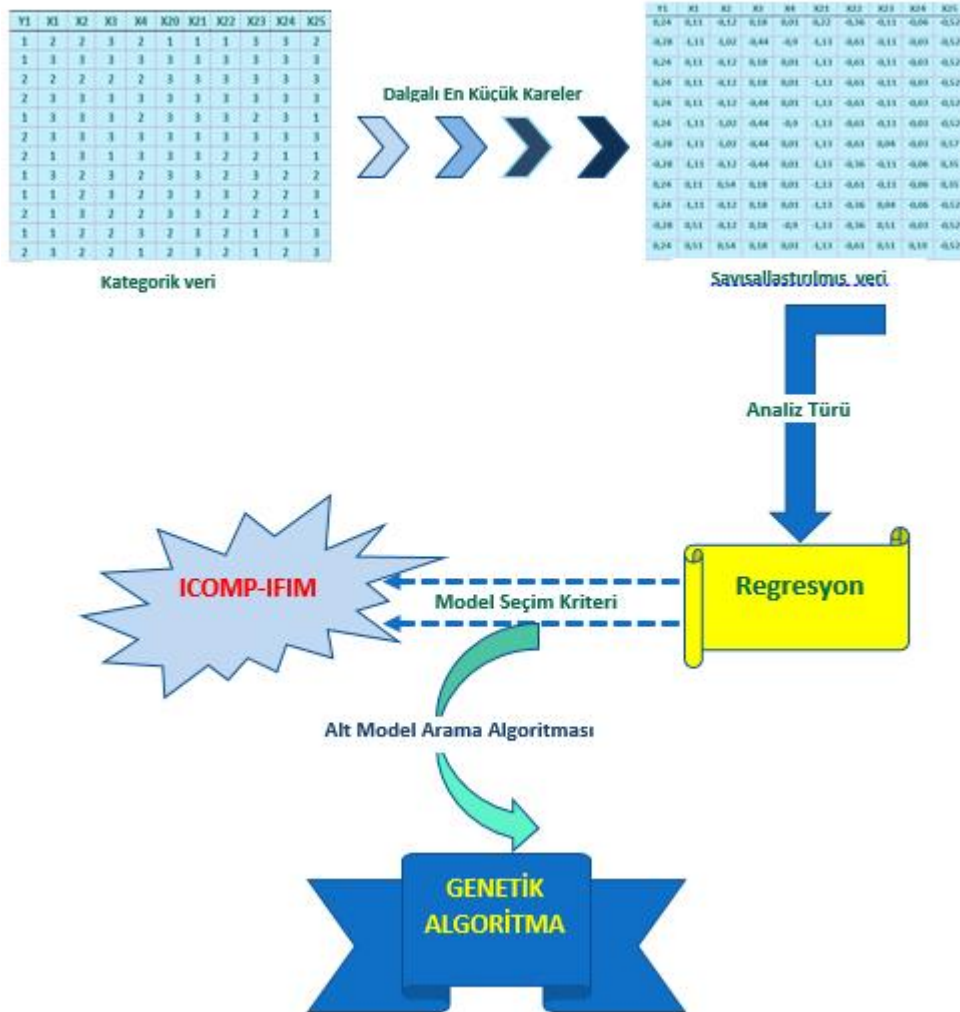
Tablo 16’da görüldüğü gibi en son kuşağın oluşturduğu popülasyon içerisinde en düşük ICOMP(IFIM) skoru listenin en başındaki gen kombinasyonudur. Sonuca göre datamızdaki 10, 12, 23, 24 ve 25 numaralı değişkenlerin oluşturduğu regresyon modeli en uygun alt set olarak tespit edilmiştir.

SONUÇ VE TARTIŞMA

Bu tezde ulaşılmak istenen temel hedef; her türden kategorik veri için uygulanabilecek evrensel bir bütünleşik analiz sistemi oluşturmaktır. İlk olarak analizlerde varsayım kısıtı sorununu aşmak için veriyi sürekli hale getirmek adına dönüşüm operatörü olarak Optimal Ölçekleme işlemi uygulandı. Bu işlem için minimum kayıpla verinin dönüşümünü sağlayan Dalgalı En Küçük Kareler Yöntemi devreye girdi. Literatürde ve özellikle ülkemizde kendine yeni uygulama alanı bulan yöntem, ilerleyen zamanlarda alanın ihtiyacını karşılayacak esnekliğe sahip olacak bir pozisyonadadır.

İkinci aşama olarak, sayısallaştırılmış veriye Regresyon Analizi uygulaması gerçekleştirilmiştir. Burada da bilindiği üzere Regresyon Analizi uygulamalarında sıklıkla tercih edilen Çoklu Belirlilik Katsayısı (R^2) Kriteri yerine, Bilgi Karmaşıklığı kriteri olarak anılan ICOMP (IFIM) model seçim kriteri baz alınmıştır. Bu da çalışmamızın orijinalliği açısından bir gösterge sunmaktadır. Bu kriter bağımlı değişkeni açıklayacak en yalın , tutarlı karmaşık olmayan modeli seçmek adına tercih edilmiştir.

Üçüncü aşamada ise, en uygun alt model seçimini aramak adına başvurulacak yöntem belirlenmiştir. Bunun için en uygun sezgisel yöntemlerden olan Genetik Algoritmaya başvurulmuştur. Geniş bir veri uzayında arama yaparak, istenilen amaç fonksiyonuna en uygun lokasyonu arayan yöntem, bize modeli en iyi temsil edebilecek çözümü sunmuştur. Şekil 12 tezin uygulama akış şemasını göstermektedir.



Şekil 12: Tez Uygulama Akış Şeması

Bu üç yöntemi kullanarak yapılan analiz yöntemi ilk kez bu çalışmada literatüre sunulmuş olup, sıralı ve sınıflayıcı verilerle çalışacak farklı alanlardaki araştırmacılara önerilmektedir.

Modelin uygulama aşaması için oluşturduğumuz verinin araştırmanın konusu, belli bir zaman ve mekânda akıllı telefon kullanımıyla alakalı davranışlardır. Zaman, somut bir biçimde, geceleyin yatak odasına çekildikten sonra başlayan, uykuya kadar geçen süredir. Mekân ise, yatak odasıdır. Bu araştırma, cep telefonunun uykuya ayrılan zamana dahil olması ve uyku öncesinde cep telefonu kullanımıyla kendini gösteren, uykuyu geciktiren ve uyku süresini kısaltan davranışlar üzerine odaklanmıştır.

Araştırmanın genel amacı, akıllı telefonu en çok kullanan genç yetişkinlerin, yatak odasında uyku öncesi akıllı telefon kullanma davranışlarını belirlemektir.

Bu çerçevede çalışmamızın temel amacı; cep telefonu kullanımının, uyku düzeni üzerine etkisinin olup olmadığını, var ise hangi değişkenlerce açıklandığını tespit etmektir. Gerekli kodlamalar ve dönüşümler yapıldıktan sonra algoritma çalıştırılmıştır. Algoritmanın ICOMP (IFIM) skoruna göre bize sunduğu çözümde 5 bağımsız değişken yer almıştır. Bu değişkenlerin bilgilerini içeren Tablo 20’de sunulmuştur.

Tablo 20: Regresyon Modeli ve Değişkenler

Sembol	Değişken adı
X10	Gece yatağımda uyumadan önce cep telefonumdan müzik dinlerim.
X12	Gece yatağımda uyumadan önce cep telefonumu tamamen kapatırım.
X23	Gece uyumadan önce cep telefonumu kolay erişebileceğim bir yere koyarım.
X24	Sabahları cep telefonumun alarmıyla uyanırım.
X25	Sabah uyandığımda ilk önce cep telefonuma göz atarım.

Bu sonuca göre Cep telefonu kullanımının, uyku düzeni üzerine olumsuz bir etki olup olmadığını belirleyen en yalın ve anlamlı değişkenlerin Tablo 20’deki beş değişken ile açıklanmıştır. Bu ifadelere verilen cevaplar üzerinden cep telefonunu bağımlılığının uyku kalitesi üzerindeki olumsuz etkisini saptayacak modele ulaşmıştır.

Bu tez çalışmasının literatürde disiplinlerarası bir katkıya sahip olabileceği düşünülmektedir. Bu açıdan bakıldığında araştırma konusu, kapsamı ve sonuçlarıyla psikoloji alanında çalışan araştırmacıların ilgisini çekebilir. Ayrıca ölçek birçok değişik kültür ve lokasyonda uygulanarak karşılaştırmalı analize tabi tutulabilir. Kısıt olarak değişkenlerin üçlü likert tipinden ziyade, daha yüksek aralıklarda (beşli ya da yedili likert) düzenlenmesi daha duyarlı sonuçlara ulaşılmasını sağlayabilir.

Son olarak alıřma, kullanılan yntemler bakımından deęerlendirildięinde ise, regresyon analizinde Genetik Algoritma Ynteminin kullanılması klasik regresyon yntemlerine (Stepwise Regresyon, Ridge Regresyon vb...) karřı etkin bir alternatif olarak nerilmekle birlikte, en uygun modelin belirlenmesi bakımından klasik uygunluk llerinin (R^2 , AIC, FPE, Cp testleri vb...) yerine ICOMP (IFIM) bilgi kriterinin kullanılması da tezin orijinal bir nitelięe sahip olduęunu ortaya koymaktadır. Bunun gibi bir ok teknikler btnleřik olarak kullanılarak yeni arařtırma yntemlerinin geliřtirilmesi ileri istatistik alıřmalarında nem arz edebilecektir.

KAYNAKÇA

- ABAR,H.,
KARAASLAN A.: 2013
“Konut Talep Edenlerin Özellikleriyle Talep Edilen Konutun Özellikleri Arasındaki İlişkinin Çoklu Uyum Analizi Yöntemi İle İncelenmesi: Atatürk Üniversitesi Personeli Örneği,” **Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi**, 27, 323–39.
- AKAIKE, H. : 1973
“Information Theory and an Extension of the Maximum Likelihood Principle,” **Proceedings of the 2nd International Symposium on Information Theory**, 267–81.
- AKÇA, N.,
IŞIKHAN S. Y.: 2017
“Hacettepe Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Öğrencilerinin İş Değerleri ve Kariyer Seçimlerinin Doğrusal Olmayan Kanonik Korelasyon Analizi İle Değerlendirilmesi,” **İşletme Bilimi Dergisi**, 5, 49–70.
- AKDİ, Y.: 2011
Matematiksel İstatistiğe Giriş, Kızılay - Ankara: Gazi Kitabevi.
- ÅKERSTEDT, K.,
HUME, D.,
MİNORS, J.,
WATERHOUSE, T.: 1994
“The Subjective Meaning of Good Sleep, An Intraindividual Approach Using the Karolinska Sleep Diary,” **Perceptual and Motor Skills**, 79, 287–96.
- AKSU, D. M.: 2010
“Birinci Basamakta Uyku Bozukluklarına Yaklaşım ve Öneriler,” **Turkish Journal Of Family Medicine And Primary Care**, 4, 11.
- AKTÜRK, D.:2015
“Geleneksel ve Organik Olarak Yetiştirilen Tarım Ürünlerine Karşı Tüketici Taleplerinin Doğrusal

- Olmayan Kanonik Korelasyon Analizi ile Değerlendirilmesi,” **ÇOMÜ Ziraat Fakültesi Dergisi**, 3, 115–21.
- ALMEIDA,A.,
GARROD. B.: 2018 “A CATREG model of destination choice for a mature Island destination,” **Journal of Destination Marketing & Management**, 8, 32–40.
- ALPAR, R.: 2003 **Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1**, Ankara: Nobel Yayın Dağıtım.
- ALTAŞ, D.,
GIRAY. S.: 2013 “Dünyadaki En Önemli Sorun Algısının Optimal Ölçeklemeli Çok Değişkenli İstatistiksel Yöntemler İle İncelenmesi,” **Öneri Dergisi**, 10, 135–42.
- ALTAŞ,D.,
KANDUR. H.:2018 “Tercih Verilerinin İdeal Nokta Ve Vektör Modelleri İle Analizi,” **Uluslararası İktisadi ve İdari İncelemeler Dergisi**, 459–72.
- ALTAY, A.: 2015 **Genetik Algoritma Ve Bir Uygulama**, Fen Bilimleri Enstitüsü.
- ANASTASIOS, M.,
KOUTSOURIS, A.
KONSTADINOS. M.: 2010 “Information and Communication Technologies as Agricultural Extension Tools: A Survey among Farmers in West Macedonia, Greece,” **The Journal of Agricultural Education and Extension**, 16, 249–63.
- ANCOLİ-ISRAEL, S.,
POCETA, J. S..
STEPNOWSKY C,
MARTİN J.,
GEHRMAN P.: 1997 “Identification And Treatment Of Sleep Problems In The Elderly,” **Sleep Medicine Reviews**, 1, 3–17.

- ARORA, T.,
BROGLIA E.,
THOMAS G. N.,
TAHERI S .:2014 “Associations Between Specific Technologies And Adolescent Sleep Quantity, Sleep Quality, And Parasomnias,” **Sleep Medicine**, 15, 240–47
- ARSLAN, U.: 2008 **Bilgi Kuramı Yaklaşımı, Kullback-Leibler Uzaklığı Ve Diğer Uzaklık Ölçüleri İle Tanı Testlerinin Değerlendirilmesi Ve Karşılaştırılması**, Hacettepe Üniversitesi / Sağlık Bilimleri Enstitüsü.
- ATA, N.: 2007 “Yaşam Verisi İçin Homojenlik Analizi,” *İnönü Üniversitesi Tıp Fakültesi Dergisi*, 14, 251–55.
- BAĞCI, İ.: 2017 **Kısmi En Küçük Kareler Yönteminin Simülasyon Verileri İle Diğer Yöntemlerle Karşılaştırılması**, Muğla: Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü
- BAKIRCI, Ç. M.: 2011 “EvrİM Mekanizmaları – 9 : Crossing-Over Kromozomal Gen Değişimi ,” *EvrİM Ağacı*, <https://evrimagaci.org/evrim-mekanizmaları-9-crossingover-kromozomal-gen-degisimi-244>, Accessed 06/21/2019. “Darwin’in Evrim Teorisi Nedir, Neler Söyler?,” **EvrİM Ağacı** .
- BALCI, K.: 2017 **Başkent Üniversitesi Ankara Hastanesi’nde Çalışan Yetişkin Bireylerin Beslenme Durumları İle Uyku Kalitesi Arasındaki İlişkinin Değerlendirilmesi**, Başkent Üniversitesi Sağlık Bilimleri Enstitüsü Beslenme Ve Diyetetik Bölümü
- BAYATA, H.,
HATTATOĞLU F.: 2014 “Yapay Sinir Ağları Ve Çok Değişkenli İstatistik Yöntemlerle Trafik Kaza Modellemesi,” **Erzincan**

- Üniversitesi Fen Bilimleri Enstitüsü Dergisi**, 3, 207–19.
- BAYRAM N,
ERTAŞ S.: 2001 “Tüketim Harcamaları Davranış Biçimi: Princals ve Overals Yaklaşımı,” **V. Ulusal Ekonometri ve İstatistik Sempozyumu**, Adana.
- BEDWARD, M.,
KEİTH, D. A.
PRESSEY. R. L.: 1992 “Homogeneity Analysis: Assessing The Utility Of Classifications And Maps Of Natural Resources,” **Australian Journal of Ecology**, 17, 133–39.
- BİÇER, H. D.,
BİÇER, C.: 2017 “Gamma ve Weibull Dağılımları Arasında Kullback-Leibler Uzaklığına Dayalı Ayrım,” **Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi**, 17, 500–505.
- BİRCAN, H.: 2004 “Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama,” **Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi**, 185–208.
- BOHACEK, S.,
JONCKHEERE, E.: 1998 “Chaotic Modeling in Network Spinal Analysis: Nonlinear Canonical Correlation with Alternating Conditional Expectation ACE : A Preliminary Report,” **Journal of Vertebral Subluxation Research**, , 188–95.
- BOZDOĞAN, H.: 1987 “Model selection and Akaike’s Information Criterion AIC : The general theory and its analytical extensions,” **Psychometrika**, 52, 345–70.
- BOZDOĞAN, H.:
1993 “Choosing the Number of Component Clusters in the Mixture-Model Using a New Informational Complexity Criterion of the Inverse-Fisher Information Matrix,” Ed. by. Otto Opitz, Berthold Lausen, and Rüdiger Klar **Information and Classification**, . Studies in

- Classification, Data Analysis and Knowledge Organization Springer Berlin Heidelberg, 40–54.
- BOZDOGAN, H.: 2003a **Statistical Data Mining and Knowledge Discovery**, New York: CRC Press.
- BOZDOGAN, H.: 2003b “Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms,” in **Statistical Data Mining and Knowledge Discovery**, ed. by Bozdogan, H. Chapman and Hall/CRC.
- BOZDOGAN, H.,
HAUGHTON M. A.: 1998 “Informational Complexity Criteria For Regression Models,” **Computational Statistics & Data Analysis**, 28, 51–76.
- BÜLBÜL, Ş.,
GİRAY, S.: 2012 “İş ve Özel Yaşam İş Dışı Yaşam Memnuniyeti Arasındaki İlişki Yapısının Doğrusal Olmayan Kanonik Korelasyon Analizi ile İncelenmesi”, *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 12/4, 101-114.
- BULUT, Y. M.: 2011 **Çoklu İç İlişki Durumunda Kısmi En Küçük Kareler Regresyonu**, Eskişehir Osmangazi Üniversitesi-Fen Bilimleri Enstitüsü.
- BURG, E.,
J. DE LEEUW.: 1988 “Use Of The Multinomial Jack-Knife And Bootstrap In Generalized Non-Linear Canonical Correlation Analysis,” **Applied Stochastic Models And Data Analysis**, 4, 159–72.
- BURG, E.,
DE LEEUW, J.,
VERDEGAAL, R.: 1988 “Homogeneity Analysis Withk Sets Of Variables: An Alternating Least Squares Method With Optimal Scaling Features,” **Psychometrika**, 53, 177–97.

- BURNHAM, K. P.,
ANDERSON, R.: 2013 **Model Selection and Inference: A Practical Information-Theoretic Approach.**, New York, NY: Springer.
- BUYSSE, D. J.: 2014 “Sleep Health: Can We Define It? Does It Matter?,” **Sleep**, 37, 9–17.
- CAPRETZ, L.,
HO, D.: 2010 “Building an OSS Quality Estimation Model with CATREG,” **International Journal on Computer Science and Engineering**, 2, 1952–58.
- CARPENTER, S.: 2018 “Ten Steps in Scale Development and Reporting: A Guide for Researchers,” **Communication Methods and Measures**, 12, 25–44.
- CARSKADON, M. A.:
2002a **Adolescent Sleep Patterns: Biological, Social, and Psychological Influences**, Cambridge University Press.
- CHAKRABARTI, A.,
GHOSH, J. K.: 2011 *AIC, BIC and Recent Advances in Model Selection, in Philosophy of Statistics*, ed. by Bandyopadhyay, P. S. and M. R. Forster. . Handbook of the Philosophy of Science Amsterdam: North-Holland, 583–605.
- CHOI, K., SON, H.
PARK, M.,
HAN, J.,
KIM, K.,
LEE, B.,
GWAK, H.:2009 “Internet Overuse And Excessive Daytime Sleepiness In Adolescents,” **Psychiatry and Clinical Neurosciences**, 63, 455–62.
- ÇILAN, Ç. A.: 2013 **Sosyal Bilimlerde Kategorik Verilerle İlişki Analizi: Kontenjans Tabloları Analizi**, Ankara: Pegem Akademi.

- ÇİLAN, Ç. A.,
CAN, M.: 2014 “Measuring Factors Effecting MBA Students’ Academic Performance by Using Categorical Regression Analysis: A Case Study of Institution of Business Economics, Istanbul University,” **Procedia - Social and Behavioral Sciences**, 122. 2nd World Conference on Design, Arts and Education DAE-2013 , 405–9.
- CLAUSEN, S. E.: 1998 **Applied Correspondence Analysis: An Introduction**, SAGE.
- COSTANTİNİ, M.,
BERNARDİ, G.: 2008 “Correlations Between Coding And Contiguous Non-Coding Sequences In Isochore Families From Vertebrate Genomes,” **Gene**, 410, 241–48.
- CRONBACH, L. J.: 1951 “Coefficient Alpha And The Internal Structure Of Tests,” **Psychometrika**, 16, 297–334.
- AKINCI, D.,E.: 2007 **Yapısal Eşitlik Modellerinde Bilgi Kriterleri**, İstanbul: Mimar Sinan Güzel Sanatlar Üniversitesi.
- DİMİTRİOS, D.: 2018 “Smoking Dependence during Economic Crisis the Case of Greece,” **IJAME**, 5/4, 18-21.
- DO, Y. K.,
SHİN, E.,
BAUTISTA, M. A.,
FOO, K.: 2013 “The Associations Between Self-Reported Sleep Duration And Adolescent Health Outcomes: What Is The Role Of Time Spent On Internet Use?,” **Sleep Medicine**, 14, 195–200.
- DOMÍNGUEZ L. F.,
MORENO SAİZ J. C.,
OLLERO.,
H. S.: 2003 : “Rarity And Threat Relationships In The Conservation Planning Of Iberian Flora,” **Biodiversity & Conservation**, 12, 1861–82.

- DUSSELDORP, E.,
MEULMAN, J. J.: 2001 "Prediction in Medicine by Integrating Regression Trees into Regression Analysis with Optimal Scaling," **Methods of Information in Medicine**, 40, 403–9.
- DWORAK, M.,
WIATER, A. H.: 2013 "Media, Sleep and Memory in Children and Adolescents", **Journal of Sleep Disorders & Therapy**, 2/7,1-4
- EMDEM, VAN H.: 1971 **An Analysis of Complexity**, Amsterdam: Mathematical Centre Tracts.
- EVERITT, B. S.: 2006 **An R and S-Plus® Companion to Multivariate Analysis**, Springer Science & Business Media.
- FİLİZ, Z.,
ÇEMREK, F.: 2007 "Üniversite Öğrencilerinin Barınma Sorunlarının Uygunluk Analizi İle İncelenmesi," **Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi**, 8, 207–24.
- FRIE, K. G.,
JANSSEN, C.: 2009 : "Social İnequality, Lifestyles And Health – A Non-Linear Canonical Correlation Analysis Based On The Approach Of Pierre Bourdieu," **International Journal of Public Health**, 54, 213–21.
- GARRISON, M. M.,
LIEKWEG, K.
CHRISTAKIS, D. A.: 2011 "Media Use and Child Sleep: The Impact of Content, Timing, and Environment," **Pediatrics**, 128, 29–35.
- GİFİ, A.: 1991 **Nonlinear Multivariate Analysis**, reprint. with corrections. Chichester: Wiley & Sons.
- GİRAY, S.: 2001 **Doğrusal Olmayan Kanonik Korelasyon Analizi Ve Yaşam Memnuniyeti Üzerine Bir Uygulama**,

- İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.
- GOHIL, A.,
HANNON, T. S.: 2018 “Poor Sleep and Obesity: Concurrent Epidemics in Adolescent Youth,” **Frontiers in Endocrinology**, 9.
- GOLDBERG, D. E.: 1988 “Genetic Algorithms and Machine Learning,” **Machine Learning**, 3, 95–99.
- GOLOB, T. F.: 1986 “A Nonlinear Canonical Correlation Analysis Of Weekly Trip Chaining Behaviour,” **Transportation Research Part A: General**, 20, 385–99.
- GRAHAM, N. A.
JNASH, ., K. L.
KOOL, J. T.: 2011 “Coral Reef Recovery Dynamics In A Changing World,” **Coral Reefs**, 30, 283–94.
- GREENACRE, M. J.: 1999 **Correspondence Analysis in the Social Sciences: Recent Developments and Applications**, Nachdr.San Diego, Calif.: Academic Press.
- GREENE, W. H.: 2012 **Econometric Analysis**, Seventh edition, international edition. Pearson series in economics Boston Columbus Indianapolis New York San Francisco Upper Saddle Rive Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo: Pearson.
- GÜÇ, K.: 2015 **Türkiye’de Resmi Kurumlara Duyulan Güvenin Kategorik Regresyon Ve Lojistik Regresyon Analizi**

İle İncelenmesi, Gazi Üniversitesi / Fen Bilimleri Enstitüsü.

GÜÇ, K.,
BAŞAR, E.: 2016 “Categorical Regression Based on Optimal Scaling and An Application,” **Eurasian Academy of Sciences Eurasian Econometrics, Statistics & Empirical Economics Journal**, 10, 14–27.

GUNDACKER, C.,
KUTALEK, R.,
GLAUNACH, R.,
DEWEİS, C.,
HENGSTSCHLÄGER, M,
PRINZ, A.: 2017 “Geophagy During Pregnancy: Is There A Health Risk For Infants?,” **Environmental Research**, 156, 145–47.

GÜNEL, A.: 2004 **Regresyon Denkleminin Başarısını Ölçmede Kullanılan Belirleme Katsayısı ve Kritiği**, Lebib Yalkın Yayınları, 1, 199–211.

GUTTMAN, L.: 1941 “The Quantification of a Class of Attributes : A Theory and Method of Scale Construction,” **The Prediction of Personal Adjustment**,
<https://ci.nii.ac.jp/naid/10015581914/>.

HARTMANN, A.,
ZEECK A.,
BARRETT, M. S.: 2010 “Interpersonal Problems In Eating Disorders,” **International Journal of Eating Disorders**, 43, 619–27.

- HOLLAND, J. H.: 1975 **Adaptation in Natural and Artificial Systems**, Nachdr. Complex adaptive systemsCambridge, Mass.: MIT Press
- HOWE, A.,
BOZDOGAN, H.: 2010 “Predictive Subset VAR Modeling Using the Genetic Algorithm and Information Complexity | Howe | European Journal of Pure and Applied Mathematics,” **European Journal of Pure and Applied Mathematics**, 3, 382–405.
- HSIEH, W. W.: 2000 “Nonlinear Canonical Correlation Analysis By Neural Networks”, **Neural Networks**, 13, 1095–1105.
- HSIEH, W. W.: 2001 :“Nonlinear Canonical Correlation Analysis of the Tropical Pacific Climate Variability Using a Neural Network Approach,” **Journal of Climate**, 14, 2528–39.
- KACZOR, M.,
SKALSKI, M.: 2016 “Prevalence And Consequences Of İnsomnia İn Pediatric Population.,” **Psychiatria Polska**, 50, 555–69.
- KARAMAN, E.,
ARICIGİL ÇİLAN. Ç.: 2018 “Uluslararası Matematik Ve Fen Eğilimleri Araştırma TIMSS Sınavının Optimal Ölçekleme Teknikleri İle Analizi,” **TURAN-SAM Uluslararası Bilimsel Hakemli Dergisi**, 10, 122–27.
- KAZAN, H.,
KARAMAN, E.: 2018 “Çalışanların Sosyal Hayat Tutumlarının, Demografik Özelliklerine Göre Optimal Ölçekleme Teknikleri İle İncelenmesi,” **Kafkas Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi**, 9, 379–89.
- KİPER, S.,
SUNAL, N.: 2009 “Romatoid Artritli Hastalarda Uyku Kalitesinin Değerlendirilmesi,” **Kocatepe Tıp Dergisi**, 10, 33–39.

- KLINE, P.: 1994 **An Easy Guide to Factor Analysis**, Routledge:London
- KOC, E.,
BOZDOGAN, H.: 2014 “Model Selection In Multivariate Adaptive Regression Splines MARS Using Information Complexity As The Fitness Function,” **Machine Learning**, 101.
- KOOIJ, A. J. V.,
MEULMAN, J. J.
HEISER, W. J.: 2006 “Local Minima In Categorical Multiple Regression,” **Computational Statistics & Data Analysis**, 50, 446–62
- KRAAIJ, V.,
DE WILDE, E. J.: 2001 “Negative Life Events And Depressive Symptoms In The Elderly: A Life Span Perspective,” **Aging & Mental Health**, 5, 84–91.
- KULLBACK, S.,
LEIBLER, R. A.: 1951 “On Information and Sufficiency,” **The Annals of Mathematical Statistics**, 22, 79–86.
- KUMAR, S.,
MARTIN, E. B.
MORRIS, J.: 2002 “Non-linear Canonical Correlation Analysis Using a RBF Network,” **European Symposium on Artificial Neural Networks Bruges Belgium**, 507–12.
- LEEuw, J. DE.: 2011 **Beyond Homogeneity Analysis.**, Department of Statistics, UCLA.
- LEEuw, J. DE,
MAIR, P. :2007 **Homogeneity Analysis in R: The Package Homals.**, Department of Statistics, UCLA.
- LIMA, A. T.,
RODRIGUES P. C.,
MEXIA, J. T.: 2010 “Heavy Metal Migration During Electoremediation Of Fly Ash From Different Wastes—Modelling,” **Journal of Hazardous Materials**, 175, 366–71.
- LINTING, M.,
MEULMAN, J. J.,
GROENEN, P. J. F.,
VAN DER KOOIJ, A. J.: 2007 “Nonlinear Principal Components Analysis: Introduction And Application,” **Psychological Methods**, 12, 336–58.

- LIU, M.,
BOZDOGAN, H.: 2008a “Multivariate Regression Models with Power Exponential Random Errors and Subset Selection Using Genetic Algorithms With Information Complexity | Liu | European Journal of Pure and Applied Mathematics,” **European Journal Of Pure And Applied Mathematics**, 1, 4–37.
- MACNISH, R.: 1834 **The Philosophy of Sleep**, D. Appleton & Company.
- MEULMAN, J.: 1982 **Homogeneity Analysis Of Incomplete Data**, DSWO Press.
- MICHAİLIDIS, G.,
DE LEEUW, J.:1998 “The Gifi System of Descriptive Multivariate Analysis,” **Statistical Science**, 13, 307–36.
- MINERVA, T.,
PATERLINI, S.: 2002 “Evolutionary Approaches for Statistical Modelling,” **Proceedings of the 2002 Congress on Evolutionary Computation. CEC’02 Cat. No.02TH8600**
- NISHISATO, S.: 2014 **Elements of Dual Scaling: An Introduction To Practical Data Analysis**, Psychology Press.
- NUNES, M. L.,
BRUNI, O.,
NUNES, M. L.,
BRUNI, O.: 2015 “Insônia Na Infância E Adolescência: Aspectos Clínicos, Diagnóstico E Abordagem Terapêutica,” **Jornal de Pediatria**, 91, S26–35.
- OLULEYE, B., LEISA, A.,
LENG, J., AND DEAN,
D.: 2014 “A Genetic Algorithm-Based Feature Selection,” **Ecu Publications Post**, 5, 7.

- ORHUNBİLGE, N.: 2010 **Çok Değişkenli İstatistik Yöntemler**, İstanbul: İşletme Fakültesi Yayın.
- ÖZDEMİR, M.: 2017 “Predictive Model Selection in Linear Regression By Genetic Algorithms,” **Pamukkale University Journal of Social Sciences Institute**, 2017, 213–33.
- PAMUKÇU, E.: 2015 **Yüksek Boyutlu Kanser Sınıflama Probleminde Bilgi Karmaşıklığı Kriteri Ile Aykırı Gözlem Tespiti ve Boyut İndirgeme**, Elazığ: Fırat Üniversitesi / Fen Bilimleri Enstitüsü.
- PAMUKÇU, E.: 2017 “Aşırı Derecede Küçük Örneklem Problemi için Hibrit Regresyon Modeli,” **Celal Bayar Üniversitesi Fen Bilimleri Dergisi**, 13, 803–13.
- PATERLINI, S.,
MINERVA, T.: 2010 “Regression Model Selection Using Genetic Algorithms,” **World Scientific and Engineering Academy and Society = WSEAS**.
- PFANZAGL, J.: 1994 **Parametric Statistical Theory**, Walter de Gruyter.
- RUGGIERO, J. S.: 2003 “Correlates Of Fatigue In Critical Care Nurses,” **Research in Nursing & Health**, 26, 434–44.
- SAHİN, S.,
CİGİZOĞLU, H. K.: 2010 “Homogeneity Analysis of Turkish Meteorological Data Set,” *Hydrological Processes*, 24, 981–92.
- SCHWARZ, G.: 1978 “Estimating the Dimension of a Model,” *The Annals of Statistics*, 6, 461–64.
- SELVİ, F. F.,
KARAKAŞ, S. A.,
BOYSAN, M.,
SELVİ, Y.: 2015 “Effects Of Shift Work On Attention Deficit, Hyperactivity, And Impulsivity, And Their Relationship

- With Chronotype,” **Biological Rhythm Research**, 46, 53–61.
- ŞENEL, F.: 2015 “Uyku ve Rüya,” **Bilim ve Teknik Dergisi**, 2, 1–15.
- ŞENOĞLU, B.,
ÖZARICI Ö.,
FİDAN, H.: 2001 “En Küçük Kareler Kestiricilerinin Etkinliklerinin Normal Olmayan Bazı Dağılımlar İçin İncelenmesi,” **Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A - Uygulamalı Bilimler ve Mühendislik**, 2, 369–73.
- ŞENOL, V.,
SOYUER, F.,
AKÇA, R. P.,
ARGÜN., M.: 2012 “Adolesanlarda Uyku Kalitesi ve Etkileyen Faktörler,” **Kocatepe Tıp Dergisi**, 13, 93–104.
- SHANNON, C.,
WEAVER, W.: 1963 **The Mathematical Theory of Communication**, IL: University of Illinois Press.
- SHRESTHA, S.: 2009 “Categorical Regression Models with Optimal Scaling for Predicting Indoor Air Pollution Concentrations inside Kitchens in Nepalese Households,” **Nepal Journal of Science and Technology**, 10, 205–11.
- SUNER, A.,
ÇELİKOĞLU, C. C.: 2010 “Toplum Tabanlı Bir Çalışmada Çoklu Uygunluk Analizi ve Kümeleme Analizi ile Sağlık Kurumu Seçimi,” **Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, 25, 43–55.
- SUNER, A.,
ÇELİKOĞLU, C. C.: 2010 “Toplum Tabanlı Bir Çalışmada Çoklu Uygunluk Analizi ve Kümeleme Analizi ile Sağlık Kurumu Seçimi,” **Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, 25, 13.

- SÜT, N.: 2001 **Doğrusal Olmayan Kanonik Korelasyon Analizi Ve Bir Uygulama**, Edirne: Trakya Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı.
- TABACHNICK, B. G., FIDELL, L. S.: 2013 **Using Multivariate Statistics**, Seventh Edition. United States: Pearson Education.
- TARI, R.: 2016 **Ekonometri**, 12. Baskı. KÜV Yayınları.
- TAYLOR, A. S., HARPER, R.: 2003 “The Gift of the Gab?: A Design Oriented Sociology of Young People’s Use of Mobiles,” **Computer Supported Cooperative Work CSCW**, 12, 267–96.
- TENENHAUS, M., YOUNG, F. W.: 1985 “An Analysis And Synthesis Of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis And Other Methods For Quantifying Categorical Multivariate Data,” *Psychometrika*, 50, 91–119.
- TOMLINSON, M.: 2003 “Lifestyle and Social Class,” *European Sociological Review*, 19, 97–111.
- TÜRKAY, H.: 2004 “Doğrusal Regresyon Analizinde M Tahminciler Ve Ekonometrik Bir Uygulama,” **Doğu Anadolu Bölgesi Araştırmaları Dergisi**, 2, 106–15.
- UCAL, M. Ş.: 2006 “Ekonometrik Model Seçim Kriterleri Üzerine Kısa Bir İnceleme,” **C.Ü. İktisadi ve İdari Bilimler Dergisi**, 7, 41–57.
- VIJAYALAXMI, C., THOMAS, R., REITER, R. J., HERMAN, T. S.: 2002 “Melatonin: From Basic Research to Cancer Treatment Clinics,” **Journal of Clinical Oncology**, 20, 2575–2601.

- VURAL, M.: 2015 **Genetik Algoritma Yöntemi İle Toplu Üretim Planlama**, Fen Bilimleri Enstitüsü. İstanbul Teknik Üniversitesi.
- WANG, D., L.
SHI, YEUNG, D. S.,
TSANG, E. C. C.:2005 “Nonlinear Canonical Correlation Analysis of FMRI Signals Using HDR Models,” **2005 IEEE Engineering in Medicine and Biology 27th Annual Conference**.
- YALTA, A. T.:2018 **Ekonometri Ders Notları** ,
<http://yalta.etu.edu.tr/econometrics-lecture-notes.html>,
Accessed 03/15/2018.
- YAZICI, A. C.,
ÖĞÜŞ, E.,
ANKARALI, H.
GÜRBÜZ, F. :2010 “An Application Of Nonlinear Canonical Correlation Analysis On Medical Data,” **Turkish Journal Of Medical Sciences**, 40, 503–10.
- YOUNG, F. W.,
TAKANE, Y.,
DE LEEUW, J.: 1978 “The Principal Components Of Mixed Measurement Level Multivariate Data: An Alternating Least Squares Method With Optimal Scaling Features,” **Psychometrika**, 43, 279–81.

EKLER

Cep telefonu kullanımı

ve

Uyku ilişkisi anketi

Cinsiyetiniz? ☐ Kadın ☐ Erkek

Yaşınız?

Eğitim durumunuz?

- ☐ Lise altı
- ☐ Lise
- ☐ Ön lisans
- ☐ Lisans
- ☐ Lisansüstü

Medeni durumunuz? ☐ Evli ☐ Bekâr

Yatak odanızda hangisi veya hangileri bulunur?

- ☐ Televizyon
- ☐ İnternet
- ☐ Bilgisayar
- ☐ Cep telefonu
- ☐ Tablet
- ☐ Sabit telefon
- ☐ Hepsi
- ☐ Hiçbiri

Ailede (evde) cep telefonu, internet, sosyal medya kullanımını düzenleyen bir politika (kural) var mıdır?

☐ Evet ☐ Hayır

Yatak odasına çekilmeden önce evinizde internet erişimini sağlayan modemi (wifi) kapatır mısınız?

☐ Evet ☐ Hayır

Cep telefonu kullanımı, uyku düzeninizi olumsuz etkiler mi?

☐ Evet ☐ Hayır

Gece yatağında uykumun daha çabuk gelmesi için cep telefonu kullanırım.

☐ Hiçbir zaman ☐ Bazen ☐ Her zaman

Yatak odanız, geceleyin uyku saatlerinde sizin için cep telefonu kullanımına uygun özerk bir alan mıdır?

☐ Tamamen özerk bir alandır

☐ Kısmen özerk bir alandır

☐ Özerk bir alan değildir

Gece kaç saat uyursunuz?

☐ 5 saatten az

☐ 5 – 5,5 saat

☐ 5,5 – 6 saat

☐ 6 – 6,5 saat

☐ 6,5 – 7 saat

☐ 7 – 7,5 saat

☐ 7,5 – 8 saat

☐ 8,5 – 9 saat

☐ 9 saatten fazla

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı	Engin	Soyadı	KARAMAN
Doğ.Yeri	Ankara	Doğ.Tar.	03/02/1982
Uyruğu	TC	TC Kim No	17387775858
Email	enginkaraman1013@gmail.com	Tel	0541 450 0982

Eğitim Düzeyi

	Mezun Olduğu Kurumun Adı	Mez. Yılı
Doktora	İstanbul Üniversitesi- Sosyal Bilimleri Enstitüsü/ Sayısal Yöntemler ABD	Devam Ediyor
Yük.Lis.	Gebze Yüksek Teknoloji- Enstitüsü/İşletme Prog.	2015
Lisans	Marmara Üniversitesi- Fenbilgisi Öğretmenliği ABD	2004
Lise	Gazi Lisesi (Ankara)	1998

İş Deneyimi (Sondan geçmişe doğru sıralayın)

	Görevi	Kurum	Süre (Yıl - Yıl)
1.	Bilgisayar İşletmeni	Kocaeli Üniversitesi Hereke MYO-	2009-Devam
2.	Matematik Öğretmenliği	İletişim Açık Öğretim Dershanesi	2004-2009

Yabancı Dilleri	Okuduğunu Anlama*	Konuşma*	Yazma*	KPDS/ÜDS Puanı	(Diğer) Puanı
İngilizce	İyi	orta	iyi	80	

Bilgisayar Bilgisi

Program	Kullanma becerisi
MATLAB	İyi

SPSS	İyi
PYHTON	İyi
QİSKİT	ORTA

Yayınları/Tebliğleri Sertifikaları/Ödülleri

Akademik Yayınlar:

Demirkaya H., Akdemir A., Karaman E., Atan., (2015), “Kuşakların Yönetim Politikası Beklentilerinin Araştırılması”, İşletme Araştırmaları Dergisi, 7/1 186-204.

Kazan H., Karaman E., Akçalı Y.B., Şişmanoğlu E., (2015), “Assessment of TEOG (Transition From Basic To Secondary Education) Examination Success: Topsis Multi-Criteria Decision-Making Method Practice ” Science Direct Procedia - Social and Behavioral Sciences, 195 (2015) 915-924.

Karaman E., Kazan H., (2015), “Performance Evaluation In Family Physician: The Application Of Topsis Multi-Criteria Decision Making Method” Alphanumeric Journal, 3(2) (2015) 001–012.

Karaman E., Kazan H., (2017), “Okulların Teog Sınavı Başarı Sıralamaları: Topsis Çok Kriterli Karar Verme Yöntem Uygulaması”, Eurasian Business & Economics Journal, 8 (2017) 165–179.

Karaman E., Çılan Ç. A., (2018), “Uluslararası Matematik Ve Fen Eğilimleri Araştırma (Timss) Sınavının Optimal Ölçekleme Teknikleri İle Analizi”. TURAN-SAM Uluslararası Bilimsel Hakemli Dergisi; Sayı: 38 (10) , 122-127

Latif, H., Kandemir, A. Ş., Gökkaya, Ö., & Karaman, E. (2018). “Predictive Analysis Of Facebook Jealousy”. Karadeniz Uluslararası Bilimsel Dergi, (40), 172-184.

Karaman E., Çılan Ç. A., Fırat S.Ü.O. (2018). “TIMSS Sınavı Sonuçlarının Kategorik Regresyon Analizi Yöntemi İle Değerlendirilmesi”. 19. Uluslararası EYİ Sempozyumu Tam Metin Bildiri Kitabı.1342-1350

Karaman E., Kazan H., (2018), “Çalışanların Sosyal Hayat Tutumlarının, Demografik Özelliklerine Göre Optimal Ölçekleme Teknikleri İle İncelenmesi”, Kafkas Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 9(18), 379-389.

Kandemir, A., Kazan, H., & Karaman, E. (2019). “İmalat Sektöründe Mesleki Eğitim Veren Girişimlere Göre Hedeflenen Beceri Türü Seçimi Üzerine Bir Uygulama: Analitik Hiyerarşi Prosesi.” TURAN-SAM Uluslararası Bilimsel Hakemli Dergisi, 11(41), 550-559.

Bildiri:

Karaman E., Kazan H., “Çalışanların Sosyal Hayat Tutumlarının, Demografik Özelliklerine Göre Optimal Ölçekleme Teknikleri ile İncelenmesi”. XVIII. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu 2017.

Karaman E., Çılan Ç. A., “(TIMSS) Sınavının Optimal Ölçekleme Teknikleri ile Analizi”. II. Uluslararası Rating Academy Congress On HOPE Çanakkale-2018.

Karaman E., Çılan Ç. A., Fırat S.Ü.O. “TIMSS Sınavı Sonuçlarının Kategorik Regresyon Analizi Yöntemi İle Değerlendirilmesi”. 19. Uluslararası EYİ Sempozyumu . 2018

Latif, H., Gökkaya, Ö., & Karaman, E..” Akıllı telefon ve sosyal medyada saygı dışı davranışların incelenmesi”. Uluslararası Marmara Fen Ve Sosyal Bilimler Kongresi. 2019