

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

**ÇOKLU DOĞRUSAL REGRESYON MODELİ İÇİN BULANIK ÜYELİK
FONKSİYONUNA DAYALI PARAMETRE TAHMİNİ**

YÜKSEK LİSANS TEZİ

Tuğba KARAN

HAZİRAN 2014

TRABZON

KARADENİZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

**ÇOKLU DOĞRUSAL REGRESYON MODELİ İÇİN BULANIK ÜYELİK
FONKSİYONUNA DAYALI PARAMETRE TAHMİNİ**

Tuğba KARAN

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
"YÜKSEK LİSANS (İSTATİSTİK)"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

Tezin Enstitüye Verildiği Tarih : 23.05.2014

Tezin Savunma Tarihi : 16.06.2014

Tez Danışmanı: Doç. Dr. Türkan ERBAY DALKILIÇ

Trabzon 2014

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalında
Tuğba KARAN tarafından hazırlanan

ÇOKLU DOĞRUSAL REGRESYON MODELİ İÇİN BULANIK ÜYELİK
FONKSİYONUNA DAYALI PARAMETRE TAHMİNİ

başlıklı bu çalışma, Enstitü Yönetim Kurulunun 27 / 05 / 2014 gün ve 1555 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda

YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.

Jüri Üyeleri:

Başkan : Doç. Dr. Tülay KESEMEN

Üye : Doç. Dr. Zafer KÜÇÜK

Üye :Doç. Dr. Türkan ERBAY DALKILIÇ


.....

.....

.....

Prof. Dr. Sadettin KORKMAZ

Enstitü Müdürü

ÖNSÖZ

Bu tez çalışmasında doğrusal regresyon analizi, bulanık mantık ile ilgili temel kavramlar, üyelik fonksiyonu belirlenmesi ve üyelik fonksiyonuna dayalı parametre tahmini üzerinde durulmuştur. Doğrusal regresyon analizinde bağımsız değişkenlerden herhangi birinin dağılımının normal dağılım dışına çıkması ve gözlemlerde aykırı değerlerin varlığı durumunda oluşacak çoklu doğrusal regresyona ilişkin varsayım bozulmalarından daha az etkilenecek bir parametre tahmini yöntemi önerilmiştir. Bu yöntemde, bağımsız değişkenlerden herhangi birinin Pareto dağılımına uygunluğu söz konusu olduğunda, bu değişkenin modele katkı miktarı Pareto dağılımına ilişkin elde edilen üyelik fonksiyonu ile belirlenmiştir. Var olan aykırı gözlemlerin hata toplamına etkisi üyelik fonksiyonundan elde edilen üyelik dereceleri ile ilişkilendirilip azaltılmaya çalışılmıştır. Önerilen yöntem ile elde edilen tahminlere ilişkin hata miktarları, En Küçük Kareler (EKK) yönteminden elde edilen parametrelere bağlı tahminlerin hata miktarları ile karşılaştırılmıştır.

Bu tez çalışmasının planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen, bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren sayın hocam Doç. Dr. Türkan ERBAY DALKILIÇ'a,

Yüksek lisans eğitimi yapma fırsatı bulduğum İstatistik ve Bilgisayar Bilimleri Bölümündeki tüm hocalarıma,

Tüm yaşamım boyunca bana her zaman güvenen, destekleyen, teşvik eden ve yalnız bırakmayan sevgili aileme,

Çalışmalarım sırasında yardımlarını esirgemeyen ve bana destek olan arkadaşlarım Çiğdem GÜNGÖR ve Ersegül ÖZTÜRK ŞAHİN'e,

Desteği ile her zaman yanımda olan, görüş ve önerilerinden her zaman yararlandığım Bilgehan KAYIŞ'a,

sonsuz teşekkürler.

Tuğba KARAN
Trabzon 2014

TEZ BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Çoklu Doğrusal Regresyon Modeli İçin Bulanık Üyelik Fonksiyonuna Dayalı Parametre Tahmini” başlıklı bu çalışmayı baştan sona kadar danışmanım Doç. Dr. Türkan ERBAY DALKILIÇ’ın sorumluluğunda tamamladığımı, verileri kendim topladığımı, analizleri ilgili laboratuarlarda yaptığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 16/06/2014

Tuğba KARAN



İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ BEYANNAMESİ	IV
İÇİNDEKİLER	V
ÖZET	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ	IX
TABLolar DİZİNİ	X
SEMBOL VE KISALTMALAR DİZİNİ.....	XI
1. GENEL BİLGİLER	1
1.1. Giriş	1
1.2. Önceki Çalışmalar	2
1.3. Regresyon Çözümlemesi	4
1.3.1 Basit Doğrusal Regresyon Modeli.....	5
1.3.1.1. Parametre Tahmini İçin En Küçük Kareler Yöntemi ve Varsayımları	5
1.3.2. Çoklu Doğrusal Regresyon Modeli	8
1.4. Bulanık Mantık Kavramı	11
1.5. Bulanık Sayı Kavramı	11
1.6. Bulanık Küme Kavramı.....	14
1.7. Bulanık Küme Teorisinde Temel Tanımlar.....	15
1.8. Bulanık Mantıkta Üyelik Fonksiyonu	17
1.9. Pareto Dağılımı.....	18
1.10. Uyum İyiği Testleri	21
1.10.1. Kolmogorov-Simirnov Uyum İyiği Testi.....	21
1.11. Aykırı Değerlerin Belirlenmesi	23
2. YAPILAN ÇALIŞMALAR.....	25
2.1. Pareto Dağılımı İçin Optimal Üyelik Fonksiyonunun Belirlenmesi	25
2.2. Parametre Tahmini İçin Bulanık Üyelik Derecesine Dayalı Algoritma.....	28

2.3.	Uygulama	29
3.	SONUÇLAR VE ÖNERİLER	46
4.	KAYNAKLAR.....	47
ÖZGEÇMİŞ		

Yüksek Lisans Tezi

ÖZET

ÇOKLU DOĞRUSAL REGRESYON MODELİ İÇİN BULANIK ÜYELİK
FONKSİYONUNA DAYALI PARAMETRE TAHMİNİ

Tuğba KARAN

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalı
Danışman: Doç. Dr. Türkan ERBAY DALKILIÇ
2014, 48 Sayfa

Basit ve çoklu doğrusal regresyon çözümlemesinde klasik metotlardan yararlanılıyorsa, modele ilişkin parametrelerin tahmini bazı varsayımlara dayanır. Doğrusal regresyon modellerinin bilinmeyen parametrelerinin tahmininde verilerin normal dağılım dışında bir dağılıma sahip olması durumu tahmin sürecinde klasik çözümlemelerin dışına çıkılmasını gerektirir. Ayrıca veri setinde aykırı gözlemlerin varlığı da kullanılacak yöntemin önemini artırmaktadır. Böyle durumlarda bulanık mantığa dayalı çözümleme yöntemleri alternatif yöntemler olarak kendini göstermektedir. Bu çalışmada, bağımsız değişkenlerden herhangi birinin Pareto dağılımına sahip olması ve veri setinde aykırı gözlemlerin mevcut olması durumunda, kurulacak çoklu doğrusal regresyon modelinin bilinmeyen parametrelerini tahmin etmek için bir algoritma önerilerek, algoritmadan elde edilen tahminler literatürde yer alan mevcut yöntemlerden elde edilen tahminler ile karşılaştırılmıştır.

Anahtar Kelimeler: Parametre Tahmini, Bulanık Üyelik Fonksiyonu, Pareto Dağılımı.

Master Thesis

SUMMARY

PARAMETERS ESTIMATION FOR MULTIPLE LINEAR REGRESSION MODEL
BASED ON FUZZY MEMBERSHIP FUNCTIONS

Tuğba KARAN

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Statistic and Computer Science Graduate Program
Supervisor: Assoc. Prof. Türkan ERBAY DALKILIÇ
2014, 48 Pages

In simple and multiple linear regression analysis, if conventional methods are using, parameter estimation for linear regression model is based on some assumptions. In the process of the estimation unknown parameters for linear regression models, if the data have different structure from normal distribution, it is requires to go beyond the classical resolution in the estimation process. In addition, the existence of contradictory observations in the data set in crease the importance of the method used in the process. In such cases, analysis methods based on fuzzy logic are known as alternative methods. In this study, in the case of any of independent variables have Pareto distribution and there were outlier observations in data set an algorithm has been suggested to define the unknown parameters of multiple linear regression model. The estimates obtained from this algorithm are compared with estimates obtained from existing methods.

Key Words: Parameter Estimation, Fuzzy Membership Function, Pareto Distribution

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Üçgensel bulanık sayı	12
Şekil 2. Yamuksal bulanık sayı	13
Şekil 3. L – R bulanık sayı	14
Şekil 4. Pareto dağılımının olasılık yoğunluk fonksiyonu	19
Şekil 5. X_1 değişkenine ilişkin histogram	30
Şekil 6. X_2 değişkenine ilişkin histogram	30
Şekil 7. X_3 değişkenine ilişkin histogram	30
Şekil 8. X_3 değişkeninin Pareto dağılımına uygunluğu	31
Şekil 9. Uygun c 'nin belirlenmesi	34
Şekil 10. Tablo 2’de yer alan veri setine ilişkin hata grafikleri	36
Şekil 11. X_2 değişkenine ilişkin histogram	38
Şekil 12. X_2 değişkeninin Pareto dağılımına uygunluğu	38
Şekil 13. Tablo 4’te yer alan veri setine ilişkin hata grafikleri	40
Şekil 14. X_3 değişkenine ilişkin histogram	42
Şekil 15. X_3 değişkeninin Pareto dağılımına uygunluğu	43
Şekil 16. Tablo 5’te yer alan veri setine ilişkin hata grafikleri	45

TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Pareto dağılımına uyum	31
Tablo 2. Veri setine ilişkin tahminler ve hatalar	32
Tablo 3. Pareto dağılımına uyum	37
Tablo 4. Veri setine ilişkin tahminler ve hatalar	37
Tablo 5. Veri setine ilişkin tahminler ve hatalar	41
Tablo 6. Pareto dağılımına uyum	43

KISALTMALAR VE SEMBOLLER DİZİNİ

EKK	: En Küçük Kareler
L-R	: Left-Right
x_m	: X rastgele değişkeninin aldığı en küçük değer
HKT	: Hata Kareler Toplamı
K-S	: Kolmogorov-Simirnov
max	: Maksimum
min	: Minimum
n	: Toplam nesne sayısı
\hat{Y}_{EKK}	: En Küçük Kareler yöntemi kullanılarak Y 'nin tahmin edicisi
$\hat{Y}_{\hat{O}A}$: Önerilen algoritma kullanılarak Y 'nin tahmin edicisi
\hat{Y}_{BER}	: Bertoni-Walfisch algoritması kullanılarak Y 'nin tahmin edicisi
$\mu_{\tilde{A}}(x)$: \tilde{A} bulanık kümesindeki x verisinin üyelik derecesi
\hat{Y}	: Y 'nin tahmin edicisi
ε_{EKK}	: En Küçük Kareler yöntemi kullanılarak hesaplanan hata terimi
$\varepsilon_{\hat{O}A}$: Önerilen algoritma kullanılarak hesaplanan hata terimi
ε_{BER}	: Bertoni-Walfisch algoritması kullanılarak hesaplanan hata terimi
\tilde{A}	: A bulanık kümesi
A_α	: A bulanık kümesinin α -kesme kümesi
<	: Küçük
=	: Eşit
>	:Büyük
\in	: Eleman
Σ	: Toplam
\cap	: Kesişim
\cup	: Birleşim
ε	: Hata terimi ve iterasyon bitirme kriteri
e_i	: Artık
p_{ii}	:Gözlem uzaklığı
r_i	:Student türü artık

1. GENEL BİLGİLER

1.1. Giriş

Regresyon çözümlemesinin genel amacı; bağımlı (açıklanan) değişken ile bağımsız (açıklayan) değişkenler arasındaki ilişkinin matematiksel modelini kurmak ve bu model aracılığı ile bağımsız değişkenin alacağı değerleri tahmin etmek şeklinde özetlenebilir. Regresyon modelinin bilinmeyen parametrelerinin tahmininde, ele alınan veri setinin yapısına uygun modelin elde edilebilmesi için geliştirilen birçok yöntem literatürde yer almaktadır.

Regresyon çözümlemesi yöntemlerinin her biri için veri analizi oldukça önemlidir. Çünkü veri setinde yer alan değişkenlerin hangi dağılımdan geldikleri regresyon modelindeki parametre tahminleri için belirleyici rol oynamaktadır. Ayrıca veri setinde aykırı gözlemlerin varlığı da kullanılacak yöntemin önemini artırmaktadır. Çünkü aykırı gözlemler yapılacak parametre tahminlerini ana veri setinden uzaklaştıracak önemde olabilir. Basit ve çoklu doğrusal regresyon modellerinin bu varsayım bozulmalarından etkilenme düzeyini en aza indirmek üzere literatürde pek çok güçlü (robust), model tahmin yöntemleri mevcuttur.

Bu çalışmada, bağımsız değişkenlerden herhangi birinin Pareto dağılımına sahip olması ve veri setinde aykırı gözlemlerin mevcut olması durumunda, kurulacak çoklu doğrusal regresyon modelinin bilinmeyen parametrelerini tahmin etmede normal dağılıma uymama ve aykırı değerlerin varlığı sorunlarından mümkün olduğu kadar az etkilenen bir tahmin yöntemi önerilmiştir. Bağımsız değişkenlerden Pareto dağılımına sahip olan değişkenin modele katkısı, Pareto dağılımına uygun elde edilen üyelik fonksiyonu ile derecelendirilmiş, böylece tahmin aşamasında aykırı gözlemlerin etkisi de en aza indirgenmeye çalışılmıştır. Bu anlamda kullanılan yöntem güçlü (robust) yöntemler sınıfında değerlendirilebilir.

Çalışmanın uygulama kısmında önerilen yöntemin geçerliliğinin irdelenebilmesi için sayısal örneklere yer verilmiş ve ele alınan veri setleri için elde edilen sonuçlar literatürde var olan klasik yöntemlerden elde edilen sonuçlar ile karşılaştırılmıştır.

1.2. Önceki Çalışmalar

Zadeh, L. A. (1965) bulanık yaklaşım konusundaki ilk ciddi adımı 1965 yılında yayınlanan “Bulanık Kümeler” adlı makale ile atmıştır. Zadeh bulanık mantığın matematik ve bilgisayar bilimleri alanlarındaki uyarlanabilirliği üzerinde durmuştur.

Takagi ve Sugeno (1985) bulanık ifadeler ve ilişkilerin kullanıldığı sistemlerin bulanık modellerinin oluşturulmasında kullanılmak üzere matematiksel gereçler önerdikleri çalışmalarında bulanık sistemlerin modellenmesi ve kontrolü için uygulamalar verilmiştir.

Civanlar ve Trussel (1986) “İstatistiksel veriler kullanılarak üyelik fonksiyonlarının oluşturulması” başlıklı çalışmalarında üyelik fonksiyonunun belirlenmesinin, bulanık küme teorisinin pratik uygulamalarında önemli olduğu belirtilmiş ve elemanları bilinen bir olasılık yoğunluk fonksiyonu ile belirleyici niteliklere sahip, bulanık kümelerle dair üyelik fonksiyonunun belirlenmesi için bir yöntem sunmuşlardır. Çalışmada üyelik fonksiyonunun sağlaması gereken koşullar da verilmiştir.

Dombi (1990) üyelik fonksiyonları üzerine yaptığı çalışmada kullanılan farklı üyelik fonksiyonlarını tanımlamış, üyelik fonksiyonlarının kurulması için gerekli özellikler ve üyelik fonksiyonlarının matematiksel formları ile ilgili bilgi vermiştir.

Ishibuchi ve Tanaka (1992) çalışmalarında sinir ağları kullanarak bulanık regresyon analizi için basit fakat güçlü metotlar önermişlerdir. Sinir ağlarının yüksek kapasitelerinden dolayı, önerilen yöntemlerin, karmaşık sistemler için var olan temel doğrusal programlama yöntemlerinden daha uygulanabilir olduğunu ileri sürmüşlerdir.

Hathaway ve Bezdek (1993) switching regresyon modelleri ve bulanık kümeleme adı altında yaptıkları çalışmalarında karma dağılımlara sahip verilerin switching regresyon modellerinin oluşturulmasında kullanılabilen bulanık c-regresyon modellerinden söz etmişler. Doğrusal ya da doğrusal olmayan modeller için iki sayısal örnek ile yeni bir yaklaşım önermişlerdir.

Jang ve Sun (1995) “Neuro-Fuzzy modelleme ve kontrol” başlıklı çalışmalarında, bulanık modelleme ve kontrol için temel ve ileri çalışmalar gözden geçirilmiş, bulanık kümeler, bulanık kurallar, bulanık muhakeme ve bulanık modeller tanımlanmıştır. Uyarlamalı ağ yapısının, sinir ağlarını ve bulanık modelleri bir araya getiren bir yapı olduğunun da belirtildiği çalışmada bulanık çıkarsama sistemine dayalı uyarlamalı ağlar için modelleme yöntemleri önerilmiştir.

Chi-Bin ve Lee (1998) bulanık regresyon analizi için bulanık uyarlamalı ağlardan faydalanmışlar ve yaptıkları çalışmada bulanık uyarlamalı ağın eğitiminde, tahmin edilen çıktılar ile beklenen çıktılar arasındaki fark olarak tanımlanan hata ölçüsünden faydalanmışlardır. Çıktıların bulanık olması durumunda hata ölçüsünün hesaplanma yöntemi de aynı çalışmada verilmiştir.

Chen ve Wang (1999) bulanık üyelik fonksiyonlarının en iyilenmesi için bulanık kümeleme analizi isimli çalışmalarında bulanık modelleme tanımını vermişlerdir. Bulanık modelleme için en önemli kriterlerden birinin bulanık sistemlerde kullanılan üyelik fonksiyonlarının parametrelerinin belirlenmesi olduğunu belirtmişlerdir. Bu parametrelerin belirlenmesinde bulanık kümelemeden faydalandıkları çalışmada bir değişkene ait optimal küme sayısının belirlenmesi için önceden önerilmiş Xie-Beni (1991) indeksini kullanmışlardır.

D’Urso (2003) “Bulanık / kesin girdi ve bulanık / kesin çıktı verileri için doğrusal regresyon analizi” başlıklı çalışmasında bulanık regresyon modelini tahmin etmek için, EKK yöntemini göz önüne alarak kesin ve bulanık girdi-çıkıtı verilerine ilişkin kesin ve bulanık regresyon modelleri önermiştir.

Erbay Dalkılıç (2005) bağımsız değişkenlerin üstel dağılımdan gelmesi durumunda switching regresyonda bulanık sinir ağları yaklaşımı ile parametre tahmini yapmış ve mevcut tahmin yöntemlerinden elde edilen sonuçlar ile karşılaştırmıştır.

Modarres, Nasrabadi ve Nasrabadi (2005) “En küçük kareler hataları ile bulanık lineer regresyon modelleri” başlıklı çalışmalarında matematiksel programlama modeli geliştirilmiştir. Gözlenen ve tahmin edilen değerlerin arasındaki farkı en aza indirmek için bir deyişle modele ilişkin hataları minimize etmek için bir model geliştirilmiştir.

Hojati, Bector ve Smimou (2005) “ Bulanık Doğrusal Regresyon Hesaplanması için Basit Bir Yöntem” adlı çalışmalarında iki olgu göz önünde bulundurulmuştur. Birincisi, bağımlı değişken bulanık olduğunda önerdikleri yaklaşımı literatürde önerilen ile karşılaştırılmıştır. İkincisi ise, hem bağımlı hem bağımsız değişken bulanık olduğunda bir yaklaşım önermişler mevcut yöntemler ile karşılaştırmalar yapmışlardır.

Bargiela, Pedrycz ve Nakashima (2007) “Bulanık verilerle çoklu regresyon” adlı çalışmalarında bulanık değişkenlere dair yinelemeli bir algoritma önerilmiş ve Bulanık C-Ortalamalar (Fuzzy C-MeansFCM) yöntemine dayalı yöntemden elde edilen sonuçlar En Küçük Kareler yönteminden elde edilen sonuçlar ile kıyaslanmıştır.

Chuang, Jeng ve Tao (2009) “Aykırı değerlerle Takagi-Sugeno-Kang TSK bulanık modelleme için güçlü bir hibrit yaklaşım” başlıklı çalışmalarında aykırı değerlerle Takagi-Sugeno-Kang TSK bulanık modelleri oluşturmak için güçlü (robust) bir hibrit yaklaşım önerilmiştir.

Mosleh, Otadi ve Abbasbady (2010) “Bulanık sinir ağı ile bulanık doğrusal regresyon modellerinin değerlendirilmesi” başlıklı çalışmalarında kesin girdi ve bulanık çıktılarla doğrusal ve doğrusal olmayan bulanık regresyon modelin yaklaşık bulanık katsayıları için bulanık sinir ağları tabanlı yeni bir hibrit yöntem sunulmuştur.

Pierpaolo, Riccardo ve Adriana, (2011) “Güçlü bulanık regresyon analizi” başlıklı çalışmalarında en küçük medyan karelere dayalı güçlü bulanık doğrusal regresyon modelini tahmin etme süreci önermişlerdir.

Chen, Hsueh ve Chang (2013) “Bulanık Regresyon Modelleri Oluşturulması İçin İki Aşamalı Bir Yaklaşım” başlıklı çalışmalarında tek olumlu katsayı odaklı mevcut yaklaşımlardan farklı olarak, bulanık regresyon modelinde negatif katsayılarla dikkate alınmıştır. Çok sayıda bulanık gözlemlerle gerçek veri seti için negatif katsayılarla iki aşamalı yaklaşımın uygulanabilirliğini göstermişlerdir.

1.3. Regresyon Çözümlemesi

Değişkenler arasındaki ilişkiyi incelemek bilimin önemli uğraşlarından birisidir. Çünkü gerek günlük hayatımızda gerekse bilimsel araştırmalarda karşılaştığımız sorunların çoğunluğu iki veya daha fazla değişken arasında bir ilişki olup olmadığının saptanması ile ilgilidir. İki değişken arasında ki ilişkinin saptanması da istatistiksel çözümlemelerde sık sık karşılaşılan bir sorundur. Değişkenler arasındaki ilişkinin incelenmesinde regresyon ilk akla gelen tekniktir.

Regresyon, bir şeyi başka bir şeye bağlama işi ve biçimidir. Bilimsel olarak regresyon terimi, bir değişken ile başka bir ya da birden çok değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır.

İstatistiksel anlamda iki değişken arasındaki ilişki, değişkenlerin değerlerinin karşılıklı değişimleri arasında bir bağımlılık şeklinde anlaşılır. Örneğin; pancar üretimi arttığında fiyat düşüyorsa veya azaldığında fiyatı yükseliyorsa bu iki değişken arasında ilişki olduğunu gösterir. Regresyon analizi, biri bağımlı değişken diğeri bağımsız değişken

olmak üzere, en az iki değişken arasındaki ortalama ilişkinin matematiksel bir fonksiyon şeklinde yazılmasıdır. Bu fonksiyona regresyon denklemi adı verilir (Şıklar, 2000).

Regresyon denklemi yardımıyla bağımsız değişkenlerin çeşitli değerlerine karşılık bağımlı değişkenin alacağı değer tahmin edilir. Bağımlı değişkeni etkileyen bağımsız değişkenlerin saptanmış olması da bağımlı değişken üzerinde geliştirilecek politikalarda hangi değişkenlerin önem kazandığını ortaya çıkarmaktadır (Apaydın, 1994; Şıklar, 2000).

1.3.1 Basit Doğrusal Regresyon Modeli

Bağımsız değişkenin, herhangi bir bağımsız değişkene bağlı olarak nasıl değiştiğini gösteren basit doğrusal regresyon denklemi genel anlamda;

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

biçiminde ifade edilir. Burada; Y , bağımlı (açıklanan) değişkeni; X , bağımsız (açıklayan) değişkeni; β_0 , denklemin sabit terimini; β_1 , bağımsız değişken katsayısını; ε , hata terimini temsil etmektedir. Regresyon denklemindeki bilinmeyen β_0 ve β_1 aynı zamanda, regresyon katsayıları olarak adlandırılır (Köseoğlu ve Yamak, 2004).

1.3.1.1. Parametre Tahmini İçin En Küçük Kareler Yöntemi ve Varsayımları

Regresyon çözümlemesinin önemli bir amacı bilinmeyen parametrelerin kestirilmesidir. Parametrelerin kestiricilerinin bulunmasında kullanılan yöntemlerden biri En Küçük Kareler (EKK) yöntemidir. Basit doğrusal regresyon denkleminin EKK yöntemiyle tahmin edilebilmesi için modelce sağlanması gereken varsayımlar aşağıdaki gibi verilebilir;

1. Bağımlı değişken Y ile bağımsız değişken X arasındaki ilişkiyi ifade eden model doğrusaldır.
2. Modelin hata terimleri sıfır ortalamalı rastgele bir değişkendir: Regresyon hata terimleri bağımsız değişken X 'in aldığı değerlere göre pozitif, negatif ve sıfır değerlerini alabilmektedir. Hata terimlerinin toplamı ve dolayısıyla aritmetik ortalamasının sıfır olması beklenmektedir $E(\varepsilon) = 0$.

3. Bağımsız değişken X ile hata terimleri arasında bir ilişki yoktur. İki değişken arasındaki kovaryans sifıra eşittir $Kov(X_i \varepsilon_i) = 0$.
4. Regresyon hata terimleri normal dağılım özelliğine sahiptir $\varepsilon_i \sim N(0, \sigma^2)$.
5. Regresyon hata terimlerinin varyansı sabittir $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_\varepsilon^2$.
6. Hata terimleri arasında ardışık bağımlılık yoktur $Kov(\varepsilon_t, \varepsilon_{t-i}) = E(\varepsilon_t, \varepsilon_{t-i}) = 0$.

Bu varsayımların sağlandığı durumda (1) eşitliğinin en küçük kareler yöntemi ile elde edilen kestirimi;

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad i = 1, \dots, n \quad (2)$$

biçiminde ifade edilir. Bu doğru denklemde $\hat{\beta}_0$ ve $\hat{\beta}_1$, sırasıyla; gerçek kesim noktası olan β_0 ve gerçek regresyon katsayısı olan β_1 parametrelerinin en küçük kareler kestiricileridir. \hat{Y}_i ise; kestirim değeri ya da kestirilen değerdir. \hat{Y}_i kestirimlerinin Y_i gözlenen değerlerini ne kadar iyi kestirebildiğinin ölçüsü, Y_i gözlenen değeri ile \hat{Y}_i kestirim değeri arasındaki farktır. Bu fark, i . gözlem için;

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \quad (3)$$

biçiminde verilir. Burada e_i artık olarak adlandırılır. Gerçek değeri bilemediğimiz için e_i , ε_i 'nin kestirimi olarak düşünülür (Alpar, 2003).

Doğrusal EKK yönteminde esas kriter, regresyon hata terimleri karelerinin toplamını en küçük yapan $\hat{\beta}_0$ ile $\hat{\beta}_1$ değerlerini bulmaktır. O halde;

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \quad (4)$$

ifadesini en küçük yapacak $\hat{\beta}_0$ ve $\hat{\beta}_1$ değerini bulmak gerekecektir. Bu amaçla ifadenin önce $\hat{\beta}_0$ 'a, daha sonra $\hat{\beta}_1$ 'e göre birinci kısmi türevleri alınarak;

$$\frac{\partial S}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i$$

biçiminde elde edilen denklemler sıfıra eşitlenerek, gerekli işlemler yapıldığında;

$$\sum_{i=1}^n Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i \quad (5)$$

$$\sum_{i=1}^n X_i Y_i = \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 \quad (6)$$

biçiminde ifade edilen denklemlere ulaşılır. Bu denklemler normal denklemler olarak adlandırılırlar. Bu iki denklemin eş anlı çözümünden,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (7)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (8)$$

ile ifade edilen EKK tahmincileri elde edilir (Köseoğlu ve Yamak, 2004; Ünver ve Gangam, 1999).

Sadece bir bağımsız değişkenin ve n adet gözlemin yer aldığı (2) eşitliğindeki örnek regresyon denkleminin matris notasyonlarıyla gösterimi;

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}_{n \times 2}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}_{2 \times 1}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_{n \times 1}$$

biçimindedir. Örnek regresyon denklemi tüm gözlemleri kapsayacak şekilde matris notasyonu;

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \quad (9)$$

biçiminde ifade edilir.

Bu genel ifade de;

Y ; $n \times 1$ boyutundaki bağımlı değişken vektörünü,
 X ; $n \times 2$ boyutundaki bağımsız değişken matrisini,
 $\hat{\beta}$; 2×1 boyutundaki katsayılar vektörünü,
 e ; $n \times 1$ boyutundaki hata terimleri vektörünü tanımlamaktadır.

Doğrusal EKK yönteminde amaç, hata terimleri kareleri toplamını ($e'e$) en aza indirecek örnek regresyon katsayılarını bulmak olduğundan, (9) eşitliğindeki örnek regresyon denkleminin tahmin edilmesi,

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (10)$$

biçiminde verilen sistemin çözülmesi ile elde edilir.

Basit doğrusal regresyon denkleminin EKK çözümünde iki değer hesaplanmaktadır. Bilinmeyenlerden birisi sabit terimi temsil eden $\hat{\beta}_0$ 'dir. Bunun anlamı, denklemin sağ tarafındaki bağımsız değişken X 'in sıfır değer alması durumunda, bağımlı değişken Y 'nin alacağı değerdir. Bu değer tüm gözlemler için aynı olacaktır. Bilinmeyenlerin ikincisi ise bağımsız değişkenin katsayısı $\hat{\beta}_1$ 'dir. Bu katsayı pozitif değerli ise, bağımsız değişkenin bağımlı değişkeni aynı yönde etkileyeceği; negatif ise ters yönde etkileyeceği sonucuna varılır. Ne kadar etkileyeceği sorusuna ise katsayının büyüklüğü ile cevap verilir. Değişkenlerin doğrusal olarak kullanıldığı modellerde değişken katsayıları marjinal etki olarak yorumlanır. Çünkü basit doğrusal regresyon denkleminde Y 'nin X 'e göre birinci derece türevi, değişken katsayısına eşit olacaktır. Bağımsız değişkendeki bir birimlik değişme sonucunda bağımlı değişken Y , $\hat{\beta}_1$ birim kadar değişecektir (Köseoğlu ve Yamak, 2004).

1.3.2. Çoklu Doğrusal Regresyon Modeli

Regresyon çözümlemesinde birden fazla bağımsız değişkenin olduğu modellere çoklu regresyon çözümlemesi denilmektedir. k sayıda bağımsız değişkenin olduğu doğrusal regresyon modeli;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (11)$$

biçimde ifade edilebilir. Bu modeldeki Y, X_1, X_2, \dots, X_k gözlenebilen değerler, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ bilinmeyen parametrelerdir (Ünver ve Gamgam, 1999).

Çoklu doğrusal regresyonda, bağımlı değişkeni etkileyen birden çok bağımsız değişken söz konusudur. Çoklu doğrusal regresyonda araştırmacıların

1. Bağımlı değişkeni etkilediği belirlenen değişkenler yardımıyla bağımlı değişken değerini kestirmek,
2. Bağımlı değişkeni etkilediği düşünülen bağımsız değişkenlerden hangisi ya da hangilerinin bağımlı değişkeni daha çok etkilediğini bulmak ve aralarındaki karmaşık yapıyı tanımlamak

ile verilen iki genel amacı vardır. Araştırmanın amacı, bu iki amaçtan her ikisine ya da herhangi birine ulaşmak olabilir (Alpar, 2003).

Çoklu doğrusal regresyon denkleminin sol tarafında Y bulunurken, sağ tarafında k adet X değişkeni bulunmaktadır. Bağımlı değişken Y 'nin, k adet X değişkeni tarafından açıklanması durumunda n tane denklem,

$$\begin{aligned} Y_1 &= \hat{\beta}_0 + \hat{\beta}_1 X_{1,1} + \hat{\beta}_2 X_{2,1} + \cdots + \hat{\beta}_k X_{k,1} + e_1 \\ Y_2 &= \hat{\beta}_0 + \hat{\beta}_1 X_{1,2} + \hat{\beta}_2 X_{2,2} + \cdots + \hat{\beta}_k X_{k,2} + e_2 \\ &\vdots \\ Y_n &= \hat{\beta}_0 + \hat{\beta}_1 X_{1,n} + \hat{\beta}_2 X_{2,n} + \cdots + \hat{\beta}_k X_{k,n} + e_n \end{aligned}$$

biçimindedir ve matrisler ile;

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} \cdots X_{k,1} \\ 1 & X_{1,2} & X_{2,2} \cdots X_{k,2} \\ \vdots & \vdots & \vdots \cdots \vdots \\ 1 & X_{1,n} & X_{2,n} \cdots X_{k,n} \end{bmatrix}_{k \times n}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{k \times 1}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_{n \times 1}$$

olarak ifade edilirler (Köseoğlu ve Yamak, 2004).

Basit doğrusal regresyon modelinde olduğu gibi burada da amaç, hata terimleri kareleri toplamını en aza indirecek $\hat{\beta}$ vektörünü bulmaktır. EKK yöntemi matris formunda ifade edilmek istendiğinde;

$$\begin{aligned}\sum_{i=1}^n e_i^2 &= e'e = (Y - \hat{Y})'(Y - \hat{Y}) \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}\end{aligned}$$

elde edilir. Burada $Y'X\hat{\beta} = \hat{\beta}'X'Y$ olduğundan;

$$\sum_{i=1}^n e_i^2 = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

eşitliği ile verilen hata kareler toplamı (HKT) $\hat{\beta}$ 'ya göre türev alınırsa;

$$\frac{\partial HKT}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta}$$

olacaktır. Burada; $-2X'Y + 2X'X\hat{\beta} = 0$ olarak alındığında,

$$X'X\hat{\beta} = X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

biçiminde elde edilir, burada; $(X'X)'$; $(k + 1) \times (k + 1)$ boyutlu $(X'Y)$ ise; $(k + 1) \times 1$ boyutlu matrislerdir (Şıklar, 2000).

1.4. Bulanık Mantık Kavramı

Bulanık Mantık yaklaşımı ilk defa Amerika Birleşik Devletleri'nde düzenlenen bir konferansta 1956 yılında duyurulmuştur. Ancak bu konudaki ilk ciddi adım 1965 yılında Lotfi A. Zadeh tarafından yayınlanan bir makalede bulanık mantık veya bulanık küme kuramı adı altında ortaya konulmuştur. Zadeh bu çalışmasında insan düşüncesinin büyük çoğunluğunun bulanık olduğunu, kesin olmadığını belirtmiştir. Bu yüzden 0 ve 1 ile temsil edilen boolean mantık bu düşünce işlemini yeterli bir şekilde ifade edememektedir. İnsan mantığı; açık, kapalı, sıcak, soğuk, 0 ve 1 gibi değişkenlerden oluşan kesin ifadelerin yanı sıra; az açık, az kapalı, serin, ılık gibi ara değerleri de göz önüne almaktadır. Bulanık mantık, klasik mantığın aksine iki seviyeli değil, çok seviyeli işlemleri kullanmaktadır (Elmas, 2003).

Klasik denetim uygulamalarında karşılaşılan zorluklar nedeniyle, bulanık mantık denetimi alternatif yöntem olarak çok hızlı gelişmiş ve modern denetim alanında geniş uygulama alanı bulmuştur. Bulanık mantığın genel özellikleri Zadeh tarafından şu şekilde ifade edilmiştir;

- 1- Bulanık mantıkta, kesin değerlere dayanan düşünme yerine, yaklaşık düşünme kullanılır.
- 2- Bulanık mantıkta her şey $[0,1]$ aralığında belirli bir derece ile gösterilir.
- 3- Bulanık mantıkta bilgi büyük, küçük, çok az gibi dilsel ifadeler şeklindedir.
- 4- Bulanık çıkarım işlemi dilsel ifadeler arasında tanımlanan kurallar ile yapılır.
- 5- Her mantıksal sistem bulanık olarak ifade edilebilir.
- 6- Bulanık mantık matematiksel modeli çok zor elde edilen sistemler için çok uygundur.
- 7- Bulanık mantık, tam olarak bilinmeyen veya eksik girilen bilgilere göre işlem yapma yeteneğine sahiptir (Erbay Dalkılıç, 2005; Elmas, 2003).

1.5. Bulanık Sayı Kavramı

Bulanık sayılar dışbükey, normalleştirilmiş, sınırlı-sürekli üyelik fonksiyonu olan ve gerçel sayılarda tanımlanmış bir bulanık küme olarak ifade edilir. Bulanık sayı normal ve dışbükey olmalıdır. Bulanık kümeler üyelik fonksiyonlarıyla tanımlandıkları için

bulanık sayılar da kendi üyelik fonksiyonları ile aynı kavramdadır. Bu nedenle üyelik fonksiyonu çeşidi kadar bulanık sayı çeşidi vardır.

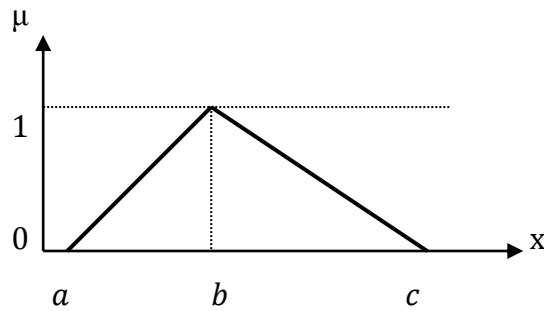
Gerçel sayılarda tanımlı, kapalı ve dışbükey bir alt küme olan $[a]$ bulanık sayısı $[a_1, a_2]$ sınırları ile de tanımlanabilir. Burada a_1 alt sınırı, a_2 ise üst sınırı gösterir. Daha genel bir ifade ile üyelik derecesi ile beraber değerlendirdiğimizde a_1 sayının sol tarafını, a_2 de sağ tarafını göstermektedir diyebiliriz. Bundan dolayı bulanık sayının sol tarafı $a_1: [0,1] \rightarrow R$ monoton artan iken $a_2: [0,1] \rightarrow R$ monoton azalandır. Üyelik dereceleri için değerlendirdiğimizde a sayısının dayanağının $(a_1(0), a_2(0))$ olduğu da açıktır. Bu a sayısı R gerçel sayılar kümesinin bir alt kümesidir (Baykal ve Beyan, 2004).

Tanım 1.1. Üçgensel Bulanık Sayı: A bir bulanık küme, $x \in A$ ve $\mu(x)$, x bulanık sayısının üyelik fonksiyonu olmak üzere, $\mu(x)$;

$$\mu(x) = \begin{cases} \frac{(x-a)}{(b-a)}, & a < x \leq b \\ 1, & x = b \\ \frac{(c-x)}{(c-b)}, & b < x \leq c \end{cases}$$

biçiminde tanımlandığında x bir üçgensel bulanık sayıdır.

Şekil 1’de x üçgensel bulanık sayısı $x = (a, b, c)$ biçiminde gösterilmiştir. Burada b merkez, $(b-a)$ sol yayılım ve $(c-b)$ sağ yayılımlardır. $(b-a) = (c-b)$ olduğunda üçgensel bulanık sayı, simetrik üçgensel bulanık sayı olarak adlandırılmaktadır (Klir ve Yuan, 1995; Erbay Dalkılıç, 2005).

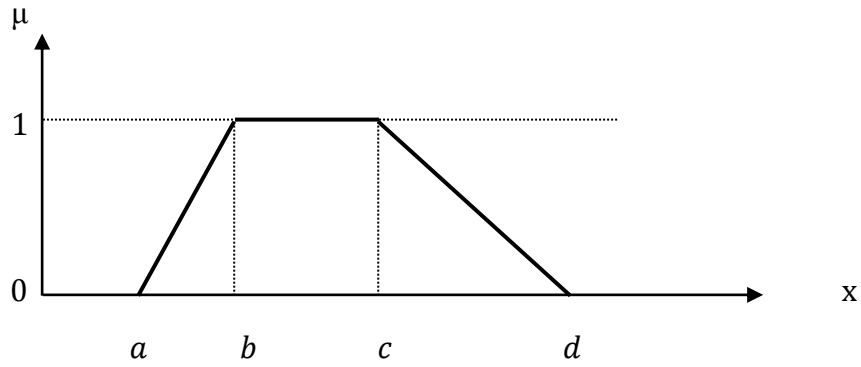


Şekil1. Üçgensel bulanık sayı

Tanım 1.2. Yamuksal Bulanık Sayı: A bir bulanık küme, $x \in A$ ve $\mu(x)$, x bulanık sayısının üyelik fonksiyonu olmak üzere, $\mu(x)$;

$$\mu(x) = \begin{cases} \frac{(x-a)}{(b-a)}, & a < x \leq b \\ 1 & , \quad b \leq x \leq c \\ \frac{(d-x)}{(d-c)}, & c < x \leq d \end{cases}$$

biçiminde tanımlandığında x , yamuksal bulanık sayıdır ve Şekil 2’de yamuksal bulanık sayı $x = (a, b, c, d)$ biçiminde gösterilmiştir. $(b-a)$ sol yayılma ve $(d-c)$ sağ yayılmalarıdır. $b = c$ olduğunda yamuksal bulanık sayı üçgensel bulanık sayıya dönüşmektedir (Klir ve Yuan, 1995).

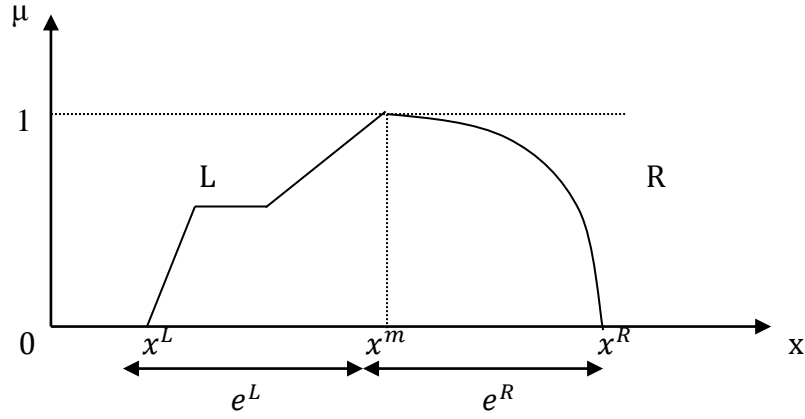


Şekil 2. Yamuksal bulanık sayı

Tanım 1.3. L – R Bulanık Sayı: A bulanık küme, $x \in A$ ve $\mu(x)$, x bulanık sayısının üyelik fonksiyonu olmak üzere, $\mu(x)$;

$$\mu(x) = \begin{cases} L((x^m - x)/e^L), & x^L < x \leq x^m \\ R((x - x^m)/e^R), & x^m \leq x \leq x^R \\ 0 & , \quad \text{diğer durumlarda} \end{cases}$$

biçiminde tanımlandığında x , L – R bulanık sayıdır ve Şekil 3’te L – R bulanık sayı $x = (x^m, x^L, x^R)_{LR}$ biçiminde gösterilmiştir. L – R bulanık sayılarda x^m merkez, e^L sol yayılma ve e^R sağ yayılma olarak adlandırılmaktadır. Üyelik fonksiyonun en büyük değeri aldığı nokta ise bulanık sayının yüksekliği olarak adlandırılır (Chen, 1999).

Şekil 3. $L - R$ bulanık sayı

1.6. Bulanık Küme Kavramı

Bulanık küme kavramı, 1960'ların ortasında Zadeh'in, klasik sistem kuramının matematiksel yöntemlerinin gerçek dünyadaki, özellikle insanları içeren kısmen karmaşık sistemlerle uğraşırken yetersiz kalmasından, hoşnut kalmayıp doğmuştur. Zadeh, niteliklerin ikili üyelik fonksiyonuyla ifade edildiği klasik kümeler yerine, dereceli üyelik fonksiyonuyla ifade edildiği bulanık kümeler tanımlamasını önermiştir. Bulanık küme kavramı, duyarlılığın artırılması açısından, klasik kümelerinkine göre daha uygun olan yeni bir araç sağlıyor olarak görülebilir. Getirdiği yaklaşım, klasik küme kuramlarında kullanılan üyelik kavramını bir kenara bırakıp yerine tamamen yenisini koymak değil, iki-değerli üyeliği çok-değerliliğe taşıyarak genelleştirmektir.

Bulanık küme, değişik üyelik derecesinde öğeleri olan bir topluluktur. Klasik küme teorisindeki siyah-beyaz ikili üyelik kavramını kısmi üyelik kavramına genelleştirir. Burada 0 değeri üye olmamayı, 1 değeri tam üye olmayı belirtirken (0,1) arası değerler de kısmi üyelik kavramına karşılık gelir.

Bulanık kümelerde üyelik dereceleri arasındaki geçiş yumuşak ve sürekli bir şekilde olmaktadır. Öğeler bulanık kümeye kısmi derecede aittir. Klasik kümelerdeki karakteristik fonksiyon,

$$\mu_A: E \rightarrow \{0,1\},$$

bulanık kümelerde yerini

$$\mu_A: E \rightarrow [0,1],$$

biçiminde ifade edilen üyelik fonksiyonuna bırakır.

Genel olarak küme üyelerini değerleri ile değişiklik gösteren eğriye üyelik fonksiyonu (önem eğrisi) adı verilir. Üyelik fonksiyonu grafiğinde x ekseni, üyeleri gösterirken y ekseni de üyelik derecelerini gösterir.

A bulanık kümesi, $\mu_A: E \rightarrow [0,1]$ A 'nın üyelik fonksiyonu ve $\mu_A(x) \in [0,1]$ $x \in E$ 'nin A bulanık kümesindeki üyelik derecesi olmak üzere;

$$A = \{(\mu_A(x), x)\}$$

olarak yazılabilir. Bu durumda E 'deki bulanık küme olan A ;

$$A = \{(\mu_A(x), x)\} = \{\mu_A(x)/x\}$$

$$A = \{\mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \dots + \mu_A(x_n)/x_n\}$$

biçimindedir ve genel olarak;

$$A = \left\{ \sum \mu_A(x_i)/x_i \right\}$$

biçiminde ifade edilebilir.

Bulanık kümenin sürekli olması durumunda gösterim;

$$A = \left\{ \int \mu_A(x_i)/x_i \right\}$$

biçiminde olacaktır.

1.7. Bulanık Küme Teorisinde Temel Tanımlar

X gözlemlerin klasik bir kümesi olmak üzere, x bu kümeye ait bir eleman ise bir A kümesinin karakteristik fonksiyonu,

$$\mu_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

biçimindedir. Burada karakteristik fonksiyonun değer kümesi $\{0,1\}$ olmaktadır.

Tanım 1.4. Destek kümesi: A bulanık kümesinin destek kümesi,

$$\begin{aligned} \text{des } A &= \{x | \mu_A(x) > 0 \text{ ve } x \in X\} \\ &= \{\text{üyelik dereceleri } 0 \text{ 'dan büyük olan } x \text{ 'ler}\} \end{aligned}$$

olarak ifade edilmektedir.

Tanım 1.5. α - Kesme Kümesi: A bulanık kümesinin α -kesme kümesi,

$$\begin{aligned} A_\alpha &= \{x | \mu_A(x) \geq \alpha \text{ ve } x \in X\} \\ &= \{\text{üyelik derecelerinin en az } \alpha \text{ kadar olan } x \text{ 'ler}\} \end{aligned}$$

ile belirlenmektedir.

Tanım 1.6. Normallik: A bulanık kümesi,

$$\text{Bulanık } A \text{ kümesi normaldir} \Leftrightarrow \sup_x \mu_A(x) = 1$$

koşulunu sağlıyorsa A bulanık kümesi normaldir. Aksi takdirde A bulanık kümesi alt normaldir. Alt normal bir kümeyi normalize etmek için A bulanık kümesinin bütün elemanları A 'nın en büyük üyelik derecesine bölünmelidir.

Tanım 1.7. Yükseklik: A bulanık kümesinin yüksekliği,

$$\begin{aligned} A &= \text{Enb } \mu_A(x) \\ &= \{A \text{ bulanık kümesinin en büyük üyelik derecesi}\} \end{aligned}$$

biçiminde gösterilir.

1.8. Bulanık Mantıkta Üyelik Fonksiyonu

Üyelik fonksiyonunun belirlenmesi bulanık küme teorisinin kullanıldığı uygulamalarda önemlidir. Ele alınan problemlerin yapılarına göre üyelik fonksiyonları farklılıklar gösterebilmektedir. Bununla birlikte üyelik fonksiyonlarının ortak birtakım özellikleri bulunmaktadır; üyelik fonksiyonları sürekli fonksiyonlardır ve bir $[a, b]$ aralığını $\mu(x)$ fonksiyonu yardımı ile $[0,1]$ aralığına dönüştürürler. Bir aralıktaki üyelik fonksiyonları dışbükey fonksiyon, içbükey fonksiyon ya da "c" noktası $[a, b]$ aralığında bir nokta olmak üzere $[a, c]$ aralığında dışbükey ve $[c, b]$ aralığında içbükey olabilmektedir. Üyelik fonksiyonunun doğrusal biçimde ya da doğrusallaşabilen bir yapıda olması büyük önem taşımaktadır. Genel anlamda problemlerin yapılarına göre farklılık gösteren üyelik fonksiyonları aşağıdaki gibi sınıflandırılabilir,

1. Sezgisel tanımlamalara dayanan üyelik fonksiyonları
2. Özel problemler için güvenilirliğe dayalı üyelik fonksiyonları
3. Teorik temele dayanan üyelik fonksiyonları

Zadeh tarafından önerilen ve daha sonra diğer araştırmacılar tarafından da kullanılan üyelik fonksiyonu, sezgisel üyelik fonksiyonlarına örnek olarak verilebilir ve x rasgele değişkeni yaşı gösterdiğinde aşağıdaki gibi tanımlanır;

$$\mu_{genç}(x) = \begin{cases} 1, & \text{eğer } x \leq 25 \\ \frac{1}{1 + \left(\frac{x-25}{5}\right)^2}, & \text{eğer } x > 25 \end{cases}$$

$$\mu_{yaşlı}(x) = \begin{cases} 0, & \text{eğer } x < 50 \\ \frac{1}{1 + \left(\frac{x-50}{5}\right)^{-2}}, & \text{eğer } x \geq 50 \end{cases}$$

Özel problemler için güvenilirliğe dayalı üyelik fonksiyonlarının en kullanışlılarından biri doğrusal bir fonksiyon olup Zimmermann tarafından önerilmiştir ve

$$\mu(x) = 1 - \frac{x}{a}, \quad x \in [0, a]$$

biçimindedir (Klir and Yuan, 1995).

Teorik temele dayanan üyelik fonksiyonları, herhangi bir dağılımdan gelen bir veri setine uygun üyelik fonksiyonunun oluşturulması için kullanılabilir yöntem Civanlar ve Trussel (1986) tarafından önerilen, olasılık yoğunluk fonksiyonu $p(x)$ 'e dayalı üyelik fonksiyonu belirleme yöntemidir.

Optimal üyelik fonksiyonunun bulunabilmesi için;

1. $E\{\mu(x) \mid x \text{ bir olasılık yoğunluk fonksiyonuna göre dağılsın}\} \geq 0$
2. $0 \leq \mu(x) \leq 1$
3. $\int \mu^2(x) d(x)$ en küçüklenmelidir.

biçiminde verilen koşulların sağlanması gerekmektedir. Bu koşullar altında optimal üyelik fonksiyonu;

$$\mu(x) = \begin{cases} \lambda p(x) & \text{eğer } \lambda p(x) < 1 \\ 1 & \text{eğer } \lambda p(x) \geq 1 \end{cases} \quad (12)$$

biçimindedir. Burada,

$p(x)$: olasılık yoğunluk fonksiyonu

λ : sabit

tir (Civanlar and Trussell, 1986).

1.9. Pareto Dağılımı

Pareto analizi, değişik sayıdaki önemli nedenleri daha az önemde olan nedenlerden ayırmak için kullanılan bir yöntemdir. Willfredo Pareto isimli İtalyan ekonomist Pareto prensibi de denilen 80/20 kuralını ortaya çıkarmış ve bu kural Pareto analizi yönteminin temelini oluşturmuştur. Willfredo Pareto ilk olarak İtalya topraklarının % 80' ine nüfusun % 20'sinin sahip olduğunu fark ederek bu prensibi kurmuştur. Daha sonra da hayatın diğer alanlarında da bu kuralın geçerli olduğunu keşfetmiştir. 80/20 kuralının uygulamalarına bazı örnekler:

- Müşteri şikâyetlerinin % 80'i ürün veya hizmetin % 20'sinden doğmaktadır.

- Herhangi bir programdaki gecikmelerin % 80'i çeşitli gecikme sebeplerinin sadece % 20'sinden kaynaklanır.
- Satış gücünün % 20'si şirket gelirinin % 80'ini etkiler.

Pareto prensibi göstermiştir ki yaşanan problemlerin büyük çoğunluğu az sayıdaki önemli faktörden kaynaklanmaktadır.

X , Pareto dağılımı gösteren rasgele bir değişken ise, X 'in olasılık değerini herhangi bir reel sayı olan x 'den daha büyük olması yani $x \geq x_m$ için aşağıdaki ifade verilebilir.

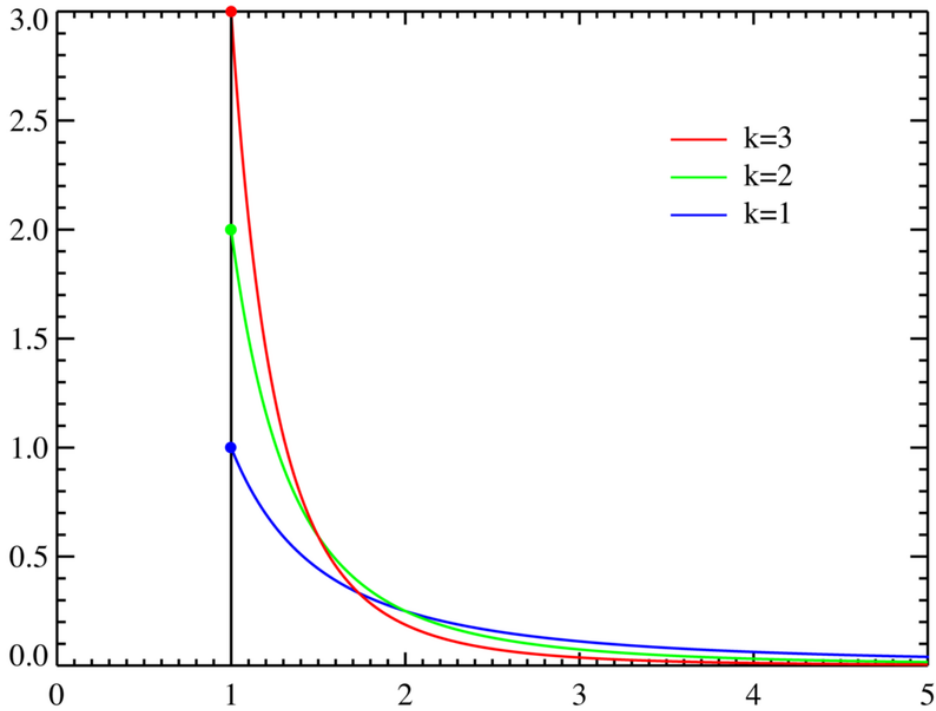
$$P(X > x) = \left(\frac{x_m}{x}\right)^k$$

burada;

x_m : X rastgele değişkenin aldığı en küçük değerdir ve ölçek parametresi olarak adlandırılır,

k : pozitif değerde bir parametredir ve şekil parametresi olarak adlandırılır.

Herhangi bir Pareto dağılımı için $x_m = 1$ olması durumunda dağılımın, olasılık yoğunluk fonksiyonunun, k şekil parametresine göre değişimi Şekil 4'te verilmiştir.



Şekil 4. Pareto dağılımının olasılık yoğunluk fonksiyonu

x_m ve k parametrelili Pareto dağılımına ilişkin dağılım fonksiyonu;

$$F(x) = 1 - \left(\frac{x_m}{x}\right)^k \quad x_m > 0, \quad k > 0; \quad x \in [x_m, +\infty)$$

biçimindedir.

Dağılım fonksiyonundan yararlanarak Pareto dağılımı için olasılık yoğunluk fonksiyonu aşağıdaki gibi elde edilebilir:

$$f(x) = \frac{d}{dx} F(x) = k \frac{x^{-k-1}}{x_m^{-k}} = \frac{kx_m^k}{x^{k+1}}$$

X rasgele değişkeni Pareto dağılımına sahip olduğunda beklenen değeri;

$$E(X) = \int_{x_m}^{\infty} x k x_m^k \frac{1}{x^{k+1}} dx = k x_m^k \frac{1}{k-1} x_m^{-k+1} = \frac{k x_m}{k-1} \quad , k > 1$$

biçiminde elde edilir. Dağılımın varyansını belirlemek için ilk olarak X^2 'nin beklenen değeri hesaplanmalıdır;

$$E(X^2) = \int_{x_m}^{\infty} x^2 k x_m^k \frac{1}{x^{k+1}} dx = k x_m^k \frac{1}{k-2} x_m^{-k+2} = \frac{k}{k-2} x_m^2 \quad , k > 2$$

biçiminde hesaplanır. Böylece, Pareto dağılımına ilişkin varyans;

$$V(X) = E(X^2) - \{E(X)\}^2$$

eşitliğinden,

$$V(X) = \frac{k}{k-2} x_m^2 - \left(\frac{k x_m}{k-1}\right)^2 = \frac{k x_m^2}{(k-1)^2(k-2)} \quad , k > 2$$

olarak elde edilir (URL-1, 2013).

1.10. Uyum İyiliği Testleri

Uyum iyiliği testlerinin amacı örneklem verilerinin öngörülen dağılıma uyup uymadığına karar vermektir. Uyum iyiliği için önerilen birçok test bulunmaktadır. Bunlar içerisinde yaygın olarak kullanılanlar ki-kare uyum iyiliği testi, daha çok sınıflama düzeyinde ölçülen değişkenler için kullanılmaktadır.

Oranlama ya da eşit aralıklı düzeyde ölçülen değişkenler için daha uygun olan uyum iyiliği testleri vardır. Bu testler ki-kare uyum iyiliğine göre daha çok bilgi kullanırlar. Bu nedenle daha güçlü testlerdir. Bu anlamda en yaygın kullanılan test Kolmogorov-Simirnov uyum iyiliği testidir. Ayrıca literatürde normallik için Shapiro-Wilk testi ve Lilliefors testi de mevcuttur.

Bu çalışmanın uygulama bölümünde ele alınacak verilerin Pareto dağılımına uygun olması ile ilgilenileceğinden her hangi bir dağılıma uygunluğu test etmek amacıyla kullanılabilen Kolmogorov-Simirnov testi örneklem verilerinin Pareto dağılımına uygunluğunu test etmede kullanılacaktır. Bu yöntem aşağıdaki gibi verilebilir.

1.10.1. Kolmogorov-Simirnov Uyum İyiliği Testi

Kolmogorov-Simirnov uyum iyiliği testi 1933'te Rus matematikçisi A.N. Kolmogorov tarafından önerilmiştir. Kolmogorov bir örnek için uyum iyiliği testini önerdikten sonra 1939'da yine bir Rus matematikçisi olan N.V. Simirnov iki bağımsız örnek için uyum iyiliği testini önermiştir. Kolmogorov testi ve Simirnov testi benzerlikleri nedeniyle, uygulamada, Kolmogorov-Simirnov uyum iyiliği testleri olarak bilinirler.

Bir örnek Kolmogorov-Simirnov uyum iyiliği testi iki birikimli dağılım fonksiyonuna dayanır. Birincisi yokluk hipotezinde belirtilen dağılımın birikimli dağılım fonksiyonudur. İkincisi örnekten elde edilen gözlenen birikimli dağılım fonksiyonudur. Birikimli dağılım fonksiyonu $F(x)$ ile gösterilir. Belirli bir X için $F(x)$, X rastgele değişkeninin değerinin x değerine eşit ya da daha küçük olması olasılığıdır. Bu matematiksel olarak

$$F(x) = P(X \leq x)$$

biçiminde ifade edilir.

Birikimli dağılım fonksiyonu bilinmeyen bir yığından rastgele bir örnek seçtiğimizi varsayalım. $F_0(x)$ yokluk hipotezinde belirlenen yığın birikimli dağılım fonksiyonu olmak üzere, her x için $F(x) = F_0(x)$ eşitliğinin doğru olup olmadığını ortaya çıkarmayı arzulamaktır. Gözlenen birikimli dağılım fonksiyonuna $S(x)$ dendiğinde, eğer $F(x) = F_0(x)$ ise $F_0(x)$ ve $S(x)$ arasında da uyum olması beklenilir. Ancak örnek değişkenliği nedeniyle bazı x değerleri için $F_0(x)$ ve $S(x)$ arasında az da olsa uyumdan sapmalar olması doğaldır. Bir örnek Kolmogorov-Simirnov uyum iyiliği testinin amacı, $F_0(x)$ ve $S(x)$ arasındaki uyuma bağlı olarak $F(x) = F_0(x)$ hipotezinin doğru olup olmadığına karar vermektir.

Birikimli dağılım fonksiyonu bilinmeyen bir yığından n hacimli rastgele bir örnek seçilmiş olsun. Bu durumda veri x_1, x_2, \dots, x_n gibi bağımsız gözlem sonuçlarından oluşacaktır. $F_0(x)$ hipotezde belirtilen dağılımın birikimli dağılım fonksiyonu iken, yokluk hipotezi ve karşıt hipotez;

$$H_0: F(x) = F_0(x), \text{ tüm } x \text{ değerleri için,}$$

$$H_1: F(x) \neq F_0(x), \text{ en az bir } x \text{ değeri için,}$$

biçiminde ifade edilir.

$S(x)$, x değerine eşit ya da daha küçük değerli örnek birimlerinin sayısının örnek hacmine oranıdır ve

$$S(x) = \frac{\text{x değerine eşit ya da daha küçük değerli örnek hacimlerinin sayısı}}{n}$$

eşitliği ile hesaplanır.

İki-yanlı bir test için tek örnek Kolmogorov-Simirnov test istatistiği

$$D = \sup |S(x) - F_0(x)|$$

biçiminde tanımlanır.

Supremum (sup) sayılabilir kümeler için en büyük olarak düşünülebilir. Bulunan test istatistiği, Kolmogorov-Smirnov uyum iyiliği testi için düzenlenen kritik değerler tablosunda yer alan D_k değeri ile karşılaştırılarak;

$D_h \geq D_k$ ise H_0 hipotezi ret edilir.

$D_h \leq D_k$ ise H_0 hipotezi ret edilemez.

H_0 hipotezinin kabulü, örnek verilerin hipotezde verilen dağılıma uygun olduğu anlamına gelir (Gamgam, 2012; Ünver, 1999).

1.11. Aykırı Değerlerin Belirlenmesi

Bu çalışmada önerilen parametre tahmini yönteminin amaçlarından biri de veri setinde yer alan aykırı gözlemlerden mümkün olduğu kadar az etkilenen parametreler elde ederek hata miktarı düşük tahmin değerlerine ulaşmaktır. Bu anlamda veri setinde var olan aykırı gözlemlerin belirlenmesi önemlidir.

Regresyon çözümlemesindeki model bozukluklarının araştırmak ve aykırı değeri belirlemek için basit ve etkin yöntem artıkların incelenmesidir. Çünkü artıklar hataların, gerçekleşen ya da gözlenen değerleri olarak düşünülebilir. Bu nedenle hatalar için söz konusu olan varsayımlara ilişkin bozulmalar artıklar yardımı ile incelenir. Bu incelemede değişik artık türlerinden yararlanılır, ham artıklar, standartlaştırılmış artıklar, student türü artıklar bunlardan bazılarıdır (Alpar, 2003).

Klasik regresyon analizinden bilindiği gibi ham artık;

$$e_i = y_i - \hat{y}_i \quad , i = 1, \dots, n$$

olarak tanımlanır. Bu tür artıklar ölçeklenmemiş artıklar olarak adlandırılır.

Ham artıkların toplamı dolayısıyla ortalaması sıfır olmakla birlikte varyansı (S^2) örneklemden örnekleme değişir. Varyansın değişiklik göstermesi sorunu artıkların standartlaştırılması ile giderilir. Bu durumda ham artıklar standart sapmaya (S) bölünerek standartlaştırılmış artıklar,

$$e_{is} = \frac{e_i}{S} \quad , i = 1, \dots, n$$

biçiminde elde edilir. e_{is} artıklarının %95'inin çoğunlukla $[-2; +2]$ sınırları arasında değiştiği kabul edilir. Bu sınırların dışına çıkan gözlemlerin aykırı değer kuşkusu ile incelenmesi önerilir. e_{is} artıkları her zaman sıfır ortalamalı ve bir standart sapma ile normal dağılım göstermeyebilir. Bu durumda artıklar kendi standart sapması ile standartlaştırıldığında elde edilen artıklara student türü artıklar denir ve

$$r_i = \frac{e_i}{\sqrt{S^2(1 - p_{ii})}} \quad , i = 1, \dots, n$$

biçiminde ifade edilir. Burada,

p_{ii} : gözlem uzaklığı değeridir, i . gözlem noktasının değişkenlerin oluşturduğu elipsoit merkezine uzaklığının bir ölçüsü olup 0 ile 1 arasında değişir ve,

$$p_{ii} = X(X'X)^{-1}X'$$

ile ifade edildiğinde p_{ii} 'nin köşegen elemanları i . gözlem noktasına ilişkin gözlem uzaklığını verir. Bu uzaklıklarında standartlaştırılmış artıklarda olduğu gibi $[-2; +2]$ aralığında kalması beklenir. Aralık dışında kalan gözlemler artık gözlem olarak adlandırılır (Apaydın, 1994; Alpar, 2003).

2. YAPILAN ÇALIŞMALAR

Bu çalışmada Pareto dağılımına sahip bağımsız değişkene ilişkin değerler Pareto dağılımına uygun elde edilen üyelik fonksiyonu ile derecelendirilmiş, böylece regresyon modelinin bilinmeyen parametrelerinin belirlenmesi aşamasında aykırı değerlerden mümkün olduğunca az etkilenen tahminler elde etmeyi amaçlayan bir algoritma önerilmiştir. Önerilen algoritmalarından elde edilen sonuçlar literatürde yer alan mevcut yöntemlerden elde sonuçlar ile karşılaştırılarak etkinliği irdelenmiştir.

2.1. Pareto Dağılımı İçin Optimal Üyelik Fonksiyonunun Belirlenmesi

Bağımsız değişkenlerden herhangi birinin Pareto dağılımı göstermesi durumunda önerilecek algoritmada kullanılacak üyelik fonksiyonu Pareto dağılımına uygun üyelik fonksiyonu olmalıdır. Dağılıma uygun üyelik fonksiyonunun belirlenmesinde izlenecek yol değişkenin olasılık yoğunluk fonksiyonu $p(x)$ 'e dayalı üyelik fonksiyonu belirleme yönteminin kullanılmasıdır. Kesim (1.8)'de Eşitlik (12) ile verilen ve olasılık yoğunluk fonksiyonuna dayalı üyelik fonksiyonunda $p(x)$, ilgilenilen dağılıma ilişkin olasılık yoğunluk fonksiyonu olduğundan formu belirlidir. Ancak λ sabiti,

$$\begin{aligned} \mathbf{P}: \min_{\mu} f(\mu) &= \frac{1}{2} \int_{-\infty}^{+\infty} \mu^2(x) d(x) \\ G(\mu) &= c - E\{\mu\} = c - \int_{-\infty}^{+\infty} \mu(x)p(x) d(x) \leq 0 \\ \mu &\in \Omega = \{\mu(x) \mid 0 \leq \mu(x) \leq 1\} \end{aligned} \quad (13)$$

ile tanımlanan problemin çözümü ile elde edilebilir (Civanlar ve Trussel 1986). \mathbf{P} ile verilen problem, optimal üyelik fonksiyonu için tanımlanan koşullardan oluşturulmaktadır. λ sabitinin elde edilmesi için, Eşitlik (13)'de \mathbf{P} ile ifade edilen problem, Langrange yöntemi ile çözümlenebilir. Problemin çözümü için oluşturulacak Lagrange fonksiyonu;

$$L(\mu, \lambda) = \frac{1}{2} \int_{-\infty}^{+\infty} \mu^2(x) d(x) + \lambda \{c - \int_{-\infty}^{+\infty} \mu(x)p(x) d(x)\} \quad (14)$$

biçimindedir. Burada, $\lambda \geq 0$ Langrange çarpanı ve $c < 1$ olacak biçimde bir sabittir.

Eşitlik (12)' de verilen üyelik fonksiyonunun değerleri Eşitlik (14)' te yerine konularak λ sabitinin değerini verebilecek fonksiyona ulaşılmaya çalışılır. Bunun için $[\lambda p(x)] < 1$ ve $[\lambda p(x)] \geq 1$ olacak şekilde çözümlenmelidir. Ancak, $[\lambda p(x)] \geq 1$ olması durumunda üyelik fonksiyonu λ 'dan bağımsız olduğundan çözümlenme sadece $[\lambda p(x)] < 1$ durumu için aşağıdaki biçimde yapılır;

$$[\lambda p(x)] < 1 \Rightarrow \mu(x) = \lambda p(x)$$

olur. Bu durumda Lagrange fonksiyonu;

$$\begin{aligned} L &= \frac{1}{2} \int_{-\infty}^{+\infty} \lambda^2 p^2(x) d(x) - \lambda \int_{-\infty}^{+\infty} \lambda p^2(x) d(x) + \lambda c \\ &= \frac{1}{2} \int_{-\infty}^{+\infty} \lambda^2 p^2(x) d(x) - \int_{-\infty}^{+\infty} \lambda^2 p^2(x) d(x) + \lambda c \\ L &= -\frac{1}{2} \int_{-\infty}^{+\infty} \lambda^2 p^2(x) d(x) + \lambda c \end{aligned}$$

biçiminde elde edilir. Bu fonksiyonun λ sabitine göre türevi alındığında,

$$\frac{\partial L}{\partial \lambda} = -\lambda \int_{-\infty}^{+\infty} P^2(x) d(x) + c \quad (15)$$

eşitliğine ulaşılır. Elde edilen (15) eşitliği kullanarak λ parametresi,

$$\frac{\partial L}{\partial \lambda} = -\lambda \int_{-\infty}^{+\infty} P^2(x) d(x) + c \quad (16)$$

Eşitliğinin sıfıra eşitlenmesi ile,

$$\lambda = \frac{c}{\int_{-\infty}^{+\infty} p^2(x) d(x)} \quad (17)$$

biçiminde elde edilir.

Bu çalışmada veri setinde yer alan değişkenlerden herhangi birinin Pareto dağılımına sahip olması ve aykırı değerlerin varlığı durumunda doğrusal regresyon modelinin parametrelerinin tahmini ile ilgilenildiğinden, Pareto dağılımına ilişkin üyelik fonksiyonu Civanlar ve Trussel (1986) tarafından önerilen bu yöntem kullanılarak aşağıdaki gibi elde edilmiştir:

Eğer X Pareto dağılımına sahip rastgele bir değişken ise olasılık yoğunluk fonksiyonu,

$$f(x; k, x_m) = k \frac{x_m^k}{x^{k+1}} \quad x \geq x_m \quad (18)$$

biçimindedir. Pareto dağılımı için optimal üyelik fonksiyonun elde edilmesi için öncelikle dağılıma uygun λ parametresi belirlenmelidir. Pareto dağılımına ilişkin Eşitlik (18) ile verilen olasılık yoğunluk fonksiyonu (7) eşitliğinde yerine konulduğunda λ sabiti,

$$\begin{aligned} \lambda &= \frac{c}{\int_{x_m}^{+\infty} p^2(x) d(x)} = \frac{c}{\int_{x_m}^{+\infty} \left[k \frac{x_m^k}{x^{k+1}} \right]^2 d(x)} \\ &= \frac{c}{k^2 x_m^{2k} \int_{x_m}^{+\infty} \frac{1}{x^{2(k+1)}} d(x)} \\ \lambda &= c \frac{2(k+1)x_m}{k^2} \end{aligned} \quad (19)$$

biçiminde elde edilir. Üyelik fonksiyonu, (19) eşitliği ile elde edilen λ parametresini kullanarak,

$$\begin{aligned} \mu(x) &= \lambda p(x) = c \frac{2(k+1)x_m}{k^2} k \frac{x_m^k}{x^{k+1}} \\ \mu(x) &= \frac{2(k+1)}{k} \left(\frac{x_m}{x} \right)^{k+1} \end{aligned} \quad (20)$$

biçiminde elde edilir.

2.2. Parametre Tahmini İçin Bulanık Üyelik Derecesine Dayalı Algoritma

Klasik regresyon analizinde parametre tahmini yapılmadan önce doğrusal regresyon analizine ilişkin çalışmanın birinci bölümünde anlatılan varsayımların sağlanması gerekir. Bu çalışmada bağımsız değişkenlerden herhangi birinin dağılımının normal dağılım dışına çıkması ve gözlemlerde aykırı değerlerin varlığı durumunda parametre tahmini için bu varsayım bozulmalarından daha az etkilenecek bir parametre tahmini yöntemi önerilmiştir. Bu yöntemde, bağımsız değişkenlerden herhangi birinin Pareto dağılımına uygunluğu söz konusu olduğunda, bu değişkenin modele katkı miktarı Pareto dağılımına ilişkin üretilen üyelik fonksiyonu ile belirlenmiştir. Var olan aykırı değerlerin hata toplamına etkisi üyelik fonksiyonundan elde edilen üyelik dereceleri ile ilişkilendirilip azaltılmaya çalışılmıştır. Önerilen yöntem adımsal olarak aşağıdaki gibi tanımlanabilir;

Adım 1: Bağımsız değişkenlerin histogramları incelenir ve kurulacak modele katkılarını belirlemek için, normal dağılım dışında bir dağılıma sahip olmaları durumu belirlenir.

Adım 2: Pareto dağıldığı düşünülen değişkenin istatistiksel olarak bu dağılıma sahip olduğu uyum iyiliği testi ile belirlenir ve hangi parametreler ile bu dağılıma sahip olduğu ortaya konur.

Adım 3: Bölüm 2.3'te Pareto dağılımı için elde edilen

$$\mu(x) = \frac{2(k+1)}{k} \left(\frac{x_m}{x}\right)^{k+1}$$

biçiminde tanımlanan üyelik fonksiyonu ile bu dağılıma sahip bağımsız değişkene ilişkin gözlem değerlerinin modele katkı dereceleri hesaplanır.

Adım 4: Pareto dağılımına sahip j . değişken belirlenen katkı dereceleri ile

$$x_{ij}^* = \mu(x_j) * x_{ij}$$

biçiminde ağırlıklandırılır.

Adım 5: Regresyon modeline ilişkin parametreler

$$B = (X^T X)^{-1} X^T Y$$

eşitliği kullanılarak belirlenir. Burada X matrisi Pareto dağılımına ilişkin değişkenin ağırlıklandırılmış biçimi olan X_{ij}^* 'yi içerir ve

$$Y = [y_1, y_2, \dots, y_n]^T$$

biçiminde tanımlanır.

Adım 6: Modele ilişkin hata

$$\varepsilon = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}$$

biçiminde hesaplanır.

Adım 7: Veriler için En Küçük Kareler yöntemi kullanılarak elde edilen modelden ve bu çalışmanın uygulama kısmında yer verilen özel örnek için literatürde teorik bir model olarak yer alan Bertoni-Walfisch modelinden elde edilen tahminlere ilişkin hatalar belirlenerek önerilen tahmin yönteminin etkinliği hataların karşılaştırılması yolu ile yorumlanır.

2.3. Uygulama

Bu bölümde, çalışmanın (2.1) Kesimi'nde önerilen algoritmanın kullanılabilirliğini irdelenmek amacı ile çeşitli örneklere yer verilmiştir. Örnek 2 ve Örnek 3 te yer alan veriler için tahminler, önerilen algoritma ve EKK yönteminden elde edilmiş ve karşılaştırma bu yöntemden elde edilen tahminlere ilişkin hataların karşılaştırılması ile gerçekleştirilmiştir. Örnek 1'de yer alan veriler diğer örneklerde yer alan verilerden farklı, özel bir probleme ait veriler olduğundan algoritmanın etkinliğinin incelenmesi aşamasında EKK yöntemine ek

olarak literatürde yer alan ve Bertoni-Walfisch (1988) tarafından önerilen teorik modelden elde edilen tahminler de karşılaştırmalar için kullanılmıştır.

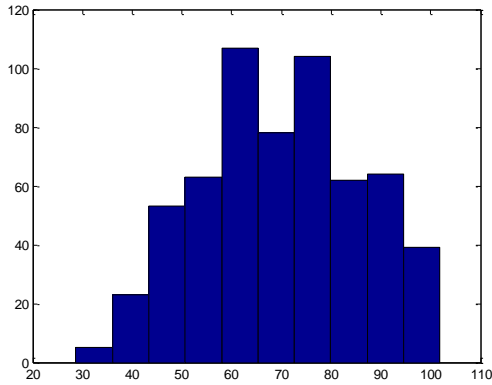
Örnek1: Önerilen algoritmanın geçerliliğini sınamak için bu örnekte ele alınan veriler mobil iletişim sistemleri ile ilgili olup yol kaybı modellerinin oluşturulması için kullanılan ölçümlerdir ve 598 gözlemden oluşan verilerin bir kısmı Tablo 2’de yer almaktadır. Veri setinde bir bağımlı ve üç bağımsız değişken yer almaktadır. Bağımlı değişken Y ; mobil istasyona gelen sinyal seviyesi miktarını, bağımsız değişkenler,

X_1 : Bina merkezleri arası uzaklığı,

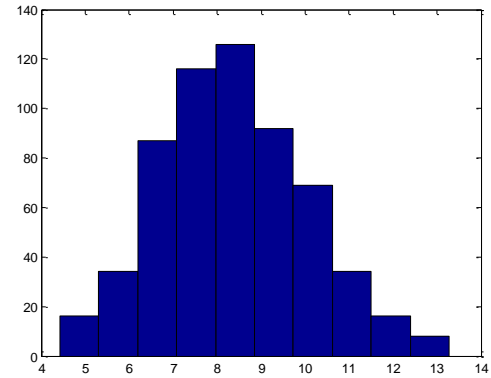
X_2 : Bina yüksekliklerini,

X_3 : Baz istasyonu ile mobil anten arasındaki yayılma açısını,

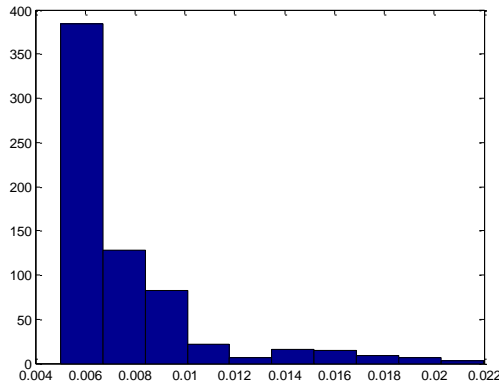
göstermektedir.



Şekil 5. X_1 değişkenine ilişkin histogram



Şekil 6. X_2 değişkenine ilişkin histogram



Şekil 7. X_3 değişkenine ilişkin histogram

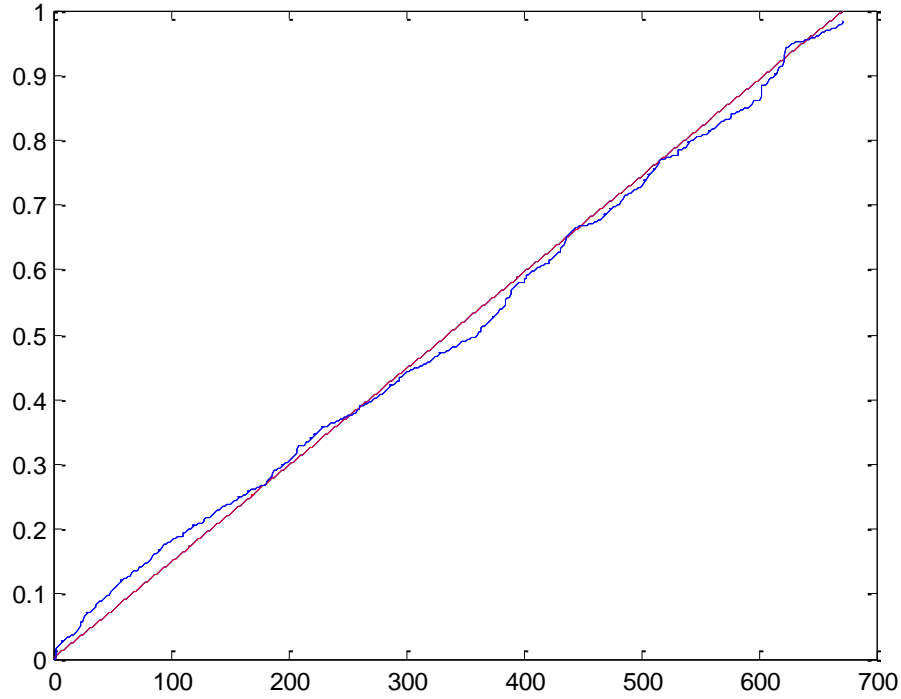
Bağımsız değişkenlerden X_3 Kolmogrov-Smirnov uyum iyiliği testine göre ($k = 2,75$, $x_m = 0,005$) parametreleri ile pareto dağılımına sahip olduğu Tablo 1'de gösterilmiştir. Tabloda hesaplanan K-S değeri ve $\alpha=0.01$ için K-S tablosundan elde edilen kritik değer yer almaktadır.

$$K - S_{Hesaplanan\ Değer} (= 0,0565) < K - S_{Tablo\ Değeri} = (0,066)$$

olduğundan X_3 rasgele değişkeninin dağılımının Pareto dağılımından farkı yoktur biçiminde kurulan yokluk hipotezi kabul edilir. Ayrıca Pareto dağılımına uygunluğa ilişkin grafik Şekil 8'de gösterilmektedir. Şekil 8'de kırmızı ile gözlenen birikimli dağılım fonksiyonu $S(x)$ ve mavi ile gösterilen birikimli dağılım fonksiyonu $F_0(x)$ 'dir.

Tablo 1. Pareto dağılımına uyum

N	k	K-S (hesaplanan değer)	K-S (tablo)
598	2,75	0,0565	< 0,0666



Şekil 8. X_3 değişkeninin Pareto dağılımına uygunluğu

Veri setine ilişkin tahminlerin yer aldığı Tablo 2’de önerilen algoritmadan ve EKK yönteminden elde edilen tahminlerin ve bu tahminlere ilişkin hataların yanında bu örnek için literatürde yer alan ve Bertoni-Walfisch tarafından önerilen modele ilişkin tahminler ve hata değerleri de yer almaktadır. Ayrıca tabloda veri seti için student türü artıklarda yer almakta ve aykırı gözlemlere ilişkin değerler koyu renk ile gösterilmektedir.

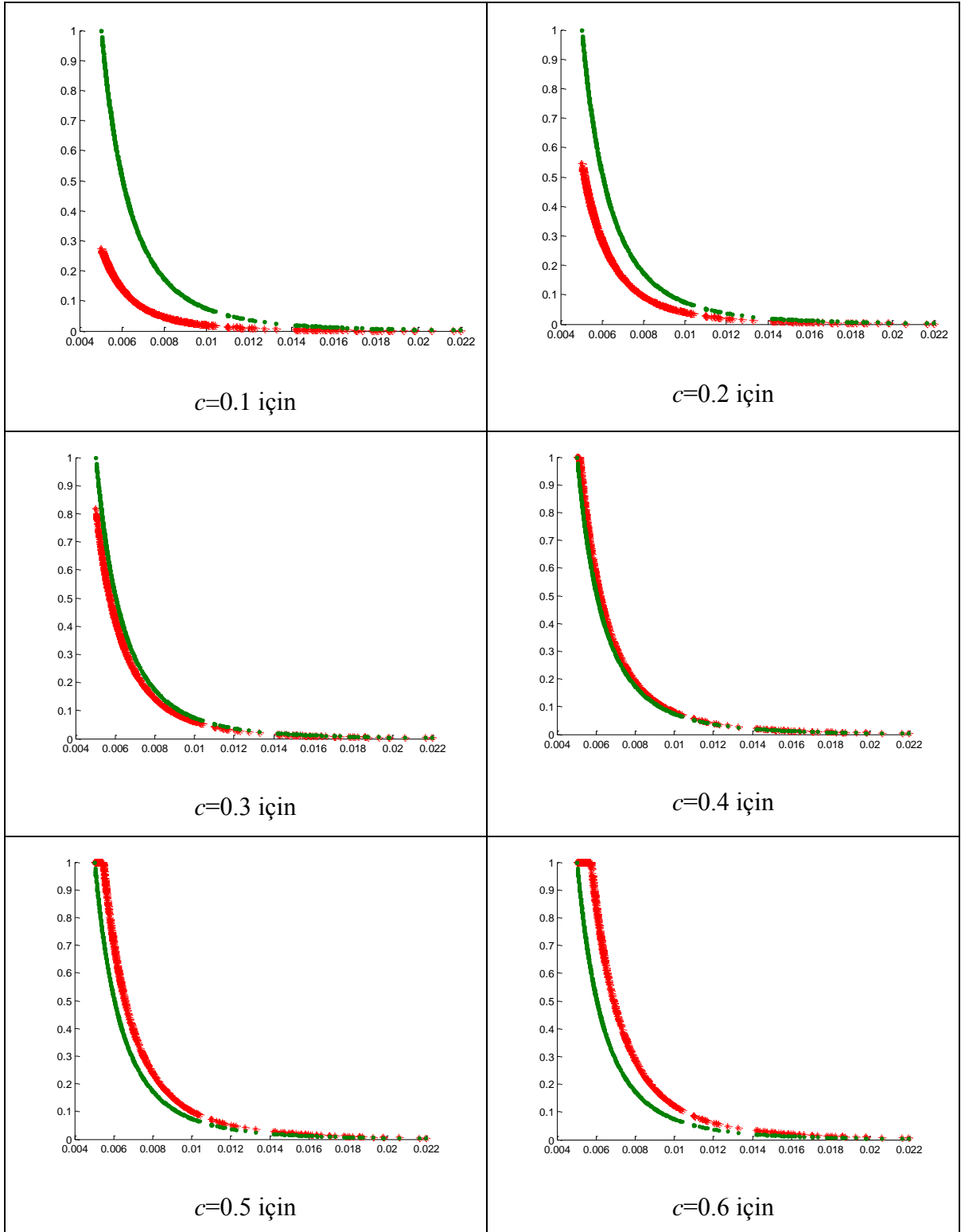
Tablo 2. Veri setine ilişkin tahminler ve hatalar

Gözlem No	X_1	X_2	X_3	Y	$\hat{Y}_{(OA)i}$	$e_{(OA)i}$	$\hat{Y}_{(EKK)i}$	$e_{(EKK)i}$	$\hat{Y}_{(ber)i}$	$e_{(ber)i}$	Student Türü Artıklar	
1	28.6900	7.0800	0.0098	17.1510	41.9515	-24.8005	40.1383	-22.9873	33.9300	-16.7790	-2,13340	
2	30.8900	7.2700	0.0099	24.2520	42.1260	-17.8740	40.0500	-15.7980	33.5100	-9.2580	-1,46533	
3	34.2000	5.8300	0.0075	40.9530	39.9428	1.0102	42.2753	-1.3223	32.7000	8.2530	-0,12250	
4	34.7900	4.8400	0.0060	31.7880	38.7924	-7.0044	43.6555	-11.8675	32.2400	-0.4520	-1,10108	
5	35.2100	7.9200	0.0061	40.6780	43.6217	-2.9437	46.6059	-5.9279	36.9700	3.7080	-0,54831	
6	50.5400	9.0400	0.0012	155.6589	126.6912	28.9677	54.8124	100.8465	155.2200	0.4389	9,32907	
7	38.7100	5.3700	0.0072	45.4170	39.0285	6.3885	42.0444	3.3726	31.1900	14.2270	0,31234	
8	38.3200	4.7700	0.0062	27.9600	38.4022	-10.4422	43.0684	-15.1084	31.1000	-3.1400	-1,40102	
9	37.2200	5.5400	0.0059	38.7620	39.8192	-1.0572	44.3963	-5.6343	33.3600	5.4020	-0,52188	
10	36.0700	8.0300	0.0076	36.6670	43.3186	-6.6516	44.2508	-7.5838	35.1200	1.5470	-0,70156	
11	41.0200	5.9400	0.0079	43.4360	39.6726	3.7634	41.3688	2.0672	30.9900	12.4460	0,19123	
12	40.9600	6.4800	0.0075	43.0870	40.6060	2.4810	42.5670	0.5200	32.2500	10.8370	0,04808	
13	40.6900	5.8300	0.0072	44.2140	39.6514	4.5626	42.4042	1.8098	31.6300	12.5840	0,16744	
14	40.3600	6.8200	0.0062	41.5770	41.5495	0.0275	45.0432	-3.4662	34.3400	7.2370	-0,32039	
15	40.7400	7.5400	0.0077	44.6000	42.2656	2.3344	43.3352	1.2648	33.5200	11.0800	0,11688	
16	40.4600	6.5700	0.0076	41.1750	40.7572	0.4178	42.5246	-1.3496	32.4700	8.7050	-0,12478	
17	40.8500	7.5800	0.0017	185.7895	73.7014	112.0881	53.0500	132.7395	185.2100	0.5795	12,29169	
18	40.5200	7.3400	0.0063	45.7740	42.3283	3.4457	45.4024	0.3716	34.9300	10.8440	0,03434	
19	44.0700	6.5200	0.0077	36.3130	40.4617	-4.1487	42.1148	-5.8018	31.5600	4.7530	-0,53609	
20	42.6100	6.4400	0.0066	44.9820	40.6787	4.3033	43.8880	1.0940	32.9000	12.0820	0,10111	
21	44.1600	5.2600	0.0060	28.8540	38.9471	-10.0931	43.5702	-14.7162	31.2400	-2.3860	-1,36227	
22	43.1800	7.9700	0.0095	40.9870	42.5960	-1.6090	40.7353	0.2517	31.8700	9.1170	0,02327	
23	43.8500	7.9100	0.0091	34.7490	42.4992	-7.7502	41.2829	-6.5339	32.0200	2.7290	-0,60382	
24	44.3000	7.1600	0.0078	43.0630	41.4491	1.6139	42.5923	0.4707	32.2900	10.7730	0,04348	
25	43.3300	6.1700	0.0076	43.1100	39.9641	3.1459	41.9605	1.1495	31.2600	11.8500	0,10627	
26	43.8400	8.2100	0.0065	36.6640	43.4580	-6.7940	45.7835	-9.1195	34.8200	1.8440	-0,84209	
27	42.2200	6.2700	0.0067	48.1140	40.3988	7.7152	43.5750	4.5390	32.5600	15.5540	0,41961	
28	44.9400	7.2600	0.0062	42.1820	41.9982	0.1838	45.2404	-3.0584	34.0300	8.1520	-0,28241	
29	43.3900	7.6000	0.0056	44.8370	42.9169	1.9201	46.6393	-1.8023	35.4300	9.4070	-0,16645	
30	44.4400	5.1000	0.0060	31.7320	38.6774	-6.9454	43.3921	-11.6601	30.9400	0.7920	-1,07965	
.	
.	
.	
589	100.210	5.9400	0.0059	49.8990	37.0060	12.8930	41.3567	8.5423	25.8800	24.0190	0,79295	
590	100.060	6.8500	0.0053	40.5670	38.8179	1.7491	43.2591	-2.6921	28.1600	12.4070	-0,24958	
591	101.870	7.8900	0.0121	19.0050	39.1186	-20.1136	33.2477	-14.2427	22.8600	-3.8550	-1,31821	
592	38.5400	8.0400	0.0010	105.6589	143.2661	-37.6072	54.7739	50.8850	125.5600	-19.9011	0,42113	
593	100.050	8.1600	0.0082	40.3410	39.9241	0.4169	39.9142	0.4268	26.3000	14.0410	4,71570	
594	99.9500	8.3700	0.0198	22.1680	39.8797	-17.7117	21.4184	0.7496	19.7200	2.4480	0,07016	
595	100.170	9.4000	0.0116	45.6820	41.6300	4.0520	35.6842	9.9978	24.8400	20.8420	0,92500	
596	100.570	5.3900	0.0113	24.3580	35.2453	-10.8873	32.0651	-7.7071	19.8400	4.5180	-0,71498	
597	101.530	7.6600	0.0099	35.7490	38.8780	-3.1290	36.5816	-0.8326	24.1700	11.5790	-0,07700	
598	99.4400	9.9900	0.0070	47.4810	43.0962	4.3848	43.7461	3.7349	29.2800	18.2010	0,34530	
	Hata				$\epsilon_{OA} = 78.6845$	$\epsilon_{EKK} = 117.2110$	$\epsilon_{BER} = 174.8910$					

Verilere ilişkin çoklu doğrusal regresyon modelini elde etmeden önce Pareto dağılımı gösterdiği belirlenen X_3 değişkenine ilişkin üyelik dereceleri,

$$\mu(X_{3i}) = \frac{2(k+1)}{k} \left(\frac{X_{3m}}{X_{3i}} \right)^{k+1}$$

ile belirlenir, burada $k = 2,75$ ve $X_{3m} = 0,005$ 'dir. $[0,1]$ aralığında değerler alabilen c sabiti 0.4 olarak belirlenmiştir. Bu sabit, üyelik fonksiyonunun olasılık yoğunluk fonksiyonuna uyumunun kontrolü ile elde edilmiştir. Uygun c sabitinin belirlenme süreci Şekil 9'da yer almaktadır.



Şekil 9. Uygun c 'nin belirlenmesi

Önerilen algoritma ile çoklu doğrusal regresyon modeli;

$$\hat{Y}_{\text{ÖA}} = 32.0091 - 0.05475x_1 + 1.5897x_2 + 10058.56x_3 \quad (21)$$

biçiminde ve En Küçük Kareler yöntemine ilişkin model;

$$\hat{Y}_{\text{EKK}} = 50.3132 - 0.05471x_1 + 1.0177x_2 - 1613.356x_3 \quad (22)$$

biçiminde elde edilmiştir.

Modelin geçerliliğinin testi için klasik yöntemlerden biri olan En Küçük Kareler Yönteminin dışında, küresel mobil iletişim sistemlerinde modelleme anlamında geçerliliğe sahip olan Bertoni-Walfisch (1988) tarafından önerilen teorik modelden elde edilen tahmin ve bu tahminlere ilişkin hata değerlerine de yer verilmiştir. Sözü edilen teorik model;

- 1) Alıcı ve verici anten arasındaki serbest uzay yayılım yol kaybı
- 2) Bina yüksekliğinden dolayı işaret seviyesi düşmesi ile oluşan yol kaybı
- 3) Bina çatısından cadde seviyesine kırınımın etkisi ile oluşan yol kaybı

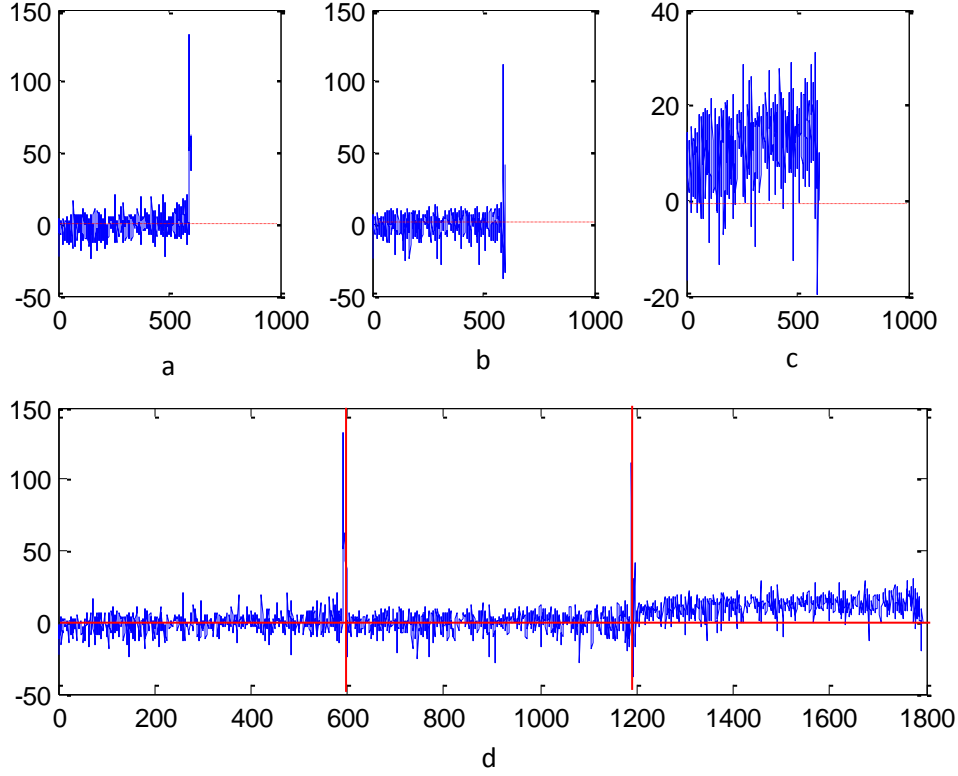
biçiminde üç ana bileşenden oluşur. Toplam yol kaybının dB cinsinden ifadesi

$$PL_{B-W}(dB) = 89.5 + 21 \log f + 38 \log R - 18 \log(h_t - h_b) + A_b \quad (23)$$

ile verilmektedir. Bina yükseklikleri h_b , binalar arası uzaklık d_c , ve sinyal açısı α bağımsız değişkenleri ile ifade edilmektedir (Hancı 2005).

Önerilen algoritmadan elde edilen modelin tahmin değerlerine ilişkin hata $\varepsilon_{\text{ÖA}} = 78.684$, En Küçük Kareler yönteminden elde edilen modelin tahmin değerlerine ilişkin hata $\varepsilon_{\text{EKK}} = 117.2110$ ve Bertoni-Walfisch tarafından önerilen teorik modelden elde edilen tahminler için hata $\varepsilon_{\text{BER}} = 174.8910$ olarak elde edilmiştir. Şekil 10-a En Küçük Kareler modelinden elde edilen hata miktarlarını, Şekil 10-b önerilen algoritmadan elde edilen tahminlerin hata miktarlarını, Şekil 10-c ise Bertoni-Walfisch tarafından önerilen

teorik modelden elde edilen tahminlerin hata miktarlarını göstermektedir. Tüm modellerden elde edilen hatalar grafikleri karşılaştırmalı olarak Şekil 10-d'de yer almaktadır.



Şekil 10. Tablo 2’de yer alan veri setine ilişkin hata grafikleri

Önerilen algoritma ile elde edilen tahminlerin En Küçük Kareler yönteminden elde edilen modele ilişkin tahminlerden ve Bertoni-Walfisch modelinden elde edilen tahminlerden elde edilen hatalara göre daha küçük hatalara sahip olduğu gözlenmiştir.

Örnek2: Bu örnekte yer alan veri seti 66 gözlemden oluşmakta ve verilerin bir kısmı Tablo 3’te yer almaktadır. Veri setinde bir bağımlı ve iki bağımsız değişken bulunmaktadır. Bağımlı değişken Y , bağımsız değişkenler X_1 ve X_2 ile ifade edilmiştir. Bağımsız değişkenlerden X_2 Kolmogrov-Smirnov uyum iyiliği testine göre ($k = 2,5$, $x_m = 5,1$) parametreleri ile pareto dağılımına sahip olduğu Tablo 3’te gösterilmiştir.

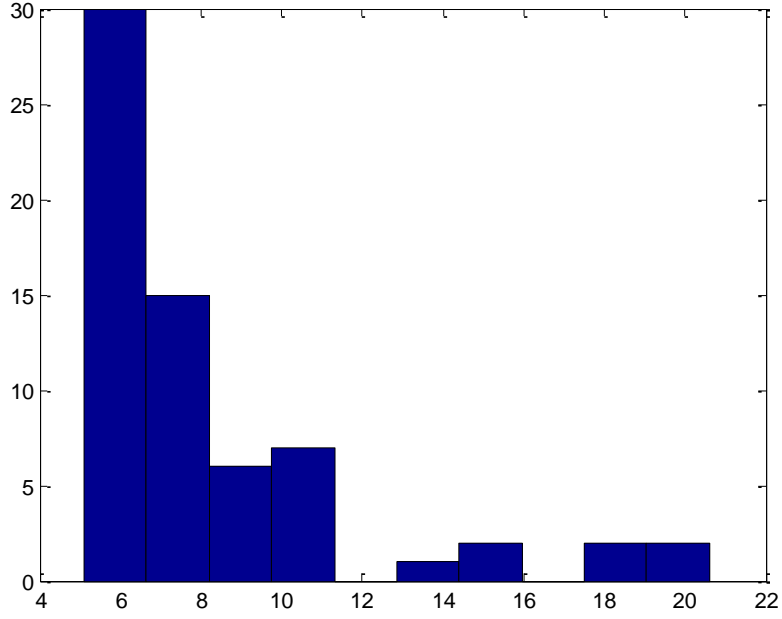
Tablo 3. Pareto dağılımına uyum

N	k	K-S (hesaplanan değer)	K-S (tablo)
66	2,5	0,0615	<0,2006

Tabloda hesaplanan K-S değeri ve $\alpha=0.01$ için K-S tablosundan elde edilen kritik değer yer almaktadır ve $K - S_{\text{Hesaplanan Değer}}(0,0869) < K - S_{\text{Tablo Değeri}} = (0,2006)$ olduğundan X_2 rasgele değişkeninin dağılımının Pareto dağılımından farkı yoktur biçiminde kurulan yokluk hipotezi kabul edilir.

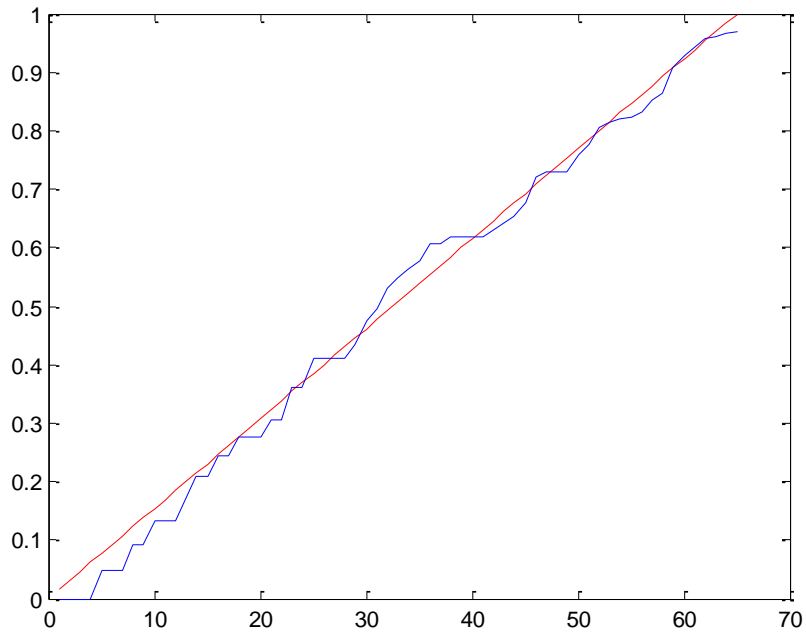
Tablo 4. Veri setine ilişkin tahminler ve hatalar

Gözlem No	X_1	X_2	Y	$\hat{Y}_{(\hat{O}A)i}$	$e_{(\hat{O}A)i}$	$\hat{Y}_{(EKK)i}$	$e_{(EKK)i}$
1	81.9200	7.5000	43.0870	39.9112	3.1758	46.7020	-3.6150
2	85.2200	6.6000	44.9820	40.2316	4.7504	47.9731	-2.9911
3	86.8000	8.5000	40.9500	39.7624	1.1876	44.8739	-3.9239
4	91.8200	7.5000	45.2670	40.0486	5.2184	46.2205	-0.9535
5	95.1800	5.4000	40.8660	41.1396	-0.2736	49.3975	-8.5315
6	97.3600	5.8000	32.9150	40.9058	-7.9908	48.6552	-15.7402
7	97.6800	8.6000	38.4840	39.9968	-1.5128	44.1857	-5.7017
8	97.4200	17.9000	44.1730	39.5798	4.5932	29.4048	14.7682
9	106.2800	7.7000	50.9860	40.5682	10.4178	45.1990	5.7870
10	103.5000	6.3000	44.4380	40.8835	3.5545	47.5612	-3.1232
11	108.9000	10.0000	20.3920	40.4085	-20.0165	41.4130	-21.0210
12	110.9200	7.5000	36.2030	40.9264	-4.7234	45.2915	-9.0885
13	111.2000	5.6000	43.9450	41.7929	2.1521	48.3002	-4.3552
14	112.5800	6.3000	53.3510	41.4744	11.8766	47.1196	6.2314
15	118.8600	7.4000	37.5110	41.6106	-4.0996	45.0644	-7.5534
16	117.2600	6.1000	27.0530	41.9673	-14.9143	47.2101	-20.1571
17	119.5000	9.8000	45.9110	41.2853	4.6257	41.2156	4.6954
18	118.5000	6.3000	52.7170	41.9745	10.7425	46.8317	5.8853
19	124.6600	6.3000	42.4380	42.5345	-0.0965	46.5321	-4.0941
20	125.9200	10.2000	47.9860	41.8333	6.1527	40.2671	7.7189
.
.
.
58	177.5400	5.4000	49.4370	41.8881	7.5489	45.3918	4.0452
59	181.2600	5.2000	34.7140	41.8281	-7.1141	45.5290	-10.8150
60	182.1400	5.3000	39.8840	41.6803	-1.7963	45.3272	-5.4432
61	188.5800	9.3000	38.3340	39.9679	-1.6339	38.6511	-0.3171
62	186.1800	19.8000	26.6650	39.7372	-13.0722	22.0655	4.5995
63	188.1800	5.2000	41.5530	41.4821	0.0709	45.1925	-3.6395
64	195.6800	14.7000	42.6410	39.4820	3.1590	29.7160	12.9250
65	196.0200	13.3000	44.7110	39.5087	5.2023	31.9265	12.7845
66	201.1400	11.3000	24.3580	39.5006	-15.1426	34.8589	-10.5009
Hata				$\epsilon_{\hat{O}A} = 61.5701$		$\epsilon_{EKK} = 106.5940$	



Şekil 11. X_2 değişkenine ilişkin histogram

Ayrıca Pareto dağılımına uygunluğa ilişkin grafik Şekil 12’de gösterilmektedir. Şekil 12’de kırmızı ile gözlenen birikimli dağılım fonksiyonu $S(x)$ ve mavi ile gösterilen birikimli dağılım fonksiyonu $F_0(x)$ ’dır.



Şekil 12. X_2 değişkeninin Pareto dağılımına uygunluğu

Verilere ilişkin çoklu doğrusal regresyon modelini elde etmeden önce Pareto dağılımı gösterdiği belirlenen X_2 değişkenine ilişkin üyelik dereceleri,

$$\mu(X_{2i}) = \frac{2(k+1)}{k} \left(\frac{X_{2m}}{X_{2i}} \right)^{k+1}$$

ile belirlenir, burada $k = 2,5$ ve $X_{2m} = 5,1$ 'dir. $[0,1]$ aralığında değerler alabilen c sabiti 0.4 olarak belirlenmiştir. Bu sabit, üyelik fonksiyonunun olasılık yoğunluk fonksiyonuna uyumunun kontrolü ile elde edilmiştir.

Önerilen algoritmadan çoklu doğrusal regresyon modeli;

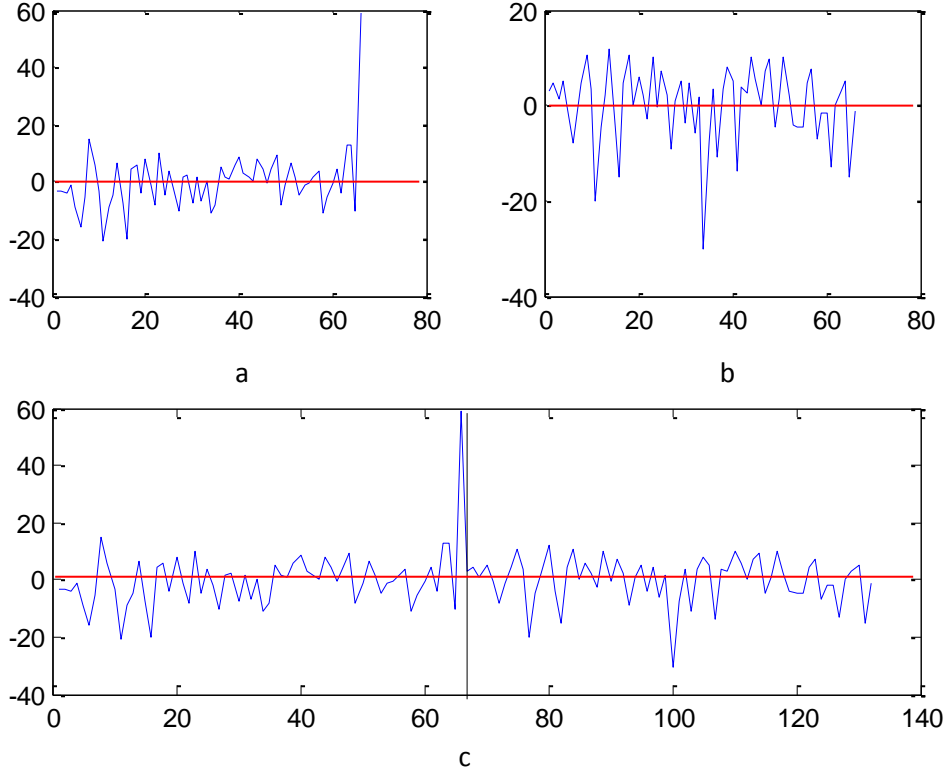
$$\hat{Y}_{\text{ÖA}} = 39.1644 + 0.0233x_1 + 64.7258x_2 \quad (24)$$

biçiminde ve En Küçük Kareler yöntemine ilişkin model;

$$\hat{Y}_{\text{EKK}} = 62.6165 - 0.0486x_1 - 1.5907x_2 \quad (25)$$

biçiminde elde edilmiştir.

Önerilen algoritmadan elde edilen modelin tahmin değerlerine ilişkin hata $\varepsilon_{\text{ÖA}} = 61.7501$, En Küçük Kareler yönteminden elde edilen modelin tahmin değerlerine ilişkin hata $\varepsilon_{\text{EKK}} = 106.5940$ olarak elde edilmiştir. Şekil 13-a En küçük Kareler modelinden elde edilen hata miktarlarını, Şekil 13-b önerilen algoritmadan elde edilen tahminlerin hata miktarlarını göstermektedir. Tüm modellerden elde edilen hatalar grafikleri karşılaştırmalı olarak Şekil 13-c'de yer almaktadır.



Şekil 13. Tablo 4’te yer alan veri setine ilişkin hata grafikleri

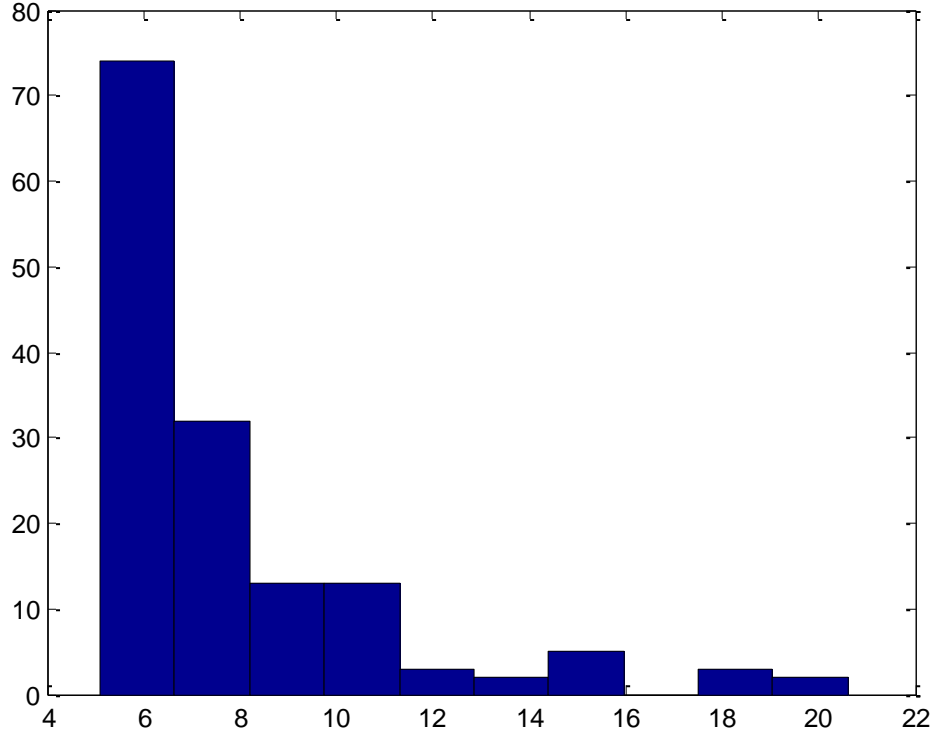
Önerilen algoritma ile elde edilen tahminlerin En Küçük Kareler yönteminden elde edilen modele ilişkin tahminlerden daha küçük hatalara sahip olduğu gözlenmiştir.

Örnek 3: Bu örnekte yer alan veri seti 149 gözlemden oluşmakta ve verilerin bir kısmı Tablo 5’te yer almaktadır. Veri setinde bir bağımlı ve üç bağımsız değişken bulunmaktadır. Bağımlı değişken Y , bağımsız değişkenler X_1 , X_2 ve X_3 ile ifade edilmiştir.

Tablo 5. Veri setine ilişkin tahminler ve hatalar

Gözlem No	X_1	X_2	X_3	Y	$\hat{Y}_{(\hat{O}A)i}$	$e_{(\hat{O}A)i}$	$\hat{Y}_{(EKK)i}$	$e_{(EKK)i}$
1	69.5800	1.6133	6.0000	31.7880	39.6349	-7.8469	45.6409	-13.8529
2	76.6400	1.5900	6.2000	27.9600	39.4771	-11.5171	45.0669	-17.1069
3	81.9200	2.1600	7.5000	43.0870	39.3869	3.7001	43.8577	-0.7707
4	80.9200	2.1900	7.6000	41.1750	39.4268	1.7482	43.7737	-2.5987
5	85.2200	2.1467	6.6000	44.9820	39.8287	5.1533	45.2810	-0.2990
6	88.6000	2.3867	7.8000	43.0630	40.1280	2.9350	43.6270	-0.5640
7	89.8800	2.4200	6.2000	42.1820	41.1277	1.0543	46.3927	-4.2107
8	86.8000	3.1467	8.5000	40.9500	40.6731	0.2769	44.0117	-3.0617
9	91.8200	2.3767	7.5000	45.2670	40.2464	5.0206	44.0345	1.2325
10	94.7200	3.0500	8.7000	41.5210	41.0775	0.4435	43.2670	-1.7460
11	93.2400	2.2733	6.2000	40.5280	40.6164	-0.0884	46.0086	-5.4806
12	95.1800	2.9967	5.4000	40.8660	43.1497	-2.2837	48.7825	-7.9165
13	98.4400	2.3633	7.5000	39.2390	40.3473	-1.1083	43.8335	-4.5945
14	97.3600	1.9100	5.8000	32.9150	40.2679	-7.3529	45.8507	-12.9357
15	101.3400	2.1567	7.9000	39.5440	39.5991	-0.0551	42.6576	-3.1136
16	100.5000	1.8733	6.0000	26.9140	40.1354	-13.2214	45.3526	-18.4386
17	97.6800	3.0433	8.6000	38.4840	41.1938	-2.7098	43.3464	-4.8624
18	97.4200	2.9000	17.9000	44.1730	40.8017	3.3713	27.1834	16.9896
19	105.4000	2.6833	6.6000	29.4420	42.2717	-12.8297	45.8328	-16.3908
20	105.1000	2.1067	5.4000	36.8190	41.3377	-4.5187	46.7270	-9.9080
21	106.2800	2.9333	7.7000	50.9860	41.9968	8.9892	44.4354	6.5506
22	105.4000	2.6700	5.7000	42.5640	42.9893	-0.4253	47.3428	-4.7788
23	103.5000	3.2400	6.3000	44.4380	41.6755	2.7625	47.5177	-3.0797
24	105.6800	2.8167	5.4000	53.7820	43.6402	10.1418	48.1436	5.6384
25	104.6800	3.1667	5.3000	42.5800	43.0708	-0.4908	49.0465	-6.4665
26	108.9000	2.5167	10.0000	20.3920	40.8239	-20.4319	39.5989	-19.2069
27	110.9200	2.3367	7.5000	36.2030	40.7849	-4.5819	43.4518	-7.2488
28	109.9400	2.8500	5.4000	49.7530	43.8709	5.8821	48.0988	1.6542
29	111.1200	2.7800	7.4000	35.1810	42.3541	-7.1731	44.5113	-9.3303
30	111.2000	2.7000	5.6000	43.9450	43.4826	0.4624	47.4216	-3.4766
.
.
.
142	188.1800	4.0633	5.2000	41.5530	40.9170	0.6360	48.8308	-7.2778
143	194.5600	2.7867	11.6000	52.6630	40.8762	11.7868	35.1598	17.5032
144	194.1800	2.6733	9.3000	45.3860	40.9183	4.4677	38.8686	6.5174
145	195.6800	1.7800	14.7000	42.6410	38.0316	4.6094	27.8065	14.8345
146	196.0200	2.9967	13.3000	44.7110	40.6469	4.0641	32.6420	12.0690
147	199.9400	3.1067	6.2000	41.6830	41.8662	-0.1832	44.8849	-3.2019
148	200.6000	2.5267	9.9000	41.1220	40.2164	0.9056	37.3797	3.7423
149	201.1400	1.7967	11.3000	24.3580	38.1484	-13.7904	33.5026	-9.1446
Hata					$\epsilon_{\hat{O}A} = 119.0887$		$\epsilon_{EKK} = 201.1761$	

Tablo 5 ayrıca önerilen algoritmadan ve EKK yönteminden elde edilen tahminleri ve tahminlere ilişkin hataları da içermektedir.



Şekil 14. X_3 değişkenine ilişkin histogram

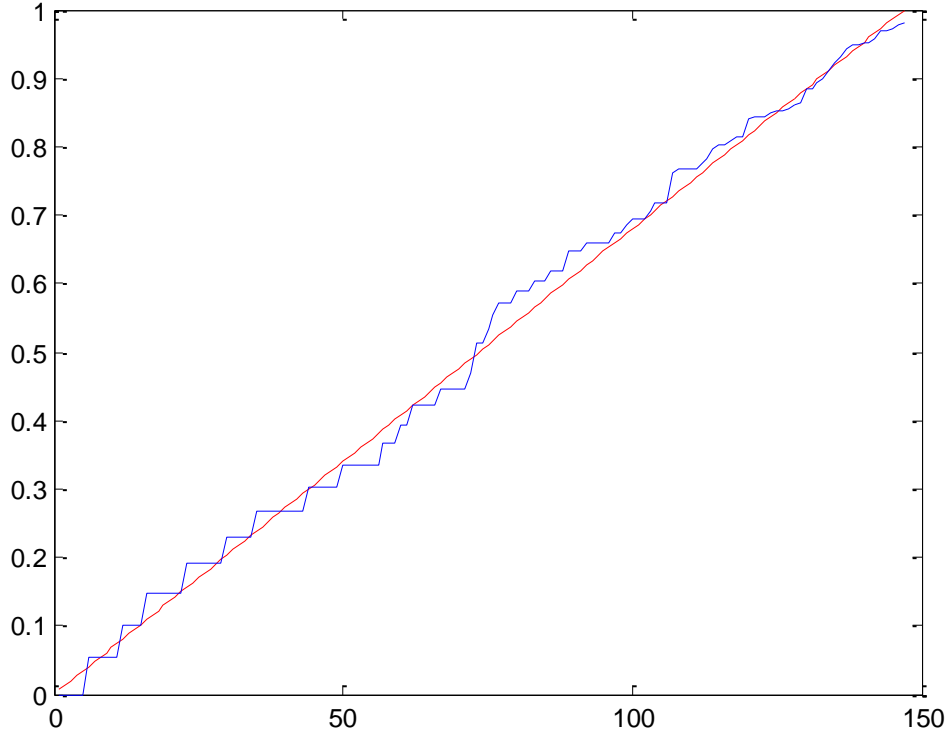
Bağımsız değişkenlerden X_3 'e ilişkin histogram Şekil 14 ile verilmiştir ve bu değişken Kolmogrov-Smirnov (K-S) uyum iyiliği testine göre ($k = 2,8$, $x_m = 5,2$) parametreleri ile Pareto dağılımına sahip olduğu Tablo 2'de gösterilmiştir. Tablo 6'da hesaplanan K-S değeri ve $\alpha = 0.01$ için K-S tablosundan elde edilen kritik değer yer almaktadır.

$$K - S_{\text{Hesaplanan Değer}} = (0,0472) < K - S_{\text{Tablo Değeri}} = (0,1335)$$

olduğundan X_3 rasgele değişkeninin dağılımının Pareto dağılımından farkı yoktur biçiminde kurulan yokluk hipotezi kabul edilir. Ayrıca X_3 rasgele değişkeninin Pareto dağılımına uygunluğa ilişkin grafik Şekil 15'te gösterilmektedir. Şekil 15'te kırmızı ile gözlenen birikimli dağılım fonksiyonu $S(x)$ ve mavi ile gösterilen birikimli dağılım fonksiyonu $F_0(x)$ 'dir.

Tablo 6. Pareto dağılımına uyum

n	k	K-S (hesaplanan değer)	K-S (tablo)
149	2,8	0,0472	<0,1335

Şekil 15. X_3 değişkeninin Pareto dağılımına uygunluğu

Verilere ilişkin çoklu doğrusal regresyon modelini elde etmeden önce Pareto dağılımı gösterdiği belirlenen X_3 değişkenine ilişkin üyelik dereceleri,

$$\mu(X_{3i}) = \frac{2(k+1)}{k} \left(\frac{X_{3m}}{X_{3i}} \right)^{k+1}$$

ile belirlenir, burada $k = 2,8$ ve $X_{3m} = 5,2$ dir. Ayrıca, $[0,1]$ aralığında değerler alabilen c sabiti 0.4 olarak belirlenmiştir. Bu sabit, üyelik fonksiyonunun olasılık yoğunluk fonksiyonuna uyumunun kontrolü ile elde edilmiştir.

Önerilen algoritma için çoklu doğrusal regresyon modeline ilişkin parametre değerleri, X matrisinde X_3 rasgele değişkeni üyelik fonksiyonundan elde edilen üyelik dereceleri ile ağırlıklandırılmış olmak üzere;

$$B = (X^T X)^{-1} X^T Y$$

eşitliğinden ve MATLAB da yazılan ve önerilen algoritmanın işleyişini sağlayan program ile;

$$\hat{Y}_{\text{öA}} = 37.6244 + 0.01785X_1 + 0.9888X_2 + 100.9175X_3 \quad (26)$$

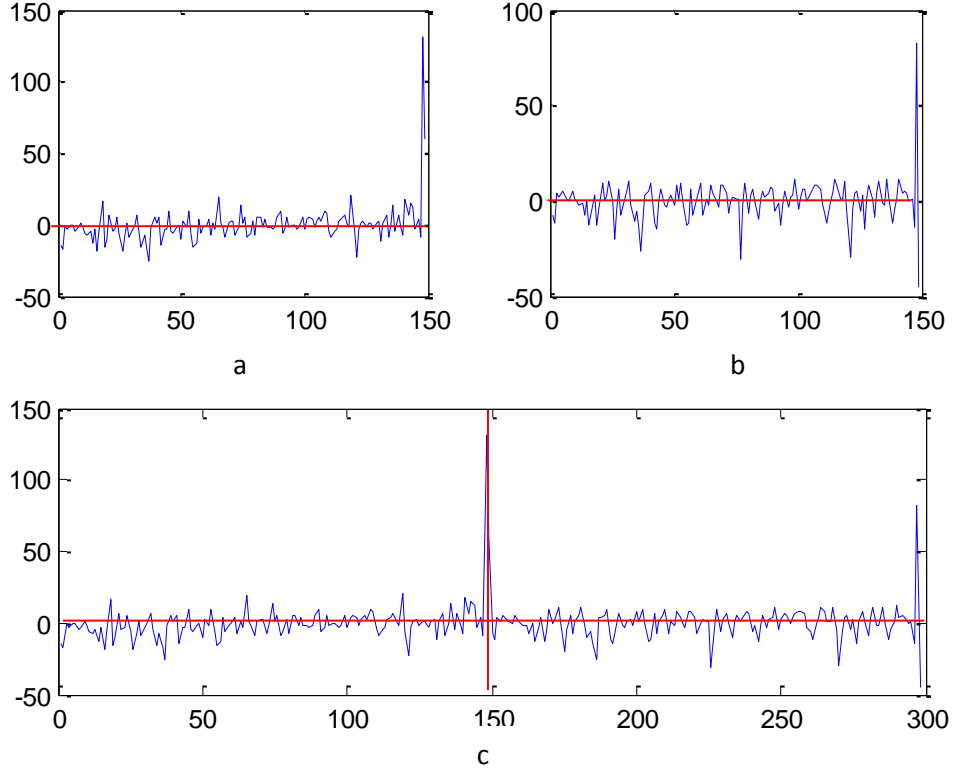
biçiminde elde edilmiştir. En Küçük Kareler yöntemine ilişkin model ise;

$$\hat{Y}_{\text{EKK}} = 54.4619 - 0.0263x_1 + 2.0167x_2 - 1.7076x_3 \quad (27)$$

biçimindedir.

Önerilen algoritmadan elde edilen modelin tahmin değerlerine ilişkin hata $\varepsilon_{\text{öA}} = 119.0887$, En Küçük Kareler yönteminden elde edilen modelin tahmin değerlerine ilişkin hata $\varepsilon_{\text{EKK}} = 201.1761$ olarak elde edilmiştir. Şekil 16-a En küçük Kareler modelinden elde edilen hata miktarlarını, Şekil 16-b önerilen algoritmadan elde edilen tahminlerin hata miktarlarını göstermektedir. Tüm modellerden elde edilen hatalar grafikleri karşılaştırmalı olarak Şekil 16-c'de yer almaktadır.

Önerilen algoritma ile elde edilen tahminlerin En Küçük Kareler yönteminden elde edilen modele ilişkin tahminlerden daha küçük hatalara sahip olduğu gözlenmiştir.



Şekil 16. Tablo 5’te yer alan veri setine ilişkin hata grafikleri

3. SONUÇLAR VE ÖNERİLER

Çoklu doğrusal regresyon analizinde kurulacak modele ilişkin bağımsız değişkenlerden herhangi biri Pareto dağılıma sahip olduğunda ve veri setinde aykırı gözlemlerin var olması durumunda literatürde yer alan ve pek çok varsayım gerektiren klasik yöntemlerin dışına çıkılması gerektiği belirtilmişti. Bu çalışmada çoklu doğrusal regresyon modeline ilişkin parametre tahmini sürecinde, Pareto dağılımı gösteren bağımsız değişkenin ağırlıklandırılmasında kullanılmak üzere, Pareto dağılımına uygun, dağılımın olasılık yoğunluk fonksiyonuna dayalı üyelik fonksiyonu elde edilmiştir. Çoklu doğrusal regresyon modelinin bilinmeyen parametreleri önerilen yöntem ve literatürde yer alan diğer yöntemler ile tahmin edilerek, belirlenerek modellerin performansları verdikleri tahmin değerleri ve bu değerlere ilişkin hata miktarları ile karşılaştırılmıştır. Elde edilen tahminlere ilişkin hata miktarlarının düşük olması, önerilen yöntemin, karşılaşılabilecek çoklu doğrusal regresyon modeli kurma problemlerinde alternatif olarak kullanılabilir olduğunu göstermektedir. Bağımsız değişkenlerden herhangi birinin Pareto dağılımı dışında başka dağılımlara sahip olmaları durumu irdelenerek bu durumlara ilişkin yöntemler geliştirilebilir.

4. KAYNAKLAR

- Alpar, R., 2003. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1, Nobel Yayın Dağıtım, Ankara.
- Apaydın, A., Kutsal, A. ve Atakan, C., 1994. Uygulamalı İstatistik, 496s., Ankara.
- Bargiela, A., Pedrycz, W. and Nakashima, T., 2007. Multiple Regression with Fuzzy Data, Fuzzy Sets and Systems 158, 2169-2188.
- Baykal, N. ve Beyan T., 2004. Bulanık Mantık İlke ve Temelleri, Bıçaklar Kitabevi, Ankara.
- Baykal, N. ve Beyan T., 2004. Bulanık Mantık Uzman Sistemler ve Denetleyiciler, Bıçaklar Kitabevi, Ankara.
- Bezdek, J., C., Ehrlich, R. ve Full, W. 1984. Fcm: The Fuzzy c-Means Clustering Algorithm, Computers&Geosciences, 10, 191-203.
- Chen, L. H., Hsueh, C. C. and Chang, C. J., 2013. A Two-Stage Approach for Formulating Fuzzy Regression Models, Knowledge-Based Systems, 52, 302-310.
- Chen, M.S. and Wang, S.W., 1999. Fuzzy Clustering Analysis for Optimizing Fuzzy Membership Functions, Fuzzy Sets and Systems, 103, 239-254.
- Chi-Bin, C. and Lee, E. S., 2001. Switching Regression Analysis by Fuzzy Adaptive Network, European Journal of Operational Research, 128, 647-663.
- Chi-Bin, C., and Lee, E. S., 1998. Applying Fuzzy Adaptive Network to Fuzzy Regression Analysis, An International Journal Computers & Mathematics With Applications, 38, 123-140.
- Chuang, C., Jeng, J. and Tao, C., 2009. Hybrid Robust Approach for TSK Fuzzy Modeling with Outliers, Expert Systems with Applications, 36, 8925-8931.
- Civanlar, M.R. and Trussell, H.J., 1986. Tructing Membership Functions Using Statistical Data, Fuzzy Sets and Systems, 18, 1-13.
- Çolak, T., 2007. İstatistiksel Süreç Kontrolü ve Uygulamalar. Yüksek Lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana.
- D'Urso, P., 2003. Linear Regression Analysis For Fuzzy/Crisp Input and Fuzzy/Crisp Output Data, Computational Statistics & Data Analysis, 42, 47-72.
- Dombi, J., 1990. Membership Functions As An Evaluation, Fuzzy Sets and Systems, 35, 1-21.
- Elmas, Ç., 2003. Bulanık Mantık Denetleyicileri, Seçkin Yayıncılık, Ankara.

- Erbay Dalkılıç, T., 2005. Switching Regresyon'da Bulanık Sinir Ağları Yaklaşımı ile Parametre Tahmini, Doktora Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Gamgam, H. ve Altunkaynak, B., 2012. SPSS Uygulamalı Parametrik Olmayan Yöntemler, Seçkin Yayıncılık, Ankara.
- Hanci, B. Y., 2005. GSM 900 MHZ Bandında Deneysel Verilere Dayalı Yeni Bir Kentsel Yayılım Modelinin Geliştirilmesi, Doktora Tezi, Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Trabzon.
- Hojati, M., Bector, C. R. and Smimou, K., 2005. A Simple Method for Computation of Fuzzy Linear Regression, European Journal of Operational Research 166, 172-184.
- Ishibuchi, H. and Tanaka, H., 1992. Fuzzy Regression Analysis Using Neural Networks, Fuzzy Sets and Systems, 50, 257-265.
- Jang, J.R. ve Sun, C.S., 1995. Neuro-Fuzzy Modeling and Control, Proceedings of The IEEE, 83, 3, 378-406.
- Klir, G.J. and Yuan, B., 1995. Fuzzy Sets and Fuzzy Logic, Prentice-Hall, USA.
- Köseoğlu, M. ve Yamak, 2004. R., Uygulamalı İstatistik ve Ekonometri, Celepler Matbaacılık, Trabzon.
- Modarres, M., Nasrabadi, E. and Nasrabadi M. M., 2005. Fuzzy Linear Regression Models with Least Square Errors, Applied Mathematics and Computation, 163, 977-989.
- Mosleh, M., Otadi, M. and Abbasbandy, S., 2010. Evaluation of Fuzzy Regression Models by Fuzzy Neural Network, Journal of Computational and Applied Mathematics, 234, 825-834.
- Pierpaolo, D., Riccardo M. and Adriana S., 2011. Robust fuzzy regression analysis, Information Sciences, 181, 19, 4154-4174.
- Şıklar, E., 2000. Regresyon Analizine Giriş, Anadolu Üniversitesi Yayınları, Eskişehir.
- Takagi, T. and Sugeno, M., 1985. Fuzzy Identification of Systems and Its Applications to Modeling and Control, IEEE Trans. On Systems, Man and Cybernetics, 15, 1, 116-132.
- URL-1, http://en.wikipedia.org/wiki/Pareto_distribution, 5 Mart 2013.
- Ünver, Ö. ve Gamgam, H., 1999. Uygulamalı İstatistik Yöntemler, Siyasal Kitabevi, Ankara.
- Walfisch, J., Bertoni, H. L., 1988. A Theoretical Model of UHF Propagation in Urban Environments, IEEE Transaction On Antennas and Propagation, 36, 12, 1788-1796
- Zadeh, L. A., 1965. Fuzzy Sets, Information and Control 8, 338-353.

ÖZGEÇMİŞ

Tuğba KARAN, 23 Haziran 1985 tarihinde Rize’de doğdu. İlköğrenimini Taşlıdere İlkokulu’nda, ortaöğrenimini 29 Ekim İlköğretim Okulu’nda, lise öğrenimini Rize Fener Lisesi’nde tamamladı. 2004–2005 öğretim yılında Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü’ne yerleşti. 2008 yılında mezun olduktan sonra 2009 yılında Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri bölümde tezli yüksek lisans programına dahil oldu. 2013 yılında Türkiye İstatistik Kurumu’nda İstatistikçi olarak çalışmaya başladı ve halen bu görevine devam etmektedir.