

ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

BULANIK ROBUST REGRESYON ÇÖZÜMLEMESİ

120325

Kamile ŞANLI

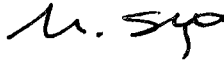
İSTATİSTİK ANABİLİM DALI

ANKARA

2005

Her hakkı saklıdır

Prof. Dr. Ayşen APAYDIN danışmanlığında, Kamile ŞANLI` tarafından hazırlanan bu çalışma 27/01/2005 tarihinde aşağıdaki jüri tarafından İstatistik Anabilim Dalı`nda Doktora tezi olarak kabul edilmiştir.

Başkan : Doç. Dr. Meral SUCU 

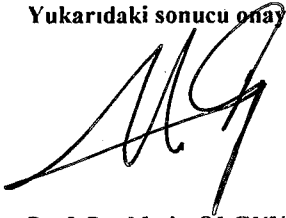
Üye : Prof. Dr. Ömer L. GEBİZLİOĞLU 

Üye : Prof. Dr. Fikri ÖZTÜRK 

Üye : Doç. Dr. Fazıl ALİOĞLU 

Üye : Prof. Dr. Ayşen APAYDIN 

Yukarıdaki sonucu onaylarım



Prof. Dr. Metin OLGUN

Enstitü Müdürü

ÖZET

Doktora Tezi

BULANIK ROBUST REGRESYON ÇÖZÜMLEMESİ

Kamile ŞANLI

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Ayşen APAYDIN

Regresyon çözümlemesinde veri analizi oldukça önemlidir. Çünkü, tek bir gözlem bile regresyon modelindeki parametre tahminleri üzerinde büyük bir etkiye sahip olabilir. Bu gözlemin veri kümesinden çıkartılması regresyon denklemini tamamen değiştirebilir. Veri kümesinde aykırı değer olması durumunda, parametre tahminlerinde robust yöntemler olarak bilinen Huber, Hampel, Andrews ve Tukey'in M yöntemleri kullanılmaktadır. Bu çalışmada girdi değişkenlerinin simetrik olmayan üçgensel bulanık sayı $(X_i = (x_i, \underline{\xi}_i, \bar{\xi}_i))$, $Y_i = (y_i, \underline{\eta}_i, \bar{\eta}_i)$; x 'in kesin sayı, $Y_i = (y_i, \underline{\eta}_i, \bar{\eta}_i)$ simetrik üçgensel bulanık sayı olması ve veri kümesinde aykırı değer olması durumunda artıklara ilişkin üyelik fonksiyonu yardımıyla ağırlık matrisi tanımlanmıştır. Regresyon model tahmininde ise bulanık regresyon çözümlemesi kullanılmış ve tanımlanan ağırlık matrisinin kullanıldığı yeni iki yöntem önerilmiştir. Klasik en küçük kareler (EKK), Huber, Hampel, Andrews ve Tukey M yöntemleri ve önerilen bulanık robust yöntemler ile regresyon model tahminleri elde edilmiş ve sonuçlar karşılaştırılmıştır.

2005, 107 sayfa

ANAHTAR KELİMELER: Aykırı değer, robust regresyon, üyelik fonksiyonu, bulanık sayı, bulanık regresyon

ABSTRACT

Ph. D. Thesis

FUZZY ROBUST REGRESSION ANALYSIS

Kamile ŞANLI

**Ankara University
Graduate Scholl of Natural and Applied Sciences
Department of Statistics**

Supervisor: Prof. Dr. Ayşen APAYDIN

Data analysis occupies an important place in regression models. Because, even one observed data may cause a large effect over parameter estimates in regression models. Omitting such data can completely change the model structure. Huber, Hampel, Andrews and Tukey M methods, called robust methods, are used in parameter estimation in the presence of outliers. Along these lines, in this thesis weighted matrices are defined with respect to membership function when $X_i = (x_i, \underline{\xi}_i, \bar{\xi}_i)$, $Y_i = (y_i, \underline{\eta}_i, \bar{\eta}_i)$, x is crisp, Y is fuzzy data, and data set has outliers. For the regression model estimation, fuzzy regression analysis is used and two new algorithms are suggested which define weighted matrices to be used. Regression model estimates are obtained by least square method (LSM), Huber, Hampel, Andrews and Tukey M methods and suggested fuzzy robust methods. Comparisons of the result are presented.

2005, 107 pages

Key Words: Outlier, robust regression, membership function, fuzzy number, fuzzy regression

TEŞEKKÜR

Bu tez konusunu bana veren ve çalışmanın her safhasında yakın ilgi, önerileri ile beni yönlendiren ve destekleyen danışman hocam, Sayın Prof. Dr. Ayşen APAYDIN (Ankara Üniversitesi, Fen Fakültesi)’a sonsuz teşekkürlerimi sunarım.

Çalışmanın her aşamasında yakın ilgileriyle beni yönlendiren Sayın Prof. Dr. Fikri ÖZTÜRK (Ankara Üniversitesi Fen Fakültesi)’e, Sayın Doç. Dr. Fazıl ALİOĞLU (Ankara Üniversitesi Fen Fakültesi)’na ve Sayın Doç. Dr. Meral SUCU (Hacettepe Üniversitesi Fen Fakültesi)’ya en içten teşekkürlerimi sunarım.

Güven, anlayış ve desteklerini daima hissettiğim sevgili babam Ömer ŞANLI, annem Melahat ŞANLI, abim Cemali ŞANLI, ablam Nazmiye ŞANLI KOPTUR ve kardeşim Yusuf ŞANLI’ya sonsuz teşekkürlerimi sunarım. Küçük yeğenim Ömerilkkan KOPTUR’a sevgiyle.

Kamile ŞANLI

ANKARA, Ocak 2005

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
SİMGELER DİZİNİ.....	vi
ŞEKİLLER DİZİNİ.....	vii
GRAFİKLER DİZİNİ.....	viii
ÇİZELGELER DİZİNİ.....	ix
1. GİRİŞ VE ÖNCEKİ ÇALIŞMALAR.....	1
1.1. Giriş.....	1
1.2. Önceki Çalışmalar.....	4
2. BULANIK KÜME TEORİSİ.....	8
2.1. Giriş.....	8
2.2. Bulanık Küme Teorisi İle İlgili Temel Kavramlar.....	9
2.3. Üyelik Fonksiyonlarının Sınıflandırılması.....	17
2.4. Bulanık Kümelerdeki Operatörler.....	22
2.5. Bulanık Kümelerdeki Aritmetik İşlemler.....	23
2.6. Bulanık Kümelerin Karşılaştırılması.....	25
3. KÜMELEME YÖNTEMLERİ.....	29
3.1. Katı ve Bulanık Parçalanma.....	30
3.2. Katı c Ortalamaya Dayalı Kümeleme.....	31
3.3. Bulanık c Ortalamaya Dayalı Kümeleme.....	32
3.4. Robust Kümeleme Yöntemleri.....	33
3.4.1 Gürültü Kümeleme Yöntemi.....	34
3.4.2. Olanaklı c Ortalamaya Dayalı Kümeleme Yöntemi.....	36
3.4.3. Karma Bulanık Olanaklı c Ortalamaya Dayalı Kümeleme.....	37
4. REGRESYON ÇÖZÜMLEMESİ.....	40
4.1. Giriş.....	40
4.2. Robust Regresyon Çözümlemesi.....	42

4.2.1. M Tahmin Edicileri.....	43
4.2.2. En Küçük Medyan Kare Yöntemi.....	50
4.2.3. En Küçük Medyan Kareye Dayanan Yeniden Ağırlıklandırılmış En Küçük Kareler Yöntemi.....	51
4.3. Bulanık Regresyon Çözümlemesi.....	52
5. BULANIK ROBUST REGRESYON ÇÖZÜMLEMESİ.....	59
5.1. $X = (x, \underline{x}, \bar{x})$ ve $Y = (y, \underline{y}, \bar{y})$ 'nin Simetrik Olmayan Üçgensel Bulanık Sayılar Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi	61
5.1.1. $X = (x, \underline{x}, \bar{x})$ ve $Y = (y, \underline{y}, \bar{y})$ Simetrik Olmayan Üçgensel Bulanık Sayılar Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi için Algoritma.....	65
5.2. x 'in Kesin Sayı ve $Y = (y, \underline{y}, \bar{y})$ 'nin Simetrik Üçgensel Bulanık Sayı Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi.....	67
5.2.1. Gözlemlenmiş $Y_i = (y_i, \underline{y}_i, \bar{y}_i)$ ve tahmin edilen $\hat{Y}_i = (\hat{y}_i, \hat{\underline{y}}_i, \hat{\bar{y}}_i)$ 'nin simetrik üçgensel bulanık sayı olması durumunda karşılaştırma.....	70
5.2.2. x 'in Kesin Sayı ve $Y = (y, \underline{y}, \bar{y})$ 'nin Simetrik Üçgensel Bulanık Sayı Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi için Algoritma.....	74
6. UYGULAMA.....	77
6.1. Giriş.....	77
6.2. Bağımlı Değişkende Aykırı Değer Olması Durumu için Uygulama.....	77
6.3. Bağımsız Değişkende Aykırı Değer Olması Durumu için Uygulama.....	89
7. SONUÇ ve TARTIŞMA.....	100
KAYNAKLAR.....	104
ÖZGEÇMİŞ.....	108

SİMGELER DİZİNİ

EKK	En Küçük Kareler
FCM	Bulanık c Ortalamaya Dayalı Kümeleme
NC	Gürültü Kümeleme
PCM	Olanaklı c Ortalamaya Dayalı Kümeleme
FPCM	Karma Bulanık c Ortalamaya Dayalı Kümeleme
LMS	En Küçük Medyan Kare
RLS	LMS'ye Dayanan Yeniden Ağırlıklandırılmış EKK

ŞEKİLLER DİZİNİ

Şekil 2.1. Bulanık küme ve kesin küme üyeliklerinin grafiksel gösterimi.....	13
Şekil 2.2. Üçgensel bulanık sayı.....	14
Şekil 2.3. Yamuksal bulanık sayı.....	15
Şekil 2.4. L - R bulanık sayısı.....	16
Şekil 2.5. İki bulanık kümenin karşılaştırılması.....	25
Şekil 4.1. Huber ρ fonksiyonu.....	46
Şekil 4.2. Huber ψ fonksiyonu.....	46
Şekil 4.3. Hampel ρ fonksiyonu.....	47
Şekil 4.4. Hampel ψ fonksiyonu.....	47
Şekil 4.5. Andrews ρ fonksiyonu.....	48
Şekil 4.6. Andrews ψ fonksiyonu.....	48
Şekil 4.7. Tukey ρ fonksiyonu.....	49
Şekil 4.8. Tukey ψ fonksiyonu.....	49
Şekil 4.9. Gözlenmiş $\tilde{Y}_i$ ve tahmin edilen $\hat{Y}_i$ arasındaki ilişki.....	58
Şekil 5.1. Y_i ve $\hat{Y}_i$ 'nin simetrik üçgensel bulanık sayı olması durumunda grafiksel gösterim.....	71

GRAFİKLER DİZİNİ

Grafik 6.1. Bağımlı değişkende aykırı değer olması durumuna ilişkin EKK, Huber, Hampel, Tukey ve Andrews yöntemleri için artıkların dağılımı.....	82
Grafik 6.2. Bağımlı değişkende aykırı değer olması durumuna ilişkin önerilen iki yöntem için artıkların dağılımı.	83
Grafik 6.3. Bağımsız değişkende aykırı değer olması durumuna ilişkin EKK, Huber, Hampel, Tukey ve Andrews yöntemleri için artıkların dağılımı.	93
Grafik 6.4. Bağımsız değişkende aykırı değer olması durumuna ilişkin önerilen iki yöntem için artıkların dağılımı.....	94

ÇİZELGELER DİZİNİ

Çizelge 6.1. Bağımlı değişkende aykırı değer olması durumu için veri kümesi.....	79
Çizelge 6.2. Bağımlı değişkende aykırı değer olması durumu için artıklar ve bunlara karşılık gelen ağırlık matrisinin köşegen elemanları.....	80
Çizelge 6.3. Bağımlı değişkende aykırı değer olması durumu için regresyon model tahminleri.....	81
Çizelge 6.4. 30 Gözlem, 3 bağımsız değişkeni içeren ve bağımlı değişkende aykırı değer olması durumu için çözümleme sonuçları.....	84
Çizelge 6.5. Bağımsız değişkende aykırı değer olması durumu için veri kümesi.....	90
Çizelge 6.6. Bağımsız değişkende aykırı değer olması durumu için artıklar ve bunlara karşılık gelen ağırlık matrisinin köşegen elemanları.....	91
Çizelge 6.7. Bağımsız değişkende aykırı değer olması durumu için regresyon model tahminleri.....	92
Çizelge 6.8. 30 Gözlem, 3 Bağımsız değişkeni içeren ve bağımsız değişkende aykırı değer olması durumu için çözümleme sonuçları.....	95

1. GİRİŞ VE ÖNCEKİ ÇALIŞMALAR

1.1. Giriş

Klasik en küçük kareler (EKK) tahmin edicisi hesaplanmasının kolay ve geleneksel olması nedeniyle regresyon çözümlemesinde geniş olarak kullanılmaktadır. Ancak veri kümesi sıklıkla farklı türlerdeki hataları içerir. Bunlar aletlerdeki yetersizlikler, kayıp veri, vb. nedenlerden dolayı olabilir. EKK tahmin edicisi hesaplanmasının kolay ve geleneksel olmasına rağmen aykırı değerlerden oldukça etkilenir ve bu durumda büyük sapmalara neden olur (<http://www.xplore-stat.de>, Stephenson 2000). Bu nedenle test istatistiklerinin ve katsayıların belirlenmesinde her bir gözlemin rolüne önem verilmeli ve veri ayrıntılı bir şekilde test edilmelidir. Çünkü elde edilen parametre tahminleri sadece bir gözleme bile bağlı olabilir ve bu gözlemin veriden çıkartılması analiz sonucunu ciddi bir şekilde değiştirebilir. Diğerlerine göre büyük artık değere sahip bu gözlemlere aykırı değer (outlier) denir (Alpar 1997). Aykırı değer olması durumunda regresyon model tahmininde EKK yöntemine göre daha az etkilenen robust yöntemler kullanılır .

Tahmin problemlerinde verideki aykırı değerın etkisini azaltma özelliğine sahip yöntemler çok önemli ve yaygındır. Bu yöntemlerden en çok kullanılanları robust yöntemler olarak bilinir. Robust yöntemlerde farklı özelliklere sahip tahmin ediciler üzerine oldukça çok çalışma vardır. Bu çalışmalar Huber tahmin edicisiyle yaygınlaşmaya başlamış ve daha sonra hızla artış göstermiştir. Son yıllarda yapılan çalışmalarda ise robust yöntemlerde bulanık teori uygulamaları artmıştır. Olabilirlik teorisindeki olabilirlik dağılımları ya da bulanık küme teorisindeki üyelik fonksiyonları kavramı robust istatistiklerdeki ağırlık fonksiyonu kavramıyla anlatılabilir ve pek çok ortak noktaya sahiplerdir. Robust istatistikler ve bulanık küme teorisi son 20 yılda bağımsız olarak yavaş yavaş gelişen iki disiplindir. Tanaka vd (1982) regresyon çözümlemesinde bulanık

teoriyi kullanmışlardır. Sakawa ve Yano (1992) etkileşimli, çok amaçlı programlama problemini formüle etmişlerdir. Watada ve Yabuuchi (1994) bulanık regresyon ve robustluğu birlikte ele alan regresyon modelini önermişlerdir. Klasik küme teorisinde bir eleman “1” üyeliği ile ya kümeye aittir ya da “0” üyeliğiyle kümeye ait değildir. Bulanık küme teorisinde ise bir elemanın kümeye ait olması $[0,1]$ aralığında değerler alır. Aykırı değerlerin kesin reddi klasik küme teorisindeki katı üyeliğe yani “0” üyeliğiyle kümeye ait olmamaya karşılık gelirken, aykırı değerlerin yumuşak reddi bulanık üyeliklere karşılık gelir. Üyelik fonksiyonlarını kullanan bulanık yaklaşımlar şüpheli veriyi daha iyi modelleyebilir (Dave ve Krishnapuram 1997).

Son yıllarda yapılan çalışmalarda robust yöntemler ve bulanık teori birlikte ele alınmıştır. Ancak bu çalışmalarda genellikle basit doğrusal regresyon incelenmiştir. Bulanık doğrusal regresyonda tanımlanan yinelemeli süreçlerde, model parametrelerinin kesin sayı, merkezlerin sıfır olması, aykırı değerlere karşı çok duyarlı, bilgi kaybı ve sınırsız çözüm gibi bazı sakıncalara rastlanmıştır. Bulanık teori ve robustluğun birlikte ele alındığı çalışmalar ise genellikle karar vericiyle etkileşime dayanan yöntemlerdir. Ancak bu yöntemlerde kullanılan etkileşimli algoritmaların aykırı değerlere karşı oldukça duyarlı olduğu ve bunun sonucunda büyük hataya sahip tahminlere ulaşıldığı görülmüştür (Redden ve Woddall 1996). Bu nedenle, bu çalışmada her bir aykırı değerın modele katkısı üyelik dereceleri ile belirlenmiş sezgiselliğe ve etkileşime izin vermeyen, bağımlı değişkenin ve bağımsız değişkenlerin simetrik olmayan üçgensel bulanık sayı, çıktının kesin sayı olması ve bağımlı değişkenin simetrik üçgensel bulanık sayı, bağımsız değişkenlerin kesin, çıktının simetrik üçgensel bulanık sayı olması durumları ele alınmıştır. Bu iki farklı durum için bulanık basit doğrusal regresyon modeli çoklu doğrusal regresyon modeli olarak genelleştirilmiştir. Tanımlanan üyelik fonksiyonu yardımıyla gözlemlerin belirli üyeliklerle modele alınması amaçlanmıştır. Veri kümesinde aykırı değer olması ve verinin bulanık olması durumunda süreci, çoklu doğrusal regresyon modeline genelleştirmek ve

aykırı değer dışındaki gözlemlerin hatalarını minimum yapmak bu çalışmanın amacıdır.

Çalışmanın İkinci Bölümü'nde, bulanık küme teorisiyle ilgili temel kavramlar ve tanımlar verilecektir.

Üçüncü Bölüm'de kümeleme yöntemleri ve robust kümeleme yöntemleri hakkında bazı temel kavramlar ve yöntemler açıklanacaktır.

Dördüncü Bölüm'de klasik regresyon çözümlemesi, bulanık regresyon çözümlemesi ve robust yöntemler açıklanacaktır.

Çalışmanın özgün yanını oluşturan Beşinci Bölüm'de, çoklu regresyon çözümlemesinde x 'in kesin Y 'nin simetrik üçgensel bulanık sayı, girdi değişkenlerinin simetrik olmayan üçgensel bulanık sayı $(X_i = (x_i, \underline{\xi}_i, \bar{\xi}_i), Y_i = (y_i, \underline{\eta}_i, \bar{\eta}_i))$ ve veri kümesinde aykırı değer olması durumunda, bulanık basit doğrusal regresyon çözümlemesi bulanık çoklu doğrusal regresyon çözümlemesine genelleştirilecektir. Üyelik fonksiyonu yardımıyla ağırlık matrisi tanımlanacaktır. Regresyon model tahmininde ise bulanık regresyon çözümlemesi kullanılacak ve tanımlanan ağırlık matrisinin kullandığı yeni iki yöntem önerilecektir. Regresyon model tahmininde yeni iki yöntem ile parametre tahminleri yapılacak ve bulunan sonuçlar en küçük kareler yöntemi ve klasik robust yöntemlerle karşılaştırılacaktır.

1. 2. Önceki Çalışmalar

Bulanık regresyon çözümlemesi ve veri kümesinde aykırı değer olması durumuyla ilgili yapılan çalışmalar aşağıdaki gibi özetlenebilir:

Tanaka vd (1982) bulanık doğrusal regresyon çözümlemesindeki ilk çalışmayı yapmışlardır. Bu yaklaşım Bardossy (1990) tarafından geliştirilmiştir. Optimizasyon problemini çözmek için kullanılan Tanaka (1987) yaklaşımı çok karmaşıktır. En küçük kareler yaklaşımı açık değildir ve artıklar açısından en iyi uygunluğun ölçümü tanımlanmamıştır. Tanaka modelinin bu sakıncalarını gidermek için Dimond (1988) bulanık en küçük kareler yaklaşımını önermiştir (Yang ve Ko 1997).

Tanaka (1987) modeli sıklıkla kesin sayılar ürettiği için Celmins (1987) tarafından eleştirilmiştir. Daha sonra Tanaka vd (1982) tarafından ileri sürülen en küçükleme probleminin ölçek bağımlı olduğunu Jo'zsef (1992) ileri sürmüştür. Celmins (1987)'in eleştirisine karşılık Tanaka ve Ishibuchi (1991) etkileşimli bulanık katsayıları elde etmek için bir yöntem geliştirmişlerdir (Redden ve Woddall 1996). Bu yöntemin aykırı değerlere güçlü bir şekilde duyarlı olduğu, verinin içerdiği kesin bilgiyi göz ardı ettiği ve sınırsız çözüm ürettiği Redden ve Woddall (1996) tarafından gösterilmiştir.

Sakawa ve Yano (1992) çok amaçlı programlama problemlerini formüle etmek için karar vericiyi tatmin edecek şekilde yinelemeli, karar vericiyle etkileşime dayanan, etkileşimli algoritmayı ileri sürmüşlerdir. Fakat bu algoritma da aykırı değerlere karşı çok fazla duyarlıdır. Redden ve Woddall (1996) ortogonal en küçük kareler yardımıyla Sakawa ve Yano'nun modelini geliştirmişler ve doğrusal programlama problemini önermişlerdir.

Watada ve Yabuuchi (1994), aykırı değer tarafından etkilenmeyen bir bulanık robust regresyon modeli önermişlerdir. Önerilen model, toplam hatayı mümkün olduğunca minimum yapacak şekilde kurulmuştur. Optimizasyon problemini çözmek için genetik algoritma kullanılmıştır.

Watada ve Yabuuchi (1995) aykırı değer tarafından etkilenmeyen bulanık robust regresyon modelini önermişlerdir. Büyük hataya sahip ya da düzensiz olan örnekleri seçmek için hyperelliptic fonksiyonu, optimizasyon problemini çözmek için genetik algoritma yöntemini kullanmışlardır.

Chang ve Lee (1996) aykırı değer olması durumu için üyelik dereceleriyle ağırlıklandırma yapan ve karar verici ile etkileşime dayanan genelleştirilmiş bulanık ağırlıklandırılmış en küçük kareler yöntemini ileri sürmüşlerdir.

Yang ve Ko (1997) basit regresyonda ağırlıklandırılmış bulanık en küçük kareler çözümlemesi için yinelemeli algoritmayı geliştirmişlerdir. Bu algoritma iki aşamalıdır. İlk olarak gözlemlerin sınıf üyeliklerini veren bulanık sınıflama yöntemi seçilir, daha sonra üyeliklerin bu değerleri ağırlıklar olarak kullanılır. Bulanık regresyon çözümlemesinde ağırlıklandırılmış bulanık en küçük kareler bir optimizasyon problemi olarak düşünülmüştür.

Özelkan ve Duckstein (2000) belirsizlik ve veri uç noktaları arasındaki ödünleşime dayanan bir baskın olmayan çözümü seçmek için karar vericiyle etkileşime dayanan çok amaçlı bulanık regresyonu önermişlerdir.

Watada vd (2000) bulanık regresyon modelini yazılım projelerindeki kusur sayılarını tahmin etmek için kullanmışlardır.

Chen (2001) veride aykırı değer olması durumunda işareti belirtilmemiş bulanık regresyon ve olanaklı (possibilistic) bulanık regresyonun yanlış sonuçlar

vereceğini ileri sürmüştür. Bu nedenle işareti belirtilmemiş bulanık regresyon ve olanaklı bulanık regresyonun birlikte ele alındığı ve yeni bir kısıtın ilave edildiği bir model önermiştir. Bu modelde esas amaç k değerinin belirlenmesidir.

Chang ve Ayyub (2001) bulanık regresyon yöntemlerini karşılaştırmıştır.

Hund vd (2002) deney tasarımında aykırı değer belirlenmesi için robust regresyonu kullanmışlardır.

Yang ve Lin (2002) bulanık girdi ve bulanık çıktı için bulanık doğrusal en küçük kareler regresyon çözümlemesini ileri sürmüşlerdir. Heterojen veri kümesi ve aykırı değerleri belirlemek için kümeleme analizinden yararlanmışlardır.

Lin ve Chen (2003) bulanık diskriminant analizine, yeni bir yaklaşım olan genetik algoritma ile aykırı değer belirlendiği bulanık diskriminant analizini önermişlerdir.

Yang ve Liu (2003) etkileşimli bulanık doğrusal regresyon modelleri için bulanık en küçük kareler algoritmasını önermişlerdir. Bu algoritma basit regresyon için aykırı değere karşı robusttur. Bu algoritmada ortogonal koşullar optimizasyon problemine kısıt olarak eklenmiştir.

Gürültüye karşı daha fazla dirençli parametre tahminleri yapmak için bulanık c ortalamaya dayalı kümelemenin (FCM) amacını yeniden düzenlemeye dayanan ilk yöntem Ohashi (1984) tarafından önerilmiştir (Dave ve Krishnapuram 1997, Nasraoui 2004)

Daha sonra Dave (1991) gürültü kümeleme (noise clustering (NC)) yöntemini ileri sürmüştür. NC yaklaşımının ana avantajı FCM algoritmasının robust şekli olmasıdır (Dave 1993, Dave ve Sen 1998).

FCM'in görelî üyelik probleminin üstesinden gelmek için Krishnapuram ve Keller (1993) olanaklı c ortalamaya dayalı kümeleme yöntemini (PCM) önermişlerdir.

Dave ve Krishnapuram (1996) robust kümeleme yöntemlerini karşılaştırmışlardır. Robust istatistikler ve bulanık küme teorisi arasındaki benzerlikleri saptayıp, robust kümeleme metotları ve M-tahmin edicileri arasındaki benzerlikleri göstermişlerdir.

Pal ve Bezdek (1997) hem bulanık kümelemedeki üyeliklerin, hem de bir gözlemin bir kümeye ne kadar uygun olduğunu gösteren küme özelliğini (typicality (t_y)) içeren karma bulanık olanaklı c ortalamaya dayalı kümeleme (FPCM) modelini tanımlamışlardır.

Dave ve Sen (1998), Dave (1991) tarafından önerilen gürültü kümeleme algoritmasını genelleştirmişlerdir (Dave ve Sen 1998).

Dave ve Krishnapuram (1997) robust kümeleme yöntemlerinin karşılaştırıldığı bir çalışma yapmışlardır.

2. BULANIK KÜME TEORİSİ

2.1 Giriş

Bulanık küme yaklaşımı ilk defa Amerika Birleşik Devletlerinde 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. İnsan mantığı, açık, kapalı, sıcak, soğuk, 0 ve 1 gibi 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.

Bulanık mantığın genel özellikleri Zadeh tarafından aşağıdaki gibi 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 (Elmas 2003).

Alt kesimlerde Bulanık Küme ile ilgili temel tanımlar, aritmetik işlemler ve karşılaştırmalar verilecektir.

2.2. Bulanık Küme Teorisi İle İlgili Temel Kavramlar

Bu kesimde bulanık küme ile ilgili bazı temel kavramlar verilecektir. Bu kavramlardan pek çoğu klasik (kesin) kümenin temel kavramlarının genişletilmiştir, fakat bazı kavramlar ise bulanık küme teorisi kapsamında taktır (Wang 1997).

Tanım 2.1. Bulanık Küme: Üyeleri kesin olarak belirli olmayan fakat aday üyelerinin bu kümeye üyelik derecelerinin bilindiğı bir kümedir.

Bir başka tanım ise; U evrensel küme, x evrensel kümenin herhangi bir elemanı ise A kümesinin karakteristik fonksiyonu,

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

bişiminde tanımlanır. Burada karakteristik fonksiyonunun değeri kümesi $\{0,1\}$ 'dir. Eğer karakteristik fonksiyonun değeri kümesi $[0,1]$ aralığı ise ve bu aralıkta her gerçel sayıyı alabilecek şekilde tanımlanırsa A kümesine bulanık küme denir (Lai ve Hwang 1992, Wang 1997, Apaydın 2000).

Tanım 2.2. Üyelik Fonksiyonu: U evrensel kümesindeki bulanık küme A , üyelik fonksiyonu $\mu_A(x)$ ile tanımlanır ve $\mu_A(x)$, $[0,1]$ aralığında değeri alır. Böylece bulanık küme $[0,1]$ aralığında herhangi bir değeri alan üyelik fonksiyonuna izin veren klasik kümenin genelleştirilmiştir (Wang 1997). Eğer x , A kümesinin elemanı değil ise $\mu_A(x)=0$ değeri, x tamamen A kümesinin elemanı ise $\mu_A(x)=1$ değeri alır. Klasik kümeler yalnızca iki üyelik ($\mu_A(x)=0$ ve $\mu_A(x)=1$) değeri alırlar (Cherkassky ve Mulier 1998). Başka bir deyişle,

klasik kümenin üyelik fonksiyonu yalnızca sıfır ve bir değerlerini alırken, bulanık kümenin üyelik fonksiyonu $[0,1]$ aralığında bir fonksiyondur. U evrensel kümesindeki bulanık küme A , eleman x ve bunun üyelik değeri sıralanmış parçaların kümesi olarak

$$A = \{x, \mu_A(x) | x \in U\} \quad (2.1)$$

biçiminde tanımlanır.

U sürekli olduğunda bulanık küme

$$A = \int_U \mu_A(x) / x \quad (2.2)$$

olarak tanımlanır.

Eşitlik (2.2)'de integral işareti integrali tanımlamaz, ortak üyelik fonksiyonu $\mu_A(x)$ 'e sahip tüm noktaları ($x \in U$) tanımlar. U kesikli olduğunda

$$A = \sum_U \mu_A(x) / x \quad (2.3)$$

biçiminde tanımlanır. Buradaki toplam işareti aritmetik toplamı tanımlamaz, ortak üyelik fonksiyonu $\mu_A(x)$ 'e sahip tüm noktaları ($x \in U$) tanımlar.

Bir bulanık küme yalnız ve yalnız bir uygun üyelik fonksiyonuna sahiptir. Bir bulanık küme ele alındığında buna karşılık gelen tek üyelik fonksiyonu vardır, bir üyelik fonksiyonuna sahip olduğunda ise bu bir bulanık kümeyi temsil eder (Lai ve Hwang 1992, Wang 1997).

Tanım 2.3. Destek Kümesi: A bulanık kümesinin destek kümesi, bulanık küme A 'da sıfırdan farklı üyelik değerine sahip x 'leri içerir ve

$$des A = \{x | \mu_A(x) > 0 \text{ ve } x \in U\}$$

biçiminde tanımlanır.

Tanım 2.4. α -Kesme Kümesi: A bulanık kümesinin α -kesme kümesi, bulanık küme A 'da en az α 'a eşit ya da daha büyük üyelik değerlerine sahip olan x 'leri içerir. A bulanık kümesinin α -kesme kümesi,

$$A_\alpha = \{x | \mu_A(x) \geq \alpha \text{ ve } x \in U\}$$

olarak tanımlanır.

Tanım 2.5. Yükseklik: A bulanık kümesinin yüksekliği,

$$h(A) = \sup_{x \in U} \mu_A(x)$$

biçiminde tanımlanır.

Eğer bulanık küme A 'nın yüksekliği 1'e eşitse ($h(A)=1$) normal bulanık küme olarak adlandırılır. Aksi takdirde A bulanık kümesi alt normaldir (Lai ve Hwang 1992, Klir ve Yuan 1995).

Tanım 2.6. Dışbükey Bulanık Küme: α -kesme kümesi A_α , $(0,1]$ aralığında her hangi bir α için dışbükey küme ise bulanık küme A dışbükeydir denir.

$x_1, x_2 \in R^n$ ve $\lambda \in [0,1]$ için dışbükey bulanık küme,

$$\mu_A[\lambda x_1 + (1-\lambda)x_2] \geq \min[\mu_A(x_1), \mu_A(x_2)],$$

biçiminde tanımlanır (Wang 1997).

Tanım 2.7. Genişleme (Extension) İlkesi: U evrensel kümesinden V evrensel kümesine bir fonksiyon $f:U \rightarrow V$ olsun. U 'daki A bulanık kümesi verildiği ve f fonksiyonu ile V 'deki $B = f(A)$ bulanık kümesinin belirlenmesi istendiğinde,

$$\mu_B(y) = \mu_A[f^{-1}(y)], \quad y \in V$$

tanımlanır. Burada $f^{-1}(y)$, f 'in tersidir. Daha genel olarak, B için üyelik fonksiyonu,

$$\mu_B(y) = \max_{x \in f^{-1}(y)} \mu_A(x), \quad y \in V \quad (2.4)$$

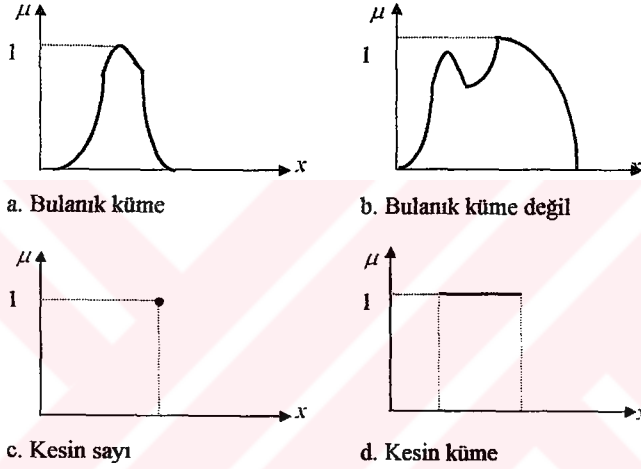
biçimindedir. Burada $f^{-1}(y)$, tüm $x \in U$ kümesi için $f(x) = y$ 'dir. (2.4) ile verilen eşitlik genişleme ilkesi olarak tanımlanır (Wang 1997, Apaydın 2000).

Tanım 2.8. Bulanık Sayı: A bulanık küme ve $x \in A$ olmak üzere,

1. A bulanık kümesi normal ise,
2. $A_\alpha \in (0,1]$ ise,
3. A 'nın destek kümesi sınırlı ise,

biçiminde belirtilen özelliklere sahip x bulanık sayı olarak adlandırılır (Klir ve Yuan 1995).

Bulanık küme ve kesin küme üyeliklerinin grafiksel gösterimi Şekil 2.1'de gösterilmiştir (Cherkassky ve Mulier 1998, Tanaka ve Guo 1999).

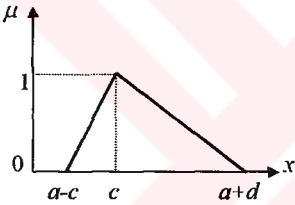


Şekil 2.1. Bulanık küme ve kesin küme üyeliklerinin grafiksel gösterimi

Tanım 2.9. Üçgensel Bulanık Sayı: Üçgensel bulanık sayı $x = (a, c, d)_T$ olarak tanımlanır. Burada a merkez, c sol yayılma, d sağ yayılma ve T üçgensel bulanık sayı anlamında kullanılır. A bulanık küme, $x \in A$ olmak üzere, üçgensel bulanık sayı x 'in üyelik fonksiyonu $\mu(x)$,

$$\mu(x) = \begin{cases} 0 & , \quad x \leq a-c \\ 1 - \frac{(a-x)}{c} & , \quad a-c \leq x \leq a \\ 1 - \frac{(x-a)}{d} & , \quad a \leq x \leq a+d \\ 0 & , \quad x \geq a+d \end{cases}$$

biçiminde tanımlanır. Üçgensel bulanık sayı x 'in üyelik fonksiyonuna ait grafik Şekil 2.2'de verilmiştir.



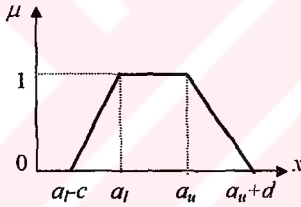
Şekil 2. 2. Üçgensel bulanık sayı

Tanım 2.10. Yamuksal Bulanık Sayı: Yamuksal bulanık sayı $x = (a_l, a_u, c, d)_{Tr}$

olarak tanımlanır. Burada c sol yayılma, d sağ yayılma ve Tr yamuksal şekle adaylığı gösterir. A bulanık küme, $x \in A$ olmak üzere, yamuksal bulanık sayı x 'in üyelik fonksiyonu $\mu(x)$,

$$\mu(x) = \begin{cases} 0 & , \quad x \leq a_l - c \\ 1 - \frac{(a_l - x)}{c} & , \quad a_l - c \leq x \leq a_l \\ 1 & , \quad a_l \leq x \leq a_u \\ 1 - \frac{(x - a_u)}{d} & , \quad a_u \leq x \leq a_u + d \\ 0 & , \quad x \geq a_u + d \end{cases}$$

olarak tanımlanır. Yamuksal bulanık sayı x 'in grafiksel gösterimi Şekil 2.3'de verilmiştir (Tanaka ve Guo 1999).



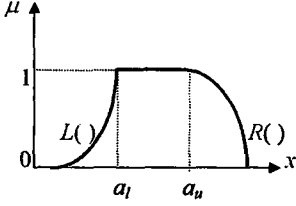
Şekil 2. 3. Yamuksal bulanık sayı

Tanım 2.11. L - R Bulanık Sayı: L - R bulanık sayıları Dubois ve Prade (1978) tarafından önerilmiştir. Bir L - R bulanık sayısı $x = (a_l, a_u, c, d)_{LR}$ olarak tanımlanır. Burada c sol yayılma, d sağ yayılma, $L(\bullet)$ ve $R(\bullet)$ bulanık sayılar fonksiyonu olarak tanımlanır. L ve R bulanık sayıların sol ve sağ yanlarını tanımlayan fonksiyonlardır ve

1. $L(x) = L(-x)$, $R(x) = R(-x)$,
2. $L(0) = 1$, $R(0) = 1$,

3. $x \geq 0$ için $L(x)$ ve $R(x)$ kesin azalandır,

koşullarını sağlarlar. L - R bulanık sayısına ait grafik Şekil 2.4’de verilmiştir.



Şekil 2. 4. L - R bulanık sayısı

Şekil 2. 4’de gösterilen L - R bulanık sayısı için üyelik fonksiyonu $\mu(x)$,

$$\mu(x) = \begin{cases} L((a_l - x)/c), & x \leq a_l \\ 1, & a_l \leq x \leq a_u \\ R((x - a_u)/d), & x \geq a_u \end{cases}$$

biçiminde tanımlanır.

$L(\bullet) = R(\bullet)$ ve $c = d$ olduğunda bulanık sayı L bulanık sayısı olarak isimlendirilir ve $x = (a_l, a_u, c)_{LL}$ ya da $x = (a, c)_L$, $a_l = a_u = a$ şeklinde gösterilir.

L bulanık sayısı $a_l = a_u$ ve $L(x) = \max(0, 1 - |x|)$ olduğunda, bulanık sayı $(a, c)_T$ şeklinde gösterilir ve simetrik üçgensel bulanık sayı olarak isimlendirilir (Tanaka ve Guo 1999).

2.3. Üyelik Fonksiyonlarının Sınıflandırılması

Üyelik fonksiyonları,

- I. Sezgisel belirlemeye dayalı üyelik fonksiyonları,
- II. Özel problemlerle ilgili güvenilirlik düşüncesine dayanan üyelik fonksiyonları,
- III. Teorik temele dayanan üyelik fonksiyonları,
- IV. Deneysel olarak modellenen üyelik fonksiyonları

olmak üzere 4 başlıkta sınıflandırılabilir. Bu sınıflandırmalarda tanımlanan üyelik fonksiyonları aşağıda verilmiştir:

I. Sezgisel belirlemeye dayalı üyelik fonksiyonları

a-) Zadeh'in tek model fonksiyonları,

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

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

dir.

b-) Dimitru ve Luban'ın güç fonksiyonları,

$$\mu(x) = \frac{x^2}{a^2} + 1, \quad x \in [0, a]$$

$$\mu(x) = -\frac{x^2}{a^2} - \frac{2x}{a} + 1, \quad x \in [0, a]$$

biçimindedir.

c-) Svarowski'nin sin fonksiyonu,

$$\mu(x) = \frac{1}{2} + \frac{1}{2} \sin \left(\frac{\pi}{b-a} \left(x - \frac{a+b}{2} \right) \right), \quad x \in [a, b]$$

dir (Lai ve Hwang 1992).

II. Özel problemlerle ilgili güvenilirlik düşüncesine dayanan üyelik fonksiyonları

a-) Zimmerman'ın doğrusal fonksiyonu,

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

dir.

b-) Tanaka, Uejima ve Asai'nin simetrik üçgensel fonksiyonu,

$$\mu(x) = \begin{cases} 1 - \frac{|b-x|}{a}, & b-a \leq x \leq b+a \\ 0 & d.d \end{cases}$$

biçimindedir.

c-) Hannan'ın parçalı doğrusal fonksiyonu,

$$\mu(x) = \sum_j \alpha_j |x - a_j| + \beta x + r, \quad j = 1, 2, \dots, N$$

$$\alpha_j = (t_{j+1} - t_j) / 2$$

$$\beta = (t_{N+1} + t_1) / 2$$

$$r = (s_{N+1} + s_1) / 2$$

dir. Burada her bir i parçası için $\mu(x) = t_i x + s_i$, $a_{i-1} \leq x \leq a_i$ 'dir ve t_i eğim, s_i ise a_{i-1} 'de başlayan ve a_i 'de biten eğrinin y -sabitidir.

d-) Leberling'in hyperbolic fonksiyonu,

$$\mu(x) = \frac{1}{2} + \frac{1}{2} \tanh(a(x-b)), \quad -\infty \leq x \leq \infty$$

biçimindedir. Burada a parametredir.

e-) Sakawa ve Yumine'nin üstel ve ters hyperbolic fonksiyonu,

$$\mu(x) = c \left(1 - e^{(b-x)/(b-a)} \right), \quad x \in [a, b]$$

$$\mu(x) = \frac{1}{2} + c \tanh^{-1}(d(x-b))$$

dir. Burada c ve d parametrelerdir.

f-) Dimitru ve Luban'ın fonksiyonu,

$$\mu(x) = \frac{1}{1 + \frac{x}{a}}$$

biçimindedir. Burada a parametredir.

g-) Dubois ve Prade'nin L-R bulanık sayısı,

$$\mu(x) = \begin{cases} L((a-x)/\alpha), & x \leq a_l \\ 1, & a_l \leq x \leq a_u \\ R((x-a)/\beta), & x \geq a_u \end{cases}$$

dır. Burada $L(\bullet)$ ve $R(\bullet)$ referans fonksiyonlarıdır (Lai ve Hwang 1992).

III. Teorik temele dayanan üyelik fonksiyonları

a-) Civanlar ve Trussel'in fonksiyonu,

$$\mu(x) = \begin{cases} ap(x) & ap(x) \leq 1 \\ 0 & d.d \end{cases}$$

biçimindedir. Burada $a \in [0,1]$ parametre ve $p(x)$ olasılık yoğunluk fonksiyonudur.

b-) Svarovski'nin fonksiyonu,

$$\mu(x) = \begin{cases} 0 & , & x < a \\ K(x-a)^2 & , & a \leq x \leq b \\ K_2x^2 + K_1x + K_0 & , & b < x \leq c \\ 1 & , & x > c \end{cases}$$

dir. Burada K, K_0, K_1 ve K_2 parametrelerdir (Lai ve Hwang 1992).

IV. Deneyisel olarak modellenen üyelik fonksiyonları

a-) Hersh ve Caramazza fonksiyonu,

$$\mu(x) = \frac{1}{2} + d(r/10)$$

biçimindedir. Burada “evet” cevabı için $d(x)=1$ ve “hayır” cevabı için $d(x)=-1$ değerini alır. r güven değeridir.

b-) Zimmermann ve Zysno'un fonksiyonu,

$$\mu(x) = \frac{1}{2} + \frac{1}{d} \left(\frac{1}{1 + e^{-d(x-a)}} - c \right)$$

dir.

c-) Dombi'nin fonksiyonu,

$$\mu(x) = \frac{(1-s)x^2}{(1-s)x^2 + s(1-x)^2}$$

dir. Burada $s, y = \mu(x)$ ve $y = x$ 'in kesişme değeridir (Lai ve Hwang 1992).

2.4. Bulanık Kümelerdeki Operatörler

Bu kesimde bulanık kümelerde kullanılan eşitlik, kapsama, tümleme, birleşim ve kesişim operatörleri tanımlanacaktır.

Tanım 2.12. Eşitlik : Tüm $x \in U$ için $\mu_A(x) = \mu_B(x)$ ise A bulanık kümesi ve B bulanık kümesi eşittir denir.

Tanım 2.13. Kapsama : $\mu_A(x) \leq \mu_B(x)$ ise B bulanık kümesi, A bulanık kümesini içeriyor denir ve $A \subset B$ biçiminde gösterilir.

Tanım 2.14. Tümleyen : Tüm $x \in U$ için A bulanık kümesinin tümleyeni $\bar{A}$. $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$, üyelik fonksiyonu ile tanımlanır.

Tanım 2.15. Birleşim : A bulanık kümesi ve B bulanık kümesinin birleşimi $A \cup B$ ile gösterilen U 'da bir bulanık kümedir ve üyelik fonksiyonu

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)]$$

dır.

Tanım 2.16. Kesişim : A bulanık kümesi ve B bulanık kümesinin kesişimi $A \cap B$ bulanık kümesidir ve üyelik fonksiyonu

$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)]$$

biçimindedir (Wang 1997).

2.5. Bulanık Kümelerdeki Aritmetik İşlemler

Üçgensel ve yamuksal bulanık sayılar için aritmetik işlemler aşağıdaki gibi tanımlanır:

İki üçgensel bulanık sayı $X = (x, \alpha, \beta)$ ve $Y = (y, r, \delta)$ ise,

Y 'nin görüntüsü (image) : $-Y = (-y, \delta, r)$

Y 'nin tersi : $Y^{-1} = (y^{-1}, \delta y^{-2}, r y^{-2})$

Toplama : $X (+) Y = (x + y, \alpha + r, \beta + \delta)$

Çıkarma : $X (-) Y = X (+) -Y = (x - y, \alpha + \delta, \beta + r)$

Çarpma : $X > 0, Y > 0 : X (.) Y = (xy, xr + y\alpha, x\delta + y\beta)$
 $X < 0, Y > 0 : X (.) Y = (xy, y\alpha - x\delta, y\beta - xr)$
 $X < 0, Y < 0 : X (.) Y = (xy, -x\delta - y\beta, -xr - y\alpha)$

Sabitler Çarpma : $a > 0, a \in R : a (.) X = (ax, a\alpha, a\beta)$
 $a < 0, a \in R : a (.) X = (ax, -a\beta, -a\alpha)$

Bölme :
 $X > 0, Y > 0 : X (/) Y = (x / y, (x\delta + y\alpha) / y^2, (xr + y\beta) / y^2)$
 $X < 0, Y > 0 : X (/) Y = (x / y, (y\alpha - xr) / y^2, (y\beta - x\delta) / y^2)$
 $X < 0, Y < 0 : X (/) Y = (x / y, (-xr - y\beta) / y^2, (-x\delta - y\alpha) / y^2)$

olarak tanımlanır (Lai ve Hwang 1992).

İki yamuksal bulanık sayı $X = (x_1, x_2, \alpha, \beta)$ ve $Y = (y_1, y_2, r, \delta)$ ise,

Y'nin görüntüsü (image) : $-Y = (-y_1, y_2, \delta, r)$

Y'nin tersi : $Y^{-1} = (y_2^{-1}, y_1^{-1}, \delta / [y_2 (y_2 + \delta)], r / [y_1 (y_1 - r)])$

Toplama : $X (+) Y = (x_1 + y_1, x_2 + y_2, \alpha + r, \beta + \delta)$

Çıkarma : $X (-) Y = X (+) - Y = (x_1 - y_2, x_2 - y_1, \alpha + \delta, \beta + r)$

Çarpma :

$$X > 0, Y > 0 : X (.) Y = (x_1 y_1, x_2 y_2, x_1 r + y_1 \alpha - \alpha r, x_2 \delta + y_2 \beta - \beta \delta)$$

$$X < 0, Y > 0 : X (.) Y = (x_1 y_2, x_2 y_1, y_2 \alpha - x_1 \delta + \alpha \delta, y_1 \beta - x_2 r - \beta r)$$

$$X < 0, Y < 0 : X (.) Y = (x_2 y_2, x_1 y_1, -x_2 \delta - y_2 \beta - \beta \delta, -x_1 r - y_1 \alpha + \alpha r)$$

Sabit Çarpma : $a > 0, a \in R : a (.) X = (ax_1, ax_2, a\alpha, a\beta)$

$a < 0, a \in R : a (.) X = (ax_1, ax_2, -a\beta, -a\alpha)$

Bölme :

$$X > 0, Y > 0 : X (:) Y = (x_1 / y_2, x_2 / y_1, (x_1 \delta + y_2 \alpha) / [y_2 (y_2 + \delta)], (x_2 r + y_1 \beta) / [y_1 (y_1 - r)])$$

$$X < 0, Y > 0 : X (:) Y = (x_1 / y_1, x_2 / y_2, (y_1 \alpha - x_1 r) / [y_1 (y_1 - r)], (y_2 \beta - x_2 \delta) / [y_2 (y_2 + \delta)])$$

$$X < 0, Y < 0 : X (:) Y = (x_2 / y_1, x_1 / y_2, (-x_2 r - y_1 \beta) / [y_1 (y_1 - r)], (x_1 \delta - y_2 \alpha) / [y_2 (y_2 + \delta)])$$

olarak tanımlanır (Lai ve Hwang 1992).

2.6. Bulanık Kümelerin Karşılaştırılması

Bulanık sayıların ya da daha genel olarak bulanık kümelerin sıralanması veya karşılaştırılması uygulamalı yaklaşımlar için çok önemlidir. Matematiksel modellerin kurulmasında bulanık küme teorisi kullanılıyorsa, bulanık sayıların karşılaştırılması veya sıralanması problemi ile karşılaşılır. Bulanık sayılar doğrusal bir sıra içinde olmadıkları için karşılaştırılmaları zordur. Literatürde farklı yaklaşımlar için farklı bulanık sıralama yöntemleri önerilmiştir.

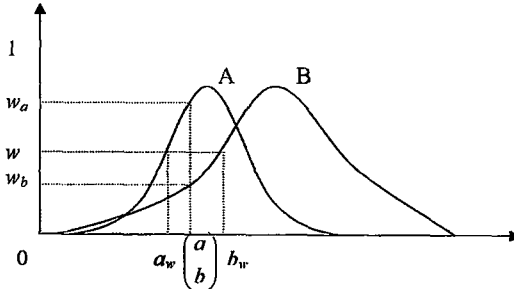
a_w : A bulanık kümesinin elemanı,

b_w : B bulanık kümesinin elemanı,

w : A ve B bulanık kümelerinin elemanı olan sırasıyla a_w ve b_w 'ya karşılık gelen üyelik derecesi olsun.

Eğer $\forall w \in [0,1]$ için $a_w < b_w$ ise (Bulanık küme A) < (Bulanık küme B) sonucuna varılabilir.

(a, b) , sırasıyla A ve B bulanık kümesinin bir elemanı olsun. w_a , a 'ya karşılık gelen üyelik derecesi; w_b , b 'ye karşılık gelen üyelik derecesi olsun. $w_a > w_b$ olması (Bulanık küme A) > (Bulanık küme B) olmasını gerektirmez. Bulanık A ve B kümelerinin karşılaştırılması Şekil 2.5'te verilmiştir.



Şekil 2.5. İki bulanık kümenin karşılaştırılması

Şekil 2.5’de a_w ve b_w , w üyelik fonksiyonunun aynı derecesine dayanır. $A < B$ tek bir w derecesi düşünüldüğünde muhtemelen doğrudur, küme tüm seviyeler açısından ele alındığında probleme başka bir bakış açısı getirilmelidir. Bu durumda tüm seviyeleri göz önünde bulundurmaya imkan veren “overall existence” kavramı söz konusudur. Bu yöntemde kullanılan indeks “overall existence index” olarak adlandırılır ve

$$I = \int_0^1 g\left(\left\{\mu_A^{-1}(w)\right\}\right)dw - g\left(\left\{\mu_B^{-1}(w)\right\}\right)dw$$

biçiminde tanımlanır. Burada

- g : Üyelik fonksiyonlarının tersinin bir fonksiyonu
- μ_A : A bulanık kümesinin üyelik fonksiyonu
- μ_B : B bulanık kümesinin üyelik fonksiyonu
- $\mu_A^{-1}(w)$: $\mu_A(x)$ üyelik fonksiyonunun ters görüntüsü
- $\mu_B^{-1}(w)$: $\mu_B(x)$ üyelik fonksiyonunun ters görüntüsü

dır. A ve B bulanık kümesine ait üyelik fonksiyonlarının ters görüntüsü sırasıyla

$$\begin{aligned}\mu_A^{-1}(w) &= \{x : \mu_A(x) = w\} \\ \mu_B^{-1}(w) &= \{y : \mu_B(y) = w\} \quad x, y \in R\end{aligned} \quad (2.5)$$

biçiminde tanımlanır.

A ve B bulanık sayıları arasındaki fark

$$d(A, B) = \int_0^{w_{\text{gt}}} g_A\left(\left\{\mu_A^{-1}(w)\right\}\right)dw - \int_0^{w_{\text{gt}}} g_B\left(\left\{\mu_B^{-1}(w)\right\}\right)dw \quad (2.6)$$

olarak tanımlanır. Burada

$$w_{hgt}^* = \min [hgt(A), hgt(B)] \quad (2.7)$$

$hgt(A)$: A bulanık sayısının alabileceği en yüksek üyelik derecesi,

$hgt(B)$: B bulanık sayısının alabileceği en yüksek üyelik derecesi

dir. Her bir bulanık küme için özel ölçü $OM(A)$

$$OM(A) = \int_0^{w_{hgt}^*} g_i(\{\mu_A^{-1}(w)\}) dw \quad (2.8)$$

biçiminde tanımlanır. Bu tanım tüm üyelik fonksiyonları (dışbükey, dışbükey olmayan, normal, normal olmayan, sürekli vb.) için uygulanabilir. Eşitlik

(2.8)'deki $g_i(\{\mu_A^{-1}(w)\})$

$$g_i(\{\mu_A^{-1}(w)\}) = W(w) [\chi_1(w)x'_i(w) + \chi_2(w)x''_i(w)] \quad (2.9)$$

olarak tanımlanır. Burada $x'_i(w)$, A bulanık sayısının sol referansının ters görüntüsü ve $x''_i(w)$ A bulanık sayısının sağ referansının ters görüntüsüdür ve

$$x'_i(w) = \mu_{A_L}^{-1}(w) \quad (2.10)$$

$$x''_i(w) = \mu_{A_R}^{-1}(w)$$

biçiminde tanımlanırlar. Tüm bu tanımlamalardan sonra OM indeksi $OM(A)$

$$OM(A) = \int_0^1 W(w) [\chi_1(w) \mu_{A_L}^{-1}(w) + \chi_2(w) \mu_{A_R}^{-1}(w)] dw \quad (2.11)$$

gibi tanımlanır.

Burada

$$W'(w) = \frac{w}{\frac{1}{2} \left(w_{lgr}^* \right)^2}$$

olarak seçilir.

Eşitlik (2.11)'de $W(w)$, $\chi_1(w)$ ve $\chi_2(w)$ ağırlık ölçüleridir ve karar verici tarafından belirlenir. $\chi_1(w)$ ve $\chi_2(w)$

$$\chi_1(w) + \chi_2(w) = 1 \text{ ve } \chi_1(w), \chi_2(w) \in (0, 1] \quad (2.12)$$

koşullarını sağlamaları gereklidir (Chang ve Lee 1994).

3. KÜMELEME YÖNTEMLERİ

Robust kümeleme yöntemleri temel olarak iki sınıfta toplanabilirler. Bunlardan birincisi doğrudan robust istatistiklere dayanan yöntemler, ikincisi ise aykırı değere daha dayanıklı parametre tahminleri yapabilmek için bulanık c ortalamaya (fuzzy c - means (FCM)) dayalı kümeleme yönteminin amaç fonksiyonunun yenilenmesine dayanan robust kümeleme yöntemleridir (Nasraoui 2004). Veriyi etkileyen dış faktörlere diğer bir ifadeyle gürültüye karşı daha fazla dayanıklı parametre tahminleri yapmak için bulanık c ortalamaya dayalı kümelemenin (FCM) amacını yeniden düzenlemeye dayanan ilk yöntem Ohashi (1984) tarafından önerilmiştir. Daha sonra Dave (1991) gürültü kümeleme (noise clustering (NC)) yöntemini ileri sürmüştür (Dave ve Krishnapuram 1997). NC yaklaşımının ana avantajı FCM algoritmasının robust şekli olmasıdır.

Katı (hard) c ortalamaya dayalı kümelemede birimler ya bir kümeye aittir ya da değildir. Fakat bazı durumlarda birimler kümelere tam üye olmamakla birlikte belirli üyeliklerle birden fazla kümeye ait olabilirler. Bu durumda bulanık c ortalamaya dayalı kümeleme yöntemi daha uygundur. Veri kümesinde aykırı değer ya da gürültü olması durumunda ise katı ve bulanık c ortalamaya dayalı kümeleme yöntemleri uygun değildir. Çünkü FCM algoritmaları aykırı değerden çok etkilenmektedir. Başka bir ifadeyle tek bir aykırı değer modeli tümüyle değiştirebilir (Dave 1993, Dave ve Krishnapuram 1997). Bu durumda robust kümeleme yöntemleri kullanılır.

Alt kesimlerde katı ve bulanık parçalanma tanımları, katı ve bulanık kümeleme yöntemleri ve robust kümeleme yöntemleri ayrıntılı olarak verilecektir.

3.1. Katı ve Bulanık Parçalanma

$X = \{x_1, x_2, \dots, x_n\}$ veri kümesi ele alınsın. Burada x_j herhangi bir eleman ve $x_i \in R^p$ 'dir. X 'in kuvvet kümesi $P(X)$ ile gösterilsin. $\bigcup_{i=1}^c A_i = X$ ve $A_i \cap A_j = \emptyset$, $1 \leq i \neq j \leq c$ için X 'in katı c parçalanması $\{A_i \in P(X) : 1 \leq i \leq c\}$ ailesidir. Her bir A_i küme olmak üzere, $\{A_1, A_2, \dots, A_c\}$ c kümeden oluşan X 'in bir parçalanmasıdır.

Katı c parçalanma A_i 'deki x_j elemanlarının üyelik fonksiyonu aracılığıyla yazılabilir ve

$$u_{ij} = \begin{cases} 1, & x_j \in A_i \\ 0, & x_j \notin A_i \end{cases}$$

olarak tanımlanır. Burada $x_j \in X$, $A_i \in P(X)$, $i = 1, 2, \dots, c$ 'dir. x_j elemanı A_i kümesine ait ise $u_{ij} = 1$, kümeye ait değilse $u_{ij} = 0$ 'dır. u_{ij}

$$u_{ij} \in [0, 1] \quad 1 \leq i \leq c, 1 \leq j \leq n \quad (3.1)$$

$$\sum_{i=1}^c u_{ij} = 1, \quad \forall j \in \{1, 2, \dots, n\} \quad (3.2)$$

$$0 < \sum_{j=1}^n u_{ij} < n, \quad \forall i \in \{1, 2, \dots, c\} \quad (3.3)$$

koşullarını sağlar.

Eşitlik (3.1) ve (3.2) her bir $x_j \in X$ 'in yalnız ve yalnız bir kümeye ait olmasını, Eşitlik (3.3) ise her bir A_i kümesinin en az bir en çok $n-1$ veri noktasını içermesini sağlar. $c \times n$ boyutlu U matrisi u_{ij} 'lerden oluşur ($1 \leq i \leq c$ ve $1 \leq k \leq n$) (Wang 1997).

Tanım 3.1. $X = \{x_1, x_2, \dots, x_n\}$ herhangi bir küme, V_{cn} $c \times n$ boyutlu $U = [u_{ij}]$ matrislerinin kümesi olsun. X 'in katı c parçalanma uzayı

$$M_c = \{U \in V_{cn} \mid \text{Eşitlik (3.1) ve Eşitlik (3.3) doğru} \}$$

kümesidir. Burada M_c

$$|M_c| = \frac{1}{c!} \left[\sum_{j=1}^c \binom{c}{j} (-1)^{c-j} j^n \right]$$

olacak şekilde tanımlanır (Wang 1997).

Tanım 3.2. $X = \{x_1, x_2, \dots, x_n\}$ herhangi bir küme, V_{cn} ; $c \times n$ boyutlu $U = [u_{ij}]$ matrislerinin kümesi olsun. X 'in bulanık c parçalanma uzayı M_{fc}

$$M_{fc} = \{U \in V_{cn} \mid u_{ik} \in [0, 1], 1 \leq i \leq c, 1 \leq k \leq n; \text{Eşitlik (3.2) doğru} \}$$

biçiminde tanımlanır. u_{ij} , A_i kümesine ait olan x_j 'nin üyeliğidir (Wang 1997).

3.2. Katı c Ortalamaya Dayalı Kümeleme

Katı c ortalamaya dayalı kümelemede amaç fonksiyonu,

$$J(B, U; X) = \sum_{i=1}^c \sum_{j=1}^n u_{ij} (d_{ij})^2 \quad (3.4)$$

olarak tanımlanır. Burada $(d_{ij})^2 = \|x_j - v_i\|^2$, $(i=1,2,\dots,c$ ve $j=1,2,\dots,n)$ 'dir ve amaç fonksiyonundaki B parametreleri temsil etmektedir. Eşitlik (3.4)'ü minimum yapacak küme merkezi (v_i) ve üyelikler (u_{ij})

$$v_i = \frac{\sum_{j=1}^n x_j u_{ij}}{\sum_{j=1}^n u_{ij}}, \quad i=1,2,\dots,c \quad (3.5)$$

$$u_{ij} = \begin{cases} 1 & \|x_j - v_i\| = \min_{1 \leq k \leq c} (\|x_j - v_k\|) \\ 0 & \text{d.d} \end{cases} \quad (3.6)$$

olarak tanımlanır (Nasraoui 2004, Wang 1997) .

3.3. Bulanık c Ortalamaya Dayalı Kümeleme

Katı c ortalamaya dayalı kümeleme ile FCM arasındaki temel fark birimlerin bir kümeye ait olması üyeliğinin belirlenmesidir. $U = [u_{ij}] \in M_{fc}$ ve $v_i \in R^p$ olan $V = (v_1, v_2, \dots, v_c)$ 'yi bulmak için FCM'de amaç

$$J(B, U; X) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m (d_{ij})^2 \quad (3.7)$$

fonksiyonunun minimum yapılmasıdır. Burada $m \in [1, \infty)$ bulanıklık olarak adlandırılan ağırlıklandırma sabitidir. m 'ye küçük değerler verilirse bulanıklık azalır, büyük değerler verilirse bulanıklık artar. Genellikle m 'ye "2" değeri verilir. Eşitlik (3.7)'de verilen amacı minimum yapan küme merkezleri (v_i) ve üyelikler (u_{ij})

$$v_i = \frac{\sum_{j=1}^n x_j (u_{ij})^m}{\sum_{j=1}^n (u_{ij})^m}, \quad i=1,2,\dots,c \quad (3.8)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{(d_{ij})^2}{(d_{ik})^2} \right)^{\frac{1}{m-1}}}, \quad 1 \leq i \leq c, 1 \leq j \leq n \quad (3.9)$$

biçiminde tanımlanır (Klir ve Yuan 1995, Dave ve Krishnapuram 1996, Pal ve Bezdek 1997, Wang 1997).

3.4. Robust Kümeleme Yöntemleri

FCM algoritmalarının sıfır kırılma noktasına sahip olduğu yani tek bir aykırı değer modelini tümüyle değiştirebileceği belirtilmiştir. En küçük kareler yöntemi ya da pek çok kümeleme algoritması gürültüye karşı güçlü değildir. Veri kümesinde gürültü olduğunda FCM tipi algoritmalar pratik değildir. Bu nedenle gürültüden etkilenmeyen robust kümeleme yöntemleri önerilmiştir (Dave 1993, Dave ve Krishnapuram 1997).

Robust kümeleme yöntemleri,

- I. Gürültü kümeleme yöntemi,

- II. Olanaklı c - ortalamaya dayalı kümeleme yöntemi,
- III. Karma bulanık olanaklı c -ortalamaya dayalı kümeleme,

olmak üzere 3 başlıkta sınıflandırılabilir. Bu sınıflandırmalarda tanımlanan yöntemler alt kesimde ayrıntılı olarak açıklanacaktır.

3.4.1. Gürültü Kümeleme Yöntemi

Gürültüye karşı daha fazla dirençli parametre tahminleri yapmak için FCM'in amacını yeniden düzenlemeye dayanan ilk yöntem Ohashi (1984) tarafından önerilmiştir. Ohashi (1984) Eşitlik (3.7)'de verilen amaç fonksiyonunu düzenleyerek aykırı değer sınıfı fikrini ileri sürmüştür. Bu durumda amaç fonksiyonu

$$J(B, U; X) = \lambda \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m (d_{ij})^2 + (1 - \lambda) \sum_{j=1}^n (u_{*j})^m \quad (3.10)$$

olarak tanımlanır. Burada u_{*j} aykırı değer sınıfındaki x_j noktasının üyeliği, λ ise önceden belirlenen bir sabittir (Dave ve Krishnapuram 1997, Nasraoui 2004).

Daha sonra Dave (1991) gürültü kümeleme (Noise clustering (NC)) yöntemini ileri sürmüştür (Dave ve Krishnapuram 1996). NC tekniğinde, gürültü tüm veri noktasından δ sabit uzaklığına sahip olarak tanımlanır. NC'de x_j 'nin üyeliği

$$u_{*j} = 1 - \sum_{i=1}^c u_{ij} \quad (3.11)$$

biçiminde tanımlanır. Gürültü sınıfındaki u_{c_j} üyelik tanımlamasında Eşitlik (3.11) kullanıldığında FCM'in alışılmış üyelik kısıtı gerekmez. Böylece iyi sınıf için üyelik kısıtı

$$0 < \sum_{i=1}^c (u_{ij}) \leq 1 \quad (3.12)$$

olarak yumuşatılır. Eşitlik (3.12)'de verilen kısıt iyi kümelerde gürültü noktalarının küçük üyelik değerlerine izin verir. Amaç fonksiyonu

$$J(B, U, X) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m (d_{ij})^2 + \sum_{j=1}^n \delta^2 \left(1 - \sum_{i=1}^c u_{ij} \right)^m \quad (3.13)$$

biçiminde tanımlanır. Eşitlik (3.13)'de verilen amaç fonksiyonun minimum yapılmasıyla üyelikler

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{(d_{ij})^2}{(d_{kj})^2} \right)^{\frac{1}{(m-1)}} + \left(\frac{(d_{ij})^2}{\delta^2} \right)^{\frac{1}{(m-1)}}}$$

olarak elde edilir. Eğer $\lambda = \frac{1}{1+\delta^2}$ olarak alınırsa Eşitlik (3.10) ve Eşitlik (3.13) aynıdır. NC yaklaşımının temel avantajı FCM algoritmasının robust şekli olması ve δ için uygun bir değer bulunarak FCM algoritmasının yerine kullanılabilir olmasıdır (Dave 1993, Dave ve Krishnapuram 1996, Dave ve Sen 1998).

3.4.2. Olanaklı c Ortalamaya Dayalı Kümeleme Yöntemi

Krishnapuram ve Keller (1993), gürültü probleminin zorluğunu ortadan kaldırmak için FCM’de kullanılan üyelik fonksiyonu için tanımlanan kısıtı yeniden düzenleyerek olanaklı c ortalamaya (possibilistic c -means (PCM)) dayalı kümeleme yöntemini ileri sürmüşlerdir.

PCM’de kümeyi temsil etmeyen gözlemlerin küçük üyelik değerine sahip olması istenirken kümeyi temsil eden gözlemlerin mümkün olduğunca yüksek üyelik değerine sahip olması istenir. Bu durumda amaç fonksiyonu

$$J(B, U; X) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m (d_{ij})^2 + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - u_{ij})^m \quad (3.14)$$

olarak tanımlanır. Eşitlik (3.14)’de ikinci terimin mümkün olduğunca büyük üyeliğe (u_{ij}) sahip olması istenirken ilk terim için uzaklığın mümkün olduğunca küçük olması istenir.

Uygulamada

$$\eta_i = K \frac{\sum_{j=1}^n (u_{ij})^m (d_{ij})^2}{\sum_{j=1}^n (u_{ij})^m} \quad K > 0, i=1,2,\dots,c \quad (3.15)$$

olduğunda iyi sonuçlar vermektedir. Burada K genellikle “1” alınır. Ayrıca Eşitlik (3.15) yerine

$$\eta_i = \frac{\sum_{x_j \in (\pi_i)_\alpha} (d_{ij})^2}{|(\pi_i)_\alpha|}, \quad i=1,2,\dots,c \quad (3.16)$$

kullanılabilir. Burada $(\pi_i)_\alpha$, π_i 'nin uygun α kesmesidir.

η_i değeri tüm iterasyonlar için sabit ya da her bir iterasyonda değişebilir. En iyi yaklaşım Eşitlik (3.15) kullanılarak bir başlangıç bulanık parçalanmaya dayanan η_i hesaplamak ve algoritma yakınsadıktan sonra Eşitlik (3.16) kullanılarak η_i için daha kesin değerler hesaplamaktır. Ayrıca kümelerin yapısı bilindiğinde η_i saptanmış bir ön bilgi (priori)de olabilir.

Eşitlik (3.14)'in minimum yapılmasıyla elde edilen üyelik

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i} \right)^{\frac{1}{m-1}}} \quad i=1, \dots, c; \quad j=1, \dots, n$$

biçiminde tanımlanır (Krishnapuram ve Keller 1993, Krishnapuram ve Keller 1996, Barni vd 1996, Dave ve Krishnapuram 1997).

NC yaklaşımındaki δ^2 gürültü uzaklığı ile PCM yaklaşımındaki η_i ağırlığı benzerdir. Fakat PCM her bir küme için ağırlığın farklı olması avantajına sahiptir. NC yaklaşımında bir gürültü sınıfı varken PCM yaklaşımında c gürültü sınıfı vardır. $c=1$ olduğunda NC ve PCM yaklaşımları $\delta^2 = \eta$ için aynıdır.

3.4.3. Karma Bulanık Olanaklı c Ortalamaya Dayalı Kümeleme

Pal ve Bezdek (1997) hem bulanık kümelemedeki üyeliklerin hem de bir gözlemin bir kümeye ne kadar uygun olduğunu gösteren küme özelliğini (typicality (t_y)) içeren karma bulanık olanaklı c ortalamaya (mixed c -means (FPCM)) dayalı kümeleme modelini tanımlamışlardır. FPCM'de amaç

$$J(B, U; X) = \sum_{i=1}^c \sum_{j=1}^n \left((u_{ij})^m + (t_{ij})^\eta \right) (d_{ij})^2 \quad (3.17)$$

olarak tanımlanır. (3.17) eşitliği aşağıda verilen kısıtlar altında minimum yapılmalıdır.

$$u_{ij} = \left(\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{(m-1)}} \right)^{-1}$$

$$t_{ij} = \left(\sum_{k=1}^n \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{(\eta-1)}} \right)^{-1}$$

$$v_i = \frac{\sum_{j=1}^n \left((u_{ij})^m + (t_{ij})^\eta \right) x_j}{\sum_{j=1}^n \left((u_{ij})^m + (t_{ij})^\eta \right)}$$

$$\sum_{i=1}^c u_{ij} = 1, \quad \sum_{j=1}^n t_{ij} = 1$$

$$m > 1, \eta > 1, u_{ij} \geq 0, t_{ij} \leq 1 \quad i=1, 2, \dots, c; \quad j=1, 2, \dots, n$$

$T = [t_{ij}]$ olmak üzere, $\sum_{j=1}^N t_{ij} = 1$ kısıtı T 'nin her bir sütun toplamının "1" olmasını

gerektirir. Böylece T 'lerin küme üyelikleri

$$\sum_{j=1}^n t_{ij} = 1, \quad \sum_{i=1}^c t_{ij} > 0$$

olmalıdır. Daha önceki yapılan çalışmalardan FPCM’de η ’nün [3, 5] aralığında olmasının iyi olacağı önerilmiştir (Pal ve Bezdek 1997).

4. REGRESYON ÇÖZÜMLEMESİ

4.1 Giriş

Bir doğrusal regresyon modeli,

$$y_i = \sum_{j=1}^p x_{ij} \beta_j + e_i \quad i=1,2,\dots,n \quad j=1,2,\dots,p \quad (4.1)$$

olarak tanımlanır. Burada y_i , i 'inci gözlem için bağımlı değişken değeri, $x_{i1}, x_{i2}, \dots, x_{ip}$, p bağımsız değişken değeri, e_i rassal hata miktarı ve $\beta_1, \beta_2, \dots, \beta_p$ regresyon katsayılarıdır. Eşitlik (4.1)'de tanımlanan regresyon modeli matris gösterimi ile

$$\underline{Y} = X \underline{\beta} + \underline{\varepsilon}$$

ya da

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

dir. Burada $\underline{Y}$: $n \times 1$ boyutlu bağımlı değişken için gözlemlerin vektörü, X ; $[n \times (m+1)]$ boyutlu bağımsız değişkenlere ilişkin matris, $\underline{\beta}$; $[(m+1) \times 1]$ boyutlu regresyon katsayılarının vektörü ve $\underline{\varepsilon}$: $(n \times 1)$ boyutlu hata vektörüdür.

En küçük kareler tahmin edicisi $\hat{\underline{\beta}}$ 'yi bulmak için

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \hat{\beta}_j \right)^2 = \varepsilon' \varepsilon = (\underline{Y} - \underline{X} \underline{\beta})' (\underline{Y} - \underline{X} \underline{\beta}) \quad (4.2)$$

eşitliğinin en küçüklenmesi gerekir. (4.2) eşitliğinden

$$\underline{\hat{\beta}} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{Y} \quad (4.3)$$

biçiminde bulunur. $\hat{\beta}$ 'nın varyansı

$$V(\hat{\beta}) = \frac{1}{n-p} \left(\sum_{i=1}^n e_i^2 \right) (\underline{X}' \underline{X})^{-1}$$

olarak tanımlanır (Huynh 1982, Alpar 1997, Apaydın vd 2002).

Regresyon çözümlemesindeki model bozukluklarını araştırmak ve aykırı değeri belirlemek için basit ve etkin bir yöntem artıkların incelenmesidir. *i.* artık

$$e_i = y_i - \hat{y}_i \quad (i = 1, 2, \dots, n)$$

olarak tanımlanır. Artıkların toplamı sıfır olmakla birlikte varyansı örneklemden örnekleme değışir ve e_i 'lerin sıfır ortalama ve s^2 varyansı ile normal dağıldıkları varsayılır. s^2 varyansının değışkenlik göstermesi sorunu artıkların standartlaştırılması ile giderilmeye çalışılır. Denek sayısının fazla olması durumunda standartlaştırılmış artıklar sıfır ortalama, bir standart sapma ile yaklaşık normal dağılırlar ve artıklar birbirleri ile ilişkili değildir. Standartlaştırılmış artıkların %95'i çoğunlukla [-2, +2] sınırları arasında değışir (Alpar 1997).

Klasik yöntemlerde, yapılan incelemeler sonucunda eğer veri kümesinde aykırı değer var ise bu gözlem veri kümesinden çıkartılarak işlemlere devam edilir. Ancak bazı çalışmalarda gözlemin çıkartılması mümkün değildir. Aykırı değerinin olduğu veri kümesi üzerinden çözümlemelerin yapılması gerekir. Bu durumda robust yöntemler kullanılır (Dave ve Krishnapuram 1997).

Alt kesimlerde Robust regresyon çözümlemesi ve Bulanık regresyon çözümlemesi ayrıntılı olarak verilecektir.

4.2. Robust Regresyon Çözümlemesi

Veri kümesinde aykırı değer olması durumunda, elde edilen regresyon modeli aykırı değerinin etkisiyle aykırı değer dışındaki gözlemlerden uzaklaşır. Aykırı değer dışındaki gözlemlerin artıkları büyür. Robust çözümlemede, bu sapmaların algoritmanın performansını çok etkilemediği varsayılır. Robust regresyon çözümlemesi ile aykırı değer olması durumunda regresyon model tahmininde EKK yöntemine göre daha az etkilenen parametre tahminleri elde edilir.

Huber (1981)'e göre bir robust yöntem aşağıda tanımlanan üç özelliği sağlamalıdır.

1. Varsayılan modelde iyi bir hız ve etkinliğe sahiptir,
2. Model varsayımlarından küçük sapmalar performansı küçük miktarda bozar,
3. Model varsayımlarından büyük sapmalar çok büyük bozulmaya neden olmaz (Dave ve Krishnapuram 1997, Nasraoui 2004).

Veri kümesinde aykırı değer olması durumunda robust yöntemler kullanılmadığında,

1. Bazı kurallar uygulanarak veri temizlenir,
2. Geri kalan veriye klasik tahminler ve test yöntemleri uygulanır.

Ancak bu adımlarda tanımlanan süreçte aşağıda verilen sakıncalar ortaya çıkabilir ve robust yöntemler kadar etkin çözümler vermeyebilir:

1. Çok değişkenli regresyon çözümlemesinde aykırı değerleri tanımlamak zordur ve parametreler için robust tahminler elde edilemez,
2. Hatalara sahip veri grubu normal olsa bile temizlenmiş veri normal olmaz. Böylece klasik teori temizlenmiş veriye uygulanamaz,
3. Küçük örneklerde verileri çıkarma işlemi sakıncalı olabilir (Huber 1981, Candan 1995).

Robust tahmin ediciler; M-tahmin edicileri, L-tahmin edicileri ve R-tahmin edicileri olmak üzere üç başlıkta toplanırlar. M tahmin edicileri en çok olabilirlik tahmin edicilerine benzer ve verilerin normal dağıldığını varsayarlar. L tahmin edicileri sıra istatistiklerinin doğrusal birleşimleridir. R tahmin edicileri Rank testlerine dayanırlar.

M-tahmin edicileri işlerlik ve uygunluk açısından daha çok tercih edilirler (Stephenson 2000). Alt kesimde M tahmin edicileri ayrıntılı olarak verilecektir.

4.2.1. M Tahmin Edicileri

Bu kesimde literatürde yaygın olarak kullanılan Huber, Hampel, Andrews ve Tukey M yöntemleri ayrıntılı olarak verilecektir.

M tahmin edicileri artıkların kareleri toplamını minimum yapmak yerine artıkların fonksiyonunu minimum yapar. Regresyon katsayıları,

$$\sum_{i=1}^n \rho \left[(y_i - \sum_{j=1}^p x_{ij} \hat{\beta}_j) / d \right] \quad (4.4)$$

toplamı minimum yapılarak elde edilir. Burada $\rho(\cdot)$ amaç fonksiyonudur. Eşitlik (4.4)'ün $\hat{\beta}_j$ 'ya göre türevi alınıp sıfıra eşitlenirse,

$$\sum_{i=1}^n x_{ij} \psi \left[(y_i - \sum_{j=1}^p x_{ij} \hat{\beta}_j) / d \right] = 0 \quad j = 1, \dots, p \quad (4.5)$$

p denklem sistemi için regresyon katsayıları elde edilir. Burada $\psi(\cdot)$ fonksiyonu $\rho(\cdot)$ fonksiyonunun türevidir. Eşitlik (4.5)'de

$$d = \text{medyan} \left| r_i - \text{medyan}(r_i) \right| / (0.6745) \quad (4.6)$$

olarak tanımlanır. Burada $r_i = y_i - \sum_{j=1}^p x_{ij} \hat{\beta}_j$ 'dir.

Veri aykırı değer içerdiğinde standart sapma değişimin iyi bir ölçüsü değildir. Bu nedenle değişimin robust ölçümleri kullanılmalıdır. Değişimin robust ölçüsü (4.6) eşitliğinde tanımlanan d 'dir. d 'nin paydası mutlak sapmaların medyanı (the Median of the Absolute Deviations, MAD) olarak tanımlanır. Eğer n geniş ve örneklem normal dağılımdan geliyorsa d standart sapma (σ)'ya yaklaşır. Örneklem standart sapması s aykırı değere karşı çok etkili ve robust değildir. Bu nedenle s , d değeri olarak kullanılamaz. Eğer σ (ya da d) biliniyorsa normallik varsayımları altında asimtotik etkinliği % 95'den daha büyüktür ve çok ağır kuyruklu durumlarda performansı oldukça iyidir (Hogg 1979).

Huber'in ρ fonksiyonu,

$$\rho(z) = \begin{cases} \frac{z^2}{2} & |z| \leq k \\ k|z| - \frac{k^2}{2} & |z| > k \end{cases} \quad (4.7)$$

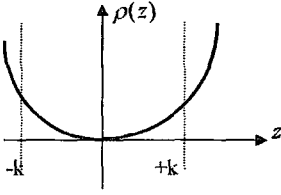
biçiminde tanımlanır. Burada k ifadesi ayar sabiti (tuning constant) olarak ifade edilir ve $k=1.5$ değerini alır. Eşitlik (4.7)'in türevi alınırsa,

$$\psi(z) = \begin{cases} -k & z < -k \\ z & |z| \leq k \\ k & z > k \end{cases} \quad (4.8)$$

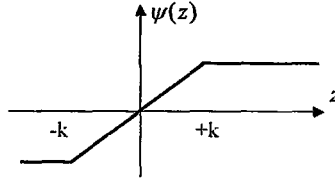
fonksiyonu elde edilir. Eşitlik (4.8)'de aykırı değere genellikle sıfır ya da sıfıra çok yakın ψ ağırlıkları verilir. Bu nedenle ψ “sıfıra geri azalan” (redescending to zero) olarak nitelendirilir. Standartlaştırılmış artıklar $z = r_i / d$ olarak tanımlanır. Ağırlık fonksiyonu (W), ψ fonksiyonunun z 'ye bölünmesiyle elde edilir. Huber ağırlık fonksiyonu (W),

$$W = \begin{cases} 1 & |z| \leq k \\ \frac{k}{z} & |z| > k \end{cases}$$

olarak tanımlanır. Huber ρ fonksiyonu grafiği Şekil 4.1'de ve ψ fonksiyonu grafiği Şekil 4.2'de verilmiştir (Hogg 1979, Huynh 1982, Hampel vd 1986, Candan 1995, Ata 1995, Dave ve Krishnapuram 1997).



Şekil 4.1. Huber ρ fonksiyonu



Şekil 4.2. Huber ψ Fonksiyonu

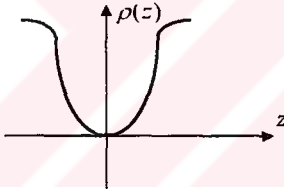
Hampel ρ , ψ ve W fonksiyonu,

$$\rho(z) = \begin{cases} \frac{z^2}{2} & 0 < |z| \leq a \\ a|z| - \frac{a^2}{2} & a < |z| \leq b \\ -\frac{a}{2(c-b)}(c-z)^2 + \frac{a}{2}(b+c-a) & b < |z| \leq c \\ \frac{a}{2}(b+c-a) & c < |z| \end{cases}$$

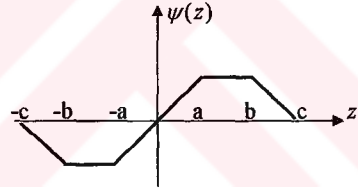
$$\psi(z) = \begin{cases} |z| & 0 < |z| \leq a \\ a \operatorname{sgn}(z) & a < |z| \leq b \\ a \left(\frac{c-|z|}{c-b} \right) \operatorname{sgn}(z) & b < |z| \leq c \\ 0 & c < |z| \end{cases}, \quad \operatorname{sgn}(z) = \begin{cases} +1, & z > 0 \\ 0, & z = 0 \\ -1, & z < 0 \end{cases}$$

$$W(z) = \begin{cases} 1 & 0 < |z| \leq a \\ \frac{a}{z} \operatorname{sgn}(z) & a < |z| \leq b \\ \frac{a}{z} \left(\frac{c-|z|}{c-b} \right) \operatorname{sgn}(z) & b < |z| \leq c \\ 0 & c < |z| \end{cases}$$

biçiminde tanımlanır. Genellikle sabitlerin değerleri $a=1.7$, $b=3.4$ ve $c=8.5$ olarak seçilir. Hampel ρ fonksiyonu grafiği Şekil 4.3'de ve ψ fonksiyonu grafiği Şekil 4.4'de verilmiştir (Hogg 1979, Huynh 1982, Candan 1995, Ata 1995, Dave ve Krishnapuram 1997).



Şekil 4.3. Hampel ρ fonksiyonu



Şekil 4.4. Hampel ψ fonksiyonu

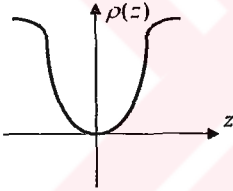
Andrews ise ρ , ψ ve W fonksiyonunu,

$$\rho(z) = \begin{cases} k^2 \left(1 - \cos \frac{|z|}{k} \right) & |z| \leq k\pi \\ 2k^2 & |z| > k\pi \end{cases}$$

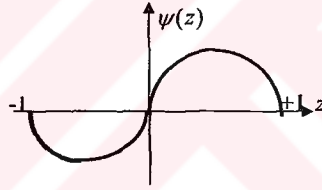
$$\psi(z) = \begin{cases} \sin(z/k) & |z| \leq k\pi \\ 0 & |z| > k\pi \end{cases}$$

$$W(z) = \begin{cases} \frac{\sin(z/k)}{z} & |z| \leq k\pi \\ 0 & |z| > k\pi \end{cases}$$

olarak tanımlanmıştır, burada $k=1.5$ ya da $k=2.1$ alınır. Andrews ρ fonksiyonu grafiği Şekil 4.5’de ve ψ fonksiyonu grafiği Şekil 4.6’da verilmiştir (Hogg 1979, Huynh 1982, Candan 1995, Ata 1995, Dave ve Krishnapuram 1997).



Şekil 4.5. Andrews ρ fonksiyonu



Şekil 4.6. Andrews ψ fonksiyonu

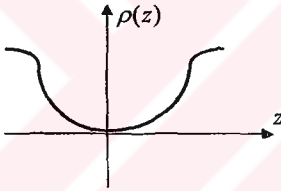
Tukey’in iki ağırlıklı tahmini için ρ , ψ ve W fonksiyonu ise,

$$\rho(z) = \begin{cases} \frac{1}{6} \left(1 - \left(1 - \left(\frac{z}{k} \right)^2 \right)^3 \right) & |z| \leq k \\ \frac{1}{6} & |z| > k \end{cases}$$

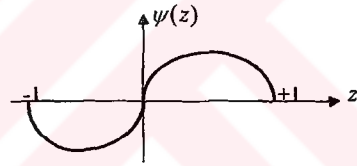
$$\psi(z) = \begin{cases} z(1 - (z/k)^2)^2 & |z| \leq k \\ 0 & |z| > k \end{cases}$$

$$W(z) = \begin{cases} \left(1 - (z/k)^2\right)^2 & |z| \leq k \\ 0 & |z| > k \end{cases}$$

biçiminde tanımlanır. k , 5 ya da 6 olarak seçilir. Tukey ρ fonksiyonu grafiği Şekil 4.7’de ve ψ fonksiyonu grafiği Şekil 4.8’de verilmiştir (Hogg 1979, Huber 1981, Huynh 1982, Candan 1995, Ata 1995, Dave ve Krishnapuram 1997).



Şekil 4.7. Tukey ρ fonksiyonu



Şekil 4.8. Tukey ψ fonksiyonu

Huber, Hampel, Andrews ve Tukey M tahmin edicileri için standart sapma,

$$\frac{\frac{d^2}{n-p} \sum (\psi(z_i))^2}{\left(\frac{1}{n} \sum \psi'(z_i)\right)^2} (XX)^{-1}$$

matrisinin köşegen elemanlarının kare köküdür (Hogg 1979, Huynh 1982).

4.2.2. En Küçük Medyan Kare Yöntemi

En küçük medyan kare (Least Median of Squares, LMS) yöntemi, artık kareler medyanını en küçük yapmayı amaçlar. Bu yöntemde amaç fonksiyonu,

$$\min_{\hat{\beta}} \text{med}_i r_i^2$$

biçiminde tanımlanır. Burada r_i , i 'inci gözlemin artığı olarak tanımlanır.

En küçük medyan kare yönteminde i 'inci gözlemin ağırlığı,

$$w_i = \begin{cases} 1 & |r_i / s_0| \leq 2.5 \\ 0 & |r_i / s_0| > 2.5 \end{cases}$$

biçiminde hesaplanır. Burada n toplam gözlem sayısı, p değişken sayısı olmak üzere ,

$$s_0 = 1.4826 \left(1 + \frac{5}{n-p} \sqrt{\text{med}_i r_i^2} \right)$$

olarak tanımlanır. LMS yöntemi, Rousseeuw ve Leroy (1987) tarafından verilen yeniden örnekleme algoritması ile hesaplanır (Rousseeuw ve Leroy 1987, Candan 1995).

4.2.3. En Küçük Medyan Kareye Dayanan Yeniden Ağırlıklandırılmış En Küçük Kareler Yöntemi

Ağırlıklandırılmış artık kare toplamını minimum yapmaya çalışan en küçük medyan kareye dayanan yeniden ağırlıklandırılmış en küçük kareler (Reweighted Least Squares Based on the LMS, RLS) yönteminde amaç fonksiyonu,

$$\min_{\beta} \sum_i w_i r_i^2$$

biçiminde tanımlanır.

RLS yönteminde ağırlıklar,

$$w_i = \begin{cases} 1 & |r_i / \hat{\sigma}| \leq 2.5 \\ 0 & |r_i / \hat{\sigma}| > 2.5 \end{cases}$$

biçiminde hesaplanır. Burada

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n w_i r_i^2}{\sum_{i=1}^n w_i - p}}$$

olarak tanımlanır (Rousseuw ve Leroy 1987, Candan 1995).

4.3. Bulanık Regresyon Çözümlemesi

Bulanık regresyon çözümlemesi klasik regresyon çözümlemesinin bulanık genişlemesi olarak düşünülebilir. Çeşitli alanlarda yaygın olarak çalışılmış ve uygulanmıştır. Genel olarak bulanık regresyon modellerinin çözümlemesi iki kategoride toplanır. Bunlardan birincisi Tanaka vd (1982)'nin doğrusal programlama yaklaşımına, ikincisi ise bulanık en küçük kareler yaklaşımına dayanır.

Tanaka vd (1982) bulanık modele sahip doğrusal regresyon çözümlemesindeki ilk çalışmayı önermişlerdir. Tanaka vd (1982) modelinde amaç, kısıtlar altında bulanık regresyon katsayılarının genişlikleri toplamını minimum yapmaktır. Bulanık doğrusal regresyon

$$y_i^* = A_0 + A_1 x_{i1} + A_2 x_{i2} + \dots + A_p x_{ip}$$

olarak tanımlanır. Burada A_j ($j=1,2,\dots,p$) parametresi α_j merkezli ve c_j genişlikli bulanık sayıdır. Matris gösterimi ile $A^T = (A_0, A_1, \dots, A_p)$ 'dır. $\{\alpha^T, c^T\}$, merkezi $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)^T$ ve genişliği $c = (c_0, c_1, \dots, c_p)^T$ olan bulanık sayıyı gösterir. A_j ($j=1,2,\dots,p$)'nin üyelik fonksiyonu,

$$\mu_{A_j}(a_j) = \begin{cases} 1 - \frac{|\alpha_j - a_j|}{c_j} & \alpha_j - c_j < a_j < \alpha_j + c_j \\ 0 & \text{dd} \end{cases} \quad (4.9)$$

dır. Burada $c_j > 0$ dır. $\mu_{A_j}(a_j)$, A_j bulanık kümesindeki a_j nin üyeliğini gösterir. Tanaka modeli,

$$\begin{aligned} \text{Min } J &= \sum_{i=1}^n c^T |x_i| \\ \text{Kısıt } x_i^T \alpha + (1-h)c^T |x_i| &\geq y_i + (1-h)e_i \\ -x_i^T \alpha + (1-h)c^T |x_i| &\geq -y_i + (1-h)e_i \\ 0 \leq h < 1 \quad c &\geq 0 \end{aligned} \tag{4.10}$$

biçiminde tanımlanır. Burada h karar verici tarafından belirlenen eşik düzeyidir. Tanaka vd (1982) tarafından ileri sürülen minimizasyon probleminin ölçek bağımlı olduğunu Jo'zsef (1992) ileri sürmüştür. Tanaka vd (1982) modelinde bağımsız değişken X kesin (crisp), bağımlı değişken merkezi y_i , yayılması e_i olan bulanık sayı ($Y = (y_i, e_i)$), parametreler merkezi a_j , yayılması c_j olan ve $A = (a_j, c_j)$ olarak tanımlanan bulanık sayıdır (Redden ve Woddall 1996).

Eşitlik (4.10) parametrelerin yayılmalarının negatif olmasına izin vermez. Değişkenler arasında negatif ilişki olması durumunda yayılmaların negatif değer almaması yalnızca yayılmaların değil, merkezlerin de yanlış yorumlamasına neden olur. Bu yanlış yorumun üstesinden gelebilmek için Chang ve Lee (1993) parametrelerin negatif yayılmalarına izin veren modeli tanımlamışlardır. Bu model eşitlik (4.9)'de tanımlanan üyelik fonksiyonuna sahiptir. Chang ve Lee (1993) modelinde tahmin edilen Y 'nin pozitif yayılması koşulu altında bulanık parametrelerin yayılmalarının işareti belirtilmez.

Dimond (1988) bulanık en küçük kareler yaklaşımını önermiştir. Dimond (1988) x 'in kesin sayı ve $Y_i = (y_i, \underline{\eta}_i, \bar{\eta}_i)$ simetrik üçgensel bulanık sayı olması durumunda bulanık regresyon modelini,

$$Y = A + xB \quad (4.11)$$

olarak tanımlamıştır. Burada $A = (a, \underline{\alpha}, \bar{\alpha})$ ve $B = (b, \underline{\beta}, \bar{\beta})$ simetrik üçgensel bulanık parametrelerdir. (4.11) eşitliğindeki parametre tahminlerini yapabilmek için en küçük kareler optimizasyon problemi,

$$\text{Minimum} \quad r(A, B) = \sum d(A + x_i B, Y_i)^2 \quad (4.12)$$

biçimindedir. Burada

$$\begin{aligned} d(A + x_i B, Y_i) = & (a + bx_i - \underline{\alpha} - \underline{\beta}x_i - y_i + \underline{\eta}_i)^2 + (a + bx_i + \bar{\alpha} + \bar{\beta}x_i - y_i - \bar{\eta}_i)^2 \\ & + (a + bx_i - y_i)^2 \end{aligned} \quad (4.13)$$

olarak tanımlanır. Eşitlik (4.13)'de verilen d fonksiyonunun a ve b 'ye göre türevi alınır ve $\frac{\partial d}{\partial a} = 0$, $\frac{\partial d}{\partial b} = 0$ denklem sistemi çözülürse,

$$a = \hat{y} - b\hat{x}$$

$$b = \frac{K}{T^2}$$

bulunur. Eşitlik (4.13)'de verilen d fonksiyonunun α ve β 'ye göre türevi alınır ve $\frac{\partial d}{\partial \alpha} = 0$, $\frac{\partial d}{\partial \beta} = 0$ denklem sistemi çözülürse,

$$\alpha = \hat{\eta} - \beta \hat{x}$$

$$\beta = k / T^2$$

bulunur. Burada

$$K = \sum_i (x_i - \hat{x})(y_i - \hat{y})$$

$$T^2 = \sum_i (x_i - \hat{x})^2$$

$$\hat{x} = \frac{\sum_i x_i}{N}$$

$$\hat{y} = \frac{\sum_i y_i}{N}$$

$$k = \sum_i (x_i - \hat{x})(\eta_i - \hat{\eta})$$

dir.

Bulanık en küçük kareler yaklaşımı için $X_i = (x_i, \underline{x}_i, \bar{x}_i)$, $Y_i = (y_i, \underline{\eta}_i, \bar{\eta}_i)$ simetrik olmayan üçgensel bulanık sayıları ele alındığında,

$$Y = a + bX$$

modeli düşünülebilir. Burada a, b kesin sayılardır. Parametrelerin kesin olduğu model ele alındığında en küçük kareler optimizasyon problemi,

$$\text{Minimum } r(a, b) = \sum d(a + bX_i, Y_i)^2 \quad (4.14)$$

olarak tanımlanır. Bu durumda

$$d(a+bx_i, Y_i) = \left[a+bx_i - y_i - (b\underline{\xi}_i - \underline{\eta}_i) \right]^2 + \left[a+bx_i - y_i + (b\bar{\xi}_i - \bar{\eta}_i) \right]^2 + (a+bx_i - y_i)^2 \quad (4.15)$$

biçiminde tanımlanır. Eşitlik (4.15)'de verilen d fonksiyonunun, a ve b 'ye göre türevi alındığında,

$$\begin{aligned} \frac{\partial d}{\partial a} &= 3Na + b \sum [3x_i + \delta(x_i)] - \sum [3y_i + \delta(Y_i)] = 0 \\ \frac{\partial d}{\partial b} &= a \sum [3x_i + \delta(x_i)] + b \left[x_i^2 + \left(x_i - \underline{\xi}_i \right)^2 + \left(x_i + \bar{\xi}_i \right)^2 \right] \\ &\quad - \sum \left[\left(x_i - \underline{\xi}_i \right) \left(y_i - \underline{\eta}_i \right) + \left(x_i + \bar{\xi}_i \right) \left(y_i + \bar{\eta}_i \right) + x_i y_i \right] = 0 \end{aligned}$$

denklem sistemi elde edilir. Bu sistemin çözülmesiyle a ve b parametreleri elde edilir. Burada tüm x_i ($i=1,2,...,n$) için

$$\delta(X) = \bar{\xi} - \underline{\xi}$$

$$\delta(Y) = \bar{\eta} - \underline{\eta}$$

olarak tanımlanır (Dimond 1988, Xu ve Li 2001, Yun-Hsi 2001).

Tahmin edilen basit regresyon modeli,

$$\hat{Y}_i = \tilde{A}_0 + \tilde{A}_1 X_i = (a_0, c_{0,L}, c_{0,R}) + (a_1, c_{1,L}, c_{1,R}) X_i$$

biçiminde tanımlanır. Burada $\tilde{A}_0$ ve $\tilde{A}_1$ bulanık parametrelerdir. $\tilde{A}_0$; merkezi a_0 , sol yayılması $c_{0,L}$, sağ yayılması $c_{0,R}$ olan L - R bulanık sayıdır. Benzer şekilde $\tilde{A}_1$; merkezi a_1 , sol yayılması $c_{1,L}$, sağ yayılması $c_{1,R}$ olan L - R bulanık sayıdır. Bulanık kümelerdeki aritmetik işlemler kullanılarak, tahmin edilen her bir üçgensel bulanık sayı,

$$\hat{Y}_i = (a_0 + a_1 X_i, c_{0,L} + c_{1,L} X_i, c_{0,R} + c_{1,R} X_i) \quad i = 1, 2, \dots, n$$

olarak tanımlanır. Bu durumda gözlenmiş $\tilde{Y}_i$; merkezi Y_i , sol yayılması $e_{i,L}$, sağ yayılması $e_{i,R}$ olan L - R bulanık sayıdır ve $\tilde{Y}_i = (Y_i, e_{i,L}, e_{i,R})$ olarak gösterilir. Burada α kesme düzeyinde tahmin edilen $\hat{Y}_i$ 'nin sol ve sağ sınırları,

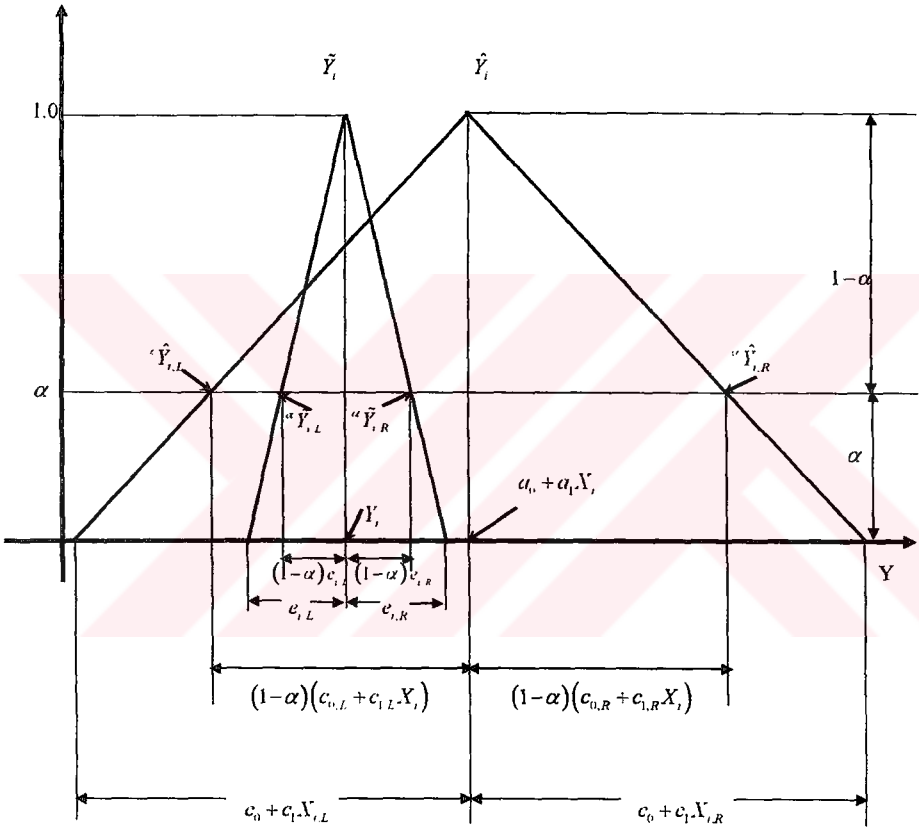
$$\begin{aligned} {}^\alpha \hat{Y}_{i,L} &= [a_0 - (1-\alpha)c_{0,L}] + [a_1 - (1-\alpha)c_{1,L}] X_i \\ &= (a_0 + a_1 X_i) - (1-\alpha)(c_{0,L} + c_{1,L} X_i) \\ {}^\alpha \hat{Y}_{i,R} &= [a_0 + (1-\alpha)c_{0,R}] + [a_1 + (1-\alpha)c_{1,R}] X_i \\ &= (a_0 + a_1 X_i) + (1-\alpha)(c_{0,R} + c_{1,R} X_i) \end{aligned}$$

olarak tanımlanır. Gözlemlenmiş $\tilde{Y}_i$ 'nin sol ve sağ sınırları ise,

$$\begin{aligned} {}^\alpha \tilde{Y}_{i,L} &= Y_i - (1-\alpha)e_{i,L} \\ {}^\alpha \tilde{Y}_{i,R} &= Y_i + (1-\alpha)e_{i,R} \end{aligned}$$

biçimindedir. Gözlenmiş $\tilde{Y}_i$ ve tahmin edilen $\hat{Y}_i$ arasındaki ilişkinin grafiksel gösterimi Şekil 4.9'da verilmiştir (Chang 2001).

Üyelik Değeri



Şekil 4.9. Gözlenmiş $\tilde{Y}_i$ ve tahmin edilen $\hat{Y}_i$ arasındaki ilişki

5. BULANIK ROBUST REGRESYON ÇÖZÜMLEMESİ

Klasik en küçük kareler yöntemi aykırı değerden etkilenir ve aykırı değer dışında kalan gözlemlerin artıklarının büyük olmasına neden olur. Bu nedenle veri kümesinde aykırı değer olması durumunda en küçük kareler yöntemi iyi sonuçlar vermemektedir. Aykırı değer olması durumunda regresyon çözümlemesinde parametre tahminleri ile ilgili oldukça çok çalışma yapılmış ve robust tahmin ediciler tanımlanmıştır. Son yıllarda yapılan çalışmalarda ise, bulanık tahminler de oldukça yaygındır.

Bulanık regresyon çözümlemesi ile ilgili ilk çalışma Tanaka vd (1982) tarafından ileri sürülmüştür. Ancak Tanaka modeli çoklu regresyon çözümlemesinde, merkezlerin sıfır olması gibi sakıncalara sahiptir. Model daha sonra yapılan çalışmalar tarafından eleştirilmiştir. Ayrıca gözlem sayısının çok fazla olması durumunda kısıt sayısı artmakta işlemler güçleşmektedir. Bu eleştirilere karşılık Tanaka ve Ishibuchi (1991) etkileşimli bulanık katsayıları elde etmek için bir yöntem geliştirmişlerdir. Bu yöntemin aykırı değerlere güçlü bir şekilde duyarlı olduğu Redden ve Woddall (1996) tarafından gösterilmiştir.

Sakawa ve Yano (1992) çok amaçlı bulanık regresyon çözümlemesi problemlerini formüle etmek için karar vericiyi tatmin edecek şekilde yinelemeli, karar vericiyle etkileşime dayanan, etkileşimli algoritmayı ileri sürmüşlerdir. Fakat bu algoritma da aykırı değerlere karşı çok fazla duyarlıdır.

Daha sonraki yıllarda yapılan çalışmalarda ise genellikle basit regresyon modeli ele alınmıştır. Karar vericiyle etkileşime dayanan modeller önerilmiş ve başlangıç parametreleri sezgisel olarak belirlenmiştir.

Bu çalışmada ise, bağımlı ve bağımsız değişkenlerin bulanık olduğu çoklu doğrusal regresyon modeli ele alınacaktır. Bu modellerin tahmin sürecinde

sezgiselliğe izin verilmeyecektir. Bulanık çoklu doğrusal regresyon modeli elde edildikten sonra veri kümesinde aykırı değer olması durumu incelenmiş ve bu amaçla üyeliklere bağlı olarak ağırlık matrisi tanımlanmıştır. Çoklu doğrusal regresyon modelinde bu ağırlıklar kullanılarak ağırlıklandırılmış bulanık regresyon çözümlemesi yapılarak aykırı değerden etkilenmeyen model tahminlerinin elde edilmesi amaçlanmıştır.

Üçüncü Bölüm’de tanımlanan robust kümeleme yöntemleri başlangıç parametrelerinin sezgisel belirlenmesi ve çözüm sürecinin uzun yinelemeli süreçlerle olması gibi sakıncalara sahiptir. Önerilen yöntemlerde tanımlanan üyelik fonksiyonu kullanılarak aykırı değer ve diğer gözlemler medyana olan uzaklıklara göre ayrı ayrı kümelenebilecektir. Bu durumda robust kümeleme yöntemlerindeki üyelikler ile tanımlanan üyelik fonksiyonundan elde edilen üyelikler eş değer olacaktır.

Bu çalışmada öncelikle bağımlı ve bağımsız değişkenlerin simetrik olmayan üçgensel bulanık sayı $(X_i = (x_i, \underline{x}_i, \overline{x}_i), Y_i = (y_i, \underline{y}_i, \overline{y}_i))$ ve çıktı parametrelerinin kesin sayı olması durumu için bulanık regresyon modeli bulanık çoklu doğrusal regresyon modeli olarak geliştirilecektir.

x ’in kesin sayı, $Y_i = (y_i, \underline{y}_i, \overline{y}_i)$ simetrik üçgensel bulanık sayı ve çıktı parametrelerinin simetrik üçgensel bulanık sayı olması durumu için, bulanık regresyon modelinde çoklu doğrusal regresyon modeli ele alınacaktır. Kesim 2.5’de verilen bulanık sayıların karşılaştırılması bulanık regresyon çözümlemesine uyarlanarak, artıkların karşılaştırılması OM indeksine göre yapılacaktır.

Veri kümesinde aykırı değer olması durumunda çoklu doğrusal regresyonda parametre tahminlerinin yapılabilmesi için

1. Girdi değişkenlerinin yani hem bağımlı hem de bağımsız değişkenlerin simetrik olmayan üçgensel bulanık sayı, çıktı parametrelerinin kesin sayı olduğu,
2. Bağımlı değişkenin simetrik üçgensel bulanık sayı, bağımsız değişkenlerin kesin sayı, çıktı parametrelerinin simetrik üçgensel bulanık sayı olduğu

durum için yeni iki yöntem önerilecektir.

5.1. $X = (x, \underline{\xi}, \bar{\xi})$ ve $Y = (y, \underline{\eta}, \bar{\eta})$ 'nin Simetrik Olmayan Üçgensel Bulanık Sayılar Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi

Bu kesimde girdi değişkenlerinin yani hem bağımlı hem de bağımsız değişkenlerin simetrik olmayan üçgensel bulanık sayı olması durumu için bulanık çoklu doğrusal regresyon çözümlemesi önerilecektir.

$X_1, X_2, \dots, X_p$ bağımsız değişkenleri merkezi x , sol ve sağ yayılımları sırasıyla $\underline{\xi}$ ve $\bar{\xi}$ olan simetrik olmayan üçgensel bulanık sayı $[X = (x, \underline{\xi}, \bar{\xi})]$ ve Y bağımlı değişkeni merkezi y , sol ve sağ yayılımları sırasıyla $\underline{\eta}$ ve $\bar{\eta}$ olan simetrik olmayan üçgensel bulanık sayı $[Y = (y, \underline{\eta}, \bar{\eta})]$ olsun. Bu durumda bulanık çoklu doğrusal regresyon modeli,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

olarak tanımlanır. Bu modeldeki $(\beta_0, \beta_1, \dots, \beta_p)$ parametrelerinin tahmini için, tanım 2.7’de verilen genişleme ilkesinden yararlanarak oluşturulan kısıtsız çok değişkenli optimizasyon problemi bir lemma olarak verildi.

Lemma: Çoklu doğrusal regresyon modeli durumunda (4.14) eşitliği ile tanımlanan optimizasyon problemi tanım 2.7 de verilen genişleme ilkesine göre,

$$P1: \quad \text{Min } r(a, b_1, b_2, \dots, b_p) = \sum d(a + b_1 X_{i1} + \dots + b_p X_{ip}, Y_i)^2$$

olarak genelleştirildi. Burada $a, b_1, \dots, b_p$ kesin sayılar, $X = (x, \underline{\xi}, \bar{\xi})$, ve $Y = (y, \underline{\eta}, \bar{\eta})$ simetrik olmayan üçgensel bulanık sayılar olmak üzere

$$\begin{aligned} d(a + b_1 X_{i1} + \dots + b_p X_{ip}, Y_i) = & \left[a + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} - y_i - (b_1 \underline{\xi}_{i1} + b_2 \underline{\xi}_{i2} + \dots + b_p \underline{\xi}_{ip} - \underline{\eta}_i) \right]^2 \\ & + \left[a + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} - y_i + (b_1 \bar{\xi}_{i1} + b_2 \bar{\xi}_{i2} + \dots + b_p \bar{\xi}_{ip} - \bar{\eta}_i) \right]^2 \\ & + (a + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} - y_i)^2 \end{aligned}$$

olarak elde edildi. P1 probleminin $a, b_1, b_2, \dots, b_p$ ’ye göre kısmi türevleri alınıp sıfıra eşitlendiğinde,

$$\begin{aligned} \frac{\partial d}{\partial a} = & 3na + b_1 \left[\sum_{i=1}^n x_{i1} + \sum_{i=1}^n (x_{i1} - \underline{\xi}_{i1}) + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1}) \right] \\ & + b_2 \left[\sum_{i=1}^n x_{i2} + \sum_{i=1}^n (x_{i2} - \underline{\xi}_{i2}) + \sum_{i=1}^n (x_{i2} + \bar{\xi}_{i2}) \right] + \dots \\ & + b_k \left[\sum_{i=1}^n x_{ik} + \sum_{i=1}^n (x_{ik} - \underline{\xi}_{ik}) + \sum_{i=1}^n (x_{ik} + \bar{\xi}_{ik}) \right] \\ & - \sum_{i=1}^n y_i + \sum_{i=1}^n (y_i - \underline{\eta}_i) + \sum_{i=1}^n (y_i + \bar{\eta}_i) = 0 \end{aligned}$$

$$\begin{aligned}
\frac{\partial d}{\partial b_1} = & a \left[\sum_{i=1}^n x_{i1} + \sum_{i=1}^n (x_{i1} - \underline{\xi}_{i1}) + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1}) \right] \\
& + b_1 \left[\sum_{i=1}^n x_{i1}^2 + \sum_{i=1}^n (x_{i1} - \underline{\xi}_{i1})^2 + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1})^2 \right] \\
& + b_2 \left[\sum_{i=1}^n x_{i2} x_{i1} + \sum_{i=1}^n (x_{i1} - \underline{\xi}_{i1})(x_{i2} - \underline{\xi}_{i2}) + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1})(x_{i2} + \bar{\xi}_{i2}) \right] + \dots \\
& + b_k \left[\sum_{i=1}^n x_{ik} x_{i1} + \sum_{i=1}^n (x_{i1} - \underline{\xi}_{i1})(x_{ik} - \underline{\xi}_{ik}) + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1})(x_{ik} + \bar{\xi}_{ik}) \right] \\
& - \sum_{i=1}^n y_i x_{i1} + \sum_{i=1}^n (x_i - \underline{\xi}_{i1})(y_i - \underline{\eta}_i) + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1})(y_i + \bar{\eta}_i) = 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial d}{\partial b_2} = & a \left[\sum_{i=1}^n x_{i2} + \sum_{i=1}^n (x_{i2} - \underline{\xi}_{i2}) + \sum_{i=1}^n (x_{i2} + \bar{\xi}_{i2}) \right] \\
& + b_1 \left[\sum_{i=1}^n x_{i1} x_{i2} + \sum_{i=1}^n (x_{i1} - \underline{\xi}_{i1})(x_{i2} - \underline{\xi}_{i2}) + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1})(x_{i2} + \bar{\xi}_{i2}) \right] \\
& + b_2 \left[\sum_{i=1}^n x_{2i}^2 + \sum_{i=1}^n (x_{2i} - \underline{\xi}_{2i})^2 + \sum_{i=1}^n (x_{2i} + \bar{\xi}_{2i})^2 \right] + \dots \\
& + b_k \left[\sum_{i=1}^n x_{ki} x_{2i} + \sum_{i=1}^n (x_{ki} - \underline{\xi}_{ki})(x_{2i} - \underline{\xi}_{2i}) + \sum_{i=1}^n (x_{ki} + \bar{\xi}_{ki})(x_{2i} + \bar{\xi}_{2i}) \right] \\
& - \sum_{i=1}^n y_i x_{i2} + \sum_{i=1}^n (x_{i2} - \underline{\xi}_{i2})(y_i - \underline{\eta}_i) + \sum_{i=1}^n (x_{i2} + \bar{\xi}_{i2})(y_i + \bar{\eta}_i) = 0
\end{aligned}$$

$\vdots$
 $\vdots$
 $\vdots$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial h_k} = & a \left[\sum_{i=1}^n x_{ik} + \sum_{i=1}^n (x_{ik} - \underline{\xi}_{ik}) + \sum_{i=1}^n (x_{ik} + \bar{\xi}_{ik}) \right] \\
& + h_1 \left[\sum_{i=1}^n x_{i1} x_{ik} + \sum_{i=1}^n (x_{i1} - \underline{\xi}_{i1}) (x_{ik} - \underline{\xi}_{ik}) + \sum_{i=1}^n (x_{i1} + \bar{\xi}_{i1}) (x_{ik} + \bar{\xi}_{ik}) \right] \\
& + h_2 \left[\sum_{i=1}^n x_{i2} x_{ik} + \sum_{i=1}^n (x_{i2} - \underline{\xi}_{i2}) (x_{ik} - \underline{\xi}_{ik}) + \sum_{i=1}^n (x_{i2} + \bar{\xi}_{i2}) (x_{ik} + \bar{\xi}_{ik}) \right] + \dots \\
& + h_k \left[\sum_{i=1}^n x_{ik}^2 + \sum_{i=1}^n (x_{ik} - \underline{\xi}_{ik})^2 + \sum_{i=1}^n (x_{ik} + \bar{\xi}_{ik})^2 \right] \\
& - \sum_{i=1}^n y_i x_{ik} + \sum_{i=1}^n (x_{ik} - \underline{\xi}_{ik}) (y_i - \underline{\eta}_i) + \sum_{i=1}^n (x_{ik} + \bar{\xi}_{ik}) (y_i + \bar{\eta}_i) = 0
\end{aligned}$$

denklem sistemi elde edilir. Bu denklem sisteminin çözümü ile parametre tahminleri elde edilir. $\hat{\beta}$ tahmin edicisi matris gösterimiyle

$$\hat{\beta} = (X'X + A'A + B'B)^{-1} (X'Y + A'C + B'D) \quad (5.1)$$

biçiminde bulunur. Eşitlik (5.1)'de bulunan tahmin edici kullanılarak çok değişkenli regresyon modeli tahmin edildi. Burada

$$X = \begin{bmatrix} 1 & x_{11} & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \cdot & \cdot & x_{np} \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad C = \begin{bmatrix} y_1 - \underline{\eta}_1 \\ y_2 - \underline{\eta}_2 \\ \cdot \\ \cdot \\ y_n - \underline{\eta}_n \end{bmatrix} \quad D = \begin{bmatrix} y_1 + \bar{\eta}_1 \\ y_2 + \bar{\eta}_2 \\ \cdot \\ \cdot \\ y_n + \bar{\eta}_n \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & (x_{11} - \underline{\xi}_{11}) & \cdot & \cdot & (x_{1p} - \underline{\xi}_{1p}) \\ 1 & (x_{21} - \underline{\xi}_{21}) & \cdot & \cdot & (x_{2p} - \underline{\xi}_{2p}) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & (x_{n1} - \underline{\xi}_{n1}) & \cdot & \cdot & (x_{np} - \underline{\xi}_{np}) \end{bmatrix} \quad B = \begin{bmatrix} 1 & (x_{11} + \bar{\xi}_{11}) & \cdot & \cdot & (x_{1p} + \bar{\xi}_{1p}) \\ 1 & (x_{21} + \bar{\xi}_{21}) & \cdot & \cdot & (x_{2p} + \bar{\xi}_{2p}) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & (x_{n1} + \bar{\xi}_{n1}) & \cdot & \cdot & (x_{np} + \bar{\xi}_{np}) \end{bmatrix}$$

olarak tanımlandı. Modelde, gözlem sayısı $i=1,...,n$ ve değişken sayısı $j=1,...,p$ 'dir. Parametre tahminlerinin elde edilebilmesi için $(X'X + A'A + B'B)$ 'nin tersinin alınabilmesi gereklidir. Klasik regresyonda X matrisi için gerekli olan koşulların sağlanması durumunda $(X'X + A'A + B'B)$ 'nin tersi alınabilir.

Girdi değişkenlerinin yani hem bağımlı hem de bağımsız değişkenlerin simetrik olmayan üçgensel bulanık sayı, çıktı parametrelerinin kesin sayı ve veri kümesinde aykırı değer olması durumunda önerilen bulanık robust regresyon çözümlemesi için algoritma alt kesimde tanımlanacaktır.

5.1.1. $X=(x, \underline{\xi}, \bar{\xi})$ ve $Y=(y, \underline{\eta}, \bar{\eta})$ Simetrik Olmayan Üçgensel Bulanık Sayılar Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi için Algoritma

Bu algoritmada $X=(x, \underline{\xi}, \bar{\xi})$ ve $Y=(y, \underline{\eta}, \bar{\eta})$ simetrik olmayan üçgensel bulanık sayılar olması durumunda başlangıç parametreleri eşitlik (5.1)'den elde edildi. Her bir gözlemin ağırlıklı olarak modele katkıda bulunabilmesi için üçgensel üyelik fonksiyonu tanımlandı. Bu fonksiyondan üyelik dereceleri belirlenerek ağırlık matrisi (W) oluşturuldu. Ağırlık matrisi, köşegen elemanları üyeliklerden oluşan diagonal bir matristir. Ağırlık matrisi yardımıyla ağırlıklandırılmış bulanık en küçük kareler tahmin edicisi bulundu ve parametre tahminleri elde edildi. Ard arda bulunan $\hat{\beta}$ tahminleri arasındaki fark sifıra yaklaşıncaya kadar sürece devam edildi.

Çoklu doğrusal regresyon çözümlemesinde $X=(x, \underline{\xi}, \bar{\xi})$ ve $Y=(y, \underline{\eta}, \bar{\eta})$ simetrik olmayan üçgensel bulanık sayı, çıktı parametrelerinin kesin sayı ve veri kümesinde aykırı değer olması durumunda regresyon modelinin tahmin

edilmesinde önerilen bulanık robust yöntem için algoritmanın adımları aşağıda verilmiştir:

Adım 1: $X = (x, \underline{x}, \overline{x})$, $Y = (y, \underline{y}, \overline{y})$ simetrik olmayan üçgensel bulanık sayıları için parametre tahmini (5.1) eşitliğinden elde edilir.

Adım 2: $\hat{y}_i$ tahmin değerleri ve artıklar (e_i) hesaplanır.

Adım 3: Artıkların mutlak değerlerine göre medyayı belirlenir ve

$$d_i = \|abs(e_i) - medyan(abs(e_i))\| \quad i = 1, 2, \dots, n$$

uzaklığı hesaplanır. Burada $\| \cdot \|$ öklid uzaklığıdır.

Adım 4: Uzaklıklara bağlı olarak üyelik fonksiyonu

$$\mu(e) = \begin{cases} 1 & , \quad abs(e) \leq a \\ \frac{h - abs(e)}{h - a} & , \quad a < abs(e) < b \\ 0 & , \quad d.d \end{cases} \quad (5.2)$$

biçiminde tanımlandı. Burada;

$$a = medyan(d_i)$$

$$b = \max(d_i) + d$$

$$d = medyan|e_i - medyan(e_i)| / (0.6745)$$

olarak alındı.

Adım 5: Eşitlik (5.2)'de tanımlanan üyelik fonksiyonundan üyelik dereceleri belirlendi ve ağırlık matrisi (W) oluşturuldu. Ağırlık matrisi, köşegen elemanları üyeliklerden oluşan **diagonal** matristir. Buradan ağırlıklandırılmış bulanık en küçük kareler tahmin edicisi

$$\hat{\beta} = (X'WX + A'WA + B'WB)^{-1}(X'WY + A'WC + B'WD) \quad (5.3)$$

biçiminde bulundu. Bu tahmin edici kullanılarak parametre tahmini yapılır.

Adım 6: Eğer $|\hat{\beta}^{k+1} - \hat{\beta}^k| < \varepsilon$ ise durulur. Aksi durumda Adım 2' ye dönlülür.

Burada $\hat{\beta}$ tahmin edilen regresyon model parametreleri, k iterasyon sayısı ve $\varepsilon > 0$ olmak üzere çok küçük bir sayıdır.

5.2. x 'in Kesin Sayı ve $Y = (y, \underline{\eta}, \bar{\eta})$ 'nin Simetrik Üçgensel Bulanık Sayı Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi

Kesim 5.1'de önerilen bulanık robust regresyon çözümlemesinde girdi değişkenleri yani hem bağımlı hem de bağımsız değişkenler simetrik olmayan üçgensel bulanık sayı olarak ele alınmıştır. Bu kesimde ise bağımsız değişkenlerin kesin sayı, bağımlı değişkenin ise simetrik üçgensel bulanık sayı olması durumunda bulanık çoklu doğrusal regresyon çözümlemesi önerilecektir.

$(x_{i1}, x_{i2}, \dots, x_{ip}) (i=1, 2, \dots, n)$ bağımsız değişkenlerinin kesin sayı, Y bağımlı değişkeninin merkezi y , sol ve sağ yayılımları sırasıyla $\underline{\eta}$ ve $\bar{\eta}$ olan $[Y = (y, \underline{\eta}, \bar{\eta})]$ simetrik üçgensel bulanık sayı olduğu bulanık çoklu doğrusal regresyon modeli

$$Y = A + B_1x_{i1} + \dots + B_px_{ip} + e_i \quad (i=1,2,\dots,n)$$

olarak tanımlandı.

Bu modeldeki A, B_i ($i=1,2,\dots,p$) parametrelerinin tahmini için, tanım 2.7’de verilen genişleme ilkesinden yararlanarak oluşturulan kısıtsız çok değişkenli optimizasyon problemi

$$P2: \quad \text{Min } r(A, B_1, B_2, \dots, B_p) = \sum d(A + B_1x_{i1} + \dots + B_px_{ip}, Y_i)^2$$

olarak tanımlandı. Burada

$$A = (a, \underline{\alpha}, \bar{\alpha}), B_1 = (b_1, \underline{\beta}_1, \bar{\beta}_1), \dots, B_p = (b_p, \underline{\beta}_p, \bar{\beta}_p)$$

biçiminde tanımlanan simetrik üçgensel bulanık parametreler olmak üzere

$$\begin{aligned} d(A + B_1x_{i1} + \dots + B_px_{ip}, Y_i) &= (a + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} - y_i)^2 \\ &+ (a + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} - \underline{\alpha} - \underline{\beta}_1x_{i1} - \underline{\beta}_2x_{i2} - \dots - \underline{\beta}_px_{ip} - y_i + \underline{\eta}_i)^2 \\ &+ (a + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} - \bar{\alpha} - \bar{\beta}_1x_{i1} - \bar{\beta}_2x_{i2} - \dots - \bar{\beta}_px_{ip} - y_i - \bar{\eta}_i)^2 \end{aligned}$$

dir. P2 probleminin a, b_i ’ya göre kısmi türevi sıfıra eşitlenerek

$$\frac{\partial d}{\partial a} = na + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} + b_3 \sum_{i=1}^n x_{i3} + \dots + b_k \sum_{i=1}^n x_{ik} - \sum_{i=1}^n y_i = 0$$

$$\frac{\partial d}{\partial b_1} = a \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1}x_{i2} + b_3 \sum_{i=1}^n x_{i1}x_{i3} + \dots + b_k \sum_{i=1}^n x_{i1}x_{ik} - \sum_{i=1}^n y_i x_{i1} = 0$$

$$\frac{\partial d}{\partial b_2} = a \sum_{i=1}^n x_{i2} + b_1 \sum_{i=1}^n x_{i2} x_{i1} + b_2 \sum_{i=1}^n x_{i2}^2 + b_3 \sum_{i=1}^n x_{i2} x_{i3} + \dots + b_k \sum_{i=1}^n x_{i2} x_{ik} - \sum_{i=1}^n y_i x_{i2} = 0$$

·
·
·

$$\frac{\partial d}{\partial b_k} = a \sum_{i=1}^n x_{ik} + b_1 \sum_{i=1}^n x_{ik} x_{i1} + b_2 \sum_{i=1}^n x_{ik} x_{i2} + b_3 \sum_{i=1}^n x_{ik} x_{i3} + \dots + b_k \sum_{i=1}^n x_{ik}^2 - \sum_{i=1}^n y_i x_{ik} = 0$$

denklem sistemi ve α , $\bar{\alpha}$, β , $\bar{\beta}$ 'ya göre türev alıp sıfıra eşitlenerek

$$\frac{\partial d}{\partial \alpha} = n\alpha + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} + \beta_3 \sum_{i=1}^n x_{i3} + \dots + \beta_k \sum_{i=1}^n x_{ik} - \sum_{i=1}^n \eta_i = 0$$

$$\frac{\partial d}{\partial \beta_1} = \alpha \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1} x_{i2} + \beta_3 \sum_{i=1}^n x_{i1} x_{i3} + \dots + \beta_k \sum_{i=1}^n x_{i1} x_{ik} - \sum_{i=1}^n \eta_i x_{i1} = 0$$

$$\frac{\partial d}{\partial \beta_2} = \alpha \sum_{i=1}^n x_{i2} + \beta_1 \sum_{i=1}^n x_{i2} x_{i1} + \beta_2 \sum_{i=1}^n x_{i2}^2 + \beta_3 \sum_{i=1}^n x_{i2} x_{i3} + \dots + \beta_k \sum_{i=1}^n x_{i2} x_{ik} - \sum_{i=1}^n \eta_i x_{i2} = 0$$

·
·
·

$$\frac{\partial d}{\partial \beta_k} = \alpha \sum_{i=1}^n x_{ik} + \beta_1 \sum_{i=1}^n x_{ik} x_{i1} + \beta_2 \sum_{i=1}^n x_{ik} x_{i2} + \beta_3 \sum_{i=1}^n x_{ik} x_{i3} + \dots + \beta_k \sum_{i=1}^n x_{ik}^2 - \sum_{i=1}^n \eta_i x_{ik} = 0$$

denklem sistemi elde edilir. Bu denklem sisteminin çözümü ile parametre tahminleri elde edilir. x 'in kesin sayı ve $\bar{Y} = (\bar{y}, \bar{\eta}, \bar{\eta})$ 'nin simetrik üçgensel bulanık sayı olması durumunda elde edilen parametre tahminleri simetrik üçgensel bulanık sayıdır. Bu durumda parametre tahmin edicisi matris gösterimiyle

$$\hat{\beta} = \left\{ \hat{\beta}; \text{Merkez} = (X'X)^{-1} X'Y; \text{Yan} = (X'X)X'\eta \right\} \quad (5.4)$$

olarak bulundu. Burada

$$X = \begin{bmatrix} 1 & x_{11} & . & . & x_{1p} \\ 1 & x_{21} & . & . & x_{2p} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_{n1} & . & . & x_{np} \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ . \\ . \\ \eta_n \end{bmatrix}$$

dir.

Eşitlik (5.4), x 'in kesin sayı ve $Y = (y, \underline{\eta}, \bar{\eta})$ simetrik üçgensel bulanık sayı olması ve veri kümesinde aykırı değerin olması durumunda bulanık çok değişkenli regresyon modelinin tahmin edilmesinde kullanılacaktır.

Modelin tahmin sürecinde, gözlemlenmiş Y_i ve tahmin edilen $\hat{Y}_i$ arasındaki fark yani artıklarının rolü oldukça önemlidir. Çözüm algoritmasında üyeliklerin oluşturulmasında uzaklık, uzaklığın tanımında da artıklar etkindir. Bu nedenle artıkların karşılaştırılması iki bulanık sayının karşılaştırılması problemine dayanır. Alt kesimlerde iki bulanık sayının karşılaştırılması regresyon çözümlemesinde tanımlanacak ve algoritmaya dayalı çözüm süreci verilecektir.

5.2.1. Gözlemlenmiş $Y_i = (y_i, \underline{\eta}_i, \bar{\eta}_i)$ ve tahmin edilen $\hat{Y}_i = (\hat{y}_i, \hat{\underline{\eta}}_i, \hat{\bar{\eta}}_i)$ 'nin simetrik üçgensel bulanık sayı olması durumunda karşılaştırma

x 'in kesin sayı ve $Y = (y, \underline{\eta}, \bar{\eta})$ simetrik üçgensel bulanık sayı olması durumunda elde edilen parametre tahminleri simetrik üçgensel bulanık sayıdır. Regresyon çözümlemesinde artıkların elde edilmesi ve karşılaştırılmasında iki bulanık sayı

Y_I 'nin üyelik fonksiyonu ve üyelik fonksiyonunun ters görüntüsü sırasıyla,

71

$$\mu^{-1}(Y) = \begin{cases} w\eta_i + (y_i - \eta_i) & y_i - \eta_i \leq x \leq y_i \\ (y_i + \eta_i) - w\eta_i & y_i \leq x \leq y_i + \eta_i \\ 0 & d.d \end{cases} \quad (5.6)$$

biçiminde tanımlandı.

$\hat{Y}_i$ 'nin üyelik fonksiyonu ve üyelik fonksiyonunun ters görüntüsü sırasıyla,

$$\mu(\hat{Y}) = \begin{cases} \frac{x_i - (\hat{y}_i - \hat{\eta}_i)}{\hat{\eta}_i} & \hat{y}_i - \hat{\eta}_i \leq x \leq \hat{y}_i \\ \frac{(\hat{y}_i + \hat{\eta}_i) - x_i}{\hat{\eta}_i} & \hat{y}_i \leq x \leq \hat{y}_i + \hat{\eta}_i \\ 0 & d.d \end{cases} \quad (5.7)$$

$$\mu^{-1}(\hat{Y}) = \begin{cases} w\hat{\eta}_i + (\hat{y}_i - \hat{\eta}_i) & \hat{y}_i - \hat{\eta}_i \leq x \leq \hat{y}_i \\ (\hat{y}_i + \hat{\eta}_i) - w\hat{\eta}_i & \hat{y}_i \leq x \leq \hat{y}_i + \hat{\eta}_i \\ 0 & d.d \end{cases} \quad (5.8)$$

olarak tanımlandı.

$\chi_1(w) + \chi_2(w) = 1$ ve $\chi_1(w), \chi_2(w) \in (0, 1]$ koşulu için $\chi_1(w) = \chi_2(w) = \chi$ alınırsa, Y_i için $OM(Y_i)$

$$\begin{aligned}
OM(Y_i) &= \int_0^{w_{hgr}^*} g_i \left(\left\{ \mu_T^{-1}(w) \right\} \right) dw \\
&= \frac{\int_0^{w_{hgr}^*} w \left[\chi(w\eta_i + (y_i - \eta_i)) + \chi((y_i + \eta_i) - w\eta_i) \right] dw}{\frac{1}{2} (w_{hgr}^*)^2} \\
&= \frac{\int_0^1 w \left[\chi w\eta_i + \chi y_i - \chi\eta_i + \chi y_i + \chi\eta_i - \chi w\eta_i \right]}{\frac{1}{2} (1)^2} dw \\
&= \frac{\int_0^1 w \left[2\chi y_i \right]}{\frac{1}{2}} dw \\
&= 2\chi y_i
\end{aligned}$$

ve $\hat{Y}_i$ için $OM(\hat{Y}_i)$

$$\begin{aligned}
OM(\hat{Y}_i) &= \int_0^{w_{hgr}^*} g_i \left(\left\{ \mu_{\hat{Y}}^{-1}(w) \right\} \right) dw \\
&= \frac{\int_0^{w_{hgr}^*} w \left[\chi(w\hat{\eta}_i + (\hat{y}_i - \hat{\eta}_i)) + \chi((\hat{y}_i + \hat{\eta}_i) - w\hat{\eta}_i) \right] dw}{\frac{1}{2} (w_{hgr}^*)^2} \\
&= \frac{\int_0^1 w \left[\chi w\hat{\eta}_i + \chi\hat{y}_i - \chi\hat{\eta}_i + \chi\hat{y}_i + \chi\hat{\eta}_i - \chi w\hat{\eta}_i \right]}{\frac{1}{2} (1)^2} dw \\
&= \frac{\int_0^1 w \left[2\chi\hat{y}_i \right]}{\frac{1}{2}} dw \\
&= 2\chi\hat{y}_i
\end{aligned}$$

olarak elde edildi. Buradan Y_i ve $\hat{Y}_i$ 'nin simetrik üçgensel bulanık sayı olması durumunda OM ölçütüne göre $d(Y, \hat{Y})$

$$\begin{aligned}
 d(Y_i, \hat{Y}_i) &= \int_0^{w_{\text{reg}}} g_Y(\{\mu_Y^{-1}(w)\})dw - \int_0^{w_{\text{reg}}} g_{\hat{Y}}(\{\mu_{\hat{Y}}^{-1}(w)\})dw \\
 d(Y_i, \hat{Y}_i) &= OM(Y_i) - OM(\hat{Y}_i) \\
 &= 2\chi y_i - 2\chi \hat{y}_i \\
 &= 2\chi (y_i - \hat{y}_i)
 \end{aligned} \tag{5.9}$$

biçiminde tanımlandı.

5.2.2. x 'in Kesin Sayı ve $Y = (y, \underline{\eta}, \bar{\eta})$ 'nin Simetrik Üçgensel Bulanık Sayı Olması Durumunda Önerilen Bulanık Robust Regresyon Çözümlemesi için Algoritma

Bu algoritmada x 'in kesin sayı ve $Y = (y, \underline{\eta}, \bar{\eta})$ simetrik üçgensel bulanık sayı olması durumunda başlangıç parametreleri eşitlik (5.4)'den elde edildi. Her bir gözlemin ağırlıklı olarak modele katkıda bulunabilmesi için üçgensel üyelik fonksiyonu tanımlandı. Bu fonksiyondan üyelik dereceleri belirlenerek ağırlık matrisi (W) oluşturuldu. Ağırlık matrisi, köşegen elemanları üyeliklerden oluşan diagonal bir matristir. Ağırlık matrisi yardımıyla ağırlıklandırılmış bulanık en küçük kareler tahmin edicisi bulundu ve parametre tahminleri elde edildi. Ard arda bulunan $\hat{\beta}$ tahminleri arasındaki fark sıfıra yaklaşıncaya kadar süreç devam edildi.

Çoklu doğrusal regresyon çözümlemesinde x kesin sayı, $Y = (\underline{y}, \underline{\eta}, \bar{\eta})$ simetrik üçgensel bulanık sayı, çıktı parametrelerinin simetrik üçgensel bulanık sayı olması ve veri kümesinde aykırı değerın olması durumunda regresyon modelinin tahmin edilmesinde önerilen bulanık robust yöntem için algoritmanın adımları aşağıda verilmiştir.

Adım 1: x 'in kesin sayı ve $Y = (\underline{y}, \underline{\eta}, \bar{\eta})$ simetrik üçgensel bulanık sayı olması durumunda başlangıç bulanık tahmini regresyon modeli Eşitlik (5.4)'den bulunur.

Adım 2: Merkezler için tahmin değerleri ($\hat{y}_i$) ve artıklar (e_i) hesaplanır. Artıklar bulanık sayıların karşılaştırılmasına göre belirlenir. Kesim 5.2.1'de tanımlandığı gibi iki bulanık sayı arasındaki fark merkezler arasındaki farka eşittir ve Eşitlik (5.9)'da verilmiştir.

Adım 3: Artıkların mutlak değerlerine göre medyanyı belirlenir ve

$$d_i = \|abs(e_i) - medyan(abs(e_i))\| \quad i = 1, 2, \dots, n$$

uzaklığı hesaplanır. Burada $\|\cdot\|$ öklid uzaklığıdır.

Adım 4: Uzaklıklara bağlı olarak üyelik fonksiyonu

$$\mu(e) = \begin{cases} 1 & , & abs(e) \leq a \\ \frac{b - abs(e)}{b - a} & , & a < abs(e) < b \\ 0 & , & d.d \end{cases} \quad (5.10)$$

biçiminde tanımlandı. Burada;

$$a = \text{medyan}(d_i)$$

$$b = \max(d_i) + d$$

$$d = \text{medyan} \left| e, -\text{medyan}(e) \right| / (0.6745)$$

dır.

Adım 5: Eşitlik (5.10)'da tanımlanan üyelik fonksiyonu ile üyelik dereceleri belirlenir ve ağırlık matrisi oluşturulur. Ağırlık matrisi, köşegen elemanları üyeliklerden oluşan diagonal matristir. Elde edilen ağırlık matrisi yardımıyla ağırlıklandırılmış bulanık en küçük kareler tahminleri bulunur.

Adım 6: Eğer $\left| \hat{\beta}^{k+1} - \hat{\beta}^k \right| < \varepsilon$ ise durulur. Aksi durumda Adım 2' ye dönlür.

Burada $\hat{\beta}$ tahmin edilen regresyon model parametreleri, k iterasyon sayısı ve $\varepsilon > 0$ olmak üzere çok küçük bir sayıdır.

6. UYGULAMA

6.1. Giriş

Bu bölümde bağımlı değişkende aykırı değer olması ve bağımsız değişkenlerde aykırı değer olması durumları için Dördüncü Bölüm'de verilen klasik çoklu robust doğrusal regresyon çözümlemesi ve Kesim 5.1 ve Kesim 5.2'de önerilen bulanık çoklu robust doğrusal regresyon çözümlemesinin karşılaştırılması amaçlandı. Bu nedenle belirtilen durumlara uygun yapay veri türetildi. Bu türetilen verilerden 11 farklı yapıdaki örnek ele alındı. Bu örnek verilerin türetilmesi ve çözümü için Matlab paket programında bir program yazıldı. Bu program EKK, Huber, Hampel, Tukey, Andrews M yöntemleri ve önerilen iki yöntemi içermektedir. Türetilen yapay veriler için elde edilen sonuçlar karşılaştırıldı.

6.2. Bağımlı Değişkende Aykırı Değer Olması Durumu için Uygulama

Bu kesimde yöntemleri karşılaştırabilmek ve sonuçları irdeleyebilmek için bir bağımlı ve üç bağımsız değişkeni içeren 11 tane regresyon problemi ele alındı. Her bir regresyon problemindeki gözlem sayısı 30 olarak belirlendi. Bu regresyon problemleri için yapay veriler, bağımsız değişkenler $X_1 \sim N(\mu=20; \sigma=3)$, $X_2 \sim N(\mu=50; \sigma=12)$ ve $X_3 \sim N(\mu=32; \sigma=13)$ olan normal dağılımdan, bağımlı değişken değerleri ise bağımsız değişken değerlerine bağlı olarak türetildi. Bağımlı değişkendeki 15'inci gözlemler $(x_{15} + 20)$ biçiminde değiştirilerek aykırı gözlem durumuna dönüştürüldü.

Kesim 5.1'de önerilen hem bağımlı hem de bağımsız değişkenlerin simetrik olmayan üçgensel bulanık sayı olması durumunda bulanık çoklu robust doğrusal regresyon için çözümleme aşamasında türetilen yapay veriler, bağımsız değişken

değerleri merkez (x_i), sol yayılma $\underline{x}_i = x_i / 9$, sağ yayılma $\overline{x}_i = x_i / 7$ ve bağımlı değişken değerleri; merkez (y_i), sol yayılma $\underline{y}_i = y_i / 8$ ve sağ yayılma $\overline{y}_i = y_i / 9$ alınarak bulanıklaştırıldı.

Kesim 5.2’de önerilen x ’in kesin sayısı, bağımlı değişkenin simetrik üçgensel bulanık sayısı olması durumunda bulanık çoklu robust doğrusal regresyon için çözümleme aşamasında türetilen yapay veriler bağımlı değişken değerleri merkez (y_i), yayılma ise $\underline{y}_i = y_i / 8$ alınarak bulanıklaştırıldı. Sol ve sağ yayılmalar literatürde belirtildiği gibi belirlendi.

Burada birinci örnek için, veri kümesi, model tahmini, çözümleme sonuçları ve tüm gözlemlere karşılık gelen artıklar, ağırlıklar ve artıkların dağılım grafikleri ayrıntılı olarak verildi. Diğer 10 örnek için model tahminleri ve sadece aykırı değere karşılık gelen artıklar ve ağırlıklar verildi. Birinci örneğin içerdiği veri kümesi Çizelge 6.1’de verildi. Bu örnek için elde edilen artıklar ve bunlara karşılık gelen ağırlık matrisinin köşegen elemanları Çizelge 6.2’de, model tahminleri Çizelge 6.3’de ve artıkların dağılım grafikleri Grafik 6.1 ve Grafik 6.2’de verildi.

Çizelge 6.1. Bağımlı değişkende aykırı değer olması durumu için veri kümesi

Gözlem No	x_1	x_2	x_3	y	Gözlem No	x_1	x_2	x_3	y
1	24.0228	66.5694	30.4791	119.1278	16	24.9026	33.8550	4.9907	64.1516
2	21.1642	65.8381	34.1903	120.1272	17	22.4756	67.7825	30.1339	122.3146
3	21.1792	39.0868	25.4843	90.7688	18	20.6923	50.3928	50.1134	120.4816
4	14.8780	22.3327	22.8340	58.3170	19	22.0149	72.4454	40.4735	133.0008
5	20.6836	71.4648	38.6061	130.5159	20	18.4758	35.4921	27.0973	80.7247
6	22.0569	54.6896	26.5280	102.7560	21	22.5691	40.6084	23.4012	90.8250
7	18.0896	50.2439	34.9787	109.1715	22	20.8055	40.7924	35.2365	104.9523
8	16.9922	45.1283	19.5265	82.8897	23	21.8749	48.7136	27.0143	99.2907
9	19.4431	31.5813	30.1014	78.8827	24	16.8580	38.2753	25.1298	80.4632
10	16.8379	52.6565	41.6790	108.7998	25	24.6070	38.4321	32.7200	98.8642
11	19.7854	33.5063	20.4236	79.9222	26	21.3033	21.4499	48.2990	90.5894
12	20.8376	39.9286	33.8078	101.3863	27	14.2486	39.9417	-0.7600	60.0177
13	24.1198	47.4963	28.9301	106.4514	28	21.4098	53.0881	39.6031	118.4216
14	20.5395	59.0710	31.0190	107.8617	29	23.8231	47.7940	18.8952	87.8295
15	18.3740	54.5088	27.3386	120.5807	30	21.9156	47.9886	44.2757	113.1703

Çizelge 6.2. Bağımlı değişkende aykırı değer olması durumu için artıklar ve bunlara karşılık gelen ağırlık matrisinin köşegen elemanları

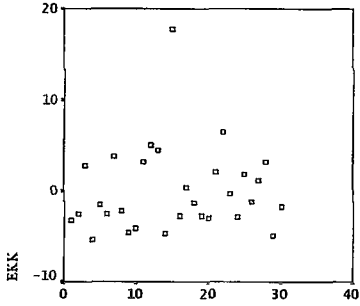
Gözlem No	EKK Artık	Huber Artık Ağırlık	Hampel Artık Ağırlık	Tukey Artık Ağırlık	Andrews Artık Ağırlık	Önerilen1 Artık Ağırlık	Önerilen2 Artık Ağırlık
1	-3.2546	-2.7534	-2.6839	-2.5357	0.9574	-0.4967	-2.5705
2	-2.5960	-1.6716	-1.5414	-1.2266	0.9900	0.6256	-1.2864
3	2.6995	2.9200	2.9120	2.9846	0.9413	0.8742	2.9607
4	-5.4076	-4.5954	-4.5269	-4.2426	0.8832	-4.5370	-4.3141
5	-1.4963	-0.3681	-0.2049	-0.1959	0.9997	0.4761	0.1284
6	-2.5206	-2.0002	-1.9377	-1.7865	0.9788	-1.9463	-1.8293
7	3.7724	4.7484	4.8620	5.2275	0.8255	0.4606	5.1641
8	-2.1846	-1.0281	-0.8624	-0.5255	0.9982	0.4750	-0.6250
9	-4.5651	-4.3169	-4.3407	-4.2157	0.8846	-4.3477	-4.2354
10	-4.1713	-2.9778	-2.8383	-2.3618	0.9630	-2.8161	-2.4308
11	3.2505	3.5893	3.6021	3.6923	0.9109	3.5745	3.6520
12	5.0459	5.2696	5.2471	5.3722	0.8162	5.2506	5.3604
13	4.3856	4.3263	4.2786	4.2642	0.8820	4.2743	4.2660
14	-4.6750	-3.8140	-3.6981	-3.4124	0.9236	-3.6964	-3.4728
15	17.7901	18.9157	19.0751	19.4342	0	19.0669	19.3491
16	-2.8054	-3.1693	-3.2402	-3.4840	0.9204	-3.2997	-3.4966
17	0.3309	1.1306	1.2513	1.4950	0.9851	1.2537	1.4383
18	-1.2603	-0.8599	-0.8698	-0.6008	0.9976	-0.8282	-0.6006
19	-2.8231	-1.9073	-1.7809	-1.4411	0.9862	-1.7550	-1.4905
20	-3.0220	-2.4684	-2.4306	-2.2297	0.9670	-2.4436	-2.2755
21	2.0207	2.0722	2.0430	2.0430	0.9723	2.0245	2.0283
22	6.4551	6.6959	6.6747	6.8131	0.7135	6.6816	6.8022
23	-0.3450	0.0263	0.0545	0.1697	0.9998	0.0455	0.1379
24	-2.8827	-1.9575	-1.8510	-1.5433	0.9841	-1.8682	-1.6184
25	1.8904	1.4779	1.3165	1.2145	0.9902	1.3186	1.2488
26	-1.2049	-1.7379	-1.9403	-1.9424	0.9749	-1.9090	-1.8799
27	1.1882	2.8451	3.1245	3.5092	0.9193	3.0501	3.3435
28	3.1503	3.6087	3.6330	3.8495	0.9033	3.6524	3.8123
29	-4.9809	-4.8867	-4.8892	-4.9167	0.8448	-4.9155	-4.9407
30	-1.7842	-1.6043	-1.6424	-1.4800	0.9854	-1.6137	-1.4727

Çizelge 6.2 incelendiğinde, aykırı değer olarak alınan 15. gözlemin ağırlığı Huber’de “0.2714”, Hampel’da “0.1684”, Tukey’de “0” ve Andrews’de “0.0844” önerilen yöntemlerde ise “0.0603” ve “0.0511” olarak bulunmuştur. Önerilen yöntemlerden bulunan ağırlıklar aslında her bir gözlemin üyelik dereceleri. Bu üyelik dereceleri gözlemlerin modele etkisini göstermektedir. Dolayısıyla Çizelge 6.2’den de görüldüğü gibi aykırı değerler çok küçük üyelik derecesi ile diğer gözlemler ise “1 ya da 1” e yakın üyelik dereceleriyle regresyon modelini etkiler.

Çizelge 6.3. Bağımlı değişkende aykırı değer olması durumu için regresyon model tahminleri

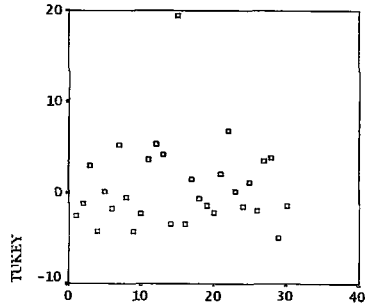
Yöntem	Sabit	Regresyon Katsayıları			Artıkların Kareleri Toplamı
		β_1	β_2	β_3	
EKK	7.9975	0.8303 (0.3505)	0.9941 (0.0740)	0.9273 (0.0877)	638.5054
Huber	5.1267	0.9981 (0.2683)	0.9648 (0.0566)	0.9367 (0.0672)	654.4245
Hampel	4.7110	1.0255 (0.2579)	0.9585 (0.0544)	0.9402 (0.0645)	659.0939
Tukey	3.8425	1.0817 (0.2641)	0.9505 (0.0557)	0.9370 (0.0661)	672.5679
Andrews	4.7978	1.0252 (0.2699)	0.9583 (0.0570)	0.9380 (0.0675)	658.8663
Önerilen 1	4.9524	0.9989	0.9402	0.9216	811.5322
Önerilen 2	4.0970	1.0696	0.9526	0.9349	669.2256
Yan	0.5121	0.1337	0.1191	0.1169	

Çizelge 6.3’de EKK, Huber, Hampel, Tukey, Andrews yöntemleri için regresyon model tahminleri ve parantez içinde parametrelerin standart sapmaları, önerilen yöntemlerden elde edilen regresyon model tahminleri ve artıkların kareleri toplamı verilmiştir. Çizelge incelendiğinde önerilen yöntemlerden elde edilen parametre tahmin değerleri klasik Robust yöntemlerle işaretçe aynı ve büyüklükçe çok yakındır.



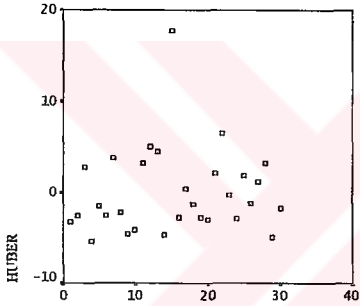
İNDEKS

(a) EKK Yöntemi için Artıkların Dağılımı



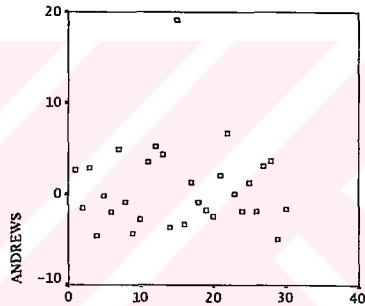
İNDEKS

(d) Tukey Yöntemi için Artıkların Dağılımı



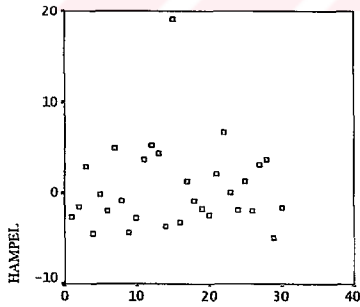
İNDEKS

(b) Huber Yöntemi için Artıkların Dağılımı



İNDEKS

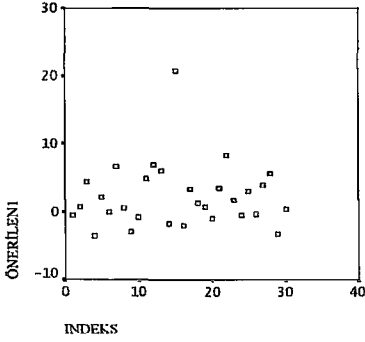
(e) Andrews Yöntemi için Artıkların Dağılımı



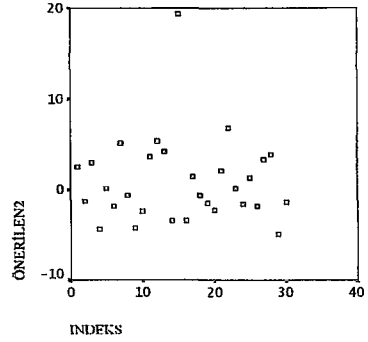
İNDEKS

(c) Hampel Yöntemi için Artıkların Dağılımı

Grafik 6.1. Bağımlı değişkende aykırı değer olması durumuna ilişkin EKK, Huber, Hampel, Tukey ve Andrews yöntemleri için artıkların dağılımı



(a) Önerilen1 Yöntemi için Artıkların Dağılımı



(b) Önerilen2 Yöntemi için Artıkların Dağılımı

Grafik 6.2. Bağımlı değişkende aykırı değer olması durumuna ilişkin önerilen iki yöntem için artıkların dağılımı

Bu kesimde ele alınan diğer 10 örneğe ilişkin model tahminleri ve Huber yöntemine göre 1'den farklı ağırlık alan gözlemlere karşılık gelen artıklar ve ağırlıklar Huber, Hampel, Tukey, Andrews ve önerilen yöntemler için Çizelge 6.4'de verildi.

Çizelge 6.4. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımlı değişkende aykırı değer olması durumu için çözümleme sonuçları

	EKK		Huber		Hampel		Tukey		Andrews		Önerilen 1		Önerilen 2	
	Artık	Ağırlık	Artık	Ağırlık	Artık	Ağırlık	Artık	Ağırlık	Artık	Ağırlık	Artık	Ağırlık	Artık	Ağırlık
Örnek 1	7.0848	7.1949	0.5092	7.0304	0.5505	7.1169	0.2460	6.9104	0.3047	8.2333	0.6065	6.9524	0.5824	
	-4.5998	-3.7751	0.9705	-3.5181	1	-3.2436	0.8016	-3.3715	0.4314	-1.2553	1	-3.4916	0.8250	
	9.7427	12.5326	0.2923	14.0030	0.1273	15.1921	0	15.0869	0	18.4573	0.0224	14.6632	0.0417	
	5.9024	4.7041	0.7788	3.7515	1	3.2148	0.8049	3.0610	0.4391	3.4193	0.8815	3.3177	0.8372	
	4.7660	4.7056	0.7786	4.4228	0.8750	4.3932	0.6528	4.1800	0.4084	5.5405	0.7603	4.2287	0.7734	
	-4.3819	-3.7170	0.9856	-3.5489	1	-3.3238	0.7922	-3.4480	0.4294	-1.5395	0.9889	-3.5294	0.8224	
Model Tahmini														
Sabit	-15.2659	-8.0014		-3.3928			0.1291		-0.0390		3.9353		-1.2660	-0.1583
$\hat{\beta}_1$	1.3278	1.1103		0.9813			0.8566		0.8891		0.7416		0.9212	0.1152
$\hat{\beta}_2$	1.1682	1.1179		1.0857			1.0673		1.0593		1.0288		1.0674	0.1334
$\hat{\beta}_3$	1.0853	1.0579		1.0425			1.0343		1.0341		0.9997		1.0406	0.1301
Örnek 2	-6.1430	-5.8264	0.8406	-5.7927	0.9630	-5.6779	0.7579	-5.9031	0.4181	-4.0957	0.8289	-5.7736	0.7047	
	14.1319	15.2957	0.3202	15.3086	0.2740	15.8828	0	15.0518	0.1701	17.3862	0.0280	15.6810	0.0721	
	-5.5141	-4.9432	0.9908	-4.9176	1	-4.6815	0.8318	-5.1098	0.4322	-3.2411	0.8804	-4.8395	0.7643	
	6.0951	7.5483	0.6489	7.5437	0.7395	8.2605	0.5272	7.2015	0.3913	9.5563	0.4998	7.9950	0.5629	
Model Tahmini														
Sabit	14.7149	10.6993		10.6654			8.6676		11.6646		7.3686		9.4384	1.1798
$\hat{\beta}_1$	0.6839	0.7369		0.7414			0.7672		0.7273		0.7957		0.7559	0.0945
$\hat{\beta}_2$	0.8948	0.9448		0.9432			0.9695		0.9326		0.9567		0.9617	0.1202
$\hat{\beta}_3$	0.9294	0.9353		0.9350			0.9371		0.9314		0.9287		0.9342	0.1168

Çizelge 6.4. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımlı değişkende aykırı değer olması durumu için çözümleme sonuçları (Devam)

	EKK Artık	Huber Artık Ağırlık	Hampel Artık Ağırlık	Tukey Artık Ağırlık	Andrews Artık Ağırlık	Önerilen 1 Artık Ağırlık	Önerilen 2 Artık Ağırlık	Yan
Örnek 3	21.4278	23.4688	23.8180	0	23.7355	0	23.6222	0.0369
	6.0202	7.4517	7.6485	0.5821	7.5354	0.3421	9.3268	0.7249
	3.8970	4.9709	5.1531	0.8640	5.1067	0.4114	5.0449	0.8272
Model Tahmini	Sabit							
	$\hat{\beta}_1$	5.2361	5.9742	6.5763	5.7136	6.1516	5.4417	0.6802
	$\hat{\beta}_2$	0.8861	0.8349	0.8140	0.8425	0.8209	0.8577	0.1072
	$\hat{\beta}_3$	0.9827	0.9733	0.9687	0.9774	0.9578	0.9782	0.1223
Örnek 4		0.9723	0.9878	0.9920	0.9913	0.9729	0.9909	0.1239
		4.4442	5.6195	5.8444	6.1371	7.4930	5.9901	0.7906
		-10.1912	-10.4888	-10.2597	-10.2066	-8.3106	-10.0601	0.5918
		-6.8749	-6.2163	-5.7686	-5.4901	-4.0285	-5.4474	0.8172
Model Tahmini	Sabit	13.6425	7.1695	5.6535	4.2624	5.1665	4.9109	0.6139
	$\hat{\beta}_1$	0.9150	0.9946	1.0122	1.0229	0.9949	1.0159	0.1270
	$\hat{\beta}_2$	0.8618	0.9420	0.9540	0.9704	0.9369	0.9593	0.1199
	$\hat{\beta}_3$	0.8843	0.9042	0.9173	0.9252	0.9124	0.9263	0.1158

Çizelge 6.4. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımlı değişkende aykırı değer olması durumu için çözümleme sonuçları
(Devam)

	EKK	Huber	Hampel	Tukey	Andrews	Önerilen 1		Önerilen 2	
	Artık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Yan
Örnek 5	8.3662	7.9948	0.6942	7.8440	0.8037	7.6935	0.6813	7.9048	0.3978
	15.8537	17.4450	0.3181	17.6870	0.2607	18.3863	0	17.3568	0.1669
	-6.2586	-5.8565	0.9476	-5.8202	1	-5.6776	0.8188	-5.9284	0.4311
	-6.4373	-5.8654	0.9462	-5.7588	1	-5.5617	0.8258	-5.9215	0.4312
Model Tahmini	6.8216	5.9970		5.6250		5.9723		5.8934	
	$\hat{\beta}_1$	0.6271	0.7269	0.7542		0.7298		0.7653	
	$\hat{\beta}_2$	1.0174	0.9841	0.9802		0.9853		0.9690	
	$\hat{\beta}_3$	1.0537	1.0530	1.0522		1.0525		1.0510	
Örnek 6	-8.4793	-7.4024	0.7845	-7.3628	0.8871	-6.8989	0.7816	-4.6081	0.8076
	12.4088	13.9188	0.4172	13.9411	0.4475	14.6094	0.2305	17.1483	0.0250
	-7.3866	-7.8693	0.7380	-7.7030	0.8479	-7.9690	0.7146	-6.8331	0.6688
								-8.3590	0.5438
Model Tahmini	9.3923	8.8776		8.6541		8.9423		8.2638	
	$\hat{\beta}_1$	0.7459	0.7417	0.7393		0.7480		0.7504	
	$\hat{\beta}_2$	1.0106	0.9937	0.9981		1.0010		0.9650	
	$\hat{\beta}_3$	0.9077	0.9507	0.9507		0.9341		0.9878	
								1.0330	
								0.7569	
								0.9772	
								0.9857	

Çizelge 6.4. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımlı değişkende aykırı değer olması durumunda için çözümleme sonuçları
(Devam)

	EKK	Huber	Hampel	Tukey	Andrews	Önerilen 1	Önerilen 2
	Artık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık Yan
Örnek 7	15.0774	15.7712	16.0563	16.1591	16.2273	17.7670	16.1438
	-4.3825	-5.1026	-4.6678	-4.5690	-4.3825	-3.1545	-4.5287
	-5.5857	-5.9677	-5.7556	-5.7717	-5.5857	-4.3166	-5.6660
	5.4582	5.3406	5.4073	5.4045	5.4582	7.0991	5.4420
							0.0475
Model Tahmini	9.2351	8.3760	6.9728	6.9022	6.0218	6.3349	6.5429
	0.5804	0.6347	0.6728	0.6779	0.6958	0.6705	0.6822
	1.0669	1.0446	1.0447	1.0412	1.0466	1.0324	1.0457
	0.8791	0.9017	0.9167	0.9202	0.9269	0.9043	0.9214
Örnek 8	5.0415	6.1046	5.8171	6.3931	6.0107	7.6816	6.1268
	6.1354	4.8949	4.5813	4.4266	4.4206	5.8601	4.5369
	18.4456	21.7385	22.5299	22.9529	22.9559	25.1725	22.6360
	-5.0485	-5.3807	-5.5826	-5.5436	-5.6221	-4.3173	-5.5418
	6.7061	5.7877	5.0519	5.3411	4.9813	6.1902	5.2769
Model Tahmini	-1.5717	3.0630	4.1462	4.8425	4.7947	5.2972	4.3430
	0.7691	0.8107	0.8296	0.8251	0.8323	0.7837	0.8260
	1.0936	1.0027	0.9892	0.9703	0.9764	0.9584	0.9817
	1.0624	1.0213	0.9974	1.0036	0.9936	0.9851	1.0033
							0.5429
							0.1033
							0.1227
							0.1254

Çizelge 6.4. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımlı değişkende aykırı değer olması durumu için çözümleme sonuçları (Devam)

	EKK	Huber	Hampel	Tukey	Andrews	Önerilen 1	Önerilen 2
	Artık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık Yan
Örnek 9	-6.8495	-7.0719	-6.9823	-7.0102	-6.9205	-5.2231	-7.0188
	3.5736	5.1892	5.0765	6.1065	4.7294	7.3078	5.8406
	12.9005	14.0648	14.0762	14.8660	13.8252	16.3065	14.6286
	-7.1604	-7.2636	-7.0911	-7.0219	-7.0617	-5.1044	-7.0919
	-5.6181	-4.7325	-4.7398	-4.1125	-4.9271	-2.4966	-4.2883
Model Tahmini							
Sabit	9.9242	7.2375	8.0207	6.3469	8.5713	6.7718	6.7048
$\hat{\beta}_1$	0.8091	0.7743	0.7466	0.7175	0.7562	0.6790	0.7371
$\hat{\beta}_2$	0.9444	1.0018	0.9967	1.0302	0.9838	1.0157	1.0216
$\hat{\beta}_3$	0.9376	0.9498	0.9486	0.9549	0.9460	0.9408	0.9531
Örnek 10	4.2774	3.5215	3.3014	3.3559	3.3341	5.7913	3.4059
	16.1186	17.6870	18.2144	18.3655	18.1155	19.4397	18.0378
	-5.8202	-5.5155	-5.2003	-5.0548	-4.9999	-4.1863	-5.0613
	-3.9681	-4.1445	-4.1636	-4.1191	-4.1354	-2.2123	-4.1126
	3.8446	5.5483	6.1249	6.3179	6.0151	7.3930	5.9329
Model Tahmini	-6.0847	-4.1511	-3.6344	-3.5048	-3.9027	-2.0379	-3.9621
	2.6480	3.7766	4.2206	4.4027	4.2045	5.5612	4.1419
Sabit	13.1329	7.0476	5.0388	4.5039	5.4147	5.6254	5.7590
$\hat{\beta}_1$	0.6271	0.8132	0.8685	0.8881	0.8490	0.8413	0.8408
$\hat{\beta}_2$	0.9086	0.9625	0.9786	0.9798	0.9743	0.9522	0.9711
$\hat{\beta}_3$	0.9784	0.9575	0.9558	0.9563	0.9628	0.9448	0.9625

6.3. Bağımsız Değişkende Aykırı Değer Olması Durumu için Uygulama

Bu kesimde yöntemleri karşılaştırmak ve sonuçları irdeleyebilmek için bir bağımlı ve üç bağımsız değişkeni içeren 11 tane regresyon problemi ele alındı. Her bir regresyon problemindeki gözlem sayısı 30 olarak belirlendi. Bu regresyon problemi için yapay veriler bağımsız değişkenler $X_1 \sim N(\mu = 2; \sigma = 9)$, $X_2 \sim N(\mu = 50; \sigma = 45)$ ve $X_3 \sim N(\mu = 32; \sigma = 20)$ olan normal dağılımdan, bağımlı değişken değerleri ise bağımsız değişken değerlerine bağlı olarak türetilti. X_1 bağımsız değişkenlerinin standart sapması ortalamadan daha büyük alınarak aykırı gözlem durumuna dönüştürüldü.

Kesim 5.1'de önerilen hem bağımlı hem de bağımsız değişkenlerin simetrik olmayan üçgensel bulanık sayı olması durumunda bulanık çoklu robust doğrusal regresyon için çözümleme aşamasında türetilen yapay veriler, bağımsız değişken değerleri merkez (x_i) , sol yayılma $\underline{\xi}_i = x_i / 7$, sağ yayılma $\bar{\xi}_i = x_i / 9$ ve bağımlı değişken değerleri; merkez (y_i) , sol yayılma $\underline{\eta}_i = y_i / 6$ ve sağ yayılma $\bar{\eta}_i = y_i / 8$ alınarak bulanıklaştırıldı.

Kesim 5.2'de önerilen x 'in kesin sayı, bağımlı değişkenin simetrik üçgensel bulanık sayı olması durumunda bulanık çoklu robust doğrusal regresyon için çözümleme aşamasında türetilen yapay veriler bağımlı değişken değerleri merkez (y_i) , yayılma ise $\underline{\eta}_i = y_i / 8$ alınarak bulanıklaştırıldı. Sol ve sağ yayılmalar literatürde belirtildiği gibi belirlendi.

Burada birinci örnek için, veri kümesi, model tahmini, çözümleme sonuçları, tüm gözlemlere karşılık gelen artıklar ve artıkların dağılım grafikleri ayrıntılı olarak verildi. Diğer 10 örnek için model tahminleri ve sadece aykırı değere karşılık gelen artıklar ve ağırlıklar verildi. Birinci örneğin içerdiği veri kümesi Çizelge

6.5’de verildi. Bu örnek için elde edilen artıklar ve bunlara karşılık gelen ağırlık matrisinin köşegen elemanları Çizelge 6.6’da, model tahminleri Çizelge 6.7’de ve artıkların dağılım grafikleri Grafik 6.2 ve Grafik 6.3’de verildi.

Çizelge 6.5. Bağımsız değişkende aykırı değer olması durumu için veri kümesi

Gözlem No	x_1	x_2	x_3	y	Gözlem No	x_1	x_2	x_3	y
1	-4.8777	137.7014	38.8186	168.0472	16	-9.8713	60.3161	34.2680	83.4114
2	-0.2649	60.2906	42.9718	103.2190	17	-4.0244	81.3879	49.6997	131.4449
3	6.7218	56.0766	11.6807	82.8562	18	-4.3061	-2.5934	20.3537	1.6819
4	-4.0585	17.3481	-0.6704	15.2539	19	-11.8636	113.2403	17.6490	118.0969
5	10.5485	77.8001	-4.2070	78.9300	20	4.0563	113.7553	18.5350	143.1857
6	13.7022	64.1522	31.3019	101.7508	21	-3.4802	2.9425	25.6062	20.0301
7	-9.0196	100.5833	45.5153	137.2468	22	10.3094	4.7964	64.1301	79.3975
8	-0.0585	100.8095	14.1827	118.8237	23	18.6609	16.7722	53.2259	104.4898
9	0.1459	30.4750	-5.8267	33.8019	24	-2.7993	7.4978	38.0110	40.9138
10	8.8687	87.1551	76.3502	161.8509	25	-2.8481	43.9067	39.0225	79.3384
11	5.0856	89.7352	13.2372	109.9495	26	17.2688	30.9818	51.0443	96.3733
12	16.3372	56.5862	1.1287	72.7973	27	0.4098	121.9038	19.3427	141.9212
13	12.5498	103.3084	35.1923	155.7732	28	9.2750	32.8525	14.8254	53.7638
14	-8.2336	35.3163	24.6232	40.5446	29	2.7285	60.5666	66.9222	151.4347
15	4.8076	-8.9089	11.7568	11.3881	30	1.9459	36.0797	20.8777	56.8167

Çizelge 6.6. Bağımsız değişkende aykırı değer olması durumu için artıklar ve bunlara karşılık gelen ağırlık matrisinin köşegen elemanları

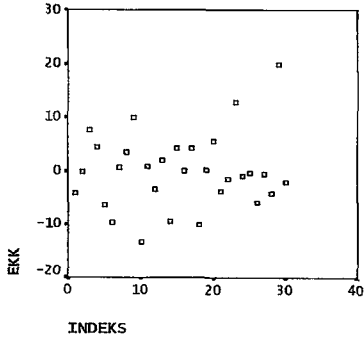
Gözlem No	EKK Artık	Huber Artık	Ağırlık	Hampel Artık	Ağırlık	Tukey Artık	Ağırlık	Andrews Artık	Ağırlık	Önerilen1 Artık	Ağırlık	Önerilen2 Artık	Ağırlık
1	-4.2364	-3.9126	1	-3.7442	1	-3.4127	0.9695	-3.9892	0.4681	-3.6288	0.9445	-3.5251	0.9568
2	-0.1346	0.8563	1	0.9630	1	1.6554	0.9928	0.3951	0.4761	1.6292	1	1.4623	1
3	7.6277	7.6067	1	7.5556	1	7.4693	0.8580	7.5613	0.4476	8.0214	0.7533	7.4780	0.7836
4	4.3946	3.3084	1	3.6700	1	2.8087	0.9793	4.0234	0.4680	3.8082	0.9367	2.8802	0.9851
5	-6.5439	-7.1539	1	-7.2981	1	-7.7867	0.8462	-6.9549	0.4519	-7.0338	0.7963	-7.6600	0.7757
6	-9.7796	-8.6809	0.9725	-8.8554	1	-8.0429	0.8364	-9.3657	0.4327	-7.9389	0.7569	-8.1632	0.7536
7	0.5691	1.1314	1	1.3971	1	1.8258	0.9912	0.9880	0.4757	1.6465	1	1.6467	1
8	3.5236	3.1438	1	3.2120	1	2.9293	0.9775	3.3576	0.4705	3.3009	0.9588	2.9543	0.9818
9	10.0242	9.0563	0.9322	9.1281	1	8.2400	0.8286	9.5602	0.4309	9.2645	0.6992	8.3521	0.7453
10	-13.3966	-10.6712	0.7911	-10.7255	0.9345	-8.6131	0.8136	-12.1027	0.4049	-9.4364	0.6917	-9.0159	0.7163
11	0.8897	0.7389	1	0.7100	1	0.5915	0.9991	0.7769	0.4759	0.9943	1	0.6128	1
12	-3.3107	-3.3572	1	-3.6031	1	-3.6887	0.9644	-3.5071	0.4699	-2.9903	0.9723	-3.5144	0.9529
13	1.8933	2.9633	1	2.8014	1	3.6730	0.9647	2.3058	0.4735	3.5588	0.9476	3.5530	0.9556
14	-9.5228	-9.5791	0.8813	-9.3233	1	-9.5117	0.7752	-9.4244	0.4322	-9.0438	0.7088	-9.5947	0.6909
15	4.2073	4.3432	1	4.3503	1	4.2428	0.9530	4.2450	0.4671	5.0531	0.8825	4.2096	0.9268
16	0.0403	0.2318	1	0.5185	1	0.5722	0.9991	0.2833	0.4761	0.7682	1	0.4421	1
17	4.1799	5.2138	1	5.3917	1	6.1549	0.9024	4.7766	0.4646	5.9400	0.8439	5.9328	0.8513
18	-10.1266	-10.0515	0.8399	-9.8630	1	-10.0226	0.7522	-10.0088	0.4267	-9.3500	0.6955	-10.1041	0.6686
19	0.3206	-0.4816	1	-0.1855	1	-0.7928	0.9983	0.0847	0.4762	-0.4682	1	-0.7765	1
20	5.5486	5.4904	1	5.4777	1	5.4801	0.9222	5.4946	0.4609	5.6949	0.8546	5.4824	0.8711
21	-3.7108	-3.3948	1	-3.4777	1	-3.1796	0.9735	-3.4815	0.4700	-2.6339	0.9878	-3.2929	0.9670
22	-1.5152	1.0544	1	0.9894	1	2.8080	0.9793	-0.3157	0.4761	2.5267	0.9925	2.4320	1
23	12.8914	15.3220	0.5510	15.0828	0.6645	16.8364	0.3912	13.9187	0.3833	16.6714	0.3768	16.5363	0.3868
24	-0.8981	-0.0382	1	0.1310	1	0.6042	0.9990	-0.4025	0.4761	0.8899	1	0.4109	1
25	-0.3759	0.3914	1	0.5506	1	1.0297	0.9972	0.0728	0.4762	1.1569	1	0.8535	1
26	-5.7978	-3.5745	1	-3.7929	1	-2.1725	0.9876	-4.8586	0.4642	-2.3419	1	-2.4480	1
27	-0.5824	-0.7982	1	-0.7420	1	-0.8517	0.9981	-0.6718	0.4760	-0.6628	1	-0.8475	1
28	-4.0772	-3.7648	1	-3.8556	1	-3.7282	0.9636	-4.0057	0.4680	-3.1622	0.9648	-3.7569	0.9467
29	19.8872	22.0343	0.3832	22.1007	0.4225	23.6956	0.0666	20.9644	0.2816	23.1905	0.0930	23.3415	0.0886
30	-1.9888	-1.7554	1	-1.7004	1	-1.6421	0.9929	-1.8681	0.4744	-1.1631	1	-1.7051	1

Çizelge 6.6 incelendiğinde aykırı değer olan 29. gözlemin ağırlığı Huber’de “0.3832”, Hampel’da “0.4225”, Tukey’de “0.0666”, Andrews’de “0.2816” önerilen yöntemlerde ise “0.0930” ve “0.0886” olarak bulunmuştur. Önerilen yöntemlerden bulunan ağırlıklar aslında her bir gözlemin üyelik dereceleri. Bu üyelik dereceleri gözlemlerin modele etkisini göstermektedir. Dolayısıyla Çizelge 6.6’da da görüldüğü gibi aykırı değerler çok küçük üyelik derecesi ile diğer gözlemler ise “1 ya da 1” e yakın üyelik dereceleriyle regresyon modelini etkiler.

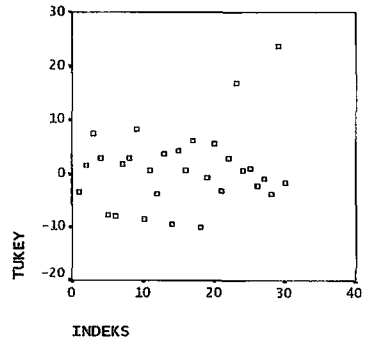
Çizelge 6.7. Bağımsız değişkende aykırı değer olması durumu için regresyon model tahminleri

Yöntem	Sabit	Regresyon Katsayıları			Artıkların
		β_1	β_2	β_3	Kareleri Toplamı
EKK	-1.3516	1.1579 (0.1665)	1.0142 (0.0342)	1.0208 (0.0656)	1434.5
Huber	-0.7383	1.1135 (0.1610)	1.0179 (0.0330)	0.9780 (0.0634)	1473.4
Hampel	-0.8270	1.1329 (0.1580)	1.0182 (0.0309)	0.9773 (0.0594)	1471.9
Tukey	-0.1484	1.0964 (0.1441)	1.0189 (0.0296)	0.9441 (0.0567)	1549.0
Andrews	-1.0692	1.1476 (0.1584)	1.0161 (0.0325)	0.9992 (0.0624)	1443.2
Önerilen 1	-1.1782	1.1021	1.0227	0.9634	1530.0
Önerilen 2	-0.2018	1.0969	1.0183	0.9508	1530.2
Yan	-0.0252	0.1371	0.1273	0.1188	

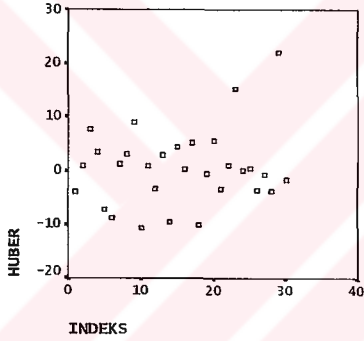
Çizelge 6.7’de EKK, Huber, Hampel, Tukey, Andrews yöntemleri için regresyon model tahminleri ve parantez içinde parametrelerin standart sapmaları, önerilen yöntemlerden elde edilen regresyon model tahminleri ve artıkların kareleri toplamı verilmiştir. Çizelge incelendiğinde önerilen yöntemlerden elde edilen parametre tahmin değerleri klasik Robust yöntemlerle işaretçe aynı ve büyüklükçe çok yakındır.



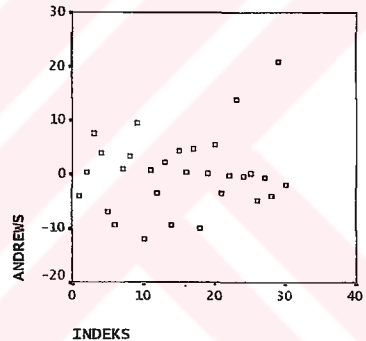
(a) EKK Yöntemi için Artıkların Dağılımı



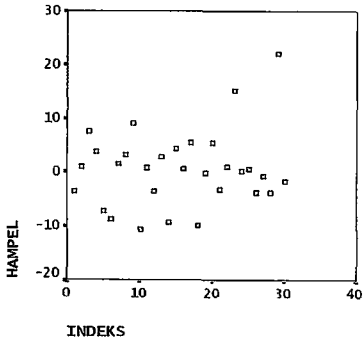
(d) Tukey Yöntemi için Artıkların Dağılımı



(b) Huber Yöntemi için Artıkların Dağılımı

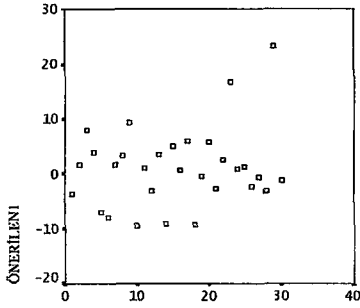


(e) Andrews Yöntemi için Artıkların Dağılımı



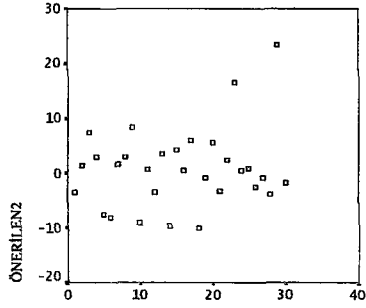
(c) Hampel Yöntemi için Artıkların Dağılımı

Grafik 6.3. Bağımsız değişkende aykırı değer olması durumuna ilişkin EKK, Huber, Hampel, Tukey ve Andrews yöntemleri için artıkların dağılımı



İNDEKS

(a) Önerilen1 Yöntemi için Artıkların Dağılımı



İNDEKS

(b) Önerilen2 Yöntemi için Artıkların Dağılımı

Grafik 6.4. Bağımsız değişkende aykırı değer olması durumuna ilişkin önerilen iki yöntem için artıkların dağılımı

Bu kesimde ele alınan diğer 10 örneğe ilişkin model tahminleri ve Huber yöntemine göre 1'den farklı ağırlık alan gözlemlere karşılık gelen artıklar ve ağırlıklar Huber, Hampel, Tukey, Andrews ve önerilen yöntemler için Çizelge 6.8'de verildi.

Çizelge 6.8. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımsız değişkende aykırı değer olması durumu için çözümleme sonuçları

	EKK	Huber	Hampel	Tukey	Andrews	Önerilen 1	Önerilen 2
	Artık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık
Örnek 1	-7.7396	-7.8005	-7.7592	-8.0278	-7.8650	-5.7614	-8.3550
	10.1695	10.6500	10.5570	10.4924	10.3140	11.7424	10.5808
	-6.1293	-5.9410	-6.0055	-6.0138	-6.0828	-4.3675	-5.9454
	-9.4333	-9.2308	-9.2853	-9.1062	-9.2988	-8.3021	-8.8071
	5.7153	6.1219	5.9818	6.2221	5.9288	7.0681	6.6607
Model Tahmini	7.6918	7.9738	7.9073	8.0202	7.8301	8.9494	8.2634
	7.1132	7.3870	7.3128	7.2828	7.1856	8.8460	7.3506
	11.9245	12.4249	12.2930	12.5426	12.1887	12.9783	13.0200
	4.0589	4.4640	4.3553	4.4881	4.2554	2.3680	4.7567
	1.1979	1.1926	1.1926	1.1811	1.1908	1.2402	1.1656
Örnek 2	1.0414	1.0354	1.0372	1.0332	1.0378	1.0468	1.0265
	0.8660	0.8565	0.8581	0.8610	0.8636	0.8624	0.8616
	9.6340	10.5635	10.4566	10.9882	10.2461	11.8822	11.4562
	14.6219	14.6622	14.7251	14.7438	14.6695	14.9597	14.7907
	-11.4741	-10.6447	-10.6978	-10.3754	-10.9723	-9.7777	-10.0640
Model Tahmini	19.7542	20.4124	20.3793	20.6912	20.1934	21.2600	20.9555
	-13.4641	-13.3375	-13.3094	-13.1536	-13.3307	-12.7510	-13.0094
	-10.8685	-10.7556	-10.7025	-10.6352	-10.7588	-10.3735	-10.5708
	0.3962	-0.2924	-0.2322	-0.6193	-0.0599	-1.3234	-0.9713
	0.6692	0.6665	0.6645	0.6736	0.6705	0.6875	0.6805
Örnek 3	1.0189	1.0157	1.0158	1.0153	1.0167	1.0155	1.0155
	1.0048	1.0237	1.0205	1.0298	1.0164	1.0417	1.0375
							0.1269
Model Tahmini							0.1297

Çizelge 6.8. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımsız değişkende aykırı değer olması durumu için çözümleme sonuçları (Devam)

	EKK	Huber	Hampel	Tukey	Andrews	Önerilen 1	Önerilen 2
	Artık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık Yan
Örnek 3	13.6397	14.9267	0.4192	15.8979	0.0934	16.1652	0.0859
	-8.2777	-7.3560	0.8506	-6.7785	0.7635	-7.8687	0.4240
	6.7618	6.3956	0.9783	6.4866	0.7646	6.7195	0.4378
	10.2686	11.1476	0.5613	11.1255	0.3942	10.6426	0.3834
	-9.5398	-9.7240	0.6434	-9.2040	0.5887	-9.4875	0.4015
	-6.5765	-6.6347	0.9430	-6.6828	0.7697	-6.6218	0.4389
	-7.7084	-7.3954	0.8461	-7.2910	0.7293	-7.6007	0.4274
Model Tahmini	7.8326	7.6777	0.8149	7.4508	0.7182	7.7185	0.4259
	7.1673	7.1304	0.8775	7.0943	0.7426	7.1272	0.4331
	-1.6980	-1.3056	-1.3473	-1.5602	-1.6154	-2.6171	-1.6693
	0.8942	0.9569	0.9594	1.0035	0.9248	1.0096	1.0024
	0.9999	0.9909	0.9902	0.9854	0.9957	0.9914	0.9865
	1.0614	1.0482	1.0507	1.0535	1.0585	1.0668	1.0554
	-22.2347	-23.0241	0.4443	-23.6778	0.1559	-23.8668	0.0781
Örnek 4	10.8363	10.8024	0.9469	11.1538	0.7495	10.9529	0.4296
	11.1286	10.5478	0.9698	10.5803	0.7894	10.7598	0.4312
	-11.2174	-11.3751	0.8992	-11.4156	0.9974	-11.3124	0.4266
	-14.2510	-14.6831	0.6966	-14.5488	0.5953	-14.3729	0.3977
	11.8949	11.5262	0.8874	11.1423	0.7499	11.6082	0.4240
	-2.8629	-2.4612	-2.3914	-1.9546	-2.5217	-2.3303	-1.7554
	0.7538	0.7793	0.7792	0.8123	0.7755	0.8165	0.8240
Model Tahmini	1.0357	1.0389	1.0386	1.0419	1.0382	1.0430	1.0431
	1.0263	1.0271	1.0235	1.0158	1.0225	1.0239	1.0125
	-0.2194	-0.2194	-0.2194	-0.2194	-0.2194	-0.2194	-0.2194
	0.1030	0.1030	0.1030	0.1030	0.1030	0.1030	0.1030
	0.1304	0.1304	0.1304	0.1304	0.1304	0.1304	0.1304
	0.1266	0.1266	0.1266	0.1266	0.1266	0.1266	0.1266
	0.1266	0.1266	0.1266	0.1266	0.1266	0.1266	0.1266

Çizelge 6.8. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımsız değişkende aykırı değer olması durumu için çözümleme sonuçları (Devam)

	EKK Artık	Huber Artık Ağırlık	Hampel Artık Ağırlık	Tukey Artık Ağırlık	Andrews Artık Ağırlık	Önerilen 1 Artık Ağırlık	Önerilen 2 Artık Ağırlık	Yan
Örnek 5	12.3232	11.9899	12.0900	11.6640	12.1053	11.9438	10.9384	0.4914
	-8.1326	-8.4905	-8.3811	-8.8433	-8.3483	-7.9965	-9.7907	0.5619
	-15.0661	-15.9148	-15.7403	-16.3602	-15.5388	-16.2092	-17.4012	0.0944
	-9.3158	-9.5344	-9.4594	-9.7905	-9.4725	-9.6335	-10.3228	0.5292
Model Tahmini								
Sabit	0.3420	1.2229	1.0343	1.7218	0.8327	1.3386	2.9368	0.3671
$\hat{\beta}_1$	0.9522	0.9599	0.9610	0.9499	0.9536	0.9565	0.9318	0.1165
$\hat{\beta}_2$	1.0022	0.9967	0.9977	0.9946	0.9992	0.9889	0.9909	0.1239
$\hat{\beta}_3$	1.0146	1.0027	1.0049	0.9971	1.0086	1.0092	0.9803	0.1225
Örnek 6	9.7480	10.2764	9.9873	10.3428	9.9462	10.3899	10.9100	0.3646
	-13.9952	-14.3321	-14.2898	-14.4006	-14.1672	-14.0076	-14.5079	0.1128
	7.6392	7.9164	7.7405	7.9215	7.7281	8.2508	8.2136	0.5533
Model Tahmini								
Sabit	1.3513	2.2741	1.9814	2.2791	1.7133	1.0940	2.6982	0.3373
$\hat{\beta}_1$	1.1885	1.1924	1.1914	1.2024	1.1934	1.2367	1.2172	0.1522
$\hat{\beta}_2$	0.9988	0.9989	0.9993	1.0001	0.9994	1.0066	1.0007	0.1251
$\hat{\beta}_3$	0.9441	0.9205	0.9295	0.9187	0.9346	0.9285	0.9030	0.1129

Çizelge 6.8. 30 gözlem, 3 bağımsız değişkeni içeren ve bağımsız değişkende aykırı değer olması durumu için çözümleme sonuçları
(Devam)

	EKK	Huber	Hampel	Tukey	Andrews	Önerilen 1	Önerilen 2
Artık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık	Artık Ağırlık Yan
Örnek 9	9.9706	9.2378	9.1750	8.8867	9.5541	9.4810	8.8351
	-9.2698	-10.5261	-10.4303	-11.0390	-9.9687	-10.4616	-11.2737
	-17.1236	-18.8692	-18.7538	-19.6301	-18.1165	-18.7446	-19.9570
	-11.2533	-11.6128	-11.5911	-11.7055	-11.4484	-11.4999	-11.7712
Model Tahmini							
Sabit	0.1135	2.1004	1.9976	2.9989	1.2499	1.9302	3.3515
$\hat{\beta}_1$	1.0097	1.0425	1.0351	1.0544	1.0274	1.0512	1.0638
$\hat{\beta}_2$	0.9721	0.9721	0.9715	0.9717	0.9722	0.9723	0.9723
$\hat{\beta}_3$	1.0304	0.9916	0.9938	0.9725	1.0077	0.9936	0.9653
Örnek 10	15.2517	16.1078	15.8438	16.1085	15.4985	16.8910	16.7719
	-18.4931	-19.0442	-18.9635	-19.3916	-18.8294	-19.3728	-19.6723
	-10.8654	-11.2360	-11.0159	-11.3028	-10.9970	-11.2991	-11.6291
	11.2010	10.9549	0.9905	11.1041	11.1323	11.1735	10.7864
	-10.8377	-10.9816	0.8883	-10.8050	-10.8328	-10.6778	-11.0548
	10.5921	9.9086	0.9845	10.1705	10.3823	9.9756	9.4540
Model Tahmini							
Sabit	-0.1783	0.2645	0.1269	0.1040	-0.1039	-0.3801	0.2655
$\hat{\beta}_1$	0.4739	0.4453	0.4619	0.4418	0.4651	0.4099	0.4112
$\hat{\beta}_2$	1.0225	1.0293	1.0279	1.0341	1.0266	1.0410	1.0393
$\hat{\beta}_3$	1.0495	1.0278	1.0328	1.0267	1.0423	1.0209	1.0140
							0.1268

7. SONUÇ VE TARTIŞMA

Verideki aykırı değer ve gürültünün etkisini azaltma özelliğine sahip yöntemler çok önemli ve yaygındır. Robust istatistikler ve bulanık küme teorisi son 20 yılda bağımsız olarak yavaş yavaş gelişen iki disiplindir. Veri kümesinde aykırı değer olması ve verilerin bulanık olması durumunda yapılan bulanık robust regresyon çalışmaları genellikle etkileşimli yöntemler ya da kümeleme analizinden faydalanan yöntemlerdir. Ancak bu yöntemler sezgiselliği ve uzun bir yineleme sürecini içermektedir. Bu nedenle veri kümesinde aykırı değer olması ve verilerin bulanık olması durumunda uzaklıklara bağlı olarak bulunan artıklara ilişkin üyelik fonksiyonu yapısının sezgiselliği ortadan kaldıracağı ve yinelemeli süreci daha kısaltacağı düşünülerek iki ayrı yöntem önerilmiştir.

Bu çalışmada girdi değişkenlerinin simetrik olmayan üçgensel bulanık sayı $(X_i = (x_i, \underline{x}_i, \bar{x}_i))$, $Y_i = (y_i, \underline{y}_i, \bar{y}_i)$; x 'in kesin sayı ve $Y_i = (y_i, \underline{y}_i, \bar{y}_i)$ simetrik üçgensel bulanık sayı olması ve veri kümesinde aykırı değer olması durumunda çoklu doğrusal regresyon çözümlemesinde parametre tahmini için yinelemeli iki yaklaşım önerilmiştir. Ağırlık matrisinin elde edilmesinde uzaklıklara bağlı olarak tanımlanan üyelik fonksiyonundan yararlanılmıştır. Her bir gözlem üyelik derecesine göre parametre tahminine katılmıştır. Elde edilen ağırlıklar ile ağırlıklandırılmış bulanık en küçük kareler analizi yapılmıştır. Parametre tahmininde aykırı değerden EKK yöntemine göre daha az etkilenen model tahmini elde edilmiştir.

Bu çalışmada tanımlanan üyelik fonksiyonu yardımıyla aykırı gözlem ve diğer gözlemler ayrı ayrı kümelenebilmektedir. Böylece Üçüncü Bölüm'de verilen kümeleme yöntemlerinde içerilen amaç fonksiyonunun en küçüklenmesi ile yapılan kümeleme bu çalışmada önerilen üyelik fonksiyonu yardımıyla daha kısa sürede ve daha az hesaplama sürecini içeren algoritmayla elde edilmiştir.

Y bağımlı değişkeninde aykırı değer olması durumu için Çizelge 6.1’de verilen veri kümesi ele alınmıştır. Bu veri kümesi için elde edilen artıklar ve ağırlıklar Çizelge 6.2’de verilmiştir. Çizelge 6.2’de verilen artıklar incelendiğinde Kesim 5.1’de önerilen yöntemden elde edilen artıklar aykırı değer için (20.8323) gibi büyük değerler, diğer gözlemlerde ise (-0.4967, 0.4054) gibi küçük değerler almaktadır. Kesim 5.2’de önerilen yöntemden elde edilen artıkların aykırı değer için (19.3491) gibi büyük değerler, diğer gözlemlerde ise (0.1284, 0.1379) gibi küçük değerler aldığı görülmektedir. Kesim 5.1’de önerilen yöntemden aykırı değer için üyelik fonksiyonundan elde edilen üyelik (ağırlık) “0.0603” iken diğer gözlemler için “1 ya da 1’e yakın” değerlerdir. Kesim 5.2’de önerilen yöntemden aykırı değer için üyelik fonksiyonundan elde edilen üyelik (ağırlık) “0.0511” iken diğer gözlemler için “1 ya da 1’e yakın” değerler aldığı görülmektedir. Üyelikler yardımıyla aykırı değerler belirlendiği gibi diğer gözlemlerin modele etki dereceleri de ayrı ayrı incelenebilmektedir. 12. gözlemin ağırlığı Huber yönteminde “0.9743”, Hampel yönteminde “1”, Tukey yönteminde “0.8162” ve Andrews yönteminde “0.4350” olarak bulunmuştur. Önerilen yöntemlerde ise bu gözlem kümeye ait olma derecesine göre “0.7503” ve “0.7818” üyelik derecelerini almıştır. Buradan bulunan üyelik derecelerinin klasik yöntemlerle örtüştüğü görülmektedir.

Çizelge 6.3 incelendiğinde önerilen yöntemlerden elde edilen parametre tahminleri ve artıkların kareleri toplamının literatürde yer alan klasik yöntemlerden bulunan parametre tahminlerine yakın olduğu görülmektedir.

Grafik 6.1’de EKK, Huber, Hampel, Tukey ve Andrews yöntemleri için artıkların dağılımları verilmiştir. Grafik 6.2’den görüldüğü gibi aykırı değer büyük artığa sahipken diğer gözlemler sıfıra yakın artığa sahiptir. Böylece elde edilen regresyon model tahmininin uygunluğunun, aykırı değer dışında kalan gözlemler için daha iyi olduğu söylenebilir. Artıkların dağılım grafiğinden,

robust yöntemlerde olduğu gibi önerilen yöntemlerde de aykırı değerler kolaylıkla belirlenebilmektedir.

Çizelge 6.4’de 30 gözlem ve 3 bağımsız değişkeni içeren ve bağımlı değişkende aykırı değer olması durumu için türetilen yapay verilere ilişkin çözümleme sonuçları verilmiştir. Girdilerin bulanık çıktı parametrelerinin kesin olması durumunda parametre tahminleri içerisinde bulanıklığı içermektedir. Elde edilen parametre tahminleri klasik robust yöntemlere benzerdir. Girdilerin bulanık çıktı parametrelerinin bulanık olması durumunda ise yine sonuçlar klasik yöntemlere çok benzerdir. Merkez $\pm$ Yan aralığı klasik yöntemlerden elde edilen parametre tahminlerini içermektedir. Önerilen yöntemlerden elde edilen artık aykırı değer için büyük diğer gözlemler için ise oldukça küçüktür. Ağırlıklar ise aykırı değer için oldukça küçük iken diğer gözlemler için büyüktür. Klasik yöntemlerden bulunan sonuçlarla örtüşmektedir.

Bağımsız değişkende aykırı değer olması durumu için Çizelge 6.5’de verilen veri kümesi ele alınmıştır. Bu veri kümesi için elde edilen artıklar ve ağırlıklar Çizelge 6.6’da verilmiştir. Önerilen her iki yöntemden elde edilen artıklar aykırı gözlem için büyük diğer gözlemler için ise oldukça küçüktür. Ağırlıklar ise aykırı gözlem için oldukça küçük diğer gözlemler için ise “1 ya da 1”e yakın değerlerdir. 18. gözlem Huber yönteminde “0.8399” ağırlığı ile aykırı gözlem iken Hampel yönteminde “1” ağırlığı ile normal gözlem olarak elde edilmiştir. Önerilen yöntemlerde ise bu gözlem kümeye ait olma derecesine göre “0.6955” ve “0.6686” üyelik derecelerini almıştır.

Çizelge 6.7 incelendiğinde önerilen her iki yöntem için elde edilen regresyon katsayılarının ve artıkların kareleri toplamalarının klasik yöntemlere benzer olduğu görülmektedir.

Grafik 6.3’de veri kümesinde aykırı değer olması durumu için Huber, Hampel, Tukey ve Andrews yöntemleri için artıkların dağılımları verilmiştir. Aykırı değerler artıkların dağılımından kolaylıkla ayırt edilebilmektedir. Grafik 6.4 incelendiğinde önerilen her iki yöntem içinde artıkların dağılımından aykırı değerler kolaylıkla görülebilmektedir.

Çizelge 6.8’de 30 gözlem ve 3 bağımsız değişken içeren ve bağımsız değişkende aykırı değer olması durumu için türetilen verilere ilişkin çözümleme sonuçları verilmiştir. Çizelge 6.8 incelendiğinde önerilen yöntemlerden elde edilen parametre tahminlerinin klasik yöntemlere benzer olduğu görülmektedir. Artıklar ise aykırı değer için büyük diğer gözlemler için oldukça küçüktür.

Sonuç olarak önerilen yöntemlerin farklılığı ve üstünlüğü aşağıdaki gibi özetlenebilir.

- Bulanık bağımlı ve bağımsız değişkenleri içeren çoklu doğrusal regresyon modellerinin tahmininde aykırı değerden etkilenmezler,
- Aykırı gözlemlerin tek tek modele etkileri ayrıntılı olarak gözlenebilir,
- Sezgiselliğe izin vermez,
- Robust kümeleme yöntemlerindeki uzun algoritmik yapıyı tanımlanan üyelik fonksiyonu yardımıyla en aza indirger.

Bundan sonra yapılacak çalışmalarda farklı dağılımlar için üyelik fonksiyonu tanımlanarak çözümlemeye gidilebilir.

KAYNAKLAR

- Alpar, R. 1997. Uygulamalı çok değişkenli istatistiksel yöntemlere giriş I. Bağırgan yayınevi, 161-303, 337s., Ankara.
- Apaydın, A. 2000. İstatistik ve optimizasyonda bulanık teori. Ders notu. Ankara.
- Apaydın, A., Kutsal, A. ve Atakan, C. 2002. Uygulamalı istatistik. Kılavuz Yayınevi, 3. Baskı, 343-357, 496s., Ankara.
- Ata, M. 1995. Uyuşumsuz ölçülerin belirlenmesinde klasik en küçük kareler yöntemi ile değişik robust kestirim yöntemlerinin uygulanması ve karşılaştırılması. Yüksek Lisans Tezi (basılmamış). Yıldız Teknik Üniversitesi, 33-35, İstanbul.
- Barni, M., Cappellini, V. and Mecocci, A. 1996. Comments on a possibilistic approach to clustering. IEEE Transactions on Fuzzy Systems, 4 (3); 393-396.
- Candan, M. 1995. Doğrusal regresyon çözümlemesinde sağlam kestiriciler. Yüksek Lisans Tezi (basılmamış). Hacettepe Üniversitesi, 17-37, Ankara.
- Chang, P.T. and Lee, E.S. 1993. Fuzzy linear regression with spreads unrestricted in sign. Computer Math. Applic., 28 (4); 61-70.
- Chang, P.T. and Lee, E.S. 1994. Ranking of fuzzy sets based on the concept of existence. Computer Math. Applic., 27 (9); 1-21.
- Chang, P.T. and Lee, E.S. 1996. A generalized fuzzy weighted least-squares regression. Fuzzy Sets and Systems, 82; 289-298.
- Chang, Y.H.O. 2001. Hybrid fuzzy least-squares regression analysis and its reliability measures. Fuzzy Sets and Systems, 119; 225-246.
- Chang, Y.H.O., and Ayyub, B.M. 2001. Fuzzy regression method-a comparative assessment. Fuzzy Sets and Systems, 119; 187-203.
- Chen, Y.S. 2001. Outliers detection and confidence interval modification in fuzzy regression. Fuzzy Sets and Systems, 119; 259-272.

- Cherkassky, V. and Mulier, F. 1998. Learning from data concepts, theory, and methods. 389-420, 441
- Dave, R.N. 1993. Robust fuzzy clustering algorithms. IEEE Transactions on Fuzzy Systems, 1281-1286.
- Dave, R.N. and Krishnapuram, R. 1996. M-estimators and robust fuzzy clustering. IEEE Transactions on Fuzzy Systems, 400-404.
- Dave, R.N. and Krishnapuram, R. 1997. Robust clustering methods: A unified view. IEEE Transactions on Fuzzy Systems, 5 (2); 270-293.
- Dave, R.N. and Sen, S. 1998. Generalized noise clustering as a robust fuzzy c-m-estimators model. IEEE Transactions on Fuzzy Systems; 256-260.
- Diamond, P. 1988. Fuzzy least squares. Information Science, 46; 141-157.
- Elmas, Ç. 2003. Bulanık mantık denetleyiciler (Kuram, Uygulama, Sinirsel Bulanık Mantık). Seçkin, 22-40, 230s., Ankara.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. 1986. Robust statistics. John Wiley & Sons, 149-172, 502s., New-York.
- Hogg, R.V. 1979. Statistical robustness: one view of its use in applications today. The American Statistician, 33 (3); 108-115.
- Huber, P.J. 1981. Robust Statistics. John Wiley & Son, 1-20, 153-243, 308s., New York.
- Hund, E., Massart, D.L. and Smeyers-Verbeke, J. 2002. Robust regression and outlier detection in the evaluation of robustness tests with different experimental designs. Analytica Chimica Acta, 22031; 1-21.
- Huynh, H. 1982. A comparison of four approaches to robust regression. Psychological Bulletin, 92 (2); 505-512.
- Klir, G.J. and Yuan, B. 1995. Fuzzy sets and fuzzy logic, Theory and applications. Prentice Hall, 357-362, 574s., USA.
- Krishnapuram, R. and Keller, M.J. 1993. A possibilistic approach to clustering. IEEE Transactions on Fuzzy Systems, 1(2); 98-110.

- Krishnapuram, R. and Keller, M.J. 1996. The possibilistic c-means algorithm: insights and recommendations. *IEEE Transactions on Fuzzy Systems*, 4 (3); 385-393.
- Lai, Y.J. and Hwang, C.L. 1992. *Fuzzy mathematical programming*. Springer-Verlag, 67-71, 301s., Germany.
- Lin, C.C. and Chen, A.P. 2003. Fuzzy discriminant analysis with outlier detection by genetic algorithm. *Computers & Operations Research*, 31 (6); 877-888.
- Nasraoui, O. 2004. A brief overview of robust clustering techniques. Web Sites: http://arcer.ee.memphis.edu/www.ee.memphis.edu/people/faculty/nasraoui/MY_TUTORIALS/RobustClustering/RobustClustering
- Özelkan, E.C. and Duckstein, L. 2000. Multi-objective fuzzy regression: a general framework. *Computers & Operations Research*, 27; 635-652.
- Pal, N.R. and Bezdek, C.J. 1997. A mixed c-means clustering model. *Fuzzy IEEE*, 11-21.
- Redden, D.T. and Woddall, W.H. 1996. Further examination of fuzzy linear regression. *Fuzzy Sets and Systems*, 79; 203-211.
- Rousseuw, P.J. and Leroy, A.M. 1997. *Robust regression and outlier detection*. John Wiley & Sons, 1-154, 329s., New York.
- Sakawa, M. and Yano, H. 1992. Multiobjective fuzzy linear regression analysis for fuzzy input-output data. *Fuzzy Sets and Systems*, 47; 173-181.
- Stephenson, D. 2000. Web Sites: <http://www.gfi-wib.bo>
- Tanaka, H., Uejima, S. and Asai, K. 1982. Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man, and Cybernetics*, 12 (6); 903-907.
- Tanaka, H. and Guo, P. 1999. *Possibilistic data analysis for operation research*. A Springer-Verlag Company, 15-35, 182s., Germany.
- Xu, R. and Li, C. 2001. Multidimensional least-squares fitting with a fuzzy model. *Fuzzy Sets and Systems*, 119; 215-223.

- Yang, M.S. and Ko, C.H. 1997. On cluster-wise fuzzy regression analysis. IEE Transaction on Systems, Man and Cybernetics Part B: Cybernetics, 27 (1); 1-13.
- Yang, M.S. and Lin, T.S. 2002. Fuzzy least-squares linear regression analysis for fuzzy input-output data. Fuzzy Sets and Systems, 126; 389-399.
- Yang, M.S., and Liu, H.H. 2003. Fuzzy least squares algorithms for interactive fuzzy linear regression models. Fuzzy Sets and Systems, 135; 305-316.
- Yun-Hsi, O.C. 2001. Hybrid fuzzy least-squares regression analysis and its reliability measures. Fuzzy Sets and Systems, 119; 225-246.
- Wang, L.X. 1997. A course in fuzzy systems and control. Prentice- Hall, 20-30, 424s., USA.
- Watada, J. and Yabuuchi, Y. 1994. Fuzzy robust regression analysis. Fuzzy/IEEE'94, Theird IEEE International Conference on Fuzzy Systems, 1-7.
- Watada, J. and Yabuuchi, Y. 1995. Fuzzy robust regression analysis based on a hyperelliptic function. Fuzzy-IEEE/IFES'95, Proceeding of The International Joint Conference of The Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium, IV; 1841-1848.
- Watada, J., Kawaura, T., Yabuuchi, Y., Ikegame, H., Sato, S., Watanebe, K., and Tohyama, M. 2000. Analysis of software bug structures based on fuzzy regression model. The Second World Congress for Software Quality; 689-694.
- <http://www.xplore.stat-de>

ÖZGEÇMİŞ

1975 yılında Antalya’da doğdu. İlk, orta ve lise öğrenimini Antalya’da tamamladı. 1993 yılında girdiği Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü’nden 1997 yılında İstatistikçi unvanıyla mezun oldu. Ağustos 1997’de Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü’nde Araştırma Görevlisi olarak göreve başladı. 1998 yılında Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında Yüksek Lisans öğrenimine başladı. 1999 yılında Yüksek Lisans öğrenimini tamamladı. Aynı yıl Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında Doktora öğrenimine başladı. Kasım 1999’da Ankara Üniversitesi Fen Fakültesi İstatistik Bölümünde Araştırma Görevlisi olarak göreve başladı.