T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

DİK AĞIRLIKLANDIRMANIN M-REGRESYON ÜZERİNE ETKİSİ

Bengü OCAK

YÜKSEK LİSANS

İstatistik Anabilim Dalı

Eylül-2010
KONYA

TEZ KABUL VE ONAYI

Bengü OCAK tarafından hazırlanan "Dik ağırlıklandırmanın M-regresyon üzerine etkisi" adlı tez çalışması 21/09/2010 tarihinde aşağıdaki jüri tarafından oy birliği ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Yrd. Doç. Dr. Necati TAŞKARA	....
Danışman Yrd. Doç. Dr. Mustafa SEMİZ	....
Üye Doç. Dr. Coşkun KUŞ	....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Bayram SADE
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Bengü OCAK

23.08.2010

ÖZET

YÜKSEK LİSANS TEZİ

DİK AĞIRLIKLANDIRMANIN M-REGRESYON ÜZERİNE ETKİSİ

Bengü OCAK

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Yrd. Doç. Dr. Mustafa SEMİZ

2010, 66 Sayfa

Jüri
Doç. Dr. Coşkun KUŞ
Yrd. Doç. Dr. Mustafa SEMİZ
Yrd. Doç. Dr. Necati TAŞKARA

Regresyon analizinde en çok kullanılan en küçük kareler yöntemi hata terimleri üzerine yapılan normallik varsayımlarının bozulması sebebiyle kullanılamamaktadır. Veri setindeki aykırı gözlemler bu sebeplerden biridir. Bu durum için geliştirilen m-regresyon yöntemi gözlemler üzerine farklı ağırlıklandırmalar kullanabilmektedir. Bu çalışmada gözlemlerin modele olan dik uzaklıklarına bağlı ağırlıklandırma temel alınmış ve dik m-regresyon yöntemi geliştirilmiştir. Monte Carlo simülasyon çalışmasıyla önerilen yöntem kısmi sağlam m-regresyon yöntemi ile karşılaştırılmıştır.

Anahtar Kelimeler: En küçük kareler, regresyon analizi, en çok olabilirlik tahmini, m-regresyon, kısmi sağlam m-regresyon.

ABSTRACT

MS THESIS

THE EFFECT OF THE PERPENDICULAR WEIGHT ON M-REGRESSION

Bengü OCAK

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF SELÇUK UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

Advisor: Asst. Prof. Dr. Mustafa SEMİZ

2010, 66 Pages

Jury
Assoc. Prof. Dr. Coşkun KUŞ
Asst. Prof. Dr. Mustafa SEMİZ
Asst. Prof. Dr. Necati TAŞKARA

The common used method in regression analysis is Least Squares regression method. This method can't used because of violation of the normality assumptions on residuals in regression. M-regression method developed for this situation can use different weight approach on observations. In this study, the weight method based on the perpendicular distance is considered and the perpendicular m-regression is developed. This suggested method and partial robust m-regression method were compared by using Monte Carlo simulation study.

Keywords: Least Squares, regression analysis, maximum likelihood estimator, m-regression, partial robust m-regression.

ÖNSÖZ

Bu tez çalışmasının konu seçiminde ve gerçekleşmesinde yardımını esirgemeyen değerli danışman hocam Yrd. Doç Dr. Mustafa SEMİZ'e ve tez çalışmam boyunca bana manevi destek olan aileme teşekkürlerimi sunarım.

Bengü OCAK

KONYA-2010

İÇİNDEKİLER

KISALTMALAR

AAKEKK	: Adımsal Ağırlıklı Kısmi En Küçük Kareler
BLUE	: Best Linear Unbiased Estimator (En iyi Doğrusal Yansız Tahmin Edici)
DMR	: Dik M-Regresyon
EKK	: En Küçük Kareler
EMS	: En Küçük Mutlak Sapma
GDR	: Genelleştirilmiş Doğrusal Regresyon
GEKK	: Genelleştirilmiş En Küçük Kareler
GKEKK	: Sağlam Kısmi En Küçük Kareler
GMM	: Genelleştirilmiş Momentler Yöntemi
KDR	: Klasik Doğrusal Regresyon
KEKK	: Kısmi En Küçük Kareler
KSMR	: Kısmi Sağlam M-Regresyon
KSM-regresyon	: Kısmi Sağlam M-Regresyon
LAD	: En Küçük Mutlak Sapmalar
MAD	: Median of Absolute Deviation (Mutlak Sapmalaların Medyanı)
MR	: M-Regresyon
MSM	: Mutlak Sapmaların Medyanı

1. GİRİŞ

Değişkenler arasındaki ilişkilerin belirlenmesinde en sık kullanılan yöntemlerden biri regresyon analizidir. Regresyon analizinde verileri oluşturan ölçümler bağımlı (*Y*) ve bağımsız ($X_1, X_2, \ldots, X_p$) değişkenler olarak ifade edilir. Ölçümlerin her zaman beklenen sınırlar içerisinde olmadığı gerçek verilerde gözlenmektedir.

Regresyon analizindeki temel amaç bağımlı değişkenin bağımsız değişkenleri uygun bir modelde kullanarak en az hata ile tahmin etmek olmasına karşın beklenenden uzak ölçümler ya da başka bir ifadeyle aykırı gözlemler (outliers) bu duruma bilinen yöntemlerin uygulanmasını önermez.

Aykırı gözlemlerin belirlenmesi, bu gözlemlerin regresyon analizine etkileri ve bu durumda izlenebilecek farklı regresyon yöntemleriyle ilgili literatür çalışması Bölüm 2'de verilmiştir.

EKK regresyon yönteminin hata terimleri üzerine birçok varsayımı ölçümlerden kaynaklanan aykırı gözlemler nedeniyle ihlal edilmiş olur ve tahmin modeli geçerliliğini kaybeder. Bu sebeple regresyon analizinde bağımlı değişkenin tahmin edilmesinden çok bağımsız değişkenlerle ilişkisini ortaya koyacak ve hata terimlerinin normallik varsayımlarına bağlı kalmayan sağlam (robust) yöntemler geliştirilmiştir. Bu yöntemler içerisinde en çok kullanılan Kantil, LAD ve M-regresyon yöntemleri Bölüm 3'de sunulmuştur.

Bu yöntemlerde amaç bağımlı değişken değerlerini değil modeldeki parametreleri daha iyi tahmin etmektir. Bu tahmin edicilerden biri de En Çok Olabilirlik (MLE) tahminidir. Bu tahmin ediciyi temel alan M-regresyon ve KSM-regresyon Bölüm 3'te aykırı gözlemler içeren tek değişkenli ve çoklu doğrusal modeller üzerinde örneklerle anlatılmıştır.

Veri seti içerisindeki aykırı gözlemlerin belirlenmesi tamamen farklı bir konu olup regresyon işlemini nasıl etkilediği incelenmiştir. Gözlemlerin EKK tahmin modeline olan dik uzaklıklarına bağlı DMR yöntemi Bölüm 4'te örneklerle verilmiştir. Bölüm 4 içerisinde son olarak en yeni olan KSM-regresyon ve önerilen Dik M-regresyon yöntemleri iki farklı model ve farklı durumlar için Monte Carlo simülasyon çalışması ile karşılaştırılmıştır.

Son olarak Bölüm 5'de DMR ve KSMR yöntemlerinin etkinlikleri için yapılan karşılaştırmalar yorumlanmış ve bu konuyla ilgili gelecekte yapılabilecek çalışmalara değinilmiştir.

2. KAYNAK ARAŞTIRMASI

Kaynak araştırması bölümü iki alt bölümde aykırı gözlemlerin belirlenmesi ve M-regresyon konularına yönelik hazırlanmıştır.

2.1. Aykırı Gözlemlerle İlgili Kaynak Araştırması

Rousseeuw ve Van Zomeren (1990) çok değişkenli verilerde aykırı gözlemlerin belirlenmesinin özellikle birkaç aykırı gözlem olduğunda belirlenmesinin son derece zor olduğunu çünkü örnek ortalaması ve kovaryans matrisinin bu aykırı gözlemlere de bağlı olduğunu ifade etmişlerdir. Bu durum aykırı gözlemlerin maskelenmesi olarak ifade edilmiştir. Maskelemenin ortadan kaldırılması için uzaklıkları belirleyecek Mahalanobis uzaklığı yerine daha sağlam bir yöntem önermişlerdir. Ortalama vektörü ve kovaryans matrisi gözlemlerin yarısını içeren elipsoid içerisine düşen gözlemlerden tahmin edilerek kullanılmaktadır.

Penny (1996) makalesinde çok değişkenli dağılımda bir aykırı gözlemin belirlenmesinde uygun kritik değerin $[p(n-1)/(n-p)]F_{p,n-p}$ olmadığını fakat $[p(n-1)^2 F_{p,n-p-1}]/[n(n-p-1+pF_{n-p-1})$ olduğunu göstermiştir.

Pena ve Yohai (1999) büyük regresyon problemleri için (çok gözlemli ve çoklu regresyon problemleri) aykırı gözlemlerin belirlenmesi için hızlı bir yöntem geliştirmişlerdir. Regresyon analizini her bir gözlemin çıkarılması sonucu elde edilen hataların ve dağılımlarının incelenmesine dayanan aykırı gözlemlerin tespiti bilgisayara üzerinde hızlı olarak tekrarlanabildiğinden daha pratik bir yol olarak önerilmiştir.

Adnan, Mohamad ve Setan (2003) bir doğrusal regresyon modelinde çoklu aykırı gözlemlerin belirlenmesi için bir yöntem geliştirmişlerdir. Öncelikle veri setindeki gözlemlerin 3 temel noktada aykırı olabilecekleri konusu ayrıntısıyla anlatılmıştır. Bir noktanın Y ekseni üzerinden, X ekseni üzerinden veya regresyon doğrusuna olan uzaklığı açısından aykırılık gösterebilmektedir. Geliştirilen yöntem sağlam bir yöntem olup karelerinin kesilmesine bağlı olarak belirlenen regresyon ve tekil bağımlı kümeleme analizi dikkate alınarak belirlenmiştir.

Liu, Shah ve Jiang (2004) bu makalede aykırı gözlemlere dirençli verilerdeki filtreleme ve temizleme yöntemi önerilmiştir. Önerilen veri filtreleme ve temizleyicisi işlem modelinin on-line aykırı direnç tahminlerini içerir ve belirleme ve temizleme için uyarlanmış Kalman filtreleme yöntemiyle birleştirir. Bu metodun diğer metotlara göre dört avantajı vardır: işlem modelinin ön bilgisine ihtiyaç yoktur; otokorelasyona sahip verilerde kullanılabilir; on-line kullanılabilir; verilerdeki sadece aykırı gözlemleri temizler (belirler ve değiştirir) verideki diğer bilgileri korur.

Saltenis (2004) bu makalede çoklu boyuttaki noktaların arasındaki ikili uzaklıkların dağılımların özelliklerini temel alan aykırı gözlem belirleme metodu önermiştir. Temel fikir her bir veri noktası için aykırılık faktörünün değerlendirilmesidir. Faktör, aykırı gözlemin değerine bakmaksızın veri setindeki yerine sıralama olarak kullanır. Minimum faktör değerleri ile seçilen noktalar aykırı gözlemleri belirleyebilir. Bu yöntemin iki avantajı vardır: aykırı gözlemlerin belirlenmesinde herhangi bir parametre seçimine gerek yoktur ve belirleme kümeleme algoritmasında bağlı değildir.

Wu ve Lee (2006) normal örneklemde örneklem ortalamasının mutlak sapmasını minimize ederek üst veya alt aykırı gözlemlerin sayısının belirlenmesi için en küçük mutlak sapma (EMS) metodunu önermişlerdir. Bu metot adımsal nümerik analiz işlemi kullanmaktadır ve makalede bu işlem için hesaplama algoritması verilmiştir.

Filzmoser, Marona ve Werner (2008) bu makalede dönüştürülmüş uzayda aykırı gözlemlerin belirlenmesi için temel bileşenlerin basit özelliklerini içeren bir algoritma sunulmuş ve bu algoritma yüksek boyutlu verilerde önemli hesaplama avantajları sağlamaktadır. Bu yaklaşım aykırı gözlemlerin belirlenmesi için kullanılan diğer mevcut yöntemlere göre daha az hesaplama işlemi gerçekleştirir ve çok daha geniş veri setleri üzerinde kullanım kolaylığı sağlar.

Fauconnier ve Haesbroeck (2009) en küçük kovaryans determinant tahmin edicisiyle aykırı gözlemlerin belirlenmesi için bir yöntem önermişlerdir. Bu yöntem hesaplama gücünün geliştirilmesinin yanı sıra çok sayıdaki mevcut sağlam yöntemlerin geliştirilmesine katkıda bulunulmuştur. Dolayısıyla bu makalenin amacı çok değişkenli veri setinde aykırı gözlemlerin güvenilir bir yolla belirlenmesi için kullanıcılara yardımcı olmaktır.

Willems, Joe ve Zamar (2009) bu makalede çok değişkenli veriler için çoklu aykırı gözlemlerin belirlenmesinde kullanılan sağlam tahmin edicilerin uygunluğu için birkaç analiz yöntemi önermişlerdir. Buradaki temel amaç, aykırı gözlemlerin ayrı kümelerden mi yoksa tesadüfi olarak mı yayıldıklarının belirlenmeye çalışılmasıdır. Burada görsel ayrımlamanın ve veri yapısının diğer özelliklerinin belirlenmesi için Mahalanobis uzaklıkları ve doğrusal projeksiyonlar kullanılmıştır. Bu yaklaşım iki boyutlu saçılım grafiklerinde görülemeyen çok değişkenli veri yapıları için uygundur.

2.2. M-Regresyonla İlgili Kaynak Araştırması

Draper ve Smith'in (1981) bu kitabı regresyon analizinin her konusunu detaylı olarak işleyen bir çalışmadır. Regresyon modellerinin yapısında değişkenlerin kendi ve birbirleri arasındaki özellikleri ve yapıları; modellerin yapıları, seçilmesi, uygunluğu ve gözlemlerin incelenmesi konuları ayrıntısıyla açıklanmıştır. Aykırı gözlemlerin tanımı da yapılmış ve modele olan etkilerinden açıkça bahsedilmiştir.

Wold ve Krishnaiah (1973) tarafından bu kitapta regresyon yöntemi ve özellikle Kısmi En Küçük Kareler (KEKK) regresyon yöntemi ayrıntısıyla incelenmiştir. KEKK yöntemi karşılaşılan çoklu bağlantı ve değişken sayısının gözlem sayısından fazla olması durumuna çözüm olarak anlatılmıştır.

Huber (1981) bu kitapta üç temel tahmin edici ayrıntısıyla anlatılmıştır. M-tahmin edicisi; Maksimum En Çok Olabilirlik tahmin edicisi, L-tahmin edicisi; Sıra istatistiklerinin Doğrusal kombinasyonu ve R-tahmin edicisi; Sıra istatistiklerinden tahmin edicilerin belirlenmesidir. Bizim çalışmamızın temelini oluşturan M-tahmin edicilerinin asimtotik ve diğer özellikleri de bu kitapta anlatılmıştır. Kitabın 7. bölümünde özellikle regresyon analizinin anlatılması M-regresyon konusuna temel oluşturmaktadır.

Dempster, Laird ve Rubin (1980) bu kitapta M-regresyon tahmin edicisinin ağırlıklı EKK yöntemiyle ifade edilebileceğini ve adımsal işlemlerle ağırlıklandırmanın tekrarlanarak hesaplanabileceğini göstermektedir.

Rousseeuw ve Leroy (1987) regresyonun M-tahmin edicileri üzerine ayrıntılı bilgi içeren en önemli kitaplardan biridir.

Wakeling ve MacFie (1992) KEKK üzerine yapılan birçok bilimsel çalışmanın yanı sıra bu makale Sağlam KEKK (GKEKK) yaklaşımının ilk önerildiği çalışmadır. Bu çalışmada bütün EKK regresyonlarının KEKK algoritmasında sağlam regresyonla GKEKK yöntemini önermişlerdir.

Cummins ve Andrews (1995) göstermiştir ki Adımsal Ağırlıklı Kısmi EKK (AAKEKK) algoritmasının yeni bir düzenlemesiyle Kısmi (Sağlam) M-regresyon hesaplanabilir. Ayrıca bu makalede sağlam çözümler için başlangıç noktasının ve ağırlıklandırma yönteminin çok iyi belirlenmesinin önemi vurgulanmıştır. Test edilen bir çok ağırlıklandırma fonksiyonundan $f(z,c) = 1 \Big/ \left(1+\left|\frac{z}{c}\right|\right)^2$ fonksiyonu "adaletli:fair" fonksiyon olarak adlandırılmıştır.

Hubert ve Vanden Branden (2003) bu makale öncelikle Adımsal Ağırlıklı Kısmi EKK (AAKEKK) regresyon yöntemini eleştirmektedir. Bağımsız değişkenlerin dikkate alınmasıyla belirlenen aykırı gözlemlerin sonucu etkilemediği ve sonucu yanlış belirlediğini ifade etmişlerdir. Her tür aykırı gözleme duyarlı olan fakat daha fazla bilgisayar zamanı harcayan Sağlam SIMPLS (GSIMPLS:RSIMPLS) algoritmasını geliştirmişlerdir.

Serneels, Croux, Filzmoser ve Van Espen (2005) bu makale hazırlanan tezin başlangıç noktasını oluşturmaktadır. M-regresyonun kısaca anlatılmasından sonra ağırlık yönteminde yapılan değişiklikle Kısmi Sağlam M-regresyon yaklaşımı ortaya konmuştur. Gözlemin regresyon doğrusuna dikey uzaklığın (w_i^r) yanı sıra gözlem noktasının veri kümesine olan yatay uzaklığı da (w_i^x) ağırlık yönteminde çarpımsal olarak ($w_i = w_i^r w_i^x$) dikkate alınmıştır. Bu makalede önerilen Kısmi Sağlam M-regresyon yaklaşımının etkinliği Monte Carlo simülasyon çalışmasıyla desteklenmiştir.

3. REGRESYON YÖNTEMLERİ

3.1. Klasik Çoklu Doğrusal Regresyon Analizi ve En Küçük Kareler Yöntemi

3.1.1. Klasik çoklu doğrusal regresyon modeli, varsayımları ve en küçük kareler yöntemi

Çoklu doğrusal regresyon analizi, bağımlı değişken (Y) ile bağımsız değişkenler $(X_1, X_2, ..., X_k)$ arasındaki model belirlemek ve bu model yardımıyla istatistiksel sonuçları elde etmek amacıyla kullanılan istatistiksel yöntemlerden biridir.

Regresyon analizinde model parametrelerini tahmin etmek amacıyla çeşitli yöntemler kullanılmaktadır. Bu yöntemlerden en çok kullanılanı, hata terimleri karesini minimum yapmayı amaçlayan "En Küçük Kareler (EKK)" yöntemidir.

Çoklu doğrusal regresyon modeli $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i$ $i = 1,2,...,n$ veya matris gösterimi ile

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{3.1}$$

şeklinde gösterilebilir. Burada *Y*, nx1 boyutlu vektörle modeldeki bağımlı değişken vektörü; *X*, nxp boyutlu matrisle bağımsız değişkenler matrisi; β, px1 boyutlu vektörle parametre vektörü ve ε, nx1 boyutlu vektörle hata terimleri vektörüdür. (3.1) modelindeki parametrelerin EKK yöntemi ile parametre tahminlerinin yapılabilmesi için model üzerinde bazı varsayımlar yapılmıştır.

Bu varsayımlar:

1) $E(\varepsilon_i) = 0$

Hata terimlerinin beklenen değeri (ortalaması) sıfıra eşittir.

2) $Var(\varepsilon_i) = \sigma^2 \quad i = 1,2,...,n$

Hata terimleri σ^2 ile sabit varyanslıdır.

$Kov(\varepsilon_i, \varepsilon_j) = 0 \; i, j = 1,2,...,n \;\; i \neq j$

Hata terimlerinin birbirini takip eden değerleri arasında ilişki yoktur. Başka bir ifade ile otokorelasyon yoktur.

$E(\varepsilon\varepsilon') = \sigma^2 I$

Hata terimi varyans-kovaryans matrisi $\sigma^2 I$ ’dır.

3) $\varepsilon_i \sim N(0, \sigma^2)$

Hata terimleri sıfır ortalama ve σ^2 varyansı ile normal dağılıma sahiptir.

4) X_i ve ε_i ’ler birbirinden bağımsızdır.

5) X ’ler sabittir.

6) (XX') matrisi singüler olmayan bir matristir.

7) X matrisinin sütun vektörleri lineer bağımsızdır.

8) Model doğru olarak belirlenmiştir.

9) $n \geq k$ ’dır.

10) $rank(X) = k$

(3.1) modelinin yukarıdaki varsayımlar doğrultusunda EKK tahmin edicisi aşağıdaki gibidir. EKK tahmin edicisi,

$$\hat{\beta} = (X'X)^{-1} X'Y \tag{3.2}$$

(3.2) eşitliği ile bulunan $\hat{\beta}$, β ’nın en iyi doğrusal yansız tahmin edicisidir. Ayrıca,

$$E(\hat{\beta}) = E[(X'X)^{-1} X'Y] = E[(X'X)^{-1} X'(X\beta + \varepsilon)] = \beta$$

$$\begin{aligned} Var(\hat{\beta}) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\ &= E\left\{ \left[(X'X)^{-1} X'(X\beta + \varepsilon) - \beta\right] \left[(X'X)^{-1} X'(X\beta + \varepsilon) - \beta\right]' \right\} \\ &= E[(X'X)^{-1} X'\varepsilon\varepsilon'X(X'X)^{-1}] \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

olarak bulunur.

(3.1) modeli için EKK tahmin edicisi, tüm lineer ve yansız tahmin ediciler arasında en küçük varyansa sahip tahmin edicidir (Gauss-Markov teoremi).

3.1.2. Genelleştirilmiş çoklu doğrusal regresyon modeli, varsayımları ve genelleştirilmiş en küçük kareler yöntemi

(3.2) eşitliği ile bulunan EKK tahmin edicileri yansızdır. Ancak etkinlik özelliği sağlanmayabilir. Bunun için farklı yöntemlerle kıyaslama yapılması gerekebilir. Bu amaçla hata terimi varyans-kovaryans matrisini de regresyona dahil ederek parametre tahminlerini yapmayı amaçlayan "Genelleştirilmiş En Küçük Kareler (GEKK)" yönteminin incelenmesi uygun olacaktır.

Genelleştirilmiş doğrusal regresyon modeli ile klasik doğrusal regresyon modeli hata terimleri üzerine yapılan varsayımlardan dolayı birbirinden farklıdır.

Genelleştirilmiş doğrusal regresyon modeli

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad i = 1,2,\ldots,n \tag{3.3}$$

veya matris gösterimi ile

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{3.4}$$

şeklinde ifade edilebilir.

Burada;

$\underline{Y} : n \times 1$ boyutlu bağımlı değişkenlere ait gözlem vektörünü

$\underline{X} : n \times k$ boyutlu bağımsız değişkenlere ait gözlem matrisini

$\underline{\beta} : n \times 1$ boyutlu regresyon katsayıları vektörünü

$\underline{\varepsilon} : n \times 1$ boyutlu hata terimleri vektörünü temsil etmektedir.

Bu model ile ilgili varsayımlar:

1) $E(\varepsilon_i) = 0$

Hata terimlerinin beklenen değeri (ortalaması) sıfıra eşittir.

2) $Var(\varepsilon_i) = \sigma_i{}^2 \qquad i = 1,2,\ldots,n$

Hata terimleri değişen varyanslıdır.

$Kov(\varepsilon_i, \varepsilon_j) = \sigma_{ij}$ $\quad i, j = 1,2,...,n \quad i \neq j$

Hata terimlerinin birbirini takip eden değerleri arasında ilişki vardır. Başka bir ifade ile otokorelasyon mevcuttur.

3) $\varepsilon_i \sim N(0, \sigma^2)$

Hata terimleri sıfır ortalama ve σ^2 varyansı ile normal dağılıma sahiptir.

4) X_i ve ε_i ’ler birbirinden bağımsızdır.

5) X ’ler sabittir.

6) (XX') matrisi singüler olmayan bir matristir.

7) X matrisinin sütun vektörleri lineer bağımsızdır.

8) Model doğru olarak belirlenmiştir.

9) $n \geq k$ ’dır.

10) $rank(X) = k$

Klasik Doğrusal Regresyon (KDR) modeli varsayımları ile genelleştirilmiş doğrusal regresyon (GDR) modeli varsayımlarına bakıldığında 2. varsayım dışındaki tüm varsayımlar özdeştir. KDR modelinde hata terimleri sabit varyanslı ve otokorelasyona sahip değilken genelleştirilmiş doğrusal regresyon modelinde hata terimleri değişen varyanslı ve/veya otokorelasyonlu olabilir. Bu bilgiler ışığında KDR modelinde, hata terimleri varyans-kovaryans matrisi;

$$E(\varepsilon\varepsilon') = \sigma^2 I_n \tag{3.5}$$

iken, GDR modelinde hata terimleri varyans-kovaryans matrisi:

$$E(\varepsilon\varepsilon') = \sigma^2 \Omega \tag{3.6}$$

şeklinde olup Ω matrisi, simetrik ve pozitif tanımlı bir matristir. Böylece, Ω^{-1} de pozitif tanımlı bir matris olacaktır.

Dikkat edilecek olursa $\Omega = I$ olursa, ki bu durum değişen varyanslılık ve otokorelasyon olmadığı zaman gerçekleşir, GEKK tahmin edicisi EKK tahmin edicisi haline dönüşür. Böylece en küçük kareler tahmin edicisinin, genelleştirilmiş en küçük kareler tahmin edicisinin özel bir hali olduğu söylenebilir.

Hata terimleri değişen varyanslı ise, GDR modelinde hata terimleri varyans-kovaryans matrisi:

$$E(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

iken, hata terimleri arasında birinci dereceden otokorelasyon olması durumunda, hata terimleri varyans-kovaryans matrisi

$$E(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}$$

şeklindedir.

Modelde değişen varyanslılık ve/veya otokorelasyon olması halinde parametre tahminleri EKK yöntemi yapılırsa yansız, tutarlı ancak etkin olmayan parametre tahminleri elde edilir ki bu da elde edilen parametre tahminlerinin en iyi doğrusal yansız tahmin edici (BLUE) olma özelliğini ortadan kaldırır.

(3.3) modeli için parametre tahminleri bulunurken EKK yöntemi uygulandığında;

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

ve

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

olarak bulunur. GDR modeli için verilen varsayımlar altında;

$$E(\hat{\beta}) = E[(X'X)^{-1}X'Y] = E[(X'X)^{-1}X'(X\beta+\varepsilon)] = \beta + (X'X)^{-1}X'\varepsilon = \beta$$

olarak bulunur ve $\hat{\beta}$'nın β için hala yansız bir tahmin edici olduğu kolaylıkla görülebilir. Bununla beraber,

$$\begin{aligned} Var(\hat{\beta}) &= E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' \\ &= E\left\{\left[(X'X)^{-1}X'(X\beta+\varepsilon)-\beta\right]\left[(X'X)^{-1}X'(X\beta+\varepsilon)-\beta\right]'\right\} \\ &= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\ &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned} \tag{3.7}$$

olarak elde edilir ki görüldüğü gibi EKK yöntemi GDR modeli söz konusu iken uygun bir tahmin edici olmamaktadır.

Aitken, GDR modelinin parametre tahminleri için alternatif bir yöntem geliştirmiştir. "Genelleştirilmiş En Küçük Kareler" veya "Aitken Tahmin Edicisi" olarak bilinen bu yöntemde genelleştirilmiş model bazı varsayımlar ve şartlar altında değişen varyanslılık ve otokorelasyonun varlığı da dikkate alınarak dönüştürülmekte ve dönüştürülen bu yeni model klasik modelin tüm varsayımlarını sağlamakta dolayısıyla modele EKK yöntemi uygulanarak en iyi doğrusal yansız tahmin edicilerin elde edilmesi sağlanmaktadır.

GDR modelinin dönüşümü aşağıdaki özelliği sağlayan bir P matrisi yardımıyla olmaktadır.

Ω^{-1} matrisi pozitif tanımlı bir matris olduğunda $n \times n$ boyutlu singüler olmayan ve $\Omega^{-1} = P'P$ özelliğini sağlayacak bir P matrisi mevcuttur. P matrisi Ω^{-1} matrisinin 'karekök matrisi' olarak ifade edilebilir.

Dönüşüm için P matrisi kullanarak (3.4) modeli aşağıdaki gibi yeniden yazılabilir.

$$\begin{aligned} PY &= PX\beta + P\varepsilon \\ Y^* &= X^*\beta + \varepsilon^* \end{aligned} \tag{3.8}$$

Burada $Y^* = PY$, $X^* = PX$ ve $\varepsilon^* = P\varepsilon$ olup (3.8) modeli ”dönüştürülmüş model” olarak adlandırılır. Görüldüğü gibi (3.4) orijinal modeli ile (3.8) dönüştürülmüş model aynı bilinmeyen regresyon katsayılarına (β) sahiptir.

Dönüştürülmüş modelin hata terimi (ε^*) ile ilgili özelliklere bakılacak olursa;

- $E(\varepsilon^*) = E(P\varepsilon) = PE(\varepsilon) = 0$
- $Var(\varepsilon^*) = E(\varepsilon^* \varepsilon^{*'})$

$$
\begin{aligned}
&= E(P\varepsilon\varepsilon' P) \\
&= PE(\varepsilon\varepsilon')P' \\
&= \sigma^2 P\Omega P' \\
&= \sigma^2 P(\Omega^{-1})^{-1} P' \\
&= \sigma^2 P(P'P)^{-1} P' && (P'P = \Omega^{-1}) \\
&= \sigma^2 PP'P'^{-1} P' && ((AB)^{-1} = B^{-1}A^{-1}) \\
&= \sigma^2 I_n
\end{aligned}
$$

olarak bulunur ki görüldüğü gibi dönüştürülmüş modelde değişen varyanslılık ve otokorelasyon sorunu ortadan kalkmıştır. Ayrıca $rank(X^*) = rank(PX) = k$ ’dır, çünkü k ranklı bir matris singüler olmayan bir matris ile çarpıldığında rankı değişmez. Bununla beraber ε^* da, ε ’nun lineer bir dönüşümü olarak elde edildiğinden, normal dağılım göstermektedir. Bunların sonucu olarak söylenebilir ki (3.4) genelleştirilmiş model klasik varsayımları sağlamamakla beraber (3.8) dönüştürülmüş model klasik varsayımları sağlamaktadır ve (3.8) modeline EKK yönteminin uygulanması sonucunda en iyi doğrusal yansız (BLUE) tahmin ediciler elde edilecektir.

Dönüştürülmüş model için EKK tahmin edicisi:

$$
\begin{aligned}
\hat{\beta}_{GEKK} &= (X^{*'} X^*)^{-1} X^{*'} Y^* \\
&= (X'P'PX)^{-1} X'P'PY \\
&= (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y \qquad (3.9)
\end{aligned}
$$

= Genelleştirilmiş En Küçük Kareler Tahmin Edicisi

=Aitken Tahmin Edicisi

şeklinde dönüştürülmüş değişkenler kullanılarak elde edilir. $\hat{\beta}_{GEKK}$, tüm lineer ve yansız tahmin ediciler arasında en küçük varyansa sahiptir. Dönüştürülmüş modelin EKK tahmin edicisi modelin GEKK tahmin edicisidir. Ayrıca,

$$\begin{aligned}
E(\hat{\beta}_{GEKK}) &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y] \\
&= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon] \\
&= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(\varepsilon) \\
&= \beta
\end{aligned}$$

olarak elde edilir ve $\hat{\beta}_{GEKK}$'in yansız olduğu gösterilebilir. Bununla beraber $\hat{\beta}_{GEKK}$'in varyans-kovaryans matrisi:

$$\begin{aligned}
Kov(\hat{\beta}_{GEKK}) &= E(\hat{\beta}_{GEKK} - \beta)(\hat{\beta}_{GEKK} - \beta)' \\
&= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}] \\
&= \sigma^2(X'\Omega^{-1}X)^{-1}
\end{aligned}$$

şeklinde bulunur.

Sonuç olarak, eğer Ω, hata terimi varyans-kovaryans matrisini biliyorsak, genelleştirilmiş en küçük kareler tahmin edicisi iki yolla bulunabilir.

1) $\Omega^{-1} = P'P$ özelliği kullanılarak P matrisi elde edilir. Böylece dönüştürülmüş değişkenler $X^* = PX$, $Y^* = PY$ bulunur ve bu değişkenler kullanılarak en küçük kareler tahmin edicisi $\hat{\beta}_{GEKK} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$ elde edilir.

2) Ω bilindiğinden Ω^{-1} bulunur ve genelleştirilmiş en küçük kareler tahmin edicisi $\hat{\beta}_{GEKK} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$ formülü ile direk elde edilir.

(3.9) eşitliğinde hata terimi varyans-kovaryans matrisinin, Ω, bilindiği varsayılır. Ancak uygulamada genellikle varyans-kovaryans matrisi bilinmez. Bu sebeple, Ω 'nın tutarlı bir tahmin edicisi olan S matrisi kullanılarak GEKK tahmin edicisi aşağıdaki şekilde elde edilir.

$$\hat{\beta}^*_{GEKK} = (X'S^{-1}X)^{-1}X'S^{-1}Y \tag{3.10}$$

3.2. Kantil Regresyon

Kantil regresyon, ortalama regresyondan daha kesin bir istatistiksel model sunar ve günümüzde yaygın uygulamaları sahiptir. Ücretlerdeki ve gelirdeki eşitsizlik gibi dağılımın bozulduğu konuların incelemesinde ekonomide, tıbbi referans çizelgeleri, yaşam analizi, finansal ekonomi, çevre modelleme, değişen varyansı algılama için yaygın olarak kullanılan Kantil Regresyon Modelleri ortalama fonksiyonları ve koşullu kantil fonksiyonları için tahmin yapılmasında kullanılır. Kantil Regresyon, Medyan Regresyon'un belirlenen kantiller için genelleştirilmiş halidir. Bu regresyon modelleri uç değerlere ve eğikliğe En Küçük Kareler Yöntemi'nden daha az hassastır (Koenter ve Bassett, 1978).

Kantil Regresyon Modeli aslında bir yerleşim modelidir. Basit yerleştirme modeli,

$$Y_t = \beta + e_t \tag{3.11}$$

olarak ifade edildiğinde, burada yer alan Y_t simetrik F dağılım fonksiyonuna sahip, bağımsız, özdeş dağılımlı, β medyanlı tesadüfi değişkendir. Bu modelde θ. örnek kantili

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i;y_i \geq \beta} \theta |y_i - x_i\beta| + \sum_{i;y_i < \beta} (1-\theta)|y_i - x_i\beta| \right\} \tag{3.12}$$

ifadesinin minimizasyonu ile elde edilir. Bunu doğrusal regresyon modeli,

$$Y_t = x_i\beta + e_t \tag{3.13}$$

için genelleştirilir. Burada x_i, bağımsız değişken vektörüdür. e_i bağımsız, sıfır etrafında simetrik ve F dağılımına sahiptir. Bu durumda θ. kantil regresyon ($0<\theta<1$),

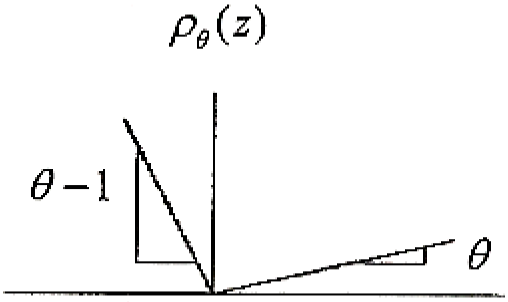
$$\min_{\beta}\frac{1}{n}\left\{\sum_{i;y_i\geq\beta}\theta|y_i-x_i\beta|+\sum_{i;y_i<\beta}(1-\theta)|y_i-x_i\beta|\right\}$$

$$\min_{\beta}\frac{1}{n}\sum_{i=1}^{n}\rho_{\theta}(y_i-x_i\beta) \qquad (3.14)$$

minimizasyonu ile tahmin edilir ve y'nin $\theta.'c\imath$ kantili olarak da adlandırılabilir. Kantil regresyonun bu şekildeki gösterimi doğrusal programlama gösterimidir. Burada I, karakteristik fonksiyondur, ρ_θ ise Kontrol (Check) Fonksiyondur ve

$$\rho_{\theta}(z)=z(\theta-I(z\leq 0)) \qquad (3.15)$$

olarak tanımlanır. Şekil 3.1.deki gibi gösterilebilir.



Şekil 3.1. ρ_θ Fonksiyonu

$\theta=0{,}5$ olması durumunda kantil regresyon amaç fonksiyonu LAD amaç fonksiyonuna ve l_l regresyona eşittir. Kantil regresyon amaç fonksiyonu mutlak sapmaların ağırlıklandırılmış toplamıdır.

$\theta.'c\imath$ kantil regresyon gözlem değerlerinin işaretlerine dayalı olarak

$$\theta\min_{\beta}\frac{1}{n}\sum_{i=1}^{n}\left(\theta-1/2+1/2\,\mathrm{sgn}(y_i-x_i'\beta)\right)(y_i-x_i'\beta) \qquad (3.16)$$

şeklinde tahmin edilir. Burada

$$\mathrm{sgn}(a)=\begin{cases}1, & a>0\\ -1, & a\le 0\end{cases} \qquad (3.17)$$

(3.17) ifadesi ile gösterilir. Tahminlerin bu şekilde, yani gözlem değerlerinin büyüklüğü yerine gözlem değerlerinin işaretlerine dayalı olması, Kantil Regresyon'un robust bir yöntem olmasını sağlamaktadır. Minimizasyon birinci mertebe koşulunun sağlanması gerekir. Birinci Mertebe Koşulu'nun $K\times 1$ vektörü,

$$\min_{\beta}\frac{1}{n}\sum_{i=1}^{n}\left(\theta-1/2+1/2\,\mathrm{sgn}(y_i-x_i\beta)\right)(y_i-x_i\beta)x_i=0 \qquad (3.18)$$

olarak gösterilir. Bu ifade, Birinci Mertebe Koşulu Genelleştirilmiş Momentler Yöntemi (GMM)'ne uyan moment fonksiyonudur. Moment fonksiyonu

$$\Psi(x_i,y_i,\beta)=\left(\theta-1/2+1/2\,\mathrm{sgn}(y_i-x_i\beta)\right)x_i \qquad (3.19)$$

olarak tanımlanır. $\psi(.)'nın$ moment fonksiyon olarak geçerli olabilmesi için belirli düzgünlük şartları altında,

$$E\left[\Psi(x_i,y_i,\beta)\right]=0 \qquad (3.20)$$

olması gerekir. Genelleştirilmiş Momentler Yöntemi kullanılarak elde edilen parametre tahmin edicileri tutarlı ve asimptotik olarak normal olacaktır.

Belirli düzgünlük şartları altında,

$$\sqrt{n}(\hat{\beta}-\beta)\xrightarrow[n\to\infty]{}N(0,\Lambda_\theta) \qquad (3.21)$$

olarak gösterilebilir. Burada,

$$\Lambda_\theta = \theta\left(1-\theta\right) E\left[f_{u_\theta}\left(0/x_i\right) x_i x_i'\right]^{-1} E\left(x_i x_i'\right) E\left[f_{u_\theta}\left(0/x_i\right) x_i x_i'\right]^{-1} \tag{3.22}$$

olarak tanımlanır. Olasılık değeri “1” olduğunda ve $f_{u_\theta}\left(0/x_i\right) = f_{u_\theta}\left(0\right)$ ise, yani hata teriminin yoğunluğu sıfır etrafındaysa ve *x*’ten bağımsızsa Λ_θ,

$$\Lambda_\theta = \frac{\theta(1-\theta)}{f_{u_\theta}^2(\theta)} E\left(x_i x_i'\right)^{-1} \tag{3.23}$$

şeklinde sadeleştirilebilir. $f_{u_\theta}^2(./x)$ *x*’ten bağımsız olduğunda, tüm kantillerin parametre vektörleri sadece kesim noktalarında farklılık gösterir.

Kantil katsayılarını yorumlayabilmek için, *y*’ nin k açıklayıcı değişkenine göre koşullu kantilinin kısmı türevi alınmaktadır. Türev alındığında,

$$\frac{\partial \Lambda_\theta \left(y_i / x_i\right)}{\partial x_i} \tag{3.24}$$

olacaktır. Bu türev, *x*’in k.’ci değerindeki marjinal değişime göre, $\theta.'c\imath$ şartlı kantildeki marjinal değişimi vermektedir.

3.2.1. Kantil regresyonun özellikleri

1) EKK ve Medyan Regresyon y’nin koşullu dağılımının ortası hakkında bilgi vermekte, Kantil Regresyon ise farklı kantil değerleri için y’nin x’e göre koşullu dağılımının tümü hakkında bilgi vermektedir.

2) Kantil Regresyon’da; $\min_b \frac{1}{n}\sum_{i=1}^{n} \rho_\theta\left(y_i - x_i'b\right)$ ’nin minimizasyonu, doğrusal programlama (LP) gösterimidir, bu durum tahmini kolaylaştırır.

3) Kantiller monoton dönüşümlere olanak verirler. Herhangi h(.) monoton fonksiyonu için $Q_{h(y/x)}\left(\tau/x\right) = h\left(Q_{h(y/x)}\left(\tau/x\right)\right)$ olur.

4) Kantiller *y*’deki aşırı değerlere karşı sağlamdırlar (robust).

5) Hata terimi normal dağılmadığında, kantil regresyon tahmincileri EKK Tahmincilerinden çok daha etkin olabilir.

6) Kantil Regresyon değişen varyansın belirlenmesine imkan verir.

7) Kantil Regresyon amaç fonksiyonu için tahmin edilen katsayı vektörü bağımlı değişkendeki aşırı değerlere duyarlı değildir ve yerleşimin robust bir ölçüsüdür.

8) Farklı kantillerde farklı sonuçlar çıkması, bağımlı değişkenin koşullu dağılımının farklı noktalarındaki açıklayıcı değişkenlerdeki değişikliklere farklı tepki vermesi olarak yorumlanabilir.

9) L-tahmin edicileri, kantil tahmincilerin doğrusal kombinasyonuna dayanır. İlk bölümde yer alan L-tahmincileri, $Y_1, Y_2, \ldots, Y_T$. sıra istatistiklerinin doğrusal kombinasyonları olan tahmincilerdi. Bu tahminciler doğrusal model için genelleştirilmiş, medyan regresyon da kantiller için geliştirilmiştir.

3.2.2. Kantil regresyonun doğrusal programlama gösterimi

Kantil Regresyon'un doğrusal programlama gösterimi sonlu sayıdaki simpleks iterasyonlarla tahmin belirleneceğini ifade eder. İterasyon sayısı doğrusal programlama algoritmasına göre küçüktür. EKK Regresyonu'ndan farklı olarak parametre vektör tahmini aşırı değerlere karşı robust' tır. Daha öncede söz ettiğimiz gibi y_i *'nin* θ.*'ncı* kantili

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i; y_i \geq \beta} \theta |y_i - x_i\beta| + \sum_{i; y_i < \beta} (1-\theta) |y_i - x_i\beta| \right\} = \min \frac{1}{n} \sum_{i=1}^{n} \rho_\theta (\theta_i) \tag{3.25}$$

ifadesinin minimizasyonu ile elde edilmekteydi. Bunu doğrusal programlama gösterimi olarak ifade edebilmek için y_i sadece pozitif elamanların bir fonksiyon olarak,

$$y_i = \sum_{j=1}^{K} x_{ij}\beta_{\theta j} + \sum_{j=1}^{K} x_{ij} \quad \beta^t_{\theta j} - \beta^2_{\theta j} \quad + \quad \varepsilon_{\theta_i} - v_{\theta_i}$$
$$\beta^t_{\theta j} \geq 0, \quad \beta^2_{\theta j} \geq 0 \quad (j = 1,...,K) \quad ve\, \varepsilon_{\theta_i} \geq 0 \quad v_{\theta_i} \geq 0\,(i = 1,...,n) \tag{3.26}$$

şeklinde yazılabilir. İlk denklem matris gösterimiyle yazıldığında doğrusal programlamanın primal problemine dönüşür ve

$$\min c'z$$
$$Az = y \tag{3.27}$$

olarak ifade edilir. Burada $A = \; X, -X, I_n, -I_n \;$, $y = \; y_1,..., y_n \;'$, $z = \; \beta^{1'}, \beta^{1'}, u', v' \;'$, $c = \; 0', 0', \theta, I', (1-\theta)I' \;'$, $X = (x_1,...,x_n)'$, I_n; n boyutlu birim matristir. $0'$; sıfırların $K \times 1$ vektörüdür, l; ise 1'lerin $n \times 1$ vektörüdür. Doğrusal programlamanın dual problemi yaklaşık olarak daha önce,

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \; \theta - 1/2 + 1/2 \operatorname{sgn} \; y_i - x_i\beta \quad y_i - x_i'\hat{\beta}_0 \;) x_i = 0 \tag{3.28}$$

Gösteriminde belirlenen önceki mertebe koşuluyla aynıdır ve

$$\max_{w} w'y$$
$$w'A \leq c' \tag{3.29}$$

olarak gösterilir. Dual problem; X matrisinin tüm sütunlarının sıralı olması durumunda hem primal hem de dual problemlerin çözümünün mümkün olduğunu ileri sürmektedir. Doğrusal programlamanın denge teoremi ancak bu durumda çözümün optimal olduğunu ifade eder.

Doğrusal programlama problemini çözmek için birçok algoritma öne sürülmüştür. Bunlardan en çok ilgi göreni $\theta = 1/2$ olan medyan regresyon için geliştirilmiştir. Bu algoritma küçük değişiklerle herhangi bir kantil regresyonuna da uyarlanabilir. Söz konusu algoritmanın en büyük avantajı, bilinen diğer algoritmalara göre anlamlı bir şekilde simpleks dönüşümlerin sayısını azaltmasıdır.

$$\hat{\beta}\left(\theta, y+X\gamma, X\right)=\hat{\beta}\left(\theta, y, X\right)+\gamma \quad , \gamma \in R^K \tag{3.30}$$

eşvaryans özelliği ile doğrusal programlama algoritması hesaplanır. $\hat{\beta}_\theta$ '$nın$ ilk değeri $\hat{\beta}_\theta^0$ biliniyorsa, gözlemleri düzlemin sağ tarafına yerleştirerek hesaplama süresini kısaltmak için kullanılabilir.

$y_R = y - X\hat{\beta}_\theta^0$ olsun

$\hat{\beta}_\theta^R$, y_R 'nin x'e göre kantil regresyon tahminidir.

$\hat{\beta}\left(\theta, y+X\gamma, X\right)=\hat{\beta}\left(\theta, y, X\right)+\gamma \quad , \gamma \in R^K$ eş varyans özelliğini kullanarak,

$\hat{\beta}_\theta = \hat{\beta}_\theta^R + \hat{\beta}_\theta^0$ bulunur. $\hat{\beta}_\theta^R$ ve $\hat{\beta}_\theta^0$ 'ı belirlemek, doğrudan $\hat{\beta}_\theta$ 'yı belirlemekten çok daha hızlıdır.

Mümkün ilk değer, sabitinin, $\hat{e}_1, \ldots, \hat{e}_n$ artıklarının $n\theta$ 'ıncı sıra istatistiğiyle değiştirildiği, düzeltilmiş EKK tahminidir. Alternatif ilk değer, gözlem değerlerinin çok büyük olması durumunda gözlem değerlerinin bir kısmına uygulanan kantil regresyonla belirlenebilir.

3.2.3. Aralık regresyonda kantil regresyon teknikleri kullanarak alt ve üst yaklaşım modelleri

Kantil Regresyon Yönteminde bağımlı değişken değerlerine Y_i ilişkin nokta tahmininin yanı sıra aralık tahmini de yapılabilmektedir.

$$\begin{aligned}
&\underset{z_j}{M}\text{ax}\sum_{i=1}^{m} z_i y_i \\
&\quad \sum_{i=1}^{m} z_j x_{ij} = 0, \qquad i = 1,\ldots,n, \\
&\;-(1-p) \le z_j \le p, \qquad j = 1,\ldots,m
\end{aligned} \tag{3.31}$$

Eşitlik (3.31)'de; z_j : j. dual değişkeni göstermektedir. Dual değişken z_j 'nin önemli özelliği örnek gözlemi ile p. Kantil hiper düzlemi arasındaki lokal ilişkiyi göstermesidir. Bu özellik aşağıdaki gibi tanımlanır.

- $z_j = p$:Regresyon hiper düzleminin üzerinde kalan j. gözlem
- $-(1-p) < z_j < p$: Regresyon hiper düzlemi üstündeki j. gözlem
- $z_j = -(1-p)$: Regresyon hiper düzlemi altında kalan j. gözlem

Artan sırada z_j dual değerleri sıralamak ile gözlemleri mp ve $m(1-p)$ olacak şekilde iki kategoriye ayırarak sınıflandırabiliriz. Bu sınıflama aşağıdaki gibi yapılır.

$$-(1-p) = \underbrace{z_1 \le z_2 \le \ldots z_{k-1} \le z_k}_{m_p} \le \underbrace{z_{k+1} \ldots \le z_{m-1} \le z_m}_{m_p} = p \qquad (3.32)$$

m örneklem hacmine sahip veri setini temel olarak % p_1 ve % p_2 $(0 < p_2 < p_1 < 1)$ şeklinde göz önünde bulundurursak. p değerini p_1 ve p_2 olarak düzenledikten sonra veri setini üç sınıfa ayırabiliriz. Örneğin problem (3.31)'i p_1 ve p_2 ile birlikte çözdüğümüzü ve iki dual değerleri z_{j1} ve z_{j2} (j=1,…,m) olarak bulduğumuzu varsayalım. Elde edilen z_{j1} ve z_{j2} (j=1,…,m) dual değerlerinin her biri ayrı ayrı artan sırada dizildikten sonra veri setini üç sınıfa ayırabiliriz. Bu sınıflama aşağıdaki gibi yapılır.

- Sınıf 1 : $C_1 =$ z_{j1}'e göre üstteki $m(1-p_1)$ sınırları içindeki j. gözlem
- Sınıf 3 : $C_3 =$ z_{j2}'ye göre alttan mp_2 sınırları içindeki j. gözlem
- Sınıf 2 : $C_2 =$ C_1 ve C_2'nin dışında kalan j. gözlem

Sınıf 1 ve Sınıf 2'ye ait büyüklük $m(1-p_1)$ ve mp_2 küsuratlı çıkması durumunda tam sayıya yuvarlanır. Örneğin elimizde 28 gözlem değerinin olduğunu düşünelim $p_1 = 0{,}7$ ve $p_2 = 0{,}2$ olsun. Bu durumda Sınıf 1, Sınıf 3 ve Sınıf 2 büyüklükleri sırasıyla $8\,(m(1-p_1) = 8{,}4)$, $6\,(mp_2 = 5{,}6)$ ve 14 $(28-8-6=14)$ bulunur.

Alt yaklaşım modeli için doğrusal programlama gösterimi:

$$
\begin{aligned}
&\min_{b,d,\theta} \quad J = \sum_{j \in c_2} d^t \left| x_j \right| + k\theta^t \left| x_j \right| \\
&\qquad y_j \geq b^t x_j + \ d^t \left| x_j \right| - \theta^t \left| x_j \right| \ , \qquad j \in C_1 \\
&\qquad b^t x_j + d^t \left| x_j \right| \geq y_j \geq b^t x_j - d^t \left| x_j \right| , \qquad j \in C_2 \\
&\qquad y_j \leq b^t x_j - \ d^t \left| x_j \right| - \theta^t \left| x_j \right| \ , \qquad j \in C_3 \\
&\qquad d_i \geq 0, \quad \theta_i \geq 0, \quad i = 1, 2, \ldots, n.
\end{aligned}
\tag{3.33}
$$

$$Y_L^* \ x_j \ = \ b_0^*, d_0^* \ + \ b_1^*, d_1^* \ \ x_{j1} + \cdots + \ b_n^*, d_n^* \ \ x_{jn}. \tag{3.34}$$

Üst yaklaşım modeli için doğrusal programlama gösterimi:

$$
\begin{aligned}
&\min_{b,d,\theta} \quad J = \sum_{j=1}^{m} c^t \left| x_j \right| . \\
&\qquad a^t x_j + c^t \left| x_j \right| \geq y_i , \\
&\qquad a^t x_j - c^t \left| x_j \right| \leq y_i , \qquad j = 1, \ldots, m \\
&\qquad B_i \subseteq A_i , \qquad i = 0, 1, \ldots, n \\
&\qquad c_i \geq 0, \qquad i = 0, 1, \ldots, n.
\end{aligned}
\tag{3.35}
$$

$$Y_U^* \ x_j \ = \ a_0^*, c_0^* \ + \ a_1^*, c_1^* \ \ x_{j1} + \cdots + \ a_n^*, c_n^* \ \ x_{jn}. \tag{3.36}$$

3.3. En Küçük Mutlak Sapma (Lad) Regresyon

En küçük mutlak Sapmalar Regresyonu (LAD)'da artıkların karelerinin mutlak değerlerinin minimize edilerek parametre tahminleri yapılır.

EKK Regresyonunda artıkların kareleri toplamını $\sum \varepsilon_i^2$ minimize eden $\hat{\beta}_i$ değerleri tahmin ediliyordu, En Küçük Mutlak Sapmalar Regresyonu'nda ise artıkların mutlak değerlerinin toplamını $\sum \left| \hat{\varepsilon}_i \right|$ minimize eden $\hat{\beta}_i$ tahminleri elde edilir.

Kantil Regresyonunda θ değerinin 0,5 alınması durumunda kantil regresyonunun amaç fonksiyonu LAD Regresyonundaki amaç fonksiyonuna eşittir.

En küçük mutlak sapma (LAD) regresyonu için simpleks yöntem algoritması aşağıdaki gibidir (Arthanari ve Dodge, 1981).

1.adım:

$$W_{p+r}=\begin{cases} Y_r, & Y_r>0 \\ 0 & d.y.\ 1\le r\le n \end{cases}$$

$$W_{p+n+r}=\begin{cases} -Y_r & Y_r<0 \\ 0 & d.y. 1\le r\le n \end{cases}$$

Temel uygun çözümle başlansın ve

$$W_j=0 \qquad j=1,2,\ldots,p$$

olsun.

Tablo biçimindeki d_{ij} 'lere uygun çözüm sunalım. Bu çözüme uygun temel B, kesinlikle e_r ya da $-e_r$, tüm $r=1,2,\ldots,n$ için kolonlarından birini içerir ve B'nin tersi kendisidir yani $B^{-1}=B$ 'dir. Bu nedenle $\alpha_j=B^{-1}a_j$; eğer W_{p+r} temelde ise (+1) ile, eğer W_{p+n+r} temelde ise (-1) ile A'nın $r.$ satırının çarpılmasıyla elde edilir ve $B^{-1}Y=W_b$ 'ye karşılık gelen uygun temel uygun çözümü verir. Temeldeki vektörleri ve onlara ait c_j 'ler aşağıya indirilir.

c_j-z_j , her bir j için c_j 'den $\sum_{i=1}^{n}C_{B_i}\alpha_{ij}$ 'nin çıkarılmasıyla elde edilir. $c_j-z_j=0$, temeldeki tüm a_j 'ler için.

2.adım:

Aşağıdaki gibi temeldeki bir vektörle değiştirmek için temelde olmayan bir a_j seçilir,

a) $c_{j1}-z_{j1}=\max_k C_k-Z_k$

$c_k-z_k>0$

W_k, sınırlandırılmamış değişken

Çizelge 3.1. 2. Adımdaki matematiksel modelin simpleks tablosu

.

C_B	**Temeldeki Vektör**	W_B	α_1	α_2	$\cdots$	α_j	$\cdots$	α_{2n+p}
1	α_{p+1} ya da α_{p+n+1}	$\lvert Y_1 \rvert$	α_{11}	α_{12}	$\cdots$	α_{1j}	$\cdots$	$\alpha_{1,2n+p}$
.	.	.	.	.		.		.
.	.	.	.	.		.		.
.	.	.	.	.		.		.
1	α_{p+n} ya da α_{p+2n}	$\lvert Y_n \rvert$	α_{n1}	α_{n2}	$\cdots$	α_{nj}	$\cdots$	$\alpha_{n,2n+p}$
	$C_k - Z_k$	$Z = \sum_{l=1}^{n} \lvert Y_l \rvert$				$C_j - Z_j$		

b) $\left| c_{jz} - z_{jz} \right| = \max_k \left| c_k - z_k \right|$ olursa

$c_k - z_k < 0$

$c_{jz} - z_{jz} = \max\left[c_{j1} - z_{j1} \ . \ \left| c_{jz} - z_{jz} \right| \right]$ olarak j , seçilir.

Bu adım, temele girmek için vektör seçimi için kriter verir. Eğer hem j_1 hem de j_2 bulunamazsa 5. adıma gidilir. Aksi halde 3.adıma gidilir.

<u>3.adım:</u>

Eğer $c_j - z_j > 0$ ise r, aşağıdaki gibi seçilir.

$$\frac{W_B}{\alpha_{rj}} = \max_{i \in R_B} \left[\frac{W_B}{\alpha_{rj}}, \quad \alpha_{rj} < 0 \right]$$

Eğer $c_j - z_j < 0$ ise r, aşağıdaki gibi seçilir.

$$\frac{W_B}{\alpha_{rj}} = \min_{i \in R_B} \left[\frac{W_B}{\alpha_{rj}}, \quad \alpha_{rj} > 0 \right]$$

Bu adım temelden ayrılacak vektör seçimi için kriteri verir ve 4. adıma gidilir.

4.adım:

Yeni temel B’ye uygun olan yeni tablo aşağıdaki gibi oluşturulur.

$$\hat{W}_{Br} = \frac{W_{Br}}{\alpha_{rj}}$$

$$\hat{\alpha}_{rl} = \frac{\alpha_{rl}}{\alpha_{rj}}$$

ve

$$\hat{W}_{Bi} = W_{Bi} - \frac{W_{Br}}{\alpha_{rj}}\alpha_{ij}\,, \quad i \neq r \quad i = 1, 2, \dots, n$$

$$\hat{\alpha}_{il} = \alpha_{il} - \frac{\alpha_{ij}\alpha_{rl}}{\alpha_{rj}}, \quad i \neq r \quad i = 1, 2, \dots, n \quad l = 1, \dots, p + 2n$$

$$\hat{Z} = Z + \ C_j - Z_j \ . \frac{W_{Br}}{\alpha_{ri}}$$

$$\hat{C}_l - \hat{Z}_l = \ C_l - Z_l \ - \ C_l - Z_l \ \frac{\alpha_{rl}}{\alpha_{rj}}, \quad l = 1, \dots, p + 2n.$$

Bu dönüşüm; ring-around-rosy metodu olarak adlandırılır. 2. adıma dönülür ve $B = \hat{B}$ ile devam edilir.

5.adım:

Geçerli temel optimaldir ve bu adımda durulur. $\mathrm{W}_1, ..., \mathrm{W}_\mathrm{p}$; $\beta_0, ..., \beta_{\mathrm{p}-1}$ parametrelerini verir ve $W_{p+r} + W_{p+n+r}$ ise r. gözlemdeki mutlak hatayı verir. Eğer alttan sınırlı olmayan amaç fonksiyonunu çözümü için bu algoritmayı kullanmak istersek, sonsuz minimum değeri olan problem için algoritmaya son verilir. Böyle bir durumda Adım 3’te r’yi bulmak için belirtilen yöntem yetersizdir.

3.4. M-Regresyon

X, $n \times p$ boyutlu bağımsız değişkenler matrisi, n:gözlem sayısı, p:değişken sayısı ve y, $n \times 1$boyutlu bağımlı değişken vektörü olmak üzere X ve Y'nin i. satırı, i. gözleme ait y_i bağımlı değişkeninin ve x_i bağımsız değişkenlerinin değerlerini ifade eder.

$$X=\begin{bmatrix} x_{11} & x_{12} & . & . & x_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ y_{i1} & y_{i2} & . & . & y_{ip} \\ x_{n1} & x_{n2} & . & . & x_{np} \end{bmatrix} \quad y=\begin{bmatrix} y_1 \\ . \\ . \\ y_i \\ y_n \end{bmatrix}$$

Bu durumda regresyon modeli aşağıdaki şekildedir.

$$y_i = x_i \beta + \varepsilon_i \tag{3.37}$$

Burada β bilinmeyen regresyon parametrelerini ifade eden p boyutlu bir vektör, ε_i ise n boyutlu hata terimleri vektörüdür.

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & . & . & x_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{n1} & . & . & . & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} \\ \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} \\ . \\ . \\ . \\ \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

β'nın en küçük kareler tahmin edicisi hata terimlerinin karelerinin toplamını minimum yapan $\hat{\beta}_{EKK}$ değerleridir. Ya da başka bir ifadeyle aşağıdaki amaç fonksiyonunu en küçük yapan parametrelerdir;

$$\hat{\beta}_{EKK} = \min \sum_{i=1}^{n} {\varepsilon_i}^2 = \min \sum_{i=1}^{n} (y_i - x_i \beta)^2 \tag{3.38}$$

Eğer ε_i hata terimleri normal dağılıma sahip ise $\hat{\beta}_{EKK}$ tahmin edicileri optimum tahmin edicilerdir. Başka bir ifadeyle en küçük varyansa sahiptirler ve yansızdırlar.

Diğer taraftan hata terimleri diğer dağılımlardan da gelebilmektedir (uç değerli dağılımlar gibi ya da tamamen dağılımdan bağımsız veya birçok aykırı gözlem içerebilir). Bu durumda en küçük kareler yöntemi optimumluğunu kaybeder ve diğer tip tahmin ediciler daha iyi performans gösterir. En çok bilinen sağlam tahmin ediciler M tahmin edicilerdir.

(3.38) numaralı eşitlikte hataların karesi yerine hataların ρ kayıp fonksiyonu (ağırlık fonksiyonu) kullanılarak aşağıdaki amaç fonksiyonunu minimum yapacak M tahmin edicileri belirlenir.

$$\hat{\beta}_M = \min \sum_{i=1}^{n} \rho(y_i - x_i\beta) \tag{3.39}$$

ρ kayıp fonksiyonu, simetrik ve azalmayan olmalıdır. ρ fonksiyonunun $\rho(u) = u^2$ olması en küçük kareler tahmin edicisine karşılık gelmektedir. Büyük hatalara daha az önem vermek isteyen biri sınırlandırılmış kayıp fonksiyonu ρ kullanabilir. $r_i = y_i - x_i\beta$ (3.39) numaralı eşitliğin amaç fonksiyonundaki hata terimini ifade etsin. i. gözleme karşılık gelen ağırlığı aşağıdaki gibi ifade edebiliriz (Cummins ve Andrews, 1995).

$$\rho(r_i) = w_i^r \tag{3.40}$$

Bu durumda (3.39) numaralı eşitliği aşağıdaki gibi tekrar yazabiliriz.

$$\hat{\beta}_M = \min \sum_{i=1}^{n} w_i^r (y_i - x_i\beta)^2 \tag{3.41}$$

Yukarıdaki tanımlamada M tahmin edicisi bir ağırlıklandırılmış en küçük kareler tahmin edicisi gibi ifade edilmiştir (Dempster ve ark., 1980).

Bu konuyla ilgili daha ayrıntılı bilgiler içeren kitaplar mevcuttur (Huber, 1981; Rousseeuw ve Zomeren, 1990).

3.4.1. Gözlemlerin ağırlıklandırma yöntemi

MR yönteminde ağırlıklandırma işlemi ya da başka bir ifadeyle gözlemlerin koordinatlarının değiştirilmesi işlemi regresyonda yapılan hata terimlerine ve hata kareler ortalamasına bağlıdır. Kullanılan ağırlık fonksiyonu yöntemin önerildiği kaynakta aşağıdaki gibi tanımlanmıştır.

$$w_i^r = f\left(\frac{r_i}{\hat{\sigma}}, c\right)$$

burada

$f(z,c) = \dfrac{1}{\left(1+\left|\dfrac{z}{c}\right|\right)^2}$ ve modelin varyansı hata terimlerinin kareler ortalaması

$\hat{\sigma}^2 = \dfrac{1}{n-p}\sum_{i=1}^{n} r_i^2$ dir. Dolayısıyla, her bir gözlemin dikey ağırlığı o gözlemin hata teriminin ve tüm hataların bir fonksiyonu aşağıdaki gibi

$w_i^r = \dfrac{1}{\left(1+\left|\dfrac{r_i}{4\hat{\sigma}}\right|\right)^2}$ olur (Cummins ve Andrews, 1995).

3.4.2. Gözlemlerin ağırlıklandırma şekli

Model ve gözlemler arasında uzaklık sadece hata terimi ile ifade edilmez. M-regresyon modelinin adımsal tahmini:

$$\hat{\beta}_M = \min \sum_{i=1}^{n} w_i^r (y_i - x_i\beta)^2$$

Amaç fonksiyonunu optimum yapacak parametre değerleri m-regresyon modeli tahmin edicileridir. Aynı fonksiyon aşağıdaki gibi yazılabilir.

$$\hat{\beta}_M = \min \sum_{i=1}^{n} \left(\sqrt{w_i^r}\, y_i - \sqrt{w_i^r}\, x_i \beta \right)^2 \tag{3.42}$$

Adım 1: Regresyon modeli EKK yönteminde parametrelerin tahminiyle belirlenir ve hata kareler ortalaması yardımıyla $\hat{\sigma}$ hesaplanır.

Adım 2: Buna bağlı olarak, her gözlem için hatalara bağlı olarak ağırlıklar hesaplanır: $w_i^r = 1/(1+|r_i / 4\hat{\sigma}|)^2$ ve bu ağırlıkların karekökü ile her y ve x ölçümü çarpılarak yeni y ve x ölçümleri belirlenir.

Adım 3: Değiştirilen y ve x ölçümleri ile tekrar EKK yapıldığında (3.42) eşitliğindeki $\hat{\beta}_M$, m-regresyon parametrelerinin tahmini belirlenir.

Adım 4: Bir önceki parametre tahminleri ile takip eden tahminler arasındaki fark 0,01'den küçük ise parametre tahmini belirlenmiş olur ve model parametrelerinin tahmin işlemi sona erer. Aksi halde Adım 2'ye dönülür.

3.4.3. Tek değişkenli m-regresyon uygulaması

M-regresyon yöntemini ilk olarak 6 gözlemli yapay verilerden oluşan tek değişkenli doğrusal model üzerinde gösterelim.

Çizelge 3.2. 6 gözlemli veri seti

GÖZLEM	Y	X
1	4	2
2	2	5
3	6	8
4	8	10
5	10	10
6	1	12

Adım 1: EKK yöntemiyle parametreler tahmin edilir.

Regresyon modelinin tahmini: $Y = 3{,}441 + 0{,}220X$ ve $\hat{\sigma}^2 =$ Hata kareler ortalaması $= 14{,}373$ ise $\hat{\sigma} = 3{,}791$ olur.

Adım 2: Hata terimlerine ($r_i = y_i - \hat{y}_i$) ve varyansa bağlı olarak ağırlıklar belirlenir.

Çizelge 3.3. 6 gözlemin ilk ağırlıklandırma işlemi sonrası koordinatları

X	**Y**	$r_i = y_i - \hat{y}_i$	$Z = \frac{r_i}{\hat{\sigma}}$	$w_i^r = \frac{1}{\left(1+\left\|\frac{r_i}{4\hat{\sigma}}\right\|\right)^2}$	$\sqrt{w_i^r}$	$X' = \sqrt{w_i^r} \times X$	$Y' = \sqrt{w_i^r} \times Y$
2	4	0,119	0,031	0,985	0,992	1,984	3,968
5	2	-2,542	-0,671	0,733	0,856	4,282	1,713
8	6	0,797	0,210	0,903	0,950	7,601	5,701
10	8	2,356	0,621	0,749	0,866	8,655	6,924
10	10	4,356	1,149	0,603	0,777	7,768	7,768
12	1	-5,085	-1,341	0,561	0,749	8,987	0,749

Değiştirilen her Y ve X ölçümleriyle tekrar EKK yapıldığında;
Regresyon modelinin tahmini:

$Y' = 2{,}798 + 0{,}256X'$ şeklinde ve

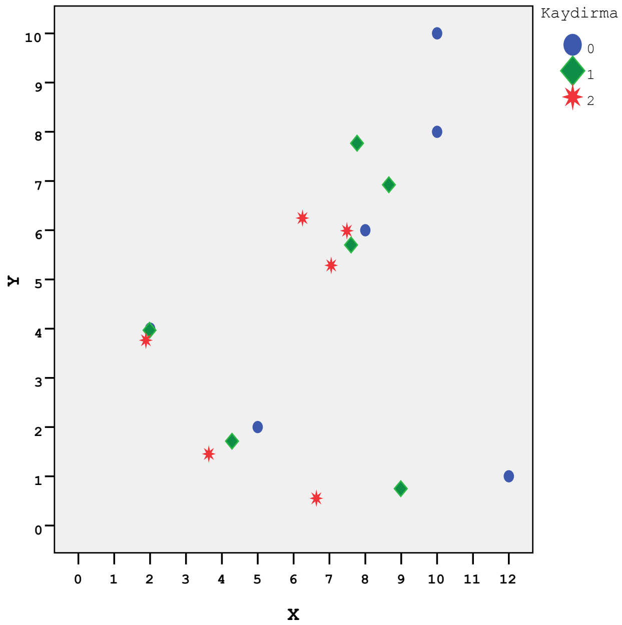
$\hat{\sigma}^2 =$ Hata kareler ortalaması $= 9{,}393$ ise

$\hat{\sigma} = 3{,}064$ olur.

Çizelge 3.4. 6 gözlemin ikinci ağırlıklandırma işlemi sonrası koordinatları

X'	Y'	$r_i = y_i - \hat{y}_i$	$Z = \frac{r_i}{\hat{\sigma}}$	$w_i^r = \frac{1}{\left(1+\left\|\frac{r_i}{4\hat{\sigma}}\right\|\right)^2}$	$\sqrt{w_i^r}$	$X'' = \sqrt{w_i^r} \times X'$	$Y'' = \sqrt{w_i^r} \times Y'$
1,984	3,968	0,664	0,217	0,900	0,949	1,882	3,765
4,282	1,713	-2,179	-0,711	0,721	0,849	3,636	1,454
7,601	5,701	0,960	0,313	0,860	0,927	7,048	5,286
8,655	6,924	1,914	0,625	0,748	0,865	7,486	5,989
7,768	7,768	2,985	0,974	0,647	0,804	6,247	6,247
8,987	0,749	-4,345	-1,418	0,545	0,738	6,634	0,553

Şekil 3.2'den de görüldüğü gibi M-regresyonda gözlemlerin adımsal ağırlıklandırılmaları sonucu veriler orijine doğru taşınmaktadır.



Şekil 3.2. İki adım ağırlıklandırma sonucunda gözlem noktalarının yer değişimi

3.4.4. Çoklu m-regresyon uygulaması

M-regresyon yöntemini ikinci olarak 6 gözlemli yapay verilerden oluşan çoklu (iki değişkenli) doğrusal model üzerinde gösterelim.

Çizelge 3.5. 6 gözlemli iki bağımsız değişkenli veri seti

GÖZLEM	Y	X_1	X_2
1	2	1	1
2	0	1	2
3	3	2	2
4	0	2	3
5	4	3	3
6	8	3	4

Adım 1: EKK yöntemiyle parametreler tahmin edilir.

Regresyon modelinin tahmini;

$Y = -2{,}000 + 2{,}833X_1 - 0{,}333X_2$ olur ve buradan

$\hat{\sigma}^2 =$ Hata kareler ortalaması $= 6{,}556$ ve

$\hat{\sigma} = 2{,}560$ olur.

Adım 2: Hata terimlerine ($r_i = y_i - \hat{y}_i$) ve varyansa bağlı olarak ağırlıklar belirlenir ve gözlem değerleri ağırlıklandırılır.

Çizelge 3.6. Gözlemlere yapılan ilk ağırlıklandırma işlemi sonrası koordinatları

Y	X_1	X_2	$r_i = y_i - \hat{y}_i$	$Z = \frac{r_i}{\hat{\sigma}}$	$w_i^r = \frac{1}{\left(1+\left\lvert\frac{r_i}{4\hat{\sigma}}\right\rvert\right)^2}$	$\sqrt{w_i^r}$	$Y' = \sqrt{w_i^r} \times Y$	$X_1' = \sqrt{w_i^r} \times X_1$	$X_2' = \sqrt{w_i^r} \times X_2$
2	1	1	1,500	0,586	0,762	0,872	1,744	0,872	0,872
0	1	2	-0,167	-0,065	0,968	0,984	0,000	0,984	1,968
3	2	2	0,000	0,000	1,000	1,000	3,000	2,000	2,000
0	2	3	-2,667	-1,042	0,629	0,793	0,000	1,587	2,380
4	3	3	-1,500	-0,586	0,761	0,872	3,489	2,617	2,617
8	3	4	2,833	1,107	0,614	0,783	6,266	2,350	3,133

Değiştirilen her Y ve X ölçümleriyle tekrar EKK yapıldığında;
Regresyon modelinin tahmini;

$Y' = -1{,}570 + 2{,}726 X_1' - 0{,}346 X_2'$ şeklinde olur.

$\hat{\sigma}^2 =$ Hata kareler ortalaması $= 4{,}393$ ise

$\hat{\sigma} = 2{,}096$'dir.

Çizelge 3.7. Ağırlıklandırılan gözlem değerlerinin ikinci kez ağırlıklandırma işlemi sonrası koordinatları

Y'	X_1'	X_2'	$r_i = y_i - \hat{y}_i$	$Z = \frac{r_i}{\hat{\sigma}}$	$w_i^r = \frac{1}{\left(1+\left\lvert\frac{r_i}{4\hat{\sigma}}\right\rvert\right)^2}$	$\sqrt{w_i^r}$	$Y'' = \sqrt{w_i^r} \times Y'$	$X_1'' = \sqrt{w_i^r} \times X_1'$	$X_2'' = \sqrt{w_i^r} \times X_2'$
1,744	0,872	0,872	1,236	0,590	0,760	0,872	1,520	0,760	0,760
0,000	0,984	1,968	-0,435	-0,207	0,904	0,951	0,000	0,935	1,871
3,000	2,000	2,000	-0,196	-0,093	0,955	0,977	2,932	1,954	1,954
0,000	1,587	2,380	-1,937	-0,924	0,660	0,812	0,000	1,289	1,934
3,489	2,617	2,617	-1,176	-0,561	0,769	0,877	3,060	2,295	2,295
6,266	2,350	3,133	2,508	1,196	0,593	0,770	4,824	1,809	2,412

3.5. Kısmi Sağlam M-Regresyon

X, $n\times p$ boyutlu bağımsız değişkenler matrisi, n:gözlem sayısı, p:değişken sayısı ve y, $n\times 1$boyutlu bağımlı değişken vektörü olmak üzere X ve Y'nin i. satırı, i. gözleme ait y_i bağımlı değişkeninin ve x_i bağımsız değişkenlerinin değerlerini ifade eder.

$$X=\begin{bmatrix} x_{11} & x_{12} & . & . & x_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ y_{i1} & y_{i2} & . & . & y_{ip} \\ x_{n1} & x_{n2} & . & . & x_{np} \end{bmatrix} \quad y=\begin{bmatrix} y_1 \\ . \\ . \\ y_i \\ y_n \end{bmatrix}$$

Bu durumda aşağıdaki regresyon modelini ele alırsak ;

$$y_i = x_i\beta + \varepsilon_i \tag{3.43}$$

Burada β bilinmeyen regresyon parametrelerini ifade eden p boyutlu bir vektör, ε_i ise n boyutlu hata terimleri vektörüdür.

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & . & . & x_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{n1} & . & . & . & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} \\ \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} \\ . \\ . \\ . \\ \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

β'nın en küçük kareler tahmin edicisi hata terimlerinin karelerinin toplamını minimum yapan $\hat{\beta}_{EKK}$ değerleridir. Ya da başka bir ifadeyle aşağıdaki amaç fonksiyonunu en küçük yapan parametrelerdir;

$$\hat{\beta}_{EKK} = \min \sum_{i=1}^{n} \varepsilon_i^{\ 2} = \min \sum_{i=1}^{n} (y_i - x_i \beta)^2 \qquad (3.44)$$

Eğer ε_i hata terimleri normal dağılıma sahip ise $\hat{\beta}_{EKK}$ tahmin edicileri optimum tahmin edicilerdir. Başka bir ifadeyle en küçük varyansa sahiptirler ve yansızdırlar. Diğer taraftan hata terimleri diğer dağılımlardan da gelebilmektedir (uç değerli dağılımlar). Bu durumda en küçük kareler yöntemi optimumluğu kaybeder ve diğer tip tahmin ediciler daha iyi performans gösterir. En çok bilinen sağlam tahmin ediciler M tahmin edicilerdir.

(3.44) numaralı eşitlikte hataların karesi yerine hataların ρ kayıp fonksiyonu kullanılarak aşağıdaki amaç fonksiyonunu minimum yapacak M tahmin edicileri belirlenir.

$$\hat{\beta}_M = \min \sum_{i=1}^{n} \rho(y_i - x_i \beta) \qquad (3.45)$$

ρ kayıp fonksiyonu, simetrik ve azalmayan olmalıdır. ρ fonksiyonunun $\rho(u) = u^2$ olması en küçük kareler tahmin edicisine karşılık gelmektedir. Büyük hatalara daha az önem vermek isteyen biri sınırlandırılmış kayıp fonksiyonu ρ kullanabilir. $r_i = y_i - x_i\beta$ (3.45) numaralı eşitliğin amaç fonksiyonundaki hata terimini ifade etsin. KSMR'de i. gözleme karşılık gelen ağırlık aşağıdaki gibi ifade edilir (Serneels ve ark., 2005).

$$w_i = w_i^r w_i^x$$

Burada, w_i ağırlığı hem dikey (w_i^r) ve hem de yatay uzaklığa (w_i^x) bağlı olarak belirlenmektedir. Bir noktanın genel veri topluluğuna ne kadar uzak olduğunu daha iyi yansıtmaktadır.

3.5.1. Gözlemlerin ağırlıklandırma yöntemi

Bu durum için ağırlık dikey (regresyon) sapmalarından (w_i^r) ve yatay gözlem sapmalarından hesaplanan ağırlıkların (w_i^x) çarpımıyla şöyle tanımlanmıştır;

$$w_i = w_i^r w_i^x$$

burada dikey uzaklığa bağlı ağırlıklandırma

$$w_i^r = f\left(\frac{r_i}{\hat{\sigma}}, c\right) = \frac{1}{\left(1+\left|\frac{r_i}{4\hat{\sigma}}\right|\right)^2} \tag{3.46}$$

ve yatay uzaklığa bağlı ağırlıklandırma

$$w_i^x = f\left(\frac{\|x_i - med_{L1}(X)\|}{median_i \|x_i - med_{L1}(X)\|}, c\right) = \frac{1}{\left(1+\left|\frac{\|x_i - med_{L1}(X)\|}{4 * median_i \|x_i - med_{L1}(X)\|}\right|\right)^2} \tag{3.47}$$

olmak üzere burada $\|.\|$ Euclid uzaklığı ve c=4'tür.

$$\hat{\sigma} = MSM(r_1,...,r_n) = median\left|r_i - median\, r_j\right| \tag{3.48}$$

olarak ifade edilir (MSM:Mutlak sapmaların medyanı-MAD: Median of Absolute Deviation) (Serneels ve ark., 2005).

3.5.2. Gözlemlerin ağırlıklandırma şekli

Model ve gözlemler arasında uzaklık sadece hata terimi ile ifade edilmez. M-regresyon modelinin adımsal tahmini (3.46) ve (3.47)'den hesaplanan ağırlıkların çarpımıyla belirlenen $w_i = w_i^r w_i^x$ ağırlığının her gözleme uygulanması sonucu belirlenir.

$$\hat{\beta}_{KGM} = \min \sum_{i=1}^{n} w_i \, (y_i - x_i \beta)^2 = \min \sum_{i=1}^{n} \left(\sqrt{w_i^r w_i^x}\, y_i - \sqrt{w_i^r w_i^x}\, x_i \beta \right)^2$$

Amaç fonksiyonunu optimum yapacak parametre değerleri m-regresyon modeli tahmin edicileridir. Aynı fonksiyon aşağıdaki gibi yazılabilir.

$$\hat{\beta}_{KGM} = \min \sum_{i=1}^{n} (\sqrt{w_i}\, y_i - \sqrt{w_i}\, x_i \beta)^2$$

Adım 1: *Y* gözlem değerlerinden noktaların veri setine dikey uzaklıkları $dy_i = y_i - medyan_j(y_j)$ olarak belirlenir. Bu değerlerden $medy = medyan_i(dy_i)$ değeri hesaplanır. (3.48) eşitliğinden y ölçümlerine ilişkin MSM değeri $\hat{\sigma}_y = MSM(dy_1,...,dy_n) = median|dy_i - medy|$ belirlenir. (3.46) eşitliğinden w_i^r ağırlıkları $w_i^r = \dfrac{1}{\left(1 + \left|\dfrac{dy_i}{4\hat{\sigma}}\right|\right)^2}$ olarak hesaplanır. (3.47) eşitliğinden gözlemlerin veri setine olan yatay uzaklıklarından w_i^x ağırlıkları belirlenir,

Adım 2: w_i^r ve w_i^x çarpımlarından gözlemin ağırlığı olan w_i $(= w_i^r\, w_i^x)$ belirlenir,

Adım 3: w_i ağırlığının karekökü ile $(y_i, x_{1i}, \ldots x_{pi})$ noktası çarpılır,

Adım 4: Belirlenen yeni noktalara EKK yöntemi uygulanır,

Adım 5: Tahmin modelinden $r_i = y_i - x_i \hat{\beta}$ hata terimleri belirlenir,

Adım 6: w_i^r ağırlıkları 5. adımdaki hata terimlerine bağlı olarak (3.48) ve (3.46) eşitlikleri kullanılarak belirlenir,

Adım 7: *X* gözlemlerinden de faydalanarak w_i^x ağırlıkları (3.47) eşitliğinden belirlenir,

Adım 8: Her gözlem için $w_i = w_i^r\, w_i^x$ ağırlıkları hesaplanır ve karekökü ile y ve x gözlem değerleri çarpılarak ağırlıklandırılır,

Adım 9: Yeni gözlem değerleri üzerinden EKK model tahmini yapılır,

Adım 10: Tahmin modelinin yeni parametreleri ile bir önceki tahmin modelinin parametrelerinin yakınsaması halinde işlem durdurulur. Aksi halde Adım 5'e devam edilir.

3.5.3. Tek değişkenli kısmi sağlam m-regresyon uygulaması

KSM-regresyon yöntemini ilk olarak 6 gözlemli yapay verilerden oluşan tek değişkenli doğrusal model üzerinde gösterelim.

Çizelge 3.8. 6 gözlemli veri seti

GÖZLEM	Y	X
1	4	2
2	2	5
3	6	8
4	8	10
5	10	10
6	1	12

Adım 1: Y gözlem değerlerinden noktaların veri setine dikey uzaklıkları $dy_i = y_i - medyan_j(y_j)$ olarak belirlenir. Burada $medyan_j(y_j) = 5$ bulunur. $\hat{\sigma}_y = MSM(dy_1,...,dy_n) = median|dy_i - medy|$ eşitliğinden $\hat{\sigma}_y$ hesaplanır.

Adım 2: Dikey uzaklıklara ve varyansa bağlı olarak w_i^r ağırlıkları hesaplanır. (3.46)'den gözlemlerin veri setine olan dikey uzaklıklarından w_i^r ağırlıkları belirlenir ve (3.47)'den gözlemlerin veri setine yatay uzaklıkları w_i^x belirlenir. Sonuç olarak her gözlem için temel ağırlık $w_i = w_i^r w_i^x$ olarak belirlenir.

Çizelge 3.9. 6 gözlemin dikey ağırlıkları

Y	X	$dy_i = y_i - medyan_j(y_j)$	$\lvert dy_i - medyan_j(dy_j)\rvert$	$Z = \frac{dy_i}{\hat{\sigma}_y}$	$w_i^r = \frac{1}{\left(1+\left\lvert\frac{dy_i}{4\hat{\sigma}_y}\right\rvert\right)^2}$
4	2	-1	1	-0,333	0,852
2	5	-3	3	-1,000	0,640
6	8	1	1	0,333	0,852
8	10	3	3	1,000	0,640
10	10	5	5	1,666	0,498
1	12	-4	4	-1,333	0,562

Çizelge 3.10. 6 gözlemin yatay ağırlıkları ve genel ağırlıkları sonucu yeni koordinatları

$\lVert x_i - medyan(X) \rVert$	$w_i^x = \dfrac{1}{\left(1+\left\lvert \dfrac{\lVert x_i - med_{L1}(X)\rVert}{4*median_i \lVert x_i - med_{L1}(X)\rVert} \right\rvert\right)^2}$	$w_i = w_i^r \times w_i^x$	$\sqrt{w_i}$	$Y' = \sqrt{w_i}Y$	$X' = \sqrt{w_i}X$
7	0,284	0,241	0,492	1,969	0,984
4	0,444	0,284	0,533	1,066	2,666
1	0,790	0,673	0,820	4,923	6,564
1	0,790	0,505	0,711	5,688	7,111
1	0,790	0,393	0,627	6,274	6,274
3	0,529	0,297	0,545	0,545	6,545

Değiştirilen her Y ve X ölçümleriyle EKK yapıldığında regresyon modelinin tahmini $Y' = 0{,}680 + 0{,}543X'$ şeklinde olur. Bu modelden belirlenen hata terimlerine bağlı olarak ve yeni x' gözlem değerlerine bağlı olarak ikinci kez ağırlıklandırma yapılır.

İlk aşamada dikey ağırlıklandırma Y gözlemleri üzerinden yapılırken daha sonraki aşamalarda dikey ağırlıklandırma EKK regresyon tahmininin hata terimleri üzerinden yapılır.

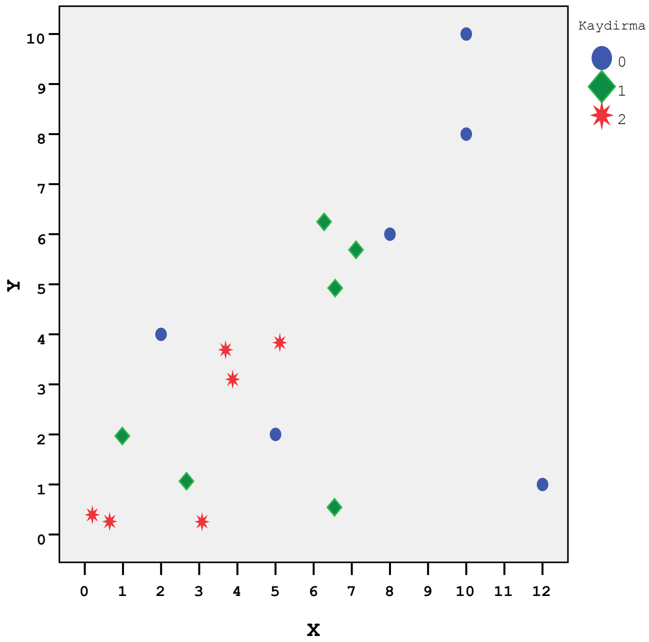
Çizelge 3.11. Gözlemlerin ikinci aşamada dikey ağırlıkları

Y'	X'	$r_i = y_i - \hat{y}_i$	$\lvert r_i - medyan_j(r_j)\rvert$	$Z = \dfrac{r_i}{\hat{\sigma}}$	$w_i^r = \dfrac{1}{\left(1+\left\lvert \dfrac{r_i}{4\hat{\sigma}}\right\rvert\right)^2}$
1,969	0,984	0,754	0,039	0,794	0,696
1,066	2,666	-1,063	1,777	-1,120	0,610
4,923	6,564	0,675	0,039	0,711	0,720
5,688	7,111	1,143	0,429	1,204	0,590
6,274	6,274	2,184	1,469	2,300	0,403
0,545	6,545	-3,692	4,406	-3,889	0,257

Çizelge 3.12. Gözlemlerin ikinci aşamada yatay ağırlıkları ve genel ağırlıkları

$\lVert x_i - medyan(X) \rVert$	$w_i^x = \dfrac{1}{\left(1+\left\lvert \dfrac{\lVert x_i - med_{L1}(X)\rVert}{4*median_i \lVert x_i - med_{L1}(X)\rVert} \right\rvert\right)^2}$	$w_i = w_i^r \times w_i^x$	$\sqrt{w_i}$	$Y'' = \sqrt{w_i}Y'$	$X'' = \sqrt{w_i}X'$
5,425	0,057	0,040	0,200	0,393	0,196
3,743	0,098	0,060	0,245	0,261	0,653
0,155	0,841	0,606	0,778	3,834	5,112
0,701	0,503	0,297	0,545	3,101	3,876
0,135	0,858	0,346	0,588	3,691	3,691
0,135	0,858	0,220	0,469	0,256	3,074

Şekil 3.3'ten de görüldüğü gibi KSM-regresyon yöntemi de noktaları her adımda ağırlıklandırma sonucu orijine taşımaktadır.



Şekil 3.3. 6 gözlemin iki ağırlıklandırma sonucunda yeni koordinatları

3.5.4. Çoklu kısmi sağlam m-regresyon uygulaması

KSM-regresyon yöntemini ikinci olarak 6 gözlemli yapay verilerden oluşan iki değişkenli (çoklu) doğrusal model üzerinde uygulayalım.

Çizelge 3.13. 6 gözlemli iki bağımsız değişkenli veri seti

GÖZLEM	Y	X_1	X_2
1	2	1	1
2	0	1	2
3	3	2	2
4	0	2	3
5	4	3	3
6	8	3	4

EKK yöntemiyle parametreler tahmin edilir.

Regresyon modelinin tahmini;

$$Y = -2{,}000 + 2{,}833X_1 - 0{,}333X_2$$

Adım 1: Y gözlem değerlerinden noktaların veri setine dikey uzaklıkları $dy_i = y_i - medyan_j(y_j)$ olarak belirlenir.

Burada $medyan_j(y_j) = 5$ bulunur.

$\hat{\sigma}_y = MSM(dy_1,...,dy_n) = median|dy_i - medy|$ eşitliğinden $\hat{\sigma}_y$ hesaplanır.

Adım 2: Dikey uzaklıklara ve varyansa bağlı olarak w_i^r ağırlıkları hesaplanır. (3.46)'den gözlemlerin veri setine olan dikey uzaklıklarından w_i^r ağırlıkları belirlenir ve (3.47)'den gözlemlerin veri setine yatay uzaklıkları w_i^x belirlenir. Sonuç olarak her gözlem için temel ağırlık $w_i = w_i^r w_i^x$ olarak belirlenir.

Çizelge 3.14. Gözlemlerin ilk dikey ağırlıklandırılmasında Y değerleri kullanılmıştır

Y	X_1	X_2	$dy_i = y_i - medyan_j(y_j)$	$\lvert dy_i - medyan_j(dy_j)\rvert$	$Z = \frac{dy_i}{\hat{\sigma}_y}$	$w_i^r = \frac{1}{\left(1+\left\lvert\frac{dy_i}{4\hat{\sigma}_y}\right\rvert\right)^2}$
2	1	1	-0,5	0,5	-0,250	0,886
0	1	2	-2,5	2,5	-1,250	0,580
3	2	2	0,5	0,5	0,250	0,886
0	2	3	-2,5	2,5	-1,250	0,580
4	3	3	1,5	1,5	0,750	0,709
8	3	4	5,5	5,5	2,750	0,351

Çizelge 3.15. Gözlemlerin ilk yatay ağırlıklandırılmasında X değerleri kullanılmıştır

$T_{1i} = \lVert x_{1i} - medyan(X_1)\rVert$	$T_{2i} = \lVert x_{2i} - medyan(X_2)\rVert$	$\sqrt{T_{1i}^2 + T_{2i}^2}$	$T_2 = \frac{\sqrt{T_{1i}^2 + T_{2i}^2}}{median(\sqrt{T_{1i}^2 + T_{2i}^2})}$	$w_i^x = \frac{1}{(1+\lvert T_2/4\rvert)^2}$
1	1,5	1,803	1,612	0,508
1	0,5	1,118	1,000	0,640
0	0,5	0,500	0,447	0,809
0	0,5	0,500	0,447	0,809
1	0,5	1,118	1,000	0,640
1	1,5	1,803	1,613	0,508

Çizelge 3.16. Temel ağırlıklandırma ve sonucunda gözlemlerin yeni koordinatları

$w_i = w_i^r w_i^x$	$\sqrt{w_i}$	$Y' = \sqrt{w_i}Y$	$X_1' = \sqrt{w_i}X_1$	$X_2' = \sqrt{w_i}X_2$
0,450	0,671	1,342	0,671	0,671
0,372	0,610	0,000	0,610	1,219
0,717	0,847	2,540	1,693	1,693
0,470	0,685	0,000	1,371	2,056
0,454	0,674	2,695	2,021	2,021
0,178	0,422	3,379	1,267	1,689

Değiştirilen her Y ve X ölçümleriyle EKK yapıldığında regresyon modelinin tahmini;

$$Y' = 0{,}951 + 3{,}058X_1{}' - 2{,}042X_2{}'$$

şeklinde olur.

İlk aşamada dikey ağırlıklandırma Y gözlemleri üzerinden yapılırken daha sonraki aşamalarda dikey ağırlıklandırma EKK regresyon tahmininin hata terimleri üzerinden yapılır.

Çizelge 3.17. İkinci aşamada dikey ağırlıklar

Y'	$X_1{}'$	$X_2{}'$	$r_i = y_i - \hat{y}_i$	$\lvert r_i - medyan_j(r_j)\rvert$	$Z = \frac{r_i}{\hat{\sigma}}$	$w_i^r = \frac{1}{\left(1+\left\lvert\frac{r_i}{4\hat{\sigma}}\right\rvert\right)^2}$
1,342	0,671	0,671	-0,291	0,009	-2,998	0,327
0,000	0,610	1,219	-0,326	0,026	-3,358	0,296
2,540	1,693	1,693	-0,132	0,168	-1,358	0,557
0,000	1,371	2,056	-0,944	0,644	-9,737	0,085
2,695	2,021	2,021	-0,310	0,010	-3,195	0,309
3,379	1,267	1,689	2,003	2,303	20,646	0,026

Çizelge 3.18. İkinci aşamada yatay ağırlıklar

$T_{1i} = \lVert x_{1i} - medyan(X_1)\rVert$	$T_{2i} = \lVert x_{2i} - medyan(X_2)\rVert$	$\sqrt{T_{1i}^2 + T_{2i}^2}$	$T_2 = \frac{\sqrt{T_{1i}^2 + T_{2i}^2}}{median(\sqrt{T_{1i}^2 + T_{2i}^2})}$	$w_i^x = \frac{1}{(1+\lvert T_2/4\rvert)^2}$
0,648	1,020	1,209	2,102	0,430
0,709	0,472	0,852	1,482	0,532
0,374	0,002	0,374	0,651	0,740
0,052	0,365	0,368	0,641	0,743
0,702	0,330	0,776	1,349	0,559
0,052	0,002	0,052	0,090	0,956

Çizelge 3.19. Gözlemlerin temel ağırlıkları ve yeni koordinatları

$w_i = w_i^r \times w_i^x$	$\sqrt{w_i}$	$Y'' = \sqrt{w_i}Y'$	$X_1{}'' = \sqrt{w_i}X_1{}'$	$X_2{}'' = \sqrt{w_i}X_2{}'$
0,140	0,375	0,503	0,251	0,251
0,157	0,397	0,000	0,242	0,484
0,412	0,642	1,631	1,087	1,087
0,063	0,251	0,000	0,344	0,516
0,173	0,416	1,120	0,840	0,840
0,025	0,159	0,536	0,201	0,268

4. DİK M-REGRESYON

Sağlam regresyon metotları içerisinde önemli bir yere sahip olan En Çok Olabilirlik (Maximum Likelihood) tahmin edicilerini kullanan M-regresyon ve KSM-regresyon Bölüm 3'te örneklerle açıklandı. Aynı tahmin edicilere dayalı fakat farklı bir yaklaşım (farklı bir ağırlıklandırma:dik ağırlıklandırma) yöntemi bu bölümde sunulmuştur.

Öncelikle neden sağlam (robust) modele ihtiyaç duyulduğunun anlaşılabilmesi için veri seti içerisindeki aykırı gözlem/ler (outlier/s) incelenecektir (Pena ve Yohai, 1999; Wu ve Lee, 2006).

Aykırı gözlemler içeren veri setine uygulanacak ağırlıklandırma yöntemi örneklerle açıklanacaktır.

Son alt bölümde ise etkinliği gösterilmiş KSM-regresyon yöntemi ile önerilen Dik M-regresyon yönteminin simülasyon karşılaştırmaları verilmiştir.

4.1. Aykırı Gözlemler

Bir gözlemin aykırılığı temelde farklı şekillerde incelenebilir. Kısaca aykırı gözlemler genel dağılımın dışında yer alan gözlemlerdir. Genellikle gözlemin merkeze olan uzaklığı ve genel yaklaşım gözlemin merkeze olan Mahalanobis uzaklığı dikkate alınarak belirlenmektedir (Rousseeuw ve Van Zomeren, 1990; Liu ve ark., 2004; Saltenis, 2004; Filzmoser ve ark., 2008; Fauconnier ve Haesbroeck, 2009; Williems ve Zamar, 2009).

Buna karşın regresyon analizinde aykırı gözlemleri 3 farklı yönüyle incelemek mümkündür. İlk olarak bir regresyon modeline beklenenden çok daha uzak gözlemler varsa bu gözlemlere aykırı gözlemler denir. Başka bir ifadeyle bu gözlemler aşırı büyük hata terimi gösteren gözlemlerdir (Penny, 1996).

Diğer bir durum ise değişkenler içerisinde aykırı gözlemlerin olmasıdır. Regresyon analizinde gözlemler bağımlı (Y) ve bağımsız değişkenler (X) açısından iki ayrı aykırılık durum açısından incelebilir. Gözlemler Y ölçümleri açısından aykırılık gösterebilir ya da X ölçümleri açısından gösterebilir.

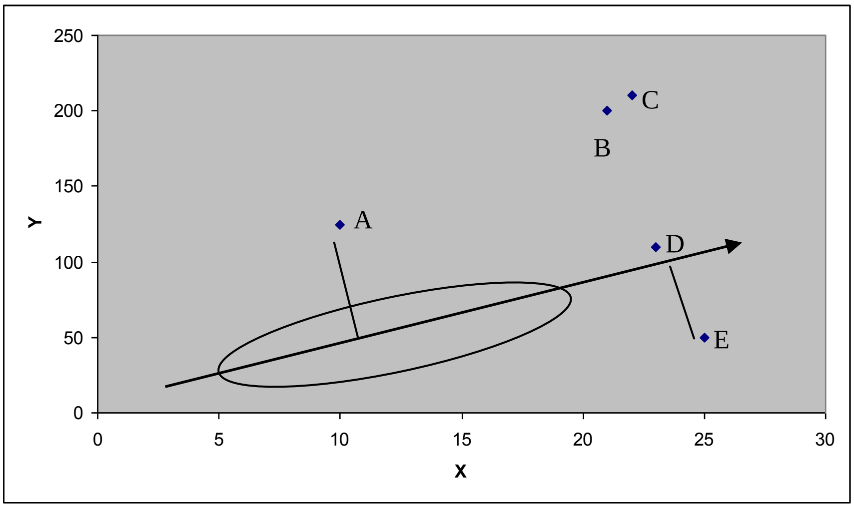
Bölüm 3'te M-regresyon yöntemi EKK yöntemine göre aşırı hatalar içeren ya da regresyon eğrisine diğer gözlemlere göre çok daha uzak duran aykırı gözlemler varsa bu uzaklıklara (hata terimlerine) bağlı w^r ağırlıklarını dikkate almıştır.

KSM-regresyon ise hem EKK regresyon modeline olan uzaklıkları w^r ile ve hem de gözlemlerin bağımsız değişken/ler içerisindeki w^x ağırlıklarını dikkate alarak regresyon modelinin M-tahminlerini adımsal olarak belirlemektedir.

Şekil 4.1'de gözlemlerin çoğunun elips içerisinde olduğunu düşünelim ve sadece 5 gözlemin (A, B, C, D ve E) bu yapı dışında olduğunu ve aykırı olduğunu görürüz (Adnan ve ark., 2003).

A gözlemi X ölçümleri açısından aykırı değildir çünkü X gözlemleri 5-20 arasında değişmektedir ve A gözleminin X değeri bu aralıktadır. Fakat, A gözlemi Y ölçümlerine, okla gösterilen regresyon doğrusuna ve diğer gözlemlere (elips içerisindeki) göre aykırıdır.

B ve C gözlemleri regresyon doğrusuna, Y ölçümlerine ve hatta X ölçümlerine göre elips içerisindeki gözlemlere nazaran aykırılık göstermektedir. D gözlemi sadece regresyon modeline göre aykırılık göstermezken E gözlemi sadece Y ölçümlerine göre aykırılık göstermemektedir. Bölüm 4.2.'de önerilen Dik M-regresyon yöntemi gözlemin aykırılığı nasıl olursa olsun regresyon doğrusuna olan dik uzaklıklarına bağlı bir ağırlıklandırma yaklaşımı kullanmaktadır.



Şekil 4.1. Gözlemlerin Y, X ve regresyon eksenlerine göre aykırılıkları

Çizelge 4.1. Gözlemlerin Y, X ve regresyon eksenlerine göre aykırılıklarının değerlendirilmesi

Noktalar	Y-ekseninde	X-ekseninde	Regresyona göre
A	AYKIRI	-	AYKIRI
B	AYKIRI	AYKIRI	AYKIRI
C	AYKIRI	AYKIRI	AYKIRI
D	AYKIRI	AYKIRI	-
E	-	AYKIRI	AYKIRI

4.2. Dik M-Regresyon

Bu yöntem gözlemlerin EKK regresyon doğrusuna olan dik uzaklıklarına bağlı gözlemlerin her adımda ağırlıklandırma esasına bağlı adımsal bir yöntemdir. Bu yöntem aykırı veya değil her gözlemi EKK regresyon doğrusuna dik olarak her gözlemi dik uzaklığına orantılı olarak yaklaştırır. Taşınan noktalara göre tekrar EKK regresyon modeli belirlenir ve gözlemlerin yeni modele göre tekrar taşınması sağlanır. Bu işlem ardışık modellerin ya da başka bir ifadeyle model parametrelerinin ardışık tahminlerinin yakınsaması halinde durulur ve Dik M-regresyon modelini oluşturur.

Bu yöntemin adımsal aşamaları tek değişkenli doğrusal ve çoklu doğrusal modellerde ayrı ayrı alt bölümlerde örneklerle açıklanmıştır.

4.2.1. Tek değişkenli doğrusal regresyonda dik m-regresyon

Adım 1 : $Y = \beta_0 + \beta_1 X + \varepsilon$ modelinin veya parametrelerinin tahmini EKK yöntemiyle belirlenir. Bu durumda tahmin modeli $y = b_0 + b_1 x$ şeklinde olur.

Adım 2 : Verilen her gözlem (noktanın) hata terimlerinden oluşan hata vektörü $r = [r_1, r_2, r_3, ... r_n]$'dir. Burada, $r_i = y_i - b_0 - b_1 x_i$ $(i = 1,...,n)$.

Adım 3: Her bir noktanın regresyon eğrisine dik uzaklığı (d_i) aşağıdaki eşitlikle hesaplanır.

$$d_i = \frac{|r_i|}{\sqrt{1 + b_1^2}} \qquad i = 1,...,n$$

Adım 4: Bulunan dik uzaklıklar yardımıyla her noktanın d_i uzaklığına orantılı olarak ağırlıklı kayma miktarı hesaplanır.

$$k_i = \frac{d_i}{\sum_{i=1}^{n} d_i} d_i$$

Adım 5: Ulaşılan yeni noktanın koordinatlarının bulunabilmesi için $A_1X + A_2Y + C = 0$ düzlemine dik ve (x, y) noktalarından geçen doğrunun denkleminin bilinmesi gerekir.

$\frac{X - x}{A_1} = \frac{Y - y}{A_2} = \ell$ buradan yeni noktanın koordinatları da aynı doğru üzerinden geçeceği için;

$$\frac{X' - x}{A_1} = \frac{Y' - y}{A_2} = \ell \tag{4.1}$$

denklemi elde edilir.

(x, y) ve (X', Y') noktaları arasındaki uzaklık

$$k = \sqrt{(x - X')^2 + (y - Y')^2} \tag{4.2}$$

olur. Bu durumda (4.1) ve (4.2)'den k_i kayma miktarı için kayma yönü

$$l_i^2 = \frac{k_i^2}{A_1^2 + A_2^2}$$

eşitliğinden l_i eğimi bulunur. Noktanın regresyon eğrisine dik olacak şekilde kayma yönü hata teriminin işareti ile aynıdır. Bu durumda, l_i eğim değeri şöyle tanımlanır;

$$l_i = \begin{cases} -l_i & r_i < 0 \\ l_i & r_i > 0 \end{cases}$$

Adım 6: (x, y) noktası l yönünde ve k birim kaydırıldığında yeni noktanın koordinatları; $(X', Y') = (A_1 l + x, A_2 l + y)$ denklemleri kullanılarak elde edilir.

Adım 7: Elde edilen yeni noktalarla aynı işlemler tekrarlanır ve parametre tahminleri elde edilir.

Adım 8: Bir önceki parametre tahminleri ile takip eden parametre tahminleri arasındaki farkların karelerinin toplamının karekökü 0,01'den küçükse parametre tahmini belirlenmiş olur ve modelin tahmin işlemi sona erer. Ya da başka bir ifadeyle durdurma kuralı aşağıdaki gibidir.

$$\left[\sum_{i=0}^{p}\left(\beta_i - \beta_i\right)^2\right]^{1/2} < 0,01$$

Örnek 4.1.

Çizelge 4.2. 6 gözlemli veri seti

GÖZLEM	Y	X
1	4	2
2	2	5
3	6	8
4	8	10
5	10	10
6	1	12

Yukarıdaki veri setine Dik M-regresyon yöntemini uygulayalım:

Adım 1: EKK yöntemiyle parametreler tahmin edilir. Modelin tahmini $Y = 3{,}441X_1 + 0{,}220X_2$ olur.

Adım 2: Hata terimlerine ($r_i = y_i - \hat{y}_i$) bağlı olarak ağırlıklar belirlenir.

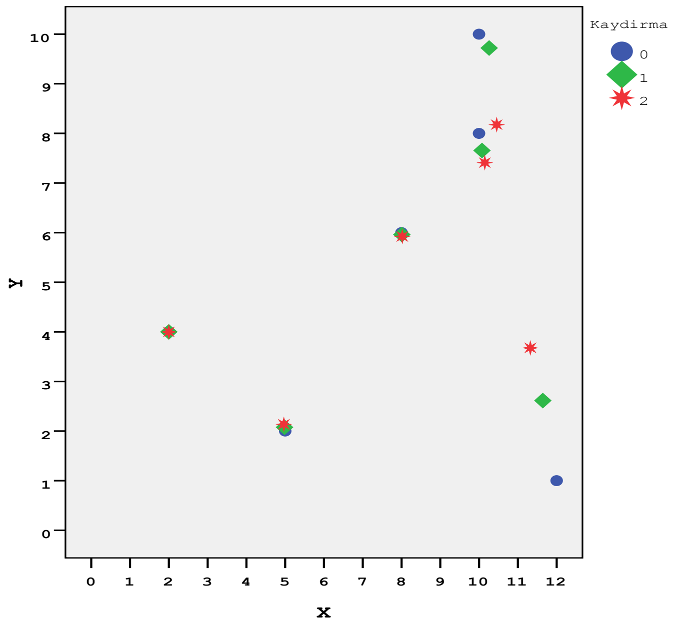
Çizelge 4.3. 6 gözlemin ilk ağırlıklandırma işlemi sonrası koordinatları

Y	X	$r_i = y_i - \hat{y}_i$	$d_i = \frac{\lvert r_i \rvert}{\sqrt{1+\sum_{i=1}^{n} b_i^2}}$	$k_i = \frac{d_i}{\sum_{i=1}^{n} d_i} d_i$	$l_i = \sqrt{\frac{k_i^2}{0{,}220^2 + (-1)^2}}$	$Y' = -1l_i + Y$	$X' = 0{,}220l_i + X$
4	2	0,119	0,116	0,0008	0,0007	3,999	2,000
2	5	-2,542	2,483	0,079	-0,077	2,077	4,983
6	8	0,797	0,778	0,040	0,039	5,961	8,008
8	10	2,356	2,301	0,354	0,345	7,655	10,075
10	10	4,356	4,254	1,212	1,183	8,817	10,260
1	12	-5,085	4,966	1,653	-1,614	2,614	11,644
		Toplam	14,898				

Adım 3: EKK yöntemiyle parametreler tahmin edilir. Modelin tahmini $Y' = 2{,}818 + 0{,}303X'$. Bir önceki EKK parametre tahminleri ile bu parametre tahminleri arasındaki fark istenilen düzeyde olmadığından tekrar ağırlıklandırma işlemi yapılır ve model tekrar tahmin edilir. İşlem bu örnekte 2 adım ilerletilmiştir.

Çizelge 4.4. 6 gözlemin ikinci ağırlıklandırma işlemi sonrası koordinatları

Y'	X'	$r_i = y_i - \hat{y}_i$	$d_i = \frac{\lvert r_i \rvert}{\sqrt{1+b_1^2}}$	$k_i = \frac{d_i}{\sum_{i=1}^{n} d_i} d_i$	$l_i = \sqrt{\frac{k_i^2}{0{,}303^2 + (-1)^2}}$	$Y'' = -1l_i + Y'$	$X'' = 0{,}303l_i + X'$
3,999	2,000	0,576	0,551	0,026	0,0006	3,998	2,000
2,077	4,983	-2,249	2,154	0,254	-0,059	2,136	4,965
5,961	8,008	0,719	0,688	0,041	0,039	5,922	8,019
7,655	10,075	1,788	1,712	0,255	0,244	7,411	10,148
8,817	10,260	2,894	2,772	0,670	0,641	8,176	10,454
2,614	11,644	-3,728	3,570	1,110	-1,062	3,676	11,322
		Toplam	11,447				



Şekil 4.2 6 gözlemin iki ağırlıklandırma sonucunda yeni koordinatları

Şekil 4.2'de görüldüğü gibi daha aykırı noktalar tahmini regresyon eğrisine dik olarak daha fazla yaklaşmaktadır.

4.2.2. Çoklu doğrusal regresyonda dik m-regresyon

Adım 1 : $Y = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$ modelinin veya parametrelerinin tahmini EKK yöntemiyle belirlenir. Bu durumda tahmin modeli $y = b_1 x_1 + b_2 x_2 + ... + b_p x_p$ olur.

Adım 2 : Verilen her noktanın hata terimlerinden oluşan hata vektörü $r = [r_1, r_2, r_3, ... r_n]$ olur. $r_i = y_i - b_1 x_{1i} - b_2 x_{2i} - ... - b_p x_{pi} = y_i - b_0 - \sum_{j=1}^{p} b_j x_{ji} \quad i = 1,...,n$

Adım 3 : Her bir noktanın regresyon eğrisine dik uzaklığı (d_i) aşağıdaki eşitlikle hesaplanır.

$$d_i = \frac{|r_i|}{\sqrt{1 + \sum_{i=1}^{p} b_i^2}} \qquad i = 1,...,n$$

Adım 4 : Bulunan dik uzaklıklar yardımıyla her noktanın regresyon doğrusuna olan dik uzaklığı ile orantılı olarak ağırlıklı kayma miktarı hesaplanır

$$k_i = \frac{d_i}{\sum_{i=1}^{n} d_i} d_i$$

Adım 5: Ulaşılan yeni noktanın koordinatlarının bulunabilmesi için $A_1 X + A_2 Y + A_3 Z + ... + A_n W + D = 0$ düzlemine dik ve $(x, y, z, ..., w)$ noktalarından geçen doğrunun denkleminin bilinmesi gerekir.

$$\frac{X - x}{A_1} = \frac{Y - y}{A_2} = \frac{Z - z}{A_3} = ... = \frac{W - w}{A_n} = \ell$$

buradan yeni noktanın koordinatları da aynı doğru üzerinden geçeceği için;

$$\frac{X' - x}{A_1} = \frac{Y' - y}{A_2} = \frac{Z' - z}{A_3} = ... = \frac{W' - w}{A_n} = \ell \qquad (4.3)$$

denklemi elde edilir.

Adım 6 : Gözlemin ilk $(x, y, z, ..., w)$ noktasındaki yeri ile taşınma sonundaki koordinatları $(X', Y', Z', ..., W')$ arasındaki uzaklık ya da kayma miktarı;

$$k = \sqrt{(x - X')^2 + (y - Y')^2 + (z - Z')^2 + ... + (w - W')^2} \quad (4.4)$$

olur. Bu durumda (4.3) ve (4.4)'den k_i kayma miktarı için

$$l_i^2 = \frac{k_i^2}{{A_1}^2 + {A_2}^2 + {A_3}^2 + ... + {A_n}^2}$$

eşitliğinden l_i eğimi bulunur. Noktanın kayma yönü hata teriminin işareti ile aynıdır. Bu durumda, l_i eğim değeri şöyle tanımlanır:

$$l_i = \begin{cases} -l_i & r_i < 0 \\ l_i & r_i > 0 \end{cases}$$

Yeni noktanın koordinatları;

$$(X', Y', Z', ..., W') = (A_1\ell + x, A_2\ell + y, A_3\ell + z, ..., A_n\ell + w)$$

denklemleri kullanılarak elde edilir.

Adım 7 : Elde edilen yeni noktalarla aynı işlemler tekrarlanır ve parametre tahminleri elde edilir.

Adım 8 : Bir önceki parametre tahminleri ile takip eden tahminler arasındaki farkların karelerinin toplamının karekökü 0,01'den küçükse parametre tahmini belirlenmiş olur ve işlem sona erer. Kısaca durdurma kuralı aşağıda ifade edilmiştir.

$$\left[\sum_{i=0}^{p} \left(\beta_i(1) - \beta_i(0)\right)^2\right]^{1/2} < 0{,}01$$

Aksi halde Adım 1'e dönülür.

Örnek 4.2.

Çizelge 4.5. 6 gözlemli iki bağımsız değişkenli veri seti

GÖZLEM	Y	X_1	X_2
1	2	1	1
2	0	1	2
3	3	2	2
4	0	2	3
5	4	3	3
6	8	3	4

Yukarıdaki veri setine Dik M-regresyon yöntemini uygulayalım:

Adım 1: EKK yöntemiyle parametreler tahmin edilir. Modelin tahmini

$Y = -2{,}000 + 2{,}833X_1 - 0{,}333X_2$ olur.

Adım 2: Hata terimlerine ($r_i = y_i - \hat{y}_i$) bağlı olarak ağırlıklar belirlenir.

Çizelge 4.6. Gözlemlerin ilk aşamadaki kayma miktarları ve eğimleri

Y	X_1	X_2	$r_i = y_i - \hat{y}_i$	$d_i = \frac{\lvert r_i \rvert}{\sqrt{1+\sum_{i=1}^{p} b_i^2}}$	$k_i = \frac{d_i}{\sum_{i=1}^{n} d_i} d_i$	$l_i = \sqrt{\frac{k_i^2}{2{,}833^2 + (-0{,}333)^2 + (-1)^2}}$
2	1	1	1,500	0,496	0,086	0,028
0	1	2	-0,167	0,055	0,001	0,000
3	2	2	0,000	0,000	0,000	0,000
0	2	3	-2,667	0,882	0,272	-0,090
4	3	3	-1,500	0,496	0,086	-0,028
8	3	4	2,833	0,938	0,307	0,101
			Toplam	2,867		

Çizelge 4.7. Gözlemlerin ilk ağırlıklandırma işlemi sonrası koordinatları

$Y' = -1l_i + Y$	$X_1{}' = 2{,}833l_i + X_1$	$X_2{}' = -0{,}333l_i + X_2$
1,972	1,089	0,991
0,000	1,001	2,000
3,000	2,000	2,000
0,090	1,745	3,030
4,028	2,919	3,009
7,899	3,287	3,966

Adım 3: EKK yöntemiyle parametreler tahmin edilir. Modeli tahmini $Y' = -1{,}865 + 3{,}789 X_1' - 1{,}162 X_2'$. Bir önceki EKK parametre tahminleri ile bu parametre tahminleri arasındaki fark istenilen düzeyde olmadığından tekrar ağırlıklandırma işlemi yapılır ve model tekrar tahmin edilir. İşlem bu örnekte 2 adım ilerletilmiştir.

Çizelge 4.8. Gözlemlerin ikinci aşamadaki kayma miktarları ve eğimleri

Y'	X_1'	X_2'	$r_i = y_i - \hat{y}_i$	$d_i = \dfrac{\lvert r_i \rvert}{\sqrt{1+\sum_{i=1}^{p} b_i^2}}$	$k_i = \dfrac{d_i}{\sum_{i=1}^{n} d_i} d_i$	$l_i = \sqrt{\dfrac{k_i^2}{3{,}789^2 + (-1{,}162)^2 + (-1)^2}}$
1,972	1,089	0,991	0,893	0,218	0,030	0,007
0,000	1,001	2,000	0,395	0,097	0,006	0,001
3,000	2,000	2,000	-0,390	0,095	0,006	-0,001
0,090	1,745	3,030	-1,139	0,279	0,050	-0,012
4,028	2,919	3,009	-1,673	0,409	0,107	-0,026
7,899	3,287	3,966	1,914	0,468	0,140	0,034
			Toplam	1,566		

Çizelge 4.9. Gözlemlerin ikinci ağırlıklandırma işlemi sonrası koordinatları

$Y'' = -1 l_i + Y'$	$X_1'' = 3{,}789 l_i + X_1'$	$X_2'' = -1{,}162 l_i + X_2'$
1,964	1,109	0,982
-0,002	1,007	1,998
3,001	1,995	2,002
0,102	1,699	3,044
4,055	2,820	3,040
7,864	3,417	3,926

4.3. Kısmi Sağlam ve Dik M-Regresyon Yöntemlerinin Monte-Carlo Simülasyon Karşılaştırması

KSMR ile DMR yöntemlerinin aykırı gözlemlere karşı duyarlılığını belirleyecek tek değişkenli ve çoklu doğrusal olmak üzere iki modelde Monte-Carlo simülasyon çalışması aşağıda verilmiştir.

4.3.1. Tek değişkenli doğrusal regresyonda Monte-Carlo simülasyonu

Gözlem değerleri X'lerin dağılımı 0-100 aralığında düzgün dağılım ve hata terimlerinin dağılımı $\varepsilon \sim N(0{,}16)$ olmak üzere $Y = 10 + 2X + \varepsilon$ modelinden üretilen $n = 10{,}20{,}30{,}40{,}50$ hacimli örnekler üzerinde aykırı gözlemlerin oluşturulmasında $\varepsilon \sim N(50{,}16)$ dağılımının kullanılmasıyla aykırı gözlem sayısı 1,2,3,4 ve 5 olmak üzere 5000 tekrar sonucu parametre tahminlerine ilişkin hata kareler ortalaması karşılaştırılmak üzere aşağıdaki eşitlikle belirlenmiş ve tabloda sunulmuştur ($m_{tek} = 5000$). Parametrelere ait varyans tahminleri $\hat{\sigma}^2_{\beta_i} = HKO_{\beta_i} = \frac{1}{m_{tek}} \sum_{j=1}^{m_{tek}} \left(\hat{\beta}_{ij} - \beta_i\right)^2$, $(i = 0{,}1 \;\; \beta_0 = 10, \beta_1 = 2)$ ise bu durumda, $\hat{\sigma}^2_{\beta_0} = HKO_{\beta_0} = \frac{1}{m_{tek}} \sum_{j=1}^{m_{tek}} \left(\hat{\beta}_{0j} - 10\right)^2$ ve $\hat{\sigma}^2_{\beta_1} = HKO_{\beta_1} = \frac{1}{m_{tek}} \sum_{j=1}^{m_{tek}} \left(\hat{\beta}_{1j} - 2\right)^2$ olur.

Çizelge 4.10. Tek değişkenli doğrusal regresyon modeli için KSMR ve DMR metodlarının Monte Carlo simülasyon karşılaştırması

		KSMR		DMR	
n	**Aykırı gözlem sayısı**	$\hat{\sigma}^2_{\beta_0}$	$\hat{\sigma}^2_{\beta_1}$	$\hat{\sigma}^2_{\beta_0}$	$\hat{\sigma}^2_{\beta_1}$
10	1	99,988	0,216	42,171	0,010
	2	99,991	0,292	145,843	0,040
	3	99,992	0,385	326,051	0,109
	4	99,996	0,606	582,825	0,187
	5	99,997	0,970	934,500	0,293
20	1	99,987	0,020	7,017	0,002
	2	99,987	0,020	14,627	0,004
	3	99,985	0,019	26,140	0,008
	4	99,985	0,020	54,988	0,018
	5	99,982	0,020	83,960	0,028
30	1	99,987	0,020	3,426	0,001
	2	99,987	0,019	5,418	0,001
	3	99,987	0,020	8,774	0,003
	4	99,986	0,019	14,257	0,004
	5	99,983	0,018	21,428	0,006
40	1	99,988	0,020	2,507	0,0007
	2	99,988	0,019	3,651	0,001
	3	99,988	0,019	5,097	0,001
	4	99,987	0,019	7,698	0,002
	5	99,986	0,019	9,397	0,003
50	1	99,989	0,019	1,992	0,0006
	2	99,988	0,019	2,379	0,0007
	3	99,988	0,019	3,319	0,001
	4	99,988	0,019	4,582	0,001
	5	99,988	0,019	5,644	0,001

4.3.2. Çoklu doğrusal regresyonda Monte-Carlo simülasyonu

Gözlem değerleri X'lerin dağılımı 0-100 aralığında düzgün dağılım ve hata terimlerinin dağılımı $\varepsilon \sim N(0{,}16)$ olmak üzere $Y = 10 + 2X_1 - X_2 + \varepsilon$ modelinden üretilen $n = 10{,}20{,}30{,}40{,}50$ hacimli örnekler üzerinde aykırı gözlemlerin oluşturulmasında $\varepsilon \sim N(50{,}16)$ dağılımının kullanılmasıyla aykırı gözlem sayısı 1,2,3,4 ve 5 olmak üzere 5000 tekrar sonucu parametre tahminlerine ilişkin hata kareler ortalaması karşılaştırılmak üzere aşağıdaki eşitlikle belirlenmiş ve tabloda sunulmuştur ($m_{tek} = 5000$). Parametrelere ait varyans tahminleri $\hat{\sigma}^2_{\beta_i} = HKO_{\beta_i} = \frac{1}{m_{tek}} \sum_{j=1}^{m_{tek}} \left(\hat{\beta}_{ij} - \beta_i\right)^2$,

$(i = 0{,}1{,}2 \;\; \beta_0 = 10, \beta_1 = 2, \beta_2 = -1)$ ise bu durumda, $\hat{\sigma}^2_{\beta_0} = HKO_{\beta_0} = \frac{1}{m_{tek}} \sum_{j=1}^{m_{tek}} \left(\hat{\beta}_{0j} - 10\right)^2$,

$\hat{\sigma}^2_{\beta_1} = HKO_{\beta_1} = \frac{1}{m_{tek}} \sum_{j=1}^{m_{tek}} \left(\hat{\beta}_{1j} - 2\right)^2$ ve $\hat{\sigma}^2_{\beta_2} = HKO_{\beta_2} = \frac{1}{m_{tek}} \sum_{j=1}^{m_{tek}} \left(\hat{\beta}_{2j} + 1\right)^2$ olur.

Çizelge 4.11. Çoklu doğrusal regresyon modeli için KSMR ve DMR metodlarının Monte Carlo simülasyon karşılaştırması

		KSMR			**DMR**		
n	**Aykırı gözlem sayısı**	$\hat{\sigma}^2_{\beta_0}$	$\hat{\sigma}^2_{\beta_1}$	$\hat{\sigma}^2_{\beta_2}$	$\hat{\sigma}^2_{\beta_0}$	$\hat{\sigma}^2_{\beta_1}$	$\hat{\sigma}^2_{\beta_2}$
10	1	99,867	0,031	0,023	105,548	0,019	0,019
	2	99,927	0,041	0,029	409,484	0,069	0,076
	3	99,954	0,164	0,081	908,575	0,156	0,141
	4	99,959	0,437	0,199	1276,896	0,214	0,234
	5	99,992	0,702	0,376	1997,342	0,331	0,357
20	1	99,851	0,008	0,008	14,060	0,002	0,002
	2	99,829	0,008	0,008	30,904	0,006	0,005
	3	99,825	0,008	0,008	52,104	0,019	0,009
	4	99,851	0,012	0,009	99,708	0,019	0,016
	5	99,871	0,021	0,012	143,663	0,032	0,024
30	1	99,829	0,007	0,007	6,197	0,001	0,001
	2	99,830	0,007	0,007	10,399	0,002	0,001
	3	99,820	0,007	0,006	16,816	0,003	0,003
	4	99,798	0,007	0,007	21,462	0,005	0,004
	5	99,805	0,007	0,007	32,924	0,008	0,006
40	1	99,822	0,007	0,006	4,617	0,0008	0,0007
	2	99,821	0,007	0,006	5,896	0,001	0,001
	3	99,789	0,006	0,006	8,191	0,001	0,001
	4	99,808	0,006	0,006	10,658	0,002	0,002
	5	99,795	0,006	0,006	15,029	0,004	0,003
50	1	99,818	0,007	0,006	3,406	0,0006	0,0006
	2	99,820	0,006	0,006	4,110	0,0008	0,0007
	3	99,820	0,006	0,006	4,993	0,001	0,001
	4	99,817	0,007	0,005	6,855	0,001	0,001
	5	99,800	0,006	0,006	9,024	0,002	0,001

5. SONUÇLAR VE ÖNERİLER

Son olarak geliştirilen KSMR ile önerilen DMR yöntemlerinin etkinliği önceden belirlenen modellerdeki parametreleri yakın tahmin eden modellerin belirlenmesi gözlem setinde aykırı gözlemler oluşturularak denenmiştir. Değerlendirme 3 temel açıdan yapılmıştır. Modeldeki sabit terimin tahminine, aykırı gözlem sayısına ve gözlem sayısına bağlı olarak karşılaştırılmaları değerlendirilmiştir.

Göründüğü gibi KSMR metodu sabit terimi kesinlikle doğru tahmin edememekte ve sıfır olarak belirlemektedir. Bunun en önemli sebebi bu yöntem her gözlemi her aşamada adım adım orijine (0,0) noktasına taşımasıdır. DMR yönteminde ise gözlem sayısı 20 veya daha büyükse aykırı gözlemlerin sayısına bağlı olarak temel modeldeki sabit terim daha iyi tahmin edilmiştir.

Her iki yöntemde aykırı gözlem sayısı artıkça temel modeldeki parametre tahminlerinde daha çok sapmaların yaşandığı gözlemlenmiştir. Ancak hatalı tahmin miktarı DMR'de KSMR'e göre her durumda daha düşüktür.

Her iki yöntemde de gözlem sayısı arttıkça parametre tahminleri daha doğru elde edilmiştir.

Sonuç olarak sabit terimin tahmin edilebilmesi ve aynı koşullarda parametre tahminlerinin daha doğru olması DMR yönteminin daha etkin bir sağlam regresyon yöntemi olduğunu göstermektedir.

İlerleyen zamanlarda bu yöntemin etkinliğinin daha farklı modellerde incelemesi yapılmalıdır.

KAYNAKLAR

Adnan, R., Mohamad, M.N. ve Setan, H., 2003, Multiple outliers detection producers in linear regression, *Matematika*, 19 (1), 29-45.

Arthanari, T.S. ve Dodge, Y., 1981, Mathematical programming in statistics, John Wiley & Sons, New York, USA.

Croux, C. ve Haesbroeck, G., 1999, Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, *Journal of Multivariate Analysis,* 71 (2), 161-190.

Cummins, D.J. ve Andrews, C., 1995, Iteratively reweighted partial least squares: A performance analysis by monte carlo simulation, *Journal of Chemometrics*, 9 (6), 489-507.

Dempster, A.P., Laird, N.M. ve Rubin, D.B., 1980, Krishnaiah P.R., Multivariate analysis V, *North-Holland,* Amsterdam.

Draper, N., ve Smith, H., 1981, Applied Regression Analysis, 2nd Edition, John Wiley & Sons, New York, USA.

Filzmoser, P., Marona, R. ve Werner, M., 2008, Outlier identification in high dimensions, *Computational Statistics & Data Analysis* , 52, 1694-1711.

Fauconnier, C. ve Haesbroeck, G., 2009, Outliers Detection with The Minimum Covariance Determinant Estimator in Practice, *Statistical Methodology*, 6 (4), 363-379.

Huber, P.J, 1981, Robust statistics, *Wiley*, New York.

Hubert, M. ve Vanden Branden, K., 2003, Robust methods for partial least squares regression, *Journal of Chemometrics,* 17, 537-549.

Koenker, R. ve Bassett, G., 1978, Regression quantiles, *Econometrica*, 146 (1), 33-50.

Liu, H., Shah, S. ve Jiang, W., 2004, On-line outlier detection and data cleaning, *Computers & Chemical Engineering*, 28, 1635-1647.

Pena, D. ve Yohai, V., 1999, A fast procedure for outlier diagnostics in large regression problems, *Journal of the American Statistical Association*, 94 (446), 434-445.

Penny, I., 1996, Appropriate critical values when testing for a single multivariate outlier by using the mahalanobis distance, *Royal Statistical Society*, 45 (1), 73-81.

Rousseeuw, P. J. ve Leroy, A. N., 1987, Robust regression and outlier detection, *Wiley*, New York.

Rousseeuw, P.J. ve Van Zomeren, B.C., 1990, Unmasking multivariate outliers and leverage points, *Journal of the American Statistical Association*, 85 (411), 633-639.

Saltenis, V., 2004, Outlier detection based on the distribution of distances between bata points, *İnformatica*, 15 (3), 399-410.

Serneels, S., Croux, C., Filzmoser, P. ve Van Espen, P.J., 2005, Partial robust m-regression, *Chemometrics and Intelligent Laboratory Systems*, 79 (1-2), 55-64.

Wakeling, I.N. ve MacFie, 1992, A robust PLS procedure, *Journal of Chemometrics*, 6 (4), 189-198.

Willems, G., Joe, H. ve Zamar, R., 2009, Diagnosing multivariate outliers detected by robust estimators, *Journal of Computational and Graphical Statistics*, 18 (1), 73-91.

Wold H. ve Krishnaiah P. R., 1973, Multivariate analysis , *Academic Pres*, New York, 383-407.

Wu, J.W. ve Lee, W.C., 2006, Computational algorithm of least absolute deviation method for determining number of outliers under normality, *Applied Mathematics and Computation*, 175, 609-617.

EKLER

EK-1 SPSS yazılımı ile tek değişkenli doğrusal regresyonda Monte-Carlo simülasyon programı:

```
set mxloop=100000.
matrix.
compute densay=5000.              /* Tekrar sayısı         */
compute n=10.                     /* Veri sayısı           */
compute kgsay=4.                  /* Aykırı gözlem sayısı  */
compute kgbekk1=make(densay,1,0).
compute kgbekk2=make(densay,1,0).
compute dmbekk1=make(densay,1,0).
compute dmbekk2=make(densay,1,0).
loop ds=1 to densay.
compute d1=uniform(n,1).
compute kokx2=d1*100.
compute d2=uniform(n,1).
compute d3=uniform(n,1).
compute e=4*(sqrt(-2*ln(d2))&*sin(2*3.14*d3)).
compute kk=0.
compute aa=1.
compute bb=n.
compute usays=kgsay.
compute torba={1:bb}.
compute issay=make(usays,1,0).
loop i=1 to usays.
compute say=trunc((bb-(i-1))*uniform(1,1))+1.
compute issay(i)=torba(say).
compute x=make(1,bb-1,0).
loop j=1 to bb-(i-1).
 do if (torba(j)<>issay(i)).
compute kk=kk+1.
 compute x(kk)=torba(j).
else.
end if.
end loop.
compute torba=x.
compute kk=0.
end loop.
loop boz=1 to kgsay.
compute a=uniform(1,1).
compute goz=issay(boz).
compute kayma=50+4*(sqrt(-2*ln(uniform(1,1)))&*sin(2*3.14*uniform(1,1))).
do if(uniform(1,1)<0.5).
    compute e(goz)=e(goz)-kayma.
else.
    compute e(goz)=e(goz)+kayma.
end if.
end loop.
compute koky=10+2*kokx2+e.
compute kokx1=make(n,1,1).
compute kokx={kokx1,kokx2}.
compute kokveri={kokx,koky}.
save kokveri/outfile=*.

/*    KISMİ-SAĞLAM M-REGRESYON   */

compute x1=kokx1.
compute x2=kokx2.
compute x=kokx.
compute y=koky.
compute veri=kokveri.
compute n=nrow(x).
compute p=ncol(x).
compute kgbekk=ginv(t(x)*x)*t(x)*y.      /*En küçük kareler*/
```

```
compute hata=y-x*kgbekk.
compute ekkhko=mssq(hata)/(n-p).
compute betao=kgbekk.
compute a=y.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medy=(a(n/2)+a(n/2+1))/2.
compute r=y-medy.                    /*R */
compute a=r.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medr=(a(n/2)+a(n/2+1))/2.
compute fark=abs(r-medr).
compute a=fark.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute sigma=(a(n/2)+a(n/2+1))/2.   /* SIGMA */
compute z=r/sigma.
compute zb4=abs(z/4).
compute wri=1/((1+zb4)&**2).         /* WRİ*/
loop k=1 to 100.                     /* ADIM SAYISI*/
compute a=x2.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medx=(a(n/2)+a(n/2+1))/2.    /* medx */
compute t1=abs(x2-medx).
compute a=t1.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medt1=(a(n/2)+a(n/2+1))/2.
compute t2=abs(t1/(4*medt1)).
compute wxi=1/((1+t2)&**2).         /* WXİ  */
compute wi=wri&*wxi.                /* Wİ   */
```

```
compute kwi=sqrt(wi).
compute x2=kwi&*x2.
compute y=kwi&*y.
compute x={x1,x2}.
compute veri={x,y}.
compute beta=ginv(t(x)*x)*t(x)*y.
compute hata=y-x*beta.
compute hko=mssq(hata)/(n-p).
compute kriter=msum(abs(betao-beta)).
/*durdurma kuralı*/
do if (kriter<0.01).
  break.
end if.
compute betao=beta.
compute a=hata.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute mhata=(a(n/2)+a(n/2+1))/2.
compute mefark=abs(hata-mhata).
compute a=mefark.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute sigma=(a(n/2)+a(n/2+1))/2.
compute z=hata/sigma.
compute z1=abs(z/4).
compute wri=1/((1+z1)&**2).
end loop.
compute kgbekk=beta.
compute kgbhko=hko.
compute kgbekk1(ds)=kgbekk(1).
compute kgbekk2(ds)=kgbekk(2).
/* DİK M-REGRESYON  */
compute sapsay=0.
compute x2=kokx2.
compute y=koky.
compute x1=kokx1.
compute x={x1,x2}.
compute n=nrow(x).
compute p=ncol(x).
compute veri={x,y}.
compute dmbekk=ginv(t(x)*x)*t(x)*y.                  /* EKK  */
compute betao=dmbekk.
compute hata=y-x*dmbekk.
compute khko=(mssq(hata)/(n-p)).
loop m=1 to 100.                                     /* DONGU */
compute dikuzak=abs(hata)/sqrt((1+mssq(dmbekk(2)))).  /* DİKUZAK */
compute k=(dikuzak/msum(dikuzak))&*dikuzak.          /* KAYDIRMA */
compute b1=dmbekk(1,:).
compute b2=dmbekk(2,:).
compute A={b2,-1}.
compute lkare=k&*k/mssq(A).
compute l=sqrt(lkare).
loop i=1 to n.
   do if(hata(i)>0).
```

```
    compute l(i)=l(i).
    else.
    compute l(i)=-l(i).
   end if.
end loop.
compute x(:,2)=A(1,1)*l+x(:,2).
compute y=A(1,2)*l+y.
compute veri={x,y}.
compute dmbekk=ginv(t(x)*x)*t(x)*y.
compute hata=y-x*dmbekk.
compute dmhko=(mssq(hata)/(n-p)).
compute kriter=msum(abs(betao-dmbekk)).
/*durdurma kuralı*/
compute sapsay=msum(k>0.07).
compute dkural=(kriter<0.01).
do if (dkural=1).
  break.
end if.
compute betao=dmbekk.
end loop.
compute dmbekk1(ds)=dmbekk(1).
compute dmbekk2(ds)=dmbekk(2).
end loop.
compute vkgmbo=mssq(kgbekk1-10)/densay.
compute vkgmb1=mssq(kgbekk2-2)/densay.
compute vdmbo=mssq(dmbekk1-10)/densay.
compute vdmb1=mssq(dmbekk2-2)/densay.
print vkgmbo.
print vkgmb1.
print vdmbo.
print vdmb1.
end matrix.
```

EK-2 SPSS yazılımı ile çoklu doğrusal regresyonda Monte-Carlo simülasyon programı:

```
set mxloop=100000.
matrix.
compute densay=5000.      /* Tekrar sayısı  */
compute n=10.             /* Gözlem sayısı  */
compute kgsay=1.          /* Aykırı gözlem sayısı */
compute kgbekk1=make(densay,1,0).
compute kgbekk2=make(densay,1,0).
compute kgbekk3=make(densay,1,0).
compute dmbekk1=make(densay,1,0).
compute dmbekk2=make(densay,1,0).
compute dmbekk3=make(densay,1,0).
loop ds=1 to densay.
compute kokx2=uniform(n,1)*100.
compute kokx3=uniform(n,1)*100.
compute d2=uniform(n,1).
compute d3=uniform(n,1).
compute e=4*(sqrt(-2*ln(d2))&*sin(2*3.14*d3)).
compute kk=0.
compute aa=1.
compute bb=n.
compute usays=kgsay.
compute torba={1:bb}.
compute issay=make(usays,1,0).
loop i=1 to usays.
compute say=trunc((bb-(i-1))*uniform(1,1))+1.
compute issay(i)=torba(say).
compute x=make(1,bb-1,0).
loop j=1 to bb-(i-1).
 do if (torba(j)<>issay(i)).
compute kk=kk+1.
 compute x(kk)=torba(j).
else.
end if.
end loop.
compute torba=x.
compute kk=0.
end loop.
loop boz=1 to kgsay.
compute a=uniform(1,1).
compute goz=issay(boz).
compute kayma=50+4*(sqrt(-2*ln(uniform(1,1)))&*sin(2*3.14*uniform(1,1))).
do if(uniform(1,1)<0.5).
    compute e(goz)=e(goz)-kayma.
else.
    compute e(goz)=e(goz)+kayma.
end if.
end loop.
compute koky=10+2*kokx2-kokx3+e.
compute kokx1=make(n,1,1).
compute kokx={kokx1,kokx2,kokx3}.
compute kokveri={kokx,koky}.
/*    KISMİ-SAĞLAM M-REGRESYON   */
compute x1=kokx1.
compute x2=kokx2.
compute x3=kokx3.
compute x=kokx.
compute y=koky.
compute veri=kokveri.
compute n=nrow(x).
compute p=ncol(x).
compute kgbekk=ginv(t(x)*x)*t(x)*y.      /*EKK*/
compute hata=y-x*kgbekk.
compute ekkhko=mssq(hata)/(n-p).
compute betao=kgbekk.
compute a=y.
```

```
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medy=(a(n/2)+a(n/2+1))/2.
compute r=y-medy.                    /*R */
compute a=r.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medr=(a(n/2)+a(n/2+1))/2.
compute fark=abs(r-medr).
compute a=fark.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute sigma=(a(n/2)+a(n/2+1))/2.      /* SİGMA       */
compute z=r/sigma.
compute zb4=abs(z/4).
compute wri=1/((1+zb4)&**2).            /* WRİ         */
loop k=1 to 100.                        /* ADIM SAYISI */
compute a=x2.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medx2=(a(n/2)+a(n/2+1))/2.      /* medx2 */
compute ti1=abs(x2-medx2).
compute a=x3.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute medx3=(a(n/2)+a(n/2+1))/2.      /* medx3 */
compute ti2=abs(x3-medx3).
compute nuz=sqrt(ti1&*ti1+ti2&*ti2).
compute a=nuz.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
```

```
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute mednuz=(a(n/2)+a(n/2+1))/2.          /* mednuz */
compute t22=abs(nuz/(4*mednuz)).
compute wxi=1/((1+t22)&**2).                 /* WXİ    */
compute wi=wri&*wxi.                         /* Wİ     */
compute kwi=sqrt(wi).
compute x2=kwi&*x2.
compute x3=kwi&*x3.
compute y=kwi&*y.
compute x={x1,x2,x3}.
compute veri={x,y}.
compute beta=ginv(t(x)*x)*t(x)*y.
compute hata=y-x*beta.
compute hko=mssq(hata)/(n-p).
compute kriter=msum(abs(betao-beta)).
/*durdurma kuralı*/
do if (kriter<0.01).
  break.
end if.
compute betao=beta.
compute a=hata.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute mhata=(a(n/2)+a(n/2+1))/2.
compute mefark=abs(hata-mhata).
compute a=mefark.
loop i=1 to n-1.
loop j=i+1 to n.
 do if(a(j)<a(i)).
   compute kutu=a(i).
   compute a(i)=a(j).
   compute a(j)=kutu.
 end if.
end loop.
end loop.
compute sigma=(a(n/2)+a(n/2+1))/2.
compute z=hata/sigma.
compute z1=abs(z/4).
compute wri=1/((1+z1)&**2).
end loop.
compute kgbekk=beta.
compute kgbhko=hko.
compute kgbekk1(ds)=kgbekk(1).
compute kgbekk2(ds)=kgbekk(2).
compute kgbekk3(ds)=kgbekk(3).
/* DİK M-REGRESYON  */
compute sapsay=0.
compute x2=kokx2.
compute x3=kokx3.
compute y=koky.
compute x1=kokx1.
compute x={x1,x2,x3}.
compute n=nrow(x).
compute p=ncol(x).
compute veri={x,y}.
compute dmbekk=ginv(t(x)*x)*t(x)*y.      /* EKK  */
compute betao=dmbekk.
```

```
compute hata=y-x*dmbekk.
compute khko=(mssq(hata)/(n-p)).
loop m=1 to 100.                               /* DONGU */
compute dikuzak=abs(hata)/sqrt((1+(dmbekk(2))**2+(dmbekk(3))**2))).
/* DİKUZAK */
compute k=(dikuzak/msum(dikuzak))&*dikuzak.  /* KAYDIRMA */
compute b1=dmbekk(1,:).
compute b2=dmbekk(2,:).
compute b3=dmbekk(3,:).
compute A={b2,b3,-1}.
compute lkare=k&*k/mssq(A).
compute l=sqrt(lkare).
loop i=1 to n.
   do if(hata(i)>0).
    compute l(i)=l(i).
    else.
    compute l(i)=-l(i).
   end if.
end loop.
compute x(:,2)=A(1,1)*l+x(:,2).
compute x(:,3)=A(1,2)*l+x(:,3).
compute y=A(1,3)*l+y.
compute veri={x,y}.
compute dmbekk=ginv(t(x)*x)*t(x)*y.
compute hata=y-x*dmbekk.
compute dmhko=(mssq(hata)/(n-p)).
compute kriter=msum(abs(betao-dmbekk)).
/*durdurma kuralı*/
compute dkural=(kriter<0.01).
do if (dkural=1).
  break.
end if.
compute betao=dmbekk.
end loop.
compute dmbekk1(ds)=dmbekk(1).
compute dmbekk2(ds)=dmbekk(2).
compute dmbekk3(ds)=dmbekk(3).
end loop.
compute vkgmbo=mssq(kgbekk1-10)/densay.
compute vkgmb1=mssq(kgbekk2-2)/densay.
compute vkgmb2=mssq(kgbekk3+1)/densay.
compute vdmbo=mssq(dmbekk1-10)/densay.
compute vdmb1=mssq(dmbekk2-2)/densay.
compute vdmb2=mssq(dmbekk3+1)/densay.
print vkgmbo.
print vkgmb1.
print vkgmb2.
print vdmbo.
print vdmb1.
print vdmb2.
end matrix.
```

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	**:**	Bengü Ocak
Uyruğu	**:**	T.C.
Doğum Yeri ve Tarihi	**:**	Ankara 25/10/1986
Telefon	**:**	05056980871
Faks	**:**	
e-mail	**:**	ocak_beng@hotmail.com

EĞİTİM

Derece		**Adı, İlçe, İl**	**Bitirme Yılı**
Lise	:	Başkent Lisesi, Mamak, Ankara	2003
Üniversite	:	Selçuk Üniversitesi, Selçuklu, Konya	2008
Yüksek Lisans	:	Selçuk Üniversitesi, Selçuklu, Konya	2010
Doktora	:		

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görevi**
2007	Türkiye İstatistik Kurumu, Ankara	Stajyer
2007	T.C Vakıflar Bankası Kızılay Şubesi, Ankara	Stajyer

UZMANLIK ALANI

İstatistik

YABANCI DİL

İngilizce

YAYINLAR

Makale:
Semiz M., Ocak B., "An Alternative Agreement Statistics with Linear Weight Between Ordinal Categorical Measurements", *Hacettepe Journal of Mathematics and Statistics*, Vol.39(1), 97-107 (2010). (SCI-Expanded)

Kitap :
Bilgisayarda İstatistik Uygulamaları (SPSS), 2008.