ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:1 - Sayı/No: 1 : 77-84 (2000)

ARAŞTIRMA MAKALESİ/**RESEARCH ARTICLE**

SAĞLAM Cp KRİTERİNİN UYGULANMASI ÜZERİNE BİR ÇALIŞMA

Meral (CANDAN) ÇETİN [1] ve Aydın ERAR [1]

ÖZ

Çoklu doğrusal regresyonda, aykırı değerler (outliers), değişken seçiminde büyük bir etkiye sahiptirler. Aykırı değerlerin bu etkisi, Huber-türü, Mallows-türü vb. ağırlık fonksiyonları ile giderilir. Bu çalışmada, değişken seçimi üzerinde durularak, aykırı değer varlığında klasik Cp kriteri, Huber ve Mallows-türü ağırlık fonksiyonlarına dayalı sağlam (robust) Cp kriterleri ile karşılıştırılacaktır.

Anahtar Kelimeler: Sağlam model seçimi, Mallows'un Cp kriteri, sağlam regresyon, M-kestiriciler

A STUDY ON THE Cp CRITERIA APPLICATIONS IN ROBUST REGRESSION

ABSTRACT

In multiple linear regression, outliers are very efficient on the variable selection. This effects of the outliers are removed by the Hubers-type and Mallows-type etc. weight functions. In this paper, the classical Cp criteria which depends on the Hubers-type and Mallows-type weight functions will be compared in the presence of outliers.

Key words: Robust models selection, Mallow's Cp criteria, robust regression, M-estimators.

1. GİRİŞ

Y, nx boyutlu yanıt vektörü, X, nx (k+1) boyutlu girdi matrisi, β, (k+1) x1 boyutlu katsayılar vektörü, ε, nx1 boyutlu hata vektörü olmak üzere,

$$Y=X\beta+\varepsilon$$

ile verilen çoklu doğrusal regresyon çözümlemesinde verilen en iyi biçimde tanımlanacak gerekli değişkenlerin modele seçilmesi, modele katkısı önemsiz değişkenlerin çıkartılması "değişken seçimi" ya da "en iyi alt küme seçimi" olarak bilinir.

Çoklu regresyonda modelden değişik çıkarmak ya da modele değişken eklemek için zaman zaman adımsal yöntemler kullanılır. Bu amaçla kullanılan "ileriye doğru seçim" ve "geriye doğru seçim" ya da bu yöntemlerin birleşiminden oluşan "adımsal teknik" yöntemlerinde farklı modellere ulaşabileceği bilinir. Her bir yöntem farklı I. tür hata oranına sahip olacağından, yöntemlerin tam bir karşılaştırmasını yapmak zordur.

Çok iyi bilindiği gibi, istatistiksel analizdeki temel zorluk uygun modelin seçimi, model boyutunun belirlenmesi ve kestirimidir. Model belirlemenin ana amaçları, parametre kestirimi ya da bağımlı değişkenin önceden kestirimidir (prediction).

Parzen'e göre istatistiksel veri modellemesi, gerçek modeli bilmeksizin (ya da bilinebilir) veri kümesine uyan modelleri araştıran istatistiksel bir alandır. Son yıllarda model seçimi ya da model belirlenmesi kavramının verilmesinin gerekliliği kabul görmektedir ve farklı boyutlu aday modeller arasından uygun bir model seçimi kriteri ile "en iyi kestirim" in nasıl yapılacağı problemi ortaya çıkmaktadır.

1 Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü, Beytepe-Ankara, **E.mail:** meral@hacettepe.edu.tr.
Geliş: 27 Ekim 1999; **Düzeltme;** 29 Ocak 2000; **Kabul:** 2 Mart 2000.

Altküme seçimi, çok sayıda açıklayıcı değişkenin kullanılması ile ortaya çıkan artan kestirim varyansı ve artan yan miktarıdır. Değişken seçim kriterinin bir çoğu, kestirimin hata kareler ortalamısının basit bir fonksiyonudur.

2. SAĞLAM MODEL SEÇİMİ

Sağlam (robust) istatistiklerin ana amaçlarından biri, model dağılımlarından küçük sapmalara karşı duyarsız olan yeni istatistiksel yöntemler bulmaktır. Son 20 yıllık dönemde En Küçük Kareler (EKK) kestiricilerine karşı pek çok sağlam alternatifler geliştirilirken, model seçim problemleri ihmal edilmiştir. Klasik yöntemler aykırıdeğer varlığında ya da normalden hafifçe sapmalarda sağlam parametre kestirimi vermemektedir. Bu durumda klasik değişken seçim yöntemlerini uygulayamayacağımızdan sağlam değişken seçim yöntemlerine gerek duyulur. Klasik seçim yöntemlerinin sağlamlaştırılmasına ilişkin bazı çalışmalar vardır: Ronchetti (1982,1985), Akaike bilgi kriterlerinin sağlam bir versiyonunu önermiş, Hampel (1983) bu kriterin değişik bir versiyonunu vermiştir.Benzer bir düşünce Martin (1980) tarafından otoregresif modeller için kullanılmıştır. Hurvich ve Tsai (1990), L_1 en küçük mutlak sapmalar regresyonu için sağlam model seçim kriterini ve Machado (1993), Schwarz kriterinin sağlam bir türünü önermişlerdir. Ronchetti, Field ve Blanchard (1997), çapraz-geçerlilik (cross-validation) yöntemini kullanarak model seçimi ile ilgili sağlam bir algoritma vermişlerdir.

Regresyon ve otoregresif modellerde uygulanan seçim yöntemlerinin başında Akaike bilgi kriteri gelir (Akaike, 1974). Son 10 yılda özellikle değişken seçiminde Akaike bilgi ölçütünün kullanımı üzerine bir yoğunluk vardır. Akaike kriteri ile ilgili çalışmaların çoğu normal dağılımlı veri kümesinde ve büyük örneklemler üzerinde incelenmiştir. Küçük örneklemli normal doğrusal regresyon modelleri için Akaike kriteri yanlı olabileceğinden, bu kriterin düzeltilmiş bir versiyonu (AICC) verilir (Hurvich ve Tsai, 1989). Ayrıca değişken seçim amacıyla AIC'nın değişik versiyonları ve farklı seçim kriterleri üzerinde durulmaktadır. Ancak burada Mallows'un C_P kriterinin sağlam bir versiyonu verilecektir.

3.SAĞLAM Cp KRİTERİ

Maslow'un kriteri, değişken ya da model seçiminde kullanılan uygun bir kriterdir. Cp kriteri,

$$C_p = \frac{AKT_p}{\hat{\sigma}^2} - n + 2p$$

biçiminde tanımlanır. Burada $\hat{\sigma}^2$, tam modelden hesaplanan artık varyansı; AKTp, p boyutlu denklemin artık kareler toplamı ve p, parametre sayısıdır. Cp istatistiği önkestirim için yeterlilik ölçüsünün bir kestirimidir. Seçim, p'ye yakın Cp'li modeller bulunarak yapılır. Cp istatistiği EKK kestiricisine dayalı olduğundan, bu istatistik aykırı değerlere ve hata dağılımının normallik varsayımından sapmalara karşı oldukça duyarlıdır. Eğer p altkümesi doğru ise, Cp kriteri p'ye yakın ya da p'den küçük olacaktır (Ryan,1998).

$Y_i = X_i\beta + \varepsilon_i$ doğrusal modelini düşünelim. Burada x_i', X matrisinin i. satırıdır.

X matrisinin satırları bir F dağılımından, aynı dağılımlı bağımsız gözlem olsun. F dağılımı $\varepsilon = y - x^T\beta$'nin bileşik dağılımını gösterir ve X'ler sabittir.

p parametreli bir P modelinin $\hat{\beta}_P$ kestiricisi, bazı $\eta(x,r)$ fonksiyonları için,

$$\Sigma\eta(x_i, y_i - x_i^T\beta)x_i = 0 \tag{1}$$

denkleminin çözümüdür. (Hampel.vd,1986).

(1) eşitliğinden elde edilen kestirici $T_p(F)$ fonksi-yoneli ile gösterilsin. $T_p(F)$,

$$E_F[\eta(x,y - X^T T_P(F))] = 0 \tag{2}$$

eşitliğini sağlar (Hampel, 1986) ve aşağıdaki koşulların geçerliliği varsayılır (Ronchetti, ve Staudte,1994).

Koşul 1, Bir P altkümesi varsa, bu altkümenin bir kestiricisi için etkinlik fonksiyonu açılımı elde edilebilir ve

$$\hat{\beta}_P - \beta = \frac{1}{n}\sum_{i=1}^{n} IF(x_i, y_i; T_P, F) + R_n \tag{3}$$

eşitliği sağlanır. Burada etkinlik fonksiyonu,

$$IF(x,y;T_p,F) = M^{-1}\eta(x,y - x'T_p(F))x \tag{4}$$

biçiminde verilir ve $M = E[\eta'(x,\varepsilon)xx']$ biçimindedir.

η',η fonksiyonun türevidir. Burada $nE\|R_n\|^2 \to 0$ varsayılır.

Koşul 2, $w(x,r)=\eta(x,r)/r$ ağırlık fonksiyonu olduğu varsayılır ve bu fonksiyonun birinci türevi $w'(x,r)$ ve ikinci türevi $w''(x,r)$ ile gösterilir.Burada, $r = y - \hat{y}$'dır.

Koşul 3, $w(x,r)$ fonksiyonu 0 ile 1 arasında sınırlıdır. w' ve w'' 'ünün sınırlı olduğu varsayılır. Koşul 2, kestiricinin asimtotik olarak normal olduğu anlamına gelir:

$$n^{1/2}\left[\hat{\beta}_P - \beta\right] \xrightarrow{d} N\left(0, M^{-1}QM^{-1}\right),$$

$$nE\left[(\hat{\beta}_p - \beta)(\hat{\beta}_p - \beta)'\right] \rightarrow M^{-1}QM^{-1}$$

Burada, $Q = E[\eta^2 (x,\varepsilon)xx']$ olarak verilmiştir.

Koşul 2' den $f(\alpha) = w(x,y - x\alpha)$ fonksiyonunun Taylor açılımı,

$$f(\alpha) = f(\alpha_0) = w'(x,y - x\alpha_0)x'(\alpha-\alpha_0) +$$

$$\frac{w''(x,y-x'\alpha^*)}{2}(\alpha - \alpha_0)'xx'(\alpha - \alpha_0)$$

biçimindedir. $(\alpha < \alpha^* < \alpha_0)$.

$\alpha_0 = \beta$ ve $\alpha = \hat{\beta}_p$' e göre bu açılım uygulanırsa, ağırlık kestirimleri elde edilir:

$$\hat{w}_i = w(x_i,y_i-x_i'\beta) - w'(x_i,y_i - x_i'\beta)\, x_i'\, (\hat{\beta}_p-\beta) + \frac{w''(x_i,y_i-x_i'\beta^*)}{2}$$

$$(\hat{\beta}_p - \beta)'\, xx'(\hat{\beta}_p - \beta), \quad \hat{\beta}_p < \beta^* < \beta$$

Koşul 3 ise Huber ağırlık fonksiyonlu $\left(w(x_i, r_i) = \frac{\psi(r_i)}{r_i}\right)$ sağlam kestiriciler ya da Mallows türü ağırlık fonksiyonlu $(w(x_i, r_i) = v(x_i)\, \psi(r_i)/r_i)$ sağlam kestiriciler ile sağlanır. x_i 'ye bağlı faktör bazı k değerleri için $|v(x)| \le k/\|x_i\|^2$ koşulunu sağlar (Hampel. vd., 1986,).

Böylece elde edilen sonuçlar yüksek bozulumlu M kestiricilere uygulanabilir. Çünkü bu kestiriciler, yukarıda verilen iteratif olarak elde edilen M-kestiriciler ile aynı asimtotik özelliklere sahiptirler (Coakley ve Hettmansperger, 1993),

(1) eşitliğinde verilen $\hat{\beta}_p$ - M kestiricisinin ağırlık fonksiyonu,

$$\hat{w}_i = w\left(x_i, y_i - x_i'\hat{\beta}_p\right) = \eta\left(x_i, y_i - x_i'\hat{\beta}_p\right) / \left(y_i - x_i'\hat{\beta}_p\right)$$

şeklindedir ve yeniden ölçeklenmiş ağırlıklı önkestirim hata kareler ortalaması

$$\Gamma_p = \frac{1}{\sigma^2} E\left[\sum \hat{w}_i^2 \left(\hat{y}_i - Ey_i\right)^2\right] \tag{5}$$

ile verilir (Ronchetti ve Staudte, 1994).

$\hat{y}_i = x_i'\hat{\beta}_p$, P altkümesi için uyum kestirim değeridir ve $E(y_i)$ de tam model altında beklenen değerdir. (5) eşitliğindeki ağırlıklar her modele göre değişmektedir. Ağırlıklandırma ile aykırı değerlerin etkisi azaltılır ve bu değerlerin Γ_p (yeniden ölçeklenmiş ağırlıklı önkestirim hata kareler ortalaması) üzerindeki etkisi sınırlanır; bunun beraberinde de model seçimi üzerindeki etkisi sınırlandırılır.

$\delta_i = \hat{y}_i - E(y_i) = x_i'\left(\hat{\beta}_p - \beta\right)$, i. önkestirimin hata terimi olmak üzere, $\sigma^2 \Gamma_p$ ifadesi,

$$\sigma^2\Gamma_p = \sum_i \mathrm{Var}\left(\hat{w}_i\delta_i\right) + \sum_i E\left[\left(\hat{w}_i\delta_i\right)\right]^2$$

biçiminde varyans ve yan^2 toplamı olarak yazılabilir:

Burada yan $E\left(\hat{w}_i\delta_i\right) = a_i - b_i$ olarak alınır. a_i , beklenen ağırlıklı hata terimi ve b_i , beklenen ağırlıklı artık terimidir:

$$a_i = E\left[\hat{w}_i\left(y_i - x_i'\beta_p\right)\right] = E\left(\hat{w}_i\, \varepsilon_i\right),$$

$$b_i = E\left[\hat{w}_i\left(y_i - \hat{y}_i\right)\right] = E\left(\hat{w}_i\, r_i\right),$$

$$V_p = \sum \mathrm{Var}\left(\hat{w}_i\, \delta_i\right),$$

$$B_p = \sum b_i^2 = \sum_i \left[E\left(\hat{w}_i\, r_i\right)\right]^2,$$

$$A_p = \sum a_i^2 = \sum_i \left[E\left(\hat{w}_i\, \varepsilon_i\right)\right]^2,$$

$$(AB)_p = \sum_i a_i b_i$$

olarak tanımlandıktan sonra,

$$\sigma^2 \Gamma_p = V_p + B_p - 2(AB)_p + A_p \tag{6}$$

eşitliği elde edilir.

Ağırlıklı artık kareler toplamının (W_p) merkezleştirilmiş ve yeniden ölçeklenmiş kestirim değeri, Γ_p nin iyi bir kestiricisi olabilir. Burada,

$$W_p = \sum_i \hat{w}_i r_i^2 = \sum_i \hat{w}_i^2 \left(y_i - \hat{y}_i\right)^2$$

olarak elde edilir.

Bu eşitliğin sağ tarafının beklenen değeri, varyans ve yan^2 terimlerinin toplamı biçiminde yazılabilir:

$$E(W_p) = U_p + B_p$$

Burada, $U_p = \sum \mathrm{Var}\left(\hat{w}_i r_i\right)$ ve B_p yukarıda verildiği gibidir. (6) eşitliğinde B_p yerine $E(W_p) - U_p$ yazılırsa,

$$\sigma^2 \Gamma_p = E(W_p) - U_p + V_p - 2(AB)_p + A_p$$

elde edilir. Gerekli kısaltma ve düzenlemelerden sonra C_p'nin sağlam bir versiyonu,

$$RC_p = \frac{W_p}{\hat{\sigma}^2} - (U_p - V_p)$$

şeklinde elde edilir (Ronchetti ve Staudte, 1994).

Burada $\hat{\sigma}^2$, tam model üzerinden σ^2'nin tutarlı ve sağlam bir kestiricisidir:

$$\hat{\sigma}^2 = W_{tam} / U_{tam}$$

U_p ve V_p sabit değerleri, Ronchetti ve Staudte tarafından,

$$U_p - V_p \approx nE\left(\psi_c^2\right) - p\,\frac{1}{E(\psi'_c)}\left\{-3E\psi_c^2 + 2E(\psi'_c)\,E(\psi'_c) + 3\,\frac{E(\psi'_c)}{E(\psi'_c)}\,E\left[\left(\frac{\psi_c}{\varepsilon}\right)^2\right]\right\}$$

eşitliğinden elde edilmiştir. Bu eşitlik Mallows-türü ağırlıklar için,

$$U_p - V_p \approx E\left(\psi^2\right)tr\left(V^2\right) - 2tr\left(N_n M_n^{-1}\right) + tr\left(L_n \textstyle\sum_n\right)$$

olarak tanımlanır.

$w(x, r) = v(x)\ \psi(r) / r$ ile verilen Mallows-türü ağırlık fonksiyonu için köşegen elemanı v_{ii} olan V matrisi tanımlanır. Burada,

$$N_n = E\left[\{\psi(\varepsilon)\}^2\ \psi'(\varepsilon)\right] X^T V^3 X/n,$$

$$L_n = E\left[\{\psi'(\varepsilon)\}^2 + 2\psi'(\varepsilon)\frac{\psi(\varepsilon)}{\varepsilon} - 3\left\{\frac{\psi(\varepsilon)}{\varepsilon}\right\}^2\right] X^T V^3 X/n,$$

$$M_n = E\left[\psi'(\varepsilon)\right] X^T V\ X/n,$$

ve $\sum_n = M_n^{-1} Q_n M_n^{-1}/n$, asimtotik kovaryans matrisidir (Staudte ve Sheather, 1990).

Sağlam C_p kriterinde, V_p ye yakın ya da V_p'den küçük RC_p değerine sahip modeller tercih edilir. Huber türü kestiriciler için V_p , P altkümesinin p boyutlu sabit bir çarpanıdır ($\approx$1) ve V_p yaklaşık olarak p'dir ($V_p \approx p$). Ağırlık fonksiyonu açıklayıcı değişkenlere dayalı kestiriciler (Mallows türü kestiriciler), V_p değeri aynı boyuttaki farklı modeller için değişebilir. Ağırlık değerleri 1 olduğunda W_p EKK'nın artık kareler toplamını verir ve $V_p = p$, $U_p = n - p$ ve RC_p Mallows'un C_p kriterine eşittir.

4. UYGULAMA

Bu bölümde değişken seçim kriterleri C_p ve RC_p arasında altküme seçimleri yönünden karşılaştırma yapabilmek için gerçek bir veri kümesi üzerinden seçilen örneklemler kullanılmıştır. Veriler Ankara Hastanesi Pediatri Bölümü'nden elde edilmiştir. Çalışmada bağımlı değişken kemik mineral yoğunluğu (KMY) olmak üzere 148 deneğe ilişkin cinsiyet, yaş, kilo ve günlük ortalama süt tüketim miktarları elde edilmiştir. Cinsiyet, yaş, kilo ve süt değişkenlerinin kemik mineral yoğunluğu üzerindeki etkisi araştırılmıştır.

Yapılan öninceleme ve artık incelemesinde aykırı değere rastlanmamıştır.

148 gözlem üzerinden yapılan regresyon çözümlemesinde doğrusal regresyon modeli,

$$\begin{aligned} KMY = {} & 0.126 + 0.01126*kilo + 0.01529*yaş \\ & \quad (0.04) \quad (0.002) \qquad\quad (0.005) \\ & + 0.03849*cins + 0.0001011*süt \\ & \quad (0.018) \qquad\quad (0.000) \end{aligned} \tag{7}$$

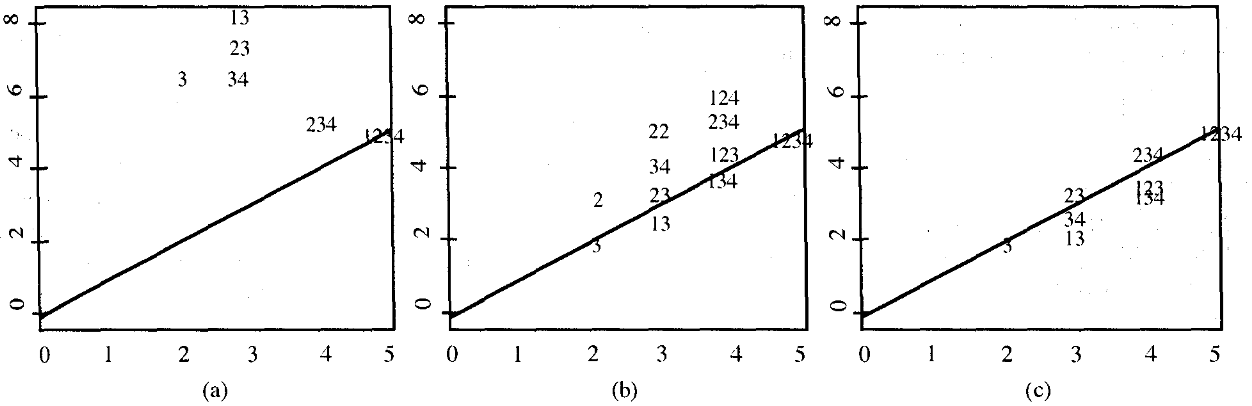
olarak elde edilmiştir. Parantez içindekiler katsayıların standart hatalarıdır. Bölüm 3'te verilen sağlam C_p kriterini küçük boyutlu örneklemlerde incelemek amacıyla, (7) denklemi gerçek model kabul edilerek 148 gözlemlik verim kümesinden 10, 14 ve 28 gözlemli rasgele örneklemler seçilmiştir. Bu rasgele örneklemlerin hepsinde bir ve iki tane olmak üzere yapay aykırı değerler yaratılarak değişken seçim kriterleri aykırı değer varlığında ve yokluğunda karşılaştırılmıştır. Çalışmada sağlam C_p kriterini elde etmek için S-Plus paket programı ve diğer hesaplamalar için SAS paket programı kullanılmıştır. Sağlam C_p kriteri Huber-türü ve Mallows-türü ağırlıklı kestiriciler için hesaplanmıştır.

C_p ve RC_p çizimleri her durum için elde edilmiştir. Ayrıca bu rasgele örneklemler üzerinden tek aykırı değer, iki aykırı değer ve aykırı değerin olmadığı durumlar gözönüne alınarak, Huber, Tukey ve Mallows ağırlık fonksiyonlu kestiriciler hesaplanmıştır. Programlar arası veri aktarımının yapılamaması, hesaplama zorluğu ve çizimler nedeniyle sonuçlar ancak üçer kez tekrarlanmıştır. Bu tekrarlamalar sonucunda, sonuçlarda belirgin bir farklılık ortaya çıkmamıştır. Bu nedenle, burada, yalnızca ilk tekrarın sonuçları verilmiştir.

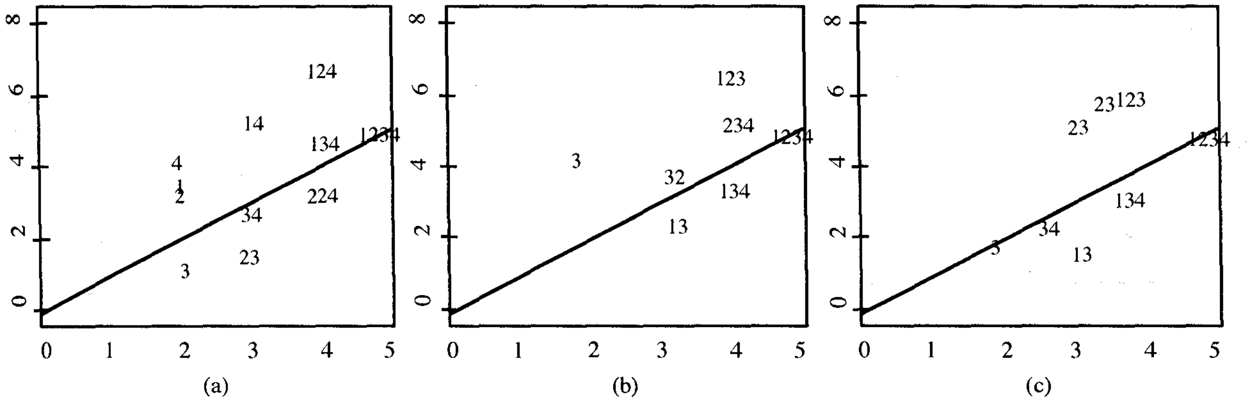
Örneklemler üzerinden çizilen PC_p-V_p ve klasik C_p-p çizimleri Şekil 1, 2, 3, 4, 5, 6, 7'de verilmiştir. Aykırı değer yokluğunda ve n=10, 14, 28 için bu çizimler Şekil 1'de verilmiştir. Bu durumda, ayrıca, Huber ve Mallows türü kestiriciler için aynı çizimler elde edilmiştir. Tek aykırı değer ve n=10, 14, 28 için çizimler Şekil 2, 3, 4'te gösterilmiştir. Şekil 2'den EKK'ya göre seçilen altkümeler sırasıyla 1234, 123, 234, 13, 24, 34 ve 3 iken Huber'e göre 1234, 34, 23, 134; Mallows'a göre 1234, 34, 3, 134 seçenek alt kümelerdir. Örneklem sayısı arttıkça Huber ve Mallowsun benzer sonuçlar verdiği Şekil 3 ve Şekil 4'ten görülmüştür. Yani bu çizimlere göre Huber ve Mallows türü kestiriciler benzer altkümeler vermekle birlikte EKK farklı alt kümeler vermektedir.

İki aykırı değer ve n=10, 14, 28 için çizimler Şekil 5, 6, 7'de gösterilmiştir. Bu çizimlerde de tek aykırı değer durumunda olduğu gibi, Huber ve Mallows kestiricileri aynı altkümeleri seçmekle birlikte Mallows kestiricilerin daha fazla altküme seçimine yöneldiği gözlenmiştir.

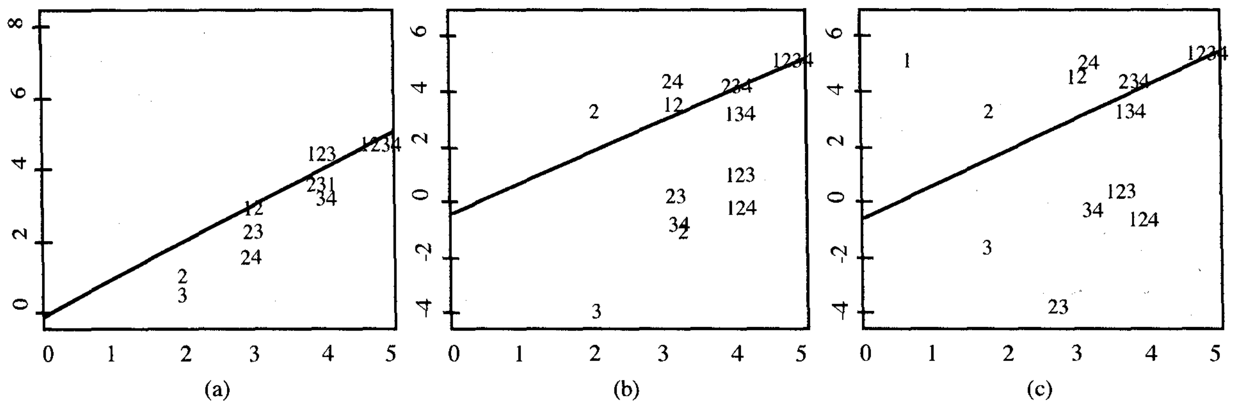
Bu çalışmada, ayrıca, örneklemlerden elde edilen kestiricilerle gerçek olarak kabul edilen (7). eşitlikteki



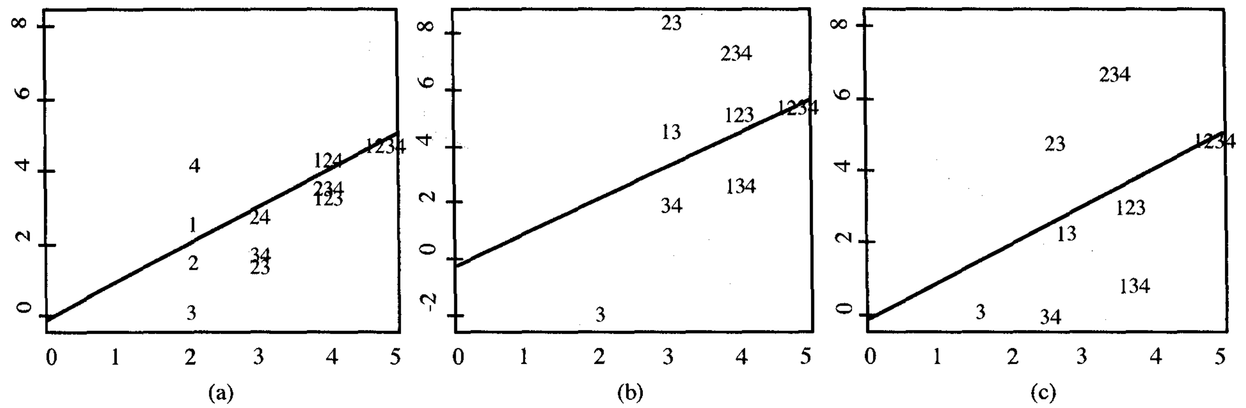
Şekil 1.



Şekil 2.

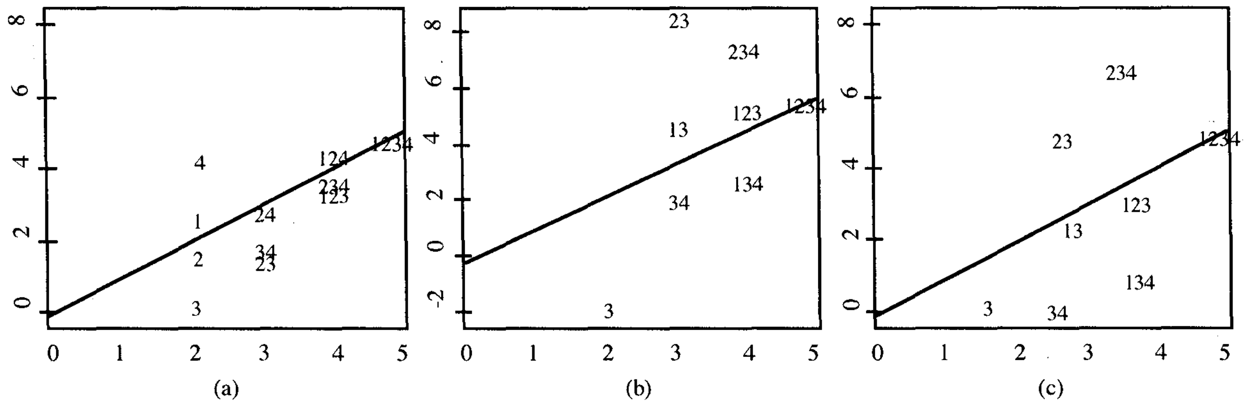


Şekil 3.

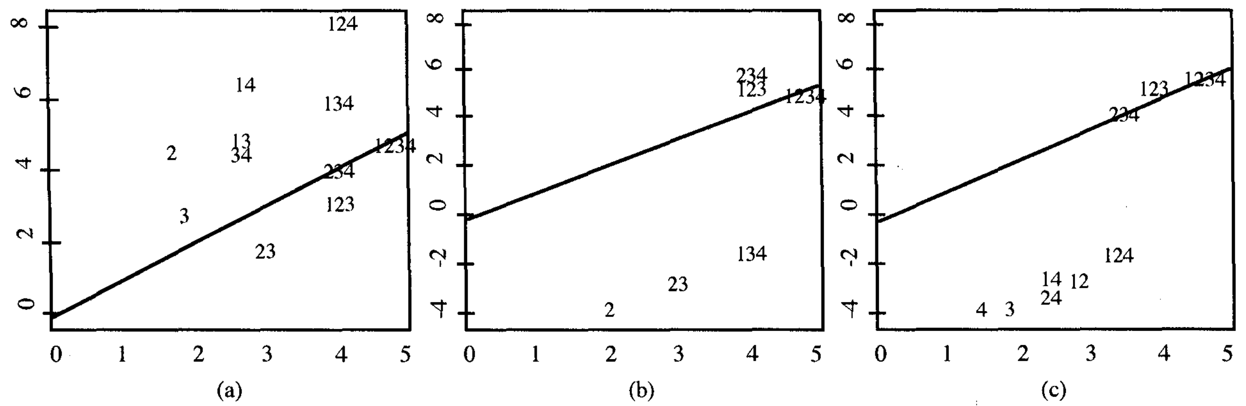


Şekil 4.

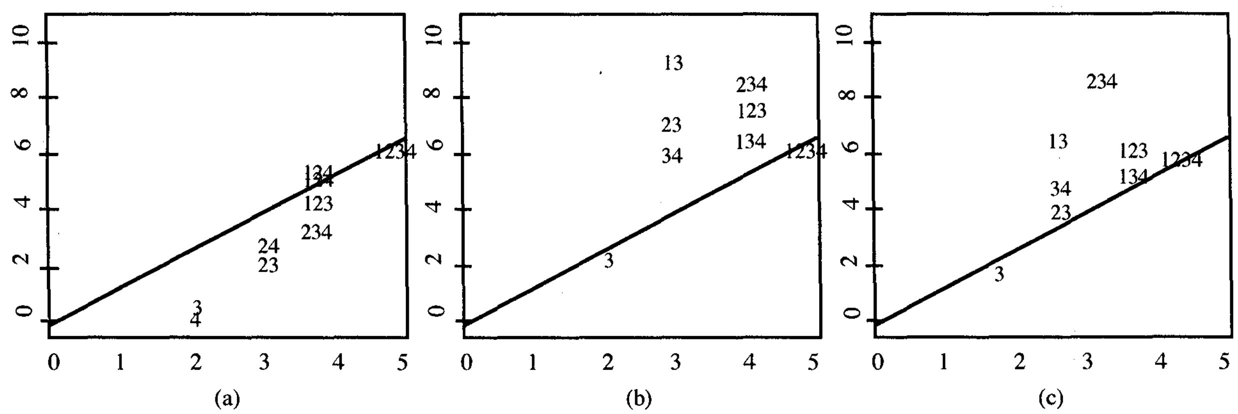
değerler karşılaştırılmıştır. Bu karşılaştırmada, özellikle seçimlerde uygun olarak görülen her altkümeler için, $M1 = \sum_{j=1}^{4} \left(\widehat{\beta}_j - \widehat{\beta}\right)^2 = \left(\widehat{\beta} - \beta\right)' \left(\widehat{\beta} - \beta\right)$ ifadesi kullanılmıştır. Ayrıca önkestirim (prediction) açısından yöntemler arasındaki farklılığı incelemek üzere de $M2 = \left(\widehat{\beta} - \beta\right)' X'X \left(\widehat{\beta} - \beta\right)$ eşitliği kullanılmıştır. Bu değerler M1 için Tablo 1'de ve M2 için Tablo 2'de verilmiştir. Buna göre, M1 açısından Huber ve Mallows kestiriciler tek ve iki aykırı değer durumunda hemen hemen benzer sonuçlar verirken, EKK'nın tutarsız sonuçlar verdiği görülmüştür. M2 hesabına göre, n=10 ve tek aykırı değer durumunda Huber ve Mallows 134, 23 alt kümelerini seçerken diğer alt kümelerde de benzerlikler gözlenmiştir. Yine aynı şekilde tek aykırı değerli ve n=14, 28 gözlemli örneklemler için de Huber ve Mallows benzer seçimler vermiştir. Aynı sonuca iki aykırıdeğer durumunda da raslanmıştır. Ancak önkestirim sonuçları, seçilen alt kümeler açısından M1 durumundan farklılık göstermiştir.



Şekil 5.



Şekil 6.



Şekil 7.

Ayrıca yapılan hesaplama ve grafiklerden Mallows-türü ağırlıklı sağlam C_p kriterinin klasik C_p ve Huber-türü ağırlıklı sağlam C_p kriterinden daha fazla sayıda altküme seçimi verdiği gözlenmiştir. 10 gözlemli örneklemde ve aykırı değerin olmadığı durumda üç seçim kriteri de benzer sonuçlar vermiştir. Aykırı değer varlığında artık kareler ortalamasının artmasından dolayı klasik Cp kriteri güvenilir sonuçlar vermemiştir. 14 ve 28 gözlemli örneklemlerde ve aykırı değerin yokluğunda klasik C_p kriteri seçenek altkümeler önermektedir.

Tablo 1.

n = 10, Aykırı değer yok			n = 10, Tek aykırı değer			n = 14, Tek aykırı değer			n = 28, Tek aykırı değer		
EKK	Huber	Mallows	EKK	Huber	Mallows	EKK	Huber	Mallows	EKK	Huber	Mallows
0.0372	0.0191	0.0191	15.3363	0.0800	0.0800	0.0733	0.0044	0.00419	0.1103	0.0088	0.0088
0.0534	0.1065	0.1129	8.5870	0.0467	0.0467	0.4364	0.0223	0.0254	0.1056	0.0089	0.0089
0.0307	0.0114	0.0160	18.3395	0.0927	0.1001	0.1593	0.00005	0.00003	0.0322	0.020	0.0130
0.0149	0.0198	0.0198	3.1591	0.0116	0.0106	0.207	0.0206	0.0235	0.4805	0.0430	0.0312
0.0290	0.0384	0.0384	11.2703	0.0312	0.0312	0.1516	0.0020	0.0029	0.1114	0.0088	0.0088
0.0286	0.0213	0.0244	17.6894	0.0276	0.0276	0.2061	0.00001	0.00003	0.0085	0.0237	0.0207
0.0205	0.0030	0.0036	2.3443	-	-	0.2951	0.0120	0.0173	0.4829	0.0430	0.0430
0.0120	0.0027	0.0027	16.1074	0.0856	0.0299	0.0318	0.0088	0.0108	0.0247	0.0268	0.0238
0.0958	0.0928	0.0915	18.6643	0.1495	0.1572	0.4210	0.0285	0.0253	0.2627	0.0089	0.00715

n = 10, İki aykırı değer			n = 14, İki aykırı değer			n = 28, İki aykırı değer		
EKK	Huber	Mallows	EKK	Huber	Mallows	EKK	Huber	Mallows
5.7258	25.046	25.046	5.3033	0.0450	0.0492	0.5961	0.0090	0.009
5.2156	0.0693	0.0693	4.8512	0.2850	0.3215	0.2340	0.0089	0.0089
16.51	8.104	8.4090	2.2251	0.0282	0.0282	0.3292	0.0207	0.0156
5.6517	0.0153	0.0153	5.3529	0.1389	0.1973	1.3601	0.0430	0.0312
5.6059	0.0493	0.0493	5.4526	0.0450	0.0548	0.5447	0.0088	0.0088
16.3521	0.0244	0.0213	1.9516	0.0275	0.0345	0.3588	0.0237	0.01795
4.9379	0.0202	0.0195	5.1144	0.0698	0.0948	1.8280	0.0430	0.472
14.2510	-	-	1.7315	0.0303	0.0338	1.2457	0.0303	0.271
6.5244	31.301	31.687	4.9215	0.1622	0.1622	0.2358	0.0088	0.0070

Tablo 2.

n = 10, Aykırı değer yok			n = 10, Tek aykırı değer			n = 14, Tek aykırı değer			n = 28, Tek aykırı değer		
EKK	Huber	Mallows	EKK	Huber	Mallows	EKK	Huber	Mallows	EKK	Huber	Mallows
0.0896	0.0126	0.0126	44.3737	0.2710	0.2710	0.6539	0.0434	0.0316	10.518	0.0131	0.9635
0.3318	0.0729	0.0673	34.778	0.061	0.061	1.1626	0.3671	0.3396	5.7948	0.0292	0.026
0.0778	0.0238	0.0425	44.226	0.2179	0.4895	0.7633	0.0479	0.0338	3.6351	0.0317	0.0225
0.1382	0.4542	0.4542	25.569	1.1151	1.1984	3.5340	0.5836	0.5460	11.469	2.5061	2.2328
0.3735	0.3665	0.3665	35.020	0.0857	0.0857	1.1533	0.0043	0.0051	5.8762	0.0324	0.1072
0.3958	0.1478	0.1293	34.570	0.1129	0.1129	1.7732	0.1057	0.1307	8.3476	0.1731	0.1354
1.6730	1.7736	1.6943	25.432	-	-	3.7080	2.3214	2.2789	11.639	3.3317	3.3317
0.0862	0.3031	0.3031	44.905	0.0707	0.2504	0.8435	0.0230	0.1557	3.7014	0.1981	0.1981
0.0840	0.0457	0.0489	43.478	0.5378	0.5777	0.7743	0.4495	0.4013	13.912	0.0772	0.0663

n = 10, İki aykırı değer			n = 14, İki aykırı değer			n = 28, İki aykırı değer		
EKK	Huber	Mallows	EKK	Huber	Mallows	EKK	Huber	Mallows
45.996	64.192	64.192	5.7549	0.8920	0.2993	15.059	0.0131	0.0131
48.300	0.3305	0.1023	13.721	3.5811	3.6739	15.496	0.0267	0.0267
42.078	53.113	44.606	12.925	1.7643	1.7643	19.361	0.5129	0.3606
46.839	1.4849	1.4849	16.987	0.3610	0.6503	19.007	0.5061	2.2328
63.418	0.2322	0.2322	7.9352	0.1381	0.1627	13.961	0.1072	0.1072
58.093	0.1293	0.1478	14.183	0.9117	1.0597	17.815	0.1052	0.0516
47.889	1.9689	1.8549	5.7045	2.1334	2.3504	23.805	3.3317	3.3317
56.394	-	-	6.0515	0.1405	0.4812	16.954	0.2473	0.1152
59.852	56.396	63.586	12.546	4.7879	4.7879	17.249	0.0470	0.0663

Gözlem sayısı az iken klasik ve sağlam C_p kötü sonuçlar verirler. Gözlem sayısı arttıkça sağlam C_p kriteri daha iyi sonuç verirken Mallows'un C_p'si büyük denek sayıları için önerilebilir.

Aykırı değer olmadığında tüm kestiricilerin benzer sonuçlar verdiği gözlenmiştir. Aykırı değer varlığında ise Huber, Tukey ve Mallows benzer sonuçlar vermekle birlikte En Küçük Kareler yöntemi tutarsızlık göstermiştir.

KAYNAKÇA

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transaction on Automatic Control*, AC-19, 716-723.

Coakley, C.W., and Hettmansperger, T.P. (1993), A Bounded Influence High Breakdown, Efficient Regression Estimator, *JASA*, 88, 872-880.

Hampel, F.R., (1983), Some aspects of model choice in robust statistics. *In Proc. 44th Sess. International Statistics Institute, Madrid, book* 2, 767-771.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A., (1986), Robust Statistics: The Approach Based on Influence Functions, *John Wiley*, New York.

Hurvich, C.F. and Tsai, C.L., (1989), Regression and time series model selection in small samples, *Biometrika*, 76, 2, 297-307.

Hurvich, C.F. and Tsai, C.L., (1990), Model selection for least absolute deviations regression in small samples, *Statistics and Probab. Letters*, 9, 259-265.

Machado, J.A.F., (1993), Robust model selection and M estimation, *Econometric Theory*, 9, 478-493.

Martin, R.D., (1980), Robust Estimation of Autoregressive Models, *In Direction in Time Series, Institute of Mathematical Statistics*, 228-262.

Ryan, P.T., (1998), Modern Regression, *John Wiley*, New York.

Ronchetti, E., (1982), Robust Testing in linear models; the infinitesimal approach, *PhD Thesis., Eidgenössische Technisches Hochschule*, Zurich.

Ronchetti, E., (1985), Robust model selection in regression, *Statistics and Probab. Letters.*, 3, 21-23.

Ronchetti, E., Field, C. and Blanchard, W., (1997), Robust linear model selection by cross- validation, *JASA*, 92, N.439, 1917-1924.

Ronchetti, E. and Staudte, R.G., (1994), A robust versions of Mallows' Cp, *J.Am. Statst. Ass.*, 89, 550-559.

Staudte, R.G. and Sheather, S.J., (1990), Robust Estimation and Testing, *John Wiley*, New York.

Meral Candan Çetin, Ankara doğumludur. Hacettepe Üniversitesi İstatistik Bölümü'nden mezun olduktan sonra yüksek lisans eğitimini aynı üniversitede tamamlamıştır. Halen İstatistik Bölümün'de Araştırma Görevlisi olarak doktora tez çalışmalarına devam etmektedir.

Aydın Erar, lisans, yüksek lisans ve doktorasını Hacettepe Üniversitesi İstatistik Bölümü'nde tamamlamıştır. Regresyon çözümlemesi, istatistiksel veri çözümlemesi ile ilgilenmektedir.