



## LOJİSTİK REGRESYON MODELİNDE KÖTÜ KALDIRAÇ NOKTALARININ BELİRLENMESİ İÇİN YENİ SAĞLAM EŞİK DEĞERLER

### NEW ROBUST CUT-OFF VALUES IN DETERMINING BAD LEVERAGE POINTS IN THE LOGISTIC REGRESSION MODEL

Ebru GÜNDOĞAN AŞIK<sup>1</sup>, Arzu ALTIN YAVUZ<sup>2</sup>, Zafer KÜÇÜK<sup>3</sup>



1. Arş. Gör., Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, ebrugundogan@ktu.edu.tr, <https://orcid.org/0000-0002-9910-6555>
2. Doç. Dr., Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, aaltin@ogu.edu.tr, <https://orcid.org/0000-0002-3277-740X>
3. Prof. Dr., Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, zkucuk@ktu.edu.tr, <https://orcid.org/0000-0001-8083-2429>

**Makale Türü** Article Type  
Araştırma Makalesi Research Article

**Başvuru Tarihi** Application Date  
20.02.2020 02.20.2020

**Yayına Kabul Tarihi** Admission Date  
24.04.2021 04.24.2021

DOI

<https://doi.org/10.30798/makuiibf.691623>

#### Öz

Yüksek kaldıraç noktası, x uzayının merkezine uzak olan değer olarak adlandırılır. İyi ya da kötü kaldıraç noktaları yüksek kaldıraç noktası olabilir. Kötü kaldıraç noktaları, yanlış sınıflandırılmış gözlemler veya x uzayındaki diğer gözlem değerleri ile uyumsuzluk gösteren aykırı değerlerdir. Kötü kaldıraç noktalarının belirlenmesinde maskeleyme ve süpürme problemini ortadan kaldırmak için kullanılan grup silme yöntemi lojistik regresyon modelinde de kullanılmaktadır. Bu çalışmada kötü kaldıraç noktalarının belirlenmesinde literatürde mevcut olan Sapma Bileşenleri (Deviance Component, DEVC) yöntemi için bazı sağlam eşik değerleri önerilmiştir. Yapılan simülasyon çalışması ile sapma bileşenleri yönteminde kullanılması için önerilen sağlam eşik değerlerin literatürde mevcut olan eşik değerden daha iyi sonuçlar verdiği ortaya konmuştur.

**Anahtar Kelimeler:** Lojistik Regresyon, Yüksek Kaldıraç Noktası, Kötü Kaldıraç, Eşik Değeri.

#### Abstract

The high leverage point is called the value far from the center of the x space. Good or bad leverage points can be high leverage points. Bad leverage points misclassified observations and outliers that are incompatible with other observation values in the x space. The group deletion method used to eliminate the masking and swamping problem in determining bad leverage points is also used in the logistic regression model. In this study, some robust cut-off points are proposed for the deviation components method available in the literature for determining bad leverage points. In this study, some robust cut-off points are proposed for the deviation components method available in the literature for determining bad leverage points.

**Keywords:** Logistic Regression, High Leverage Point, Bad Leverage, Robust, Cut-off Point.

## **EXTENDED SUMMARY**

### **Research Problem**

In the logistic regression model, cut-off point is an important parameter in determining bad leverage points with high leverage points. For the DEVC method, which is used in determining the bad leverage points, alternative robust cut-off points are proposed. These cut-off points, consisting of the location and scale parameters, are created from different combinations of robust estimators and scale estimators of the location in the literature. With the help of these cut-off points, it is aimed to determine the bad leverage points in the dataset correctly.

### **Research Questions**

Does the recommended robust cut-off points in determining bad leverage points in the logistic regression model have better results than the current cut-off point? What is the contribution of the proposed robust cut-off points in determining bad leverage?

### **Literature Review**

Extraordinary observations, which are an important concept in the regression model, are important in the logistic regression model. These observations need to be examined to determine the model correctly. An existing cut-off point is used in the methods developed to identify these observations known as outlier value, effective observation and high leverage point. In the logistic regression model, the DEVC method was used to identify bad leverage points. Norazan et al, like many researchers in the literature in the DEVC method, tried to determine bad leverage points using Median + 3MAD cut-off value (Habshah et al 2009; A.H.M.R Imon & Hadi, 2013).

### **Methodology**

In this study, the DEVC method used in determining the high leverage points in the logistic regression model is discussed. The cut-off point used for the DEVC method is known as the Median + 3MAD. Considering the Median + 3MAD cut-off point, the position parameter is the alternative Trimmed mean, Winsorized mean, Huber, Hampel, Tukey Bisquare, Andrew and Hodges-Lehmann estimators and the scale parameter MAD. Alternatively,  $Q_n$  and  $S_n$  estimators are used. By creating different combinations of these cut-off points obtained, a simulation mechanism was created with different sample sizes and different bad leverage point (outlier value in the x direction) dimensions. Accurate determination and swamping rates were obtained for each threshold value obtained.

### **Results and Conclusions**

As a result, in this study, some alternative robust cut-off points were obtained for the threshold value used in the DEVC method. It has been demonstrated with correct determination rate and swamping rate that these obtained cut-off points give better results in determining bad leverage points than the cut-off point available in the literature.

## 1. GİRİŞ

Regresyon analizi, bağımlı değişken ile bağımsız değişken/değişkenler arasındaki fonksiyonel ilişkiyi açıklamak ve bağımlı değişken değerlerini tahmin etmek için kullanılan istatistiksel bir yöntemdir. Regresyon analizinde genellikle bağımlı değişken sürekli ve ölçülebilir bir değişken olarak ele alınmaktadır. Ancak bazı durumlarda bu mümkün değildir. Bağımlı değişkenin iki düzeyli kategorik bir değişken olduğu durumda kullanılan regresyon yöntemi ikili lojistik regresyon modeli olarak adlandırılmaktadır. Çoklu regresyonda bağımsız değişken değerleri kullanılarak bağımlı değişkenin sonraki değerleri tahmin edilirken, ikili lojistik regresyon modelinde ilgilenilen olayın gerçekleşme olasılığı tahmin edilmektedir.

Çoklu doğrusal regresyonda olduğu gibi, lojistik regresyon analizinde de olağandışı gözlemlerle karşılaşmaktadır. Regresyon analizinde karşılaşılan olağandışı gözlemler; aykırı değer, etkili gözlem ve yüksek kaldıraç noktası olarak adlandırılmaktadır. Aykırı değerler, verinin büyük bir çoğunluğundan ayırt edilebilen gözlemlerdir (Maronna vd., 2006). Etkili gözlem ise veri kümesinden çıkartıldığında hesaplanan çeşitli istatistikleri büyük ölçüde değiştiren gözlemlerdir (Belsley vd., 1980). Yüksek kaldıraç noktaları ise x uzayında merkeze uzak olan noktalardır. Yüksek kaldıraç noktaları iyi kaldıraç noktaları (Good Leverage Points) ve kötü kaldıraç noktaları (Bad Leverage Points) olarak ikiye ayrılır. İyi kaldıraç noktaları, verilerin çoğunluğunun düzenini izleyen ve parametre tahminlerine katkı sağlayan noktalardır (Habshah vd., 2009). Kötü kaldıraç noktaları ise verilerin çoğunluğunun düzenine uymayan ve parametre tahminleri üzerinde olumsuz etkilere sahip etkili gözlemlerdir. Kötü kaldıraç noktaları, bağımsız değişkenlerle ilgili uzaktaki (dış, outlying) noktalardır. Yüksek kaldıraç noktalarının MLE tahmin edicilerini ve model uyumunu olumsuz etkileyen tehlikeli aykırı gözlemler olduğu belirtilmiştir (Rousseeuw ve Christophe, 1993).

Yüksek kaldıraç noktalarının varlığı, lojistik regresyon modelinde daha fazla soruna neden olmaktadır. Yüksek kaldıraç noktalarından kaynaklanan etki, diğer kötü noktalardan daha şiddetlidir. Yüksek kaldıraç noktalarının sadece yanlış parametre tahminlerini elde etmekle sorumlu olmadığı ayrıca maskeleyen probleme de neden olabildiği belirtilmiştir. (A.H.M.R Imon, 2006). Maskeleyen etkisi, başka yüksek kaldıraç noktalarının varlığı nedeniyle dış noktaların tespit edilememesi durumunda ortaya çıkar. Süpürme etkisi ise iyi noktaların yanlış tanımlanmasından dolayı kötü noktalar olarak belirlenmesi olarak ortaya çıkmaktadır.

Literatür incelendiğinde lojistik regresyon modeli için yüksek kaldıraç noktalarını belirlemede birçok yöntem bulunmaktadır. Yaygın kullanıma sahip bu yöntemlerden bazıları; Ortalamadan Uzaklık (Distance from the Mean) yöntemi, Genelleştirilmiş Ağırlıklar (Generalized Weights) ve Sağlam Lojistik Teşhis Edici (Robust Logistic Diagnostic) yöntemleridir (A.H.M.R Imon, 2006; Nurunnabi vd., 2010; Syaiba ve Habshah, 2010). Minimum kovaryans determinanı (MCD) veya minimum hacim elipsoidi (MVE) kullanılarak hesaplanan Robust Mahalanobis uzaklığı ile de yüksek kaldıraç noktaları

belirlenmektedir. İncelenen bu belirleme yöntemleri yüksek kaldıraç noktalarını iyi ya da kötü kaldıraç noktaları olup olmadığına bakılmaksızın belirlemektedir. Yüksek kaldıraç noktalarının belirlenmesinde kullanılan yöntemlerden biri olan Sapma Bileşenler yöntemi (DEVC) lojistik regresyon modeli için de kullanılmaktadır. DEVC teşhis yönteminde, artıkların belirlenmesi kadar eşik değerlerinde doğru tespiti önemlidir. Bu nedenle çalışmamızda DEVC teşhis edici yönteminde kullanılan Medyan+3MAD eşik değeri için alternatif bazı sağlam eşik değerleri sunulmuştur. Konum ve ölçek parametresinin bazı sağlam tahmin edicilerinin farklı kombinasyonları kullanılmıştır. Konum parametresi için kullanılan Medyan tahmin edicisine alternatif olarak; Trimmed, Winsorized, Huber, Hampel, Tukey Bisquare, Andrew and Hodges-Lehmann (HL) tahmin edicileri ve ölçek parametresi için kullanılan MAD değerine alternatif olarak ise;  $Q_n$  ve  $S_n$  tahmin edicileri ele alınmıştır.

Çalışmanın ikinci bölümünde lojistik regresyon modeli ve yüksek kaldıraç noktalarının belirlenmesinde kullanılan bazı yöntemlere değinilmiştir. Bölüm 3 de eşik değer için alternatif tahmin ediciler açıklanmıştır. Bölüm 4 de önerilen eşik değerler ile klasik eşik değerlerin kötü kaldıraç noktası teşhisindeki performansı bir simülasyon çalışması ile karşılaştırılmıştır. Bölüm 5 de elde edilen sonuçlar sunulmuştur.

## 2. LOJİSTİK REGRESYON VE DEVC YÖNTEMİ

Bağımlı değişkenin 0,1 değeri alan Bernoulli dağılımına sahip olduğu lojistik regresyon modeli,

$$P(Y = 1|X = x) = \pi_i = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}, \quad i = 1, 2, \dots, n \quad (1)$$

şeklinindedir. Burada,  $Y$ :  $n \times 1$  boyutlu bağımlı değişken vektörü,  $\beta$ :  $(p+1) \times 1$  boyutlu bilinmeyen parametreler vektörü,  $X$ :  $n \times (p+1)$  bağımsız değişkenler matrisi ve  $\varepsilon$   $n \times 1$  boyutlu hata terimleri vektörüdür.  $\pi_i$  niceliği  $i$ 'inci faktörün bir olasılığıdır ve  $0 \leq \pi_i \leq 1$  arasında değer almaktadır. Bilinmeyen  $\beta$  parametrelerinin tahmini için En Çok Olabilirlik Yöntemi (EÇO) kullanılmaktadır. Olabilirlik ve log-olabilirlik fonksiyonu,

$$L(\beta; y) = \prod_i^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (2)$$

$$l(\beta; y) = \sum_i^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \quad (3)$$

şeklinde yazılmaktadır. Log-olabilirlik fonksiyonunun  $\beta$ 'ya göre türevi alınarak sifıra eşitlendiğinde elde edilen denklemlerin yinelemeli çözümleri  $\beta$  'nın EÇO tahmin edicisi olarak adlandırılır.  $\hat{\beta}$  tahminleri kullanılarak  $\hat{\pi}_i$  değerleri hesaplanmaktadır.

Lojistik regresyon modelinde modelin genel uyumunu değerlendirmek için kullanılan sapma (DEV),

$$DEV = \sum_i^n d(y_i, \hat{\pi}_i)^2 = \sum_i^n 2 \left[ y_i \log\left(\frac{y_i}{\hat{\pi}_i}\right) + (1 - y_i) \log\left(\frac{1-y_i}{1-\hat{\pi}_i}\right) \right] \quad (4)$$

şeklinde belirlenmektedir. Burada  $d(y_i, \hat{\pi}_i)$  sapma artıkları olarak bilinmektedir ve sapmaların farkları üzerine oluşturulmuş bir aykırı değer belirleme yöntemidir (Sarkar vd., 2011). Sapma artıkları,

$$d_i = \text{sign}(y_i - \hat{\pi}_i) \left\{ 2 \left[ y_i \log \left( \frac{y_i}{\hat{\pi}_i} \right) + (1 - y_i) \log \left( \frac{1 - y_i}{1 - \hat{\pi}_i} \right) \right] \right\}^{1/2} \quad (5)$$

biçiminde elde edilir. Burada  $\text{sign}(y_i - \hat{\pi}_i)$  fonksiyonu,  $(y_i - \hat{\pi}_i)$  değerine göre pozitif ya da negatif yapan işaret fonksiyonudur. Eşitlik (5) incelendiğinde  $i$ ' inci sapma artığı için bir bileşen

$$dc_i = \begin{cases} 2 \log \left( \frac{1}{1 - \hat{\pi}_i} \right) & , y_i = 0 \\ 2 \log \left( \frac{1}{\hat{\pi}_i} \right) & , y_i = 1 \end{cases} \quad (6)$$

şeklinde tanımlanmaktadır ve DEVC olarak adlandırılmaktadır (Norazan vd., 2012). DEVC 'nin dağılımı bilinmediğinden aykırı değerleri belirlemede önerilen eşik değer,

$$dc_i > \text{Medyan}(dc_i) + 3\text{MAD}(dc_i) \quad (7)$$

şeklinde hesaplanmaktadır (Hadi, 1992) . Birçok araştırmacı tarafından ele alınan olağandışı gözlemlerin belirlenmesi yöntemlerinde *Medyan + 3MAD* eşik değeri kullanılmaktadır (Alguraibawi vd, 2015; Fitrianto ve Midi, 2010; Habshah vd., 2009; A. H.M.R. Imon ve Hadi, 2013; Norazan vd., 2012; Sarkar vd., 2011).

DEVC yöntemi için de bu eşik değeri kullanılmaktadır. Ancak literatür incelendiğinde artıkların hesaplanmasında kullanılan yöntemlerin iyileştirildiği birçok çalışma olmasına rağmen eşik değerlerin aynı şekilde hesaplandığı görülmektedir. Kötü kaldıraç noktası teşhisi için kullanılan eşik değerinde doğru tespit edilmesi teşhis yöntemlerinin performansı üzerinde etkilidir. Bu nedenle ilerleyen bölümde bu eşik değere alternatif olarak önerilmiş eşik değerler tanıtılmıştır.

### 3. EŞİK DEĞERLERİN BELİRLENMESİNDE SAĞLAM TAHMİN EDİCİLERİ

Eşik değer tespitinde artıklar için bir ortalama ve değişkenlik ölçüsü kullanılmaktadır. Eşitlik (7) de verilen eşik değeri hesabında ortalama olarak aykırı gözlemlerden etkilenmeyen Medyan değeri kullanılmıştır. Değişkenlik ölçüsü olarak ise Medyan Mutlak sapma (MAD) ölçüsü ele alınmıştır.

Bu çalışmada Medyan tahmin edicisine alternatif olarak yedi farklı konum tahmin edicisinin kullanımı önerilmiştir. Bu yedi farklı konum tahmin edicisi M, L ve R tipi sağlam tahmin ediciler sınıfından seçilmiştir.

#### 3.1. M Tahmin Ediciler

M tahmin edici, en çok olabirlik tipinde sağlam tahmincilerden biridir.  $X_1, X_2, \dots, X_n$  n birimlik rassal örneklem,  $\theta$  anakitle ortalaması ve  $u_i$  hata terimleri olarak ifade edilen  $x_i = \theta + u_i$  konum modelinden hareketle  $\theta$  konum parametrelili anakitle ortalamasının M tahmin edicisi,

$$\hat{\theta}_n = \arg \min_{\theta} \sum_{i=1}^n \rho(x_i - \theta) \quad (8)$$

biçiminde elde edilmektedir (Huber, 1953). Eşitlik (8) da verilen denklemin doğrudan çözümü zordur. Bu ifadenin  $\theta$ ' ya göre türevi alınarak sifira eşitlendiğinde,

$$\sum_{i=1}^n \psi(x_i - \theta) = 0 \quad (9)$$

ortalamanın M tahmin edicisi elde edilmektedir (Maronna ve Yohai, 1901). M tahmin edicisi seçilen  $\rho$  fonksiyonuna göre değişmektedir.  $\rho$  amaç fonksiyonun türevi etki fonksiyonu olarak adlandırılmaktadır. M tahmin edicilerinde DEVC yöntemi için kullanılan eşik değeri,

$$dc_i > \hat{\theta}_n(dc_i) + 3MAD(dc_i) \quad (10)$$

şeklinde hesaplanabilir. Bazı M tahmin edicilerin çeşitli özellikleri Tablo 1 'de verilmiştir. Ayrıca Tablo 1 'de yer alan ayarlama sabiti, tahmin edicinin aykırı değerlere karşı sağlamlığını ve aykırı değerlerin yokluğunda tahmincinin etkinliğini belirlemede yardımcı olmaktadır. Ayarlama sabitinin değeri, normal dağılım etkinliğini ve tahmin edicinin dayanıklılığını yüksek yapacak şekilde optimum olmalıdır (Welsch, 1977). Ele alınan diğer sabit değerler araştırmacılar tarafından önerilen literatürde yaygın kullanılan değerlerdir.

**Tablo 1.** M tahmin Edicilerin Özellikleri

M Tahmincisi	Türü	Ayarlama Sabiti	Etki Fonksiyonu $\psi$
<b>Huber</b>	Monoton	$c = 1.345$	$\psi(x) = \begin{cases} x & ,  x  \leq c \\ c \operatorname{sgn}(x) & ,  x  > c \end{cases}$
<b>Andrews</b>	Hızlı yeniden azalan	$c = 1.339$	$\psi(x) = \begin{cases} \sin(x/c) & ,  x  \leq c\pi \\ 0 & ,  x  > c\pi \end{cases}$
<b>Tukey Bisquare</b>	Hızlı yeniden azalan	$c = 4.685$	$\psi(x) = \begin{cases} x \left[ 1 - \left( \frac{x}{c} \right)^2 \right]^2 & ,  x  \leq c \\ 0 & ,  x  > c \end{cases}$
<b>Hampel</b>	Hızlı yeniden azalan	$a=1.7$ $b=3.4$ $c=8.5$	$\psi(x) = \begin{cases} x & ,  x  \leq a \\ a \operatorname{sgn}(x) & , a <  x  \leq b \\ a \frac{c -  x }{c - b} \operatorname{sgn}(x) & , b <  x  \leq c \\ 0 & ,  x  > c \\ 0 < a \leq b \leq c \end{cases}$

### 3.2. Medyan Tahmin Edici

L tahmin edicileri, sıra istatistiklerine dayalı sağlam tahmin ediciler olarak adlandırılmaktadır. L tahmin ediciler arasında yaygın kullanıma sahip olan Medyan,

$$\text{Medyan} = \begin{cases} x_{(M+1)} & , n = 2M + 1 \\ \frac{x_{(M)} + x_{(M+1)}}{2} & , n = 2M \end{cases} \quad (11)$$

şeklinde ifade edilmektedir. Örnek hacmine göre veri setinde sadece bir veya iki gözlemin farklı ağırlıklandırıldığı, diğer tüm gözlem değerlerine sıfır ağırlığının verildiği bir tahmincidir.

### 3.3. Trimmed Tahmin Edici

Veri setinde aykırı değer olduğu durumda anakitle ortalaması yerine alternatif olarak kullanılan Trimmed ortalama, n birimlik örneklemden sıra istatistiklerinin her iki ucundan belli oranlarda gözlem değerlerinin atılmasından sonra hesaplanan ortalamadır (Hoaglin ve Mosteller, 1983). L sınıfı sağlam tahmin ediciler içerisinde en yaygın kullanıma sahip olan tahmin edicidir.

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  sıra istatistikleri olmak üzere  $(\alpha, 1 - \beta)$  trimmed ortalama,

$$T_n = T_n(l_n, u_n) = \frac{1}{u_n - l_n} \sum_{i=l_n+1}^{u_n} x_{(i)} \quad (12)$$

şeklinde ifade edilmektedir (Yavuz, 2013). Burada  $l_n = [n\alpha]$  sıralanmış gözlemler içerisinde alt uçtan atılan gözlem sayısını,  $u_n = [n\beta]$  ise üst uçtan atılan gözlem sayısını göstermektedir.  $[\cdot]$  operatörü en büyük tamsayı fonksiyonunu göstermektedir. Genellikle  $0 \leq \alpha \leq 0.25$ ,  $1 - \beta \leq 0.25$  alınmaktadır (Hoaglin ve Mosteller, 1983).

### 3.4. Winsorize Tahmin Edici

Winsorize ortalama, veri setinde aykırı değer olması durumunda alternatif olarak kullanılan bir diğer L sınıfı ortalamadır.

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  sıra istatistikleri olmak üzere  $(\alpha, 1 - \beta)$  Winsorize ortalama,

$$W_n = W_n(l_n, u_n) = \frac{1}{n} \left\{ l_n x_{(l_n+1)} + \sum_{i=l_n+1}^{u_n} x_{(i)} + (n - u_n) x_{(u_n)} \right\} \quad (13)$$

şeklinde ifade edilmektedir. Burada  $l_n$  ve  $u_n$  ifadeleri Eşitlik (12) de ifade edildiği gibi belirlenmektedir. Trimmed ortalama olduğu gibi kuyrukların ağırlığına göre  $0 \leq \alpha \leq 0.25$ ,  $1 - \beta \leq 0.25$  alınmaktadır (Hastings vd., 1947).

### 3.5. Hodges-Lehmann (HL) Tahmin Edici

R tahmin edicileri, örneklemden ranklara dayalı tahmin edicilerdir. R tahmin edicileri arasında en yaygın kullanıma sahip tahminci Hodges-Lehmann tahmincisidir. HL tahmin edicisi tüm ikili gözlem çiftlerinin ortalamalarının medyanı olarak tanımlanmaktadır.

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  sıra istatistiklerini göstermek üzere ve HL tahmincisi,

$$\hat{\theta}_{HL} = \text{medyan} \left\{ \frac{x_{(i)} + x_{(j)}}{2} \right\}, \quad 1 \leq i < j \leq n \quad (14)$$

şeklinde ifade edilmektedir (Hodges ve Lehmann, 1963).

Çalışmada ayrıca eşik değerlerin belirlenmesinde kullanılan MAD 'ın da sağlam alternatifleri ele alınmıştır. Ölçek parametresinin yaygın kullanılan bir tahmincisi olan varyans, veri kümesindeki olağandışı gözlemlerden oldukça fazla etkilenmektedir. Bu nedenle ölçek parametresi için sağlam tahminciler önerilmiştir. MAD,  $S_n$  ve  $Q_n$  tahmin edicileri varyansa alternatif sağlam değişkenlik

ölçüleridir. Bu tahmin edicilerinin kırılma noktaları ve etkinlikleri göz önüne alınarak kısaca açıklanmıştır.

### 3.6. Medyan Mutlak Sapma (MAD)

Medyan mutlak sapma, hesaplama açısından kolay olması ve etki fonksiyonu içinde dayanıklı olmasından dolayı yaygın olarak kullanılan sağlam ölçek tahmin edicisidir. MAD,

$$MAD = \frac{\text{medyan}(|x_i - \text{medyan}(x_i)|)}{0.6745} \quad (15)$$

şeklinde hesaplanmaktadır (Hampel, 1974). Paydadaki ölçek sabiti, ölçegin tahminini normal dağılımın standart sapması ile tutarlı hale getirmek için MAD' ı standardize etmektedir. %50 kırılma noktasına sahip olan MAD yöntemi simetrik dağılımlar için geliştirilmiştir ve etkinliği %37 'dir (Rousseeuw ve Christophe, 1993).

### 3.7. $S_n$ Tahmin Edicisi

MAD ölçek tahmin edicisine alternatif bir tahmin edici olan  $S_n$  tahmin edicisi,

$$S_n = c_n 1.1926 \text{ medyan}_i \{ \text{medyan}_j |x_i - x_j| \}, \quad i, j = 1, 2, \dots, n \quad (16)$$

biçiminde elde edilmektedir. Burada  $c_n$  ve 1.1926 sabitleri  $S_n$  tahmin edicisinin yansız olmasını sağlamaktadır.  $c_n$  sabitinin değeri ise örneklem hacmine göre değişmektedir (Rousseeuw ve Christophe, 1993).  $S_n$  tahmin edicisinin konumdan bağımsız olması asimetrik dağılımlar için de kullanılmasını sağlamaktadır. Kırılma noktası %50 ve etkinliği ise MAD tahmin edicisine göre daha yüksektir (%58).

### 3.8. $Q_n$ Tahmin Edicisi

Önerilen bir diğer ölçek tahmin edicisi olan  $Q_n$  tahmin edicisi,

$$Q_n = d_n 2.2219 \{ |x_i - x_j|; i < j \}_{(k)}, \quad i, j = 1, 2, \dots, n \quad (17)$$

şeklinde hesaplanmaktadır (Rousseeuw ve Christophe, 1993). Burada  $k = \binom{h}{2} \approx \binom{n}{2}/4$  ve  $h$  değeri  $\left(\frac{n}{2} + 1\right)$  dir. Yansızlığı sağlayan  $d_n$  sabit değeri örneklem hacmine göre farklı değerler almaktadır. Yüksek etkinliğe sahip olan (%82)  $Q_n$  tahmin edicisi asimetrik dağılımlar içinde kullanılmaktadır ve kırılma noktası MAD ve  $S_n$  tahmin edicileri ile aynıdır.  $\text{Medyan} + 3MAD$  eşik değerine konum parametresinin alternatif olarak Trimmed, Winsorize, Huber, Hampel, Tukey Bisquare, Andrew ve HL tahmin edicileri ele alınmıştır. Ayrıca ölçek tahmin edicisi MAD için de alternatif olarak  $Q_n$  ve  $S_n$  tahmin edicileri kullanılmıştır. Konum ve ölçek parametrelerinin sağlam tahmin edicilerinin farklı kombinasyonları ile yeni eşik değerler önerilmiştir. Mevcut olan eşik değer ve önerdiğimiz sağlam eşik değerleri Tablo 2 de gösterilmiştir.



**Tablo 2.** M Tahmin Edicilerin Özellikleri

<b>CT:</b> <i>Medyan</i> + 3 <i>MAD</i>	<b>CT8:</b> <i>Hampel</i> + 3 <i>Q<sub>n</sub></i>	<b>CT16:</b> <i>Trimmed</i> + 3 <i>S<sub>n</sub></i>
<b>CT1:</b> <i>Medyan</i> + 3 <i>S<sub>n</sub></i>	<b>CT9:</b> <i>Andrew</i> + 3 <i>MAD</i>	<b>CT17:</b> <i>Trimmed</i> + 3 <i>Q<sub>n</sub></i>
<b>CT2:</b> <i>Medyan</i> + 3 <i>Q<sub>n</sub></i>	<b>CT10:</b> <i>Andrew</i> + 3 <i>S<sub>n</sub></i>	<b>CT18:</b> <i>Tukey</i> + 3 <i>MAD</i>
<b>CT3:</b> <i>Huber</i> + 3 <i>MAD</i>	<b>CT11:</b> <i>Andrew</i> + 3 <i>Q<sub>n</sub></i>	<b>CT19:</b> <i>Tukey</i> + 3 <i>S<sub>n</sub></i>
<b>CT4:</b> <i>Huber</i> + 3 <i>S<sub>n</sub></i>	<b>CT12:</b> <i>Winsorize</i> + 3 <i>MAD</i>	<b>CT20:</b> <i>Tukey</i> + 3 <i>Q<sub>n</sub></i>
<b>CT5:</b> <i>Huber</i> + 3 <i>Q<sub>n</sub></i>	<b>CT13:</b> <i>Winsorize</i> + 3 <i>S<sub>n</sub></i>	<b>CT21:</b> <i>HL</i> + 3 <i>MAD</i>
<b>CT6:</b> <i>Hampel</i> + 3 <i>MAD</i>	<b>CT14:</b> <i>Winsorize</i> + 3 <i>Q<sub>n</sub></i>	<b>CT22:</b> <i>HL</i> + 3 <i>S<sub>n</sub></i>
<b>CT7:</b> <i>Hampel</i> + 3 <i>S<sub>n</sub></i>	<b>CT15:</b> <i>Trimmed</i> + 3 <i>MAD</i>	<b>CT23:</b> <i>HL</i> + 3 <i>Q<sub>n</sub></i>

CT: Eşik Değer CT1-23: Önerilen Eşik Değerler

#### 4. SİMÜLASYON ÇALIŞMASI

Bu bölümde Monte Carlo simülasyon çalışması ile lojistik regresyon modelinde kötü kaldıraç noktalarının belirlenmesi üzerine durulmuştur. Çalışmada, literatürde var olan *Medyan* + 3*MAD* eşik değeri için önerilen alternatif eşik değerlerinin performansları karşılaştırılmıştır. Kötü kaldıraç noktalarının tespitinde yeni eşik değerlerinin kullanılmasıyla, kötü kaldıraç noktalarının ne kadarının doğru belirlendiğinin tespiti için doğru belirlenme oranı (CIR) kullanılmıştır. Ayrıca aslında iyi olan gözlemlerin ne kadarının kötü gözlem olarak belirlendiğinin tespiti için de swamping oranı (SR) verilmiştir. CIR, doğru bir şekilde tanımlanmış şüpheli gözlem sayısının verilerdeki toplam gerçek şüpheli gözlem sayısına oranıdır. SR ise şüpheli olarak tanımlanan iyi gözlem sayısının toplam iyi gözlem sayısına oranını göstermektedir.

İyi bir simülasyon çalışması için tahmin edicilerin özelliklerini etkileyen faktörlerin doğru seçilmesi ve sonuçların değerlendirilmesinde doğru kriterin belirlenmesi gerekmektedir. Bağımsız değişkenlerin sayısı ve örneklem hacmi tahmin edicilerin değerlendirilmesinde önemli bir etkiye sahiptir. Bu çalışma için bağımsız değişken sayısı 2 ve 3 olarak ele alınmıştır. Ele alınan diğer bir etken ise örneklem hacmidir. Farklı tahmin yöntemlerinin karşılaştırılmasında örneklem hacminin artırılması varyansı düşüreceğinden olumlu bir etken olarak düşünülmektedir. Anlamlı sonuçlar elde etmek için örneklem hacminin doğru belirlenmesi gerekmektedir. Bu çalışma için örneklem hacmi  $n$ ; 50,80,100,120,150 olarak belirlenmiştir.

Farklı bağımsız değişken sayıları ve başlangıç parametre değerleri kullanılarak bağımlı değişken,

$$y_i = \begin{cases} 0, & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon_i < 0, \quad i = 1, 2, \dots, n, \quad x_i \sim N(0,1) \\ 1, & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon_i \geq 0, \quad \varepsilon_i \sim \Lambda(0,1), \quad \beta_i = (1, \dots, 1) \end{cases}$$

şeklinde oluşturulmuştur. Burada  $p$  bağımsız değişken sayısını ve  $\varepsilon$  lojistik dağılımdan elde edilen hata terimlerini göstermektedir.,

Veri seti,  $\%(\alpha/2)$  iyi ve  $\%(\alpha/2)$  kötü kaldıraç noktası olacak şekilde oluşturulmuştur. Burada  $\alpha$  yüksek kaldıraç noktalarının farklı yüzdelikleridir ve  $\alpha = 4,6,10$  olarak belirlenmiştir. Veri setine eklenen yüksek kaldıraç noktaları Düzgün [2,4] dağılımdan üretilmiştir. Bu yüksek kaldıraç noktalarına

karşılık gelen  $y$  değerleri ise; iyi kaldıraç noktaları için  $y = 1$ , kötü kaldıraç noktaları için  $y = 0$  şeklinde düzenlenmiştir.

Simülasyon için R programı kullanılmıştır. Çalışmada, göz önünde bulundurulmuş bağımsız değişken sayısı, örneklem hacmi ve kaldıraç noktası değerlerinin tüm kombinasyonları için 10000 deneme gerçekleştirilmiştir ve elde edilen sonuçlar Tablo 3-8 de verilmiştir.

**Tablo3.** Bağımsız Değişken Sayısı P=2, Kötü Kaldıraç Noktası 2 İçin Simülasyon Sonuçları.

n		MAD		$S_n$		$Q_n$	
		CRI	SW	CRI	SW	CRI	SW
50	Median	0,94335	0,04462	0,94550	0,03985	0,95035	0,03523
	Huber	0,92865	0,03283	0,93010	0,02831	0,93405	0,02386
	Hampel	0,92885	0,02616	0,93000	0,02138	0,93365	0,01676
	Andrew	0,94460	0,04060	0,94555	0,03546	0,95010	0,03054
	Winsorize	0,92775	0,02534	0,92865	0,02062	0,93220	0,01583
	Trimmed	0,92850	0,02621	0,92925	0,02155	0,93260	0,01698
	Tukey	0,94385	0,03884	0,94475	0,03374	0,94935	0,02893
	HL	0,92685	0,02710	0,92845	0,02266	0,93175	0,01798
80	Median	0,99765	0,02691	0,99770	0,02487	0,99825	0,02385
	Huber	0,99665	0,01855	0,99680	0,01688	0,99755	0,01574
	Hampel	0,99690	0,01789	0,99685	0,01608	0,99770	0,01497
	Andrew	0,99785	0,02820	0,99770	0,02589	0,99825	0,02500
	Winsorize	0,99680	0,01685	0,99690	0,01521	0,99760	0,01397
	Trimmed	0,99670	0,01690	0,99705	0,01531	0,99770	0,01404
	Tukey	0,99785	0,02662	0,99765	0,02443	0,99825	0,02352
	HL	0,99670	0,01795	0,99765	0,01629	0,99755	0,01520
100	Median	0,99965	0,02116	0,99960	0,02140	0,99990	0,02591
	Huber	0,99945	0,01296	0,99955	0,01316	0,99990	0,01648
	Hampel	0,99950	0,01338	0,99955	0,01360	0,99990	0,01744
	Andrew	0,99970	0,02300	0,99980	0,02338	0,99990	0,02856
	Winsorize	0,99950	0,01274	0,99955	0,01294	0,99990	0,01655
	Trimmed	0,99950	0,01231	0,99955	0,01251	0,99990	0,01606
	Tukey	0,99970	0,02175	0,99980	0,02216	0,99990	0,02719
	HL	0,99945	0,01295	0,99955	0,01316	0,99990	0,01663
120	Median	0,99995	0,04174	0,99995	0,04069	0,99995	0,04709
	Huber	0,99990	0,03086	0,99995	0,02996	0,99995	0,03526
	Hampel	0,99995	0,03168	0,99995	0,03079	0,99995	0,03638
	Andrew	0,99995	0,04490	0,99995	0,04386	0,99995	0,05103
	Winsorize	0,99995	0,03005	0,99995	0,02915	0,99995	0,03455
	Trimmed	0,99995	0,02971	0,99995	0,02881	0,99995	0,03412
	Tukey	0,99995	0,04311	0,99995	0,04210	0,99995	0,04909
	HL	0,99995	0,03063	0,99995	0,02969	0,99995	0,03502
150	Median	1,00000	0,04108	1,00000	0,03779	1,00000	0,04185
	Huber	1,00000	0,03238	1,00000	0,02949	1,00000	0,03295

Hampel	1,00000	0,03380	1,00000	0,03082	1,00000	0,03461
Andrew	1,00000	0,04394	1,00000	0,04056	1,00000	0,04498
Winsorize	1,00000	0,03219	1,00000	0,02930	1,00000	0,03288
Trimmed	1,00000	0,03209	1,00000	0,02918	1,00000	0,03273
Tukey	1,00000	0,04293	1,00000	0,03963	1,00000	0,04396
HL	1,00000	0,03238	1,00000	0,02946	1,00000	0,03296

**Tablo 4.** Bağımsız Değişken Sayısı P=2, Kötü Kaldıraç Noktası 3 İçin Simülasyon Sonuçları.

n		MAD		$S_n$		$Q_n$	
		CRI	SW	CRI	SW	CRI	SW
50	Median	0,80490	0,00691	0,80793	0,00245	0,81490	0,00673
	Huber	0,77600	0,00530	0,77870	0,00941	0,78507	0,01103
	Hampel	0,76863	0,01526	0,77060	0,01936	0,77623	0,02126
	Andrew	0,80120	0,00211	0,80367	0,00695	0,81257	0,00642
	Winsorize	0,77280	0,01440	0,77523	0,01878	0,78077	0,02071
	Trimmed	0,77267	0,01295	0,77483	0,01719	0,78087	0,01918
	Tukey	0,80020	0,00268	0,80293	0,00738	0,81197	0,00896
	HL	0,77110	0,01511	0,77393	0,01953	0,78113	0,02129
80	Median	0,99087	0,02758	0,99183	0,02590	0,99333	0,01982
	Huber	0,98697	0,01576	0,98813	0,01442	0,99007	0,00906
	Hampel	0,98733	0,00984	0,98790	0,00817	0,98953	0,00274
	Andrew	0,99173	0,02238	0,99217	0,02036	0,99357	0,01439
	Winsorize	0,98733	0,00936	0,98777	0,00770	0,98980	0,00228
	Trimmed	0,98753	0,01071	0,98817	0,00911	0,98983	0,00369
	Tukey	0,99153	0,02073	0,99210	0,01882	0,99343	0,01288
	HL	0,98677	0,01151	0,98777	0,01011	0,98963	0,00486
100	Median	0,99827	0,03256	0,99813	0,03000	0,99820	0,02522
	Huber	0,99730	0,02187	0,99723	0,01941	0,99720	0,01529
	Hampel	0,99707	0,01710	0,99717	0,01457	0,99710	0,01038
	Andrew	0,99837	0,02902	0,99830	0,02643	0,99847	0,02163
	Winsorize	0,99727	0,01682	0,99730	0,01428	0,99727	0,01008
	Trimmed	0,99727	0,01763	0,99723	0,01514	0,99723	0,01091
	Tukey	0,99833	0,02733	0,99827	0,02472	0,99843	0,02003
	HL	0,99713	0,01923	0,99720	0,01673	0,99717	0,01263
120	Median	0,99967	0,04739	0,99973	0,04486	0,99987	0,04088
	Huber	0,99960	0,03401	0,99960	0,03193	0,99967	0,02802
	Hampel	0,99960	0,03022	0,99957	0,02807	0,99963	0,02414
	Andrew	0,99970	0,04410	0,99970	0,04169	0,99987	0,03758
	Winsorize	0,99960	0,02870	0,99957	0,02658	0,99967	0,02263
	Trimmed	0,99960	0,02982	0,99960	0,02764	0,99967	0,02372
	Tukey	0,99970	0,04194	0,99970	0,03961	0,99987	0,03544
	HL	0,99960	0,03133	0,99960	0,02924	0,99963	0,02546
150	Median	1,00000	0,03181	1,00000	0,03653	1,00000	0,04079
	Huber	0,99997	0,01910	1,00000	0,02285	1,00000	0,02610

Hampel	0,99997	0,01945	1,00000	0,02346	1,00000	0,02708
Andrew	1,00000	0,03276	1,00000	0,03767	1,00000	0,04266
Winsorize	1,00000	0,01856	1,00000	0,02252	1,00000	0,02598
Trimmed	1,00000	0,01793	1,00000	0,02178	1,00000	0,02516
Tukey	1,00000	0,03135	1,00000	0,03615	1,00000	0,04098
HL	0,99997	0,01890	1,00000	0,02273	1,00000	0,02607

**Tablo 5.** Bağımsız Değişken Sayısı P=2, Kötü Kaldıraç Noktası 5 İçin Simülasyon Sonuçları.

n		MAD		$S_n$		$Q_n$	
		CRI	SW	CRI	SW	CRI	SW
50	Median	0,50486	0,08634	0,49186	0,08902	0,50686	0,08238
	Huber	0,44722	0,10463	0,42866	0,10804	0,43654	0,10366
	Hampel	0,40462	0,12016	0,38336	0,12426	0,39212	0,11976
	Andrew	0,45494	0,10551	0,43604	0,10946	0,44958	0,10354
	Winsorize	0,41186	0,11903	0,39092	0,12322	0,39992	0,11875
	Trimmed	0,41904	0,11599	0,3996	0,12012	0,40828	0,11539
	Tukey	0,45454	0,10562	0,43566	0,10957	0,44904	0,10368
	HL	0,4032	0,12435	0,38122	0,12862	0,39058	0,12447
80	Median	0,86792	0,02838	0,86140	0,02140	0,87270	0,02048
	Huber	0,83066	0,01128	0,81852	0,00452	0,83184	0,00327
	Hampel	0,82046	0,00311	0,80634	0,01043	0,82070	0,01224
	Andrew	0,85744	0,01255	0,84928	0,00508	0,86164	0,00345
	Winsorize	0,82594	0,00297	0,81272	0,01031	0,82664	0,01217
	Trimmed	0,82766	0,00005	0,81538	0,00726	0,82902	0,00902
	Tukey	0,85718	0,01239	0,84914	0,00494	0,86132	0,00329
	HL	0,82258	0,00643	0,80880	0,01370	0,82420	0,01566
100	Median	0,97346	0,04302	0,97200	0,03398	0,97498	0,03136
	Huber	0,96116	0,02696	0,95884	0,01861	0,96204	0,01615
	Hampel	0,95858	0,01368	0,95586	0,00453	0,95932	0,00209
	Andrew	0,97022	0,02831	0,96874	0,01872	0,97178	0,01590
	Winsorize	0,96022	0,01379	0,95770	0,00434	0,96098	0,00184
	Trimmed	0,96072	0,01661	0,95814	0,00729	0,96150	0,00481
	Tukey	0,97012	0,02780	0,96852	0,01823	0,97166	0,01543
	HL	0,95936	0,01281	0,95658	0,00346	0,96032	0,00101
120	Median	0,99122	0,02122	0,99224	0,01618	0,99296	0,01134
	Huber	0,98748	0,00857	0,98872	0,00390	0,99020	0,00039
	Hampel	0,98710	0,00049	0,98828	0,00418	0,98980	0,00845
	Andrew	0,99060	0,01425	0,99188	0,00890	0,99300	0,00391
	Winsorize	0,98758	0,00091	0,98894	0,00377	0,99038	0,00816
	Trimmed	0,98762	0,00249	0,98892	0,00221	0,99042	0,00658
	Tukey	0,99056	0,01291	0,99186	0,00758	0,99298	0,00267
	HL	0,98734	0,00266	0,98866	0,00189	0,99018	0,00621
150	Median	0,99944	0,02639	0,99940	0,02063	0,99954	0,01588
	Huber	0,99892	0,01329	0,99910	0,00847	0,99920	0,00436

Hampel	0,99894	0,00819	0,99904	0,00348	0,99912	0,00042
Andrew	0,99950	0,02200	0,99950	0,01624	0,99954	0,01156
Winsorize	0,99898	0,00739	0,99912	0,00274	0,99920	0,00114
Trimmed	0,99894	0,00866	0,99912	0,00393	0,99918	0,00005
Tukey	0,99948	0,01964	0,99948	0,01401	0,99954	0,00946
HL	0,99892	0,00941	0,99912	0,00477	0,99920	0,00085

**Tablo 6.** Bağımsız Değişken Sayısı P=3, Kötü Kaldıraç Noktası 2 İçin Simülasyon Sonuçları.

n		MAD		$S_n$		$Q_n$	
		CRI	SW	CRI	SW	CRI	SW
50	Median	0,94670	0,02235	0,94630	0,01717	0,94790	0,01304
	Huber	0,93530	0,01294	0,93430	0,00808	0,93595	0,00448
	Hampel	0,93480	0,00751	0,93410	0,00250	0,93570	0,00104
	Andrew	0,94855	0,02015	0,94715	0,01465	0,94995	0,01022
	Winsorize	0,93550	0,00723	0,93380	0,00212	0,93515	0,00150
	Trimmed	0,93510	0,00786	0,93330	0,00284	0,93530	0,00064
	Tukey	0,94790	0,01834	0,94625	0,01286	0,94935	0,00869
	HL	0,93505	0,00883	0,93340	0,00390	0,93510	0,00035
80	Median	0,96033	0,01727	0,96300	0,01351	0,96767	0,01016
	Huber	0,95067	0,00730	0,95267	0,00432	0,95700	0,00058
	Hampel	0,95133	0,00301	0,95367	0,00012	0,95867	0,00382
	Andrew	0,96433	0,01578	0,96400	0,01207	0,96833	0,00809
	Winsorize	0,95267	0,00239	0,95367	0,00077	0,95867	0,00455
	Trimmed	0,95133	0,00303	0,95333	0,00032	0,95833	0,00389
	Tukey	0,96167	0,01389	0,96400	0,01032	0,96700	0,00655
	HL	0,95133	0,00453	0,95333	0,00154	0,95800	0,00251
100	Median	0,97346	0,04302	0,97200	0,03398	0,97498	0,03136
	Huber	0,96116	0,02696	0,95884	0,01861	0,96204	0,01615
	Hampel	0,95858	0,01368	0,95586	0,00453	0,95932	0,00209
	Andrew	0,97022	0,02831	0,96874	0,01872	0,97178	0,01590
	Winsorize	0,96022	0,01379	0,95770	0,00434	0,96098	0,00184
	Trimmed	0,96072	0,01661	0,95814	0,00729	0,96150	0,00481
	Tukey	0,97012	0,02780	0,96852	0,01823	0,97166	0,01543
	HL	0,95936	0,01281	0,95658	0,00346	0,96032	0,00101
120	Median	0,99122	0,02122	0,99224	0,01618	0,99296	0,01134
	Huber	0,98748	0,00857	0,98872	0,00390	0,99020	0,00039
	Hampel	0,98710	0,00049	0,98828	0,00418	0,98980	0,00845
	Andrew	0,99060	0,01425	0,99188	0,00890	0,99300	0,00391
	Winsorize	0,98758	0,00091	0,98894	0,00377	0,99038	0,00816
	Trimmed	0,98762	0,00249	0,98892	0,00221	0,99042	0,00658
	Tukey	0,99056	0,01291	0,99186	0,00758	0,99298	0,00267
	HL	0,98734	0,00266	0,98866	0,00189	0,99018	0,00621
150	Median	1,00000	0,03409	1,00000	0,03949	1,00000	0,05044
	Huber	1,00000	0,02176	1,00000	0,02598	1,00000	0,03484

Hampel	1,00000	0,02580	1,00000	0,03058	1,00000	0,04048
Andrew	1,00000	0,03929	1,00000	0,04514	1,00000	0,05753
Winsorize	1,00000	0,02223	1,00000	0,02663	1,00000	0,03579
Trimmed	1,00000	0,02192	1,00000	0,02627	1,00000	0,03543
Tukey	1,00000	0,03782	1,00000	0,04351	1,00000	0,05572
HL	1,00000	0,02161	1,00000	0,02590	1,00000	0,03484

**Tablo 7.** Bağımsız Değişken Sayısı 3, Kötü Kaldıraç Noktası 3 İçin Simülasyon Sonuçları.

n		MAD		$S_n$		$Q_n$	
		CRI	SW	CRI	SW	CRI	SW
50	Median	0,72077	0,02164	0,71897	0,02282	0,73603	0,02334
	Huber	0,66960	0,03611	0,66683	0,03748	0,67677	0,03871
	Hampel	0,66043	0,04290	0,65787	0,04414	0,66750	0,04545
	Andrew	0,70883	0,03030	0,70863	0,03143	0,72293	0,03223
	Winsorize	0,66153	0,04300	0,65870	0,04430	0,66807	0,04578
	Trimmed	0,66270	0,04184	0,65987	0,04314	0,66913	0,04454
	Tukey	0,70713	0,03085	0,70683	0,03206	0,72120	0,03276
	HL	0,65987	0,04275	0,65673	0,04423	0,66647	0,04573
80	Median	0,96033	0,01727	0,96300	0,01351	0,96767	0,01016
	Huber	0,95067	0,00730	0,95267	0,00432	0,95700	0,00058
	Hampel	0,95133	0,00301	0,95367	0,00012	0,95867	0,00382
	Andrew	0,96433	0,01578	0,96400	0,01207	0,96833	0,00809
	Winsorize	0,95267	0,00239	0,95367	0,00077	0,95867	0,00455
	Trimmed	0,95133	0,00303	0,95333	0,00032	0,95833	0,00389
	Tukey	0,96167	0,01389	0,96400	0,01032	0,96700	0,00655
	HL	0,95133	0,00453	0,95333	0,00154	0,95800	0,00251
100	Median	0,99633	0,01029	0,99647	0,01004	0,99690	0,00864
	Huber	0,99523	0,00081	0,99537	0,00041	0,99587	0,00098
	Hampel	0,99550	0,00028	0,99557	0,00016	0,99607	0,00151
	Andrew	0,99677	0,01190	0,99707	0,01143	0,99733	0,01010
	Winsorize	0,99557	0,00053	0,99547	0,00092	0,99600	0,00226
	Trimmed	0,99537	0,00073	0,99533	0,00108	0,99587	0,00247
	Tukey	0,99657	0,00998	0,99693	0,00956	0,99720	0,00823
	HL	0,99537	0,00036	0,99537	0,00006	0,99590	0,00147
120	Median	0,99973	0,02437	0,99977	0,02590	0,99990	0,02505
	Huber	0,99963	0,01187	0,99970	0,01283	0,99983	0,01192
	Hampel	0,99967	0,01165	0,99970	0,01274	0,99980	0,01190
	Andrew	0,99980	0,02244	0,99977	0,02394	0,99990	0,02321
	Winsorize	0,99967	0,00976	0,99967	0,01074	0,99983	0,00991
	Trimmed	0,99967	0,01005	0,99970	0,01106	0,99983	0,01019
	Tukey	0,99980	0,02131	0,99977	0,02280	0,99990	0,02208
	HL	0,99967	0,01045	0,99970	0,01140	0,99983	0,01050
150	Median	1,00000	0,04815	1,00000	0,04966	1,00000	0,06103
	Huber	1,00000	0,03512	1,00000	0,03637	1,00000	0,04602

Hampel	1,00000	0,04033	1,00000	0,04170	1,00000	0,05231
Andrew	1,00000	0,05689	1,00000	0,05872	1,00000	0,07120
Winsorize	1,00000	0,03543	1,00000	0,03669	1,00000	0,04656
Trimmed	1,00000	0,03480	1,00000	0,03605	1,00000	0,04585
Tukey	1,00000	0,05420	1,00000	0,05603	1,00000	0,06839
HL	1,00000	0,03519	1,00000	0,03651	1,00000	0,04625

**Tablo 8.** Bağımsız Değişken Sayısı 3, Kötü Kaldıraç Noktası 5 İçin Simülasyon Sonuçları.

n		MAD		$S_n$		$Q_n$	
		CRI	SW	CRI	SW	CRI	SW
50	Median	0,31976	0,14600	0,31136	0,14820	0,32306	0,14361
	Huber	0,26874	0,16205	0,25526	0,16498	0,26388	0,16203
	Hampel	0,23974	0,17224	0,22568	0,17503	0,23414	0,17228
	Andrew	0,28192	0,15964	0,27060	0,16219	0,28304	0,15821
	Winsorize	0,24626	0,17091	0,23348	0,17371	0,24224	0,17090
	Trimmed	0,24984	0,16953	0,23678	0,17229	0,24636	0,16926
	Tukey	0,28104	0,15992	0,26990	0,16243	0,28226	0,15847
	HL	0,24132	0,17357	0,22728	0,17660	0,23560	0,17405
80	Median	0,71564	0,02810	0,71000	0,03108	0,73074	0,02998
	Huber	0,65924	0,04412	0,65100	0,04697	0,66882	0,04632
	Hampel	0,64692	0,05305	0,63848	0,05599	0,65606	0,05567
	Andrew	0,70020	0,03986	0,69342	0,04292	0,71444	0,04204
	Winsorize	0,65144	0,05297	0,64326	0,05597	0,66064	0,05570
	Trimmed	0,65452	0,05099	0,64692	0,05391	0,66388	0,05358
	Tukey	0,69928	0,04006	0,69296	0,04307	0,71370	0,04225
	HL	0,64932	0,05373	0,64094	0,05669	0,65824	0,05656
100	Median	0,92808	0,00256	0,92706	0,00284	0,93074	0,00553
	Huber	0,90278	0,01089	0,90042	0,01558	0,90192	0,01823
	Hampel	0,89998	0,01906	0,89734	0,02379	0,89810	0,02642
	Andrew	0,92688	0,00644	0,92578	0,01190	0,93004	0,01468
	Winsorize	0,90222	0,01894	0,89912	0,02378	0,90074	0,02637
	Trimmed	0,90218	0,01726	0,89950	0,02215	0,90098	0,02473
	Tukey	0,92630	0,00690	0,92538	0,01235	0,92954	0,01508
	HL	0,90084	0,01864	0,89860	0,02340	0,89992	0,02593
120	Median	0,98678	0,01001	0,98610	0,00590	0,98720	0,00123
	Huber	0,98170	0,00169	0,98056	0,00541	0,98202	0,00929
	Hampel	0,98124	0,00823	0,98012	0,01177	0,98140	0,01556
	Andrew	0,98698	0,00416	0,98660	0,00006	0,98728	0,00449
	Winsorize	0,98154	0,00834	0,98052	0,01194	0,98176	0,01571
	Trimmed	0,98162	0,00692	0,98066	0,01055	0,98202	0,01436
	Tukey	0,98672	0,00272	0,98630	0,00139	0,98708	0,00580
	HL	0,98138	0,00702	0,98060	0,01049	0,98176	0,01425
150	Median	0,99964	0,00255	0,99964	0,00571	0,99974	0,00730
	Huber	0,99930	0,00972	0,99948	0,01254	0,99956	0,01391
	Hampel	0,99936	0,01177	0,99948	0,01456	0,99960	0,01588

Andrew	0,99962	0,00157	0,99968	0,00509	0,99980	0,00678
Winsorize	0,99936	0,01156	0,99948	0,01441	0,99962	0,01571
Trimmed	0,99934	0,01150	0,99950	0,01429	0,99960	0,01562
Tukey	0,99962	0,00383	0,99964	0,00712	0,99980	0,00873
HL	0,99936	0,01070	0,99950	0,01350	0,99960	0,01485

**\*Bağımsız değişken sayısı iki olduğu durumda;**

Veri setinde iki kötü kaldıraç noktası olması durumunda Tablo 3 incelendiğinde örneklem hacmi  $n=50$  için CT11 eşik değeri en yüksek CRI değerine sahip olduğu görülmektedir. SW değeri, ölçek tahmin edicisi  $Q_n$  ve  $S_n$  olarak kullanıldığında MAD' a göre daha düşük değere sahip olmaktadır.  $n=80$  için ölçek tahmin edicisi olarak  $Q_n$  kullanıldığında en yüksek CRI değeri Tukey, Medyan ve Andrew tahmin edicileri ile elde edilen eşik değerler için belirlenmiştir. Bu tahmin edicilerinin SW değerleri birbirine yakındır.  $n=100$  için en yüksek CRI değeri CT5, CT14 ve CT17 eşik değerleri kullanıldığında elde edilmiştir. Bu eşik değerler kullanıldığında hesaplanan SW değerleri yaklaşık aynıdır. Örneklem hacmi 120 ve 150 olduğu durumda ise tüm eşik değerler yardımıyla elde edilen CRI değerlerinin aynı olduğu görülmektedir. Ancak CT16 eşik değeri kullanıldığında en düşük SW değerine sahip olduğu görülmektedir. Ayrıca örneklem hacmi arttıkça CRI değerinin 1'e yaklaştığı Tablo 3' de görülmektedir. Genel olarak; yeni önerilen eşik değerlerin CT eşik değerinden daha iyi sonuç verdiği görülmektedir.

Veri setinde üç kötü kaldıraç noktası olması durumunda elde edilen sonuçlar Tablo 4 de verilmiştir. Örneklem hacmi 50 için 1'e en yakın CRI değeri CT2 eşik değeri ile elde edildiği görülmektedir. Ölçek parametresi için  $S_n$  ve MAD tahmin edicileri kullanıldığında da  $Q_n$  tahmin edicisine yakın CRI değerlerinin elde edildiği görülmektedir. Andrew ve Tukey tahmin edicilerinin de Medyan tahmin edicisine yakın CRI değerine sahip tahmin ediciler olduğu tablodan görülmektedir. SW değeri, yüksek CRI değerlerine sahip olan tahmin edicilerde oldukça düşük olduğu görülmektedir. Örneklem hacmi 80,100 için en yüksek CRI değeri CT11 eşik değerinde elde edilmiştir.  $Q_n$  tahmin edicisini  $S_n$  ve MAD tahmin edicileri takip ettikleri görülmektedir. Ölçek parametresi için  $Q_n$  tahmin edicisi kullanıldığında Andrew ve Tukey tahmin edicilerinin hemen hemen yakın SW değerlerini aldıkları görülmektedir.  $n=120$  olduğu durumda ölçek parametresi için  $Q_n$  tahmin edicisi kullanıldığında Tukey, Andrew ve Medyan tahmin edicilerinin CRI değerlerinin diğer tahmin edicilere göre yüksek olduğu ve Tukey tahmin edicisinin SW değerinin daha düşük olduğu görülmektedir.  $n=120$  için en etkin eşik değerin CT20 olduğu görülmektedir. Örneklem hacmi 150 olduğu durumda konum ve ölçek parametreleri için kullanılan tahmin edicilerinin hemen hemen hepsinin CRI değeri 1' e yakın olduğu ve SW değerleri incelendiğinde ise CT15 eşik değerinin daha düşük SW değerine sahip olduğu görülmektedir.

Son olarak bağımsız derişken sayısı iki iken veri setinde beş kötü kaldıraç noktası olması durumunda Tablo 5 incelendiğinde  $n=50$  için CT2 eşik değerinin CRI değeri diğer eşik değerlere oranla daha yüksek olduğu görülmektedir.  $Q_n$  tahmin edicisini MAD ve  $S_n$  tahmin edicileri takip etmektedir.



SW değeri Medyan tahmin edicisi kullanıldığı durumda ölçek parametresi için önerilen tahmin edicilerde çok yakın değerler aldığı görülmektedir. Örneklem hacmi 80 ve 100 olduğunda CRI değerinin CT2 kullanıldığında en yüksek olduğu,  $n=120$  için CT11 eşik değerinin yüksek CRI değerine sahip olduğunu görülmektedir. Bu eşik değerinden sonra CT20 eşik değerinin en yüksek CRI değerinin verdiği görülmektedir.  $n=150$  için 1'e en yakın CRI değeri, ölçek parametresi için  $Q_n$  tahmin edicisi kullanıldığı durumda konum parametresi için ise sırasıyla Tukey, Andrew, Medyan (CT18,CT11,CT2) tahmin edicilerinin kullanıldığı eşik değerlerde elde edilmiştir. Medyan ve Andrew tahmin edicilerinin SW değerlerinin çok yakın olduğunu ve Tukey tahmin edicisinin SW değerinin bu tahmin edicilerden daha düşük olduğu Tablo 5' den görülmektedir.

**\*Bağımsız değişken sayısı üç olduğu durumda;**

Veri setinde iki kötü kaldıraç noktası olduğu durumda elde edilen sonuçlar Tablo 6 verilmiştir. Örneklem hacmi 50 ve 80 için CT11 eşik değerinin en yüksek CRI değerine sahip olduğu ve SW değeri ise diğer ölçek tahmin edicileri kullanıldığı durumdan daha düşük olduğu görülmektedir.  $n=100$  için CT2 eşik değerinin 1'e en yakın CRI değerine sahip olduğu görülür. MAD ve  $S_n$  tahmin edicileri sırasıyla  $Q_n$  tahmin edicisinden sonra en yüksek CRI değerine sahip ölçek tahmin edicileri olduğu görülmektedir. Benzer şekilde  $n=120$  için konum tahmin edicisi Andrew ve ölçek tahmin edicisi  $Q_n$  kullanıldığı durumda 1'e yakın CRI değerlerine sahip olduğu görülür. Örneklem hacmi 150 olduğu durumda ise konum ve ölçek parametreleri için kullanılan tahmin edicilerinin hepsinin CRI değerlerinin 1 olduğu ve SW değerleri incelendiğinde ise CT21 eşik değerinin düşük SW değerine sahip olduğu görülmektedir. Ayrıca ölçek parametresi MAD kullanıldığında Huber ve Trimmed konum tahmin edicilerinin de düşük SW değerlerine sahip olduğu görülmektedir.

Veri setinde üç kötü kaldıraç noktası olduğu durumda CT2 eşik değeri en yüksek CRI değerine sahip olduğu görülmektedir. Andrew ve Tukey tahmin edicilerinin de yüksek CRI değerine sahip olduğunu ve bu üç tahmin edicinin diğer tahmin edicilerden daha düşük SW değerine sahip olduğu görülmektedir. Örneklem hacmi 80 ve 100 olduğu durumda CT11 eşik değerinin en yüksek CRI değerine sahip olduğu görülmektedir. Andrew tahmin edicisinin  $S_n$  ve MAD tahmin edicileri ile kıyaslandığı durumda  $Q_n$  tahmin edicisinin daha düşük SW değerine sahip olduğu görülmektedir. Tukey tahmin edicisi de diğer konum tahmin edicilere kıyasla daha yüksek CRI değerine sahip olduğu görülmektedir.  $n=120$  için ölçek tahmin edicisi  $Q_n$  kullanıldığı durumda konum parametresi için ise sırasıyla Tukey, Andrew, Medyan ve tahmin edicilerinin kullanıldığı eşik değerlerde SW değerinin daha düşük olduğu görülmektedir. Örneklem hacmi 150 olduğu durumda konum ve ölçek parametreleri için kullanılan tahmin edicilerinin hepsinin CRI değeri 1'e yakın olduğu ve SW değerleri incelendiğinde ise CT15 eşik değerinin daha düşük SW değerine sahip olduğu görülmektedir. Huber ve HL konum tahmin edicileri de düşük SW değerine sahip olduğu Tablo 7' den görülmektedir.

Son olarak Tablo 8 incelendiğinde veri setinde beş kötü kaldıraç noktası olduğu durumda, CT2 eşik değerinin diğer tahmin edicilere kıyasla daha yüksek CRI değere sahip olduğu görülmektedir. Örneklem hacmi 80,100,120 için elde edilen sonuçların bağımsız değişken sayısı iki olduğu durumla benzer olduğu görülmektedir. n=150 için, CT11 ve CT20 eşik değerleri kullanıldığında elde edilen CRI değeri en yüksektir. Bu eşik değerlerinin SW değerleri incelendiğinde CT20 eşik değerinin daha düşük SW değerine neden olduğu görülmektedir.

## 5. GERÇEK VERİ UYGULAMASI

Çalışmamızın bu bölümünde lojistik regresyon modelinde kötü kaldıraç noktalarının belirlenmesi için gerçek veri seti uygulaması üzerine durulmuştur. İflas veri seti, iki yıllık süre sonunda batan ve iş hayatına devam eden 64 firmaya ilişkin verilerden oluşmaktadır (Chatterjee ve Hadi, 1974). Bağımlı ve bağımsız değişkenler aşağıdaki gibi tanımlanmıştır.

$$y = \begin{cases} 0, & \text{firma iki yıl sonra iflas etmiş} \\ 1, & \text{firmanın iki yıl sonra hala iş hayatına devam ediyor} \end{cases}$$

$x_1$  = Alıkonulmuş kazaçlar/Toplam varlıklar  
 $x_2$  = Faiz ve vergilerden önceki kazançlar/Toplam varlıklar  
 $x_3$  = Satış hasılatı /Toplam varlıklar

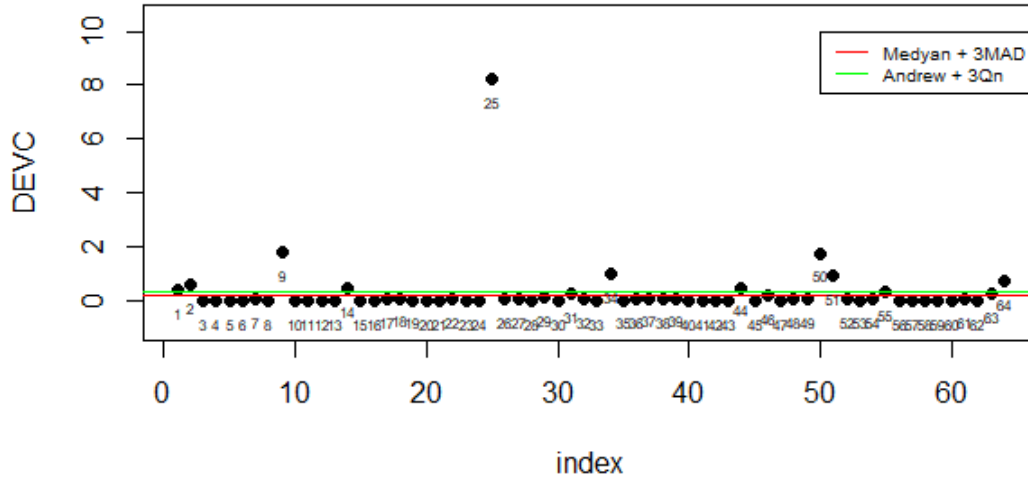
DEVC yöntemi kullanılarak elde edilen eşik değerler Tablo 9 da verilmiştir.

**Tablo 9.** Klasik eşik değer ve önerilen eşik değerler

Konum tahmin ediciler	Ölçek tahmin ediciler		
	Mad	$S_n$	$Q_n$
Medyan	0,0974	0,0861	0,0909
Huber	0,1076	0,0963	0,1011
Hampel	0,2395	0,2282	0,2330
Andrew	0,2190	0,2077	0,2125
Winsorize	0,1379	0,1266	0,1314
Trimmed	0,1534	0,1421	0,1469
Tukey	0,2223	0,2110	0,2158
HL	0,1141	0,1028	0,1076

Tablo 9 incelendiğinde DEVC sapma artıkları kullanılarak literatürde mevcut olan eşik değer ile çalışmamızda önerdiğimiz eşik değerlerin değerleri görülmektedir.

Şekil 1. DEVC Sapma Artıklarının Gözlem Numaralarına Göre Grafiği



Şekil 1 incelendiğinde literatürde mevcut olan *Medyan + 3MAD* eşik değer ile simülasyon çalışmamızda yüksek CRI değerine sahip olan *Andrew + 3Q<sub>n</sub>* eşik değeri için DEVC sapma artıklar grafik ile gösterilmiştir. *Medyan + 3MAD* eşik değeri için eşik değer çizgisi üzerinde kalan 12 gözlem ile *Andrew + 3Q<sub>n</sub>* eşik değer çizgisinin üzerinde kalan 8 gözlem bu veri seti için kötü kaldıraç noktası olarak belirlenmiştir.

## 6. SONUÇ

Teşhis yöntemleri veri setlerindeki şüpheli gözlemlerin belirlenmesinde önemli bir yere sahiptir. Lojistik regresyon modeli için veri setinde bulunan yüksek kaldıraç noktalarının belirlenmesinde DEVC yöntemi yaygın olarak kullanılmaktadır. Bu çalışmada kötü kaldıraç noktalarının belirlenmesinde DEVC yöntemi için yeni sağlam eşik değerleri ele alınmıştır. Literatürde mevcut olan *Medyan + 3MAD* eşik değeri için alternatif eşik değerlerin simülasyon sonuçları incelendiğinde bağımsız değişken sayısı iki olduğu durumda konum parametresi için Andrew, Medyan ve Tukey tahmin edicilerin değişen örneklem hacimlerinde yüksek CRI değerine sahip olduğu görülmüştür.

Bağımsız değişken sayısı üç olduğu durumda da ise  $Q_n$  tahmin edicisinin tüm örneklem hacimlerinde mevcut olan MAD tahmin edicisinden daha yüksek CRI değerine sahip olduğu görülmüştür. Veri setinde bulunan kötü kaldıraç noktası iki ve üç olduğu ve örneklem hacmi 50,80,100 ve 120 için Andrew, Medyan ve Tukey tahmin edicilerinin yüksek CRI değerlerine sahip olduğu görülmüştür.  $n=150$  için CRI değerlerinin tüm kombinasyonlarda 1 değerine eşit olduğu ve SW değerleri incelendiğinde Trimmed, Huber ve HL tahmin edicilerinin diğer tahmin edicilere göre daha düşük olduğu görülmüştür.

Simülasyon sonuçları incelendiğinde örneklem hacminin artmasıyla CRI değerlerinin de arttığı ve  $n=150$  olduğu durumda hemen hemen 1 değerini aldığı görülmektedir. SW değerlerinin ise bağımsız

değişken sayısı, örneklem hacmi ve kötü kaldıraç değerlerinin değiştiği her bir durumunda çok yüksek olmamakla birlikte farklı değerler aldığı görülmektedir.

Gerçek veri uygulamasında gözlemler incelendiğinde kötü kaldıraç noktalarına rastlanmış ve DEVC yöntemi ile kötü kaldıraç noktaları belirlenmiştir. DEVC yönteminde kötü kaldıraç noktalarını belirlenmesinde önerdiğimiz tüm eşik değerler Tablo 9 da verilmiştir. Literatürde mevcut olan *Medyan + 3MAD* eşik değer ile simülasyon çalışmamızda yüksek CRI değerine sahip olan *Andrew + 3Q<sub>n</sub>* eşik değerleri kullanarak gözlemlerin bu eşik değer çizgisinden uzaklıkları da Şekil 1 de gösterilmiştir. Böylece bu eşik değerler için DEVC yöntemi kullanılarak kötü kaldıraç noktalarının belirlenmesi sağlanmıştır.

Sonuç olarak bu çalışma ile literatürde mevcut olan *Medyan + 3MAD* eşik değerine alternatif olarak 23 farklı eşik değerinin CRI ve SW değerleri ortaya konmuştur. *Q<sub>n</sub>* ölçek tahmin edicisine dayalı eşik değerlerinin CRI değerleri literatürde kullanılan MAD ölçek tahmin edicisine dayalı eşik değerlerden daha iyi sonuçlar verdiğini görülmektedir. Farklı örneklem boyutlarında Medyan tahmin edicisi yerine alternatif olarak Andrew, Tukey tahmin edicilerin kullanılabilceğini de yine CRI ve SW değerleri ile görülmektedir.

Genel olarak simülasyon sonuçları incelendiğinde CT11 ve CT20 eşik değerlerinin kötü kaldıraç noktası tespitinde daha iyi sonuçlar verdiği görülmektedir. Özellikle örneklem hacminin 100 ‘den küçük olduğu durumlarda eşik değerlerinin elde edilmesi, kötü kaldıraç noktası tespitinde oldukça etkili olmaktadır. Bu nedenle yapılacak çalışmalarda bu eşik değerlerinin kullanılması, kötü kaldıraç noktalarının doğru bir şekilde belirlenmesini sağlayacaktır.

## KAYNAKÇA

- Alguraibawi, M., Midi, H., ve Imon, A. H. M. R. (2015). *A new robust diagnostic plot for classifying good and bad high leverage points in a multiple linear regression model*. Mathematical Problems in Engineering, 2015. <https://doi.org/10.1155/2015/279472>
- Belsley, D. A., Kuh, E., ve Welsch, R. E. (1980). *Regression Diagnostics*. <https://doi.org/10.1002/0471725153>.
- Chatterjee, S., ve Hadi, A. S. (2006). *Regression Analysis by Example*. 4th Editio. Wiley-Interscience.
- Fitrianto, A., ve Midi, H. (2010). Diagnostic-Robust Generalized Potentials for Identifying High Leverage Points in Mediation Analysis. *World Applied Sciences Journal*, 11(8), 979–987.
- Habshah, M., Norazan, M. R., ve Imon, A. H. M. R. (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics*, 36(5), 507–520. <https://doi.org/10.1080/02664760802553463>
- Hadi, A. S. (1992). A New Measure of Overall Potential Influence in Linear Regression. *Computational Statistics and Data Analysis*, 14(1), 1–27. [https://doi.org/10.1016/0167-9473\(92\)90078-T](https://doi.org/10.1016/0167-9473(92)90078-T)

- Hampel, F. R. (1974). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69(346), 383–393.
- Hastings, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. (1947). Low Moments for Small Samples: A Comparative Study of Order Statistics. *The Annals of Mathematical Statistics*, 18(3), 413–426. <https://doi.org/10.1214/aoms/1177730388>
- Hoaglin, D. C., ve Mosteller, F. (1983). *Understanding robust and exploratory data analysis*,. New York, NY: John Wiley & Sons.
- Hodges, J. L. J., ve Lehmann, E. L. (1963). Estimates of Location Based on Ranks Tests. *Annals of Mathematical Statistics*, 34(2), 598–611.
- Huber, P. j. (1953). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101. <https://doi.org/10.1214/aoms/1177705148>
- Imon, A. H. M. . (2006). Identification of High Leverage Points in Logistic Regression. *Pakistan Journal of Statistics*, 22(2), 147–156.
- Imon, A. H. M. R., ve Hadi, A. S. (2013). Identification of Multiple High Leverage Points in Logistic Regression. *Journal of Applied Statistics*, 40(12), 2601–2616. <https://doi.org/10.1080/02664763.2013.822057>
- Maronna, R. A., ve Yohai, V. J. (1901). Estimates of Regression and Scale The Breakdown Point of Simultaneous General M Estimates of Regression and Scale. (July 2015), 1–6. <https://doi.org/10.1080/01621459.1991.10475097>
- Maronna, R. A., Martins, D. R., ve Yohai, V. J (2006). *Robust Statistics: Theory and Methods*, 1st Editio. Wiley.
- Norazan, M. R., Sanizah, A., ve Habshah, M. (2012). Identifying Bad Leverage Points in Logistic Regression Model Based on Robust Deviance Components. *Mathematical Models and Methods in Modern Science*, 62–67. Retrieved from <http://www.wseas.us/e-library/conferences/2012/Porto/MAMECTIS/MAMECTIS-09.pdf>
- Nurunnabi, A. A. M., Imon, A. H. M. R., ve Nasser, M. (2010). Identification of Multiple Influential Observations in Logistic Regression. *Journal of Applied Statistics*, 37(10), 1605–1624. <https://doi.org/10.1080/02664760903104307>
- Rousseeuw, P. J., ve Christophe, C. (1993). Alternatives to the Median Absolute Deviation. 88(424).
- Sarkar, K., Midi, H., ve Rana, S. (2011). Detection of outliers and influential observations in binary Logistic regression: An empirical study. *Journal of Applied Sciences*, 11(1), 26–35. <https://doi.org/10.3923/jas.2011.26.35>
- Syaiba, B. A., ve Habshah, M. (2010). Robust Logistic Diagnostic for the Identification of High Leverage Points in Logistic Regression Model. *Journal of Applied Sciences*, 10(23), 3042–3050.
- Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Commun. Stat. - Theory Methods*, 6(9), 813–827. doi: 10.1080/03610927708827533.
- Yavuz, A. A. (2013). Estimation of the shape parameter of the Weibull distribution using linear regression methods: Non-censored samples. *Quality and Reliability Engineering International*, 29(8), 1207–1219. <https://doi.org/10.1002/qre.1472>
- Yüksel, G. ve Çetin, M. (2016). Outlier detection in a preliminary test estimator of the mean. *J. Stat. Manag. Syst.*, 19(4), 605–615. doi: 10.1080/09720510.2016.1139851.