GÜFBED/*GUSTIJ* (2019) 9 (3): 393-400
DOI: 10.17714/gumusfenbil.449411 Araştırma Makalesi / *Research Article*

Çoklu Doğrusal Regresyon Analizinde Etkili Gözlemlerin Belirlenmesine Yönelik Bir Yöntem

A Method to Detect Influential Observations in Multiple Linear Regression Analysis

Osman Ufuk EKİZ*
Gazi University, Faculty of Sciences, Department of Statistics, 06500, Ankara

• Geliş tarihi / *Received*: 31.07.2018 • Düzeltilerek geliş tarihi / *Received in revised form*: 15.12.2018 • Kabul tarihi / *Accepted*: 21.12.2018

Öz
Çoklu doğrusal regresyon analizinde aykırı, etkili ve kaldıraç noktaları belirlemek istatistiksel çıkarsamaların doğruluğu açısından son derece önemlidir. Nurunnabi vd. (2016) tarafından sağlam etkili uzaklık *(EU)* ölçüsü regresyon analizinde etkili gözlemlerin belirlenmesi için önerilmiştir. Ancak bu yöntemde hesaplamalarda kullanılmayacak gözlemlerin belirlenmesi sağlam olmayan istatistiklere dayanmaktadır. Dolayısıyla bu yöntem aykırı gözlemlerden etkilenecektir. Bu çalışmada sağlam tahmin edicilere dayalı etkili uzaklık *(SEU)* ölçüsünün etkili gözlemleri belirlemekte kullanılması önerilmiştir. Ayrıca etkili gözlemleri belirlemekte *EU* ve *SEU* 'ların iyi bilinen iki gerçek veriye uygulanması ve simülasyon çalışması ile karşılaştırılmaları gerçekleştirilmiştir. Bu yöntemler içerisinde en iyi sonuçlar yeniden ağırlıklandırılmış en küçük kareler (*YEKK*) sağlam tahmin edicisine dayalı *SEU* 'lar üzerinden elde edilmiştir.

Anahtar kelimeler: Aykırı Gözlem, Etkili Gözlem, Kaldıraç, Sağlam Tahmin Edici

Abstract
It is so important to determine outlier, influence and leverage points in multiple linear regression analysis for the accuracy of statistical inferences. To detect the influence observations, Nurunnabi et al. (2016) proposed a robust influence distance (ID). However, the determination of observations that would not be used in the calculations of this distance are based on non-robust statistics. Thus, it is affected by outliers. In this paper, it is suggested that influence distance based on robust estimators (RID) could be used for detecting influence observations. Moreover ID and RID's which were used to determine outliers, are applied to two known data sets and are compared based on simulation studies. The results show that RID based on RLS performs the best.

Keywords*: Outlier, Influential observations, Leverage, Robust Estimator*

* Osman Ufuk EKİZ; ufukekiz@gazi.edu.tr; Tel: (0505) 319 6087; orcid.org/0000-0002-4004-0336

1. Giriş

Çoklu doğrusal regresyon, istatistikte bağımlı değişken ile açıklayıcı değişkenler arasındaki ilişkinin modellenmesi yöntemidir. Bu yöntem, mühendislik, fizik, kimya, biyoloji, eğitim ve sosyal bilimler gibi farklı alanlarda kullanılmaktadır. Bu model

$$Y = X\beta + \varepsilon \qquad (1)$$

şeklinde ifade edilmektedir. Burada Y, $n\times 1$ boyutlu bağımlı değişken vektörü, X, $(n\times p)$ boyutlu tam ranklı açıklayıcı değişkenler matrisi, ε ise $E(\varepsilon)=0$ ve $Cov(\varepsilon)=\sigma^2 I$ olan $n\times 1$ boyutlu normal dağılıma sahip hata vektörüdür. β, $(p\times 1)$ boyutlu bilinmeyen parametreler vektörü ve σ^2 bilinmeyen varyanstır. Burada ℓ açıklayıcı değişken sayısı olmak üzere $p=\ell+1$ dir. Tüm bu varsayımların geçerliliği altında β parametresinin en çok olabilirlik ($EÇO$) tahmin edicisi,

$$\hat{\beta}_{EÇO} = (X'X)^{-1}X'Y \qquad (2)$$

şeklinde ifade edilir. Bu tahmin edicinin en iyi, doğrusal ve sapmasız olduğu bilinmektedir (Graybill, 1976). Ancak aykırı gözlemler modelin yukarıda ifade edilen varsayımlardan sapmalar göstermesine, dolayısıyla (2) 'deki tahmin edicinin kötü sonuçlar vermesine neden olabilmektedir (Rousseeuw ve Leroy, 1987). Bu sebeple aykırı gözlemler regresyon analizi literatüründe oldukça geniş yer tutmaktadır (Barnett vd., 1994; Cook vd., 1982). (2)'deki tahmin edici üzerinde kötü etkilere neden olan aykırı gözlemler etkili gözlemler (EG) olarak isimlendirilmektedir. EG 'ler bağımlı değişken değerleri diğer gözlemlerden oldukça farklı olan gözlemlerdir. Bu gözlemler açıklayıcı değişken değerlerinin diğerlerinden çok farklılık gösterip göstermemesine göre Y yönünde veya X yönünde aykırı olarak da isimlendirilmektedir. X yönünde aykırı olan gözlemler literatürde kötü kaldıraç gözlem olarak da bilinir (Rousseeuw ve Leroy, 1987). EG 'lerin model üzerindeki etkisini en aza indirgemek için literatürde önerilmiş yöntemler iki temel gruba ayrılır. Bunlardan birincisi teşhise dayalı yöntemlerdir. Bu yöntemler, aykırı gözlemler veriden çıkarıldıktan sonra (2)'deki tahmin edicinin kullanılması fikrine dayanır. İkinci grupta yer alan yöntemler ise sağlam olarak isimlendirilir. Burada temel fikir aykırı olabilecek gözlemlerin model parametre tahmin edicisi üzerinde en az etkiye sebep olacak şekilde ağırlıklandırılmasıdır.

Nurunnabi vd. (2016) EG 'lerin belirlenmesinde etki uzaklığı (EU) 'yu önermişlerdir. i. gözlem için etki uzaklığı,

$$EU_i = \sqrt{(G_i-\bar{G}_R)\hat{\Sigma}_{G_R}^{-1}(G_i-\bar{G}_R)},\ i=1,2,3,...,n \qquad (3)$$

şeklinde tanımlanmaktadır. Burada R, D 'nin elemanı olmayan gözlemlerden oluşmaktadır. D ise genelleştirilmiş student tipi artıklar (r_i), genelleştirilmiş kaldıraç değerleri (h_i) ve Welsch-Kuh uzaklığı ($DIFITS_i$) gibi sırasıyla aykırı, kaldıraç ve EG 'leri belirlemek üzere önerilmiş çeşitli istatistiklerin kullanılmasıyla tespit edilen gözlemler kümesini ifade etmektedir. Bu istatistiklerin tanımları Cook vd., (1982) çalışmada yer almaktadır. Bu kümelemeye göre X ve Y,

$$X=\begin{bmatrix} X_R \\ X_D \end{bmatrix} \quad \text{ve} \quad Y=\begin{bmatrix} Y_R \\ Y_D \end{bmatrix}$$

şeklinde tanımlanır. R 'de yer alan gözlemler üzerinden parametre tahmini,

$$\hat{\beta}_R = (X_R'X_R)^{-1}X_R'Y_R \qquad (4)$$

olarak ifade edilmektedir. G, ilk sütunu (4)'e dayalı elde edilmiş genelleştirilmiş student tipi artıklardan (r_i), ikinci sütunu genelleştirilmiş kaldıraç (h_i) değerlerinden oluşan $n\times 2$ boyutlu matristir. $\bar{G}_R$ ve $\hat{\Sigma}_{G_R}^{-1}$ sırasıyla G 'nin R kümesinde yer alan gözlemler üzerinden hesaplanan ortalama ve kovaryans matrisinin tersini ifade etmektedir. Buna göre eğer

$$EU_i > \sqrt{\frac{(n-1)}{(n-p)}F_{\alpha,(p,n-p)}} \quad ,\ i=1,2,3,...,n$$

sağlandığında $i.$ gözlem etkili olarak değerlendirilir. Burada $F_{\alpha,(p,n-p)}$, p ve $(n-p)$ serbestlik dereceli F dağılımı üzerinden birinci tip hata miktarı α 'ya karşılık gelen kritik değeri ifade etmektedir. Ayrıca, r_i 'ler üzerinden aykırı gözlem değerlendirmesinde $medyan(r_i) \mp 3(medyan(|r_i - medyan(r_i)|)/0.6745)$, h_i değerleri üzerinden kaldıraç gözlemlerin değerlendirmesinde

$medyan(h_i) \mp 3(medyan(|h_i - medyan(h_i)|)/0.6745)$ kritik değerleri kullanılmaktadır (Nurunnabi vd., 2016).

Ancak D'de yer alan gözlemlerin belirlenmesinde kullanılan yöntemlerin aykırı gözlemlerden etkilenmesi, yanlış gözlemlerin D kümesinde yer almasına neden olabilir. Bu sebeple gerçek aykırı, kaldıraç ve etkili gözlemlerin D kümesinde yer alması olasılığı düşük olacaktır. Bu çalışmada söz konusu olasılığı daha yüksek tutmak için (4) nolu tahmin ediciye dayalı r_i ve h_i değerlerini kullanmak yerine en küçük medyan kareler (EMK) ve yeniden ağırlıklandırılmış en küçük kareler ($YEKK$) gibi sağlam tahmin edicilere dayalı elde edilen r_i ve h_i değerlerinin kullanılması önerilmiştir. Böylece D kümesinde sağlam bir tahmin edici tarafından aykırı ya da kaldıraç olarak belirlenmiş gözlemler yer almış olur. Önerilen bu yöntemle D kümesinde yer alan gözlemlerin gerçek aykırı ve kaldıraç olan gözlemlerden oluşması olasılığı artar. Dolayısıyla sağlam istatistiklere dayalı oluşturulan R kümesi üzerinden hesaplanan sağlam etkili uzaklık (SEU) değerleri ile EG'ler daha doğru tespit edilir.

Yöntemin uygulamasını ve geçerliliğini ortaya koymak açısından ikinci bölümde EMK ve $YEKK$ sağlam tahmin edicilerine yer verilmiştir. 3. bölümde önerilen yöntem ve bu yöntem üzerinden etkili gözlemlerin nasıl belirlendiği anlatılmıştır. Yöntemin gerçek veriler üzerine uygulanmasına 4. bölümde ve performansının değerlendirilmesi için bir simülasyon çalışmasına da son bölümde yer verilmiştir.

2. Çoklu Doğrusal Regresyonda Bazı Sağlam Tahmin Ediciler

(1) nolu eşitlikteki β parametre vektörüne ilişkin bilinen sağlam tahmin edicilerden ikisi EMK ve $YEKK$ aşağıdaki gibi özetlenmektedir (Rousseeuw ve Leroy, 1987).

2.1. EMK Tahmin Edici

r_i, i. Gözleme ilişkin artık olmak üzere, $(i = 1,2,...,n)$,

i. n adet gözlemden p tanesi rastgele seçilir.

ii. Amaç fonksiyonu $\min_{\hat{\beta}_t} \left\{ medyan_i \left(r_i^2 \right) \right\}$,

iii. $(t = 1,2,...,m)$ olmak üzere $\hat{\beta}_t$, n'nin p'lik kombinasyonlarından rastgele seçilen m tanesinden t. İçin elde edilen $EÇO$ tahmini ifade etmektedir. Amaç fonksiyonunu minimize eden $\hat{\beta}_t$, $\hat{\beta}_{EMK}$ olarak kabul edilir.

iv. $S^0 = 1.4826(1 + \frac{5}{n - p}) \sqrt{medyan_i(r_i^2)}$

hesaplanır.

v. $w_i = \begin{cases} 1, & |r_i / S^0| \le 2.5 \\ 0, & diğer \quad hallerde \end{cases}$

ağırlıkları elde edilir.

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

ile ölçek parametresinin EMK tahmini belirlenir.

2.2. YEKK Tahmin Edici

$\hat{\beta}_{EMK}$ tahmin başlangıç değeri olarak kullanıldığında, amaç fonksiyonu $\min_{\hat{\beta}_k} \sum_{i=1}^{n} w_i r_i^2$ olmak üzere k. Adımda elde edilen $\hat{\beta}_k$, $\hat{\beta}_{YEKK}$ olarak kabul edilir. Burada w_i, EMK'nin iv. Adımından belirlenir. Ancak bu adımda S^0 yerine v. Adımda bulunan σ^* (σ'nın EMK tahmini) kullanılır. Yani,

$$w_i = \begin{cases} 1, & |r_i / \sigma^*| \le 2.5 \\ 0, & diğer \quad hallerde \end{cases}$$

olur. Ölçek parametresi σ'nın $YEKK$ tahmin edicisi de

$\hat{\sigma}_{YEKK} = \sqrt{\frac{\sum_{i=1}^{n} w_i r_i^2}{\sum_{i=1}^{n} w_i - p}}$ şeklinde belirlenir.

3. Sağlam Tahmin Edicilere Dayalı *EU (SEU)* Üzerinden Gözlemlerin Sınıflandırılması

Nurunnabi vd. (2016), D'de yer alacak şüpheli gözlemleri robust olmayan yöntemler kullanarak belirlemektedir. Bu yöntemlerin kendileri zaten aykırı gözlemlerden etkilenir olduklarından, yanlış gözlemlerin D kümesinde yer almasına neden olabilirler. Dolayısı ile aykırı, kaldıraç ve etkili gözlemlerin doğru belirlenmesi olasılığı düşük olacaktır.

Bu çalışmada söz konusu olasılığı daha yüksek tutmak için (4) nolu tahmin ediciye dayalı r_i ve h_i değerlerini kullanmak yerine *EMK* ve *YEKK* gibi sağlam tahmin edicilere dayalı olanları kullanmak önerilmiştir. Bu değerler, sağlam tahmin edicilerin hesaplanmasında kullanılan ve en son iterasyonda ağırlığı $w_i = 1$ olan, gözlemlerden oluşan küme R olmak üzere,

$$r_i = Y_i - X_i\hat{\beta}_R$$
$$h_i = X_i'\left(X_R{}'X_R\right)^{-1}X_i \qquad (5)$$

ile hesaplanır (Georgios, 2013). *SEU* 'da (3) nolu eşitlikteki *EU* 'ya benzer şekilde oluşturulmaktadır. Ancak burada G matrisinin sütunları sırasıyla (5) nolu eştlikteki r_i ve h_i ' ler den oluşmaktadır. Buna göre eğer

$$SEU_i > \sqrt{\frac{(n-1)}{(n-p)}F_{\alpha,(p,n-p)}} \quad , \quad i = 1,2,3,...,n$$

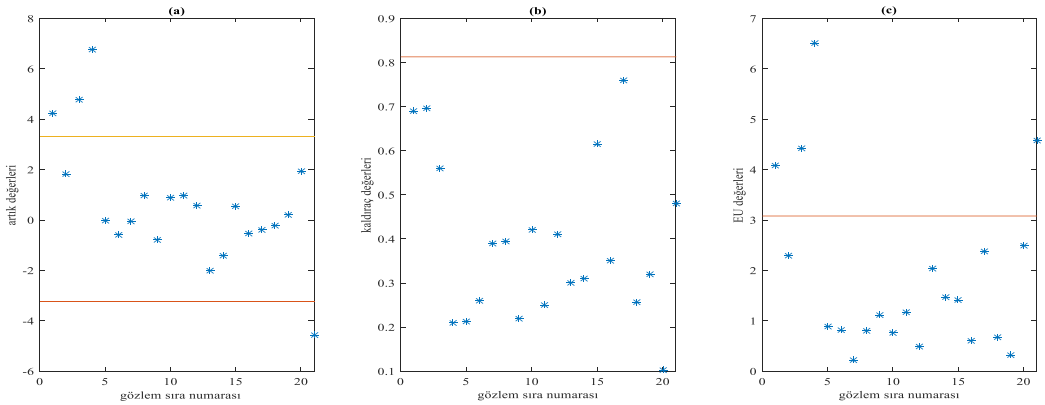
oluyorsa $i.$ gözlem etkili olarak değerlendirilmektedir. r_i ve h_i değerleri için kritik değerler Bölüm 1'de verildiği gibidir.

4. Gerçek Veri Uygulamaları

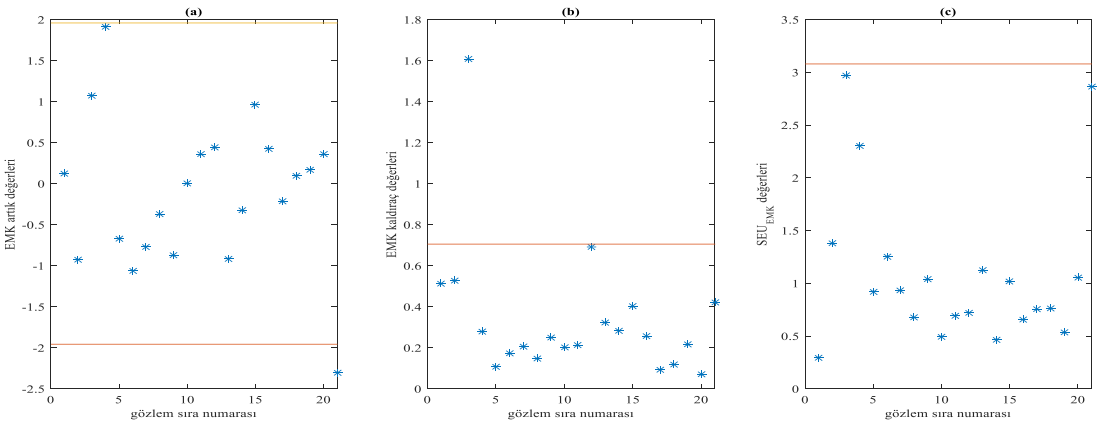
Bu bölümde *EU* ve *SEU* 'lar literatürde sık kullanılan veriler üzerine uygulanmıştır. *SEU* 'lar *EMK* ve *YEKK* sağlam tahmin edicilere dayalıdır.

4.1. Stack Loss Verisi

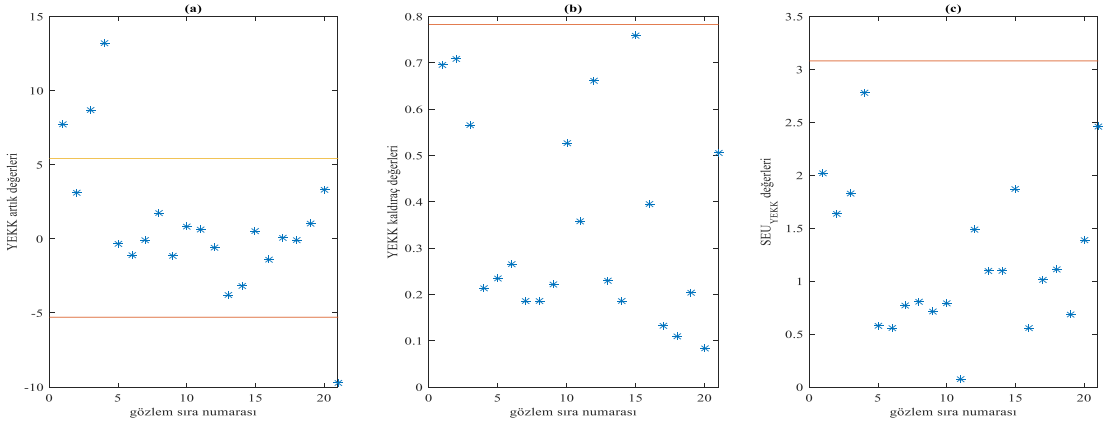
İlk örnek literatürde sıkça kullanılan ve Rousseeuw ve Leroy (1987) çalışmasında yer alan "stack loss" verisidir. Bu veri seti 21 gözlemden oluşmaktadır. Ayrıca hava akımı, soğutma suyu ve asit yoğunluğu isimli üç açıklayıcı değişken içermektedir. Rousseeuw ve van Zomeren (1990) tarafından gerçekleştirilen çalışmada 3 tane X yönünde (1, 3 ve 21 numaralı gözlemler) ve bir tane Y yönünde aykırı gözlemin olduğunu ancak etkili gözlemin varlığına dair bir göstergenin olmadığını belirtmişlerdir. Nurunnabi vd. (2016) tarafından yapılan çalışmada D kümesinde 1-4 ve 21 sıra numaralı beş gözleme yer verilmiştir. Bu beş gözlem r_i, h_i ve $DIFITS_i$ gibi farklı yöntemlerce tespit edilmiş gözlemlerin bir araya getirilmesiyle oluşturulmuştur. Ancak, bu şekilde oluşturulan bir küme içerisinde gerçekte aykırı yada kaldıraç olmayan gözlemlerin bulunması ihtimali yüksek olabilir. Çünkü kullanılan yöntemler sağlam değildir. Bu çalışmada *EU* 'lar üzerinden 1, 3, 4 ve 21 numaralı gözlemler etkili olarak belirlenmiştir (Şekil 1). Önerilen yönteme göre Şekil 2'deki *EMK* 'ya dayalı sağlam etkili uzaklık (SEU_{EMK}) değerlerinden veride etkili gözlem bulunmadığı, *EMK* 'nın r_i değerlerinden 21 numaralı gözlemin aykırı olduğu tespit edilmiştir. *EMK* 'nın h_i değerlerinden ise 3. Gözlemin kaldıraç olduğu anlaşılmaktadır. Şekil 3 'teki *YEKK'nın* r_i değerlerinden görüleceği üzere 1, 3, 4 ve 21 numaralı dört gözlem aykırı olarak tespit edilmiştir. Ayrıca *YEKK* 'ya dayalı sağlam etkili uzaklık (SEU_{YEKK}) değerlerinden veride etkili gözlem bulunmadığı anlaşılmaktadır. *YEKK* tahmin edici üzerinden elde edilen bu sonuçlar Rousseeuw ve van Zomeren (1990)'nin yapmış oldukları çalışmadakilerle aynıdır.



Şekil 1. Nurunnabi vd. (2016)'de önerilen tahmin ediciden elde edilen **(a)** artık (r_i), **(b)** kaldıraç (h_i) ve **(c)** *EU* değerlerine ilişkin serpilme diyagramları.



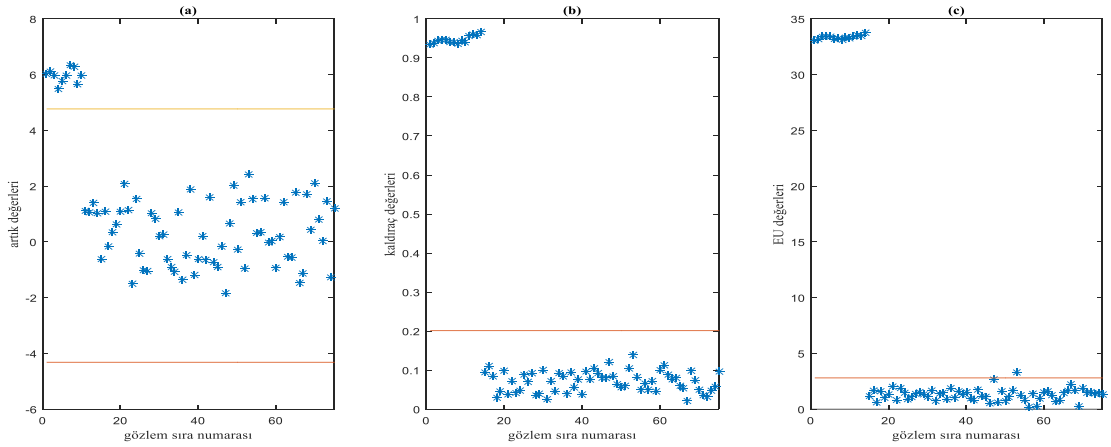
Şekil 2. *EMK* tahmin edici üzerinden elde edilen **(a)** *EMK* artık (r_i), **(b)** *EMK* kaldıraç (h_i) ve **(c)** SEU_{EMK} değerlerine ilişkin serpilme diyagramları.



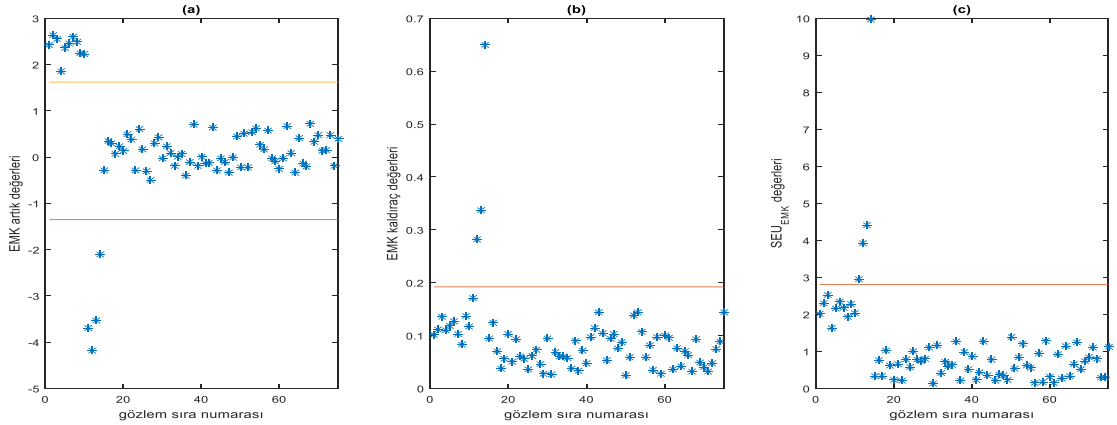
Şekil 3. *YEKK* tahmin edici üzerinden elde edilen **(a)** *YEKK* artık (r_i), **(b)** *YEKK* kaldıraç (h_i) ve **(c)** SEU_{YEKK} değerlerine ilişkin serpilme diyagramları.

4.2. Hawkins Bradu Kass Verisi

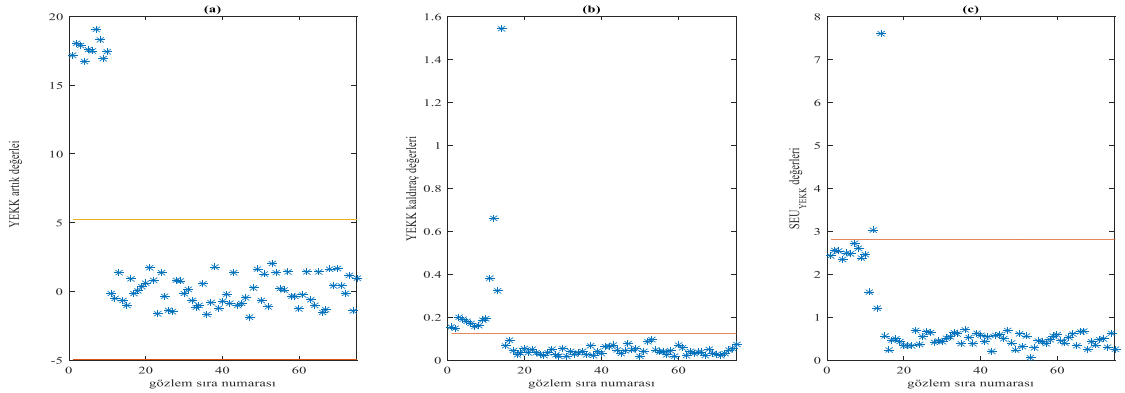
Literatürde sık kullanılan ve Rousseeuw ve Leroy (1987)'de yer alan ikinci veri seti "Hawkins Bradu Kass" olarak isimlendirilmektedir. 75 gözlemden oluşan bu veri setinde ilk 10 gözlem X yönünde aykırı (kötü kaldıraç nokta) ve 11-14 sıra numaralı gözlemler ise kaldıraç gözlem olarak bu yapay veriye eklenmiştir. Nurunnabi vd. (2016)'de yapılan çalışmada 1-14 numaralı gözlemler D kümesine alınmıştır. Bu çalışmada hangi gözlemlerin D kümesine alınması gerektiğine dair bir keyfiyet söz konusudur. Sonuçta 1-14 numaralı gözlemlerin etkili oldukları Şekil 4'teki *EU* değerlerinden anlaşılmaktadır. Şekil 5'teki *EMK'nın* r_i değerlerinden 1-14 numaralı gözlemler aykırı olarak tespit edilmiştir. *EMK'nın* h_i değerlerinden 12, 13 ve 14 numaralı gözlemler kaldıraç olarak belirlenmiştir. 11, 12, 13 ve 14 numaralı gözlemlerin ise SEU_{EMK}'lar üzerinden etkili bulundukları Şekil 5'te görülmektedir. Şekil 6'daki *YEKK'nın* r_i değerlerinden 1-14 numaralı gözlemlerin aykırı, YEKK'nın h_i değerlerine göre de 1-14 numaralı gözlemlerin kaldıraç olduğu görülmektedir. Ancak SEU_{YEKK} değerlerinden sadece 12 ve 13 numaralı gözlemlerin etkili olduğu sonucuna ulaşılmaktadır.



Şekil 4: Nurunnabi vd. (2016)'de önerilen tahmin ediciden elde edilen **(a)** artık (r_i), **(b)** kaldıraç (h_i) ve **(c)** *EU* değerlerine ilişkin serpilme diyagramları.



Şekil 5: *EMK* tahmin edici üzerinden elde edilen **(a)** *EMK* artık (r_i), **(b)** *EMK* kaldıraç (h_i) ve **(c)** SEU_{EMK} değerlerine ilişkin serpilme diyagramları.



Şekil 6: *YEKK* tahmin edici üzerinden elde edilen **(a)** *YEKK* artık (r_i), **(b)** *YEKK* kaldıraç (h_i) ve **(c)** SEU_{YEKK} değerlerine ilişkin serpilme diyagramları.

Yukarıda gerçekleştirilen iki farklı uygulamadan görüleceği üzere, yöntemler hangi gözlemlerin etkili olduğu hususunda farklılıklar göstermektedir. Hangisinin daha iyi olduğunu ortaya koyabilmek için bir sonraki bölümde bir simülasyon çalışması gerçekleştirilmiştir.

5. Simulation

Bu bölümde *EU*, SEU_{EMK} ve SEU_{YEKK}'nın karşılaştırmasını doğru belirleme oranı (*DBO*) bakımından yapabilmek için bir simülasyon çalışması gerçekleştirilmiştir. Burada *DBO*, eklenen etkili gözlemlerin tam olarak doğru tespit edildiği, *n* çaplı örneklerin adedinin 10000 deneme içindeki oranını ifade etmektedir. Bu çalışmada *EU*'ların elde edilmesinde *D* kümesinde yer alacak gözlemler (4) nolu eşitliğe bağlı r_i, h_i ve $DIFITS_i$ değerleri üzerinden belirlenmiştir. Ayrıca örnek çapları 50, 100 ve 500 açıklayıcı değişken sayısı 2, 4 ve 6, örneğe eklenen *EG* oranı λ 0.05 ve 0.10 olarak alınmıştır. Gözlemlerin $n-[n\lambda]$ adedinin *X* değişken değerleri $Normal(3,1)$ dağılımdan üretilmiştir. Bunlara bağlı bağımlı değişken değerleri de,

$$Y=\beta_0+\beta_j X_j+\varepsilon \quad , \quad j=1,2,...,\ell$$

modeli üzerinden üretilmiştir. $[.]$ virgülden sonraki birinci haneye göre bir alt yada üst tamsayıya yuvarlamayı ifade etmektedir. $\beta_0=1$ ve $\beta_j=1$ olmak üzere ε rastgele değişkeninin dağılımı da $Normal(0,0.5)$ olarak alınmıştır. Örneğe; λ_X her bir döngüde $0<\lambda_X<\lambda$ aralığında tekdüze dağılımdan üretilmiş bir değer olmak üzere $[n\lambda_X]$ adet *X* yönünde aykırı (kötü kaldıraç); λ_Y her bir döngüde $0<\lambda_Y<\lambda-\lambda_X$ aralığında tekdüze dağılımdan üretilmiş bir değer olmak üzere $[n\lambda_Y]$ adet *Y* yönünde aykırı; $[n\lambda]-[n\lambda_X]-[n\lambda_Y]$ adet kaldıraç (iyi kaldıraç) gözlem eklenmiştir. *X* yönünde aykırı gözlemler de *X* değerleri $Normal(K,0.5)$, buna bağlı *Y* değerleri de $Normal(\beta_0+\beta_i X_i+K,0.5)$ dağılımından üretilmiştir. Burada *K* değerleri,

$$P(K=k)=\begin{cases}0.5, & k=-10\\ 0.5, & k=10\end{cases}$$

dağılımından üretilmiştir. *Y* yönünde aykırı gözlemler de, *X* değerleri $Normal(3,1)$, buna bağlı *Y* değerleri de $Normal(\beta_0+\beta_i X_i+K,0.5)$ dağılımından üretilmiştir. Kaldıraç gözlemlerde de *X* değerleri $Normal(K,0.5)$, buna bağlı *Y* değerleri de $Normal(\beta_0+\beta_i X_i,0.5)$ dağılımından üretilmiştir. Tablo 1'deki *DBO* sonuçları 10000 döngü üzerinden elde edilmiştir. Tablo 1'de *SEU* 'ların ve bunlar arasında da SEU_{YEKK}'nin daha iyi sonuçlar verdiği görülmektedir. Açıklayıcı değişken sayısı arttıkça *DBO* 'larının azaldığı ve örnek çapı arttığında bu oranda az da olsa bir azalmanın olduğu yine Tablo 1'deki sonuçlardan gözlenmektedir.

Tablo 1: *DBO* değerleri.

		$\ell=2$			$\ell=4$			$\ell=6$		
n	λ	*EU*	SEU_{EMK}	SEU_{YEKK}	*EU*	SEU_{EMK}	SEU_{YEKK}	*EU*	SEU_{EMK}	SEU_{YEKK}
50	0.05	0.1821	0.2121	0.2692	0.1583	0.1804	0.2001	0.1229	0.1404	0.1594
	0.10	0.1595	0.1987	0.2463	0.1207	0.1553	0.1862	0.1008	0.1353	0.1503
100	0.05	0.1727	0.1845	0.2567	0.1471	0.1612	0.1837	0.1149	0.1282	0.1472
	0.10	0.1625	0.2128	0.2483	0.1388	0.1591	0.1752	0.0927	0.1201	0.1388
500	0.05	0.1585	0.2204	0.2584	0.1441	0.1754	0.1791	0.1094	0.1154	0.1254
	0.10	0.1461	0.1739	0.2327	0.1285	0.1437	0.1608	0.0858	0.1137	0.1177

6. Sonuçlar

Bu çalışmada sağlam tahmin edicilere dayalı etkili uzaklık (*SEU*) ölçüsünün etkili gözlemlerin belirlenmesinde kullanılması önerilmiştir. Gerçek veri üzerinde yapılan uygulamalarda *EU* ve önerilen *SEU*'ların farklı sonuçlar verdiği gözlenmiştir. *SEU* 'ların elde edilmesinde sağlam tahmin edicilerin hesaplanmasında kullanılan (yani sağlam bir tahmin edici üzerinden aykırı olmadığı tespit edilen) gözlemlere yer verildiği için *DBO* 'larının daha yüksek çıktığı yapılan simülasyon çalışması üzerinden görülmüştür. *EMK* ve *YEKK* sağlam tahmin edicilerine dayalı *SEU* 'lar üzerinden elde edilen *DBO* 'lar daha yüksek çıkmıştır. Ayrıca *DBO* bakımından SEU_{YEKK}, SEU_{EMK}'ya göre daha iyi sonuç vermektedir.

Referanslar

Barnett, V. and Lewis, T., 1994, Outliers in Statistical Data. New York: John Wiley & Sons.

Cook, R. D. and Weisberg, S., 1982, Residuals and Influence in Regression (Chapman & Hall, New York.

Georgios Pitselis, 2013, A review on robust estimators applied to regression credibility., Journal of Computational and Applied Mathematics. 239, 231-249.

Graybill, F. A. 1976. Theory and Application of the Linear Model, North Scituate, Mass. Duxbury Press.

Nurunnabi A.A.M., M. Nasser and A.H. M. R. Imon, 2016. Identification and classification of multiple outlers, high leverege points and influential observations in linear regresion., Journal of Applied Statistics, Vol. 43, No. 3, 509-525.

Rousseeuw P. J. and B.C. van Zomeren. 1990, Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85:633–651.

Rousseeuw PJ, Leroy AM., 1987, Robust regression and outlier detection. NewYork: Wiley Interscience.