Volume: 09 Number: 01 Category: 02 Page: 45-52 ISSN No: 1303-6319 Cilt: 09 Sayı: 01 Kategori: 02 Sayfa: 45-52

REGRESYON ANALİZİNDE GÖZLEMLERİN AYKIRI DEĞER HARİTASI İLE SINIFLANDIRILMASI

Yasemin KAYHAN ATILGAN* **Süleyman GÜNAY****

ÖZET

Uygulamalarda, üzerinde çalışılan çok boyutlu veri kümeleri, genellikle verinin çoğunluğuna uymayan aykırı gözlemler içerir. Regresyon analizinin önemli aşamalarından bir tanesi de artık analizi ile bu aykırı gözlemleri doğru belirlemektir. Ancak, bu amaçla kullanılan klasik istatistiksel yöntemler aykırı değerlerden çok fazla etkilenir. Dolayısıyla klasik tahmin edicilere dayalı, artık analiz teknikleri araştırmacıyı yanlış yönlendirebilir. Bu çalışmada, çok boyutlu veri kümesindeki gözlemleri incelemek için kullanılan ve klasik tahmin ediciler yerine sağlam tahmin edicilere dayalı olarak oluşturulan aykırı değer haritası basitçe açıklanmıştır. Çalışmanın amacı ise, farklı tahmin ediciler kullanılarak oluşturulan regresyon modelleri ve bu modellere bağlı olarak elde edilen haritaları karşılaştırarak, hangi tahmin edicinin daha güvenilir aykırı değer haritası oluşturacağını tartışmaktır.

Anahtar Kelimeler: Aykırı değer, Sağlam regresyon, Sağlam tahmin ediciler, Uç gözlem.

1. GİRİŞ

Regresyon analizinde veriyi modellemeye geçmeden önce uygulanacak istatistiksel analiz yöntemlerinin geçerliliğini garanti altına alan bir takım varsayımlar vardır. Varsayımların sağlanmadığı durumlarda ilk olarak 'artık analizi' ile varsayım bozulumu yaratan gözlem / gözlemlerin belirlenmesi amaçlanır, daha sonra cevap değişkeni ve açıklayıcı değişkenlere uygun dönüşümler uygulanarak ya da modele yüksek dereceden terimler eklenerek sorun çözülmeye çalışılır. Bu nedenle aykırı değerlerin ya da uç gözlemlerin doğru olarak belirlenmesi regresyon analizinin önemli aşamalarından biridir.

İki değişkenli veri kümeleri ile çalışırken gözlemlerin saçılım grafiklerinden yararlanarak aykırı değerler görsel olarak belirlenebilir, ancak çok boyutlu veri kümelerine geçildiğinde benzer grafikler elde edilemediği için, aykırı değerlerin görsel olarak saptanması iki boyutlu durumdaki kadar kolay değildir. Bu nedenle araştırmacılar çok boyutlu veri kümelerinde şüpheli gözlemleri kolay ve doğru bir biçimde saptayacak yöntemler geliştirilmiştir ve bunlardan bir tanesi de 'aykırı değer haritası / outlier map'dır. Klasik tahmin ediciler yerine sağlam tahmin ediciler kullanılarak oluşturulan harita ile veriler düzenli gözlemler, dikey aykırı değerler, iyi uç gözlemler ve kötü uç gözlemler olarak tek bir grafik yardımıyla sınıflandırılmaktadır. Amaç, araştırmacıya veri kümesindeki muhtemel aykırı değerleri görsel olarak değerlendirme imkanı sunmaktır. Bu çalışmada ilk olarak, Mahalanobis uzaklık ile Sağlam uzaklık kavramlarının farklılığına değinilmiştir.

*Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: ykayhan@hacettepe.edu.tr

**Prof. Dr., Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: sgunay@hacettepe.edu.tr

Daha sonra, klasik tahmin edici - Mahalanobis uzaklık ile oluşturulacak aykırı değer haritası ile, sağlam tahmin ediciler - sağlam uzaklık kullanılarak oluşturulacak haritalar karşılaştırılmıştır. Son olarak, "hangi sağlam tahmin edici kullanılırsa elde edilecek aykırı değer haritası daha güvenilir olur" sorusunu araştırmak amacıyla üç farklı sağlam tahmin edici ile aykırı değer haritaları elde edilerek sonuçlar tartışılmıştır.

2. AYKIRI DEĞER ANALİZİ

En genel tanımı ile aykırı değerler, verinin çoğunluğu ile aynı yapıyı göstermeyen ya da çözümlemelerde kullanılan genel varsayımlardan sapmalar gösteren gözlemlerdir (Hubert vd., 2008). Bu gözlemler genel olarak iki grupta incelenir; y cevap değişkeni doğrultusunda gözlenen, veri kümesinde yer alan diğer gözlem değerlerine göre pozitif ya da negatif yönde daha büyük değerli gözlemler 'dikey aykırı değerler / vertical outliers', x açıklayıcı değişken doğrultusunda gözlenen büyük değerli gözlemler ise 'uç gözlemler / leverage points' olarak adlandırılır (Croux, 2007). Bir uç gözlem, x-uzayında verinin çoğunluğunun yer aldığı düzlemden farklı bir doğrultuda yer alıyorsa 'kötü uç gözlem / bad leverage point', verinin çoğunluğunun yer aldığı düzlem ile aynı doğrultuda yer alıyorsa 'iyi uç gözlem / good leverage point' denir (Rousseeuw ve Zomeren, 1990).

2.1. Sağlam Uzaklık

Analizlerde veri kümesindeki aykırı gözlemleri belirlemek amacıyla sıkça kullanılan bir yöntem 'Mahalanobis Uzaklık / Mahalanobis Distance / MD'dır. Bu uzaklık, veri kümesinde yer alan bir gözlemin, örneklem ortalama vektörüne olan uzaklığının örneklem kovaryans matrisi ile standartlaştırılmış ölçüsüdür. Veri kümesinde ortaya çıkabilecek bir grup aykırı değer, örneklem ortalama vektörünü kendi doğrultusunda çekebilir ve varyans kovaryans matrisini şişirerek varlıklarını gizleyebilir. Dolayısıyla örneklem ortalamasına ve kovaryansına dayalı olarak hesaplanan ve sıkça kullanılan MD yanıltıcı olabilir. Dolayısıyla MD yerine aykırı değerlerden etkilenmeyen ya da daha az etkilenen sağlam yöntemlerin kullanılmasını tercih etmek doğal bir yaklaşımdır. Bu amaçla Campell 1980 yılında, MD de yer alan konum ve ölçeğin tahmini için M tahmin edicilerini kullanmayı önermiştir. Ancak, M tahmin edicisinin kırılma noktası veri kümesindeki değişken sayısına bağlıdır ve değişken sayısı arttıkça sıfıra yakınsamaktadır. Bu da veri kümesindeki değişken sayısı arttığında tahmin edicinin aykırı değerlere karşı olan direncini azaltmaktadır. Bu soruna bir çözüm olarak 1985 yılında Rousseeuw yüksek kırılma noktasına sahip 'En Küçük Hacimli Elipsoid / Minimum Volume Elipsoid / MVE' tahmin edicisinin kullanılmasını önermiştir (Rousseeuw ve Zomeren, 1990). Böylece aykırı değerlerden MD kadar etkilenmeyen 'Sağlam Uzaklık / Robust Distance / RD' kavramı ortaya çıkmıştır. MVE tahmin edicisi veri kümesinde ortaya çıkabilecek aykırı değerlere karşı dirençlidir. Kırılma noktası %50'ye ulaşabilir ancak tahmin edici asimtotik olarak normal dağılıma yakınsamaz, etkinliği düşüktür. Bu nedenle MVE tahmin edicisine alternatif olarak daha yüksek etkiliğe sahip 'En Küçük Kovaryans Determinant / Minimum Covariance Determinant /MCD' tahmin edicisinin önerilmiştir. MCD tahmin edicisine dayalı olarak hesaplanan RD'ler aşağıdaki eşitlik ile hesaplanır,

$$RD_i = \sqrt{(\mathbf{x}_i - \mathbf{T}(\mathbf{X}))\mathbf{C}(\mathbf{X})^{-1}(\mathbf{x}_i - \mathbf{T}(\mathbf{X}))'} \quad (1)$$

MCD tahmin edicisi n gözlemli örneklemde, kovaryans matrisinin determinantı minimum olan ve h gözlemi içeren elipsoidi bulmayı amaçlar. Bu h gözlemin ortalaması veri kümesinin MCD konum tahmini, $\mathbf{T}(\mathbf{X})$, olacaktır. Ölçek tahmini, $\mathbf{C}(\mathbf{X})$, ise yine belirlenen bu h gözlemin kovaryans matrisi ile hesaplanır. Literatürde yer alan FAST-MCD algoritması (Verboven ve Hubert, 2005) ile de, MCD tahminlerinin kolaylıkla elde edilmesi mümkündür (Hubert vd., 2008). MCD konum ve ölçek tahmin edicileri aykırı değerlerden etkilenmedikleri için bu tahmin edicilere dayalı olarak hesaplanan RD de aykırı değerlerden etkilenmez ve veri kümesi ile benzer yapı göstermeyen gözlemler kolayca saptanabilir (Dallal ve Rousseeuw, 1992).

2.2. Bazı Sağlam Tahmin Ediciler

Regresyon analizi iki ya da daha çok değişken arasında bir ilişki olup olmadığının araştırılması ve ilişki varsa bunun matematiksel bir fonksiyon ile tanımlanması olarak açıklanabilir. Amaç, bilinmeyen regresyon katsayılarını tahmin ederek veri kümesine en iyi uyum gösteren regresyon modelini belirlemektir. Bu amaçla en çok kullanılan yöntem klasik 'En Küçük Kareler / Least Squares / LS' regresyondur. Bilinmeyen regresyon parametreleri, 'artık / residual' kareler toplamının minimum yapılması esasına dayalı olarak hesaplanır. LS regresyon, hatalar sıfır ortalama ve sabit varyans ile normal dağılıma sahip olduğu durumda optimal çözümü üretir. Ancak bu varsayımların geçerli olmadığı durumlarda hesaplanan parametre tahminleri yanıltıcı olabilmektedir. Bu nedenle LS regresyona seçenek olacak sağlam tahmin ediciler türetilmiştir. Geliştirilen birçok sağlam tahmin edici ilk bakışta aykırı değerlere karşı dirençli gibi görülebilir. Ancak bu tahmin ediciler y doğrultusunda ortaya çıkacak aykırı değerlere karşı sağlam iken, x doğrultusundaki uç gözlemlerden olumsuz etkilenebilir. Regresyon analizinde ise, veri kümeleri genellikle uç gözlemler içerir. Bu sebeple regresyon modelinden elde edilen artıklar ile aykırı değer analizi yapılırken kullanılan sağlam tahmin edicinin hangi tür aykırı değerlerin varlığında güvenilir sonuçlar ürettiğine dikkat edilmeli ve tahmin edicinin yüksek kırılma noktasına sahip olup olmadığı dikkate alınmalıdır.

Dikey aykırı değerlere karşı sağlamlık özelliğine sahip olmalarına rağmen kötü uç gözlemlerden olumsuz etkilenen tahmin edicilere örnek olarak Huber tarafından 1973 yılında geliştirilen M tahmin edicileri verilebilir. M tahmin edicileri uç gözlemlerden fazla etkilendikleri etkilendikleri için kırılma noktaları $1/n$'dir. Sonraki yıllarda araştırmacılar hem dikey aykırı değerlere hem de kötü uç gözlemlere karşı sağlam, yüksek kırılma noktasına sahip sağlam tahmin ediciler araştırmaya başlamışlardır. Bunlara tipik bir örnek 'En küçük Ortanca Kareler / Least Median of Squares / LMS' tahmin edicisidir. LMS tahmin edicisinin kırılma noktası %50'ye yakınsar, dolayısıyla hem dikey aykırı değerlerin hem de kötü uç gözlemlerin varlığından verinin çoğunluğuna uyan regresyon modelini oluşturur (Rousseeuw ve Leroy, 1987). Ancak asimtotik olarak normal dağılıma yakınsamaz ve etkinliği düşüktür. Bu sebeple istatistiksel çıkarsamalarda tercih edilmese de aykırı değer belirlemede kullanılmaktadır. LMS tahmin edicisinin etkiliğinin düşük olması nedeniyle ona seçenek olarak geliştirilen, kırılma noktası ve etkinliği yüksek bir başka tahmin edici de 'En küçük kırpılmış kareler / Least trimmed squares / LTS' tahmin edicisidir. Son

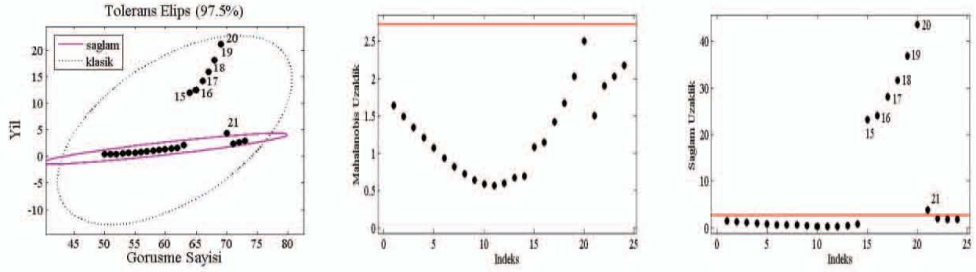
zamanlarda hem hesaplama kolaylığı hemde istatistiksel çıkarsamlarda kullanılabilmesi sebebiyle regresyon analizinde sıklıkla tercih edilmektedir.

2.3. Aykırı Değer Haritası

Veri kümesine regresyon analizi uygulandıktan sonra elde edilen modelden yararlanarak, standartlaştırılmış artık grafiği oluşturulur. Bu grafik sayesinde verinin çoğunluğu ile aynı yapıyı göstermeyen gözlemler yani aykırı değerler belirlenebilir. Ancak bu grafik yardımıyla, belirlenen noktaların dikey aykırı değer mi yoksa uç gözlem mi olduğu kararına varılamaz. Benzer bir problem de RD grafikleri için geçerlidir. RD grafiği ile veri kümesindeki uç gözlemler belirlenebilir. Ancak RD'ler hesaplanırken y_i değerleri dikkate alınmadığından, bu gözlemlerin iyi uç gözlem mi yoksa kötü uç gözlem mi olduğu anlaşılamaz. Bu nedenle hem dikey aykırı değerleri, hem iyi uç gözlemleri, hem de kötü uç gözlemleri tek bir grafik üzerinde gözlemleme olanağı sunan aykırı değer haritası geliştirilmiştir. Sağlam standartlaştırılmış model artıkları, $r_i/\hat{\sigma}$, ve RD'leri kullanarak oluşturulan bu haritada gözlemler, "düzenli (regular) gözlemler - küçük RD ve küçük $r_i/\hat{\sigma}$" , "dikey aykırı değerler - küçük RD ve büyük $r_i/\hat{\sigma}$", "iyi uç gözlemler - büyük RD ve küçük $r_i/\hat{\sigma}$" ve "kötü uç gözlemler - büyük RD ve büyük $r_i/\hat{\sigma}$" olmak üzere dört kategoriye ayrılır (Rousseeuw ve Zomeren, 1990).

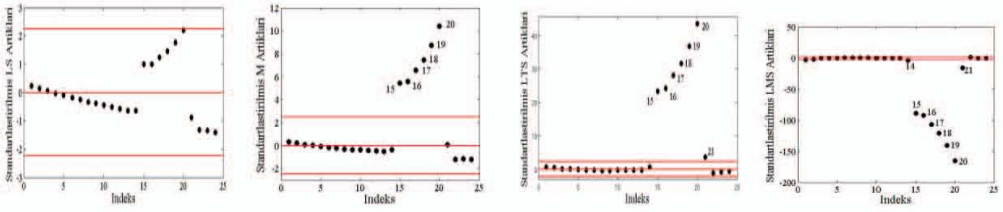
3. UYGULAMA

Bu bölümde, veri kümesinde hem x hem de y doğrultusunda ortaya çıkacak aykırı değerlerin, tahmin edicileri ve dolayısıyla bu tahmin edicilere dayalı olarak elde edilen aykırı değer haritalarını nasıl etkileyeceği ortaya koyulmuş ve yüksek kırılma noktasına sahip tahmin edicilerin kullanılmasının gerekliliği vurgulanmıştır. İlk örnek için, 1950-1973 yılları arasında Belçika'da yapılan uluslararası telefon görüşmelerinin sayısının yer aldığı veri kümesi kullanılmaktadır (Rousseeuw ve Leroy, 1987). Bu veri kümesini kullanmaktaki amaç aykırı değerlerin y cevap değişkeni doğrultusunda gözlenmesi durumunda kullanılan sağlam tahmin edicileri ve bu tahmin edicilere dayalı olarak elde edilen aykırı değer haritalarını karşılaştırmaktır. Şekil 1'de klasik örneklem ortalama ve varyansı ile elde edilen tolerans elips, MCD tahmin edicisi kullanılarak elde edilen tolerans elips, RD ve MD grafikleri verilmiştir. Açıkça görüldüğü gibi veri kümesinde yer alan bir grup aykırı değer, klasik tahminleri etkilemiş ve varlıklarını gizlemiştir. Ancak sağlam tahmin ediciler kullanılarak elde edilen grafikler, bu gözlemlerin aykırı değer olduğuna işaret etmektedir.



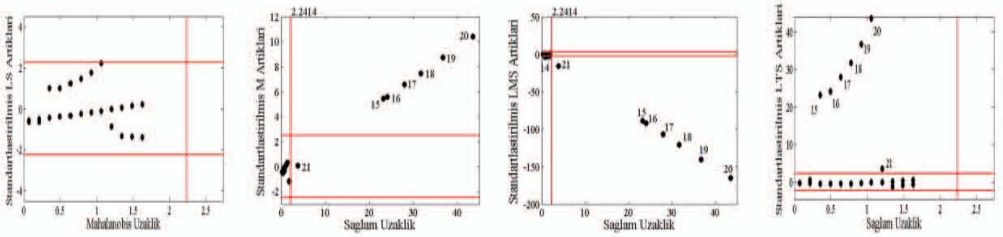
Şekil 1. Sağlam ve klasik tahmin ediciler ile hesaplanan tolerans elipsler, Mahalanobis uzaklık ve sağlam uzaklık.

Uzaklık grafiklerine bakıldığında MD ile veri kümesinde hiçbir aykırı değer tespit edilemezken, RD grafiğinden 15, 16, 17, 18, 19, 20 ve 21 no'lu gözlemlerin verinin çoğunluğu ile aynı yapıyı göstermeyen gözlemler olarak saptamıştır. Daha sonra veri kümesine sırasıyla LS, M, LMS ve LTS regresyon uygulanmıştır. Oluşturulan modellerden standartlaştırılmış artık grafikleri çizdirilmiş ve Şekil 2'de verilmiştir. LS regresyonla veri kümesinde aykırı değer saptanamamış, M tahmin edicisi kullanıldığında, 15, 16, 17, 18, 19 ve 20 no'lu gözlemler, bu gözlemlere ek olarak, LTS tahmin edicisi ile 21, LMS tahmin edicisi ile de 21 ve 14 no'lu gözlemler aykırı değer olarak belirlenmiştir. Daha öncede değinildiği gibi sadece uzaklık grafiklerine ya da sadece artık grafiklerine bakarak belirlenen bu şüpheli gözlemlerin dikey aykırı değer mi, iyi uç gözlem mi yoksa kötü uç gözlem mi olduğu anlaşılamamaktadır.



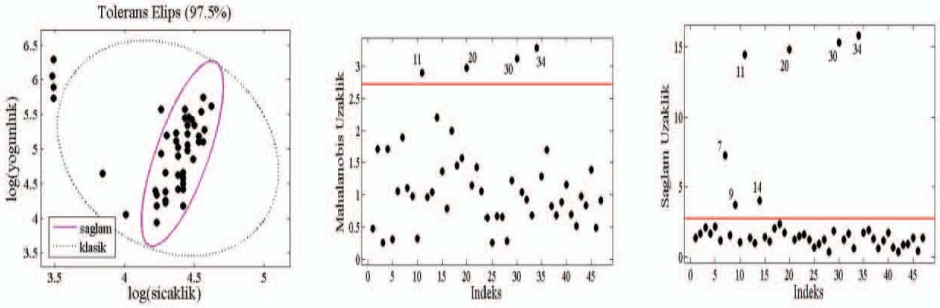
Şekil 2. LS, M, LTS ve LMS regresyon modelleri için standartlaştırılmış artık grafikleri.

Bu sebeple, her bir regresyon modeli için hem hesaplanan uzaklıkların, hem de standartlaştırılmış artıkların tek bir grafik üzerinde birleştirildiği aykırı değer haritaları oluşturulmuş ve Şekil 3'de verilmiştir. Haritalar incelendiğinde, LS regresyon, veri kümesindeki gözlemlerin tamamını düzenli gözlem, M regresyon ise 21 no'lu gözlemi iyi uç gözlem, 15, 16, 17, 18, 19 ve 20 no'lu gözlemleri kötü uç gözlem olarak saptamıştır. LMS regresyon sonucu 14 no'lu gözlem dikey aykırı değer, 15, 16, 17, 18, 19, 20 ve 21 no'lu gözlemler kötü üç gözlem, LTS regresyon ile de 15, 16, 17, 18,19,20 ve 21 no'lu gözlemlerin hepsi dikey aykırı değer olarak tespit edilmiştir.



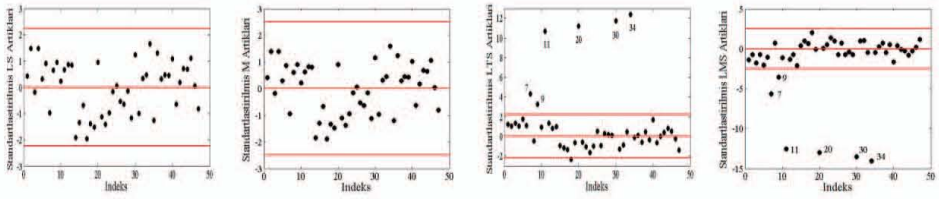
Şekil 3. Aykırı değer haritaları.

İkinci örnekte, x açıklayıcı değişken doğrultusunda gözlenen kötü uç gözlemlerin tahmin ediciler ve dolayısıyla aykırı değer haritaları üzerindeki etkisi incelenmektedir. Bu amaçla, Hertzsprung-Russell'in 47 yıldızdan oluşan CYG OB1 veri kümesi kullanılmıştır (Rousseeuw, Leroy, 1987). Klasik ve sağlam tahmin ediciler ile oluşturulan tolerans elipsler, MD ve RD grafikleri Şekil 4'de verilmektedir. Uzaklık grafikleri incelendiğinde MD ile 11, 20, 30 ve 34 no'lu gözlemler, RD ile bunlara ek olarak 7, 9 ve 14 no'lu gözlemler aykırı gözlem olarak belirlenmiştir.

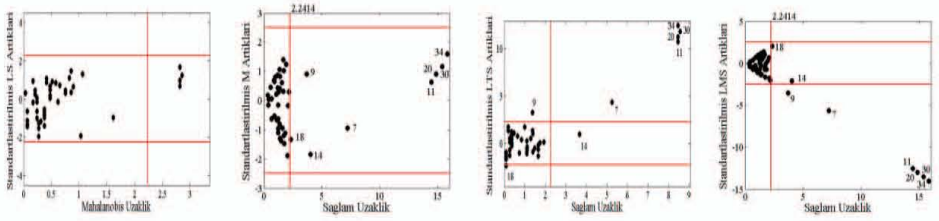


Şekil 4. Sağlam ve klasik tahmin ediciler ile hesaplanan tolerans elipsler, Mahalanobis uzaklık ve sağlam uzaklık.

Şekil 5 ile verilen standartlaştırılmış artık grafikleri değerlendirildiğinde LS ve M tahmin edicileri ile veri kümesindeki aykırı gözlemlerin belirlenemediği gözlenmektedir. LMS ve LST tahmin edicileri ise 7, 9, 11, 20, 30 ve 34 nolu gözlemlerin veri kümesinin çoğunluğu ile aynı yapıyı göstermeyen gözlemler olduğunu tespit etmiştir. Şekil 6 ile verilen aykırı değer haritaları karşılaştırıldığında ise, LS ve M tahmin edicileri veri kümesinde yer alan aykırı değerleri iyi uç gözlem olarak sınıflandırmaktadır. LMS tahmin edicisi ile 7, 9, 11, 20, 30, 34 nolu gözlemler kötü uç gözlem olarak belirlenmiştir. LTS tahmin edicisi ile de 9 ve 18 nolu gözlemler dikey aykırı değer, 7, 11, 20, 30 ve 34 nolu gözlemler de kötü uç gözlem olarak saptanmıştır.



Şekil 5. LS, M, LTS ve LMS regresyon modelleri için standartlaştırılmış artık grafikleri.

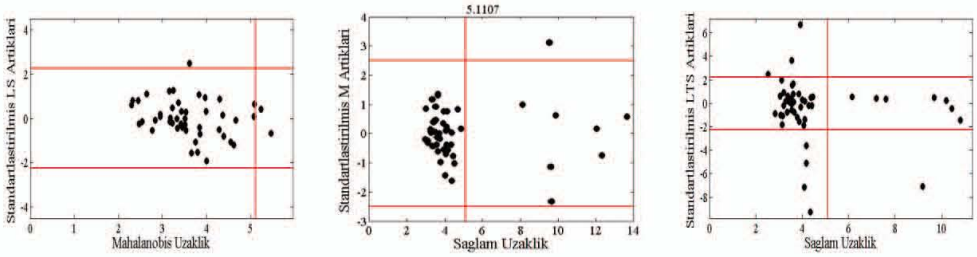


Şekil 6. Aykırı değer haritaları.

Bu iki örnek değerlendirildiğinde klasik bir tahmin edici olan LS'in veri kümesinde hem x hem de y doğrultusunda ortaya çıkabilecek aykırı değerlerden fazlasıyla etkilendiğini ve aykırı değer haritası oluşturulurken kullanılmaması gerektiği açıktır. Benzer biçimde M sağlam tahmin edicisi kullanılarak bu haritaların oluşturulmasının, kötü uç gözlemlerin varlığında LS tahmin edicisi gibi yanıltıcı olabileceği söylenebilir. LMS tahmin edicisinin, veri kümesinden ufak sapmalar gösteren gözlemleri de aykırı değer olarak sınıflandırabileceği görülmektedir. Ayrıca veri kümesinden alt örneklem seçmeye dayalı olarak hesaplanan LMS tahmin edicisi, veri kümesindeki değişken sayısı arttığında hesaplama süresi açısından etkin olmayacaktır. Sonuç olarak, hem x

hem de y doğrultusundaki aykırı değerlerden etkilenmeyen ve iteratif olarak hesaplanması daha kolay olan LTS tahmin edicisi kullanılarak aykırı değer haritası oluşturmak tercih edilebilir.

Son olarak, çok boyutlu veri kümelerine örnek olması amacıyla 1973 yılında Amerika Birleşik Devletleri'nin 47 eyaletindeki suç oranlarına ilişkin bir çalışmadan alınan 47 gözlem ve 14 değişkenli veri kümesine üzerinden LS, M ve LTS regresyona dayalı aykırı değer haritaları elde edilmiş ve Şekil 7'de verilmiştir.



Şekil 7. LS, M ve LTS ile oluşturulan aykırı değer haritaları.

LS tahmin edicisi ve MD kullanılarak elde edilen harita ile, veri kümesinde sadece 1 adet aykırı gözlem belirlenmiş, M tahmin edicisi ve RD kullanıldığında 1 adet kötü uç gözlemin varlığına işaret edilmiş, LTS tahmin edicisi ve RD kullanıldığında ise veri kümesinde 7 adet dikey aykırı değer ve 1 tane de kötü uç gözlem olabileceği tespit edilmiştir.

4. SONUÇ VE TARTIŞMA

İki değişkenli veri kümeleri ile çalışırken, gözlemlerin saçılım grafiğinin ya da standartlaştırılmış artıklar grafiğinin incelenmesi ile görsel olarak veri kümesinin çoğunluğu ile aynı yapıyı göstermeyen gözlemlerin belirlenmesi mümkün iken, çok boyutlu veri kümelerine geçildiğinde benzer grafikler elde edilememektedir. Bu gözlemleri saptamak için kullanılan klasik yöntemler yanıltıcı sonuçlar üretebilmektedir. Bu problemin çözümüne ilişkin önerilen aykırı değer haritaları ile veriyi 4 grupta kategorize etmek ve değerlendirmek mümkündür. Çoklu konum ve ölçeğin sağlam tahmin edicisine dayalı olarak hesaplanan sağlam uzaklıklar ile, sağlam regresyon sonucu elde edilen standartlaştırılmış artıkların kullanıldığı bu yöntemdeki tahmin ediciler aykırı değerlerden etkilenmediği için veri kümesinin çoğunluğu ile aynı yapıyı göstermeyen gözlemler kolaylıkla saptanabilmektedir. Ancak haritalar oluşturulurken veri kümesine uygulanacak sağlam regresyon yönteminin yüksek kırılma noktasına sahip olması, oluşturulacak haritanın ve belirlenen şüpheli gözlemlerin doğruluğunu arttıracaktır. Bu sebeple hem x hem de y doğrultusunda gözlenecek aykırı değerlerden etkilenmeyen LTS gibi yüksek kırılma noktasına sahip tahmin edicilerin kullanılması tercih edilmelidir. Elbette bir çalışmada, gözlemlerin tek bir tanı aracı ile kesin olarak aykırı değer kabul edilmesi yanıltıcı olabilir. Bu çalışmadaki amaç, araştırmacıya veri kümesinde incelenmesi gereken şüpheli gözlemleri gösterecek, kullanımı kolay bir yöntemi tanıtmak ve hangi durumlarda hangi sağlam tahmin edicinin kullanılmasının daha güvenilir sonuçlar vereceği konusunda karşılaştırmalı bir uygulama sunmaktır.

5. KAYNAKLAR

Croux, C., 2007. An Introduction to Robust Statistics: Mathematics and Practice. Lecture Notes, Faculty of Economics and Management, University Center of Statistics.

Dallal, G.E., Rousseeuw, P. J., 1992. LMSMVE A Program for Least Median of Squares Regression and Robust Distances, Computers and Biomedical Researches, Vol 25, 384-391.

Hubert, M., Rousseeuw, P. J., Aelst, S. V., 2008. High-Breakdown Robust Multivariate Methods, Statistical Science, Vol.23, No.1, 92-119.

Rousseeuw, P. J., Leroy, A. M., 1987. Robust Regression and Outlier Detection, John Wiley & Sons, New York.

Rousseeuw, P. J., Van Zomeren, B. C., 1990. Unmasking Multivariate Outliers and Leverage Points. Journal of the American Statistical Association, Theory and Methods, Vol.85, No.411.

Verboven, S., Hubert, M., 2005. LIBRA: A MATLAB Library for Robust Analysis. http://wis.kuleuven.be/stat/robust/LIBRA.html.

CLASSIFICATION OF THE OBSERVATIONS IN REGRESSION ANALYSIS BY OUTLIER MAP

ABSTRACT

In practice, multidimensional data sets generally contain observations that deviate from the majority of data. One of the important stages of regression analysis is to correctly determine these observations by using residual analysis. However, conventional statistical methods used for this purpose are too much influenced by outliers. Therefore, the outlier analysis techniques based on classical estimators may mislead the investigator. In this study, outlier map which is used to examine observations in multidimensional data sets and generated by robust estimators instead of the classical estimators is briefly explained. The aim of this study is to compare outlier maps of different regression models generated by using different robust estimators and to discuss which robust estimator will create more reliable map.

Keywords: Outliers, Robust regression, Robust estimators, Extreme observation.