

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

İKİLİ LOJİSTİK REGRESYONDA ROBUST YÖNTEMLER

DOKTORA TEZİ

Ebru GÜNDOĞAN AŞIK

**OCAK 2023
TRABZON**



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

İKİLİ LOJİSTİK REGRESYONDA ROBUST YÖNTEMLER

Ebru GÜNDOĞAN AŞIK

ORCID : - - -

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünce
"DOKTOR (İSTATİSTİK)"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.

Tezin Enstitüye Verildiği Tarih : 14 / 12 / 2022

Tezin Savunma Tarihi : 11 / 01 / 2023

Tez Danışmanı : Prof. Dr. Zafer KÜÇÜK

ORCID : - - -

İkinci Danışman : Prof. Dr. Arzu ALTIN YAVUZ

Trabzon 2023

ÖNSÖZ

Doktora tezi olarak sunulan bu çalışmada, veri setinde çoklu iç ilişki ve aykırı değer problemi olduğu durumda bu iki problemi çözmeye yardımcı olacak robust yanlı parametre tahmin edicisi önerilmesi amaçlanmıştır.

Doktora tez çalışmamda danışmanlığımı üstlenerek, konu seçimi ve çalışmaların yürütülmesi sırasında bilgisini, tecrübesini benimle paylaşan deneyimleri ile desteklerini esirgemeyen, her konuda bana yol gösteren, kıymetli danışmanlarım Sayın Prof. Dr. Arzu ALTIN YAVUZ'a ve Prof. Dr. Zafer KÜÇÜK'e teşekkür ederim. Tez çalışması boyunca değerli fikir ve görüşleri ile beni yönlendirerek katkı sağlayan tez izleme komitesi üyelerim Sayın Prof. Dr. Türkan ERBAY DALKILIÇ ve Prof. Dr. Yüksel TERZİ hocalarıma, lisans ve lisansüstü eğitim hayatım boyunca yardımlarını esirgemeyen Sayın Doç. Dr. Yakup Murat BULUT hocama teşekkür ederim.

Tezin her aşamasında bana desteklerini esirgemeyen her zaman yanımda olduklarını bildiğim değerli arkadaşlarım Arş. Gör. Dr. Yeşim AKBAŞ'a, Arş. Gör. Dr. Serkan AKBAŞ'a, Dr. Öğr. Üyesi Erdiç KARAKULLUKÇU'ya ve Melike IŞILAR'a teşekkür ederim. Desteklerini esirgemeyen İstatistik ve Bilgisayar Bilimleri Bölümü'ndeki tüm değerli hocalarıma ve asistan arkadaşlarıma şükranlarımı sunarım.

Ayrıca bu tez aşamasına gelene kadar eğitim hayatım boyunca bana maddi manevi destek olan kuzenlerim, arkadaşlarım, öğretmenlerim ve isimlerini yazmakla bitiremeyeceğim manevi ailelerime teşekkürü bir borç bilirim.

Son olarak her zaman yanımda olan, sevgisini her zaman hissettiğim babama, hayattaki diğer yarım olan ikizim Emre GÜNDOĞAN'a, bize her şey olan canım abim Hüseyin GÜNDOĞAN'a, bana her zaman inanan, lisans eğitimime adım atmamı sağlayan abim Hakan GÜNDOĞAN'a, bu süreçte destek olan babaannemiz Nezaat AŞIK'a, kocaman güzel aileme ve beni bir yerlerden izleyip gurur duyduğunu düşündüğüm anneme teşekkür ederim. Hayatımda oldukları için şükrettiğim, bu stresli dönemde bana güç veren, sonsuz destek olan sevgili eşim Zeki AŞIK'a ve bizimle oldukları andan itibaren hayatımızı daha da güzelleştiren ikizlerim Öykü ve Mete'ne sonsuz sevgi ve teşekkürlerimi sunarım.

Ebru GÜNDOĞAN AŞIK

Trabzon 2023

TEZ ETİK BEYANNAMESİ

Doktora Tezi olarak sunduğum “İkili Lojistik Regresyonda Robust Yöntemler” başlıklı bu çalışmayı baştan sona kadar danışmanlarım Prof. Dr. Arzu ALTIN YAVUZ ve Prof. Dr. Zafer KÜÇÜK’ün sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 11/01/2023

Ebru GÜNDOĞAN AŞIK

İÇİNDEKİLER

Sayfa No

ÖNSÖZ.....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ	IX
TABLolar DİZİNİ.....	X
SEMBOLLER DİZİNİ	XI
1. GENEL BİLGİLER.....	1
1.1. Giriş	1
1.2. Literatür Çalışması.....	3
1.3. Lojistik Regresyon Modeli	8
1.3.1. En Çok Olabilirlik Yöntemi	11
1.4. Lojistik Regresyon Modelinde Yanlı Tahmin Ediciler	13
1.4.1. Çoklu İç İlişki Problemi.....	13
1.4.2. Çoklu İç İlişki Problemi İçin Belirleme Yöntemleri	13
1.4.3. Çoklu İç İlişki Problemi İçin Çözüm Yöntemleri.....	14
1.4.4. Yanlı Tahmin Ediciler	14
1.4.4.1. Lojistik Ridge Tahmin Edicisi.....	15
1.4.4.2. İki Parametrelili Ridge Tahmin Edicisi	17
1.4.4.3. İki Parametrelili Lojistik Ridge Tahmin Edicisi	19
1.5. Lojistik Regresyon Modelinde Olağan Dışı Gözlemler	21
1.5.1. Aykırı Değer Belirleme Yöntemleri	25
1.5.1.1. Pearson Artıklar	25
1.5.1.2. Standart Pearson Artıklar.....	26
1.5.1.3. Sapma Artıklar.....	26
1.5.2. Yüksek Kaldıraç Noktası Belirleme Yöntemleri.....	27
1.5.2.1. DEVC ve RobDEVC Teşhis Yöntemleri	27
1.6. Lojistik Regresyon Modelinde Robust Tahmin Ediciler	29

1.6.1.	Genelleştirilmiş M Tahmin Edicisi.....	30
1.6.1.1.	Mallows Tahmin Edicisi.....	31
1.6.1.2.	CUBIF Tahmin Edicisi	31
1.6.2.	Bianco ve Yohai Tahmin Edicisi.....	32
1.6.3.	Ağırlıklandırılmış Bianco ve Yohai Tahmin Edicisi	32
1.6.4.	Robust Lojistik Tahmin Edicisi	33
2.	YAPILAN ÇALIŞMALAR.....	35
2.1.	Lojistik Regresyon Modelinde Robust Teşhis Yöntemleri ve Robust Eşik Değerler	35
2.1.1.	RobDEVC Teşhis Yöntemleri	35
2.1.2.	Robust Eşik Değerler.....	36
2.1.2.1.	Konumun Robust Tahmin Edicileri.....	36
2.1.2.2.	Ölçek Parametresinin Robust Tahmin Edicileri	41
2.2.	Robust Yanlı Tahmin Ediciler.....	44
2.2.1.	Robust Lojistik Ridge Tahmin Edicisi	44
2.2.2.	İki Parametrelili Robust Lojistik Ridge Tahmin Edicisi	45
2.2.3.	Tahmin Edicilerin Karşılaştırılması.....	48
2.2.3.1.	RLE ve RLRE Tahmin Edicilerinin Karşılaştırılması.....	48
2.2.3.2.	RLRE ve TPRLRE Tahmin Edicilerinin Karşılaştırılması.....	49
3.	BULGULAR VE İRDELEME	51
3.1.	Simülasyon Çalışması I	51
3.2.	Sayısal Örnek I	66
3.3.	Simülasyon Çalışması II.....	70
3.4.	Sayısal Örnek II	78
4.	SONUÇ VE ÖNERİLER.....	82
5.	KAYNAKÇA	84
ÖZGEÇMİŞ		

Doktora Tezi

ÖZET

İKİLİ LOJİSTİK REGRESYONDA ROBUST YÖNTEMLER

Ebru GÜNDOĞAN AŞIK

Karadeniz Teknik Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik ve Bilgisayar Bilimleri Anabilim Dalı

Danışman: Prof. Dr. Zafer KÜÇÜK - Prof. Dr. Arzu ALTIN YAVUZ

2023, 89 Sayfa

Bağımlı değişkenin kesikli olması durumunda uygulanan regresyon analizi yöntemine lojistik regresyon analizi denir. Lojistik regresyon analizinin amaçlarından biri sınıflandırma, diğeri ise en az değişkeni kullanarak en iyi uyuma sahip olacak şekilde bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayabilen ve genel olarak kabul edilebilen bir model kurmaktır.

Doğrusal regresyon analizinde olduğu gibi lojistik regresyon analizinde de çoklu iç ilişki ve aykırı değer problemleriyle karşılaşılabilir. Bağımsız değişkenler arasında çoklu iç ilişkinin var olması durumunda elde edilen regresyon katsayıları gerçek değerlerini yansıtmayacaktır. Veri kümesinde aykırı değerler olması durumunda ise bu değerlerin belirlenmesi ve parametre tahminleri üzerinde etkilerinin incelenmesi gerekmektedir. Bu problemlerin çözümleri için literatürde ayrı ayrı tahmin ediciler önerilmiştir.

Bu tez çalışmasında veri kümesinde hem çoklu iç ilişki hem de aykırı değer problemlerinin olması durumunda bu iki problemin birlikte çözümü için robust yanlı bir tahmin edici önerilmiştir. Önerilen bu tahmin edici literatürde mevcut olan diğer tahmin edicilerle hata kareler ortalaması (MSE) kriterine göre karşılaştırılmıştır. Gerçek veri uygulaması ile simülasyon çalışması desteklenmiştir. Ayrıca veri setinde bulunan aykırı değerler için literatürde mevcut olan yöntemler yardımıyla robust teşhis yöntemleri ve bu yöntemler için robust eşik değerler önerilmiştir.

Anahtar Kelimeler: Lojistik regresyon, Robust, Çoklu iç ilişki, Aykırı değer

PhD. Thesis

SUMMARY

ROBUST METHODS IN BINARY LOGISTIC REGRESSION

Ebru GÜNDOĞAN AŞIK

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Statistics and Computer Sciences Graduate Program
Supervisor: Prof. Dr. Zafer KÜÇÜK - Prof. Dr. Arzu ALTIN YAVUZ
2023, 89 Pages

The regression analysis method applied when the dependent variable is discrete is called logistic regression analysis. One of the purposes of logistic regression analysis is classification, and the other is to establish a generally acceptable model that can describe the relationship between dependent and independent variables in a way that has the best fit by using the fewest variables.

As in linear regression analysis, multicollinearity and outlier problems can be encountered in logistic regression analysis. In case of multicollinearity between independent variables, the obtained regression coefficients will not reflect their true values. In case of outliers in the data set, these values should be determined and their effects on parameter estimations should be examined. For the solutions of these problems, different estimators have been proposed in the literature.

In this thesis study, a robust biased estimator is proposed to solve these two problems together in case of both multicollinearity and outlier problems in the data set. This proposed estimator was compared with other estimators available in the literature according to the mean square error (MSE) criterion. Simulation study was supported with real data application. In addition, robust diagnostic methods and robust cut-off values for these methods were proposed with the help of the methods available in the literature for the outliers in the data set.

Key Words: Logistic regression, Robust, Multicollinearity, Outlier

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.	Düzenlenmiş Brown verisi için saçılım grafiği	23
Şekil 2.	DEVC (a), RobDEVC (b), RobDEVC ₁ (c), RobDEVC ₂ (d) ve RobDEVC ₃ (e) sapma artıklarının gözlem numaralarına göre grafikleri	69
Şekil 3.	Pena verisi korelasyon grafiği	79
Şekil 4.	DEVC (a) ve RobDEVC ₃ (b) sapma artıklarının gözlem numaralarına göre grafiği	80



TABLolar DİZİNİ

Sayfa No

Tablo 1. M tahmin ediciler için ψ fonksiyonları.....	39
Tablo 2. $n \leq 9$ için c_n değerleri	42
Tablo 3. $n \leq 9$ için d_n değerleri	43
Tablo 4. Oluşturulan yeni eşik değerler.....	43
Tablo 5. $p = 2, n = 80$ ve kötü kaldırıcı noktası 2 için simülasyon sonuçları	52
Tablo 6. $p = 2, n = 120$ ve kötü kaldırıcı noktası 2 için simülasyon sonuçları	53
Tablo 7. $p = 2, n = 80$ ve kötü kaldırıcı noktası 3 için simülasyon sonuçları	54
Tablo 8. $p = 2, n = 120$ ve kötü kaldırıcı noktası 3 için simülasyon sonuçları	55
Tablo 9. $p = 2, n = 80$ ve kötü kaldırıcı noktası 5 için simülasyon sonuçları	56
Tablo 10. $p = 2, n = 120$ ve kötü kaldırıcı noktası 5 için simülasyon sonuçları	57
Tablo 11. $p = 3, n = 80$ ve kötü kaldırıcı noktası 2 için simülasyon sonuçları	58
Tablo 12. $p = 3, n = 120$ ve kötü kaldırıcı noktası 2 için simülasyon sonuçları	59
Tablo 13. $p = 3, n = 80$ ve kötü kaldırıcı noktası 3 için simülasyon sonuçları	60
Tablo 14. $p = 3, n = 120$ ve kötü kaldırıcı noktası 3 için simülasyon sonuçları	61
Tablo 15. $p = 3, n = 80$ ve kötü kaldırıcı noktası 5 için simülasyon sonuçları	62
Tablo 16. $p = 3, n = 120$ ve kötü kaldırıcı noktası 5 için simülasyon sonuçları	63
Tablo 17. Teşhis edici yöntemlerin sapma artık değerleri.....	67
Tablo 18. Hesaplanan eşik değerler	68
Tablo 19. Tahmin ediciler.....	71
Tablo 20. $p = 3$ ve %5 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri ...	72
Tablo 21. $p = 3$ ve %10 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri .	73
Tablo 22. $p = 4$ ve %5 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri ...	74
Tablo 23. $p = 4$ ve %10 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri .	75
Tablo 24. $p = 6$ ve %5 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri ...	76
Tablo 25. $p = 6$ ve %10 aykırı değer olduğu durumda tahmin edicilerin, MSE değerleri	77
Tablo 26. Pena veri setinin parametre tahminleri, tahminlerin standart hataları ve MSE değerleri	81

SEMBOLLER DİZİNİ

OLS	: Sıradan En Küçük Kareler (Ordinary Least Squares)
MLE	: En Çok Olabilirlik Tahmin Edicisi (Maximum Likelihood Estimator)
MSE	: Hata Kareler Ortalaması (Mean Square Error)
TPRE	: İki Parametrelili Ridge Tahmin Edicisi (Two Parameter Ridge Estimator)
RMSE	: Artık Kök Hata Kareler Ortalaması Matrisi (Residual Root Mean Square Error)
MMSE	: Matris Hata Kareler Ortalaması (Matrix Mean Square Error)
LRE	: Lojistik Ridge Tahmin Edicisi (Logistic Ridge Estimator)
BY	: Bianco Yohai Tahmin Edicisi (Bianco and Yohai Estimator)
WBY	: Ağırlıklandırılmış Bianco Yohai Tahmin Edicisi (Weighted Bianco and Yohai Estimator)
CUBIF	: Koşullu Yansız Sınırlı Etki Fonksiyon Tahmin Edicisi (Conditionally Unbiased Bounded Influence Function Estimator)
Mallows	: Mallows Tip Kaldıraç Bağımlı Ağırlık Tahmin Edicisi (Mallows Type Leverage Dependent Weights Estimator)
RLRE	: Robust Lojistik Ridge Tahmin Edicisi (Robust Logistic Ridge Estimator)
IRLS	: İteratif Yeniden Ağırlıklı En Küçük Kareler Yöntemi (Iterative Re-weighted Least Squares Method)
LE	: Lojistik Tahmin Edicisi (Logistic Estimator)
DEVC	: Sapma Bileşenler (Deviance Components)
RobDEVC	: Robust Sapma Bileşenler (Robust Deviance Components)
GM	: Genelleştirilmiş M (Generalized M)
RMD	: Robust Mahalanobis Uzaklığı (Robust Mahalanobis Distance)
MCD	: Minimum Kovaryans Belirleyicisi (Minimum Covariance Determinant)
RLE	: Robust Lojistik Tahmin Edicisi (Robust Logistic Estimator)
MAD	: Medyan Mutlak Sapma (Median Absolute Deviation)
CIR	: Doğru Belirleme Oranı (Correct Identification Rate)
SR	: Swamping Oranı (Swamping Rate)
TPLRE	: İki Parametrelili Lojistik Ridge Tahmin Edicisi (Two Parameter Logistic Ridge Estimator)
TPRLRE	: İki Parametrelili Robust Lojistik Ridge Tahmin Edicisi (Two Parameter Robust Logistic Ridge Estimator)

1. GENEL BİLGİLER

1.1. Giriş

Günümüzde teknolojinin gelişmesi ile birlikte veri boyutları artmaktadır. Artan bu verilerin çok boyutlu olması çok değişkenli istatistiksel yöntemlerin kullanılmasını gerekli kılmaktadır. Ayrıca çok değişkenli istatistiksel yöntemler verilerin sınıflandırılması için de kullanılmaktadır. Sınıflandırma için uygulamada çeşitli yöntemler bulunmaktadır ve bunlar gözlemleri yapılarına göre gruplara atamak için kullanılmaktadır. Bu yöntemlerden bazıları kümeleme, diskriminant (ayırıcı fonksiyon) ve lojistik regresyon yöntemidir [1].

Gözlemlerin gruplara atanmasında, grup sayısı bilinmiyor ise kümeleme analizi yapılarak gözlemler uzaklık ya da benzerlik ölçütlerine göre kümelenmektedir. Burada amaç yalnızca gözlemlerin oluşturduğu küme yapısını belirlemektir ve kovaryans matrisi için herhangi bir varsayım yoktur [2]. Gözlemlerin gruplara atanmasında grup sayısı biliniyor ise diskriminant ve lojistik regresyon analizi ile bir model elde edilmektedir ve veri kümesine yeni alınan gözlemlerin gruplara atamaları yapılmaktadır. Diskriminant analizinde verilerin çok değişkenli normal dağılıma sahip olmaları koşulu bulunmaktadır. Ancak uygulamada normallik varsayımının sağlanmadığı durumlar ile karşılaşilmektedir. Bu durumda diskriminant analizi yerine alternatif olarak görülen lojistik regresyon analizinin kullanılması önerilir [3]. Bağımlı değişkenin iki ya da ikiden fazla düzey içeren değişken olduğu durumda bazı varsayımsal bozulmalardan dolayı lojistik regresyon analizi, doğrusal regresyon analizi yerine alternatif olarak uygulanmaktadır [4].

Lojistik regresyon analizi, bağımlı ve bağımsız değişkenler arasında en iyi uyuma sahip bir ilişkiyi tanımlayan model kurmayı amaçlamaktadır. Kurulan modeldeki parametreler Sıradan En Küçük Kareler (OLS), En Çok Olabilirlik (MLE), momentler vb. tahmin yöntemleri ile tahmin edilmektedir. OLS yöntemi kullanılarak oluşturulan modelin etkin ve çıkarsama amaçlı kullanılması için gerekli varsayımların sağlanması gerekmektedir. Bu varsayımlar hataların sıfır ortalamalı, eşit varyanslı (σ^2) normal dağılıma sahip olması ve bağımsız değişkenler arasında çoklu iç ilişkinin olmaması olarak sıralanabilir.

Lojistik regresyon analizinde de bağımsız değişkenler arasında ilişkinin olmaması gerekmektedir. Bağımsız değişkenler arasında çoklu iç ilişkinin var olması durumunda elde edilen tahmin edicilerin minimum varyanslılık özelliğini sağlayamadıkları bilinmektedir.

Çoklu iç ilişkinin var olduğu durumda ele alınan değişkenlerin indirgenmesi yapılmakta veya yanlı regresyon yöntemlerine başvurulmaktadır. Yanlı regresyon yöntemleri kullanılarak yanlı fakat daha küçük varyansa sahip tahminler elde edilir. Yanlı regresyon yöntemlerini kullanmak daha fazla tercih edilen bir durumdur. Yanlı regresyon yöntemleri arasında yaygın kullanıma sahip yöntemlerden biri de ridge regresyon yöntemidir. Lojistik regresyon modelindeki parametreler ridge regresyon yöntemi kullanılarak da tahmin edilmektedir [5].

Lojistik regresyon analizinde ele alınması gereken bir diğer konu ise olağan dışı gözlemler olarak bilinen aykırı değer, etkili gözlem ve yüksek kaldıraç noktalarının varlığının kontrol edilmesidir. Lojistik regresyon analizinde MLE yöntemi kullanılarak elde edilen parametre tahminleri, verinin büyük çoğunluğundan önemli derecede farklılık gösteren olağan dışı gözlemlerin varlığından oldukça etkilenmektedir. Doğrusal regresyon analizinde bağımlı değişkenin alacağı değerler tahmin edilirken, lojistik regresyon analizinde bağımlı değişkenin alabileceği değerlerin ortaya çıkabilme olasılıkları tahmin edilmektedir. Bu sebepten dolayı lojistik regresyon analizinde olağan dışı gözlem yani aykırı değer var ise, bu gözlemler; modelden tahmin edilen başarı olasılığı küçük olduğu halde gözlemlenen durumu başarılı olan ya da modelden elde edilen başarı olasılığı yüksek olduğu halde gözlemlenen durumu başarısız olan gözlemler olarak tanımlanabilir [3]. Veri kümesinde aykırı değerler var ise modelden geçerli ve güvenilir sonuçlar elde edilemeyeceğinden aykırı değerlerin belirlenmesi ve parametre tahminleri üzerinde etkilerinin incelenmesi gerekmektedir [6]. Bu durumda bu değerlerin etkilerini en aza indirmeye çalışan robust yöntemler tercih edilmelidir.

Bu tez çalışmasında, bağımlı değişkenin kesikli değişken olması durumunda uygulamada sıklıkla karşılaşılabilecek iki önemli problem ele alınmıştır. Bu problemlerden ilki bağımsız değişkenler arasında çoklu iç ilişki olmasıdır. Diğer problem ise veri setinde olağan dışı gözlem olan aykırı değer bulunması durumudur. Bu problemler için öncelikle literatürde sıkça kullanılan yöntemler tanıtılmıştır. Veri setinde bulunan olağan dışı gözlemlerler için robust teşhis yöntemleri ve bu yöntemler için robust eşik değerler önerilmiştir. Ayrıca ele alınan bu problemlerin çözümü için yeni bir tahmin edici önerilerek alternatif tahmin edicileri ile karşılaştırılmıştır. Yöntemlerin karşılaştırılması MSE kriterine göre yapılmıştır.

1.2. Literatür Çalışması

Değişkenler arasında ilişkinin varlığını ortaya koyan ve çıkarsama yapmak için model sunan regresyon analizi, istatistiksel analiz yöntemleri arasında en geniş kullanım alanına sahip yöntemlerden biridir. Bağımlı değişkenin kategorik olması durumunda kullanılan yöntemlerden biri olan lojistik regresyon analizi, bağımlı değişkenin kategori sayısına göre adlandırılmaktadır. Bu tez çalışmasında kullanılan ikili lojistik regresyon analizi literatürde oldukça yaygın bir kullanım alanına sahiptir.

Lojistik regresyon analizi tıpta, mühendislikte, biyomedikal araştırmalarda, iş ve finans alanlarında ve sınıflandırmanın olduğu her alanda son yıllarda kullanılmaktadır. İlk olarak Berkson [6] tarafından biyolojik deneylerin analizinde lojistik regresyon modeli kullanılmıştır. Cox [7], Berkson tarafından önerilen bu modeli geliştirmiştir. Pregibon [8] iki kategorili lojistik modelde etkinliği, aykırı gözlemleri ve belirleme ölçülerini incelemiştir. Lee [9] basit dönüşümlü deneme planları için lojistik modeller üzerinde çalışmıştır. Bonney [10] lojistik regresyon modelinin geliştirilmesi için çalışmalar yapmıştır. Duffy [11] lojistik regresyonda hata terimlerinin dağılışı ve parametre değerlerinin gerçek değerlere yaklaşımını incelemiştir. Tester vd. [12] yollardaki hız kesici kasislerin yaralanan çocuk sayısını azaltma üzerindeki etkisini incelerken lojistik regresyon modelinden yararlanmışlardır. Arı ve Önder [13] yaptıkları çalışmalarında lojistik regresyon ve diğer farklı regresyon modellerini karşılaştırarak bu regresyon modellerinin hangi durumlarda kullanılabileceğini incelemişlerdir.

Regresyon analizinde, regresyon katsayılarının tahmini için OLS en klasik yöntemlerden biridir. OLS, yansız bir tahmin edicidir ve Gauss-Markov teoremine göre yansız ve en küçük varyansa sahip olduğu bilinmektedir. Bağımsız değişkenler arasında yüksek korelasyon olması çoklu iç ilişki problemine neden olmaktadır. Bu durumda OLS tahmin edicisi hala yansızdır fakat varyansı çok büyüktür. Bu problemi ortadan kaldırmak için yanlı regresyon yöntemleri önerilmiştir. Yanlı regresyon yöntemlerinden biri olan ridge regresyon yöntemi ilk olarak Hoerl ve Kennard [14] tarafından doğrusal modeller için önerilmiştir. Schaefer vd. [15] ridge regresyon modelini ikili lojistik regresyon modeli için düzenleyip uygulamışlardır. Önerilen ridge lojistik tahmin edicisinin MSE kriterine göre MLE tahmin edicisinden daha iyi performansa sahip olduğunu göstermişlerdir. Lipovetsky ve Conklin [16] tarafından önerilen iki parametrelili ridge tahmin edicisi (TPRE), çoklu iç ilişki problemi ile karşılaşıldığı durumda kullanılmıştır. Bu çalışmada ayrıca ridge tahmin

edicisinin kullanım amacına ek olarak modelin uyumu için belirlilik katsayısını en büyük yapacak şekilde ikinci bir parametre kullanmışlardır. Böylece eş zamanlı olarak hataların en küçük yapılmasını ve istenilen özelliklere sahip ridge tahmin edicisinden daha iyi (istenilen özelliklere sahip) performans gösteren bir tahmin edici elde edilmesini sağlamışlardır. Çalışmada TPRE ve ridge tahmin edicisi artık kök hata kareler ortalama (RMSE) kriterine göre karşılaştırılmıştır. Çalışma sonucunda TPRE tahmin edicisinin diğer tahmin edicilerden RMSE kriterine göre daha iyi performans gösterdiğini ortaya koymuşlardır. Lipovetsky [17] iki parametrelili ridge tahmin edicisinin yorumlanabilir katsayılarını, yan karakteristiğini, etkinliğini ve modelin önemli diğer özelliklerini incelemiştir. OLS, ridge ve TPRE'nin performans karşılaştırmalarını belirlilik katsayısı ve RMSE kriterlerine göre gerçek veri yardımıyla yapmıştır. Özkale ve Kaçıranlar [18] tarafından önerilen iki parametrelili ridge tahmin edicisi OLS, ridge, Liu ve özel bir durum olan yeni tahmin edicilerini içeren genel bir tahmin edicidir. Önerilen tahmin edici matris hata kareler ortalaması (MMSE) kriterine göre karşılaştırılmıştır. Bu çalışmada yanlılık parametrelerinin tahminleri ve karşılaştırmaları simülasyon yardımıyla yapılmıştır.

Mansson ve Shukur [19] çoklu iç ilişki problemi durumunda farklı koşullarda lojistik ridge parametrelerinin MLE yöntemi kullanarak elde edilen tahminlerini incelemişlerdir. MSE kriterine göre yapılan karşılaştırmada lojistik ridge tahmin edicilerinin MLE tahmin edicilerden daha iyi performansa sahip olduklarını ortaya koymuşlardır. Kibria vd. [20] lojistik ridge yöntemi için Muniz vd. [21] tarafından önerilen ridge parametrelerinin tahmin yöntemlerinin geliştirilmiş farklı tiplerini ele almışlardır. Monte Carlo simülasyonu ile elde edilen tahmin edicilerin performansları incelendiğinde lojistik ridge regresyon modeli ile elde edilen tahmin edicilerin, MLE yöntemi ile elde edilen tahmin edicilerden MSE kriterine göre daha iyi olduğunu söylemişlerdir. Mansson vd. [22], Liu [23] tarafından önerilen tahmin edicinin geliştirilmiş bir tahmin edicisi olan yeni bir büzülme (shrinkage) tahmin edicisi önermişlerdir. Önerdikleri lojistik Liu tahmin edicisinin MSE ve ortalama mutlak hata kriterlerine göre MLE tahmin yönteminden daha iyi sonuçlar verdiğini belirtmişlerdir. Huang [24] lojistik ridge ve lojistik Liu tahmin edicilerinin bir kombinasyonu olan yanlı bir tahmin edici önermiştir. Önerilen bu tahmin edicinin MLE ve ridge tahmin edicileri elde edilerek bazı özellikleri için MSE kriterine göre karşılaştırılması yapılmıştır.

İnan ve Erdoğan [25] çoklu iç ilişki probleminin çözümü için lojistik Liu-tipi tahmin edicisi önermişlerdir. Önerilen bu tahmin edici ridge ve Liu tahmin edicilerinin

kombinasyonudur ve iki parametresi bulunmaktadır. Asar vd. [26] lojistik ridge regresyon modeli için bir kısıtlı ridge regresyon tahmin edicisi geliştirmişlerdir. Çalışmalarında bu yanlı tahmin edicinin özelliklerini de ortaya koymuşlardır. Simülasyon çalışması ve gerçek veri uygulamasıyla ridge parametresinin farklı seçimleri için önerilen bu tahmin edicinin etkin sonuçlar verdiğini göstermişlerdir. Ertaş vd. [27] doğrusal regresyon modelinde çoklu iç ilişki ve y yönünde aykırı değer olduğu durumda M tahmin edicisine dayalı iki parametrelili ridge- M tahmin edicisini önermişlerdir. Regresyon modelinde önemli iki problemi birlikte ele alan bu çalışmada önerilen tahmin ediciyle birlikte TPPE ve ridge- M tahmin edicisi MSE kriterine göre karşılaştırılmıştır. Diğer tahmin edicilere kıyasla etkin performans gösterdiği ortaya konmuştur. Simülasyon ve gerçek veri uygulaması ele alınan çalışmada veri seti aykırı değersiz, bir, iki ve dört aykırı değerli olarak belirlenmiştir. Karşılaştırmalarda ridge parametreleri literatürde mevcut olan yöntemlerden seçilmiştir. Asar ve Erişoğlu [28] doğrusal regresyon modelinde veri setinde etkili gözlem ve çoklu iç ilişki olduğu durumda klasik regresyon ve ridge regresyon modellerine alternatif olarak iki parametrelili ridge regresyon modeline dayalı teşhis yöntemlerinden olan DFFITS ve Cook uzaklıklarını tanımlamışlardır. Etkili gözlemleri belirlemede yeni ölçümlerin kullanışlı olduklarını gerçek veri uygulamasıyla ortaya koymuşlardır. Özkale vd. [29] lojistik ridge regresyon yöntemi ile veri setinde olağan dışı gözlemleri belirlemek için istatistikler önermişlerdir. Hem Monte Carlo simülasyonu hem de sayısal bir örnek ile etkilerini azaltma yoluyla, lojistik ridge tahmin edicisinin (LRE) bu olağan dışı gözlemlerle başa çıktığını göstermişlerdir. Asar [30] çoklu iç ilişki problemini gidermek için yeni kısıtlı bir tahmin edici önermiştir. Kısıtlı Liu tahmin edicisinin, kısıtlı en çok olabilirlik tahmin edicisinin, MLE ve Liu tipi tahmin edicilerinin performanslarını karşılaştırmıştır. Bu yöntemler MSE kriterine göre karşılaştırıldığında önerilen kısıtlı Liu tahmin edicisinin daha iyi performans gösterdiğini ortaya koymuştur. Huang vd. [31] doğrusal regresyon modelinde çoklu iç ilişki problemini ele almak için Özkale ve Kaçıranlar'ın [18] geliştirdiği kısıtlı iki parametrelili tahmin ediciye dayalı kısıtlı hemen hemen yansız iki parametrelili tahmin ediciyi önermişlerdir. Çetinkaya ve Kaçıranlar [32] bağımsız değişkenler arasında yüksek derecede ilişki olduğu durumda negatif binom ve poisson regresyon modelleri için iki parametrelili tahmin ediciler geliştirmişlerdir. MSE kriterine göre MLE, ridge ve Liu tahmin edicileri kıyaslandığında yeni iki parametrelili tahmin edicilerin daha iyi performans sergilediğini göstermişlerdir. Bulut ve Işıl [33] ters Gauss regresyon modelinde çoklu iç ilişki problemi ile başa çıkmak için yanlı tahmin edici önermişlerdir. Önerilen tahmin edici literatürdeki diğer tahmin edicilerle MSE kriteri

bakımından karşılaştırılmış ve gerçek veri uygulamasıyla da iyi sonuçlar verdiği ortaya konulmuştur.

Regresyon analizde aykırı değerlerin bulunması karşılaşılan bir diğer önemli problemdir. Aykırı değerlerin varlığında önerilen yöntemlerden biri de robust regresyon yöntemidir. Robustluk kavramı ilk olarak Box [34] tarafından ele alınarak Huber [35] ve Hampel [36] tarafından yapılan robustluk kavramı ile ilgili öncü çalışmalar literatüre önemli katkılar sağlamıştır. Nevitt ve Tam [37] veri setinde bulunan aykırı değerlerle OLS yöntemi ile parametrik olmayan robust tahmin yöntem olan Theil ve ağırlıklandırılmış Theil tahmin edicileri ile robust tahmin yöntemlerinden olan en küçük mutlak sapma, en küçük medyan kareler, budanmış en küçük kareler yöntemlerini karşılaştırmışlardır. Yapılan simülasyon sonucunda hataların dağılımına göre tahmin edicilerinin performanslarında farklılıklar olduğunu göstermişlerdir. Wisnowski vd. [38] doğrusal regresyon analizinde birden çok aykırı değer olması durumunda bu aykırı değerleri belirlemek için var olan teknikleri kullanmışlardır. MM tahmin edicisinin ve en küçük medyan kareler, budanmış en küçük kareler tahmin edicilerinin etkin performans sergilediklerini belirlemişlerdir. Schumacker vd. [39] çoklu doğrusal regresyon analizinde OLS yönteminin aykırı değerlerden, çoklu iç ilişkiden ve normal olmayan dağılımdan etkilendiğini söylemişlerdir. Bu sebepten dolayı robust tahmin yöntemleri kullanarak tahminlerin geçerliliğini vurgulamak istemişlerdir. Çalışmalarında MM ve budanmış en küçük kareler tahmin edicilerini OLS yöntemine göre karşılaştırmışlar ve kullandıkları veri seti için MM yönteminin daha etkin performans sergilediğini belirtmişlerdir. Rousseeuw ve Christmann [40] lojistik regresyon modelinde bağımsız değişkenler uzayında ayrılma olduğu durumda parametrelerin tanımlanmasında problemler yaşandığını belirtmişlerdir. Bu durumda mevcut yöntemlerle yapılacak olan yakınsamanın başarısız olacağından veya yanlış cevaplar vereceğinden gizli lojistik regresyon modelinin kullanılmasını önermişlerdir. Croux ve Haesbroeck [41] lojistik regresyon modelinde robust bir tahmin edici olan Bianco ve Yohai (BY) tahmin edicisi için hızlı ve daha kararlı bir algoritma önermişlerdir. Çalışmalarında BY tahmin edicisi ile birlikte ağırlıklandırılmış Bianco Yohai (WBY), MLE, ağırlıklandırılmış en çok olabilirlik, koşullu olarak tarafsız sınırlı etki fonksiyonu tahmin edicisi (CUBIF) ve Mallows tipi kaldırıcı bağımlı ağırlık tahmin edicisi (Mallows) yöntemlerini karşılaştırmışlardır. Karşılaştırma sonucunda bu tahmin edicilerin aykırı değer varlığında etkin sonuçlar elde ettiklerini belirtmişlerdir. Ergül [42] regresyon analizinin varsayımlarının sağlanmadığı durumda robust regresyon yöntemlerini üç farklı veri seti kullanarak karşılaştırmıştır. Ayrıca

kullanılan birçok robust regresyon yönteminin belirtme katsayısı bakımında OLS yönteminden daha iyi performans gösterdiğini ortaya koymuştur. Imon ve Hadi [43] çalışmalarında literatürde yar alan ve tek tek belirleme yöntemi kullanılarak elde edilen aykırı değerlerin veri setinde birden fazla olması durumunda maskeleye ve süpürme problemlerinden etkilenmeyen grup silmeye dayalı yeni bir yöntem olan genelleştirilmiş standart Pearson artıklar yöntemini ele almışlardır. Yıldırım [44] regresyon modeli için çoklu iç ilişki ile aykırı değer problemini birlikte incelemiştir. Çoklu iç ilişki problemi için ridge regresyonu ve aykırı değer belirlenmesi için de bazı hesaplama yöntemlerini önermiştir. Karşılaştırma kriteri olarak belirtme katsayısı, hata kareler ortalaması, F ve t değerlerini kullanarak analizler sonucunda aykırı değerlerin çoklu iç ilişki problemini maskeleyebildiklerini belirtmiştir. Syaiba ve Habshah [45] lojistik regresyon modelinde yüksek kaldırma noktalarının tespiti için yeni bir robust tespit yöntemi önermişlerdir. Önerilen robust lojistik tespit yönteminin performansını Monte Carlo simülasyon yöntemi ve gerçek bir örnekle ortaya koymuşlardır. Bu yöntem sadece doğru yüksek kaldırma noktalarını belirlediğini ve süpürme ile maskeleye etkilerinden de arınmış olduğunu belirtir. Habshah ve Syaiba [46] lojistik regresyon modelinde parametrelerin tahmininde kullanılan MLE tahmin edicilerinin aykırı değerlerden etkilendiği bilindiğinden bu tahmin edicilere alternatif robust lojistik regresyon tahmin edicilerini (Mallows, CUBIF, BY, WBY) kullanmışlardır. Gerçek bir veri seti ve simülasyon çalışmasıyla elde edilen sonuçlarda aykırı değerlerin varlığında MLE tahmin edicisinin MSE kriterine göre etkin olmadığı ve WBY tahmin edicisinin var olan diğer tahmin edicilerden daha etkin performans sergilediğini göstermişlerdir. Imon ve Hadi [47] bağımsız değişkenler uzayında olağan dışı gözlemler olarak tanımladıkları yüksek kaldırma noktalarını lojistik regresyon analizinde birden fazla yüksek kaldırma noktasının belirlenmesi için yeni bir yöntem önermişlerdir. Önerilen robust grup silme tekniğini literatürde yaygın kullanılan birkaç veri seti ve Monte Carlo simülasyon yöntemi ile incelemişlerdir.

Arifin ve Midi [48] hem çoklu iç ilişki hem de yüksek kırılma noktalarının varlığında LRE tahmin edicisini incelemişlerdir ve robust lojistik ridge tahmin edicisi (RLRE) geliştirmişlerdir. Kullanılan tahmin ediciler MSE ve yan kriterlerine göre MLE tahmin edicisi ile karşılaştırılmıştır. Çalışma sonucunda RLRE'nin MLE tahmin edicisinin performansından daha etkin olduğunu göstermişlerdir. Feng vd. [49] lojistik regresyon modelinin değişkenler matrisinde aykırı değer olması durumunda yeni bir robust lojistik regresyon algoritması önermişlerdir. Yu ve Yao [50] yaptıkları çalışma ile doğrusal

regresyon modelinde aykırı değerlerin varlığında robust tekniklerin tanımlanmasını ve yeniden gözden geçirilmesini sağlamışlardır. Bu yöntemlerin kırılma noktası ve etkinlik kavramlarını da kullanarak karşılaştırmalar yapmışlardır. Shariff ve Ferdaos [51] regresyon modelinde tahmin edicilerin güvenilir olmayan sonuçlar vermesine sebep olan çoklu iç ilişki ve aykırı değer problemini ele almışlardır. Genelleştirilmiş M (GM) tahmin edicisi ve ridge parametresinin kombinasyonları ile bu iki problemi çalışmışlardır. Önerilen robust ridge regresyon tahmin edicisinin gerçek veri uygulamasında güvenilir sonuçlar verdiğini göstermişlerdir. İki parametrelili ridge tahmin edicisinin ele alındığı tez çalışmaları incelendiğinde şu sonuçlara ulaşılmıştır. Toker [52], yüksek lisans tez çalışmasında doğrusal regresyon modelinde çoklu iç ilişki olduğu durumda iki parametrelili ridge tahmin edicisi ile birlikte OLS, ridge ve Stein tipi tahmin edicileri MMSE kriterine göre karşılaştırmıştır. Ertaş [53], doktora tez çalışmasında doğrusal regresyonda dayanıklı yanlı tahmin edicileri ele almış ve Liu tipi M tahmin edici, hemen hemen yansız genelleştirilmiş Liu tipi M tahmin edici ve iki parametrelili Ridge-M tahmin edicilerini incelemiştir. İki parametrelili Ridge-M tahmin edicisi için k ve q 'nun seçimini de ele almıştır. Asar [54], doktora tez çalışmasında bazı yanlı tahmin edicilerin lojistik versiyonlarını ele almıştır. Ayrıca, yeni bir genelleştirme yaparak literatürde mevcut olan bazı tahmin ediciler ile birlikte MSE kriteri bakımından performans karşılaştırması yapmıştır. Yeni önerilen tahmin edici iki parametrelili olduğundan parametrelerin seçimi için tekrarlı (iterative) bir metot önermiştir. Işıl [55], yüksek lisans tez çalışmasında doğrusal regresyon modelinde hem çoklu iç ilişki bulunması hem de bağımlı ve bağımsız değişkenlerde aykırı değer bulunması durumunu birlikte ele alıp önerilen robust tahmin edicilerden daha etkin sonuçlar veren yeni tahmin ediciler önermiştir. Önerilen Liu tipi GM tahmin edicisinin diğer tahmin edicilerle karşılaştırıldığında MMSE kriterine göre etkin sonuçlar verdiği görülmüştür. Ayrıca Liu tipi GM için farklı parametre seçimleri kendi içlerinde karşılaştırılmıştır.

1.3. Lojistik Regresyon Modeli

Regresyon analizi, bağımsız değişkenler aracılığı ile bağımlı değişkeni tahmin etmek, değişkenler arasındaki yapıyı matematiksel bir model ile ifade etmektir. Çoklu doğrusal regresyon modeli birden fazla bağımsız değişken olduğu durumda bağımlı değişkenin değerini tahmin etmekte kullanılır. Çoklu doğrusal regresyon modeli,

$$\underline{Y} = \underline{X}\beta + \underline{\varepsilon} \quad (1)$$

şeklinde ifade edilir. Bu modelde n gözlem sayısını; p bağımsız değişken sayısını göstermek üzere; $\underline{Y}$, $n \times 1$ boyutlu bağımlı değişken vektörü; $\underline{X}$, $n \times k$ ($k=p+1$) boyutlu bağımsız değişken değerlerinden oluşmuş gözlem matrisi; $\underline{\beta}$, $k \times 1$ boyutlu bilinmeyen parametreler vektörü; $\underline{\varepsilon}$, $n \times 1$ boyutlu normal dağılıma sahip ($\underline{\varepsilon} \sim N(0, \sigma^2 I)$) hata terimleri vektörüdür. Çoklu doğrusal regresyon modelinde hata terimi üzerinde bazı varsayımları vardır. Bu varsayımlar aşağıdaki gibi verilmiştir.

1. $\underline{Y} = \underline{X}\beta + \underline{\varepsilon}$ fonksiyonunda Y , X matrisinin doğrusal bir fonksiyonudur ve $\underline{\varepsilon}$ bir hata terimi vektörüdür.
2. $E(\varepsilon_i) = 0$, ($i = 1, 2, \dots, n$) hata terimleri ortalaması 0'dır.
3. $Var(\varepsilon_i) = \sigma^2$, hataların varyansı bütün gözlemler için aynıdır ve bu özellik eş varyanslıktır.
4. $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$, hata terimleri ilişkisizdir.
5. $X_1, X_2, \dots, X_k$ bağımsız değişkenleri doğrusal bağımsızdır.
6. $\underline{\varepsilon} \sim N(0, \sigma^2 I)$, hataların dağılımı normaldir.

Regresyon analizinde bağımlı değişken genellikle sürekli ve ölçülebilir bir değişken olarak ele alınır. Ancak bazı durumlarda bağımlı değişken kesikli (kategorik) bir değişken olabilir ve sınırlı sayıda değer alabilir. Doğrusal regresyon analizinin kullanılmadığı durumlardan biri olan bağımlı değişkenin iki ya da ikiden çok düzey içeren değişken olduğu durumda, normallik varsayımı bozulmaktadır. Bu durumda yaygın bir şekilde kullanılan regresyon analizi, lojistik regresyon veya lojit regresyondur. Lojistik regresyon analizi olasılıklı istatistiksel sınıflandırma modelidir ve kategorik olan bağımlı değişkenin sonucunu tahmin etmek için kullanılır [56].

Lojistik regresyon analizi bağımlı değişkenin kategorisine göre ikili veya çoklu olabilmektedir. Bağımlı değişkenin iki kategorili olduğu ikili (binary) lojistik regresyon analizinde bağımsız değişken sürekli, kategorik veya her ikisi de olabilir. Lojistik regresyon analizinde bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin doğrusal olması varsayımı yoktur ve doğrusal olmayan bir yöntemdir. Ayrıca modelin parametrelerinin kolay yorumlanabilmesi ve matematiksel olarak kolay olup anlamlı sonuçlar vermesi lojistik regresyon modelini daha çok tercih edilir olmasını sağlamaktadır [57].

Lojistik regresyon modelinin amacı, istatistikte kullanılan diğer model kurma tekniklerinde olduğu gibi, en az değişkeni kullanarak en iyi uyuma sahip olacak şekilde

bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayabilen ve genel olarak kabul edilebilen bir model kurmaktır [2]. Diğer bir amacı ise verilerin sınıflandırılmasıdır. Bağımlı değişken iki kategorili olduğu durumda doğrusal regresyon modelinin yerine lojistik regresyon modelinin kullanılmasının daha iyi bir yaklaşım olduğu bilinmektedir. Doğrusal regresyon analizinin yukarıda bahsedilen temel varsayımları karşılanırsa, bilinmeyen parametreler için OLS tahmin edicileri yansız ve minimum varyansa sahip oldukları Gauss-Markov teoremi ile bilinmektedir [58].

Bağımlı değişkenin ikili değer olduğu durumda;

1. Hataların normal dağılıma sahip olmaması,
2. Hataların değişen varyanslı olması,
3. Tahmin edilen Y değerlerinin $[0,1]$ aralığının dışında değer alması

gibi problemler ile karşılaşmaktadır. İlk iki problemin doğrusal regresyon modeli varsayımlarını sağlamadığı görülmektedir. Bu problemler karşısında kullanılacak regresyon modellerinden biri de lojistik regresyon modelidir. Modelin, geçerli olduğunu söyleyebilmek için aşağıdaki genel varsayımların sağlanması gerekmektedir [59].

1. Koşullu olasılıklar bağımsız değişkenlerin lojistik bir fonksiyonudur,
2. Bağımsız değişkenler hatasız ölçülür,
3. Bağımsız değişkenler arasında çoklu ilişki yoktur.

Bağımlı değişkenin iki kategorili olması durumunda lojistik regresyon modeli,

$$P(Y = 1|X = x) = \pi(x) = \frac{e^{X\beta + \varepsilon}}{1 + e^{X\beta + \varepsilon}} \quad (2)$$

şeklinde ifade edilir [59]. 0 veya 1 değerini alan Y_i , *Bernoulli* (π_i) dağılımına sahiptir. Lojistik regresyon modelinde bağımlı değişkenin alacağı değerleri $[0,1]$ aralığına çekebilmek için dönüşüm yapılması gerekmektedir [3]. Lojit dönüşümü,

$$\text{lojit } \pi(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = X\beta \quad (3)$$

şeklinde ifade edilir. Lojit dönüşüm X değişkeni ve β parametrelerin tüm olası değerlerine karşı π 'nin $[0,1]$ aralığında kalmasını sağlar. Bu dönüşümde π bir olayın gerçekleşme olasılığını, $1 - \pi$ ise bir olayın gerçekleşmeme olasılığını göstermektedir. $\pi/(1 - \pi)$ oranı,

odds oranı olarak adlandırılmaktadır ve lojistik regresyon modelinde katsayıların yorumlanmasında kullanılmaktadır.

Doğrusal regresyon analizinde bilinmeyen parametrelerin tahmininde OLS yöntemi kullanılmaktadır. Doğrusal regresyon analizi varsayımları sağlandığında OLS yöntemi ile elde edilmiş tahmin ediciler iyi istatistiksel özelliklere sahip tahmin edicilerdir. OLS yöntemi lojistik regresyon analizinde uygulandığında bağımlı değişkenin kategorik olmasından dolayı tutarlı sonuçlar vermemektedir. Daha önceden de bahsedildiği gibi hata terimleri normal dağılım göstermeyeceğinden OLS yöntemini kullanmak doğru sonuçlar vermeyecektir. Bilinmeyen parametre tahminleri için uygulamada en yaygın kullanılan yöntem MLE yöntemidir. Parametre tahminlerinde yaygın olarak kullanılan diğer bir yöntem ise yeniden ağırlıklandırılmış iteratif en küçük kareler yöntemidir [60].

1.3.1. En Çok Olabilirlik Yöntemi

Lojistik regresyon analizinde MLE yöntemi parametre tahmini için en sık kullanılan yöntemlerden biridir. MLE yöntemi, gözlemlenen veri kümesinin elde edilme olasılığını en büyük yapan bilinmeyen parametre değerlerinin tahminlerini elde etmeyi amaçlamaktadır. MLE tahmin edicileri, olabilirlik fonksiyonunu en büyük yapan parametre değerleri elde edecek şekilde belirlenmektedir.

Lojistik regresyon analizinde MLE yöntemi uygulandığında olabilirlik fonksiyonu elde edilirken Y bağımlı değişkenin Bernoulli dağılımına sahip olmasından dolayı, bağımsız değişken değeri verildiğinde $y_i = 0$ olduğunda gözlemlerin olabilirlik fonksiyonuna katkısı $1 - \pi_i$, $y_i = 1$ olduğunda ise olabilirlik fonksiyonuna katkısı π_i kadardır.

Her bir gözlemi Bernoulli dağılımından gelen bağımsız rasgele değişkenlerin olabilirlik fonksiyonu,

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (4)$$

şeklinde elde edilir. İşlem kolaylığı açısından olabilirlik fonksiyonunun logaritması alınır ve log-olabilirlik fonksiyonu,

$$\begin{aligned}
l(\beta; y) &= \sum_{i=1}^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \\
&= \sum_{i=1}^n y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + \sum_{i=1}^n \log(1 - \pi_i) \\
&= \sum_{i=1}^n y_i (X\beta) - \sum_{i=1}^n \log[1 + \exp(X\beta)]
\end{aligned} \tag{5}$$

şeklinde ifade edilir. En çok olabilirlik tahmin edicilerini elde etmek için Eşitlik (5)'te verilen ifadenin ilgilenilen parametrelere göre türevlerinin alınması gerekmektedir. β parametrelerine göre kısmi türev alınıp sıfıra eşitlenirse,

$$\begin{aligned}
\sum_{i=1}^n (y_i - \pi_i) &= 0 \\
\sum_{i=1}^n x_i (y_i - \pi_i) &= 0
\end{aligned} \tag{6}$$

denklemleri elde edilir. Elde edilen bu denklemler için türev işlemleriyle doğrudan çözülerek kapalı bir çözüm elde edilememektedir. Bu nedenle bu denklemin çözümü için iteratif yeniden ağırlıklı en küçük kareler (IRLS) yöntemi uygulanabilir. IRLS algoritması kullanılarak elde edilen lojistik tahmin edici (LE),

$$\hat{\beta}_{LE} = (X' \hat{W} X)^{-1} X' \hat{W} \hat{Z} = (M)^{-1} v \tag{7}$$

biçiminde elde edilir. Burada $M = X' \hat{W} X$, $v = X' \hat{W} \hat{Z}$, $\hat{W} = \text{diag}(\hat{\pi}_i(1 - \hat{\pi}_i))$ ve $\hat{Z}$, $n \times 1$ sütun vektörünün i . elemanı $\hat{z}_i = \text{logit}(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)}$ şeklinde elde edilir. $\hat{\beta}_{LE}$ tahmin edicisinin MMSE ve MSE fonksiyonları ise,

$$MMSE(\hat{\beta}_{LE}) = (M)^{-1} \tag{8}$$

$$MSE(\hat{\beta}_{LE}) = \text{tr}(MMSE) = \sum_{j=1}^{p+1} \frac{1}{\lambda_j} \tag{9}$$

şeklinde elde edilir.

1.4. Lojistik Regresyon Modelinde Yanlı Tahmin Ediciler

Bu bölümde genel olarak çoklu iç ilişki problemi tanıtılacak ve çoklu iç ilişki problemi için çözüm yöntemlerinden bahsedilecektir. Ayrıca lojistik regresyon modeli için yanlı tahmin ediciler ele alınacaktır.

1.4.1. Çoklu İç İlişki Problemi

Çoklu doğrusal regresyon analizinde olduğu gibi, lojistik regresyon analizinde de model kurulurken ele alınan bağımsız değişkenler arasında ilişki olmaması gerekmektedir. Fakat uygulamada bu varsayım her zaman sağlanmayabilir. Bu durumda bağımsız değişkenler arasında bağımsızlık varsayımı geçerli olmaz ve çoklu iç ilişki problemine neden olur. Bağımsız değişkenler arasında çoklu iç ilişki söz konusu olduğunda elde edilen tahmin ediciler minimum varyanslılık gibi önemli bir özelliği sağlayamamaktadır. Eşitlik (1)'deki regresyon modeli ele alındığında;

$\underline{X}_j$, X matrisinin j 'inci sütun vektörü olsun: $X = [\underline{X}_1, \dots, \underline{X}_p]$

$$\sum_{j=1}^p a_j \underline{X}_j = 0 \quad (10)$$

olacak şekilde hepsi sıfır olmayan $a_1, \dots, a_p$ sabitler varsa $\underline{X}_1, \dots, \underline{X}_p$ vektörleri doğrusal bağımlıdır. Bu durumda bağımsız değişkenler arasında çoklu iç ilişki vardır denir.

Çoklu iç ilişki probleminin birçok nedeni olabilir. Bunlar; uygulanan veri toplama metodu, modeldeki zorluklar, modelin belirlenememesi ve modelin aşırı tanımlanması ($n < p$) olarak söylenebilir [61].

1.4.2. Çoklu İç İlişki Problemi İçin Belirleme Yöntemleri

Çoklu iç ilişkinin belirlenmesi için çeşitli yöntemler geliştirilmiştir. Bu yöntemlerden bazıları aşağıda maddeler halinde verilmiştir.

1. Korelasyon matrisinin incelenmesi: $X'X$ matrisinin köşegen dışı elemanlarının (r_{ij}) incelenmesi ile çoklu iç ilişkinin varlığı tespit edilir. Bire yakın $|r_{ij}|$ değerlerinde x_i ve x_j doğrusal bağımlıdır.

2. Varyans şişirme faktörü (VIF): $(X'X)^{-1}$ matrisinin t 'inci köşegen elemanına k 'ıncı varyans şişirme faktörü (VIF_t) denir. $VIF_t = (1 - R_t^2)^{-1}$ olarak ifade edilir [62]. VIF değeri 10'dan büyükse çoklu iç ilişki vardır denir.
3. $X'X$ matrisinin özdeğerlerinin analizi: Çoklu iç ilişki anlamında ilk araştırmacılardan biri olan Ragnar Frisch, çoklu iç ilişkiyi öz değerlerle ilişkilendirmiştir. $X'X$ matrisinin $\lambda_{\min}$ ve $\lambda_{\max}$ minimum ve maksimum öz değerleri olmak üzere; $K = \lambda_{\max}/\lambda_{\min}$ olarak belirlenen değer 100'den büyük ise çoklu iç ilişki vardır denir.
4. Diğer tanımlama yöntemleri: Regresyon katsayılarının işaretlerinin veya büyüklüklerinin tahmin edilenden farklı olması, bağımsız değişkenlerin eklenmesi veya çıkartılmasında regresyon katsayılarının tahmininde büyük değişikliklere neden olması ve modelin tümüyle yüksek anlam gücüne sahipken tahmin edilen parametrelerin anlamsız çıkması durumunda çoklu iç ilişkinin varlığından söz edilebilir.

1.4.3. Çoklu İç İlişki Problemi İçin Çözüm Yöntemleri

Çoklu iç ilişki problemini ortadan kaldırmak için; ek verinin toplanması, modelin yeniden belirlenmesi ve alternatif regresyon modellerinin uygulanması kullanılan yöntemler arasında yer alabilir. Uygulamada daha çok çoklu iç ilişki problemini ortadan kaldırmak için alternatif regresyon modelleri kullanılmaktadır. Hoerl ve Kennard [14] tarafından önerilen ridge regresyon, çoklu iç ilişki problemi ile karşılaşıldığı durumda kullanılan yanlı regresyon yöntemlerinden birisidir. Yanlı regresyon tahmin edicileri arasında ridge tahmin edicisi yaygın kullanıma sahip bir tahmin edicidir.

1.4.4. Yanlı Tahmin Ediciler

Çoklu iç ilişki problemi ile karşılaşıldığında yapılan analizler tahmin edicinin varyans değerini olduğundan büyük olarak belirlemekte ve regresyon katsayıları gerçek değerlerinden farklı çıkmaktadır. Bu duruma çözüm olarak Hoerl ve Kennard [14] OLS tahmin edicisi matrisinin köşegen elemanlarına bir k sabiti ekleyerek ridge tahmin edicisini önermişlerdir. Ridge tahmin edicisi yansızlıktan ödün verdiği için ele alınan parametreleri

yanlı olarak tahmin etmektedir ancak tahmin ediciler daha küçük varyansa sahip olmaktadır. Çoklu doğrusal regresyon modeli için ridge tahmin edicisinin amaç fonksiyonu,

$$\begin{aligned} S(\underline{\beta}_R)^2 &= \|\underline{Y} - X\underline{\beta}_R\|^2 + k \|\underline{\beta}_R\|^2 = \underline{Y}'\underline{Y} - 2\underline{\beta}_R'X'\underline{Y} + \underline{\beta}_R'X'X\underline{\beta}_R + k\underline{\beta}_R'\underline{\beta}_R \\ &= 1 - 2\underline{\beta}_R's + \underline{\beta}_R'N\underline{\beta}_R + k\underline{\beta}_R'\underline{\beta}_R \end{aligned} \quad (11)$$

şeklinde ifade edilir. Burada $N = X'X$, $s = X'\underline{Y}$, $\underline{Y}'\underline{Y} = 1$ ve k ($k > 0$) yanalılık parametresidir. Denklem sisteminin en küçüklenmesinden

$$(N + kI)\underline{\beta}_R = s \quad (12)$$

β bilinmeyen parametrelerinin ridge tahmin edicisi

$$\hat{\beta}_R = (N + kI)^{-1}s \quad (13)$$

biçiminde elde edilir. k yanalılık parametresinin seçimi ridge tahmin edicisinin performansını etkilemektedir.

1.4.4.1. Lojistik Ridge Tahmin Edicisi

Bağımlı değişkenin iki kategorili olduğu durumda çoklu iç ilişkinin giderilmesi için Schaefer vd. [15] tarafından önerilen LRE tahmin edicisi,

$$\begin{aligned} \hat{\beta}_{LRE} &= (X'\widehat{W}X + kI)^{-1}X'\widehat{W}\underline{\hat{Z}} \\ &= (M + kI)^{-1}v \end{aligned} \quad (14)$$

şeklinde ifade edilir. Burada M ve v Eşitlik (7)'de tanımlandığı gibidir.

Lojistik ridge regresyon yönteminde MSE hesaplamasında varyans ve yan (bias) fonksiyonlarının elde edilmesinde kolaylık açısından kanonik form kullanılır. $M = Q\Omega Q'$ matrisinin kanonik formu olsun ve burada Q , sütunları M matrisinin öz vektörlerini içeren bir matris ve Ω , M matrisinin özdeğerlerini içeren diagonal (köşegen) bir matristir. $\alpha = Q'\underline{\beta}$ olacak şekilde yan ve kovaryans,

$$Bias(\hat{\beta}_{LRE}) = -kQ\Omega_k^{-1}\alpha \quad (15)$$

$$Cov(\hat{\beta}_{LRE}) = Q\Omega_k^{-1}\Omega\Omega_k^{-1}Q' \quad (16)$$

şeklinde elde edilir.

Burada $\Omega = diag(\lambda_1, \lambda_2, \dots, \lambda_{p+1})$ ve $\Omega_k = diag(\lambda_1 + k, \lambda_2 + k, \dots, \lambda_{p+1} + k)$ olarak elde edilir. $\hat{\beta}_{LRE}$ tahmin edicisinin MMSE ve MSE fonksiyonları ise,

$$\begin{aligned} MMSE(\hat{\beta}_{LRE}) &= Cov(\hat{\beta}_{LRE}) + Bias(\hat{\beta}_{LRE})Bias(\hat{\beta}_{LRE}) \\ &= Q\Omega_k^{-1}\Omega\Omega_k^{-1}Q' + (-kQ\Omega_k^{-1}\alpha)(-kQ\Omega_k^{-1}\alpha)' \end{aligned} \quad (17)$$

$$MSE(\hat{\beta}_{LRE}) = tr(MMSE) = \sum_{j=1}^{p+1} \frac{\lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^{p+1} \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2} \quad (18)$$

biçiminde elde edilir. Burada ilk terim asimptotik varyans ve ikinci terim ise tahmin edicinin yan karesidir.

Literatürde k yanlılık parametresinin seçimi için birçok çalışma bulunmaktadır ve çalışmalar hala devam etmektedir. En uygun k seçimi için genel bir kural bulunmamaktadır. Lojistik regresyon modelinde k seçimi için önerilen yöntemlerden bazıları aşağıdaki şekildedir:

➤ Schaefer vd. [15] tarafından önerilen k yanlılık parametreleri,

$$\hat{k}_{SRW1} = \frac{p+1}{\hat{\alpha}'\hat{\alpha}} \quad (19)$$

$$\hat{k}_{SRW2} = \frac{p}{\hat{\alpha}'\hat{\alpha}} \quad (20)$$

$$\hat{k}_{SRW2} = \frac{1}{\hat{\alpha}'\hat{\alpha}} \quad (21)$$

şeklinde elde edilir. Burada $\hat{\alpha} = Q'\hat{\beta}_{MLE}$ olarak hesaplanır. Q , sütunları $X'X$ matrisinin özvektörlerinden oluşan bir matristir.

➤ Kibria vd. [20] tarafından önerilen bazı k parametresi seçimleri ise

$$k_6 = \max(m_j), \quad k_8 = \left(\prod_{j=1}^{p+1} m_j\right)^{1/p+1}, \quad k_{11} = \max\left(\frac{1}{q_j}\right) \quad \text{ve} \quad k_{16} = \text{median}(q_j)$$

şeklinde elde edilir. Burada $m_j = \frac{s^2}{\hat{\alpha}_{max}^2}$, $s^2 = \frac{(y-\hat{\pi})'(y-\hat{\pi})}{n-p-1}$ ve $q_j = \frac{\lambda_{max}}{(n-p-1)s^2 + \lambda_{max}\hat{\alpha}_j^2}$ biçiminde ifade edilir.

1.4.4.2. İki Parametrelili Ridge Tahmin Edicisi

Lipovetsky ve Conklin [16], doğrusal regresyon modelinde ridge tahmin edicisinin asimptotik özelliklerini inceleyerek modelin çoklu belirleyicilik katsayısını (R^2) en büyük yapmak için ikinci bir parametre ekleyerek modeli geliştirmişlerdir. Bu tahmin edici hem varyansı azaltmak için kullanılan yanlılık parametresini hem de regresyon uyumunun kalitesini göz önünde bulunduracak ikinci bir parametreden oluşmaktadır. Regresyon model uyumunu en büyük yapmaya çalışan doğrusal regresyon modeli için TPRES'nin amaç fonksiyonunu elde etmeden önce OLS tahmin edicisinin amaç fonksiyonu,

$$\begin{aligned} S(\beta)^2 &= \left\| \underline{Y} - X\underline{\beta} \right\|^2 = \underline{Y}'\underline{Y} - 2\underline{\beta}'X'\underline{Y} + \underline{\beta}'X'X\underline{\beta} \\ &= 1 - 2\underline{\beta}'\underline{s} + \underline{\beta}'N\underline{\beta} \end{aligned} \quad (22)$$

şeklinde yazılabilir. Burada N ve s Eşitlik (11)'de tanımlandığı gibidir. β parametrelerinin OLS tahmin edicisi,

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'\underline{Y} = N^{-1}s \quad (23)$$

şeklinde de ifade edilebilir. Ayrıca MSE fonksiyonunun hesaplanmasında kolaylık sağlayacak Eşitlik (22) modelinin kanonik formu,

$$\underline{Y} = C\underline{\alpha} + \underline{\varepsilon} \quad (24)$$

şeklinde ifade edilir. Burada $C = QX$ ve $\underline{\alpha} = Q\underline{\beta}'$ 'dir. $C'C = Q'X'XQ = \Omega = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{p+1})$ şeklindedir. Ω , $X'X$ matrisinin özdeğerlerini içeren köşegen bir matristir ve λ , $X'X$ matrisinin özdeğerleridir.

OLS tahmin edicisinin kanonik formu,

$$\hat{\alpha} = \Omega^{-1} C' Y \quad (25)$$

biçiminde yazılabilir. Tahmin edilen modelin gerçek modele uyumluluğunu belirlemek için hesaplanan belirlilik katsayısı ise,

$$R^2 = 1 - S(\beta)^2 = 2\beta' s + \beta' N \beta = \beta' (2s - N \beta) \quad (26)$$

biçiminde elde edilir.

Lipovetsky ve Conklin [16], Eşitlik (22)'den mantık yürütme yoluyla ortogonalite varsayımlarını içeren daha genel bir amaç fonksiyonu tanımlamışlardır. Bu amaç fonksiyonu,

$$\begin{aligned} S(\beta)^2 &= \|Y - X\beta\|^2 + q_1 \|\beta\|^2 + q_2 \|X'Y - \beta\| + q_3 \|Y' (Y - X\beta)\|^2 \\ &= (1 - 2\beta' s + \beta' N \beta) + q_1 (\beta' \beta) + q_2 (s' s - 2\beta' s + \beta' \beta) + \\ &\quad + q_3 (1 - 2\beta' s + \beta' s s' \beta) \end{aligned} \quad (27)$$

şeklinde elde edilir.

Burada ilk terim modelden sapmaların OLS minimizasyonuna, q_1 terimi tahminlerin normunu en aza indirmeye, q_2 terimi bağımlı değişken ile s korelasyon çiftlerinden β parametre tahminlerinin sapmalarını en aza indirmeye ve q_3 terimi de $Y' \varepsilon = Y' Y - Y' \hat{Y} = 1 - Y' \hat{Y} = 1 - R^2$ ilişkiye karşılık gelir. Eşitlik (27)'nin en küçüklemesinden elde edilen matris denklemi,

$$\frac{\partial S(\beta)^2}{\partial \beta} = -2s + 2N\beta + q_1 2\beta - q_2 2s + q_2 2\beta - q_3 2s + q_3 s s' \beta = 0 \quad (28)$$

biçimindedir. Eşitlik (28) düzenlenirse

$$\begin{aligned} (N + q_1 + q_2 + q_3 s' s) \beta &= q s \\ (N + kI) \beta &= q s \end{aligned} \quad (29)$$

şeklinde ifade edilir ve gerekli düzenlemelerden sonra TPRE tahmin edicisi,

$$\hat{\beta}_{TPRE} = q(N + kI)^{-1}s \quad (30)$$

biçiminde elde edilmiş olur. Burada k ve q pozitif parametrelerdir. Modelin uyumunu belirlemek için Eşitlik (26) yardımıyla,

$$R^2 = 2qs'(N + kI)^{-1}s - q^2s'(N + kI)^{-1}N(N + kI)^{-1}s \quad (31)$$

belirlilik katsayısı elde edilir ve bu belirlilik katsayısı q parametresinin karesel bir fonksiyonudur. Bu fonksiyonun en küçüklenmesiyle q parametresinin gerçek değeri elde edilir. q için en uygun değer,

$$q = \frac{s'(N + kI)^{-1}s}{s'(N + kI)^{-1}N(N + kI)^{-1}s} \quad (32)$$

biçiminde elde edilir ve burada $q > 1$ olduğu da görülmektedir.

1.4.4.3. İki Parametrelili Lojistik Ridge Tahmin Edicisi

Doğrusal regresyon modelinde tanımlanan iki parametrelili ridge tahmin edicisinin lojistik regresyon genelleştirilmesi olan iki parametrelili lojistik ridge tahmin edicisi (TPLRE),

$$\hat{\beta}_{TPLRE} = q(M + kI)^{-1}v \quad (33)$$

biçiminde tanımlanmıştır [54]. Burada tanımlanan M ve v Eşitlik (7)'de tanımlandığı gibidir. Eşitlik (26) yardımıyla modelin belirlilik katsayısı,

$$R_{TPLRE}^2 = 2qv'(M + kI)^{-1}v - q^2v'(M + kI)^{-1}M(M + kI)^{-1}v \quad (34)$$

şeklinde elde edilir. Eşitlik (34)'ün en küçüklenmesi ile q parametresinin en uygun değeri,

$$q = \frac{v'(M + kI)^{-1}v}{v'(M + kI)^{-1}M(M + kI)^{-1}v} \quad (35)$$

biçiminde hesaplanır. TPLRE'nin yan ve kovaryansı,

$$Bias(\hat{\beta}_{TPLRE}) = (qQ\Omega_k^{-1}\Omega Q' - I)\underline{\beta} \quad (36)$$

$$Cov(\hat{\beta}_{TPLRE}) = q^2 Q\Omega_k^{-1}\Omega\Omega_k^{-1}Q' \quad (37)$$

şeklinde elde edilir. Diğer tahmin edicilerle kıyaslama kriteri olan MMSE ve MSE fonksiyonları ise,

$$MMSE(\hat{\beta}_{TPLRE}) = q^2 Q\Omega_k^{-1}\Omega\Omega_k^{-1}Q' + (qQ\Omega_k^{-1}\Omega Q' - I)\underline{\beta} \left[(qQ\Omega_k^{-1}\Omega Q' - I)\underline{\beta} \right]' \quad (38)$$

$$MSE(\hat{\beta}_{TPLRE}) = \sum_{j=1}^{p+1} \frac{q^2 \lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^{p+1} \frac{\alpha_j^2 (q\lambda_j - \lambda_j - k)^2}{(\lambda_j + k)^2} \quad (39)$$

biçiminde elde edilir.

Eşitlik (39)'da k ve q parametrelerinin seçimi modeli doğrudan etkilediği için önemlidir. İki parametrelili lojistik ridge tahmin yönteminde Lipovetsky ve Conklin [16] tarafından önerilen q parametresi lojistik regresyon modeli için genelleştirilmiştir ve mevcut parametre değeri olarak kullanılmaktadır. Asar ve Genç [63], yaptıkları çalışmada iteratif bir yöntem kullanarak iki parametrelili lojistik ridge tahmin edicisi için k ve q parametrelerinin seçimini önermişlerdir. Önerilen bu yöntem aşağıdaki gibidir:

$$\hat{k} = \frac{\sum_{j=1}^{p+1} (q\lambda_j + (q-1)\lambda_j^2 \hat{\alpha}_j^2)}{\sum_{j=1}^{p+1} \lambda_j \hat{\alpha}_j^2} \quad (40)$$

Adım1: $q_0 \in R$, $q_0 > 1$ olacak şekilde $q = q_0$ belirle,

Adım 2: $q = q_0$ ve Eşitlik (40) kullanılarak k 'nın başlangıç değerini hesapla,

Adım 3: Eşitlik (35) ile q elde et,

Adım 4: TPLRE'yi hesapla,

Adım 5: Adım 2'ye dön ve k 'yı 3.adımda elde edilen q ve Eşitlik (40) kullanılarak hesapla,

Adım 6: Adım 5'te elde edilen k 'yı kullanarak Eşitlik (35) ile q elde et,

Adım 7: TPLRE'yi tekrar hesaplayarak eski ve yeni TPLRE değerleri arasındaki mesafenin tolerans değerinin 10^{-6} 'dan küçük olup olmadığını kontrol et.

Asar ve Genç elde edilen k ve q parametreleri kullanarak TPLRE tahmin edicisini elde etmişlerdir.

Bu tez çalışmasında ele alınan diğer bir problem ise veri setinde aykırı değer bulunması durumudur. Olağan dışı gözlem olarak da nitelendirilen aykırı değer alt bölümde ayrıntılı bir şekilde ele alınmıştır.

1.5. Lojistik Regresyon Modelinde Olağan Dışı Gözlemler

Lojistik regresyon modelinde olağan dışı gözlemler; aykırı değer, yüksek kaldıraç noktası ve etkili gözlem olmak üzere üç kategoride sınıflandırılmaktadır [64]. Regresyon analizinde olduğu gibi lojistik regresyon analizinde de olağan dışı gözlemler birlikte incelenerek ele alınır.

Aykırı değer problemi istatistiksel analizlerin var oluşu kadar eski bir problemdir. İstenmeyen, temsil etmeyen, veri kümesi içerisinde diğer gözlemler ile tutarsız görünen gözlem olarak tanımlanmıştır [65]. Doğrusal regresyon modelinde ise en büyük artık değere sahip veya veri kümesindeki diğer gözlemlere uzak olan değer aykırı değer olarak adlandırılmıştır.

Barnett ve Levis [66] aykırı değer ölçme hatasından, örnekleme hatasından ve doğal çeşitlilikten ortaya çıkabileceğini söylemişlerdir. Aykırı değerler x yönünde (bağımsız değişkenler yönünde), y yönünde (bağımlı değişken yönünde) ve hem x hem y yönünde (hem bağımlı hem bağımsız değişkenler yönünde) olabilirler.

y yönünde aykırı değerler; büyük artık değere sahip olan gözlemler ya da regresyon doğrusundan y yönünde uzakta olan değerlerdir.

x yönünde aykırı değerler; veri kümesindeki gözlem değerlerinin ortalama x değerinden uzakta bulunan gözlem değerlerine denir. x yönündeki aykırı değerler iki şekilde adlandırılır.

- İyi kaldıraç (good leverage) noktası: Regresyon katsayılarının doğruluğunu arttıran x yönündeki noktalara iyi kaldıraç noktası denir.

- Kötü kaldıraç (bad leverage) noktası: Regresyon doğrusunun eğimini çok fazla değiştiren x yönünde uzakta olan değerlere kötü kaldıraç noktası denir.

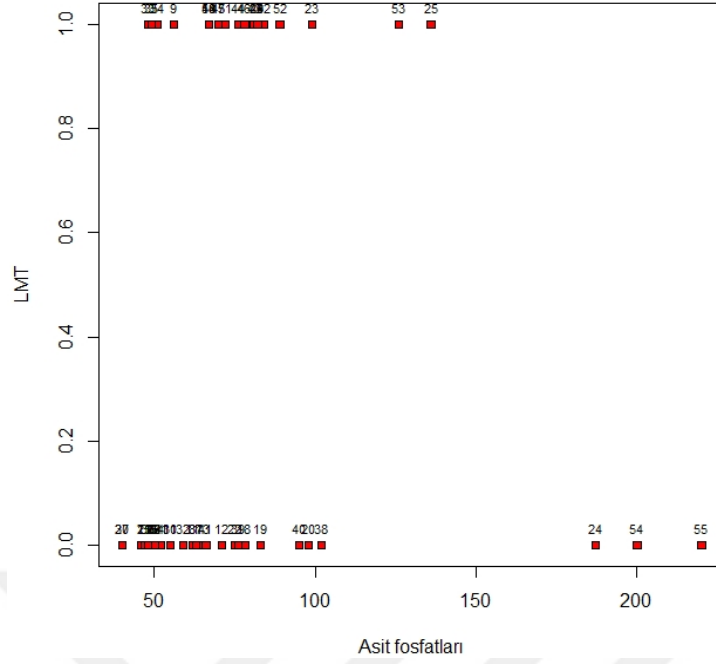
Hem x hem de y yönündeki aykırı değerler; veri kümesinde hem x hem de y yönünde uzakta bulunan değerlere denir.

Etkili gözlem (influential observation), veri kümesinden çıkartıldığında hesaplanan çeşitli istatistikleri (katsayıları, standart hataları vb.) büyük ölçüde değiştiren gözlemlere denir [67].

Chatterjee ve Hadi [68] etkili gözlemi, regresyon modeli üzerinde diğer gözlemlere göre oldukça büyük etkisi olan gözlemler olarak tanımlamışlardır. Bu durumda, etkili gözlemlerin veri kümesinden çıkartılması, parametre tahminleri üzerinde büyük etkiye sahip olmasından dolayı yapılan analiz sonuçlarını önemli derecede etkileyecektir.

Yüksek kaldıraç noktası (uç değer), veri kümesinden uzakta bulunan değer olarak adlandırılır. x yönündeki diğer gözlem noktalarıyla kıyaslandığında uyumsuzluk gösteren yüksek kaldıraç noktaları, x yönündeki aykırı değerlerdir. Kaldıraç ifadesi bağımsız değişkene ilişkin bir kavramdır.

Lojistik regresyon modelinde aykırı değerleri tanımlamak ve teşhis etmek doğrusal regresyon modeline göre oldukça zordur [69]. Modelde aykırı değerlerin olması durumunda, lojistik regresyon modeli bağımlı değişkenin 0 veya 1 değerlerini alma olasılıklarını modellediğinden yapılan analiz ile elde edilen başarı olasılığı küçük olduğu halde gözlenen durum başarılı, elde edilen başarı olasılığı yüksek olduğu halde gözlenen durum başarısız olarak belirlemektedir. Lojistik regresyon modelinde olağandışı gözlemler içeren Düzenlenmiş Brown verisinin saçılım grafiği Şekil 1’de verilmiştir.



Şekil 1.Düzenlenmiş Brown verisi için saçılım grafiği

Şekil 1 incelendiğinde 24, 25, 53, 54 ve 55'inci gözlemlerin verinin geri kalanından uzak olduğu görülmektedir. Bu olağan dışı gözlemlerin belirlenebilmesi için lojistik regresyon modelinde teşhis (tanı, belirleme) yöntemleri kullanılır.

Doğrusal regresyon modelinde aykırı değer, etkili gözlem ve yüksek kaldıraç noktalarının belirlenmesi için birçok tanı yöntemi literatürde ayrıntılı olarak incelenmiştir [70],[65],[66].

Olağan dışı gözlemlerin varlığı doğrusal regresyon modelinde olduğu gibi lojistik regresyon modelinde de parametre tahminlerinin ve yorumlarının yanlış yapılmasına neden olmaktadır. Bu nedenle veri setinde var ise olağan dışı gözlemlerin belirlenmesi ve buna bağlı olarak analizlerin yapılması gerekmektedir.

Bu bölümde literatürde mevcut olan lojistik regresyon modeli için aykırı değer ve yüksek kaldıraç noktaları belirleme yöntemlerinden birkaçı ele alınacaktır.

Çalışmada teşhis yöntemlerinin açıklanmasında artıklar ve Hat matrisi önemli rol oynamaktadır. Eşitlik (2) de verilen lojistik regresyon modeli için hata terimi,

$$\varepsilon = \begin{cases} 1 - \pi(x), & y = 1 \\ -\pi(x), & y = 0 \end{cases} \quad (41)$$

fonksiyonunun bir parçasıdır ve ortalaması 0, varyansı $\pi(x)[1 - \pi(x)]$ olan binom dağılımına sahiptir [69]. i 'inci değişkenin tahmin değeri $\hat{y}_i = \hat{\pi}_i$ olan lojistik regresyon modeli için i 'inci artık,

$$e_i = y_i - \hat{\pi}_i \quad (42)$$

şeklindedir. Eşitlik (42)'de görüldüğü gibi artıklar ölçekten bağımsız değildir ve aykırı değerleri belirlemede doğrudan kullanılamazlar [43].

Olağan dışı gözlemleri belirleme yöntemlerinin tespitinde diğer bir istatistik olan Hat matrisi (H), her bir gözlemin her bir tahmin değeri üzerindeki etkisini göstermektedir. Hat matrisinin köşegen elemanları h_i kaldıraç olarak adlandırılır ve bu değerler,

$$h_i = x_i'(X'X)x_i \quad (43)$$

şeklinde ifade edilir. Her bir bağımsız X değişkeninin ortalaması arasındaki uzaklık olarak da değerlendirilebilir.

Hat matrisi hem doğrusal regresyon hem de lojistik regresyon analizinde aykırı değer, yüksek kaldıraç noktası ve etkili gözlem teşhis ölçülerinin temelini oluşturmaktadır. Pregibon [8] IRLS yöntemi kullanarak lojistik regresyon modeli için önerdiği Hat matrisini,

$$H = W^{\frac{1}{2}}X(X'WX)^{-1}X'W^{\frac{1}{2}} \quad (44)$$

şeklinde tanımlamıştır. Burada W köşegen elemanları,

$$W: w_i = \hat{\pi}_i(1 - \hat{\pi}_i) \quad (45)$$

biçiminde tanımlanmış $n \times n$ köşegen bir matristir. Pregibon [8] lojistik regresyon modeli için artıkların, Eşitlik (46) şeklinde tanımlanacağını göstermiştir.

$$e_i = y_i - \hat{\pi}_i \approx (1 - h_{ii})y_i \quad (46)$$

Burada h_{ii} , Hat matrisinin köşegen elemanlarını gösterir ve aşağıdaki gibi doğrudan hesaplanır. Eşitlik (47)'de $x_i = [1, x_{1i}, x_{2i}, \dots, x_{pi}]$ $1 \times k$ boyutlu vektördür.

$$h_{ii} = \hat{\pi}_i(1 - \hat{\pi}_i)x_i'(X'WX)^{-1}x_i \quad (47)$$

1.5.1. Aykırı Değer Belirleme Yöntemleri

Lojistik regresyon modelinde aykırı değerleri belirleme yöntemlerinin amacı modelin kötü uyum gösterdiği ya da parametre tahminleri üzerinde paylaştıkları etkiden daha fazlasını gösteren gözlemleri tespit etmektir [69]. Aykırı değerlerin varlığı parametre tahminlerinin ve yorumlarının yanlış yapılmasına neden olmaktadır. Lojistik regresyon modelinde bağımlı değişkenin 0 veya 1 değerini alma olasılığı hesaplanırken aykırı değerlerin bulunması halinde tahmin edilen başarı olasılığı küçük olduğu durumda gözlenen durum başarılı, başarı olasılığı yüksek olduğu durumda gözlenen durum başarısız olarak değerlendirilebilmektedir. Lojistik regresyon modeli için aykırı değer belirleme yöntemleri aşağıda incelenmiştir.

1.5.1.1. Pearson Artıklar

Pearson artıkları, gözlenen ve tahmin edilen değerler arasındaki farkın standart hatasına oranıdır ve bu artıklar,

$$r_{pi} = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}} \quad (48)$$

şeklinde hesaplanır. Burada $\hat{\pi}_i$ lojistik regresyon modelinde i 'inci değişkenin tahmin değeridir. r_{pi} değerinin karesi i 'inci gözlemin Pearson'ın ki-kare istatistiğine olan katkısıdır ve adı bu yüzden Pearson artığıdır. Pearson ki-kare istatistiği aşağıdaki formül yardımıyla hesaplanır [69].

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\pi}_i)^2}{\hat{\pi}_i(1 - \hat{\pi}_i)} \quad (49)$$

Hesaplanan bir gözlemin Pearson artık değeri, kabul edilen kritik değeri mutlak değer olarak aşıyorsa o değere aykırı değerdir denir. Parmak kuralı ile aykırı değer belirlenmek istenirse, $|r_{pi}| > 3$ ise ilgili gözleme aykırı değerdir denir [59].

1.5.1.2. Standart Pearson Artıklar

Pearson artıklarını ölçekten bağımsız hale getirmek yani istenilen özelliklere sahip olmalarını sağlamak için artıklar standartlaştırılır. Eşitlik (48) Pearson artıklarının, $\sqrt{1 - h_{ii}}$ 'ye bölünmesi ile elde edilir.

$$r_{spi} = \frac{r_{pi}}{\sqrt{1 - h_{ii}}} \quad (50)$$

Standartlaştırılmış Pearson artıklarının, Pearson artıklarına tercih edildiği literatür çalışmaları da bulunmaktadır [69]. Ayrıca Standart Pearson artıklarının varyans değeri 1'dir ve n örneklem hacmi büyüdükçe artıklar standart normal dağılıma sahip olacaktır. Bu özellik, Pearson artık değeri ile karşılaştırıldığı durum için bir avantajdır. $|r_{spi}| > 3$ olduğu durumda bu gözlemin aykırı değer olduğu söylenir [47].

1.5.1.3. Sapma Artıklar

Sapma artığı, sapmaların farkları üzerine oluşturulmuş bir aykırı değer belirleme yöntemidir. Lojistik regresyon modeli için sapma artığı,

$$rd_i = \text{sign}(y_i - \hat{\pi}_i) \left\{ 2 \left[y_i \ln \left(\frac{y_i}{\hat{\pi}_i} \right) + (1 - y_i) \ln \left(\frac{1 - y_i}{1 - \hat{\pi}_i} \right) \right] \right\}^{\frac{1}{2}} \quad (51)$$

şeklinde verilmiştir. Burada $\text{sign}(y_i - \hat{\pi}_i)$ fonksiyonu, $(y_i - \hat{\pi}_i)$ değerine göre pozitif ya da negatif değer alan bir fonksiyondur. Eşitlik (51) de verilen i 'inci sapma artıkları değeri $|rd_i| > 2$ olduğunda, ilgili gözlemin aykırı değer olduğu söylenmektedir [69].

Sapma artıkları $\sqrt{1 - h_{ii}}$ ile bölünerek birim varyansa sahip olacak şekilde de standartlaştırılabilir ve standartlaştırılmış sapma artıkları,

$$rsd_i = \frac{rd_i}{\sqrt{1 - h_{ii}}} \quad (52)$$

şeklinde elde edilir.

1.5.2. Yüksek Kaldıraç Noktası Belirleme Yöntemleri

Yüksek kaldıraç noktası, veri kümesinde diğer gözlem değerlerinden uzakta bulunan gözlem olarak adlandırılır. İyi ya da kötü kaldıraç noktaları yüksek kaldıraç noktası olabilir. Fakat kötü kaldıraç noktaları yanlış sınıflandırılmış gözlemler ve de x yönündeki diğer gözlem noktalarına uyumsuzluk gösteren aykırı değerlerdir [69]. Lojistik regresyon modeli için yüksek kaldıraç noktası belirleme yöntemi aşağıda verilmiştir.

1.5.2.1. DEVC ve RobDEVC Teşhis Yöntemleri

Norazan, Sanizah ve Habshah [71] lojistik regresyon modelinde kötü kaldıraç noktalarının belirlenmesi için robust sapma belirleyici (RobDEVC) ölçüsünü önermişlerdir. Robust sapma belirleyicileri sapma artıkları (rd_i) belirleyicisine bağlı olarak geliştirilmiştir.

Eşitlik (2)'de verilen lojistik regresyon modeli de β parametrelerinin tahmini için MLE fonksiyonu Eşitlik (5)'te gösterilmiştir.

Lojistik regresyon modelinde sapma olarak adlandırılan DEV genel modelin uyumunu değerlendirmek için kullanılır. Sapma istatistiği, sapma artıklarının (rd_i) kareleri alınarak,

$$DEV = \sum_{i=1}^n (rd_i)^2 = \sum_{i=1}^n 2 \left[y_i \log \left(\frac{y_i}{\hat{\pi}_i} \right) + (1 - y_i) \log \left(\frac{1 - y_i}{1 - \hat{\pi}_i} \right) \right] \quad (53)$$

şeklinde ifade edilmektedir. i 'inci gözlem için sapma artığı,

$$DR_i = \sum_{i=1}^n rd_i = \pm \sqrt{-2 \left[y_i \log \left(\frac{y_i}{\hat{\pi}_i} \right) + (1 - y_i) \log \left(\frac{1 - y_i}{1 - \hat{\pi}_i} \right) \right]} \quad (54)$$

şeklinde ifade edilir. Burada $y_i = 0$ için $DR_i = -\sqrt{-2|\log(1 - \hat{\pi}_i)|}$ ve $y_i = 1$ için $DR_i = \sqrt{-2|\log \hat{\pi}_i|}$ elde edilir. i 'inci gözlemin artık değeri mutlak değerce 2'den büyük ise aykırı değer olduğu söylenir. Eşitlik (54)'te ifade edilen i 'inci sapma artığının bir bileşeni,

$$DEVC = dc_i = \begin{cases} 2 \log \left(\frac{1}{1 - \hat{\pi}_i} \right) & , \quad y_i = 0 \\ 2 \log \left(\frac{1}{\hat{\pi}_i} \right) & , \quad y_i = 1 \end{cases} \quad (55)$$

şeklinde ifade edilir. Burada dc_i değerleri negatif değerler almayan ve aykırı değerlerin belirlenmesi için kullanışlı belirleyicilerdir. DEVC'nin dağılımı bilinmediğinden DEVC'ye dayalı aykırı değerlerin belirlenmesi için bir eşik değeri önerilmiştir. Bu eşik değeri,

$$dc_i > Medyan(dc_i) + 3MAD(dc_i) \quad (56)$$

biçiminde belirlenmiştir [72]. Literatürde mevcut olan Eşitlik (56)'da verilen eşik değeri için medyan mutlak sapma (MAD) değeri, $MAD(dc_i) = medyan\{|dc_i - medyan(dc_i)|\}/0.6745$ olarak elde edilir.

DEVC belirleyici ölçüsü için β parametre tahminleri, MLE yöntemi kullanılarak Eşitlik (7)'de elde edilmiştir. MLE yöntemi ile elde edilen tahmin edicilerin aykırı değerlere karşı dirençli olmadığı bilinmektedir. Bu düşünceden yola çıkarak, DEVC belirleyici yöntemi geliştirilerek robust sapma bileşenine dayalı olarak RobDEVC belirleyici yöntemi geliştirilmiştir [69]. RobDEVC yöntemi için ilk olarak Mallows tahmin edicisi kullanılarak robust bir başlangıç tahmini $\hat{\beta}_{Mallows}$ elde edilmektedir. $\hat{\beta}_{Mallows}$ parametre tahminleri kullanılarak olasılık değerleri,

$$\hat{\pi}_{i(Mallows)}(x_i) = \frac{\exp(x_i' \hat{\beta}_{Mallows})}{1 + \exp(x_i' \hat{\beta}_{Mallows})} \quad (57)$$

biçiminde hesaplanır. Robust başlangıç değerleri $\hat{\beta}_{Mallows}$ kullanılarak elde edilen RobDEVC belirleyicisi,

$$RobDEVC = rdc_i = \begin{cases} 2\log\left(\frac{1}{1 - \hat{\pi}_{i(Mallows)}}\right) & , \quad y_i = 0 \\ 2\log\left(\frac{1}{\hat{\pi}_{i(Mallows)}}\right) & , \quad y_i = 1 \end{cases} \quad (58)$$

şeklinde ifade edilir. RobDEVC belirleyici ölçüsü için eşik değeri, DEVC teşhis yönteminde kullanılan eşik değeri formülasyonu ile hesaplanmaktadır ve eşik değeri,

$$rdc_i > Medyan(rdc_i) + 3MAD(rdc_i) \quad (59)$$

şeklinde önerilmiştir. Eşitlik (59)'u sağlayan gözlemler kötü kaldıraç noktaları olarak ifade edilmiştir [69].

Bu çalışmada RobDEVC teşhis edici yönteminde β parametrelerin tahmininde kullanılan $\hat{\beta}_{Mallows}$ tahmin edicisine alternatif olarak robust tahmin ediciler kullanarak yeni sonuçlar elde etmeye çalışılmıştır. Kullanılan bu robust tahmin ediciler bir sonraki bölümde tanıtılmıştır. Ayrıca DEVC ve RobDEVC yöntemleri için literatürde kullanılan eşik değerler Eşitlik (56) ve Eşitlik (59)'da verilmiştir. Yapılan bu çalışmada bu eşik değerlere alternatif robust eşik değerler önerilerek kötü kaldıraç noktalarını daha iyi bir şekilde tespit etmek amaçlanmıştır. Eşik değer parametreleri göz önüne alındığında konum ve ölçek parametresi için literatürde yaygın kullanılan tahmin ediciler göz önünde bulundurulmuştur. Oluşturulan alternatif eşik değerler için konum ve ölçeğin robust tahmin edicileri ilerleyen kısımda kısaca tanıtılmıştır.

1.6. Lojistik Regresyon Modelinde Robust Tahmin Ediciler

Robust kelimesi istatistik literatürüne Box [35] tarafından kazandırılmıştır. Box, robust kelimesini varsayımlardaki sapmalara karşı önerilen istatistiksel yöntemler için kullanmıştır. Robust yöntemler, klasik yöntemlere kıyasla büyük hatalara karşı koruma sağlamaktadır. Elde edilen tahmin edicilerin; verideki küçük bozulmalara küçük tepkiler vermesi, büyük bozulmalarda güvenilir sonuçlar vermesi, verideki aykırı değerleri ayırabilmesi ve olabildiğince küçük varyansa sahip olması istenen özellikleridir.

Robust istatistik, istatistiksel yöntemlerin sağlamlık teorisidir. OLS, regresyon modeli normal dağılıma sahipken iyi istatistiksel özelliklere sahip tahminler vermektedir. OLS yöntemi kullanılarak elde edilen tahmin edicilerin etkin olması ve modelin çıkarsama amaçlı kullanılabilmesi için gerekli olan modelin varsayımlarının sağlanması gerekir. Fakat normallikten sapmalar olduğu durumlarda OLS yöntemi ile yapılan tahminlerin etkinliği azaldığından böyle durumlarda alternatif olarak robust yöntemler önerilmiştir.

Robust yöntemlerin önerildiği diğer bir durum ise veri setinde şüpheli gözlemlerin var olmasıdır. Bu gözlemler aykırı değer, etkili gözlem ve yüksek kaldıraç noktaları olarak ortaya çıkabilir. Bu durum regresyon modelinde OLS yöntemi ile yapılan analizlerde sıkça karşılaşılan önemli sorunlardan biridir. Veri kümesinde şüpheli gözlemler varsa, bu gözlem ya da gözlemlerin tahmin üzerinde etkilerini belirlemek için teşhis yöntemleri kullanılmaktadır. Alternatif olarak önerilen robust yöntemler ise bu gözlemlere karşı duyarlı tahminler elde etmektedir.

Lojistik regresyon modelinde aykırı değer probleminin çözümü için iki farklı yaklaşım bulunmaktadır. İlk yaklaşım olabilirlik skor fonksiyonlarının ağırlıklandırılmasına dayanan Mallows sınıfı tahmin edicilerin kullanılması, ikinci yaklaşım ise farklı amaç fonksiyonu kullanılarak tahmin edicilerin elde edilmesidir.

Robust yöntemlerden elde edilen bir tahmin edicinin sağlamlığına karar vermede bozulma noktası ve etkinlik fonksiyonunun önemi büyüktür. Bozulma noktası tahmin edicinin etkilenmediği aykırı değer miktarını göstermektedir. İyi bir robust tahmin edicinin yüksek bozulma noktasına sahip olması gerektiği ve en yüksek değer olarak da %50 bozulma noktasına sahip olması gerektiği söylenmektedir [36]. Etkinlik fonksiyonu, bir gözlemin tahmin edicisi veya test istatistiği üzerindeki göreceli etkisini değerlendirir [73]. Bu fonksiyon robustluk ölçülerinin elde edilmesinde kullanılmaktadır.

Lojistik regresyon modelinde β parametreleri için tahminler MLE yöntemi ile yapılmaktadır. Veri setinde aykırı değer olduğu durumda MLE yerine aykırı değere daha dirençli olduğu bilinen robust yöntemler kullanılabilir. Bu durumda β parametrelerinin iyi bir tahmininde robust yöntem kullanılması ile güvenilir $\hat{\pi}_i$ tahmini olasılıkları elde edilebilir. Bu çalışmada lojistik regresyon modelin de x yönündeki aykırı değerlerin etkilerini azaltmak için GM, BY ve WBY tahmin edicileri β parametrelerinin tahminlerinde kullanılmıştır.

1.6.1. Genelleştirilmiş M Tahmin Edicisi

M tahmin edicisi y yönünde aykırı değerlere karşı dirençli tahmin edicidir. M tahmin edicisi x yönündeki aykırı değerlere karşı dirençli olmadığından GM tahmin edicisi önerilmiştir [73]. GM tahmin edicisi bağımsız değişkenlerdeki her bir gözleme ağırlık vererek x yönündeki aykırı değerlerin etkisini sınırlamaktadır. GM tahmin edicisi,

$$\min \sum_{i=1}^n \sigma^2 \rho \left(\frac{\varepsilon_i}{\sigma w_i^a} \right) w_i^{a+1} \quad (60)$$

şeklinde ifade edilir. Eşitlik (60) amaç fonksiyonundaki w_i , gözlemlere bağlı ağırlıkları göstermektedir.

Aykırı değer probleminde önerilen robust tahmin ediciler lojistik regresyon modelinde yapılarına göre iki gruba ayrılmaktadır: MLE tahmininde gözlemleri ağırlıklandırmaya dayanan tahmin ediciler ve farklı amaç fonksiyonu kullanmaya dayanan tahmin ediciler.

Gözlemleri ağırlıklandırmaya dayanan tahmin ediciler, GM tahmin edicisinin özel durumları olan Mallows ve CUBIF tahmin edicileridir.

1.6.1.1. Mallows Tahmin Edicisi

Genelleştirilmiş M tahmin edicilerinden biri olan ve gözlemleri ağırlıklandırmaya dayanan Mallows tahmin edicisi,

$$\sum_{i=1}^n w_i x_i \left[y_i F(x_i \underline{\beta}) - c(x_i \underline{\beta}) \right] = 0 \quad (61)$$

şeklinde ifade edilmektedir [74]. Eşitlik (61)'de w_i gözlemlere bağlı ağırlıkları ve $c(x_i \underline{\beta})$ düzeltme terimini göstermektedir. Burada $w_i = w(x_i, x_i \underline{\beta})$ ve $c(x_i, \underline{\beta}) = 0$ olarak alınır. Robust Mahalanobis uzaklığı (RMD) kullanılarak gözlemler ağırlıklandırılır. Ağırlık w_i , yüksek kaldıraçlı gözlemin, düşük kaldıraçlı gözlemde daha az ağırlık almasını sağlamaktadır.

1.6.1.2. CUBIF Tahmin Edicisi

CUBIF tahmin edicisi, lojistik regresyon modelinde MLE tahmin edicisini daha güçlü hale getirmek için mevcut olan bir diğer alternatif yaklaşım olarak bilinmektedir. Künsch vd. [75] tarafından önerilen CUBIF tahmin edicisi aşağıdaki eşitlik yardımıyla elde edilebilir.

$$\sum_{i=1}^n w_i x_i \left[y_i F(x_i \underline{\beta}) - c(x_i \underline{\beta}) \right] = 0 \quad (62)$$

Bağımlı değişkenin etkisini yansıtarak oluşturulan ağırlıklar $w_i = w(x_i, x_i \underline{\beta}, y_i)$ şeklinde ifade edilir. Model etkinliği yüksek olan CUBIF tahmin edicisinin kullanıldığı durumlarda aykırı değer oranı az olduğunda aykırı değerlere karşı duyarlı bir tahmin edicidir.

1.6.2. Bianco ve Yohai Tahmin Edicisi

Pregibon [8] tarafından önerilen BY tahmin edicisi, farklı amaç fonksiyonu kullanarak robust tahmin edici elde etmeyi amaçlamaktadır. Önerilen bu robust tahmin edici,

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \lambda [d(x_i \underline{\beta}, y_i)] \quad (63)$$

şeklinde elde edilir. Burada λ artan bir Huber fonksiyonu ve $d(x_i \underline{\beta}, y_i)$ sapma bileşenleri fonksiyonudur. Eşitlik (63)'teki amaç fonksiyonu ile elde edilen sonuçların etkili gözlemlere karşı tutarlı olmadıkları görülmüş ve Pregibon'un önerisinden yola çıkılarak 1996 yılında Bianco ve Yohai, BY tahmin edicisini,

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \rho [d(x_i \underline{\beta}, y_i) + G(F(x_i \underline{\beta})) + G(1 - F(x_i \underline{\beta}))] \quad (64)$$

şeklinde geliştirmişlerdir. Burada

$$G(t) = \int_0^t \rho'(-\ln u) du. \quad (65)$$

biçiminde ifade edilir ve ρ fonksiyonu,

$$\rho(t) = \begin{cases} t - \left(\frac{t^2}{2c}\right), & t \leq c \\ \frac{kc}{2}, & \text{diğer yerlerde} \end{cases} \quad (66)$$

şeklinde tanımlanmaktadır ve c pozitif bir sayıdır.

1.6.3. Ağırlıklandırılmış Bianco ve Yohai Tahmin Edicisi

Croux ve Haesbroeck [41] yüksek kaldıraç noktalarının etkisini düşürmek için BY tahmin edicisine yeniden ağırlık ekleyerek WBY tahmin edicisini önermişlerdir. Farklı amaç fonksiyonu yaklaşımı kullanılarak elde edilen WBY tahmin edicisi,

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n w(x_i) \left\{ \rho \left[d(x_i \underline{\beta}, y_i) + G \left(F(x_i \underline{\beta}) \right) + G \left(1 - F(x_i \underline{\beta}) \right) \right] \right\} \quad (67)$$

biçiminde ifade edilmektedir. Burada ağırlıklar $w(x_i)$, RMD'nin azalan bir fonksiyonu olması için Minimum Kovaryans Belirleyicisi (MCD) tahmin edicisi kullanılarak uzaklıklar

$$\rho(t) = \begin{cases} 1, & RMD_i^2 \leq \chi_{(p,0.975)}^2 \\ 0, & \text{diğer yerlerde} \end{cases} \quad (68)$$

şeklinde hesaplanmaktadır [75]. WBY tahmin edicisinde ağırlık sadece x değişkenlerinde kullanıldığında tutarlı kalmaktadır. Ancak ağırlıklar iyi kaldıraç noktalarının ağırlığını da azaltmaktadır [76].

Açıklanan robust yöntemler kullanılarak DEVC ve RobDEVC yöntemlerine alternatif robust teşhis yöntemleri ve bu teşhis yöntemleri için robust eşik değerler bir sonraki bölümde tanıtılmıştır.

1.6.4. Robust Lojistik Tahmin Edicisi

Lojistik regresyon modelinde robust tahmin ediciler Pregibon'un çalışmasıyla literatüre kazandırılmıştır. Pregibon parametre tahminlerinin aykırı değer varlığından ciddi şekilde etkilenebileceğini belirtmiş ve bu sebeple daha az etkilenen robust tahmin ediciler önermiştir. Literatüre bakıldığında da Pregibon [8], Copas [77], Kunsch, et al., [78], Carroll ve Pederson [77], Bianco ve Yohai [79], Croux ve Haesbroeck [41] tarafından lojistik regresyon modeli için robust tahmin ediciler önerilmiştir.

Lojistik regresyon modeli için önerilen ve yapılan çalışmada incelenen robust lojistik tahmin edicilerin (RLE) performans karşılaştırmaları için genel olarak aşağıdaki denklem kullanılmıştır.

$$\hat{\beta}_{RLE} = (X' \hat{W}^* X)^{-1} X' \hat{W}^* \hat{Z}^* = (M^*)^{-1} v^* \quad (69)$$

Burada $M^* = X' \hat{W}^* X$ ve $v^* = X' \hat{W}^* \hat{Z}^*$ şeklinde ifade edilir. Ayrıca ağırlık matrisi

$\hat{W}^* = \text{diag} \left(\hat{\pi}_{i(rob)}^* (1 - \hat{\pi}_{i(rob)}^*) \right)$ şeklinde elde edilir. $\hat{\pi}_{i(rob)}^*(x_i) = \frac{\exp(x_i' \hat{\beta}_{rob})}{1 + \exp(x_i' \hat{\beta}_{rob})}$ ve

$\hat{Z}_i^* = \text{lojit}(\hat{\pi}_{i(rob)}^*) + \frac{y_i - \hat{\pi}_{i(rob)}^*}{\hat{\pi}_{i(rob)}^* (1 - \hat{\pi}_{i(rob)}^*)}$ biçiminde gösterilir. Burada $\hat{\beta}_{rob}$ tahmin edicisi

$\hat{\beta}_{CUBIF}$, $\hat{\beta}_{Mallows}$, $\hat{\beta}_{BY}$ ve $\hat{\beta}_{WBY}$ tahmin edicilerini ifade etmektedir. $\hat{\beta}_{RLE}$ tahmin edicisinin MMSE ve MSE fonksiyonları,

$$MMSE(\hat{\beta}_{RLE}) = (M^*)^{-1} \quad (70)$$

$$MSE(\hat{\beta}_{RLE}) = tr(MMSE) = \sum_{j=1}^{p+1} \frac{1}{\lambda_j^*} \quad (71)$$

şeklinde elde edilir.

2. YAPILAN ÇALIŞMALAR

2.1. Lojistik Regresyon Modelinde Robust Teşhis Yöntemleri ve Robust Eşik Değerler

Regresyon teşhis yöntemleri, olağan dışı gözlemleri (aykırı değer, yüksek kaldıraç noktası, etkili gözlem) belirlemede yaygın olarak kullanılmaktadır. Bu kısımda birinci bölümde bahsedilen ve literatürde mevcut olan regresyon teşhis yöntemlerine alternatif olarak robust teşhis yöntemler ve bu yöntemler için robust eşik değerler önerilmiştir.

2.1.1. RobDEVC Teşhis Yöntemleri

Yüksek kaldıraç noktalarını belirlemede kullanılan DEVC yöntemi Eşitlik (55)'te verilmiştir. Norazan vd. [71], DEVC teşhis yönteminden hareketle β parametrelerinin tahmininde robust yöntem kullanılmasının kötü kaldıraç noktalarının belirlenmesinde güvenilir tahminlere olanak tanıyacağını ileri sürmüşlerdir. Bu durumda DEVC yönteminde GM tahmin edicilerinden Mallows tahmin ediciyi kullanarak ($\hat{\beta}_{Mallows}$) Eşitlik (58)'de verilen RobDEVC teşhis ediciyi elde etmişlerdir.

Bu bölümde tez çalışmasında ele alınan robust tahmin edicileri β parametrelerinin tahminlerinde kullanarak yeni teşhis yöntemler elde etmek amaçlanmıştır. β parametrelerinin robust tahminlerinde GM tahmin edicilerinden CUBIF tahmin edici ($\hat{\beta}_{CUBIF}$) kullanılarak RobDEVC₁ teşhis yöntemi, BY tahmin edicisi ($\hat{\beta}_{BY}$) kullanılarak RobDEVC₂ teşhis yöntemi ve WBY tahmin edicisi ($\hat{\beta}_{WBY}$) kullanılarak RobDEVC₃ teşhis yöntemi genel bir ifadeyle aşağıdaki şekilde önerilmektedir.

$$RobDEVC^* = rdc_i^* = \begin{cases} 2\log\left(\frac{1}{1 - \hat{\pi}_{i(rob)}^*}\right) & , \quad y_i = 0 \\ 2\log\left(\frac{1}{\hat{\pi}_{i(rob)}^*}\right) & , \quad y_i = 1 \end{cases} \quad (72)$$

β parametrelerinin robust tahmini için Mallows tahmin edicisine alternatif olarak ele alınan ve yukarıda açıklanan CUBIF, BY ve WBY tahmin ediciler yardımıyla Eşitlik (57)'de

verilen tahmin olasılıkları hesaplanır. Daha sonra Eşitlik (72) yardımıyla robust tahmin edicilerinin kullanıldığı robust sapma artıklar; $RobDEVC_1$, $RobDEVC_2$ ve $RobDEVC_3$ elde edilir. Bu robust sapma artık değerlerin geleneksel olarak *Medyan* + $3MAD$ eşik değerini aşması durumunda ilgili gözlemler kötü kaldıraç noktası olarak adlandırılır. Elde edilen robust teşhis yöntemleri için literatürde kullanılan *Medyan* + $3MAD$ eşik değerinde iyileştirmeye gidilerek yeni robust eşik değerler önerilmiştir.

2.1.2. Robust Eşik Değerler

Birçok teşhis yönteminde olduğu gibi klasik yöntemlerden DEVC ve RobDEVC yöntemlerinde de elde edilen değerlerin kötü kaldıraç noktası olduğunu belirlemek için *Medyan* + $3MAD$ eşik değeri kullanılmaktadır. Teşhis yöntemlerinin incelendiği birçok çalışmada artıkların iyileştirildiği çalışmalar mevcut olmasına rağmen, bu çalışmalarda eşik değerlerin aynı şekilde hesaplandığı görülmektedir. Olağan dışı gözlemlerin teşhisinde doğru eşik değerin kullanılması, teşhis yöntemlerinin performansları üzerinde etkili olacaktır. Bu nedenle tez çalışmasında *Medyan* + $3MAD$ eşik değerine alternatif robust eşik değerler önerilmiştir.

Eşik değerlerin tespitinde artıklar için ortalama ve değişkenlik ölçüsü kullanılmaktadır. Klasik eşik değeri olan *Medyan* + $3MAD$ eşik değerinde ortalama olarak aykırı değerlerden etkilenmeyen medyan ve değişkenlik ölçüsü olarak MAD ele alınmıştır. Bu eşik değere alternatif konum ve ölçek tahmin edicileri aşağıda kısaca açıklanmıştır.

Bu çalışmada medyan tahmin edicisine alternatif olarak yedi farklı konum tahmin edicisinin kullanılması önerilmiştir. Bu konum tahmin edicileri M, L, R robust tahmin ediciler sınıfından seçilmiştir. Ölçek tahmin edicisi MAD için ise alternatif olarak Q_n ve S_n tahmin edicileri kullanılmıştır.

2.1.2.1. Konumun Robust Tahmin Edicileri

Lojistik regresyon modelinde teşhis yöntemleriyle elde edilen sapmaların kötü kaldıraç noktası olup olmadığını belirlemek için *Medyan* + $3MAD$ eşik değeri kullanılmaktadır. Medyan tahmin edicisine alternatif olarak yedi farklı konum tahmin edicisinin kullanılması önerilmiştir. Bu konum tahmin edicileri M, L, R robust tahmin ediciler sınıfından seçilmiştir. M tahmin ediciler en çok olabilirlik tipi tahmin ediciler, L

tahmin ediciler sıra istatistiklerinin doğrusal bir fonksiyonu ve R tahmin ediciler ise sıralanmış örneklem gözlemlerinin ranklarından oluşan tahmin edicilerdir [13]. Ayrıca çalışmada kullanılan M tahmin edicisi Huber, Andrews, Tukey Bisquare, Hampel; L tahmin edicisi Medyan, Trimmed, Winsorize ve R tahmin edicisi ise Hodges Lehmann tahmin edicileridir. Tez çalışmasında kullanılan konum ve ölçek parametreleri aşağıda kısaca açıklanmıştır.

M Tahmin Edicileri: Huber [36] tarafından önerilen M tahmin edicisi, en çok olabilirlik tipinde robust tahmin edicisidir. Bilinmeyen parametre θ 'nın M tahmin edicisi,

$$\min_{\theta} \sum_{i=1}^n \rho(X_i, \theta) \quad (73)$$

biçiminde tanımlanmaktadır. Burada ρ amaç fonksiyonudur ve aşağıdaki özelliklere sahiptir.

1. $\rho(0) = 0$
2. $0 < x_1 < x_2$ için $\rho(x_1) < \rho(x_2)$.

$f(x, \theta)$ yoğunluk fonksiyonu olmak üzere $\rho = -\log(f)$ olarak seçilirse M tahmin edicisi,

$$\min_{\theta} \sum_{i=1}^n \rho(-\log f(x_i, \theta)) \quad (74)$$

biçiminde elde edilir ve MLE tahmin edicisini verir. Amaç fonksiyonu $\rho(x) = x^2$ şeklinde seçildiğinde ise OLS tahmin edicisi elde edilir. Eşitlik (73)'teki ifadenin θ parametresine göre türevi alındığında,

$$\sum_{i=1}^n \psi(x_i, \theta) \quad (75)$$

M tahmin edicisi elde edilir. Burada $\psi(x) = \frac{\partial}{\partial \theta} \rho(x)$ şeklinde hesaplanır ve etki fonksiyonu olarak adlandırılmaktadır. Etki fonksiyonu sürekli, simetrik ve parçalanabilir bir fonksiyondur.

Konum parametresi için ise M tahmin edicisi;

$$x_i = \theta + u_i \quad (76)$$

ifadesi yardımıyla hesaplanabilir. Burada $x_1, x_2, \dots, x_n$ n birimlik rastgele örneklem, θ ana kitle ortalaması ve u_i hata terimleri olarak ifade edilmektedir. Bu konum modelinden hareketle θ konum parametrelili ana kitle ortalamasının M tahmin edicisi,

$$\hat{\theta}_n = \arg \min_{\theta} \sum_{i=1}^n \rho(x_i - \theta) \quad (77)$$

şeklinde ifade edilmektedir [80]. Eşitlik (77)'de verilen denklemin analitik çözümü zordur. Bu ifadenin θ 'ya göre türevi alınarak sıfıra eşitlendiğinde ana kitle ortalamasının M tahmin edicisi Eşitlik (78)'deki gibi elde edilir [81].

$$\sum_{i=1}^n \psi(x_i - \theta) = 0 \quad (78)$$

M tahmin edicileri ψ fonksiyonun yapısına göre monoton M tahmin edicileri ve yeniden azalan M tahmin edicileri olarak adlandırılmaktadır [38]. Orijinden itibaren belirlenen bir noktaya kadar doğrusal ya da azalarak artan, bu noktadan sonra ise sabit ya da çok az artışlarla ağırlık veren tahmin edicilere monoton M tahmin edicileri denir. Yeniden azalan M tahmin edicilerinin ψ fonksiyonu, belirli bir noktadan sonra çok küçük ağırlık veriyorsa (sıfıra yakın) yavaş yeniden azalan; belirli bir noktadan sonra sıfır ağırlık veriyorsa hızlı yeniden azalan M tahmin edicileri olarak adlandırılmaktadır. Yeniden azalan M tahmin edicileri aykırı değerlere karşı daha dayanıklıdır ve aykırı değer teşhisinde de kullanılmaktadır [82]. M tahmin ediciler için bazı ψ fonksiyonları Tablo 1'de verilmiştir.

Tablo 1. M tahmin ediciler için ψ fonksiyonları

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

ψ fonksiyonu, ağırlık fonksiyonu w yardımıyla gözlemdeki her bir verinin nasıl ağırlıklandırılacağını kontrol etmektedir. Ağırlık fonksiyonu,

$$w(x) = \begin{cases} \frac{\psi(x)}{x} & , x \neq 0 \\ \psi'(0) & , x = 0 \end{cases} \quad (79)$$

şeklinde ifade edilmektedir. Eşitlik (78) yeniden yazılacak olursa,

$$\sum_{i=1}^n w(x_i - \theta)(x_i - \theta) = 0 \quad (80)$$

şekline dönüşmektedir. Konumun M tahmin edicisi,

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (81)$$

şeklinde ifade edilmektedir. Burada $w_i = w(x_i - \theta)$ şeklindedir.

M tahmin edicilerinin hesaplanması kolay olduğundan dolayı yaygın kullanıma sahip robust yöntemlerdir. Fakat $\rho(\cdot)$ 'nin seçimi için bir koşul bulunmaması ve x yönündeki aykırı değerlere karşı duyarlı olması dezavantaj olarak görülmektedir.

L Tahmin Edicileri: Sıra istatistiklerine dayalı L tahmin edicileri ilk olarak Daniel [83] tarafından önerilmiştir. $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ sıra istatistiklerini göstermek üzere L tahmin edicisi,

$$T_n(X_1, X_2, \dots, X_n) = \sum_{i=1}^n a_i x_{(i)} \quad (82)$$

sıra istatistiklerinin doğrusal bir biçimidir. Burada a_i sabitleri gerçel sayılardır ($0 \leq a_i \leq 1; i = 1, \dots, n$) ve $\sum_{i=1}^n a_i = 1$ koşulunu sağlaması gerekmektedir.

Örneklem ortalaması, örneklem medyanı, trimmed ve winsorize ortalamalar L tahmin ediciler arasında yaygın kullanıma sahip tahmin edicilerdir. L tahmin ediciler kolay hesaplanmaktadır [84].

Medyan: L tahmin ediciler arasında en yaygın tahmin edicilerden biri olan medyan, örnek hacmine göre veri setinde sadece bir veya iki gözlemin farklı ağırlıklandırıldığı diğer tüm gözlem değerlerine sıfır ağırlığının verildiği bir tahmindir. Örneklem hacmi n tek sayı olduğunda medyan merkezi sıra istatistiği, n çift sayı olduğunda ise medyan iki merkezi sıra istatistiğinin ortalamasıdır ve medyan değeri,

$$Medyan = \begin{cases} x_{(\frac{n+1}{2})} & , \quad n \text{ tek sayı ise} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2} & , \quad n \text{ çift sayı ise} \end{cases} \quad (83)$$

şeklinde elde edilmektedir.

Trimmed (Budanmış) Ortalama: Veri setinde olağandışı gözlemler olduğu durumda ortalamaya alternatif olarak kullanılan trimmed ortalama, n birimlik örneklemdeki sıra istatistiklerinin her iki ucundan belli oranlarda gözlem değerlerinin atılmasından sonra hesaplanan ortalama olarak hesaplanmaktadır [94].

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

$$T_n = T_n(l_n, u_n) = \frac{1}{u_n - l_n} \sum_{i=l_n+1}^{u_n} x_{(i)} \quad (84)$$

şeklinde ifade edilmektedir. Burada α alt uçtan budanma oranı, β üst uçtan budanma oranı, $l_n = [n\alpha]$ 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 $0 \leq \alpha \leq 0.25$, $1 - \beta \leq 0.25$ olarak alınmaktadır.

Winsorize Ortalama: Winsorize ortalama, trimmed ortalama benzer bir tahmin edicidir ancak trimmed ortalama gibi verinin uç kısımları atılmaz, en küçük ve en büyük değerlere winsorized yüzdesi kadar veri setindeki değerlerin eklemesi yapılmaktadır.

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

$$W_n = W_n(l_n, u_n) = \frac{1}{n} \left\{ l_n x_{(l_n+1)} + \sum_{i=l_n+1}^{u_n} x_{(i)} + (n - u_n) x_{(u_n)} \right\} \quad (85)$$

şeklinde ifade edilmektedir. Burada l_n ve u_n ifadeleri Eşitlik (84)'te verilenlerle aynıdır. Trimmed ortalama olduğu gibi kuyrukların ağırlığına göre $0 \leq \alpha \leq 0.25$, $1 - \beta \leq 0.25$ alınmaktadır.

R Tahmin Edicileri: Hodges ve Lehmann [85] tarafından önerilen R tahminleri, rank testlerinden üretildiğinden R tahmin edicisi olarak adlandırılmaktadır. R tahmin edicileri arasında en yaygın kullanıma sahip tahmin edici Hodges-Lehmann tahmin edicisidir.

Hodges-Lehmann Tahmin Edicisi: Hodges-Lehmann tahmin edicisi Wilcoxon sıra işaret test istatistiğinden elde edilmiştir ve tüm ikili gözlem çiftlerinin ortalamalarının medyanı olarak tanımlanmaktadır.

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ sıra istatistiklerini göstermek üzere, Hodges-Lehmann tahmin edicisi,

$$\hat{\theta}_{HL} = \text{medyan} \left\{ \frac{x_{(i)} + x_{(j)}}{2} \right\}, \quad 1 \leq i < j \leq n \quad (86)$$

şeklinde ifade edilmektedir [85].

2.1.2.2. Ölçek Parametresinin Robust Tahmin Edicileri

Yayılma ölçülerinden biri olan varyans, ana kitle ölçek parametresinin yaygın kullanılan bir tahmin edicisidir. Ancak veri kümesinde aykırı değerlere karşı hassastır. Bu durumda ölçek parametresi için robust tahmin ediciler önerilmektedir. Önerilen bu tahmin edicilerden bazıları; MAD, S_n ve Q_n 'dir. Bu robust ölçek tahmin edicilerinin kırılma noktaları ve etkinlikleri göz önüne alınarak avantaj ve dezavantajları karşılaştırılmıştır.

Medyan Mutlak Sapma Tahmin Edicisi: Hesaplama açısından kolay olması ve etki fonksiyonu içinde dayanıklı olmasından dolayı yaygın olarak kullanılan robust ölçek tahmini MAD,

$$MAD = \frac{\text{medyan}(|x_i - \text{medyan}(x_i)|)}{0.6745} \quad (87)$$

şeklinde hesaplanmaktadır [36]. Paydadaki ölçek sabiti, ölçeğin tahminini normal dağılımın standart sapması ile tutarlı hale getirmek için MAD'ı standardize etmektedir. %50 kırılma noktasına sahip olan MAD yönteminin bazı dezavantajları vardır. Bunlar simetrik dağılımlar için geliştirilmesi ve Gaussian etkinliğinin (%37) diğer robust ölçek tahmin edicilere göre düşük olmasıdır [86].

S_n Tahmin Edicisi: Rousseeuw ve Croux [20] tarafından önerilen alternatif S_n tahmin edicisinin, hem simetrik dağılımlar hem de simetrik olmayan dağılımlar için etkinliği yüksektir. S_n tahmin edicisi,

$$S_n = c_n 1.1926 \text{ medyan}_i \{ \text{medyan}_j |x_i - x_j| \}, \quad i, j = 1, 2, \dots, n \quad (88)$$

biçiminde elde edilmektedir. Burada c_n ve 1.1926 sabitleri S_n tahmin edicisinin yansız olmasını sağlamaktadır. c_n sabitinin değeri ise Tablo 2'de verilmiştir [86].

Tablo 2. $n \leq 9$ için c_n değerleri

n	2	3	4	5	6	7	8	9
c _n	0.743	1.851	0.954	1.351	0.993	1.198	1.005	1.131

Ayrıca $n > 9$ için, n tek ise $c_n = \frac{n}{n-0.9}$ ve n çift ise $c_n = 1$ değerini almaktadır. S_n tahmin edicisi hesaplanırken örneklemdeki değerlerin kendi aralarındaki uzaklıkları hesaba katılmaktadır ve örneklemin konum tahminine ihtiyaç yoktur. Bu durum S_n tahmin edicisinin MAD tahmin edicisinin aksine asimetrik dağılımlar için de geçerli olmasını sağlamaktadır [86]. %50 kırılma noktasına sahip olan S_n tahmin edicisinin etkinliği MAD ölçeğine göre daha yüksektir (%58).

Q_n Tahmin Edicisi: MAD ve S_n tahmin edicilerinden daha yüksek etkinliğe (%82) sahip olan Q_n tahmin edicisi, sürekli bir fonksiyona sahiptir. Q_n tahmin edicisi,

$$Q_n = d_n 2.2219 \{ |x_i - x_j|; i < j \}_{(k)} , \quad i, j = 1, 2, \dots, n \quad (89)$$

şeklinde hesaplanır. Burada $k = \binom{h}{2} \approx \binom{n}{2}/4$ ve h değeri $\left(\frac{n}{2} + 1\right)$ şeklindedir. d_n sabitinin değeri Tablo 3'te verilmiştir.

Tablo 3. $n \leq 9$ için d_n değerleri

n	2	3	4	5	6	7	8	9
d_n	0.399	0.994	0.512	0.844	0.611	0.857	0.669	0.872

Ayrıca $n > 9$ için; n tek olduğu durumda $d_n = \frac{n}{n+1.4}$ ve n çift olduğu durumda ise $d_n = \frac{n}{n+3.8}$ olarak hesaplanmaktadır [86]. Q_n tahmin edicisi S_n tahmin edicisi gibi simetrik olmayan dağılımlar için de kullanılan bir tahmin edicidir. Ancak tahmin edicideki algoritmalar hesaplanma açısından zorluklara neden olabilmektedir [86].

Konum ve ölçek parametrelerinin robust tahmin edicilerinin farklı kombinasyonları ile önerilen yeni eşik değerler Tablo 4'te gösterilmiştir.

Tablo 4. Oluşturulan yeni eşik değerler

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

CT: Eşik Değer CT1-23: Önerilen Eşik Değerler

2.2. Robust Yanlı Tahmin Ediciler

Veri setinde çoklu iç ilişki problemi olduğu durumda tahmin edicilerin minimum varyansa sahip olmadığı bilinmektedir. Yanlı tahmin ediciler kullanılarak yanlı ve daha küçük varyanslı tahmin ediciler elde edilmektedir. Aykırı değer problemi olduğu durumda ise parametre tahminlerinden güvenilir sonuçlar elde edilemeyeceği bilinmektedir. Robust tahmin ediciler kullanılarak elde edilen parametrelerin etkin sonuçlar verdiği görülmüştür. Bu iki problem veri setinde ayrı ayrı olabildiği gibi birlikte de görülmektedir. Veri setinde hem çoklu iç ilişki hem aykırı değer bulunması durumuna yanlı tahmin ediciler çoklu iç ilişki problemi ile başa çıksa da tahmin edicilerin hesaplamasında kullanılan MLE tahmin edicileri etkin olmayan sonuçlar verecektir. Böylece yanlı tahmin ediciler kullanarak çoklu iç ilişki problemini ve robust tahmin ediciler kullanarak aykırı değer problemini çözmeye çalışan, etkin sonuçlar veren robust yanlı tahmin ediciler elde edilmiştir. Lojistik regresyon modelinde bu iki problemi birlikte ele alan RLRE ve bu tez çalışmasında önerilen iki parametrelili robust lojistik ridge tahmin edicisi alt bölümde anlatılmıştır.

2.2.1. Robust Lojistik Ridge Tahmin Edicisi

Arifin ve Midi [48] veri setinde hem yüksek kaldıraç noktalarının hem de çoklu iç ilişkinin olması durumunda lojistik ridge regresyon modelinde WBY tahmin edicisine bağlı RLRE tahmin edicisini aşağıdaki gibi tanımlamışlardır.

$$\hat{\beta}_{RLRE_{WBY}} = (X' \hat{W}_{WBY} X + kI)^{-1} X' \hat{W}_{WBY} \hat{Z}_{WBY} \quad (90)$$

Burada $\hat{W}_{WBY}$ ve $\hat{Z}_{WBY}$ ifadeleri Eşitlik (69)'daki gibi hesaplanır ve Eşitlik (69)'daki $\hat{\beta}_{rob}$ yerine $\hat{\beta}_{WBY}$ tahmin edicisi kullanılmıştır. Ayrıca Eşitlik (90)'da k yanlılık parametresi de $\hat{\beta}_{WBY}$ tahmin edicisi ile elde edilir.

Eşitlik (90) dikkate alınarak bu çalışmada lojistik regresyon modelinde önerilen diğer robust tahmin ediciler [41] için de genel olarak aşağıdaki denklem kullanılmıştır.

$$\hat{\beta}_{RLRE} = (X' \hat{W}^* X + k^* I)^{-1} X' \hat{W}^* \hat{Z}^* = (M^* + k^* I)^{-1} v^* \quad (91)$$

Burada M^* ve v^* ifadeleri Eşitlik (69)'da verildiği gibidir. k^* yanlılık parametresi robust tahmin edicilerine bağlı hesaplanmaktadır.

RLRE tahmin edicisinin yan ve kovaryans fonksiyonları,

$$Bias(\hat{\beta}_{RLRE}) = -k^* Q \Omega_k^{*-1} \alpha^* \quad (92)$$

$$Cov(\hat{\beta}_{RLRE}) = Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q' \quad (93)$$

şeklinde elde edilir. Burada Eşitlik (15)'ten farklı olarak $\alpha^* = Q' \underline{\beta}_{rob}$ şeklinde elde edilir. $\Omega_k^* = diag(\lambda_1^* + k^*, \lambda_2^* + k^*, \dots, \lambda_{p+1}^* + k^*)$ olarak hesaplanır. $\hat{\beta}_{RLRE}$ tahmin edicisinin MMSE ve MSE fonksiyonları ise,

$$\begin{aligned} MMSE(\hat{\beta}_{RLRE}) &= Cov(\hat{\beta}_{RLRE}) + Bias(\hat{\beta}_{RLRE}) Bias(\hat{\beta}_{RLRE})' \\ &= Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q' + (-k^* Q \Omega_k^{*-1} \alpha^*) (-k^* Q \Omega_k^{*-1} \alpha^*)' \end{aligned} \quad (94)$$

$$MSE(\hat{\beta}_{RLRE}) = tr(MMSE) = \sum_{j=1}^{p+1} \frac{\lambda_j^*}{(\lambda_j^* + k^*)^2} + \sum_{j=1}^{p+1} \frac{k^{*2} \alpha_j^{*2}}{(\lambda_j^* + k^*)^2} \quad (95)$$

şeklinde elde edilir.

2.2.2. İki Parametrelili Robust Lojistik Ridge Tahmin Edicisi

Lipovetsky ve Conklin [16] tarafından geliştirilen TPPE, çoklu iç ilişki problemi ile karşılaşıldığı durumda kullanılması önerilen yeni yöntemlerden biridir. TPPE yöntemi yansızlıktan ödün vererek hem tahmin edicinin varyansını azaltmakta hem de modelin genel uyumunu göz önünde bulundurmaktadır. Ertaş [53] tarafından önerilen robust yanlı tahmin edici olan iki parametrelili ridge-M tahmin edicisi doğrusal regresyon modeli için ortaya konulmuştur. Asar [54], Lipovetsky ve Conklin [16] tarafından önerilen TPPE tahmin yöntemini lojistik regresyon modeli için genelleştirip çoklu iç ilişki problemini ele almaya çalışmıştır.

Bu tez çalışmasında literatürde mevcut olmayan lojistik regresyon modeli için veri setinde çoklu iç ilişki ve x yönünde aykırı değer olduğu durumda bu iki problemi birlikte

çözmek için iki parametrelili robust lojistik ridge tahmin edicisi (TPRLRE) tanımlanmıştır. Asar [54] tarafından geliştirilen iki parametrelili lojistik ridge tahmin edicisinden yola çıkarak önerilen tahmin edici,

$$\hat{\beta}_{TPRLRE} = q^* (X' \hat{W}^* X + k^* I)^{-1} X' \hat{W}^* \underline{\hat{Z}}^* \quad (96)$$

veya

$$\hat{\beta}_{TPRLRE} = q^* (M^* + k^* I)^{-1} v^* \quad (97)$$

biçiminde yazılabilir. Burada M^* , $\hat{W}^*$, k^* , $\underline{\hat{Z}}^*$ ve v^* Eşitlik (69)'da verildiği gibi hesaplanır. Ayrıca modelin parametreleri,

$$q^* = \frac{v^{*'} (M^* + k^* I)^{-1} v^*}{v^{*'} (M^* + k^* I)^{-1} M (M^* + k^* I)^{-1} v^*} \quad (98)$$

ve

$$k^* = \frac{\sum_{j=1}^{p+1} (q^* \lambda_j^* + (q^* - 1) \lambda_j^{*2} \alpha_j^{*2})}{\sum_{j=1}^{p+1} \lambda_j^* \alpha_j^{*2}} \quad (99)$$

şeklinde ifade edilir. k yarıllık parametresi Eşitlik (40)'taki gibi hesaplanır. Bu çalışmanın temel problemlerinden biri olan veri setinde aykırı değer olması durumunda, β parametre tahminlerinde MLE tahmin yöntemi yerine aykırı değerlere daha dirençli olduğu bilinen robust tahmin yöntemleri kullanılmıştır. Böylece robust yöntemlerle elde edilen tahminlerin, MLE yöntemi ile elde edilen tahminlerden daha iyi sonuçlar vereceği öngörülmektedir.

TPRLRE için yan ve kovaryans fonksiyonları,

$$\begin{aligned} Bias(\hat{\beta}_{TPRLRE}) &= E \left[q^* (X' \hat{W}^* X + k^* I)^{-1} X' \hat{W}^* \underline{\hat{Z}}^* \right] - \underline{\beta}_{rob} \\ &= q^* (X' \hat{W}^* X + k^* I)^{-1} X' \hat{W}^* E(\underline{\hat{Z}}^*) - \underline{\beta}_{rob} \\ &= q^* (X' \hat{W}^* X + k^* I)^{-1} X' \hat{W}^* X \underline{\beta}_{rob} - \underline{\beta}_{rob} \\ &= \left[q^* (X' \hat{W}^* X + k^* I)^{-1} X' \hat{W}^* X - I \right] \underline{\beta}_{rob} \end{aligned} \quad (100)$$

$$\begin{aligned}
Cov(\hat{\beta}_{TPRLRE}) &= q^{*2} X' \hat{W}^* Var(\hat{\underline{Z}}^*) \hat{W}^{*'} X (X' \hat{W}^* X + k^* I)^{-1} \\
&= q^{*2} (X' \hat{W}^* X + k^* I)^{-1} X' \hat{W}^* \hat{W}^{*-1} \hat{W}^{*'} X (X' \hat{W}^* X + k^* I)^{-1}
\end{aligned} \tag{101}$$

biçiminde elde edilir. MSE hesaplamasında kovaryans ve yan fonksiyonlarının elde edilmesinde kolaylık sağlayacağı düşünülen kanonik form kullanılabilir. Bu yaklaşımda $M^* = Q \Omega^* Q'$, $Q' Q = Q Q' = I$ ve $\alpha^* = Q' \beta_{rob}$ olacak şekilde yan ve kovaryans,

$$\begin{aligned}
Bias(\hat{\beta}_{TPRLRE}) &= [q^*(M^* + k^* I)^{-1} M^* - I] \underline{\beta}_{rob} \\
&= [q^*(Q \Omega^* Q' + k^* I)^{-1} Q \Omega^* Q' - I] \underline{\beta}_{rob} \\
&= [q^* Q (\Omega^* + k^* I)^{-1} Q' Q \Omega^* Q' - I] \underline{\beta}_{rob} \\
&= [q^* Q \Omega_k^{*-1} \Omega^* Q' - I] \underline{\beta}_{rob}
\end{aligned} \tag{102}$$

$$\begin{aligned}
Cov(\hat{\beta}_{TPRLRE}) &= q^{*2} (M^* + k^* I)^{-1} M^* (M^* + k^* I)^{-1} \\
&= q^{*2} (Q \Omega^* Q' + k^* I)^{-1} Q \Omega^* Q' (Q \Omega^* Q' + k^* I)^{-1} \\
&= q^{*2} Q (\Omega^* + k^* I)^{-1} Q' Q \Omega^* Q' Q (\Omega^* + k^* I)^{-1} Q' \\
&= q^{*2} Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q'
\end{aligned} \tag{103}$$

şeklinde hesaplanır.

TPRLRE için MMSE ve MSE fonksiyonları aşağıdaki gibi hesaplanır.

$$\begin{aligned}
MMSE(\hat{\beta}_{TPRLRE}) &= Cov(\hat{\beta}_{TPRLRE}) + Bias(\hat{\beta}_{TPRLRE}) Bias(\hat{\beta}_{TPRLRE})' \\
&= q^{*2} Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q' + \\
&\quad + [q^* Q \Omega_k^{*-1} \Omega^* Q' - I] \underline{\beta}_{rob} \underline{\beta}_{rob}' [q^* Q \Omega_k^{*-1} \Omega^* Q' - I]
\end{aligned} \tag{104}$$

$$\begin{aligned}
MSE(\hat{\beta}_{TPRLRE}) &= tr(MMSE(\hat{\beta}_{TPRLRE})) \\
&= \sum_{j=1}^{p+1} \left(\frac{q^{*2} \lambda_j^*}{(\lambda_j^* + k^*)^2} + \frac{\alpha_j^{*2} (q^* \lambda_j^* - \lambda_j^* - k^*)^2}{(\lambda_j^* + k^*)^2} \right)
\end{aligned} \tag{105}$$

Bu bölümde lojistik regresyon modeli için β parametrelerinin robust tahminlerinde sıkça kullanılan ve birinci bölümde tanıtılan; Mallows, CUBIF, BY ve WBY tahmin

yöntemleri ele alınmıştır. Lojistik regresyon, lojistik ridge regresyon ve iki parametrelili ridge regresyon modellerinde MLE ve robust tahmin ediciler kullanılarak bu yöntemlerin simülasyon çalışmasıyla MSE kriterine göre karşılaştırılması da yapılmıştır.

2.2.3. Tahmin Edicilerin Karşılaştırılması

Ele alınan yöntemlerin teorik karşılaştırılması için bazı yardımcı teoremlerin verilmesi gerekmektedir. $\hat{\beta}_1$ ve $\hat{\beta}_2$, β 'nın iki farklı tahmin edicisi olmak üzere $\hat{\beta}_2$ tahmin edicisini $\hat{\beta}_1$ tahmin edicisinden üstünlüğü ancak ve ancak $MMSE(\hat{\beta}_1) - MMSE(\hat{\beta}_2) \geq 0$ koşulu sağlandığında söylenebilir. Tahmin edicilerin teorik karşılaştırmaları için Farebrother [87] ve Trenkler ve Toutenberg [88] teoremleri referans olarak kullanılmıştır.

Yardımcı Teorem 1 (Farebrother, 1976): A pozitif tanımlı bir matris, α sıfır olmayan sabitlerin bir vektörü ve d pozitif bir sabit olsun. $\alpha' A \alpha < d$ olması durumunda $dA - \alpha \alpha' > 0$ 'dır.

Yardımcı Teorem 2 (Trenkler ve Toutenburg, 1990): $\hat{\beta}_1$ ve $\hat{\beta}_2$, β 'nın iki farklı tahmin edicisi olsun. Ayrıca $S = Var(\hat{\beta}_1) - Var(\hat{\beta}_2)$ pozitif tanımlı bir matris ve $b_1 = Bias(\hat{\beta}_1)$, $b_2 = Bias(\hat{\beta}_2)$ olsun. Ancak ve ancak $b_2'(S - b_1 b_1')^{-1} b_2 < 1$ koşulunda $MMSE(\hat{\beta}_1) - MMSE(\hat{\beta}_2) \geq 0$ 'dır.

2.2.3.1. RLE ve RLRE Tahmin Edicilerinin Karşılaştırılması

RLE ve RLRE tahmin edicilerinin MMSE kriterine göre karşılaştırılması Teorem 1'de verilmiştir.

Teorem 1: $j = 1, 2, \dots, p + 1$ ve $k^* > 0$ için eğer $b_{RLRE}' [\Omega^* - \Omega_k^{*-1} \Omega^* \Omega_k^{*-1}] b_{RLRE} < 1$ ise $MMSE(\hat{\beta}_{RLE}) - MMSE(\hat{\beta}_{RLRE}) > 0$ 'dır.

İspat: $\hat{\beta}_{RLE}$ ve $\hat{\beta}_{RLRE}$ tahmin edicilerin MMSE farkları Eşitlik (70) ve Eşitlik (91) kullanılarak

$$\begin{aligned}
MMSE(\hat{\beta}_{RLRE}) - MMSE(\hat{\beta}_{RLRE}) &= Q\Omega^*Q' - Q\Omega_k^{*-1}\Omega^*\Omega_k^{*-1}Q' - b_{RLRE}b'_{RLRE} \\
&= Q[\Omega^* - \Omega_k^{*-1}\Omega^*\Omega_k^{*-1}]Q' - b_{RLRE}b'_{RLRE} \\
&= Qdiag\left[\frac{1}{\lambda_j^*} - \frac{\lambda_j^*}{(\lambda_j^* + k^*)^2}\right]_{j=1}^{p+1} Q'
\end{aligned} \tag{106}$$

şeklinde elde edilir. $(\lambda_j^* + k^*)^2 - \lambda_j^* > 0$ olduğunda ve $k^* > 0$ için $[\Omega^* - \Omega_k^{*-1}\Omega^*\Omega_k^{*-1}]$ pozitif tanımlıdır ve ispat Yardımcı Teorem 1 kullanılarak tamamlanır.

2.2.3.2. RLRE ve TPRLRE Tahmin Edicilerinin Karşılaştırılması

RLRE ve TPRLRE tahmin edicilerini MMSE kriterine göre karşılaştırmak için gerekli teoremler aşağıda verilmiştir.

Teorem 2: Eğer $b'_{RLRE}Q[(q^{*2} - 1)\Omega_k^{*-1}\Omega^*\Omega_k^{*-1}]^{-1}Q'b_{RLRE} < (q^{*2} - 1)$, $q^* > 1$ ve $k^* > 0$ ise $MMSE(\hat{\beta}_{TPRLRE}) - MMSE(\hat{\beta}_{RLRE}) > 0$ 'dır.

İspat: RLRE ve TPRLRE tahmin edicilerinin MMSE farkları Eşitlik (91) ve Eşitlik (96) yardımıyla aşağıdaki gibi yazılır.

$$\begin{aligned}
MMSE(\hat{\beta}_{TPRLRE}) - MMSE(\hat{\beta}_{RLRE}) &= \Lambda \\
\Lambda &= q^{*2}Q\Omega_k^{*-1}\Omega^*\Omega_k^{*-1}Q'b_{TPRLRE}b'_{TPRLRE} - Q\Omega_k^{*-1}\Omega^*\Omega_k^{*-1}Q' - b_{RLRE}b'_{RLRE} \\
&= Q[q^{*2}\Omega_k^{*-1}\Omega^*\Omega_k^{*-1} - \Omega_k^{*-1}\Omega^*\Omega_k^{*-1}]^{-1}Q' + b_{TPRLRE}b'_{TPRLRE} - b_{RLRE}b'_{RLRE} \\
&= Q[(q^{*2} - 1)\Omega_k^{*-1}\Omega^*\Omega_k^{*-1}]^{-1}Q' + b_{TPRLRE}b'_{TPRLRE} - b_{RLRE}b'_{RLRE}
\end{aligned} \tag{107}$$

Burada $q^* > 1$ olduğu durumda $k^* > 0$ için $(q^{*2} - 1)\Omega_k^{*-1}\Omega^*\Omega_k^{*-1}$ matrisi pozitif tanımlıdır ve $b_{TPRLRE}b'_{TPRLRE}$ de pozitif tanımlıdır. $Q[(q^{*2} - 1)\Omega_k^{*-1}\Omega^*\Omega_k^{*-1}]^{-1}Q' - b_{RLRE}b'_{RLRE}$ matrisinin pozitif tanımlı olduğunu ispatlamak yeterli olacaktır ve ispat Yardımcı Teorem 1 ile yapılır.

Teorem 3: Hesaplanan $q^* > 1$ ve $k^* > 0$ parametreleri için eğer

$$\left(b'_{RLRE} Q [(q^{*2} - 1) \Omega_k^{*-1} \Omega^* \Omega_k^{*-1}]^{-1} Q' b_{TPRLRE} b'_{TPRLRE} \right)^{-1} b_{RLRE} < 1 \quad \text{ise}$$

$MMSE(\hat{\beta}_{TPRLRE}) - MMSE(\hat{\beta}_{RLRE}) > 0$ olur.

İspat: TPRLRE ve RLRE tahmin edicilerinin kovaryans farkları alınırsa,

$$\begin{aligned} Cov(\hat{\beta}_{TPRLRE}) - Cov(\hat{\beta}_{RLRE}) &= q^{*2} Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q' - Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q' \\ &= (q^{*2} - 1) Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q' \end{aligned} \quad (108)$$

$Cov(\hat{\beta}_{TPRLRE}) - Cov(\hat{\beta}_{RLRE}) = (q^{*2} - 1) Q \Omega_k^{*-1} \Omega^* \Omega_k^{*-1} Q'$, $q^* > 1$ ve $k^* > 0$ için pozitif tanımlıdır ve Yardımcı Teorem 2 ile ispat tamamlanmış olur.

3. BULGULAR VE İRDELEME

3.1. Simülasyon Çalışması I

Bu bölüm de Monte Carlo simülasyon çalışması kullanılarak, lojistik regresyon modelinde kötü kaldıraç noktalarının belirlenmesinde var olan yöntemlerle önerilen robust teşhis edici yöntemler karşılaştırılmıştır. Ayrıca literatürde mevcut olan *Medyan + 3MAD* eşik değeri için önerilen robust eşik değerlerin performansları da incelenmiştir. Yeni robust eşik değerlerinin kullanılmasıyla kötü kaldıraç noktalarının ne kadarının doğru bir şekilde belirlendiğinin tespiti için doğru belirleme oranı (CIR) ve aslında iyi olan gözlemin 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 oranını ve SR; şüpheli olarak tanımlanan iyi gözlem sayısının, toplam iyi gözlem sayısına oranını göstermektedir.

Simülasyon düzeneği için bu doktora tez çalışmasında bağımsız değişken sayısı $p = (2, 3)$ olarak ele alınmıştır ve örneklem hacmi $n = (80, 120)$ olarak belirlenmiştir. Bağımlı değişken,

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

şeklinde oluşturulmuştur. Burada ε lojistik dağılımdan elde edilen hata terimlerini göstermektedir. Veri seti $\%(\alpha/2)$ iyi ve $\%(\alpha/2)$ kötü kaldıraç noktası olacak şekilde oluşturulmuş ve α yüksek kaldıraç noktalarının farklı yüzdeliklerini ifade etmektedir. Simülasyon düzeneği için $\alpha = (4, 6, 10)$ olarak belirlenmiştir. Veri setine eklenen yüksek kaldıraç noktaları düzgün dağılımdan $[2, 4]$ üretilmiştir. Eklenen bu yüksek kaldıraç noktalarına karşılık gelen y değerleri ise; iyi kaldıraç noktaları için $y = 1$, kötü kaldıraç noktaları için $y = 0$ şeklinde düzenlenmiştir. R programlama dili kullanılarak yapılan simülasyon çalışmasında, farklı değişken sayısı, örneklem hacmi ve kaldıraç noktası değerlerinin tüm kombinasyonları için 10.000 deneme gerçekleştirilmiştir ve sonuçlar Tablo 5-16'da verilmiştir.

Tablo 5. $p = 2$, $n = 80$ ve kötü kaldıraç noktası 2 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,9483	0,0269	0,9701	0,0327	0,9798	0,0389	0,9848	0,0577	0,9850	0,0577
	Huber	0,8470	0,0031	0,8944	0,0073	0,9211	0,0117	0,9496	0,0267	0,9498	0,0268
	Humpel	0,8976	0,0122	0,9375	0,0180	0,9551	0,0240	0,9706	0,0438	0,9705	0,0438
	Andrew	0,9567	0,0326	0,9768	0,0389	0,9841	0,0449	0,9877	0,0632	0,9878	0,0632
	Winsor	0,8603	0,0043	0,9047	0,0084	0,9302	0,0127	0,956	0,0266	0,9561	0,0266
	Trimmed	0,8546	0,0032	0,8999	0,0073	0,9266	0,0117	0,9533	0,0255	0,9533	0,0256
	Medyan	0,8997	0,0135	0,9363	0,0184	0,9522	0,0238	0,9692	0,0414	0,9693	0,0415
	HodgesLeh	0,8570	0,0040	0,8998	0,0080	0,9268	0,0125	0,9534	0,0265	0,9535	0,0265
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,9488	0,0271	0,9850	0,0341	0,9808	0,0420	0,9851	0,0676	0,9852	0,0675
	Huber	0,8459	0,0032	0,9522	0,0083	0,9254	0,0141	0,9522	0,0349	0,9524	0,0350
	Humpel	0,9024	0,0127	0,9738	0,0196	0,9605	0,0273	0,9738	0,0530	0,9739	0,0530
	Andrew	0,9564	0,0327	0,9870	0,0402	0,9836	0,0480	0,9874	0,0734	0,9875	0,0734
	Winsor	0,8603	0,0045	0,9572	0,0095	0,9340	0,0154	0,9572	0,0349	0,9571	0,0349
	Trimmed	0,8541	0,0035	0,9541	0,0085	0,9291	0,0141	0,9541	0,0337	0,9544	0,0338
	Median	0,9056	0,0141	0,9719	0,0201	0,9570	0,0270	0,9719	0,0505	0,9718	0,0505
	HodgesLeh	0,8550	0,0041	0,9549	0,0092	0,9298	0,0149	0,9549	0,0347	0,9548	0,0347
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,9636	0,0357	0,9805	0,0433	0,9869	0,0509	0,9899	0,0720	0,9899	0,0719
	Huber	0,8827	0,0094	0,9254	0,0150	0,9475	0,0211	0,9656	0,0381	0,9655	0,0382
	Humpel	0,9268	0,0200	0,9564	0,0275	0,9710	0,0353	0,9791	0,0568	0,9791	0,0568
	Andrew	0,9691	0,0415	0,9837	0,0496	0,9901	0,0572	0,9912	0,0782	0,9913	0,0781
	Winsor	0,8921	0,0108	0,9327	0,0165	0,9523	0,0224	0,9678	0,0383	0,9677	0,0383
	Trimmed	0,8875	0,0096	0,9297	0,0153	0,9504	0,0211	0,9666	0,0371	0,9665	0,037
	Median	0,9291	0,0211	0,9574	0,0277	0,9687	0,0348	0,9800	0,0544	0,9800	0,0544
	HodgesLeh	0,8880	0,0104	0,9298	0,0161	0,9505	0,0220	0,9673	0,0382	0,9671	0,0381

CIR: Doğru sınıflandırma oranı, SR: Swamping oranı

Tablo 6. $p = 2$, $n = 120$ ve kötü kaldıraç noktası 2 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,9799	0,0444	0,9872	0,0491	0,9908	0,0543	0,9956	0,0705	0,9957	0,0703
	Huber	0,8883	0,0201	0,9157	0,0238	0,9372	0,0277	0,9633	0,0393	0,9632	0,0393
	Humpel	0,9465	0,0325	0,9642	0,0370	0,9761	0,0421	0,9872	0,0573	0,9872	0,0573
	Andrew	0,9863	0,0487	0,9912	0,0534	0,9951	0,0586	0,9979	0,0756	0,9979	0,0757
	Winsor	0,9051	0,0219	0,9313	0,0254	0,9488	0,0290	0,9708	0,0394	0,9707	0,0392
	Trimmed	0,9018	0,0214	0,9265	0,0249	0,9462	0,0285	0,9687	0,0389	0,9688	0,0387
	Median	0,9367	0,0319	0,9537	0,0360	0,9672	0,0405	0,9818	0,0544	0,9819	0,0545
	HodgesLeh	0,8952	0,0203	0,9209	0,0237	0,9421	0,0273	0,9655	0,0378	0,9654	0,0378
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,9838	0,0501	0,9906	0,0566	0,9950	0,0642	0,9976	0,0875	0,9977	0,0875
	Huber	0,9184	0,0248	0,9409	0,0297	0,9600	0,0350	0,9782	0,0509	0,9780	0,0509
	Humpel	0,9615	0,0377	0,9782	0,0437	0,9863	0,0506	0,9925	0,0717	0,9926	0,0717
	Andrew	0,9878	0,0548	0,9934	0,0612	0,9961	0,0690	0,9985	0,0935	0,9985	0,0935
	Winsor	0,9326	0,0266	0,9536	0,0313	0,9687	0,0362	0,9828	0,0507	0,9828	0,0507
	Trimmed	0,9297	0,0261	0,9498	0,0307	0,9669	0,0357	0,9817	0,0502	0,9816	0,0502
	Median	0,9534	0,0369	0,9702	0,0425	0,9806	0,0489	0,9889	0,0683	0,9888	0,0683
	HodgesLeh	0,9248	0,0250	0,9465	0,0297	0,9636	0,0346	0,9801	0,0489	0,9801	0,0489
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,9982	0,0649	0,9994	0,0706	0,9998	0,0771	0,9999	0,0955	0,9999	0,0953
	Huber	0,9757	0,0377	0,9850	0,0413	0,9911	0,0452	0,9948	0,0567	0,9947	0,0567
	Humpel	0,9925	0,0513	0,9962	0,0562	0,9984	0,0619	0,9996	0,0783	0,9996	0,0780
	Andrew	0,9988	0,0701	0,9995	0,0759	0,9998	0,0826	1,0000	0,1019	1,0000	0,1017
	Winsor	0,9821	0,0395	0,9893	0,0429	0,9942	0,0465	0,9966	0,0567	0,9966	0,0567
	Trimmed	0,9800	0,0390	0,9874	0,0424	0,9934	0,0460	0,9965	0,0562	0,9966	0,0560
	Median	0,9882	0,0504	0,9943	0,0551	0,9968	0,0601	0,9986	0,0747	0,9987	0,0748
	HodgesLeh	0,9787	0,0379	0,9866	0,0413	0,9926	0,0448	0,9957	0,0549	0,9958	0,0548

Tablo 7. $p = 2$, $n = 80$ ve kötü kaldıraç noktası 3 için simülasyon sonuçları

		DEVc		RobDEVc		RobDEVc ₁		RobDEVc ₂		RobDEVc ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,7787	0,0177	0,8477	0,0118	0,8975	0,0056	0,9465	0,0158	0,9472	0,0157
	Huber	0,6321	0,0382	0,7229	0,0332	0,7907	0,0285	0,8927	0,0132	0,8925	0,0132
	Humpel	0,7000	0,0303	0,7862	0,0244	0,8471	0,0187	0,9243	0,0017	0,9238	0,0016
	Andrew	0,7967	0,0132	0,8644	0,0068	0,9101	0,0004	0,9516	0,0218	0,9552	0,0217
	Winsor	0,6509	0,0368	0,7402	0,0319	0,8052	0,0274	0,9009	0,0136	0,9007	0,0136
	Trimmed	0,6457	0,0375	0,7347	0,0326	0,8005	0,0281	0,8986	0,0144	0,8983	0,0143
	Median	0,7095	0,0276	0,7861	0,0224	0,8429	0,0172	0,9221	0,0011	0,9219	0,0011
	HodgesLeh	0,6384	0,0380	0,7295	0,0330	0,7964	0,0285	0,8950	0,0145	0,8948	0,0145
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,7752	0,0171	0,8475	0,0100	0,8973	0,0024	0,9470	0,0266	0,9472	0,0266
	Huber	0,6308	0,0374	0,7253	0,0314	0,7959	0,0258	0,8967	0,0052	0,8968	0,0052
	Humpel	0,6965	0,0297	0,7869	0,0229	0,8495	0,0156	0,9234	0,0115	0,9231	0,0115
	Andrew	0,7948	0,0125	0,8639	0,0049	0,9104	0,0032	0,9528	0,0328	0,9527	0,0327
	Winsor	0,6470	0,0361	0,7412	0,0301	0,8084	0,0247	0,9035	0,0057	0,9034	0,0057
	Trimmed	0,6415	0,0367	0,7357	0,0309	0,8040	0,0255	0,9009	0,0065	0,9010	0,0065
	Median	0,7037	0,0268	0,7857	0,0206	0,8450	0,0139	0,9211	0,0107	0,9209	0,0107
	HodgesLeh	0,6353	0,0373	0,7322	0,0313	0,7996	0,0258	0,8993	0,0067	0,8992	0,0067
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,8551	0,0003	0,9139	0,0066	0,9475	0,0137	0,9671	0,0374	0,9673	0,0374
	Huber	0,7385	0,0257	0,8184	0,0203	0,8748	0,0150	0,9347	0,0009	0,9342	0,0009
	Humpel	0,7923	0,0159	0,8655	0,0092	0,9133	0,0024	0,9529	0,0200	0,9522	0,0200
	Andrew	0,8698	0,0048	0,9258	0,0123	0,9549	0,0198	0,9708	0,0444	0,9711	0,0443
	Winsor	0,7533	0,0237	0,8305	0,0186	0,8845	0,0134	0,9389	0,0011	0,9384	0,0011
	Trimmed	0,7473	0,0247	0,8257	0,0195	0,8810	0,0143	0,9367	0,0002	0,9363	0,0002
	Median	0,7987	0,0135	0,8659	0,0075	0,9098	0,0015	0,9516	0,0183	0,9511	0,0183
	HodgesLeh	0,7434	0,0252	0,8218	0,0200	0,8780	0,0147	0,9356	0,0002	0,9352	0,0002

Tablo 8. $p = 2$, $n = 120$ ve kötü kaldıraç noktası 3 için simülasyon sonuçları

		DEV C		RobDEV C		RobDEV C ₁		RobDEV C ₂		RobDEV C ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,8365	0,0114	0,8703	0,0164	0,9027	0,0219	0,9498	0,0368	0,9498	0,0368
	Huber	0,7528	0,0023	0,7845	0,0011	0,8181	0,0052	0,8848	0,0166	0,8849	0,0167
	Humpel	0,7897	0,0034	0,8267	0,0078	0,8611	0,0130	0,9240	0,0271	0,9239	0,0271
	Andrew	0,8472	0,0136	0,8803	0,0186	0,9115	0,0241	0,9547	0,0395	0,9547	0,0395
	Winsor	0,7599	0,0013	0,7925	0,0022	0,8275	0,0064	0,8937	0,0175	0,8940	0,0175
	Trimmed	0,7587	0,0015	0,7909	0,0019	0,8263	0,0061	0,8922	0,0172	0,8925	0,0172
	Median	0,7961	0,0047	0,8289	0,0088	0,8587	0,0136	0,9167	0,0266	0,9166	0,0266
	HodgesLeh	0,7535	0,0022	0,7861	0,0012	0,8205	0,0053	0,8869	0,0162	0,8871	0,0162
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,8412	0,0117	0,8788	0,0170	0,9128	0,0233	0,9568	0,0422	0,9568	0,0422
	Huber	0,7483	0,0022	0,7828	0,0016	0,8203	0,0062	0,8922	0,0200	0,8921	0,0201
	Humpel	0,7881	0,0036	0,8279	0,0085	0,8664	0,0141	0,9327	0,0315	0,9327	0,0316
	Andrew	0,8532	0,0139	0,8883	0,0194	0,9221	0,0256	0,9622	0,0452	0,9623	0,0452
	Winsor	0,7559	0,0012	0,7908	0,0028	0,8294	0,0075	0,9012	0,0209	0,9011	0,0209
	Trimmed	0,7547	0,0014	0,7902	0,0025	0,8271	0,0072	0,8989	0,0205	0,8985	0,0205
	Median	0,7956	0,0049	0,8301	0,0093	0,8633	0,0147	0,9233	0,0310	0,9233	0,0310
	HodgesLeh	0,7497	0,0020	0,7845	0,0017	0,8217	0,0062	0,8939	0,0194	0,8937	0,0194
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,9185	0,0252	0,9517	0,0313	0,9718	0,0380	0,9873	0,0559	0,9873	0,0559
	Huber	0,8172	0,0075	0,8597	0,0124	0,8979	0,0176	0,9499	0,0307	0,9499	0,0307
	Humpel	0,8657	0,0148	0,9107	0,0208	0,9428	0,0272	0,9741	0,0435	0,9741	0,0435
	Andrew	0,9305	0,0279	0,9600	0,0341	0,9763	0,0408	0,9898	0,0595	0,9901	0,0595
	Winsor	0,8259	0,0089	0,8719	0,0140	0,9101	0,0192	0,9563	0,0317	0,9562	0,0318
	Trimmed	0,8249	0,0086	0,8688	0,0136	0,9080	0,0188	0,9549	0,0314	0,9549	0,0314
	Median	0,8658	0,0159	0,9050	0,0214	0,9345	0,0273	0,9670	0,0427	0,9670	0,0427
	HodgesLeh	0,8183	0,0077	0,8628	0,0126	0,9017	0,0178	0,9519	0,0302	0,9518	0,0302

Tablo 9. $p = 2$, $n = 80$ ve kötü kaldıraç noktası 5 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,4943	0,0791	0,5918	0,0697	0,6096	0,0676	0,7781	0,0469	0,7791	0,0470
	Huber	0,3516	0,0993	0,4412	0,0913	0,4580	0,0894	0,6730	0,0687	0,6718	0,0689
	Humpel	0,3901	0,0941	0,4864	0,0851	0,5060	0,0828	0,7130	0,0601	0,7115	0,0602
	Andrew	0,5122	0,0761	0,6128	0,0661	0,6308	0,0637	0,7942	0,0428	0,7929	0,0429
	Winsor	0,3644	0,0978	0,4548	0,0898	0,4725	0,0878	0,6838	0,0672	0,6826	0,0674
	Trimmed	0,3559	0,0988	0,4462	0,0910	0,4633	0,0890	0,6766	0,0687	0,6753	0,0689
	Median	0,4282	0,0887	0,5163	0,0805	0,5333	0,0786	0,7256	0,0581	0,7242	0,0581
	HodgesLeh	0,3542	0,0991	0,4448	0,0911	0,4617	0,0892	0,6768	0,0683	0,6754	0,0684
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,4634	0,0827	0,5593	0,0736	0,5750	0,0719	0,7573	0,0501	0,7584	0,0502
	Huber	0,3264	0,1020	0,4122	0,0944	0,4274	0,0928	0,6465	0,0717	0,6456	0,0718
	Humpel	0,3604	0,0973	0,4540	0,0886	0,4719	0,0867	0,6880	0,0633	0,6867	0,0634
	Andrew	0,4826	0,0798	0,5818	0,0700	0,5986	0,0679	0,7728	0,0460	0,7741	0,0461
	Winsor	0,3374	0,1006	0,4244	0,0930	0,4410	0,0912	0,6570	0,0704	0,6556	0,0704
	Trimmed	0,3298	0,1016	0,4164	0,0940	0,4318	0,0924	0,6492	0,0718	0,6482	0,0719
	Median	0,3969	0,0915	0,4837	0,0838	0,4984	0,0822	0,6994	0,0613	0,6980	0,0614
	HodgesLeh	0,3280	0,1019	0,4149	0,0942	0,4302	0,0926	0,6499	0,0714	0,6489	0,0714
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,5852	0,0665	0,6730	0,0584	0,6883	0,0566	0,8276	0,0380	0,8291	0,0381
	Huber	0,4043	0,0910	0,5017	0,0829	0,5173	0,0813	0,7170	0,0617	0,7156	0,0618
	Humpel	0,4567	0,0841	0,5599	0,0753	0,5746	0,0735	0,7614	0,0523	0,7600	0,0524
	Andrew	0,6024	0,0633	0,6930	0,0545	0,7086	0,0526	0,8415	0,0337	0,8420	0,0338
	Winsor	0,4241	0,0888	0,5192	0,0808	0,5358	0,0791	0,7284	0,0600	0,7274	0,0601
	Trimmed	0,4132	0,0901	0,5096	0,0822	0,5262	0,0804	0,7218	0,0615	0,7205	0,0616
	Median	0,4889	0,0788	0,5812	0,0710	0,5961	0,0693	0,7713	0,0505	0,7691	0,0507
	HodgesLeh	0,4093	0,0907	0,5070	0,0825	0,5225	0,0808	0,7205	0,0612	0,7193	0,0613

Tablo 10. $p = 2$, $n = 120$ ve kötü kaldıraç noktası 5 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,5579	0,0475	0,6400	0,0413	0,7084	0,0353	0,8524	0,0160	0,8523	0,0161
	Huber	0,3914	0,0628	0,4734	0,0581	0,5463	0,0535	0,7206	0,0383	0,7200	0,0383
	Humpel	0,4546	0,0577	0,5442	0,0520	0,6200	0,0464	0,7914	0,0278	0,7911	0,0278
	Andrew	0,5752	0,0452	0,6597	0,0385	0,7286	0,0320	0,8668	0,0121	0,8667	0,0121
	Winsor	0,4091	0,0617	0,4931	0,0568	0,5648	0,0522	0,7365	0,0371	0,7360	0,0372
	Trimmed	0,4072	0,0618	0,4906	0,0570	0,5622	0,0524	0,7335	0,0375	0,7330	0,0375
	Median	0,4728	0,0553	0,5559	0,0499	0,6263	0,0445	0,7853	0,0275	0,7848	0,0275
	HodgesLeh	0,3976	0,0624	0,4809	0,0577	0,5525	0,0532	0,7257	0,0383	0,7253	0,0383
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,5791	0,0461	0,6565	0,0399	0,7202	0,0338	0,8614	0,0104	0,8611	0,0105
	Huber	0,4032	0,0618	0,4894	0,0567	0,5628	0,0518	0,7417	0,0341	0,7414	0,0341
	Humpel	0,4711	0,0563	0,5602	0,0506	0,6365	0,0447	0,8051	0,0229	0,8047	0,0229
	Andrew	0,5952	0,0439	0,6735	0,0372	0,7401	0,0305	0,8737	0,0061	0,8734	0,0061
	Winsor	0,4227	0,0604	0,5083	0,0554	0,5806	0,0504	0,7571	0,0329	0,7567	0,0330
	Trimmed	0,4205	0,0606	0,5056	0,0556	0,5781	0,0507	0,7546	0,0333	0,7540	0,0333
	Median	0,4873	0,0541	0,5702	0,0486	0,6413	0,0428	0,8004	0,0225	0,8000	0,0225
	HodgesLeh	0,4108	0,0613	0,4965	0,0563	0,5695	0,0514	0,7479	0,0341	0,7473	0,0342
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,6740	0,0336	0,7441	0,0266	0,8113	0,0190	0,9204	0,0059	0,9202	0,0058
	Huber	0,5208	0,0526	0,6048	0,0470	0,6746	0,0414	0,8302	0,0232	0,8298	0,0232
	Humpel	0,5860	0,0458	0,6648	0,0396	0,7352	0,0327	0,8775	0,0098	0,8773	0,0098
	Andrew	0,6876	0,0309	0,7600	0,0232	0,8281	0,0148	0,9289	0,0110	0,9290	0,0109
	Winsor	0,5402	0,0509	0,6226	0,0453	0,6909	0,0396	0,8415	0,0218	0,8413	0,0218
	Trimmed	0,5383	0,0511	0,6206	0,0456	0,6882	0,0400	0,8395	0,0222	0,8393	0,0222
	Median	0,5992	0,0434	0,6741	0,0373	0,7413	0,0308	0,8753	0,0097	0,8752	0,0097
	HodgesLeh	0,5292	0,0519	0,6120	0,0464	0,6813	0,0408	0,8340	0,0232	0,8338	0,0232

Tablo 11. $p = 3$, $n = 80$ ve kötü kaldıraç noktası 2 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,8821	0,0063	0,9395	0,0140	0,9835	0,0328	0,9850	0,0825	0,9860	0,0828
	Huber	0,8040	0,0118	0,8880	0,0064	0,9615	0,0053	0,9800	0,0424	0,9810	0,0427
	Humpel	0,8445	0,0044	0,9185	0,0030	0,9755	0,0190	0,9830	0,0666	0,9840	0,0667
	Andrew	0,8950	0,0099	0,9485	0,0181	0,9865	0,0392	0,9865	0,0908	0,9875	0,0908
	Winsor	0,8135	0,0112	0,8925	0,0061	0,9635	0,0042	0,9810	0,0356	0,9820	0,0359
	Trimmed	0,8125	0,0112	0,8925	0,0063	0,9635	0,0043	0,9805	0,0363	0,9815	0,0366
	Median	0,8420	0,0031	0,9130	0,0044	0,9705	0,0194	0,9830	0,0623	0,9840	0,0624
	HodgesLeh	0,8080	0,0121	0,8900	0,0069	0,9615	0,0034	0,9800	0,0347	0,9810	0,0350
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,8995	0,0113	0,9520	0,0215	0,9875	0,0470	0,9885	0,1082	0,9895	0,1081
	Huber	0,8265	0,0086	0,9035	0,0018	0,9710	0,0141	0,9815	0,0590	0,9825	0,0590
	Humpel	0,8610	0,0005	0,9285	0,0085	0,9815	0,0300	0,9865	0,0892	0,9875	0,0894
	Andrew	0,9090	0,0154	0,9580	0,0263	0,9875	0,0541	0,9885	0,1173	0,9895	0,1173
	Winsor	0,8350	0,0080	0,9090	0,0015	0,9735	0,0132	0,9825	0,0519	0,9835	0,0519
	Trimmed	0,8320	0,0081	0,9075	0,0017	0,9735	0,0134	0,9825	0,0526	0,9835	0,0526
	Median	0,8630	0,0014	0,9215	0,0099	0,9795	0,0303	0,9865	0,0835	0,9875	0,0836
	HodgesLeh	0,8275	0,0088	0,9060	0,0025	0,9715	0,0120	0,9820	0,0503	0,9830	0,0505
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,9345	0,0255	0,9695	0,0354	0,9840	0,0601	0,9885	0,1099	0,9895	0,1097
	Huber	0,8760	0,0004	0,9355	0,0060	0,9825	0,0199	0,9860	0,0561	0,9870	0,0561
	Humpel	0,9030	0,0104	0,9550	0,0187	0,9895	0,0394	0,9870	0,0891	0,9880	0,0890
	Andrew	0,9400	0,0299	0,9745	0,0410	0,9882	0,0689	0,9890	0,1205	0,9900	0,1203
	Winsor	0,8800	0,0015	0,9390	0,0068	0,9830	0,0192	0,9860	0,0492	0,9870	0,0493
	Trimmed	0,8790	0,0012	0,9395	0,0065	0,9830	0,0193	0,9860	0,0503	0,9870	0,0503
	Median	0,9055	0,0121	0,9530	0,0197	0,9875	0,0381	0,9870	0,0819	0,9880	0,0817
	HodgesLeh	0,8770	0,0003	0,9360	0,0054	0,9825	0,0179	0,9860	0,0481	0,9870	0,0480

Tablo 12. $p = 3$, $n = 120$ ve kötü kaldıraç noktası 2 için simülasyon sonuçları

		DEV C		RobDEV C		RobDEV C ₁		RobDEV C ₂		RobDEV C ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,9830	0,0483	0,9936	0,0580	0,9985	0,0786	0,9993	0,1174	0,9995	0,1174
	Huber	0,9550	0,0188	0,9766	0,0248	0,9927	0,0384	0,9967	0,0688	0,9967	0,0689
	Humpel	0,9747	0,0336	0,9889	0,0422	0,9970	0,0616	0,9990	0,1000	0,9991	0,1003
	Andrew	0,9862	0,0546	0,9946	0,0653	0,9993	0,0870	0,9997	0,1252	0,9997	0,1252
	Winsor	0,9588	0,0192	0,9786	0,0247	0,9931	0,0372	0,9970	0,0632	0,9971	0,0632
	Trimmed	0,9581	0,0189	0,9784	0,0245	0,9930	0,0370	0,9967	0,0631	0,9969	0,0633
	Median	0,9719	0,0326	0,9865	0,0403	0,9955	0,0576	0,9980	0,0925	0,9985	0,0922
	HodgesLeh	0,9565	0,0179	0,9769	0,0233	0,9929	0,0353	0,9966	0,0607	0,9967	0,0606
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,9879	0,0635	0,9961	0,0772	0,9994	0,1048	0,9996	0,1512	0,9997	0,1512
	Huber	0,9655	0,0283	0,9848	0,0370	0,9951	0,0562	0,9985	0,0931	0,9985	0,0933
	Humpel	0,9807	0,0462	0,9927	0,0585	0,9982	0,0847	0,9992	0,1309	0,9994	0,1310
	Andrew	0,9904	0,0709	0,9967	0,0856	0,9995	0,1147	0,9998	0,1602	0,9998	0,1602
	Winsor	0,9674	0,0288	0,9867	0,0373	0,9954	0,0548	0,9985	0,0871	0,9987	0,0871
	Trimmed	0,9672	0,0285	0,9866	0,0370	0,9953	0,0546	0,9983	0,0871	0,9986	0,0870
	Median	0,9775	0,0450	0,9913	0,0562	0,9976	0,0797	0,9991	0,1213	0,9993	0,1215
	HodgesLeh	0,9657	0,0273	0,9853	0,0354	0,9951	0,0526	0,9984	0,0840	0,9985	0,0840
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,9937	0,0797	0,9981	0,0909	0,9998	0,1117	0,9999	0,1447	0,9999	0,1449
	Huber	0,9797	0,0380	0,9916	0,0453	0,9975	0,0600	0,9992	0,0872	0,9993	0,0871
	Humpel	0,9895	0,0592	0,9960	0,0697	0,9995	0,0901	0,9997	0,1248	0,9999	0,1247
	Andrew	0,9951	0,0881	0,9985	0,1002	0,9998	0,1221	1,0000	0,1541	1,0000	0,1541
	Winsor	0,9807	0,0392	0,9930	0,0460	0,9977	0,0591	0,9994	0,0816	0,9994	0,0814
	Trimmed	0,9808	0,0387	0,9931	0,0456	0,9977	0,0588	0,9993	0,0812	0,9993	0,0817
	Median	0,9875	0,0572	0,9954	0,0665	0,9994	0,0845	0,9998	0,1150	0,9997	0,1151
	HodgesLeh	0,9796	0,0372	0,9922	0,0439	0,9975	0,0565	0,9991	0,0786	0,9993	0,0786

Tablo 13. $p = 3$, $n = 80$ ve kötü kaldıraç noktası 3 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,4913	0,0351	0,6257	0,0268	0,8487	0,0066	0,9350	0,0404	0,9317	0,0399
	Huber	0,3607	0,0495	0,4923	0,0431	0,7260	0,0287	0,8960	0,0063	0,8893	0,0060
	Humpel	0,4167	0,0442	0,5507	0,0369	0,7997	0,0184	0,9223	0,0262	0,9177	0,0258
	Andrew	0,5083	0,0326	0,6423	0,0237	0,8640	0,0022	0,9413	0,0462	0,9387	0,0455
	Winsor	0,3733	0,0486	0,5030	0,0423	0,7387	0,0283	0,9010	0,0029	0,8950	0,0027
	Trimmed	0,3723	0,0487	0,5043	0,0423	0,7363	0,0285	0,9003	0,0028	0,8943	0,0027
	Median	0,4287	0,0418	0,5573	0,0347	0,7900	0,0174	0,9153	0,0228	0,9103	0,0225
	HodgesLeh	0,3650	0,0494	0,4933	0,0431	0,7277	0,0294	0,8957	0,0016	0,8897	0,0014
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,5103	0,0324	0,6470	0,0234	0,8813	0,0011	0,9410	0,0592	0,9420	0,0588
	Huber	0,3783	0,0399	0,5150	0,0318	0,7583	0,0119	0,9080	0,0386	0,9010	0,0382
	Humpel	0,4337	0,0479	0,5750	0,0410	0,8250	0,0241	0,9280	0,0187	0,9243	0,0183
	Andrew	0,5277	0,0422	0,6707	0,0339	0,8933	0,0131	0,9465	0,0425	0,9467	0,0420
	Winsor	0,3893	0,0295	0,5277	0,0198	0,7717	0,0056	0,9113	0,0666	0,9067	0,0662
	Trimmed	0,3883	0,0471	0,5270	0,0401	0,7693	0,0238	0,9120	0,0155	0,9063	0,0153
	Median	0,4443	0,3803	0,5803	0,5160	0,8200	0,7580	0,9277	0,9090	0,9223	0,9027
	HodgesLeh	0,3803	0,0481	0,5160	0,0411	0,7580	0,0251	0,9090	0,0137	0,9027	0,0133
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,6030	0,0187	0,7430	0,0083	0,9310	0,0161	0,9590	0,0642	0,9591	0,0640
	Huber	0,4703	0,0392	0,6060	0,0319	0,8423	0,0154	0,9323	0,0195	0,9300	0,0193
	Humpel	0,5310	0,0315	0,6740	0,0230	0,8963	0,0010	0,9480	0,0456	0,9470	0,0451
	Andrew	0,6240	0,0148	0,7653	0,0038	0,9403	0,0223	0,9603	0,0716	0,9609	0,0714
	Winsor	0,4837	0,0380	0,6197	0,0306	0,8537	0,0145	0,9373	0,0170	0,9360	0,0166
	Trimmed	0,4827	0,0380	0,6180	0,0308	0,8530	0,0147	0,9377	0,0168	0,9370	0,0166
	Median	0,5410	0,0290	0,6760	0,0206	0,8890	0,0005	0,9487	0,0407	0,9477	0,0403
	HodgesLeh	0,4737	0,0390	0,6090	0,0319	0,8457	0,0162	0,9323	0,0150	0,9310	0,0149

Tablo 14. $p = 3$, $n = 120$ ve kötü kaldıraç noktası 3 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,9617	0,0207	0,9857	0,0286	0,9986	0,0462	0,9993	0,0838	0,9993	0,0837
	Huber	0,9078	0,0018	0,9548	0,0045	0,9898	0,0174	0,9971	0,0458	0,9969	0,0458
	Humpel	0,9415	0,0089	0,9744	0,0164	0,9963	0,0331	0,9986	0,0695	0,9985	0,0694
	Andrew	0,9673	0,0247	0,9891	0,0324	0,9988	0,0510	0,9992	0,0897	0,9993	0,0896
	Winsor	0,9147	0,0001	0,9593	0,0060	0,9916	0,0183	0,9976	0,0436	0,9975	0,0436
	Trimmed	0,9139	0,0006	0,9587	0,0056	0,9913	0,0177	0,9975	0,0430	0,9973	0,0430
	Median	0,9361	0,0090	0,9694	0,0159	0,9947	0,0312	0,9983	0,0643	0,9982	0,0642
	HodgesLeh	0,9108	0,0014	0,9560	0,0047	0,9907	0,0168	0,9972	0,0419	0,9971	0,0418
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,9676	0,0246	0,9891	0,0341	0,9988	0,0579	0,9994	0,1071	0,9993	0,1070
	Huber	0,9192	0,0014	0,9613	0,0086	0,9931	0,0255	0,9978	0,0628	0,9977	0,0627
	Humpel	0,9477	0,0123	0,9787	0,0214	0,9974	0,0433	0,9990	0,0905	0,9989	0,0904
	Andrew	0,9735	0,0289	0,9922	0,0384	0,9991	0,0632	0,9995	0,1141	0,9995	0,1140
	Winsor	0,9254	0,0031	0,9653	0,0101	0,9941	0,0265	0,9979	0,0605	0,9979	0,0604
	Trimmed	0,9245	0,0026	0,9647	0,0096	0,9939	0,0259	0,9979	0,0598	0,9978	0,0597
	Median	0,9417	0,0124	0,9738	0,0208	0,9960	0,0411	0,9986	0,0842	0,9985	0,0841
	HodgesLeh	0,9211	0,0018	0,9625	0,0087	0,9934	0,0249	0,9979	0,0585	0,9978	0,0583
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,9854	0,0396	0,9968	0,0491	0,9997	0,0698	0,9997	0,1057	0,9997	0,1056
	Huber	0,9566	0,0138	0,9826	0,0204	0,9977	0,0343	0,9990	0,0616	0,9989	0,0615
	Humpel	0,9727	0,0258	0,9922	0,0345	0,9991	0,0533	0,9995	0,0890	0,9995	0,0889
	Andrew	0,9891	0,0445	0,9975	0,0542	0,9997	0,0756	0,9998	0,1127	0,9999	0,1126
	Winsor	0,9605	0,0157	0,9847	0,0221	0,9979	0,0354	0,9992	0,0595	0,9991	0,0594
	Trimmed	0,9600	0,0152	0,9843	0,0216	0,9979	0,0347	0,9992	0,0588	0,9991	0,0587
	Median	0,9703	0,0259	0,9906	0,0339	0,9989	0,0510	0,9995	0,0827	0,9995	0,0826
	HodgesLeh	0,9580	0,0142	0,9833	0,0207	0,9977	0,0336	0,9991	0,0575	0,9990	0,0574

Tablo 15. $p = 3, n = 80$ ve kötü kaldıraç noktası 5 için simülasyon sonuçları

		DEVC		RobDEVC		RobDEVC ₁		RobDEVC ₂		RobDEVC ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,3134	0,0891	0,3874	0,0828	0,5192	0,0708	0,6810	0,0467	0,6800	0,0470
	Huber	0,2036	0,1097	0,2750	0,1043	0,4002	0,0946	0,5820	0,0731	0,5862	0,0731
	Humpel	0,2418	0,1025	0,3194	0,0965	0,4596	0,0843	0,6282	0,0607	0,6314	0,0605
	Andrew	0,3292	0,0858	0,4060	0,0786	0,5402	0,0662	0,6966	0,0421	0,6952	0,0425
	Winsor	0,2138	0,1081	0,2882	0,1028	0,4100	0,0931	0,5896	0,0729	0,5932	0,0728
	Trimmed	0,2134	0,1084	0,2866	0,1031	0,4080	0,0935	0,5880	0,0734	0,5914	0,0735
	Median	0,2606	0,0991	0,3332	0,0936	0,4644	0,0827	0,6274	0,0608	0,6300	0,0605
	HodgesLeh	0,2084	0,1091	0,2816	0,1037	0,4032	0,0941	0,5824	0,0740	0,5872	0,0739
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,3214	0,0868	0,3996	0,0799	0,5338	0,0668	0,6902	0,0401	0,6896	0,0402
	Huber	0,2140	0,1076	0,2862	0,1023	0,4164	0,0914	0,6004	0,0679	0,6002	0,0681
	Humpel	0,2536	0,0999	0,3336	0,0934	0,4690	0,0812	0,6440	0,0542	0,6452	0,0542
	Andrew	0,3366	0,0832	0,4208	0,0756	0,5560	0,0616	0,7050	0,0355	0,7050	0,0354
	Winsor	0,2244	0,1061	0,2994	0,1004	0,4274	0,0895	0,6070	0,0676	0,6072	0,0678
	Trimmed	0,2228	0,1064	0,2958	0,1008	0,4242	0,0903	0,6040	0,0683	0,6044	0,0684
	Median	0,2694	0,0967	0,3476	0,0908	0,4768	0,0792	0,6412	0,0545	0,6412	0,0546
	HodgesLeh	0,2178	0,1071	0,2904	0,1016	0,4204	0,0908	0,6030	0,0686	0,6040	0,0688
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,3848	0,0736	0,4554	0,0666	0,5904	0,0533	0,7310	0,0300	0,7322	0,0299
	Huber	0,2674	0,0986	0,3448	0,0926	0,4700	0,0821	0,6396	0,0607	0,6402	0,0610
	Humpel	0,3102	0,0898	0,3934	0,0825	0,5238	0,0702	0,6852	0,0455	0,6856	0,0458
	Andrew	0,3968	0,0695	0,4682	0,0620	0,6132	0,0473	0,7440	0,0243	0,7468	0,0243
	Winsor	0,2798	0,0966	0,3556	0,0908	0,4804	0,0801	0,6474	0,0602	0,6490	0,0605
	Trimmed	0,2790	0,0968	0,3540	0,0912	0,4778	0,0807	0,6452	0,0608	0,6456	0,0611
	Median	0,3286	0,0862	0,4018	0,0800	0,5286	0,0688	0,6852	0,0461	0,6854	0,0464
	HodgesLeh	0,2722	0,0979	0,3490	0,0920	0,4736	0,0815	0,6434	0,0613	0,6438	0,0616

Tablo 16. $p = 3$, $n = 120$ ve kötü kaldıraç noktası 5 için simülasyon sonuçları

		DEVc		RobDEVc		RobDEVc ₁		RobDEVc ₂		RobDEVc ₃	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0,4998	0,0374	0,6974	0,0242	0,9369	0,0031	0,9686	0,0469	0,9682	0,0465
	Huber	0,3100	0,0572	0,4810	0,0472	0,8158	0,0232	0,9346	0,0117	0,9310	0,0114
	Humpel	0,3874	0,0496	0,5848	0,0373	0,8931	0,0096	0,9562	0,0332	0,9550	0,0328
	Andrew	0,5342	0,0334	0,7315	0,0192	0,9508	0,0078	0,9724	0,0524	0,9722	0,0519
	Winsor	0,3227	0,0561	0,4988	0,0460	0,8289	0,0224	0,9382	0,0092	0,9351	0,0087
	Trimmed	0,3227	0,0561	0,4973	0,0461	0,8278	0,0226	0,9375	0,0087	0,9343	0,0083
	Median	0,4025	0,0476	0,5831	0,0362	0,8749	0,0107	0,9516	0,0289	0,9500	0,0285
	HodgesLeh	0,3118	0,0571	0,4837	0,0472	0,8175	0,0236	0,9348	0,0086	0,9313	0,0082
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
S _n	Tukey Biq.	0,5524	0,0323	0,7556	0,0184	0,9621	0,0114	0,9752	0,0668	0,9750	0,0664
	Huber	0,3309	0,0548	0,5151	0,0441	0,8543	0,0175	0,9461	0,0260	0,9436	0,0256
	Humpel	0,4212	0,0461	0,6304	0,0329	0,9248	0,0026	0,9638	0,0510	0,9630	0,0506
	Andrew	0,5906	0,0278	0,7914	0,0129	0,9708	0,0164	0,9784	0,0729	0,9785	0,0724
	Winsor	0,3447	0,0536	0,5344	0,0427	0,8677	0,0167	0,9497	0,0231	0,9476	0,0227
	Trimmed	0,3442	0,0537	0,5330	0,0428	0,8663	0,0170	0,9493	0,0227	0,9470	0,0223
	Median	0,4343	0,0442	0,6276	0,0320	0,9084	0,0038	0,9611	0,0460	0,9596	0,0455
	HodgesLeh	0,3332	0,0546	0,5194	0,0439	0,8560	0,0181	0,9468	0,0224	0,9442	0,0219
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Q _n	Tukey Biq.	0,7321	0,0123	0,8892	0,0007	0,9860	0,0257	0,9878	0,0727	0,9879	0,0717
	Huber	0,4710	0,0423	0,6747	0,0302	0,9321	0,0066	0,9668	0,0292	0,9665	0,0285
	Humpel	0,5869	0,0301	0,7910	0,0163	0,9717	0,0103	0,9782	0,0559	0,9781	0,0549
	Andrew	0,7631	0,0069	0,9110	0,0069	0,9880	0,0316	0,9882	0,0793	0,9901	0,0783
	Winsor	0,4952	0,0402	0,6970	0,0281	0,9412	0,0055	0,9692	0,0269	0,9690	0,0259
	Trimmed	0,4930	0,0404	0,6956	0,0283	0,9404	0,0057	0,9690	0,0264	0,9688	0,0256
	Median	0,5957	0,0282	0,7845	0,0157	0,9655	0,0085	0,9770	0,0500	0,9769	0,0491
	HodgesLeh	0,4775	0,0418	0,6798	0,0299	0,9340	0,0070	0,9674	0,0261	0,9672	0,0251

Tablo 5-16 incelendiğinde aşağıdaki sonuçlara ulaşılabilir:

Bağımsız değişken sayısı iki olduğu durumda; Tablo 5 incelendiğinde, $n = 80$ ve veri setinde iki kötü kaldırma noktası olduğunda DEVC ve RobDEVC klasik teşhis yöntemleri yerine bu çalışmada önerilen RobDEVC₁, RobDEVC₂ ve RobDEVC₃ teşhis yöntemlerinin konum ve ölçek parametrelerinin tüm kriterlerinde yüksek bir başarı sağladığı görülmektedir. RobDEVC₃ teşhis edici diğer yöntemlere kıyasla daha yüksek bir başarı göstermiştir. Ele alınan bu yöntemlerin CIR değerlerine bakılarak hangi eşik değerlerde kötü kaldırma noktalarını doğru bir şekilde belirlediği söylenebilir. RobDEVC teşhis edicisi dışındaki tüm teşhis edici yöntemlerinin $Andrew + 3Q_n$ eşik değerinde yüksek CIR değerine sahip olduğu görülmektedir. SR değerleri konum ve ölçek parametresine göre değişkenlik göstermektedir ve Huber konum parametresinde düşük değerler aldığı söylenebilir. Tablo 6 incelendiğinde, $n = 120$ için önerilen RobDEVC₂ ve RobDEVC₃ teşhis edici yöntemlerinin $Andrew + 3Q_n$ eşik değerinde kötü kaldırma noktalarını belirlemede daha iyi performans gösterdiği görülmektedir.

Tablo 7 incelendiğinde, $n = 80$ için ve veri setinde üç kötü kaldırma noktası olduğu durumda önerilen robust yöntemlerin mevcut yöntemlerden daha etkin sonuçlar verdiği $Andrew + 3Q_n$ eşik değerinin kullanıldığı durumda görülmektedir. DEVC ve RobDEVC teşhis yöntemlerinin $Andrew + 3Q_n$ eşik değerinde yüksek CIR değerine sahip olduğu ve $Andrew + 3MAD$ ile $Andrew + 3S_n$ eşik değerlerinin takip ettiği görülmektedir. Tablo 8 incelendiğinde $n = 120$ için tüm yöntemlerin kötü kaldırma noktalarını belirlemede başarılı oldukları ve RobDEVC₃ teşhis yönteminin diğerlerine kıyasla daha yüksek başarıya sahip olduğu görülmektedir. Tüm teşhis yöntemleri için $Andrew + 3Q_n$ eşik değerinin kullanıldığı durumlarda yüksek CIR değerine ulaşılmıştır.

Tablo 9 incelendiğinde, $n = 80$ için ve veri setinde beş kötü kaldırma noktası bulunduğu durumda mevcut olan yöntemlerin kötü kaldırma noktalarını belirlemede kullanılan tüm eşik değerlerde düşük başarı oranına sahip olduğu görülürken, önerilen RobDEVC₁, RobDEVC₂ ve RobDEVC₃ teşhis edici yöntemlerinin daha başarılı olduğu görülmektedir. Ayrıca $Andrew + 3Q_n$ eşik değeri kullanıldığı durumda SR değerleri de tüm yöntemler için düşüktür. Tablo 10 incelendiğinde, $n = 120$ için önerilen robust teşhis edici yöntemlerinin mevcut yöntemlerden daha yüksek CIR değerine sahip olduğu ve kıyaslandığında RobDEVC₃ teşhis edici yönteminin daha etkin sonuçlar verdiği görülmektedir. $Andrew + 3Q_n$ eşik değerinin kullanıldığı durumlarda yüksek başarı olduğu görülmektedir.

Bağımsız değişken sayısı üç olduğu durumda; Tablo 11 incelendiğinde, $n = 80$ için veri setinde iki kötü kaldırıcı noktası olduğu durumda önerilen robust yöntemlerde daha etkin sonuçlar verdiği görülmektedir. Ayrıca tüm teşhis edici yöntemlerin $Andrew + 3Q_n$ eşik değerinin kötü kaldırıcı noktaları belirlemede daha iyi sonuçlar verdiği görülmektedir. Tablo 12 incelendiğinde, $n = 120$ için tüm teşhis edici yöntemlerinin kötü kaldırıcı noktalarını belirlemede $Andrew + 3Q_n$ eşik değeri kullanıldığı durumda yüksek başarı gösterdiği söylenebilir. RobDEVC₂ ve RobDEVC₃ teşhis edici yöntemlerinin diğer yöntemlere kıyasla daha etkin olduğu görülmektedir. Ayrıca SR değerleri eşik değer kombinasyonu için konum parametresi Hodges-Lehmann olduğu durumda en düşük değerleri almıştır.

Tablo 13 incelendiğinde, $n = 80$ için veri setinde üç kötü kaldırıcı noktası olduğu durumda önerilen robust yöntemlerin daha yüksek CIR değerlerine sahip olduğu ve tüm teşhis edici yöntemlerinin $Andrew + 3Q_n$ eşik değerinde daha iyi sonuçlar verdiği görülmektedir. Tablo 14 incelendiğinde, $n = 120$ için tüm teşhis edici yöntemlerin kötü kaldırıcı noktalarını belirlemede başarılı olduğu ve kıyaslama yapıldığında RobDEVC₂ ve RobDEVC₃ teşhis edici yöntemlerinin daha yüksek başarıya sahip olduğu görülmektedir. $Andrew + 3Q_n$ eşik değeri tüm teşhis edici yöntemlerinde yüksek CIR değerine sahipken bu sıralamayı $Andrew + 3S_n$ ve $Andrew + 3MAD$ eşik değerlerinin takip ettiği söylenebilir.

Tablo 15 incelendiğinde, $n = 80$ için veri setinde beş kötü kaldırıcı noktası olduğu durumda mevcut yöntemlerin başarısının oldukça düşük olduğu ve önerilen teşhis edici yöntemlerinin yüksek CIR değerine sahip olduğu görülmektedir. Ayrıca $Andrew + 3Q_n$ eşik değeri tüm teşhis edici yöntemlerinde yüksek CIR değeri verirken, en düşük SR değerine de sahiptir. Son olarak Tablo 16 incelendiğinde, $n = 120$ için önerilen robust teşhis edici yöntemlerin daha etkin olduğu ve $Andrew + 3Q_n$ eşik değerinin kullanıldığı durumlarda yüksek CIR değerlerine sahip olduğu görülmektedir.

Simülasyon tabloları incelendiğinde ve tüm parametreler göz önüne alındığında önerilen RobDEVC₁, RobDEVC₂ ve RobDEVC₃ teşhis yöntemlerinin kötü kaldırıcı noktalarını belirlemede mevcut olan DEVC ve RobDEVC yöntemlerden daha yüksek başarı sağladığı görülmektedir. Önerilen teşhis edici yöntemler kıyaslandığında ise RobDEVC₃ teşhis edici yönteminin daha başarılı bir orana sahip olduğu CIR değerinden görülmektedir. Ayrıca tüm yöntemler incelendiğinde $Andrew + 3Q_n$ eşik değerinin diğer tüm eşik değer kombinasyonlarında yüksek CIR değerine sahip olduğu belirlenmiştir.

3.2. Sayısal Örnek I

Yapılan tez çalışmasının bu bölümünde kötü kaldıraç noktalarının belirlenmesinde önerilen robust teşhis edici yöntemler ele alınmıştır ayrıca bu yöntemler için önerilen robust eşik değerleri gerçek veri seti yardımıyla incelenmiştir. Modified Finney veri seti, parmak derisindeki geçici bir damar daralması üzerindeki solunan hava hızının ve hacminin etkisini incelemek için kullanılmaktadır [43]. Modified Finney veri setinde 4, 10, 11 ve 18'inci gözlemlerin kötü kaldıraç noktası olduğu bilinmektedir.

Kötü kaldıraç noktalarının belirlenmesi için mevcut teşhis edici yöntemler ile önerilen robust teşhis edici yöntemlerin sapma artık değerleri Tablo 17'de verilmiştir. Tablo 4'te verilen eşik değerler DEVC, RobDEVC, RobDEVC₁, RobDEVC₂ ve RobDEVC₃ teşhis yöntemleri için ayrı ayrı hesaplanmıştır. Mevcut eşik değer ile önerilen robust eşik değerlerin hesaplanan değerleri de Tablo18'de verilmiştir.

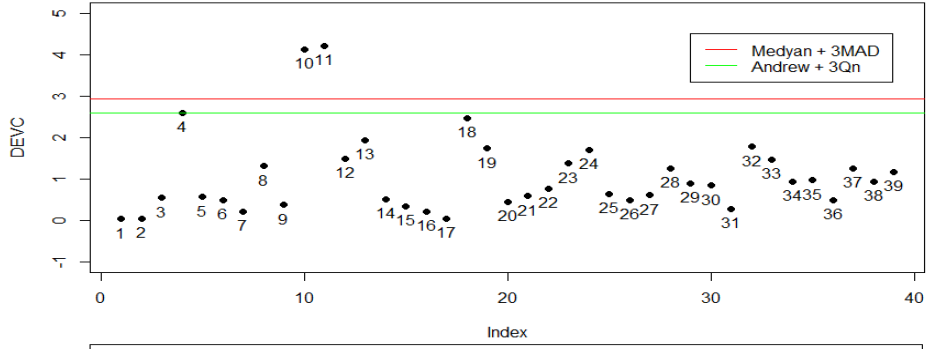
Tablo 17'de her bir gözlem için mevcut olan ve önerilen robust sapma artık değerleri hesaplanmıştır. Tablo 18'de ise ele alınan yöntemlerin sapma artıkları kullanılarak literatürde mevcut olan eşik değer ile çalışmada önerilen eşik değerler görülmektedir. Bu çalışmada ele alınan teşhis edici yöntemlerin ve literatürde mevcut olan *Medyan + 3MAD* eşik değeri ile simülasyon çalışmasında yüksek doğru sınıflandırma oranına sahip olan *Andrew + 3Q_n* eşik değerinin grafikleri Şekil 2'de verilmiştir.

Tablo 17. Teşhis edici yöntemlerin sapma artık değerleri

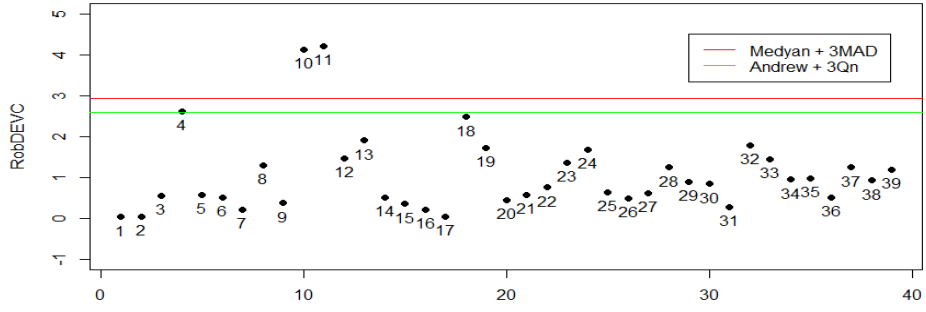
Gözlemler	DEVC	RobDEVC	RobDEVC ₁	RobDEVC ₂	RobDEVC ₃
1	0,0410	0,0412	0,0511	0,0229	0,0231
2	0,0441	0,0443	0,0542	0,0245	0,0247
3	0,5505	0,5604	0,5658	0,5272	0,5261
4	2,6003	2,6137	2,5524	3,4965	3,4836
5	0,5703	0,5843	0,5721	0,5223	0,5206
6	0,4953	0,5092	0,4954	0,4266	0,4252
7	0,2205	0,2203	0,2335	0,0895	0,0904
8	1,3108	1,2998	1,3083	1,0208	1,0250
9	0,3763	0,3760	0,3880	0,1796	0,1811
10	4,1258	4,1218	4,0574	5,7783	5,7608
11	4,2146	4,2128	4,1338	5,8785	5,8602
12	1,4908	1,4668	1,5199	1,2978	1,3037
13	1,9370	1,9053	1,9598	1,8482	1,8547
14	0,5097	0,5181	0,5288	0,4848	0,484
15	0,3486	0,3596	0,3517	0,2671	0,2665
16	0,2201	0,2225	0,2444	0,1775	0,1779
17	0,0430	0,0435	0,0520	0,0227	0,0229
18	2,4685	2,4803	2,4324	3,3207	3,3088
19	1,7332	1,7277	1,6814	1,4223	1,4252
20	0,4483	0,4535	0,4765	0,4301	0,4299
21	0,5921	0,5846	0,6211	0,3619	0,3648
22	0,7653	0,7604	0,7765	0,4840	0,487
23	1,3775	1,3697	1,3589	1,0651	1,0688
24	1,6975	1,6884	1,6604	1,4097	1,4131
25	0,6403	0,6472	0,668	0,6726	0,6716
26	0,4953	0,4913	0,5162	0,2745	0,2767
27	0,6110	0,6161	0,6447	0,6481	0,6474
28	1,2610	1,2482	1,2678	0,9828	0,9872
29	0,8927	0,8953	0,9377	1,0717	1,0702
30	0,8436	0,8454	0,8276	0,5122	0,5144
31	0,2859	0,2864	0,3227	0,2659	0,2665
32	1,7755	1,7830	1,6774	1,3770	1,3781
33	1,4633	1,4499	1,4581	1,1926	1,1970
34	0,9407	0,9536	0,9498	1,0488	1,0455
35	0,9736	0,9851	0,9871	1,1085	1,1052
36	0,4988	0,5119	0,5015	0,4362	0,4349
37	1,2610	1,2482	1,2678	0,9828	0,9872
38	0,9387	0,9285	0,9579	0,6627	0,6666
39	1,1800	1,1897	1,1974	1,4305	1,4263

Tablo 18. Hesaplanan eşik değerler

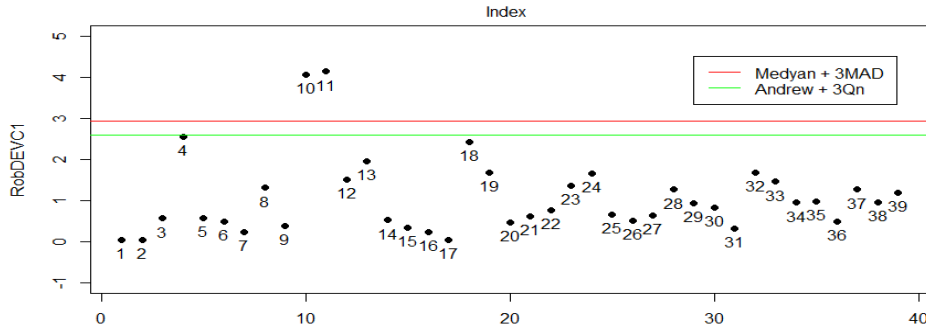
		Ölçek tahmin ediciler		
Konum tahmin ediciler		Mad	S_n	Q_n
DEVC	Medyan	2,9217	3,0768	2,9084
	Huber	3,0207	3,1757	3,0074
	Hampel	3,0559	3,2109	3,0426
	Andrew	2,9077	3,0628	2,8944
	Winsorize	3,0026	3,1576	2,9893
	Trimmed	3,0228	3,1778	3,0095
	Tukey	2,9150	3,0700	2,9017
	HodgesLeh.	3,0093	3,1643	2,9960
RobDEVC	Medyan	2,8664	3,0626	2,8926
	Huber	2,9598	3,1559	2,9860
	Hampel	2,9985	3,1947	3,0247
	Andrew	2,8515	3,0476	2,8777
	Winsorize	2,9426	3,1387	2,9688
	Trimmed	2,9655	3,1616	2,9917
	Tukey	2,8586	3,0548	2,8849
	HodgesLeh.	2,9487	3,1449	2,9750
RobDEVC ₁	Medyan	2,7855	2,9633	2,8551
	Huber	2,8996	3,0774	2,9692
	Hampel	2,9427	3,1204	3,0123
	Andrew	2,8006	2,9783	2,8701
	Winsorize	2,8950	3,0727	2,9645
	Trimmed	2,9082	3,0859	2,9777
	Tukey	2,8760	2,8760	2,8760
	HodgesLeh.	2,8984	3,0762	2,9680
RobDEVC ₂	Medyan	2,4526	2,6258	2,6119
	Huber	2,5951	2,7683	2,7544
	Hampel	2,6360	2,8092	2,7953
	Andrew	2,4313	2,6045	2,5906
	Winsorize	2,5818	2,7550	2,7411
	Trimmed	2,6385	2,8117	2,7977
	Tukey	2,4990	2,6722	2,6583
	HodgesLeh.	2,4990	2,6722	2,6583
RobDEVC ₃	Medyan	2,4553	2,6370	2,6254
	Huber	2,5947	2,7764	2,7648
	Hampel	2,6361	2,8178	2,8062
	Andrew	2,4318	2,6135	2,6018
	Winsorize	2,5819	2,7636	2,7519
	Trimmed	2,6380	2,8197	2,8081
	Tukey	2,4991	2,6808	2,6692
	HodgesLeh.	2,5857	2,7674	2,7558



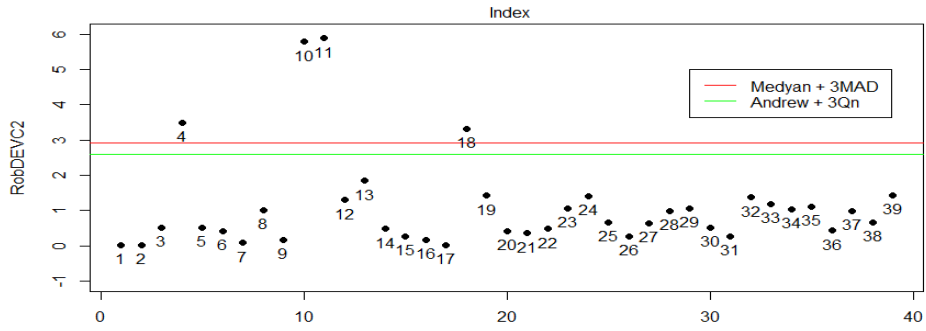
(a)



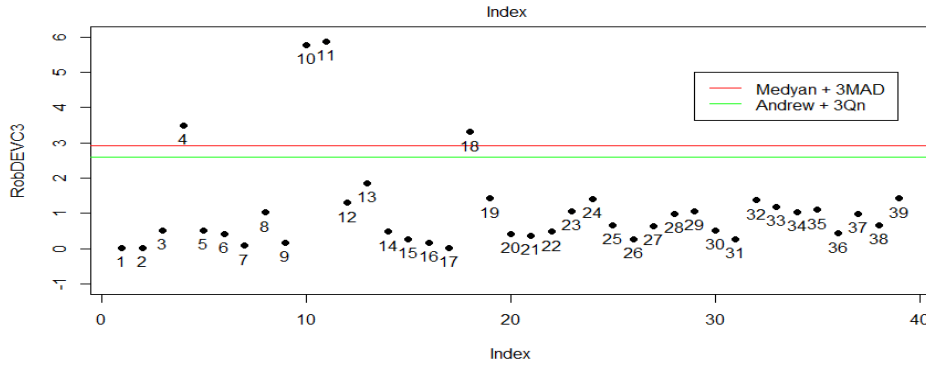
(b)



(c)



(d)



(e)

Şekil 2. DEVC (a), RobDEVC (b), RobDEVC₁ (c), RobDEVC₂ (d) ve RobDEVC₃ (e) sapma artıklarının gözlem numaralarına göre grafikleri

Şekil 2 incelendiğinde önerilen RobDEVC₂ ve RobDEVC₃ teşhis edici yöntemlerinin kullanıldığı durumda kötü kaldıraç noktalarının her iki eşik değeri de doğru bir şekilde tespit edildiği görülmektedir. DEVC ve RobDEVC₁ teşhis edici yöntemlerinde 4. ve 18. gözlemler her iki eşik değeri de belirlenemezken, RobDEVC teşhis edici yöntemi kullanıldığında $Andrew + 3Q_n$ eşik değerinin sadece 18. gözlemi doğru bir şekilde belirleyemediği görülmektedir.

3.3. Simülasyon Çalışması II

Bu bölümde Monte Carlo simülasyon çalışması ile önerilen robust yanlı tahmin edici TPRLRE ile literatürde mevcut olan LE, LRE ve bu yöntemlerin robust performansları ele alınmıştır. İncelenen bu yöntemlerin performansları MSE kriterine göre karşılaştırılmıştır. Ayrıca aykırı değer ve çoklu iç ilişki bulunan veri seti üretilmiştir. Simülasyon çalışmasında üretilen veri setinde bağımsız değişkenler McDonald ve Galarneau [89] çalışması örnek alınarak üretilmiştir. İlişkili bağımsız değişkenler,

$$x_{ij} = (1 - \rho^2)^{1/2} z_{ij} + \rho z_{ip+1}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, p + 1 \quad (109)$$

eşitliği ile elde edilmiştir. Burada n gözlem sayısını, p bağımsız değişken sayısını, ρ^2 bağımsız değişkenler arasındaki korelasyon katsayısını ve z_{ij} standart normal dağılımdan üretilen rastgele sayıları göstermektedir. Bağımlı değişken $Y_i \sim \text{Bernoulli}(\pi_i)$ dağılımı kullanılarak oluşturulmuştur. Hata terimleri ε_i , lojistik dağılımdan elde edilmiştir ($\varepsilon_i \sim A(0,1)$). Ayrıca simülasyon çalışması için parametreler,

- $n_1 = (100, 150, 200, 400)$
- $n_2 = (200, 250, 300, 400, 500)$
- $\rho^2 = (0.85, 0.9, 0.95)$
- $p = (3, 4, 6)$
- $\beta = (1, \dots, 1)$

olarak belirlenmiştir. Bağımsız değişkendeki aykırı değerler örneklem hacminin %5 ve %10 oranı kadar veriye eklenmiştir. Veri setinde aykırı değerler $x_{ij}^* = x_{ij} + \delta$ kadar ötelenerek oluşturulmuştur ($\delta = 4$) [90]. Simülasyonlar R programında 5000 kez tekrar edilmiştir ve tahmin edicilerin performans karşılaştırmaları için MSE değerleri,

$$MSE(\hat{\beta}) = \frac{1}{5000} (\hat{\beta} - \beta)' (\hat{\beta} - \beta) \quad (110)$$

ile hesaplanmıştır.

LRE_1 , $RLRE_2$, $RLRE_3$, $RLRE_4$, $RLRE_5$ tahmin edicileri için k parametresi Eşitlik (19) kullanılarak hesaplanmıştır. $TPLRE_1$, $TPRLRE_2$, $TPRLRE_3$, $TPRLRE_4$, $TPRLRE_5$ tahmin edicileri için k^* ve q^* parametreleri sırasıyla Eşitlik (99) ve Eşitlik (98) kullanılarak hesaplanmıştır.

Simülasyon çalışmasında kullanılan tahmin ediciler Tablo 19’da kısaca açıklanmıştır. Simülasyon çalışması sonuçları Tablo 20-25’te verilmiştir.

Tablo 19. Tahmin ediciler

Lojistik regresyon modeli için kullanılan tahmin ediciler	Lojistik ridge regresyon modeli için kullanılan tahmin ediciler	İki parametrelili lojistik ridge regresyon modeli için kullanılan tahmin ediciler
LE_1 : lojistik tahmin edici $RLRE_2$ $RLRE_3$ $RLRE_4$ $RLRE_5$	LRE_1 : lojistik ridge tahmin edici $RLRE_2$ $RLRE_3$ $RLRE_4$ $RLRE_5$	$TPLRE_1$: iki parametrelili ridge tahmin edici $TPRLRE_2$ $TPRLRE_3$ $TPRLRE_4$ $TPRLRE_5$

1:MLE ; 2:CUBIF ; 3:Mallows ; 4:BY ; 5:WBY

Tablo 20. $p = 3$ ve %5 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri

q	LE_1	$RLRE_2$	$RLRE_3$	$RLRE_4$	$RLRE_5$	LRE_1	$RLRE_2$	$RLRE_3$	$RLRE_4$	$RLRE_5$	$TPLRE_1$	$TPRLRE_2$	$TPRLRE_3$	$TPRLRE_4$	$TPRLRE_5$
0.85	3.2621	2.1986	2.0755	1.6518	1.5993	2.2879	1.2353	1.2027	1.1678	1.1250	1.3196	0.9817	0.9713	0.9283	0.9151
n=100 0.90	4.6604	2.8126	2.2039	2.1617	2.1127	2.7822	1.3785	1.3279	1.2462	1.2294	1.3972	0.9664	0.9731	0.9404	0.9110
0.95	10.7397	5.5623	5.5981	4.5267	4.4569	5.3006	2.1526	2.1296	1.8184	1.8189	1.4127	0.9655	0.9850	0.9836	0.9260
0.85	2.6743	1.2682	1.3224	1.1243	1.1029	2.1451	0.9223	0.8958	0.9168	0.8853	1.2280	0.8300	0.8284	0.8180	0.7980
n=150 0.90	3.7885	1.5564	1.6674	1.5049	1.4769	2.6863	1.0271	0.9856	1.0143	0.9990	1.1925	0.8318	0.8366	0.8112	0.8038
0.95	5.6280	2.9926	3.1874	2.7630	2.7375	4.9269	1.4334	1.3797	1.2784	1.2769	1.1179	0.8402	0.8437	0.8235	0.8196
0.85	2.4104	1.0269	0.9704	0.9382	0.8497	2.1133	0.9040	0.8644	0.7849	0.7657	1.0983	0.8201	0.8250	0.7106	0.7424
n=200 0.90	3.3511	1.3107	1.2626	1.2228	1.0946	2.6699	0.9894	0.9572	0.8177	0.8474	1.0905	0.8001	0.8083	0.7804	0.7462
0.95	7.9172	2.2772	2.3382	2.3457	1.9708	4.8074	1.2333	1.2726	1.0754	1.2058	1.0582	0.8114	0.8138	0.8175	0.7557
0.85	1.9757	0.7859	0.7286	0.7281	0.6811	1.5935	0.7708	0.7175	0.5672	0.5749	1.0968	0.7373	0.7219	0.5523	0.5378
n=400 0.90	2.7807	0.9286	0.8886	0.6859	0.7008	1.8687	0.8256	0.7945	0.7435	0.6783	1.0607	0.7184	0.7228	0.6326	0.6005
0.95	6.8106	1.4114	1.4788	1.1624	1.1681	2.8891	0.9795	1.0269	0.8264	0.7901	0.9847	0.7159	0.7265	0.7451	0.6726

Tablo 21. $p = 3$ ve %10 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri

q	LE_1	RLE_2	RLE_3	RLE_4	RLE_5	LRE_1	$RLRE_2$	$RLRE_3$	$RLRE_4$	$RLRE_5$	$TPLE_1$	$TPRLRE_2$	$TPRLRE_3$	$TPRLRE_4$	$TPRLRE_5$
0.85	8.0935	7.4204	7.2102	2.8716	2.6518	4.5434	1.9634	1.9533	1.5900	1.5703	1.9365	1.1950	1.1846	1.2199	1.1518
n=100 0.90	11.7280	8.1893	8.3651	4.2408	3.8654	6.3976	2.5667	2.5705	1.7932	1.7993	2.0301	1.1772	1.1685	1.2320	1.1750
0.95	55.9794	58.4294	47.4031	29.4433	49.9263	40.9295	32.6495	39.2935	34.3168	16.9181	5.0431	4.6053	4.4188	5.0452	2.7270
0.85	6.0140	2.4004	1.7594	2.2449	1.6986	4.3772	1.3361	1.2156	1.2955	1.1749	1.3631	1.0084	0.9890	0.9519	0.9459
n=150 0.90	8.8728	3.4706	2.5782	3.0541	2.4784	5.9272	1.6446	1.3915	1.2327	1.3769	1.4597	0.9953	0.9831	0.9592	0.9574
0.95	20.8594	6.6147	5.3552	5.8011	5.0846	13.8343	2.4747	2.0374	2.2647	2.0169	1.5064	1.0288	0.9820	0.9899	0.9801
0.85	5.3088	1.6862	1.6219	1.4520	1.5167	4.3741	1.1478	1.1966	1.1347	1.0668	1.5797	0.9484	0.9249	0.8809	0.8623
n=200 0.90	8.2905	2.2343	2.2037	1.8338	1.9558	6.4972	1.5597	1.3087	1.1847	1.2435	1.6967	0.9286	0.9752	0.8897	0.8745
0.95	18.2987	3.8242	4.3128	3.6432	4.0081	13.2168	1.6743	1.8515	1.6145	1.8559	1.7968	0.9246	0.9742	0.9374	0.9025
0.85	4.0392	1.0405	0.9350	1.1203	0.8168	2.8040	0.9840	0.9717	0.8829	0.8142	1.3534	0.8509	0.8310	0.7856	0.7604
n=400 0.90	6.1703	1.4169	1.1845	1.5181	1.1056	3.6125	1.1490	1.1388	0.9664	0.8355	1.5545	0.8877	0.8216	0.7947	0.7786
0.95	14.4949	2.4094	1.8201	2.2998	1.9302	7.1373	1.4424	1.2857	1.0880	1.1035	1.6892	0.9376	0.8208	0.8008	0.7842

Tablo 22. $p = 4$ ve %5 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri

q	LE_1	$RL E_2$	$RL E_3$	$RL E_4$	$RL E_5$	$LR E_1$	$RL R E_2$	$RL R E_3$	$RL R E_4$	$RL R E_5$	$TPL R E_1$	$TPRL R E_2$	$TPRL R E_3$	$TPRL R E_4$	$TPRL R E_5$
0.85	9.2126	7.6607	4.6936	4.3089	4.1141	7.5974	5.8903	3.5897	3.1104	2.8505	3.4087	2.6643	1.4174	1.4442	1.2214
n=100 0.90	13.3643	11.8353	6.3045	6.7322	5.9565	10.8573	9.7657	4.5348	4.8273	3.6334	3.4111	2.9526	1.6628	1.5557	1.2705
0.95	27.5404	26.3976	10.8297	17.7489	12.3131	21.4501	21.1016	5.6529	11.0255	6.6833	3.4271	3.2973	1.7993	1.6248	1.3087
0.85	8.5690	6.7601	4.3564	3.5216	2.8587	7.4896	4.7778	3.4655	2.7644	2.4768	3.3393	2.6359	1.3273	1.3288	1.2047
n=150 0.90	12.3759	10.7830	5.6763	4.8713	4.7139	10.5040	9.1972	4.1558	3.6321	3.9693	3.3355	3.0247	1.4731	1.3528	1.2539
0.95	25.1790	23.9346	9.3246	9.6650	11.6327	20.0004	20.0869	5.2099	6.5830	9.7198	3.3485	3.2113	1.5990	1.4166	1.2960
0.85	8.5564	6.1946	4.1187	3.3312	2.8160	7.1632	5.7620	3.2827	2.7274	2.1368	3.3121	2.5605	1.2231	1.3012	1.1131
n=200 0.90	12.2675	10.6105	5.4669	4.7366	4.2153	10.0832	8.7752	3.9463	3.4692	3.2909	3.3292	2.8896	1.4806	1.3292	1.1297
0.95	24.5665	23.1144	9.0227	9.6602	10.9000	20.6407	19.0027	5.9723	6.2992	8.9255	3.3541	3.3249	1.6688	1.2444	1.2178
0.85	8.1612	5.5402	4.0270	3.1687	1.8574	7.1167	5.7557	3.0206	2.4937	1.7607	3.3851	2.2206	1.1602	1.2804	1.1043
n=400 0.90	11.7793	7.5213	5.3873	4.4900	3.4318	10.0463	7.7174	3.6476	3.1295	3.1174	3.3892	2.5413	1.3515	1.1865	1.1148
0.95	23.4866	14.0864	8.8586	8.8371	10.0261	19.9149	18.3061	6.7357	5.5698	7.9335	3.3917	3.0627	1.5267	1.1440	1.1247

Tablo 23. $p = 4$ ve %10 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri

q	LE_1	RLE_2	RLE_3	RLE_4	RLE_5	LRE_1	$RLRE_2$	$RLRE_3$	$RLRE_4$	$RLRE_5$	$TPLE_1$	$TPRLRE_2$	$TPRLRE_3$	$TPRLRE_4$	$TPRLRE_5$
0.85	22.9877	14.9139	12.3318	6.6431	5.3809	11.9610	10.5344	4.0694	3.7223	3.6952	4.6231	3.9783	2.0188	1.8540	1.6435
n=100 0.90	28.2095	23.1550	20.1949	10.7629	7.5412	17.8747	16.2735	5.8299	4.2830	5.6247	4.6218	4.2688	1.9179	1.6476	1.6974
0.95	50.5543	49.5583	25.7260	24.2148	13.3124	35.6932	34.8508	11.9930	5.2728	9.0074	4.6353	4.5118	1.8256	1.4068	1.8326
0.85	14.1515	12.4880	5.6298	4.7848	3.0346	11.6550	9.2455	3.7685	3.4666	2.0638	4.5707	3.5387	1.7495	1.6259	1.1817
n=150 0.90	21.4263	19.6998	8.7136	6.3597	5.4194	17.2034	16.0502	5.1048	4.0721	3.3881	4.5836	3.9420	1.6558	1.4569	1.2936
0.95	45.2072	44.2496	18.5595	9.4750	13.2084	35.3277	34.4408	11.9729	5.0527	8.7839	4.5904	4.4414	1.5058	1.2298	1.5232
0.85	13.8589	11.5125	5.6078	4.7667	2.4482	11.3221	10.7091	4.0514	3.3361	1.9039	4.6210	3.2839	1.7327	1.6055	1.1110
n=200 0.90	20.8194	19.2331	8.5661	6.1942	4.4445	17.6149	16.0030	5.7763	3.9165	3.2589	4.6235	3.5008	1.6459	1.4324	1.2435
0.95	44.1208	43.4763	19.1815	9.4160	12.2077	36.6505	36.0639	11.8421	4.8091	8.3146	4.6156	3.5753	1.5012	1.1656	1.5176
0.85	13.3373	12.4261	5.2198	4.5687	1.1567	11.3558	10.4996	4.2601	3.2447	1.0529	4.5729	4.0931	1.5718	1.4718	0.8128
n=400 0.90	19.6580	19.1887	7.6249	5.7165	1.9269	17.0373	15.5989	5.8770	3.6002	1.6223	4.5809	4.3214	1.4327	1.2618	0.8759
0.95	40.7825	40.2448	16.2149	8.0384	6.2670	37.0721	36.5807	10.8787	4.2422	5.0521	4.5922	4.5548	1.2363	1.1437	0.9954

Tablo 24. $p = 6$ ve %5 aykırı değer olduğu durumda tahmin edicilerin MSE değerleri

q	LE_1	$RL E_2$	$RL E_3$	$RL E_4$	$RL E_5$	LRE_1	$RLRE_2$	$RLRE_3$	$RLRE_4$	$RLRE_5$	$TPLRE_1$	$TPRLRE_2$	$TPRLRE_3$	$TPRLRE_4$	$TPRLRE_5$
0.85	25.6453	25.5240	12.0619	5.2382	3.4797	22.2343	21.9390	8.3499	2.1556	2.3010	6.8074	6.7796	2.5627	1.4296	1.2295
n=200 0.90	40.0953	40.0625	19.8522	8.0247	4.8288	33.3177	32.8593	12.2842	2.2407	3.2905	6.8132	6.7635	2.6193	1.4341	1.2938
0.95	87.0740	86.3985	46.7265	24.2915	10.4023	70.8844	69.4894	28.1689	3.1493	7.1895	6.8204	6.8036	2.6478	1.4605	1.3609
0.85	24.8551	24.5412	11.2220	3.1581	2.9046	21.8884	21.6600	8.1182	2.0046	1.6773	6.8011	6.7705	2.2883	1.2156	1.1418
n=250 0.90	38.6336	37.6562	17.7469	4.2563	3.8813	33.2412	32.7669	12.3126	2.0173	3.1659	6.8031	6.7881	2.3048	1.2217	1.1609
0.95	83.3010	82.0416	42.6026	10.3657	7.4300	70.7140	69.6286	28.0282	2.5409	3.8737	6.8052	6.8080	2.4316	1.2262	1.2611
0.85	24.8499	24.1686	10.8220	2.5784	2.5859	21.8105	21.5026	7.9693	1.8575	1.4655	6.7926	6.7309	2.2029	1.0899	1.0832
n=300 0.90	37.9782	38.2137	17.5762	3.7890	3.1937	32.7588	32.4748	11.9406	1.9019	1.9391	6.7934	6.7775	2.3004	1.1325	1.1191
0.95	83.0330	82.4991	40.7973	8.0343	6.0525	70.3268	69.1819	27.9531	2.2996	3.3579	6.7965	6.7965	2.4163	1.1873	1.1588
0.85	24.1603	23.9094	10.7023	1.6626	2.2456	21.2602	20.5790	7.7572	1.7278	1.1357	6.7824	6.6675	2.0547	0.9989	1.0079
n=400 0.90	37.1563	36.6320	17.0074	2.5205	2.6490	32.3686	31.9840	11.9146	1.7972	1.5216	6.7850	6.7750	2.1592	1.0125	1.0176
0.95	80.3184	78.8811	39.5578	5.3986	4.2654	68.2699	67.6133	27.3888	1.8740	2.5747	6.7898	6.7882	2.3353	1.0575	1.0786
0.85	24.0721	23.7517	10.5431	1.2603	2.1582	21.1340	20.8716	7.3116	1.5900	0.9460	6.7745	6.5598	1.8703	0.9291	0.9073
n=500 0.90	36.4297	35.9219	16.1829	1.9535	2.2316	32.1750	31.8764	11.0575	1.6320	1.2851	6.7872	6.7207	2.0714	0.9473	0.9183
0.95	78.1563	76.4486	37.5628	4.0439	3.3493	68.0240	67.4307	24.6245	1.7957	2.1117	6.7887	6.7878	2.2624	0.9917	0.9763

Tablo 25. $p = 6$ ve %10 aykırı değer olduğu durumda tahmin edicilerin, MSE değerleri

q	LE_1	RE_2	RE_3	RE_4	RE_5	LRE_1	RLE_2	RLE_3	RLE_4	RLE_5	$TPLRE_1$	$TPRLRE_2$	$TPRLRE_3$	$TPRLRE_4$	$TPRLRE_5$
0.85	27.1512	27.1408	24.5787	10.0759	6.8793	23.2260	22.9225	20.7732	6.2350	3.8945	7.1089	7.1063	6.7347	2.5381	1.3839
n=200 0.90	41.9917	41.8059	37.9801	17.5385	8.5805	34.9715	34.3580	31.3956	10.7205	4.3017	7.1274	7.1250	6.7648	2.6905	1.7179
0.95	91.0321	89.9508	82.8790	81.2284	12.1378	75.1702	73.4894	67.7520	24.9662	4.8295	7.1339	7.1375	6.7948	2.9946	1.9943
0.85	26.3482	26.3172	23.7500	8.8526	6.3728	23.0515	22.7647	20.6436	6.0380	3.3876	7.1044	7.1030	6.7285	2.4348	1.2960
n=250 0.90	40.9541	40.7005	36.9481	9.0032	6.4344	34.7941	34.3318	31.3032	7.4488	4.1751	7.1244	7.1211	6.7635	1.8159	1.6797
0.95	88.2288	87.2734	80.2994	34.3418	9.2290	73.9968	72.5990	66.6989	24.4739	4.7901	7.1354	7.1312	6.7938	2.9271	1.8735
0.85	25.8043	25.6003	23.2730	6.5919	6.3000	22.6285	22.4899	20.2987	4.9319	3.1992	7.1022	7.0993	6.7281	2.1290	1.2422
n=300 0.90	27.2579	40.5215	35.1434	13.6518	7.4610	34.1792	34.0271	30.7520	9.9008	4.1815	7.1223	7.1191	6.7554	2.5296	1.5732
0.95	86.5314	85.1261	78.2565	31.6578	7.9216	73.2813	71.5138	65.9332	24.2969	4.7380	7.1286	7.1236	6.7931	2.8846	1.8365
0.85	25.4547	25.1314	22.8921	6.2844	5.4059	22.3130	22.4530	20.1835	4.6188	2.8952	7.1010	7.0978	6.7266	1.9107	1.1594
n=400 0.90	38.8890	38.3657	34.4576	10.2777	6.4167	33.5414	33.9391	30.2376	8.6058	4.1185	7.1106	7.1086	6.7548	2.2220	1.3972
0.95	83.8135	82.2136	75.8606	26.9494	6.4528	70.9613	72.0993	65.4088	22.2992	4.4524	7.1275	7.1235	6.7865	2.6912	1.7534
0.85	25.1517	24.8172	22.5738	5.9182	4.7414	22.5238	22.3069	20.0590	4.1886	2.8855	7.0996	7.0977	6.7222	1.7626	1.1182
n=500 0.90	38.2397	37.5600	22.6885	8.1202	6.2158	23.3242	33.6462	20.8795	7.0597	4.1927	7.1075	7.1064	6.7341	1.9528	1.3426
0.95	83.0904	81.2160	75.1464	26.2866	6.2544	72.3009	70.1060	64.1503	22.8005	4.4263	7.1214	7.1181	6.7860	2.6902	1.6805

Tablo 20-25 incelendiğinde aşağıdaki sonuçlar özet olarak çıkartılabilir.

- ✓ Tüm tahmin edicilerin $p = \{3, 4, 6\}$ bağımsız değişken sayısı için MSE değerleri örneklem hacmi arttıkça azalmaktadır.
- ✓ Tüm tahmin edicilerin p bağımsız değişken sayısı arttıkça MSE değerleri artmaktadır.
- ✓ Veri setinde aykırı değer arttığı durumda tüm tahmin edicilerin MSE değerleri artmaktadır.
- ✓ Veri setinde aykırı değer olması durumunda kullanılan robust tahmin ediciler MLE tahmin edicisinden daha iyi sonuçlar vermiştir. Genelleştirilmiş M tahmin edicilerinden CUBIF ve Mallows tahmin edicileri ve robust BY ve WBY tahmin edicilerinin MSE değerleri incelendiğinde WBY ve BY tahmin edicilerinin bağımsız değişken sayısına göre değişkenlik göstererek etkin sonuçlar verdiği görülmüştür.
- ✓ $p = 3$ olduğu durumda tüm parametrelerde genel olarak WBY tahmin edicisinden sonra etkin tahmin edicilerin BY ve Mallows tahmin ediciler olduğu görülmektedir.
- ✓ $p = 4$ olduğu durumda tüm parametrelerde WBY tahmin edicisinden sonra daha küçük MSE değerine ise BY ve Mallows tahmin edicisinin sahip olduğu görülmektedir. Özellikle de düşük korelasyon değerlerinde çok yakın değerler aldığı görülmektedir.
- ✓ $p = 6$ olduğu durumda WBY tahmin edicisinin etkin olduğu görülmüştür.
- ✓ Önerilen TPRLRE tahmin edicilerin hem çoklu iç ilişki hem de aykırı değer olduğu durumda MSE kriterine göre daha etkin olduğu görülmüştür.

3.4. Sayısal Örnek II

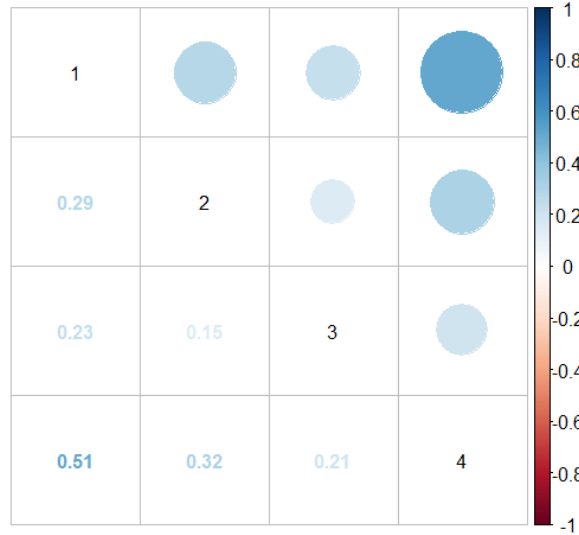
Bu bölümde önerilen tahmin edicilerin uygulamada kullanılabilirliğini göstermek için gerçek veri örneği üzerinde durulmuştur. Pena vd. [91] tarafından elde edilen bu veri setindeki yazarlar elma suyunun yeniden kontaminasyonunu azaltmak için kullanılabilecek bir lojistik regresyon modeli geliştirmeyi amaçlamışlardır. Ayrıca bu veri seti Özkale [92] ve Asar [63] tarafından da kullanılmıştır. Değişkenler pH (x_1), nisin konsantrasyonu (x_2), sıcaklık (x_3) ve çözünür katı konsantrasyonun (x_4) elma suyu kontaminasyonunun kontrolünde önemli bir rol oynadığını göstermişlerdir.

Bağımlı değişken,

$$y_i = \begin{cases} 1, & \text{Alicyclobacillus acidoterrestris büyüme gösterir} \\ 0, & \text{Alicyclobacillus acidoterrestris büyüme göstermez} \end{cases}$$

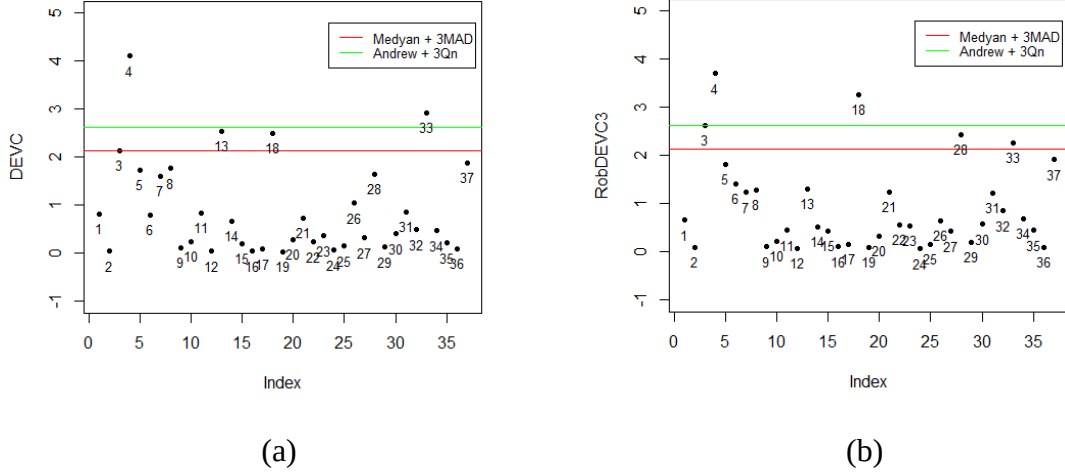
şeklinde oluşturulmuştur. Veri setinde bağımsız değişken matrisine üç aykırı değer simülasyon düzeneğine benzer şekilde oluşturularak eklenmiştir. $X'\hat{W}X$ matrisinin özdeğerleri $\lambda_1 = 5.3409$, $\lambda_2 = 0.2402$, $\lambda_3 = 0.1276$, $\lambda_4 = 0.0906$, $\lambda_5 = 0.0481$ biçiminde hesaplanmıştır. Koşul sayısı ise $K = \lambda_{\max}/\lambda_{\min} = 111.1212$ olarak bulunmuştur.

Pena verisi için korelasyon grafiği Şekil 3'te verilmiştir.



Şekil 3. Pena verisi korelasyon grafiği

Veri setinde aykırı değerler literatürde mevcut olan DEVC ve önerilen RobDEVC₃ teşhis yöntemleri ile belirlenmiştir. Teşhis yöntemleri için $Medyan + 3MAD$ ve $Andrew + 3Q_n$ eşik değerleri hesaplanmıştır ve sırasıyla 2.1416 ve 2.6946 olarak bulunmuştur. Teşhis edici yöntemlerinde hesaplanan artıkların grafiği Şekil 4'te verilmiştir.



Şekil 4. DEVC (a) ve RobDEVC₃ (b) sapma artıklarının gözlem numaralarına göre grafiği

Şekil 4 incelendiğinde veri setine eklediğimiz üç aykırı değer için, DEVC teşhis edici yönteminde *Medyan + 3MAD* eşik değeri kullanıldığı durumda veri setinde beş aykırı değer hesaplanmış, *Andrew + 3Q_n* eşik değeri kullanıldığı durumda ise iki aykırı değer hesaplanmıştır. RobDEVC₃ teşhis edici yönteminde *Medyan + 3MAD* eşik değeri kullanıldığı durumda yine beş aykırı değer belirlenirken, *Andrew + 3Q_n* eşik değeri kullanıldığı durumda ise üç aykırı değer belirlenmiştir. Önerilen RobDEVC₃ teşhis edici yöntemi ve *Andrew + 3Q_n* eşik değerinde veri setine eklenen aykırı değerler tespit edilmiştir.

Aykırı değer ve çoklu iç ilişki bulunan Pena veri seti için LE, RLE, LRE, RLRE, TPLRE ve TPRLRE tahmin edicileri MSE kriterine göre karşılaştırılmış ve bu tahmin edicilerin parametre tahminleri ve tahminlere ait standart hataları Tablo 26’da verilmiştir.

Tablo 26 incelendiğinde TPRLRE₅’in diğer tahmin edicilere kıyasla daha küçük MSE değerine sahip olduğu görülmektedir. TPRLRE modelinde robust tahmin edicilerin kullanıldığı durumda MSE değerlerinin lojistik ve lojistik ridge regresyon modellerine göre daha iyi olduğu Tablo 26’dan da görülmektedir. Bu sayısal örnek çalışmasından elde edilen sonuçlar simülasyon sonuçlarını desteklemektedir. Ayrıca standart hatalar incelendiğinde en küçük standart hataya sahip tahmin edicinin TPRLRE₅ olduğu görülmektedir.

Tablo 26. Pena veri setinin parametre tahminleri, tahminlerin standart hataları ve MSE değerleri

	LE_1	$RLRE_2$	$RLRE_3$	$RLRE_4$	$RLRE_5$	LE_1	$RLRE_2$	$RLRE_3$	$RLRE_4$	$RLRE_5$	$TPRLRE_1$	$TPRLRE_2$	$TPRLRE_3$	$TPRLRE_4$	$TPRLRE_5$
$\hat{\beta}_0$	-1.0111	-0.7397	-0.7820	-0.9132	-1.0699	-0.7555	-0.5695	-0.5888	-0.6993	-0.8410	-0.6196	-0.5560	-0.5265	-0.6352	-0.8017
$\hat{\beta}_1$	3.6967	10.2251	8.2210	3.0937	3.9749	1.6153	5.9002	4.8997	1.2039	1.7571	-0.0885	0.3278	0.2882	-0.0882	-0.0029
$\hat{\beta}_2$	-7.7075	-7.6120	-7.6093	-7.0503	-6.8323	-5.2310	-5.6523	-5.5104	-4.7244	-4.3960	-0.4766	-1.0225	-0.8039	-0.4593	-0.3953
$\hat{\beta}_3$	4.3601	4.5253	4.6179	3.8239	1.5733	2.7172	3.1279	3.1232	2.3316	0.8324	0.1304	0.3507	0.2934	0.1225	0.0122
$\hat{\beta}_4$	-7.0912	-8.1506	-7.4230	-6.3480	-4.9924	-4.1136	-4.7516	-4.4801	-3.6459	-2.7182	-0.2967	-0.5275	-0.4535	-0.2927	-0.2130
$se(\hat{\beta}_0)$	0.4826	0.4793	0.4926	0.4638	0.4621	0.4615	0.4524	0.4502	0.4327	0.4273	0.5842	0.7743	0.6367	0.7279	0.2889
$se(\hat{\beta}_1)$	4.5761	3.8071	3.1520	3.0713	3.0040	2.3880	2.3158	2.1867	2.0611	1.9010	0.3286	0.2624	0.2138	0.2414	0.1294
$se(\hat{\beta}_2)$	3.2692	3.2329	3.2075	3.0290	2.9578	2.4397	2.3178	2.1453	2.0149	1.8744	0.3198	0.2708	0.2215	0.2520	0.1295
$se(\hat{\beta}_3)$	3.1098	3.0202	2.8884	2.7641	2.6308	2.4374	2.3312	2.1702	2.0344	1.8830	0.3199	0.2649	0.2177	0.2468	0.1283
$se(\hat{\beta}_4)$	4.0610	3.7915	3.9010	3.6511	3.4201	2.4174	2.3132	2.2441	2.0957	1.9214	0.3239	0.2524	0.2098	0.2350	0.1297
MSE	58.0239	48.6726	44.0263	39.7940	36.6043	42.9612	36.1458	28.7455	24.2362	20.2933	7.4053	7.1467	5.8167	6.3327	4.7166

4. SONUÇ VE ÖNERİLER

Lojistik regresyon modelinde parametre tahminleri genellikle MLE yöntemi kullanılarak elde edilmektedir. Çoklu doğrusal regresyon analizinde olduğu gibi lojistik analizinde de karşılaşılan problemlerden biri bağımsız değişkenler arasında çoklu iç ilişkinin var olmasıdır. Çoklu iç ilişki probleminin varlığında parametre tahmin edicileri yansız fakat yüksek varyansa sahiptir. Ayrıca elde edilen tahmin değerleri teorik beklentilerden farklı işaretlerde elde edilebilmektedir. Lojistik regresyon analizindeki diğer bir problem ise veri setinde aykırı değer olması durumudur. Bu durumda MLE tahmin edicileri etkin olmayan, tutarsız sonuçlar vermektedir. Literatürde lojistik regresyon modeli için bu iki problemi ayrı ayrı ele alan birçok çalışma olmasına rağmen bu iki problemin birlikte bulunduğu durumları ele alan sınırlı sayıda çalışma söz konusudur.

Bu tez çalışmasında lojistik regresyon modelinde çoklu iç ilişki ve aykırı değer probleminin birlikte bulunması durumunda yanlı ve robust tahmin ediciler ele alınmıştır. Bölüm 1’de lojistik regresyon modeli ile ilgili genel bilgiler üzerinde durulmuştur. Literatürde var olan çalışmalardan bahsedilmiş ve kısaca açıklanmıştır. Lojistik regresyon modelinin tanımı ve parametre tahmin edicilere ait bilgilere yer verilmiştir. Tez çalışmasında ele alınan problemlerden biri olan değişkenler arasında çoklu iç ilişkinin var olması durumu incelenmiş ve lojistik regresyon modeli için yanlı tahmin edicilerden bahsedilmiştir. Lojistik regresyon modelinde karşılaşılabilen olağan dışı gözlemler ele alınmış ve literatürde mevcut olan aykırı değer ve yüksek kaldıraç noktalarının belirlenme yöntemleri açıklanmıştır. Belirleme yöntemlerinden RobDEVC teşhis edicisine alternatif teşhis edici yöntemler önermek için, aykırı değerlere karşı daha dirençli olan robust tahmin ediciler tanıtılmıştır. Mallows, CUBIF, BY ve WBY tahmin edicilerin veri setinde aykırı değer olduğu durumda MLE tahmin edicisinden daha etkin sonuçlar verdiği literatürde bilinmektedir. Bölüm 2’de yukarıda belirtilen robust tahmin ediciler yardımıyla robust teşhis yöntemleri ve robust eşik değerler önerilmiştir. Önerilen RobDEVC₁, RobDEVC₂ ve RobDEVC₃ teşhis edici yöntemleri literatürde mevcut olan DEVC ve RobDEVC teşhis edici yöntemleri ile CIR değerlerine göre karşılaştırılmıştır. Yöntemlerde kullanılan ve belli bir eşik değerin üzerindeki değerleri olağan dışı gözlemler olarak belirlediğimiz eşik değerler için de yeni robust eşik değerler önerilmiştir. Simülasyon çalışması I ile elde edilen Tablo 5-16 incelendiğinde önerilen robust yöntemlerin ve kullanılan robust eşik değerlerin mevcut

yöntemlerden daha yüksek CIR değerine sahip olduğu görülmüştür. Bu teşhis edici yöntemlerin $Medyan + 3MAD$ eşik değerine alternatif olarak önerilen $Andrew + 3Q_n$ eşik değerinde yüksek CIR değerine sahip olduğu ortaya konmuştur.

Son olarak bu tez çalışmasının amacı olan veri setinde hem aykırı değer hem de çoklu iç ilişki olması durumunda TPLRE tahmin edicisinden yola çıkarak lojistik regresyon modeli için yanlı robust TPRLRE tahmin edicileri önerilmiştir. Bu yaklaşımdan hareketle parametre tahminleri için robust teknikler kullanılmıştır. Simülasyon çalışması II ile öncelikle lojistik regresyon modelinde aykırı değer olması durumunda ele alınan robust tahmin ediciler parametre tahminlerde kullanılarak RLE_2 , RLE_3 , RLE_4 ve RLE_5 tahmin edicileri elde edilmiştir. Çoklu iç ilişki problemi için robust tahmin ediciler kullanılarak $RLRE_2$, $RLRE_3$, $RLRE_4$ ve $RLRE_5$ tahmin edicileri elde edilmiştir. Bu tahmin ediciler hem aykırı değer hem de çoklu iç ilişki problemini birlikte ele almaktadır. Asar ve Genç [63] tarafından önerilen TPLRE modeli robust tahmin ediciler kullanılarak geliştirilmiştir. Böylelikle hem aykırı değer hem de çoklu iç ilişki probleminin çözümüne öneri sağlanmıştır. Tüm tahmin edicilerin MSE kriterlerine göre karşılaştırılması Tablo 20-25'te verilmiştir. Tablolar incelendiğinde veri setinde hem aykırı değer hem de çoklu iç ilişki olduğu durumda TPRLRE tahmin edicilerinin diğer tahmin edicilere kıyasla daha etkin sonuçlar verdiği görülmüştür. TPRLRE₅ tahmin edicisinin diğer tüm tahmin edicilerle kıyaslandığında ise MSE kriteri bakımından etkin sonuçlar verdiği gerçek veri uygulamasıyla da desteklenmiştir.

Sonuç olarak veri setinde hem aykırı değer hem de çoklu iç ilişki probleminin varlığında TPRLRE₅ tahmin edicisi etkin sonuçlar vermektedir. Önerilen diğer TPRLRE tahmin edicileri de farklı durumlarda diğer tahmin edicilere göre etkin sonuçlar vermektedir. Her iki problemin varlığında TPRLRE₅ tahmin edicisinin tercih edilmesi daha tutarlı ve etkin sonuçlar elde edilmesine olanak sağlayacaktır.

İlerleyen dönemde farklı k ve q parametreleri için TPRLRE₅ tahmin edicisinin etkinliği artırılmaya çalışılabilir. Ayrıca bağımlı değişkenin kategori sayısının ikiden fazla olduğu durumda multinomial lojistik regresyon modeli ve ordinal veri setleri için tüm bu yaklaşımlar geliştirilebilir.

5. KAYNAKÇA

1. Alpar, R., Uygulamalı Çok Değişkenli İstatistiksel Yöntemler, Detay Yayıncılık, 2017.
2. Coşkun, S., Kartal, M., Coşkun, A. ve Bircan, H., Lojistik Regresyon Analizinin İncelenmesi ve Dişhekimliğinde Bir Uygulama, Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi, 7, 1 (2004) 41–50.
3. Atabey, Ö., Lojistik Regresyon Modeli ve Geriye Doğru Eliminasyon Yöntemiyle Değişken Seçiminin Hipertansiyon Riski Üzerine Uygulamasında Bootstrap Yöntemi, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2010.
4. Yavuzkanat, M.S., Lojistik Regresyonda Çoklu Aykırı Gözlemlerin Belirlenmesi ve Etkilerinin İncelenmesi. Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2011.
5. Bilgin, M., Türetilmiş İkili Heterojen Veri Yapılarında Genel, Sağlam ve Kesin Lojistik Regresyon Yöntemlerinin Karşılaştırılması, Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir, 2012.
6. Berkson, J., Application of the Logistic Function to Bio-Assay, Journal of the American Statistical Association, 39, 227 (1944) 357–365.
7. Cox, D.R., The Analysis of Multivariate Binary Data, Journal of the Royal Statistical Society, 21, 2 (1972) 113–120.
8. Pregibon, D., Logistic Regression Diagnostics, The Annals of Statistics, 9, 4 (1981) 705–724.
9. Lee, A.H., ve Silvapulle, M.J., Communications In Statistics -Simulation and Computation Ridge Estimation in Logistic Regression, Communications in Statistics -Simulation and Computation, 174, (1988) 1231–1257.
10. Bonney, G.E., Logistic Regression for Dependent Binary Observations, International Biometric Society, 43, 4 (1987) 951–973.
11. Duffy, D.E., ve Santner, T.J., On The Small Sample Properties of Norm-Restricted Maximum Likelihood Estimators for Logistic Regression Models, Communications in Statistics - Theory and Methods, 18, 3 (1989) 959–980.
12. Tester, J.M., Rutherford, G.W., Wald, Z. and W., R.M., A Matched Case–Control Study Evaluating the Effectiveness of Speed Humps in Reducing Child Pedestrian Injuries, American Journal of Public Health, 94, 4 (2004) 646–650.
13. Arı, A. ve Önder, H., Regression Models Used for Different Data Structures, Anadolu Journal of Agricultural Sciences, 28, 3 (2013) 168–174.
14. Hoerl, A.E. ve Kennard, R.W., Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, 1 (1970) 55–67.

15. Schaefer, R.L., Roi, L.D. ve Wolfe, R.A., A Ridge Logistic Estimator, Communications in Statistics - Theory and Methods, 13, 1 (1984) 99–113.
16. Lipovetsky, S., ve Conklin, W.M., Ridge Regression in Two-Parameter Solution. Applied Stochastic Models in Business and Industry, 21, 6 (2005) 525–540.
17. Lipovetsky, S., Two-Parameter Ridge Regression and Its Convergence to the Eventual Pairwise Model, Mathematical and Computer Modelling, 44, 3–4 (2006) 304–318.
18. Özkale, M.R. ve Kaçiranlar, S., The Restricted and Unrestricted Two-Parameter Estimators, Communications in Statistics - Theory and Methods, 36, 15 (2007) 2707–2725.
19. Månsson, K. ve Shukur, G., On Ridge Parameters in Logistic Regression. Communications in Statistics - Theory and Methods, 40, 18 (2011) 3366–3381.
20. Kibria, B.M.G., Månsson, K. and Shukur, G., Performance of Some Logistic Ridge Regression Estimators, Computational Economics, 40, 4 (2012) 401–414.
21. Muniz, G., Kibria, B.M.G., Mansson, K. ve Shukur, G., On Developing Ridge Regression Parameters: A Graphical Investigation, Statistics and Operations Research Transactions, 36, 2 (2012) 115–138.
22. Månsson, K., Kibria, B.M.G. ve Shukur, G., On Liu Estimators for the Logit Regression Model, Economic Modelling, 29, 4 (2012) 1483–1488.
23. Kejian, L., A New Class of Biased Estimate in Linear Regression, Communications in Statistics - Theory and Methods, 22, 2 (1993) 393–402.
24. Huang, J., A Simulation Research on a Biased Estimator in Logistic Regression Model, Springer, Berlin, 2012.
25. Inan, D. ve Erdogan, B.E., Liu-Type Logistic Estimator, Communications in Statistics: Simulation and Computation, 42, 7 (2013) 1578–1586.
26. Asar, Y., Karaibrahimoğlu, A. ve Genç, A., Modified Ridge Regression Parameters: A Comparative Monte Carlo Study, Hacettepe Journal of Mathematics and Statistics, 43, 5 (2014) 827–841.
27. Ertaş, H., Toker, S. ve Kaçiranlar, S., Robust Two Parameter Ridge M-Estimator for Linear Regression, Journal of Applied Statistics, 42, 7 (2015) 1490–1502.
28. Asar, Y. ve Erişoğlu, M., Influence Diagnostics in Two-Parameter Ridge Regression, Journal of Data Science, 14 (2016) 33–52.
29. Özkale, M.R., Lemeshow, S. ve Sturdivant, R., Logistic Regression Diagnostics in Ridge Regression, Computational Statistics, 33, 2 (2017) 563–593.
30. Asar, Y., Performance of a New Restricted Biased Estimator in Logistic Regression, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 22, 1 (2018) 53–59.
31. Huang, H., Wu, J. ve Yi, W., On the Restricted Almost Unbiased Two-Parameter Estimator in Linear Regression Model, Communications in Statistics - Theory and Methods, 46, 4 (2017) 1668–1678.

32. Kandemir Çetinkaya, M. ve Kaçıranlar, S., Improved Two-Parameter Estimators for the Negative Binomial and Poisson Regression Models, Journal of Statistical Computation and Simulation, 89, 14 (2019) 2645–2660.
33. Bulut, Y. M. ve Işılar, M., Two parameter Ridge estimator in the inverse Gaussian regression model, Hacettepe Journal of Mathematics and Statistics, 50, 3 (2021) 895–910.
34. Box, G.E.P., Non-Normality and Tests on Variances, Biometrika Trust, 40, 3 (1953) 318–335.
35. Huber, P. j., Robust Estimation of A Location Parameter, The Annals of Mathematical Statistics, 35, 1 (1953) 73–101.
36. Hampel, F.R., The Influence Curve and its Role in Robust Estimation, Journal of the American Statistical Association, 69, 346 (1974) 383–393.
37. Nevitt, J. and Tam, H., A Comparison of Robust and Nonparametric Estimators Under the Simple Linear Regression Model, Multiple Linear Regression Viewpoints, 25 (1998) 54–69.
38. Wisnowski, J.W., Montgomery, D.C. ve Simpson, J.R., A Comparative Analysis of Multiple Outlier Detection Procedures in the Linear Regression Model, Computational Statistics and Data Analysis, 36, 3 (2001) 351–382.
39. Schumacker, R.E., Monahan, M.P. ve Mount, R.E., A Comparison of OLS and Robust Regression Using S-PLUS, Multiple Linear Regression Viewpoints, 28, 2 (2002) 10–13.
40. Rousseeuw, P.J. ve Christmann, A., Robustness Against Separation and Outliers in Logistic Regression, Computational Statistics and Data Analysis, 43, 3 (2003) 315–332.
41. Croux, C. ve Haesbroeck, G., Implementing the Bianco and Yohai Estimator for Logistic Regression, Computational Statistics and Data Analysis, 44, 1–2 (2003) 273–295.
42. Ergül, B., Robust Regresyon ve Uygulamaları, Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir, 2006.
43. Imon, A.H.M.R. ve Hadi, A.S., Identification of Multiple Outliers in Logistic Regression, Communications in Statistics - Theory and Methods, 37, 11 (2008) 1697–1709.
44. Yıldırım, N., En Küçük Kareler, Ridge Regresyon ve Robust Regresyon Yöntemlerinde Analiz Sonuçlarına Aykırı Değerlerin Etkilerinin Belirlenmesi, Yüksek Lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana, 2010.
45. Syaiba, B.A. ve Habshah, M., Robust Logistic Diagnostic for the Identification of High Leverage Points in Logistic Regression Model, Journal of Applied Sciences, 10, 23 (2010) 3042–3050.
46. Habshah, M. ve Syaiba, B.A., The Performance of Classical and Robust Logistic Regression Estimators in The Presence of Outliers, Pertanika Journal of Science and Technology, 20, 2 (2012) 313–325.

47. Imon, A.H.M.R. ve Hadi, A.S., Identification of Multiple High Leverage Points in Logistic Regression, Journal of Applied Statistics, 40, 12 (2013) 2601–2616.
48. Ariffin, S.B. ve Midi, H., The Effect of High Leverage Points on the Logistic Ridge Regression Estimator Having Multicollinearity, AIP Conference Proceedings, (2014) 1105–1111.
49. Feng, J., Xu, H., Mannor, S. ve Yan, S., Robust Logistic Regression and Classification, Proceeding, 1 (2014) 253–261.
50. Yu, C. ve Yao, W., Robust Linear Regression: A Review and Comparison. Communications in Statistics: Simulation and Computation, 46, 8 (2017) 6261–6282.
51. Md Shariff, N.S. ve Ferdaos, N.A., An Application of Robust Ridge Regression Model In The Presence of Outliers to Real Data Problem, Journal of Physics: Conference Series, 890, 1 (2017) 1-6.
52. Toker, S., Ridge Regresyon Tahmin Edici ve İlişkili Tahmin Ediciler, Yüksek Lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana, 2011.
53. Ertaş, H., Dayanıklı Yanlı Tahmin Ediciler, Doktora Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana, 2015.
54. Asar, Y., Liu Type Logistic Estimator, Doktora Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya, 2015.
55. Işılar, M., Çoklu Doğrusal Regresyon Modelinde Liu Tipi Tahmin Edici GM Tahmin Edicisi, Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir 2020.
56. Bishop, C.M., Pattern Recognition and Machine Learning, Springer, 2006.
57. Akçay, A., Broiler Coccidiosis’inde Risk Faktörlerinin Lojistik Regresyon Analizi İle Belirlenmesi, Doktora Tezi, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara 2009.
58. Gujerati, D.N. ve Porter, D.C., Temel ekonometri, Beşinci Basım, Literatür Yayıncılık, 2014.
59. Ryan, T.P., Modern regression methods, Second Edition, Wiley, 1997.
60. Şahin, M., Lojistik Regresyon ve Biyolojik Alanlarda Kullanımı, Yüksek Lisans Tezi, KSÜ, Fen Bilimleri Enstitüsü, Kahramanmaraş, 1999.
61. Montgomery, D.C., Peck, E.A., and Vining, G.G., Introduction to linear regression analysis, Wiley, 2012.
62. Theil, H., Principles of Econometrics, John Wiley and Sons Inc., New York, 1971.
63. Asar, Y. ve Genç, A., Two-Parameter Ridge Estimator in the Binary Logistic Regression Model, Communications in Statistics: Simulation and Computation, 46, 8 (2017) 7088–7099.
64. Nurunnabi, A.A.M., Nasser, M. ve Imon, A.H.M.R., Identification and Classification Of Multiple Outliers, High Leverage Points and Influential Observations in Linear Regression, Journal of Applied Statistics, 43, 3 (2016) 509–525.

65. Walfish, S., A Review of Statistical Outlier Methods, Pharmaceutical Technology, (2006) 1-5.
66. Barnett, V. ve Lewis, T., Outliers in statistical data, Three Edition, Wiley, 1994.
67. Belsley, D.A., Kuh, E. ve Welsch, R.E., Regression Diagnostics, John Wiley & Sons, Inc., USA, 1980.
68. Chatterjee, S. ve Hadi, A.S., Sensitivity Analysis in Linear Regression, John Wiley & Sons, Ltd, New York, 1988.
69. Cook, R.D. ve Weisberg, S., Graphs in Statistical Analysis: Is the Medium the Message? The American Statistician, 53, 1 (1999) 29-37.
70. Krogh, A., The Respiratory Exchange of Animals and Man, London: Longmans Green & Co., 1916.
71. Norazan, M.R., Sanizah, A. ve Habshah, M., Identifying Bad Leverage Points in Logistic Regression Model Based on Robust Deviance Components, Mathematical Models and Methods in Modern Science, (2012) 62–67.
72. Kutner, M.H., Nachtsheim, C.J., Neter, J. ve Li, W., Applied Linear Statistical Models, Five Edition, McGraw-Hill/Irwin, 1996.
73. Rousseeuw, P.J. ve Leroy, A.M., Robust Regression and Outlier Detection, Hoboken, NJ, John Wiley & Sons, Inc., USA, 1987.
74. C. L. Mallows, Some Comments On Cp, Technometrics, 15 (1973) 87–94.
75. Rousseeuw, P.J. ve Leroy, A.M., Robust Regression And Outlier Detection, John Wiley, 1987.
76. Ahmad, S. ve Ahmad, S., Robust Estimators in Logistic Regression : A Comparative Simulation Study, Journal of Modern Applied Statistical Methods 9, 2 (2010) 502-511.
77. Carroll, R.J. ve Pederson, S., On Robustness in the Logistic Regression Model, Journal of the Royal Statistical Society: Series B (Methodological), 55, 3 (1993) 693–706.
78. Künsch, H.R., Stefanski, L.A. ve Carroll, R.J., Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, With Applications to Generalized Linear Models, Journal of the American Statistical Association, 84, 406 (1989) 460–466.
79. Bianco, A.M. ve Yohai, V.J., Robust Estimation in the Logistic Regression Model, Springer V, New York, 1996.
80. Huber, P. j., Robust Statistics, John Wiley and Sons Inc., New York, 1981.
81. Maronna, R.A. ve Yohai, V.J., Estimates of Regression and Scale the Breakdown Point of Simultaneous General M Estimates of Regression and Scale, (1901) 1–6.
82. Croux, C.; Flandre, C. ve Haesbroeck, G., The Breakdown Behavior of The Maximum Likelihood Estimator in the Logistic Regression Model, Statistics and Probability Letters, 60, 4 (2002) 377–386.

83. Daniel, C., Observations Weighted According to Order, Amer. J. Math., 42 (1920) 222–236.
84. Hoaglin, D.C. ve Mosteller, F., Understanding robust and explanatory data analysis, John Wiley & Sons, New York, 1983.
85. Hodges, J.L.J. ve Lehmann, E.L., Estimates of Location Based on Ranks Tests, Annals of Mathematical Statistics, 34, 2 (1963) 598–611.
86. Rousseeuw, P.J. ve Christophe, C., Alternatives to The Median Absolute Deviation, 88, 424 (1993) 1273-1283.
87. Farebrother, R., Further Results on the Mean Square Error of Ridge Regression, Journal of the Royal Statistical Society, 36, 1 (1976) 103–106.
88. Trenkler, G. ve Toutenburg, H., Mean Squared Error Matrix Comparisons Between Biased Estimators -An Overview of Recent Results, Statistical Papers, 31, 1 (1990) 165–179.
89. McDonald, G.C. ve Galarneau, D.I., A Monte Carlo Evaluation of Some Ridge-Type Estimators, Journal of the American Statistical Association, 70, 350 (1975) 407–416.
90. Midi, H. ve Ariffin, S.B., Modified Standardized Pearson Residual for The Identification of Outliers in Logistic Regression Model, Journal of Applied Sciences, 13, 6 (2013) 828–836.
91. Peña, W.E.L., De Massaguer, P.R., Zuñiga, A.D.G. ve Saraiva, S.H., Modeling The Growth Limit of Alicyclobacillus Acidoterrestris Cra7152 In Apple Juice: Effect Of Ph, Brix, Temperature And Nisin Concentration, Journal of Food Processing and Preservation, 35, 4 (2011) 509–517.
92. Özkale, M.R., Iterative algorithms of biased estimation methods in binary logistic regression, Springer, 57, 4 (2016) 991-1016.

ÖZGEÇMİŞ

Ebru GÜNDÖĞAN AŞIK ilköğretimini Ahmet Mesut Yılmaz İlköğretim Okulu'nda bitirdikten sonra lise eğitimini 75.Yıl İMKB Anadolu Lisesi'nde tamamladı. 2008 yılında kazandığı Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü'nden 2012 yılında mezun oldu ve 2013 yılında İstatistik Anabilim dalında yüksek lisans eğitimine başladı. 2015 yılında KTÜ, Fen Bilimleri Enstitüsü, İstatistik ve Bilgisayar Bilimleri Anabilim dalına araştırma görevlisi olarak atandı ve aynı yıl “Çoklu İç İlişki ve Aykırı Değer Olduğu Durumda Lojistik Regresyon Yaklaşımları” başlıklı tezini tamamlayarak KTÜ, Fen Bilimleri Enstitüsü, İstatistik ve Bilgisayar Bilimleri Anabilim dalında doktora eğitimine başladı. Halen KTÜ, Fen Bilimleri Enstitüsü, İstatistik ve Bilgisayar Bilimleri Anabilim dalında araştırma görevlisi olarak çalışmaktadır. Evli ve ikiz çocuk annesidir. Orta derecede İngilizce bilmektedir.

Tezden üretilen ESCI dizininde makale:

Gündoğan Aşık, E., Altın Yavuz, A. & Küçük, Z. (2021). LOJİSTİK REGRESYON MODELİNDE KÖTÜ KALDIRAÇ NOKTALARININ BELİRLENMESİ İÇİN YENİ SAĞLAM EŞİK DEĞERLER. Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 8 (2), 630-650. DOI: 10.30798/makuiibf.691623.