ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

LOJİSTİK REGRESYONDA ROBUST TAHMİN YÖNTEMLERİNİN KULLANILMASI

Tuğçe PARLAK

İSTATİSTİK ANABİLİM DALI

ANKARA
2019

TEZ ONAYI

Tuğçe PARLAK tarafından hazırlanan **"Lojistik Regresyonda Robust Tahmin Yöntemlerinin Kullanılması"** adlı tez çalışması 04/10/2019 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Prof. Dr. Olcay ARSLAN
Ankara Üniversitesi İstatistik Anabilim Dalı

Jüri Üyeleri:

Başkan: Doç. Dr. Esin KÖKSAL BABACAN
Ankara Üniversitesi İstatistik Anabilim Dalı

Üye : Prof. Dr. Olcay ARSLAN
Ankara Üniversitesi İstatistik Anabilim Dalı

Üye : Dr. Öğr. Üyesi Fulya GÖKALP YAVUZ
Orta Doğu Teknik Üniversitesi İstatistik Anabilim Dalı

Yukarıdaki sonucu onaylarım.

Prof. Dr. Özlem YILDIRIM
Enstitü Müdürü

ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

04/10/2019

Tuğçe PARLAK

ÖZET

Yüksek Lisans Tezi

LOJİSTİK REGRESYONDA ROBUST TAHMİN YÖNTEMLERİNİN KULLANILMASI

Tuğçe PARLAK

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Olcay ARSLAN

En çok olabilirlik tahmin edicisi (MLE), parametrik bir model altında etkinliği nedeniyle lojistik regresyon modellerinin parametre tahminleri için sıklıkla kullanılır. Fakat en çok olabilirlik yöntemi aykırı değerlerin varlığında parametrelerin tahminlerinde doğru olmayan sonuçlar verebilmektedir. Bu tezde, parametre tahmini yaparken aykırı değerlerin meydana getirdiği bozucu etkinin en aza indirilebilmesi için robust yöntemler araştırılmıştır. En çok olabilirlik tahmin edicisinin performansı ile alternatif olarak öne sürülen ağırlıklandırılmış Bianco-Yohai tahmin edicisi (WBYE), ağırlıklandırılmış Mallows tahmin edicisi ve ağırlıklandırılmış en çok olabilirlik tahmin edicisinin (WMLE) performanslarını karşılaştırmak için simülasyon çalışması ve gerçek veri üzerinde çalışmalar yapılmıştır.

Eylül 2019, 72 sayfa

Anahtar Kelimeler: Lojistik regresyon, Bağımlı değişken, Bağımsız değişken, Robust lojistik regresyon, Aykırı değer, Bozulma oranı, Yan, Ağırlıklandırılmış en çok olabilirlik, Tahmin edici, Hata Kareler Ortalaması

ABSTRACT

Master Thesis

USING ROBUST ESTIMATION METHODS IN LOGISTIC REGRESSION

Tuğçe PARLAK

Ankara University
Graduate School of Nature and Applied Sciences
Department of Statistics

Supervisor: Prof. Dr. Olcay ARSLAN

To estimate the parameters of logistic regression models are often used the maximum likelihood estimator (MLE) owing to its good property under a parametric model. However, the maximum likelihood method can give inefficient parameter estimations in the presence of outliers. In this thesis, robust methods are considered to estimate the parameters of a logistic regression model when there are outliers in data. A simulation study and a real data example are afforded to contrast the performance of the maximum likelihood estimator with the performances of the weighted Bianco-Yohai estimator, the weighted Mallows estimator and the weighted maximum likelihood estimator.

September 2019, 72 Pages

Key Words: Logistic regression, Dependent variable, Independent variable, Robust logistic regression, Outlier, Contamination rate, Bias, Weighted maximum likelihood, Estimator, Mean Squared Error

TEŞEKKÜR

Yüksek lisans eğitimim ve tez çalışmam sırasında rehberliği ile beni yönlendiren, araştırmalarım sırasında bilgi, öneri ve desteklerini esirgemeyen, yetişmem ve gelişmemde katkı sağlayan sayın danışmanım Prof. Dr. Olcay ARSLAN ‘ a (Ankara Üniversitesi İstatistik Anabilim Dalı) teşekkürü borç bilirim.

Beni bu süreçte destekleyen ve cesaretlendiren arkadaşlarım Melike Özlem Karaduman, Gülşah Tekatlı, Selma Keçeli, Nurdan Helvacıoğlu ve Yağmur Kibar’a, çalışmalarım boyunca beni maddi ve manevi olarak destekleyen kıymetli annem Arife Parlak, babam Abbas Parlak ve ablam Zeynep Parlak’a derinden teşekkür ederim.

Tuğçe PARLAK
Ankara, Eylül 2019

İÇİNDEKİLER

KISALTMALAR DİZİNİ

BY	Bianco ve Yohai
LR	Olabilirlik oranı
MCD	En küçük kovaryans determinantı
MLE	En çok olabilirlik tahmin edicisi
MSE	Hata kareler ortalaması
WBYE	Ağırlıklandırılmış Bianco ve Yohai tahmin edicisi
WMLE	Ağırlıklandırılmış en çok olabilirlik tahmin edicisi

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

1. GİRİŞ

Regresyon analizi, değişkenler arasındaki bağlantıyı araştırmak ve modellemek için kullanılan istatistiksel yöntemlerden biridir. Regresyon yöntemi fizik, mühendislik ve biyoloji, kimya, sosyal bilimler ve ekonomiyi kapsayan birçok alanda sayısız kullanıma sahiptir.

Lojistik regresyonda, doğrusal regresyona benzer olarak tahminler değişkenlere bağlı olarak yapılır. Lakin bu iki metot birbirinden 3 farklı şekilde ayrılır. Bunlar:

1. Lojistik regresyonda tahmin edilen bağımlı değişken kesikli değerler alırken, doğrusal regresyonda bağımlı değişken sürekli değer almaktadır.
2. Lojistik regresyonda bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilirken, doğrusal regresyonda bağımlı değişkenin değeri tahmin edilir.
3. Lojistik regresyonda bağımsız değişkenin belli bir dağılımdan gelmesi gerekmezken, doğrusal regresyonda bağımsız değişkenlerin çoklu normal dağılım göstermesi koşulu aranır (Elhan, A. H. 1997).

Lojistik modelleme, özellikle normallik ve ortak kovaryansa sahip olma gibi gerekli varsayımların yerine getirilmediği durumlarda, regresyon yöntemlerine bir alternatiftir. Kesikli ve sürekli değişkenler içeren geniş bir parametrik dağılım ailesi için kullanılır (Day ve Kerridge 1967, Anderson 1972). Fakat uygulamada ve model oluşturma sürecinde, verilerdeki aykırı değerlerden oldukça etkilendiği için en çok olabilirlik yöntemi ile yinelemeli olarak parametre tahmini yapılırken bazı sorunlarla karşı karşıya kalınır. Lojistik regresyon varsayımını kontrol etmek için uygunluk testleri (Hosmer ve Lemeshow 1980, Tsiatis 1980) gibi çeşitli yöntemler geliştirilmiştir.

Çok değişkenli lojistik regresyon modeli literatürde sıklıkla kullanılmasına rağmen yine de aykırı değerlerin varlığında parametre tahmini yaparken karşılaştığı bir takım problemler vardır.

Lojistik regresyon yönteminin uygulanabilmesi, büyük örnek hacmine ve olabilirlik fonksiyonunun kullanılmasına bağlıdır. Ancak lojistik regresyon yöntemi değişkenin

beklenenin dışında aykırı değerler alması durumunda geçerli ve güvenilir sonuçlar vermeyebilir.

Bunun yanı sıra robust lojistik regresyon yöntemi ise, veri setinde aykırı değer olması durumunda lojistik regresyon modeline alternatif olarak kullanılır. Bu tezin amacı, lojistik regresyonda kullanılan klasik parametre tahmin yöntemiyle parametre tahmini yaparken x yönünde aykırı gözlemlerin olması durumda, robust tahmin yöntemlerini kullanmaktır. Bu yöntemler; ağırlıklandırılmış Bianco-Yohai tahmin yöntemi, ağırlıklandırılmış Mallows tahmin yöntemi ve özel bir ağırlık matrisi seçilerek oluşturulan ağırlıklandırılmış en çok olabilirlik tahmin yöntemidir. Ayrıca, kullanılan robust yöntemlerinin performanslarını karşılaştırarak; hangi yöntemin modeli daha iyi şekilde açıkladığına karar verilecektir.

Bu tezde, ilk olarak lojistik regresyon modeli ve tez çalışmasında kullanılacak olan robust yöntemlerden kısaca bahsedilmiştir.

İkinci kısımda ise tek değişkenli lojistik regresyon modelinin nasıl olduğu anlatılıp, parametre tahmini, parametrelerin nasıl yorumlanacağı, parametre anlamlılık testleri ve son olarak güven aralıkları ile ilgili kısa bilgiler verilmiştir.

Tezin üçüncü bölümünde çok değişkenli lojistik regresyona ilişkin parametre tahminlerinin nasıl yapıldığından, parametrelerin anlamlılık testlerinden, model katsayısı üzerine testlerden ve güven aralıklarının nasıl oluşturulacağından bahsedilmiştir.

Tezin dördüncü kısmında ise veride x yönünde aykırı değer olması durumda, lojistik regresyondaki parametre tahmininden daha tutarlı olan robust lojistik regresyon yöntemleri ile parametre tahminlerinin nasıl yapıldığı anlatılmıştır.

Tezin beşinci bölümünde ise rasgele aykırı değerlerler üretilip lojistik (MLE) ve robust lojistik regresyon için kullanılan parametre tahmin yöntemlerinin (Mallows, WMLE, WBYE) etkinliği simülasyon çalışması ile karşılaştırılmıştır. Ayrıca, simülasyon çalışmasında kullandığımız yöntemlerin etkinliği gerçek veri seti kullanarak da karşılaştırılmıştır.

Tezin son kısmında ise simülasyon ve gerçek veri çalışmalarından elde edilen sonuçlar yorumlanmıştır.

2. TEK DEĞİŞKENLİ LOJİSTİK REGRESYON

Lojistik regresyon modeli bağımlı değişkenin 0 ve 1 gibi iki sonuç ya da ikiden fazla sonuca sahip kesikli bir değişken olduğunda uygulanan, matematiksel açıdan esnek ve yorumu kolay bir tekniktir (Hosmer ve Lemeshow 2000). Bağımsız değişkenlerden yararlanarak bağımlı değişkene ait beklenen değerin olasılık değeri olarak elde edildiği bir regresyon yöntemidir.

Tek değişkenli ve çok değişkenli lojistik regresyon modellerinde bağımlı değişken iki olası sonuca sahip olduğu durumlar çalışılmıştır. Tek ve çok değişkenli lojistik regresyon modelleri arasındaki fark bağımsız değişken sayısıdır. Tek değişkenli lojistik regresyonda 1, çok değişkenli lojistik regresyonda ise 2 veya daha fazla bağımsız değişken bulunmaktadır.

Basit doğrusal regresyon modelinin çeşitli gösterim biçimleri vardır. Genel olarak,

$$y = \beta_0 + \beta_1 x + \varepsilon$$

şeklinde gösterilir. Burada x bağımsız değişken, $\beta' = [\beta_0, \beta_1]$ regresyon parametreleridir. Bağımlı değişken olarak adlandırılan y, 0 ve 1 değerlerini alan kategorik bir değişkendir. Ayrıca hatalara ait beklenen değer $E(\varepsilon) = 0$'dır.

Çizelge 2.1 y_i nin olasılık dağılımı

y	Olasılık
1	$P(y = 1) = \pi(x)$
0	$P(y = 0) = 1 - \pi(x)$

Bağımlı değişken y_i Bernoulli dağılımına sahip bir rasgele değişkendir. Bundan dolayı y nin gerçekleşmesi ya da gerçekleşmemesi Çizelge 2.1' deki gibi ifade edilir. Hatalara ait beklenen değer $E(\varepsilon) = 0$ olduğundan bağımlı değişkeninin beklenen değeri,

$$E(y) = 1[\pi(x)] + 0[1 - \pi(x)] = \pi(x) \text{ ve}$$

$$E(y) = \beta_0 + \beta_1 x$$

olarak ifade edilirken bağımlı değişkenin varyansı da,

$$Var(y) = \pi(x)[1 - \pi(x)]$$

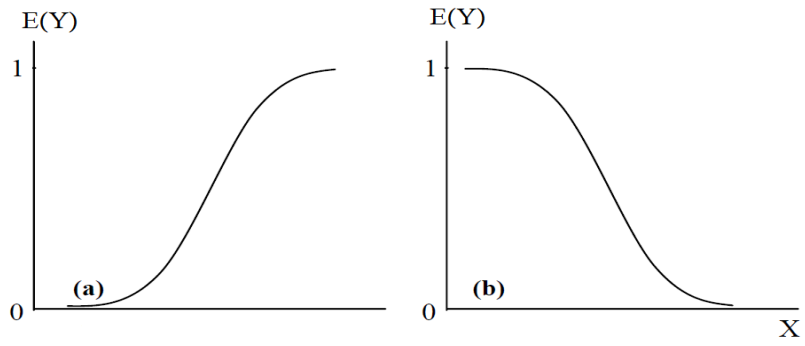
şeklinde ifade edilir.

Bağımlı değişken ve bağımsız değişkenler arasındaki ilişkinin doğrusal olmadığı durumlarda lojistik regresyon modeli tercih edilebilir.

Genel olarak, bağımlı değişkenin ikili olduğu durumlar için regresyon modelinin doğrusal olmadığına dair varsayımlar vardır. Tüm x bağımsız değişkenleri için $\pi(x)$ Şekil 2.1 'de gösterildiği gibi $0 \leq \pi(x) \leq 1$ arasında aldığı değerler ile monoton artan ya da azalan S biçimli (ya da ters S biçimli) bir yanıt fonksiyonunu verir (Agresti 2007). Bu yanıt fonksiyonuna lojistik yanıt fonksiyonu denir ve

$$\pi(x) = E(y) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \tag{2.1}$$

(2.1) ile verilen eşitlik ile ifade edilir.



Şekil 2.1 Yanıt fonksiyonu (Hosmer ve Lemeshow 2000)

Lojistik fonksiyon kolayca doğrusal hale getirilebilir. Doğrusal yanıt fonksiyonu,

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)}$$

dönüşümü ile tanımlanır. Bu dönüşüme $\pi(x)$ olasılığının logit dönüşümü denir ve dönüşümdeki $\pi(x)/\big(1 - \pi(x)\big)$ oranı odds oranı adını alır. Bazen logit dönüşüme log-odds denir. Odds oranı olayın gerçekleşme olasılığının gerçekleşmeme olasılığına oranı olarak ifade edilir.

2.1 Parametrelerin Tahmin Edilmesi

Lojistik regresyon modelinin genel biçimi,

$$y_i = E(y_i) + \varepsilon_i$$

olarak yazılır. Burada y_i eşitlik (2.2) de verilen beklenen değer ile bağımsız Bernoulli rasgele değişkenidir.

$$E(y_i) = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \tag{2.2}$$

$x_i'\beta'$ daki parametrelerin tahminini yapmak için en çok olabilirlik yöntemi kullanılır. Burada $(x_i, y_i),\ i = 1,2,\cdots,n$ olmak üzere y_i, i. gözleme ait bağımlı değişken ve x_i, i. gözleme ait bağımsız değişkendir.

Lojistik regresyonda denklemler doğrusal olmadığından $\hat{\beta}$ tahmin edicisini hesaplamak için özel yöntemler kullanılır. Bunlardan ilki en çok olabilirlik tahmin yöntemidir.

Her bir gözlem Bernoulli dağılımına sahiptir, dolayısıyla her bir örneklem gözleminin olasılık dağılımı,

$$f_i(y_i) = \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \quad i = 1,2,3,\ldots,n \tag{2.3}$$

olacaktır ve her y_i gözlemi 0 ve 1 değerini alacaktır. Gözlemler bağımsız olduğundan olabilirlik fonksiyonu:

$$\prod_{i=1}^{n} f_i(y_i) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \tag{2.4}$$

şeklinde yazılır. Eşitlik (2.5) ile verilen log-olabilirlik ile çalışmak daha uygundur.

$$L(\beta) = \ln \prod_{i=1}^{n} f_i(y_i) \tag{2.5}$$

$$= \sum_{i=1}^{n} l(y_i, \beta)$$

$$= \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\}$$

$$= \sum_{i=1}^{n} y_i \ln[\pi(x_i)] + \sum_{i=1}^{n} (1 - y_i) \ln[1 - \pi(x_i)]$$

$$= \sum_{i=1}^{n} \left[y_i \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right)\right] + \sum_{i=1}^{n} \ln[1 - \pi(x_i)]$$

$L(\beta)$' yı maksimize eden β değerini bulmak için, $L(\beta)$' nın β ' ya göre türevi alınır ve elde edilen ifadeler sıfıra eşitlenir.

$$\hat{\beta}_{MLE} = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} l(y_i, \beta)$$

Sırasıyla β_0 ve β_1 'e ait parametre tahmin değerleri,

$$\sum [y_i - \pi(x_i)] = 0 \tag{2.6}$$

$$\sum x_i [y_i - \pi(x_i)] = 0 \tag{2.7}$$

eşitlik (2.6) ve eşitlik (2.7) çözülerek elde edilir.

Lojistik regresyon modellerinde sık sık x değişkeninin her bir düzeyinde gözlemler ya da denemeler tekrar edilir. Bu çoğunlukla tasarlanmış deneylerde olur. y_i, *i.* gözlem için gözlenen 1'lerin sayısını temsil etsin ve n_i her bir gözlemde denemelerin sayısı olsun. Bu durumda log olabilirlik

$$ln\, L(y, \beta) = \sum_{i=1}^{n} y_i \ln(\pi_i) + \sum_{i=1}^{n} n_i \ln(1 - \pi_i) - \sum_{i=1}^{n} y_i \ln(1 - \pi_i)$$

$$= \sum_{i=1}^{n} y_i \ln(\pi_i) + \sum_{i=1}^{n} (n_i - y_i) \ln(1 - \pi_i)$$

biçiminde yazılır.

$\hat{\beta}$ tahmin edicisini hesaplamak için alternatif olarak, MLE sapma istatistiği β'ya göre minimize edilir (Ahmad vd. 2010). Yani,

$$d_i = \left[-y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) - (1-y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \tag{2.8}$$

$$\hat{\beta}_{MLE} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} d_i$$

olarak bulunur.

2.2 Parametre Tahminlerinin Yorumlanması

Lojistik regresyon modelinde parametreleri yorumlamak lineer regresyon modeline göre daha zordur. $\hat{\beta}$ parametresinin tahmin değerini yorumlamak için x de meydana gelen bir birim değişikliğin lojistik yanıt fonksiyonunu nasıl etkilediği araştırılır. Doğrusal yanıt fonksiyonunun tek bir bağımsız değişkene sahip olduğu durum göz önüne alındığında x 'in belirli bir x_i değeri için tahmin değeri,

$$\hat{g}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

olarak ifade edilir. $x_i + 1$' de elde edilen tahmin değeri ise,

$$\hat{g}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1 (x_i + 1)$$

olur ve iki değer arasındaki fark,

$$\hat{g}(x_i + 1) - \hat{g}(x_i) = \hat{\beta}_1$$

$\hat{\beta}_1$ in tahmin değerini verir. Bağımsız değişken x_i'ye eşit olduğunda $\hat{g}(x_i)$ ifadesi ve bağımsız değişken $x_i + 1$'e eşit olduğunda $\hat{g}(x_i + 1)$ ifadesi log-odds oranı verir. Bundan dolayı, tahmin edilen iki değer arasındaki fark,

$$\hat{g}(x_i + 1) - \hat{g}(x_i) = \ln\left(odds_{x_i+1}\right) - \ln\left(odds_{x_i}\right)$$

$$= \ln\left(\frac{odds_{x_i+1}}{odds_{x_i}}\right) = \hat{\beta}_1$$

olarak bulunur. Eğer antiloglar alınırsa odds oranı,

$$\hat{O}_R = \frac{odds_{x_i+1}}{odds_{x_i}} = e^{\hat{\beta}_1}$$

biçiminde elde edilir. Odds oranları, bağımsız değişken değerinde meydana gelen bir birimlik değişime karşılık olarak başarı olasılığında meydana gelen artıştır.

2.3 Parametrelerin Anlamlılık Testi

Literatürde parametrelerin anlamlılıklarını test etmek için birden fazla yöntem ileri sürülmüştür. Olabilirlik oran testi kullanılarak parametrelerin anlamlılıkları test edilebilir. Olabilirlik oran testi, tahmin edilen ve gözlenen modelin kıyaslanmasında kullanılır. Bağımlı değişkeninin gözlenen bir değeri, sabit başarı olasılığına sahip modelden alınır. Bu model doymuş model olarak adlandırılır. Tahmin edilen ve gözlenen değerlerin karşılaştırılmasında olabilirlik fonksiyonu kullanılır.

$$D = -2\,ln\frac{(tahmin\ edilen\ modelin\ olabilirliği)}{(doymuş\ modelin\ olabilirliği)}$$

$$D = \sum_{i=1}^{n} {d_i}^2 = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \qquad (2.9)$$

(2.9) ile verilen eşitlikte $\hat{\pi}_i = \hat{\pi}(x_i)$ dir ve D ile gösterilen bu test istatistiği McCullagh ve Nelder (1989) tarafından sapma olarak adlandırılmaktadır. Sapma istatistiğinin doğrusal regresyondaki karşılığı hata kareler toplamıdır.

Sapma istatistiği, olabilirlik oran testi olarak adlandırılan hipotez testinde kullanılır. Olabilirlik oran testi lojistik regresyonda regresyonun anlamlılığını test etmek için kullanılan yöntemlerden bir tanesidir. Bu test doymuş model olarak, sabit başarı olasılığına sahip bir modeli kullanır. Bu sabit başarı olasılığı,

$$E(y) = \pi = \frac{e^{\beta_0}}{1+e^{\beta_0}}$$

şeklinde verilen, yani bağımsız değişkeni olmayan bir lojistik regresyon modelidir. Bu testte, bağımsız değişkene sahip olan modelin sapması ile bağımsız değişkene sahip olmayan modelin sapması karşılaştırılır (Anderson 1990).

D değerinde meydana gelen bu farklılık *G* istatistiği olarak ifade edilmektedir. *G* istatistiği aşağıdaki gibi tanımlanabilir:

$$G = -2\ln\frac{(değişkensiz\ olabilirlik)}{(değişkenli\ olabilirlik)}$$

$$G = -2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_o}{n}\right)^{n_0}}{\prod_{i=1}^{n}\hat{\pi}_i{}^{y_i}\,(1-\hat{\pi}_i)^{(1-y_i)}}\right]$$

Bağımsız değişkenin modelde olmadığı durumda, β_0'ın en çok olabilirlik tahmini $\ln(n_1/n_0)$ dır. Burada, $n_1 = \sum y_i$, $n_o = \sum(1-y_i)$ ve tahmin edilen değer sabiti n_1/n 'dir.

$$G = 2\left\{\sum_{i=1}^{n}[y_i\ln(\hat{\pi}_i) + (1-y_i)\ln(1-\hat{\pi}_i)] - [n_1\ln(n_1) + n_0\ln(n_0) - n\ln(n)]\right\}$$

β_1'in sıfıra eşit olduğu hipotez için, *G* istatistiği, 1 serbestlik derecesi ile ki-kare dağılımına sahiptir. Ayrıca *G* istatistiği, bütün β katsayılarının anlamlılığını test etmek için kullanılabilir.

Parametre anlamlılığını test etmek için ikinci bir yöntem olarak Wald (1943) testi kullanılabilir. Wald testi, $\hat{\beta}_1$' in en çok olabilirlik tahminine ve bu tahmine ait standart hatanın karşılaştırılmasına dayanmaktadır. Bu model için test istatistiği:

$$W = \frac{\hat{\beta}_1}{\widehat{SE}(\hat{\beta}_1)}$$

olacak şekilde tanımlanır. W istatistiği standart normal dağılıma sahiptir. $\hat{\beta}_1$'in standart hatası, kovaryans matrisinin köşegen elemanlarının kareköklerinin alınması sonucu elde edilir. $\beta_1 = 0$ hipotezi için test istatistiğin yorumlanmasında standart normal dağılımdan yararlanılır (Hosmer ve Lemeshow 2000).

2.4 Güven Aralıkları

Lojistik regresyonda aralık tahmini yaparken, modelin anlamlılığını test ederken kullanılan metotlardan faydalanılır. Aralık tahmini için, Wald (1943) test istatistiği kullanılabilir.

β_1 regresyon katsayısı (regresyon eğim parametresi) için $\%100(1-\alpha)$ lık güven aralığı:

$$\hat{\beta}_1 \pm z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_1)$$

ve β_0 regresyon katsayısı (regresyon sabit terimi) için $\%100(1-\alpha)$ lık güven aralığı:

$$\hat{\beta}_0 \pm z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_0)$$

şeklinde ifade edilir. Burada $z_{1-\alpha/2}$, standart normal dağılımın $\%100(1-\alpha/2)$ lık değeridir ve $\widehat{SE}(.)$ ilgili parametreye ait tahmin edicinin standart hatasının model tabanlı bir tahminini gösterir.

3. ÇOK DEĞİŞKENLİ LOJİSTİK REGRESYON

Çok değişkenli lojistik regresyonda bağımsız değişken sayısı 1'den fazladır.

Çok değişkenli regresyon modeli genel olarak,

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

şeklinde gösterilir. Burada bağımsız değişken $x' = [1, x_1, x_2, \dots, x_p]$, regresyon parametreleri $\beta' = [\beta_0, \beta_1, \dots, \beta_p]$ ve bağımlı değişken y' de 0 ya da 1 değerini alır. Ayrıca hatalara ait beklenen değer $E(\varepsilon) = 0$'dır.

Çok değişkenli durum için lojistik yanıt fonksiyonu:

$$\pi(x) = E(y) = \frac{e^{g(x)}}{1 + e^{g(x)}} \tag{3.1}$$

$$= \frac{\exp(x'\beta)}{1 + \exp(x'\beta)} = \frac{1}{1 + \exp(-x'\beta)}$$

eşitlik (3.1) ile ifade edilir.

Genellikle, nominal ölçeklendirilmiş bir değişken k olası değere sahip olduğunda k-1 tane bağımsız değişkene ihtiyaç duyulur. Tüm regresyon modellerinde genel olarak bir tane sabit terim bulunur. j. bağımsız değişken x_j nin k_j tane seviyesi bulunmaktadır. $k_j - 1$ tane bağımsız değişken D_{ij} ve bu bağımsız değişkenlere ait katsayılar β_{ij} olarak ifade edilir. Böylece, p değişkenli bir modelin j. değişkene ait logit değeri

$$g(x) = \beta_0 + \beta_1 x_1 + \cdots + \sum_{i=1}^{k_j - 1} \beta_{ij} x_{ij} + \beta_p x_p \tag{3.2}$$

eşitlik (3.2) ile gösterilir.

3.1 Parametrelerin Tahmin Edilmesi

Örneklem büyüklüğü n olan veriye ait gözlemler $(x \quad , y \quad)$, $i = 1,2, \cdots, n$ şeklindedir. Tek değişkenli durumdakine benzer olarak, modelin yorumlanması için $\beta' = [\beta_0, \beta_1, \dots, \beta_p]$ vektörünün tahminlerinin elde edilmesi lazımdır. Çok değişkenli durumda parametre

tahmini yaparken, tek değişkenli durumdakine benzer olarak en çok olabilirlik tahmin yönteminden yararlanılır. Burada ki tek farklılık, olabilirlik fonksiyonunda eşitlik (3.1) ile ifade edilen $\pi(x)$ istatistiğinin kullanmasıdır.

Parametre tahmini için, log olabilirlik fonksiyonunun, p + 1 tane katsayıya göre türevini alarak elde edilen p + 1 tane olasılık denklemi olacaktır. Regresyon katsayılarına ait tahmin değerleri eşitlik (3.3) ve eşitlik (3.4) çözülerek elde edilir.

$$\sum_{i=1}^{n}[y_i - \pi(x_i)] = 0 \tag{3.3}$$

$$\sum_{i=1}^{n} x_{ij}[y_i - \pi(x_i)] = 0 \tag{3.4}$$

Burada $j = 1,2,\cdots,p$ dir.

Rao (1973), tarafından öne sürülen bir teoriye göre tahmin edilen β katsayılarına ait varyans ve kovaryansları bulmak için, gelişmiş bir en çok olabilirlik tahmininden yararlanılır. Bu teoriye göre tahmin ediciler, log olabilirlik fonksiyonunun ikinci kısmi türevinin alınması ile elde edilir. Kısmi türevler,

$$\frac{\partial^2 L(\beta)}{\partial {\beta_j}^2} = -\sum_{i=1}^{n} {x_{ij}}^2 \pi_i(1-\pi_i) \tag{3.5}$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n} x_{ij} x_{i1} \pi_i(1-\pi_i) \tag{3.6}$$

eşitlik (3.5) ve eşitlik (3.6) de ki gibi ifade edilir. Burada $j, l = 0,1,2,\ldots,p$ ve π_i ise $\pi(x_i)$ yi belirtmektedir. (3.5) ile verilen eşitlik ile oluşturulan $(p+1)$x$(p+1)$ boyutlu matris $I(\beta)$ ile gösterilir. Bu matrisin tersi alınarak varyans ve kovaryans matrisi edilir $\mathrm{Var}(\beta) = I^{-1}(\beta)$. β nın j. köşegen elemanı β_j ye ait varyans $\mathrm{Var}(\beta_j)$ ile gösterilirken, β_j ve β_l ye ilişkin kovaryans $\mathrm{Cov}(\beta_j, \beta_l)$ ile gösterilir. $\widehat{\mathrm{Var}}\,(\hat{\beta})$ ile gösterilen , $\hat{\beta}$ ya ait varyans ve kovaryanslar Var (β) den elde edilir.

$\hat{\beta}_j$ ye ait standart hata,

$$\widehat{\mathrm{SE}}(\hat{\beta}_j) = \left[\widehat{\mathrm{Var}}(\hat{\beta}_j)\right]^{1/2}$$

şeklinde ifade edilir. Burada $j = 0,1,2,\ldots,p$ dir.

Parametrelerin anlamlılığını test ederken $\hat{I}(\hat{\beta}) = \mathrm{X}'\mathrm{VX}$ matrisi kullanılacaktır. Burada X her bir x gözlemine ait n x $(p+1)$boyutlu bir matristir ve V de köşegen elemanları $\hat{\pi}_i(1-\hat{\pi}_i)$ den oluşan n x n boyutlu diyagonal bir matristir.

$$\mathrm{X} = \begin{bmatrix} 1 & x_{11} & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix}$$

$$\mathrm{V} = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \ldots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \ldots & 0 & \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}$$

3.2 Parametrelerin Anlamlılık Testi

Parametrelerin anlamlılığını test etmek için birçok model öne sürülmüştür. İlk olarak lojistik regresyon modelinde bulunan parametrelerin anlamlılığını test etmek için, olabilirlik oran testi kullanılabilir. Bu test doymuş model olarak, sabit başarı olasılığına sahip bir modeli kullanır. Sabit başarı olasılığı y_i/n_i' dir. Burada y_i başarıların sayısı ve n_i gözlemlerin sayısıdır. Bu yöntem için kullanılacak test istatistiği, eşitlik (3.7) ile verilir.

$$D(\beta) = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i(1-\hat{\pi}_i)}\right)\right] \tag{3.7}$$

Bu istatistiği hesaplarken eğer $y = 0$ ise $y\ln(y/n\hat{\pi}) = 0$ ve $y = n$ ise $(n-y)\ln[(n-y)/n(1-\hat{\pi})] = 0$ olur. Lojistik regresyon modeli verilerle yeterli bir uyum sağladığında ve örneklem genişliği büyük olduğunda sapma $n-p$ serbestlik dereceli ki-kare dağılımına sahiptir; burada p modeldeki parametre sayısıdır. Sapma istatistiğinin büyük değerleri tahmin edilen modelin uygun olmadığını belirtirken, küçük değerleri (ya da büyük bir p-değeri) modelin verilere uyum sağladığı anlamına gelir. Geçerli ve pratik bir kural, sapmayı kendi serbestlik derecesi sayısına bölmektir.

Eğer $D(\beta)/(n-p)$ oranı, birden büyükse tahmin edilen model verilere yeterli bir uyum sağlamamaktadır.

Sapma istatistiğinin, doğrusal regresyonda benzeri vardır. Doğrusal regresyon modelinde $D(\beta) = SS_{Res}/\sigma^2$ dir. Bu değer, eğer gözlemler normal ve bağımsız dağılmışsa $n-p$ serbestlik derecesi ile ki-kare dağılımına sahiptir. Bununla birlikte, doğrusal regresyonda sapma, bilinmeyen parametre σ^2 ye sahip olduğu için doğrudan hesaplanamayabilir.

Uyum iyiliği, her bir gözlemde başarı ve başarısızlığın gözlenen ve beklenen olasılıklarını karşılaştıran Pearson ki-kare ile de incelenebilir. Başarılar beklenen sayısı $n_i\hat{\pi}_i$ ve başarısızlıkların beklenen sayısı $n_i(1-\hat{\pi}_i)$ 'dir $(i = 1,2,\dots,n)$. Pearson ki-kare istatistiği,

$$\chi^2 = \sum_{i=1}^{n}\left\{\frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i} + \frac{[(n_i - y_i) - n_i(1-\hat{\pi}_i)]^2}{n_i(1-\hat{\pi}_i)}\right\} \tag{3.8}$$

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

eşitlik (3.8) ile verilir (McCullagh ve Nelder 1989). Pearson ki-kare uyum iyiliği istatistiği $\mathrm{n}-\mathrm{p}$ serbestlik dereceli ki-kare dağılımına sahiptir. İstatistiğin küçük değerleri (ya da büyük p değeri) modelin verilere uyum sağladığı anlamına gelir. Pearson ki-kare istatistiği $\mathrm{n}-\mathrm{p}$ serbestlik derecesi sayısına bölünebilir ve oran 1 ile karşılaştırılır. Eğer oran 1 i çok aşarsa modelin uyum iyiliği sorgulanır.

Bağımsız değişkende tekrarlar olmadığı zaman gözlemler, Hosmer-Lemeshow (1980) testi denilen bir uyum iyiliği testi için gruplanabilir. Bu yöntemde gözlemler tahmin edilen başarı olasılıklarına dayalı olarak g grupta sınıflandırılır. Genel olarak, yaklaşık 10 grup kullanılır ve gözlenen başarılar sayısı O_j ve başarısızlıklar $N_j - O_j$, her bir gruptaki beklenen frekans olan $N_j\bar{\pi}_j$ ve $N_j(1-\bar{\pi}_j)$ ile karşılaştırılır. Burada N_j, j. gruptaki gözlemlerin sayısıdır ve j. grupta tahmin edilen ortalama başarı olasılığı $\bar{\pi}_j = \sum_{i\,\epsilon\,grup\;j}\hat{\pi}_j/N_j$'dir. Hosmer – Lemeshow (1980) istatistiği gözlenen ve beklenen frekanslarla Pearson ki-kare uyum iyiliği istatistiğidir:

$$HL = \sum_{j=1}^{g} \frac{(O_j - N_j\bar{\pi}_j)^2}{N_j\bar{\pi}_j(1-\bar{\pi}_j)}$$

Eğer tahmin edilen lojistik regresyon modeli doğru ise HL istatistiği, $g-2$ serbestlik dereceli ki-kare dağılımına sahiptir. HL istatistiğinin sonucunun büyük değerlere sahip olması, modelin verilere yeterli uyum sağlamadığına işaret eder.

3.3 Model Katsayısı Üzerine Testler

Tek tek model katsayıları için,

$$H_0 : \beta_j = 0, \quad H_1 : \beta_j \neq 0$$

gibi hipotezlerin testleri, sapma istatistiğine bakılarak test edilebilir. Aynı zamanda en çok olabilirlik tahminine dayalı bir yaklaşım daha vardır. Büyük örneklemler için en çok olabilirlik tahmin edicisinin dağılımı, küçük bir yanla ya da yan olmadan yaklaşık olarak normaldir. Ayrıca, en çok olabilirlik tahmin edicisinin varyans ve kovaryansaları, en çok olabilirlik tahmin yöntemi ile hesaplanan model parametrelerine göre log olabilirlik fonksiyonunun ikinci kısmi türevlerinden bulunabilir. O zaman yukarıdaki hipotezleri test etmek için t istatistiğine benzer bir istatistik kullanılabilir. Bu test istatistiği, Wald (1943) testi olarak bilinir.

I, log olabilirlik fonksiyonunun ikinci kısmi türevlerinin p× p boyutlu matrisini göstersin:

$$\mathrm{I}_{il} = \frac{\partial^2 L(\beta)}{\partial\beta_i \partial\beta_l}, \quad i,l = 0,1,\ldots,p$$

I'ye Hessian matris denir. Hessian matrisinin elemanlarını oluşturan $\beta = \hat{\beta}$ katsayıları, en çok olabilirlik tahmin yönteminden yararlanarak hesaplanır. Regresyon parametrelerine ait örneklem kovaryans matrisi,

$$\mathrm{Var}(\hat{\beta}) = -\mathrm{I}(\hat{\beta})^{-1} = (\mathrm{X}'\mathrm{V}\mathrm{X})^{-1}$$

şeklinde ifade edilir. Bu kovaryans matrisinin köşegen elemanlarının karekökleri, regresyon parametrelerinin örneklem standart hatalarıdır. Böylece,

$$H_0 : \beta_j = 0, \quad H_1 : \beta_j \neq 0$$

hipotezi için test istatistiği,

$$W = Z_0 = \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)}$$

biçiminde olacaktır. Bu istatistiğe, Wald test istatistiği denir. Bu istatistiğin yorumlanması için standart normal dağılımdan yararlanılır (Hosmer ve Lemeshow 2000). Bazı bilgisayar paket programları Z_0 istatistiğinin karesini alır ve onu 1 serbestlik dereceli ki-kare dağılımıyla karşılaştırır.

3.4 Güven Aralıkları

Lojistik regresyonda güven aralıklarını oluşturmak için Wald (1943) test istatistiğini kullanmak mümkündür. Doğrusal yanıt fonksiyonundaki tek tek regresyon parametreleri için güven aralıkları bulunur. j. model katsayısı için yaklaşık $\%100(1-\alpha)$ güven aralığı aşağıda verilmiştir:

$$\hat{\beta}_j - Z_{\alpha/2}\widehat{SE}(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + Z_{\alpha/2}\widehat{SE}(\hat{\beta}_j)$$

Regresyon katsayısı β_j aynı zamanda odds oranının logaritmasıdır. β_j için güven aralığının nasıl bulunacağı bilindiğinden odds oranı için bir güven aralığı bulmak kolaydır. Odds oranının nokta tahmini $O_R = exp(\hat{\beta}_j)$ 'dir ve odds oranı için $\%100(1-\alpha)$ güven aralığı da aşağıdaki gibidir:

$$\exp[\hat{\beta}_j - Z_{\alpha/2}\widehat{SE}(\hat{\beta}_j)] \leq O_R \leq \exp[\hat{\beta}_j + Z_{\alpha/2}\widehat{SE}(\hat{\beta}_j)]$$

Odds oranı için güven aralığı, nokta tahmini yaparken genellikle simetrik dağılmaz. $\hat{O}_R = \exp(\hat{\beta}_j)$ nokta tahmini, aslında O_R'nin örneklem medyanını tahmin eder.

4. ROBUST REGRESYON

İstatistiksel yöntemlerin hepsi varsayımlar içermektedir. Verilerin normal dağılımdan gelmesi en çok istenen varsayımlardan bir tanesidir. Teorik olarak uygun olmasına rağmen gerçek verilerle çalışıldığında bu şartın sağlanması oldukça zor olabilir. Veride aykırı gözlemler olabilir. Bu tip sorunları çözmek için de robust yöntemlere başvurulur.

4.1 Aykırı Gözlem Problemi

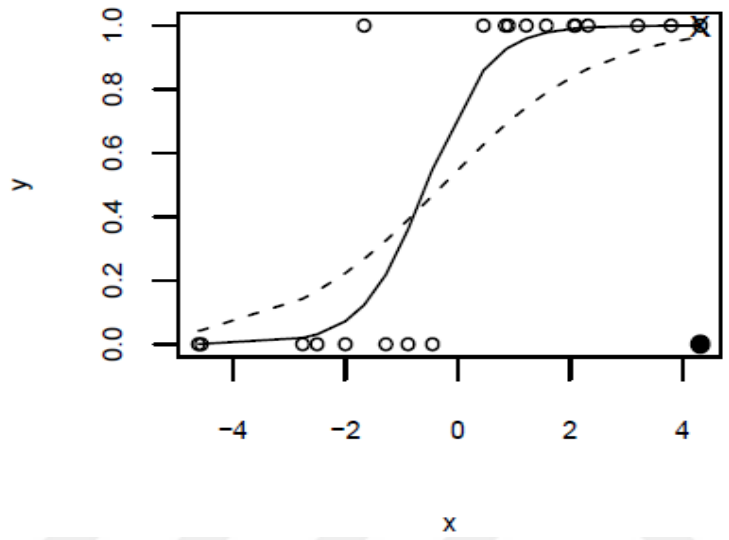
Lojistik regresyonda aykırı gözlemlerin farklı durumlarını ayırt etmek önemlidir. Lojistik modelde, farklılıklar Y-yönünde, X-yönünde veya her iki alanda da oluşabilir. Bağımlı değişkenin tek olduğu durumlar da, tüm *y*'ler 0 veya 1'dir. Bu nedenle *y* yönündeki bir hata yalnızca $0 \rightarrow 1$ veya $1 \rightarrow 0$ geçişi olarak gerçekleşebilir (Copas, 1988). Bu tür aykırı değerler ayrıca artık aykırı değerler veya yanlış sınıflandırma hatası olarak da bilinir. X yönünde ki aykırı gözlemler genellikle kaldıraç noktası (leverage point) olarak adlandırılır. X yönünde bulunan bu kaldıraç noktaları iyi ya da kötü olarak gruplandırılır. İyi bir kaldıraç noktası, $P\ (Y\ = 1|x\ )$ in büyük bir değerine sahip olduğunda $Y\ = 1$ ya da $P\ (Y\ = 1|x\ )$ nin küçük bir değerine sahipse $Y\ = 0$ şeklindedir. Kötü bir kaldıraç noktası ise bunun tam tersidir. Victoria-Feser (2002), MLE' nin X yönünde ki aykırı değerlerden etkilenebileceğini ve Pregibon (1982) ve Copas (1988) ise aykırı değerlerin yanlış sınıflandırmaya sebep olabileceğini göstermiştir. Croux vd. (2002), kötü kaldıraç noktaları şeklinde isimlendirilen en tehlikeli aykırı değerlerin, aynı zamanda x değişkenlerinin tasarım alanında yanlış sınıflandırılmış gözlemler olduğunu bulmuştur.

Bu tezde x yönünde aykırı değer olduğu durumlar ele alınarak, bu durumdan etkilenmeyen robust yöntemlerle parametre tahminleri yapılacaktır.

4.2 Lojistik Regresyonda Kullanılan Robust Tahmin Yöntemleri

Veri basit rastgele örnekleme ile elde edilemediğinde, standart lojistik regresyon geçerli değildir. Veriler tabakalama, kümeleme ve / veya eşit olmayan ağırlıklandırma ile yapılan karmaşık bir araştırma tasarımından geldiğinde, olağan tahminler uygun değildir (Rao ve Scott 1984). Bu durumlarda uygun tahminleri ve standart hataları

bulmak için özel teknikler uygulanmalıdır. Kategorik sonuçlu regresyon modellerinde parametrelerin tahmininde en çok olabilirlik yöntemi yaygın olarak kullanılır.Bu yöntemin bir dezavantajı aykırı değerlere karşı çok duyarlı olmasıdır (Şekil 4.1). Düz çizgi orijinal veriden elde edilen MLE, kesikli çizgi ise bir tane aykırı değer (sağ köşedeki siyah daire) olan veriden elde edilen MLE'yi gösterir.



Şekil 4.1 Aykırı değerlerden etkilenen MLE örneği (Simeckova 2005)

Pregibon (1981), lojistik regresyonda parametre tahminlerinin aykırı değerlerden ciddi şekilde etkilenebileceğini belirmiştir. Bu nedenle aykırı değer sorununu çözmek için, aykırı değerlerden çok daha az etkilenen MLE' nin birçok robust alternatifi literatürde önerilmiştir (Pregibon 1981, Copas 1988, Kunsch vd. 1989, Carroll ve Pederson 1993, Bianco ve Yohai 1996, Croux ve Haesbroeck 2003). Bu robust yöntemlerden bazıları takip eden bölümlerde incelenecektir.

4.2.1 Ağırlıklandırılmış en çok olabilirlik tahmin edicisi (WMLE)

Bu yöntem, tahminlerin aykırı değerlerden en az etkilenmesi sağlamak için kullanılır. Bu nedenle, veri setinin bozulmasına sebep olan aykırı değerler tespit edilerek bu değerlerin ağırlığını sıfıra eşitlenir.

$Y_1, \dots, Y_n$ gözlemlerinin bağımsız olduğu ve Y_i'nin yoğunluk fonksiyonun $f_i(y_i; \beta)$ olduğu kabul edilesin. $y = (y_1, \dots y_n)'$, $(Y_1, \dots, Y_n)$ e ait gözlemleri içeren vektör ve

$w = (w_1, \ldots, w_n)$ ise ağırlıkları içeren vektör olmak üzere, en çok olabilirlik tahmin edicisi

$$l_i(\beta) = \ln f_i(y_i; \beta)$$

şeklinde ifade edilir. Ağırlıklandırılmış en çok olabilirlik tahmin edicisi (WMLE),

$$l(\beta) = \sum_{i=1}^{n} w_i . l_i(\beta) \tag{4.1}$$

$$= \sum_{i=1}^{n} w_i \{y_i x_i' \beta - \ln[1 + exp(x_i' \beta)]\}$$

$l(\beta)$ fonksiyonunun β' ya göre minimum yapılması ile bulunur. Burada $\beta \in R^p$ ' dir.

Carroll ve Pederson (1993), yüksek kaldıraç noktaları içeren verilerde, bu noktaların parametre tahminine olan olumsuz etkisini zayıflatmak amacı ile MLE' yi ağırlıkla sınırlandırmayı önermiştir. Bir başka deyişle, aykırı gözlemlere denk gelecek olan ağırlıkları sıfır seçerek bu gözlemlerin parametre tahminindeki etkisini en aza indirmeyi amaçlamışlardır.

Robust mahalanobis uzaklığını hesaplamak için en küçük kovaryans determinantından (MCD) yararlanılmıştır. MCD yaklaşımı Rousseeuw (1984, 1985) tarafından önerilmiştir. Bu yaklaşımda amaç, aykırı değer kabul edilmeyen gözlemleri (h) bularak bu gözlemlere ait örneklem ortalamasını ve kovaryansını hesaplamaktır. Başka bir deyişle, n büyüklüğüne sahip bir örneklem için, h'nin $n / 2$ ile n arasında olduğu bir h alt kümesi oluşturulup işlemler yapılmaktadır (Rousseeuw ve Van Zomeren 1990). Bu temel altküme civarında ki-kare dağılımına göre kritik uzaklık tespit edilir ve bu alanın dışında kalan gözlemler 0 ile ağırlıklandırılır. Aykırı değerlerin etkisi azaldığı için örneklem merkezine konum ve değer olarak yakın olan gözlemler tahmin yapmak için kullanılmış olunur. $x_1, \ldots, x_n$ gözlemlerine ait konum ve kovaryans matrislerinin robust tahmin edicileri sırasıyla $\hat{\mu}_{MCD}$ ve $\hat{\Sigma}_{MCD}$ olmak üzere,

$$Rd_i^{\,2} = h(x) = ((x - \hat{\mu}_{MCD})' \hat{\Sigma}_{MCD}^{\,-1} (x - \hat{\mu}_{MCD}))^{1/2} \tag{4.2}$$

eşitliği ile x_i gözlemlerine ait robust mahalanobis uzaklığı hesaplanır. Eşitlik (4.2) sayesinde, aykırı değerlerinden oldukça az etkilenen alt kümenin etrafında ki-kare dağılımına bakılarak kritik uzaklık tespiti yapılır ve ağırlıklar belirlenir.

$$w_i = I(Rd_i{}^2 \leq \chi^2{}_{p,0.95})$$

Bu tezde Rousseeuw'un (1984) En Küçük Kovaryans Determinantı (MCD) ile hesaplanan robust uzaklıklar kullanılmıştır. Bu algoritma, S-Plus (cov.mcd fonksiyonu olarak) paketine dahil edilmiştir ve çıktılarında robust uzaklıklardan yararlanılmıştır. Eşitlik (4.1) de ki uzaklık hesaplanırken robust konum vektörü ve robust kovaryans matrisi olarak MCD'den elde edilen değerler kullanılmıştır.

4.2.2 Ağırlıklandırılmış Bianco ve Yohai tahmin edicisi (WBYE)

Pregibon (1982), sapma istatistiğini minimize etmeyi amaçlayan, sapma istatistiğine dayalı robust bir tahmin yöntemi önermiştir.

$$\beta = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \lambda(d_i)$$

Burada λ artan bir Huber kayıp fonksiyonudur. Bu tahmin edici, aykırı gözlemlere daha az ağırlık vermek üzere tasarlanmıştır, ancak bu tahmin edici x-yönündeki aykırı gözlemleri zayıflatamamıştır ve tutarlı değildir. Bianco ve Yohai (1996),

$$\beta = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\big[(d_i) + g\big(\pi(x_i)\big) + g\big(1 - \pi(x_i)\big)\big] \qquad (4.3)$$

eşitlik (4.3) ile verilen tahmin ediciyi tanımlayarak Pregibon'un tahmin edicisinden daha tutarlı ve daha robust olan bir yöntem geliştirmiştir.

Bianco ve Yohai (1996) tarafından seçilen ρ,

$$\rho(x) = f(x) = \begin{cases} x - (x^2/2k), & x \leq k\ ise \\ k/2\,, & aksi\ halde \end{cases}$$

ile tanımlanan sınırlandırılmış, türevlenebilir ve azalan bir fonksiyondur. k pozitif bir sayıdır. $g(x) = \int_0^x \psi(-\ln u)\, du$ ve $\psi(x) = \rho'(x)$ şeklindedir.

Croux ve Haesbroeck (2003) Huber kayıp fonksiyonu ile çalışırken, daha önce Bianco ve Yohai (1996) tarafından önerilen $\rho\ (x)$ 'nun, sık sık ortaya çıkan bozulmamış veriler için bile bulunmadığını fark etmişler. Bu nedenle, Croux ve Haesbroeck (2003) yüksek kaldıraç noktalarını düşürmek için Bianco Yohai (BY) tahmin edicisine ekstra bir ağırlık ekleyerek genişletmeyi önermiştir. Croux ve Haesbroeck tahmin edicisi olarak da adlandırılan bu ağırlıklı BY (WBYE) tahmin edicisi eşitlik (4.4) ile tanımlanır.

$$\beta = \underset{\beta}{argmin} \sum_{i=1}^{n} w(x_i)\{\rho[(d_i) + g(\pi(x_i)) + g(1-\pi(x_i))]\} \tag{4.4}$$

Burada ağırlık $w(x_i)$, robust Mahalanobis mesafelerinin azalan bir fonksiyonu olan MCD kullanılarak hesaplanan mesafelerdir (Rousseeuw ve Leroy 1987).

$$w(x_i) = \begin{cases} 1, & Rd_i{}^2 \leq \chi^2{}_{p,0.975} \\ 0, & değilse \end{cases}$$

Bu WBYE tutarlıdır çünkü ağırlık sadece x değişkenlerine bağlıdır.

Bu tezde Bianco ve Yohai (1996) tarafından öne sürülen ağırlıklı BY tahmin edicisini bulmak için R da bulunan glmrob fonksiyonu kullanılmıştır. Bu fonksiyon, robust yöntemlerden faydalanılarak, genelleştirilmiş doğrusal modeller elde etmek için kullanmıştır. Bu fonksiyonda metot olarak WBYE seçilmiştir. Bu da robust bir Bianco-Yohai tahmin edicisinin hesaplanması için R programında Bylogreg adı verilen bir fonksiyonu çalıştmıştır.

4.2.3 Mallows ağırlığına göre ağırlıklandırılmış tahmin edici (Mallows)

Mallows tipi tahmin edici, ağırlıkların bağımsız değişkene bağlı olduğu ağırlıklı log-olabilirlik fonksiyonu en aza indirilerek elde edilir. Carrol ve Pederson (1993) Mallows tipi tahmin ediciyi araştırmışlar ve X uzayındaki aykırı değerlerin ağırlığını azaltarak MLE'yi sınırlı etkiye sahip bir tahmine dönüştürmeyi önermişlerdir.

Mallows tahmin edicisi, belirli bir ağırlık kullanılarak log-olabilirlik fonksiyonunun en aza indirilmesi ile elde edilir.

Eşitlik (2.2) de verilen lojistik model için robust bir β tahmini,

$$\sum_{i=1}^{n} w_i x_i [y_i - \pi(x_i) - c(x_i)] = 0 \qquad (4.5)$$

eşitlik (4.5) ile verilen ifadenin çözülmesi ile elde edilir (Carroll and Pederson 1993) . Burada $c(x_i)$ tutarlılığı sağlamak için verilen bir düzeltme fonksiyonudur. Ağırlıklar y_i, x_i ya da her ikisine bağlı olabilir. $w_i = w(x_i, x_i'\beta)$ ve $c(x_i) = 0$ ise, ağırlıklar sadece bağımsız değişkene bağlıdır ve tahmin ediciye Mallows sınıfı denir. Bu nedenle bu tahmin edici, ağırlıklı en çok olabilirlik tahmin edicisini temsil eder. Stefanski (1985), x yönündeki ağırlıkları hesaplamak için, robust kovaryans matrisi kullanılarak bulunan Mahalonobis uzaklığından yararlanmayı önermiştir. Bu uzaklık:

$$h_n(x) = ((x - \hat{\mu}_n)' \hat{\Sigma}_n^{-1} (x - \hat{\mu}_n))^{1/2}$$

şeklindedir. Burada $\hat{\mu}_n$ ve $\hat{\Sigma}_n$ sırasıyla $x_1, \ldots, x_n$ gözlemlerine ait konum ve kovaryans matrislerinin robust tahmin edicileridir.

$$\sum_{i=1}^{n} w_i \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \qquad (4.6)$$

Robust tahmin edici eşitlik (4.6) in minimize edilmesi ile bulunur.

Burada, $w_i = W\big(h_n(x_i)\big)$ ağırlık fonksiyonudur. W, $W(u)$ parametresine bağlı olarak sınırlandırılmış ve artmayan bir fonksiyondur. Carroll ve Pederson (1993), c> 0 parametresine bağlı olarak

$$W(u) = \left(1 - \frac{u^2}{c^2}\right)^3 \; I(|u| \leq c)$$

biçiminde ifade edilen bir W seçmeyi önermiştir.

Bu tezde MCD tahmin edicisi kullanılarak hesaplanan robust uzaklıklar kullanılmıştır. Mallows tipi robust tahmin edicilerden yararlanarak lojistik regresyon modeline uygun tahminler yapılmıştır. Bu algoritma için R da bulunan glmrob fonksiyonu kullanılmış ve Mqle metot olarak seçilmiştir.

5. NÜMERİK ÇALIŞMALAR

Parametre tahminini yaparken x- yönünde aykırı değer olması durumunda parametre tahmininde meydana gelecek olacak değişmeleri incelemek ve aykırı değer varlığında hangi tahmin yönteminin daha robust olduğunu tespit etmek amacıyla simülasyon çalışması ve gerçek verilerle bir çalışma yapılmıştır. Bu tahmin yöntemleri şunlardır. En çok olabilirlik tahmin yöntemi, ağırlıklandırılmış en çok olabilirlik tahmin yöntemi, ağırlıklandırılmış Bianco-Yohai tahmin yöntemi ve ağırlıklandırılmış Mallows tahmin yöntemidir.

5.1 Simülasyon Çalışması

Simülasyon çalışmasında yöntemlerin birbirleri ile karşılaştırılması için kullanılacak verilerin üretimi için lojistik regresyon modeli kullanılmıştır. Verilerin türetilme şekli aşağıda gösterilmiştir:

- Gerçek parametre değerleri olarak sabit bir $\beta = (1,1)$ değeri seçildi.
- Bütün yöntemlerin özelliklerini incelemek için homojen olmayan veri seti oluşturularak bağımsız değişkenler sırayla %1, %2, %3, %4 ve %5 oranında bozulmaya uğratıldı. Bozulmaya uğramamış olan veri seti %0 olarak gösterildi.
- Veri setinde bozulmaya uğramayan n_1 tane bağımsız değişken 0 ortalamalı 1 varyanslı standart normal dağılımdan üretildi, $x_i \sim N(0,1)$.
- Veri setinde bozulmaya neden olan n_2 tane bağımsız değişken 100 ortalamalı 1 varyanslı normal dağılımdan üretildi, $x_i \sim N(100,1)$.
- $n = n_1 + n_2$ tane hata terimi ε_i 'ler ise, konum parametresi 0 ve ölçek parametresi 1 olan bir lojistik dağılımdan rasgele üretildi, $\varepsilon_i \sim \text{Logistic}\,(0,1)$
- n tane bağımlı değişken ise,

$$y_i = \begin{cases} 0, & x_i'\beta + \varepsilon_i \leq 0 \\ 1, & x_i'\beta + \varepsilon_i > 0 \end{cases}$$

  olacak şekilde oluşturuldu.

- Simülasyon çalışması için örneklem büyüklüğü sırasıyla n=100, 200, 300, 400 ve 500 olarak seçildi.

Dört yöntem içinde yukarda belirtilen süreçler uygulandı. Her simülasyon çalışması 1000 kez tekrar edildi. Dört yöntemin performansı, yan (bias) ve hata kareler ortalamasına (MSE) bakılarak değerlendirildi. Her parametre için yan ve hata kareler ortalaması sırasıyla:

$$Bias = \frac{1}{1000}\left\| \sum_{i=1}^{1000} \hat{\beta}_i - \beta \right\| \tag{5.1}$$

ve

$$MSE = \frac{1}{1000}\left(\sum_{\text{İ}=1}^{1000} \left\| \hat{\beta}_i - \beta \right\|^2 \right) \tag{5.2}$$

eşitlik (5.1) ve eşitlik (5.2) den hesaplanır. Burada $\|.\|$ öklid normunu göstermektedir (Croux ve Haesbroeck, 2003).

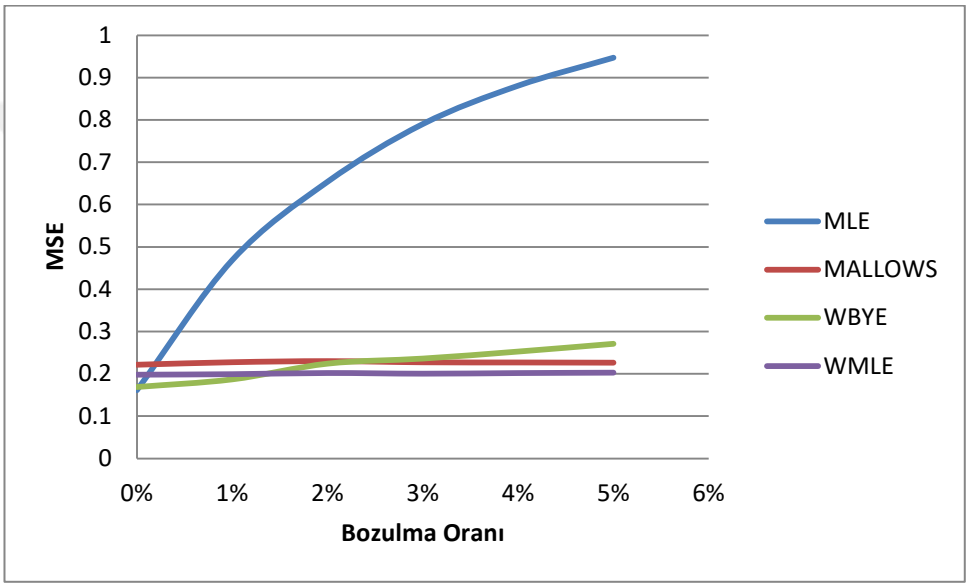
Dört tahminin yan ve MSE' si Çizelge 5.1'de gösterilmektedir. İyi bir tahmin edici, nispeten küçük veya sıfıra yakın bir yana ve MSE' ye sahip olandır. Aykırı değerler ile bozulmanın olmadığı verilerde, dört tahmin edicinin hepsinin yan ve MSE değerleri birbirine oldukça yakın olduğu görülmektedir.

Çizelge 5.1 n=100 olduğu durumda tahmin edicilerin Yan ve MSE oranları

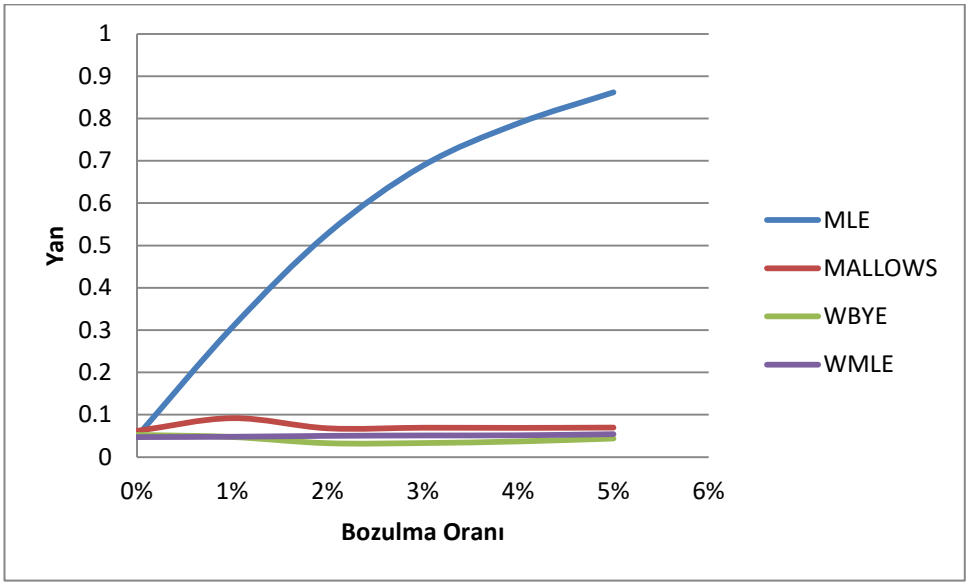
		MLE	MALLOWS	WBYE	WMLE
%0	Yan	0.04852377	0.06177093	0.05233679	0.04711922
	MSE	0.1613389	0.2214444	0.168816	0.1979205
%1	Yan	0.3066143	0.0918703	0.04735668	0.04797145
	MSE	0.4685382	0.2274064	0.1868893	0.1991236
%2	Yan	0.5276827	0.0679307	0.03272833	0.05005479
	MSE	0.6538023	0.2299855	0.2240265	0.2018755
%3	Yan	0.6883467	0.06932172	0.0330736	0.05097963
	MSE	0.7902735	0.2272044	0.236036	0.2003275
%4	Yan	0.7884755	0.06895091	0.0369085	0.0513654
	MSE	0.8808219	0.2267937	0.2525883	0.2017632
%5	Yan	0.8618038	0.06961693	0.04378603	0.05410475
	MSE	0.9469481	0.2262938	0.2709066	0.2027326

Çizelge 5.1 örneklem büyüklüğü 100 seçildiği zaman aykırı değerlerin meydana getirdiği farklı bozulma oranları karşısında 4 yöntemde meydan gelen değişmeleri göstermektedir. Bu 4 yönteme ait MSE ve yan değerleri Çizelge 5.1'de ki gibi hesaplanmıştır. Çizelge 5.2' de görüldüğü gibi MLE yöntemi, meydana gelen %1 oranında bozulmadan hızlıca etkilenmiştir. Bu veriler ve örneklem büyüklüğü için WMLE yöntemi MSE değerine bakarak en iyi tahmin yöntemi olarak seçilebilir. Çünkü bozulma oranlarındaki değişikliğe rağmen, en küçük MSE değerine sahiptir.

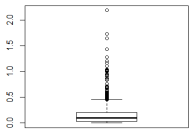
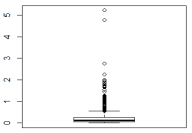
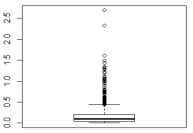
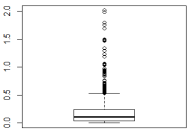
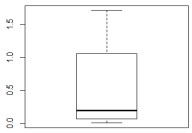
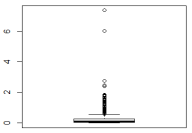
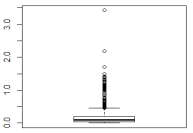
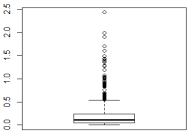
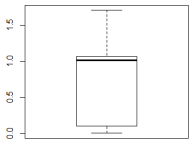
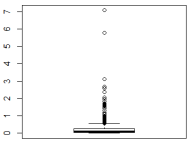
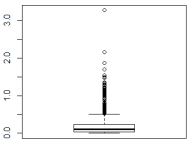
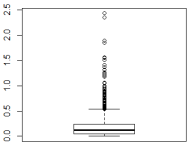
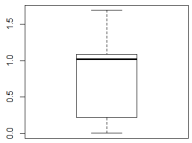
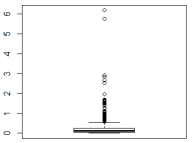
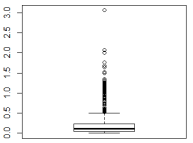
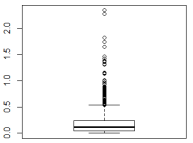
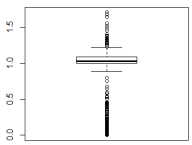
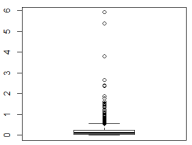
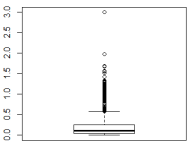
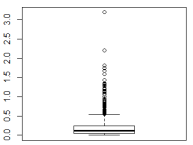
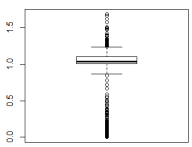
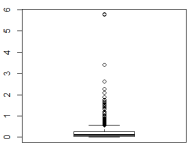
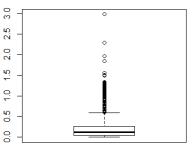
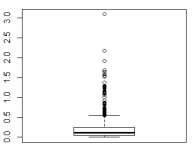
Çizelge 5.2 n=100 olduğunda tahmin edicilere ait MSE'ler



Çizelge 5.3 n=100 olduğunda tahmin edicilere ait Yan'lar



Çizelge 5.4 n=100 için bozulma oranlarına göre box- plot çizimleri

	MLE	MALLOWS	WBYE	WMLE
%0				
%1				
%2				
%3				
%4				
%5				

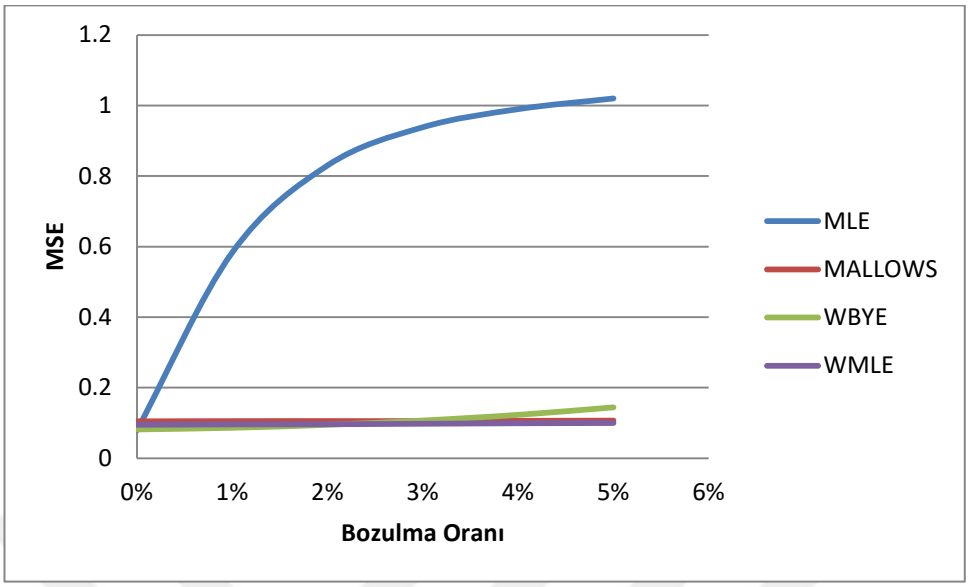
Çizelge 5.4 örneklem büyüklüğü 100 seçildiği durumda, MSE değerini hesaplamak için kullanılan uzaklıklara ait box-plot çizimlerini göstermektedir. Bu uzaklıkların oluşturduğu box-plot grafikleri aykırı değer olmadan önce 0 civarında bir dağılım göstermektedir. Aykırı değer olmadığı durumda en iyi dağılımı MLE gösterirken, aykırı değerlerin meydana getirdiği %1 oranında bir bozulmadan MLE yöntemi hızlıca etkilenmiştir. Diğer 3 tahmin yöntemi bu aykırı değerlerden önemli ölçüde etkilenmemiştir.

Çizelge 5.5 n=200 olduğu durumda tahmin edicilerin Yan ve MSE oranları

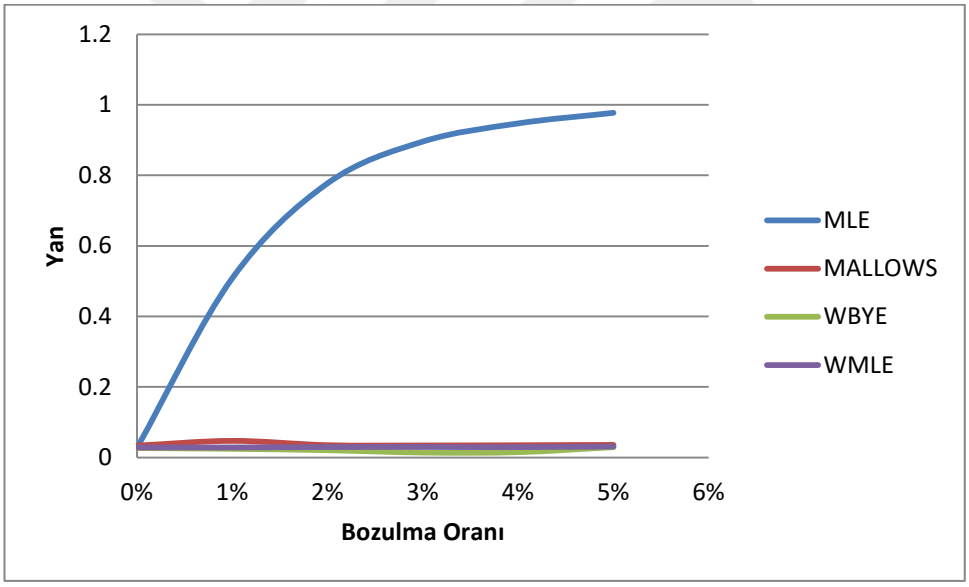
		MLE	MALLOWS	WBYE	WMLE
%0	Yan	0.02692466	0.03344315	0.02667738	0.0282067
	MSE	0.07890039	0.104466	0.08146216	0.09465652
%1	Yan	0.507991	0.04700985	0.0244034	0.02830118
	MSE	0.5832058	0.1050472	0.08617274	0.09575514
%2	Yan	0.7773063	0.03472052	0.02049548	0.02946133
	MSE	0.8300336	0.1052311	0.09462762	0.09628042
%3	Yan	0.8956775	0.033987	0.01382878	0.02973195
	MSE	0.938438	0.1050652	0.1067736	0.09756991
%4	Yan	0.9471713	0.03458098	0.01494296	0.02930093
	MSE	0.9896905	0.1058631	0.1228337	0.09896727
%5	Yan	0.9771021	0.03558627	0.02917479	0.03083677
	MSE	1.01868	0.1064654	0.1441057	0.09970036

Çizelge 5.5 örneklem büyüklüğü 200 seçildiği zaman aykırı değerlerin meydana getirdiği farklı bozulma oranları karşısında 4 yöntemde meydan gelen değişmeleri göstermektedir. Bu veriler için en uygun tahmin yöntemi WMLE seçilebilir. Çünkü bozulma oranlarında ki farklılığa rağmen her durumda en küçük MSE değerine sahiptir (Çizelge 5.6 da görülmektedir).

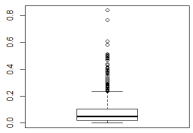
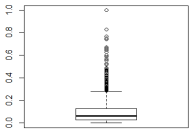
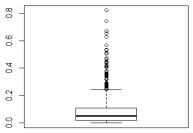
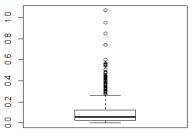
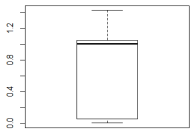
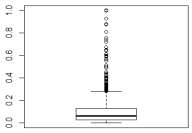
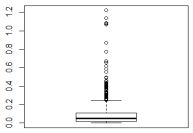
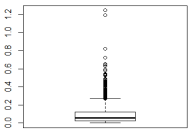
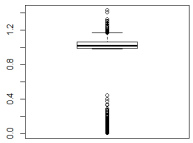
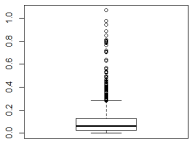
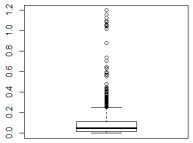
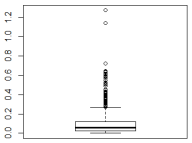
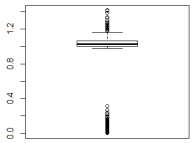
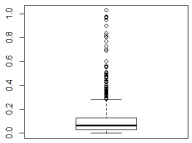
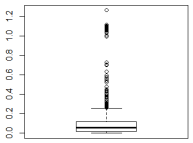
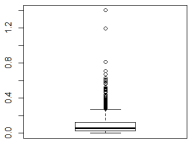
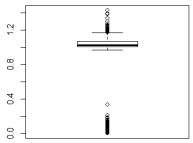
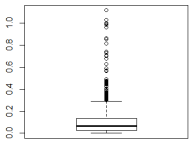
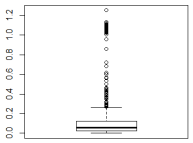
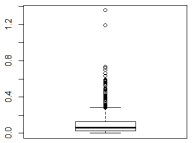
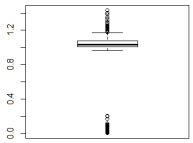
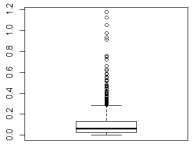
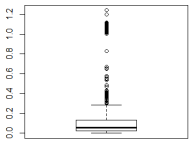
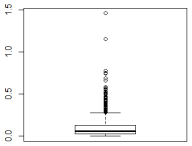
Çizelge 5.6 n=200 olduğunda tahmin edicilere ait MSE'ler



Çizelge 5.7 n=200 olduğunda tahmin edicilere ait Yan'lar



Çizelge 5.8 n=200 için bozulma oranlarına göre box- plot çizimleri

	MLE	MALLOWS	WBYE	WMLE
%0				
%1				
%2				
%3				
%4				
%5				

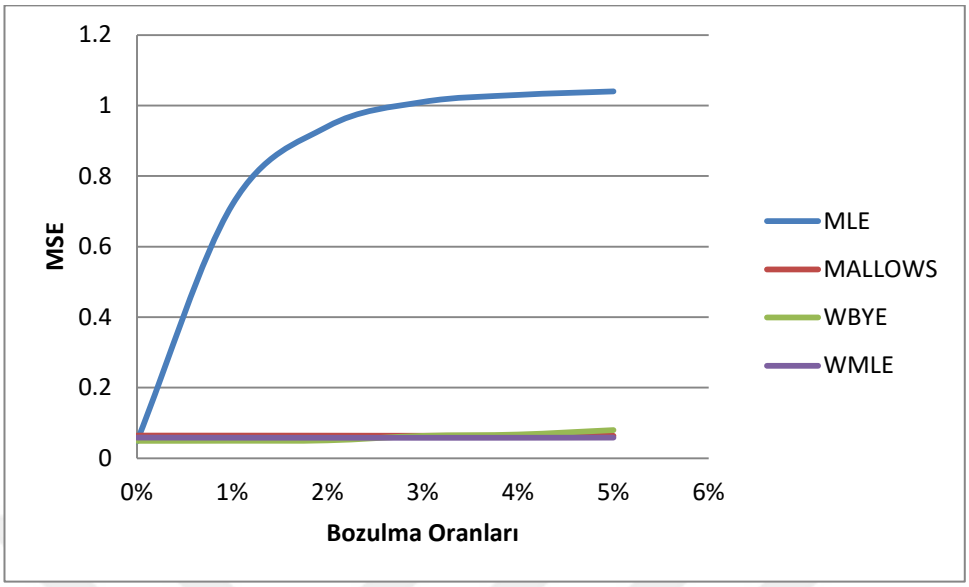
Çizelge 5.8 örneklem büyüklüğü 200 seçildiği durumda, MSE değerini hesaplamak için kullanılan uzaklıklara ait box-plot çizimlerini göstermektedir. Aykırı değer olmadan önce box-plot grafiklerini oluşturan uzaklıklar 0 civarında bir dağılım gösterirken, aykırı değerlerin varlığında uzaklıkların dağılımda MLE yöntemi için önemli bir değişim meydana gelmiştir. Diğer 3 yöntem aykırı değerlerden önemli ölçüde etkilenmemiştir.

Çizelge 5.9 n=300 olduğu durumda tahmin edicilerin Yan ve MSE oranları

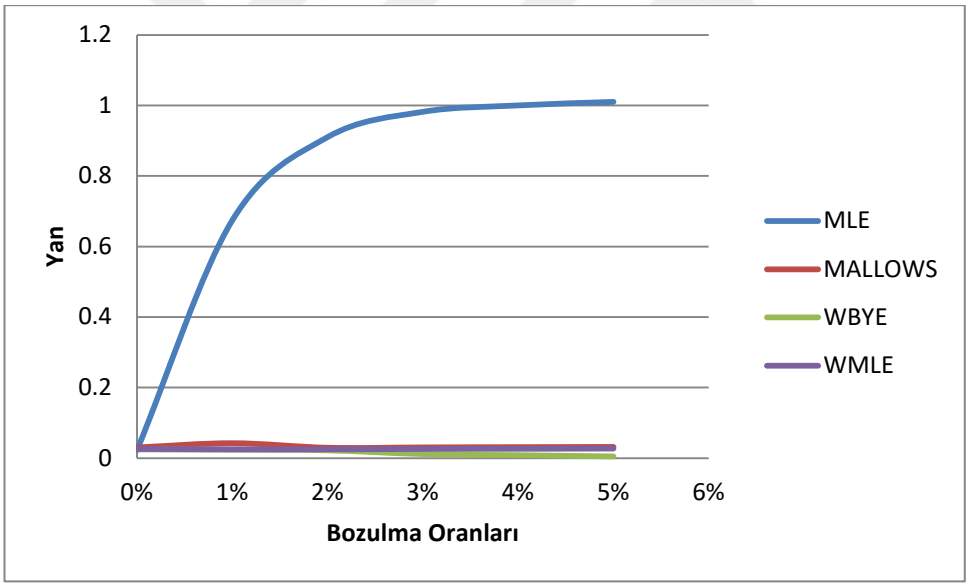
		MLE	MALLOWS	WBYE	WMLE
%0	Yan	0.02195008	0.03010524	0.02455262	0.02592647
	MSE	0.04739687	0.06409687	0.04911486	0.0583962
%1	Yan	0.6737981	0.04220615	0.02353196	0.02486617
	MSE	0.7183559	0.06401548	0.04937427	0.05805015
%2	Yan	0.9100832	0.0293298	0.02197101	0.02485177
	MSE	0.9403468	0.06387399	0.0505571	0.05789987
%3	Yan	0.9818596	0.03029224	0.01142002	0.02589688
	MSE	1.010138	0.06355606	0.06323055	0.05776038
%4	Yan	1.00076	0.03104697	0.008251136	0.02636667
	MSE	1.029685	0.06349113	0.06703754	0.0578796
%5	Yan	1.010408	0.03160366	0.004620659	0.02698047
	MSE	1.039683	0.06338242	0.0796058	0.05830602

Çizelge 5.9 örneklem büyüklüğü 300 seçildiği zaman aykırı değerlerin farklı bozulma oranlarında meydana getirdiği değişmelerini göstermektedir. Bu 4 yönteme ait MSE ve yan değerleri Çizelge 5.9'daki gibi hesaplanmıştır. WMLE yöntemi MSE değerine bakarak en iyi tahmin yöntemi olarak kullanılabilir. Bozulma oranlarında değişiklik olmasına rağmen, en küçük MSE değerine sahip olan yöntemdir. Ayrıca, bozulma oranı değişmesine rağmen WMLE'nin MSE oranlarında pek fazla değişiklik meydana gelmediği çizelge 5.10 da gösterilmektedir.

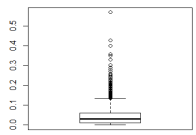
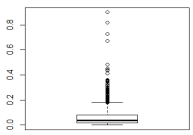
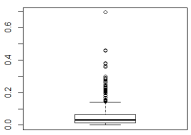
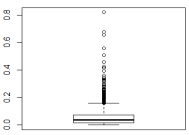
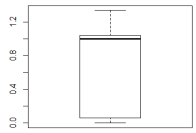
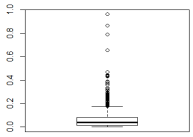
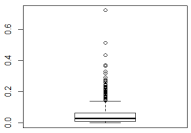
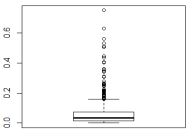
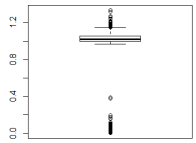
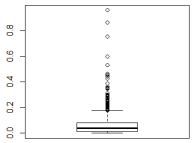
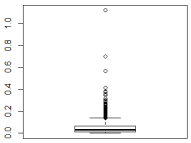
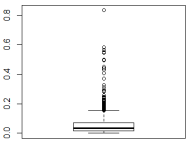
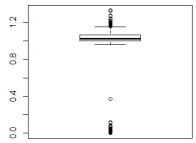
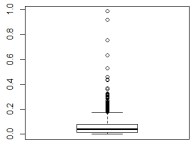
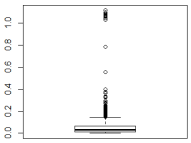
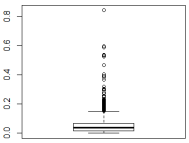
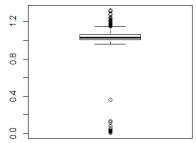
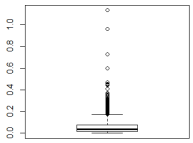
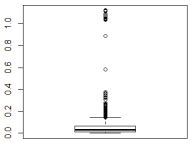
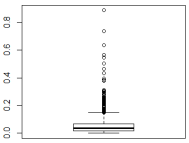
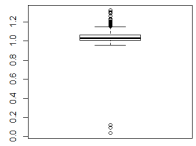
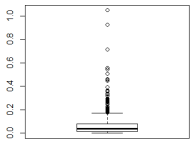
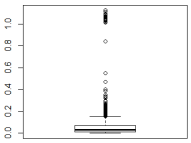
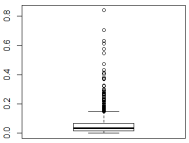
Çizelge 5.10 n=300 olduğunda tahmin edicilere ait MSE'ler



Çizelge 5.11 n=300 olduğunda tahmin edicilere ait Yan'lar



Çizelge 5.12 n=300 için bozulma oranlarına göre box- plot çizimleri

	MLE	MALLOWS	WBYE	WMLE
%0				
%1				
%2				
%3				
%4				
%5				

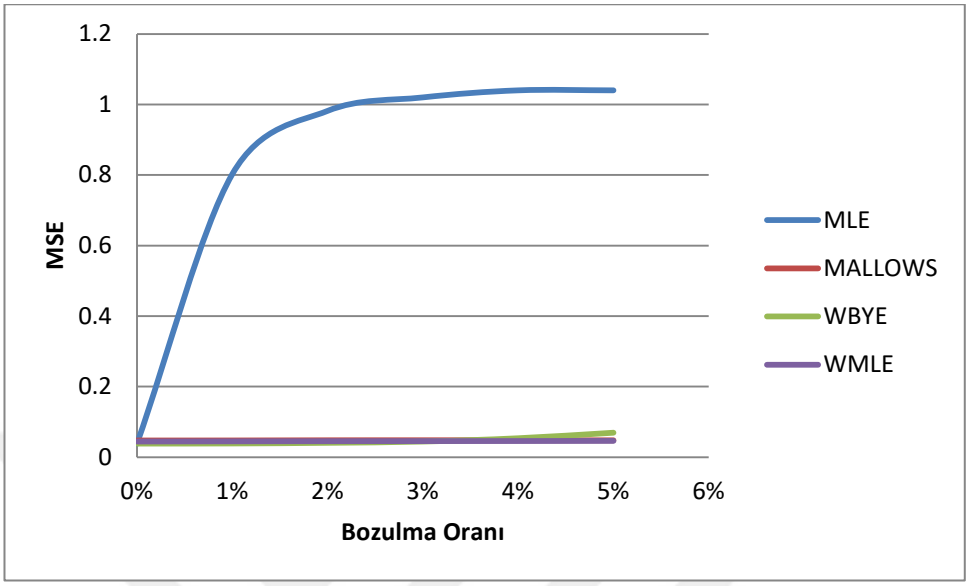
Çizelge 5.12 örneklem büyüklüğü 300 seçildiği durumda, MSE değerini hesaplamak için kullanılan uzaklıklara ait box-plot çizimlerini göstermektedir. Aykırı değer olmadan önce oluşturulan box-plot grafiklerinde uzaklıklar 0 civarında bir dağılım gösterirken, aykırı değerlerin varlığında uzaklıkların dağılımda MLE yöntemi için önemli bir değişim meydana gelmiştir. Aykırı değer olmadığı durumda en iyi dağılımı MLE gösterirken, aykırı değerlerin meydana getirdiği %1 oranında bir bozulmadan MLE yöntemi hızlıca etkilenmiştir. Diğer 3 yöntem aykırı değerlerden önemli ölçüde etkilenmemiştir.

Çizelge 5.13 n=400 olduğu durumda tahmin edicilerin Yan ve MSE oranları

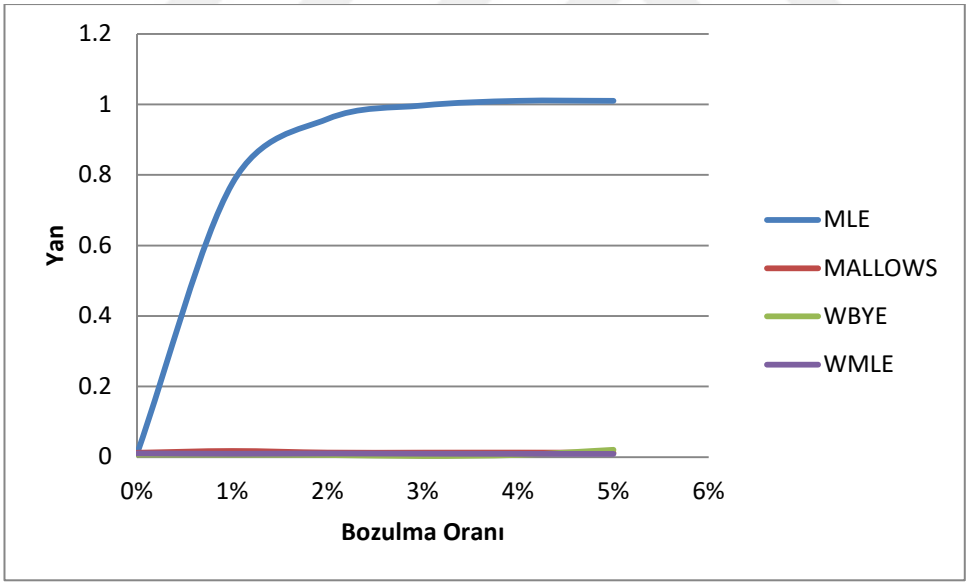
		MLE	MALLOWS	WBYE	WMLE
%0	Yan	0.008152358	0.01214062	0.006986907	0.01034201
	MSE	0.0377607	0.04721988	0.03840407	0.04430425
%1	Yan	0.7753138	0.01713209	0.006596716	0.009291438
	MSE	0.8020221	0.0472497	0.03873706	0.04478039
%2	Yan	0.9586177	0.01262547	0.005746986	0.01007001
	MSE	0.9817028	0.04775281	0.04054897	0.04519569
%3	Yan	0.9969022	0.0125234	0.002783096	0.009510001
	MSE	1.02157	0.04784577	0.04416259	0.04546452
%4	Yan	1.009585	0.01267815	0.006029788	0.009084117
	MSE	1.035283	0.04778733	0.05353006	0.04564579
%5	Yan	1.013151	0.01229725	0.02052129	0.008833621
	MSE	1.039919	0.04765783	0.0692336	0.04604233

Örneklem büyüklüğü 400 seçildiğinde aykırı değerlerin meydana getirdiği farklılıklar Çizelge 5.13 de gösterilmektedir. WMLE yöntemi MSE değerinin küçük olmasından dolayı en iyi tahmin yöntemi olarak seçilebilir. Ayrıca, WMLE yöntemine ait MSE oranlarında çizelge 5.14 den görüleceği gibi neredeyse hiç değişiklik yoktur.

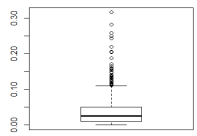
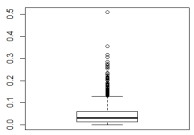
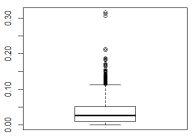
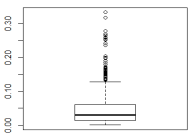
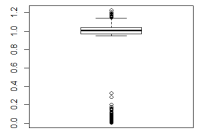
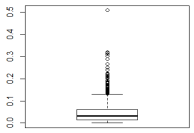
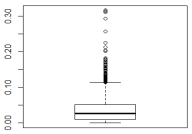
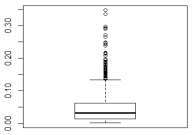
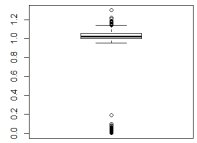
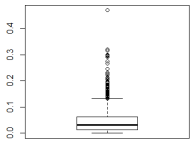
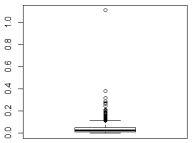
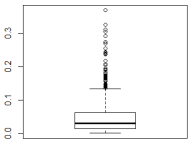
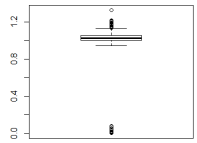
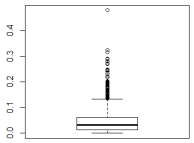
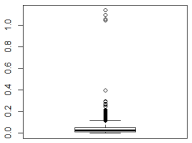
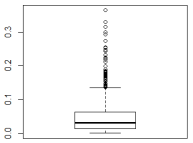
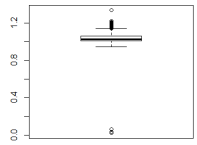
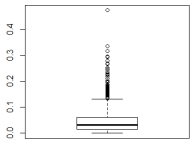
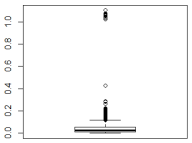
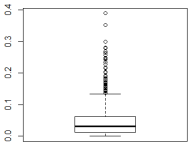
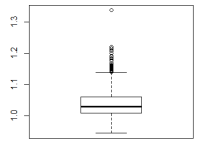
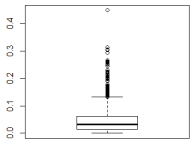
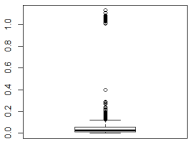
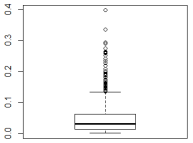
Çizelge 5.14 n=400 olduğunda tahmin edicilere ait MSE'ler



Çizelge 5.15 n=400 olduğunda tahmin edicilere ait Yan'lar



Çizelge 5.16 n=400 için bozulma oranlarına göre box- plot çizimleri

	MLE	MALLOWS	WBYE	WMLE
%0				
%1				
%2				
%3				
%4				
%5				

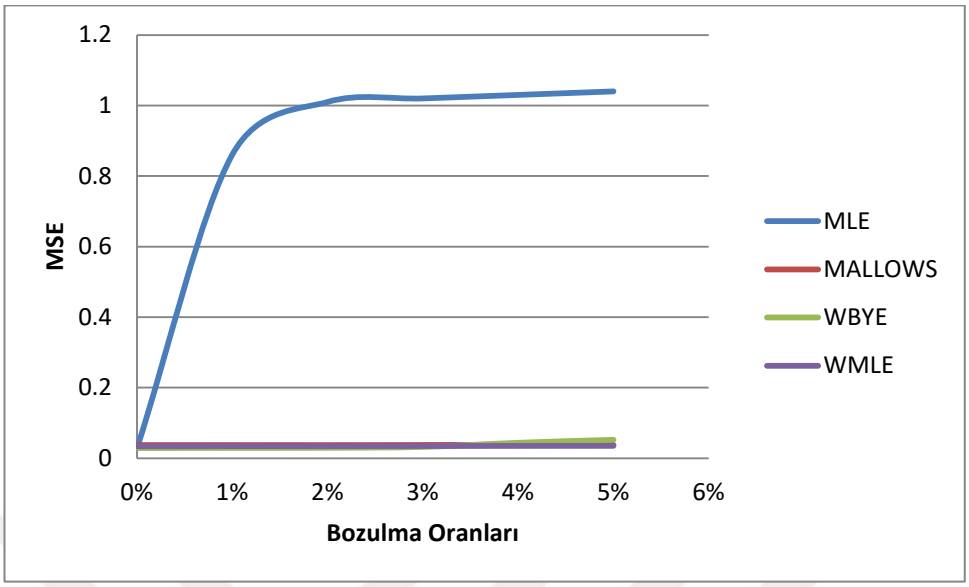
Örneklem büyüklüğü 400 seçildiği durumda, MSE değerini hesaplamak için kullanılan uzaklıklara ait box-plot grafiği Çizelge 5.16'da gösterilmektedir. Bu grafiklere göre aykırı değerler olmadan önce hesaplanan uzaklıklar 0 civarında bir dağılım gösterirken, meydana gelen %1 oranında bir bozulma MLE tahmin yöntemine bağlı tahmin ediciler ile hesaplanan uzaklıkların dağılımında önemli derecede farklılık meydana getirmiştir. Fakat diğer yöntemlerin tahmin edicileri ile hesaplanan uzaklıklara ait box-plot grafiklerinde önemli bir değişiklik olmamıştır.

Çizelge 5.17 n=500 olduğu durumda tahmin edicilerin Yan ve MSE oranları

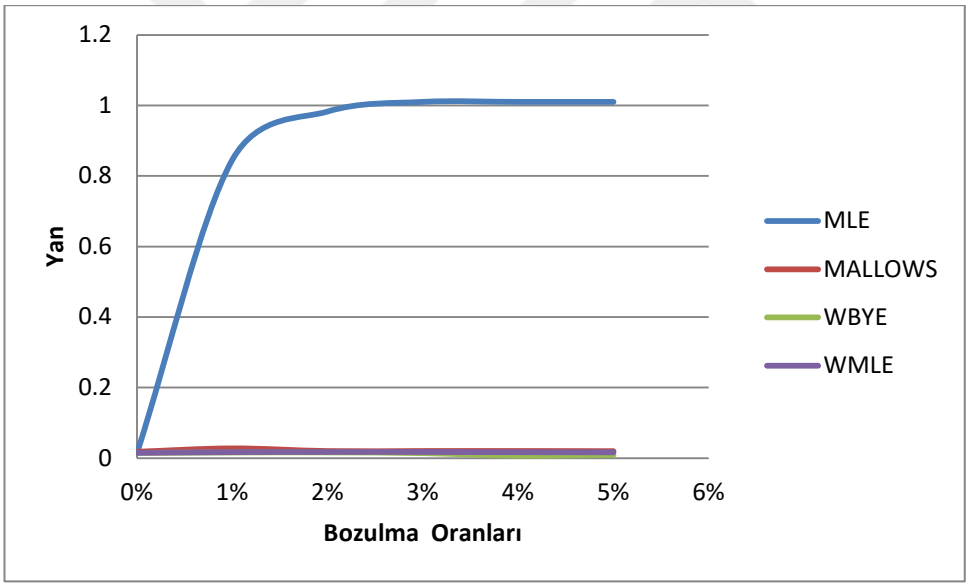
		MLE	MALLOWS	WBYE	WMLE
%0	Yan	0.01367054	0.01798302	0.01478182	0.0148006
	MSE	0.02769391	0.03721419	0.02867107	0.03380333
%1	Yan	0.845299	0.02748836	0.01577877	0.01688358
	MSE	0.8603176	0.03722385	0.02902524	0.03421026
%2	Yan	0.9826084	0.01988424	0.01611419	0.01736536
	MSE	1.002702	0.03707203	0.02938694	0.03414546
%3	Yan	1.008632	0.02006581	0.01400332	0.01733257
	MSE	1.029414	0.03744316	0.03188083	0.03447588
%4	Yan	1.011046	0.02008892	0.00653908	0.01666754
	MSE	1.033344	0.03765694	0.04365332	0.03477185
%5	Yan	1.012675	0.01959476	0.008917443	0.01640913
	MSE	1.035723	0.03751797	0.05207516	0.03492906

500 örneklem büyüklüğüne sahip olacak şekilde üretilen veri setinde, farklı bozulma oranları karşısında 4 tahmin yönteminde oluşan değişiklikler gösterilmiştir. Bu 4 yönteme ait MSE ve yan değerleri Çizelge 5.17'deki gibi hesaplanmıştır. Bu örneklem büyüklüğünde de WMLE yöntemi MSE değerinden dolayı en iyi yöntem olarak tercih edilebilir. Çizelge 5.18 de WMLE yöntemine ait MSE'lerin değişiminin en az olduğu açıkça belirtilmektedir.

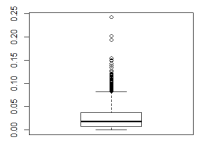
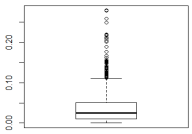
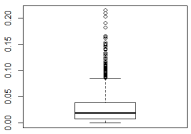
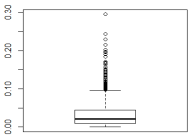
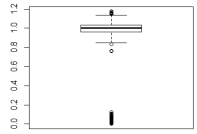
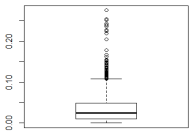
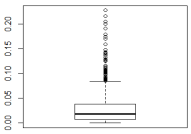
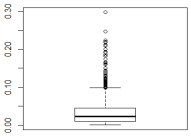
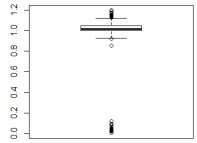
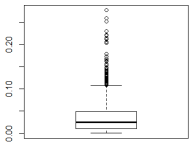
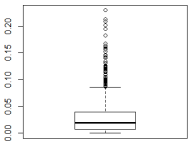
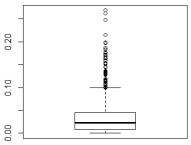
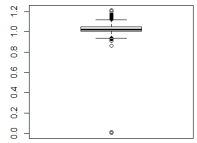
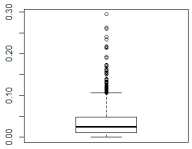
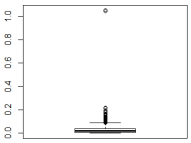
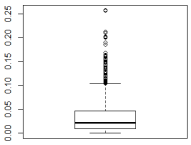
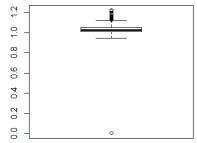
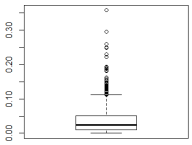
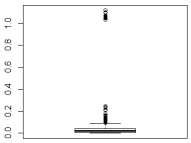
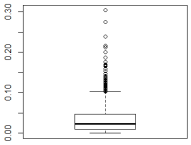
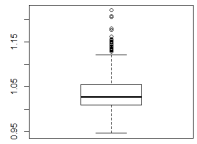
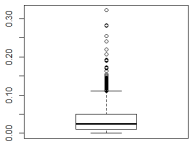
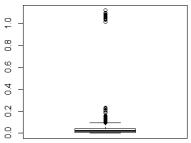
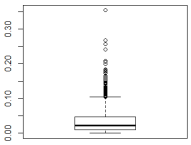
Çizelge 5.18 n=500 olduğunda tahmin edicilere ait MSE'ler



Çizelge 5.19 n=500 olduğunda tahmin edicilere ait Yan'lar



Çizelge 5.20 n=500 için bozulma oranlarına göre box- plot çizimleri

	MLE	MALLOWS	WBYE	WMLE
%0				
%1				
%2				
%3				
%4				
%5				

Çizelge 5.20 örneklem büyüklüğü 500 seçildiği durumda, MSE değerini hesaplamak için kullanılan uzaklıklara ait box-plot grafikleri göstermektedir. Aykırı değerlere bağlı bozulmalar MLE yönteminden elde edilen tahmin ediciler ile hesaplanan uzaklıklara ait box plot çizimlerini önemli ölçüde etkilemesine rağmen, aykırı değerlerden kaynaklanan bozulmalar diğer 3 tahmin yöntemi ile hesaplanan tahmin değerlerinde ve bu değerlere bağlı olarak hesaplanan uzaklıklarda önemli ölçüde bir değişim meydana getirmemiştir.

5.2 Gerçek Veri Uygulaması

Bu uygulamada kullanılan veriler, Pimalı Hintliler Diyabet Veri tabanının bir parçası olarak "Ulusal Diyabet Sindirim ve Böbrek Hastalıkları Enstitüsü" tarafından toplanıp sunulmuştur. Özellikle, veri setindeki tüm hastalar 21 yaş ve üstü Pimalı Hintli kadınlardır. Bu veri setinde 700 hastanın çeşitli özellikleri ile diyabet olması olasılığı arasındaki ilişkiye bakılmıştır (Anonymus 2019).

Bu özellikler:

- Vücut kütle indeksi (BMI)
- Diyabet soyağacı işlevi (diapedi)

şeklindedir. Yukarda verilen değişkenler x bağımsız değişkenleri olarak ele alınmıştır. Diyabet olma olasılığı olarak ifade edilen y bağımlı değişkenine ise 0 ya da 1'lerden oluşan bir değişken atanmıştır.

İlk olarak bozulma olmayan veri de hesaplamalar yapılmıştır. Daha sonra x yönündeki veriler sırasıyla %1, %2, %3, %4 ve %5 oranında bozulmuştur. Bozulmuş veri setinde hangi yöntemin daha anlamlı ve yansız sonuçlar verdiği karşılaştırılmıştır.

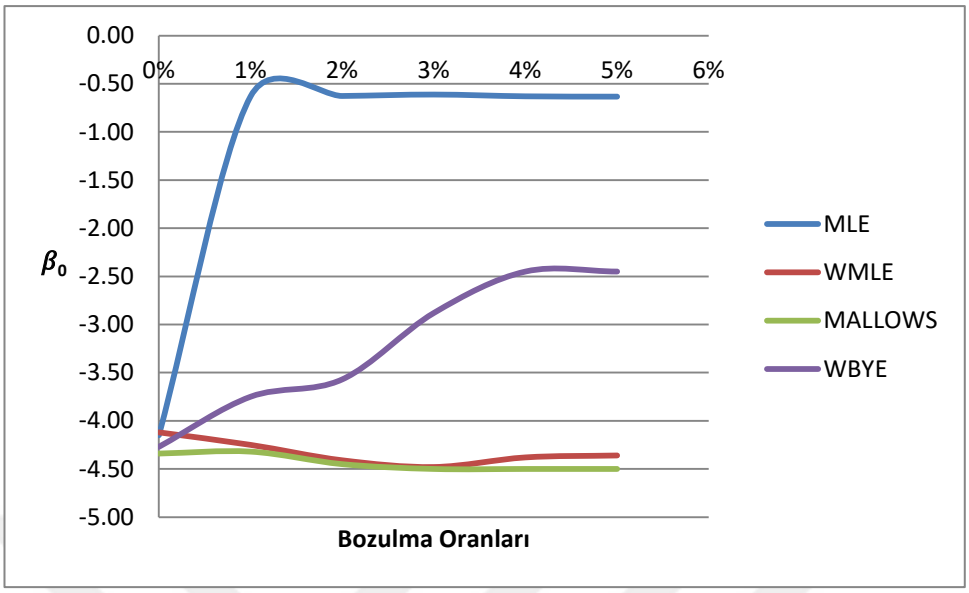
Aşağıda bozulma oranlarına göre parametre tahminlerinde meydana gelen değişimler 4 yönteme göre gösterilmektedir. Burada β parametrelerine ilişkin parametre tahminleri göz önüne alınarak değerlendirme yapılmıştır.

Çizelge 5.21 Farklı bozulma oranlarına göre parametre tahmin değerleri

		%0	%1	%2	%3	%4	%5
MLE	β_0	-4.1579	-0.6313	-0.6269	-0.6121	-0.6304	-0.6337
	β_1	0.0943	-0.0002	-0.0003	-0.0003	-0.0001	-0.0001
	β_2	0.8836	-0.0027	-0.0032	-0.0027	-0.0084	-0.0067
WMLE	β_0	-4.1175	-4.2548	-4.4085	-4.4787	-4.3778	-4.3757
	β_1	0.0859	0.0956	0.0995	0.1018	0.0987	0.0985
	β_2	1.1422	1.0141	1.1253	1.1379	1.1268	1.1371
MALLOWS	β_0	-4.3309	-4.316	-4.4532	-4.5048	-4.4974	-4.4960
	β_1	0.0973	0.0964	0.1002	0.1021	0.1018	0.1017
	β_2	1.1300	1.1635	1.2418	1.2166	1.2235	1.2274
WBYE	β_0	-4.2732	-3.7510	-3.5677	-2.8843	-2.4473	-2.4496
	β_1	0.0972	0.0815	0.0777	0.0603	0.0488	0.0487
	β_2	1.0905	0.9572	0.9269	0.7183	0.6094	0.6133

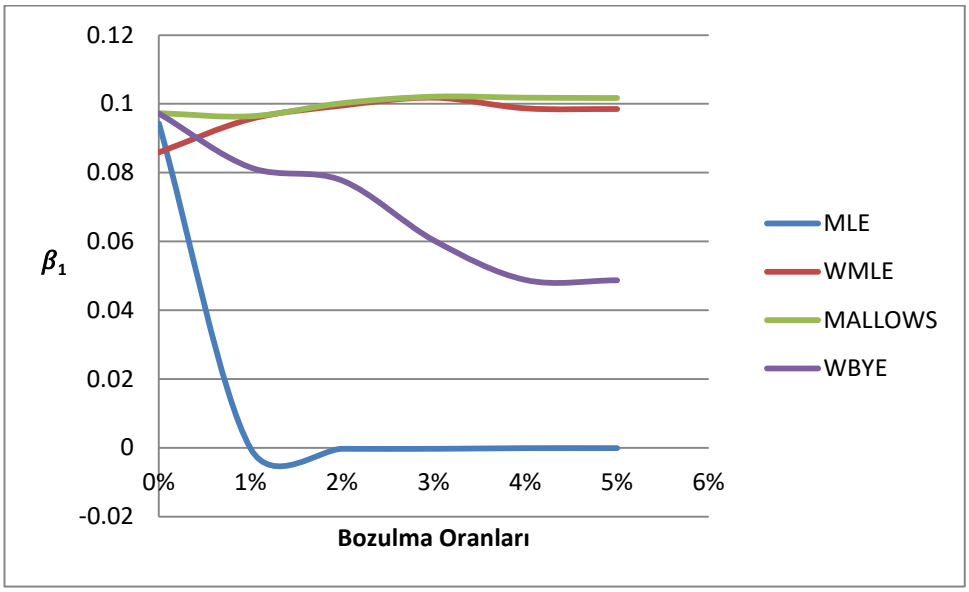
Çizelge 5.21 e baktığımızda 4 yönteme ait β parametre tahminleri görülmektedir. Veri de aykırı değer yokken 4 yöntemde de birbirine oldukça yakın tahmin değerleri vermektedir. Fakat veri setinde aykırı değerler olduğu zaman MLE yöntemi %1 bozulma oranında bile oldukça hızla etkilenmiştir ve β parametre tahminleri hızlı bir şekilde değişmiştir. MLE' nin β tahmin değerlerinde hızlı bir değişme varken. WBYE, MALLOWS ve WMLE yönteminde parametre tahmin değerleri çok değişmemiştir. Bozulma oranı artmasına rağmen WMLE ile tahmin edilen parametre değerleri birbirine oldukça yakın değerler almıştır.

Çizelge 5.22 Farklı bozulma oranlarına göre β_0 parametre tahmini



Çizelge 5.22 de görüldüğü üzere MLE' nin β_0 tahmin değerinde oldukça fazla bir değişme vardır. MALLOWS ve WMLE yönteminde parametre tahmin değerleri çok değişmemiştir. Bozulma oranı artmasına rağmen WMLE ile tahmin edilen β_0 parametre değerleri bu değişimden oldukça az etkilenmiştir ve son derece anlamlı tahmin değerleri vermiştir.

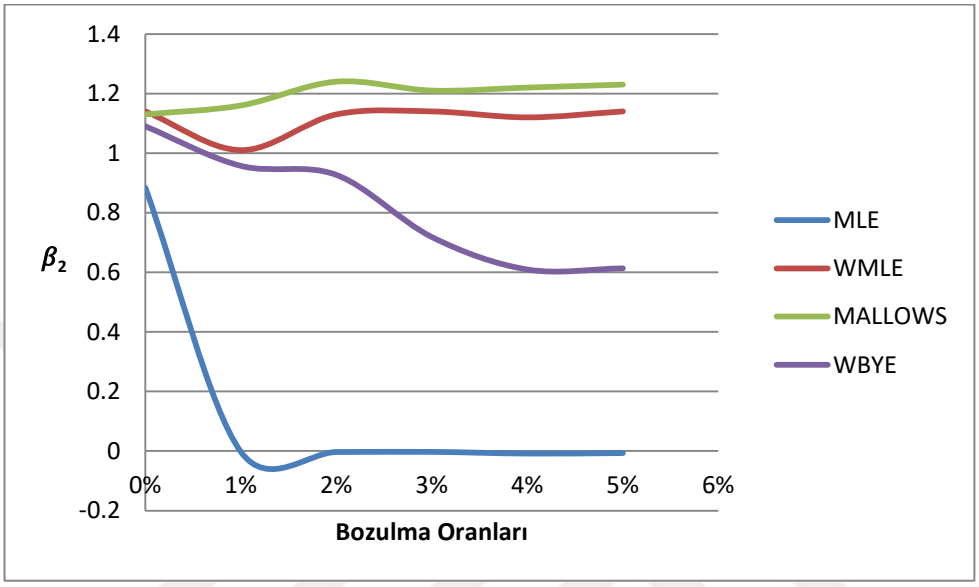
Çizelge 5.23 Farklı bozulma oranlarına göre β_1 parametre tahmini



MLE' ye ait β_1 tahmin değeri %1 oranındaki bir bozulmadan bile oldukça etkilenmiştir. Çizelge 5.23 de MLE nin β_1 tahmin değerindeki değişikler açıkça görülmektedir.

MALLOWS ve WMLE yönteminde parametre tahmin değerleri çok değişmemiştir. Bozulma oranı artmasına rağmen WMLE ile tahmin edilen β_1 parametre değerleri birbirine oldukça yakın değerler almıştır.

Çizelge 5.24 Farklı bozulma oranlarına göre β_2 parametre tahmini



β_2 parametresinin tahmininde MLE yöntemi bozulmalardan oldukça etkilenip farklı sonuçlar vermiştir. Çizelge 5.24 ile verilen grafikte MLE nin β_2 tahmin değerindeki değişikler açıkça görülmektedir. Parametre tahmin değeri bozulmalardan en az etkilenen yöntemler MALLOWS ve WMLE yöntemleridir. Bozulma oranlarındaki değişmeye rağmen WMLE ile tahmin edilen β_2 parametre değerleri birbirine oldukça yakındır.

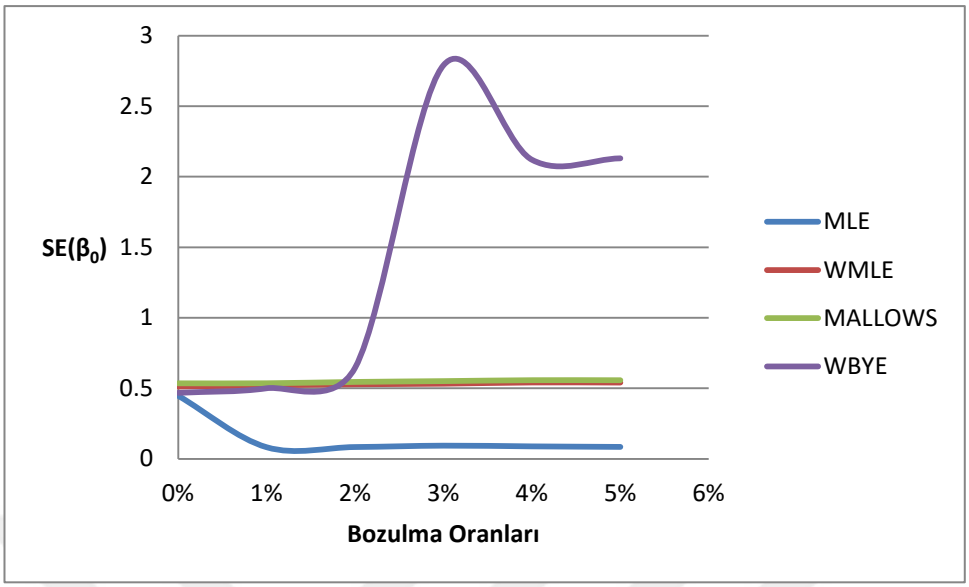
Çizelge 5.25 Farklı bozulma oranlarına göre parametrelere ait standart hatalar

		%0	%1	%2	%3	%4	%5
MLE	$SE(\beta_0)$	0.4473	0.0822	0.0835	0.0933	0.0882	0.0844
	$SE(\beta_1)$	0.0128	0.0006	0.0007	0.0008	0.0005	0.0004
	$SE(\beta_2)$	0.2582	0.0102	0.0168	0.0826	0.0659	0.0423
WMLE	$SE(\beta_0)$	0.5126	0.5210	0.5284	0.5324	0.5412	0.5405
	$SE(\beta_1)$	0.0144	0.0143	0.0145	0.0147	0.0150	0.0150
	$SE(\beta_2)$	0.2893	0.4807	0.4748	0.4663	0.4664	0.4667
MALLOWS	$SE(\beta_0)$	0.5352	0.5356	0.5450	0.5508	0.5573	0.5570
	$SE(\beta_1)$	0.0154	0.0154	0.0156	0.0157	0.0159	0.0160
	$SE(\beta_2)$	0.3678	0.3759	0.3797	0.3764	0.3767	0.3769
WBYE	$SE(\beta_0)$	0.4676	0.5001	0.6451	2.7892	2.1165	2.1258
	$SE(\beta_1)$	0.0129	0.0132	0.0164	0.0557	0.0506	0.0507
	$SE(\beta_2)$	0.3065	0.3332	0.4411	2.1600	1.353	1.3455

Çizelge 5.25 deki standart hatalarda meydana gelen değişmeye bakarsak MLE tahmin yöntemi aykırı değerlerden çok etkilendiği için parametrelerin standart hatalarında ciddi bir değişiklik meydana gelmiştir. Fakat WMLE yöntemine bakacak olursak aykırı değerlerden etkilenmesine rağmen standart hatalarında büyük bir değişiklik meydana gelmemiştir. Bu durum aykırı değer sayısı artmasına rağmen, WMLE yönteminin bu değerlerden az etkilenen robust bir yöntem olduğunu göstermektedir.

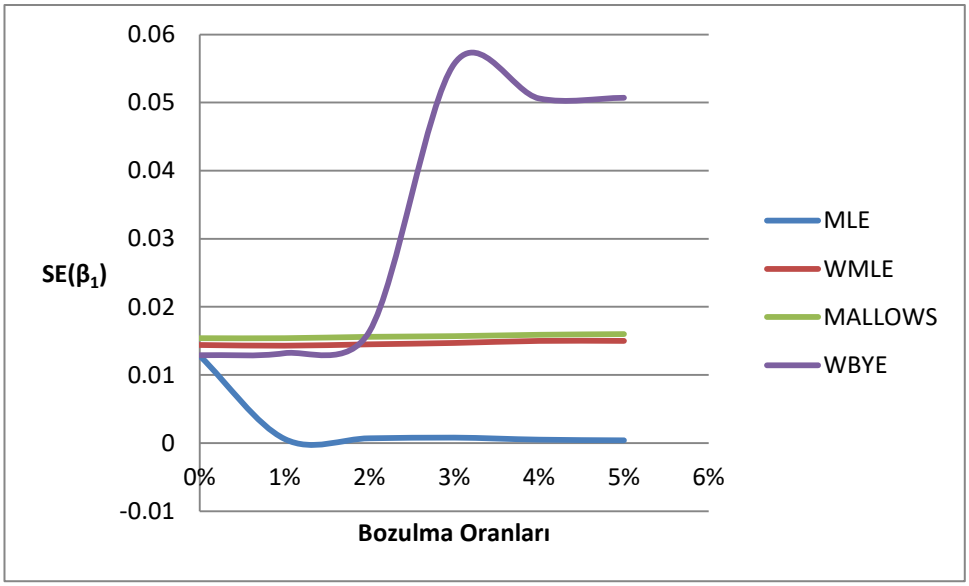
Yukarda bahsedilen sebeplerden ötürü WMLE yöntemi veride x yönünde aykırı değer olması durumunda etkin olarak tercih edilebilir. Bu durum WMLE yönteminin, bozulma oranı değişse bile robust bir alternatif olarak öne sürülebileceği fikrini desteklemektedir.

Çizelge 5.26 Farklı bozulma oranlarına göre β_0 parametresine ait standart hata



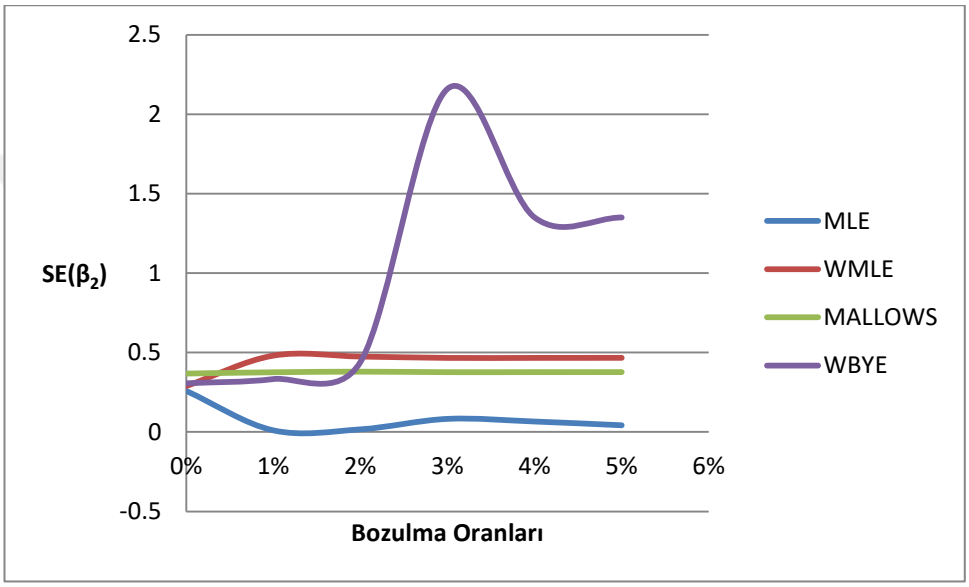
Çizelge 5.26 da verilen MLE ve WBYE yöntemi ile tahmin edilen β_0 parametresine ait standart hatalarda ciddi bir değişiklik meydana gelmiştir. Fakat WMLE ve MALLOWS yöntemine bakacak olursak aykırı değerlerden etkilenmesine rağmen standart hatalarında büyük bir değişiklik meydana gelmemiştir. Bu durum aykırı değer sayısı artmasına rağmen, WMLE yönteminin bu değerlerden az etkilenen robust bir yöntem olduğunu göstermektedir.

Çizelge 5.27 Farklı bozulma oranlarına göre β_1 parametresine ait standart hata



MLE ve WBYE yöntemi ile tahmin edilen çizelge 5.27 ile verilen β_1 parametresine ait standart hatalar %1 oranında bir bozulmadan bile oldukça etkilenip farklı sonuçlar vermiştir. Fakat WMLE ve MALLOWS yöntemi ile tahmin edilen β_1 parametresine ait standart hatalara bakacak olursak aykırı değerlerden etkilenmesine rağmen büyük bir değişiklik meydana gelmemiştir. WMLE yöntemi bu değerlerden az etkilendiğinden dolayı robust bir alternatif olarak kullanılabilir.

Çizelge 5.28 Farklı bozulma oranlarına göre β_2 parametresine ait standart hata



β_2 parametresine ait standart hatalar çizelge 5.28 ile verilmiştir. Bu grafikten de anlaşılacağı üzere MLE ve WBYE tahmin yöntemleri ile tahmin edilen β_2 parametresinin standart hataları %1 oranında bir bozulmadan bile oldukça etkilenip farklı sonuçlar vermiştir. Buna rağmen, WMLE ve MALLOWS yöntemi ile tahmin edilen β_2 parametresine ait standart hatalarda büyük bir değişiklik meydana gelmemiştir. Bundan dolayı, WMLE yöntemi aykırı değerlerden önemli bir ölçüde etkilenmeyen robust bir yöntem olarak kullanılabilir.

6. SONUÇ

Bu tezde, x-yönünde aykırı değere sahip lojistik regresyon modelinin parametre tahminleri MLE ve üç robust tahmin yöntemi kullanılarak yapılmıştır. En çok olabilirlik tahmin edicisi, ağırlıklandırılmış Bianco-Yohai tahmin yöntemi, ağırlıklandırılmış en çok olabilirlik tahmin yöntemi ve Mallows ağırlığına göre ağırlıklandırılmış olan tahmin yöntemi kullanılmıştır. Veri setinde bulunan aykırı değerlerin oranı arttıkça, özellikle MLE yönteminde ciddi tahmin hataları meydana gelmiştir. Bu bozulmalara ilişkin veriler ve diğer yöntemlerle alakalı değerler tablolar ve grafikler halinde tezde sunulmuştur.

Simülasyon bulgularına göre, MLE' nin aykırı değer varlığında yanlı olabileceği, oysa diğer robust tahmin edicilerin daha iyi sonuçlar verdiği ortaya çıkmıştır. Bozulma, verilerin kaldıraç noktalarında meydana geldiğinde, simülasyon sonuçları robust yöntemlerin küçük yana ve küçük hata kareler ortalamasına sahip olduğunu göstererek, robust parametre tahminlerinin bu bozulmadan az etkileneceğini göstermektedir.

Veri setinde bozulma olduğu durumlarda, WMLE tahmin yöntemi tercih edilir çünkü daha güvenilir parametre tahminleri yapmaya yardımcı olur. Bu çalışmada aykırı değerlerin verilerde meydana getirdiği her bozulma oranı için WMLE yöntemi, test edilen diğer tahmin yöntemlerinden daha sağlamdır. Bu tahmin yöntemini MALLOWS ve WBYE tipi tahmin yöntemleri takip eder.

Bu tez çalışmasında veri setinde oluşan bozulma durumlarında dört tahmin metodu birbiri ile kıyaslanmıştır. İlerde yapılacak araştırmalar için, metotların karşılaştırılması bir fikir sunmaktadır.

Simülasyon çalışmasında bozulma olmadan kullanılan veri setinin yanı sıra, aynı veri setine aykırı değerler eklenip veri setinde faklı oranda bozulmalar meydana getirilmiştir. Tez çalışmasında, kullanılan robust yöntemleri birbirleriyle karşılaştırmak için Monte Carlo simülasyon metodu kullanılıp birbirinden ayrı örneklem büyüklüğü ve bozulma oranına sahip veri setleri türetilmiştir. Her metot için, oluşturulan bu veri setlerinden yararlanılmıştır. Yöntemlerin performansları hakkında yorum yapabilmek için, hata

kareler ortalaması, yan ve parametre tahmin değerleri ve parametre tahminlerine ait standart hataları değerlendirilmiştir.

Bozulma oranının, kullanılacak metodu seçme konusunda en önemli faktör olduğu anlaşılmıştır. Örneklem büyüklüklerinin metotların etkinliklerinde çok bir farklılığa sebep olmadığı tespit edilmiştir. Veri de bozulma olmadığında her dört yöntemde birbirine oldukça yakın sonuçlar vermesine rağmen %1 oranında bir bozulma bile yöntemlerin performanslarını etkilediği gözlenmiştir.

Ayrıca bu tezde sadece simülasyon ile değil, aynı zamanda gerçek veri seti ile de aynı çalışma yapılmış ve sonuçlar incelenmiştir. Gerçek veri setinin sonuçlarına göre de WMLE en robust sonuçları verdiği saptanmıştır.

KAYNAKLAR

Agresti, A. 2007. An Introduction to Categorical Data Analysis. John Wiley and Sons, 7, New Jersey.

Ahmad, S., Ramli, N.M. and Midi, H. 2010. Robust Estimators in Logistic Regression: A Comparative Simulation Study. Journal of Modern Applied Statistical Methods, 9(2), 502-511.

Anderson, J. A. 1972. Separate Sample Logistic Discrimination. Biometrika, 59(1), 19-35.

Anderson Erling, B. 1990. The Statistical Analysis of Categorical Data. Springer Verlag, Berlin.

Anonymus. 2019. Web Sitesi: https://www.kaggle.com/kandij/diabetes-dataset#diabetes2.csv. Erişim tarihi: 02/07/2019

Bianco, A. M. and Yohai, V. J. 1996. Robust estimation in the logistic regression model, In: Robust statistics, Data analysis and computer intensive methods. Reider, H. (ed), Springer Verlag, 17-34, New York.

Carroll, R. J. and Pederson, S. 1993. On robustness in the logistic regression model. Journal of the Royal Statistical Society,Series B, 55(3), 693-706.

Copas, J. B. 1988. Binary regression model for contaminated data. Journal of the Royal Statistical Society, Series B, 50(2), 225-265.

Croux, C., Flandre, C. and Haesbroeck,G. 2002. The breakdown behavior of the maximum likelihood estimator in the logistic regression model. Statistics and Probability Letters, 60(4), 377-386.

Croux, C. and Haesbroeck, G. 2003. Implementing the Bianco and Yohai estimator for logistic regression. Computational Statistics and Data Analysis Journal, 44(1-2), 273-295.

Day, N. E. and Kerridge, D. F. 1967. A general maximum likelihood discriminant. Biometrics, 23(2), 313-323.

Elhan, A. H. 1997. Lojistik Regresyon Analizinin İncelenmesi ve Tıpta Bir Uygulaması. Yüksek Lisans Tezi. Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, 4, Ankara.

Hosmer, D. W. and Lemeshow, S. 1980. Goodness-of-fit tests for the multiple logistic regression model. Communications in Statistics - Theory and Methods, 9(10), 1043-1069.

Hosmer, D. and Lemeshow, S. 2000. Applied Logistic Regression (Wiley Series in Probability and Statistics). Wiley-Interscience, 2-4, New York.

Kunsch, H. R., Stefanski, L. A. and Carroll, R. J. 1989. Conditionally unbiased bounded influence estimation in general regression models, with applications to

generalized linear models. Journal of American Statistical Association, 84(406), 460-466.

McCullagh, P. and Nelder, J.A. 1989.Generalized Linear Models (Second Edition). Chapman and Hall, 33-37, London.

Pregibon, D. 1981. Logistic regression diagnostics.The Annals of Statistics, 9(4), 705-724.

Pregibon, D. 1982. Resistant fits for some commonly used logistic models with medical applications. Biometrics, 38(2), 485-98.

Rao, J. N. K. and Scott, A. J. 1984. On Chi-Squared Tests for Multiway Contingency Tables with Cell Properties Estimated from Survey Data. The Annals of Statistics, 12(1), 46–60.

Rao, C.R. 1973. Linear Statistical Inference and its Applications (Second Editon). John Wiley and Sons, Inc., New York.

Rousseeuw, P.J 1984. Least median of squares regression. Journal of the American Statistical Association, 79 (388), 871–880.

Rousseeuw, P.J 1985. Multivariate estimation with high breakdown point. Mathematical Statistics and Applications, B, 283–297.

Rousseeuw, P. J. and Leroy, A. M., 1987. Robust regression and outlier detection. John Wiley and Sons Inc, 347, United States of America.

Rousseeuw, P. J. and Van Zomeren, B. C., 1990. Unmasking multivariate outliers and leverage points (with discussion). Journal of the American Statistical Association, 85(411), 633–651.

Simeckova, M. 2005. Maximum Weighted Likelihood Estimator in Logistic Regression. WDS'05 Proceedings of Contributed Papers, 144-148.

Stefanski, L. A. 1985. The effects of measurement error on parameter estimation. Biometrika, 72(3), 583-592.

Tsiatis, A. A. 1980. A note on a goodness-of-fit test for the logistic regression model. Biometrika, 67(1), 250-251.

Victoria-Feser, M-P. 2002. Robust inference with binary data. Psychometrika, 67(1), 21-32.

Wald, A. 1943. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. Transactions of the American Mathematical Society, 54(3), 426-482.

EK 1 R KODLARI

```
rm(list = ls())
library(MASS)
library(robust)
r<-1000
n<-100
mu<-0
sigma<-1
location<-0
scale<-1
beta=matrix(1,2,1)
beta0matrix_a0<-rep(NA,r)
beta1matrix_a0<-rep(NA,r)
betasonmatrix_a0<-rep(NA,r)
beta0matrix_b0<-rep(NA,r)
beta1matrix_b0<-rep(NA,r)
betasonmatrix_b0<-rep(NA,r)
beta0matrix_c0<-rep(NA,r)
beta1matrix_c0<-rep(NA,r)
betasonmatrix_c0<-rep(NA,r)
beta0matrix_d0<-rep(NA,r)
beta1matrix_d0<-rep(NA,r)
betasonmatrix_d0<-rep(NA,r)
beta0matrix_a1<-rep(NA,r)
beta1matrix_a1<-rep(NA,r)
betasonmatrix_a1<-rep(NA,r)
beta0matrix_b1<-rep(NA,r)
beta1matrix_b1<-rep(NA,r)
betasonmatrix_b1<-rep(NA,r)
beta0matrix_c1<-rep(NA,r)
beta1matrix_c1<-rep(NA,r)
betasonmatrix_c1<-rep(NA,r)
```

```
beta0matrix_d1<-rep(NA,r)
beta1matrix_d1<-rep(NA,r)
betasonmatrix_d1<-rep(NA,r)
beta0matrix_a2<-rep(NA,r)
beta1matrix_a2<-rep(NA,r)
betasonmatrix_a2<-rep(NA,r)
beta0matrix_b2<-rep(NA,r)
beta1matrix_b2<-rep(NA,r)
betasonmatrix_b2<-rep(NA,r)
beta0matrix_c2<-rep(NA,r)
beta1matrix_c2<-rep(NA,r)
betasonmatrix_c2<-rep(NA,r)
beta0matrix_d2<-rep(NA,r)
beta1matrix_d2<-rep(NA,r)
betasonmatrix_d2<-rep(NA,r)
beta0matrix_a3<-rep(NA,r)
beta1matrix_a3<-rep(NA,r)
betasonmatrix_a3<-rep(NA,r)
beta0matrix_b3<-rep(NA,r)
beta1matrix_b3<-rep(NA,r)
betasonmatrix_b3<-rep(NA,r)
beta0matrix_c3<-rep(NA,r)
beta1matrix_c3<-rep(NA,r)
betasonmatrix_c3<-rep(NA,r)
beta0matrix_d3<-rep(NA,r)
beta1matrix_d3<-rep(NA,r)
betasonmatrix_d3<-rep(NA,r)
beta0matrix_a4<-rep(NA,r)
beta1matrix_a4<-rep(NA,r)
betasonmatrix_a4<-rep(NA,r)
beta0matrix_b4<-rep(NA,r)
beta1matrix_b4<-rep(NA,r)
```

```
betasonmatrix_b4<-rep(NA,r)
beta0matrix_c4<-rep(NA,r)
beta1matrix_c4<-rep(NA,r)
betasonmatrix_c4<-rep(NA,r)
beta0matrix_d4<-rep(NA,r)
beta1matrix_d4<-rep(NA,r)
betasonmatrix_d4<-rep(NA,r)
beta0matrix_a5<-rep(NA,r)
beta1matrix_a5<-rep(NA,r)
betasonmatrix_a5<-rep(NA,r)
beta0matrix_b5<-rep(NA,r)
beta1matrix_b5<-rep(NA,r)
betasonmatrix_b5<-rep(NA,r)
beta0matrix_c5<-rep(NA,r)
beta1matrix_c5<-rep(NA,r)
betasonmatrix_c5<-rep(NA,r)
beta0matrix_d5<-rep(NA,r)
beta1matrix_d5<-rep(NA,r)
betasonmatrix_d5<-rep(NA,r)
#Tekrar için döngü oluşturma
for(i in 1:r)
{
#hata matrisi,beta hata matrisi, bağımlı ve bağımsız değişkenlere ilişkin matrisleri
oluşturma
e=matrix (rlogis(n,location,scale),n,1)
x=matrix(rnorm(n,mean=0,sd=1),n,1)
x1=as.matrix(cbind(rep(1,n),x))
denk= (x1%*%beta)+e
y<-ifelse(denk<=0,0,1)
y<- ifelse(x>95,0,y)
#aykırı değerlere ait matrisleri oluşturma
c= matrix(x[-n,])
```

```
b1=matrix(rnorm(1,mean=100,sd=1),1,1)
xad1=rbind(c,b1)
c= matrix(x[-n,])
c1=matrix(c[-n+1,])
b2=matrix(rnorm(2,mean=100,sd=1),2,1)
xad2=rbind(c1,b2)
c= matrix(x[-n,])
c1=matrix(c[-n+1,])
c2=matrix(c1[-n+2,])
b3=matrix(rnorm(3,mean=100,sd=1),3,1)
xad3=rbind(c2,b3)
c= matrix(x[-n,])
c1=matrix(c[-n+1,])
c2=matrix(c1[-n+2,])
c3=matrix(c2[-n+3,])
b4=matrix(rnorm(4,mean=100,sd=1),4,1)
xad4=rbind(c3,b4)
c= matrix(x[-n,])
c1=matrix(c[-n+1,])
c2=matrix(c1[-n+2,])
c3=matrix(c2[-n+3,])
c4=matrix(c3[-n+4,])
b5=matrix(rnorm(5,mean=100,sd=1),5,1)
xad5=rbind(c4,b5)
#karşılaştılacak yöntemlere ait kodlar
mle.out0<-glm(y~x,family=binomial("logit"))
mlebeta0<-summary(mle.out0)$coefficients
beta0matrix_a0[i]<- mlebeta0[1,1]
beta1matrix_a0[i]<-mlebeta0[2,1]
betasonmatrix_a0<- matrix(1,2,r)
mallows.out0 <- glmrob(y~x,family=binomial("logit"),method = "Mqle", weights.on.x
="covMcd")
```

```
mallowsbeta0<-summary(mallows.out0)$coefficients
beta0matrix_b0[i]<- mallowsbeta0[1,1]
beta1matrix_b0[i]<- mallowsbeta0[2,1]
betasonmatrix_b0<- matrix(1,2,r)
wby.out0<-glmrob(y~x,family=binomial,method="WBY")
wbybeta0<-summary(wby.out0) $coefficients
beta0matrix_c0[i]<- wbybeta0[1,1]
beta1matrix_c0[i]<- wbybeta0[2,1]
betasonmatrix_c0<- matrix(1,2,r)
p0=ncol(x)+1
mcdx0=cov.mcd(x, quan=((3*n/4)+1),method=c("mcd"))
rdx0=mahalanobis(x,center=mcdx0$center,cov=mcdx0$cov)
w0<-(rdx0<= qchisq(0.95,p0-1))
wml.out0<-glmrob(y~x,family=binomial,subset=w0)
wmlbeta0<-summary(wml.out0)$coefficients
beta0matrix_d0[i]<- wmlbeta0[1,1]
beta1matrix_d0[i]<- wmlbeta0[2,1]
betasonmatrix_d0<- matrix(1,2,r)
mle.out1<-glm(y~xad1,family=binomial("logit"))
mlebeta1<-summary(mle.out1)$coefficients
beta0matrix_a1[i]<- mlebeta1[1,1]
beta1matrix_a1[i]<-mlebeta1[2,1]
betasonmatrix_a1<- matrix(1,2,r)
mallows.out1 <- glmrob(y~xad1,family=binomial("logit"),method = "Mqle",
weights.on.x ="covMcd")
mallowsbeta1<-summary(mallows.out1)$coefficients
beta0matrix_b1[i]<- mallowsbeta1[1,1]
beta1matrix_b1[i]<- mallowsbeta1[2,1]
betasonmatrix_b1<- matrix(1,2,r)
wby.out1=glmrob(y~xad1,family=binomial,method="WBY")
wbybeta1<-summary(wby.out1) $coefficients
beta0matrix_c1[i]<- wbybeta1[1,1]
```

```
beta1matrix_c1[i]<- wbybeta1[2,1]
betasonmatrix_c1<- matrix(1,2,r)
p1=ncol(xad1)+1
mcdx1=cov.mcd(xad1, quan=((3*n/4)+1),method=c("mcd"))
rdx1=mahalanobis(xad1,center=mcdx1$center,cov=mcdx1$cov)
w1<-(rdx1<= qchisq(0.95,p1-1))
wml.out1<-glmrob(y~xad1,family=binomial,subset=w1)
wmlbeta1<-summary(wml.out1)$coefficients
beta0matrix_d1[i]<- wmlbeta1[1,1]
beta1matrix_d1[i]<- wmlbeta1[2,1]
betasonmatrix_d1<- matrix(1,2,r)
mle.out2<-glm(y~xad2,family=binomial("logit"))
mlebeta2<-summary(mle.out2)$coefficients
beta0matrix_a2[i]<- mlebeta2[1,1]
beta1matrix_a2[i]<-mlebeta2[2,1]
betasonmatrix_a2<- matrix(1,2,r)
mallows.out2 <- glmrob(y~xad2,family=binomial("logit"),method = "Mqle",
weights.on.x ="covMcd")
mallowsbeta2<-summary(mallows.out2)$coefficients
beta0matrix_b2[i]<- mallowsbeta2[1,1]
beta1matrix_b2[i]<- mallowsbeta2[2,1]
betasonmatrix_b2<- matrix(1,2,r)
wby.out2=glmrob(y~xad2,family=binomial,method="WBY")
wbybeta2<-summary(wby.out2) $coefficients
beta0matrix_c2[i]<- wbybeta2[1,1]
beta1matrix_c2[i]<- wbybeta2[2,1]
betasonmatrix_c2<- matrix(1,2,r)
p2=ncol(xad2)+1
mcdx2=cov.mcd(xad2, quan=((3*n/4)+1),method=c("mcd"))
rdx2=mahalanobis(xad2,center=mcdx2$center,cov=mcdx2$cov)
w2<-(rdx2<= qchisq(0.95,p2-1))
wml.out2<-glmrob(y~xad2,family=binomial,subset=w2)
```

```
wmlbeta2<-summary(wml.out2)$coefficients
beta0matrix_d2[i]<- wmlbeta2[1,1]
beta1matrix_d2[i]<- wmlbeta2[2,1]
betasonmatrix_d2<- matrix(1,2,r)
mle.out3<-glm(y~xad3,family=binomial("logit"))
mlebeta3<-summary(mle.out3)$coefficients
beta0matrix_a3[i]<- mlebeta3[1,1]
beta1matrix_a3[i]<-mlebeta3[2,1]
betasonmatrix_a3<- matrix(1,2,r)
mallows.out3 <- glmrob(y~xad3,family=binomial("logit"),method = "Mqle",
weights.on.x ="covMcd")
mallowsbeta3<-summary(mallows.out3)$coefficients
beta0matrix_b3[i]<- mallowsbeta3[1,1]
beta1matrix_b3[i]<- mallowsbeta3[2,1]
betasonmatrix_b3<- matrix(1,2,r)
wby.out3=glmrob(y~xad3,family=binomial,method="WBY")
wbybeta3<-summary(wby.out3) $coefficients
beta0matrix_c3[i]<- wbybeta3[1,1]
beta1matrix_c3[i]<- wbybeta3[2,1]
betasonmatrix_c3<- matrix(1,2,r)
p3=ncol(xad3)+1
mcdx3=cov.mcd(xad3, quan=((3*n/4)+1),method=c("mcd"))
rdx3=mahalanobis(xad3,center=mcdx3$center,cov=mcdx3$cov)
w3<-(rdx3<= qchisq(0.95,p3-1))
wml.out3<-glmrob(y~xad3,family=binomial,subset=w3)
wmlbeta3<-summary(wml.out3)$coefficients
beta0matrix_d3[i]<- wmlbeta3[1,1]
beta1matrix_d3[i]<- wmlbeta3[2,1]
betasonmatrix_d3<- matrix(1,2,r)
mle.out4<-glm(y~xad4,family=binomial("logit"))
mlebeta4<-summary(mle.out4)$coefficients
beta0matrix_a4[i]<- mlebeta4[1,1]
```

```
beta1matrix_a4[i]<-mlebeta4[2,1]
betasonmatrix_a4<- matrix(1,2,r)
mallows.out4 <- glmrob(y~xad4,family=binomial("logit"),method = "Mqle",
weights.on.x ="covMcd")
mallowsbeta4<-summary(mallows.out4)$coefficients
beta0matrix_b4[i]<- mallowsbeta4[1,1]
beta1matrix_b4[i]<- mallowsbeta4[2,1]
betasonmatrix_b4<- matrix(1,2,r)
wby.out4=glmrob(y~xad4,family=binomial,method="WBY")
wbybeta4<-summary(wby.out4) $coefficients
beta0matrix_c4[i]<- wbybeta4[1,1]
beta1matrix_c4[i]<- wbybeta4[2,1]
betasonmatrix_c4<- matrix(1,2,r)
p4=ncol(xad4)+1
mcdx4=cov.mcd(xad4, quan=((3*n/4)+1),method=c("mcd"))
rdx4=mahalanobis(xad4,center=mcdx4$center,cov=mcdx4$cov)
w4<-(rdx4<= qchisq(0.95,p4-1))
wml.out4<-glmrob(y~xad4,family=binomial,subset=w4)
wmlbeta4<-summary(wml.out4)$coefficients
beta0matrix_d4[i]<- wmlbeta4[1,1]
beta1matrix_d4[i]<- wmlbeta4[2,1]
betasonmatrix_d4<- matrix(1,2,r)
mle.out5<-glm(y~xad5,family=binomial("logit"))
mlebeta5<-summary(mle.out5)$coefficients
beta0matrix_a5[i]<- mlebeta5[1,1]
beta1matrix_a5[i]<-mlebeta5[2,1]
betasonmatrix_a5<- matrix(1,2,r)
mallows.out5<- glmrob(y~xad5,family=binomial("logit"),method = "Mqle",
weights.on.x ="covMcd")
mallowsbeta5<-summary(mallows.out5)$coefficients
beta0matrix_b5[i]<- mallowsbeta5[1,1]
beta1matrix_b5[i]<- mallowsbeta5[2,1]
```

```
betasonmatrix_b5<- matrix(1,2,r)
wby.out5=glmrob(y~xad5,family=binomial,method="WBY")
wbybeta5<-summary(wby.out5) $coefficients
beta0matrix_c5[i]<- wbybeta5[1,1]
beta1matrix_c5[i]<- wbybeta5[2,1]
betasonmatrix_c5<- matrix(1,2,r)
p5=ncol(xad5)+1
mcdx5=cov.mcd(xad5, quan=((3*n/4)+1),method=c("mcd"))
rdx5=mahalanobis(xad5,center=mcdx5$center,cov=mcdx5$cov)
w5<-(rdx5<= qchisq(0.95,p5-1))
wml.out5<-glmrob(y~xad5,family=binomial,subset=w5)
wmlbeta5<-summary(wml.out5)$coefficients
beta0matrix_d5[i]<- wmlbeta5[1,1]
beta1matrix_d5[i]<- wmlbeta5[2,1]
betasonmatrix_d5<- matrix(1,2,r)
}
betamatrix_a0<- matrix(rbind(beta0matrix_a0,beta1matrix_a0),2,r)
beta_a0<- (betamatrix_a0- betasonmatrix_a0)
betasum_a0<-matrix(rowSums (beta_a0, na.rm = FALSE, dims = 1),2,1)
betaort_a0<-matrix(betasum_a0/r)
bias_a0<-norm(betaort_a0,type="f")
bias_a0
normkare_a0<-rep(NA,r)
#normhesaplama
for(i in 1:r)
{
normkare_a0[i]= (norm(matrix(beta_a0[,i]),type="f"))^2
  }
d_a0=matrix(normkare_a0)
boxplot(d_a0)
normkaresum_a0<-colSums (d_a0, na.rm = FALSE, dims = 1)
mse_a0<- normkaresum_a0/r
```

```
mse_a0
betamatrix_b0<-matrix(rbind(beta0matrix_b0,beta1matrix_b0),2,r)
beta_b0<- (betamatrix_b0- betasonmatrix_b0)
betasum_b0<-matrix(rowSums (beta_b0, na.rm = FALSE, dims = 1),2,1)
betaort_b0<-matrix(betasum_b0/r)
bias_b0<-norm(betaort_b0,type="f")
bias_b0
normkare_b0<-rep(NA,r)
for(i in 1:r)
{
normkare_b0[i]= (norm(matrix(beta_b0[,i]),type="f"))^2
  }
d_b0=matrix(normkare_b0)
boxplot(d_b0)
normkaresum_b0<-colSums (d_b0, na.rm = FALSE, dims = 1)
mse_b0<- normkaresum_b0/r
mse_b0

betamatrix_c0<-matrix(rbind(beta0matrix_c0,beta1matrix_c0),2,r)
beta_c0<- (betamatrix_c0- betasonmatrix_c0)
betasum_c0<-matrix(rowSums (beta_c0, na.rm = FALSE, dims = 1),2,1)
betaort_c0<-matrix(betasum_c0/r)
bias_c0<-norm(betaort_c0,type="f")
bias_c0
normkare_c0<-rep(NA,r)
for(i in 1:r)
{
normkare_c0[i]= (norm(matrix(beta_c0[,i]),type="f"))^2
  }
d_c0=matrix(normkare_c0)
boxplot(d_c0)
normkaresum_c0<-colSums (d_c0, na.rm = FALSE, dims = 1)
```

```
mse_c0<- normkaresum_c0/r
mse_c0
betamatrix_d0<-matrix(rbind(beta0matrix_d0,beta1matrix_d0),2,r)
beta_d0<- (betamatrix_d0- betasonmatrix_d0)
betasum_d0<-matrix(rowSums (beta_d0, na.rm = FALSE, dims = 1),2,1)
betaort_d0<-matrix(betasum_d0/r)
bias_d0<-norm(betaort_d0,type="f")
bias_d0
normkare_d0<-rep(NA,r)
for(i in 1:r)
{
normkare_d0[i]= (norm(matrix(beta_d0[,i]),type="f"))^2
  }
d_d0=matrix(normkare_d0)
boxplot(d_d0)
normkaresum_d0<-colSums (d_d0, na.rm = FALSE, dims = 1)
mse_d0<- normkaresum_d0/r
mse_d0
```

```
betamatrix_a1<- matrix(rbind(beta0matrix_a1,beta1matrix_a1),2,r)
beta_a1<- (betamatrix_a1- betasonmatrix_a1)
betasum_a1<-matrix(rowSums (beta_a1, na.rm = FALSE, dims = 1),2,1)
betaort_a1<-matrix(betasum_a1/r)
bias_a1<-norm(betaort_a1,type="f")
bias_a1
normkare_a1<-rep(NA,r)
for(i in 1:r)
{
normkare_a1[i]= (norm(matrix(beta_a1[,i]),type="f"))^2
  }
d_a1=matrix(normkare_a1)
boxplot(d_a1)
normkaresum_a1<-colSums (d_a1, na.rm = FALSE, dims = 1)
```

```
mse_a1<- normkaresum_a1/r
mse_a1
betamatrix_b1<-matrix(rbind(beta0matrix_b1,beta1matrix_b1),2,r)
beta_b1<- (betamatrix_b1- betasonmatrix_b1)
betasum_b1<-matrix(rowSums (beta_b1, na.rm = FALSE, dims = 1),4,1)
betaort_b1<-matrix(betasum_b1/r)
bias_b1<-norm(betaort_b1,type="f")
bias_b1
normkare_b1<-rep(NA,r)
for(i in 1:r)
{
normkare_b1[i]= (norm(matrix(beta_b1[,i]),type="f"))^2
  }
d_b1=matrix(normkare_b1)
boxplot(d_b1)
normkaresum_b1<-colSums (d_b1, na.rm = FALSE, dims = 1)
mse_b1<- normkaresum_b1/r
mse_b1
betamatrix_c1<-matrix(rbind(beta0matrix_c1,beta1matrix_c1),2,r)
beta_c1<- (betamatrix_c1- betasonmatrix_c1)
betasum_c1<-matrix(rowSums (beta_c1, na.rm = FALSE, dims = 1),2,1)
betaort_c1<-matrix(betasum_c1/r)
bias_c1<-norm(betaort_c1,type="f")
bias_c1
normkare_c1<-rep(NA,r)
for(i in 1:r)
{
normkare_c1[i]= (norm(matrix(beta_c1[,i]),type="f"))^2
  }
d_c1=matrix(normkare_c1)
boxplot(d_c1)
normkaresum_c1<-colSums (d_c1, na.rm = FALSE, dims = 1)
```

```
mse_c1<- normkaresum_c1/r
mse_c1
betamatrix_d1<-matrix(rbind(beta0matrix_d1,beta1matrix_d1),2,r)
beta_d1<- (betamatrix_d1- betasonmatrix_d1)
betasum_d1<-matrix(rowSums (beta_d1, na.rm = FALSE, dims = 1),2 ,1)
betaort_d1<-matrix(betasum_d1/r)
bias_d1<-norm(betaort_d1,type="f")
bias_d1
normkare_d1<-rep(NA,r)
for(i in 1:r)
{
normkare_d1[i]= (norm(matrix(beta_d1[,i]),type="f"))^2
  }
d_d1=matrix(normkare_d1)
boxplot(d_d1)
normkaresum_d1<-colSums (d_d1, na.rm = FALSE, dims = 1)
mse_d1<- normkaresum_d1/r
mse_d1
```

```
betamatrix_a2<- matrix(rbind(beta0matrix_a2,beta1matrix_a2),2,r)
beta_a2<- (betamatrix_a2- betasonmatrix_a2)
betasum_a2<-matrix(rowSums (beta_a2, na.rm = FALSE, dims = 1),2,1)
betaort_a2<-matrix(betasum_a2/r)
bias_a2<-norm(betaort_a2,type="f")
bias_a2
normkare_a<-rep(NA,r)
normkare_a2<-rep(NA,r)
for(i in 1:r)
{
normkare_a2[i]= (norm(matrix(beta_a2[,i]),type="f"))^2
  }
d_a2=matrix(normkare_a2)
boxplot(d_a2)
```

```
normkaresum_a2<-colSums (d_a2, na.rm = FALSE, dims = 1)
mse_a2<- normkaresum_a2/r
mse_a2
betamatrix_b2<-matrix(rbind(beta0matrix_b2,beta1matrix_b2),2,r)
beta_b2<- (betamatrix_b2- betasonmatrix_b2)
betasum_b2<-matrix(rowSums (beta_b2, na.rm = FALSE, dims = 1),2,1)
betaort_b2<-matrix(betasum_b2/r)
bias_b2<-norm(betaort_b2,type="f")
bias_b2
normkare_b2<-rep(NA,r)
for(i in 1:r)
{
normkare_b2[i]= (norm(matrix(beta_b2[,i]),type="f"))^2
  }
d_b2=matrix(normkare_b2)
boxplot(d_b2)
normkaresum_b2<-colSums (d_b2, na.rm = FALSE, dims = 1)
mse_b2<- normkaresum_b2/r
mse_b2
betamatrix_c2<-matrix(rbind(beta0matrix_c2,beta1matrix_c2),2,r)
beta_c2<- (betamatrix_c2- betasonmatrix_c2)
betasum_c2<-matrix(rowSums (beta_c2, na.rm = FALSE, dims = 1),2,1)
betaort_c2<-matrix(betasum_c2/r)
bias_c2<-norm(betaort_c2,type="f")
bias_c2
normkare_c2<-rep(NA,r)
for(i in 1:r)
{
normkare_c2[i]= (norm(matrix(beta_c2[,i]),type="f"))^2
  }
d_c2=matrix(normkare_c2)
boxplot(d_c2)
```

```
normkaresum_c2<-colSums (d_c2, na.rm = FALSE, dims = 1)
mse_c2<- normkaresum_c2/r
mse_c2
betamatrix_d2<-matrix(rbind(beta0matrix_d2,beta1matrix_d2),2,r)
beta_d2<- (betamatrix_d2- betasonmatrix_d2)
betasum_d2<-matrix(rowSums (beta_d2, na.rm = FALSE, dims = 1),2,1)
betaort_d2<-matrix(betasum_d2/r)
bias_d2<-norm(betaort_d2,type="f")
bias_d2
normkare_d2<-rep(NA,r)
for(i in 1:r)
{
normkare_d2[i]= (norm(matrix(beta_d2[,i]),type="f"))^2
  }
d_d2=matrix(normkare_d2)
boxplot(d_d2)
normkaresum_d2<-colSums (d_d2, na.rm = FALSE, dims = 1)
mse_d2<- normkaresum_d2/r
mse_d2
```

```
betamatrix_a3<- matrix(rbind(beta0matrix_a3,beta1matrix_a3),2,r)
beta_a3<- (betamatrix_a3- betasonmatrix_a3)
betasum_a3<-matrix(rowSums (beta_a3, na.rm = FALSE, dims = 1),2,1)
betaort_a3<-matrix(betasum_a3/r)
bias_a3<-norm(betaort_a3,type="f")
bias_a3
normkare_a3<-rep(NA,r)
for(i in 1:r)
{
normkare_a3[i]= (norm(matrix(beta_a3[,i]),type="f"))^2
  }
d_a3=matrix(normkare_a3)
boxplot(d_a3)
```

```
normkaresum_a3<-colSums (d_a3, na.rm = FALSE, dims = 1)
mse_a3<- normkaresum_a3/r
mse_a3
betamatrix_b3<-matrix(rbind(beta0matrix_b3,beta1matrix_b3),2,r)
beta_b3<- (betamatrix_b3- betasonmatrix_b3)
betasum_b3<-matrix(rowSums (beta_b3, na.rm = FALSE, dims = 1),2,1)
betaort_b3<-matrix(betasum_b3/r)
bias_b3<-norm(betaort_b3,type="f")
bias_b3
normkare_b3<-rep(NA,r)
for(i in 1:r)
{
normkare_b3[i]= (norm(matrix(beta_b3[,i]),type="f"))^2
  }
d_b3=matrix(normkare_b3)
boxplot(d_b3)
normkaresum_b3<-colSums (d_b3, na.rm = FALSE, dims = 1)
mse_b3<- normkaresum_b3/r
mse_b3
betamatrix_c3<-matrix(rbind(beta0matrix_c3,beta1matrix_c3),2,r)
beta_c3<- (betamatrix_c3- betasonmatrix_c3)
betasum_c3<-matrix(rowSums (beta_c3, na.rm = FALSE, dims = 1),2,1)
betaort_c3<-matrix(betasum_c3/r)
bias_c3<-norm(betaort_c3,type="f")
bias_c3
normkare_c3<-rep(NA,r)
for(i in 1:r)
{
normkare_c3[i]= (norm(matrix(beta_c3[,i]),type="f"))^2
  }
d_c3=matrix(normkare_c3)
boxplot(d_c3)
```

```
normkaresum_c3<-colSums (d_c3, na.rm = FALSE, dims = 1)
mse_c3<- normkaresum_c3/r
mse_c3
betamatrix_d3<-matrix(rbind(beta0matrix_d3,beta1matrix_d3),2,r)
beta_d3<- (betamatrix_d3- betasonmatrix_d3)
betasum_d3<-matrix(rowSums (beta_d3, na.rm = FALSE, dims = 1),2,1)
betaort_d3<-matrix(betasum_d3/r)
bias_d3<-norm(betaort_d3,type="f")
bias_d3
normkare_d3<-rep(NA,r)
for(i in 1:r)
{
normkare_d3[i]= (norm(matrix(beta_d3[,i]),type="f"))^2
  }
d_d3=matrix(normkare_d3)
boxplot(d_d3)
normkaresum_d3<-colSums (d_d3, na.rm = FALSE, dims = 1)
mse_d3<- normkaresum_d3/r
mse_d3
```

```
betamatrix_a4<-matrix(rbind(beta0matrix_a4,beta1matrix_a4),2 ,r)
beta_a4<- (betamatrix_a4- betasonmatrix_a4)
betasum_a4<-matrix(rowSums (beta_a4, na.rm = FALSE, dims = 1),2,1)
betaort_a4<-matrix(betasum_a4/r)
bias_a4<-norm(betaort_a4,type="f")
bias_a4
normkare_a4<-rep(NA,r)
for(i in 1:r)
{
normkare_a4[i]= (norm(matrix(beta_a4[,i]),type="f"))^2
  }
d_a4=matrix(normkare_a4)
boxplot(d_a4)
```

```
normkaresum_a4<-colSums (d_a4, na.rm = FALSE, dims = 1)
mse_a4<- normkaresum_a4/r
mse_a4
betamatrix_b4<-matrix(rbind(beta0matrix_b4,beta1matrix_b4),2,r)
beta_b4<- (betamatrix_b4- betasonmatrix_b4)
betasum_b4<-matrix(rowSums (beta_b4, na.rm = FALSE, dims = 1),2,1)
betaort_b4<-matrix(betasum_b4/r)
bias_b4<-norm(betaort_b4,type="f")
bias_b4
normkare_b4<-rep(NA,r)
for(i in 1:r)
{
normkare_b4[i]= (norm(matrix(beta_b4[,i]),type="f"))^2
  }
d_b4=matrix(normkare_b4)
boxplot(d_b4)
normkaresum_b4<-colSums (d_b4, na.rm = FALSE, dims = 1)
mse_b4<- normkaresum_b4/r
mse_b4
betamatrix_c4<-matrix(rbind(beta0matrix_c4,beta1matrix_c4),2,r)
beta_c4<- (betamatrix_c4- betasonmatrix_c4)
betasum_c4<-matrix(rowSums (beta_c4, na.rm = FALSE, dims = 1),2,1)
betaort_c4<-matrix(betasum_c4/r)
bias_c4<-norm(betaort_c4,type="f")
bias_c4
normkare_c4<-rep(NA,r)
for(i in 1:r)
{
normkare_c4[i]= (norm(matrix(beta_c4[,i]),type="f"))^2
  }
d_c4=matrix(normkare_c4)
boxplot(d_c4)
```

```
normkaresum_c4<-colSums (d_c4, na.rm = FALSE, dims = 1)
mse_c4<- normkaresum_c4/r
mse_c4
betamatrix_d4<-matrix(rbind(beta0matrix_d4,beta1matrix_d4),2,r)
beta_d4<- (betamatrix_d4- betasonmatrix_d4)
betasum_d4<-matrix(rowSums (beta_d4, na.rm = FALSE, dims = 1),2 ,1)
betaort_d4<-matrix(betasum_d4/r)
bias_d4<-norm(betaort_d4,type="f")
bias_d4
normkare_d4<-rep(NA,r)
for(i in 1:r)
{
normkare_d4[i]= (norm(matrix(beta_d4[,i]),type="f"))^2
  }
d_d4=matrix(normkare_d4)
boxplot(d_d4)
normkaresum_d4<-colSums (d_d4, na.rm = FALSE, dims = 1)
mse_d4<- normkaresum_d4/r
mse_d4
betamatrix_a5<- matrix(rbind(beta0matrix_a5,beta1matrix_a5),2,r)
beta_a5<- (betamatrix_a5- betasonmatrix_a5)
betasum_a5<-matrix(rowSums (beta_a5, na.rm = FALSE, dims = 1),2,1)
betaort_a5<-matrix(betasum_a5/r)
bias_a5<-norm(betaort_a5,type="f")
bias_a5
normkare_a5<-rep(NA,r)
for(i in 1:r)
{
normkare_a5[i]= (norm(matrix(beta_a5[,i]),type="f"))^2
  }
d_a5=matrix(normkare_a5)
boxplot(d_a5)
```

```
normkaresum_a5<-colSums (d_a5, na.rm = FALSE, dims = 1)
mse_a5<- normkaresum_a5/r
mse_a5
betamatrix_b5<-matrix(rbind(beta0matrix_b5,beta1matrix_b5),2,r)
beta_b5<- (betamatrix_b5- betasonmatrix_b5)
betasum_b5<-matrix(rowSums (beta_b5, na.rm = FALSE, dims = 1),2,1)
betaort_b5<-matrix(betasum_b5/r)
bias_b5<-norm(betaort_b5,type="f")
bias_b5
normkare_b5<-rep(NA,r)
for(i in 1:r)
{
normkare_b5[i]= (norm(matrix(beta_b5[,i]),type="f"))^2
  }
d_b5=matrix(normkare_b5)
boxplot(d_b5)
normkaresum_b5<-colSums (d_b5, na.rm = FALSE, dims = 1)
mse_b5<- normkaresum_b5/r
mse_b5
betamatrix_c5<-matrix(rbind(beta0matrix_c5,beta1matrix_c5),2,r)
beta_c5<- (betamatrix_c5- betasonmatrix_c5)
betasum_c5<-matrix(rowSums (beta_c5, na.rm = FALSE, dims = 1),2,1)
betaort_c5<-matrix(betasum_c5/r)
bias_c5<-norm(betaort_c5,type="f")
bias_c5
normkare_c5<-rep(NA,r)
for(i in 1:r)
{
normkare_c5[i]= (norm(matrix(beta_c5[,i]),type="f"))^2
  }
d_c5=matrix(normkare_c5)
boxplot(d_c5)
```

```
normkaresum_c5<-colSums (d_c5, na.rm = FALSE, dims = 1)
mse_c5<- normkaresum_c5/r
mse_c5
betamatrix_d5<-matrix(rbind(beta0matrix_d5,beta1matrix_d5),2,r)
beta_d5<- (betamatrix_d5- betasonmatrix_d5)
betasum_d5<-matrix(rowSums (beta_d5, na.rm = FALSE, dims = 1),2,1)
betaort_d5<-matrix(betasum_d5/r)
bias_d5<-norm(betaort_d5,type="f")
bias_d5
normkare_d5<-rep(NA,r)
for(i in 1:r)
{
normkare_d5[i]= (norm(matrix(beta_d5[,i]),type="f"))^2
  }
d_d5=matrix(normkare_d5)
boxplot(d_d5)
normkaresum_d5<-colSums (d_d5, na.rm = FALSE, dims = 1)
mse_d5<- normkaresum_d5/r
mse_d5
```

ÖZGEÇMİŞ

Adı Soyadı: Tuğçe PARLAK

Doğum Yeri: ANKARA

Doğum Tarihi: 15.06.1992

Medeni Hali: Bekar

Yabancı Dili: İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise: Aliye Yahşi Anadolu Kız Meslek ve Meslek Lisesi (2006-2010)

Lisans: Ankara Üniversitesi Fen Fakültesi Biyoloji Bölümü (Anadal 2010-2014)

Lisans: Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü (Çift Anadal 2011-2015)

Yüksek Lisans: Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı (2015-2019)

Çalıştığı Kurum/Kurumlar ve Yıl

Eskihisar İnşaat Malzemeleri A.Ş., Finans Departmanı, Finans Personeli (2018-2019)