



**DAĞILIMI BELLİ OLMAYAN DAİRESEL VERİLERDE AYKIRI DEĞER
TESPİTİ İÇİN BİR YÖNTEM ÖNERİSİ**

Sümevra SERT

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

HAZİRAN 2019

Sümeýra SERT tarafından hazırlanan “DAĞILIMI BELLİ OLMAYAN DAİRESEL VERİLERDE AYKIRI DEĞER TESPİTİ İÇİN BİR YÖNTEM ÖNERİSİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İstatistik Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Filiz KARDİYEN

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. M. Akif BAKIR

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Çağdaş Hakan ALADAĞ

İstatistik Ana Bilim Dalı, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 21/06/2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Sena YAŞYERLİ
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Sümeyra SERT

21/06/2019

DAĞILIMI BELLİ OLMAYAN DAİRESEL VERİLERDE AYKIRI DEĞER TESPİTİ İÇİN BİR YÖNTEM ÖNERİSİ

(Yüksek Lisans Tezi)

Sümeyra SERT

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2019

ÖZET

Açısal veya yönsel olarak ölçülmüş dairesel veri tipi, biyoloji, coğrafya, jeoloji, tıp, meteoroloji, jeofizik, okyanus bilimi gibi pek çok alanda karşımıza çıkmaktadır. Aykırı gözlemlerin varlığı, doğrusal verilerde olduğu gibi dairesel veriler için de modelleme ve tahmin sürecinde karşılaşılan önemli sorunlardan bir tanesidir ve yapısı nedeniyle bu tip verilerde aykırı gözlem olması durumu özel bir inceleme gerektirmektedir. Bu tez çalışmasında, dairesel verilerde aykırı gözlem tespiti için, dairesel medyana dayalı bir uzaklık ölçüsü tanımlayan parametrik olmayan bir yöntem önerilmiştir. Önerilen yöntemin performansı tek boyutlu dairesel veriler ve parametrik olmayan dairesel regresyon bağlamında incelenmiştir. Yöntem, tek değişkenli simülatif ve gerçek veri setlerinde literatürde yer alan diğer tespit yöntemleriyle karşılaştırıldığında yüksek performans göstermiştir. Açıklayıcı değişkeni doğrusal ve yanıt değişkeni dairesel olan, doğrusal-dairesel parametrik olmayan regresyon modellerinden Nadaraya Watson ve Yerel Lineer Kernel yöntemleri ele alınarak, yanıt değişkeninde aykırı gözlem bulunması halinde, ilgili yöntemlerin ve önerilen tespit yönteminin performansı, gerçek veri setleri ve kapsamlı bir simülasyon çalışması ile incelenmiştir. Önerilen yöntemin parametrik olmayan dairesel regresyon bağlamında da aykırı gözlemleri yüksek oranlarda tespit ettiği gözlenmiştir.

Bilim Kodu : 20513
Anahtar Kelimeler : Dairesel veri, tek değişkenli, aykırı gözlem, parametrik olmayan dairesel regresyon, sağlam aykırı gözlem tespiti
Sayfa Adedi : 163
Danışman : Doç. Dr. Filiz KARDİYEN

A PROPOSAL METHOD FOR DETECTION OF OUTLIERS IN CIRCULAR DATA
WITH UNKNOWN DISTRIBUTION

(M. Sc. Thesis)

Sümeýra SERT

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

June 2019

ABSTRACT

Angular or directional circular data type appears in many fields such as biology, geography, geology, medicine, meteorology, geophysics, and ocean science. The existence of outliers is one of the most important problems that are encountered in modelling and forecasting process for circular data as well as linear data, and because of the nature of such observations, outliers require a special examination. In this thesis, a non-parametric outlier detection procedure for circular data is proposed which defines a distance based on circular median. The performance of the proposed method was examined in the context of one dimensional circular data and non-parametric circular regression. The proposed method performs better than the other detection methods in the literature for simulative and real data sets in univariate circular data. Linear-circular non-parametric regression models Nadaraya Watson and Local Linear Kernel methods are discussed when the explanatory variable is linear and the response variable is circular, and the performance of the proposed method was examined with a comprehensive simulation study and a real data set when the the response variable contains outlier. It has been observed that the proposed method has high detection rates in the context of non-parametric circular regression.

Science Code : 20513
Key Words : Circular data, univariate, outlier, nonparametric circular regression, robust outlier detection
Page Number : 163
Supervisor : Assoc. Prof. Dr. Filiz KARDIYEN

TEŞEKKÜR

Tez çalışmam sırasında yoğun çalışma temposuna rağmen kıymetli bilgi, birikim ve tecrübeleri ile sabırla bana yol gösteren, sadece akademik anlamda değil her konuda destek olan, hakkını hiçbir zaman ödeyemeyeceğim değerli danışman hocam sayın Doç. Dr. Filiz KARDİYEN'e sonsuz teşekkür ve saygılarımı sunarım. Tez jüri üyeleri değerli hocalarım Prof. Dr. M. Akif BAKIR ve Doç. Dr. Ç. Hakan ALADAĞ'a, değerli öneri ve katkıları için sonsuz teşekkür ve saygılarımı sunarım. Lisans eğitimimin başladığı günden itibaren her koşulda yanımda olup motivasyon ve sonsuz destek kaynağım olan Elif BİLGİN'e, Alparslan DOĞAN'a, lisans ve yüksek lisans eğitimim boyunca yardım, bilgi ve tecrübeleri ile bana sürekli destek olan Atacan ERDİŞ'e, değerli dostlukları için teşekkürü borç bilirim. Tez çalışmam esnasında her türlü kolaylık ve desteği sağlayan Selçuk Üniversitesi Fen Fakültesi Dekanlığı, İstatistik Bölümü ve Aktüerya Bölümü araştırma görevlisi arkadaşlarım ve değerli hocalarıma sonsuz teşekkür ve saygılarımı sunarım. Son olarak, bizleri bugünlere getirebilmek için hiçbir fedakarlıktan kaçınmayan, her zaman arkamızda duran annem Yeter ve babam Adem'e, bana verdikleri en değerli hediyeler olan kardeşlerim Esmâ ve Ayşe'ye sonsuz destekleri için en derin saygı ve sevgilerimi sunarım.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xv
SİMGELER VE KISALTMALAR.....	xv
1. GİRİŞ.....	1
2. DAİRESEL VERİLER	7
2.1. Dairesel Veriler için Tanımlayıcı İstatistikler.....	9
2.2. Önemli Dairesel Olasılık Dağılımları	13
2.2.1. Dairesel düzgün dağılım	14
2.2.2. Von Mises dağılımı.....	15
2.2.3. Cardioid dağılım	16
2.2.4. Wrapped Cauchy dağılımı	17
2.2.5. Wrapped Normal dağılım	19
3. PARAMETRİK OLMAYAN REGRESYON	23
3.1. Kernel Fonksiyonu ve Özellikleri.....	25
3.2. Nadaraya-Watson Tahmin Edicisi	31
3.3. Yerel Lineer Kernel Tahmin Edicisi.....	32
3.4. Tahminin İstatistiksel Özellikleri.....	33
3.5. Kernel Fonksiyonunun Seçimi.....	35
3.6. Bant Genişliği Seçimi	35
3.6.1. Çapraz doğrulama (Cross Validation) yöntemi	37

	Sayfa
3.6.2. Plug-in yöntemi.....	38
3.6.3. Yinelemeli yöntem.....	39
3.7. Parametrik Olmayan Regresyonda Model Seçimi	41
4. DOĞRUSAL-DAİRESEL PARAMETRİK OLMAYAN REGRESYON	43
4.1. Basit Doğrusal-Dairesel Regresyon Analizi	43
4.2. Parametrik Olmayan Doğrusal-Dairesel Regresyon Analizi	45
4.2.1. Model yapısı	46
4.2.2. Doğrusal-Dairesel nadaraya-watson düzgünleştiricisi.....	47
4.2.3. Doğrusal-Dairesel yerel lineer kernel düzgünleştiricisi.....	47
4.2.4. Dairesel-Doğrusal parametrik olmayan regresyonda bant genişliği seçimi.....	48
4.2.5. Gerçek veri uygulaması	49
5. DAİRESEL VERİLERDE AYKIRI GÖZLEM	57
5.1. Tek Değişkenli Dairesel Verilerde Aykırı Gözlem Kavramı.....	59
5.1.1. Mardia istatistiği	59
5.1.2. M istatistiği	59
5.1.3. A ve A* istatistiği.....	60
5.1.4. Chord (Kiriş) istatistiği	61
5.1.5. RCD _u istatistiği	62
5.1.6. Önerilen istatistik.....	63
5.1.7. Simülasyon çalışması.....	64
5.1.8. Gerçek veri seti uygulaması.....	85
5.2. Dairesel Regresyonda Aykırı Gözlem Kavramı ve Tespiti için Önerilen Yöntemler	87
5.2.1. COVRATIO istatistiği	88
5.2.2. DMCE _C istatistiği	88

	Sayfa
5.2.3. DMCEs istatistiđi.....	89
5.2.4. RCD _y istatistiđi	90
5.2.5. Önerilen istatistik.....	91
5.2.6. Simölasyon çalıřması.....	93
5.2.7. Gerçek veri seti uygulaması.....	119
6. SONUÇ VE ÖNERİLER	123
KAYNAKLAR	127
EKLER.....	133
EK-1. n=20 için simölasyon sonuçları.....	134
EK-2. n=40 için simölasyon sonuçları.....	138
EK-3. n=50 için simölasyon sonuçları.....	142
EK-4. n=100 için simölasyon sonuçları.....	146
EK-5. n=200 için simölasyon sonuçları.....	152
ÖZGEÇMİř	163

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Bazı Kernel fonksiyonlarının Epanechnikov Kernel fonksiyonuna göre görelî etkinlikleri	28
Çizelge 4.1. Mavi Deniz Salyangozu (“Periwinkles”) veri seti.....	51
Çizelge 4.2. Mavi Deniz Salyangozu veri setine NW ve LL yöntemlerinin uygulanmasıyla elde edilen risk fonksiyonları	54
Çizelge 5.1. $X \sim VM(3,20)$, $n=20$ olduğunda elde edilen sonuçlar	65
Çizelge 5.2. $n=30$, $X \sim VM(3,20)$ olduğunda elde edilen sonuçlar	66
Çizelge 5.3. $n=60$, $X \sim VM(3,20)$ olduğunda elde edilen sonuçlar	66
Çizelge 5.4. $n=20$, $X \sim VM(3,10)$ olduğunda elde edilen sonuçlar	67
Çizelge 5.5. $n=30$, $X \sim VM(3,10)$ olduğunda elde edilen sonuçlar	67
Çizelge 5.6. $n=60$, $X \sim VM(3,10)$ olduğunda elde edilen sonuçlar	68
Çizelge 5.7. Kurbağalar veri seti.....	85
Çizelge 5.8. Kurbağa (frogs) veri seti için farklı istatistiklerin performanslarının karşılaştırılması	87
Çizelge 5.9. $\gamma=0,70$ olduğunda farklı örneklemeler için simülasyon sonuçları.....	95
Çizelge 5.10. $\gamma=0,80$ olduğunda farklı örneklemeler için simülasyon sonuçları.....	96
Çizelge 5.11. $\gamma=0,90$ olduğunda farklı örneklemeler için simülasyon sonuçları.....	97

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Dairesel verilerin birim çember üzerinde gösterimi	7
Şekil 2.2. Bir gün boyunca (24 saat) bir hastanenin acil servisine gelen hastaların saçılım grafiği (a) ve gül diyagramı ile birlikte gösterimi (b)	8
Şekil 2.3. Dairesel verilerde başlangıç noktası seçimi.....	9
Şekil 2.4. Bir dizi dairese gözlem için ortalama yön ve ortalama bileşke uzunluk	13
Şekil 2.5. Dairesel düzgün dağılımın saçılım grafiği (a) ve gül diyagramı (b)	14
Şekil 2.6. Dairesel düzgün dağılımın yoğunluğu.....	14
Şekil 2.7. Farklı parametreler için VM dağılımının yoğunluğu, (a) $\mu = \pi, \kappa = 0,5$, (b) $\mu = \pi, \kappa = 1$, (c) $\mu = \pi, \kappa = 2$	16
Şekil 2.8. Cardioid dağılımının saçılım grafiği (a) ve gül diyagramı (b)	17
Şekil 2.9. Wrapped Cauchy dağılımının saçılım grafiği (a) ve gül diyagramı (b).....	18
Şekil 2.10. Farklı rho parametreleri için Wrapped Cauchy dağılımının yoğunlukları, (a) $\mu = \pi, \rho = 0.7$, (b) $\mu = \pi, \rho = 0.4$	18
Şekil 2.11. Wrapped Normal dağılımının saçılım grafiği (a) ve gül diyagramı (b).....	19
Şekil 2.12. Wrapped Normal dağılımının yoğunluğu ($\mu = \pi, \rho = 0,4$).....	19
Şekil 2.13. Wrapped Cauchy dağılımının yoğunluğu ($\mu = \pi, \rho = 0,6$).....	20
Şekil 2.14. Farklı parametrelere sahip iki VM dağılımından oluşan karma dağılımın yoğunlukları (a) $\mu = 0, \pi, \kappa = 0.5, 0.5$, (b) $\mu = 3, 5, \kappa = (0.5, 0.5)$	20
Şekil 2.15. Farklı parametrelere sahip üç VM dağılımından oluşan karma dağılımın yoğunlukları (a) $\mu = \pi, \pi, 5\pi, \kappa = 6, 3, 6, p = (0.25, 0.5, 0.25)$, (b) $\mu = \pi, \pi, 5\pi, \kappa = 6, 6, 2, p = (0.4, 0.4, 0.2)$	20
Şekil 2.16. Farklı parametrelere sahip dört VM dağılımından oluşan karma dağılımın yoğunluğu ($\mu = 0, \pi, 2\pi, 3\pi, \kappa = 10, 10, 10, 10, p = (0.25, 0.25, 0.25, 0.25)$)	21
Şekil 2.17. VM ve Wrapped Cauchy dağılımlarından oluşan karma dağılımın yoğunluğu $\mu = \pi, 4\pi/3, \kappa = 5, 0.9, p = (0.25, 0.75)$	21
Şekil 2.18. Cardioid ve Wrapped Cauchy dağılımlarından oluşan karma dağılımın yoğunluğu ($\mu = \pi, 2\pi, \kappa = 0.5, 0.9, p = (0.75, 0.25)$)	21

Şekil	Sayfa
Şekil 2.19. Farklı oran ve parametrelerde VM dağılımından oluşan karma dağılımlarının yoğunlukları (a) $\mu = 2,3,3.5,4,4.5$, $\kappa = 3,3,50,50,50$, $p = 2036,536,536,236,436$ (b) $\mu = \pi 5,3\pi 5, \pi, 7\pi 5,9\pi 5$, $\kappa = 10,10,10,10,10$, $p = (0.2,0.2,0.2,0.2,0.2)$	22
Şekil 3.1. Farklı bant genişlikleri ile normal dağılımdan üretilmiş veriler için histogramlar	26
Şekil 3.2. Farklı bant genişlikleri ile Kernel yoğunlukları.....	36
Şekil 4.1. Mavi deniz salyangozu veri setinin şekil üzerinde gösterimi	50
Şekil 4.2. Uzaklık değişkeninin saçılım grafiği	52
Şekil 4.3. Hareket yönü değişkeninin dairesel saçılım grafiği.....	52
Şekil 4.4. NW ve LL düzgünleştiricileri uygulandığında elde edilen eğriler	53
Şekil 4.5. Elde edilen eğrilerin silindir üzerinde gösterimi	54
Şekil 5.1. Örnek veri setinin doğrusal gösterimi.....	57
Şekil 5.2. Örnek veri setinin dairesel düzlem üzerinde gösterimi	58
Şekil 5.3. Dairesel verilerde uzaklık kavramı	58
Şekil 5.4. Tek değişkenli veri setleri için önerilen yöntemin akış diyagramı.....	64
Şekil 5.5. $\mu = 3, \kappa = 20$, $n=20$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi	68
Şekil 5.6. $\mu = 3, \kappa = 20$, $n=30$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi	68
Şekil 5.7. $\mu = 3, \kappa = 20$, $n=60$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi	69
Şekil 5.8. $\mu = 3, \kappa = 10$, $n=20$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi	69
Şekil 5.9. $\mu = 3, \kappa = 10$, $n=30$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi	69
Şekil 5.10. $\mu = 3, \kappa = 10$, $n=60$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi.....	69
Şekil 5.11. $\mu = 3, \kappa = 20$ ve $n=20$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları.....	70

Şekil	Sayfa
Şekil 5.12. $\mu = 3, \kappa = 20$ ve $n = 20$ olduğunda ilk dört aykırı gözlem için maskeleme oranları.....	71
Şekil 5.13. $\mu = 3, \kappa = 20$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları.....	72
Şekil 5.14. $\mu = 3, \kappa = 20$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için maskeleme oranları.....	73
Şekil 5.15. $\mu = 3, \kappa = 20$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları.....	74
Şekil 5.16. $\mu = 3, \kappa = 20$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için maskeleme oranları.....	75
Şekil 5.17. $\mu = 3, \kappa = 20$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için sahte aykırı gözlem oranları.....	76
Şekil 5.18. $\mu = 3, \kappa = 10$ ve $n = 20$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları.....	77
Şekil 5.19. $\mu = 3, \kappa = 10$ ve $n = 20$ olduğunda ilk dört aykırı gözlem için maskeleme oranları.....	78
Şekil 5.20. $\mu = 3, \kappa = 10$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları.....	79
Şekil 5.21. $\mu = 3, \kappa = 10$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için maskeleme oranları.....	80
Şekil 5.22. $\mu = 3, \kappa = 10$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için sahte aykırı gözlem oranları.....	81
Şekil 5.23. $\mu = 3, \kappa = 10$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları.....	82
Şekil 5.24. $\mu = 3, \kappa = 10$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için maskeleme oranları.....	83
Şekil 5.25. $\mu = 3, \kappa = 10$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için sahte aykırı gözlem oranları.....	84
Şekil 5.26. Kurbağalar veri setinin dairesel düzlemde gösterimi	85
Şekil 5.27. Parametrik olmayan dairesel regresyon analizinde aykırı gözlem tespiti için önerilen yöntemin akış diyagramı	93

Şekil	Sayfa
Şekil 5.28. Bir aykırı gözlem varlığında $n = 20$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL) (a) aykırı gözlem yokken (b) aykırı gözlem varken ...	98
Şekil 5.29. Bir aykırı gözlem varlığında $n = 40$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL) (a) aykırı gözlem yokken (b) aykırı gözlem varken	98
Şekil 5.30. Bir aykırı gözlem varlığında $n = 50$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL) (a) aykırı gözlem yokken (b) aykırı gözlem varken	98
Şekil 5.31. İki aykırı gözlem varlığında $n = 100$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL), (a) aykırı gözlem yokken (b) aykırı gözlem varken ...	99
Şekil 5.32. İki aykırı gözlem varlığında $n = 200$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL), (a) aykırı gözlem yokken (b) aykırı gözlem varken ...	99
Şekil 5.33. $n=20$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları	100
Şekil 5.34. $n=20$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleye oranları	101
Şekil 5.35. $n=20$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları	102
Şekil 5.36. $n=40$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları	103
Şekil 5.37. $n=40$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleye oranları	104
Şekil 5.38. $n=40$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları	105
Şekil 5.39. $n=50$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları	106
Şekil 5.40. $n=50$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleye oranları	107
Şekil 5.41. $n=50$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları	108
Şekil 5.42. $n=100$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları	109
Şekil 5.43. $n=100$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleye oranları	110

Şekil	Sayfa
Şekil 5.44. n=100 olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları	111
Şekil 5.45. n=200 olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları	112
Şekil 5.46. n=200 olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleme oranları	113
Şekil 5.47. n=200 olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları	114
Şekil 5.48. n=20 iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri	115
Şekil 5.49. n=40 iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri	116
Şekil 5.50. n=50 iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri	117
Şekil 5.51. n=100 iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri	118
Şekil 5.52. n=200 iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri	119
Şekil 5. 53. Mavi deniz salyangozu verisine LL yöntemi uygulandığında artıkların dağılımı	120
Şekil 5. 54. Mavi deniz salyangozu verisine LL yöntemi uygulandığında artıkların dağılımı	121

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar açıklamalarıyla birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
AIC	Akaike Bilgi Kriteri
AISB	Asymptotic Integrated Squared Bias
AIV	Asymptotic Integrated Variance
AMISE	Asimptotik Bütünleşik Hata Kareler Ortalaması
BCV	Yanlı Çapraz Doğrulama
CV	Çapraz Doğrulama
DTO	Doğru Tespit Oranı
EÇO	En Çok Olabilirlik Yöntemi
EKK	En Küçük Kareler Yöntemi
EM	Expectation Maximization
F&L	Fisher ve Lee Regresyon Modeli
HL	Hodges-Lehmann Tahmini
ISE	Bütünleşik Hata Kareler
LL	Yerel Lineer Kernel
LRT	Olabilirlik Oran Testi
LSCV	En Küçük Kareler Çapraz Doğrulama
M	Maskeleme Oranı
MISE	Bütünleşik Hata Kareler Ortalaması
MSE	Hata Kareler Ortalaması
NW	Nadaraya Watson
R	Risk fonksiyonu
S	Sahte aykırı gözlem oranı
SPML	Küresel olarak yansıtılan çok değişkenli doğrusal model
VM	Von Mises dağılımı

1. GİRİŞ

Varsayımların sağlanmadığı (normallik vb.) durumlarda regresyon analizi yapabilmek için, verideki rasgeleliği azaltıp aykırı değerleri de dikkate alarak veri setine en uygun eğriyi elde etmek amacıyla çeşitli parametrik olmayan yöntemler önerilmiştir. Bu yöntemlere genel olarak “Düzgünleştirme Yöntemleri” (Smoothing Methods) adı verilmektedir.

$Y = m(.) + e$ regresyon modelinde kullanışlı ve düzgün (smooth) bir $m(.)$ elde etmek için kullanılan yöntemlerde, parametrik yöntemlerdeki gibi kesin varsayımlar kullanılmamakta ve veri seti temel alınarak işlem yapılmaktadır. Literatürde parametrik olmayan regresyon analizi ile ilgili Härdle (1990), Wand ve Jones (1994), Bowman ve Azzalini (1997), Takezawa (2005), Ivanka ve diğerleri (2012) gibi geniş kapsamlı birçok çalışma bulunmaktadır. Dairesel veri analizi, günümüzde gittikçe popüler hale gelen bir konudur. Genellikle biyoloji, jeoloji, coğrafya, meteoroloji gibi alanlarda rastlanan daireysel veya diğer bir deyişle yönsel veriler için istatistiksel analiz yöntemlerinin sürekli olarak geliştiği görülmektedir. Bilgisayar kullanımı ve paket programların kullanımının yaygınlaşması da yöntemlerin geliştirilmesini ve uygulanmasını kolaylaştırmaktadır. Dairesel verilerle ilgili ilk önemli çalışma, Mardia (1972) tarafından yapılmıştır. İlgili çalışmada istatistik teorisinin daireysel veri analizine uygulanması üzerine yoğunlaşmıştır. Ardından Fisher (1993), Mardia ve Jupp (2000), Jammalamadaka ve SenGupta (2001) tarafından daireysel verilerde analiz teknikleri teorik ve uygulamalı olarak geliştirilmiştir. Pewsey ve diğerleri (2013), daireysel veri analiz yöntemlerini hem teorik olarak incelemiş hem de R paket programında bu yöntemlerin uygulamasını detaylı bir şekilde ele almıştır.

Kernel yöntemlerinin daireysel verilere uygulanması, ilk olarak Fisher (1993: 25) tarafından gerçekleştirilmiştir. Bu çalışmada Fisher, Kernel yoğunluk tahmin edicisini parametrik olmayan daireysel veriler için kullanmıştır. Yoğunluk tahmininde en önemli olan parametrenin bant genişliği olduğunu vurgulamış, bant genişliğini etkileyen faktörlerden bahsetmiş ve çok modlu dağılımlar ve simetrik veri setleri için Kernel yöntemlerini düzenlemiştir. Fisher (1993:173), Kernel düzgünleştiricilerini zaman serisi modellerine uyarlamıştır.

Parametrik olmayan dairesel regresyon için ilk uygulama, Di Marzio ve diğerleri (2009) tarafından yapılmıştır. Bu çalışmada dairesel açıklayıcı değişkenler için yerel polinomiyal regresyon yöntemi incelenmiştir. Dairesel Kernel fonksiyonları ile ilgili bilgi verilmiş ve en küçük kareler yöntemi, yerel polinomiyal regresyon modeline uyması için genişletilmiştir. Di Marzio ve diğerleri (2012) tarafından yapılan çalışmada, yanıt değişkeni dairesel olduğunda parametrik olmayan regresyon modellerinden Nadaraya Watson (NW) ve Yerel Lineer Kernel (LL) yöntemleri için model ve ağırlık fonksiyonları tanımlanmıştır. Açıklayıcı değişken doğrusal ve dairesel iken bu yöntemler ayrı ayrı incelenmiştir. Dairesel regresyon ve dairesel zaman serilerinde parametrik olmayan otoregresyon için yaklaşık güven aralıkları oluşturulmuştur. Gerçek veri uygulamaları yapılarak NW, LL, Fisher ve Lee (1992) tarafından önerilen Fisher ve Lee (F&L) regresyon modeli, Presnell ve diğerleri (1998) tarafından önerilen küresel olarak yansıtılan çok değişkenli doğrusal model (SPML) ve doğrudan trigonometrik eğri yöntemlerinin davranışları karşılaştırılmıştır. Salyangoz hareketleri veri seti, Mavi Deniz Salyangozu veri seti ve damlacık hareketi veri setleri, uygulamada kullanılan veri setleridir. Bu çalışma, dairesel yanıtlarda parametrik olmayan regresyonla ilgili ilk çalışmadır. İlgili çalışmanın öncesinde, Di Marzio ve diğerleri (2011), sadece örnek uzay küre olduğunda, çok değişkenli dairesel veri setleri için Kernel yoğunluk tahminini formülize etmiştir. Bant genişliği seçimi için çapraz doğrulama (CV) yöntemine dayalı bootstrap ve plug-in yöntemleri türetilmiştir. Simülasyon çalışması ve gerçek veri seti uygulamaları yapılmış ve bant genişliği seçimi için R programı kullanılmıştır. Taylor (2008), dairesel yoğunluk tahminleri için otomatik bant genişliği seçim yöntemi önermiştir. Yoğunlaşma parametresinin bilinmediği durumda, plug-in yönteminde kullanılan referans yoğunluğa alternatif sağlam yöntemler araştırılmıştır. Yapılan simülasyonlar sonucu, sarılmış (wrapped) dağılımların araştırılan özellikleri taşıdığı görülmüştür. Protein yapısı veri seti ile gerçek veri seti uygulaması yapılmıştır. García-Portugués ve diğerleri (2013), rasgele vektör dairesel ve doğrusal bileşenlerden oluşan bir vektör olduğunda, Kernel yoğunluk tahmini elde etmiştir. Yan, varyans ve MISE türetilmiş ve asimptotik özellikleri incelenmiştir. Oliveira ve diğerleri (2012), dairesel Kernel yoğunluk tahmini için yeni bir plug-in bant genişliği seçim yöntemi önermiştir. Referans dağılımı karma Von Mises dağılımı olarak belirlenmiş, bileşen sayısı Akaike Bilgi Kriteri (AIC) kullanarak belirlenmiştir. Parametre tahminleri ise “Expectation Maximization” (EM) yöntemi ile elde edilmiştir.

Önerilen yöntem Parmak Kuralı (Rule of Thumb) ve En Çok Olabilirlik Çapraz Doğrulama (Likelihood Cross Validation) yöntemleri ile, simülasyon çalışmaları ve gerçek veri seti uygulamaları yapılarak karşılaştırılmıştır. Oliveira ve diğerleri (2013), açıklayıcı değişken dairesel, yanıt değişkeni doğrusal olduğunda yerel trigonometrik ve Nadaraya Watson Kernel tahmin edicilerinin davranışları simülasyonlar yardımıyla karşılaştırmıştır. Bant genişlikleri çapraz doğrulama yöntemi ile elde edilen bu çalışmada doğrusal durum için sözü edilen tahmin edicilerin davranışları karşılaştırılmıştır. Dört farklı model kullanılarak simülasyon çalışmaları yapılmış ve basitten karmaşığa doğru tanımlanan bu modellerin davranışları yorumlanmıştır. Gerçek veri setleriyle uygulamalar yapılmış, yapılan uygulamalar için R programı kullanılmıştır. Qin ve diğerleri (2011), çok değişkenli dairesel-doğrusal regresyon için yeni bir yerel doğrusal Kernel modeli önermiştir. Bant genişliği seçiminde Parmak Kuralı (Rule of Thumb) kullanılmış, bu bant genişliğinin hesaplanmasının CV'ye göre daha kısa sürede gerçekleştiğini ifade etmiştir. Oliveira ve diğerleri (2014), parametrik olmayan yoğunluk tahmini ve regresyon analizi için kullanılacak R program paketi geliştirmişlerdir. Bant genişliği seçimi, parametrik olmayan regresyon analizi yöntemleri, yoğunluk tahmini gibi analizler, bu pakette yer alan komutlarla rahatlıkla uygulanabilmektedir.

Aykırı değer (gözlem) kavramı ile ilgili ilk çalışma, Bernoulli (1977) tarafından yapılmıştır. Bernoulli bu çalışmasında aynı dağılımlı hatalar varsayımını sorgulamıştır. Aykırı değerlerle ilgili istatistiksel bir method ilk olarak Peirce (1852) tarafından önerilmiştir. Wright (1884) tarafından bu çalışma genişletilmiş ve artıkları standart sapmanın 3,37 katını aşan gözlemlerin aykırı gözlem olarak nitelendirilebileceğini öne sürmüştür. 1950'lerin sonuna kadar aykırı gözlem tespiti konusunda fazla bir gelişme yaşanmamıştır. Bunun sebebi, yüksek performansa sahip bilgisayarların mevcut olmamasıdır. Beckman ve Cook (1983) ve Lewis (1984)'in aykırı gözlem üzerine yaptıkları literatür taraması ile birlikte aykırı gözlem ile ilgili yapılan çalışmalar hız kazanmıştır. Bu tarihe kadar yapılan çalışmalar Öklid uzayında çalışılan doğrusal istatistiksel yöntemler içindir. Dairesel verilerde aykırı gözlem ile ilgili yapılan ilk çalışma Graedel (1977)'in rüzgar gülü diyagramını geliştirmek için kutu grafiği çizimini önerdiği çalışmadır. Collett (1980), tek boyutlu dairesel veri setlerinde aykırı değer tespiti için C, D, L ve M istatistiklerini önermiştir. M istatistiği, ilk olarak Mardia (1975) tarafından önerilmiş olup, Collett (1980) bu testi geliştirmiştir.

Fisher (1993), aykırı değere neden olabilecek etkenleri özetlemiş ve M istatistiğini kullanmıştır. Jammalamadaka ve SenGupta (2001), aykırı değer tespiti için P-P plot kullanımını önermiştir.

Ayrıca bazı bilinen dağılımlarda kullanılmak üzere LMPI istatistiğini kullanmış ve LRT (Likelihood Ratio Testing) yaklaşımının aykırı değer olarak kullanılmasını önermiştir.

Veri setinin dağılımı bilindiğinde ve Von Mises olduğunda, ilk olarak μ ve κ parametreleri bilindiğinde ve sadece κ parametresi bilindiğinde LMPI ve LRT istatistiklerinin performanslarını Collett (1980) tarafından önerilen L istatistiği ve M istatistiği ile karşılaştırmış ve önerdikleri yöntemin diğerlerine göre daha iyi sonuç verdiğini vurgulamışlardır. Ancak κ parametresinin bilinmediği durumlar için bir yöntem önermemişlerdir. Otenio ve Anderson-Cook (2005), yaptıkları çalışmada Hodges-Lehmann (HL) tahmininin kullanımının dairesel ortalama ve dairesel medyana göre daha avantajlı olduğunu belirtmişlerdir.

Abuzaid ve diğerleri (2008), dairesel artıkları kullanarak dairesel regresyonda tek bir aykırı gözlem tespiti için sayısal ve grafiksel yöntemler geliştirmiştir. Abuzaid ve diğerleri (2008), dairesel veriler için dayanıklılık katsayısı olarak adlandırılan ve kutu grafiği çizimi için gerekli olan katsayı seçimiyle ilgili çalışmalar yapmış ve aykırı gözlem tespiti için önerdiği kutu grafiğini kullanmış ve S-Plus programına komut olarak eklemiştir. Abuzaid ve diğerleri (2009), tek boyutlu dairesel veri setlerinde aykırı gözlem tespiti için A istatistiğini önermiştir. Abuzaid (2010), tek boyutlu dairesel veri setlerinde tek bir aykırı gözlem tespiti için Chord istatistiğini önermiştir. Abuzaid ve diğerleri (2012), tek boyutlu ve çok boyutlu dairesel veri setlerinde aykırı gözlem tespiti için bir yöntem önermişler ancak önerilen yöntemi herhangi bir istatistiksel ölçü kullanarak test etmemişlerdir. Abuzaid ve diğerleri (2013), dairesel regresyonda aykırı gözlem tespiti için satır silme işlemlerine dayalı MCE ve DMCE istatistiklerini önermiştir. Mohamed ve diğerleri (2016) dairesel veride örnekleme teorisine dayalı bir aykırı gözlem tespit yöntemi önermiştir ancak bu yöntem yüksek oranda sahte aykırı gözlem tespit etmektedir. Mahmood ve diğerleri (2017), tek boyutlu dairesel veri setlerinde aykırı gözlem tespiti için dairesel uzaklığa dayalı sağlam bir yöntem önermiş ve bu yöntemi Mardia, M, A ve Chord yöntemleriyle karşılaştırmıştır. Önerilen yöntemin birden fazla aykırı gözlem tespitinde de başarılı sonuçlar verdiği ifade edilmiştir. Mahmood ve diğerleri (2017), dairesel

regresyonda aykırı gözlem tespiti için dairesel medyana dayalı sağlam bir yöntem önermiştir.

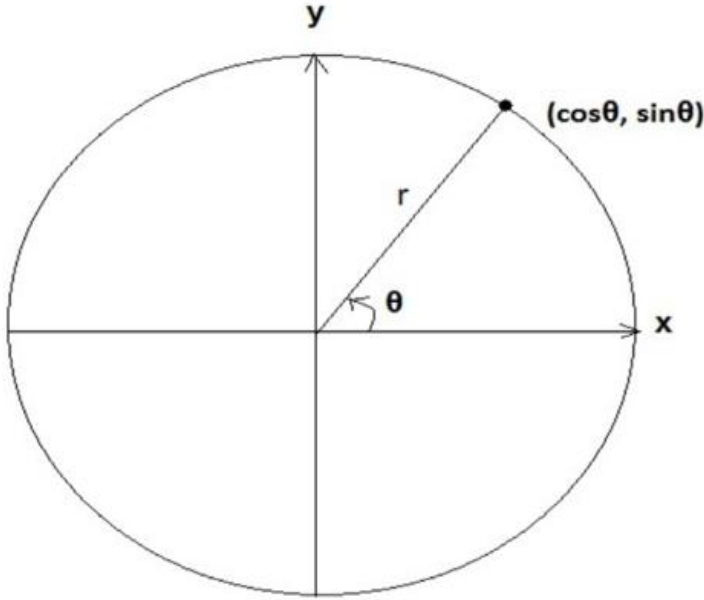
Önerilen yöntemin birden fazla aykırı gözlem varlığında da güzel sonuçlar verdiği ifade edilmiştir. Yukarıda bahsedilen yöntemlerin hepsi, veri setinin dağılımının bilindiği ve Von Mises (VM) olduğu durumlar için önerilmiştir.

Dairesel verilerde aykırı gözlem tespiti ile ilgili yapılan en kapsamlı çalışmalardan biri olan Mahmood ve diğerleri (2017), aykırı gözlem tespiti için kullanılacak kesim noktasının (cut-off point) belirlenmesi için, dağılımın Von Mises olduğu varsayımından yararlanmıştır. Veri setinin dağılımıyla ilgili herhangi bir önsel bilgi yoksa veya varsayımlar sağlanmıyorsa, yukarıda önerilen yöntemlerin kullanılması doğru sonuçlar vermeyecektir.

Bu tez çalışmasında, doğrusal-dairesel parametrik olmayan regresyon analiz yöntemlerinden Nadaraya Watson ve Yerel Lineer Kernel tahmin edicilerinin aykırı gözlem varlığında davranışları incelenmiştir. Tek değişkenli dairesel veri setleri ve parametrik olmayan dairesel regresyon analizinde aykırı gözlem tespiti için uygulanabilecek bir yöntem önerilmiştir. Önerilen yöntemin geçerliliği simülasyon çalışmaları ve gerçek veri seti uygulamaları ile gösterilmiştir. Uygulamalar için MATLAB ve açık kaynak kodlu “R” programı kullanılmıştır. R programında “circular”, “CircStats”, “Directional”, “NPCirc”, “stats”, “CircOutlier”, “boot”, “ggplot” paketlerinden yararlanılmıştır.

2. DAİRESEL VERİLER

Dairesel veri, istatistiğin alt dalı olup yön, açı ve rotasyonlarla ilgilidir. Geleneksel istatistiksel yöntemlerde verilerin gerçek sayı dizisi üzerinde yer aldığı varsayılmaktadır. Dairesel istatistikte ise veriler, belli bir referans noktası temel alınarak ölçülen ve birim çember üzerinde olduğu varsayılan açılardan oluşmaktadır. Veri seti dairesele olmasa bile, çeşitli dönüşümler kullanılarak dairesele hale getirilebilir. Yapılan dönüşümler yorumlama aşamasında dikkate alınmalıdır. Dairesel veriler, problemin tanımına göre radyan cinsinden $[0, 2\pi)$ aralığında veya derece cinsinden $[0, 360)$ aralığındadır (Jammalamadaka ve SenGupta, 2001:1-6).



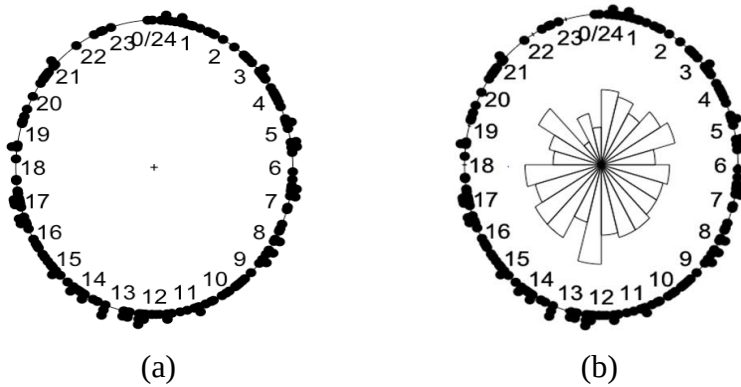
Şekil 2.1. Dairesel verilerin birim çember üzerinde gösterimi

Doğası gereği dairesele veri analizi bilindik standart istatistiksel yöntemler kullanılarak yapılamamaktadır. Örneğin, 310° ve 50° olarak ölçülen iki açısal gözlem değeri olsun. Standart istatistiksel yöntemlere göre bu iki gözlemin aritmetik ortalamasının 180° olması gerekir, oysa ki dairesele ortalamalar için doğru sonuç 0° derece olmalıdır. Bu yüzden dairesele veri analizi yapabilmek için olasılık dağılımları, tanımlayıcı istatistikler ve istatistiksel tahmin yöntemleri için özel teknikler gerekmektedir. Dairesel veriler biyoloji, jeoloji, coğrafya, sağlık, meteoroloji, okyanus bilimi, sinirbilim, tıp, inşaat mühendisliği gibi birçok alanda kullanılmaktadır (Di Marzio ve diğerleri, 2012).

Dairesel veriler iki alt gruba ayrılmaktadır. Birincisi yönsel verilerdir. Bu tür veriler, belli bir nokta referans noktası seçilerek, birim çember gibi bir düzlem üzerinde genel olarak yön analizi yapmak için kullanılmaktadır. Örneğin, Ankara’da herhangi bir günde esen rüzgarın yönü hakkında bilgi sahibi olabilmek için yönsel (directional) veri kullanılır. Diğer bir tür daireysel veri ise periyodik verilerdir. Bilinen bir periyotta, bir periyot, bir ay veya bir yıl olan periyodik bir olay, döngünün bu rotasyona karşılık geldiği bir çember üzerinde birkaç kişi veya periyot üzerinden veri toplayarak temsil edilebilir (Jammalamadaka ve SenGupta, 2001:1-6). Bu tür veriler için en yaygın örnek, bir gün içerisinde çağrı merkezine gelen -telefonların geliş zamanı (saat ve dakika cinsinden), mobese kameralarının görüntü açısına göre kayıt kalitesi olarak verilebilir.

Dairesel istatistikte başlıca veri toplama araçları pusula, saat ve takvimdir. Pusula yön içeren verileri ölçmek için kullanılırken, saat ve takvim periyodik verileri ölçmek için kullanılmaktadır.

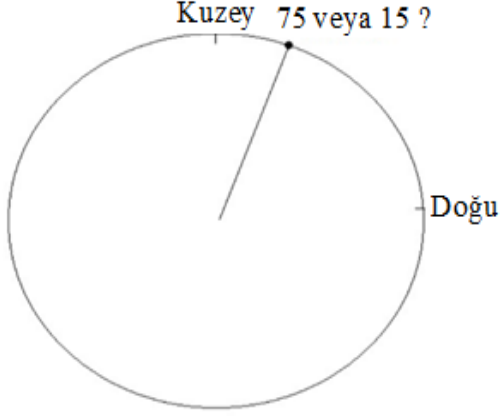
Dairesel verilerin saçılım grafikleri ve histogramların gösterimi, doğrusal durumda kullanılan verilerden farklıdır. Şekil (2.2)’de, bir hastanenin acil servisine gelen hastaların geliş saatlerine göre (24 saat boyunca) elde edilen verilerin saçılım grafiği (a) ve saçılım grafiğinin gül diyagramı ile birlikte gösterimi verilmiştir. Gül diyagramı, doğrusal veriler için kullanılan histograma karşılık gelmektedir.



Şekil 2.2. Bir gün boyunca (24 saat) bir hastanenin acil servisine gelen hastaların saçılım grafiği (a) ve gül diyagramı ile birlikte gösterimi (b)

Dairesel verilerle analiz yapılırken dikkat edilmesi gereken en önemli noktalar, başlangıç noktası ve yön seçimidir.

Veriler saat yönünde ve saat yönünün tersinde olmak üzere iki şekilde yerleştirilebilmekte ve bu iki analiz farklı sonuçlar vermektedir (Pewsey ve diğerleri, 2013:2). Şekil (2.3)'te, verilen açısal bir gözlemin değerinin, seçilen başlangıç noktası ve yöne göre değişimi görülmektedir.



Şekil 2.3. Dairesel verilerde başlangıç noktası seçimi

Şekil (2.3)'te, başlangıç noktası kuzey ve yön olarak saat yönü alındığında bu açısal gözlemin değeri 15° olarak elde edilirken, başlangıç noktası doğu ve yön olarak saat yönünün tersi alındığında ise aynı açısal gözlemin değerinin 75° olarak elde edildiği görülmektedir.

2.1. Dairesel Veriler için Tanımlayıcı İstatistikler

Bir veri kümesi kullanarak analiz yapabilmek için, veri kümesinin dağılımı çarpıklığı, basıklığı, mod, medyan, bir aykırı değer ya da değerler olup olmadığı hakkında bilgi sahibi olmak gerekir. Bu istatistikleri dikkate alarak işlemlerin başlaması, analistin yanlış sonuçlara ulaşmasını önlemektedir. Yukarıda verilen dairesel verilerin ortalamasının doğrusal verilerin ortalamasından farklı olduğu örneğinden de anlaşılacağı üzere, dairesel veri analizi bilinen istatistiksel yöntemlerden farklı şekillerde yapılmaktadır. Bu tür verilerin analizinin yapılabilmesi için farklı istatistiksel yöntemler geliştirilmiştir (Fisher, 1993; Mardia ve Jupp, 2000; Jammalamadaka ve SenGupta, 2001). $\theta_1, \theta_2, \dots, \theta_n$, n boyutlu birim çember üzerinden alınan rasgele dairesel gözlemler olsun. Her bir gözlem için $x_i = \cos(\theta_i)$, $y_i = \sin(\theta_i)$ kutupsal dönüşümü yapıldığında elde edilen $C = \sum_{i=1}^n \cos(\theta_i)$, $S = \sum_{i=1}^n \sin(\theta_i)$ eşitlikleri kullanılarak bileşke vektörün uzunluğu eşitlik

(2.1)'de verildiği gibi elde edilmektedir. Burada tanımlanan (C, S) , ağırlık merkezini temsil etmektedir.

$$R = \|R\| = \sqrt{C^2 + S^2} \quad (2.1)$$

R bileşke uzunluk, $\bar{R} = \frac{R}{n}$ ise ortalama bileşke uzunluk olarak adlandırılır. $0 \leq R \leq n$ iken, $0 < \bar{R} < 1$ 'dir. $\bar{R} = 1$ olması, bütün veri noktalarının çakışık olduğu anlamına gelmektedir. Ancak $\bar{R} = 0$ olması, veri setinin çember etrafında düzgün bir şekilde dağıldığını göstermemektedir. $\bar{R}$, veri seti tek bir gruptan oluşmadığı sürece yayılım hakkında bilgi edinmek için kullanılabilecek yararlı bir gösterge değildir (Fisher, 1993:32).

R bileşke vektörünün yönü “dairesel ortalama yön” olarak adlandırılmakta ve $\bar{\theta} = \arctan^*\left(\frac{S}{C}\right)$ olmak üzere aşağıdaki gibi tanımlanmaktadır:

$$\bar{\theta} = \begin{cases} \arctan\left(\frac{S}{C}\right) & , & C > 0, S \geq 0 \\ \arctan\left(\frac{S}{C}\right) + \pi & , & C < 0 \\ \frac{\pi}{2} & , & C = 0, S > 0 \\ \arctan\left(\frac{S}{C}\right) + 2\pi & , & C \geq 0, S > 0 \\ \text{tanımsız} & , & C = 0, S = 0 \end{cases} \quad (2.2)$$

Eşitlik (2.2)'den anlaşılacağı üzere, dairesel yön aritmetik ortalamaya eşit değildir. Ortalama yön ile ortalama bileşke uzunluk arasındaki ilişki, eşitlik (2.3)'te ifade edilmiştir:

$$\cos(\bar{\theta}) = \frac{C}{R}, \sin(\bar{\theta}) = \frac{S}{R} \quad (2.3)$$

Ortalama yön $\bar{\theta}$, açıların ortalamasıdır. Dairesel verinin konum parametresi ölçüsü olarak kullanılmaktadır. (Jammalamadaka ve SenGupta, 2001: 14) ‘te yer alan önerme 1.1’de elde edilen sonuçlara göre, başlangıç noktası ve yön seçimi, dairesel ortalama yönü etkilemektedir.

Önemli konum parametrelerinden birisi de dairesel medyandır. Dairesel medyayı elde etmek için veri setini iki eşit parçaya bölen eksen seçilir ve ortalama yönden daha küçük değer veren uç seçilir. Eğer örneklem boyutu (n) tek sayı ise dairesel medyan bir veri üzerinden geçmektedir. Eğer çift sayı ise, Öklid uzayındaki veri setlerinde uyguladığımız gibi iki nokta arasındaki nokta dairesel medyayı vermektedir. Dairesel medyanın değeri tek modlu veri setleri için tek bir değer iken, çok modlu ya da izotropik veri setlerinde tek bir değer almayabilir. Dairesel medyan, eşitlik (2.4)'ü minimize eden değer olarak tanımlanmıştır (Fisher, 1993:36):

$$d(\bar{\theta}) = \pi - \frac{1}{n} \sum_{i=1}^n |\pi - |\theta_i - \bar{\theta}|| \quad (2.4)$$

Kullanılan veri setinin dağılımı bilindiğinde ise Mardia (1972), dairesel medyayı(φ) aşağıdaki eşitliğin çözümü olarak tanımlamıştır:

$$\int_{\varphi}^{\varphi+\pi} f(\theta) d\theta = \int_{\varphi+\pi}^{\varphi+2\pi} f(\theta) d\theta = 0,5 \quad (2.5)$$

Lenth(1981) ve Wherly ve Shine (1981), etki eğrilerini kullanarak hem dairesel ortalama yön hem de dairesel medyanın sağlamlık özelliklerini incelemiş ve dairesel ortalama yönün doğrusal veri setlerindeki ortalamaya göre daha sağlam olduğunu ifade etmiştir. Durcharme ve Milasevic (1987), aykırı gözlem varlığında dairesel medyanın ortalama yöne göre daha etkin olduğunu göstermiştir. He ve Simpson (1992), özellikle veri setinin dağılımının Von Mises olmadığı durumlarda, dairesel medyanın tercih edilen yönün tahmini olarak kullanılmasını savunmaktadır (Otieno ve Anderson-Cook, 2006).

Dairesel varyans V ise, açıların ortalama yön etrafındaki varyasyonunu ölçmektedir. V değeri küçüldükçe dağılım yoğunlaştığından, doğrusal veriler için kullanılan varyans kavramıyla benzer özellikler göstermektedir. α açısı sıfır (başlangıç) yönü olmak üzere, θ_i noktalarının α açısı etrafında yayılımının ölçüsü olarak $D = \frac{1}{n} \sum_{i=1}^n (1 - \cos(\theta_i - \alpha))$ alınıp, bu eşitlikte sıfır yönü $\bar{\theta}$ alındığında, eşitlik (2.6)'daki varyans ifadesi elde edilmektedir.

$$V = \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) \quad (2.6)$$

Eşitlik (2.6)'da kosinüs açılımı kullanılarak eşitlik (2.7) elde edilmektedir.

$$V = 1 - \frac{1}{n} \sum_{i=1}^n (\cos(\theta_i) \cos(\bar{\theta}) + \sin(\theta_i) \sin(\bar{\theta})) \quad (2.7)$$

$\sum_{i=1}^n \cos(\theta_i) = C$ ve $\sum_{i=1}^n \sin(\theta_i) = S$ dönüşümü kullanılarak eşitlik (2.7), aşağıdaki gibi yeniden yazılabilir:

$$V = 1 - \frac{1}{n} (C \cos(\bar{\theta}) + S \sin(\bar{\theta})) \quad (2.8)$$

Eşitlik (2.3) kullanılarak eşitlik (2.8) yeniden düzenlenirse,

$$\begin{aligned} V &= 1 - \frac{1}{n} (R \cos(\theta) \cos(\theta) + R \sin(\theta) \sin(\theta)) \\ &= 1 - \frac{1}{n} (R \cos^2(\theta) + R \sin^2(\theta)) \\ &= 1 - \frac{1}{n} (R (\cos^2(\theta) + \sin^2(\theta))) \\ &= 1 - \frac{1}{n} R \quad \text{elde edilir. Böylece eşitlik (2.9)'da yer alan ifade elde edilmiştir (Mardia, 1972:23-24).} \end{aligned}$$

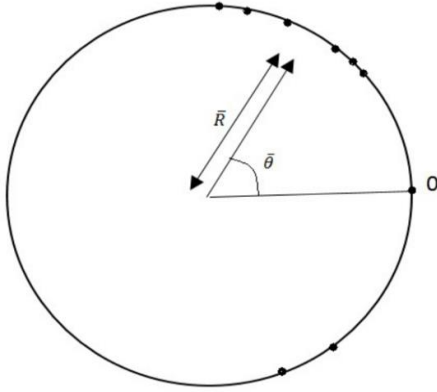
$$V = 1 - \bar{R} \quad (2.9)$$

Eşitlik (2.9)'dan yola çıkarak, ortalama bileşke uzunluk ve dairesel varyans arasında yakın bir ilişki olduğu görülmektedir. Eğer V sıfıra yakınsa, doğrusal verilerdeki gibi verinin oldukça konsantre olduğu sonucuna varabiliriz. Ancak doğrusal verilerin aksine, $0 \leq V \leq 1$ 'dir (Fisher, 1993:32).

Dairesel örnek standart sapması eşitlik (2.10)'da ifade edilmiştir. Dairesel varyans, $[0,1]$ aralığında değer aldığından, standart sapma değerinin $(0, \infty)$ aralığına uygun bir dönüşümü eşitlik (2.10)'de verilmiştir (Mardia, 1972:74).

$$v = \sqrt{-2 \ln(\bar{R})} \quad (2.10)$$

Şekil (2.4)'te, ortalama yön ve ortalama bileşke uzunluk kavramları ifade edilmektedir(Mardia,1972).



Şekil 2.4. Bir dizi dairesel gözlem için ortalama yön ve ortalama bileşke uzunluk

2.2. Önemli Dairesel Olasılık Dağılımları

Dairesel verilerin yapısı gereği olasılık dağılımları başlangıç noktasına bağlıdır. Olasılık dağılımlarını belirlemede dikkat edilmesi gereken iki önemli nokta vardır. Bunlardan ilki, seçilen başlangıç yönü, yönün belirlenmesi ve kullanılan ölçü birimleridir. İkincisi, periyodikliğe bağlı olarak, birim çember üzerinde θ açısının $\theta + 2k\pi$, $k = 0, \pm 1, \pm 2, \dots$ noktalarında aynı değerlere sahip olmasıdır. Yani 40° ile 760° , birim çember üzerinde aynı noktaya tekabül etmektedir. Bu hususlar göz önüne alındığında dairesel dağılım fonksiyonu aşağıdaki gibi tanımlanmaktadır:

$F(\theta) = P(0 < \Theta < \theta)$, $0 \leq \theta \leq 2\pi$ ve $F(\theta + 2\pi) - F(\theta) = 1$, $-\infty \leq \theta \leq \infty$ 'dur. Ayrıca $F(0) = 0$ ve $F(2\pi) = 1$ 'dir. Aşağıdaki özellikleri sağlayan bir θ değişkeni f olasılık yoğunluk fonksiyonuna sahiptir:

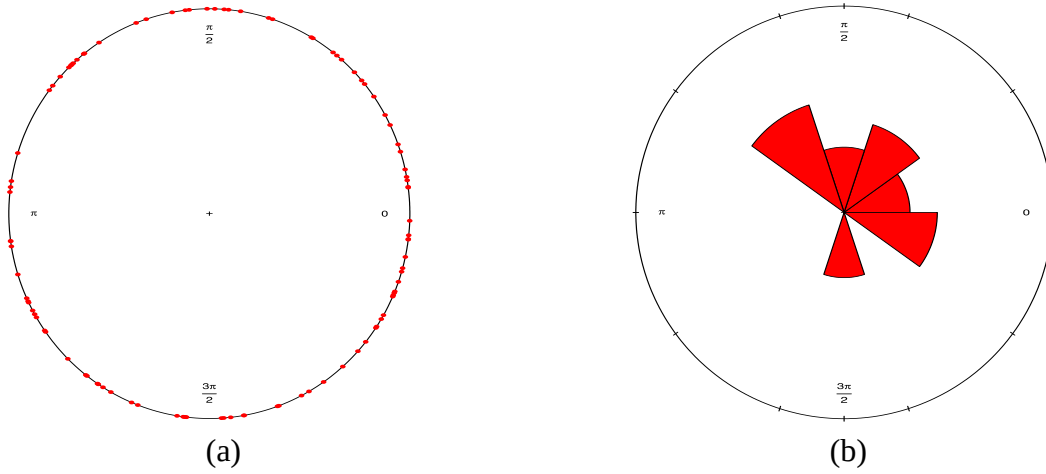
1. $(-\infty, \infty)$ aralığında $f(\theta) \geq 0$,
2. $(-\infty, \infty)$ aralığında $f(\theta + 2\pi) = f(\theta)$,
3. $\int_0^{2\pi} f(\theta) d\theta = 1$ 'dir (Pewsey ve diğerleri, 2013: 36-37).

Bu bölümde, uygulamada en çok kullanılan dairesel dağılımlara yer verilmiştir.

2.2.1. Dairesel düzgün dağılım

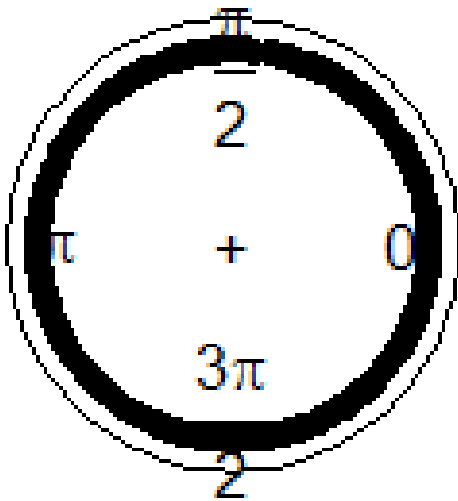
Dairesel düzgün dağılımda, 0 ile 360 derece veya 0 ile 2π arasındaki bütün ortalama yönler, eşit olasılık değerini almaktadır (Fisher 1993: 43). Dairesel Düzgün dağılım eşitlik (2.11) ile tanımlanan olasılık yoğunluk fonksiyonuna sahiptir.

$$f(\theta) = \frac{1}{2\pi}, 0 \leq \theta < 2\pi \quad (2.11)$$



Şekil 2.5. Dairesel düzgün dağılımın saçılım grafiği (a) ve gül diyagramı (b)

Dairesel düzgün dağılımdan üretilen verilerin, daireesel yoğunluğu ile birlikte gösterimi şekil (2.6)'da verilmiştir.



Şekil 2.6. Dairesel düzgün dağılımın yoğunluğu

2.2.2. Von Mises dağılımı

Dairesel istatistikte en çok kullanılan dağılım olup, daireysel veriler için doğrusal verilerde kullanılan normal dağılıma benzer özellik göstermektedir. Dairesel normal dağılım da benzer özelliklere sahiptir ancak Von Mises dağılımının parametrelerinin tahmin edilmesi daha kolay olduğundan daireysel verilerle ilgili uygulamalar yapılırken varsayımlar genellikle bu dağılım üzerine kurulmuştur (Fisher, 1993 :48). Olasılık yoğunluk fonksiyonu eşitlik (2.12)'te, dağılım fonksiyonu ise eşitlik (2.13)'te verilmiştir.

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)], 0 \leq \theta \leq 2\pi, 0 \leq \kappa < \infty \quad (2.12)$$

$$F(\theta) = [2\pi I_0(\kappa)]^{-1} \int_0^\theta \exp[\kappa \cos(\phi - \mu)] d\phi \quad (2.13)$$

Burada kullanılan Bessel fonksiyonu, normalleştirici bir sabit olup yoğunlaşma parametresine bağlıdır. Bu dağılım, Jones-Pewsey ailesinin sınırlayıcı bir örneğidir (Pewsey ve diğerleri, 2013: 56). Eşitlik (2.14)'te, modifiye edilmiş sıfırıncı sıradan Bessel fonksiyonu verilmiştir (Pewsey ve diğerleri, 2013: 56).

$$I_0(\kappa) = (2\pi)^{-1} \int_0^{2\pi} \exp[\kappa \cos(\phi - \mu)] d\phi \quad (2.14)$$

Von Mises dağılımının μ ve κ olmak üzere bilinmeyen iki parametresi vardır. Parametre tahminleri genel olarak En Çok Olabilirlik Yöntemi ile yapılmaktadır. Burada μ parametresi ortalama yön, κ parametresi ise yoğunlaşma parametresidir. Dairesel gözlemlerde ortalama yön tahminini veren eşitlikler, Fisher (1993: 34) tarafından eşitlik (2.15)'teki gibi tanımlanmıştır.

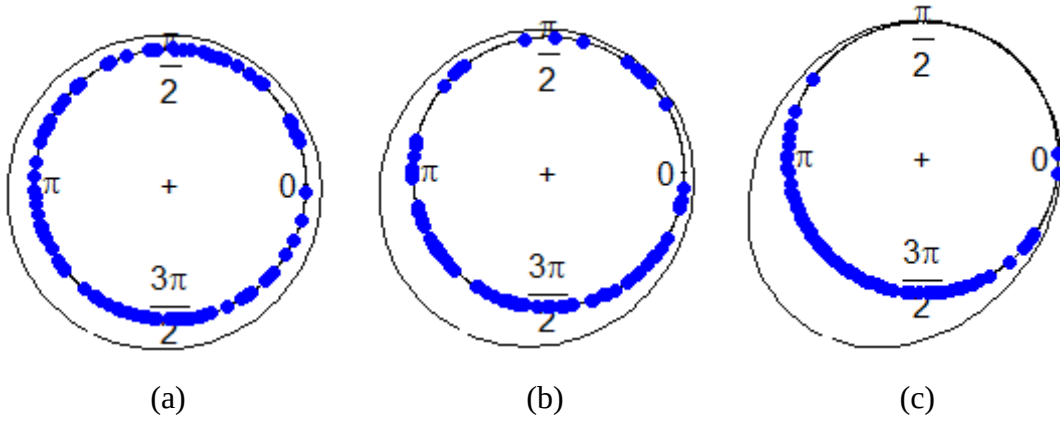
$$\hat{\mu} = \begin{cases} \tan^{-1}\left(\frac{s}{c}\right) & , s > 0, c > 0 \\ \tan^{-1}\left(\frac{s}{c}\right) + \pi & , c < 0 \\ \tan^{-1}\left(\frac{s}{c}\right) + 2\pi & , s < 0, c > 0 \end{cases} \quad (2.15)$$

Burada $s = \sum_{i=1}^n \sin(\theta_i)$, $c = \sum_{i=1}^n \cos(\theta_i)$ 'dir.

κ parametresi ise eşitlik (2.16) kullanılarak tahmin edilebilmektedir:

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + \frac{5}{6}\bar{R}^5 & , \quad \bar{R} < 0,53 \\ -0,4 + 1,39\bar{R} + \frac{0,43}{(1-\bar{R})} & , \quad 0,53 \leq \bar{R} < 0,85 \\ (\bar{R}^3 - 4\bar{R}^2 + 3\bar{R})^{-1} & , \quad \bar{R} > 0,85 \end{cases} \quad (2.16)$$

Şekil (2.7)'de farklı μ ve κ parametreleri kullanılarak Von Mises dağılımından elde edilen şekiller verilmiştir.



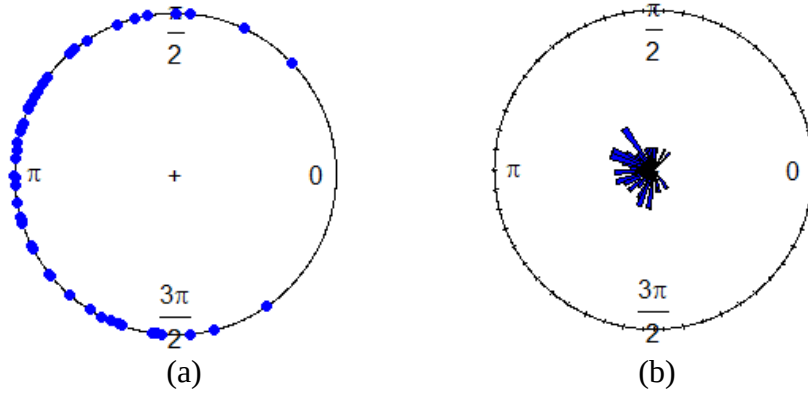
Şekil 2.7. Farklı parametreler için VM dağılımının yoğunluğu, (a) $\mu = \pi, \kappa = 0,5$, (b) $\mu = \pi, \kappa = 1$, (c) $\mu = \pi, \kappa = 2$

Şekillerden de görüldüğü üzere, κ parametresi azaldıkça veriler yayılmaya ve dairesel düzgün dağılıma yakınsamaya, κ parametresi arttıkça ise veriler ortalama yön etrafında yoğunlaşmaya başlamıştır. Bu durum κ parametresinin, normal dağılımın varyans parametresinin tam tersi davranış sergilediğini göstermektedir.

2.2.3. Cardioid dağılım

Bu dağılım Jeffreys tarafından önerilmiştir (Jeffreys, 1961: 328). Tek modlu ve μ etrafında simetrik bir dağılımdır. $C(\mu, \rho)$ ile gösterilir. Burada ρ , ortalama bileşke uzunluğu temsil etmektedir. $\rho > 0$ olduğunda ortalama yön μ ile ifade edilir. $\rho = 0$ olduğunda ise bu dağılım düzgün dairesel dağılıma dönüşür. Olasılık yoğunluk fonksiyonu eşitlik (2.17)'de verilmiştir.

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} (1 + 2\rho \cos(\theta - \mu)), 0 \leq \theta < 2\pi, |\rho| \leq \frac{1}{2} \quad (2.17)$$



Şekil 2.8. Cardioid dağılımının saçılım grafiği (a) ve gül diyagramı (b)

Sarma dağılımlar, doğrusal dağılımların birim çember etrafında sarılmasıyla elde edilmektedir. X , Öklid uzayında $g(x)$ olasılık yoğunluk fonksiyonuna ve $F(x)$ dağılım fonksiyonuna sahip bir rasgele değişken olsun. Dairesel bir rasgele değişken olan Θ 'yı ; $\Theta \equiv X[\text{mod } 2\pi]$ şeklinde tanımlayarak elde edebiliriz.

Θ 'nın $f(\theta)$ olasılık yoğunluk fonksiyonu, $g(x)$ 'in birim yarıçaplı çember etrafında sarılmasıyla eşitlik (2.18)'de verildiği gibidir.

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2k\pi) \quad (2.18)$$

Θ 'nın dağılım fonksiyonu ise eşitlik (2.19)'da verilmiştir.

$$F(\theta) = \sum_{k=-\infty}^{\infty} [G(\theta + 2k\pi) - G(2\pi k)] \quad , 0 \leq \theta < 2\pi \quad (2.19)$$

Birim çember üzerindeki dağılımların birçoğu bu yöntemle elde edilebilmektedir (Fisher,1993).

2.2.4. Wrapped Cauchy dağılımı

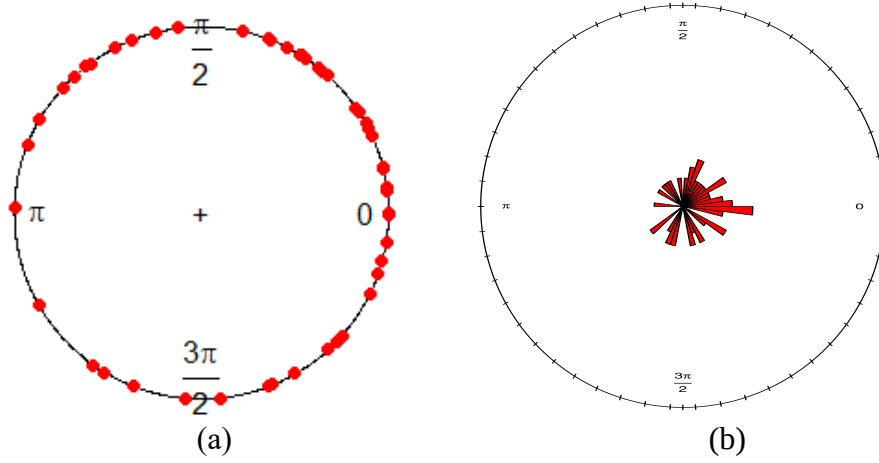
Tek modlu ve simetrik bir dağılımdır. Cauchy dağılımının birim çember etrafında sarılmasıyla elde edilebilir. Olasılık yoğunluk fonksiyonu eşitlik (2.20)'de tanımlanmıştır.

$$f(\theta) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}, \quad 0 \leq \theta < 2\pi, 0 \leq \rho \leq 1 \quad (2.20)$$

Dağılım fonksiyonu ise eşitlik (2.21)'de verildiği gibidir.

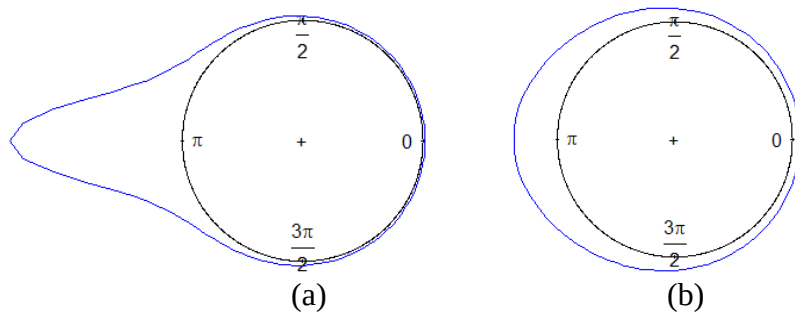
$$F(\theta) = \frac{1}{2\pi} \cos^{-1} \left(\frac{(1+\rho^2)\cos(\theta-\mu)-2\rho}{1+\rho^2-2\rho\cos(\theta-\mu)} \right), \quad 0 \leq \theta < 2\pi \quad (2.21)$$

$\rho \rightarrow 0$ iken Wrapped Cauchy dağılımı Düzgün dağılıma yakınsar. $\rho \rightarrow 1$ iken ise dağılım μ yönü etrafında yoğunlaşma eğilimindedir (Fisher, 1993: 46).



Şekil 2.9. Wrapped Cauchy dağılımının saçılım grafiği (a) ve gül diyagramı (b)

Farklı parametreler kullanılarak elde edilen WC dağılımının yoğunlukları Şekil 2.10.'da verilmiştir.



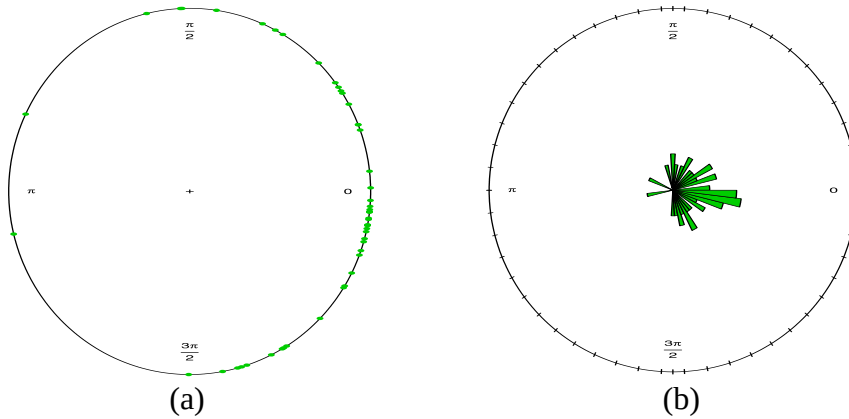
Şekil 2.10. Farklı rho parametreleri için Wrapped Cauchy dağılımının yoğunlukları, (a) $\mu = \pi, \rho = 0.7$, (b) $\mu = \pi, \rho = 0.4$

2.2.5. Wrapped Normal dağılım

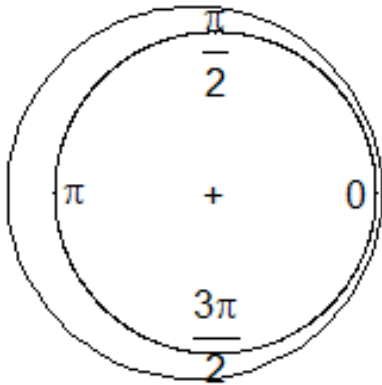
Simetrik, tek modlu ve iki parametrelili bir dağılımdır. Normal dağılımın birim çember etrafında sarılmasıyla elde edilebilir. Olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu aşağıdaki gibi verilmiştir:

$$f(\theta) = \frac{1}{2\pi} \left(1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu) \right), 0 \leq \theta < 2\pi, 0 \leq \rho \leq 1 \quad (2.22)$$

Dağılım fonksiyonu, yukarıdaki ifadenin integralinin alınmasıyla elde edilebilir. $f(\theta)$, varyansı σ^2 olan bir normal dağılımın birim çember etrafında sarılmasıyla elde edilmiş ise, $\rho = e^{-\frac{1}{2}\sigma^2}$ veya $\sigma^2 = -2\log \rho$ olur. $\rho \rightarrow 0$ iken dağılım düzgün dağılıma yakınsarken, $\rho \rightarrow 1$ iken dağılım μ yönü etrafında yoğunlaşmış noktasal dağılıma yakınsama eğilimindedir (Fisher, 1993).

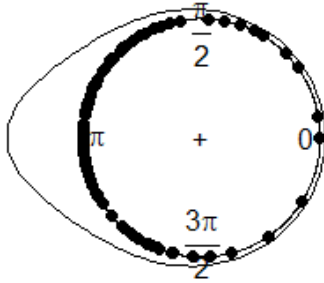


Şekil 2.11. Wrapped Normal dağılımının saçılım grafiği (a) ve gül diyagramı (b)

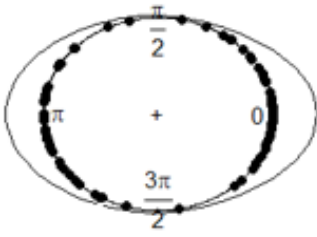


Şekil 2.12. Wrapped Normal dağılımının yoğunluğu ($\mu = \pi, \rho = 0,4$)

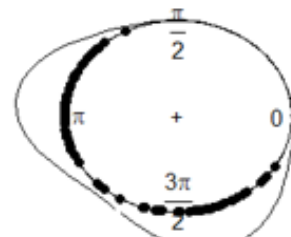
Aşağıdaki şekillerde, farklı dağılımların farklı parametreler için yoğunlukları verilmiştir. Yukarıda bahsedilen dağılımlardan karma dağılımlar oluşturulmuş ve Oliveira (2014)'te kullanılan modellerden yararlanılmıştır.



Şekil 2.13. Wrapped Cauchy dağılımının yoğunluğu ($\mu = \pi, \rho = 0,6$)

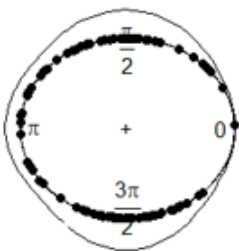


(a)

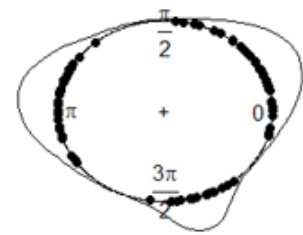


(b)

Şekil 2.14. Farklı parametrelere sahip iki VM dağılımından oluşan karma dağılımın yoğunlukları (a) $\mu = (0, \pi), \kappa = (0.5, 0.5)$, (b) $\mu = (3, 5), \kappa = (0.5, 0.5)$

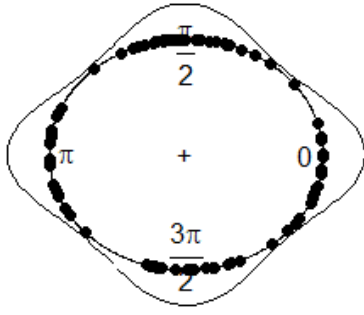


(a)

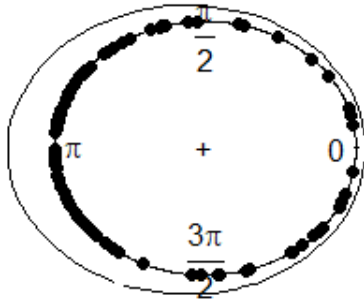


(b)

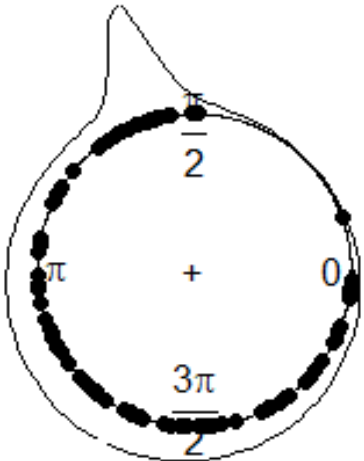
Şekil 2.15. Farklı parametrelere sahip üç VM dağılımından oluşan karma dağılımın yoğunlukları (a) $\mu = \left(\pi, \pi, \frac{5\pi}{3}\right), \kappa = (6, 3, 6), p = (0.25, 0.5, 0.25)$, (b) $\mu = \left(\pi, \pi, \frac{5\pi}{3}\right), \kappa = (6, 6, 2), p = (0.4, 0.4, 0.2)$



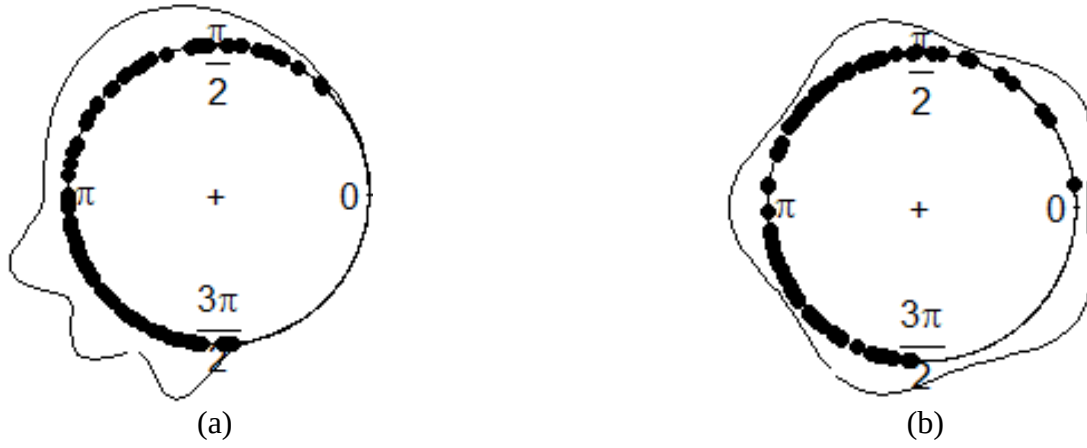
Şekil 2.16. Farklı parametrelere sahip dört VM dağılımından oluşan karma dağılımın yoğunluğu ($\mu = (0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}), \kappa = (10, 10, 10, 10), p = (0.25, 0.25, 0.25, 0.25)$)



Şekil 2.17. VM ve Wrapped Cauchy dağılımlarından oluşan karma dağılımın yoğunluğu $\mu = (\pi, 4\pi/3), \kappa = (5, 0.9), p = (0.25, 0.75)$



Şekil 2.18. Cardioid ve Wrapped Cauchy dağılımlarından oluşan karma dağılımın yoğunluğu ($\mu = (\pi, 2\pi), \kappa = (0.5, 0.9), p = (0.75, 0.25)$)



Şekil 2.19. Farklı oran ve parametrelerde VM dağılımından oluşan karma dağılımlarının yoğunlukları

(a) $\mu = (2, 3, 3.5, 4, 4.5), \quad \kappa = (3, 3, 50, 50, 50),$
 $p = \left(\frac{20}{36}, \frac{5}{36}, \frac{5}{36}, \frac{2}{36}, \frac{4}{36}\right)$ (b) $\mu = \left(\frac{\pi}{5}, \frac{3\pi}{5}, \pi, \frac{7\pi}{5}, \frac{9\pi}{5}\right), \quad \kappa = (10, 10, 10, 10, 10),$
 $p = (0.2, 0.2, 0.2, 0.2, 0.2)$

3. PARAMETRİK OLMAYAN REGRESYON

Dairesel verilerin modellenmesinde kullanılan istatistiki araçlardan biri de regresyon analizidir. Dairesel veriler de doğrusal veriler gibi parametrik ve parametrik olmayan regresyon yöntemleri ile modellenenbilmektedir. Yanıt değişkeni ve açıklayıcı değişkenin farklı formları için farklı dairesel regresyon tipleri mevcuttur. Bu bölümde parametrik olmayan dairesel regresyonun anlaşılrlılığını sağlamak için parametrik olmayan regresyon hakkında temel bilgiler verilecektir.

Regresyon analizi, regresyon fonksiyonu hakkında istatistiksel çıkarımda bulunan bir analizdir ve temelde iki değişken arasındaki ilişkinin incelenmesinde kullanılmaktadır. Regresyon analizinin temeli; ilk olarak Francis Galton tarafından 19. yüzyılın sonlarında atılmıştır. Galton yaptığı çalışmada; anne-babaların boyu ile çocuklarının boyları arasındaki ilişkiyi incelemiş ve kısa boylu anne- babaların çocuklarının boylarının kısa, uzun boylu anne-babaların çocuklarının boylarının uzun olmasına rağmen, çocuklarının boylarının anakütle boy ortalamasına doğru yaklaşma eğiliminde olduğunu görmüştür. Bu eğilimi “ortaya doğru çekilme = regression to mediocrity” olarak adlandırmıştır (Galton, 1886: 246-263). Galton’un çalışmaları bugün, değişkenler arasındaki istatistik ilişkileri inceleyen “Regresyon Analizi (Regression Analysis – Relationship Analysis)”nin başlangıcı olmuştur.

Varsayımların sağlanmadığı (normallik vb.) durumlarda regresyon analizi yapabilmek için çeşitli parametrik olmayan metotlar önerilmiştir. Bu metotlara genel olarak “Düzgünleştirme Yöntemleri” (Smoothing Methods) adı verilmektedir.

Parametrik ve parametrik olmayan ayrımı “parametre” kavramına dayanmaktadır. Parametre, ana kütleleri birbirinden ayırmaya yarayan ortalama, oran, varyans gibi ölçülerin genel ifadesidir. Genel olarak temeli ölçülere dayanan yöntemlere parametrik yöntemler denilmektedir. Bu yöntemin dışında kalan ve uygulanması için çeşitli parametrelerin hesaplanmasına gerek olmayan yöntemlere ise parametrik olmayan yöntemler denilmektedir (Freund, 1962). Uygulamada parametrik yöntemlerin, parametrik olmayan yöntemlere göre daha fazla tercih edildiği ve daha yaygın olarak kullanıldığı görülmektedir. Bunun nedeni parametrik yöntemlerin uygulanması ve yorumlanmasının daha kolay olmasıdır.

Ancak parametrik yöntemlerin uygulanabilmesi için gerekli tüm varsayımların sağlanması gerekmektedir. Herhangi bir varsayımın sağlanmaması durumunda modeller eksik belirlenecek, modelin tutarlılık ve etkinlik gibi özelliklerinde kayıplar ortaya çıkacak ve araştırmacı bu yöntemi kullanamayacaktır. Bu nedenle parametrik olmayan yaklaşımlar, varsayımların az olması nedeniyle parametrik yaklaşımlara göre daha fazla alanda uygulanabilmektedir.

Parametrik ve parametrik olmayan regresyon yöntemleri, regresyon analizi için her ne kadar farklı yaklaşımlar olarak kabul edilseler de, bu durum bir yöntemin diğerini dışlayacağı anlamına gelmez. Parametrik olmayan regresyon sadece olayların altında yatan nedenler tanımlı olmadığında kullanılan bir yöntem değildir. Olayı temsil eden denklem bilinse de parametrik olmayan regresyon yöntemleri bizi yine doğru sonuca götürür. Bununla birlikte parametrik olmayan tahmin ediciler, parametrik model geçerli olduğunda, parametrik tahmin edicilere göre daha az etkindir. Regresyon analizinde performans amaçtan daha önemlidir. Regresyonun amacı, regresyon denkleminin değerlerini etkilemez (Takezawa, 2005:3).

Regresyon analizinin amacı, bilinmeyen yanıt fonksiyonu “ m ” için uygun analizler üretmektir. Gözlemsel hataları azaltarak yorumlama yapmak, y 'nin x 'e bağımlılığının önemli ayrıntılarına konsantre olmamızı sağlar. Bu eğri yaklaştırma prosedürüne genel olarak “düzgünleştirme” denir. “Parametrik olmayan” terimi, “regresyon eğrisinin esnek fonksiyonel formu” anlamına gelmekte, dağılımdan bağımsız yöntemler kast edilmemektedir. Bir regresyon eğrisini tahmin etmek için parametrik olmayan yaklaşımların kullanılması, 4 temel amaca dayanmaktadır:

- 1) İki değişken arasındaki genel bir ilişkiyi keşfetmek için çok yönlü bir yöntem sunar.
- 2) Belli bir modele atıfta bulunulmaksızın gözlemlerin tahminlerinin elde edilmesini sağlar.
- 3) İzole edilmiş noktaların etkisini inceleyerek sahte gözlem bulmak için bir araç görevi görür.
- 4) Eksik değerler için esnek bir değiştirme yöntemi oluşturur veya bitişik x değerleri arasında interpolasyon yapar.

Parametrik olmayan düzgünleştirme, belli bir parametrik modele atıfta bulunmadan, x yönünde sapmalar için çok yönlü bir ön tarama yöntemi sunmaktadır. Parametrik olmayan modeller, bitişik veri noktaları arasında interpolasyon yardımıyla eksik veri boşluğu arasında köprü kurarken, parametrik modeller interpolasyonda tüm gözlemleri içermektedir (Härdle, 1990:13). Parametrik olmayan modellerinin en büyük faydası bağımsız değişken ile bağımlı değişken arasındaki ilişkinin şekli hakkında kesinlik olmaması durumunda kullanılabilmesidir. Bu modeller düşünilemeyen fonksiyonel şekillerin ortaya çıkarılmasını da sağlamaktadır (D.Dole,1999).

Bu çalışmada parametrik olmayan regresyon yöntemlerinden Nadaraya-Watson ve Yerel Lineer Kernel tahmin edicileri ele alınmıştır.

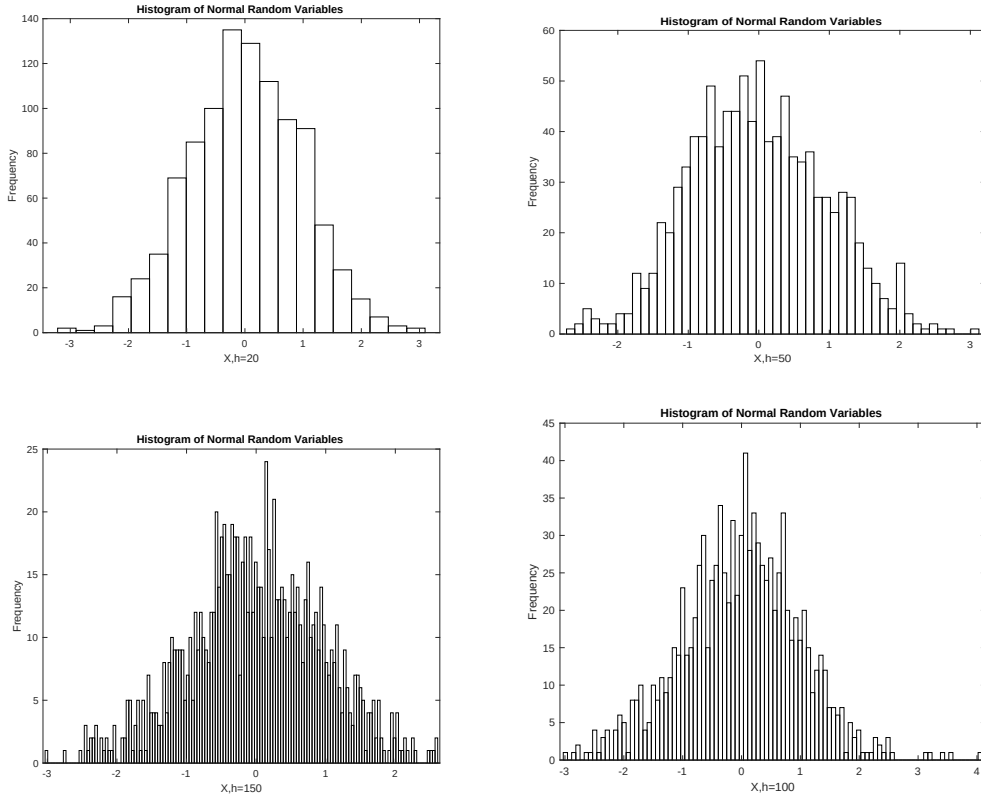
3.1. Kernel Fonksiyonu ve Özellikleri

Kernel fonksiyonu sadece düzgünleştirme için değil, yoğunluk tahmininde de sıklıkla kullanılan yöntemlerden biridir. Kullanılan en eski yoğunluk tahmin edicilerinden biri histogramdır. Kernel düzgünleştiricisi, histogramdan esinlenerek elde edilmiş bir yöntemdir. Histogram, gerçek doğruyu, genellikle küçük kutular (bin) olarak adlandırılan eşit büyüklükteki aralıklara bölerek oluşturulur. h kutu genişliğini ifade edecek olursa, eşitlik (3.1), x 'in bir noktasında histogram tahminini vermektedir.

$$\hat{f}_H(x; b) = \frac{x' \text{ i içeren kutulardaki gözlem sayısı}}{nb} \quad (3.1)$$

Histogram oluştururken dikkat edilmesi gereken en önemli nokta, kutu genişliği (bin width) seçimi ve kutu sınırlarının belirlenmesidir (Wand ve Jones,1995: 5). Hatalı kutu genişliği seçimi, yanlış sonuçlara ulaşılmasına neden olur. Şekil (3.1)'de, normal dağılımdan üretilen bir veri seti için farklı bant genişlikleri ile çizilen histogramlar yer almaktadır.

h : kutu genişliği



Şekil 3.1. Farklı bant genişlikleri ile normal dağılımdan üretilmiş veriler için histogramlar

Şekillerden de görüldüğü üzere, aynı veri seti kullanılmasına rağmen farklı kutu genişlikleri kullanılarak histogramlar elde edilmiştir. Kutu genişliği azalıp kutu sayısı arttıkça, dağılım çarpıklaşmaya ve çok modlu bir dağılıma dönüşmeye başlamıştır. Kutu genişliğinin azalması aşırı düzgünleşmeye (oversmoothing) neden olmuştur.

Histogram oluşturulurken, başlangıç noktası x_0 ve kutu genişliğinin (bin width) seçimi önemlidir. Özellikle kutu genişliği seçimi problemi histogramın dezavantajlarından biridir. Ayrıca histogramın bir basamak fonksiyonu olması, yoğunluğun temelini düzgün yapısının ortaya çıkmasını zorlaştırmaktadır. Rosenblatt (1956), Whittle (1958) ve Parzen (1962) , yukarıdaki zorlukları ortadan kaldıran bir yaklaşım geliştirmişlerdir. Bu yaklaşımın adımları aşağıdaki gibidir:

- Temel yapı taşı olarak bir kutu yerine düzgün bir Kernel fonksiyonu kullanılır.
- Bu düzgün fonksiyonlar her bir gözlemin etrafında doğrudan merkezlenir.

Bu adımlara uyacak şekilde Kernel tahmin edicisi (Rozenblatt-Parzen) eşitlik (3.2)'deki gibi tanımlanmıştır:

$$\hat{f}(x,h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \quad (3.2)$$

Burada $K_h(t)$, t noktasındaki Kernel fonksiyonunun değeri olup, $K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$, $K \in S_{0,k}$, $h > 0$ olarak tanımlanmıştır.

Pozitif h değeri, düzgünleştirme parametresi veya bant genişliği olarak adlandırılmaktadır. Bant genişliği h , örneklem büyüklüğü n ' e bağlıdır (Ivanka ve diğerleri, 2012).

Buradaki Kernel fonksiyonu K_h , aşağıdaki özelliklere sahiptir:

- $\int K(u)du = 1$,
- $\int uK(u)du = 0$,
- $\int u^2K(u)du = \sigma_K^2 > 0$

Parametrik olmayan regresyon analizlerinde düzgünleştirme işleminin başarılı sonuçlar vermesi, uygun Kernel fonksiyonun kullanılmasına ve uygun düzgünleştirme parametresinin seçimine bağlıdır (Bowman ve Azzalini, 1997:17). Kernel fonksiyonu kullanımında ortaya çıkabilecek en önemli problem, kullanılacak Kernel fonksiyonunun nasıl seçileceğidir. Seçilecek Kernel fonksiyonunun simetrik olması ve sıfır noktasında odaklanmış olması gerekir (Gasser ve diğerleri, 1985). Kernel fonksiyonları genel olarak eşitlik (3.3)'teki gibi ifade edilmektedir.

$$K(x,p) = \{2^{2p+1} * B(p+1,p+1)^{-1} * (1-x^2)^p \{|x| < 1\}\} \quad (3.3)$$

Buradaki B , beta fonksiyonunu temsil etmektedir. $[-1,1]$ aralığında bahsi geçen Kernel fonksiyonlarının hepsi simetrik beta fonksiyonlarıdır.

$p = 0$ olduğunda Uniform,

$p = 1$ olduğunda Epanechnikov,

$p = 2$ olduğunda İki Ağırlıklı,

$p = 3$ olduğunda Üç ağırlıklı Kernel fonksiyonları elde edilmektedir.

p sonsuza yaklaştığında ise Merkezi Limit Teoremi gereğince Kernel fonksiyonları Standart Normal Dağılıma yakınsamaktadır.

En çok kullanılan Kernel fonksiyonlarından bazıları ve Epanechnikov Kernel fonksiyonuna göre görelî etkinlikleri çizelge (3.1)'de verilmiştir (Silverman, 1998:43).

Çizelge 3.1. Bazı Kernel fonksiyonlarının Epanechnikov Kernel fonksiyonuna göre görelî etkinlikleri

Kernel	Şekli	Etkinlik
Epanechnikov	$\frac{3}{4}(1 - u^2)$	1
İki Ağırlıklı (Biweight)	$\frac{15}{16}(1 - u^2)^2$	0,9939
Üç Ağırlıklı (Triweight)	$\frac{35}{32}(1 - u^2)^3$	0,9859
Normal (Gaussian)	$(2\pi)^{-\frac{1}{2}} e^{-\frac{u^2}{2}}$	0,9512
Düzgün (Uniform)	$\frac{1}{2}$	0,9295

Kernel fonksiyonları, k . dereceden türevlenebilir fonksiyonlar olmalıdır. Çizelge (3.1)'de yer alan etkinlik değerleri, yanlılık ile varyans arasındaki dengenin en iyi şekilde sağlanabilmesi amacıyla, AMISE (Asymptotic Mean Integrated Squared Error)'yi minimize etmek amacıyla kullanılan $\int K(u)^2 du$ ifadesinden yola çıkılarak elde edilmiştir. AMISE'nin minimizasyonu aslında $\int K(u)^2 du$ ifadesinin minimizasyonun problemine indirgenmektedir. Bu minimizasyonu sağlayan Kernel fonksiyonu eşitlik (3.4) ile Epanechnikov tarafından tanımlanan Kernel'dir.

$$f(x) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}u^2\right), & -\sqrt{5} \leq u \leq \sqrt{5} \\ 0, & \text{diğer yerlerde} \end{cases} \quad (3.4)$$

Buradan, herhangi bir simetrik Kernel fonksiyonunun Epanechnikov Kernel fonksiyonuna göre etkinlik değerleri eşitlik (3.5) kullanılarak hesaplanabilir (Yolsal, 2017: 50-51).

$$E(K) = \left\{ \frac{C(K_e)}{C(K)} \right\}^{5/4} \quad (3.5)$$

Burada $C(K_e)$, Epanechnikov Kernel fonksiyonunun etkinliğini temsil etmektedir. Gaussian Kernel dışındaki Kernel fonksiyonları $[-1,1]$ aralığında değer almakta ve bu aralık dışında sıfır olmaktadır. Çizelge (3.1)'den açıkça görüldüğü üzere kullanılan Kernel fonksiyonunun çeşidi, AMISE'yi önemli derecede etkilememektedir. Bu nedenle

kullanılacak Kernel fonksiyonunun çeşidi, hesaplama kolaylığı veya $\hat{f}$ 'nin istatistiksel özellikleri gibi ölçüler göz önüne alınarak seçilmelidir (Simonof, 1996:44). Kernel tahmin edicileri 3 parametreye bağlıdır:

- Ağırlık fonksiyonu olarak rol oynayan Kernel
- Tahminin düzgünlüğünü kontrol eden bant genişliği
- Modelde var olduğu varsayılan türevlerin sayısına bağlı olarak Kernel' in sırası.

$\hat{f}$ tahmini, Kernel'in süreklilik ve diferansiyellenebilirlik gibi bütün özelliklerini taşımaktadır. Olasılık yoğunluk tahmininde olduğu gibi Kernel tahmin edicisi, h ile temsil edilen bir bant genişliğine (düzgünleştirme parametresi) sahiptir. h , yerel eğri üzerindeki etki noktalarının derecesini kontrol etmektedir. Eğer h küçükse, eğri yüksek değişkenlik gösterir çünkü tahmin, başlangıç noktası x_0 'a en yakın noktalara büyük ölçüde bağlıdır. Bu durumda, komşuluk küçük olduğu için model yerel değerlere uymaya çalışır ve aşırı düzgünleştirme (oversmoothing) ile karşılaşılır. h ' in büyük olması ise komşuluğun büyük olmasına neden olacağından eğri fazla değişkenlik göstermez. Hatta yeterince büyük bir h ile eğri, tüm veri setine uydurulabilmektedir. Ancak bu durumda da eksik düzgünleştirme (undersmoothing) problemi ile karşılaşılmaktadır (Wand ve Jones, 1995:12-13).

Aşağıda yerel polinomiyal Kernel tahmin edicisi için tanımlamalar verilmiştir. Herhangi bir x noktasında uydurulan polinomiyalin derecesi d ile temsil edilmektedir. $\hat{Y} = f(x)$ tahmini, eşitlik (3.6)'da yer alan polinomiyalde (X_i, Y_i) noktaları kullanılarak ve ağırlıklı EKK'dan yararlanarak elde edilebilmektedir.

$$\beta_0 + \beta_1(X_i - x_0) + \dots + \beta_d(X_i - x_0)^d \quad (3.6)$$

Kernel fonksiyonu için ağırlıklar eşitlik (3.7)'deki gibi tanımlanmaktadır.

$$K(X_i - x) = \frac{1}{h} K\left(\frac{X_i - x}{h}\right) \quad (3.7)$$

$\hat{\beta}_1$, eşitlik (3.8)'i minimize etmekte ve x noktasındaki tahmin değeri $\hat{\beta}_0$ olarak elde edilmektedir.

$$\sum_{i=1}^n K_h(X_i - x) (Y_i - \beta_0 - \beta_1(X_i - x) - \dots - \beta_d(X_i - x)^d)^2 \quad (3.8)$$

Modeli tahmin etmek için kullanılan noktalar eşitlik (3.6)'dan da açıkça görüldüğü üzere x merkezli olduğundan, x noktasındaki tahmin değeri, modelin elemanlarının sıfıra eşitlenmesiyle elde edilir. Böylece geriye sadece modelin sabit terimi olan β_0 parametresi kalmaktadır.

Fonksiyonun kestirilen değeri, Kernel fonksiyonlarının x noktası etrafında merkezlenmesiyle elde edilmektedir. Ağırlıklı EKK (En küçük kareler) yöntemi matris notasyonları kullanılarak ifade edilebilmektedir. Standart ağırlıklı EKK teoremine göre çözüm eşitlik (3.9) kullanılarak elde edilebilir (Draper ve Smith, 1981).

$$\hat{\beta} = (X_x^T W_x X_x)^{-1} X_x^T W_x Y, \quad (3.9)$$

$$X_x = \begin{bmatrix} 1 & X_1 - x & \dots & (X_1 - x)^d \\ \vdots & \vdots & \dots & \vdots \\ 1 & X_n - x & \dots & (X_n - x)^d \end{bmatrix} \quad (3.10)$$

Y , $nx1$ boyutlu vektör ve W nxn boyutunda ağırlıkları diagonal bir matristir. Ağırlıklar $w_{ii}(x) = K_h(X_i - x)$ şeklinde ifade edilmektedir. Bazı ağırlıklar kullanılan Kernel fonksiyonunun türüne göre sıfır olarak elde edilebilir.

$\hat{f}(x)$ tahmin edicisi yerel eğrinin β_0 kesim (intercept) katsayısı olduğundan, bu değer eşitlik (3.11) kullanılarak elde edilmektedir.

$$\hat{f}(x) = e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x Y \quad (3.11)$$

Burada e , ilk elemanı bir ve diğer elemanları sıfır olan, $(d + 1) \times 1$ boyutunda bir vektördür (Martinez ve Martinez, 2007:529-531).

Bileşik olasılık yoğunluk fonksiyonu $f(x, y)$ olan X ve Y gibi iki rasgele değişken ele alındığında, bu rasgele değişkenler bağımsız ve özdeş dağılmış gözlem çiftleri (X_i, Y_i) iken, çekilen rasgele örneklemden n adet (x_i, y_i) gözlem çifti alındığında rasgele tasarım

oluşmaktadır. Burada hem X hem de Y rasgele değişkendir. Bu nedenle kurulacak olan regresyon modelinde Y 'nin koşullu ortalaması ve varyansı X 'e bağlı olacaktır. Ancak bazı durumlarda X açıklayıcı değerinin aldığı değere karşılık Y bağımsız değişkeninin alacağı değer araştırılmaktadır.

Örneğin, kansere yakalanma oranı ile sigara içme durumu arasındaki ilişki incelendiğinde, kanserin belli bir sigara içme oranının üzerinde artacağı düşüncesiyle araştırmacı öncelikle sigara içme durumu ile ilgili bir eşik değeri belirlemelidir. Bu durumda tekrarlı örneklemelerde Y yanıt değişkeni hala rasgele değişken iken X açıklayıcı değişkeni rasgele değişken olmayacaktır. Bu durumda açıklayıcı değişken ve yanıt değişkeninin özdeş olduğunda dair bir varsayımda bulunulmayan sabit tasarım söz konusudur (Yolsal, 2017: 120-121).

3.2. Nadaraya-Watson Tahmin Edicisi

En basit Kernel düzgünleştirme yaklaşımı olup Nadaraya (1964) ve Watson (1964) tarafından birbirlerinden bağımsız bir şekilde geliştirilmiştir.

$(X_i, Y_i) = (x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$, bir yığından rasgele seçilen gözlemler olsun.

Nadaraya Watson Kernel tahmin edicisi eşitlik (3.12)'de verildiği gibidir.

$$\hat{f}_{NW}(x) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i}{\sum_{i=1}^n K_h(X_i - x)} = \sum_{i=1}^n w_i Y_i \quad (3.12)$$

Burada w , ağırlıklar olup toplamaları bire eşittir. Burada eğri uydurulurken sabit bir bant genişliği kullanılmaktadır. Eşitlik (3.7)'ye bakıldığında, NW tahmin edicisinin aslında yanıt değişkeninin yerel ağırlıklı ortalaması olduğu açıkça görülmektedir (Härdle ve diğerleri, 2005: 90). Eşitlik (3.7)'deki ifade, eşitlik (3.4)'te $d = 1$ olduğunda ortaya çıkmaktadır. $d = 0$ olduğunda ise, verilen x noktasında sabit bir yerel fonksiyon uydurulmaktadır. Yukarıda tanımlanan eşitliğin rasgele tasarım durumları için olduğunu not etmek gerekmektedir.

Tasarım noktaları sabitlendiğinde X_i , x_i yerine kullanılır, aksi durumda ifade değişmez (Wand and Jones, 1995:130). Eşitlik (3.7)'te tanımlanan x noktası, uygulamada genellikle açıklayıcı değişkenlerin yerel bölgedeki ortalaması olarak kullanılmaktadır.

3.3. Yerel Lineer Kernel Tahmin Edicisi

x noktasına düz bir çizgiyle doğru uydurulduğunda, yerel lineer tahmin edici kullanılmaktadır.

Bu, $d = 1$ olduğu duruma karşılık gelir. Bu yüzden tahmin değerleri, eşitlik (3.13)'ün β_0 ve β_1 'e göre minimize edilmesiyle elde edilir:

$$\sum_{i=1}^n K_h(X_i - x)(Y_i - \beta_0 - \beta_1(X_i - x))^2 \quad (3.13)$$

Bu durumda Yerel Lineer Kernel Tahmin Edicisi aşağıdaki gibi tanımlanmaktadır:

$$\widehat{f}_{LL}(x) = \frac{1}{n} \sum_{i=1}^n \frac{\{\hat{s}_2(x) - \hat{s}_1(x)(X_i - x)\} K_h(X_i - x) Y_i}{\hat{s}_2(x) \hat{s}_0(x) - \hat{s}_1(x)^2} \quad (3.14)$$

$$\hat{s}_r(x) = \frac{1}{n} \sum_{i=1}^n (X_i - x)^r K_h(X_i - x) \quad (3.15)$$

Buradaki $\hat{s}_r(x)$, ağırlıkları temsil etmektedir ve toplamaları NW tahmin edicisinde olduğu gibi bire eşittir. Bu tahmin edici kullanıldığında ağırlıklandırma, veri setindeki bütün gözlemler dikkate alınarak yapılmaktadır. Fonksiyon, NW tahmin edicisinde de olduğu gibi sabit bir bant genişliği ile uydurulmaktadır. Eşitlik (3.9), rasgele tasarım durumları için kullanılmaktadır. Sabit tasarım durumu, rasgele değişen X_i 'nin sabit nokta x_i ile değiştirilmesi ile elde edilir.

Kernel düzgünleştirme yöntemlerini kullanırken sınırlarda ya da örneklemin uç noktalarında, veri yokluğundan kaynaklanan sorunlar ortaya çıkabilir. Yani, ağırlık vardır ancak bununla ilişkilendirilecek veri yoktur. Bu durumda ağırlıklar sıfır değerini almakta ve bant genişliği tahminlerinde sorunlara yol açmaktadır. Wand ve Jones (1995:129), Yerel Lineer Kernel tahmin edicilerinin çoğu durumda hatta sınırlarda bile iyi sonuçlar verdiğini ifade etmiştir. Scott (1992), Wand ve Jones (1995) ise yukarıdaki durumlar mevcut iken

Nadaraya-Watson tahmin edicisi kullanılırsa, modifiye edilmiş Kernel'e ihtiyaç duyulacağını belirtmiştir.

3.4. Tahminin İstatistiksel Özellikleri

$\hat{f}$ Kernel yoğunluk fonksiyonu tahmin edicisinin gerçek f yoğunluk fonksiyonuna yakın sonuçlar vermesi, istatistiksel analizlerde elde edilmesi gereken en önemli sonuçlardan biridir. Tahminin gerçek yoğunluğa yakınlığı, kullanılan yöntemin doğruluğunu kanıtlar niteliktedir. Tahminin gerçek yoğunluğa yakın sonuçlar vermesi, yan (bias) değerinin azalmasını ve varyans - yan arasındaki dengenin korunmasını sağlar. Yoğunluğun tek bir noktadaki tahmininin hatası bulunurken genel olarak eşitlik (3.16)'de tanımlanan Hata Kareler Ortalaması (Mean Squared Error-MSE) kriteri kullanılır.

$$MSE\{\hat{f}(x, h)\} = E\{\hat{f}(x, h) - f(x)\}^2 \quad (3.16)$$

Bu ifade açık bir şekilde yazılacak olursa eşitlik (3.17) elde edilir.

$$MSE\{\hat{f}(x, h)\} = (E[\hat{f}(x, h)] - f(x))^2 + E(\hat{f}(x, h))^2 - E^2(\hat{f}(x, h)) \quad (3.17)$$

Burada $E\hat{f}(x, h) - f(x)$ yan, $\hat{f}(x, h)$ ve $E(\hat{f}(x, h))^2 - E^2(\hat{f}(x, h))$ ifadesi $\hat{f}(x, h)$ 'nin varyansının tahmin edicisi olarak bilinmektedir. Buradan anlaşılacağı üzere, hata kareler ortalaması, varyans ve yan'dan oluşmaktadır. Bu ayrışmadan yararlanarak Kernel yoğunluk tahmini için eşitlik (3.18)'de yer alan ifade elde edilebilir:

$$MSE\{\hat{f}(x, h)\} = \{(K_h * f)(x) - f(x)\}^2 + \frac{1}{n}\{(K_h^2 * f)(x) - (K_h * f)^2(x)\} \quad (3.18)$$

Bu ifadede 1.terim $\hat{f}(x, h)$ 'in yan'ın karesini, 2.ve 3. Terim ise varyansını ifade etmektedir. Sadece tek bir nokta değil, bütün bir eğrinin yoğunluğunun hata kareler ortalamasını bulmak için ise daha global bir kriter geliştirilmiştir. ISE (Bütünleşik hata kare) aşağıda tanımlandığı gibidir.

$$ISE\{\hat{f}(\cdot, h)\} = \int \{\hat{f}(x, h) - f(x)\}^2 dx \quad (3.19)$$

Ancak bu rasgele niceliğin beklenen değeri alınarak analiz edilmesi daha uygundur. Bu nicelik, eşitlik (3.20)'de tanımlanmıştır ve MISE (Bütünleşik Hata Kareler Ortalaması) olarak adlandırılmaktadır.

$$MISE\{\hat{f}(\cdot, h)\} = E \int \{\hat{f}(x, h) - f(x)\}^2 dx \quad (3.20)$$

Bazı hesaplamalar yapılarak eşitlik (3.21)'deki gibi daha kullanışlı ifadeler elde edilebilir.

$$\begin{aligned} MISE\{\hat{f}(\cdot, h)\} &= \frac{1}{nh} \int K^2(x) dx + (1 - n^{-1}) \int (K_h * f)^2(x) dx \\ &\quad - 2 \int (K_h * f)(x) dx + \int f^2(x) dx \end{aligned} \quad (3.21)$$

MSE ve MISE kullanımı ile ilgili karşılaşılan problemlerden biri, dolaylı olarak bant genişliğine bağlı olmalarıdır. Bu durum, bant genişliğinin Kernel yoğunluk tahmini üzerindeki performansını tahmin etme konusunda zorluklar yaşanmasına neden olmaktadır. Varyans ve yanlılık terimleri için büyük örneklem yaklaşımlarının türetilip kullanıldığı yöntemler aşağıda verilmiştir. Bu yöntemler çok basit ifadelere sahip olmakla birlikte, bant genişliğinin rolünün daha iyi anlaşılmasını sağlamaktadır. Bu yöntemler ayrıca Kernel yoğunluk tahmin edicisi ve MISE optimal bant genişliğinin yakınsama oranını elde etmek için de kullanılabilir (Ivanka ve diğerleri, 2012).

$$h_{MISE} = \arg \min MISE\{\hat{f}(x, h)\} \quad (3.22)$$

$f \in C^{k_0}, 0 < k \leq k_0, K \in S_{0,k}, \lim_{n \rightarrow \infty} h = 0, \lim_{n \rightarrow \infty} nh = \infty$ ve $f^{(k)}$ integrallenebilir karesel fonksiyon olmak üzere, MISE ifadesi eşitlik (3.23)'teki gibi yeniden yazılabilmektedir.

$$MISE\{\hat{f}(\cdot, h)\} = \frac{1}{n} V(K) + h^{2k} \beta_k^2 D_k + o\{h^{2k} + (nh)^{-1}\} \quad (3.23)$$

Burada $D_k = \int \left(\frac{f^{(k)}(x)}{k!} \right)^2 dx$, $V(K)$ ise kullanılan Kernel fonksiyonunun varyansını ifade etmektedir.

MISE ifadesi matematiksel olarak çözülebilir olmadığından, AIV (Asymptotic Integrated Variance) ve AISB (Asymptotic Integrated Squared Bias)'ın toplamı olarak yazılan

AMISE (Asymptotic MISE) eşitlik (3.24)'de tanımlanmıştır. Burada eşitliğin sağındaki ilk ifade AIV, 2. İfade ise AISB'dir.

$$AMISE\{\hat{f}(\cdot, h)\} = \frac{V(K)}{nh} + h^{2k} \beta_k^2 D_k \quad (3.24)$$

Yukarıdaki ifadeden de açıkça görüldüğü üzere, h, k ve K , AMISE ölçüsünün hesaplanmasında önemli rol oynayan parametrelerdir. AMISE değerini elde edebilmek için öncelikle bu parametrelerin tahmin edilmesi gereklidir (Ivanka ve diğerleri, 2012).

Analizlerde, AMISE değeri küçük olan model seçilir. Ayrıca bu değeri en küçük yapan bant genişliği, optimal bant genişliği olarak seçilir. Yukarıdaki parametreler ve AMISE değerinin tahmini için birçok algoritma geliştirilmiştir (Ivanka ve diğerleri, 2012).

3.5. Kernel Fonksiyonunun Seçimi

Kernel fonksiyonunun seçimi, tahminlerin asimptotikliğini bant genişliğinin seçimi kadar etkilememektedir. Genel olarak uygulamada Epanechnikov Kernel ya da Gaussian Kernel fonksiyonları kullanılmaktadır.

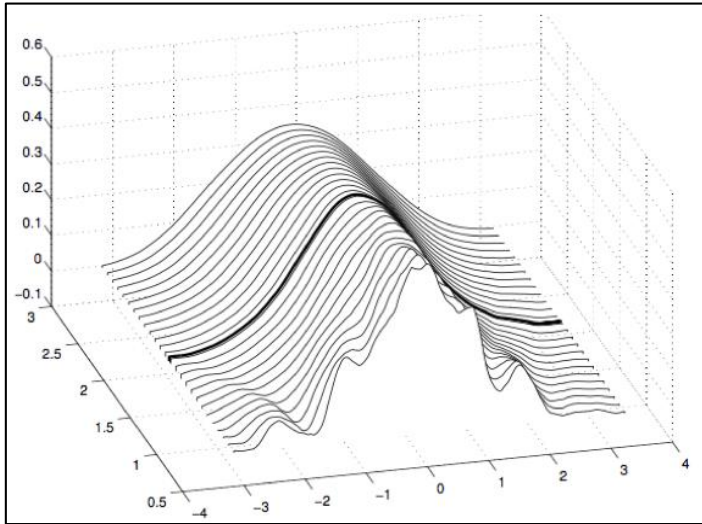
3.6. Bant Genişliği Seçimi

Kernel düzgünleştirme yaparken bant genişliğinin seçimi önemli bir problemdir. Evrensel olarak kabul edilmiş kesin bir yöntem henüz bulunamamıştır. Bant genişliği seçimi, kullanılan Kernel fonksiyonunu da etkilemektedir. Bant genişliği seçimi yapılırken genellikle otomatik yöntemler kullanılmaktadır. Bu yöntemlerin avantajı, farklı Kernel düzgünleştiricileri için karşılaştırma yapılmasına olanak sağlamasıdır. Bu yöntemler uygulanırken, küresel hata kriterini minimize eden değerler tercih edilir (Ivanka ve diğerleri, 2012). Hardle, bu konuyla ilgili kestirilen artıklara dayalı bootstrap yöntemi önermiştir. En basit bant genişliği seçimi yöntemi, farklı birçok bant genişliği değeri için grafikler çizdirip sezgisel olarak karar vermektir. Bu yöntem “Judge by eye” olarak adlandırılmaktadır (Ivanka ve diğerleri, 2012). Ancak, bu yöntemin uygulanması tavsiye edilmemektedir, çünkü istatistiksel olarak teorik bir temeli yoktur ve araştırmacı doğru

sonuca ulaşamayabilir. Yanlış bant genişliği seçimi, modelin hatalı kurulmasına neden olmaktadır (Ivanka ve diğerleri, 2012).

Elde edilmesi amaçlanan $Y = m(.) + e$ eşitliğindeki bilinmeyen m 'nin ya da türevinin tahmini olan Kernel düzgünleştiricilerinin doğruluğu (accuracy), K ve h 'nin bir fonksiyonudur. Doğruluk, temel olarak düzgünleştirme parametresi h 'ye bağlıdır.

Bu bölümde tanıtılan bant genişliği (düzgünleştirme parametresi) seçimi yöntemleri regresyon eğrisi ve türevleri için karesel hata oranlarının optimize edilmesine dayalıdır (Härdle,1994). Bütün bant genişliği seçimlerinin kullanışlı yoğunluk tahminleri elde edilmesini sağladığı dikkat çekmektedir. Büyük bant genişlikleri, bilinmeyen yoğunluğun küresel yapısı hakkında bir öngörü sağlarken; küçük bant genişlikleri, gerçek yoğunlukta mevcut olmayan yerel yapıları ortaya çıkarabilmektedir (Ivanka ve diğerleri, 2012). Yani optimal bant genişliği olmasa dahi bütün bant genişlikleri ile elde edilen tahminler, bilinmeyen yoğunluğun yapısıyla ilgili bilgiler vermektedir. Bu yaklaşım, küçük bir simülasyon çalışmasıyla gösterilmiştir. Standart normal dağılımdan $n=100$ olacak şekilde üretilen örnekler ve farklı bant genişlikleri ile Kernel yoğunlukları Şekil (3.3)'teki gibi elde edilmiştir (Chaudhuri ve Marron ,1999).



Şekil 3.2. Farklı bant genişlikleri ile Kernel yoğunlukları

Burada kalın çizgiyle çizilmiş eğri, optimal bant genişliği ile elde edilen eğriyi temsil etmektedir.

- Eğer h çok küçükse, küçük yana sahip fakat değişkenliği yüksek olan bir tahmin edici elde edilir.
- Eğer h çok büyükse, büyük yana sahip fakat değişkenliği az olan bir tahmin edici elde edilir.

Bant genişliği seçimi için birçok yöntem geliştirilmesine rağmen burada en temel yöntemler tanıtılacaktır.

3.6.1. Çapraz doğrulama (Cross Validation) yöntemi

En eski tamamen otomatik ve veri odaklı bant genişliği seçimi yöntemleri CV (Çapraz Doğrulama-Cross Validation) fikrine dayalıdır (Bowman (1984); Silverman (1986); Scott (1992); Wand ve Jones (1995)). Bütün CV bant genişliği seçicilerinin amacı, MISE veya AMISE'yi minimize etmektir. En küçük kareler çapraz doğrulama (LSCV) MISE'yi hedef alır ve minimizasyon için eşitlik (3.20)'de verilen amaç fonksiyonunu kullanır:

$$LSCV(h) = \int \hat{f}^2(x, h) dx - 2n^{-1} \sum_{i=1}^n \hat{f}_{-i}(X_i, h) \quad (3.25)$$

Burada $\hat{f}_{-i}$ yoğunluğu, eşitlik (3.26) ile ifade edilmiştir.

$$\hat{f}_{-i}(X_i, h) = (n-1)^{-1} \sum_{\substack{j=1 \\ i \neq j}}^n K_h(X_i - X_j) \quad (3.26)$$

Bu yoğunluk fonksiyonu, Jackknife yönteminde olduğu gibi, X_i 'nin silindiği örnekleme dayalıdır ve genellikle “leave one out” yöntemi olarak adlandırılır.

$\hat{f}_{-i}(X_i, h) = \frac{n}{n-1} \hat{f}(X_i, h) - \frac{K_h(0)}{n-1}$ olması dikkat çekici bir özelliktir.

LSCV(h)'ın MISE ölçüsünün yansız bir tahmin edicisi olduğu, eşitlik (3.27) ile kolayca gösterilebilir.

$$E\{LSCV(h)\} = MISE\{\hat{f}(\cdot, h)\} - \int f^2(x) dx \quad (3.27)$$

$\hat{h}_{LSCV}$, LSCV(h)'ın minimizasyonunu gösterebilir. Bu durumda optimal bant genişliği eşitlik (3.28) kullanılarak elde edilmektedir.

$$\hat{h}_{LSCV} = \arg \min_{h \in H_n} LSCV(h) \quad (3.28)$$

Yanlı çapraz doğrulama (BCV-Biased Cross Validation), AMISE'nin tahminini içerir ve amaç fonksiyonu olarak AMISE'yi kullanır (Scott and Terrell, 1987).

$$BCV(h) = (nh)^{-1}V(K) + \frac{1}{(k!)^2} h^{2k} \beta_k^2(K) \tilde{V}(f^{(k)}) \quad (3.29)$$

$$\tilde{V}(f^{(k)}) = \frac{1}{n^2} \sum_{\substack{i,j \\ i \neq j}}^n (K_h^{(k)} * K_h^{(k)})(X_i - X_j) \quad (3.30)$$

$V(f^{(k)}) = \int (f^{(k)}(x))^2 dx$ 'in tahminidir. BCV bant genişliği seçicisi eşitlik (3.31)'de elde edilmiştir.

$$\hat{h}_{BCV} = \arg \min_{h \in H_n} BCV(h) \quad (3.31)$$

Bant genişliği seçicilerinin performansı, görelî yakınsama oranı ile değerlendirilebilir. $\hat{h}$, veri odaklı bir bant genişliği seçicisi olsun. $\hat{h}$ 'nın $h_{opt,0,k}$ 'ya $n^{-\alpha}$ görelî yakınsama oranı ile yakınsadığı söylenirse eşitlik (3.32) elde edilir.

$$\frac{\hat{h} - h_{opt,0,k}}{h_{opt,0,k}} = O(n^{-\alpha}) \quad (3.32)$$

$k=2$ için görelî yakınsama oranı $\frac{\hat{h} - h_{opt,0,2}}{h_{opt,0,2}} = O(n^{-0.1})$ olarak elde edilir (Hall ve Marron (1987); Scott ve Terrell (1987)).

3.6.2. Plug-in yöntemi

Bu method, bilinmeyen $V(f^{(k)}) = \int (f^{(k)}(x))^2 dx$ 'in tahminine dayalıdır.

$$h_{opt,0,k}^{2k+1} = \frac{\gamma_{0k}^{2k+1}}{2nkD_k} = \frac{\gamma_{0k}^{2k+1} (k!)^2}{2nkV(f^{(k)})} \quad (3.33)$$

f üzerinde yeterli düzgünlük varsayımları altında, parçalarla integral alarak; $V(f^{(k)}) = (-1)^k \int f^{(2k)}(x)f(x)dx$ elde edilir. Bu nedenle $\psi_k = \int f^{(2k)}(x)f(x)dx$ tahmin edicisini incelemek yeterlidir. Burada $\psi_k = E\{f^{(2k)}(X)\}$ olduğu gözden kaçmamalıdır.

ψ_k ölçüsü, g bant genişliği ve L Kerneli ile $\widehat{\psi}_k(g) = \frac{1}{n^2} \sum_{i,j=1}^n L_g^{(2k)}(X_i - X_j)$ şeklinde tahmin edilebilir. Böylece eşitlik (3.34)'te yer alan yakınsama oranı elde edilir.

$$\hat{h}_{PI} = \left(\frac{\gamma_{0k}^{2k+1} (k!)^2}{2nk(-1)^k \widehat{\psi}_k(g)} \right)^{\frac{1}{2k+1}} \quad (3.34)$$

Bu kural $\hat{h}_{PI}$ 'nın seçimi pilot bant genişliği g 'ye bağlı olduğundan tamamen otomatik değildir. $k=2$ için göreceli yakınsama oranı eşitlik (3.35)'te verilmiştir.

$$\frac{\hat{h}_{PI} - h_{opt,0,2}}{h_{opt,0,2}} = O(n^{-\frac{5}{14}}) \quad (3.35)$$

3.6.3. Yinelemeli yöntem

Aşağıdaki denkleme bağlı bu yöntem, AMISE için uygun bir tahmin elde edilmesini sağlamaktadır:

$$AIV\{\hat{f}(\cdot, h_{opt,0,k})\} - 2kAISB\{\hat{f}(\cdot, h_{opt,0,k})\} = 0 \quad (3.36)$$

Bu denklem eşitlik (3.37)'deki gibi yeniden yazılabilir ve $\arg \min_{h \in H_n} AMISE(h)$ minimizasyon problemi bu denklemin çözümüne eşit hale gelir.

$$\frac{V(K)}{nh} - 2kh^{2k} \frac{\beta_k^2}{(k!)^2} V(f^{(k)}) = 0 \quad (3.37)$$

Varyans tahmin edicisinin eşitlik (3.38) ve yanının eşitlik (3.39)'daki gibi tanımlandığı varsayılın.

$$\widehat{var}\{\hat{f}(x, h)\} = \frac{1}{nh} \int K^2(y) \hat{f}(x - hy, h) dy \quad (3.38)$$

$$\overline{bias}\{\hat{f}(x, h)\} = (K_h * \hat{f})(x, h) - \hat{f}(x, h) = \int \hat{f}(x - hy) K(y) dy - \hat{f}(x, h) \quad (3.39)$$

Buradan $\widehat{AV}\{\hat{f}(\cdot, h)\} = \frac{1}{nk} V(K)$ olduğu açıkça görülmektedir. Bu konvolüsyon kullanılarak tahmin edilen yan için eşitlik (3.35)'te elde edilmiştir.

$$\begin{aligned} \overline{AISB}\{\hat{f}(\cdot, h)\} &= \int \overline{bias}^2\{\hat{f}(x, h)\} dx \\ &= \frac{1}{n^2 h} \sum_{i,j=1}^n (K * K * K * K - 2K * K * K + K * K) \left(\frac{X_i - X_j}{h} \right) \\ &= \frac{1}{n^2 h} \sum_{i,j=1}^n \Lambda \left(\frac{X_i - X_j}{h} \right) \end{aligned} \quad (3.40)$$

Yukarıdaki ifadenin yanlı olmasından kaçınmak için, AISB' nin yansız tahmin edicisi

$$\widehat{AISB}\{\hat{f}(\cdot, h)\} = \frac{1}{n^2 h} \sum_{i \neq j}^n \Lambda \left(\frac{X_i - X_j}{h} \right) \text{ olup böylece AMISE tahmini eşitlik (3.41)'deki}$$

gibi elde edilir.

$$\widehat{AMISE}\{\hat{f}(\cdot, h)\} = \frac{1}{nh} V(K) + \frac{1}{n^2} \sum_{i \neq j}^n \Lambda_h(X_i - X_j) \quad (3.41)$$

$\hat{h}_{IT,0,k}$, AMISE'yi minimize eden bant genişliği olsun. Bu durumda $\hat{h}_{IT} = \arg \min_{h \in H_n} \widehat{AMISE}\{\hat{f}(\cdot, h)\}$ ifadesine eşittir. $\widehat{AMISE}$ 'nin minimizasyonu ise eşitlik (3.42)'nin çözümüne eşittir:

$$\frac{1}{nh} V(K) + \frac{2k}{n^2} \sum_{i \neq j}^n \Lambda_h(X_i - X_j) = 0 \quad (3.42)$$

Horová ve Zelinka (2007a)'da, bu doğrusal olmayan denklem Steffensen'in metoduyla çözülmüştür. $k = 2$ olduğu durumda yakınsama oranı eşitlik (3.43)'te elde edilmiştir.

$$\frac{\hat{h}_{IT} - h_{opt,0,2}}{h_{opt,0,2}} = O(n^{-\frac{1}{10}}) \quad (3.43)$$

CV ve yinelemeli yöntemlerinde bant genişlikleri için göreceli yakınsama oranları aynı olsa da, $O(n^{-\frac{1}{10}})$ 'daki sabit yinelemeli yöntem için daha hızlı yakınsama sağlamaktadır. Dahası, Λ içindeki konvolüsyon fonksiyonu kolaylıkla hesaplanabilir. Bu gerçeklerden dolayı, yinelemeli metodunun optimal bant genişliği için daha iyi ve çok hızlı tahminler verdiği görülmektedir.

Horová ve Zelinka (2007a)'da, yinelemeli (iterative) metodunun CV'ye göre daha az zaman kaybına yol açtığı gösterilmiştir.

3.7. Parametrik Olmayan Regresyonda Model Seçimi

Parametrik olmayan düzgünleştirme yöntemlerinde model seçimi için kesin kurallar olmamakla beraber, parametrik yöntemler için kullanılan ölçüler geliştirilerek model seçimi yöntemleri belirlenmektedir. Konuyla ilgili Yuhong Yang (1999), "Model Selection for Nonparametric Regression" makalesinde AIC'ye benzer ABC ölçüsü kullanmıştır. Bu çalışmada σ' 'nin bilindiği ve bilinmediği durumlar için ölçü değişmektedir. σ bilinmiyorsa k-NN yöntemi kullanılarak elde edilebileceği ifade edilmiştir. Rasgele tasarım durumu için hataların birbirinden bağımsız ve aynı dağılımlı normal $(0, \sigma^2)$ olduğu varsayılmaktadır.

Wegkamp (2003), veri setini iki parçaya ayırmıştır. İlk parçada her bir model için bir en küçük kareler tahmini üretilip, 2. parça kullanılarak daha önce elde edilen modellerden biri seçilmiştir. Tahmin için bütün veri seti kullanılmamış ancak her bir seferde veri farklı şekillerde seçilip bu prosedür tekrarlanarak seçilen tahminlerin ortalaması alınmıştır. Böylece bütün veri seti hesaplanmıştır. Seçilen indislerin histogramı çizilerek verilerin rasgele bölümlerini alarak ortaya çıkan değişkenlik hakkında bilgiler elde edilebileceği ifade edilmiştir.

4. DOĞRUSAL-DAİRESEL PARAMETRİK OLMAYAN REGRESYON

Dairesel verilerde regresyon analizi, özellikle biyoloji, jeoloji, inşaat mühendisliği ve meteoroloji gibi alanlarda sıklıkla kullanılmaktadır. Dairesel verilerde regresyon analizi, açıklayıcı değişken ve yanıt değişkeninin türüne göre üç farklı yöntem olarak karşımıza çıkmaktadır:

- 1) Dairesel-Dairesel Regresyon (Circular-Circular Regression): Hem yanıt değişkeni hem de açıklayıcı değişken daireysel değişkenlerdir.
- 2) Dairesel-Doğrusal Regresyon (Circular-Linear Regression): Yanıt değişkeni doğrusal, açıklayıcı değişken daireysel değişkendir.
- 3) Doğrusal-Dairesel Regresyon (Linear-Circular Regression): Yanıt değişkeni daireysel, açıklayıcı değişken doğrusaldır.

Bu çalışmada Doğrusal-Dairesel Regresyon yöntemi incelenmiştir. İlk olarak Basit Doğrusal-Dairesel regresyon analizi ile ilgili kısa bilgiler verilecek, daha sonra parametrik olmayan doğrusal-dairesel regresyon yöntemleri incelenecektir.

4.1. Basit Doğrusal-Dairesel Regresyon Analizi

Dairesel verilerde regresyon analizi yapılırken çoğunlukla yanıt değişkeninin Von-Mises dağılımından geldiği varsayımıyla analiz yapılmaktadır. Daha önce de bahsedildiği üzere Von Mises dağılımı, doğrusal veri setleri için kullanılan normal dağılım ile aynı işlevi görmektedir. Normal dağılım gibi tek modlu ve simetrik bir dağılımdır. VM dağılımının yoğunluğu, eşitlik (2.9)'da verildiği üzere $f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)]$, $0 \leq \theta \leq 2\pi, 0 \leq \kappa < \infty$ olarak tanımlanmıştır.

Gould (1969), Laycock, P.J. (1975), Johnson, R.A. ve Wehrly, T. E. (1978), Fisher ve Lee (1992), çalışmalarında yanıt değişkeninin VM dağılımına sahip olduğu varsayımını kullanmışlardır.

Gould (1969)'un tanımladığı regresyon modeli, Fisher (1992) tarafından “Type A” olarak adlandırılmaktadır. Bu modelin ortalama yön için model yapısı eşitlik (4.1)'de verildiği gibi tanımlanmaktadır.

$$\mu = \mu_0 + \sum \beta_j x_j, j = 1, 2, \dots, n \quad (4.1)$$

Eşitlik (4.1)'de tanımlanan model “Barber’s Pole” olarak adlandırılmaktadır. Bu modelde μ_0 herhangi bir sabit, x_j açıklayıcı değişken vektörü ve β_j regresyon katsayısıdır. Bu ifadeye göre, açıklayıcı değişken doğrusal ve yanıt değişkeni dairesel olduğunda elde edilen model eğrisi, sonsuz bir silindirin etrafında sarılmıştır. Önerilen modelin parametrelerinin yinelemeli olarak EÇO yöntemi ile elde edilebileceği ifade edilmiştir. Laycock (1975), bu modeli yorumlamış ve EÇO tahminlerinin EKK tahminlerine eşit olduğunu belirtmiştir. Johnson ve Wehrly (1978), bu modele alternatif olarak, sadece bir doğrusal değişken olduğu durumlarda, açıklayıcı değişken ve yanıt değişkeninin ortak dağılımını kullanarak yeni bir model önermişlerdir.

$\theta_1, \theta_2, \dots, \theta_n$ açısal gözlemlerinin $\mu_1, \mu_2, \dots, \mu_n$ ortalama yön ve $\kappa_1, \kappa_2, \dots, \kappa_n$ yoğunlaşma parametrelili VM dağılımından geldiği varsayılın. Johnson ve Wehrly (1978) tarafından önerilen modelin yapısı aşağıdaki gibidir:

$$\mu_i = \mu_0 + 2\pi F(x_i), \quad i = 1, 2, \dots, n \quad (4.2)$$

Bu modelde $F(x_i)$, tamamen bilinen birikimli dağılım fonksiyonu olup μ ve κ parametreleri doğrudan tahmin edilebilmektedir. (4.2)'de tanımlanan model, Fisher ve Lee (1992) tarafından birden fazla açıklayıcı değişken olduğu durumlar için genişletilmiştir. Önerilen model eşitlik (4.3)'te tanımlanmıştır.

$$\mu_i = \mu_0 + g(\beta^T X_i) \quad (4.3)$$

Burada X_i , i. gözlem için açıklayıcı değişken vektörü, β bilinmeyen regresyon katsayısı ve g “link” yani bağlantı fonksiyonudur. Link fonksiyonu, reel sayılar doğrusunu daireye eşlemektedir. x $(-\infty, +\infty)$ aralığında değiştiğinden, $g(x)$ link fonksiyonunun $(-\pi, +\pi)$ aralığında değerler alması gerekmektedir.

Bu nedenle sadece monoton link fonksiyonları dikkate alınmaktadır. μ 'nün orijin olarak yorumlanabilmesi için $g(0) = 0$ olduğu varsayılacaktır. Bu durumdan ötürü link fonksiyonları “açısal monoton” fonksiyonlar olmalıdır. En yaygın kullanılan link fonksiyonları sırasıyla, $g(x) = 2\pi F(x)$, $g(x) = 2\tan^{-1}(x)$, $g(x) = 2\tan^{-1}(\text{sgn}(x)|x|^d)$

şeklinde olup, $d=0$ olduğu durum, logaritmik dönüşüme tekabül etmektedir. Bu parametre, veri setinden yola çıkılarak kolayca tahmin edilebilir.

Açıklayıcı değişken birden fazla olduğunda da Gould (1969)'un tanımladığı “Barber’s Pole” modeli geçerlidir. Bu durumda, silindirin etrafında birden fazla eğri yer alacağı belirtilmiştir (Fisher ve Lee, 1992).

4.2. Parametrik Olmayan Doğrusal-Dairesel Regresyon Analizi

Bu bölümde açıklayıcı değişken doğrusal, yanıt değişkeni dairesel olduğunda parametrik olmayan regresyon analizi yöntemlerinden NW ile LL yöntemleri incelenmiştir. Kernel düzgünleştiricileri, ilk olarak Fisher (1993: 25) tarafından parametrik olmayan yoğunluk tahmininde kullanılmıştır. $\theta_1, \theta_2, \dots, \theta_n$, n boyutlu bir örneklemden alınan dairesel gözlemler olmak üzere, yoğunluk fonksiyonunun tahmini eşitlik (4.4)'teki gibi elde edilmektedir.

$$\hat{f}(\theta) = \frac{1}{(nh)} \sum_{i=1}^n w\left(\frac{\theta - \theta_i}{h}\right) \quad (4.4)$$

Burada h düzgünleştirme parametresi, w ise ağırlık fonksiyonudur. Doğrusal durumda olduğu gibi burada da düzgünleştirme parametresi h , yoğunluk fonksiyonunun davranışını belirlemede önemli rol oynamaktadır. Burada düzgünleştirme parametresi olan h , örneklem boyutu n 'e ve veri setinin yayılımına bağlıdır. Ardından Fisher (1993: 173), zaman serisi modelleri için herhangi bir “ t ” noktasındaki Kernel tahmin edicisinin değerini aşağıdaki eşitlikleri kullanarak tanımlamıştır.

$$\hat{x}(t) = \frac{\sum_{i=1}^n K_h(t-t_i)x_i}{\sum_{i=1}^n K_h(t-t_i)} \quad (4.5)$$

$$\hat{y}(t) = \frac{\sum_{i=1}^n K_h(t-t_i)y_i}{\sum_{i=1}^n K_h(t-t_i)} \quad (4.6)$$

Burada $K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$ ‘dır. $\theta(t)$ ’nin düzgün tahmini ise eşitlik (4.7) kullanılarak tahmin edilmektedir.

$$\hat{\theta}(t) = \tan^{-1} \left(\hat{y}(t) / \hat{x}(t) \right) \quad (4.7)$$

4.2.1. Model yapısı

Kernel düzgünleştiricilerinin doğrusal-dairesel parametrik olmayan dairesel regresyona uygulanması Di Marzio ve diğerleri (2012) tarafından gerçekleştirilmiştir. Mevcut Kernel düzgünleştiricisi yöntemleri, dairesel veri yapısına uyarlanarak modeller kurulmuştur. $\{(\Delta_i, \Theta_i), i = 1, 2, \dots, n\}$, Δ doğrusal veya dairesel bir rasgele değişkenden alınan bir örneklem, Θ ise dairesel bir rasgele değişkenden alınan bir örneklem olsun. Burada Δ , kapsamlı tanım kümesi olup hem doğrusal hem de açıklayıcı dairesel değişkenleri içermektedir. Bu değişkenler arasındaki bağımlılık, eşitlik (4.8)'de verilen risk fonksiyonunun mümkün olduğunda küçük tutulmasıyla elde edilmektedir.

$$E[1 - \cos(\Theta - m(\Delta))] \quad (4.8)$$

$\delta \in \Delta$ ve m ölçülebilir bir fonksiyon olmak üzere,

$$\hat{m}_1(\delta) = E[\sin(\Theta) | \Delta = \delta], \quad (4.9)$$

$$\hat{m}_2(\delta) = E[\cos(\Theta) | \Delta = \delta], \quad (4.10)$$

ve f yoğunluğun tasarımı olmak üzere, $g_i = m_i f$ olsun. Yukarıda tanımlanan risk fonksiyonu $m(\delta) = \text{atan}^2[g_1(\delta), g_2(\delta)]$ ile minimize edilir. Burada verilen “atan²” fonksiyonu, x eksenini ile vektör arasındaki açıyı, orijin noktasından koordinat eksenindeki (x, y) noktasına taşımaktadır. Bu yaklaşımdan yararlanarak açıklayıcı değişken ile yanıt değişkeni arasındaki ilişkiyi tanımlayan model eşitlik (4.11)'deki gibi kurulmuştur.

$$\Theta_i = [m(\Delta_i) + \epsilon_i] \bmod(2\pi), i = 1, 2, \dots, n \quad (4.11)$$

ϵ_i açısız hataları, Δ_i 'lerden bağımsız, sıfır ortalama yöne ve sınırlı yoğunluğa sahiptir.

Düzgünleştiriciler tanımlanmadan önce, $g_1(\delta) / g_2(\delta)$ oranı için $\hat{g}_1(\delta) / \hat{g}_2(\delta)$ oranının asimptotik olarak yansız olabilmesi için tasarlanan örneklem istatistikleri

$\hat{g}_1(\delta) = \frac{1}{n} \sum_{i=1}^n \sin(\Theta_i) W(\Delta_i - \delta)$ ve $\hat{g}_2(\delta) = \frac{1}{n} \sum_{i=1}^n \cos(\Theta_i) W(\Delta_i - \delta)$ olarak tanımlanmıştır. Burada W , yerel ağırlıkları temsil etmektedir. Sonuç olarak, bu bölümde bahsedilen tahmin düzgünleştiricilerin herhangi bir δ noktasındaki regresyon fonksiyonunun değeri eşitlik (4.12) ile elde edilmektedir.

$$\hat{m}(\delta) = \text{atan } 2[\hat{g}_1(\delta), \hat{g}_2(\delta)] \quad (4.12)$$

4.2.2. Doğrusal-Dairesel Nadaraya-Watson düzgünleştiricisi

$\{(X_i, \Theta_i), i = 1, 2, \dots, n\}$, X doğrusal rasgele değişkenden ve Θ ise dairesel bir rasgele değişkenden alınan bir örnekleme temsil ediyor olsun. Bu durumda dairesel yanıt değişkeni Θ , doğrusal açıklayıcı değişken X olmaktadır. Di Marzio ve diğerleri (2012)'de bahsedildiği üzere, açıklayıcı değişken doğrusal olduğunda, ağırlıklandırma için Öklidyen Kernel temelli fonksiyonlar kullanılabilir. Yani Gaussian Kernel, Epanechnikov Kernel gibi fonksiyonlar kullanılarak ağırlıklandırma yapılabilir. Kernel fonksiyonunun $[-1,1]$ aralığında tanımlı, sıfır noktasında en büyük değerini alan simetrik bir yoğunluk fonksiyonu olduğundan daha önceki bölümlerde bahsedilmişti. Bu çalışmada Kernel ağırlık fonksiyonlarının oluşturulmasında Gaussian Kernel fonksiyonu kullanılmıştır.

Herhangi bir x noktasında doğrusal-dairesel NW tahmin edicisi aşağıdaki eşitlik kullanılarak elde edilmektedir.

$$\hat{f}_{NWLC}(x) = \frac{\sum_{i=1}^n \Theta_i K_{\vartheta}(x-x_i)}{\sum_{i=1}^n K_{\vartheta}(x-x_i)} \quad (4.13)$$

Eşitlik (4.13)'te yer alan ϑ parametresi, düzgünleştirme parametresini (bant genişliği) ve K_{ϑ} ise Gaussian Kernel fonksiyonunu temsil etmektedir. Bu ifade sonucunda elde edilen değerler, eşitlik (4.12) ile tanımlanan regresyon fonksiyonu ile aynı sonucu vermektedir.

4.2.3. Doğrusal-Dairesel Yerel Lineer Kernel düzgünleştiricisi

Di Marzio ve diğerleri (2012), açıklayıcı değişken doğrusal olduğunda LL Kernel düzgünleştiricisinin tahmin değerlerini elde etmek için eşitlik (4.12)'de tanımlanan

fonksiyondan yararlanılacağını belirtmiştir. Burada da NW düzgünleştiricisinde olduğu gibi, Kernel fonksiyonu olarak Gaussian Kernel kullanılmıştır. Ağırlıklandırma fonksiyonu eşitlik (4.4)'te tanımlandığı gibidir.

$$W(x_i - x) = \frac{1}{n} K_g(x_i - x) \left\{ \sum_{j=1}^n K_g(x_j - x) (x_j - x)^2 - (x_j - x) \sum_{j=1}^n K_g(x_j - x) (x_j - x) \right\} \quad (4.14)$$

$\hat{g}_1(x) = \frac{1}{n} \sum_{i=1}^n \sin(\Theta_i) W(x_i - x)$ ve $\hat{g}_2(x) = \frac{1}{n} \sum_{i=1}^n \cos(\Theta_i) W(x_i - x)$ olmak üzere, herhangi bir x noktasındaki doğrusal-dairesel LL tahmin edicisi eşitlik (4.15) kullanılarak elde edilmektedir.

$$\hat{f}_{LLCL}(x) = \text{atan } 2[\hat{g}_1(x), \hat{g}_2(x)] \quad (4.15)$$

Burada da ϑ parametresi, düzgünleştirme parametresini (bant genişliği) ve K_g ise Gaussian Kernel fonksiyonunu temsil etmektedir.

4.2.4. Dairesel-Doğrusal parametrik olmayan regresyonda bant genişliği seçimi

Bant genişliği seçimi, doğrusal parametrik olmayan regresyonda olduğu kadar dairesel parametrik olmayan regresyonda da oldukça önemlidir. Dairesel verilerde de genellikle hata kareler ortalamasını minimize eden değer bant genişliği olarak tercih edilmektedir. Di Marzio ve diğerleri (2012), eşitlik (4.2)'de tanımlanan risk ölçüsünü minimize eden değeri bant genişliği olarak kullanmıştır. Ayrıca bu risk ölçüsünün asimptotik olarak hata kareler ortalamasına yakınsadığını da ifade etmiştir. En eski ve uygulanması en kolay bant genişliği seçim yöntemlerinden birinin CV olduğundan önceki bölümlerde bahsedilmişti. Bu çalışmada da bant genişlikleri, CV yöntemi ile elde edilmiştir. Oliveira ve diğerleri (2013) ve Di Marzio ve diğerleri (2012)'de kullanıldığı üzere, Oliveira ve diğerleri (2014), dairesel yanıt değişkenleri içeren modeller için CV kullanarak bant genişliğini elde etmek için eşitlik (4.16)'dan yararlanmıştır.

$$\vartheta = \min \left(\sum_{i=1}^n \left(-\cos(\Theta_i - \hat{f}^{-i}(x_i)) \right) \right), i = 1, 2, \dots, n \quad (4.16)$$

Eşitlik (4.16)'yı minimize eden değer, optimal bant genişliği olarak seçilmektedir. Burada $\hat{f}^{-i}, (X_i, \Theta_i)$ gözlem çifti veri setinden çıkarıldıktan sonra elde edilen tahmin değerini

ifade etmektedir. Her bir gözlem çifti teker teker veri setinden çıkarılarak hesaplamalar yapıldığından bu yöntem “Leave-One-Out CV” olarak adlandırılmaktadır.

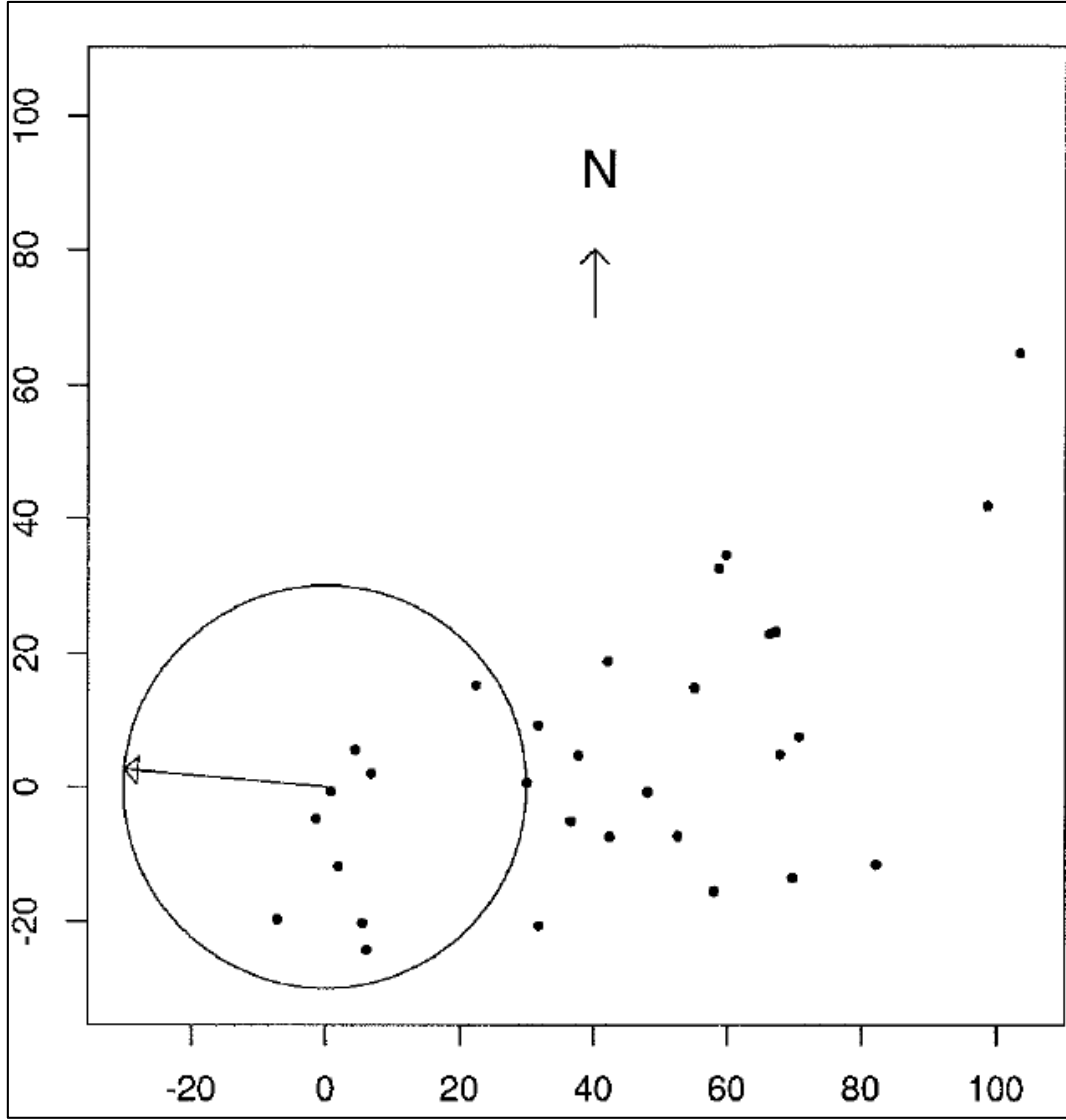
4.2.5. Gerçek veri uygulaması

Bu çalışmada Mavi Deniz Salyangozu (Blue Periwinkles) veri seti kullanılmıştır. Chapman (1986), Chapman ve Underwood (1992) ve Underwood ve Chapman (1985, 1989, 1992) tarafından mavi deniz salyangozlarının yaşadıkları yükseklikten kıyıya bırakıldıktan sonraki davranışlarını incelemek üzere bir dizi deney yapılmış ve mavi deniz salyangozlarının kıyıya bırakıldıktan sonraki hareket yönü ve bırakıldıkları noktaya uzaklıkları gözlemlenmiştir. Toplam 31 adet deniz salyangozunun davranışları incelenmiştir. Bunlardan 15 tanesinin davranışlarıyla ilgili ölçümler yeni yaşam alanına bırakıldıktan bir gün sonra, kalan 16 tanesinin davranışları ile ilgili ölçümler ise dört gün sonra alınmıştır. İki farklı konumdan elde edilen deney sonuçları, doğrusal-dairesel regresyon örneğinde kullanılmak üzere birleştirilmiştir (Fisher 1993:136). Mavi Deniz Salyangozu veri seti, daha önce Fisher (1993), Di Marzio ve diğerleri (2012) ve Oliviera(2014) tarafından parametrik olmayan regresyon analizinde kullanılmış ancak aykırı gözlem durumu incelenmemiştir. Bu veri setinde uzaklıklar açıklayıcı değişken, hareket yönü ise yanıt değişkenidir. Yanıt değişkeninin VM dağılımından gelip gelmediğini test etmek için “Watson’s Test” uygulanmış ve $\alpha = 0,05$ anlam düzeyinde, veri setinin VM dağılımından gelmediği sonucuna ulaşılmıştır.

Watson's Test for the von Mises Distribution

Test Statistic: 0.1669
Level 0.05 Critical Value: 0.101
Reject Null Hypothesis

Yukarıda yer alan program çıktısından yola çıkarak, açıklayıcı değişkeni doğrusal, bağımlı değişkeni dairesel olan Doğrusal-Dairesel (Linear-Circular) parametrik olmayan regresyon analizi uygulanmıştır.



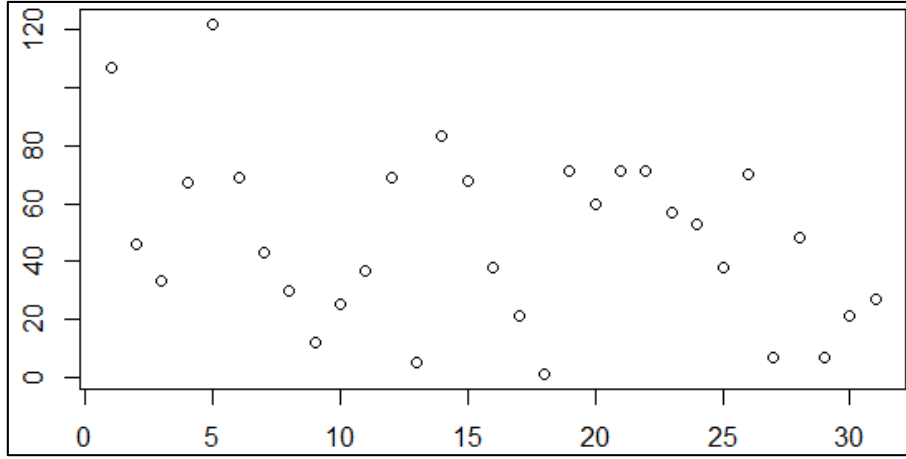
Şekil 4.1. Mavi deniz salyangozu veri setinin şekil üzerinde gösterimi (Fisher,1993:138)

Şekil (4.1)'de gösterilen yön, her bir deniz salyangozu için bırakılma noktasına göre pusula yönüdür (Fisher,1993:138). Ok ise denizin yaklaşık yönünü (275°) göstermektedir. Mavi deniz salyangozu veri seti aşağıdaki çizelgede verilmiştir. x rasgele değişkeni uzaklığı (cm), θ rasgele değişkeni ise hareket yönünü temsil etmektedir.

Çizelge 4.1. Mavi Deniz Salyangozu (“Periwinkles”) veri seti

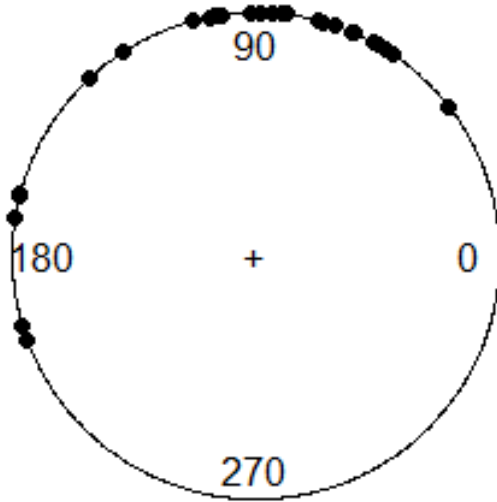
n	x (cm)	θ
1	107	67
2	46	66
3	33	74
4	67	61
5	122	58
6	69	60
7	43	100
8	30	89
9	12	171
10	25	166
11	37	98
12	69	60
13	5	197
14	83	98
15	68	86
16	38	123
17	21	165
18	1	133
19	71	101
20	60	105
21	71	71
22	71	84
23	57	75
24	53	98
25	38	83
26	70	71
27	7	74
28	48	91
29	7	38
30	21	200
31	27	56

Değişkenlerin saçılım grafikleri şekil (4.2) ve şekil (4.3)’te verilmiştir.



Şekil 4.2. Uzaklık değişkeninin saçılım grafiği

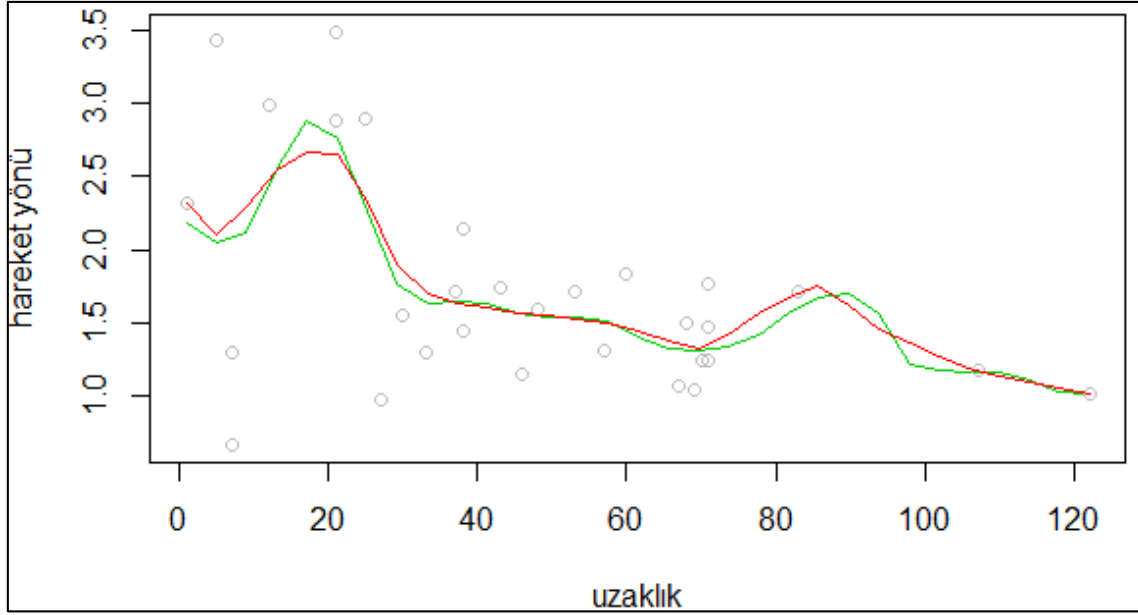
Şekil 4.2. 'deki x eksenini gözlemleri, y eksenini ise uzaklığı göstermektedir.



Şekil 4.3. Hareket yönü değişkeninin dairesel saçılım grafiği

Bant genişliği seçimi için eşitlik (4.16)'da verilen çapraz doğrulama yöntemi uygulanmıştır. NW yöntemi için bant genişliği $\vartheta_{NW} = 5,35935$, LL yöntemi için bant genişliği ise $\vartheta_{LL} = 6,924234$ olarak elde edilmiştir.

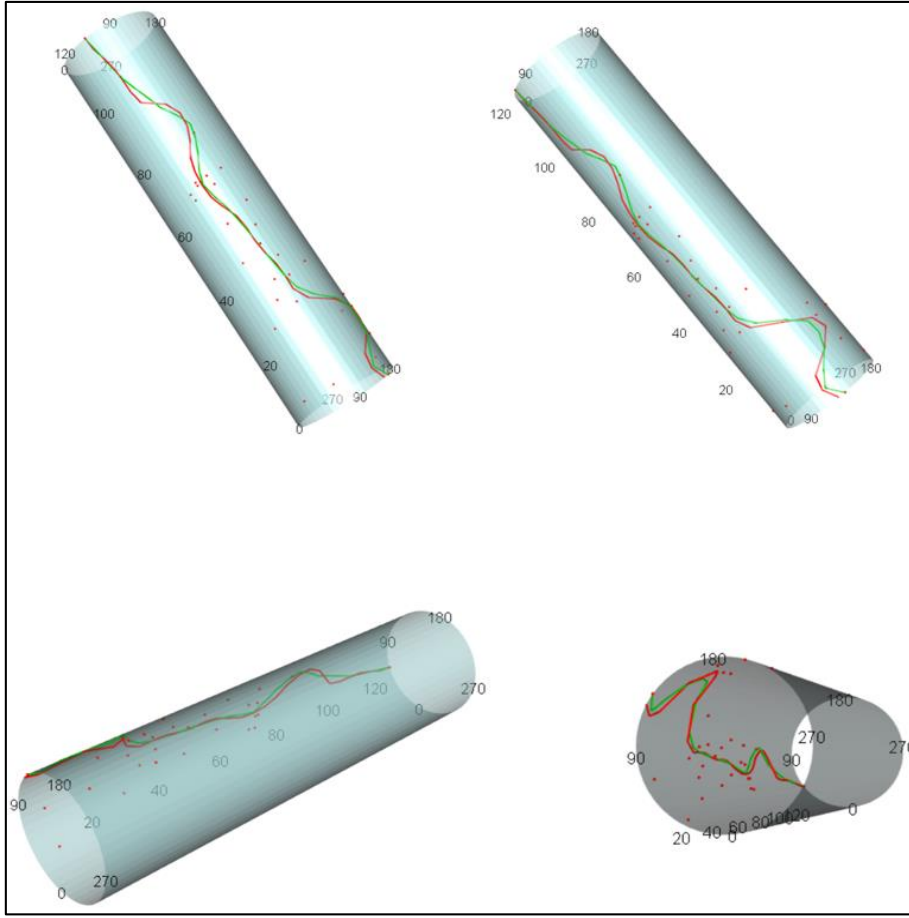
Parametrik olmayan dairesel regresyon yöntemlerinden NW ile LL yöntemleri uygulandığında elde edilen eğriler şekil (4.4)'te verilmiştir.



Şekil 4.4. NW ve LL düzgünleştiricileri uygulandığında elde edilen eğriler

“R” paket programında parametrik olmayan doğrusal-dairesel regresyon analizi için kullanılan komut, radyan cinsinden verilerle çalıştığından, hareket yönü değişkeni dereceden radyana dönüştürülmüştür. Şekil 4.4.’te yeşil eğri NW düzgünleştiricisi kullanıldığında elde edilen eğriyi, kırmızı eğri ise LL düzgünleştiricisi kullanıldığında elde edilen eğriyi göstermektedir.

Açıklayıcı değişken doğrusal, yanıt değişkeni ise dairesel olduğunda elde edilen eğrilerin bir silindir etrafında sarılı eğriler olduğundan, 4.1. bölümde bahsedilmiştir. Bu veri seti için elde edilen eğrilerin silindir üzerinde gösterimi Şekil (4.5)’te verilmiştir. Program çıktısı üç boyutlu olduğundan Şekil (4.5), farklı açılardan ekran görüntüsü alınarak elde edilmiştir.



Şekil 4.5. Elde edilen eğrilerin silindir üzerinde gösterimi

Kullanılan yöntemlerin karşılaştırılması için, Di Marzio (2012) tarafından kullanılan ve hataların bir fonksiyonu olan $R = \sum_{i=1}^n (1 - \cos(\theta_i - \hat{\theta}_i))$ risk fonksiyonu kullanılmıştır. Elde edilen sonuçlar aşağıda verilmiştir:

Çizelge 4.2. Mavi Deniz Salyangozu veri setine NW ve LL yöntemlerinin uygulanmasıyla elde edilen risk fonksiyonları

	Yöntem	
	NW	LL
R	29,1048	30,65747

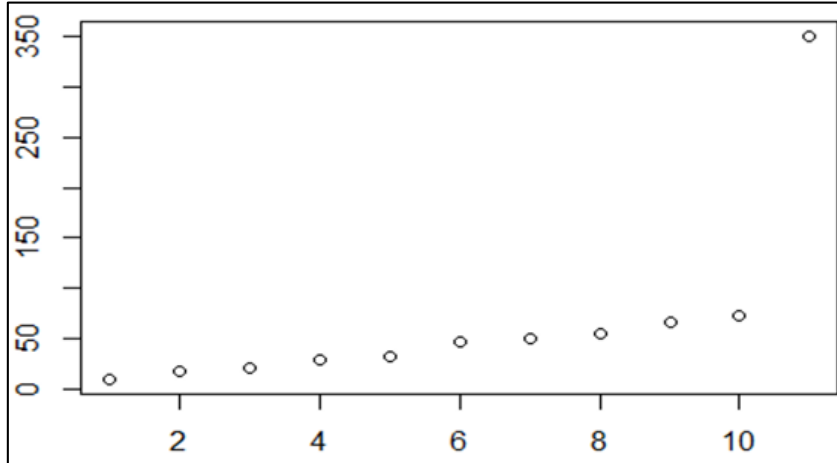
Elde edilen sonuçlardan da açıkça görüldüğü üzere, NW düzgünleştiricisi ile LL düzgünleştiricisi aynı veri setine uygulandığında, NW düzgünleştiricisinin risk ölçüsü daha küçük çıktığı görülmektedir. İki yöntem arasındaki fark çok büyük olmamasına rağmen

NW düzgünleştiricisinin, LL düzgünleştiricisine kıyasla veri setinin gerçek fonksiyonel yapısına daha yakın sonuç verdiği gözlenmiştir.

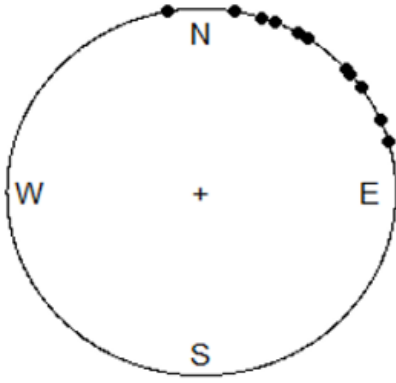
5. DAİRESEL VERİLERDE AYKIRI GÖZLEM

Aykırı gözlem kavramı, istatistikte bir yığından seçilen rasgele bir örneklemdaki diğer gözlemlerden farklı davranış sergileyen ve diğer gözlem değerleri ile aralarında belirgin ölçüde mesafe olan gözlemler için kullanılmaktadır. Aykırı gözlemler, kullanılan veri setinin veya modelin parametreleri üzerinde güçlü bir etkiye sahiptir. Bu nedenle bu gözlemlerin varlığında elde edilen tahminler güvenilir sonuçlar vermemektedir (Mahmood ve diğerleri, 2017).

Dairesel verilerde aykırı gözlem kavramı, doğrusal durumdaki aykırı gözlem kavramından farklıdır. Doğrusal bir veri seti ile çalışırken, çok büyük veya çok küçük değerler aykırı gözlem olabilecekken, dairesele veri setleri için bu durum geçerli olmayabilir. Örnek olarak (10, 18, 22, 30, 33, 48, 50, 55, 67, 74, 350) veri seti ele alındığında, doğrusal bir veri setiyle çalıştığımızı varsayarsak, 11. gözlem değerinin (350) aykırı gözlem olduğunu söyleyebiliriz. Ancak dairesele istatistikte, dairesele uzaklık kavramı ve veri setinin birim çember üzerinde yer almasından dolayı bu gözlemin aykırı gözlem olduğunu söyleyemeyiz.



Şekil 5.1. Örnek veri setinin doğrusal gösterimi



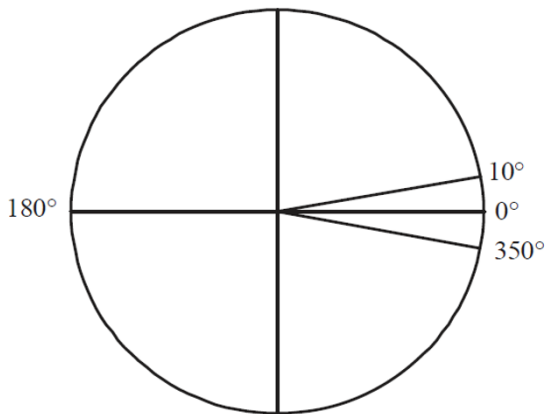
Şekil 5.2. Örnek veri setinin dairesel düzlem üzerinde gösterimi

Şekil 5.1., yukarıda verilen veri setinin doğrusal gösterimini ifade etmektedir. Şekilden de anlaşılacağı üzere 11. gözlem değeri aykırı değerdir. Ancak Şekil 5.2.'ye bakıldığında 11. gözlemin veri setinin kalan kısmıyla uyumlu olduğu görülmektedir. Bu iki şekilden yararlanarak, dairesel verilerde dairesel uzaklık kavramının doğrusal durumlar için tanımlanan uzaklık kavramından farklı olduğu söylenebilir.

$\theta_1, \theta_2, \dots, \theta_n$, n boyutlu bir örneklemden alınan açısal gözlemler olsun. Jammalamadaka ve Sengupta (2001), dairesel uzaklığı aşağıdaki gibi tanımlamıştır:

$$cd = \pi - \left| \pi - |\theta_i - \theta_j| \right| \quad (5.1)$$

Bu tanıma göre, iki açısal gözlem arasındaki maksimum uzaklığın π 'den büyük olamayacağı açıkça görülmektedir.



Şekil 5.3. Dairesel verilerde uzaklık kavramı

Şekil 5.3.'te görüldüğü üzere, iki açısal gözlemin değerleri $\theta_1 = 10^\circ$, $\theta_2 = 350^\circ$ olsun. Bu iki açısal gözlem arasındaki uzaklık, 340° olarak düşünülse de, iki açısal gözlem arasındaki uzaklık π 'den büyük olamayacağından, gerçek dairesel uzaklık 20° 'dir (Mahmood ve diğerleri, 2017).

5.1. Tek Değişkenli Dairesel Verilerde Aykırı Gözlem Kavramı

Dairesel verilerde aykırı değer tespiti ile ilgili bu zamana kadar yapılan çalışmaların tamamı, veri setinin dağılımı bilindiğinde uygulanabilmektedir.

Çalışmaların büyük çoğunluğunda $\theta_1, \theta_2, \dots, \theta_n$ dairesel gözlemlerinin Von Mises dağılımından geldiği varsayılmaktadır. Von Mises dağılımı, 2.4.2'de ayrıntılı olarak incelenmiştir. Bu bölümde tek değişkenli veri setlerinde aykırı gözlem tespiti ile ilgili önerilen yöntemlerden kısaca bahsedilmiştir. Bahsedilen yöntemlerin çoğu, ortalama bileşke uzunluk gibi tanımlayıcı istatistikleri kullanmaktadır.

5.1.1. Mardia istatistiği

Mardia (1975), tek değişkenli dairesel bir veri setinde tek bir aykırı gözlem tespiti için Mardia istatistiğini önermiştir. Bu istatistiğe göre, bileşke uzunluk üzerinde en fazla etkiye sahip olan gözlem aykırı gözlemdir.

$$Mar = \min \left(\frac{n-1-R_{(-i)}}{n-R} \right) \quad (5.2)$$

Burada $R_{(-i)}$, i. gözlem veri setinden çıkarıldığında elde edilen bileşke uzunluk ve R ise tüm gözlemleri içeren bileşke uzunluktur. n , veri setindeki gözlem sayısını ifade etmektedir.

5.1.2. M istatistiği

Collett (1980) tarafından önerilen M istatistiği de Mardia istatistiği gibi bileşke uzunluğa bağlıdır. Bu yöntem de tek değişkenli bir dairesel veri setinde tek bir aykırı gözlem tespiti için kullanılmaktadır. M istatistiği eşitlik (5.3)'teki gibi tanımlanmıştır.

$$M = maks\left(\frac{R_{(-i)}-R+1}{n-R}\right) = \frac{R_k-R+1}{n-R} \quad (5.3)$$

Burada $R_k = maks(R_{(-i)})$ 'dir.

5.1.3. A ve A* istatistiği

Abuzaid (2009), aykırı gözlem tespiti için Rao (1969) tarafından önerilen dairesel uzaklığı kullanmıştır. θ_i ve θ_j gözlemleri arasındaki dairesel uzaklık eşitlik (5.4.)'te tanımlanmıştır.

$$d_{ij} = 1 - \cos(\theta_i - \theta_j) \quad (5.4)$$

Kosinüs fonksiyonu $[-1,1]$ aralığında değer alabileceğinden, d_{ij} uzaklığı $[0,2]$ aralığında tanımlı olacaktır. Veri setindeki tüm dairesel gözlemlerin dairesel uzaklıklarının toplamı eşitlik (5.5) kullanılarak elde edilmektedir.

$$D_j = \sum_{i=1}^n (1 - \cos(\theta_i - \theta_j)), j = 1, 2, \dots, n \quad (5.5)$$

Yukarıdaki dairesel uzaklık toplamına göre, eğer θ_j gözlemi aykırı değer ise, diğer gözlemlerden uzakta yer alacak ve bu durum D_j istatistiğinin değerinin artmasına yol açacaktır. Bu durumdan yola çıkarak, tek değişkenli bir veri setinde tek bir aykırı gözlem tespiti için geliştirilen A istatistiği, θ_j gözlem veri setinden çıkarıldığında elde edilen $\frac{D_j}{n-1}$ ortalama dairesel uzaklığına bağlıdır. A istatistiği eşitlik (5.6.)'da tanımlanmıştır.

$$A = maks\left(\frac{D_j}{2(n-1)}\right), j = 1, 2, \dots, n \quad (5.6)$$

Burada $A \in [0,1]$ olduğu açıktır. Jammalamadaka ve SenGupta (2001), herhangi iki dairesel nokta arasındaki uzaklığın, birim çember üzerinde bu iki nokta arasındaki en kısa yay uzunluğu olduğunu dikkate alarak, eşitlik (5.7)'deki dairesel uzunluk ölçüsünü tanımlamıştır.

$$d_{ij}^* = \pi - \left| \pi - |\theta_i - \theta_j| \right|, d_{ij}^* \in [0, \pi] \quad (5.7)$$

Abuzaid (2010), d_{ij}^* eşitliğini kullanarak D_j istatistiğini ve A istatistiğine alternatif olarak A^* istatistiğini yeniden hesaplamıştır.

$$D_j^* = \sum_{i=1}^n \left(\pi - \left| \pi - |\theta_i - \theta_j| \right| \right), j = 1, 2, \dots, n \quad (5.8)$$

$$A^* = \max \left(\frac{D_j^*}{n-1} \right), j = 1, 2, \dots, n \quad (5.9)$$

A^* istatistiği, $[0, \pi]$ aralığında değer almaktadır. Uzaklık tanımları güncellense de A istatistiği ile A^* istatistiğinin benzer performans sergiledikleri gözlenmiştir.

5.1.4. Chord (Kiriş) istatistiği

Abuzaid (2010) tarafından geliştirilmiş olup, tek değişkenli bir veri setinde sadece bir aykırı gözlem tespiti için kullanılmaktadır. Çember üzerinde yer alan farklı iki noktayı birleştiren parçaya kiriş adı verilmektedir. Bu yöntem, kirişin geometrik özelliklerinden yararlanarak geliştirilmiştir. İki nokta arasındaki kirişin uzunluğu eşitlik (5.10) kullanılarak hesaplanmaktadır:

$$crd(d_{ij}) = 2r \sin \frac{d_{ij}^*}{2r} \quad (5.10)$$

Burada r , yarıçapı temsil etmektedir. Birim çember üzerinde yarıçap bire eşit olur ve d_{ij}^* eşitlik (5.7) kullanılarak elde edilir. Birim çember üzerinde yer alan θ_j noktasından geçen bütün kirişlerin toplamı eşitlik (5.11)'de ifade edilmiştir.

$$S_j = \sum_{i=1}^n crd(d_{ij}) = 2 \sum_{i=1}^n \sin \frac{d_{ij}^*}{2}, j = 1, 2, \dots, n \quad (5.11)$$

Burada $0 \leq S_j \leq 2(n-1)$ olduğu açıktır.

$d_{ij}^* = 0$ olduğunda, $\sum_{i=1}^n \sin \frac{d_{ij}^*}{2} = 0$, $d_{ij}^* = \pi$ olduğunda ise, $\sum_{i=1}^n \sin \frac{d_{ij}^*}{2} = n-1$ 'dir. Eğer S_j en büyük değerine ulaştıysa, θ_j gözleminin aday aykırı gözlem olduğu sonucuna varılmaktadır. Chord istatistiği eşitlik (5.12)'de tanımlanmıştır.

$$Chord = maks\left(\frac{s_j}{2(n-1)}\right), j = 1, 2, \dots, n \quad (5.12)$$

5.1.5. RCD_u istatistiği

Mahmood ve diğerleri (2017), tek değişkenli veri setindeki aykırı gözlemleri tespit edebilmek için sağlam (robust) dairesel uzaklığa dayalı bir yöntem önermişlerdir.

Dairesel medyanın sağlamlık özelliğinden yararlanarak geliştirilen bu yöntemde, veri setinin VM dağılımından geldiği varsayılarak kesim noktaları (cut-off points) elde edilmiştir. Test istatistiğinin hesaplanması için, her bir dairesel gözlem ile veri setinin dairesel medyanı arasındaki uzaklık ölçü olarak kullanılmıştır. Dairesel uzaklık, dairesel veri setinin geometrisinden dolayı birden fazla şekilde hesaplanabilir. Önerilen yöntemde, dairesel medyan aşağıdaki gibi elde edilmektedir:

i) $0 \leq medyan \leq \pi$ ise;

$$dist_{(i)} = \begin{cases} |\theta_i - medyan|, & |\theta_i - medyan| \leq \pi \\ 2\pi - \theta_i + medyan, & |\theta_i - medyan| > \pi \end{cases} \quad (5.13)$$

ii) $\pi \leq medyan \leq 2\pi$

$$dist_{(i)} = \begin{cases} |\theta_i - medyan|, & |\theta_i - medyan| \leq \pi \\ 2\pi - medyan + \theta_i, & |\theta_i - medyan| > \pi \end{cases} \quad (5.14)$$

Eğer θ_i gözlemi aykırı gözlem ise, $dist_{(i)}$ uzaklığının büyük olması beklenir. Bu nedenle kesim noktası eşitlik (5.15) ile belirlenir.

$$RCD_u = maks(dist) \quad (5.15)$$

θ_i gözleminin uzaklığı $dist_{(i)}$, kesim noktasından büyükse bu gözlem aykırı gözlem olarak tanımlanmaktadır.

Aykırı gözlem tespit yöntemlerinin başarısı, üç farklı ölçü ile belirlenmektedir (Mahmood ve diğerleri (2017)). Bunlar;

i) Maskeleme oranı (M): Tespit edilemeyen aykırı gözlemlerin oranı

- ii) Sahte aykırı gözlem oranı (S): Aykırı gözlem olmayan gözlemlerin aykırı gözlem olarak tespit edilmesi oranı
- iii) Doğru tespit oranı (DTO): Doğru tespit edilen aykırı gözlem oranı.

$$M = \frac{\text{aykırı gözlem olduğu halde tespit edilemeyen gözlem sayısı}}{\text{aykırı gözlem sayısı}} \quad (5.16)$$

$$S = \frac{\text{aykırı gözlem olmadığı halde aykırı gözlem olarak tespit edilen gözlem sayısı}}{n-M} \quad (5.17)$$

Eğer kullanılan yöntemde doğru tespit oranı bire, maskeleye ve sahte aykırı gözlem oranları sıfıra yakın değerler alıyorsa, kullanılan yöntemin başarılı olduğu sonucuna varılır. Bu bölümde bahsedilen yöntemlerin performansları, yukarıdaki ölçüler kullanılarak incelenmiştir. Önerilen yöntemin performansı için de aynı ölçüler kullanılacaktır.

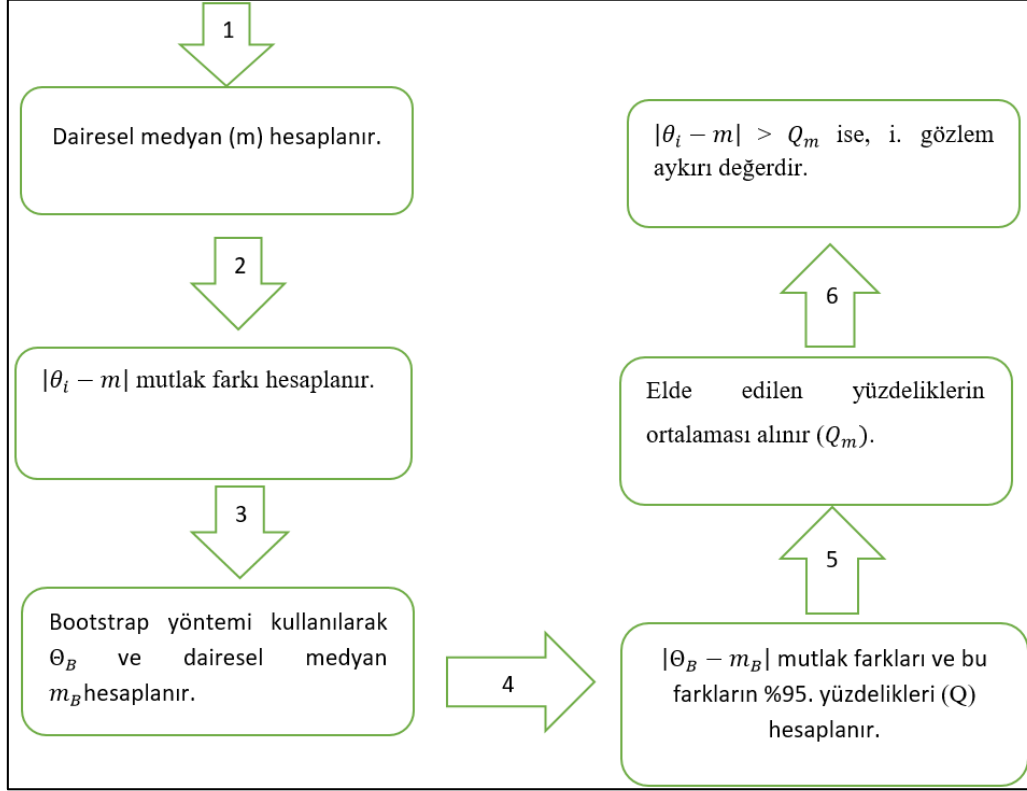
5.1.6. Önerilen istatistik

Yukarıda bahsedilen yöntemlerin hepsi, veri setinin dağılımının bilindiği ve VM olduğu varsayımı altında uygulanabilmektedir. Bugüne kadar veri setinin dağılımıyla ilgili herhangi bir önsel bilgi olmadığında uygulanabilecek bir yöntem geliştirilmemiştir. Bu bölümde önerilen istatistik için herhangi bir dağılım varsayımı gerekmemektedir. Ayrıca, dairesel medyanın sağlamlık özelliğinden yararlanıldığından sağlam bir yöntemdir. Önerilen yöntemde amaç, gözlemlerin dairesel medyana olan uzaklıklarından yararlanarak aykırı gözlem tespiti için eşik değeri (“cut off point”) elde etmektir. Herhangi bir gözlem değerinin dairesel medyandan farkı eşik değerinden büyükse, o gözlemin aykırı gözlem olduğu tespit edilmiştir. Burada eşik değerleri bootstrap tekniği ve yüzdeliklerden yararlanarak hesaplanmaktadır. Önerilen yöntemin algoritması aşağıda verildiği gibidir:

$\Theta = \theta_1, \theta_2, \dots, \theta_n$, n boyutlu bir örneklemden alınan açısal gözlemler olmak üzere;

- 1) Veri setinin dairesel medyanı hesaplanır. (m)
- 2) $|\theta_i - m|$ mutlak farkı hesaplanır.
- 3) $|\theta_i - m|$ mutlak farkı için, bootstrap yöntemi kullanılarak n boyutlu Θ_B örneklemeleri elde edilir. Elde edilen her bir örneklem vektörünün dairesel medyanı m_B hesaplanır.
- 4) $|\Theta_B - m_B|$ mutlak farkları ve bu farkların %95. Yüzdelikleri (Q) hesaplanır.

- 5) Elde edilen yüzdeliklerin ortalaması alınır (Q_m). Bu değer eşik değeridir.
 $|\theta_i - m| > Q_m$ ise θ_i gözlemi aykırı gözlemdir.



Şekil 5.4. Tek değişkenli veri setleri için önerilen yöntemin akış diyagramı

5.1.7. Simülasyon çalışması

Tek değişkenli veri setlerinde önerilen aykırı gözlem tespit yöntemiyle ilgili bir dizi simülasyon çalışması yapılmıştır. İlk olarak VM dağılımından farklı κ parametreleri için veri setleri üretilmiştir. $n=20, 30, 60$ ve bozulma dereceleri $0,70, 0,75, 0,80, 0,85, 0,90, 0,95$ alınarak 1000 deneme yapılmıştır. Üretilen veri setlerine farklı derece ve farklı oranlarda aykırı gözlem eklenmiştir. Veri setine tek bir aykırı gözlem eşitlik (5.18) kullanılarak eklenmiştir (Mahmood ve diğerleri, 2017).

$$\theta_k = \theta + \gamma\pi \bmod(2\pi) \quad (5.18)$$

Burada θ_k kirletilen gözlemin yeni değeri iken γ parametresi bozulma derecesi olup,

$0 \leq \gamma \leq 1$ 'dir. γ değeri sıfıra yaklaştıkça verideki bozulma derecesi azalır. Yani orijinal gözlem, diğer gözlemlerden fazla uzaklaşmış olmaz. γ değeri 1'e yaklaştıkça verideki bozulma derecesi artar, bozulan gözlem değeri diğer gözlemlerden uzaklaşır. Simülasyon çalışması sonucunda elde edilen sonuçlar aşağıdaki çizelgelerde verilmiştir.

DTO doğru tespit oranı, M maskeleme, S sahte aykırı gözlem oranını ve Sd standart sapmayı temsil etmektedir.

Çizelge 5.1. $X \sim VM(3,20)$, $n=20$ olduğunda elde edilen sonuçlar

γ	aykırı gözlem	DTO	M	Sd(M)	S	Sd(s)
0,70	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,9326	0,0674	0,1347	0	0
	4	0,6712	0,3287	0,1728	0	0
0,75	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,942	0,058	0,1264	0	0
	4	0,679	0,321	0,1742	0	0
0,80	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,96	0,04	0,1083	0	0
	4	0,7012	0,2988	0,172	0	0
0,85	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,9653	0,0347	0,1018	0	0
	4	0,7137	0,2862	0,1711	0	0
0,90	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,9803	0,0196	0,0785	0	0
	4	0,728	0,272	0,1685	0	0
0,95	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,985	0,015	0,0691	0	0
	4	0,741	0,259	0,1759	0	0

Çizelge 5.2. $n=30$, $X \sim VM(3,20)$ olduğunda elde edilen sonuçlar

γ	aykırı gözlem	DTO	M	Sd(M)	S	Sd(s)
0,70	1	1	0	0	0,0001	0,0021
	2	1	0	0	0	0
	3	0,995	0,005	0,0405	0	0
	4	0,884	0,116	0,1371	0	0
0,75	1	1	0	0	0,0001	0,0018
	2	1	0	0	0	0
	3	0,9983	0,0017	0,0235	0	0
	4	0,9005	0,0995	0,1289	0	0
0,80	1	1	0	0	0,0001	0,0018
	2	1	0	0	0	0
	3	0,999	0,001	0,0182	0	0
	4	0,914	0,086	0,1234	0	0
0,85	1	1	0	0	0	0
	2	1	0	0	0	0
	3	1	0	0	0	0
	4	0,9327	0,0673	0,1153	0	0
0,90	1	1	0	0	0	0
	2	1	0	0	0	0
	3	1	0	0	0	0
	4	0,9417	0,0582	0,1074	0	0
0,95	1	1	0	0	0	0
	2	1	0	0	0	0
	3	1	0	0	0	0
	4	0,9515	0,0485	0,102	0	0

Çizelge 5.3. $n=60$, $X \sim VM(3,20)$ olduğunda elde edilen sonuçlar

γ	aykırı gözlem	DTO	M	Sd(M)	S	Sd(s)
0,70	1	1	0	0	0,013	0,0117
	2	1	0	0	0,0003	0,0024
	3	1	0	0	0	0
	4	0,9997	0,0003	0,0079	0	0
0,75	1	1	0	0	0,0121	0,0119
	2	1	0	0	0,0001	0,0017
	3	1	0	0	0	0
	4	1	0	0	0	0
0,80	1	1	0	0	0,0119	0,0122
	2	1	0	0	0,0001	0,0015
	3	1	0	0	0	0
	4	1	0	0	0	0
0,85	1	1	0	0	0,0107	0,0112
	2	1	0	0	0,0001	0,0016
	3	1	0	0	0	0
	4	1	0	0	0	0
0,90	1	1	0	0	0,0105	0,0113
	2	1	0	0	0,00008	0,0012
	3	1	0	0	0	0
	4	1	0	0	0	0
0,95	1	1	0	0	0,0096	0,0111
	2	1	0	0	0,000003	0,0007
	3	1	0	0	0	0
	4	1	0	0	0	0

Çizelge 5.4. $n=20$, $X \sim VM(3,10)$ olduğunda elde edilen sonuçlar

γ	aykırı gözlem	DTO	M	Sd(M)	S	Sd(s)
0,70	1	1	0	0	0,0001	0,0028
	2	0,9935	0,0065	0,0566	0	0
	3	0,8293	0,1707	0,1764	0	0
	4	0,584	0,416	0,1672	0	0
0,75	1	1	0	0	0,0002	0,0037
	2	0,9965	0,0035	0,0417	0	0
	3	0,858	0,142	0,176	0	0
	4	0,5912	0,4088	0,1639	0	0
0,80	1	1	0	0	0	0
	2	0,998	0,002	0,0315	0	0
	3	0,8743	0,1256	0,167	0	0
	4	0,6117	0,3883	0,1638	0	0
0,85	1	1	0	0	0	0
	2	0,999	0,001	0,0223	0	0
	3	0,889	0,111	0,1613	0	0
	4	0,6257	0,3742	0,1718	0	0
0,90	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,9083	0,0916	0,1496	0	0
	4	0,638	0,362	0,1728	0	0
0,95	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,9046	0,0954	0,1521	0	0
	4	0,639	0,361	0,1752	0	0

Çizelge 5.5. $n=30$, $X \sim VM(3,10)$ olduğunda elde edilen sonuçlar

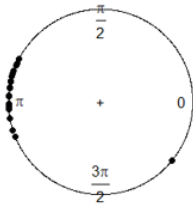
γ	aykırı gözlem	DTO	M	Sd(M)	S	Sd(s)
0,70	1	1	0	0	0,0016	0,0074
	2	0,9995	0,0005	0,0158	0	0
	3	0,9523	0,0477	0,1176	0	0
	4	0,7692	0,2308	0,1548	0	0
0,75	1	1	0	0	0,001	0,0058
	2	1	0	0	0,00003	0,0011
	3	0,972	0,028	0,0925	0	0
	4	0,7872	0,2128	0,1539	0	0
0,80	1	1	0	0	0,0006	0,0045
	2	1	0	0	0	0
	3	0,9733	0,0267	0,0904	0	0
	4	0,8055	0,1945	0,1567	0	0
0,85	1	1	0	0	0,0005	0,0041
	2	1	0	0	0	0
	3	0,9793	0,0207	0,0804	0	0
	4	0,8382	0,1618	0,1476	0	0
0,90	1	1	0	0	0,0003	0,0034
	2	1	0	0	0	0
	3	0,9933	0,0066	0,0466	0	0
	4	0,8435	0,1565	0,1467	0	0
0,95	1	1	0	0	0	0
	2	1	0	0	0	0
	3	0,9936	0,0064	0,0455	0	0
	4	0,8546	0,1355	0,1411	0	0

Çizelge 5.6. $n=60$, $X \sim VM(3,10)$ olduğunda elde edilen sonuçlar

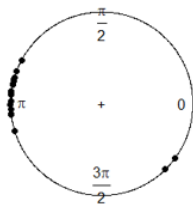
γ	aykırı gözlem	DTO	M	Sd(M)	S	Sd(s)
0,70	1	1	0	0	0,0157	0,03
	2	1	0	0	0,0021	0,0058
	3	0,9986	0,0014	0,021	0,00003	0,0007
	4	0,99	0,01	0,049	0	0
0,75	1	1	0	0	0,014	0,0126
	2	1	0	0	0,002	0,0057
	3	1	0	0	0,00007	0,0011
	4	0,9965	0,0035	0,0293	0	0
0,80	1	1	0	0	0,0137	0,0127
	2	1	0	0	0,0012	0,0046
	3	1	0	0	0,00001	0,0005
	4	0,9982	0,0018	0,0208	0	0
0,85	1	1	0	0	0,0141	0,013
	2	1	0	0	0,0008	0,0038
	3	1	0	0	0	0
	4	0,9995	0,0005	0,0111	0	0
0,90	1	1	0	0	0,0138	0,0125
	2	1	0	0	0,0009	0,0039
	3	1	0	0	0	0
	4	1	0	0	0	0
0,95	1	1	0	0	0,0131	0,012
	2	1	0	0	0,0004	0,0028
	3	1	0	0	0	0
	4	1	0	0	0	0

Aşağıdaki şekillerde, simülasyon çalışmasında kullanılan dağılımlar için farklı bozulma dereceleri ve bozulma oranlarında aykırı değerler görsel olarak verilmiştir.

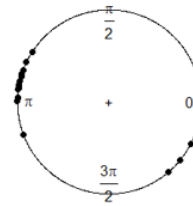
0.70 bozulma derecesi, 1 aykırı değer



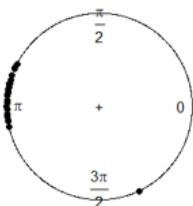
0.80 bozulma derecesi, 2 aykırı değer



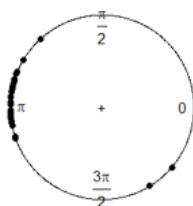
0.90 bozulma derecesi, 3 aykırı değer

Şekil 5.5. $\mu = 3, \kappa = 20$, $n=20$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi

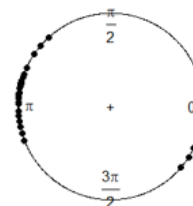
0.70 bozulma derecesi, 1 aykırı değer



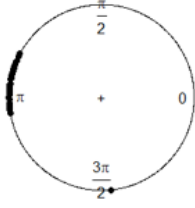
0.80 bozulma derecesi, 2 aykırı değer



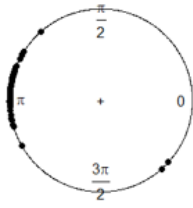
0.90 bozulma derecesi, 3 aykırı değer

Şekil 5.6. $\mu = 3, \kappa = 20$, $n=30$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi

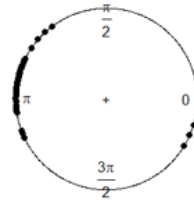
0.70 bozulma derecesi, 1 aykırı değer



0.80 bozulma derecesi, 2 aykırı değer

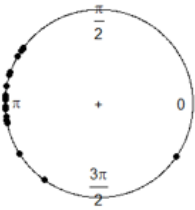


0.90 bozulma derecesi, 3 aykırı değer

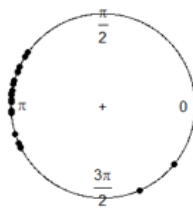


Şekil 5.7. $\mu = 3, \kappa = 20, n=60$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi

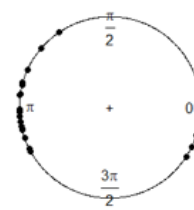
0.70 bozulma derecesi, 1 aykırı değer



0.80 bozulma derecesi, 2 aykırı değer

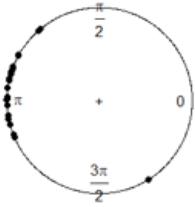


0.90 bozulma derecesi, 3 aykırı değer

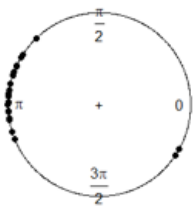


Şekil 5.8. $\mu = 3, \kappa = 10, n=20$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi

0.70 bozulma derecesi, 1 aykırı değer



0.80 bozulma derecesi, 2 aykırı değer

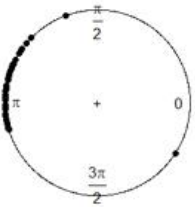


0.90 bozulma derecesi, 3 aykırı değer

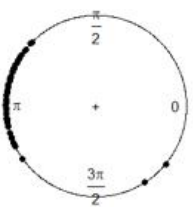


Şekil 5.9. $\mu = 3, \kappa = 10, n=30$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi

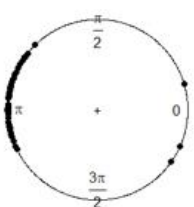
0.70 bozulma derecesi, 1 aykırı değer



0.80 bozulma derecesi, 2 aykırı değer



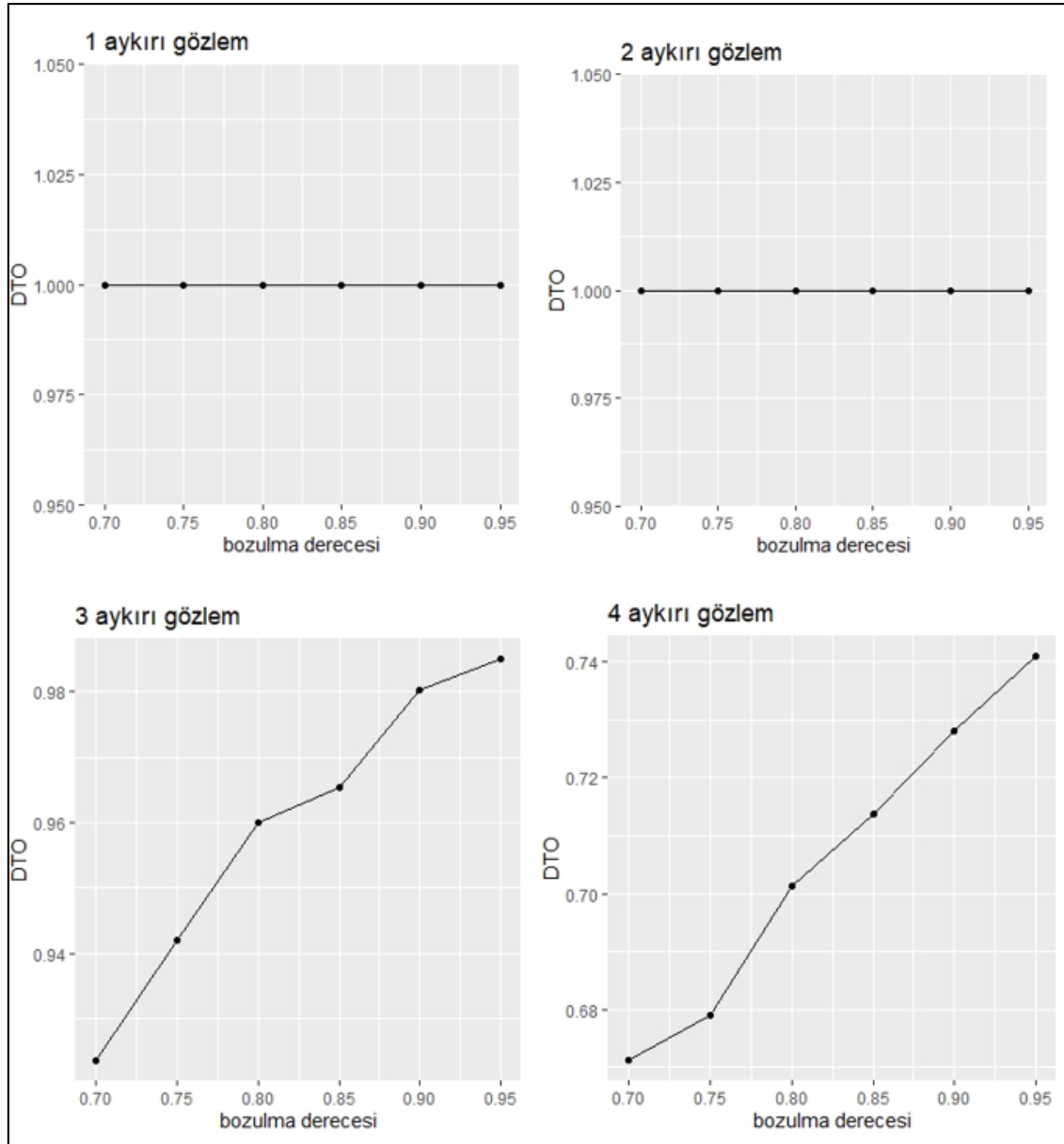
0.90 bozulma derecesi, 3 aykırı değer



Şekil 5.10. $\mu = 3, \kappa = 10, n=60$ ve farklı bozulma dereceleri için aykırı gözlemlerin dairesel düzlemde gösterimi

Yukarıdaki şekillerde, veri setinin bozulma derecesi arttıkça, üretilen aykırı gözlemin (gözlemlerin) veri setinden uzaklaştığı görülmektedir.

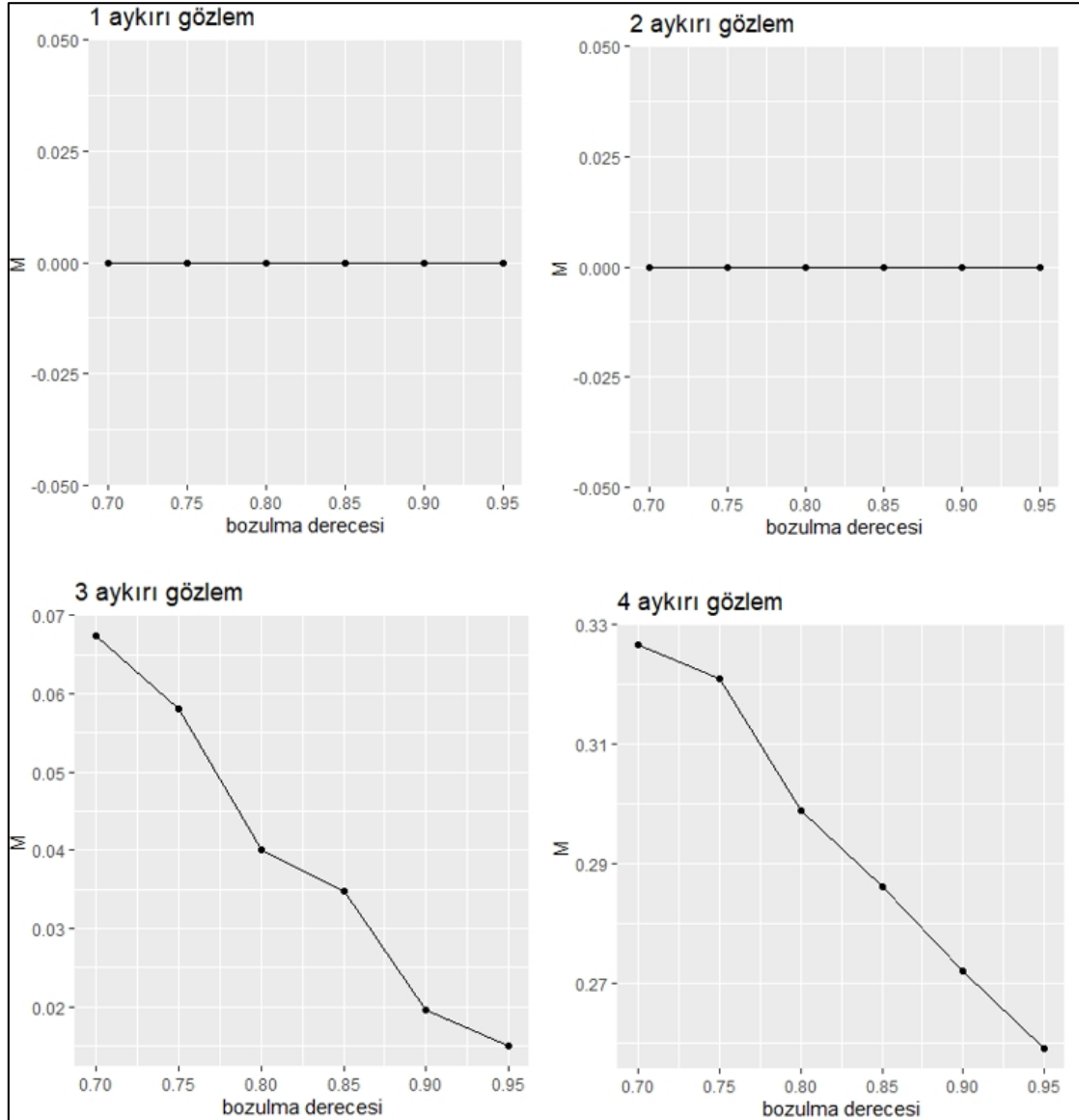
Yapılan simülasyon sonuçlarından elde edilen bulgular, aşağıdaki şekiller ile görsel olarak desteklenmiştir. Farklı örneklem boyutlarında ilk dört aykırı gözlem için doğru tespit oranları, maskeleye ve sahte aykırı gözlem oranları incelenmiştir. DTO doğru tespit oranı, M maskeleye, S sahte aykırı gözlem oranını ve Sd standart sapmayı temsil etmektedir.



Şekil 5.11. $\mu = 3, \kappa = 20$ ve $n=20$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları

$n=20$ iken önerilen yöntemin %70 bozulma derecesinde dahi iyi sonuçlar verdiği gözlenmiştir. İlk iki aykırı gözlem düşük bozulma oranlarında dahi başarılı sonuçlar

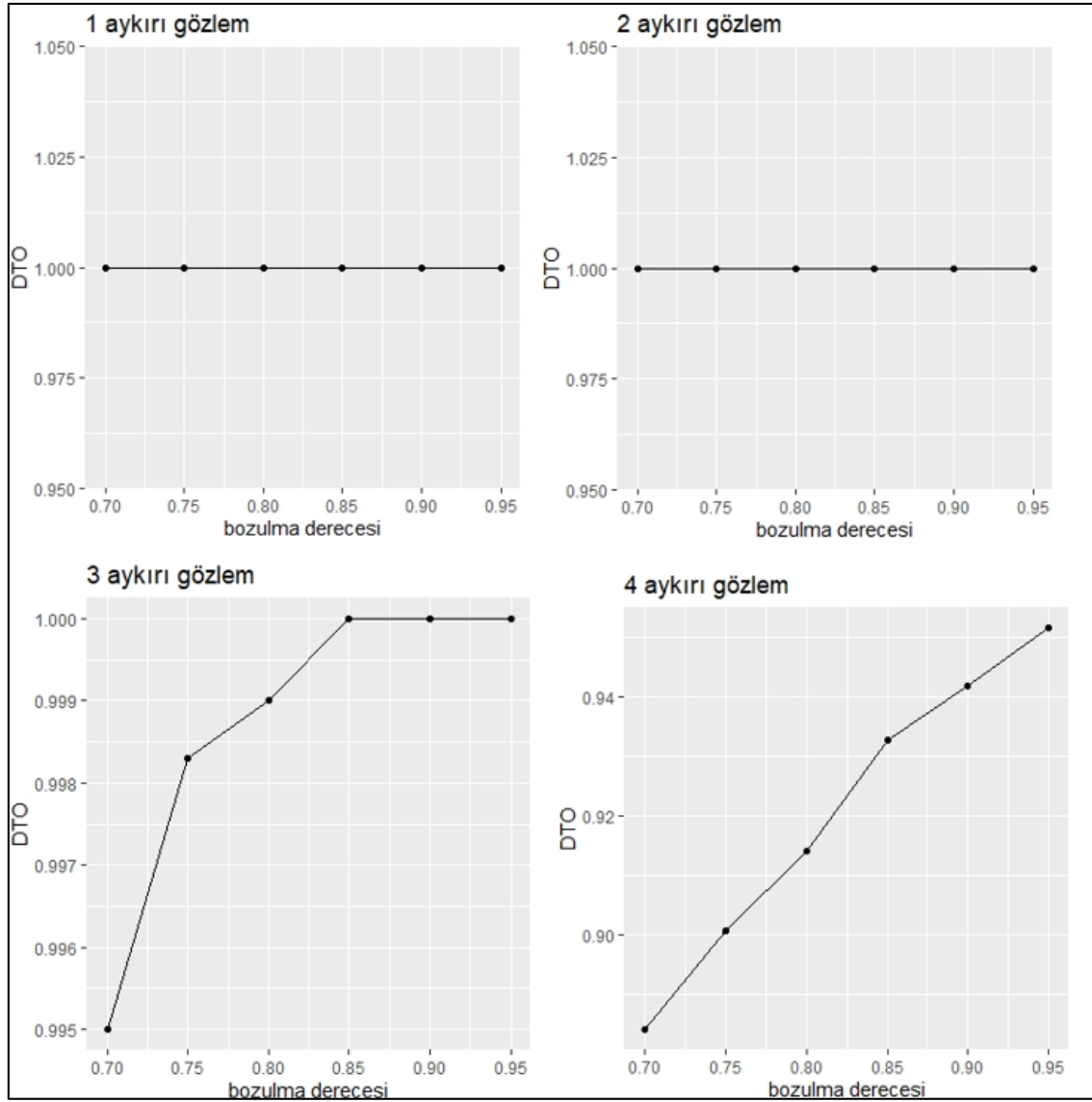
vermişken, veri setinin %10'unu aşacak şekilde aykırı gözlem eklendiğinde beklendiği üzere aykırı gözlem tespit oranı azalmıştır.



Şekil 5.12. $\mu = 3, \kappa = 20$ ve $n = 20$ olduğunda ilk dört aykırı gözlem için maskeleme oranları

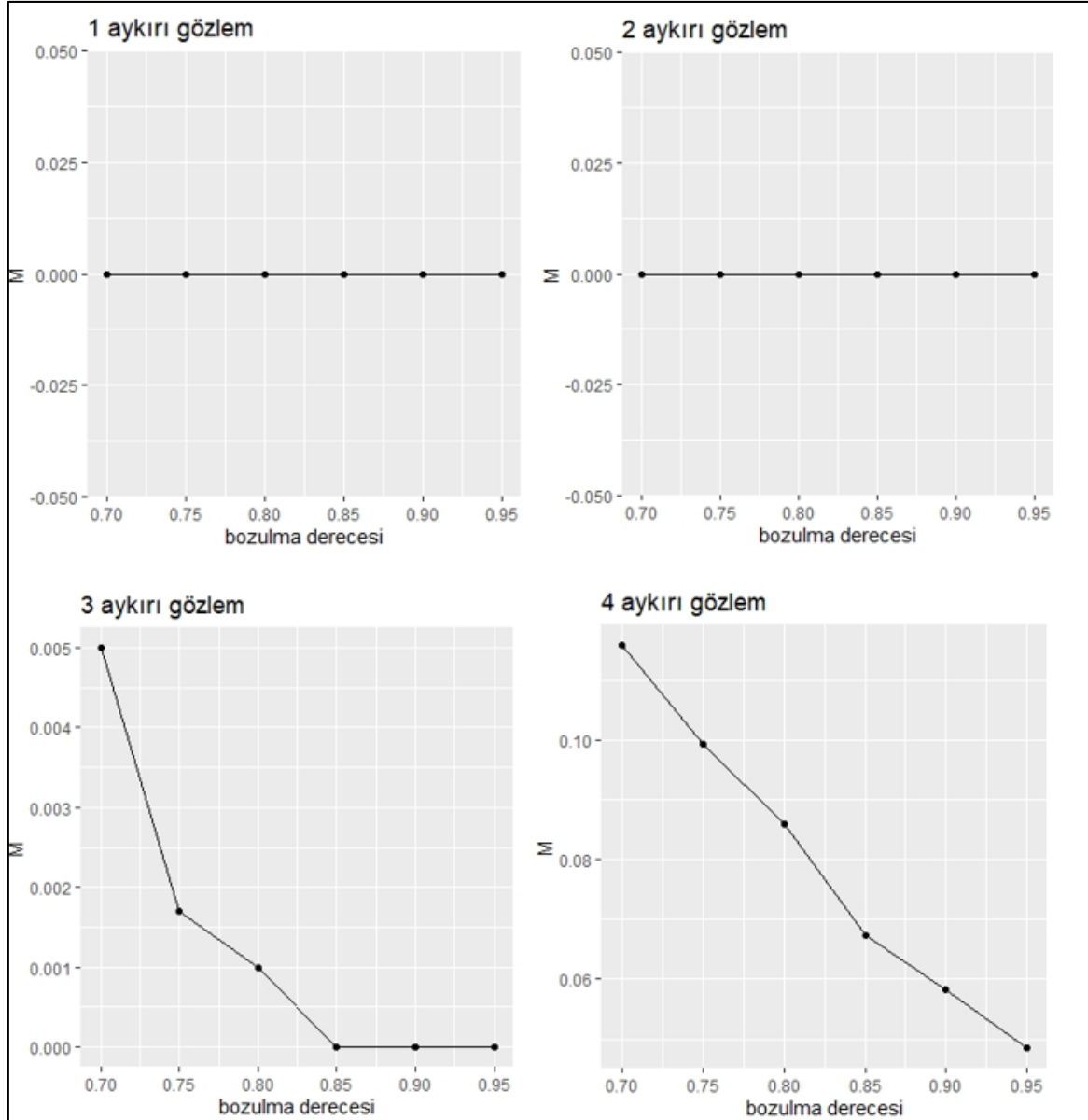
$n=20$ iken, ilk iki aykırı gözlem bütün bozulma derecelerinde tespit edildiğinden, maskeleme oranları sıfır olarak elde edilmiştir. Yani önerilen yöntemde bütün aykırı gözlemler tespit edilmiştir.

Yapılan simülasyonlar sonucunda $\mu = 3, \kappa = 20$ ve $n = 20$ olduğu durumda sahte aykırı gözleme rastlanmamıştır.



Şekil 5.13. $\mu = 3, \kappa = 20$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları

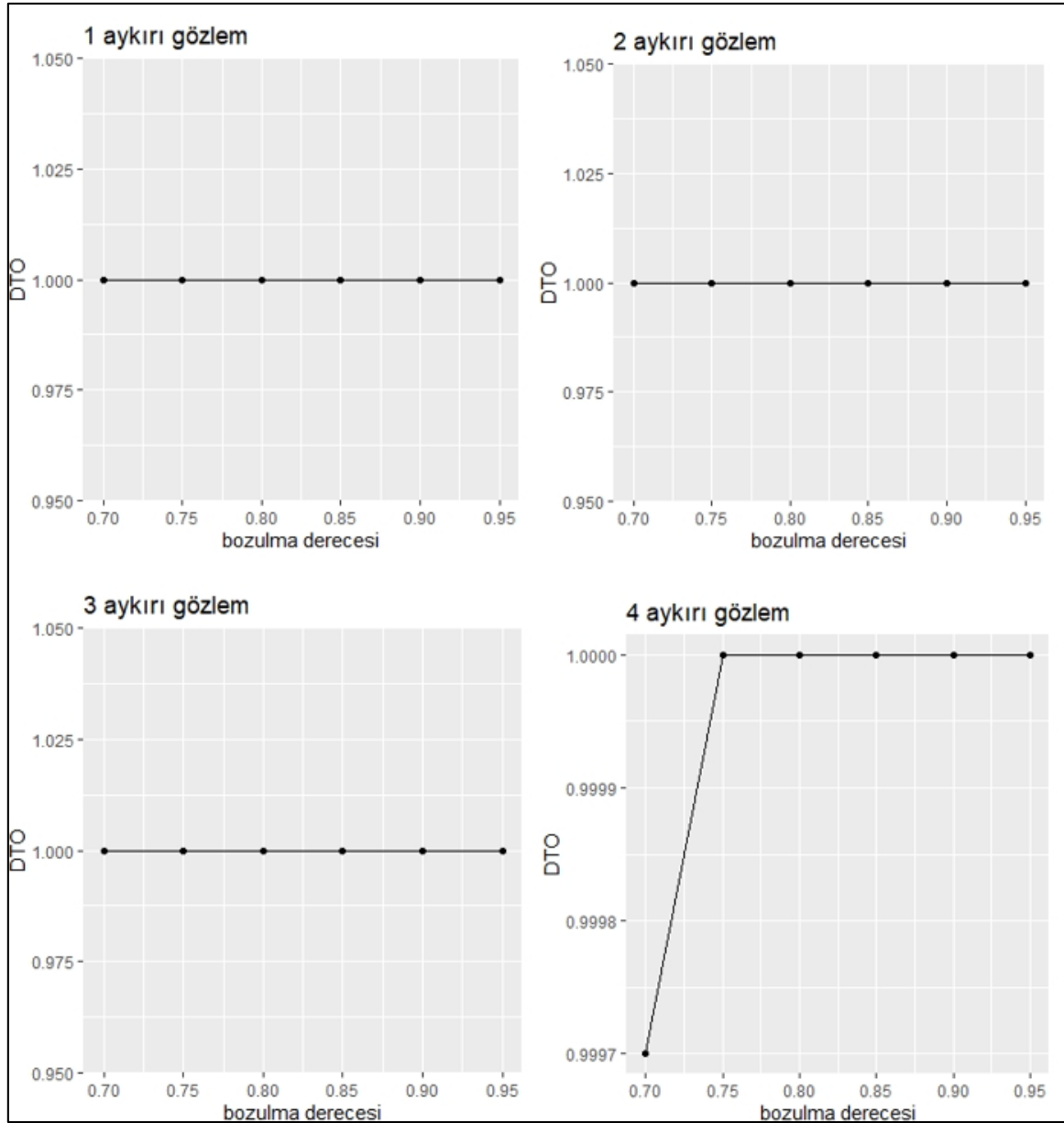
$n=30$ olduğunda ilk iki aykırı gözlem tamamen tespit edilmiş, 3. aykırı gözlem ise en düşük bozulma derecesi olan %70'de dahi %99 oranında doğru tespit oranına ulaşmıştır. Bozulma derecesi arttıkça doğru tespit oranının da arttığı gözlenmiştir.



Şekil 5.14. $\mu = 3, \kappa = 20$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için maskeleye oranları

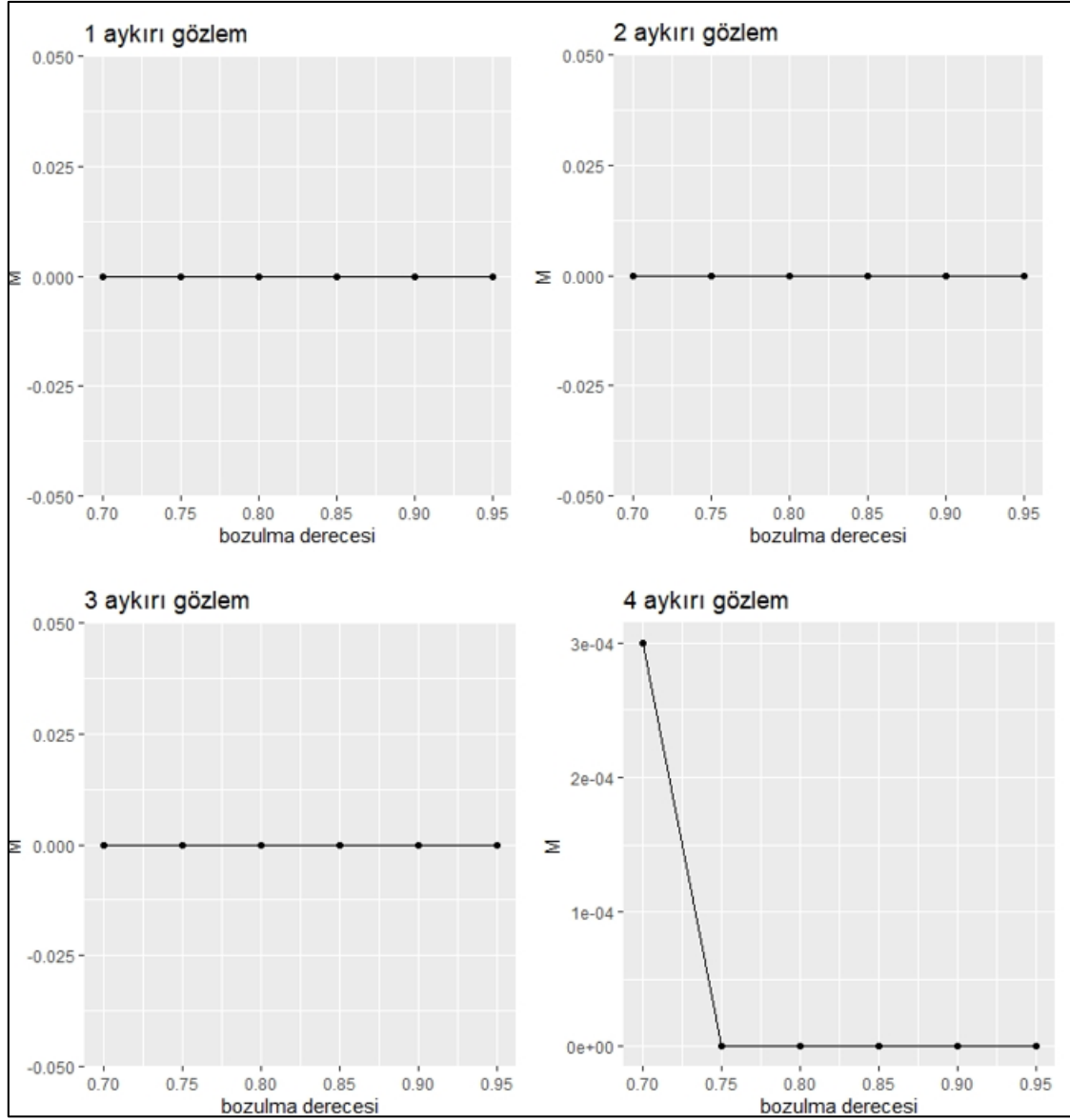
$n=30$ için önerilen yöntem düşük maskeleye oranına sahiptir. Veri setinin %10'unu açacak şekilde aykırı gözlem eklenmesine rağmen, doğru tespit oranının yüksek ve maskeleye oranının düşük olduğu gözlenmiştir.

Yapılan simülasyonlar sonucunda $\mu = 3, \kappa = 20$ ve $n = 30$ olduğu durumda sahte aykırı gözleme rastlanmamıştır.



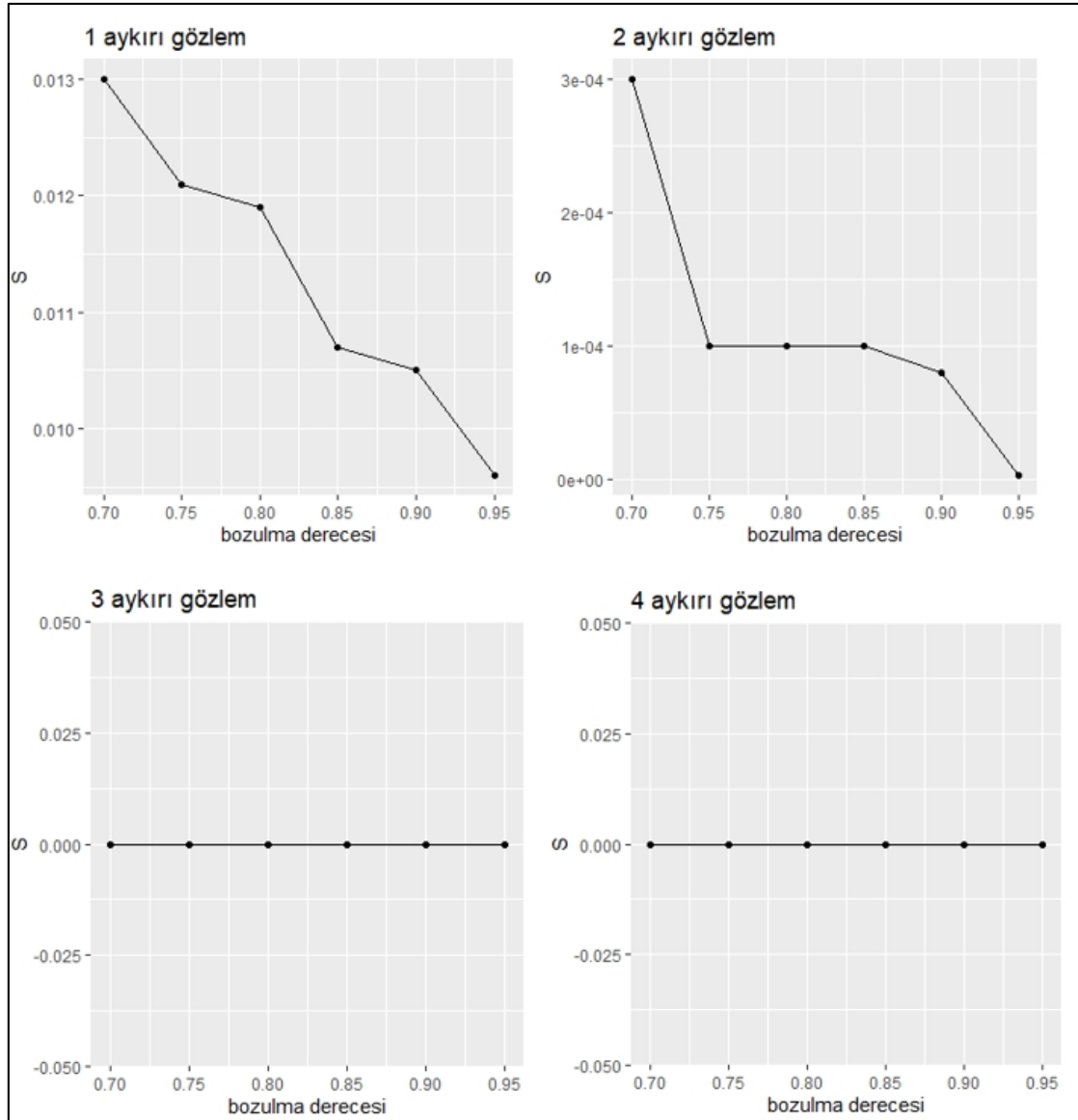
Şekil 5.15. $\mu = 3, \kappa = 20$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları

$n=60$ olduğunda önerilen yöntemin doğru tespit oranının yüksek olduğu görülmektedir. Dört aykırı gözlemde dahi, aykırı gözlem tespit oranı %70 bozulma derecesi ile %99.97 olarak elde edilmiştir. Örneklem boyutu arttıkça, doğru tespit oranının arttığı görülmektedir.



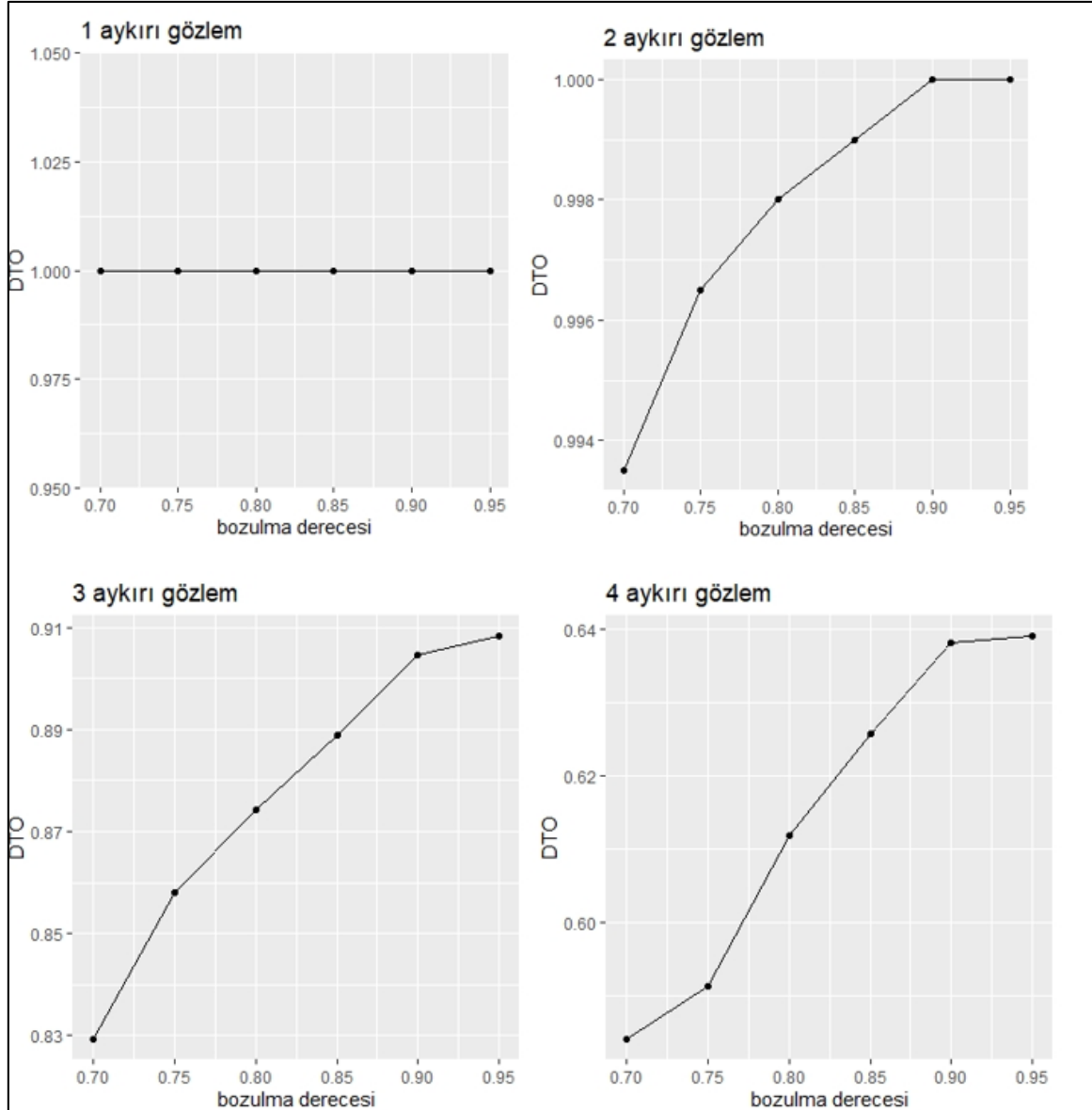
Şekil 5.16. $\mu = 3, \kappa = 20$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için maskeleme oranları

$n=60$ için maskeleme oranlarının, doğru tespit oranıyla uyumlu olacak şekilde sıfır ve sıfıra yakın olduğu görülmektedir.



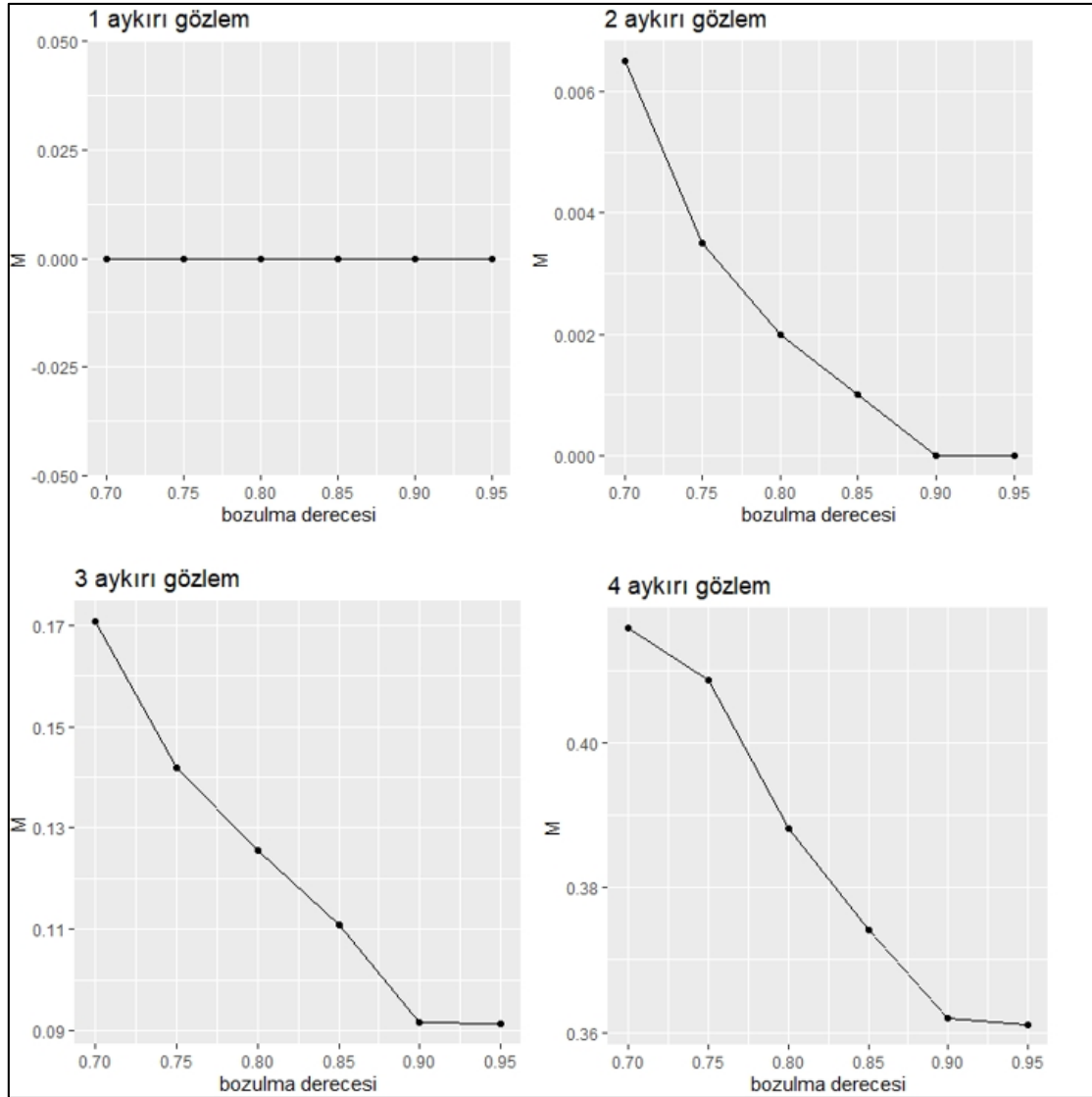
Şekil 5.17. $\mu = 3, \kappa = 20$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için sahte aykırı gözlem oranları

$n=60$ olduğunda, sahte aykırı gözlem gözlenmeye başlamıştır. Ancak ilk iki aykırı gözlem tespitinde sahte aykırı gözlem durumuna rastlanmıştır, 3. ve 4. aykırı gözlemler eklendiğinde sahte aykırı gözleme rastlanmamıştır. Bozulma derecesi arttıkça sahte aykırı gözlem oranının azaldığı görülmektedir.



Şekil 5.18. $\mu = 3, \kappa = 10$ ve $n = 20$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları

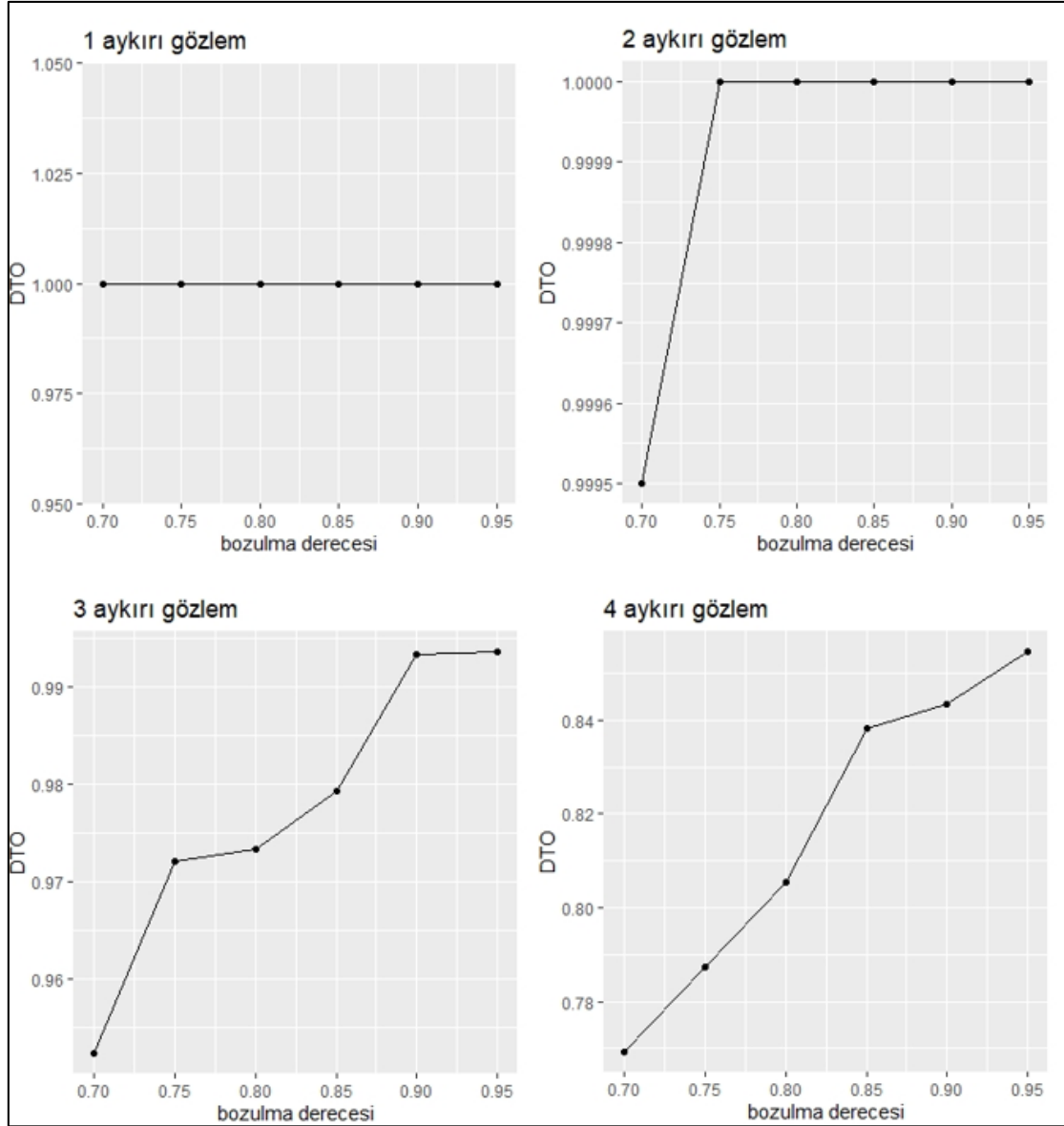
Veri setinin yayılımı arttıkça aykırı gözlem tespitinin güçleşmesi beklenen bir durumdur. Ancak iki aykırı gözlem durumu ele alındığında, en düşük bozulma derecesinde dahi doğru tespit oranının %99.4 olması, önerilen yöntemin performansını destekler niteliktedir. Veri setindeki aykırı gözlem oranı %10'dan fazla olduğunda, aykırı gözlem tespit oranında beklendiği üzere düşüş yaşanmıştır.



Şekil 5.19. $\mu = 3, \kappa = 10$ ve $n = 20$ olduğunda ilk dört aykırı gözlem için maskeleme oranları

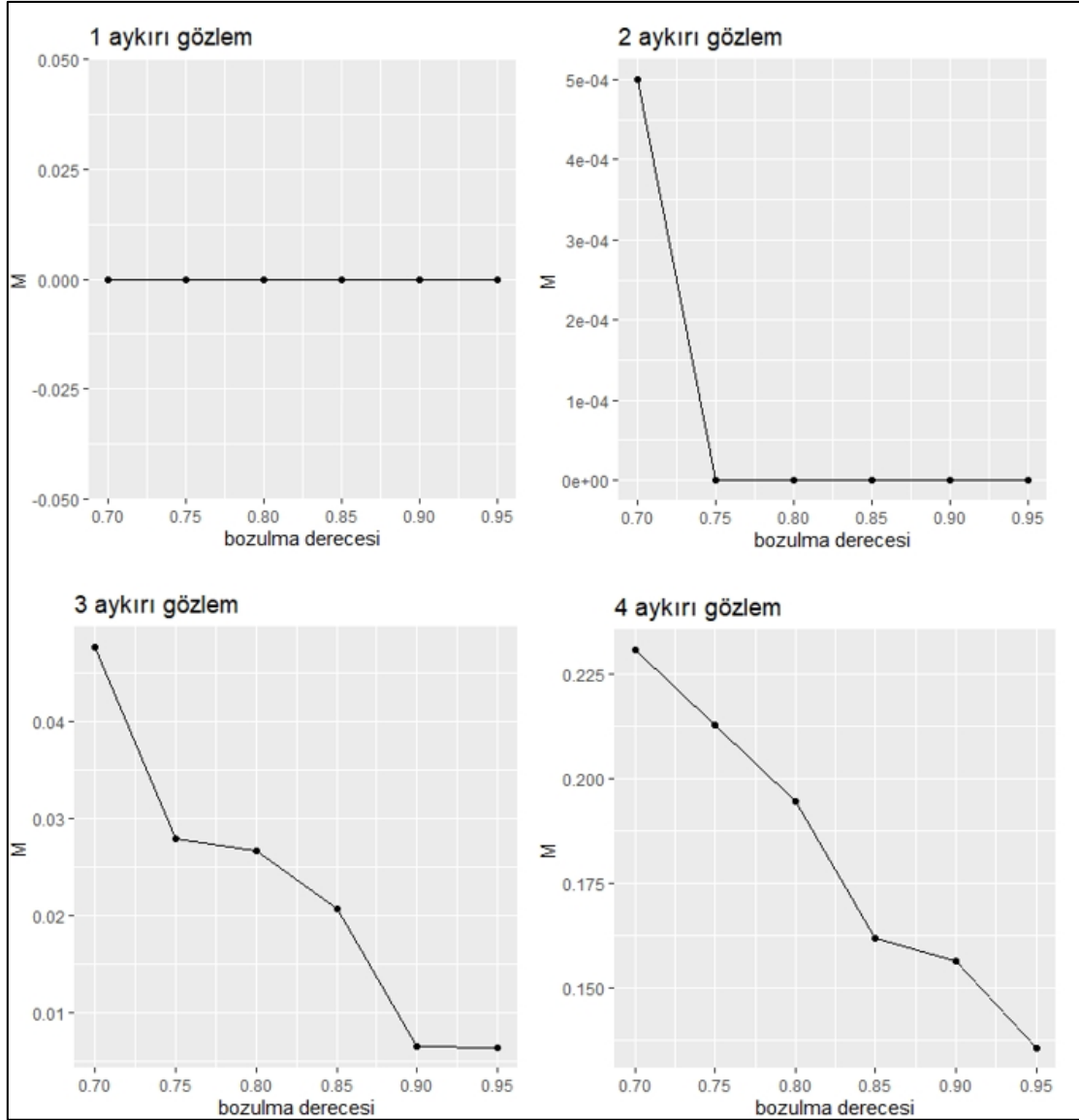
$n=20$ olduğunda maskeleme ile ilgili elde edilen sonuçlar, doğru tespit oranı ile uyumludur. İlk iki aykırı gözlem için maskeleme oranları sıfıra yakınken, aykırı gözlem sayısı veri setinin %10'unu aştığında maskeleme oranlarında artış gözlenmiştir. Veri setinin bozulma derecesi arttıkça maskeleme oranının azaldığı açıkça görülmektedir.

Yapılan simülasyonlar sonucunda $\mu = 3, \kappa = 10$ ve $n = 20$ olduğu durumda sahte aykırı gözleme rastlanmamıştır.



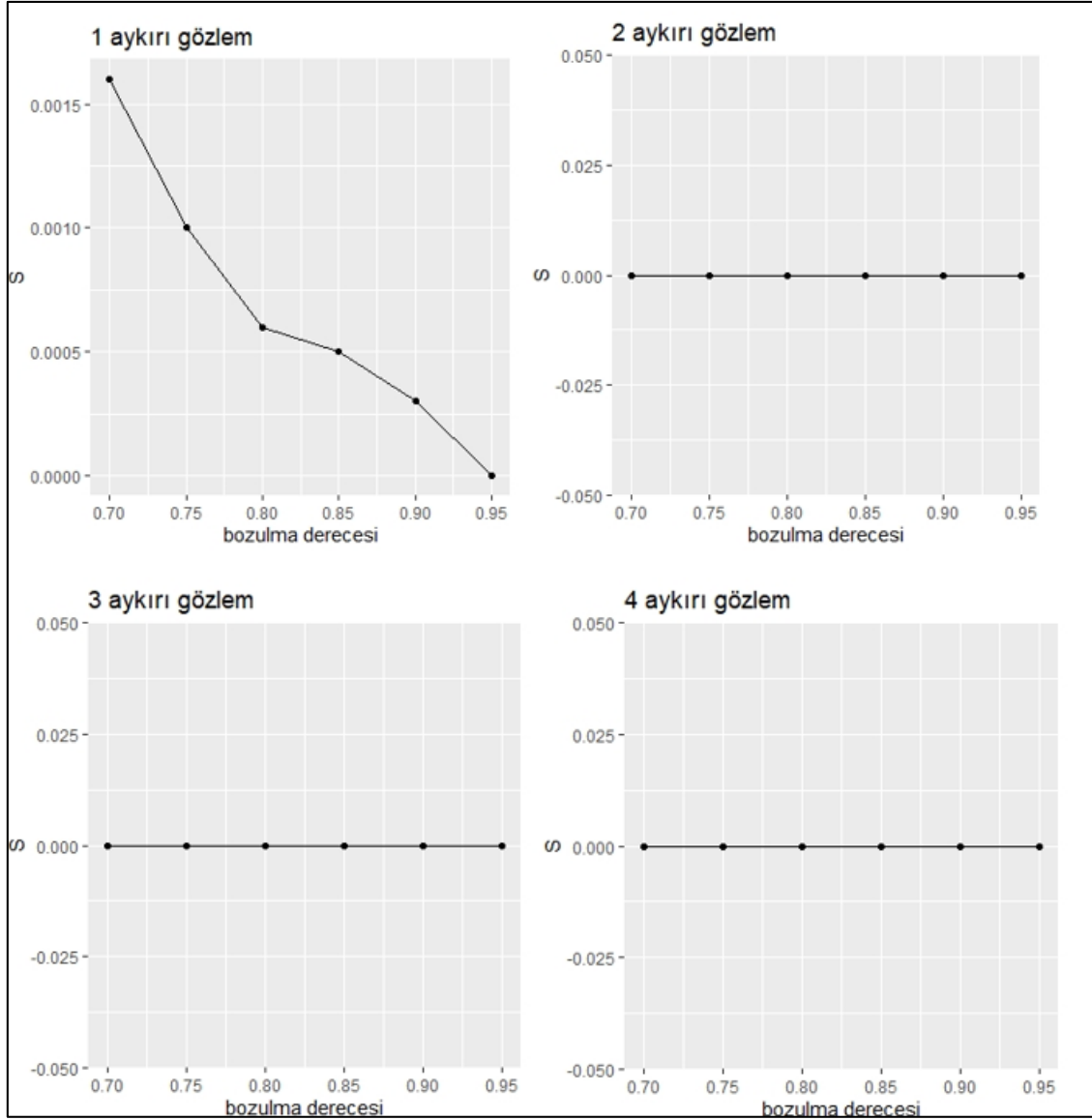
Şekil 5.20. $\mu = 3, \kappa = 10$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları

$n=30$ olduğunda, bozulma derecesi arttıkça doğru tespit oranının arttığı gözlenmiştir. Birden fazla aykırı gözlem olduğunda ve bozulma derecesi %75 ve üstü değerler aldığında, doğru tespit oranının %97'nin üzerinde olduğu görülmektedir. Bu durum, veri setinin %10'u kadar aykırı gözlem olduğu durumlar için geçerlidir.



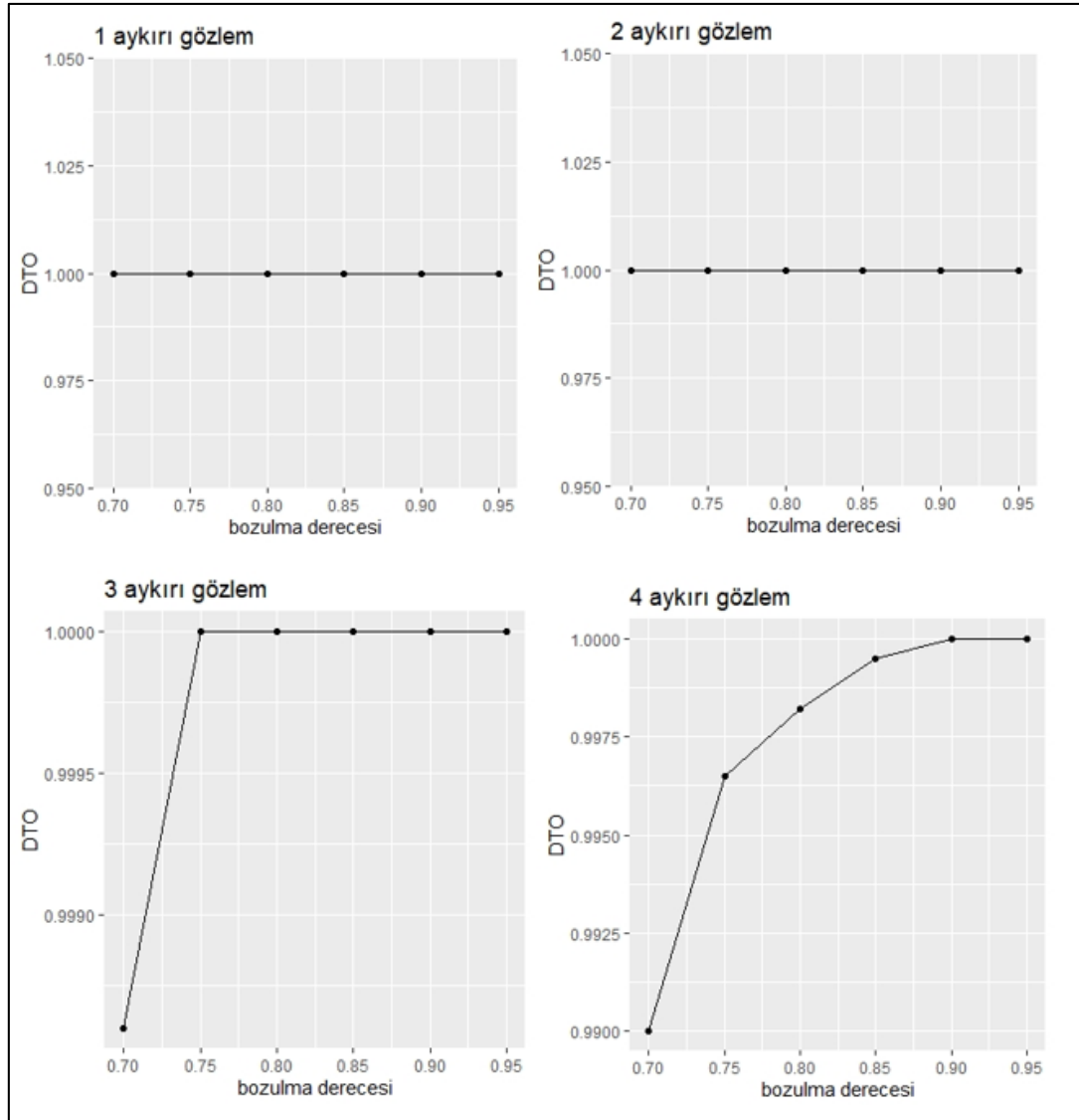
Şekil 5.21. $\mu = 3, \kappa = 10$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için maskeleye oranları

$n=30$ olduğunda maskeleye ile ilgili elde edilen sonuçlar, doğru tespit oranı ile uyumludur. İlk üç aykırı gözlem için maskeleye oranları sıfıra yakinken, aykırı gözlem sayısı veri setinin %10'unu aştığında maskeleye oranlarında artış gözlenmiştir. Veri setinin bozulma derecesi arttıkça maskeleye oranının azaldığı açıkça görülmektedir. Aykırı gözlem sayısı arttıkça maskeleye oranı da artmaktadır.



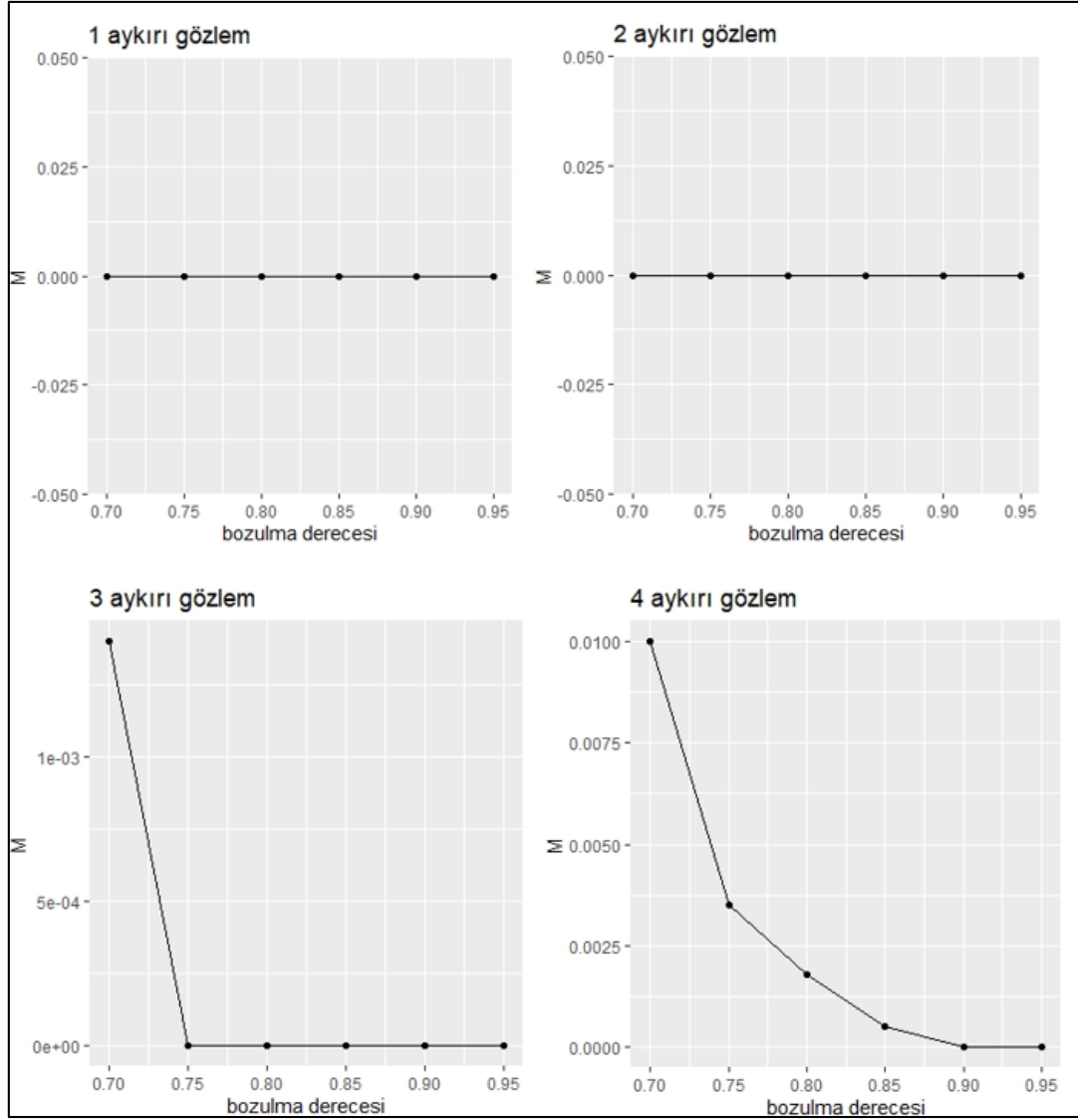
Şekil 5.22. $\mu = 3, \kappa = 10$ ve $n = 30$ olduğunda ilk dört aykırı gözlem için sahte aykırı gözlem oranları

$n=30$ olduğunda, sadece bir aykırı gözlem tespitinde sahte aykırı gözlem durumuyla karşılaşmıştır. Veri setinin bozulma derecesi arttıkça sahte aykırı gözlem oranı sıfıra yaklaşmıştır. Veri setinin yayılımı arttıkça küçük örneklerde dahi sahte aykırı gözlem durumuyla karşılaşılabileceği gözlenmiştir.



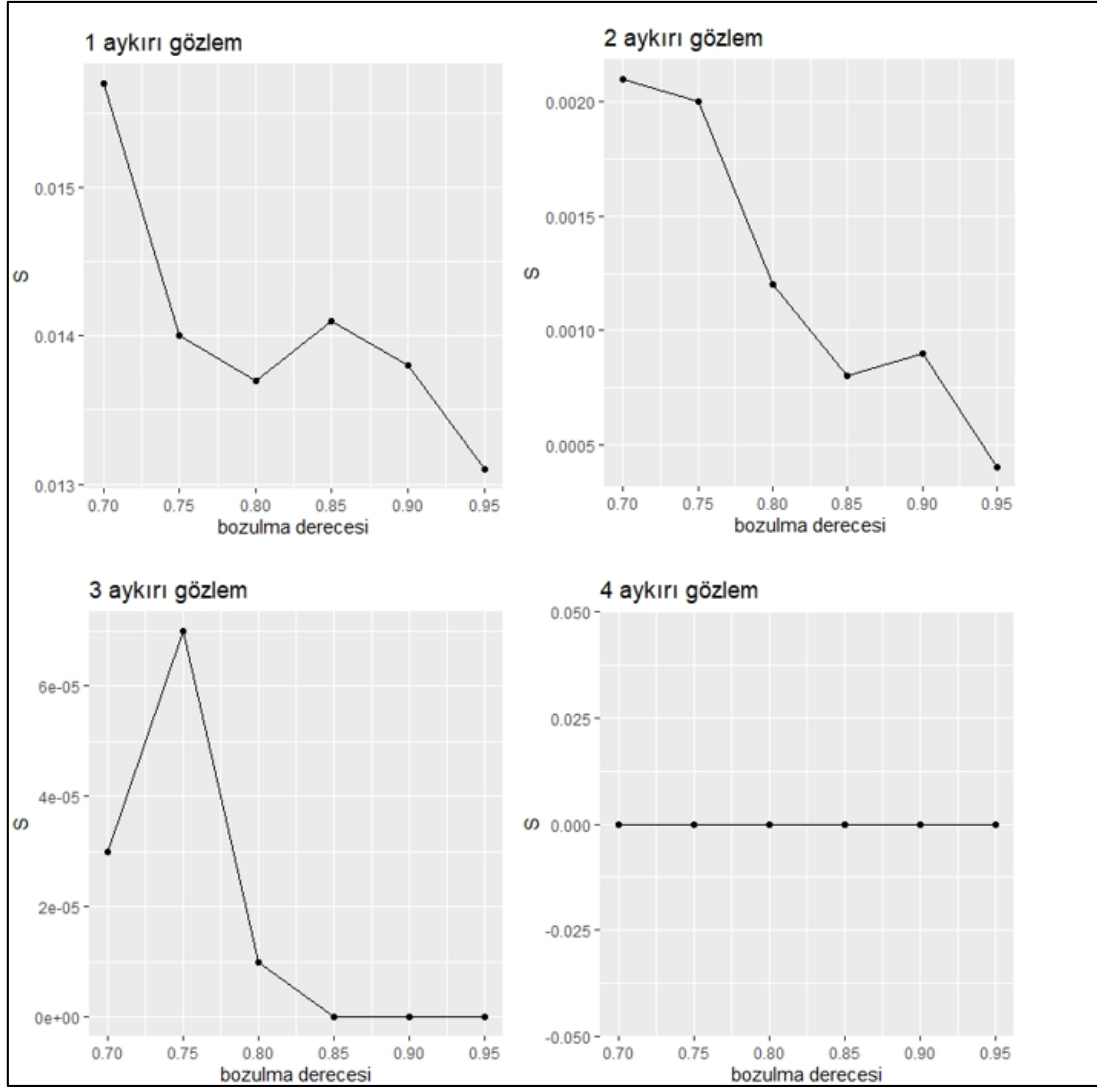
Şekil 5.23. $\mu = 3, \kappa = 10$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için doğru tespit oranları

$n=60$ olduğunda, doğru tespit oranının incelenen bütün durumlarda yüksek olduğu gözlenmiştir. $n=20$ ve $n=30$ 'a kıyasla, doğru tespit oranında artış gözlenmiştir.



Şekil 5.24. $\mu = 3, \kappa = 10$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için maskeleye oranları

$n=60$ olduğunda, doğru tespit oranıyla uyumlu olarak, düşük maskeleye oranları elde edilmiştir. Veri setinin bozulma oranı arttıkça maskeleye oranının azaldığı gözlenmiştir.



Şekil 5.25. $\mu = 3, \kappa = 10$ ve $n = 60$ olduğunda ilk dört aykırı gözlem için sahte aykırı gözlem oranları

$n=60$ olduğunda, sahte aykırı gözlem durumu ortaya çıkmaya başlamıştır. Şekil 5.25'ten de görüldüğü üzere, sahte aykırı gözlem varlığı herhangi bir düzen takip etmemektedir.

Yukarıdaki tablo ve şekillerden elde edilen sonuçlara göre, önerilen yöntemin yüksek doğru tespit, düşük maskeleyme ve sahte aykırı gözlem oranına sahip olduğu görülmektedir. Veri setinin bozulma derecesi arttıkça aykırı değer tespit oranı da artmaktadır. Veri setinin yayılımı aykırı gözlem tespitini etkileyen önemli parametrelerden biridir ve veri setinin çember etrafındaki yayılımı arttıkça aykırı değer tespit oranının düştüğü ancak buna rağmen önerilen yöntemin hala yüksek oranlarda aykırı değer tespit oranına sahip olduğu gözlenmiştir.

Örneklem boyutu küçük ve veri setinin yoğunluğu küçük iken sahte aykırı gözleme rastlanmamış, boyut ve veri setinin yayılımı arttıkça sahte aykırı gözlem durumu ortaya çıkmaya başlamıştır.

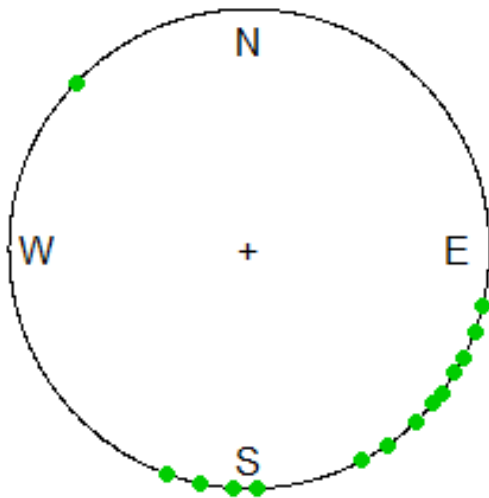
5.1.8. Gerçek veri seti uygulaması

Collett(1980) tarafından kullanılan Kurbağalar (Frogs) veri seti, önerilen yöntemi sınamak için kullanılmıştır. Ferguson, LanTOeth ve McKeown (1967), kuzeyli çekirge kurbağalarının yurda dönüş kabiliyetini araştırmak için bir dizi deney yapmıştır. Bu deneylerin birinde, Mississippi yakınlarında terkedilmiş bir akıntı menderesinin çamurundan bir miktar kurbağa toplanmıştır. Toplanan kurbağalar, toplandıkları noktanın kuzeybatısındaki bir odaya alınmıştır. 30 saat boyunca karanlık odada kalan kurbağalardan 14'ü serbest bırakılarak ve başlangıç referans noktası “Kuzey” alınarak gittikleri yönler derece cinsinden kaydedilmiştir. Elde edilen sonuçlar aşağıdaki çizelgede verildiği gibidir:

Çizelge 5.7. Kurbağalar veri seti

n_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
θ_i	104	110	117	121	127	130	136	145	152	178	184	192	200	316

Veri setinin dairesel düzlemde gösterimi aşağıdaki gibidir:



Şekil 5.26. Kurbağalar veri setinin dairesel düzlemde gösterimi

Başlangıç noktası kuzey ve hareket yönü saat yönü alınmıştır. Şekil (5.26)'da, 14. kurbağanın diğer kurbağalardan farklı davranış sergilediği açıkça görülmektedir ancak bu durum 14. gözlemin kesin olarak aykırı gözlem olduğunu belirtmemektedir. Önerilen yöntem uygulanmadan önce, veri setinin VM dağılımından gelip gelmediğini test etmek için “Watson’s Test” uygulanmış ve $\alpha = 0,05$ anlam düzeyinde aşağıdaki sonuç elde edilmiştir.

Watson's Test for the von Mises Distribution

Test Statistic: 0.0534

Level 0.05 Critical Value: 0.101

Do Not Reject Null Hypothesis

Watson’s testinin sonucuna göre, veri setinin VM dağılımından geldiği sonucuna varılmıştır. Tarafımızca önerilen yöntem ile bahsedilen diğer yöntemler bu veri setine uygulanmıştır. Önerilen yöntem kullanıldığında, diğer yöntemlerde olduğu gibi 14. gözlem aykırı gözlem olarak tespit edilmiştir.

Mahmood ve diğerleri (2017) çalışmasından yola çıkarak, frogs veri seti %20 oranında (3 aykırı gözlem olacak şekilde) kirletilmiştir. Veri setini kirletme işlemi eşitlik (5.18) kullanılarak yapılmıştır. Önerilen yöntem %20 bozulma oranı için uygulanıp diğer yöntemlerle karşılaştırıldığında elde edilen sonuçlar çizelge 5.8’de verilmiştir.

Çizelge 5.8. Kurbağa (frogs) veri seti için farklı istatistiklerin performanslarının karşılaştırılması

		Mardia	M	A	Chord	RCDu	Önerilen Yöntem
Orijinal veri seti	Aykırı gözlem tespit oranı	1	1	1	1	1	1
	Maskeleme oranı	0	0	0	0	0	0
	Sahte aykırı gözlem oranı	0	0	0	0	0	0
%20 bozulma oranı	Aykırı gözlem tespit oranı	0	0	0	0	1	1
	Maskeleme oranı	1	1	1	1	0	0
	Sahte aykırı gözlem oranı	0	0	0	0	0	0

Elde edilen bulgulara göre önerilen yöntem, RCDu istatistiğinde olduğu gibi %20 bozulma oranında aykırı gözlemleri tespit edebilmiştir. RCDu istatistiğinde kesim noktaları, farklı κ parametreleri ve farklı örnek boyutları kullanılarak, veri setinin dağılımının bilindiği ve Von-Mises olduğu durumlar için elde edilmişti. Ancak kullanılacak veri setinin dağılımı bu varsayımları sağlamıyorsa, veri setinin dağılımı hakkında elimizde önsel bilgi yoksa, yukarıda söz edilen Mardia, M, A, Chord ve RCDu istatistiklerinin aykırı gözlem tespiti için kullanımı mümkün olmayacaktır. Önerilen yöntem ise, veri setiyle ilgili herhangi bir varsayım veya önsel bilgi gerekmeden kullanılabilen parametrik olmayan bir yöntem olup aykırı gözlem tespitinde başarılı sonuçlar vermektedir.

5.2. Dairesel Regresyonda Aykırı Gözlem Kavramı ve Tespiti için Önerilen Yöntemler

Dairesel regresyonda da bugüne kadar yapılan çalışmaların hepsi, hataların sıfır ortalamalı ve dağılımının bilindiği durum için yapılmıştır. Literatürde parametrik olmayan dairesel regresyonda aykırı gözlem tespiti için yapılan herhangi bir çalışma bulunmamaktadır. Önerilen yöntemler, bilinen dağılımın tahmin edilebilen parametrelerine göre çalışmakta ve bu nedenle parametrik olmayan modellerde doğru sonuçlar vermemektedir.

Bu bölümde öncelikle basit dairesel regresyonda aykırı gözlem tespiti için literatürde yer alan yöntemler incelenecek, daha sonra parametrik olmayan dairesel regresyonda aykırı gözlem tespiti için önerilen yöntemin algoritması verilecektir.

5.2.1. COVRATIO istatistiği

Basit dairesel regresyonun kovaryans matrisinin kullanımına dayalı bu yöntem, yanıt değişkenindeki aykırı gözlemlerin tespiti için Abuzaid ve diğerleri (2011) tarafından önerilmiştir. Bu istatistik, doğrusal regresyon modellerinde uygulanan COVRATIO istatistiğinin dairesel regresyona uyarlamasıdır. Mardia istatistiğinde olduğu gibi i . gözlemin veri setinden çıkarılması mantığıyla çalışmaktadır. COVRATIO istatistiği eşitlik (5.19)'da verilmiştir:

$$COVRATIO_{(-i)} = \frac{|COV_{(-i)}|}{|COV|} \quad (5.19)$$

Burada $|COV| = \frac{1}{\hat{\kappa}_{A\kappa}}$ olup, tüm gözlemleri içeren veri seti kullanılarak hesaplanır. $|COV_{(-i)}|$ ise, i . gözlem veri setinden çıkarıldığında elde edilen kovaryans matrisini ifade etmektedir.

Kesim noktası, farklı κ parametresi ve farklı n değerleri için Monte Carlo simülasyonu yardımıyla belirlenmiştir. Herhangi bir örneklem boyutu “ n ” için kesim noktasının $\left(\frac{16}{n}\right)$ 'e yakın sonuçlar verdiği gözlemlenmiştir. Eğer $|COVRATIO_{(-i)} - 1|$ değeri kesim noktasından büyük ise, i . gözlem aykırı gözlemdir sonucuna varılmıştır.

5.2.2. DMCE_C istatistiği

Abuzaid (2010) tarafından önerilen bu istatistik, ortalama dairesel hataya alternatif bir ölçü olarak kosinüs fonksiyonlarını kullanmaktadır. Kosinüs fonksiyonu yardımıyla elde edilen ortalama hata istatistiği eşitlik (5.20)'de tanımlanmıştır.

$$MCE_c = 1 - \frac{1}{n} \sum \cos(y_i - \hat{y}_i) \quad (5.20)$$

Kosinüs fonksiyonunun tanım aralığının $[-1,1]$ olduğu göz önüne alındığında, MCE_c istatistiğinin $[0,2]$ aralığında tanımlı olduğu açıkça görülmektedir. Eğer y_i gözlemi aykırı gözlem ise, y_i ile $\hat{y}_i$ arasındaki dairesel uzaklığın büyük olması ve dolayısıyla MCE_c değerinin artması beklenir. y_i aykırı gözleminin çıkarılması, MCE_c istatistiğinin değerini düşürecektir. Aykırı gözlem tanımlamak için geliştirilen istatistik eşitlik (5.21)'de verilmiştir.

$$DMCE_{c(i)} = |MCE_c - MCE_{c(-i)}| \quad (5.21)$$

Burada $MCE_{c(-i)}$, i . gözlem veri setinden çıkarıldığında elde edilen istatistiğin değerini vermektedir. Aykırı gözlem tespiti için gerekli olan kesim noktası ise eşitlik (5.22) yardımıyla belirlenmektedir.

$$DMCE_c = maks|MCE_c - MCE_{c(-i)}| \quad (5.22)$$

Eğer $DMCE_{c(i)}$ gözlemi kesim noktasından büyükse, i . gözlem aykırı gözlem olarak adlandırılır. Kesim noktasını belirlemek için farklı κ parametreleri, n değerleri ve yüzdelikler kullanılarak MC simülasyon çalışması yapılmıştır.

5.2.3. DMCE_s istatistiği

Abuzaid ve diğerleri (2010), DMCE_c istatistiğinde olduğu gibi, $[0, \frac{\pi}{2}]$ aralığında artan bir fonksiyon olan sinüs fonksiyonu kullanılarak ortalama dairesel hataya alternatif bir ölçü geliştirmiştir. Sinüs fonksiyonu yardımıyla önerilen ortalama hata istatistiği eşitlik (5.23)'te tanımlanmıştır.

$$MCE_s = \frac{1}{n} \sum \sin\left(\frac{d_i}{2}\right) \quad (5.23)$$

Burada dairesel uzaklık olarak eşitlik (5.24) kullanılmaktadır.

$$d_i = \pi - |\pi - |y_i - \hat{y}_i|| \quad (5.24)$$

Sinüs fonksiyonunun tanım aralığı göz önüne alındığında, $MCEs$ istatistiğinin $[0,1]$ aralığında tanımlı olduğu açıkça görülmektedir. $MCEc$ istatistiğinde de olduğu gibi, aykırı gözlem varlığında $MCEs$ istatistiğinin değerinin artması beklenmektedir. Bu nedenle, $MCEc$ istatistiğine benzer olarak, i . gözlemin veri setinden çıkarılmasına dayalı olarak elde edilen aykırı gözlem tespiti için kullanılacak istatistik eşitlik (5.25)'te verilmiştir.

$$DMCEs_i = |MCEs - MCEs_{(-i)}| \quad (5.25)$$

Burada $MCEs_{(-i)}$, i . gözlem veri setinden çıkarıldıktan sonra elde edilen ortalama hata istatistiğidir. Aykırı gözlem tespiti için kullanılacak kesim noktası ise eşitlik (5.26)'da tanımlanmıştır.

$$cut DMCEs = maks|MCEs - MCEs_{(-i)}| \quad (5.26)$$

Eşitlik (5.26)'da tanımlanan kesim noktasından büyük değer alan gözlem, aykırı gözlem olarak adlandırılmaktadır.

5.2.4. RCD_y istatistiği

Dairesel ortalamalarından uzakta bulunan bir dairesel gözlem aykırı gözlem olarak adlandırılabilir ancak dairesel ortalama da aritmetik ortalama gibi aykırı gözlem varlığında sağlam bir davranış sergilememektedir. Dairesel medyanın ise, dairesel ortalamaya göre daha sağlam bir davranış sergilediği bilinmektedir. Bu nedenle Mahmood ve diğerleri (2017) tarafından önerilen RCD_y ("Robust Circular Distance") istatistiği, dairesel artıkların dairesel medyana olan uzaklığını kullanarak aykırı gözlem tespiti yapmaktadır. Bu yöntemde Jammalamadaka ve SenGupta (2001) tarafından önerilen ve eşitlik (5.1)'de tanımlanan dairesel uzaklık kullanılmaktadır. RCD_y istatistiğinin hesaplanması için ilk olarak eşitlik (5.27)'de verilen tahmin edilen dairesel artıkların uzaklığı, daha sonra eşitlik (5.28) ile verilen tahmin edilen dairesel artıkların uzaklığı $\hat{c}r_i$ ile dairesel medyan arasındaki uzaklık hesaplanır.

$$\hat{c}r(\hat{c}r_i = \pi - |\pi - |y_i - \hat{y}_i||) \quad (5.27)$$

$$\hat{c}r[RCD_i]_y = \pi - |\pi - |\hat{c}r - med(\hat{c}r)|| \quad (5.28)$$

Eğer y_i gözlemi aykırı gözlem ise, $\hat{c}r$ 'nın değerinin artacağı, ve dolayısıyla $\hat{c}r_i$ ile $\hat{c}r$ 'nın nispeten büyük değerler alacağı açıktır. Bu nedenle, kesim noktası eşitlik (5.29) ile tanımlanmaktadır.

$$cut\ RCD_y = maks[RCD]_y \quad (5.29)$$

Kesim noktasından büyük herhangi bir $[RCD_i]_y$ gözlemi, aykırı gözlem olarak tespit edilmektedir. Burada kesim noktaları yine VM dağılımı varsayımı altında farklı parametreler için Monte Carlo simülasyonu yapılarak elde edilmiştir.

5.2.5. Önerilen istatistik

Parametrik olmayan dairesel regresyonda aykırı gözlem tespiti için, bilgimiz dahilinde, bu güne kadar için yapılmış çalışma bulunmamaktadır. Bu bölümde, herhangi bir varsayım (dağılım vb.) veya veri setinin fonksiyonel şekli ile ilgili önsel herhangi bir bilgi olmadığında aykırı gözlem tespitine yönelik bir yöntem önerilmiştir. Yöntemin geçerliliğinin sınanması için simülasyon çalışmaları yapılmış, ardından gerçek veri setleriyle yöntemin işlevselliği ispatlanmıştır. Önerilen yöntemde amaç, mutlak artıklardan yararlanarak aykırı gözlem tespiti yapabilmektir. Bu amaçla, ortalamaya göre daha dayanıklı (robust) olan dairesel medyan kullanılmıştır. Mutlak artıkların dairesel medyanlarına olan uzaklıkları hesaplanmış ve eşik değeri tespiti için “bootstrap” yönteminden yararlanılmıştır. Bootstrap yöntemi ile mutlak artıklar için veri setinin boyutunda (n) örneklemeler üretilmiş ve her bir örneğin dairesel medyanı hesaplanmıştır. Ardından üretilen n boyutlu her bir örneklemin dairesel medyanı ile mutlak farkı hesaplanmış, hesaplanan bu farkların %95. yüzdelikleri elde edilmiştir. Eşik değeri, elde edilen yüzdeliklerin ortalaması olarak belirlenmiştir. Herhangi bir gözlemin mutlak artığının dairesel medyana olan uzaklığı eğer kesim noktasından büyükse, bu gözlem aykırı gözlem olarak adlandırılmaktadır. Önerilen algoritmanın adımları aşağıda verilmiştir.

- 1) Mutlak artıklar hesaplanır.

$$res = |y_i - \hat{y}_i|$$

- 2) Mutlak artıklar ile artıkların dairesel medyanının (med) mutlak farkı hesaplanır.

$$mf = |res - med|$$

- 3) Bootstrap yöntemi ile, mutlak artıklar için yeniden örnekleme yaparak “n” boyutlu res_B örneklemleri üretilir.

- 4) Üretilen res_B örneklemlerinin her birinin dairesel medyanı med_B hesaplanır.

- 5) res_B örneklemleri ile dairesel medyanları (med_B) arasındaki mutlak fark hesaplanır.

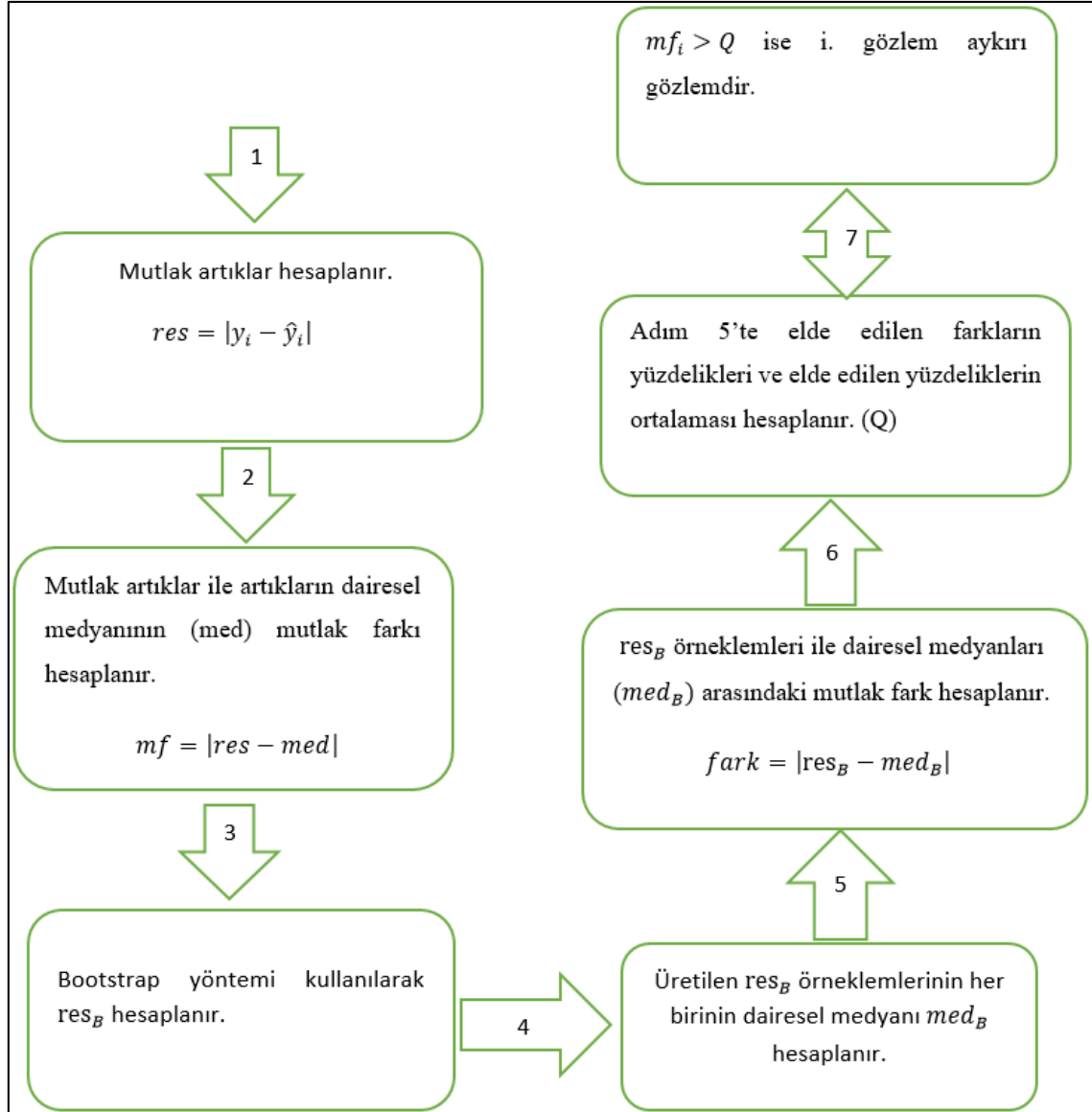
$$fark = |res_B - med_B|$$

- 6) Adım 5’te elde edilen farkların yüzdelikleri ve elde edilen yüzdeliklerin ortalaması hesaplanır. (Q)

- 7) Adım 6’da elde edilen Q değeri, aykırı gözlem tespiti için kullanılacak kesim noktasıdır. Eğer $mf_i > Q$ ise i. gözlemin aykırı gözlem olduğu tespit edilmiştir.

Yapılan simülasyonlar sonucunda maskeleye ve sahte aykırı gözlem oranları göz önüne alındığında, eşik değerinin belirlenmesinde %95. yüzdeliğin kullanılması yeterli görülmüştür.

Şekil 5.27’de, önerilen yöntemin akış diyagramı verilmiştir.



Şekil 5.27. Parametrik olmayan dairesel regresyon analizinde aykırı gözlem tespiti için önerilen yöntemin akış diyagramı

5.2.6. Simülasyon çalışması

Bu çalışmanın amacı, açıklayıcı değişken doğrusal ve yanıt değişkeni dairesel olduğunda ve veri setinde aykırı gözlem olduğunda parametrik olmayan dairesel regresyon yöntemlerinin davranışlarını incelemektir. Bu amaçla, Nadaraya Watson ve Yerel Lineer Kernel tahmin edicilerinin aykırı gözlem varlığında davranışları karşılaştırılmıştır. Önerilen yöntemin doğruluğunu araştırmak üzere, bir dizi simülasyon çalışması yapılmıştır. Bu çalışma yapılırken aşağıdaki adımlar takip edilmiştir:

- Açıklayıcı değişken olan X rasgele değişkeni, normal dağılımdan üretilmiştir.

$$X \sim N(3, 0.25)$$

- Dairesel yanıt değişkeni Y , aşağıdaki gibi üretilmiştir (Oliveira,2014):

$$y = \sin\left(1,5 \times \left(x - \frac{\pi}{2}\right)\right) + \left(2 \times \sqrt{2/3}\right) \times \cos\left(\frac{x}{3}\right) + \epsilon_i$$

olup $\epsilon_i \sim \text{WrappedCauchy}(0,1)$ 'dir.

- Uyum iyiliği ölçüsü olarak Di Marzio ve diğerleri (2012) tarafından önerilen risk fonksiyonu kullanılmıştır.
- Açıklayıcı değişkenimiz doğrusal olduğundan, ağırlıklandırma için "Gaussian Kernel" fonksiyonu kullanılmıştır.
- Parametrik olmayan regresyon yöntemlerinden NW ve LL tahmin edicileri kullanılmıştır.
- Bant genişlikleri "Çapraz Doğrulama" (Cross Validation) yöntemi ile aşağıdaki fonksiyonu minimize eden değer olarak belirlenmiştir:

$$\sum_{i=1}^n (-\cos(y_i - \hat{f}^{-i}(x_i)))$$

- $n=20, 40, 50, 100, 200$ boyutlu veri setleri üretilmiştir.
- Bozulma derecesi $\gamma = 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95$ değerlerini almış ve veri setlerine farklı oranlarda aykırı gözlemler eklenmiştir.
- Üretilen yanıt değişkenine aykırı gözlem eklenmesi, tek değişkenli veri setlerinde olduğu gibi aşağıdaki eşitlik yardımıyla gerçekleştirilmiştir (Mahmood ve diğerleri,2017).

$$y_k = y + \gamma \pi \bmod(2\pi)$$

y_k : kirlenilen yanıt değişkeni

- Aykırı gözlem tespitinde kullanılacak eşik değeri (kesim noktası) için 95. yüzdelikler elde edilmiştir.
- NW ile LL tahmin edicilerinin risk değerleri, aykırı gözlem tespit oranları, maskeleye, sahte aykırı gözlem oranları incelenmiştir.
- Monte Carlo simülasyonu ile 1000 tekrarda elde edilen sonuçlar raporlanmıştır.

Yapılan simülasyonlar sonucunda, veri setinin bozulma derecesi %70 ve %70'ten daha büyük olduğunda önerilen yöntemin aykırı gözlem tespit oranının yüksek olduğu görülmüştür. $\gamma = 0.70, 0.80, 0.90$ olduğunda elde edilen sonuçlar aşağıdaki çizelgelerde verilmiştir:

Çizelge 5.9. $\gamma=0,70$ olduğunda farklı örneklemeler için simülasyon sonuçları

0,70		NW							LL						
N	aykırı gözlem	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
20	1	0,9989	0,001	0,0323	0,0004	0,0047	1,4682	0,3163	1	0	0	0	0	1,5260	0,1138
	2	0,9792	0,0207	0,0997	0	0	2,9441	0,4593	0,9938	0,0061	0,0551	0	0	2,977	0,3449
	3	0,9206	0,0793	0,1653	0,0001	0,0029	4,322	0,7135	0,9326	0,0673	0,151	0,0002	0,0036	4,3053	0,6345
	4	0,7515	0,2484	0,2189	0,000009	0,0024	5,5862	1,0452	0,7680	0,2319	0,2212	0,0001	0,0034	5,5507	0,8348
40	1	0,9994	0,0006	0,0705	0,0011	0,0061	1,4815	0,3003	1	0	0	0,0005	0,0038	1,55	0,0641
	2	0,9914	0,0085	0,0647	0,00002	0,0008	2,9884	0,4338	1	0	0	0	0	3,0617	0,0231
	3	0,9882	0,0117	0,0633	0,00005	0,0012	4,4993	0,4466	0,9966	0,0033	0,0365	0,00002	0,0008	4,5807	0,2929
	4	0,9574	0,0425	0,1113	0,00005	0,0012	5,9676	0,6063	0,9832	0,0167	0,0701	0,00002	0,0008	6,0567	0,4102
50	1	0,995	0,005	0,0944	0,0023	0,0087	1,4871	0,2956	1	0	0	0,0016	0,0058	1,551	0,0693
	2	0,9945	0,0055	0,0473	0,0001	0,0022	3,0286	0,3344	0,999	0,001	0,0223	0,00002	0,0006	3,0857	0,191
	3	0,9919	0,0081	0,0532	0,00002	0,0006	4,5188	0,4404	0,9973	0,0027	0,0332	0,00002	0,0006	4,6026	0,2728
	4	0,9814	0,0185	0,0745	0,00004	0,0009	5,9823	0,557	0,9924	0,0076	0,0507	0,00004	0,0009	6,0836	0,3969
100	1	0,998	0,002	0,0834	0,043	0,0384	1,5227	0,2249	1	0	0	0,0291	0,0387	1,5664	0,0522
	2	0,9975	0,0025	0,0352	0,0084	0,053	3,0612	0,3113	1	0	0	0,0055	0,0439	3,1172	0,1635
	3	0,997	0,003	0,0314	0,0016	0,0294	4,5972	0,3743	1	0	0	0,0005	0,0023	4,675	0,1596
	4	0,9965	0,0035	0,0333	0,0001	0,0011	6,1315	0,4205	1	0	0	0,00003	0,0007	6,2271	0,1804
200	1	1	0	0	0,0516	0,0395	1,5468	0,1738	1	0	0	0,0513	0,0053	1,5735	0,0313
	2	0,999	0,001	0,0223	0,0462	0,043	3,1074	0,211	1	0	0	0,0453	0,0091	3,1416	0,0677
	3	0,9986	0,0014	0,0315	0,0383	0,0495	4,6511	0,2919	1	0	0	0,0339	0,0157	4,7113	0,1049
	4	0,9985	0,0015	0,0193	0,0245	0,0755	6,2132	0,2956	1	0	0	0,0036	0,0448	6,2724	0,1977

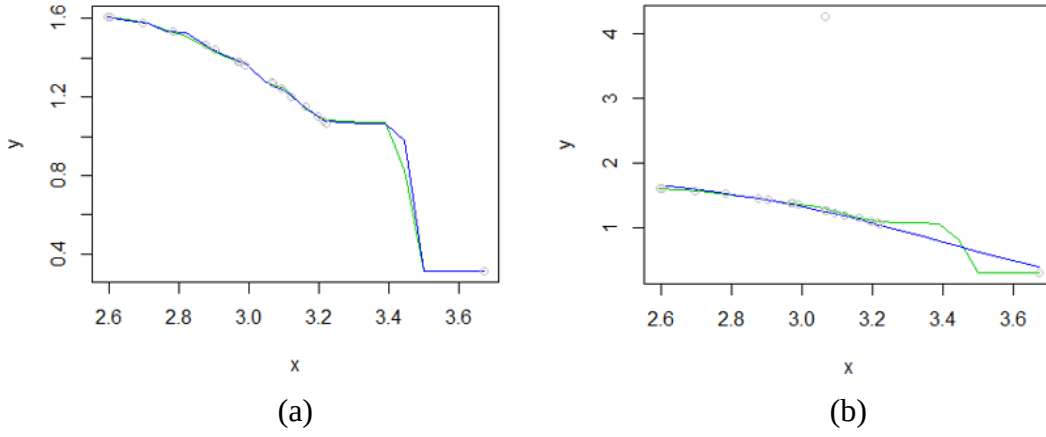
Çizelge 5.10. $\gamma=0,80$ olduğunda farklı örneklemeler için simülasyon sonuçları

$\gamma = 0,80$		NW							LL						
N	aykırı gözlem	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
20	1	1	0	0	0,0001	0,0023	1,6692	0,4338	1	0	0	0	0	1,7675	0,1485
	2	0,9844	0,0156	0,0944	0	0	3,4277	0,5256	0,9975	0,0025	0,0519	0	0	3,487	0,3808
	3	0,9405	0,0594	0,159	0	0	5,1025	0,7873	0,9519	0,048	0,136	0	0	5,1038	0,7284
	4	0,8135	0,1864	0,2103	0	0	6,6967	1,0799	0,816	0,1839	0,2131	0,00008	0,0023	6,6474	0,9339
40	1	0,995	0,005	0,0705	0,0008	0,0061	1,7186	0,3436	1	0	0	0,0002	0,0026	1,784	0,0934
	2	0,9935	0,0065	0,0566	0,00002	0,0008	3,4728	0,4424	0,9985	0,0015	0,0273	0	0	3,5447	0,2273
	3	0,9925	0,0165	0,071	0	0	5,1921	0,5868	0,9963	0,0036	0,0348	0	0	5,293	0,362
	4	0,9704	0,0295	0,1034	0	0	6,977	0,5886	0,9906	0,0093	0,0548	0	0	7,0629	0,3802
50	1	0,997	0,003	0,0891	0,0013	0,0071	1,7197	0,3310	1	0	0	0,0008	0,0041	1,7873	0,0815
	2	0,9964	0,0036	0,0566	0,00004	0,0009	3,4636	0,4922	0,9996	0,0004	0,0158	0,00002	0,0006	3,5626	0,1825
	3	0,9933	0,0067	0,0467	0	0	5,233	0,4907	0,9983	0,0017	0,0235	0	0	5,3227	0,3192
	4	0,9844	0,0156	0,0708	0	0	6,968	0,5899	0,9959	0,0041	0,0369	0	0	7,0953	0,284
100	1	0,997	0,003	0,0547	0,0428	0,0599	1,7622	0,2407	1	0	0	0,0274	0,0223	1,7975	0,0398
	2	0,9985	0,0015	0,0273	0,0051	0,0432	3,5232	0,3602	1	0	0	0,0017	0,0057	3,5893	0,0428
	3	0,997	0,003	0,0314	0,0032	0,0519	5,3101	0,351	1	0	0	0,0011	0,0316	5,3782	0,1381
	4	0,9967	0,0033	0,0304	0,00003	0,0005	7,0768	0,4101	1	0	0	0,00001	0,0003	7,1605	0,1634
200	1	1	0	0	0,0510	0,0368	1,7888	0,1264	1	0	0	0,0517	0,0049	1,8008	0,044
	2	1	0	0	0,0502	0,0758	3,5589	0,2764	1	0	0	0,0463	0,0312	3,5930	0,1367
	3	0,9986	0,0014	0,0278	0,0388	0,0587	5,3528	0,2928	1	0	0	0,0335	0,0158	5,3975	0,0744
	4	0,998	0,002	0,0222	0,0218	0,0661	7,1322	0,3227	1	0	0	0,0114	0,0441	7,1827	0,1722

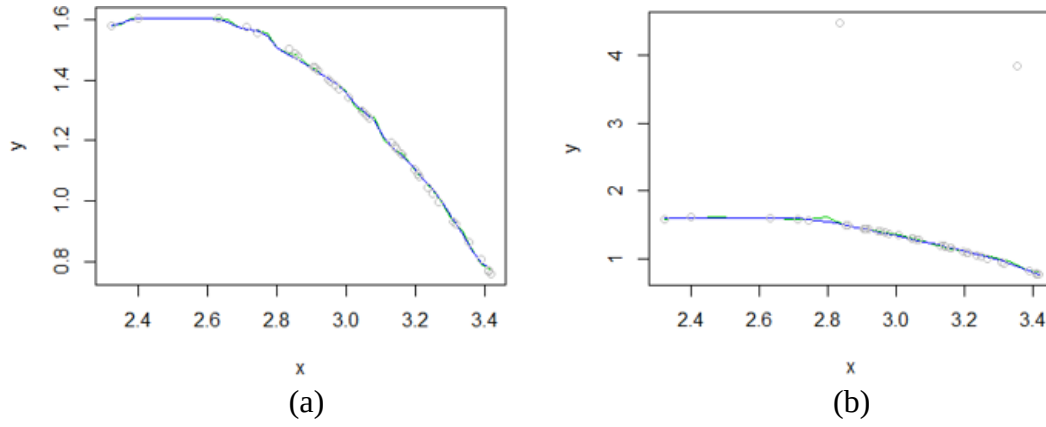
Çizelge 5.11. $\gamma=0,90$ olduğunda farklı örneklemeler için simülasyon sonuçları

$\gamma = 0,90$		NW							LL						
N	aykırı gözlem	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
20	1	1	0	0	0,00005	0,0016	1,8247	0,4746	1	0	0	0	0	1,9124	0,2495
	2	0,9880	0,002	0,1095	0	0	3,7151	0,6148	0,987	0,0129	0,0793	0	0	3,7981	0,4641
	3	0,9474	0,0526	0,1616	0	0	5,6664	0,6593	0,956	0,0439	0,1409	0	0	5,7127	0,5369
	4	0,8513	0,1486	0,2084	0	0	7,5375	0,8191	0,8594	0,1405	0,213	0,0024	0,0038	7,5461	0,7634
40	1	0,9998	0,0002	0,0446	0,0004	0,004	1,8354	0,4552	1	0	0	0,00005	0,0011	1,9383	0,1187
	2	0,997	0,003	0,0386	0	0	3,7683	0,5106	0,9995	0,0005	0,0158	0	0	3,8722	0,2025
	3	0,9933	0,0066	0,049	0	0	5,7058	0,4944	0,9979	0,002	0,0257	0	0	5,7966	0,2808
	4	0,9783	0,0216	0,0954	0	0	7,5918	0,6664	0,9914	0,0085	0,0554	0,00002	0,0008	7,6992	0,5021
50	1	1	0	0	0,005	0,0035	1,8563	0,4111	1	0	0	0,0005	0,0031	1,9424	0,0772
	2	0,997	0,003	0,0386	0,0004	0,0144	3,7667	0,5149	1	0	0	0	0	3,8791	0,1762
	3	0,995	0,005	0,0431	0	0	5,7013	0,5432	1	0	0	0	0	5,8007	0,2988
	4	0,9865	0,0135	0,0607	0	0	7,5632	0,7263	0,9969	0,0041	0,0333	0	0	7,7104	0,4756
100	1	1	0	0	0,0464	0,0075	1,2707	0,4147	1	0	0	0	0	1,3407	0,1221
	2	0,998	0,002	0,0315	0,0081	0,0766	3,8327	0,3473	1	0	0	0	0	3,8865	0,1278
	3	0,9973	0,0026	0,0297	0,0001	0,0014	5,7598	0,4179	1	0	0	0	0	5,8203	0,2373
	4	0,9967	0,0032	0,0283	0,0019	0,04258	7,6837	0,4635	1	0	0	0	0	7,7607	0,2524
200	1	1	0	0	0	0	1,9204	0,2298	1	0	0	0	0	1,9406	0,1236
	2	1	0	0	0	0	3,8619	0,261	1	0	0	0	0	3,8888	0,12
	3	0,9998	0,0002	0,0257	0,0451	0,1035	5,7728	0,3817	1	0	0	0	0	5,8265	0,183
	4	0,9985	0,0015	0,0193	0,0229	0,0905	7,7231	0,3733	1	0	0	0	0	7,7744	0,2084

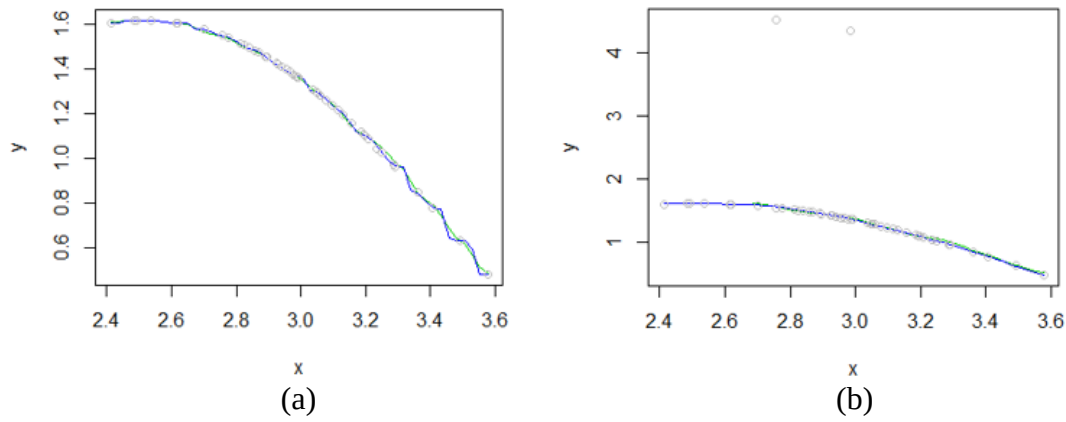
Yapılan simülasyonlar sonucunda elde edilen eğrilerin bir kısmı aşağıda verilmiştir.



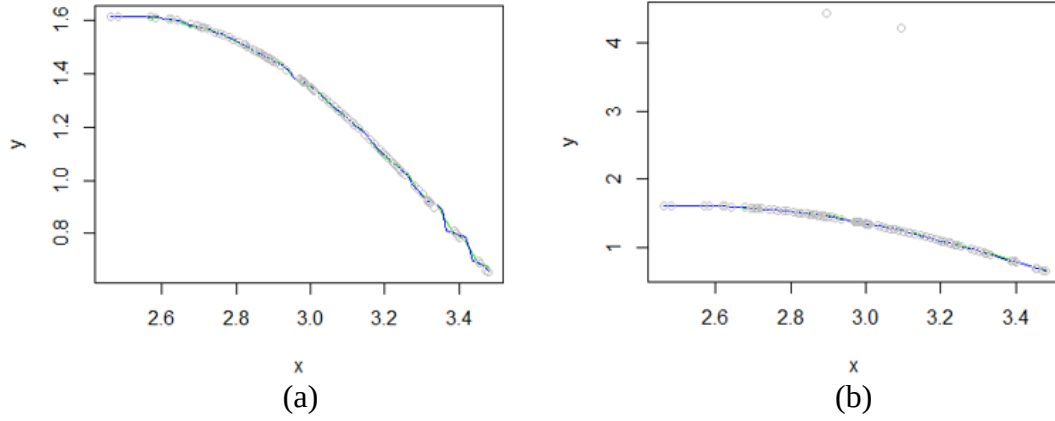
Şekil 5.28. Bir aykırı gözlem varlığında $n = 20$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL) (a) aykırı gözlem yokken (b) aykırı gözlem varken



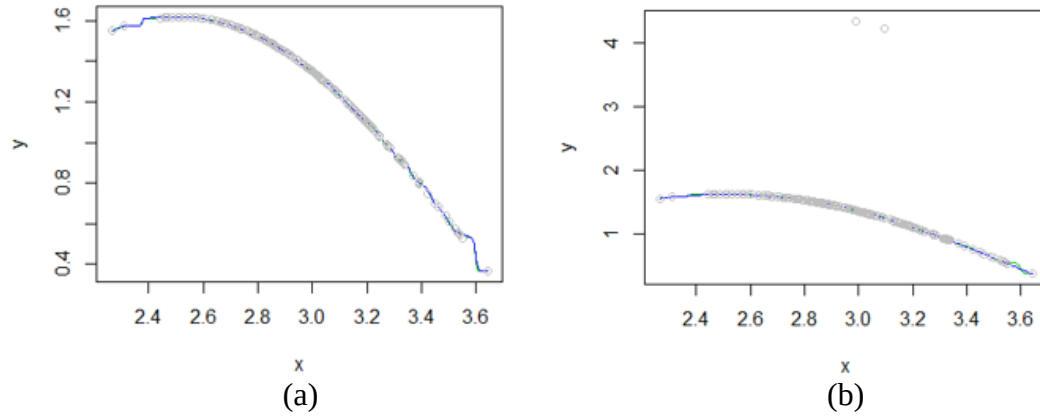
Şekil 5.29. Bir aykırı gözlem varlığında $n = 40$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL) (a) aykırı gözlem yokken (b) aykırı gözlem varken



Şekil 5.30. Bir aykırı gözlem varlığında $n = 50$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL) (a) aykırı gözlem yokken (b) aykırı gözlem varken

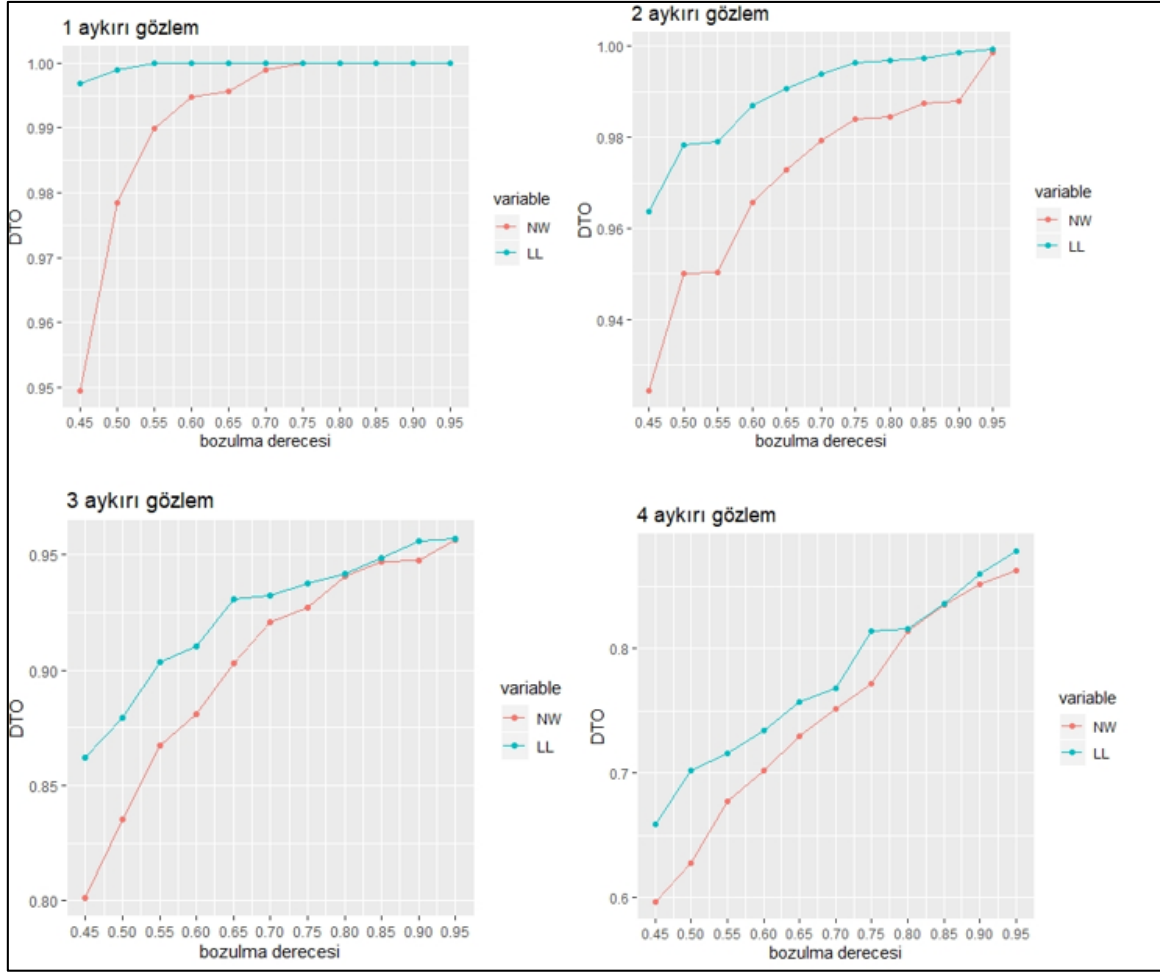


Şekil 5.31. İki aykırı gözlem varlığında $n = 100$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL), (a) aykırı gözlem yokken (b) aykırı gözlem varken



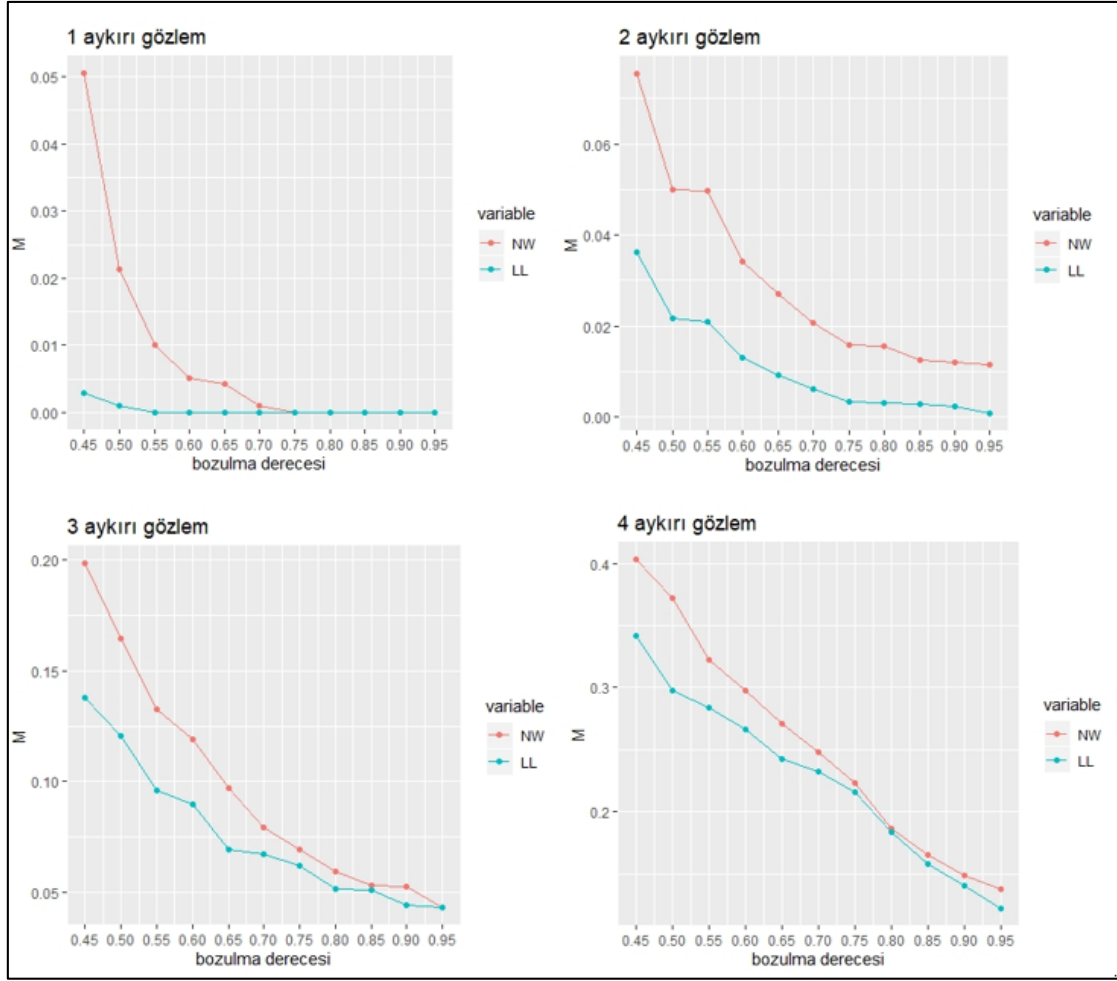
Şekil 5.32. İki aykırı gözlem varlığında $n = 200$ iken elde edilen eğriler (yeşil eğri NW, mavi eğri LL), (a) aykırı gözlem yokken (b) aykırı gözlem varken

Yapılan simülasyonlar sonucu NW ve LL tahmin edicileri için farklı bozulma dereceleri ve farklı örneklem boyutlarında doğru tespit oranı (DTO), maskeleye oranı (M) ve sahte aykırı gözlem oranı (S) ve risk fonksiyonu değerleri (R) aşağıdaki şekillerde verildiği gibi elde edilmiştir.



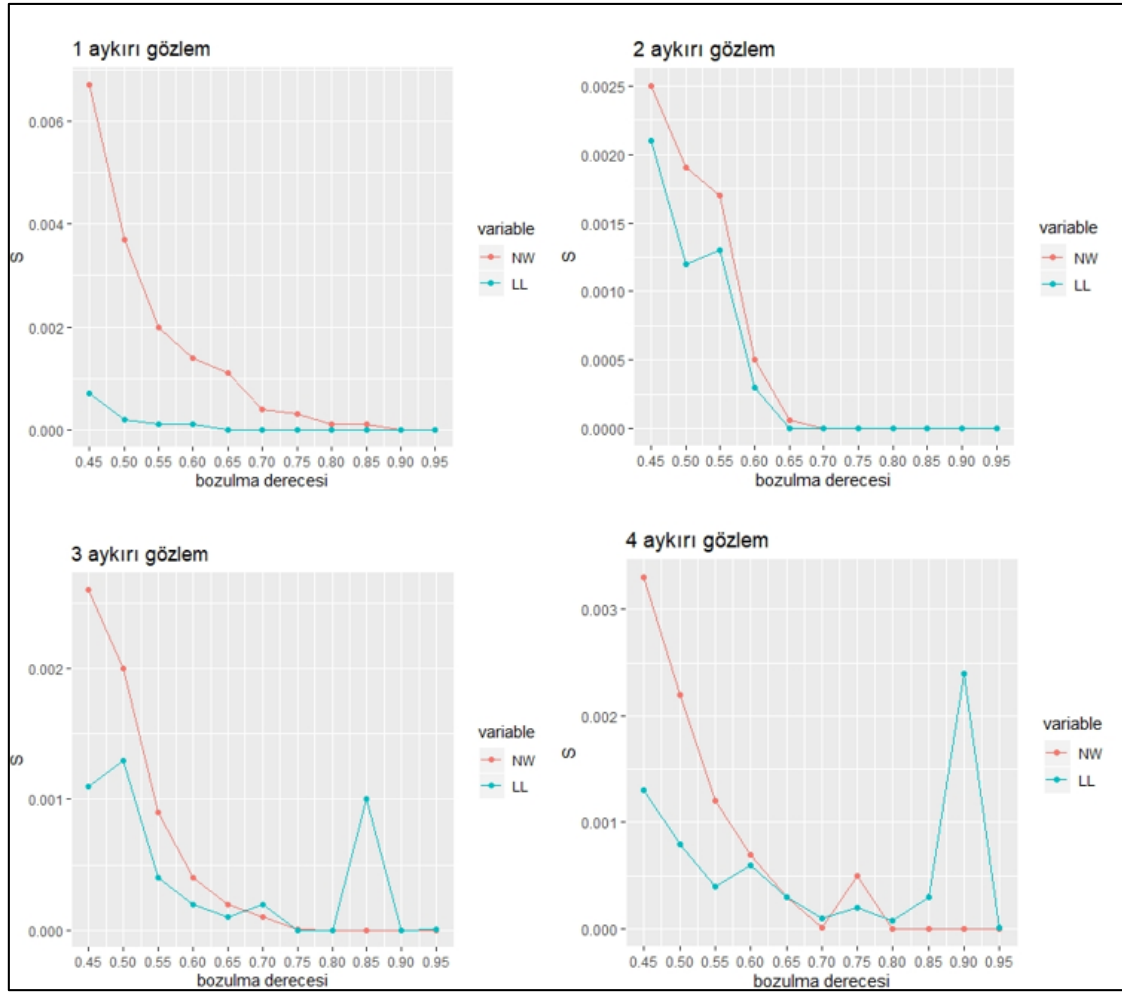
Şekil 5.33. $n=20$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları

$n=20$ olduğunda, bozulma derecesi arttıkça, her iki yöntem için de aykırı gözlem tespit oranının arttığı görülmektedir. Aykırı gözlem sayısı arttıkça doğru tespit oranı beklendiği üzere azalmaktadır. Veri setinin %10'u kadar aykırı gözlem, yüksek oranda doğru olarak tespit edilmiştir. LL yönteminde doğru tespit oranının, NW yöntemine göre daha yüksek olduğu gözlenmiştir. Bu durum, LL yönteminin aykırı gözlem varlığından daha fazla etkilendiğini göstermektedir.



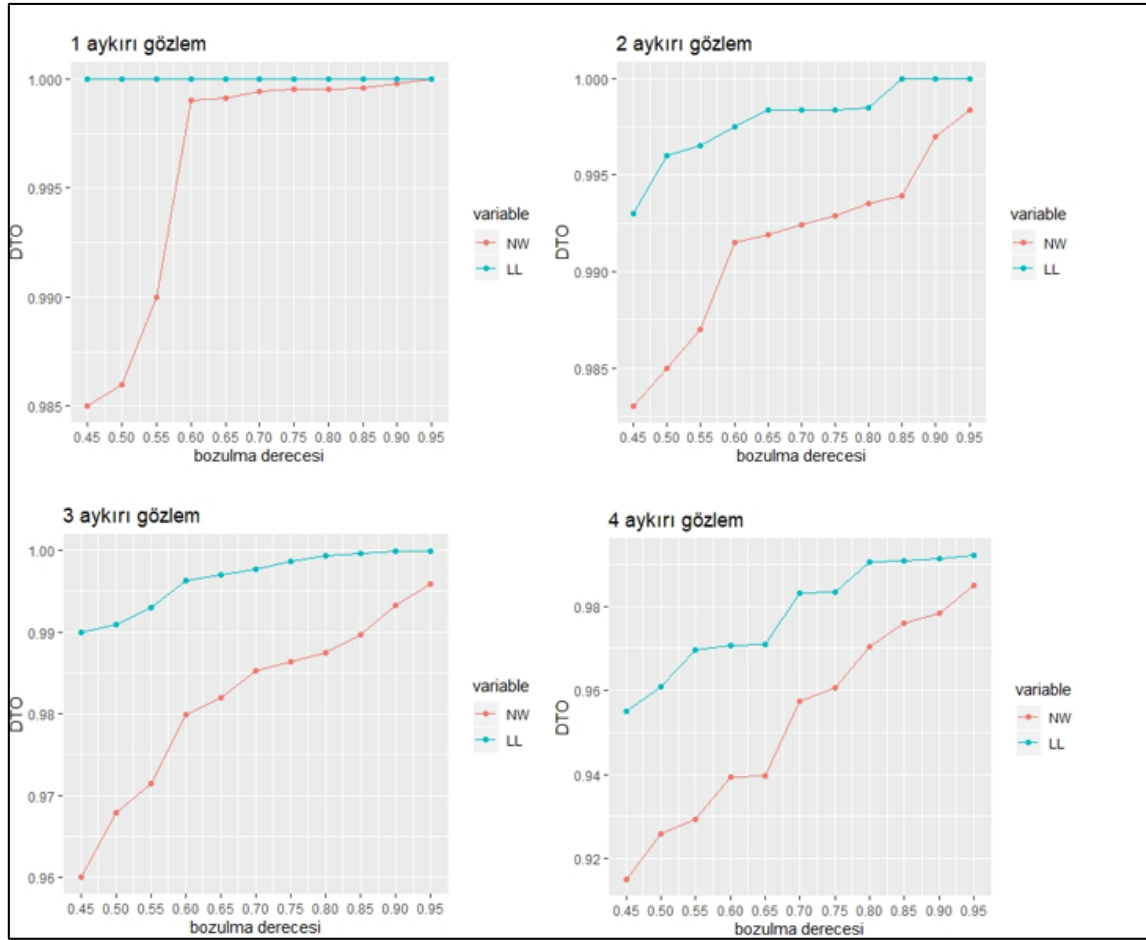
Şekil 5.34. $n=20$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleme oranları

$n=20$ için maskeleme oranları, doğru tespit oranı ile uyumlu sonuçlar vermiştir. Veri setindeki bozulma oranı arttıkça maskeleme oranı sıfıra yaklaşmıştır. LL yönteminin NW yöntemine göre daha düşük maskeleme oranına sahip olması, doğru tespit oranında olduğu gibi, aykırı gözlem varlığından daha fazla etkilendiğini göstermektedir.



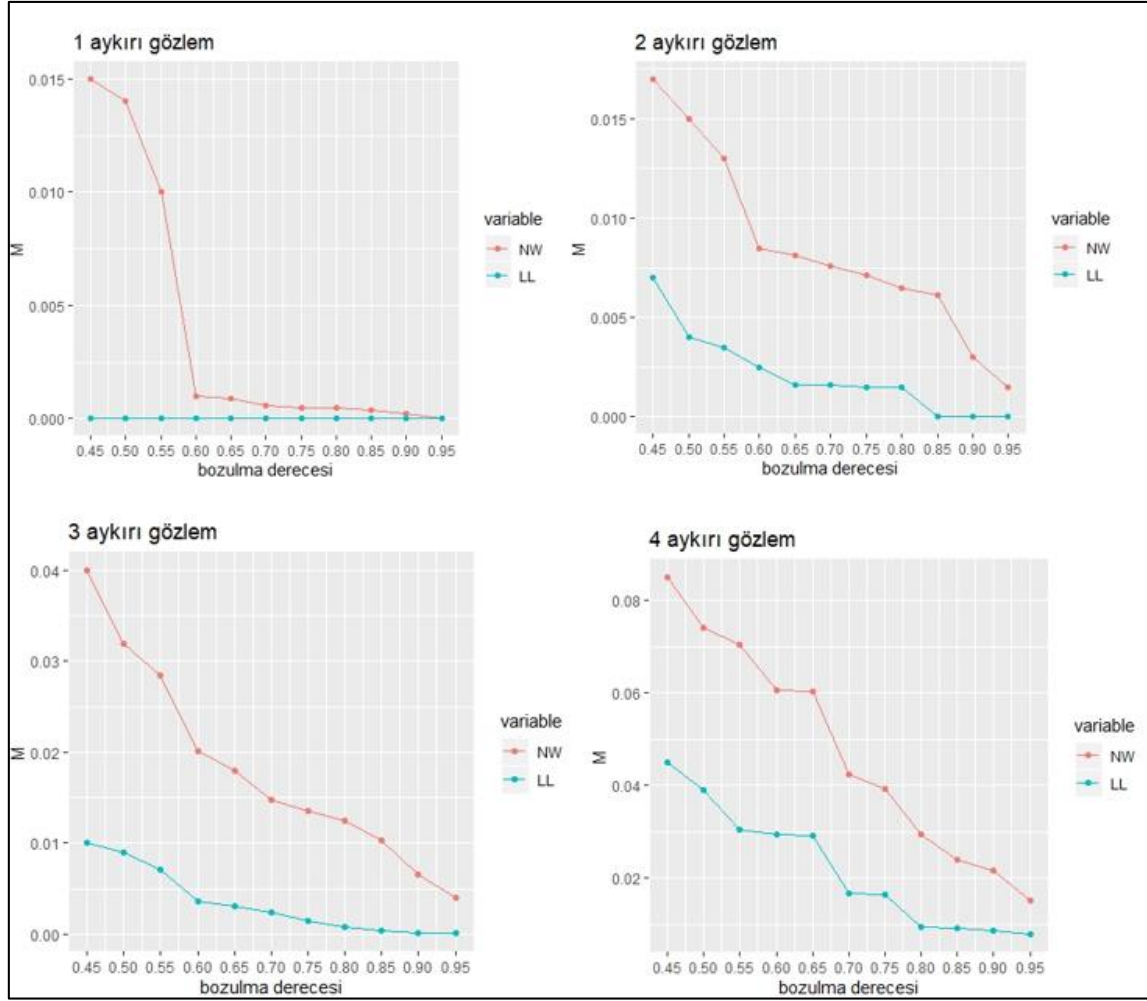
Şekil 5.35. $n=20$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları

$n=20$ için sahte aykırı gözlem oranlarına bakıldığında, her iki yöntem için de sıfıra yakın değerler aldığı gözlenmiştir. Ancak sahte aykırı gözlem varlığının, doğru tespit oranı ve maskeleye ölçüleri gibi belli bir düzene sahip olmadığı görülmektedir.



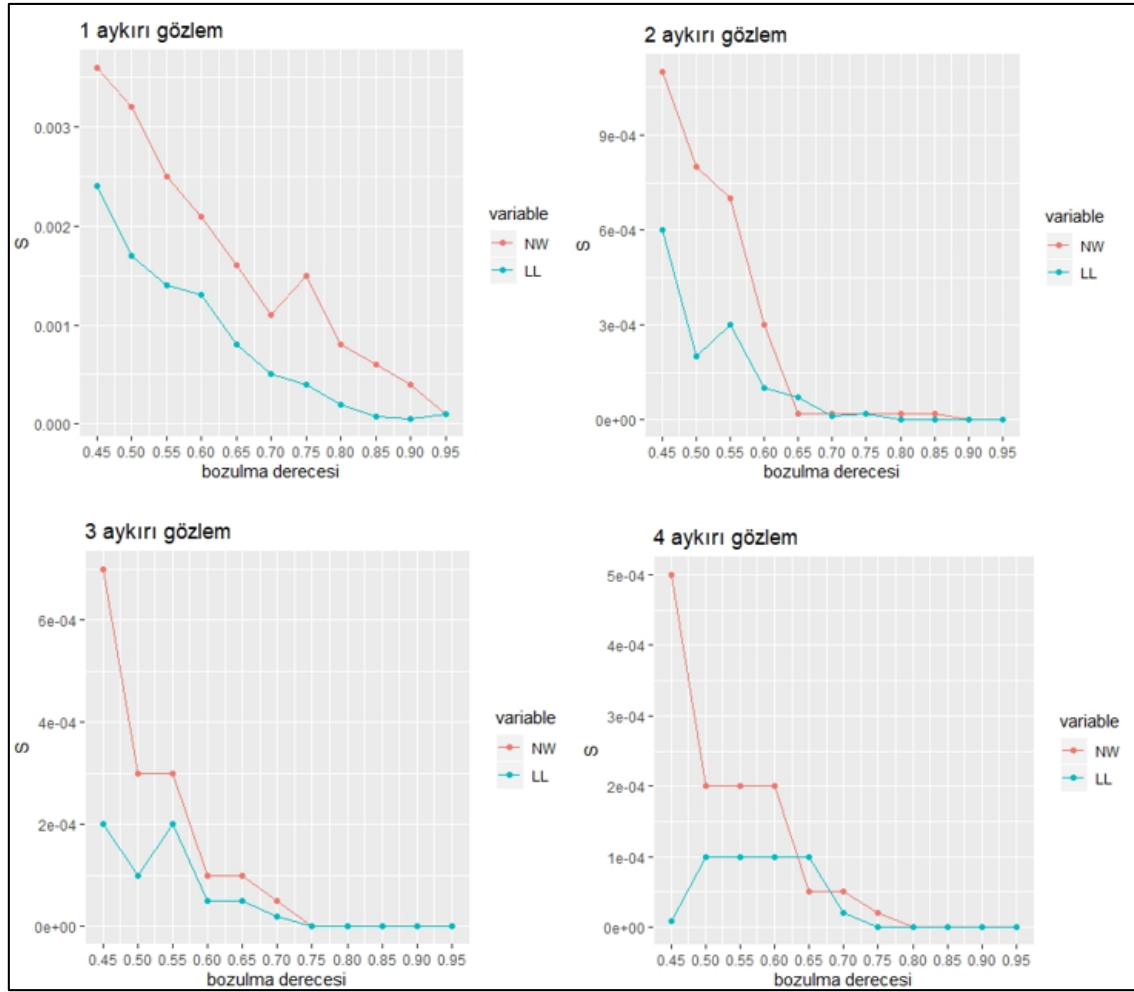
Şekil 5.36. $n=40$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları

$n=40$ olduğunda, LL yönteminin aykırı gözlem tespit oranının daha yüksek olduğu görülmektedir. Her iki yöntem de düşük bozulma derecesinde dahi yüksek doğru tespit oranına sahiptir. LL yönteminin doğru tespit oranının yüksek olması, aykırı gözlemden daha fazla etkilendiğini göstermektedir.



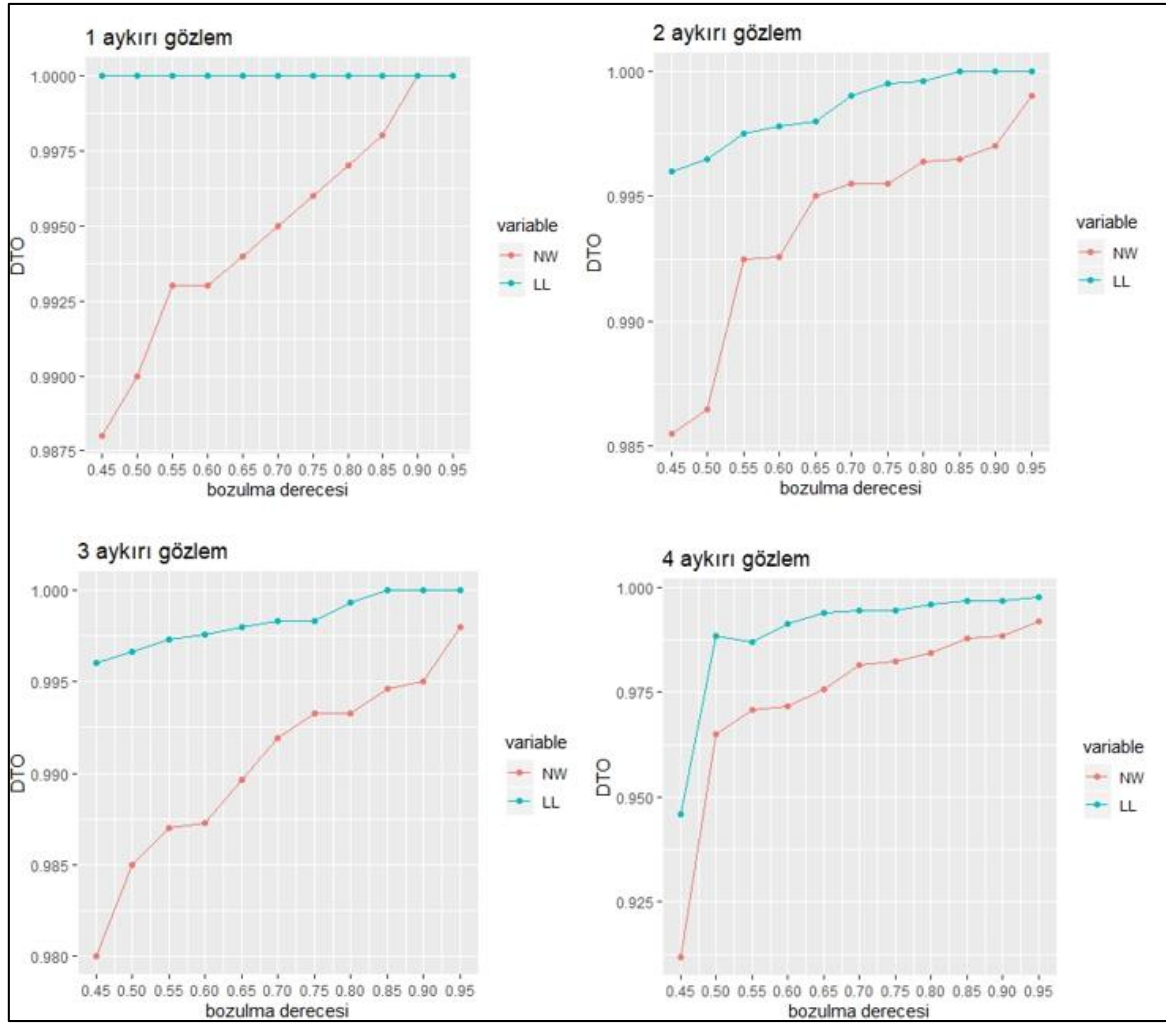
Şekil 5.37. $n=40$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleme oranları

$n=40$ için maskeleme oranları, doğru tespit oranı ile uyumlu sonuçlar vermiştir. Bozulma derecesi arttıkça maskeleme oranının azaldığı ve %85'den fazla bozulma oranında sıfır olduğu gözlenmiştir.



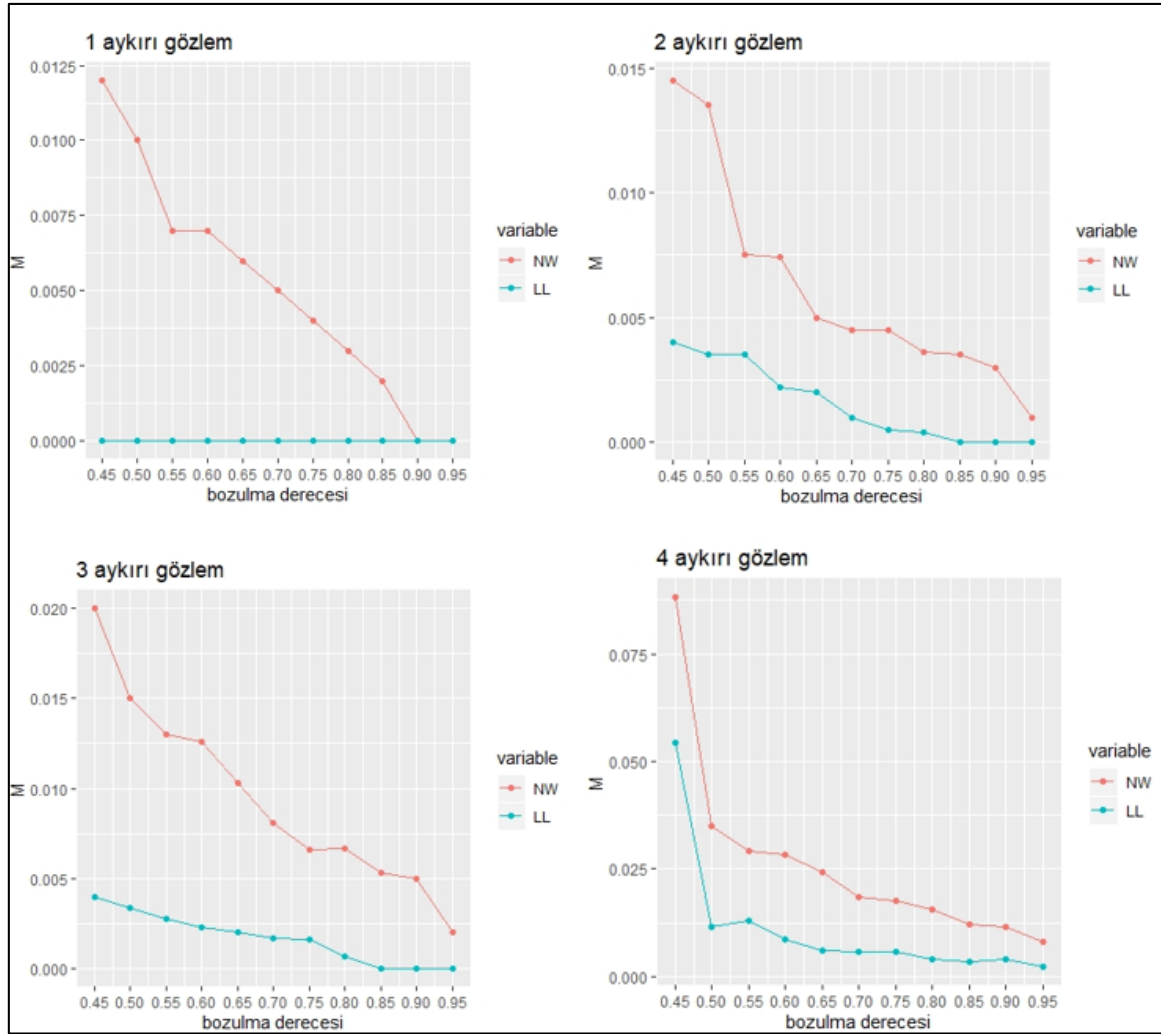
Şekil 5.38. $n=40$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları

$n=40$ için sahte aykırı gözlem oranlarına bakıldığında, her iki yöntem için de sıfıra yakın değerler aldığı gözlenmiştir. Ancak sahte aykırı gözlem varlığının, doğru tespit oranı ve maskeleye ölçüleri gibi belli bir düzene sahip olmadığı ve $n=20$ 'ye kıyasla, sahte aykırı gözlem oranının azaldığı görülmektedir.



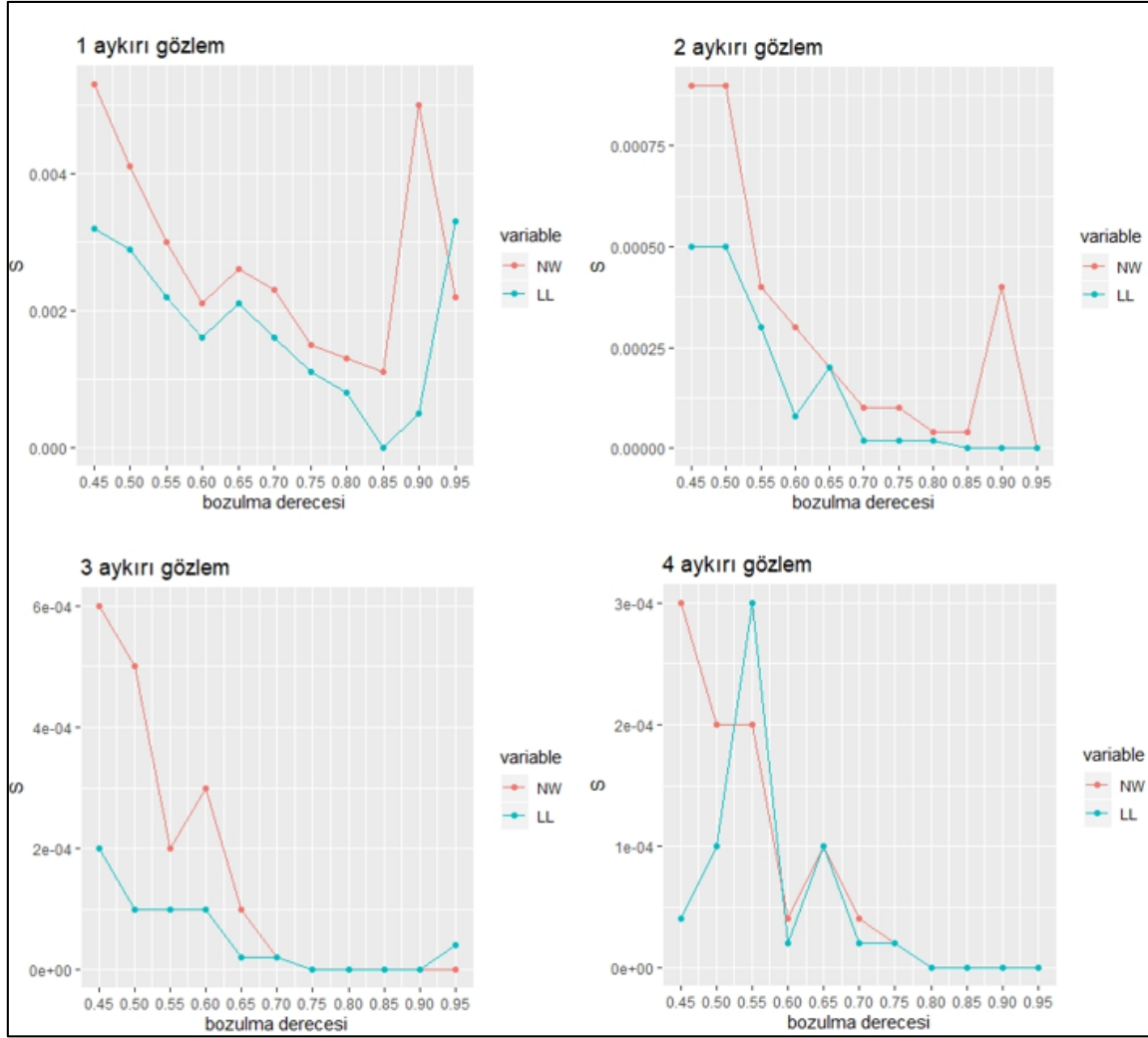
Şekil 5.39. $n=50$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları

$n=50$ olduğunda, LL yönteminin aykırı gözlem tespit oranının daha yüksek olduğu görülmektedir. Her iki yöntem de düşük bozulma derecesinde dahi yüksek doğru tespit oranına sahiptir. LL yönteminin doğru tespit oranının yüksek olması, aykırı gözlemde daha fazla etkilendiğini göstermektedir.



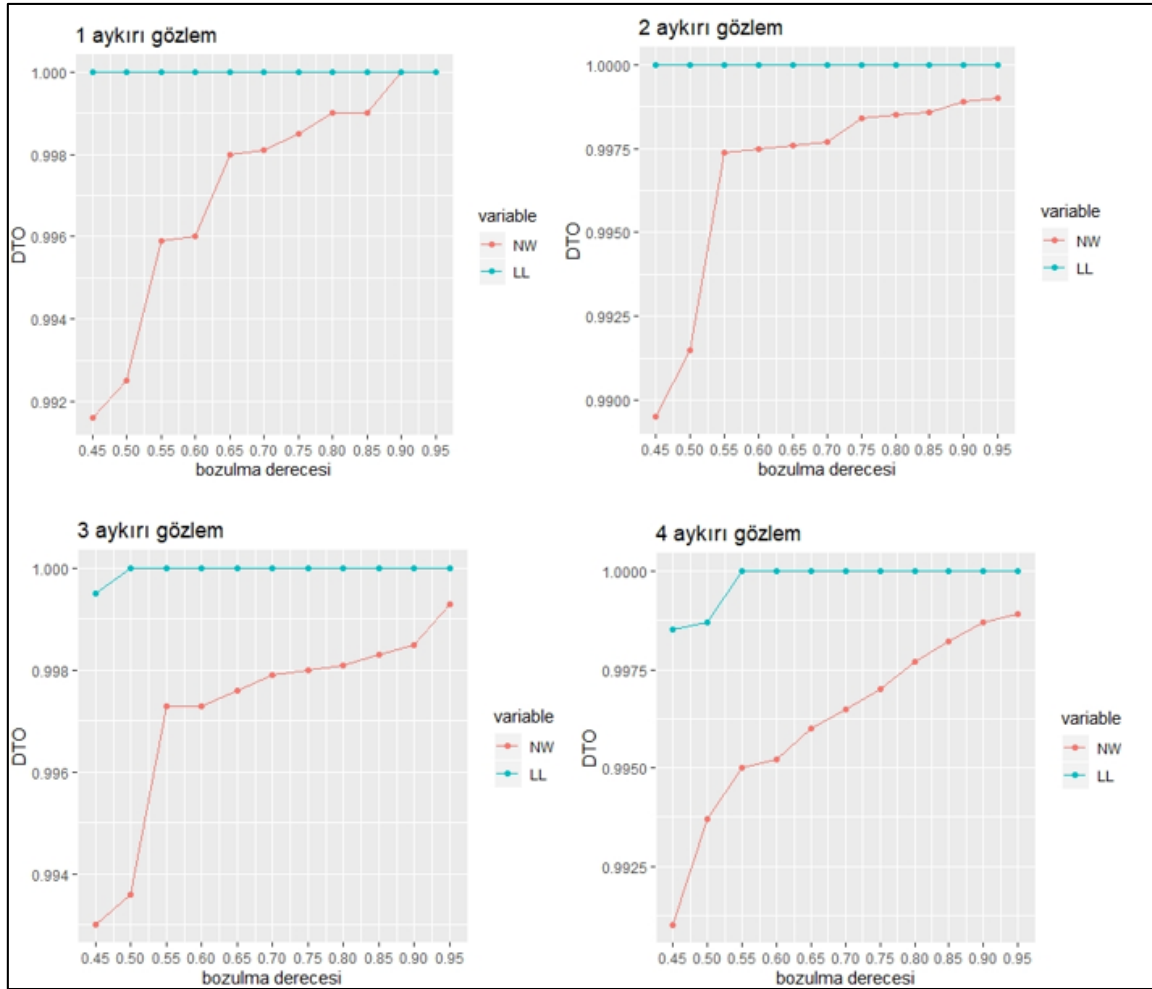
Şekil 5.40. $n=50$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleme oranları

$n=50$ için maskeleme oranları, veri setinin bozulma derecesi arttıkça azalmaktadır. LL yönteminin maskeleme oranlarının NW yöntemine göre daha düşük olduğu gözlenmiştir.



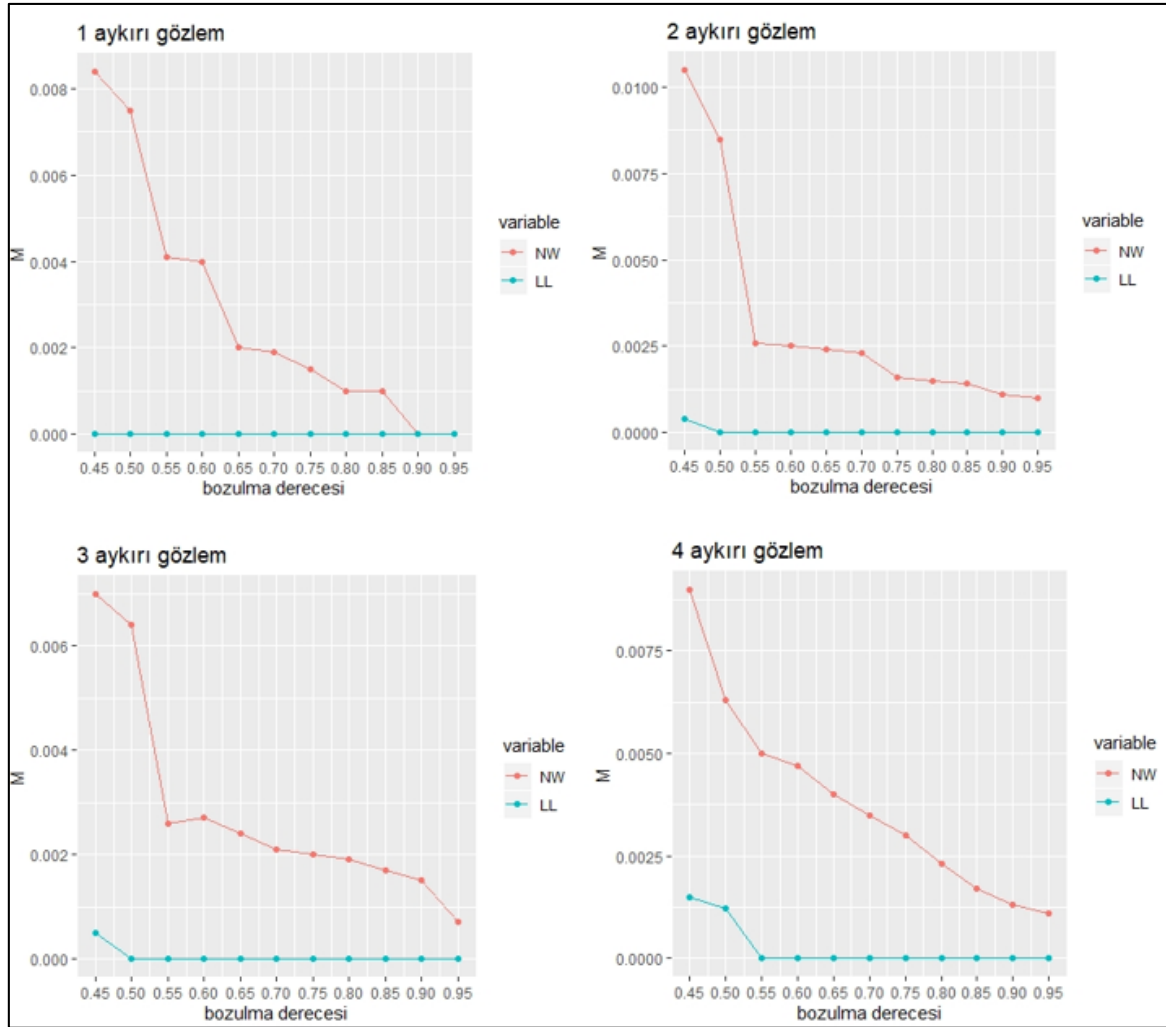
Şekil 5.41. $n=50$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları

$n=50$ olduğunda sahte aykırı gözlem oranlarına bakıldığında, her iki yöntem için de sıfıra yakın değerler aldığı gözlenmiştir. Ancak sahte aykırı gözlem varlığının, doğru tespit oranı ve maskeleye ölçüleri gibi belli bir düzene sahip olmadığı ve $n=40$ 'a kıyasla, sahte aykırı gözlem oranının azaldığı görülmektedir.



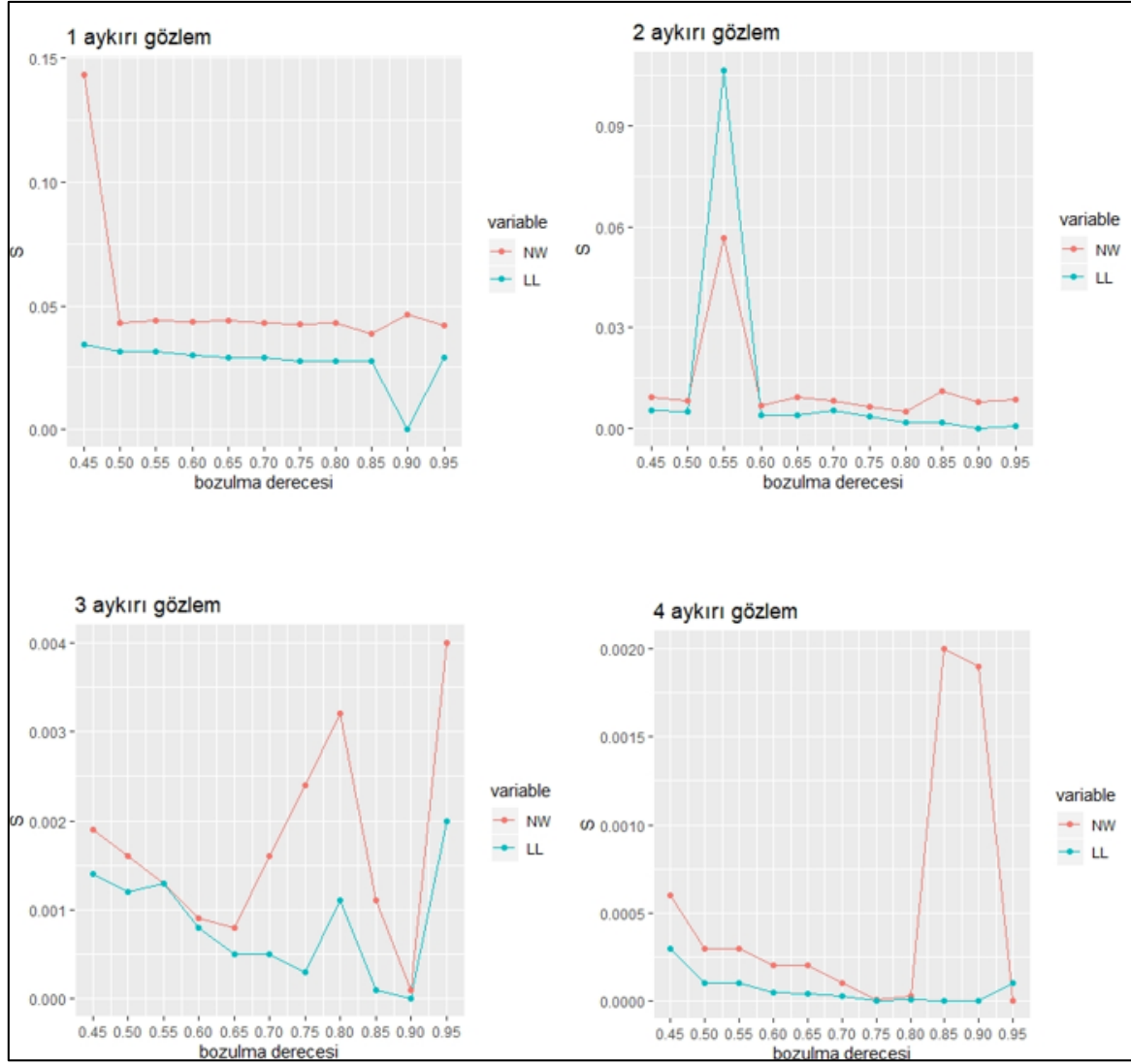
Şekil 5.42. $n=100$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları

$n=100$ olduğunda, LL yönteminin doğru tespit oranının daha yüksek olduğu görülmektedir. Her iki yöntem de düşük bozulma derecesinde dahi yüksek doğru tespit oranına sahiptir.



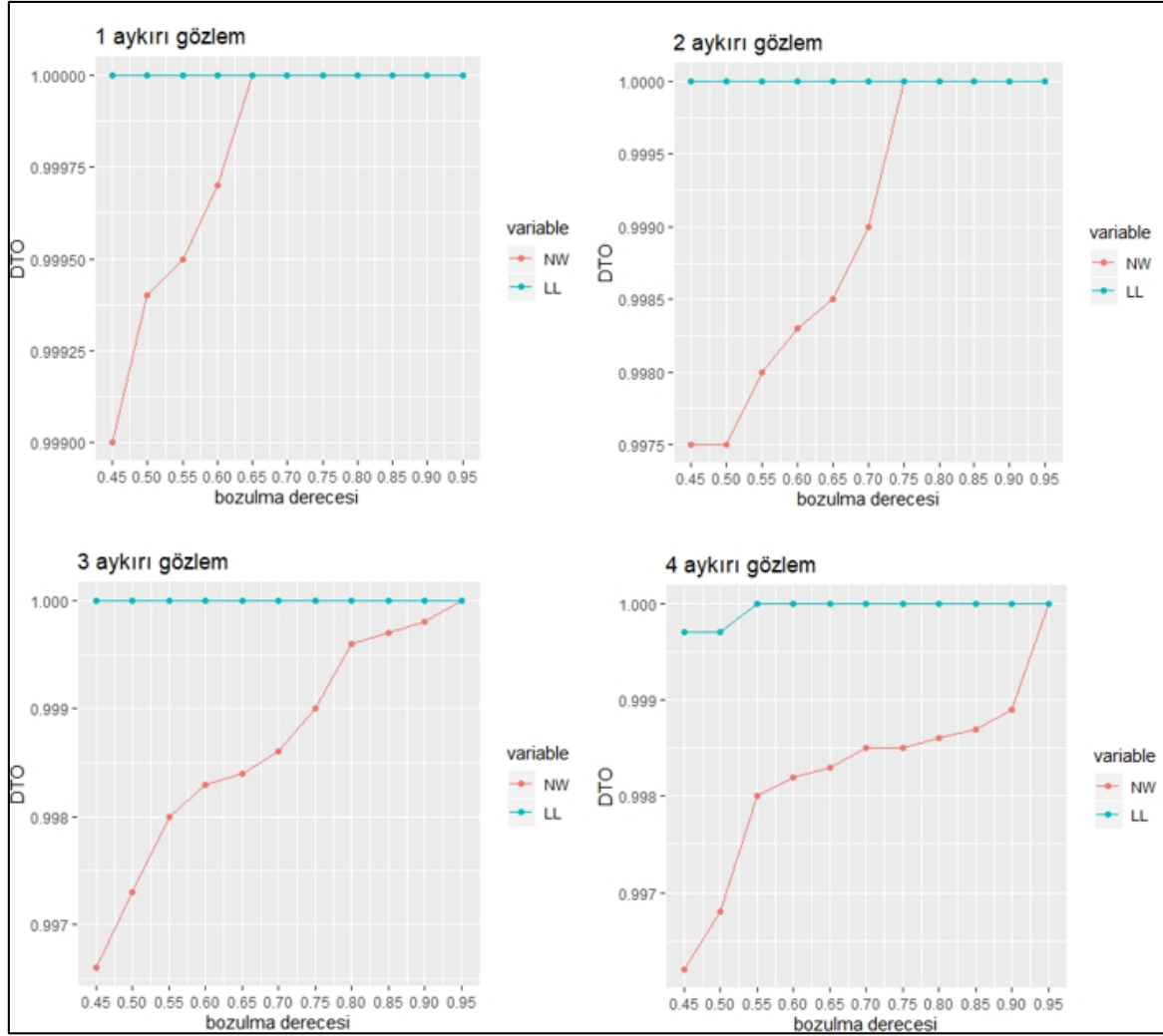
Şekil 5.43. $n=100$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleme oranları

$n=100$ için maskeleme oranları, veri setinin bozulma derecesi arttıkça azalmaktadır. LL yönteminin maskeleme oranlarının NW yöntemine göre daha düşük olduğu gözlenmiştir.



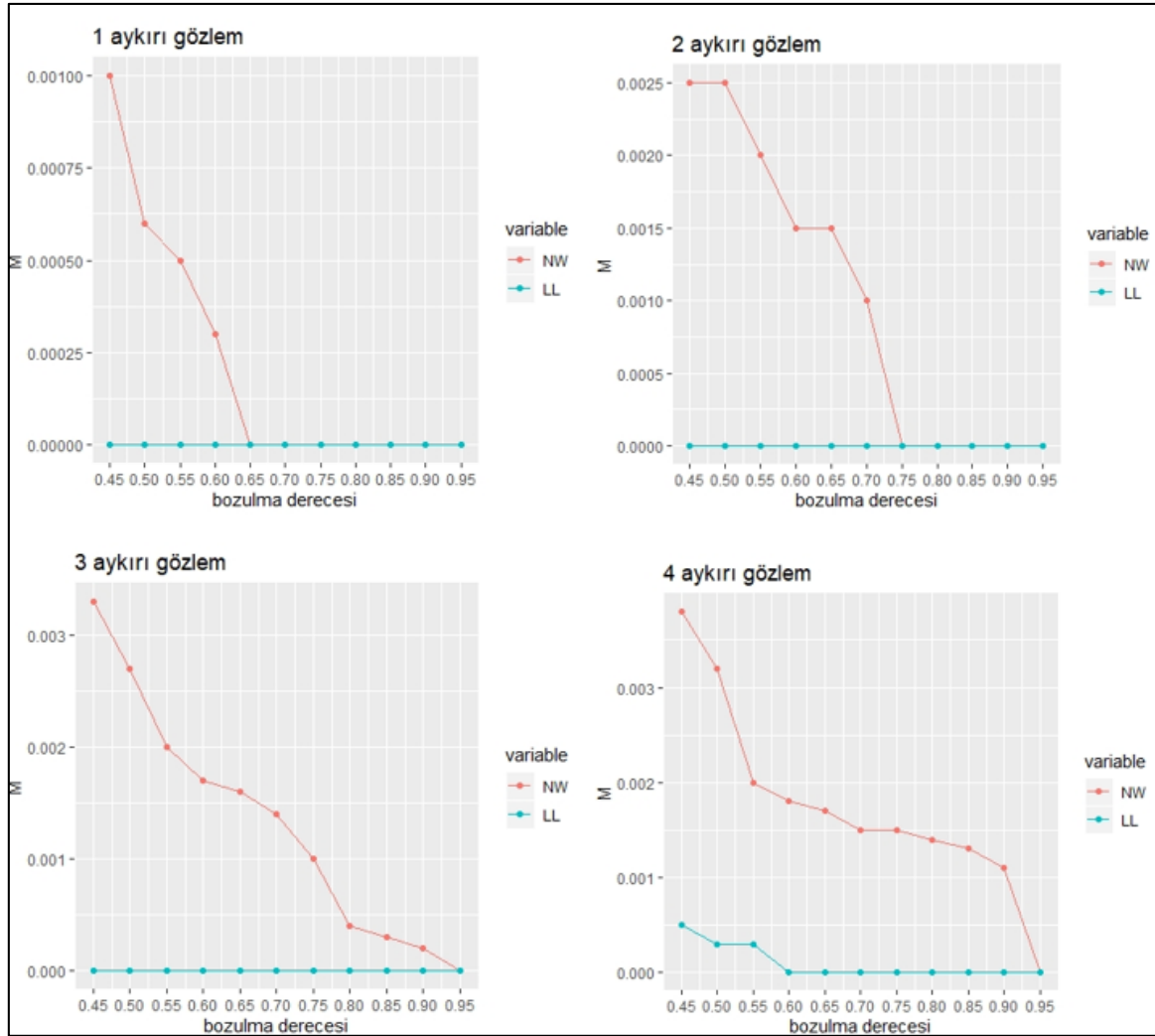
Şekil 5.44. $n=100$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları

$n=100$ için sahte aykırı gözlem oranlarına bakıldığında, her iki yöntem için de sıfıra yakın değerler aldığı gözlenmiştir. Sahte aykırı gözlem varlığının, doğru tespit oranı ve maskeleye ölçüleri gibi belli bir düzene sahip olmadığı görülmektedir.



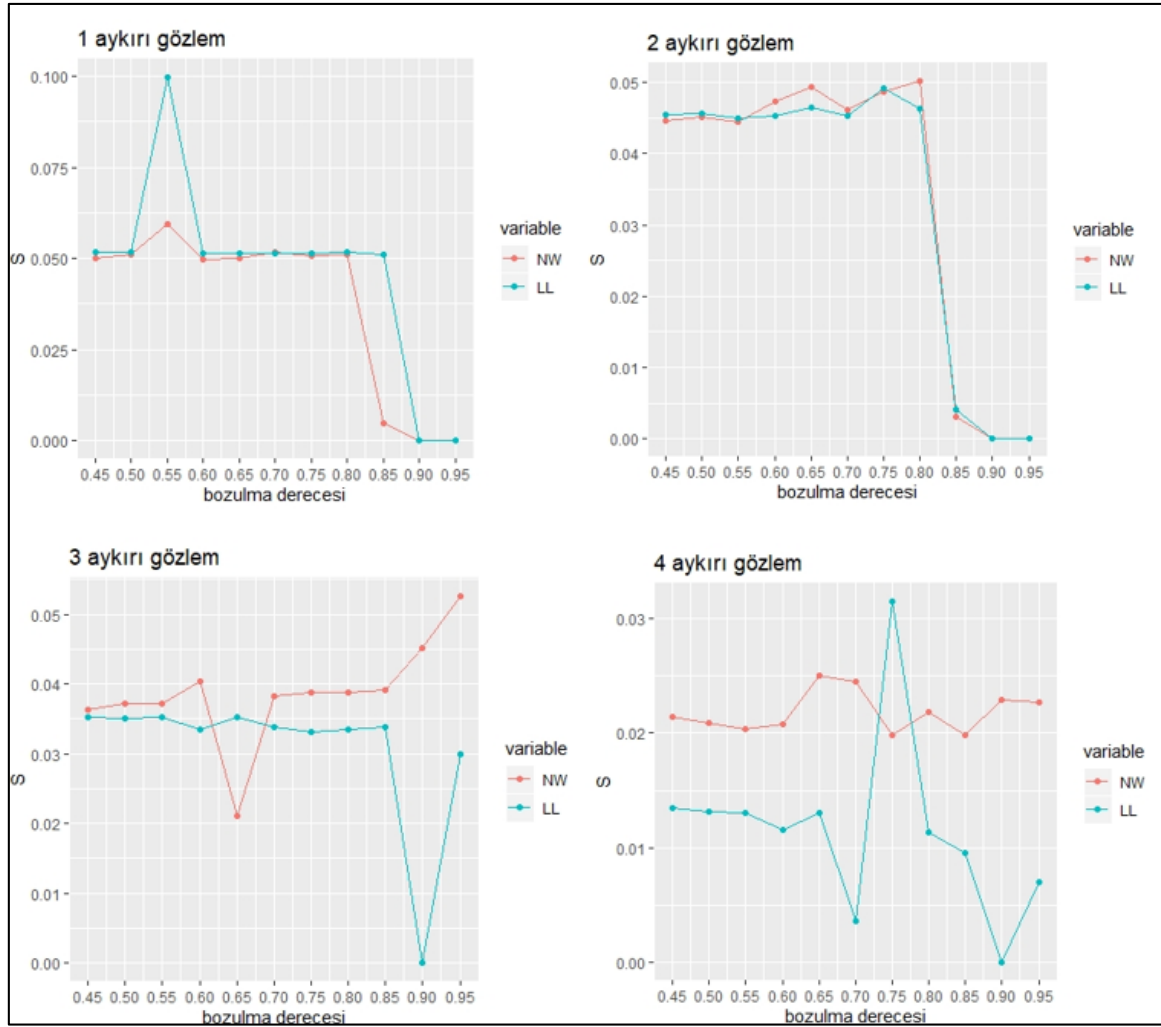
Şekil 5.45. $n=200$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin doğru tespit oranları

$n=200$ olduğunda, LL yöntemi kullanıldığında dört aykırı değer bütün bozulma parametrelerinde doğru tespit oranının daha yüksek olduğu görülmektedir. Her iki yöntem de düşük bozulma derecesinde dahi yüksek doğru tespit oranına sahiptir. LL yönteminin doğru tespit oranının yüksek olması, aykırı gözlemden daha fazla etkilendiğini göstermektedir.



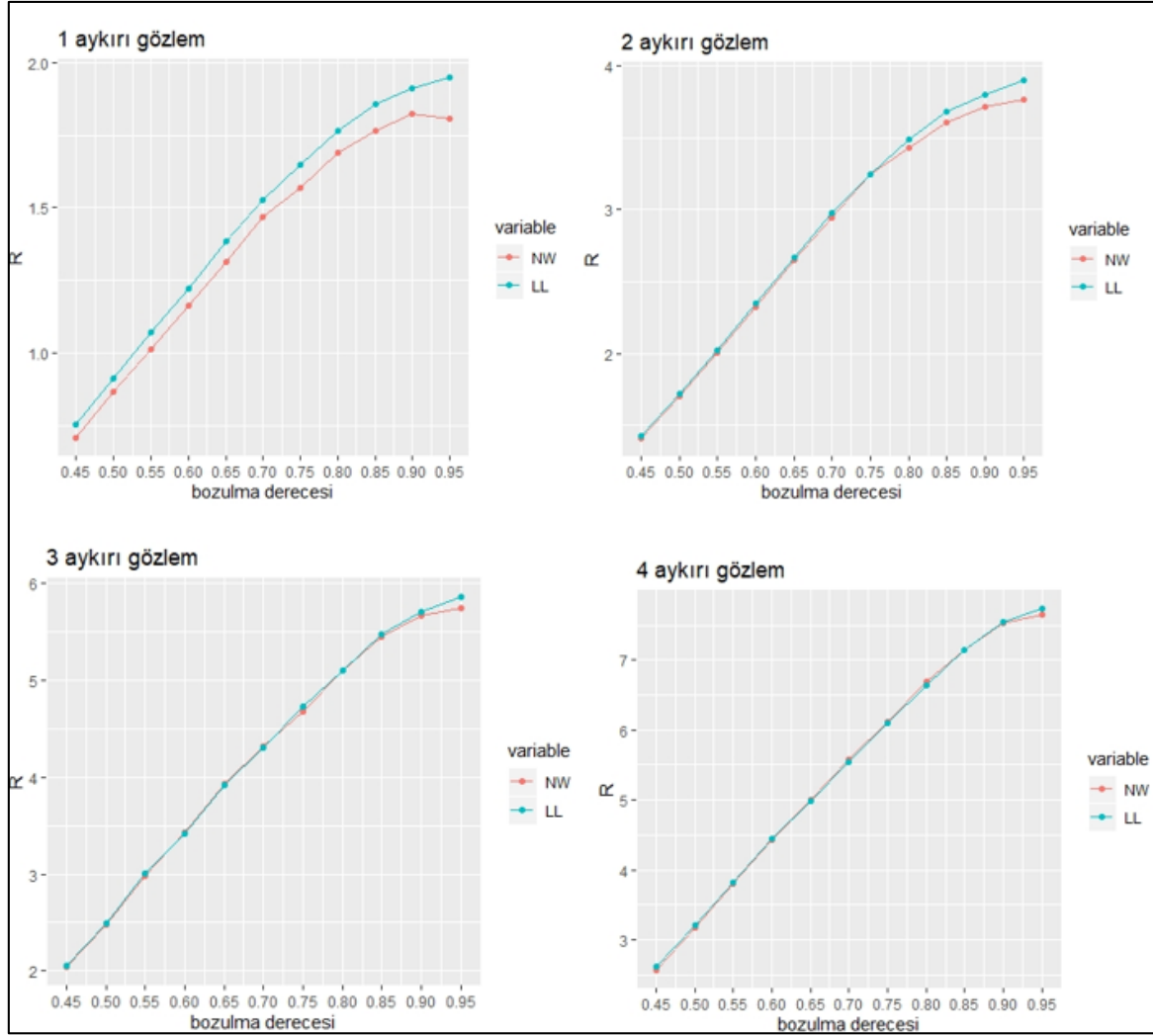
Şekil 5.46. $n=200$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin maskeleme oranları

$n=200$ olduğunda LL yöntemi kullanıldığında maskeleme oranlarının sıfır, NW yönteminin ise sıfıra yakın değerler aldığı gözlenmiştir.



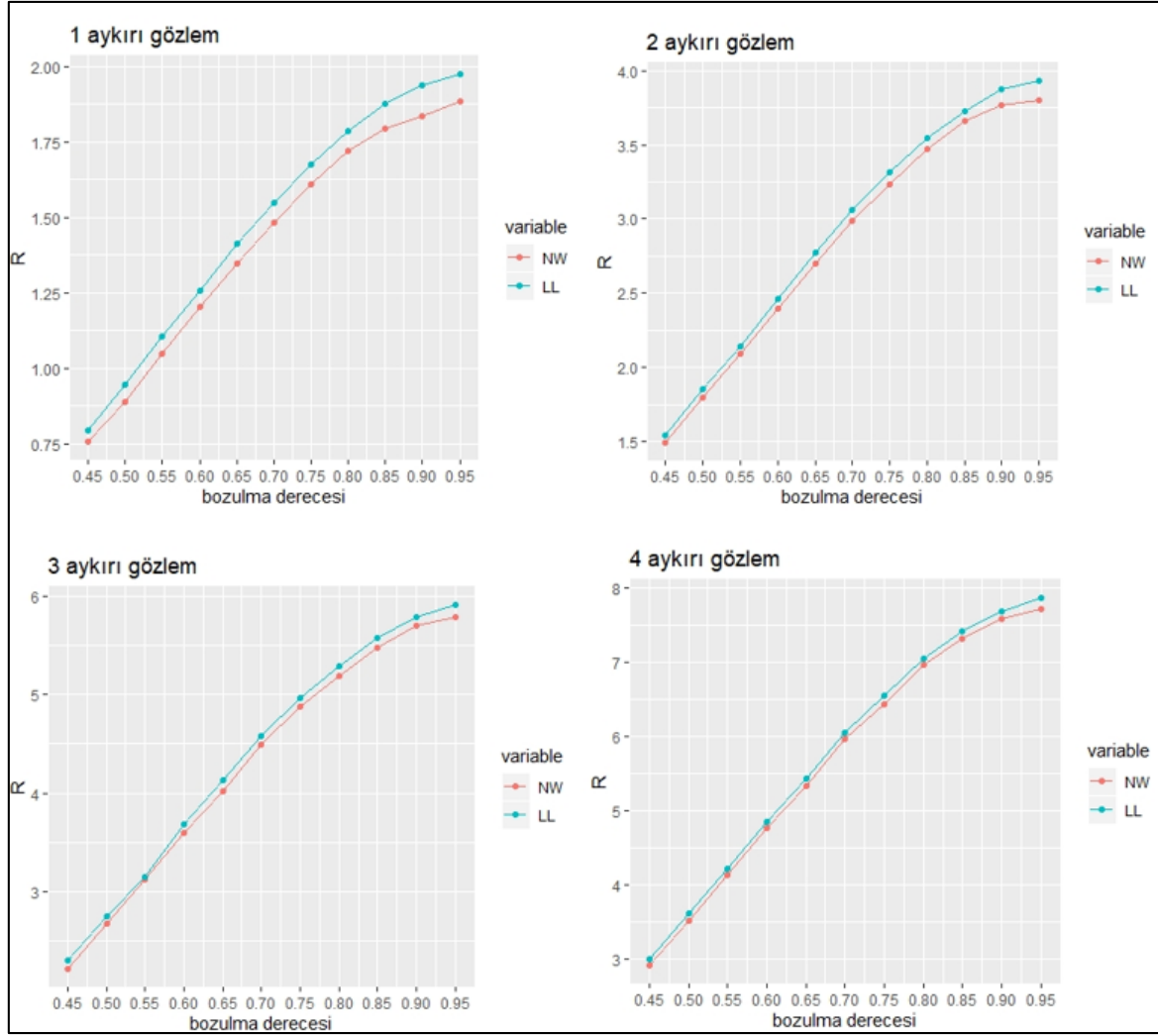
Şekil 5.47. $n=200$ olduğunda farklı bozulma parametreleri için NW ve LL yöntemlerinin sahte aykırı gözlem oranları

$n=200$ için sahte aykırı gözlem oranlarına bakıldığında, her iki yöntem için de sıfıra yakın değerler aldığı gözlenmiştir. Ancak sahte aykırı gözlem varlığının, doğru tespit oranı ve maskeleye ölçüleri gibi belli bir düzene sahip olmadığı açıkça görülmektedir.



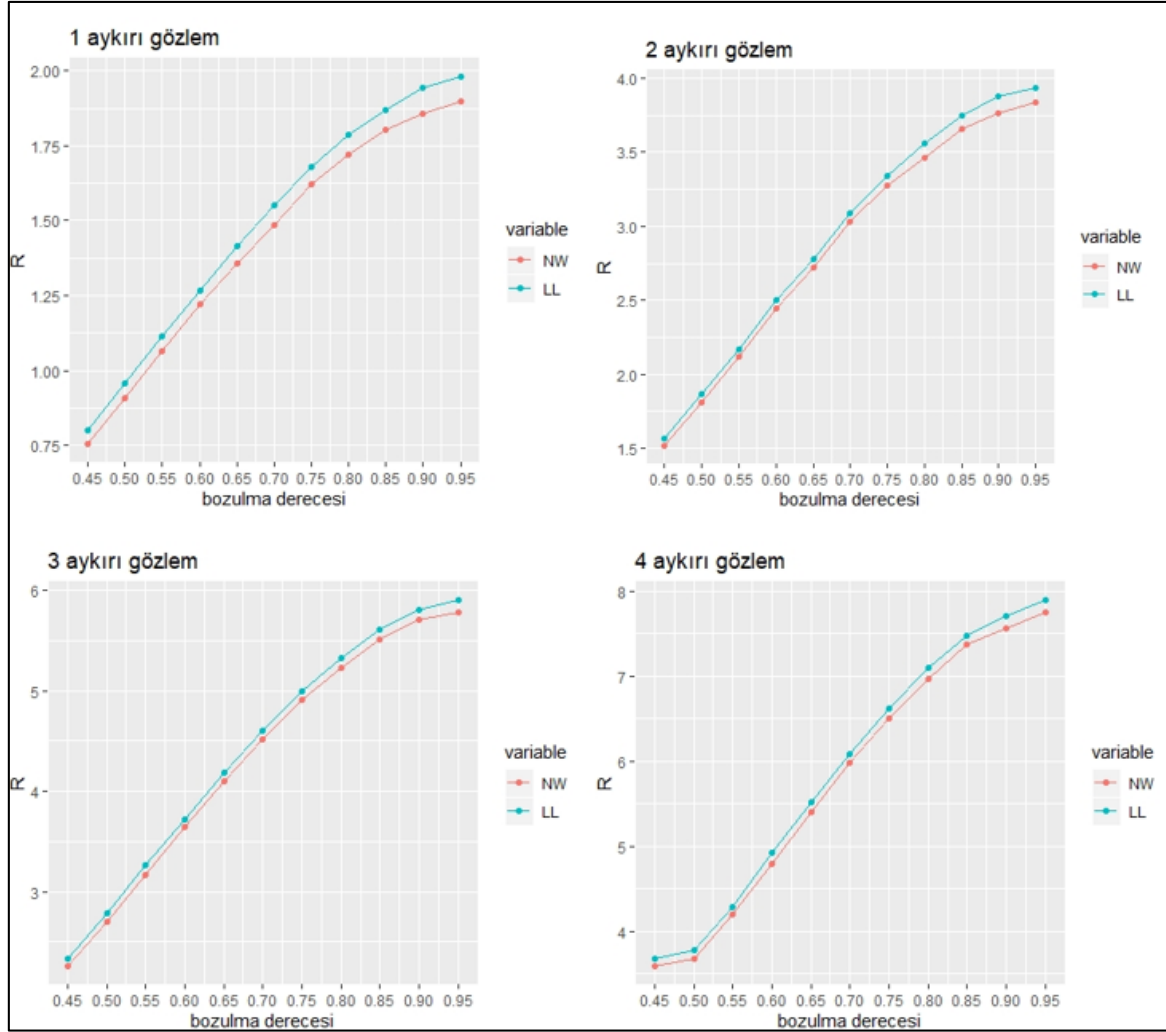
Şekil 5.48. $n=20$ iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri

$n=20$ olduğunda, NW ve LL yöntemleri kullanılarak elde edilen eğrilerin risk değerleri kıyaslandığında, bir ve iki aykırı gözlem varlığında NW eğrisi ile elde edilen risk değerlerinin daha düşük olduğu, diğer durumlarda neredeyse aynı değerler aldığı gözlenmiştir. Veri setinin bozulma derecesi arttıkça risk değerlerinin de arttığı görülmektedir.



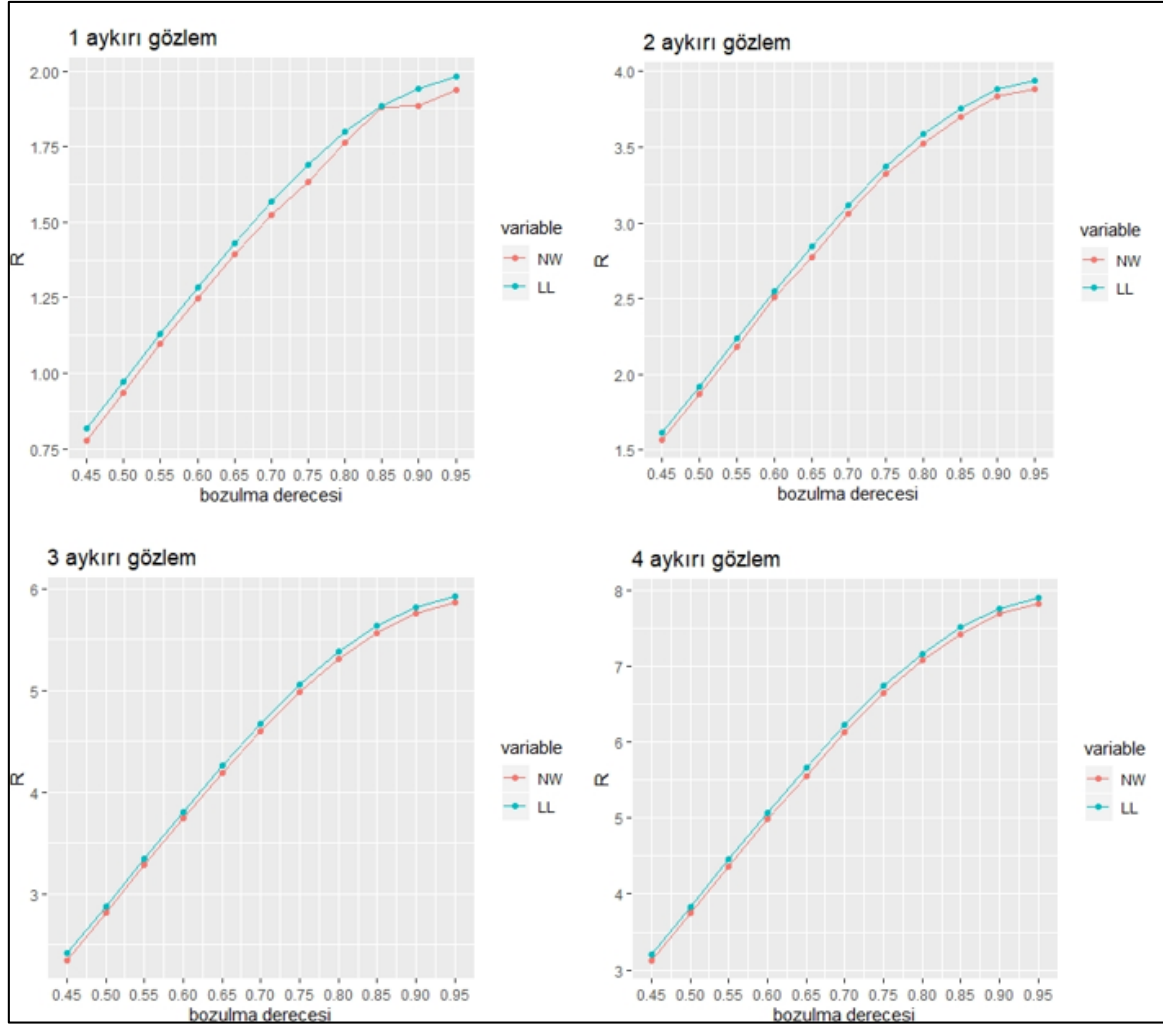
Şekil 5.49. $n=40$ iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri

$n=40$ olduğunda, NW ve LL yöntemleri kullanılarak elde edilen eğrilerin risk değerleri kıyaslandığında, NW yöntemi kullanılarak elde edilen risk değerlerinin daha düşük olduğu gözlenmiştir.



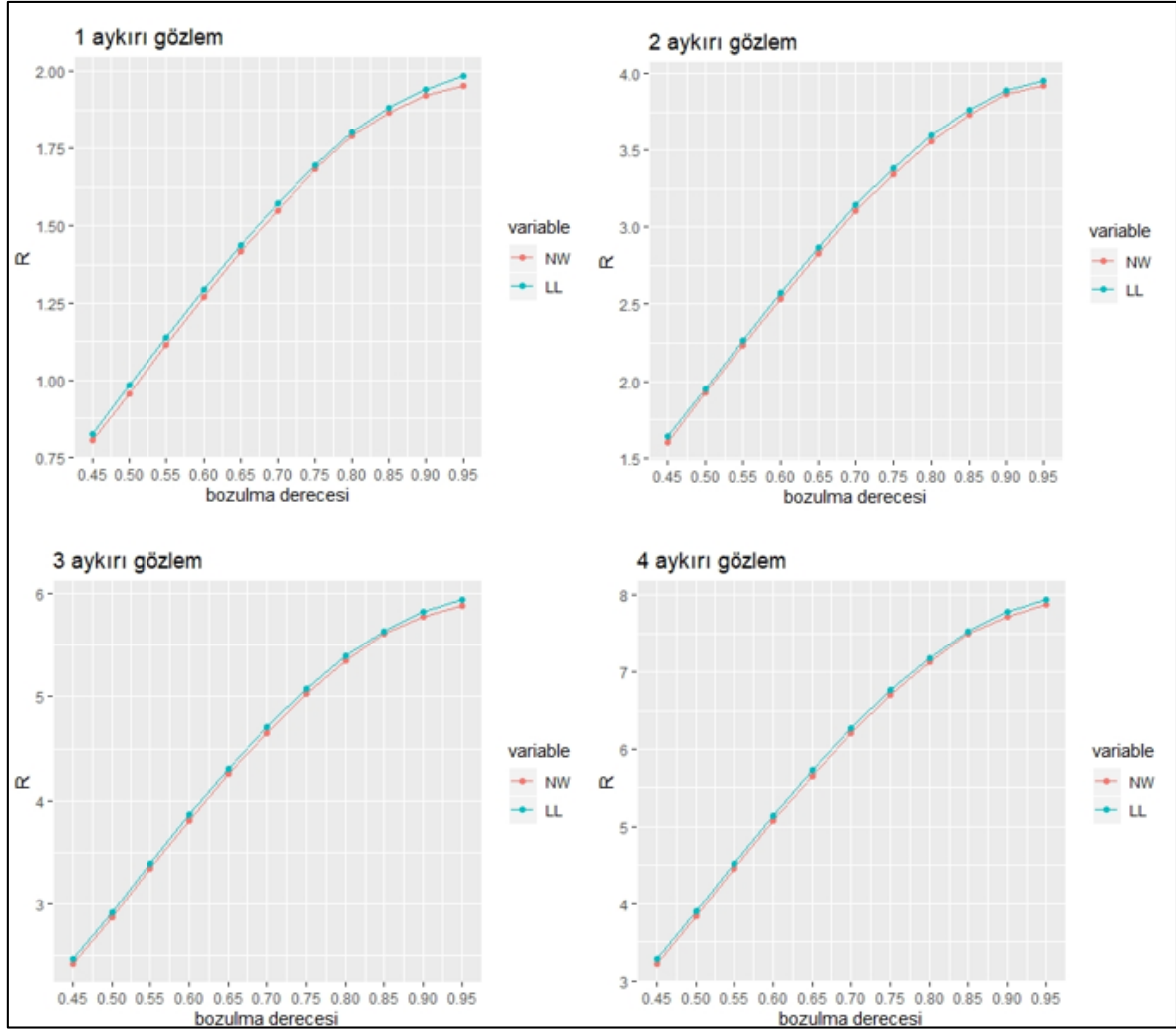
Şekil 5.50. $n=50$ iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri

$n=50$ olduğunda, NW ve LL yöntemleri kullanılarak elde edilen eğrilerin risk değerleri kıyaslandığında, NW yöntemi kullanılarak elde edilen risk değerlerinin daha düşük olduğu gözlenmiştir.



Şekil 5.51. $n=100$ iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri

$n=100$ olduğunda, NW ve LL yöntemleri kullanılarak elde edilen eğrilerin risk değerleri kıyaslandığında, NW yöntemi kullanılarak elde edilen risk değerlerinin daha düşük olduğu gözlenmiştir.



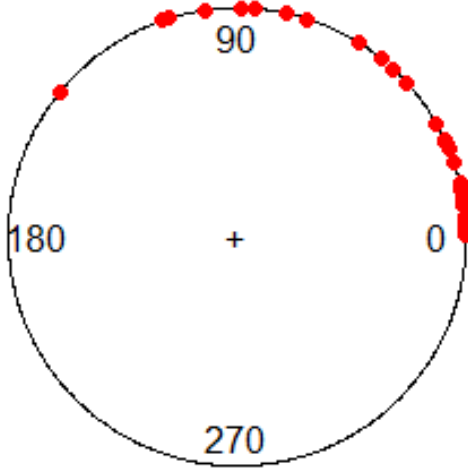
Şekil 5.52. $n=200$ iken farklı bozulma parametreleri için NW ve LL yöntemlerinin risk değerleri

$n=200$ olduğunda, NW ve LL yöntemleri kullanılarak elde edilen eğrilerin risk değerleri kıyaslandığında, NW yöntemi kullanılarak elde edilen risk değerlerinin daha düşük olduğu gözlenmiştir.

5.2.7. Gerçek veri seti uygulaması

Bölüm 4'te kullanılan "Blue Periwinkles" veri seti, daha önce Di Marzio ve diğerleri (2012) ve Oliviera ve diğerleri (2014) tarafından parametrik olmayan regresyon analizinde kullanılmış ancak aykırı gözlem durumu incelenmemiştir. Bu veri setinde uzaklıklar açıklayıcı değişken, hareket yönü ise bağımlı değişkendir. Açıklayıcı değişken doğrusal, bağımlı değişken dairesel olduğundan Doğrusal-Dairesel (Linear-Circular) parametrik olmayan regresyon analizi uygulanmıştır.

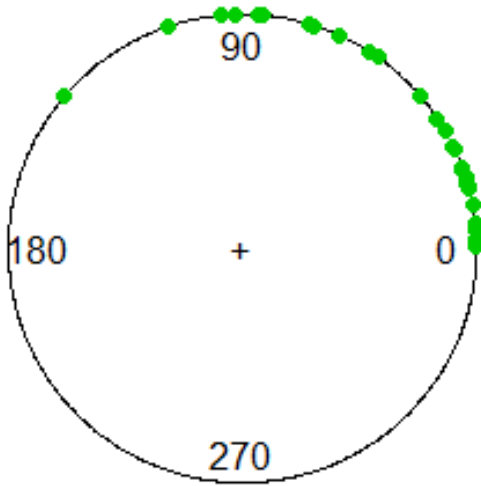
İlk olarak NW yöntemi kullanılarak eğri elde edilmiş ve önerilen yöntem uygulanmıştır. Artıkların grafiği Şekil 5.53.'teki gibi elde edilmiştir.



Şekil 5.53. Mavi deniz salyangozu verisine LL yöntemi uygulandığında artıkların dağılımı

Artıkların dairesel medyanı 25,16237, eşik değeri ise 103.5656 olarak elde edilmiştir. Eşik değerinden büyük olan $mf_{(i)}$ gözlemi, aykırı gözlem olarak adlandırılacaktır. Yukarıdaki şekilden de anlaşılacağı üzere, 30. gözlem artıkların dairesel medyanından oldukça uzaktır. Önerilen yöntem uygulandığında, $mf_{(30)} = 155,5755$ olduğu ve sadece bu gözlemin kesim aykırı gözlemktasından büyük olduğu gözlenmiştir. Dolayısıyla 30. gözlem aykırı gözlem olarak tespit edilmiştir.

LL yöntemi ile eğri uydurulduğunda ise, artıkların grafiği şekil (5.54)'teki gibi elde edilmiştir.



Şekil 5.54. Mavi deniz salyangozu verisine LL yöntemi uygulandığında artıkların dağılımı

Önerilen yöntem uygulandığında artıkların dairesel medyanı 25,72971, eşik değeri ise 103,5419 olarak elde edilmiştir. $mf_{(30)} = 113,8536$ değeri eşik değerinden büyük olduğundan 30. gözlem değeri aykırı gözlem olarak tespit edilmiştir.

6. SONUÇ VE ÖNERİLER

Bu çalışmada dairesel yanıtlarda parametrik olmayan regresyon yöntemlerinden NW ile LL yöntemlerinin aykırı değer varlığında davranışlarının incelenmesi amaçlanmıştır. Bu amaçla ikinci bölümde dairesel verilerle ilgili temel bilgilere ve en yaygın kullanılan dağılımlara yer verilmiştir. Üçüncü bölümde, doğrusal veri setleri için parametrik olmayan regresyon yöntemleri incelenmiştir. Dördüncü bölümde açıklayıcı değişken doğrusal ve yanıt değişkeni dairesel olduğunda basit dairesel regresyon ve parametrik olmayan dairesel regresyon yöntemleri tanıtılmıştır. Beşinci bölümde ise dairesel veri setlerinde aykırı gözlem kavramı incelenmiştir. Tek değişkenli veri setlerinde aykırı gözlem tespiti ile ilgili önerilen yöntemler incelenmiş, aykırı gözlem tespiti ile ilgili yeni bir yöntem önerilmiştir. Ardından dairesel regresyonda aykırı gözlem tespiti ile ilgili yöntemler ile ilgili bilgiler verilmiş ve yeni bir aykırı gözlem tespit yöntemi önerilmiştir.

Dördüncü bölümde yapılan uygulamanın sonucuna göre, aykırı gözlem olmadığı durumda NW düzgünleştiricisinin LL düzgünleştiricisine göre gerçek eğriye daha yakın sonuçlar verdiği gözlenmiştir. Farklı parametrelerde aykırı gözlem yokken yapılan simülasyon sonuçları ile bu bulgular desteklenmektedir. Yapılan simülasyon çalışmalarının sonuçlarına göre, örneklemdaki veri sayısı arttıkça veri setinin fonksiyonel yapısının daha iyi ortaya çıktığı görülmüştür.

Beşinci bölümde ilk olarak tek değişkenli veri setleri için aykırı gözlem tespit yöntemleri incelenmiştir. Burada bahsedilen yöntemlerin hepsi, dağılım varsayımı altında uygulanabilen yöntemlerdir. Veri setinin dağılımıyla ilgili önsel bir bilgi olmadığında veya veri seti sağlanması gereken varsayımları sağlamadığında bu yöntemlerin uygulanmasının mümkün olmadığı görülmektedir. Herhangi bir varsayım veya önsel bilgi gerekmeden aykırı değer tespiti yapabilmek için sağlam (robust) bir yöntem önerilmiştir. Bu yöntem, veri setinde yer alan gözlemlerin dairesel medyana olan uzaklığına dayalıdır. Eşik değerinin belirlenmesi için bootstrap yönteminden yararlanılmıştır. $n=20, 30, 60$ olduğunda, farklı dağılımlardan veri setleri üretilip bu veri setleri belirli oranlarda kirletilmiştir. Monte Carlo simülasyon çalışması ile doğru tespit oranı, maskeleye ve sahte aykırı gözlem durumları incelenmiştir. Elde edilen sonuçlar maddeler halinde verilmiştir.

- Veri setinin bozulma derecesi %70 ve üzerinde olduğunda yüksek doğru tespit oranı, düşük maskeleyme ve sahte aykırı gözlem oranları elde edilmiştir.
- Önerilen yöntemin örneklem boyutu arttıkça ve veri setinin yayılımı azaldıkça yüksek oranda aykırı gözlem tespiti yapabildiği gözlenmiştir.
- Veri setinin çember etrafındaki yayılımı arttıkça aykırı gözlem tespit oranının düşmesi beklenen bir durumdur ancak bu durumda önerilen yöntemde aykırı gözlem tespit oranı, diğer yöntemlere nispeten daha iyi sonuçlar vermektedir.
- Küçük örneklemelerde sahte aykırı gözlem tespit oranı sıfıra yakinken, örneklem boyutu arttıkça çok az bir oranda da olsa sahte aykırı gözleme rastlanmıştır.
- “Kurbağa” veri seti ile gerçek veri seti uygulaması yapılmış ve önerilen yöntemin performansı sözü edilen yöntemlerin performansı ile kıyaslanmıştır. Elde edilen sonuçlara göre önerilen yöntemin diğer yöntemlere kıyasla daha iyi sonuçlar verdiği gözlenmiştir.

Beşinci bölümde sözü edilen dairesel regresyonda aykırı gözlem tespit yöntemlerinin de varsayımlar altında uygulanabildiği görülmektedir. Bugüne kadar, bilğimiz dahilinde, parametrik olmayan dairesel regresyon analizi için uygulanabilen aykırı gözlem tespit yöntemi bulunmamaktadır. Bu nedenle, artıkların dairesel medyana olan uzaklığından yararlanarak aykırı gözlem tespit yöntemi geliştirilmiştir. Rasgele $n=20, 40, 50, 100$ ve 200 birimlik veri setleri üretilerek $\gamma = 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95$ bozulma dereceleri dikkate alınarak veri setine farklı oranlarda (en fazla veri setinin %10'u kadar) aykırı gözlemler eklenmiştir. Eşik değer tespiti için %95. yüzdeler bulunmuştur. MC simülasyonu kullanılarak, 1000 tekrarda elde edilen sonuçlar raporlanmıştır. Aykırı değer tespit oranları, risk fonksiyonu değerleri, maskeleyme(masking) ve sahte aykırı gözlem (swamping) durumları incelenmiştir. Elde edilen sonuçlara göre,

- Örneklem boyutu arttıkça veri setinin fonksiyonel yapısının daha iyi ortaya çıktığı görülmüştür. Bozulma derecesi ve örneklem boyutu arttıkça doğru tespit oranının arttığı ve dolayısıyla maskeleymenin azaldığı görülmüştür.
- Daha önce önerilen RCD_y istatistiği %80 bozulma derecesinde iyi sonuç verirken, önerilen yöntemin %70 bozulma derecesinde dahi iyi sonuçlar verdiği gözlenmiştir.

- Beklendiği üzere veri setindeki aykırı gözlem oranı arttıkça doğru tespit oranının azaldığı görülmüştür.
- Önerilen yöntem artıkların dairesel medyana olan uzaklığına dayalı olduğundan, doğru bant genişliği kullanılarak tahmin değerlerinin gerçeğe yakın sonuçlar vermesi, artıkların sıfır etrafında yoğunlaşmasını ve dolayısıyla aykırı gözlemin rahatlıkla tespit edilmesini sağlayacaktır. Dolayısıyla parametrik olmayan regresyonda veri setindeki yanıt değişkeninin yayılımı, tek değişkenli veri setlerindeki kadar etkili olmamaktadır.
- NW ve LL tahmin edicilerinin aykırı gözlem varlığında performansları, yukarıda sözü edilen simülasyonlar yardımıyla incelenmiştir. Öncelikle her iki yöntemde de bant genişliklerinin aykırı gözlem varlığında arttığı gözlenmiştir.
- Bu yöntemlerin risk fonksiyonlarının aykırı gözlem varlığında değerleri kıyaslandığında, NW tahmin edicisinin daha küçük risk değerlerine sahip olduğu belirlenmiştir.
- Risk fonksiyonunun aykırı değer oranıyla doğru orantılı olduğu ve bozulma derecesinden etkilenmediği gözlenmiştir.
- Veri setinin yayılımı arttıkça sahte aykırı gözleme rastlama oranının yükseldiği belirlenmiştir.
- Yöntemlerin risk fonksiyonlarının davranışlarının incelenmesi amacıyla, fit edilen eğrilerden doğru tespit edilen aykırı gözlemler çıkarılmış ve eğriler yeniden fit edilmiştir. Elde edilen sonuçlara göre, LL eğrilerinin aykırı gözlemler çıkarıldıktan sonra risk değerlerinde NW yöntemine göre daha fazla düşüş olduğu gözlenmiştir. Bu davranış, LL yönteminin aykırı gözlem varlığından daha fazla etkilendiğini göstermektedir.
- Simülasyon çalışmalarından sonra önerilen yöntem, gerçek bir veri setine uygulanmıştır. “Mavi deniz salyangozu” veri setine önerilen yöntem uygulandığında, 30. gözlem değerinin aykırı gözlem olduğu gözlenmiştir.

Önerilen yöntemin dağılımı bilinmeyen veri setlerinde aykırı gözlem tespitinde literatürdeki veri setlerine kıyasla yüksek performans sergilediği gözlenmiştir. NW tahmin edicisinin aykırı gözlem varlığında LL tahmin edicisine göre daha sağlam bir davranış sergilediği görülmüştür. Bu nedenle aykırı gözlem varlığında NW tahmin edicisinin kullanılması önerilmektedir.

KAYNAKLAR

- Abuzaid, A. H. (2010). *Some problems of outliers in circular data*, PhD Thesis, University of Malaya, Kuala Lumpur .
- Abuzaid, A. H., Hussin, A. G., and Mohamed, I. B. (2008). Identifying single outlier in linear circular regression model based on circular distance. *Journal of Applied Probability and Statistics*, 3(1), 107-117.
- Abuzaid, A. H., Hussin, A. G., and Mohamed, I. B. (2013). Detection of outliers in simple circular regression models using the mean circular error statistic. *Journal of Statistical Computation and Simulation*, 83(2), 269-277.
- Abuzaid, A. H., Hussin, A. G., Rambli, A., and Mohamed, I. (2012). Statistics for a new test of discordance in circular data. *Communications in Statistics-Simulation and Computation*, 41(10), 1882-1890.
- Abuzaid, A. H., Mohamed, I. B., and Hussin, A. G. (2009). A new test of discordancy in circular data. *Communications in Statistics-Simulation and Computation*, 38(4), 682-691.
- Abuzaid, A. H., Mohamed, I. B., and Hussin, A. G. (2012). Boxplot for circular variables. *Computational Statistics*, 27(3), 381-392.
- Abuzaid, A., Mohamed, I., Hussin, A. G., & Rambli, A. (2011). COVRATIO statistic for simple circular regression model. *Chiang Mai J. Sci*, 38(3), 321-330.
- Agostinelli, C., and Lund, U. (2011). Circular: Circular Statistics. Web: <https://cran.rproject>.
- Barnett, V. (1978). The study of outliers: purpose and model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 27(3), 242-250.
- Batschelet, E. (1981). *Circular statistics in biology*. New York: Academic Press , 388.
- Beckman, R. J., and Cook, R. D. (1983). Outlier..... s. *Technometrics*, 25(2), 119-149.
- Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2), 353-360.
- Bowman, A. W., and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations* (18). OUP Oxford.
- Chaloner, K., and Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika*, 75(4), 651-659.
- Chaudhuri, P., and Marron, J. S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447), 807-823.
- Collett, D. (1980). Outliers in circular data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(1), 50-57.

- De Sá, J. P. M. (2007). *Applied statistics using SPSS, Statistica, MatLab and R*. Verlag Berlin Heidelberg: Springer Science and Business Media, 375-399.
- Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statistics and Probability Letters*, 79(19), 2066-2075.
- Di Marzio, M., Panzera, A., and Taylor, C. C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, 141(6), 2156-2173.
- Di Marzio, M., Panzera, A., and Taylor, C. C. (2013). Non-parametric Regression for Circular Responses. *Scandinavian Journal of Statistics*, 40(2), 238-255.
- Dole, D. (1999). Cosmo: A constrained scatterplot smoother for estimating convex, monotonic transformations. *Journal of Business & Economic Statistics*, 17(4), 444-455.
- Ducharme, G. R., and Milasevic, P. (1987). Some asymptotic properties of the circular median. *Communications in Statistics-Theory and Methods*, 16(3), 659-664.
- Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.
- Fisher, N. I., and Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, 665-677.
- Fisher, N. I., and Powell, C. M. (1989). Statistical analysis of two-dimensional palaeocurrent data: Methods and examples. *Journal of the Geological Society of Australia*, 36(1), 91-107.
- Freund, J. E. (1962). *Mathematical Statistics*, N. J :Englewood Cliffs.
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246-263.
- García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *Journal of Multivariate Analysis*, 121, 152-175.
- Gasser, T., Muller, H. G., and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 238-252.
- Hall, P., and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statistics & Probability Letters*, 6(2), 109-115.
- Härdle, W. (1990). *Applied nonparametric regression* (No. 19). Cambridge: Cambridge university press.
- Horová, I., and Zelinka, J. (2007). Contribution to the bandwidth choice for kernel density estimates. *Computational Statistics*, 22(1), 31-47.

- Ivanka, H., Jan, K., and Jiri, Z. (2012). *Kernel Smoothing in MATLAB: theory and practice of kernel smoothing*. World scientific.
- Jammalamadaka, S. R., and Sengupta, A. (2001). *Topics in circular statistics*. Singapore: World Scientific.
- Jones, M. C., Davies, S. J., and Park, B. U. (1994). Versions of kernel-type regression estimators. *Journal of the American Statistical Association*, 89(427), 825-832.
- Kleiner, B., and Graedel, T. E. (1980). Exploratory data analysis in the geophysical sciences. *Reviews of Geophysics*, 18(3), 699-717.
- Kong, D., Bondell, H., and Shen, W. (2018, March). *Outlier Detection and Robust Estimation in Nonparametric Regression*. In International Conference on Artificial Intelligence and Statistics, 208-216.
- Lenth, R. V. (1981). Robust measures of location for directional data. *Technometrics*, 23(1), 77-81.
- Lund, U., and Agostinelli, C. (2007). Circstats: circular statistics, from “topics in circular statistics”(2001). S-plus original by Lund, U. R port by Agostinelli, C. R package version 0.2-3.
- Mahmood, E. A., Rana, S., Midi, H., and Hussin, A. G. (2017). Detection of Outliers in Univariate Circular Data using Robust Circular Distance. *Journal of Modern Applied Statistical Methods*, 16(2), 22.
- Mardia, K. V. (1972). *Statistics of Directional Data*, London and New York: Academic Press.
- Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(3), 349-371.
- Mardia, K. V., and Jupp, P. E. (2009). *Directional statistics* (494). New York: John Wiley & Sons.
- Martinez, W. L., and Martinez, A. R. (2007). *Computational statistics handbook with MATLAB*. London: Chapman and Hall/CRC.
- Mohamed, I. B., Rambli, A. D. Z. H. A. R., Khaliddin, N. U. R. L. I. Z. A., and Ibrahim, A. I. N. (2016). A new discordancy test in circular data using spacings theory. *Communications in Statistics-Simulation and Computation*, 45(8), 2904-2916.
- Muller, H. G., and Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 61-76.
- Pérez, M. O. (2013). *Nonparametric circular methods for density and regression*, PhD Thesis, Universidade de Santiago de Compostela, A Coruña.

- Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Computational Statistics & Data Analysis*, 56(12), 3898-3908.
- Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2013). Nonparametric circular methods for exploring environmental data. *Environmental and ecological statistics*, 20(1), 1-17.
- Oliveira, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2014). NPCirc: An R package for nonparametric circular methods. *Journal of Statistical Software*, 61(9), 1-26.
- Otieno, B. S., and Anderson-Cook, C. M. (2003). Hodges-Lehmann estimator of preferred direction for circular. *Virginia Tech Department of Statistics Technical Report*, Virginia, 03-3.
- Otieno, B., and Anderson-Cook, C. M. (2005). Effect of position of an outlier on the influence curve of the measures of preferred direction for circular data. *Journal of Modern Applied Statistical Methods*, 4(1), 9.
- Otieno, B. S., and Anderson-Cook, C. M. (2006). Measures of preferred direction for environmental and ecological circular data. *Environmental and Ecological Statistics*, 13(3), 311-324.
- Pewsey, A., Neuhaus, M., and Ruxton, G. D. (2013). *Circular statistics in R*. New York: Oxford University Press.
- Otieno, B., and Anderson-Cook, C. M. (2003). A more efficient way of obtaining a unique median estimate for circular data. *Journal of Modern Applied Statistical Methods*, 2(1), 15.
- Qin, X., Zhang, J. S., and Yan, X. D. (2011). A nonparametric circular-linear multivariate regression model with a rule-of-thumb bandwidth selector. *Computers & Mathematics with Applications*, 62(8), 3048-3055.
- Rambli, A., Mohamed, I., Abuzaid, A. H., and Hussin, A. G. (2010). *Identification of influential observations in circular regression model*. In Proceedings of the Regional Conference on Statistical Sciences 2010 (RCSS'10), 195-203.
- Rao, J. S. (1969). *Some contributions to the analysis of circular data*, PhD Thesis, Indian Statistical Institute, Kolkata.
- Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400), 1131-1146.
- Sheather, S. J., and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3), 683-690.
- Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. New York: Routledge.

- Simonoff, J. S. (2012). *Smoothing methods in statistics*. New York: Springer Science & Business Media.
- Takezawa, K. (2005). *Introduction to nonparametric regression* (606). New Jersey: John Wiley & Sons.
- Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Computational Statistics & Data Analysis*, 52(7), 3493-3500.
- Terrell, G. R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association*, 85(410), 470-477.
- Terrell, G. R., and Scott, D. W. (1985). Oversmoothed nonparametric density estimates. *Journal of the American Statistical Association*, 80(389), 209-214.
- Terrell, G. R and Scott, D. W. (1992). Variable kernel density estimation. *The Annals of Statistics*, 20(3), 1236-1265.
- Wand, M. P., and Jones, M. C. (1994). *Kernel smoothing*. New York: Chapman and Hall/CRC.
- Wegkamp, M. (2003). Model selection in nonparametric regression. *The Annals of Statistics*, 31(1), 252-273.
- Wehrly, T. E., and Shine, E. P. (1981). Influence curves of estimators for directional data. *Biometrika*, 68(1), 334-335.
- Yang, Y. (1999). Model selection for nonparametric regression. *Statistica Sinica*, 9(2), 475-499.
- Yolsal, H. (2017). *Parametrik Olmayan Yoğunluk Tahmincileri ve Regresyon Analizi*, Ankara: Detay Yayıncılık.
- Zhao, G., and Ma, Y. (2016). Robust nonparametric kernel regression estimator. *Statistics & Probability Letters*, 116, 72-79.

EKLER

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,45	1	0,9494	0,0505	0,2192	0,0067	0,023	0,7064	0,2097	0,9969	0,003	0,055	0,0007	0,007	0,7548	0,0959
	2	0,9244	0,0755	0,2041	0,0025	0,0143	1,4091	0,3207	0,9638	0,0361	0,1487	0,0021	0,011	1,4271	0,2655
	3	0,8012	0,1987	0,2326	0,0026	0,0151	2,0422	0,4779	0,8619	0,1380	0,2003	0,0011	0,0081	2,057	0,3683
	4	0,5969	0,4030	0,2171	0,0033	0,0157	2,56	0,6382	0,6586	0,3413	0,216	0,0013	0,01	2,6145	0,4483
0,50	1	0,9785	0,0214	0,1448	0,0037	0,018	0,8647	0,2235	0,9989	0,001	0,0319	0,0002	0,0033	0,9099	0,1048
	2	0,95	0,05	0,1751	0,0019	0,0118	1,7066	0,3785	0,9782	0,0217	0,1097	0,0012	0,0083	1,7208	0,3209
	3	0,8353	0,1646	0,221	0,002	0,0127	2,4745	0,5539	0,8794	0,1205	0,1943	0,0013	0,0091	2,4995	0,4233
	4	0,6274	0,3725	0,2155	0,0022	0,014	3,1719	0,7345	0,7019	0,298	0,2231	0,0008	0,0071	3,2083	0,4833
0,55	1	0,9898	0,0101	0,1	0,002	0,012	1,0124	0,2664	1	0	0	0,0001	0,0029	1,0712	0,1011
	2	0,9503	0,0496	0,1722	0,0017	0,0127	2,0064	0,4495	0,979	0,021	0,1347	0,0013	0,0091	2,0227	0,3912
	3	0,8674	0,1325	0,2010	0,0009	0,0074	2,9853	0,5397	0,9037	0,0962	0,1748	0,0004	0,005	3,006	0,3658
	4	0,6777	0,3222	0,229	0,0012	0,0093	3,7957	0,7293	0,7158	0,2841	0,2208	0,0004	0,0062	3,8095	0,5247

EK-1. (devam) n=20 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,60	1	0,9948	0,0051	0,0716	0,0014	0,0108	1,1629	0,2979	1	0	0	0,0001	0,0023	1,2235	0,1291
	2	0,9657	0,0342	0,1307	0,0005	0,0061	2,3227	0,54667	0,987	0,0129	0,0793	0,0003	0,0045	2,3505	0,3452
	3	0,8811	0,1188	0,1918	0,0004	0,0066	3,426	0,6299	0,9103	0,0896	0,1684	0,0002	0,0036	3,4252	0,487
	4	0,7019	0,298	0,2246	0,0007	0,0075	4,4243	0,8473	0,7337	0,2662	0,2217	0,0006	0,0071	4,4379	0,6049
0,65	1	0,9956	0,0043	0,0854	0,0011	0,0092	1,3130	0,3277	1	0	0	0	0	1,3836	0,1065
	2	0,9729	0,027	0,1131	0,00006	0,0018	2,6477	0,4480	0,9906	0,0093	0,0678	0	0	2,6715	0,3449
	3	0,9028	0,0971	0,1751	0,0002	0,0036	3,9328	0,6075	0,9306	0,0693	0,1541	0,0001	0,0029	3,9173	0,4942
	4	0,7291	0,2708	0,2242	0,0003	0,0046	4,9967	0,9475	0,7574	0,2425	0,2247	0,0003	0,0046	4,9792	0,7304
0,70	1	0,9989	0,001	0,0323	0,0004	0,0047	1,4682	0,3163	1	0	0	0	0	1,5260	0,1138
	2	0,9792	0,0207	0,0997	0	0	2,9441	0,4593	0,9938	0,0061	0,0551	0	0	2,977	0,3449
	3	0,9206	0,0793	0,1653	0,0001	0,0029	4,322	0,7135	0,9326	0,0673	0,151	0,0002	0,0036	4,3053	0,6345
	4	0,7515	0,2484	0,2189	0,000009	0,0024	5,5862	1,0452	0,7680	0,2319	0,2212	0,0001	0,0034	5,5507	0,8348

EK-1. (devam) n=20 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,75	1	1	0	0	0,0003	0,0041	1,5701	0,3911	1	0	0	0	0	1,6487	0,1625
	2	0,9840	0,0159	0,0878	0	0	3,251	0,4688	0,9965	0,0034	0,0411	0	0	3,2491	0,3699
	3	0,9169	0,083	0,1801	0,000007	0,0021	4,6771	0,8817	0,9378	0,0621	0,453	0	0	4,7303	0,6973
	4	0,772	0,2227	0,2281	0,0005	0,0082	6,1222	1,1070	0,841	0,2158	0,2213	0,0002	0,0041	6,097	0,9262
0,80	1	1	0	0	0,0001	0,0023	1,692	0,4338	1	0	0	0	0	1,7675	0,1485
	2	0,9844	0,056	0,0944	0	0	3,4277	0,5256	0,9970	0,003	0,0519	0	0	3,487	0,3808
	3	0,9405	0,0594	0,159	0	0	5,1025	0,7873	0,9519	0,0512	0,136	0	0	5,1038	0,7284
	4	0,8135	0,1864	0,2103	0	0	6,6967	1,0799	0,816	0,1839	0,2131	0,00008	0,0023	6,6474	0,9339
0,85	1	1	0	0	0,0001	0,0029	1,764	0,4442	1	0	0	0	0	1,8564	0,1804
	2	0,9874	0,0126	0,1038	0	0	3,6045	0,5723	0,9975	0,0028	0,0689	0	0	3,6870	0,3419
	3	0,9468	0,0531	0,1522	0	0	5,4558	0,6437	0,9488	0,0511	0,148	0,0001	0,0028	5,4696	0,5525
	4	0,8349	0,165	0,2113	0	0	7,142	1,0831	0,8217	0,1782	0,2181	0,0003	0,0045	7,1419	0,9023

EK-1. (devam) n=20 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,90	1	1	0	0	0,0005	0,0016	1,8247	0,4746	1	0	0	0	0	1,9124	0,2495
	2	0,9880	0,012	0,1095	0	0	3,7151	0,6148	0,9987	0,0023	0,0793	0	0	3,7981	0,4641
	3	0,9474	0,526	0,1616	0	0	5,6664	0,6593	0,956	0,0439	0,1409	0	0	5,7127	0,5369
	4	0,8513	0,1486	0,2084	0	0	7,5375	0,8191	0,8594	0,1405	0,213	0,0024	0,0038	7,5461	0,7634
0,95	1	1	0	0	0,00005	0,0016	1,8063	0,594	1	0	0	0	0	1,95	0,2877
	2	0,9885	0,0115	0,0808	0	0	3,77	0,7008	0,9993	0,0007	0,0529	0	0	3,9019	0,4275
	3	0,9567	0,0433	0,1784	0	0	5,751	0,7601	0,957	0,043	0,1523	0,000006	0,0019	5,862	0,5435
	4	0,8621	0,178	0,2167	0	0	7,6559	1,0327	0,8776	0,1223	0,2108	0,000007	0,0021	7,7409	0,8832

EK-2. n=40 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,45	1	0,985	0,015	0,14	0,0036	0,0119	0,7574	0,1462	1	0	0	0,0024	0,0081	0,7966	0,0423
	2	0,983	0,017	0,1011	0,0011	0,0064	1,4909	0,2421	0,993	0,007	0,0703	0,0006	0,0049	1,537	0,1954
	3	0,96	0,04	0,1289	0,0007	0,0058	2,2083	0,331	0,99	0,01	0,0768	0,0002	0,0031	2,2945	0,198
	4	0,9149	0,085	0,1524	0,0005	0,0039	2,9146	0,4127	0,9549	0,045	0,1090	0,000008	0,0015	3,002	0,2836
0,50	1	0,986	0,014	0,1175	0,0032	0,0110	0,8909	0,2003	1	0	0	0,0017	0,0067	0,9446	0,065
	2	0,985	0,015	0,0882	0,0008	0,0051	1,7983	0,2733	0,996	0,004	0,0445	0,0002	0,0026	1,8496	0,1977
	3	0,9679	0,032	0,114	0,0003	0,0029	2,6743	0,3709	0,9909	0,009	0,0599	0,0001	0,0019	2,7493	0,2205
	4	0,9259	0,074	0,1418	0,0002	0,0027	3,5253	0,4126	0,9609	0,039	0,1067	0,0001	0,0023	3,6162	0,2851
0,55	1	0,99	0,01	0,0995	0,0025	0,0101	1,0475	0,2188	1	0	0	0,0014	0,0058	1,1046	0,057
	2	0,987	0,013	0,094	0,0007	0,0062	2,0912	0,346	0,9965	0,0035	0,0472	0,0003	0,0035	2,1438	0,2870
	3	0,9715	0,0284	0,1044	0,0003	0,0034	3,1174	0,4407	0,9929	0,0071	0,0752	0,0002	0,0029	3,1969	0,3352
	4	0,9294	0,0705	0,1498	0,0002	0,0024	4,1269	0,4941	0,9696	0,0304	0,1125	0,0001	0,0017	4,2221	0,3451

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,60	1	0,999	0,001	0,1089	0,0021	0,0094	1,2035	0,2446	1	0	0	0,0013	0,006	1,2592	0,0847
	2	0,9915	0,0085	0,0646	0,0003	0,0029	2,3957	0,3679	0,9975	0,0025	0,0352	0,0001	0,0016	2,4641	0,2723
	3	0,9799	0,02	0,0873	0,0001	0,0019	3,5971	0,4348	0,9963	0,0036	0,0407	0,00005	0,0012	3,682	0,2883
	4	0,9393	0,0606	0,1453	0,0002	0,0026	4,7662	0,5478	0,9707	0,0293	0,092	0,0001	0,0021	4,8604	0,3415
0,65	1	0,9991	0,0009	0,0631	0,0016	0,0079	1,3466	0,269	1	0	0	0,0008	0,0045	1,4111	0,051
	2	0,9919	0,0081	0,0544	0,00002	0,0008	2,703	0,3865	0,9984	0,0016	0,0273	0,00007	0,0018	2,7757	0,2347
	3	0,9819	0,018	0,0878	0,0001	0,0019	4,0207	0,5538	0,9969	0,0031	0,0481	0,00005	0,0012	4,1309	0,3402
	4	0,9397	0,0603	0,1393	0,00005	0,0012	5,33	0,6767	0,9709	0,0291	0,0992	0,0001	0,0017	5,4453	0,4496
0,70	1	0,9994	0,0006	0,0705	0,0011	0,0061	1,4815	0,3003	1	0	0	0,0005	0,0038	1,55	0,0641
	2	0,9924	0,0076	0,0647	0,00002	0,0008	2,9884	0,4338	0,9984	0,0016	0,0272	0,00001	0,0008	3,0617	0,0231
	3	0,9852	0,0148	0,0633	0,00005	0,0012	4,4993	0,4466	0,9976	0,0024	0,0365	0,00002	0,0008	4,5807	0,2929
	4	0,9574	0,0425	0,1113	0,00005	0,0012	5,9676	0,6063	0,9832	0,0167	0,0701	0,00002	0,0008	6,0567	0,4102

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,75	1	0,9995	0,0005	0,0995	0,0015	0,0086	1,6108	0,3107	1	0	0	0,0004	0,0033	1,675	0,0754
	2	0,9929	0,0071	0,0497	0,00002	0,0008	3,2359	0,4205	0,9984	0,0015	0,0273	0,00002	0,0008	3,3115	0,2505
	3	0,9864	0,0135	0,0691	0	0	4,8767	0,4634	0,9986	0,0014	0,0366	0	0	4,9666	0,256
	4	0,9607	0,0392	0,1123	0,00002	0,0008	6,4461	0,7409	0,9834	0,0165	0,0690	0	0	6,5518	0,5229
0,80	1	0,9995	0,0005	0,0705	0,0008	0,0061	1,7186	0,3436	1	0	0	0,0002	0,0026	1,784	0,0934
	2	0,9935	0,0065	0,0566	0,00002	0,0008	3,4728	0,4424	0,9985	0,0015	0,0273	0	0	3,5447	0,2273
	3	0,9875	0,0125	0,071	0	0	5,1921	0,5868	0,9993	0,0008	0,0348	0	0	5,293	0,362
	4	0,9704	0,0295	0,1034	0	0	6,977	0,5886	0,9906	0,0094	0,0548	0	0	7,0629	0,3802
0,85	1	0,9996	0,0004	0,0631	0,0006	0,0055	1,7942	0,3822	1	0	0	0,00007	0,0014	1,8762	0,0645
	2	0,9939	0,0061	0,0544	0,00002	0,0008	3,6562	0,4492	1	0	0	0	0	3,7297	0,267
	3	0,9896	0,0103	0,0634	0	0	5,484	0,5259	0,9996	0,0004	0,0364	0	0	5,5824	0,3005
	4	0,976	0,0239	0,0907	0	0	7,3288	0,6094	0,9908	0,0092	0,0582	0	0	7,4295	0,4164

EK-2. (devam) n=40 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,90	1	0,9998	0,0002	0,0446	0,0004	0,004	1,8354	0,4552	1	0	0	0,00005	0,0011	1,9383	0,1187
	2	0,997	0,003	0,0386	0	0	3,7683	0,5106	1	0	0	0	0	3,8722	0,2025
	3	0,9933	0,0066	0,049	0	0	5,7058	0,4944	0,9999	0,0001	0,0257	0	0	5,7966	0,2808
	4	0,9783	0,0216	0,0954	0	0	7,5918	0,6664	0,9914	0,0086	0,0554	0,00002	0,0008	7,6992	0,5021
0,95	1	1	0	0	0,0001	0,0021	1,8849	0,45	1	0	0	0,0001	0,0032	1,9742	0,1574
	2	0,9984	0,0015	0,0273	0	0	3,8037	0,6164	1	0	0	0	0,0316	3,9315	0,3064
	3	0,9959	0,004	0,0392	0	0	5,7892	0,6974	0,9999	0,0001	0,0257	0	0	5,9167	0,3554
	4	0,9849	0,015	0,0728	0	0	7,7303	0,7695	0,9922	0,0078	0,0697	0	0	7,8752	0,473

EK-3. n=50 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,45	1	0,988	0,012	0,1678	0,0053	0,0143	0,7545	0,1563	1	0	0	0,0032	0,0084	0,8013	0,0489
	2	0,9855	0,0145	0,0897	0,0009	0,0051	1,5148	0,2231	0,996	0,004	0,0498	0,0005	0,0042	1,5613	0,1704
	3	0,98	0,02	0,0957	0,0006	0,0054	2,2562	0,3037	0,996	0,004	0,0445	0,0002	0,0025	2,3329	0,1867
	4	0,9118	0,0882	0,1418	0,0003	0,0032	3,6728	0,4389	0,9458	0,0542	0,1084	0,00004	0,0009	3,7853	0,2895
	5	0,9041	0,0958	0,142	0,0002	0,0025	3,6815	0,4549	0,9387	0,0612	0,1125	0,00004	0,0009	3,7835	0,2837
0,50	1	0,99	0,01	0,0995	0,0041	0,0112	0,9083	0,1764	1	0	0	0,0029	0,0081	0,9557	0,0446
	2	0,9865	0,0135	0,0926	0,0009	0,0061	1,8116	0,270	0,9965	0,0035	0,0417	0,0005	0,0036	1,8639	0,2056
	3	0,985	0,015	0,0984	0,0005	0,0042	2,6975	0,3638	0,9966	0,0034	0,0535	0,0001	0,0023	2,7887	0,2361
	4	0,965	0,035	0,1085	0,0002	0,0024	3,5907	0,42	0,9885	0,0115	0,0651	0,0001	0,0018	3,6806	0,2819
	5	0,9111	0,0888	0,1419	0,0003	0,003	4,4238	0,5141	0,9509	0,0490	0,0957	0,00002	0,0007	4,5524	0,259
0,55	1	0,993	0,007	0,0891	0,003	0,0092	1,0658	0,1948	1	0	0	0,0022	0,007	1,1127	0,0612
	2	0,9925	0,0075	0,0647	0,0004	0,0032	2,1188	0,3198	0,9975	0,0025	0,0352	0,0003	0,0026	2,168	0,2616
	3	0,987	0,013	0,0695	0,0002	0,0028	3,1689	0,3671	0,9973	0,0026	0,0364	0,0001	0,0021	3,2612	0,2253
	4	0,9708	0,0291	0,0972	0,0002	0,0023	4,2011	0,4622	0,9871	0,0128	0,0761	0,0003	0,003	4,2935	0,3624
	5	0,9212	0,0787	0,1292	0,00006	0,0012	5,2161	0,52	0,9508	0,0491	0,1073	0,00002	0,0007	5,3232	0,3636

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,60	1	0,993	0,007	0,0834	0,0021	0,0087	1,2208	0,2042	1	0	0	0,0016	0,0058	1,2685	0,0493
	2	0,9926	0,0074	0,0666	0,0003	0,0034	2,4415	0,3021	0,9978	0,0022	0,0456	0,00008	0,0013	2,5029	0,1607
	3	0,9873	0,0126	0,0831	0,0003	0,0031	3,6415	0,4159	0,9976	0,0023	0,0326	0,0001	0,0015	3,7238	0,2562
	4	0,9716	0,0283	0,0896	0,00004	0,0009	4,7986	0,5383	0,9912	0,0087	0,0523	0,00002	0,0006	4,9181	0,3201
	5	0,9334	0,0665	0,1293	0,00008	0,0017	5,9958	0,5596	0,9592	0,0407	0,0978	0,00008	0,0017	6,1064	0,3746
0,65	1	0,994	0,006	0,1133	0,0026	0,0094	1,3587	0,26	1	0	0	0,0021	0,0067	1,415	0,0711
	2	0,995	0,005	0,0589	0,0002	0,004	2,7233	0,3934	0,998	0,002	0,0386	0,0002	0,0024	2,7802	0,3049
	3	0,9896	0,0103	0,0633	0,0001	0,0017	4,1014	0,4121	0,998	0,002	0,0257	0,00002	0,0006	4,1841	0,2335
	4	0,9756	0,0243	0,0873	0,0001	0,002	5,4054	0,598	0,9939	0,006	0,0509	0,0001	0,0018	5,5217	0,4055
	5	0,9377	0,0622	0,1323	0,00008	0,0014	6,7091	0,7498	0,9646	0,0353	0,0971	0,00008	0,0014	6,8469	0,4627
0,70	1	0,995	0,005	0,0944	0,0023	0,0087	1,4871	0,2956	1	0	0	0,0016	0,0058	1,551	0,0693
	2	0,9955	0,0045	0,0473	0,0001	0,0022	3,0286	0,3344	0,999	0,001	0,0223	0,00002	0,0006	3,0857	0,191
	3	0,9919	0,0081	0,0532	0,00002	0,0006	4,5188	0,4404	0,9983	0,0017	0,0332	0,00002	0,0006	4,6026	0,2728
	4	0,9814	0,0185	0,0745	0,00004	0,0009	5,9823	0,557	0,9944	0,0056	0,0507	0,00004	0,0009	6,0836	0,3969
	5	0,9577	0,0422	0,1038	0,00002	0,0007	7,5181	0,5158	0,9775	0,0226	0,0726	0	0	7,6107	0,3214

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,75	1	0,996	0,004	0,0891	0,0015	0,0072	1,6206	0,2891	1	0	0	0,0011	0,0047	1,6804	0,059
	2	0,9955	0,0045	0,0522	0,0001	0,0021	3,2765	0,3702	0,9995	0,0005	0,0158	0,00002	0,0006	3,3421	0,1994
	3	0,9933	0,0066	0,0467	0	0	4,9036	0,4561	0,9983	0,0026	0,0332	0	0	4,9887	0,2874
	4	0,9824	0,0175	0,0747	0,00002	0,0006	6,5038	0,5708	0,9944	0,0056	0,043	0,00002	0,0006	6,6143	0,3774
	5	0,9594	0,0405	0,1119	0	0	8,1147	0,7212	0,9778	0,0221	0,0833	0	0	8,2455	0,4066
0,80	1	0,997	0,003	0,0891	0,0013	0,0071	1,7197	0,3310	1	0	0	0,0008	0,0041	1,7873	0,0815
	2	0,9964	0,0036	0,0566	0,00004	0,0009	3,4636	0,4922	0,9996	0,0004	0,0158	0,00002	0,0006	3,5626	0,1825
	3	0,9933	0,0067	0,0467	0	0	5,233	0,4907	0,9993	0,0007	0,0235	0	0	5,3227	0,3192
	4	0,9844	0,0156	0,0708	0	0	6,968	0,5899	0,9959	0,0041	0,0369	0	0	7,0953	0,284
	5	0,9677	0,0322	0,0947	0	0	8,7018	0,6708	0,9828	0,0172	0,0756	0	0	8,8305	0,3796
0,85	1	0,998	0,002	0,0891	0,0011	0,0065	1,8016	0,3729	1	0	0	0	0	1,8701	0,1340
	2	0,9965	0,0035	0,0445	0,00004	0,0009	3,655	0,4526	1	0	0	0	0	3,7491	0,1187
	3	0,9946	0,0053	0,0418	0	0	5,5158	0,4999	1	0	0	0	0	5,6095	0,2455
	4	0,9879	0,012	0,0601	0	0	7,3683	0,5438	0,9967	0,0033	0,0324	0	0	7,471	0,2823
	5	0,9634	0,0366	0,1129	0,00002	0,0007	9,1081	0,8511	0,9786	0,0213	0,0891	0	0	9,2592	0,5784

EK-3. (devam) n=50 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,90	1	1	0	0	0,005	0,0035	1,8563	0,4111	1	0	0	0,0005	0,0031	1,9424	0,0772
	2	0,997	0,003	0,0386	0,0004	0,0144	3,7667	0,5149	1	0	0	0	0	3,8791	0,1762
	3	0,995	0,005	0,0431	0	0	5,7013	0,5432	1	0	0	0	0	5,8007	0,2988
	4	0,9885	0,0115	0,0607	0	0	7,5632	0,7263	0,9969	0,0041	0,0333	0	0	7,7104	0,4756
	5	0,9697	0,0302	0,1006	0,00002	0,0007	9,4674	0,8594	0,9847	0,0152	0,0717	0	0	9,6448	0,4606
0,95	1	1	0	0	0,0022	0,0447	1,8971	0,4191	1	0	0	0,0033	0,0547	1,9806	0,6529
	2	0,999	0,001	0,022	0	0	3,8416	0,5654	1	0	0	0	0	3,9358	0,3045
	3	0,998	0,002	0,0235	0	0	5,7808	0,6825	1	0	0	0,00004	0,0013	5,9034	0,4070
	4	0,9919	0,0081	0,0518	0	0	7,7538	0,6816	0,9977	0,0022	0,0284	0	0	7,8945	0,3575
	5	0,9786	0,0214	0,0946	0	0	9,6814	0,8784	0,9872	0,0128	0,0726	0	0	9,8672	0,4577

EK-4. n=100 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,45	1	0,9916	0,0084	0,0547	0,1434	0,0445	0,7772	0,1407	1	0	0	0,0344	0,0227	0,8168	0,0276
	2	0,9895	0,0105	0,0751	0,0094	0,0132	1,5656	0,1897	0,9996	0,0004	0,0158	0,0054	0,0105	1,6133	0,1154
	3	0,993	0,007	0,0635	0,0019	0,0058	2,3464	0,2379	0,9995	0,0005	0,0105	0,0014	0,0042	2,4205	0,1060
	4	0,991	0,009	0,0455	0,0006	0,0031	3,1206	0,2843	0,9985	0,0015	0,0223	0,0003	0,0021	3,2002	0,1614
	5	0,9888	0,0112	0,0582	0,0003	0,0028	3,8746	0,3409	0,9976	0,0024	0,0267	0,0001	0,0012	3,9716	0,2070
	6	0,9861	0,0138	0,0543	0,00008	0,0012	4,6434	0,3229	0,9963	0,0036	0,0276	0,00002	0,0006	4,7484	0,1896
	7	0,9734	0,0265	0,0727	0,0001	0,0016	5,3738	0,4172	0,9895	0,0104	0,0428	0,00003	0,0007	5,5072	0,2297
	8	0,9483	0,0516	0,0884	0,00005	0,0009	6,1128	0,4463	0,9658	0,0341	0,0687	0,00003	0,0007	6,2495	0,2824
	9	0,9174	0,0825	0,1043	0,00001	0,0003	6,8311	0,4449	0,9335	0,0664	0,0966	0,00001	0,0003	6,9852	0,3161
	10	0,8371	0,1629	0,1284	0	0	7,5605	0,4487	0,8718	0,1282	0,1268	0	0	7,7139	0,2937
0,50	1	0,9925	0,0075	0,0547	0,0431	0,0147	0,9369	0,1361	1	0	0	0,0313	0,0235	0,9724	0,0273
	2	0,9915	0,0085	0,0753	0,0083	0,0125	1,8706	0,2175	1	0	0	0,005	0,0101	1,922	0,1326
	3	0,9936	0,0064	0,0544	0,0016	0,0053	2,8154	0,2471	1	0	0	0,0012	0,004	2,876	0,1652
	4	0,9937	0,0063	0,0359	0,0003	0,0024	3,747	0,26	0,9987	0,0012	0,0208	0,0001	0,0018	3,8268	0,1595
	5	0,9902	0,0098	0,0458	0,0001	0,0012	4,6545	0,3245	0,9982	0,0018	0,0209	0,00002	0,0004	4,7548	0,2037
	6	0,9868	0,0131	0,0574	0,0001	0,0016	5,5521	0,4223	0,9955	0,0045	0,0373	0,00007	0,001	5,6699	0,283
	7	0,9798	0,0201	0,0562	0,00004	0,0006	6,4602	0,4062	0,9908	0,0091	0,0404	0,00004	0,0006	6,5809	0,2601
	8	0,9546	0,0453	0,0815	0,00003	0,0005	7,3236	0,5137	0,9688	0,0311	0,0683	0	0	7,4647	0,358
	9	0,9158	0,0841	0,1058	0,00002	0,0006	8,2075	0,5176	0,9404	0,0595	0,0904	0,00004	0,0009	8,3746	0,3499
	10	0,8641	0,1539	0,1284	0,00002	0,0004	9,0838	0,5428	0,8832	0,1168	0,123	0,00002	0,0007	9,246	0,3749

EK-4. (devam) n=100 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,55	1	0,9959	0,0041	0,0547	0,0438	0,0148	1,0999	0,1346	1	0	0	0,0312	0,0228	1,1318	0,0202
	2	0,9974	0,0026	0,0386	0,0569	0,2494	2,1836	0,2324	1	0	0	0,1066	0,2084	2,2412	0,115
	3	0,9973	0,0026	0,0332	0,0013	0,0041	3,2909	0,2849	1	0	0	0,0013	0,0043	3,3425	0,2218
	4	0,995	0,005	0,0444	0,0003	0,0026	4,3602	0,3438	1	0	0	0,0001	0,0012	4,4525	0,1805
	5	0,9912	0,0088	0,0551	0,0002	0,0023	5,4197	0,4692	0,9968	0,0032	0,0388	0,0001	0,0021	5,5342	0,2826
	6	0,9906	0,0093	0,0462	0,00009	0,0012	6,5011	0,4265	0,9951	0,0049	0,0374	0,00007	0,001	5,6599	0,283
	7	0,9812	0,0187	0,0599	0,00008	0,0011	7,5346	0,5261	0,9932	0,0067	0,0346	0,00002	0,0004	7,6696	0,3177
	8	0,9672	0,0327	0,0653	0	0	8,6007	0,4273	0,9978	0,0221	0,0533	0,00001	0,0003	8,745	0,2694
	9	0,9271	0,0728	0,0974	0,00001	0,0003	9,6243	0,5257	0,9463	0,0536	0,0816	0,00001	0,0003	9,7826	0,3573
	10	0,8539	0,1461	0,1275	0,00004	0,0008	10,6216	0,667	0,8833	0,1167	0,1206	0,00003	0,0006	10,7823	0,4578
0,60	1	0,996	0,004	0,0772	0,0433	0,0342	1,2472	0,1782	1	0	0	0,0299	0,023	1,2846	0,0255
	2	0,9975	0,0025	0,0158	0,0067	0,0108	2,5082	0,2181	1	0	0	0,004	0,009	2,549	0,1213
	3	0,9973	0,0027	0,0297	0,0009	0,0036	3,7398	0,3102	1	0	0	0,0008	0,003	3,8041	0,2091
	4	0,9952	0,0047	0,0392	0,0002	0,0022	4,9827	0,3488	1	0	0	0,00005	0,0007	5,0757	0,142
	5	0,9968	0,0032	0,0281	0,00007	0,0009	6,2236	0,3639	0,9996	0,0004	0,0126	0,00002	0,0006	6,3243	0,1785
	6	0,9918	0,0081	0,043	0,00004	0,0006	7,4109	0,5218	0,9995	0,0005	0,0117	0,00001	0,0003	7,5472	0,2544
	7	0,9818	0,0181	0,0573	0,00003	0,0005	8,6061	0,5516	0,9941	0,0058	0,0387	0,00005	0,001	8,7582	0,3966
	8	0,973	0,027	0,0626	0,00003	0,0005	9,8319	0,5091	0,9827	0,0172	0,0514	0,00002	0,0004	9,9804	0,3263
	9	0,9307	0,0692	0,0968	0	0	11,0021	0,6085	0,9493	0,0506	0,0801	0	0	11,1667	0,3786
	10	0,8665	0,1335	0,1282	0,00003	0,0007	12,1743	0,7098	0,8928	0,1072	0,1191	0,00002	0,0007	12,3684	0,3853

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,65	1	0,998	0,002	0,0705	0,0442	0,0395	1,3955	0,1962	1	0	0	0,0289	0,023	1,4317	0,0307
	2	0,9976	0,0024	0,0386	0,0095	0,053	2,7722	0,3083	1	0	0	0,004	0,008	2,8417	0,1008
	3	0,9976	0,0024	0,0331	0,0008	0,0037	4,1936	0,2997	1	0	0	0,0005	0,0026	4,2621	0,1461
	4	0,996	0,004	0,0333	0,0002	0,0017	5,5586	0,4392	1	0	0	0,00004	0,0006	5,6651	0,1696
	5	0,9956	0,0044	0,0319	0,00004	0,0006	6,9628	0,4207	0,9996	0,0004	0,0126	0,00002	0,0006	7,0576	0,2498
	6	0,9955	0,0045	0,029	0,00001	0,0003	8,3492	0,3856	0,9983	0,0016	0,021	0,00003	0,0007	8,4421	0,2775
	7	0,9888	0,0111	0,0428	0,00002	0,0004	9,6991	0,4951	0,9972	0,0027	0,0224	0,00001	0,0003	9,8414	0,2205
	8	0,9748	0,0251	0,066	0,00003	0,0007	11,0156	0,6626	0,9852	0,0147	0,0503	0	0	11,1836	0,3246
	9	0,9388	0,0611	0,0976	0,00004	0,0006	12,3475	0,6821	0,9541	0,0458	0,0801	0	0	12,5216	0,4465
	10	0,8824	0,1176	0,1284	0	0	13,6839	0,7403	0,903	0,097	0,1137	0	0	13,8727	0,3733
0,70	1	0,9981	0,0019	0,0834	0,043	0,0384	1,5227	0,2249	1	0	0	0,0291	0,0387	1,5664	0,0522
	2	0,9977	0,0023	0,0352	0,0084	0,053	3,0612	0,3113	1	0	0	0,0055	0,0439	3,1172	0,1635
	3	0,9979	0,0021	0,0314	0,0016	0,0294	4,5972	0,3743	1	0	0	0,0005	0,0023	4,675	0,1596
	4	0,9965	0,0035	0,0333	0,0001	0,0011	6,1315	0,4205	1	0	0	0,00003	0,0007	6,2271	0,1804
	5	0,9952	0,0048	0,0365	0,00004	0,0006	7,6497	0,5044	0,999	0,001	0,0189	0,00006	0,0009	7,7417	0,3543
	6	0,995	0,005	0,0303	0,00001	0,0003	9,1953	0,4386	0,9991	0,0008	0,0157	0,00001	0,0003	9,2968	0,3104
	7	0,9871	0,0128	0,046	0	0	10,6309	0,6740	0,9961	0,0038	0,0326	0,00002	0,0004	10,8001	0,4131
	8	0,978	0,022	0,0613	0,00002	0,0006	12,167	0,6020	0,9881	0,0118	0,0464	0	0	12,3088	0,4233
	9	0,9486	0,0513	0,0906	0,00002	0,0004	13,667	0,6311	0,9594	0,0405	0,0725	0	0	13,8367	0,3233
	10	0,8895	0,1105	0,1205	0	0	15,1389	0,6706	0,9148	0,0852	0,1077	0,00001	0,0003	15,3089	0,4187

EK-4. (devam) n=100 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,75	1	0,9985	0,0015	0,0834	0,0423	0,0454	1,6327	0,29	1	0	0	0,0274	0,023	1,6887	0,0585
	2	0,9984	0,0016	0,0445	0,0066	0,0426	3,323	0,2693	1	0	0	0,0035	0,0316	3,3756	0,084
	3	0,998	0,002	0,0257	0,0024	0,0422	4,987	0,3441	1	0	0	0,0003	0,0018	5,0562	0,1344
	4	0,997	0,003	0,0294	0,00001	0,0003	6,6522	0,3693	1	0	0	0	0	6,7382	0,0994
	5	0,9956	0,0044	0,0319	0,001	0,0316	8,2939	0,4615	1	0	0	0	0	8,40760	0,1289
	6	0,9923	0,0076	0,0414	0,00001	0,0003	9,9102	0,585	0,9993	0,0007	0,0128	0	0	10,0494	0,3519
	7	0,992	0,008	0,038	0,00001	0,0003	11,5842	0,53	0,9974	0,0026	0,0237	0,00001	0,0003	11,7040	0,3514
	8	0,9831	0,0168	0,0531	0,00001	0,0003	13,2251	0,5633	0,992	0,008	0,0335	0	0	13,3591	0,2892
	9	0,9545	0,0454	0,0981	0,00007	0,001	14,8108	0,7893	0,9702	0,0297	0,0648	0,00001	0,0003	14,9941	0,4235
	10	0,9007	0,0993	0,1245	0,00003	0,0006	16,415	0,7727	0,9237	0,0763	0,1179	0	0	16,5862	0,5338
0,80	1	0,999	0,001	0,0547	0,0428	0,0599	1,7622	0,2407	1	0	0	0,0274	0,0223	1,7975	0,0398
	2	0,9985	0,0015	0,0273	0,0051	0,0432	3,5232	0,3602	1	0	0	0,0017	0,0057	3,5893	0,0428
	3	0,9981	0,0019	0,0314	0,0032	0,0519	5,3101	0,351	1	0	0	0,0011	0,0316	5,3782	0,1381
	4	0,9977	0,0023	0,0304	0,00003	0,0005	7,0768	0,4101	1	0	0	0,00001	0,0003	7,1605	0,1634
	5	0,9962	0,0038	0,0287	0,00001	0,0003	8,829	0,5944	1	0	0	0,00001	0,0003	8,9265	0,3675
	6	0,9961	0,0039	0,0291	0,00001	0,0003	10,6048	0,4930	0,9991	0,0009	0,0139	0	0	10,7122	0,29
	7	0,9921	0,0079	0,035	0,00001	0,0003	12,3464	0,5589	0,9974	0,0025	0,0254	0,00002	0,0004	12,4797	0,3263
	8	0,9826	0,0173	0,055	0,00002	0,0004	14,0798	0,7413	0,9928	0,0071	0,0366	0	0	14,2474	0,4119
	9	0,9655	0,0334	0,0853	0	0	15,8594	0,7035	0,9764	0,0235	0,0625	0,00001	0,0003	15,9936	0,4410
	10	0,9127	0,0873	0,1272	0,00003	0,0006	17,6028	0,7087	0,9376	0,0624	0,1	0,00001	0,0003	17,7451	0,4891

EK-4. (devam) n=100 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,85	1	0,999	0,001	0,0631	0,0387	0,0341	1,8795	0,2552	1	0	0	0,0278	0,0226	1,8826	0,0365
	2	0,9986	0,0014	0,0315	0,0111	0,0885	3,6986	0,3598	1	0	0	0,0019	0,0058	3,757	0,1605
	3	0,9983	0,0017	0,0314	0,0011	0,0316	5,5615	0,4394	1	0	0	0,0001	0,0012	5,6404	0,1483
	4	0,9982	0,0017	0,0208	0,002	0,044	7,4166	0,4752	1	0	0	0	0	7,5097	0,2154
	5	0,9952	0,0048	0,0319	0,0003	0,0113	9,2689	0,4908	1	0	0	0,00001	0,0003	9,3815	0,2011
	6	0,9955	0,0045	0,029	0	0	11,1459	0,5474	0,999	0,001	0,0166	0	0	11,2473	0,2727
	7	0,9945	0,0054	0,028	0	0	13,0339	0,4752	0,9995	0,0004	0,0078	0	0	13,14	0,1712
	8	0,9858	0,0141	0,0647	0	0	14,8492	0,6189	0,9941	0,058	0,0427	0	0	14,9897	0,2343
	9	0,9648	0,0351	0,0967	0	0	16,6883	0,6884	0,9776	0,0224	0,0635	0	0	16,8385	0,3325
	10	0,9306	0,0694	0,1108	0	0	18,5523	0,6953	0,9473	0,0527	0,0891	0	0	18,7027	0,3551
0,90	1	1	0	0	0,0464	0,0075	1,8827	0,4147	1	0	0	0	0	1,9407	0,1221
	2	0,9989	0,0011	0,0315	0,0081	0,0766	3,8327	0,3473	1	0	0	0	0	3,8865	0,1278
	3	0,9985	0,0015	0,0297	0,0001	0,0014	5,7598	0,4179	1	0	0	0	0	5,8203	0,2373
	4	0,9987	0,0013	0,0283	0,0019	0,04258	7,6837	0,4635	1	0	0	0	0	7,7607	0,2524
	5	0,9974	0,0026	0,0226	0	0	9,6085	0,5889	1	0	0	0	0	9,7065	0,2488
	6	0,9961	0,0038	0,0249	0	0	11,5246	0,5965	0,9991	0,0009	0,0117	0	0	11,6421	0,2788
	7	0,9937	0,0062	0,03	0	0	13,4504	0,5885	0,998	0,002	0,0179	0	0	13,5682	0,3735
	8	0,9862	0,0137	0,0632	0	0	15,3686	0,6848	0,9941	0,0058	0,0389	0	0	15,4853	0,4987
	9	0,9722	0,0278	0,076	0,00002	0,0004	17,2548	0,9477	0,9826	0,0173	0,0525	0	0	17,4366	0,4643
	10	0,9418	0,0582	0,1149	0,00002	0,0004	19,2097	0,7916	0,9594	0,0406	0,0895	0	0	19,3721	0,4897

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,95	1	1	0	0	0,0419	0,0847	1,9378	0,3143	1	0	0	0,029	0,0484	1,9819	0,1020
	2	0,999	0,001	0,0223	0,0087	0,0849	3,88746	0,437	1	0	0	0,0007	0,0033	3,9402	0,264
	3	0,9993	0,0007	0,0148	0,0040	0,0631	5,8633	0,4926	1	0	0	0,002	0,0447	5,9269	0,2659
	4	0,9989	0,0011	0,0193	0	0	7,8123	0,5226	1	0	0	0,0001	0,0032	7,9001	0,3607
	5	0,998	0,002	0,0154	0	0	9,7774	0,654	1	0	0	0,00001	0,0003	9,8803	0,4055
	6	0,9953	0,0046	0,0294	0	0	11,5276	0,6007	0,9988	0,0012	0,0174	0	0	11,6404	0,3415
	7	0,9925	0,0074	0,0348	0,00001	0,0003	13,4341	0,7243	0,9984	0,0016	0,0266	0	0	13,5705	0,3946
	8	0,9921	0,0079	0,0514	0	0	15,7178	0,7444	0,9968	0,0032	0,0251	0	0	15,8125	0,5234
	9	0,9803	0,0196	0,0728	0,00003	0,0006	17,6232	0,9476	0,988	0,012	0,0625	0	0	17,8005	0,5058
	10	0,9617	0,0383	0,0999	0,00001	0,0003	19,6232	0,8323	0,9708	0,0292	0,0889	0	0	19,7845	0,5406

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,45	1	0,999	0,001	0,0834	0,0499	0,0052	0,8042	0,0922	1	0	0	0,0516	0,0053	0,8258	0,01924
	2	0,9975	0,0025	0,0753	0,0446	0,0074	1,6001	0,1786	1	0	0	0,0455	0,0105	1,6396	0,0918
	3	0,9966	0,0033	0,0363	0,0363	0,0101	2,4213	0,1513	1	0	0	0,0352	0,0154	2,4641	0,093
	4	0,9962	0,0038	0,0283	0,0214	0,0125	3,2139	0,1828	0,9995	0,0005	0,0111	0,0135	0,0148	3,273	0,1009
	5	0,9956	0,0044	0,0366	0,0089	0,0102	4,0105	0,2454	0,999	0,001	0,0189	0,0057	0,0089	4,0844	0,142
	6	0,9951	0,0049	0,0189	0,003	0,0054	4,8135	0,2104	0,9998	0,0001	0,0052	0,0021	0,0043	4,8883	0,1072
	7	0,9942	0,0057	0,032	0,0012	0,0032	5,5863	0,2877	0,9985	0,0014	0,0168	0,0006	0,0021	5,6817	0,1811
	8	0,9963	0,0036	0,0244	0,0003	0,0015	6,3804	0,2986	0,9987	0,0013	0,0147	0,0001	0,0011	6,4769	0,1683
	9	0,9951	0,0049	0,0238	0,0001	0,0009	7,1543	0,2824	0,9986	0,0014	0,006	0,00004	0,0005	7,2763	0,1507
	10	0,9929	0,0071	0,0303	0,00007	0,0008	7,9306	0,3672	0,9985	0,0015	0,0137	0,00002	0,0003	8,0572	0,1912
	11	0,9905	0,0094	0,0341	0,00003	0,0005	8,665	0,3989	0,9974	0,0026	0,0184	0,00001	0,0002	8,8178	0,2377
	12	0,9917	0,0083	0,0287	0,00001	0,0003	9,483	0,3349	0,9966	0,0034	0,0179	0	0	9,6147	0,196
	13	0,9849	0,015	0,0373	0	0	10,219	0,3701	0,9928	0,0071	0,0269	0	0	10,374	0,239
	14	0,9701	0,0299	0,0538	0,00001	0,0002	10,9592	0,4744	0,981	0,019	0,0438	0,000005	0,0001	11,1181	0,3289
	15	0,9524	0,0476	0,0635	0,000005	0,0001	11,7264	0,4238	0,9662	0,0338	0,0581	0,000005	0,0001	11,8858	0,3134
	16	0,9218	0,0781	0,0839	0,000005	0,0001	12,4733	0,4577	0,9398	0,0601	0,0782	0	0	12,633	0,3199
	17	0,8667	0,1332	0,1001	0	0	13,1782	0,5226	0,8978	0,1021	0,1019	0	0	13,37	0,337
	18	0,8038	0,1961	0,1114	0	0	13,9071	0,5254	0,8451	0,1548	0,1198	0	0	14,119	0,29856
	19	0,7252	0,2747	0,1062	0	0	14,6326	0,549	0,7723	0,2276	0,123	0	0	14,8525	0,3488
	20	0,651	0,349	0,0915	0	0	15,3379	0,5523	0,6808	0,3192	0,1161	0	0	15,5716	0,3209

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,50	1	0,9994	0,0006	0,0631	0,0512	0,0223	0,9575	0,1139	1	0	0	0,0516	0,0053	0,9819	0,0186
	2	0,9975	0,0025	0,0352	0,0451	0,0073	1,9229	0,1329	1	0	0	0,0457	0,0087	1,9527	0,0797
	3	0,9973	0,0027	0,0391	0,0372	0,0109	2,8702	0,2132	1	0	0	0,035	0,0157	2,9241	0,1097
	4	0,998	0,002	0,0222	0,0209	0,013	3,8403	0,2063	0,9997	0,0079	0,0003	0,0132	0,0149	3,8952	0,1295
	5	0,9966	0,0034	0,0258	0,0087	0,0099	4,7828	0,2429	0,9996	0,0004	0,0089	0,0053	0,0087	4,8501	0,1515
	6	0,9958	0,0041	0,027	0,0027	0,0054	5,7227	0,3081	0,9996	0,0003	0,0074	0,0019	0,0042	5,8173	0,1527
	7	0,9968	0,0031	0,0219	0,0008	0,0024	6,6711	0,318	0,9992	0,0007	0,01	0,0005	0,002	6,7656	0,2009
	8	0,9966	0,0033	0,0231	0,0003	0,0014	7,6195	0,3381	0,999	0,001	0,0136	0,0001	0,001	7,7163	0,2412
	9	0,995	0,005	0,0278	0,0001	0,0011	8,5443	0,3674	0,9994	0,0006	0,0092	0,00003	0,0004	8,6699	0,1895
	10	0,9944	0,0056	0,0266	0,00003	0,0004	9,4647	0,3903	0,9983	0,0017	0,0169	0,00002	0,0004	9,5864	0,3145
	11	0,9931	0,0068	0,0292	0,00002	0,0005	10,3945	0,4158	0,9978	0,0022	0,0189	0,00002	0,0005	10,5278	0,2576
	12	0,991	0,009	0,0318	0,00001	0,0023	11,3079	0,4185	0,9963	0,0037	0,0207	0,00001	0,0002	11,4529	0,2614
	13	0,9861	0,0138	0,035	0,000005	0,0001	12,2219	0,4193	0,9935	0,0064	0,0239	0	0	12,3804	0,2594
	14	0,9751	0,0248	0,0464	0,000005	0,0001	13,1313	0,445	0,9853	0,0146	0,0355	0,000001	0,0002	13,3027	0,2925
	15	0,9554	0,0446	0,0616	0,000005	0,0001	14,0228	0,5386	0,968	0,032	0,055	0,000005	0,0001	14,1953	0,3723
	16	0,9272	0,0727	0,0771	0	0	14,9106	0,4952	0,9463	0,0536	0,0695	0	0	15,0954	0,3171
	17	0,8795	0,1204	0,0973	0,000005	0,0001	15,78	0,6299	0,9077	0,0922	0,095	0	0	15,9857	0,3736
	18	0,8173	0,1827	0,1065	0	0	16,6953	0,5256	0,8556	0,1443	0,117	0	0	16,8891	0,3431
	19	0,731	0,269	0,1069	0	0	17,5556	0,5706	0,7803	0,2196	0,1245	0	0	17,772	0,3696
	20	0,6572	0,3427	0,0939	0	0	18,391	0,7069	0,6932	0,3067	0,12	0	0	18,6369	0,4092

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,55	1	0,9995	0,0005	0,0705	0,0594	0,0118	1,1129	0,1266	1	0	0	0,0997	0,0096	1,1384	0,03
	2	0,998	0,002	0,0315	0,0445	0,0073	2,2373	0,1235	1	0	0	0,045	0,0111	2,2657	0,0763
	3	0,998	0,002	0,0257	0,0372	0,0237	3,3464	0,2105	1	0	0	0,0352	0,0337	3,3986	0,1016
	4	0,998	0,002	0,0222	0,0204	0,0167	4,4623	0,2024	0,9997	0,0003	0,0079	0,013	0,014	4,5229	0,0945
	5	0,9978	0,0022	0,0208	0,0085	0,0215	5,5691	0,2884	0,9992	0,0008	0,0126	0,0048	0,0082	5,6365	0,196
	6	0,9985	0,0015	0,0157	0,0023	0,0047	6,6881	0,2734	0,999	0,001	0,0052	0,0015	0,0036	6,7646	0,1484
	7	0,9977	0,0022	0,022	0,0008	0,0028	7,7724	0,3574	0,9997	0,0003	0,0063	0,0003	0,0015	7,8767	0,1503
	8	0,9968	0,0031	0,0217	0,0002	0,0014	8,8412	0,4367	0,9992	0,0008	0,0111	0,0001	0,0009	8,9674	0,2443
	9	0,9966	0,0034	0,0202	0,00004	0,0004	9,9401	0,4084	0,9992	0,0008	0,0116	0,00001	0,0002	10,0693	0,2375
	10	0,9961	0,0039	0,0218	0,00003	0,0004	11,0374	0,433	0,9986	0,0014	0,0140	0,00001	0,0003	11,1663	0,2613
	11	0,9942	0,0058	0,0271	0,00002	0,0003	12,1178	0,4422	0,9986	0,0014	0,0137	0,00002	0,0003	12,2636	0,2646
	12	0,9933	0,0067	0,0278	0,00001	0,0003	13,1927	0,4686	0,9968	0,0032	0,0194	0,00002	0,0004	13,3247	0,3488
	13	0,9882	0,0118	0,0343	0,00001	0,0003	14,2513	0,489	0,9943	0,0057	0,0231	0,000005	0,0001	14,415	0,2716
	14	0,9793	0,0207	0,0436	0	0	15,2921	0,5504	0,9876	0,0124	0,0332	0	0	15,4752	0,3646
	15	0,9623	0,0376	0,0565	0,000005	0,0001	16,3859	0,5243	0,9723	0,0276	0,0505	0	0	16,564	0,3205
	16	0,9343	0,0656	0,0742	0,000005	0,0001	17,413	0,5715	0,9508	0,0491	0,0681	0	0	17,6137	0,312
	17	0,8991	0,1109	0,0967	0	0	18,4458	0,6124	0,9138	0,0861	0,0937	0	0	18,6585	0,3574
	18	0,8222	0,1778	0,1027	0	0	19,5163	0,5416	0,8580	0,1419	0,1131	0	0	19,7038	0,3432
	19	0,7440	0,2559	0,112	0	0	20,5322	0,6102	0,7927	0,2073	0,1293	0	0	20,7523	0,3503
	20	0,6633	0,3367	0,0981	0	0	21,5645	0,6567	0,7026	0,2974	0,1209	0	0	21,787	0,3981

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,60	1	0,9997	0,0003	0,0547	0,0498	0,0054	1,2682	0,1305	1	0	0	0,0515	0,0051	1,2917	0,0247
	2	0,9985	0,0015	0,0417	0,0473	0,0508	2,5352	0,1912	1	0	0	0,0453	0,0094	2,5746	0,0753
	3	0,9983	0,0017	0,0148	0,0403	0,0592	3,8142	0,2309	1	0	0	0,0334	0,0167	3,8632	0,0603
	4	0,9982	0,0018	0,0248	0,0208	0,0443	5,073	0,277	1	0	0	0,0116	0,0149	5,1457	0,0918
	5	0,9984	0,0016	0,0236	0,0087	0,0376	6,3484	0,2843	0,9998	0,0002	0,0014	0,0042	0,0077	6,4196	0,13874
	6	0,9985	0,0015	0,0157	0,0024	0,007	7,6183	0,228	0,9998	0,0002	0,0026	0,0014	0,0033	7,6945	0,0858
	7	0,9982	0,0017	0,0179	0,0015	0,0311	8,8496	0,3945	0,9997	0,0003	0,0063	0,0003	0,0014	8,9459	0,188
	8	0,9976	0,0023	0,0196	0,0001	0,0012	10,0954	0,4349	0,9997	0,0003	0,0055	0,00006	0,0006	10,2126	0,1806
	9	0,998	0,002	0,0163	0,00003	0,0004	11,3638	0,3684	0,9997	0,0003	0,0049	0,00003	0,0005	11,4661	0,2264
	10	0,997	0,003	0,0182	0,00001	0,0002	12,5899	0,3903	0,9995	0,0005	0,0083	0,00001	0,0003	12,7279	0,2123
	11	0,9959	0,0041	0,0197	0,000005	0,0001	13,817	0,4347	0,9993	0,0007	0,0086	0	0	13,9708	0,2126
	12	0,9932	0,0068	0,0286	0,00002	0,0004	15,0231	0,5439	0,9983	0,0017	0,0133	0	0	15,1898	0,3243
	13	0,9912	0,0088	0,0297	0,000006	0,0001	16,2798	0,4834	0,997	0,003	0,0166	0	0	16,4463	0,2527
	14	0,9826	0,0174	0,0371	0	0	17,4673	0,5656	0,9964	0,0036	0,0273	0	0	17,666	0,2898
	15	0,9673	0,0327	0,0543	0,000001	0,0003	18,6805	0,6388	0,9759	0,0241	0,0467	0,000005	0,0001	18,8632	0,4374
	16	0,9468	0,0531	0,0666	0	0	19,9296	0,4852	0,9578	0,0421	0,0611	0	0	20,1104	0,3005
	17	0,8987	0,1013	0,093	0,000005	0,0001	21,08	0,7122	0,9212	0,0787	0,0904	0	0	21,2976	0,4728
	18	0,8347	0,1653	0,1081	0	0	22,2805	0,6604	0,8683	0,1317	0,1141	0	0	22,5002	0,4744
	19	0,7458	0,2541	0,1129	0	0	23,5022	0,6966	0,7907	0,2092	0,1318	0	0	23,7217	0,3842
	20	0,6706	0,3293	0,1087	0	0	24,683	0,6531	0,715	0,2849	0,1267	0	0	24,912	0,3834

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,65	1	1	0	0	0,0499	0,0054	1,4183	0,1154	1	0	0	0,0515	0,0051	1,4387	0,0224
	2	0,9985	0,0015	0,0353	0,0493	0,0689	2,8265	0,2196	1	0	0	0,0464	0,031	2,8685	0,1076
	3	0,9984	0,0016	0,0013	0,0210	0,04	4,2563	0,2225	1	0	0	0,0352	0,0342	4,3023	0,1113
	4	0,9983	0,0017	0,0222	0,025	0,0702	5,6483	0,3551	0,9999	0,0001	0,0024	0,013	0,0333	5,7237	0,1506
	5	0,9982	0,0018	0,0188	0,0093	0,0523	7,0616	0,336	0,9998	0,0002	0,0044	0,0051	0,0278	7,1413	0,2283
	6	0,9981	0,0019	0,0203	0,0049	0,0471	8,4841	0,3314	0,9997	0,0003	0,0032	0,0013	0,0034	8,5654	0,2581
	7	0,9984	0,0157	0,0149	0,0014	0,0218	9,8692	0,412	0,9995	0,0005	0,0078	0,0011	0,0288	9,9717	0,2558
	8	0,9983	0,0017	0,0162	0,0001	0,0008	11,3026	0,3744	0,9996	0,0004	0,0088	0,00003	0,0004	11,3909	0,2488
	9	0,9972	0,0028	0,0193	0,00002	0,0004	12,6674	0,4858	0,9996	0,0004	0,006	0,000005	0,0001	12,8033	0,2319
	10	0,9967	0,0033	0,0218	0,00001	0,0002	14,0939	0,4118	0,9992	0,0008	0,0109	0,000005	0,0001	14,2037	0,2676
	11	0,9952	0,0048	0,0271	0,00002	0,0004	15,4461	0,6155	0,9986	0,0014	0,0148	0,000005	0,0001	15,6043	0,2626
	12	0,994	0,006	0,0231	0	0	16,8107	0,0509	0,9983	0,0017	0,0122	0	0	16,9727	0,318
	13	0,9924	0,0075	0,0271	0	0	18,2122	0,5975	0,9969	0,0031	0,0169	0	0	18,3757	0,3322
	14	0,9864	0,0136	0,0354	0,00001	0,0003	19,5852	0,5245	0,9923	0,0077	0,028	0,00001	0,0003	19,7545	0,3030
	15	0,9742	0,0258	0,0464	0	0	20,9259	0,6033	0,9818	0,0182	0,0395	0	0	21,1144	0,4162
	16	0,9508	0,0492	0,0666	0	0	22,2901	0,6727	0,9621	0,0378	0,0668	0	0	22,4946	0,3861
	17	0,9059	0,0941	0,0912	0	0	23,6726	0,6586	0,9297	0,0703	0,0855	0	0	23,88	0,3617
	18	0,8406	0,1593	0,1165	0,000005	0,0001	25,001	0,7092	0,8757	0,1243	0,1102	0,000005	0,0001	25,2194	0,484
	19	0,7623	0,2377	0,1227	0	0	26,364	0,7523	0,8097	0,1902	0,1321	0	0	26,575	0,4020
	20	0,6763	0,3237	0,1157	0,000005	0,0001	27,646	0,8484	0,7251	0,2749	0,1339	0	0	27,9212	0,4465

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,70	1	1	0	0	0,0516	0,0395	1,5468	0,1738	1	0	0	0,0513	0,0053	1,5735	0,0313
	2	0,999	0,001	0,0223	0,0462	0,043	3,1074	0,211	1	0	0	0,0453	0,0091	3,1416	0,0677
	3	0,9986	0,0014	0,0315	0,0383	0,0495	4,6511	0,2919	1	0	0	0,0339	0,0157	4,7113	0,1049
	4	0,9985	0,0015	0,0193	0,0245	0,0755	6,2132	0,2956	0,9999	0,0001	0,0024	0,0036	0,0448	6,2724	0,1977
	5	0,9986	0,0014	0,0166	0,0087	0,0527	7,7685	0,3695	0,9999	0,0001	0	0,0125	0,0068	7,8402	0,1454
	6	0,9991	0,0009	0,0117	0,0043	0,0522	9,3211	0,3341	0,9999	0,0001	0	0,0009	0,0028	9,3976	0,1956
	7	0,9991	0,0009	0,011	0,0003	0,0014	10,8783	0,3501	0,9998	0,0002	0,0045	0,0002	0,0012	10,9575	0,1352
	8	0,9981	0,0019	0,0171	0,00006	0,0005	12,3926	0,4508	0,9998	0,0002	0,0039	0,00001	0,0002	12,4977	0,2533
	9	0,998	0,002	0,0163	0,00001	0,0002	13,9319	0,4614	0,9993	0,0007	0,0099	0,00002	0,0004	14,0428	0,2796
	10	0,9967	0,0034	0,0204	0,00001	0,0002	15,4596	0,4506	0,9987	0,0013	0,0144	0,000005	0,0001	15,5848	0,2845
	11	0,9959	0,0041	0,0242	0,00002	0,0004	16,9705	0,6487	0,9988	0,0012	0,0124	0,00005	0,0001	17,1194	0,3610
	12	0,994	0,006	0,0266	0,00001	0,0003	18,4923	0,6138	0,9985	0,0015	0,0125	0,00005	0,0001	18,6556	0,3271
	13	0,9936	0,0064	0,0244	0,0001	0,004	20,0387	0,5168	0,9976	0,0024	0,0153	0,00009	0,0028	20,1963	0,2845
	14	0,9882	0,0118	0,0307	0,000005	0,0001	21,5949	0,5268	0,9927	0,0072	0,0242	0,000005	0,0001	21,7182	0,3924
	15	0,977	0,023	0,0424	0,000005	0,0001	23,0481	0,5958	0,9816	0,0183	0,0469	0	0	23,2289	0,3417
	16	0,9504	0,0496	0,0831	0,00001	0,0002	24,5834	0,6644	0,9653	0,0346	0,058	0,000001	0,0002	24,7445	0,4632
	17	0,9146	0,0853	0,0997	0	0	26,0846	0,6524	0,9379	0,062	0,0848	0	0	26,2745	0,361
	18	0,8562	0,1438	0,1204	0	0	27,544	0,7403	0,8887	0,1113	0,1109	0	0	27,7808	0,3552
	19	0,7684	0,2316	0,1346	0,000005	0,0001	29,0361	0,7899	0,8163	0,1837	0,1372	0	0	29,2513	0,5448
	20	0,6815	0,3139	0,1233	0,00001	0,0002	30,5435	0,795	0,7405	0,2594	0,1336	0,00001	0,0002	30,7763	0,4234

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,75	1	1	0	0	0,0506	0,0305	1,6833	0,1342	1	0	0	0,0515	0,005	1,6957	0,0542
	2	1	0	0	0,0487	0,0645	3,3465	0,2596	1	0	0	0,0491	0,0607	3,3791	0,1578
	3	0,999	0,001	0,0235	0,0387	0,0541	5,0272	0,3214	1	0	0	0,0331	0,0164	5,0797	0,1238
	4	0,9985	0,0015	0,0294	0,0198	0,0536	6,7019	0,315	1	0	0	0,0315	0,0135	6,7675	0,1209
	5	0,9984	0,0016	0,0178	0,009	0,0579	8,3939	0,318	0,9995	0,0005	0,0012		0,0315	8,4553	0,1254
	6	0,9983	0,0017	0,0181	0,0047	0,0552	10,0452	0,4503	0,9993	0,0007	0,006		0,0312	10,1306	0,1899
	7	0,9982	0,0018	0,0168	0,0022	0,0424	11,7212	0,4917	0,999	0,001	0,0063	0,0001	0,0008	11,8204	0,178
	8	0,9977	0,0023	0,0175	0,0015	0,0338	13,4002	0,4272	0,9989	0,0011	0,0039	0,00001	0,0002	13,5054	0,1351
	9	0,9982	0,0018	0,0139	0,00002	0,0003	15,0501	0,4708	0,9988	0,0012	0,0351	0	0	15,1745	0,1752
	10	0,9973	0,0027	0,0185	0,000005	0,0001	16,722	0,4836	0,9993	0,0007	0,0104	0	0	16,8308	0,3266
	11	0,9967	0,0032	0,0183	0	0	18,3961	0,4538	0,9996	0,0004	0,007	0	0	18,521	0,1864
	12	0,996	0,004	0,0191	0,000005	0,0001	20,0436	0,5656	0,9989	0,011	0,0114	0,00001	0,0002	20,1635	0,3599
	13	0,995	0,005	0,0210	0	0	21,6906	0,5597	0,9982	0,0018	0,012	0	0	21,8349	0,3277
	14	0,9902	0,0097	0,0302	0	0	23,3746	0,5458	0,9962	0,0037	0,0174	0	0	23,499	0,3172
	15	0,9625	0,0375	0,0158	0,0009	0,0063	24,7101	0,6929	0,977	0,023	0,0881	0,0004	0,0047	24,6817	
	16	0,9553	0,0446	0,0843	0,00001	0,0002	26,5961	0,8182	0,9686	0,0313	0,0714	0,00001	0,0002	26,7664	0,5352
	17	0,9221	0,0778	0,1076	0	0	28,3146	0,5854	0,9438	0,0561	0,0834	0	0	28,4699	0,3001
	18	0,8577	0,1423	0,1274	0,000005	0,0001	29,8858	0,7811	0,8989	0,1011	0,1116	0,000005	0,0001	30,0891	0,3811
	19	0,7841	0,2159	0,1388	0	0	31,5516	0,6522	0,8376	0,1624	0,1299	0	0	31,7516	0,3495
	20	0,7093	0,2906	0,1302	0,000005	0,0001	33,1711	0,6682	0,753	0,247	0,1385	0	0	33,3639	0,3662

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,80	1	1	0	0	0,0510	0,0368	1,7888	0,1264	1	0	0	0,0517	0,0049	1,8008	0,044
	2	1	0	0	0,0502	0,0758	3,5589	0,2764	1	0	0	0,0463	0,0312	3,5930	0,1367
	3	0,9996	0,0004	0,0278	0,0388	0,0587	5,3528	0,2928	1	0	0	0,0335	0,0158	5,3975	0,0744
	4	0,998	0,002	0,0222	0,0218	0,0661	7,1322	0,3227	1	0	0	0,0114	0,0441	7,1827	0,1722
	5	0,9992	0,0008	0,0126	0,0124	0,0858	8,919	0,3633	0,9999	0,0001	0,0049	0,0034	0,032	8,9828	0,1590
	6	0,9983	0,0017	0,0181	0,0045	0,0564	10,7006	0,3812	0,9998	0,0002	0,0031	0,0015	0,0316	10,7724	0,1965
	7	0,998	0,002	0,019	0,0019	0,0396	12,4795	0,4827	0,9997	0,0003	0,0063	0,00007	0,0006	12,5668	0,166
	8	0,9986	0,0014	0,013	0,0007	0,023	14,2635	0,5050	0,9997	0,0003	0,0055	0,0009	0,0312	14,3461	0,2526
	9	0,9979	0,0021	0,1075	0,0013	0,0064	16,4658	0,5074	0,9997	0,0003	0,0072	0,0007	0,0043	16,9537	0,2522
	10	0,9978	0,0022	0,0153	0,00001	0,0002	17,8092	0,5091	0,9997	0,0003	0,0054	0	0	17,9157	0,2529
	11	0,9973	0,0027	0,0152	0	0	19,5756	0,5514	0,9991	0,0009	0,0085	0	0	19,6989	0,2312
	12	0,9964	0,0036	0,0195	0,000005	0,0001	21,3572	0,5502	0,9985	0,0015	0,0153	0,000005	0,0001	21,4573	0,3994
	13	0,9951	0,0049	0,0221	0	0	23,112	0,6223	0,9981	0,0018	0,0122	0	0	23,2513	0,3255
	14	0,9895	0,0105	0,0448	0	0	24,8815	0,691	0,9948	0,0052	0,032	0	0	25,0192	0,3764
	15	0,9819	0,0181	0,018	0,0496	0,00001	26,6487	0,6521	0,9888	0,0112	0,0487	0	0	26,7837	0,3883
	16	0,9645	0,0355	0,0759	0	0	28,4181	0,63	0,9768	0,0231	0,0637	0,000005	0,0001	28,5773	0,3359
	17	0,9344	0,0655	0,0938	0,000005	0,0001	30,1714	0,6459	0,9503	0,0497	0,0884	0,000005	0,0001	30,3309	0,3739
	18	0,8841	0,1159	0,1173	0,000005	0,0001	31,8621	0,8702	0,92	0,08	0,1009	0	0	32,0848	0,4817
	19	0,8042	0,1957	0,1257	0,000005	0,0001	33,7202	0,6283	0,8563	0,1436	0,1358	0	0	33,8529	0,4682
	20	0,7162	0,0238	0,1267	0,000005	0,0001	35,4391	0,7051	0,7703	0,2297	0,1452	0,000005	0,0001	35,5942	0,5242

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,85	1	1	0	0	0,0046	0,047	1,8676	0,1781	1	0	0	0,0512	0,0053	1,8827	0,0845
	2	1	0	0	0,0031	0,004	3,7326	0,2703	1	0	0	0,0041	0,007	3,75830	0,154
	3	0,9997	0,0003	0,0278	0,0391	0,0673	5,6072	0,3085	1	0	0	0,0339	0,0341	5,6381	0,2058
	4	0,9987	0,0013	0,0176	0,0198	0,058	7,4865	0,3195	1	0	0	0,0096	0,013	7,5337	0,1155
	5	0,998	0,002	0,0199	0,0104	0,078	9,3483	0,4575	0,9994	0,0006	0,0109	0,0052	0,0538	9,4088	0,1586
	6	0,9985	0,0015	0,0157	0,0035	0,0524	11,2453	0,345	0,9996	0,0004	0,0074	0,0013	0,0282	11,2972	0,1399
	7	0,9988	0,0012	0,0127	0,0038	0,0589	13,1116	0,4151	0,9994	0,0005	0,009	0,001	0,0312	13,1703	0,1913
	8	0,9978	0,0022	0,0161	0,002	0,0445	14,968	0,4195	0,9993	0,0007	0,0088	0,0009	0,0312	15,0424	0,216
	9	0,9978	0,0022	0,0159	0,000005	0,0001	16,834	0,5382	0,9996	0,0004	0,006	0	0	16,9227	0,2312
	10	0,9979	0,0021	0,0143	0,000001	0,0002	18,7164	0,4836	0,9997	0,0003	0,0054	0	0	18,8095	0,151
	11	0,9972	0,0028	0,0155	0,00001	0,0002	20,5608	0,5397	0,9992	0,0008	0,0081	0,000005	0,0001	20,6537	0,3446
	12	0,9971	0,0029	0,0158	0	0	22,4452	0,4887	0,9991	0,0009	0,0082	0,000005	0,0001	22,5399	0,2736
	13	0,9454	0,0546	0,0871	0	0	31,7619	0,6072	0,9624	0,0375	0,0826	0	0	31,8651	0,39
	14	0,9936	0,0064	0,0236	0,000005	0,0001	26,1245	0,6951	0,9958	0,0042	0,0331	0	0	26,2604	0,3644
	15	0,9835	0,0165	0,0709	0	0	27,9828	0,6719	0,9901	0,0099	0,0542	0	0	28,1241	0,4439
	16	0,9728	0,0272	0,0834	0,000005	0,0001	29,9012	0,5944	0,9851	0,0149	0,045	0,000001	0,0002	30,0279	0,2876
	17	0,9229	0,0771	0,0889	0	0	30,5009	0,5651	0,9346	0,0654	0,0704	0	0	31,5266	0,4493
	18	0,8926	0,1074	0,1366	0,00002	0,0004	33,5451	0,9162	0,9291	0,0709	0,1024	0,00001	0,0003	33,695	0,6531
	19	0,8227	0,1773	0,1379	0	0	35,4102	0,8227	0,8736	0,1263	0,1384	0	0	35,5743	0,4919
	20	0,7419	0,2581	0,1516	0	0	37,2494	0,9043	0,8086	0,1914	0,1359	0	0	37,4507	0,3908

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,90	1	1	0	0	0	0	1,9204	0,2298	1	0	0	0	0	1,9406	0,1236
	2	1	0	0	0	0	3,8619	0,261	1	0	0	0	0	3,8888	0,12
	3	0,9998	0,0002	0,0257	0,0451	0,1035	5,7728	0,3817	1	0	0	0	0	5,8265	0,183
	4	0,9989	0,0011	0,0193	0,0229	0,0905	7,7231	0,3733	1	0	0	0	0	7,7744	0,2084
	5	0,9982	0,0018	0,0209	0,0057	0,0548	9,6547	0,0472	1	0	0	0,0001	0,0072	9,7132	0,2553
	6	0,9988	0,0012	0,0157	0,0032	0,0515	11,5886	0,5000	0,9998	0,0002	0,0052	0,0022	0,0439	11,6617	0,254
	7	0,9985	0,0015	0,0155	0,003	0,0536	13,5585	0,4045	0,9994	0,0006	0,011	0,001	0,0316	13,609	0,2541
	8	0,9978	0,0022	0,0171	0,01	0,0314	15,4747	0,4972	0,9987	0,0012	0,0136	0,0008	0,0268	15,5358	0,3068
	9	0,9983	0,0017	0,0135	0,001	0,0316	17,4369	0,461	0,9991	0,0009	0,0099	0,0008	0,0279	17,4928	0,2646
	10	0,9971	0,0029	0,0173	0	0	19,313	0,6147	0,9993	0,0007	0,0094	0,000005	0,0001	19,4179	0,3772
	11	0,9975	0,0025	0,0152	0	0	21,2837	0,5606	0,9991	0,0009	0,0085	0	0	21,3678	0,2945
	12	0,9973	0,0027	0,0146	0	0	23,2096	0,5995	0,999	0,001	0,0072	0	0	23,3059	0,3501
	13	0,9951	0,0049	0,019	0	0	24,2783	0,5807	0,998	0,002	0,012	0	0	24,4014	0,2991
	14	0,9928	0,0072	0,0344	0,000005	0,0001	27,0196	0,7314	0,9954	0,0046	0,0336	0	0	27,1575	0,4797
	15	0,9864	0,0136	0,0634	0	0	28,9885	0,7070	0,9919	0,0081	0,0531	0,000005	0,0001	29,121	0,4156
	16	0,9716	0,0284	0,0931	0,000005	0,0001	30,9239	0,8539	0,9823	0,0176	0,0729	0	0	31,0607	0,4249
	17	0,9567	0,0433	0,1047	0	0	32,832	0,8653	0,9735	0,0264	0,0701	0,000005	0,0001	32,9663	0,5203
	18	0,9077	0,0923	0,1043	0	0	33,5252	0,3439	0,9164	0,0836	0,0868	0	0	34,5345	0,2797
	19	0,8517	0,1483	0,1418	0	0	36,7481	0,6842	0,902	0,098	0,1253	0	0	36,8698	0,423
	20	0,7669	0,2331	0,1451	0,000005	0,0001	38,6719	0,7616	0,8277	0,1723	0,1447	0	0	38,8093	0,4633

EK-5. (devam) n=200 için simülasyon sonuçları

γ	aykırı gözlem	NW							LL						
		DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)	DTO	M	Sd(M)	S	Sd(S)	R	Sd(R)
0,95	1	1	0	0	0	0	1,9548	0,2597	1	0	0	0	0	1,9856	0,0533
	2	1	0	0	0	0	3,9217	0,3366	1	0	0	0	0	3,9525	0,2091
	3	1	0	0	0,0526	0,1348	5,8809	0,429	1	0	0	0,0299	0,0142	5,9415	0,2189
	4	1	0	0	0,0227	0,1021	7,878	0,4111	1	0	0	0,007	0,0108	7,94	0,1411
	5	0,9994	0,0006	0,0126	0,0039	0,0461	9,8611	0,4620	1	0	0	0,001	0,0032	9,9154	0,2075
	6	0,9993	0,0007	0,0117	0,0037	0,0572	11,8273	0,4734	1	0	0	0,0011	0,0316	11,8811	0,2734
	7	0,9992	0,0008	0,0127	0,001	0,0316	13,7707	0,6411	0,9999	0,0001	0,0078	0,00001	0,0002	13,853	0,3541
	8	0,9995	0,0005	0,0078	0,0026	0,0485	15,7853	0,5268	0,9996	0,0004	0,0068	0,0009	0,0309	15,8412	0,3771
	9	0,999	0,001	0,007	0,0019	0,0445	17,7709	0,5422	0,9997	0,0003	0,0049	0,0009	0,0312	17,8451	0,2915
	10	0,9989	0,0011	0,0104	0,001	0,0316	19,7017	0,7218	0,9992	0,0008	0,0089	0	0	19,8036	0,4344
	11	0,9981	0,0019	0,0127	0	0	21,705	0,6557	0,9988	0,0012	0,011	0	0	21,7688	0,5045
	12	0,9981	0,0019	0,021	0	0	23,6898	0,6144	0,9989	0,0011	0,0094	0	0	23,7651	0,4110
	13	0,9978	0,0022	0,0167	0	0	25,6655	0,6454	0,999	0,0009	0,0083	0	0	25,7416	0,4364
	14	0,9961	0,0038	0,0274	0	0	27,6464	0,6797	0,9973	0,0027	0,0312	0	0	27,7374	0,4081
	15	0,9927	0,0073	0,0368	0	0	29,6257	0,7327	0,9963	0,0036	0,0181	0	0	29,7206	0,4906
	16	0,9833	0,0166	0,0789	0,000005	0,0001	31,5767	0,7613	0,9914	0,0086	0,0487	0,000005	0,0001	31,7028	0,5493
	17	0,963	0,037	0,1068	0	0	33,5193	0,8809	0,9797	0,0303	0,0802	0	0	33,6635	0,5646
	18	0,9386	0,0613	0,1238	0	0	35,4939	0,8532	0,9651	0,0349	0,0932	0	0	35,6336	0,5737
	19	0,8828	0,1172	0,1618	0	0	37,4784	0,871	0,9341	0,0658	0,1070	0	0	37,6332	0,5134
	20	0,8053	0,1947	0,1562	0	0	39,4938	0,8777	0,8758	0,1242	0,1319	0	0	39,6363	0,4835

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : SERT, Sümeyra
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 01.01.1990, Ankara
 Medeni hali : Bekar
 Telefon : 0 (507) 701 7218
 e-mail : sumeyra.sert@selcuk.edu.tr



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek Lisans	Gazi Üniversitesi / İstatistik	Devam Ediyor
Lisans	Ankara Üniversitesi / İstatistik	2015
Lise	Fethiye Kemal Mumcu Anadolu Lisesi	2008

İş Deneyimi

Yıl	Yer	Görev
2018-Halen	Selçuk Üniversitesi	Araştırma Görevlisi
2016-2018	Bilgi Evi Kişisel Gelişim Kursu	Kurum Müdürü

Yabancı Dil

İngilizce, Almanca (başlangıç)

Yayınlar

Sert, S. ve Kardiye, F. , (2019, 28 Mart). *Parametrik Olmayan Doğrusal-Dairesel Regresyonda Aykırı Değer Problemi Üzerine Bir Çalışma*, 2nd International Congress on Statistics Mathematics and Analytical Methods, 1(1), 335, İstanbul.

Hobiler

Tiyatro, Gitar, Dans



GAZİ GELECEKTİR..