

**DOĞRUSAL KARMA MODELDE VARYANS
BİLEŞENLERİNİN SAĞLAM KESTİRİCİLER İLE
TAHMİNİ**

**ESTIMATION OF VARIANCE COMPONENTS IN
LINEAR MIXED MODEL WITH ROBUST ESTIMATORS**

FATMA AYOĞLU ÇELTİKÇİ

DOÇ. DR. SEMRA TÜRKAN

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

ÖZET

DOĞRUSAL KARMA MODELDE VARYANS BİLEŞENLERİNİN SAĞLAM KESTİRİCİLER İLE TAHMİNİ

Fatma AYOĞLU ÇELTİKÇİ

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Doç. Dr. Semra TÜRKAN

Temmuz 2020, 95 sayfa

Bu çalışmada doğrusal karma modelde varyans bileşenlerinin sağlam kestiriciler ile nasıl tahmin edildiği açıklanmış ve gerçek bir veri kümesi üzerinde uygulama yapılmıştır. Bu amaçla ilk olarak doğrusal karma modelde sabit, rasgele etki ve varyans bileşenlerinin tahmin yöntemlerine değinilmiştir. Daha sonra, doğrusal karma modelde aykırı değerler ve aykırı değerleri ortaya çıkarmak için kullanılan ölçütler tanıtılmıştır. Aykırı değerlerin olması durumunda aykırı değerleri çıkarmak yerine aykırı değerlerin parametre kestirimleri üzerindeki etkisini azaltan sağlam tahmin yöntemleri kullanılır. Tez çalışmasında doğrusal karma modelde aykırı değer olması ya da verilerin dağılımının çarpık olması durumunda kullanılan sağlam kestiriciler üzerinde durulmuştur. Tasarım uyarlamalı ölçek tahmini ile doğrusal karma modelde varyans bileşenlerinin sağlam tahminlerinin elde edilmesi için gerekli algoritma adımları açıklanmıştır.

Uygulamanın ilk bölümünde literatürde yer alan ve aykırı değerler içeren küçük bir veri setinin sağlam doğrusal karma model tahminleri verilmiştir. Bu verinin analiz edilmesinde temel amaç sağlam tahminin özellikle küçük örneklerde aykırı değerler olması durumunda daha iyi sonuçlar verdiğini göstermektedir. Uygulamanın ikinci bölümünde, daha önce Orman Genel Müdürlüğü tarafından yürütülen “Milli Ağaç Islah Programı” nda Antalya Bölgesi’ndeki Kızılcım ağaçları projesinde doğrusal karma model ile tahmin edilen sekizinci yaş boy değerlerine ilişkin parametre tahminleri, varyans bileşenleri, genetik parametreler ve ıslah değerleri sağlam doğrusal karma model ile tahmin edilmiştir.

Anahtar Kelimeler: Doğrusal Karma Model, Varyans Bileşenleri, Sağlam Tahmin Ediciler, Islah Değeri, Genetik Parametreler

ABSTRACT

ESTIMATION OF VARIANCE COMPONENTS IN LINEAR MIXED MODEL WITH ROBUST ESTIMATORS

Fatma AYOĞLU ÇELTİKÇİ

Master's Degree, Department of Statistics

Supervisor: Doç. Dr. Semra TÜRKAN

July 2020, 95 pages

In this study, how to estimate the variance components of the linear mixed model with robust estimators is explained and an application is made on a real dataset. For this purpose, firstly, estimation methods of fixed, random effect and variance components in linear mixed model are mentioned. Then, outliers and diagnostics used to reveal outliers are introduced in the linear mixed model. In case of outliers, instead of removing outliers, robust estimation methods are used to reduce the impact of outliers on parameter estimates. In this thesis study, if there is an outlier in the linear mixed model or the distribution of the data is skewed, the robust estimators are emphasized. The algorithm

steps required to obtain robust estimates of variance components in linear mixed model with design adaptive scale estimation (DAS) are explained.

In the first part of the application, robust linear mixed model estimates of a small data set containing outliers in the literature are given. The main purpose of analyzing this data is to show that the robust prediction gives better results, especially in small samples, in case of outliers. In the second part of the application, the parameter estimates, variance components, genetic parameters and breeding values related to the eighth age values estimated with a robust linear mixed model in the Red Pine trees project in Antalya Region in the "National Tree Breeding Program" previously carried out by the General Directorate of Forestry.

Keywords: Linear Model, Variance Components, Robust Estimators, Breeding Value, Genetic Parameters

TEŐEKKÜR

Yüksek lisans tez hazırlığıım süresince rehberliđi, bilgisi, katkıları, manevi desteđi ve yön göstericiliđi ile her zaman güven veren ve yanımda olan deđerli danışmanım Doç.Dr.Semra TÜRKAN'a,

Tez çalışmam sırasında ormancılık alanında bilgi ve deneyimlerini esirgemeyen deđerli mesai arkadaşım Orman Mühendisi Kubilay ÖZYALÇIN'a, çalışmamda kullanmak üzere veri sağladığı için Orman Genel Müdürlüğü Orman Ağaçları ve Tohumları İslah Araştırma Enstitüsü'ne,

Eleştiri ve önerileriyle tezime katkı sağladıkları için deđerli jüri üyelerim Prof.Dr.Kürşad ÖZKAN, Prof.Dr.Gamze ÖZEL KADILAR, Dr.Öğr. Üyesi Esra POLAT ve Dr.Öğr. Üyesi Ebru GÜL'e,

Ve eğitim hayatım boyunca yoluma ışık olan başta annem olmak üzere aileme ve tez hazırlığıım süresince her zaman destek olan eşime,

Sonsuz teşekkürler...

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT	iii
TEŞEKKÜR	v
ŞEKİLLER DİZİNİ.....	ix
ÇİZELGELER DİZİNİ	x
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ	1
2. DOĞRUSAL KARMA MODEL	6
2.1. Doğrusal Karma Modellerde Parametre Kestirimi.....	7
2.1.1. En İyi Önkestirici (BP) Yöntemi.....	8
2.1.2. En İyi Doğrusal Önkestirici (BLP) Yöntemi.....	8
2.1.3. En İyi Doğrusal Yansız Önkestirici (BLUP) Yöntemi.....	9
2.1.4. Henderson’ın Karma Model Eşitlikleri	11
2.2. Varyans Bileşenlerinin Kestirimi	14
2.2.1. En Çok Olabilirlik Yöntemi ile Varyans Bileşenlerinin Kestirimi	14
2.2.1.1. Hartley –Rao Formu.....	17
2.2.2. Kısıtlanmış En Çok Olabilirlik Yöntemi ile Varyans Bileşenlerinin Tahmini.....	18
2.2.3. En Çok Olabilirlik Yöntemi (ML) ile Kısıtlanmış En Çok Olabilirlik Yöntemi (REML) Arasındaki Farklar.....	20
2.3. İteratif Yöntemler	21
2.3.1. Newton-Raphson Algoritması.....	21
2.3.2. Skorlama Algoritması	22
2.3.3. EM Algoritması.....	23
2.3.1.1. En Çok Olabilirlik Tahmini için EM Algoritması.....	23

2.3.1.2. Kısıtlanmış En Çok Olabilirlik Tahmini için EM Algoritması.....	24
3. DOĞRUSAL KARMA MODELDE SAĞLAM KESTİRİCİLER	26
3.1. Aykırı Değer	27
3.1.1. Sabit Etkiler Modeli için Aykırı Değer Ölçütleri	27
3.1.1.1. Şapka (Hat) Matrisinin Köşegen Elemanları	30
3.1.1.2. Student Türü Artıklar	31
3.1.1.3. Cook Uzaklığı	31
3.1.1.4. DFFITS Ölçütü	32
3.1.1.5. COVRATIO Ölçütü.....	33
3.1.2. Karma Model için Aykırı Değer Ölçütleri.....	33
3.1.2.1. Şapka Matrislerinin Köşegen Elemanları.....	34
3.1.2.2. Student Türü Artıklar.....	35
3.1.2.3. Cook Uzaklığı Ölçütü.....	35
3.1.2.4. DFFITS Ölçütü.....	36
3.1.2.5. COVTRACE ve COVRATIO Ölçütleri.....	36
3.2. Sağlam (Robust) Yaklaşım	37
3.3. Sağlamlık Kriterler	37
3.3.1. Bozulma Noktası.....	38
3.3.2. Etki Fonksiyonu.....	40
3.3.3. Göreli Etkinlik	41
3.4. Sağlam Regresyon	41
3.4.1. M Tahmin Edicileri.....	42
3.4.2. S Tahmin Edicileri	44
3.4.3. MM Tahmin Edicileri	45
3.4.4. Tasarım Uyarlamalı Ölçek Tahmini (Design Adaptive Scale Estimation).....	47
3.4.5. SMDM Tahmin Edicileri	49
3.5. Doğrusal Karma Modelde Sabit ve Rasgele Etkinin Sağlam Tahmini.....	49
3.5.1. Tahmin Denklemlerinin Sağlamlaştırılması	51
3.5.1.1. Ölçek ve Kovaryans Parametrelerinin Sağlam Tahmini.....	53
3.5.1.2. Tasarım Uyarlamalı Ölçek (DAS) Tahmini ile Ölçek ve Kovaryans Parametrelerinin Sağlam Tahmini.....	55
3.5.1.3. Tahmin Algoritması.....	57

4. UYGULAMA.....	61
4.1. Küçük Veri Seti Uygulaması.....	61
4.2. Gerçek Veri Seti Uygulaması.....	65
4.2. Tahmin Sonuçları.....	77
5. SONUÇ VE ÖNERİLER	84
KAYNAKLAR.....	87
ÖZGEÇMİŞ	95

ŞEKİLLER DİZİNİ

Şekil 1.1.	Sabit ve rasgele etkiler için karar verme.....	1
Şekil 3.1.	(a) Beş gözleme ilişkin en küçük kareler kestirimi. (b) y-yönünde aykırı değer olduğunda en küçük kareler kestirimi.....	28
Şekil 3.2.	(a) Beş gözleme ilişkin en küçük kareler kestirimi. (b) y-yönünde aykırı değer olduğunda en küçük kareler kestirimi. (c) İyi kaldıraç noktası olduğunda en küçük kareler kestirimi.....	30
Şekil 4.1.	Aeresol veri seti için aykırı değer belirleme ölçütlerine ilişkin saçılım grafikleri a) Student türü artıkların saçılım grafiği, b) Cook Uzaklığı saçılım grafiği, c) COVRATIO değerleri saçılım grafiği.....	63
Şekil 4.2.	İbrelî Ağaç Türlerinin Yayılış Alanları Haritası.....	67
Şekil 4.3.	Tohum meşceresinde plus ağaç seçimi ve işaretlenmesi.....	69
Şekil 4.4.	Tohum Bahçesi Görüntüsü.....	70
Şekil 4.5.	Aşılı Fidanlar.....	73
Şekil 4.6.	Aykırı değer belirleme ölçütlerine ilişkin saçılım grafikleri a) Student türü artıkların saçılım grafiği, b) Cook Uzaklığı saçılım grafiği, c) COVRATIO değerleri saçılım grafiği.....	80

ÇİZELGELER DİZİNİ

Çizelge 4.1.	Aerosol veri seti.....	61
Çizelge 4.2.	Doğrusal Karma Model ve Sağlam Doğrusal Karma Model Sonuçları...	64
Çizelge 4.3.	Gözlem değerlerinin sağlam ağırlık değerleri.....	65
Çizelge 4.4.	İbrelî Ağaç Türlerinin Alansal Dağılımı.....	66
Çizelge 4.6.	Altı Adet Klonal Tohum Bahçesi Dağılımı.....	70
Çizelge 4.6.	Antalya Deneme Alanı Özellikleri.....	71
Çizelge 4.7.	Boy Karakterine İlişkin Betimleyici İstatistikler.....	77
Çizelge 4.8.	İç İçe Karma Doğrusal Modele İlişkin ANOVA Tablosu.....	78
Çizelge 4.9.	Doğrusal Karma Model ve Sağlam Doğrusal Karma Model Sonuçları...	81
Çizelge 4.10.	Genetik Parametrelerin Tahmin Sonuçları.....	81
Çizelge 4.11.	İslah Değerleri, Mutlak İslah Değerleri Tahmini ve Genetik Kazanç....	83

SİMGELER VE KISALTMALAR

Kısaltmalar

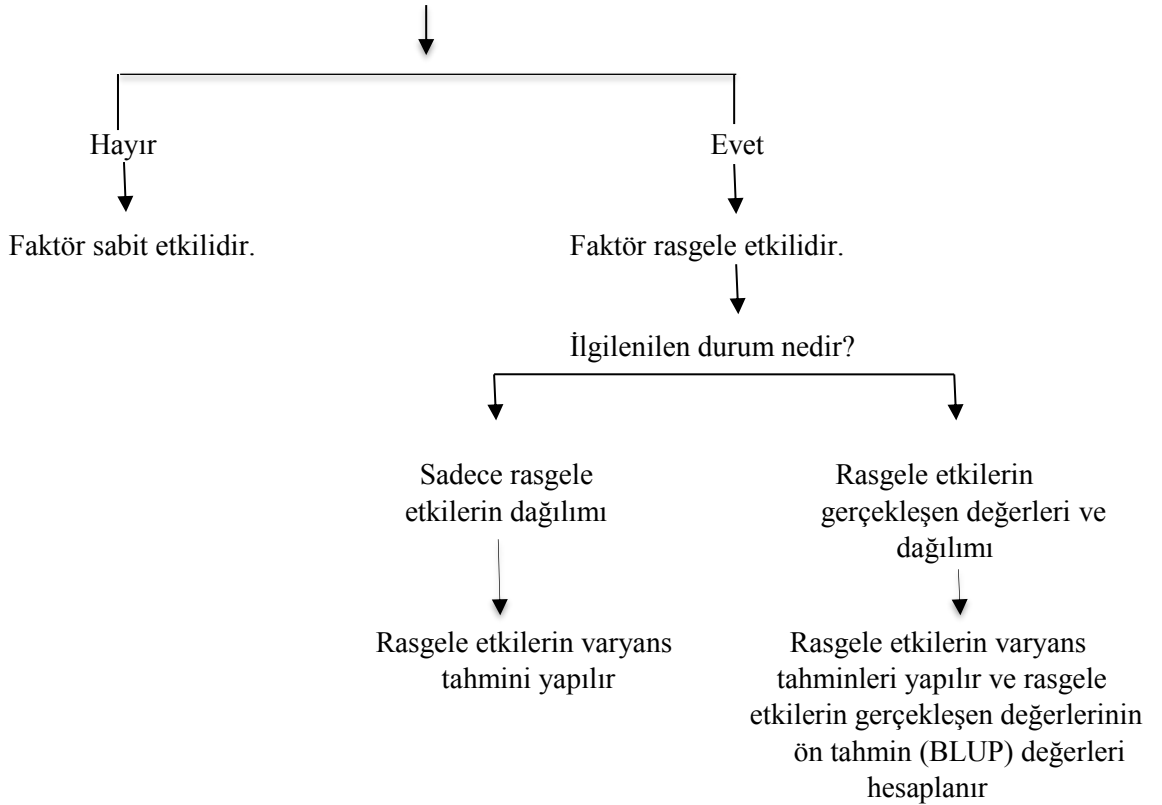
LS	En Küçük Kareler (Least Square)
BP	En İyi Ön Tahmin (Best Prediction)
BLP	En İyi Doğrusal Ön Tahmin (Best Linear Prediction)
BLUE	En İyi Doğrusal Yansız Tahmin (Best Linear Unbiased Estimation)
BLUP	En İyi Doğrusal Yansız Ön Tahmin (Best Linear Unbiased Prediction)
BV	Islah Değeri (Breeding Value)
DAS	Tasarım Uyarlamalı Ölçek (Design Adaptive Scale)
EM	Beklenti-Maksimizasyon (Expectation-Maximization)
GK	Genetik Kazanç (Genetic Gain)
IF	Etki Fonksiyonu (Influence Function)
Kov	Kovaryans (Covariance)
MBV	Mutlak Islah Değeri (Absolute Breeding Value)
ML	En Çok Olabilirlik (Maximum Likelihood)
REML	Kısıtlanmış En Çok Olabilirlik (Restricted Maximum Likelihood)
Sim.	Simetrik
Var	Varyans (Variance)

1. GİRİŞ

İstatistiksel modelleme, veriyi tanımlama, veri yapısını temsil eden matematiksel formu belirleme ve değişkenler arasındaki ilişkinin fonksiyonel yapısını belirleme aşamalarını içerir [1].

İstatistiksel modeller oluşturulurken sabit etki ve rasgele etki olmak üzere iki farklı etki düzeyi vardır. Sabit etkiler bir faktörün sonlu düzeyler kümesini belirtirken, rasgele etkiler bir faktörün sonsuz düzeyler kümesini belirtir. Bir faktörün sabit mi rasgele mi olduğuna karar vermek için izlenen yol haritası Şekil 1.1 de kısaca verilebilir:

Faktörün düzeylerinin bir olasılık dağılımından geldiği varsayımı mantıklı mıdır?



Şekil 1.1. Sabit ve rasgele etkilere karar verme [2]

Sabit etkiler, ilgilenilen faktörün düzeylerinin sonlu kümesinden özel olarak seçilen etkilerdir. Rasgele etkiler ise ilgilenilen faktör düzeylerinin genellikle sonsuz kümesinden rasgele seçilen etkilerdir. Örneğin, üç farklı sıcaklık derecesinde pişirilmiş (sonsuz)

ekmek yığımları içerisinde rasgele seçilen altı ekmeğin her birinden 4 somun ekmeği seçilsin. Bu örnekte ekmekler belirli sıcaklık derecelerinde pişirildiği için sıcaklık sabit etki, sonsuz sayıdaki ekmeğin rasgele seçilen ekmeğin ise rasgele etkidir. Burada sabit etki olan sıcaklık derecelerinin etkisi tahmin edilirken, rasgele etki olan ekmeğin varyansı tahmin edilir. Bu nedenle bu veride rasgele değişimin iki kaynağı vardır: Ekmek yığın varyansı ve hata varyansı. Bu iki varyans, varyans bileşenleri olarak bilinir [1] [2].

Sadece sabit etkilerin yer aldığı modele sabit etkili model, sadece rasgele etkilerin yer aldığı modele rasgele etkili model, her iki etki düzeyinin yer aldığı modele ise karma model denir. Karma model hem sabit etkileri hem de rasgele etkileri içeren model olduğu için uygulamada daha çok kullanılmaktadır. Çünkü çoğu zaman hem sabit etkiler hem de rasgele etkiler aynı modelde yer alır. Aslında, ortalamayı içeren her model karma bir modeldir. Çünkü her model hata terimi de içereceğinden ortalama sabit etkiyi, hata da rasgele etkiyi ifade etmek üzere sabit ve rasgele etkiler aynı anda modelde yer almış olur. Ancak, uygulamada karma model genellikle hem sabit etkiler hem de rasgele artıkların yanı sıra rasgele etkilere sahip olan modeller için kullanılmaktadır[2][17].

Çalışmanın birinci bölümü olan giriş bölümünde genişletilmiş özet halinde tezin alt bölümleri kısaca açıklanmıştır ve literatür taraması verilmiştir.

Çalışmanın ikinci bölümünde doğrusal karma modeller açıklanmıştır. İlk olarak sabit etkili doğrusal model hakkında kısaca bilgi verilmiş daha sonra doğrusal karma modellerde sabit etkilerin tahmini, rasgele etkili parametrelerin ön kestirimi ve rasgele etkilerin varyans bileşenlerinin tahmini için kullanılan yöntemler açıklanmıştır.

Çalışmanın üçüncü bölümünde, doğrusal karma modellerde varyans bileşenlerinin sağlam yaklaşım ile tahmini üzerinde durulmuştur. Aykırı değer kavramı açıklanmış, doğrusal karma modelde aykırı değerleri ortaya çıkarmak için önerilen ölçütlere değinilmiştir. Aykırı değerlerin varlığında daha iyi sonuç veren sağlam yöntemler tanıtılmıştır. Doğrusal karma modelde varyans bileşenlerinin tasarım uyarlamalı ölçek tahmini ile elde edilmesi için önerilen algoritma adımları açıklanmıştır.

Çalışmanın dördüncü bölümü olan uygulama bölümünde ilk olarak sağlam tahmin yönteminin özellikle küçük örneklerde aykırı değerler olması durumunda klasik yöntemlere göre daha doğru sonuçlar verdiğini göstermek için literatürde yer alan ve aykırı değerler içeren küçük bir veri seti sağlam karma model ile tahmin edilmiştir. Daha sonra Orman Genel Müdürlüğü Orman Ağaçları ve Tohumları İslah Araştırma Enstitüsü tarafından yürütülen “Türkiye Milli Ağaç İslahı Programı” kapsamında aykırı değerler içeren ve doğrusal karma model ile tahmin edilen kızılçamla ilişkin veriler incelenmiştir. Proje kapsamında veriler analiz edilirken aykırı değerler çıkarılarak sonuçlar elde edilmiştir. Bu çalışmada kızılçam verilerinin analizinde, aykırı değerleri çıkarmak yerine aykırı değerlerin etkisini azaltan sağlam doğrusal karma model kullanılmıştır. Bu bölümde ayrıca ağaç ıslahının önemi, ıslah programı kapsamında hesaplanan genetik parametreler, ıslah değeri ve genetik kazanç kavramları açıklanmıştır. İslah çalışması kapsamında Antalya Bölgesi’nden elde edilen kızılçam ağaçlarına ilişkin sekinci yaş boy verileri analiz edilmiştir. Boy verilerinin analizinde kullanılan doğrusal karma modelde blok, set ve aile faktörleri bağımsız değişkenler olup; blok, set etkisi sabit ve aile etkisi rasgele etki olarak alınmıştır. Doğrusal karma model tahminleri üzerinde etkili olan gözlemlerin varlığı, literatürde önerilen aykırı değer ölçütleri ile belirlenmiştir. Verilerde aykırı değerler olması nedeniyle veriler sağlam doğrusal karma model tahmin edicisi ile tahmin edilmiştir. Doğrusal karma model ve sağlam doğrusal karma model ile elde edilen sonuçlar incelenmiştir. Genetik parametreler, ıslah değerleri ve genetik kazanç değerleri doğrusal karma model ve sağlam doğrusal karma model tahminlerine göre hesaplanmıştır. Tüm hesaplamalarda SAS 9.4 ve R programları kullanılmıştır.

Çalışmanın beşinci bölümünde, dördüncü bölümde elde edilen sonuçlar tartışılmıştır. Bu bölümde Orman Genel Müdürlüğü Orman Ağaçları ve Tohum İslahı Araştırma Enstitüsü Müdürlüğü tarafından yürütülen ıslah çalışmalarında veride aykırı değerler olması durumunda sağlam doğrusal karma modellerin kullanılması önerilmiştir.

Bu çalışmada doğrusal karma modelde varyans bileşenlerinin sağlam kestiriciler ile tahmini ele alınmıştır. Bu alanda son yıllarda yapılan diğer çalışmalar aşağıda verilmiştir:

Paramjit ve Gill[3], boylamsal verilerde karma doğrusal model parametrelerini tahmini için sağlam yöntemleri ele almışlardır.

Waterman[4], karma doğrusal modelde sağlam regresyonu ele almış ve doğrusal karma modelde sağlam regresyon (MMRR) tahminleri, doğrusal karma model ve yerel ağırlıklı en çok olabilirlik tahmininden elde edilen tahminlerin dışbükey kombinasyonlarını inceleyerek gerçek bir veri üzerinde uygulama yapmıştır.

Koller[5], doktora tezinde doğrusal karma etkili modelleri sağlam bir şekilde tahmin etmek için bir yöntem geliştirmiştir ve bu yöntem çeşitli hiyerarşik, iç içe veya çapraz veri yapıları için uygulanabilir. Ayrıca sağlam ölçek tahmini olan “Tasarım Uyarlamalı Ölçek Tahmini (DAS)” ni doğrusal karma modelde varyans bileşenlerinin tahminini elde etmek için karma model eşitliklerine uyarlamıştır. Daha sonra Koller[6] önerdiği sağlam tahmin yöntemlerinin elde edildiği “robustlmm” R paket programını yazmıştır.

Ayrıca çalışmanın uygulama kısmında ele alınan kızılçam ağaç ıslahı konusunda literatürde doğrusal karma model kullanılarak yapılan birçok çalışma bulunmaktadır. Ancak bu alanda doğrusal karma model tahminlerinin sağlam yöntem ile elde edildiği bir çalışmaya rastlanmamıştır. Kızılçam ağaç ıslahı konusunda son yıllarda yapılan çalışmalar aşağıdaki gibi özetlenmiştir:

Işık[7], kızılçamda (Pinus burtia Ten.) genetik çeşitlilik, kalıtım derecesi ve genetik kazancı hesaplamak için doğrusal karma modelden yararlanmıştır.

Öztürk [8], kızılçam (Pinus brutia Ten.) ağaç türü için ıslah değerlerini BLUP yöntemi ile tahmin etmiştir. ((Arka plan gri kalmıştı beyaza çevirdim)

Öztürk vd. [9] kızılçam (Pinus brutia Ten.) ağaç türü için ıslah değerlerini, genetik kazancı ve kalıtım derecelerini doğrusal karma modeli kullanarak tahmin etmiştir.

Ercanlı vd. [10] ormancılık alanında artım ve büyümenin modellenmesinde karma model eşitlikleri kullanmışlardır.

Sergio, German ve Pukkala [11] Güney Amerika’da yerli bir tür olan *Centrolobium tomentosum* için büyüme ve verimlilik için sabit etkili, karma etkili model için marjinal etkileri, karma etkili model için koşullu tahminleri elde etmişlerdir.

Banh ve Hug[12], ormancılık verisini kullanarak R programında otokorelasyonun olduğu ve olmadığı durumda rasgele etkiler incelemiştir.

Doğrusal karma model uygulama alanı oldukça geniş, analizlerde çok tercih edilen bir yöntemdir çünkü neredeyse her model sabit ve rasgele etkilerin ikisini de içerir. Sağlam tahmin ediciler ise veri seti küçük olduğunda, normallik sağlanmadığında ya da aykırı değer varlığında güvenilir sonuçlar elde edilmesini sağlar. Gerçek veri setlerinde her zaman normalliğin sağlanamaması, ya da aykırı değerlerin varlığı çok sık karşılaşılan bir sorundur, doğrusal karma modelde sağlam kestiriciler ise aykırı değer varlığında aykırı değerleri atmadan güvenilir sonuçlar vermek için kullanılır. Bu çalışmada aykırı değerlerin varlığında daha güvenilir sonuçlar veren sağlam yöntemin doğrusal karma modellerdeki uygulamasının gösterilmesi amaçlanmıştır. Doğrusal karma modellerde sağlam tahmin yöntemleri ile ilgili son yıllarda bazı çalışmalar yapılmıştır. Ancak bu tez çalışmasında doğrusal karma modelde, Koller [5] tarafında önerilen sağlam tahmin yöntemi ele alınmıştır. Bunun temel nedeni, Koller [6] tarafından yazılan “robustlmm” R paket programında yöntemin kolaylıkla uygulanabilir olmasıdır. Sağlam doğrusal karma modelin ormancılık alanında bir veri setine uygulanmasının temel nedeni ise, bu alanda doğrusal karma modellerin özellikle ıslah değerlerinin tahmininde sıkça kullanılıyor olmasıdır. Bu verilerde de genellikle aykırı değerler olduğundan, aykırı değerlerin varlığında daha güvenilir sonuçlar veren sağlam doğrusal karma modeller ile verinin tahmin edilebileceğinin gösterilmesidir.

Karma doğrusal modelde sağlam kestiriciler ile parametre tahmini doğrusal karma modelin uygulandığı her veri setinde, yani sabit ve rasgele etkinin olduğu her modelde uygulanabilir.

2. DOĞRUSAL KARMA MODEL

Sadece sabit etkinin olduğu model genel olarak

$$y = X\beta + \varepsilon \quad (2.1)$$

şeklinde ifade edilir. Burada y , $n \times 1$ boyutlu gözlemlenebilir yanıt vektörü, X $n \times p$ boyutlu sabit etkilere ilişkin tasarım matrisi, β $p \times 1$ boyutlu sabit etkilere ilişkin parametreler vektörü, $\varepsilon = y - E(y) = y - X\beta$ hata vektörüdür. Burada $E(\varepsilon) = 0$ ve $\text{var}(\varepsilon) = \sigma^2 I_n$ 'dir. Sabit etkilerde parametre tahminleri genellikle en küçük kareler (LS) yöntemi kullanılarak elde edilir.

Sabit etkili bir modele rasgele etkili bir parametre eklendiğinde ise karma etkili bir model elde edilir. Doğrusal karma model (linear mixed model)

$$y = X\beta + Zu + \varepsilon \quad (2.2)$$

şeklinde ifade edilir. Burada y $n \times 1$ boyutlu gözlemlenebilir yanıt vektörü, X $n \times p$ boyutlu sabit etkilere ilişkin tasarım matrisi, β $p \times 1$ boyutlu sabit etkilere ilişkin parametreler vektörü, Z $n \times q$ boyutlu rasgele etkilere ilişkin tasarım matrisi, u $q \times 1$ boyutlu rasgele etkilere ilişkin parametreler vektörü ve ε $n \times 1$ boyutlu rasgele hatalara ilişkin hata vektörüdür. u rasgele etkiler vektörü $u' = [u'_1, u'_2, \dots, u'_r]$ şeklinde r tane alt vektöre bölünebilir ve her bir alt vektör u_i $q_i \times 1$, tek bir faktörün tüm düzeylerini göstermektedir. Bu durumda r , rasgele faktörlerin sayısıdır. Buna bağlı olarak $Z = [Z_1, Z_2, \dots, Z_r]$ olup i . rasgele faktör için Z_i $n \times q_i$ 'dir [13].

Doğrusal karma modelde u ve ε 'nin

$$\begin{bmatrix} u \\ \varepsilon \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} D & 0 \\ 0 & R \end{bmatrix} \right)$$

şeklinde bağımsız ve normal dağılıma sahip olduğu varsayılır. u 'nin varyans-kovaryans matrisi D , $\text{var}(\varepsilon) = \sigma^2 I_n$, $\text{var}(u_i) = \sigma_i^2 I_{q_i}$ $\forall i$, $\text{kov}(u_i, u_j) = 0$ $\forall i \neq j$ ve $\text{kov}(u, e^i) = 0$ olmak üzere,

$$D = \text{var}(u) = \begin{bmatrix} \sigma_i^2 \mathbf{I}_{q_1} & 0 & 0 \\ & \sigma_i^2 \mathbf{I}_{q_2} & 0 \\ \text{Sim.} & & \sigma_i^2 \mathbf{I}_{q_r} \end{bmatrix}, \quad i = 1, 2, \dots, r \quad (2.4)$$

dir. Buna göre y yanıt değişkeninin beklenen değeri,

$$\begin{aligned} E(y) &= E(X\beta + Zu + \varepsilon) = E(X\beta) + E(Zu) + E(\varepsilon) = XE(\beta) + ZE(u) + E(\varepsilon) \\ &= XE(\beta) + ZE(u) + E(\varepsilon) = X_{n \times p} \beta_{p \times 1} + Z_{n \times q} \mathbf{0}_{q \times 1} + \mathbf{0}_{n \times 1} = (X\beta)_{n \times 1} \end{aligned} \quad (2.5)$$

ve varyansı,

$$\begin{aligned} V(y) &= V(X\beta + Zu + \varepsilon) = V(X\beta) + V(Zu) + V(\varepsilon) = \mathbf{0} + ZV(u)Z' + V(\varepsilon) \\ &= Z_{n \times q} D_{q \times q} Z'_{q \times n} + R_{n \times n} = V_{n \times n} \end{aligned} \quad (2.6)$$

olarak elde edilir. (2.6)'daki varyans-kovaryans matrisi

$$V(y) = V = ZDZ' + \sigma_e^2 \mathbf{I} = \sum_{i=1}^r \sigma_i^2 Z_i Z_i' + \sigma_e^2 \mathbf{I}_n \quad (2.7)$$

şeklinde de ifade edilir. $u_0 \equiv e$, $Z_0 \equiv \mathbf{I}_n$ ve $\sigma_0^2 \equiv \sigma_e^2$ olmak üzere (2.1)'deki doğrusal karma model

$$y = X\beta + \sum_{i=0}^r Z_i u_i \quad (2.8)$$

şeklinde ifade edilebilir. Bu durumda (2.7)'deki varyans-kovaryans matrisi

$$V = \sum_{i=0}^r Z_i Z_i' \sigma_i^2 \quad (2.9)$$

şeklinde gösterilir [2] [14] [16].

2.1. Doğrusal Karma Modellerde Parametre Kestirimi

Doğrusal karma modellerde hem sabit etkilerin hem de rasgele etkilerin kestirimlerinin elde edilmesi gerekir. Sabit etkiler için β 'nin kestirimi, rasgele etkiler için u ve varyans-kovaryans matrisinin (V) bilinmediği durumlarda varyans bileşenlerinin de kestirimlerinin elde edilmesi gerekir. Doğrusal karma modellerde en iyi ön tahmin (BP)

ve en iyi doğrusal ön tahmin (BLP) ile sadece rasgele etkili parametrelerin önkestirimi elde edilebilirken, en iyi doğrusal yansız ön tahmin (BLUP) yöntemi ve Henderson'ın doğrusal karma model eşitlikleri ile hem sabit hem de rasgele etkilerin parametre kestirimleri aynı anda elde edilebilmektedir. Varyans bileşenleri tahmininde ise en yaygın kullanılan tahmin yöntemleri en çok olabilirlik yöntemi(ML) ve kısıtlanmış en çok olabilirlik yöntemidir(REML).

2.1.1. En İyi Önkestirici (BP) Yöntemi

u ve y 'nin bileşik olasılık yoğunluk fonksiyonu $f(u, y)$ olmak üzere u 'nun en iyi önkestiricisi \tilde{u} 'nın hata kareler ortalaması,

$$E(\tilde{u} - u)^2 = \iint (\tilde{u} - u)^2 f(u, y) dy du \quad (2.10)$$

biçimindedir. Raslantı değişkenler vektörü u 'nun hata kareler ortalaması matris-vektör gösterimi ile

$$E(\tilde{u} - u)' A(\tilde{u} - u) = \iint (\tilde{u} - u)' A(\tilde{u} - u) f(u, y) dy du \quad (2.11)$$

şeklinde ifade edilir. A herhangi bir pozitif tanımlı simetrik matristir. Burada "en iyi" ifadesi, ön tahminin hata kareler ortalamasının minimum olması anlamına gelmektedir.

$\tilde{u} = E(u|y)$, y bilindiğinde u ' nun koşullu beklenen değeridir. \tilde{u} 'nın tahmini $f(u, y) = f(u|y)f(y)$ bileşik yoğunluk fonksiyonundan elde edilir. Buna göre

$$\begin{bmatrix} u \\ y \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_U \\ \mu_Y \end{bmatrix}, \begin{bmatrix} D & C \\ C' & V \end{bmatrix} \right) \quad (2.12)$$

varsayımı altında,

$$\tilde{u} = E(u|y) = \mu_U + CV^{-1}(y - \mu_Y) \quad (2.13)$$

şeklinde elde edilir. Burada $C = DZ'$ dir [1].

2.1.2. En İyi Doğrusal Önkestirici (BLP) Yöntemi

En iyi doğrusal önkestirici (BLP) yönteminde u rasgele vektörünün en iyi doğrusal önkestiricisi \tilde{u} , y 'nin doğrusal bir fonksiyonu, $\tilde{u} = a + By$, olmak üzere hata kareler

ortalamasını en küçük yapan değerdir. Buna göre $\tilde{u} = a + By$ rasgele etkiler vektörünün hata kareler ortalaması

$$q = E(\tilde{u} - u)' A(\tilde{u} - u) \quad (2.14)$$

$$= E(a + By - u)' A(a + By - u)$$

şeklinde ifade edilir. (2.14)'teki hata kareler ortlamasının u'ya göre türevi alınıp sıfıra eşitlenirse en iyi doğrusal önkestirim,

$$\tilde{u} = a + By = \mu_u + CV^{-1}(y - \mu_y) \quad (2.15)$$

olarak elde edilir. Normallik varsayımına bağlı olmadan elde edilen BLP ile normallik varsayımı altında elde edilen BP aynıdır [2].

2.1.3. En İyi Doğrusal Yansız Önkestirici (BLUP) Yöntemi

En iyi doğrusal yansız önkestirici (BLUP), $w = L'\beta + u$ 'nun kestirimidir. Burada $E(u) = 0$, L' $q \times p$ boyutlu bilinen bir matris ve $L'\beta$ tahmin edilebilir bir fonksiyondur. w hem sabit etkileri hem de rasgele etkileri içerdiği için, w için tahmin (estimate) mi yoksa kestirim (prediction) ifadesinin mi kullanılacağı tartışma konusu olabilir. Çünkü karma modellerde sabit etkiler için tahmin, rasgele etkiler için kestirim ifadesi kullanılmaktadır. \tilde{w} 'nun yansız, en iyi ve doğrusal olmasından dolayı kestirim ifadesi kullanılacaktır. \tilde{w} 'nın yansızlık özelliğini sağlaması için $E(\tilde{w}) = E(w)$ olmalıdır [2] [14].

$$E(w) = E(L'\beta + u) = E(L'\beta) + E(u) = L'\beta + 0 = L'\beta \quad (2.16)$$

dır. Burada

$$\begin{bmatrix} w \\ y \end{bmatrix} \sim N \left(\begin{bmatrix} L'\beta \\ X\beta \end{bmatrix}, \begin{bmatrix} D & C \\ C' & V \end{bmatrix} \right)$$

olup $E(\tilde{w}) = E(w)$ olduğundan \tilde{w} 'nin yansızlığı $a + BX\beta = L'\beta$ 'ya bağlıdır ve eğer a değeri β 'ya bağlı değilse $a=0$ ve $L' = BX$ 'dir. Bu durumda

$$E(w) = E(L'\beta + u) = L'\beta = BX\beta$$

Şeklinde elde edilir. $E(\tilde{w})$ ise

$$E(\tilde{w}) = E(a + By) = E(a) + BE(y) = 0 + BXB \quad (2.17)$$

olur. \tilde{w} 'nin en küçük özelliğini sağlaması için

$$E(\tilde{w} - w)'A(\tilde{w} - w) = \text{Var}(\tilde{w} - w)$$

hata kareler ortalamasını en küçük yapan \tilde{w} değeri bulunur.

$$\text{Var}(a + By - L'\beta - u) = B\text{Var}(y)B' + \text{Var}(u) - 2\text{Kov}(u, y) = BVB' + D - 2BC \quad (2.18)$$

hata kareler ortalamasını $L' = BX$ kısıtı altında en küçük yapan değeri bulmak için Lagrange çarpanı m kullanılır.

$$E(\tilde{w} - w)'(\tilde{w} - w) = \text{Var}(\tilde{w} - w) = \theta = BVB' + D - 2BC + 2m'(X'B' - L) \quad (2.19)$$

(2.19) eşitliğinin sırasıyla B ve m 'e göre türevi alınıp sıfıra eşitliğinde B ve m değerleri aşağıdaki gibi bulunur:

$$\frac{\partial E(\tilde{w} - w)'(\tilde{w} - w)}{\partial B} = 0 \quad (2.20)$$

$$2VB' - 2C + 2Xm = 0 \quad \Longrightarrow \quad B' = V^{-1}C - V^{-1}Xm$$

ve

$$\frac{\partial E(\tilde{w} - w)'(\tilde{w} - w)}{\partial m} = 0 \quad (2.21)$$

$$2X'B' - 2L = 0 \quad \Longrightarrow \quad L = X'B'$$

şeklindedir. Buradan,

$$m = (X'V^{-1}X)^{-1}(X'V^{-1}C) - (X'V^{-1}X)^{-1}L \quad (2.22)$$

$$B' = V^{-1}C - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}C + V^{-1}X(X'V^{-1}X)^{-1}L \quad (2.23)$$

olarak elde edilir. Bu durumda en iyi doğrusal yansız ön tahmin edici $\tilde{w} = a + By$ eşitliğinde B değişkeni yerine yazılırsa;

$$\begin{aligned} \tilde{w} &= C'V^{-1}y - C'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y + L'(X'V^{-1}X)^{-1}X'V^{-1}y \\ &= C'V^{-1}[y - X(X'V^{-1}X)^{-1}X'V^{-1}y] + L'(X'V^{-1}X)^{-1}X'V^{-1}y \end{aligned} \quad (2.24)$$

şeklinde elde edilir.

β 'nın genelleştirilmiş en küçük kareler yöntemiyle bulunan en iyi doğrusal yansız kestiricisi β^0 ,

$$\beta^0 = (X'V^{-1}X)^{-1}X'V^{-1}y \quad (2.25)$$

olup, (2.24)'te yerine yazılırsa \tilde{w}

$$\tilde{w} = L'\beta^0 + C'V^{-1}(y - X\beta^0) \quad (2.26)$$

olarak elde edilir [2] [16]. (2.26)'da görüldüğü gibi w 'nun önkestiricisi olan \tilde{w} iki kısımdan oluşmaktadır: $L'\beta^0$, $L'\beta$ 'nın en iyi doğrusal yansız tahmini ve $C'V^{-1}(y - X\beta^0)$, $E(u) = 0$ olduğunda ve $X\beta$ yerine en iyi doğrusal yansız tahmini $X\beta^0$ yazıldığında, u 'nun en iyi doğrusal önkestiricisidir. Buna göre en iyi doğrusal yansız önkestirici

$$\tilde{u}^0 = BLUP(u) = C'V^{-1}(y - X\beta^0) \quad (2.27)$$

şeklinde elde edilir. (2.15)'te $\mu_U = 0$ ve $\mu_Y = E(y) = X\beta$ olduğunda (2.27)'deki ifade elde edilir. (2.26)'daki \tilde{w} , β^0 kullanılarak elde edilen $L'\beta$ 'nın en iyi doğrusal yansız tahmininin (BLUE) ve u 'nun en iyi doğrusal yansız önkestiricisinin (BLUP) toplamıdır.

u ve y 'nin $f(u, y)$ bileşik dağılımının tüm parametreleri bilindiğinde BP kullanılabilir. BLP ve BLUP, $f(u, y)$ 'nin bazı parametreleri bilindiğinde en iyi yöntemlerdir. BLP için sadece birinci ve ikinci momentlerin bilindiği varsayılır. BLUP için birinci momentlerin bilinmediği ancak ikinci momentlerin bilindiği varsayılır [2].

2.1.4. Henderson'ın Karma Model Eşitlikleri

Henderson'ın karma model eşitliklerine göre en iyi doğrusal yansız önkestirici

$$\begin{bmatrix} u \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ X\beta \end{bmatrix}, \begin{bmatrix} D & C \\ C' & V \end{bmatrix}\right) \text{ varsayımı altında } (u, y) \text{ nin bileşik en çok olabilirlik}$$

fonksiyonunun maksimum yapılması ile elde edilebilir. Buna göre $f(u, y)$

$$f(u, y) = f(y|u)f(u)$$

$$= \frac{e^{\left(-\frac{1}{2}(y-X\beta-Zu)'R^{-1}(y-X\beta-Zu)+u'D^{-1}u\right)}}{(2\pi)^{\frac{1}{2}(N+q)}|R|^{\frac{1}{2}}|D|^{\frac{1}{2}}} \quad (2.28)$$

şeklinde elde edilir. (2.28) en çok olabilirlik fonksiyonudur ve logaritması alındıktan sonra β ve u 'ya göre türevi alınıp sifira eşitlenirse Henderson'ın karma model eşitlikleri aşağıdaki gibi elde edilir:

$$\begin{aligned} \log f(u, y) &= \log \{f(y|u)f(u)\} \\ &= -\frac{1}{2} \left\{ n \log(2\pi) + n \log \sigma^2 + \log |R| + (y - X\beta - Zu)' R^{-1} (y - X\beta - Zu) / \sigma^2 \right\} \\ &\quad - \frac{1}{2} \left\{ q \log(2\pi) + q \log \sigma^2 + \log |D| + u' D^{-1} u / \sigma^2 \right\} \\ &= -\frac{1}{2} \left\{ (n+q) \log(2\pi) + (n+q) \log \sigma^2 + \log |R| + \log |D| \right\} \\ &\quad \left\{ + (y - X\beta)' R^{-1} (y - X\beta) / \sigma^2 \right\} \\ &\quad - \frac{1}{2\sigma^2} \left\{ u' (ZR^{-1}Z' + D^{-1}) u - 2(y - X\beta)' R^{-1} Zu \right\} \end{aligned} \quad (2.29)$$

(2.29)'un β ve u 'ya göre türevi alınıp sifira eşitlenirse;

$$\frac{\partial \log f(u, y)}{\partial \beta} = -\frac{1}{2} (-2X'R^{-1}y + 2X'R^{-1}X\beta + 2X'R^{-1}Zu) = 0$$

$$\frac{\partial \log f(u, y)}{\partial u} = -\frac{1}{2} \left\{ (-2Z'R^{-1}y + 2Z'R^{-1}X\beta + (Z'R^{-1}Z + Z'R^{-1}Z)u + (D^{-1} + D^{-1})u \right\} = 0$$

bulunur. β ve u yerine tahmin edicileri $\hat{\beta}$ ve u yazılarak elde edilen eşitlikler yazılırsa

$$X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{u} = X'R^{-1}y \quad (2.30)$$

$$Z'R^{-1}X\hat{\beta} + (ZR^{-1}Z' + D^{-1})\hat{u} = Z'R^{-1}y \quad (2.31)$$

olur [2] [16]. (2.30) ve (2.31)'deki eşitlikler Henderson'ın doğrusal karma model eşitlikleridir, bu eşitlikler matris gösterimi ile

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & ZR^{-1}Z' + D^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (2.32)$$

şeklinde yazılabilir. (2.31)'den u

$$u = (Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\hat{\beta}) \quad (2.33)$$

dir. (2.33)'teki u değeri (2.30)'da yerine yazılırsa;

$$X'R^{-1}X\hat{\beta} + X'R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\hat{\beta}) = X'R^{-1}y \quad (2.34)$$

$$X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]X\hat{\beta} = X'[R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}]y \quad (2.35)$$

elde edilir. Burada $B = R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}$ için (2.35) yeniden düzenlenirse $X'BX\hat{\beta} = X'By$ olur. Buradan

$$\hat{\beta} = (X'BX)^{-1}X'By \quad (2.36)$$

olarak elde edilir. $V = ZDZ' + R$ ifadesinin tersi Schur tümleyenine göre $V^{-1} = (ZDZ' + R)^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}$ 'dir. Bu durumda $B = V^{-1}$ 'dir. (2.35)'te B yerine V^{-1} yazılırsa

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \quad (2.37)$$

olarak elde edilir. Burada elde edilen $\hat{\beta}$, β 'nin genelleştirilmiş en iyi tahmin edicisine (BLUE) eşittir. D ve R bilindiğinde yani V^{-1} bilindiğinde $\hat{\beta}$, β 'nin en iyi doğrusal yansız kestiricisidir.

Aynı şekilde \hat{u} , u 'nun en iyi doğrusal yansız kestiricisidir. Bunu göstermek için (2.33), bazı düzenlemeler yapılarak yeniden yazılırsa,

$$u = (Z'R^{-1}Z + D^{-1})^{-1}Z'R^{-1}(y - X\hat{\beta}) \quad (2.38)$$

şeklinde elde edilir. (2.38)'de $(Z'R^{-1}Z + D^{-1})^{-1}$ terimi Schur tümleyenlerine göre

$(Z'R^{-1}Z + D^{-1})^{-1} = D - DZ'(R + ZDZ')^{-1}ZD = D - DZ'V^{-1}ZD$ olarak elde edilir. Bu değer (2.38)'de yerine yazılırsa,

$$\begin{aligned} \hat{u} &= (D - DZ'V^{-1}ZD)Z'R^{-1}(y - X\hat{\beta}) \\ &= DZ'R^{-1}(y - X\hat{\beta}) - DZ'V^{-1}ZDZ'R^{-1}(y - X\hat{\beta}) \\ &= DZ'R^{-1}(y - X\hat{\beta}) - DZ'V^{-1}(V - R)R^{-1}(y - X\hat{\beta}) \\ &= DZ'R^{-1}(y - X\hat{\beta}) - DZ'V^{-1}VR^{-1}(y - X\hat{\beta}) + DZ'V^{-1}RR^{-1}(y - X\hat{\beta}) \end{aligned}$$

$$\begin{aligned}
&= DZ'V^{-1}(y - X\hat{\beta}) = CV^{-1}(y - X\hat{\beta}) \\
&= DZ'V^{-1}(y - X\hat{\beta}) \tag{2.39}
\end{aligned}$$

olarak elde edilir.

Henderson'ın karma model eşitliklerinde β ve u vektörü eş zamanlı kestirilir. Ayrıca doğrusal yansız önkestirici (BLUP) için V matrisinin tersini alınırken, Henderson'ın karma model eşitliklerinde tahminler elde edilirken D ve R köşegen matrislerinin tersi alınır. V matrisine göre köşegen matrislerin tersinin alınması daha kolaydır [14].

2.2. Varyans Bileşenlerinin Kestirimi

Doğrusal karma modelde parametrelerin önkestirimleri bulunurken varyans-kovaryans matrisin V 'nin bilindiği varsayılır. V bilinmediğinde, V 'yi oluşturan (2.6)'da verilen D ve R matrisinin kestirilmesi gerekir. Bu matrislerin kestirimi (2.7)'de verilen σ_i^2 'nin ve σ_e^2 'nin kestirilmesi anlamına gelmektedir. Bu iki varyansın kestirimine varyans bileşenlerinin kestirimi denir.

2.2.1. En Çok Olabilirlik Yöntemi ile Varyans Bileşenlerinin Kestirimi

(2.2)'de verilen doğrusal karma modelde Zu , $Z_u = [Z_1 \dots Z_r]$ $\begin{bmatrix} u_1 \\ \vdots \\ u_r \end{bmatrix} = \sum_{i=1}^r Z_i u_i$ şeklinde

yazılabilir. Rasgele modelde, u rasgele etkilere ilişkin beklenen değer, varyans ve kovaryans değeri $E(u_i) = 0$, $Var(u_i) = \sigma_i^2 I_{q_i} \forall i$ ve $i \neq h$ için $Kov(u_i, u_h) = 0$ 'dır. Artık terimleri vektörüne ilişkin beklenen değer ve varyans $E(e) = 0$, $V(e) = \sigma_e^2 I_n$ olup rasgele etkiler ve artık terimleri arasındaki kovaryans $Kov(u_i, e) = 0 \forall i$ 'dir. Bu varsayımlara göre y yanıt vektörüne ilişkin ortalama vektör $E(y) = X\beta$ ve varyans kovaryans matrisi $V = Var(y) = \sum_{i=1}^r Z_i Z_i' \sigma_i^2 + \sigma_e^2 I_n$ şeklinde yazılabilir. Buna göre $y \sim N(X\beta, V)$ varsayımı altında

$$L = L(\beta, V | y) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} e^{-\frac{1}{2}(y-X\beta)'V^{-1}(y-X\beta)} \quad (2.40)$$

olabilirlik fonksiyonu olarak adlandırılır. (2.40)'taki olabilirlik fonksiyonunu maksimum yapan β ve V değerlerinin tahminleri en çok olabilirlik tahminleridir. Daha doğrusu, olabilirlik fonksiyonunun β ve $\sigma^2 = [\sigma_e^2 \quad \sigma_1^2 \quad \dots \quad \sigma_r^2]$ vektörlerine göre maksimize edilmesidir. (2.40)'taki olabilirlik fonksiyonunu maksimum yapan değerleri bulmak için fonksiyonun logaritması alınır ve aşağıdaki gibi elde edilir:

$$I = \text{Log}(L) = -\frac{1}{2}n \log(2\pi) - \frac{1}{2} \log|V| - \frac{1}{2}(y-X\beta)'V^{-1}(y-X\beta) \quad (2.41)$$

(2.41)'daki fonksiyonu maksimize etmek için sırasıyla β ve σ_i^2 'ye göre türev alınırsa aşağıdaki eşitlikler elde edilir:

$$I_{\beta} = \frac{\partial I}{\partial \beta} = X'V^{-1}y - X'V^{-1}X\beta \quad (2.42)$$

$$I_{\sigma_i^2} = \frac{\partial I}{\partial \sigma_i^2} = -\frac{1}{2} \text{tr}(V^{-1}Z_i Z_i') + (y-X\beta)'V^{-1}Z_i Z_i'V^{-1}(y-X\beta) \quad (2.43)$$

Burada $\frac{\partial V}{\partial \sigma_i^2} = Z_i Z_i'$, $\frac{\partial V}{\partial \sigma_i^2} \log|V| = I_z(V^{-1} \frac{\partial V}{\partial \sigma_i^2})$ ve

$$\frac{\partial V^{-1}}{\partial \sigma_i^2} = -V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} = -V^{-1} Z_i Z_i' V^{-1} \text{ dir.}$$

(2.42) ve (2.43)'ten en çok olabilirlik eşitlikleri aşağıdaki gibi elde edilir:

$$X'V^{-1}X\beta^0 = X'V^{-1}y \quad (2.44)$$

$$\text{tr}(\tilde{V}^{-1}Z_i Z_i') = (y-X\beta^0)'\tilde{V}^{-1}Z_i Z_i'\tilde{V}^{-1}(y-X\beta^0) \quad i = 0, 1, \dots, r \quad (2.45)$$

(2.45)'teki ifadeyi daha basit ifade edebilmek için

$$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \quad (2.46)$$

şeklinde tanımlandığında

$$\tilde{V}^{-1}(y-X\beta^0) = Py \quad (2.47)$$

olur. Bu tanımlamaya göre (2.45),

$$\left\{ {}_s \text{tr}(\tilde{V}^{-1} Z_i Z_i') \right\}_{i=0}^r = \left\{ {}_s y' P Z_i Z_i' P y \right\}_{i=0}^r \quad (2.48)$$

şeklinde ifade edilir. $\{ {}_s a_i \}_{i=0}^r$ gösterimi $rx1$ boyutlu sütun vektörünü ifade etmektedir.

Bu eşitlikler yardımı ile β ve σ_i^2 'lerin en çok olabilirlik tahmin edicileri bulunur.

Elde edilen tahminlerin en çok olabilirlik tahminleri olması için ikinci türevlerinin negatif olup olmadığının ve olabilirlik fonksiyonunun parametre uzayının sınırları içinde olup olmadığının kontrol edilmesi gerekmektedir. Çünkü maksimizasyon parametre uzayı ile sınırlandırılmalıdır. Ancak birçok durumda bu kısıtlayıcı bir gereklilik değildir. Örneğin Eşitlik (2.44)'te β 'nın tahmini olarak β^0 gösterilirse, V matrisi bilindiğinde $X\beta$ 'nin en çok olabilirlik tahmini $X\beta^0$ 'dır. β vektörünün herhangi bir elemanı β_k için parametre uzayı genellikle $-\infty < \beta_k < \infty$ 'dir. Eşitlik (2.40)'da β^0 'ın elemanlarının negatif, pozitif ya da sıfır olması ile ilgilenilmez. Ancak bu durum σ^2 vektörünün elemanları için geçerli değildir. Çünkü karma modelde varyans bileşenleri için parametre uzayı $\sigma_e^2 > 0$ ve $\sigma_i^2 \geq 0$ $i=1,2,\dots,r$ biçimindedir. Dolayısıyla, $\tilde{\sigma}_e^2$ ve $\tilde{\sigma}_i^2$ 'nin en çok olabilirlik tahmin edicileri olması için $\tilde{\sigma}_e^2 > 0$ ve $\tilde{\sigma}_i^2 \geq 0$ koşulunu sağlaması gerekmektedir. σ^2 vektörü için (2.44) ve (2.45)'ten elde edilen çözümler $\hat{\sigma}^2$ ile gösterilsin. Buna göre $i=1,2,\dots,r$ için $\hat{\sigma}_e^2 > 0$ ve $\hat{\sigma}_i^2 \geq 0$ koşulu sağlanırsa $\hat{\sigma}^2 = \tilde{\sigma}^2$ 'dir. Ancak genellikle bir ya da daha fazla $\hat{\sigma}^2$ negatif olabilir. Bu durumda negatif değerler yerine sıfır değeri alınır, bu da negatif değere karşılık gelen rasgele faktörün modelden çıkarılması anlamına gelmektedir. Yeni $\hat{\sigma}^2$ vektörü bu azaltılmış model üzerinden yeniden elde edilir ve $\hat{\sigma}_e^2 > 0$ ve $\hat{\sigma}_i^2 \geq 0$ koşulunun sağlanıp sağlanmadığı tekrar kontrol edilir. Eğer koşul sağlanmıyorsa işlemler tekrarlanır.

(2.45) ya da (2.48) eşitlikleri, varyans bileşenlerinin karmaşık polinom fonksiyonlarıdır ve en çok olabilirlik çözümlerinin kapalı formları elde edilememektedir. Bu nedenle her bir veri kümesi için çözümler dolayısıyla tahminler iteratif yöntemler ile elde edilir [2][15].

2.2.1.1. Hartley –Rao Formu

Hartley ve Rao (1967) en çok olabilirlik fonksiyonunu, varyans-kovaryans matrisi $V = \sigma_e^2 H$ 'e göre tanımlamışlardır. V matrisinde σ_e^2 yerine H matrisinde 1 değeri ve $i = 1, 2, \dots, r$ için σ_i^2 yerine varyans oranları olarak bilinen $\gamma_i = \sigma_i^2 / \sigma_e^2$ gelmiştir. Buna göre en çok olabilirlik yönteminde kestirilecek parametreler β , σ_e^2 ve γ_i 'dir [2][16].

Hartley ve Rao'nun önerdiği en çok olabilirlik yönteminin kullanılabilmesi için tasarım matrislerinin bazı varsayımları sağlaması gerekmektedir. X ve $i = 1, 2, \dots, r$ için Z_i tasarım matrislerinin tam ranklı olduğu ($rank(X) = p$ ve $rank(Z_i) = q_i$) varsayılr. Genellikle varyans analizinde her bir y_i rasgele etkenin bir düzeyi ile ilişkilidir. Bu nedenle Z_i tasarım matrisinin her satırının bir elemanı 1, diğer $q_i - 1$ elemanı sıfır değerini alacaktır. Hartley ve Rao, Z_i 'nin bu özelliği sağladığını varsayarlar. Z_i 'nin bu özelliği sağlaması Z_i matrisinin $q_i \times q_i$ boyutlu köşegen bir matris olduğunu gösterir [17].

y yanıt vektörünün normal dağıldığı varsayımı altında $V = \sigma_e^2 H$ 'e göre olabilirlik fonksiyonu ve olabilirlik fonksiyonunun logaritması sırasıyla aşağıdaki gibi yazılır:

$$L = (2\pi)^{-N/2} (\sigma_e^2)^{-N/2} |H|^{-1/2} \exp\left[\frac{-(y - X\beta)' H^{-1} (y - X\beta)}{2\sigma_e^2}\right] \quad (2.49)$$

$$l = \log L = -\frac{1}{2} n \log(2\pi) - \frac{1}{2} n \log \sigma_e^2 - \frac{1}{2} \log |H| - \frac{1}{2\sigma_e^2} (y - X\beta)' H^{-1} (y - X\beta) \quad (2.50)$$

Eşitlik (2.50)'nin β , σ_e^2 ve γ_i 'ye göre türevi alınırsa, sırasıyla

$$L_{\beta} = \frac{\partial L}{\partial \beta} = \frac{1}{\sigma_e^2} (X' H^{-1} y - X' H^{-1} X \beta) \quad (2.51)$$

$$L_{\sigma_e^2} = \frac{\partial L}{\partial \sigma_e^2} = -\frac{n}{2\sigma_e^2} + \frac{1}{2\sigma_e^4} (y - X\beta)' H^{-1} (y - X\beta) \quad (2.52)$$

$$\frac{\partial L}{\partial \gamma_i} = -\frac{1}{2} \text{tr}(Z_i' H^{-1} Z_i) + \frac{1}{2\sigma_e^2} (y - X\beta)' H^{-1} Z_i Z_i' H^{-1} (y - X\beta) \quad (2.53)$$

eşitlikleri elde edilir. Burada $\frac{\partial H}{\partial \gamma_i} = Z_i Z_i'$ $i = 1, \dots, r$, $\frac{\partial}{\partial \gamma_i} \log |H| = \dot{z}(H^{-1} \frac{\partial H}{\partial \gamma_i})$ ve

$\frac{\partial H^{-1}}{\partial \gamma_i} = -H^{-1} \frac{\partial H}{\partial \gamma_i} H^{-1}$, dir. Buna göre

$$X' \tilde{H}^{-1} X \beta^0 = X' \tilde{H}^{-1} y \quad (2.54)$$

$$\tilde{\sigma}_e^2 = (y - X \beta^0)' \tilde{H}^{-1} (y - X \tilde{\beta}) / n \quad (2.55)$$

$$\left\{ {}_c \text{tr}(\tilde{H}^{-1} Z_i Z_i') \right\}_{i=1}^r = \left\{ {}_c (y - X \beta^0)' \tilde{H}^{-1} Z_i Z_i' \tilde{H}^{-1} (y - X \beta^0) / \tilde{\sigma}_e^2 \right\}_{i=1}^r \quad (2.56)$$

Eşitlik (2.56)'da tahmin edilen σ_i^2 değil, $i = 1, 2, \dots, r$ için varyans oranları $\gamma_i = \sigma_i^2 / \sigma_e^2$ 'dir. Eşitlik (2.56)'nın iteratif çözümü Eşitlik (2.41)'e göre daha kolaydır.

Ancak bu tahmin edicilerin en çok olabilirlik tahmin edicisi olması için ikinci türev matrisi olan Hessian matrisinin negatif tanımlı bir matris olması ve parametre uzayının sınırları içinde olması gerekir.

2.2.2. Kısıtlanmış En Çok Olabilirlik Yöntemi ile Varyans Bileşenlerinin Tahmini

En çok olabilirlik tahmininin bir özelliği, varyans bileşenlerinin tahmininde sabit etkiler tahminindeki serbestlik derecesini dikkate almamasıdır. İlk olarak Anderson ve Bancroft [18] ve Russell ve Bradley [19] tarafından dengeli veri seti için geliştirilmiştir. Daha sonra, Thompson [20] tarafından genel olarak dengeli veriler için genişletilmiştir. Patterson ve Thompson [21] tarafından da karma modeller için genişletilmiştir. Kısıtlanmış en çok olabilirlik yöntemi ile en çok olabilirlik yönteminde varyans bileşenleri tahmininde sabit etkiler tahmin edilirken kullanılan serbestlik derecelerinin dikkate alınmaması kısıtının üstesinden gelinmiş olunur. Kısıtlanmış en çok olabilirlik (REML) kestirimi ile ilgili temel fikir, modelin sadece sabit etkiler kısmı en küçük kareler ile hesaplandıktan sonra artıklara dayalı olarak varyans bileşenlerini tahmin etmektir.

REML yönteminde, doğrudan y vektörünü kullanmak yerine, değerleri ne olursa olsun herhangi bir sabit etki içermeyecek şekilde seçilen y 'nin doğrusal kombinasyonu kullanılır. Bu doğrusal kombinasyonlar, sabit etkiler kestirildikten sonra elde edilen

artıklara eşdeğerdir. $k'y = k'X\beta + k'Zu$ doğrusal birleşiminin β terimini içermemesi için, yani $k'X\beta = 0$ olduğu k' vektörleri seçilir. Bu durumda $\forall \beta$ için $k'X\beta = 0$ ve $k'X = 0$ olur.

k' , herhangi bir c' için $k' = c'(I - XX^{-1})$ veya

$$k' = c' [I - X(X'X)^{-1}X'] = c'(I - XX^+) = c'M \text{ olmalıdır.}$$

Burada $M = I - X(X'X)^{-1}X' = I - XX^+$ olarak tanımlanır. $k'y$ 'nin beklenen değeri $E(k'y) = k'X\beta = 0$ 'dir. Doğrusal olarak bağımsız birleşimlerin sayısı X matrisine bağlıdır. $n \times p$ boyutlu X matrisi ve r tane $k'X\beta = 0$ eşitliği sadece k' vektörünün $n-r$ doğrusal olarak bağımsız değerleri ile mümkündür. Bu nedenle, K' matrisinin satırları olan k' doğrusal bağımsız vektörler kümesi kullanıldığında doğrusal birleşimler $K'y$ ile ifade edilir. Burada $K' = TM$ olup K' ve T $n-r$ boyutlu tam ranklı matrislerdir.

$y \sim N(X\beta, V)$ varsayımı ve $K'X = 0$ kısıtı altında $K'y \sim N(0, K'VK)$ 'dir ve $K'y$ 'nin kısıtlanmış en çok olabilirlik fonksiyonunun logaritması aşağıdaki gibi ifade edilir:

$$I_R = \log L_R = -\frac{1}{2}(N-r) \log 2\pi - \frac{1}{2} \log |K'VK| - \frac{1}{2} y'K(K'VK)^{-1}K'y \quad (2.57)$$

Eşitlik (2.57) daha basit ifade edebilmek için

$$P = \tilde{V}^{-1} - \tilde{V}^{-1}X(X'\tilde{V}^{-1}X)^{-1}X'\tilde{V}^{-1} = K(K'\tilde{V}K)^{-1}K' \quad (2.58)$$

şeklinde tanımlanırsa

$$I_R = \log L_R = -\frac{1}{2}(N-r) \log 2\pi - \frac{1}{2} \log |K'VK| - \frac{1}{2} y'Py \quad (2.59)$$

olur. Eşitlik (259)'un varyans bileşenlerine göre türevi alınıp sifıra eşitlenirse REML tahminleri elde edilir. Eşitlik (259)'un σ_i^2 'ye göre türevini almak için öncelikle P eşitliğinin σ_i^2 'ye göre türevi alınır ve

$$\frac{\partial P}{\partial \sigma_i^2} = \frac{\partial}{\partial \sigma_i^2} K(K'VK)^{-1}K'$$

$$\begin{aligned}
&= -K(K'VK)^{-1} \frac{\partial K'VK}{\partial \sigma_i^2} (K'VK)^{-1} K' \\
&= -K(K'VK)^{-1} K' \frac{\partial V}{\partial \sigma_i^2} K(K'VK)^{-1} K' \\
&= -P \frac{\partial V}{\partial \sigma_i^2} P = -PZ_i Z_i' P
\end{aligned}$$

şeklindedir. Bulunan değer Eşitlik (2.59)'da yerine konulursa,

$$\begin{aligned}
\frac{\partial I_R}{\partial \sigma_i^2} &= -\frac{1}{2} \text{tr} \left[(K'VK)^{-1} K' Z_i Z_i' K \right] - \frac{1}{2} y'(-1) PZ_i Z_i' P y \\
&= -\frac{1}{2} \text{tr}(PZ_i Z_i') + \frac{1}{2} y' PZ_i Z_i' P y = 0
\end{aligned}$$

olarak elde edilir. Buna göre REML tahminleri

$$\left\{ {}_c \text{tr}(\tilde{P}Z_i Z_i') \right\}_{i=0}^r = \left\{ {}_c y' \tilde{P}Z_i Z_i' \tilde{P}y \right\}_{i=0}^r \quad (2.60)$$

şeklinde bulunur.

REML denklemlerinin çözümleri, karma modellere uyan dengeli veriler için, ANOVA tahmin edicileriyle aynıdır ve bu sonuç, normalliğin kabul edilip edilmediğine bakılmaksızın doğrudur. Başka bir deyişle normallik göz ardı edilir, ancak yine de normallik altındaki REML denklemleri olan denklemler çözümlerse, çözümler tüm karma modellere uyan dengeli veriler için ANOVA tahminleri ile özdeştir. Bu çözümler, normallik varsayımı olmadıkça ve varyans bileşenlerini tahmin etmek için en çok olabilirlik yönteminin negatif olmama şartı dikkate alınmadıkça REML tahmin edicileri değillerdir. REML çözümlerinin normalliğe dayanmadan ANOVA tahmin edicileri olduğuna dair ayrıntılı kanıt Anderson'ın [22] yayınladığı çalışmasında verilmiştir.

2.2.3. En Çok Olabilirlik Yöntemi (ML) ile Kısıtlanmış En Çok Olabilirlik Yöntemi (REML) Arasındaki Farklar

Varyans bileşenlerinin tahmininde dengesiz veriler için ML ve REML'nin herhangi biri ANOVA yöntemine tercih edilir. Çünkü ML ve REML en çok olabilirlik yöntemine dayanmaktadır. En çok olabilirlik yöntemi ile tutarlı, asimptotik normal ve asimptotik

örneklem yayılım matrisinin bilindiği tahmin ediciler elde edilmektedir. Tahmin edicilerin bu özellikleri güven aralıklarının oluşturulmasını ve parametrelerle ilgili hipotezlerin test edilmesini sağlar. Buna karşın ANOVA tahmin edicileri sadece yansızdır ve örneklem yayılım matrislerinin türetilmesi genellikle zordur. ML ve REML tahmin edicileri verilerin normalliğini varsayımına dayanmaktadır. ML ve REML, bilgisayar gereksinimleri nedeniyle başlangıçta pratik yöntemler değildi. Ancak günümüzde büyük ve daha hızlı bilgisayarların gelişimi ile varyans bileşenlerinin tahmininde yaygın olarak kullanılan yöntemlerdir.

ML veya REML yöntemlerinden hangisinin tercih edileceği konusuna gelince öncelikli olarak her iki yöntem de en çok olabilirlik ilkesine dayanmaktadır ve varyans bileşenlerine göre doğrusal değildir. Bu nedenle tahminleri elde etmek için iteratif yöntemler kullanılır. ML ile sabit etkilerin tahmin edicileri elde edilirken, REML ile bu tahmin ediciler elde edilemez. REML yöntemi ile dengeli verilerde yansız ve en küçük varyanslı tahmin ediciler elde edilirken, ML ile elde edilen tahmin ediciler yanlıdır. REML tahmin edicileri ML tahmin edicilerine göre verideki aykırı değerlere karşı hassas değildir [2] [14].

2.3. İteratif Yöntemler

En çok olabilirlik eşitlikleri parametrelere göre doğrusal olmadığından, eşitliklerin çözümü zordur. Bu nedenle varyans bileşenleri kestirimlerini elde etmek için iteratif yöntemler kullanılır. Bu iteratif yöntemlerden yaygın olarak kullanılanları Newton-Raphson algoritması, Fisher'in skorlama yöntemi algoritması ve EM algoritmasıdır [2] [23].

2.3.1. Newton-Raphson Algoritması

Doğrusal olmayan fonksiyonları maksimize etmek için yaygın olarak kullanılan bir yöntem Newton-Raphson yöntemidir. Burada amaç Taylor teoremi yardımıyla θ 'nın bir fonksiyonu olan $f(\theta)$ 'yı maksimize eden $\frac{\partial f(\theta)}{\partial \theta} = 0$ türev ifadesinin bir kökünü bulmaktır.

$\frac{\partial f(\theta)}{\partial \theta} = 0$ türevi için θ parametre vektörünün başlangıç değeri olarak alınan θ_0 için

Taylor açılımı,

$$\frac{\partial f(\theta)}{\partial \theta} = f'(\theta) \cong f'(\theta_0) + \frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} (\theta - \theta_0) \quad (2.61)$$

şeklinde yazılır. Eşitlik (2.61) sıfıra eşitlenirse

$$f'(\theta_0) + \frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} (\theta - \theta_0) = 0 \quad (2.62)$$

$$\theta = \theta_0 - \left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} \right]^{-1} f'(\theta_0) \quad (2.63)$$

elde edilir. Newton—Raphson algoritmasının genel iterasyon denklemi herhangi bir $(m+1)$. iterasyon adımı için θ parametre vektörünün tahmin değeri

$$\theta^{(m+1)} = \theta^{(m)} - \left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} \right]^{-1} f'(\theta^{(m)}) \quad (2.64)$$

olarak elde edilir. Burada $\left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} \right]$ ifadesi ikinci türevler matrisi olarak bilinen Hessian matrisidir.

Newton-Raphson yönteminin avantajı, hızlı yakınsamasıdır. Özellikle, θ minimuma yakın olduğunda çok hızlı yakınsar. Yöntemin dezavantajları ise, parametrelerin sayısı çok fazla olduğunda Hessian matrisinin tersinin bulunmasının zorlaşması ve θ minimuma yakın değilse, yakınsamanın olmaması durumunda Hessian matrisinin negatif tanımlı olmasıdır [24].

2.3.2. Skorlama Algoritması

İkinci türev matrisin ağır hesaplama yükünden kurtulmak için kullanılan bir başka yöntem, Eşitlik (2.64)'te $\left[\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'} \right]$ tersinin bilgi matrisinin tersi ile değiştirildiği skorlama yöntemidir. Başka bir deyişle Hessian matrisinin beklenen değeri ile değiştirilir. Jennrich ve Sampson [12], skorlama yönteminin Newton-Raphson yöntemine göre zayıf

başlangıç değerlerine göre daha sağlam olduğunu göstermişlerdir. İlk birkaç adımda skorlama kullanarak başlayıp, daha sonra Newton-Raphson'a geçerek tekrarlayan bir algoritma önermişlerdir. Skorlama yöntemine göre iterasyon adımları aşağıdaki gibi tanımlanır:

$$\theta^{(m+1)} = \theta^{(m)} + [I(\theta^{(m)})]^{-1} \left. \frac{\partial l}{\partial \theta} \right|_{\theta^{(m)}} \quad (2.65)$$

Burada $I(\theta^{(m)})$, $\theta = \theta^{(m)}$ için hesaplanan bilgi matrisidir. Bilgi matrisi, ikinci türevler matrisinin negatif beklenen değeridir.

2.3.3. EM Algoritması

EM algoritması, Newton-Raphson ve Skorlama yöntemi gibi yöntemler ile elde edilen ML ve REML tahminlerinden farklı ML ve REML tahminlerini hesaplamak için kullanılan iteratif bir algoritmadır. EM, beklenti-maksimizasyonu anlamına gelir. Çünkü koşullu beklenen değerlerinin hesaplanması ve basitleştirilmiş olabilirlik fonksiyonunun maksimize edilmesi arasında işlemler tekrarlanır. EM algoritması ile yalnızca tahminler elde edilir ve Newton-Raphson ve Skorlama yöntemlerinde olduğu gibi varyans tahminleri elde edilmez. Varyans tahminlerini elde etmek için ekstra hesaplamalar yapılmalıdır [2].

Beklenti-Maksimizasyon (EM) algoritması, eksik veri olduğunda model parametrelerini en çok olabilirlik tahmini ile elde etmenin iteratif bir yoludur.

2.3.1.1. En Çok Olabilirlik Tahmini için EM Algoritması

EM algoritması kullanılarak en çok olabilirlik tahmini dört adımda elde edilir. EM algoritması adımları aşağıdaki gibidir:

Adım 0: $m = 0$ için $\sigma^{2(0)}$ ve $\beta^{(0)}$ başlangıç değerlerine karar verilir.

Adım 1(Adım-E): Yeterli istatistiklerin koşullu beklenen değeri hesaplanır:

$$\begin{aligned}
\hat{t}_i^{(m)} &= E(u_i' u_i | y) \Big|_{\beta=\beta^{(m)}, \sigma^2=\sigma^{2(m)}} \\
&= \sigma_i^{4(m)} (y - X\beta^{(m)})' (V^{(m)})^{-1} Z_i Z_i' (V^{(m)})^{-1} (y - X\beta^{(m)}) \\
&\quad + iz \left[\sigma_i^{2(m)} I_{q_i} - \sigma_i^{4(m)} Z_i' (V^{(m)})^{-1} Z_i \right]
\end{aligned}$$

ve

$$\begin{aligned}
\hat{s}^{(m)} &= E \left(y - \sum_{i=1}^r Z_i u_i \mid y \right) \Big|_{\beta=\beta^{(m)}, \sigma^2=\sigma^{2(m)}} \\
&= y - \sum_{i=1}^r Z_i Z_i' \sigma_i^{2(m)} (V^{(m)})^{-1} (y - X\beta^{(m)}) \\
&= y - (V^{(m)} - \sigma_0^{2(m)} I) (V^{(m)})^{-1} (y - X\beta^{(m)}) \\
&= X\beta^{(m)} + \sigma_0^{2(m)} (V^{(m)})^{-1} (y - X\beta^{(m)})
\end{aligned} \tag{2.66}$$

Adım 2(Adım M): En çok olabilirlik tahminleri elde edilir:

$$\sigma_i^{2(m+1)} = \hat{t}_i^{(m)} / q_i, \quad i = 0, 1, 2, \dots, r \text{ için} \tag{2.67}$$

$$X\beta^{(m+1)} = X(X'X)^{-1} X' \hat{s}^{(m)} \tag{2.68}$$

Adım 3: Yakınsama sağlandığında, $\tilde{\sigma}^2 = \sigma^{2(m+1)}$ ve $\tilde{\beta} = \beta^{(m+1)}$, iterasyon durur; aksi takdirde m bir birim artırılarak 1. adıma tekrar dönülür [2].

2.3.1.2. Kısıtlanmış En Çok Olabilirlik Tahmini için EM Algoritması

REML'de ise sabit değerler tahmin edilmediği için β değeri hesaplanmamakla birlikte genel adımlar aynıdır. REML tahmininde ise gerekli düzenlemeler yapılarak aynı adımlar tekrar edilir.

Adım 0: $m = 0$ olarak $\sigma^{2(0)}$ ın başlangıç değerine karar verilir.

Adım 1(Adım-E): Yeterli istatistiklerin koşullu beklenen değeri hesaplanır:

$$\begin{aligned}
\hat{t}_i^{(m)} &= \sigma_i^{4(m)} (K'y)' (K'V^{(m)}K)^{-1} K'Z_i Z_i' K (K'V^{(m)}K)^{-1} K'y \\
&\quad + iz \left[\sigma_i^{2(m)} I_{q_i} - \sigma_i^{4(m)} Z_i' K (K'V^{(m)}K)^{-1} K'Z_i \right] \\
&= \sigma_i^{4(m)} y' P^{(m)} Z_i Z_i' P^{(m)} y + tr \left[\sigma_i^{2(m)} I_{q_i} - \sigma_i^{4(m)} Z_i' P^{(m)} Z_i \right]
\end{aligned} \tag{2.69}$$

Adım 2(Adım M): Kısıtlanmış en çok olabilirlik tahminleri elde edilir:

$$\sigma_i^{2(m+1)} = \hat{t}_i^{(m)} / q_i, \quad i = 0, 1, 2, \dots, r,$$

Adım 3: Yakınsama sağlandığında, $\tilde{\sigma}^2 = \sigma^{2(m+1)}$, iterasyon durur; aksi takdirde m bir birim artırılarak 1. adıma tekrar dönülür [2].

3. DOĞRUSAL KARMA MODELDE SAĞLAM KESTİRİCİLER

Parametrik istatistiksel yöntemler varsayımlara dayanmaktadır. Parametrik istatistiksel yöntemler için en yaygın bilinen varsayım gözlemlenen verilerin normal dağılıma sahip olduğu varsayımdır. Normal dağılım varsayımının temel nedeni dağılımın teorik olarak uygun olmasıdır. Normal dağılım varsayımı sayesinde t istatistiği gibi çıkarsamalı istatistiklerin örneklem dağılımının bulunabilmesinin yanı sıra en çok olabilirlik ve olabilirlik oranı testleri gibi istatistiksel yöntemler için açık formüller elde edilebilir.

Normal dağılım varsayımına dayanan modellerde bazı veriler için gözlemlerin çoğu normal dağılıma uyum gösterdiği halde gözlemlerin çoğunluğundan uzakta olan bazı gözlemler normal dağılıma uyum olmamasına neden olmaktadır. Bu tür gözlemlere aykırı gözlemler/değerler denir ve tek bir aykırı değer bile normallik ya da doğrusallık varsayımının bozulmasına neden olabilir. Bazı durumlarda, aykırı değerlerin varlığı verilerin dağılımının yaklaşık normal dağılmasına neden olmaktadır. Bu durumda, dağılım normal dağılıma göre daha uzun ya da daha geniş kuyruklara sahiptir. Verilerin yaklaşık olarak normal dağılması durumunda normal dağılım teorisine dayanan sonuçların da yaklaşık olarak geçerli olması beklenir. Ancak verilerin dağılımı normal dağılıma göre uzun kuyruklu ise en çok olabilirlik yöntemine dayalı tahminler en iyi tahmin özelliğini sağlamaz, ayrıca kuyruklar simetrik ise tahminler olduğundan büyük varyansa sahip olabilir veya kuyruklar asimetrik ise çok büyük bir yanlılığa neden olabilirler. Bu durumda hipotez testi sonuçlarına güvenilmez ve güven aralıkları için güven düzeyleri güvenilir olmayabilir ve beklenen güven aralığı uzunlukları oldukça büyük olabilir. Eğer veriler herhangi bir aykırı değer içermiyorsa ya da aykırı değer oranı küçükse, sağlam yöntem ile klasik yöntem yaklaşık olarak aynı sonuçları verir [25]. İstatistiksel modellemeye sağlam yaklaşım, sadece veriler normal dağılıma uymadığı durumlarda değil, aynı zamanda yaklaşık normal dağıldığında da güvenilir parametre tahminleri, hipotez testleri ve güven aralıkları elde etmek için kullanılır. Sağlam yöntemler ile ilgili literatürde çok sayıda çalışma yayınlanmıştır: Huber [26], Hampel, Ronchetti, Rousseeuw ve Stahel [27], Rousseeuw ve Leroy [28] ve Staudte ve Sheather [29].

3.1. Aykırı Değer

Yapılan çalışmalarda toplanan veriler genellikle aykırı değer (outlier) olarak adlandırılan, verinin saçılım grafiğinde verilerin büyük çoğunluğunun dışında kalan veya verinin genel modelinden sapan gözlemler içerir. Örneklem ortalaması, örneklem varyansı, örneklem kovaryansları ve korelasyonları veya regresyon modelinin en küçük kareler tahminleri, tek bir aykırı değerden etkilenebilir [25]. Veride aykırı değerler olduğunda klasik yöntemler ile elde edilen kestirime güvenilmez.

Aykırı değerlere gerçek verilerde çok sık rastlanabilir ve çoğu zaman fark edilmezler çünkü günümüzde çok fazla veri bilgisayarlara girilirken yanlış girilebilmektedir. Aykırı değerler hem yanıt değişkeninde hem de açıklayıcı değişkenlerde olabilir. Yanıt değişkenindeki aykırı değerler y-yönünde aykırı ya da dikey aykırı değer olarak adlandırılır. Açıklayıcı değişkenlerdeki aykırı değerler ise kaldıraç noktaları (leverage points) olarak adlandırılır. Her iki aykırı değer de, en küçük kareler (LS) analizini tamamen bozabilir.

Aykırı değerleri ortaya çıkarmak için literatürde önerilen aykırı değer ölçütlerinden ve grafiksel gösterimlerden yararlanır. Ancak ikiden fazla açıklayıcı değişken olması durumunda, bu noktaların basit grafiksel yöntemlerle ortaya çıkarılması zordur. Bu sorunu çözmek için aykırı değerlerden etkilenmeyen sağlam (robust) yöntemler önerilmiştir.

3.1.1. Sabit Etkiler Modeli için Aykırı Değer Ölçütleri

Regresyon analizinin amacı, gözlemlenen değişkenleri en iyi temsil eden modeli tahmin etmektir. Klasik doğrusal regresyon modeli

$$y_i = x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \varepsilon_i \quad i = 1, 2, \dots, n \quad (3.1)$$

şeklinindedir. Burada x_{i1}, \dots, x_{ip} açıklayıcı değişkenler, y_i yanıt değişkenidir. ε_i hata teriminin ortalaması sıfır ve standart sapması bilinmeyen σ ile normal dağıldığı varsayılmaktadır. Eşitlik (3.1)'deki modelin tahmini

$$\hat{y}_i = x_{i1}\hat{\beta}_1 + \dots + x_{ip}\hat{\beta}_p \quad (3.2)$$

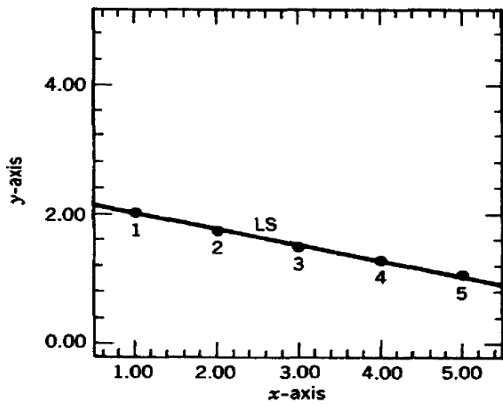
şeklindedir. Burada \hat{y}_i , y_i 'nin tahmin edilen veya kestirilen değeridir. Gözlenen değer ile modelden tahmin edilen değer arasındaki fark artık terimidir ve

$$e_i = y_i - \hat{y}_i \quad (3.3)$$

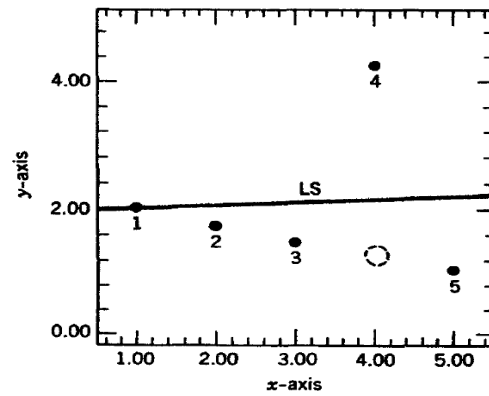
şeklinde elde edilir. LS tahmin yöntemi ile Eşitlik (3.3)'deki artık kareler toplamını en küçük yapan parametre değerleri elde edilir:

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial \beta_i} \Big|_{\beta_i = \hat{\beta}_i} = 0 \quad (3.4)$$

Eşitlik (3.4)'den elde edilen en küçük kareler tahminlerini önemli ölçüde etkileyen aykırı değerler olabilir. Bu değerlerin LS tahminleri üzerindeki etkileri Şekil 3.1-3.2 ile gösterilmektedir. Şekil 3.1 (a), $(x_1, y_1), \dots, (x_5, y_5)$ beş gözlemlili bir verinin en küçük kareler kestirimini göstermektedir. Bu veride dördüncü gözlemin yanlış değer aldığı varsayılınsın. Bu durumda Şekil 3.1 (b)'de görüldüğü gibi dördüncü gözleme ilişkin tahmin değeri en küçük kareler kestirimine ilişkin doğrunun üzerinde yer almamakta ve en küçük kareler kestirim doğrusunu kendisine doğru çekmektedir. Bu gözleme y-yönünde ya da dikey aykırı değer denir. Dikey aykırı değerler büyük pozitif ya da büyük negatif artıklara sahiptir. Şekil 3.1 (b)'den de görüldüğü gibi dördüncü gözlem LS doğrusunun en uzağındadır ve büyük artık değerine sahiptir.



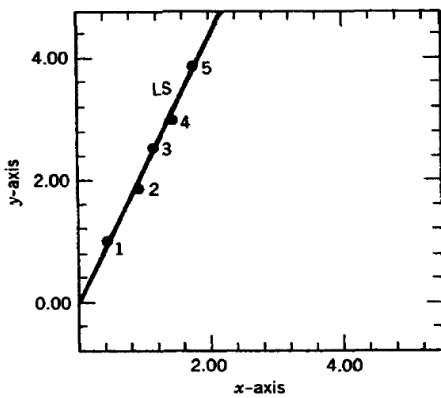
(a)



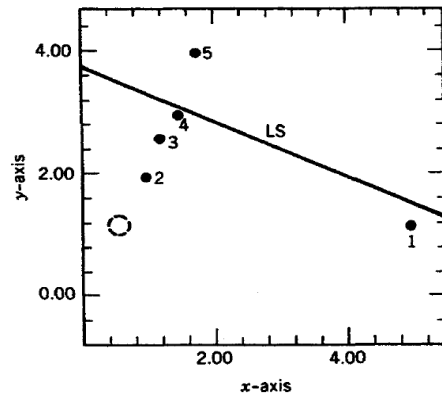
(b)

Şekil 3.1. (a) Beş gözleme ilişkin en küçük kareler kestirimi. (b) y-yönünde aykırı değer olduğunda en küçük kareler kestirimi [30]

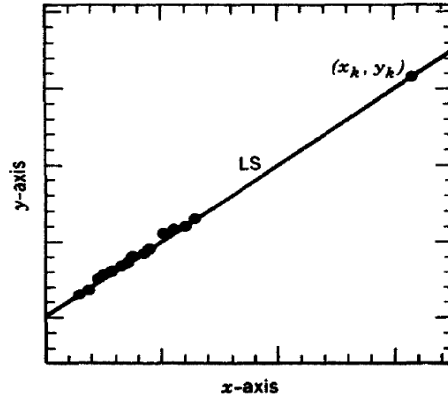
Şekil 3.2 açıklayıcı değişkenlerde aykırı değer olduğunda LS tahminleri üzerindeki etkisini göstermektedir. Şekil 3.2 (a), $(x_1, y_1), \dots, (x_5, y_5)$ beş gözlemleri veriye ilişkin LS tahminine ilişkin kestirim doğrusunu göstermektedir. x_1 değişken değerinin yanlış kaydedildiği varsayılırsa, Şekil 3.2 (b) elde edilir. Şekil 3.2 (b)'de görüldüğü gibi birinci gözlemin LS tahmini üzerindeki etkisi oldukça büyüktür çünkü LS kestirim doğrusunun eğimini değiştirmiştir. Bu gözlem x yönünde aykırı değer olarak adlandırılır. Doğrunun eğimini değiştirdiği için (x_1, y_1) noktası mekanikteki kaldıraç kavramına benzetilerek kaldıraç noktası (leverage points) olarak adlandırılır. x_1 değeri diğer gözlemlerin x değerlerinden çok uzakta olduğundan, Şekil 3.2 (a)'daki LS tahminine göre e_1 artık değeri çok büyük negatif değer olacaktır ve tahmin edilen modelin artık kareler toplamı yani model ile açıklanamayan kısım fazla olacaktır. Şekil 3.2 (b)'de görüldüğü gibi e_1^2 değerini azaltmak için ve en küçük $\sum_{i=1}^5 e_i^2$ toplamını elde etmek için kestirim doğrusu (x_1, y_1) gözlemine doğru eğilmiştir. Diğer dört gözleme ilişkin artık terimleri e_2^2, \dots, e_5^2 biraz artmış olsa bile e_1^2 'nin değeri oldukça azalmıştır ve veriyi daha iyi temsil eden bir LS tahmini elde edilmiştir. Herhangi bir (x_k, y_k) gözlemi Şekil 3.2 (c)'de olduğu gibi verilerin çoğunluğu tarafından belirlenen LS doğrusu üzerinde olup diğer gözlemlere uzak olduğunda artık değeri e_k sıfır olduğu için iyi kaldıraç noktası (good leverage) olarak adlandırılır.



(a)



(b)



(c)

Şekil 3.2. (a) Beş gözleme ilişkin en küçük kareler kestirimi (b) Kaldıraç noktası olduğunda en küçük kareler kestirimi (c) İyi kaldıraç noktası olduğunda en küçük kareler kestirimi [30]

Verinin çoğunluğundan farklı olan bir gözlemi veya gözlem gruplarını ortaya çıkarmak için aykırı değer ölçütleri önerilmiştir. Ayrıca aykırı değerleri ortaya çıkarmanın bir diğer yolu da grafiksel gösterimlerdir. Doğrusal regresyon modelinde yaygın olarak kullanılan aykırı değer ölçütleri şapka matrisinin köşegen elemanları, Student türü artıklar, Cook Uzaklığı, DFFITS ve COVRATIO ölçütleridir.

3.1.1.1. Şapka (Hat) Matrisinin Köşegen Elemanları

(3.2)'deki kestirim denklemini matris-vektör formunda,

$$\hat{y} = X\hat{\beta}$$

şeklinde yazılır. Burada $\hat{\beta}$ değeri yerine yazılırsa,

$$\hat{y} = X'(X'X)^{-1}Xy = Hy \quad (3.5)$$

bulunur. (3.5)'deki H matrisine, y vektörü ile çarpıldığında \hat{y} kestirim değerlerini verdiği için şapka matrisi (hat matrix) denilmektedir. X matrisi sabit terimi içerirse, $1/n \leq h_{ii} \leq 1$ ve $\text{iz}(H) = \text{rank}(X)$ dır [31].

Herhangi bir gözlemin x değeri, diğer gözlemlerin x değerlerinin merkezine uzakta ise bu gözlem, x -yönünde aykırı değer olarak tanımlanır. Şapka matrisi $H = X'(X'X)^{-1}X$ 'in köşegen elemanı h_{ii} değerine kaldıraç (leverage) değeri denir. Büyük h_{ii} değerine sahip gözlemlere büyük kaldıraç değeri (high leverage) denir. Büyük kaldıraç değeri kestirilen

regresyon doğrusunun altında ise doğruyu aşağı, kestirilen regresyon doğrusunun üstünde ise doğruyu yukarı çeker. x-yönündeki aykırı değerler, kaldıraç değeri h_{ii} 'den yararlanarak ortaya çıkarılabilir [16]. n gözlem sayısı p parametre sayısı olmak üzere, $h_{ii} > 2p/n$ ise i . gözlem x-yönünde aykırı değerdir [32].

3.1.1.2. Student Türü Artıklar

Doğrusal regresyon modelinde, y-yönünde ya da dikey aykırı değerleri ortaya çıkarmak için artıklara dayanan ölçütler önerilmiştir. Çünkü y-yönündeki aykırılık artık değerlerinin büyük olmasından kaynaklanmaktadır. y-yönündeki aykırı değerleri ortaya çıkarmak için önerilen ölçütler standartlaştırılmış artıklar ve student türü artıklardır. Standartlaştırılmış artıklar z ile gösterilir ve

$$z = \frac{e_i}{s} \quad (3.6)$$

şeklinde ifade edilir. Burada $s^2 = \frac{1}{n-p} \sum_{i=1}^n e_i^2$ olup σ^2 'nin yansız bir tahmin edicisidir.

Standartlaştırılmış artıkların (-2, 2) arasında olması istenir. Bu değerlerin dışında kalan artık değerlerine sahip gözlemler y-yönünde aykırı değer olarak belirlenir.

Student türü artıklar t_i ile gösterilir ve,

$$t_i = \frac{e_i}{s\sqrt{1-h_{ii}}} \quad (3.7)$$

şeklinde tanımlanır. Burada h_{ii} şapka matrisi olarak bilinen $H = X(X'X)^{-1}X'$ matrisinin köşegen elemanlarıdır. Student türü artıkların (-3, 3) arasında olması istenir. Bu değerlerin dışında kalan artık değerlerine sahip gözlemler y-yönünde aykırı değer olarak belirlenir. Ayrıca standartlaştırılmış kestirim değerlerine karşılık student türü artıkların saçılım grafiğine bakılarak da aykırı değerler görsel olarak belirlenebilir.

3.1.1.3. Cook Uzaklığı

Cook [33], herhangi bir gözlemin parametre tahminleri üzerindeki etkisini ölçmek için bir ölçüt önermiştir. Bu ölçüt i . gözlem veri kümesinden çıkarıldıktan sonra elde edilen

$\hat{\beta}(i)$ ve tüm veriler kullanılarak elde edilen $\hat{\beta}$ arasındaki farka dayanır. Cook Uzaklığı ölçütü olarak bilinen bu ölçüt $CD(i)$ ile gösterilir ve

$$CD^2(i) = \frac{(\hat{\beta} - \hat{\beta}(i))' M (\hat{\beta} - \hat{\beta}(i))}{c} \quad (3.8)$$

şeklinde tanımlanır. Burada $M = X'X$ ve $c = ps^2$ 'dir. $CD^2(i) > 1$ olması i 'nci gözlemin parametre kestirimleri üzerinde etkili olduğunu gösterir [34] [35]. Ayrıca $CD^2(i)$ 'nin n küçük olduğunda $4/n-p$ değerinden ve n büyük olduğunda $4/n$ değerinden büyük olması da i 'nci gözlemin parametre kestirimleri üzerinde etkili olduğunu gösterir [36] [37].

Görsel olarak $CD^2(i)$ değerlerinin saçılım grafiğine bakılarak da etkili gözlemler belirlenebilir. $CD^2(i)$ 'nin saçılım grafiğinde diğer gözlemlerin uzağında kalan gözlemler etkili gözlemler olarak belirlenir. Bu sayede her bir gözlemin $CD^2(i)$ değeri diğer gözlemin $CD^2(i)$ değerine göre değerlendirilmiş olur.

Cook Uzaklığı, kestirim değerleri $\hat{y} = X\hat{\beta}$ kullanılarak,

$$CD^2(i) = \frac{(\hat{y} - \hat{y}(i))' M (\hat{y} - \hat{y}(i))}{ps^2} \quad (3.9)$$

şeklinde de yazılabilir. Bu gösterim ile i 'nci gözlemin kestirim vektörü üzerindeki etkisi ölçülür.

3.1.1.4. DFFITS Ölçütü

Welsch ve Kuh [32] herhangi bir gözlemin kestirim değerleri üzerindeki etkisini ölçmek için DFFITS ölçütünü önermişlerdir. DFFITS ölçütü

$$DFFITS(i) = \frac{(\hat{y} - \hat{y}(i))^2}{s(i)\sqrt{h_{ii}}} \quad (3.10)$$

şeklinde tanımlanır. $DFFITS > 2(p/n)^{1/2}$ olması i 'nci gözlemin kestirim değerleri üzerinde etkili olduğunu gösterir. Görsel olarak DFFITS (i) değerlerinin saçılım grafiğine bakılarak da kestirim değerleri üzerinde etkili olan gözlemler belirlenebilir.

DFFITS (i) 'nin saçılım grafiğinde diğer gözlemlerin uzağında kalan gözlemler etkili gözlemler olarak belirlenir.

3.1.1.5. COVRATIO Ölçütü

Doğrusal regresyon modelinde i.'nci gözlem değerinin parametre tahminlerine ilişkin varyans-kovaryans matrisi üzerindeki etkisini araştırmak için COVRATIO ölçütü önerilmiştir. COVRATIO ölçütü CR(i) ile

$$CR(i) = \frac{|\hat{\sigma}^2(i)(X(i)X(i))^{-1}|}{|\hat{\sigma}^2(X'X)^{-1}|} = \left(\frac{\hat{\sigma}^2(i)}{\hat{\sigma}^2}\right)^p \frac{|(X'X)|}{|(X(i)'X(i))|} \quad (3.11)$$

şeklinde ifade edilir. Burada $\hat{\sigma}^2(X'X)^{-1}$ parametrelere ilişkin varyans-kovaryans matrisidir. CR(i) ölçütü aynı zamanda parametre tahminlerinin doğruluğu üzerinde i.'nci gözlemin etkisini gösterir [38]. CR(i) ölçütü 1 ise i.'nci gözlemin parametre tahminlerinin doğruluğu üzerinde etkili olmadığı, 1'den büyük ise parametre tahminlerinin doğruluğunu artırdığı, 1'den küçük ise parametre tahminlerinin doğruluğunu azalttığı anlamına gelir [36]. Görsel olarak CR(i) değerlerinin saçılım grafiğine parametrelerin doğruluğu üzerinde etkili olan gözlemler belirlenebilir. CR(i) 'nin saçılım grafiğinde diğer gözlemlerin uzağında kalan gözlemler etkili gözlemler olarak belirlenir.

3.1.2. Karma Model için Aykırı Değer Ölçütleri

Karma modelde de sabit modelde olduğu gibi aykırı değerleri ortaya çıkarmak için benzer ölçütler önerilmiştir. Karma modelde de y-yönündeki aykırı değerleri ortaya çıkarmak için artıklara dayanan ölçütler kullanılır.

Karma modellerde artıklar, marjinal artıklar e_m ve koşullu artıklar e_c olarak ele alınır. Marjinal artık, gözlemlenen veri ile tahmini (marjinal) ortalama arasındaki farktır ve

$$e_{mi} = y_i - x_i' \hat{\beta} \quad (3.12)$$

şeklinde tanımlanır. Koşullu artık, gözlemlenen veriler ile gözlemin tahmin edilen değeri arasındaki farktır ve

$$e_{ci} = y_i - x_i' \hat{\beta} - z_i' \hat{u} \quad (3.13)$$

şeklinde tanımlanır. Rasgele etki olmayan bir modelde, iki artık aynıdır. Koşullu artık denmesinin sebebi, $x_i' \hat{\beta} - z_i' \hat{u}$ teriminin y_i 'nin koşullu ortalaması olmasından kaynaklanmaktadır.

Artıklar model varsayımlarını incelemek, aykırı değerleri ve parametre kestirimleri üzerinde etkili gözlemleri ortaya çıkarmak için kullanılır. Ham artıklar e_{mi} ve e_{ci} genellikle bu amaçlar için uygun değildir. Gerçek model hataları ilişkisiz ve eşit varyansa sahip olsa bile, artıklar arasında korelasyon olacaktır ve varyansları farklı olacaktır. Gözlemlerin varyansları farklıysa, ham artıkların yorumlanması daha da zorlaşır. Artıkların eşit varyansının değişkenlik göstermesi sorunu artıkların standartlaştırılması ile çözülür. Rasgele bir değişkenin ortalamadan farkı alınıp standart sapması ile ölçeklendirilmesi standartlaştırma olarak bilinir.

Doğrusal karma modelde ham artıklar sıfır ortalamaya sahiptir, ancak varyansları bilinmemektedir. Dolayısıyla uygulamada standartlaştırma mümkün değildir. Bunun yerine, bir artığı standart sapmasının tahminine bölerek student türü artıkları hesaplanabilir.

3.1.2.1. Şapka Matrislerinin Köşegen Elemanları

Doğrusal karma modellerde de şapka matrisinin köşegen elemanları kaldıraç noktalarını ortaya çıkarmak için kullanılır. Doğrusal karma modellerde iki şapka matrisi tanımlanmıştır.

$$H_1 = X(X'V(\hat{\theta})X)^{-1}X'V(\hat{\theta})^{-1} \quad (3.14)$$

$$H_2 = C^{-1}X(X'V(\hat{\theta})^{-1}X)^{-1}X'C^{-1} \quad (3.15)$$

(3.15)'de θ varyans-kovaryans parametreler vektörünü ifade etmektedir ve $V(\theta) = ZD(\theta)Z' + R(\theta)$ olup $CC' = V(\hat{\theta})$ 'dir. H_1 şapka matrisi idempotent olmasına rağmen, simetrik değildir [39]. Her iki şapka matrisi için de $iz(H_1) = iz(H_2) = rank(X)$ 'dir. Ancak bu iki şapka matrisi için alt ve üst sınır değerleri farklıdır. Örneğin H_1 'in

köşegen elemanları negatif olabilir. Değişkenlerin yüksek derecede ilişkili olduğu modellerde negatif değerler gözlemlenebilir.

$X^* = C^{-1}X$ tanımlandığında H_2 matrisi doğrusal regresyon modelinde olduğu gibi ifade edilebilir:

$$H_2 = X^* (X^{*'} X^*)^{-1} X^{*'} \quad (3.16)$$

Şapka matrisinin büyük köşegen değerleri x yönündeki aykırı değerleri ortaya çıkarırken, H_2 matrisinin büyük köşegen değerleri için bu yorum geçerli değildir. H_2 matrisinin büyük köşegen değerleri dönüştürülmüş x yönündeki aykırı değerleri ortaya çıkarır. X matrisi sabit terim içerdiğinde X^* matrisi sabit terim içermeyeceğinden H_2 'nin köşegen elemanları için alt sınır 1/n olmayacaktır.

3.1.2.2. Student Türü Artıklar

Doğrusal regresyonda olduğu gibi karma modellerde de y-yönündeki aykırı değerleri ortaya çıkarmak için student türü artıklar kullanılır. Student türü artıklar marjinal ve koşullu artıklar için aşağıdaki gibi tanımlanır:

$$e_{mi}^{student} = \frac{e_{mi}}{\sqrt{\text{var}(e_{mi})}} \quad (3.17)$$

$$e_{ci}^{student} = \frac{e_{ci}}{\sqrt{\text{var}(e_{ci})}} \quad (3.18)$$

Doğrusal karma modellerde Student türü artık değerleri (-2,2) aralığının dışında olan gözlemler aykırı değer olarak belirlenir. Diğer bir yaklaşım ise student türü artıkların saçılım grafiklerinin incelenmesidir. Saçılım grafikleri incelendiğinde diğer gözlemlerin uzağında olan gözlemler aykırı değer olarak değerlendirilir [39].

3.1.2.3. Cook Uzaklığı Ölçütü

Karma modellerde sabit etkiler için Cook uzaklığı

$$D(\beta) = (\hat{\beta} - \hat{\beta}(i))' \text{var}[\hat{\beta}]^{-1} (\hat{\beta} - \hat{\beta}(i)) / \text{rank}(X) \quad (3.19)$$

şeklinde tanımlanmıştır. Cook uzaklığı ölçütünün büyük değerleri, parametre tahminleri üzerinde etkili olan gözlemleri belirtir [39]. Cook uzaklığı ölçütünün saçılım grafiği çizdirildiğinde, saçılım grafiğinde diğer gözlemlerin uzağında olan gözlemler parametre kestirimi üzerinde etkili gözlemler olarak belirlenir.

3.1.2.4. DFFITS Ölçütü

Karma modellerde sabit etkiler için DFFITS ölçütü

$$DFFITS(\beta) = (\hat{\beta} - \hat{\beta}(i))' \text{var}[\hat{\beta}(i)]^{-1} (\hat{\beta} - \hat{\beta}(i)) / \text{rank}(X) \quad (3.20)$$

şeklinde tanımlanır.

Karma modellerde herhangi bir gözlemin kestirim değerleri üzerindeki etkisini ölçmek için önerilen DFFITS ölçütü,

$$DFFITS_i = (\hat{y}_i - \hat{y}_i(i)) / \sqrt{\text{var}(\hat{y}_i)} \quad (3.21)$$

şeklinde tanımlanır. DFFITS ölçütünün büyük değerleri kestirim değerleri üzerinde etkili olan gözlemler belirtir [39]. DFFITS ölçütü saçılım grafiği çizdirilerek, saçılım grafiğinde diğer gözlemlerin uzağında olan gözlemler kestirim değerleri üzerinde etkili gözlemler olarak belirlenir.

3.1.2.5. COVTRACE ve COVRATIO Ölçütleri

Karma modellerde de doğrusal regresyon modelinde olduğu gibi parametre tahminlerine ilişkin varyans-kovaryans matrisi üzerinde etkili olan gözlemleri belirlemek için COVTRACE ve COVRATIO ölçütleri önerilmiştir. Bu ölçütler

$$COVTRACE(\beta) = \left| \text{iz} \left(\text{var}[\hat{\beta}]^{-1} \text{var}[\hat{\beta}(i)] \right) - \text{rank}(X) \right| \quad (3.22)$$

$$COVRATIO(\beta) = \frac{\det_{ns} \left(\text{var}[\hat{\beta}(i)] \right)}{\det_{ns} \left(\text{var}[\hat{\beta}] \right)} \quad (3.23)$$

şeklinde tanımlanmıştır. Burada $\det_{ns}(M)$, M matrisinin tekil olmayan kısmının determinantını belirtir. COVTRACE değeri sıfır ve COVRATIO değeri 1 ise gözlemin sabit etkili parametre tahminlerinin doğruluğu üzerinde etkili olmadığı anlamına gelir.

Varyans bileşenleri vektörü θ için COVTRACE ve COVRATIO ölçütleri aşağıdaki gibidir:

$$COVTRACE(\theta) = \left| \text{trace} \left(\text{var} \left[\hat{\theta} \right]^{-1} \text{var} \left[\hat{\theta}(i) \right] \right) - q \right| \quad (3.24)$$

$$COVRATIO(\theta) = \frac{\det_{ns} \left(\text{var} \left[\hat{\theta}(i) \right] \right)}{\det_{ns} \left(\text{var} \left[\hat{\theta} \right] \right)} \quad (3.25)$$

Burada q , $\text{var} \left[\hat{\theta} \right]$ 'nin rankını ifade eder. Kovaryans parametreleri için COVTRACE ve COVRATIO nun hesaplanmasında kullanılan varyans matrisi, ters Hessian matrisinden elde edilir.

3.2. Sağlam (Robust) Yaklaşım

İstatistiksel çıkarımlar bazı varsayımlara dayanmaktadır: Rasgelelik ve bağımsızlık, dağılım hakkındaki varsayımlar, bazı bilinmeyen parametreler için önsel dağılımlar hakkındaki varsayımlar vb gibi. Bu varsayımlar genellikle bulanık bir bilgi veya inancın matematiksel olarak rasyonelleştirmeleridir. Uygulamalı matematiğin diğer tüm dallarında olduğu gibi, bu tür rasyonelleştirmeler veya basitleştirmeler önemlidir. Matematiksel modeldeki küçük bir hata, nihai sonuçlarda sadece küçük bir hataya neden olmalıdır gibi düşünülse de ama bu durum her zaman geçerli değildir. Yaygın olarak kullanılan istatistiksel yöntemlerin bazıları (özellikle normal dağılım varsayımına göre optimize edilenler) varsayımlardan küçük sapmalara karşı aşırı duyarlıdır ve bu sorunu çözmek için çok sayıda alternatif sağlam yöntemler önerilmiştir. Burada sağlamlık, varsayımlardan küçük sapmalara karşı duyarsızlığı ifade etmektedir. Öncelikli olarak dağılımsal sağlamlık ile ilgilenilir. Verinin dağılımının şekli varsayılan modelden biraz sapar. Bu varsayım bozulumu en önemli ve en iyi bilinen durumdur. Dağılım varsayımının sağlanmaması ya da veride aykırı değerler olması durumunda dağılımsal olarak sağlam ya da aykırı değerlere karşı sağlam (dirençli) yöntemler önerilmiştir.

3.3. Sağlamlık Kriterler

Sağlam yaklaşımda bozulma noktası, etki fonksiyonu ve göreceli etkinlik olmak üzere üç kriter vardır. Bu kriterler sağlamlık ifadesini açıklamak için kullanılan ölçütlerdir.

3.3.1. Bozulma Noktası

Bozulma noktası, bir tahmin edicinin anormal değerler almasına neden olabilecek aykırı gözlemlerin en küçük kısmını ifade eder. Başka bir deyişle, bozulma noktası, bir tahmin edicinin büyük anormal değerler almasına neden olabilecek en düşük aykırı değer yüzdesidir. Bozulma noktası sağlamlığın bir ölçüsüdür; bozulma noktası ne kadar büyük olursa tahmin edici o kadar iyidir. Bir tahmin edicinin yüksek bir bozulma noktası varsa, bu tahmin ediciye dirençli tahmin edici denir. Bozulma noktası için elde edilebilecek en yüksek değer, 0,5 değeridir. Çünkü 0,5 değerini aşan bozulma noktasında, normal gözlemlerle aykırı değerler arasında ayırım yapılamamaktadır [40]. Bozulma noktası, istatistiksel bir yöntemin sağlamlığının test edilmesinin bir ölçüsüdür.

n gözlem varsa, örneklem ortalamasının sonlu bozulma noktası $1/n$ 'dir. n sonsuza gittikçe $1/n$ değeri sıfıra yaklaşmaktadır. Bu nedenle ortalamanın (asimptotik) bozulma noktası sıfırdır. Bozulma noktasının sıfır olması ortalamanın sağlam veya dirençli bir tahmin olmadığı anlamına gelmektedir. Bu, olası en yüksek bozulma noktasına (0,5) sahip olan ortancanın tam tersidir [41]. Beş gözlemlili bir veri setini ele alalım. Beş gözlem değerinden en büyük iki değeri aykırı değer olacak şekilde artırırsak, bu gözlemlerin örneklem ortancası üzerinde bir etkisi olmaz. Örneklem ortancasının değişmesi için beş gözlemlili bir veride en az üç aykırı değer olmalıdır. Buradan hareketle genel olarak, n gözlem varsa, örneklem ortanca değerinin değişmesi için gereken minimum aykırı değerlerin sayısı yaklaşık olarak $n/2$ 'dir. Bu nedenle ortancanın sonlu bozulma noktası yaklaşık $1/2$ 'dir [41].

Bozulma noktası, nicel sağlamlılığın bir göstergesidir ve bir kestiricinin etkilenmediği aykırı değerlerin sınırlı bir miktarını verir ve uzak aykırı değerlerin ne kadarının reddedilebileceğini gösterir. Genel olarak bozulma sınırı 0-1 arasındadır [42].

Bozulma noktası kavramını ilk kez 1967 yılında Hodges tarafından kullanılmıştır. Hodges [43] bozulma noktasını “aşırı değerlere tolerans” olarak tanımlamıştır. Daha sonra Hampel [44] tarafından istatistik fonksiyonları için geliştirilmiştir. Huber [45], bozulma noktasını sonlu örneklem için ele almıştır. Buna göre $X = (x_1, \dots, x_n)$ n

boyutlu bir örneklem olsun. Bu örneklem değerleri birçok şekilde bozulabilir. Bunlardan üçü aşağıda verilmiştir:

(1) ε -kirlenme: X örneğine m tane aykırı değer $Y = (y_1, \dots, y_m)$ eklensin. Bu durumda $X' = X \cup Y$ bozulmuş örnekteki aykırı değerlerin oranı $\varepsilon = m / (n + m)$ olur.

(2) ε -yer değiştirme: X örneğinde m tane rasgele değer y_1, \dots, y_m ile değiştirilsin. X' bozulmuş örnekteki aykırı değerlerin oranı $\varepsilon = m / n$ 'dir.

(3) ε -değiştirme: π deneysel ölçümler uzayında tanımlanan keyfi bir uzaklık fonksiyonu olsun. F_n , verilen X örneğine karşılık gelen deneysel ölçüm olsun. X' de G_n deneysel ölçümü ile diğer örneklem olsun, $\pi(F_n, G_n) \leq \varepsilon$. (1)'de olduğu gibi n' , n 'den farklı olabilir. (1) durumunda olduğu gibi, n' örneklem büyüklüğü n 'den farklı olabilir.[46]

$Z = \{(x_{11}, \dots, x_{1p}, y_1), \dots, (x_{n1}, \dots, x_{np}, y_n)\}$ örneğini ele alalım. T bir regresyon tahmin edicisi olsun. Z örneğine ilişkin regresyon tahmin edicisi regresyon katsayılar vektörüdür. Bu durumda

$$T(Z) = \hat{\theta} \quad (3.26)$$

dır. Z örneğindeki m tane değer rasgele değerlerle (aykırı değerlerle) değiştirilerek elde edilen tüm olası bozulmuş Z örneklemi Z' ile gösterilsin. Verideki böyle bir kirlenmenin neden olabileceği maksimum yanlılık

$$yan(m; T, Z) = \sup_{Z'} \|T(Z') - T(Z)\| \quad (3.27)$$

şeklinde ifade edilir. $Yanlılık(m; T, Z)$ değerinin sonsuz olması m aykırı değer T üzerinde çok büyük bir etkiye sahip olabileceği anlamına gelir. Bu da tahmin edicinin bozulduğu anlamına gelmektedir. Bu nedenle, Z örneği için T tahmin edicisinin (sonlu örnek) bozulma noktası,

$$\varepsilon_n^*(T, Z) = \min \left\{ \frac{m}{n}; yan(m; T, Z) = \infty \right\} \quad (3.28)$$

şeklinde gösterilir. Başka bir deyişle, T tahmin edicisinin, $T(Z)$ 'den uzak değerler almasına neden olabilecek en küçük kirlenme oranıdır. Buna göre, bozulma noktası

$$\varepsilon_n^*(T, Z) = 1/n \quad (3.29)$$

şeklinde elde edilir. Bu da örneklem büyüklüğü n arttıkça sifira yaklaşır, bu nedenle en küçük kareler tahmin edicisinin bozulma noktası sıfırdır. Bu en küçük kareler yönteminin aykırı değerlere karşı aşırı duyarlılığını göstermektedir [28].

3.3.2. Etki Fonksiyonu

Bozulma noktası, bir tahmin edicinin anlamsız sonuçlar elde edilmemesi için veride ne kadar bozulma olacağına dair genel bir fikir verirken; etki fonksiyonu, tahmin edicinin herhangi bir noktada az miktardaki bozulmaya nasıl tepki vereceğine dair kesin bir fikir verir [29].

x noktasında aykırı gözlemler ε oranının $T(F)$ üzerindeki göreceli etkisi

$$\frac{T(F_{x,\varepsilon}) - T(F)}{\varepsilon} \quad (3.30)$$

olarak ifade edilir. F 'de T 'nin etki fonksiyonu her x için

$$IF(x) = \lim_{\varepsilon \rightarrow 0} \left[\frac{T(F_{x,\varepsilon}) - T(F)}{\varepsilon} \right] \quad (3.31)$$

şeklinde tanımlanır. Burada $F_{x,\varepsilon}$ fonksiyonu $F_{x,\varepsilon} = (1 - \varepsilon)F + \varepsilon\Delta_x$ olarak tanımlanır.

Bazen T ve F üzerindeki bağımlılığı vurgulamak için $IF(x) = IF_{T,F}(x)$ yazılır [29].

Ayrıca Hampel etki fonksiyonunu

$$IF(x; T, F) = \lim_{t \rightarrow 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t}$$

şeklinde ifade etmiştir.

Etki fonksiyonunun önemi sezgisel yorumunda yatmaktadır: x noktasındaki sonsuz küçük bir bozulmanın tahmin üzerindeki etkisini açıklar. Gözlemlerdeki bozulmanın neden olduğu asimptotik yanlılığı ölçer [27].

3.3.3. Göreli Etkinlik

Belirli bir parametre için bir tahmin edicinin etkinliği olası minimum varyansının gerçek varyansına oranı olarak tanımlanır. Bu oran bir değerine eşit olduğunda tahmin edici etkin olarak kabul edilir. Büyük örneklem olduğunda kabul edilebilir bir etkinliğe sahip tahmin ediciler asimptotik olarak etkindir. Andersen [47], örneklem verisi kullanılıp kitle hakkında çıkarılacak yapılmak istendiğinde bir tahmin edicinin etkinliğinin yüksek olmasının önemli olduğunu vurgulamıştır.

Görelilik etkinlik bir tahmin edicinin etkinliğini karşılaştırmak için bilinen bir yöntemdir. β parametresi için iki tahmin edici T_1 ve T_2 olsun. T_1 daha etkin bir tahmin edici olduğunda T_2 'nin göreli etkinliği T_1 'in hata kareler ortalamasının T_2 'nin hata kareler ortalamasına oranıdır:

$$Etkinlik(T_1, T_2) = \frac{E[(T_1 - \beta)^2]}{E[(T_2 - \beta)^2]} \quad (3.32)$$

Andersen [47]'ye göre göreli etkinlik genellikle asimptotik etkinlik olarak hesaplanır, bu nedenle küçük örneklem için dikkate alınması gereken bir özellik değildir. Bir regresyon tahmin edicisinin göreli etkinliği, hataları normal dağılıma sahip sonsuz örneklem için tahmin edicinin hata kareler ortalaması ile en küçük kareler tahmin edicisinin hata kareler ortalaması arasındaki oran olarak verilir [48].

3.4. Sağlam Regresyon

Sağlam regresyon artıkların dağılımını normal olmadığında ya da aykırı değerler olduğunda kullanılan regresyon yöntemidir. Bu yöntem, aykırı değerlerin olduğu verilerin analizi için elde edilen modellerin aykırı değerlere karşı dayanıklı olmasını sağlar [49]. Araştırmacılar, regresyon modelleri varsayımları sağlamadığı durumlarda ilk olarak verilere dönüşüm uygularlar. Ancak dönüşüm ile aykırı değerlerin etkisinin azaltması ya

da ortadan kaldırılması pek olası değildir. Bu durumda da yanlış tahminler elde edilir. Bu durumda sağlam regresyon aykırı değerlerin etkisine dirençli en iyi yöntemdir. Sağlam regresyon aykırı değerlere karşı dayanıklı sonuçlar elde etmek için kullanılır [50].

Aykırı değerlere karşı dirençli bir yöntemin geliştirilmesi fikri ilk olarak Edgeworth [40] tarafından ortaya konulmuştur. Aykırı değerlerin en küçük kareler yöntemi üzerinde çok büyük etkisi olmasının temel nedeni bu yöntemin e_i artıklarının karesine dayanıyor olmasıdır. Bu nedenle Edgeworth [51] artıkların mutlak değerine dayanan en küçük mutlak sapma yöntemini (least absolute deviation) önermiştir. Bu yöntemde amaç mutlak artık toplamını en küçük yapan parametrelerin bulunmasıdır. Mutlak artık toplamı L_1 ile gösterilir ve

$$L_1 = \sum_{i=1}^n |e_i| \quad (3.33)$$

şeklinde ifade edilir. Bu yöntemde parametre tahminlerini hesaplamak için açık bir formül yoktur. Mutlak sapma tahminleri iteratif algoritma ile hesaplanabilmektedir. L_1 regresyon tahmin edicisi en küçük kareler tahmin edicisinden daha etkindir. Çünkü L_1 regresyon tahmin edicisi y -yönündeki aykırı değerlere karşı dayanıklıdır. Ancak L_1 regresyon tahmin edicisinin bozulma noktası en küçük kareler tahmin edicisi ile aynıdır yani $1/n$ 'dir. Bunun nedeni L_1 regresyon tahmin edicisinin tek bir kötü kaldıraç noktasından bile etkilenmesidir. Literatürde L_1 regresyonundan sonra M, S, MM vb. sağlam tahmin ediciler önerilmiştir [52].

3.4.1. M Tahmin Edicileri

M-tahmin edicileri Huber [53] tarafından önerilmiştir. M tahmin edicileri artıkların karelerinin yerine artıkların bir fonksiyonunun en küçük yapılmasına dayanır. M tahmin edicisi

$$\sum_{i=1}^n \rho(e_i) = \sum_{i=1}^n \rho(y_i - x_i' \beta) \quad (3.34)$$

ifadesinin β 'ya göre en küçük yapılması ile elde edilir. Burada $\rho(\cdot)$ sıfır noktasında tek minimumu olan simetrik bir fonksiyondur. M tahmin edicilerinde M, en çok olabilirlik anlamına gelmektedir. Çünkü $\rho(\cdot)$ fonksiyonu, artıkların dağılımının olabilirlik

fonksiyonu ile ilişkilidir. Artıklar normal dağılıma sahip olduğunda $\rho(e) = \frac{1}{2}e^2$ 'dir. M tahmin edicilerinin ölçeğe göre değişmez olması gerekmektedir yani bu tahmin ediciler artıkların ölçeğinden etkilenir. Ölçeğe göre değişmeyen tahmin edicisi

$$\sum_{i=1}^n \rho\left(\frac{e_i}{s}\right) = \sum_{i=1}^n \rho\left(\frac{y_i - x_i' \beta}{s}\right) \quad (3.35)$$

ifadesinin β 'ya göre en küçük yapılması ile elde edilir. Burada s sağlam ölçek tahminidir. s 'nin yaygın olarak kullanılan sağlam ölçek tahmini

$$s = \frac{\text{medyan}|e_i - \text{medyan}(e_i)|}{0.6745} \quad (3.36)$$

şeklindedir [28]. Bu s tahmini, n büyük olduğunda ve hata dağılımı normal olduğunda, artıkların standart sapmasının yaklaşık olarak yansız tahmin edicisini verir. (3.35)'deki fonksiyonun β 'ya göre türevi alınıp sıfıra eşitlenirse $p+1$ tane doğrusal olmayan eşitlik elde edilir:

$$\sum_{i=1}^n \psi\left(\frac{y_i - x_i' \beta}{s}\right) x_{ij} = 0, \quad j = 1, 2, \dots, p \quad (3.37)$$

Burada ψ , ρ 'nun türevidir ve etki fonksiyonudur. M tahmin edicileri için Huber ve Tukey tarafından ρ fonksiyonları önerilmiştir. Huber'in tanımladığı ρ fonksiyonu,

$$\rho(e) = \begin{cases} e^2/2 & |e| \leq c \\ |2e|c - c^2 & |e| > c \end{cases} \quad (3.38)$$

Şeklindedir [28]. Bu fonksiyonun türevi alınırsa

$$\psi(e) = \begin{cases} e & |e| \leq c \\ c \text{sign}(e) & |e| > c \end{cases} \quad (3.39)$$

elde edilir [28]. Burada $c = 1,345$ 'tir. Tukey'in tanımladığı ρ fonksiyonu

$$\rho(e) = \begin{cases} \frac{c^2}{6} \left\{ 1 - \left[1 - \left(\frac{e}{c} \right)^2 \right]^3 \right\} & |e| \leq c \\ \frac{c^2}{6} & |e| > c \end{cases} \quad (3.40)$$

şeklindedir [28]. Bu fonksiyonun türevi alınırsa

$$\psi(e) = \begin{cases} [1 - (e/c)^2]^2 & |e| \leq c \\ 0 & |e| > c \end{cases} \quad (3.41)$$

olarak elde edilir.

Burada c veri setine bağlı olan sabit bir değerdir. Ayarlama sabiti olarak da bilinen bu değer tahmin edicilerin yüksek etkinlik veya yüksek bozulma noktası gibi istenen özelliklere sahip olmasını ayarlamak için kullanılır. M-tahminlerinin avantajı, diğer sağlam tahminlerden daha hızlı hesaplanabilmeleridir. Dezavantajı, yüksek kaldıraç noktalarına duyarlı olmaları ve bu nedenle yüksek bir bozulma noktasına sahip olmamalarıdır. M tahmin edicilerinin bozulma noktası $1/n$ 'dir. M tahminleri iteratif yöntemler ile elde edilir [28].

3.4.2. S Tahmin Edicileri

S tahmin edicileri Rousseeuw ve Yohai [43] tarafından önerilmiştir. S tahmin edicileri, M tahmin edicileri ile aynı asimptotik özelliklere sahip olup verilerde aykırı değerlerin % 50'sini tolere edebilir. Bu nedenle S tahmin edicileri yüksek bozulma noktasına sahip sağlam regresyon tahmin edicileridir. S tahmin edicileri artıkların yayılımının minimize edilmesine dayanır:

$$\min_{\beta} s(e_1(\beta), \dots, e_n(\beta))$$

Burada ölçek tahmini $\hat{\sigma} = s(e_1(\hat{\beta}), \dots, e_n(\hat{\beta}))$ 'dir. En küçük kareler yönteminde artıkların varyansını minimum yapılırken, S tahmininde artıkların ölçeğinin sağlam M tahmini minimum yapılır. Bu nedenle en küçük kareler tahmini, S tahminin daha az sağlam, özel bir durumu olarak görülebilir. Yayılım $s(e_1(\hat{\beta}), \dots, e_n(\hat{\beta}))$ olmak üzere $\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{e_i}{s}\right) = K$ ifadesinin çözümüdür [28]. Burada K sabit olup yaygın olarak $E_{\phi}[\rho]$ ifadesine eşit olacak biçimde alınır ve ϕ standart normal dağılım fonksiyonudur. ρ fonksiyonu aşağıdaki özellikleri sağlamalıdır:

1. ρ fonksiyonu simetrik ve sürekli türevlenebilen bir fonksiyondur ve $\rho(0) = 0$ 'dır.
2. ρ fonksiyonunun $[0, c]$ aralığında kesinlikle arttığı, $[c, \infty)$ aralığında sabit olduğu bir $c > 0$ değeri vardır.
3. $\frac{K}{\rho(c)} = \frac{1}{2}$

Üçüncü koşul % 50'lik bir bozulma noktası elde etmek için gereklidir. S tahmini için Tukey'in çiftkare (bisquare) ρ fonksiyonu kullanıldığında Rousseeuw and Yohai [54] üçüncü koşulun sağlanması için $c=1.547$ olarak alınması gerektiğini belirtmişlerdir. Tukey'in çiftkare (bisquare) ρ fonksiyonu aşağıdaki gibi tanımlanmıştır:

$$\rho(e) = \begin{cases} \frac{e^2}{2} - \frac{e^4}{2c^2} + \frac{e^6}{6c^4} & |e| \leq c \\ \frac{e^2}{6} & |e| > c \end{cases} \quad (3.41)$$

[48]. S tahmin edicilerin bozulma noktası $n \rightarrow \infty$ için 0.5'dir. Bu tür tahmin edicilerin asimptotik etkinliği tanımlanan ağırlık fonksiyonuna bağlıdır. Ayar sabitleri (c) tahmin edicinin hem yüksek bozulma noktasına sahip hem de etkinliği yüksek olacak şekilde seçilememektedir. Rousseeuw ve Yohai [54], c ayar sabitinin seçiminin Tukey çiftkare ağırlık fonksiyonu kullanılarak tanımlanan S-tahmin edicisinin bozulma noktası ve etkinliği üzerindeki etkisini göstermişlerdir. S tahmin edicisi yüksek bozulma noktasına sahip olsa da etkinliğinin düşük olması nedeniyle genellikle sağlam regresyon yöntemi olarak kötü bir tercihtir. Ancak, yüksek bozulma noktasına sahip olduğu için daha karmaşık sağlam regresyon tahmin süreçleri içinde başlangıç tahmini olarak kullanılmaktadır [28].

3.4.3. MM Tahmin Edicileri

MM-tahmin edicileri etkinliğinin yüksek olmasının yanı sıra yüksek bir bozulma noktasına da sahiptir. MM tahmin edicilerinin, hatalar normal dağıldığında bozulma noktası 0,5 olup, asimptotik etkinliği % 85 ile % 95 arasında değişmektedir.

MM tahminleri üç aşamada elde edilir. İlk aşamada, $\hat{\beta}$ başlangıç tahminlerini bulmak için yüksek bozulma noktasına sahip bir tahmin edici kullanılır. Tahmin edicinin etkin olmasına gerek yoktur. Artık değerleri $e_i(\hat{\beta}) = y_i - x_i' \hat{\beta}$ hesaplanır. Bu aşamada genellikle Huber ya da çiftkare ağırlıklarının kullanıldığı S-tahmin edicisi tercih edilir. İkinci aşamada hesaplanan $e_i(\hat{\beta})$ artıkları, (3.35) eşitliğinde yerine konularak artık ölçeğinin M-tahmini $\hat{\sigma}_s = s(e_1(\hat{\beta}), \dots, e_n(\hat{\beta}))$ elde edilir. İkinci aşamada kullanılan ağırlık fonksiyonu ρ_0 ile gösterilsin. Son aşamada MM tahmin edicisi yeniden azalan $\psi_1(e) = \frac{\partial \rho_1(e)}{\partial e}$ fonksiyonunu kullanarak elde edilen β 'nin M tahmin edicisi ve ikinci aşamada elde edilen $\hat{\sigma}_s$ ölçek tahmini olarak tanımlanır. Dolayısıyla MM tahmin edicisi $\hat{\beta}_{MM}$;

$$\sum_{i=1}^n x_{ij} \psi_1\left(\frac{y_i - x_i' \beta}{\hat{\sigma}_s}\right) = 0, \quad j = 1, 2, \dots, p \quad (3.42)$$

eşitliğinin bir çözümüdür. Bu aşamada kullanılan ağırlık fonksiyonu ρ_1 , ilk aşamada kullanılan ağırlık fonksiyonu ρ_0 ile aynı olmak zorunda değildir ancak aşağıdaki özellikleri sağlamalıdır:

1. ρ fonksiyonu simetrik ve sürekli türevlenebilen bir fonksiyondur ve $\rho(0) = 0$ 'dır.
2. ρ fonksiyonunun $[0, c]$ aralığında kesinlikle arttığı, $[c, \infty)$ aralığında sabit olduğu bir $c > 0$ değeri vardır
3. $\rho_1(e) \leq \rho_0(e)$

[48]. MM tahmininde ikinci aşamada ölçek tahmini değişmemektedir. Ölçek tahmininin etkinliğinin düşük olması, ikinci aşamada hesaplanan regresyon tahmininin asimptotik etkinliği üzerinde bir etkisinin olmadığını gösterir. Ölçek tahmininin sadece yansız olması gerekir [28].

3.4.4. Tasarım Uyarlamalı Ölçek Tahmini (Design Adaptive Scale Estimation)

Klasik en küçük kareler yönteminde artıkların örneklem varyansının, σ^2 için yansız bir tahmin olduğu bilinmektedir. σ^2 'nin yansız tahmin edicisi artık kareler toplamının n yerine $(n-p)$ değerine bölünmesiyle

$$\hat{\sigma}_{LS}^2 = \frac{1}{n-p} \sum_{i=1}^n e_i^2 \quad (3.43)$$

olarak elde edilir. S-tahmin edicileri için de yansız tahmin edici, en küçük kareler yönteminde olduğu gibi, $(n-p)$ değerine bölünerek

$$\frac{1}{n-p} \sum_{i=1}^n \rho(e_i / \hat{\sigma}_s) = \kappa_s \quad (3.44)$$

şeklinde elde edilir[5].

Ölçek tahminindeki herhangi bir yanlılık, ψ fonksiyonundaki ayarlama sabitinin değişimi olarak görülebilir. Bu nedenle, yanlılığın miktarına bağlı olarak, ayarlama sabiti ve bununla birlikte regresyon tahmininin etkinliği kullanıcı tarafından ayarlanandan farklı olabilir.

β katsayılarının önem kontrolünün yapılabilmesi için ölçek tahmini gereklidir. β 'nin k . bileşeninin sıfır olduğu yokluk hipotezini test etmek için t-istatistiği,

$$t_k = \frac{\hat{\beta}_k}{\sqrt{\text{kov}(\hat{\beta}_k)}} \quad (3.45)$$

şeklinde dir. Paydadaki kovaryans matris tahmini, ölçek tahmini $\hat{\sigma}$ 'yi da içerir. $\hat{\sigma}$ ölçek tahmini yanlı ise parametlerin önem kontrolü de yanıltıcı olacaktır. Ayrıca, ölçek tahmininin etkinliğinin düşük olması, testlerin gücünü olumsuz yönde etkileyecektir. Sonuç olarak, ölçeğin sadece sağlam bir şekilde tahmin edilmesi değil, aynı zamanda yansız ve etkin olması da önemlidir [5].

Maronna ve Yohai [55] tarafından S ölçek tahmini $\hat{\sigma}_s$ 'in yanlı olduğu ve bunun sonucunda MM-regresyon tahmini $\hat{\beta}_{MM}$ 'in de etkinliğinin azaldığını fark etmişlerdir ve

yalnızca p/n oranına ve kullanılan ρ fonksiyonuna bağlı olan ölçek tahmininin yanlılığını ortadan kaldırmak için bir düzeltme önermişlerdir. Bu düzeltme en küçük kareler yöntemindekine benzer olarak paydadaki n değerinin $(n-p)$ değeri ile değiştirilmesidir. Ancak, Koller ve Stahel [56] önerilen düzeltmenin yanlılığı güvenilir bir şekilde ortadan kaldırmak için yeterli olmadığını ve çok basit bir yaklaşım olduğunu belirtmişlerdir.

Koller ve Stahel [56] daha doğru bir düzeltme bulmak için, (3.44)' deki artıkların sağlam bir tahminden geldiğini ve farklı, sağlam bir ölçek tahmininin örneklem varyansının yerine kullanılması gerektiğini belirtmişlerdir.

Doğrusal regresyonda, artık değerleri değişen varyanslılığa sahip olabilir. Doğrusal regresyonda artıkların varyansı, kaldıraç değerlerine bağlıdır ($Var(e_i) = \sigma^2(1-h_{ii})$). Benzer bir eşitlik, sağlam regresyon için de geçerlidir. Koller ve Stahel [57] çalışmasında, ölçegi tahmin ederken değişen varyanslılık sorununu düzeltmeyi önermişlerdir. Ölçeğin M-tahminleri

$$\sum_{i=1}^n (w_{ölçek} \left(\frac{e_i}{\hat{\sigma}}\right) \left(\frac{e_i}{\hat{\sigma}}\right)^2 - \kappa) = 0 \quad (3.46)$$

eşitliğinin çözümüdür. Burada $w_{ölçek}$ kullanılan ρ fonksiyonuna bağlıdır ve κ tahminin tutarlılığını sağlayan sabittir [5].

Klasik doğrusal regresyon modeli için $w_{ölçek}$ değerleri ve κ , 1 değerine eşittir. Koller ve Stahel [56] önerdiği yaklaşım ile ölçegin yansız tahminin nasıl elde edilebileceğini göstermek için klasik regresyon modelini ilk olarak ele almış ve ölçegi tahmin etmeden önce artıkları (e_i) τ_i sabitleri ile normalleştirmeyi önermişlerdir. Buna göre (3.46) eşitliği klasik regresyon modeline göre düzenlenirse ölçek tahmini

$$\sum_{i=1}^n \tau_i^2 \left(\left(\frac{e_i}{\tau_i \hat{\sigma}_{LS}} \right)^2 - 1 \right) = 0$$

eşitliğinin çözümü olarak elde edilir. Burada $\tau_i = \sqrt{1-h_{ii}}$ alınırsa doğrusal regresyondaki yansız ölçek tahmini elde edilmiş olur. Koller ve Stahel [56] doğrusal regresyonda ölçek

tahminini bu şekilde ifade ederek, sağlam doğrusal regresyon modelinde de sağlam bir ölçeği tanımlamak için iyi bir başlangıç noktası elde etmişlerdir. Buradan hareketle Koller ve Stahel [56], sağlam regresyon için tasarım uyarlanabilir ölçeği (Design Adaptive Scale, DAS), $\hat{\sigma}_D$, önermişlerdir [5]. Buna göre $\hat{\sigma}_D$

$$\sum_{i=1}^n \tau_i^2 w \left(\frac{e_i}{\tau_i \hat{\sigma}_D} \right) \left[\left(\frac{e_i}{\tau_i \hat{\sigma}_D} \right)^2 - \kappa_D \right] = 0 \quad (3.47)$$

eşitliğinin çözümüdür. τ_i sabiti, (3.47) eşitliğinde toplamın beklenen değerini yaklaşık olarak sıfır yapan bir değer olarak tanımlanmıştır. Buna göre

$$E \left[w \left(\frac{e_i}{\tau_i \hat{\sigma}_D} \right) \left[\left(\frac{e_i}{\tau_i \hat{\sigma}_D} \right)^2 - \kappa_D \right] \right] = 0 \quad (3.48)$$

dır. Artıkların değişen varyanslılığı dikkate alındığı için, modeldeki sabit κ_D
 $\kappa_D = E_0 [w(\varepsilon)\varepsilon^2] / E_0 [w(\varepsilon)] \quad (3.49)$

şeklinde gösterilir.

3.4.5. SMDM Tahmin Edicileri

DAS-tahmini, MM-tahmin edicisi elde edildikten sonraki ölçeğin tahminidir. MM tahmini $\hat{\beta}_{MM}$, yanlı S-tahmini $\hat{\sigma}_s$ 'e bağlı olduğundan, $\hat{\beta}_{MM}$ 'nin etkinliği istenenden daha düşük olabilir. Bu sorunu çözmek için ilk olarak DAS-tahmini bulunur ve ardından $\hat{\sigma}_s$ yerine $\hat{\sigma}_D$ ile bir başka M-tahmini hesaplanır. Sonuçta ortaya çıkan tahmin, bir S- ve ardından bir M- sonra D- (DAS-tahmin) ve son olarak da M-tahmininden oluşan bir tahmin zinciri olarak görülebilir. Bu tahmin zinciri SMDM olarak adlandırılmıştır [56].

3.5. Doğrusal Karma Modelde Sabit ve Rasgele Etkinin Sağlam Tahmini

Doğrusal karma modeller genel olarak (2.2)'deki gibi gösterilir. Bu gösterime göre parametre tahminleri elde edilirken rasgele etkilerin kovaryans matrisinin tersi, V^{-1} , bulunmalıdır. Matris tersi bulunurken varyans bileşenlerinden birinin sıfır ya da sıfıra yakın tahmin edilmesi gibi sorunlar ortaya çıkmaktadır. Bu sorunu ortadan kaldırmak için

karma modeller küresel rasgele etkilere göre tanımlanmıştır. Buna göre (2.2)'deki model küresel rasgele etkiler kullanıldığında aşağıdaki gibi ifade edilir:

$$y = X\beta + ZB_u u^* + B_e \varepsilon^* \quad (3.50)$$

Burada küresel rasgele etkiler vektörü u^* olup, sıfır varyans bileşenlerine karşılık gelen u^* bileşenlerinin sıfır olduğu varsayılır. Ayrıca $u = B_u u^*$ dır.

Bu model altında, gözlemlerin dağılımı $y \sim N(X\beta, \sigma^2 V(\theta))$, rasgele etkilerin dağılımı $u \sim N(0, \sigma^2 V_u(\theta))$ olup rasgele küresel etkilerin dağılımı $u^* \sim N(0, \sigma^2 I_q)$ ve hataların dağılımı $\varepsilon^* \sim N(0, \sigma^2 I_n)$ 'dir. Ayrıca $V(\theta) = ZV_u(\theta)Z' + V_e$, $V_u(\theta) = B_u B_u'$, $V_e = B_e B_e'$ olup $\varepsilon^* = B_e^{-1} \varepsilon$ 'dur. Burada toplam gözlem sayısı n ($i = 1, \dots, n$), rasgele etkilerin sayısı q ($j = 1, 2, \dots, q$), $V_u(\theta)$ blok köşegen matris olup blokların sayısı K ($k = 1, \dots, K$) ve tahmin edilecek kovaryans parametrelerinin (θ) sayısı r 'dir. Bu gösterimde B_u 'nin tersinin hesaplanmasına gerek yoktur. Bu nedenle varyans bileşenlerinin herhangi birinin sıfır olması sorun olmamaktadır.

(3.50) modeli için en çok olabilirlik fonksiyonunun logaritması

$$\ell(\beta, \sigma, \theta | y) = n \log 2\pi + \log |\sigma^2 V(\theta)| + (y - X\beta)' V(\theta)^{-1} (y - X\beta) / \sigma^2 \quad (3.51)$$

şeklinde yazılır. Verilen θ , β ve σ için rasgele etkilerin BLUP tahmini, (2.35) eşitliğine benzer biçimde

$$u^* = \sigma^2 B_u' Z' V(\theta)^{-1} (y - X\beta) \quad (3.52)$$

olarak elde edilir. (3.52) eşitliği (3.51)'de yerine konulursa en çok olabilirlik fonksiyonunun logaritması (2.25) eşitliğine benzer biçimde

$$\begin{aligned} \ell(\theta, \beta, \sigma, u^* | y) &= n \log 2\pi + \log |\sigma^2 V(\theta)| + ((y - X\beta - ZB_u u^*)' (B_e^{-1})' B_e^{-1} (y - X\beta - ZB_u u^*) \\ &\quad + u^{*'} u^*) / \sigma^2) \\ &= n \log 2\pi + \log |\sigma^2 V(\theta)| + ((y - X\beta - ZB_u u^*)' V_e^{-1} (y - X\beta - ZB_u u^*) \\ &\quad + u^{*'} u^*) / \sigma^2) \end{aligned} \quad (3.53)$$

olarak elde edilir. (3.53) eşitliğinin β ve u^* a göre kısmi türevi alınarak Henderson'un karma model eşitlikleri (2.28) eşitliğine benzer biçimde aşağıdaki gibi elde edilir:

$$\begin{bmatrix} X'V_e^{-1}X & X'V_e^{-1}ZB_u \\ B_u'Z'V_e^{-1}X & B_u'Z'V_e^{-1}ZB_u + I_q \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{u}^* \end{bmatrix} = \begin{bmatrix} X'V_e^{-1}y \\ Z'V_e^{-1}y \end{bmatrix} \quad (3.54)$$

Bu denklemlerin çözümleri, $\hat{\beta}$ ve \tilde{u}^* , (3.3) eşitliğindeki küresel rasgele etkilerin, u^* , BLUP tahmini ve sabit etkilerin, β , en çok olabilirlik tahminine karşılık gelir.

3.5.1. Tahmin Denklemlerinin Sağlamaştırılması

Koller [5], karma modelde olabilirlik fonksiyonunun logaritması yerine (3.54)'deki Henderson'ın karma model eşitliklerinden elde edilen tahmin denklemlerini sağlamaştırmıştır. Tahmin denklemleri sağlamaştırılırken yeniden azalmayan ψ (örneğin Huber fonsiyonu) fonksiyonları kullanılmıştır.

Varyans bileşenleri θ ve σ 'nın bilindiği varsayımı altında, (3.53)'ün β ve u^* 'ya göre kısmi türevleri alınırsa aşağıdaki eşitlikler elde edilir:

$$\begin{aligned} X'(B_e^{-1})'B_e^{-1}(y - X\hat{\beta} - ZB_u\tilde{u}^*) / \sigma &= 0 \\ (B_u'Z'(B_e^{-1})'(y - X\hat{\beta} - ZB_u\tilde{u}^*) - \tilde{u}^*) / \sigma &= 0 \end{aligned} \quad (3.55)$$

(3.55)'deki eşitliklerde $e^* = (y - X\hat{\beta} - ZB_u\tilde{u}^*)$ olmak üzere, e^* ve \tilde{u}^* yerine $\psi(e^*)$ ve $\psi(\tilde{u}^*)$ fonksiyonları yazılarak eşitlikler sağlamaştırılır. B_u matrisinin köşegen bir matris olduğu durumda sağlamaştırma basittir ancak köşegen olmadığına işlemler zorlaşmaktadır [5].

B_u matrisi köşegen bir matris olduğunda sağlam tahmin denklemleri

$$\begin{aligned} X'(B_e^{-1})'\psi(e^* / \sigma) / \lambda_e &= 0 \\ B_u'Z'(B_e^{-1})'\psi(e^* / \sigma) / \lambda_e - \psi(\tilde{u}^* / \sigma)\lambda_u &= 0 \end{aligned} \quad (3.56)$$

olarak elde edilir. Burada $\lambda = E_0[\psi']$ 'dir. $1/\lambda_e$ ve $1/\lambda_b$ ölçeklendirme faktörleri, bir M -tahmininin etki fonksiyonu ile aynıdır. M tahminleri için, ölçeklendirme faktörünün

olması tahminleri deęiřtirmez ve bu nedenle genellikle denklemlerde göz ardı edilir. Ancak, karma modeller için, rasgele etkilerin uygun şekilde cezalandırılmasını sağlamak için ölçeklendirme faktörleri gerekir. $\psi(e^*) \equiv \psi(\tilde{u}^*)$ ise, ölçeklendirme faktörleri göz ardı edilir [5].

$$\Delta\beta = \hat{\beta} - \beta, \Delta u^* = \tilde{u}^* - u^* \text{ olsun ve}$$

$$\hat{\psi}_e = \psi(e^* / \sigma), \psi_e = \psi(\varepsilon^* / \sigma), D_e = \text{Diag}(\psi'_e(\varepsilon^* / \sigma))$$

$$\hat{\psi}_u = \psi(\tilde{u}^* / \sigma), \psi_u = \psi_u(u^* / \sigma), D_u = \text{Diag}(\psi'_u(u^* / \sigma))$$

tanımlansın. $\hat{\psi}_e$ ve $\hat{\psi}_u$ 'nun β ve u^* deęerlerine göre birinci dereceden açılımları

$$\hat{\psi}_e \approx \psi_e - D_e B_e^{-1} (X\Delta\beta + ZB_u \Delta u^*) / \sigma$$

$$\hat{\psi}_u \approx \psi_u - D_u \Delta u^* / \sigma$$

řeklindedir. $B_e = I$ olmak üzere (3.54) yeniden düzenlenirse

$$\begin{bmatrix} X'D_e X / \lambda_e & X'D_e ZB_u / \lambda_e \\ B_u' Z'D_e X / \lambda_e & B_u' Z'D_e ZB_u / \lambda_e + D_u / \lambda_u \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{u}^* \end{bmatrix} \approx \begin{bmatrix} XD_e y / \lambda_e - XD_e(\varepsilon^* - \psi_e) / \lambda_e \\ B_u' Z'D_e y / \lambda_e - B_u' Z'D_e(\varepsilon^* - \psi_e) / \lambda_e - (D_u u^* + \psi_u) / \lambda_u \end{bmatrix} \quad (3.57)$$

elde edilir.

B_u matrisi köřegen bir matris olmadığında, u^* bileřenini saęlamlařtırmak kolay deęildir. Rasgele etkinin sabit terimi ve eęiminin iliřkili olduęu bir model ele alınsın (blok büyüklüęü iki olan). Sabit terim ve eęim rasgele etkinin (u) bileřenleridir. Küresel rasgele etkiler u^* , rasgele etkilerin doęrusal bir dönüşümüdür; $u^* = B_u^{-1}u$. Dolayısıyla u^* bileřeninin saęlamlařtırılması, sabit terim ve eęimin doęrusal bir kombinasyonunun deęiřtięi anlamına gelir. Ayrıca, bu doęrusal kombinasyon iki parametrenin korelasyonuna baęlıdır. Bu nedenle, bir ψ -fonksiyon bileřeninin basitçe uygulanması mantıklı olmadıęı için blok aęırlıkları (blokwise) kullanılmıřtır.

$k(j)$, j 'nci rasgele etkinin k 'nci bloęunun bir fonksiyonu olsun. Tahmin edilen rasgele etkilerin karesel Mahalanobis uzaklıęı

$$d = d(u_{k(j)} / \sigma), \quad j = 1, \dots, q,$$

olup $d(u_k) = u_k^* u_k^*$ 'dir. j'nci rasgele etkinin sağlamlık ağırlığı $w_u(d_j)$

$$w_u(d) = \begin{cases} \psi_u(\sqrt{d}) / \sqrt{d} & d \neq 0 \\ \psi_u'(0) & d=0 \end{cases}$$

şeklinde yazılır [5]. B_u matrisi köşegen olduğunda, ağırlık fonksiyonunda sıfırdan daha uzak değerler için daha küçük bir ağırlık atanır. B_u matrisi köşegen olmadığına köşegen olduğu durumdakine benzer şekilde uzak değerlere daha küçük ağırlık atanır ancak ağırlıklar orijinden uzaklığa göre atanır. Kovaryans matrislerinin ve konum parametresinin sağlam tahmini Stahel [58] tarafından çalışılmıştır. Bu konu ile ilgili temel sonuçlar biraz farklı biçimde Hampel vd. [27] çalışmasında da bulunabilir. Bu çalışmalarda optimal B-sağlam tahmin edicileri veren ağırlıklandırma fonksiyonlarının yanı sıra, tahmin edicilerin asimptotik etkinlikleri için basit ifadeler de elde edilmiştir.

Sağlamlılık ağırlıkları köşegen ağırlık matrisi olarak

$$W_u(d) = \text{Diag}(w_u(d_{k(j)})), \quad j = 1, 2, \dots, q$$

şeklinde gösterilir. λ_e ile çarpılan sağlam tahmin denklemleri,

$$X'(B_e^{-1})' \psi_e(e^* / \sigma) = 0 ,$$

$$B_u' Z'(B_e^{-1})' \psi_e(e^* / \sigma) - \wedge_u w_u(\hat{d}) \tilde{u}^* / \sigma = 0 \quad (3.58)$$

olarak elde edilir. Burada $\wedge_u = \text{Diag}(\lambda_e / \lambda_{u,j})$, $j=1, \dots, q$, köşegen matris olup elemanları

blok büyüklüğü $s_{k(j)}$ bağlıdır. $\lambda_{u,j} = \tilde{\lambda}(s_{k(j)})$ ve $u^* \sim N(0, I_s)$ olup

$$\tilde{\lambda}(s) = E_0 \left[\frac{\partial}{\partial u_1^*} (\omega_u(u^* u^*) u_1^*) \right]$$

şeklinde gösterilir.

3.5.1.1. Ölçek ve Kovaryans Parametrelerinin Sağlam Tahmini

Daha önceki bölümlerde kovaryans parametreleri θ ve σ ölçeğinin bilindiği varsayılmıştı. Bu bölümde bu parametrelerin sağlam tahminlerinin nasıl elde edildiği

anlatılmıştır. Kovaryans parametrelerinin ve ölçek parametresinin tahminin de kısıtlanmış en çok olabilirlik yöntemi ele alınmıştır. Bu yönteme göre (REML) tahmin denklemlerinin sağlamaştırılması yaklaşımı incelenmiştir.

(3.53)'deki olabilirlik fonksiyonunun logaritmasının θ ve σ parametrelerine göre kısmi türevleri alınırsa

$$e^{*t} e^* + \tilde{u}^{*t} \tilde{u}^* = n\hat{\sigma}^2 \quad (3.59)$$

$$e^{*t} B_e^{-1} Z \frac{\partial B_u(\hat{\theta})}{\partial \theta_l} \tilde{u}^* = \frac{\hat{\sigma}^2}{2} iz \left(V(\hat{\theta})^{-1} Z \frac{\partial V_u(\hat{\theta})}{\partial \theta_l} Z' \right), \quad l=1,2,\dots,r \quad (3.60)$$

eşitlikleri elde edilir. Bu tahmin denklemleri, ε^* ve u^* terimlerini içerdiği için bu denklemler sağlamaştırma için uygun değildir. (3.59) denklemini küresel etkilerin dağılımı σ 'ya göre tanımlandığı için tahmin edilmiş küresel rasgele etkileri içerir. Mutlak parametrelendirme kullanılarak daha basit tahmin denklemleri elde edilir. En çok olabilirlik tahminleri parametre dönüşümüne göre değişmezdir, bu nedenle bir parametrelendirme ile bulunan çözüm aynı zamanda başka bir parametrelendirme için de bir çözümdür [5]. Bu nedenle (3.59)'daki tahmin denklemini mutlak parametrelendirme ile

$$e^{*t} e^* = \hat{\sigma}^2 iz(V(\hat{\theta})^{-1} V_e)$$

olarak elde edilir. (3.55)'ten $\tilde{u}^* = B_u' Z'(B_e^{-1})' e^*$ olup (3.60)'daki tahmin denkleminde yerine yazılırsa

$$\tilde{u}^{*t} Q_l(\hat{\theta}) \tilde{u}^* = \frac{\hat{\sigma}^2}{2} tr \left(V(\hat{\theta})^{-1} Z \frac{\partial V_u(\hat{\theta})}{\partial \theta_l} Z' \right), \quad l=1,2,\dots,r.$$

olarak elde edilir. Burada $Q_l(\theta) = B_u(\theta)^{-1} \frac{\partial B_u(\theta)}{\partial \theta_l}$ 'dir.

$\sigma^2 V_u(\theta)$ ve σ^2 'nin en çok olabilirlik tahminlerinin yanlı olduğu bilinmektedir. Bu yanlılık, en çok olabilirlik yerine kısıtlanmış en çok olabilirlik (REML) kullanılarak önlenebilir. REML tahmin denklemleri farklı şekillerde elde edilebilir ve burada elde edilen denklemlerin tahmin edilmesi için basit bir yol sağ tarafı, sol taraftaki tahminlerin

beklenen değerleri ile değiştirmektedir [36]. Buna göre REML tahmin denklemlerinin sağ tarafı

$$e^{*'} e^* = E \left[e^{*'} e^* \right] \quad (3.61)$$

$$\tilde{u}^{*'} Q_l(\hat{\theta}) \tilde{u}^* = tr \left(E \left[\tilde{u}^* \tilde{u}^{*'} \right] Q_l(\hat{\theta}) \right), \quad l=1,2,\dots,r \quad (3.62)$$

olarak ifade edilir.

3.5.1.2. Tasarım Uyarlamalı Ölçek (DAS) Tahmini ile Ölçek ve Kovaryans Parametrelerinin Sağlam Tahmini

Tasarım uyarlamalı ölçek yaklaşımı (3.47)'e uygulandığında

$$\sum_{i=1}^n \tau_{e,i}^2 \omega_e^{(\sigma)} \left(\frac{e_i^*}{\tau_{e,i} \hat{\sigma}} \right) \left[\left(\frac{e_i^*}{\tau_{e,i} \hat{\sigma}} \right)^2 - \kappa_e^{(\sigma)} \right] = 0 \quad (3.63)$$

elde edilir. Burada “ $\cdot^{(\sigma)}$ ” üst simgesi ölçek ve kovaryans parametreleri için kullanılan ağırlık fonksiyonlarını, sabit etkiler için kullanılan ağırlık fonksiyonlarından ayırt etmek için kullanılmıştır. Doğrusal regresyon modelinde olduğu gibi, $\tau_{e,i}$ terimi (3.47)'deki i . toplamın beklenen değerini sıfır yapan değer olarak tanımlanmıştır. Buna göre

$$E \left[\omega_e^{(\sigma)} \left(\frac{e_i^*}{\tau_{e,i} \hat{\sigma}} \right) \left(\frac{e_i^*}{\tau_{e,i} \hat{\sigma}} \right)^2 - \kappa_e^{(\sigma)} \right] = 0 \quad (3.64)$$

ve

$$\kappa_e^{(\sigma)} = E_0 \left[\omega_e^{(\sigma)}(\varepsilon) \varepsilon^2 \right] / E_0 \left[\omega_e^{(\sigma)}(\varepsilon) \right] \quad (3.65)$$

dır. Ölçek tahminleri için kullanılan ağırlıklandırma fonksiyonları, sabit ve rasgele etkilerin tahmininde kullanılan karesel sağlamlık ağırlıklarını verir, dışbükey ρ fonksiyonlar için, $\omega_e^{(\sigma)}(x) = \left(\psi_e^{(\sigma)}(x) / (x) \right)^2$, $\omega_e^{(\sigma)}(0) = \psi_e^{(\sigma)'}(0)$.

B_u matrisi köşegen olduğunda, $Q_l(\hat{\theta})$ terimi birlere ve sıfırlara indirgenir ve bu nedenle (3.62) eşitliğinde ortadan kalkar. Geriye kalan denklem için tasarım uyarlamalı ölççek yaklaşımı

$$\sum_{j=1}^q \tau_{u,j}^2 \omega_u^{(\sigma)} \left(\frac{\tilde{u}_j^*}{\tau_{u,j} \hat{\sigma}} \right) \left[\left(\frac{\tilde{u}_j^*}{\tau_{u,j} \hat{\sigma}} \right)^2 - \kappa_u^{(\sigma)} \right] = 0 \quad (3.66)$$

şeklinde elde edilir ve

$$E \left[\omega_u^{(\sigma)} \left(\frac{\tilde{u}_j^*}{\tau_{u,j} \hat{\sigma}} \right) \left[\left(\frac{\tilde{u}_j^*}{\tau_{u,j} \hat{\sigma}} \right)^2 - \kappa_u^{(\sigma)} \right] \right] = 0 \quad (3.67)$$

$$\text{dır. Burada } \kappa_u^{(\sigma)} = E_0 \left[\omega_u^{(\sigma)} (u^*) u^{*2} \right] / E_0 \left[\omega_u^{(\sigma)} (u^*) \right], \text{ dır.} \quad (3.68)$$

B_u matrisi köşegen olmadığında blok yapısı dikkate alınmalıdır. Normalleştirme sabiti $\tau_{u,i}^2$, her k blok için tanımlanan bir $T_{u,k}$ matris ile değiştirilmelidir. Kovaryans matrisi ve konum parametresi tahmin edisicine benzer olarak iki farklı ağırlık fonksiyonu kullanılır: Matrisin boyutu için $\left(\omega_u^{(\tau)} \right)$ ve şekli için $\left(\omega_u^{(\eta)} \right)$. Gösterimi basitleştirmek için üçüncü bir ağırlık fonksiyonu $\omega_u^{(\delta)}$ tanımlanmıştır [6] [27] [59]. $s > 1$ boyutlu blokları için

$$\omega_u^{(\delta)}(d) = \left(d \omega_u^{(\eta)}(d) - (d - s \kappa_u^{(\tau)}) \omega_u^{(\tau)}(d - s \kappa_u^{(\tau)}) \right) / s \quad (3.69)$$

olarak tanımlanmıştır. Burada $\kappa_u^{(\tau)} = E \left[(b - s \kappa_u^{(\tau)}) \omega_u^{(\tau)}(b - s \kappa_u^{(\tau)}) \right] = 0$ ve $u \sim \chi_s^2$ dir.

B_u matrisi köşegen olmadığında sağlam tahmin denklemi aşağıdaki şekilde tanımlanabilir:

$$\sum_{k=1}^K \left[\omega_u^{(\eta)} \left(d \left(T_{u,k}^{-1/2} \tilde{u}_k^* / \hat{\sigma} \right) \right) \tilde{u}_k^{*'} Q_{l,k}(\hat{\theta}) \tilde{u}_k^* / \hat{\sigma}^2 - \omega_u^{(\delta)} \left(d \left(T_{u,k}^{-1/2} \tilde{u}_k^* / \hat{\sigma} \right) \right) \text{iz} \left(T_{u,k} Q_{l,k}(\hat{\theta}) \right) \right] = 0 \quad (3.70)$$

Burada $Q_{l,k}(\hat{\theta})$, $Q_l(\hat{\theta})$ 'nın k bloğuna karşılık gelen $s \times s$ boyutlu alt matrisidir ve $Q_l(\hat{\theta})$, $s \times s$ boyutlu $T_{u,k}$ matrisinin herhangi bir karekökünün tersidir.

Köşegen durumda olduğu gibi, $T_{u,k}$ matrisini her bir toplamda beklenen değeri sıfır olacak şekilde tanımlanır.

$$E \left[\omega_u^{(\eta)} \left(d \left(T_{u,k}^{-1/2} \tilde{u}_k^* / \sigma \right) \right) \tilde{u}_k^{*'} Q_{l,k} \left(\hat{\theta} \right) \tilde{u}_k^* / \sigma^2 - \omega_u^{(\delta)} \left(d \left(T_{u,k}^{-1/2} \tilde{u}_k^* / \sigma \right) \right) i_z \left(T_{u,k} Q_{l,k} \left(\hat{\theta} \right) \right) \right] = 0 \quad (3.71)$$

Yapılandırılmamış kovaryans matrisleri için, $Q_l(\hat{\theta})$ matrisi sıfır-bir matrisine indirgenir ve bu indirgeme bu denklemlerin çözümünü kolaylaştırır.

Simetrik matris $T_{u,k}$ tamamen yapılandırılmamış kovaryans matrisleri için tanımlanmıştır, burada $r = s(s+1)/2$ 'dir. Diğer kovaryans matris yapıları için, $T_{u,k}$, u^*/σ 'nın doğrusal yaklaşımının varyansı ile yer değiştirilir [5].

$$T_{u,k} = \hat{\sigma}^{-2} E \left[\tilde{u}_k^* \tilde{u}_k^{*'} \right] \approx I - LE_0 \left[\psi_u u^{*'} \right] - E_0 \left[u^* \psi_u' \right] L' + LE_0 \left[\psi_u \psi_u' \right] L' + E_0 \left[\psi_e^2(\varepsilon^*) \right] KK' \quad (3.72)$$

3.5.1.3. Tahmin Algoritması

(3.58), (3.63) ve (3.66) tahmin denklemlerinin köklerini eş anlolu bulmak için aşağıdaki dört adım izlenir:

Adım 1. Başlangıç tahminleri hesaplanır.

Adım 2. Verilen $\hat{\theta}$ ve $\hat{\sigma}$ için (3.58)'in çözümü olan $\hat{\beta}$ ve \tilde{u}^* bulunur.

Adım 3. $\hat{\beta}$ ve \tilde{u}^* sabit tutulur, (3.66) eşitliği sağlanacak şekilde $\hat{\sigma}$ bulunur.

Adım 4. $\hat{\theta}$ için (3.63) tahmin eşitliğinin yakınsayıp yakınsamadığı kontrol edilir. Yakınsama olmadığında $\hat{\theta}$ güncellenir ve adım 2'ye dönülür.

Dört adım için algoritmalar birbirinden bağımsız olarak seçilir. Bu algoritma durduğunda bir bloğa karşılık gelen bazı bileşenleri parametre uzayının sınırında değer aldığı blok-köşegen durumu dışında tüm tahmin eşitliklerinin eş anlolu çözümü bulunmuş olur. Bu parametreler için tahmin eşitlikleri güvenilir değildir. Tahmin eşitliklerinin doğru olmayan bu çözümlerinden kaçınmak için her yeni $\hat{\theta}$ için $\hat{\beta}$ ve \tilde{u}^* güncellenir ve bu durumda θ 'nın başlangıç tahminleri yeteri kadar büyük olur. Aksi takdirde, algoritma

yanlış bir bloğa karşılık gelen θ 'nın bileşenlerinin tümünü sıfıra veya sıfıra yakın bulabilir [6].

Adım 1. Doğru çözüm ne olursa olsun, algoritmanın doğru yerel çözüme yakınsamasını garanti eden başlangıç değerlerini sağlamak başlangıç tahmincisine bağlıdır. Sabit etkiler modeli için MM tahminleri durumunda, başlangıçtaki S tahmini, son tahminin istenen yüksek bozulma noktasına sahip olmasını sağlar. Aynı durum, karma etkili modeller için de kesinlikle istenir. Ancak, literatürde böyle bir tahmin edici yoktur.

Yeniden azalan ρ -fonksiyonları, bazı gözlemlere veya rasgele etki seviyelerine sıfır ağırlık atama avantajına sahiptir. Bu, böyle gözlemlerin tahminler üzerinde hiçbir etkisinin olmamasını mümkün kılar. Dışbükey ρ -fonksiyonları kullanıldığında, bir gözlem pratikte her zaman tahminler üzerinde bir etkiye sahiptir, çünkü artık veya rasgele etki seviyesi artı veya eksi sonsuza yakınsadığında limitte sadece sıfır ağırlığına ulaşılır. Aykırı gözlemlerin etkisinin ortadan kaldırılması istenirse ilk olarak, dışbükey bir ρ fonksiyonu kullanılarak kestirimler hesaplanır. Daha sonra sonuçlar, ikinci adımda yeniden azalan ρ fonksiyonu kullanarak kestirim için başlangıç değeri olarak kullanılır. Yeniden azalan ρ -fonksiyonları için iyi başlangıç tahmin edicilerinin olmaması, bu yaklaşımı yeniden azalan ρ -fonksiyonlarının istenilen özelliklerinin en azından bir kısmını elde etmek için kullanılabilir [6].

Adım 2. Sabit ve rasgele etkilerin tahmini:

Verilen θ ve σ için, sabit ve rasgele etkilerin tahmini iteratif olarak yeniden ağırlıklandırılmış en küçük kareler kullanılarak elde edilebilir. Buna göre w_u 'ya benzer biçimde w_e

$$w_e = \text{Diag}(w_e(\varepsilon_i^* / \sigma)), \quad i = 1, 2, \dots, n$$

şeklinde tanımlanır. Burada $w_e(\varepsilon^*) = \begin{cases} \psi_e(\varepsilon^*) / \varepsilon^*, & \varepsilon^* \neq 0 \\ \psi_e'(0), & \varepsilon^* = 0 \end{cases}$ dır.

Bu tanımlara göre (3.60) yeniden düzenlenirse aşağıdaki eşitlikler elde edilir:

$$\begin{bmatrix} X'(B_e^{-1})'w_e B_e^{-1} X & X'(B_e^{-1})'w_e B_e^{-1} ZB_u / \lambda_e \\ B_u'Z'(B_e^{-1})'w_e B_e^{-1} X & B_u'Z'(B_e^{-1})'w_e B_e^{-1} ZB_u + \wedge_u w_u \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{u}^* \end{bmatrix} = \begin{bmatrix} X'(B_e^{-1})'w_e y \\ B_u'Z'(B_e^{-1})'w_e y \end{bmatrix} \quad (3.73)$$

Verilen ağırlıklar dikkate alındığında $\hat{\beta}$ ve \tilde{u}^* hesaplamaları arasında geçiş yaparak verilen tahminler için ağırlıkları güncelleyerek sabit ve rasgele etkilerin hesaplanmasında basit ve etkili bir algoritma elde edilir. Algoritma önceden tanımlanmış ağırlık değerleri ile başlatılır veya tüm ağırlıklara 1 değeri verilerek başlatılır. Tahminlerin göreceli değişikliği yeterince küçük olduğunda algoritma durdurulur.

$\hat{\beta}$ ve \tilde{u}^* için tahminlerin $\hat{\theta}$ 'nin her yeni üyesi için güncellenmesi ve θ için başlangıç tahmininin yeterince büyük olması çok önemlidir. Aksi takdirde algoritma, θ 'nin bir veya daha fazla bileşenini yanlışlıkla sıfıra veya sıfıra yakın ayarlayabilir, bu da her zaman bir çözümdür [6].

Adım 3. Varyans Parametreleri

(3.63) eşitliği,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \omega_e^{(\sigma)} \left(\frac{\hat{\varepsilon}_i^*}{\tau_{e,i} \hat{\sigma}} \right) \hat{\varepsilon}_i^{*2}}{\kappa_e^{(\sigma)} \sum_{i=1}^n \tau_{e,i}^2 \omega_e^{(\sigma)} \left(\frac{\hat{\varepsilon}_i^*}{\tau_{e,i} \hat{\sigma}} \right)} \quad (3.74)$$

olarak yeniden yazılabilir

(3.74)'deki gösterim $\hat{\sigma}$ hesaplanması ve $\hat{\sigma}$ 'ya verilen ağırlıkların güncellenmesi arasında değişen basit iki aşamalı algoritmayı ifade eder. Bu algoritmanın, özellikle tüm algoritma yakınsaksa ve $\hat{\sigma}$, $\hat{\theta}$ 'nin iterasyonları arasında çok az değişiyorsa, hızlı ve güvenilir olduğu kanıtlanmıştır.

Benzer bir yaklaşım, $\tau_{e,i}$ 'nin hesaplanması için de verilebilir. $\tau_{e,i}$ için çözüm (3.64),

$$\tau_{e,i}^2 = \frac{\text{E} \left[\omega_e^{(\sigma)} \left(\frac{\hat{\varepsilon}_i^*}{\tau_{e,i} \hat{\sigma}} \right) \left(\frac{\hat{\varepsilon}_i^*}{\hat{\sigma}} \right)^2 \right]}{\text{E} \left[\kappa_e^{(\sigma)} \omega_e^{(\sigma)} \left(\frac{\hat{\varepsilon}_i^*}{\tau_{e,i} \hat{\sigma}} \right) \right]} \quad (3.75)$$

dir. $\tau_{e,i}$ değerleri, her yeni $\hat{\theta}$ değeri için yeniden hesaplanmalıdır. Tercihen son başlangıç değerleri $\hat{\theta}$ için hesaplanan değerleri başlangıç değerleri olarak kullanarak yeniden hesaplanmalıdır [6].

Adım 4: Kovaryans parametreleri

Sadece bir blok tipi olduğu varsayalım. Bu adımdaki algoritmalar, çoklu blok tiplerine kolayca genelleştirilebilir. Bir iterasyon sonrası her blok tipi için güncellemelerin (hepsini birlikte uygulamadan önce) ayrı ayrı hesaplanmasından oluşur.

$B_u(\theta)$ 'nin köşegen olması durumunda $\hat{\theta}$, Adım 3'deki algoritmanın benzeri kullanılarak hesaplanabilir. Bunun, genel bir tahmin algoritması yaklaşımına göre daha etkin ve sağlam olduğu kanıtlanmıştır. Aynı durum $B_u(\theta)$ 'nin köşegen olmadığına da geçerlidir. Bununla birlikte, eğer özel bir kovaryans yapısı varsa, tek seçenek Newton-Raphson gibi kapsamlı tahmin algoritmasının kullanılmasıdır. Ancak, kovaryans yapısının özel olması değişken olabilir ve çoğu zaman birçok parametreyle ilgili problemler için algoritma yakınsamaz.

Yapılandırılmamış kovaryans matrislerinde, daha iyi EM türü algoritma vardır. $L(A)$ fonksiyonu A'nın alt üçgen Cholesky faktörüdür ve L^{-1} 'in de faktör tersidir. Daha sonra, yapılandırılmamış kovaryans matrisleri için ve B_u nin $B_{u,1}$ ilk blok olması açısından, güncelleme

$$B_{u,1}(\hat{\theta}^{[it]}) = B_{u,1}(\hat{\theta}^{[it-1]}) \frac{1}{\sigma} L \left(\sum_{k=1}^K \hat{\omega}_{u,k}^{(\eta)} \hat{u}_k^* (\hat{u}_k^*)' \right) L^{-1} \left(\sum_{k=1}^K \hat{\omega}_{u,k}^{(\delta)} \Gamma_{u,k} \right) \quad (3.76)$$

dir. Burada köşeli parantez içindeki üst simge iterasyonunu belirtir. Sağ taraf, son iterasyondan gelen değer olan $\hat{\theta}^{[it-1]}$ kullanılarak hesaplanır ve $\hat{\omega}_{u,k}^{(\cdot)}$ k . sağlamlık ağırlığına karşılık gelir.[6]

4. UYGULAMA

Bu bölümde, ilk olarak küçük bir veri setinde aykırı değerlerin bulunması durumunda doğrusal karma modelde, REML ve sağlam yöntem ile varyans bileşenlerinin tahmini elde edilmiştir. Burada küçük örneklerde aykırı değer olması durumunda sağlam tahmin yönteminin REML yöntemine göre daha iyi sonuç verdiğini göstermek amaçlanmıştır. Daha sonra aykırı değerler içeren gerçek bir veri seti üzerinde REML ve sağlam yöntem ile doğrusal karma modelde varyans bileşenleri tahmin edilmiştir.

4.1. Küçük Veri Seti Uygulaması

Bu bölümde literatürde yer alan ve aykırı değerler içeren aerosol veri setinde varyans bileşenleri sağlam yöntem ile tahmin edilmiştir. Aerosol veri setinin kullanılmasının temel amacı küçük örnekte aykırı değerler olduğunda daha iyi sonuç veren sağlam karma model tahmininin etkinliğini göstermektir. Aerosol veri seti iki farklı filtre imalatçısı tarafından üretilen iki aerosol türünden elde edilen penetrasyon yüzde değerlerini (polen, toz vb çekim gücü) içermektedir. Her bir filtre imalatçısı, iki aerosol türünün penetrasyon yüzdesini ölçerken üç farklı filtre türünden yararlanmıştır [1] [60] [13]. Aerosol veri seti Çizelge 4.1’de verilmiştir:

Çizelge 4.1. Aerosol veri seti

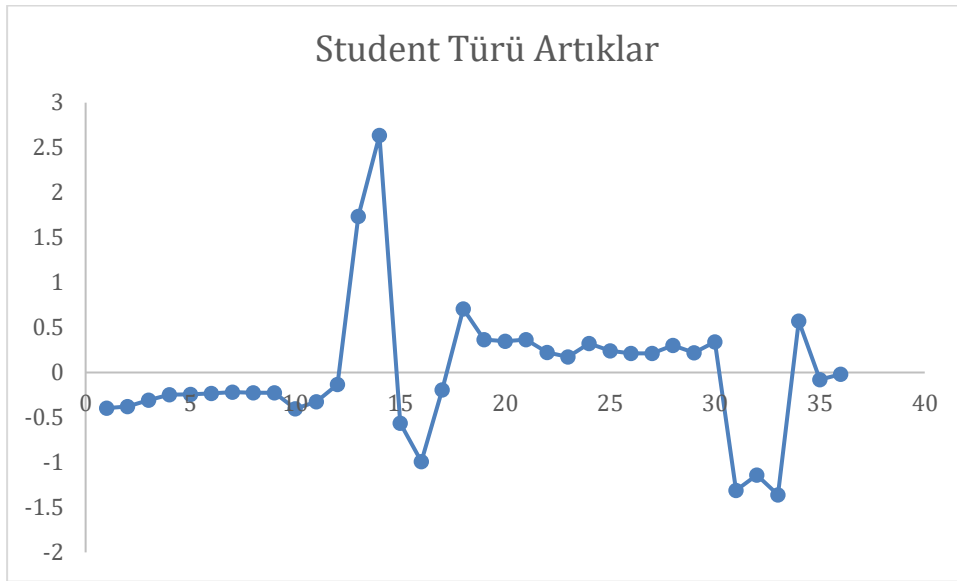
Aerosol 1		Aerosol 2	
İmalatçı 1	İmalatçı 2	İmalatçı 1	İmalatçı 2
Filtre 1	Filtre 4	Filtre 1	Filtre 4
0.750	0.600	1.120	0.910
0.770	0.680	1.100	0.830
0.840	0.870	1.120	0.950
Filtre 2	Filtre 5	Filtre 2	Filtre 5
0.082	4.100	0.160	0.660
0.085	5.000	0.110	0.830
0.096	1.800	0.260	0.610
Filtre 3	Filtre 6	Filtre 3	Filtre 6
0.082	1.000	0.150	2.170
0.076	1.800	0.120	1.520
0.077	2.700	0.120	1.580

Filtrelerin penetrasyon yüzdesi (toz, polen vb çekim güçleri) üzerinde aerosol türünün, filtre imalatçısının ve filtre türlerinin etkili olduğu düşünülmektedir. Bu veri setinde aerosol türü ve imalatçı faktörleri sabit, filtre türü faktörü rasgeledir. Bu nedenle filtrelerin penetrasyon yüzdesi tahmininde hem sabit etkilerin hem de rasgeleri etkilerin bir arada bulunduğu doğrusal karma model kullanılmıştır. Filtre türü faktörü imalatçı faktörünün içinde yer aldığı için iç içe (nested) model kullanılmıştır. Aerosol veri seti için doğrusal karma model aşağıdaki gibi ifade edilmiştir:

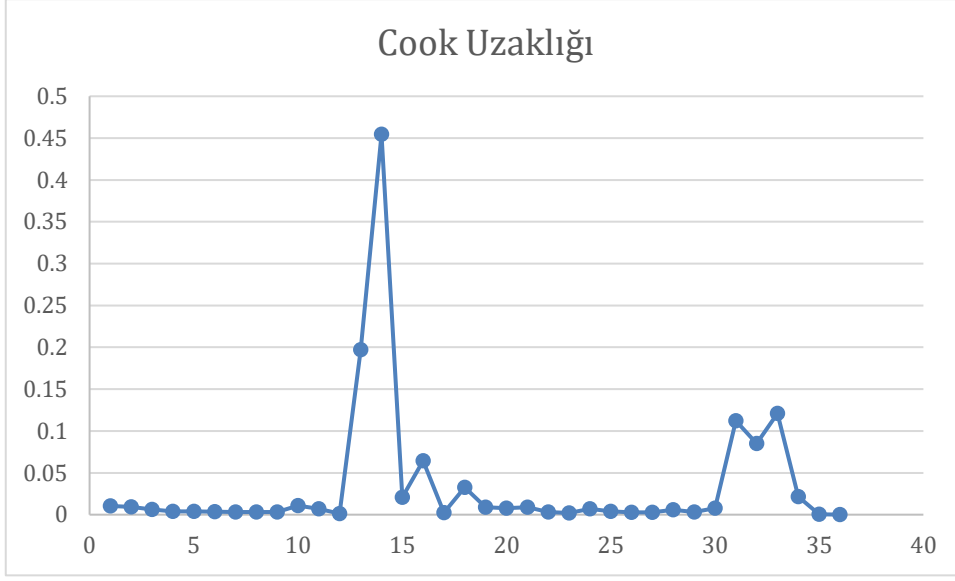
$$y_{ijk} = \mu + A_i + I_j + F_{k(j)} + e_{ijk}, \quad i = 1, 2; \quad j = 1, 2; \quad k = 1, 2, 3, 4, 5, 6$$

Verideki aykırı değerleri ortaya çıkarmak için aykırı değer ölçütlerine ilişkin saçılım grafikleri Şekil 4.1'deki gibi elde edilmiştir:

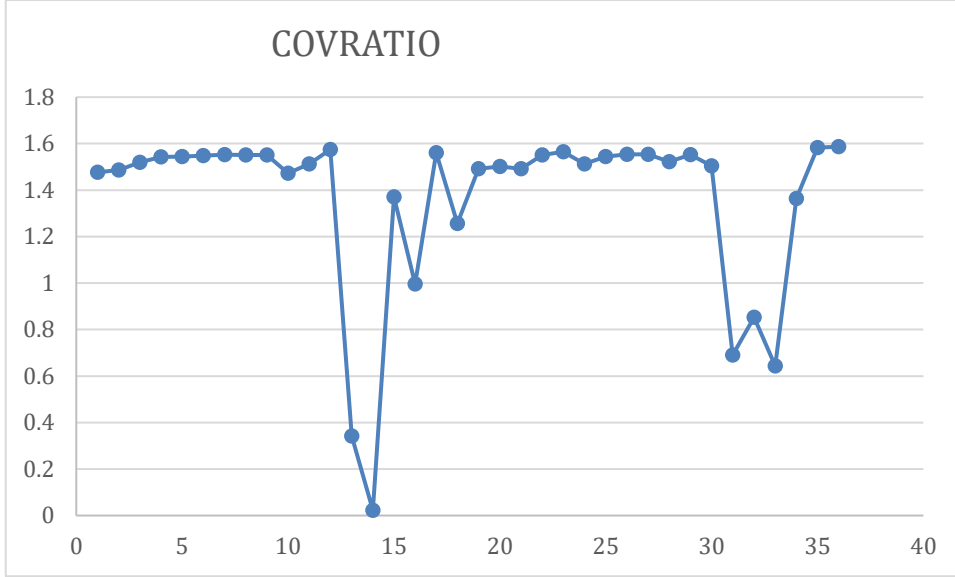
(a)



(b)



(c)



Şekil 4.1. Aeresol veri seti için aykırı değer belirleme ölçütlerine ilişkin saçılım grafikleri a) Student türü artıkların saçılım grafiği, b) Cook Uzaklığı saçılım grafiği, c) COVRATIO değerleri saçılım grafiği

Şekil 4.1 incelendiğinde aeresol verisinde aykırı değer belirleme ölçütlerine göre literatürde de belirtildiği gibi 13. ve 14. gözlem değerlerinin saçılım grafikleri incelendiğinde bu gözlemlerin student türü artık, Cook Uzaklığı ve COVRATIO ölçüt değerlerine göre, diğer gözlemlerin uzağında yer aldığı görülmektedir. Bu gözlemlerin student türü artık değerlerinin (-2,2) aralığı dışında değer aldığı görülmektedir. Bu

gözlemler y yönüde aykırı değerlerdir. Cook Uzaklığı ölçütüne ilişkin saçılım grafiği incelendiğinde 13. ve 14. gözlem değerlerinin diğer gözlemlere göre Cook uzaklığı değerinin çok farklı olduğu görülmektedir. Bu gözlemler aynı zamanda parametre kestirimleri üzerinde etkili gözlemlerdir. Bu gözlemlerin COVRATIO değerlerinin 0 değerine çok yakın olduğu görülmektedir. COVRATIO değeri 1'den küçük olan gözlemler parametre tahminlerinin doğruluğunu azaltmaktadır. Bu nedenle bu gözlemler parametre kestirimlerinin doğruluğunu önemli ölçüde azaltmaktadır. Sonuç olarak Şekil 4.1. incelendiğinde veride aykırı değerlerin olduğu görülmektedir.

Küçük örneklerde veride aykırı değerler olduğunda bu değerleri çıkarmak yerine, bu değerlerin parametre kestirimleri üzerindeki etkisini azaltmak için sağlam tahmin yönteminin kullanılması önerilmektedir. Aeresol veri setinde 13. ve 14. gözlemler aykırı değerler olduğu için bu gözlemlerin etkisini azaltmak için varyans bileşenlerinin tahmininde sağlam yöntem kullanılmıştır. Sağlam tahmin yönteminin küçük örneklerde aykırı değerlerin varlığında REML yöntemine göre daha doğru sonuçlar verdiğini göstermek için aeresol verisi hem REML ve hem de sağlam yöntem ile tahmin edilmiştir. Tahmin sonuçları Çizelge 4.2.'de verilmiştir:

Çizelge 4.2. Doğrusal Karma Model ve Sağlam Doğrusal Karma Model Sonuçları

	Doğrusal Karma Model Tahminleri			Sağlam Doğrusal Karma Model Tahminleri		
	Tahmin	Standart Hata	t değeri	Tahmin	Standart Hata	t değeri
	Sabit Etkiler			Sabit Etkiler		
Sabit	-0,2079	0,8762	-0,237	-0,4395	0,7555	-0,582
Aeresol türü	-0,3938	0,2697	-1,460	-0,0352	0,1013	-0,348
İmalatçı	1,1940	0,4916	2,429	0,8878	0,4681	1,897
	Rasgele Etkiler			Rasgele Etkiler		
	Varyans	Standart Hata	-	Varyans	Standart Hata	-
Sabit(Filtre)	0,2534	0,5033	-	0,2978	0,5457	-
Artık	0,6547	0,8091	-	0,0878	0,2964	-

Çizelge 4.2 incelendiğinde sabit etkilere ilişkin doğrusal karma model tahminleri ve sağlam doğrusal karma model tahminlerinin farklı olduğu görülmektedir. Bunun nedeni veride aykırı değerlerin olması ve bu aykırı değerlerin küçük örneklerde tahmin

sonuçları üzerinde önemli derecede etkili olmasıdır. REML ve sağlam yöntem ile elde edilen varyans tahminleri incelendiğinde sağlam tahmin ile elde edilen artık varyansının (0,0878) REML ile tahmin edilen artık varyansına (0,6547) göre oldukça küçük olduğu görülmektedir. Bu da sağlam yöntem ile tahmin edilen modelin hatasının oldukça küçük olduğunu ve tahmin sonuçlarının daha güvenilir olduğunu göstermektedir. Gözlem değerlerine ilişkin sağlam ağırlık değerleri Çizelge 4.3’de verilmiştir:

Çizelge 4.3. Gözlem değerlerinin sağlam ağırlık değerleri

Gözlem No	Ağırlık Değeri	Gözlem No	Ağırlık Değeri
1	1	19	1
2	1	20	1
3	1	21	1
4	1	22	1
5	1	23	1
6	1	24	1
7	1	25	1
8	1	26	1
9	1	27	1
10	1	28	1
11	1	29	1
12	1	30	1
13	0.143	31	0.646
14	0.108	32	0.883
15	0.812	33	0.598
16	0.535	34	0.858
17	1	35	1
18	0.417	36	1

Çizelge 4.3 incelendiğinde aykırı değer olarak belirlenen gözlemlere ilişki ağırlık değerlerin en küçük olduğu görülmektedir. Sağlam tahminde aykırı değerlerin etkisini azaltmak için aykırı gözlem ağırlıkları küçük değerler almaktadır.

4.2. Gerçek Veri Seti Uygulaması

Bu çalışmada Orman Genel Müdürlüğü Orman Ağaçları ve Tohumları Islah Araştırma Enstitüsü Müdürlüğü tarafından yürütülen Türkiye Milli Ağaç Islahı ve Tohum Üretimi Programı sonucunda elde edilen veriler incelenmiştir. Bu verilerin tez çalışması kapsamında incelenmesinin iki nedeni vardır. Birincisi, Orman Genel Müdürlüğü’nün,

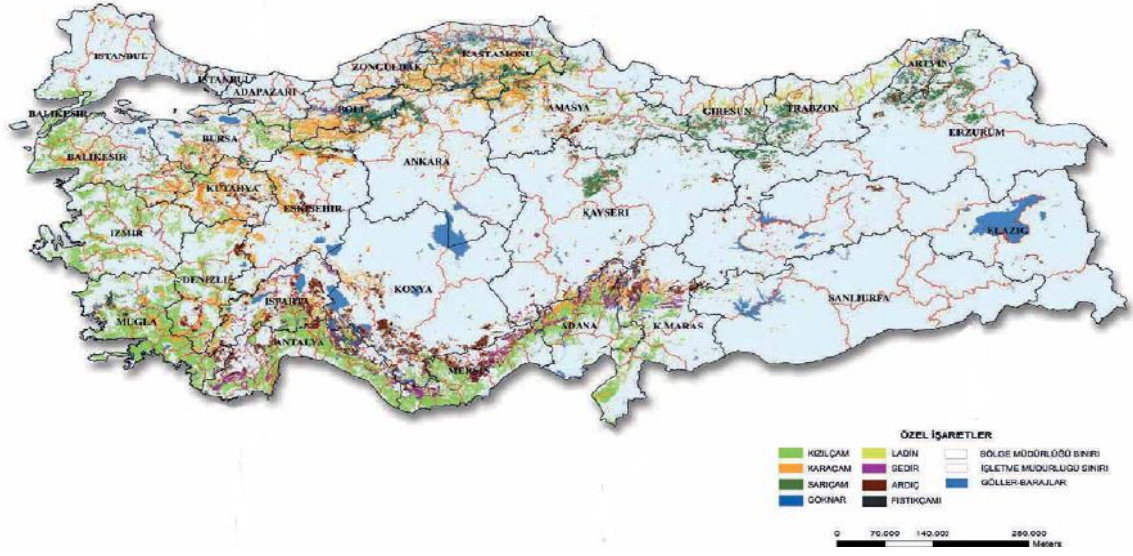
verilerin analizinde doğrusal karma modeli kullanmasıdır. İkincisi ise verilerin aykırı değerler içermesidir. Bu kapsamda aykırı değerlerin varlığında varyans bileşenlerinin tez çalışmasında anlatılan sağlam doğrusal karma model ile tahmin edilmesi için uygun bir veri setidir.

Orman Genel Müdürlüğü Orman Ağaçları ve Tohum Islah Araştırma Enstitüsü Müdürlüğü tarafından yürütülen çalışmada kızılçam seçilmesinin nedenleri aşağıdaki gibi sıralanabilir:

- Genetik ıslah çalışmalarında türün yayılış alanının geniş olması önemlidir çünkü bir türün orman ürün ve hizmetlerine olan ihtiyacın türün yayılış alanı ile doğru orantılıdır[7] . Kızılçamın, Türkiye Orman Atlası verilerine göre 5.8 milyon ha. ile ülkemizde en geniş yayılış alanına sahip ibrelili ağaç türü olduğu Çizelge 4.4 ve Şekil 4.2’de görülmektedir.

KOD	AĞAÇ TÜRÜ	KORU			BALTALIK			GENEL ORMANLIK		
		Normal Ha.	Bozuk Ha.	TOPLAM Ha	Normal Ha.	Bozuk Ha.	TOPLAM Ha	Normal Ha.	Bozuk Ha.	TOPLAM Ha
01	Kızılçam	3.202.943,4	2.324.860,9	5.527.804,3	4.970,5	321.898,0	326.868,5	3.207.913,9	2.646.758,9	5.854.672,8
02	Karaçam	2.564.719,5	1.724.059,1	4.288.778,6	15.473,5	388.807,5	404.281,0	2.580.193,0	2.112.866,6	4.693.059,6
03	Sarıçam	738.494,7	553.386,3	1.291.881,0	12.565,0	175.201,6	187.766,6	751.059,7	728.587,9	1.479.647,6
04	Göknar	406.497,6	229.014,0	635.511,6	491,5	34.386,5	34.878,0	406.989,1	263.400,5	670.389,6
05	Ladin	228.785,9	95.666,5	324.452,4	1.426,0	8.594,0	10.020,0	230.211,9	104.260,5	334.472,4
06	Sedir	220.328,4	237.133,5	457.461,9	0,0	6.059,1	6.059,1	220.328,4	243.192,6	463.521,0
07	Ardıç	89.473,8	457.873,1	547.346,9	1.760,5	26.207,9	27.968,4	91.234,3	484.081,0	575.315,3
08	Fistikçamı	60.888,5	26.234,0	87.122,5	0,0	1.905,0	1.905,0	60.888,5	28.139,0	89.027,5
09	Servi	419,8	1.050,0	1.469,8	0,0	0,0	0,0	419,8	1.050,0	1.469,8
11	Halepçamı	465,0	250,0	715,0	0,0	0,0	0,0	465,0	250,0	715,0
12	Sahilçamı	54.461,2	5.367,9	59.829,1	0,0	3.839,0	3.839,0	54.461,2	9.206,9	63.668,1
13	P.Radiata	57,8	0,0	57,8	0,0	0,0	0,0	57,8	0,0	57,8
14	Duglaz	84,2	3,2	87,4	0,0	0,0	0,0	84,2	3,2	87,4
20	Diğer ibrelili	57,3	0,0	57,3	0,0	0,0	0,0	57,3	0,0	57,3

Çizelge 4.4. İbrelili Ağaç Türlerinin Alansal Dağılımı [61]



Şekil 4.2. İbrelili Ağaç Türlerinin Yayılış Alanları Haritası [61]

- Genetik ıslah çalışmalarına konu olacak türlerin seçiminde bir diğer önemli etken ormanları gençleştirmek için kullanılan tohum ve fidana olan ihtiyaçtır. Ormanlık çalışmalarında gençleştirme materyaline ihtiyaç ağaç türünün temeli gençleştirme şekline bağlıdır. Yaygın olarak yapay yolla gençleştirilen türler ile plantasyon ormancılığına konu olan türler için gençleştirme materyali ihtiyacı fazladır. Kızılçam Türkiye’de ağaçlandırmalarda en yaygın olarak kullanılan türdür. Ayrıca kızılçam doğal gençleştirme çalışmalarında bile çoğu zaman tohum takviyesi yapılmaktadır. Kızılçamın ağaçlandırma potansiyelinin yüksek olması, genetik ıslah çalışmalarının daha düşük maliyetle gerçekleşmesi açısından da önemli bir özelliktir [9].
- Orman ağaçları için çiçeklenme yaşı ıslah çalışmalarının etkinliği açısından önemlidir. Eğer türün olgunlaşma yaşı uzarsa ıslah programının süresi uzar ve birim sürede elde edilecek genetik kazanç azalır. Kızılçam Türkiye’deki doğal çam türleri arasında en erken çiçeklenen türdür, 2. yaştan itibaren çiçeklenmektedir.[9]
- Islah çalışmalarında başarılı olabilmenin ön şartlarından biri de türdeki genetik çeşitliliktir. Tüm yukarıda açıklanan özellikler uygun olsa bile, eğer türde genetik çeşitlilik az ise, genetik ıslah çalışmaları ile sağlanacak kazanç (genetik kazanç)

sınırlı olacaktır. Kızılçamda gerek populasyonlar arası, gerekse populasyon içi genetik çeşitlilik, genetik ıslah çalışmalarıyla önemli miktarda genetik kazanç elde edilebileceğini göstermektedir [9].

- Kızılçam odununun inşaat malzemesi, ambalaj, maden direği, selüloz ve kağıt endüstrisi gibi geniş bir kullanım alanı vardır[7]. Ayrıca doğal çamlarımız olan kızılçamlardan, reçine, terebentin yağı, çam yağı, odun katranı ve kalofan gibi yan ürünlerde üretilmektedir.

Orman Genel Müdürlüğü Orman Ağaçları ve Tohum Islah Araştırma Enstitüsü Müdürlüğü tarafından yürütülen çalışmada, Akdeniz Bölgesi Alçak Yükselti Kuşağı (0-400m) Islah Zonunda kurulan Çizelge 4.4' te verilen altı adet klonal tohum bahçesinde yer alan 168 ailenin açık tozlaşma ürünü tohumlarla kurulan Antalya Bölgesi döl denemesinin sekizinci yaş verileri incelenmiştir. Bu veriler kullanılarak genetik parametreler, ailelerin ıslah değeri ve genetik kazanç değerleri doğrusal karma model ve sağlam doğrusal karma model ile tahmin edilerek sonuçlar elde edilmiş ve yorumlanmıştır.

Yapılan çalışmadan Akdeniz Bölgesi Alçak Yükselti Kuşağı (0-400m) Islah Zonunda kurulan Çizelge 4.4' te verilen altı adet klonal tohum bahçesinde yer alan 168 ailenin açık tozlaşma ürünü tohumlarla kurulan Antalya Bölgesi döl denemesinin sekizinci yaş verileri ele alınarak incelenmiştir. Bu veriler kullanılarak genetik parametreler, ailelerin ıslah değeri ve genetik kazanç değerleri doğrusal karma model ve sağlam doğrusal karma model ile tahmin edilerek sonuçlar elde edilmiş ve yorumlanmıştır.

Orman Ağaçları ve Tohumları Islah Araştırma Müdürlüğü tarafından yürütülen bu çalışmada tüm tohum bahçelerinde kozalaklar toplama işlemi 1996 yılı Nisan ayında, tohum meşcerelerinde ise 1997 yılı Nisan ayında yapılmıştır [9].

Tohum Meşcereleri, bulunduğu ekosistemde ekonomik önemi olan odun özellikleri bakımından üstün özelliklere sahip ağaçlardan oluşması nedeniyle tohum kaynağı olarak seçilen ve özel bir silvikültür planıyla işletilen meşcerelerdir. [62]

Tohum bahçeleri, genetik bakımdan daha yüksek nitelikli tohum elde etmek üzere, bir anlamda damızlık olarak seçilen üstün ağaçlardan alınan aşı kalemleriyle aşılanan fidanlardan meydana gelen bir çeşit meyve bahçesidir [63].

Klon, bir ağaçtan çelik, aşı kalemi gibi genetik materyallerle üretilen aynı genotipe sahip fidanlar topluluğuna "klon" denmektedir. Böylece, istenen genlere ve dolayısıyla morfolojik yapıya sahip o bitkinin özelliklerinin bozulmadan, çelikle veya aşılama yapılarak sürdürülmesi mümkün olmaktadır [64].

Doğadaki populasyonlardan genotiplerine göre seçilen ve her biri bir klonu temsil eden bireylerden oluşan tohum bahçelerine "Klonal Tohum Bahçeleri" denir [63].



Şekil 4.3. Tohum meşçeresinde plus ağaç seçimi ve işaretlenmesi

Şekil 4.3 tohum meşçeresinde plus ağaç seçimini göstermektedir. Plus ağaç, diğer bireylere göre daha üstün ve daha iyi görünüm bireylerdir. Bu yöntem ile ıslaha yön verecek olan ağaçlar işaretlenerek Şekil 4.4'te meşçerelerden alınan klonların tohum bahçelerine aktarılmış şekli gösterilmektedir.



Şekil 4.4. Tohum Bahçesi Görüntüsü

Antayla Bölgesi deneme alanı kurulurken alınan klonların tohum bahçeleri ve özellikleri Çizelge 4.5’ te yer almaktadır.

Çizelge 4.5. Altı Adet Klonal Tohum Bahçesi Dağılımı [9]

Tohum Bahçesinin Özellikleri	Tohum Bahçesi Numarası					
	3	4	5	7	11	16
Tesis Yeri	Kadirli	Antakya	Antalya	Kadirli	Antalya	Antalya
Tesis Yılı	1977	1977	1978	1978	1982	1984
Klon Sayısı	24	35	29	30	25	25
Ramet Sayısı	857	1032	927	516	1200	2710
Alanı(ha)	7	2.6	7.5	4.2	4.3	18.6
Dikim Aralığı	9x9	5x5	9x9	9x9	6x6	8x8
Dağıtım Şekli	Rasgele	Sistemantik	Rasgele	Rasgele	Rasgele	Rasgele

Ele alınan Antalya bölgesi deneme alanı ile ilgili başlıca özellikler Çizelge 4.6’de yer almaktadır.

Çizelge 4.6. Antalya Deneme Alanı Özellikleri

Deneme Alanı Özellikleri	Antalya 1B Deneme Alanı
Dikim Tarihi	Mart 1998
Arazi Tipi	Orman Alanı
Rakım	350
Meyil	% 1
Enlem	36°58'10"
Boylam	30°22'45"
Ortalama Yağış(mm)	1043
Ortalama Sıcaklık(°C)	18,4
Anakaya	Traverten
Toprak Türü	Kumlu Balçık
Toprak pH	7.7

Seleksiyon, genetik ıslahı sağlamak üzere istenilen özelliklere göre populasyonların ve bireylerin seçimine denir. Bu çalışmada genetik ıslah stratejileri belirlenmesinde kullanılan genetik test tipi, selektif ıslah yöntemlerinden biri olan açık tozlaşma döl denemeleridir. Selektif ıslaha populasyondaki amaca uygun fenotiplerin seçimi ile başlanır, açık tozlaşma döl denemeleri ile çok miktarda ebeveynlerden tohum toplanır ve alındıkları bireylere göre karıştırılmadan ayrı ayrı ekilir ve fidan elde edilir. Bu fidanlar karıştırılmadan ayrı ayrı ve tekrarlı olarak deneme plantasyonlarına aktarılır. Her bir üstün (plus) ağaçtan yetiştirilen tüm aşılı fidanlar ve çeliklerin köklendirilmesiyle elde edilen tüm fidanlar genetik bakımdan o ağacın aynı özelliklerini taşımaktadırlar. Bu işlemlerden sonra, birbirini izleyen yıllarda bu fidanların bazı karakterlerine ilişkin ölçümler yapılır. [64]

Orman Genel Müdürlüğü tarafından yürütülen bu çalışmada ayrı bir işlem birimi olarak kontrol grubu bulunmaktadır. Kontrol grubu genetik kazanç hesabıyla ıslah edilen

ağaçlardan ne kadar genetik kazanç sağladığını göstermek için kullanılmaktadır. Antalya bölgesi deneme alanı için F testi ile ailelerin anlamlı farklılıklar gösterip göstermediğini tespit etmek için varyans analizi (ANOVA) yapılmıştır. Aileler arasında anlamlı farklılık olduğu tespit edilmiştir. İslah çalışmalarında aileler arasında fark olması çalışmanın temeli için önemlidir aksi takdir incelenmesi gereken bir durum kalmamaktadır. Çünkü amaç aileler arasındaki farkı ortaya çıkararak iyi olan aileleri bulmaktır.

Tüm deneme alanlarında rasgele blok düzeni kullanılmıştır. Rasgele blok düzeninde sahada mevcut çevresel değişim, daha küçük homojen bloklarda kontrol edilmektedir. Ancak orman ağaçları büyük hacimlidirler. Bu nedenle test edilecek aile sayısı, blok adedi ve her bir parseldeki ağaç sayısına bağlı olarak bir blok için gerekli olan alan çok büyük olabilmektedir. Bu durumda çevresel varyans artmakta, genetik varyansın bir kısmının çevresel varyans tarafından örtülmesine ve sonuçta yapılacak seleksiyonun etkinliğinin azalmasına yol açmaktadır. Bu problemin ortadan kaldırılması için önerilerden bir tanesi, alt bloklama yapılmasıdır. Çalışmada döl denemelerinde Schutz ve Cockerham [66] tarafından B (set in rep) ve R (rep in set) deseni olarak adlandırılan iki tip alt bloklama yapılmıştır. Bu düzenleme ile tüm deneme materyali belirli ailelerden oluşan sabit setlere bölünmüştür. Her bir set içine kontrol materyallerinin tümü dahil edilmiştir. Ancak alınan Antalya Bölgesi için analizlere sadece Antalya Bölgesinin kontrol grupları dahil edilmiştir. B tipi alt bloklama tercih edilmiştir. Her bir set içinde yer alan aileler birbirinden farklı olduğu için setlerin üniform olmadıkları düşünülmüş ve set etkisinin anlamlı farklılıklar gösterip göstermediği test edilmiştir. Test sonucunda her bir deneme alanında setler ($p < 0,05$ seviyesinde) anlamlı farklılıklar gösterdiği için set etkisi modellere bir faktör olarak dahil edilmiştir. Burada deneme deseni 25 bloktan oluşmaktadır, her bir blok içinde 4 set bulunmaktadır her set içinde yer alan ailelerin boy uzunlukları kaydedilmiştir [9]. Burada kontrol grubu meşcereden alınan ağaçları ifade etmektedir.

Orman ağaçları genetik ıslahında; blok, deneme alanı, yıl etkileri varyans bileşenlerinin tahmininde rasgele etki olarak alınır ancak bu değişkenler kalıtım derecesinin hesabında analizlerden hariç tutulur. Örneğin; bir aile beş tekrarlı bir denemede yetiştirme ortamı bakımından en uygun iki blokta yer alsın ve diğer bloklarda gözlemlenmesin. Bu ailenin boy ortalamaları, blokların hepsinde bulunan ailelerin birçoğunun boy ortalamasından

daha yüksek olabilir. Bu durumda bu ailenin tüm bloklarda yer alan ancak ortalaması düşük ailelerden daha iyi olduğu fikri yanıltıcıdır. Bu olumsuzluk seleksiyon aşamasında orijinal verilerin sabit etkilerden arındırılması ile giderilebilir. Bu nedenle blok, set ve deneme alanı etkileri genetik parametrelerin tahmininde rasgele etki, ıslah değerlerinin tahmininde ise sabit etki olarak alınmıştır. [9].

Burada kovaryansın sıfır kabul edilmesinin sebebi Şekil4.5'ta gösterilen aşılı fidanların ekimden önce boylarının eşit kabul edilmesi anlamına gelmektedir.



Şekil 4.5. Aşılı Fidanlar

Birçok ıslah çalışması aile seçimiyle ilgilenmektedir. Aile seçimi çeşitli nedenlerden dolayı tercih edilir. Canlı bireylerde ölçülemeyen özellikler (çiftlik hayvanlarında et kalitesi gibi), seçim adaylarının döllerinde veya kardeşlerinde ve akrabaların ortalama fenotiplerine dayalı olarak yapılan seçimlerde ölçülebilir. İlgilenilen birçok özellik, küçük etkileri olan birçok gen tarafından kontrol edilir ve düşük veya orta kalıtım derecesi gösterir. Bir bireyin fenotipi, büyük çevresel etkilerden dolayı bir ıslah ebeveyni olarak genetik değerinin zayıf bir kestiricisidir, bu nedenle tek bir birey üzerinde fenotipik seçim

(toplu seçim veya aile seçimi ile sınırlı) etkili değildir. Soy içi üremenin test edilmesi bu soruna bir çözümdür, ancak soy içi üremenin mümkün olmadığı türler için aile ortalama fenotipine dayalı seçim alternatif bir çözümdür. Spesifik bir çaprazlama dölünün seri üretimi de, uygun spesifik birleştirme kabiliyetinden (eklemeli olmayan genetik etkiler) yararlanmak için bir amaç olabilir. Örneğin, orman ağaçlarında, tohum bahçesi kurucuları tohum bahçeleri kurmak için aile seçimini tercih ederler. Buna bazen geriye doğru seçim (önceki neslin bireylerinin seçimi) veya ileriye doğru seçim ile aile arasında seçim denir. İslah çalışmalarında, ailelerin veya ebeveynlerin ıslah değerlerini daha doğru hesaplamak için birçok dölün fenotipik verilerini kullanırlar [67].

İslah değerleri, doğrusal karma modelde rasgele etkilerin BLUP tahminleridir. İslah değerlerinin tahmininde BLUP yönteminin kullanılmasının bazı avantajları vardır. İslah değerleri hesaplanırken sabit etkiler de dikkate alınır. İslah değerlerini hesaplarırken tüm akrabalardan gelen bilgileri kullanmak, ölçülen bir fenotipe sahip olmayan bireyler için bile ıslah değerlerinin tahmin edilmesini sağlamak için etkili bir yöntemdir. BLUP verilerin çok dengesiz olduğu durumda da kullanılabilir. BLUP ile yapılan ıslah değeri tahminleri sabit ve rastlantısal etkileri en sağlıklı şekilde (unbiased) tahmin etmekte; rastlantısal aile etkilerinin sıralamasının gerçek değere yakın olma olasılığını ve tahmin edilen ıslah değeri ile gerçek ıslah değeri arasındaki korelasyonu maksimize etmektedir. İslah değerlerinin BLUP ile tahmini genetik kazancın hesaplanmasını basitleştirir.

BLUP yöntemi ile her bir aile için tahmin edilen ıslah değeri aynı zamanda ölçülen özelliğin birimi cinsinden genetik kazancı ifade eder. Ancak bu kazanç gerçekleşen kazanç değildir. Gerçekleşen genetik kazanç, ıslah edilmiş materyal ile ıslah edilmemiş materyal karşılaştırılarak bulunur. Bu karşılaştırmayı yapmak için kontrol materyali de ayrı bir aile olarak düşünülür ve ıslah değerleri tahmin edilir. BLUP değerleri ile tahmin edilen ıslah değerlerinin ortalaması sıfır olduğundan, genetik kazancın kontrol materyaline oranı hesaplanamamaktadır. Bunu sağlamak için her bir ailenin tahmin edilen ıslah değerine, denemenin genel ortalama değeri eklenerek mutlak ıslah değeri (MBV) hesaplanmıştır. Genetik kazanç

$$GK = \frac{(\overline{BV}_f - BV_K)}{MBV_K} \cdot 100 \quad (4.1)$$

formülü ile hesaplanır. \overline{BV}_f döl denemesinde test edilen bir bahçedeki klonların ıslah değeri ortalamasını, BV_K : Kontrol materyali için hesaplanan ıslah değerini, MBV_K : Kontrol materyali için hesaplanan mutlak ıslah değerini ifade etmektedir [9].

Kalıtım derecesi, bir populasyonda ailelerin belirli bir özelliği yavrulara aktarma oranıdır [68], bir başka deyişle fenotipik farklılıklara atfedilebilen varyasyon oranıdır. Kalıtım derecesi, yapay seçim yöntemleri ile ne kadar genetik kazanç sağlanabileceğini gösteren önemli bir parametredir. Bu sayede ıslahçı, hangi canlı karakterlerinin yapay seleksiyon ile ıslah edilebileceğini, seleksiyon sonucu elde edilebilecek genetik kazanç miktarını belirler ve ıslah stratejisini buna göre oluşturur [69]. Bu nedenlerle üzerinde çalışılan bitki veya hayvan populasyonunda ekonomik önemi olan karakterlerin kalıtım derecelerinin güvenilir bir şekilde tahmini büyük bir önem kazanmaktadır [7]. Bir canlı karakteri üzerinde yapılan ölçümler için aşağıdaki eşitlik geçerlidir:

$$P = G + E \quad (4.2)$$

Burada P, karakterin ölçülen veya gözlenen fenotipik değeri; G, genotipik değer ve E, çevrenin neden olduğu sapmadır [7]. Buna göre bir popülasyonda görülen fenotipik farklılık, genotip ve çevreden kaynaklanan farklılıkların toplamı olarak düşünülür. Fenotipik varyans, genetik ve çevre varyanslarının toplamıdır. Bu toplamda genetik varyansın payını ifade eden kalıtım derecesi (h^2), genetik varyansın fenotipik varyansa bölünmesi ile elde edilir [70].

Populasyonda gözlenen genetik varyans bileşenlerine göre bireysel kalıtım derecesi (h_i^2) ve aile ortalamaları kalıtım derecesi (h_f^2) olmak üzere iki farklı kalıtım derecesi tahmin edilebilmektedir [7]. Bireysel kalıtım derecesi (h_i^2)

$$h_i^2 = \frac{\sigma_f^2}{k(\sigma_{pi}^2)} \quad (4.3)$$

dir. Burada $\sigma_{pi}^2 = \sigma_f^2 + \sigma_e^2$ olup fenotipik varyansı, σ_f^2 aile varyansını, k döl ile ebeveyn arasındaki genetik kovaryansı ifade etmektedir. Yarım kardeş ailelerde k, 1/4 değerine eşittir [9]. Aile ortalamaları kalıtım derecesi (h_f^2)

$$h_f^2 = \frac{\sigma_f^2}{\sigma_{pfam}^2} \quad (4.4)$$

dir. Burada $\sigma_{pfam}^2 = \sigma_f^2 + \sigma_e^2/c_1$ aile ortalamaları fenotipik varyansı ve c_1 aile varyansının katsayısını göstermektedir.

Kalıtım derecesi 0 ile 1 arasında değer alır. Bu oran ne kadar büyükse kalıtım o kadar yüksektir. Ayrıca yaşa, ve uygulanan yöntemlere göre farklılık göstermektedir [7].

Bireysel kalıtım derecesinin standart hatası

$$S.H.(h_i^2) = 4\sqrt{\frac{2(n-1)(1-t)^2 [1+(k-1)t]^2}{(n-s)(s-1)k^2}} \quad (4.5)$$

şeklinde ifade edilir. Burada

$$n = n_1 + n_2 + \dots + n_f$$

$$t, \text{ sınıflar arası korelasyon değeri } t = \frac{\sigma_f^2}{\sigma_{pi}^2};$$

k , ailelerdeki ortalama birey sayısı;

s , aile sayısıdır.

Aile kalıtım derecesinin standart hatası

$$S.H.(h_f^2) = \frac{S.H.(\sigma_f^2)}{\sigma_p^2} \quad (4.6)$$

şeklinde ifade edilir. Burada $S.H.(\sigma_f^2) = \sqrt{\frac{2}{c_f^2} \sum_g \frac{MS_g^2}{df_g + 2}}$ olup c_f^2 aile varyansının

katsayısı, MS_g^2 aile varyansının kareler ortalaması ve df_g aile varyansının serbestlik

derecesi, $\sigma_p^2 = \sigma_f^2 + \frac{\sigma_e^2}{c_1}$ aile ortalamaları fenotipik varyansı olup c_1 beklenen kareler

ortalaması tablosunda aile varyansının katsayısına karşılık gelmektedir.

4.2. Tahmin Sonuçları

Antalya Bölgesi döl denemesi sekizinci yaş boy değerlerine ilişkin modelde aile, blok, set faktörleri bağımsız değişkenler, boy uzunluğu ise bağımlı değişken olarak ele alınmıştır. Alt Bölüm 4.1’de de bahsedildiği gibi ıslah değerlerinin tahmininde blok ve set faktörleri sabit etki, aile ise rasgele etki olarak alınmaktadır. Bu nedenle boy uzunluğunun tahmininde hem sabit etkilerin hem de rasgeleri etkilerin bir arada bulunduğu doğrusal karma model kullanılmıştır. Set faktörü blok faktörünün içinde yer aldığı için iç içe (nested) model kullanılmıştır. Bu çalışmada Antalya Bölgesi Kızılçam ağaçlarının ıslah değerlerinin hesaplanması için Orman Ağaçları Islah Genel Müdürlüğü tarafından kullanılan doğrusal karma model

$$y_{ijk} = \mu + B_i + S_{j(i)} + A_k + e_{ijk} \quad (4.7)$$

şeklinde [9]. Bu modelde; y_{ijk} , i . bloktaki j . sette k . ailenin gözlem değeri; μ , genel ortalama; B_i , i . blok etkisi ($i=1,2,\dots,25$); $S_{j(i)}$, i . bloktaki j . set alanının çevresel etkisi ($j=1,2,\dots,s_i$); A_k , k . ailenin etkisi ($k=1,2,\dots,f$) ve e_{ijk} hata terimidir. Tez çalışmasında kullanılan verilere ilişkin betimleyici istatistikler Çizelge 4.7’de verilmiştir:

Çizelge 4.7. Boy Karakterine İlişkin Betimleyici İstatistikler

Ortalama	494,45
Ortanca	500
Varyans	9916,329
Standart Sapma	99,581
Minimum	85
Maksimum	828
Çarpıklık Katsayısı	-,321
Basıklık Katsayısı	,362
n	3815

Çizelge 4.7 incelendiğinde Antalya bölgesinde 3815 kızılçama ilişkin boy uzunluğu değerlerinin ortalaması 494,45cm, varyansı 9916,329cm², standart sapması 99,581cm,

gözlenen en kısa boy uzunluğu 85cm, gözlenen en uzun boy uzunluğu ise 828cm' dir. Kızılçam boy uzunluğu değerlerinin dağılımının çarpıklık katsayısına göre sola doğru çarpık olduğunu söyleyebiliriz.

İlk olarak (4.1)'deki modelde yer alan faktörlerin istatistiksel olarak anlamlı olup olmadığı test edilmiştir. Faktörlerin etkilerinin anlamlı olup olmadığı test edildiği iç içe karma doğrusal modele ilişkin ANOVA tablosu Çizelge 4.8'de verilmiştir:

Çizelge 4.8. İç İçe Karma Doğrusal Modele İlişkin ANOVA Tablosu

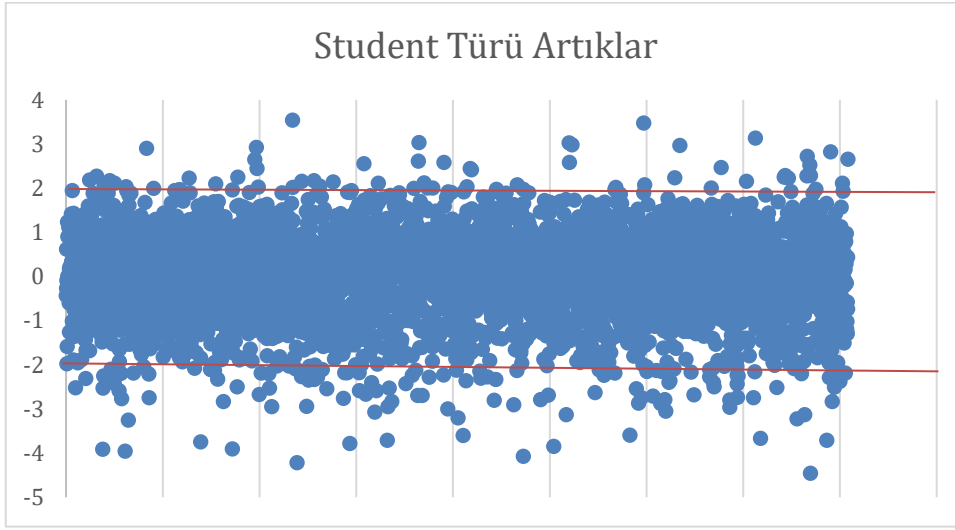
Varyasyon Kaynağı	Serbestlik Derecesi	Kareler Ortalaması	Beklenen Kareler Ortalaması	F Değeri	p değeri
Aile	164	21147.009	$\sigma_e^2 + 21.134(\sigma_a^2)$	2.78	0.000
Set(Blok)	70	44223.395	$\sigma_e^2 + 36.09(\sigma_{s(b)}^2)$	5.81	0.000
Blok	24	110833.657	$\sigma_e^2 + 35.09(\sigma_{s(b)}^2) + 139.59(\sigma_b^2)$	14.57	0.000
Hata	3301	7606.23			
Toplam	3562				

Çizelge 4.8' de aile, set ve blok faktörlerin ağaçların boy uzunluğu üzerinde istatistiksel olarak anlamlı etkileri olup olmadığı araştırılmıştır. İlk olarak “H₀: Aile faktörünün boy uzunluğu üzerinde etkisi yoktur” şeklindeki yokluk hipotezi test edilmiş ve test sonucunda yokluk hipotezi reddedilmiştir (p=0.000<0.05). Buna göre “Aile” faktörünün boy uzunluğu üzerinde istatistiksel olarak anlamlı bir etkisi vardır. Daha sonra “H₀: Blok içinde set faktörünün boy uzunluğu üzerinde etkisi yoktur” şeklindeki yokluk hipotezi test edilmiştir ve test sonucunda yokluk hipotezi reddedilmiştir (p=0.000<0.05). Buna göre blok içinde set faktörünün boy uzunluğu üzerinde istatistiksel olarak anlamlı bir etkisi vardır. Son olarak “H₀: “Blok” faktörünün boy uzunluğu üzerinde etkisi yoktur” şeklindeki yokluk hipotezi test edilmiş ve test sonucunda yokluk hipotezi reddedilmiştir (p=0.000<0.05). Buna göre “Blok” faktörünün boy uzunluğu üzerinde istatistiksel olarak anlamlı bir etkisi vardır.

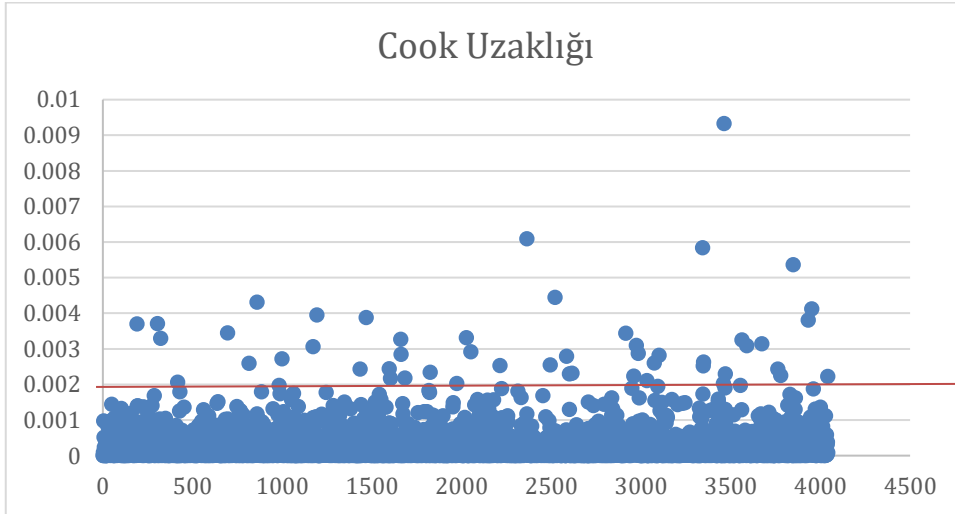
Çizelge 4.8'deki doğrusal karma modelde parametre ve varyans bileşenleri kestirimleri kısıtlanmış en çok olabilirlik (REML) yöntemi ile elde edilmiştir. Antalya döl denemesi sekizinci yaş boy uzunluğu verilerinde aykırı değerler olabilir. Bu nedenle (4.7)'deki

doğrusal karma model için literatürde önerilen aykırı değer ölçütleri kullanılarak incelenen veri setinde aykırı değerlerin olup olmadığı araştırılmıştır. Doğrusal karma model için önerilen aykırı değer belirleme ölçütlerinden Student türü artık değerleri, Cook Uzaklığı ölçüt değerleri ve COVRATIO ölçüt değerlerine ilişkin saçılım grafikleri SAS 9.4 “proc mixed” komutu ve “influence” fonksiyonu kullanılarak Şekil 4.6’ daki gibi elde edilmiştir:

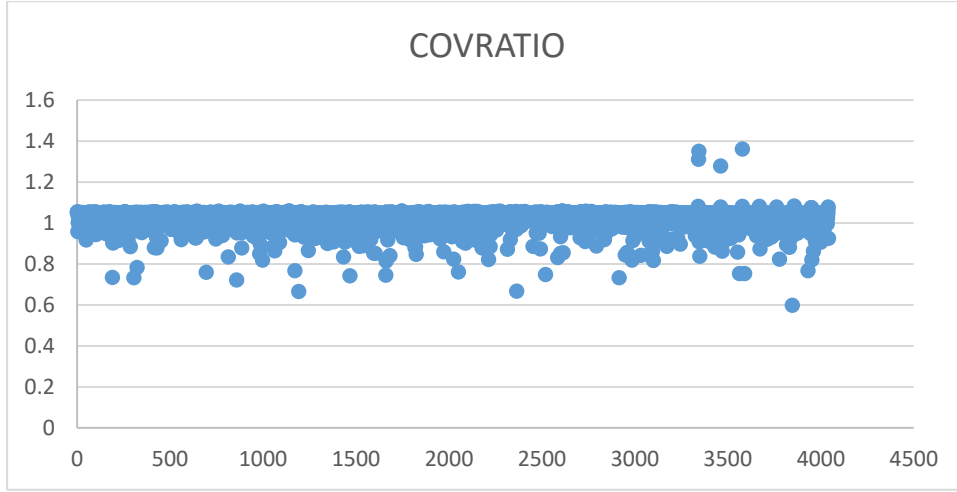
(a)



(b)



(c)



Şekil 4.6. Aykırı değer belirleme ölçütlerine ilişkin saçılım grafikleri a) Student türü artıkların saçılım grafiği, b) Cook Uzaklığı saçılım grafiği, c) COVRATIO değerleri saçılım grafiği

Şekil 4.6 incelendiğinde bazı gözlemlere ilişkin Student türü artık değerlerinin (-2,2) aralığı dışında değer aldığı görülmektedir. Bu gözlemler veride y yönünde aykırı değerlerdir. Cook Uzaklığı ölçütüne ilişkin saçılım grafiği incelendiğinde, gözlemlerin çoğunluğunun Cook Uzaklığı değerinin 0-0.002 aralığında olduğu görülmektedir. Bu değerlerin dışında değer alan gözlemler parametre kestirimleri üzerinde etkili olan gözlemler olarak değerlendirilir. COVRATIO değerlerine ilişkin saçılım grafiği incelendiğinde çoğunluğunun 1 değeri etrafında yoğunlaştığı görülmektedir. Ancak bazı gözlemlerin COVRATIO değerlerinin 1 değerinden küçük olduğu görülmektedir. COVRATIO değeri 1'den küçük olan gözlemler parametre tahminlerinin doğruluğunu azaltmaktadır. Sonuç olarak Şekil 4.2 incelendiğinde veride aykırı değerlerin olduğu söylenebilir.

Veride aykırı değerlerin olması durumunda alternatif tahmin yöntemlerinden birisi sağlam yaklaşımdır. Veride aykırı değerler olduğu için (4.2)'de verilen doğrusal karma modeli, R programında “robustlmm” paketi kullanılarak sağlam yöntem ile tahmin edilmiştir. (4.7)'de verilen doğrusal karma model, REML ve sağlam yöntem ile tahmin edilmiştir ve bu iki tahmin yöntemine ilişkin sonuçlar Çizelge 4.9'da verilmiştir.

Çizelge 4.9. Doğrusal Karma Model ve Sağlam Doğrusal Karma Model Sonuçları

	Doğrusal Karma Model Tahminleri			Sağlam Doğrusal Karma Model Tahminleri		
	Sabit Etkiler			Sabit Etkiler		
	Tahmin	Standart Hata	t değeri	Tahmin	Standart Hata	t değeri
Sabit	493,8414	3,9437	125,223	496,5923	3,8903	127,65
Blok	1,8429	0,3789	4,864	2,0175	0,3742	5,39
Set (Blok)	-0,6904	0,1211	-5,7	-0,7565	0,1182	-6,49
	Rasgele Etkiler			Rasgele Etkiler		
	Varyans	Standart Hata	-	Varyans	Standart Hata	-
Sabit(Aile)	597,6	24,45	-	469,50	21,67	-
Artık	9177,5	95,8	-	8871,50	94,19	-

Çizelge 4.9 incelendiğinde sabit etkilere ilişkin doğrusal karma model tahminleri ve sağlam doğrusal karma model tahminleri birbirine yakındır. Rasgele etkilere ilişkin varyans tahminlerini incelendiğinde artık ve aile varyansının sağlam tahminlerinin daha küçük olduğu görülmektedir.

Doğrusal karma model REML ve sağlam doğrusal karma model tahminleri kullanılarak Alt Bölüm 4.1’de açıklanan genetik parametre tahminleri Çizelge 4.10’daki gibi elde edilmiştir:

Çizelge 4.10. Genetik Parametrelerin Tahmini

Parametreler	Doğrusal Karma Model	Sağlam Doğrusal Karma Model
σ_f^2	662,1	594,3
σ_b^2	444,2	500,8
σ_s^2	1176,8	1151,4
σ_e^2	7605,5	7007,4
σ_{pi}^2	8267,6	7601,7
σ_{pfam}^2	1021,97	925,9
h_i^2	0,32	0,31
h_f^2	0,6478	0,6418
$h_i^2 \mp$ S.E.	0,32 \mp 0,032	0,32 \mp 0,32
$h_f^2 \mp$ S.E.	0,6478 \mp 0.11	0,6418 \mp 0,12

Çizelge 4.10 incelendiğinde sağlam doğrusal karma model ile tahmin edilen varyans parametreleri daha küçüktür. Her iki yönteme göre hesaplanan bireysel kalıtım derecesi ve aile katılım derecesi birbirine çok yakındır.

Doğrusal karma model ve sağlam doğrusal karma model ile elde edilen ıslah değerleri tahminlerinden en iyi ıslah değerine sahip ilk otuz aileye ilişkin ıslah değerleri ve mutlak ıslah değerleri tahminleri ve bu tahminler kullanılarak hesaplanan genetik kazanç değeri Çizelge 4.11’de verilmiştir.

Çizelge 4.11. Islah Değerleri, Mutlak Islah Değerleri Tahmini ve Genetik Kazanç

Karma Doğrusal Model Islah Değerleri (BV) ve Mutlak Islah Değerleri (MBV) Tahmini				Sağlam Karma Doğrusal Model Islah Değerleri (BV) ve Mutlak Islah Değerleri (MBV) Tahmini			
	AİLE	BV	MBV		AİLE	BV	MBV
1	8537	43.82396	582.0825	1	8455	43.88156	584.957
2	8496	42.84184	580.1183	2	8537	42.93914	583.0722
3	8455	40.85169	576.138	3	8652	36.34625	569.8864
4	8501	39.26765	572.9699	4	8501	35.43569	568.0653
5	8652	37.96557	570.3657	5	8496	34.94676	567.0874
6	8460	34.43541	563.3054	6	8474	30.33026	557.8544
7	8419	33.92343	562.2815	7	8481	29.26805	555.73
8	8481	33.78135	561.9973	8	8421	28.88021	554.9543
9	8476	32.43907	559.3127	9	8419	26.83891	550.8717
10	8437	31.51981	557.4742	10	8482	26.26956	549.733
11	8559	30.36448	555.1636	11	8437	25.31072	547.8153
12	8482	30.35609	555.1468	12	8460	25.10454	547.403
13	8474	29.90536	554.2453	13	8559	23.65418	544.5023
14	8547	28.16832	550.7712	14	8495	22.23899	541.6719
15	8495	26.82108	548.0768	15	8547	21.21119	539.6163
16	8421	26.73564	547.9059	16	8476	20.86132	538.9165
17	8546	25.19532	544.8252	17	8417	20.56311	538.3201
18	8472	24.6912	543.817	18	8472	20.20726	537.6084
19	8466	24.3672	543.169	19	8440	20.16728	537.5285
20	8498	24.02355	542.4817	20	8498	20.08318	537.3603
21	8417	23.91954	542.2737	21	8546	19.69656	536.587
22	8440	22.06213	538.5589	22	8470	19.0467	535.2873
23	8475	21.60702	537.6486	23	8499	17.5521	532.2981
24	8544	20.5698	535.5742	24	8475	17.41361	532.0211
25	8649	20.49484	535.4243	25	8552	16.81314	530.8202
26	8485	18.99013	532.4149	26	8485	16.73385	530.6616
27	8431	18.63673	531.7081	27	8649	16.39478	529.9835

28	8542	18.21082	530.8562	28	8466	16.2702	529.7343
29	8787	18.03909	530.5128	29	8431	15.03492	527.2637
30	8470	17.56116	529.5569	30	8544	14.98566	527.1652
Genetik kazanç: %10				Genetik kazanç: %8,84			

Çizelge 4.11 incelendiğinde ıslah değerlerine göre en yüksek ıslah değerine sahip 30 klonla kurulacak genotipik tohum bahçelerinden elde edilecek genetik kazanç karma doğrusal model sonuçlarına göre %10 iken sağlam doğrusal karma model sonuçlarına göre %8,84 olarak hesaplanmıştır [59]. Burada verilen genetik kazanç değerleri iki yöntemi kıyaslamak için bir bulgu olmamakla beraber gerçekte gerçekleşen genetik kazancı daha güvenilir tahmin edilmesi yapılacak olan çalışmanın sonuç beklentisinin daha güvenilir olmasını sağlamakta kullanılabilir.

5. SONUÇ VE ÖNERİLER

Bu çalışmada doğrusal karma modellerde parametre tahmin yöntemleri, aykırı değer belirleme ölçütleri ve varyans bileşenlerinin sağlam tahminlerinin elde edilmesi incelenmiştir.

Aykırı değerlerin varlığında, verilerin dağılımı normal dağılıma uymadığında ya da örneklem çok küçük olduğunda daha güvenilir sonuçlar (parametre tahminleri ve ilişkili testler ile güven aralıkları) elde edildiği için sağlam tahmin yöntemleri önerilmiştir. Bu tez çalışmasında sağlam tahmin yöntemleri ile doğrusal karma modellerde varyans bileşenleri tahminlerinin nasıl elde edildiği açıklanmıştır.

Bu amaç doğrultusunda öncelikle ıslah programlarında sıkça kullanılan doğrusal karma model tanıtılarak, doğrusal karma modellerde aykırı değerlerin tespitine ilişkin ölçütlere yer verilmiştir. Birçok istatistikî analiz yöntemi normal dağılım varsayımına ya da merkezi limit teoremine dayandırılrsa da gerçek hayatta ele alınan verilerin normal dağılıma uyması ya da aykırı değer içermemesi neredeyse mümkün değildir. Yapılan çalışmalar göstermektedir ki aykırı değer sonuçları çarpıtarak yanlışlığa ya da güvensiz sonuçlara sebep olabilmektedir. Doğrusal karma modelde sağlam tahmin ediciler, veride aykırı değer olduğunda ya da veri herhangi bir dağılıma uyum göstermediğinde kullanılabilir.

Tez çalışmasında ilk olarak özellikle küçük örneklemelerde aykırı değer olduğunda iyi sonuçlar veren sağlam tahmin yöntemleri kullanılarak doğrusal karma modelde sabit etkili parametreler, rasgele etkili parametreler ve varyans bileşenleri tahmin edilmiştir. Bu amaçla literatürde yer alan küçük veri seti tezde anlatılan sağlam tahmin yöntemi ile analiz edilmiştir. Analiz sonucunda küçük örneklemelerde aykırı değerlerin varlığında klasik yöntem ile elde edilen hata varyansı tahmininin sağlam yöntem ile elde edilen hata varyansı tahmininde oldukça büyük olduğu görülmüştür. Bunun nedeni küçük örneklemelerde aykırı değerlerin parametre tahminlerini önemli ölçüde değiştirmesidir. Bu durumda, aykırı değerleri çıkarmak yerine aykırı değerlerin parametre tahminleri üzerindeki etkisini azaltan sağlam tahmin yöntemlerinin kullanılması önerilmektedir.

Tez çalışmasında, sağlam tahmin yöntemlerinin küçük örneklerde aykırı değerlerin varlığında daha güvenilir sonuçlar verdiği görülmüştür.

Bu tez çalışmasında Orman Genel Müdürlüğü tarafından yürütülen Türkiye Milli Ağaç Islahı Programı'nda elde edilen veriler kullanılarak ıslah programının konusu olan aile seçimi BLUP ve sağlam BLUP yöntemi kullanılarak yapılmıştır.

Islah çalışmalarında yetiştirme ortamı verimliliği ile büyüme arasında önemli bir ilişki olduğu, ıslah edilmiş ağaç seçimi ile birlikte elde edilen verimliliğin çok daha fazla arttığı yapılan önceki çalışmalarda gösterilmiştir [7] [9].

Bu nedenle ıslah değerlerinin doğru tahmini ağaçlandırma çalışmaları için oldukça önemlidir. Bu tez çalışmasında amaç Türkiye Milli Ağaç Islahı Programı tarafından yürütülen ıslah çalışmalarında veride aykırı değer olduğunda varyans parametrelerinin ve ıslah değerlerinin sağlam BLUP tahminlerinin elde edilmesidir. Bu amaçla ilk olarak veride aykırı değer olup olmadığı doğrusal karma model aykırı değer ölçütleri kullanılarak araştırılmıştır. İncelenen veride aykırı değerlerin olduğu görülmüştür. Aykırı değerler olduğunda daha güvenilir sonuçlar veren sağlam BLUP yöntemi ile veriler analiz edilmiştir.

Elde edilen sonuçlar incelendiğinde sağlam tahmin ediciler ile hata varyansının küçülmesinin yanı sıra elde edilen ıslah değerlerinde ilk 10 aile sıralamasında değişimler meydana gelmiştir.

Tez çalışmasında, belli periyotlar ile Türkiye Milli Ağaç Islahı Programı tarafından tekrarlanan ıslah çalışmalarında verilerin analizinde kullanılan doğrusal karma modelde aykırı değerlerin doğrusal karma model ölçütleri ile belirlenebileceği ve verilerde aykırı değerlerin tespit edilmesi durumunda BLUP yöntemine alternatif olarak daha güvenilir sonuçlar veren sağlam BLUP yönteminin de kullanılabileceği önerilmektedir.

Bu tez çalışmasında incelenen sağlam tahmin yöntemleri, R paket programında uygulanabilir olması [6] ve bu sayede kullanıcı kolaylığı sağlaması, sonuçlar elde edilmeden önce aykırı değerleri çıkarma zorunluluğu bulunmaması, verilerin normal

dağılıma uygun olması varsayımı gerektirmemesi ve her doğrusal karma model için kullanılabilir olması nedeniyle verilerin analizinde tercih edilmiştir.

KAYNAKLAR

- [1] Ö. Kuran, Linear Karma Modellerde Parametre Tahmini ve Sonuç Çıkarım, Doktora Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana, 2017.
- [2] S.R. Searle, C.E. McCulloch, G. Casella, Variance Components, John Wiley & Sons, New Jersey, **1992**.
- [3] P.S. Gill, "A Robust Mixed Linear Model Analysis For Longitudinal Data." *Statistics in Medicine*, **2000**, 975-987.
- [4] M.J.T. Waterman, "Linear mixed model robust regression" Doctoral Thesis, Virginia Tech, **2002**.
- [5] Koller M., Robust Estimation of Linear Mixed Models, Doctoral Thesis, Eth Zurich, **2013**.
- [6] M. Koller, "Robustlmm: An R Package For Robust Estimation Of Linear Mixed-Effects Models", *Journal Of Statistical Software*, 75(6)(**2016**)1-24.
- [7] F. Işık, Kızılçamda Genetik Çeşitlilik, Kalıtım Derecesi ve Genetik Kazancın Belirlenmesi Teknik Bülten No:7, T.C. Orman Bakanlığı Batı Akdeniz Araştırma Enstitüsü Müdürlüğü, Antalya, **1998**, 264-6.
- [8] H. Öztürk, "Kızılçam (Pinus Brutia Ten.) Tohum Bahçelerinde Yer Alan Klonların Islah Değerlerinin Açık Tozlaşma Döl Denemeleri ile Belirlenmesi", Doktora Tezi, Akdeniz Üniversitesi Fen Bilimleri Enstitüsü, **2003**.
- [9] H. Öztürk, S. Şıklar, M. Alan, T. Ezen, B. Korkmaz, A.G. Gülbaba, R. Sabuncu, M. Tulukçu, S.I. Derilgen, Akdeniz Bölgesi Alçak Islah Zonunda (0-400 m) Kızılçam (Pinus brutia Ten.) Döl Denemeleri (4. Yaş Sonuçları), Çevre ve Orman Bakanlığı Orman Ağaçları ve Tohumları Islah Araştırma Müdürlüğü Teknik Bülten No:12, Ankara, **2004**.
- [10] İ. Ercanlı, H. Yavuz, A. Kahrıman. "Ormancılıkta Artım Ve Büyümenin Modellenmesinde Yeni Bir Regresyon Analizi Yaklaşımı: Karışık Model Eşitlikleri", *I. Ulusal Akdeniz Orman ve Çevre Sempozyumu* (**2011**)26-28.
- [11] S. De-Miguel, G. Guzmán, T. Pukkala, "A Comparison Of Fixed-And Mixed-Effects Modeling In Tree Growth And Yield Prediction Of An İndigenous Neotropical Species (Centrolobium Tomentosum) In A Plantation System.", *Forest Ecology And Management*, 291(**2013**)249-258.
- [12] H. Bui Manh, "Applying Linear Mixed Model (Lmm) To Analyze Forestry Data, Checking Autocorrelation and Random Effects, Using R.", **2017**.
- [13] Z. Temesgen, J. Galpin, Influence Diagnostics for Linear Mixed Models:South Africa: *Journal of Data Science* 3.2, **2005**, 153-177.

- [14] B. Dağ, Karışık Doğrusal Modellerde Kestirimler, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen bilimleri Enstitüsü, Ankara, **2006**.
- [15] F. Can, Doğrusal Karma Modellerde Ridge ve Kısıtlı Tahmin Metotları: Bir Monte Carlo Simülasyon Çalışması, Çukurova Üniversitesi Fen bilimleri Enstitüsü, Adana, **2015**.
- [16] S. Türkan, Karışık Doğrusal Modellerde Artık Analizi ve Etki Analizi, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen bilimleri Enstitüsü, Ankara, **2008**.
- [17] H.O. Hartley, J. Rao, Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model, *Biometrika* 54.1-2, **1967**, 93-108.
- [18] R.L. Anderson, T.A. Bancroft, *Statistical Theory in Research*, **1952**.
- [19] S.T. Russel, R.A. Bradley, One-way variances in a Two-way Classification, *Biometrika* 45.1-2, **1958**, 111-129.
- [20] J.E.S. Thampson, G.E. Stuart, *A Catalog of Maya Hieroglyphs*, Norman: University of Oklahoma Press, 1962.
- [21] H. D. Patterson, R. Thompson, Recovery of inter-block information when block sizes are unequal, *Biometrika*, 58.3(**1971**), 545-554.
- [22] R.D. Anderson, *Studies on the estimation of variance components*, Cornell University, **1978**.
- [23] R. I. Sampson, P. F. Jennrich, Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation, *Technometrics*, **1976**, 11-17.
- [24] B. Everitt, *Introduction to Optimization Methods and their Application in Statistics*, Springer Science & Business Media, **1987**.
- [25] R.A. Maronna, R.D. Martin, V.J. Yohai, *Robust Statistics Theory and Methods*, John Wiley & Sons, England, **2006**.
- [26] P.J. Huber, *Robust Statistics (Vol. 523)*, John Wiley & Sons, **1981**.
- [27] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, *Robust statistics: the approach based on influence functions (Vol. 196)*, John Wiley&Sons, **2011**.
- [28] P. Rousseeuw, A. Leroy,. *Robust Regression and Outlier Detection*. New York, **1987**.
- [29] R.G. Staudte, S.J. Sheather, *Robust Estimation and Testing*, John Wiley & Sons, **1990**.
- [30] P. J. Rousseeuw, Tutorial to Robust Statistics, *Journal of Chemometrics*, **1991**.
- [31] S. Chatterjee, A.S. Hadi, *Sensitivity Analysis in Linear Regression*, John Wiley&Sons, **1988**.

- [32] R.E. Welsch, E. Kuh, Linear Regression Diagnostics, National Bureau of Economic Research, **1977**.
- [33] R.D. Cook, Detection of Influential Observation in Linear Regression, *Technometrics*, 19.1(**1977**)15-18.
- [34] R.D. Cook, S. Weisberg, Residuals and Influence in Regression, New York, Chapman and Hall, **1982**.
- [35] D.C. Montgomery, E.A. Peck, Introduction to Linear Regression Analysis, John Wiley&Sons, **2012**.
- [36] J.O. Rawlings, S.G. Pantula, D. A. Dickey, Applied Regression Analysis: A Research Tool, Springer, New York, **1998**.
- [37] J. Fox, Applied regression analysis, linear models, and related methods, Sage Publications, Thousand Oaks, 1997.
- [38] D.A. Belsley, E. Kuh, R.E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, **1980**, 6-84.
- [39] O. Schabenberger, Mixed model influence diagnostics, In *SUGI*, **2005**, 189-29.
- [40] M. Candan, A. Orsoy, Doğrusal Regresyonda Sağlam Tahmin Ediciler ve Bir Uygulama: Anadolu Üniversitesi Bilim ve Teknoloji Dergisi - A : Uygulamalı Bilimler ve Mühendislik, **2001**, 265-270.
- [41] R.R. Wilcox,. Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy,. Los Angeles, Springer,Second Edition, **2010**.
- [42] M. Candan, Doğrusal Regresyon Çözümlemesinde Sağlam Kestiriciler, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, **1995**.
- [43] J.L. Hodges, Efficiency in Normal Samples and Tolerance of Extreme, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, **1967**. 163-186.
- [44] F.R. Hampel, Contributions to the theory of robust estimation, Doctoral Thesis, University of California, **1968**.
- [45] P.J. Huber, Minimax Aspects of Bounded-Influence Regression: Journal of the American Statistical Association, **1983**, 66-72.
- [46] P.J. Huber, E. Ronchetti, Robust Statistics, John Wiley & Sons, New York, **1981**.
- [47] R. Andersen, Modern Methods For Robust Regression, Thousand Oaks: SAGE Publications, **2008**.
- [48] C. Stuart, Robust Regression, Department of Mathematical Sciences, Durham University, **2011**.
- [49] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, United States, **1998**.

- [50] C. Chen, Robust Regression And Outlier Detection With The ROBUSTREG Procedure: Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference, **2002**, 265-27.
- [51] F.Y. Edgeworth, On Observations Relating to Several Quantities, **1887**, 279-285.
- [52] A. Toy, Sağlam Regresyon Tahmin Edicilerinin İncelenmesi ve Bir Uygulama, Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ, **2014**.
- [53] P.J. Huber, Robust estimation of a Location Parameter, Ann. Math. Statist, **1964**, 73-101.
- [54] P. Rousseeuw, V. Yohai, Robust Regression By Means Of S-estimators, Robust and Nonlinear Time Series Analysis, Springer, New York, **1984**, 256-272.
- [55] R. A. Maronna, V. J. Yohai, Correcting MM Estimates For “Fat” Data Sets: Computational Statistics & Data Analysis, **2010**, 3168-3173.
- [56] M. Koller, W.A. Stahel, Sharpening Wald-type Inference in Robust Regression for Small Samples, Switzerland, **2011**.
- [57] M. Koller, S. Werner, Sharpening Wald-type Inference in Robust Regression for Small Samples, Computational Statistics & Data Analysis, Switzerland, **2011**, 2504-2515.
- [58] W. Stahel, Estimation of a covariance matrix with location: Asymptotic formulas and optimal b-robust estimators. Journal of multivariate analysis, **1987**, 296-312.
- [59] W. Stahel, A. Welsh, Approaches to Robust Estimation in the Simplest Variance Components Model: Journal of Statistical Planning and Inference, **1997**, 295-319.
- [60] R.J. Beckman, C.J. Nachtsheim, R.D. Cook, Diagnostics for Mixed Model Analysis of Variance. Technometrics 29, 413-426, **1987**.
- [61] Orman Atlası, Orman ve Su İşleri Bakanlığı Orman Genel Müdürlüğü, Ankara **2013**.
- [62] Orman Genel Müdürlüğü, <https://ortohum.ogm.gov.tr/Sayfalar/Tohum-Kaynaklari-Basmuhendisligi>. (Erişim Tarihi: 12 Ağustos 2020).
- [63] A.Ö. Üçler, <http://www.ktu.edu.tr/dosyalar>, (Erişim Tarihi: 12 Ağustos 2020).
- [64] İ. Turna, http://www.ktu.edu.tr/dosyalar/silvikultur_ab27d.pdf, (Erişim Tarihi: 12 Ağustos 2020).
- [65] W.E. Nyquist, Estimation Of Heritability And Prediction Of Selection Response In Plant Populations: Critical Reviews in Plant Sciences, 235-322, **1991**.
- [66] W.M. Schutz, C. C. Cockerham, The Effect of Field Blocking on Gain From Selection: Biometrics, **1966**, 843-863.

- [68] R.W. Stonecypher, The Loblolly Pine Heritability Study. Technical Bulletin No:5. Southlands Experiment Forest, Bainbridge, Georgia, **1966**.
- [67] F. Işık, J. Holland, M. Christian, Genetic Data Analysis for Plant and Animal Breeding, Springer, North Carolina, **2017**.
- [69] B. Zolbert, J. Talbert, Applied Forest Tree Improvement, John Wiley&Sons, New York, **1984**, 505.
- [70] R.W. Stonecypher, The Loblolly Pine Heritability Study. Technical Bulletin No 5. Southlands Experiment Forest, Bainbridge, Georgia, **1966**, 128 pp.
- [71] N. Akman, Hayvan Islahı Hayvan Yetiştirme, Ankara, **1993**.

EK F – R Programı Kodları

```
veri=read.csv(file.choose(), header=TRUE, sep=";")

veri

attach(veri)

install.packages("lme4")

library(lme4)

install.packages("robustlmm")

library(robustlmm)

mixedmodel=lmer(BOY~(1|BLOK)+(1|BLOK:SET)+(1|AILE),data=veri)

summary(mixedmodel)

robustmixedmodel=rlmer(BOYY~1+(1|BLOK)+(1|BLOK:SET)+(1|AILE),data=veri)

summary(robustmixedmodel)

veriblup=read.csv(file.choose(), header=TRUE, sep=";")

veriblup

attach(veriblup)

mixedmodelblup=lmer(BOY~BLOK+(BLOK/SET)+(1|AILE),data=veriblup)

summary(mixedmodelblup)

ranef(mixedmodelblup)

fixef(mixedmodelblup)

robustmixedmodelblup=rlmer(BOY~BLOK+BLOK:SET+(1|AILE),data=veriblup)

plot(robustmixedmodelblup)

summary(robustmixedmodelblup)

ranef(robustmixedmodelblup)

fixef(robustmixedmodelblup)
```